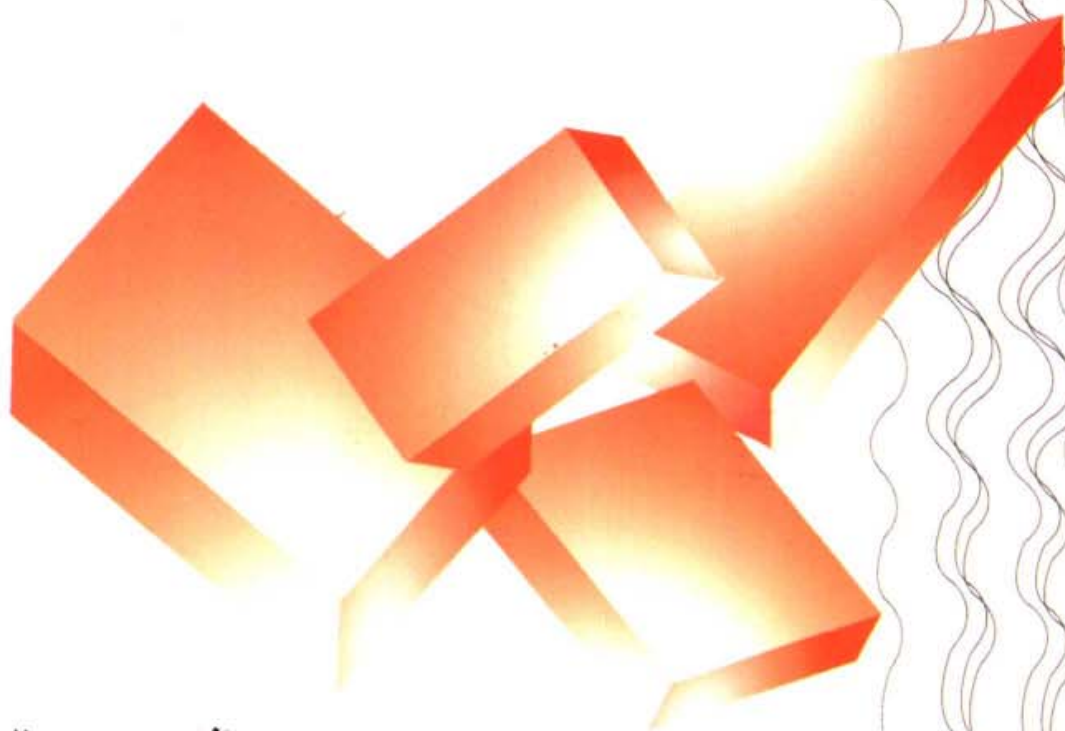


An Introduction to Statistical Methods  
and Data Analysis

# 统计学方法与 数据分析引论 (上)

〔美〕 R.L. 奥特 / M. 朗格内克 著

张忠占 等 译



科学出版社

[www.sciencep.com](http://www.sciencep.com)

An Introduction to Statistical Methods  
and Data Analysis

# 统计学方法与 数据分析引论 (下)

〔美〕 R.L. 奥特 / M. 朗格内克 著

张忠占 等 译



 科学出版社  
[www.sciencep.com](http://www.sciencep.com)



(O-1684.0101)

# 统计学方法与 数据分析引论

An Introduction to Statistical  
Methods and Data Analysis

ISBN 7-03-010815-9



ISBN 7-03-010815-9

定价: 118.00 元 (上、下册)

# 统计学方法与数据分析 引 论

(原书第 5 版)

(上 册)

[美] R. L. 奥特 著  
M. 朗格内克

张忠占 王建稳 译  
王 强 杨中华  
张忠占 校

科 学 出 版 社

北 京



# 统计学方法与数据分析 引 论

(原书第 5 版)

(下 册)

[美] R. L. 奥特 著  
M. 朗格内克

张忠占 王建稳 译  
王 强 杨中华  
张忠占 校

科 学 出 版 社

北 京

## 内 容 简 介

本书内容分为八个部分,共20章,分上、下两册,每册10章。各章均有大量习题。本书给出了大量的实际例子,这些例子涉及众多的学科和实际领域,但又不过于专门,容易理解。在大部分章节中都使用实例来引入主题,并把统计概念和这些非常实际的问题联系在一起进行讲解,深入浅出,从而可以避免许多人对统计所抱有的粗浅的感性认识,即认为统计仅仅是另一门数学课程。作者把统计数据的收集与分析过程总结成“四步法”,并把“四步法”的讲解贯穿始终,利用实例逐步展开并阐明在设计调查研究或试验时所需要的统计技术和思路,然后讲解用直观、有效的“四步法”来收集并分析数据,非常利于初学者和实际工作人员抓住有关统计方法和模型的本质。书中提供了多种多样的图示,如正态概率图、盒形图、散点图、矩阵图和残差图等,通过这些图,读者可以一方面理解数据的特点和概括数据的方法,一方面进一步理解有关统计方法的基本思想和特点。作者很重视统计在解决实际问题中的作用,在全书中用许多篇幅讨论如何解释数据分析的结果,并专门用一章讲述了如何写数据分析报告。

本书适用于作为我国文科各专业的统计学引论教程,以及理工科各专业应用统计学课程的教材或教学参考书;也可作为有关方面实际工作人员的统计入门书。阅读本书不需要其他统计方面的基础,也不需要高等数学知识。

图字:01-2002-1678号

图书在版编目(CIP)数据

统计学方法与数据分析引论(原书第5版)/〔美〕奥特(R. Lyman Ott)、〔美〕朗格内克(Michael Longnecker)著;张忠占等译. —北京:科学出版社,2003

ISBN 7-03-010815-9

书名原文:An Introduction to Statistical Methods and Data Analysis

I. 统… II. ①奥…②张… III. ①统计-方法 ②统计分析 IV. C8

中国版本图书馆CIP数据核字(2002)第079445号

责任编辑:杨 波、刘晓炜/责任校对:柏连海

责任印制:安春生/封面设计:耕者工作室

First published by Duxbury Press, a division of Thomson Learning.

All Rights Reserved

Authorized Translation Adaptation of the edition by Thomson Learning and SP.

No part of this book may be reproduced in any form without the express written permission of Thomson Learning and SP.

THOMSON

http://www.thomsonlearning.com

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

2003年6月第 一 版 开本:B5(720×1000)

2003年6月第一次印刷 印张:37

印数:1—3 000 字数:709 000

定价:118.00元(上、下册)

(如有印装质量问题,我社负责调换(路通))



## 序 言

《统计学方法与数据分析引论》(原书第五版)是为各种学科的高年级本科生和研究生而写的一本教科书,目的在于为学生解决研究项目中遇到的问题、基于数据做出决策以及理解课堂和大学环境以外的一般生活经验打下一个基础。我们假定使用本书的学生具有最少的数学知识(高中代数),而且没有修过统计的先行课程。本书前11章包括统计引论课程中的典型内容。更进一步,我们给出了把统计概念和非常实际的问题联系在一起的案例和例子。其余的章节包括回归模型和试验设计。我们在这些章节中逐步展开并阐明在设计调查研究或试验时所需要的统计技术和思路,然后用直观、有效的四步法来收集并分析数据。

### 主要特点

**从数据中学习** 我们通过使用从数据中获取知识的“四步法”来学习统计:

1. 设计数据收集过程;
2. 准备用于分析的数据;
3. 分析数据;
4. 解释数据分析的结果。

**案例** 为了展示统计在解决客观实际问题时的适用性和重要性,我们在本书大部分章节中都使用案例来引入主题。这样做的目的是为了克服许多人对统计所抱有的粗浅的感性认识,即认为统计仅仅是另一门数学课程。通过案例来介绍主题可以让读者着眼于应用统计的本质特点。我们希望这些各种各样的、关于探索以及与商业问题有关的例子会激发读者广泛地运用统计和我们发现的“四步法”的热情,这个“四步法”在我们多年的教学、咨询及研究与开发管理中一直使用着。下面是本书通篇使用的几个典型案例。

- **脂肪中卡路里的百分比** 评估和确定一个人饮食中的各种含量,对评价饮食和疾病之间的相关程度是很关键的。
- **评价财产评估者之间的一致性** 确定各县的财产评估员在财产评估中是否有系统差别的一项研究。
- **激光治疗葡萄酒色痣的效果与时间的关系** 一种判定在年轻时治疗是否比在年老时治疗效果好的前瞻性研究。
- **评价石油泄漏对植物生长的影响** 研究被泄漏的石油污染的沼泽在清除

后一年,植物生长所受的影响。

对于每一案例,如何使用四步法学习数据处理。许多案例研究还包含有样本量确定、数据图示的讨论,以及一份完整的统计研究结果的报告的要点摘要。

### 数据的图示

整个课本中,我们提供了多种多样的图示,这对评价关键的假定是十分必要的,正是这些关键的假定构成了运用如正态概率图、盒形图、散点图、矩阵图和残差图等统计方法的基础。更进一步,我们强调用图示法概括数据,是为了给读者提供工具,利用它们读者可以说明各种处理之间的差异。用这些图还可以把统计上不同的处理与实际中的不同处理区别开来。例如,在“面试人员的决定受到不同类型身体障碍的影响吗”的案例中,我们用盒形图来概括评价的差异。正态概率图用于检验残差正态性的假定。这里所示的图出现在图 9.1 和图 9.2 中。

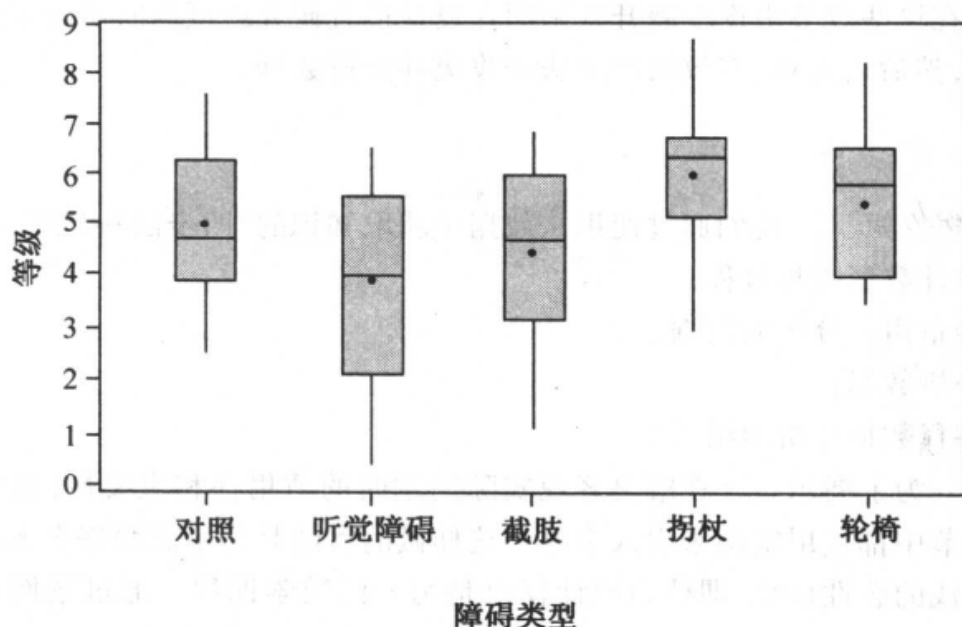


图 9.1 对不同障碍等级的盒形图(实心圆点表示均值)

### 例题和习题

我们利用出自期刊论文、报纸及我们多次咨询经历的例题和习题来进一步强调统计学的实践特性。这些例题和习题进一步向学生说明了统计学方法在解决跟他们日常生活相关的实际问题中的应用。这一版中包括许多新的例题和习题,丰富多样的练习题无论对于教师还是学生都将成为学习统计的巨大财富。在许多练习题中,我们给出了一些计算机输出的结果,以供学生做练习时使用。例如,练习 9.7 中给出由 SAS 输出的方差分析表,用四种不同方法对均值进行区分的结果以及残差分析的结果,然后向学生提出研究人员在试图概括研究结果时会问的各种



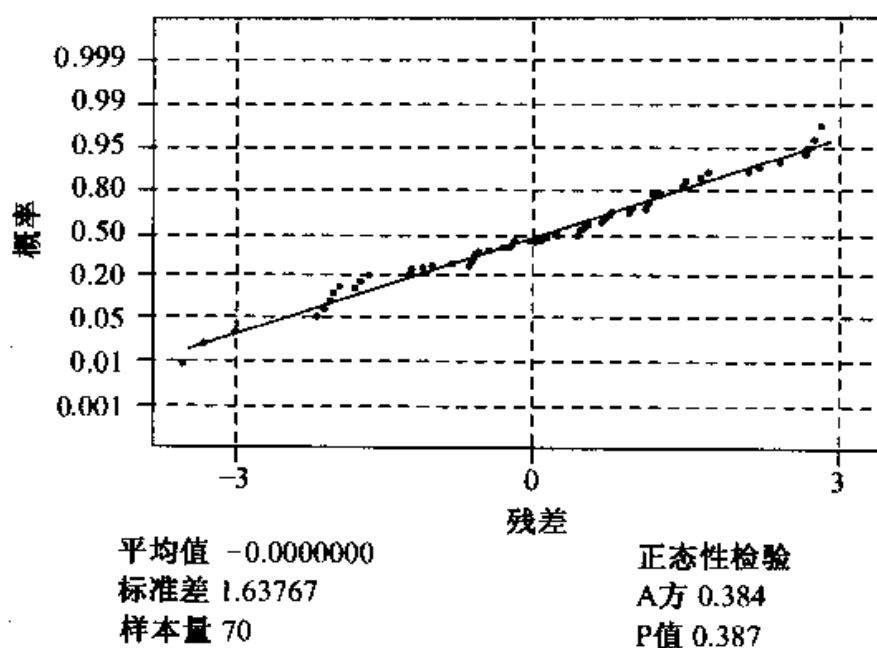


图 9.2 残差的正态概率图

问题。

### 包含的主题

本书分成八部分,适用于一学期或两学期课程的学习。书中八个部分的内容与从数据中学习的各个步骤及全书 20 章之间的联系参见表 1.1。从第一部分到第五部分及第十一章和第二十章构成一学期课时的内容:

**第一部分:** 引论

**第二部分:** 收集数据

**第三部分:** 概括数据

**第四部分:** 工具和概念

**第五部分:** 数据分析

**第十一章:** 线性回归和相关

**第二十章:** 分析结果的传达和备案

两学期课时的第二学期包括建模、多元回归分析推断、Logistic 回归、试验设计和方差分析:

**第六部分:** 分析数据: 回归方法和模型的建立

**第七部分:** 试验设计与方差分析

## 着重解释而不是计算

在初等统计课程中,标准的做法是定义一个统计方法,给出该统计方法的计算形式,然后利用这个计算形式进行数据分析。然而,我们发现这样做对于大型数据的分析十分不利。进一步说,学生按照这种统计方法的计算形式进行手头计算要花费太多时间,而计算形式本身却不能从直观上表明为何如此计算。所以,我们提供例题和练习,以便学生能够学习如何依据统计方法的定义形式去计算统计估计量和检验统计量的值。当学生对该统计方法所反映的数据的特征熟悉以后,我们再强调在分析大型数据集时,运用计算机软件做计算的好处。我们给出了一些主要统计软件包的输出结果:SAS, Minitab, Stata, Systat, JMP 和 SPSS。我们发现这样做使学生有了用定义计算统计方法中各量值的体会,从而学到了每一统计方法背后的基本知识。在统计课程以外大多数的场合下,学生应利用计算机软件进行计算,不仅方便而且计算准确。在许多练习和例题中,利用计算机使得我们有更多的时间用于强调对计算结果的解释,而不是把大量的时间和精力花费在实际的计算上。

通过例子和习题,我们揭示了假设检验以下诸方面的重要性:

1. 假设的陈述,即通过概括研究者的目标形成关于总体参数的假设;
2. 选择最合适的检验统计量,包括许多统计方法中的样本量计算;
3. 在讨论统计假设检验的结果时,考虑第 I 类和第 II 类错误( $\alpha$  和  $\beta$ )的必要性;
4. 既考虑检验结果的统计意义也考虑其实际意义,即我们阐明了估计效用大小和建立总体参数的置信区间的重要性;
5. 除“拒绝  $H_0$ ”和“不能拒绝  $H_0$ ”外,用非统计方面的专业术语来陈述统计结果。

我们用醒目的方式排印有关一个假设检验的内容,包括原假设、备择假设、检验统计量、拒绝域及使得此假设检验有效的条件。当假设检验所需要的条件没有得到满足时,则建议可供选择的其它检验统计量。例如,第五章中有如下所示的醒目的排印:

### 正态总体( $\sigma$ 已知)或大样本容量 $n$ 时 $\mu$ 的统计检验

假设:

情形 1.  $H_0: \mu \leq \mu_0$      $H_a: \mu > \mu_0$  (右侧检验)

情形 2.  $H_0: \mu \geq \mu_0$      $H_a: \mu < \mu_0$  (左侧检验)

情形 3.  $H_0: \mu = \mu_0$      $H_a: \mu \neq \mu_0$  (双边检验)

T.S. (检验统计量):  $Z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}$

R.R: 对于犯第一类错误的概率  $\alpha$



情形 1. 如果  $z \geq z_\alpha$  则拒绝  $H_0$

情形 2. 如果  $z \leq -z_\alpha$  则拒绝  $H_0$

情形 3. 如果  $|z| \geq z_{\alpha/2}$  则拒绝  $H_0$

注意: 如果总体是  $\sigma$  已知的正态分布, 则这些方法是适用的。在大多数情况下, 如果  $n \geq 30$ , 由中心极限定理知, 在总体非正态分布时, 也可用这些方法。同样, 如果  $n \geq 30$ , 那么, 我们可以用样本标准差  $s$  来代替  $\sigma$ 。  $n < 30$  的情形在这章的后面再讨论。

### 新版的變化

- 大多数章节中的案例帮助学生理解应用统计在解决实际问题时所扮演的角色。侧重点放在阐明从数据处理中学习的步骤。
- 书中强调了统计方法所基于的假定的重要性。我们通过计算机模拟来讨论和说明许多估计和检验方法是稳健的或不稳健的。在许多例题和练习中, 我们给出了一些方法以说明如何考察一项研究是否满足这些必要的假定。另外, 在许多问题中, 我们还给出了当这种假定不满足时可供选择的其他方法。
- 强调从练习题和例题所包含的研究中学习解释结果和总结结论。
- 在用小数据集讲清了如何通过确定的方式来进行相应的统计计算后, 我们鼓励大家利用计算机来做大部分计算。
- 大部分例题都有数据的图示, 计算机的使用使得我们利用复杂的图形工具来描述统计结果极为方便。
- 提供了样本量及  $t$  检验和  $F$  检验中犯第二类错误的概率的计算技巧。
- 我们扩充和更新了练习题, 而且习题和例题涉及众多学科, 其中包括许多实用的、现实生活中的问题。这些练习可为学生运用求知于数据的四个步骤积累一些经验。
- 在有关试验和研究的设计的新一章中, 讨论了一些在数据收集之前需要考虑的重要因素。我们强调了如何防止在设计阶段因没考虑所有因素而导致不能回答研究者所关心的重要问题这种事情的发生。
- 关于线性回归和相关分析的两章已被合并成一章。
- 尽管把表达和撰写数据分析结果放在最后一章, 但我们在全书中利用案例使其中许多思想得以体现。
- 网站: <http://www.Suxbury.com> 上有所有习题的数据集。
- 精选练习题的答案同样也可在该网站得到。

### 前几版中保留下来的一些特点

- 书中包含了统计方法和数据分析的许多实际应用,这些应用来自农业、商业、经济、教育、工程、医药、法律、政治学、心理学、环境研究及社会学。
- 练习题被分成基本技巧型和应用型两组。
- 在大多数章节后都有供复习用的练习题。
- 我们在大量例子和习题中给出了 Minitab, SAS, Systat, JMP, Stata 和 SPSS 等统计分析软件的计算机输出结果。
- 注意基本的假定。用图示方法和检验方法判断假定是否成立。
- 第一章讨论了“什么是统计”,指出了学生为什么要学习统计,讨论了几种重要的调查问题,以说明统计在解决现实问题中的应用。
- 花一定篇幅讨论了如何进行数据整理和为数据分析做准备。

### 辅助材料

- 学生解题手册(0-534-37123-X)包含了书中问题的部分解答。
- 解题手册(0-534-37121-3)为教师提供了书中所有问题的答案。
- 试题库(0-534-37122-1)包含了书中所讲内容的一些试题。
- Ott/Longnecker 网站资源中心包括了数据集,勘误表和供教师和学生用的其他资源。要进入资源中心,请到 [www.Duxbury.com](http://www.Duxbury.com), 选择“Online Book Companions.”

### 致谢

在原稿付梓和随后几版的编写过程中,许多人提出了宝贵的、有建设性的建议。Duxbury 出版社的编辑 Carolyn Creckett 在该版写作的整个过程中一直给予我们巨大的鼓励。同时,我们也深深的感谢乔治亚大学的 Chris Franklin,感谢她关于这版所需变动的内容作出的有创见的评论以及她对各章修订稿的审阅。我非常感谢以下评论者富有洞察力的、建设性的评论,他们是 Deborah J. Rumsey(俄亥俄州州立大学), Larry J. Ringer(得克萨斯州 A&M 大学), Mosuk Chow(宾夕法尼亚州州立大学), Christine Franklin(乔治亚大学)和 Darcy P. Mays(弗吉尼亚 Commonwealth 大学)。我们要特别感谢 Felicita Longnecker, Michael 之妻在准备初稿素材、校订及打印该版初稿方面所给予的帮助。

R.L. 奥特  
M. 朗格内克

# 目 录

## 上 册

### 序言

## 第一部分 引 论

第一章 什么是统计 .....	( 3 )
1.1 引言 .....	( 3 )
1.2 为什么学习统计 .....	( 7 )
1.3 当前统计的一些应用 .....	( 7 )
1.4 统计学家做什么 .....	( 11 )
1.5 质量和工序改进 .....	( 12 )
1.6 学生注意 .....	( 14 )
1.7 小结 .....	( 15 )
补充练习 .....	( 15 )

## 第二部分 收集数据

第二章 利用调查和科学研究来收集数据 .....	( 19 )
2.1 引言 .....	( 19 )
2.2 调查 .....	( 20 )
2.3 科学研究 .....	( 28 )
2.4 观察研究 .....	( 35 )
2.5 数据整理:为概括和分析准备数据 .....	( 36 )
2.6 小结 .....	( 40 )

## 第三部分 概括数据

第三章 数据的描述 .....	( 43 )
3.1 引言 .....	( 43 )
3.2 计算器、计算机及软件系统 .....	( 44 )

3.3	单个变量数据的描述:图表法 .....	( 46 )
3.4	单个变量数据的描述:中心趋势的度量 .....	( 76 )
3.5	单个变量数据的描述:变异性度量 .....	( 88 )
3.6	盒形图 .....	( 103 )
3.7	多变量数据的概括 .....	( 109 )
3.8	小结 .....	( 119 )
	重要公式 .....	( 120 )
	补充练习 .....	( 120 )

## 第四部分 工具和概念

第四章	概率和概率分布 .....	( 135 )
4.1	如何应用概率进行推断 .....	( 135 )
4.2	确定一个事件的概率 .....	( 138 )
4.3	基本的事件关系和概率法则 .....	( 141 )
4.4	条件概率和独立性 .....	( 144 )
4.5	Bayes 公式 .....	( 150 )
4.6	离散变量和连续变量 .....	( 155 )
4.7	离散随机变量的概率分布 .....	( 157 )
4.8	一个常用的离散随机变量:二项分布 .....	( 158 )
4.9	连续随机变量的概率分布 .....	( 169 )
4.10	一个常用的连续随机变量:正态分布 .....	( 172 )
4.11	随机抽样 .....	( 184 )
4.12	抽样分布 .....	( 189 )
4.13	二项分布的正态逼近 .....	( 201 )
4.14	Minitab 指令 .....	( 204 )
4.15	小结 .....	( 205 )
	重要公式 .....	( 206 )
	补充练习 .....	( 206 )

## 第五部分 数据分析:中心值,方差和比例

第五章	关于总体中心值的推断 .....	( 215 )
5.1	引言和案例 .....	( 215 )
5.2	$\mu$ 的估计 .....	( 218 )
5.3	估计 $\mu$ 时样本容量的选取 .....	( 228 )
5.4	关于 $\mu$ 的统计检验 .....	( 231 )
5.5	对于 $\mu$ 进行检验时样本容量的选取 .....	( 242 )

5.6	统计检验的显著性水平 .....	( 248 )
5.7	正态总体均值 $\mu$ 的统计推断, $\sigma$ 未知 .....	( 253 )
5.8	关于中位数的推断 .....	( 267 )
5.9	小结 .....	( 276 )
	重要公式 .....	( 277 )
	补充练习 .....	( 278 )
第六章	两总体中心值的比较 .....	( 290 )
6.1	引言和案例 .....	( 290 )
6.2	关于 $\mu_1 - \mu_2$ 的推断: 独立样本 .....	( 294 )
6.3	非参数推断方法: Wilcoxon 秩和检验 .....	( 315 )
6.4	关于 $\mu_1 - \mu_2$ 的推断: 成对数据 .....	( 328 )
6.5	非参数推断方法: Wilcoxon 符号秩检验 .....	( 336 )
6.6	推断 $\mu_1 - \mu_2$ 时样本容量的选取 .....	( 343 )
6.7	小结 .....	( 345 )
	重要公式 .....	( 346 )
	补充练习 .....	( 348 )
第七章	关于总体方差的推断 .....	( 371 )
7.1	引言和案例 .....	( 371 )
7.2	单个总体方差的估计和检验 .....	( 374 )
7.3	比较两个总体方差时的估计和检验 .....	( 386 )
7.4	比较多个总体方差时的检验 .....	( 397 )
7.5	小结 .....	( 406 )
	重要公式 .....	( 406 )
	补充练习 .....	( 407 )
第八章	两个以上总体的中心值的推断 .....	( 413 )
8.1	引言和案例 .....	( 413 )
8.2	两个以上总体均值的统计检验: 方差分析 .....	( 418 )
8.3	完全随机化设计中观测值的模型 .....	( 428 )
8.4	方差分析条件的检查 .....	( 431 )
8.5	其他的分析方法: 数据变换 .....	( 439 )
8.6	另一种非参数方法: Kruskal-Wallis 检验 .....	( 446 )
8.7	小结 .....	( 451 )
	重要公式 .....	( 452 )
	补充练习 .....	( 453 )
第九章	多重比较 .....	( 465 )
9.1	引言和案例 .....	( 465 )
9.2	线性对照 .....	( 469 )
9.3	控制哪个错误率 .....	( 476 )

---

9.4	Fisher(费舍尔)最小显著差异法 .....	( 479 )
9.5	Tukey 的 W 方法 .....	( 483 )
9.6	Student-Newman-Keuls 方法 .....	( 486 )
9.7	Dunnett 方法;处理组与对照组的比较 .....	( 489 )
9.8	Scheffé 的 S 方法 .....	( 491 )
9.9	小结 .....	( 498 )
	重要公式 .....	( 499 )
	补充练习 .....	( 499 )
第十章	类型数据 .....	( 511 )
10.1	引言和案例 .....	( 511 )
10.2	总体比例 $\pi$ 的推断 .....	( 513 )
10.3	两总体比例之差 $\pi_1 - \pi_2$ 的推断 .....	( 525 )
10.4	多比例的推断;卡方拟合优度检验 .....	( 532 )
10.5	Poisson(泊松)分布 .....	( 541 )
10.6	列联表;独立性检验和齐性检验 .....	( 546 )
10.7	相关程度的度量 .....	( 557 )
10.8	几率和优比 .....	( 563 )
10.9	小结 .....	( 566 )
	重要公式 .....	( 567 )
	补充练习 .....	( 568 )

# 目 录

## 下 册

### 第六部分 数据分析:回归方法和模型的建立

第十一章 线性回归和相关 .....	( 583 )
11.1 引言和案例 .....	( 583 )
11.2 估计模型中的参数 .....	( 592 )
11.3 回归参数的推断 .....	( 612 )
11.4 利用回归预测新的 $y$ 值 .....	( 623 )
11.5 线性回归中拟合不足的考察 .....	( 633 )
11.6 逆回归问题(校准) .....	( 640 )
11.7 相关 .....	( 649 )
11.8 小结 .....	( 660 )
重要公式 .....	( 661 )
补充练习 .....	( 663 )
第十二章 多元回归与一般线性模型 .....	( 679 )
12.1 引言和案例 .....	( 679 )
12.2 一般线性模型 .....	( 688 )
12.3 估计多元回归系数 .....	( 690 )
12.4 多元回归中的推断 .....	( 713 )
12.5 回归系数子集的检验 .....	( 726 )
12.6 用多元回归进行的预测 .....	( 736 )
12.7 比较几条回归线的斜率 .....	( 741 )
12.8 Logistic 回归 .....	( 747 )
12.9 多元回归的一些理论结果(任选) .....	( 756 )
12.10 小结 .....	( 759 )
重要公式 .....	( 760 )
补充练习 .....	( 761 )
第十三章 多元回归续论 .....	( 781 )



13.1 引言和案例 .....	( 781 )
13.2 变量的挑选(第一步) .....	( 784 )
13.3 模型形式的确定(第二步) .....	( 806 )
13.4 模型假设的检查(第三步) .....	( 840 )
13.5 小结 .....	( 866 )
重要公式 .....	( 867 )
补充练习 .....	( 867 )

## 第七部分 试验设计与方差分析

第十四章 试验和研究的设计概念 .....	( 923 )
14.1 引言 .....	( 923 )
14.2 研究的类型 .....	( 924 )
14.3 设计的试验:术语 .....	( 925 )
14.4 控制试验误差 .....	( 929 )
14.5 试验单元对处理的随机化 .....	( 934 )
14.6 确定重复试验的次数 .....	( 938 )
14.7 小结 .....	( 942 )
第十五章 标准设计的方差分析 .....	( 948 )
15.1 引言和案例 .....	( 948 )
15.2 单因子的完全随机化设计 .....	( 951 )
15.3 随机化完全区组设计 .....	( 955 )
15.4 拉丁方设计 .....	( 974 )
15.5 完全随机化设计中的因子处理结构 .....	( 988 )
15.6 随机化完全区组设计中的因子处理结构 .....	( 1014 )
15.7 处理差异的估计和处理均值的比较 .....	( 1016 )
15.8 小结 .....	( 1023 )
重要公式 .....	( 1023 )
补充练习 .....	( 1024 )
第十六章 协方差分析 .....	( 1048 )
16.1 引言和案例 .....	( 1048 )
16.2 具有一个协变量的完全随机化设计 .....	( 1051 )
16.3 外推问题 .....	( 1064 )
16.4 多维协变量和更复杂的设计 .....	( 1068 )
16.5 小结 .....	( 1077 )
补充练习 .....	( 1077 )
第十七章 一些固定效应、随机效应和混合效应模型的方差分析 .....	( 1083 )
17.1 引言和案例 .....	( 1083 )

17.2	具有随机处理效应的单因子试验:随机效应模型 .....	(1086)
17.3	随机效应模型的扩充 .....	(1091)
17.4	混合效应模型 .....	(1100)
17.5	计算期望均方的规则 .....	(1110)
17.6	套抽样和裂区设计 .....	(1121)
17.7	小结 .....	(1132)
	补充练习 .....	(1132)
第十八章	重复测量与交叉设计 .....	(1139)
18.1	引言和案例 .....	(1139)
18.2	有重复观测的单因子试验 .....	(1144)
18.3	一个因子有重复观测的两因子试验 .....	(1146)
18.4	交叉设计 .....	(1157)
18.5	小结 .....	(1161)
	补充练习 .....	(1161)
第十九章	一些非平衡设计的方差分析 .....	(1168)
19.1	引言和案例 .....	(1168)
19.2	有一个或多个缺失观察值的随机化区组设计 .....	(1170)
19.3	有缺失数据的拉丁方设计 .....	(1176)
19.4	平衡不完全区组(BIB)设计 .....	(1181)
19.5	小结 .....	(1191)
	重要公式 .....	(1192)
	补充练习 .....	(1193)
第二十章	分析结果的传达和备案 .....	(1198)
20.1	引言 .....	(1198)
20.2	做好传达沟通工作所面临的困难 .....	(1198)
20.3	传达的障碍:图形的歪曲 .....	(1200)
20.4	传达的障碍:有偏抽样 .....	(1203)
20.5	传达的障碍:样本容量 .....	(1205)
20.6	为统计分析准备数据 .....	(1206)
20.7	统计分析的指导原则和报告 .....	(1209)
20.8	文档和结果的保存 .....	(1210)
20.9	小结 .....	(1211)
	补充练习 .....	(1211)
附录	统计表 .....	(1212)
	参考文献 .....	(1281)
	索引 .....	(1286)
	译后记 .....	(1303)

# 第一部分 引 论

## 第一章 什么是统计



# 第一章 什么是统计

- 1.1 引言
- 1.2 为什么学习统计
- 1.3 当前统计的一些应用
- 1.4 统计学家做什么
- 1.5 质量和工序改进
- 1.6 学生注意
- 1.7 小结

## 1.1 引言

什么是统计？它是数字的加法吗？它是图表、击球手的平均得分、保送上垒的成功率、失业率吗？总之，它是不是对社会和自然的数值描述？

统计是一套科学原理和技术，用于在可得到的信息既有限又富于变化时，从中得出关于总体的和过程的结论。也就是说，统计是关于从数据中学习的科学。几乎每个人，包括社团主席、市场代理、社会科学家、工程师、医学研究人员、消费者，都与数据打交道。这些数据以各种形式出现，如季销售量、青少年犯罪的增长率、水样的污染等级、经过某种治疗后病人的存活率、人口普查数据，或用来帮助决定购买哪种品牌的汽车的数据等。在本书中，我们通过考虑从数据中获取知识的四步来学习统计。这四步是：(1)设计收集数据的过程；(2)准备分析用的数据(概括、建模)；(3)分析数据；(4)报告从数据分析中得到的结论。

本书分为八个部分。这八个部分的内容与从数据中学习的各个步骤及全书各章之间的联系如表 1.1 所示。从表中可见，我们花费许多时间讨论如何用基本的方法(如中心值、方差和比例)、回归方法和方差分析方法来分析数据。然而，你必须记住，对于我们要分析的每一数据集合，已经有人给出了一种收集数据的方案(第 1 步)。在准备好分析用的数据(第 2 步)和分析完数据(第 3 步)后，必须有人把分析的结果以书面或口头的形式传达给有关人员。所有这四步对从数据中获取知识都很重要。事实上，如果第一步执行不当，试验或研究的目标往往会因为收集的数据集不完全或含有不恰当的信息而实现不了。自始至终，我们将使你集中于从数据获取知识的大图景上。当然不时的查阅下面这张表将有助于你清楚每章在整个安排中所处的位置。

表 1.1 本书的编排结构

书中各个部分	从数据获取知识的步骤	书的章节
1. 引论		1. 什么是统计?
2. 收集数据	1	2. 利用调查和科学研究来收集数据
3. 概括数据	2	3. 数据的描述
4. 工具和概念		4. 概率和概率分布
5. 数据分析:中心值、方差和比例	3	5. 关于总体中心值的推断 6. 两总体中心值的比较 7. 总体方差的推断 8. 两个以上总体的中心值的推断 9. 多重比较 10. 类型数据
6. 数据分析:回归方法和模型的建立	3	11. 线性回归和相关 12. 多元回归和一般线性模型 13. 用诊断建立回归模型
7. 试验设计和方差分析	3	14. 试验和研究的设计概念 15. 标准设计的方差分析 16. 协方差分析 17. 某些固定、随机、及混合效应模型的方差分析 18. 重复测量和交叉设计 19. 一些非平衡设计的方差分析
8. 表达和撰写分析结果	4	20. 分析结果的传达和备案

在进入统计的学习之前,让我们考虑四个例子。在这些例子中,运用统计能帮助我们解决实际问题。

1. 一灯泡厂每天大约生产 50 万只灯泡,质量控制部门必须检验灯泡的次品率。这个任务可以通过检验每一只灯泡来完成,但是这样做的花费巨大而且会造成每只灯泡价格的猛涨。另一方法是从每天生产的 50 万只灯泡中选出 1 000 只,然后检验这 1 000 只灯泡。如果这 1 000 只灯泡是以

正确的方式被选出的,那么从中检测出的次品比例,可被用于估计全天所有产品的次品比例。我们将在以后的章节中论证被检查的灯泡的次品率可能与全天生产的 50 万只灯泡的次品率很接近。

2. 为了调查研究“人在戒烟后体重会增加”这一断言,研究人员选择了一个由 400 个参与者构成的样本,他们都成功地参与了戒烟运动。每个人在活动开始前和一年后都称量了体重。参与者体重的平均变化是增加了 5 磅。研究人员由此总结说有证据表明这一断言是正确的。我们将在后续章节介绍一些方法,并用之说明这个变化是真由于戒烟造成的而不是由于随机因素造成的。
3. 为了研究氮肥对小麦产量的影响,研究人员一共可以利用 15 块田地。他随机地选择 3 块田分配给五种被考察的氮肥中的一种。在所有这 15 块田中种上同种小麦。所有这些田都以同样的方式耕作,直到收获的季节,记录下这 15 块田的亩产量。试验者想决定任何麦田的最佳氮肥施肥量,但显然他只能在有限的一些田块中做试验。当得出使试验田有最大产量所需的氮肥量后,研究人员断言,这个施肥量对于与试验田具有相同特征的麦田也适用。试验者做出这样的结论有足够的理由吗?
4. 纽约时报(New York Times)/哥伦比亚广播公司新闻(CBS News)、华盛顿邮报(Washington Post)/全美广播公司新闻(ABC News)、有线新闻网(CNN)、哈里斯和盖洛普(Gallup)民意调查机构等频繁地使用统计,也使我们想起统计的类似应用。民意调查人员如何判断 19 500 多万具有选举权的美国人的意见?他们当然不会接触每一个潜在的美国选民。相反,他们调查一小部分,比如 1 500 个潜在选民的意見,然后据此估计整个国家所有具有选举权的人的意见。令人惊讶的是,被访人中持有特定观点的人的比例与整个总体在那个特殊时刻持有相同观点的人的比例会很接近。在以后的章节中我们将提供令人信服的有力的证据来支持这个论断。

这些问题都说明了从数据中学习的四个步骤。首先每个问题都涉及设计一个试验或调查。质量控制小组必须决定需要检测多少灯泡,如何从所有的产品中选出 1,000 个灯泡样本以得到正确的结果。民意调查组织必须决定为得到能够代表全部选民的信息,须调查多少选民以及如何挑选这些选民。类似地,在体重是否会增加的研究中必须认真设计需要选择多少参与者,如何从所有这样的参与者中选择这些人。另外,研究人员应该测量每一参与者的什么量?是不是有必要知道每一参与者的年龄,性别,健康状况及其他与健康相关的量,还是只测量体重?如果被研究的参与者中许多人都有特殊的健康状况,则研究结果可能就不适用于一般



总体。在小麦试验中,为了得到可以推广到试验田以外的结果,测量田块的土壤特征和环境条件(如温度和降雨量)很重要。

最后,收集、概括、分析完数据后,用清楚的措辞向有关的人报告你的结果也是很重要的。对于灯泡的例子,管理人员和技术人员需要知道他们成批产品的质量状况。基于这些信息,他们将决定是否有必要对生产做出相应的调整。因而,统计分析的结果不可以用模棱两可的术语表达,必须基于已知的知识做出判决。戒烟后是否增重的研究结果,对那些有病人参与戒烟运动的医生来说是感兴趣的。如果纪录表明那些戒烟的个人体重明显增加,那么医生可能不得不向他们提出饮食建议,免得以前吸烟的人从一种健康问题(吸烟)转向另一健康问题(由于过重而血压升高)。在报告中有关参与者的仔细描述如年龄、性别和其他与健康有关的信息是很关键的。在小麦试验中,试验者要给农民提供一些可以帮助他们选择对自己的田块最经济、最佳的施氮肥量的信息。所以,报告中必须包括有关他们所研究的田地的湿度和土壤类型的信息。否则,有关最优小麦产量的结论将可能无益于那些在截然不同的条件下种植小麦的农民。

为了使调查结果能够适用于比被调查的部分更大的群体,我们必须仔细地定义**总体**的概念(参见定义 1.1),把问题归结为对于这个总体的推断问题,设计一个调查并依之从该总体中恰当地选择**样本**(参见定义 1.2)。我们将在第二章讨论这些问题。

**定义 1.1** 总体是样本收集人员感兴趣的所有测量值的集合(见图 1.1)。

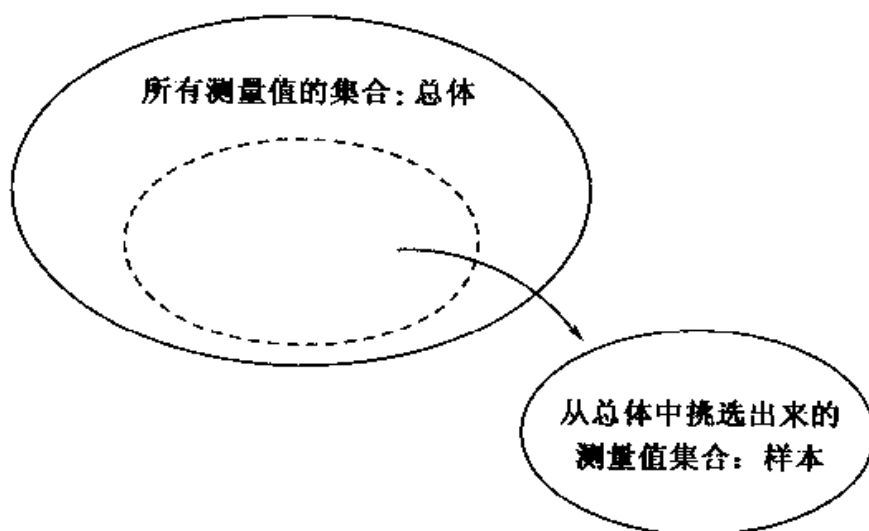


图 1.1 样本和总体

**定义 1.2** 样本是选自总体的任何子集(见图 1.1)。

## 1.2 为什么学习统计

我们可以为学习统计入门课程想出两条好的理由。一条是你需要知道如何去评价公开发表的用数字说明的事实。我们的周围充斥着生产厂家对其产品的声明;充斥着有关社会问题的、消费的、政治的民意测验结果;充斥着出版了的科研结果。其中许多结果都是基于样本做出的推断,有些推断是正确的,有些是不正确的;有些是基于足够大的样本量,有些不是。然而,所有这些出版的结论都带着真理的光环。有些人(尤其是统计学家)说统计可用于支持几乎任何东西,也有一些人说用统计很容易造假。两种说法都对。在把抽样的结论应用于没有获得信息的范围时,很容易有意或无意地利用统计歪曲事实。

学习统计的第二条理由,是你的职业或工作可能要求你解释抽样的结果(如调查或试验),或者运用统计分析法做出推断。例如,开业医生得到许多广告,描述新药的优点。这些广告经常给出新药和旧药比较试验的数据结果。这样的数据是否真的意味着新药更有效,结果中观察到的差别是不是仅仅是试验中测量值的随机变动引起的呢?

在法庭审判中,最近的趋势表明,在评定证据的真实性时运用概率统计推断的方法在增加。在社会、生物、物理科学中运用统计是必然的,因为这些科学都要利用抽样调查或试验获得的对自然现象的观察来发展和检验新的理论。统计方法在商业中的运用体现在用抽样数据来预测销售额和利润。另外,统计还用于工程和制造业的产品质量监督。账目的抽样还可以支持主管会计进行账目的稽查。由此可以看出,统计在科学、商业、工业等几乎所有领域中都扮演着重要的角色。在这些领域中的工作人员需要了解统计的基本概念、威力和局限性。

## 1.3 当前统计的一些应用

### 酸雨:对我们环境的一种威胁

普遍接受的酸雨的成因是硫酸和硝酸。雨中这些酸性成分的来源是碳氢化合物燃料,当它们燃烧后把硫、氮氧化物释放到大气中去。酸雨有许多危害,下面是其中的几种:

- 当酸雨出现在春天雪融期时,会侵入许多鱼类的繁殖地区阻碍成功繁殖。许多依赖池塘、湖泊的生物种类由于池塘或湖泊受到酸雨污染而逐渐消失。
- 在森林中,酸雨会削弱一些树种的抵抗力,使他们更容易遭受昆虫的侵害和罹病。

- 在被受酸雨影响的水包围的地方,主要养分将从土壤中分离出去。
- 人工建筑物也受到酸雨的影响。美国的专家估计酸雨给房屋或其他建筑物造成的损失至今已达 15 亿美元左右。

解决与酸雨相关的问题并不容易。美国国家科学基金会(NSF)建议我们努力减少 50% 的氧化硫排放量。说来容易做起来难。因为含硫量高的煤是主要的排放源,但在一个依靠煤提供能源的国家转用含硫量低的煤并不总是可能的。相反,必须发明更好的过滤装置,从燃烧过程中把这些有污染的氧化物在它们排入大气之前除掉。内燃机所用的油料也是酸雨中硫氧化物的主要来源。显然,需要对汽车和卡车的排放进行更好的控制。

要减少燃煤锅炉和内燃机交通工具氧化物的排放量,需要大量使用现有的过滤器和排放控制装置,并开发新的技术使我们能利用可利用的能源。开发可替代的、更清洁的能源对我们达到 NSF 提出的预定目标也是很重要的。统计和统计学家将在监察大气状况,检验排放控制装置的有效性,以及发展新的控制技术和替代能源中担当举足轻重的角色。

### 判定新药的疗效

预防小儿麻痹症的脊髓灰质炎疫苗的发明和检验为说明统计是如何被用于解决实际问题提供了一个极好的例子。大多数父母和 1954 年以前出生的孩子都能想起那个夏天由于小儿麻痹症爆发而导致的恐慌。尽管每年患小儿麻痹症的孩子相对较少,但是由于该病的爆发不可预测而且由于该病会导致瘫痪甚至是死亡而被引起极大的关注。如今的年轻人很少有人听说过小儿麻痹症,这一事实表明疫苗及其投放市场之前所做的检验工作的巨大成功。

确定一种药品疗效的通常做法是在人身上进行试验(常称为临床试验)。对于某些临床试验,药物随机地分配给参与试验的人,其中一半人真正获得药品,其余一半人获得不含该药的溶液或药片(称为安慰剂)。统计问题之一是确定临床试验参与者的总数。这一问题在脊髓灰质炎疫苗检验试验中尤为重要,因为以往年份的数据表明小儿麻痹症的发病率为每 10 万个孩子中不足 50 例。这样,为了测出接受疫苗的人和得到安慰剂的人之间发病率的差异,临床试验必须包括大量的参与者。

在统计学家的帮助下,确定该疫苗的临床试验需要共 40 万个孩子参与。以往从没尝试做过由如此多人参与的临床试验。试验从 1954 年开始。在这些参与者中随机地抽取了一半人分配了疫苗,其余的人给以安慰剂。通过公立学校的接种活动,对这 40 万参与者进行了处理(接种疫苗或安慰剂),然后经过一个夏天的观察得出患小儿麻痹症的人数。尽管临床试验中的 40 万人只出现了不到 200 例的小儿麻痹症,但其中获得安慰剂的患病人数是获得疫苗的患病人数的 3 倍还要多。这些结果加上一些统计计算,足以说明疫苗的效果。然而,如果没有统计学家和科

学家计划和指导这样庞大的临床试验,那么,这样的结果是不可能得到的。

脊髓灰质炎疫苗的开发并不是统计应用于检验和研制药品的特例。近年来,美国食品和药品管理署(FDA)一直严格地要求药厂建立新药品的疗效试验说明。因而,统计在避孕药、风疹疫苗、癌症的化疗剂和其他医药、食品的研制和检验中都发挥着重要的作用。

### 统计在法庭上的应用

我们中的每一个人都能碰到与消费品有关的投诉事件,你可能在这样的事件中充当原告或被告,或者你可能听说某人被牵涉进这样的官司。当然我们都间接地通过增加了的保险费和商品的费用帮助支付这种官司的费用。投诉中有关特定商品(如汽车、药品等)的证词常在很大程度上依赖于对一个或几个有关该产品调查得到的数据的解释,这就是为什么统计和统计学家进入法庭的原因。

例如,流行病学家已经把统计概念用于数据来判断一种特征(如硅酮乳房移植中的泄露)和疾病(如自体免疫疾病)之间是否有某种统计关联。发现这种关联的流行病学家还得努力辨别研究所观察到的统计联系是由于随机变动引起的,还是确实反映了特征与疾病间的真正联系。法庭论据中有关对这类联系的解释涉及用统计概念分析数据和数据的临床解释。当然,还有许多其他在法院案例中利用统计模型的例子。在薪水歧视的案例中,诉讼状声称某雇主基于年龄、种族、性别少付钱给雇员。统计模型被用于解释基于许多因素如工作经验、所受教育程度和工作表现等造成的工资差异。然后,利用根据工作表现等相关因素调整后的工资数据在年龄组或各种族之间进行比较,以决定这些调整后的工资数据之间是否还存在显著的差异。

### 估计弓头鲸的总数

Raftery 和 Zeh(1998)讨论了有关弓头鲸(Bowhead whale)的总数和增长率的问题。这种研究之所以重要是由于弓头鲸是已经禁止商用捕捉的第一种巨鲸;它们的状况对于其他巨鲸的恢复前景具有启示作用。此外,国际鲸调查委员会利用这一估计决定阿拉斯加的爱斯基摩人维持土著生计所需的捕鲸配额。为了获得必要的的数据,研究人员利用视觉和声学法在阿拉斯加的巴鲁角(Point Barrow)外进行了普查。然后,把统计模型和估计方法用于普查所得的数据,由此判断在商用捕鲸停止后,弓头鲸是增加了还是减少了。统计估计表明弓头鲸的总数以正常的比例增长。这一结果说明,因商用需要而捕捉十分之一后余下的巨鲸将会在捕捉终止后恢复正常。

### 臭氧暴露和人口密度

城市周围地区臭氧的污染是每个国家最普遍的环境问题。虽然同温层臭氧的减少可能导致皮肤癌的增发,但周围环境中高浓度的臭氧也会造成对人类呼吸系统、农作物和树木的破坏。得克萨斯州的休斯敦是仅次于洛杉矶的第二大臭氧浓

度超过国家周边环境空气质量标准(National Ambient Air Quality Standard)的地区。Carroll 等人(1997)描述了如何分析休斯敦的臭氧时测数据,这些数据是由设在休斯顿的 9 至 12 个监测站在 1980~1993 年间收集来的。除了臭氧浓度外,每一监测站还记录了另外三个气象变量:温度、风速及风向。

这一项目在统计方面有三个主要目标:

1. 提供有关缺失数据的总量和模式的信息(且/或提供获得这种信息的工具),同时提供有关臭氧水平和三个气象变量的测量质量的信息。
2. 建立一个有关臭氧强度的模型,由此来预测休斯顿地区 1980 年到 1993 间任何时刻任何给定的地区的臭氧浓度。
3. 运用此模型来估计描述长期暴露或短期高浓度暴露的暴露指数。同时,把普查信息和不同的暴露指数相结合得到人口总体的暴露指数。

研究人员建立的空间—时间模型所提供的估计表明,最高浓度的臭氧水平发生在儿童人口相对较少的地方。还有,据该模型估计,从 1980 到 1993 年儿童在臭氧中的暴露减少了约 20%。所得到的这一人口总体暴露的分布蕴涵着一些方针政策上的解释。具体说来,由此得出,如果想评估人口总体的暴露情况,那么,现今监测站的安置并不理想。这一项目涉及从数据中获取知识的所有四个组成部分:设计监测站应建立在城中的何处,决定以何种频繁的程度来采集数据以及需要记录什么量;描出数据的空间—时间图形;创建臭氧数据、气象数据和人口数据的空间—时间模型;最后,写一份报告以能够帮助地区和联邦官员制定关于降低臭氧浓度的政策。

### 民意测验和偏好调查

公众意见、消费者的偏好及关于选举的民意测验常被用于评估一部分公民对某问题、某产品或感兴趣的候选人等的观点和偏爱。美国公众每天都能从报纸、杂志、收音机、电视中听到或看到这些民意测验的结果。例如,每隔两天当地报纸就登出有关以下问题的调查结果:

- 消费者对美国经济未来前景的信心。
- 对即将到来的选举和政党预备会议中候选人的偏爱。
- 对联邦所得税返还中的欺骗行为的态度。
- 对指定产品偏爱的调查(例如,其他国家的汽车对美国汽车;可口可乐(Coke)对百事可乐(Pepsi);麦当劳(McDonald)对 Wendy 快餐。
- 北卡罗来纳(North Carolina)居民对有关烟草道德争论的反映。
- 投票者对增税提案和国防预算变动提案的观点。

关于民意测验可以提出大量的问题。假定我们考虑一个民意测验,以了解公众对于提高密歇根州所得税的动议的意见。*什么是民意调查员所关心的总体?* 民意调查员对密歇根的所有居民都感兴趣,还是仅对那些近来缴纳所得税的公民感

兴趣?样本真是抽自这个总体吗?如果所关心的总体是所有近来缴纳所得税的人,民意调查员是否确信抽样得到的个人都是近来纳税的人?问什么样的问题,提问时如何措辞?是否问每个人同样的问题?是不是问题的表达方式会使回答产生偏差?我们能否相信这些民意测验的结果?这些结果代表一般公众当前对调查问题的认识吗?

民意测验和偏好调查对消费者来说是重要的、看得见的应用统计的实例。我们将在第十章中更详细地讨论这个主题。我们希望等学完这些内容后,你能更好地理解这些民意测验的结果。

## 1.4 统计学家做什么

统计学家是做什么的?就从数据中学习而言,统计学家要进行试验或调查的设计,用图表或数据概括准备分析用的数据、分析数据及报告分析结果等所有方面的工作。收集数据的方法有好有坏。统计学家运用他们现有的有关调查技术和科学调查设计的知识,或者开发新的技术为收集数据提供好的指导方法。我们将在第二章中进一步探讨这些方面。

一旦数据收集完后,必须先对数据进行概括,然后才可能做出有意义的解释。统计学家可以建议或运用有效的方法,以图表、表格、数字等形式来概括数据。精心设计的图、表是了解数据的第一步。同时,平均值(或典型的值)和一些反映数据的极差或分散性特征指标的量对于解释也很有帮助。这些内容将在第三章中详细讨论。

*统计学的目标在于基于从感兴趣的总体抽得的样本的测量信息对该总体做出推断。*在从数据中学习的四个阶段中,推断是在分析这一阶段中进行的。例如,市场调研中只对很少一部分新产品的潜在购买者进行调查,但是,所有潜在购买者(总体)的可能反应必须从被调查的潜在购买者(样本)的反应中推断出来。如果该市场调研被认真地设计并很好地执行了,那么样本中那些人的反应应该与总体的反应吻合得较好(但是未必严格一致)。之所以如此,是因为概率的基本概念允许我们对感兴趣的总体做出推断,这一推断包含我们对总体的最佳推测以及最佳推测可能存在的误差。

我们通过一个例子来说明如何作出推断。假定查账人从一个有2.5万多笔财务账目的集合中随机地选取了2,000笔,发现84(占4.2%)份有错。对这25,000多笔账目的集合我们能说什么?基于2,000笔账目样本得到的信息,我们对25,000笔账目的出错率能作出怎样的推断?我们将在第十章说明对账目总体出错率的最佳推测是4.2%(推断),并且这个最佳推测值介于总体账目的实际出错率(未知) $\pm 0.9\%$ 之间。这个加减因子称为推断的可能误差。任何人都可以对账目

出错率作出推测,而概率的概念允许我们计算出这种推测可能出现的误差。

在进行数据分析时,统计工作者可能利用现有的方法作出推断,而有些理论统计工作者运用更高等的数学和概率的理论来发展新的方法。我们将在第四章讨论有关概率的基本概念和抽样分布,然后从第五章开始学习分析样本数据的方法。

最后,统计工作者要报告分析结果,这是理解数据的最后一个阶段。分析结果报告的形式从非正式的谈话到正式的报告多种多样。在一个比较正式的用可视化工具等来做的口头讲解或一份研究报告中,可以利用图形、表格、数据图表、以及对数据所做的分析来帮助传达在数据中所发现的内容;而在非正式的交谈中往往就做不到这一点。报告或交流应该注意向感兴趣的听众传达从样本数据所能获取的总体的信息,并且应尽可能地用非技术性术语来表达以避免造成含糊不清。有关如何报告、解释分析结果的更详细的内容在第二十章中给出。在一些章节中,结合阐明统计概念的实例,我们也将指出分析报告中所应包括的重要组成部分。

值得指出的是,前面所讨论的思想观点跟每一个搞调查研究和试验的人都是密切相关的。具有统计学学位的人相对说是较少的,许多组织机构中只有很少几个统计工作者,有的甚至根本没有。这样,在许多研究中数据收集方案的设计,数据的概括和统计分析,解释研究结果等都要由参与研究的人员在没有或几乎没有真正的统计学家帮助的情况下而完成。即使统计工作者是研究小组中的活跃分子,其他的小组成员具有统计设计和数据分析概念的一般知识仍很重要。事实上,小组中每一成员都把某一方面的专业知识和经验带进所面对的问题中去,因而,在如何进行调查的设计以及如何来解释分析结果方面,可以群策群力。

## 1.5 质量和工序改进

看到这个题目,或许有人会疑惑,我们在统计课本中怎么会提到质量和工序改进的问题呢?这样做是为了让读者了解商业和科学领域中一些涉及求知于数据的更广泛的内容。

第二次世界大战后,美国商业和经济在世界商业中处于支配地位,而且这一状况一直持续了将近 30 年。那时,人们很少试图去改变做事的方式,主要精力集中于大规模作业和完善批量生产。然而,从 20 世纪 70 年代中期至今,许多企业都必须面对来自日本,最近还有来自远东国家,如中国和韩国等同类企业的激烈竞争。

质量而非数量已经成为消费者购买商品时主要考虑的标准,美国企业为了调整经营方式以适应这一变化而经历了一段艰难的时刻。除非美国企业积极地改变他们的企业经营方式,否则,在这场质量革命中将会有更多的企业遭遇不幸。

日本人是第一个懂得质量至上这一理念的。他们积极地采纳 Deming(1981)等人提出的统计质量管理和工序控制的建议,并且启动了全面质量管理的程序。



整个组织,从上层管理部门到下层质量管理部门,他们都有一个提高产品质量、改进生产工序的约定。他们从不满足于现状,不断寻求更好、更新的方法。

许多美国公司现在开始在质量改进机构的协助下朝着好的方向发展。以下十项基本要求是成功地进行质量改进的基础:

#### 成功实施质量改进过程的基本要求

1. 以顾客为中心是该过程的最重要部分。
2. 使质量改进过程作为管理系统的一部分成为管理部门的长期责任。
3. 相信总有改进的余地。
4. 相信防止问题发生总比问题发生后再改正要好。
5. 管理部门重视、领导并参与。
6. 把不出任何错误作为工作标准(目标)。
7. 全体员工或作为个人,或组成小组积极参与。
8. 专注于改进工序,而不是改造人。
9. 相信如果供应厂商理解你的需要,他们定会跟你合作。
10. 表彰成功。

让全公司上下都齐心协力参与质量改进过程旨在于改进工序。多年来,大公司们为了扩大和改进生产,通常的做法是通过增加员工或引进技术来加速生产过程,而没有察觉到工作流程中的不足之处。在 Michael Hammer 及 James Champy 所著的具有重大突破的 *Reengineering the Corporation* (1995) 一书及 Hammer 后来的 *Beyond Reengineering* (1996) 一书中, Hammer 和 Champy 指出,许多公司的商业过程在激烈的商业和技术环境变化中一直保持不变;一个公司可以通过完全重新思考他们的商业过程来实现质量、效率、效益的巨大提高。我们把一个任务定义为一个工作单元、把一个工序定义为一系列为顾客创造价值的相关任务,而 Hammer 和 Champy 则建议公司把他们谋求变化的注意力集中于创造价值的活动上。

事实胜于雄辩。几乎所有大企业——服装业(如 Nike, 耐克), 化工业(如 Dupont, 杜邦), 计算机设备企业(如 Dell, 戴尔), 计算机软件业(如 Microsoft, 微软), 电子业(如 General Electric, 通用电器), 食品业(如 Nestlé, 雀巢), 日用商品业(如 WalMart), 网络通讯业(如 Cisco), 石油业(如 Exxon Mobil), 医药业(如 Merck) 等等, 这些企业中像部门领导这样有竞争力的职位已经、正在或即将受到挑战。在许多例子中, 企业领导未能跟上市场中令人眼花缭乱的变化步伐。随着管理人员和股东们对兼并能带来市场份额的增加、成本的降低、利润的增加抱有很高的期望, 企业间的兼并激增。然而不幸的是, 成功实现合并的既定目标的企业少得可怜。

由此可见, 还应考虑其他的一些东西。Christopher Meyer 在他著的 *Fast Cycle*

Time(1993)中举例说明了在瞬息万变的市場条件下,只有那些能够“持续地、可靠地和能获利地”向消费者提供最大价值的竞争者才能获胜。Meyer的基本观点是公司必须缩短整个商业循环周期,这个循环始于对消费者需求的识别,终止于消费者对公司提供的产品或服务的支付。如果一个公司能够随着需求和竞争环境的变化而始终做好这一环节,它就会在竞争中获胜。

不管公司是集中于改进商业过程还是加快循环,变化的依据将是有关消费者需求的基本数据、当前企业的循环时间,以及企业中可用来比较的基准数据。那些明确他们正在努力做什么,建立当前所需数据以用来评价消费者需求和当前公司的运作状况,迅速及时地实现被推荐的变革并且对经验教训进行整理备案的企业能够在竞争中成为胜者。这四点,与我们前面章节中讨论的从数据中学习的四个步骤非常类似,能使人理解统计(从数据中学习)在商业环境中的适用性。这儿列出了许多有助于改进这些商业过程的统计工具和技术。

#### 用于质量改进和再造工程的统计工具、技术和方法

- 直方图
- 数值描述量(均值、标准差、比例等)
- 散点图
- 线图(在散点图中用线连接各点)
- 控制图:  $\bar{y}$  (样本均值),  $r$  (样本极差), 及  $s$  (样本标准差)
- 抽样方案
- 试验设计

此处所列出的及本书中将要讨论的统计工具和概念只是改进商业过程或加快循环时间的技术中很小的一部分。当你在本书的各个部分遇到这些工具和概念时,不要忘记它们可能对商业改进有用。质量改进、工序重新设计及加快循环时间显然已成为20世纪90年代的美国企业在以竞争加剧、更加联合、专业化程度加强为特点的世界市场中谋求发展的焦点。不管我们是作为消费者还是作为企业参与者,这些转变将会对我们所有人带来影响,所以我们了解一些作为这场革命的一部分的统计工具是有益的。最后,近年来质量控制的观点和原则已经被用于制造业以外的其他领域。服务行业如宾馆、饭店和百货商店等已成功地把质量控制的原理用于自己的经营管理中。许多联邦机构例如国内税务总局(IRS)、国防部和农业部(USDA)也已采用质量控制的原理来改善他们机构的工作。

## 1.6 学生注意

我们通过词语和概念进行思考。统计学科的学习要求我们记忆新的术语和概念(正如学习一门外语一样),务必记住这些定义、定理和概念。

与此同时,要着眼于理解数据所包含的信息这一大概概念,不要让细节模糊了研究对象的由数据信息反映出来的主要特征。本书的教学目标就在于明晰和发挥这些统计的大概念。

## 1.7 小 结

统计学科及那些运用统计工具的学科研究如何从数据中获取知识。医学研究人员、社会学家、会计、农学家、消费者、政府领导和职业统计学家都从事与数据收集、数据概括、数据分析和数据分析结果诠释等有关的工作。

补充练习

### 基本技能

**1.1 (生物)** 选择合适的饲料来喂虾或其他的海洋生物是海水养殖的一个重要方面。水产研究人员想估计喂了 6 个月指定饲料后虾的平均重量。从一个人工挖掘的鱼塘中随机地选择了 100 只虾并称了每一只的重量。

- 给出研究人员感兴趣的总体。
- 给出相应的样本。
- 研究人员对总体的什么特征感兴趣。
- 如果用样本的测量值对总体的某些特征进行推断,为什么说对推断的可靠性的度量是重要的?

**1.2 (环境)** 放射性废料的处理及一些矿业运作中放射性材料的生产给美国的一些地区造成了严重的污染问题。州卫生部官员决定对某可疑地区进行放射性水平的调查。在这个地区随机地选择了 200 个点,并测出每个点处的放射性水平。就这个抽样情况,回答练习 1.1 中 a, b, c, d 四问。

**1.3 (社会)** 某城市中的某社会研究人员想得到城市家庭中接受社会福利支持的孩子个数的有关信息。从该城市的福利卷宗中随机地抽取了 400 户家庭进行查证,得到每户家庭接受福利孩子的数量。对这个抽样调查回答练习 1.1 中 a, b, c, d 四问。

**1.4 (调查科学)** 查查你们当地报纸或新闻杂志上的一些热点问题,找出对这些问题最近民意调查的结果。

- 说出为了获得样本测量值而观察的研究对象。
- 指出对每一研究对象所进行的测量。
- 明确指出所要调查的总体。
- 调查人员对总体的哪个或哪些特征感兴趣?

- e. 调查文章有没有解释样本是如何抽取的?
- f. 调查文章有没有包括样本中测量值的数目?
- g. 关于总体特征做出了何种推断?
- h. 调查文章有没有告诉你对总体特征推断的可信度?

**1.5(政府)** 由于最近高中橄榄球运动员颈部受伤的数量有所增加,商业部设计了一个调查来估计高中橄榄球运动员戴的防护帽的强度。从五个近期生产防护帽的公司收集了共 540 个防护帽。该机构把防护帽送到一独立检测机构去评估防护帽的缓冲作用及面罩被扭转时传给颈部的冲击力。

- a. 感兴趣的总体是什么?
- b. 样本是什么?
- c. 应该测量什么量?
- d. 关于高中运动员所戴的防护帽的安全性,这项调查有哪些主要局限?例如,运动员颈部力量是否与传递到颈部的冲击力以及运动员是否会受伤有关?

**1.6(教育)** 一个拥有 35 000 名学生的大学中的教员评议会正考虑把现行的评分政策由 A, B, C, D, E 改为加减系统——即 B+, B, B- 而不仅仅是 B, 教员评议会对学生们关于这种变化的意见感兴趣,将抽取由 500 名学生组成的样本。

- a. 什么是感兴趣的总体?
- b. 什么是样本?
- c. 如何抽取样本?
- d. 调查问卷中应包括什么类型的问题?

## 第二部分 收集数据

### 第二章 利用调查和科学研究来收集数据



## 第二章 利用调查和科学研究来收集数据

### 2.1 引言

### 2.2 调查

### 2.3 科学研究

### 2.4 观察研究

### 2.5 数据整理:为概括和分析准备数据

### 2.6 小结

## 2.1 引言

正如在第一章中提到的那样,从数据中学习的第一步是仔细地思考研究的目标(先思而后行)。在*善于智慧地数据收集*中关键的一步是数据收集过程的设计。在这个过程中,需要有意地做好以下几步:

- 详细说明研究、调查或试验的目标。
- 确定所关心的变量。
- 为调查或科学研究选择适当的设计方案。
- 收集数据。

为了明确研究的目标,你必须理解所面对的问题。例如,某大城市的交通部门为了增加市内公共汽车的使用率而想评估一下公众对该市公交系统的认识。这样,交通部门需要确定公交系统的哪些方面决定着人们是否乘车。该项研究的目标是,找出那些交通部门为增加乘坐公共汽车的人数可以改变的因素。

为了确定所关心的变量,你必须仔细分析研究的目标。对于公交系统,通过参考其他城市所做的调查或和公交系统的员工集体讨论可以确定出一些主要因素。其中一些可能的因素是安全性、车费、公交车的整洁度、在家或工作地点附近是否有停靠站以及公交车晚点的频繁程度。调查中要获得的测量值包括被确定出的因素的重要程度(非常重要、重要、没意见、不太重要、很不重要)。人口统计学方面的信息如年龄、性别、收入、住址等也要测量。最后,有关一个人当前乘车频繁程度的量的测量值很重要。一旦明确了调查目标和所关心的变量,你必须选择最适当的方法去收集数据。数据收集过程包括:调查、试验和对已有的来自商业纪录、人口普查、政府记录及过去的调查记录等数据的检查。抽样调查和试验设计的理论为数据收集提供了极好的方法。调查的目标是收集有关现有条件、态度或行为的数



据。因此,上述城市的交通部门需要设计一份调查问卷,然后抽取该市中当前的公共汽车乘客及使用其他交通工具的人作为样本。

另一方面,科学研究似乎更积极主动:试验人员通过改变试验条件来研究试验条件对试验结果的影响。例如,交通部门可以通过在一些选出来的线路上降低票价来评估公共汽车的利用率是否因此而增加。然而,在此例中,一些公交系统控制不了的其他因素也可能在这期间发生变化。比如,公共汽车使用率的增加可能是由于地铁工人罢工或石油价格上涨而造成的。票价的降低只是造成乘坐公交车人数增加的几种因素中的一种因素。

在大多数科学试验中,影响测量值的尽可能多的因素都在试验者的控制之下。一个花匠想确定一种新的植物激素对商品花生长的影响。该花匠在一个温度、湿度、水分和光照都被控制好了的温室中做试验,选定几个生长激素的用量,其中包括一个对照——即不施加生长激素,然后按照每个用量,给一组有同样数量的植物施用生长激素。当试验结束时,测量出植物的大小和生长状况。由此可以决定出对植物生长最优的生长激素水平,因为在理想的状态上,所有影响植物大小和生长状况的其他因素对试验中的所有植物来说都是一样的。

在这一章,我们将考虑一些调查方法和科学研究的设计方案,也将分别对科学研究和观察研究进行讨论。

## 2.2 调 查

来自抽样调查的信息几乎影响我们日常生活的每一方面。这些调查决定了诸如经济调控和促进社会进步等的国家政策。民意测验是众多新闻媒体所报道的许多新闻的基础。电视节目的排行榜决定着将来可以看到哪些节目。

谁做调查?我们都熟悉公众的民意测验:纽约时报(*New York Times*)/哥伦比亚广播公司新闻(*CBS News*),华盛顿邮报(*Washington Post*)/全美广播公司新闻(*ABC News*),哈里斯(*Harris*),新闻周刊(*Newsweek*)的盖洛普民意测验(*Gallup*)和CNN的民意测验。然而,绝大多数调查是为了工业、政府、管理或科学等方面的特定目的而进行的。例如,汽车制造厂利用调查来找出顾客对他们生产的汽车的满意程度;我们在购买某种产品后,经常被要求填写一份调查表,以此作为获得保修登记程序的一部分;许多涉及健康问题的重要研究常用调查来完成,例如,食物中的脂肪含量、被动吸烟的情况、避孕套的使用与艾滋病的预防以及青春期抑郁症的流行等。

美国宪法规定美国人口普查局每十年普查一次人口。随着政府对居民生活介入的增长,人口普查局已经扩展它的职能,不再仅仅是计数人口数量了。他们曾经做过这样的尝试,给美国每户居民邮寄一份人口调查问卷。从1940年的人口普查

开始,在完全计数人口数量的基础上,又从国民的代表样本获得了更多的信息。在2000年的人口普查中,使用了变化抽样率。对大多数地区来说,大约六分之五的家庭被要求回答列在简表上的14个问题,其余的家庭回答包括增加了另外45个问题的一份长长的问卷。许多机构和个人把所得到的信息用于各种目的。联邦政府用它来给各州和城市分配基金;商业部门用它来预测销售量、管理员工和构思将来的销售点;城市和地方规划人员用它来计划土地的使用、交通网络的建设和能源的消耗等;社会科学家利用它来研究经济状况、种族均衡及生活质量的其他方面。

美国劳动力资源统计局(BLS)例行20多项调查,其中最著名、使用最广泛的是建立消费价格指数(CPI)的调查。消费价格指数用来测量市场上一个固定的货物和服务业范围内的价格变化。一方面它是反映通货膨胀的量,另一方面,它也是反映政府政策的一项经济指标。商业部门认为工资额度及退休金计划与CPI紧紧联系在一起。联邦卫生和福利机构以及许多州或地方部门把对他们项目适当与否的评价基础与CPI紧紧联在一起。租借和抵押价格的调整也要基于CPI。这个由抽样调查确定的指标,在我们的社会中起着十分重要的作用。

BLS所做的其他许多调查对促进社会的发展也是很重要的。每月当前人口的调查提供了有关劳动力、就业和失业等方面的基本信息。消费支出调查收集了有关家庭日常生活中用于商品和服务支出的数据。基层单位调查(The Establishment Survey)收集有关非农业商业机构的雇用时间和工资收入的信息。有关职业前景的调查提供了有关将来各种工作就业机会的信息,它可以反映未来将近十年的情况。BLS的其他一些活动可参见 *BLS Handbook of Methods* (1982)。

民意测验经常地出现在新闻中,Gallup(盖洛普)和Harris也已经成为家喻户晓的民意调查机构。这些民意测验和抽样调查反映了公民对从政治和宗教到体育和娱乐几乎所有事情的观点和态度。Nielsen收视率决定着电视节目的成败。Nielsen零售指数提供了有关食品、化妆品、药品、饮料和其他许多类产品的最新销售数据。这些数据来自全美1,600家商店每60天的销售和库存清单。

商业部门为了内部经营而进行抽样调查。另外他们在作重大管理决策时也利用政府部门的调查。审计员用抽取账目的方法来估计账上的收支平衡情况以及是否按规操作。生产过程中的质量控制很大程度上依赖于抽样技术。

商业活动中依赖于细致的抽样活动的另一领域是市场销售。决定销售什么产品,在哪儿销售及如何为它们做广告等常常都要基于抽样调查得来的数据。这些数据可能来自生产该产品的公司所做的调查或从专业从事市场数据调查的调查公司购买得来。

### 抽样的方法

在任何调查中,如何从总体中挑选样本都是一个关键的因素。如果其中个体的选择仅仅是出于方便,那么抽样调查可能会产生偏差,这将妨碍调查准确地反映

作为一个整体的总体。例如,一个市场营销学专业的研究生发明了一种做广告的新方法。为了评价这一新方法,他在一个大的本科生商学课程班中抽取学生,让他们来评价该新方法是否对标准的广告方法有所改善。该班学生的观点能代表将来接收到这一新方法广告的全体对象的观点吗?上述学生的收入水平、种族、受教育程度及其他的一些社会经济特点可能与所关心的总体有很大的区别。此外,那些学生有可能是被他们老师强迫去参加调查的,因而他们可能对调查中的问题不给以坦诚的回答。因此,抽样所用的方法对研究结果的可信度和适用程度至关重要。

**简单随机抽样** 基本的设计方法(简单随机抽样)是如此抽取  $n$  个个体组成的样本,使得总体中每组样本量为  $n$  的样本都有相同的机会被抽选出来。在民意测验中,我们从选民登记列表中按照某种使得每个样本量为  $n$  的样本有同样的概率被选中的方法选择名单,这样我们得到由合格选民组成的随机样本。简单随机抽样的详细内容在 4.11 节中讨论,这里我们只说明一点,即如果社区中所有选民具有相似的社会经济背景,那么简单随机抽样与任何其他抽样方案一样包含同样多的有关社区偏好的信息。

**分层随机样本** 然而,假定该社区由高、低两类不同收入层次的人构成。高额收入阶层中的选民对调查问题所持有的观点可能与低收入阶层中选民的观点很不相同。所以,为了得到有关总体的确切信息,我们要从每一“阶层”中抽取选民。我们可以把该总体中的元素按照收入分成两组,或者说两层,然后,从每组中抽取一个简单随机样本。这样抽得的样本称为**分层随机样本**。(参见 Scheffer 等,1996,第五章)注意,分层是利用一个辅助变量,即个人收入的知识来实现的。通过把收入分为高低两层,我们增加了估计的精度。

**比估计** 比估计是第二个利用辅助变量所含信息的方法。比估计不仅使用对所关心的响应变量的测量值,而且把它与辅助变量的测量值结合起来。比估计同样可用于分层随机抽样。

**整群抽样** 尽管我们在抽样调查中希望得到的是个人的见解,但尤其是在城市地区,抽取家庭、公寓楼或城市街区而不是抽取单个的选民可能会是一个更经济的方法。抽取完这些单元以后,在抽中的单元内我们可以得到每一位合格选民的见解。这一方法称为**整群抽样**。尽管在整群抽样和分层抽样中我们都对总体进行分组,但这是两种不同的方法。在分层随机抽样中,我们在每一组中抽取简单随机样本,而在整群抽样中,先抽取有关组的简单随机样本,然后调查选中组(群)中的所有元素(详细内容参见 Scheffer 等,1996,第八、九章)。

**系统样本** 有时,所关心的总体中人的名单可以从诸如登记表,或抽屉中的文件卡等这样的列表中得到。对这种情况,一种经济的抽样方法是在列表的起始处选择一个名字,然后依次每隔 10 或 15 个名字选一个人,从而得到一个样本。以这种方式抽取的样本我们称之为**系统样本**。正如你期望的那样,系统抽样提供了一

种获得抽样信息的便利方法。不幸的是,在一定费用的前提下,我们未必能得到最多的信息。(详见 Scheffer 等,1996,第七章)

此处要理解的重点是,有多种不同的收集样本数据的抽样调查方法。在本书中,我们将讨论简单随机抽样和由此得到的样本数据的分析和概括的方法。更复杂的抽样调查导致在统计概括和分析数据阶段更复杂的问题。

美国统计学会(ASA,网址: <http://www.amstat.org>, email: [asainfo@amstat.org](mailto:asainfo@amstat.org))出版了一系列有关抽样调查的文件: *What Is a Survey (什么是抽样调查)?*, *How to Plan a Survey (如何设计调查)*, *How to Collect Survey Data (如何收集调查数据)*, *Judging the Quality of a Survey (判定调查的质量)*, *How to Conduct Pretesting (如何进行预备调查)*, *What Are Focus Groups (什么是焦点组)?* 以及 *More About Mail Survey (有关通信调查的进一步讨论)*。这些文件描述了许多对获得有效和有用的调查起决定作用的因素,列出了许多在调查中常见错误的潜在根源,并且给出了如何避免这些错误的指导性原则。下面列出其中的一些内容。

### 与抽样调查有关的问题

即使恰当地选取了样本,调查结果是否代表了从中取样的总体仍可能并不确定。这种不确定性的两个主要来源是无回答和测量问题。无回答发生在一部分被抽中的个人不能或不愿参与调查。测量问题发生在对问卷中问题的回答不能提供设计调查时所预期得到的那类数据。

**调查中的无回答** 调查中出现无回答可能会因为样本不代表总体而使得调查产生偏差。*Judging the Quality of a Survey* 中指出,在对一般人的调查中,女人比男人更愿意参与调查,也就是说,男人的无回答率要比女人高。这样,在有关政治问题的民意调查中,如果总体中赞成某特定问题的女人的百分比大于总体中支持该问题的男人的百分比,该调查可能出现偏差。因为样本中女人的比例比总体中女人的比例大,所以这项民意调查将会过高估计总体中赞成该问题的人的比例。在所有的调查中,必须对无回答组进行详细检查,以断定是否总体的某特定部分在样本中可能有过多或过少的出现。对付无回答的措施有:

1. 为参与调查的人提供某种奖励物品以增加他们参与调查的积极性。
2. 对没有回答第一次联络的个人给以提醒或打电话跟踪采访。
3. 利用统计方法,在考虑到样本不同于总体的侧面的基础上,对调查结果予以调整。

**测量问题** 测量问题是由于回答者没有提供调查所寻求的信息而造成的。这些问题常常可归因于调查设计问题中的某些措辞、被调查者回答调查问题的方式,及在采访中采访人表达问题的方式。特殊问题的例子及可能的措施如下:

1. **回想不起问题的答案**: 被采访者被问及去年他或她参观某个城市公园的次

数。这类问题常常导致低估一个家庭在去年参观该公园的平均次数。因为人们往往低估普通事情或距采访时间很久前发生的事情的发生次数。一个可能的措施是要求被采访人在作出回答前查看原先的纪录或与家庭中其他成员商讨一下。

2. **对答案有诱导性的提问**：有关观点问题的提问方式可能会使回答并不真正的代表被访者的观点。这样，调查结果就会朝着问题本身所倾向的那个方向偏倚。例如，把有关州里是否要对违反环保规定的化工厂征收大笔罚款的问题表述为：“考虑到处罚会导致该厂转移到另一个州，而该厂是我们地区主要的就业单位，你会支持州里处罚该化工厂吗？”这类问题往往诱使人们回答：“不支持”，从而，造成对整个地区中人们关于征收化工厂罚金之观点的歪曲。预防措施是以客观的方式仔细地写出问题。
3. **问题中不明确的措辞**：某运动俱乐部试图确定人们每周锻炼的次数。问题是这样问的：“上星期你运动了多少次？”“运动”一词对不同的个人有不同的含义。允许人们对调查问卷中问题的主要词句给以不同的理解只会使得调查结果的精确性大大降低。一些可能的预防措施是：在进行调查之前，把问题在各种人中进行测试，以判断问题中是否有含糊不清或会引起误导的词句；在对采访人员进行培训的过程中，让所有采访人员知道所有关键词的“正确的”定义，并且建议他们把这些定义提供给回答者。

ASA 出版的一系列小册子中还有许多其他注意事项、问题及其对应措施。

图 2.1 中包含了一项调查的设计、实施和分析的各个阶段。该图的早期版本可从 Cryer 和 Miller(1991)著的 *What Is a Survey?* 中得到。该图表为恰当地进行成功的调查提供了指导。

### 数据收集方法

选择了某个抽样调查方法后，实际中如何来收集数据呢？抽样调查中最常用的收集数据的方法是当面采访和电话采访。如果适当地培训采访者，谨慎地计划再次访问，那么这些方法可达到 60% 至 75% 有时甚至更高的回答率。向我们感兴趣的特定的人群邮寄调查问卷也能收得到较好的结果，但是一般来说，这种收集数据的方法的回答率是如此之低，以至于所有报告出来的结果都值得怀疑。通常，直接观察比采访或邮寄问卷更容易得到客观的信息。

**当面采访** 调查数据经常通过**当面采访**得到。例如，我们可以利用对有资格选民的个人采访得到有关公众对社区公债问题看法的样本。这一程序通常要求采访者问一些预先准备好的问题并记录下回答者的回答。这类采访的最大优点是人们通常会因面对面的采访而作出回答。另外，采访者可以注意到回答人的特殊反映，排除他们对所问问题的误解。当面采访的主要局限(除了费用高外)涉及到采访人员。如果对他们没有进行充分地培训，他们可能背离所要求的有关事项，从而

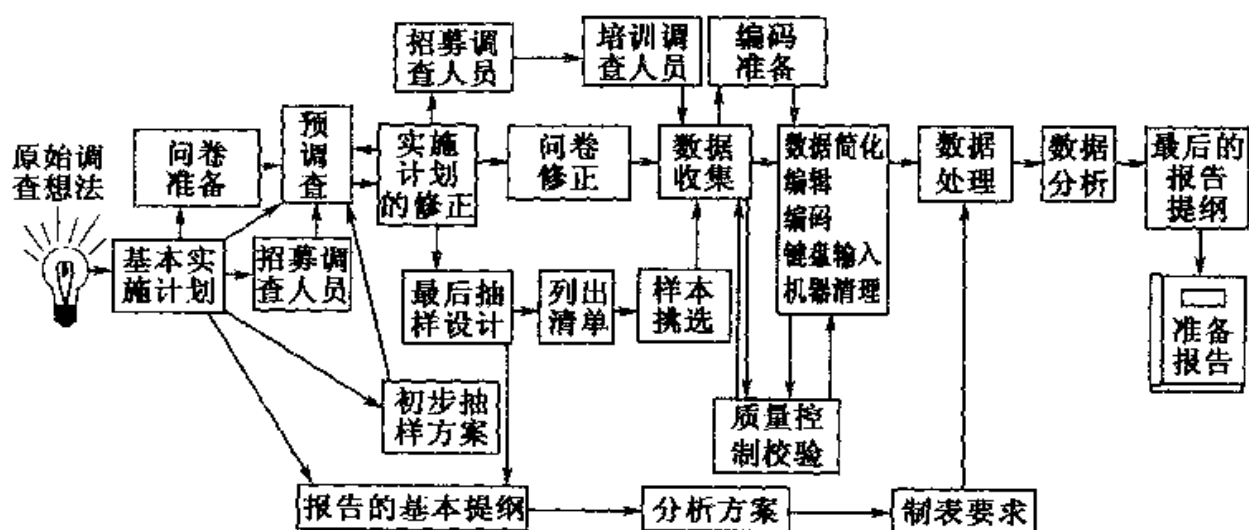


图 2.1 调查的各个阶段

把偏差带进样本数据。调查员的任何动作、表情或陈述都会影响得到的回答。例如,像“你也支持公债吗?”这样对答案有诱导性的提问会导致肯定的回答。最后,回答纪录中的误差也会导致错误的结果。

**电话采访** 也可以通过电话采访样本中的个人而得到信息。随着电话服务机构之间竞争的加剧,采访者不用花费多少钱就可向国内的某个或某些地区打上无数次电话。由于省去了旅行费用,通过电话采访进行调查通常比当面采访的费用少。调查员还能监控调查,以确信定好的采访方案被采访人员采用。

电话调查的主要问题是很难找到与总体紧相符合的电话号码本或电话号码单。电话簿中有许多非家庭的电话号码,而许多家庭电话的号码又未列入电话本中。尽管有些家庭没有电话,但在美国对大多数调查来说因缺少电话而给大多数调查带来不便只是个小问题而已。避免由于未列入电话本而造成的问题的方法是随机拨号,即先选定交换机号码(七位电话号码的前三位),然后随机地选择后四位号码进行拨叫,直至接通的指定类型的家庭达到预先规定的个数为止。用此方法不但从目标总体中产生了样本,而且避免了许多直接对电话簿进行抽样所固有的问题。

电话采访的时间一般应短于当面采访的时间,因为回答者往往更容易在电话交谈中变得不耐烦。如果问卷设计恰当,调查员经过了良好的培训,那么电话采访会像当面采访一样的成功。

**自主执行的问卷调查** 另一种收集数据的有用方法是自主执行的问卷调查,问卷由回答人自己完成。尽管可以使用其他的发送方法,但这些调查表通常是邮寄给包括在样本中的个人。如果要想鼓励回答者积极参与,那么必须谨慎地设计调查表。

由于自主执行的问卷调查不需要调查人员,所以它的使用可以节省调查费用。这一费用的节省通常是以降低回答率为代价换来的。尽管在任何形式的数据收集方法中,都会出现无回答这样的问题,但由于邮寄问卷的形式使得我们最小限度的跟回答者接触,故我们经常得到最低的回答率。低回答率会因为回答问卷的人不能代表研究人员所关心的总体而把偏差引进样本。为消除这样的一些偏差,调查员经常通过作为追加信件、电话采访或当面采访等形式与无回答者进行联系。

**直接观察** 收集数据的第四种方法是**直接观察**。如果有兴趣估计下午 4—6 时交通高峰期间使用某特定路段的卡车的数量,我们可以派一个人去数一下这段时间内经过某指定地点的卡车数,或者也可以使用电子计数装置。选用观察者的缺点在于观察中可能出现误差。

直接观察在许多不涉及对人进行测量的调查中使用。美国农业部为了得出作物产量的估计而直接测量田块中跟作物有关的一些量。野生生物学家会通过数出动物、动物的足迹、卵或巢的数量来估计动物种群的数量。

与直接观察密切相关的概念是从没有受回答者影响的客观物源处获得数据。例如,有时可以从医院的记录中得到健康信息,收入信息(特别是州或联邦政府的工作人员)可以从雇主的记录中得到。这种方法可能花费更多的时间,但它会为重要调查带来大量令人满意的回报。

## 练习

### 基本技能

**2.1(社会)** 某实验者想估计一城市每户家庭的平均用水量。讨论分别选择单个的家庭、公寓楼和街区作为样本单元时,各调查相对的优点。

**2.2(环境)** 一个林业工作人员想估计一个林场中直径超过 12 英寸的树的总数。林场的地图可以得到。讨论抽取什么样本及如何抽取它们的问题。

**2.3(工程)** 一个安全设备专家想估计具有不安全花纹的汽车轮胎的比例。在他的样本中,他应选用单辆的汽车还是选用像停车场这样的汽车的集合。

**2.4(人力资源)** 由遍布全美的各小分工厂组成的一家企业的执行官员想调查该企业员工对企业休假政策的意见。你将为她的抽样提供什么建议?

**2.5(农业)** 某州农业部门想估计该州玉米的种植面积。怎样进行这样的调查?

**2.6(政治科学)** 政治学者想估计某州成人居民赞成单院立法的百分比。抽取什么样本?同时讨论当面采访、电话采访、及邮寄式的问卷调查等作为收集数据方法的相对优点。

**2.7** 对下面的每一种情况讨论使用如当面采访、电话采访、及邮寄式的调查问卷等作为数据收集方法的相对优点。

- a. 一电视经理人员想估计某时刻全国正在观看他们的节目的观众的比例。
- b. 某报纸的编辑想调查公众对该报纸所刊登的新闻报道的类型的态度。
- c. 某城市一政府部门的官员想确定房主对提议的城市分区变动的看法。
- d. 某县卫生部门想估计去年一年发狂犬病咬人的狗的比例。

**2.8(社会)** 1984年秋季 Yankelovich、Skelly 和 White 所做的一项民意调查表明,被调查的 2207 人中有  $1/5$  的人承认在交纳联邦所得税时有欺骗行为。你认为这一比例与真正的有欺骗行为的人的比例接近吗?为什么?(讨论获得有关这类问题的精确信息的困难之处。)

**2.9(商业)** 为衡量低脂花生酱广告的效果做了两个调查,在一个调查中,调查员上门询问他们是否购买了此种低脂花生酱。在另一调查中,调查员在被访者说他或她购买了低脂花生酱时,要求他们拿盛花生酱的罐给调查员看。

- a. 你认为这两种调查会产生类似的有关选用该产品的家庭的百分比的结果吗?
- b. 每种调查中会引入什么类型的偏差?

**2.10(教育)** 时代杂志在 20 世纪 50 年代末期的一篇文章中声称“耶鲁大学 1924 年的那届毕业生的平均收入是每年 25111 美元”,这相当于如今的 150000 多美元。时代杂志的这一估计是基于 1924 年那届学生对邮寄给他们的抽样调查问卷所做的回答,而收到问卷的这些人,是其地址记录在 20 世纪 50 年代后期耶鲁大学管理部门的文档中的人。

- a. 什么是所关心的抽样总体?
- b. 抽选样本所使用的方法能产生一组可代表所关心的总体的样本吗?
- c. 获得样本的程序中什么是产生偏差的可能的来源?
- d. 基于样本偏差的来源,你认为时代杂志 20 世纪 50 年代对耶鲁大学 1924 年那届毕业生的年薪估计过高、过低还是接近正确值?

**2.11(人力资源)** 一家大型卫生保健机构想了解在其雇员当中,有多少人花费相当多时间照看年岁大的亲戚。该机构要制定一项有关雇员为照料年岁大的亲戚所需要的病假天数的政策。该机构中有数以千计的雇员,因此决定抽取一部分雇员来填写一份调查表。

- a. 你如何来定义“雇员”?是不是只考虑全职的人?
- b. 如何选取雇员样本?
- c. 应收集雇员的什么信息?

**2.12(商业)** 某大学的护理学院正在制订一项长期的计划以确定该学院将来几年所需的教师人员的数量。为此,学院需要确定往年许多毕业生于之就业的几个地区中未来对护士的需求。于是学院决定调查医疗设施和私人医生以帮助确定未来护理人员的需求。



- a. 你如何得到一张有关私人医生和医疗设施的表, 从中抽取一些医生来填写问卷?
- b. 调查问卷中应包括什么问题?
- c. 你如何确定有执照而近期又没被雇用的那些护士的数量?
- d. 在有許多护理人员就业的这些地区, 什么可能是决定人口增长和影响健康因子的根源?
- e. 你如何抽查卫生保健设施和各类私人医生的总体而不把任何医疗方面的人员排除在调查之外?

## 2.3 科学研究

科学研究中的试验设计这一主题不可能在一门统计方法课程的开始阶段就融会贯通。全面理解试验设计的方法和概念需要开设本科生和研究生水平的一系列课程。即便如此, 我们将努力对这一主题给你一个简略的概览, 因为许多需要分析或概括的数据都来自于包含某种试验设计的科学研究。

某商品检测机构决定评估四种主要品牌轮胎的使用性能。对于该研究, 这一机构选取了标准车型的四辆车, 且每一品牌选了 4 个轮胎。这些轮胎将被安在 4 辆车上, 然后这些车在 2 英里的跑道上开 30,000 英里的路程。在开了 30,000 英里后, 各轮胎胎面的厚度磨损是研究中所关心的量。由 4 个不同的司机来开这四辆车, 但他们都是具有同等的训练经历和经验的职业司机。在研究期间, 天气条件、路面状况及 4 辆车的维护情况对于所有这四种轮胎品牌没有本质差别, 其他会影响轮胎磨损的外部因素也都基本相同。这样, 检测机构有自信认为, 如果在最后的观察中发现四种品牌的轮胎的耐磨性之间有差异, 那么这一差异是真正源于品牌间的差异, 而非由进行研究的方式而造成的。检测机构有兴趣纪录其他因素, 如轮胎的价格、厂家提供的保修期限、研究过程中轮胎是否失衡及轮胎断面磨损的均匀性。在这个例子中, 我们将只考虑胎面的磨损。应该记录下每一轮胎的胎面磨损度, 每一品牌 4 个轮胎共 16 个。为了在四种轮胎品牌中做出比较(或推断), 要利用第八章和第十五章提出的方法来分析和概括胎面磨损数据。一个可能感兴趣的推断是选出胎面磨损最少的轮胎品牌。如果同样的试验重复进行的话, 能否期望在这次的样本数据中表现最佳的轮胎品牌下次还呈现最佳的胎面耐磨性? 这些研究结果适用于一般司机的驾车习惯吗?

### 试验设计

有许多方法把轮胎分配给那 4 辆车。我们将考虑做这样的试验, 其中四个品牌中每一品牌有 4 个轮胎。首先, 我们需要决定如何把轮胎分配给汽车。我们可以随机地给每辆车分配单一品牌的轮胎。但这样做将使得测量单元变成每辆车上

四个轮胎总的胎面磨损而不是单个轮胎的磨损。因此我们必须随机地把十六个轮胎分给 4 辆车。在第十五章我们将说明这一随机化如何实施。一种可能的分配方案如表 2.1 所示。

表 2.1 轮胎磨损的完全随机化设计

汽车 1	汽车 2	汽车 3	汽车 4
品牌 B	品牌 A	品牌 A	品牌 D
品牌 B	品牌 A	品牌 B	品牌 D
品牌 B	品牌 C	品牌 C	品牌 D
品牌 C	品牌 C	品牌 A	品牌 D

一般来说,完全随机化设计用在我们对比较  $t$  种“处理”(上例中  $t = 4$ , 处理是轮胎品牌)感兴趣时。对每一种处理我们得到一组观察样本。对于各个处理,样本量可以不同。例如我们可以对品牌 A、B、C 各检测 20 个而对品牌 D 只检测 12 个。设由一种处理产生的所有可能值构成假定的总体,则从该处理得到的观察值样本被认为是从该总体得到的简单随机样本的观察结果。在我们的例子中, A 品牌的 4 个轮胎磨损厚度样本被认为是从由标准型号的汽车使用 A 牌轮胎运行 30,000 英里后所有可能的轮胎磨损厚度构成的假定总体中抽出的有 4 个观测的简单随机样本的观测值。

试验中可能有某些变量的效应与如何进行试验有关。我们有可能改变试验的设计,使其适应于这种情况。在我们的例子中,我们假定了不同汽车、天气、司机和各种其他因素的效应对四种品牌都是相同的。现在,如果 4 号车轮胎的磨损程度较其他三辆车的小,那么我们的设计有没有把这一效应考虑在内了呢? 因为 4 号车上的 4 个轮胎都是 D 品牌的, D 品牌观察到的磨损可能会由于 4 个 D 品牌的轮胎均装在“最好的”车上而比其他 3 个品牌的小。在某些情况下,分配处理之前观测对象就已有所差别。例如,在评价一些降血压药效果的试验里,参与者的年龄或身体条件会降低药的疗效。为了避免这些因素掩盖药物疗效,我们希望在设计试验时把这些因素考虑在内。同样,试验过程中所遇到的环境条件也可能降低处理的效应。

在我们的例子中,我们想避免因四辆车的差异而歪曲了四种品牌的比较结果。用于实现这一目标的试验设计方法称为**随机化区组设计**,这是因为我们要用划分区组的方法排除四辆汽车差别的影响以得到四种轮胎的精确比较。在随机化区组设计中,每一处理出现在每一组内。在血压的例子中,我们将按照他们血压问题的严重程度对病人分组,然后在每一组内随机分配药物给病人。由此看出,随机化区

组设计与抽样调查中用的分层抽样相似。在轮胎磨损例子中,我们将把四辆车看成四个组,并给四辆车中的每一辆车,在每一品牌的轮胎中随机地分配一个轮胎。如表 2.2 所示。

表 2.2 轮胎磨损的随机化区组设计

汽车 1	汽车 2	汽车 3	汽车 4
品牌 A	品牌 A	品牌 A	品牌 A
品牌 B	品牌 B	品牌 B	品牌 B
品牌 C	品牌 C	品牌 C	品牌 C
品牌 D	品牌 D	品牌 D	品牌 D

好了,如果汽车间有任何影响轮胎的磨损情况的差别,那么这一效应会同样的作用到所有四种品牌上。

如果车胎在汽车上的安装位置影响轮胎的磨损会怎么样呢?车胎的位置有右前(RF)、左前(LF)、右后(RR)、左后(LR)。在表 2.2 中,我们假定 4 个 A 品牌的轮胎被安装在车的 RF 位置,B 品牌安装在 RR 位置,C 品牌安装在 LF 位置,D 品牌安装在 LR 位置。如果在 RF 位置上的轮胎磨损最严重,那么相对其他 3 种品牌而言,A 牌轮胎处于极其不利的地位。在这种情况下,我们说品牌效应和车上位置效应混杂在一起了。也就是说,使用研究中的数据,我们不能把两个或两个以上因子的效应明确地归因于单个因子。如果我们观察到四种轮胎品牌之间的平均磨损存在很大差异,那么这一差异是由于品牌不同造成的还是由于轮胎安在车上的不同位置造成的?利用表 2.2 给出的设计,无法回答这一问题。这样,我们现在需要两个区组变量:测试轮胎的汽车和车胎安在汽车上的位置。有两个区组变量的设计称为是拉丁方设计。对于我们的例子,一个拉丁方设计如表 2.3 所示。

表 2.3 轮胎磨损的拉丁方设计

轮胎安装位置	汽车 1	汽车 2	汽车 3	汽车 4
RF	品牌 A	品牌 B	品牌 C	品牌 D
RR	品牌 B	品牌 C	品牌 D	品牌 A
LF	品牌 C	品牌 D	品牌 A	品牌 B
LR	品牌 D	品牌 A	品牌 B	品牌 C

注意,对于这一设计,每种品牌的轮胎都分别被安在四辆车上的四个不同位置。因此,如果位置或汽车对轮胎磨损有影响的话,那么位置效应和/或汽车效应对四种品牌来说将相同。观测到的磨损差异就能被归因于四种汽车品牌之间的差异。

随机化区组设计和拉丁方设计都是以比较  $t$  个处理为目标的完全随机化设计方法的扩展。有关对完全随机化设计收集来的数据的分析及由此作出推断的讨论,我们将在第十五、十七章给出。第六章给出了随机化区组设计的一个特例,其中处理的个数  $t=2$ ,并且还讨论了数据的分析及如何从这些分析作出推断。

### 析因试验

假定我们想考察两个(或两个以上的)变量(因子)对于某响应的效应。例如,假定一试验员要考察两个变量,即氮和磷对选定的一个玉米品种的产量的影响。同时假定每一因子有 3 个水平,即氮肥每块地施 40、50 或 60 磅;磷肥每块地施 10、20 或 30 磅。对这个研究,试验单元是一些面积较小而又相当类似的小块土地,它们是把一定面积的农场进行划分而成的。

一种考察两种或两种以上的因子对响应的影响的方法是一次一因子法(one at a time approach)。为了考察单个变量的效应,试验人员在保持其他自变量水平恒定的同时,改变该变量的水平。这一过程一直继续到每一变量对响应的效应都被考察完毕。

例如假定我们想确定使玉米产量达最高的氮肥和磷肥的组合。我们将选定磷肥的一个水平比如 20 磅,变动氮肥的水平,然后观察出哪种组合得到最大的产量,其中产量以每英亩多少浦式耳计。其次,我们将使用有最大产量的组合中氮肥的水平,变动磷肥的施肥量,然后观察出最大产量的氮肥和磷肥的组合。这一组合称为最佳处理。我们将用表 2.4 中给出的假定的产量值来说明与该方法有关的问题。假定对各种处理做了许多次重复试验,这样试验的结果几乎与产量的真值一样。

表 2.4 假定的总产量(浦式耳/英亩)

氮	磷		
	10	20	30
40	125	145	190
50	155	150	140
60	175	160	125

起先,我们用 20 磅磷肥分别和氮肥的三个水平 40、50、60 来做试验。我们会得到用 60 磅氮肥和 20 磅磷肥产出最大产量为 160 浦式耳/英亩。然后我们把氮

肥的水平固定在 60 磅,变化磷的水平,这将得到它与水平为 10 磅的磷肥有最高产量 175 浦式耳/英亩。因此,我们可能断定 10 磅磷肥和 60 磅氮肥产生最大的产量。这些试验的结果概括在表 2.5 中。

表 2.5 试验结果中的产量

磷	20	20	20	10	30
氮	40	50	60	60	60
产量	145	155	160	175	125

基于用一次一因子法得到的试验结果,我们可能作出这样的结论,60 磅氮肥和 10 磅磷肥是最优的组合。细细检查表 2.4 中的产量我们发现最优的组合是 40 磅氮肥和 30 磅磷肥,这一组合的产量是 190 浦式耳/英亩。由此看出,只要一因子对响应的效应在另一因子的所有水平上不保持相同,此类试验就可能得出不正确的结果。在这种情况下,我们就说因子间有交互作用。图 2.2 描出了氮肥和磷非对玉米产量的交互作用。注意当磷肥保持 10 磅,而氮肥量从 40 磅增加到 60 磅时,产量随之增加;在磷肥的水平为 20 磅处,增加氮肥的量,产量也随之增加,但增量较小。在磷肥的水平为 20 磅处,当氮的水平从 40 磅变到 60 磅时,产量增加了 15 浦式耳。然而,在磷肥的水平为 10 磅时,当氮的水平从 40 磅变到 60 磅时,产量增加了 50 浦式耳。此外,在磷肥的水平为 30 磅处,氮肥水平的增加实际上导致产量的减少。当因子间没有交互作用时,对于磷的各个水平,增加氮肥量的水平将得到相同的产量变化。

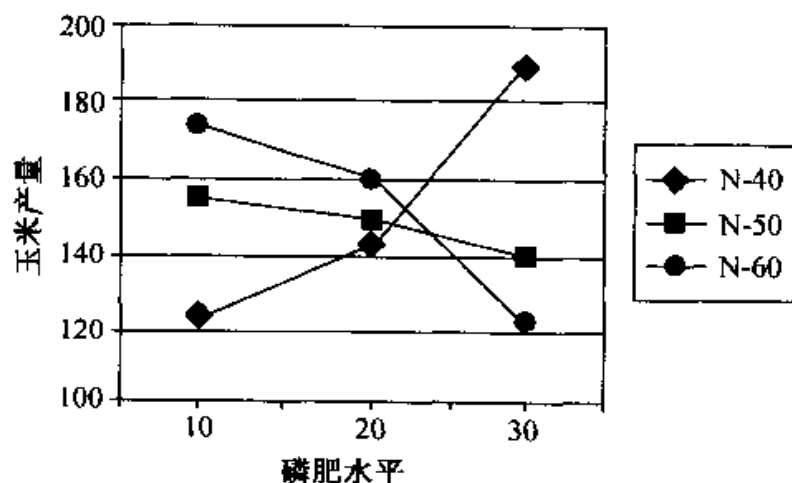


图 2.2 各种氮-磷处理的产量

表 2.6 和图 2.3 描绘出一种两个因子没有交互作用的情况。在这种情况下,磷肥对玉米产量的效应对于氮肥的三个水平都是一样的。也就是说,当我们增加

磷肥的量时,对氮肥的三个水平来说,相应的玉米产量的变化是完全相同的。注意在氮肥的所有水平,对于磷的给定的变化相应的产量变化是相同的。然而,氮肥量越高,产量也越高。因此,图中的三条线虽不相同但互相平行。当因子间有交互作用存在时,这些线或者相交或者分道扬镳。

表 2.6 假定的总产量(无交互作用)

氮	磷		
	10	20	30
40	125	145	150
50	145	165	170
60	165	185	190

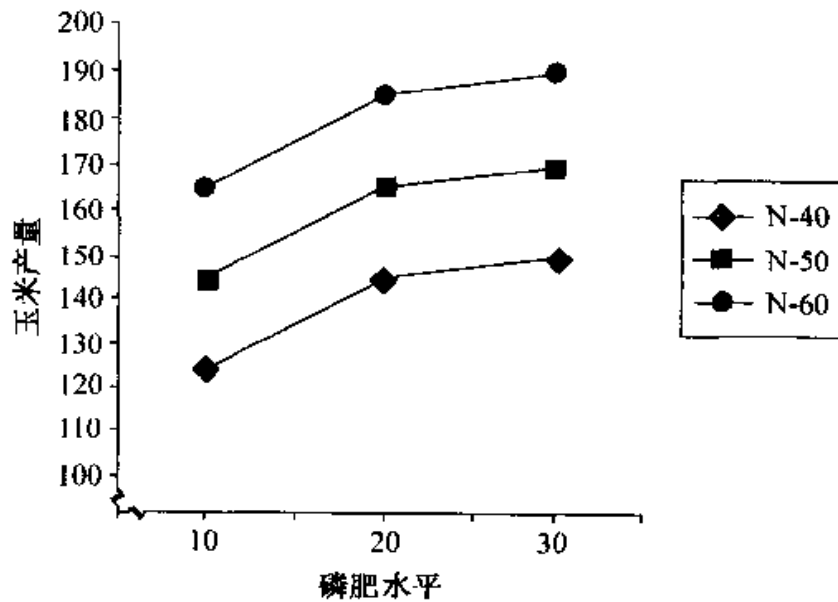


图 2.3 氮-磷各种处理的产量  
(无交互作用)

从图 2.3 我们可以看出对于两个因子没有交互作用的情况,一次一因子法是适用的。不论磷肥最初选择何种水平,一次一因子法总会得出最优产量。然而,在大多数情况下,试验之前并不知道两个因子之间是否存在交互作用。如果事实上两因子间确有交互作用而假定其没有交互作用并执行了一次一因子法,那么试验的结果将不能识别出最佳处理的组合。

不论交互作用存在与否,析因试验对于考察两个或两个以上因子对于响应值的效应都是很有用的。与前面一样,每个变量选取几个水平、这些变量的实际背景是什么都很重要。当把因子水平的组合随机地分配给试验单元时,我们就得到

了一个完全随机化设计, 处理就是这些因子水平的组合。

**定义 2.1 析因试验**是指在自变量的所有因子水平组合上都对响应变量进行了观测的试验。

使用我们前面提到的例子, 我们对考察氮肥和磷肥各水平对玉米产量的效应感兴趣, 氮肥的水平为每小块儿地 40, 50 和 60 磅, 磷肥的水平为每小块地 10, 20 和 30 磅。我们可以用完全随机化设计, 把表 2.7 中的 9 个因子混合水平(处理)随机地分配给试验单元(种植玉米的小块土地)。

两个因子不必具有相同的水平个数。例如, 我们可以做一个试验, 其中磷肥有 2 个水平, 氮肥有 3 个水平, 即  $2 \times 3$  的析因试验。因子的个数也可以不止两个。玉米产量试验可能是由钾肥 4 水平、磷肥 3 水平、氮肥 3 水平构成的, 即一个  $4 \times 3 \times 3$  的析因试验。因此, 我们会有  $4 \cdot 3 \cdot 3 = 36$  个因子水平的组合或者说处理。在第十四、十五、十七、十八章我们将讨论随机化方法及分析和推断各种试验设计中由析因试验获得的数据的方法。

表 2.7  $3 \times 3$  析因试验的因子水平组合

处理	1	2	3	4	5	6	7	8	9
磷肥	10	10	10	20	20	20	30	30	30
氮肥	40	50	60	40	50	60	40	50	60

### 更复杂的设计

有时研究的目标是我们想考察当把某些其他外部的变异根源通过“化分区组”排除在外时, 某些因子对响应值的效应。这种情况需要对来自析因试验的处理进行区组设计, 这一点可以用下例来阐明。

某调查人员想要考察两种抑制幼犬犬恶丝虫的药物(A 和 B)的效果。兽医猜想药物的效果可能依赖于小狗吃的食物。三种不同的食物(因子 1)和两种药物(因子 2)组合起来我们得到一个有 6 个处理的  $3 \times 2$  的析因试验。同时, 药物的效果还依赖于小狗从母犬那儿遗传来的对犬恶丝虫的抵抗能力。这样, 在试验中选取每窝有 6 只幼犬的 4 窝小狗作为区组因子, 因为对于特定的一窝小狗它们有相同的母亲。这 6 因子的水平组合(处理)被随机地分配到 4 窝中任一窝内的 6 只小狗, 这一方案见表 2.8。注意该设计是真正的随机化区组设计, 其中区组是窝, 处理是  $3 \times 2$  析因试验的 6 个因子水平组合。

表 2.8 犬恶丝虫试验的区组设计

幼犬	窝			
	1	2	3	4
1	A - D1	A - D3	B - D3	B - D2
2	A - D3	B - D1	A - D2	A - D2
3	B - D1	A - D1	B - D2	A - D1
4	A - D2	B - D2	B - D1	B - D3
5	B - D3	B - D3	A - D1	A - D3
6	B - D2	A - D2	A - D2	B - D1

其他更复杂的区组设计和析因试验的组合也是可能的。尽管如此,类似于对抽样调查的讨论,在本书中我们只处理最简单的试验设计。我们想强调的一点是:科学研究中有许多不同的试验设计方案可用来收集样本数据,每一设计方案各有利弊。我们将在第十四~十九章展开对试验设计的讨论。到时,我们将集中讨论如何分析由这些设计产生的数据。对于那些需要更复杂的设计的情况,为了得到最恰当的调查或试验设计方案,我们需要向专业统计学家咨询。

### 练习

#### 基本技能

**2.13** 考虑表 2.6 给出的产量。在这种情况下,没有交互作用。说明试验人员使用一次一因子法可以找到氮肥和磷肥的最佳水平组合,即得到最大产量相应的氮肥和磷肥的组合。你的解答应包括你用于试验的 5 种组合。

**2.14** 做一个  $2 \times 3$  的析因试验,得到下表中给出的总体的值。注意有两个值丢失了。如果两个因子间无交互作用,试确定丢失的值。

因子 1	因子 2		
	I	II	III
A	25	45	
B		30	50

## 2.4 观察研究

在我们离开样本数据收集这一话题之前,我们将在观察研究和科学研究之间作一区分。在科学研究的试验设计中,观测条件是固定的或被控制的。例如,对用完全随机化设计方法做的析因试验,我们在每一因子水平组合处都作了观测。类



似地,对于随机化区组设计,我们得到了每个区组中每一处理的观察。这些“受控的”研究非常不同于观察研究。我们有时使用观察研究是因为无法进行真正的科学研究。下例可以说明这一点。

许多研究和公众的兴趣集中于吸烟对肺癌及心血管病的影响。一个可能的试验设计是我们随机地把一定数量的个人(如,1000名)分成两组,要求其中一组在被研究期(如,10年)内坚持吸烟,而另一组人在整个被调查过程中不准吸烟。在研究的最后,比较这两组人患肺癌和心血管病的情况。即使我们忽略道德问题,这类试验也不可能做成。因为研究时间很长,我们不可能跟踪所有参与者并保证他们都遵守研究协定。同样,也难于找到愿意冒险去吸烟组的不吸烟者。

另一可能的研究是抽取一定量的吸烟者和非吸烟者,比较这两组间患肺癌和心血管病的情况。如果我们能找到志愿参与者,这一研究在很短的时间内就可以完成。

这样做,我们牺牲了什么呢?实际上,观察研究和科学研究之间的根本区别在于能够从研究中得出的推断是不同的。对于比较吸烟者和非吸烟者的科学研究,假设两组中的个人都遵守了研究协议,那么观察到的吸烟组和非吸烟组间的患病区别可以归因于吸烟的结果,因为这些人是被随机地分成了两组。因此从一开始就认为这两组是可比的。

这一推理过程对有关吸烟的观察研究并不适用。观察结果中显示的两组间的区别未必能归因于吸烟的结果。原因是,比如有些遗传因素会使人易染上抽烟且使人易患肺癌和/心血管病。这样组间区别可能是由于遗传因素、吸烟或两者的综合因素造成的。典型的情况是,观察研究的结果常常就变量之间的关联作出报告。对于我们的例子,如果观察研究表明,与非吸烟者相比吸烟者肺癌和心血管病的发病率高,我们会说该研究表明吸烟与肺癌和心血管病的发病率增加有关联,这种谨慎的措辞是为了避开“吸烟导致肺癌和心血管病”这样的说法。

然而,观察研究时常是我们所能做的惟一的一类研究。我们的任务是确实理解所做研究的类型,从而理解数据是如何收集来的。然后,我们可以评论从这类研究数据的分析中作出的推断。

## 2.5 数据整理:为概括和分析准备数据

这一部分,我们集中讨论在数据收集完成以后,为把这些数据变成计算机可读的形式并进而进行分析所要遵循的一些重要的数据整理方法。这里给出的当然不是一本包含所有必备工具的完整的手册,而只是对一位整理人员应该了解的各个步骤的综述。作为一个例子,本节讨论制药工业中已经严格正规化了的标准方法。这些方法在其他企业和其他背景下会有所不同。

我们从处理研究数据开始对这些方法的讨论。在实践中,这些步骤可能要花费从原始数据的获取到分析结果的得出所需总精力的 75%。这些方法是什么呢,它们为什么如此重要,为什么如此的耗费时间呢?

为了回答这些问题,我们来列出从数据的获取起,到统计分析开始之前,这一过程中的主要数据处理步骤。然后我们将分别讨论每一步骤。

**为概括和分析而处理数据的步骤:**

1. 获得原始数据资源。
2. 从原始数据资源创建数据库。
3. 编辑数据库。
4. 校正和澄清原始数据资源。
5. 数据库的最终确定。
6. 从数据库创建数据文件。

1. **获得原始数据资源**。对于每一个将要作出分析和概括的研究,数据总是以某种形式到达我们手中的。我们将把这些形式的数据统称为**原始数据资源**。对于一个临床试验,原始数据资料的形式通常是病历报告,即记录着进入研究的每一位患者的调查数据的规格为  $8\frac{1}{2} \times 11$  英寸的纸。对于其他类型的研究,原始数据资源可以是实验室笔记本上的几页纸,储存数据的磁带(或其他机器可读的数据形式),手工绘制的表格等等。

保留原始数据资料是重要的,因为它是**数据轨迹**的开始。数据轨迹起始于原始数据终止于从研究中得出的结论。涉及分析和概括许多不同研究的很多咨询工作都保留一份日志,记录与研究 and 原始数据有关的重要信息。在某些正规化的环境中(例如制药工业),可能必须在以前工作的基础上重新产生数据并重新进行数据分析。药厂以外的其他场合中,也可能会要求对研究分析所做的一切写一份回顾总结。在这些场合下,研究日志将会成为获取研究信息的宝贵资源。研究日志中一般包括如下一些信息:

**研究数据日志**

- (1) 获得数据的时间,数据从何处而来。
- (2) 调查研究者。
- (3) 指定来做研究的统计学家和其他工作人员。
- (4) 研究的简单描述。
- (5) 所研究的处理(组合、准备等)。
- (6) 原始数据资源。
- (7) 所要测量的响应及测量方法(如何测量)。
- (8) 用于研究的参考数字。

(9)估计得到的(或实际的)完成日期。

(10)其他的一些相关信息。

随后,当研究分析完毕并进行结果交流以后,可以附加一些内容,说明研究结果是如何交流的,这些结果记录在何处,保存了什么数据文件及这些文件保存在何处等。

**2. 从原始数据资源创建数据库。**对大多数计划做统计分析的研究,一般都建立一个计算机可读的数据库。创建数据库所采取的步骤及数据的最终形式随具体情况而异,依赖于统计分析所用的统计软件系统。然而,我们可以基于输入系统的形式给出一些指导性原则。

当数据将要被输入某终端时,首先要检查原始数据的易读性。任何难以辨认的数字或字母或其他的问题都应引起协作者的注意。然后,写一份指定数据列数和变量名的编码规则,其中也要给缺失值(例如得不到的数值)定义相应的代码。另外,对每一变量做一简单描述也是有帮助的。从终端键入的数据文件被称为**机器可读的数据库**。应该打印出一张数据库的清单,并且把它与原始数据资源进行细细地检查。其中出现的任何错误都应当在终端上进行改正并对更新的列表进行核对。

有时我们得到的数据已是机器可读的形式。对于这种情况,磁带和磁盘文件就被认为是数据库。然而你要读取数据库就必需要有一个编码指导书。利用该编码指导书,得到数据库的清单,然后仔细检查一下,看一看是否所有数字和符号看上去是合理的,生成文件的格式是否是适当的。在进一步做下去之前,必须解决掉出现的任何问题。

有些数据集是如此之小以至于我们没有必要从原始数据资源创建机器可读的数据文件。相反取而代之的是手算或用电子计算器计算。对这种情况,要检查任何计算,看它们是否有意义。不要轻信你看到的任何东西,重新验算一下不是个坏主意。

**3. 编辑数据库。**做什么样的编辑、何时算完成了编辑过程与研究的类型以及在分析之前你对数据的准确性和完全性的关心程度有关。例如,在使用统计软件包(如 SAS, Minitab)过程中,观察每一变量的最小值、最大值及频率分布以确认没有不合理之处不失为明智之举。

应该考虑使用某些其他手段进行核对,如通过数据图去寻找问题。同样,根据数据结构的不同,应该做一些**逻辑检查**。例如,如果有关病人的数据是由几次访问记录下的,那么第二次访问记录的数据不可能早于第一次访问记录的数据。类似的,如果第二次访问后病人失去跟踪,那就不该有任何该病人后来再被访问的数据记录。

对于小的数据集,数据的管理可由手工完成。但是对于大的数据集,这样的工

作费时良多而且枯燥乏味。如果需要用计算机编辑,那么应寻找一种允许用户指定某种数据编辑方式的软件。即便如此,对于更复杂的编辑和逻辑检查,可能有必要编写一个专门的编辑程序来编辑数据。编程是一项纷繁耗时的工作,在获得数据之前应预先做好计划。

**4. 校正和澄清原始数据资源。**在从收到原始数据到交出统计分析结果的任何一步中都时常出现有关原始数据的易读性或精确性的问题。我们发现为了便于在研究中确定数据的轨迹,把有关这些问题和差错列为一表保留下来是有益的。如果需要对原始数据作出修正(或澄清),就把它记在表上,并对原始数据资料作适当的变化。如果不需作任何修正,也在表上记下来。记住,对应于原始数据资料的任何变动,应对机器可读的数据做相应的变动。

**5. 数据库的最终确定。**可能有人一直告诉你说一项研究的所有数据总是一次到齐的。当然,情况并不总是这样。例如,对于市场销售的调查,不同地理位置可能调查的时间不同,因此那些负责数据处理的人不是同时得到所有的数据的。然而所有这些数据子集必须经过一个创建、编辑和修正数据库的处理过程。最后,该项研究完成并把这些数据进行编辑,写进数据库。此时应再检查一下数据库,并在分析之前作最后的修正,因为对于大型数据集,分析和概括此类的零碎工作要花费大量的人力和计算时间。对最终的数据库进行分析当然比以后不得不对作了变动的数据库重复所有的分析要强的多。

**6. 从数据库创建数据文件。**一般来说,有一组或两组数据文件是从机器可读的数据库创建来的。第一组文件称为**原始文件**,反映数据库的基本结构。列出一份这些文件的清单并与数据库清单进行核对,以验证变量是否以正确的形式被读入,并且缺省值代码被保留了下来。对于某些研究,原始文件实际上被用于编辑数据库。

第二组数据文件称为**工作文件**,它可以从原始文件产生出来。创建工作文件是为了便于分析。它们可能要通过插入、计算或变换来重新组织原始文件,选择重要变量,创建或增加新变量。列出一份工作文件的清单并与原始文件进行核对以确认已经进行的重新组织和变量选择是对的。通过手工计算来检查计算和变换得来的变量以核对程序代码。

如果原始文件和工作文件是 SAS 数据集,你应该尽量利用 SAS 提供的文件管理特性。在创建 SAS 数据集时,应该给它们分配一个长度不超过 40 个字符的描述标识。该标识能和数据集一起储存,而数据集的内容可在程序中随时用 SAS 过程“contents”打印出来。可以给所有变量起描述性的名字,长度不超过 8 个字符,这些名字应具有与项目中有关变量相应的含义。另外,长度可达 40 个字符的变量标识可用来提供附加的信息,“title”语句可以包括在 SAS 代码中以标识该项目及描述每一项工作。对每一文件可保留一份清单(用 proc print)和分析字典(用 proc

contents)。

对于用其他软件包创建文件的情况,你应该使用该计算机程序中可用的标识和文件管理特性。

即使运用了适当的统计方法,从研究中得出的结论也不能僭越分析所使用的数据,因此你得作出判断。在分析之前,你在这些数据处理的零碎工作上所花费的时间总量,真切地依赖于研究的特性、原始数据资料的质量及你对数据的精度和完全性所要达到的信度。

## 2.6 小 结

从数据中学习的第一步是聪明地收集数据:指定数据收集活动的目标、确定所关心的变量和为抽样调查和科学研究选择恰当的设计方案。这一章,我们讨论了科学研究中各种各样的调查设计和试验设计。有了对进行调查或科学研究所需考虑的问题的基本了解,就能弄清楚为了达到数据收集活动规定的目标,该如何收集所关心的变量的数据。

我们还依据从样本数据所能得出的推断结论的不同,对观察研究和科学研究作出了区分。在观察研究中发现的处理组间的区别被说成是与处理的使用有关~~群~~;另一方面,科学研究中发现的处理组间的区别则被说成是~~由于~~处理造成的。在下一章,我们将考察用来概括我们收集到的数据的方法。

## 第三部分 概括数据

### 第三章 数据的描述



## 第三章 数据的描述

- 3.1 引言
- 3.2 计算器、计算机及软件系统
- 3.3 单个变量数据的描述:图表法
- 3.4 单个变量数据的描述:中心趋势的度量
- 3.5 单个变量数据的描述:变异性度量
- 3.6 盒形图
- 3.7 多变量数据的概括
- 3.8 小结

### 3.1 引言

在前一章里,我们讨论了从数据中学习的第一步——如何智慧地从试验或调查中收集数据。现在我们转到第二步——概括数据。

统计领域可以分为两个主要分支:描述统计与推断统计。在两个分支中,我们都讨论一个测量值的集合。在以数据描述为主要目标的情形,我们所拥有的测量值常常来自于整个总体。例如,假设要描述在2000年人口普查中登记的所有家庭的年收入的分布,由于所有这些数据都记录在计算机磁带上,并可以随时调用,我们不需要从这个总体中获取随机样本,而是对整个总体的测量值进行处理。我们面临的主要问题是组织、概括和描述这些数据,即理解这些数据。同样,大量的月份、季度及年份医疗费用数据可以从HMO(保健组织)管理的健康护理业得到,这些数据由于包含疾病的类型,患者的年龄,住院和门诊的护理,处方的费用,外界的补偿以及许多其他类型的花销等而庞杂纷繁。然而,为了把这些数据以有用的形式提供给HMO的管理者、国会、医生以及消费者,需要对其进行组织、概括和描述。适当的概括性度量可以为原始测量值的集合提供一幅良好的、粗线条的描绘。通过把一大堆测量值缩减到几个这样的描述性统计量,我们可以理解数据所包含的信息。

当我们关注统计推断时,通常惟一可以利用的测量值集合就是一个样本。我们利用样本信息对从中抽取样本的总体做出结论。当然,在推断的过程中,我们也需要组织、概括及描述这个样本数据。



例如,与一些孤立的产品损害事件有关的悲剧导致了联邦法案的产生,这些法案要求对某些供柜台上销售的药品进行抗损包装。这些事件也促使工业界认识到了产品及包装质量的严格标准的必要性,认识到即使是在把这些产品运送到储仓的过程中也必须保持产品及包装的质量。某公司要了解在所有产品中不能被正确密封及运送过程中被损的产品比例。显然不可能在所有出售这种产品的商店中检查所有这种商品中的包装,但是可以得到该产品的一个随机样本,并根据样本中有缺陷的比例来估计密封不当或包装被损的实际比例。

同样,在对下一年的新住宅供应进行经济预测时,要用来自各种经济指标的样本数据。

第三个例子是某食物学家的一个试验。在该试验中食物学家要研究两种因素对面包块的体积的影响。这两种因素是脂肪类型和表面活性剂的类型(表面活性剂是这样一种物质,把它掺入制作面包用的面团中,能降低面团的表面张力,从而使生产出的面包块增大一定的体积)。试验中,在烤面包之前,把一种类型的脂肪与一种类型的表面活性剂掺入面团中,然后,测定所得面包的体积。每一对脂肪与活性剂的组合都被重复试验几次。在这个试验中,科学家想从试验的结果对这种面包的商业生产做出推断。在许多这种类型的试验中,科学家不仅给出他们的推断,而且附之以适当的图形显示。

无论我们是在描述观测到的总体还是在使用样本数据对总体进行推断,在从数据得到结论的过程中,重要的一步是对其进行有见地的描述。无论我们的目的是统计推断还是总体描述,首先我们必须充分描述我们手头所有的测量值集合。

描述一组测量值有两个主要方法,即图形方法和数值描述方法。在 3.3 节中讲解描述单变量的数据的图形方法。在 3.4, 3.5 和 3.6 节,讨论了描述数据的数值方法。3.7 节是关于数据描述的最后一个题目,即关于多变量数据的一些描述(概括)方法。

## 3.2 计算器、计算机及软件系统

电子计算器对于完成本章后面提出的一些计算有极大的帮助,尤其是对小规模的数据集合。对大规模的数据集合,由于输入数据所需要的时间太长,很少使用手持计算器,这种情况下计算机可以帮得上忙。即使对于很大的数据集合,在利用计算机终端、磁带或磁盘把数据输入计算机后,使用特定的程序或更一般的软件系统,几乎在一瞬间即可完成对数据的统计分析。使用这些特定程序或软件系统进行统计分析,不必了解计算机编程——它们大多都附有使用指南,对于如何使用给予了详细指导。另外一些程序则开发了用户终端界面,可以指导使用者通过选择完成分析。

使用计算机时,有许多统计软件包可供选择。其中常用的三个系统是 Minitab, Sas 和 Spss, 每一个系统都有个人计算机版本及主机版本。由于软件系统是一组共同工作的程序,我们可以一举得到图形、数据描述和复杂的统计分析。虽然大多数人最初遇到过挫折,但他们会发现自己可以很轻松地使用任何特定的系统。这些系统对于大量数据进行复杂的数据分析的能力,远胜于对最初的时间花费和挫折的回报。

通常,使用一个系统只需了解你感兴趣的程序。在工作中典型的步骤包括:对所使用的软件系统描述你的数据;若你的数据没有适当的形式,或想得到原始数据的子集合,就要对数据进行一些操作;然后,用你所使用的软件的特有的关键词,去调用适当的一组程序或过程;最后,从调用的程序中得到的结果,显示在计算机终端上或由打印机输出。

如果你有一些计算机的入门知识,并对使用计算机感兴趣,不妨试试弄清楚如何得到一个账户,什么程序和软件系统可用来做统计分析,以及从何处能够得到关于使用这些程序和软件系统时,如何进行数据输入的指示。

由于计算机的配置、操作系统以及文本编辑器因地而异,最好与一些计算机系统知识比较丰富的人进行讨论,以得到访问系统的方法。一旦你掌握了在一个软件系统中执行程序命令,你就会发现,在不同的地点用同一个软件系统处理一项工作是很相似的。

由于这不是一本关于计算机使用的教科书,我们没必要在这方面花费过多的时间和空间。计算机的使用方法还要在使用中学习,实践出真知。我们的主要兴趣在于解释这些程序的结果。由于程序的设计者想在输出中囊括一切使用者了解的方面,结果,在任何特定的情况下,会有部分输出结果是不相干的。当阅读计算机输出结果时,要寻求你想要的东西;如果你不需要,或是不理解某个输出统计量,不必担心。当然,你学的统计知识越多,就能理解更多的输出结果。同时,要寻求你所要的,忽视其余的。

粗心地使用这些统计软件是危险的。计算机是一个没有头脑的东西,无论结果多么荒谬,它都会做任何被要求的事情。例如,假设一组数据包括年龄、性别(1=女,2=男)、宗教(1=天主教,2=犹太教,3=耶稣教,4=其他或无宗教)以及一组人的月收入。就像我们计算年龄和月收入的平均一样,若我们要求计算机去计算宗教和性别的平均,我们就会得到这个平均数,即使它没有任何意义。如果使用得聪明,这些软件包是方便、强有力而实用的,但必须检查从计算机得到的任何结果,确信它是有意义的。结果是不是有错误?是不是有何疏漏?换言之,要保持警觉。计算机技术的一个重要缩写词仍然适用,即, GIGO: garbage in, garbage out (如果输入是有问题的,那么输出也是无意义的)。

在本书中,我们总是在介绍了如何计算之后,再用计算机软件包作一些十分冗

长而乏味的统计计算。通过使用计算机以及有关的图形和统计分析软件包,使得我们把更多的时间用来解释分析的结论,而不是去作分析。

### 3.3 单个变量数据的描述:图表法

在收集到感兴趣的测量值后,可以通过各种图形技术,对数据进行合理地组织、显示及考察。就常规而言,应当对数据进行分类,使每一个测量值都被分入这些类中某一类,且是惟一一类。这一方法可以消除不这样做时测量值分类中可能出现的模糊之处。例如,假设有一起性别歧视的案件,原告的代理律师需要总结一个大公司所有雇员的工资。为了检查是否存在工资上的不公现象,律师事务所决定把 1997 年所有女雇员的年收入以美元为单位舍入到整数后,按如下分类进行总结:

收入水平	工 资
1	少于 20,000 美元
2	20,000~39,999 美元
3	40,000~59,999 美元
4	60,000~79,999 美元
5	80,000~99,999 美元
6	100,000 美元以上

每一个女雇员的收入落入其中一类,且惟一一类中。然而,若各类定义为:

收入水平	工 资
1	少于 20,000 美元
2	20,000~40,000 美元
3	40,000~60,000 美元
4	60,000~80,000 美元
5	80,000~100,00 美元
6	100,000 美元以上

那么,当把一个收入归类时就可能发生混乱。例如,收入 40,000 美元的雇员可以被分在第 2 类或第 3 类。重申一下,若将数据整理分类,重要的是定义其中的类,使每一个测量值都只可以归入其中的一类。

当按以上准则整理数据时,有几种图表法可以把数据显示出来。首先,按这种方式整理数据的最简单的方法是饼图法。通过把一个圆分割(近似于切一张饼),饼图法很有效的展示了测量值落进每一类的总数的百分比。

表 3.1 是一项研究的总结,该项研究统计了在重要的公益组织中,负最高责任

的人通向权利顶峰的路径。利用他们的传记中的信息,根据他或她是如何被选拔到当前的中坚位置,把 1,345 个人进行分类。

表 3.1 最高公益位置的选拔

从何位置上选拔上来	人数	百分比
社团	501	37.2
公益	683	50.8
政府	94	7.0
其他	67	5.0

虽然你可以浏览表 3.1 中的数据,但用饼图更容易解释这些结果。通过图 3.1,我们可以对人们通向官宦之路作出某种推断。例如,与从其他组织的中坚位置出身的人数比起来,更多的人来自公益组织。

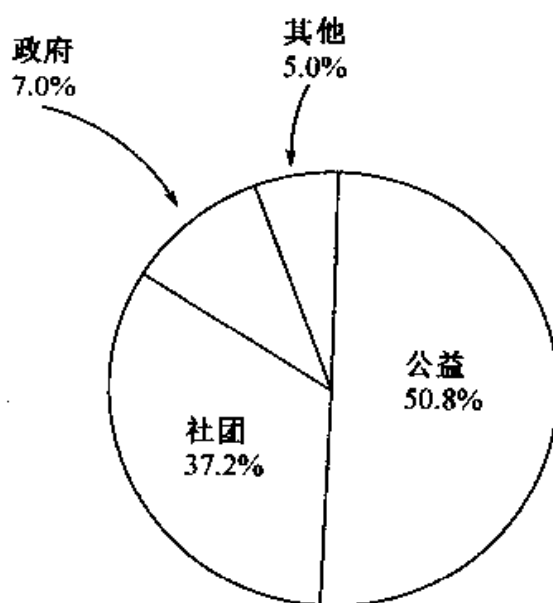


图 3.1 表 3.1 中数据的饼图

图 3.2 和图 3.3 给出了饼图的其他变化形式。可以从图 3.2 明显地看出,从 1980 年到 1990 年,可乐软饮料的人气上升,而一些其他的软饮料的市场份额下降。类似,从图 3.3 中可见,丢失了一家主要的食品连锁店使得百事可乐公司的销量受到影响。总之,饼图可以用来显示变量中每一类所占的百分比。下面的指南可以使你把饼图做得更清晰易看。

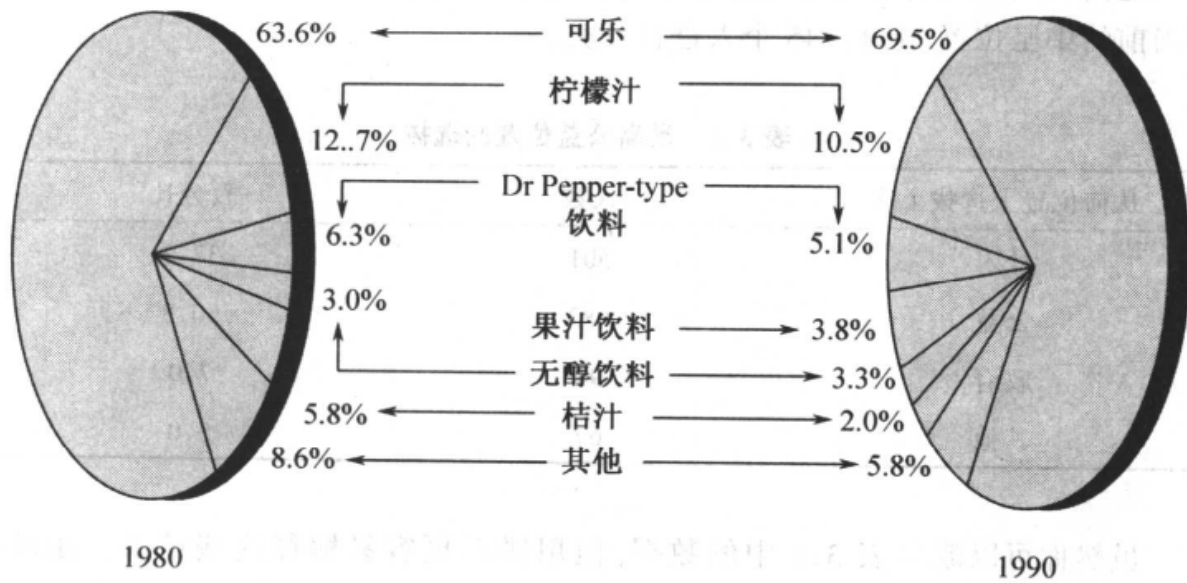


图 3.2 1980 年和 1990 年各种软饮料所占市场的大致份额

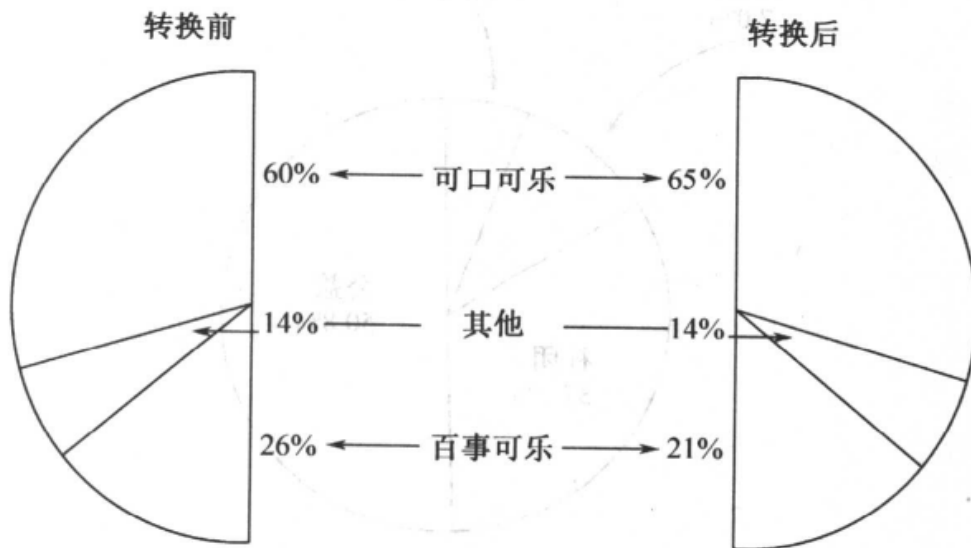


图 3.3 连锁店转换供货商前后占美国市场份额的估计\*

### 饼图的作法指南

1. 选择较少的变量分类数, 因为分类太多会使得饼图难以解释。
  2. 在任何可能的情况下, 按百分比上升或下降的顺序作饼图。
- 按前面介绍的将数据分类的准则组织数据的第二种图表法是条形图。图 3.4

\* 一家大型快餐连锁店的供货, 从百事可乐转移到了可口可乐。

给出了在 Cincinnati(辛辛那提)州及 Ohio(俄亥俄)州最大的五家外国投资者的工人数目。据估计,总的劳动力是 680,000 人。条形图有许多形式。有时,条形图平行地排列,如图 3.5(a)和 3.5(b)。条形图也可根据时间排列数据,如图 3.6。用下面给出的指南,条形图是很容易作成的。

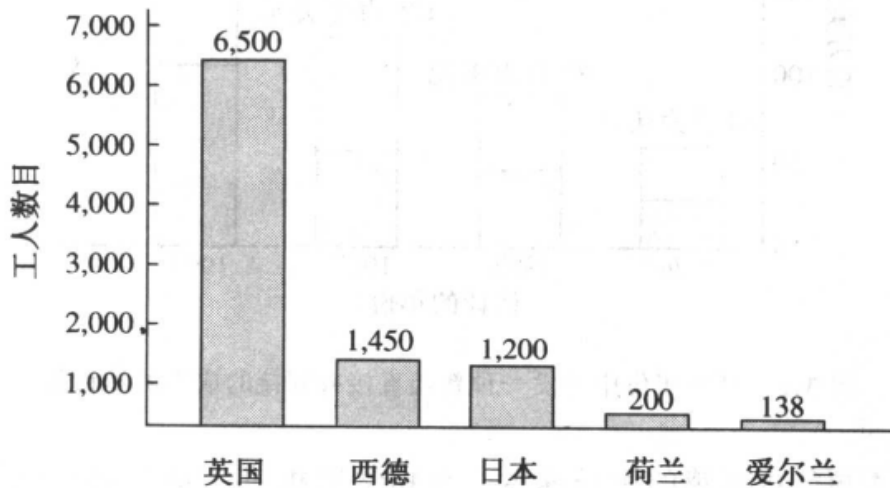


图 3.4 主要外国投资者的工人数目

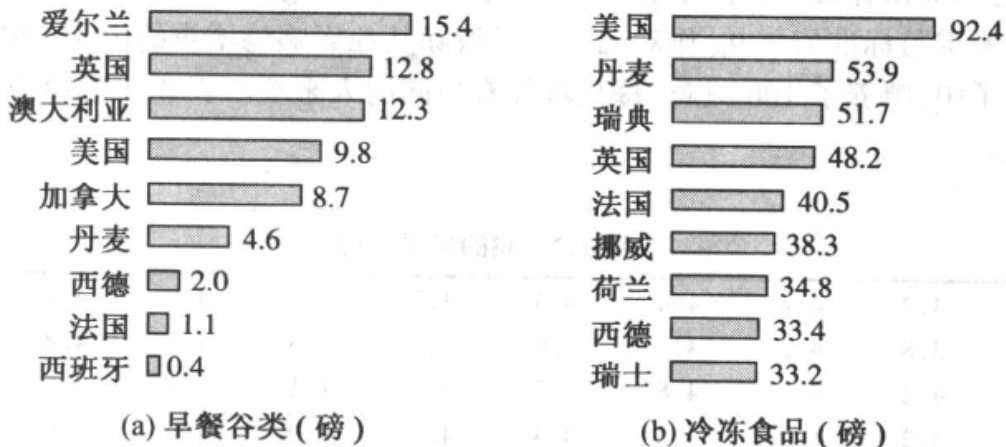


图 3.5 各个国家人均消费的最大值

### 条形图作法指南

1. 把频数列在一个轴上,变量的类别列在另一个轴上。
2. 在变量值的每一类别上,利用频数(观察值的数目)作相应高度的矩形。
3. 为了加以区别、分开各类以及使描述更清晰,在各类之间留出空隙。

下面我们将讨论**频数直方图**和**频率直方图**。这两个图表法都仅适用于定量的(可测量的)数据。像饼图中一样,我们必须在作图之前组织数据。

一名动物学家正在做一个试验,调查在鸡食中加入抗生素后比没有抗生素的

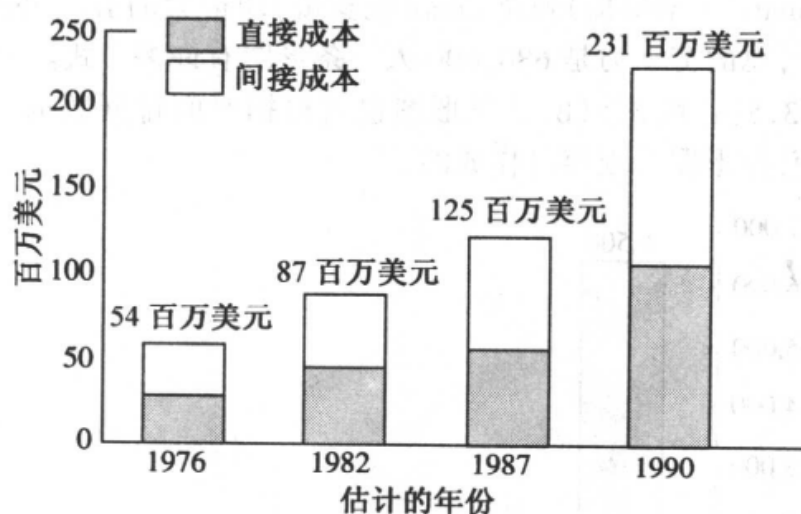


图 3.6 某些年份中开发一种新药直接或间接的成本的估计值

标准食物是否更能提高鸡的生长速度。动物学家相信, 100 只鸡可以为证实这个试验的结论提供足够的信息。(在第五章, 我们将给出若干方法, 以确定在一项研究中, 为了达到某个指定的目标所需要的适当的样本容量)。从以前的研究中知道, 通过八周用标准饲料的喂养, 一只鸡平均增重 3.9g。在这个研究中, 科学家想把鸡的增重与标准值 3.9g 作对比。为了清除其他影响鸡增重的因素, 动物学家在同一房子中, 喂养了 100 只鸡, 每只鸡都有不同的人来喂。表 3.2 记录了 100 只鸡的增重。

表 3.2 鸡的增重(克)

3.7	4.2	4.4	4.4	4.3	4.2	4.4	4.8	4.9	4.4
4.2	3.8	4.2	4.4	4.6	3.9	4.3	4.5	4.8	3.9
4.7	4.2	4.2	4.8	4.5	3.6	4.1	4.3	3.9	4.2
4.0	4.2	4.0	4.5	4.4	4.1	4.0	4.0	3.8	4.6
4.9	3.8	4.3	4.3	3.9	3.8	4.7	3.9	4.0	4.2
4.3	4.7	4.1	4.0	4.6	4.4	4.6	4.4	4.9	4.4
4.0	3.9	4.5	4.3	3.8	4.1	4.3	4.2	4.5	4.4
4.2	4.7	3.8	4.5	4.0	4.2	4.1	4.0	4.7	4.1
4.7	4.1	4.8	4.1	4.3	4.7	4.2	4.1	4.4	4.8
4.1	4.9	4.3	4.4	4.4	4.3	4.6	4.5	4.6	4.0

对增重数据的初步考察显示最大增重是 4.9g, 最小增重是 3.6g。虽然我们可以进一步考察表中的数据, 确定鸡的增重是否显著地大于 3.9g, 但很难描述测量值是如何在 3.6 到 4.9 这一区间上分布的。大多数的测量值大于 3.9? 集中于

3.6 附近? 还是集中于 4.9 附近? 或是在这一区间上均匀分布? 对于这些问题, 得到答案的一种方法是把数据组织在一**频数表**中。

要构建频数表, 首先我们把 3.6 到 4.9 这一区间分成任意个数的子区间, 称之为**分组区间**。所选择的子区间的数量依赖于集合中测量值的数量, 但我们通常提倡用 5~20 个分组区间。数据越多, 我们用的组数越多。这里给出的规则可用来构建合适的分组区间。

#### 分组区间构造指南

1. 利用所希望的分组区间的大致数目, 除测量值的极差 (最大与最小测量值之间的差)。通常, 要求 5~20 个区间。
2. 利用所希望的子区间的数目除极差之后, 把结果中的数据四舍五入到便利 (易于处理) 的单位。这个单位表示分组区间的共同宽度。
3. 选择第一个分组区间, 使其包含最小的测量值。建议为第一个区间选择一个始点, 使得没有测量值落在两个子区间的分点上, 从而避免在确定一个测量值的分组区间时发生模棱两可的情况 (一种方法是选取这些边界点, 使得它们比数据多一位小数)。

对于表 3.2 中的数据, 极差 =  $4.9 - 3.6 = 1.3$ 。

假设我们想取大约十个区间, 把这个极差除以 10 并四舍五入到一个方便的单位, 得到  $1.3/10 = 0.13 \approx 0.1$ 。分组区间的宽是 0.1。

很方便地选择第一个区间是 3.55~3.65, 第二个区间是 3.65~3.75 等等。注意, 最小的测量值 3.6 落入了第一个区间, 并且没有测量值落入分组区间的端点上 (见表 3.3)。

表 3.3 小鸡数据的频数表

分组	分组区间	频数 $f_i$	频率 $f_i/n$
1	3.55~3.65	1	0.01
2	3.65~3.75	1	0.01
3	3.75~3.85	6	0.06
4	3.85~3.95	6	0.06
5	3.95~4.05	10	0.10
6	4.05~4.15	10	0.10
7	4.15~4.25	13	0.13
8	4.25~4.35	11	0.11
9	4.35~4.45	13	0.13
10	4.45~4.55	7	0.07
11	4.55~4.65	6	0.06
12	4.65~4.75	7	0.07
13	4.75~4.85	5	0.05
14	4.85~4.95	4	0.04
总和		$n = 100$	1.00



确定了分组区间, 我们开始作数据的频数表。第一列列出了组号, 第二列说明了分组区间。然后, 检查表 3.2 中的 100 个测量值, 记录落入每一个区间的数据个数。落入某一个区间的测量值的个数叫做**分组频数**。这个数据记录在频数表第三列中(见表 3.3)。

一组的**频率**定义为此组的频数除以数据集合中测量值的总个数(总频数)。因此, 若  $f_i$  表示第  $i$  组的频数,  $n$  表示测量值总数, 则第  $i$  组的频率为  $f_i/n$ 。所有组的频率都在表 3.3 中的第四列列出。

我们已经把表 3.2 中的数据整理为一张频数表, 现在可以用频数表作频数直方图或频率直方图。要作一个频数直方图, 先画两个轴: 横轴标出分组区间, 纵轴标出频数。然后, 用与落入某个子区间的测量值数目相对应的高度, 在这个分组区间上作矩形。表 3.3 中数据的频数直方图如图 3.7(a) 中所示。

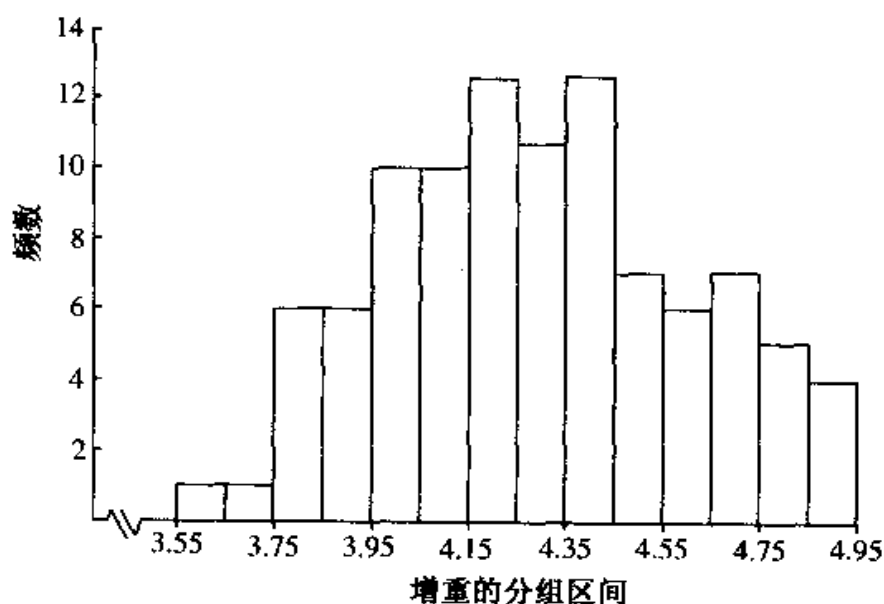
频率直方图的作法与频数直方图类似。所不同的是, 在频率直方图中, 纵轴标出频率。在每个分组区间中, 取与频率(表 3.3 中的第四列)对应的高度作矩形。表 3.3 中数据的频率直方图如图 3.7(b) 所示。显然, 直方图 3.7(a) 与 3.7(b) 形状相同。若纵轴相同, 那么它们是完全相同的。我们通常把两者都简称为直方图。

在此, 指出作直方图时应注意的几点。首先, 条形图和直方图之间的区别在于定性变量和定量变量之间的区别。定性变量值的不同在于其种类不同, 而不在于程度上的不同, 因此不是测量值。例如, 党派这一变量可按共和党, 民主党和其他党派分类。虽然我们可用 1, 2, 3 标出每一党派的人士, 但这些值仅是一种代码, 没有数量意义。相反, 定量变量有实际测量单位。例如, 每亩谷类的产量(以蒲式耳为单位)这一变量可取特定的数值。*饼图和条形图用来表示来自定性变量的频数数据, 直方图用来表示定量变量的频数数据。*

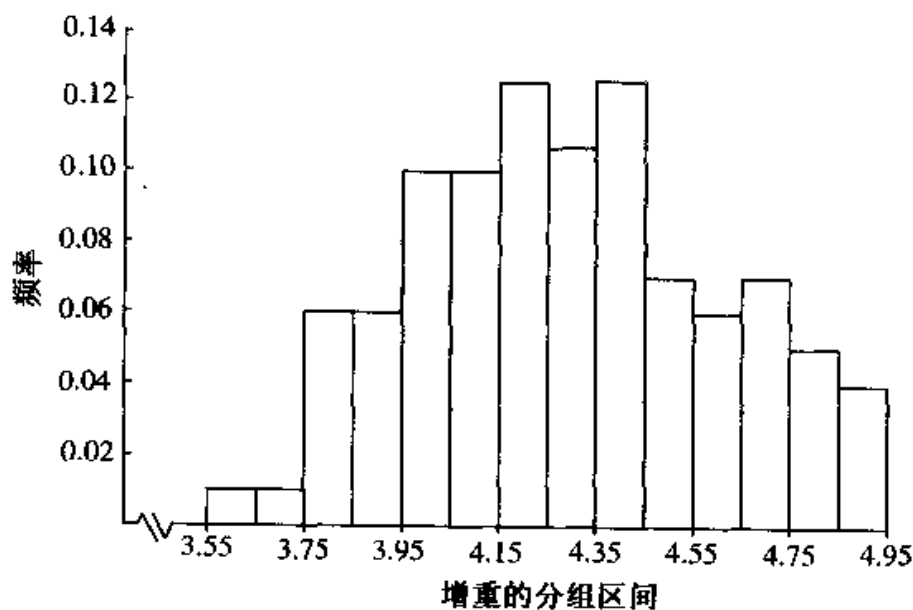
第二, 在后面章节中, 我们将要讨论统计推断。而直方图在统计推断中具有重要作用, 因此它是我们介绍的最重要的图形方法之一。第三, 对于一个特别大的测量值集合, 如果用许多宽度很小的分组区间作一个频数直方图, 直方图将近似于一条平滑曲线。第四, 在一个区间中的测量值的个数所占总数的比例, 等于对应的矩形的面积所占整个区间上直方图面积的比例。例如, 若在表 3.3 小鸡数据中, 考虑增重大于 3.9 克的区间, 会看到这些区间包含有 100 个度量中的 86 个。因此, 落入这些区间中所有测量值的比例 0.86, 等于直方图中这些区间上的矩形的面积所占整个直方图总面积的比例。

第五, 若从样本测量值集合中随机选择一个测量值, 它落在特定区间的机会或概率, 等于落在那个区间的样本测量值个数占测量值总数的比例。这一比例也可用来估计从总体中随机选择一个测量值落入所感兴趣区间的概率。例如, 从表 3.2 中的数据中, 选择一只增重大于 3.9g 的小鸡机会或概率是 0.86。0.86 可看作所有喂含抗生素的食物的鸡增重大于 3.9g 的比例的近似值, 3.9g 是喂标准食

物所得的增重。



(a)表 3.3 中小鸡数据的频数直方图



(b)表 3.3 中数据的频率直方图

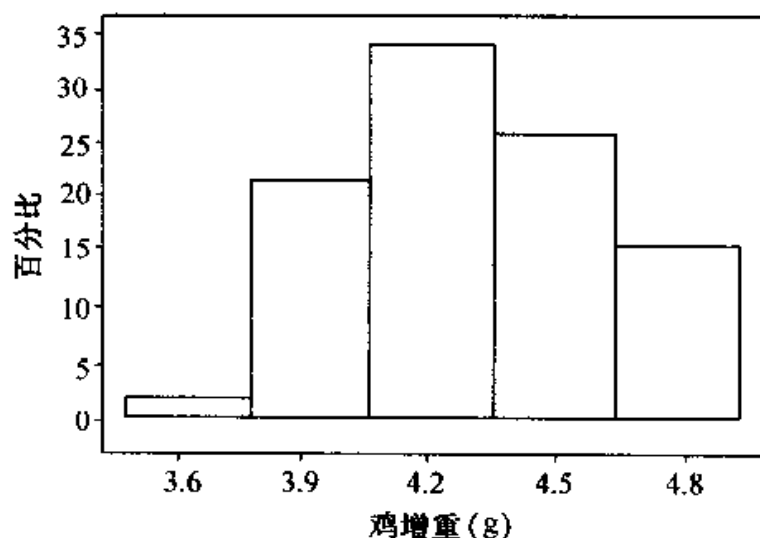
图 3.7

由于区间数、起始值和区间长度选择的任意性,对于同一个数据集合,尤其是小的数据集合,直方图可以被做成不同的形状。当数据很多时,比如 50 或更多时,用直方图来描述数据集合是很有用的。在图 3.8(a)~(d)中,用 5, 10, 14 和 18 个分组区间建立的小鸡数据的一组直方图,可以说明在建立直方图过程中所遇到的

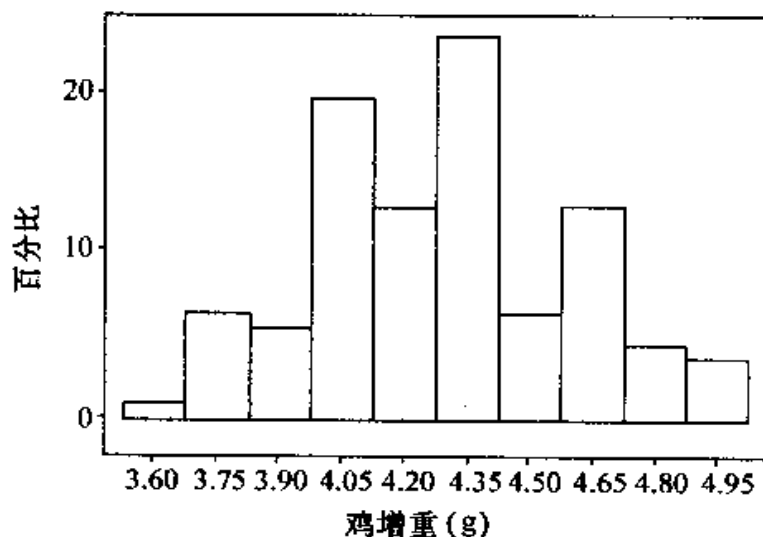
一些问题。这些图是用 Minitab 软件得到的。

当端点数相当小且区间数目很多时, 直方图会有若干区间没有数值。见图 3.8(d), 这会导致不能对总体做出真实描绘的直方图。当分组区间的数目很小时, 数据中的大多数模式或趋势都无法反映出来, 见图 3.8(a)。在 3.8 的图集中, 有 14 个分组区间的直方图可能是最恰当的一个。

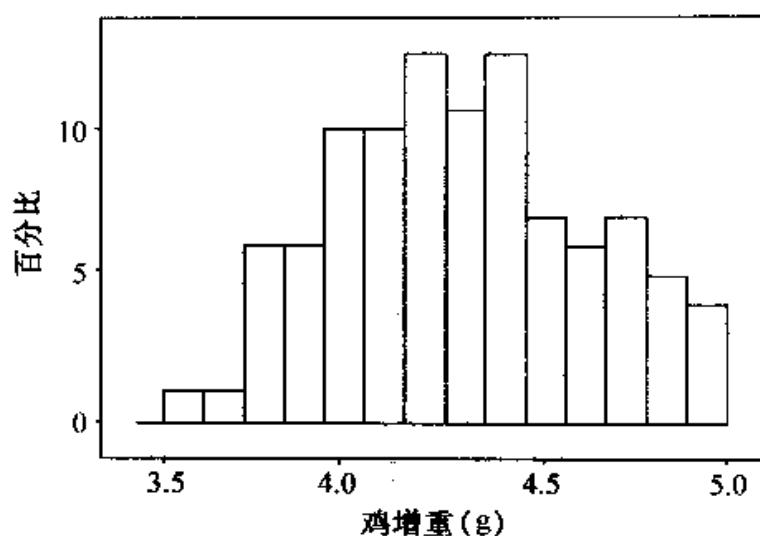
最后, 由于在频率直方图中用到的是比例而不是频数, 所以对两个不同的样本 (或总体), 即使它们的容量不同, 也可通过检查它们的频率直方图, 对它们加以比较。当描述来自许多样本的频率直方图或比较这些图形时, 我们检查直方图的整体形状。图 3.9 描述了频率直方图的一般形状。



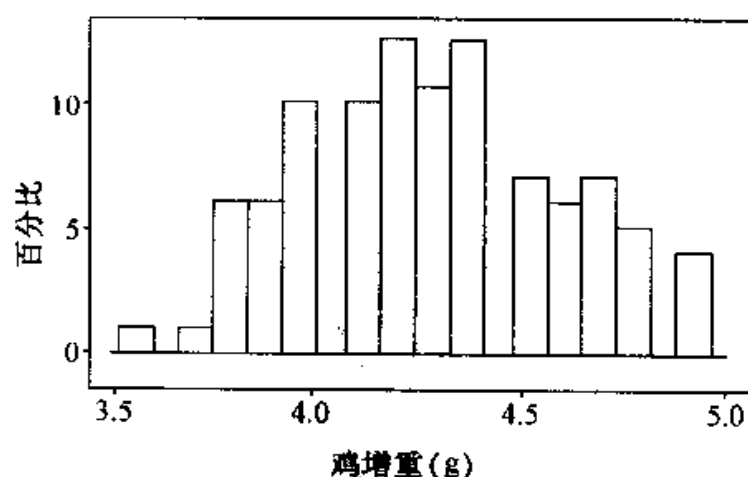
(a) 小鸡数据的频率直方图 (5 个区间)



(b) 小鸡数据的频率直方图 (10 个区间)



(c) 小鸡数据的频率直方图 (14个区间)



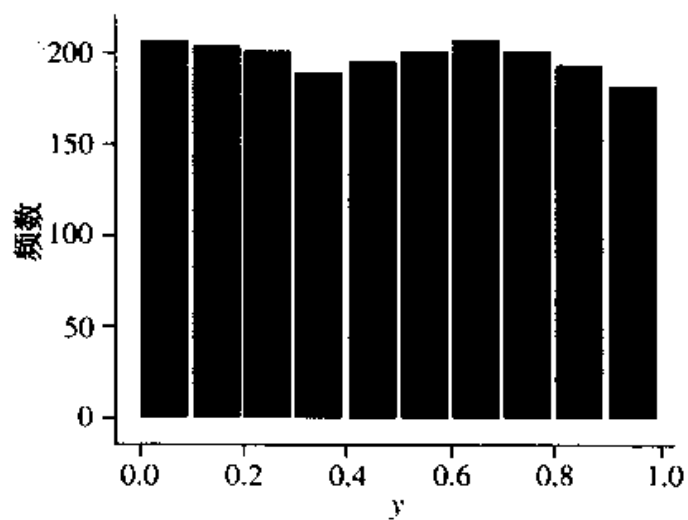
(d) 小鸡数据的频率直方图 (18个区间)

图 3.8 小鸡数据直方图

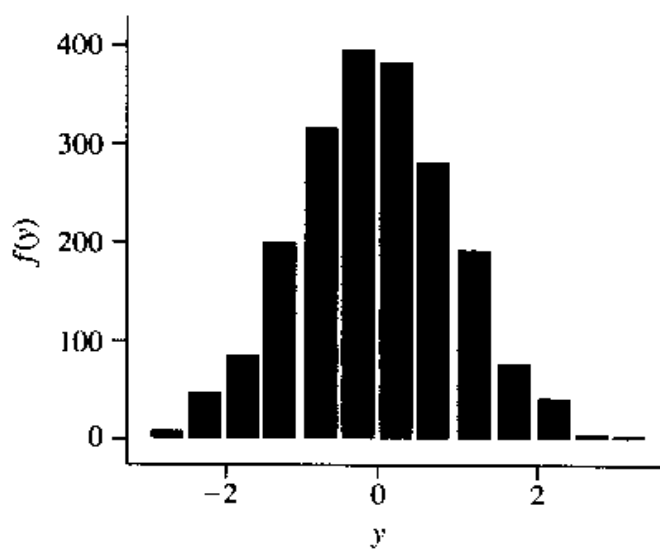
有一个峰的直方图称为**单峰直方图**, 见图 3.9(b), (c)和(d)。如图 3.9(e)和(f)中有两个主峰的直方图称之为**双峰直方图**。在许多情况下, 双峰直方图实际上是样本数据来自两个截然不同的总体的一种暗示。最后, 当在每个区间上都有基本相同的个数的观测值时, 这种直方图称为**均匀直方图**, 见图 3.9(a)。

如果一个直方图的左边和右边有基本相同的形状, 则直方图是**对称的**, 因此, 图 3.9(a), (b)和(e)有对称形状。当直方图的右边包含观察数据的大多半时, 比左边伸展的远, 则称之为**右偏的**, 见图 3.9(c)。同样, 当直方图的左边比右边伸展的更远时, 称它是**左偏的**, 见图 3.9(d)。由本书后面的论述可知, 若知道分布的形

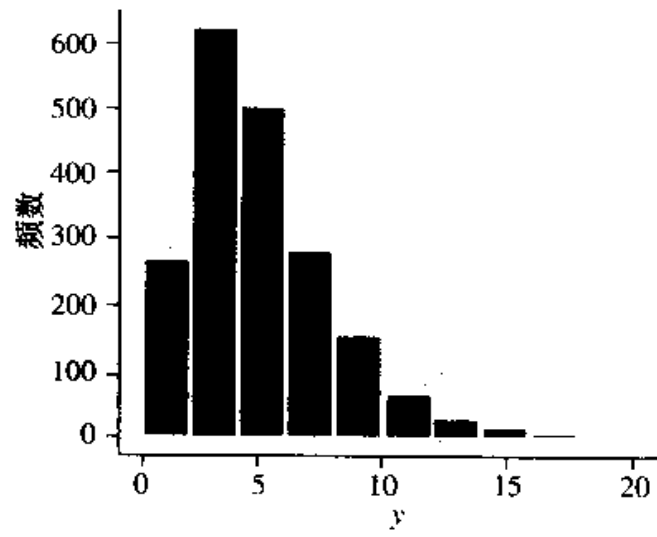
状,将有助于我们选择适当的概括数据(3.4~3.7节)和分析数据的方法(第五章及其以后的章节)。



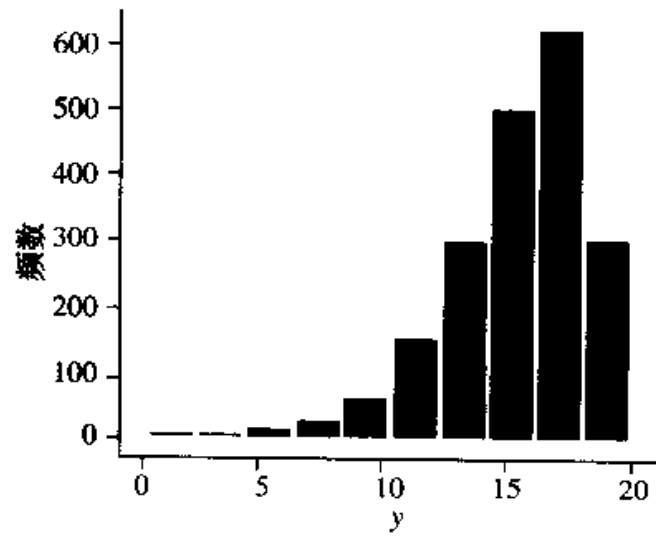
(a) 均匀分布



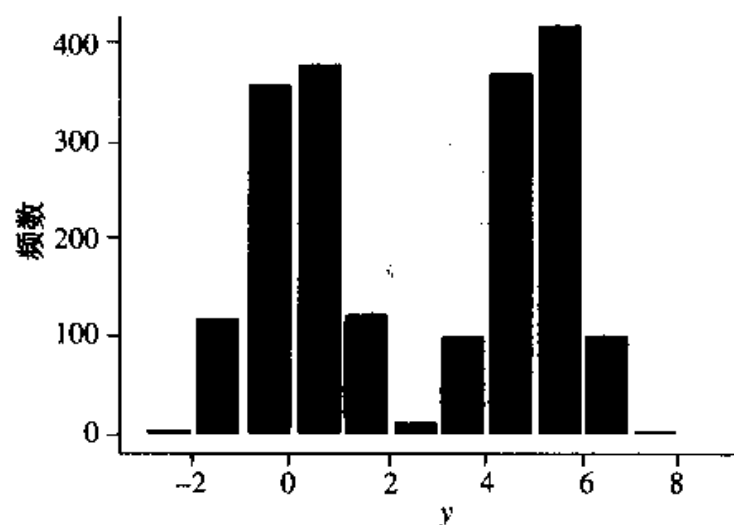
(b) 对称, 单峰 (正态) 分布



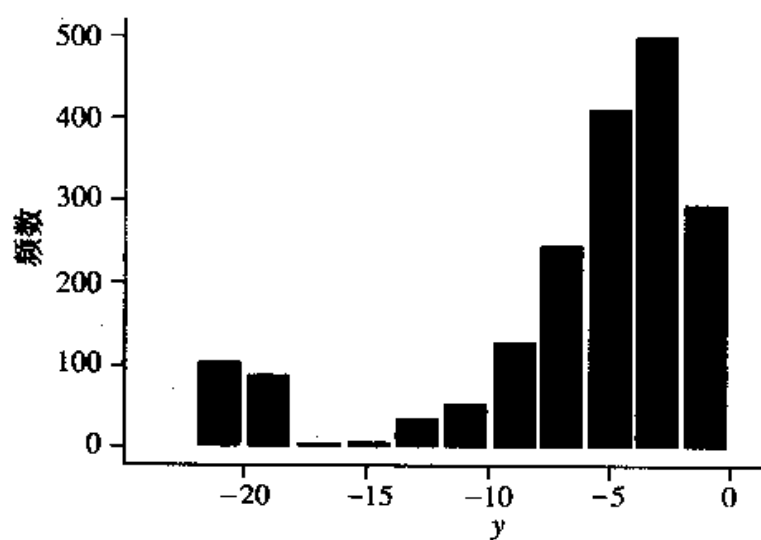
(c) 右偏分布



(d) 左偏分布



(e) 双峰分布



(f) 左偏的双峰分布

图 3.9 一些常见的分布形状

表 3.4 来自南部、北部和西部 90 个标准大城市的暴力犯罪率

南 部	犯罪率	北 部	犯罪率	西 部	犯罪率
Albany, GA	876	Allentown, PA	189	Abilene, TX	570
Anderson, SC	578	Battle Creek, MI	661	Albuquerque, NM	928
Anniston, AL	718	Benton Harbor, MI	877	Anchorage, AK	516
Athens, GA	388	Bridgeport, CT	563	Bakersfield, CA	885
Augusta, GA	562	Buffalo, NY	647	Brownsville, TX	751
Baton Rouge, LA	971	Canton, OH	447	Denver, CO	561
Charleston, SC	698	Cincinnati, OH	336	Fresno, CA	1,020
Charlottesville, VA	298	Cleveland, OH	526	Galveston, TX	592
Chattanooga, TN	673	Columbus, OH	624	Houston, TX	814
Columbus, GA	537	Dayton, OH	605	Kansas City, MO	843
Dothan, AL	642	Des Moines, IA	496	Lawton, OK	466
Florence, SC	856	Dubuque, IA	296	Lubbock, TX	498
Fort Smith, AR	376	Gary, IN	628	Merced, CA	562
Gadsden, AL	508	Grand Rapids, MI	481	Modesto, CA	739
Greensboro, NC	529	Janesville, WI	224	Oklahoma City, OK	562
Hickory, NC	393	Kalamazoo, MI	868	Reno, NV	817
Knoxville, TN	354	Lima, OH	804	Sacramento, CA	690
Lake Charles, LA	735	Madison, WI	210	St. Louis, MO	720
Little Rock, AR	811	Milwaukee, WI	421	Salinas, CA	758
Macon, GA	504	Minneapolis, MN	435	San Diego, CA	731
Monroe, LA	807	Nassau, NY	291	Santa Ana, CA	480
Nashville, TN	719	New Britain, CT	393	Seattle, WA	559
Norfolk, VA	464	Philadelphia, PA	605	Sioux City, IA	505
Raleigh, NC	410	Pittsburgh, PA	341	Stockton, CA	703
Richmond, VA	491	Portland, ME	352	Tacoma, WA	809
Savannah, GA	557	Racine, WI	374	Tucson, AZ	706
Shreveport, LA	771	Reading, PA	267	Victoria, TX	631
Washington, DC	685	Saginaw, MI	684	Waco, TX	626
Wilmington, DE	448	Syracuse, NY	685	Wichita Falls, TX	639
Wilmington, NC	571	Worcester, MA	460	Yakima, WA	585

注: 犯罪率指每 100,000 居民中暴力(谋杀、强奸、抢劫及暴力袭击)犯罪的数量, 并四舍五入到整数。  
数据来源: 司法部, 美国统一犯罪报告, 1990 年。

本章介绍的下一个图形方法, 是来自于称为探索性数据分析(EDA)领域的一种显示数据的技巧。John Tukey 教授(1977)是这一数据分析的实用方法的倡导者, 这种方法的目的旨在探索和理解数据。

茎叶图是一种简单、灵活的图形方法, 即作一个像直方图一样的频数分布图。



可以通过包含在频数分布中的信息,知道记分(这里指观测值—译者注)的范围,记分的集中之处,分布的形状以及是否有一些特殊值或记分未表示出来,是否有离群值或异常值。茎叶图不遵循前面提到的直方图规则。我们将用表 3.4 中的数据说明如何作茎叶图。

表 3.4 中的原始数据是 3 位或 4 位数字。用第一个或首位上的数字作茎(见图 3.10),其后位上的数作为叶。例如,奥尔巴尼(Albany 市,美国纽约州的首府)的暴力犯罪率是 876,首位是 8,后位是 76。而对于 Fresno 市,则首位数字是 10,后位是 20。若有 6 位数字,如 104,328,可用前两位作茎,其后的两位作叶,最后两位可忽略。

在这些有关暴力犯罪的数据中,最小的犯罪率是 189,最大是 1,020。首位数字分别是 1, 2, 3, ..., 10。与分组区间确定一个测量值在频数表中的位置类似,首位(记分的茎)确定一个记分在茎叶图中所处的行。其后位上的数写在相应的行中。按照这种方法,把每一个记分都记入茎叶图,就像暴力犯罪数据的茎叶图 3.10 一样。

```

1 89
2 98 96 24 10 91 67
3 88 76 93 54 36 93 41 52 74
4 64 10 91 48 47 96 81 21 35 60 66 98 80
5 78 62 37 08 29 04 57 71 63 26 70 16 61 92 62 62 59 05 85
6 98 73 42 85 61 47 24 05 28 05 84 85 90 31 26 39
7 18 35 19 71 51 39 20 58 31 03 06
8 76 56 11 07 77 68 04 85 14 43 17 09
9 71 28
10 20
  
```

图 3.10 表 3.4 中暴力犯罪率的茎叶图

可以看到,每一个茎决定一个分组区间,这个分组区间的上下限分别是该组中可能的最大与最小记分。由每个叶所代表的数值必须在区间的上下限之间。

注意,茎叶图看起来像横向放置的直方图,如图 3.10。若将一行(茎)中的数据(叶)按从小到大的顺序排列(图 3.11),茎叶图可能更有用。茎叶图比直方图更有利,它不仅反映了频数,记分的集中,分布的形状,而且它描述了实际记分。不利处是对于大量的数据集合,茎叶图会很庞大。

```

1 89
2 10 24 67 91 96 98
3 36 41 52 54 74 76 88 93 93
4 10 21 35 47 48 60 64 66 80 81 91 96 98
5 04 05 08 16 26 29 37 57 59 61 62 62 62 63 70 71 78 85 92
6 05 05 24 26 28 31 39 42 47 61 73 84 85 85 90 98
7 03 06 18 19 20 31 35 39 51 58 71
8 04 07 09 11 14 17 43 56 68 76 77 85
9 28 71
10 20

```

图 3.11 按叶序排列的茎叶图

### 茎叶图的作法指南

1. 把每一个记分或数值分为两套数字。第一或前几位数字的集合是茎,第二或其后的数字是叶。
2. 从小到大排列所有可能的茎数字。
3. 对于大量数据中的每一记分,在适当的茎数字所确定的行中写下叶数字。
4. 如果排列太拥挤,可以把每个茎的位置扩为两行,例如,叶数字 1, 2, 3 和 4 放在茎的第一行, 5, 6, 7, 8 和 9 放在茎的第二行。
5. 如果出现太多的位数,如 6 位或 7 位记分,可去掉最右面的尾数字,使排列更清晰。
6. 作茎叶图的规则和建分组区间的规则有所不同,分组区间的规则用在传统的频数分布及后面章节将考虑的各种过程。在某种意义上讲,茎叶图的分组区间是不规则的。

下面的茎叶图是从 Minitab 中获得的。数据描述了威斯康星州 Wisconsin 批发和零售商业中雇员的数目,是在 5 年中每月测一次而得到的。

### 数据排列

Trade

```

322 317 319 323 327 328 325 326 330 334
337 341 322 318 320 326 332 334 335 336
335 338 342 348 330 326 329 337 345 350
351 354 355 357 362 368 348 345 349 355
362 367 366 370 371 375 380 385 361 354
357 367 376 381 381 383 384 387 392 396

```

### 字符茎叶图

Stem-and-leaf of Trade      N = 60

Leaf Unit = 1.0

```

31  789
32  0223
32  5666789
33  00244
33  556778
34  12
34  55889
35  0144
35  5577
36  122
36  6778
37  01
37  56
38  01134
38  57
39  2
39  6

```

注意,由于叶数字分为两组:0~4及5~9,大多数茎都重复了两次。

本节介绍的最后一个图形法用来显示某个变量是如何随时间而变化的。对于宏观经济学数据,如可自由支配的收入,以及微观经济学数据,如特定商店中特定商品的周销售量等,它们随时间变化的数据图对于商业管理是十分重要的。同样,社会研究者常感兴趣于证实变量如何随时间而变化;或感兴趣于对各种种族团体的态度、美国的储蓄率、城市的犯罪率如何随时间而变化。表示变量如何随时间而变化的图形法称为**时间序列**。图 3.12 是一个时间序列,它表明了从 1970 年到 1986 年,年龄在 30~34 岁之间没有孩子的白人妇女的百分比。

通常,时间点按年代顺序排列在横轴上(横坐标),感兴趣的变量的数值(频数、百分比、比率等)沿纵轴标出(纵坐标)。时间可按年、月、日或任何合适的单位计。凭经验而言,一个时间序列的时间点不少于四、五个时间点;一般地,这些时间点之间有相等的间隔。然而,为了表明变量随时间变化的更完整的图像,希望有比这多得多的时间点。

时间序列中时间轴的排列通常依赖于得到数据的时间间隔。例如,美国人口普查局以年为基础报告美国平均家庭收入。当要研究的变量的信息可按不同的时间单位获得时,必须确定哪个或哪些时间单位在该项研究中是最恰当的。在大选之年,政治学家很可能按周或月检查候选人在登记的选民中受欢迎程度的变化。

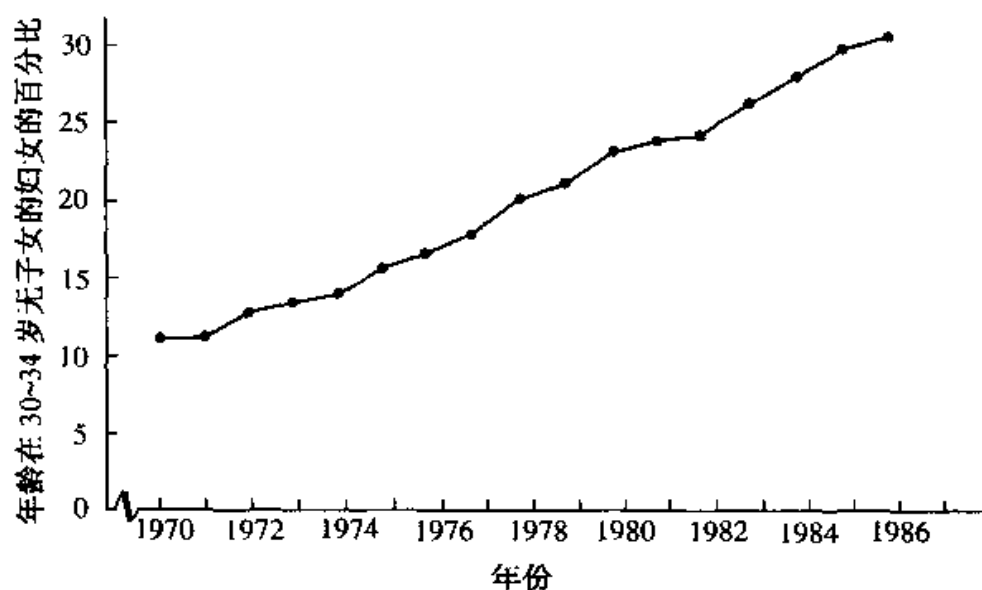


图 3.12 1970~1986 年间, 年龄在 30~34 岁无子女的妇女的百分比

另一方面, 一个机械工具的生产商可能会记录月、季度、年的销售情况(按美元和销售数量)。图 3.13 表明了过去 3 年中机械工具的季度销售量(以千为单位)。从时间序列看出, 在过去 3 年中, 公司经历了销售数量的渐进而稳定的增长。

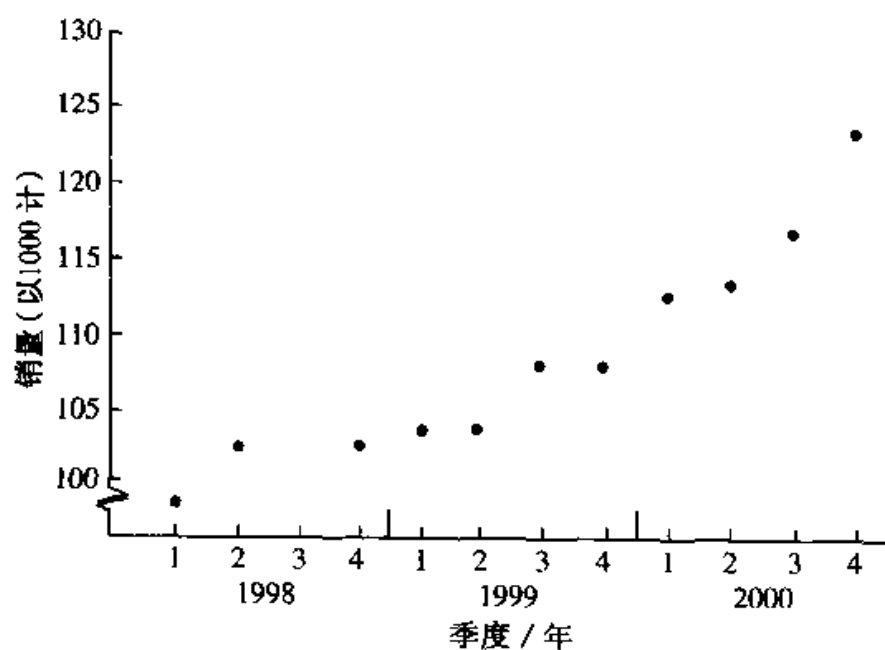


图 3.13 季度销售量(以 1,000 计)

30 STOCKS IN INDUSTRIAL AVERAGE-NYSE CHG.					
AT&T	- 3/8	Dupont	+ 1/8	McDonalds	...
AllSgnl	+ 3/16	EKodak	- 1/8	Merck	+ 5/16
Alcoa	+ 3/16	Exxon	+ 1 3/4	MinnMnMf	- 1 1/2
AmExprss	- 1/8	GenElec	- 1 5/8	Morgan JP	- 1 5/8
Boeing	- 1 1/2	GenMotor	- 1/2	PhilipMor	- 3/16
Caterpillar	- 5/16	Goodyear	+ 3/16	Proct Gam	- 1
Chevron	+ 5 1/4	HewlettPk	- 7/16	Scars	+ 1 1/2
Citigroup	+ 9/16	IBM	+ 3 3/8	UnCarbide	+ 1/4
CocaCola	- 1 7/8	IntPaper	+ 1/16	UtdTch	+ 3/8
Disney	+ 1 1/2	JohnsJohns	- 13/16	WalMart	+ 9/16

COMPONENT VOL. 29,009,600 PREV. 58,201,700  
MARKET CAP. \$ 2,430,764,995,688

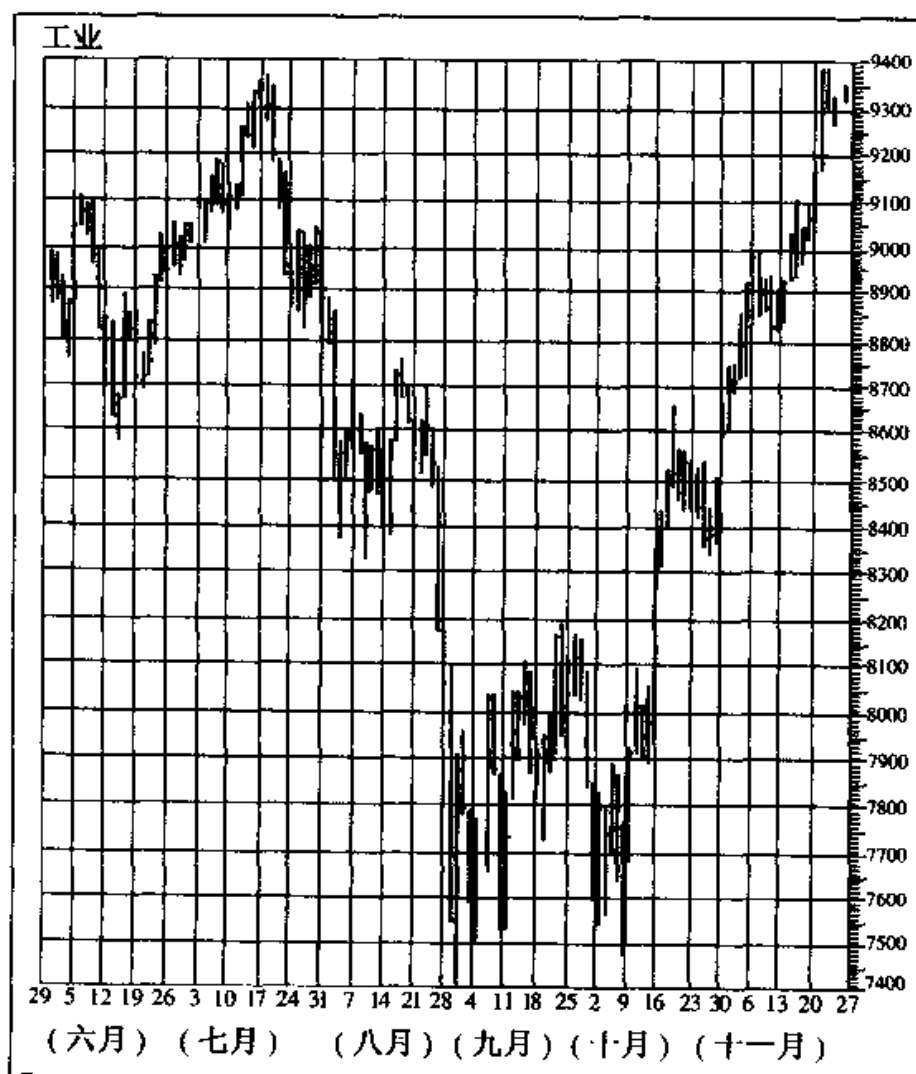
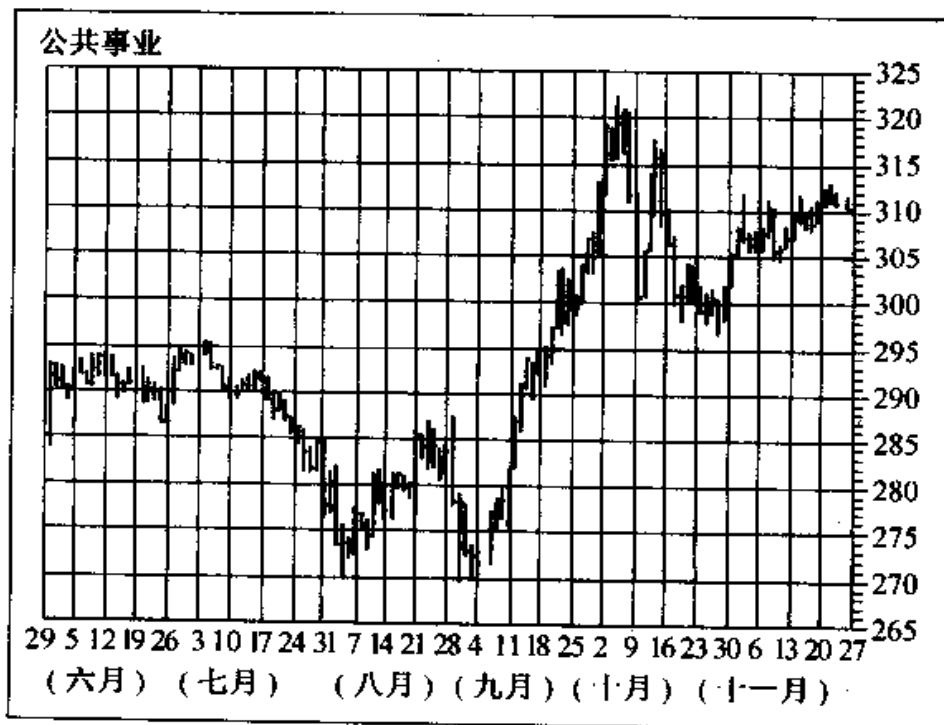


图 3.14 道琼斯工业、公共事业和运输平均指数的时间序列图

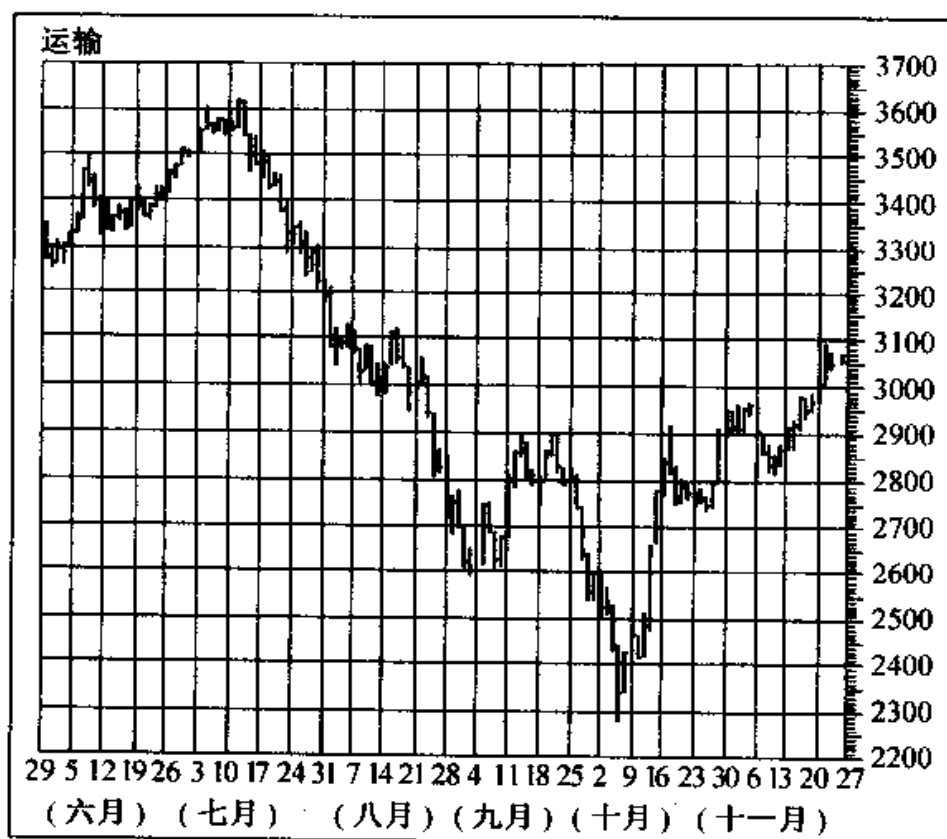
时间序列对于检查变化的总趋势、季节性及周期性是很有用的。例如,华尔街杂志(*Wall Street Journal*)的“金融与投资”栏目以六个月为一期,给出道琼斯工业、运输和公共事业平均指数的每个工作日的值。图 3.14 是以某一期的数据为典型画出的图。从图中可以看出,在 1998 年中,运输指数和工业指数从六月到七月中旬呈增长趋势,随后直到九月份呈下降趋势,从十月到十一月则呈急剧上升趋势。与这两个指数相反,公共事业指数则从六月到九月相当平稳,至十月早期有所上升,然后直到十一月又保持平稳。研究季度性或周期性,必须有数年的每周或每月的数据。

有时,比较同一变量在两组或多组中随时间变化的趋势是很重要的。图 3.5 记录了从 1976 年到 1980 年的两组比值:非洲裔美国人中位家庭收入与英裔美国人中位家庭收入的比值,及拉丁美洲人中位家庭收入与英裔美国人中位家庭收入的比值。

15 STOCKS IN UTILITY AVERAGE-NYSE CHG.					
AEP	- 1/16	Edisonint	- 1/8	PubSvcEnt	- 1/8
ColmEngy	+ 1	EnronCp	- 1/2	SouthernCo	+ 1/16
ConEd	- 9/16	Houstrind	- 5/16	Tex Util	- 1/2
ConsNat	+ 15/16	PECO Engy	- 7/16	UnicomCp	+ 3/16
DukeEngy	- 5/8	PG&E	- 3/8	WillmsCos	+ 3/8
COMPONENT VOL. 3,283,900 PREV. 7,217,700					
MARKET CAP. \$ 175,964,191,750					



20 STOCKS IN TRANSPORTATION AVERAGE-NYSE CHG.					
AMR	+	1/4	FDX Cp	+	1/16
AirbnFrt	-	1/16	GATX	+	7/8
AlxBldwn*	-	1/8	HuntJB*	+	5/16
BurlNthSF	+	3/8	NorflkSo	-	1
CNF Trnsp	...		NowestAir*	+	7/16
CSX	+	1/16	RoadwEx*	+	9/16
DeltaAir	+	1/8	RyderSys	-	5/16
			SowestAir	+	5/16
			UAL Cp	+	5/8
			UnPacific	-	1/8
			US Airways	+	1 1/2
			USFrghtr*	+	3/8
			YellowCp*	+	1/4
			* Nasdaq		NMS
COMPONENT VOL. 3,598,500 PREV. 9,572,100					
MARKET CAP. \$ 107,744,107,125					



来源: 华尔街杂志, 1998 年 11 月 30 日

图 3.14 道琼斯工业、公共事业和运输平均指数的时间序列图(续)

中位家庭收入是把家庭收入分为两组的一个收入水平。这两组分别是高收入的一半和低收入的一半。1987 年非洲裔美国人的中位家庭收入为 \$ 18,098, 即指所有非洲裔美国人的家庭中, 50% 的家庭收入高于 \$ 18,098, 50% 的家庭收入低于 \$ 18,098。中位数是中心趋势的度量方法中的一种, 将在以后的章节中作充分的讨论。

图 3.15 表明, 在 1976 年至 1988 年间, 非洲裔美国人和拉丁美洲美国人相对

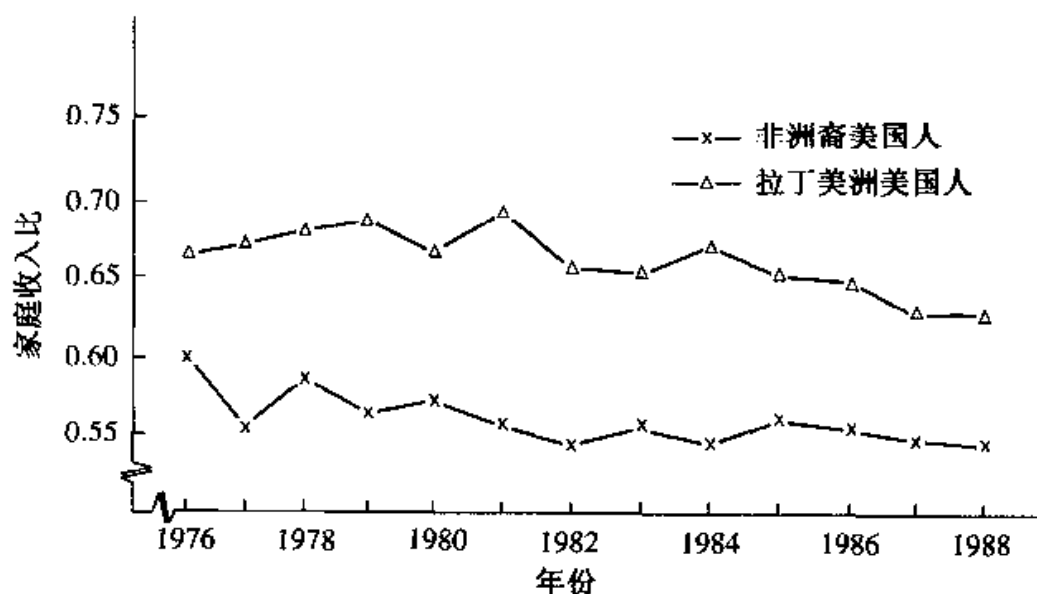


图 3.15 非洲裔美国人和拉丁美洲美国人相对于英裔美国人中位家庭收入的比值:1976~1988

于英裔美国人中位家庭收入的比值有波动,但两个比值的整体趋势说明,整个时间段中的比值是呈下降趋势的。社会研究者将其解释为,非洲裔美国人和拉丁美洲美国人的家庭收入相对于英裔美国人家庭收入逐渐下滑。

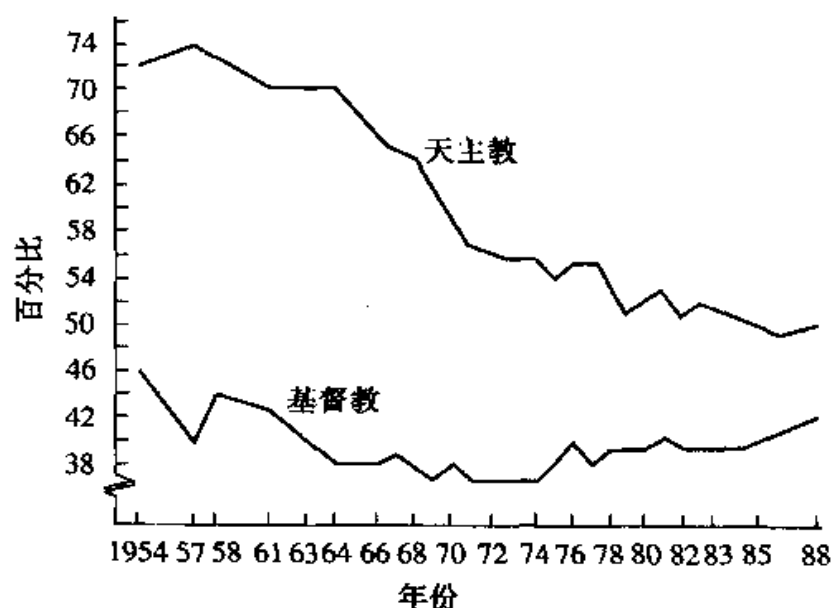
有时候信息不是在等时间间隔中获得的。例如,民意测验组织,像盖洛普(Gallup)民意测验中心或国家公众意见研究中心(National Opinion Research Center)等,不必每年都问美国公民对同一个问题的态度或行为。有时,再问这个问题时,与前次的时间间隔已大于两年。

当信息不是在等时间间隔中获得的时候,横轴上时间点间的间隔宽度能够反映这一点是很重要的。例如,社会研究者绘制一个变量在1985,1986,1987及1990年的数值图,横轴上在1987和1990间的间隔宽度应是其他年份间隔宽度的3倍。若均匀地隔开,所得到的趋势线会使人误解。图形变形的其他例子将在第二十章加以讨论。

图3.16描绘了从1954年至1988年,美国天主教徒和新教徒参加教堂活动的人数的趋势。时间点的间隔宽度反映出并不是每年都对天主教徒调查了他的出席情况。

在结束描述数据的图形方法的介绍之前,列出几个有益于作图的一般规则。这些规则注重于图的设计和绘制技巧,将有助于你做出更好、含有更多信息的图形。





资料来源: Gallup (盖洛普)

图 3.16 1954 年~1988 年间, 某一周中天主教徒和新教徒的教堂出席率

### 成功作图的一般规则

1. 在作图前确定出你的重点。看图的人应当在图中得到什么信息?
2. 选择图形类型(饼图, 条形图, 直方图等)。
3. 注意题目。图形最重要的一个方面是它的题目。题目将使看图的人了解图形的主旨, 起到画龙点睛的作用。
4. 防止过多的字体型号、风格及颜色改变的应用。不加选择地过度地应用不同的字体类型, 风格和颜色将使看图者感到混乱。一般地, 我们提倡用两种字体, 颜色变化和斜体字仅在一、两个地方用到。
5. 通过颜色和格式的应用传达图形的基调。热烈、温暖的颜色(黄色、橙色和红色)比蓝色、紫色更加生动, 并有助于激起看图者的热情。相反, 柔和的淡色(尤其是灰色)传达一种保守的商业基调。同样, 简单的形式显示保守的调子。然而, 繁复的样式则给人以刺激。
6. 不要低估了一个简单易懂的图形的效果。
7. 经常练习作图, 像做任何事情一样, 熟能生巧。

### 练习

#### 基本技能

3.1(教育) 大学行政官员定期考察各个学院中在校大学生的专业分布, 以便

在学院内的各系之间公平地分配资源。在一次考察中得到如下数据:

系	学生人数
农业	1,500
文理	11,000
工商管理	7,000
教育	2,000
工程	5,000

- a. 做出这些数据的饼图。
- b. 做出这些数据的条形图。

**3.2(商业)** 因为基本原料的进口是美国经济实力的一个指标,商业部对钢铁进口进行管理。下面是 1985 年到 1996 年钢铁进口量的有关数据(以百万吨计):

年份	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
进口量	27.6	22.7	21.9	20.4	19.7	21.9	21.8	21.9	21.8	32.7	27.3	32.1

- a. 饼图是否是描述这些数据的适当的图形方法?
- b. 做出这些数据的条形图。
- c. 你能观察到钢铁进口的有关趋势吗?

**3.3(农业)** 绘出下表中关于生活费用在各类别中的分配的数据图,尝试一下饼图和条形图,哪个更好?

其生活费的去向	百分比
日常产品	13.4
软饮料	8.9
蔬菜和水果	15.6
其他食物	17.5

**3.4(商业)** 俄亥俄州立大学基于对在 45,000 个商业的调查,做了一项大规模的关于就业趋势的研究。若假设失业率为 5% 或更少,该项研究曾经预言,在 1980 年到 1990 年间将有 210 万个就业机会被创造出来。就业的增长可由主要工业群体显示出来。作饼图显示以上数据。

工业群体	1980~1990 年就业增长百分比
服务业	33.2
制造业	25.0
零售业	17.9
金融、保险、不动产	6.6
批发业	4.8
建筑业	4.6
政府	2.7
其他	1.3

**3.5(社会)** 在练习 3.4 所提到的研究中,得到从 1980 年到 1990 年间就业机会的数据,用这些数据做出条形图。

职业	1980~1990 年间就 业机会的百分比
办事员	20.9
管理者	9.5
劳工	3.7
技工	13.1
售货员	7.3
专业技术人员	16.3
服务人员	18.1
手艺人	11.1

**3.6(环境)** 某个州的健康委员会规定:氟化物的含量不能超过 1.5ppm。这儿给出了一个 25 天的样本的氟化物含量。虽然每天都不止一次地测量氟化物的含量,但这些数据都是指抽到的 25 天中早晨的氟化物含量。

0.75	0.86	0.84	0.85	0.97
0.94	0.89	0.84	0.83	0.89
0.88	0.78	0.77	0.76	0.82
0.72	0.92	1.05	0.94	0.83
0.81	0.85	0.97	0.93	0.79

a. 确定这些测量值的极差。

- b. 选择子区间的数目为 7, 把极差除以 7 并四舍五入, 得到宽度为 0.05 的一组分组区间, 用 0.705 作为第一个区间的下限(左端点), 作频数直方图。
- c. 计算每个区间的频率, 做频率直方图。注意到频数和频率直方图的形状相同。
- d. 如果 25 天中随机选择一天, 读数大于 0.90ppm 的机会(概率)是多少? 试猜想(预言)明年有多大比例的天数的氟化物含量大于 0.90ppm?

**3.7(政府)** 国家公路交通安全局研究了汽车后座安全腰带和肩带的应用情况。按照两种带子使用的百分率, 把可能因使用两种安全带而得以救助的人员数量列出如下:

使用的百分率	使用两种安全带而得以救助的人员数	
	只使用腰带	腰带和肩带
100	529	678
80	423	543
60	318	407
40	212	271
20	106	136
10	85	108

用不同的方法作数据图, 并指出哪一种更恰当, 为什么?

**3.8** 作表 3.4 中数据的频数直方图,

- a. 把直方图与 3.11 中的茎叶图作比较。
- b. 用直方图的标准术语描述直方图的形状。
- c. 解释怎样设计一个直方图, 使它和这些数据的茎叶图有相同的形状。

**3.9(社会)** 作下表中数据的一个频率直方图。用直方图的标准术语描述直方图的形状。

人均公共福利消费, 按州的个数列出

美元	州数
50~74	3
75~79	6
100~124	14
125~149	11
150~174	2
175~199	5
200~224	2
225~249	5
250~274	1
275~299	1
总和	50

**3.10(社会)** 下表给出了美国 24 个大城市 1994 年人均城市税的数额。

2470	520	561	488	986	359
1305	512	467	270	360	451
4904	572	498	382	271	634
1682	784	298	643	947	686

资料来源: Statistical Abstract of the United States, 1997

- 用合适的分组区间数作一个频数分布图。
- 作数据的频率直方图。
- 描述频率直方图的形状。
- 若你因调动工作而进入 24 个城市中的一个, 你的城市税多于 900 美元的概率是多少?

**3.11** 作练习 3.10 中数据的茎叶图, 频率直方图和茎叶图哪一个可提供更多的信息? 并解释你的理由。

**3.12** 用一个计算机软件做出练习 3.3 中数据的频率直方图和茎叶图。叙述一下用计算机作图和人工作图的区别。

**3.13(医学)** 严重的慢性左心室疾病有两种不同的治疗方法, 患者用这两种方法治疗的存活时间(月)如下表:

标准疗法							新疗法						
4	15	24	10	1	27	31	5	20	29	15	7	32	36
14	2	16	32	7	13	36	17	15	19	35	10	16	39
29	6	12	18	14	15	18	27	14	10	16	12	13	16
6	13	21	20	8	3	24	9	18	33	30	29	31	27

- 对两种治疗的存活时间分别作频率直方图。
- 对比两个直方图, 说明新疗法是否能延长存活时间, 并解释你的理由。

**3.14** 把练习 3.13 两种疗法中的数据放在一起, 作联合数据的频率直方图。这个图能否暗示数据来自两个不同的总体? 说明你的理由。

**3.15(政府)** 许多公益组织宣称, 联邦政府在国防方面的开支在国家资产中的比重呈上升趋势。下表给出了自 1980 年以来国防部的费用(以 10 亿美元为单位)以及所占国民生产总值(GNP)的百分比(%GNP)。

年份	1980	1981	1982	1983	1984	1985	1986	1987	1988
费用	134	158	185	210	227	253	273	282	290
% GNP	4.9	5.2	5.8	6.1	6.0	6.2	6.2	6.1	5.9
年份	1989	1990	1991	1992	1993	1994	1995	1996	1997
费用	304	299	273	298	291	282	272	266	267
% GNP	5.7	5.3	4.7	4.9	4.5	4.1	3.8	3.6	3.4

资料来源: Statistical Abstract of the United States, 1997

- 绘出开支的时间序列图,并描绘开支的趋势。
- 绘出 %GNP 的时间序列图,并描绘出 %GNP 的趋势。
- 这两个时间序列趋势相同吗? 这两个图都支持公共利益团体的论点吗?

**3.16(教育)** 教育研究者调查了 SAT 分数的趋势,以评估人们关于男女生的考试成绩有所差别的说法。分别做出四个时间序列图。

性别/类别	年 份									
	1967	1970	1975	1980	1985	1990	1993	1994	1995	1996
男/语汇	540	536	515	506	514	505	504	501	505	507
女/语汇	545	538	509	498	503	496	497	497	502	503
男/数学	535	531	518	515	522	521	524	523	525	527
女/数学	495	493	479	473	480	483	484	487	490	492

资料来源: Statistical Abstract of the United States, 1997

- 分别做出四个时间序列图并描述它们的趋势。
- 男生和女生的趋势一样吗?
- 你从图上观察到男女生有何差别?

**3.17(教育)** 有关公立学校系统如何提高办学质量的研究常常指出,教师是需要增加工资收入的一个主要领域。下列是公立学校教师的平均工资(以 1,000 美元计)。用这些数据,作一个时间点均匀分布在横轴上的时间序列图。

年份	1970	1975	1980	1985	1986	1987	1988	1989
工资	8.6	11.7	16.0	23.6	25.2	26.6	28.0	29.6
年份	1990	1991	1992	1993	1994	1995	1996	
工资	31.4	33.1	34.1	35.0	35.7	36.6	37.7	

**3.18** 用练习 3.17 表中的数据作一个时间序列图,要考虑到其中不等的时

间点,并将其与 3.17 中的时间序列图作比较。就“若时间点间的间隔宽度与数据点间的实际长度不符,时间序列将如何使人误解”这一点进行讨论。

**3.19(社会)** 公益调查中为了得到人们的随机样本,任意地拨打电话是常用的抽样方法。通常认为,50 个州内,每千户居民的电话数是基本相同的。用下列附表中的电话数据(每千户居民的电话数)作一个频数直方图。这 50 个州有主要区别吗?

州	电话	州	电话
Alabama	500	Montana	540
Alaska	350	Nebraska	590
Arizona	550	Nevada	720
Arkansas	480	New Hampshire	590
California	610	New Jersey	650
Colorado	570	New Mexico	470
Connecticut	620	New York	530
Delaware	630	N. Carolina	530
Florida	620	N. Dakota	560
Georgia	570	Ohio	550
Hawaii	480	Oklahoma	580
Idaho	550	Oregon	560
Illinois	650	Pennsylvania	610
Indiana	580	Rhode Island	560
Iowa	570	S. Carolina	510
Kansas	600	S. Dakota	540
Kentucky	480	Tennessee	540
Louisiana	520	Texas	570
Maine	540	Utah	560
Maryland	610	Vermont	520
Massachusetts	570	Virginia	530
Michigan	580	Washington	570
Minnesota	560	W. Virginia	450
Mississippi	470	Wisconsin	540
Missouri	570	Wyoming	580

**3.20(社会)** 下表给出了 1985 年及 1996 年各州拥有房产权住户的百分比。这个比例是指所有拥有房产权的居住者与居住者总数的比例。

州	1985	1996	州	1985	1996
Alabama	70.4	71.0	Montana	66.5	68.6
Alaska	61.2	62.9	Nebraska	68.5	66.8
Arizona	64.7	62.0	Nevada	57.0	61.1
Arkansas	66.6	66.6	New Hampshire	65.5	65.0
California	54.2	55.0	New Jersey	62.3	64.6
Colorado	63.6	64.5	New Mexico	68.2	67.1
Connecticut	69.0	69.0	New York	50.3	52.7
Delaware	70.3	71.5	North Carolina	68.0	70.4
Dist. of Columbia	37.4	40.4	North Dakota	69.9	68.2
Florida	67.2	67.1	Ohio	67.9	69.2
Georgia	62.7	69.3	Oklahoma	70.5	68.4
Hawaii	51.0	50.6	Oregon	61.5	63.1
Idaho	71.0	71.4	Pennsylvania	71.6	71.7
Illinois	60.6	68.2	Rhode Island	61.4	56.6
Indiana	67.6	74.2	South Carolina	72.0	72.9
Iowa	69.9	72.8	South Dakota	67.6	67.8
Kansas	68.3	67.5	Tennessee	67.6	68.8
Kentucky	68.5	73.2	Texas	60.5	61.8
Louisiana	70.2	64.9	Utah	71.5	72.7
Maine	73.7	76.5	Vermont	69.5	70.3
Maryland	65.6	66.9	Virginia	68.5	68.5
Massachusetts	60.5	61.7	Washington	66.8	63.1
Michigan	70.7	73.3	West Virginia	75.9	74.3
Minnesota	70.0	75.4	Wisconsin	63.8	68.2
Mississippi	69.6	73.0	Wyoming	73.2	68.0
Missouri	69.2	70.2			

资料来源:美国人口普查局 网址:<http://www.census.gov/ftp/pub/hhes/www/hvs.html>.

- 作出表中列出的 1985 年及 1996 年的数据的频率直方图。
- 这两年的图形存在什么主要区别?
- 你认为, 11 年中图形为什么会有所变化?
- 国会如何利用图中的有关信息, 制定一部允许房产税大幅减免的法律?

**3.21** 做出练习 3.20 的数据的茎叶图。

**3.22** 描述练习 3.20 及 3.21 中房产拥有者比例的直方图及茎叶图的形状, 要用到有关术语如“形态”, “偏度”及“对称”。

**3.23(商业)** 一名汽车设备供应商, 记录了扬声器及扩音器五年中的月销售纪录。如下表中的数据(千台/月)。做出数据图, 你能观察出数据的整体趋势吗? 是



否有周期或季节影响?

年份	J	F	M	A	M	J	J	A	S	O	N	D
1	101.9	93.0	93.5	93.9	104.9	94.6	105.9	116.7	128.4	118.2	107.3	108.6
2	109.0	98.4	99.1	110.7	100.2	112.1	123.8	135.8	124.8	114.1	114.9	112.9
3	115.5	104.5	105.1	105.4	117.5	106.4	118.6	130.9	143.7	132.2	120.8	121.3
4	122.0	110.4	110.8	111.2	124.4	112.4	124.9	138.0	151.5	139.5	127.7	128.0
5	128.1	115.8	116.0	117.2	130.7	117.5	131.8	145.5	159.3	146.5	134.0	134.2

**3.24(商业)** 一家为制造商提供各种产品的机械工具公司以季节为单位记录了前8年的生产活动。表中的数据反映了有关活动而不是价格,与通货膨胀无关。

年份	季 节			
	1	2	3	4
1	97.2	100.2	102.8	102.6
2	106.1	107.8	110.5	110.6
3	116.5	117.3	119.9	119.3
4	126.1	125.7	128.3	132.1
5	133.2	133.8	141.1	142.1
6	144.2	146.1	151.6	154.0
7	155.8	158.6	165.8	167.0
8	171.1	172.6	176.5	179.7

- 作出上述数据对时间的图(第1—第32个季度)。
- 是否有清晰的趋势?若有,你应选择什么样的趋势方程?
- 你能观察出其中的周期或季节特征吗?

### 3.4 单个变量数据的描述:中心趋势的度量

数值的描述方法常用来传达对图形、物体及其他现象的理性反映。之所以如此,有两个主要原因。首先,图形描述方法对于统计推断是不太恰当的,因为它很难描述样本频数直方图和相应的总体频数直方图之间的类似之处。用数据描述的第二个理由是出于便利——我们不可能随身带着适当的图形或直方图,必须借助于语言表达能力去进行恰当的描绘。我们寻求几个数字,称之为数值描述性度量,这些量将使我们在一组测量值的频率分布有一个大致的印象。

最常用的两类数值描述性度量是**中心趋势度量**和**变异性度量**。也就是说,我

们希望描述测量值分布的中心,并弄清测量值是如何相对于分布中心变化的。为了把总体的数值描述性度量和样本的数值描述性度量区别开来,称前者为**参数**,后者为**统计量**。在统计推断的有关问题中,不能计算各种参数的数值,但可以计算来自样本的相应的统计量,并用所得到的数值去估计相应的总体参数。

在这一节中,我们将讨论各种中心趋势度量,下一节(3.5节)将讨论变异性度量。

第一种中心趋势度量是**众数**。

**定义 3.1** 一组测量值的**众数**是指出现次数最多(有最高频数)的测量值。

下面举例说明众数的应用及其确定方法。

### 例 3.1

抽取 15 头食用牛作为一个样本,记录下它们的屠宰重量(单位:英镑)。数据如下:

962	1, 005	1,033
980	965	1,030
975	989	955
1,015	1, 000	970
1,042	1,005	995

求屠宰重量的众数。

**解答** 对以上数据,1,005 这一重量出现两次,其他的重量只出现一次,因此,众数是 1,005。

例 3.1 中众数的确定是很容易的,因为我们可以数出每一个测量值出现的次数。当处理分组数据——以频数表的形式出现的数据的时候,我们可以把具有最高频数的分组区间定义为众数区间。但是,由于只知道有多少测量值落入每一个区间,而不知道实际的测量值,就把众数区间的中点当作众数。它是实际样本测量值的众数的近似值。

众数常被认为是人气所在,反映中心倾向或观点。例如,我们可以谈论最受欢迎的股票,最受喜爱的洗衣机样式或最受爱戴的候选人。此中均指分布的众数。在前一节图 3.9 中,频数直方图(b)、(c)及(d)都有众数,它位于有最高频数的区间的中心。因此,直方图(b)的众数是 0.25,(c)是 3,(d)是 17。还应注意,有些分布不止一个具有最高频数的测量值。比如,我们可能遇到双峰、三峰分布等。图 3.8 给出的小鸡数据的频率直方图是双峰分布,众数是 4.2g 和 4.4g。在图 3.9 中,直方图(e)和(f)都是双峰的。

我们要谈到的第二个中心趋势的度量是**中位数**。

**定义 3.2** 一组测量值的**中位数**是指把这些测量值从小往大排列时的中间值。

中位数最常用于衡量一大组测量值的中点。例如,我们常谈到工会成员争取到的工资增长的中位数,收到社会福利保障的人员的年龄的中位数,在某个月中屠宰的牛的重量的中位数。这些情况中的每一种都涉及到一大组数据,中位数反映了数据的中心值,即中位数把这些测量值分为两组,每一组都有相等数目的测量值。

然而,对于小的测量值集合,可以据中位数的定义做如下约定:对于偶数个测量值的中位数,是将其从小到大排列后中间两个数值的平均;奇数个测量值的中位数仍是中间值。因此,无论在偶数个还是奇数个测量值的情况,在中位数两侧都有等个数的测量值。

### 例 3.2

对二年级的十个孩子做阅读能力测试,分数如下:

95 86 78 90 62 73 89 92 84 76

写出测试分数的中位数。

**解答** 首先,按大小顺序排列这些分数,由于是偶数个测量值,中位数是中间两数的平均:

$$\text{中位数} = (84 + 86) / 2 = 85$$

### 例 3.3

作一个试验以考察修剪葡萄的新方法的效果。有 13 个工人,每人分了修剪一英亩的任务。生产纪录如下(以人-小时/英亩计)。

4.4 4.9 4.2 4.4 4.8 4.9 4.8 4.5 4.3 4.8 4.7 4.4 4.2

写出这些值的众数和中位数。

**解答** 按从小到大的顺序排列这些测量值:

4.2 4.2 4.3 4.4 4.4 4.4 4.5 4.7 4.8 4.8 4.8 4.9 4.9

这组数中有两个测量值出现了三次,因此,数据是双峰的,众数是 4.4 和 4.8。奇数个测量值的中位数是中间的值,即 4.5。

**分组数据的中位数**较难以计算。中位数位于一个特定的分组区间中,但由于测量的实际值未知,我们无法确定中位数位于这个区间的何处。若假设测量值在整个区间上是均匀分散的,可得到如下结论。令

$L$  = 包含中位数的分组区间的下限

$n$  = 总频数

$cf_b$  = 包含中位数的区间前所有区间的频数之和(累积频数)

$f_m$  = 包含中位数的分组区间的频数

$w$  = 区间宽度

则对于分组数据,

$$\text{中位数} = L + \frac{w}{f_m}(0.5n + cf_b)$$

下例说明如何寻找分组数据的中位数.

### 例 3.4

表 3.5 是表 3.3 中小鸡数据的频数表.由这些数据计算小鸡增重的中位数。

表 3.5 表 3.3 的小鸡数据频数表

分组区间	$f_i$	累加 $f_i$	$f_i/n$	累加 $f_i/n$
3.55~3.65	1	1	0.01	0.01
3.65~3.75	1	2	0.01	0.02
3.75~3.85	6	8	0.06	0.08
3.85~3.95	6	14	0.06	0.14
3.95~4.05	10	24	0.10	0.24
4.05~4.15	10	34	0.10	0.34
4.15~4.25	13	47	0.13	0.47
4.25~4.35	11	58	0.11	0.58
4.35~4.45	13	71	0.13	0.71
4.45~4.55	7	78	0.07	0.78
4.55~4.65	6	84	0.06	0.84
4.65~4.75	7	91	0.07	0.91
4.75~4.85	5	96	0.05	0.96
4.85~4.95	4	100	0.04	1.00
总计	$n = 100$		1.00	

**解答** 令  $j$  组的累加频率等于第 1 组到第  $j$  组的频率之和。为了确定包含中位数的区间,需要找出累加相关频数超过 0.50 的第一个区间,此即含中位数的区间。对于以上数据,区间 4.25~4.35 是第一个频率超过 0.50 的区间,如表 3.5 中第 5 列所示。所以这个区间包含中位数。于是

$$L = 4.25 \quad f_m = 11$$

$$n = 100 \quad w = 0.1$$

$$cf_b = 47$$

从而

$$\text{中位数} = L + w(0.5n - cf_b)/f_m = 4.25 + 0.1(50 - 47)/11 = 4.28$$

注意从表 3.2 中的未分组数据得到的中位数是 4.3, 因此, 近似值与来自未分组数据的值几乎相等。样本中位数的这两个值之间的差异会随分组区间的增加而减小。

第三个, 也是最后要讨论的中心趋势度量是算术平均值, 简称之为均值。

**定义 3.3** 一组测量值的算术平均值或均值定义为: 测量值的总和除以测量值的总个数。

当人们谈论平均数时, 常指均值, 它是数据集合的重心。由于均值在后面章节的统计推断中的重要角色, 我们给样本均值及总体均值以特定的符号, 总体均值用希腊字母  $\mu$  表示(按英文读作 mu), 样本均用符号  $\bar{y}$  表示(读作 y-bar)。如第一章中所指出的, 一个测量总体是我们感兴趣的测量值的整个集合, 而一个测量样本是从感兴趣的总体中选取的测量值的子集合。如果以  $y_1, y_2, \dots, y_n$  表示容量为  $n$  的测量样本的观察值, 则样本均值记为

$$\bar{y} = \frac{\sum y_i}{n}$$

其中出现在分子中的符号  $\sum y_i$  表示  $n$  个测量值  $y_i$  的总和:

$$\sum_i y_i = y_1 + y_2 + \dots + y_n$$

对应的总体均值是  $\mu$ 。

在多数情况下, 总体均值未知, 此时常用样本均值对相应的总体均值作出推断。例如, 一家大型连锁店的会计科对拖欠的账目进行清查。这家商店有几千笔这样的账目, 这就产生了拖欠数额均值为  $\mu$  的拖欠账目的总体。 $\mu$  的数值只有通过花几天的时间进行大规模地查账才能确定。通过在每天的拖欠账目中随机抽取  $n$  笔账目, 计算样本均值  $\bar{y}$ , 并用这个样本均值  $\bar{y}$  估计当日所有的拖欠账目的均值  $\mu$ , 会计可以检查每天的逾期账目。估计的准确度及确定合适的样本容量的方法将在第五章加以讨论。

### 例 3.5

一家大型商店抽取了容量  $n = 15$  的拖欠账目的样本, 下表给出了这些账目拖欠款的数目(美元):

55.20	4.88	271.95
18.06	180.29	365.29
28.16	399.11	807.80
44.14	97.47	9.98
61.61	56.89	82.73

- 求出所抽取的 15 笔账目的平均拖欠款。
- 若共有 150 笔拖欠账目,用样本均值去估计所有 150 笔拖欠账目的总数。

### 解答

- 样本均值计算如下:

$$\bar{y} = \frac{\sum y_i}{15} = \frac{55.20 + 18.06 + \cdots + 82.73}{15} = \frac{2483.56}{15} = 165.57 (\text{美元})$$

- 从(a)中看出,抽取到的 15 笔账目拖欠款的平均数为 165.57 美元,由此可以推断,150 笔账目的总计拖欠款为:  $150(165.57) = 24,835.50$  美元。

分组数据的样本均值公式比刚提到的未分组数据的公式稍微复杂些。在某些情况下,原始数据会在频数表或直方图中给出,因此,个体的样本测量值未知,只能知道它所在的区间。在这种情况下,对于分组数据的均值公式只是实际样本均值的近似公式。因此,当样本测量值已知时,应使用未分组数据的公式。若有  $k$  个分组区间,且

$y_i$  = 第  $i$  个区间的中点

$f_i$  = 第  $i$  个区间的频数

$n$  = 测量值的总数

那么,

$$\bar{y} \cong \frac{\sum f_i y_i}{n}$$

这儿,  $\cong$  表示大约相等。

### 例 3.6

表 3.6 是用例 3.4 中的数据经分组作成的,并在其中增添了三栏—— $y_i$ ,  $f_i y_i$  和  $f_i (y_i - \bar{y})^2$ ——这将有助于计算概括统计量的近似值。试计算这组分组数据的样本均值。

表 3.6 小鸡数据

分组区间	$f_i$	$y_i$	$f_i y_i$	$f_i (y_i - \bar{y})^2$
3.55~3.65	1	3.6	3.6	0.478864
3.65~3.75	1	3.7	3.7	0.350464
3.75~3.85	6	3.8	22.8	1.452384
3.85~3.95	6	3.9	23.4	0.921984
3.95~4.05	10	4.0	40.0	0.852640
4.05~4.15	10	4.1	41.0	0.368640
4.15~4.25	13	4.2	54.6	0.110032
4.25~4.35	11	4.3	47.3	0.000704
4.35~4.45	13	4.4	57.2	0.151632
4.45~4.55	7	4.5	31.5	0.302848
4.55~4.65	6	4.6	27.6	0.569184
4.65~4.75	7	4.7	32.9	1.165248
4.75~4.85	5	4.8	24.0	1.290320
4.85~4.95	4	4.9	19.6	1.478656
总和	100		429.2	9.493600

**解答** 把  $f_i y_i$  栏中的值相加之后, 将其带入公式, 得到样本均值是

$$\bar{y} = \frac{\sum f_i y_i}{100} = \frac{429.2}{100} = 4.292$$

样本均值的实际值为

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{429.2}{100} = 4.292$$

这正是用分组数据公式得到的实际值。从未分组数据得到的样本均值和从分组数据得到的样本均值的数值并不总是相同的。然而, 当分组区间的数目相当大时, 这两个值相当接近。

均值是对一组测量值中心的常用的度量, 但它会由于在集合中一个或多个极端值的出现而发生失真。在这样的情况下, 极端值(又称作离群值)会使均值偏向自己一方以找到数据的平衡点, 因此而歪曲了均值作为中心值度量的意义。对均值的一种变通的方法是截尾均值, 即去掉最大和最小的若干数值, 对其余的数作平均。例如, 5%的截尾均值就是去掉 5% 最大的测量值和 5% 最小的测量值, 然后对其余的 90% 的测量值求平均。同样, 10% 的截尾均值则是去掉 10% 最大和最小的数值, 平均其余的数值。在例 3.5 中, 要求 10% 的截尾均值, 就要去掉最大和最小值, 得到均值:

$$\bar{y} = \frac{2,483.56 - 4.88 - 807.8}{13} = 128.53(\text{美元})$$

通过剪裁数据,可以减少特别大(或特别小)的值对均值的影响,得到集合中心的更可靠的度量。当样本均值用来预测总体均值时,这一点尤其重要。

注意,在极限的意义上,中位数是 50% 的截尾均值。因此,当数据集合中有极端值时,中位数常被用来代替均值。例 3.5 中,值 807.80 比起其他值是相当大的,这导致 15 个数中有 10 个小于均值,仅有 5 个大于均值。15 个数的中位数是 61.61。有 7 个数比中位数小,7 个数比中位数大。因此,在选择有代表性的拖欠账目时,中位数比均值更合适。然而,若我们想估计 150 笔账目的拖欠款总数时,将用均值而不是中位数。当估计总体中所有测量值之和时,不必排除样本中的极值。假设样本中含有一些极大的值,若这些极大的值被截掉了,那么,用样本截尾均值或用样本中位数来代替样本均值估计总体之和,将严重低估被估计的量。

在本节中,我们讨论了众数,中位数,均值和截尾均值。对于某组测量值,这些中心趋势度量之间有何联系呢? 答案依赖于数据的**偏倚程度(偏度)**。若分布是丘形的,并且关于单峰对称,则众数( $M_0$ ),中位数( $M_d$ ),均值( $\mu$ )和截尾均值(TM)是相同的。这可以用图 3.7(a)中的光滑曲线及总体的各中心趋势度量来说明。若分布是偏斜的、有一长尾且单峰的,均值被拖向长尾的方向,中位数落在众数和均值之间,截尾均值常落于中位数和均值中间,其位置依赖于截尾的比例。图 3.17(b)和(c)分别给出了左偏和右偏的分布。

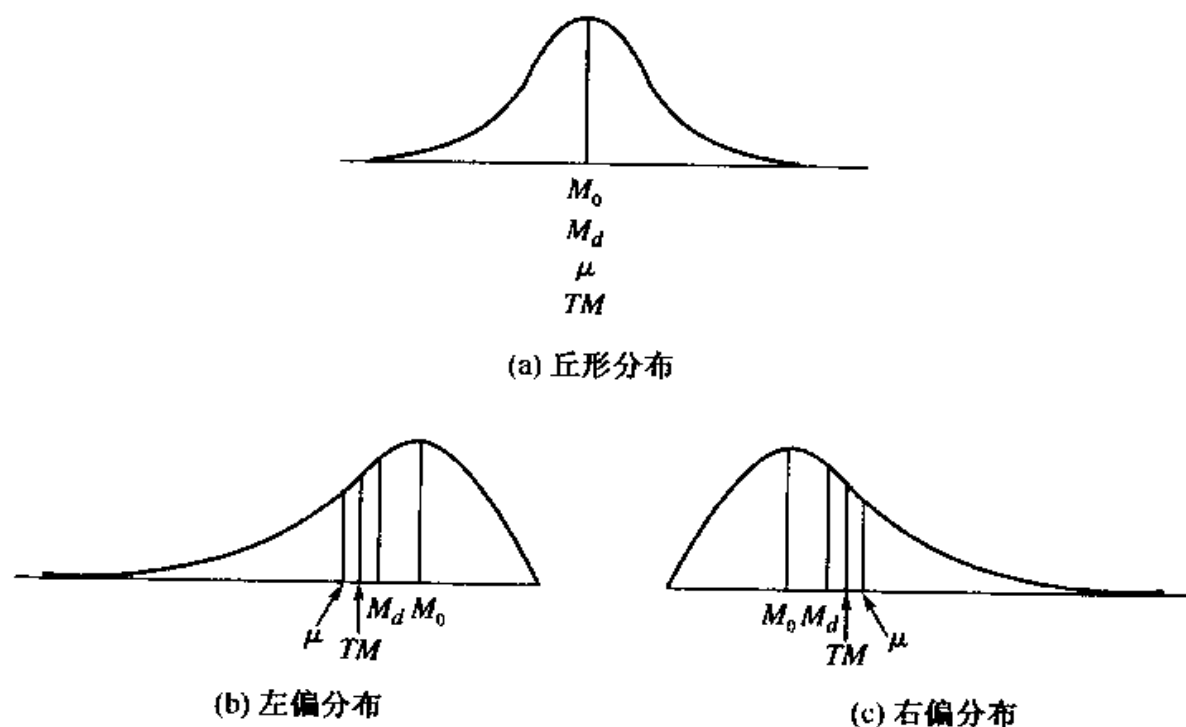


图 3.17 均值、截尾均值、中位数、众数间的关系



要记住的重要一点是：我们不能局限于仅用一种中心趋势度量。对某些数据集合，有必要用多种度量，才能对数据的中心趋势作出准确的描述性的概括。

### 中心趋势的各种度量的性质

#### 众数

1. 它是数据集合中出现最多，或者说出现的可能性最大的测量值。
2. 一个数据集合可能存在不止一个众数。
3. 不受极端测量值的影响。
4. 子集的众数不能联合起来确定整个集合的众数。
5. 对于分组数据，众数的值随所用的类别而改变。
6. 对于定性数据和定量数据均适用。

#### 中位数

1. 它是中心值，50%的测量值比它大，50%的测量值比它小。
2. 一个数据集合仅有一个中位数。
3. 不受极端测量值的影响。
4. 子集合的中位数放在一起不能用来确定整个集合的中位数。
5. 对分组数据，即使用不同的方法对数据进行分类，它的值也是相当稳定的。
6. 仅适用于定量数据。

#### 均值

1. 它是数据集合中测量值的算术平均值。
2. 一个数据集合仅有一个均值。
3. 其值受极端测量值的影响。截尾有助于减少这种影响。
4. 子集合的均值放在一起可以决定整个集合的均值。
5. 仅适用于定量数据。

中心趋势度量并未为我们提供一幅关于一组测量值频数分布的完整的描绘。除了确定分布的中心，我们还必须有数据的分散性的度量。在下一节，我们将讨论变异性度量或离中趋势。

### 练习

#### 基本技能

3.25 计算下列数据的均值，中位数及众数。

11 17 18 10 22 23 15 17  
14 13 10 12 18 18 11 14

3.26 将练习 3.25 中的测量数据 22 和 23 用 42 和 43 代替，重新计算均值，中位数及众数。讨论这些极端值对三个中心趋势度量的影响。

3.27 参见练习 3.25 和 3.26。计算两组数据的 10% 的截尾均值。极端值

对 10% 的截尾均值有影响吗? 5% 的截尾均值呢?

**3.28** 确定下列测量值的中位数、众数和均值。

10 2 1 5  
1 5 7 10  
3 4 8 12  
5 6 8 9

**3.29** 确定下列频数表中数据的均值, 中位数及众数。

分组区间	频数
0~2	1
3~5	3
6~8	5
9~11	4
12~14	2

### 应用

**3.30(工程)** 在一项汽车的可靠性试验中, 检查了 191 辆汽车的可靠性[参见“Large sample simultaneous confidence intervals for the multinomial probabilities on transformation of the cell frequencies,” *Technometrics* (技术度量)(1980)22:588], 记录了第一次重大发动机故障之前所行驶的里程(以 1,000 英里计), 并将之分类, 得到下表:

行驶距离(1,000 英里)	频数
0~20.0	6
20.1~40.0	11
40.1~60.0	16
60.1~100.0	59
100.1~120.0	46
120.1~140.0	33
140.1~160.0	16
160.1~200.0	4

- 作里程数的频率直方图并描述其形状。
- 估计 191 辆汽车行驶里程的众数, 中位数和均值。
- 三种数据中心度量趋势间的关系对于其频数直方图的形状暗示了什么?
- 你认为三种度量中哪一种是这 191 辆汽车行驶里程的最恰当的描述? 并

解释你的理由。

**3.31(医药)** 在美国统计学家 (*American Statistician*) [(1974), 28: 115 ~ 122] 刊登的一篇报告中, 调查了 1,329 名男性, 将其按血清胆固醇的含量及血压分类。把 408 名血压低于 127mmHg 的人又按血清中胆固醇的含量分类, 结果如下表。

血清胆固醇(mg/100cc)	频数
0.0~199.9	119
200.0~219.9	88
220.0~259.9	127
大于 259.9	74

- 估计血清胆固醇读数的众数, 中位数及均值(若可能)。
- 三种概括统计量中, 哪一种关于这群人的血清胆固醇的含量包含更多的信息? 说明你的理由。

**3.32(环境)** 鸟蛋中 DDE(与 DDT 有关)与 PCB 的浓度比在生态学上是一个很有意义的量。这一比例可用作污染通过食物链传播的指标。文章“大湖的银鸥蛋中 DDE 与 PCB 的浓度比及其在解释环境污染数据中的应用”[五大湖研究杂志 (*Journal of Great Lakes Research*) (1998)24(1): 12~31] 报告了从五大湖的十三个地点收集的鸟蛋中的这一比值。这些鸟蛋来自于陆地及水生鸟类。

DDE 与 PCB 的比											
陆地喂养	76.50	6.03	3.51	9.96	4.24	7.74	9.54	41.70	1.84	2.50	1.54
水生喂养	0.27	0.61	0.54	0.14	0.63	0.23	0.56	0.48	0.16	0.18	

- 忽略饲养方式, 计算 21 个数值的均值和中位数。
- 分别计算每种饲养方式下的均值和中位数。
- 通过(a)(b)中的结论, 评论一下均值和中位数对于数据集合中极端值的相对敏感性。
- 对于这两种饲养方式, 你认为中位数和众数哪一个是最合适的测量? 为什么?

**3.33(医学)** 在 Woolson 和 Lachenbruch [生物统计学 (*Biometrika*) (1980) 67: 597~606] 的文章中, 调查了烧伤患者移植皮肤的存活时间。有两个患者在移

殖皮肤败坏前退出,他们的移植皮肤的存活时间比所给值要长。

存活时间(天):37, 19, 57\*, 93, 16, 22, 20, 18, 63, 29, 60\*

(“\*”表示患者在败坏前退出,给出的值是患者退出研究时的天数)

- a. 若可能的话,计算 11 个患者的中心趋势度量。
- b. 若退出研究的患者的存活时间可得到,这些新值如何改变了(a)中算出的概括统计量的值?

**3.34(机械)** 为了研究柴油发动机的可靠性,对 14 台发动机进行了试验。在实验室中运转这些发动机,试验进行了 300 天后结束。下面给出了各台发动机到失效时运转的天数(天),其中那些在试验中没有失效的发动机用 300\* 记,即这些发动机失效天数大于 300 天。

失效时间(天)130, 67, 300\*, 234, 90, 256, 87, 120, 201, 178, 300\*, 106, 289, 74

- a. 计算 14 台发动机的中心度量。
- b. 当一部分发动机的准确失效天数未知,计算的中心度量值意味着什么?

**3.35(农业)** 氮是许多植物结实的制约因素,尤其是苹果树的生长直接与树叶中氮的含量有关,必须仔细管理以保护果园中的果树。研究表明,氮含量大致为 2.5% 可获得最好的收成。注意,一些研究人员在报告他们的结论时用了 ppm (百万分之一), $1\% = 10,000\text{ppm}$ 。

为了确定果园里树中氮的含量,从整个果园中的树上剪下生长有 150 个叶子的树梢。把这些叶子磨碎形成混合样品,研究人员通过化验这些样品来获得氮的百分比。在全州随机抽取了 36 个果园,这些果园的混合样品的含氮量给出如下:

2.0968	2.8220	2.1739	1.9928	2.2194	3.0926
2.4685	2.5198	2.7983	2.0961	2.9216	2.1997
1.7486	2.7741	2.8241	2.6691	3.0521	2.9263
2.9367	1.9762	2.3821	2.6456	2.7678	1.8488
1.6850	2.7043	2.6814	2.0596	2.3597	2.2783
2.7507	2.4259	2.3936	2.5464	1.8049	1.9629

- a. 将每个测量值四舍五入到百分位。
- b. 确定四舍五入后的样本众数。
- c. 确定四舍五入后的样本中位数。
- d. 确定四舍五入后的样本均值。

**3.36** 对于练习 3.35 中四舍五入到百分位后的数据,用 29.40 代替 2.94,计算这些数据的样本方差,中位数及众数,并与练习 3.35 得到的结论作比较。

**3.37(政府)** 对于三组大城市中居民财产的有效税率(每 100 美元),按居民财产税率的大小排列,给出如下数据。

第 1 组	税率	第 2 组	税率	第 3 组	税率
Detroit, MI	4.10	Burlington, VT	1.76	Little Rock, AR	1.02
Milwaukee, WI	3.69	Manchester, NH	1.71	Albuquerque, NM	1.01
Newark, NJ	3.20	Fargo, ND	1.62	Denver, CO	0.94
Portland, OR	3.10	Portland ME	1.57	Las Vegas, NV	0.88
Des Moines, IA	2.97	Indianapolis, IN	1.57	Oklahoma City, OK	0.81
Baltimore, MD	2.64	Wilmington, DE	1.56	Casper, WY	0.70
Sioux Falls, IA	2.47	Bridgeport, CT	1.55	Birmingham, AL	0.70
Providence, RI	2.39	Chicago, IL	1.55	Phoenix, AZ	0.68
Philadelphia, PA	2.38	Houston, TX	1.53	Los Angeles, CA	0.64
Omaha, NE	2.29	Atlanta, GA	1.50	Honolulu, HI	0.59

资料来源:哥伦比亚区政府,金融税收部,哥伦比亚特区的税率及税务负担:全国对比,年鉴(Tax Rates and Tax Burdens in the District of Columbia: A Nationwide Comparison, annual)

- 分别计算三组数据的均值,中位数和众数。
- 计算全体 30 个测量值的均值,中位数和众数。
- 哪些度量最好地概括了这些分布的中心? 给出解释。

3.38 参见练习 3.37。对三个组均值,三个组中位数和三个组众数,分别求平均,并把所得结果与(b)作比较,并阐述你的发现。

### 3.5 单个变量数据的描述:变异性度量

仅用中心度量如均值和中位数描述数据是不充分的。例如,我们要检验一名义上厚度为 3mm 的塑料布,若随机从工厂产品中选择 100 张发现平均厚度为 3mm。这能说明所有的塑料布厚度如我们所希望的为 3mm 吗? 有可能出现这样的情况,即有 50 张塑料布的厚度为 1mm,其余 50 张的厚度为 5mm。此时得到的平均厚度为 3mm。因此,我们需要确定塑料布的厚度是如何在均值 3mm 的周围分散开来的。

利用图 3.18 中的图形,我们可以看出引入变异性度量的必要性。在图 3.18 的几个频率直方图中,其均值都相同,但每个图都有不同的分散程度,或者说变异性都不同。在图中,为了便于说明,我们已经把直方图用光滑曲线表示出来。假设三个直方图描绘了大量一升水的样本中 PCB(多氯化联二苯,这里以 ppb 为单位)的含量,这些样本分别取自于靠近化工厂的三个湖中。三个湖中一升水里 PCB 的平均数  $\mu$  是相等的。然而,PCB 含量的变异性的差别是相当大的。与直方图(c)所描述的 PCB 值相比,直方图(a)中有较少的样本含有很少或很多的 PCB。仅知道三个湖中 PCB 含量的均值将误导调查者对当前三个湖中 PCB 含量的认识。

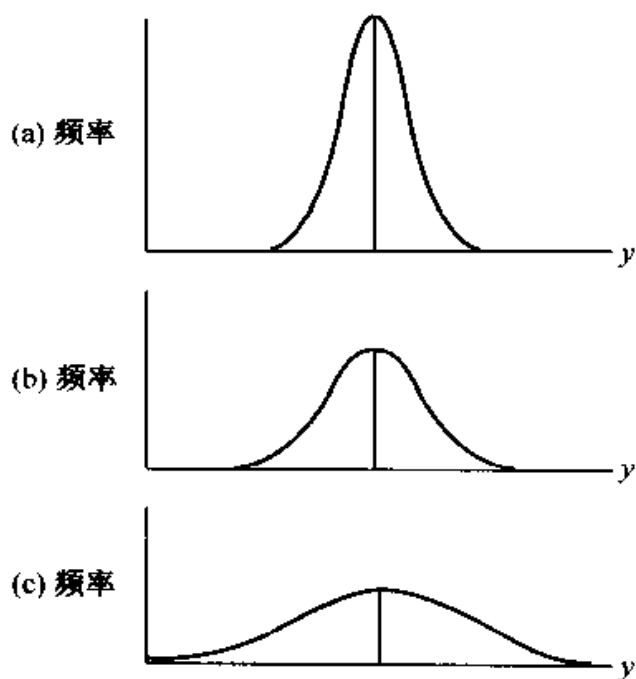


图 3.18 变异性不同但均值相同的频率直方图

**极差**是最简单但也用得最少的数据变异性度量,在 3.2 节中曾提到过,现给出其定义。

**定义 3.4** 一组测量值的**极差**是指此集合中最大与最小测量值之间的差。

### 例 3.7

确定例 3.5 中 15 笔拖欠款项的极差。

**解答** 最小测量值是 4.88, 最大测量值是 807.80。因此,极差是

$$807.80 - 4.88 = 802.92 (\text{美元})$$

对于**分组数据**,由于不知道个体的测量值,**极差**取为最后一个区间上的上限和第一个区间的下限之间的差。

虽然极差易于计算,但它依赖于最极端的值,所以很易受离群值的影响。极差并未给出变异模式的更多信息。对于例 3.5 描述的情形,若在目前的预算期内 15 笔拖欠账目中有 10 笔为 4.88 美元, 3 笔为 807.80 美元, 1 笔为 5.36 美元, 1 笔为 6.00 美元,则均值将是 165.57 美元,,极差是 802.92 美元,均值和极差都与例 3.5 中相等。然而,目前预算期中的数据相对于均值的分散程度比以前预算期中的数据大。我们要寻求的是一种能够辨别数据在均值附近集中程度的变异性度量。

第二种变异性的度量方法是用**百分位数**。

**定义 3.5**  $n$  个按大小排列的测量值集合的  $p\%$  分位数是指这样一个数值, 集合中有至多  $p\%$  的测量值比它小, 有至多  $(100 - p)\%$  的测量值比它大。

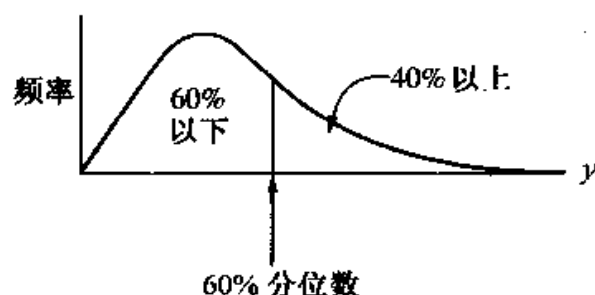


图 3.19 一组测量值的 60% 分位数

例如, 图 3.19 说明了一组测量值的 60% 的分位数。百分位数常用来说明考试成绩的结果以及和其他参加考试的人相比某个人的排名。人们特别感兴趣的百分位数是 25%, 50%, 75% 分位数, 分别称之为 *四分之一分位数*, *二分之一分位数* (*中位数*), *四分之三分位数*。(见图 3.20)

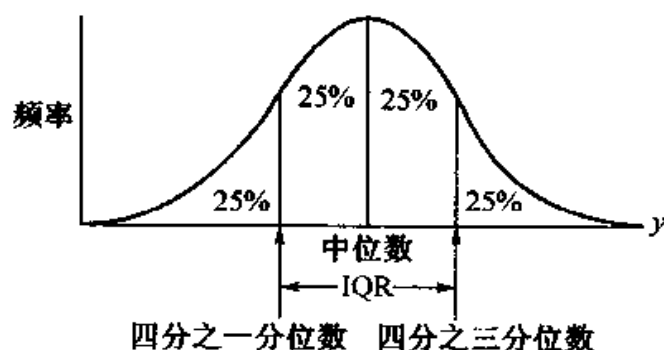


图 3.20 分布的分位数

百分位数计算如下: 每一个数据值都对应于一个百分位数, 这个百分位数的百分比等于数据集合中值小于或等于它的数据值所占的百分比。记  $y_{(1)}, y_{(2)}, \dots, y_{(n)}$  为数据集合的顺序观察值, 也就是,

$$y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$$

第  $j$  个观察值  $y_{(j)}$  对应于  $100(j - 0.5)/n$  百分位数, 用公式  $100(j - 0.5)/n$  代替  $100j/n$ , 以避免  $y_{(n)}$  作为 100% 分位数, 否则, 将意味着我们在这个数据集合里观察到了总体中可能的最大数据取值, 而这是不大可能发生的事。例如, 在一项血清胆固醇含量 (mg/l) 的研究中, 记录了 20 个成年患者的胆固醇含量, 数据如下表。于是, 每个顺序观察值都对应于  $100(j - 0.5)/n = 100(2j - 1)/2n = 100(2j - 1)/40$  的一个百分位数。

观察值( $j$ )	胆固醇含量(mg/l)	百分位数
1	133	2.5
2	137	7.5
3	148	12.5
4	149	17.5
5	152	22.5
6	167	27.5
7	174	32.5
8	179	37.5
9	189	42.5
10	192	47.5
11	201	52.5
12	209	57.5
13	210	62.5
14	211	67.5
15	218	72.5
16	238	77.5
17	245	82.5
18	248	87.5
19	253	92.5
20	257	97.5

22.5%的分位数是 152(mg/l)。因此,在研究中 22.5%的人血清胆固醇的含量小于等于 152。同样,上面集合的中位数,即 50%的分位数,是 192 和 201 的中点。也就是说,中位数  $= (192 + 201)/2 = 196.5$ 。因此,在本项研究中大约一半的人胆固醇含量低于 196.5,一般的人胆固醇含量高于 196.5。

当处理大量数据时,常将百分位数推广到分位数。所谓分位数,记作  $Q(u)$ ,是这样的一个数,它把容量为  $n$  的样本分为两组,使得在  $n$  个数据值中有  $u$  这么大的比例小于或等于该数,即  $Q(u)$ 。分位数  $Q(u)$  作为  $u$  的函数的图像,提供了估计样本所在的总体的分位数的方法。我们可以用下列步骤作出分位数图:

1. 在图形横轴上标上包含  $(0, 1)$  区间的刻度。
2. 在图形纵轴上标上刻度,并使之包含从最小值  $y_1$  到最大值  $y_n$  的范围。
3. 画出点  $(u_i, y_{(i)})$ , 其中  $u_i = (i - 0.5)/n = (2i - 1)/(2n)$ ,  $i = 1, \dots, n$ 。

用 Minitab 软件我们可以得到胆固醇数据的图 3.21。注意,用 Minitab 软件画出的图,纵轴标注的是  $Q(u)$ ,而不是  $y_{(i)}$ 。对  $u$  画出  $y_{(i)}$  得到分位数图。可以从图上读出特定的分位数。

对任意值  $u$ , 用如下方法可得到相应的分位数。首先,通过分位数图上的点



作出一条光滑曲线, 然后读出所希望的  $u$  值所对应的数值。

为了解释这个方法, 假设我们想确定胆固醇的 80% 的分位数——也就是说, 在总体中有 80% 的人胆固醇的含量低于该值, 即  $Q(0.80)$ 。

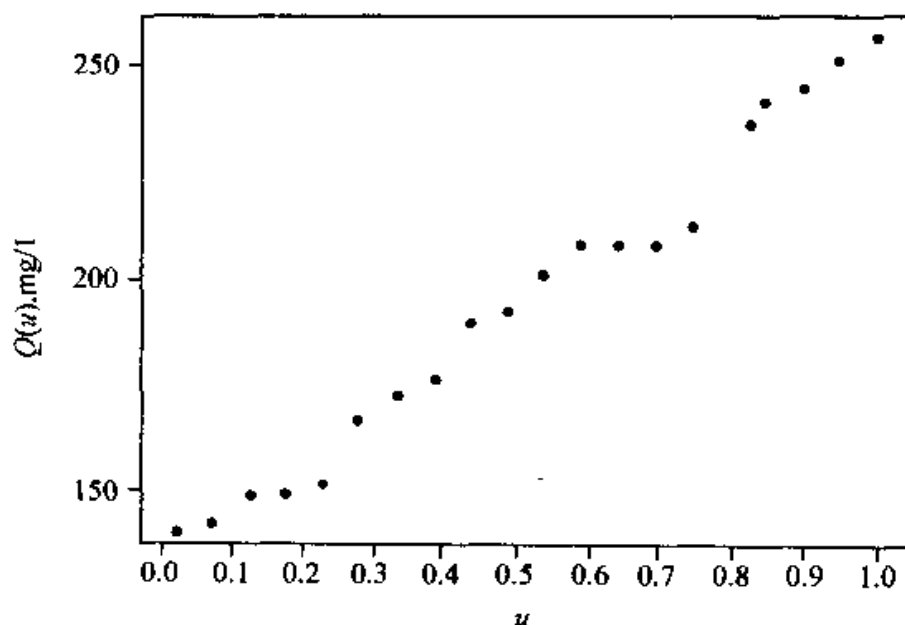


图 3.21 胆固醇数据的分位数图

在图 3.21 中, 找出横轴上  $u = 0.80$  的点, 然后作垂线与分位数图相交, 再作水平线与纵轴相交, 则交点就是 80% 的分位数的估计值。(见图 3.22)。由此可以粗略地说, 大约 80% 的人胆固醇含量低于 243。

对于分组以后的数据, 可用下列公式近似原始数据的百分位数。令

$P$  = 要求的百分位数

$L$  = 包含所求的百分位数的分组区间的下限

$n$  = 总频数

$cf_b$  = 百分位数所在区间前的所有分组区间上的累积频数

$f_p$  = 包括所要求的百分位数的分组区间上的频数

$w$  = 区间宽度

那么, 以 65% 的分位数为例, 将用下面这个公式计算:

$$P = L + \frac{w}{f_p}(0.65n - cf_b)$$

为了确定  $L$ ,  $f_p$  和  $cf_b$ , 从最低区间开始, 找出累积频率超过 0.65 的第一个区间, 这个区间将包含 65% 分位数。

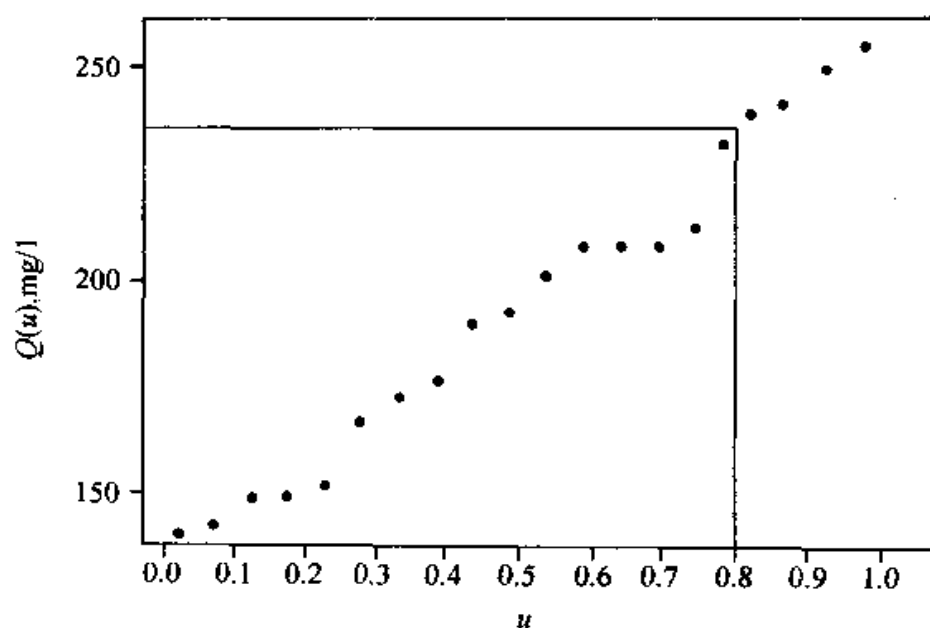


图 3.22 胆固醇数据的 80% 的分位数

**例 3.8**

对表 5.3 中的小鸡数据计算 90% 分位数。

**解答** 由于第 12 个区间是第一个累积频率超过 0.90 的区间, 我们有,

$$L = 4.65 \quad n = 100 \quad cf_b = 84 \quad f_{90} = 7 \quad w = 0.1$$

因此, 90% 的分位数是:

$$P_{90} = 4.65 + 0.1 [0.9(100) - 84] / 7 = 4.74$$

这意味着 90% 的测量值在该值以下, 10% 的测量值在该值以上。

现定义变异性的第二个度量——**四分位数间距**。四分位数间距的另一个略有区别的定義随盒形图给出(3.5 节)。

**定义 3.6** 一组测量值的**四分位数间距(IQR)**是指在四分之三和四分之一分位数之间的差异, 即

$$IQR = 75\% \text{ 的分位数} - 25\% \text{ 的分位数}$$

虽然四分位数间距对于数在中点的集中比极差更敏感, 但对于我们的研究目的而言仍是不够充分的。实际上, 当数据在中位数附近高度集中时, IQR 可能会使人产生误解。例如, 假设由 10 个数据组成的样本: 20, 50, 50, 50, 50, 50, 50, 50, 50, 80。均值, 中位数, 四分之一分位数和四分之三分位数都等于 50。因此,  $IQR = 50 - 50 = 0$ , 这是很易使人产生误解的, 因为一个变异性度量等于 0, 将意味着  $n$  个数据是相等的, 而例子中并非如此, IQR 完全忽略了数据集合上的极端值。事实上, IQR 仅度量了为覆盖中间的 50% 的数据所需要的距离, 因此, 完全忽略了

最低端的 25% 和最高端的 25% 的数据的变异性。总之, IQR 不能就单独一组测量值的变异性提供太多的有用信息, 但对于比较两组或多组测量值的变异性则相当有用。尤其当数据集合有偏倚时, 就更是如此。我们将把 IQR 作为盒形图的一部分, 在 3.6 节做进一步的讨论。

在大多数情况下, 为了对数据集合进行最低限度的描述, 至少需要五个概括性的量, 即最小值  $y_{(1)}$ , 四分之一分位数  $Q(0.25)$ , 中位数, 四分之三分位数  $Q(0.75)$ , 及最大值  $y_{(n)}$ 。当数据集合有一个单峰、钟形并对称的频率直方图时, 只用样本均值及一个变异性度量即样本方差就可以描述该数据集合。现在介绍样本方差。

现在来寻找变异性的一个灵敏的测量方法, 不仅可以用来比较两组测量值的变异性, 而且, 也可以用来解释单个测量值集合的变异性。为此, 我们使用  $y - \bar{y}$ , 即测量值  $y$  与这组测量值的均值  $\bar{y}$  的离差。

为了说明起见, 假设有五个样本测量值  $y_1 = 68$ ,  $y_2 = 67$ ,  $y_3 = 66$ ,  $y_4 = 63$ , 及  $y_5 = 61$ 。这些值表示五个城市中过去至少行使过一次选举权的注册选民的百分比。这些测量值在点图 3.23 中标出。每个测量值在图中横轴上方用一个点标出, 用样本均值

$$\bar{y} = \frac{\sum y_i}{n} = \frac{325}{5} = 65$$

作为集合的中心, 在图 3.23 上作横线代表样本测量值与它的均值之间的离差。这些测量值的离差是用  $y - \bar{y}$  来计算的。这五个测量值及其离差如图 3.23 所示。

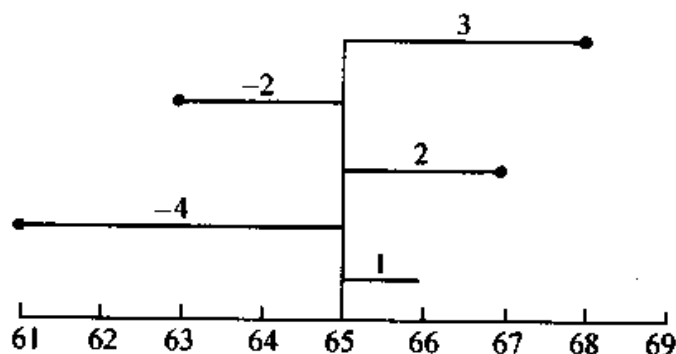


图 3.23 五个城市中注册选民的百分比点图

一个变异性很小的数据集, 其大多数测量值都位于分布的中心附近。而对于一组有更大变异性的测量值, 其相对于均值的离差就会相当大。

许多不同的变异性的度量可以用离差  $y - \bar{y}$  来构造。第一种想法是用平均离差。但如例中所示, 它总是等于 0。第二种可能是省略负号, 计算绝对值的平均。

然而,一个更易解释的离差的函数,是这些测量值相对于均值的离差的平方和,称之为**方差**。

**定义 3.7** 一组均值为  $\bar{y}$  的  $n$  个的测量值的**方差**定义为离差的平方和除以  $n - 1$ , 即

$$\frac{\sum_i (y_i - \bar{y})^2}{n - 1}$$

像样本均值和总体均值一样,对于样本方差和总体方差,我们也分别用特定的符号记之。样本方差记为  $s^2$ , 对应的总体方差记为  $\sigma^2$ 。

一组测量值的方差的定义依赖于视这组数据为测量值的样本还是总体。在上述定义中,我们是针对样本而言的,因为通常得不到总体的测量值。许多统计学家把样本方差定义为离差平方和的均值,即  $\sum_i (y_i - \bar{y})^2 / n$ 。然而,用  $(n - 1)$  作  $s^2$  的分母不是随意的,这个样本方差的定义使其是总体方差  $\sigma^2$  的一个无偏估计。这大致意味着,若从研究的总体中抽取大量样本,每组容量为  $n$ , 计算每个样本的  $s^2$ , 则平均的样本方差将等于总体方差  $\sigma^2$ 。若在样本方差的定义中除以  $n$ , 从大量样本计算的平均样本方差将小于总体方差,因此,相应的  $s^2$  将低估  $\sigma^2$ 。

另一个有用的变异性度量是**标准差**,指方差的平方根。定义标准差的一个原因是,它产生了与原始数据具有相同量纲的变异性度量,然而方差的单位却是测量值单位的平方。

**定义 3.8** 一组测量值的**标准差**定义为其方差的正的平方根。样本标准差记为  $s$ , 相应的总体标准差记为  $\sigma$ 。

### 例 3.9

在某种试验中,记录下五只老鼠在用电火花刺激和为了避免震惊而按压一下压条之间的时间间隔如下,计算样本方差和标准差。

电击避免时间(秒): 5, 4, 3, 1, 3

**解答** 样本均值为  $\bar{y} = 3.2$ , 离差及其平方如表所示。

	$y_i$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
	5	1.8	3.24
	4	0.8	0.64
	3	-0.2	0.04
	1	-2.2	4.84
	3	-0.2	0.04
总和	16	0	8.80

用离差平方的总和,可知样本方差为:

$$s^2 = \frac{\sum_i (y_i - \bar{y})^2}{n - 1} = \frac{8.80}{4} = 2.2$$

若只有分组数据时,可对样本方差公式作简单调整求得近似样本方差。回忆求分组数据的样本均值的近似值的方法,令  $y_i, f_i$  分别表示第  $i$  个分组区间的中点和频数。用这些记号,分组数据的样本方差是  $s^2 = \sum f_i (y_i - \bar{y})^2 / (n - 1)$ , 样本标准差是  $\sqrt{s^2}$ 。

### 例 3.10

关于例 3.6 中表 3.6 的小鸡数据,计算这些数据的样本方差及其标准差。

**解答** 从表 3.6 中可知,  $f_i (y_i - \bar{y})^2$  的总和是 9.4936。用这个值可近似求得  $s^2$  和  $s$ 。

$$s^2 = \frac{1}{n - 1} \sum_i f_i (y_i - \bar{y})^2 = \frac{1}{99} (9.4936) = 0.095895$$

$$s = \sqrt{0.095895} = 0.3097$$

若有最初的 100 个数值计算  $s$ ,  $s$  的值到第四位小数都和上述结果相同。从原始数据和从分组数据计算所得的  $s$  值一般不同。然而,当频数表中的区间很多时,从频数表中得到的  $s$  的近似值是很精确的。

我们已经讨论了变异性的几个度量。每一个都可用来对比两组或多组测量值之间的变异性。其中标准差尤其受欢迎,原因有二:(1) 可以用标准差对比两个以上的数据集合之间的变异性;(2) 可用从下面的准则所得到的结论解释一组测量值的标准差。这个准则适用于大体上有“丘形”直方图的数据——也就是说,其直方图单峰、对称并在尾部逐渐降低。由于很多数据集都可归为“丘形”类中,这个准则应用得很广泛,由此称之为经验准则。

#### 经验准则:

若一个含有  $n$  个测量值的集合拥有“丘形”直方图,那么,

区间  $\bar{y} \pm s$  大约包含这组测量值的 68%,

区间  $\bar{y} \pm 2s$  大约包含这组测量值的 95%,

区间  $\bar{y} \pm 3s$  大约包含这组测量值的 99.7%。

### 例 3.11

来自某个牲畜围场的年度报告给出了每磅牛肉的平均日批发价格为 0.61 美元,标准差为 0.07 美元。对畜场的每日牛肉价格可得到什么结论? 由于不能得到

原始的日价格数据,不能提供更多的每天牛肉价格的信息。然而,由过去的经验可知,日价格有一个“丘形”频率直方图。应用经验准则,可以从日牛肉价格分布图中得到什么结论?

**解答** 应用经验准则可知,区间

$0.61 \pm 0.07$  或者 0.54 美元至 0.68 美元

含有大约 68% 的测量值。区间

$0.61 \pm 0.14$  或者 0.47 美元至 0.75 美元

含有大约 95% 的测量值。区间

$0.61 \pm 0.21$  或者 0.40 美元至 0.82 美元

含有大约 99.7% 的测量值。

在英国,大约 2/3 的牛肉售价在每磅 0.54 美元和 0.68 美元之间,95% 的牛肉售价在每磅 0.47 美元和 0.75 美元之间,最低及最高售价大约为每磅 0.54 美元和 0.68 美元。

为了增加对经验准则的信赖,看一下它对图 3.24 中的五个频数分布的描述情况。计算五个数据集合(这里没有给出)的均值和标准差,这些值已标在每一个频数直方图的旁边。图 3.24 (a) 给出了关于一个变量的测量值的频数分布,该变量可以取值  $y = 0, 1, 2, \dots, 10$ , 并具有对称丘形分布,均值和标准差分别是  $\bar{y} = 5.50$  及  $s = 1.49$ 。由此计算出区间  $\bar{y} \pm 2s$ , 并在图形中横轴的下方标出。我们发现,94% 的测量值落在这个区间中——即落在均值的两个标准差的范围内。注意,这个百分比很接近于经验准则给出的 95%。我们还计算了落在均值的一个标准差的范围内的测量值的百分比,大约为 60%, 与经验准则中给出的 68% 相差不大。因此,可认为经验准则为图形 3.34(a) 中的数据提供了一个充分的描述。

图形 3.24(b) 给出了另一个丘形频数分布,但它比图 3.24(a) 的分布的峰略低,如图的右侧所标注的,这个分布的均值和标准差分别是 5.50 和 2.67。位于均值的--和两个标准差范围内的测量值的百分比分别是 64% 和 96%。这些百分比也和经验准则中提到的很好地相符。

现在看一下另外的三个分布,图 3.24(c) 中的分布完全是平的,然而图 3.24 (b) 和 (d) 中的分布是不对称且向右偏的。这三个分布中,位于均值的两个标准差之内的测量值的百分比分别是 100%, 96% 和 95%。所有这些百分比都合理地接近于经验准则中给出的 95%。在均值的一个标准差范围之内的测量值的百分比分别是 60%, 75%, 87%, 与准则中 68% 有一定程度的差异。

总之,可以看出,在图 3.24 所有五个分布中,经验准则精确地预测了落在均值的两个标准差中的测量值的百分比,尽管图 3.24(c) 中的分布是平的,或图 3.24 (e) 中的分布右偏很严重。经验准则对预测在均值的一个标准差范围之内的测量

值的百分比精确性略差。但是,预测 68% 对于 3.24(a), 3.24(b) 及 3.24(d) 中的  
 丘形分布相比较之下是合理的。

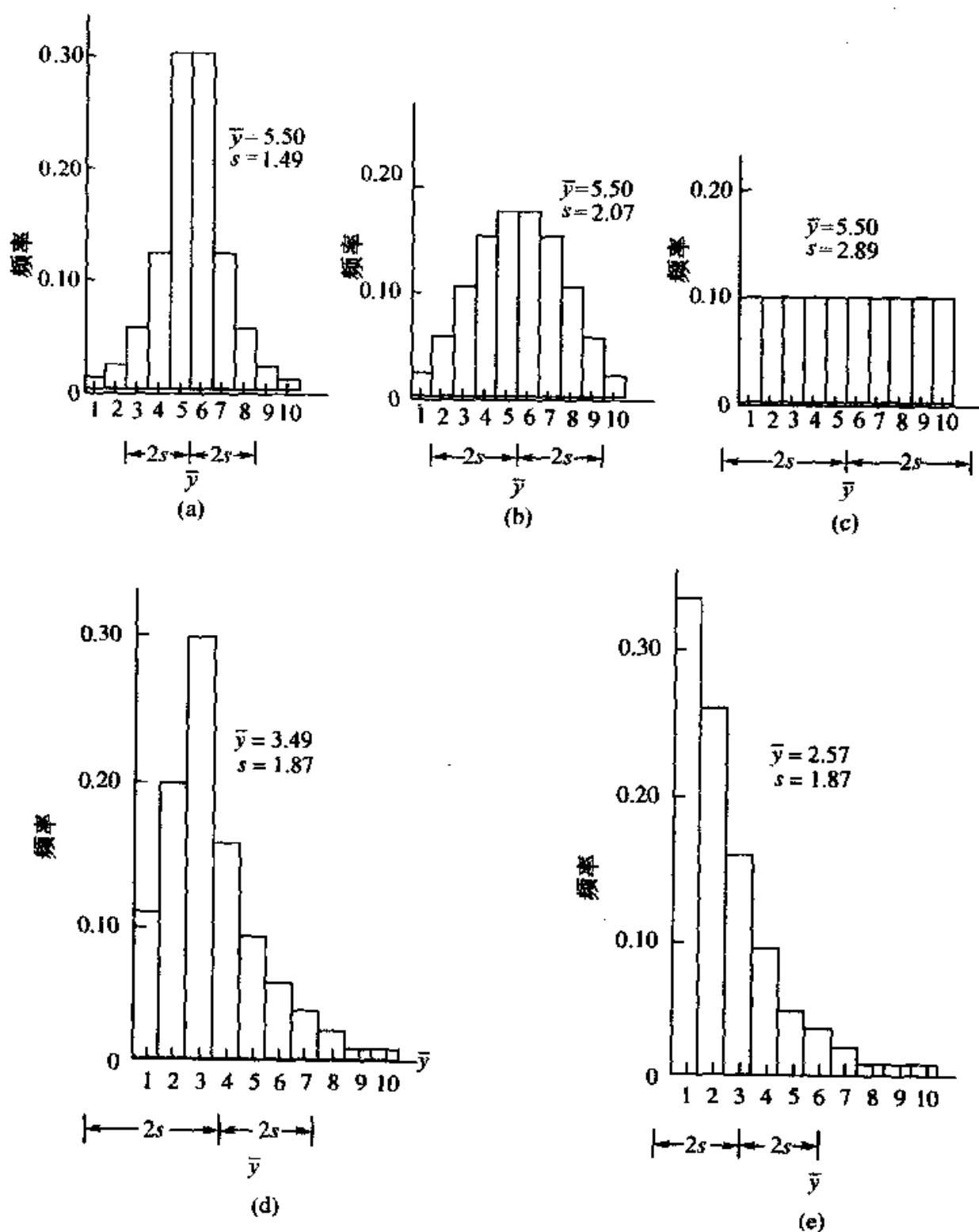


图 3.24 经验准则的应用

经验准则的结论可使我们很快地得到样本标准差  $s$  的近似值。经验准则提到大约 95% 的测量值位于区间  $\bar{y} \pm 2s$  内。因此, 这个区间的长度是  $4s$ 。由于测量值的极差约是  $4s$ , 可以通过用极差除以 4 得到  $s$  的近似值:

$$s \text{ 近似值} = \text{极差}/4$$

有人可能奇怪, 由于区间  $\bar{y} \pm 3s$  包含差不多所有测量值, 为什么不令极差等于  $6s$ ? 这个方法得到的  $s$  的近似值将比前面方法得到的值小。若注定要犯错误(像我们处理任何近似值一样)的话, 我们宁愿过度估计样本标准差, 以便不要被误导为比实际情况有更小的变异性。

### 例 3.12

下面的数据描述了家庭月收入在食品支出上的百分比, 是从一个由 30 名顾客组成的样本得到的。

26	28	30	37	33	30
29	39	49	31	38	36
33	24	34	40	29	41
40	29	35	44	32	45
35	26	42	36	37	35

对于这些数据,  $\sum y_i = 1,043$  且  $\sum (y_i - \bar{y})^2 = 1069.3667$ 。计算花在食物上的收入百分比的均值, 方差和标准差, 并检查  $s$  的计算结果。

**解答** 样本均值是

$$\bar{y} = \frac{\sum y_i}{30} = \frac{1,043}{30} = 34.77$$

对应的样本方差和标准差是

$$s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2 = \frac{1}{29}(1,069.3667) = 36.8747$$

$$s = \sqrt{36.8747} = 6.07$$

用极差近似值可以检查  $s$  的计算结果。最大测量值是 49, 最小测量值是 24。因此,  $s$  的近似值是

$$s \approx \text{极差}/4 = (49 - 24)/4 = 6.25$$

可见近似值很接近计算值。

虽然不能总是如例 3.12 中近似得很好, 极差近似值提供了一个有用且快捷的检查  $s$  的计算结果的方法。

当比较不同类型的总体的变异性的太小时, 标准差可能使人发生误解。在某些总体中, 一个单位大小的方差可能被认为是很小的, 而对于另一些总体, 同等大



小的变异性则可能被认为很大。例如,假若要比向容器中装产品的两个生产流程。流程 A 是装化肥袋,要求每袋重 80 磅,而该流程封装的袋子,平均重量为 80.6 磅,标准差为 1.2 磅。流程 B 是装 24 盎司的玉米片盒子,要求装入的重量是 24 盎司。而流程 B 装盒的实际平均重量为 24.3 盎司,标准差是 0.4 盎司。1.2 是 0.4 的 3 倍,我们是否可由此认为流程 A 比流程 B 有更强的变异性呢?为了比较两个很不相同的流程或总体的变异性,需定义另一个变异性度量——**变异系数**,它在数量上度量了一个总体的变异性相对于其总体均值的大小。在一个均值为  $\mu$ , 标准差为  $\sigma$  的流程或总体中,**变异系数**定义为

$$CV = \frac{\sigma}{|\mu|}$$

此处假设  $\mu \neq 0$ 。因此,变异系数是单位均值上总体或流程的标准差。若两个封装流程有相同的 CV,那么,它们就有相同程度的变异性。对于装化肥的流程,  $CV = 1.2/80 = 0.015$ 。而对于装玉米片的流程,  $CV = 0.4/24 = 0.017$ 。两个流程相对于它们均值的大小,有很相近的变异性。在许多实际应用中, CV 用百分比的形式表达:  $CV = 100(\sigma/|\mu|)\%$ 。因此,若一个流程的 CV 是 15%,那么,这个流程的输出标准差就是其输出的均值的 15%。通过来自总体的样本数据,我们可以用  $100(s/|\bar{y}|)\%$  来估计 CV。

## 练习

### 基本技能

**3.39 (工程)** 在 Goosty 和 Campbel 的一篇论文中[适应体育活动季刊 (*Adapted Physical Activity Quarterly*)(1998)15:36~50],调查了机动脚踏车的推进系统的性能。为此,在一个机动脚踏车比赛中,考察了 8 名轮椅参赛者。8 名参赛者的参赛年数为

参赛经历(年): 6, 3, 10, 4, 4, 2, 4, 7

- 试验证平均参赛经历是 5 年,这一数值是否充分描述了数据集合的中心?
- 证明  $\sum_i (y_i - \bar{y})^2 = \sum_i (y_i - 5)^2 = 46$ 。
- 计算这些经历数据的样本方差和标准差。你如何相对于样本均值来解释标准差的数值?

**3.40** 在练习 3.39 描述的研究中,调查者也记录了 8 名参赛者的年龄。

年龄(年): 39, 38, 31, 26, 18, 36, 20, 31

- 计算 8 名参赛者年龄的标准差。
- 为什么你能够期望参赛者年龄标准差比他们的经历标准差大?

**3.41** 对于练习 3.39 的比赛经历数据,用极差除以 4 估计其标准差。这个估计值与你在练习 3.39 计算所得的结果接近吗?对练习 3.40 中的数据也进行同

样的计算。

**3.42(医学)** 一家健康诊所中,对患者的治疗时间(分钟)如下:

21	20	31	24	15	21	24	18	33	8
26	17	27	29	24	14	29	41	15	11
13	28	22	16	12	15	11	16	18	17
29	16	24	21	19	7	16	12	45	24
21	12	10	13	20	35	32	22	12	10

计算治疗时间的标准差。(提示:可验证对于上述 50 个治疗时间  $\sum y_i = 1,016$  及  $\sum (y_i - \bar{y})^2 = 3,434.88$ )

**3.43** 用练习 3.42 中的数据,作该诊所中治疗时间的分位数图。

- 找出治疗时间的 25% 百分位数,并解释这一数值。
- 该健康诊所的广告宣称:90% 的患者治疗时间不超过 40 分钟,这些数据是否支持其宣传?

**3.44(环境)** 为了帮助估计一片林地中的木材数量,林地主人决定在随机选择的  $50 \times 50$  英尺正方形的面积中,数出直径超过 12 英寸的树木的数量。从林地中随机选择了 70 个  $50 \times 50$  英尺正方形的面积,并数出直径超过 12 英寸的树的数量,数据如下:

7	8	6	4	9	11	9	9	9	10
9	8	11	5	8	5	8	8	7	8
3	5	8	7	10	7	8	9	8	11
10	8	9	8	9	9	7	8	13	8
9	6	7	9	9	7	9	5	6	5
6	9	8	8	4	4	7	7	8	9
10	2	7	10	8	10	6	7	7	8

- 作出频率直方图描述这些数据。
- 计算样本均值  $\bar{y}$  以估计  $\mu$ , 这里  $\mu$  是林地中所有  $50 \times 50$  方块面积中直径超过 12 英寸的树木棵数的平均值。
- 计算数据  $s$ , 作出区间  $(\bar{y} \pm s)$ ,  $(\bar{y} \pm 2s)$  和  $(\bar{y} \pm 3s)$ 。分别算出其树木数量落在三个区间中的方块的百分比,并与经验准则给出的对应区间的百分比作比较。

**3.45(政府)** 一个城市计划为运输部门买汽车,为了正确预算汽车的维修费,必须确定汽车的可靠性。用练习 3.30 的数据,这个城市可以估计汽车出重大发动机故障前的平均行驶里程。

- 估计 191 辆车行驶里程的极差。

- b. 估计 191 辆车行驶里程的标准差。
- c. 用均值和标准差的估计值, 作出区间  $(\bar{y} \pm s)$ ,  $(\bar{y} \pm 2s)$  和  $(\bar{y} \pm 3s)$ , 分别计算其行驶里程落在三个区间中的汽车的百分比, 并与经验准则中提到的相应的百分比作比较。为什么在此例中, 经验准则是适用的?

**3.46(商业)** 在 *Consumer Reports* 杂志 1998 年 6 月号中, 报告了 6 个豪华宾馆及 9 个廉价宾馆的日住宿费。住宿费如下表。

豪华宾馆(美元)	175	180	120	150	120	125			
廉价宾馆(美元)	50	50	49	45	36	45	50	50	40

- a. 分别计算豪华宾馆和廉价宾馆的均值和标准差。
- b. 验证豪华宾馆比廉价宾馆的住宿费有更大的变异性。
- c. 就为什么豪华宾馆比廉价宾馆的费用有更大的变异性, 给出一种符合实际的解释。
- d. 有没有一个更好的变异性度量, 能用来对豪华宾馆和廉价宾馆的住宿费作比较? 并给出你的理由。

**3.47(环境)** 许多海洋显花生物对于环境条件的变化极为敏感, 在文章 “*Posidonia oceanica*: A biological indicator of past and present mercury contamination in the Mediterranean Sea” [海洋环境研究 (*Marine Environmental Research*), 45: 101~111] 中, 调查者报告了过去 20 年的时期内, 地中海中几个地点的水银含量。*Posidonia oceanica* 的样本是由佩带水下呼吸器的潜水员潜入水中 10 米处收集的。在每一个地点, 采集了 45 个直生性嫩芽, 并测定其水银含量。下表中记录了每一个样本年的平均水银含量。

水银含量 (ng/g 干重)		
年份	地点 1	地点 2
1992	14.8	70.2
1991	12.9	160.5
1990	18.0	102.8
1989	8.7	100.3
1988	18.3	103.1
1987	10.3	129.0

续表

年份	水银含量(ng/g 干重)	
	地点 1	地点 2
1986	19.3	156.2
1985	12.7	117.6
1984	15.2	170.6
1983	24.6	139.6
1982	21.5	147.8
1981	18.2	197.7
1980	25.8	262.1
1979	11.0	123.3
1978	16.5	363.9
1977	28.1	329.4
1976	50.5	542.6
1975	60.1	369.9
1974	96.7	705.1
1973	100.4	462.0
1972	*	556.1
1971	*	461.4
1970	*	628.8
1969	*	489.2

- 作水银含量的时间序列图,把两个地点上的时间序列连线画在同一个图中;说明跨越这些年的连线所具有的趋势;对于两个地点其趋势相同吗?
- 选择最合适的水银含量的中心度量,并比较两个地点水银含量的中心度量的值。
- 用 CV 比较两个地点中水银含量的变异性。并说明为何比用标准差更恰当?
- 当比较两个地点的中心和变异性时,地点 2 中的年份 1969~1972 的数据能用吗?

### 3.6 盒 形 图

如本章前面所提到的,茎叶图是一组记分的图形表示,可用来检查分布的形状、记分的极差及记分聚集的位置。**盒形图**(又译为箱线图)是建立在茎叶图所显示的信息的基础上的,它更多地关注分布的对称性,并结合中心趋势及位置的数字度量研究记分的变异性以及在分布的尾部记分的分布情况。

在讲解如何做盒形图及解释盒形图之前,需解释几个探索性数据分析(EDA)的特定术语。我们已经熟悉了本章前面提到的第一,第二(中位数)和第三个四分位数。盒形图用到了一个分布的中位数和四分位数。

现用例子介绍骨架盒形图。

### 例 3.13

图 3.25 是表 3.4 中 90 个大城市的暴力犯罪率的茎叶图,由此作一骨架盒形图。

```

1  89
2  10 24 67 91 96 98
3  36 41 52 54 75 76 88 93 93
4  10 21 35 47 48 60 64 66 80 81 91 96 98
5  04 05 08 16 26 29 37 57 59 61 62 62 63 70 71 78 85 92
6  05 05 24 26 28 31 39 42 47 61 73 84 85 85 90 98
7  03 06 18 19 20 31 35 39 51 58 71
8  04 07 09 11 14 17 43 56 68 76 77 85
9  28 71
10 20
  
```

图 3.25 茎叶图

**解答** 当记分按从小到大的顺序排列时,中位数为第 45 个和第 46 个记分的平均值,对于这些数据,第 45 个记分(在图 3.25 中从小到大计数)是 571,第 46 个记分是 578,因此,中位数是

$$M = \frac{571 + 578}{2} = 574.5$$

为寻找这个分布的四分之一和四分之三分位数,需要确定 25% 和 75% 分位数。我们可以用 3.5 节介绍的方法计算  $Q(0.25)$  和  $Q(0.75)$ 。然而,通过下列步骤,可以很快得到这两个百分位数的近似值,并且所得值与上述方法近乎相同。

1. 从最小值到最大值将这些数据排序。
2. 用中位数作为分点,将这些有序数据分为两组。
3. 把由较小数据组成的集合的中位数,作为四分之一分位数。
4. 把由较大数据组成的集合的中位数,作为四分之三分位数。

在本例中,数据集合有 90 个值。因此,可以分为两组,一组含有  $90/2 = 45$  个最小的值,一组含有 45 个最大的值。四分之一分位数是第  $(45 + 1)/2 = 23$  个最小值,四分之三分位数从最大值数起的第 23 个数值。第 23 个最小记分和第 23 个最

大计分分别是 464 和 719。

四分之二分位数,  $Q_1 = 464$

四分之三分位数,  $Q_3 = 719$

数据集中的这三个描述性度量和最大最小值一起被用来做骨架盒形图(见图 3.26)。**骨架盒形图**的构造方法是, 在四分之二分和四分之三分位数之间画一个盒子, 并在中位数的位置上画一条实线穿过盒子; 然后, 用一条直线连接盒子与最大值, 第二条直线连接盒子与最小值。这些直线有时称为须, 整个图形则称为**盒形图**。

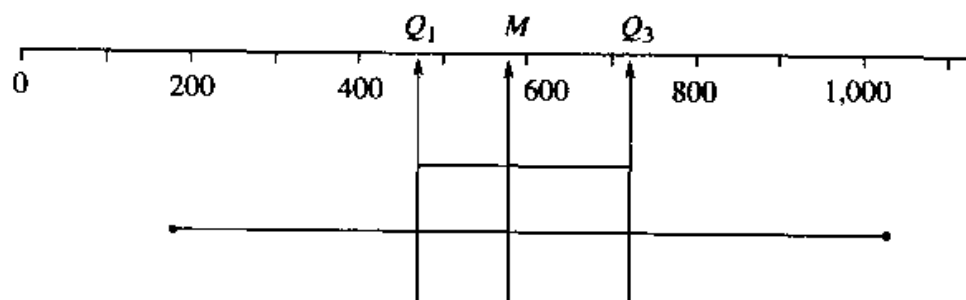


图 3.26 图 3.25 数据的骨架盒形图

浏览一下骨架盒形图, 很容易对数据的如下方面得到大致印象:

1. 四分之二分和四分之三分位数, 即  $Q_1$  和  $Q_3$ ;
2. 四分位间距(IQR), 即四分之二分和四分之三分位数间的距离;
3. 极值(最大和最小值);
4. 记分分布的对称与否。

如果未见到原始数据, 而仅凭图 3.26 的描述, 我们将观察到

$$Q_1 \approx 475$$

$$Q_3 \approx 725$$

$$IQR \approx 725 - 475 = 250$$

$$M \approx 575$$

极值: 175 和 1025。

而且, 由于中位数更接近于四分之二分位数, 且上半部分的须比下半部分须略长, 分布稍微有些不对称。要看这些结论是否是对的, 可作出这组数据的频数直方图(或参考练习 3.8 的结果)。

骨架盒形图可以进行拓展, 使之包含分布尾部的极值的更多信息。为此, 还需要下面的量:

$$\text{下侧内分线: } Q_1 - 1.5(IQR)$$

$$\text{上侧内分线: } Q_3 + 1.5(IQR)$$

下侧外分线:  $Q_1 - 3(IQR)$

上侧外分线:  $Q_3 + 3(IQR)$

任何一个超出下侧或上侧内分线的记分称之为**中度离群值** (mild outlier)。

任何一个超出下侧或上侧外分线的记分称之为**极端离群值** (extreme outlier)。

### 例 3.14

计算例 3.13 数据的内分线 and 外分线, 找出中度离群值和极端离群值。

**解答** 对这些数据, 可发现四分之一和四分之三分位数分别是 464 和 719。  
 $IQR = 719 - 464 = 255$ 。于是,

下侧内分线  $= 464 - 1.5 \times 255 = 81.5$

上侧内分线  $= 719 + 1.5 \times 255 = 1,101.5$

下侧外分线  $= 464 - 3 \times 255 = -301$

上侧外分线  $= 719 + 3 \times 255 = 1,484$

从茎叶图我们还可看出, 最低和最高边界值是 189 和 1,020。由于上侧内分线和下侧内分线分别是 1,101.5 和 81.5, 没有观测值在内分线之外。因此, 无中度离群值和极端离群值。

现在已经有了作盒形图必需的所有数据。

#### 作盒形图的步骤

1. 与作骨架盒形图一样, 从四分之一到四分之三分位数划出一个盒子。
2. 过中位数画一条实线穿过盒子。
3. 用 X 标出上下边界值。
4. 在每一个四分位数和它所对的边界值间画一条虚线。
5. 用记号 0 标出每个极端离群值。

### 例 3.15

由例 3.13 的数据作一个盒形图。

**解答** 盒形图如图 3.27 所示。

可从图中得到什么结论呢? 首先, 在盒形图中, 记分分布的中心被盒形图里的中位数线所指明。第二, 记分的变异性度量由四分位间距给出, 即盒子长度。回忆一下, 盒子是由四分之一和四分之三分位数构造出来的, 因而, 它包含了分布中间的 50% 的记分, 在盒子里中位数线的每一边有 25%。第三, 通过检查中位数线的相对位置, 可以测定中间 50% 的记分的对称性。例如, 若中位数线比起四分之三分位数来, 更接近于四分之一分位数, 则记分更多地集中在盒内比中位数低的一

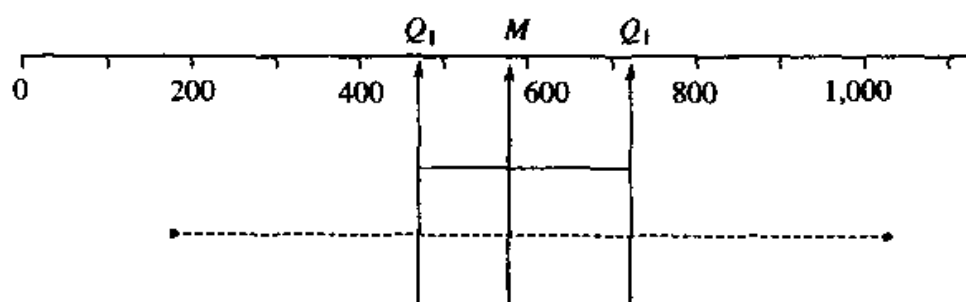


图 3.27 例 3.13 中数据的盒形图

边。记分的对称分布将使中位数线位于盒子的中间。第四,关于偏斜的其他信息可以从须的长度得到。一个须相对于另一个越长,在有长须的尾部就有越多的偏斜。第五,通过检查划为中度离群值和极端离群值的记分的个数,可以对离群值情况进行一般的讨论。

盒形图为对比来自几个不同的处理或总体的样本提供了有效的图形方法。我们将用下面的例子来解释这些概念。几种新的过滤系统被建议用于小城市的供水体系。正在考虑的三个系统的初装费和操作费用大体一致,现在比较通过这些系统后留在水中的杂质含量。在仔细地评估后,决定监控 20 天的操作,这将为确定三个系统间任何有意义的区别提供充分的信息。水的样本按 1 小时为间隔来收集。记录通过三个系统后留在水中的杂质含量(以 ppm 即百万分之一为单位)。把三个系统每天的平均值数据,作成并立(并列)形式的盒形图,如图 3.28 所示。

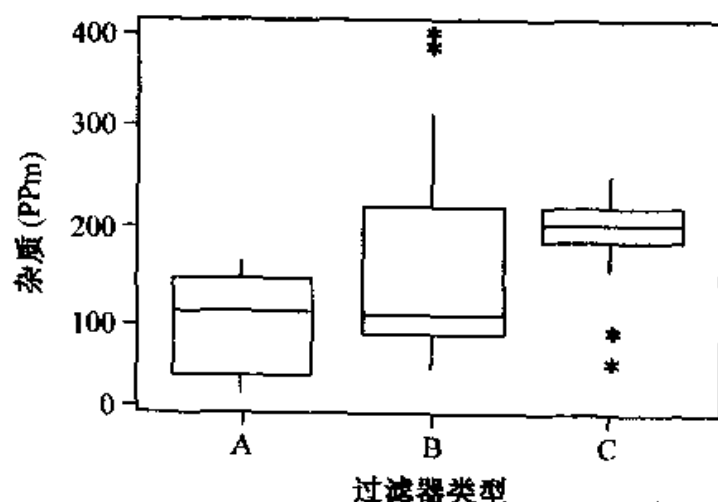


图 3.28 用三种过滤器除掉杂质

检查图 3.28 中的盒形图,可以得到由盒形图所表示的三种过滤器的频率直方图的形状。过滤器 A 所对应的分布是单峰、对称的;过滤器 B 的分布是一个单峰



右偏的分布;过滤器 C 的分布是一个单峰左偏的分布。过滤器 A、B 有几乎相同的中位数。然而,过滤器 B 比起过滤器 A、C 来,有更大的变异性。过滤器 C 有比过滤器 A、B 都大的中位数,但除了用过滤器 C 得到两个很小的值外,它比过滤器 A 有更小的变异性。从过滤器 B、C 得到的极端值,图中用 \* 表示出来。对于这些极端值,要予以考证,以确认它们都是有效的测量值。这些测量值可能是由记录错误或操作错误造成的。由于它们对于概括统计量有很大的影响,故应仔细检查。过滤器 A 比过滤器 B 更稳定,比过滤器 C 过滤得更彻底。我们将在第八章介绍一些统计技术,这些技术将有助于我们区分三种过滤器。

### 练习

#### 基本技能

**3.48** 找出下列测量值的中位数及四分之一和四分之三分位数:5, 9, 3, 6, 5, 7, 9, 2, 8, 10, 4, 3, 9。

**3.49** 用下列测量值重复作 3.48 中的练习:29, 22, 26, 20, 19, 16, 24, 11, 22, 23, 20, 29, 17, 18, 15。

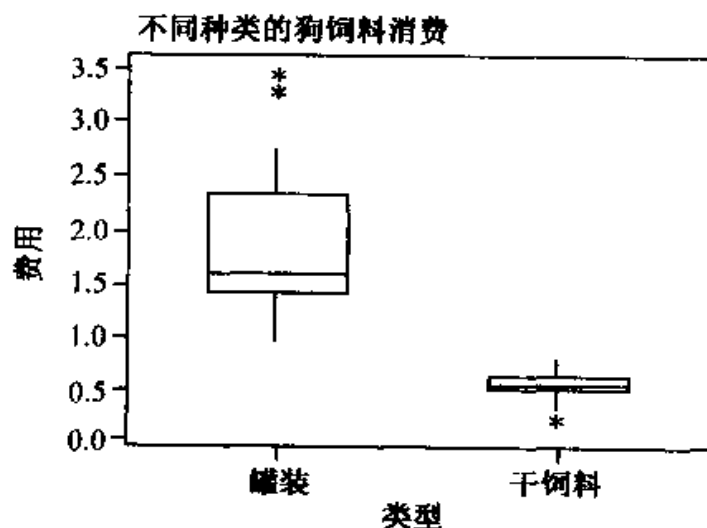
### 应用

**3.50(医药)** 在献血中心,记录了连续 20 个周五自愿献一品脱血的人的数目。数据如下:

320	370	386	334	325	315	334	301	270	310
274	308	315	368	332	260	295	356	333	250

- 作出茎叶图。
- 作出盒形图并描述出献血人数分布的形状。

**3.51(商业)** 在杂志消费者报告(*Consumer Reports*)1998 年 5 月号中,给出



了每天的狗饲料消费数据。这些饲料包括 28 种干饲料和 23 种罐装饲料。用 Minitab 计算机程序,作出了这些数据的并立盒形图。

- a. 从图中,确定干饲料和罐装饲料消费的中位数及四分之一和四分之三分位数。
- b. 指出两种狗饲料日消费的分布的相似和不同之处。

### 3.7 多变量数据的概括

在前几节,我们讨论了用来概括来自单个变量的数据的数值描述方法和图形方法。研究多个变量的情况也很常见,此时我们可能要对各个变量的数据分别进行概括,也可能要研究变量之间的相互关系。例如,我们可能对基本利润率和消费价格指数感兴趣,同时,也关注两者之间的联系。在本节中,我们将讨论概括来自两个(或多个)变量的数据的一些技巧。本节材料将为以后要介绍的  $\chi^2$  方法(第十章),方差分析(第八章及第十五~第十九章),和回归(第十一、十二和十三章)提供一个简单的预览和导引。

首先,我们考虑来自于两个定性变量的数据的概括问题。对于这样的数据,我们可以构造交叉列表,形成**列联表**。表中的行给出其中一个变量取值的种类,而表中的列给出另一个变量取值的种类,表中的数值是两个变量各取其一可能的值的次数。例如,调查 1,500 个人的电视收看情况。要求每一个被调查者说出他(或她)的居住地和喜爱的新闻网。调查结果列在表 3.7 中,如你所见,144 个城市居民喜欢 ABC,135 个城市居民喜欢 CBS,等等。

表 3.7 电视收视率调查数据

所喜欢的电视网	居住地			
	市区	郊区	农村	总和
ABC	144	180	90	414
CBS	135	240	96	471
NBC	108	225	54	387
其他	63	105	60	228
总和	450	750	300	1500

在列联表中寻求变量之间关系的最简单的方法是比较基于行的总和,列的总和或所有数的总和的百分比。如果计算出表 3.7 中每一行内的百分比,我们便可以比较喜欢每一个电视网的居民的分布。表 3.8 给出了一个这样的基于行总和上的百分比的比较。

表 3.8 各个电视网的居民分布的比较

所喜欢的电视网	居住地			总和
	市区	郊区	农村	
ABC	34.8	43.5	21.7	100( $n = 414$ )
CBS	28.7	50.9	20.4	100( $n = 471$ )
NBC	27.9	58.1	14.0	100( $n = 387$ )
其他	27.6	46.1	26.3	100( $n = 228$ )

喜欢收看 ABC 的市区居民的百分比在各个电视网中最高。除此以外,各个电视网居民分布的差异主要在郊区和农村。收看 ABC 的郊区居民的百分比为 43.5%,而对于 NBC,这一比例上升到 58.1%,农村居民的百分比则有相应的下降。在第十章,我们将用  $\chi^2$  方法进一步探求两个(或多个)定性变量之间的关系。

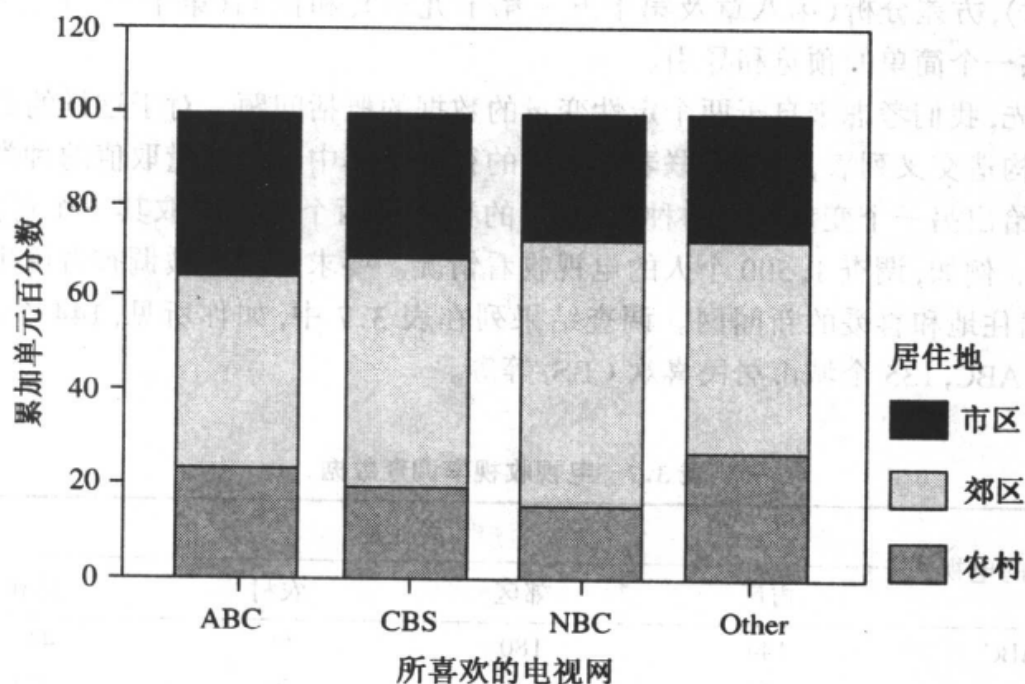


图 3.29 各个电视网居民分布的对比

对于来自于两个定性变量的数据,一种广义的条形图提供了方便的显示数据的方法,这就是**堆叠条形图**。图 3.29 中的堆叠条形图是由表 3.8 中的数据作出的。

在上图中,基于电视观众所居住的位置,显示了每一个主要电视网新闻节目观众的分布。广告商常用这种信息去决定在哪个电视网上作商业广告。

一个定性变量和定量变量之间的关系可以用条形图的另一种扩充形式很方便地显示出来。一个食物学家正在研究不同类型的脂肪和不同类型的表面活性剂的组合对于烘烤的面包的体积所产生的影响。在这个试验的设计中,使用了3种表面活性剂和3种脂肪,因而形成了一个 $3 \times 3$ 析因试验。在试验中,她用9种处理中的每一种,烘烤出不同数量的面包(处理指一种表面活性剂和一种脂肪的组合,关于析因试验,请参见第十五章——译者著)。试验时,她在面团中按3种表面活性剂和3种脂肪的9种组合把表面活性剂和脂肪掺入面团来烘烤面包。数据和概括统计量在表3.9中给出。

在这个试验中,科学家想从试验结果推断商业生产过程。图3.30是该试验的**集类条形图**。利用这种集类条形图,试验者可以考查这两个因子,即脂肪类型和表面活性剂类型对所得面包体积的同时的影响。因此,调查者可以检查出烤面包的9种不同方式对面包体积的影响上的差异。

表 3.9 因变量面包体积的描述性统计量

脂肪	表面活性剂	均值	标准差	N
1	1	5.567	1.206	3
	2	6.200	0.794	3
	3	5.900	0.458	3
	总和	5.889	0.805	9
2	1	6.800	0.794	3
	2	6.200	0.849	2
	3	6.000	0.606	4
	总和	6.311	0.725	9
3	1	6.500	0.849	2
	2	7.200	0.668	4
	3	8.300	1.131	2
	总和	7.300	0.975	8
总和	1	6.263	1.023	8
	2	6.644	0.832	9
	3	6.478	1.191	9
	总和	6.469	0.997	26

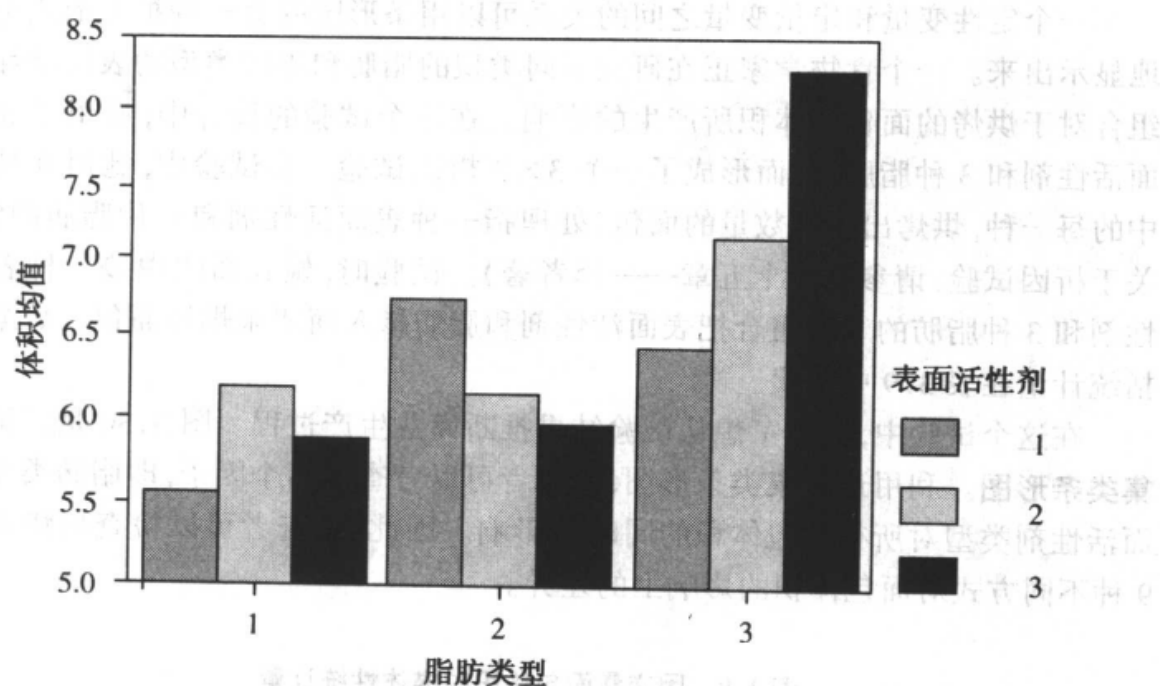


图 3.30 烘烤面包试验中的体积

我们也可以作图去总结两个定量变量之间的关系。考虑下面的例子。一家小机械商店的经理考查他付给有  $x$  年工作经历的机械师的每小时的起点工资  $y$ , 数据如下。

$y$ (美元)	8.09	8.70	9.10	9.00	9.79	9.45	10.00	10.65	11.10	11.05
$x$ (年)	1.25	1.50	2.00	2.00	2.75	4.00	5.00	6.00	8.00	12.00

所付的小时工资和工作经历之间有关系吗？概括这些数据的一种方法是用散点图, 如图 3.31 所示。图上每一点都代表一个有着特定起始工资和工作经历的机械师。图中拟合数据用的光滑曲线, 称之为最小二乘线, 概括描述了  $x$  和  $y$  之间的关系。利用这条线, 可以预测其工作经历年数未包含在数据集合中的机械师的小时起点工资。我们将在第十一和十二章讨论如何得到这条曲线。总体说来, 这条拟合的曲线表明, 随着工作经历  $x$  的增加, 每小时起点工资会增加至某一点, 然后趋于稳定。我们将在有关回归的章节(第十一至第十三章)中, 讨论关于多个定量变量之间关系分析的基本思想。

最后, 我们可以用数据图来概括几个定量变量之间的关系。考虑下例。Thall 和 Vail(1990)对抗癫痫药物氟柳双胺作为标准化疗的辅助药物的效果作了一个评估。在临床试验中, 有 59 个癫痫患者参与试验。这些受轻微或严重的间歇发作折

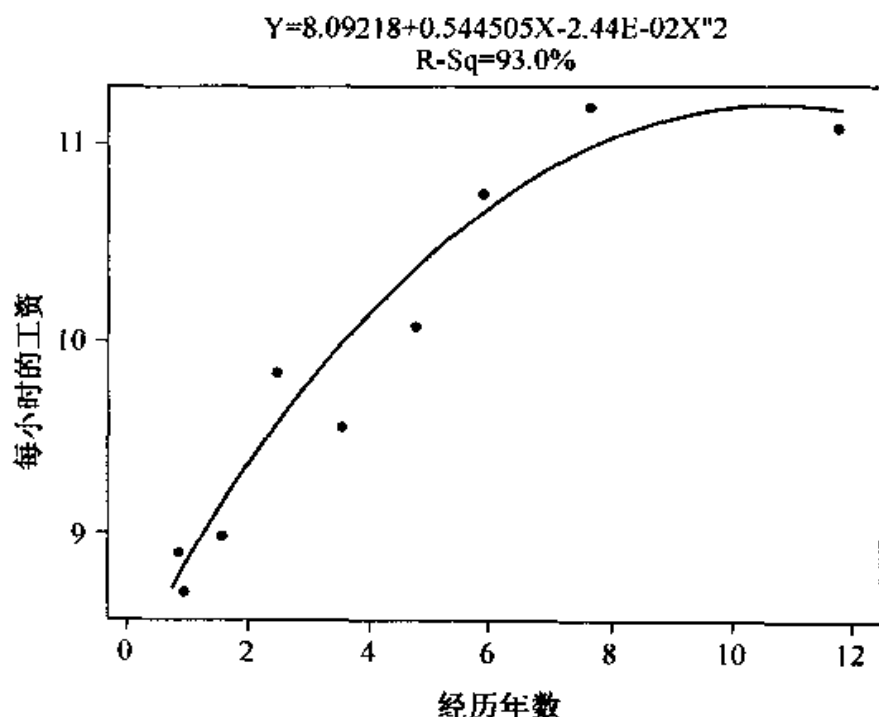


图 3.31 起点小时工资和工作经历的散点图

磨的患者被随机指定去接受抗癫痫药物或安慰剂。连续做了四次后随机化(pos-trandomization)临床诊断,每次都记录了前两周的发作次数。所观测的变量是  $y_i$  ( $i=1, 2, 3, 4$ )——四次临床诊断中记录的发作次数,  $\text{Trt}(x_1)$ ——0 指安慰剂, 1 指辅助药物;  $\text{Base}(x_2)$ , 即基线发作次数,  $\text{Age}(x_3)$ , 即患者的年龄。在表 3.10 和表 3.11 中给出了数据和概括统计量。

第一种图形是**并立盒形图**,我们用这种图来对比治疗组患者和安慰剂组患者的基线发作次数和年龄。利用并立盒形图的视觉效果,我们可以判断,在临床试验之前,治疗组患者和安慰剂组的患者的年龄和基线发作频率是否有相同的分布。观察 3.32(a)可知,在临床试验之前,两组患者的发作次数有相似的模式。两组中只有一个患者的发作基数大于 100。安慰剂组比治疗组的基线发作次数有更大的变异性——它的盒子比治疗组的盒子更宽。描述统计表与这一观察结果相矛盾:安慰剂组的样本标准差为 26.10,治疗组的样本标准差是 27.37。这种看上去不一致的结论,是由于在治疗组中有一个患者有很大的基线发作次数。治疗组的基线发作次数的中位数比安慰剂组大,而两组的均值几乎相等。由于安慰剂组中中间 50%的数据右偏,而治疗组中中间 50%的数据几乎是对称的,故两组的均值比它们的中位数有更好的一致性。图 3.32(b)显示这两组中的年龄几乎是同分布的。仅有的区别是:治疗组的中位数年龄稍微小一点,变异性稍微大一点。因此,这两组在临床试验之前有相似的年龄和基线发作次数的分布。

表 3.10 癫痫研究的数据:59 名癫痫患者连续两周的发作次数,协变量是辅助治疗  
(0 = 安慰剂,1 = 氟柳双胺),8 周的基线发作次数和年龄(岁)

ID	$y_1$	$y_2$	$y_3$	$y_4$	Trt	Base	Age
104	5	3	3	3	0	11	31
106	3	5	3	3	0	11	30
107	2	4	0	5	0	6	25
114	4	4	1	4	0	8	36
116	7	18	9	21	0	66	22
118	5	2	8	7	0	27	29
123	6	4	0	2	0	12	31
126	40	20	23	12	0	52	42
130	5	6	6	5	0	23	37
135	14	13	6	0	0	10	28
141	26	12	6	22	0	52	36
145	12	6	8	4	0	33	24
201	4	4	6	2	0	18	23
202	7	9	12	14	0	42	36
205	16	24	10	9	0	87	26
206	11	0	0	5	0	50	26
210	0	0	3	3	0	18	28
213	37	29	28	29	0	111	31
215	3	5	2	5	0	18	32
217	3	0	6	7	0	20	21
219	3	4	3	4	0	12	29
220	3	4	3	4	0	9	21
222	2	3	3	5	0	17	32
226	8	12	2	8	0	28	25
227	18	24	76	25	0	55	30
230	2	1	2	1	0	9	40
234	3	1	4	2	0	10	19
238	13	15	13	12	0	47	22
101	11	14	9	8	1	76	18
102	8	7	9	4	1	38	32
103	0	4	3	0	1	19	20
108	3	6	1	3	1	10	30
110	2	6	7	4	1	19	18
111	4	3	1	3	1	24	24
112	22	17	19	16	1	31	30
113	5	4	7	4	1	14	35

续表

ID	y <sub>1</sub>	y <sub>2</sub>	y <sub>3</sub>	y <sub>4</sub>	Trt	Base	Age
117	2	4	0	4	1	11	27
121	3	7	7	7	1	67	20
122	4	18	2	5	1	41	22
124	2	1	1	0	1	7	28
128	0	2	4	0	1	22	23
129	5	4	0	3	1	13	40
137	11	14	25	15	1	46	33
139	10	5	3	8	1	36	21
143	19	7	6	7	1	38	35
147	1	1	2	3	1	7	25
203	6	10	8	8	1	36	26
204	2	1	0	0	1	11	25
207	102	65	72	63	1	151	22
208	4	3	2	4	1	22	32
209	8	6	5	7	1	41	25
211	1	3	1	5	1	32	35
214	18	11	28	13	1	56	21
218	6	3	4	0	1	24	41
221	3	5	4	3	1	16	32
225	1	23	19	8	1	22	26
228	2	3	0	1	1	25	21
232	0	0	0	0	1	13	36
236	1	4	3	2	1	12	37

表 3.11 描述统计量:癫痫病例子的 Minitab 输出结果  
(工作单大小: 100,000 个单元)

0 = PLACEBO

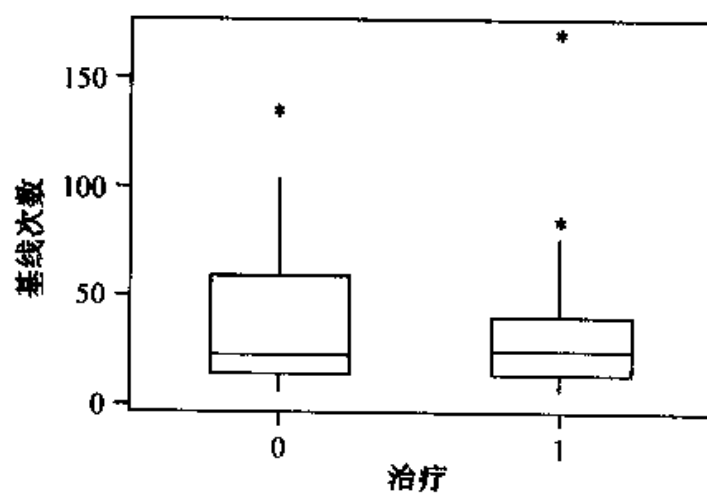
1 = TREATED

Variable	TREATMENT	N	Mean	Median	Tr Mean	StDev	SE Mean
Y1	0	28	9.36	5.00	8.54	10.14	1.92
	1	31	8.58	4.00	5.26	18.24	3.28
Y2	0	28	8.29	4.50	7.81	8.16	1.54
	1	31	8.42	5.00	6.37	11.86	2.13
Y3	0	28	8.79	5.00	6.54	14.67	2.77
	1	31	8.13	4.00	5.63	13.89	2.50

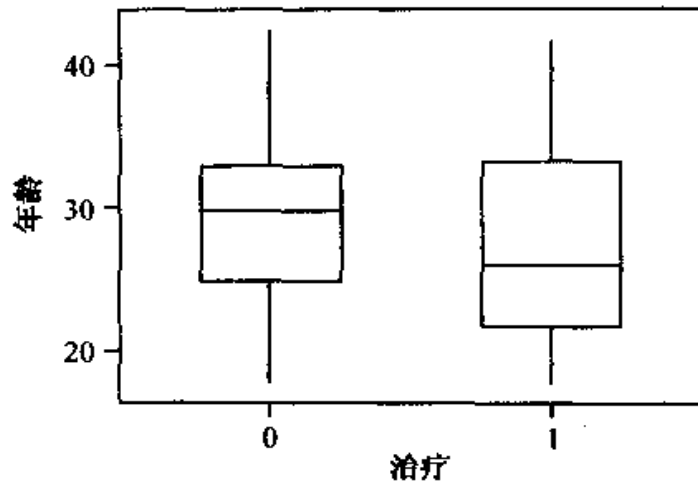


续表

Variable	TREATMENT	N	Mean	Median	Tr Mean	StDev	SE Mean
Y4	0	28	7.96	5.00	7.46	7.63	1.44
	1	31	6.71	4.00	4.78	11.26	2.02
BASE	0	28	30.79	19.00	28.65	26.10	4.93
	1	31	31.61	24.00	27.37	37.98	5.03
AGE	0	28	29.00	29.00	28.88	6.00	1.13
	1	31	27.74	26.00	27.52	6.60	1.19
Variable	TREATMENT	Min	Max	Q1	Q3		
Y1	0	0.00	40.00	3.00	12.75		
	1	0.00	102.00	2.00	8.00		
Y2	0	0.00	29.00	3.00	12.75		
	1	0.00	65.00	3.00	10.00		
Y3	0	0.00	76.00	2.25	8.75		
	1	0.00	72.00	1.00	8.00		
Y4	0	0.00	29.00	3.00	11.25		
	1	0.00	63.00	2.00	8.00		
BASE	0	6.00	111.00	11.00	49.25		
	1	7.00	151.00	13.00	38.00		
AGE	0	19.00	42.00	24.25	32.00		
	1	18.00	41.00	22.00	33.00		



(a) 两个治疗组的基线次数的盒形图



(b) 两个治疗组的年龄盒形图

图 3.32

## 练习

**3.52(社会)** 论文“Demographic implications of socioeconomic transition among the tribal populations of Manipur, India(印度 Manipur 部落群体社会经济变迁的人口统计学分析)”[人类生物学(*Human Biology*)(1998) 70(3):597~619]中,作者描述了从 20 世纪初开始,印度所有 Manipur 部落的群体发生的巨大变化。Manipur 部落正处于在从传统的生计型经济向市场经济的社会经济转变过程中。下表中的数据来自于印度 Manipur 部落中 614 个已婚男性和女性组成的样本,数据显示了他们的文化水平和生计群体的关系。

生计群体	文化水平		
	无文化	基础教育	中学以上
变迁的种植者	114	10	45
安居的农业主	76	2	53
城镇居民	93	13	208

a. 用集类条形图描述表中的数据。

b. 基于行和列的总和,作百分数的对比。对于文化水平和生计群体之间的关系你能得到什么结论。

**3.53(人力资源)** 一个大公司记录了 250 次工作变更发生的原因及工作变更时人员的年龄,数据由下表给出。

变更原因	年龄(岁)				总和
	$\leq 29$	30~39	40~49	$\geq 50$	
辞职	30	6	4	20	60
调动	12	45	4	5	66
退休/解雇	8	9	52	55	124
总和	50	60	60	80	250

基于行的总和,作百分数对比,并由此描述这组数据。

**3.54** 在练习 3.53 中,用基于列总和的百分数对比,你会得到什么不同的结论? 实际做一下,并给出你的结果。

**3.55(工程)** 在隐形眼镜镜片制造中,镜片的强度必须很接近规定的值。在论文“An ANOM-type test for variance from normal populations(正态总体方差的一个 ANOM 型检验)”[技术度量(*Technometrics*)(1997)39:274~283]中,比较了几个供应商的透片强度的一致性。下表给出了用来自三个供应商的材料生产的镜片强度与规定值的偏差:

供应商	与规定值的偏差									
1	189.9	191.9	190.9	183.8	185.5	190.9	192.8	188.4	189.0	
2	156.6	158.4	157.7	154.1	152.3	161.5	158.1	150.9	156.9	
3	218.6	208.4	187.1	199.5	202.0	211.1	197.6	204.4	206.8	

- 对每个供应商,计算相应的偏差的均值和标准差。
- 画出偏差样本数据的图。
- 描述三种镜片对于指定强度的偏差。
- 使用哪个供应商提供的材料,可以生产出强度与规定值最接近的镜片?

**3.56(商业)** 联邦政府极其关注货币增长与制定的增长目标之间的关系。下面列出反映美国货币供应量的两种指标,即 M2(私人支票存款,现金及储蓄)和 M3(M2 加上某些投资)的连续 20 月的数据。

- 散点图能描述 M2 和 M3 间的关系吗?
- 作散点图。二者有明显的关系吗?

月份	货币供应(单位:1 万亿美元)	
	M2	M3
1	2.25	2.81
2	2.27	2.84
3	2.28	2.86
4	2.29	2.88
5	2.31	2.90
6	2.32	2.92
7	2.35	2.96
8	2.37	2.99
9	2.40	3.02
10	2.42	3.04
11	2.43	3.05
12	2.42	3.05
13	2.44	3.08
14	2.47	3.10
15	2.49	3.10
16	2.51	3.13
17	2.53	3.17
18	2.53	3.18
19	2.54	3.19
20	2.55	3.20

**3.57** 在练习 3.56 中,其他什么数据图可用来描述和概括这些数据?作出这种图并解释你的结果。

### 3.8 小 结

本章讨论了数据的图形描述和数值描述方法。饼图和条形图尤其适合于定性变量数据的图形显示,而频数直方图、频率直方图及茎叶图是只适用于定量变量数据的图形方法。

数据的各个数值描述性度量使我们对测量值分布的形态有一个理性的想象。中心趋势度量包括众数、中位数和算术平均值。变异性度量包括一组测量值的极差,四分位数间距和标准差。

我们还把数据描述的概念推广到概括两个定性变量之间的关系。其中用交叉列表法导出了百分数对比的方法。我们还介绍了可以概括定性变量与定量变量及两个定量变量之间的关系的各种图。这里提到的这些内容(即变量间的关系的概

括)将在有关  $\chi^2$  方法,方差分析和回归的章节中加以讨论和扩充。

### 重要公式

$$1. \text{中位数, 分组数据} \quad \text{中位数} = L + \frac{W}{f_m}(0.5n - cf_b)$$

$$2. \text{样本均值} \quad \bar{y} = \frac{\sum y_i}{n}$$

$$3. \text{样本均值, 分组数据} \quad \bar{y} \cong \frac{\sum f_i y_i}{n}$$

$$4. \text{样本方差} \quad s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$$

$$5. \text{样本方差, 分组数据} \quad s^2 \cong \frac{1}{n-1} \sum f_i (y_i - \bar{y})^2$$

$$6. \text{样本标准差} \quad s = \sqrt{s^2}$$

$$7. \text{样本变异系数。} \quad CV = \frac{s}{|\bar{y}|}$$

### 补充练习

**3.58(环境)** 为了控制商业核电站在发生熄火时造成严重的反应堆心损坏的风险,在需要时启动的应急柴油发电机必须有很高的可靠性。论文“Empirical Bayes estimation of the reliability of nuclear-power emergency diesel generators (核能应急柴油发电机可靠性的经验 Bayes 估计)”[技术度量(*Technometrics*)(1996)38: 11-23]收集有七个核电站发生事故时的数据。下列数据是从 1982 年到 1998 年某一个站的柴油发电机在失效之间成功启动的次数。

28 50 193 55 4 7 147 76 10 0 84 0 9 1 0 62

26 15 226 54 46 128 4 105 40 273 164 7 55 41 26 6

(注意:柴油发电机的失败并不一定造成核电站反应堆心的损坏,因为所有的核电站都备有若干应急柴油发电机。)

- 计算在失效之间成功启动次数的均值和中位数。
- 哪一个度量看上去最能代表数据的中心?
- 计算极差和标准差  $s$ 。
- 用极差近似的方法来估计  $s$ 。近似值与真实值如何接近?
- 作区间

$$\bar{y} \pm s \quad \bar{y} \pm 2s \quad \bar{y} \pm 3s,$$

分别数出在两次失效间隔中,落入三个区间的成功启动次数。把这些数变为百分数,并与经验准则中的百分比作比较。

- 为什么你认为你的百分比与经验准则中相符不好?

**3.59(环境)** 佛罗里达大学的口腔医学院作出一项承诺,利用学生可自己掌

握进度的教学材料,例如录像带,幻灯片和教学大纲来教授所有课程,以期所有学生都能以适合于自己能力的进度进行学习,并且指导人员在与学生接触的过程中有更多的时间留给个人咨询。已经开发了一套这样的教学材料,并首批对 50 个从头至尾学习了这门课的学生进行了试验。下面的测量值是这些学生完成被要求的所有材料所花费的时间(小时)。

16 8 33 21 34 17 12 14 27 6  
 33 25 16 7 15 18 25 29 19 27  
 5 12 29 22 14 25 21 17 9 4  
 12 15 13 11 6 9 26 5 16 5  
 9 11 5 4 5 23 21 10 17 15

- 计算完成时间的众数,中位数和均值。
- 猜测  $s$  的值。
- 利用快捷的公式计算  $s$  的值,并与(b)中答案作比较。
- 你认为经验准则充分地描述了这些数据的变异性吗? 解释你的理由。

**3.60(商业)** 消费者报告杂志(*Consumer Report*)1998年2月号提供了24种牌子的纸巾的价格。由于各种品牌的纸巾每卷纸的张数不同,下面给出了每种纸巾每张纸和每卷纸的价格。

品牌	每卷价格	每卷纸的张数	每张纸的价格
1	1.59	50	0.0318
2	0.89	55	0.0162
3	0.97	64	0.0152
4	1.49	96	0.0155
5	1.56	90	0.0173
6	0.84	60	0.0140
7	0.79	52	0.0152
8	0.75	72	0.0104
9	0.72	80	0.0090
10	0.53	52	0.0102
11	0.59	85	0.0069
12	0.89	80	0.0111
13	0.67	85	0.0079
14	0.66	80	0.0083
15	0.59	80	0.0074
16	0.76	80	0.0095

续表

品牌	每卷价格	每卷纸的张数	每张纸的价格
17	0.85	85	0.0100
18	0.59	85	0.0069
19	0.57	78	0.0073
20	1.78	180	0.0099
21	1.98	180	0.0011
22	0.67	100	0.0067
23	0.79	100	0.0079
24	0.55	90	0.0061

- 计算每张纸和每卷纸价格的标准差。
- 每张纸和每卷纸的价格,哪个有更大的变异?
- 在(b)的对比中,你应用  $s$  还是用 CV 做对比?说明你的答案。

**3.61** 画出每卷纸的价格和每卷纸张数的散点图。

- 24 个点在一条直线上吗?
- 若不在一条直线上,每卷纸的张数与每卷纸的价格有其他类型的关系吗?
- 什么因素可以解释,为何每卷纸的价格和纸张数的比率不是常数?

**3.62** 作每卷纸的价格和每卷纸张数的盒形图。在这些数据中有些“不一般”的品牌吗?

**3.63(生物)** 作一项研究,给绵羊静脉注射抑制尿分泌的荷尔蒙 ADH 后,测量一下绵羊的尿流量(毫升/分钟)。10 只绵羊的尿流量如下。

0.7 0.5 0.5 0.6 0.5 0.4 0.3 0.9 1.2 0.9

- 计算这些样本数据的均值,中位数和众数。
- 假设最大的测量值是 6.8 而不是 1.2,将如何影响到均值,中位数和众数?

**3.64** 参见练习 3.63。

- 计算极差和样本标准差。
- 用极差给出的近似值,检查你计算所得的  $s$ 。
- 若最大的测量值是 6.8 而不是 1.2,将如何影响到极差和标准差?若是 68 呢?

**3.65(环境)** 论文“Conditional simulation of waste-site performance (垃圾厂性状的条件模拟)”[技术度量(*Technometrics*)(1994)36:129~16]讨论了一种控制装置的评估问题。这种控制装置用来演示对放射性超铀垃圾的安全管理、存储及处理。研究人员已经知道,放射性核物质一种可能的释放方式是通过地下水的污染转移。最近的研究集中于可传送性(transmissivity)分析。所谓可传送性,是含水

土层性质和厚度的一个函数,它反映了水通过含水土层移转的速度。下表含有 41 个可传递性  $T$  的测量值。

9.345	6.302	24.609	10.093	0.939	354.81	15399.27
88.17	1253.43	0.75	312.10	1.94	3.28	1.32
7.68	2.31	16.69	2772.68	0.92	10.75	0.000753
1.08	741.99	3.23	6.45	2.69	3.98	2876.07
12201.13	4273.66	207.06	2.50	2.80	5.05	3.01
462.38	5515.69	118.28	10752.27	956.97	20.43	

- 画出  $T$  的 41 个测量值的频率直方图。
- 描述直方图的形状。
- 当频率直方图右偏很大时,经验准则可能不会产生很精确的结论。利用所给的数据验证此说法。
- 数据分析人员发现,丘形频率直方图比较容易处理。对数据进行某个变换后有时会得到这个形状。把上表中的 41 个  $T$  值都取以 10 为底的对数,并利用所得值重新作频率直方图。该图的形状比起原始数据的频率直方图来更近似于丘形吗?对变换以后的数据应用经验准则,并由此验证,比起对于原始数据的应用来,经验准则此时产生了更精确的结论。

**3.66** 在练习 3.20 中,对拥有房产权的数据,计算其均值,中位数和标准差。

- 对于 1996 年的数据,比较其均值和中位数。对于这个数据集合,哪一个最合适?解释你的理由。
- 这两个年份的概括统计量有本质区别吗?从这些概括统计量,对于在 11 年间拥有房产权的家庭比例数的变化,你能得到什么结论?

**3.67** 在练习 3.20 中,对于两个年份的数据,画出盒形图。

- 这些年来,拥有房产权的家庭的百分比变化了吗?
- 是否有某些州,其房产拥有率极低?
- 有一些州有极高的房产拥有率吗?
- 房产权拥有率低的一类中,各州存在相似之处吗?房产权拥有率高的一类中呢?

**3.68(社会)** 为了获得谋杀率的有关信息,研究了由 90 个标准大型城市统计地区(SMSA)的组成的随机样本,记录其谋杀率(每 100,000 人中谋杀发生的次数),所得数据总结成下面的频数表。



分组区间	$f_i$	分组区间	$f_i$
-0.5~1.5	2	13.5~15.5	9
1.5~3.5	18	15.5~17.5	4
3.5~5.5	15	17.5~19.5	2
5.5~7.5	13	19.5~21.5	1
7.5~9.5	9	21.5~23.5	1
9.5~11.5	8	23.5~25.5	1
11.5~13.5	7		

作这些数据的频率直方图。

**3.69** 对练习 3.68 中的数据,

- 计算样本中位数和众数。
- 计算样本均值。
- 你将用哪一个中心趋势度量描述谋杀率分布的中心?

**3.70** 对练习 3.68 中的数据,

- 计算四分位数间距。
- 计算样本标准差。

**3.71** 用练习 3.20 中两个年份各州拥有房产权家庭比例的数据,作出这两个年份的数据的分位数图。

- 找出 1996 年拥有房产权的比例数据的 20% 的分位数,并解释这个值。
- 国会想要找出 1996 年拥有房产权百分比最高的州。哪些州拥有房产权的比例落入了前 10% 的行列?
- 找出 1985 年中,房权拥有率落入前 10% 的各州。这些州和 1996 年的那些州有什么不同?

**3.72(工程)** 在生产线上,每 20 分钟抽取一个由 10 只晶体管组成的成品样本,并对其加以检测。从前 500 个样本获得的数据总结如下。

$y_i$	0	1	2	3	4	5	6	7	8	9	10
$f_i$	170	185	75	25	15	10	8	5	4	2	1

作一个频率直方图,并描绘其四分位数间距。(注:上表中  $y_i$  的值是由 10 个晶体管组成的样本中出现的缺陷的数目。)

**3.73** 参见练习 3.72。

- 计算样本中位数和众数。
- 计算样本均值。

c. 根据均值、中位数和众数, 你能指出这个分布是如何偏斜的吗?

**3.74** 可以用经验准则描述练习 3.72 中的这组测量值吗? 利用频率分布证实你的答案。

**3.75(政府)** 把健康和医疗服务的人均支出(美元)按州进行整理, 得到如下数据。

美元	$f$
45~59	1
60~74	4
75~89	9
90~104	9
105~119	12
120~134	6
135~149	4
150~164	1
165~179	3
180~194	0
195~209	1
总和	50

a. 作一个频率直方图。

b. 由这些分组支出数据计算  $\bar{y}$  和  $s$  的近似值。

**3.76** 在表 3.4 中的数据中, 从北部除去费城(Philadelphia), 西部除去圣何塞(San Jose)和西雅图(Seattle)。

a. 对修改后的各组计算  $\bar{y}_i$ 。

b. 用公式

$$\bar{y} = \frac{\sum_i n_i \bar{y}_i}{n},$$

把各组中的均值  $\bar{y}_i$  结合起来得到总体均值, 其中  $n_i$  是子分组  $i$  中的观察值的数目。

c. 证明在(b)中得到的样本均值等于从(a)中使用的 87 个测量值得到的样本均值。

**3.77(工程)** 高速公路安全保险协会公布了小型汽车在一系列受控、低速碰撞中的总损失数据, 去掉车牌名称的数据(美元)如下:

361 393 430 543 566 610 763 851  
886 887 976 1,039 1,124 1,267 1,328 1,415  
1,425 1,444 1,476 1,542 1,544 2,048 2,197

- 用 6 或 7 个分组作频数直方图。
- 由直方图,你能猜到均值是多少吗?
- 计算中位数和均值。
- 均值和中位数间的关系,关于数据的形状说明了什么?

**3.78(商业)** 汽车制造商的生产记录表明了每个班的产品数量(每个班的最大产量是 720 辆汽车)。

668 711 625 701 688 667 694 630 547 703 688 697 703  
656 677 700 702 688 691 664 688 679 708 699 667 703

- 对于这些数据,众数是一个有用的概括统计量吗?
- 找出中位数。
- 找出均值。
- 均值和中位数间的关系关于数据的形状说明了什么?

**3.79** 作出练习 3.78 中数据的茎叶图,茎中应当包括(从高到低)71, 70, 69, …… ,茎叶图排列的形状可以证明你在练习 3.78(d)中的判断吗?

**3.80** 对练习 3.79,

- 找出中位数和 IQR。
- 找出内分线、外分线。有离群值吗?
- 画出数据的盒形图。

**3.81(社会)** 收集了城市家庭每周食物开支(包括饭店开支)的样本数据。数据是从每户户主的日常记录中得到的,将其按每个家庭成员的数目分类。开支数据如下:

1 个成员: 67 62 168 128 131 118 80 53 99 68  
76 55 84 77 70 140 84 65 67 183  
2 个成员: 129 116 122 70 141 102 120 75 114 81 106 95  
94 98 85 81 67 69 119 105 94 94 92  
3 个成员: 79 99 171 145 86 100 116 125  
82 142 82 94 85 191 100 116  
4 个成员: 139 251 93 155 158 114 108  
5+ 个成员: 111 106 99 132 62 129 91  
121 128 129 140 206 111 104 109 135 136

- 对于每一个家庭成员数,计算该类家庭开支的均值。

b. 对于每一个家庭成员数, 计算该类家庭开支的中位数。

**3.82** 对练习 3.81 中的数据, 回答下列问题:

- 用原始数据, 计算由所有数据组成的数据集合的均值。
- 所有数据的均值可以通过结合由家庭成员数目所确定的每个类的均值来得到吗?
- 用原始数据计算所有数据的中位数。
- 通过由每个成员数所确定的类的中位数可以计算所有数据的中位数吗?

**3.83(人力资源)** 一家公司为了取消对于未受豁免的雇员的时间钟点和打卡限制, 修改了一项长期执行的政策。随着这一变化, 所有雇员(豁免的和未被豁免的)都要记录他们的出勤时间, 以及由于生病、周末、假日等原因而造成的缺勤。以前的制度规定了一定数目的病休日, 这项制度已被取消。在新制度下, 若一个雇员生病了, 他可以带薪缺勤, 否则, 他就要出勤。

为了看新制度的效果如何, 抽取了 15 个雇员的随机样本, 并记录下他们今年的病休天数(在新制度下), 以及对应的去年的天数。数据如下:

雇员	今年(新制度)	去年(老制度)
1	0	2
2	0	2
3	0	3
4	0	4
5	2	5
6	1	2
7	1	6
8	3	8
9	1	5
10	0	4
11	5	5
12	6	12
13	1	3
14	2	4
15	12	4

- 对每一列计算均值和标准差。
- 由样本数据, 你对新制度得到什么结论或推断? 解释你的理由。

**3.84** 在练习 3.83 中, 若在每一列中都除去 12, 以 7 替代, 对  $\bar{y}$  和  $s$  将会有何影响? 新旧制度的极差受这一替换的影响吗?

**3.85(政府)** 联邦当局捣毁了大量的野生和栽培的大麻。下表给出了 15 个州在 12 个月的时间内捣毁的大麻植株数量和逮捕的人数。

州	植株	逮捕人数
1	110,010	280
2	256,000	460
3	665	6
4	367,000	66
5	4700,000	15
6	4,500	8
7	247,000	36
8	300,200	300
9	3,100	9
10	1,250	4
11	3900,200	14
12	68,100	185
13	450	5
14	2,600	4
15	205,844	33

a. 讨论用样本均值描述这两个变量的适用性。

b. 计算样本均值, 10% 的截尾均值和 20% 的截尾均值。对每个变量, 哪个截尾均值更合适, 为什么?

**3.86** 参见练习 3.85。在毁掉的植株数量与逮捕的人数之间有关系吗? 你如何检查这个问题呢? 有什么其他的变量可能与捣毁的大麻数量有关系?

**3.87(商业)** FDC 指数每月的读数是制药业景气状况公认的晴雨表。对这一指数有如下的数据表。从表中我们看到, FDC 指数有几个分量——这些分量分别是针对制药公司、多种经营公司、连锁药店和药物与医疗设备批发商的, 这几种公司各有一项。

	药物公司	多种经营公司	连锁药店	批发商
一月	123.1	154.6	393.3	475.5
二月	122.4	146.0	407.6	504.1
三月	125.2	169.2	405.0	476.6
四月	136.1	156.7	415.1	513.3
五月	149.3	177.0	418.9	543.5
六月	145.7	158.1	443.2	552.6

续表

	药物公司	多种经营公司	连锁药店	批发商
七月	162.4	156.6	419.1	526.2
八月	168.0	178.6	404.0	516.3
九月	155.6	170.4	391.8	482.1
十月	177.0	162.9	410.9	484.0
十一月	196.6	182.4	459.8	522.6
十二月	195.2	195.4	431.9	536.8

a. 在一张图中画出这些数据。

b. 讨论每一分量的趋势, 以及 FDC 指数内各个不同的分量之间的明显的关系。

**3.88** 参见练习 3.87。计算 FDC 指数的每一个分量每月变化的百分比(假设一月份这四个分量各自变化的百分比是 12.3, -0.7, 12.1, 16.1)。作出这些数据的图。比起原始数据, 它们揭示了更多信息吗?

**3.89(商业)** 最受关注的关于纽约股票交易所(NYSE)交易状况的指数是道琼斯工业平均指数(DJIA)。这个指数是由 30 个公司的股票价格计算得到的。在 1896 年, DJIA 刚被提出的时候, 它是 12 个股票的平均价格。多年来, 新的公司被加入其中, 也有公司被从中排除, 从而指数有所调整; 同时, 当公司分股时, 指数会有所变动。下表给出 DJIA 中的 30 个公司的股票在纽约股票交易所(NYSE)收盘时(1999 年 5 月 20 日)的价格。

a. 计算 DJIA 中 30 个股票价格的平均价格。

b. 计算 DJIA 中 30 个股票价格的极差。

c. DJIA 不再是一个平均数, 名称中包括“平均”的字样仅是由于历史的原因。这个指数是通过指数中所包含的所有股票价格的总和除以一个常数  $C$  而得到, 这个常数  $C$  随指数中股票的增减以及分裂而变化。

$$DJIA = \frac{\sum_{i=1}^{30} y_i}{C},$$

其中  $y_i$  为股票  $i$  的收盘价格,  $C = 0.211907$ 。利用这里给出的股票价格, 计算 1999 年 5 月 20 日的 DJIA。

d. DJIA 是一种数据的概括。利用样本数据计算得到的 DJIA 提供了某个总体的信息吗? 若是, 该总体是什么总体? 样本是随机样本吗?

DJIA 的各个成分

公 司	DJIA 的百分比	NYSE 的股票价格
Allied-Signal	2.640	60.8125
Alcoa	2.532	58.3125
American Express	5.357	123.3750
AT&T	2.570	59.1875
Boeing	1.948	44.8750
Caterpillar	2.559	58.9375
Chevron	4.060	93.5000
Citigroup	2.996	69.0000
Coca-Cola	2.950	67.9375
duPont	2.999	69.0625
Eastman Kodak	3.281	75.5625
Exxon	3.474	80.0000
General Electric	4.619	106.3750
General Motors	3.479	80.1250
Goodyear	2.578	59.3750
Hewlett-Packard	4.090	94.1875
IBM	2.380	232.5000
International Paper	2.380	54.8125
J. P. Morgan	6.049	139.3125
Johnson & Johnson	4.125	95.0000
McDonald's	1.739	40.0625
Merck	3.126	72.0000
Minnesota Mining	3.884	89.4375
Phillip Morris	1.745	40.1875
Procter & Gamble	4.198	96.6875
Sears, Roebuck	2.125	48.9375
Union Carbide	2.421	55.7500
United Technologies	2.640	60.8125
Wal-Mart Stores	1.935	44.5625
Walt Disney	1.294	29.8125

**3.90** 在练习 3.20 中, 已经对 1985 年和 1996 年的拥有房产权家庭的比例的数据作了频率直方图。

- 通过检查这些图, 你认为用经验准则描述这些数据合适吗?
- 计算 1996 年的  $\bar{y}$  和  $s$ , 计算落入区间  $\bar{y} \pm s$ ,  $\bar{y} \pm 2s$ ,  $\bar{y} \pm 3s$  的测量值的百分比, 这些数据与经验准则提供的百分比一致吗?

**3.91** 参见练习 3.90。数据中有许多影响  $\bar{y}$  的极端值吗? 从练习 3.21 中

画出的数据图中,能预先看出这一点吗? 计算这些数据 10% 的截尾均值。

**3.92(人力资源)** 作为中层管理者选拔过程考察的一部分,作一项调查以研究雇佣来源(内部提升,来自相关行业,来自不相关行业)和 3 年工作经历(晋升,同一职务,辞职,解雇)之间的关系。120 个中层管理者数据如下。

工作经历	来 源			总和
	公司内部	相关行业	不相关行业	
提升	13	4	10	27
统一职务	32	8	18	58
辞职	9	6	10	25
解雇	3	3	4	10
总和	57	21	42	120

a. 计算每一来源中的各种工作经历所占的百分比。

b. 你能断定在来源和工作历史间有很强的相关性吗?

**3.93(环境)** 调查了主要产煤的州中的 150 个居民,主要产石油和天然气的州中的 200 个居民以及其他州的 450 个居民,每个居民选择一项最喜欢的国家能源政策。下面给出的这个调查的结果是由 Spss 打印输出的。

COUNT ROW PCT COL PCT TOT PCT	STATE			ROW TOTAL
	COAL	OIL AND GAS	OTHER	
OPINION	62	25	102	189
COAL ENCOURAGED	32.8	13.2	54.0	23.6
	41.3	12.5	22.7	
	7.8	3.1	12.8	
	3	12	26	41
FUSION DEVELOP	7.3	29.3	63.4	5.1
	2.0	6.0	5.8	
	0.4	1.5	3.3	
	8	6	22	36
NUCLEAR DEVELOP	22.2	16.7	61.1	4.5
	5.3	3.0	4.9	
	1.0	0.8	2.8	



续表

COUNT ROW PCT COL PCT TOT PCT	STATE			ROW TOTAL
	COAL	OIL AND GAS	OTHER	
	19	79	53	
	12.6	52.3	35.1	
	12.7	39.5	11.8	
OIL DEREGULATION	2.4	9.9	6.6	151
	58	78	247	383
	15.1	20.4	64.5	47.9
SOLAR DEVELOP	38.7	39.0	54.9	
	7.3	9.8	30.9	
COLUMN	150	200	450	800
TOTAL	18.8	25.0	56.3	100.0

- 解释上述交叉列表中左上方的数值 62, 32.8, 41.3 和 7.8 的意义, 注意表的左上方列出的 COUNT, ROW PCT, COL PCT TOT PCT。
- 哪些百分比的计算结果看上去对你更有意义?
- 根据你选择的百分比计算结果, 在州和居民的观点之间有很强的相关性吗?

**3.94(商业)** 在许多小的中西部城市, 代表环卫工人的市政联合会研究了前几年签署的合同, 合同被细分为: 未经罢工而谈判签署的, 未罢工经仲裁而签署的以及罢工后签署的。每一个合同都确定了第一年工资增长的百分比。总结数据如下。

合同类型	谈判	仲裁	罢工后
工资增长的平均百分比	8.20	9.42	8.40
方差	0.87	1.04	1.47
标准差	0.93	1.02	1.21
样本容量	38	16	6

在合同类型和工资增长的平均百分比间有一定的关系吗? 若你是经理而不是工会成员, 在将来的合同签署中你将采取什么姿态?

## 第四部分 工具和概念

### 第四章 概率和概率分布



## 第四章 概率和概率分布

- 4.1 如何应用概率进行推断
- 4.2 确定一个事件的概率
- 4.3 基本的事件关系和概率法则
- 4.4 条件概率和独立性
- 4.5 Bayes 公式
- 4.6 离散变量和连续变量
- 4.7 离散随机变量的概率分布
- 4.8 一个常用的离散随机变量:二项分布
- 4.9 连续随机变量的概率分布
- 4.10 一个常用的连续随机变量:正态分布
- 4.11 随机抽样
- 4.12 抽样分布
- 4.13 二项分布的正态逼近
- 4.14 Minitab 指令
- 4.15 小结

### 4.1 如何应用概率进行推断

我们在第一章曾提到,一位科学家应用统计推断,基于从总体中抽选的样本中所包含的信息,得到关于总体的阐述。我们还在第三章提出了图表和数值描述技术,作为概括和描述一个样本的工具。然而,一个从总体中选取的样本并不等同于其总体。我们需要说明以样本均值、样本标准差,以及样本比例来代表与其相应的总体值,其精度如何。

大多数管理决策必须面对不确定性。新型汽车的价格和设计,必须在对消费者偏爱、国家经济的发展趋势以及竞争行为的不确定性预测的基础上来选择。医院职员队伍的规模和分配必须在关于病人负担只有有限信息的情况下来决定。产品库存的设置得面临需求的不确定性。概率就是用来表达不确定性的语言。现在让我们来考察概率这一推理的机制。其中的思想最好用一个例子说明。

Newsweek(《新闻周刊》)在 1998 年 6 月 20 日的一期中,提出了这样的问题:

“谁需要医生？家庭检查的兴隆。”文章就家用医学筛选检验的戏剧性增加展开了讨论。家用检查的市场，从两种最频繁使用的检查——怀孕和糖尿病患者的葡萄糖监测，扩展到五花八门的以前仅仅由医生和合格的实验室才能做的诊断检查。有决定双胞胎是同卵双生还是异卵双生的 DNA 检查，有胆固醇水平的检查，有结肠癌的筛选检查，也有确认你在少年时是否是一个毒品使用者的检查。然而，需要回答的主要问题是，这些用于检查的器械在多大程度上是可靠的？当检查显示一个女性没有怀孕时，检查不正确和女性真正怀孕的机率是多少？这类不正确的家庭检查结果可能导致一个女性没有在她怀孕的初期去寻求合适的产前护理。

假定一个公司在宣传材料中说它的怀孕检测方法应用于怀孕妇女时能提供 75% 的正确结果。我们想要评估这种说法，因此选择了 20 位妇女，她们的医生通过可能是最好的检测方法，确诊她们已经怀孕。这 20 位妇女都接受了这个公司的检测，并且她们的检测结果都呈阴性，这表明 20 人中没有一人怀孕。假如你再次确认了这 20 位妇女事实上都已经怀孕，就象在这次检测做完后几个月所能确定的那样，你对该公司关于其检测的可靠性的说法有什么样的结论？

如果该公司关于可靠性为 75% 的说法是正确的，我们可以期望样本中阳性的检测结果的比例在某种程度上与 75% 接近。然而，没有一个检测结果是阳性的。这样，我们就会得出结论，认为公司的说法可能是错的。为什么我们不能肯定地说公司的说法是错的呢？让我们来考虑一下上述检测结果的可能性。假定我们有一个由百万个个体组成的大总体，其中 75% 的个体为 P 即阳性的，而 25% 的个体为 N 即阴性的。我们随机地从该总体中选取 20 个个体，并且记数样本中 P 的个数。要获得一个包含 0 个 P 和 20 个 N 的样本可能吗？当然是有可能的，但可能性非常小。在本章后面的内容中，我们将计算这样的样本出现的概率。

为了更好地看清楚概率在利用样本的结果进行推断从而得到关于总体的结论这一过程中所起的作用，假设这 20 个检测结果中有 14 是阳性的——即 70% 的正确反应率。你会认为这个结果极不可能，从而拒绝这家公司的 75% 的正确反应率的说法吗？如果检测结果是 12 个阳性和 8 个阴性或 16 个阳性和 4 个阴性，又会如何呢？假若公司的说法是正确的，我们在什么情况下，认为所观察到的样本结果很不可能出现，以至我们不同意公司的说法呢？为了回答这个问题，我们必须知道如何去算出观察到某个特殊样本结果的概率。知道了这个概率，我们就可以决定是同意还是不同意公司的说法。概率是一种工具，利用它能够作出推断。在这一章的后面我们将详细讨论 FDA 和私人公司如何确定筛选检查的可靠性。

因为概率是进行推断的一种工具，我们需要给出它的定义。在前面的讨论中，我们在其日常的意义下使用了 *概率* 这个名词。现在我们更仔细地来考察这个概念。

对现象的观察能产生许多不同的结果，其中一些比其他有更大的可能性。为

了对一个试验结果出现的概率给出精确定义,人们进行了大量的尝试。我们将引用其中的三条。

概率的第一种解释叫做**概率的古典解释**,是从碰运气的游戏中产生的。这种类型的解释的典型说法是,例如,“将一枚均匀的硬币抛起显示出正面朝上的概率是  $1/2$ ”,“从 52 张标准扑克牌中抽出一张纸牌,得到 A 的概率是  $4/52$ ”。这些概率的数值是从游戏的特性中得来的。一个硬币抛起有 2 种可能的结果(正面朝上或反面朝上);正面朝上的概率应该是  $1/2$ (两者之一)。同样,在 52 张标准扑克牌中有 4 张 A,这样,从中抽一张牌,抽出的牌是 A 的概率是  $4/52$ ,即从 52 中取 4。

在概率的古典解释中,每种可能的不同结果被称为一个基本事件;一个事件被认为是许多基本事件的集合。在概率的古典解释下,一个事件  $E$  的概率是用有利于事件  $E$  的基本事件数  $N_e$  与所有可能的基本事件总数  $N$  的比值来计算的:

$$P(\text{事件 } E) = \frac{N_e}{N}$$

这种解释的适用性取决于所有基本事件都是等可能的假设。如果这种假设不成立,由概率的古典解释确定的概率将是错误的。

概率的第二种解释被称之为**概率的频率概念**,是定义概率的一种经验方法。如果试验重复了很多次,事件  $E$  发生的次数占 30%,则 0.30 应该是事件  $E$  的概率的很好的近似。用符号表示,如果试验进行了  $n$  次,并且在这些试验中事件  $E$  发生了  $n_e$  次,则事件  $E$  的概率近似于

$$P(\text{事件 } E) \approx \frac{n_e}{n}$$

我们之所以说“近似于”,是因为我们认为,确切的概率  $P(\text{事件 } E)$  是在对现象进行了大量的观察或重复时事件  $E$  发生的频率。由于我们可以对概率的频率解释(通过大量重复的模拟试验)进行验证,这种解释不仅生动而且也很实用。

概率的第三种解释可以应用于不大可能进行重复试验的问题。现实中有许多“一锤子买卖”的情况。例如,州福利机构的主任评估州立法机关通过对资格认证条款的修改提议的概率时,就很难想象去做一系列这样的尝试。相反,他使用**个人或主观的概率**就所提议的资格认证条款的修改被通过的可能性给出一个一次性的估量。主观概率的问题是因人而异并且不能被检查。

在以上三种概率的解释中,频率概念当为最合理的一个,因为它为大多数要研究的事件提供了概率的实际解释。尽管我们不去做那么大量的重复试验以确定一个事件的准确概率,但对一个事件的概率可以进行验证这一点赋予了频率以意义。在本书后面的部分中,我们将倚重对概率的这种解释。

## 练习

### 应用

4.1 指出下列哪个关于概率解释的陈述是最恰当的。

- 养牛行业协会声称在这个夏季牛肉的批发价格将有 60/40 的机会上涨,也就是说,以 0.60 的概率涨价,0.40 的概率降价。
- 一家大化学药品生产公司的质量控制部门进行了一项工序校验研究。从这项研究中,质量控制部门声称一批最新的化学药品的保存期限超过指定的最小期限的概率是 0.998。
- 一个大公司的市场营销部将推出一种新的混合咖啡。初步的市场调查结果表明,随机抽取的 1 000 名潜在顾客中有 550 个人认为新的混合咖啡比一种品牌咖啡好。假定消费者对两种品牌的偏爱没有差别,那么上述结果发生的概率近似于 0.001。
- 用 24 小时邮递业务投寄的包裹,顾客在第二天收到的概率是 0.92。
- 在得克萨斯州的大学橄榄球比赛中,现场解说员说安吉斯(Aggies)球队战胜佛罗里达大学队的概率是 0.75。
- 在给定的一天内,一个核电厂熔融的概率是 0.00001。
- 如果一个顾客购买得克萨斯州的一张彩票,这张票中奖的概率是  $1/15\,890,700$ 。

4.2 对下列各种情况,给出你自己的主观概率。把全班的这些概率做成表将是有益的。对于哪些情况,你们给出的概率有很大的差别?

- 联邦所得税将被取消。
- 你的这门课的得分会是 A。
- 在一个班里有 2 人或 2 人以上同一天生日。
- 明年将有一颗流星撞击地球。
- 下届总统选举中,一位妇女将被选为美国的副总统或总统。

## 4.2 确定一个事件的概率

在前面的一节中,我们讨论了概率的三种不同的解释。在本节,我们将用概率的古典解释和频率的概念说明基本事件或事件的概率的计算方法。假设我们有一枚一分和一枚一角硬币,考虑掷这两枚硬币组成的试验,并且观察朝上的那面。有 4 个可能的基本事件:

TT:两枚硬币都是反面朝上;

TH:一分是反面朝上,一角硬币是正面朝上;

HT:一分是正面朝上,一角硬币是反面朝上;

HH:两枚硬币都是正面朝上。

观察到 2 枚硬币恰有一个正面朝上的事件的概率是多少?

如果我们能够假设所有的 4 个基本事件是等可能的,这个概率很容易被求出。在这个例子中,这种等可能的假设似乎是相当合理的。有  $N=4$  个可能的基本事件,其中  $N_e=2$  个有利于我们关心的事件,即观察到恰有一个正面朝上的。因此,由概率的古典解释,有

$$P(\text{恰有一个是正面朝上}) = \frac{2}{4} = \frac{1}{2}$$

因为所关心的事件的概率有一个频率的解释,使用频率的概念,我们也能获得同样的经验结果。为了阐明如何用频率来获得一个事件的概率,我们将应用模拟的思想。模拟是一种能产生与真实情况下的事件具有相同的发生概率的基本事件的技术,计算机是产生这些基本事件的一个方便工具。假定我们要模拟 2 枚硬币的 1 000 次投掷。我们可以使用一个计算机程序,比如 SAS 或 Minitab 来模拟这 2 枚硬币的投掷。该程序有一个随机数发生器。我们指定偶数为 H,奇数为 T。由于在一位数中有 5 个偶数和 5 个奇数,获得一个偶数的概率是  $5/10=0.5$ ,获得奇数的概率也一样。这样,我们需要给出 500 对数字,这个 500 对数字的集合将代表 2 枚硬币 500 次投掷的结果,每一对中的第一个数字代表投掷一分硬币的结果,而第二个数字代表掷一角硬币的结果。例如,一对数字(3,6)将代表一分出现反面而一角的硬币出现正面。使用版本 13 的 Minitab,下列步骤将产生 1 000 个范围在 0~9 的随机数字:

1. 从工具栏中选择“Calc(计算)”。
2. 从列表中选择“Random Data(随机数)”。
3. 从列表中选择“Integer(整数)”。
4. 生成 20 个数据行。
5. 存到列:c1~c50 中。
6. 最小值:0。
7. 最大值:9。

上述步骤将产生 1,000 个随机的一位数字,然后我们可以把这些数字组成 500 对一位数字。(大多数计算机软件都包含有随机数发生器,利用这些随机数发生器可以得到类似的结果。)

```
25 32 70 15 96 87 80 43 15 77 89 51 08 36 29 55 42 86 45 93 68 72 49 99 37
82 81 58 50 85 27 99 41 10 31 42 35 50 02 68 33 50 93 73 62 15 15 90 97 24
46 86 89 82 20 23 63 59 50 40 32 72 59 62 58 53 01 85 49 27 31 48 53 07 78
15 81 39 83 79 21 88 57 35 33 49 37 85 42 28 38 50 43 82 47 01 55 42 02 52
```



66 44 15 40 29 73 11 06 79 81 49 64 32 06 07 31 07 78 73 07 26 36 39 20 14  
 48 20 27 73 53 21 44 16 00 33 43 95 21 08 19 60 68 30 99 27 22 74 65 22 05  
 26 79 54 64 94 01 21 47 86 94 24 41 06 81 16 07 30 34 99 54 68 37 38 71 79  
 86 12 83 09 27 60 49 54 21 92 64 57 07 39 04 66 73 76 74 93 50 56 23 41 23  
 18 87 21 48 75 63 09 97 96 86 85 68 65 35 92 40 57 87 82 71 04 16 01 03 45  
 52 79 14 12 94 51 39 40 42 17 32 94 42 34 68 17 39 32 38 03 75 56 79 79 57  
 07 40 96 46 22 04 12 90 80 71 46 11 18 81 54 95 47 72 06 07 66 05 59 34 81  
 66 79 83 82 62 20 75 71 73 79 48 86 83 74 04 13 36 87 96 11 39 81 59 41 70  
 21 47 34 02 05 73 71 57 64 58 05 16 57 27 66 92 97 68 18 52 09 45 34 80 57  
 87 22 18 65 66 18 84 31 09 38 05 67 10 45 03 48 52 48 33 36 00 49 39 55 35  
 70 84 50 37 58 41 08 62 42 64 02 29 33 68 87 58 52 39 98 78 72 13 13 15 96  
 57 32 98 05 83 39 13 39 37 08 17 01 35 13 98 66 89 40 29 47 37 65 86 73 42  
 85 65 78 05 24 65 24 92 03 46 67 48 90 60 02 61 21 12 80 70 35 15 40 52 76  
 29 11 45 22 38 33 32 52 17 20 03 26 34 18 85 46 52 66 63 30 84 53 76 47 21  
 42 97 56 38 41 87 14 43 30 35 99 06 76 67 00 47 83 32 52 42 48 51 69 15 18  
 08 30 37 89 17 89 23 58 13 93 17 44 09 08 61 05 35 44 91 89 35 15 06 39 27

500 次投掷硬币的模拟结果在表 4.1 中给出。

表 4.1 投掷一枚一分和一枚一角硬币的 500 次模拟

事件	模拟基本事件	频数	频率
TT	(奇数, 奇数)	129	$129/500 = 0.258$
TH	(奇数, 偶数)	117	$117/500 = 0.234$
HT	(偶数, 奇数)	125	$125/500 = 0.250$
HH	(偶数, 偶数)	129	$129/500 = 0.258$

注意, 这种模拟方法产生的概率很接近我们的直觉; 也就是说, 对于这些基本事件, 我们可能从直观上预期到它们是等可能的。这样, 4 个基本事件发生的概率都等于  $1/4$ , 或  $0.25$ 。这正是我们在概率的古典解释中所作出的假设。我们将在第十章说明, 为了使模拟得到的概率落在真实概率的  $0.01$  范围内的把握为  $95\%$ , 投掷的次数至少应该是  $7500$ , 而不是我们前面所用的  $500$  次。

如果希望求出投掷 2 枚硬币恰好观察到一个硬币正面朝上的概率, 从表 4.1 我们得到

$$P(\text{恰有一个是正面朝上}) \approx \frac{117 + 125}{500} = 0.484$$

这个值很接近于理论上的概率,也就是我们曾经给出的 0.5。

注意,我们可以很容易地把上面的例子进行修改,使之适用于不均匀硬币的投掷。假定我们投掷一个加重的一分硬币,抛一次时该硬币正面朝上的概率是 0.70,而反面朝上的概率是 0.30。我们可以指定  $H$  这个基本事件当取得随机数为 0, 1, 2, 3, 4, 5, 6 之一时发生,而  $T$  这个基本事件当取得随机数为 7, 8, 9 之一时发生。象前面一样,运行同一个模拟程序,但我们对输出结果所作的解释是不同的。

### 练习

### 应用

**4.3(教育)** 假定考试由 20 道判断题组成。学生在参加考试时,对每个问题猜测其答案。学生正确回答 15 道题或 15 道题以上的概率是多少? [提示:使用模拟方法。产生多组(2 000 组或更多)一位数,每组 20 个。每个一位数代表考试中一个问题的答案,其中偶数代表正确的回答,奇数代表错误的回答。确定这些组中有 15 题或 15 题以上答对的组的频率。]

**4.4(医药)** 4.1 节中的例子考察了一项筛选检查的可靠性问题。假定当阳性结果的概率是 0.75 时,我们想要模拟在一组 20 个检查结果中,观察到至少 15 个阳性结果、5 个阴性结果的概率。使用随机数发生器去模拟进行 20 个筛选检查。

- 让一个 2 位数表示对一个人进行筛选检查的结果。哪个数字代表检查结果是阳性的? 哪个数字代表检查的结果是阴性的?
- 如果我们产生 2 000 个由 20 个 2 位数构成的集合,怎样用模拟的结果去近似求出在 20 个检查中,出现至少 15 个阳性结果的概率?

## 4.3 基本的事件关系和概率法则

一个事件,比如说事件  $A$  的概率总满足性质

$$0 \leq P(A) \leq 1,$$

即一个事件的概率在 0(该事件的发生是不可能的)到 1(该事件的发生是必然的)的范围内。

假定  $A$  和  $B$  代表试验中的两个事件,而你关心一个新的事件,即由  **$A$  或  $B$**  发生构成的事件。例如,假定我们掷一双骰子并且定义下列事件:

$A$ :总点数是 7

$B$ : 总点数是 11

则事件“ $A$  或  $B$  发生”是指这样的事件: 你掷一双骰子, 所得总点数是 7 或 11。

注意, 在这个例子中, 事件  $A$  和  $B$  是相互排斥的; 即如果你观察到事件  $A$  (总数是 7), 你就不能同时观察到事件  $B$  (总数是 11)。这就是说, 如果  $A$  发生, 则  $B$  不能发生 (并且反过来也是如此)。

**定义 4.1** 在进行一次试验时, 如果一个事件  $A$  的发生排除了另一个事件  $B$  发生的可能, 称事件  $A$  和  $B$  是相互排斥的。

事件的概率必须满足的第二条性质是用事件相互排斥的概念来给出的。当两个事件相互排斥时, 则“两事件之一发生”这个事件的概率是两个事件概率的和。

**定义 4.2** 如果两个事件  $A$  和  $B$  相互排斥, 则事件“ $A$  或  $B$  发生”的概率是

$$P(A \text{ 或 } B) = P(A) + P(B).$$

后面我们将定义两个事件的并, 定义 4.2 是两个事件并的一个特例。

相互排斥事件概率的可加性的定义可推广到两个以上的事件。例如, 当我们掷一双骰子时, 骰子出现的点数的和设为  $S$ ,  $S$  可取 2, 3, 4, ..., 11, 12 中任一值。在投掷一次骰子时, 我们仅能观察到其中一值。因此, 数值 2, 3, ..., 12 代表了相互排斥的事件。如果我们想要得出投掷骰子得到的总点数小于或等于 4 的概率, 这个概率是

$$P(S \leq 4) = P(2) + P(3) + P(4)$$

对于这个试验, 骰子的落下有 36 种不同的方式, 这些方式是等可能的。当我们观察到第一个骰子 1 朝上, 第二个骰子也是 1 朝上时, 用符号 (1, 1) 表示。当我们观察到第一个骰子 1 朝上, 第二个骰子是 2 朝上时, 用符号 (1, 2) 表示。换句话说, 对这个试验, 可能的结果是

$$\begin{array}{cccccc} (1, 1) & (2, 1) & (3, 1) & (4, 1) & (5, 1) & (6, 1) \\ (1, 2) & (2, 2) & (3, 2) & (4, 2) & (5, 2) & (6, 2) \\ (1, 3) & (2, 3) & (3, 3) & (4, 3) & (5, 3) & (6, 3) \\ (1, 4) & (2, 4) & (3, 4) & (4, 4) & (5, 4) & (6, 4) \\ (1, 5) & (2, 5) & (3, 5) & (4, 5) & (5, 5) & (6, 5) \\ (1, 6) & (2, 6) & (3, 6) & (4, 6) & (5, 6) & (6, 6) \end{array}$$

由此可见, 其中仅有一个事件 (1, 1) 使得总点数为 2。因此, 在大量的重复试验中, 我们期望总点数为 2 发生的概率为频率  $1/36$ , 从而我们令  $P(2) = 1/36$ 。如果我们观察到结果 (1, 2) 或 (2, 1), 则总点数为  $S = 3$  的事件发生。因此,  $P(3) = 2/36 = 1/18$ 。同样地, 我们求出  $P(4) = 3/36 = 1/12$ 。由此得到

$$P(S \leq 4) = P(2) + P(3) + P(4) = \frac{1}{36} + \frac{1}{18} + \frac{1}{12} = \frac{1}{6}$$

事件概率的第三条性质关系到事件和它的补。

**定义 4.3** 事件  $A$  的补定义为“事件  $A$  不发生”这样一个事件。 $A$  的补用符号  $\bar{A}$  表示。

这样,如果我们定义事件  $A$  的补作为一个新的事件,也就是“ $A$  不发生”,则得出以下结论

$$P(A) + P(\bar{A}) = 1.$$

作为一个例子,再次回到掷两枚硬币的试验。如果,在多次重复试验中,你观察到事件  $A$ :“2 枚硬币正面朝上”发生次数的比例是  $1/4$ ,那么这就意味着你观察到  $\bar{A}$  事件,即“2 枚硬币正面不都朝上”发生次数的比例是  $3/4$ 。这样, $P(A)$  和  $P(\bar{A})$  的和总是为 1。

我们把事件的概率必须满足的三条性质总结如下:

#### 概率的性质

如果在一个试验中, $A$  和  $B$  是两个相互排斥的事件,那么  $P(A)$  和  $P(B)$  必须满足下列性质:

1.  $0 \leq P(A) \leq 1$  和  $0 \leq P(B) \leq 1$
2.  $P(A \text{ 或 } B \text{ 发生}) = P(A) + P(B)$
3.  $P(A) + P(\bar{A}) = 1$  和  $P(B) + P(\bar{B}) = 1$

现在我们可以定义另外两个事件之间的关系:两个事件的**并**和**交**。

**定义 4.4** 两个事件  $A$  和  $B$  的**并**是包含在  $A$  或  $B$  (或两者)中的所有基本事件的集合,表示为  $A \cup B$ 。

**定义 4.5** 两个事件  $A$  和  $B$  的**交**是既包含在  $A$  也包含在  $B$  中的所有基本事件的集合,表示为  $A \cap B$ 。

这些定义连同事件的补的定义一起,使一些简单的概念公式化。当事件  $A$  不发生时,事件  $\bar{A}$  发生;当  $A$  或  $B$  发生时, $A \cup B$  发生;当  $A$  和  $B$  同时发生时, $A \cap B$  发生。

相互排斥事件的概率可加性,叫做**相互排斥事件的加法法则**,这个法则可以推广到一般的情形。

**定义 4.6** 考虑事件  $A$  和  $B$ 。 $A$  和  $B$  **并**的概率是

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

#### 例 4.1

事件和事件的概率在图 4.1 中用 Venn (维恩)图表示出来。使用这些图确定下列概率:

- a.  $P(A), P(\bar{A})$
- b.  $P(B), P(\bar{B})$

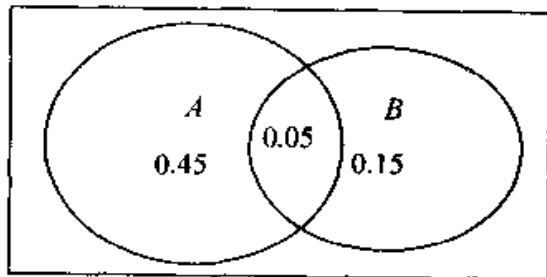


图 4.1

c.  $P(A \cap B)$

d.  $P(A \cup B)$

**解答** 从 Venn 图, 我们能得到下列概率:

a.  $P(A) = 0.5$ , 因此  $P(\bar{A}) = 1 - 0.5 = 0.5$

b.  $P(B) = 0.2$ , 因此  $P(\bar{B}) = 1 - 0.2 = 0.8$

c.  $P(A \cap B) = 0.05$

d.  $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.2 - 0.05 = 0.65$

#### 4.4 条件概率和独立性

考虑下列情况: 在对很多保险索赔的分析中, 根据保险的类型以及索赔是否属于欺诈对索赔进行分类, 得到的结果见表 4.2。假定你负责审核保险索赔——具体地说, 是要识别出欺诈索赔——并且正在处理一桩索赔, 那么, 事件  $F$  既“该桩索赔为欺诈索赔”的概率是多少? 为了回答这个问题, 你考察表 4.2, 并且注意到在所有的索赔中有 10% 是欺诈索赔。于是, 假定在表中给出的各个百分比与收到特定类型的索赔的真实概率充分接近, 就得出  $P(F) = 0.10$ 。你会说你面对个欺诈索赔风险的概率有 0.10 吗? 我们想不会, 因为你有一些可以影响估计  $P(F)$  的附加信息。这些附加信息与你正在核审的保险单的类型(火灾, 汽车, 或其他)有关。

表 4.2 保险索赔的分类

类型	保险单的类型 (%)			总和 %
	火灾	汽车	其他	
欺诈索赔	6	1	3	10
非欺诈索赔	14	29	47	90
总 和	20	30	50	100

假定你的附加信息是这桩索赔与一张火灾保险单有关。检查表 4.2, 我们看到所有的索赔中有 20% (或 0.20) 与火灾保险单有关, 有 6% (或 0.06) 是欺诈性火灾保险索赔。因此, 可以得到在已知是火灾保险单的情况下, 该桩索赔是欺诈索赔的概率为

$$\begin{aligned}
 P(F | \text{火险单}) &= \text{欺诈火险单索赔的比例} / \text{火险单索赔的比例} \\
 &= \frac{0.06}{0.20} = 0.30
 \end{aligned}$$

概率  $P(F|\text{火险单})$  称为事件  $F$  的**条件概率**——即在给定“火险单”这个事件已经发生的条件下, 事件  $F$  的概率。这告诉你所有的火险单索赔中 30% 是欺诈性的。在表达式  $P(F|\text{火险单})$  中, 垂直的线代表短语“给定……”, 或简化为“给定”。这样, 该表达式读作“给定事件‘火险单’时, 事件  $F$  的概率。”

$\%P(F)=0.10$  称为事件  $F$  的**无条件概率**或**边缘概率**, 它给出了索赔是欺诈性索赔的次数的比例——即在非常多(无穷地多)的重复试验(这里, 试验指收到一项保险索赔并且确定索赔是否是欺诈性索赔)中事件  $F$  发生次数的比例。与此相对, 在给定索赔为火险单索赔时,  $F$  的条件概率  $P(F|\text{火险单})$  给出的是火险单索赔是欺诈性索赔的比例。很显然, 在给定保险单的类型条件下,  $F$  的条件概率在测定欺诈的风险时将比  $F$  的无条件概率有更大的帮助。

**定义 4.7** 考虑两个有非零概率  $P(A)$  和  $P(B)$  的事件  $A$  和  $B$ , 给定事件  $B$  时事件  $A$  的**条件概率**是

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

给事件  $A$  时事件  $B$  的条件概率是

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

利用条件概率的定义, 我们可以给出所谓乘法法则。

**定义 4.8** 两个事件  $A$  和  $B$  的**交的概率**是

$$\begin{aligned} P(A \cap B) &= P(A)P(B|A) \\ &= P(B)P(A|B) \end{aligned}$$

定义 4.7 和定义 4.8 中都包含条件概率, 它们之间的惟一差别在于什么概率是已知的, 什么概率需要计算。当交的概率  $P(A \cap B)$  和事件  $A$  的概率  $P(A)$  已知时, 我们可以求出  $P(B|A)$ 。当我们已知  $P(A)$  和  $P(B|A)$  时, 可以求出  $P(A \cap B)$ 。

#### 例 4.2

挑选 2 名管理员作为公司内的安全代表。已知 6 名管理员在研究部门, 4 名管理员在开发部门, 并且在挑选时, 每个组中的管理员被选中的机会相同。求从研究部门选择 2 个管理员的概率。

**解答** 设  $A$  是第 1 个管理员从研究部门选出的事件, 设  $B$  是第 2 个管理员也从研究部门选出的事件。显然, 我们要求出  $P(A \text{ 并且 } B) = P(A \cap B) = P(B|A)P(A)$ 。

对这个例子,

$$P(A) = \frac{\text{研究部门管理员数}}{\text{管理员数}} = \frac{6}{10}$$

并且

$$\begin{aligned} P(B|A) &= \frac{\text{选出一个研究部门管理员后的研究管理员数}}{\text{选出一个研究部门管理员后的管理员数}} \\ &= \frac{5}{9} \end{aligned}$$

于是

$$P(A \cap B) = P(A)P(B|A) = \frac{6}{10} \left( \frac{5}{9} \right) = \frac{30}{90} = 0.333$$

因此,若假定两组中的人被选中的机会相同,则从研究部门选出 2 个管理员的概率是 0.333。

假定事件  $A$  的概率与事件  $B$  是否发生无关,即设

$$P(A|B) = P(A)$$

则我们说事件  $A$  的发生不依赖于事件  $B$  的发生,或简单地说事件  $A$  和  $B$  是**相互独立的事件**。当  $P(A|B) \neq P(A)$  时,事件  $A$  的发生依赖于事件  $B$  的发生,此时说事件  $A$  和  $B$  是**相关的事件**。

**定义 4.9** 两事件  $A$  和  $B$  是**相互独立的事件**,如果

$$P(A|B) = P(A) \text{ 或 } P(B|A) = P(B)$$

(注意:你可以证明,如果  $P(A|B) = P(A)$ ,则  $P(B|A) = P(B)$ ,反之亦然。)

从定义 4.9 可以导出  $P(A \cap B)$  的一个特殊情况。当事件  $A$  和  $B$  相互独立时,有

$$P(A \cap B) = P(A)P(B|A) = P(A)P(B)$$

相互独立的概念在抽样中有特别的重要性。在本书后面的章节中,我们将讨论从两个(或两个以上)总体中抽取样本以比较总体的均值,方差,或者其他的总体参数。对于这些应用中的大多数,在抽取样本时,我们将使在一个样本中的观察值与另一个样本中的值独立。我们把这样的样本叫做**独立样本**。

练习

**基本技能**

**4.5** 把一枚硬币掷 3 次。在表格中列出所有可能的基本事件(第一次投掷的结果,第二次投掷的结果,第三次投掷的结果)。

**4.6** 在练习 4.5 中,假设每一个基本事件发生的概率是  $1/8$ ,求下列概率:

- a.  $A$ : 观察到恰好有 1 次正面朝上;
- b.  $B$ : 观察到有 1 次或 1 次以上正面朝上;
- c.  $C$ : 观察到正面均不朝上。

**4.7** 对练习 4.6:

- a. 计算事件  $A, B, C$  的补的概率;  
 b. 判断事件  $A$  和  $B$  是否是相互排斥的。

4.8 对练习 4.6 中的事件求下列条件概率。

- a.  $P(A|B)$     b.  $P(A|C)$     c.  $P(B|C)$

4.9 参照练习 4.8。事件  $A$  和  $B$  相互独立吗? 解释其原因。对事件  $A$  和  $C$ , 事件  $B$  和  $C$  又如何?

4.10 滚动一枚骰子观察其停止时朝上的数字。求下列事件的概率:

- a.  $A$ : 观察到的是一个 6;  
 b.  $B$ : 观察到的是一个偶数;  
 c.  $C$ : 观察到的是一个大于 2 的数;  
 d.  $D$ : 观察到的是一个比 2 大的偶数。

4.11 参照练习 4.10。事件  $A, B$  和  $C$  中哪些是相互独立的? 哪些是互相排斥的?

4.12 考虑一个试验中的下列基本事件:

基本事件	1	2	3	4	5
概率	0.20	0.25	0.15	0.10	0.30

令事件  $A$  由基本事件 1, 3 和 5 组成, 且事件  $B$  由基本事件 4 和 5 组成。

- a. 求  $P(A)$  和  $P(B)$ 。  
 b. 求  $P(\text{事件 } A \text{ 和 } B \text{ 都发生})$ 。  
 c. 求  $P(\text{事件 } A \text{ 或 } B \text{ 发生})$ 。

4.13 参照练习 4.12。  $P(\text{事件 } A \text{ 或 } B \text{ 发生}) = P(A) + P(B)$  成立吗? 为什么?

## 应用

4.14(教育) 一个学生下学期必须修会计课和经济课。假设时间表没有冲突, 如果有 4 个会计课课堂和 3 个经济课课堂可选, 描述他选择一个会计课课堂和一个经济课课堂的可能的基本事件。

4.15(工程) 一所医院的急诊室有 2 台备用发电机, 其中任何一台能为医院的基本运转供应足够的电。我们定义事件  $A$  和  $B$  如下:

事件  $A$ : 1 号发电机工作正常

事件  $B$ : 2 号发电机工作正常

用语言描述下列事件:



a. 事件  $A$  的补;      b.  $B|A$ ;      c. 事件  $A$  或  $B$  发生。

**4.16(人力资源)** 通过对很多大公司的调查,得到有关对升职或调动工作的意见的事件的概率如下表所示。

升职/调动	已婚		未婚	总和
	双方有职业	一方有职业		
拒绝	0.184	0.0555	0.0170	0.2565
接受	0.276	0.3145	0.1530	0.7435
总和	0.46	0.37	0.17	

利用上述概率回答下列问题:

a. 一个(随机选择的)专职人员接受升职的概率是多少? 拒绝升职的概率是多少?

b. 一个(随机选择的)专职人员为双方均有职业的夫妇中的一方的概率是多少? 夫妇双方仅这一个人有职业的概率呢?

**4.17(商业)** 一个机构投资者考虑对 5 个公司中的 2 个公司进行一项大的投资。假设投资者不知道 5 个公司中的 2 个公司关于新产品的开发的基础不稳定。

a. 列出所有可能的基本事件。

b. 确定从 3 个基础更好的公司中选出 2 个公司的概率。

c. 所选公司中包含 1 个基础不稳定的公司的概率是多少?

d. 选出 2 个基础最不稳定公司的概率是多少?

**4.18(社会)** 对一个公司两个车间的工人进行调查,调查包括下列问题:响应于工人的合理投诉的管理工作有效吗? 结果显示如下。

	已调查数字	回答是“差”的数字
车间 1	192	48
车间 2	248	80

设事件  $A$  为“工人来自于 1 号车间”这个事件,事件  $B$  为“回答为‘差’”的事件。计算  $P(A)$ ,  $P(B)$  和  $P(A \cap B)$ 。

**4.19** 参考练习 4.16。

a. 事件  $A$  和  $B$  相互独立吗?

b. 求  $P(B|A)$  和  $P(B|\bar{A})$ 。它们相等吗？

**4.20** 一家大公司为了评估其雇员在日常工作中的表现,花了相当多的时间开发出一套雇员表现等级的评估办法。这样,可以把应当被安排在重点岗位上的人确定下来,并在需要时进行重大调整。确定重点岗位人员的关键是体现雇员能力的指标,即可以负荷的工作量,以及雇员所接受的正规工作训练。

工作量	正规训练			
	无	很少	一定程度	全面
低	0.01	0.02	0.02	0.04
中等	0.05	0.06	0.07	0.10
高	0.10	0.15	0.16	0.22

由负荷的工作量以及所接受的正规工作训练把所有雇员分成 12 个类。各个类中的雇员被安排在重要岗位上的概率如表所示。下面定义了 3 个事件( $A, B$  和  $C$ ):

$A$ : 一个雇员负荷的工作量是属于高的;

$B$ : 一个雇员具有最高的(全面)正规训练水平;

$C$ : 一个雇员很少或没有正规训练并且工作量为中低档。

a. 求  $P(A), P(B)$ , 和  $P(C)$ 。

b. 求  $P(A|B), P(\bar{A}|B)$  和  $P(B|C)$ 。

c. 求  $P(A \cup B), P(A \cap C)$  和  $P(B \cap C)$ 。

**4.21(商业)** 某个大城市的公用事业公司发现其 70% 的顾客付清每月的账单。

a. 假定从所有顾客的列表中随机选择 2 名顾客。两个顾客都付清每月账单的概率是多少？

b. 至少一个顾客付清每月账单的概率是多少？

**4.22** 参照练习 4.21。对公司记录更详细的检查表明,付清当月账单的顾客中,95% 的顾客也会付清下一个月的账单。没有付清当月账单的顾客中,仅有 10% 的顾客会付清下一个月的账单。

a. 求随机选出的一位顾客在连续的两个月中都付清账单的概率。

b. 求随机选出的一位顾客在连续的两个月中都不付清账单的概率。

c. 求随机选出的一位顾客在连续的两个月中只付清一个月账单的概率。

## 4.5 Bayes 公式

在本节中,我们将说明如何用样本数据,通过 Bayes(贝叶斯)公式来更新条件概率。这些“更新”了的条件概率可应用在决策过程中。这些方法的一个特殊应用是诊断检查的评估。假定一名肉类的检查员必须决定一个随机选择的肉的样本中是否含有大肠杆菌(E.coli)。检查员进行了一项诊断检查。理论上,阳性结果(Pos)意味着肉样本中确实含有大肠杆菌,而阴性结果(Neg)意味着肉样本中不含大肠杆菌。然而,诊断检查偶尔有误差。检查的结果可能是假阳性,即检查结果错误地显示含有大肠杆菌,或假阴性,即检查结果错误地显示不含有大肠杆菌。为了评估这种诊断检查的精确性,进行了大量的筛选检验。例如,在 10 000 个肉的样本中放入大肠杆菌(E),诊断检查的结果是 9 500 个样本呈阳性和 500 个样本呈阴性;也就是说,10 000 个检查结果中有 500 个是假阴性。另有 10 000 个样本都清除了大肠杆菌(NE),并且诊断检查结果 100 个样本为阳性,9 900 个样本为阴性;即在 10 000 个检查结果中有 100 个假阳性。我们在下表中把这些结果归纳出来:

诊断检查 结 果	肉样本状况	
	E	NE
阳性	9 500	100
阴性	500	9 900
总和	10 000	10 000

检查结果的评估如下:

$$\text{真实的阳性率} = P(\text{Pos}|\text{E}) = \frac{9500}{10000} = 0.95$$

$$\text{假的阳性率} = P(\text{Pos}|\text{NE}) = \frac{100}{10000} = 0.01$$

$$\text{真实的阴性率} = P(\text{Neg}|\text{NE}) = \frac{9900}{10000} = 0.99$$

$$\text{假的阴性率} = P(\text{Neg}|\text{E}) = \frac{500}{10000} = 0.05$$

诊断检验的**灵敏度**是真阳性率,而诊断检验的**精确度**是真阴性率。

检查员面临的主要问题是当检查结果为阳性时,判断肉样本中出现大肠杆菌的概率,即检查员需要知道  $P(\text{E}|\text{Pos})$ 。如下列计算所示, Bayes 公式回答了这个问题。计算时,我们需要知道被检查的这类肉中含有大肠杆菌的比率。在此例中,假定在所有肉的样本中有 4.5% 的样本含大肠杆菌,即大肠杆菌的感染率  $P(\text{E}) =$

0.045。我们可以如下计算  $P(E|Pos)$ ：

$$\begin{aligned}
 P(E|Pos) &= \frac{P(E \cap Pos)}{P(Pos)} = \frac{P(E \cap Pos)}{P(E \cap Pos) + P(NE \cap Pos)} \\
 &= \frac{P(Pos|E)P(E)}{P(Pos|E)P(E) + P(Pos|NE)P(NE)} \\
 &= \frac{(0.95)(0.045)}{(0.95)(0.045) + (0.01)(1 - 0.045)} \\
 &= 0.817
 \end{aligned}$$

这样，在检查结果为阳性的被检样本中，有 81.7% 是真正含大肠杆菌的。另外，我们也可以得出，当在肉的样本并没有大肠杆菌时，被检样本的 18.3% 显示了含有大肠杆菌。

### 例 4.3

图书俱乐部把其成员分为三组：大量购买者、中等和少量购买者，给俱乐部成员邮寄时，对各组分开进行。总体而言，20% 的成员是大量购买者，30% 为中等，50% 为少量购买者。成员加入俱乐部 18 个月以后，才给他分组，但在此之前要做一个检验，以考察用其前 3 个月的购买量来给他分组的可行性。下面是从已经被分组后的大量购买者、中等和少量购买的会员的现存记录中得到的百分比。

前 3 个月 购买量	组 (%)		
	大量	中等	少量
0	5	15	60
1	10	30	20
2	30	40	15
3+	55	15	5

如果一位成员在前 3 个月没有购买书，该成员是少量购买者的概率是多少？

(注意：这个表的每一列的内容是“条件”百分比)

**解答** 用表中的条件概率、各个基础订购量组的概率和 Bayes 公式，我们可以计算出所求的条件概率为

$$\begin{aligned}
 P(\text{少量} | 0) &= \frac{P(0 | \text{少量})P(\text{少量})}{P(0 | \text{少量})P(\text{少量}) + P(0 | \text{中等})P(\text{中等}) + P(0 | \text{大量})P(\text{大量})} \\
 &= \frac{(0.60)(0.50)}{(0.60)(0.50) + (0.15)(0.30) + (0.05)(0.20)}
 \end{aligned}$$

$$= 0.845$$

上述例子说明了 Bayes 公式的基本思想。有  $k$  个可能的、互相排斥的基础事件  $A_1, \dots, A_k$ , 有时这些事件称为**自然的状态**; 已知无条件的概率  $P(A_1), \dots, P(A_k)$ , 通常叫做**先验概率**; 有  $m$  个可能的、互相排斥的可观测事件  $B_1, \dots, B_m$ ; 给定每个自然状态下的每个可观测事件的条件概率  $P(B_i | A_j)$  也是已知的, 这些概率称为**似然**。问题是求**后验概率**  $P(A_i | B_j)$ 。先验概率和后验概率是指观察到事件  $B_j$  之前和以后的概率。

(作者这里所说的基础事件, 蕴涵着  $A_1, \dots, A_k$  满足条件“ $\bigcup_{i=1}^k A_i$  为必然发生的事件”——译者注)

**贝叶斯公式** 如果  $A_1, \dots, A_k$  是互相排斥的自然状态, 并且如果  $B_1, \dots, B_m$  是  $m$  种可能的互相排斥的可观测事件, 则

$$\begin{aligned} P(A_i | B_j) &= \frac{P(B_j | A_i)P(A_i)}{P(B_j | A_1)P(A_1) + P(B_j | A_2)P(A_2) + \dots + P(B_j | A_k)P(A_k)} \\ &= \frac{P(B_j | A_i)P(A_i)}{\sum_i P(B_j | A_i)P(A_i)} \end{aligned}$$

#### 例 4.4

在电路板的制造中, 不合格板主要有三种类型。这三种类型, 以及有这三种缺陷的电路板在所有电路板中占的百分比分别是 (1) 不恰当的电焊范围 ( $D_1$ ), 2.8%; (2) 电镀脱层 ( $D_2$ ), 1.2%; (3) 蚀刻问题 ( $D_3$ ), 3.2%。一个电路板至多包含三种缺陷之一。用破坏性试验肯定可以检测出成品电路板中的缺陷, 然而, 要检测电路板成品中的一个较大的百分比, 这不是一个很实际的方法。开发了一种非破坏性的检测过程, 其检测结果如下:  $A_1$ , 电路板仅有缺陷  $D_1$ ;  $A_2$ , 电路板仅有缺陷  $D_2$ ;  $A_3$ , 电路板仅有缺陷  $D_3$ ;  $A_4$ , 电路板没有缺陷。通过对已知有三种类型的缺陷之一的大量电路板的评估, 分别确定出了非破坏性测试的 4 种结果的似然, 并由下表给出:

测试结果	缺陷类型			
	$D_1$	$D_2$	$D_3$	无
$A_1$	0.90	0.06	0.02	0.02
$A_2$	0.05	0.80	0.06	0.01
$A_3$	0.03	0.05	0.82	0.02
$A_4$ (无缺陷)	0.02	0.09	0.10	0.95

如果使用非破坏性测试方法对电路板测试, 结果显示无缺陷( $A_4$ ), 那么电路板没有缺陷或缺陷类型为  $D_1, D_2$ , 或  $D_3$  的概率是多少?

设  $D_4$  代表电路板没有缺陷的状况。

$$P(D_1 | A_4)$$

$$\begin{aligned} &= \frac{P(A_4 | D_1)P(D_1)}{P(A_4 | D_1)P(D_1) + P(A_4 | D_2)P(D_2) + P(A_4 | D_3)P(D_3) + P(A_4 | D_4)P(D_4)} \\ &= \frac{(0.02)(0.028)}{(0.02)(0.028) + (0.09)(0.012) + (0.10)(0.032) + (0.95)(0.928)} \\ &= \frac{0.00056}{0.88644} = 0.00063 \end{aligned}$$

$$P(D_2 | A_4)$$

$$\begin{aligned} &= \frac{P(A_4 | D_2)P(D_2)}{P(A_4 | D_1)P(D_1) + P(A_4 | D_2)P(D_2) + P(A_4 | D_3)P(D_3) + P(A_4 | D_4)P(D_4)} \\ &= \frac{(0.09)(0.012)}{(0.02)(0.028) + (0.09)(0.012) + (0.10)(0.032) + (0.95)(0.928)} \\ &= \frac{0.00108}{0.88644} = 0.00122 \end{aligned}$$

$$P(D_3 | A_4)$$

$$\begin{aligned} &= \frac{P(A_4 | D_3)P(D_3)}{P(A_4 | D_1)P(D_1) + P(A_4 | D_2)P(D_2) + P(A_4 | D_3)P(D_3) + P(A_4 | D_4)P(D_4)} \\ &= \frac{(0.10)(0.032)}{(0.02)(0.028) + (0.09)(0.012) + (0.10)(0.032) + (0.95)(0.928)} \\ &= \frac{0.0032}{0.88644} = 0.0036 \end{aligned}$$

$$P(D_4 | A_4)$$

$$\begin{aligned} &= \frac{P(A_4 | D_4)P(D_4)}{P(A_4 | D_1)P(D_1) + P(A_4 | D_2)P(D_2) + P(A_4 | D_3)P(D_3) + P(A_4 | D_4)P(D_4)} \\ &= \frac{(0.95)(0.928)}{(0.02)(0.028) + (0.09)(0.012) + (0.10)(0.032) + (0.95)(0.928)} \\ &= \frac{0.8816}{0.88644} = 0.9945 \end{aligned}$$

因此, 如果新的测试表明, 电路板没有这三种类型的缺陷, 那么电路板事实上无缺陷的概率是很高的: 0.9945。在练习 4.25 中, 我们将请你评定非破坏性测试方法对三种类型缺陷测试的敏感性。

## 练习

## 应用

**4.23(商业)** 在一个金融公司的借贷中,有 1% 是违约的(没有完全偿还)。公司通常对所有的贷款申请作信用检查。发现违约贷款中有 30% 是高风险的,40% 是一般风险的,30% 是低风险的。不违约贷款中有 10% 是高风险的,40% 是一般风险的,50% 是低风险的。使用 Bayes 公式计算一项高风险贷款会成为违约贷款的概率。

**4.24** 参照练习 4.23。证明,在已知贷款为一般风险贷款的条件下,违约的后验概率等于违约的先验概率。解释这为什么是合理的结果。

**4.25** 在例 4.4 中,我们描述了一个确定电路板缺陷的新测试方法。计算这种测试方法准确认定是缺陷  $D_1$ ,  $D_2$ , 和  $D_3$  的概率,即计算  $P(D_1|A_1)$ ,  $P(D_2|A_2)$  和  $P(D_3|A_3)$ 。

**4.26** 在例 4.4 中,计算新测试方法不能正确识别出缺陷  $D_1$ ,  $D_2$  和  $D_3$  的概率,即计算  $P(D_1|A_1)$ ,  $P(D_2|A_2)$  和  $P(D_3|A_3)$ 。

**4.27(商业)** 一个从事家庭保险的保险商研究由烧木头的炉子引起的家庭火灾问题。在所有有这类炉子的家庭中,30% 有第一类炉子,25% 有第二类炉子,15% 有第三类炉子,30% 有其他类炉子。在以后的 3 年中,第一类炉子的 5%,第二类炉子的 3%,第三类炉子的 2%,和其他类炉子的 4% 导致了火灾。如果有一个特定的家庭发生了火灾,这个家庭用的是第一类炉子的概率多少?

**4.28(医药)** 1998 年 1 月 15 日, *New England Journal of Medicine* 杂志刊登的一篇文章中,报导了利用计算机 X 线断层摄影术(CT)作为临床上诊断怀疑有阑尾炎的患者检查的效用。在有阑尾炎的患者中,至少有 20% 没有正确诊断出来。另一方面,检查时显示阑尾正常的患者中,15% 到 40% 是经历过急性阑尾切除手术的。为了改进对患者的治疗,设计出一项研究来确定使用 CT 作为诊断检查的预期效果。研究中相继检查了 100 个怀疑有急性阑尾炎的患者,这些患者或者被送到急诊部的,或者是从某个诊所转院来的。为这 100 个患者做了 CT 扫描,并请外科医生为每个患者是否有阑尾炎作出了诊断。最终的临床结果是在外科通过阑尾切除手术后的阑尾病理检查或在 CT 扫描后临床跟踪至少 2 个月得到的。

扫描的结果	有无阑尾炎	
	已证实(C)	已排除(RO)
一定有阑尾炎(DA)	0.943	0.023
怀疑有阑尾炎(EA)	0.038	0.045
一定无阑尾炎(DNA)	0.019	0.936

在 1996 年阑尾炎的发病率近似于  $P(C) = 0.00108$ 。

- a. 求阑尾炎 CT 诊断的灵敏度和精确度。
- b. 求一个阑尾炎患者, 用 CT 诊断的结果为“一定有阑尾炎(DA)”的概率。
- c. 求一个未患阑尾炎的患者, 用 CT 诊断的结果为“一定有阑尾炎(DA)”的概率。
- d. 求一个未患阑尾炎的患者, 用 CT 诊断的结果为“一定无阑尾炎(DNA)”的概率。

**4.29(医药)** 条件概率能应用于诊断疾病。假设有三种密切相关的不同疾病 ( $A_1, A_2$  和  $A_3$ ), 在人群中分别有 25%, 15% 和 12% 的发生率。进一步, 假设三种可能与这些疾病相联系的症状 ( $B_1, B_2$  和  $B_3$ ) 互相排斥。经验表明当患有疾病时, 出现给定症状的可能性  $P(B_j | A_i)$  如下表所示。分别求给定症状是  $B_1, B_2, B_3$  和  $B_4$  时患有疾病  $A_2$  的概率。

症状 $B_j$	疾病 $A_i$		
	$A_1$	$A_2$	$A_3$
$B_1$	0.08	0.17	0.10
$B_2$	0.18	0.12	0.14
$B_3$	0.06	0.07	0.08
$B_4$ (无症状)	0.68	0.64	0.68

## 4.6 离散变量和连续变量

本章中介绍的概率的基本语言用来处理许多不同类型的事件。我们既对定量事件概率的计算感兴趣, 也对定性事件的概率的计算感兴趣。例如, 我们开发了一项技术, 用来决定从一家大的汽车制造厂中随机选择的一名机械制造工人在 8 小时工作期间出现事故的概率。这样的技术也可用来计算一名随机选择的机械制造工人无事故工作时间超过 80 个小时的概率。

对于这些定性和定量的事件, 可以作为与定性和定量变量相联系的事件(或基本事件)而进行分类。例如, 在汽车制造业事故研究中, 对于随机选择的机械制造工人, 其事故报告中包含下列检查结果之一: 无事故, 小事故, 或重大事故。这样, 在研究中 100 名机械制造工人的数据是一个定性变量的观察值, 因为对于这些工人, 可能的响应值之间只是事故类型不同, 而不是测量数值上的区别。因为我们不



能确切地预测对于一名特定的机械制造工人会发生哪种类型的事故,故把这种变量称为**定性随机变量**。用定性随机变量测量的其他常见的例子有:所属党派、社会经济地位,在苹果叶上发现的昆虫的种类,以及顾客偏爱的商标。与定性变量相关的可能的基本事件数是有限的(并且通常相当小)。使用本章的方法,可以计算与这些事件相关的概率。

在很多情况下,试验中感兴趣的事件是与**定量随机变量**相联系的定量基本事件,因为可能的响应值在数值的大小上是不同的。例如,在汽车制造业事故研究中,随机选择的机械制造工人连续无事故的工作日(8小时为一个工作日)这一数字就是一个定量随机变量的观察值。此处,感兴趣的事件就是随机选择的机械制造工人在两次事故之间按8小时一天算得的工作日数,它是一个定量随机变量的观察值。定量随机变量的其他例子有:下一个季度每张股票的收入的变化,在癌症治疗以后病人病情缓解的时间长度,小麦新品种的每亩产量和在即将来临的选举中有义务投票的人数。本章的方法能用来计算任何与特定的事件相联系的概率。

定量随机变量有明显的优点。对于定量变量,其数量尺度使得均值和标准差等有意义。对于定性随机变量,本章中的方法可以用来计算各种不同事件的概率,但基本上仅此而已。而对于定量随机变量,我们能做得更多;我们能求出平均的结果,标准偏差,并且估量随机变量的可能误差,等等。以后,我们用**随机变量**这个术语表示定量随机变量。

感兴趣的大多数事件源于数值型观察值或测量值。如果试验中定量变量的测量值(或观察值)用 $y$ 表示,我们对 $y$ 能取哪些值感兴趣。这些值被称为**数值型试验结果**。在一个经过开垦改造以后的露天煤矿上,每亩种植的不同的植物种数就是一个数值型试验结果。在某选举中,注册选民中参加投票的百分比也是一个数值型试验结果。因为在给定的试验中 $y$ 的值是一个偶然的或随机的试验结果,定量变量 $y$ 称为**随机变量**。

**定义 4.10** 当一个定量随机变量的观察值只可能取可数个数值时,该变量称为**离散随机变量**。

离散变量的例子有:

1. 每棵某种转基因苹果树上所结苹果的蒲式耳数。
2. 在安装一台新的信号设备以后,某十字路口每月交通事故数的变化。
3. 在中西部的一个主要城市的上次市长选举中“死人”投票的数字。

注意,在这些例子中,每个随机变量能够取到的值的个数都可以计数出来。

**定义 4.11** 当一个定量随机变量的观察值可以取到一个区间上的任何一个数值时,该变量称为**连续随机变量**。此时,该随机变量所能取到的值的个数是不可数的。

例如,在纽约州罗彻斯特地区每天的最高温度,可以取到一个区间上所有那无

限多个数值中的任一个。它可以是 89.6, 89.799 或 89.7611114。典型的连续随机变量有温度、压力、高度、重量和距离。

当我们寻求一个与随机变量的某个特定值有关的概率时,需要区分离散随机变量和连续随机变量。在本章的以后章节中讨论概率分布时,进行这种划分的必要性会看得更清楚些。

## 4.7 离散随机变量的概率分布

如前所述,为了推断从中抽出样本的总体,我们需要求出观察到特殊样本值的概率。为此,我们需要知道与变量  $y$  的每个值有关的概率。按照概率的频率解释,这些概率产生一个理论频率的分布,称之为  $y$  的**概率分布**。概率分布对于离散随机变量和连续随机变量是不同的。对离散随机变量,我们能计算出各个特定值发生的概率。而对连续随机变量,取值落在一个区间的概率是要研究的对象。

**离散随机变量的概率分布**用与  $y$  的每个取值有关的概率  $P(y)$  表示。这种表示可以通过表格,图形,或公式来表达。考虑 4.2 节中投掷两枚硬币的例子。设  $y$  为观察到正面朝上硬币的个数。则  $y$  可取值为 0, 1 或 2。从表 4.1 的数据中,我们可以确定  $y$  的每个值的近似概率,如表 4.3 所示。我们指出,表中列出的频率与理论概率非常接近,用概率的古典解释可以说明,这些理论频率分别为 0.25, 0.50 和 0.25。如果我们投掷了 2 000 000 次硬币而不是 500 次,  $y=0, 1$  和 2 的频率值与理论上的概率将是没有任何区别的。

表 4.3  $y$  的经验抽样结果:2 枚硬币的 500 次投掷中正面朝上的个数

$y$	频数	频率
0	129	0.258
1	242	0.484
2	129	0.258

表 4.4 给出了  $y$  (即投掷 2 枚硬币时正面朝上的个数)的概率分布,并且在图 4.2 中用图形给出了**概率直方图**。

表 4.4 投掷 2 枚硬币时正面朝上的个数的概率分布

$y$	$P(y)$
0	0.25
1	0.50
2	0.25

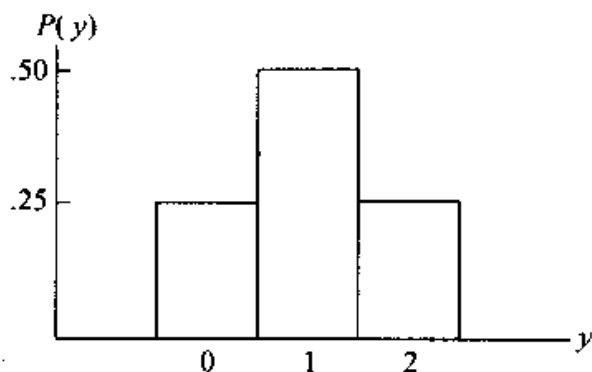


图 4.2 投掷 2 枚硬币时正面朝上的个数的概率分布

这个简单的离散随机变量的概率分布说明了离散随机变量的 3 个重要的性质。

1.  $y$  取每个值的概率在 0 和 1 之间。
2.  $y$  的所有值上的概率的和等于 1。
3. 离散随机变量的概率是可加的, 即“ $y = 1$  或 2”这一事件的概率等于  $P(1) + P(2)$ 。

当讨论二项随机变量的概率分布时, 我们将强调统计推断和概率分布之间的关系。

## 4.8 一个常用的离散随机变量: 二项分布

对生意人和科学家感兴趣的人群可以看成是 0 和 1 组成的一个大集合。例如, 考虑美国所有成年人对如下问题的回答的集合: “你赞成开发利用核能吗?” 如果我们不允许回答“不发表意见”, 那么, 所有回答将组成一个由“是”和“不”构成的集合。如果我们指定 1 表示“是”, 0 表示“不”, 总体将由若干 0 和 1 组成的集合构成, 其中 1 的个数等于赞成开发核能的人的总数, 1 的个数除以美国成年人的总数将等于赞成开发核能的人数所占的比例。

Gallup(盖洛普)和 Harris(哈里斯)民意测验就是从 0 和 1 的总体中抽样的例子。通过对人们进行调查并记录他们的意见, 在样本应答的基础上, Gallup 和 Harris 估计总体中喜欢某种事情或拥有某些特征的人在总体中所占的比例。

类似的调查在生物科学, 工程技术和商业中也有, 但是它们可能被称之为试验而不是民意测验。例如, 为了确定新药的效果, 在对更大的动物直至最后对人类参与者进行试验之前, 往往先对小动物例如老鼠进行试验。在很多这样的试验中, 试验者往往仅仅记录药物是否有效, 因而这些试验与民意调查有着显著的相似性。

比如,如果对 300 只老鼠注射一种药物且 230 个老鼠反应良好,这就像试验者进行了一项“民意测验”——测验老鼠对药物的反应,有 230 个“赞成”和 70 个“反对”。

类似的“民意测验”被大多数制造商用来决定产品中优质产品的比例。在装运以前收集工业产品的样本,由公司的质量控制部门根据既定的标准对样本中的各个样品进行判定,看是否为“次品”或“合格品”。根据样本中次品的个数,公司可以决定此批产品是否适合装运。注意,如前面的那些例子一样,在这个例子中,以基于样本所包含的信息对总体做出推断为实际的目标。

公众民意测验、消费者偏爱调查、药物试验,以及工业中的次品率抽样都是常见的二项试验的例子。二项试验应用于所有科学和商业领域,且各种情况之间的区别仅在于作为样本对象的种类(人,老鼠,电灯,橘子)的不同。这样,定义二项试验的特征,对于我们把这个类型的试验的知识应用于各种不同的抽样试验中将是有益的。

就各个方面的实用目的而言,二项试验等同于前面章节中的投掷硬币的例子。这里,投掷  $n$  枚不同的硬币(或一枚硬币投掷  $n$  次),并且我们对观察到的正面朝上的个数感兴趣。我们假设在试验中抛一次硬币正面朝上的概率是  $\pi$ (当硬币均匀时, $\pi$  等于 0.50,但是在许多实际的情况中, $\pi$  可以取 0 到 1 之间的其他值)。我们同样假设任何一次抛硬币的结果不受前面任何抛硬币结果的影响。这些特征可以总结如下。

**定义 4.12 二项试验** 是一个有下列性质的试验:

1. 整个试验由  $n$  次相同的试验组成。
2. 每次试验的结果是两个基本事件之一。我们把其中一种基本事件记为成功,把另外一种基本事件记为失败。
3. 在一次试验中成功的概率等于  $\pi$ ,并且在不同的试验中  $\pi$  的值不变。
4. 各次试验是互相独立的;就是说,一次试验的结果不影响任何另外试验的结果。
5. 随机变量  $y$  是在  $n$  次试验中观察到的成功的次数。

#### 例 4.5

1998 年 3 月 5 日 *New England Journal of Medicine* 杂志刊登的一篇文章中,讨论了一次肺结核大爆发的问题。一个称为指标病人的人,在 1995 年被诊断有肺结核。对该指标病人的 232 个同事进行了肺结核的筛选检验。在检验中读数为阳性记录的同事的人数是要研究的随机变量。该项研究满足二项试验的性质吗?

**解答** 为了回答这个问题,我们对二项试验的 5 个性质中的每一个性质进行检查,验证其是否得到满足。

1.  $n$  次试验是相同的吗? 是的。这  $n = 232$  个工人与指标病人的关系是近似等同的。
  2. 每次试验的结果是两个基本事件之一吗? 是的。从试验记录中得到的每个同事的读数不是阳性就是阴性。
  3. 不同试验中成功的概率是一样的吗? 是的, 如果这些同事与指标病人有相同的风险因子和暴露程度。
  4. 试验是互相独立的吗? 是的。一次检验的结果不受其他检验结果的影响。
  5. 在 232 次检验中成功的次数  $y$  是试验者感兴趣的随机变量吗? 是的。在检验中读数是阳性的同事的人数是要研究的变量。
- 所有的 5 条性质都满足, 因此肺结核筛选检验是二项试验。

#### 例 4.6

一个经济学家在一个 100 人的班中会见了 75 个学生, 以估计在一门课程中期望获得“C”或更好成绩的学生的比例。这是二项试验吗?

**解答** 验证这个试验不满足二项试验的 5 个性质。

1. 试验是相同的吗? 是的, 会见了 75 个学生中的每一个。
2. 每次试验的结果是两个基本事件之一吗? 是的。每个学生期望或不期望获得“C”或更高的成绩。
3. 在不同的试验中成功的概率相同吗? 不是, 如果我们假设成功表示一个学生期望获得“C”或更高的成绩, 则在不同试验中成功的概率可能改变很多。例如, 在教授不知情的前提下, 假定 100 个学生中有 75 个学生期望获得“C”或更高的成绩。则对第一个学生会见成功的概率  $\pi$  是  $75/100 = 0.75$ 。如果这学生是失败(不期望“C”或更高的成绩), 下一个学生成功的概率是  $75/99 = 0.76$ 。假设在会见 70 个学生以后有 60 个成功和 10 个失败。则下一个学生(第 71 个)成功的概率  $15/30 = 0.50$ 。

这个例子表明, 在样本大小占总体大小一个相当大的比例的情况下, 不同的试验成功的概率会有巨大的改变。这样的试验不满足二项试验的性质。

注意, 很少有现实生活中的情形能完全地满足定义 4.12 中陈述的要求, 但在许多情形, 与这些要求的不一致之处是非常小的, 从而二项试验仍然为现实提供一个很好的模型。

我们已经定义了二项试验, 并提出了几个实际的应用, 现在来考察二项随机变量  $y$ , 即  $n$  次试验中观察到的成功次数的概率分布。尽管可以通过频率的途径来逼近  $P(y)$ , 即二项试验中与  $y$  的一个值相对应的概率, 我们对二项概率可以用一个更容易的公式来计算。

**二项试验中计算  $P(y)$  的公式**

在二项试验中的  $n$  次试验中, 观察  $y$  次成功的概率是

$$P(y) = \frac{n!}{y!(n-y)!} \pi^y (1-\pi)^{n-y}$$

其中

$n$  = 试验次数,

$\pi$  = 单个试验中成功的概率,

$1 - \pi$  = 单个试验中失败的概率,

$y = n$  次试验中成功的次数,

$n! = n(n-1)(n-2)\cdots(3)(2)(1)$ ,

如上面所解释的, 记号  $n!$  (叫做  $n$  的阶乘) 用于表示下列乘积

$$n! = n(n-1)(n-2)\cdots(3)(2)(1)$$

对  $n = 3$

$$n! = 3! = (3)(3-1)(3-2) = (3)(2)(1) = 6$$

类似地, 对  $n = 4$

$$4! = (4)(3)(2)(1) = 24$$

我们也注意到  $0!$  定义为 1。

为了了解二项概率的公式如何应用于计算  $y$  的特殊值的概率, 考虑下列例子。

**例 4.7**

开发出一种在高尔夫球场上使用的新草皮品种, 目的是获得 85% 的发芽率。为了评估这种草皮, 把 20 粒种子种在一个温室里, 并使得每粒种子被暴露在相同的条件下。如果 85% 的发芽率是正确的, 20 粒种子中有 18 粒或 18 粒以上种子发芽的概率是多少?

$$P(y) = \frac{n!}{y!(n-y)!} \pi^y (1-\pi)^{n-y}$$

用  $n = 20$ ,  $\pi = 0.85$ ,  $y = 18, 19$  和  $20$  代入上式, 我们得

$$\begin{aligned} P(y = 18) &= \frac{20!}{18!(20-18)!} (0.85)^{18} (1-0.85)^{20-18} \\ &= 190(0.85)^{18} (0.15)^2 = 0.229 \end{aligned}$$

$$\begin{aligned} P(y = 19) &= \frac{20!}{19!(20-19)!} (0.85)^{19} (1-0.85)^{20-19} \\ &= 20(0.85)^{19} (0.15)^1 = 0.137 \end{aligned}$$

$$P(y = 20) = \frac{20!}{20!(20-20)!} (0.85)^{20} (1-0.85)^{20-20}$$

$$= (0.85)^{20} = 0.0388$$

$$P(y \geq 18) = P(y = 18) + P(y = 19) + P(y = 20) = 0.405$$

在例 4.7 的计算中, 尽管  $n$  仅仅是 20, 我们还是承担了相当大的计算工作量。对于那些  $n$  的取值很大的情况, 可以使用计算机软件做出准确的计算。我们将在后面的章节中给出一个方法, 这种方法在许多情况下, 能够获得相当精确的结果而不需要使用计算机。

#### 例 4.8

为了估计某城市中失业的户主的百分比, 我们从所有家庭中随机选择一个由多个家庭构成的样本。为了举例说明二项概率的计算, 假定未知的百分比实际为 10%, 从总体中选择出  $n = 5$  的一个样本(我们选择一个小样本以便于计算)。所有的 5 个家庭的户主均有工作的概率是多少?

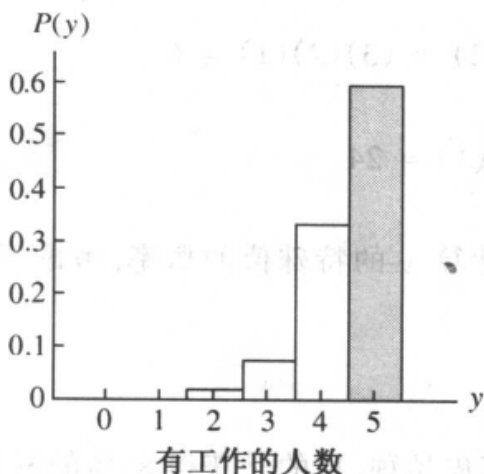


图 4.3  $n = 5, \pi = 0.9$  的二项概率分布

**解答** 在把哪个基本事件定义为成功时, 我们必须谨慎从事。对这个例子,

我们定义有工作为成功。从总体选出一人成功的概率是  $\pi = 0.9$  (因为失业的概率是 0.1)。我们要求出在 5 次试验中  $y = 5$  (所有的 5 个人都有工作) 的概率。

$$\begin{aligned} P(y = 5) &= \frac{5!}{5! (5-5)!} (0.9)^5 (0.1)^0 \\ &= \frac{5!}{5! 0!} (0.9)^5 (0.1)^0 \\ &= (0.9)^5 = 0.590 \end{aligned}$$

把  $n = 5, \pi = 0.9$  的二项概率分布表示在图 4.3 中。注意由 5 个家庭组成的样本中 5 个户主均有工作的概率也在图中给出。

#### 例 4.9

参考例 4.8 并且计算 5 个家庭的样本中有一个人失业的确切的概率。一个人或更少的人失业的概率是多少?

**解答** 因为  $y$  是 5 个户主的样本中有工作的人数, 一个人失业对应于 4 个人有工作 ( $y = 4$ ), 于是

$$P(4) = \frac{5!}{4! (5-4)!} (0.9)^4 (0.1)^1$$

$$\begin{aligned}
 &= \frac{(5)(4)(3)(2)(1)}{(4)(3)(2)(1)} (0.9)^4 (0.1) \\
 &= 5(0.9)^4 (0.1) \\
 &= 0.328
 \end{aligned}$$

因此,在 5 个家庭的样本中有 4 个家庭,其户主有工作的概率是 0.328,或粗略地说有三分之一的机会。

“一个人或更少的人失业”的事件与“4 个人或 5 个人有工作”的基本事件相同。因为  $y$  表示被雇用的人数,我们求  $y = 4$  或 5 的概率。因为与随机变量相关的这些值代表互相排斥的事件,离散随机变量的概率是可加的。于是,我们有

$$\begin{aligned}
 P(y = 4 \text{ 或 } 5) &= P(4) + P(5) \\
 &= 0.328 + 0.590 \\
 &= 0.918
 \end{aligned}$$

这样,在 5 个家庭组成的随机样本中,4 个或 5 个户主有工作的概率是 0.918。这个高概率与我们的直觉是一致的:如果这个城市中,全部家庭的户主 90% 有工作,我们自然可以期望在样本中有工作的人数是一个很大的数值。

与任何频率直方图一样,二项概率分布有其均值  $\mu$  和标准差  $\sigma$ 。尽管我们省略了推导过程,我们给出这些参数的公式如下。

#### 二项概率分布的均值和标准差

$$\mu = n\pi \quad \text{和} \quad \sigma = \sqrt{n\pi(1-\pi)}$$

其中  $\pi$  是在给定试验中成功的概率, $n$  是在二项试验中试验的次数。

如果我们知道  $\pi$  和样本大小  $n$ ,我们就能计算  $\mu$  和  $\sigma$ ,从而找出特定的二项的概率分布的中心,并且描述其变异性。这样,我们就能快速决定哪些  $y$  的取值是可能的,哪些取值是不大可能的。

#### 例 4.10

我们以草皮种子的例子来说明均值和标准差的计算。假定生产草皮的公司为了监管种子的质量,从一般的种子中抽取了 20 粒种子。如果种子的发芽率保持在 85% 不变,那么在 20 粒种子的样本中种子发芽的平均数是

$$\mu = n\pi = 20(0.85) = 17$$

标准差是

$$\sigma = \sqrt{n\pi(1-\pi)} = \sqrt{20(0.85)(1-0.85)} = 1.60$$

假定我们记录了大量由 20 粒种子构成的样本的发芽情况,并对之进行考察。如果发芽率始终保持在 85% 不变,那么每个样本发芽种子的平均数应该接近 17。如果在某个 20 粒种子的特定样本中,仅有 12 粒种子发芽,那么我们的观察结果与 85% 的发芽率一致吗? 使用一个计算机软件程序,我们可以产生 20 粒种子的样本



中发芽种子的棵数的概率分布,如图 4.4 所示。

尽管该分布有向左偏倚的倾向(见图 4.4),由于分布的形状接近于丘形,可以对之使用经验准则。 $y = 12$  粒种子比发芽种子的平均数  $\mu = 17$  减去 3 倍的标准差还少;因而,如果  $\pi$  确实等于 0.85,我们在 20 粒种子中仅观测到 12 粒种子发芽这种事是非常不可能的。很可能发芽率比 0.85 要小很多。

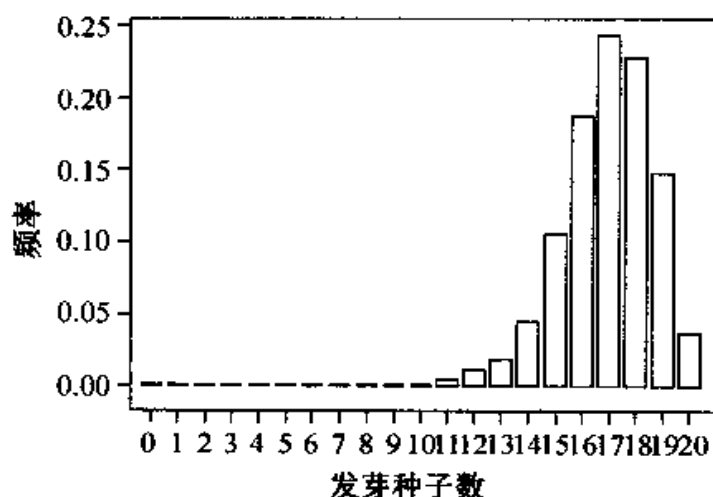


图 4.4  $n = 20$  和  $\pi = 0.85$  的二项分布

#### 例 4.11

一项民意测验表明,在 1 218 位选民中有 516 位赞成某位政党候选人竞选连任。你认为该候选人会赢吗?

**解答** 要赢得选举,候选人将需要至少 50% 的选票。现在我们来查看,  $y = 516$  作为  $y$  的值是否太小,以至于我们不能认为  $\pi$  的值(赞成该候选人的选民的比例)等于 0.5 或更大。如果  $\pi = 0.5$ , 则

$$\begin{aligned}\mu &= n\pi = (1218)(0.5) = 609 \\ \sigma &= \sqrt{n\pi(1-\pi)} = \sqrt{(1218)(0.5)(0.5)} \\ &= \sqrt{304.5} = 17.45\end{aligned}$$

从而  $3\sigma = 52.35$ 。

从图 4.5 读者可以看出,  $y = 516$  与  $\mu = 609$  的距离大于  $3\sigma$ , 或 52.35。事实上,如果读者愿意做进一步的检查,可以发现如果  $\pi$  真等于 0.5 的话,  $y = 516$  与  $\mu = 609$  的距离甚至大于  $5\sigma$ 。这样,如果事实上大多数选民赞成该候选人再次竞选的话,那么在选民的样本中,赞成该候选人的选民的人数显然太少了。因而,我们得出结论,他或她将落选。(注意这个结论基于这样一个假定,即从中抽取样本

的选民的集合与将参加投票的选民的集合是一样的。我们还必须假定选民的意见将在抽样的时间和选举的日期之间不发生变化。)

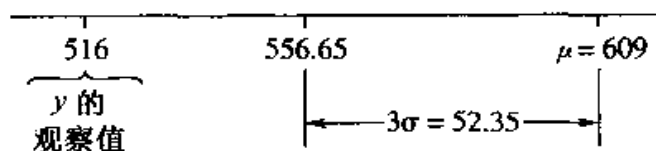


图 4.5  $y$  的观察值( $y = 516$ )相对于  $\mu$  的位置

本节的目的是给出二项概率分布,从而使读者能够了解二项概率的计算方法,并且当  $n$  的值不大时能够计算出二项概率。在实际中,  $n$  的值通常很大(在全国范围的调查中,样本容量大到 1 500 是常见的),此时二项概率的计算是乏味的。在本章的后面,我们将介绍一个简单的方法,利用该方法,可以得到我们在进行推断时所需要的概率的近似值。另有一些粗略的近似方法,通过应用经验准则以及二项随机变量  $y$  的均值和标准差,也可以用来计算二项分布的概率的近似值。

本书中要讨论的惟一的另外一个离散随机变量是 Poisson(泊松分布,第十章)。我们建议感兴趣的读者从 Hildebrand 和 Ott(1998)的书和 Devore(2000)的书中获得关于离散随机变量的更多信息。在下一节,我们将重点讨论正态分布。

### 练习

#### 基本技能

**4.30** 考虑下列课堂试验:抛 3 枚硬币并且观察正面朝上的枚数  $y$ 。让每个学生重复 10 次试验,结合全班的结果,构造一张  $y$  的频率表。注意,这些频率给出了  $y = 0, 1, 2$  或 3 的实际概率的近似值。(注意:用二项公式计算实际的概率  $P(y)$  并与近似值相比较。)

**4.31** 设  $y$  是二项随机变量,计算下列每种情况下的  $P(y)$ :

a.  $n = 10, \pi = 0.2, y = 3$

b.  $n = 4, \pi = 0.4, y = 2$

c.  $n = 16, \pi = 0.7, y = 12$

**4.32** 设  $y$  是二项随机变量且  $n = 8$  和  $\pi = 0.4$ , 求下列各值:

a.  $P(y \leq 4)$       b.  $P(y > 4)$

c.  $P(y \leq 7)$       d.  $P(y > 6)$

### 应用

**4.33(商业)** 一家器材店一天中卖出的主要器材的个数  $y$  有如下分布:

$y$	$P(y)$
0	0.100
1	0.150
2	0.250
3	0.140
4	0.090
5	0.080
6	0.060
7	0.050
8	0.040
9	0.025
10	0.015

a. 构造  $P(y)$  的一张图。

b. 求  $P(y \leq 2)$ 。

c. 求  $P(y \geq 7)$ 。

d. 求  $P(1 \leq y \leq 5)$ 。

**4.34(商业)** 一家计算机商店每周卖出某流行的文字处理程序的拷贝数量的概率分布如下。

$y$	$P(y)$
0	0.06
1	0.14
2	0.16
3	0.14
4	0.12
5	0.10
6	0.08
7	0.07
8	0.06
9	0.04
10	0.03

a. 在一个特定的星期内需要 3 个或更多拷贝的概率是多少？

b. 需要至少 2 个但不超过 6 个拷贝的概率是多少？

c. 如果商店在每个星期的开始有该程序的 8 个拷贝可用, 在指定的一周需求超过供应的概率是多少？

**4.35(生物)** 一位生物学家从当地的水库随机地选择了 10 份水样,每份水样体积为 0.1 立方厘米。生物学家计数了每份水样中的细菌数,然后用这 10 份水样中细菌的数量得出水库中每立方厘米存在细菌个数的估计。这是一个二项试验吗?

**4.36(政治学)** 检查下面的报纸剪辑。这种抽样满足二项试验的特征吗?

民意测验发现人们反对窃听电话

纽约——在近期的一个民意测验中,被调查的人中以 81% 对 13% 反对在没有法院指令的情况下窃听他们的电话。

在调查中人们以 68% 对 27% 反对政府对犯罪嫌疑人进行窃听,除非有法院的指令。

这项测验调查了 1 495 个家庭,同时发现了下列结果:

——调查中人们以 80% 对 12% 反对在没有法院指令的情况下使用任何电子间谍设施。

——市民中以 77% 对 14% 反对在没有法院指令的情况下政府拆看他们的信件。

——他们以 80% 对 12% 反对电话公司公开他们的长途电话记录,除非有法院的指令。

对于每个问题,调查中有少部分人没有回答。

**4.37(环境)** 进行一项调查,估计在一个森林里受松树蛾(pine shoot moth)侵害的松树的百分比。在该森林的一张地图上,按照  $25 \times 25$  英尺的方形区域划分打上格子。随机选择 100 个方形区域并记录每个方形区域受到侵害的松树的数字。这是二项试验吗?

**4.38(政治学)** 进行一项调查,以了解在美国退伍军人管理局的医院里工作的护士的态度。对于一个由 1 000 个护士组成的样本,使用邮寄问卷的方法进行联系,并且记录下赞成或反对某个论点的人数。如果我们把注意限制在护士对单个问题的回答中,这种抽样是二项试验吗?如大多数邮寄调查一样,一些护士没有回应。在估计美国退伍军人管理局的医院里全部护士中赞成某个提案的百分比时,这种没有回应的现象会有什么样的影响?

**4.39(环境)** 在对洛杉矶市汽车的检查中,所有的汽车中有 60% 汽车的排放不符合 EPA 标准。对一个由 10 辆汽车构成的随机样本,计算下列概率:

- a. 所有 10 辆汽车都通不过检查。
- b. 10 辆汽车中恰好有 6 辆汽车通不过检查。
- c. 6 辆或更多的汽车通不过检查。
- d. 所有 10 辆汽车都通过检查。

使用下列 Minitab 的输出结果回答这些问题。注意, 在 Minitab 中, 二项概率  $\pi$  用  $P$  表示, 而二项变量  $y$  用  $x$  表示。

$x$	$P(X = x)$	$P(X \leq x)$
0.00	0.0001	0.0001
1.00	0.0016	0.0017
2.00	0.0106	0.0123
3.00	0.0425	0.0548
4.00	0.1115	0.1662
5.00	0.2007	0.3669
6.00	0.2508	0.6177
7.00	0.2150	0.8327
8.00	0.1209	0.9536
9.00	0.0403	0.9940
10.00	0.0060	1.000

**4.40** 参照练习 4.39。

a. 如果  $\pi = 0.3$ , 计算(a)~(d)中的概率。

b. 当  $n = 1\ 000$ ,  $\pi = 0.3$  时, 你如何计算  $P(y \leq 100)$ 。

**4.41(生物)** 用老鼠进行一项试验, 以测试一种抗凝血剂药的效果。在试验中使用了由 4 只老鼠组成的一个随机样本。如果该药的制造商宣称, 药物会对 80% 的老鼠产生所希望的效果, 那么药物对 4 只试验鼠都无效的概率是多少? 4 个中只对一个有效的概率是多少? 4 个中至多有一个有效的概率呢?

**4.42(社会学)** 犯罪学家宣称侵占公款的初犯者改过自新的概率是 0.9。假定我们规定改过自新意味着此人在 5 年内不再进行犯罪。从监狱记录中随机选择 3 个获得假释的侵占公款犯人, 并且对他们释放后 5 年的行为历史记录进行检查。如果犯罪学家的说法是正确的, 所有 3 个罪犯均改过自新的概率是多少? 其中至少有 2 个罪犯改过自新的概率呢?

**4.43** 考虑下列试验: 抛 3 个硬币并且观察正面朝上的个数  $y$ 。重复试验 100 次, 并构造一张  $y$  的频率表。注意这些频率给出  $y = 0, 1, 2$  和 3 的准确概率值的近似。(注意: 可以证明这些概率分别为  $1/8, 3/8, 3/8$ , 和  $1/8$ 。)

**4.44** 参照练习 4.43。应用二项概率分布的公式, 证明  $P(0) = 1/8$ ,  $P(1) = 3/8$ ,  $P(2) = 3/8$ , 和  $P(3) = 1/8$ 。

**4.45** 假定你和另一人各抛一枚硬币看是否匹配,共抛 1 000 次。匹配次数的均值是多少?标准差是多少?计算区间 $(\mu \pm 3\sigma)$ (提示:抛一对硬币匹配的概率是  $\pi = 0.5$ 。)

**4.46** 参照练习 4.39。说明如果  $n = 1\ 000$ ,  $\pi = 0.6$ , 你如何计算  $P(y \leq 100)$ 。

**4.47(商业)** 一个大的跨国公司在一段很长的时期内,所有实习销售员中有 10% 被评为杰出,75% 为优秀或好,10% 为满意,5% 为不满意。对一个随机选择的含有 10 个实习销售员的样本,求下列概率:

- a. 2 个被评为杰出。
- b. 2 个或 2 个以上被评为杰出。
- c. 10 个中有 8 个被评为杰出或优秀或好。
- d. 没有销售员被评为不满意。

**4.48(医药)** 一项新技术,即血管通栓气球(balloon angioplasty)被广泛地应用于打开栓塞的心脏瓣膜和血管。把气球通过一根导管插入心脏并且使之膨胀,从而打开栓塞,因此,不需要进行外科手术。未经治疗的心脏瓣膜疾病患者有 50% 的人在大约 2 年以内死亡。如果这项新技术的使用经验表明,约有 70% 的人存活超过 2 年的时间,那么,在一所医院内用血管通栓气球方法治疗的下 5 个病人构成一个  $n = 5$ ,  $\pi = 0.70$  的二项试验吗?为什么是或为什么不是?

**4.49(医药)** 一家处方药公司宣称,所有在动物试验中表明有效的药物中,仅有 12% 通过临床试验程序并且投放到市场。如果一个公司有 15 种新的化合物在动物试验中是有效的,求下列概率:

- a. 没有一种药物能投放市场。
- b. 一种或一种以上的药物能投放市场。
- c. 二种或二种以上的药物能投放市场。

**4.50** 练习 4.49 满足二项试验的性质吗?为什么?

**4.51(商业)** 在一家大型超级市场,从大量列表中选择了 50 个价格变化的随机样本。如果一个价格变化被准确公布的概率为 0.93,

- a. 写出 3 个或更少价格变化不被准确公布的概率的表达式。
- b. 对(a)应做出怎样的假设?

## 4.9 连续随机变量的概率分布

离散随机变量(例如二项分布)的可能的取值是孤立和分散的,例如 0, 1, 2 或 3。另外一类随机变量是连续随机变量:他们可能的值形成一个区间(或范围,或连续统)。例如,一种普通股票每美元投资的 1 年回报可能取值为 0 与某个相当大的值所构成的区间中的任何值。在实践中,实际上所有的随机变量的取值范围都呈

现为一个离散的集合;每一张百万美元普通股票每美元的投资回报是 1.06219423 美元或 1.06219424 美元或 1.06219425 美元或……然而,当随机变量的可能值很多时,把这样的随机变量当作连续随机变量来处理在数学上有时是有益的。

从理论上讲,一个连续随机变量的取值与数轴上一个区间中的无穷多个点相对应。粗略地说,我们不可能把小数值的概率分配到  $y$  的各个值上(象对离散随机变量所做的那样)并且保持概率的总和等于 1 的性质。

为了克服这个困难,我们再回到第三章中的频率直方图的概念,在讲解这个概念时,我们讨论的是  $y$  落在给定的区间内的概率。因为分类区间的个数可以变大而区间的长度可以减小,对于一个包含大量测量值的总体,其频率直方图几乎是光滑的曲线。这样,我们可以设想,通过重复观察一个连续随机变量得到一条光滑的曲线,并将其作为总体的频率分布的一个模型。所得到的曲线类似于图 4.6 中给出的曲线。

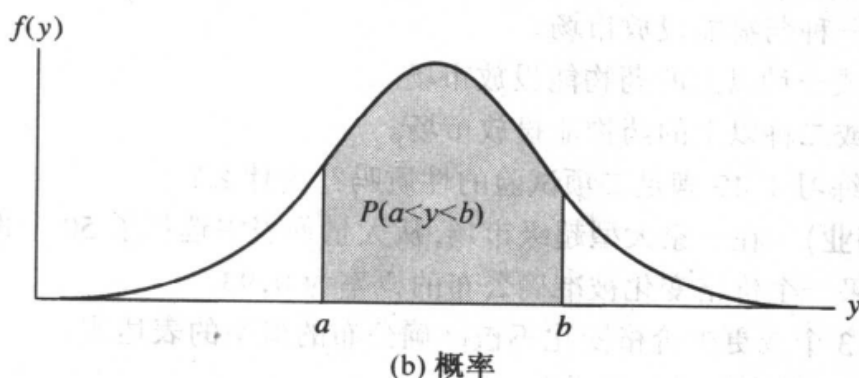
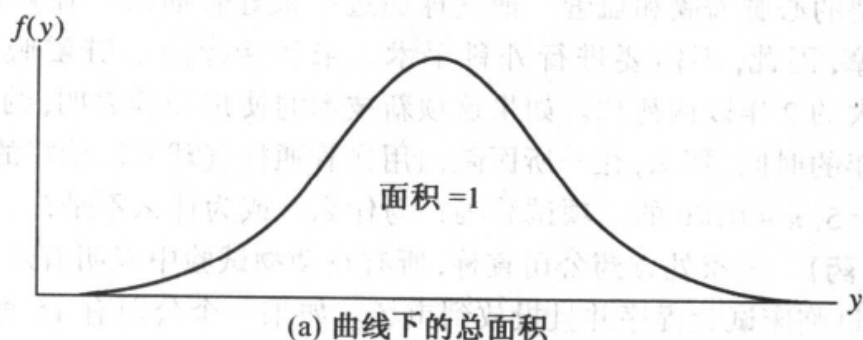


图 4.6 连续随机变量的概率分布

回想第三章中所讲的,频率直方图中,频率的大小与对应的分类区间上的面积成比例,且这些面积有其概率解释。也就是说,如果从观测值的集合中随机地选取一个,那么,该值落在一个分类区间上的概率与该区间上方对应的直方图面积的大小成比例。由于一个总体是指全部(100%或 1),我们在画出这条光滑概率曲线时,使其下面的总面积的大小等于 1。如果我们令曲线下的总面积为 1,则每个区

间上的面积恰等于相应的概率。

一个连续随机变量的概率分布图如图 4.7 中所示。对给定的  $y$  的值, 纵坐标 (即曲线的高度) 用记号  $f(y)$  表示。许多人喜欢这样说, 就象  $P(y)$  为二项随机变量的概率一样,  $f(y)$  是表示连续随机变量  $y$  的概率的一个量。然而, 如我们以前所指出的, 不可能为一个连续随机变量的无穷多个可能的取值中的每一个指定一个概率。因此, 我们只能说  $f(y)$  表示在给定的  $y$  值处, 这个概率分布的高度。

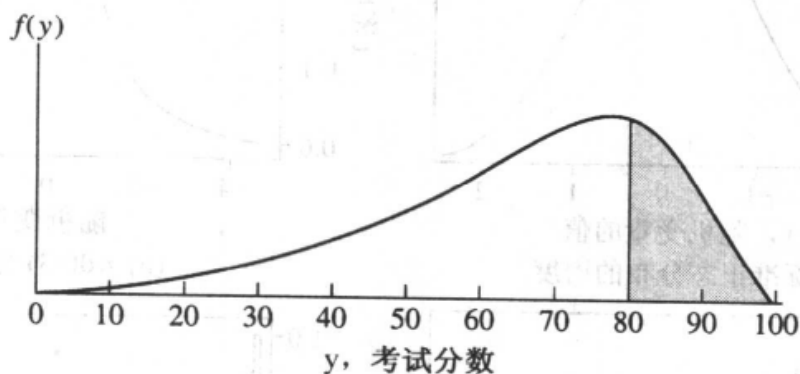


图 4.7 假设的学生考试成绩的概率分布

一个连续随机变量落在一个区间上, 比如说在 2 个点  $a$  和  $b$  之间的概率, 可以直接从频率直方图 (3.3 节) 中区间上的面积的概率解释得出, 它等于从  $a$  到  $b$  这一区间上曲线下方的面积, 如图 4.6 中所示。这个概率写为  $P(a < y < b)$ 。

有许多种形状的曲线能用来表示作为连续随机变量的测量指标的总体频率分布。幸运的是, 对于大多数这些曲线, 其下面的区间上的面积已制成表格, 可供查用。这样, 如果我们知道学生的考试成绩有特定的概率分布, 如图 4.7 所示, 并且如果曲线下的面积已制成表, 那么, 我们通过查表中的面积值, 能求出某个特定的学生取得 80 以上成绩的概率, 它等于图 4.7 中阴影部分的面积。

图 4.8 描绘了 4 种重要的概率分布, 我们将在后面的章节中频繁地使用这些分布。对于一个特定的状况, 我们使用哪个概率分布是很重要的, 因为概率陈述由曲线下面的面积决定。如从图 4.8 中所见, 选取不同的分布, 将得到迥异的答案。例如, 在图 4.8(a) 和 (b) 的概率分布中, 随机变量取值小于 5.0 的概率实质上是 1.0, 但在图 4.8(c) 和 (d) 的概率分布中, 它的概率分别是 0.584 和 0.947。在某些状况下, 我们不知道在特定的研究中随机变量的确切分布。这时, 我们可以使用随机变量的观察值构造频率直方图, 它是真实概率分布的一个样本估计。就统计推断而言, 在许多情况中选择一个连续随机变量的概率分布的准确的形状不是关键的, 因为我们的大部分推断方法对概率分布形状的精确表述并不敏感。

我们将会发现, 对于连续变量收集的数据常常有近似于钟形的频率分布, 如在图 4.8(a) 中所看到的那样。一个连续的变量 (即正态变量) 和它的概率分布 (钟形



曲线)为这种类型的数据提供了一个良好的模型。正态分布的变量在统计推断中也是很重要的。我们将在下一节详细地研究正态分布。

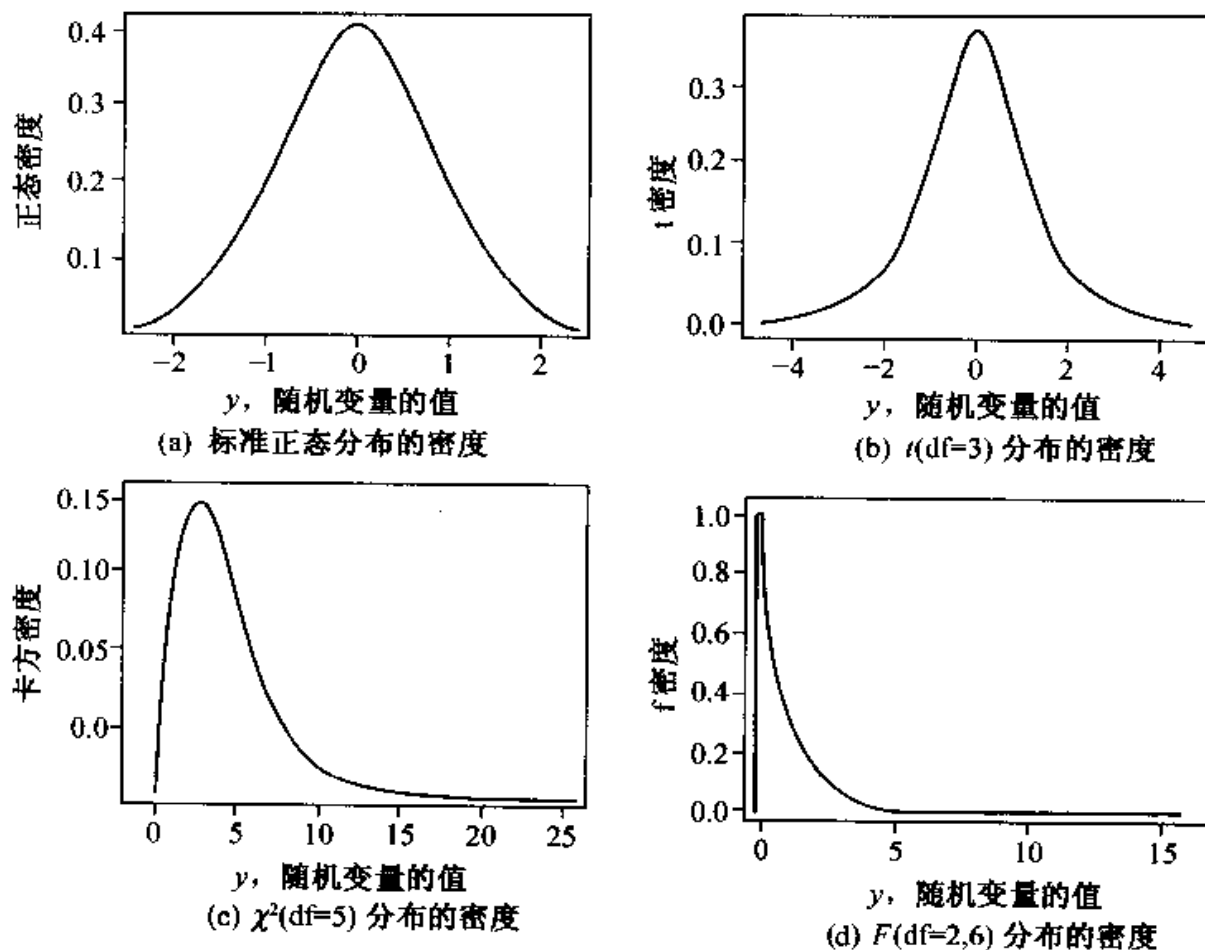


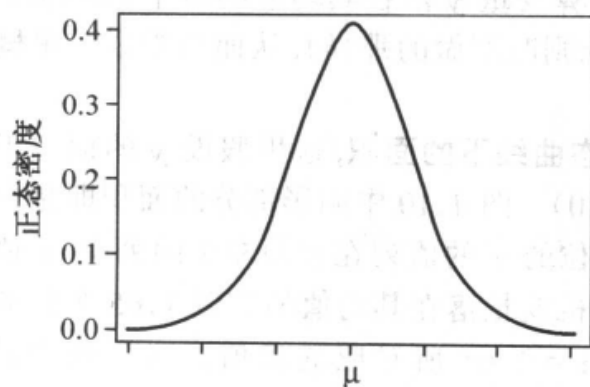
图 4.8

## 4.10 一个常用的连续随机变量:正态分布

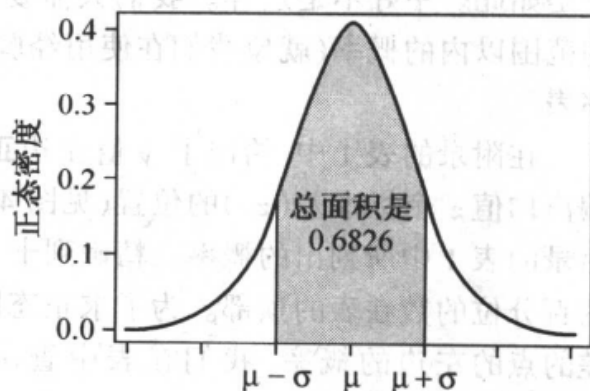
许多要研究的变量,包括在以后的章节要讨论的一些统计量,有丘状的频率分布,这些分布能用正态曲线近似。例如,近期有重复攻击行为历史的门诊病人按简要精神病分级标准(brief psychiatric rating scale)所得总分数的分布是丘状的。丘状分布的其他实际的例子包括:从特定的社会经济背景中挑选的学龄前儿童的对社会的理解能力得分,循环型狂躁抑郁病患者(circular-type manic-depressive)的精神运动延迟(psychomotor retardation)得分,特定品种的牛的牛奶产量以及一个社区的居民焦虑感觉的得分。这些丘状的分布都能用一条正态曲线近似。

既然正态分布已被很好地制成表格,我们就可以使用正态曲线下的面积——它对应于概率——来近似与在试验中要研究的变量有关的概率。因此,正态随机变量及其分布在统计推断中扮演一个重要的角色。

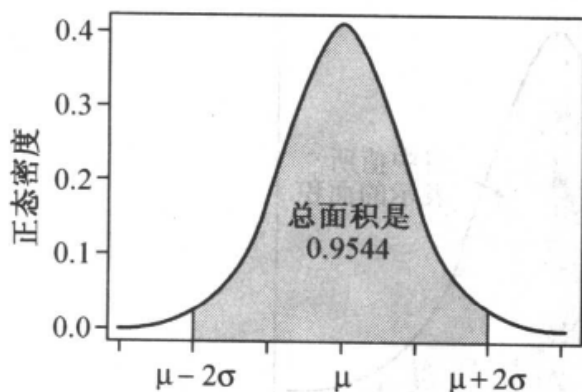
正态随机变量的频率直方图,又叫做正态曲线或正态概率分布,是一条光滑的钟形曲线。图 4.9(a)中给出了一条正态曲线。如果我们用  $y$  表示正态随机变量,那么在  $y$  的特定值处概率分布的高度用  $f(y)^*$  表示。正态曲线下的各个概率值形成了第三章中所讲的“经验准则”的基础。



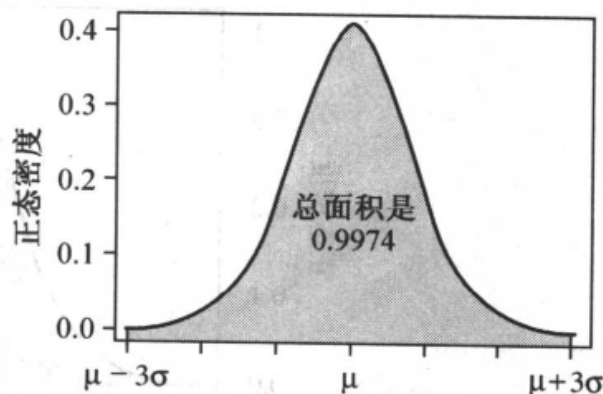
(a) 正态分布的密度



(b) 均值 1 个标准差以内正态曲线下的面积



(c) 均值 2 个标准差以内正态曲线下的面积



(d) 均值 3 个标准差以内正态曲线下的面积

图 4.9

如我们从图 4.9(a)中所见,正态概率分布是钟形的,并且关于均值  $\mu$  对称。尽管从理论上讲正态随机变量  $y$  可能取到从  $-\infty$  到  $+\infty$  的任何值,但由经验准则我们知道,近乎所有的测量值落在  $\mu$  的 3 个标准差( $3\sigma$ )的范围以内。从经验准则,我们还知道,如果从一个有丘形分布的测量值的总体中随机选择一个测量值,这个测量值落在该总体均值的 1 个标准差范围以内的概率近似为 0.68(见图 4.9(b))。同样,我们知道一个值落在区间  $\mu \pm 2\sigma$  内的概率近似为 0.954,而落在区间  $\mu \pm 3\sigma$  内的概率近似为 0.997(见图 4.9(c)和(d))。然而,我们还不知道一个测量值落在其均值的 1.65 个标准差范围以内的概率,也不知道它落在其均值的 2.58

\* 对于正态分布,  $f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}$ , 其中  $\mu$  和  $\sigma$  分别是  $y$  值总体的均值和标准差。

个标准差范围以内的概率。在本节余下的部分,我们将讨论计算正态曲线下一个测量值落在均值  $\mu$  的任何距离以内的概率。

因为有许多不同的正态曲线(取决于参数  $\mu$  和  $\sigma$ ),为所有的正态曲线下的面积(概率)制表似乎是一项不可能的任务,特别当每条曲线要求一张单独的表时就更是如此。幸好不是这样。我们只需要计算变量  $y$  落在其均值的若干倍标准差的范围以内的概率(就象我们在使用经验准则时所做的那样),从而只需要一张概率表。

在附录的表 1 中,给出了  $y$  值左边正态曲线下的面积,这里假设  $y$  的值位于偏离均值  $z$  个标准差( $z\sigma$ )的位置(见图 4.10)。图 4.10 中阴影部分的面积即是在附录的表 1 中所列出的概率。精确到十分位的  $z$  的值列在表的左手的列中, $z$  值的百分位的数在表的顶部。为了求正态随机变量落在其均值的右侧 1.65 个标准差的点的左边的概率,我们在表中查出  $z = 1.65$  所对应的表值。这个概率是 0.9505(见图 4.11)。

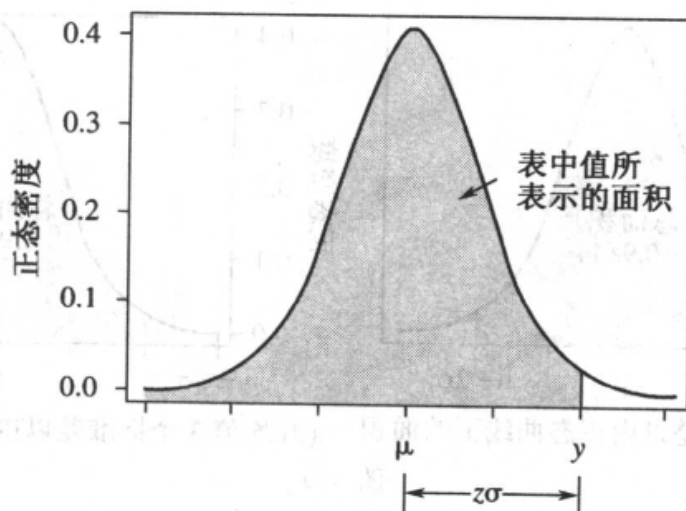


图 4.10 附录的表 1 中给出的正态曲线下的面积

要求一个测量值小于某个  $y$  值的概率,我们首先运用下面的公式计算  $y$  与均值的偏离是标准差的多少倍:

$$z = \frac{y - \mu}{\sigma}$$

用这个公式计算所得到  $z$  的值,有时称为这个  $y$  值的  $z$  得分。使用计算出的  $z$  值,我们用附录中的表 1 来确定所求的概率。注意我们仅是通过减去  $\mu$  除以  $\sigma$  对  $y$  的值进行编码。(换句话说,  $y = z\sigma + \mu$ 。)图 4.12 说明了  $y$  值与  $z$  值之间的对应关系。例如,低于  $\mu$  ( $\mu$  的左边)2 个标准差的  $y$  值,对应于  $z = -2$ 。

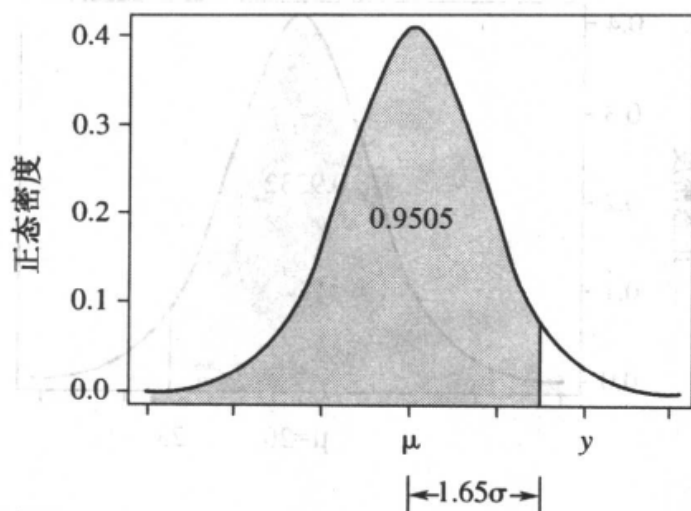


图 4.11 正态曲线下落在其均值  $\mu$  的右侧 1.65 个标准差的点的左边的面积

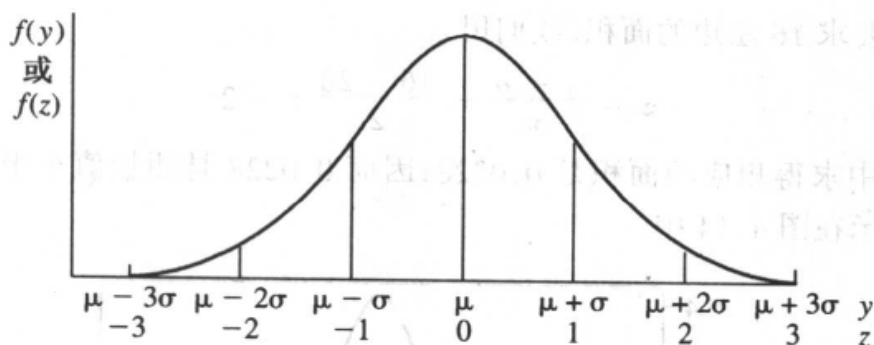


图 4.12  $y$  与  $z = (y - \mu) / \sigma$  之间的关系

#### 例 4.12

考虑一个  $\mu = 20$  和  $\sigma = 2$  的正态分布。确定一个测量值小于 23 的概率。

**解答** 面对这样的问题, 首先画一张图可能有利于看清所求的面积, 如在图 4.13 中所看到的。

为了求出曲线下方  $y = 23$  左边的面积, 我们首先计算  $y = 23$  与均值的偏离是标准差的多少倍。

$$z = \frac{y - \mu}{\sigma} = \frac{23 - 20}{2} = 1.5$$

这样,  $y = 23$  位于  $\mu = 20$  右侧 1.5 个标准差。参考附录中的表 1, 我们求出对应于  $z = 1.5$  的面积是 0.9332。这就是测量值小于 23 的概率。

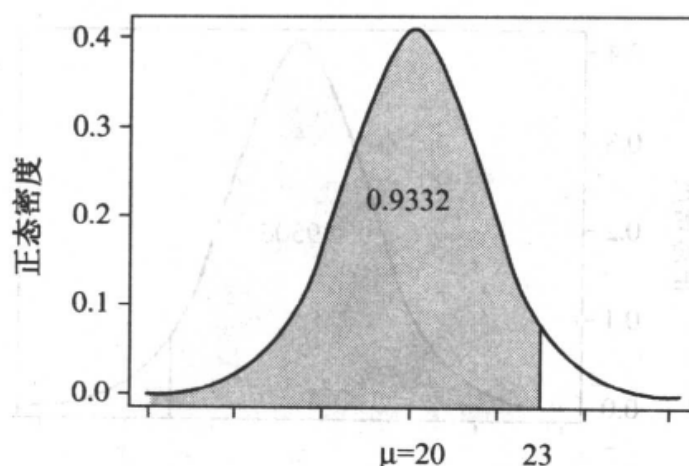


图 4.13  $\mu = 20, \sigma = 2$  的正态曲线下,  $y = 23$  左边的面积

#### 例 4.13

对例 4.12 中  $\mu = 20, \sigma = 2$  的正态分布, 求  $y$  小于 16 的概率。

**解答** 要求 16 左边的面积, 我们用

$$z = \frac{y - \mu}{\sigma} = \frac{16 - 20}{2} = -2$$

我们从表 1 中求得相应的面积是 0.0228; 因而 0.0228 是测量值小于 16 的概率。这个面积表示在图 4.14 中。

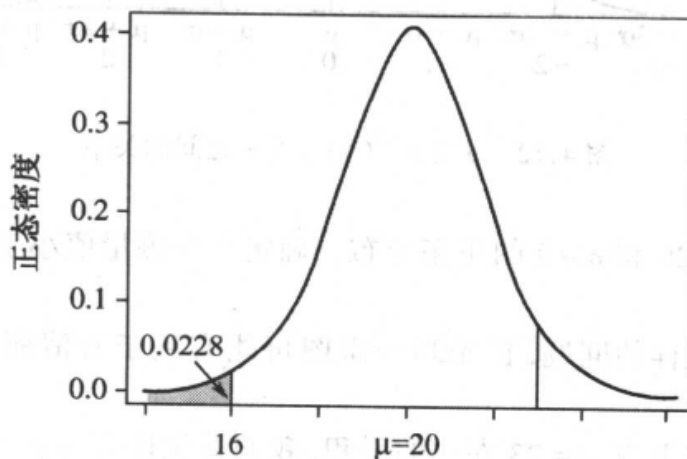


图 4.14  $\mu = 20, \sigma = 2$  的正态曲线下方  $y = 16$  左边的面积

#### 例 4.14

一群格恩西(Guernsey)奶牛日产奶量的均值服从正态分布, 其中  $\mu = 70$  磅,  $\sigma = 13$  磅。

- 随机选择的奶牛产奶量小于 60 磅的概率是多少?
- 随机选择的奶牛产奶量大于 90 磅的概率是多少?
- 随机选择的奶牛产奶量在 60 磅至 90 磅之间的概率是多少?

**解答** 首先我们画出所要求的面积的图形(图 4.15(a)~(c))。为了回答(a),我们必需计算与值 60 相对应的  $z$  的值。值  $y = 60$  对应的  $z$  得分为

$$z = \frac{y - \mu}{\sigma} = \frac{60 - 70}{13} = -0.77$$

从表 1 中得知,60 左边的面积是 0.2206(见图 4.15(a))。

为了回答(b),值  $y = 90$  对应的  $z$  得分为

$$z = \frac{y - \mu}{\sigma} = \frac{90 - 70}{13} = 1.54$$

从表 1 中我们得出 90 左边的面积是 0.9382。于是,大于 90 的面积必定为  $1 - 0.9382 = 0.0618$ ,因为曲线下的总面积等于 1(见图 4.15(b))。

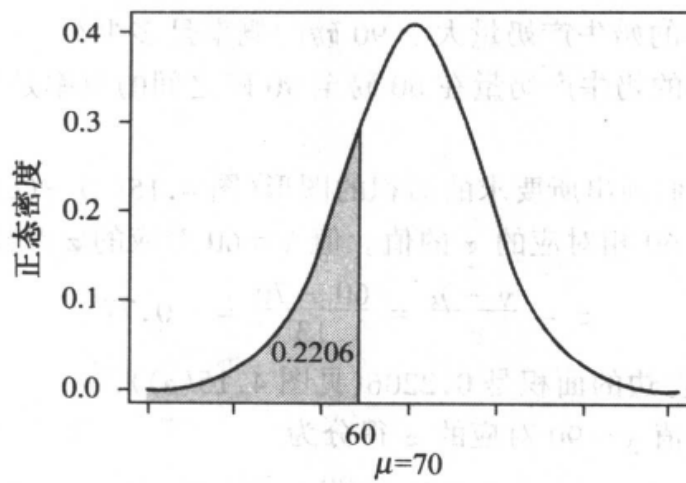
要回答(c),我们可以使用(a)和(b)的结果。 $y_1$  和  $y_2$  两个值之间的面积由在这 2 个值左边的面积的差来决定(见图 4.15(c))。我们已经求出小于 60 的面积是 0.2206,小于 90 的面积是 0.9382。因此,在 60 和 90 之间的面积是  $0.9382 - 0.2206 = 0.7176$ 。我们由此得出,奶牛产量的 22.06% 小于 60 磅,6.18% 大于 90 磅,71.76% 在 60 和 90 磅之间。

正态分布的一个重要的方面是我们可以很容易得到该分布的百分位数。一个分布的  $100p$  百分位数是这样一个值  $y_p$ ,总体值的  $100p\%$  小于  $y_p$ ,  $100(1-p)\%$  大于  $y_p$ 。例如,总体的中位数是 50 百分位数  $y_{0.50}$ ,两个四分位数分别是 25 和 75 个百分位数。正态分布是对称的,所以其中位数和均值相等:  $y_{0.50} = \mu$ (见图 4.16(a))。

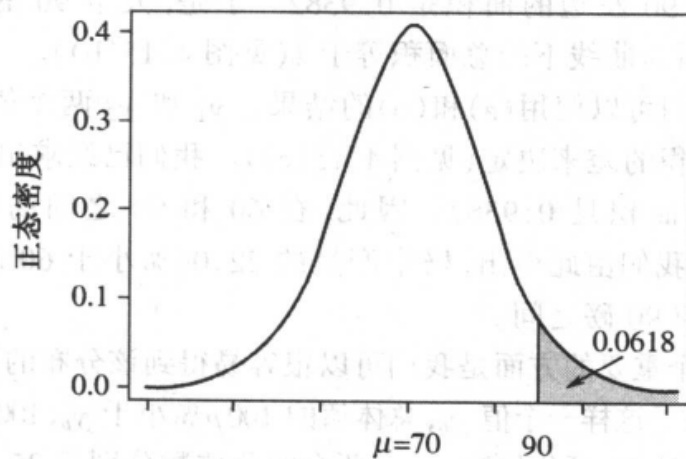
为了求出标准正态分布的百分位数,我们将表 1 的用法反过来。要求出  $100p$  百分位数  $z_p$ ,我们在表 1 中找到概率  $p$ ,然后沿着表的边缘读出与它相应的数字  $z_p$ 。例如,求 80 百分位数  $z_{0.80}$  时,我们在表 1 中寻找概率  $p = 0.8000$ 。最接近 0.8000 的值是 0.7995,与它对应的  $z$  的值是 0.84。于是,  $z_{0.80} = 0.84$ (见图 4.16(b))。现在,我们来求均值为  $\mu$ ,标准差为  $\sigma$  的正态分布的  $100p$  百分位数  $y_p$ 。为此,我们需要反向使用标准化公式。

$$y_p = \mu + z_p \sigma$$

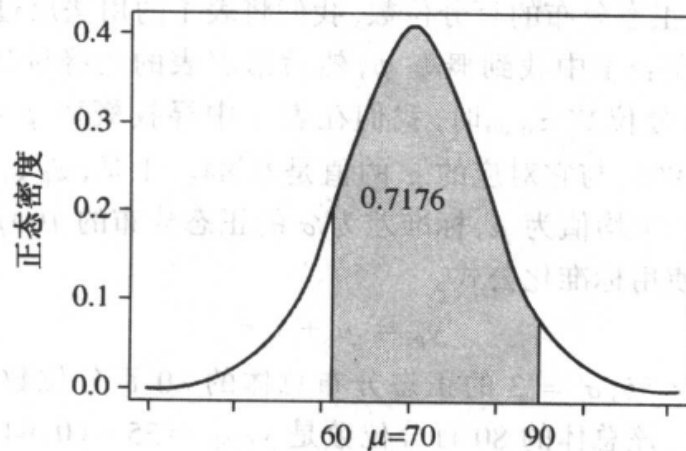
假定我们要求  $\mu = 55$ ,  $\sigma = 3$  的正态分布总体的 80 百分位数。我们已经求出了  $z_{0.80} = 0.84$ ,于是,该总体的 80 百分位数是  $y_{0.80} = 55 + (0.84)(3) = 57.52$ 。



(a)  $\mu = 70, \sigma = 13$  的正态曲线下方  $y = 60$  左边的面积

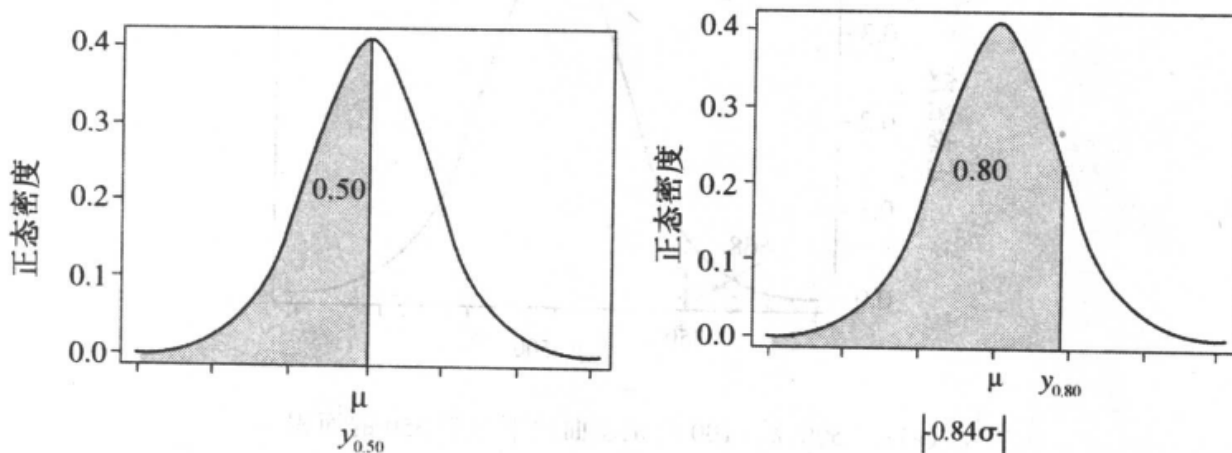


(b)  $\mu = 70, \sigma = 13$  的正态曲线下方  $y = 90$  右边的面积



(c)  $\mu = 70, \sigma = 13$  的正态曲线下方 60 到 90 之间的面积

图 4.15



(a) 对于正态曲线,均值和中位数相同

(b) 正态曲线的 80 百分位点

图 4.16

**例 4.15**

学校评估测验(SAT)是一种用于测试个人为上大学所做准备的考试。数学得分有均值为 500, 标准差为 100 的正态分布。参加 SAT 的人得分低于 350 的比例是多少? 要确定一组学生是否需要辅导, 我们要确定所有分数中最低的 10%, 即我们要确定 10 百分位数  $y_{0.10}$ 。

**解答** 要求出得分低于 350 的比例(见图 4.17(a)), 我们需要求出 350 左边的面积:

$$z = \frac{y - \mu}{\sigma} = \frac{350 - 500}{100} = -1.5$$

在正态分布中, 均值 -1.5 倍标准差那一点左边的面积, 可以从表 1 中查得, 为 0.0668。因此, 6.68% 或近似于 7% 的参加考试的人其得分低于 350。得分 350 近似于所有得分总体的 7 百分位数  $y_{0.07}$ 。

为求 10 百分位数(见图 4.17(b)), 我们首先在表 1 中求出  $z_{0.10}$ 。因为 0.1003 是最接近于 0.1000 的值, 而它的对应的  $z$  值是 -1.28, 我们取  $z_{0.10} = -1.28$ 。然后计算

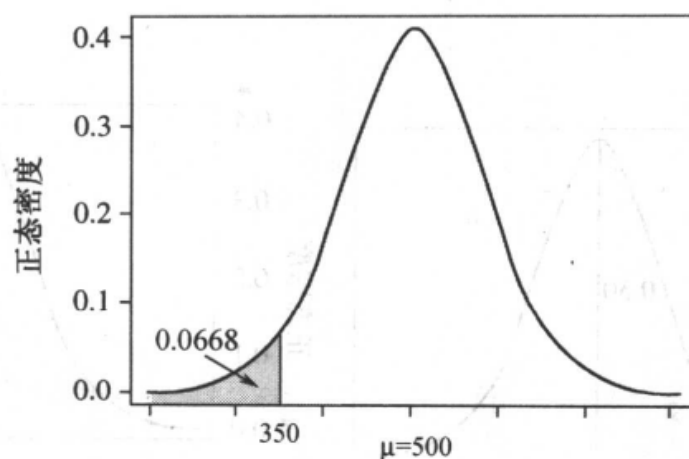
$$y_{0.10} = \mu + z_{0.10}\sigma = 500 + (-1.28)(100) = 500 - 128 = 372$$

这样, 在 SAT 中有 10% 的分数不到 372。

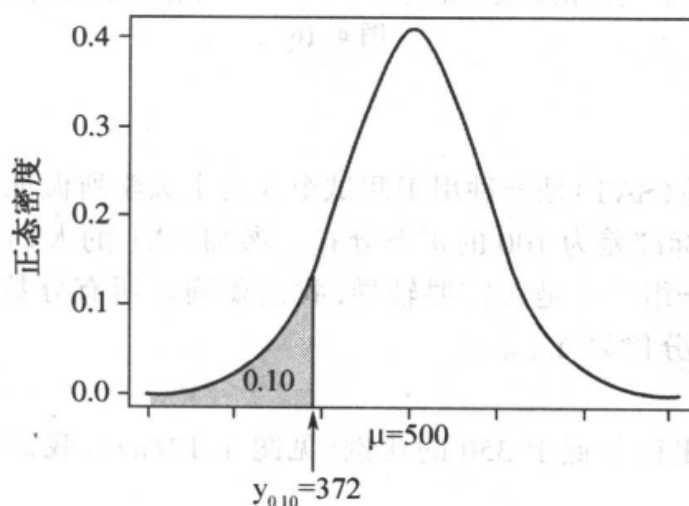
**例 4.16**

对上一年所得税纳税申报单的分析表明, 对一个给定的收入阶层, 在头三笔





(a)  $\mu = 500, \sigma = 100$  的正态曲线下小于 350 的面积



(b)  $\mu = 500, \sigma = 100$  的正态曲线的 10 百分位数

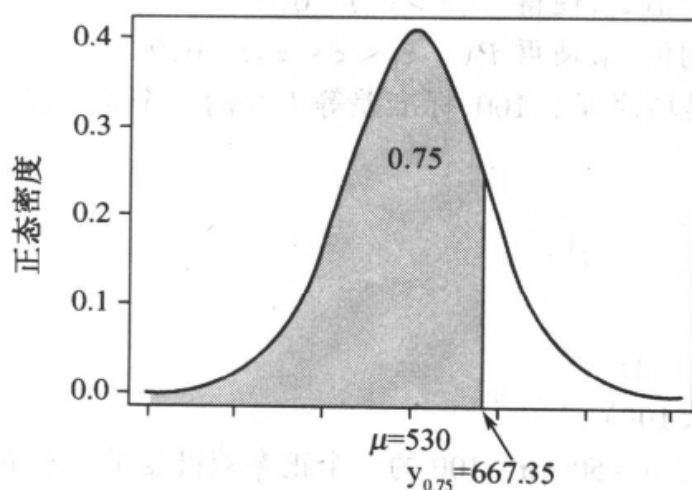
图 4.17

支付中,除已付的纳税额估计凭单以外,拖欠政府的税额近似服从均值是 530 美元,标准差是 205 美元的正态分布。对这个测量值的分布,求其 75 百分位数。政府想要找出所欠额度中最大的 25% 的那些申报单。

**解答** 我们需要决定 75 百分位数  $y_{0.75}$  (图 4.18)。从表 1 中,我们求出  $z_{0.75} = 0.67$ , 因为最接近 0.7500 的概率是 0.7486, 与它对应的  $z$  得分为 0.67。我们然后计算

$$y_{0.75} = \mu + z_{0.75}\sigma = 530 + (0.67)(205) = 667.35$$

这样,在这个收入阶层中,纳税申报单的 25% 拖欠政府的税额超过了 667.35 美元。

图 4.18  $\mu = 530, \sigma = 205$  的正态曲线的 75 百分位数

## 练习

## 基本技能

4.52 使用附录中的表 1 求正态曲线下这些值之间的面积:

a.  $z = 0$  和  $z = 1.3$ 。

b.  $z = 0$  和  $z = -1.9$ 。

4.53 对下面的值重复练习 4.52:

a.  $z = 0$  和  $z = 0.7$ 。

b.  $z = 0$  和  $z = -1.2$ 。

4.54 对下面的值重复练习 4.52:

a.  $z = 0$  和  $z = 1.29$ 。

b.  $z = 0$  和  $z = -0.77$ 。

4.55 对下面的值重复练习 4.52:

a.  $z = -0.21$  和  $z = 1.35$ 。

b.  $z = 0.37$  和  $z = -1.20$ 。

4.56 对下面值重复练习 4.52:

a.  $z = 1.43$  和  $z = 2.01$ 。

b.  $z = -1.74$  和  $z = -0.75$ 。

4.57 求  $z$  大于 1.75 的概率。

4.58 求  $z$  小于 1.14 的概率。

4.59 求  $z$  的值  $z_0$ , 使得  $P(z > z_0) = 0.5$ 。

4.60 求  $z$  的值  $z_0$ , 使得  $P(z > z_0) = 0.025$ 。

4.61 求  $z$  的值  $z_0$ , 使得  $P(z > z_0) = 0.0089$ 。

4.62 求  $z$  的值  $z_0$ , 使得  $P(z > z_0) = 0.05$ 。

4.63 求  $z$  的值  $z_0$ , 使得  $P(-z_0 < z < z_0) = 0.95$ 。

4.64 设  $y$  是均值等于 100, 标准差等于 8 的一个正态随机变量。求下列概率:

a.  $P(y > 100)$ 。

b.  $P(y > 110)$ 。

c.  $P(y < 115)$ 。

d.  $P(88 < y < 112)$ 。

e.  $P(100 < y < 108)$ 。

4.65 设  $y$  是  $\mu = 500, \sigma = 100$  的一个正态随机变量。求下列概率:

a.  $P(500 < y < 696)$ 。

b.  $P(y > 696)$ 。

c.  $P(304 < y < 696)$ 。

d. 设  $P(500 - k < y < 500 + k) = 0.60$ , 求  $k$ 。

4.66 设  $y$  是  $\mu = 100, \sigma = 15$  的一个正态随机变量。

a. 证明  $y < 130$  等价于  $z < 2$ 。

b. 把  $y > 82.5$  转换为等价的  $z$  得分。

c. 求  $P(y < 130)$  和  $P(y > 82.5)$ 。

d. 求  $P(y > 106)$ ,  $P(y < 94)$  和  $P(94 < y < 106)$ 。

e. 求  $P(y < 70)$ ,  $P(y > 130)$  和  $P(70 < y < 130)$ 。

4.67 用附录中的表 1 计算曲线下方下列值之间的面积:

a.  $z = 0$  和  $z = 1.5$ 。

b.  $z = 0$  和  $z = 1.8$ 。

4.68 对下列值重复练习 4.67:

a.  $z = -1.96$  和  $z = 1.96$ 。

b.  $z = -2.33$  和  $z = 2.33$ 。

4.69 当  $z$  右边的面积是 0.05 时, 求  $z$  的值是多少? 左边的面积为 0.05 时  $z$  的值是多少? (提示: 利用附录中的表 2)

4.70 对下列面积求  $z$  的值。

a.  $z$  右边的面积是 0.01。

b.  $z$  左边的面积是 0.10。

4.71 求观察到的  $z$  值大于下列值的概率。

a. 1.96。

b. 2.21。

c. 2.86。

d. 0.73。

4.72 求观察到的  $z$  值小于下列值的概率。

a. -1.20。

b. -2.62。

c. 1.84。

d. 2.17。

### 应用

4.73(政府) 由某个州的预算部门所保存的记录表明,从递交传票到最终支付资金所经过的总时间近似服从均值为 39 天和标准差为 6 天的正态分布。

a. 从递交传票到最终支付资金之间经过的时间超过 50 天的概率是多少?

b. 如果你在 55 天以前或更早递交了一张传票,你能得出什么结论?

4.74(教育) 每年对数以千计的中学生实施的大学入学考试的成绩按照均值为 500 且标准差是 100 的分布来评定。这些成绩近似于服从正态分布。求满足下列条件的成绩所占的百分比。

a. 大于 600。

b. 大于 700。

c. 小于 450。

d. 在 450 和 600 之间。

4.75(商业) 某一食品厂每月的销售额近似服从均值是 150(千美元)标准差是 35(千美元)的正态分布。计算下列概率:

a.  $P(y > 200)$ 。

b.  $P(y > 220)$ 。

c.  $P(y < 120)$ 。

d.  $P(100 < y < 200)$ 。

4.76 参考练习 4.74。一个限于某些人参加的俱乐部希望邀请在大学入学考试中成绩在最高的 10% 中的学生加入。

a. 要被邀请加入该俱乐部,得分需为多少?

b. 把总体的最高的 60% 与最低的 40% 分开的分值是多少? 我们把这个值称为什么?

4.77 一个正态分布的均值是 50, 标准差是 10。

a. 什么百分位数是 38? 选择适当的答案。

88.49   38.49   49.99   0.01   11.51

b. 下列哪一个  $z$  得分对应于 67 百分位数?

1.00   0.95   0.44   2.25 以上都不是

**4.78(社会)** 一大群高中男生的体重的分布是  $\mu = 120$  磅,  $\sigma = 10$  磅的正态分布。下列哪个是正确的?

- a. 大约 16% 的男生将超过 130 磅。
- b. 大概少于 2.5% 的男生低于 100 磅。
- c. 可以预期男生的一半体重小于 120 磅。
- d. 以上都是正确的。

## 4.11 随机抽样

迄今为止,在本书中我们已经讨论了随机样本,并在第二章介绍了各种抽样方案。随机抽样的重要性是什么?我们必须知道样本要如何选择,才能确定与各种抽样结果有关的概率。用*随机方法选择的样本*的概率是可以确定出来的,并且我们可以使用这些概率,从抽取的样本对总体做出推断。

用非随机的方法选择的样本数据经常受到*抽选偏倚*的干扰。无论何时,只要在抽选时存在过分或过少代表总体的某些部分的系统倾向,就会有抽选偏倚产生。例如,如果在进行家庭调查时,调查时间选在整个一星期内从上午 9 点到下午 5 点的时间段内,那么,调查结果就会向着至少有一个成员在家的家庭严重地偏倚。因此,从这个调查的样本数据所做出的任何推断将偏向于至少一个成员在家的那些家庭的态度或意见,从而对于该区域家庭的总体不具有真正的代表性。

现在,我们转而定义从一个包含  $N$  个测量值的总体中抽取的含有  $n$  个测量值的*随机样本* ( $N > n$ )。(注意,这是在第二章中讨论过的一种简单随机样本。因为本书讨论的大多数随机样本是简单随机样本,除非为了澄清的需要,我们将省略“简单”这一形容词。)

**定义 4.13** 从总体中抽选大小为  $n$  的样本时,如果所有不同的大小为  $n$  的样本被选择的概率都相同,那么,这样抽取的样本称为一个*随机样本*。

### 例 4.17

计划在美国对 10 个最大的城市进行一项有关手枪犯罪的研究。在进行初步调查后,该研究将从 10 个最大的城市中随机选择 2 个作为深入研究的对象。要研究的总体是 10 个最大的城市  $\{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}\}$ 。列出从 10 个城市的总体中选取 2 个城市时,所有可能的不同的样本。给出从总体中抽取  $n = 2$  个城市的随机样本时,每个样本被抽中的概率。

**解答** 所有可能的样本列表如下。

样本	城市	样本	城市	样本	城市
1	$C_1, C_2$	16	$C_2, C_9$	31	$C_5, C_6$
2	$C_1, C_3$	17	$C_2, C_{10}$	32	$C_5, C_7$
3	$C_1, C_4$	18	$C_3, C_4$	33	$C_5, C_8$
4	$C_1, C_5$	19	$C_3, C_5$	34	$C_5, C_9$
5	$C_1, C_6$	20	$C_3, C_6$	35	$C_5, C_{10}$
6	$C_1, C_7$	21	$C_3, C_7$	36	$C_6, C_7$
7	$C_1, C_8$	22	$C_3, C_8$	37	$C_6, C_8$
8	$C_1, C_9$	23	$C_3, C_9$	38	$C_6, C_9$
9	$C_1, C_{10}$	24	$C_3, C_{10}$	39	$C_6, C_{10}$
10	$C_2, C_3$	25	$C_4, C_5$	40	$C_7, C_8$
11	$C_2, C_4$	26	$C_4, C_6$	41	$C_7, C_9$
12	$C_2, C_5$	27	$C_4, C_7$	42	$C_7, C_{10}$
13	$C_2, C_6$	28	$C_4, C_8$	43	$C_8, C_9$
14	$C_2, C_7$	29	$C_4, C_9$	44	$C_8, C_{10}$
15	$C_2, C_8$	30	$C_4, C_{10}$	45	$C_9, C_{10}$

现在,我们假定从 45 个可能的样本中选择一个  $n=2$  个城市的随机样本。如果每个样本有相等的被选择的概率,即  $1/45$ ,选取的样本就称为一个**随机样本**。

从总体中选择容量为  $n$  的随机样本,最简单、最可靠的方法之一是使用一张随机数表(见附录中的表 13)。**随机数表**是这样的一张数字的表格,不管你从表的哪儿开始,也不管你向哪个方向移动,数字是随机出现的并且有相等的概率。这样,如果我们希望从包含 100 个测量值的总体中选择由  $n=10$  个测量值组成的一个随机样本,可以把总体中的测量值标记为  $0\sim 99$ (或  $1\sim 100$ )。然后参考附录中的表 13,并且从中选择一个随机起点,从该起点开始向前接下来的 10 个 2 位数就作为包含在选取的随机样本中的那些测量值的标记。同样,通过向上移动或向下移动,我们也将获得一个随机样本。

对于所有可能的样本进行列表,只有当样本容量  $n$  和总体大小  $N$  都不大时才可行。从含有  $N$  个测量值的总体中选取样本容量为  $n$  的样本时,所有不同的样本的个数  $M$  可以由下列的公式确定:

$$M = \frac{N!}{n!(N-n)!}$$

在例 4.17 中,  $N=10, n=2$ 。于是,

$$M = \frac{10!}{2!(10-2)!} = \frac{10!}{2!8!} = 45$$

甚至当  $N$  不大时,  $M$  的值也会很大。例如, 如果  $N=50$  而  $n=5$ , 那么  $M=2\,118\,760$ 。这样, 如果要从含有  $N=50$  个测量值的总体中抽取由  $n=5$  个测量值组成的随机样本, 要列出所有的 2 118 760 种可能的样本是很不切实际的。在实践中, 我们把总体中的  $N$  个元素编号为从 1 到  $N$ , 并构造一张表, 列出这些编号, 该表称为**抽样框架**。然后使用随机数表(见附录表 13)或计算机程序, 从整数  $(1, 2, \dots, N)$  中随机选取  $n$  个整数。大多数统计软件程序包含从整数  $(1, 2, \dots, N)$  中随机选取  $n$  个整数的例程, 其中  $N > n$ 。练习 4.86 中包含了使用 Minitab 产生随机样本时需要的命令。

#### 例 4.18

一个小社区由 850 个家庭组成。我们希望获得 20 个家庭的一个随机样本, 以查明公众对工资和价格限定的接受情况。参考附录中的表 13, 决定哪些家庭作为样本。

**解答** 假设有社区中所有的家庭的列表(例如一个电话目录), 因而我们可以把这些家庭标记为 0~849(或等价地标记为 1~850)。然后, 参考附录中表 13, 我们选择一个起点。假定我们决定从第 1 行第 3 列开始。沿着这一页, 我们选取头 20 个介于 000 和 849 之间的 3 位数。从表 13, 我们得到

015	110	482	333
255	564	526	463
225	054	710	337
062	636	518	224
818	533	524	055

这 20 个数字指明了在我们的样本中包括的 20 个家庭。

电话目录并不总是最好的名字的来源, 特别在调查与经济或政治有关的问题时。在 1936 年的总统选举中, 富兰克林·罗斯福作为民主党的候选人与共和党候选人堪萨斯的州长阿尔弗雷德·兰登竞选。此时正是美国的困难时期; 美国还没有从 20 世纪 30 年代的大萧条中恢复过来, 仍然有 900 万人失业。

读者文摘(*Literary Digest*)为了预测选举的获胜者进行了一项投票公众的抽样。使用从电话本和俱乐部的会员表中拿到的名字和地址, *Literary Digest* 向

外送出了 1 000 万份问卷调查表,并收回了 240 万份。基于调查表的回答,他们预言兰登将以 57% 对 43% 取得胜利。

这时候,乔治·盖洛普(George Gallup)正在开始他的调查业务。他进行了两项调查。第一项基于对 3 000 人的调查,在 *Literary Digest* 的结果发表之前很久就预言了 *Literary Digest* 调查的结果会是什么;第二项调查基于 50 000 人的样本,正确地预测到了罗斯福的胜利。

盖洛普怎么能正确地预言 *Literary Digest* 的调查结果呢?他又怎么能在其后的另一项调查中,正确预言选举的结果呢? *Literary Digest* 哪里出现了错误?第一问题是严重的抽选偏倚。从电话目录和俱乐部会员表中选取名字和地址,其调查系统地排除了穷人。不幸的是,投票沿着经济水平分离;穷人大多数投票给了罗斯福,而富人趋于投票给兰登。犯错误的第二个原因可能是由于无回答偏倚。因为在 1 000 万人中仅仅有 20% 返回了他们的调查表,并且在那些返回的表中有近似一半是赞成兰登的,人们可能猜想也许没返回调查表的人与做了回答的人者有不同的偏爱。事实正是如此。

那么,怎样得到一个随机样本呢?这需要细心计划,机敏从事,甚至于要将近就进行近似的随机抽样时也需要如此。当要研究的问题包括人时,就更是这样。有些人可能很难与其打交道;他们容易丢弃邮寄的调查问卷并拒绝参与访问调查。除非非常谨慎,我们获得的数据可能包含各种偏倚,从而对我们要进行的推断造成未知的影响。

我们没有足够的时间在本书中进一步探讨随机抽样的问题;样本调查研究方法论的内容可以构成大学生和研究生的一系列课程。要记住的重要的一点是,从随机样本得出的数据将作为以后的各章中进行统计推理的基础。随机样本虽然是不容易获得的,但是小心一些,我们就能避免许多潜在的能影响我们所做推断的偏倚。

## 练习

### 基本技能

**4.79** 详细说明随机样本的意义。能否抽出一个真正的随机样本?给出解释。

**4.80(政府)** 假定我们想要从 800 个人的总体中选择  $n = 10$  个人的一个随机样本。使用附录中的表 13 找出被选在样本中的人。

**4.81(政治学)** 参考练习 4.80。确定从  $N = 1\,000$  的总体中抽取的一个  $n = 15$  的随机样本中的元素。



## 应用

**4.82(社会)** 某城市要提高地方税额来改进公立学校的质量。城市官员要通过抽样来了解一个社区中拥有房产的人关于此事的意见。如果用一张随机数表来确定样本中的家庭,并且当访问者访问时房主不在家的时候,就放弃这一家,这个过程会接近于随机抽样吗?解释其原因。

**4.83(社会)** 一家地方电视网想要对居住在一个当地投票选区的人进行的一项非正式调查,以初步了解他们对于一个筹措资金以把市属的历史博物馆迁到一个新地点的提案的意见。这家电视网怎样才可能对这些选民进行近似的随机抽样?

**4.84(社会)** 一个心理学家要研究在获得离婚的过程中的妇女,以确定这些妇女在完成离婚以后是否经历了重要的观念转变。待研究的地理区域的现存记录表明,近期有 798 对夫妇离婚。假设研究中需要 25 位妇女的一个样本,试用附录中的表 13 决定应要求哪些妇女参加到研究中来。(提示:在第 2 列第 1 行开始向下走。)

**4.85** 参考练习 4.84。在大多数调查的案例中,不是所有被选择参加研究的人都同意参与。假定选择的 25 位妇女中有 5 位妇女拒绝参与。确定在研究中要包括另外 5 位妇女。

**4.86(社会)** 假定要求你就即将来临的选举进行民意测验。城市里有 230 个选区,而你需要从每个选区中随机选取 50 个登记的选民。假定每个选区有 1 000 个登记的选民,并且你可以获得一张这些人的表。你在每张表上把 1 000 个人标以数字 1~1 000,以 1 记表中的第一人,以 1 000 记表中的最后一人。接下来你需要从 1 到 1 000 这 1 000 个数字中获得 50 个数字的一个随机样本。抽样框架上相应于这 50 个数字的名字就是你在民意测验中选取的 50 个人。为了说明这一过程,这里给出一个 Minitab 程序。注意你需要分别运行这个程序 230 次,以在 230 个选区中获得每个选区的一个新的随机样本。

按照以下步骤:

单击“Calc”。

单击“Random Data”。

单击“Integer”。

在“Generate rows of data”对话框中输入 5。

在“Store in Column(s):”对话框中输入 c1~c10。

在“Minimum value:”对话框中输入 1。

在“Maximum value:”对话框中输入 1000。

单击“OK”。

单击“File”。

单击“Print Worksheet”。

- a. 使用一张随机数表或一个计算机程序,从数字 1~1 000 中产生 50 个数字的第 2 个随机样本。
- b. 给出为什么你需要每个选区产生一个不同的随机数集合的若干原因。为什么不在所有的 230 个选区中使用同样的 50 个数字的集合?

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
1	340	701	684	393	313	312	834	596	321	739
2	783	877	724	498	315	282	175	611	725	571
3	862	625	971	30	766	256	40	158	444	546
4	974	402	768	593	980	536	483	244	51	201
5	232	742	1	861	335	129	409	724	340	218

## 4.12 抽样分布

我们在第三章讨论了中心趋势和变异性的若干不同的度量,并且讨论了总体的数值描述量(参数)与样本的数值描述量(统计量)之间的区别。例如, $\mu$  和  $\sigma$  是参数,而  $\bar{y}$  和  $s$  是统计量。

样本统计量的数值不能预先确切地预测出来。即使我们知道总体的均值  $\mu$  是 216.37 美元,总体的标准差  $\sigma$  是 32.90 美元——就算我们知道完全的总体分布——我们也不能说样本均值  $\bar{y}$  将恰好等于 216.37 美元。一个样本统计量是一个随机变量;由于它基于从感兴趣的总体中抽取的测量值的一个随机样本,统计量的取值有随机的变异。像其他随机变量一样,一个样本统计量也有其概率分布。我们把一个样本统计量的概率分布称为这统计量的*抽样分布*。换句话说,统计量的抽样分布是该统计量的所有的可能值的总体。

抽样分布的实际的数学推导是数理统计学的基本问题之一。我们将举例说明对简单的总体如何获得  $\bar{y}$  的抽样分布。在以后章节中,我们将给出一些一般的结果。

### 例 4.19

样本均值  $\bar{y}$  是从容量为 2 的一个随机样本计算得到的,该随机样本是从由 10 个值(2, 3, 4, 5, 6, 7, 8, 9, 10, 11)组成的总体中取出的。基于容量为 2 的随机样本,求  $\bar{y}$  的抽样分布。

**解答** 求抽样分布的一种方法是通过计数。从 10 个个体中选取 2 个个体共有 45 个可能的样本。列出如下：

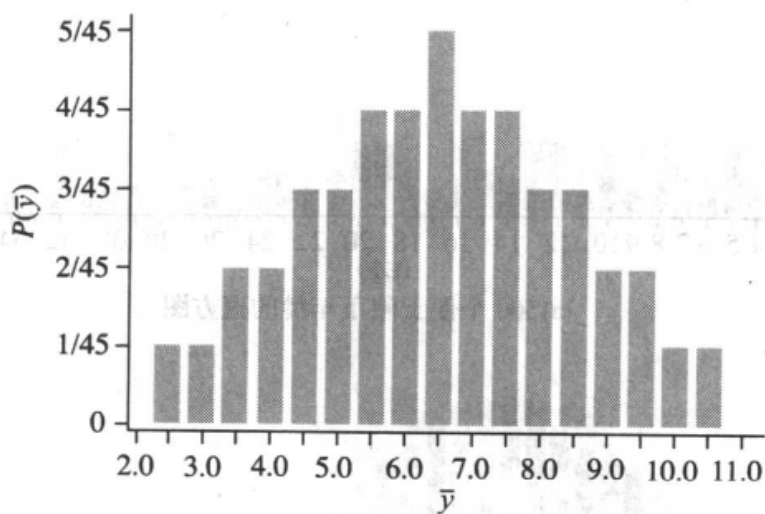
样本	$\bar{y}$ 的值	样本	$\bar{y}$ 的值	样本	$\bar{y}$ 的值
2,3	2.5	3,10	6.5	6,7	6.5
2,4	3	3,11	7	6,8	7
2,5	3.5	4,5	4.5	6,9	7.5
2,6	4	4,6	5	6,10	8
2,7	4.5	4,7	5.5	6,11	8.5
2,8	5	4,8	6	7,8	7.5
2,9	5.5	4,9	6.5	7,9	8
2,10	6	4,10	7	7,10	8.5
2,11	6.5	4,11	7.5	7,11	9
3,4	3.5	5,6	5.5	8,9	8.5
3,5	4	5,7	6	8,10	9
3,6	4.5	5,8	6.5	8,11	9.5
3,7	5	5,9	7	9,10	9.5
3,8	5.5	5,10	7.5	9,11	10
3,9	6	5,11	8	10,11	10.5

假定每个容量为 2 的样本是等可能的,则基于从总体  $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$  中抽出的  $n=2$  个观察值,  $\bar{y}$  的抽样分布如下表所示。

$\bar{y}$	$P(\bar{y})$	$\bar{y}$	$P(\bar{y})$
2.5	1/45	7	4/45
3	1/45	7.5	4/45
3.5	2/45	8	3/45
4	2/45	8.5	3/45
4.5	3/45	9	2/45
5	3/45	9.5	2/45
5.5	4/45	10	1/45
6	4/45	10.5	1/45
6.5	5/45		

该抽样分布显示在图 4.19 中。注意分布是对称的,其中均值是 6.5 和标准差约为 2.0(极差除以 4)。

例 4.19 说明,对很小的总体,我们能做到枚举从总体选出的容量为 2 的所有

图 4.19  $\bar{y}$  的抽样分布

可能的样本,并且算出样本均值的所有可能的值。下一个例子将说明当从较大的总体抽样时,样本均值  $\bar{y}$  的性质。该例子说明,  $\bar{y}$  作为  $\mu$  的估计,其性态取决于样本的大小  $n$ 。在本章的后面,我们还将说明总体分布的形状对于  $\bar{y}$  的抽样分布的影响。

#### 例 4.20

在这个例子中,总体的值是已知的,因而我们能准确地计算总体均值  $\mu$  和总体的标准差  $\sigma$ 。我们将从总体选取样本容量为  $n=5, 10$  和  $25$  的样本,并以此来考察  $\bar{y}$  的性态。总体由 500 个便士组成,我们可以计算出每个便士的年龄:年龄 = 2000 - 便士上的日期。这 500 个年龄的直方图显示在图 4.20(a)中。图形向右偏倚,并且右边有一条很长的尾巴。算得均值和标准差分别为  $\mu = 13.468$  年和  $\sigma = 11.164$  年。为了对  $n=5$  生成  $\bar{y}$  的抽样分布,我们需要产生容量为  $n=5$  的所有可能的样本,然后对这些样本的每一个计算  $\bar{y}$ 。因为从 500 个元素的总体中抽选的容量为 5 的可能的样本有 255,244,687,600 个,这些计算将是一项庞大的任务。样本容量为 10 或 25 的可能的样本的数就更是大得连国债数额也相形见绌。因此,我们将使用计算机程序 S-plus,从 500 个便士的总体中选出容量为 5 的 25,000 个样本。例如,第一个样本由年龄是 4, 12, 26, 16 和 9 的便士组成,相应的样本均值  $\bar{y} = (4 + 12 + 26 + 16 + 9)/5 = 13.4$ 。我们重复 25,000 次选取 5 个便士的过程,记录它们的年龄,  $y_1, y_2, y_3, y_4, y_5$ , 然后计算  $\bar{y} = (y_1 + y_2 + y_3 + y_4 + y_5)/5$ 。用所得到的  $\bar{y}$  的这 25,000 个值作频数直方图,这个图称为  $n=5$  时  $\bar{y}$  的抽样分布。对样本容量  $n=10$  和  $25$  进行类似的过程,获得的抽样分布显示在图 4.20(b) ~ (d) 中。

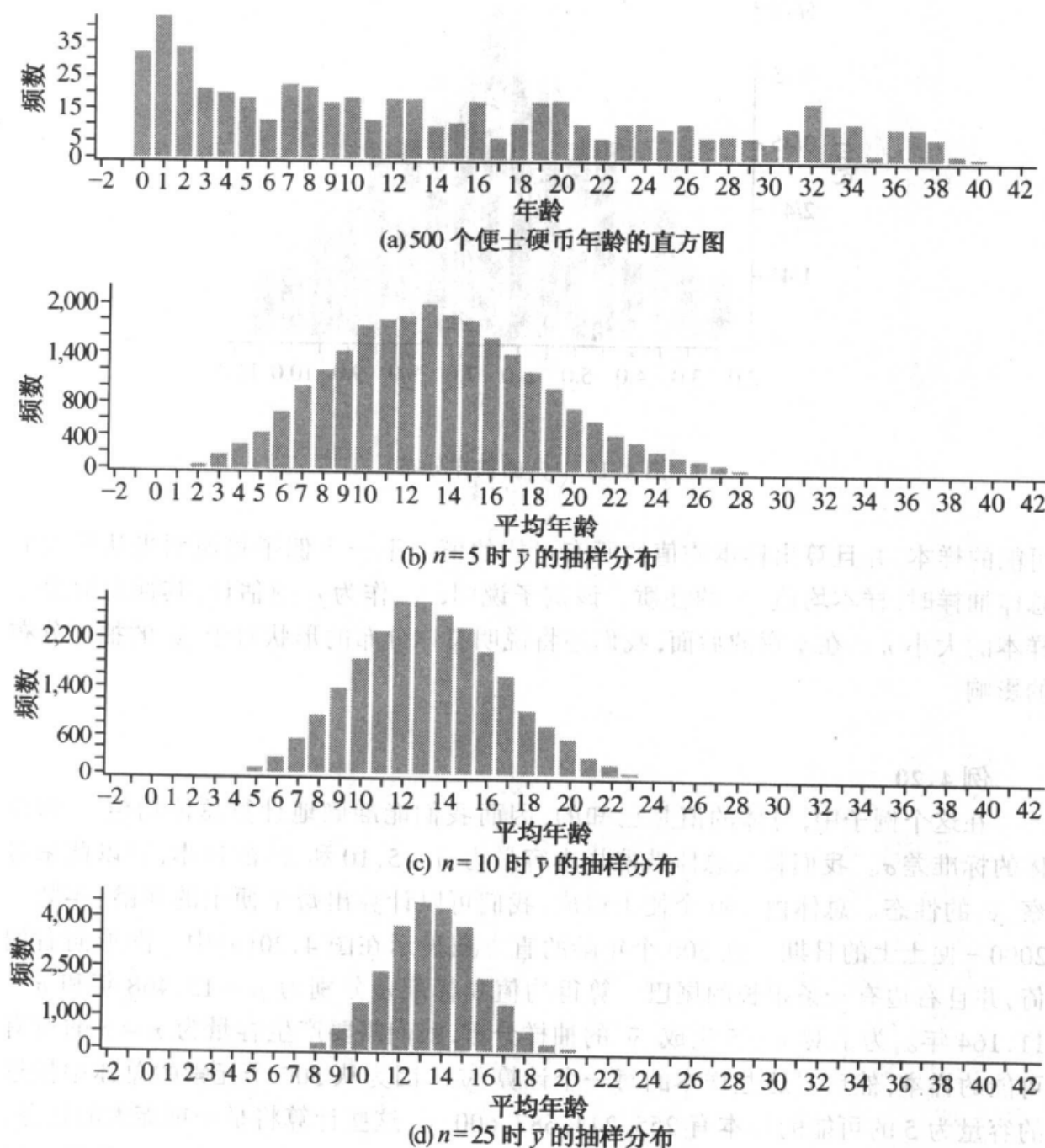


图 4.20

注意,所有三种抽样分布有近似相同的中心值,近似于 13.5。(见表 4.5)对于三种样本,  $\bar{y}$  的均值大体上与总体均值  $\mu = 13.468$  一致。事实上,如果我们对  $n$  的所有三个值生成所有可能的样本,那么  $\bar{y}$  的可能值的均值将准确地等于  $\mu$ 。

对于这三张直方图,下一个值得注意的特征是它们的形状。所有三张图在形状上具有某种程度的对称性,当  $n = 25$  时达到了接近于正态分布的形状。然而,

基于样本容量  $n=5$  的  $\bar{y}$  的直方图,比基于  $n=10$  的直方图张得更开些,而基于  $n=10$  的直方图比  $n=25$  的直方图张得更开。当  $n$  较小时,比起  $n$  较大时我们更可能获得远离  $\mu$  的  $\bar{y}$  的值。是什么原因增加了  $\bar{y}$  的值分散程度呢?在样本中一个极端的  $y$ ,不论相对于  $\mu$  是大还是小,当  $n$  小时比  $n$  大时,在  $\bar{y}$  的大小上有更大的影响。因此,基于小的  $n$  的样本均值作为  $\mu$  的估计时,不如那些大样本的均值来得精确。

表 4.5 包含了  $\bar{y}$  的抽样分布的概括统计量。 $\bar{y}$  的抽样分布有均值  $\mu_{\bar{y}}$  和标准差  $\sigma_{\bar{y}}$ ,它们与总体均值  $\mu$  和标准差  $\sigma$  有下列关系:

$$\mu_{\bar{y}} = \mu, \quad \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

表 4.5  $\bar{y}$  的抽样分布的均值和标准差

样本容量	$\bar{y}$ 的均值	$\bar{y}$ 的标准差	$\frac{11.1638}{\sqrt{n}}$
1(总体)	13.468( $\mu$ )	11.1638( $\sigma$ )	11.1638
5	13.485	4.9608	4.9926
10	13.438	3.4926	3.5303
25	13.473	2.1766	2.2328

从表 4.5 中,我们注意到三种抽样均值近似等于总体的均值;三种抽样的标准差近似等于  $\sigma/\sqrt{n}$ 。如果我们产生了  $\bar{y}$  的所有的可能的值,那么  $\bar{y}$  的标准差将准确地等于  $\sigma/\sqrt{n}$ 。数量  $\sigma_{\bar{y}} = \sigma/\sqrt{n}$  称为  $\bar{y}$  的标准误差。

在常见的样本统计量中,有不少统计量,如样本中位数和样本标准差,对于中等以上的样本容量  $n$  的值,有近似于正态分布的抽样分布。通过下面的计算,我们可以看到这一点。对于从 500 个便士的年龄总体中抽取的由 25,000 个样本组成的三个样本集合(分别对应于  $n=5, 10, 25$ ),计算出各个样本的中位数和样本标准差。所得样本中位数结果的样本分布表示在图 4.21(a)~(d)中,图 4.22(a)~(d)则表示样本标准差的计算结果。样本中位数和标准差的抽样分布比样本均值的抽样分布倾斜得更厉害。事实上,样本中位数和标准差的抽样分布要近似于正态分布的形状所需要的  $n$  值比样本均值所需要的  $n$  值大得多。我们称之为**中心极限定理**的数理统计学的一系列定理,为我们用正态分布去近似许多样本统计量的真实的抽样分布提供了理论上的保证。我们将讨论关于样本均值的一个这样的定理。对样本中位数,样本标准差和样本比例有类似的定理。

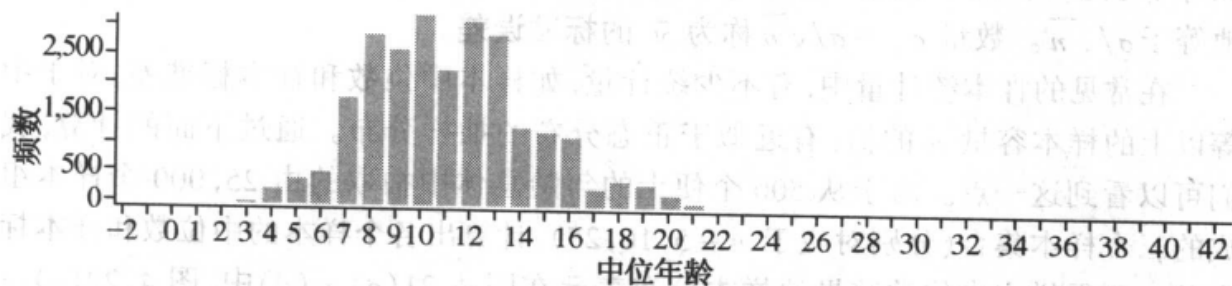
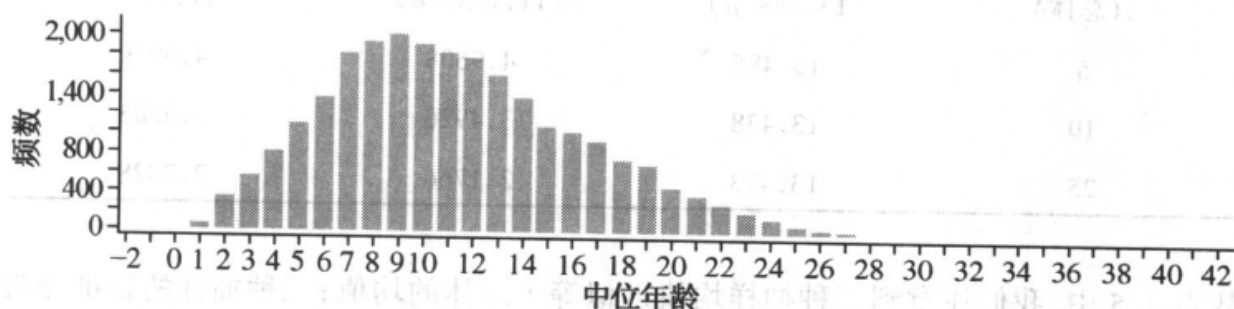
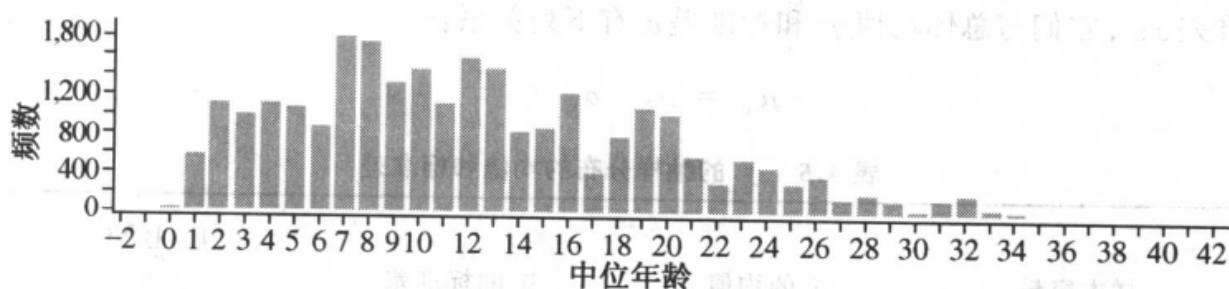
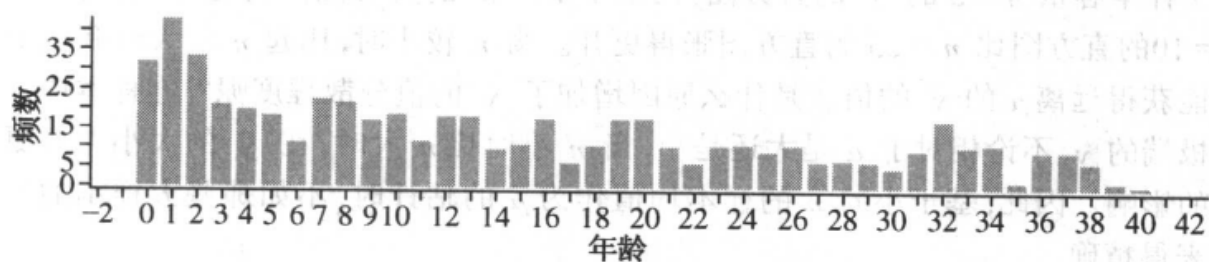


图 4.21

### 定理 4.1 关于 $\bar{y}$ 的中心极限定理

设  $\bar{y}$  表示从抽自于均值为  $\mu$  并且有有限的标准差  $\sigma$  的总体的、容量为  $n$  的随机样本计算出来的样本均值。设  $\mu_{\bar{y}}$  和  $\sigma_{\bar{y}}$  分别表示  $\bar{y}$  的抽样分布的均值和标准差。基于从总体中重复抽取的容量为  $n$  的随机样本,我们可以得出下列结论:

1.  $\mu_{\bar{y}} = \mu$ 。

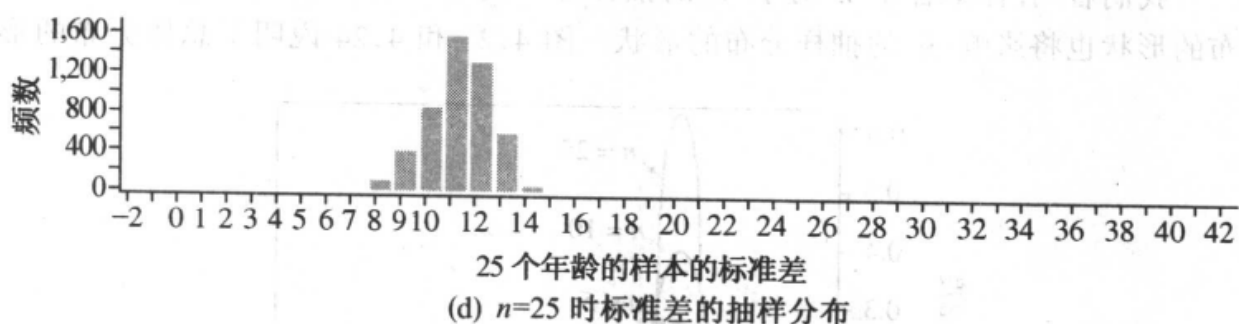
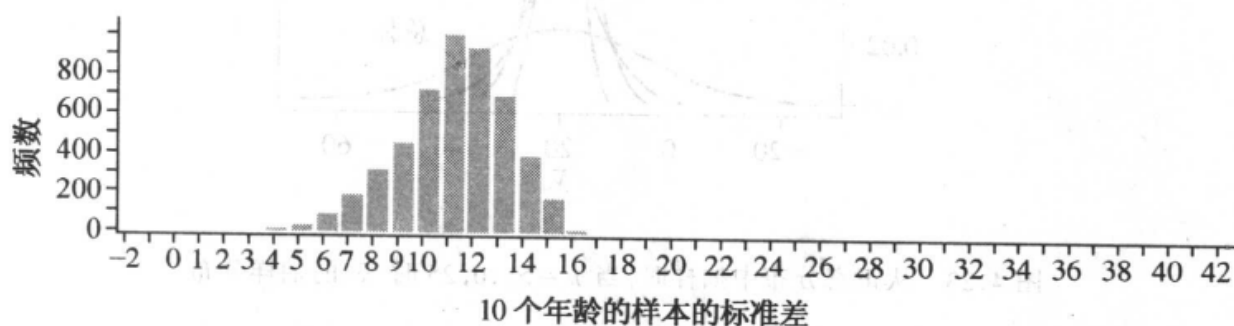
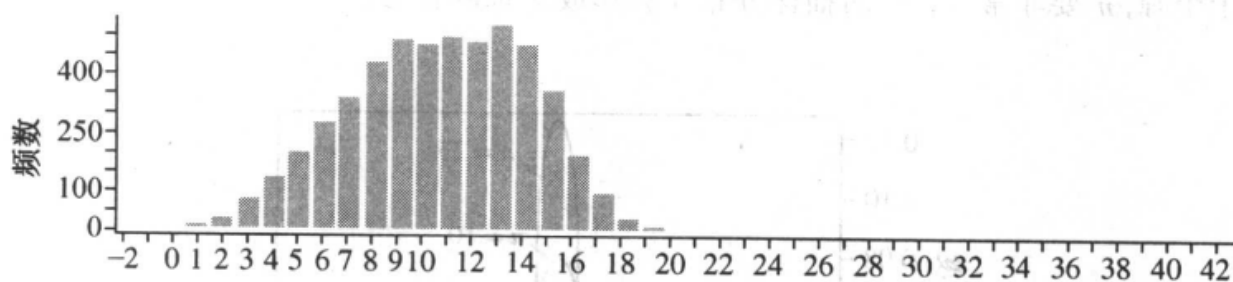
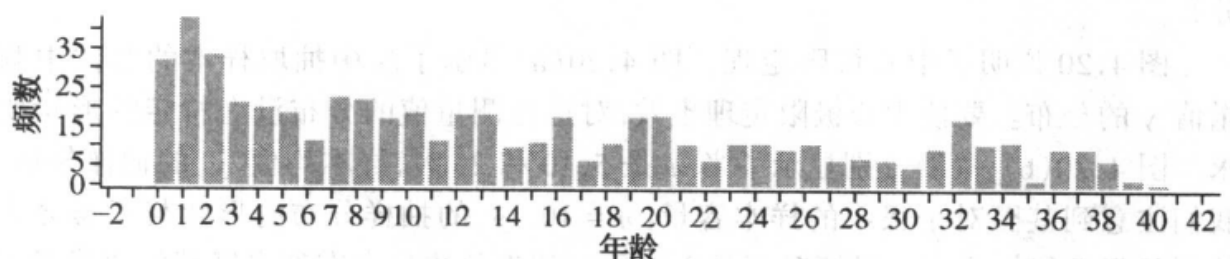


图 4.22

2.  $\sigma_{\bar{y}} = \sigma / \sqrt{n}$ 。

3. 当  $n$  取大值时,  $\bar{y}$  的抽样分布将逼近正态分布(随着  $n$  的增加这种近似将愈加精确)。

4. 当总体分布是正态分布时,对任意的样本容量  $n$ ,  $\bar{y}$  的样本分布均为正态



分布。

图 4.20 说明了中心极限定理。图 4.20(a)显示了从中抽取样本的总体中测量值  $y$  的分布。要使中心极限定理生效,对这些测量值的分布没有特定的形状要求。图 4.20(b)~(d)分别显示了当  $n$  是 5, 10 和 25 时,样本均值  $\bar{y}$  的抽样分布。我们注意到甚至对于很小的样本容量,  $n = 10$ ,  $\bar{y}$  的抽样分布的形状与正态分布的形状很类似。这在一般情况下是不对的。如果总体分布有许多极端的值或者若干个峰,  $n$  要非常大,  $\bar{y}$  的抽样分布才会形成对称的钟形。

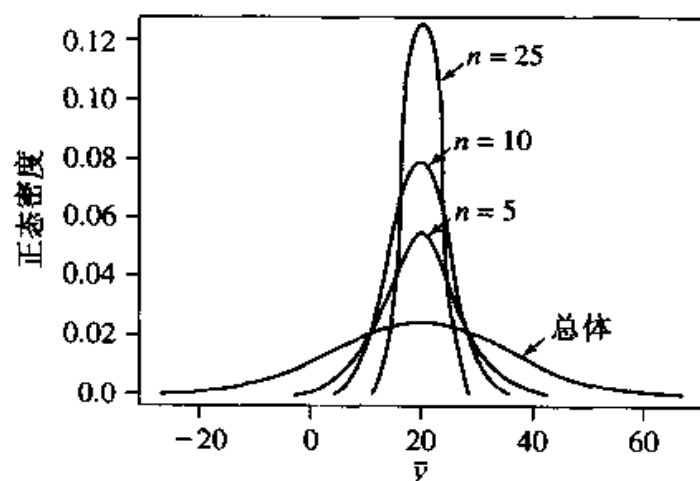


图 4.23 从正态分布中抽样时,当  $n = 5, 10, 25$  时  $\bar{y}$  的抽样分布

我们看到,样本容量  $n$  对于  $\bar{y}$  的抽样分布的形状有影响。总体测量值的分布的形状也将影响  $\bar{y}$  的抽样分布的形状。图 4.23 和 4.24 说明了总体分布的形

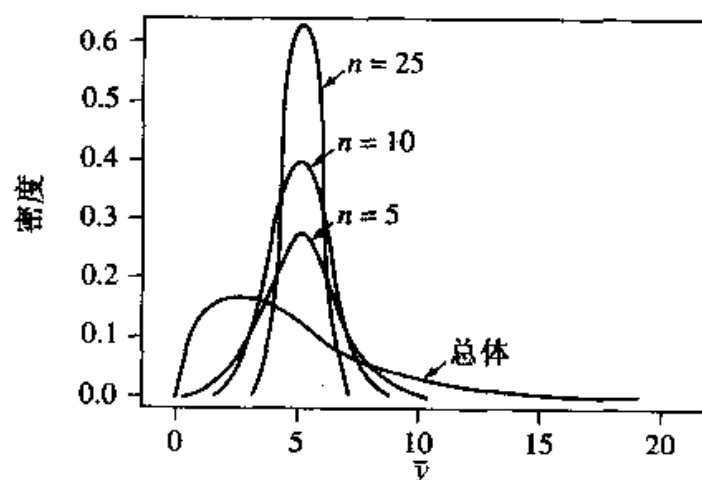


图 4.24 从偏倚的分布中抽样时,当  $n = 5, 10, 25$  时  $\bar{y}$  的抽样分布

状对  $\bar{y}$  的抽样分布的形状上的影响。在图 4.23 中, 总体测量值有正态分布。对  $n$  的所有的值,  $\bar{y}$  的抽样分布是准确的正态分布, 如图 4.23 对于  $n = 5, 10$  和  $25$  所显示的那样。当总体分布是非正态分布时, 如在图 4.24 中所描绘的, 对于小的  $n$ ,  $\bar{y}$  的抽样分布将不再有正态分布的形状(见图 4.24  $n = 5$  的情况)。然而, 当  $n = 10$  和  $25$  时, 抽样分布在形状上接近于正态分布, 这在图 4.24 中可以看到。

已知总体分布的确切形状是不大可能的。因此,  $\bar{y}$  的抽样分布的确切形状也将是未知的。重要的一点是要记住  $\bar{y}$  的抽样分布近似于均值为  $\mu_{\bar{y}} = \mu$ , 即总体均值, 以及标准差为  $\sigma_{\bar{y}} = \sigma/\sqrt{n}$  的正态分布。随着每个样本的样本容量  $n$  的增加, 或者随总体分布的形状向着正态分布的形状变化, 这种近似将更为精确。

一个显然的问题是: 为使中心极限定理成立, 样本容量应为多大? 多年以来进行了众多的模拟研究。总的来说, 这些研究的结果建议中心极限定理当  $n > 30$  时成立。然而, 我们不应盲目地使用这条规则。如果总体分布严重地偏倚, 那么  $\bar{y}$  的抽样分布即使当  $n > 30$  时也将仍然是偏倚的。另一方面, 如果总体是对称的, 中心极限定理当  $n < 30$  时也可用。

因此, 要看一看数据。如果样本的直方图明显地偏倚, 那么总体分布也可能是偏倚的。因而, 要使  $\bar{y}$  的抽样分布近似正态分布, 所需要的  $n$  的值可能要比 30 大得多。此时, 对于在  $n = 30$  时, 基于  $\bar{y}$  的正态性所作出的任何推断都要进行仔细地推敲。

如在图 4.21 和 4.22 中所看到的那样, 中心极限定理可以被推广到许多不同的样本统计量。关于样本中位数和样本标准差的中心极限定理的形式, 在某种程度上比样本均值的中心极限定理的形式更复杂。我们在以后的章节中会遇到许多统计量, 它们是变量的平均值或总和。总和的中心极限定理容易从样本均值的中心极限定理中得到。假定我们从总体中抽出含有  $n$  个测量值  $y_1, \dots, y_n$  的一个随机样本, 令  $\sum y = y_1 + \dots + y_n$ 。

#### 定理 4.2 $\sum y$ 的中心极限定理

设  $\sum y$  表示从总体中抽出的容量为  $n$  的一个随机样本的和, 且总体的均值为  $\mu$ , 总体的标准差  $\sigma$  有限。设  $\mu_{\sum y}$  和  $\sigma_{\sum y}$  分别表示  $\sum y$  的抽样分布的均值和标准差。基于从总体中重复抽取的容量为  $n$  的随机样本, 我们可以得出下列结论:

1.  $\mu_{\sum y} = n\mu$ 。

2.  $\sigma_{\sum y} = \sqrt{n}\sigma$ 。

3. 当  $n$  取大值时,  $\sum y$  的抽样分布近似于正态分布(随着  $n$  的增加这种近似将愈加精确)。

4. 当总体分布是正态分布时, 对任意的样本容量  $n$ ,  $\sum y$  的抽样分布是准确

的正态分布。

一个样本统计量通常用来作为一个总体参数的估计。例如,样本均值  $\bar{y}$  可以用于估计从中抽取样本的总体的均值  $\mu$ 。同样,样本中位数和样本标准差可估计相应的总体的中位数和标准差。因而,一个样本统计量的抽样分布常常用来决定该统计量作为总体参数估计的精度。在例 4.19 中,已知总体均值  $\mu$  是 6.5。显然,在任何实际的研究或试验中我们不知道  $\mu$  的值。然而,对于从总体中抽出的容量为  $n$  的随机样本,我们可以使用  $\bar{y}$  的抽样分布来确定  $\bar{y}$  的值大于比  $\mu$  低 3 个单位的值的概率。当  $n=2$  时,使用例 4.19 中的数据,知这个概率为

$$P(2.5) + P(3) + P(10) + P(10.5) = \frac{4}{45}$$

在一般的情况下,样本统计量的抽样分布很少知道,因而我们根据中心极限定理使用正态分布做近似的计算。第五章将讨论这类计算。由于我们使用样本统计量对一个总体参数进行推断,统计量的抽样分布是决定推断的精确性的关键所在。

**抽样分布**至少可以用两种方法解释。一种方法是用大量重复抽样时的频率来解释。设想从给定的总体中重复抽取固定容量的样本,并且对每个样本计算所考虑的样本统计量的值。在大量的重复抽样中,该样本统计量的可能值的频率将接近于相应的抽样分布的概率。例如,如果从相应于例 4.19 中的概率的总体分布中抽取大量的样本,并对每个样本计算其样本均值,则在这些样本均值的结果中,约有 9% 满足  $\bar{y} = 5.5$ 。

抽样分布的另外一种解释方法是使用概率的古典解释。设想列出从给定的总体抽出的所有的可能的样本。那么,一个样本统计量取特定值的概率(比如说,  $\bar{y} = 5.5$ )对每一个样本计算该统计量时从中得出这一个值的所有可能的样本的比例。在例 4.19 中,  $P(\bar{y} = 5.5) = 4/45$  对应于在 45 个样本中有 4 个样本的样本均值等于 5.5 这一事实。无论用重复抽样方法还是用古典概率的方法求样本统计量的概率都是合理的。

然而,在实践中,一个样本仅仅抽取一次,因而也只计算一个样本统计量的值。抽样分布是不能在实际中看见的;它不是一个经验的、观察到的分布。相反,它是一个理论上的概念,是一个从关于总体和抽样方法的假设中导出的概率的集合。

短语“抽样分布”和“样本分布”字面相近,容易造成误解,然而,它们的意义却迥然不同。“抽样分布”意味着理论上导出的统计量的概率分布,而“样本分布”,是指在一个特定的样本中实际观察到的个体值的直方图。为了避免混乱,我们将样本值的分布叫做**样本直方图**,而不称其为样本分布。

## 练习

## 基本技能

**4.87** 从均值为 60, 标准差为 5 的总体中, 抽出容量为 16 的一个随机样本。描述样本均值  $\bar{y}$  的抽样分布。你预期  $\bar{y}$  的大约 95% 的取值会落在什么样的范围内?

**4.88** 参考练习 4.87。描述样本总和  $\sum y_i$  的抽样分布。 $\sum y_i$  距离 960 超出 70 个单位是不大可能(不可能)的吗? 解释原因。

**4.89** 在练习 4.87 中, 从  $\mu = 60$  和  $\sigma = 5$  的总体中抽取出一个 16 个观察值的随机样本。假设测量值总体的分布是正态分布。使用一个计算机程序, 通过抽取 500 个由 16 个观察值组成的样本来模拟  $\bar{y}$  的抽样分布。模拟时可以使用本章最后一节(4.14)中给出的一个 Minitab 程序。

## 应用

**4.90 (心理学)** 对于一大群狂躁型抑郁症患者, 其精神运动延迟得分是近似于正态分布的, 其中均值为 930, 标准差为 130。

- 得分落在 800 和 1,100 之间的患者的比例是多少?
- 不到 800 的患者的比例是多少?
- 大于 1,200 的患者的比例呢?

**4.91** 参考练习 4.90。

- 对于狂躁型抑郁症患者的精神运动延迟得分的分布。求其第 90 个百分位数。(提示: 在表达式  $z = (y - \mu)/\sigma$  中求出  $y$ , 此处  $z$  表示第 90 个百分位数超过均值  $\mu$  的部分是标准差的多少倍。)
- 求四分位数间距。

**4.92 (社会)** 联邦政府的管理部门已经原则上同意一个门诊医院的建设计划。为了在有限的预算内, 把该医院设计成为能够满足患者就诊需求量要求的设施, 设计人员调查了患者的需求。他们调查了该区域内一家类似的设施, 发现在一个星期期间需要住院的病的人的数量的分布可以用一个均值为 125, 标准差为 32 的正态分布来近似。

- 使用经验准则来描述在一个星期内请求服务的病人的数量  $y$  的分布。
- 如果该设施的容纳能力为 160 患者, 这家医院在一个星期内由于容纳能力所限而不能接待的患者占有多大比例?

**4.93** 参考练习 4.92。为使得前来就医的患者人数超过医院容量的概率是 0.05 或 0.01, 该医院分别应建造多大?

**4.94 (社会)** 基于 1990 年的普查结果, 成年人每天花在看电视的小时数近似于均值为 5 个小时, 标准差为 1.3 小时的正态分布。

- a. 成年人中每天看电视的时间超过 7 个小时的人占多大比例?
- b. 在 1998 年关于电视收视的一项研究中, 由 500 个成年人组成的一个随机样本表明, 每天用在看电视上的平均时间大于 5.5 小时。这项调查的结果与 1990 年的普查结果看起来是一致的吗? (提示: 如果普查的结果仍然是正确的, 那么看电视的平均时间超过 5.5 小时的概率是多少?)

**4.95(环境)** 一种假设的型号(称为普鲁特, Polluter 意为污染源)的汽车, 行驶在城市里时, 其排放的尾气中一种特别的污染物质, 即一氧化氮的水平近似服从均值为  $2.1\text{g/m}$ (克每英里)和标准差为  $0.3\text{g/m}$  的正态分布。

- a. 如果美国环保署(EPA)规定, 一氧化氮的水平不能超过  $2.7\text{g/m}$ , 那么违背这项规定的这种汽车占多大比例?
- b. 找出一氧化氮的一个水平, 使得至多有 25% 的普鲁特车超过这一水平(也就是求 75 百分位数)。
- c. 生产普鲁特车的公司必须降低一氧化氮的水平, 以使得排放超过 EPA 规定的水平  $2.7\text{g/m}$  的汽车至多占 5%。如果标准差仍然是  $0.3\text{g/m}$ , 平均排放水平要减少到多少, 才能使一氧化氮排放超过  $2.7\text{g/m}$  的汽车至多占 5%?

**4.96** 参考练习 4.95。一家公司有一个 150 辆普鲁特车的车队, 这些车供该公司的销售人员使用。试以  $\text{g/m}$  为单位, 描述该车队的排放的尾气中所含的一氧化氮的总量的分布。在这个车队排放的尾气中所含一氧化氮的总量的均值和标准差是多少(以  $\text{g/m}$  为单位)? (提示: 一氧化氮的总量可表示为  $\sum_{i=1}^{150} W_i$ , 其中  $W_i$  是第  $i$  个车辆排出的一氧化氮的数量。由此, 可以利用对总和的中心极限定理。)

**4.97(社会)** 一架飞机的行李限制是每位旅客不超过 100 磅。因此, 对于有 200 个座位的一架飞机, 行李不应超过 20,000 磅。单个的旅客的行李的重量是一个具有均值 95 磅, 标准差 35 磅的随机变量。如果对于某次特定的飞行, 所有的 200 个座位都卖出去了, 那么旅客的行李的全部重量超过 20,000 磅的限制的概率是多少?

**4.98(医学)** 一位病人因为担心她的血压访问其医生。如果血液的收缩压超过 150, 则认为这个病人有高血压并且可以开处方药。问题是在给定的一天内, 一个病人的血液的收缩压的读数有很大的变化。

- a. 如果一个病人在给定的一天内的收缩压的读数服从均值为 160 毫米汞柱, 标准差为 20 毫米汞柱的正态分布, 对该有高血压的病人的一次测量没能检测出其患有高血压的概率是多少?
- b. 如果在一天内不同的时间做 5 次测量, 那么血压读数的平均值小于 150, 并因此没能显示该患者有高血压问题的概率是多少?

- c. 要使得对这位有高血压的病人没能检测出的概率不超过 1%, 需要测量多少次?

### 4.13 二项分布的正态逼近

前面定义的二项随机变量  $y$ , 是在一项随机试验内进行的  $n$  次独立的试验中观察到的成功次数, 其中所有的  $n$  次试验的试验结果是成功(S)或失败(F), 并且  $P(S) = \pi$ 。我们下面将说明, 如何根据总和的中心极限定理, 使用适当的正态曲线作为二项分布的近似, 从而能够计算一个二项随机变量的概率。我们在 4.8 节说过, 对于  $n$  或  $\pi$  为任何值的二项试验, 总能计算出  $y$  取任一值的概率。但是当  $n$  变得大时, 这项任务愈加困难。例如, 假定调查由 1,000 个选民组成的一个样本, 以确定人们对市政府和县政府合并的意见。如果我们假定整个总体的 50% 赞成这一改变, 那么观察到 460 或更少人赞成合并的概率是多少? 这是一个  $n = 1,000$ ,  $\pi$  (挑选到一个赞成合并的人的概率) 等于 0.5 的二项试验。为了确定在 1,000 个选民的随机样本中, 观察到 460 或更少的人赞成合并的概率, 我们可以使用二项式公式来对于  $y = 460, 459, \dots, 0$  计算  $P(y)$ , 然后, 计算所要求的概率

$$P(y = 460) + P(y = 459) + \dots + P(y = 0)$$

这里有 461 个概率需要计算, 其中的每一个概率的计算都因为要计算阶乘而有些困难。例如, 观察到 460 人赞成合并的概率是

$$P(y = 460) = \frac{1000!}{460! 540!} (0.5)^{460} (0.5)^{540}$$

对  $y$  的所有其他的值也需要类似的计算。

为了阐述中心极限定理的使用, 我们需要定义如下  $n$  个随机变量,  $I_1, \dots, I_n$ ,

$$I_i = \begin{cases} 1 & \text{如果第 } i \text{ 次试验结果为成功} \\ 0 & \text{如果第 } i \text{ 次试验结果为失败} \end{cases}$$

二项随机变量  $y$  是在  $n$  次试验中成功的次数。现在, 考虑随机变量  $I_1, \dots, I_n$  的总和  $\sum_{i=1}^n I_i$ 。每当 S 发生时在总和中加 1, 而每当 F 发生在总和中加 0。这样,  $\sum_{i=1}^n I_i$  是在  $n$

试验期间 S 发生的次数。因此, 我们得到  $y = \sum_{i=1}^n I_i$ 。因为二项随机变量  $y$  是独立的随机变量的总和, 其中每一个随机变量有同样的分布, 我们可以对  $y$  使用总和的中心极限定理。于是, 当  $n$  的大小适当时, 正态分布可以用来近似二项分布。所使用的正态分布的均值和标准差由下列公式给出:

$$\mu = n\pi \quad \sigma = \sqrt{n\pi(1-\pi)}$$

这正是二项随机变量  $y$  的均值和标准差。

### 例 4.21

使用二项分布的正态逼近计算, 在 1,000 人的样本里面观察到 460 个或更少的人赞成合并的概率, 假定整个总体中 50% 的人赞成合并。

**解答** 用来近似二项分布的正态分布的均值和标准差分别为

$$\mu = n\pi = 1000(0.5) = 500$$

$$\sigma = \sqrt{n\pi(1-\pi)} = \sqrt{1000(0.5)(0.5)} = 15.8$$

所求的概率在图 4.25 中用阴影区域表示出来。要计算阴影的面积, 我们需要首先计算

$$z = \frac{y - \mu}{\sigma} = \frac{460 - 500}{15.8} = -2.53$$

参考附录中的表 1, 我们发现, 在正态分布曲线下 460 左边 (相应于  $z = -2.53$ ) 的面积是 0.0057。这样, 观察到 460 或更少人赞成合并的概率近似等于 0.0057。

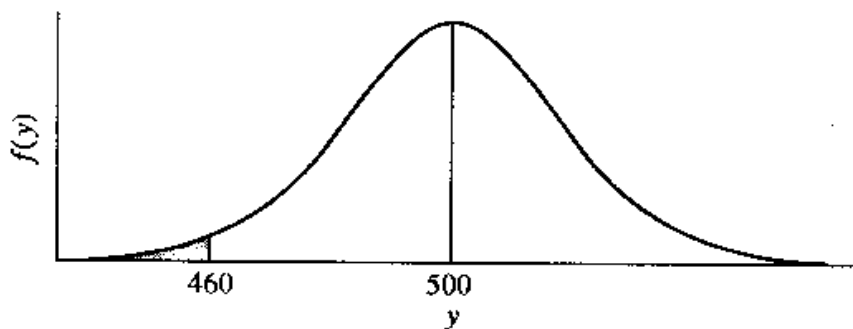


图 4.25 二项分布的正态分布近似, 其中  $\mu = 500$  和  $\sigma = 15.8$

如果  $n\pi < 5$  或  $n(1-\pi) < 5$ , 二项分布的正态逼近可能是不能令人满意的。如果成功的概率  $\pi$  是小的, 并且样本容量  $n$  是中等的, 实际的二项分布严重地向右偏倚。在这种情况下, 对称的正态分布曲线不能给出令人满意的近似。如果  $\pi$  在 1 附近, 从而  $n(1-\pi) < 5$ , 实际的二项分布将向左偏倚, 因此正态分布近似也不很精确。如上例中那样, 当  $n\pi$  和  $n(1-\pi)$  超过大约 20 时, 正态近似是相当好的。在中间的区域, 即  $n\pi$  或  $n(1-\pi)$  在 5 和 20 之间时, 一种称为**连续性校正**的调整使得正态近似的质有显著的提高。

连续性校正的要点是注意到我们是用连续的正态分布的曲线去近似离散的二项分布。图 4.26 显示出了一种我们所面临的情况。

二项分布下  $y \leq 5$  的概率是在 5, 4, 3, 2, 1, 和 0 上方矩形的面积的总和。在做正态近似时, 这一概率 (面积) 是用在 5 的左边正态分布的曲线下方双重阴影的面积来近似的。因而, 这一近似忽略了 5 上面的半个矩形。连续性校正就是把在  $y$

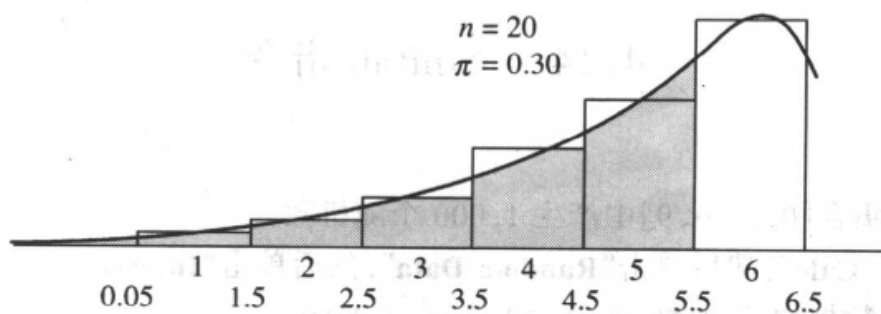


图 4.26 二项分布的正态近似

$y=5$  和  $y=5.5$  之间的面积也包括进去。对于  $n=20$  和  $\pi=0.30$  的二项分布(图 4.26 中所示),连续性校正就是把  $P(y \leq 5)$  换为  $P(y \leq 5.5)$ ,即不是用

$P(y \leq 5) = P[z \leq (5 - 20(0.3)) / \sqrt{20(0.3)(0.7)}] = P(z \leq -0.49) = 0.3121$  而是使用

$$\begin{aligned} P(y \leq 5.5) &= P[z \leq (5.5 - 20(0.3)) / \sqrt{20(0.3)(0.7)}] \\ &= P(z \leq -0.24) = 0.4052 \end{aligned}$$

可以算出,对于该二项分布,概率  $P(y \leq 5)$  实际上为 0.4164。连续性校正的一般的想法是在使用正态分布的概率前,把二项分布的值增加或减去 0.5。决定是增加还是减去的最好的方法是画出如图 4.26 一样的图。

#### 二项概率分布的正态近似

对于大的  $n$  和不太接近 0 或 1 的  $\pi$ ,二项随机变量  $y$  的分布可以由  $\mu = n\pi$  和  $\sigma = \sqrt{n\pi(1-\pi)}$  的正态分布近似。仅当  $n\pi \geq 5$  且  $n(1-\pi) \geq 5$  时,才能使用这一近似。在  $n$  不是特别大的情况下,连续性校正会改进这种近似的精度。

#### 例 4.22

一个大的药品公司有 100 种有潜力的新的处方药正在进行临床试验。到达这个阶段的所有药中大约 20% 最后得到出售许可证。问这 100 种药中至少有 15 种最后得到出售许可证的概率是多少? 这里假定二项分布的假设是满足的。使用有连续性校正的正态近似来计算。

**解答**  $y$  的均值是  $\mu = 100(0.2) = 20$ ; 标准差是  $\sigma = \sqrt{100(0.2)(0.8)} = 4.0$ 。要求的是 15 种或 15 种以上的药品被同意出售的概率。因为包括  $y = 15$ , 连续性校正使用的事件是  $y$  大于或等于 14.5。

$$\begin{aligned} P(y \geq 14.5) &= P\left(z \geq \frac{14.5 - 20}{4.0}\right) = P(z \geq -1.38) = 1 - P(z < -1.38) \\ &= 1 - 0.0838 = 0.9162 \end{aligned}$$



## 4.14 Minitab 指令

### 产生随机数

为了从集合 $[0, 1, \dots, 9]$ 中产生 1,000 个随机数,

1. 单击“Calc”, 然后单击“Random Data”, 然后单击“Integer”。
2. 输入数据的行数: **Generate 20 rows of data**。
3. 输入存储数据的列: **Store in column(s): c1 ~ c50**。
4. 输入列表的第一个数: **Minimum value: 0**。
5. 输入列表的最后一个数: **Maximum: 9**。
6. 单击“OK”。

注意由此我们已经产生了 $(20)(50) = 1,000$  随机数字。

### 计算二项概率

计算二项概率, 其中  $n = 10, \pi = 0.6$ :

1. 在列 c1 中输入  $x$  的值: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10。
2. 单击“Calc”, 然后单击“Probability Distributions”, 然后单击“Binomial”。
3. 选择“Probability”[计算  $P(X = x)$ ] 或“Cumulative probability”[计算  $P(X \leq x)$ ]。
4. 输入  $n$  的值: **Number of trials: 10**。
5. 输入  $\pi$  的值: **Probability of success: 0.6**。
6. 单击“Input column”。
7. 输入  $x$  的值所在的列号: **C1**。
8. 单击“Optional storage”。
9. 输入存储概率的列号: **C2**。
10. 单击“OK”。

### 计算正态概率

计算  $P(X \leq 18)$ , 其中  $X$  服从  $\mu = 23, \sigma = 5$  的正态分布:

1. 单击“Calc”, 然后单击“Probability Distributions”, 然后单击“Normal”。
2. 单击“Cumulative probability”。
3. 输入  $\mu$  的值: **Mean: 23**。
4. 输入  $\sigma$  的值: **Standard deviation: 5**。
5. 单击“Input constant”。
6. 输入  $x$  的值: **18**。
7. 单击“OK”。

### 产生 $\bar{y}$ 的抽样分布

从  $\mu = 60$  和  $\sigma = 5$  的正态分布中抽取 500 个容量为  $n = 16$  的样本,并基于这些样本生成的  $\bar{y}$  的抽样分布:

1. 单击“Calc”,再单击“Random Data”,然后单击“Normal”。
2. 输入样本的个数:Generate 500 rows。
3. 用列号输入样本容量  $n$ :Store in column(s)c1~c16。
4. 输入  $\mu$  的值:Mean:60。
5. 输入  $\sigma$  的值:Standard deviation:5。
6. 单击“OK”。现在在列 c1~c16 中,每列有 500 行,每一行中在列 c1~c16 的 16 个值为一个样本,从这些样本可以得出 500 个  $\bar{y}$  的值。
7. 单击“Calc”,再单击“Row Statistics”,然后单击“Mean”。
8. 输入数据的位置:Input Variables c1~c16。
9. 输入存储 500 个均值的列号:Store Results in c17。
10. 为了获得 500 个  $\bar{y}$  值的均值,单击“Calc”,然后“Column Statistics”,然后“Mean”。
11. 输入 500 个均值所在的位置:Input Variables c17。
12. 单击“OK”。
13. 为了获得 500 个  $\bar{y}$  的值的标准差,单击“Calc”,然后“Column Statistics”,然后“Standard deviation”。
14. 输入 500 个均值所在的位置:Input Variables c17。
15. 单击“OK”。
16. 为了获得  $\bar{y}$  的抽样分布,单击“Graph”,然后“Histogram”。
17. 在图形框中输入 c17。
18. 单击“OK”。

## 4.15 小 结

在这一章中,我们介绍了概率,概率分布和抽样分布。与抽样结果的概率有关的知识对于统计推断是至关重要的。我们给出了基本事件概率的三种不同的解释:古典解释,频率解释和主观的解释。尽管各种解释在统计学中有它的地位,但因为频率的解释方法可以得到验证,所以这个解释方法最有直观感染力。

定量的随机变量被分为两类:离散或连续的随机变量。离散随机变量  $y$  的概率分布是  $y$  的每个值联系的概率  $P(y)$  的一个罗列。这个罗列可以以一张直方图,表或公式的形式出现。

二项分布是一个很重要也很常用的离散的随机变量。科学家们进行的许多试

验类似于投掷硬币的试验,即累积二歧(是或不是)类型的数据。二项试验经常为计算各种各样的抽样结果的概率提供一个良好的模型。

与一个连续的随机变量有关的概率,对应于在概率分布下方的面积。我们通过计算正态曲线下方的面积来说明了这种概率的计算。正态曲线下方的面积的计算之所以重要,是由于中心极限定理:任何一个随机变量,如果它是从有有限标准差的总体中抽出的一个随机样本的总和或平均值,那么当样本容量足够大时,其分布就是正态分布。通过直接应用中心极限定理,我们给出了样本均值的抽样分布。因为许多样本统计量是随机变量的总和或平均值,中心极限定理的使用为我们提供了抽样结果的概率的信息。这些概率对我们想要做的统计推断是至关重要的。

### 重要公式

#### 1. 二项概率分布

$$P(y) = \frac{n!}{y!(n-y)!} \pi^y (1-\pi)^{n-y}$$

#### 2. $\bar{y}$ 的抽样分布

均值:  $\mu$

标准差:  $\sigma_{\bar{y}} = \sigma / \sqrt{n}$

#### 3. 二项分布的正态近似

$$\mu = n\pi \quad \sigma = \sqrt{n\pi(1-\pi)}$$

如果  $n\pi$  和  $n(1-\pi)$  大于或等于 5 或等价地, 如果

$$n \geq \frac{5}{\min(\pi, 1-\pi)}$$

### 补充练习

**4.99(商业)** 一家大的咨询公司审计其消费账目的一种方法,是在每个月的最后一天的所有报告中抽样。讨论如此得到的一个样本是否为一个随机样本。

**4.100(商业)** 一家大型地区性银行,在处理数据的操作过程中,发生的重大键盘输入错误占键盘输入总次数的大约 0.1%。如果考察 10,000 个键盘输入的一个随机样本,求解下列问题:

- 错误数的数学期望。
- 观察到小于 5 个错误的概率。
- 观察到小于 2 个错误的概率。

**4.101** 用  $n=20$  和  $\pi=0.5$  的二项分布,来比较二项分布的正态近似的精度。

- 对  $y < 5$  计算准确的概率和相应的正态近似值。

b. 用  $P(y \leq 4.5)$  可以稍微改进正态近似的精度。为什么有此改进? 比较你的结果。

c. 对  $P(8 < y < 14)$  计算准确的概率和相应的有连续性校正的正态近似值。

**4.102** 设  $y$  是  $n = 20$  和  $\pi = 0.5$  的一个二项随机变量。

a. 计算  $P(4 \leq y \leq 6)$ 。

b. 使用没有连续性校正的正态近似计算同样的概率。比较你的结果。正态近似值的效果如何?

**4.103** 参考练习 4.102。使用有连续性校正的正态近似计算概率  $P(4 \leq y \leq 6)$ 。连续性校正有效吗?

**4.104(商业)** 一家市场调查公司相信, 如果在一个给定的区域邮寄了 5,000 份赛马的报价单, 那么在所有被邮寄给报价单的人中约有 25% 的人回应。

a. 回应 1,000 份或更少的概率是多少?

b. 回应 3,000 份或更多的概率是多少?

**4.105(工程)** 某种合成纤维的 1 平方英尺大小的样本的强度, 通常近似于均值为 2,250psi(磅/每平方英寸)和标准差为 10.2psi 的正态分布。

a. 随机选取这种试验材料的一块 1 平方英尺的样本, 求其强度超过 2,265psi 的概率。

b. 基于 15 块这种试验材料 1 平方英尺大小的随机样本, 描述  $\bar{y}$  的抽样分布。

**4.106** 参考练习 4.105。假定开发了一种新的合成纤维, 其强度的分布可能有不同的均值。抽取 15 块 1 平方英尺的随机样本并且对每块测试其强度。如果我们假设这种新型纤维的强度总体的标准差与原来纤维的总体标准差相同, 在新型纤维强度的分布下给出  $\bar{y}$  的抽样分布的标准差。

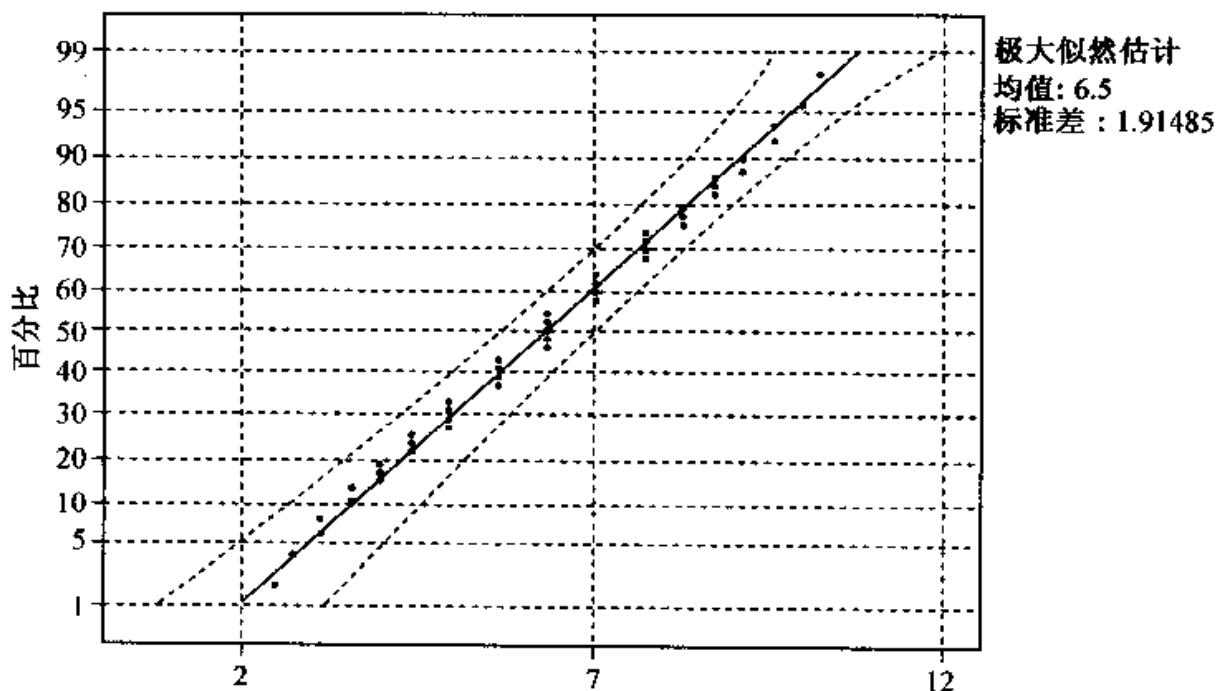
**4.107** 参考练习 4.106。假定新的合成纤维的 15 块 1 平方英尺的样本的强度的均值是 2,268 磅每平方英寸。如果新型纤维的强度的均值与原来的一样, 即为 2,250, 那么观察到  $\bar{y}$  的值等于或大于 2,268 的概率是多少?

**4.108** 在练习 4.107 中的答案的基础上, 你相信新型纤维与原来的纤维有一样的平均强度吗? (假定  $\sigma = 10.2$ 。)

**4.109** 在图 4.19 中, 我们利用视觉考察了基于容量为 2 的样本均值的频率直方图, 并且注意到它具有钟形的形状。另一种确定一组测量值是否服从钟形分布(正态)的方法, 是构造样本数据的一张**正态概率散点图**。如果散点图中的点接近于一条直线, 我们就说这些测量值是从正态总体中抽选的。我们可以使用下列 Minitab 代码, 来产生正态概率散点图。如果散点图中的点落在图中两条弯曲的虚线以内, 我们认为这些数据是从正态分布中抽出的一个随机样本。

Minitab 代码:

1. 把 45 个测量值输入到数据表的 c1 列中。
2. 单击“Graph”，然后单击“Probability Plot”。
3. 在“Variables:”对话框里输入 c1。
4. 单击“OK”。



45 个数据点清楚地落在两条曲线之间，而且相当接近于直线。这样，我们得出结论：正态概率图有力地表明，这些数据是服从正态分布的。

- a. 假定我们的总体是由 10 个测量值(2, 3, 6, 8, 9, 12, 25, 29, 39, 50)组成的。产生 45 个样本，每个样本包含  $n = 2$  个观察值，从这些样本计算得到 45 个样本均值，通过构造 45 个样本均值的直方图和正态概率散点图，确定样本均值的抽样分布是否近似于正态分布。
- b. 为什么你认为，从总体(2, 3, 6, 8, 9, 12, 25, 29, 39, 50)中得到的均值的正态概率散点图，与从总体(2, 3, 4, 5, 6, 7, 8, 9, 10, 11)中得到的均值的正态概率散点图会有很大的不同？

**4.110(人力资源)** 一个工会的学徒挑选审查委员会的记录表明，所有的申请人中有 70% 满足一系列基本的要求而被录取。一个少数派组织的 5 个成员最近经委员会审查，5 个人中的 4 人被拒绝。如果记录确实是 0.70，求至多一个人被录取的概率。当考察少数派组织的 5 个成员时，委员会使用了更低的接受概率吗？

**4.111(政府)** 假定你是美国国税局在一个地区的主管，并且你负责抽取总收入水平在 15,000 美元以上的纳税申报单的 1% 作为样本。对此你怎么做？你要使用随机抽样吗？怎样使用？

**4.112(医药)** 专家认为血清胆固醇的高水平与冠心病的发生率的增加有联系。假定在一个给定的年龄范围内,男性的胆固醇水平的对数服从均值为 2.35,标准差为 1.2 的正态分布。

- 在这年龄范围内的男性血清胆固醇水平大于 250mg/ml 的百分比是多少? (250mg/ml 是临床上正常范围的上限)
- 男性的血清胆固醇水平,能落在临床上的正常范围 150~250mg/ml 以内的占多大比例?
- 如果水平超过 300mg/ml,则被认为是很危险的。在这年龄范围内的成年男性,其血清胆固醇水平超过 300 的比例有多大?

**4.113(商业)** 一家主要的软饮料公司,为了吸引新的顾客,改变了其主要饮料的生产秘方。最近,一家市场调查公司访问了 1,000 名潜在的新顾客,并且给他们品尝了按照新秘方重新生产的饮料,得到了在不久的将来计划购买这种按新方法生产的饮料的个人的数量。

- 确定一个随机变量,来表示感兴趣的  $y$  值的总体。
- 你能计算该随机变量的均值和方差吗?为什么能计算或为什么不能计算?
- 你如何计算  $P(y \leq 250)$ ?

**4.114(商业)** 许多公司正在使用或探索使用电话推销产品技术的可能性——即经由电话来出售他们的产品,以作为许多传统型销售策略的补充。假设一个公司发现,在每 100 次电话推销中大约有 1 次销售成功。

- 求第一次销售成功发生在头 5 次电话推销中的概率。
- 求第一次销售成功发生在 10 次呼叫后的某次呼叫的概率。

**4.115(商业)** 市场分析家认为,一次特定的广告活动将使得成年人总体的至少 20% 注意到广告中的产品。在一个最近的活动以后,抽取的 400 个成年人中有 25 人表明他们看见了广告并且注意到了宣传的新产品。

- 假定总体中 20% 的人通过广告活动注意到了新产品,求观察到  $y \leq 25$  的概率的近似值。
- 基于你的(a)中的答案,你认为广告活动是成功的吗?解释原因。

**4.116(医药)** 一种或一种以上特定的、轻微的出生缺陷以概率 0.0001 发生(即在 10,000 个出生中发生 1 例)。如果在给定的一年里一个给定地理区域有 20,000 个婴儿出生,使用二项分布或二项分布的正态近似,我们能计算出观察到至少一例有轻微的出生缺陷的概率吗?解释原因。

**4.117** 总体由 8 个测量值(2, 6, 9, 12, 25, 29, 39, 50)组成,从中抽出样本容量  $n=4$  的一个随机样本,并计算其均值。求  $\bar{y}$  的抽样分布。(提示:从有 8 个测量值的总体中抽取的样本容量为 4 的样本共有 70 个。)

**4.118** 画出练习 4.117 中的  $\bar{y}$  的抽样分布。

a. 该抽样分布看上去近似于正态分布吗?

b. 证明  $\bar{y}$  的抽样分布的均值等于总体中 8 个值的平均值。

**4.119** 参考练习 4.117。使用同样的总体, 基于容量为  $n = 4$  的样本, 求样本中位数的抽样分布。

**4.120** 画出练习 4.119 中样本中位数的抽样分布的图。

a. 该抽样分布看上去近似于正态分布吗?

b. 计算样本中位数的抽样分布的均值, 并把这个值与总体中位数的值相比较。

**4.121** 从一个均值  $\mu = 100$ , 标准差  $\sigma = 15$  的总体中, 抽出样本容量为 5, 20 和 80 的随机样本。

a. 对于 5, 20 和 80 中的每个样本容量, 给出  $\bar{y}$  的抽样分布的均值。

b. 对于 5, 20 和 80 中的每个样本容量, 给出  $\bar{y}$  的抽样分布的标准差。

c. 在(a)和(b)获得的结果基础上, 关于用样本均值  $\bar{y}$  作为总体均值  $\mu$  的估计的精度, 你的结论是什么?

**4.122** 参考练习 4.121。为了评估样本均值  $\bar{y}$  作为总体均值  $\mu$  的估计的精度, 我们需要知道获得一个远离  $\mu$  的  $\bar{y}$  值的机会有多大。假定样本均值  $\bar{y}$  落在距离总体均值  $\mu$  5 个单位以内的地方这一点是重要的。对于上述三个样本容量中的每一个, 求下列概率, 并评价使用  $\bar{y}$  去估计  $\mu$  的精确度。

a.  $P(\bar{y} \geq 105)$ 。

b.  $P(\bar{y} \leq 95)$ 。

c.  $P(95 \leq \bar{y} \leq 105)$ 。

**4.123** 从一个均值等于 40, 标准差等于 12 的总体中选取  $n = 36$  个测量值的一个随机样本。

a. 描述  $\bar{y}$  的抽样分布。

b. 求  $P(\bar{y} > 36)$ 。

c. 求  $P(\bar{y} < 30)$ 。

d. 如果  $k$  满足  $P(\bar{y} > k) = 0.05$ , 求  $k$  的值。

**4.124** 参考练习 4.123。

a. 描述样本总和  $\sum y_i$  的抽样分布。

b. 求  $P(\sum y_i > 1440)$ 。

c. 求  $P(\sum y_i > 1540)$ 。

d. 如果  $k_1, k_2$  满足  $P(k_1 < \sum y_i < k_2) = 0.95$ , 求  $k_1, k_2$  的值。

**4.125** 对下列每种情况, 基于从均值  $\sigma$  和标准差  $\mu$  的总体中抽出的样本容量为  $n$  的一个随机样本, 求  $\bar{y}$  的数学期望值与标准误差。

- a.  $n = 25, \mu = 10, \sigma = 10$ 。
- b.  $n = 100, \mu = 10, \sigma = 10$ 。
- c.  $n = 25, \mu = 10, \sigma = 20$ 。
- d.  $n = 100, \mu = 10, \sigma = 20$ 。

**4.126** 在练习 4.125 的结果的基础上, 推测增加样本容量, 以及增大  $\sigma$  对于  $\bar{y}$  的标准误差会有什么影响。





## 第五部分 数据分析:中心值, 方差和比例

- 第五章 关于总体中心值的推断
- 第六章 两总体中心值的比较
- 第七章 关于总体方差的推断
- 第八章 两个以上总体的中心值的推断
- 第九章 多重比较
- 第十章 类型数据



## 第五章 关于总体中心值的推断

- 5.1 引言和案例
- 5.2  $\mu$  的估计
- 5.3 估计  $\mu$  时样本容量的选取
- 5.4 关于  $\mu$  的统计检验
- 5.5 对于  $\mu$  进行检验时样本容量的选取
- 5.6 统计检验的显著性水平
- 5.7 正态总体均值  $\mu$  的统计推断,  $\sigma$  未知
- 5.8 关于中位数的推断
- 5.9 小结

### 5.1 引言和案例

推断,具体来说就是决策和预测,已有数百年的历史,且在我们的生活中起着非常重要的作用。我们每个人都面临着日常的个人决策,面临着对未来的预测。美国政府要考虑保持与欧洲和亚洲国家的贸易平衡问题。一个投资顾问想知道未来的6个月是否要发生通货膨胀。冶金学家意欲根据一个实验的结果来决定一种新型的轻合金是否具有制造汽车所必需的强度。兽医欲了解一种新的化学药物的效力,以杀死狗身上的犬恶丝虫。这些单位或个人都是根据相关的事实来作出各自的推断,我们称这些事实为**观察值或数据**。

在很多实际问题中,这些相关事实往往是大量的,似乎还是不自一致的,并且在许多方面,人们对这些事实显得无能为力。结果,往往一个花很大心思作出的决策或预测并不比一个直接的猜测好多少。你只要看一下华尔街杂志(*Wall Street Journal*)中的市场了望部分或有线电视上的某一金融消息,你就会发现专家对未来股票市场行为的观点是千差万别的。同样地,由科学家和工程师根据对数据直观分析所得出的观点往往与实验的结论相互矛盾。

许多人觉得他们自己天生地具有相当好的推断才能。然而,经验告诉我们,大多数人不能利用大量的数据资料,理智地权衡每一点相关信息,得出一个好的推断。(你可以利用第五至第十章的练习检验自己的推断能力。在使用适当的统计方法之前浏览一下数据,作出推断,然后比较所得结果。)统计学家并不是凭借于他或她的直觉,而是根据统计结果作推断。虽然,在前几章中,我们接触到一些有关统计推断的概念,但是现在我们仍需集中思路介绍有关统计推断的基本思想。

统计的目的就是根据包含在样本中的信息对总体进行推断。总体的特征是用

数字进行描述的,称这种特征为参数。典型的总体参数有均值  $\mu$ 、中位数  $M$ 、标准差  $\sigma$  和某一比例  $\pi$ 。大多数推断问题可以归结为对总体的一个或多个参数进行推断。例如,由 Wisconsin(威斯康星)教育部门做的一项研究,欲评估小学阶段孩子的阅读能力。总体是由威斯康星州所有小学各年级孩子参加标准阅读考试的分数所构成的一个整体。我们欲估计总体平均分数  $\mu$  和在总体中分数低于某一标准的学生所占的比例  $\pi$ ,若某一学生的分数低于该标准,则表明这个学生需要一些帮助措施。

参数推断无外乎就是下面两种情况之一,要么估计(预测)总体参数的值,要么对参数值进行检验。这两种统计推断——估计和检验——的方法步骤不同。更重要的是,它们回答了参数的两个不同问题。估计总体参数实际上就是回答这样一个问题,即“总体参数的值是多少”,而检验一个假设就是回答“总体参数是否等于这个特定的值?”

考虑下面一项研究,研究者希望检验一种药物在减轻焦虑病人焦虑程度方面的效力。研究者用筛选(screening procedure)的方法挑选出一组焦虑的病人,在病人被接受参加这项研究之后,测量每个人的焦虑等级,紧接着让其服第一剂药,经一周药物治疗后,再测量他们焦虑的等级。这些数据可用来对样本所来自的总体进行估计或统计检验:

**估计:** 样本中的信息可以用来估计(或预测)所有可以被认可使用这种药物的焦虑患者焦虑程度下降的平均值。

**统计检验:** 样本中的信息可以用来确定焦虑程度下降的总体平均值是否大于 0。

注意,与估计相联系的推断目的在于回答这样的问题:“患者焦虑程度下降的总体平均值是什么?”与此不同,统计检验却是要回答这样的问题,“焦虑程度下降的总体平均值大于 0 吗?”

### 案例:从脂肪中所摄入的卡路里的百分比

对于饮食和许多疾病之间的关系已经有许多研究。例如,从饮食的脂肪中摄入的热量所占的百分比可能与某种类型的癌症和心脏病的发病率有关。要探索饮食与疾病之间的关系,对一个人的日常饮食的评估和量化是很关键的。这是一项很困难的工作,但在监测人们的日常饮食行为的努力中,这是很重要的。Rosner, Willett 和 Spiegelman 在发表的论文“Correction of logistic regression relative risk estimates and confidence intervals for systematic within-persons measurement error”(“在有系统的个人测量误差的情况下,Logistic 回归模型中相对风险的估计和置信区间的修正”,刊登在 *Statistics in Medicine* (1989), 8:1051 - 1070)中,描述了称为“护士健康的研究”的一项研究,其中考察了由很多妇女组成的样本的饮食。

**设计如何收集数据** 研究目的之一就是确定一群妇女从日常饮食的脂肪中所

摄取的卡路里的百分数。有许多饮食评定方法。在很多关于营养的流行病学的研究中,最常用的方法是饮食情况问卷调查法(food frequency questionnaire, FFQ),即通过精心设计的一系列问题来确定被研究者的饮食情况。在护士健康的研究中,从一群女性护士中随机地抽取 168 名护士,进行了一次问卷调查。根据问卷所收集的信息计算出脂肪中所含卡路里的百分数(PCF)。人们感兴趣的是对于由护士组成的这一总体,其 PCF 的平均值  $\mu$ 、标准差  $\sigma$ , PCF 值大于 50% 的护士所占的比例  $\pi$  以及其他一些参数。数据软盘中有包含护士年龄和其他变量的完整的数据。

被调查的样本人数是由估计  $\mu$ 、 $\sigma$  和  $\pi$  的估计精度所确定的。下一节,我们将讨论确定样本容量的几种方法。据此,我们知道:上述容量为 168 的样本将是足够的。

**数据的整理** 研究者需要仔细地审核来自调查问卷的数据,以确定被调查者的响应是否被正确地记录,然后,在分析之前,将数据按照第 2.5 节中的方法转换成计算机文件以备分析之用。

```

1 5
2 0 0 4 4
2 5 5 6 6 6 6 7 7 8 8 8 9 9 9 9 9 9 9
3 0 0 0 0 0 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4
3 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9
4 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 4 4 4 4 4
4 5 5 5 5 5 6 6 6 7 7 8 9 9
5 0 3 4
5 5 7

```

图 5.1 168 名护士的 PCF 数据图

Descriptive Statistics for Percentage Calories from Fat Data						
Variable	N	Mean	Median	TrMean	StDev	SE Mean
PCF	168	36.919	36.473	36.847	6.728	0.519
Variable	Minimum	Maximum	Q1	Q3		
PCF	15.925	57.847	32.766	41.295		

脂肪中的卡路里百分比数据的描述性统计量

**数据分析** 研究中的下一步是通过图和概括统计量来概括数据。168 名护士的 PCF 值用茎叶图表示出来如图 5.1 所示。随后的表则给出了概括统计量。

从茎叶图来看,数据似乎近似服从正态分布,且 PCF 值落在 15% 到 57% 之间。PCF 值大于 50% 的护士所占的比例为  $\hat{\pi} = 4/168 = 2.4\%$ 。根据概括统计量的表,样本均值  $\bar{y} = 36.919$ , 样本标准差  $s = 6.728$ 。研究者欲根据 168 名护士这

一随机样本推断其所属的总体。因此,我们需要将点估计附以边界,以反映总体参数估计的置信度。同时,研究者或许对关于总体 PCF 的平均值  $\mu$  和方差  $\sigma^2$  的假设检验感兴趣。例如,很多营养学家建议人们每天的饮食所摄取的卡路里中,来自于脂肪的卡路里不应超过 30%。因此,我们要检验  $\mu > 30$  这一假设以确定总体的 PCF 平均值是否超过营养学家建议的值。本章,我们将建立总体均值  $\mu$  的区间估计和统计检验。第七章和第十章将分别讨论参数  $\sigma$  和  $\pi$  的估计和检验问题。

## 练习

### 基本技能

**5.1(政治学)** 一名研究人员欲估计在她所在的州中,已登记的、在过去两年中至少参加过一次选举投票的选举人所占的百分比。

- a. 指出研究人员感兴趣的总体。
- b. 你如何选择—个样本以收集所要的信息?

**5.2** 在关于脂肪中所含卡路里的百分数的案例中

- a. 什么是总体?
- b. 除了变量 PCF 外,还有哪些变量影响一个人的健康?
- c. 在护士的健康状况研究中,除了饮食以外,护士还有哪些重要的特征?
- d. 试叙述一种随机选择护士参加试验的方法。
- e. 给出研究人员可能感兴趣的几种假设。

**5.3(工程)** 一家厂商声称某一保险丝的平均寿命是 1,500 小时。根据容量为 35 的保险丝样本得到的信息表明:平均寿命是 1,380 小时。对厂商的断言作何评价?

- a. 指出我们感兴趣的总体。
- b. 对所提问题解决办法是一个估计问题,还是假设检验问题?

**5.4** 参看习题 5.3,你如何从保险丝厂抽取一组样本,以检验厂商的断言?

## 5.2 $\mu$ 的估计

统计推断的第一步就是点估计,即根据样本数据计算出一个值(统计量)作为总体参数的估计。假设我们欲估计一个总体的均值,且认为该总体是正态的。自然地,样本均值可被用来估计总体均值,但我们也可以用样本中位数和截尾均值。那么,究竟使用哪一个样本统计量呢?

数理统计学的一个完整的分支涉及到对各种不同总体的参数建立相应的点估计量(即根据样本数据计算具体点估计的计算公式),以及一个特定的点估计量是否具有某些合意的性质。幸运的是,我们不必推导这些点估计量——每一个参数

的点估计都会被给出。对于所给的参数,我们一旦知道使用那一个点估计(公式),就能建立这些参数的置信区间,即区间估计。

在本节中,我们将讨论总体均值  $\mu$  的点估计和区间估计。第 5.4 节将讨论关于  $\mu$  的假设检验问题。

在本书中,大多数情况下,我们都将用样本均值  $\bar{y}$  估计总体均值  $\mu$ ;同样也用它构造  $\mu$  的区间估计。根据第四章给出的样本均值的中心极限定理,当样本容量  $n$  充分大时(粗略地,  $n \geq 30$ ),  $\bar{y}$  近似于均值为  $\mu$ , 标准差为  $\sigma_{\bar{y}}$  的正态分布。根据经验准则和正态曲线下面积的有关结论,我们知道:在重复抽样时,所有  $\bar{y}$  中有 95% 落在区间  $\mu \pm 2\sigma_{\bar{y}}$ , 或更精确的区间  $\mu \pm 1.96\sigma_{\bar{y}}$  内,见图 5.2。

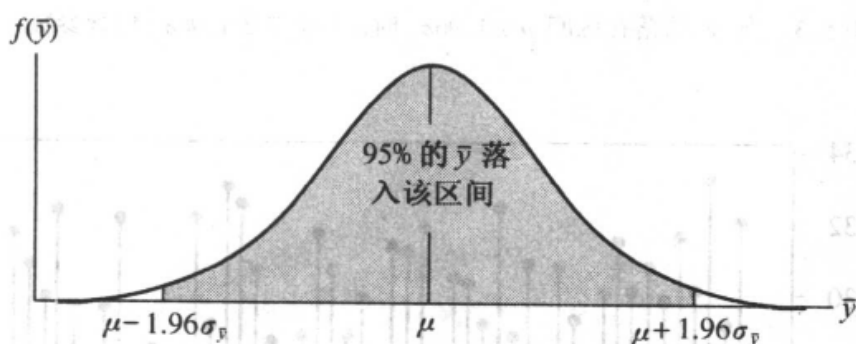


图 5.2  $\bar{y}$  的抽样分布

从图 5.2 可以看出,样本均值  $\bar{y}$  可能并不非常接近于其所要估计的总体均值  $\mu$ 。因此,当得到  $\bar{y}$  值时,我们也应该提供  $\bar{y}$  作为  $\mu$  的估计的精度。为此,考虑  $\mu$  的可能值的区间,而不是仅仅使用单个  $\bar{y}$  值。考虑区间  $\mu \pm 1.96\sigma_{\bar{y}}$ 。一旦  $\bar{y}$  落在区间  $\mu \pm 1.96\sigma_{\bar{y}}$ , 那么区间  $\bar{y} \pm 1.96\sigma_{\bar{y}}$  将包含参数  $\mu$  (见图 5.3)。 $\bar{y}$  落在区间  $\mu \pm 1.96\sigma_{\bar{y}}$  的概率是 0.95, 所以,我们说区间  $\bar{y} \pm 1.96\sigma_{\bar{y}}$  是  $\mu$  的置信水平为 0.95 的区间估计。

为了评价区间估计的好坏,我们在重复抽样的情况下,考察所有区间中那些包含所估计参数的区间的比例,称其为置信系数。区间  $\bar{y} \pm 1.96\sigma_{\bar{y}}$  的置信系数为 0.95, 即在重复抽样时,用  $\bar{y} \pm 1.96\sigma_{\bar{y}}$  计算出的区间中有 95% 包含均值  $\mu$ 。

这种思想可用图 5.4 表示。假设我们考察从虾的生产到销售给饭馆这一商业过程。为监测虾的大小,从池里随机捞出 40 只虾,并测量其长度。我们考虑一次模拟监测。假设池中虾的长度服从均值  $\mu = 27\text{cm}$ 、标准差  $\sigma = 10\text{cm}$  的正态分布。从虾的总体中选出容量为 40 的 50 个样本。由于  $\sigma_{\bar{y}} = \sigma/\sqrt{n}$  (见表 5.1), 对于每一个样本,我们计算其区间估计  $\bar{y} \pm 1.96\sigma_{\bar{y}} = \bar{y} \pm 1.96(10/\sqrt{40})$ 。可以看出,虽然区间的位置不同,但是 50 个区间中只有 2 个区间没有覆盖均值  $\mu$ 。有两个样本得到的区间不包括  $\mu$  这一事实并不表示这种区间估计的方法是失败的。因为置



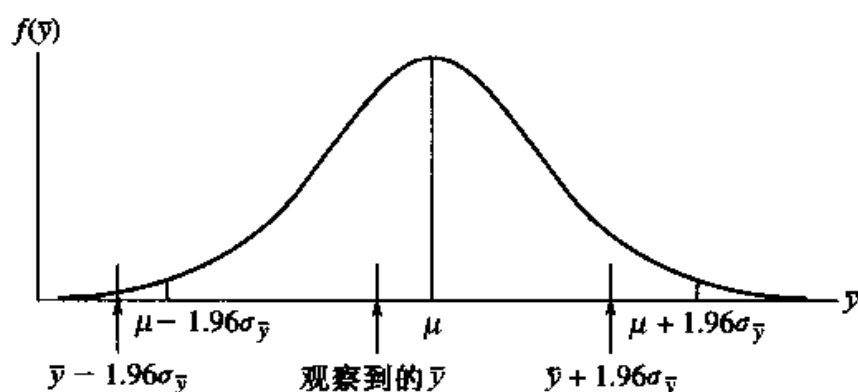


图 5.3 当  $\bar{y}$  值落在区间  $\mu \pm 1.96\sigma_{\bar{y}}$  时, 区间  $\bar{y} \pm 1.96\sigma_{\bar{y}}$  包含参数  $\mu$

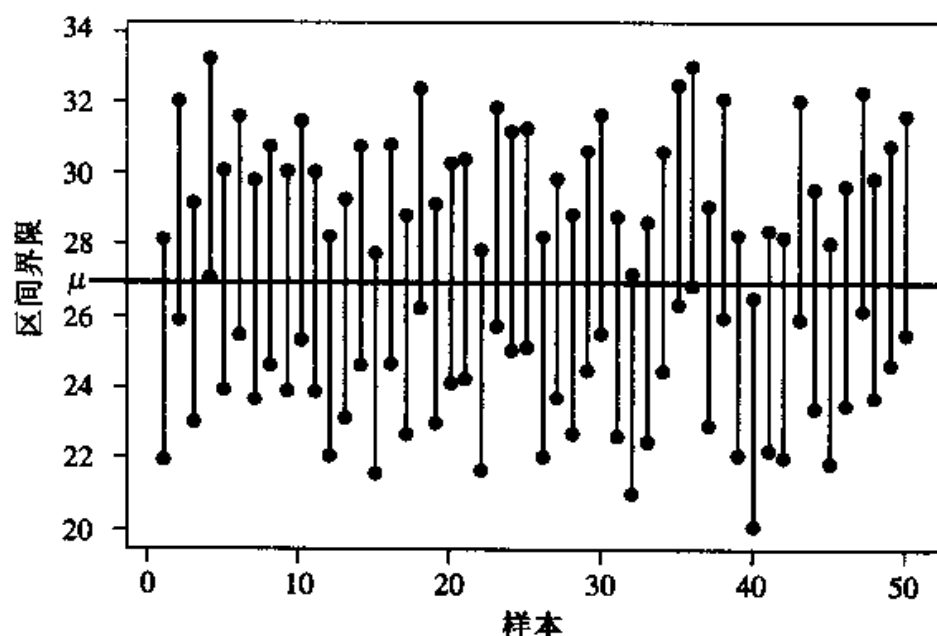


图 5.4 总体均值(27)的 50 个区间估计

信水平为 95%, 所以, 可以预见: 在大量的置信水平为 95% 的置信区间中, 差不多有 5% 的区间不包含  $\mu$ 。因此在 50 个区间中, 有两个或三个 (即 50 中的 5%) 区间不包含  $\mu$ 。关键要理解即使试验按规操作, 但仍然有一些试验所得的结果在某种意义上是有误差的。之所以如此, 是因为我们只做了少量的试验或抽取了容量较小的样本。在我们的例子中, 我们从总体中随机地作了 40 次观测, 然后构造总体均值  $\mu$  的 95% 置信区间。如果大量重复这一过程——例如, 重复 10,000 次而不是 50 次——那么, 不包含  $\mu$  的区间将大约是 5%。

表 5.1 总体均值(27)的 50 个 95%的置信区间

	样本均值	下限	上限	区间包含样本均值
1	25.0080	21.9089	28.1070	是
2	28.8373	25.7382	31.9363	是
3	26.1587	23.0597	29.2578	是
4	30.1301	27.0310	33.2291	否
5	26.9420	23.8430	30.0411	是
6	28.5148	25.4158	31.6139	是
7	26.6456	23.5465	29.7446	是
8	27.6168	24.5178	30.7158	是
9	26.9287	23.8297	30.0278	是
10	28.3338	25.2348	31.4329	是
11	26.9008	23.8017	29.9998	是
12	25.0978	21.9988	28.1969	是
13	26.1866	23.0875	29.2856	是
14	27.6062	24.5072	30.7053	是
15	24.6580	21.5590	27.7570	是
16	27.6427	24.5437	30.7418	是
17	25.7136	22.6146	28.8127	是
18	29.2075	26.1084	32.3065	是
19	26.0411	22.9421	29.1402	是
20	27.1937	24.0947	30.2928	是
21	27.2848	24.1858	30.3838	是
22	24.7320	21.6330	27.8310	是
23	28.8036	25.7046	31.9026	是
24	28.0333	24.9343	31.1324	是
25	28.1065	25.0075	31.2055	是
26	25.1701	22.0711	28.2691	是
27	26.7039	23.6049	29.8029	是
28	25.7587	22.6597	28.8577	是
29	27.4835	24.3844	30.5825	是
30	28.5009	25.4019	31.5999	是
31	25.7142	22.6152	28.8133	是

续表

	样本均值	下限	上限	区间包含样本均值
32	24.0557	20.9567	27.1547	是
33	25.5259	22.4269	28.6249	是
34	27.5036	24.4046	30.6026	是
35	29.3654	26.2664	32.4645	是
36	29.9348	26.8358	33.0338	是
37	25.9826	22.8835	29.0816	是
38	29.0128	25.9138	32.1118	是
39	25.1266	22.0276	28.2256	是
40	23.2452	20.1462	26.3442	否
41	25.3202	22.2212	28.4192	是
42	25.0905	21.9914	28.1895	是
43	28.9345	25.8354	32.0335	是
44	26.4079	23.3089	29.5069	是
45	24.9458	21.8467	28.0448	是
46	26.5274	23.4284	29.6265	是
47	29.1770	26.0780	32.2760	是
48	26.7114	23.6123	29.8104	是
49	27.6640	24.5650	30.7631	是
50	28.5054	25.4063	31.6044	是

## 例 5.1

假设我们考察前面例子中关于脂肪中所含卡路里百分数的数据。为了构造  $\mu$  的置信区间, 需要样本容量足够大, 或者总体服从正态分布。从茎叶图来看, 数据好像服从正态分布。为了进一步证实我们的看法, 我们利用正态概率图(见第四章练习 4.107)来判断总体分布曲线是否真正具有正态的形状。从图 5.5 可见, 所描的点非常接近于一条直线。因此我们相信数据服从正态分布。由于样本容量较大, 我们利用样本标准差  $s$  以代替  $\sigma$  计算均值  $\mu$  的 95% 的置信区间。

**解答** 由数据得

$$\bar{y} = 36.92 \quad s = 6.73$$

利用  $\bar{y} \pm 1.96\sigma_{\bar{y}} = \bar{y} \pm 1.96\sigma/\sqrt{n}$  计算均值  $\mu$  的 95% 的置信区间, 同时以  $s$  以代替  $\sigma$ , 得置信区间为

$$36.92 \pm 1.96 \frac{6.73}{\sqrt{168}} \text{ 或 } 36.92 \pm 1.02,$$

即区间(35.90, 37.94)形成了  $\mu$  的 95% 置信区间。换句话说, 我们有 95% 的把握相信来自脂肪中的平均卡路里含量的百分数位于 35.90 到 37.94 之间。下一步, 研究人员将这些值和其他总体的 PCF 均值或 PCF 建议值进行比较以考察它们的相对大小。

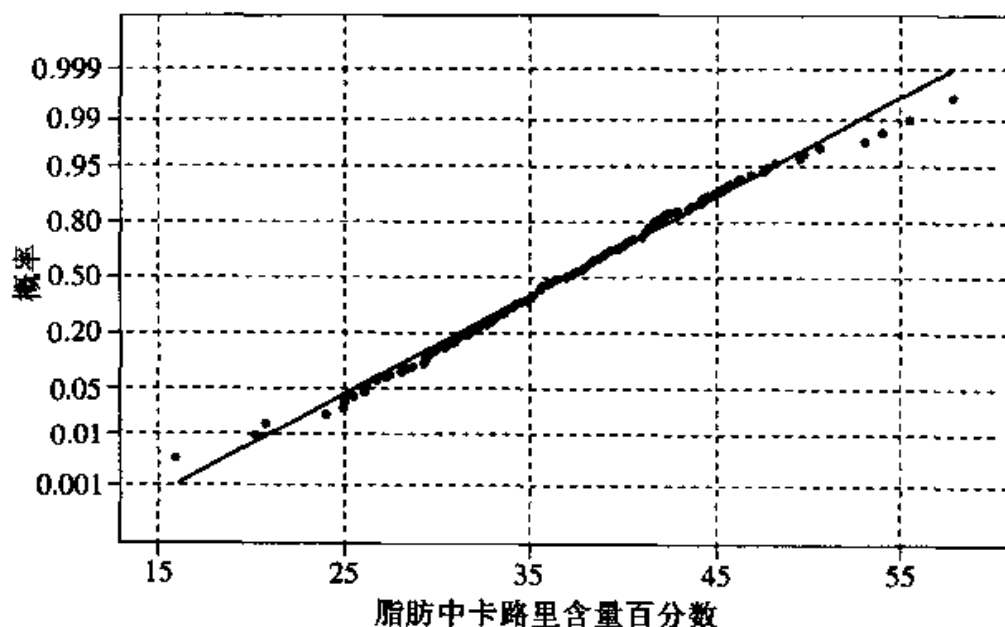


图 5.5 脂肪中卡路里含量百分数(PCF)的正态概率图

在第 5.7 节, 我们将给出当  $\sigma$  未知时构造  $\mu$  的置信区间的方法。然而, 如果样本容量充分大——比如  $n$  大于 30——我们就能在计算置信区间的公式中用样本标准差  $s$  估计总体标准差  $\sigma$ 。另外, 根据中心极限定理, 如果总体分布不是太偏离正态分布, 并且样本容量很大, 那么, 所得区间的置信水平近似等于根据来自正态总体的样本而得到的置信水平。

依据所选置信系数的不同, 有很多关于  $\mu$  的置信区间。例如, 在重复抽样情况下, 区间  $\mu \pm 2.58\sigma_{\bar{y}}$  包含 99% 的  $\bar{y}$  值, 因此区间  $\bar{y} \pm 2.58\sigma_{\bar{y}}$  是  $\mu$  的 99% 置信区间。

我们可以给出  $\mu$  的置信系数为  $1 - \alpha$  的置信区间的通用计算公式, 其中  $\alpha$  在 0 和 1 之间。对于给定的  $1 - \alpha$ ,  $\mu$  的  $100(1 - \alpha)\%$  的置信区间由下式给出, 这里假设  $\sigma$  已知或样本容量充分大以至于可以用  $s$  代替  $\sigma$ 。

$$\bar{y} \pm z_{\alpha/2}\sigma_{\bar{y}}, \quad \text{其中 } \sigma_{\bar{y}} = \sigma/\sqrt{n}$$

$z_{\alpha/2}$  是右侧尾部面积为  $\alpha/2$  所对应的  $z$  值。换言之, 正态曲线下方, 距离  $\mu$   $z_{\alpha/2}$  倍标准差处的右侧面积是  $\alpha/2$ 。在附表 1 中, 查对应面积为  $1 - (\alpha/2)$  的  $z$  值可得到  $z_{\alpha/2}$  值(见图 5.6)。表 5.2 给出了常用的置信系数  $1 - \alpha$  和  $z_{\alpha/2}$  值。

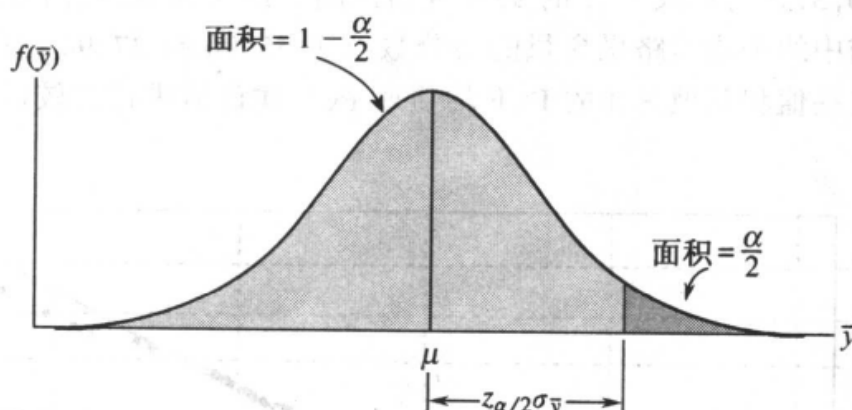


图 5.6 置信区间公式中的  $z_{\alpha/2}$  的图释

表 5.2 常用置信系数  $1-\alpha$  和相应的  $z$  值  $z_{\alpha/2}$

置信系数 $(1-\alpha)$	$\alpha/2$ 的值	表 1 中的面积 $1-\alpha/2$	相应的 $z$ 值 $z_{\alpha/2}$
0.90	0.05	0.95	1.645
0.95	0.025	0.975	1.96
0.98	0.01	0.99	2.33
0.99	0.005	0.995	2.58

### 例 5.2

林业员希望估计 2,000 英亩的某一林场中每英亩的“可计入的树”(大于特定尺寸的树)的平均数量。然后,她能利用这一信息算出林场总的木材蓄积量。随机选择 50 块 1 英亩的林地,并数出“可计入的树”的数量。经计算得每英亩林地可计入树的平均数量为 27.3,标准差为 12.1。根据此信息构造整个林场平均每英亩可计入树数  $\mu$  的 99% 的置信区间。

**解答** 我们利用通用的置信区间公式,其中置信系数等于 0.99,  $z_{\alpha/2}$  等于 2.58(见表 5.2)。代入公式  $\bar{y} \pm 2.58\sigma_{\bar{y}}$ , 且在  $\sigma_{\bar{y}} = \sigma/\sqrt{n}$  中用  $s$  代替  $\sigma$  得  $27.3 \pm 2.58\frac{12.1}{\sqrt{50}}$ , 即  $27.3 \pm 4.41$ ——也就是说,区间从 22.89 到 31.71。因此,我们有 99% 的把握相信每英亩可计入树的平均数量位于 22.89 和 31.71 之间。

在本节的讨论中,我们作了一个相当不现实的假定——即总体标准差已知。实际上,很难找到一种情况,其总体的均值未知,但标准差已知。通常情况下,均值和标准差都要根据样本进行估计。因为用样本标准差  $s$  估计  $\sigma$ , 所以均值的实际

标准误差,  $\sigma/\sqrt{n}$ , 自然用  $s/\sqrt{n}$  进行估计。这一估计又产生另一随机误差源( $s$  围绕着  $\sigma$  从一个样本到另一个样本随机地变化), 而且, 严格来说, 使得我们的置信区间公式变得无效。幸运的是, 对于大样本来说, 前面的区间估计公式仍是一个很好的近似。大体说来, 当  $n$  大于 30 时, 我们能够利用这一公式; 第 5.7 节将介绍一种更好的处理这一问题的方法。

统计决策方法不同于普通的决策方法, 因为它不仅进行推断, 而且还告知推断的好坏情况。对于区间估计, 置信区间的宽度和置信系数表明了其估计的优劣。对于给定的置信系数, 区间宽度越窄, 推断越精确。另一方面, 置信系数是由试验人员给定的, 以体现他或她要以多大的把握保证区间估计包含所感兴趣的参数。当样本容量一定时, 提高置信水平将导致区间变宽。因此, 试验人员一般要求一个合意的置信水平和区间宽度。下面我们将讨论如何确定样本容量以满足这些要求。

## 练习

### 基本技巧

**5.5(工程)** 一经营谷物的公司每 10 分钟随机选择 25 箱装玉米片的 12 盎司盒子, 并称其重量。假设重量服从  $\sigma = 0.2$  盎司的正态分布。根据某一样本得到  $\bar{y} = 12.3$  盎司。

- 计算在抽取样本期间所生产的包装箱的平均重量  $\mu$  的 95% 置信区间。
- 不用统计术语, 对置信区间作出小心的解释。

**5.6** 谷物公司从事加工的工程人员关心  $\mu$  的置信区间不要太宽而失去了实用价值。

- 如果样本容量从 25 加倍到 50, 那么对 95% 置信区间的宽度有何影响?
- 如果将置信水平从 95% 提高到 99%, 那么对置信区间的宽度有何影响?

**5.7** 公司正在每 10 分钟收集包含 25 箱的样本, 那么, 5 天时间就得到 720 个置信区间。

a. 如果 720 个置信区间中每个区间的置信水平都是 95%, 那么, 你预计有多少区间是错误的——即不包含  $\mu$ ?

b. 如果样本容量从 25 增加到 50, 且仍然每 10 分钟抽取一个样本, 那么, 你预计 5 天中有多少个 99% 置信区间是错误的?

c. 如果每个样本的样本容量仍保持为 25, 但置信水平从 95% 增加到 99%, 那么, 你预计 5 天中有多少个 95% 置信区间是错误的?

## 应用

**5.8(工程)** 现有一台新机器分发黑咖啡, 从中随机选取 50 杯黑咖啡(单位: mg), 测量它们的咖啡因含量。均值和标准差分别是 110mg 和 7.1mg。试用这些

数据构造由该机器分发的平均每杯咖啡的咖啡因含量  $\mu$  的 98% 的置信区间。

**5.9(商业)** 某城市的商会欲估计他们城市小公司(销售额在 500,000 美元以下)的毛利润。根据 10 个小公司组成的随机样本所得的年终报告表明:销售的平均毛利润是 5.2%, 标准差为 7.5%。

- 构造该城市所有小公司平均毛利润  $\mu$  的 99% 置信区间。
- 在使用(a)中构造的置信区间时有哪些限制? 例如, 由于样本容量较小, 你认为数据来自于正态分布总体吗? 用  $s$  代替  $\sigma$  有效吗?

**5.10(工程)** 随着大量充气站从全方位服务向自我服务的转移, 某消费者组织担心许多汽车在充气不足的状态下行驶。这样会导致大量的轮胎磨损, 汽车转向和刹车的安全隐患。如果一个轮胎压力低于推荐水平 10psi 以上, 那么就被认为是严重充气不足。随机抽取 400 部汽车, 其平均充气不足量是  $\bar{y} = 10.4\text{psi}$ , 标准差  $s = 4.2\text{psi}$ 。

- 构造平均充气不足量  $\mu$  的 99% 置信区间。
- 基于你的置信区间, 你是否推荐该消费者组织发表一份报告, 宣称平均轮胎压力严重不足? 解释你的答案。
- 如果决策是用 90% 置信区间, 那么, 你在(b)中的答案改变吗?

**5.11(社会学)** 某社会工作者欲估计那些初犯从监狱释放出来后到第二次犯罪并重新被投入监狱的犯人在狱外的平均时间。随机从县法院抽取  $n = 150$  份监狱记录, 表明第一、第二次犯罪之间监外生活的平均时间为 3.2 年, 标准差为 1.1 年。用该样本信息估计县法院所有记录在册的犯人在第一、第二次犯罪之间监外生活的平均时间  $\mu$ , 构造  $\mu$  的 95% 的置信区间。假设能够用  $s$  代替  $\sigma$ 。

**5.12(商业)** 纽约市某快递公司声称它投递到该城市任何一个地方的投递时间不足 3 小时。为了评价其说法, 质量控制人员随机选取了 50 件投递业务, 算得其平均投递时间为  $\bar{y} = 2.8$  小时, 标准差  $s = 0.6$  小时。

- 用 95% 的置信区间估计平均投递时间  $\mu$ 。
- 根据 95% 的置信区间, 公司的说法听起来合理吗?
- 如果利用  $\mu$  的 99% 的置信区间, 你在(b)中的答案改变吗?

**5.13(农业)** 赤褐色螨虫(rust mite)是佛罗里达州危害柑橘的主要害虫, 它刺穿树叶和果实的细胞。由于受害柑橘, 呈褐色, 并且随着受害程度而相应变小, 因此, 人们很容易知道柑橘是否受到损害。如果这种螨虫得不到控制, 那么受侵害果园的柑橘产量和质量都会大大地下降。无论哪种情况发生, 柑橘种植者都会遭受经济损失, 因为这种柑橘的等级较低, 在新鲜水果市场上只能低价出售。今年, 越来越多的柑橘种植者给柑橘喷洒农药, 养护柑橘, 以防螨虫的侵害。为了评价这项措施, 从 60 个果园的每一个果园中随机地选取一块 10 英亩的果园地作为样本。样本数据显示平均每 10 英亩果园地的产量为 850 箱, 标准差为 100 箱。对于所有

喷过药的果园,试给出平均每 10 英亩果园的产量  $\mu$  的 95% 的置信区间。假定可以用  $s$  代替  $\sigma$ 。

**5.14(农业)** 对于种类繁多的柠檬树,现做一个实验,以检验其根茎对某种幼虫的易感染性。现有 40 棵柠檬树遭受这种幼虫的侵害,经过一段时间以后,检验它们的易感染性。我们感兴趣的是每克根茎所含幼虫数目的对数。对于这 40 棵树,样本均值为 9.02,样本标准差为 1.12。试用这些数据构造  $\mu$  的 90% 的置信区间,其中  $\mu$  表示样本所来自的柠檬树根茎总体对幼虫的平均易感染性。假定可以用  $s$  代替  $\sigma$ 。

**5.15(教育)** 从某一州过去 10 年毕业的高中生中随机抽取 900 名作为样本,研究高中生的流动性。对于抽取的每一名学生,记录其高中学校与其目前的永久居住地之间的距离。由这些数据得  $\bar{y} = 430$  英里,  $s = 262$  英里。利用 95% 置信区间,估计该州过去 10 年高中毕业生每个学生的高中学校与其目前的永久居住地的平均英里数。假定可以用  $s$  代替  $\sigma$ 。

**5.16(政府)** 美国、其他政府机构以及关心世界范围内食品严重短缺的一些世界理事会,他们感兴趣的问题就是寻找一种方法以估计在某一特定年份世界范围内所生产的谷物的总产量。

一种方法就是根据地球表面的卫星照片预测谷物的总收成。由于扫描设备读出的某种谷物的总种植面积具有一定的误差,因此,有必要对种植某种谷物且具有相同面积的多块土地,记录下其用这种扫描设备测出的产量,以校正这种扫描仪的误差。用扫描仪读出 100 块 50 英亩的小麦地的卫星照片,得到样本平均值和标准差为

$$\bar{y} = 3.27 \quad s = 0.23$$

对于所有 50 英亩小麦地构成的总体,求平均每块地收成的 95% 的置信区间。解释所得区间的含义。

**5.17(农业)** 另一个农业问题就是关于蛋白质的产量,蛋白质是人类和动物饮食的重要成分之一。虽然众所周知,谷物和豆类中蛋白含量较高,但是少有人知道某些草还提供丰富的蛋白质资源。例如,以重量计,百慕大(Bermuda)草中蛋白含量约占 20%。为验证这些结果,现进行一项研究,分析 100 根重 1 磅的小草中蛋白含量。样本均值和标准差为

$$\bar{y} = 0.18 \text{ 磅} \quad S = 0.08 \text{ 磅}$$

利用 95% 置信区间,估计百慕大草总体平均每磅的蛋白质含量,并解释所得区间的含义。



### 5.3 估计 $\mu$ 时样本容量的选取

我们如何确定样本中观察的个数?这一问题的意义是显而易见的。收集数据需要成本。样本过大会浪费时间和才智。相反,如果样本过小,由于花费时间和精力而得到的是不充分的信息,因此这也是一种浪费,并且有时不可能在后来再增加样本容量。所以,样本容量的确定既要考虑作为总体参数的样本估计统计量的期望精度又要兼顾为了得到这一精度所花费的时间和成本。

在第 5.1 节介绍的饮食研究中,研究人员必须确定调查多少名护士才能得到切实可行的结论。为了确定抽取多少名护士,我们必须确定研究人员究竟要以多高的精确度来估计脂肪中卡路里含量百分数(PCF)的平均值。如果研究人员规定样本估计值和总体均值  $\mu$  的偏离应在 1.5 以内,那么我们就是要  $\mu$  的置信区间为  $\bar{y} \pm 1.5$ 。或者说,如果研究人员规定估计的容许误差是 3,那么我们得到的区间同样为  $\bar{y} \pm 1.5$ ,因为容许误差就是置信区间的宽度。

为了估计  $\mu$  的置信区间,确定合适的样本容量时需要考虑两个方面。首先,容许误差决定了区间宽度;其次就是置信水平。在选择区间的范围时,我们需要考虑是否置信区间太宽,以至于  $\mu$  的估计不精确,不能提供很多的信息。同样地,在一个很低的置信水平(如 50%)下所得到的置信区间很可能是错误的,即区间不包含  $\mu$ 。然而,为了获得一个宽度较窄、置信水平较高的置信区间,样本的容量就要足够大,因此,就时间或成本来讲,这是不合理的。

什么是合理的可信度呢?大多数情况下,置信水平被设定为 90% 或 95%,部分是因为传统习惯,部分是因为对某些人来讲,这些水平代表了一个合理的可信度水平。从长期来看,95%(或 90%)说明 20 次中有 1 次(或 10 次中有 1 次)置信区间可能不包含总体参数。这似乎是合理的,也是可理解的,而 1,000 次中有 1 次或 10,000 次中有 1 次这样的机率太小了。

容许误差与所考虑的问题密切相关,只有那些对问题非常熟悉的人才能对容许误差的大小作出合理的判断。

当考虑总体均值的置信区间时,其置信区间的加减项为  $z_{\alpha/2}\sigma_{\bar{y}}$ ,其中  $\sigma_{\bar{y}} = \sigma/\sqrt{n}$ 。决定加减项的有三个量:期望的置信水平(决定了  $z$  值),标准差( $\sigma$ ),以及样本容量(和  $\sigma$  共同决定标准误差  $\sigma_{\bar{y}}$ )。通常情况下需要对总体的标准差进行估测。(有时,需要根据初始样本估计标准差;这种估计为进一步确定所需另外的样本容量奠定基础。)给定了容许误差和置信水平,以及  $\sigma$  的估计,则样本容量可通过下面的公式进行计算。

假定我们欲用容许误差为  $W$  的  $100(1-\alpha)\%$  置信区间来估计  $\mu$ ,则置信区间形如  $\bar{y} \pm E$ ,其中  $E = W/2$ ,  $W$  是置信区间的宽度。为了确定样本容量  $n$ ,关于  $n$

解方程

$$E = z_{\alpha/2} \sigma_{\bar{y}} = z_{\alpha/2} \sigma \sqrt{n}$$

得到  $\mu$  的形如  $y \pm E$  的  $100(1 - \alpha)\%$  置信区间所需要的样本容量为

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$$

为了估计  $\mu$ , 确定样本容量需要知道总体方差  $\sigma^2$  (或标准差  $\sigma$ )。用下面两种方法之一, 通过估计  $\sigma^2$  可得到样本容量的近似值:

1. 根据以前的试验数据计算样本方差  $s^2$ 。可以用它作为  $\sigma^2$  的近似估计值。
2. 利用观测值的极差得到  $\sigma$  的估计。

在样本容量的等式中, 将  $\sigma^2$  用其估计值代替, 可得到  $n$  的近似值。下面我们举两个例子来说明如何选择样本容量。

### 例 5.3

在饮食的营养摄取量例子中, 研究者要估计容许误差为 3 的平均 PCF 的 95% 置信区间。由以往的研究得知, PCF 值在 10% 到 50% 之间。研究者究竟要抽取多少名护士才能达到他们的要求?

**解答** 由于 95% 置信区间的宽度为 3, 因此  $E = 3/2 = 1.5$ , 且  $z_{\alpha/2} = z_{0.025} = 1.96$ 。 $\sigma$  估计为 极差/4 =  $(50 - 10)/4 = 10$ 。代入  $n$  的公式中, 得

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2} = \frac{(1.96)^2 (10)^2}{(1.5)^2} = 170.7$$

因此, 若 10 是  $\sigma$  的一个合理估计, 那么, 171 名护士的随机样本应该给出宽度为 3 的  $\mu$  的 95% 置信区间。

### 例 5.4

一联邦机构决定调查标明在某品牌谷物包装箱上谷物的重量。涉及到的公司定期从生产线上抽取装有谷物的纸板箱, 该联邦机构总共得到 1,500 个重量数据, 据此算得平均每箱重 11.80 盎司, 标准差为 0.75 盎司。根据这一信息, 为了估计现在生产的平均每箱的重量, 联邦机构必须检查多少箱谷物, 已知 99% 置信区间的宽度为 0.50。

**解答** 联邦机构规定置信区间的宽度为 0.50, 所以  $E = 0.25$ 。假定公司提供的重量是精确的, 于是可取  $\sigma = 0.75$ 。由于  $z_{\alpha/2} = 2.58$ , 因此样本容量为

$$n = \frac{(2.58)^2 (0.75)^2}{0.25^2} = 59.91$$

由此,联邦机构必须抽取 60 箱谷物所构成的一个随机样本,才能使得估计平均重量的误差在  $\pm 0.25$  之内。

### 练习

#### 基本技能

**5.18** 参见例 5.3, 假设我们继续估计  $\sigma$ , 得  $\hat{\sigma} = 10$ 。

- 如果置信水平仍然是 95%, 但容许宽度为 2, 那么需要多大的样本?
- 如果置信水平增加到 99%, 但规定的宽度仍为 3, 那么需要多大的样本?
- 如果置信水平减少到 90%, 但规定的宽度仍是 3, 那么样本容量需要多大?

**5.19** 根据你在 5.18(a), (b) 和 (c) 中所得到的结果。

- 若减少置信水平, 区间宽度不变, 那么样本容量如何?
- 若增加置信水平, 区间宽度不变, 那么样本容量如何?
- 若减少区间宽度, 置信水平不变, 那么样本容量如何?

**5.20** 一般地, 如果保持置信水平不变, 那么样本容量增加多少, 才能将区间宽度缩短一半?

#### 应用

**5.21(生物)** 生物学家欲估计抗生素对某种特殊细菌生长的影响, 为此将固定量的抗生素涂于细菌培养片上, 然后观察平均每个培养片上出现的细菌量。以前的试验表明所出现的细菌量的标准差近似为  $13\text{cm}^2$ 。利用这一信息确定所要观察值(即进行培养然后检测)的数量, 以便使用 99% 的置信区间估计平均细菌量, 且区间半长为  $3\text{cm}^2$ 。

**5.22(社会)** 市房管部门想估计其管理的公寓的平均租金。为了用 95% 置信区间估计平均租金, 且偏差在 50 美元之内, 他们必须确定其调查所需抽取租户的数量。根据过去的结果, 其管理的公寓的租金平均每月在 200 美元到 1500 美元之间。需要调查多少户才能满足其要求?

**5.23** 参见练习 5.22。假定市长重新考虑所提议的调查, 然后决定进行如下变动:

- 如果置信水平增加到 99%, 平均租金估计偏差在 25 美元之内, 则样本容量为多少?
- 假定项目预算并不支持置信水平的提高和区间宽度的减少。你如何向市长解释不把置信水平由 95% 提高到 99% 对平均租金估计所造成的影响?

**5.24(商业)** 保险公司关心由于食品杂货店装袋工后背受伤所造成的工人索赔的数量。为此, 他们想估计向他们投保的杂货店中装袋工的工作能力, 即: 考察工人在不产生后背重压下所能提起的货物重量。根据其他保险公司的研究,  $\sigma \approx$

25 磅。

- a. 需要调查多少名工人才能使得工人提起的平均重量估计的置信水平达到 99%, 偏差在 8 磅之内?
- b. 如果  $\sigma$  未知, 试提出保险公司能够获得  $\sigma$  值的几种方法。

## 5.4 关于 $\mu$ 的统计检验

统计决策方法的第二种形式就是统计检验(或假设检验)。正如估计一样,我们要对总体的参数进行推断,但这里的推断不尽相同。就点估计和区间估计而言,在收集数据之前,并不对参数的真值进行假定。利用来自总体的样本数据,我们就可简单地确定参数值。在假设检验中,对总体参数有一个先验的思想。例如,为了研究某试验性复方药中安定药的特性,我们也许会问:服过特定剂量复方药的老鼠的平均避免休克反应(shock-avoidance response)值是否大于 60, 即  $\mu > 60$ , 其中 60 为使用合适的标准药进行广泛试验所观察到的值。因此,包含在统计研究中的有两种理论或假设。第一种就是进行试验的人员所提出的假设,称之为**研究假设(research hypothesis)**,在上述例子中即  $\mu > 60$ 。第二种就是研究假设的反面,称之为**零假设**,在上述例子中即  $\mu \leq 60$ 。研究的目的是要确定数据是否支持研究假设。

**统计检验**建立在反证法基础之上,由下面 5 个部分组成。

1. 研究假设(也称备择假设), 记为  $H_a$ 。
2. 零假设, 记为  $H_0$ 。
3. 检验统计量, 记为  $T.S.$ 。
4. 拒绝域, 记为  $R.R.$ 。
5. 核对假定, 并得出结论。

例如, 德克萨斯的 A&M 农业服务部门想确定今年某一特定品种大豆的平均每英亩产量(单位: 蒲式耳)是否比前两年有所增加, 其中前两年平均每英亩产量  $\mu = 520$  蒲式耳。建立统计检验的第一步就是要确立适当的假设  $H_0$  和  $H_a$ 。下面的原则将是有益的:

1.  $\mu$  等于某一特定值这一假定总是包含在  $H_0$  之中。 $\mu$  所取的特定值被称之为零值, 记为  $\mu_0$ 。
2. 研究人员根据观测数据企图支持或探索的关于  $\mu$  的假设是研究假设  $H_a$ 。
3.  $H_a$  的对立假设是零假设  $H_0$ 。
4. 零假设被认为是正确的, 除非数据中有足够充分的证据支持研究假设。

在我们的例子中,  $\mu_0$  为 520。研究假设就是今年产量大于 520; 即  $\mu > 520$ 。(注意: 520 包含在零假设中。)因此, 零假设, 即  $H_a$  的对立假设为  $H_0: \mu \leq 520$ 。

为了评价研究假设,我们从样本中挖掘信息,试图确定数据是支持研究假设还是支持零假设,当我们没有充分证据否定零假设时,就承认零假设是正确的。

在建立了零假设和研究假设之后,我们从分布在全州的农田中得到 1 英亩产量的样本数据。判定数据是否支持研究假设要根据通过样本数据所算得的某个量而定,称这个量为**检验统计量**。如果总体分布是丘形分布,则关于  $\mu$  的一个合理的检验统计量应是  $\bar{y}$  或  $\bar{y}$  的函数。

如果我们选择  $\bar{y}$  作为检验统计量,那么我们知道:在总体服从正态分布或者样本容量充分大的情况下,  $\bar{y}$  近似服从均值为  $\mu$ , 标准差  $\sigma_{\bar{y}} = \sigma/\sqrt{n}$  的正态分布。我们试图在  $H_a: \mu > 520$  和  $H_0: \mu \leq 520$  之间作出选择,要么拒绝假设  $H_0$ , 要么不拒绝假设  $H_0$ 。在作出判决的过程中,我们假设  $\mu = 520$ , 其中 520 为  $\mu$  的零值。现在我们要决定,  $\bar{y}$  的哪些值在  $\mu = 520$  (或如果  $\mu$  取  $H_0$  中其他值)时不太可能观测到,这些值的全体称为**拒绝域**。拒绝域包含支持研究假设,并和零假设相抵触的那些  $\bar{y}$  值,因此,一旦  $\bar{y}$  值落在这一区域,则拒绝零假设。拒绝域就是  $\bar{y}$  的零分布( $\mu = 520$ )的右侧尾部区域,见图 5.7。

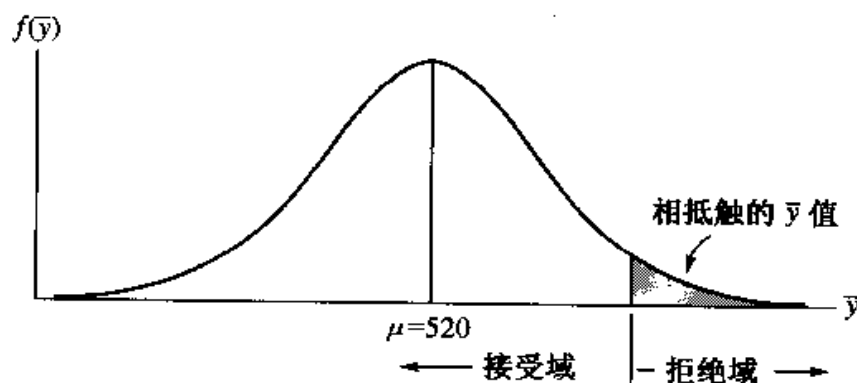


图 5.7 假设  $H_0$  为真,相抵触的  $\bar{y}$  值在右侧尾部

正如任何双向判决过程一样,我们也可能犯错误,即错误地拒绝零假设,或错误地接受零假设。我们特别地称它们为**第一类错误**和**第二类错误**。

**定义 5.1 第一类错误** 是指当零假设为真时拒绝零假设所犯的误差。犯第一类错误的概率用  $\alpha$  表示。

**定义 5.2 第二类错误** 是指当零假设不成立时接受零假设所犯的误差。犯第二类错误的概率用  $\beta$  (希腊字母 Beta) 表示。

表 5.3 列出了这个双向判决过程及相应的概率。

表 5.3 双向判决过程

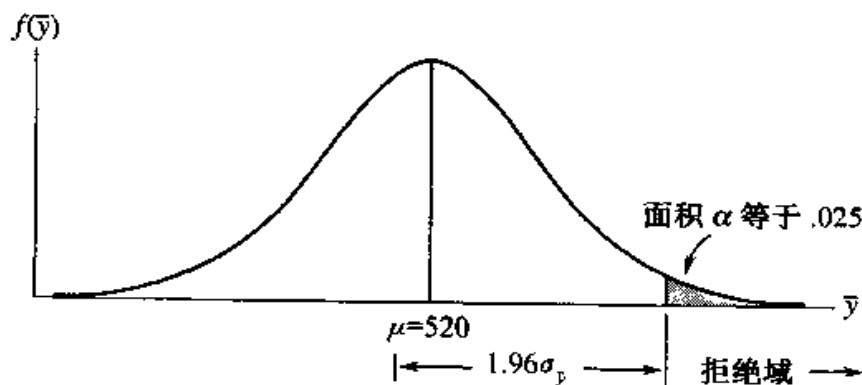
判决	零假设	
	真	假
拒绝 $H_0$	第一类错误 $\alpha$	正确 $1 - \beta$
接受 $H_0$	正确 $1 - \alpha$	第二类错误 $\beta$

虽然希望确定接受域和拒绝域使得  $\alpha$  和  $\beta$  同时达到最小,但这是不可能的。犯第一类错误的概率和犯第二类错误的概率之间存在负向关系。样本容量一定时,当我们改变拒绝域以增加  $\alpha$  时,  $\beta$  值减小,反之亦然。

为了减缓这种难受的处境,试验者事先给定统计检验犯第一类错误的概率。因此,试验者选择  $\alpha$  为 0.01, 0.05, 0.10 等。根据给定的  $\alpha$  值,然后确定拒绝域。计算犯第二类错误的概率相当复杂,本章后面部分将作介绍。

现在来看看如何根据  $\alpha$  的选择确定拒绝域。回到大豆的例子中,当样本均值  $\bar{y}$  较大时,拒绝零假设。假定我们取  $n = 36$  的一个 1 英亩农田样本,并据此算得  $\bar{y} = 573, s = 124$ 。我们能够得出结论,认为平均每亩农田产量超过 520 吗?

在回答这一问题之前,我们必须给定  $\alpha$  的值。如果我们愿意在 40 次中有 1 次冒险错误地拒绝零假设的话,那么  $\alpha = 1/40 = 0.025$ 。对于  $\alpha$  的这一取值,根据  $\bar{y}$  的样本分布即可确定相应的拒绝域。假设  $\mu = 520, \sigma$  可用  $s$  代替,那么  $\bar{y}$  服从  $\mu = 520, \sigma_{\bar{y}} = 124/\sqrt{36} = 20.67$  的正态分布。因为,图 5.8 中标出的区域面积相当于  $\alpha$ ,所以确定拒绝域使之落在  $\bar{y}$  分布右侧尾部,且面积等于 0.025 就相当于确定  $z$  值使之右侧面积等于 0.025。查附录中表 1,得  $z$  为 1.96。因此,本例中拒绝

图 5.8 当  $\alpha = 0.025$  时大豆例子中的拒绝域

域就是均值  $\mu = 520$  加 1.96 倍标准差 ( $1.96\sigma_{\bar{y}}$ ) 处的右侧区域。如果  $\bar{y}$  的观测值大于均值  $\mu = 520$  加 1.96 倍标准差, 则拒绝零假设, 如图 5.8 所示。

### 例 5.5

写出关于大豆的统计检验的各部分内容, 并根据样本数据判断接受还是拒绝零假设。令  $\alpha = 0.025$ , 并假设可用  $s$  估计  $\sigma$ 。

**解答** 检验的 5 部分内容如下:

$$H_0: \mu \leq 520$$

$$H_a: \mu > 520$$

$$\text{T.S.: } \bar{y}$$

R.R.: 对于  $\alpha = 0.025$ , 若  $\bar{y}$  大于均值 ( $\mu = 520$ ) 加 1.96 倍标准差, 则拒绝零假设。

经计算, 得  $\bar{y}$  等于 573。为了计算  $\bar{y}$  位于  $\mu = 520$  右侧多少个标准差, 我们用下面公式计算  $\bar{y}$  的  $z$  记分值

$$z = \frac{\bar{y} - \mu_0}{\sigma_{\bar{y}}},$$

其中  $\sigma_{\bar{y}} = \sigma/\sqrt{n}$ , 把各值代入公式得

$$z = \frac{\bar{y} - \mu_0}{\sigma_{\bar{y}}} = \frac{573 - 520}{124/\sqrt{36}} = 2.56$$

检查假定, 得出结论: 当样本容量  $n = 36$  时, 可以利用中心极限定理, 知  $\bar{y}$  的分布近似于正态。因为  $\bar{y}$  观测值位于零假设均值 ( $\mu = 520$ ) 加 1.96 倍标准差处的右侧, 所以我们拒绝零假设而接受研究假设, 从而得出结论: 平均每亩大豆产量大于 520。

例 5.5 所作的统计检验称之为**单侧检验**, 因为拒绝域仅位于  $\bar{y}$  的分布的一个尾部。如果我们的研究假设是  $H_a: \mu < 520$ , 那么当  $\bar{y}$  较小时拒绝零假设。这种检验也叫单侧检验, 但拒绝域位于  $\bar{y}$  分布的左侧尾部。图 5.9 显示了当  $\alpha = 0.025$  时备择假设  $H_a: \mu < 520$  的拒绝域。

我们简要介绍**双侧检验**, 其研究假设为  $H_a: \mu \neq 520$ , 其中我们感兴趣的是平均每亩大豆产量是大于 520 或小于 520。显然,  $\bar{y}$  较大和较小都与零假设相抵触, 因此我们的拒绝域位于  $\bar{y}$  分布的两侧尾部。关于  $H_a: \mu \neq 520$ ,  $\alpha = 0.05$  的双侧拒绝域如图 5.10 所示。

### 例 5.6

某公司为其销售人员供养了一支庞大的车队。为了检测平均每月每辆车的行

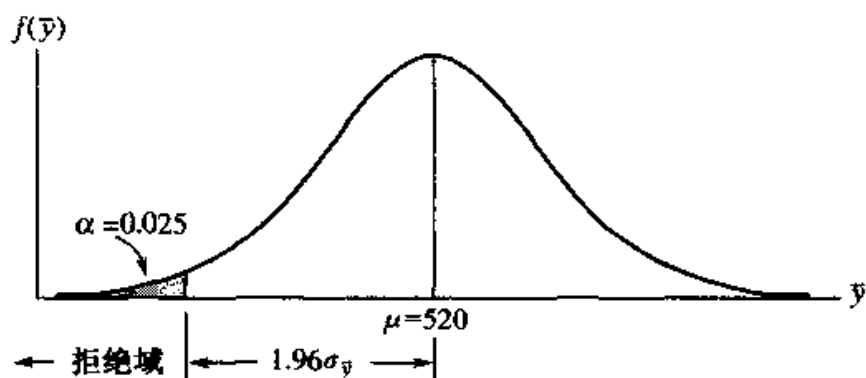


图 5.9 大豆例子中, 当  $\alpha = 0.025$ , 备择假设为  $H_a: \mu < 520$  时的拒绝域

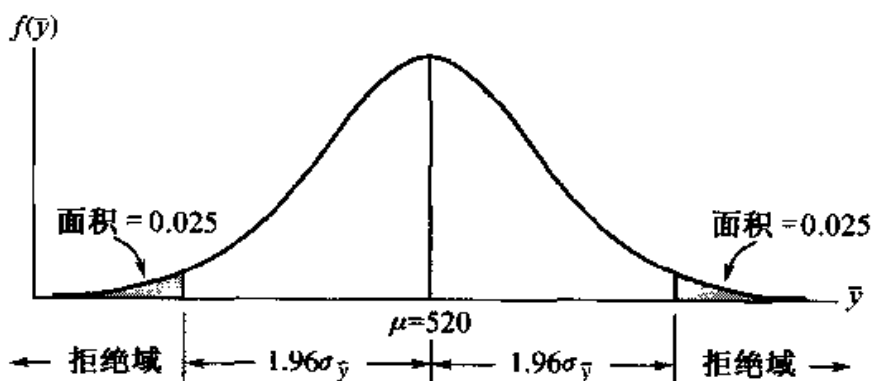


图 5.10 大豆例子中, 当  $\alpha = 0.05$  时, 备择假设为  $H_a: \mu \neq 520$  时的双侧拒绝域

驶里程, 随机抽取  $n = 40$  辆车的样本进行检测。样本均值和样本标准差分别为 2,752 和 350 英里。前几年的记录表明平均每月每辆车的行驶里程是 2,600 英里。利用样本数据检验研究假设, 即现在的均值不等于 2,600。令  $\alpha = 0.05$ , 并假设可用  $s$  估计  $\sigma$ 。

**解答** 该统计检验的研究假设是  $H_a: \mu \neq 2,600$ , 零假设是  $H_0: \mu = 2,600$ 。给定  $\alpha = 0.05$ , 则该检验的双侧拒绝域如图 5.11 所示。

当样本容量  $n = 40$  时, 对于  $\bar{y}$ , 中心极限定理应该成立。为了确定检验统计量  $\bar{y}$  距离  $\mu = 2,600$  多少个标准差。我们计算

$$z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}} = \frac{2,752 - 2,600}{350 / \sqrt{40}} = 2.75$$

$\bar{y}$  的观测值大于均值加上 1.96 个标准差的值, 所以, 拒绝零假设而接受备择假设



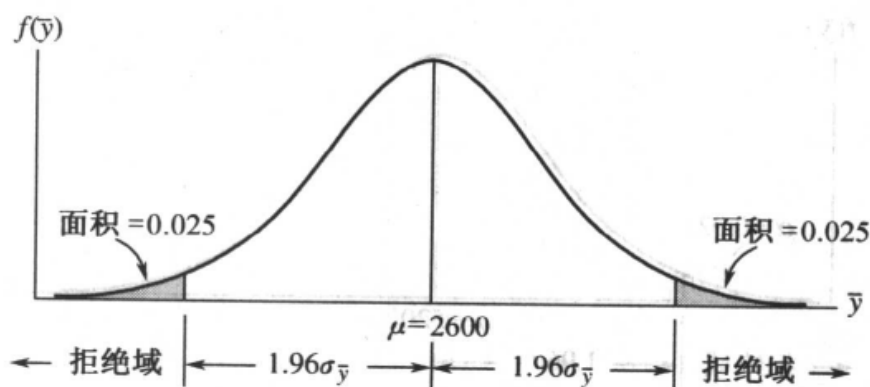


图 5.11 当  $\alpha = 0.05$  时关于  $H_a: \mu \neq 2600$  的拒绝域

$H_a: \mu \neq 2,600$ 。从而得出结论: 平均行驶里程数不等于 2,600。

如果用  $z$  而不是用  $\bar{y}$  作为检验统计量, 那么关于总体均值统计检验的步骤就能大大简化。对于

$$H_0: \mu \leq \mu_0 \text{ (其中 } \mu_0 \text{ 为某一特定值)}$$

$$H_a: \mu > \mu_0$$

检验统计量

$$z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}$$

那么对于  $\alpha = 0.025$ , 若  $z \geq 1.96$ , 即若  $\bar{y}$  位于均值右侧 1.96 个标准差之外, 则拒绝零假设。同样地, 对于  $\alpha = 0.05$ ,  $H_a: \mu \neq \mu_0$ , 如果  $z \geq 1.96$  或  $z \leq -1.96$ , 则拒绝零假设。这等价于当  $|z| \geq 1.96$  时, 拒绝零假设。

关于总体均值  $\mu$  的统计检验可概括如下。给出了三组不同的假设以及它们相应的拒绝域。对于给定的情形, 你只能从三组中选择一组假设以及相应的拒绝域。所给出的检验只适用于正态总体且  $\sigma$  已知。如果样本容量较大, 那么即使总体为非正态分布, 拒绝域也近似于正确。大多数情况下,  $n \geq 30$  就足够了, 此时我们可以使用中心极限定理的结果, 即用样本标准差  $s$  代替  $\sigma$ ,  $z = (\bar{y} - \mu_0) / (s / \sqrt{n})$  近似服从正态分布。

#### 正态总体分布或样本容量 $n$ 很大时均值的统计检验小结

假设:

情形 1.  $H_0: \mu \leq \mu_0$   $H_a: \mu > \mu_0$  (右侧检验)

情形 2.  $H_0: \mu \geq \mu_0$   $H_a: \mu < \mu_0$  (左侧检验)

情形 3.  $H_0: \mu = \mu_0$   $H_a: \mu \neq \mu_0$  (双侧检验)

$$\text{T.S.: } z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}$$

R.R.: 对于犯第一类错误的概率  $\alpha$ ,

情形 1. 若  $z \geq z_\alpha$ , 则拒绝  $H_0$ 。

情形 2. 若  $z \leq -z_\alpha$ , 则拒绝  $H_0$ 。

情形 3. 若  $|z| \geq z_{\alpha/2}$ , 则拒绝  $H_0$ 。

注: 这些方法适用于正态总体且  $\sigma$  已知。大多数情况下, 如果  $n \geq 30$ , 根据中心极限定理, 当总体为非正态分布时, 这些方法也可以使用。并且, 若  $n \geq 30$ , 我们也可用样本标准差  $s$  代替  $\sigma$ 。 $n < 30$  的情形将在本章的后面进行讨论。

### 例 5.7

作为市政管理人员对市政雇员进行评价的一部分, 市政管理人员要审查停车场管理员开据的违章停车罚款通知单以确定那些受到车主的质疑且被发现罚款不当的罚款通知单的数量。在过去几年, 每人开据不当罚款通知单的数量服从,  $\mu = 380$ ,  $\sigma = 35.2$  的正态分布。由于最近对城市停车规章制度作了一些改动, 因此市政管理人员怀疑平均每人开据的不当罚款通知单的数量会增加。随机审查 50 名停车场管理员以检验不当罚款通知单的数量是否增加。利用所给样本数据及  $\alpha = 0.01$  检验研究假设: 平均每人开据的不当罚款通知单的数量大于 380。根据审查得到下面数据:  $n = 50$ ,  $\bar{y} = 390$ 。

**解答** 利用样本数据及  $\alpha = 0.01$ , 统计检验的 5 个部分如下:

$$H_0: \mu \leq 380$$

$$H_a: \mu > 380$$

$$\text{T.S.: } z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} = \frac{390 - 380}{35.2/\sqrt{50}} = \frac{10}{35.2/7.07} = 2.01$$

R.R.: 给定  $\alpha = 0.01$ , 对于右侧检验, 若  $z \geq z_{0.01}$ , 则拒绝零假设, 其中  $z_{0.01} = 2.33$ 。

结论: 由于  $z$  的观察值 2.01 并没有超过 2.33, 因此, 我们打算接受零假设  $\mu \leq 380$ 。该结论存在的惟一问题就是我们并不知道错误地接受零假设的概率  $\beta$ 。为避免正面回答, 当  $z$  没有落在拒绝域中, 且不能算出  $\beta$  值时, 我们建议这样下结论, 即没有充分证据拒绝零假设。要想得出究竟是否接受  $H_0$ , 试验者不得不计算出  $\beta$  值。如果对于备择假设的某些  $\mu$  值,  $\beta$  值较小, 那么接受  $H_0$ 。否则, 试验者所下的结论应为: 尚没有充分证据拒绝零假设。

利用例 5.7 中的数据, 我们画图说明如何计算犯第二类错误的概率  $\beta$ 。如果零假设是  $H_0: \mu \leq 380$ , 那么不正确地接受  $H_0$  所犯错误的概率依赖于均值的真值接近于 380 的程度。例如, 若平均每人开据不当罚款单的真正数是 400, 我们可以期望当真值为 387 时,  $\beta$  值更小。均值的真值离  $\mu_0$  越近, 则我们越有可能得到数

据使得  $\bar{y}$  落在接受域。确定  $\beta$  的整个过程就是一个“如果…就会…”的过程。实际上,我们要针对备择假设  $H_a$  中的很多  $\mu$  值计算  $\beta$ ,并画出  $\beta$  对  $\mu$  的曲线图,称之为 **OC 曲线**。我们也可以通过计算当零假设不成立时拒绝零假设的概率(称为**检验的功效**)来评价检验的好坏。我们注意到功效  $= 1 - \beta$ 。称功效对  $\mu$  的曲线图为**功效曲线**。我们希望设计一种检验使得功效较大,从而  $\beta$  值较小。

假设平均每位管理员开据不当罚款单的真正数是 395,那么  $\beta$  值等于多少? 零假设和研究假设同前?

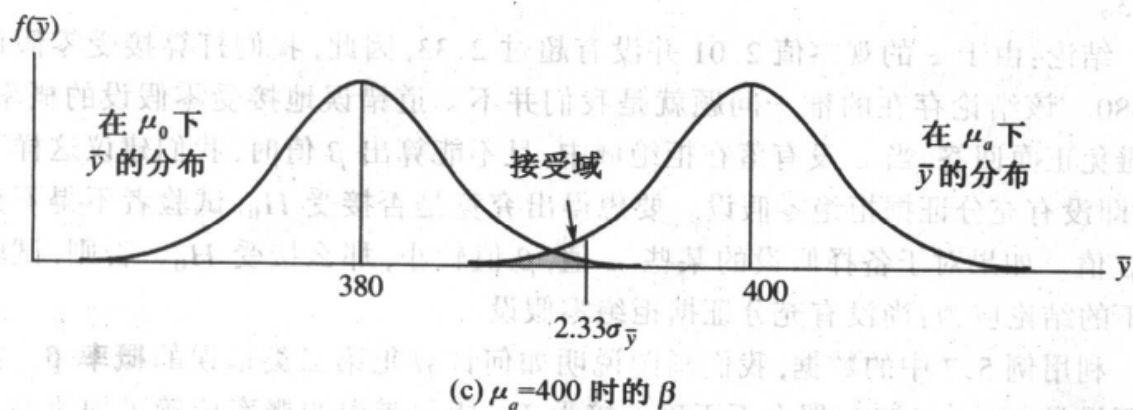
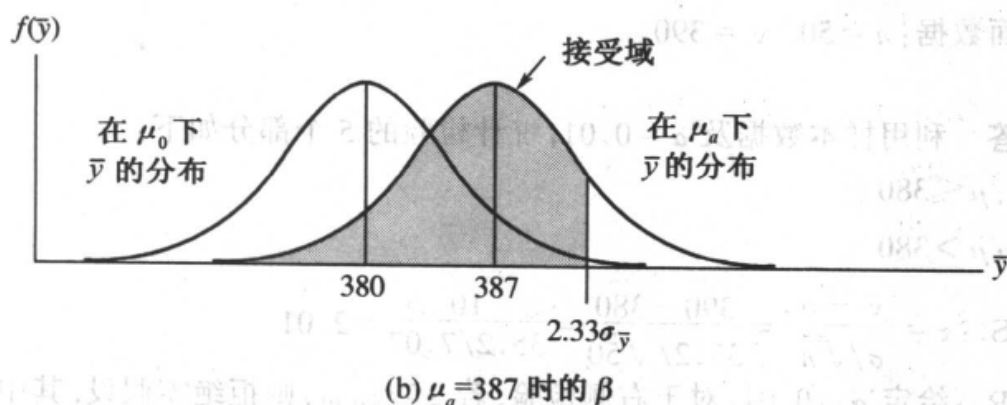
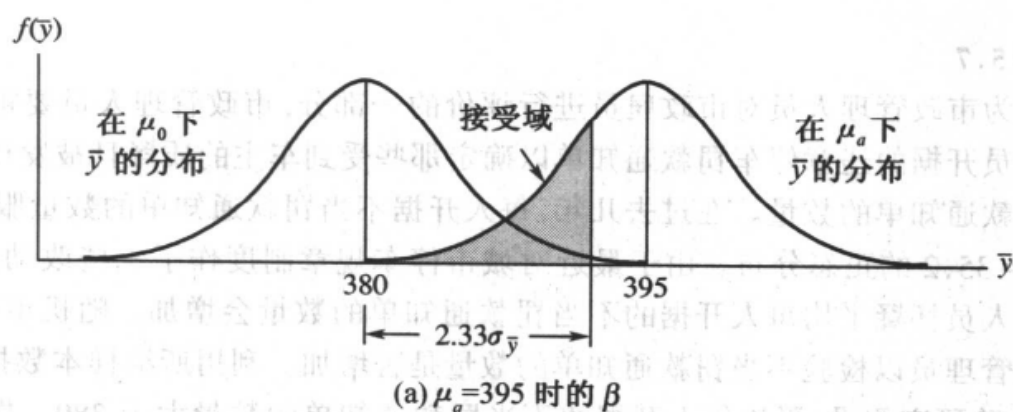


图 5.12 第二类错误的概率  $\beta$

$$H_0: \mu \leq 380$$

$$H_a: \mu > 380$$

给定  $\alpha = 0.01$ , 图 5.12(a) 显示了  $\beta$  值。图 5.12(a) 中的阴影部分代表  $\beta$ , 这就是当零假设不成立, 且  $\mu$  的真值等于 395 时  $\bar{y}$  落在接受域的概率。检出  $\mu$  的真值为 395 的检验功效等于  $1 - \beta$ , 即拒绝域面积。

考虑  $\mu$  的两个其他值, 即 387 和 400。相应的  $\beta$  值分别如图 5.12(b) 和图 5.12(c) 中的阴影部分所示; 功效是指图 5.12(b) 和图 5.12(c) 中拒绝域的非阴影部分。图 5.12 所示的三种情形均证实了我们早先提及的, 即  $\mu$  距离  $H_0$  假设下的均值越远, 第二类错误的概率  $\beta$  越小 (因此功效增加)。

下面的符号将有助于计算  $\beta$ 。记  $\mu_0$  为  $\mu$  的零值,  $\mu_a$  为  $H_a$  假设下均值的真值。当均值的真值为  $\mu_a$  时,  $\beta(\mu_a)$  表示第二类错误的概率,  $PWR(\mu_a)$  表示  $\mu_a$  处的功效。显然  $PWR(\mu_a)$  等于  $1 - \beta(\mu_a)$ 。虽然我们从来也不知道均值的真值究竟是多少, 但是我们选择  $\mu$  所可能取到的值, 然后计算相应的  $\beta$  值。这样, 一旦  $\mu$  的其中一个值正好等于真值, 那么我们就确定第二类错误的概率。对于一个或一个以上  $\mu_a$  的值, 是否接受  $H_0$  与  $\beta$  的大小有关。同样, 研究者也可以计算假设检验的功效曲线。回顾一下,  $\mu_a$  处的检验功效  $PWR(\mu_a)$  就是当  $\mu$  的真值为  $\mu_a$  时拒绝  $H_0$  的概率。因此, 我们希望  $\mu_a$  距离  $\mu$  越远时, 假设检验的功效  $PWR(\mu_a)$  越大。

对于单侧检验  $H_0: \mu \leq \mu_0$  或  $H_0: \mu \geq \mu_0$ ,  $\beta$  等于  $z$  小于  $z_\alpha - \frac{|\mu_0 - \mu_a|}{\sigma_{\bar{y}}}$  的概率, 写作

$$\beta(\mu_a) = P\left(z < z_\alpha - \frac{|\mu_0 - \mu_a|}{\sigma_{\bar{y}}}\right)$$

$\beta(\mu_a)$  的值可通过查附录中的表 1 与  $z_\alpha - |\mu_0 - \mu_a|/\sigma_{\bar{y}}$  对应的概率得到。

下面给出单侧和双侧检验下  $\beta$  的计算公式。应用这些公式的例子在后面给出。

#### 1. 单侧检验:

$$\beta(\mu_a) = P\left(z \leq z_\alpha - \frac{|\mu_0 - \mu_a|}{\sigma_{\bar{y}}}\right) \quad PWR(\mu_a) = 1 - \beta(\mu_a)$$

#### 2. 双侧检验

$$\beta(\mu_a) \approx P\left(z \leq z_{\alpha/2} - \frac{|\mu_0 - \mu_a|}{\sigma_{\bar{y}}}\right) \quad PWR(\mu_a) = 1 - \beta(\mu_a)$$

### 例 5.8

若平均每人开据不当罚款单数的真值是 395, 试计算例 5.7 中的  $\beta$  及检验功效。

**解答** 例 5.7 的研究假设为  $H_a: \mu > 380$ 。给定  $\alpha = 0.01$ , 当  $\mu_0 = 380$ ,  $\mu_a = 395$  时, 根据  $\beta$  的计算公式, 我们有

$$\begin{aligned}\beta(395) &= P\left(z < z_{0.01} - \frac{|\mu_0 - \mu_a|}{\sigma_{\bar{y}}}\right) = P\left(z < 2.33 - \frac{|380 - 395|}{35.2/\sqrt{50}}\right) \\ &= P(z < 2.33 - 3.01) = P(z < -0.68)\end{aligned}$$

查附录中的表 1, 对应于  $z = -0.68$  的面积是 0.2483。于是,  $\beta(395) = 0.2483$ ,  $\text{PWR}(395) = 1 - 0.2483 = 0.7517$ 。

以前, 当  $\bar{y}$  未落在拒绝域, 我们认为没有充分证据拒绝  $H_0$ , 因为  $\beta$  未知。现在, 当  $\bar{y}$  落入接受域时, 对于在该试验中看上去合理的对立假设值下  $\mu$  的一个(或多个)值, 我们可以计算其对应的  $\beta$  值。假如我们愿意容忍错误地接受零假设的概率等于与这个对立假设下的  $\mu$  值相对应的  $\beta$ , 那么我们就接受零假设。因此, 在例 5.8 中, 如果平均每人开据不当罚款单数的真值是 395, 那么大约有 0.25(4 次有 1 次机会)的可能性接受  $\mu$  小于或等于 380 这一假设, 而实际上  $\mu$  等于 395。市政管理人员必须分析作出这一决定的后果。如果这样的风险可以接受, 那么她可能会得出这样的结论, 即审查的结果是平均每人开据的不当罚款单数没有增加。如果风险太大, 那么市政管理人员必须扩大审查范围, 抽取 50 个以上的停车场管理员。下一节, 我们将讨论如何选取适当的  $n$ 。

### 例 5.9

百科全书公司未来的销售人员正在参加一个销售培训计划。以前的数据表明那些没有参加该培训的人员平均每月的销售量是 33。为了确定培训计划是否有效, 随机选择 35 名新雇员进行销售培训, 然后让他们从事销售。一个月后, 百科全书销售量的均值和标准差分别为 35 套和 8.4 套。数据是否提供了充分证据显示培训计划增加了销售量? 给定  $\alpha = 0.05$ 。

**解答** 统计检验的 5 个部分如下:

$$H_0: \mu \leq 33$$

$$H_a: \mu > 33$$

$$\text{T.S.: } z = \frac{\bar{y} - \mu_0}{\sigma_{\bar{y}}} \approx \frac{35 - 33}{8.4/\sqrt{35}} = 1.41$$

R.R.: 给定  $\alpha = .05$ , 若  $z \geq z_{.05} = 1.645$ , 则拒绝零假设。

核对假定并作出结论:  $n = 35$  时, 中心极限定理成立。因为  $z$  没有落入拒绝域, 所以在算出  $\beta$  以前, 我们暂不作出接受  $H_0$  的判断。换句话说, 我们尚没有充分证据拒绝零假设; 参加销售培训计划的人员平均每月的销售量不大于那些未参

加培训计划的人员平均每月的销售量。

### 例 5.10

见例 5.9。假定百科全书公司认为：如果参加培训的人员平均每月销售量是为 38，那么培训计划的财务成本将被抵消。计算对应于  $\mu_a = 38$  的  $\beta$  值，然后根据  $\beta(38)$  的值，决定你是否接受零假设。

**解答**  $\mu_0 = 33, \mu_a = 38, \alpha = 0.05$ ，利用  $\beta$  的计算公式，得

$$\begin{aligned}\beta(38) &= P\left[z \leq z_{0.05} - \frac{|\mu_0 - \mu_a|}{\sigma_{\bar{y}}}\right] = P\left[z \leq 1.645 - \frac{|33 - 38|}{8.4/\sqrt{35}}\right] \\ &= P[z \leq -1.88]\end{aligned}$$

查附录中表 1，对应于  $z = -1.88$  的面积是 0.0301。因此，

$$\beta(38) = 0.0301 \quad \text{PWR}(38) = 1 - 0.0301 = 0.9699$$

因为  $\beta$  相对较小，于是我们接受零假设，认为培训计划增加的平均每月销售量尚不能抵消培训成本。

对应于  $H_a$  假设下  $\mu$  的其他几个值，百科全书公司欲计算其第二类错误的机会，这样，根据新雇员的随机样本数据，他们就能大致地知道犯第二类错误的概率。重新计算，得  $\beta(38)$  如表 5.4 所示。

表 5.4 对应于  $H_a$  假设下的  $\mu$  值，第二类错误的概率及功效

$\mu$	33	34	35	36	37	38	39	40	41
$\beta(\mu)$	0.9500	0.8266	0.5935	0.3200	0.1206	0.0301	0.0049	0.0005	0.0000
$\text{PWR}(\mu)$	0.0500	0.1734	0.4065	0.6800	0.8794	0.9699	0.9951	0.9995	0.9999

图 5.13 的曲线是根据表 5.4 中  $\beta(\mu)$  的值进行描点，然后用光滑的曲线连接而成的。可见，随着  $\mu$  值的增加，第二类错误的概率减少到 0，相应的功效增加到 1.0。公司想通过考察曲线来决定，对应于  $H_a$  假设下的对公司至关重要的  $\mu$  值，第二类错误的概率是否合理。从表 5.4 或图 5.13 得知， $\beta(38) = 0.0301$  是一个相对小的值。从例 5.10 的结果，我们发现检验统计量并没有落入拒绝域。由于  $\beta(38)$  较小，因此我们接受零假设，从而得出结论：培训计划增加的平均每月销售量尚不能抵消培训成本。

在 5.2 节中，我们讨论了如何评价区间估计的有效性。统计检验的有效性可通过第一类错误和第二类错误的大小，即  $\alpha$  和  $\beta(\mu)$  来衡量。 $\alpha$  是事先给定的试验者的容忍水平，而对于固定的  $\mu_a$ ， $\beta(\mu_a)$  是样本容量的函数。样本容量越大，我们

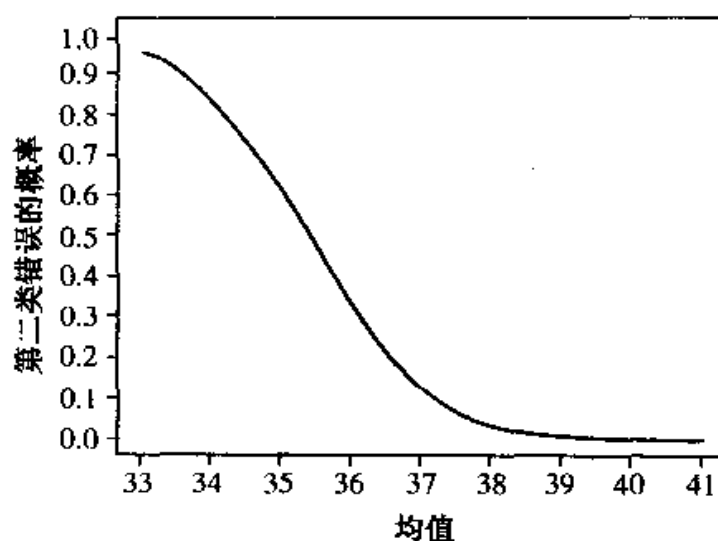


图 5.13 第二类错误的概率

所关心的  $\mu$  的信息越多, 我们越不太可能犯第二类错误, 从而  $\beta(\mu_a)$  越小。为了用图来说明这种思想, 假定我们正在检验假设  $H_0: \mu \leq 84$  对  $H_a: \mu > 84$ , 其中  $\mu$  是  $\sigma = 1.4$  的正态分布总体的均值, 若取  $\alpha = 0.05$ , 则图 5.14(a) 画出了针对三种样本容量  $n = 10, 18, 25$ , 第二类错误概率的曲线。可见, 随着  $n$  从 10 增加到 25,  $\beta(84.6)$  变得越来越小。另一个有趣的关系就是  $\alpha$  和  $\beta(\mu)$  之间的关系。对于固定的样本容量  $n$ , 如果我们改变拒绝域以增加  $\alpha$  值, 则  $\beta(\mu_a)$  将减小。这种关系可从图 5.14(b) 中看出。固定样本容量为 25, 针对  $\alpha$  的三个不同值  $\alpha = 0.05, 0.01, 0.001$ , 画出  $\beta(\mu)$  曲线。我们看到, 随着  $\alpha$  从 0.001 增加到 0.05,  $\beta(84.6)$  变得越来越小。对于检验功效可画出同样的曲线图, 只要简单地将  $\text{PWR}(\mu) = 1 - \beta(\mu)$  对  $\mu$  画出即可。对于功效曲线, 上面的关系正好颠倒过来; 也就是说, 对于固定的  $\alpha$  增加样本容量将增加  $\text{PWR}(\mu)$  值, 对于固定的样本容量,  $\alpha$  值增加  $\text{PWR}(\mu)$  值也将增加。现在, 我们考虑当  $\alpha$  给定, 对于  $\mu_a, \beta(\mu_a)$  也事先给定时, 如何设计一个试验对  $\mu$  进行检验。这个问题可简化为如何确定样本容量以达到固定的  $\alpha$  和  $\beta(\mu_a)$ 。值得注意的是, 对于初始给定的  $\alpha$  和  $\beta$ , 若  $n$  值太大, 则我们通过增加  $\alpha$  的值, 用较小的样本容量达到期望的  $\beta(\mu_a)$  值。

## 5.5 对于 $\mu$ 进行检验时样本容量的选取

我们可根据第一类和第二类错误的大小, 即  $\alpha$  和  $\beta(\mu)$  来衡量从  $\mu$  的假设检验中获取的信息的质量, 其中  $\beta(\mu)$  是对应于备择假设  $H_a$  下各个  $\mu$  的  $\beta$  值。假定我们欲检验假设  $H_0: \mu \leq \mu_0$   $H_a: \mu > \mu_0$ 。首先, 我们必须给定  $\alpha$ 。其次, 确定一

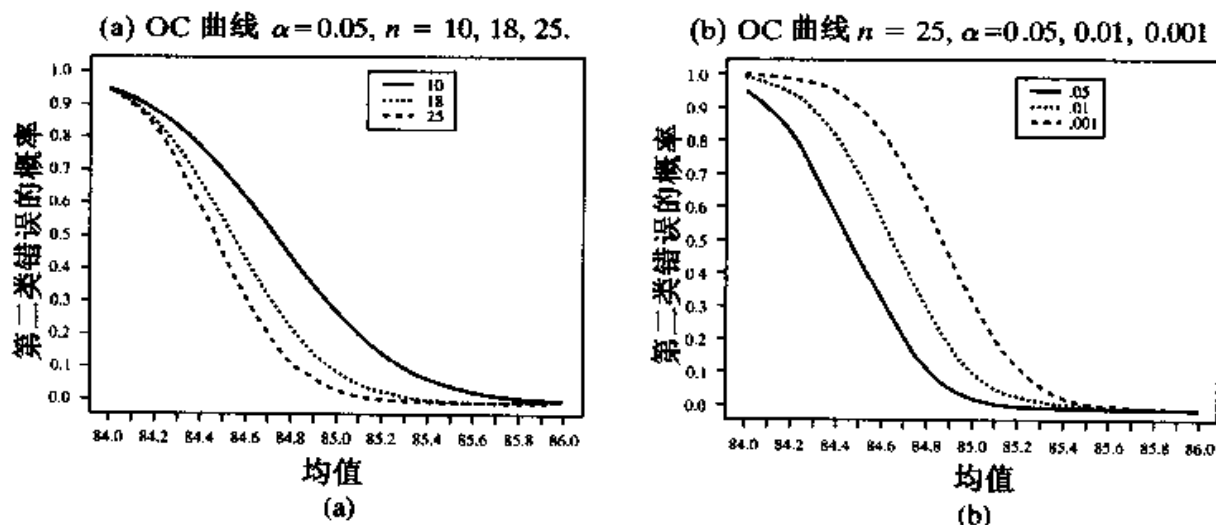


图 5.14

个备择假设中的  $\mu$  值, 记为  $\mu_1$ , 如果均值的真值大于  $\mu_1$ , 则犯第二类错误的后果将是严重的。最后, 我们必须为  $\beta(\mu_1)$  选择一个  $\beta$  值。对于大于  $\mu_1$  的任何  $\mu$ , 其第二类错误的概率小于  $\beta(\mu_1)$ ; 即,

$$\beta(\mu) < \beta(\mu_1) \quad \text{对于所有的 } \mu > \mu_1$$

令  $\Delta = \mu_1 - \mu_0$ 。为满足这些要求, 需要的样本容量为

$$n = \sigma^2 \frac{(z_\alpha + z_\beta)^2}{\Delta^2}$$

注: 若  $\sigma^2$  未知, 则可以用从以前的研究中所得到的估计值代替, 或者通过一个试探性的研究得到样本容量的近似值。

同样的公式也适用于检验  $H_0: \mu \geq \mu_0$   $H_a: \mu < \mu_0$ , 只不过此时当  $\mu$  的真值小于  $H_a$  中的某个均值  $\mu_1$  时, 我们期望第二类错误的概率不大于  $\beta$ , 即,

$$\beta(\mu) < \beta(\mu_1) \quad \text{对于所有的 } \mu < \mu_1$$

且  $\Delta = \mu_1 - \mu_0$ 。

### 例 5.11

谷物加工厂商生产一种食品, 并将其装入标有 16 盎司的盒子。用机器进行装填, 机器被设定为平均每盒装 16.37 盎司。因为这些机器所装盒子的真正重量服从标准差为 0.225 盎司的正态分布, 重量不足 16 盎司的盒子占 5%。厂商关心是否其中有一台机器装填不足, 于是从这台机器装填的成品中抽取几盒, 以检查平均重量  $\mu$  是否小于 16.37, 也就是说, 给定  $\alpha = 0.05$ , 检验

$$H_0: \mu \geq 16.37$$



$$H_a: \mu < 16.37$$

如果平均重量的真值是 16.27 或更小, 那么厂商要求未能发现装填不足的概率最多为 0.01, 或者说招致民事赔偿的风险不超过 0.01。因此, 我们需要选取样本容量, 使得对于任何小于 16.37 盎司的  $\mu$  值,  $H_0$  对  $H_a$  的检验满足  $\alpha = 0.05$ ,  $\beta(\mu)$  小于 0.01。

**解答** 由于  $\alpha = 0.05$ ,  $\beta = 0.01$ ,  $\Delta = 16.37 - 16.27 = 0.1$ ,  $\sigma = 0.225$ , 将  $z_{0.05} = 1.645$ ,  $z_{0.01} = 2.33$  代入公式得

$$n = \frac{(0.225)^2(1.645 + 2.33)^2}{(0.1)^2} = 79.99 \approx 80$$

因此, 在给定条件下, 厂商必须抽取  $n = 80$  盒的随机样本进行检验。

假定根据样本资料算得  $\bar{y} = 16.35$ 。检验统计量的值为

$$z = \frac{\bar{y} - 16.37}{\sigma_{\bar{y}}} = \frac{16.35 - 16.37}{0.225/\sqrt{80}} = -0.795$$

由于拒绝域是  $z < -1.645$ , 因此  $z$  未落入拒绝域。如何下结论呢? 前几节中在同样情况下, 我们的结论是: 尚没有充分证据拒绝  $H_0$ 。然而, 现在当  $\mu \leq 16.27$  时, 我们知道  $\beta(\mu) \leq 0.01$ , 因此我们可以更放心地作出结论: 接受  $H_0: \mu \geq 16.37$ 。这样, 厂商也可放心地作出结论: 机器的平均装填量至少为 16.37 盎司。

对单侧检验下样本容量的计算公式进行稍微修正, 对于给定的  $\alpha, \beta, \Delta$  我们就能检验

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

其中  $\beta, \Delta$  满足只要  $|\mu - \mu_0| \geq \Delta$ , 都有  $\beta(\mu) < \beta$ 。因此, 当均值的真值距离  $\mu_0$  至少  $\Delta$  时, 第二类错误的概率最多为  $\beta$ 。检验关于  $\mu$  的双侧假设时, 样本容量的近似公式为

$$n = \frac{\sigma^2}{\Delta^2} (z_{\alpha/2} + z_{\beta})^2$$

注: 若  $\sigma^2$  未知, 则用其估计值代替以得到样本容量的近似值。

## 练习

### 基本技能

5.25 参见例 5.11 中的数据, 假定我们要检验均值是否不等于 16.37。

a. 选取样本容量以检验  $H_0: \mu = 16.37$   $H_a: \mu \neq 16.37$ , 使得当  $\mu$  的真值距  $\mu_0 = 16.37$  有 0.1 个单位以上的距离时  $\alpha = 0.05$ ,  $\beta(\mu)$  小于 0.01。

b. 和例 5.11 中单侧检验相比, 该样本容量如何?

**5.26** 给定  $\alpha = 0.05$ , 某研究人员要检验  $H_0: \mu \leq 38$   $H_a: \mu > 38$ 。从总体随机抽取容量为 50 的样本, 得到  $\bar{y} = 40.1, s = 5.6$ 。

- 根据样本数据, 你能对假设得出什么结论?
- 这种情况下, 你会犯第二类错误吗? 试解释之。
- 若  $\mu$  的真值至少为 39, 试计算第二类错误的概率。

**5.27** 根据练习 5.26 中的数据, 画出通过确定  $PWR(\mu_a)$  而拒绝  $H_0: \mu \leq 38$  的功效曲线图, 其中  $PWR(\mu_a)$  中的  $\mu_a$  取备择假设中的  $\mu$  值: 39, 40, 41, 42, 43, 44。

- 解释曲线上的值。
- 不必真正计算  $PWR(\mu_a)$ , 画出  $\alpha = 0.025, n = 50$  的功效曲线草图。
- 不必真正计算  $PWR(\mu_a)$ , 画出  $\alpha = 0.05, n = 20$  的功效曲线草图。

**5.28** 利用计算机软件程序, 模拟 100 个容量为 16 的来自于  $\mu = 40, \sigma = 8$  的正态总体的样本。给定  $\alpha = 0.01$ , 我们希望检验假设  $H_0: \mu = 40$   $H_a: \mu \neq 40$ 。用 100 个样本中的每一个样本数据分别进行检验时, 均假定  $\sigma = 8$ , 且总体服从正态分布。

- 100 次检验中有多少次你不正确地拒绝  $H_0$ 。
- 平均来说, 给定  $\alpha = 0.01$ , 如果你作了 100 次检验, 那么你预料有多少次得出拒绝  $H_0$  这样的结论?
- 如果你不正确地拒绝  $H_0$ , 那么你犯了哪一类错误?

**5.29** 参看练习 5.28, 假定总体均值是 41.5, 而不是 40。模拟 100 个容量为 16 的来自于  $\mu = 41.5, \sigma = 8$  的正态总体的样本。在检验假设  $H_0: \mu = 40$   $H_a: \mu \neq 40$  时, 针对每个样本, 均设定  $\alpha = 0.10$ 。

- 100 次检验中, 决策正确, 即拒绝  $H_0$  的次数所占的比例是多少? 易见, 你正在计算检验功效, 即检验方法发现原假设不成立的能力。
- 当  $\mu = 41.5$  时, 计算检验的理论功效。
- 基于(b)中结论, 你预期 100 次检验中有多少次正确地拒绝了  $H_0: \mu = 40$ 。
- 当  $\mu = 38$  和  $\mu = 43$  时, 重复模拟, 并回答问题(a)~(c)。

**5.30** 参看练习 5.28 和 5.29。

- 给定  $\alpha = 0.01$ , 而不是  $\alpha = 0.10$ , 回答这些习题中提出的问题。你也可以使用习题 5.28 和 5.29 所模拟的数据集, 但精确的功效还需重新计算。
- 将  $\alpha$  从 0.10 减少到 0.01, 检验功效是增加还是减少? 并解释为什么会发生这样的变化。

## 应用

**5.31(商业)** 疗养院的管理人员希望做一项工时效率研究, 以便了解职工每天花费在非紧急任务上的时间。在没有采用有效方法之前, 平均每天花费在这些

任务上的工时为  $\mu = 16$ 。管理人员欲检验有效方法的实施是否减小了  $\mu$  值。当均值的真值不超过 12 小时(即有效方法的实施至少降低工时 25%)时,她希望其检验使得  $\alpha = 0.05$ , 第二类错误的概率最多为 0.10, 为此需要选取多少天? 给定  $\sigma = 7.64$ 。

**5.32** 某项研究需要对 90 位成年男性病人实施一项新的治疗充血性心力衰竭的方法。对病人进行测量的变量之一就是经过 4 周的一个疗程, 病人锻炼能力(以分钟计)的增加。以前的治疗方法使得锻炼能力增加  $\mu = 2$  分钟。研究人员想评价一下, 和以前的治疗方法相比, 新方法是否增加了  $\mu$  值。由样本数据得  $\bar{y} = 2.17$ ,  $s = 1.05$ 。

- a. 给定  $\alpha = 0.05$ , 你如何对研究假设作出结论?
- b. 如果  $\mu$  的真值等于 2.1, 那么犯第二类错误的概率是多少?

**5.33** 参看练习 5.32。就  $\mu_a = 2.1, 2.2, 2.3, 2.4, 2.5$ , 计算检验功效  $PWR(\mu_a)$ 。用一条光滑的曲线连接点  $(\mu_a, PWR(\mu_a))$ 。

- a.  $\alpha$  从 0.05 减小到 0.01 将对功效曲线产生什么样的影响?
- b. 样本容量从 90 减小到 50 会对功效曲线产生什么样的影响?

**5.34** 某个 1 年期的试验计划旨在提高贫困的中学毕业班学生的数学成绩。为了评价该计划的成果, 从参加该计划的学生中随机抽取一部分学生, 然后将他们的数学成绩与前一年州贫困学生的平均成绩 525 进行对比。研究人员想搞清是否该试验计划使得平均成绩比前一年州平均成绩有所提高。给定  $\alpha = 0.05$ , 如果均值的真值增加到 550, 样本容量取多少才能使得第二类错误的概率不超过 0.025? 由以前的结果知  $\sigma \approx 80$ 。

**5.35** 参看练习 5.34。假定选取 100 名学生作为随机样本得到  $\bar{y} = 542$ ,  $s = 76$ 。是否有充分证据表明平均数学成绩得到了提高? 试解释之。

**5.36** 为了研究减肥剂的效用, 进行一次临床试验, 固定 35 位超重男性的食谱。两周以后, 称一下每位体重, 然后给他们提供减肥剂。饮食不变; 另外, 每天服用一剂减肥剂。又过了两周, 再称一次体重。根据以前的研究, 仅仅固定饮食, 平均体重就会减少 5 磅。研究人员欲检验当  $\alpha = 0.05$  时服用减肥剂平均体重减少会超过 5 磅。进一步地, 他们需要其检验至少有 90% 的可能观察到平均体重减少不少于 8 磅。

- a. 进入研究之前, 35 位男性的体重在 195 至 415 磅之间。他们之间的体重差别如此之大会影响研究所得出的结论吗? 试解释之。你如何设计一个不同的研究方法?
- b. 由以前的研究知  $\sigma \approx 6.8$  磅。样本容量是否大到足以满足研究人员对  $\alpha$  和功效的要求? 如果不是, 样本容量需要多大?

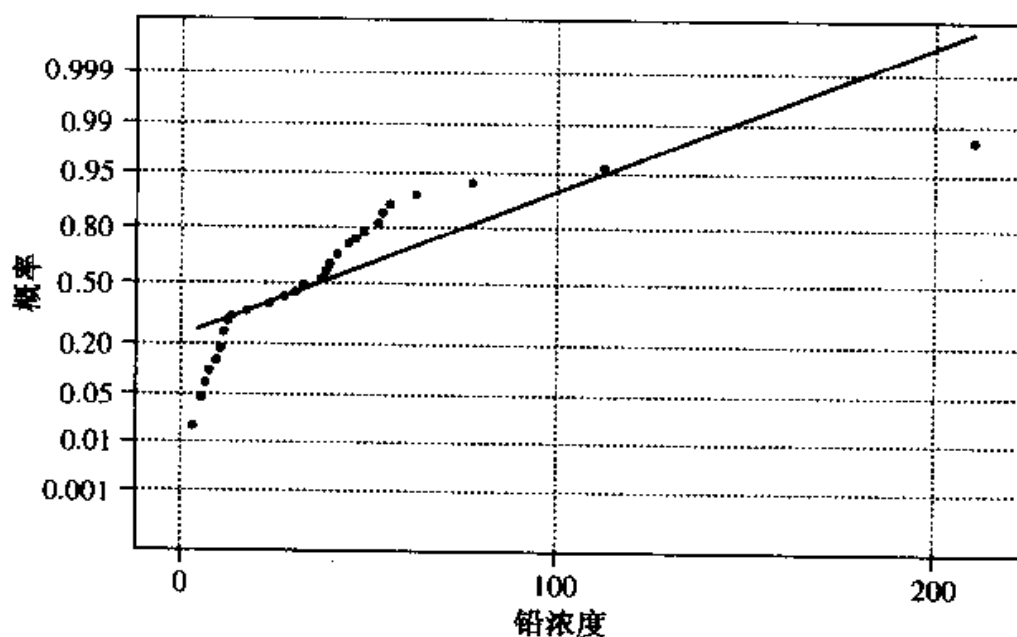
**5.37** 参看练习 5.36。

- a. 在第二个两周期间, 35 位男性的平均体重减少量为  $\bar{y} = 7.3$  磅, 且  $s = 4.6$  磅。给定  $\alpha = 0.05$ , 是否有充分证据表明减肥剂是有效的?
- b. 基于你在(a)中的结论, 犯第一类错误的机率有多大? 第二类错误呢?

**5.38(工程)** 在论文“Metals, petroleum hydrocarbons and organochlorines in inshore sediments and waters on Mombasa, Kenya” (*Marine Pollution Bulletin*, 1997, pp.570~577)中, 作者讨论了在蒙巴萨(Mombasa)地区, 由于城市的扩张和工业化造成的污染使得近海岸的环境变得很脆弱。在 1995 年 9 月至 1996 年 1 月期间, 对肯尼亚的蒙巴萨近海水域进行了一次地质化学和海洋学的调查。在调查中, 从位于蒙巴萨河口地区的 48 个观测站收集了悬浮的颗粒物和沉淀物。在每个观测站不同数量的活性区测量了主要氧化物和 13 种微量元素的浓度。特别地, 在 37 个观测站测量了悬浮的微粒物中铅的浓度( $\text{mg kg}^{-1}$ 干重)。研究人员想知道平均铅浓度是否大于  $30\text{mg kg}^{-1}$ 干重。下表列出了收集的数据, 并且给出了概括统计量及正态概率图。

肯尼亚 37 个观测站铅的浓度

48	53	44	55	52	39	62	38	23	27
41	37	41	46	32	17	32	41	23	12
3	13	10	11	5	30	11	9	7	11
77	210	38	112	52	10	6			



- a. 数据中是否有充分证据显示平均铅浓度超过  $30\text{mg kg}^{-1}$ 干重?
- b. 如果平均铅浓度的真值是 50, 那么第二类错误的概率等于多少?

c. 数据看上去具有正态分布吗?

d. 基于(c)中的答案, 样本容量是否大到使得检验方法有效? 试解释之。

**5.39(医学)** 一般来说, 首先是那些不规则形状的牙齿(常称之为臼齿)发生腐蚀。这些牙上最容易感染的牙面是咀嚼面。这些表面的釉质通常含有微小的空洞, 他们容易藏有食物的微粒。细菌开始吃这些食物微粒, 从而创造一个牙面腐蚀的环境。

除了牙齿的自然硬化以外, 影响牙齿腐蚀速度的特别重要的因素就是每个人所吃食物的种类。一些碳水化合物对牙齿的健康特别有害。人们做了很多研究来证实这些发现, 我们可以想象如何进行这样的研究。从某一地区选取 60 名男性作为随机样本。每人都接受了牙齿检查, 并且按照指定的食谱进餐, 每餐中有一杯糖水。一年以后, 该组平均每人新腐蚀的牙齿数为 0.70, 标准差为 0.4。

a. 给定  $\alpha = 0.05$ , 这些数据是否提供了充分的证据表明那些每餐喝糖水的人平均每人新腐蚀的牙齿数超过 0.30? 0.30 是一个感兴趣的值, 因为它被证明是那些每餐不喝糖水的人平均每人新腐蚀的牙齿数。

b. 为什么双侧检验不适合?

## 5.6 统计检验的显著性水平

在 5.4 节, 我们按照相当传统的方法介绍了假设检验: 确定假设检验的 5 个步骤以及两类错误和相应的概率  $\alpha, \beta(\mu_a)$ 。问题是如果其他人应用你的研究结果时使用不同的  $\alpha$  值, 那么在作出有关  $H_0$  和  $H_a$  的结论之前, 他们必须计算新的拒绝域。假设检验还有另一种方法, 其步骤如下: 确立零假设和备择假设, 给定  $\alpha$  值, 收集样本数据, 最后, 求出拒绝零假设的证据的权重。权重通常用概率的形式来表示, 称之为统计检验的**显著性水平**(或  $p$  值)。显著性水平的更正式的定义如下: 显著性水平是指在假定零假设为真的条件下, “检验统计量的取值与已经观测到的检验统计量的值相比同样或更倾向于拒绝  $H_0$ ”这样一个事件发生的概率。因此, 如果显著性水平较小, 那么样本数据不能支持假设  $H_0$ , 从而我们拒绝  $H_0$ 。另一方面, 如果显著性水平较大, 那么我们不能拒绝  $H_0$ 。接下来, 我们必须界定显著性水平究竟等于多少才称之为“大”或“小”。下面的判决规则所得出的结果总是和 5.5 节介绍的检验方法一致。

**用  $p$  值进行假设检验的判决规则**

1. 若  $p$  值  $\leq \alpha$ , 那么拒绝  $H_0$ 。

2. 若  $p$  值  $> \alpha$ , 那么不能拒绝  $H_0$ 。

我们举几个例子来说明如何计算显著性水平。

**例 5.12**

见例 5.7。

- a. 求统计检验的显著性水平( $p$  值)。给定  $\alpha = 0.01$ , 作出关于研究假设的结论。
- b. 如果事先确定  $\alpha$  为 0.05 而不是 0.01, 那么你关于  $H_a$  的结论改变吗?

**解答**

- a. 零假设和备择假设为

$$H_0: \mu \leq 380$$

$$H_a: \mu > 380$$

根据样本数据, 检验统计量的值为

$$z = \frac{\bar{y} - 380}{s/\sqrt{n}} = \frac{390 - 380}{35.2/\sqrt{50}} = 2.01$$

该检验的显著性水平(即拒绝  $H_0$  的权重)就是在假定零假设为真(即  $\mu = 380$ )的条件下, 得到一个大于或等于 390 的  $\bar{y}$  观测值的概率。此概率可借助于检验统计量的  $z$  值 2.01 来计算, 因为

$$p \text{ 值} = p(\bar{y} \geq 390, \text{假定 } \mu = 380) = p(z \geq 2.01)$$

查附录中的表 1,  $P(z \geq 2.01) = 1 - P(z < 2.01) = 1 - 0.9778 = 0.0222$ 。如图 5.15 阴影部分所示。因为  $p$  值大于  $\alpha$  ( $0.0222 > 0.01$ ), 所以我们拒绝  $H_0$ , 从而我们得出结论: 样本数据不支持研究假设。

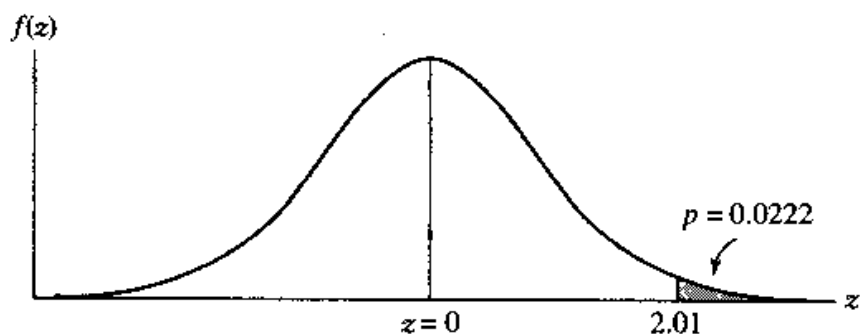


图 5.15 例 5.12 中的显著性水平

- b. 另一个人分析同样的数据, 但事先确定  $\alpha = 0.05$ 。这个人愿意冒更大的风险犯第一类错误, 因此他的结论是拒绝  $H_0$ , 因为  $p$  值小于  $\alpha$  ( $0.0222 < 0.05$ )。重要的是在判决规则中的  $\alpha$  是事先确定的, 而不是计算  $p$  值之后再选择。

正如我们从例 5.12 所看到的, 显著性水平表示观测到比已观测到的样本更

不利于  $H_0$  的样本的概率。这个概率越小, 拒绝  $H_0$  的证据越充分。例如, 和另一个显著性水平  $p=0.20$  的统计检验相比, 显著性水平  $p=0.01$  的统计检验具有更充分的证据拒绝  $H_0$ 。

### 例 5.13

见例 5.11。若事先给定  $\alpha=0.05$ , 样本数据有充分的证据支持研究假设吗?

**解答** 零假设和备择假设为

$$H_0: \mu \geq 16.37$$

$$H_a: \mu < 16.37$$

根据样本数据, 检验统计量的计算值为

$$z = \frac{\bar{y} - \mu_0}{s / \sqrt{n}} = \frac{16.35 - 16.37}{0.225 / \sqrt{80}} = -0.795$$

计算显著性水平就是求观测到比已有的观测值  $\bar{y} = 16.35$  更不利于  $H_0$  的  $\bar{y}$  值的概率。因为  $H_a$  假设  $\mu$  小于 16.37, 所以更不利于  $H_0$  的  $\bar{y}$  值就是那些小于已有的  $\bar{y}$  值 16.35 的那些值。因此,

$$p \text{ 值} = P(\bar{y} \leq 16.35, \text{假设 } \mu = 16.37) = P(z \leq -0.795) = 0.2119$$

从而有可观的证据支持  $H_0$ 。更精确地说,  $p \text{ 值} = 0.2119 > 0.05 = \alpha$ , 因此我们不能拒绝  $H_0$ 。故我们的结论是: 没有充分证据( $p \text{ 值} = 0.2119$ )支持研究假设。可见这恰和用传统方法得出的结论一样。

对于双侧检验,  $H_a: \mu \neq \mu_0$ , 为了求显著性水平, 我们仍然要计算观测到比已有的检验统计量的值更不利于  $H_0$  的样本的概率。然而, 对于双侧假设检验, 我们要根据  $\bar{y}$  到  $\mu$  的零值的距离来计算这个概率, 因为无论  $\bar{y}$  比  $\mu_0$  小得太多还是  $\bar{y}$  比  $\mu_0$  大得太多, 都与  $\mu = \mu_0$  相对立。因此, 显著性水平为

$$\begin{aligned} p \text{ 值} &= p(|\bar{y} - \mu_0| \geq \text{已观测到的 } |\bar{y} - \mu_0|) \\ &= p(|z| \geq |z \text{ 的计算值}|) = 2p(z \geq |z \text{ 的计算值}|) \end{aligned}$$

将显著性水平( $p \text{ 值}$ )的计算概括如下:

情形 1	情形 2	情形 3
$H_0: \mu \leq \mu_0$	$H_0: \mu \geq \mu_0$	$H_0: \mu = \mu_0$
$H_a: \mu > \mu_0$	$H_a: \mu < \mu_0$	$H_a: \mu \neq \mu_0$
$p \text{ 值}: p(z \geq z \text{ 的计算值})$	$p(z \leq z \text{ 的计算值})$	$p(z \geq  z \text{ 的计算值} )$

**例 5.14**

见例 5.6。若事先给定  $\alpha = 0.01$ , 样本数据有充分的证据支持研究假设吗?

**解答** 零假设和备择假设为

$$H_0: \mu = 2,600$$

$$H_a: \mu \neq 2,600$$

根据样本数据, 检验统计量的计算值为

$$z = \frac{\bar{y} - \mu_0}{s / \sqrt{n}} = \frac{2,752 - 2,600}{350 / \sqrt{40}} = 2.75$$

检验统计量的显著性水平可根据前面的公式进行计算。

$$p \text{ 值} = 2p(z \geq |z \text{ 的计算值}|)$$

$$= 2p(z \geq |2.75|) = 2p(z \geq 2.75) = 2(1 - 0.9970) = 0.006$$

因为  $p$  值很小, 所以没有什么证据支持  $H_0$ 。更精确地说,  $p \text{ 值} = 0.006 \leq 0.05 = \alpha$ , 从而我们拒绝  $H_0$ 。因此, 有充分证据 ( $p \text{ 值} = 0.006$ ) 支持研究假设, 从而我们得出结论: 平均行驶里程不等于 2,600。可见这恰与用传统方法得出的结论一样。

这种假设检验方法的好处很多。统计学家(或从事统计检验的人)给予试验者拒绝零假设证据的权重, 而不是直接得出结论。然后, 试验者可据此得出他或她自己的结论。一些试验者当  $p \leq 0.10$  时就拒绝零假设, 而其他一些试验者拒绝零假设时需要  $p \leq 0.05$  或  $p \leq 0.01$ 。试验者在他或她认为有充分证据拒绝零假设时, 作出自己的决策。

许多专业期刊已经用显著性水平报告统计检验的结果。因此, 我们也许会读到: 某检验在  $p = 0.05$  或  $p \leq 0.01$  的水平下是显著的。按照这种方法报告结果后, 读者可以得出自己的结论。

需要警告的是, 0.05 的  $p$  值已经是一个具有魔力的水平。很多人都觉得除非检验达到了 0.05 水平或更低, 否则零假设不应该被拒绝。之所以如此, 部分是由于基于判决的方法通常事先给定  $\alpha$  为 0.05。当你在期刊杂志上读到这类文章或报告你自己的统计检验时, 不要落入这样的圈套。毕竟, 在某一水平下的统计显著性并不表明其实际的显著性。相反地, 它意味着在特别低的风险下拒绝零假设。例如, 假设某公司想知道平均每月每辆车的行驶里程是否增加到 2,600 英里。400 辆车的样本数据表明  $\bar{y} = 2,640$ ,  $s = 35$ 。根据这些数据及  $\sigma = 35$ , 我们得到关于  $H_0: \mu = 2,600$  的  $z$  统计量为  $z = 22.86$ ; 显著性水平为  $p = 0.0000000001$ 。因此即使平均每月每辆车的行驶里程只增加 1.5%, 从统计上来说, 其结果也是高度显著的。但这种增长具有实际意义吗? 可能不。我们检验的结果表明  $\mu$  稍微增加了一点。

以后我们在做统计检验时, 即使用基于判决的方法, 也使用显著性水平的方



法, 以使你熟悉这两种思路。无论使用哪种方法, 都应考虑在你根据统计检验得出结论之后, 该结论的实际意义。

### 练习

#### 基本技能

**5.40** 由关于  $\mu$  的统计检验的样本数据得  $n = 50$ ,  $\bar{y} = 48.2$ ,  $s = 12.57$ 。求检验  $H_0: \mu \leq 45$   $H_a: \mu > 45$  时的显著性水平。数据中是否有明显的证据支持假设:  $\mu$  大于 45? 给定  $\alpha = 0.05$ 。

**5.41** 参看练习 5.40。如果研究人员用  $\alpha = 0.025$ , 而不是  $\alpha = 0.05$ , 关于  $\mu$  的结论有变化吗? 试解释同样的数据如何得出一个关于  $\mu$  的不同结论。

**5.42** 由关于  $\mu$  的统计检验的样本数据得  $n = 30$ ,  $\bar{y} = 5.8$ ,  $s = 4.11$ 。求检验  $H_0: \mu = 4$   $H_a: \mu \neq 4$  时的显著性水平。数据中是否有明显的证据支持研究假设:  $\mu$  不等于 4? 给定  $\alpha = 0.01$ 。

**5.43** 参看练习 5.42, 如果研究假设是单侧假设  $H_0: \mu \leq 4$   $H_a: \mu > 4$ , 而不是双侧假设, 你关于  $\mu$  的结论有变化吗?

**5.44** 研究人员想知道  $\mu$  是否小于 14。由样本数据得  $n = 40$ ,  $\bar{y} = 13.5$ ,  $s = 3.62$ 。求检验  $H_0: \mu \geq 14$   $H_a: \mu < 14$  时的显著性水平。数据中是否有明显的证据支持假设:  $\mu$  小于 14? 给定  $\alpha = 0.05$ 。

**5.45** 假定练习 5.44 中的假设被错误地表述为  $H_0: \mu \leq 14$   $H_a: \mu > 14$ , 求这一假设检验的显著性水平, 并证明所得关于  $\mu$  的结论与习题 5.44 中的结论相反。

### 应用

**5.46(医学)** 某烟草公司广告声称其香烟中平均尼古丁含量最多为 14 毫克。消费者保护组织欲了解他们的平均尼古丁含量是否实际上多于 14 毫克。随机抽取 300 支该公司品牌的香烟, 算得平均尼古丁含量为 14.6, 标准差为 3.8 毫克。求检验该组织的断言: “ $\mu$  大于 14” 时统计检验的显著性水平。如果  $\alpha = 0.01$ , 数据中是否有明显的证据支持该组织的断言?

**5.47(心理学)** 为了进行一项心理学实验, 需要调查从实施刺激至观察到某个特定的反应所经过的时间(时间迟滞)。随机抽取 36 人, 对其实施刺激, 然后记录下时间迟滞。样本均值和样本标准差分别为 2.2 秒和 0.57 秒。是否有明显的证据显示, 对于那些可能被实施刺激的人所构成的假定总体, 其平均时间迟滞不等于 1.6 秒吗? 给定  $\alpha = 0.05$ 。检验的显著性水平是多少?

## 5.7 正态总体均值 $\mu$ 的统计推断, $\sigma$ 未知

早先本章所介绍的关于  $\mu$  的估计和检验方法都是建立在这样一个基础上, 即假定总体方差已知或者我们能够观测到足够的样本使得  $s$  可作为  $\sigma$  的一个很好的估计。在本节中, 我们要介绍在总体近似服从正态总体分布的条件下, 当  $\sigma$  未知时的一种检验方法, 而不管样本容量的大小。在第 5.8 节, 我们将提供总体分布非正态时的一种检验方法。考虑下面的例子。研究人员想知道患有某种罕见疾病的病人服药 1 小时之后血液中药物的平均浓度。对于这种情况, 在给定时间内, 我们不可能得到 30 个或更多的观测。用什么样的检验方法对  $\mu$  作出推断呢?

上世纪初, W. S. Gosset 遇到了同样的问题。作为 Guinness 酿酒厂的一名化学家, 他要判断各种啤酒的平均质量, 但是他无法得到大样本来作出结论。

Gosset 认为, 当他在小样本情况下使用检验统计量

$$z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}},$$

并用  $s$  代替  $\sigma$  时, 他错误地拒绝零假设  $H_0: \mu = \mu_0$  的比率比给定的  $\alpha$  略高一些。

这一问题激起了他的兴趣, 他开始推导检验统计量  $z = \frac{\bar{y} - \mu_0}{s / \sqrt{n}}$  的分布及  $n < 30$  时的分位数。

例如, 假设试验者规定  $\alpha$  为标准的水平, 比如 0.05, 那么他或她预料错误地拒绝零假设的几率接近于 1:20。然而, Gosset 证明该检验犯第一类错误的概率略高于标准水平  $\alpha$ 。他以笔名 Student 发表了他的研究成果, 因为在当时以真实姓名发表自己的成果违反了公司的政策。检验统计量

$$\frac{\bar{y} - \mu_0}{s / \sqrt{n}}$$

被称为  $t$  统计量, 其分布被称为**学生  $t$  分布**, 或者简称为**学生  $t$** (见图 5.16)。

虽然统计量  $\frac{\bar{y} - \mu_0}{s / \sqrt{n}}$  仅在样本来自于正态总体时才服从  $t$  分布, 但是当样本来

自于服从丘形分布的总体时,  $t$  分布提供了统计量  $\frac{\bar{y} - \mu_0}{s / \sqrt{n}}$  分布的一个合理近似。

我们将  $t$  分布的性质概括如下。

### $t$ 分布的性质

1.  $t$  分布有很多, 每一个  $t$  分布对应一个参数, 称该参数为自由度 ( $df$ ) (见图 5.16)。

2. 和  $z$  分布一样,  $t$  分布关于 0 点对称, 因此均值等于 0。

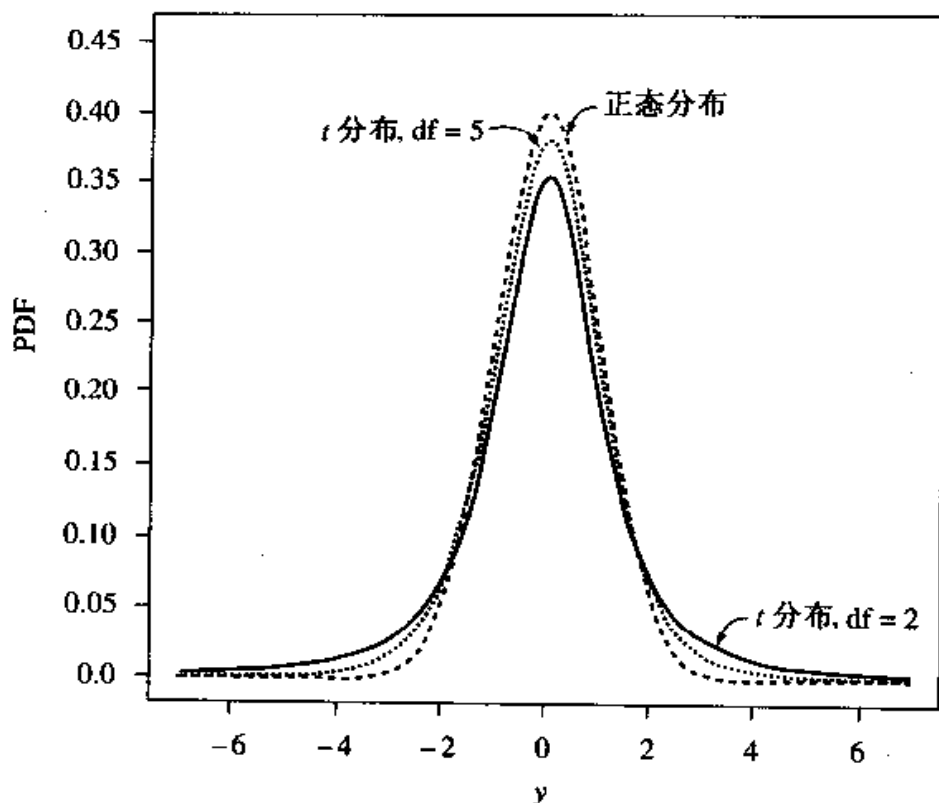


图 5.16  $t$  分布和标准正态分布的概率密度函数(PDF)

3.  $t$  分布的方差等于  $df/(df-2)$ , 因此比方差为 1 的  $z$  分布离散程度更大。(见图 5.16)

4. 随着自由度  $df$  的增加,  $t$  分布将逼近于  $z$  分布。(易见随着自由度  $df$  的增加, 方差  $df/(df-2)$  接近于 1.)

5. 令  $t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$ , 则  $t$  服从自由度为  $df = n - 1$  的  $t$  分布, 且随着  $n$  增加,  $t$  的分布逼近  $z$  分布。

“自由度”一词现在听起来好像难以理解, 但其中的思想是很有用的。“自由度”一词的严格定义需要高深的数学知识, 我们予以回避; 从不太严格的意义上讲, 自由度表示用  $s$  估计  $\sigma$  时使用了多少条信息。对于容量为  $n$  的样本, 其标准差  $s$  是根据离差  $y_i - \bar{y}$  计算的。由于  $\sum (y_i - \bar{y}) = 0$  总是成立, 因此, 如果  $n - 1$  个离差已知, 那么可以根据总和等于 0 直接求出第  $n$  个离差, 这样第  $n$  个离差就没有提供任何信息。因此, 容量为  $n$  的样本包含了  $n - 1$  条关于  $\sigma$  的信息。对自由度的第二种解释就是  $\sigma$  度量了总体关于  $\mu$  的离散程度, 所以在估计  $\sigma$  之前, 我们必须首先估计  $\mu$ 。因此样本数据中关于估计  $\sigma$  的信息有  $n - 1$  条。这相当于原始数据的个数减去在估计  $\sigma$  之前所要估计的参数的个数。

由于  $t$  分布的对称性, 附录中表 2 只列出了  $t$  分布的上分位数。自由度列于表的左侧, 表中的数为  $t$  值, 记为  $t_a$ , 它表示其右侧面积等于  $a$ , 见图 5.17。各种各样的  $a$  值列于附录中的表 2 的顶部。因此, 例如当  $df=7$ , 右侧面积为 0.05 的  $t$  值是 1.895 (查表中  $a=0.05$  列,  $df=7$  行)。由于当  $df$  近似于  $\infty$  时,  $t$  分布近似于  $z$  分布, 因此, 表 2 的最后一行数等于  $z_a$ 。这样, 利用表 2 的最后一行值, 我们可以很快地求出  $z_a$ 。

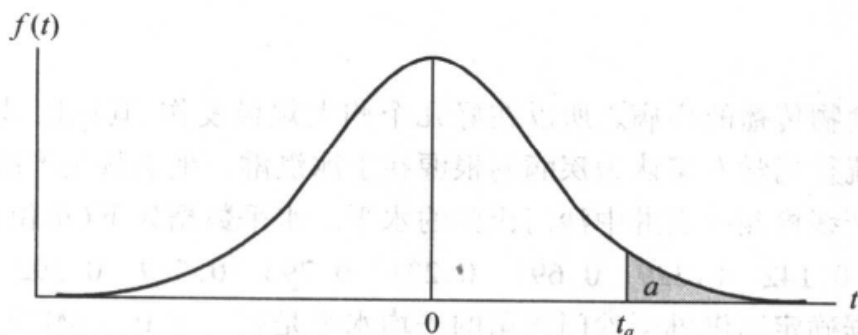


图 5.17 附录表 2 中的  $t$  分布的面积示意图

我们可以用  $t$  分布对总体均值  $\mu$  进行统计推断。下面对关于  $\mu$  的样本检验进行简单总结。本章早先讨论的  $z$  检验和这里的  $t$  检验惟一不同在于, 这里用  $s$  代替了  $\sigma$ 。任何时候, 只要  $\sigma$  未知, 并且  $y$  值的分布是丘形的, 就应该用  $t$  检验而不是  $z$  检验。

### $\sigma$ 未知时, 正态总体均值 $\mu$ 的统计检验

假设:

情形 1.  $H_0: \mu \leq \mu_0$        $H_a: \mu > \mu_0$       (右侧检验)

情形 2.  $H_0: \mu \geq \mu_0$        $H_a: \mu < \mu_0$       (左侧检验)

情形 3.  $H_0: \mu = \mu_0$        $H_a: \mu \neq \mu_0$       (双侧检验)

$$\text{T.S.: } t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

R.R.: 第一类错误的概率为  $\alpha$ ,  $df = n - 1$

情形 1. 如果  $t \geq t_a$ , 则拒绝  $H_0$ 。

情形 2. 如果  $t \leq -t_a$ , 则拒绝  $H_0$ 。

情形 3. 如果  $|t| \geq t_{a/2}$ , 则拒绝  $H_0$ 。

显著性水平( $p$  值)

情形 1.  $p$  值 =  $p(t \geq t \text{ 的计算值})$

情形 2.  $p$  值 =  $p(t \leq t \text{ 的计算值})$

情形 3.  $p$  值 =  $2p(t \geq |t \text{ 的计算值}|)$

注意  $\alpha$  表示  $t$  分布尾部的面积。对于第一类错误概率等于  $\alpha$  的单侧检验, 我们用附录表 2 中对应于  $\alpha = \alpha$ ,  $df = n - 1$  的  $t$  值确定拒绝域。然而, 对于双侧检验, 我们应利用表 2 中对应于  $\alpha = \alpha/2$ ,  $df = n - 1$  的  $t$  值。

因此, 对于右侧检验, 如果  $t$  的计算值大于表 2 中对应于  $\alpha = \alpha$ ,  $df = n - 1$  的  $t$  值, 那么就拒绝零假设。同样地, 对于双侧检验, 如果  $|t|$  大于表 2 中对应于  $\alpha = \alpha/2$ ,  $df = n - 1$  的  $t$  值, 就拒绝零假设。

### 例 5.15

一种经食物传播的疾病之所以在好几个州大规模发作, 其原因被归咎于肠炎沙门氏菌。流行病学专家认为疾病的根源在于冰淇淋。他们从生产冰淇淋的公司抽取 9 条生产线检测冰淇淋中沙门氏菌的水平。水平数据如下(单位: MPN/g)。

0.593 0.142 0.329 0.691 0.231 0.793 0.519 0.392 0.418

利用这些数据确定冰淇淋中沙门氏菌的平均水平是否大于 0.3 MPN/g, 而这一水平被认为是危险的。令  $\alpha = 0.01$ 。

**解答** 本例中的零假设和研究假设为

$$H_0: \mu \leq 0.3$$

$$H_a: \mu > 0.3$$

因为样本容量较小, 我们需要验证数据是否随机抽自于正态总体。图 5.18 是该数据的正态概率图。所有 9 个点几乎在一条直线上。我们认为正态性条件看起来是

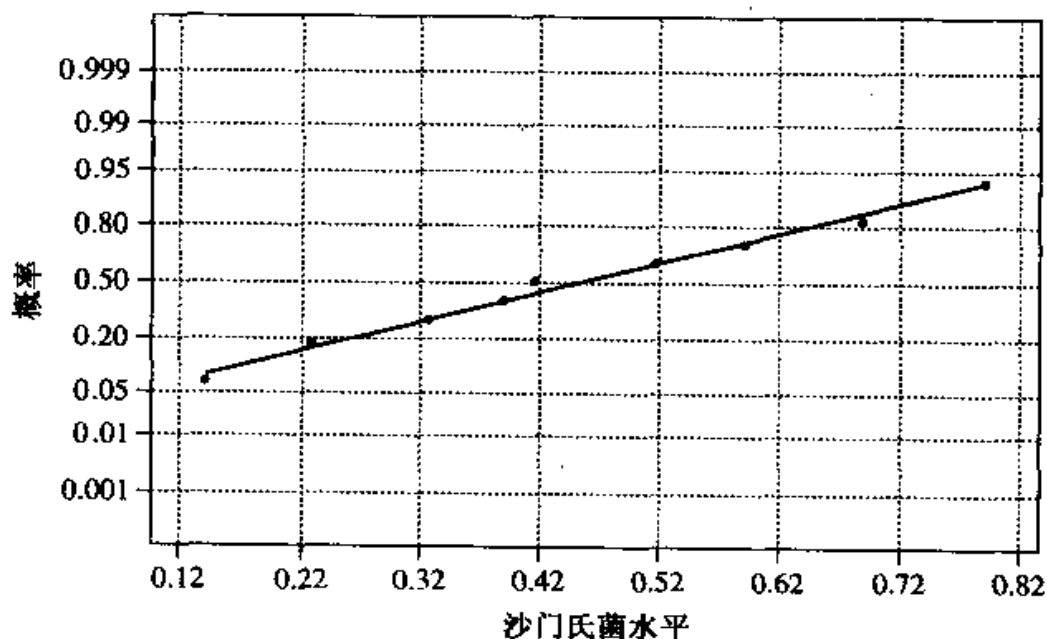


图 5.18 沙门氏菌数据的正态概率图

满足的。在确定拒绝域, 计算检验统计量之前, 我们首先必须计算样本均值和标准差。经计算得:

$$\bar{y} = 0.456 \quad s = 0.2128$$

对应于  $\alpha = 0.01$  的拒绝域为

R.R.: 若  $t > 2.896$ , 则拒绝  $H_0$ 。

其中, 根据附录表 2, 当自由度  $df = 9 - 1 = 8$  时,  $t_{0.01}$  值为 2.896。  $t$  的计算值为

$$t = \frac{\bar{y} - \mu_0}{s / \sqrt{n}} = \frac{0.456 - 0.3}{0.2128 / \sqrt{9}} = 2.21$$

$t$  的观察值并不大于 2.896, 所以我们没有充分的证据显示冰淇淋中沙门氏菌的平均水平大于 0.3 MPN/g。检验的显著性水平为

$$p \text{ 值} = p(t > t \text{ 的计算值}) = p(t > 2.21)$$

由于对于每个  $df$ ,  $t$  表只列出少数几个面积  $\alpha$ , 所以我们至多只能给出  $p$  值的一个界限。从表 2 得, 当  $df = 8$  时,  $t_{0.05} = 1.860$ ,  $t_{0.025} = 2.306$ 。由于  $t$  的计算值 = 2.21, 因此  $0.025 < p \text{ 值} < 0.05$ 。然而,  $\alpha = 0.01 < 0.025 < p \text{ 值}$ 。我们仍然能得出结论:  $p \text{ 值} > \alpha$ , 因此我们不能拒绝  $H_0$ 。下面给出的用 Minitab 软件算得的结果表明  $p \text{ 值} = 0.029$ 。

#### T-Test of the Mean

Test of  $\mu \leq 0.3000$  vs  $\mu > 0.3000$

Variable	N	Mean	StDev	SE Mean	T	P
Sal.Lev	9	0.4564	0.2128	0.0709	2.21	0.029

#### T Confidence Intervals

Variable	N	Mean	StDev	SE Mean	95.0% CI
Sal.Lev	9	0.4564	0.2128	0.0709	(0.2928, 0.6201)

正如我们以前所说, 为了作出沙门氏菌的水平小于或等于 0.3 的结论, 我们必须计算对应于  $H_a$  中某些关键的  $\mu$  值的第二类错误概率。这些计算比  $z$  检验的计算更复杂一些。我们将用一组图来确定  $\beta(\mu_a)$ 。  $\beta(\mu_a)$  依赖于三个量, 即  $df = n - 1$ ,  $\alpha$  以及以  $\sigma$  为单位的  $\mu_a$  至  $\mu_0$  的距离  $d$ ,

$$d = \frac{|\mu_a - \mu_0|}{\sigma}$$

因此, 为了确定  $\beta(\mu_a)$ , 我们必须给定  $\alpha$ ,  $\mu_a$ , 并提供  $\sigma$  的估计。然后计算  $d$  和  $df = n - 1$ , 据此在图上确定  $\beta(\mu_a)$ 。无论是单侧假设还是双侧假设, 对于  $df = 8$ ,  $\alpha = 0.01$  和 0.05, 以及不同的  $d$  和  $df$  值, 附录表 3 均列出了  $\beta(\mu_a)$  的曲线图。

### 例 5.16

见例 5.15。我们有  $n = 9$ ,  $\alpha = 0.01$  并作了一个单侧检验。因此  $df = 8$ , 如果我们估计  $\sigma \approx 0.25$ , 那么我们能够计算对应于某些  $\mu_a$  的距离  $d$ 。然后根据附录表 3 中的曲线图求出  $\beta(\mu_a)$ 。图 5.19 即是本例中所需要的曲线图。为了说明用图计算的过程, 令  $\mu_a = 0.45$ , 那么

$$d = \frac{|\mu_a - \mu_0|}{\sigma} = \frac{|0.45 - 0.3|}{0.25} = 0.6$$

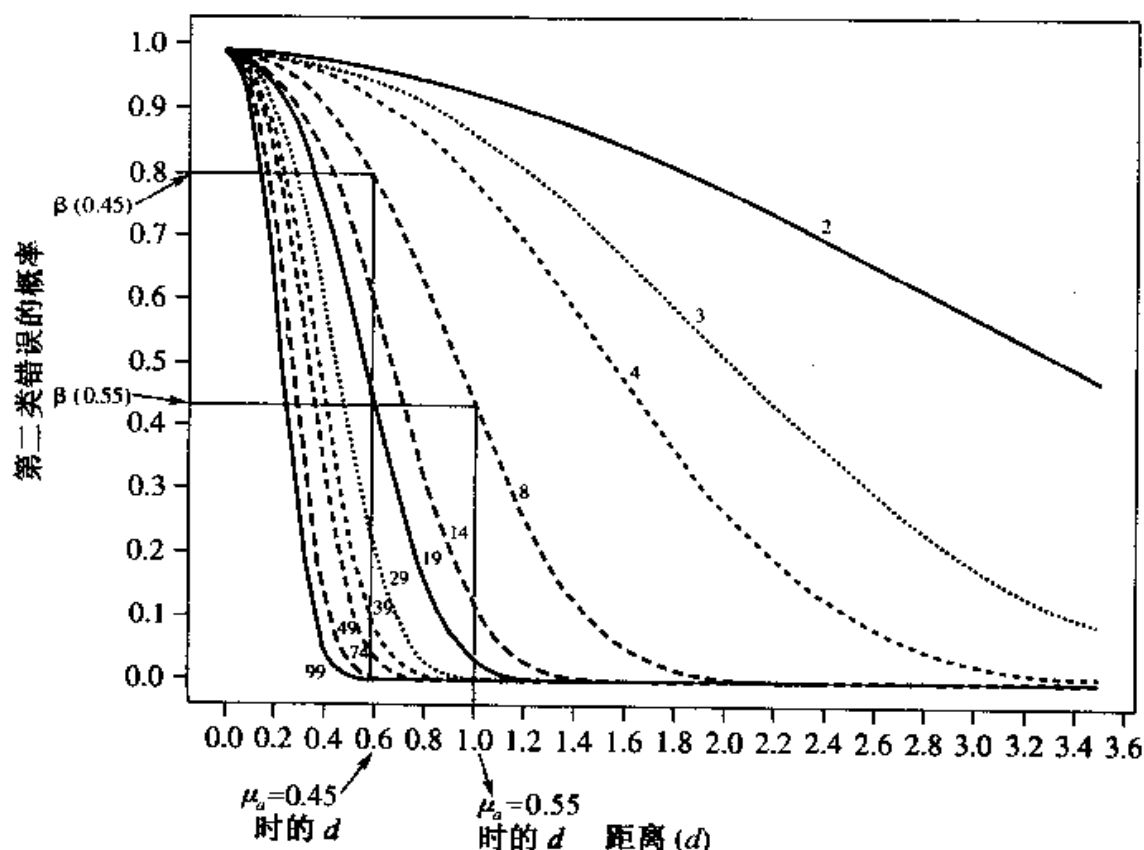


图 5.19  $\alpha = 0.01$ , 单侧检验时第二类错误的概率

从水平轴上  $d = 0.6$  开始画一条垂直线至标有自由度等于 8 的这条线。然后以交点的高度在垂直轴上确定  $\beta(\mu_a)$  值, 于是  $\beta(0.45) = 0.79$ 。同样地, 为了求  $\beta(0.55)$ , 首先计算出  $d = 1.0$ , 从  $d = 1.0$  处画一条垂直线至标有 8 的这条线, 在垂直轴上找到 0.43 这点, 因此  $\beta(0.55) = 0.43$ 。表 5.5 列出了一些  $\beta(\mu_a)$  值。因为对于显著地大于  $\mu_0 = 0.3$  的某些  $\mu_a$ ,  $\beta(\mu_a)$  的值较大, 例如  $\beta(0.6) = 0.26$ , 因此, 我们不能说  $\mu$  小于或等于 0.3, 但是我们可以说数据不支持  $\mu$  大于 0.3 这一论点。

当  $\sigma$  未知时, 除了能够对  $\mu$  进行统计检验外, 我们还能利用  $t$  构造置信区间。

$\sigma$  未知时  $\mu$  的置信区间和  $\sigma$  已知时相应的  $\mu$  置信区间是同样的, 只要将其中的  $z$  用  $t$  代替, 将  $\sigma$  用  $s$  代替即可。  $\sigma$  未知时  $\mu$  的置信区间为

表 5.5 第二类错误的概率

$\mu_a$	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8
$d$	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
$\beta(\mu_a)$	0.97	0.91	0.79	0.63	0.43	0.26	0.13	0.05	0.02	0.00

$$\bar{y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

注:  $df = n - 1$ , 置信系数为  $(1 - \alpha)$ 。

### 例 5.17

某航空公司欲评价 50 岁以上的飞行员的判断能力。随机抽取  $n = 14$  名 50 岁以上的飞行员, 要求他们判断两个放置在实验室两端相距 20 英尺的标记之间的距离。下面列出的样本数据是指飞行员的判断误差(以英尺计)。

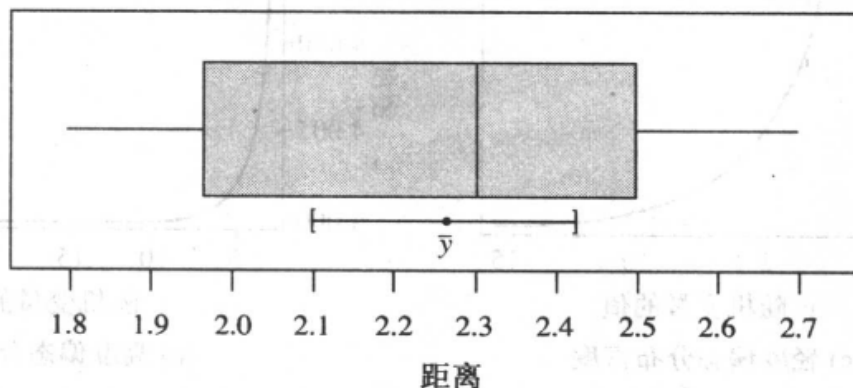
2.7    2.4    1.9    2.6    2.4    1.9    2.3  
2.2    2.5    2.3    1.8    2.5    2.0    2.2

利用样本数据求  $\mu$  的 95% 置信区间, 其中  $\mu$  表示 50 岁以上飞行员对距离的平均判断误差。

**解答** 在求  $\mu$  的 95% 置信区间之前, 我们首先要通过作正态概率图或盒形图来考察数据的正态性。图 5.20 是 14 个数据的盒形图。中位数线接近于盒的中心, 左右两边的须长近似相等, 没有异常值。数据可以看成从正态分布抽取的样本。因此, 用  $t$  分布构造置信区间是合适的。经计算得:

$$\bar{y} = 2.26 \quad s = 0.28$$

查附录表 2 得, 对应于  $\alpha = 0.02$ ,  $df = 13$  的  $t$  值为 2.160。因此  $\mu$  的 95% 置信区间

图 5.20 距离的盒形图(附有均值的 95%  $t$  置信区间)



为

$$\bar{y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \quad \text{或} \quad 2.26 \pm 2.160 \frac{0.28}{\sqrt{14}}$$

即置信区间为  $2.26 \pm 0.16$ , 或 2.10 至 2.42。因此我们有 95% 的把握相信飞行员对距离的平均判断误差在 2.10 英尺和 2.42 英尺之间。

本节中, 我们已经从数学上作了正式的假定, 即假定总体服从正态分布。在实际中没有哪个总体严格服从正态分布。总体分布的非正态性如何影响基于  $t$  分布的统计推断呢?

当总体被认为服从非正态分布时, 有两个问题要考虑。首先, 它服从哪种形式的非正态分布? 其次, 这些非正态分布的形式对  $t$  分布的方法有何影响? 和正态分布最重要的偏离就是偏态分布和厚尾分布。厚尾分布基本上是对称的, 但相对于正态分布, 它具有更多的异常值。图 5.21 显示了四种分布: 图 5.21(a) 是标准正态分布, 图 5.21(b) 是厚尾分布 ( $df=3$  的  $t$  分布), 图 5.21(c) 是中等程度向右偏

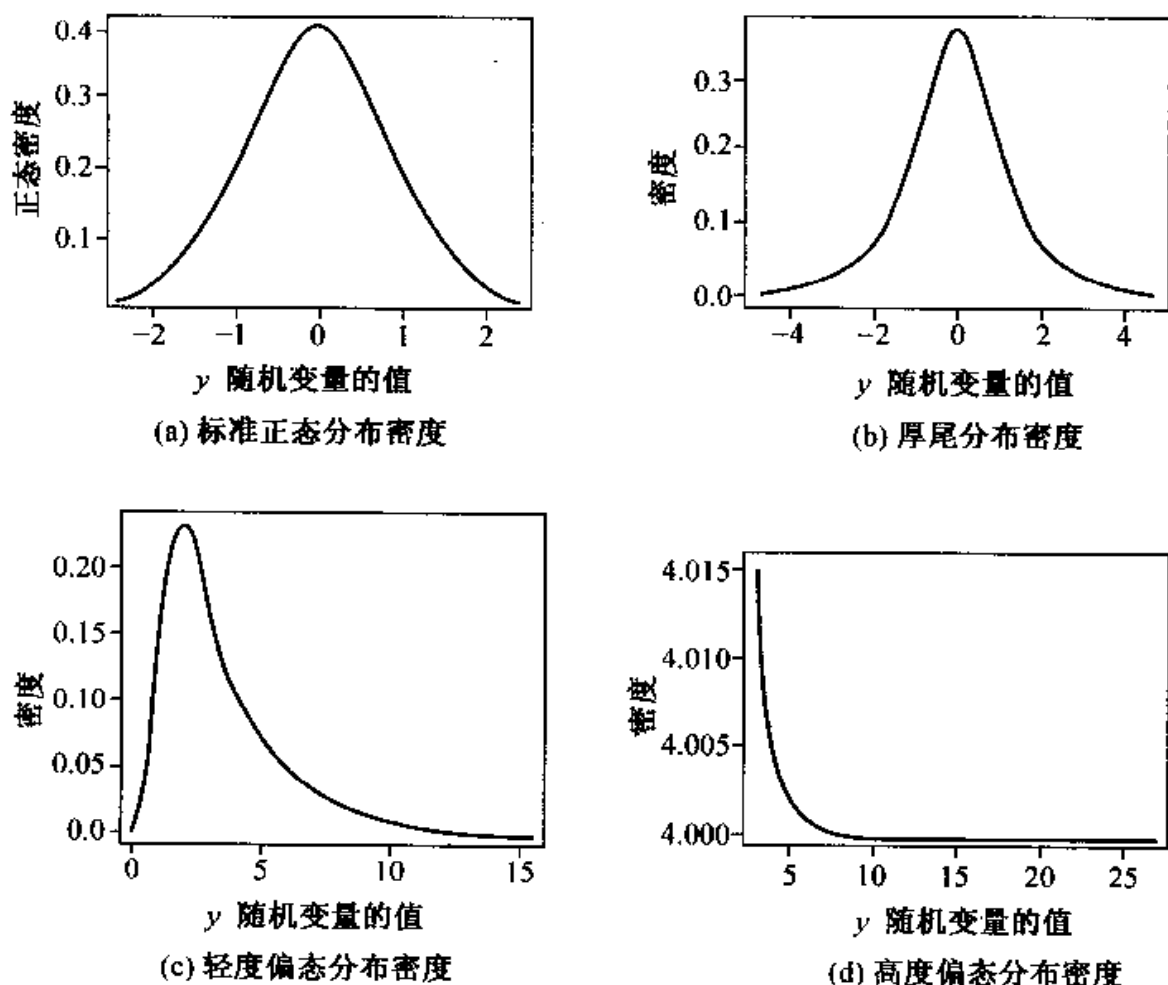


图 5.21

倚的分布,图 5.21(d)是严重向右偏倚的分布。

为了评估如偏态或厚尾这样的非正态性的影响,我们考察对于这些形式的非正态性, $t$  分布方法是否仍然基本正确,是否还有其他更有效的方法。例如,不妨说对于厚尾分布,即使基于  $t$  分布的检验方法给出近似正确的结果,如果我们对总体中位数而不是对总体均值  $\mu$  进行检验,那么,也可能获得一种检验方法,具有更精确的第一类错误概率和更大的功效。并且,在厚尾分布或高度偏倚的分布的情形下,用中位数代表总体的中心比  $\mu$  更合适。

$t$  分布的近似正确性问题已经被广泛研究。一般来说,即使总体分布是厚尾的, $t$  方法的概率,特别是置信区间的置信水平和统计检验的第一类错误概率已经被发现是相当精确的。然而,若如图 5.21(b)所示,总体分布是严重厚尾的,则假设检验的第一类错误的概率比给定水平偏小,这样导致检验的功效更低,从而有更大的机会犯第二类错误。而偏态,特别是在小样本的情况下,甚至对第一类和第二类错误的概率有更大的影响。当我们从正态总体中抽样时, $t$  统计量的抽样分布是对称的。然而,当我们从严重偏态的总体分布中抽样时, $t$  统计量的抽样分布是偏态的,而不是对称的。虽然,随着样本容量的增加,偏态的程度逐渐减少,但是没有哪种方法能够确定样本容量为多少时  $t$  统计量的样本分布变成对称分布。

因此,当样本容量小于 20,并且总体分布如图 5.21(b)(c)(d)时,名义上  $\alpha = 0.05$  的检验水平实际上可能等于 0.01 或更小。进一步地,其检验功效将大大低于当总体分布为正态分布时的检验功效,因此导致第二类错误增加。表 5.6 列出了偏态和厚尾影响检验水平和检验功效的模拟研究结果。表中的值是在  $\alpha = 0.05$  下,对于假设  $H_0: \mu \leq \mu_0$   $H_a: \mu > \mu_0$  的检验功效。功效值是根据漂移量  $d = |\mu_a - \mu_0|/\sigma$  而计算的。表中列出了  $d = 0, 0.2, 0.6, 0.8$  且样本容量分别为  $n = 10, 15, 20$  时的功效。当  $d = 0$  时,对于每一类型的总体分布均给出了相应的检验水平,我们当将其和 0.05 进行比较。当  $d > 0$  时,将功效与相应的正态总体下的功效进行对比。通过对比我们发现,当样本来自于轻度偏态分布和厚尾分布时,检验水平稍微小于 0.05,且  $n = 20$  时,水平几乎等于 0.05。然而,当样本来自于重度偏态分布时,即使  $n = 20$ ,水平也仅仅为 0.011。样本来自于厚尾分布和重度偏态分布的功效明显小于样本来自于正态分布时的功效。因此,这种检验不太可能正确地检测出备择假设为真,甚至当  $n = 20$  时,功效仍然很低。当样本来自于轻度偏态分布时,其功效非常接近于样本来自于正态分布的功效。

因为当样本取自于偏态分布总体且样本容量较小时, $t$  方法的功效较低,所以人们研究出一些新方法,它们不受总体分布的偏态和极端厚尾的影响。这些方法被称为估计和推断的**稳健方法**。第 5.6 节和第六章将分别介绍两种稳健方法,即符号检验和 Wilcoxon 符号秩检验。当总体分布严重偏离正态时,这两种检验均比  $t$  检验更有效。并且,它们保持给定的水平  $\alpha$  不变,而不像  $t$  检验那样,当数据严

重非正态时,其真实的  $\alpha$  值与给定的  $\alpha$  值差别较大。同样的说法也适用于均值的置信区间。当总体分布严重偏态时,名义上  $100(1-\alpha)\%$  置信区间的覆盖概率比  $100(1-\alpha)$  要小很多。

表 5.6  $t$  检验的水平和功效

总 体 分 布	$n = 10$				$n = 15$				$n = 20$			
	漂移量 $d$				漂移量 $d$				漂移量 $d$			
	0	0.2	0.6	0.8	0	0.2	0.6	0.8	0	0.2	0.6	0.8
正态	0.05	0.145	0.543	0.754	0.05	0.182	0.714	0.903	0.05	0.217	0.827	0.964
厚尾	0.035	0.104	0.371	0.510	0.049	0.115	0.456	0.648	0.045	0.163	0.554	0.736
轻度偏倚	0.025	0.079	0.437	0.672	0.037	0.129	0.614	0.864	0.041	0.159	0.762	0.935
严重偏倚	0.007	0.055	0.277	0.463	0.006	0.078	0.515	0.733	0.011	0.104	0.658	0.873

那么,一个非专业人员如何做呢?首先,用图对数据进行检查。盒形图或正态概率图能揭示出任何严重的偏态或极端异常值。如果从图形中并没有发现严重偏态或很多异常值,那么名义上的  $t$  分布概率应基本正确。因此,假设检验的水平和功效,以及置信区间覆盖的概率也应是相当精确的。如果图形揭示出严重偏态或厚尾,那么基于  $t$  分布的  $t$  检验方法和置信区间将非常值得怀疑。在这种情况下,中位数比均值更适合作为总体中心的度量。在第 5.8 节,我们将研究总体中位数的假设检验和置信区间。当总体分布严重偏态或厚尾时,第 5.8 节所介绍的方法将保持置信区间覆盖的名义概率以及给定的检验水平  $\alpha$  不变。

现在,我们来完成关于脂肪中卡路里含量百分比的案例分析。

**分析案例中的数据** 研究的其中之一就是估计女性护士从其食用的脂肪中摄取的卡路里的平均百分数。并且,研究人员希望估计平均数是否大于推荐的值 30%。回想一下,我们有 168 名妇女的随机样本,并且记录了用问卷调查所得的卡路里百分数(PCF)。在构造置信区间或检验假设之前,首先,我们必须检查数据是否来自于正态总体。从图 5.1 的茎叶图和图 5.5 的正态概率图可见,可以认为数据来自于正态分布。PCF 数据的均值和标准差分别为  $\bar{y} = 36.92$ ,  $s = 6.73$ 。下面我们构造护士总体平均 PCF 的 95% 置信区间,置信区间为:

$$36.92 \pm t_{0.025, 167} \frac{6.73}{\sqrt{168}}, 36.92 \pm 1.974 \frac{6.73}{\sqrt{168}} \text{ 或 } 36.92 \pm 1.02$$

我们有 95% 的把握相信护士中平均 PCF 值在 35.90 与 37.94 之间。因此,有证据表明护士平均 PCF 值大于推荐的值 30%。下一步,我们将正式检验假设

$$H_0: \mu \leq 30 \quad H_a: \mu > 30$$

因为认为数据来自于正态分布,并且在任何情况下,只要样本容量充分大,我们都能利用  $t$  统计量得拒绝域:

R.R.: 对于单侧  $t$  检验, 给定  $\alpha = 0.05$ , 如果

$$t = \frac{\bar{y} - 30}{s/\sqrt{16}} \geq t_{0.05, 167} = 1.654, \text{ 则拒绝 } H_0.$$

因为  $t = (36.92 - 30)/(6.73/\sqrt{168}) = 13.33$ , 所以我们拒绝  $H_0$ 。由于检验的  $p$  值基本上等于 0, 因此, 我们认为平均 PCF 显著地大于 30, 从而我们有充分的理由认为护士的平均 PCF 大于推荐值 30。该领域的专家面对高于推荐值 5.90 到 7.94 个单位的平均 PCF 值, 必须确定下一步怎么办。

**报告结论** 我们必须写一份报告对我们的研究结果进行总结, 报告包括以下几项内容:

1. 叙述研究目的。
2. 说明研究的设计及数据收集方案。
3. 用数字和图的形式对数据集进行归纳总结。
4. 说明所有的推断方法:
  - $t$  检验。
  - 基于  $t$  方法的总体均值的置信区间。
  - 验证使用推断方法所必要的条件。
5. 讨论所得的结果和做出的结论。
6. 相对于以前的研究, 解释本次研究中的发现。
7. 对未来的研究给出建议。
8. 列出数据集合。

## 练习

### 基本技能

**5.48** 当  $n < 30$ ,  $\sigma$  未知时, 5.4 节中的  $z$  检验为什么不适合检验  $H_0: \mu \geq \mu_0$ ?

**5.49** 给定  $\alpha = 0.05$ , 试针对下面的情况建立基于  $t$  检验的拒绝域:

a.  $H_0: \mu < \mu_0, n = 15$ 。

b.  $H_0: \mu \neq \mu_0, n = 23$ 。

c.  $H_0: \mu > \mu_0, n = 6$ 。

**5.50** 给定  $\alpha = 0.01$ , 重复练习 5.49。

**5.51** 关于假设  $H_0: \mu \leq 15$   $H_0: \mu > 15$  的  $t$  检验的样本数据为  $\bar{y} = 16.2$ ,  $s = 3.1$ ,  $n = 18$ 。给定  $\alpha = 0.05$ , 试给出你的结论。

## 应用

**5.52(教育)** 某小学正在评估在四年级实施的一项新的阅读计划。随机抽取 20 名学生,对他们进行一次测验,以考察他们的阅读速度和阅读理解能力。测试内容为阅读一篇固定长度的用于标准测验的短文。经测验,阅读速度(以分钟计)和阅读理解成绩(百分制)如下:

学生	1	2	3	4	5	6	7	8	9	10	11	12
速度	5	7	15	12	8	7	10	11	9	13	10	6
理解	60	76	76	90	81	75	95	98	88	73	90	66
学生	13	14	15	16	17	18	19	20	$n$	$\bar{y}$	$s$	
速度	11	8	10	8	7	6	11	8	20	9.10	2.573	
理解	91	83	100	85	76	69	91	78	20	82.05	10.88	

- 利用阅读速度数据,求该小学所有四年级学生平均阅读速度  $\mu$  的 95% 置信区间。
- 画出阅读速度的正态概率图和盒形图,考察数据是否来自于正态总体。
- 解释(a)中的区间估计。
- 如果用 98% 置信区间而不是 95% 置信区间,你的结论有变化吗?

**5.53** 参看练习 5.52。根据阅读成绩数据,是否有充分证据表明四年级的阅读成绩大于 80 分? 这里的 80 分是前一年全州范围内可比学生的平均成绩。试给出你的检验的显著性水平,并解释你的结果。

**5.54** 参看练习 5.53。

- 根据数据,你能发现学生的阅读能力和阅读速度之间存在什么关系吗?
- 你对关于新的阅读计划的评价结果有何看法?

**5.55(商业)** 某廉价轮胎生产厂家声称他们的轮胎在磨破之前至少能够行使 35,000 英里,现有一消费者检测机构欲评价其真实性。为了知道该厂家轮胎的平均行驶里程,检测机构从其仓库中随机抽取 60 只轮胎,将他们安装到 15 辆小汽车上,然后在 2 英里长的椭圆型试车跑道上进行试车。下表列出了轮胎在磨破之前的行驶里程数(以千英里计)。

汽车	1	2	3	4	5	6	7	8	9
行驶里程	25	27	35	42	28	37	40	31	29

汽车	10	11	12	13	14	15	$n$	$\bar{y}$	$s$
行驶里程	33	30	26	31	28	30	15	31.47	5.04

- 求在轮胎在磨破之前, 其平均行驶里程  $\mu$  的 99% 置信区间。
- 给定  $\alpha = 0.01$ , 是否有充分证据显示厂家的声音是错误的? 检验的显著性水平是多少? 并对你的结果加以解释。

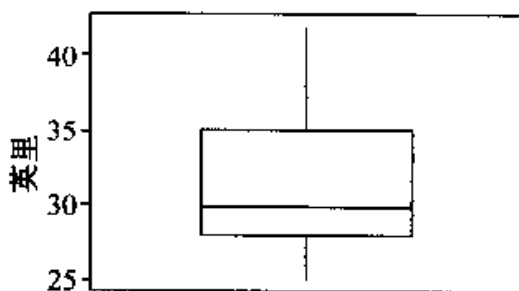
**5.56** 参看练习 5.55。下面是 Minitab 输出结果, 试将你的结果和计算机程序给出的结果进行比较。

- 正态假定有效吗?
- 你所得到的关于  $p$  值的界限与准确的  $p$  值相差多少?
- $\mu$  的区间估计和假设检验的结论是否相互矛盾?

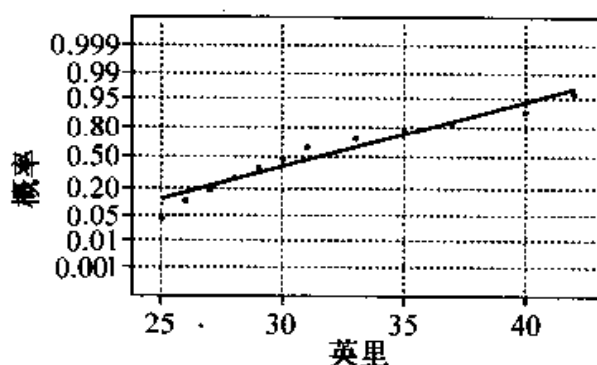
Test of  $\mu = 35.00$  vs  $\mu < 35.00$

Variable	N	Mean	StDev	SE Mean	T	P	99.0 % CI
Miles	15	31.47	5.04	1.30	-2.71	0.0084	27.59, 35.3

轮胎数据的盒形图



轮胎数据的正态检验



**5.57(环境)** 水体中的污水和工业污染的多少会通过减少水中被溶解的氧气而影响水体的水质, 生物的生长与生存有赖于这种氧气。两个月内, 从污水处理厂下游 1 英里处的一条小河里取得 8 个水样。检测水样中溶解的氧气含量, 数据列于下表。根据最近的研究, 为了保证鱼的生存, 水中溶解的氧气的平均含量必须达到百万分之五, 即 5.0ppm。

水样	1	2	3	4	5	6	7	8	$n$	$\bar{y}$	$s$
氧(ppm)	5.1	4.9	5.6	4.2	4.8	4.5	5.3	5.2	8	4.95	0.45

- 求两个月期间平均氧气含量的 95% 置信区间。
- 根据(a)中的置信区间, 平均氧气含量看起来小于 5.0ppm 吗?
- 检验研究假设: 平均氧气含量小于 5.0ppm。检验的显著性水平等于多少? 并对你的结果加以解释。

**5.58(工程)** 某从事纸张回收的商贩将空挂车放置在各个地点。一些人把旧报纸和旧杂志扔进车里, 逐渐将其填满, 然后, 商贩每隔不同的时间将这些挂车运走。其中一种安排就是每两周回收一次。如果平均每两周回收纸张超过 1,600 立方英尺, 那么这种安排就是可取的。该商贩记录了某地点每两周回收一次, 共 18 次回收量的数据如下:

1,660 1,820 1,590 1,440 1,730 1,680 1,750 1,720 1,900  
 1,570 1,700 1,900 1,800 1,770 2,010 1,580 1,620 1,690  
 $\bar{y} = 1,718.3$        $s = 137.8$

- 假定这 18 次的记录在全年的两周一次的回收记录中是典型的, 是否有充分的证据显示平均回收量  $\mu$  超过 1,600 立方英尺?
- 求  $\mu$  的 95% 置信区间。
- 计算检验统计量的  $p$  值。是否有实足的证据表明  $\mu$  大于 1,600?

**5.59** 观赏性灌木的商业种植者经常要延缓灌木的成长以便在它们售出之前不至于长得太大。对某高凉菜属观赏性灌木施用了一种叫 dikegulac 的生长延缓剂, 以评价该延缓剂的效用。文章“dikegulac alters growth and flowering of klanchoe”(Hort Science(1985), 20:722~724)记述了这些实验的结果。对 10 株灌木施用了 dikegulac, 而另对 10 株灌木未加以施用, 以便确定 dikegulac 对植物生长的影响。在经过处理 13 周之后, 测量了 20 株灌木的高度, 并计算了有关统计量, 列表如下:

	$n$	$\bar{y}$	$s$
未使用	10	43.6	5.7
使用	10	36.1	4.9

- a. 分别求经过处理的灌木的平均高度和未经过处理的灌木的平均高度的置信区间。并解释这些区间。
- b. 两置信区间有重叠吗？你能对作为生长延缓剂的 dikegulac 的效用作何结论？

**5.60(政府)** 某联邦管理机构调查一则广告, 该广告声称一种装置能够增加汽车的汽油行驶里程(单位:mpg, 英里/每加仑)。该联邦机构购买了 10 套这样的装置, 并安装在该机构的汽车上。记录安装前后每辆车的汽油行驶里程。数据如下:

	汽车												
	1	2	3	4	5	6	7	8	9	10	$n$	$\bar{x}$	$s$
安装前(mpg)	19.1	29.9	17.6	20.2	23.5	26.8	21.7	25.7	19.5	28.2	10	23.22	4.25
安装后(mpg)	25.8	23.7	28.7	25.4	32.8	19.2	29.6	22.3	25.7	20.1	10	25.33	4.25
变化(mpg)	6.7	-6.2	11.1	5.2	9.3	-7.6	7.9	-3.4	6.2	-8.1	10	2.11	7.54

试分别求出安装前后每个装置的平均 mpg 的 90% 置信区间, 并解释这些区间。该种装置是否明显地提高了平均行驶里程?

**5.61** 参看练习 5.60。

- a. 安装了装置之后, 汽车的平均每加仑汽油行驶里程似乎明显不同于安装之前。利用里程的改变量数据, 检验在安装了装置之后行驶里程是否有明显增加? 给定  $\alpha = 0.05$
- b. 构造行驶里程平均改变量的 90% 置信区间。在这一区间基础上, 人们能够拒绝假设: 平均改变量为零或为负吗? (注意: 90% 双侧置信区间对应于一个  $\alpha = 0.05$  的单侧检验, 即如果  $\mu_0$  大于置信区间上限, 那么拒绝  $H_0: \mu \leq \mu_0$ )

**5.62** 参看练习 5.60。

- a. 对于几个不同的  $\mu_C$  值, 计算第二类错误的概率, 其中  $\mu_C$  为行驶里程的平均改变量。这些值如何影响你在练习 5.61 中所得的结论?
- b. 试对练习 5.60 中的研究方法提出一些改变的建议。

## 5.8 关于中位数的推断

当总体分布严重偏倚或严重厚尾时, 中位数比均值更适合表示总体的中心。更进一步地, 正如 5.7 节所指出的, 当样本取自这样的总体, 且样本容量较小时, 不



适合用  $t$  方法构造置信区间和作均值的假设检验。本节我们将介绍关于总体中位数的假设检验和区间估计问题,且适用于各种总体分布。

总体中位数  $M$  的估计基于第三章所讨论的顺序统计量。回想一下,如果用  $y_1, y_2, \dots, y_n$  表示容量为  $n$  的随机样本,那么将其从小到大排序所得到的各个统计量就是顺序统计量。令  $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$  表示数据按顺序排列。因此  $y_{(1)}$  是最小值,  $y_{(n)}$  是最大值。总体中位数的估计量是样本中位数  $\hat{M}$ 。 $\hat{M}$  的计算如下。

如果  $n$  为奇数,那么  $\hat{M} = y_{(m)}$ , 其中  $m = (n+1)/2$ 。

如果  $n$  为偶数,那么  $M\hat{M} = (y_{(m)} + y_{(m+1)})/2$ , 其中  $m = n/2$ 。

为了考虑作为  $M$  的估计  $\hat{M}$  的变异性,下面我们构造  $M$  的置信区间。可以用  $\pi = 0.5$  的二项分布构造总体中位数  $M$  的置信区间。

**置信水平至少为  $100(1-\alpha)\%$  的  $M$  的置信区间为**

$$(M_L, M_U) = (y_{(L_{\alpha/2})}, y_{(U_{\alpha/2})})$$

其中  $L_{\alpha/2} = C_{\alpha(2), n}$ ,  $U_{\alpha/2} = n - C_{\alpha(2), n} + 1$

附表 4 列出了  $\pi = 0.5$  的二项分布的分位数  $C_{\alpha(2), n}$  的值。

因为置信限是根据二项分布计算的,而二项分布是离散分布,因此  $(M_L, M_U)$  的置信水平一般来说比给定的  $100(1-\alpha)\%$  略大一些。精确的置信水平为

$$\text{水平} = \Pr(\text{Bin}(n, 0.5) \leq U_{\alpha/2} - 2) - \Pr(\text{Bin}(n, 0.5) \leq L_{\alpha/2})$$

下面将通过例子来说明如何构造置信区间。

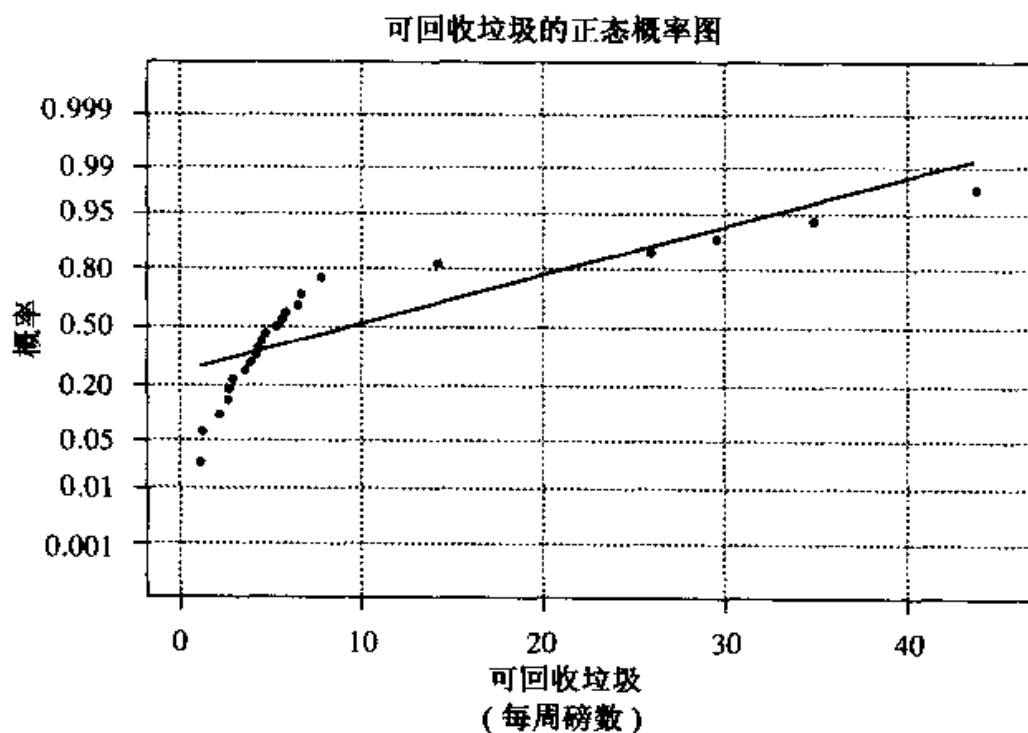
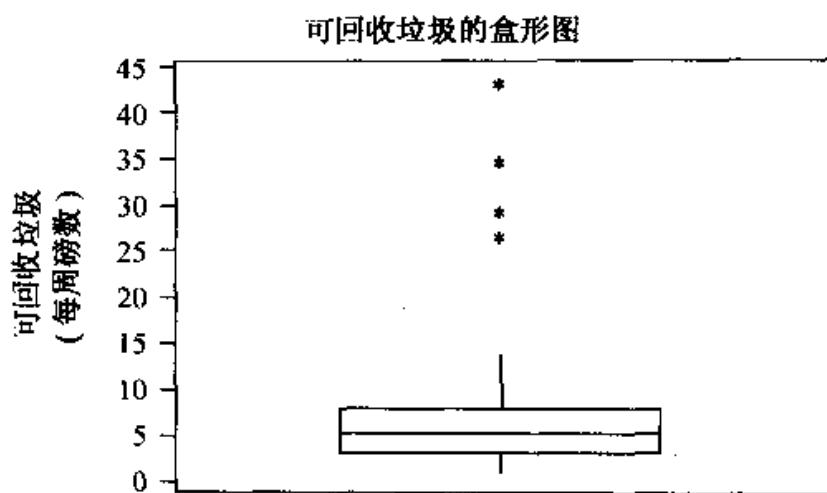
### 例 5.18

某大城市的环卫部门欲研究如何减少那些放在城市垃圾掩埋场的可回收物的数量。通过从垃圾中分离出可回收物,可延长垃圾掩埋场的生命。更重要的是,可大大减少为了纸张生产而砍伐的树的数量以及罐装品所用的铝的数量。根据其他城市回收记录的分析,如果平均每周从每户回收 5 磅以上的可回收物,那么商业回收公司就可通过回收垃圾获利。为了论证回收计划的可行性,随机抽取 25 户。每户每周可回收物的重量(以磅/周计)数据如下。

14.2 5.3 2.9 4.2 1.2 4.3 1.1 2.6 6.7 7.8 25.9 43.8 2.7

5.6 7.8 3.9 4.7 6.5 29.5 2.1 34.8 3.6 5.8 4.5 6.7

试给出一种适当的方法以确定该城市家庭可回收垃圾的数量。



**解答** 可回收垃圾的盒形图和正态概率图均显示数据严重右偏。因此,均值并不适合代表典型家庭潜在可回收垃圾的数量。下面计算样本中位数和置信区间。首先,我们将数据从小到大排序:

1.1 1.2 2.1 2.6 2.7 2.9 3.6 3.9 4.2 4.3 4.5 4.7 5.3  
5.6 5.8 6.5 6.7 7.8 7.8 14.2 25.9 29.5 34.8 43.8

数据的个数为奇数, 所以样本中位数为

$$\hat{M} = y_{((25+1)/2)} = y_{(13)} = 5.3$$

样本均值为  $\bar{y} = 9.53$ 。因此, 25 户中有 20 户每周的可回收垃圾小于样本均值。然而, 25 户中有 12 户垃圾小于样本中位数, 25 户中有 12 户垃圾大于样本中位数。因此, 样本中位数比样本均值更适合代表典型家庭潜在可回收垃圾的数量。下面我们将构造总体中位数的 95% 置信区间。

查表 4, 我们得到  $C_{\alpha(2), n} = C_{0.05, 25} = 7$   
因此,

$$L_{0.025} = C_{0.05, 25} = 7$$

$$U_{0.025} = n - C_{0.05, n} + 1 = 25 - 7 + 1 = 19$$

于是总体中位数的 95% 置信区间为

$$(M_L, M_U) = (y_{(7)}, y_{(19)}) = (3.6, 7.8)$$

根据二项分布, 置信区间的精确水平为

$$\begin{aligned} & \Pr(\text{Bin}(25, 0.5) \leq 19 - 2) - \Pr(\text{Bin}(25, 0.5) \leq 7) \\ &= 0.9784 - 0.0216 = 0.9568, \end{aligned}$$

比期望的水平 95% 略大一点。因此, 我们至少有 95% 的把握相信每户每周可回收垃圾的中位数在 3.6 到 7.8 磅之间。

### 大样本近似

当样本容量  $n$  很大时, 我们可以用正态分布近似二项分布, 从而得到  $C_{\alpha(2), n}$  的近似值。近似值为

$$C_{\alpha(2), n} \approx \frac{n}{2} - z_{\alpha/2} \sqrt{\frac{n}{4}}$$

由于  $C_{\alpha(2), n}$  的近似值一般情况下并不是一个整数。因此, 我们令  $C_{\alpha(2), n}$  为小于或等于其近似值的最大整数。

### 例 5.19

根据例 5.18 的数据, 利用  $C_{\alpha(2), n}$  的近似值, 求中位数的 95% 置信区间。

**解答** 由于  $n = 25, \alpha = 0.05$ 。因此,  $z_{0.05/2} = 1.96$

$$C_{\alpha(2), n} \approx \frac{n}{2} - z_{\alpha/2} \sqrt{\frac{n}{4}} = \frac{25}{2} - 1.96 \sqrt{\frac{25}{4}} = 7.6$$

于是, 令  $C_{\alpha(2), n} = 7$ , 所得置信区间与例 5.18 构造的置信区间相同。如果  $n$  大于 30, 则  $C_{\alpha(2), n}$  的近似值和精确值通常都是相等且是整数。

在例 5.18 中, 城市想知道每户每周可回收物的中位数是否大于 5 磅。我们构造了中位数的置信区间, 但仍不能回答中位数是否大于 5。因此, 我们需要研究中

位数的假设检验问题。

我们将利用中位数置信区间的构造方法来研究关于总体中位数的假设检验问题。事实上,总体中位数  $M$  的  $100(1-\alpha)\%$  置信区间可以被用来检验关于  $M$  的双侧假设。如果我们要在水平  $\alpha$  下检验假设  $H_0: M = M_0$   $H_1: M \neq M_0$ , 那么我们就构造  $M$  的  $100(1-\alpha)\%$  置信区间。若  $M_0$  被包含于置信区间之中, 那么我们就不能拒绝  $H_0$ 。若  $M_0$  落在置信区间之外, 那么我们就拒绝  $H_0$ 。

对于  $M$  的单侧假设检验问题, 我们将利用二项分布来确定拒绝域。所用的检验方法被称之为**符号检验**, 具体构造如下。令  $y_1, y_2, \dots, y_n$  为取自于中位数等于  $M$  的总体的随机样本。 $M$  的零值为  $M_0$ , 定义  $W_i = y_i - M_0$ 。符号检验的检验统计量就是正的  $W_i$  的个数。因为  $M$  是总体中位数, 所以, 有约 50% 的数据大于  $M$ , 50% 的数据小于  $M$ 。现在, 如果  $M = M_0$ , 那么  $y_i$  就有 50% 的机会大于  $M_0$ , 因此  $W_i$  有 50% 的机会为正。因为这些  $W_i$  相互独立, 且当  $M = M_0$  时, 每一个  $W_i$  有 50% 的机会为正,  $B$  表示在假设  $H_0$  下正的  $W_i$  的个数, 那么  $B$  是服从  $\pi = 0.5$  的二项分布随机变量, 于是, 附表 4 中  $\pi = 0.5$  的二项分布分位数可以被用来构造假设检验的拒绝域。下面将**总体中位数  $M$  的统计检验**进行一下总结, 其中给出了三种不同的假设以及相应的拒绝域。该检验适用于任何总体分布。

#### 总体中位数 $M$ 的统计检验小结

假设:

情形 1:  $H_0: M \leq M_0$   $H_1: M > M_0$  (右侧检验)

情形 2:  $H_0: M \geq M_0$   $H_1: M < M_0$  (左侧检验)

情形 3:  $H_0: M = M_0$   $H_1: M \neq M_0$  (双侧检验)

T.S.: 令  $W_i = y_i - M_0$ ,  $B =$  正的  $W_i$  个数。

R.R.: 给定第一类错误概率  $\alpha$ ,

情形 1: 若  $B \geq n - C_{\alpha(1), n}$ , 则拒绝  $H_0$

情形 2: 若  $B \leq C_{\alpha(1), n}$ , 则拒绝  $H_0$

情形 3: 若  $B \leq C_{\alpha(2), n}$ , 或  $B \geq n - C_{\alpha(2), n}$ , 则拒绝  $H_0$

下面将举例说明总体中位数的假设检验问题。

#### 例 5.20

参看例 5.18。环卫部门想知道每户每周可回收垃圾的中位数是否大于 5 磅。利用例 5.18 的数据, 在水平下  $\alpha = 0.05$  检验研究假设。

**解答** 假设为

$$H_0: M \leq 5 \quad H_a: M > 5$$

数据集为  $n=25$  户的一个随机样本。查附表 4, 得  $C_{\alpha(2), n} = C_{0.05, 25} = 7$ 。因此, 如果  $B \geq n - C_{\alpha(1), n} = 25 - 7 = 18$ , 我们将拒绝  $H_0: M \leq 5$ 。令  $W_i = y_i - M_0 = y_i - 5$ , 即得:

-3.9	-3.8	-2.9	-2.4	-2.3	-2.1	-1.4	-1.1	-0.8
-0.7	-0.5	-0.3	0.3	0.6	0.8	1.5	1.7	1.7
2.8	2.8	9.2	20.9	24.5	29.8	38.8		

所有 25 个  $W_i$  中有 13 个为正。因此  $B=13$  并不大于 18。于是, 我们得出结论: 数据尚不能证明每户每周可回收垃圾的中位数大于 5 磅。

### 大样本近似

当样本容量  $n$  大于附表 4 所列的值时, 我们可以用正态分布近似二项分布以确定拒绝域。符号检验的标准化形式为

$$B_{ST} = \frac{B - (n/2)}{\sqrt{n/4}}$$

当  $M$  等于  $M_0$  时,  $B_{ST}$  近似于服从标准正态分布。因此, 对于三种不同的研究假设, 我们有下面的判决规则。

情形 1: 若  $B_{ST} \geq z_{\alpha}$ , 则拒绝  $H_0: M \leq M_0$ , 且  $p$  值 =  $Pr(Z \geq B_{ST})$

情形 2: 若  $B_{ST} \leq -z_{\alpha}$ , 则拒绝  $H_0: M \geq M_0$ , 且  $p$  值 =  $Pr(Z \leq B_{ST})$

情形 3: 若  $|B_{ST}| \geq z_{\alpha/2}$ , 则拒绝  $H_0: M = M_0$ , 且  $p$  值 =  $2Pr(Z \geq |B_{ST}|)$

其中  $z_{\alpha}$  是标准正态分布的分位数。

### 例 5.21

利用例 5.20 中的信息, 构造符号检验的大样本近似, 并将你的结论与精确符号检验的结论相比较。

**解答** 参看例 5.20,  $n=25$ ,  $B=13$ 。我们构造符号检验的大样本近似如下。如果  $B_{ST} \geq z_{0.05} = 1.96$ , 那么我们拒绝  $H_0: M \leq 5$ , 支持  $H_a: M > 5$ 。

$$B_{ST} = \frac{B - (n/2)}{\sqrt{n/4}} = \frac{13 - (25/2)}{\sqrt{25/4}} = 0.2$$

因为  $B_{ST}$  不大于 1.96, 所以我们不能拒绝  $H_0$ 。查附表 1, 得

$$p \text{ 值} = Pr(Z \geq 0.2) = 1 - Pr(Z < 0.2) = 1 - 0.5793 = 0.4207$$

因此, 我们得出和精确检验同样的结论。

在 5.7 节, 我们知道: 当总体分布严重厚尾或严重偏倚时,  $t$  检验的效果变差。表 5.7 列出了符号检验的水平 and 功效, 并针对 5.7 节图 5.21 所描绘的四种分布, 将这些值和  $t$  检验的相应值进行对比。理想地, 对于所有总体分布, 检验水平应该

一样。同时,我们希望检验功效达到最大,因为检验功效表示发现错误零假设的能力。当总体分布严重厚尾或严重偏倚时, $t$  检验的水平发生了变化,而不是给定的值 0.05。在这些情况下,符号检验的水平保持不变,因为符号检验的水平对于所有的分布都是一样的。当样本来自于正态总体时, $t$  检验的功效大于符号检验的功效。然而,当样本来自于严重厚尾分布或严重偏倚分布的总体时,符号检验的功效大于  $t$  检验的功效。

表 5.7  $t$  检验对符号检验的水平和功效

总体分布	检验 统计量	$n = 10$				$n = 15$				$n = 20$			
		$(M_a - M_0)/\sigma$				$(M_a - M_0)/\sigma$				$(M_a - M_0)/\sigma$			
		水平	0.2	0.6	0.8	水平	0.2	0.6	0.8	水平	0.2	0.6	0.8
正态	$t$	0.05	0.145	0.543	0.754	0.05	0.182	0.714	0.903	0.05	0.217	0.827	0.964
	符号检验	0.055	0.136	0.454	0.642	0.059	0.172	0.604	0.804	0.058	0.194	0.704	0.889
厚尾	$t$	0.035	0.104	0.371	0.510	0.049	0.115	0.456	0.648	0.045	0.163	0.554	0.736
	符号检验	0.055	0.209	0.715	0.869	0.059	0.278	0.866	0.964	0.058	0.325	0.935	0.990
轻度偏倚	$t$	0.055	0.140	0.454	0.631	0.059	0.178	0.604	0.794	0.058	0.201	0.704	0.881
	符号检验	0.025	0.079	0.437	0.672	0.037	0.129	0.614	0.864	0.041	0.159	0.762	0.935
严重偏倚	$t$	0.007	0.055	0.277	0.463	0.006	0.078	0.515	0.733	0.011	0.104	0.658	0.873
	符号检验	0.055	0.196	0.613	0.778	0.059	0.258	0.777	0.912	0.058	0.301	0.867	0.964

### 练习

#### 基本技能

**5.63** 假定我们从中位数为  $M$  的总体中抽取容量为  $n$  的随机样本。我们要构造  $M$  的 90% 置信区间。

- 若  $n = 20$ , 查附表 4 求  $L_{\alpha/2}$  和  $U_{\alpha/2}$ 。
- 用大样本近似求  $L_{\alpha/2}$  和  $U_{\alpha/2}$ , 并将它们和 (a) 中的值进行比较。

**5.64** 假定我们从中位数为  $M$  的总体中抽取容量为  $n$  的随机样本。我们要构造  $M$  的 90% 置信区间。

- 若  $n = 20$ , 查附表 4, 求  $L_{\alpha/2}$  和  $U_{\alpha/2}$ 。
- 用大样本近似, 求  $L_{\alpha/2}$  和  $U_{\alpha/2}$ , 并将它们和 (a) 中的值进行比较。

**5.65** 假定我们从中位数为  $M$  的总体中抽取容量为 30 的随机样本。我们要在  $\alpha = 0.05$  水平下检验  $H_0: M \leq M_0$   $H_a: M > M_0$ 。利用附表 4 中的值, 构造拒绝域以检验这些假设。

**5.66** 参看练习 5.65。用大样本近似构造拒绝域,并将你的结果与练习 5.65 所得的拒绝域进行比较。

**5.67** 假定我们从中位数为  $M$  的总体中抽取容量为 50 的随机样本。我们要在  $\alpha = 0.05$  水平下检验  $H_0: M = M_0$   $H_a: M \neq M_0$ 。利用附表 4 中的值,构造拒绝域以检验这些假设。

**5.68** 参看练习 5.67。用大样本近似构造拒绝域,并将你的结果与练习 5.67 所得的拒绝域进行比较。

## 应用

**5.69(商业)** 对于工人来说,在保健上究竟要花多少钱是一个重要的问题,因为许多公司提供的健康保险只覆盖很多医疗项目的一部分。一中型公司的职工福利主管想了解公司一般小时工用于保健的费用。随机抽取 35 名工人,过去一年他们家庭的保健费用如下。

400 345 248 1,290 398 218 197 342 208 223 531 172 4,321  
143 254 201 3,142 219 276 326 207 225 123 211 108

- 用盒形图和正态概率图考察总体是否服从正态分布。
- 根据你在(a)中的答案,每户花费的均值还是中位数更适合度量一般工人用于保健的费用。
- 构造工人保健费用的 95% 置信区间。解释置信区间告诉了我们关于保健的费用的什么信息。
- 给定  $\alpha = 0.05$ ,工人每年的保健费用的典型值超过 400 美元吗?

**5.70(政府)** 许多州企图减少血液中酒精含量的法定水平,只要不超过这一水平,司机饮酒就被认为是合法的。法律上作这样的修改遇到来自于某些商业组织的阻力,他们宣称目前的限制水平就足够了。为此,需要作一项研究以证明血液中 0.1% 酒精含量对反应时间的影响,而 0.1% 酒精含量是目前许多州规定的法定水平。随机抽取 25 名法定驾驶年龄的司机,让他们饮足够量的酒直至血液中酒精含量达到 0.1%。通过一项标准的实验室测试方法分别测量他们饮酒前后的反应时间。饮酒后与饮酒前反应时间(以秒计)之差如下:

0.01 0.02 0.04 0.05 0.07 0.09 0.11 0.26 0.27 0.27 0.28 0.28 0.29  
0.29 0.30 0.31 0.31 0.32 0.33 0.35 0.36 0.38 0.39 0.39 0.40

- 对于那些血液中酒精含量为 0.1% 的司机,构造其反应时间之差均值的置

信区间和中位数的置信区间。

- b. 是否有充分的证据表明 0.1% 酒精含量的水平导致了平均反应时间的增加?
- c. 是否有充分的证据表明 0.1% 酒精含量的水平导致了反应时间中位数的增加?
- d. 对反应时间之差的哪一种结论更适合, 均值还是中位数? 给出理由。

**5.71** 参看练习 5.70。商业组织的说客让他们的专家检查实验设备, 并认为反应时间的测量可能有误。除非反应时间之差至少为 0.25, 否则, 专家声称两个时间基本上是相同的。

- a. 是否有充分的证据表明反应时间之差的中位数大于 0.25 秒?
- b. 在企图确定饮适量的酒是否影响反应时间时, 还有其他什么有关司机的因素是重要的?

**5.72(社会)** 美国国会在 1999 年会上辩论社会保险基金是否应该部分投资到股市, 以增加人们退休时从社会保险机构所领得的保障金。共有基金的倡导者为了让公众放心, 他们声称绝大多数共有基金比社会保险提供了更大的退休收益。从 1989 到 1999 年重点推荐的两大共有基金的年回报率数据如下(年回报率被定义为  $(P_1 - P_0)/P_0$ , 其中  $P_0$  和  $P_1$  分别为年初和年末基金的价格):

年份	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
基金 A	25.4	17.1	-8.9	26.7	3.6	-8.5	-1.3	32.9	22.9	26.6
基金 B	31.9	-8.4	41.8	6.2	17.4	-2.1	30.5	15.8	26.8	5.7

- a. 对于基金 A 和基金 B, 试估计年回报率的均值和中位数, 并分别构造相应的 95% 置信区间。
- b. 均值或中位数中, 你认为哪一个更适合表示 1989 到 1999 年基金 A 和基金 B 的年回报率? 给出你的理由。

**5.73** 参看练习 5.72。

- a. 是否有充分的证据表明两种共有基金年回报率的中位数大于 10%?
- b. 是否有充分的证据表明两种共有基金年回报率的均值大于 10%?

**5.74** 对于那些选择退休计划的人来说, 还有哪些年回报率的概括统计量是重要的?

**5.75** 利用表 5.7 的信息, 回答下列问题。

- a. 如果总体服从正态分布, 那么总体均值和总体中位数相等。因此均值和中位数都可以表示总体的中心。在这种情况下, 为什么  $t$  检验比符号检验更



适合于检验总体的分布中心?

- b. 假定总体分布严重右偏。研究人员在  $\alpha = 0.05$  下用  $t$  检验来检验关于总体均值的假设。如果样本容量  $n = 10$ , 那么检验的第一类错误概率是 0.05 吗? 给出你的理由。
- c. 当检验严重偏倚分布总体的均值或中位数时, 符号检验和  $t$  检验的功效之差随着  $(M_a - M_0)$  的增加而减少。用表 5.7 中的值验证这一命题。为什么会出现这种情况?
- d. 当检验轻度偏倚分布总体的均值或中位数时, 符号检验和  $t$  检验的功效之差比检验严重偏倚分布总体时更小。用表 5.7 中的值验证这一命题。为什么会出现这种情况?

## 5.9 小 结

对总体均值和总体中位数能够用点估计和区间估计的方法进行估计。是否选择中位数而不选择均值作为总体中心的代表依赖于总体分布的形状。区间估计的好坏取决于区间的宽度和置信系数。我们给出了均值  $\mu$  和中位数  $M$  的  $100(1 - \alpha)\%$  置信区间的公式。在事先给定  $\mu$  的置信区间宽度和置信水平前提下, 提供了确定样本容量的公式。

根据传统的假设检验方法, 统计检验包括 5 个部分: 零假设, 检验统计量, 拒绝域, 检查假设以及得出结论。统计检验利用反证法的思想。我们做试验和分析, 收集数据, 通过获得与零假设  $H_0$  相矛盾的论据来验证研究假设。与任何双向决策过程一样, 假设检验也可能犯两种类型的错误。第一类错误就是当  $H_0$  为真但拒绝  $H_0$  所犯的误差, 第二类错误就是当备择假设  $H_a$  为真但接受  $H_0$  所犯的误差。犯第一类错误的概率用  $\alpha$  来表示。对于  $H_a$  中均值  $\mu_a$  的某个已知值, 犯第二类错误的概率用  $\beta(\mu_a)$  表示。随着  $\mu_a$  到  $\mu_0$  距离的增加,  $\beta(\mu_a)$  减少。检验功效就是当  $\mu$  值位于  $H_a$  一边时, 检验拒绝  $H_0$  的概率。因此,  $\mu_a$  处的功效等于  $1 - \beta(\mu_a)$ 。

我们也指出, 给定样本容量和均值  $\mu_a$ ,  $\alpha$  和  $\beta(\mu_a)$  成反向关系; 随着  $\alpha$  增加,  $\beta(\mu_a)$  减少, 反之亦然。对于任一检验方法, 给定样本容量  $n$  和检验水平  $\alpha$ , 我们都可以计算  $\beta(\mu_a)$ , 其中均值  $\mu_a$  为备择假设中的值。在很多研究中, 我们需要确定样本容量  $n$  使得检验具有给定的水平  $\alpha$ , 并且  $\beta(\mu_a)$  以某一值为界。为此, 我们给出了  $n$  的计算公式, 使得水平为  $\alpha$  的检验满足  $\beta(\mu_a) \leq \beta$ , 而无论  $\mu_a$  距离  $\mu_0$  有多远。

我们给出了另外一种并非基于传统判决方法的统计假设检验方法。我们不是依赖于事先给定的水平  $\alpha$ , 而是计算包含在数据中拒绝零假设的证据的权重。这种权重用概率来表示, 我们称之为检验的显著性水平。大多数专业期刊都用显著

性水平总结统计检验的结果。我们讨论了如何使用显著性水平以便获得与传统方法同样的结果。

我们也讨论了当  $\sigma$  未知时  $\mu$  的统计推断问题, 而  $\sigma$  未知是常见的情况。通过使用  $t$  分布, 我们能够构造  $\mu$  的置信区间, 进行统计检验。当总体分布严重偏倚或严重厚尾, 且样本容量较小时,  $t$  检验和置信区间往往达不到规定的水平和功效。在这些情况下, 我们可以用中位数代替均值来表示总体的中心。我们也提供了构造总体中位数置信区间和进行假设检验的方法。

### 重要公式

#### $\mu$ 和中位数的估计和检验

1. 当样本来自于正态总体或样本容量较大时,  $\mu(\sigma$  已知) 的  $100(1-\alpha)\%$  置信区间

$$\bar{y} \pm z_{\alpha/2} \sigma_{\bar{y}} \text{ 其中 } \sigma_{\bar{y}} = \sigma / \sqrt{n}$$

2. 当样本来自于正态总体或样本容量较大时,  $\mu(\sigma$  未知) 的  $100(1-\alpha)\%$  置信区间

$$\bar{y} \pm t_{\alpha/2} s / \sqrt{n}, df = n - 1$$

3.  $\mu$  的  $100(1-\alpha)\%$  置信区间估计为  $\bar{y} \pm E$  所需的样本容量

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$$

4. 当样本来自于正态总体或样本容量较大时,  $\mu(\sigma$  已知) 的统计检验

$$\text{统计量: } z = \frac{\bar{y} - \mu_0}{\sigma / \sqrt{n}}$$

5. 当样本来自于正态总体或样本容量较大时,  $\mu(\sigma$  未知) 的统计检验

$$\text{统计量: } t = \frac{\bar{y} - \mu_0}{s / \sqrt{n}}, df = n - 1$$

6. 当样本来自于正态总体或样本容量较大时, 针对  $\mu(\sigma$  已知) 的统计检验,  $\beta(\mu_a)$  (和相应的功效) 的计算

a. 水平为  $\alpha$  的单侧检验

$$\beta(\mu_a) = P\left(z < z_{\alpha} - \frac{|\mu_0 - \mu_a|}{\sigma_{\bar{y}}}\right), \text{ 其中 } \sigma_{\bar{y}} = \sigma / \sqrt{n}$$

b. 水平为  $\alpha$  的双侧检验

$$\beta(\mu_a) \approx P\left(z < z_{\alpha/2} - \frac{|\mu_0 - \mu_a|}{\sigma_{\bar{y}}}\right), \text{ 其中 } \sigma_{\bar{y}} = \sigma / \sqrt{n}$$

7. 当样本来自于正态总体或样本容量较大时, 针对  $\mu(\sigma$  未知) 的统计检验, 计

算 $\beta(\mu_a)$ (和相应的功效):利用附录表 3。

8. 当样本来自于正态总体, $\mu$ ( $\sigma$  已知)的统计检验所需的样本容量。

a. 水平为  $\alpha$  的单侧检验

$$n = \frac{\sigma^2}{\Delta^2} (z_\alpha + z_\beta)^2$$

b. 水平为  $\alpha$  的双侧检验

$$n \approx \frac{\sigma^2}{\Delta^2} (z_{\alpha/2} + z_\beta)^2$$

9. 总体中位数  $M$  的  $100(1-\alpha)\%$  置信区间

$$(y_{(L_{\alpha/2})}, y_{(U_{\alpha/2})}), \text{ 其中 } L_{\alpha/2} = C_{\alpha(2), n}, U_{\alpha/2} = n - C_{\alpha(2), n} + 1$$

10. 关于中位数的统计检验

检验统计量:

$$\text{令 } W_i = y_i - M_0, B = \text{正的 } W_i \text{ 的个数}$$

补充练习

**5.76(商业)** 一生产油漆的厂商在广告中声称:1 加仑油漆平均可覆盖 400 平方英尺以上的面积。为了使其广告语更有效,该厂商聘请了一家独立检测实验室对其进行评价,为此,从该厂商的仓库中随机抽取 50 罐 1 加仑油漆。

- 感兴趣的参数是什么?
- 如果厂商希望第一类错误的概率最多为 0.05,那么研究假设和相应的拒绝域是什么?
- 如果根据 50 罐油漆的样本得到平均覆盖面积为 412 平方英尺,标准差为 38 平方英尺,那么数据支持厂商的广告语吗?
- 构造 1 加仑该厂商油漆的平均可覆盖面积的 95% 置信区间。
- 求检验的  $p$  值。如果  $\alpha$  减小到 0.01,你对厂商广告所下的结论会改变吗?

**5.77(工程)** 某大城市交通部门要重新改造其中的一个停车库,增加每小时的停车数量。根据城市的记录,过去 5 年每辆车的平均停车时间是 220 分钟。交通部门想知道重新改造以及每小时停车数的增加是否会改变平均停车时间。改造之后的 3 个月内,随机抽取 100 辆汽车,得到平均停车时间为 208 分钟,标准差为 55 分钟。

- 什么是研究假设?
- 如果  $\alpha = 0.05$ ,那么数据支持研究假设吗?
- 检验的显著性水平( $p$  值)等于多少?
- 构造改造之后平均停车时间的 95% 置信区间。

**5.78(人力资源)** 某办公室主任实施了一项激励计划,他认为该激励计划能

够减少处理顾客投诉所需的平均时间。在执行这一计划前,处理投诉平均所需时间是 30 分钟。计划实施几个月之后,随机抽取 38 份处理顾客投诉的记录,发现平均所需时间为 28.7 分钟,标准差为 3.8 分钟。

- a. 给出处理顾客投诉平均所需时间的点估计。
- b. (a) 中的点估计量的标准差是多少?
- c. 构造计划实施之后处理顾客投诉平均所需时间的 95% 置信区间,并替办公室主任解释所得的置信区间。
- d. 是否有充分证据显示激励计划减少了处理顾客投诉的平均时间?

**5.79(工程)** 人们监测湖中汞的浓度已有很多年了。根据每月的记录得到平均浓度为  $1.20\text{mg}/\text{m}^3$  (每立方米毫克),标准差为  $0.32\text{mg}/\text{m}^3$ 。湖岸边的一个熔炉发生了一次事故之后,紧接着抽取了 15 个水样,得汞的浓度数据如下。

1.60	1.77	1.61	1.08	1.07	1.79	1.34	1.07
1.45	1.59	1.43	2.07	1.16	0.85	2.11	

- a. 给出事故之后平均汞的浓度的点估计。
- b. 构造事故之后平均汞的浓度的 95% 置信区间,并解释之。
- c. 给定  $\alpha = 0.05$ , 是否有充分证据表明自从事故之后平均汞的浓度增加了?
- d. 假定汞的浓度的标准差为  $0.32\text{mg}/\text{m}^3$ , 对于汞的浓度 1.28, 1.32, 1.36 和 1.40, 分别计算相应的检验功效。

**5.80** 对于下面每个问题,回答“是”或“错”。

- a. 给定任何一个特定的样本,如果我们构造了样本均值的 95% 置信区间,那么总体均值有 95% 的机会落在置信区间之内。
- b. 如果选取很多组随机样本,且分别构造 95% 置信区间,那么,这些区间中将有 95% 包含总体均值。
- c. 如果样本容量大于 30,那么,样本均值将有 95% 的机会等于总体均值。
- d. 如果选取很多组随机样本,那么,有 95% 的可能性其中一个样本均值等于总体均值。
- e. 根据同一样本均值所得的 95% 置信区间比 90% 置信区间更宽。
- f. 为了证明  $\mu = \mu_0$  且犯第一类错误概率为 0.05,那么,我们必须选取一个样本,且在  $\alpha = 0.05$  下未能拒绝零假设  $H_0: \mu = \mu_0$ 。
- g. 为了求第一类错误概率为 0.04 的双侧假设的临界值,我们可以查附表 1 中对应于面积 0.4800 的  $z$  值。
- h. 为了求第一类错误概率为 0.02 的单侧假设的临界值,我们可以查附表 1 中对应于面积 0.4800 的  $z$  值。

i. 如果我们在  $\alpha = 0.05$  下拒绝零假设, 那么, 在  $\alpha = 0.01$  下我们也应该拒绝零假设。

**5.81** 对于下面每个问题, 回答“是”或“错”。如果回答“错”, 请修改命题使之成立。仅修改画线部分。

- 当零假设  $H_0$  实际上不成立但我们未能拒绝  $H_0$  时所犯的错误是第一类错误。
- 如果我们犯了第二类错误, 那么, 我们一定未能发现一个事件或影响, 而它确实存在。
- 犯第一类错误的概率等于  $\beta$ 。
- 如果我们增加第二类错误的概率, 那么, 犯第一类错误的概率也将增加。

**5.82(医学)** 过去几年, 都市的某大医院对孕妇预产期进行预测, 效果相当差。医生参加了一项在职培训计划以提高技术, 改进他们的预测效果。在最近一次调查中, 随机选取 100 名母亲, 她们都是在培训计划之后于这家医院分娩的。由样本数据得, 超过预测的预产期的平均天数为 9.2 天, 标准差为 12.4 天。

- 请叙述如何随机选取 100 名母亲。
- 用 95% 置信区间估计超过预产期的平均天数, 并解释之。
- 如果在培训之前, 超过预产期的平均天数是 13 天, 那么, 是否有充分的证据表明平均天数减少了? 检验的显著性水平等于多少?
- 医生预测的预产期并不接近于真正的分娩期, 有哪些重要的因素可解释这一现象?

**5.83(商业)** 某公司为其销售人员供养了大批车辆。为了调查全体销售人员每月驾车行驶的里程数, 随机抽取 70 个记录。平均行使里程数及标准差分别为 3,250 和 420。利用 99% 置信区间估计公司内全体销售人员平均每月驾车行驶的里程数  $\mu$ 。

**5.84(工程)** 机器挤压出的塑料管的直径围绕着均值而变化, 而均值是由机器的某一装置所控制。从该台机器一天的产量中随机抽取 50 根管子。测量其直径得平均直径为 4.05 英寸, 标准差为 0.12 英寸。

- 请叙述如何随机抽取 50 根管子。
- 有哪些有关塑料管生产的重要因素可解释为什么塑料管的直径产生变化?
- 用 99% 置信区间估计塑料管的平均直径。
- 数据是否提供了充分的证据显示平均直径不等于名义直径 4 英寸? 已知  $\alpha = 0.01$ 。

**5.85(工程)** 某自动控制装置的厂商声称该装置将能使房间的平均湿度保持在 80%。记录了 30 天装有该控制装置的房间的湿度。均值和标准差分别为 78.3% 和 2.9%。数据是否提供了充分的证据足以反对厂商的声言? 已知

$\alpha = 0.05$ 。

**5.86(农业)** 一采购员希望了解从某一柑橘林运来的柑橘中平均每个柑橘所含的糖分是否小于 0.027 磅。随机选取 50 个柑橘, 得平均含糖量为 0.025 磅, 标准差为 0.003 磅。数据是否提供了充分的证据显示平均含糖量小于 0.027 磅? 已知  $\alpha = 0.05$ 。

**5.87(教育)** 某州立大学的行政官员担心住在校外的一年级学生的平均成绩 (GPA) 明显低于整个年级的 GPA。在秋季学期之后, 所有的一年级学生平均 GPA 为 2.3 (以 4 分制)。因为不可能从大学的成绩单中分离出所有住在校外的学生, 于是通过他们的永久家庭地址追踪学生, 从而得到 200 名住在校外新生的随机样本。样本平均 GPA 为 2.24, 标准差为 0.31。

- 数据是否提供了充分的证据显示所有住在校外的新生的平均 GPA 低于全体新生的平均 GPA 2.3, 已知  $\alpha = 0.05$ 。
- 住在校外新生的平均 GPA 与全体新生的平均 GPA 之差在实际上是显著的, 还是仅仅在统计上是显著的? 试解释之。

**5.88(医学)** 厂商需要对某特殊药品的片剂进行标准的溶解检验, 以检测出该批药片的溶解率, 然后才能投放市场。假定随机抽取 24 片重 25mg 的药片, 溶解检验就是对这些药片进行分析和鉴定。每次检验时, 将药片悬浮在酸液中, 30 分钟后进行分析和鉴定。24 个鉴定结果如下。

19.5	19.7	19.7	20.4	19.2	19.5	19.6	20.8
19.9	19.2	20.1	19.8	20.4	19.8	19.6	19.5
19.3	19.7	19.5	20.6	20.4	19.9	20.0	19.8

- 用图示法检验样本数据是否来自于正态总体。
- 用点估计和 99% 置信区间估计, 估计该批药片的平均溶解率。
- 给定  $\alpha = 0.01$ , 是否有充分证据表明该批药片的平均溶解率小于 20mg (即药片上标注量的 80%)?
- 如果真正的溶解率为 19.6mg, 计算犯第二类错误的概率。

**5.89(商业)** 对于审计师来说, 统计是一个有用的工具, 特别是在审计大量的清单时。如果一个审计师要对大量清单进行逐项审计, 那将是既费钱又费时的事。因此, 审计师通常采用抽样的方式从清单中随机抽取一组样本, 用样本的结果来检验公司财务报告的真实性的。例如, 某医院财务报告声称其一份财产清单中平均每份清单 300 美元。审计师从中随机抽取 20 份得均值和标准差分别为 160 美元和 90 美元。给定  $\alpha = 0.05$ , 样本数据与医院财务报告所声称的相矛盾吗? 数据是否表明均值小于 300 美元?

**5.90(商业)** 过去 5 年,某仓库交付定货所需的时间平均为 25 分钟。公司职员认为最近交付定货的时间有所增加要么是由于劳动力的变化,要么是由于顾客购买策略发生了变化。随机抽取 15 份上个月处理的定单,记录其交付定货所需的时间如下。

28	25	27	31	10
26	30	15	55	12
24	32	28	42	38

数据是否提供了充分的证据表明交付定货的时间有所增加?

**5.91(工程)** 如果一项新的铜矿采掘方法被全日制地付诸实施,那么平均每天生产铜矿 50 吨以上。下表列出了 15 天试验期间的铜矿产量。

日期	1	2	3	4	5	6	7	8
产量(吨)	57.8	58.3	50.3	38.5	47.9	157.0	38.6	140.2
日期	9	10	11	12	13	14	15	
产量(吨)	39.3	138.7	49.2	139.7	48.3	59.2	49.7	

- 用点估计和 95% 置信区间估计,估计实施新的采掘法后平均每天铜矿的产量。
- 是否有充分证据表明实施新的采掘法后平均每天铜矿的产量超过 50 吨? 已知  $\alpha = 0.05$ 。

**5.92(教育)** 对学生进行一次测验以检查他们阅读某规定量的材料所需的时间。学生被告知在理解文章的基础上以最快的速度阅读,因为测验试题与他们阅读的材料有关。从大学校区随机抽取 16 名三年级学生,他们的测验结果如下(以分钟计):

25	18	27	29	20	19	25	24
32	21	24	19	23	28	31	22

- 用 95% 置信区间估计所有三年级学生阅读所需的平均时间。
- (a) 中的置信区间所适用的总体是什么?
- 解释(a)中的置信区间。

**5.93(医学)** 某药厂用大的发酵缸生产抗生素。为了检验准备好的一批抗生

素的平均效力,从发酵缸的 12 个不同位置抽样。记录抗生素的效力如下:

8.9	9.0	9.1	8.3	9.9	9.0
8.4	9.2	9.0	8.7	9.3	9.1

- 估计该批抗生素平均效力的 95% 置信区间,并解释之。
- 你如何从发酵缸中选取 12 个样品?
- 如果抗生素效力被认为是 9.0,那么,是否有充分证据显示抗生素的平均效力与所认为的不一致?

**5.94** 在关于  $\mu$  的统计检验中,拒绝了零假设。基于这一结论,下面哪些命题是正确的?

- 犯了第一类错误。
- 犯了第二类错误。
- 可能犯了第一类错误。
- 可能犯了第二类错误。
- 不可能同时犯第一类错误和第二类错误。
- 不可能既不犯第一类错误,也不犯第二类错误。
- 不知道是否犯了错误,但如果犯了错误,那一定是第一类错误。
- 不知道是否犯了错误,但如果犯了错误,那一定是第二类错误。

**5.95** 对下面每一个命题,回答“对”或“错”。

- 在水平  $\alpha = 0.05$  的假设检验中,增加样本容量并不影响检验水平。
- 在水平  $\alpha = 0.05$  的假设检验中,增加样本容量并不影响检验功效。
- 样本容量  $n$  在检验假设过程中起着重要的作用,因为,它是衡量数据(从而也是衡量信息)量大小的尺度,而数据是我们作判决的基础。如果数据变异性很大,且  $n$  值较小,那么我们将不太可能拒绝零假设,甚至当零假设不成立时,我们也不太可能拒绝。
- 假定我们检验关于总体均值  $\mu$  的如下假设,  $H_0: \mu \leq \mu_0$   $H_a: \mu > \mu_0$ 。如果样本容量  $n$  很大,且数据的变异性也不太大,那么甚至当  $\mu$  的真值比  $\mu_0$  只大一点点,我们也很可能拒绝零假设。
- 当对分布严重右偏的总体进行统计推断时,和  $\alpha = 0.05$  的符号检验相比,  $\alpha = 0.05$  的  $t$  检验不太可能犯第二类错误。
- 当对分布严重右偏的总体进行统计推断时,和  $\alpha = 0.05$  的符号检验相比,  $\alpha = 0.05$  的  $t$  检验不太可能犯第一类错误。

**5.96** 完成下面的命题。

- 如果我们从总体中抽取所有可能的样本(样本容量固定),那么,样本均值



的分布趋向于\_\_\_\_\_且这些样本均值的均值等于\_\_\_\_\_。

- b. 如果其他保持不变, 样本容量越大, 那么, 置信区间\_\_\_\_\_。
- c. 如果其他保持不变, 置信系数越大, 那么, 置信区间\_\_\_\_\_。
- d. 命题“如果从任一总体(不管总体分布的形状)抽取容量相同的随机样本, 那么, 随着  $n$  变大, 样本均值的分布接近于正态分布”被认为是\_\_\_\_\_。
- e. 当零假设不成立, 但没有拒绝零假设时, 我们犯了第\_\_\_\_\_类错误。

**5.97(医学)** 假定香烟中焦油的含量服从均值为 10mg, 方差为 2.4mg 的正态分布。现开发了一种新的香烟制造技术以减少焦油的含量。随机抽取 16 根利用新技术生产的香烟, 得平均焦油含量为 8.8mg。给定  $\alpha = 0.05$ 。

- a. 试利用以下要点, 制定一假设检验以检验新的技术是否明显地减少了焦油含量。

零假设

备择假设

假定

拒绝域

检验统计量和计算

用统计语言给出结论

用简单直观的语言给出结论

- b. 基于你的结论, 你是可能犯第一类错误? 还是可能会犯了第二类错误? 还是两类错误都没犯? 还是同时可能犯两类错误?

**5.98(环境)** 某州卫生机构被要求调查居民的投诉, 投诉声称厂家将污染物未经处理就被排放到流经一居民区的小河里。通过在财政上施加压力, 该州能够迫使违反规定的厂家作出较多的让步, 安装新的水净化装置。在过渡时期, 可引进另外的生产装置以降低排放到河流的水污染水平。为了监测过渡装置的效果, 在过渡期的月份, 从生产厂家下游的某个地点随机抽取 15 个水样。如果  $\bar{y} = 5.0$ ,  $s = 0.70$ , 用样本数据检验水中被溶解氧气的平均含量(以 ppm 计)是否小于 5.2, 而 5.2 是上一年该地点的平均值。

- a. 给定  $\alpha = 0.05$ , 列出统计检验的 5 个步骤。
- b. 构造统计检验, 并给出结论。

**5.99(工程)** 一自动装置被安装在进入某一高速公路的坡道处。安装之前, 调查人员发现司机的平均紧张水平为 8.2(10 分制)。安装之后, 随机抽取 50 名司机, 得  $\bar{y} = 7.6$ ,  $s = 1.8$ 。试构造一统计检验, 其研究假设为: 在新装置下, 车流高峰期司机的平均紧张水平小于 8.2, 其中 8.2 是并轨装置安装之前司机的平均紧张水平。求统计检验的显著性水平, 并解释之。

**5.100(环境)** 寻找石油替代品作为主要燃料和能源的主要来源将不可避免

地给环境带来很多挑战。这些挑战包括需要解决诸如露天开采等所造成的环境问题。我们不妨集中讨论一个问题。如果煤被认为是燃料和能源的主要来源,那么,我们必须考虑如何避免大量的二氧化硫( $\text{SO}_2$ )和颗粒物进入空气。这一点对政府部门和工业生产特别重要。下面是一些可能的解决办法。

1. 建造特别高的烟囱。

2. 燃烧之前从煤里分离出  $\text{SO}_2$  和颗粒物。

3. 在煤燃烧所产生的煤气被排放之前,从中分离出  $\text{SO}_2$ 。这可通过煤气洗涤器来达到目的。

最近一种新型煤气洗涤器被制造出来,并被安装在电厂进行测试。在为期 15 天的测试期间,每天从烟囱冒出的气体中抽样 3 次。 $\text{SO}_2$  排放量(单位:磅/百万 BTU)如下:

时间	日期							
	1	2	3	4	5	6	7	8
上午 6 点	0.158	0.129	0.176	0.082	0.099	0.151	0.084	0.155
下午 2 点	0.066	0.135	0.096	0.174	0.179	0.149	0.164	0.122
下午 10 点	0.128	0.172	0.106	0.165	0.163	0.200	0.228	0.129

时间	日期						
	9	10	11	12	13	14	15
上午 6 点	0.163	0.077	0.116	0.132	0.087	0.134	0.179
下午 2 点	0.063	0.111	0.059	0.118	0.134	0.066	0.104
下午 10 点	0.101	0.068	0.100	0.119	0.125	0.182	0.138

- 用 95% 置信区间分别估计在三次抽样的每一时间段  $\text{SO}_2$  的平均排放量。
- 三个时间段  $\text{SO}_2$  平均排放量是否有显著差异?
- 将数据混合得到全天的  $\text{SO}_2$  排放量数据,那么使用新型煤气洗涤器之后,  $\text{SO}_2$  平均排放量是否小于 0.145,其中 0.145 是使用旧煤气洗涤器时平均  $\text{SO}_2$  排放量?

**5.101(社会)** 作为对训练方法进行总体评价的一部分,我们通过做试验来测定健康男性新兵所能承受的平均训练量。为此,在这些士兵中随机抽取 35 名,让他们在一定的负荷下用自行车测功计(一种测量肌肉做功的设备)进行训练直到疲劳为止。仔细监测他们的血压,脉率以及其他指标以确保没有人的健康处于危险之中。下面列出了 35 名士兵的训练能力(平均时间,以分钟计)数据。

23	19	36	12	41	43	19
28	14	44	15	46	36	25
35	25	29	17	51	33	47
42	45	23	29	18	14	48
21	49	27	39	44	18	13

- a. 用这些数据构造  $\mu$  的 95% 置信区间, 并解释之, 其中  $\mu$  为应召入伍的健康男性士兵的平均训练能力。
- b. 如果置信系数为 99%, 那么, 置信区间如何变化?

**5.102** 用练习 5.101 的数据, 求所需的样本观测次数, 使得  $\mu$  的置信系数为 95% 的置信区间长度在 1 分钟内(提示: 在你的结论中用  $s = 12.36$  代替  $\sigma$ )。

**5.103(农业)** 一项研究旨在调查花叶病毒制剂对烟叶的影响。由  $n = 32$  片烟叶的样本知, 平均每片烟叶上的病变有 22 处, 标准差为 3。用这些数据和 95% 置信区间估计由于受制剂影响平均每片烟叶上的病变数。

**5.104** 参看练习 5.103。用样本数据构造  $\mu$  的 99% 置信区间, 其中  $\mu$  为由于受花叶病毒制剂影响平均每片烟叶上的病变数。

**5.105(医学)** 我们都记得曾被告知, “你的烧已经退了, 你的体温已正常”。我们说“正常”意味着什么? 大多数人都以  $98.6^{\circ}\text{F}$  为标准体温, 但这并不适用于所有的人, 仅适用于“平均”人。不必硬要某人说出自己的体温, 我们可以定义一个人的正常体温就是他或她健康时的平均体温, 但即使这样的定义也是模糊的, 因为一天当中一个人的体温是变化的。为了测出一个人的正常体温, 随机选取 30 天, 每天早晨 7 点钟测量其体温。30 天体温的样本均值和样本标准差分别为 98.4 和 0.15。假定该人在检查期间身体是健康的, 利用这些数据估计该人每天早晨 7 点钟时平均体温的 90% 置信区间。

**5.106** 参看练习 5.101 中的数据。假定从一大群参加过新的(即可望有所改进的)身体适应性训练项目的士兵中随机选取 35 名士兵, 并假定过去几年对数千名士兵进行的测试表明平均训练能力为 29 分钟。对研究假设进行统计检验, 其中研究假设为: 新的适应性训练项目提高了平均训练能力。给出检验的显著性水平, 并解释之。

**5.107** 参看练习 5.106。

- a. 如果我们感兴趣的是新的适应性训练比以前的适应性训练更好或更坏, 那么, 研究假设将如何变化?
- b. 检验的显著性水平等于多少?

**5.108** 从具有 100 个以上半专用病床的医院中随机选取 40 家医院。这些

医院的病人中有一些病人,他们的费用由投保于一家大型医疗保险公司的团体保险支付。研究人员收集了每家医院这部分人所占的比例。样本数据如下。

0.67	0.74	0.68	0.63	0.91	0.81	0.79	0.73
0.82	0.93	0.92	0.59	0.90	0.75	0.76	0.88
0.85	0.90	0.77	0.51	0.67	0.67	0.92	0.72
0.69	0.73	0.71	0.76	0.84	0.74	0.54	0.79
0.71	0.75	0.70	0.82	0.93	0.83	0.58	0.84

利用样本数据构造  $\mu$  的 90% 置信区间,并解释之,其中  $\mu$  为平均每个医院由团体保险支付费用的病人所占的比例。

**5.109(人力资源)** 某州立大学系统的教职员工如果在第一个 10 年服务期内辞职,他们就有资格得到已经向退休机构缴纳的钱,且每年增加 4%。不幸的是,经验表明州当局返还这笔钱非常慢。考虑到这一实际情况,一地方教师组织决定进行调查。从过去 5 年辞职的教职员工中随机抽取 50 名进行调查,结果发现这笔钱到期的时间与返还的时间平均相差 75 天,标准差为 15 天。利用这些数据和 95% 置信区间估计从期满到返还的平均时间。

**5.110** 参看练习 5.109。在和教师工会面对面协商之后,州当局答应 60 天内返还。随后观察了 40 例,得到平均 58 天,标准差 10 天。如果我们假定这 40 例是一随机样本,代表了州当局未来的表现,试估计平均返还时间的 99% 置信区间。

**5.111(社会)** 随机选取 40 个内陆城市,调查它们的出生率,结果表明平均每个内陆城市的出生率为千分之 35,标准差为千分之 6.3。试用 95% 置信区间估计内陆城市的平均出生率。

**5.112(社会)** 随机选取 30 个标准的大城市统计区域,记录每个区已登记选民与年龄在 18 岁或 18 岁以上的选民总人数之比。利用所给的数据(每 1,000 人)检验研究假设,即  $\mu$  不等于 675,其中  $\mu$  为今年的平均比例,675 为去年的平均比例。给出检验的显著性水平。

802	497	653	600	729	812
751	730	635	605	760	681
807	747	728	561	696	710
641	848	672	740	818	725
694	854	674	683	695	803

**5.113(商业)** 对于邮购部门来说,定单被不正确填写将付出高昂的代价。为了估计平均每份错误定单所带来的损失,某大公司计划随机抽取  $n$  份被不正确填

写的定单,调查每份定单所带来的额外花费。公司估计到额外花费在 40 美元至 400 美元之间。为了使得平均额外花费的 95% 置信区间的宽度为 20 美元,问:必须抽取多少份被不正确填写的定单?

**5.114** 通过对某医院的病历进行调查以了解平均每位肺癌病人住院的时间长度。100 份病历的样本数据表明  $\bar{y} = 2.1$  月,  $s = 2.6$  月。

- a. 用  $t$  方法估计  $\mu$  的置信区间合适吗? 为什么?
- b. 指出另一种估计总体中心的方法。

**5.115(政府)** 发给失业者或贫民粮票过程中存在欺诈行为,调查人员准备对此进行调查,为此,他们希望用 95% 置信区间估计粮票享有者的年平均总收入,并且误差在 750 美元以内。如果我们假定粮票享有者年收入的极差为 20,000 美元,问:应调查多少名粮票享有者?

**5.116(商业)** 正如前面所指出的,大号洗衣剂的新包装盒上标明的重量是 42 盎司,并且包装箱上写有以下声明:“由于高速包装机的原因,每箱洗衣剂的重量可能略高于标重,也可能略低于标重,这是正常的,但是一天的产量中,平均每箱洗衣剂的重量会略高于标重。”试讨论你如何能检验这一声明。为了检验的目的,对该声明作稍微修改可以使之更简单吗? 叙述该检验的所有步骤。是否可以事先确定样本容量使得检验出备择假设为真的检验功效等于 0.90? 给定  $\alpha = 0.05$ 。

**5.117(医学)** 充血性心力衰竭病人的死亡率很高。对总共 182 例患有慢性左心室衰竭的病人进行跟踪调查,他们尽管接受治疗,但仍有症状。通过调查得知,这些病人的存活时间从 1 个月到 41 个月不等,且平均存活 12 个月,标准差为 10。构造这些病人平均存活月数的置信区间合适吗? 为什么?

**5.118(商业)** 自动取款机(ATM)的销售经过十多年稳步增长,现开始下滑。最近一个月,通过随机抽取 40 家厂商,发现其发货量比 1 年前同期水平平均减少 20%。假定其标准差为 6.2%,数据基本上呈现出丘形分布。利用这些数据构造厂商 ATM 发货量减少的百分数平均值的 99% 置信区间。

**5.119** 假定练习 5.118 中 40 个厂商发货量减少的百分数从 -40% (即减少 40%) 到 +16% (即增加 16%),且样本均值为 -20%,中位数为 -10%,截尾均值为 -12%。讨论用  $t$  方法检验 ATM 发货量减少的百分数的适应性。

**5.120(医学)** 医生建议我们每天摄入咖啡因的量应不超过 200 mg。随机抽取 35 位上班族,要求他们根据下面的表记录自己 7 天期间的咖啡因摄入量。

咖啡(6 盎司)	100~150mg
茶(6 盎司)	40~110mg
可乐类饮料(12 盎司)	30mg
巧克力蛋糕	20~30mg
可可类饮料(6 盎司)	5~20mg

奶油巧克力(1 盎司)                      5~10mg

7 天之后, 我们得到每位上班族平均每天的摄入量。样本均值和样本标准差分别为 560mg 和 160mg。用这些数据估计  $\mu$  的 90% 置信区间, 其中  $\mu$  为上班族平均每位每天的摄入量。

**5.121** 参看练习 5.120。为了用 90% 置信区间估计  $\mu$ , 使之误差在  $\pm 10$ , 那么另外还需要多少个观测?

**5.122(农业)** 俄亥俄州农业部门的调查人员在该地区选取了一个初中, 并抽取了用于学生午餐的半品脱(8 盎司)盒装奶的样本。基于 25 盒奶, 调查人员发现平均每盒奶比标准的半品脱少 0.067 盎司, 标准差为 0.02。

- a. 用这些数据检验假设“平均短缺量是 0”, 备择假设为单边假设。给出检验的  $p$  值。
- b. 虽然 0.067 只是一个很小的量, 如果每年卖出 300 万盒 8 盎司的盒装奶, 并且这些盒装奶都是短缺 0.067 盎司的, 那么一年的销售能省下多少品脱出奶?

**5.123** 参见数据盘中的临床试验数据库, 构造处理组 C 的 HAM-D 总积分的 95% 的置信区间。99% 的置信区间与之相比会有什么变化?

**5.124** 用临床试验数据库中的数据, 求处理 A 的 Hopkins Obrist 分类积分的 90% 的置信区间。

## 第六章 两总体中心值的比较

- 6.1 引言和案例
- 6.2 关于  $\mu_1 - \mu_2$  的推断:独立样本
- 6.3 非参数推断方法:Wilcoxon 秩和检验
- 6.4 关于  $\mu_1 - \mu_2$  的推断:成对数据
- 6.5 非参数推断方法:Wilcoxon 符号秩检验
- 6.6 推断  $\mu_1 - \mu_2$  时样本容量的选取
- 6.7 小结

### 6.1 引言和案例

到现在为止,我们所做的都是关于一个总体的参数推断。但是更多时候我们所做的推断涉及到不同总体参数的比较问题。比如,我们希望比较两种不同玉米的平均产量,两个不同种族人群的平均年收入,两个不同湖泊的平均含氮量,或者两种不同防眩晕药品从服用到最终见效的平均时间。

在许多抽样情况下,我们分别从两个总体中抽取独立随机样本来比较总体的参数。很多时候,我们都是用相应的样本统计量的差来进行这些推断。假定我们分别独立地从第一个总体中抽取含  $n_1$  个观察值的随机样本,从第二个总体中抽取含  $n_2$  个观察值的随机样本。我们将用样本均值差  $\bar{y}_1 - \bar{y}_2$  对总体均值差  $\mu_1 - \mu_2$  进行推断。

#### 案例研究:石油溢出对植物生长的影响

1992年1月7日,一个地下输油管道破裂,造成了对得克萨斯州(Texas)San Patricio 县 Chiltipin Creek 河沿岸的一个沼泽地的污染。之后,用火烧的方法对这些污染进行了清除。为了评价石油溢出对植物的影响,研究人员设计了一个研究方案,以便对一年之后植物生长进行研究。在一篇未发表的得克萨斯州(Texas), A&M 大学的博士论文中,Newman(1997)介绍了研究人员对 *Distichlis spicata*(一种盐草属植物)的研究结果,其中,*Distichlis spicata* 是对石油溢出地区非常重要的一种植物。

**设计数据收集方案** 研究人员需要确定可能受石油溢出影响的植物的重要特征。在研究之前,需要回答以下问题:

1. 哪些因素决定了这种植物的存活能力?

2. 石油溢出如何影响这些因素?
3. 在石油溢出之前, 是否有关于这些植物因素的资料?
4. 研究者们应该怎样测量石油溢出地区的这些植物因素?
5. 需要多少个观察值来确定在石油溢出之后该植物确实发生了变化?
6. 需要什么样的试验设计或研究?
7. 在对石油溢出之后植物参数所发生的变化进行推断时, 哪些统计方法是可行的?
8. 在最终报告中需要包含哪些信息以证明我们观察到了植物参数的变化?

经过长时间的讨论, 阅读相关文献以及查询类似地区和类似植物的数据库, 研究人员发现没有该种植物在石油溢出之前的具体资料。他们确定有用的参数包括火烧之后 *Distichlis spicata* 的平均密度  $\mu$ , 密度的标准差  $\sigma$ , 以及密度非常接近零的区域所占石油溢出地区的比例  $\pi$ 。由于没有在溢出之前的这种植物的密度的相关资料, 因此, 为了确定植物密度在石油溢出后是否发生了变化, 需要评价一个在这个沼泽地中未受影响的区域的植物密度。研究人员选择了几处未污染的区域, 并把受污染和未受污染的区域都分成了几乎同等大小的地段。研究所需地段的数量由参数  $\mu, \sigma, \pi$  的 95% 置信区间的宽度和假设检验的功效所决定。根据这些计算并考虑到经费预算和时间限制, 研究人员决定从溢出区域和未受影响区域分别随机选取 40 块地段, 并记录其 *Distichlis spicata* 的密度。在沼泽地溢出区域也作同样的测量。

对照区域				溢出区域			
均值:	38.48	000	0		均值:	26.93	
中位数:	41.50	7	0	59	中位数:	26.00	
标准差:	16.37	1	1	14	标准差:	9.88	
$n$ :	40	6	1	77799	$n$ :	40	
		4	2	2223444			
		9	2	555667779			
		0	3	11123444			
		55678	3	5788			
		000111222233	4	1			
		57	4				
		0112344	5	02			
		67789	5				

图 6.1 在石油溢出区域和对照区域所观察到的地段的植物株数。  
数据以茎叶图的方式展示



**整理数据** 这些数据包括未受污染(对照)区域 40 个植物密度测量值和溢出(燃烧)区域 40 个植物密度测量值。然后研究人员通过实地考察认真审核这些数据以检查它们是否被正确地记录。随后将这些数据转换成计算机文档,为以后的分析做好准备。这些分析步骤已在第 2.5 节作了介绍。

**分析数据** 下一步将通过图和统计量的方式对这些数据进行概括。图 6.1 中展示了这些数据。表 6.1 列出了一些统计量。

表 6.1 石油溢出数据的概括统计量

描述统计量						
变量	区域类型	N	均值	中位数	截尾均值	标准差
植株数量	对照	40	38.48	41.50	39.50	16.37
	溢出	40	26.93	26.00	26.69	9.88
变量	区域类型	均方	极小值	极大值	Q1	Q2
植株数量	对照	2.59	0.00	59.00	35.00	51.00
	溢出	1.56	5.00	52.00	22.00	33.75

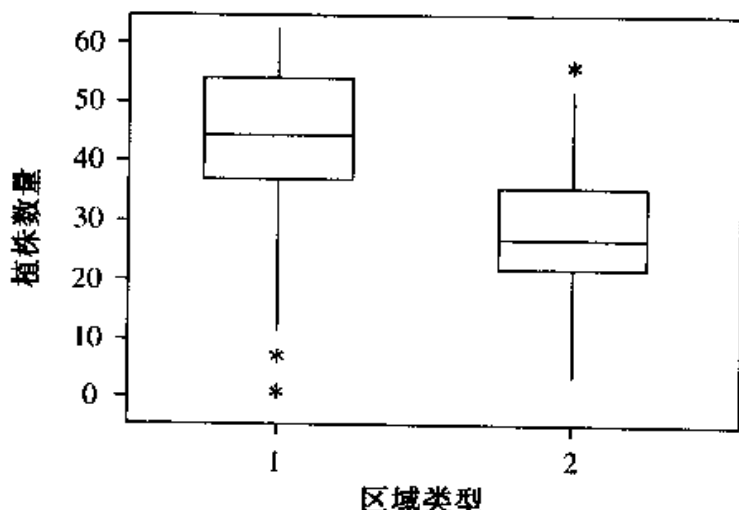


图 6.2 对照区和溢出区地段上的植物数量

图 6.2 中的盒形图表明对照区的植物密度大于溢出区。根据统计量,我们得到对照区植物密度的均值  $\bar{y}_{\text{con}} = 38.48$ , 标准差  $s_{\text{con}} = 16.37$ , 溢出区植物密度均值  $\bar{y}_{\text{spil}} = 26.93$ , 标准差  $s_{\text{spil}} = 9.88$ 。因此,对照区植物密度的均值和变异均大于溢出区。为了确定此差异是否能代表所有地区而不仅是研究中所涉及的区域,研究人员需要对数据进行统计分析。在本章的以下部分,我们将讨论置信区间的建立和

总体均值  $\mu_{\text{con}}$  与  $\mu_{\text{spil}}$  之差的检验问题。在第七章和第十章中讨论总体标准差  $\sigma$  和总体比例  $\pi$  的估计和检验问题。

在求由独立随机样本产生的样本统计量差的抽样分布时,我们会发现下述定理是很有用的。

**定理 6.1** 如果两个独立随机变量  $y_1$  和  $y_2$  服从正态分布,均值和方差分别为  $(\mu_1, \sigma_1^2)$  和  $(\mu_2, \sigma_2^2)$ ,那么它们的差服从均值为  $\mu_1 - \mu_2$ ,方差为  $\sigma_1^2 + \sigma_2^2$  的正态分布,它们的和  $y_1 + y_2$  也服从正态分布,其均值为  $\mu_1 + \mu_2$ ,方差为  $\sigma_1^2 + \sigma_2^2$ 。

定理 6.1 可直接应用于求两个独立样本均值差或比例差的抽样分布。根据中心极限定理(见第四章),如果两个容量分别为  $n_1$  和  $n_2$  的随机样本是从两个总体中独立地抽出的,那么若  $n_1$  和  $n_2$  足够大,则样本均值  $\bar{y}_1$  和  $\bar{y}_2$  近似地服从正态分布,均值和方差分别为  $(\mu_1, \sigma_1^2/n_1)$  和  $(\mu_2, \sigma_2^2/n_2)$ 。这样,因为  $\bar{y}_1$  和  $\bar{y}_2$  是独立的且都服从正态分布的随机变量,根据定理 6.1,  $\bar{y}_1 - \bar{y}_2$  也近似服从正态分布,其均值为

$$\mu_{\bar{y}_1 - \bar{y}_2} = \mu_1 - \mu_2,$$

方差为

$$\sigma_{\bar{y}_1 - \bar{y}_2}^2 = \sigma_{\bar{y}_1}^2 + \sigma_{\bar{y}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

标准差为

$$\sigma_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

两个独立的正态分布样本均值差的抽样分布见图 6.3。

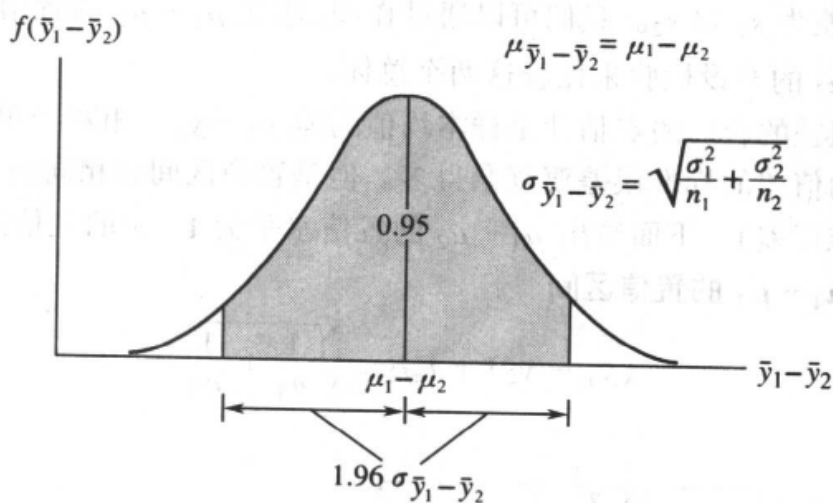


图 6.3 两样本均值差的抽样分布

两个样本均值差 $(\bar{y}_1 - \bar{y}_2)$ 的抽样分布的性质:

1. 当样本容量充分大时,  $(\bar{y}_1 - \bar{y}_2)$ 近似地服从正态分布。

2. 抽样分布的均值  $\mu_{\bar{y}_1 - \bar{y}_2}$  等于两总体的均值差  $\mu_1 - \mu_2$ 。

3. 抽样分布的标准为  $\sigma_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

根据两样本均值差 $(\bar{y}_1 - \bar{y}_2)$ 的抽样分布, 我们可以回答与在第四章中提出的关于  $\bar{y}$  的抽样分布同样的问题。因为样本统计量是用于对相应的总体参数进行推断的, 我们可以根据一个统计量的抽样分布来计算其值落在距总体参数一定距离内的概率。比如我们可根据样本均值的抽样分布来计算 $(\bar{y}_1 - \bar{y}_2)$ 落在距未知总体均值差  $\mu_1 - \mu_2$  某一距离内的概率。本章后面的几节我们将讨论对  $\mu_1 - \mu_2$  的推断(估计和检验)问题。

## 6.2 关于 $\mu_1 - \mu_2$ 的推断: 独立样本

基于独立地取自于两总体的随机样本, 当我们要对  $\mu_1 - \mu_2$  进行推断时, 需考虑以下三种情形:

情形 1. 两个总体均服从正态分布, 且  $\sigma_1 = \sigma_2$ 。

情形 2. 样本容量  $n_1$  和  $n_2$  很大。

情形 3. 样本容量  $n_1$  或  $n_2$  很小, 两总体分布均不是正态分布。

本节中, 我们将考虑这样一种情形, 即, 两随机样本独立地取自于均值分别为  $\mu_1, \mu_2$ , 标准差相同  $\sigma_1 = \sigma_2 = \sigma$  的两正态总体。数据用统计量来概括: 样本均值  $\bar{y}_1$  和  $\bar{y}_2$ , 样本标准差为  $s_1$  和  $s_2$ 。我们可以通过作图, 建立  $\mu_1 - \mu_2$  的置信区间, 以及进行关于  $\mu_1 - \mu_2$  的假设检验来比较这两个总体。

总体均值差的合理的点估计是样本均值的差  $\bar{y}_1 - \bar{y}_2$ 。相对于单样本的均值, 计算两样本均值差的标准误差要复杂得多。但是置信区间有相同的形式, 即: 点估计  $\pm t_{\alpha/2}$ (标准误差)。下面给出  $\mu_1 - \mu_2$  的置信水平为  $1 - \alpha$  的置信区间。

**独立样本下  $\mu_1 - \mu_2$  的置信区间**

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2, s_p} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

其中

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad \text{自由度 df} = n_1 + n_2 - 2$$

$\bar{y}_1 - \bar{y}_2$  的抽样分布是正态分布。因为我们要求两总体具有相同的标准差  $\sigma$ , 所以该正态分布标准差为

$$\sigma_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

如果我们知道  $\sigma$  的值, 就可以在计算置信区间的公式中用  $z_{\alpha/2}$ 。因为在多数情况下  $\sigma$  未知, 所以我们必须先对它进行估计。该估计通常被记作  $s_p$ , 它由  $\sigma$  的两个独立估计  $s_1$  和  $s_2$  混合而得。事实上,  $s_p^2$  是样本方差  $s_1^2$  和  $s_2^2$  的加权平均。由于我们必须估计  $\mu_1 - \mu_2$  的点估计的标准差, 因此需用  $t$  分布的分位数  $t_{\alpha/2}$  来代替正态分布的分位数  $z_{\alpha/2}$ 。因为共有  $n_1 + n_2$  个观察数据, 且在估计标准差  $\sigma$  之前需估计两个参数  $\mu_1$  和  $\mu_2$ , 所以  $t$ -分位数的自由度为  $df = n_1 + n_2 - 2$ 。注意在计算  $s_1^2$  和  $s_2^2$  的公式中, 应该用  $\bar{y}_1$  和  $\bar{y}_2$  分别代替  $\mu_1$  和  $\mu_2$ 。

我们曾假定两个抽样总体服从正态分布且有相同的方差。如果以上给出的置信区间公式只有当这个假定严格满足时才成立, 那么这个估计方法的应用便十分有限。幸运的是如果两总体的分布都是钟形的, 并且样本容量大体相同, 那么置信系数相当稳定。当这些条件不能满足时, 我们将在本节和 6.3 中讨论其他的方法。

### 例 6.1

一个公司的领导往往很关心一种药品的有效时间。在某种药品的生产线上随机抽取  $n_1 = 10$  瓶进行药效分析。

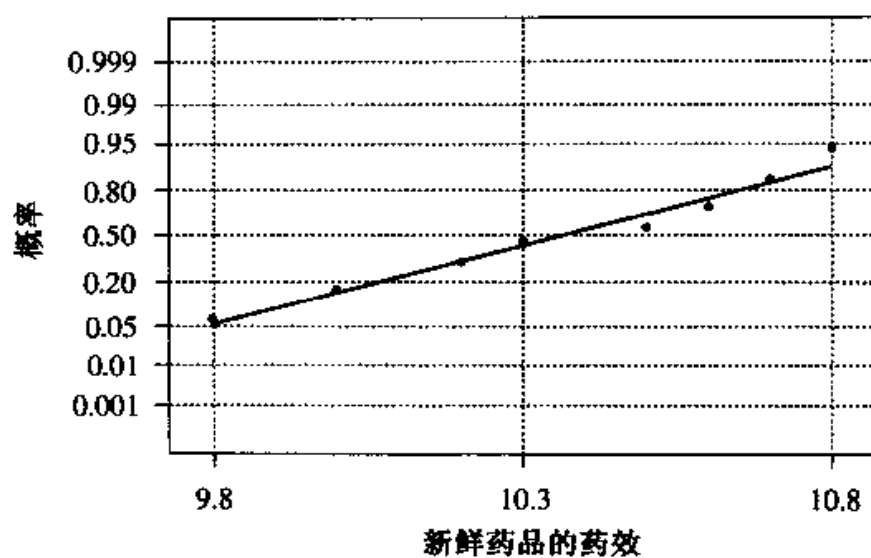
另一个容量为  $n_2 = 10$  瓶的样本被在规定的存储条件下存储一年。两个样本经测试所得的结果见表 6.2。

表 6.2 两样本的药效数据

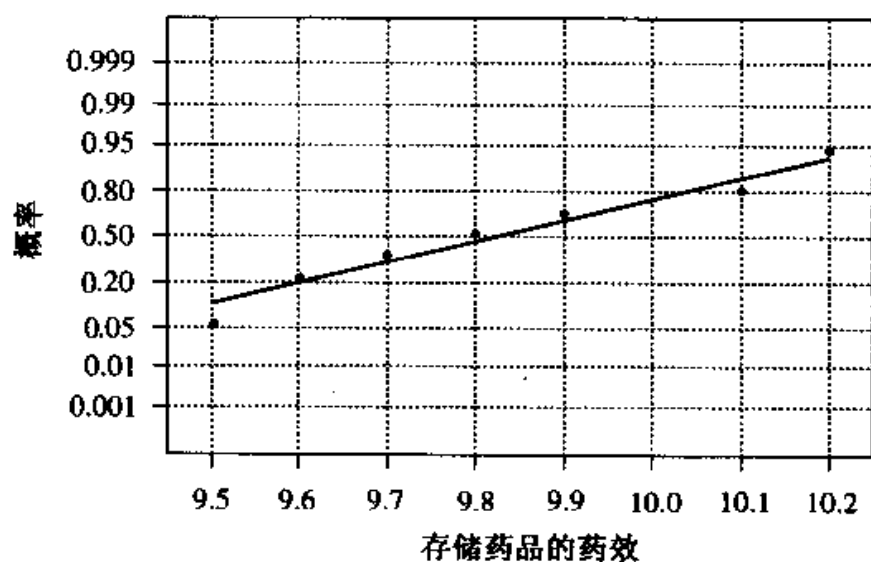
新鲜		存储	
10.2	10.6	9.8	9.7
10.5	10.7	9.6	9.5
10.3	10.2	10.1	9.6
10.8	10.0	10.2	9.8
9.8	10.6	10.1	9.9

假定令  $\mu_1$  表示第一个样本所来自的生产线生产的所有瓶药的平均药效,  $\mu_2$  表示所有那些可能储存一年的药品的平均药效。求  $\mu_1 - \mu_2$  的 95% 置信区间。

**解答** 根据以上新鲜药品和存储药品的数据分别作正态概率图, 见图 6.4(a) 和 (b), 据此来确定它们是否服从正态分布。我们发现这些点非常接近一条直线, 两个样本的正态性条件都满足。根据这两个样本数据的统计量如下。



(a) 正态概率图：新鲜药品的药效



(b) 正态概率图：存储药品的药效

图 6.4

新鲜药品

$$n_1 = 10$$

$$\bar{y}_1 = 10.37$$

$$s_1 = 0.3234$$

存储药品

$$n_2 = 10$$

$$\bar{y}_2 = 9.83$$

$$s_2 = 0.2406$$

在第七章,我们将给出关于两总体方差相等的检验。但是对于以上数据,考虑到样本容量很小,计算出的两个样本标准差近似相等。这样,我们构造  $\mu_1 - \mu_2$  置信区间所需的条件——即正态性,等方差性和独立随机样本——都满足。共同标准差  $\sigma$  的估计是

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{9(0.3234)^2 + 9(0.2406)^2}{18}} = 0.285$$

对于  $df = n_1 + n_2 - 2 = 18$ ,  $\alpha = 0.025$ ,  $t$ -值为 2.101。平均药效差的 95% 置信区间为

$$(10.37 - 9.83) \pm 2.101(0.285) \sqrt{1/10 + 1/10} \\ 0.54 \pm 0.268 \text{ 或 } (0.272, 0.808)$$

于是,我们估计新鲜药品与存储一年期药品的平均药效差  $\mu_1 - \mu_2$  在 0.272 到 0.808 之间。这样,公司领导需要评估平均药效下降幅度在 0.272 到 0.808 之间是否会对此药品的实际效用产生影响。

### 例 6.2

某学区认为在他们的高中就读的学生数量几乎难以控制,所以将此街区分成两部分:街区 1 和街区 2。街区 1 的学生仍在原学校就读,而街区 2 的学生去一个新建的学校。一些家长认为两个街区是根据收入水平划分的。因此,进行了一项调查以判定街区 1 和街区 2 的居民的平均收入水平是否存在差异。在街区 1 内随机抽取 20 个户主,但在街区 2 的即将访问的 20 个户主中,尽管研究人员保证访问的保密性,仍有一人拒绝提供要求的信息。这样在街区 2 的样本中只有 19 个观察值。通过这些数据(以 1000 美元为单位),我们可得出样本均值和样本方差,见表 6.3。根据这些数据构造  $\mu_1 - \mu_2$  的 95% 置信区间。

表 6.3 例 6.2 中的收入数据

	街区 1	街区 2
样本容量	20	19
样本均值	18.27	16.78
样本方差	8.74	6.58

**解答** 针对两个样本的数据,用直方图进行初步分析表明总体分布是丘形的(接近正态),并且,样本方差也是相近。样本均值的差为

$$\bar{y}_1 - \bar{y}_2 = 18.27 - 16.78 = 1.49$$

共同标准差  $\sigma$  的估计为

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{19(8.74) + 18(6.58)}{20 + 19 - 2}} = 2.77$$

对于  $\alpha = \alpha/2 = 0.025$  和  $df = 20 + 19 - 2 = 37$ ,  $t$ -分位数没有在附表 2 中列出。但是取小于 37 且最接近 37 的  $df$  值(35), 我们可得出  $t_{0.025} = 2.030$ 。两街区的平均收入差的 95% 置信区间可由以下公式给出

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

代入数据可得到

$$1.49 \pm 2.030(2.77) \sqrt{\frac{1}{20} + \frac{1}{19}} \quad \text{或} \quad 1.49 \pm 1.80$$

这样我们估计出平均收入的差会落在区间  $(-0.31, 3.29)$  的某处。将上下限均乘以 1000, 置信区间就变为  $(-310, 3290)$ 。这个区间表示  $\mu_1 - \mu_2$  可能为负, 也可能为正, 故不能认为街区 1 的收入大于或小于街区 2 的收入。

我们可以对两个总体均值差进行假设检验。正如任何一种检验方法一样, 我们都是先对总体均值差给出一个研究假设, 比如我们可假设  $\mu_1 - \mu_2$  大于某个值  $D_0$  (注:  $D_0$  通常是 0)。整个检验过程如下。

#### 独立样本下 $\mu_1 - \mu_2$ 的统计检验

使得检验方法有效所需的假定条件等同于构造  $\mu_1 - \mu_2$  置信区间所需的条件, 即两个总体分布是正态的、方差相等的, 且两随机样本独立。

$$H_0: 1. \mu_1 - \mu_2 \leq D_0 \quad (D_0 \text{ 是给定的值, 通常为 } 0)$$

$$2. \mu_1 - \mu_2 \geq D_0$$

$$3. \mu_1 - \mu_2 = D_0$$

$$H_a: 1. \mu_1 - \mu_2 > D_0$$

$$2. \mu_1 - \mu_2 < D_0$$

$$3. \mu_1 - \mu_2 \neq D_0$$

$$T. S.: \quad t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

R. R.: 对于第一类错误概率  $\alpha$  及自由度  $df = n_1 + n_2 - 2$

1. 若  $t \geq t_\alpha$ , 则拒绝  $H_0$ 。

2. 若  $t \leq -t_\alpha$ , 则拒绝  $H_0$ 。

3. 若  $|t| \geq t_{\alpha/2}$ , 则拒绝  $H_0$ 。

检查假定, 并作出结论。

#### 例6.3

为了评价一种被用于杀死羊胃中绦虫的疗法的效用, 进行了一个试验。随机抽取 24 只被绦虫感染的健康的年龄大致相同的羊, 并把它们随机平均分成两组。

一组被注射了某种药品, 另一组不加治疗。6 个月后, 这些羊被屠宰。我们记录了这些羊胃中绦虫的数量:

治疗	18	43	28	50	16	32	13	35	38	33	6	7
未治疗	40	54	26	63	21	37	39	23	48	58	28	39

- 检验治疗过的羊胃中绦虫的平均数量是否小于未经治疗的羊。( $\alpha = 0.05$ )
- 该检验的显著性水平是多少?
- 给出  $\mu_1 - \mu_2$  的 95% 的置信区间以估计两组均值差的大小。

### 解答

a. 盒形图如图 6.5 所示。图中给出处理(治疗)和未处理两组羊胃中所含绦虫的数据。从图中, 我们可以看出未处理组数据是对称的, 且没有异常值, 处理组数据则有点左偏倚, 且也没有异常值。同样两个盒的宽度也几乎相等。因此总体服从正态分布且方差相等这一条件能够得到满足。两组之间和组内绦虫数量的独立性这一条件是否满足尚需考虑这些羊是如何被选出, 如何被分配到两个组, 以及 6 个月试验期内如何被照料。这 24 只羊是从一群有代表性的被感染的羊中随机

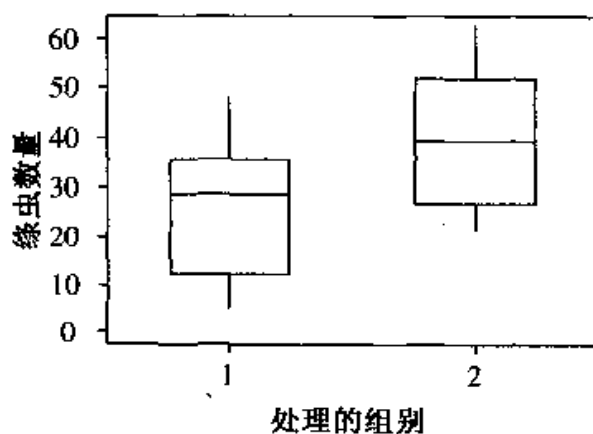


图 6.5 绦虫数量的盒形图

1 处理组 2 未处理组

抽取的, 也被随机地分配到处理组和未处理组中, 在 6 个月期间被适当地隔开, 并得到了良好照顾, 所以这 24 个数据可视为来自两总体的独立随机样本。最后, 我们可以看出未处理组羊的绦虫数量高于处理组羊, 因此前者的中值线较靠上。下面的检验确认了我们的想法。处理组和未处理组的数据概括如下:



处理组羊	未处理组羊
$n_1 = 12$	$n_2 = 12$
$\bar{y}_1 = 26.58$	$\bar{y}_2 = 39.67$
$s_1 = 14.36$	$s_2 = 13.86$

因为样本标准差几乎相同, 据此并根据盒形图, 我们可以计算共同的总体标准差  $\sigma$  的混合估计:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{11(14.36)^2 + 11(13.86)^2}{22}} = 14.11$$

研究假设为: 处理组羊的胃中所含绦虫的平均数量( $\mu_1$ )小于未处理组羊的胃中所含绦虫平均数量( $\mu_2$ )。检验该研究假设的步骤如下:

$H_0: \mu_1 - \mu_2 \geq 0$  (也就是说, 药并没有减少平均绦虫数量)

$H_a: \mu_1 - \mu_2 < 0$  (也就是说, 药减少了平均绦虫数量)

$$T. S.: t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(26.58 - 39.67) - 0}{14.11 \sqrt{\frac{1}{12} + \frac{1}{12}}} = -2.272$$

R.R.: 对于  $\alpha = 0.05$ , 查附表 2 可得自由度  $df = n_1 + n_2 - 2 = 22$  的单侧检验的  $t$ -临界值。若  $t \leq -1.717$ , 则拒绝  $H_0$ 。

结论: 因为观察值  $t = -2.272$  小于  $-1.717$ , 落在拒绝域内, 所以, 我们有足够的理由认为服用此药品确实能减少绦虫数量。

b. 根据附表 2, 当  $t = -2.272$ ,  $df = 22$  时, 我们可确定显著性水平落在  $0.01 < p < 0.025$  的范围内, 根据下面列出的计算结果, 我们可看出精确的显著性水平  $p$  值  $= 0.017$ 。

#### Two-Sample T-Test and Confidence Interval

##### Two-sample T for Treated vs Untreated

	N	Mean	StDev	SE Mean
Treated	12	26.6	14.4	4.1
Untreated	12	39.7	13.9	4.0

95% CI for  $\mu$  Treated -  $\mu$  Untreated:  $(-25.0, -1.1)$

T-Test  $\mu$  Treated =  $\mu$  Untreated (vs  $<$ ):  $T = -2.27$   $P = 0.017$   $DF = 22$

Both use Pooled StDev = 14.1

c.  $\mu_1 - \mu_2$  的 95% 置信区间为试验者提供了使用药物后绦虫减少数量的估计。区间估计计算如下:

$$(\bar{y}_1 - \bar{y}_2) \pm t_{0.025, s_p} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$(26.58 - 39.67) \pm (2.074)(14.11) \sqrt{\frac{1}{12} + \frac{1}{12}} \quad \text{或} \quad -13.09 \pm 11.95$$

这样我们有 95% 的把握认为服用药物带来的绦虫减少量在 1.1 到 25.0 之间。

在这一节中, 为比较两个总体均值而构造的置信区间和假设检验需满足三个条件。第一个, 也是最关键的一个就是两随机样本相互独立。在实际中, 这种独立性是指两个样本是从两个不同总体中随机抽取的, 并且一个样本中的元素与另一个样本中的元素在统计意义上相互独立。在试验和研究中经常会出现两种类型的相关(数据并非独立)。数据可能会有类聚效应, 之所以会出现这种情况, 一般是因为数据是从子群中选出的。比如在一个试验中, 从 5 个班级中选出 50 个孩子, 对两种教学方式进行比较, 这些孩子被随机指定一种教学方式。因为同一个班的孩子有同一个老师, 因此从同一班内抽取的同学的成绩就会比从不同班级抽取的同学的成绩更接近。这样, 参加者的独立性就不能被满足。

另一种相关是由于序列和空间相关。当数据是在一定时间内测量得到时, 在较接近的时间内的观察值之间往往比较远的时间内的观察值之间更为接近。一个类似的相关是数据取自不同地点——例如, 为了估计一个化工厂是否排放了污染物, 而在一湖的不同位置取水样。较接近的地点所得的数据往往比较远的地点所得的数据更为接近, 这种相关叫空间相关。当数据相关时, 那么基于  $t$  分布方法构造的置信区间的覆盖概率不同于预期的概率值, 假设检验犯第 I 类错误的概率也不同于给定的值。有适于处理这类数据的统计方法, 但他们都太深奥了。关于纵向或重复测量数据分析或空间数据分析的书会介绍分析相关性数据的详细方法。

当总体分布严重厚尾或严重偏倚时, 置信区间的覆盖概率以及  $t$  检验的水平和功效就会远远偏离预定的值。下一节将介绍可替代  $t$  检验的一个非参数检验, 这种检验不要求总体的正态性。

第三个也是最后一个假定就是两个总体方差  $\sigma_1^2$  和  $\sigma_2^2$  相等。现在我们仅考察样本方差, 看他们是否近似相等。后面(第七章)我们再对这一假定进行检验。关于偏离等方差性假定对独立样本的  $t$  方法的影响, 人们已经做了很多研究。一般结论为: 对于容量相同的样本, 总体方差之间相差 3 倍时(比如,  $\sigma_1^2 = 3\sigma_2^2$ ),  $t$  检验方法仍然适用。

为了说明不等方差的影响, 我们在计算机上进行了模拟, 从相同均值, 不等方差:  $\sigma_1 = k\sigma_2$ , 其中  $k = 0.25, 0.5, 1, 2, 4$  的正态总体中产生两个独立随机样本。对于每个样本容量和标准差的组合进行 1,000 次模拟。对每次模拟, 进行水平为 0.05 的检验。在表 6.4 中列出了 1,000 次检验中错误地拒绝  $H_0$  的比率。如果混

合  $t$  检验不受不等方差的影响, 那么得到的比率就应接近于 0.05, 即在所有情况下的预定值。

表 6.4 不等方差对混合  $t$  检验犯第 I 类错误比率的影响

$n_1$	$n_2$	$\sigma_1 = k\sigma_2$				
		$k = 0.25$	0.50	1	2	4
10	10	0.065	0.042	0.059	0.045	0.063
10	20	0.016	0.017	0.049	0.114	0.165
10	40	0.001	0.004	0.046	0.150	0.307
15	15	0.053	0.043	0.056	0.060	0.060
15	30	0.007	0.023	0.066	0.129	0.174
15	45	0.004	0.010	0.069	0.148	0.250

根据表 6.4 的结果, 我们可以得出这样的结论: 当样本容量相等时, 犯第 I 类错误的比率总是接近于 0.05 (从 0.042 到 0.065)。当样本容量不等时, 犯第 I 类错误的比率就大大偏离 0.05。当样本容量较小且总体方差较大时, 偏离更为严重。这时错误率高于 0.05。例如, 当  $n_1 = 10, n_2 = 40, \sigma_1 = 4\sigma_2$  时, 错误率为 0.307。当  $n_1 = 10, n_2 = 10, \sigma_1 = 4\sigma_2$  时, 错误率为 0.063, 非常接近于 0.05。这一点很关键, 它为我们作  $t$  检验时使用容量相等的样本提供了令人信服的证据。

当样本方差 ( $s_1^2$  和  $s_2^2$ ) 显示  $\sigma_1^2 \neq \sigma_2^2$  时, 用下面的检验统计量作近似的  $t$  检验

$$t' = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Welch(1938)证明了具有修正后自由度的  $t$  分布分位点, 称作 Satterthwaites 近似, 可用于确定  $t'$  的拒绝域。近似  $t$  检验步骤如下:

对于独立样本, 不等方差时的近似  $t$  检验

$$\begin{array}{ll} H_0: 1. \mu_1 - \mu_2 \leq D_0 & H_a: 1. \mu_1 - \mu_2 > D_0 \\ 2. \mu_1 - \mu_2 \geq D_0 & 2. \mu_1 - \mu_2 < D_0 \\ 3. \mu_1 - \mu_2 = D_0 & 3. \mu_1 - \mu_2 \neq D_0 \end{array}$$

$$T. S.: t' = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

R. R.: 对于第一类错误概率  $\alpha$  及自由度  $df$

1. 若  $t' \geq t_\alpha$ , 则拒绝  $H_0$ 。
2. 若  $t' \leq -t_\alpha$ , 则拒绝  $H_0$ 。

3. 若  $|t'| \geq t_{\alpha/2}$ , 则拒绝  $H_0$ 。

$$\text{其中 } df = \frac{(n_1 - 1)(n_2 - 2)}{(1 - c)^2(n_1 - 1) + c^2(n_2 - 1)}, \text{ 而 } c = \frac{s_1^2/n_1}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

注: 如果计算出  $df$  不是整数, 四舍五入, 取其最接近的整数。

$t'$  检验有时被称为分离方差  $t$  检验 (separate-variance  $t$  test), 因为所用的是各自的样本方差  $s_1^2$  和  $s_2^2$ , 而不是混合样本方差。

当  $\sigma_1^2$  和  $\sigma_2^2$  相差很大时, 我们必须修正  $\mu_1 - \mu_2$  的置信区间。由分离方差  $t$  检验给出以下置信区间公式:

对于独立样本, 方差不等时  $\mu_1 - \mu_2$  的近似置信区间

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\text{其中 } df = \frac{(n_1 - 1)(n_2 - 2)}{(1 - c)^2(n_1 - 1) + c^2(n_2 - 1)}, \text{ 而 } c = \frac{s_1^2/n_1}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

下面我们继续对案例中的数据进行分析。

**分析石油溢出案例中的数据** 研究人员假定被污染区域的植物密度要小于对照区域。这样, 我们就可分别对对照区域的植物密度 ( $\mu_1$ ) 和溢出区的密度 ( $\mu_2$ ) 构建置信区间来估计他们的平均密度。同样可以构建  $\mu_1 - \mu_2$  的置信区间来检验研究假设“ $\mu_1$  大于  $\mu_2$ ”。从图 6.1 可看出, 溢出区的数据看似正态分布, 而对照区的数据则呈左偏倚。根据图 6.6(a) 和 (b) 所示的正态概率图可进一步判定总体是否服从正态分布。我们可以看出溢出区的数据服从正态分布, 而对照区的数据不服从正态分布, 因为图中的点不近似成一条直线。并且, 对照区植物密度变异性大于溢出区。这样近似  $t$  检验就应是最合适的推断方法。由样本数据得结果如下:

对照区	溢出区
$n_1 = 40$	$n_2 = 40$
$\bar{y}_1 = 38.48$	$\bar{y}_2 = 26.93$
$s_1 = 16.37$	$s_2 = 9.88$

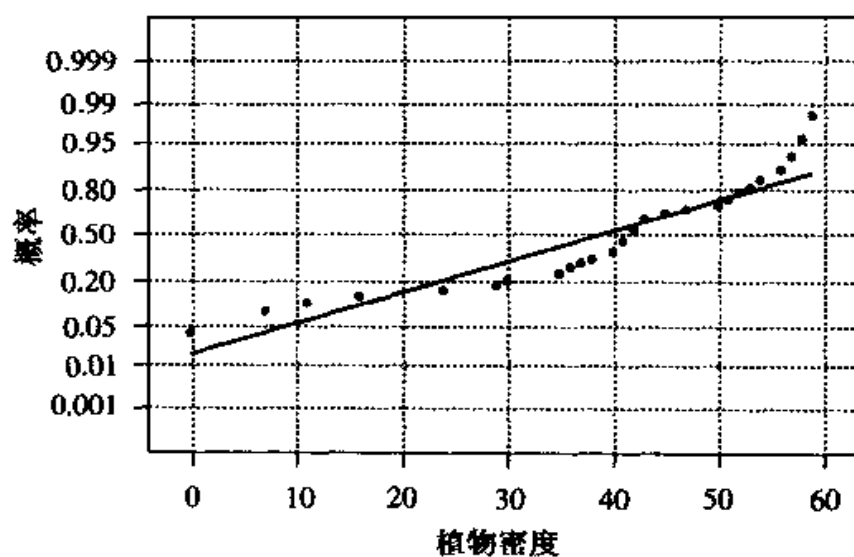
研究假设为对照区的平均植物密度超过溢出区。近似  $t$  检验的步骤如下:

$$H_0: \mu_1 \leq \mu_2 \quad H_a: \mu_1 > \mu_2$$

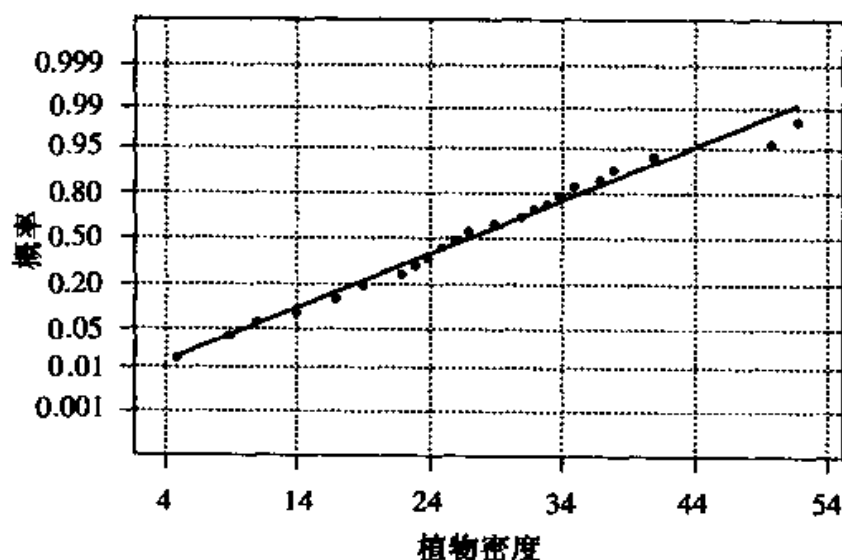
$$\text{即, } H_0: \mu_1 - \mu_2 \leq 0$$

$$H_a: \mu_1 - \mu_2 > 0$$

$$\text{T. S.: } t' = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(38.48 - 26.93) - 0}{\sqrt{\frac{(16.37)^2}{40} + \frac{(9.88)^2}{40}}} = 3.82$$



(a) 对照区正态概率图



(b) 溢出区正态概率图

图 6.6

为了计算拒绝域和  $p$  值, 我们需要计算  $t'$  的近似自由度  $df$ :

$$c = \frac{\frac{s_1^2/n_1}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(16.37)^2/40}{(16.37)^2/40 + (9.88)^2/40} = 0.73$$

$$df = \frac{(n_1 - 1)(n_2 - 2)}{(1 - c)^2(n_1 - 1) + c^2(n_2 - 1)}$$

$$= \frac{(39)(39)}{(1 - 0.73)^2(39) + (0.73)^2(39)} = 64.38$$

取整为 64

附表 2 中没有  $df=64$  的情况,所以我们将用  $df=60$ 。事实上,当自由度逐渐变大时,分位数的差异很小:对于  $df=60$  和  $64$ ,  $t_{0.05}$  分别等于 1.671 和 1.669。

R. R.: 对于  $\alpha=0.05$ ,  $df=60$ , 若  $t' > 1.671$ , 则拒绝  $H_0$ 。

因为  $t' = 3.82 > 1.671$ , 所以我们拒绝  $H_0$ 。我们可根据附表 2 中  $df=60$  的情况界定  $p$  值。当  $t' = 3.82$  时,显著性水平  $p < 0.001$ 。因此我们可以得出结论:有明显的证据表明  $\mu_1$  大于  $\mu_2$ 。尽管我们已确定对照区和溢出区的植物密度在统计意义上存在显著差异,但问题是这些差异是否具有实际意义。我们可以用 95% 置信区间估计平均值差异的大小。

利用以下公式计算  $\mu_1 - \mu_2$  的 95% 置信区间,其中  $df=64$ ,该公式与 R. R 中所用的公式相同:

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\text{或 } (38.48 - 26.93) \pm 2.0 \sqrt{\frac{(16.37)^2}{40} + \frac{(9.88)^2}{40}}$$

$$\text{或 } 11.55 \pm 6.05$$

这样我们有 95% 的把握相信平均植物密度的差异在 5.5 到 17.6 之间。植物学家可根据这个差异来评价其是否具有实际意义。

**报告结论** 我们需对此研究写一份总结报告,内容包括以下几项:

1. 说明研究目的。
2. 叙述如何设计研究方案,如何收集数据。
3. 数据集的数值和图形概括。
  - 均值,中位值,标准差,分位数和极差表。
  - 盒形图。
  - 茎叶图。
4. 描述所有推断方法。
  - 均值差的近似  $t$  检验。
  - 基于近似  $t$  方法的总体均值置信区间。
  - 用盒形图和正态概率图验证推断方法所需的所有条件是否满足。
5. 讨论所得的结果和结论。
6. 相对于以前的研究结果解释新的发现。
7. 提出未来研究的建议。
8. 列出数据集。

为了说明不等方差对分离方差  $t$  检验的影响较对混合  $t$  检验的影响要小, 用分离方差  $t$  检验的方法分析表 6.4 中的计算机模拟数据。表 6.5 列出了 1,000 次检验中, 错误地拒绝  $H_0$  的比率。如果分离方差  $t$  检验不受不等方差的影响, 那么, 我们期望得到的比率就应接近于 0.05, 即在所有情况下的预定值。

表 6.5 在分离方差  $t$  检验中不等方差对犯第 I 类错误比率的影响

$n_1$	$n_2$	$\sigma_1 = k\sigma_2$				
		$k = 0.25$	0.50	1	2	4
10	10	0.055	0.040	0.056	0.038	0.052
10	20	0.055	0.044	0.049	0.059	0.051
10	40	0.049	0.047	0.043	0.041	0.055
15	15	0.044	0.041	0.054	0.055	0.057
15	30	0.052	0.039	0.051	0.043	0.052
15	45	0.058	0.042	0.055	0.050	0.058

从表 6.5 我们可以看出, 在所考察的每一种情况下, 分离方差  $t$  检验犯第 I 类错误的比率都很接近 0.05。另一方面, 当样本容量不同且来自具有不等方差的总体时, 混合  $t$  检验犯第 I 类错误的比率与 0.05 相差很多。

在本节中, 针对独立样本来自于等方差正态总体的情况, 我们研究了混合  $t$  检验。对于方差不等的情况, 引入了分离方差  $t'$  检验统计量。基于  $t$  和  $t'$  两种方法所得到的置信区间和假设检验的结果不必一致。标准统计软件一般都给出两种检验结果。在你的报告中应该用哪一种?

如果样本容量相同, 总体方差也相同, 分离方差  $t$  检验和混合  $t$  检验从代数上给出了相同的结果, 即  $t = t'$ 。既然如此为什么不在  $n_1 = n_2$  的时候总是用  $t'$  检验呢? 我们选  $t$  而不是  $t'$  是因为  $t$  的自由度几乎总是大于  $t'$  的自由度。所以当方差相等时,  $t$  检验的功效大于  $t'$ 。当两样本容量以及方差都差别很大时,  $t$  和  $t'$  检验的结果可能会大不相同。在这种情况下, 有依据表明分离方差方法所得的结果一般比混合  $t$  方法更可靠和保守一些。但是如果总体的方差和均值都不同时, 仅对总体均值差  $\mu_1 - \mu_2$  检验远不足以说明总体的差别。我们应同时检验被比较的总体均值之间的差异和方差之间的差异。第七章将讨论两方差差异的推断问题。

## 练习

### 基本技能

6.1 试针对下列情况, 分别给出检验  $H_0: \mu_1 - \mu_2 = 0$  的拒绝域

a.  $H_a: \mu_1 - \mu_2 \neq 0, n_1 = 12, n_2 = 14, \alpha = 0.05$ 。

b.  $H_a: \mu_1 - \mu_2 > 0, n_1 = n_2 = 8, \alpha = 0.01$ 。

c.  $H_a: \mu_1 - \mu_2 < 0, n_1 = 6, n_2 = 4, \alpha = 0.05$ 。

应用两样本  $t$  检验前应作哪些假定?

6.2 给定  $\alpha = 0.05$ , 利用下面的样本数据检验零假设  $H_0: \mu_1 - \mu_2 \geq 0$  对备择假设  $H_a: \mu_1 - \mu_2 < 0$ 。

	总 体	
	1	2
样本容量	16	13
样本均值	71.5	79.8
样本方差	68.35	70.26

6.3 参见习题 6.2 中的数据。给出检验的显著性水平。

## 应用

6.4 (医学) 为了研究寒冷环境对高血压患者的影响, 首先对老鼠进行试验, 调查低温对老鼠的高血压的影响, 随机抽取 12 只老鼠并把它们分成两组, 一组被置于室温(26℃)下, 另一组置于 5℃ 的环境中。测试两组老鼠的血压和心率。下表给出了这 12 只老鼠的血压。

26℃		5℃	
老鼠	血压	老鼠	血压
1	152	7	384
2	157	8	369
3	179	9	354
4	182	10	375
5	176	11	366
6	149	12	423

a. 这些数据是否能充分地说明 5℃ 环境中老鼠的平均血压高于室温下的老鼠? 已知  $\alpha = 0.05$ 。

b. 评估(a)中检验所需的三个条件是否满足?



c. 给出两个总体均值差的 95% 置信区间。

**6.5(环境)** 一个环境控制检验员怀疑一个河边的社区往河里排放半处理的污水, 这会导致河水中被溶解氧气的变化。为了证实这一怀疑, 他分别在这个城镇的上下游各随机抽取了 15 个水样, 下表给出这 30 处水样中被溶解氧气(单位 ppm)的数据。

上游	5.2	4.8	5.1	5.0	4.9	4.8	5.0	4.7
下游	4.2	4.4	4.7	4.9	4.6	4.8	4.9	4.6
上游	4.7	5.0	4.7	5.1	5.0	4.9	4.9	
下游	5.1	4.3	5.5	4.7	4.9	4.8	4.7	

根据计算机输出的结果回答下列问题

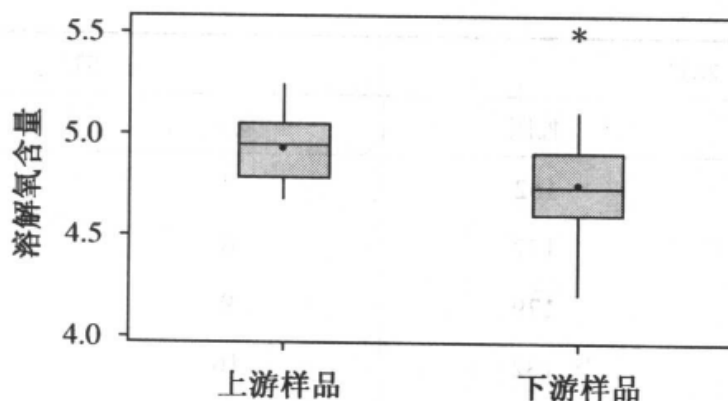
Two-Sample T-Test and Confidence Interval

Two-sample T for Above Town vs Below Town

	N	Mean	StDev	SE Mean
Above To	15	4.92	0.157	0.042
Below To	15	4.74	0.320	0.084

95 % CI for mu Above To - mu Below To: (-0.013, 0.378)

T-Test mu Above To = mu Below To (vs not =): T = 1.95 P = 0.065 DF = 20



上游和下游样品数据盒形图

(均值以实心圆点标出)

a. 这些数据是否能充分证明上下游被溶解氧气的平均值存在差异?

$\alpha = 0.05$

b. 计算机输出结果所用的是混合  $t$  检验还是分离方差  $t$  检验?

c. (a)中检验所需的条件是否能满足? 为什么?

d. 上下游被溶解氧气的平均值差异有多大?

**6.6(工业)** 一个工程师猜测, 成功和不成功公司的一个很大的不同就在于它们所加工的产品因缺陷而被退回的百分比。为评价这个猜测, 这个工程师调查了 50 个成功和 50 个不成功的公司(以年利润为标准)。这些公司生产的产品性能和成本相似, 下面图表中列出了顾客退回产品的百分比。

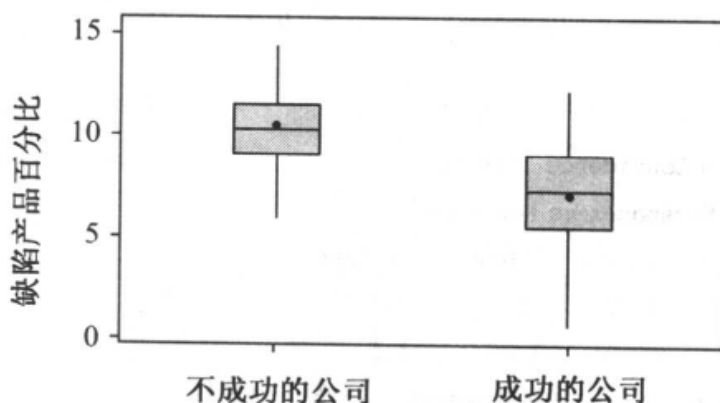
Two-Sample T-Test and Confidence Interval

Two-sample T for Unsuccessful vs Successful

	N	Mean	StDev	Se Mean
Unsuccessful	50	9.08	1.97	0.28
Successful	50	5.40	2.88	0.41

95 % CI for  $\mu$  Unsuccessful- $\mu$  Successful: (2.70, 4.66)

T-Test  $\mu$  Unsuccessful =  $\mu$  Successful (vs  $>$ ):  $T = 7.46$   $P = 0.0000$   $DF = 86$



成功和不成功的公司的盒形图(均值以实心圆点标出)

a. 这些数据是否能充分证明成功公司的产品退回率较低?  $\alpha = 0.05$

b. 计算机输出结果所用的是混合  $t$  检验还是分离方差  $t$  检验?

c. (a)中检验所需的条件是否能满足? 为什么?

d. 成功和不成功公司的产品退回率的差异有多大?

**6.7(社会)** 在过去的十年中, 看报纸家庭的数量有所减少。许多人称他们现在通过电视和网络获取信息。为了调查读报纸的人是否对时事较其他人有更多了解, 分别随机抽取 15 个订报纸的人和 15 个不订报纸的人, 对他们进行一个关于时事的测验。下面的茎叶图给出了每组的分数。问: 看报纸的人是否对时事有更多

的了解? 注意保证检验所需的条件都被满足。

#### Character. Stem-and-Leaf Display

Stem-and-leaf of No Newspaper Deliver

N = 30

Leaf Unit = 1.0

0 000

0

1 3

1 59

2 334

2 57

3 00234

3 5589

4 00124

4 5

5 0

5 55

6 2

Stem-and-leaf of Newspaper Subscribers

N = 25

Leaf Unit = 1.0

2 2

2 99

3 2

3 66889

4 000112333

4 55666

5 2

5 9

#### Two-Sample T-Test and Confidence Interval

Two-sample T for No Newspaper vs Newspaper

	N	Mean	StDev	SE Mean
No Newspaper	30	32.0	16.0	2.9
Newspaper	25	40.91	7.48	1.5

95 % CI for mu No Newspaper-mu Newspaper: (-15.5, -2.2)

T-Test mu No Newspaper = mu Newspaper (vs <): T = -2.70 P = 0.0049 DF = 42

**6.8(环境)** 为了确定一小时内汽车排放氮氧化物的平均排放量, 检测两种不同的排放控制装置。选择同年出厂、同型号的 20 辆车, 随机选出 10 辆配备第 1 种排放控制装置, 另外 10 辆配备第 2 种排放控制装置。对每一辆车进行 1 小时的监测, 测出氮氧化物排放量。

根据以下数据检验研究假设: 第一种装置的平均排放量  $\mu_1$  大于第二种装置的平均排放量  $\mu_2$ , 给定  $\alpha = 0.001$ 。

第一种装置		第二种装置	
1.35	1.28	1.01	0.96
1.16	1.21	0.98	0.99
1.23	1.25	0.95	0.98
1.20	1.17	1.02	1.01
1.32	1.19	1.05	1.02

**6.9(医学)** 据估计每年对一些物质(如颜料)的异常偏好(异食癖)导致的铅中毒可能会影响多达 25 万儿童的健康,使他们有严重的不能恢复的发育障碍。对儿童为何愿意食用铅的解释有“父母的不正确指导”“孩子口含东西的需要”等。一些研究者调查了是否这种情况可从营养学上进行解释。在一个研究中,对正常饮食的老鼠和饮食缺钙的老鼠所消化的醋酸铅溶液进行了对比。20 只老鼠被随机分到试验组或控制组。控制组正常饮食,试验组的食物中缺钙。每个老鼠被单独放在一个笼子中,对其在研究期间所食用的 1.5% 醋酸铅溶液进行监测。结果如下:

控制组	5.4	6.2	3.1	3.8	6.5	5.8	6.4	4.5	4.9	4.0
试验组	8.8	9.5	10.6	9.6	7.5	6.9	7.4	6.5	10.5	8.3

a. 将两样本的数据分别绘成图表。是否有理由认为  $t$  检验所需的条件不成立?

b. 检验研究假设:试验组食用的醋酸铅溶液的平均量大于控制组。给定  $\alpha = 0.05$ 。

**6.10(医学)** 一个为期三年的研究调查了几种速食早餐麦片对少儿龋齿的影响。Rowe, Anderson 和 Wanninger(1974)对这一研究结果进行了报道。来自 Michigan 州 Ann Arbor 公立学校的 375 个男孩和女孩(经父母允许)参加了这项研究。发给每个孩子牙刷和不同种类的麦片,尽管这些麦片品牌不同,但都被装在普通白色 7 盎司的盒子中,且都被标上了小麦、玉米、燕麦、水果味、玉米蓬松、可可、加糖燕麦片。注意后四种麦片事先都已加糖。

在开始研究之前,对每个孩子进行一次牙齿检查,研究期间两次,研究结束时一次。我们感兴趣的是变质的,缺失的,以及装填的(DMF)表面积增加量——即,研究前后 DMF 牙齿表面积的差。在 3 年中对每个孩子进行了认真的记录。在研

究的最后,若一个孩子所食用的麦片少于 28 盒,他就被归类为“非食用者”,否则被归为“食用者”。两组孩子增加的 DMF 表面积数据如下。用这些数据检验研究假设:非食用者的 DMF 面积大于相应的食用者的 DMF 面积。给出检验的显著性水平,并解释之。

	样本容量	样本均值	样本标准差
非食用者	73	6.41	5.62
食用者	302	5.20	4.67

**6.11** 参看习题 6.10。虽然原始的研究细节并没有被公布,请对以上的检验方法进行评价。

**6.12 (环境)** 对荒芜地带大气中微量金属的研究得到了极大的重视,因为有人担心改变大气中微量金属的含量和分布可能会改变地球的气候。考虑一项在南极进行的研究,在两个月中随机抽取 10 个时间段,测得 10,000 标准立方米(scm)空气中金属含量,得到镁和钨数据如下。(注:镁的单位为  $10^{-9}$ g/scm,钨的单位为  $10^{-15}$ g/scm)。注意:对于镁来说  $s > \bar{y}$ ,你预计这些数据服从正态分布吗?为什么?

	样本容量	样本均值	样本标准差
镁	10	1.0	2.21
钨	10	17.0	12.65

**6.13** 参看习题 6.12。你是否可以进行一个检验来比较镁和钨的平均含量?为什么?

**6.14(环境)** 从 1929 年开始 PCB 就已被应用,主要用于电力产业中,然而一直到 19 世纪 60 年代才知它是一个重要环境污染物,在论文“The ratio of DDE to PCB concentrations in Great Lakes herring gull eggs and its use in interpreting contaminants data”,[刊登在 *Journal of Great Lakes Research* 24(1):12~31]中,研究者报告了下面的研究情况。从五大湖地区选取了 13 个研究区域,在每一区域,连续几年内每年随机收集 9~13 个银鸥蛋,并检测 PCB 含量。1982 和 1996 年,每个区域的平均 PCB 含量见下表:

年份	地 点						
	1	2	3	4	5	6	7
1982	61.48	64.47	45.50	59.70	58.81	75.86	71.57
1996	13.99	18.26	11.28	10.02	21.00	17.36	28.20

年份	地 点						
	8	9	10	11	12	13	
1982	38.06	30.51	39.70	29.78	66.89	63.93	
1996	7.30	12.80	9.41	12.63	16.83	22.74	

- 在 20 世纪 70 年代通过了一项立法,限制 PCB 的生产和应用。因此,原 PCB 产地供应的 PCB 大大减少,这些数据是否可以充分地证明银鸥蛋的平均 PCB 含量显著减少?
- 用 95% 置信区间,估计从 1982 年到 1996 年平均 PCB 的减少量。
- 评价假设检验和构造置信区间的条件是否满足。
- 独立性条件被破坏了吗?

**6.15** 参看习题 6.14。可以看出 13 个区域的 PCB 含量差别很大。怎样才能减小地点不同对评价两年内 PCB 含量均值差的影响?

**6.16(人力资源)** 一家公司的低层管理人员年终奖金制度很复杂。关键是对“公司目标的贡献”的主观评定。一个人事部门的官员随机抽取 24 个女性和 36 个男性管理人员来判定奖金之间是否存在差异,用占年收入的百分比表示。数据如下:

性别	奖金百分比								
F	9.2	7.7	11.9	6.2	9.0	8.4	6.9	7.6	7.4
	8.0	9.9	6.7	8.4	9.3	9.1	8.7	9.2	9.1
	8.4	9.6	7.7	9.0	9.0	8.4			
M	10.4	8.9	11.7	12.0	8.7	9.4	9.8	9.0	9.2
	9.7	9.1	8.8	7.9	9.9	10.0	10.1	9.0	11.4
	8.7	9.6	9.2	9.7	8.9	9.2	9.4	9.7	8.9
	9.3	10.4	11.9	9.0	12.0	9.6	9.2	9.9	9.0

### Two-Sample T-Test and Confidence Interval

Two-sample T for Female vs Male

	N	Mean	StDev	SE Mean
Female	24	8.53	1.19	0.24
Male	36	9.68	1.00	0.17

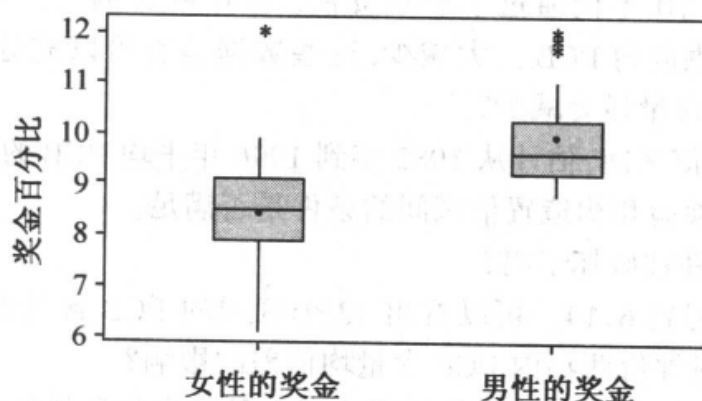
95 % CI for mu Female-mu Male: (-1.74, -0.56)

T-Test mu Female = mu Male (vs <): T = -3.90 P = 0.0002 DF = 43

95 % CI for mu Female-mu Male: (-1.72, -0.58)

T-Test mu Female = mu Male (vs <): T = -4.04 P = 0.0001 DF = 58

Both use Pooled StDev = 1.08



男性和女性管理人员奖金盒形图  
(均值以实心圆点表示)

- 求混合方差  $t$  统计量的值( $t$  检验一般建立在等方差性假定基础上)。
- 求  $t'$  统计量值。
- 分别在  $\alpha = 0.05$  和  $\alpha = 0.01$  水平下, 检验不等均值的研究假设, 结论与用那个统计量有关吗?

**6.17(商业)** 由于住院费用、医疗事故保险、医生收诊费和租金的不同, 各州之间外科手术费用差别很大。在加州和蒙大拿州进行了一个子宫切除手术费用的研究。从每个州随机抽取 200 个病人的病历作为样本, 以下是样本统计量的观察值:

	样本容量	样本均值	样本标准差
蒙大拿州	200	6,458 美元	250 美元
加州	200	12,690 美元	890 美元

- a. 是否有足够的理由认为加州子宫切除手术费用高于蒙大拿州?
- b. 用 95% 置信区间估计两州平均费用的差异。
- c. 在(a)中你是用混合  $t$  检验还是分离方差  $t$  检验? 为什么?

**6.18(教育)** 美国的一个教育机构对学生的阅读能力进行了测量, 分数范围为 0~500。每一组 500 名学生, 测量结果如下。根据这些数据进行推断。假定混合标准差为 100。

年龄	性别	样本均值*
9	男	210
	女	216
13	男	253
	女	262
17	男	283
	女	293

\* 等级: 150—初级阅读水平, 可达到最基本的要求。

200—基本水平, 可说出简单段落的内容。

250—中级水平, 可用长文章组织信息。

300—适应水平, 可理解和解释复杂的信息。

350—高级水平, 可理解和解释专业文献。

- a. 绘制一个图表来表示学生的年龄、性别和平均分数。
- b. 根据样本数据估计 17 岁男女生平均分数差异的 95% 置信区间。
- c. 用 90% 置信区间, 比较 13 和 17 岁女生的平均分数, 此区间包括 0 吗? 为什么两个均值是不同的?

### 6.3 非参数推断方法: Wilcoxon 秩和检验

上一节讲到的两样本  $t$  检验基于以下几个条件: 独立样本、正态性和方差相等。当方差相等和正态性两个条件不具备但样本容量很大时,  $t$  (或  $t'$ ) 检验结果基本上是正确的。但是另一个检验方法却不需要如此严格的条件。这个方法就是下面将要讨论的 Wilcoxon 秩和检验。

在这个检验中, 假定有两个样本独立, 分别取自于两个总体, 这两个总体的分布形状相同, 只是其中一个总体的分布曲线可能是由另一个总体的分布曲线向右



平移得到的,如图 6.7 所示。Wilcoxon 秩和检验并不要求总体服从正态分布。因此我们就从  $t$  检验所需的条件中剔除了一个。另外的两个条件,即同方差和独立性依然需要满足。因为在零假设下,两个总体的分布被假定是相同的,因此,如果零假设为真,从两个总体中随机抽取的独立样本也应是相似的。因为我们允许总体不服从正态分布,秩和检验法就必然涉及到极端观察值出现的可能性。一个办法就是在混合样本——即由来自两个总体的样本组成的样本数据中,用数据对应的秩(从最小到最大)替换数据本身。其中最小的数据秩为 1,最大数据秩为  $N = n_1 + n_2$ 。秩的大小与最小次小两值(或最大次大两值)之间的距离无关。这样,极端值对秩和统计量的影响就不像对  $t$  检验影响那么大。

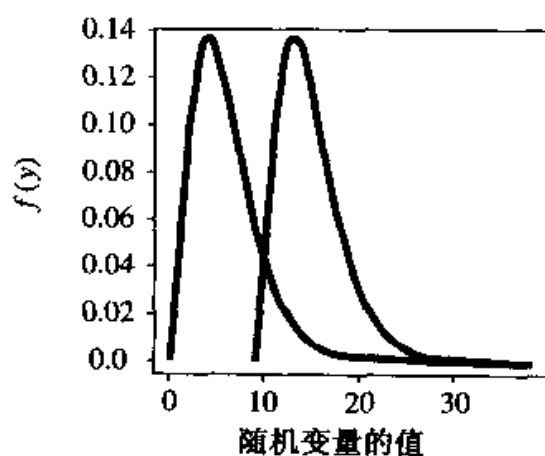


图 6.7 形状相同但位置不同的  
偏态总体分布

秩和检验统计量的计算包括以下几个步骤:

1. 将两个样本的数据一起从小到大进行排序。
2. 在下一列中,将数据从小到大进行编号,最小值对应于 1,最大值对应于  $N$ 。这些序号就是相应观察值的秩。
3. 如果在混合样本数据中有相同的值,那么它们的秩就是它们几个秩和的均值。
4. 令  $T$  表示总体 1 观察值的秩和。

如果零假设为真,即两总体同分布,那么,对应于总体 1 的  $n_1$  个秩就是  $N$  个整数  $1, 2, \dots, N$  的一个随机样本。因此在零假设下,秩和分布就由样本容量  $n_1$  和  $n_2$  惟一决定,而与总体的分布形状无关。在零假设下,  $T$  的抽样分布的均值和方差分别为

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \quad \sigma_T^2 = \frac{n_1 n_2}{12} (n_1 + n_2 + 1)$$

直观上, 如果  $T$  比  $\mu_T$  小(或大)很多时, 我们就有理由相信零假设不成立, 即两个总体分布是不同的。秩和检验的拒绝域具体给出了, 当零假设被拒绝时  $T$  和  $\mu_T$  差异的大小。由于在零假设条件下,  $T$  分布和总体分布无关, 因此, 表 5 给出的检验临界值与总体分布形状无关。现将 Wilcoxon 秩和检验步骤列出如下:

### Wilcoxon 秩和检验

$H_0$ : 两个总体分布相同。

$H_a$ : 1. 总体 1 是总体 2 的一个右平移。

2. 总体 1 是总体 2 的一个左平移。

3. 总体 1 和 2 互为平移。

( $n_1 \leq 10, n_2 \leq 10$ )

T. S.:  $T$ , 样本 1 的秩和

R. R.: 给定  $\alpha = 0.05$ , 查附表 5 得临界值  $T_U$  和  $T_L$ ;

1. 若  $T > T_U$ , 则拒绝  $H_0$ 。

2. 若  $T < T_L$ , 则拒绝  $H_0$ 。

3. 若  $T > T_U$  或  $T < T_L$ , 则拒绝  $H_0$ 。

检查假设, 并得出结论。

### 例 6.4

许多州都在考虑降低司机开车时所允许的血液中酒精含量的标准。一个立法机构的研究人员设计了以下检验来研究酒精对反应时间的影响。10 个参加者饮用指定量的酒精, 另外 10 个饮用同样多的不含酒精的饮料——一种安慰剂。两组成员都不知道他们所喝的是酒精还是安慰剂。下表列出了这 20 个人对一系列模拟驾驶场景的反应时间(以秒计)。能否看出酒精使反应时间延长了。

安慰剂	0.90	0.37	1.63	0.83	0.95	0.78	0.86	0.61	0.38	1.97
酒精	1.46	1.45	1.76	1.44	1.11	3.07	0.98	1.27	2.56	1.32

a. 在此研究中, 为什么  $t$  检验方法不适合于分析这些数据?

b. 用 Wilcoxon 秩和检验来检验以下假设:

$H_0$ : 对应于安慰剂和对应于酒精的两个反应时间总体分布相同。

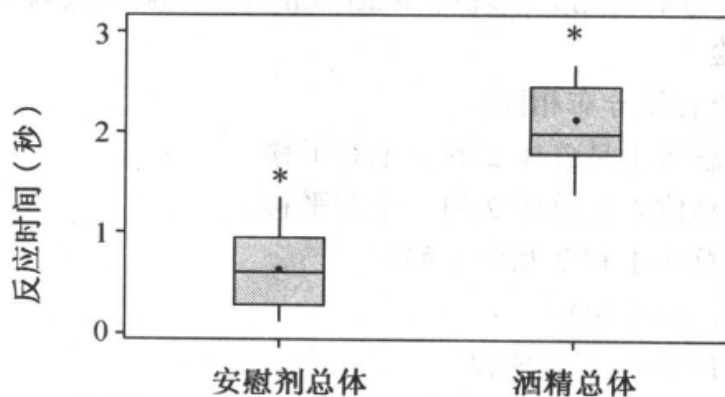
$H_a$ : 对应于安慰剂的反应时间总体分布是对应于酒精的反应时间总体分布的左平移(反应时间较长与饮用酒精有关。)

c. 对于两总体, 分别给出反应时间中位数的 95% 置信区间。

d. 将(b)的结果与 Minitab 的结果进行比较。

### 解答

a. 两样本的盒形图如下。此图表明两总体分布均右偏, 因为 10% 的数据是大的异常值且上须比下须长。样本容量都很小, 因此,  $t$  检验不适合。



安慰剂和酒精总体的盒形图  
(均值以实心圆点表示)

b. 下面将用 Wilcoxon 秩和检验来评价酒精是否延长了反应时间。表 6.6 给出了按顺序排列的混合样本数据以及相应的秩。我们令对应于安慰剂的总体为总体 1, 对应于酒精的总体为总体 2。

根据附表 5, 对于  $\alpha = 0.05$ , 单侧,  $n_1 = n_2 = 10$ , 若  $T < 83$ , 则拒绝  $H_0$ 。  $T$  的值为总体 1 样本的秩和, 即  $T = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 16 + 18 = 70$ 。因为 70 小于 83, 所以, 我们拒绝  $H_0$ , 并认为安慰剂总体的反应时间小于酒精总体。

表 6.6 反应时间排序和秩

顺序数据	组	秩	顺序数据	组	秩
1     0.37	1	1	11    1.27	2	11
2     0.38	1	2	12    1.32	2	12
3     0.61	1	3	13    1.44	2	13
4     0.78	1	4	14    1.45	2	14
5     0.83	1	5	15    1.46	2	15
6     0.86	1	6	16    1.63	1	16
7     0.90	1	7	17    1.76	2	17
8     0.95	1	8	18    1.97	1	18
9     0.98	2	9	19    2.56	2	19
10    1.11	2	10	20    3.07	2	20

c. 因为样本容量小且总体分布呈右偏态, 所以, 我们构造反应时间中位数的置信区间以代替均值的置信区间。利用第 5.8 节的方法, 并根据附表 4, 我们发现

$$C_{\alpha(2), n} = C_{0.05, 10} = 1$$

因此,

$$L_{0.025} = C_{0.50, 10} + 1 = 2$$

$$U_{0.025} = n - C_{0.50, 10} = 10 - 1 = 9$$

总体中位数的 95% 置信区间由以下公式给出

$$(M_L, M_U) = (y_{(2)}, y_{(9)})$$

因此, 安慰剂总体中位数的 95% 置信区间为 (0.38, 1.63), 酒精总体中位数的 95% 置信区间为 (1.11, 2.56)。由于样本容量很小, 因此, 这两个置信区间的意义不大。

d. Minitab 输出结果如下。

Mann-Whitney Confidence Interval and Test

PLACEBO N = 10 Median = 0.845

ALCOHOL N = 10 Median = 1.445

Point estimate for ETA1-ETA2 is -0.600

95.5 Percent CI for ETA1-ETA2 is (-1.080, -0.250)

W = 70.0

Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.0046

Minitab 中提到的是 Mann-Whitney 检验统计量。这个检验等同于 Wilcoxon 检验。事实上检验统计量的值  $W = 70$  与 Wilcoxon 检验统计量的值  $T = 70$  相同。输出结果给出了  $p$  值 = 0.0046, 两总体中位数之差的 95.5% 置信区间为 (-1.08, -0.25)。

当两个样本容量都大于 10 时,  $T$  的抽样分布近似于正态; 这就使得我们可以在 Wilcoxon 秩和检验中用  $z$  统计量代替  $T$ :

$$z = \frac{T - \mu_T}{\sigma_T}$$

理论上 Wilcoxon 秩和检验要求总体分布是连续的, 所以任意两个数值相等的概率为 0。因为在许多研究中, 我们只记录数据的几个小数位, 这样, 我们会常常遇上数据相等的情况, 即相等的观察值。在这种情况下, 每个相等数据的秩为这些相等数据秩的平均值。当有相等数据时, 必须调整  $T$  的方差。调整后的方差  $\sigma_T^2$  为

$$\sigma_T^2 = \frac{n_1 n_2}{12} \left[ (n_1 + n_2 + 1) - \frac{\sum_{j=1}^k t_j (t_j^2 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)} \right]$$

这里  $k$  是相等数据的组数,  $t_j$  是第  $j$  组相等观察值中数据的个数。注意, 当没有相

等数据时,对所有的  $j, t_j = 1$ , 这时

$$\sigma_T^2 = \frac{n_1 n_2}{12} (n_1 + n_2 + 1)$$

从实用的角度看,除非有许多相等数据,否则,调整对  $\sigma_T^2$  的影响不大。将正态近似的 Wilcoxon 秩和检验步骤总结如下:

#### Wilcoxon 秩和检验:正态近似

$n_1 \leq 10$  且  $n_2 \leq 10$

$H_0$ : 两个总体分布相同。

$H_a$ : 1. 总体 1 是总体 2 的一个右平移。

2. 总体 1 是总体 2 的一个左平移。

3. 总体 1 和 2 互为平移。

T. S.:  $z = \frac{T - \mu_T}{\sigma_T}$ , 其中  $T$  为样本 1 的秩和

R. R.: 给定  $\alpha$ ,

1. 若  $z \geq z_\alpha$ , 则拒绝  $H_0$ 。

2. 若  $z \leq -z_\alpha$ , 则拒绝  $H_0$ 。

3. 若  $|z| \geq z_{\alpha/2}$ , 则拒绝  $H_0$ 。

检查假设,并得出结论。

#### 例 6.5

环境工程师想要确定附近一个湖泊的清除工程是否有效。在工程开始前,他们从湖中随机抽取 12 个水样,并测出水样中溶解氧的含量(ppm),因为溶解氧的含量在夜间有所波动,故所有的测量均在下午 2:00 高峰期进行,清除工程前后的数据见表 6.7。

表 6.7 溶解氧的含量(ppm)

清除前		清除后	
11.0	11.6	10.2	10.8
11.2	11.7	10.3	10.8
11.2	11.8	10.4	10.9
11.2	11.9	10.6	11.1
11.4	11.9	10.6	11.1
11.5	12.1	10.7	11.3

a. 在  $\alpha = 0.05$  的水平下检验下列假设:

$H_0$ : 清除工程前和工程开始 6 个月后数据的分布相同。

$H_a$ : 清除工程前被溶解氧含量的分布是工程开始 6 个月后的一个右平移。

(注意从某种意义上讲, 如果氧气溶解量在一段时期内降低, 则说明清除工程有效。)

为了方便起见, 表 6.7 中的数据按从小到大列出。

b. 根据相等数据所做的修正有没有很大的意义?

### 解答

a. 首先, 我们混合 24 个样本观察值, 并赋予相应的秩, 最小观察值的秩为 1, 次小观察值的秩为 2, 依次类推。当两个或两个以上观察值相同时, 赋予它们平均秩。样本观察值和相应的秩见表 6.8。

表 6.8 被溶解氧的含量及相应的秩

清除前		清除后	
11.0	(10)	10.2	(1)
11.2	(14)	10.3	(2)
11.2	(14)	10.4	(3)
11.2	(14)	10.6	(4.5)
11.4	(17)	10.6	(4.5)
11.5	(18)	10.7	(6)
11.6	(19)	10.8	(7.5)
11.7	(20)	10.8	(7.5)
11.8	(21)	10.9	(9)
11.9	(22.5)	11.1	(11.5)
11.9	(22.5)	11.1	(11.5)
12.1	(24)	11.3	(16)
$T = 216$			

因为  $n_1$  和  $n_2$  的值都大于 10, 所以, 我们可以用检验统计量  $z$ 。如果想要检测出清理之后观察值的分布向左平移, 那么, 我们期望样本 1 的秩和就应较大。因此, 如果  $z = (T - \mu_T) / \sigma_T$  值较大, 就应拒绝  $H_0$ 。

按秩相等进行分组, 我们得到 18 个组。将这些组以及相应的每组等秩个数  $t_j$  列表如下。

秩	组	$t_j$	秩	组	$t_j$
1	1	1	14, 14, 14	10	3
2	2	1	16	11	1
3	3	1	17	12	1
4.5, 4.5	4	2	18	13	1
6	5	1	19	14	1
7.5, 7.5	6	2	20	15	1
9	7	1	21	16	1
10	8	1	22.5, 22.5	17	2
11.5, 11.5	9	2	24	18	1

对于所有  $t_j = 1$  的组, 因为  $t_j^2 - 1 = 0$ , 所以, 它们对  $\sigma_T^2$  中的

$$\frac{\sum_{j=1}^k t_j(t_j^2 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)}$$

没有贡献。因此, 我们仅需要  $t_j = 2, 3$ 。

将数据代入公式, 我们可得到

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{12(12 + 12 + 1)}{2} = 150$$

$$\sigma_T^2 = \frac{n_1 n_2}{12} \left[ (n_1 + n_2 + 1) - \frac{\sum_{j=1}^k t_j(t_j^2 - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)} \right]$$

$$= \frac{12(12)}{12} \left[ 25 - \frac{6 + 6 + 6 + 24 + 6}{24(23)} \right]$$

$$= 12(25 - 0.0870) = 298.956$$

$$\sigma_T = 17.29$$

$z$  的计算值为

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{216 - 150}{17.29} = 3.82$$

这个值大于 1.645, 所以我们拒绝  $H_0$ , 从而得出结论: 清除前的分布是清除后分布的一个右平移; 即, 清除后被溶解氧的含量趋向于小于清除前的含量。

b. 没有修正过的  $\sigma_T^2$  值是

$$\sigma_T^2 = \frac{12(12)(25)}{12} = 300 \quad \sigma_T = 17.32$$

对于这个  $\sigma_T$ ,  $z = 3.81$  而不是修正后得到的 3.82。这有助于我们理解除非相等数

据的个数很多,否则,修正对最终结果的影响是很小的。

Wilcoxon 秩和检验可代替两样本  $t$  检验,它比  $t$  检验所需的条件要少。特别地, Wilcoxon 秩和检验不要求总体服从正态分布,它只要求两总体分布形状相同,其中一个分布是另一个的平移。当两个分布都是正态时,  $t$  检验更能检测出存在的差异,即此时  $t$  检验的功效大于秩和检验。这是可以理解的,因为  $t$  检验用的是观察值本身的大小,而 Wilcoxon 秩和检验用的是观察值的相对大小(秩)。但是,当两总体不服从正态分布时, Wilcoxon 秩和检验的功效大于  $t$  检验的功效;即更可能检测出总体分布的平移。并且,对于所有分布类型的总体, Wilcoxon 秩和检验犯第 I 类错误的水平或概率均等于给定的值。当总体不服从正态分布时,  $t$  检验的实际水平与给定的水平将有所偏离。尤其当总体分布严重偏态或有极端离群值时,更是如此。

表 6.9  $\alpha = 0.05$  水平下,  $t$  检验( $t$ )和 Wilcoxon 秩和检验( $T$ )的功效

分布		正态			双指数			柯西			威布尔		
$n_1, n_2$	平移检验	0	0.6	1.2	0	0.6	1.2	0	0.6	1.2	0	0.6	1.2
5, 5	$t$	0.044	0.213	0.523	0.045	0.255	0.588	0.024	0.132	0.288	0.049	0.221	0.545
	$T$	0.046	0.208	0.503	0.049	0.269	0.589	0.051	0.218	0.408	0.049	0.219	0.537
5, 15	$t$	0.047	0.303	0.724	0.046	0.304	0.733	0.056	0.137	0.282	0.041	0.289	0.723
	$T$	0.048	0.287	0.694	0.047	0.351	0.768	0.046	0.284	0.576	0.049	0.290	0.688
15, 15	$t$	0.052	0.497	0.947	0.046	0.507	0.928	0.030	0.153	0.333	0.046	0.488	0.935
	$T$	0.054	0.479	0.933	0.046	0.594	0.962	0.046	0.484	0.839	0.046	0.488	0.927

Randles 和 Wolfe(1979)研究了偏态和厚尾分布对  $t$  检验和 Wilcoxon 秩和检验功效的影响。表 6.9 是他们模拟研究结果的一部分。针对每一分布类型的总体、每一样本容量和每一平移量,选取了 5000 个样本,记录在  $\alpha = 0.05$  水平下,  $t$  检验或 Wilcoxon 秩和检验拒绝  $H_0$  的次数所占的比例。考虑的分布包括正态分布,双指数分布(对称,厚尾),柯西(Cauchy)分布(对称,极端厚尾),威布尔(Weibull)分布(右偏态)。考虑的平移量有  $0, 0.6\sigma$  和  $1.2\sigma$ ,其中,除柯西分布外,  $\sigma$  表示总体标准差,而当总体服从柯西分布时,  $\sigma$  是一般的刻度参数。

当总体服从正态分布时,  $t$  检验只比 Wilcoxon 秩和检验稍好一点,即  $t$  检验有更高的功效。对双指数分布来说, Wilcoxon 秩和检验比  $t$  检验有更高的功效。对于柯西分布,  $t$  检验的水平显著偏离 0.05,它的功效要比 Wilcoxon 秩和检验低得多。当分布偏态时,如 Weibull 分布,两个检验有相似的表现;另外,  $t$  检验的水平和功效此时与分布为正态时的值几乎相同。除非有大量的异常值,否则,  $t$  检验对



偏态分布具有相当强的稳健性。

### 练习

### 应用

**6.19(商业)** 某管道工程承包人希望减少服务呼叫地点之间的平均距离,但仍然维持至少同样多的生意,这样他的工作效率就得到了提高。某管道工(管道工1)被指派给一个调度员,该调度员负责监听所来到的服务请求并制订一个当日的服务方案。管道工2仍同原来一样,基本按照呼叫顺序提供服务。两个管道工18天内(3个工作周)每天所走的英里数记录如下。

管道工1	88.2	94.7	101.8	102.6	89.3	95.7
	78.2	80.1	83.9	86.1	89.4	71.4
	92.4	85.3	87.5	94.6	92.7	84.6
管道工2	105.8	117.6	119.5	126.8	108.2	114.7
	90.2	95.6	110.1	115.3	109.6	112.4
	104.6	107.2	109.7	102.9	99.1	115.5

- 将每个管道工的样本数据绘制成图,并计算  $\bar{y}$  和  $s$ 。
- 在(a)的基础上,哪种方法更适合于比较这两个分布?

**6.20(医学)** 论文“Serum beta-2-microglobulin (SB2M) in patients with multiple myeloma treated with alpha interferon”[*Journal of Medicine* (1997) 28:311~318]报告了在治疗多骨髓瘤(MM)病人过程中  $\alpha$  干扰素治疗法的影响。指定新近被诊断为 MM 的 20 个病人为研究对象。研究人员将 20 个病人随机分成两组,对其中 10 位病人采用间歇地服用(左旋)苯丙氨酸氮芥(melphalan)和 sumiferon 治疗(处理组),对另外 10 位仅采用间歇地服用(左旋)苯丙氨酸氮芥治疗。分别在采用此疗法之前,之后第 3, 8, 15 天和第 1, 3, 6 个月测量 SB2M 的水平。用放射性免疫测定方法对 SB2M 值进行测定。治疗之前的测量值如下:

处理组	2.9	2.7	3.9	2.7	2.1	2.6	2.2	4.2	5.0	0.7
控制组	3.5	2.5	3.8	8.1	3.6	2.2	5.0	2.9	2.3	2.9

- 绘出两组样本数据的盒形图和正态概率图。
- 根据(a)的结果,哪种方法更适合于比较这两个 SB2M 分布?
- 是否有足够的理由认为两组 SB2M 分布有差异?

d. 讨论(c)的结果对评价  $\alpha$  干扰素的影响有何意义。

**6.21(农业)** 为比较吃两种不同维他命添加剂饲料的公鸡的鸡冠重量, 进行了一个试验。将 28 个健康的公鸡随机分成两个组。一组吃第 I 种饲料, 另一组吃第 II 种饲料。研究期过后, 每只公鸡的鸡冠重量(微克)如下:

饲料 I	73	130	115	144	127	126	112	76	68	101	126	49	110	123
饲料 II	80	72	73	60	55	74	67	89	75	66	93	75	68	76

a. 分别就对应于饲料 I 和饲料 II 的两组样本数据绘制成图, 并评价混合  $t$  检验的条件能否满足?

b. 给定  $\alpha = 0.05$ , 利用合适的  $t$  检验法确定两组鸡冠重量分布是否有差异?

c. 给定  $\alpha = 0.05$ , 用 Wilcoxon 秩和检验法检验两组鸡冠重量分布是否有差异?

d. 哪一种方法, Wilcoxon 或  $t$  检验法, 更适合于评价试验的结果? 并解释之。

**6.22(商业)** 参看练习 6.19。一年后又进行了一个研究。由 18 位管道工组成的处理组配备了一名调度员, 监听该组所有管道工所收到的服务请求并制订当天的服务计划。对照组的 18 位管道工同原来一样基本按照呼叫顺序提供服务。将 30 天内这 36 位管道工每人平均每日走的英里数记录如下:

处理组	62.2	79.3	83.2	82.2	84.1	89.3
	95.8	87.9	91.5	96.6	90.1	98.6
	85.2	87.9	86.7	99.7	101.1	88.6
对照组	87.1	70.2	94.6	182.9	85.6	89.5
	109.5	101.7	99.7	193.2	105.3	92.9
	63.9	88.2	99.1	95.1	92.4	87.3

a. 将样本数据绘制成图如下。根据这些图, 哪种检验方法更适合于比较这两总体的分布?

b.  $t$  检验和 Wilcoxon 秩和检验(等价于输出结果中的 Mann-Whitney 检验)计算机输出结果如下。试比较这两个检验结果, 并就该调度员计划的有效性作出结论。

c. 根据这些数据的图表和计算机输出结果, 评价  $t$  检验是否合适。

b. 不同检验的结果是否有所不同? 是不是在有些情况下应进行两种检验? 为什么?

Two-Sample T-Test and Confidence Interval

Two-sample T for Treatment vs Control

	N	Mean	StDev	SE Mean
Treatment	18	88.33	9.06	2.1
Control	18	102.1	33.2	7.8

95 % CI for  $\mu$  Treatment- $\mu$  Control: (- 30.3, 2.7)

T-Test  $\mu$  Treatment =  $\mu$  Control (vs<); T = - 1.70 P = 0.049 DF = 34

Both use Pooled StDev = 24.3

Two-Sample T-Test and Confidence Interval

Two-sample T for Treatment vs Control

	N	Mean	StDev	SE Mean
Treatment	18	88.33	9.06	2.1
Control	18	102.1	33.2	7.8

95 % CI for  $\mu$  Treatment- $\mu$  Control: (- 30.8, 3.2)

T-Test  $\mu$  Treatment =  $\mu$  Control (vs<); T = - 1.70 P = 0.053 DF = 19

Mann-Whitney Confidence Interval and Test

Treatment	N = 18	Median =	88.25
Control	N = 18	Median =	93.75

Point estimate for ETA1-ETA2 is - 5.20

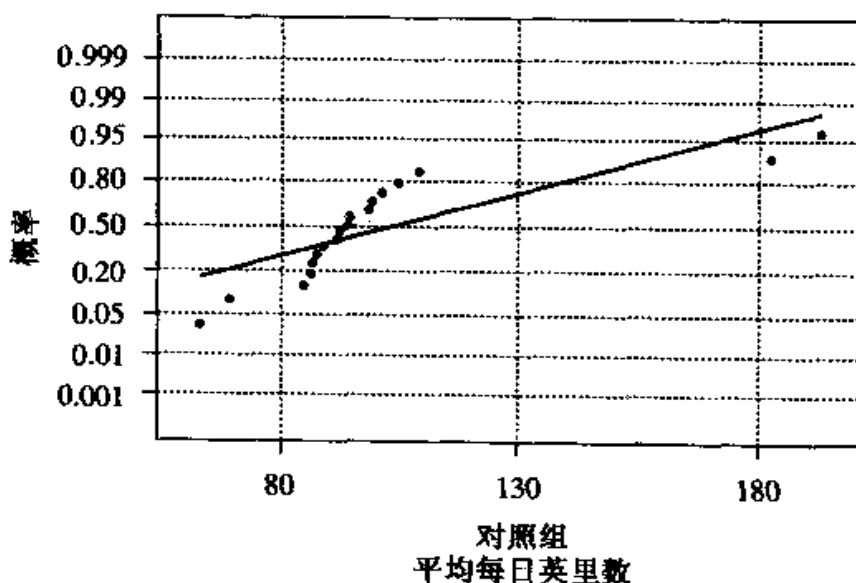
95.2 Percent CI for ETA1-ETA2 is (- 12.89, 0.81)

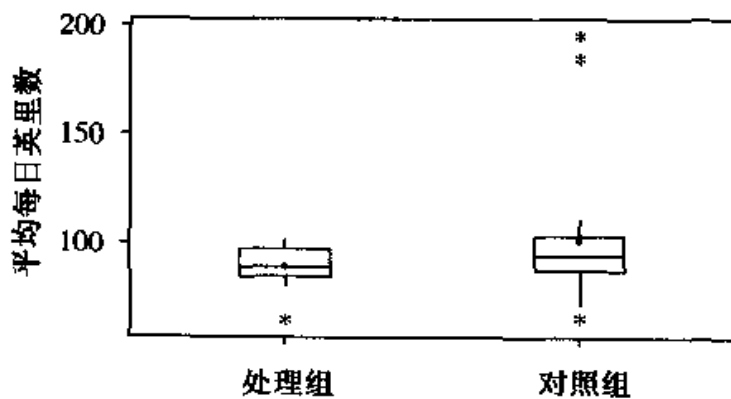
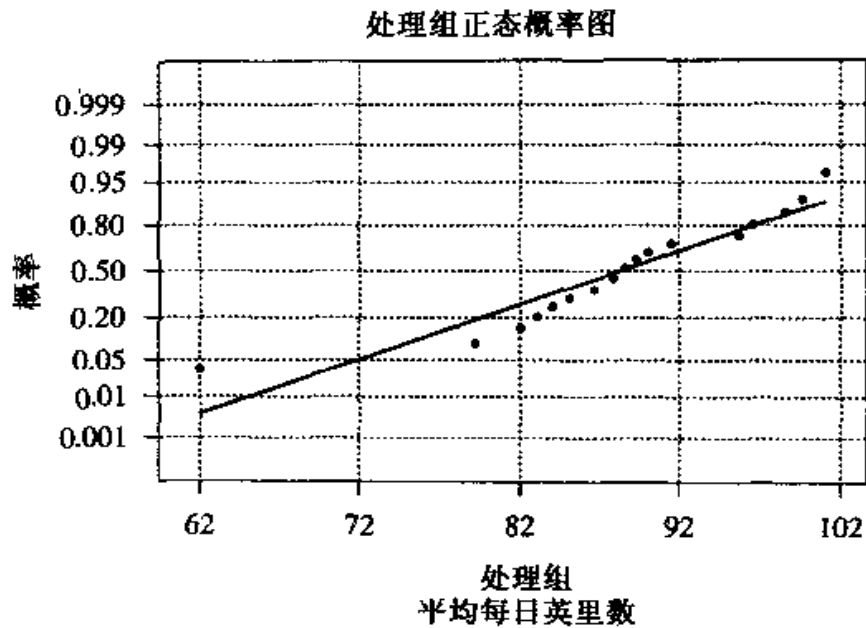
W = 278.5

Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.0438

The test is significant at 0.0438 (adjusted for ties)

对照组正态概率图





处理组和对照组盒形图(均值以实心圆点表示)

**6.23** 表 6.9 给出了厚尾分布和偏态分布影响  $t$  检验和 Wilcoxon 秩和检验效果的模拟研究结果。

- a. 如果有的话,对于哪种类型的分布, Wilcoxon 秩和检验犯第 I 类错误的概率远远偏离于给定的值  $\alpha = 0.05$ 。
- b. 如果有的话,对于哪种类型的分布,  $t$  检验犯第 I 类错误的概率远远偏离于给定的值  $\alpha = 0.05$ 。

**6.24** 参看习题 6.23。

- a. 比较三种非正态分布条件下和正态分布条件下  $t$  检验的功效。偏态或厚尾是不是似乎影响最大?
- b. 对于四种不同类型的分布, Wilcoxon 秩和检验的功效是不是差异很大? 并

解释之。

c. 你认为 Wilcoxon 秩和检验适用于哪种类型的分布？并解释之。

d. 你认为  $t$  检验适用于哪种类型的分布？并解释之。

## 6.4 关于 $\mu_1 - \mu_2$ 的推断：成对数据

前三节中我们给出的方法适用于来自于两总体的独立随机样本。然而，这些方法不适用于这样一种情形，其中一个样本的每个观测值均与另一样本中某一观测值相匹配或成对。本节，我们将讨论成对数据的分析方法。先看一个例子。

### 例 6.6

保险公司的评估员认为汽车修理厂 I 比汽车修理厂 II 对汽车修理费用的评估要高。为了证实他们的猜想，将最近出现事故的 15 辆汽车分别在两个汽车修理厂进行维修费用评估。所得数据如表 6.10。

表 6.10 维修费用评估(以百元计)

汽车	修理厂 I	修理厂 II
1	17.6	17.3
2	20.2	19.1
3	19.5	18.4
4	11.3	11.5
5	13.0	12.7
6	16.3	15.8
7	15.3	14.9
8	16.2	15.3
9	12.2	12.0
10	14.8	14.2
11	21.3	21.0
12	22.1	21.0
13	16.9	16.1
14	17.6	16.7
15	18.4	17.5
总和	$\bar{y}_1 = 16.85$	$\bar{y}_2 = 16.23$
	$s_1 = 3.20$	$s_2 = 2.94$

用两样本  $t$  检验对数据进行初步分析。

**解答** 计算机输出结果如下:

Two-Sample T-Test and Confidence Interval

Two-sample T for Garage I vs Garage II

	N	Mean	StDev	SE Mean
Garage I	15	16.85	3.20	0.83
Garage II	15	16.23	2.94	0.76

95 % CI for  $\mu$  Garage I -  $\mu$  Garage II: (-1.69, 2.92)

T-Test  $\mu$  Garage I =  $\mu$  Garage II (vs not =): T = 0.55 P = 0.59 DF = 27

从输出结果看,样本均值确实存在差异( $\bar{y}_1 - \bar{y}_2 = 0.62$ ),但是与这些观察值的变异性( $s_1 = 3.20, s_2 = 2.94$ )相比,这个差异很小。事实上,计算出的  $t$  值(0.55)具有  $p$  值 0.59,这表明很少有证据显示,平均来说两个修理厂的维修费用评估存在差异。

仔细观察这些数据,我们会发现例 6.10 所得出的结论与我们的直观感觉不相符。因为对于所有 15 辆车,除一辆车以外,修理厂 I 对事故车的维修费用评估高于修理厂 II。根据二项分布的知识,假定修理厂 I 和 II 对维修费用评估无差异( $\pi = 0.5$ ),那么在 15 对观察值中, I 有 14 个或 14 个以上大于 II 的概率是

$$\begin{aligned} P(y = 14 \text{ 或 } 15) &= P(y = 14) + P(y = 15) \\ &= \binom{15}{14} (0.5)^{14} (0.5) + \binom{15}{15} (0.5)^{15} = 0.000488 \end{aligned}$$

这样,如果两修理厂评估的维修费用具有相同的分布,那么有大约 5/10000 的机会, I 的 14 个或 14 个以上值大于 II 的相应值。根据此概率,我们认为观察到的评估值与两修理厂评估的维修费用具有相同分布这样的零假设非常矛盾。 $t$  检验和二项分布计算的结果为什么会发生如此冲突呢?

两种方法所得的结论之所以不同,主要在于  $t$  检验要求两样本互相独立,而在这一研究中恰好违反了这一独立性要求。评估员分别在两个修理厂得到关于一辆车的两个维修费用值。若要两样本独立,评估员需随机抽取 15 辆车由修理厂 I 给出维修费用,再抽取 15 辆车由修理厂 II 给出维修费用。

从图 6.8 可以看出,对于同一辆车,修理费的估价大体相等。但是对每个修理厂来说不同车的修理费用相差很大。这个大的 15 辆车之间的变异性使得两个修车厂之间的差异相对变小。在设计研究方案时,评估员发现汽车受损害的程度差异越大,每个修车厂的 15 个修理费用的差别也越大。因此他计算出每一辆车的修理费用的差值,这样便减少了车与车之间的差异。

这个例子说明了一个一般的设计原则。许多情况下,在随机指定给处理之前,

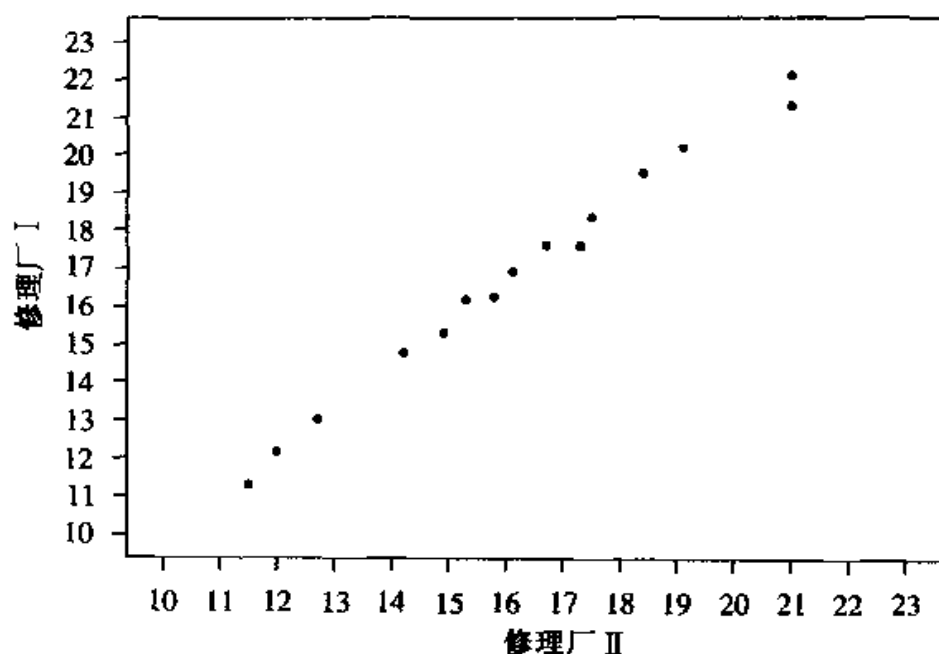


图 6.8 来自两修理厂的维修费用评估

试验单元之间在许多方面差别很大,而这些方面的差异又会影响到试验结果,从而会掩盖真正的处理所带来的差别。在前面的例子中,车辆在事故中的受损程度本来就差别很大,而不论由哪家修车厂评估其维修费用,所得估计都会有很大差异。当比较两个处理或两个组时,如果在将试验单元分配到两个处理或两个组之前,试验单元之间就有很大的差别,就应当把样本配对。有许多产生成对数据的试验设计方法。其中一种就是使同一组试验单元经过两种处理,如上例。第二种方法是观测试验单元在经过处理前后所得的数据。比如,假设我们要考察一种降压药品的效果,应记录参加者在服药前后的血压值。第三种方法是利用自然出现的成对事物,比如双胞胎或夫妻。最后一种方法是按照那些可能掩盖处理差别的因素,将试验单元排成对。比如,为比较两种阅读教学方法的效果,进行一个研究。在研究之前,应根据参加者现有的阅读能力将他们成对分组。分组后,两种方法随机地用于每一对的两个参加者之上。设计原则见第十四章。

对成对数据的合理分析需考虑两样本之间是否缺乏独立性。样本均值差  $(\bar{y}_1 - \bar{y}_2)$  的抽样分布的均值和标准差为

$$\mu_{\bar{y}_1 - \bar{y}_2} = \mu_1 - \mu_2 \quad \sigma_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho}{n}}$$

其中  $\rho$  表示两样本的相关程度。当配对的两样本观测值相似时,  $\rho$  为正,且  $\bar{y}_1 - \bar{y}_2$  的标准差比由两独立样本计算出的标准差要小,维修费用就是如此。通过考察成对数据的散点图可确定  $\rho$  的大小和符号。当散点图中的点接近一条直线时,  $\rho$

较大。当此直线呈上升趋势时,  $\rho$  为正。当此直线呈下降趋势时,  $\rho$  为负。从图 6.8 可以看出, 这些观察值的点接近一条上升的直线, 因此,  $\rho$  为正。在此修理费用研究中, 应用成对数据与应用独立样本相比可减少样本均值差的标准差。

在实际分析中, 我们需计算  $n$  对数据的差  $d_i = y_{1i} - y_{2i}$ , 得到  $d_i$  的均值  $\bar{d}$  和标准差  $s_d$ 。同时, 我们需要将关于  $\mu_1$  和  $\mu_2$  的假设转化为关于  $\mu_d = \mu_1 - \mu_2$  的假设。应用  $t$  方法来检验假设和构造  $\mu_d$  的置信区间需满足下列条件:

1.  $d_i$  的抽样分布是正态的。
2.  $d_i$  相互独立; 也就是说, 观察值对与对之间相互独立。

将检验步骤概括如下。

### 配对 $t$ 检验

$$H_0: 1. \mu_d \leq D_0. \quad (D_0 \text{ 为给定的值, 通常为 } 0)$$

$$2. \mu_d \geq D_0.$$

$$3. \mu_d = D_0.$$

$$H_a: 1. \mu_d > D_0.$$

$$2. \mu_d < D_0.$$

$$3. \mu_d \neq D_0.$$

$$T. S.: t = \frac{\bar{d} - D_0}{s_d / \sqrt{n}}$$

R. R.: 对于给定的水平, 即犯第 I 类错误概率  $\alpha$ , 及自由度  $df = n - 1$

1. 如果  $t \geq t_\alpha$ , 则拒绝  $H_0$ 。
2. 如果  $t \leq -t_\alpha$ , 则拒绝  $H_0$ 。
3. 如果  $|t| \geq t_{\alpha/2}$ , 则拒绝  $H_0$ 。

检查假定, 并得出结论。

基于成对数据的  $\mu_d = \mu_1 - \mu_2$  的  $100(1 - \alpha)\%$  置信区间为

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

这里,  $n$  是观察值的对数(即差值的个数),  $df = n - 1$ 。

### 例 6.7

对例 6.6 中的数据进行配对  $t$  检验。在  $\alpha = 0.05$  的水平下给出结论。

**解答** 对于这些数据统计检验的步骤如下

$$H_0: \mu_d = \mu_1 - \mu_2 \leq 0$$

$$H_a: \mu_d > 0$$



$$T. S. : t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

R. R. : 对于自由度  $df = n - 1$ , 如果  $t \geq t_{0.05}$ , 则拒绝  $H_0$ 。

在计算  $t$  之前, 必须先算出  $\bar{d}$  和  $s_d$ 。对于表 6.10 中的数据, 差值  $d_i =$  修理厂 I 的评估价 - 修车厂 II 的评估价。

汽车	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$d_i$	0.3	1.1	1.1	-0.2	0.3	0.5	0.4	0.9	0.2	0.6	0.3	1.1	0.8	0.9	0.9

均值和标准差为

$$\bar{d} = 0.61, s_d = 0.394$$

代入检验统计量  $t$  得

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{0.61}{0.394/\sqrt{15}} = 6.00$$

实际上, 对于  $df = 14$ ,  $t = 6.00$  远大于表中的  $t$  值, 所以  $p$  值小于 0.005; 事实上,  $p$  值等于 0.000016。于是, 我们可得出结论: 修理厂 I 的维修费用估价大于修理厂 II。这个结论与我们在二项分布基础上直观的感觉相一致。

作这些讨论不是为了说明我们通常会有两种或两种以上的分析, 得出相矛盾的结论。而是要强调分析必须适于具体的试验环境。对于这个试验, 两个样本不独立, 我们需要用一种适用于不独立(配对)数据的分析方法。

确定了均值在统计上具有显著差异之后, 我们应估计此差异的大小。 $\mu_1 - \mu_2 = \mu_d$  的 95% 置信区间为两修理厂维修费用估价的平均差异提供了一种估计:

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

$$0.61 \pm 2.145 \frac{0.394}{\sqrt{15}} \quad \text{或} \quad 0.61 \pm 0.22$$

这样, 我们有 95% 的把握相信, 两修理厂维修费用估价的平均差异在 390 美元到 830 美元之间。据此, 保险公司评估员可判定这一范围内的差异是否具有实际意义。

通过用差值  $d_i$  代替观察值  $y_{1i}$  和  $y_{2i}$ , 可减少  $\bar{y}_1 - \bar{y}_2$  的标准差, 从而使得新的  $t$  检验具有更高的功效和更窄的置信区间。利用配对数据试验时会产生损失吗? 当然, 用  $d_i$  的  $t$  检验, 自由度为  $df = n - 1$ , 而用单个观察值的  $t$  检验, 自由度为  $df = n_1 + n_2 - 2 = 2(n - 1)$ 。因此, 在设计研究或试验方案时, 究竟是用独立样本试验还是用配对数据试验, 要考虑试验单元之间在分配到处之前所存在差异的大小。如果差异很小, 那么, 设计独立样本试验更加有效。如果差异极大, 则配对数

据试验更有效。

练习

基本技能

### 6.25 考虑下列成对数据

对	$y_1$	$y_2$
1	21	29
2	28	30
3	17	21
4	24	25
5	27	33

- 进行配对  $t$  检验并给出检验的  $p$  值。
- 根据二项分布,你的结论会是什么?它与(a)中的结论一致吗?这两种方法在什么时候不一致?

应用

**6.26(工程)** 研究人员正在研究用于防止天然气输送管道腐蚀的两种保护膜,他们对至少在地下使用过5年的管道进行了检测。保护膜的有效性依赖于土壤的 pH 值,所以,研究人员在测量管道腐蚀之前,首先记录下埋藏管道的20处土壤的 pH 值, pH 值如下。试说明研究人员如何进行研究以减少土壤 pH 值之间的差异对比较两种保护膜性能的影响。

20 处研究地点的 pH 值										
保护膜 A	3.2	4.9	5.1	6.3	7.1	3.8	8.1	7.3	5.9	8.9
保护膜 B	3.7	8.2	7.4	5.8	8.8	3.4	4.7	5.3	6.8	7.2

**6.27(医学)** 假设你参加了一个治疗高胆固醇新方法的研究项目,为了将此新疗法与现有疗法进行比较,记录10周治疗期内的胆固醇水平的变化。治疗的效果可能与患者的年龄、体内脂肪百分比、饮食和健康情况有关。由于费用限制,此研究只能包括至多30位参加者。

- 说明用独立样本你如何进行研究?
- 说明用配对样本你如何进行研究?
- 你如何确定哪种方法,配对样本还是独立样本,对评价胆固醇变化更有效?

**6.28(农业)** 一个农业试验站想要比较两个新的玉米品种的产量。研究人员认为不同农场之间产量的差异可能会很大,故在 7 个农场的每一个农场都选择了两块 1 英亩的土地随机播种这两种玉米。成熟时收获玉米,试验结果(以蒲式耳为单位)如下表所示:

农场	1	2	3	4	5	6	7
品种 A	48.2	44.6	49.7	40.5	54.6	47.1	51.4
品种 B	41.5	40.1	44.0	41.2	49.8	41.7	46.8

- 根据这些数据检验两种玉米平均产量是否存在差异,  $\alpha = 0.05$ 。
- 估计两品种玉米平均产量差的大小。

**6.29(医学)** 论文“Effect of long-term blood pressure control on salt sensitivity”[*Journal of Medicine*(1997)28:147~156]的研究目的是评价经一段时期抗高血压治疗后,病人对盐的敏感性(SENS)。经过至少 18 个月抗高血压治疗后,对其 10 位高血压病人(舒张压在 90 至 115mmHg 之间)进行研究。记录治疗前后的 SENS 数据如下:

病人											
治疗前	22.86	7.74	15.49	9.97	1.44	9.39	11.40	1.86	-6.71	6.42	
治疗后	6.11	-4.02	8.04	3.29	-0.77	6.99	10.19	2.09	11.40	10.70	

- 是否有充分的理由证明经过治疗后,平均 SENS 值下降了?
- 估计平均 SENS 值的变化量。
- 对于这些数据,应用  $t$  检验的条件是否满足? 并解释之。

**6.30(人力资源)** 假设我们要估计男女销售代表平均月工资的差异。由于各公司间工资的差别很大,为消除这些差异的影响,仅对同一公司的男女代表进行对比。依据背景和工作经验,从每个公司中选择男女代表各一名。如果同一公司男女代表月工资的差异在 300 美元之内,为了估计男女平均月工资差异,试确定此研究所需公司的数量,使得 95% 置信区间宽度的一半为 5 美元。(提示:参考第 5.3 节)

**6.31** 参看练习 6.30。若  $n = 35$ ,  $\bar{d} = 120$ ,  $s_d = 250$ , 构造男女代表平均月工资差异  $\mu_d$  的 90% 置信区间。

**6.32(教育)** 一项研究旨在衡量家庭环境对 12 岁儿童学习成绩的影响。由于遗传因素可能也会影响学习成绩,因此,研究人员要对这一因素进行控制。找到

了 30 对同卵双胞胎,他(她)们在 1 岁生日之前就已被分别收养。其中一个双胞胎所在的家庭重视学术,另一个双胞胎所在的家庭则不重视学术。这 60 名学生的期末成绩(百分制)如下。

双胞胎	重视学术	不重视学术	双胞胎	重视学术	不重视学术
1	78	71	16	90	88
2	75	70	17	89	80
3	68	66	18	73	65
4	92	85	19	61	60
5	55	60	20	76	74
6	74	72	21	81	76
7	65	57	22	89	78
8	80	75	23	82	78
9	98	92	24	70	62
10	52	56	25	68	73
11	67	63	26	74	73
12	55	52	27	85	75
13	49	48	28	97	88
14	66	67	29	95	94
15	75	70	30	78	75

- 根据计算机输出结果,评价两种家庭孩子的平均成绩是否存在差异。
- 估计学术家庭和非学术家庭孩子平均成绩差异的大小。
- 就这些数据而言, $t$  检验的条件是否满足?
- 在研究中,为了控制学习成绩的变化,用双胞胎做样本是否比从两种家庭中随机抽取 30 个孩子作为样本更为有效? 并解释之。

Paired T-Test and Confidence Interval

Paired T for Academic-Nonacademic

	N	Mean	StDev	SE Mean
Academic	30	75.23	13.29	2.43
Nonacademic	30	71.43	11.42	2.09
Difference	30	3.800	4.205	0.768

95 % CI for mean difference: (2.230, 5.370)

T-Test of mean difference = 0 (vs not = 0): T-value = 4.95 P-Value = 0.000

Two-Sample T-Test and Confidence Interval

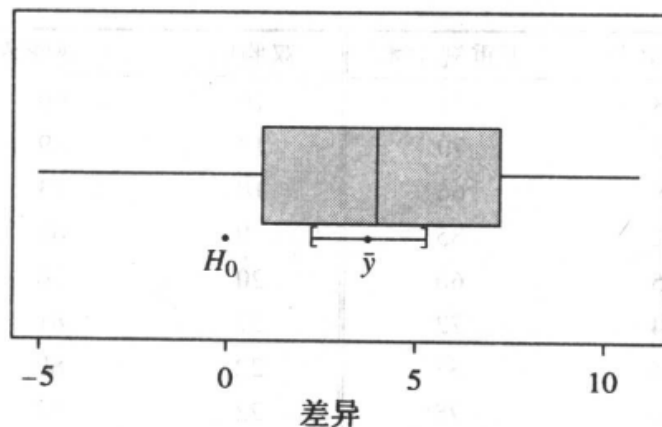
Two-sample T for Academic vs Nonacademic

	N	Mean	StDev	SE Mean
Academic	30	75.2	13.3	2.4

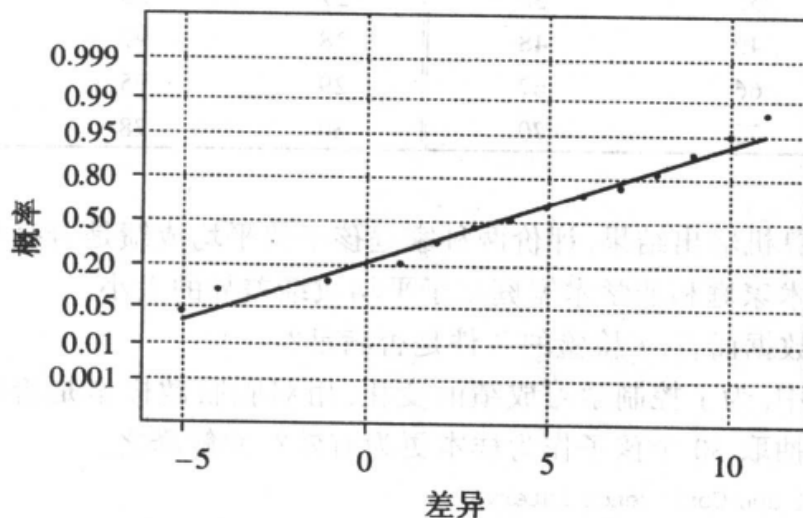
Nonacademic 30 71.4 11.4 2.1

95% CI for  $\mu$  Academic- $\mu$  Nonacademic: (-2.6, 10.2)

T-Test  $\mu$  Academic =  $\mu$  Nonacademic (vs not =): T = 1.19 P = 0.24 DF = 56



差异的盒形图( $H_0$  和均值的 95%  $t$  置信区间)



差异的正态概率图

## 6.5 非参数推断方法: Wilcoxon 符号秩检验

Wilcoxon 符号秩检验利用配对观察数据差值秩的符号和大小, 提供了另外一种当差值分布为非正态时的配对  $t$  检验。Wilcoxon 符号秩检验要求差值的总体分布关于一个未知中位数  $M$  对称。令  $D_0$  为  $M$  的一个给定的假设值。此检验是为了评价差值分布的中心是在  $D_0$  的左侧或右侧; 大多数情况下,  $D_0$  为 0。符号秩

检验的步骤如下:

1. 计算  $n$  对观察值的差。
2. 将这些差值减去  $D_0$ 。
3. 去掉所有 0 值, 令  $n$  表示非零值的个数。
4. 按从小到大的顺序排列这些差值的绝对值, 将它们赋予秩  $1, \dots, n$  (若绝对值相等, 则取平均秩)。

在叙述 Wilcoxon 符号秩检验之前, 先定义以下符号:

$n$  = 差值不为 0 的观察数据对数

$T_+$  = 正数秩的和; 如果没有正数秩, 则  $T_+ = 0$

$T_-$  = 负数秩的和, 如果没有负数秩, 则  $T_- = 0$

$T = T_+$  和  $T_-$  中较小者

$$\mu_T = \frac{n(n+1)}{4}$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

如果我们将具有相同秩的差值分成一组, 共有  $g$  组, 则  $T$  的方差为

$$\sigma_T^2 = \frac{1}{24} \left[ n(n+1)(2n+1) - \frac{1}{2} \sum_j t_j(t_j-1)(t_j+1) \right]$$

这里  $t_j$  为第  $j$  组中相等秩的个数。注意如果没有相等秩, 则  $g = n$ , 并且, 对所有组来说, 均有  $t_j = 1$ 。于是, 公式可简化为

$$\sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$$

Wilcoxon 符号秩检验步骤如下。

#### Wilcoxon 符号秩检验

$H_0$ : 差值的分布关于  $D_0$  对称。( $D_0$  给定, 通常情况下为 0。)

$H_a$ : 1. 差值大于  $D_0$ 。

2. 差值小于  $D_0$ 。

3. 1 或 2 为真(双侧  $H_a$ )。

( $n \leq 50$ )

T. S.: 1.  $T = T_-$ 。

2.  $T = T_+$ 。

3.  $T = T_+$  和  $T_-$  中较小者。

R. R.: 对于给定的  $\alpha$  (单尾 0.05, 0.025 或 0.005; 双尾 0.10, 0.05, 0.02, 0.01) 和非零差值个数  $n$ , 如果  $T$  小于或等于附表 6 中相应的值, 则拒绝  $H_0$

( $n > 50$ )

T. S.: 计算检验统计量

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

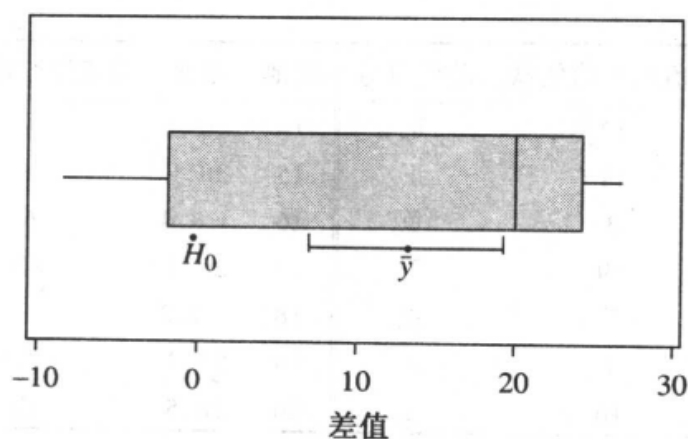
R. R.: 对于情形 1 和情形 2, 如果  $z < -z_\alpha$ , 则拒绝  $H_0$ ; 对于情形 3, 如果  $z < -z_{\alpha/2}$ , 则拒绝  $H_0$   
检查假定, 并得出结论。

### 例 6.8

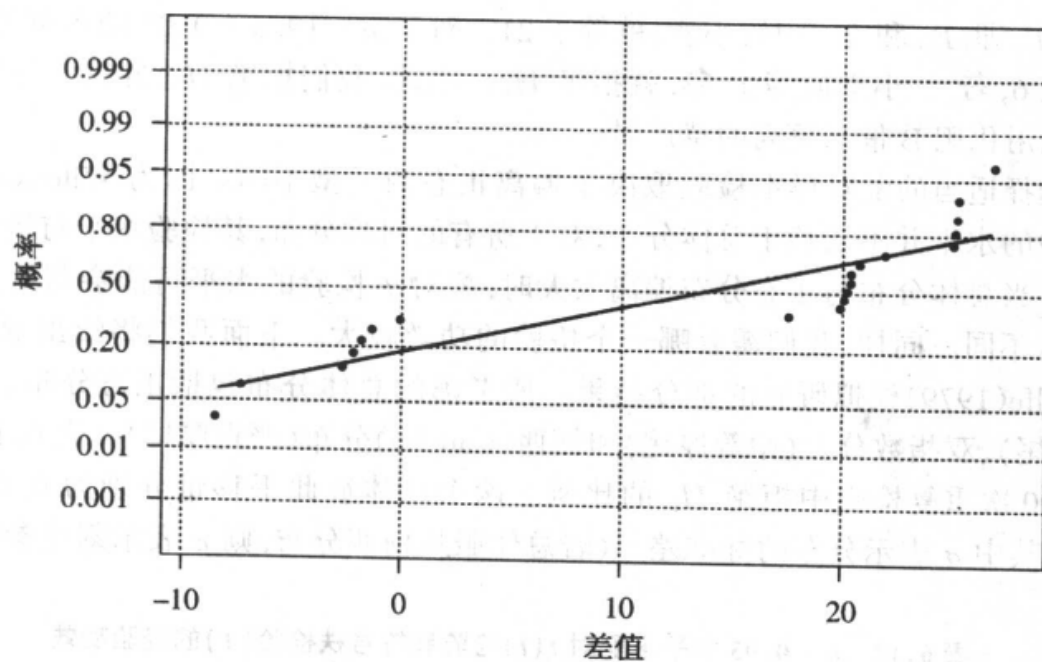
某城市公园管理部门在 20 个不同全球场地比较一种新配方的化肥 A 与以前使用的化肥 B。每个场地分为两半, 随机地选出一半施用化肥 A, 另一半施用化肥 B。每英亩场地施 60 磅化肥。用 1 个月后割下的草的重量(磅)来衡量两种化肥对草皮生长的影响。试评价施用化肥 A 是否比施用化肥 B 产的草更多。数据列出如下。

场地	A 区	B 区	差值	场地	A 区	B 区	差值
1	211.4	186.3	25.1	11	208.9	183.6	25.3
2	204.4	205.7	-1.3	12	208.7	188.7	20.0
3	202.0	184.4	17.6	13	213.8	188.6	25.2
4	201.9	203.6	-1.7	14	201.6	204.2	-2.6
5	202.4	180.4	22.0	15	201.8	181.6	20.1
6	202.0	202.0	0	16	200.3	208.7	-8.4
7	202.4	181.5	20.9	17	201.8	181.5	20.3
8	207.1	186.6	20.4	18	201.5	208.7	-7.2
9	203.6	205.7	-2.1	19	212.1	186.8	25.3
10	216.0	189.1	26.9	20	203.4	182.9	20.5

**解答** 要评价施用化肥 A 是否比施用化肥 B 产更多的草。将 20 个场地内青草产量差值数据绘制成图 6.9(a)和(b)。可以看出, 这些差值似乎并不服从正态分布, 而且形成了两个不同的群。因此, 我们用 Wilcoxon 符号秩检验评价因施化肥 A 和化肥 B 所带来的草产量的差异。零假设为差值的分布关于 0 对称, 备择假设为差值大于 0, 首先, 我们必须将  $n = 20 - 1 = 19$  个差值的绝对值从小到大进行排序。这些秩见表 6.11。



(a) 差值的盒形图  
( $H_0$  和均值的 95%  $t$  置信区间)



(b) 差值的正态概率图

图 6.9

表 6.11 草产量数据的秩

场地	差值	差值绝对值的秩	差值符号	场地	差值	差值绝对值的秩	差值符号
1	25.1	15	正	4	-1.7	2	负
2	-1.3	1	负	5	22.0	8	正
3	17.6	7	正	6	0	无	正



续表

场地	差值	差值绝对值的秩	差值符号	场地	差值	差值绝对值的秩	差值符号
7	20.9	13	正	14	-2.6	4	负
8	20.4	11	正	15	20.1	9	正
9	-2.1	3	负	16	-8.4	6	负
10	26.9	19	正	17	20.3	10	正
11	25.3	17.5	正	18	-7.2	5	负
12	-1.3	1	正	19	25.3	17.5	正
13	25.2	16	正	20	20.5	12	正

正秩与负秩之和分别为

$$T_- = 1 + 2 + 3 + 4 + 5 + 6 = 21$$

$$T_+ = 7 + 8 + 9 + 10 + 11 + 12 + 13 + 14 + 15 + 16 + 17.5 + 17.5 + 19 = 169$$

这样,  $T$ , 即  $T_+$  和  $T_-$  中较小者, 就等于 21。对于  $n=19$ ,  $\alpha=0.05$  的单侧检验, 根据附表 6, 若  $T$  小于或等于 53, 就拒绝  $H_0$ 。因此, 我们拒绝  $H_0$ , 并认为施用化肥 A 比施用化肥 B 带来更高的草产量。

选择适当的配对样本检验取决于偏离正态的类型不同。因为 Wilcoxon 符号秩检验的水平并不依赖于总体分布, 对于所有的对称分布, 其检验水平与给定的值相同。当总体分布与正态分布偏离太大时, 配对  $t$  检验的水平可能会与给定的水平有所不同。同样, 我们需看哪一个检验的功效较大。下面我们将给出 Rangdles 和 Wolfe(1979)模拟研究的部分结果。所考虑的总体分布包括正态分布、均匀分布(短尾)、双指数分布(中等厚尾)和柯西(Cauchy)分布(严重厚尾)。表 6.12 列出了 5000 次重复检验中拒绝  $H_0$  的比例。两个总体彼此平移量分别为  $0, 0.4\sigma$  和  $0.8\sigma$ , 其中  $\sigma$  表示分布的标准差。(若总体服从柯西分布, 则  $\sigma$  表示刻度参数。)

表 6.12  $\alpha=0.05$  水平下配对  $t(t)$  检验和符号秩检验( $T$ )的经验功效

分布	平移	正态			双指数			柯西			均匀		
		0	$0.4\sigma$	$0.8\sigma$	0	$0.4\sigma$	$0.8\sigma$	0	$0.4\sigma$	$0.8\sigma$	0	$0.4\sigma$	$0.8\sigma$
$n=10$	$t$	0.049	0.330	0.758	0.047	0.374	0.781	0.028	0.197	0.414	0.051	0.294	0.746
	$T$	0.050	0.315	0.741	0.048	0.412	0.804	0.049	0.332	0.623	0.049	0.277	0.681
$n=15$	$t$	0.048	0.424	0.906	0.049	0.473	0.898	0.025	0.210	0.418	0.051	0.408	0.914
	$T$	0.047	0.418	0.893	0.050	0.532	0.926	0.050	0.423	0.750	0.051	0.383	0.852
$n=20$	$t$	0.048	0.546	0.967	0.044	0.571	0.955	0.026	0.214	0.433	0.049	0.522	0.971
	$T$	0.049	0.531	0.962	0.049	0.652	0.975	0.049	0.514	0.849	0.050	0.479	0.935

据表 6.12, 我们可以得出以下结论: 对于均匀分布和双指数分布, 配对  $t$  检验的水平仍然接近 0.05, 但是对于严重厚尾的柯西分布, 检验水平比 0.05 小得多。正如我们所预期的, 对于所有四种分布, Wilcoxon 符号秩检验的水平几乎都是 0.05, 这是因为 Wilcoxon 检验只要求总体的分布是对称的。当分布为正态时,  $t$  检验功效只比 Wilcoxon 符号秩检验稍高一点。当总体分布为短尾和均匀分布时, 配对  $t$  检验比 Wilcoxon 符号秩检验的功效稍高一点; 此时, 我们还看到,  $t$  检验的功效比起当总体服从正态分布时的  $t$  功效值来稍微大一些。对于双指数分布, Wilcoxon 检验的功效略高于  $t$  检验。对于柯西分布,  $t$  检验的水平明显偏离 0.05, 其功效也比 Wilcoxon 检验功效低。根据其他研究, 如果差值分布严重偏态时, 名义上的  $t$  概率可能令人误解。偏态对 Wilcoxon 检验的水平影响不大。

虽然经过上述讨论, 你也许对在特定情况下, 究竟采用哪种统计检验和置信区间仍感到迷惑。首先, 把数据绘制成图, 以确定总体分布是严重厚尾还是严重偏态。在这种情况下, 建议采用基于秩的 Wilcoxon 检验。如果从图中难以看出它们的非正态性, 建议两种检验都试一下。如果不同的检验得到不同的结果, 那么认真检查这些数据, 找出任何特殊之处, 想想为什么会有不同的结果。如果两个结论一致, 并且所需条件没有明显地被破坏, 你就应对你的结论非常自信。这个特殊的“规避”策略不仅适用于成对数据而且适用于很多有若干种分析方法的情形。

## 练习

### 基本技能

**6.33** 题目参见 6.29。

- 根据表中的数据, 进行 Wilcoxon 符号秩检验。求  $p$  值并给出结论。
- 将此处的结论与练习 6.29 的结论进行比较。两种检验法( $t$  检验和符号秩检验)有区别吗?

## 应用

**6.34** 根据表 6.12 中的数据回答下列问题。

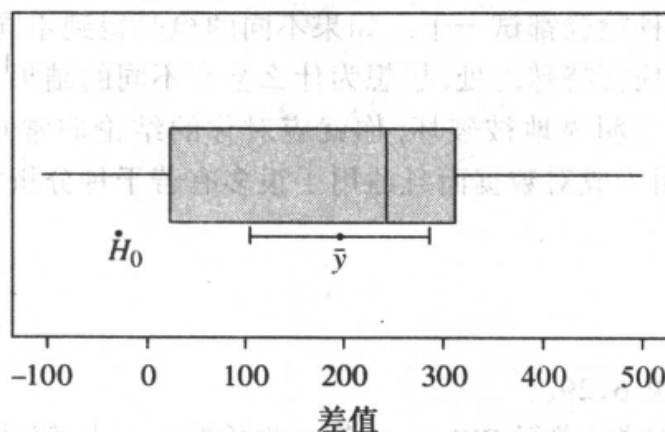
- 样本容量  $n$  对  $t$  检验的真实水平与给定值 0.05 的接近程度有影响吗?
- 总体分布形状对  $t$  检验的真实水平与给定值 0.05 的接近程度有影响吗?
- 总体分布形状对 Wilcoxon 符号秩检验的真实水平与给定值 0.05 的接近程度有影响吗?
- 假设对一个严重右偏的数据进行水平为 0.05 的检验, Wilcoxon 符号秩和配对  $t$  检验, 哪一个的实际水平更接近 0.05? 并解释之。

**6.35(社会)** 进行一项研究旨在确定女性顾客的汽车修理费是否大于男性顾客。从电话簿中随机选取 10 家汽车修理铺。试验中所用的两辆车同车龄, 同品

牌,发动机故障也相同。对于每家修理铺,将这两辆车随机分给男女两个参加者,由他们将车带至修理厂进行修理费用评估。修理费用(单位:美元)如下。

修理铺	1	2	3	4	5	6	7	8	9	10
女性顾客	871	684	795	838	1,033	917	1,047	723	1,179	707
男性顾客	792	765	511	520	618	447	548	720	899	788
修理铺	11	12	13	14	15	16	17	18	19	20
女性顾客	817	846	975	868	1,323	791	1,157	932	1,089	770
男性顾客	927	657	851	702	918	528	884	702	839	878

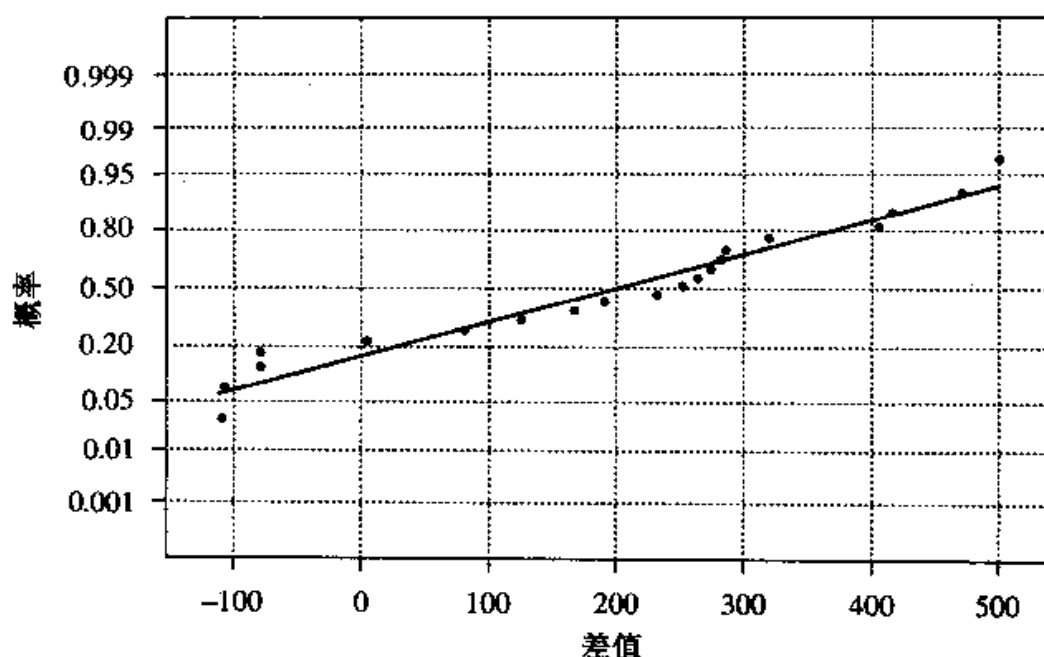
- $t$  检验或 Wilcoxon 检验,哪种检验更适合于本例?为什么?
- 总的来看,女性顾客的修理费用是否比男性顾客高?  $\alpha = 0.05$



差值的盒形图(对于  $H_0$  和均值的 95%  $t$  置信区间)

**6.36(生物)** 选择 14 只狗进行一项试验,以检测苯丙胺对狗的心率(每分钟心跳数)的影响。每只狗以其自身作为对照。其中 7 只在第一个试验期内服用苯丙胺,另 7 只服用安慰剂(盐溶液)。服用两小时后测量每只狗的心率。此后两周,这些狗不再服用苯丙胺或安慰剂。之后,为第二试验期。将两组狗所服之物互换,之前服用苯丙胺的现服用安慰剂,原服用安慰剂的现服用苯丙胺。两小时后记录心率。

以下样本数据不是按得到数据的顺序,而是按服用药物的种类列出。根据这些数据检验研究假设:服用苯丙胺组狗的心率分布是服用安慰剂组狗的心率分布的右平移。给定  $\alpha = 0.05$ ,进行单侧 Wilcoxon 符号秩检验。



费用差值的正态概率图

狗	安慰剂组	苯丙胺组	狗	安慰剂组	苯丙胺组
1	250	258	8	296	305
2	271	285	9	301	319
3	243	245	10	298	308
4	252	250	11	310	320
5	266	268	12	286	293
6	272	278	13	306	305
7	293	280	14	309	313

## 6.6 推断 $\mu_1 - \mu_2$ 时样本容量的选取

第 5.3 和第 5.5 节讨论了如何选取样本容量使得置信区间具有事先给定的宽度和置信水平, 或者使得关于  $\mu$  的统计检验具有预先给定的水平  $\alpha$  和  $\beta$ 。在独立样本或成对情况下, 也可类似地选择样本容量对  $\mu_1 - \mu_2$  进行统计推断。通过解下列等式可得到样本容量  $n$  使得基于独立样本条件下  $\mu_1 - \mu_2$  的  $100(1 - \alpha)\%$  置信区间的宽度为  $2E$ 。这里, 我们假定两样本容量相等。

$$z_{\alpha/2}\sigma \sqrt{\frac{1}{n} + \frac{1}{n}} = E$$

注意, 在这个公式中  $\sigma$  是总体的共同标准差, 并且假定了样本容量相等。

独立样本下形如  $\bar{y}_1 - \bar{y}_2 \pm E$  的  $\mu_1 - \mu_2$  的  $100(1 - \alpha)\%$  置信区间所需要的样本容量

$$n = \frac{2z_{\alpha/2}^2 \sigma^2}{E^2}$$

(注意: 如果  $\sigma$  未知, 则用其估计值来代替以得到近似的样本容量。)

用这个公式计算出的样本容量一般都是近似的, 因为公式中总体的共同标准差  $\sigma$  需用其估计值来代替。这个估计值可能是根据以前研究资料或根据总体值范围的经验得到的。

利用下面的公式可计算给定  $\alpha$  和  $\beta$  时  $\mu_1 - \mu_2$  的单侧和双侧检验所需的样本容量。其中, 当  $|\mu_1 - \mu_2| \geq \Delta$  时, 我们期望水平  $\alpha$  的检验犯第 II 类错误的概率  $\beta(\mu_1 - \mu_2) \leq \beta$ 。

独立样本下  $\mu_1 - \mu_2$  的假设检验所需要的样本容量

$$\text{单侧检验: } n = 2\sigma^2 \frac{(z_{\alpha} + z_{\beta})^2}{\Delta^2}$$

$$\text{双侧检验: } n = 2\sigma^2 \frac{(z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$$

这里,  $n_1 = n_2 = n$ , 且当差的真值  $|\mu_1 - \mu_2| \geq \Delta$  时, 犯第二类错误的概率  $\leq \beta$ 。(注意: 如果  $\sigma$  未知, 则用其估计值来代替以得到近似的样本容量。)

### 例 6.9

进行一项试验旨在确定饲料中添加液体乳清对奶牛的影响。虽然吃标准饲料(7.5 公斤谷物外加草料, 草料的量为可选)加清水的奶牛与吃标准饲料加乳清的奶牛在产奶量上并无明显差异, 但他们食用的干草量却多少不同。我们检验零假设: 两组奶牛平均食用干草量并无差异。对于  $\alpha = 0.05$  的双尾检验, 若当  $|\mu_1 - \mu_2| \geq 0.5$  时, 我们要求  $\beta \leq 0.10$ , 求每组所需奶牛大约多少头? 根据以前试验结果,  $\sigma$  约等于 0.8。

**解答** 根据题目, 对于  $\Delta = |\mu_1 - \mu_2| \geq 0.5$ ,  $\sigma = 0.8$ , 我们要求  $\alpha = 0.05$ ,  $\beta \leq 0.10$ , 附表 1 给出  $z_{0.025} = 1.96$ ,  $z_{0.10} = 1.28$ 。代入公式可得:

$$n \approx \frac{2(0.8)(1.96 + 1.28)^2}{(0.5)^2} = 53.75, \text{ 或 } 54$$

这样, 为得到我们期望的检验, 每组需要 54 头奶牛。

当样本容量不相等, 即  $n_1 \neq n_2$  时, 也可计算样本容量。令  $n_2$  是  $n_1$  的  $m$  倍, 即,  $n_2 = mn_1$ 。例如, 如果我们希望  $n_1$  是  $n_2$  的 3 倍, 那么,  $n_2 = \frac{1}{3}n_1$ 。以上公式

仍适用,只是将倍数 2 用  $(m+1)/m$  代替,将  $n$  用  $n_1$  代替。解出  $n_1$  后,可求出  $n_2 = mn_1$ 。

### 例 6.10

题目见例 6.8。假定试验人员希望得到饲料加乳清组(第 II 组)的信息比饲料加水组(第 I 组)更为详细。具体地,试验人员希望第 II 组的奶牛数比第 I 组多 40%,即,  $n_2 = 1.4n_1$ 。其他条件同例 6.9。

**解答** 在样本容量公式中,用与  $m = 1.4$  相应的值代替倍数 2,可得

$$\begin{aligned} n &\approx \frac{(m+1)\sigma^2(z_{\alpha/2} + z_{\beta})^2}{m\Delta^2} \\ &= \frac{(1.4+1)(0.8)^2(1.95+1.28)^2}{1.4(0.5)^2} \\ &= 46.07, \text{ 或 } 47 \end{aligned}$$

也就是说,第 I 组需 47 头奶牛,第 II 组需  $1.4(47) = 65.8$ ,即 66 头奶牛。

用第五章关于  $\mu$  的公式可计算基于成对数据(差)来估计  $\mu_d$  和进行关于  $\mu_d$  的检验所需的样本容量。惟一的不同就是我们所用的是差值的单样本而不是  $y$  值的单样本。为方便起见,将公式给出如下。

**配对样本下形如  $\bar{d} \pm E$  的  $\mu_1 - \mu_2$  的  $100(1-\alpha)\%$  置信区间所需要的样本容量**

$$n = \frac{z_{\alpha/2}^2 \sigma_d^2}{E^2}$$

(注:若  $\sigma_d$  未知,则用其估计值来代替以得到近似的样本容量。)

**配对样本下  $\mu_1 - \mu_2$  的假设检验所需要的样本容量**

$$\text{单侧检验: } n = \frac{\sigma_d^2(z_{\alpha} + z_{\beta})^2}{\Delta^2}$$

$$\text{双侧检验: } n = \frac{\sigma_d^2(z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$$

其中,若真正的差值  $\mu_d \geq \Delta$ ,则犯第 II 类错误的概率不超过  $\beta$ 。(注:若  $\sigma_d$  未知,则用其估计值来代替以得到近似的样本容量。)

## 6.7 小 结

在本章中,我们讨论了关于  $\mu_1 - \mu_2$  的统计推断问题。第 I 类方法是基于来自我们感兴趣总体的独立随机样本。还学习了如何对样本数据用  $t$  方法对  $\mu_1 - \mu_2$  进行统计检验,或构造  $\mu_1 - \mu_2$  的置信区间。Wilcoxon 秩和检验不要求总体服

从正态分布,可以用来代替  $t$  检验。

当样本观察值是成对数据时,可用第二种方法来比较两总体。在这种情况下,我们不再有独立随机样本,因此第 6.2 和第 6.3 节中的方法( $t$  方法和 Wilcoxon 秩和法)不再适用。成对数据的检验和估计是基于成对数据的差值或这些差值的秩。我们介绍了基于差值的配对  $t$  检验和相应的置信区间,发现它们与第五章单样本  $t$  方法相同。Wilcoxon 符号秩检验是一个可以用来代替配对  $t$  检验的非参数方法。

第五章和第六章的内容为后面的统计推断(估计和检验)奠定了基础。当你研究一个新问题时,应定期复习本章的内容,以便记住统计推断的基本原理。

### 重要公式

1. 对于独立样本,  $y_1$  和  $y_2$  近似正态,  $\sigma_1^2 = \sigma_2^2$ ,  $\mu_1 - \mu_2$  的  $100(1 - \alpha)\%$  置信区间为

$$\bar{y}_1 - \bar{y}_2 \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

其中

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad \text{且} \quad df = n_1 + n_2 - 2$$

2. 对于独立样本,  $y_1$  和  $y_2$  近似正态,  $\sigma_1^2 = \sigma_2^2$ ,  $\mu_1 - \mu_2$  的  $t$  检验为

$$\text{T. S. : } t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad df = n_1 + n_2 - 2$$

3. 对于不等方差,独立样本,  $y_1$  和  $y_2$  近似正态,  $\mu_1 - \mu_2$  的  $t'$  检验为

$$\text{T. S. : } t' = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad df = \frac{(n_1 - 1)(n_2 - 2)}{(1 - c)^2(n_1 - 1) + c^2(n_2 - 1)}$$

$$\text{其中 } c = \frac{\frac{s_1^2/n_1}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

4. 对于不等方差,独立样本,  $y_1$  和  $y_2$  近似正态,  $\mu_1 - \mu_2$  的  $100(1 - \alpha)\%$  置信区间为

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

其中,  $t$  分位数具有自由度

$$df = \frac{(n_1 - 1)(n_2 - 2)}{(1 - c)^2(n_1 - 1) + c^2(n_2 - 1)}$$

$$\text{而 } c = \frac{\frac{s_1^2/n_1}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

### 5. 基于独立样本的 Wilcoxon 秩和检验

$H_0$ : 两个总体分布相同。

$(n_1 \leq 10, n_2 \leq 10)$

T. S. :  $T$ , 样本 1 的秩和

$(n_1, n_2 > 10)$

$$\text{T. S. : } z = \frac{T - \mu_T}{\sigma_T}$$

其中,  $T$  表示样本 1 的秩和, 在没有等秩的情况下

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \quad \sigma_T = \sqrt{\frac{n_1 n_2}{12}(n_1 + n_2 + 1)}$$

### 6. 差值近似于正态时的配对 $t$ 检验

$$\text{T. S. : } t = \frac{\bar{d} - D_0}{s_d/\sqrt{n}} \quad df = n - 1$$

其中  $n$  是差值的个数。

7. 对于配对数据, 差值近似正态时,  $\mu_d$  的  $100(1 - \alpha)\%$  置信区间为

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

### 8. 基于配对样本数据的 Wilcoxon 符号秩检验

$H_0$ : 差值分布关于  $D_0$  对称

T. S. :  $(n \leq 50) T_-$ , 或  $T_+$ , 或  $T_+$  与  $T_-$  中较小者, 依赖于  $H_a$  的形式

T. S. :  $n > 50$

$$z = \frac{T - \mu_T}{\sigma_T}$$

其中, 在没有等秩的情况下  $\mu_T = \frac{n(n+1)}{4}$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

9. 独立样本:  $\mu_1 - \mu_2$  的  $100(1 - \alpha)\%$  置信区间估计形如  $\bar{y}_1 - \bar{y}_2 \pm E$  时所需的样本容量

$$n = \frac{2z_{\alpha/2}^2 \sigma^2}{E^2}$$

10. 独立样本: 检验  $\mu_1 - \mu_2$  所需的样本容量

$$\text{a. 单侧检验: } n = 2\sigma^2 \frac{(z_\alpha + z_\beta)^2}{\Delta^2}$$



b. 双侧检验:  $n = 2\sigma^2 \frac{(z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$

11. 配对样本:  $\mu_1 - \mu_2$  的  $100(1 - \alpha)\%$  置信区间估计形如  $\bar{d} \pm E$  时所需的样本容量

$$n = \frac{z_{\alpha/2}^2 \sigma_d^2}{E^2}$$

12. 配对样本: 检验  $\mu_1 - \mu_2$  所需的样本容量

a. 单侧检验:  $n = \frac{2\sigma_d^2 (z_{\alpha} + z_{\beta})^2}{\Delta^2}$

b. 双侧检验:  $n = \frac{2\sigma_d^2 (z_{\alpha/2} + z_{\beta})^2}{\Delta^2}$

补充练习

6.37 (工程) 有人建议用一种新合金制造钢梁。公司的冶金专家设计了一个研究方案来比较新合金和目前使用的合金之间的强度。分别用每种合金制造 10 根梁。测定这 20 根梁的承重能力, 结果如下:

梁	1	2	3	4	5	6	7	8	9	10
老合金	23.7	24.0	21.1	23.1	22.8	25.0	25.3	22.6	23.3	22.8
新合金	26.6	32.3	28.0	31.1	29.6	28.5	31.2	23.6	29.1	28.5

- 我们感兴趣的总体是什么?
- 分别求出用新旧合金制造的梁其平均承重能力的 90% 置信区间。
- 给定  $\alpha = 0.01$ , 问: 新合金的平均承重能力是否大于旧合金的平均承重能力? 并求出检验的  $p$  值。
- 在(b)和(c)中使用的统计方法, 其所需的前提条件是否满足?
- 用新合金制造大梁要比用旧合金费用高很多, 所以只有当新合金的承重能力比旧合金高至少 5 吨时, 才选用新合金。根据这个标准, 你是否建议公司用新合金?

6.38 (医学) 长跑运动员认为适度地暴露于臭氧环境可增加肺容量。为了调查这种可能性, 研究人员将老鼠放在浓度为 2ppm 的臭氧环境中 30 天。在研究前和置于臭氧中 30 天后, 分别测出老鼠的肺活量(ml)。见下表

老鼠	1	2	3	4	5	6	7	8	9	10	11	12
暴露前	8.7	7.9	8.3	8.4	9.2	9.1	8.2	8.1	8.9	8.2	8.9	7.5
暴露后	9.4	9.8	9.9	10.3	8.9	8.8	9.8	8.2	9.4	9.9	12.2	9.3

- 已知  $\alpha = 0.05$ , 问: 是否有足够的理由相信臭氧可增加肺容量? 并给出检验的  $p$  值。
- 估计暴露于臭氧环境后老鼠肺活量增加量的 95% 置信区间。
- 经研究, 调查人员称臭氧能增加肺活量。该试验支持这一论断吗?

**6.39 (环境)** 在关于新机场对环境影响的研究中, 研究人员测量了各种喷气式飞机在轮子刚离开地面几秒钟之后的噪声水平。这些飞机包括窄体和宽体, 下表列出了 15 架宽体飞机和 12 架窄体飞机的噪声水平(单位: 分贝)

宽体飞机

109.5 107.3 105.0 117.3 105.4 113.7 121.7 109.2 108.1 106.4 104.6 110.5 110.9 111.0 112.4

窄体飞机

131.4 126.8 114.1 126.9 108.2 122.0 106.9 116.3 115.5 111.6 124.5 116.2

- 两种飞机的噪声水平是否不同? 给出检验的显著性水平。
- 估计两种飞机平均噪音水平差值的 95% 置信区间。
- 你怎样选择研究所需的飞机?

**6.40 (农业)** 某昆虫学家正在研究  $F_1$  和  $F_2$  两种熏剂中哪种熏剂能够更有效地控制烟草上的寄生虫。为了比较这两种熏剂, 在 9 块土质, 排水、防风性皆不同的土壤上种植烟草。再将每块土地分成大小相同的两小块。随机地将  $F_1$  施用于其中一小块土地上,  $F_2$  施用于另一小块土地上。从每块地上选取 50 株烟草, 每小块各 25 株。数一下寄生虫的个数, 数据如下表所示:

地块	1	2	3	4	5	6	7	8	9
熏剂 $F_1$	77	40	11	31	28	50	53	26	33
熏剂 $F_2$	76	38	10	29	27	48	51	24	32

- 我们感兴趣的总体是什么?
- 给定  $\alpha = 0.01$ , 问: 这些数据是否提供了充分的证据表明施用不同的熏剂其平均寄生虫水平存在差异? 并求出检验的  $p$  值。
- 估计施用两种熏剂所导致的平均寄生虫水平差值的 90% 置信区间。

**6.41** 见练习 6.40。在另一个试验设计中, 随机地将  $F_1$  施用于某 9 块土地上,  $F_2$  施用于另外 9 块土地上, 而不考虑土地的类型。这种试验方法会导致什么问题?

**6.42 (环境)** 1989 年 3 月 24 日, *Exxon valdez* 号油轮在阿拉斯加搁浅之后, 大约有 3500 吨原油被泄露到 Prince William Sound 海区。在论文“The deep ben-

thos of Prince William Sound, Alaska, 16 months after the Exxon Valdez oil spill" [Marine Pollution Bulletin (1998), 36:118~130]中,作者报告了关于原油泄露之后对深海生物的评价。研究中选取了13个地点,其中7个点位于石油流经路线,另6个点在路线之外。在1990年7月1日到30日一段时间内,采集了水下40米和100米处环境和生物数据。其中一个变量是生物丰富度(每平方米生物个体数)。数据如下表所示:

地点	石油流经路线之内						
	1	2	3	4	5	6	7
40 米	5,124	2,904	3,600	2,880	2,587	4,416	1,408
100 米	3,228	2,034	3,256	3,816	2,438	4,897	1,346

地点	石油流经路线之外					
	1	2	3	4	5	6
40 米	1,336	394	7,370	6,762	744	1,874
100 米	1,676	2,008	2,224	1,234	1,598	2,182

- 将两深度数据混合,路线内外的生物丰富度是否有差异?
- 估计两类地点生物丰富度差的95%置信区间。已知  $\alpha = 0.05$
- (a)(b)中方法所需条件是什么?
- 检查这些条件是否已满足。

6.43 题目见练习6.42。利用6.42中两深度的混合数据回答下列问题:

- 用 Wilcoxon 秩和检验法估计路线内外生物丰富度是否存在差异?
- (a)中方法所需条件是什么?
- 这些条件是否已满足?
- 讨论  $t$  检验和 Wilcoxon 秩和检验所得结论有什么不同。

6.44 题目见练习6.42。研究人员同时也检验了深度对生物丰富度的影响。

- 将四个数据集画成并立盒形图以证实深度对生物丰富度的影响。
- 对于每个深度,评价路线内外生物丰富度的差异。已知  $\alpha = 0.05$
- 在40米处和100米处你的结论一致吗?

6.45 题目见练习6.42~6.44。

- 讨论命题“石油泄漏对生物丰富度没有负面影响,事实上,它似乎增加了生物丰富度”的正确性。
- 对这项研究的一个可能的批评是:路线外6个点与路线内7个点在许多方面不具备可比性。假定研究人员现有路线内7个地点泄漏前生物丰富度

的数据。可用哪种方法评价原油泄漏对生物丰富度的影响？用这种数据而不是用练习 6.43 的数据有哪些优点？

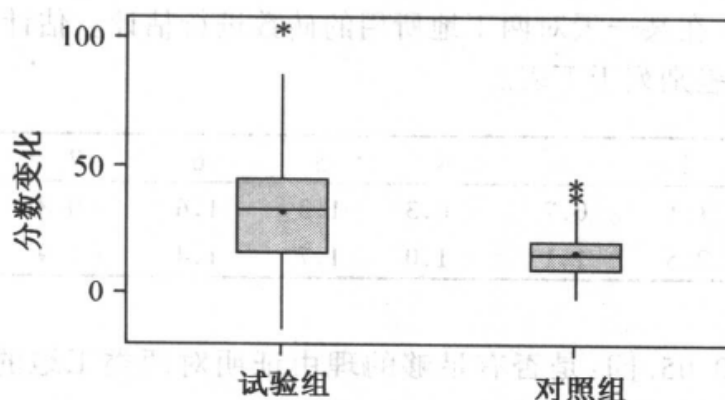
- c. 用原油泄漏前后的数据评价原油泄漏对生物密度产生的影响可能会出现哪些问题？

6.46 数据见练习 3.13。

- 新疗法是否看起来比标准疗法使得平均生存时间更长？ $\alpha = 0.05$ 。
- 估计使用新疗法时平均生存时间的 95% 置信区间。
- 估计使用新疗法与标准疗法时平均生存时间差的 95% 置信区间。该置信区间与(a)的结论一致吗？
- (a)~(c)所用方法的条件是否满足？

6.47 (教育) 某知名大学教育学院的研究小组设计了一个研究方案用来评价以下猜想：老师的期望会对学生的成绩产生影响。将 200 名成绩相当的学生随机地分到试验组或对照组。试验组的老师被告知这些学生都是班级中成绩很好的人，且最有可能在标准化考试中取得较大进步。而对照组的老师却不被告知学生的任何情况。在学期初和学期末对这些学生进行了标准化考试。两次考试的分数变化情况见下列图表。

	样本容量	均值	标准差
试验组	100	26.5	24.2
对照组	100	17.0	9.1



试验组和对照组盒形图(均值以实心圆点表示)

- 是否可看出试验组平均成绩变化较大？
- 给出检验的  $p$  值。

- c. 估计两组学生平均分数变化的差。
- d. (a)~(c)所用方法的条件是否满足?
- e. 用非技术术语,也就是说,用非统计术语叙述关于两组结果差异的结论。

**6.48** 题目见练习 6.47。有批评认为,差别很大的教学方法和学生能力可能会影响研究的结果。因两组中老师和学生都不同,所以不可能知道分数的差异是源于老师的期望,学生能力的内在差距还是不同老师的教学方法。给出其他研究方法以解决其中一部分问题。

**6.49 (医学)** 某试验旨在比较癫痫病患者对两种药品  $D_1$ 、 $D_2$  的平均吸收时间。随机抽取 20 个癫痫病患者进行研究。其中 10 个服用  $D_1$ , 另外 10 个服用  $D_2$ 。将血管中药物达到某一特定水平的时间记录如下:

患者	1	2	3	4	5	6	7	8	9	10
药品 $D_1$	19.8	45.4	32.0	24.5	47.2	18.1	50.2	47.2	16.8	41.2
药品 $D_2$	23.0	37.3	11.4	60.6	72.1	41.4	42.6	43.8	42.8	65.3

- a. 是否有足够的理由认为两种药品的平均吸收时间存在差异?
- b. 给出检验的显著性水平。
- c. 试估计两种药物平均吸收时间的差。
- d. (a)~(c)所用方法的条件是否满足?

**6.50 (工程)** 通过每天目测所消耗的原材料数量, 建筑公司经常需要估计工地施工的进度。公司希望知道这种估计的精确度是否与工地类型有关。考虑两类工地: 高层建筑和一层建筑。从公司数百个工地负责人中随机选出 9 个参加这项研究。每个负责人在某一天对两工地所用的砖数进行估计。估计数量与实际数量(单位: 千块砖)的差别列于下表。

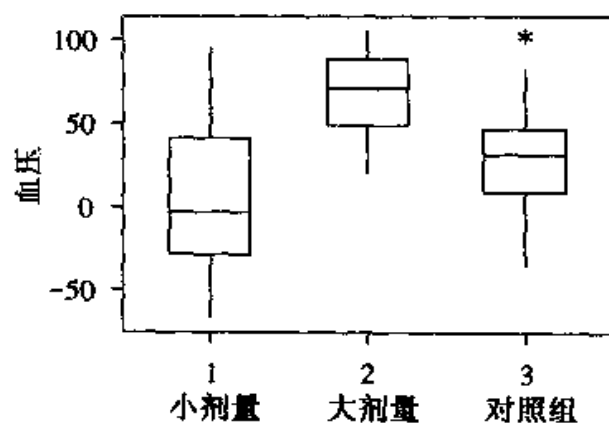
负责人	1	2	3	4	5	6	7	8	9
高层	0.9	1.1	0.7	0.3	1.3	1.6	-0.8	1.4	1.7
一层	-1.6	2.5	1.1	-1.0	1.7	1.4	1.9	1.8	1.9

- a. 已知  $\alpha = 0.05$ , 问: 是否有足够的理由证明对两类工地的估计存在差异? 给出检验的  $p$  值。
- b. 试估计对两类工地进行目测估算的平均精确度的差。
- c. (a)和(b)中所用方法的条件是否满足? 并解释之。

**6.51 (生物)** 为了评价一种抗高血压药物的有效性, 进行了一项研究。从同一品种的高血压老鼠中随机选出 3 组, 每组 20 只。第一组服用小剂量的药物, 第二组服用大剂量的药物, 第三组为无治疗对照组。注意, 负值表示血压的升高。根

据计算机的输出结果回答下列问题：

- 已知  $\alpha = 0.05$ , 试比较对照组和大剂量组的平均降压量, 并给出检验的显著性水平。
- 估计大剂量组和对照组平均降压量差的 95% 置信区间。
- (a) 和 (b) 中所用方法的条件是否满足? 并解释之。



血压数据的盒形图

Row	Low Dose	High Dose	Control
1	-45.1	54.2	18.2
2	-59.8	89.1	17.2
3	58.1	89.6	34.8
4	-23.7	98.8	3.2
5	64.9	107.3	42.9
6	12.1	65.1	-27.2
7	10.5	75.6	42.6
8	42.5	52.0	10.0
9	48.5	50.2	102.3
10	-1.7	80.9	61.0
11	-65.4	92.6	-33.1
12	-17.5	55.3	55.1
13	22.1	103.2	84.6
14	-15.4	45.4	40.3
15	96.5	70.9	30.5
16	-27.7	29.7	18.5
17	-16.7	40.3	29.3
18	39.5	73.3	-19.7

19	-4.2	21.0	37.2
20	-41.3	73.2	48.8

Two-sample T for Low Dose vs Control

	N	Mean	StDev	SE Mean
Low Dose	20	3.8	44.0	9.8
Control	20	29.8	34.0	7.6

95% CI for  $\mu$  Low Dose- $\mu$  Control: (-51.3, -0.8)

T-Test  $\mu$  Low Dose =  $\mu$  Control (vs not =); T = -2.09 P = 0.044 DF = 35

Two-sample T for High Dose vs Control

	N	Mean	StDev	SE Mean
High Dose	20	68.4	24.5	5.5
Control	20	29.8	34.0	7.6

95% CI for  $\mu$  High Dose- $\mu$  Control: (19.5, 57.6)

T-Test  $\mu$  High Dose- $\mu$  Control (vs not =); T = 4.12 P = 0.0002 DF = 34

Two-sample T for Low Dose vs High Dose

	N	Mean	StDev	SE Mean
Low Dose	20	3.8	44.0	9.8
High Dose	20	68.4	24.5	5.5

95% CI for  $\mu$  Low Dose- $\mu$  High Dose: (-87.6, -41.5)

T-Test  $\mu$  Low Dose =  $\mu$  High Dose (vs not =); T = -5.37 P = 0.0000 DF = 29

6.52 题目见练习 6.51。

- 已知  $\alpha = 0.05$ , 试比较对照组和小剂量组的平均降压量, 并给出检验的显著性水平。
- 估计小剂量组和对照组平均降压量差的 95% 置信区间。
- (a) 和 (b) 中所用方法的条件是否满足? 并解释之。

6.53 题目见练习 6.51。

- 已知  $\alpha = 0.05$ , 试比较大剂量组和小剂量组的平均降压量, 并给出检验的显著性水平。
- 估计小剂量组和大剂量组平均降压量差的 95% 置信区间。
- (a) 和 (b) 中所用方法的条件是否满足? 并解释之。

6.54 在练习 6.51~6.53 中, 我们根据同一数据集中相应的部分数据进行了三组假设检验。定义整个试验的第 I 类错误率为检验任意一组假设时至少犯一次第 I 类错误的概率。

- 如果在  $\alpha = 0.05$  水平下检验每一假设, 试估计整个试验的第 I 类错误的概率。

b. 给出一种检验方法以确保整个试验的第 I 类错误的概率不超过 0.05。

**6.55 (工程)** 某回收铝罐加工者很关心从两种来源回收的铝罐其杂质(主要是其他金属)的含量。在试验室对样本进行分析得以下数据(每百千克产品中杂质的千克数):

来源 I: 3.8 3.5 4.1 2.5 3.6 4.3 2.1 2.9 3.2 3.7 2.8 2.7  
(均值 = 3.617, 标准差 = 0.676)

来源 II: 1.8 2.2 1.3 5.1 4.0 4.7 3.3 4.3 4.2 2.5 5.4 4.6

a. 计算混合方差和标准差。

b. 估计平均杂质水平差的 95% 置信区间。

c. 该加工者在  $\alpha = 0.05$  水平下能否得出这样的结论, 即, 平均杂质水平差并不等于零?

**6.56** 为比较两种计算机电子制表程序的性能, 将学生分成每三人一组, 每组均根据自己的喜好来选择电子制表程序, 以解决相同的标准会计和金融方面的问题。记录每个小组所花的时间(分钟)。下表列出了利用应用最广泛的两种制表程序解决问题所花费的时间, 同时列出的还有样本均值, 样本标准差和样本容量。

程序	时间										$\bar{y}$	$s$	$n$
A	39	57	42	53	41	44	71	56	49	63	51.50	10.46	10
B	43	38	35	45	50	28	50	54	37	29			
	36	27	52	33	31	30					38.00	8.67	16

a. 计算混合方差。

b. 根据这个方差, 求总体均值差的 99% 置信区间。

c. 依据此区间, 在  $\alpha = 0.01$  水平下, 均值相等的零假设是否应被拒绝?

**6.57** 用分离方差法( $t'$ 方法)重做练习 6.56 中的(b)和(c)。在此实例中哪种方法更为合适? 方法的选用很关键吗?

**6.58** 题目见练习 3.32。对于陆地和水生鸟类, DDE 与 PCB 的平均比值是否存在显著差异? 利用  $\alpha = 0.05$ , 并给出检验的显著性水平。

**6.59** 为评价有学位的护士是否比注册护士接受过更为综合的训练, 研究人员作了一项研究来对此进行比较。州护士执照办公室分别从两组中随机选出 50 名护士, 并对她们进行了护士资格考试。分数见以下图表。

#### Two-Sample T-Test and Confidence Interval

Two-sample T for Degreed vs RN

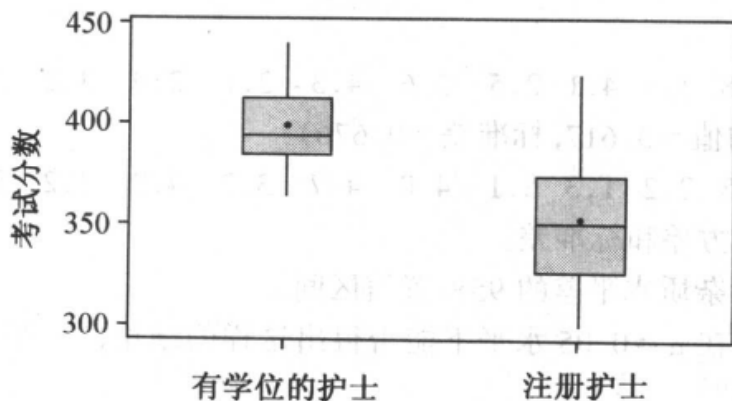
	N	Mean	StDev	SE Mean
Degreed	50	399.9	17.2	2.4



RN 50 354.7 30.9 4.4

95% CI for  $\mu_{\text{Degreed}} - \mu_{\text{RN}}$ : (35.3, 55.2)

T-Test  $\mu_{\text{Degreed}} = \mu_{\text{RN}}$  (vs >):  $T = 9.04$   $P = 0.0000$   $DF = 76$



有学位的护士和注册护士成绩的盒形图  
(均值以实心圆点表示)

- 护士执照办公室是否能得出这样的结论, 即, 有学士学位护士的平均分数高于注册护士? 给定  $\alpha = 0.05$ 。
- 给出检验的近似  $p$  值。
- 估计两组护士平均分数差的 95% 置信区间。
- 只有当分数差别大于 40 时, 才视为有实际意义。平均分数的观察差异值是否有实际意义?

**6.60 (社会)** 设计了一项研究旨在比较市中心的食物价格是否高于市郊。将一篮食品分别在市中心的  $n$  个店和市郊的  $n$  个店进行估价。假设两处的价格服从正态分布, 市中心店的价格范围为 115 美元至 135 美元, 市郊店的价格范围为 100 美元至 120 美元。如果我们以 95% 的把握认为平均价格差的估计值与真正值相差不超过 4 美元, 那么, 需要的样本容量是多少?

**6.61** 题目见练习 6.60。随机选出 13 个市中心店和 14 个市郊店进行研究。每店对该篮的估价列出如下。

食品店							
郊区	109.11	106.94	118.37	111.09	117.74	109.48	113.59
市中心	131.76	121.39	122.37	130.47	125.81	134.47	122.15
食品店							
郊区	117.49	117.74	108.68	113.00	122.97	111.87	113.00
市中心	135.47	128.26	129.90	129.24	134.72	126.17	

- a. 估计在两地该篮食品的平均价格差。
- b. 该研究是否实现了练习 6.60 中规定的目标?
- c. 是否有充分的理由认为市中心食品的价格较高? 用非统计术语叙述你的结论。
- d. 你可能犯了哪类错误, 第 I 类还是第 II 类?

**6.62** 考虑第 6.1 节关于石油泄漏的案例。

- a. 我们感兴趣的总体是什么?
- b. 除了植物密度, 还有那些因素可表明石油泄漏对沼泽有影响?
- c. 描述如何随机地选取研究中所用的土地。
- d. 假定研究人员在石油泄漏 5 年前曾测量过石油泄漏地区 20 块成片土地上的植物密度数据。根据这一信息, 试给出另一种分析方法。

**6.63 (政治科学)** 公共职位的竞选者必须公布他们在竞选中的花费。政治科学家认为女候选人一般很难募集到款项从而她们的花费要比男候选人低。假定下表中列出的数据为从某州立法机关职位竞选者中随机选出的男女候选人的花费数据。这些数据是否支持上述论断?

竞选费用(单位:千美元)

候选人	1	2	3	4	5	6	7	8	9	10
男性	169	206	257	294	252	283	240	207	230	183
女性	289	334	278	268	336	438	388	388	394	394
候选人	11	12	13	14	15	16	17	18	19	20
男性	298	269	256	277	300	126	318	184	252	305
女性	425	386	356	342	305	365	355	312	209	458

- a. i. 用非专业术语叙述零假设和备择假设。  
ii. 用统计术语或符号叙述零假设和备择假设。
- b. 估计女性候选人与男性候选人竞选费用的差。
- c. 在 0.05 水平下, 此差异在统计意义上是否显著?
- d. 此差异有实际意义吗?

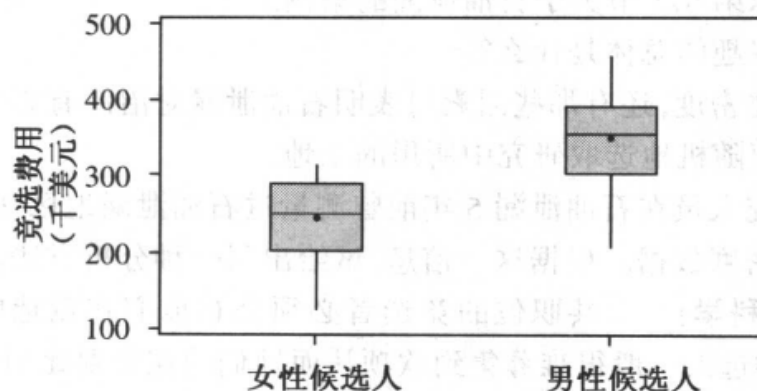
**6.64** 题目见练习 6.63。若要用  $t$  方法分析数据, 需满足那些条件? 根据图表确定在练习 6.63 中的数据是否满足这些条件。

Two-Sample T-Test and Confidence Interval

Two-sample T for Female vs Male

N	Mean	StDev	SE Mean
---	------	-------	---------

Female 20 245.4 52.1 12  
 Male 20 350.9 61.9 14  
 95 % CI for mu Female-mu Male: (-142, -69)  
 T-Test mu Female = mu Male(vs not = ): T = -5.83 P = 0.0000 DF = 38  
 Both use Pooled StDev = 57.2



女性候选人与男性候选人的盒形图  
 (实心圆点表示均值)

**6.65 (环境)** 对一地区进行煤的露天开采后, 开采公司必须将土地恢复到开采前水平。因 pH 值是决定土地适于播种哪种植物的重要因素, 故应考虑 pH 值的变化。在开采之前, 将这片地划分为格子地块, 并从中随机抽取 15 块, 测出它们的 pH 值。开采完之后, 将土地复原, 在同样的 15 块地上测量 pH 值。见下表:

地块	1	2	3	4	5	6	7
开采前	10.02	10.16	9.96	10.01	9.87	10.05	10.07
开采后	10.21	10.16	10.11	10.10	10.07	10.13	10.08

地块	8	9	10	11	12	13	14	15
开采前	10.08	10.05	10.04	10.09	10.09	9.92	10.05	10.13
开采后	10.30	10.17	10.10	10.06	10.37	10.24	10.19	10.13

- 州土地办公室对数据进行了分析, 并给出了计算机输出结果。对土壤复原后的平均 pH 值变化进行检验, 问, 其检验的显著性水平是多少?
- 州土地办公室所检验的研究假设是什么?
- 估计开采前后平均 pH 值变化的 99% 置信区间。
- 因  $t$  检验显示在土壤复原后 pH 值具有明显差异, 故土地办公室欲对此公

司进行罚款。你认为这些数据是否支持这一决定？并根据(a)和(c)的结果解释之。

**6.66** 题目见练习 6.65。基于土地办公室在假设检验中所作的结论,他们可能犯了以下哪类错误(从 a, b, c, d 中选一项)。

- 第 I 类错误。
- 第 II 类错误。
- 第 I 类错误和第 II 类错误。
- 两类错误都没有。

**6.67** 题目见练习 3.13。用下列图表分析对应于标准疗法和新疗法的生存时间。

#### Two-Sample T-Test and Confidence Interval

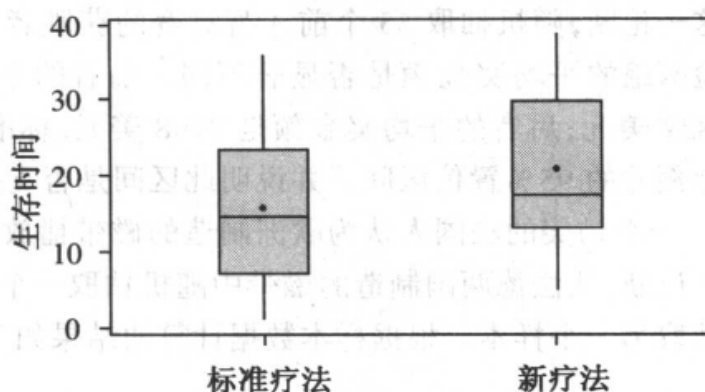
Two-sample T for Standard vs New

	N	Mean	StDev	SE Mean
Standard	28	15.68	9.63	1.8
New	28	20.71	9.81	1.9

95 % CI for  $\mu$  Standard- $\mu$  New: (-10.2, 0.2)

T-Test  $\mu$  Standard =  $\mu$  New (vs <): T = -1.94 P = 0.029 DF = 54

Both use Pooled StDev = 9.72



标准疗法和新疗法盒形图(均值以实心圆点表示)

- 这些数据是否支持研究假设:新疗法比标准疗法有更长的生存时间?
- 估计新疗法增加的生存时间的 95% 置信区间?
- 若要用  $t$  检验分析数据,需满足哪些条件? 这些条件满足吗?

**6.68 (社会)** 为确定中产阶级家庭中不工作的妇女是否比工薪阶层家庭中不工作的妇女有更多的志愿协会成员资格,进行了一项研究。随机抽取家庭主妇

作为样本,并询问她们丈夫的职业及本人的会员资格情况,根据其丈夫的职业将这些妇女分为中产阶级和工薪阶层,计算平均每组具有志愿协会成员的资格数。15个中产阶级妇女中,平均每名妇女的成员资格数为  $\bar{y}_1 = 3.4$ , 方差  $s_1 = 2.5$ , 15个工薪阶层妇女中,平均每名妇女的成员资格数为  $\bar{y}_2 = 2.2$ , 方差  $s_2 = 2.8$ 。

- a. 利用这些数据构造两组妇女平均成员资格数差的 95% 置信区间。
- b. 这些数据是否支持研究假设: 中产阶级妇女成员资格数要多于工薪阶层?

**6.69 (政府)** 某地区美国国税的审计员以 3 月 15 日之前的退税申报的样本进行检验,以判断今年纳税人的退税额是否大于去年。随机抽取每年 100 个退税额作为样本。数据如下:

	去年	今年
均值	320	410
方差	300	350
样本容量	100	100

- a. 这个审计员的研究假设是什么?
- b. 进行假设检验,你能得到什么结论?
- c. 假设检验所需的条件是什么? 该数据集是否满足这些条件? 为什么?

**6.70 (社会)** “美国小姐”大会的官员称:若要在大会中胜出,才智比美貌更为重要。为评价这一论断,随机抽取 55 个前才智竞赛的获胜者和 53 个前泳装竞赛的获胜者。检验两组的平均奖金额是否显著不同。前者的平均奖金额是 8645 美元,标准差为 5829 美元;后者的平均奖金额是 9198 美元,标准差为 8185 美元。计算两组平均奖金额差的 95% 置信区间。并说明此区间是否支持该官员的论断。

**6.71 (商业)** 一个访美的法国人认为欧洲制造的磁带播放时间比美国的长。为证实或反驳这一论断,从法德两国制造的磁带中随机抽取一个样本,再从美国制造的磁带中随机抽取另一个样本。根据样本数据计算的结果如下:

	外国产	美国产
样本容量	14	14
平均播放时间(秒)	207.45	182.54
标准差	41.43	37.32

国外生产的磁带播放时间是否更长? 给定  $\alpha = 0.05$ 。

**6.72 (政府)** 一个位于华盛顿的联邦机构定期提供 PL/1 课程,其中 PL/1

为在该机构内部编写程序时所用的一种计算机编程语言。第一周由外部咨询公司的人授课,第二周由本机构的一计算机职员授课。结果如下:

外部教师授课	38	42	53	37	36	48	47	47	44
本机构人员授课	46	33	38	60	58	52	44	45	51

这些数据是学员 1 周内的总分数(最大可能值为 64)。假定分数可反映教学成果,这些数据是否能充分证明两个教师的教学成果存在差异?

**6.73 (医学)** 公司领导很关心某种药品药效保持的时间。从当前生产线上随机抽取 10 瓶药品,记为样本 1,分析它们的药效。将它们储存一年,再进行分析,这样得到第二个样本,记为样本 2。所得数据如下:

样本 1	10.2	10.5	10.3	10.8	9.8	10.6	10.7	10.2	10.0	10.6
样本 2	9.8	9.6	10.1	10.2	10.1	9.7	9.5	9.6	9.8	9.9

用标准统计软件 SAS 对这些数据进行分析。相关结果输出如下:

```
Variable: POTENCY          TTEST PROCEDURE

SAMPLE  N      Mean      Std Dev   Std Error  Variances    T      DF    Prob>|T|
  1      10  10.37000000  0.32335052  0.10225241  Unequal    4.2368  16.6   0.0006
  2      10   9.83000000  0.24060110  0.07608475   Equal    4.2368  18.0   0.0005

For H0: Variances are equal, F' = 1.81  DF = (9,9)  Prob>F' = 0.3917
```

- 研究假设是什么?
- $t$  和  $t'$  统计量的值等于多少? 对于该数据集来说,为什么它们相等?
- $t$  和  $t'$  统计量的  $p$  值为多少? 为什么它们不相等?
- 在  $\alpha=0.05$  水平下,两种检验对此研究假设所作的结论是否相同?
- $t$  和  $t'$  检验,哪一个更适合于分析该数据集?

**6.74** 考虑两种方法改进喷气式飞机的发动机以减少噪音。将两个系统安装在同样的飞机上。为比较两架飞机的噪音,在主体机场的跑道正下方安装了噪音记录装置。每当飞机降落到机场时,就记录其噪音水平。用 SAS 软件对这些数据进行分析,相关结果如下:

TTEST PROCEDURE

Variable: NOISE

System	N	Mean	Std Dev	Std Error	Variances	T	DF	Prob> T
H	42	99.67299428	11.30050001	1.74370501	Unequal	2.2236	41.3	0.0317
R	20	93.30597443	10.15740464	2.27126473	Equal	2.1400	60.0	0.0364

For HO: Variances are equal,  $F' = 1.24$  DF = (41, 19) Prob>F' = 0.6290

- 研究假设是什么?
- $t$  和  $t'$  统计量的值等于多少? 对于该数据集来说, 为什么它们不相等?
- $t$  和  $t'$  统计量的  $p$  值是什么?
- 在  $\alpha = 0.05$  水平下, 两种检验对此研究假设所作的结论是否相同?
- $t$  和  $t'$  检验, 哪一个更适合于分析该数据集?

**6.75 (农业)** 对 16 头奶牛进行研究。随机抽取 8 头(第一组), 让其仅饮用清水, 其余 8 头(第二组)饮用液体乳清。另外每头牛每天可获得 7.5kg 谷物, 并且可以任意食用青草。尽管两组奶牛的产奶情况(比如说产量或奶中脂肪含量)并无显著差异, 但每日青草的食用量却引起了研究人员的兴趣。数据如下:

组 1	15.1	14.9	14.8	14.2	13.1	12.8	15.5	15.9
组 2	6.8	7.5	8.6	8.4	8.9	8.1	9.2	9.5

- 根据这些数据检验研究假设, 即两组牛的平均青草食用量存在差异。
- 估计两组牛平均青草食用量的差。

**6.76** 题目见练习 6.75。下表列出了在研究之前 16 头奶牛的体重(千磅)。

奶牛	1	2	3	4	5	6	7	8
样本 1	12.3	16.0	15.6	12.4	14.2	18.6	16.6	15.8
样本 2	22.7	20.1	17.9	21.0	20.5	19.5	20.6	21.4

- 这些数据是否能充分证明在研究之前两组牛的体重存在显著差异?
- 假定我们要根据 16 头牛的体重数据重新设计研究方案。你怎样将两种食物分配给这些牛以减少它们体重差异的影响?

**6.77 (工程)** 一个工业公司欲对几种用于火箭推进剂的混合物进行试验, 这些混合物是由四种成分——镁, 硝酸盐, 硝酸铝, 黏合剂——按不同比例混合而成的。试验发现有两种混合物发出的火花亮度比其他的要高。混合物 1 中四种原料比例分别为 0.40, 0.10, 0.42 和 0.08, 混合物 2 中四种原料比例分别为 0.60,

0.27, 0.10 和 0.05。准备 20 份不同的混合物(每种混合物 10 份), 并检测出其火花亮度。数据如下(以 1000 个烛光单位计)。

混合物 1	185	192	201	215	170	190	175	172	198	202
混合物 2	221	210	215	202	204	196	225	230	214	217

a. 将数据绘成图。哪种检验可用于比较两种混合物的平均亮度。

b. 给出检验的显著性水平, 并解释之。

**6.78** 题目见练习 6.77。根据样本数据而不是统计检验回答, 两种混合物的平均亮度差是多少?

**6.79** 题目见练习 6.77。对于这两种混合物, 假定我们要检验研究假设:  $\mu_1 < \mu_2$ , 并假定总体服从正态分布, 具有共同标准差  $\sigma = 12$ , 若要求检验满足: 当  $\mu_1 - \mu_2 \geq 15$  时,  $\alpha = 0.05$ ,  $\beta(\mu_a) < 0.10$ 。则所需样本容量是多少?

**6.80** 参数数据盘中癫痫病研究的数据。对这些数据进行分析得到以下输出结果。研究中需测量的变量是服用安慰剂(Placebo)的病人和服用药品氟柳双胺(progabide)的病人 8 个星期后的发病次数。

Two-Sample T-Test and Confidence Interval

Two-sample T for Placebo vs Progabide

	N	Mean	StDev	SE Mean
Placebo	28	7.96	7.63	1.4
Progabide	31	6.7	11.3	2.0

95% CI for mu Placebo-mu Progabide: (-3.8, 6.3)

T-Test mu Placebo = mu Progabide(vs >): T = 0.50 P = 0.31 DF = 57

Both use Pooled StDev = 9.71

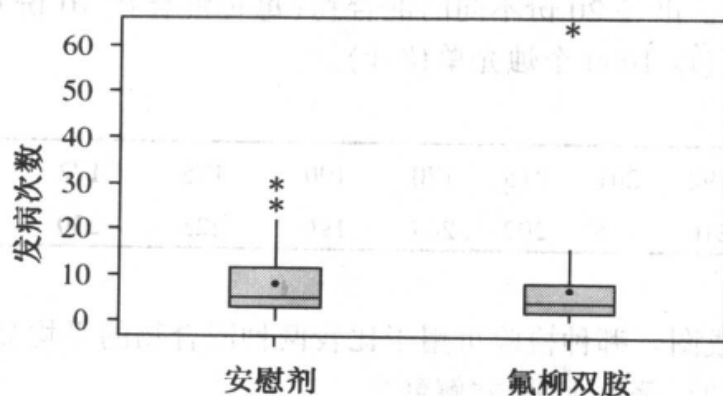
a. 数据是否支持猜想: 药品可以减少癫痫病人的发病次数? 在水平  $\alpha = 0.05$  下进行  $t$  检验和 Wilcoxon 检验。

b. 哪种检验更适合于此项研究, 为什么?

c. 估计两种病人平均发病次数的差。

**6.81 (社会)** 在一个南部大城市的居民中进行一项关于焦虑的研究, 参与研究的居民在 12 个焦虑症状前依据有无选择“是”, 或“不是”。每个人的焦虑得分从 1 到 12, 最高分表示出现了每一种焦虑症状。随机选取 50 个居民, 按性别分类, 将他们的测试结果列于下表。根据这些数据检验研究假设: 男性和女性的平均焦虑分数存在差异。给出检验的显著性水平。





安慰剂和药品的盒形图(均值以实心圆点表示)

	样本容量	均值	标准差
女性	26	5.26	3.2
男性	24	7.02	3.9

**6.82 (医学)** 某临床试验旨在研究药品 A 对于治疗有关戒酒症状的疗效。随机抽取 30 位病人,在其不知晓的情况下服用此药物,另外 30 位病人服用外表看起来和该药物一样的安慰剂。经过 1 周的治疗以后,两组的平均症状分数分别为 1.5 和 6.3。(注意:高分表示更多的戒酒“问题”)。相应的标准差为 3.1 和 4.2。

a. 比较两组的平均总症状分数。求双样本  $t$  检验  $H_0: \mu_1 - \mu_2 = 0$   $H_a: \mu_1 - \mu_2 < 0$  的  $p$  值。并给出结论。

b. 假定治疗之前的平均总症状分数是 6.8 和 12.2,这会影响你的结论吗? 你怎样防止可能存在的(处理前)基准分数的差?

**6.83 (环境)** 两个分析员,假定能力相同,分别对饮用水中某种化学物质的含量(ppm)进行测量。有人认为分析员 1 给出的测量值往往高于分析员 2。为了检验这种说法,将六个水样每个分成两份,然后请两个分析员对它们进行测量,所得结果如下。

水样	分析员 1	分析员 2
1	31.4	28.1
2	37.0	37.1
3	44.0	40.6
4	28.8	27.3
5	59.9	58.4
6	37.6	38.9

a. 有无理由认为分析员 1 测得的平均结果高于分析员 2? 给出检验的显著性水平。

b. 若用 Wilcoxon 检验, 结论是什么? 将你的结果与(a)作比较。

**6.84(农业)** 从 11 种不同烟草中各取下一片叶子, 每片叶子分成两半, 随机抽取一半用第 I 种制品处理, 另一半用第 II 种制品进行处理。试验目的是比较一段固定的时间后两种马赛克病毒制品对叶子裂痕个数的影响。试验数据在下表列出。在水平  $\alpha = 0.05$  下用 Wilcoxon 符号秩检验, 以评价研究假设: 两个总体裂痕数目分布不同。

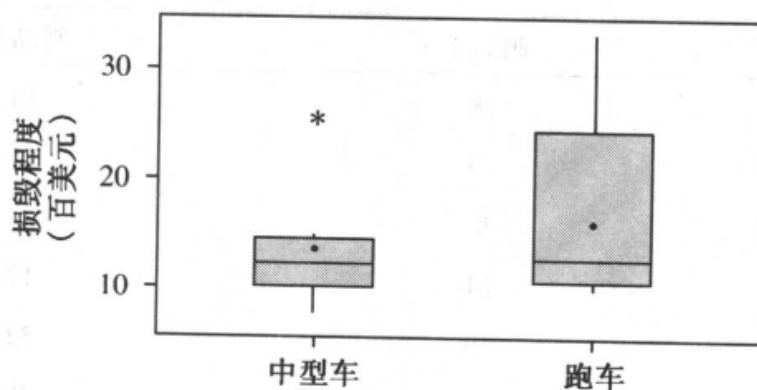
烟草	半片叶子上裂痕数目	
	制品 I	制品 II
1	18	14
2	20	15
3	9	6
4	14	12
5	38	32
6	26	30
7	15	9
8	10	2
9	25	18
10	7	3
11	13	6

**6.85(环境)** 某调查人员欲比较取自于一水处理厂两个不同地点排出的水样中微粒的平均个数。如果在每个地点取出的水样中微粒个数的标准差约为 6, 那么, 若微粒均值差的 99% 置信区间估计的宽度为 1, 则所需样本容量为多少?

**6.86(商业)** 许多人买跑车是因为他们认为跑车(SUV)更为结实, 从而比普通车更安全。但是以前的数据显示同一事故中, 跑车的修理费用要高于中型(Midsize)车。随机抽取 8 辆新跑车和 8 辆中型车, 对它们进行车头碰撞抵抗力的测试, 这些车以 20mph 的速度撞在一个固定的障碍物上。下表列出了车的毁损程度的测量值(单位: 百美元)。

汽车	1	2	3	4	5	6	7	8
跑车	14.23	12.47	14.00	13.17	27.48	12.42	32.59	12.98
中型车	11.97	11.42	13.27	9.87	10.12	10.36	12.65	25.23

- 将数据绘成图以确定  $t$  检验的条件是否满足。
- 数据是否支持这样的猜想, 即, 跑车的平均毁损程度大于中型车? 在  $\alpha = 0.05$  下进行  $t$  和 Wilcoxon 两种检验。
- 哪种方法更适合于分析该数据?
- 两种方法所得的结论相同吗? 为什么?



中型车和跑车毁损程度盒形图(均值以实心圆点表示)

#### Two-Sample T-Test and Confidence Interval

Two-sample T for Midsize vs SUV

	N	Mean	StDev	SE Mean
Midsize	8	13.11	5.05	1.8
SUV	8	17.42	7.93	2.8

95 % CI for  $\mu$  Midsize- $\mu$  SUV: (-11.4, 2.8)

T-Test  $\mu$  Midsize =  $\mu$  SUV (vs <): T = -1.30 P = 0.11 DF = 14

Both use Pooled StDev = 6.65

#### Mann-Whitney Confidence Interval and Test

Midsize	N = 8	Median = 11.69
SUV	N = 8	Median = 13.59

Point estimate for ETA1-ETA2 is -2.32

95.9 Percent CI for ETA1-ETA2 is (-14.83, -0.33)

W = 48.0

Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.0203

**6.87** 参见练习 6.86。这个研究因所测试的车辆太少而受到批评。为此需要在大样本基础上重新研究。假定毁损程度总体均服从正态分布,具有共同的标准差  $\sigma = 700$  美元。

- a. 确定样本容量,以便我们能够以 95% 的把握相信平均修理费用差的估计值与真正值相差不到 500 美元。
- b. 对于研究假设  $H_a: \mu_{\text{SUV}} > \mu_{\text{MID}}$ , 确定样本容量,使得当  $\mu_{\text{SUV}} - \mu_{\text{MID}} \geq 500$  时,检验具有  $\alpha = 0.05, \beta(\mu_d) < 0.05$ 。

**6.88 (法律)** 在一个涉及许多科学问题的案件中,法官发布了关于以下统计显著性的见解备忘录,这些见解中没有使用法律术语,且脱离了上下文。但它仍然可帮助我们理解他人怎样处理关于确定统计显著性意义的问题。阅读该备忘录并对统计显著性引出的问题作一评论。

### 见解备忘录

在审判之前法庭要处理的事情是能预期到的在审判中会出现的两个与证据相关的问题。第一,必须确定法庭在接受科学证据时,其合理的统计显著性水平应是多少。

关于统计显著性,在审判过程中,只有当统计证据的置信水平达到 95% 时,才可被法庭接受。

法庭上的每个相关研究使用的置信水平至少为 95%。而且原告承认社会科学家一般使用 95% 置信水平。最后,所有法律权威达成一致,即,统计证据除非达到统计人员所要求的 95% 置信水平,否则,不被接受。所以,由于原告不能提出合理的根据来改变数学家对统计显著性检验的现有态度,因此,除非达到 95% 置信水平,否则,对任何统计证据法庭不予接受。

**6.89 (医学)** 在一个生存研究中,观察 182 位患充血性心脏衰竭的男性病人,得到他们的基线测定值。在对这些资料进行总结时,其中 88 位病人已死亡。下表列出了生存者和死亡者的基线数据。在“心率”下面列出的变量表示心脏衰竭的严重程度,每个变量左边的箭头表示症状变化的方向。

- a. 对这些基线数据进行讨论。
- b. 进行  $t$  检验时,作者作了哪些假定?

由于心肌症造成的严重慢性左心室衰竭患者的基线特征数据			
变量	死亡者 ( $n = 88$ )	生存者 ( $n = 94$ )	$t$ 检验 $p$ 值
年龄(岁)	$57 \pm 10$	$56 \pm 8$	NS

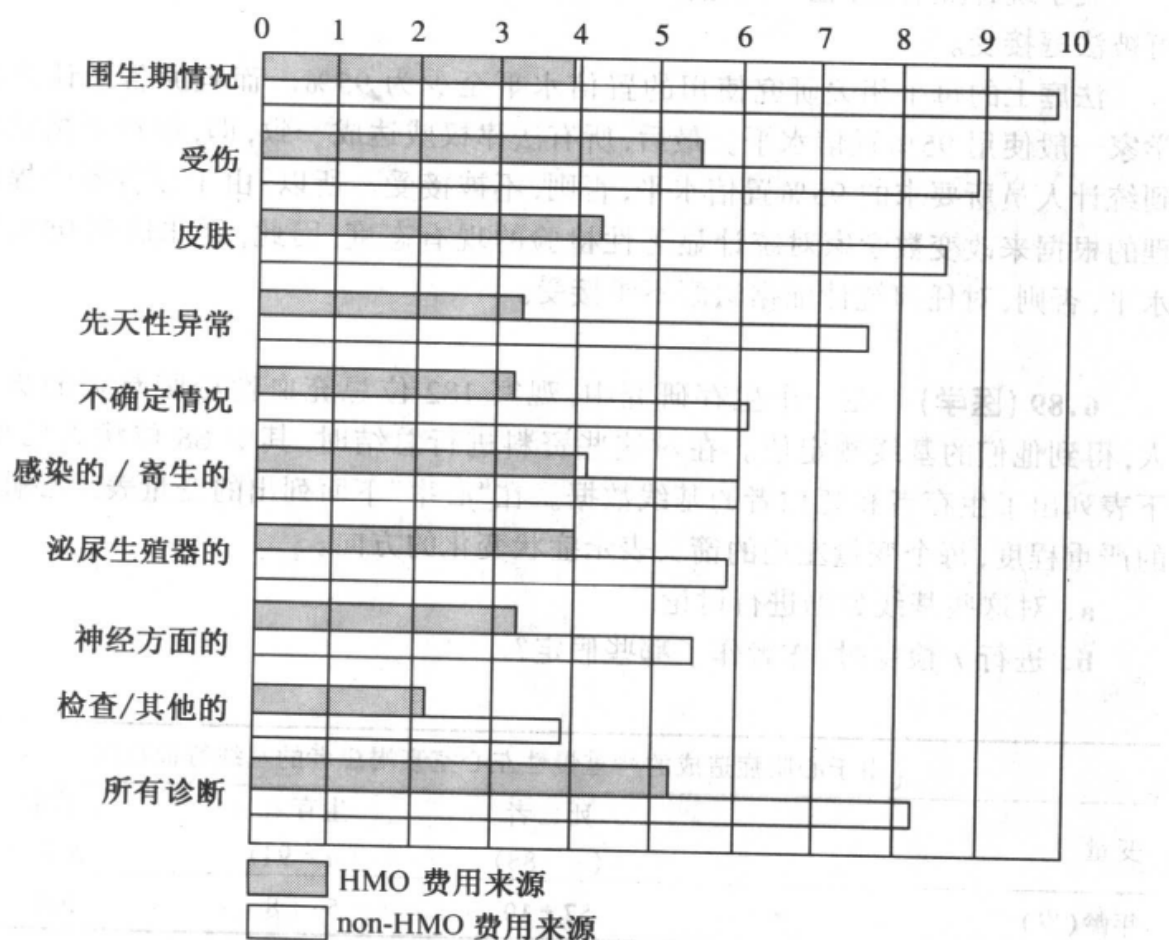
续表

由于心肌症造成的严重慢性左心室衰竭患者的基线特征数据

变量	死亡者 (n = 88)	生存者 (n = 94)	t 检验 p 值
症状持续期(月)	45 ± 43	39 ± 27	NS
心率(次/分钟)	87 ± 15	83 ± 16	NS
↓ 平均血压(毫米汞柱)	87 ± 13	94 ± 13	<0.001
↓ 左心室填充压(毫米汞柱)	29 ± 7	24 ± 9	<0.001
↑ 心脏指数[升/(分·米 <sup>2</sup> )]	2.0 ± 0.7	2.5 ± 0.8	<0.001
↑ 心搏量(毫升/搏动)	45 ± 16	59 ± 5	<0.001
↓ 系统血液阻力(单位)	25 ± 10	21 ± 8	<0.01
搏动功(克·米)	35 ± 19	56 ± 33	<0.001

表中所列的值为均值 ± 标准差

**6.90(商业)** 医院管理人员对拥有健康维护组织(HMO)费用来源的特护病人和那些拥有非健康维护组织(non-HMO)费用来源的特护病人的住院时间情况进行了研究。样本数据如下图所示。



来源: American Hospital Assn.; Twin Cities Metropolitan Health Board

- a. 你能从图中得出什么样的一般性结论? 若要作出更明确的结论, 你还需要哪些信息?
- b. 假定表中列出了所有诊断的样本统计量。用这些数据检验  $H_0: \mu_1 - \mu_2 = 0$   $H_a: \mu_1 - \mu_2 \neq 0$ , 并给出检验的  $p$  值。

	样本均值	样本容量	样本标准差
HMO	5.0	120	1.3
Non-HMO	8.1	130	1.9

**6.91** 题目见练习 6.90。进行  $t'$  检验并比较所得结果, 哪一种检验(如果有)更适合于分析这些数据?

**6.92 (医学)** 下面是一份对 10 个充血性心脏衰竭病人研究结果的摘要。阅读此摘要并试着解释这些结果。

### 摘 要

将 10 位患有充血性心脏衰竭的病人作为研究对象, 以研究某种试验性药物的疗效。分别测量了某些变量的基线数据和静脉注射此药物四小时后的数据, 数据显示此药物使心脏指数从基线平均水平  $2.41 \pm 0.49 \text{ L/min/m}^2$   $11.90\% \sim 34.3\%$  ( $p < 0.01$ ), 心率从  $72 \pm 12$  次/分钟提高了  $6\% \sim 10\%$  ( $p < 0.02$ )。使脉搏毛细血管楔压从 18.7 降低了  $15.3\% \sim 24.2\%$  ( $p < 0.001$ )。

**6.93 (医学)** 曾经对几种抗抑郁症药物进行研究, 以考察其在治疗可卡因吸食者方面的效果。最近一个研究显示作为门诊病人的 20 个可卡因吸食者在服用抗抑郁症药物的两个星期后对可卡因的渴望程度降低了且对可卡因的实际吸食量有所减少。试对此结果进行评论, 它们可信吗? 为什么?

**6.94 (教育)** 1986 年 4 月 *Australian Journal of Statistics* (30:1, 23~44) 公布了 S. R. Butler 和 H. W. Maisd 对来自非英语家庭学生的阅读和数学成绩的研究成果。原始样本为澳大利亚悉尼市七个公立学校的所有 392 个幼儿园学生。分别在试验开始的幼儿园阶段和小学 1, 2, 3 和 6 年级时期对他(她)们进行阅读和数学测验。

下表列出了最初的 392 个学生中参加了小学第六年测验的 286 个学生 ( $n = 226$  个来自英语家庭,  $n = 60$  个来自非英语家庭)的一些特征数据。

- a. 你能给出一个概括这些基线数据的更好的方法吗?
- b. 可能用了哪种检验对这些特征进行比较?
- c. 为了直接比较阅读和数学能力, 还有哪些特征可以或应该测定?
- d. 样本的损耗率对研究结果可能会有什么影响(如果有)? 记住最初 392 个

学生中有 106 个(27%)未能参加 6 年级的测验。

特 征	组	
	英语家庭 ( $n = 226$ ) $\bar{y}$	非英语家庭 ( $n = 60$ ) $\bar{y}$
年龄(月)	67.17	67.15
性别(1=男,2=女)	1.50	1.55
家庭中孩子数	2.54	2.62
在家庭中的排位 (1=最小的孩子,等)	1.89	1.82
父亲的职业 (1=技术性最强, 17=技术性最差)	8.26 *	11.50
词汇量智商	99.26 *	74.45

\* 统计显著,  $p < 0.01$

**6.95** 参考数据盘中的临床试验数据库。根据 HAM-D 总分数数据检验假设  $H_0: \mu_D - \mu_A = 0$   $H_a: \mu_D - \mu_A > 0$ ; 即, 我们想知道在研究结束之时, 安慰剂组(D)的平均抑郁分数是否高于 A 治疗组, 给定  $\alpha = 0.05$ 。你的结论是什么?

**6.96** 题目见练习 6.95, 重复地将安慰剂组和 B 治疗组, C 治疗组进行比较。给出每个检验的  $p$  值。三个治疗组(A, B 或 C)中, 哪一组的 HAM-D 总分数最低?

**6.97** 根据 HAM-D 焦虑分数数据构造  $\mu_D - \mu_A$  的 95% 置信区间。由此区间你可得出关于  $\mu_D - \mu_A$  的什么结论?

**6.98** 参考数据盘中的临床试验数据库。用双侧统计检验比较 B 治疗组与 C 治疗组的平均年龄。对于所有检验, 检验水平  $\alpha = 0.05$ , 试给出结论。当研究不同药品对抑郁症的疗效时, 为什么不同治疗组的病人年龄相仿这一点很重要?

**6.99** 题目见练习 6.98, 为了对用于治疗抑郁症的药物的疗效作出结论, 还应该在几个处理组间比较其他哪些变量?

## 第七章 关于总体方差的推断

- 7.1 引言和案例
- 7.2 单个总体方差的估计和检验
- 7.3 比较两个总体方差时的估计和检验
- 7.4 比较多个总体方差时的检验
- 7.5 小结

### 7.1 引言和案例

当人们考虑统计推断时,他们通常想到的是总体均值的推断。但是,针对试验者的实际问题的总体参数可能随情况而不同。在许多情况下,总体值的变异性与总体均值同样重要。在涉及产品改进问题的案例中,产品质量被定义为产品均值等于目标值且围绕均值的变异较小。比如说,药品的生产者肯定要关心如何控制药片的平均药效,但他或她一定也担心药片之间在药效上的变异。过量的药效和药效不足对病人来说可能都是有害的。因此,厂商愿意生产的药片具有期望的平均药效并且药片之间药效的差异(用  $\sigma$  或  $\sigma^2$  来度量)尽可能地小。另外一个例子来自投资策略领域。投资者要寻找具有较低风险的股票,证券,不动产和其他投资的资产组合。投资者通常用一定时期内投资价值的方差来度量某一特定投资组合的不确定性。具有高平均值和较大标准差的投资组合获得的收益往往比平均值低很多。因此投资者在决定投资组合的风险度时,需要考察投资价值围绕其平均值的变化情况。

#### 案例:关于检测大肠杆菌方法的评估

最近几年由于消费被感染的肉产品而引起的细菌疾病的爆发要求在肉类食品监管程序上使用更新更快捷的病原体检测方法。文章“Repeatability of the petri-film HEC test and agreement with a hydrophobic grid membrane filtration method for the enumeration of *Escherichia coli* 0157:H7 on beef carcasses,”[*Journal of Food Protection*(1998)61:402~408]将一种新的大肠杆菌细菌检测法——HEC法和HGMF法进行对比。相对于传统方法,HEC检测法更易于接种,更便于培养,且操作起来更安全。但是,研究人员必须比较HEC法和HGMF法的效用以决定HEC是否是一种用于检测大肠杆菌的可行的方法。

**数据收集的设计** HEC方法的开发者寻找以下问题的答案:



1. 与 HEC 和 HGMP 读数有关的什么参数必须要进行对比?
2. 为了有效地比较 HEC 和 HGMP 法,必须要多少个观察值?
3. 进行什么样的试验设计可最有效地比较 HEC 和 HGMP 法?
4. 什么是用于比较的有效统计方法?
5. 在关于 HEC 和 HGMP 的最终评价报告中应包括什么类型的信息?

对研究者来说,大肠杆菌计数的哪些方面是有趣的? 仅仅对比平均浓度就会知道两种方法在大量检测的平均读数上是否一致。但是,我们并不知道在大肠杆菌检测过程中 HEC 法的读数是否比 HGMP 法更易变化。例如,考虑图 7.1 的两个分布。假定它们分别代表利用 HEC 和 HGMP 法检测大肠杆菌浓度所得的测量值的分布,而真正的大肠杆菌浓度为  $7 \log_{10} \text{CFU/ml}$ 。分布表明:对于给定的肉类样本,HEC 法的检测值与大肠杆菌的真实值相差较大,而 HGMP 法的检测值可能更接近真实值。在这种情况下,关键是既要比较两种方法的均值又要比较它们的标准差。事实上,为了评价这两种方法,我们需要考察 HEC 和 HGMP 检测值之间关系的其他方面。我们将在关于建立模型和进行方差分析的章节中讨论这一思想。

试验设计分为两个阶段,阶段 1 将两种方法应用于纯培养出的大肠杆菌,标明浓度为  $10^7 \text{CFU/ml}$  的 E318N 品种。基于估计大肠杆菌水平(见练习 7.4)的给定

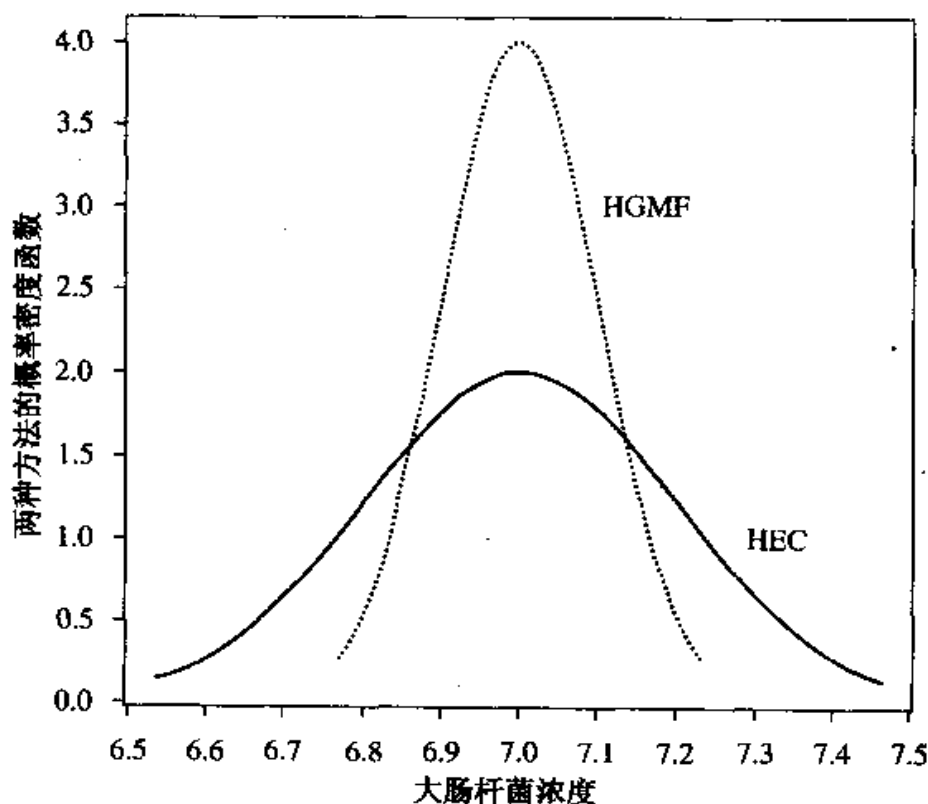


图 7.1 假设的根据 HEC 和 HGMP 法大肠杆菌浓度的分布

精度,决定将 HEC 和 HGMF 方法应用于 24 个纯培养细菌。将大肠杆菌浓度数据进行变换( $\log_{10}\text{CFU/ml}$ ),表 7.1 列出了变换后的检测值(表中的值是利用文章中给出的概括统计量得出的模拟值)。

表 7.1 HEC 和 HGMF 的大肠杆菌读数( $\log_{10}\text{CFU/ml}$ )

样本	HGMF	HEC	样本	HGMF	HEC	样本	HGMF	HEC
1	6.65	6.67	9	6.89	7.08	17	7.07	7.25
2	6.62	6.75	10	6.90	7.09	18	7.09	7.28
3	6.68	6.83	11	6.92	7.09	19	7.11	7.34
4	6.71	6.87	12	6.93	7.11	20	7.12	7.37
5	6.77	6.95	13	6.94	7.11	21	7.16	7.39
6	6.79	6.98	14	7.03	7.14	22	7.28	7.45
7	6.79	7.03	15	7.05	7.14	23	7.29	7.58
8	6.81	7.05	16	7.06	7.23	24	7.30	7.54

阶段 2 将两种方法应用于人工污染的牛肉上。部分牛肉块取自于三头经过检验为大肠杆菌阴性的 Holstein 母牛。从这些牛身上共取出 18 份牛肉块,然后将它们感染上大肠杆菌。分别将 HEC 和 HGMF 方法应用于这 18 份样本牛肉中的每一份。对两种方法得到的大肠杆菌浓度数据进行变换( $\log_{10}\text{CFU/ml}$ )。本例中共得到 18 对样本数据。研究人员欲建立模型将两种方法得到的大肠杆菌浓度检测值联系起来。本章我们只讨论阶段 1,阶段 2 将在十一章讨论。

**数据整理** 下一步,研究人员要按照第 2.5 节介绍的步骤准备数据,以便进行统计分析。

**数据分析** 研究人员感兴趣的是两种方法得到的大肠杆菌浓度检测值是否一致。图 7.2 给出了试验数据的盒形图。就盒形图的宽度和须的长度而言,两种方法所得的结果非常相似,但 HEC 的中位数比 HGMF 大。样本概括统计量如下。

#### Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev	SE Mean
HEC	24	7.1346	7.1100	7.1373	0.2291	0.0468
HGMF	24	6.9529	6.9350	6.9550	0.2096	0.0428
Variable	Minimum	Maximum	Q1	Q3		
HEC	6.6700	7.5400	6.9925	7.3250		
HGMF	6.5600	7.3000	6.7900	7.1050		

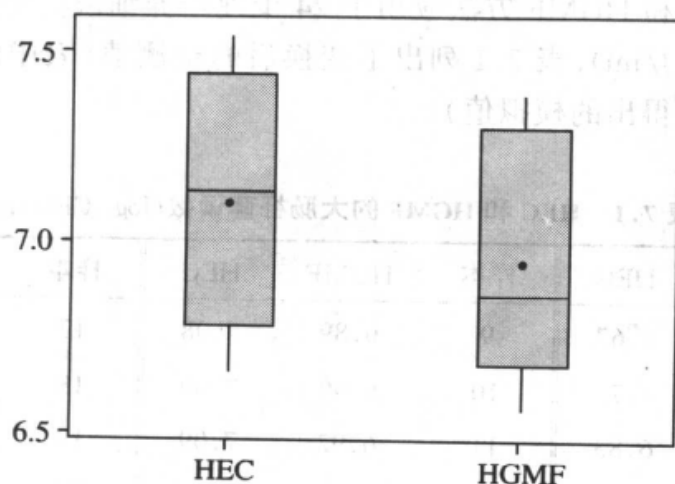


图 7.2 HEC 和 HGMF 盒形图  
(均值以实心圆点表示)

从概括统计量中,我们可以看到利用 HEC 法得到的平均浓度比 HGMF 法大。并且,在浓度读数的变异性上 HEC 比 HGMF 稍大。我们初步认为两种方法得到的大肠杆菌浓度读数的分布不同。但是我们需要决定是否可以根据样本均值和标准差的不同来推断总体相应值之间也存在差异。在介绍了用以比较总体方差的适当方法以后,我们将给出相应的结果。

总体方差的推断问题类似于总体均值的推断。首先,我们必须根据随机样本数据构造点估计、置信区间和检验统计量以对总体值的变异性进行推断。然后,我们就能说我们能够以多大的把握相信观察到的样本数据之间的差异就代表了相应总体参数之间的差异。

## 7.2 单个总体方差的估计和检验

### 样本方差

$$s^2 = \frac{\sum (y - \bar{y})^2}{n - 1}$$

可以用来推断总体方差  $\sigma^2$ 。对于来自均值为  $\mu$ , 方差为  $\sigma^2$  总体的一个容量为  $n$  的随机样本,  $s^2$  是  $\sigma^2$  的**无偏估计量**。如果总体分布为正态, 那么  $s^2$  的抽样分布可以确定如下。根据容量为  $n$  且取自于方差为  $\sigma^2$  的正态总体的重复样本数据, 我们可以计算统计量  $(n-1)s^2/\sigma^2$ , 并画出直方图。直方图的形状与图 7.3 相似, 因为可以证明统计量  $(n-1)s^2/\sigma^2$  服从**自由度  $df = n - 1$  的卡方分布**。卡方 ( $\chi^2$ , 其中  $\chi$  是希腊字母 chi) 概率分布的数学表达式很复杂, 因此我们并不将它写出来。但

是,我们将卡方分布的有关性质列出如下:

1.  $\chi^2$  分布为正的偏态分布,其取值在 0 到  $\infty$  之间(见图 7.3)。
2.  $\chi^2$  分布很多,分布的参数为自由度(df)。图 7.3 给出了自由度分别为 5, 15 和 30 的三种  $\chi^2$  分布。
3.  $\chi^2$  分布的均值和方差分别为  $\mu = df, \sigma^2 = (2)df$ 。例如,如果  $\chi^2$  分布的自由度  $df = 30$ ,那么分布的均值和方差为  $\mu = 30, \sigma^2 = 60$ 。

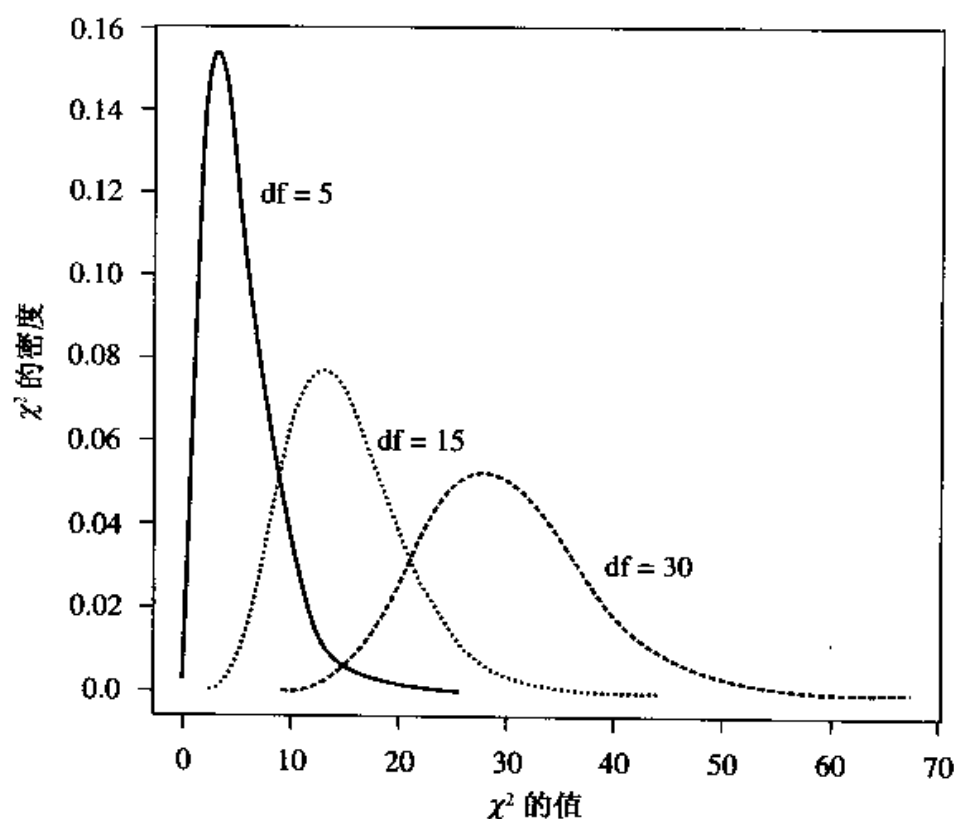


图 7.3  $\chi^2$  ( $df = 5, 15, 30$ ) 分布的概率密度曲线

在附表 7 中可查到  $\chi^2$  分布的上分位数。表中的每一个数均表示  $\chi^2$  值,其值右侧曲线下方的面积为  $\alpha$ 。表左栏为自由度,顶部为  $\alpha$  值。因此,当  $df = 14$  时,对应于曲线下方右边面积为  $\alpha = 0.025$  的  $\chi^2$  值是 26.12。(见图 7.4)。为了确定曲线下方左侧面积为 0.025 对应的  $\chi^2$  值,我们计算  $\alpha = 1 - 0.025$ ,从附表 7 查得  $\chi^2$  值等于 5.629。利用这两个值,我们得到曲线下方介于 5.629 和 26.12 之间的面积是  $1 - 0.025 - 0.025 = 0.95$ (见图 7.4)。根据这一点,我们就能构造  $\sigma^2$  的置信区间。因为  $\chi^2$  分布不对称,所以基于这一分布的置信区间也不具有通常的形式,即没有我们看到的  $\mu$  和  $\mu_1 - \mu_2$  置信区间所具有的形式:估计值  $\pm$  误差,  $\sigma^2$  的  $100(1 - \alpha)\%$  置信区间可通过用  $\sigma^2$  的估计量  $s^2$  除以下  $\alpha/2$  分位数  $\chi^2_L$  和上  $\alpha/2$

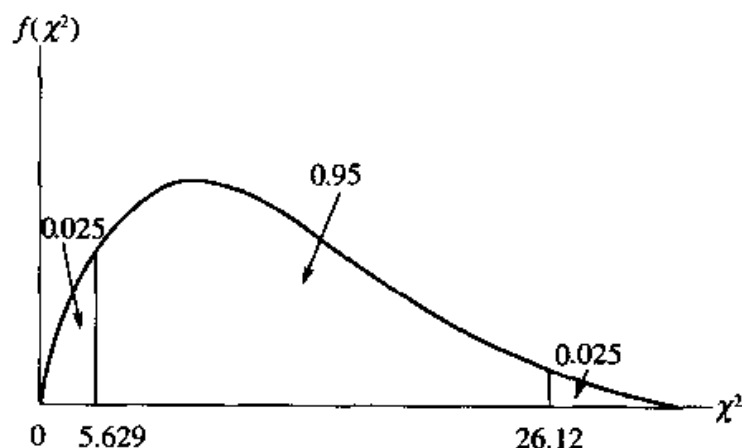


图 7.4 自由度  $df=14$  的  $\chi^2$  分布临界值

分位数  $\chi^2_U$  而得到。

$\sigma^2$  (或  $\sigma$ ) 的置信系数  $1-\alpha$  的置信区间

$$\frac{(n-1)s^2}{\chi^2_U} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}$$

其中  $\chi^2_U$  为自由度  $df=n-1$  的  $\chi^2$  分布下, 右侧面积等于  $\alpha/2$  的上尾值,  $\chi^2_L$  为  $\chi^2$  分布的左侧面积等于  $\alpha/2$  的下尾值 (见图 7.5)。对于给定的自由度  $df$ , 我们可以通过查附表 7 中对应于  $\alpha = \alpha/2$  和  $\alpha = 1 - \alpha/2$  的临界值得到  $\chi^2_U$  和  $\chi^2_L$ 。(注意: 通过开平方根可得到  $\sigma$  的置信区间。)

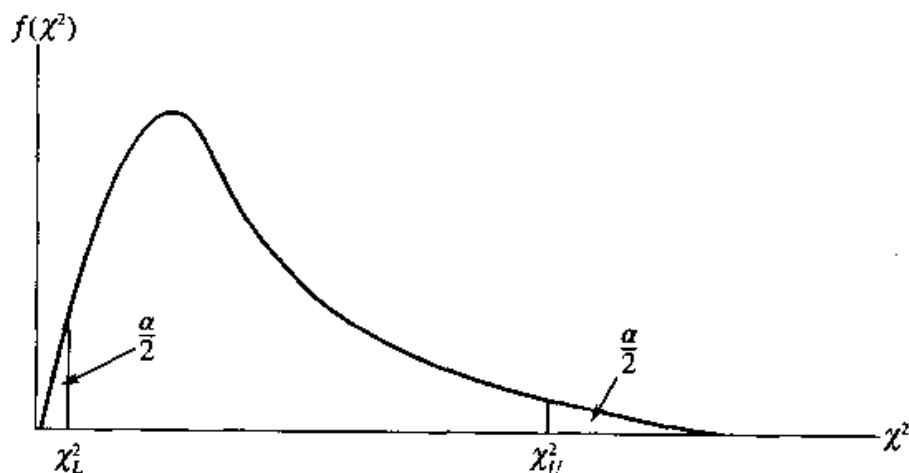


图 7.5  $\chi^2$  分布的上尾值和下尾值

### 例 7.1

某大型食品厂的质量控制部门对其咖啡装填机器进行监控, 该机器欲将每个

容器装满 500 克咖啡。理想情况下,每个容器内的咖啡量应该围绕着额定值 500 克略有波动。如果波动很大,那么大部分的容器可能因没有装满而欺骗了顾客,或者因装得过量而导致公司的经济损失。设计机器使得容量 500 克的容器内所装咖啡的重量服从均值为 506.6 克,标准差为 4 克的正态分布。这将产生一个容器总体,使之最多有 5% 的容器内咖啡的重量少于 500 克。为了保证至多有 5% 的容器分量不足,每小时抽取 30 个容器作为随机样本。根据这些数据来确定均值和标准差是否维持在额定值上。其中的一个一小时样本数据如下。

501.4 498.0 498.6 499.2 495.2 501.4 509.5 494.9 498.6 497.6  
505.5 505.1 499.8 502.4 497.0 504.3 499.7 497.9 496.5 498.9  
504.9 503.2 503.0 502.6 496.8 498.2 500.1 497.9 502.2 503.2

试估计对于该小时内装填的所有容器,其咖啡重量均值和标准差的 99% 置信区间。

**解答** 由样本数据得:  $\bar{y} = 500.453$ ,  $s = 3.433$

为了利用我们的方法建立  $\mu$  和  $\sigma$  的置信区间,首先,我们必须检查数据是否是来自正态总体的一个随机样本。图 7.6 为 30 个重量数据的正态概率图。这 30 个值几乎落在一条直线上。因此正态性条件得到满足。本例的置信系数  $1 - \alpha = 0.99$ 。对于  $df = n - 1 = 29$ ,  $\alpha = \alpha/2 = 0.005$  查附表 7 得到上尾  $\chi^2$  值,同样地,对于  $\alpha = 1 - \alpha/2 = 0.995$  查附表 7 得到下尾  $\chi^2$  值。因此

$$\chi_L^2 = 13.12, \chi_U^2 = 52.34。$$

那么,  $\sigma$  的 99% 置信区间为

$$\sqrt{\frac{29(3.433)^2}{52.34}} < \sigma < \sqrt{\frac{29(3.433)^2}{13.12}}$$

或

$$2.56 < \sigma < 5.10$$

因此,我们有 99% 的把握相信咖啡重量的标准差落在 2.56 与 5.10 克之间。从而,  $\sigma$  的设计值, 4 克,落在我们的置信区间以内。根据第五章的结果,  $\mu$  的 99% 置信区间为

$$500.453 \pm 2.756 \frac{3.433}{\sqrt{30}} \quad 500.453 \pm 1.73$$

或

$$498.7 < \mu < 502.2$$

因此,这台机器似乎使得容器装填不足,因为 506.6 克并没有落在置信区间以内。

除了估计总体方差以外,我们还能对  $\sigma^2$  等于给定值  $\sigma_0^2$  的零假设进行统计检验。检验步骤如下。

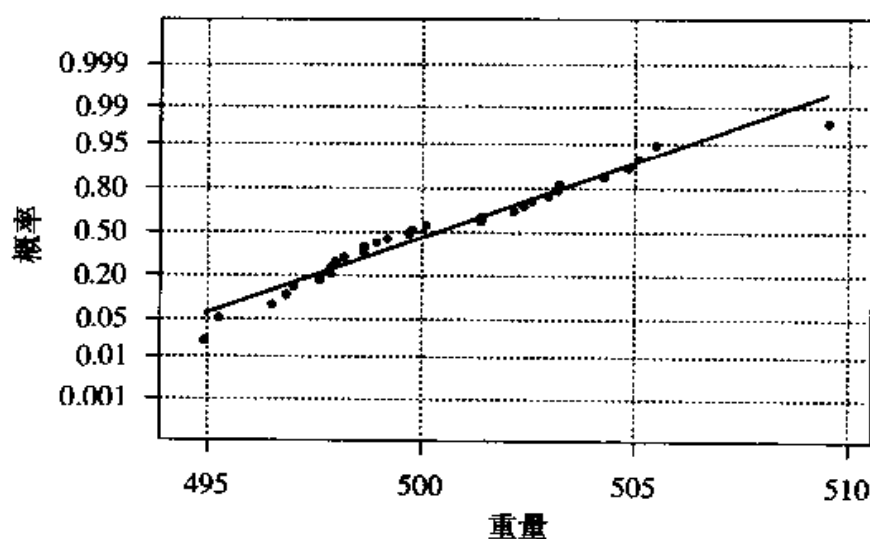


图 7.6 容器内重量的正态概率图

$\sigma^2$  (或  $\sigma$ ) 的统计检验:

$$H_0: 1. \sigma^2 \leq \sigma_0^2$$

$$H_a: 1. \sigma^2 > \sigma_0^2$$

$$2. \sigma^2 \geq \sigma_0^2$$

$$2. \sigma^2 < \sigma_0^2$$

$$3. \sigma^2 = \sigma_0^2$$

$$3. \sigma^2 \neq \sigma_0^2$$

$$T. S.: \chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

R. R.: 对于给定的值  $\alpha$ ,

1. 如果  $\chi^2$  大于  $\chi_U^2$ , 则拒绝  $H_0$ , 其中  $\chi_U^2$  为对应于  $\alpha = \alpha$ ,  $df = n - 1$  的上尾  $\chi^2$  值。
2. 如果  $\chi^2$  小于  $\chi_L^2$ , 则拒绝  $H_0$ , 其中  $\chi_L^2$  为对应于  $\alpha = 1 - \alpha$ ,  $df = n - 1$  的下尾  $\chi^2$  值。
3. 如果  $\chi^2$  大于  $\chi_U^2$  (对应于  $\alpha = \alpha/2$ ,  $df = n - 1$ ), 或小于  $\chi_L^2$  (对应于  $\alpha = 1 - \alpha/2$ ,  $df = n - 1$ ), 则拒绝  $H_0$ 。

### 例 7.2

某种控制屋内臭虫的杀虫剂的生产商声称他的产品能将其大部分药效至少保持 6 个月, 并且声称从 0 到 6 个月, 其产品的药效下降幅度在 0% 到 8% 之间。为了检验这一声称, 一消费者组织从生产商那里获得了 20 罐杀虫剂随机样本。先检验每一罐的药效, 然后在室温下储存 6 个月。储存期过后, 再检验每一罐的药效。记录每罐药效下降的百分比, 数据如下。

0.5 3.5 4.4 6.0 6.6 5.4 7.9 4.6 5.4 5.7

2.5 1.1 5.9 2.7 2.3 1.4 1.8 5.8 0.2 7.1

利用这些数据确定是否有充足的证据表明药效下降百分比总体比生产商声称的有更大的变异性。给定  $\alpha = 0.05$ 。

**解答** 生产商声称药效下降百分比总体的极差为 8。将极差除以 4, 我们便得到总体标准差的近似值  $\sigma = 2$  (或  $\sigma^2 = 4$ )。

零假设和备择假设为:

$H_0: \sigma^2 \leq 4$  (即, 假设生产商的声称正确)

$H_a: \sigma^2 > 4$  (即, 比生产商声称的有更大的变异性)

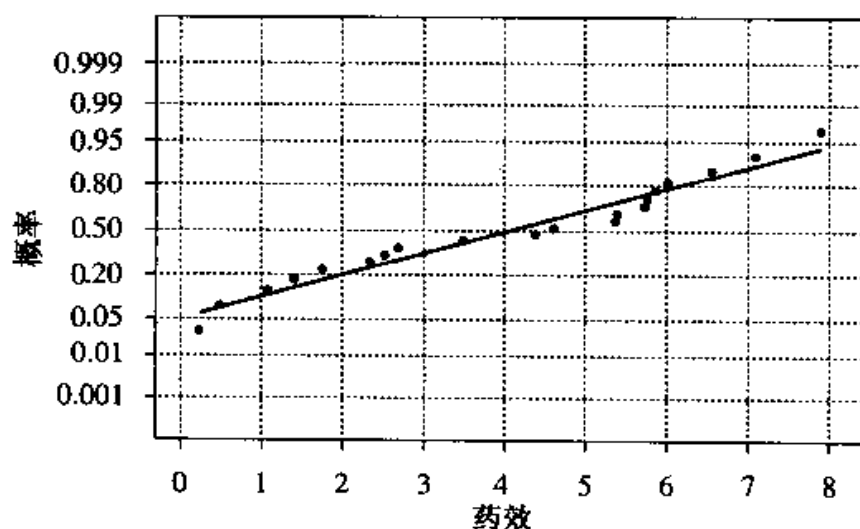


图 7.7 药效数据的正态概率图

为了利用我们关于方差的推断方法, 首先, 我们必须检查数据的正态性。从图 7.7 我们看到, 描的点大致落在一条直线上, 因此正态性条件似乎满足。根据这 20 个数据, 我们得到  $s^2 = 5.45$ 。检验统计量和拒绝域如下:

$$T. S. (\text{检验统计量}): \chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{19(5.45)}{4} = 25.88$$

R. R. (拒绝域): 对于  $\alpha = 0.05$ , 如果  $\chi^2$  的计算值大于 30.14, 则拒绝  $H_0$ , 其中 30.14 为附表 7 中对应于  $\alpha = 0.05$ ,  $df = 19$  的  $\chi^2$  临界值。

结论: 因为  $\chi^2$  的计算值 25.88 小于临界值 30.14, 所以在  $\alpha = 0.05$  水平下, 我们并没有充足的证据拒绝生产商的声称。利用计算机程序可以得出  $p$  值  $= P(\chi_{19}^2 > 25.88) = 0.14$ 。利用附表 7, 我们仅能得出  $p$  值  $> 0.10$ , 因  $p$  值  $= P(\chi_{19}^2 > 25.88) > P(\chi_{19}^2 > 27.20) = 0.10$ 。由于样本容量相对较小, 且  $p$  值也不是很大, 因此消费者并不准备接受  $H_0: \sigma^2 \leq 4$ 。更为明智的选择就是在作出明确的结论之前



利用大样本再作一次补充的检验。

关于  $\sigma$  的推断方法是基于这样一个条件, 即, 随机样本来自于一个正态总体, 这与基于  $t$  分布的推断方法相类似。但是, 当样本容量适当大时 ( $n \geq 30$ ), 即使正态性条件不满足, 也可利用基于  $t$  分布的方法来对  $\mu$  进行统计推断, 因为对于相对大的样本容量来说, 由中心极限定理可知样本均值近似服从正态分布。不幸的是, 在用基于  $\chi^2$  分布的方法对  $\sigma$  进行推断时, 同样的结论并不能成立; 也就是说, 如果总体分布明显偏离正态, 那么, 即使样本容量很大, 用这些方法对  $\sigma$  进行推断也是不合适的。总体的非正态性, 或是偏态, 或是厚尾, 对名义显著性水平和  $\sigma$  的置信概率都有严重影响。如果样本数据的盒形图和正态概率图显示严重偏态或者发现大量异常值, 那么就不能应用基于  $\chi^2$  分布的推断方法。此时可利用另外一些推断方法, 这些方法在计算上进行了精心的设计。其中一种方法就是自助 (bootstrap) 法。自助法为估计像样本方差这样的样本统计量的不确定性提供了一种简单、实用的方法。我们可以利用自助方法估计样本方差的抽样分布。然后, 根据估计出的抽样分布产生  $\sigma$  的置信区间和关于  $\sigma$  假设检验的拒绝域。关于自助法的有关内容可参看两本书, 即, *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993 和 *Randomization, Bootstrap and Monte Carlo Methods in Biology*, Chapman and Hall, New York, 1998, 前者由 Efron 和 Tibshirani 撰写, 后者由 Manly 所著。

### 例 7.3

为了解当样本来自于厚尾分布或偏态分布而不是来自于所需的正态分布时  $\chi^2$  检验水平所受到的影响, 进行了一项模拟研究。研究中使用了五种分布, 即正态分布, 均匀分布 (短尾分布), 自由度为 5 的  $t$  分布的 (厚尾分布), 两个伽马 ( $\Gamma$ , Gamma) 分布, 其中一个轻度偏态, 另一个严重偏态。表 7.2 列出了以上分布的一些统计量。

表 7.2 模拟中各个分布的概括统计量

统计量	分 布				
	正态	均匀	$t(df=5)$	伽马(形状=1)	伽马(形状=0.1)
均值	0	17.32	0	10	3.162
方差	100	100	100	100	100
偏度	0	0	0	2	6.32
峰度	3	1.8	9	9	63

注意到每一个分布都具有同样的方差  $\sigma^2 = 100$ , 但是分布的偏度和峰度不同。从

每一个分布中,选择样本容量分别为 10, 20 和 50 的 2500 个随机样本。给定  $\alpha = 0.05$  时,检验  $H_0: \sigma^2 \leq 100$   $H_a: \sigma^2 > 100$  和  $H_0: \sigma^2 \geq 100$   $H_a: \sigma^2 < 100$ 。对于来自 5 个分布中的每一个分布的不同样本容量的每 2500 个样本均进行方差的卡方检验。结果列于表 7.3, 这些结果显示了非正态性样本对检验的灵敏性吗?

表 7.3  $H_0$  被拒绝次数的比例( $\alpha = 0.05$ )

样本容量	$H_a: \sigma^2 > 100$				
	分 布				
	正态	均匀	$t$	$\Gamma(1)$	$\Gamma(0.1)$
$n = 10$	0.047	0.004	0.083	0.134	0.139
$n = 20$	0.052	0.006	0.103	0.139	0.175
$n = 50$	0.049	0.004	0.122	0.156	0.226

样本容量	$H_a: \sigma^2 < 100$				
	分 布				
	正态	均匀	$t$	伽马(1)	伽马(0.1)
$n = 10$	0.046	0.018	0.119	0.202	0.213
$n = 20$	0.050	0.011	0.140	0.213	0.578
$n = 50$	0.051	0.018	0.157	0.220	0.528

**解答** 表 7.3 的值是关于方差的卡方检验的犯第 I 类错误概率  $\alpha$  的估计。当样本来自于正态总体时,犯第 I 类错误的实际概率非常接近于名义概率  $\alpha = 0.05$ 。当总体分布为对称但比正态分布更短尾时,实际概率比 0.05 要小,但是对于对称厚尾分布,犯第 I 类错误的实际概率比 0.05 要大很多。并且,对于两个偏态分布,实际  $\alpha$  值比 0.05 要大很多。而且总体分布越偏态,实际  $\alpha$  值与 0.05 的偏差越大。从上面这些结果中可以看出,有很强证据表明总体方差的卡方检验的  $\alpha$  值对于非正态性是很敏感的。这就更加使得我们推荐:在进行总体方差的卡方检验之前要检验数据的正态性。

### 练习

#### 基本技能

7.1 假设  $Y$  服从自由度为 27 的  $\chi^2$  分布

- 求  $P(Y > 46.96)$ 。
- 求  $p(Y > 18.81)$ 。
- 求  $P(Y < 12.88)$ 。

d.  $P(12.88 < Y < 46.96)$  等于多少?

7.2 对于自由度为 11 的  $\chi^2$  分布,

a. 求  $\chi_{0.025}^2$ .

b. 求  $\chi_{0.975}^2$ .

7.3 对于很多自由度  $df$ , 利用附表 7 我们可以查到  $\chi^2$  分布的分位数。但是, 当所需自由度没有列在表中且自由度大于 40 时, 我们可以用下面的公式求出相应分位数的近似值。

$$\chi_a^2 \approx v \left( 1 - \frac{2}{9v} + z_a \sqrt{\frac{2}{9v}} \right)^3$$

其中  $\chi_a^2$  是  $df = v$  的  $\chi^2$  分布的上分位数,  $z_a$  是标准正态分布的上分位数。

a. 对于自由度为 80 的  $\chi^2$  分布, 将表 7 中  $\chi_{0.025}^2$  和  $\chi_{0.975}^2$  的实际值与其近似值进行对比。

b. 假定  $Y$  服从自由度为 277 的  $\chi^2$  分布, 求  $\chi_{0.025}^2$  和  $\chi_{0.975}^2$  的近似值。

## 应用

7.4 (生物) 在关于评估用于检出大肠杆菌的方法的案例分析中, 使用了 24 个培养细菌作为随机样本。证明这是一个合适的样本容量, 它满足如下条件: 即, 研究者需要一个样本容量, 使其能以 95% 置信度相信样本均值在实际 HGMF 均值的 0.1 个单位以内。根据过去对 HGMF 方法的经验, 标准差近似等于 0.25 个单位。

7.5 (工程) 包装线上放着名义上为 32 盎司番茄汁的罐子, 这些罐子内果汁数量满足均值为 32.30 盎司的正态分布, 工序的标准差应该小于 0.15 盎司。(标准差大会导致太多罐子内的番茄汁不足且有太多罐子过满。)为了评价该工序, 每小时随机抽取 50 罐。对其中一个样本的数据概括分析如下, 且画出图 7.8。

Descriptive Statistics for Juice Data

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Juice Jars	50	32.267	32.248	32.270	0.135	0.019
Variable	Minimum	Maximum	Q1	Q3		
Juice Jars	31.874	32.515	32.177	32.376		

a. 如果工序加工出的每罐番茄汁服从满足均值为 32.30 盎司, 标准差为 0.15 盎司的正态分布, 那么在包装线上不满的罐子比例是多少?

b. 图 7.8 是否表明利用卡方方法构造  $\sigma$  的置信区间和对  $\sigma$  进行假设检验所需的条件被违反了?

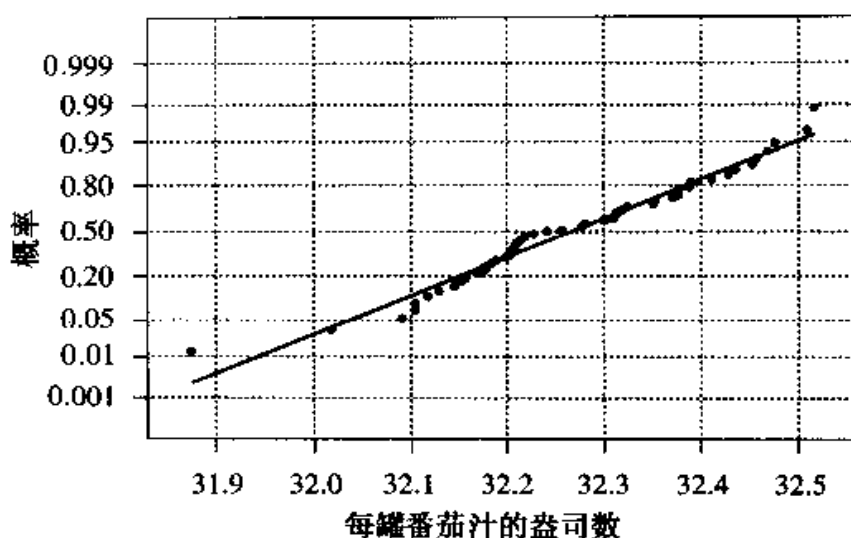
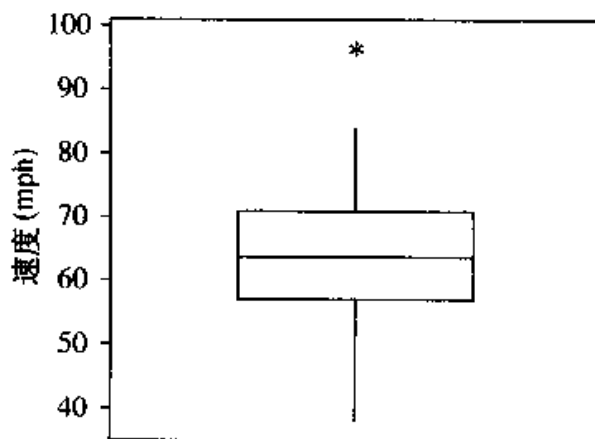


图 7.8 番茄汁数据的正态概率图

- c. 构造工序标准差的 95% 置信区间。
- d. 数据是否表明工序标准差大于 0.15? 取  $\alpha = 0.05$ 。
- e. 求检验的  $p$  值的界限。

**7.6 (工程)** 一位研究人员在研究州际高速公路上发生的交通事故时指出州际高速公路上发生的许多撞车事故, 其主要原因并不是汽车的速度和汽车速度之间的差别。当一些汽车速度很慢而另一些汽车以极限速度快速行驶时, 快行的汽车必须迅速改变车道, 这就增加了事故发生的可能性。因此, 在州际高速公路上的某个地点, 与行驶速度更一致时相比, 当汽车行驶速度差别很大时, 事故发生的次数会更多。研究人员相信当汽车速度标准差超过 10mph(英里/小时)时, 事故发生率会大大增加。在一个小时的时间段内, 从一段事故发生率较高的州际公路上随机选择了 100 辆汽车, 用雷达枪记录它们的速度。将数据总结如下, 并画出盒形



图。

Descriptive Statistics for Vehicle Speeds

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Speed(mph)	100	64.48	64.20	64.46	11.35	1.13
Variable	Minimum	Maximum	Q1	Q3		
Speed(mph)	37.85	96.51	57.42	71.05		

- 盒形图是否表明利用卡方方法构造  $\sigma$  的置信区间和对  $\sigma$  进行假设检验所需的条件被违反了?
- 用 95% 置信区间估计汽车速度的标准差。
- 数据是否以 5% 的水平表明汽车行驶速度的标准差超过 10mph?

**7.7 (工程)** 一小型装置上某零件的直径应该为 4.000mm, 且说明书允许最大标准差为 0.011mm。26 个零件直径的随机样本数据如下:

3.952 3.978 3.979 3.984 3.987 3.991 3.995 3.997 3.999 3.999 3.999  
 4.000 4.000 4.000 4.001 4.001 4.002 4.002 4.003 4.004 4.006 4.009  
 4.010 4.012 4.023 4.041

- 计算样本均值和标准差。
- 研究假设  $\sigma > 0.011$  能够在  $\alpha = 0.05$  的水平上被数据支持吗? 叙述统计假设检验的所有部分。

**7.8** 利用练习 7.7 的数据, 分别计算实际方差和实际标准差的 90% 置信区间。

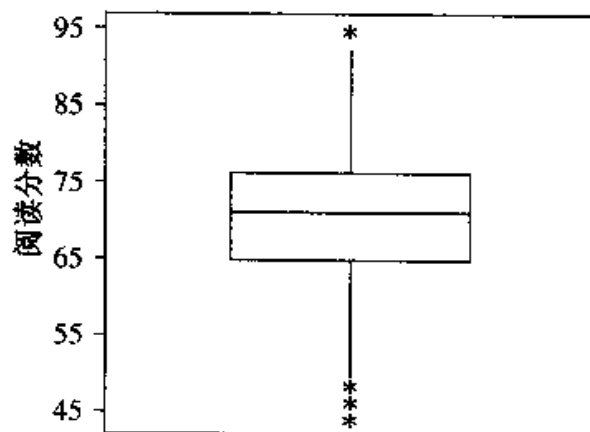
**7.9** 对练习 7.7 的数据作图。图是否表明你在练习 7.7 和 7.8 中答案所基于的假设被违背了吗? 这样的违背对于答案的有效性有严重影响吗?

**7.10 (教育)** 某一大型公众教育系统正在评估它的小学阅读计划。特别地, 教育工作者对学生参加面向州内所有三年级学生的标准化阅读考试所取得的分数感兴趣。将他们的平均分数与州的平均分数进行对比可确定该教育系统的等级。同时, 教育工作者也关心分数之间的差异。如果平均分数可以接受, 但分数之间的差异较大, 这表明有很大一部分学生将需要纠正阅读计划。同时, 分数差异这样大也表明需要设计能满足有才华学生需要的阅读计划。如果没有加速的阅读计划, 这些学生会对阅读课失去兴趣。为了获得学生在学年早期的信息(在学年的最后一个月进行全州范围的考试), 随机抽取 150 名三年级学生参加考试, 试卷为上学年的考试试卷。考试分数在 0 到 100 之间, 将数据描述如下。

Descriptive Statistics for Reading Scores

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Reading	150	70.571	71.226	70.514	9.537	0.779
Variable	Minimum	Maximum	Q1	Q3		
Reading	44.509	94.570	65.085	76.144		

- 数据图是否表明利用卡方方法构造  $\sigma$  的置信区间和对  $\sigma$  进行假设检验所需的条件被违反了?
- 用 99% 置信区间估计阅读分数的方差。
- 数据是否表明阅读分数的方差大于 90, 其中 90 为所有参加上学年考试学生的阅读分数的方差?



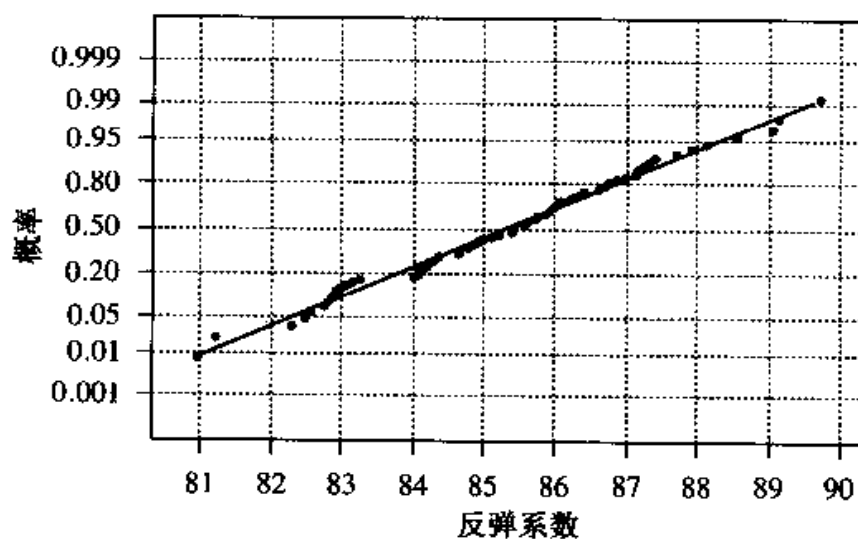
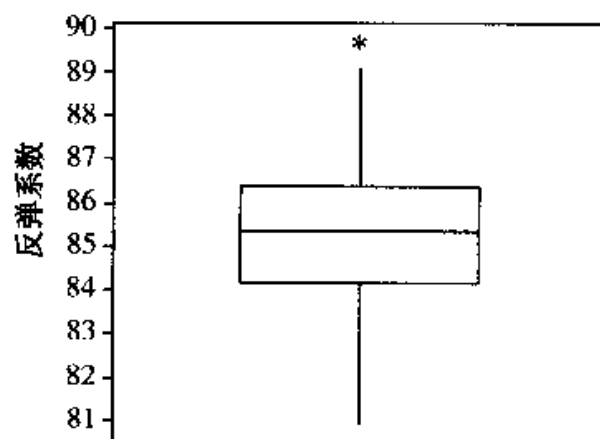
7.11 给出练习 7.10 中检验  $p$  值的界限。

7.12 (工程) 棒球之间的反弹系数往往有所差别。当用同样的力作用于棒球时, 反弹系数较大的棒球比反弹系数较小的棒球飞得远。为了获得一场比赛, 其中每一个击球手都有同样的机会击中本垒打, 球应该具有同样的反弹系数。现已设计出一种标准检测来测量棒球的反弹系数。一买主欲购买很多棒球, 他要求平均反弹系数为 85 单位且标准差小于比 2 单位。从批量较大的一批棒球中随机选取 81 个球进行检测, 检测数据的概括统计量如下。

Descriptive Statistics for Rebound Coefficient Data

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Rebound	81	85.296	85.387	85.285	1.771	0.197
Variable	Minimum	Maximum	Q1	Q3		
Rebound	80.934	89.687	84.174	86.352		

- 下图是否表明利用卡方方法构造  $\sigma$  的置信区间和对  $\sigma$  进行假设检验所需的条件被违反了？
- 是否有充分证据表明该批棒球反弹系数的标准差小于 2？
- 用 95% 置信区间估计反弹系数标准差。



### 7.3 比较两个总体方差时的估计和检验

在关于大肠杆菌检测方法的案例研究中,我们关心如何比较两总体的标准差。在很多情况下,比如比较两道工序,或比较产品的两个供应商,我们都需要比较与工序测量相关的总体的标准差。两总体方差的齐性检验的另一个主要应用就是对于两样本的  $t$  检验要评价方差相等这一条件的正确性(即  $\sigma_1^2 = \sigma_2^2$ )。本节所介绍的检验要求两总体的分布为正态。我们欲将总体 1 的方差  $\sigma_1^2$  和总体 2 的方差  $\sigma_2^2$

进行对比。

当容量分别为  $n_1$  和  $n_2$  两个随机样本独立地取自于两个正态分布总体时, 比率

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = \frac{s_1^2/s_2^2}{\sigma_1^2/\sigma_2^2}$$

具有  $F$  分布。这里略去其概率分布的公式, 但是我们将列出它的性质。

#### $F$ 分布的性质

1. 不像  $t$  或  $z$ , 但像  $\chi^2$  一样,  $F$  仅能取到正值。
2. 不像正态分布或  $t$  分布, 但像  $\chi^2$  分布那样,  $F$  分布是非对称的。(见图 7.9)。
3. 有许多  $F$  分布, 每一个都有不同的形状。我们规定每一个特定的  $F$  分布的自由度为相应的  $s_1^2$  和  $s_2^2$  的自由度, 分别记为  $df_1$  和  $df_2$ 。(见图 7.9)。
4.  $F$  分布的尾部值被编制成表, 见附表 8。

附表 8 列出了对应于面积  $\alpha = 0.25, 0.10, 0.05, 0.025, 0.01, 0.005$  和  $0.001$  的  $F$  分布的上尾值; 将  $s_1^2$  的自由度  $df_1$  列于表的顶部,  $s_2^2$  的自由度  $df_2$  列于表左侧的第一列,  $\alpha$  的值被列在下一列。因此当  $df_1 = 5, df_2 = 10$  时, 对应于面积  $\alpha = 0.25, 0.10, 0.05, 0.025, 0.01, 0.005$  和  $0.001$  的  $F$  分布临界值分别为 1.59,

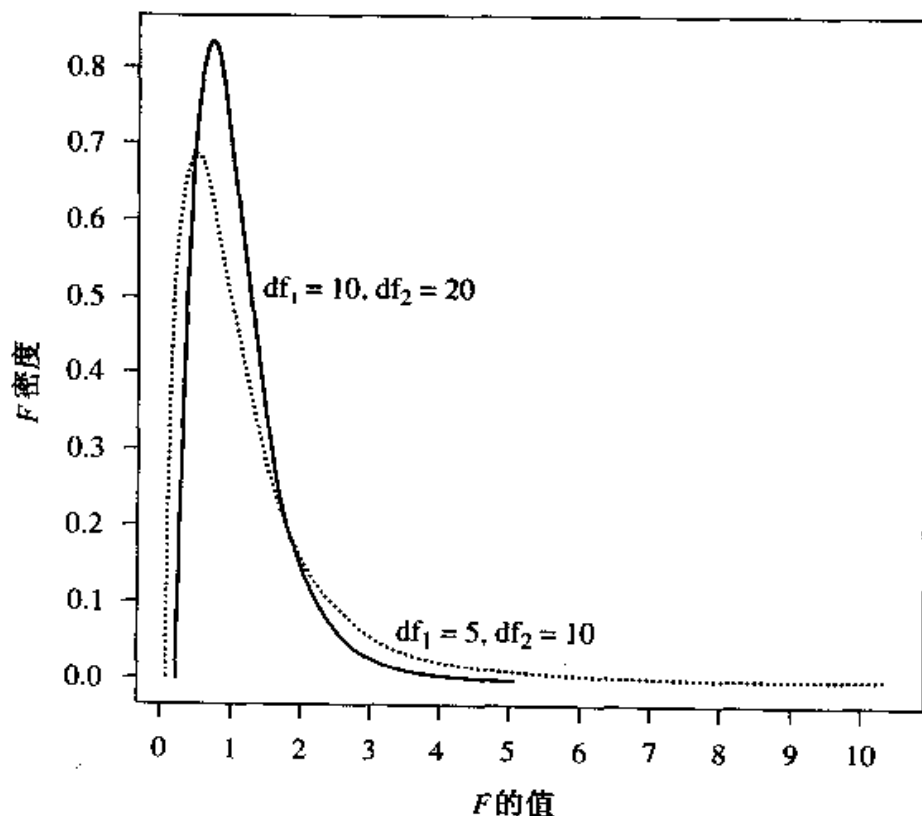


图 7.9 两个  $F$  分布的密度函数



2.52, 3.33, 4.24, 5.64, 6.78 和 10.48。于是, 在重复抽样中, 来自于自由度  $df_1 = 5, df_2 = 10$  的  $F$  分布的观测值中只有 5% 的观测值超过 3.33 (见图 7.10)。同样, 在当  $df_1 = 24, df_2 = 10$  时, 对应于尾部面积  $\alpha = 0.01$  和 0.001 的  $F$  临界值分别为 4.33 和 7.64。

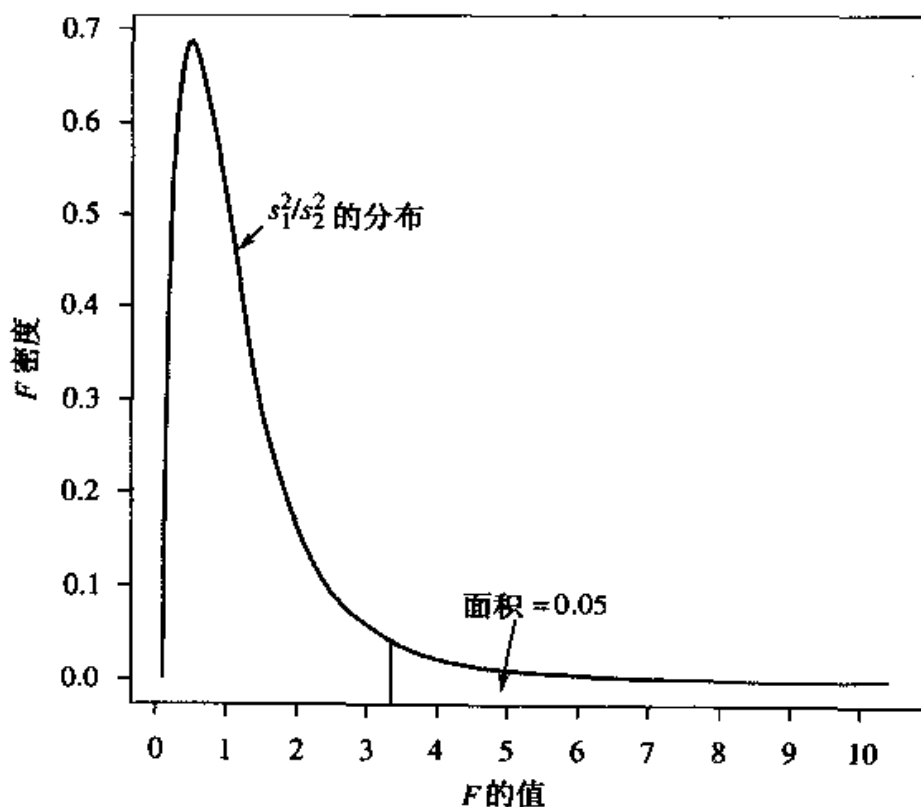


图 7.10  $F$  分布的临界值 ( $df_1 = 5, df_2 = 10$ )

利用检验统计量  $s_1^2/s_2^2$  可进行关于比较  $\sigma_1^2$  和  $\sigma_2^2$  的统计检验。当  $\sigma_1^2 = \sigma_2^2$  时,  $\sigma_1^2/\sigma_2^2 = 1$  且  $s_1^2/s_2^2$  服从自由度  $df_1 = n_1 - 1, df_2 = n_2 - 1$  的  $F$  分布, 附表 8 给出了  $F$  分布的上尾值。下尾值可以通过下面的关系求出: 分别令  $F_{\alpha, df_1, df_2}, F_{1-\alpha, df_1, df_2}$  为自由度为  $df_1, df_2$  的  $F$  分布的上  $\alpha$  分位数和下  $\alpha$  分位数。则,

$$F_{1-\alpha, df_1, df_2} = \frac{1}{F_{\alpha, df_2, df_1}}$$

注意等式两边  $F$  分位数的自由度正好相反。

#### 例 7.4

给出自由度为  $df_1 = 4, df_2 = 9$  的  $F$  分布的下 0.025 分位数。

**解答** 从附表 8 可以查到, 自由度  $df_1 = 9, df_2 = 4$  的  $F$  分布的下 0.025 分位数

为 8.90。因此,

$$F_{0.975, 9, 4} = \frac{1}{F_{0.025, 9, 4}} \quad \text{或} \quad F_{0.975, 4, 9} = \frac{1}{8.90} = 0.11$$

对于单侧备择假设, 记两总体为总体 1 和总体 2, 这样  $H_a$  的形式为:  $\sigma_1^2 > \sigma_2^2$ , 拒绝域落在  $F$  分布的上尾区域。

我们将检验方法总结如下。

#### 比较 $\sigma_1^2$ 和 $\sigma_2^2$ 的统计检验

$$H_0: 1. \sigma_1^2 \leq \sigma_2^2 \quad H_a: 1. \sigma_1^2 > \sigma_2^2$$

$$2. \sigma_1^2 = \sigma_2^2 \quad 2. \sigma_1^2 \neq \sigma_2^2$$

$$T. S. (\text{检验统计量}): F = s_1^2 / s_2^2$$

R. R. (拒绝域): 对于给定的  $\alpha$  值, 且  $df_1 = n_1 - 1, df_2 = n_2 - 1$ ,

1. 如果  $F \geq F_{\alpha, df_1, df_2}$ , 则拒绝  $H_0$ 。

2. 如果  $F \leq F_{1-\alpha/2, df_1, df_2}$ , 或  $F \geq F_{\alpha/2, df_1, df_2}$ , 则拒绝  $H_0$ 。

#### 例 7.5

以前, 我们讨论了一项试验, 其中公司职员所关心的是一种特定药品的药效维持的时间。为此, 从生产线上随机抽取 10 瓶药品作为随机样本且对每一瓶进行分析以确定它们的药效。同时抽取另外 10 瓶作为第二个样本, 并在规定的环境下将其储存一年。年底再观测这些药品的药效。然后, 根据这些样本数据构造两个时期平均药效差  $\mu_1 - \mu_2$  的置信区间。

我们在第六章提到, 在构造  $\mu_1 - \mu_2$  的置信区间或在作关于  $\mu_1 - \mu_2$  的统计检验时为了使用  $t$  方法, 我们要求两个样本分别来自于两个正态总体, 其均值可以不同, 但方差应该相同。利用下面根据样本数据算得的结果检验总体方差是否相等。给定  $\alpha = 0.05$ 。样本 1 的数据是刚下生产线的药品的药效读数, 样本 2 的数据是储存一年以后的药品的药效读数, 试给出结论。

样本 1:  $\bar{y}_1 = 10.37, s_1^2 = 0.105$ 。

样本 2:  $\bar{y}_2 = 9.83, s_2^2 = 0.058$ 。

**解答** 关于  $H_0: \sigma_1^2 = \sigma_2^2$  统计检验的四个部分如下。

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_a: \sigma_1^2 \neq \sigma_2^2$$

$$\text{检验统计量: } F = s_1^2 / s_2^2 = 0.105 / 0.058 = 1.81$$

在确定拒绝域之前, 我们必须检查两个样本是否均来自于正态总体。在这一条件能够满足之后, 我们才能确定下面的拒绝域。

拒绝域: 对于  $\alpha = 0.05$  的双侧检验, 如果  $F \geq F_{0.025, 9, 9} = 4.03$  或  $F \leq F_{0.975, 9, 9} = 1/F_{0.025, 9, 9} = 1/4.03 = 0.25$ , 我们将拒绝  $H_0$ 。因为 1.81 没有落在拒绝域中, 所以, 我们并不能拒绝  $H_0: \sigma_1^2 = \sigma_2^2$ 。对于  $t$  方法所需要的方差齐性, 从这些数据看来是成立的。

现在我们能求比率  $\sigma_1^2/\sigma_2^2$  的置信区间。

$\sigma_1^2/\sigma_2^2$  的置信系数为  $1 - \alpha$  的置信区间

$$\frac{s_1^2}{s_2^2} F_L \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} F_U$$

其中,  $F_U = F_{\alpha/2, df_2, df_1}$ ,  $F_L = F_{1-\alpha/2, df_2, df_1} = 1/F_{\alpha/2, df_1, df_2}$ , 自由度  $df_1 = n_1 - 1$ ,  $df_2 = n_2 - 1$ 。(注意:  $\sigma_1/\sigma_2$  的置信区间可以通过将  $\sigma_1^2/\sigma_2^2$  的置信区间端点值开平方得到。)

应该注意到, 尽管我们对  $\sigma_1^2/\sigma_2^2$  的估计方法适用于任何置信系数  $(1 - \alpha)$ , 但表 8 只允许我们使用常用的置信系数如 0.90, 0.95, 0.98, 0.99 等去构造  $\sigma_1^2/\sigma_2^2$  的置信区间。更详细的  $F$  分布分位数表可参见 Pearson and Hartley(1966)。

### 例 7.6

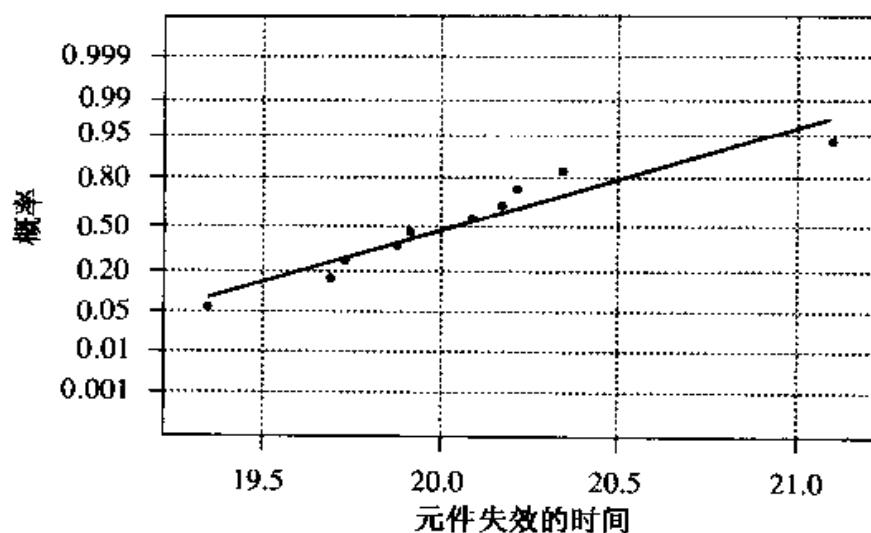
为研究一种电子元件在 110 和 220 伏两种电压下工作的寿命, 取 10 只不同的电子元件随机安排在 110 伏电压下工作, 而另外 16 只不同元件被随机安排在 220 伏电压下工作。记录 26 只元件失效的时间(以百小时计), 并得到如下统计量和正态概率图。

电压	样本容量	均值	标准差
110	10	20.04	0.474
220	16	9.99	0.233

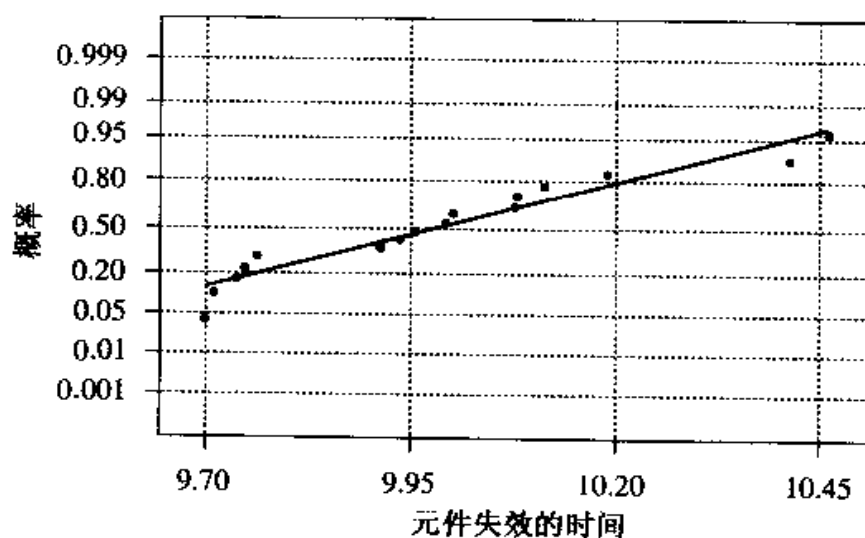
研究人员欲估计 110 和 220 伏电压下使用寿命变异程度的相对大小, 利用样本数据构造两种工作电压下元件使用寿命标准差比率  $\sigma_1/\sigma_2$  的 90% 置信区间。

**解答** 在构造置信区间之前, 有必要检查一下两个使用寿命总体是否服从正态分布。从正态概率图来看, 两个样本数据似乎均来自于正态总体。接下来, 我们需要查  $df_1 = 10 - 1 = 9$ ,  $df_2 = 16 - 1 = 15$  的  $F$  分布的上和下  $\alpha/2 = 0.10/2 = 0.05$  分位数。从附表 8 我们可以查得:

$$F_U = F_{0.05, 15, 9} = 3.01 \quad \text{和} \quad F_L = F_{0.95, 15, 9} = 1/F_{0.05, 9, 15} = 1/2.59 = 0.386$$



110 伏电压下使用寿命的正态概率图



220 伏电压下使用寿命的正态概率图

代入置信区间公式, 我们得到  $\sigma_1^2/\sigma_2^2$  的 90% 置信区间:

$$\frac{(0.474)^2}{(0.233)^2} 0.386 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{(0.474)^2}{(0.233)^2} 3.01$$

$$1.5975 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq 12.4569$$

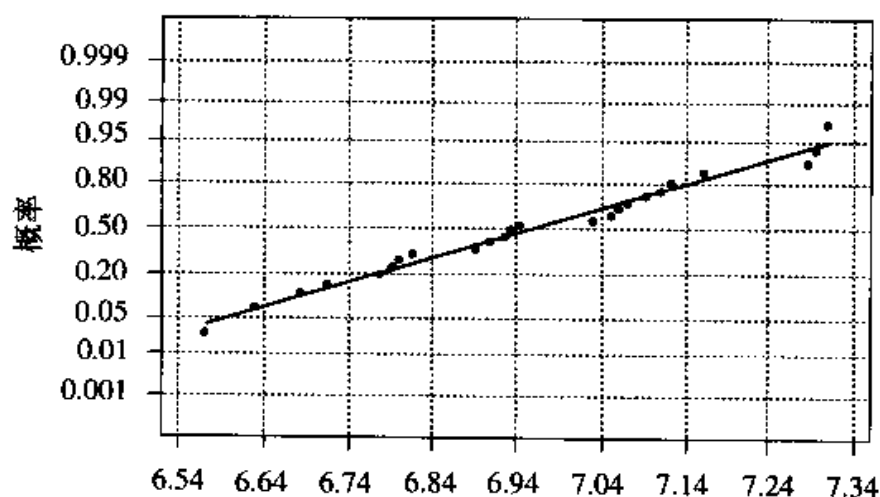
于是,  $\sigma_1/\sigma_2$  的 90% 置信区间为:

$$\sqrt{1.5975} \leq \frac{\sigma_1}{\sigma_2} \leq \sqrt{12.4569} \quad \text{或} \quad 1.26 \leq \frac{\sigma_1}{\sigma_2} \leq 3.53$$

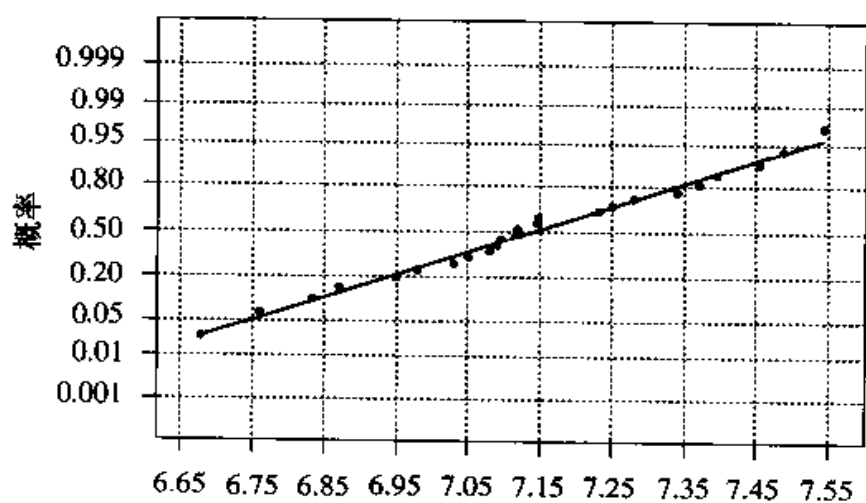
因此,我们有 90%把握相信  $\sigma_1$  在  $\sigma_2$  的 1.26 倍到 3.53 倍之间。

### 例 7.7

现在我们来完成大肠杆菌案例研究的第一阶段的分析。



(a) HGMF 测得的大肠杆菌浓度



(b) HEC 测得的大肠杆菌浓度

图 7.11 HEC 和 HGMF 的正态概率图

**案例中数据的分析** 由于研究的目的是要评价 HEC 方法在检测大肠杆菌方面的效果,因此,有必要将 HEC 方法和一种已经被接受了的方法—HGMF 法进行比较,以评价其可重复性和一致性。因此我们需要比较这两种方法检测大肠杆菌浓度的水平和变异性,即我们需要检验关于 HEC 和 HGMF 的大肠杆菌浓度均值和标准差的假设。对于一个具有特定水平  $7\log_{10}\text{CFU/ml}$  的大肠杆菌纯培养细菌,曾利用 HEC 和 HGMF 方法进行检测分别得到 24 个独立观察值。在构造置信区

间和进行检验假设之前,我们必须首先检查样本数据是否来自正态总体。从盒形图 7.2 和正态概率图 7.11 中可以看出,两种方法的样本数据均来自于正态总体。下面我们检验假设:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad H_a: \sigma_1^2 \neq \sigma_2^2$$

其中,记 HEC 为总体 1, HGMP 为总体 2。有关统计量如下。

方法	样本容量	均值	标准差
HEC	24	7.1346	0.2291
HGMF	24	6.9529	0.2096

R. R. (拒绝域): 对于  $\alpha = 0.05$  的双侧检验, 如果

$$F_0 = s_1^2/s_2^2 \leq F_{0.975, 23, 23} = 1/F_{0.025, 23, 23} = 1/2.31 = 0.43$$

$$\text{或 } F \geq F_{0.025, 23, 23} = 2.31$$

则我们将拒绝  $H_0$ 。

由于  $F_0 = (0.2291)^2/(0.2096)^2 = 1.19$  既不比 0.43 小也不比 2.31 大, 因此我们不能拒绝  $H_0$ 。利用计算机软件, 我们可以得出该统计检验的  $p$  值为 0.672。因此我们认为在检测大肠杆菌浓度时, HEC 和 HGMP 法有相似的变异程度。为了获得 HEC 和 HGMP 检测值变异程度的估计, 它们的标准差  $\sigma_{\text{HEC}}$  和  $\sigma_{\text{HGMP}}$  的 95% 置信区间分别为 (0.17, 0.23) 和 (0.16, 0.21)。

HEC 和 HGMP 两种大肠杆菌检验浓度读数是两个独立随机样本, 分别来自有着公共标准差的两个正态总体, 因此我们利用混合  $t$  检验来检验假设:

$$H_0: \mu_1 = \mu_2 \quad H_a: \mu_1 \neq \mu_2$$

R. R. (拒绝域): 对于  $\alpha = 0.05$  的双尾检验, 如果

$$|t| = \frac{|\bar{y}_1 - \bar{y}_2|}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{0.025, 46} = 2.01$$

则我们拒绝  $H_0$ 。

因为  $t = 2.87$  比 2.01 大, 所以我们拒绝  $H_0$ 。  $p$  值等于 0.0062。因此有充分证据表明 HEC 法检测大肠杆菌浓度的读数均值与 HGMP 法不同。为了估计平均读数, 可得  $\mu_{\text{HEC}}$  和  $\mu_{\text{HGMP}}$  的 95% 置信区间分别为 (7.04, 7.23) 和 (6.86, 7.04)。HEC 法的平均读数比 HGMP 的高。这些结果为研究的第二阶段做好了准备。在这一阶段中, HEC 和 HGMP 法被应用于实地研究中同样的肉类样本, 这种实地研究类似于我们将要遇到的对肉类质量的监控。两种方法具有同样的变异水平, 但 HEC 产生的大肠杆菌浓度读数比 HGMP 的高。因此阶段 2 的目的是参照 HGMP 读数校准 HEC 的读数。在本章的后面部分, 我们将讨论这一阶段的分析。

**报告结论** 我们需要写一份报告,总结我们关于阶段 1 的研究结果。报告应包括以下几部分:

1. 介绍研究目的。
2. 叙述研究设计和数据收集方案。
3. 用数字和图表对数据集合进行概括。
4. 叙述所有的推断方法。
  - $t$  检验和  $F$  检验。
  - 基于  $t$  方法的均值的置信区间。
  - 基于卡方方法的标准差的置信区间。
  - 证实所用推断技术的所有必要条件能够满足。
5. 对结论和结果的讨论。
6. 相对于以前的研究,对新发现的研究结果进行解释。
7. 对于未来研究的建议。
8. 将数据列表。

通过模拟研究考察了当样本来自于厚尾分布和偏态分布而不是来自于需要的正态分布时  $F$  检验的水平所受到的影响。模拟中所用到的五种分布在例 7.3 中已经介绍过了。

对于每一对样本容量  $(n_1, n_2) = (10, 10), (10, 20)$  或  $(20, 20)$ 。从每一个分布总体中抽取了给定样本容量的随机样本。在  $\alpha = 0.05$  水平下,对假设  $H_0: \sigma_1^2 = \sigma_2^2$   $H_a: \sigma_1^2 \neq \sigma_2^2$  进行  $F$  检验。对于五种分布中的每一个分布和三对样本容量都重复进行 2500 次这种检验,结果列于表 7.4 中。

表 7.4  $H_0: \sigma_1^2 = \sigma_2^2$  被拒绝次数的比例( $\alpha = 0.05$ )

样本容量	分 布				
	正态	均匀	$t$ ( $df=5$ )	伽马(形状=1)	伽马(形状=0.1)
(10,10)	0.054	0.010	0.121	0.225	0.693
(10,20)	0.056	0.0068	0.140	0.236	0.671
(20,20)	0.050	0.0044	0.150	0.264	0.673

表 7.4 中给出的值是两总体方差齐性  $F$  检验的犯第 I 类错误概率的估计值。对于三对样本容量来说,当样本来自于正态总体时,  $\alpha$  值都几乎等于其名义值 0.05,这是预料之中的。因为  $F$  检验就是在总体分布为正态时被构造用来检验假设。但是当总体分布是类似于均匀分布的对称短尾分布时,  $\alpha$  值比给定的值 0.05 小很多。因此,此时  $F$  检验犯第 II 类错误的概率很可能比当样本来自于正态总体

时要大很多。当总体分布是对称厚尾分布时,如自由度  $df=5$  的  $t$  分布,  $\alpha$  值比给定的值 0.05 大两到三倍。因此当总体分布具有这种形式时,  $F$  检验犯第 I 类错误的概率比预料的更大。同样的问题也发生在样本来自于偏态分布总体,如两个伽马分布。事实上,在这些情况下犯第 I 类错误的概率非常大,因此致使对于这些类型分布的  $F$  检验无效。

### 练习

#### 基本技能

7.13 在下列条件下,求  $F$  分布上尾面积为  $\alpha$  的  $F$  临界值。

- a.  $\alpha = 0.05$ ,  $df_1 = 7$ ,  $df_2 = 12$ 。
- b.  $\alpha = 0.05$ ,  $df_1 = 3$ ,  $df_2 = 10$ 。
- c.  $\alpha = 0.05$ ,  $df_1 = 10$ ,  $df_2 = 20$ 。
- d.  $\alpha = 0.01$ ,  $df_1 = 8$ ,  $df_2 = 15$ 。
- e.  $\alpha = 0.01$ ,  $df_1 = 12$ ,  $df_2 = 25$ 。

(注意:你的答案可能与书后面的答案不一致,只要你的答案接近于书中的答案,就可以认为是正确的。)

7.14 求出下列条件下  $F_\alpha$  的近似值

- a.  $\alpha = 0.05$ ,  $df_1 = 11$ ,  $df_2 = 24$ 。
- b.  $\alpha = 0.05$ ,  $df_1 = 14$ ,  $df_2 = 14$ 。
- c.  $\alpha = 0.05$ ,  $df_1 = 35$ ,  $df_2 = 22$ 。
- d.  $\alpha = 0.01$ ,  $df_1 = 22$ ,  $df_2 = 24$ 。
- e.  $\alpha = 0.01$ ,  $df_1 = 17$ ,  $df_2 = 25$ 。

7.15 分别从总体 1 和总体 2 中抽取  $n_1 = 8$ ,  $n_2 = 10$  的随机样本。相应的样本方差是  $s_1^2 = 7.4$ ,  $s_2^2 = 12.7$ 。数据是否提供了充分的证据表明  $\sigma_1^2$  与  $\sigma_2^2$  之间存在差异? 给定  $\alpha = 0.10$ , 试检验之。你做了什么样的假设?

7.16 通过做试验以确定是否有充分的证据表明一个总体(记为总体 A)内数据的变异程度,超过了另一个总体(记为总体 B)内数据的变异程度。从两个总体中分别抽取  $n_A = n_B = 8$  的随机样本,且计算样本方差如下:

$$s_A^2 = 2.87 \quad s_B^2 = 0.91$$

数据是否提供了充分的证据表明  $\sigma_A^2$  大于  $\sigma_B^2$ ? 试在  $\alpha = 0.05$  水平下进行检验。

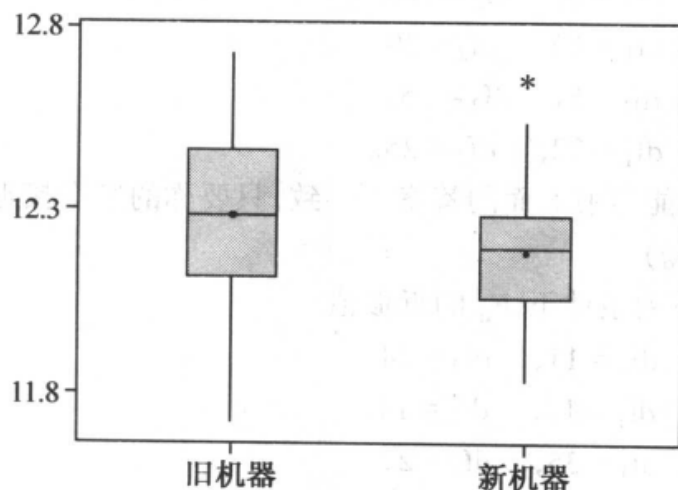
### 应用

7.17(工程) 一家软饮料公司正在评估一项购买新型罐装机器的投资。该公司已经了解到如果安装了新机器,那么花费同样的成本每天能生产更多罐饮料。但是,公司也必须搞清楚使用新机器进行罐装时各罐饮料重量之间的变异性,并且



希望新机器罐装的饮料重量变异性小于或等于旧机器罐装的饮料重量变异性。他们设计了一个研究方案,就是从两种机器生产的产品中各抽取 61 罐作为随机样本,并且规定了每罐饮料的重量(以盎司为单位)。将数据用表和盒形图描述如下。

罐装试验的数据			
机器类型	样本容量	均值	标准差
旧	61	12.284	0.231
新	61	12.197	0.162



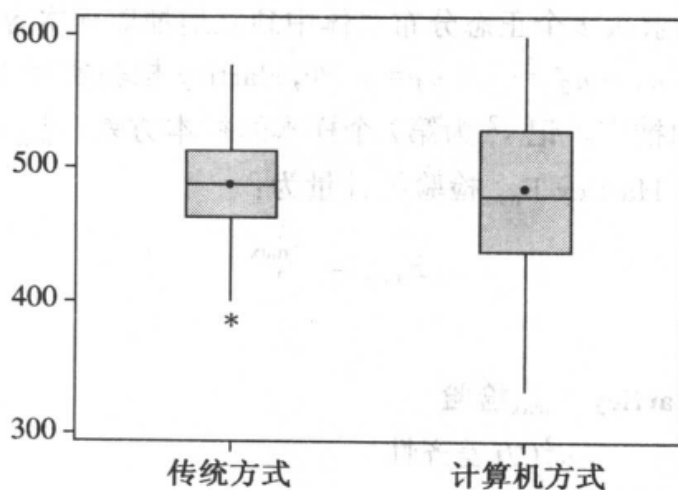
旧机器和新机器的盒形图  
(均值以实心圆点表示)

- 用 95% 置信区间估计新旧两种机器罐装饮料重量的标准差。
- 数据是否提供了充分的证据表明新机器比旧机器罐装饮料重量的变异性更小?
- 进行(a)和(b)的统计推断所需的条件是否满足? 试证明你的答案。

**7.18 (教育)** SAT (Scholastic Assessment Test) 考试作为大学入学条件的一部分是大多数高中生都要参加的一种考试。有人建议想改变这项考试的方式,让学生在计算机上进行考试。考试题目由学生按下面的方式进行选择。对于给定的一部分试题,如果学生把最初的问题答对了,后面的题就会更难;如果学生把最初的问题答错了,后面的题难度不会增加。根据每个考试的难度水平给出考试的最终标准分数。考试机构欲比较采用新方法和当前使用的旧方法之间考试分数的差异。随机选择了 182 名学生参与研究。其中随机安排 91 名采用新方法参加考试,

另 91 名学生采用旧方法参加考试。数学考试的分数以表和盒形图表述如下。

SAT 数学考试分数数据			
考试方式	样本容量	均值	标准差
计算机	91	484.45	53.77
传统	91	487.38	36.94



传统方式和计算机方式的盒形图  
(均值以实心圆点表示)

试评估两种 SAT 考试方式, 进行假设检验并构造置信区间。两种方式下考试分数的均值和标准差皆相等吗? 并证实你的结论, 取  $\alpha = 0.05$ 。

7.19 利用表 7.3 和 7.4 回答下列问题:

- 偏态和厚尾对卡方检验和  $F$  检验的影响相同吗?
- 对于给定的总体分布, 增加样本容量会使得  $\alpha$  值与其名义值 0.05 更接近吗? 并解释之。
- 对于短尾分布, 犯第 I 类错误的真实概率比给定值 0.05 确实小吗? 如果  $\alpha$  值的减小对  $F$  检验有负面影响的话, 那么有什么负面影响?

## 7.4 比较多个总体方差时的检验

在前面的部分, 我们讨论了根据独立地取自于两个正态分布总体的随机样本来比较两总体方差的方法。在许多情况中, 我们需要比较两个以上的总体。例如, 我们欲比较 5 个不同的供应商所供应食品的营养水平的变异性或比较使用三个主

要出版社出版的 SAT 考试材料的考生分数的变异性。因此我们需要设计一种统计检验方法,以便能比较  $t > 2$  个总体的方差。我们将考虑两种方法。第一个是 Hartley 检验,该方法应用简单,但要求总体必须服从正态分布且样本容量相等。第二个是 Levine 检验,该方法计算较复杂,但对总体分布和样本容量没有限制。许多统计软件都有 Levine 检验功能。例如 SAS 和 Minitab 都使用 Levine 检验比较总体方差。

H. O. Hartley(1950)建立了 **Hartley  $F_{\max}$  检验**用来检验假设

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_t^2 \quad H_a: \sigma_i^2 \text{ 不全相等}$$

Hartley  $F_{\max}$  检验要求从  $t$  个正态分布总体中独立地抽取样本容量均为  $n$  的  $t$  个随机样本。除了要求  $n_1 = n_2 = \cdots = n_t = n$  外, Hartley 检验实际上是前面检验  $t = 2$  个方差的  $F$  检验的推广。记  $s_i^2$  为第  $i$  个样本的样本方差,  $s_{\min}^2 = s_i^2$  中最小者,  $s_{\max}^2 = s_i^2$  中最大者。则 Hartley  $F_{\max}$  检验统计量为:

$$F_{\max} = \frac{s_{\max}^2}{s_{\min}^2}$$

检验步骤如下:

**总体方差齐性的 Hartley  $F_{\max}$  检验**

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_t^2 \text{ (方差齐性)}$$

$$H_a: \text{各总体方差不全相等}$$

$$T. S. \text{ (检验统计量): } F_{\max} = \frac{s_{\max}^2}{s_{\min}^2}$$

R. R. (拒绝域): 对于给定的  $\alpha$ , 如果对于  $\alpha = \alpha, t, df_2 = n - 1, F_{\max}$  超过表 12 中的  $F$  值, 则拒绝  $H_0$ , 其中,  $n$  是  $t$  个随机样本的共同样本容量。

检查假定并得出结论。

### 例 7.8

Wludyka 和 Nelson[ *Technometrics* (1997), 39: 274 ~ 285 ] 描述了下面这个实验。在生产隐形眼镜时, 一种单体被注入到塑料框中, 将这种单体用紫外线光照射并加热(时间, 温度, 和光的强度是变化的), 再拿走塑料框, 将镜片与水化合。据说能通过控制温度来达到所要求的镜片强度。因此, 我们感兴趣的是比较镜片强度的变异性。下面的数据是使用三个不同供应商的单体所生产的镜片, 其强度距目标值偏差的编码值。我们希望检验  $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ 。

对应于三个供应商的镜片的强度距目标值的偏差											
供应商	样 本									$n$	$s_i^2$
	1	2	3	4	5	6	7	8	9		
1	191.9	189.1	190.9	183.8	185.5	190.9	192.8	188.4	189.0	9	8.69
2	178.2	174.1	170.3	171.6	171.7	174.7	176.0	176.6	172.8	9	6.89
3	218.6	208.4	187.1	199.5	202.0	211.1	197.6	204.4	206.8	9	80.22

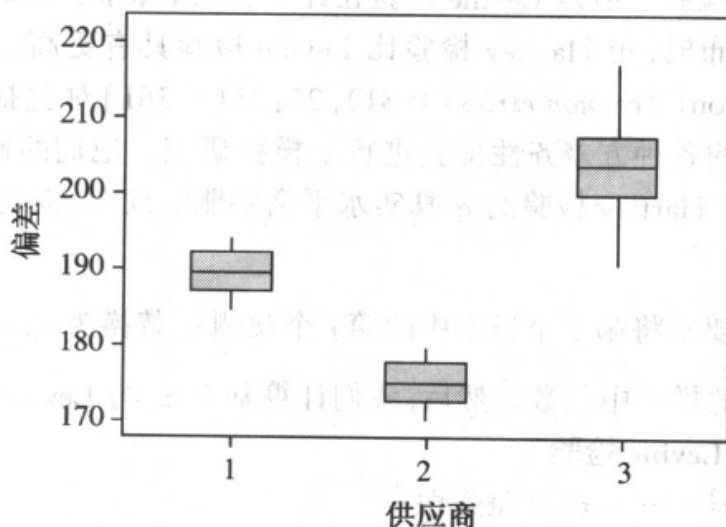


图 7.12 对应于三个供应商的镜片的强度距目标值的偏差的盒形图

**解答** 在进行 Hartley 检验之前,我们必须检查正态性条件是否满足。用图 7.12 中的盒形图观察数据的正态性。所有三组数据似乎都来自于正态分布总体。因此我们可以应用 Hartley  $F_{\max}$  检验。当  $\alpha = 0.05$ ,  $t = 3$ ,  $df_2 = 9 - 1 = 8$  时,查表 12,我们得到  $F_{\max, 0.05} = 6.00$ 。

因此拒绝域为:

拒绝域:如果  $F_{\max} \geq F_{\max, 0.05} = 6.00$ , 则拒绝  $H_0$ 。

$$s_{\min}^2 = \min(8.69, 6.89, 80.22) = 6.89$$

$$s_{\max}^2 = \max(8.69, 6.89, 80.22) = 80.22$$

因此

$$F_{\max} = \frac{s_{\max}^2}{s_{\min}^2} = \frac{80.22}{6.89} = 11.64 > 6.00$$

于是,我们拒绝  $H_0$ , 并且认为方差不全相等。

如果样本容量不全相等,我们取  $n = n_{\max}$ , 其中,  $n_{\max}$  是最大样本容量。这时

$F_{\max}$  不再具有精确的水平  $\alpha$ 。事实上,从犯第 I 类错误的概率略大于名义值  $\alpha$  的意义上讲,该检验不够准确。因此,和所有  $n_i$  相等,且样本均来自于方差相等的正态总体时的检验相比,该检验更容易错误地拒绝  $H_0$ 。

Hartley  $F_{\max}$  检验对于正态性的偏离十分敏感。因此,如果样本所来自的总体的分布稍微偏离正态但方差相等时,  $F_{\max}$  也会拒绝  $H_0$ , 从而认为方差不相等。此时检验探索的是总体的非正态性,而不是方差的不等。因此,当总体分布非正态时,  $F_{\max}$  检验不适合用来作为方差的齐性检验。另一种不需要总体服从正态分布的方法是 Levine 检验。但是 Levine 检验在计算上比 Hartley 检验更复杂,并且,当总体服从正态分布时,用 Hartley 检验比 Levine 检验具有更高的功效。Conover, Johnson, 和 Johnson [ *Technometrics* (1981), 23: 351 ~ 361 ] 对包括 Hartley 检验和 Levine 检验在内的各种方差齐性检验进行了模拟研究。他们的研究表明,当总体分布严重偏态时, Hartley 检验的  $\alpha$  真实水平会膨胀。此时,他们推荐使用 Levine 检验。

Levine 检验要求将第  $i$  个样本中的第  $j$  个观测  $y_{ij}$  替换为  $z_{ij} = |y_{ij} - \tilde{y}|$ , 其中,  $\tilde{y}$  是第  $i$  个样本的样本中位数。然后,我们计算基于  $z_{ij}$  的 Levine 检验统计量。

#### 总体方差齐性的 Levine 检验

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_t^2$  方差齐性

$H_a$ : 总体方差不全相等

$$T. S. (\text{检验统计量}): L = \frac{\sum_{i=1}^t n_i (\bar{z}_{i.} - \bar{z}_{..})^2 / (t-1)}{\sum_{i=1}^t \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_{i.})^2 / (N-t)}$$

R. R. (拒绝域): 对于给定的  $\alpha$ , 如果  $L \geq F_{\alpha, df_1, df_2}$ , 则拒绝  $H_0$ , 其中

$df_1 = t-1, df_2 = N-t, N = \sum_{i=1}^t n_i$ ,  $F_{\alpha, df_1, df_2}$  是  $F$  分布的上  $\alpha$  分位数, 见表 8。

检查假定并作出结论。

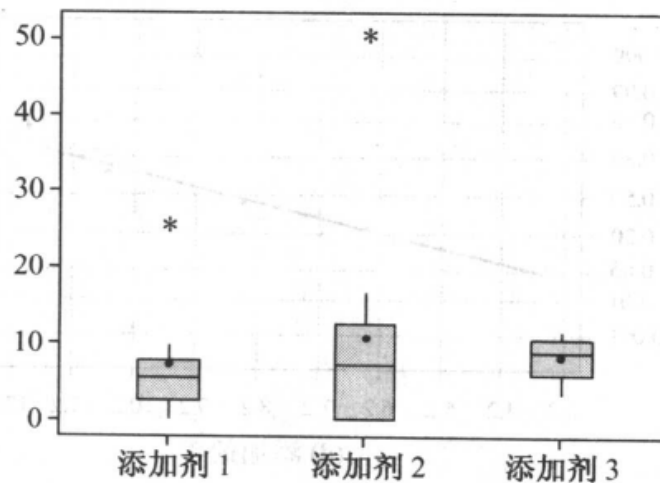
我们将在下面这个例子中举例说明 Levine 检验的计算。但是,大多数情况下,我们推荐使用计算机软件进行该检验,如 SAS 和 Minitab 软件。

#### 例 7.9

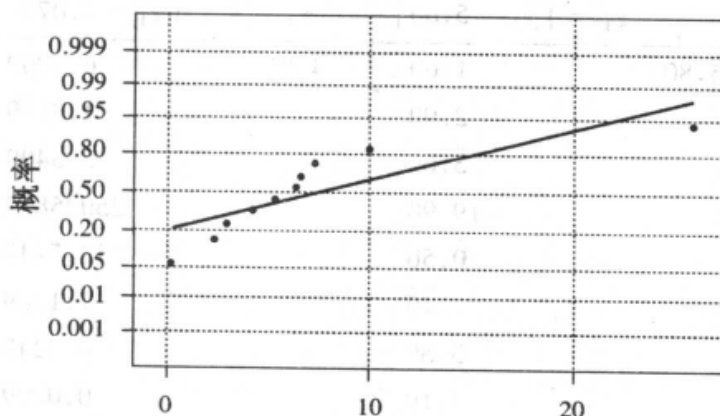
一消费者检测机构对市场上销售的用于增加汽车每加仑英里数 (mpg) 的三种添加剂进行评估。过去的研究表明,对于经济型汽车使用该产品行驶 250 英里,每加仑英里数可以平均增加 8%。该检测机构欲考察各种品牌的经济型小汽车增加行驶里程数的变异性。他们随机选取了使用年限相同,里程表读数相同,并且动力装置总的状况相同的 30 辆经济型小汽车。对于每一种添加剂,随机分配 10 辆小

汽车进行试验。每辆车行驶 250 英里,记录他们每加仑英里增加的百分数。他们想知道三种添加剂在增加行驶里程的变化方面是否存在差异。下面列出了原始数据以及计算 Levine 检验统计量所需的中间结果。

**解答** 从图 7.13(a)~(d)我们可以看出,来自添加剂 1 和 2 的样本似乎并非来自正态分布总体。因此,本例中,我们不能利用 Hartley  $F_{\max}$  检验来检验方差是否相等。表 7.5 中的信息会帮助我们计算 Levine 检验统计量的值。对应于三种添加剂,里数增加百分数  $y_{ij}$  的中位数分别为 5.80, 7.55 和 9.15。下面,我们计算每个数据相对于各自中位数的绝对偏差。记  $z_{1j} = |y_{1j} - 5.80|$ ,  $z_{2j} = |y_{2j} - 7.55|$ ,  $z_{3j} = |y_{3j} - 9.15|$ ,  $j = 1, 2, \dots, 10$ 。见表第三列。



(a) 添加剂 1, 2 和 3 的盒形图  
(均值以实心圆点表示)



(b) 添加剂 1

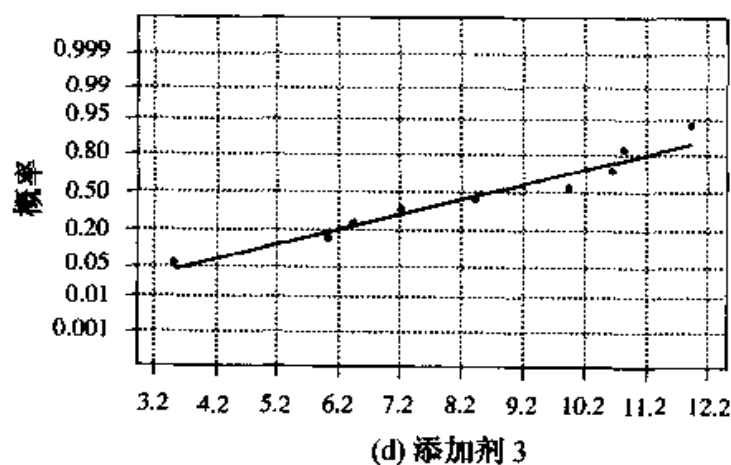
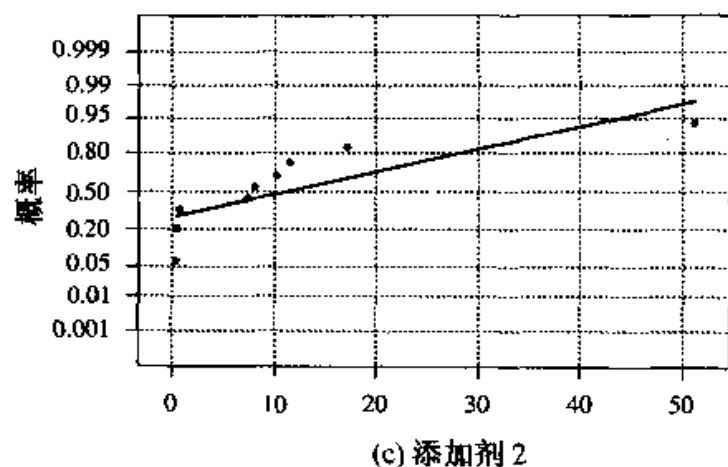


图 7.13(b)~(d) 添加剂 1,2 和 3 的正态概率图

表 7.5 使用三种添加剂使得小汽车行驶里程增加的百分数

添加剂	$y_{1j}$	$\bar{y}_1$	$z_{1j} =  y_{1j} - 5.80 $	$\bar{z}_1$	$(z_{1j} - 4.07)^2$	$(z_{1j} - 5.06)^2$
1	4.2	5.80	1.60	4.07	6.1009	11.9716
1	2.9		2.90		1.3689	4.6656
1	0.2		5.60		2.3409	0.2916
1	25.7		19.90		250.5889	220.2256
1	6.3		0.50		12.7449	20.7936
1	7.2		1.40		7.1289	13.3956
1	2.3		3.50		0.3249	2.4336
1	9.9		4.10		0.0009	0.9216
1	5.3		0.50		12.7449	20.7936
1	6.5		0.70		11.3569	19.0096

续表

添加剂	$y_{2j}$	$\bar{y}_2$	$z_{2j} =  y_{2j} - 7.55 $	$\bar{z}_2$	$(z_{1j} - 8.88)^2$	$(z_{2j} - 5.06)^2$
2	0.2	7.55	7.35	8.88	2.3409	5.2441
2	11.3		3.75		26.3169	1.7161
2	0.3		7.25		2.6569	4.7961
2	17.1		9.55		0.4489	20.1601
2	51.0		43.45		1,195.0849	1,473.7921
2	10.1		2.55		40.0689	6.3001
2	0.3		7.25		2.6569	4.7961
2	0.6		6.95		3.7249	3.5721
2	7.9		0.35		72.7609	22.1841
2	7.2		0.35		72.7609	22.1841
添加剂	$y_{3j}$	$\bar{y}_3$	$z_{3j} =  y_{3j} - 9.15 $	$\bar{z}_3$	$(z_{1j} - 2.23)^2$	$(z_{3j} - 5.06)^2$
3	7.2	9.15	1.95	2.23	0.0784	9.6721
3	6.4		2.75		0.2704	5.3361
3	9.9		0.75		2.1904	18.5761
3	3.5		5.65		11.6964	0.3481
3	10.6		1.45		0.6084	13.0321
3	10.8		1.65		0.3364	11.6281
3	10.6		1.45		0.6084	13.0321
3	8.4		0.75		2.1904	18.5761
3	6.0		3.15		0.8464	3.6481
3	11.9		2.75		0.2704	5.3361
Total			5.06		1,742.6	1,978.4

其次,我们计算三个平均值,  $\bar{z}_1 = 4.07$ ,  $\bar{z}_2 = 8.88$ ,  $\bar{z}_3 = 2.23$ 。再其次,我们计算  $z_{ij}$  关于各自均值的偏差平方  $(z_{ij} - \bar{z}_{i.})^2$ , 即  $(z_{1j} - 4.07)^2$ ,  $(z_{2j} - 8.88)^2$ ,  $(z_{3j} - 2.23)^2$ 。这些值列于表中的第六列。然后,我们计算  $z_{ij}$  关于总均值  $\bar{z}_{..} = 5.06$  的偏差平方, 即  $(z_{ij} - \bar{z}_{..})^2 = (z_{ij} - 5.06)^2$ 。见表中最后一列。最后,分别加总第六列和第七列。得

$$T_1 = \sum_{i=1}^3 \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_{i.})^2 = 1742.6, T_2 = \sum_{i=1}^3 \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_{..})^2 = 1978.4$$

Levine 检验统计量的值也可以用下面的形式求出:

$$L = \frac{(T_2 - T_1)/(t-1)}{T_1/(N-t)} = \frac{(1978.4 - 1742.6)/(3-1)}{1742.6/(30-3)} = 1.827$$

Levine 检验的拒绝域为: 如果  $L \geq F_{\alpha, t-1, N-t} = F_{0.05, 3-1, 30-3} = 3.35$ , 则拒绝  $H_0$ 。



因为  $L = 1.827$ , 所以, 我们不能拒绝  $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ 。从而我们得出结论, 认为没有充分的证据表明三种添加剂增加里数百分数的总体的方差存在差异。

### 练习

### 应用

**7.20** 在例 7.9 中, 我们指出 Hartley 检验不适合, 因为有证据表明其中两个总体分布非正态, 然后应用了 Levine 检验。通过检验获知, 在  $\alpha = 0.05$  水平下数据并不支持总体方差存在差异。根据数据得到如下统计量:

添加剂	样本容量	均值	中位数	标准差
1	10	7.05	5.80	7.11
2	10	10.60	7.55	15.33
3	10	8.53	9.15	2.69

- 利用例 7.9 中的图, 证明总体分布非正态。
- 使用 Hartley 检验法检验总体方差的差异性。
- Hartley 检验与 Levine 检验结果一致吗?
- 哪种检验更合适? 并解释之。
- 你将哪家单体供应商推荐给隐形眼镜的制造商? 并解释之。

**7.21** 参见例 7.8。用 Levine 检验法检验总体方差的差异性。

- 在例 7.8 中, 我们说总体分布是正态的, 试证明之。
- Levine 检验与 Hartley 检验结果一致吗?
- 哪种检验更合适? 并解释之。
- 哪种添加剂较好? 试解释你的选择。

**7.22 (生物)** 一位野生生物学家对研究圈养对鹿的大小所产生的影响感兴趣。她考虑三个总体: 野外放养的鹿, 在大型狩猎农场中饲养的鹿, 在动物园中饲养的鹿。在每一种环境下饲养的鹿中, 她随机选了 8 只鹿, 且在它们一岁时称量它们的体重。体重数据(单位: 磅)如下表所示。

环境	鹿的重量(磅)							
野外	114.7	128.9	111.5	116.4	134.5	126.7	120.6	129.59
农场	120.4	91.0	119.6	119.4	150.0	169.7	100.9	76.1
公园	103.1	90.7	129.5	75.8	182.5	76.8	87.3	77.3

- 生物学家假设圈养鹿的体重变异性大于野外放养鹿的体重变异性。数据支持这一论点吗?

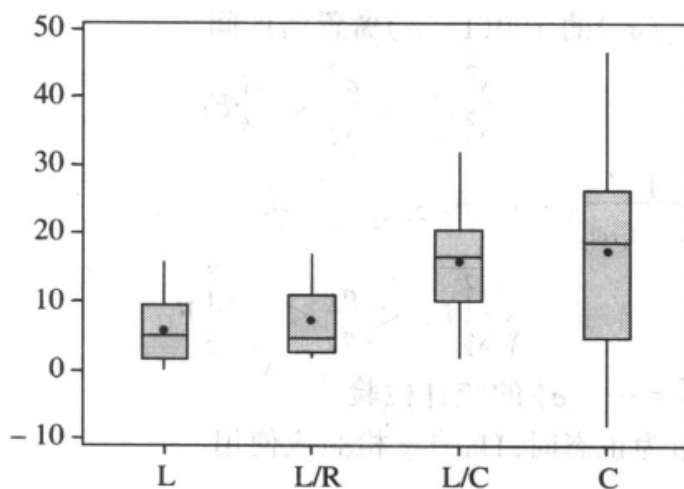
b. (a)所用检验的必要条件满足吗? 作图以支持你的答案。

**7.23** 叙述一项你所参加的实验或一个来自你研究领域文献的研究项目, 在其中到对被测响应的方差进行对比, 且研究处理本身的变异性与研究处理均值的差异性同等重要的。

**7.24** 为什么你认为当总体分布严重偏态时, Levine 检验比 Hartley 检验更合适于检验总体方差之间的差异性呢? (提示: 总体中心位置的哪一种度量更容易受到偏态分布的影响, 均值还是中位数?)

**7.25 (教育)** 许多学区企图通过利用计算机进行辅助教学以达到既能减少成本又能激励学生的目的。将一所学校的一群学生随机分成 4 组进行分数的加法和乘法运算的教学。每组使用一种教学方法, 这四种方法是: 仅仅讲授(L), 讲授并配以补习教科书辅助教学(L/R), 讲授并利用计算机进行辅助教学(L/C), 仅通过计算机进行教学(C)。15 周之后, 对这些学生进行一次考试。15 周开始时也对学生进行过一次考试。两次考试分数之差列于下表。校管方想知道哪种方法使得考试分数增加最多且增加量最一致。哪一种教学方法最成功? 给出所有的相关检验, 置信区间和图以证明你的结论。

方法	学 生									
	1	2	3	4	5	6	7	8	9	10
L	7	2	2	6	16	11	9	0	4	2
L/R	5	2	3	11	16	11	3			
L/C	9	12	2	17	12	20	20	31	21	
C	17	19	26	1	47	27	-8	10	20	



L, L/R, L/C 和 C 的盒形图(均值以实心圆点表示)

## 7.5 小 结

本章,我们讨论了关于总体方差,或等价地,总体标准差的推断方法。关于  $\sigma$  的估计和统计检验利用了  $df = n - 1$  的  $\chi^2$  分布。在关于两个总体方差或标准差比率的推断中利用了  $df_1 = n_1 - 1, df_2 = n_2 - 1$  的  $F$  分布。最后,我们利用 Hartley 检验统计量和 Levine 检验统计量建立了关于  $t \geq 2$  个总体方差的齐性检验。

关于一个或多个总体方差的推断问题可以追溯到第三章我们讨论过的关于总体的数值描述性度量方法。为了描述或推断一个总体,我们不能总是依赖于均值,即集中趋势的度量。在评价或比较心理测试中个人的表现,取自于某生产线的产品的一致性,或某品种玉米的产量时,很多次我们通过研究总体方差获得了重要信息。

## 重要公式

1.  $\sigma^2$  (或  $\sigma$ ) 的  $100(1 - \alpha)\%$  置信区间

$$\frac{(n-1)s^2}{\chi_U^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2} \quad \text{或} \quad \sqrt{\frac{(n-1)s^2}{\chi_U^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_L^2}}$$

2.  $\sigma^2$  的统计检验 ( $\sigma_0^2$  已知)

$$\text{检验统计量: } \chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

3.  $\sigma_1^2/\sigma_2^2$  的统计检验

$$\text{检验统计量: } F = \frac{s_1^2}{s_2^2}$$

4.  $\sigma_1^2/\sigma_2^2$  (或  $\sigma_1/\sigma_2$ ) 的  $100(1 - \alpha)\%$  置信区间

$$\frac{s_1^2}{s_2^2} F_L < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} F_U$$

其中,  $F_L = \frac{1}{F_{\alpha/2, df_1, df_2}}, F_U = F_{\alpha/2, df_2, df_1}$  或

$$\sqrt{\frac{s_1^2}{s_2^2} F_L} < \frac{\sigma_1}{\sigma_2} < \sqrt{\frac{s_1^2}{s_2^2} F_U}$$

5.  $H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_t^2$  的统计检验

- a. 当总体分布为正态时, Hartley 检验应使用

$$\text{检验统计量: } F_{\max} = \frac{s_{\max}^2}{s_{\min}^2}$$

- b. 当总体分布非正态时, Levine 检验应使用

$$\text{检验统计量: } L = \frac{\sum_{i=1}^t n_i (\bar{z}_{i.} - \bar{z}_{..})^2 / (t-1)}{\sum_{i=1}^t \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_{i.})^2 / (N-t)}$$

其中,  $z_{ij} = |y_{ij} - \tilde{y}_{i.}|$ ,  $\tilde{y}_{i.} = y_{i1}, \dots, y_{in_i}$  的中位数,  $\bar{z}_{i.} = z_{i1}, \dots, z_{in_i}$  的均值,  $\bar{z}_{..} = z_{11}, \dots, z_{tn_i}$  的均值

### 补充练习

**7.26** 参见练习 6.21。在那里我们对比较两种公鸡鸡冠的重量感兴趣, 其中每种公鸡的饲料中均添加了两种维他命添加剂之一。Wilcoxon 秩和检验被建议用来检验假设两个总体同分布。用  $t$  检验来比较两个总体均值合适吗? 试解释之。

**7.27 (商业)** 一家保护消费者权益的杂志欲比较从两家不同的公司购买的轮胎, 其中每家公司都宣称他们的轮胎能承受 40,000 英里的行驶路程。在模拟的道路状况下, 随机抽取每种品牌的 10 个轮胎进行测试。当轮胎面磨损到一定厚度时记下它们行驶的英里数, 数据(单位: 英里)如下。

品牌 I	38.9	39.7	42.3	39.5	39.6	35.6	36.0	39.2	37.6	39.5
品牌 II	44.6	46.9	48.7	41.5	37.5	33.1	43.4	36.5	32.5	42.0

- 根据数据作图, 并比较两种品牌轮胎使用寿命的分布。
- 对于两种品牌的轮胎, 分别构造直到轮胎面被磨破时轮胎行驶里程均值和标准差的 95% 置信区间。
- 两种品牌轮胎的耐磨性是否有差异? 用适当的图, 假设检验和置信区间来证明你的论点。

**7.28 (医药)** 从某精神病院随机抽取 20 位病人, 每人均患有抑郁症。对每人进行简单的精神病等级评估。评估等级中包括一系列形容词, 表示这些病人的情绪。过去更广泛的检测表明具有某些情绪形容词的等级倾向于一致, 因此将他们放在一起用以联合评价病人情绪的一个或多个组成部分。例如, 一群具有某些形容词的病人可能被认为患有抑郁症。如果一个病人总体的标准差大于 4, 那么, 为了治疗, 一般将这个总体分成至少两组使得同组病人的病情更一致。下面列出这 20 位病人的数据。

10 16 15 13 19 12 16 16 14 11  
21 21 18 18 24 16 13 16 24 9

- 求病人得分总体标准差  $\sigma$  的 95% 置信区间。

- b. 需要将病人总体分成几个组吗?
- c. 上面推断所需要的条件是什么? 它们能够满足吗? 用适当的数据图说明之。

**7.29 (医药)** 一家药品公司生产某种品牌的抗组胺药片。其质量控制部门定期对产品进行某些检测以检查产品在投放市场之前是否满足特定的标准。特别地, 公司要求药力应在标定药量的 90% 到 110% 范围之内。

- a. 如果公司正在生产 25mg 的药片, 那么药力应在什么范围。
- b. 从最近生产的一批抗组胺药片中随机抽取了 30 片。药力数据如下。对总体方差进行推断时的正态假定能保证吗?
- c. 将公司关于药力范围在 90% 到 110% 之间这一说明翻译成关于药力总体方差的统计检验。试在  $\alpha = 0.05$  水平下给出你的结论。

24.1	27.2	26.7	23.6	26.4	25.2
20.8	27.3	23.2	26.9	27.1	26.7
22.7	26.9	24.8	24.0	23.4	25.0
24.5	26.1	25.9	25.4	22.9	24.9
26.4	25.4	23.3	23.0	24.3	23.8

**7.30 (工程)** 一项研究旨在比较两种不同工序下生产的 1 寸方人工合成纤维强度的变异性。从每个工序中随机选取 9 个 1 寸方纤维, 并进行检测。

- a. 分别对每个样本作图。
- b. 正态假定能得到保证吗?
- c. 如果正态假定能得到保证, 那么利用下面的数据(psi)检验研究假设: 对应于两种工序的总体方差存在差异, 取  $\alpha = 0.05$ 。

工序 1	74	90	103	86	75	102	97	85	69
工序 2	59	66	73	68	70	71	82	69	74

**7.31** 参见例 7.2。构造  $\sigma^2$  的 95% 置信区间, 并利用这一区间帮助解释消费者组织所得的结果。例 7.2 的检验具有较高的功效来检验出  $\sigma^2$  比宣称的值增加了 25% 吗? 试解释之。

**7.32 (商业)** 投资风险通常用投资回报的方差来进行度量, 而这种回报是能够观测得到的。从两个不同的投资组合中各随机抽取 10 个年回报额作为样本。数据如下(以千美元为单位)。

投资组合 1	130	135	135	131	129	135	126	136	127	132
投资组合 2	154	144	147	150	155	153	149	139	140	141

- a. 投资组合 2 比投资组合 1 有更大的风险吗?
- b. 给出检验的  $p$  值, 并构造两个标准差比率的置信区间。
- c. 证明(a)和(b)中推断所需的条件已经满足。

**7.33** 参见练习 7.32。两种投资组合的平均回报有差别吗? 叙述在得出结论过程中你所用的方法, 并解释为什么你用这种方法。

**7.34 (商业)** 考虑两种不同的用来评估房屋转售价值的建模技术。随机抽取 12 幢正挂牌上市的房屋作为样本。对每幢房屋均采用两种建模技术分别估价。数据如下。

待售房屋	评 估 价	
	1	2
1	155	138
2	137	128
3	248	230
4	136	146
5	102	95
6	87	82
7	63	67
8	129	134
9	144	149
10	270	292
11	157	150
12	51	48

- a. 作图。两种不同的建模技术得到同样的结果吗?
- b. 估计两种建模技术估价差的均值和标准差。

**7.35** 参见练习 7.34。求两种方法估价差方差的 90% 置信区间, 并给出  $\sigma$  的相应的置信区间。

**7.36** 参见练习 7.34 和 7.35。关于样本数据的关键性假定是什么? 如何检查这一假定? 数据表明该假定成立吗? 你对练习 7.35 中的推断有什么需要注意的吗?

**7.37 (医药)** 在评价一种药品的效力时, 一个重要因素就是药品放在药架上一段时间以后, 药力下降的幅度。特别地, 药力下降幅度的变化是非常重要的。研究人员研究两种药品经过六个月时间的放置后药力下降的情况。假定药品 1 是试

验药品, 药品 2 是市场上销售的药品。数据如下。

试验药品	59.4	52.5	69.0	63.9	62.6	55.8	52.7	50.8
销售药品	51.9	52.5	50.0	55.2	51.3	52.7	59.0	64.8
试验药品	53.1	76.9	62.1	53.6	51.2	88.3	51.4	
销售药品	50.1	52.7	50.6	54.8	53.1	55.8	50.0	

研究人员欲确定试验药品和市场上销售药品的药力下降均值和标准差是否存在差异。在确定两种药品是否存在差异时, 应给出检验的  $p$  值, 所有相关参数的置信区间和图形以便确定为了使用各种推断方法, 这些数据是否满足所需的条件。

**7.38 (医药)** 随机选择若干位具有相同病史的病人, 比较在食用两种不同饮食的情况下他们血液中的胆固醇水平。一组食用低脂肪饮食, 另一组食用正常饮食。将数据汇总如下。

	低脂肪	正常
样本容量	19	24
样本均值	170	196
样本方差	198	435

a. 这些数据是否显示出了充分的证据表明食用两种不同的饮食其胆固醇水平的变异性存在差异? 给定  $\alpha = 0.05$ 。

b. 在比较两种饮食时, 是否还有其他什么检验?

**7.39 (医药)** 在美国市场上减肥药品的销售给许多生产这些药品的公司带来了可观的收入。一种减肥方法对一个人的减肥效果既受这个人身体条件的影响又受心理条件的影响。比较两种减肥药品 A 和 B。特别地, 考虑人们持续使用一种治疗方法的时间长度。随机将总共 26 名身体条件相当的超重男子分成两组。第一组服用药品 A, 第二组服用药品 B。数据如下(以天为单位)。

药品 A	42	47	12	17	26	27	28	26	34	19	20	27	34
药品 B	35	38	35	36	37	35	29	37	31	31	30	33	44

比较两组人持续用药时间的长度。给出所有相关图形, 检验, 置信区间, 并写一份关于两种治疗方法的书面总结。

**7.40** 参见练习 7.39, 如果药品 A 是一个在市场上用了多年的旧产品, 而药品 B 是新产品, 那么你的推断方法改变吗? 我们想知道和药品 A 相比, 人们持续使用 B 的时间是否将更长。

**7.41 (工程)** 某铁矿厂的一名化学工作者怀疑每磅矿石含铁氧化物量(重量, 以盎司为单位)的方差随着每磅矿石含铁氧化物均值的增加而增加。为了检验这一理论, 从两个地点分别随机挑选 10 个 1 磅重的铁矿石标本。地点 1 相对于地点 2 有着较高的平均铁氧化物。矿石标本中含有铁氧化物的重量数据如下。

地点 1	8.1	7.4	9.3	7.5	7.1	8.7	9.1	7.9	8.4	8.8
地点 2	3.9	4.4	4.7	3.6	4.1	3.9	4.6	3.5	4.0	4.2

数据是否提供了充分的证据表明地点 1 比地点 2 每磅矿石含铁氧化物数量的变化程度更大一些? 给定  $\alpha = 0.05$ 。试给出一个置信区间以显示两个方差之差的大小。

**7.42 (人力资源)** 某人事部门官员计划用  $t$  检验比较一跨国公司两个部门平均每月未经允许而缺席的人数, 但是发现可能存在一些困难。两个部门月缺席人数的方差似乎不同。为此, 从每个部门随机选取 5 个月, 记录月缺席人数。

部门 A	20	14	19	22	25
部门 B	37	29	51	40	26

- 什么样的假定令人事部门官员烦恼?
- 数据是否提供了充分的证据表明两部门缺席职员总体的方差存在差异? 给定  $\alpha = 0.05$ 。

**7.43 (环境)** 某研究人员对凤凰城和西雅图的天气形式很感兴趣。作为研究的一部分, 他随机选取 20 天(7 月份), 记录每天的平均气温。所收集的数据跨越好几年, 以保证日温度的独立性。数据如下(华氏):

凤凰城	96	95	85	98	94	93	98	98	96	96
西雅图	60	66	65	62	67	69	64	65	61	57
凤凰城	93	102	99	85	102	83	93	107	104	87
西雅图	55	58	75	67	63	60	71	59	50	63



数据是否表明两城市 7 月份日平均气温的变化存在差异? 7 月份两城市日平均气温有差异吗? 已知两检验的水平均为  $\alpha = 0.05$ 。

**7.44** 根据数据盘中的临床试验数据库, 计算每个处理组内部焦虑得分的样本方差。利用这些数据作单独检验, 分别将处理 A, B 和 C 与安慰剂组 D 进行比较。用  $\alpha = 0.05$  作双侧检验。

**7.45** 练习 7.44 中的任一检验是否均否定了在分别将 A, B 和 C 的处理均值与安慰剂组 D 的处理均值进行比较时使用  $t$  检验的可能性? 试解释之。

**7.46** 利用临床数据库中睡眠失调得分构造  $\sigma_B^2/\sigma_C^2$  的 98% 置信区间。同样地, 构造  $\sigma_B^2/\sigma_A^2$  的 98% 置信区间。

## 第八章 两个以上总体的中心值的推断

- 8.1 引言和案例
- 8.2 两个以上总体均值的统计检验:方差分析
- 8.3 完全随机化设计中观测值的模型
- 8.4 方差分析条件的检查
- 8.5 其他的分析方法:数据变换
- 8.6 另一种非参数方法:Kruskal-Wallis 检验
- 8.7 小结

### 8.1 引言和案例

在第六章,基于独立随机样本,我们提出了比较两个总体均值的方法。通常情况下,双样本问题是实际问题的简单化。例如,我们要比较分别来自于三个不同民族(非洲裔美国人、英国裔美国人、西班牙裔美国人)并同受雇于一家大型生产公司且不属于工会组织的农场工人的平均小时工资。分别从三个民族(总体)的农场工人中独立地选取随机样本。那么,利用三个样本均值的信息,我们可以对相应的总体平均小时工资进行推断。样本均值很可能不同,但这并不一定意味着三个总体的均值也不同。怎样由样本均值的不同去推断相应总体均值的不同呢?我们将用方差分析这一统计检验方法回答如上问题。

#### 案例:用激光对深紫色胎痣进行处理的时间效应

脸部的深紫色胎痣是一种先天性血管畸形,在新生儿中约占千分之三,胎痣对儿童的社会和心理调节有很大的影响。1985年,一种激光技术(即 flash-pumped, pulsed-dye laser)被用于儿童深紫色胎痣的治疗,由于儿童的皮肤细薄且痣较小,我们可以假设用这种激光治疗儿童比治疗成人更有效,达到最优清理所需要的治疗次数较少,因此认为要及早开始治疗。

在文章“Effect of the timing of treatment of port-wine stains with the flash-lamp-pumped-dye laser”(1998), *The New England Journal of Medicine*, 338: 1028—1033 所描述的前瞻性研究中,研究者讨论了是否治疗的年纪越轻越好。

**数据收集的设计** 研究人员考虑了如下与最有效治疗相关的问题:

1. 在降低深紫色胎痣的可见度方面,什么样的客观度量被用于评估治疗的有效性?
2. 评价一种治疗方案需要考虑多少不同的年龄组?

3. 什么样的试验设计对不同的治疗方案能产生最有效的比较?
4. 作比较的有效统计方法是什么?
5. 什么样的信息应该包括在记录不同年龄组用激光进行治疗的有效性的最后报告中?

挑选了 100 个从前没治疗过的、年龄不超过 31 岁的深紫色胎痣病人作为研究对象。在首次就诊时,记录下胎痣的范围和位置,为了评价激光治疗是否对较年轻的病人更有效,按照开始就诊的次序每 25 人分成一组,每一组有 4 个年龄段。要达到对胎痣的最优清除需要一系列的治疗。首次治疗前,由照相师在标准条件下对每一个病人进行彩色照相,用染色计(chromometer)对皮肤的颜色作测试,颜色测试的再生性可由治疗前在同一个位置的两次测试来分析。对每一个病人,以后的颜色测试都在同一个位置进行。如果胎痣已经消失或三次治疗后没有减轻就停止治疗,我们的目的是测试每一个病人的胎痣的皮肤和健康皮肤间在颜色的差异上是否降低。

由于研究期间情况的变化,有 11 人没有包括在最后的分析中。记录了 89 个病人的各种基本特征:性别、胎痣的表面积和位置、对治疗的有效性有影响的其他医疗条件,也包括诸如看医生的平均次数、放射暴露的水平、每次治疗的激光脉冲数、治疗后头疼的发生这些有关治疗特点的变量。相对于这些特征,在 4 个年龄组之间没有大的区别。

研究者主要感兴趣以下两个方面,一是治疗前胎痣的皮肤和健康皮肤间颜色的差异,二是一系列治疗后这种颜色差异的改变。图 8.1 表示治疗前颜色的差异,盒形图表明 4 个年龄组在颜色上没有很大的差异,这是重要的,如果治疗前组别不同,那么年龄组对治疗的有效性的影响就可能被先前的差异所掩盖。(表 8.1 中的

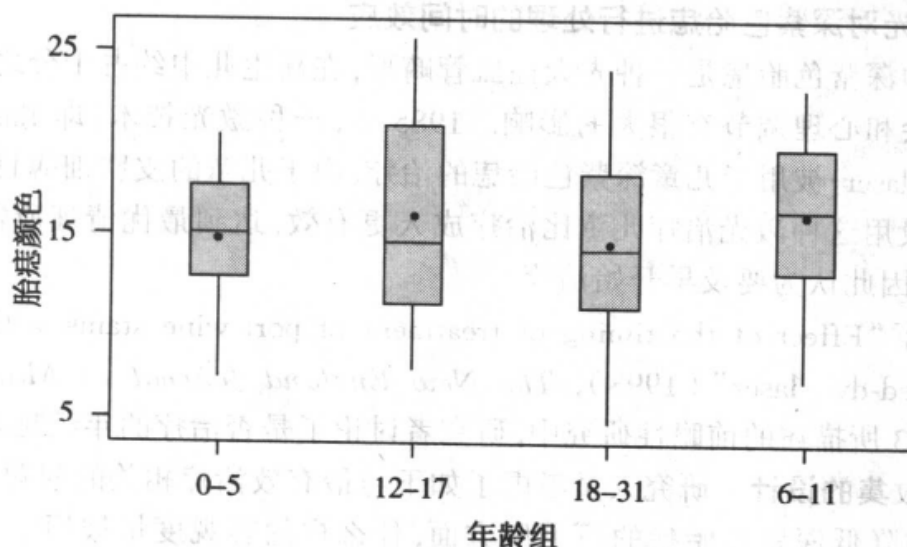


图 8.1 不同年龄组胎痣颜色的盒形图(均值由实心圆点表示)

数值是用文章所给的概括统计量模拟得到的。)

表 8.1 不同年龄组胎痣颜色的改善

患者	0~5 岁	6~11 岁	12~17 岁	18~31 岁
1	9.6938	13.4081	10.9110	1.4352
2	7.0027	8.2520	10.3844	10.7740
3	10.3249	12.0098	6.4080	8.4292
4	2.7491	7.4514	13.5611	4.4898
5	0.5637	6.9131	3.4523	13.6303
6	8.0739	5.6594	9.5427	4.1640
7	0.1440	8.7352	10.4976	5.4684
8	8.4572	0.2510	4.6775	4.8650
9	2.0162	8.9991	24.7156	3.0733
10	6.1097	6.6154	4.8656	12.3574
11	9.9310	6.8661	0.5023	7.9067
12	9.3404	5.5808	7.3156	9.8787
13	1.1779	6.6772	10.7833	2.3238
14	1.3520	8.2279	9.7764	6.7331
15	0.3795	0.1883	3.6031	14.0360
16	6.9325	1.9060	9.5543	0.6678
17	1.2866	7.7309	5.3193	2.7218
18	8.3438	7.9143	3.0053	2.3195
19	9.2469	1.8724	11.0496	1.6824
20	0.7416	12.5082	2.8697	1.8150
21	1.1072	6.2382	0.1082	5.9665
22		11.2424		0.5041
23		6.8404		5.4484
24		11.2774		

**数据的整理** 研究人员按 2.5 节的步骤准备数据以便进行统计分析,他们需要确认胎痣的颜色被正确记录以及所有的计算机文件和现场数据一致。

**数据的分析** 概括统计量在下面的表中给出,四个年龄组的盒形图如图 8.2。可以看出,12~17 岁组有最大的改善,而 6~11 岁组的改善稍微小一点。另外两个组至少比 12~17 岁组小 2 个单位。然而,从盒形图我们看到四个组在改善方面没有出现巨大的差异。我们将利用方差分析的过程进一步证实在四个年龄组是否存在很大的差异。

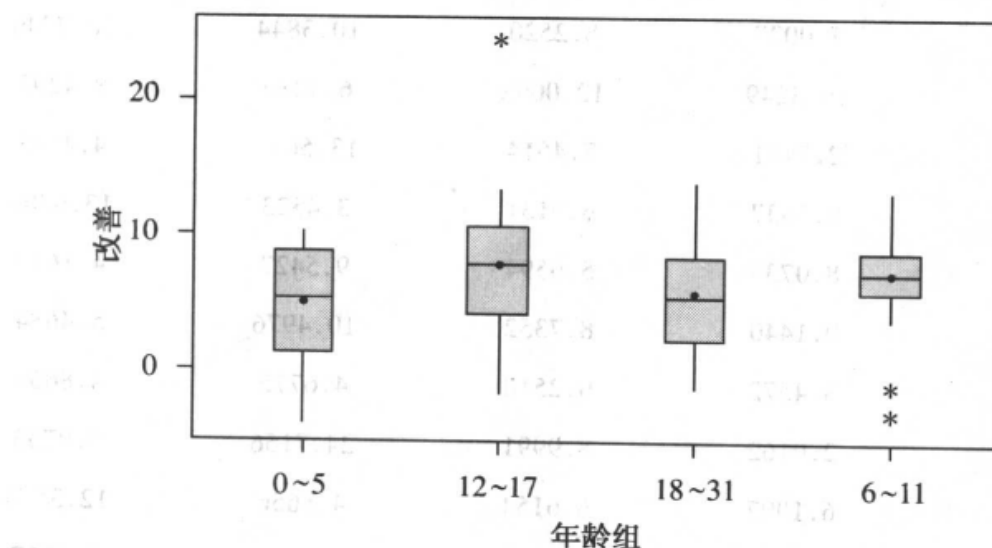


图 8.2 不同年龄组颜色改善的盒形图(均值由实心圆点表示)

Descriptive Statistics for Port-Wine Stain Case Study

Variable	N	Mean	Median	TrMean	StDev	SE Mean
0 - 5 Years	21	4.999	6.110	4.974	3.916	0.855
6 - 11 Years	24	7.224	7.182	7.262	3.564	0.727
12 - 17 Years	21	7.757	7.316	7.270	5.456	1.191
18 - 31 Years	23	5.682	4.865	5.531	4.147	0.865

Variable	Minimum	Maximum	Q1	Q3
0 - 5 Years	0.144	10.325	1.143	8.852
6 - 11 Years	0.188	13.408	5.804	8.933
12 - 17 Years	0.108	24.716	3.528	10.640
18 - 31 Years	0.504	14.036	2.320	8.429

从本节开始引入的例子,我们可以看出这种检验方法叫做方差分析的原因。假定希望从每个种族取 5 个工人的样本去比较三个种族小时工资的均值,我们从

每个总体中取容量为 5 的样本去介绍基本的思想,虽然这个样本容量太小。

假定样本数据(小时工资,以美元计)如表 8.2。这些数据能充分反映三个总体均值的差异吗?直接检查数据可看出在每个样本内数据变化很小,但各样本均值间的差异较大。由于和**样本内变差**比较,样本均值间的变化较大,我们直观地认为对应的均值是不同的。

表 8.2 三个样本均值的比较(样本内变差小)

来自于总体的样本		
1	2	3
5.90	5.51	5.01
5.92	5.50	5.00
5.91	5.50	4.99
5.89	5.49	4.98
5.88	5.50	5.02
$\bar{y}_1 = 5.90$	$\bar{y}_2 = 5.50$	$\bar{y}_3 = 5.00$

表 8.3 列出的样本均值与表 8.2 相同,但在每一个样本内数据的变差较大,且相对于样本内的变化**样本间变差**较小,依据这些数据就不大可能说对应总体的均值是不同的。

表 8.3 三个样本均值的比较(样本内变差大)

来自于总体的样本		
1	2	3
5.90	6.31	4.52
4.42	3.54	6.93
7.51	4.73	4.48
7.89	7.20	5.55
3.78	5.72	3.52
$\bar{y}_1 = 5.90$	$\bar{y}_2 = 5.50$	$\bar{y}_3 = 5.00$

图 8.3 表示了表 8.2 和表 8.3 中两个数据集的变化。图 8.3(a)明显表明表 8.2 中数据能反映总体的均值是不同的,而图 8.3(b)缺乏证据说明表 8.3 中数据

能反映总体的均值是不同的。

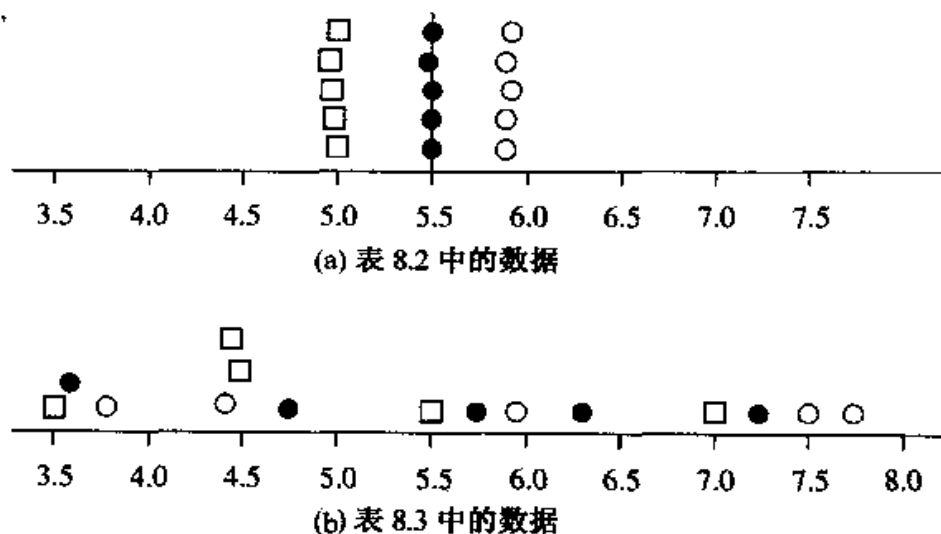


图 8.3 表 8.2 和表 8.3 中数据的点图:○, 样本 1 的观察值;  
●, 样本 2 的观察值;□, 样本 3 的观察值

借助于图 8.3, 从如上的讨论我们可以看出**方差分析**就是将样本均值之间的差异和样本内的变差进行比较, 从而判断样本均值之间的差异在统计上是否显著。详细的检验过程下节讨论。

## 8.2 两个以上总体均值的统计检验: 方差分析

在第六章, 我们提出了检验两个总体均值相等的方法。假定有两个正态总体 (1 和 2), 它们的方差均为  $\sigma^2$ , 均值分别为  $\mu_1, \mu_2$ , 用取自两个正态总体的容量为  $n_1$  和  $n_2$  的独立随机样本去检验零假设  $\mu_1 = \mu_2$ , 利用样本数据计算检验统计量

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p \sqrt{(1/n_1) + (1/n_2)}}$$

这里

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

是正态总体方差  $\sigma^2$  的一个组合估计, 犯第一类错误的概率为  $\alpha$  的拒绝域可由附录中的表 2 查得。

我们希望推广这种方法用来检验多于两个正态总体均值的相等, 而这里使用的检验方法仅对于两个均值的情形, 因此是不适当的, 所以我们将使用更一般的数据分析的方法, 即方差分析。通过下面的例子介绍它的应用。

调查全国五所不同学校的学生对工业污染的态度, 每个被选学生需回答指定

的一些问题,然后给出一个总分。假设每个学校有 25 个学生被调查,我们想去估计每个学校的学生平均分。

设学校 I 所有学生得分集记为总体 I,其均值为  $\mu_1$ ,容量为  $n_1 = 25$  的随机样本来自于总体 I。依此类推,有总体 II、III、IV、V,均值分别是  $\mu_2, \mu_3, \mu_4, \mu_5$ ,样本容量分别是  $n_2 = 25, n_3 = 25, n_4 = 25, n_5 = 25$ 。

从这 5 个样本,我们可以计算它们各自的样本均值和样本方差,结果如表 8.4。

表 8.4 关于 5 个总体的样本的结果

	总 体				
	I	II	III	IV	V
样本均值	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	$\bar{y}_5$
样本方差	$s_1^2$	$s_2^2$	$s_3^2$	$s_4^2$	$s_5^2$
样本容量	25	25	25	25	25

如果要检验总体均值的等式,即  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ ,那么我们可以比较所有可能的两个总体的均值。所以,若进一步确定 5 个总体均服从方差为  $\sigma^2$  的正态分布,我们可用 10 次  $t$  检验以比较总体的均值,将假设列出如下(见 6.2 节)。

#### 零假设

$$\begin{aligned} \mu_1 = \mu_2 \quad \mu_1 = \mu_3 \quad \mu_1 = \mu_4 \quad \mu_1 = \mu_5 \quad \mu_2 = \mu_3 \quad \mu_2 = \mu_4 \quad \mu_2 = \mu_5 \quad \mu_3 = \mu_4 \\ \mu_3 = \mu_5 \quad \mu_4 = \mu_5 \end{aligned}$$

这个检验方法的一个明显的缺点是既乏味又费时。然而,一个重要且不明显的缺点是随着  $t$  检验次数的增加,错误地拒绝至少一个假设的概率就增加。所以,虽然我们对每一个检验固定其犯第一类错误的概率为  $\alpha = 0.05$ ,但错误地拒绝至少一个假设的概率大于 0.05。换句话说,对 10 个假设的整体,犯第一类错误的概率大于 0.05。确实,可以证明这个概率可大到 0.40。

我们需要的是假设“所有五个总体的均值是相等的”的单个的假设检验,并要求这个检验比起那些  $t$  检验来不那么单调乏味,犯第一类错误的概率等于指定的值(如 0.05),这个检验方法即是方差分析。

方差分析方法要满足如下条件:

1. 五个总体都服从正态分布。
2. 五个总体的方差是相等的;即  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma^2$ 。
3. 五个独立随机样本分别来自于五个总体。

由条件 2,我们考虑量



$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 + (n_4 - 1)s_4^2 + (n_5 - 1)s_5^2}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1) + (n_4 - 1) + (n_5 - 1)}$$

$$= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 + (n_4 - 1)s_4^2 + (n_5 - 1)s_5^2}{n_1 + n_2 + n_3 + n_4 + n_5 - 5}$$

注意,这个量是

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

的推广,而此量可作对假设  $\mu_1 = \mu_2$  进行检验时两个总体的公共方差的估计(看 6.2 节)。所以,  $s_w^2$  表示公共方差  $\sigma^2$  的组合估计,它反映了五个总体内观察值的变异性大小(下标  $w$  是样本内的变异性“within-sample variability”的字头)。

下面我们考虑度量总体均值间变异性的一个量。若零假设  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$  为真,则 5 个总体是相同的,且均值为  $\mu$ 、方差为  $\sigma^2$ ,所以从 5 个总体中取出的单个样本相当于从一个总体中取出 5 个不同的样本。我们希望这些样本均值有什么变化? 如果这种变化太大,我们就拒绝假设  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ 。

为了评估 5 个样本均值的变化,我们需要知道来自正态总体的容量为 25 的随机样本得到的样本均值的抽样分布。由第四章,我们知容量为 25 的随机样本的样本均值  $\bar{y}$  的抽样分布的均值与总体的均值  $\mu$  相同,而方差为  $\sigma^2/25$ , 5 个容量为 25 的随机样本均如此,我们可以利用公式

$$5 \text{ 个样本均值的样本方差} = \frac{\sum_{i=1}^5 (\bar{y}_i - \bar{\bar{y}})^2}{5 - 1}$$

估计样本均值的分布的方差  $\sigma^2/25$ , 这里  $\bar{\bar{y}} = \sum_{i=1}^5 \bar{y}_i / 5$  是 5 个样本均值的平均。

注意,我们仅仅是把各个  $\bar{y}$  作为有 5 个观测值的一个样本,并计算了该样本的“样本方差”。如上式子是对  $\sigma^2/25$  的估计,从而  $25 \times (\text{均值的样本方差})$  是对  $\sigma^2$  的估计,将其记为  $s_B^2$ ; 下标  $B$  表示 5 个总体的样本均值间的变异性。

在零假设下, 5 个总体的均值是相等的,我们有  $\sigma^2$  的两个估计,分别为  $s_w^2$ ,  $s_B^2$ 。假定用比值

$$s_B^2 / s_w^2$$

作为检验假设  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$  的检验统计量,那么如果我们一次又一次重复试验并计算  $s_w^2$  及  $s_B^2$ ,  $s_B^2 / s_w^2$  的分布是什么样的呢?

对于我们的例子,  $s_B^2 / s_w^2$  服从  $F$  分布,  $s_B^2$  的自由度是  $df_1 = 4$ ,  $s_w^2$  的自由度是  $df_2 = 120$ , 关于该结论的证明超出了本书的讨论范围。我们将利用这个结果去检验零假设  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ 。

检验总体均值相等的检验统计量是

$$F = s_B^2 / s_w^2$$

当零假设为真时,  $s_w^2$  和  $s_B^2$  可作为  $\sigma^2$  的估计且  $F$  的取值在 1 附近。当零假设为假时, 由于总体均值间的不同,  $s_B^2$  将大于  $s_w^2$ 。所以, 在  $F = s_B^2 / s_w^2$  的分布的上尾部就拒绝零假设。对  $\alpha = 0.05$ ,  $F = s_B^2 / s_w^2$  的临界值是 2.45。(看图 8.4) 如果  $F$  的计算值落在拒绝域内, 我们就认为不是所有 5 个总体的均值是相等的。

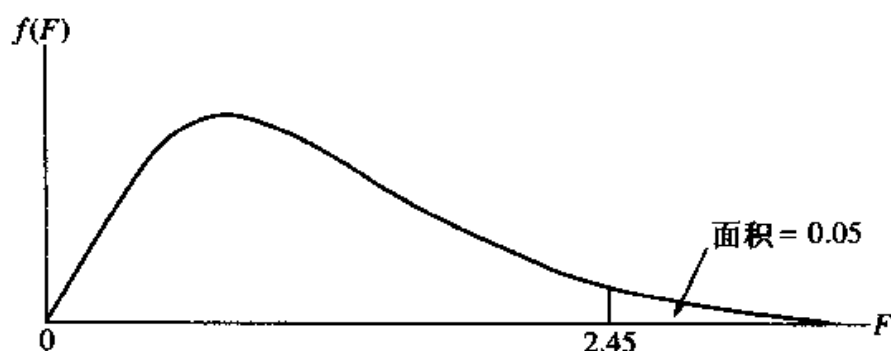


图 8.4  $\alpha = 0.05$ ,  $df_1 = 4$ ,  $df_2 = 120$  时,  $F$  的临界值

稍微修改一下公式, 这个方法就可以推广到检验方差相等的  $t$  (大于等于 2 的整数) 个正态总体均值的相等的假设。来自于各个总体的样本容量分别为  $n_1, n_2, \dots, n_t$ , 然后我们可以计算样本均值和样本方差。取备择假设为“至少有一个总体的均值和其他是不相同的”, 来检验零假设  $\mu_1 = \mu_2 = \dots = \mu_t$ 。

在表述一般化的检验方法之前, 我们引入一些关于  $s_B^2$  和  $s_w^2$  的公式中使用的记号。

对  $t$  个不同总体的每一个进行观察而得到  $t$  个随机样本的试验方法称为一个**完全随机化设计**。考虑从 5 个总体中的每一个进行 4 次观察的一个完全随机化设计。如果我们用  $y_{ij}$  表示总体  $i$  的第  $j$  次观察, 则可以把这个完全随机化设计的样本数据列在表 8.5 中。使用表 8.5, 在进行一个完全随机化设计的**方差分析 (AOV)**时, 我们可以引入一些有用的记号。

表 8.5 完全随机化设计的样本数据

总体	数据				均值
1	$y_{11}$	$y_{12}$	$y_{13}$	$y_{14}$	$\bar{y}_{1.}$
2	$y_{21}$	$y_{22}$	$y_{23}$	$y_{24}$	$\bar{y}_{2.}$
3	$y_{31}$	$y_{32}$	$y_{33}$	$y_{34}$	$\bar{y}_{3.}$
4	$y_{41}$	$y_{42}$	$y_{43}$	$y_{44}$	$\bar{y}_{4.}$
5	$y_{51}$	$y_{52}$	$y_{53}$	$y_{54}$	$\bar{y}_{5.}$

### 一个完全随机化设计的方差分析所需要的记号

$y_{ij}$ : 总体  $i$  的第  $j$  个样本观察值, 例如,  $y_{23}$  表示来自于总体 2 的第三个样本观察值。

$n_i$ : 总体  $i$  的样本观察值的个数。在我们的数据集中,  $n_1 = 4$  是总体 1 的样本观察值的个数, 类似地有  $n_2 = n_3 = n_4 = n_5 = 4$ 。然而, 样本容量不一定是相等的, 例如我们可以有  $n_1 = 12, n_2 = 3, n_3 = 6, n_4 = 10$  等等。

$n_T$ : 总样本容量,  $n_T = \sum n_i$ 。对表 8.5 中所给的数据,  $n_T = n_1 + n_2 + n_3 + n_4 + n_5 = 20$ 。

$\bar{y}_{i.}$ : 总体  $i$  的  $n_i$  个样本观察值的平均,  $\bar{y}_{i.} = \sum_j y_{ij} / n_i$ 。

$\bar{y}_{..}$ : 所有样本观察值的平均,  $\bar{y}_{..} = \sum_i \sum_j y_{ij} / n_T$ 。

用这些记号, 我们可以建立下列代数恒等式(虽然在下列计算  $s_B^2$  和  $s_w^2$  时要用到这些结果, 但这些恒等式的证明不在本书中列出), 令  $s_T^2$  是  $n_T$  个样本观察值  $y_{ij}$  的样本方差, 我们可以用量

$$\text{TSS} = \sum_{i=1}^t \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = (n_T - 1)s_T^2$$

来表示  $n_T$  个样本观察值  $y_{ij}$  关于总的均值  $\bar{y}_{..}$  的变异性。这个量被称作是关于总的均值的总平方和(记为 TSS), 在 TSS 中的双连加号表示的是对所有行和列的求和。

总平方和可以分解如下:

$$\sum_{i,j} (y_{ij} - \bar{y}_{..})^2 = \sum_{i,j} (y_{ij} - \bar{y}_{i.})^2 + \sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

等式右边的第一个量反映了观察值  $y_{ij}$  关于它的样本均值  $\bar{y}_{i.}$  的变异性。所以,

$$\text{SSW} = \sum_{i,j} (y_{ij} - \bar{y}_{i.})^2 = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_t - 1)s_t^2$$

是一个样本内的变异性的度量, SSW 被称为样本内平方和且用于计算  $s_w^2$ 。

总平方和等式中的第二个表达式反映了样本均值  $\bar{y}_{i.}$  关于总的均值  $\bar{y}_{..}$  的变异性, 这个量说明了样本均值间的差异, 被称为样本间平方和(SSB)且用于计算  $s_B^2$ 。

$$\text{SSB} = \sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

虽然关于 TSS, SSW, SSB 的公式很容易解释, 但计算却不太方便, 我们可以利用计算机软件包来计算。

关于  $t$  个总体的完全随机化设计的方差分析, 有如下的零假设和备择假设:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_t$$

$H_a$ :  $t$  个总体的均值至少有一个和其余的不同

量  $s_B^2$  和  $s_w^2$  可由下列公式计算

$$s_B^2 = \text{SSB}/(t-1) \quad s_w^2 = \text{SSW}/(n_T - t)$$

这里  $t-1$  和  $n_T - t$  分别表示  $s_B^2$  及  $s_w^2$  的自由度。

历史上,人们曾经用自由度去除平方和,并把所得的值称为均方,所以  $s_B^2$  经常被叫作样本间的均方,  $s_w^2$  被叫作样本内的均方。由于这些量都是偏差的平方的平均,所以被叫作均方。因为对  $t$  个样本的每一个都有  $\sum_j (y_{ij} - \bar{y}_{i.}) = 0$ , 在 SSW 中仅有  $n_T - t$  个线性无关的偏差  $y_{ij} - \bar{y}_{i.}$ , 所以 SSW 用  $n_T - t$  去除而不用  $n_T$ 。类似地,因为  $\sum_i (\bar{y}_{i.} - \bar{y}_{..}) = 0$ , 在 SSB 中仅有  $t-1$  个线性无关的偏差  $\bar{y}_{i.} - \bar{y}_{..}$ , 所以 SSB 用  $t-1$  去除。

如果

$$F = s_B^2/s_w^2$$

的值超出了对  $\alpha = \alpha$ ,  $df_1 = t-1$  及  $df_2 = n_T - t$  查表所得的值, 则拒绝零假设。

在完成了  $F$  检验后, 我们把方差分析总结在一张表中, 称为方差分析表, 简记为 AOV 表。AOV 表的格式如表 8.6, AOV 表的第一列列出了方差的来源, 第二列是与方差来源相关的平方和。我们已经证明了总平方和 (TSS) 能被分成两部分, 所以在 AOV 表中, SSW 和 SSB 相加必等于 TSS。表中的第三列是和方差来源相关的自由度, 我们可验证  $(t-1) + (n_T - t) = n_T - 1$ 。均方被列在第四列, 关于  $t$  个总体均值相等的  $F$  检验列在第五列。

表 8.6 完全随机化设计方差分析的一个例子

来源	平方和	自由度	均方	F 检验
样本间	SSB	$t-1$	$s_B^2 = \text{SSB}/(t-1)$	$s_B^2/s_w^2$
样本内	SSW	$n_T - t$	$s_w^2 = \text{SSW}/(n_T - t)$	
总 和	TSS	$n_T - 1$		

### 例 8.1

一个园艺学家调查三种不同种类苹果树(1, 2 和 3)树叶的含磷量, 从三种不同种类苹果树中的每一类取 5 个树叶的样本进行含磷量分析, 数据见表 8.7, 用这些数据检验假设“三种不同种类苹果树的平均含磷量水平是一样的”, 取  $\alpha = 0.05$ 。

表 8.7 三种不同种类苹果树树叶的含磷量

种类	含 磷 量					样本容量	均值	方差
1	0.35	0.40	0.58	0.50	0.47	5	0.460	0.00795
2	0.65	0.70	0.90	0.84	0.79	5	0.776	0.01033
3	0.60	0.80	0.75	0.73	0.66	5	0.708	0.00617
总和						15	0.648	

**解答** 零假设和备择假设是

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$H_a$ : 3 个总体的均值不全相等

样本容量是  $n_1 = n_2 = n_3 = 5$ , 且  $n_T = 15$ 。利用样本均值和样本方差, 样本内和样本间平方和是

$$\begin{aligned} \text{SSB} &= \sum_{i=1}^3 n_i (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &= 5(0.46 - 0.648)^2 + 5(0.776 - 0.648)^2 + 5(0.708 - 0.648)^2 = 0.277 \end{aligned}$$

$$\begin{aligned} \text{SSW} &= \sum_{i=1}^3 (n_i - 1) s_i^2 \\ &= (5 - 1)(0.00795) + (5 - 1)(0.01033) + (5 - 1)(0.00617) = 0.0978 \end{aligned}$$

从而  $\text{TSS} = \text{SSB} + \text{SSW} = 0.277 + 0.0978 = 0.3748$

关于这些数据的 AOV 表见表 8.8。由附录中的表 8 可查得对  $\alpha = 0.05$ ,  $df_1 = 2$ ,  $df_2 = 12$ ,  $F = s_B^2/s_w^2$  的临界值是 3.89。而我们计算得  $F$  的值为 17.25, 大于 3.89, 所以拒绝零假设。从数据中可以看出, 种类 1 的均值小于种类 2 和种类 3 的均值。

表 8.8 例 8.1 的方差分析表

来 源	平方和	自由度	均方	F 检验
样本间	0.277	2	$0.277/2 = 0.138$	$0.138/0.008 = 17.25$
样本内	0.0978	12	$0.0978/12 = 0.008$	
总 和	0.3748	14		

### 例 8.2

一个诊所的心理医生想要比较减少大学生敌意水平的三种方法, 他使用了某种测试(HLT)以度量敌意程度。测试中的高分表示敌意度大, 心理医生取出了试验中得到高分及分数较接近的 24 个学生。24 个学生中随机选取 8 个用第一种方法治疗, 余下的 16 个中随机选取 7 个用第二种方法治疗, 另外 9 个学生用第三种方法治疗, 所有的治疗均连续进行一个学期, 每一个学生在学期末都作一个 HLT 测试, 结果见表 8.9。利用这些数据做方差分析以确定是否三种方法的平均分有差异, 取  $\alpha = 0.05$ 。

表 8.9 HLT 测试得分

方法	测试分								均值	标准差	样本容量
1	96	79	91	85	83	91	82	87	86.750	5.625	8
2	77	76	74	73	78	71	80		75.571	3.101	7
3	66	73	69	66	77	73	71	70 74	71.000	3.674	9

**解答** 零假设和备择假设是

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$H_a$ : 3 个总体的均值不全相等

对  $n_1 = 8, n_2 = 7, n_3 = 9$ , 我们有总样本容量  $n_T = 24$ 。使用表中所给的样本均值, 可计算 24 个数据的总体平均值

$$\begin{aligned}\bar{y}_{..} &= \sum_{i=1}^3 n_i \bar{y}_{i.} / n_T = (8(86.750) + 7(75.571) + 9(71.000)) / 24 \\ &= 1861.997 / 24 = 77.5832\end{aligned}$$

利用此值及表 8.9 中的均值和标准差, 三个平方和计算如下:

$$\begin{aligned}\text{SSB} &= \sum_{i=1}^3 n_i (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &= 8(86.750 - 77.5832)^2 + 7(75.571 - 77.5832)^2 + 9(71 - 77.5832)^2 \\ &= 1090.6311\end{aligned}$$

$$\begin{aligned}\text{SSW} &= \sum_{i=1}^3 (n_i - 1) s_i^2 \\ &= (8 - 1)(5.625)^2 + (7 - 1)(3.101)^2 + (9 - 1)(3.674)^2 = 387.1678\end{aligned}$$

从而  $\text{TSS} = \text{SSB} + \text{SSW} = 1090.6311 + 387.1678 = 1477.80$ 。方差分析表见表 8.10。

表 8.10 例 8.2 中数据的方差分析表

来 源	平方和	自由度	均方	F 检验	p 值
样本间	1090.6311	2	545.316	545316/18.4366 = 29.58	<0.001
样本内	387.1678	21	18.4366		
总 和	1477.80	23			

对  $\alpha = 0.05$ ,  $df_1 = 2$ ,  $df_2 = 21$ , 由附录中表 8 可查得  $F$  的临界值为 3.47, 而  $F$  的计算值为 29.58, 大于临界值 3.47。所以, 拒绝零假设  $H_0$ 。由于从表 8 知, 对  $df_1 = 2$ ,  $df_2 = 21$ ,  $F$  的最大值为 9.77, 其对应的  $\alpha = 0.001$ , 因而, 本题相当强地拒绝  $H_0$ 。从三个样本均值中可看出, 方法 1 的均值大于方法 2 及方法 3 的均值。研究人员还需进一步确定是所有三个总体的均值不相同, 或方法 2 和方法 3 的均值相同, 亦即我们要给出三种方法的均值以及它们差的置信区间, 这将提供给研究人员所关心的三种方法差异程度的信息。在下一章, 将给出构造这些类型推断的技巧。下面的计算机输出结果与我们这里得到的结果是一致的。注意在这些计算机打印出的结果里, 平方和的名字与上面表中的不同, 样本间的平方和是用总体的名字标出的, 在此例中为“METHOD”, 样本内平方和则常标为“ERROR”(误差平方和)。

#### General Linear Models Procedure

#### Class Level Information

Class	Levels	Values
METHOD	3	1 2 3

Number of observations in data set = 24

Dependent Variable: SCORE

Source	DF	Sum of Squares	F Value	Pr > F
Model	2	1090.61904762	29.57	0.0001
Error	21	387.21428571		
Corrected Total	23	1477.83333333		

#### 练习

#### 应用

**8.1(农业)** 一个大的实验室有四种测试土壤样本 pH 值的装置。实验室想要确定四种装置测得的平均值是否有差异, 实验人员选取了 24 种在研究中 pH 值已知的土壤样品, 每个装置随机指定 6 个样品进行测试, 记录下的响应值为测试值与已知值的差。这些值见下表

装置	样 本						样本 容量	均值	标准差
	1	2	3	4	5	6			
A	-0.307	-0.249	-0.079	-0.019	-0.136	-0.324	6	-0.1605	0.1767
B	-0.176	0.125	-0.013	0.082	0.091	0.459	6	0.0947	0.2091
C	0.137	-0.603	0.240	-0.050	0.318	0.154	6	0.1227	0.1532
D	-0.042	0.690	0.201	0.166	0.219	0.407	6	0.2735	0.2492

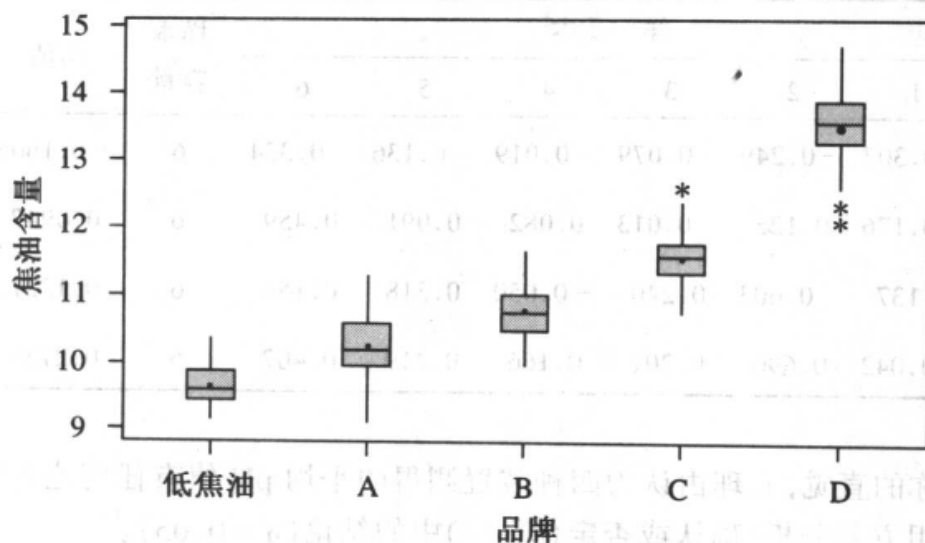
- 凭你的直觉,有理由认为四种装置测得的平均 pH 值有任何差异吗?
- 利用方差分析,确认或否定你在(a)中的结论( $\alpha = 0.05$ )。
- 计算(b)中  $F$  检验的  $p$  值。
- 在什么条件下,你在(b)、(c)得到结论才是有效的?
- 假设 24 个土壤样品的 pH 值有很大的不同,随机指定土壤样品到不同的装置会发生什么问题?

**8.2(商业)** 一个香烟制造商做广告称,发明了一种新品牌的香烟,此香烟的焦油含量比其他主要品牌的低。为验证他的说法,一个消费者测试机构随机从 4 个主要品牌香烟中选择了 100 支,从新品牌中选择了 100 支,香烟的焦油含量结果如下:

品牌	$\bar{y}_i$	$s_i$	$n_i$
低焦油	9.64	0.291	100
A	10.22	0.478	100
B	10.77	0.372	100
C	11.57	0.352	100
D	13.59	0.469	100

- 从盒形图提供的信息看,新品牌的香烟平均焦油含量比其他品牌低吗?
- 利用计算机输出的结果,5 种品牌的香烟平均焦油含量有很大的不同吗( $\alpha = 0.05$ )?
- (b)中检验统计量的  $p$  值是多少?
- 对(b)中的检验,犯第一类错误的实际含义是什么?





表中数据的盒形图见上(均值用实圆点表示)

#### One-Way Analysis of Variance for Exercise 8.2

##### Analysis of Variance for Tar Cont

Source	DF	SS	MS	F	P
Brand	4	941.193	235.298	1478.39	0.000
Error	495	78.784	0.159		
Total	499	1019.976			

##### Individual 95 % CIs for Mean

Based on Pooled StDev

Level	N	Mean	StDev	-	+	-	+	-	+	-	+	-	+	-	+
1	100	9.644	0.291	*	)										
2	100	10.221	0.478			*	)								
3	100	10.775	0.372					(	*						
4	100	11.570	0.352							*	)				
5	100	13.592	0.469											*	)

Pooled StDev = 0.399

9.6 10.8 12.0 13.2

### 8.3 完全随机化设计中观测值的模型

在这一节,我们考虑完全随机化设计的模型(有时称为单向分类模型)。这个

模型将说明方差分析的检验方法是恰当的;我们可以把此模型认为是一个物理背景的数学描述,这也可使我们对物理过程产生的数据进行计算机模拟。

我们假定涉及的样本观察值及从中抽取样本的总体满足如下条件:

1. 样本是独立随机样本,一个样本的结果对另一个样本的观察值没有任何影响。

2. 每一个样本来自于正态分布。

3. 总体  $i$  的均值和方差分别为  $\mu_i$  和  $\sigma^2 (i=1, 2, \dots, t)$ 。

图 8.5 描绘了三个条件均满足时的一种情形。总体的分布是正态分布且具有相同的标准差,其中总体 III 和 IV 的均值相同, I 和 II 的均值不相同。总之,我们已假定  $t$  个总体是独立的正态分布,它们的均值不相同,方差均为  $\sigma^2$ 。

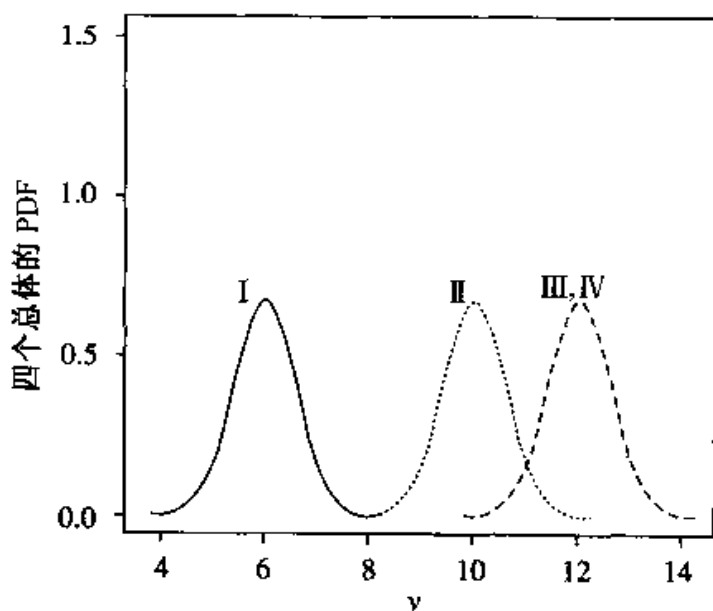


图 8.5 满足方差分析假设的 4 个总体的分布

现在我们能写出满足三个假设的一个模型(方程),如从前,令  $y_{ij}$  表示总体  $i$  的第  $j$  个样本观察值。

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

这个模型表明  $y_{ij}$  是三项的和,项  $\mu$  表示总平均值,是一个未知常数;  $\alpha_i$  表示总体  $i$  的效应,它也是一个未知常数。这里  $\mu$  表示的是所有  $t$  个总体的总均值,也就是所有  $t$  个总体包含的观察值构成的总体的均值,  $\alpha_i$  表示的是总体  $i$  对观察值的总变异的效应,  $\mu$ 、 $\alpha_i$  是未知常数,它们可由研究或实验的数据来估计。  $\epsilon_{ij}$  表示的是  $y_{ij}$  关于第  $i$  个总体的均值  $\mu_i$  的随机偏差,  $\epsilon_{ij}$  经常被作为误差项,但这个“误差”不能被解释成试验中所犯的误差,其实  $\epsilon_{ij}$  表示  $y_{ij}$  关于它们的均值  $\mu_i$  的随机变异。“误差”这个词只不过表明了来自于  $t$  个总体的观察值之间的差异不只是它们的均值

之间的差异这一事实。假设  $\epsilon_{ij}$  服从独立的正态分布, 均值为 0, 标准差为  $\sigma_\epsilon$ , 独立性这一条件可被解释如下:  $y_{ij}$  的观察值与  $\mu_i$  的偏差的大小对其他观察值的偏差的大小没有影响。

由于  $y_{ij}$  是来自于第  $i$  个总体的观察值, 所以它的均值为  $\mu_i$ , 而  $\epsilon_{ij}$  服从均值为 0 的分布,  $y_{ij}$  的均值记为  $E(y_{ij})$  且有

$$\mu_i = E(y_{ij}) = E(\mu + \alpha_i + \epsilon_{ij}) = \mu + \alpha_i + E(\epsilon_{ij}) = \mu + \alpha_i$$

也就是说  $y_{ij}$  是从均值为  $\mu + \alpha_i$  的总体随机挑选的观察值。所以, 效应  $\alpha_i$  表示了第  $i$  个总体的均值  $\mu_i$  和总体均值  $\mu$  的偏差, 因而  $\alpha_i$  可以是正数、零或负数, 即总体  $i$  的均值可以大于、等于或小于总体的均值  $\mu$ 。可以证明,  $t$  个总体中的每一个的方差均为  $\sigma_\epsilon^2$ 。最后, 因为  $\epsilon$  是正态分布, 所以  $t$  个总体均服从正态分布。关于单向分类的假设综合如下表 8.11。

表 8.11 关于完全随机化设计的假设

总体	总体均值	总体方差	样本观测值
1	$\mu + \alpha_1$	$\sigma_\epsilon^2$	$y_{11}, y_{12}, \dots, y_{1n_1}$
2	$\mu + \alpha_2$	$\sigma_\epsilon^2$	$y_{21}, y_{22}, \dots, y_{2n_2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$t$	$\mu + \alpha_t$	$\sigma_\epsilon^2$	$y_{t1}, y_{t2}, \dots, y_{tn_t}$

单向方差分析的零假设是  $\mu_1 = \dots = \mu_t$ 。利用我们的模型, 它等价于零假设

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_t = 0$$

如果  $H_0$  为真, 则所有总体均有未知均值  $\mu$ 。确实许多教科书把后者作为一个完全随机化设计的方差分析的零假设。对应的备择假设是

$$H_a: \text{至少有一个 } \alpha_i \text{ 不为零。}$$

在这一节, 我们已经对一个完全随机化设计的方差分析的模型做了简略的描述, 虽然一些作者从其他渠道考虑模型, 但我们相信, 这是方差分析讨论所必需的部分。

我们已经对选择数据的总体, 或等价地说对数据被生成的试验施加了几个条件。所以在用方差分析表做推断前, 需要验证这些条件。在第七章, 我们讨论了如何利用 Hartley  $F_{\max}$  检验或 Levine 检验去检验方差是相等的这一条件。除非总体是严重偏倚的或是厚尾的, 在大样本情形下, 正态性这一条件不象方差是相等的这一条件要求那么严格; 在小样本条件下, 正态性和方差是相等的这些条件变得更加关键。由于通常对每个总体没有足够的观察值用于检验正态性和方差是相等的这些条件是否满足, 这就提出了一个问题。在下一节, 我们将讨论一个技巧, 至少能

部分地克服这个问题。另外,在本章最后一节,将提出当总体的方差不相等以及总体不服从正态分布时,用来代替方差分析的一些方法。如第六章中所指出的,我们讨论的三个条件中最关键的是数据的独立性。这个条件可以通过仔细地实施研究或试验,避免观察值之间的相依性得到满足。在从  $t$  个总体中随机选择数据时,一定要小心,以保证数据是随机的且从一个总体中得到的样本值不依赖于从另一个总体中得到的样本值。如在试验中,  $t$  个处理是随机地分配给试验单元的,我们就需要确保每个处理真是随机地被分配的。还有,在试验中必须保证试验单元之间不相互影响,以免影响它们的响应值。

## 8.4 方差分析条件的检查

总体方差是相等的及总体服从正态分布的假设在本书中的几个地方被用到,如比较两个总体均值的  $t$  检验及完全随机化设计中方差分析的  $F$  检验。

让我们先考虑比较  $t$  个总体均值的试验,每个总体的随机样本是独立的。我们重申这里的总体均服从方差为  $\sigma_i^2$  的正态分布,但均值可能不同。我们利用第七章中的 Hartley 检验或 Levine 检验去检验方差是相等的这一假设。

这里几点需要提及,许多实际工作者日常不使用 Hartley 检验,一个原因是此检验对正态性的违反特别敏感,所以在检验方差相等时,不得不很关心方差分析的另一个假设(总体的正态性)。幸运的是,在第六章我们提到,在样本容量几乎相等的情况下,总体方差齐性(相等性)的假设不是很关键的,方差可以差别很大而方差分析的  $p$  值稍微有一点变化。所以我们只在更为极端的情形下推荐使用 Hartley 检验或 Levine 检验。在这些总体方差齐性是一个问题的极端情形下,利用数据的变换可能稳定这些方差,此时可以用方差分析进行推断。

同我们在第五章及第六章讨论的一样,当样本容量相对较大时,我们可以利用正态概率图或盒形图去检测总体分布的正态性。然而,在许多试验中,来自于每个总体的样本容量为 5 到 10,在这种情形下,就总体分布是否是正态的这一点来说,图所揭示的结果就不大可靠。利用前节所介绍的模型去考虑,正态条件的评估可使用残差分析来进行。

从这个模型,我们有  $y_{ij} = \mu + \alpha_i + \epsilon_{ij} = \mu_i + \epsilon_{ij}$ , 可以写成  $\epsilon_{ij} = y_{ij} - \mu_i$ 。如果方差相等的条件是可用的,则这些  $\epsilon_{ij}$  就是来自于一个正态总体的随机样本。虽然  $\mu_i$  是一个未知常数,但如果我们用  $\bar{y}_i$  估计  $\mu_i$  并令

$$e_{ij} = y_{ij} - \bar{y}_i.$$

那么我们可以用  $e_{ij}$  去估价正态性假设。即使每个  $n_i$  是很小的,我们也有  $n_T$  个残差,它们可以提供充分多的值去评价正态性条件,我们可以在盒形图里或正态图里画  $e_{ij}$  去评价数据是否是由正态总体产生的。

## 例 8.3

因为许多 HMO 组织或者不付心理健康的费用,或者只提供最低限度的关照,牧师和神父们经常需要给遭受心理疾病的人以忠告。一个由不同宗教派别所共有的组织想要确定不同宗教派别的神职人员相对于心理疾病的原因的认识水平,抽取了三个随机样本,第一个样本包括了 10 个卫里工会的牧师,第二个包括了 10 个天主教的神父,第三个包括了 10 个犹太教的神父。30 个人中每一个都被测试,使用标准的笔试以检测他们对心理疾病原因的知识,测验分数列于表 8.12,三个宗教人员的平均测试分是否有显著的不同?

表 8.12 神职人员心理疾病知识的测试分数

职员	卫里工会	天主教	犹太教
1	62	62	37
2	60	62	31
3	60	24	15
4	25	24	15
5	24	22	14
6	23	20	14
7	20	19	14
8	13	10	5
9	12	8	3
10	6	8	2
$\bar{y}_i$	30.50	25.90	15.00
$s_i$	21.66	20.01	11.33
$n_i$	10	10	10
中位数( $\tilde{y}_i$ )	23.5	21	14

**解答** 在进行三个均值的方差分析之前,我们需要评估方差分析所要求的条件是否被满足。图 8.6 是关于心理疾病分数的盒形图,可看出数据稍微有点向右偏。因而,我们要评价正态性条件。为此,需要知道残差  $e_{ij} = y_{ij} - \bar{y}_i$ , 例如  $e_{11} = y_{11} - \bar{y}_1 = 62 - 30.50 = 31.50$ , 其余的  $e_{ij}$  在表 8.13 中给出。

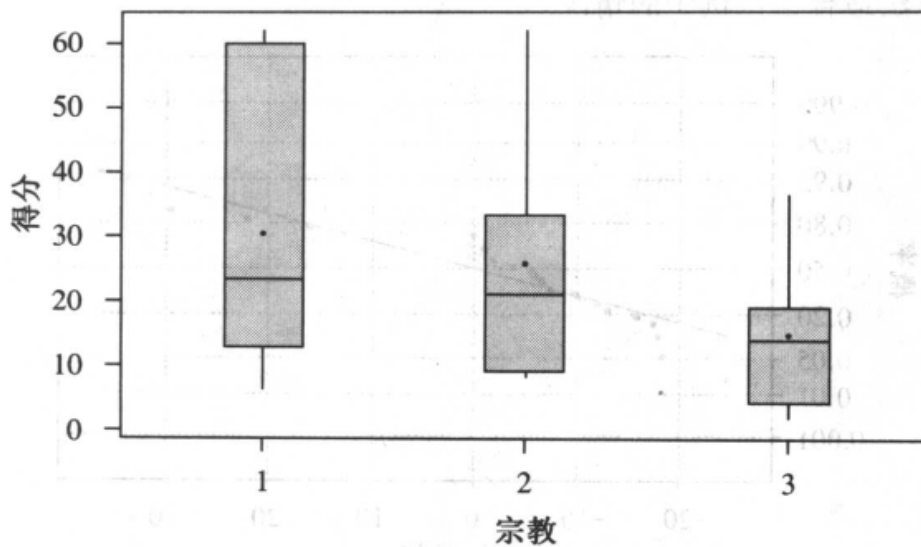


图 8.6 三个宗教得分的盒形图(均值用实心圆点表示)

表 8.13 神职人员关于心理疾病知识得分的残差  $e_{ij}$ 

职员	卫里工会	天主教	犹太教
1	31.5	36.1	22.0
2	29.5	36.1	16.0
3	29.5	-1.9	0.0
4	-5.5	-1.9	0.0
5	-6.5	-3.9	-1.0
6	-7.5	-5.9	-1.0
7	-10.5	-6.9	-1.0
8	-17.5	-15.9	-10.0
9	-18.5	-17.9	-12.0
10	-24.5	-17.9	-13.0

残差被画在图 8.7 和图 8.8 中。图 8.8 的盒形图表明 30 个残差中有三个离群值。如果残差是正态分布的一个随机样本,数据值中有 10% 为离群值是不太可能的,这也被图 8.7 所示的正态概率图所证实。图 8.7 表明残差没有集中在直线附近。更进一步,正态性检验的  $p$  值小于 0.001,这表明不符合正态性,所以结论是数据有非正态特征。在 8.6 节,我们将提供一个方法,以代替方差分析中的  $F$

检验,这个方法适合于该例中的情况。

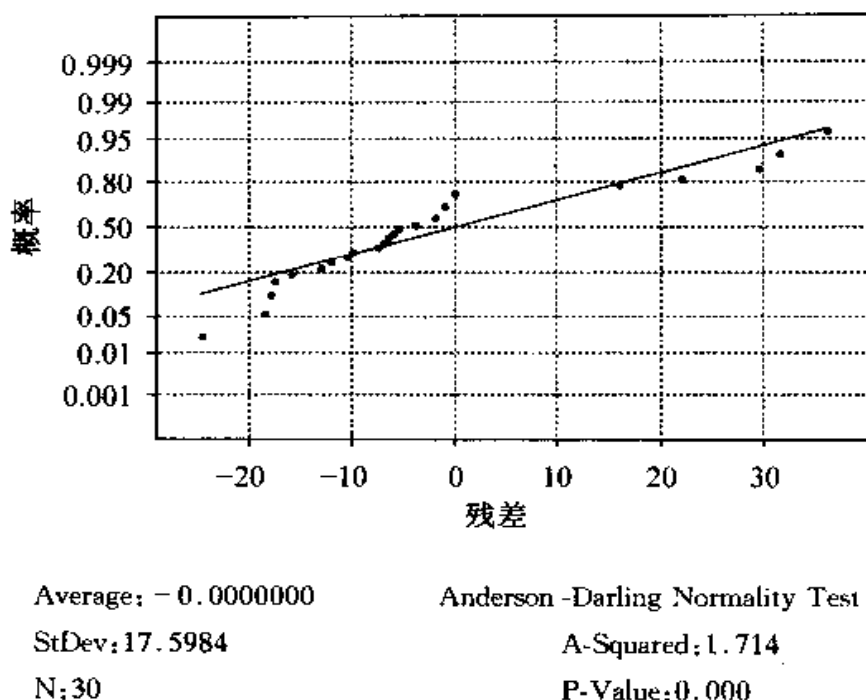


图 8.7 残差的正态概率图

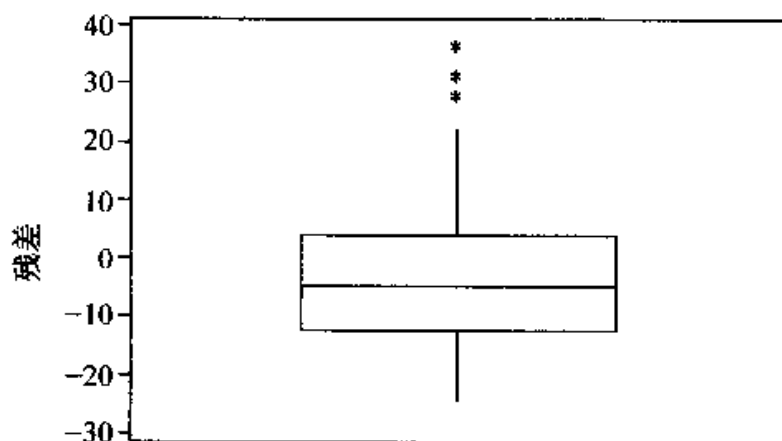


图 8.8 残差的盒形图

由于数据可能是非正态的,使用 Hartley  $F_{\max}$  检验去检验方差的相等是不适当的,所以我们将使用 Levine 检验。对 Levine 检验公式进行考察发现,如果我们把数据  $y_{ij}$  换为  $z_{ij} = |y_{ij} - \bar{y}_i|$ , 这里  $\bar{y}_i$  是第  $i$  个数据集的样本中位数, Levine 检验等价于把方差分析应用到  $z_{ij}$  的  $F$  检验,因而我们可以简单地使用方差分析的公式去对 Levine 检验进行计算。用表 8.12 的中位数,  $z_{ij}$  列在表 8.14。

表 8.14 变换后的数据集:  $z_{ij} = |y_{ij} - \bar{y}_i|$ 

职员	卫里工会	天主教	犹太教
1	38.5	41	23
2	36.5	41	17
3	36.5	3	1
4	1.5	3	1
5	0.5	1	0
6	0.5	1	0
7	3.5	2	0
8	10.5	11	9
9	11.5	13	11
10	17.5	13	12
$\bar{z}_{i.}$	15.70	12.90	7.40
$s_i$	15.80	15.57	8.29

使用在表中给出的样本均值, 我们来计算 30 个数据值的总均值:

$$\bar{z}_{..} = \sum_{i=1}^3 n_i \bar{z}_{i.} / n_T = [10(15.70) + 10(12.90) + 10(7.40)] / 30 = 360 / 30 = 12$$

用该值及表 8.14 中的均值及标准差, 我们可以计算下列平方和:

$$\begin{aligned} \text{SSB} &= \sum_{i=1}^3 n_i (\bar{z}_{i.} - \bar{z}_{..})^2 \\ &= 10(15.70 - 12)^2 + 10(12.90 - 12)^2 + 10(7.40 - 12)^2 = 356.6 \end{aligned}$$

$$\begin{aligned} \text{SSW} &= \sum_{i=1}^3 (n_i - 1) s_i^2 \\ &= (10 - 1)(15.80)^2 + (10 - 1)(15.57)^2 + (10 - 1)(8.29)^2 = 5,047.10 \end{aligned}$$

均方是  $\text{MSB} = \text{SSB} / (t - 1) = 178.3$  和  $\text{MSW} = \text{SSW} / (n_T - t) = 186.9$ 。最后, 我们从  $L = \text{MSB} / \text{MSW} = 178.3 / 186.9 = 0.95$  得到 Levine 检验统计量的值, 对  $\alpha = 0.05$ , 从  $\text{df}_1 = 2, \text{df}_2 = 27$  的  $F$  表中可查得  $L$  的临界值是 3.35, 所以我们接受标准差相等这一零假设。因为在  $\text{df}_1 = 2, \text{df}_2 = 27$  的  $F$  表中, 最小值是 1.46, 对应的概率是 0.25, 因而  $p$  值大于 0.25。这样, 我们有很高的置信度认为三个总体有相同的方差。

在 8.6 节, 我们将提出 Kruskal-Wallis 检验, 它能用于总体是非正态的, 但在零假设下有相同分布的情形。这个检验至少要求总体有相同的方差, 所以



Kruskal-Wallis检验对总体有不同方差的情形是不适当的。下一节将提出一个在总体方差不相等时检验总体均值差异的检验方法。

**案例中的数据分析** 研究的目的是评价胎痣的治疗对于年轻的人是否比对年长的人更有效。数据的概括统计量如下。

Descriptive Statistics for Port-Wine Stain Case Study

Variable		N	Mean	Median	TrMean	StDev	SE Mean
0 - 5	Years	21	4.999	6.110	4.974	3.916	0.855
6 - 11	Years	24	7.224	7.182	7.262	3.564	0.727
12 - 17	Years	21	7.757	7.316	7.270	5.456	1.191
18 - 31	Years	23	5.682	4.865	5.531	4.147	0.865

Variable		Minimum	Maximum	Q1	Q3
0 - 5	Years	0.144	10.325	1.143	8.852
6 - 11	Years	0.198	13.408	5.804	8.933
12 - 17	Years	0.108	24.716	3.528	10.640
18 - 31	Years	0.504	14.036	2.320	8.429

从图 8.1 看到, 盒形图有相同的宽度, 没有离群值且须长相同。对四个年龄组来说, 均值和中位数相近, 所以方差分析的假设似乎是满足的。为验证这个观察, 我们计算了残差并把它们画在一个概率图上(见图 8.9)。从此图可看到, 除一个

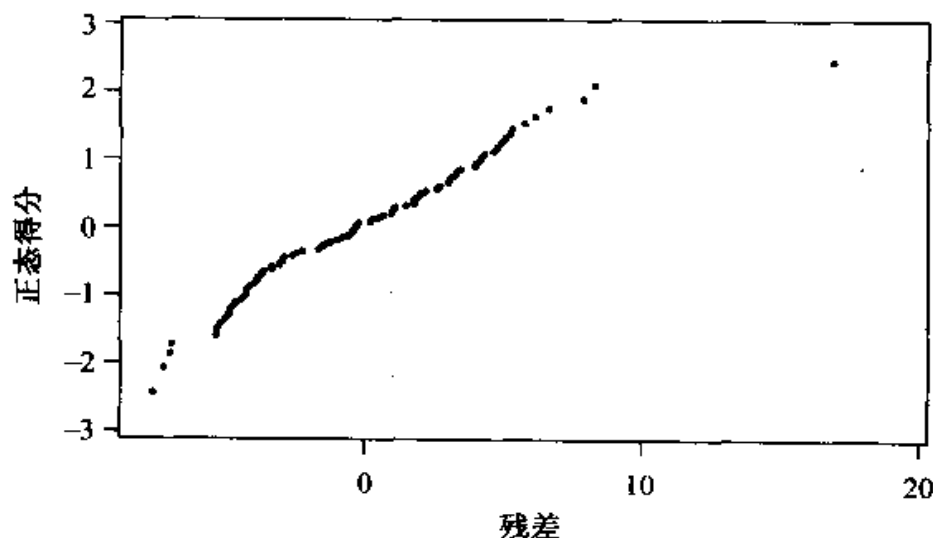


图 8.9 案例中残差的正态概率图

数据点外都落在直线附近,这足以说明,皮肤颜色改进的四个总体服从正态分布。

其次,我们使用 Hartley 检验或 Levine 检验去检查方差相等这一假设。对 Hartley 检验,我们有

$$F_{\max} = \frac{(5.46)^2}{(3.564)^2} = 2.35$$

对  $\alpha = 0.05$ 、 $df_2 = 20$ 、 $t = 4$ ,  $F_{\max}$  的临界值是 3.29。由于样本容量不等,这个检验仅是一个近似。然而,样本容量非常接近,分别为 21、21、23 和 24。因为  $F_{\max}$  不大于 3.29,没有足够的证据表明四个总体的方差是不同的。对于 Levine 检验,有  $L = 1.050$ ,其  $p$  值是 0.375,这说明与 Hartley 检验是一致的。我们有理由认为方差分析中的正态性和方差相等的条件是满足的。数据的独立性可以通过就研究进行的方式与研究者进行讨论来检查。治疗的次序以及胎痣颜色的评估应当确保一个病人皮肤颜色的改善不影响另一个病人颜色的改善。

在这种类型的试验中,可能会由于设备、技术人员的偏差、病人间的任何关系及其他相似的因素引起数据相关的问题出现。

研究假设是治疗后胎痣颜色的平均改善对四个年龄组是不同的:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$H_a$ : 至少有两个均值不相等

关于方差分析表的计算结果如下:

One-Way Analysis of Variance for Improvement in Stain Color

Source	DF	SS	MS	F	P
Age Group	3	108.0	36.0	1.95	0.128
Error	85	1572.5	18.5		
Total	88	1680.5			

Individual 95% CIs for Mean  
Based on Pooled StDev

Level	N	Mean	StDev	
0 - 5	21	4.999	3.916	{-----*-----}
06 - 11	24	7.224	3.564	{-----*-----}
12 - 17	21	7.757	5.456	{-----*-----}
18 - 31	23	5.682	4.147	{-----*-----}

Pooled StDev = 4.301

4.0      6.0      8.0      10.0

从中看出,  $F$  检验的  $p$  值是 0.128, 所以四个年龄组的平均改善没有显著的不同。我们可以计算置信度为 95% 的平均改善的置信区间, 四个区间在计算机的输出结果中已给出。它们是用组合标准差  $\sigma = \sqrt{MSW} = \sqrt{18.5} = 4.30$  得到的,  $df = 85$ , 所以区间有形式

$$\bar{y}_{i.} \pm \frac{t_{0.025, 85} \sigma}{\sqrt{n_i}} = \bar{y}_{i.} \pm \frac{(1.99)(4.30)}{\sqrt{n_i}}$$

四个区间表示如下:

年龄组	$\bar{y}_{i.}$	95% 置信区间
0~5	4.999	(3.13, 6.87)
6~11	7.224	(5.48, 8.97)
12~17	7.757	(5.89, 9.62)
18~31	5.682	(3.90, 7.47)

从置信区间能得到四个组中胎痣颜色的平均改善估计的总效应, 最年轻一组的改善最小, 但其上界大于改善最大的年龄组的下界。这种类型的判决的问题在于这些置信区间不是同时置信区间, 所以不能给出结论的置信水平。在下一章, 将提出均值的同时置信区间, 因而可确定不同年龄组是否有显著的差异。但是, 在我们研究的情形中, 因为方差分析中  $F$  检验接受零假设, 所以结论是不同年龄组没有显著差异。

研究者不能确证早期治疗比晚期治疗有效的假设。他们的确作出结论, 暗示治疗时间应选择在儿童期。虽然面部的胎痣可以在早期有效、安全地得到治疗, 但迟后的治疗也有相似的结果。所以开始治疗的年龄应基于对预期的益处及治疗时的不适的谨慎衡量。

**报告结论** 我们需要写一份报告, 以总结在关于胎痣治疗的前瞻性研究中所得到的发现。其内容应包括:

1. 研究目的的陈述。
2. 研究设计及数据收集过程的描述。
3. 为什么 100 个病人中的 11 个没有包括在数据分析中的讨论。
4. 数据集的数值和图形概括。
5. 所有推断方法论的描述:
  - 方差分析表和  $F$  检验;
  - 均值的基于  $t$  的置信区间;
  - 所有推断方法使用的必要条件被满足的验证。

6. 结果和结论的讨论。
7. 相对于从前的研究,新发现的解释。
8. 进一步研究的建议。
9. 列出数据集。

## 8.5 其他的分析方法:数据变换

**样本数据的变换** 是把原始尺度下的测量值系统地转化为新尺度的一个过程。例如,如果原始变量是  $y$ , 且与各个处理下该变量的方差是不相等(非齐)的,那么,就可能需要引进一个新的变量,诸如  $\sqrt{y}$ 、 $\log y$  或其他变换产生的变量。

如何选择适当的变换呢?这不是一件容易的事,需要实验者在应用领域内有大量的经验。尽管有这样的困难,但对选择合适的变换我们有几条指导原则。

经常我们感兴趣的各个总体的方差是不同的,并且随总体均值的大小而变化。例如,可能总体均值越大,总体的方差也越大。当我们能识别方差怎样随总体均值变化时,我们就可以定义一个合适的变换从变量  $y$  到新变量  $y_T$ 。三个特别情形见表 8.15。

表 8.15 的第一行说明,如果  $y$  是一个泊松随机变量,则  $y$  的方差等于  $y$  的均值,所以均值不同,方差也不同。稳定化方差的变换是  $y_T = \sqrt{y}$ , 或者如果泊松分布的均值是小的(5 以下),变换  $y_T = \sqrt{y + 0.375}$  更好。

表 8.15 获得一致方差的变换

$\mu$ 和 $\sigma^2$ 间的关系	$y_T$	$y_T$ 的方差(对给定的 $k$ )
$\sigma^2 = k\mu$ (当 $k = 1$ , $y$ 是一个泊松分布随机变量)	$y_T = \sqrt{y}$ 或 $\sqrt{y + 0.375}$	$1/4; (k = 1)$
$\sigma^2 = k\mu^2$	$y_T = \log y$ 或 $\log(y + 1)$	$1; (k = 1)$
$\sigma^2 = k\pi(1 - \pi)$ (当 $k = 1/n$ , $y$ 是二项分布随机变量)	$y_T = \arcsin \sqrt{y}$	$1/4n; (k = 1/n)$

### 例 8.4

海洋生物学家正在研究墨西哥湾大量的虾及商用鱼的重要变化,密西西比河进入海湾的区域是最受关注的区域之一。生物学家假设富营养的水流入海湾,这些水主要包含来自中西部农田的氮,这导致了浮游生物赖以生存的藻类的快速生长。细菌则以浮游生物排泄物和死藻类为生,消耗了水中的氧。不大移动的海洋生物因缺氧而死亡时,善于移动的海洋生物则逃离了这一区域。为了调查这一状

况以确定从密西西比河入口处开始的四个区域的平均溶解氧含量(以 ppm 计), 在四个区域的每一个中, 从海底 12 米处取 10 个水样的一个随机样本, 样本数据见表 8.16。生物学家想要检验平均氧含量是否越靠近密西西比河入口处越低。

表 8.16 离入口处 4 个距离的平均溶解氧的含量(以 ppm 计)

水 样	距入海口的距离			
	1 公里	5 公里	10 公里	20 公里
1	1	4	20	37
2	5	8	26	30
3	2	2	24	26
4	1	3	11	24
5	2	8	28	41
6	2	5	20	25
7	4	6	19	36
8	3	4	19	31
9	0	3	21	31
10	2	3	24	33
均值	$\bar{y}_{1.} = 2.2$	$\bar{y}_{2.} = 4.6$	$\bar{y}_{3.} = 21.2$	$\bar{y}_{4.} = 31.4$
标准差	$s_1 = 1.476$	$s_2 = 2.119$	$s_3 = 4.773$	$s_4 = 5.522$

- 对  $\alpha = 0.05$ , 作总体方差相等的检验。
- 如果必要的话, 对数据进行变换, 使得变换后的新数据集中的观察值有相同的方差。

### 解答

- 图 8.10 描述了数据的盒形图, 数据没有明显的偏倚或厚尾, 所以我们使用 Hartley  $F_{\max}$  检验,  $\alpha = 0.05$

$$F_{\max} = 5.222^2 / 1.476^2 = 14.0$$

对  $\alpha = 0.05$ ,  $t = 4$ ,  $df_2 = 10 - 1 = 9$ ,  $F_{\max}$  的临界值是 6.31。因为  $F_{\max}$  的值大于 6.31, 所以拒绝总体方差相等这一假设。

- 检查一下样本均值  $\bar{y}_{i.}$  和样本方差  $s_i^2$  间的关系。

$$s_1^2 / \bar{y}_{1.} = 0.99 \quad s_2^2 / \bar{y}_{2.} = 0.97 \quad s_3^2 / \bar{y}_{3.} = 1.06 \quad s_4^2 / \bar{y}_{4.} = 0.97$$

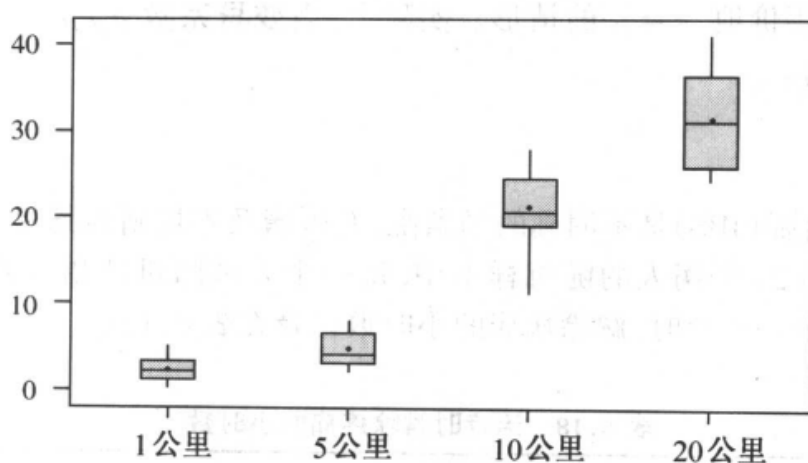


图 8.10 1~20 公里的盒形图(均值用实圆点表示)

所以  $\sigma_i^2 = k\mu_i$ ,  $k \approx 1$ 。从表 8.15, 取变换为  $y_T = \sqrt{y + 0.375}$ ,  $y_T$  的值以及它们的均值和标准差见表 8.17。虽然原数据有不同的方差, 但正象表 8.17 指出的那样, 样本方差均接近于 0.25。

表 8.17 表 8.16 中数据的变换:  $y_T = \sqrt{y + 0.375}$ 

水 样	距入海口的距离			
	1 公里	5 公里	10 公里	20 公里
1	1.173	2.092	4.514	6.114
2	2.318	2.894	5.136	5.511
3	1.541	1.541	4.937	5.136
4	1.173	1.837	3.373	4.937
5	1.541	2.894	5.327	6.432
6	1.541	2.318	4.514	5.037
7	2.092	2.525	4.402	6.031
8	1.837	2.092	4.402	5.601
9	0.612	1.837	4.623	5.601
10	1.541	1.837	4.937	5.777
均 值	1.54	2.19	4.62	5.62
方 差	0.24	0.22	0.29	0.24

表 8.15 中的第二个变换是针对总体方差近似等于总体均值的平方这样一个

试验的情形,或等价地, $\sigma = \mu$ 的情形。实际上,当变异系数  $\sigma_i/\mu_i$  为常数时,对数变换总是合适的。

### 例 8.5

急性肠综合症( IBS)是不明确的肠紊乱,有腹痛及不规则的肠运动的特征。在周期性得 IBS 的 24 个病人的随机样本中,每一个人被随机地指定到三个治疗组 A、B、C 中的一个。治疗时,解除病痛的小时数记录在表 8.18。

表 8.18 治疗时解除疼痛的小时数

处理		
A	B	C
4.2	4.1	38.7
2.3	10.7	26.3
6.6	14.3	5.4
6.1	10.4	10.3
10.2	15.3	16.9
11.7	11.5	43.1
7.0	19.8	48.6
3.6	12.6	29.5
$\bar{y} = 6.46$	$\bar{y} = 12.34$	$\bar{y} = 27.35$
$s = 3.22$	$s = 4.53$	$s = 15.66$

- 用  $\alpha = 0.05$ , 检验个总体方差间的差异。
- 因为  $y$  值均非零,使用变换  $y_T = \ln y$  ( $\ln$  表示以  $e$  为底的自然对数) 稳定化方差。
- 对变换后的数据,计算样本均值和样本标准差。

### 解答

- 关于零假设  $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$  的 Hartley  $F_{\max}$  检验是

$$F_{\max} = 15.66^2/3.22^2 = 245.24/10.37 = 23.63$$

$F_{\max}$  的计算值超过了 6.94 (对  $\alpha = 0.05$ ,  $t = 3$ ,  $df_2 = 7$  的查表值), 所以拒绝  $H_0$ , 即总体的方差是不相等的。

- 变换后的数据列在表 8.19。注: 自然对数使用计算器或计算机程序来计算。

表 8.19 表 8.18 中数据的自然对数

处理		
A	B	C
1.435	1.411	3.656
0.833	2.370	3.270
1.887	2.660	1.686
1.808	2.342	2.332
2.322	2.728	2.827
2.460	2.442	3.764
1.946	2.986	3.884
1.281	2.534	3.384

c. 变换后数据的样本均值和标准差列在表 8.20。关于变换后数据方差齐性的 Hartley  $F_{\max}$  检验是

$$F_{\max} = 0.77^2 / 0.46^2 = 2.80$$

$F_{\max}$  的计算值是 2.80, 它比查表值 6.94 小, 所以我们接受  $H_0$ , 即没有充分的证据说明总体的方差有差异, 因而变换产生的数据中, 三个方差接近相等。

表 8.20 表 8.19 中数据的样本均值和标准差

	处理		
	A	B	C
样本均值	1.75	2.43	3.10
样本方差	0.54	0.46	0.77

在表 8.15 中所列的第三种变换对以比率或分数记录的数据是特别合适的。在第四章, 我们介绍了二项分布, 这里  $y$  指的是  $n$  次试验中的成功数,  $\hat{\pi} = y/n$  是  $\pi$  的估计, 这里  $\pi$  表示的是总体中具有此特征的试验单元的比例。虽然在研究二项分布时没有提到  $\hat{\pi}$ , 但它的方差由  $\pi(1-\pi)/n$  给出。所以, 如果对应的变量是  $\hat{\pi}$ , 它是由  $n$  个观察值组成的随机样本中的成功率, 则  $\hat{\pi}$  的方差将随抽取样本的总体的  $\pi$  值而变化, 见表 8.21。



表 8.21 对  $\pi$  的几个值及  $n=20$  时  $\hat{\pi}$  的方差

$\pi$ 的值	$\pi(1-\pi)/n$	$\pi$ 的值	$\pi(1-\pi)/n$
0.01	0.0005	0.3	0.0105
0.05	0.0024	0.4	0.0120
0.1	0.0045	0.5	0.0125
0.2	0.0080		

因为  $\hat{\pi}$  的方差关于  $\pi=0.5$  是对称的, 所以对  $\pi=0.7, n=20, \hat{\pi}$  的方差是 0.0105, 与  $\pi=0.3$  时一样。类似地, 对  $\pi>0.5$  的值, 可确定  $\pi(1-\pi)/n$ 。需注意的一件重要的事情是, 如果总体中  $\pi$  的值在 0.3~0.5 附近, 则  $\hat{\pi}$  的方差有很小的差异。然而, 对  $\pi$  的值较大或较小时,  $\hat{\pi}$  的方差有很大的不同。对这些情形, 应考虑对样本比率进行变换以稳定方差的可能性。

我们介绍的变换是  $\arcsin \sqrt{\pi}$ , 即变换样本比率成为一个角, 此角的正弦是  $\sqrt{\pi}$ , 一些试验者用度表示这些角, 而另外一些用弧度表示, 为了一致, 我们总是用弧度表示角。附录中表 9<sup>①</sup> 给出了对  $\hat{\pi}$  的不同值反正弦的计算结果。

### 例 8.6

为了评估选民就 FBI 领导者是否应该有一个固定的任期(比如 10 年)的意见, 进行了一项全国的民意测验。由于地理位置的不同, 意见可能有一定差异。对这个民意测验, 全国被分为四个地区(西北 NW、西南 SW、东北 NE、东南 SE)。在四个区域中, 每个区域有 6 个城市统计区(SMSA), 从每个统计区中抽取 100 个注册选民的随机样本。下列数据是 24 个 SMSA 的样本比例, 使用  $y_T = 2\arcsin \sqrt{\pi}$  变换数据。

地区	SMSA						均值	标准差
	1	2	3	4	5	6		
东北	0.13	0.20	0.23	0.05	0.14	0.31	0.177	0.0903
东南	0.57	0.47	0.47	0.51	0.53	0.20	0.458	0.1321
西北	0.30	0.10	0.07	0.13	0.17	0.23	0.167	0.0860
西南	0.53	0.72	0.70	0.63	0.79	0.87	0.707	0.1191

① 附录中的表 9 给出的是  $2\arcsin \sqrt{\pi}$ 。

**解答** 使用计算器、计算机或附录中的表 9, 变换数据如下表:

地区	SMSA						均值	标准差
	1	2	3	4	5	6		
东北	0.74	0.93	1.00	0.45	0.77	1.18	0.845	0.2515
东南	1.71	1.51	1.51	1.59	1.63	0.93	1.480	0.2799
西北	1.16	0.64	0.54	0.74	0.85	1.00	0.822	0.2307
西南	1.63	2.03	1.98	1.83	2.19	2.40	2.010	0.2693

使用方差分析的方法, 可对四个区域的意见进行比较。

我们需对观察到的**样本比例为 0 或 1**的情形进行说明。在这些情形下, 建议分别用  $1/4n$  和  $1 - 1/4n$  代替在计算中对应的样本比。

在本节, 我们讨论了在进行方差分析前, 怎样通过数据变换解决非常数值方差的问题。这样做的另一个附加的好处是本节中的变换有时也减少了数据的非正态性。还有, 有时用这些变换, 由于严重偏倚或离群值所引起的非正态性也可消除。对两个非正态的独立随机样本, 可用 Wilcoxon 秩和检验(第六章)比较两个总体。对多于两个的独立样本的数据, 可用 Kruskal-Wallis 检验(8.6 节)处理非正态性问题。注意这些检验也都是建立在样本数据的变换(秩变换)的基础上。

### 练习

**8.3** 参考例 8.6, 使用反正弦变换分析样本数据以确定在四个地理区域是否有差异, 取  $\alpha = 0.05$ 。

**8.4** 参考例 8.4, 作变换分析样本数据, 确定氧含量是否与距密西西比河入海口的距离有关。

**8.5** 参见例 8.5。在许多的方差差异不太大的情况下, 变换后的数据总体均值的方差分析比较结果同由原始数据得到的结果相似。在这些情况下, 研究人员倾向于忽略变换, 因为变换后的数据对于研究人员来说不大好接受, 从而用被换后的数据得到的变换后总体均值的置信区间也不好解释。对这个问题, 一个可能补救的办法是利用变换后的数据构造一个置信区间, 然后对于该区间的端点进行逆变换, 从而得到一个其中的值与原始数据的测量值有相同意义的置信区间。

- 检验假设: 三种治疗病人病痛的平均小时数是否有差异,  $\alpha = 0.05$ , 用原始数据。
- 对三种治疗方案, 求均值的置信度为 95% 的置信区间。
- 用变换后的数据重新做分析(a)和(b)。

- d. 评价假设检验的结果有何不同。  
e. 在(c)中, 作一区间端点的逆变换, 同(b)中的区间做一比较。

## 8.6 另一种非参数方法: Kruskal-Wallis 检验

秩和检验的概念可以推广到多于两个总体的比较。具体说来, 设  $n_1$  个观察值随机取自总体 1,  $n_2$  个取自总体 2,  $\dots$ ,  $n_k$  个取自总体  $k$ 。我们想要检验假设:  $k$  个样本来自相同的分布。下列的检验方法是合适的, 该方法有时被称作 Kruskal-Wallis 检验。

**两个以上总体的秩和检验的推广**

$H_0$ :  $k$  个分布是相同的

$H_a$ : 不是所有的分布都相同

$$\text{T.S.: } H = \frac{12}{n_T(n_T + 1)} \sum_i \frac{T_i^2}{n_i} - 3(n_T + 1)$$

这里  $n_i$  是样本  $i$  的观察值数 ( $i = 1, 2, \dots, k$ ),  $n_T$  是总的样本容量, 即

$$n_T = \sum_i n_i, \quad T_i \text{ 为样本 } i \text{ 在总的样本观察值中的秩的和。}$$

R.R.: 对于指定的  $\alpha$  值, 如果  $H$  超过  $\alpha = \alpha, df = k - 1$  的  $\chi^2$  的临界值, 则拒绝  $H_0$ 。

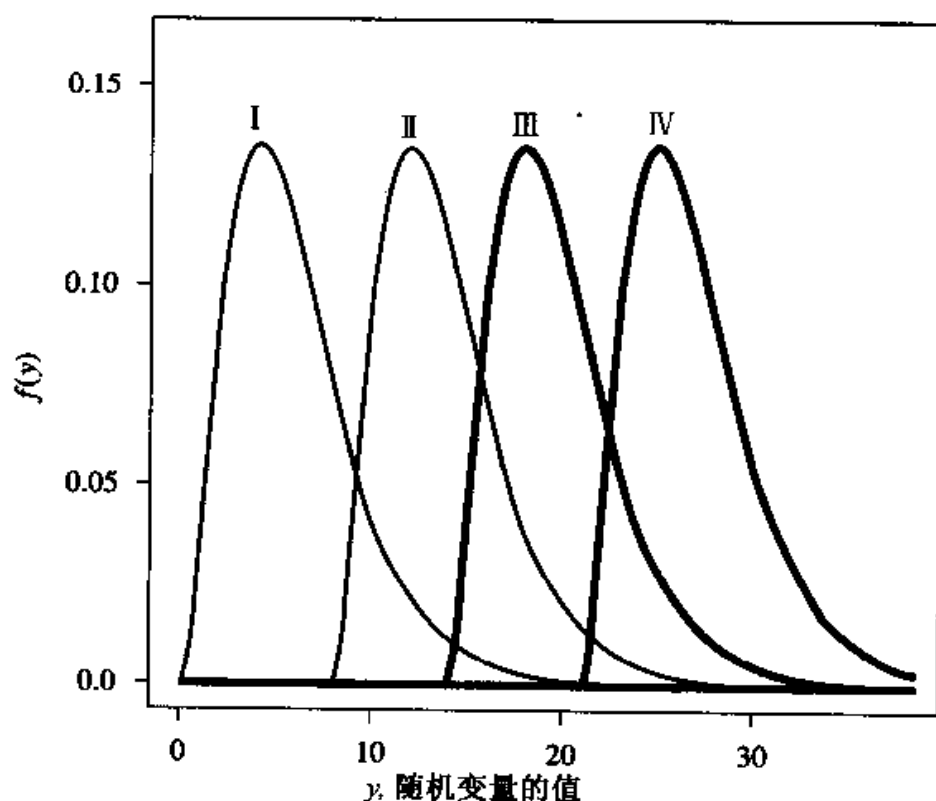


图 8.11 形状相同、位置不同的四个偏态的总体分布

注:当样本观察值的秩有大量相等时,用

$$H' = \frac{H}{1 - \left[ \sum_j (t_j^3 - t_j) / (n_T^3 - n_T) \right]}$$

这里  $t_j$  是第  $j$  个相等秩组中的观察值数。

图 8.11 列出了在 Kruskal-Wallis 检验的备择假设下总体的分布。

### 例 8.7

参看例 8.3。在那里我们已确定笔试成绩不是正态分布,所以把 Kruskal-Wallis 检验用于表 8.12 中的数据集。

用数据确定对于心理疾病起因的了解,三个测试组是否有差异,取  $\alpha = 0.01$ 。

**解答** 关于此例的研究假设和零假设如下:

$H_a$ : 三组中至少有一个和其他组不同

$H_0$ : 三组没有差异(即分数的样本来自同一总体)

在计算  $H$  之前,首先从低到高排列 30 个测试分。从表 8.22 可看到,最小分数是 2,其秩为 1,相似地,给分数 3、4、6 的秩分别为 2、3、4,有两个都是 8 分,由于这两个分数有秩 5 和 6,我们指定这两个分数的秩都为 5.5,也就是 5 和 6 的平均数。用相同的方法,可指定其他测试分数的秩。表 8.22 列出了 30 个测试分数及其秩(圆括号里的数)。

表 8.22 心理疾病知识的笔试成绩

职员	卫理公会派教徒	天主教徒	犹太教徒
1	62(29)	62(29)	37(25)
2	60(26.5)	62(29)	31(24)
3	60(26.5)	24(21)	15(13.5)
4	25(23)	24(21)	15(13.5)
5	24(21)	22(18)	14(11)
6	23(19)	20(16.5)	14(11)
7	20(16.5)	19(15)	14(11)
8	13(9)	10(7)	5(3)
9	12(8)	8(5.5)	3(2)
10	6(4)	8(5.5)	2(1)
秩和	182.5	167.5	115

从表 8.22, 三组的秩和分别是 182.5、167.5、115, 所以  $H$  的计算值是

$$H = \frac{12}{30(30+1)} \left( \frac{(182.5)^2}{10} + \frac{(167.5)^2}{10} + \frac{(115)^2}{10} \right) - 3(30+1)$$

$$= \frac{12}{930} (3330.625 + 2805.625 + 1322.5) - 93 = 3.24$$

因为含有相同大小的数据, 我们用  $H'$ , 并和  $H$  相比较。为此, 列出  $g$  组相同的秩, 见下表:

秩	组	$t_i$
1	1	1
2	2	1
3	3	1
4	4	1
5.5, 5.5	5	1
7	6	1
8	7	1
9	8	1
11, 11, 11	9	1
13.5, 13.5	10	1
15	11	1
16.5, 16.5	12	1
18	13	1
19	14	1
21, 21, 21	15	1
23	16	1
24	17	1
25	18	1
26.5, 26.5	19	1
29, 29, 29	20	1

从这些信息, 计算量

$$\begin{aligned}
& \sum_i \frac{(t_i^3 - t_i)}{n_T^3 - n_T} \\
&= \frac{(2^3 - 2) + (3^3 - 3) + (2^3 - 2) + (2^3 - 2) + (3^3 - 3) + (2^3 - 2) + (3^3 - 3)}{30^3 - 30} \\
&= 0.0036
\end{aligned}$$

代入此值到  $H'$  得

$$H' = H / (1 - 0.0036) = 3.24 / 0.9964 = 3.25$$

所以,即使有多于一半的值涉及秩的相等,  $H'$  和  $H$  的值几乎相等。对  $\alpha = 0.05$ ,  $df = k - 1 = 2$ ,  $\chi^2$  的临界值可由附录中的表 7 查得,此值是 5.991。因此,接受  $H_0$ , 即三个组的测试分数没有显著差异。有趣的是 Kruskal-Wallis 检验的  $p$  值是 0.198, 而方差分析中应用到原始数据的  $F$  检验的  $p$  值是 0.168。因而,即使数据不服从正态分布,方差分析中的  $F$  检验对于正态分布的偏离是稳健的。只有当数据有很大的偏倚或厚尾时, Kruskal-Wallis 检验和方差分析中的  $F$  检验才是不同的。

### 练习

**8.6(园艺)** 一组研究人员想要比较一个果园中树龄 4 年的 5 个不同品种 (A、B、C、D、E) 的橘子树的产量(磅), 每个品种取一个容量为 7 的随机样本, 这些树的产量如下:

A	B	C	D	E
13	27	40	17	36
19	31	44	28	32
39	36	41	41	34
38	29	37	45	29
22	45	36	15	25
25	32	38	13	31
10	44	35	20	30

- 用检验和数据图, 确定是否满足方差分析的条件。
- 作方差分析, 检验零假设: 5 个品种的平均产量相同, 取  $\alpha = 0.01$ 。
- 用 Kruskal-Wallis 检验去检验零假设: 5 个品种有相同的产量分布, 取  $\alpha = 0.01$ 。

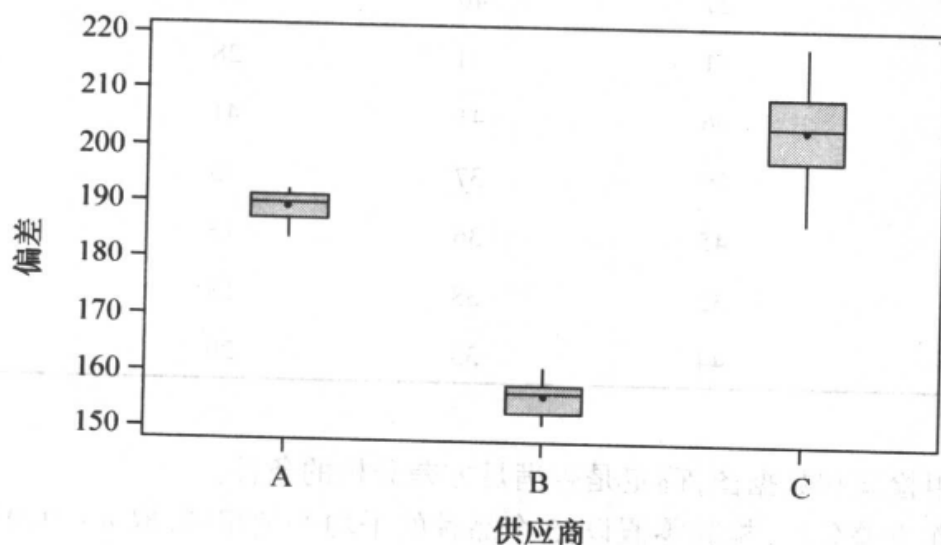
d. 在(b)和(c)中得到的结论是否一致?

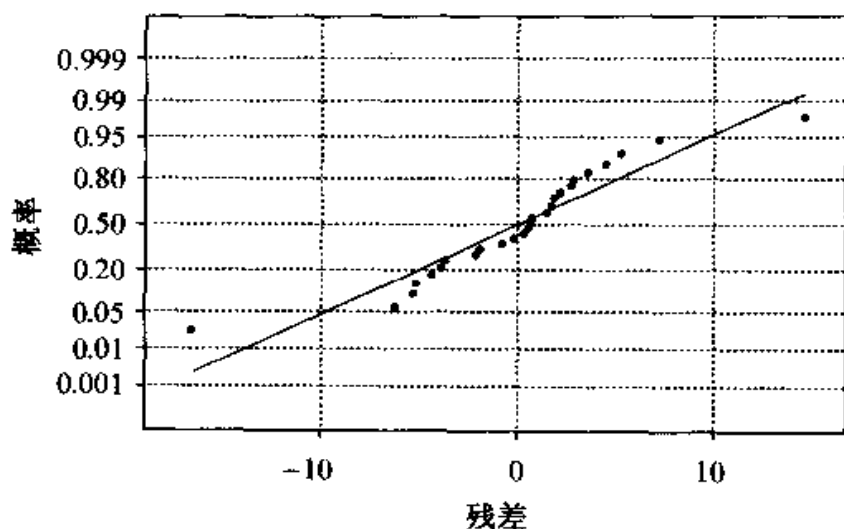
8.7 用方差分析和 Kruskal-Wallis 检验检验的研究假设有何不同?

8.8(工程) 在柔软的隐形眼镜镜片生产中, 镜片的实际强度需要十分接近消费者所适用的指标值。在文章“An ANOM-type test for variances from normal populations,”[ *Technometrics* (技术度量)(1997), 39:274~283]里, 就镜片强度的一致性, 对用三个供应商提供的原材料生产的镜片作了比较。下表列出的是使用三个供应商的材料生产的镜片强度指标值的偏差。

供应商	镜片								
	1	2	3	4	5	6	7	8	9
A	189.9	191.9	190.9	183.8	185.5	190.9	192.8	188.4	189.0
B	156.6	158.4	157.7	154.1	152.3	161.5	158.1	150.9	156.9
C	218.6	208.4	187.1	199.5	202.0	211.1	197.6	204.4	206.8

- 对这三个供应商, 偏差的分布有显著的差异吗? 取  $\alpha = 0.01$ 。
- 用下面给出的合适的检验或图, 评价当对数据用方差分析以确定对于三个供应商平均偏差是否有显著的差异时, 方差分析的条件是否满足。
- 用  $\alpha = 0.05$  作方差分析, 然后把你的结果与(a)的结论作比较。
- 假定如果镜片的平均偏差之间的差别超过 20 个单位时, 就会对于镜片的生产造成商业上的后果, 那么这三个供应商的原材料之间的差异有实际意义吗?





Average: -0.0000000

Anderson-Darling Normality Test

StDev: 5.53971

A-Squared: 0.718

N: 27

P-Value: 0.054

## 8.7 小 结

在这一章,我们提出了一些用于对  $t$  个总体均值进行比较的方法,推广了第六章的结果。从  $t$  个总体的每一个中取出的是独立随机样本,样本内的变异用  $s_w^2 = \text{SSW}/(n_T - t)$  来衡量,类似地,样本间的变异由  $s_B^2 = \text{SSB}/(t - 1)$  来衡量。

是接受还是拒绝  $t$  个总体的均值相等这一零假设,依赖于  $F = s_B^2/s_w^2$  的计算值。在  $H_0$  下,  $s_B^2$  和  $s_w^2$  都是  $\sigma_e^2$  的估计,  $\sigma_e^2$  是所有  $t$  个总体的共同方差。在备择假设下,  $s_B^2$  是  $\sigma_e^2 + \theta$  的估计,这里  $\theta$  是一正的量,  $s_w^2$  还是  $\sigma_e^2$  的估计量。所以对于  $F$  的大值就拒绝  $H_0$ ,  $F$  的临界值可从附录中的表 8 用  $df_1 = t - 1$ 、 $df_2 = n_T - t$  查得。这个检验方法被称作方差分析,通常总结在一个方差分析表中。

你可能对如下问题还有疑问,假定我们拒绝了  $H_0$ ,则至少有一个均值和其他的不同,是哪些与别的不同呢?本章没有回答这个问题,在第九章将通过多重比较去讨论此问题。

在这一章,对完全随机化设计,我们讨论了进行方差分析的假定。其中样本的独立随机性是绝对必要的。因为我们讨论的是均值,并且对适当的样本容量中心极限定理成立,所以对于正态性的假定要求不严。只有当样本容量有大的差别时,方差相等的假定才是重要的;样本容量相等或几乎相等时,对该假定可以放松要求。检验方差相等用  $F_{\max}$  统计量  $s_{\max}^2/s_{\min}^2$  或 Levine 检验。



有时,样本数据表明总体的方差是不同的,那么当总体均值与总体标准差间的关系是已知的或可猜测时,可以利用表 8.15 中提供的变换,把样本观测值  $y$  转化为一个新值  $y_T$  以稳定化总体方差。这些变换包括方根、对数、反正弦以及许多其他变换。

本章中的讨论还有很多细致的内容没有包括进来。但这些材料对于训练初级研究人员,以使他们认识到研究项目中的假定以及如何选择一种合适的分析方法(诸如非参数统计方法、Kruskal-Wallis 检验等)或应用一个变换到样本数据来说是足够的。

### 重要公式

#### 1. 完全随机化设计的方差分析

$$SSB = \sum_i n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SSW = \sum_{i,j} (y_{ij} - \bar{y}_{i.})^2 = \sum_i (n_i - 1) s_i^2$$

$$TSS = \sum_{i,j} (y_{ij} - \bar{y}_{..})^2 = SSB + SSW$$

#### 2. 完全随机化设计的模型

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad \text{这里 } \mu_i = \mu + \alpha_i$$

#### 3. 模型应满足的条件

- $t$  个总体都服从正态分布。
- $\sigma_1^2 = \cdots = \sigma_t^2 = \sigma^2$ 。
- 数据由  $t$  个独立随机样本构成。

#### 4. 检查是否满足条件

- 正态性:画残差  $e_{ij} = y_{ij} - \bar{y}_{i.}$  的图。
- 方差的齐性:Hartley 检验或 Levine 检验。
- 独立性:仔细回顾试验或研究是怎样进行的。

#### 5. 关于总体均值 $\mu_i$ 的置信度为 $100(1-\alpha)\%$ 的置信区间

$$\bar{y}_{i.} \pm t_{\frac{\alpha}{2}, n_T - t} \frac{\hat{\sigma}}{\sqrt{n_i}}, \text{ 其中 } \hat{\sigma} = \sqrt{MSW}$$

#### 6. Kruskal-Wallis 检验(总体分布是非正态的)

$H_0$ :  $k$  个总体的分布相同

$H_a$ :  $k$  个总体的分布不全相同

$$T.S. = \frac{12}{n_T(n_T + 1)} \sum_{i=1}^k \frac{T_i^2}{n_i} - 3(n_T + 1)$$

## 补充练习

**8.9(制造业)** 公司的研究人员作了一个试验,以比较正在改造中的五条生产线中的每一条有重要缺陷的产品数目。他们在改造期间连续监视生产,记录下每条生产线每天有重要缺陷的数目,数据如下:

生产线				
1	2	3	4	5
34	54	75	44	80
44	41	62	43	52
32	38	45	30	41
36	32	10	32	35
51	56	68	55	58

- 对每一个样本计算  $\bar{y}$  和  $s^2$ , 有非常数方差问题出现吗? 使用 Hartley 检验, 取  $\alpha = 0.05$ 。
- 作平方根变换, 并对变换后的数据作一分析。
- 对生产线间的差异, 写出你的结论。

**8.10** 对练习 8.9 中的数据作 Kruskal-Wallis 检验, 此检验支持练习 8.9 中的结论吗? 若结果不同, 你相信哪一个分析呢? 取  $\alpha = 0.05$ 。

**8.11(农业)** 一个大学的农业试验站测试了两种不同的除草剂及它们对农作物产量的影响。他们选择了 90 英亩做试验, 除草剂 1 用于 30 英亩的一随机样本, 除草剂 2 用于第二个 30 英亩的随机样本, 余下的 30 英亩作为对照。在生长期结束后, 产量(蒲式耳/英亩)如下表。

	样本均值	样本标准差	样本容量
除草剂 1	90.2	6.5	30
除草剂 2	89.3	7.8	30
对照 3	85.0	7.4	30

- 用这些数据作一个单向分类方差分析, 以检验是否在平均产量上有差异, 取  $\alpha = 0.05$ 。
- 求平均产量  $\mu_i$  的置信度为 95% 的置信区间。
- 哪些平均产量看起来是不同的?

**8.12(园艺)** 一个大学的果类农作物系的研究人员比较了用于草莓保鲜的四种不同的防腐剂。研究人员准备了一定量的用于保鲜的草莓,并随机地分成相同的四组。在每一组里,用适当的防腐剂对草莓进行处理,并装在 8 个塑料袋中在  $0^{\circ}\text{C}$  下冷冻起来。把组 I 中的袋子作为对照组,而组 II、III、IV 的袋子分别用三种新生产的防腐剂。32 个草莓袋准备好后,在  $0^{\circ}\text{C}$  下储藏 6 个月。到时间后,去掉每个袋子的包装,按变色程度从 1 到 10 打分(低分意味着变色度小),等级如下:

组 I	10	8	7.5	8	9.5	9	7.5	7
组 II	6	7.5	8	7	6.5	6	5	5.5
组 III	3	5.5	4	4.5	3	3.5	4	4.5
组 IV	2	1	2.5	3	4	3.5	2	2

- 使用下列残差图以及方差齐性的检验,检查是否满足对数据进行方差分析的条件。
- 取  $\alpha = 0.05$ , 检验平均等级是否有差异。
- 对每一组,求置信度为 95% 的平均等级的置信区间。
- 用给出的计算机输出结果验证你的结论。

练习 8.12 中的方差分析

Analysis of Variance for Ratings

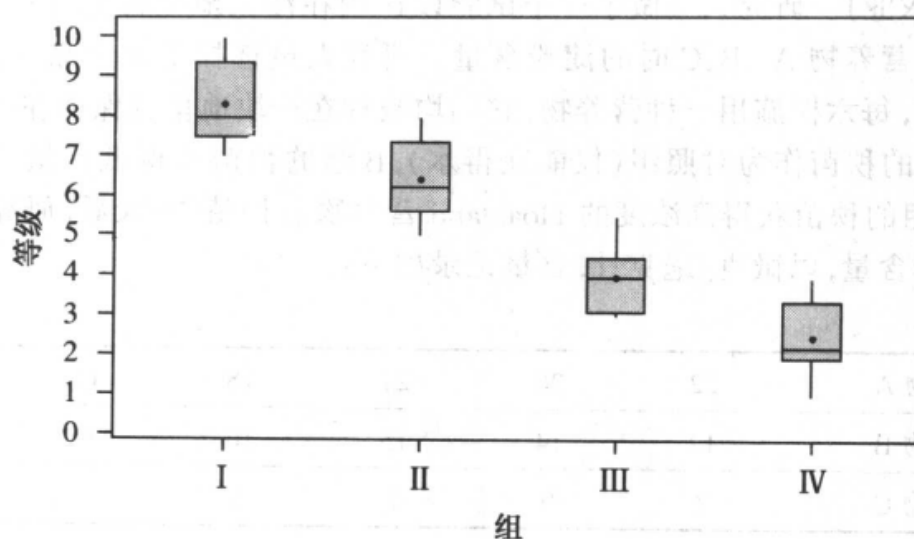
Source	DF	SS	MS	F	P
Group	3	159.187	53.062	55.67	0.000
Error	28	26.687	0.953		
Total	31	185.875			

Individual 95% CIs for Mean

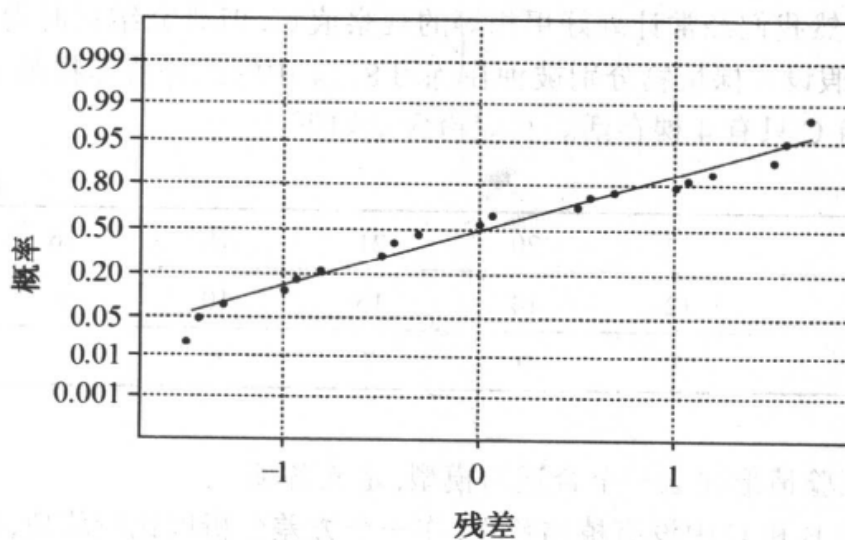
Based on Pooled StDev

Group	N	Mean	StDev	
I	8	8.3125	1.0670	(---*---)
II	8	6.4375	1.0155	(---*---)
III	8	4.0000	0.8452	(---*---)
IV	8	2.5000	0.9636	(---*---)

Pooled StDev = 0.9763      2.0      4.0      6.0      8.0



练习 8.12 中按组的等级盒形图



Average:0

Anderson-Darling Normality Test

StDev:0.927840

A-Squared:0.503

N:32

P-Value:0.191

练习 8.12 中残差的正态概率图

**8.13** 同练习 8.12, 许多情形下, 给出的是一个等级而不是实际的观察值, 此时推荐使用 Kruskal-Wallis 检验。

- 用 Kruskal-Wallis 检验确定四个组的等级分布是否有变化。
- 用 Kruskal-Wallis 检验得到的结论与练习 8.12 用方差分析得到的结论一致吗?

**8.14(农业)** 研究人员做了一个试验以比较在沙土地生长的蕃茄类植物,在补充不同的营养物 A、B、C 时的淀粉含量。研究人员选择了某个品种的 18 棵秧苗用于研究,每六棵施用一种营养物,它们均被种在一块地里且保持在一个可控环境内。A 组的秧苗作为对照组(仅能获得水),B 组的秧苗获得低含量的 Hoagland 营养物,C 组的秧苗获得高浓度的 Hoagland 营养物。种植 25 天后,研究人员测试了茎的淀粉含量,以微克/毫克计,含量记录如下:

营养物 A	22	20	21	18	16	14
营养物 B	12	14	15	10	9	6
营养物 C	7	9	7	6	5	3

- 作方差分析,检验是否对三个营养物组淀粉含量有差异( $\alpha = 0.05$ )。
- 得出你的结论。

**8.15** 虽然我们经常计划好用相等的观察值数,但研究结束时得到的观察值数可能不等。假设 6 棵植物分别被种在练习 8.14 中的三种营养物组中,但组 B 仅有 5 棵存活,组 C 只有 4 棵存活。其淀粉含量如下:

营养物 A	22	20	21	18	16	14
营养物 B	12	14	15	10	9	
营养物 C	7	9	7	6		

- 对此试验情形建立一个合适的模型,定义各项。
- 假若组 B 和 C 中没有植物死亡,作一个方差分析以比较均值,取  $\alpha = 0.05$ 。

**8.16(人力资源)** 工资的争论以及最后的解决的全部痛苦经历,使得职员和雇主之间留下了怨恨。为了评价最近雇员对薪水和福利计划的反应,人事部在公司的三个部门(生产、市场、研发)中的每一个部门随机地取 15 个人的一个样本。人事部的工作人员请抽取到的每一个人真实回答一系列问题。有一些雇员拒绝回答,因而实际样本容量不相等,具体数据如下:

	生产	市场	研发
样本容量	12	14	11
样本均值	25.2	32.6	28.1
样本方差	3.6	4.8	5.3

a. 建立这种试验情形的模型。

b. 用上述得分响应值的概括统计量, 比较三个部门的均值(分数越高雇员的满意度越高), 取  $\alpha = 0.01$ 。

**8.17(农业)** 对四个不同种类(A、B、C、D)的谷物, 研究人员记录下了作物的产量(蒲式耳/每小块土地)。在可控制的温室实验里, 研究人员随机地把每类谷物放在可用作研究的 32 块土地中的 8 块, 产量列表如下:

A	2.5	3.6	2.8	2.7	3.1	3.4	2.9	3.5
B	3.6	3.9	4.1	4.3	2.9	3.5	3.8	3.7
C	4.3	4.4	4.5	4.1	3.5	3.4	3.2	4.6
D	2.8	2.9	3.1	2.4	3.2	2.5	3.6	2.7

a. 写出合适的统计模型。

b. 对这些数据作一个方差分析, 并给出你的结论, 取  $\alpha = 0.05$ 。

**8.18** 同练习 8.17, 利用秩作一个 Kruskal-Wallis 方差分析, 取  $\alpha = 0.05$ 。把你的结果与练习 8.17 的结果做比较。

**8.19(商业)** 许多公司利用每月有固定租金的广域电话系统(WATS)进行长距离的通话。根据公司所在国的区域, 对一定的地理带可租用一条 WATS 线路。例如公司在 Ohio(俄亥俄)州, 这些地理带可包括下列州:

带 I: Ohio(俄亥俄州)

带 II: Indiana(印第安纳州)

Kentucky(肯塔基州)

Maryland(马里兰州)

Michigan(密歇根州)

North Carolina(北卡罗来纳州)

Pennsylvania(宾夕法尼亚州)

Tennessee(田纳西州)

Virginia(弗吉尼亚州)

West Virginia(西弗吉尼亚州)

Washington, D.C.(华盛顿特区)

带 III: 32 个东部和中西部州, 加上华盛顿特区

为了监督 WATS 线路的使用, 一个公司在某个月里从下列地区中的每一个选择了一个 12 次通话的随机样本, 记录下每一次的通话时间(分钟)(带 III 不包括带 II 的州及俄亥俄州)。

俄亥俄州	2	3	5	8	4	6	18	19	9	6	7	5
带 II	6	8	10	15	19	21	10	12	13	2	5	7
带 III	12	14	13	20	25	30	5	6	21	22	28	11

作一个方差分析,比较三个区域的平均通话时间,取  $\alpha = 0.05$ 。

**8.20(教育)** 大多数学龄儿童做家庭作业是晚上的常规任务。文章“Family involvement with middle-grades homework: effect of differential prompting”, *Journal of Experimental Education*, 66:31—48 讨论了父母参与孩子的家庭作业是否与提高孩子的学习水平有关。研究中包括 74 个六年级的学生及其家长。研究人员把学习能力和学习背景相似的学生安排在由同一教师所教的三个数学班中,随机指定班级到三个测试组中。

组 I 学生/家庭激励:学生被鼓励寻求家庭成员的帮助,家庭也被鼓励提供帮助给学生。

组 II 学生激励:学生被鼓励寻求家庭成员的帮助,但不特别鼓励家庭给学生帮助。

组 III 没有激励:不鼓励学生寻求家庭成员的帮助,家庭成员也不给学生以帮助。

这样,每个班被指定到三个组之一。试验后研究人员对学生进行一个测试,结果如下:

处理组	学生数	平均分
学生/家庭激励	22	68%
学生激励	22	66%
没有激励	25	67%

在这个研究中,研究人员的结论是较高程度的家庭参与和较高的学生成绩之间没有关系。

- 在这个研究中,相关的总体是什么?
- 基于上述数据,研究的结果能对什么总体具有参考意义?
- 对于三个组中的每一个,有效的样本是什么?即有多少试验单元是随机指定到三个处理组中?
- 对此研究的设计,你有什么评论?
- 在家庭参与能改进数学班学生的学习成绩的假设下,提出一个改进的设计。

**8.21(政府)** 在一个 1994 年的参议院小组委员会的听证会上,一个大烟草公司的执行官声明,尼古丁被人为地加到烟草中的指控是错误的。烟草公司的科学家称香烟中大量的尼古丁完全是由烟叶的大小确定的,较小的烟叶尼古丁含量较高。所以,香烟中所发生的尼古丁含量的变化是由于烟叶大小的变化,而不是公司在香烟中加进去的。更进一步,烟草公司认为烟叶大小的改变依赖于在生长期的气候条件,这些条件是不可控制的。为了研究是否小烟叶有较高的尼古丁含量,一

个消费者健康组织做了如下实验,控制叶子大小的主要因素是温度和在生长期植物获得的水量。实验者考虑了四种生长条件,条件 A 是平均气温和降雨量;条件 B 是比平均气温和降雨量低;条件 C 是高气温、低降雨量;条件 D 是高于正常的气温和降雨量。科学家在温室里,四种条件下分别种植了 10 株烟草,温室里的温度和降雨量是可控制的,成熟后,记录下叶子的大小及尼古丁含量,列表如下:

烟草	A 叶子大小	B 叶子大小	C 叶子大小	D 叶子大小
1	27.7619	4.2460	15.5070	33.0101
2	27.8523	14.1577	5.0473	44.9680
3	21.3495	7.0279	18.3020	34.2074
4	31.9616	7.0698	16.0436	28.9766
5	19.4623	0.8091	10.2601	42.9229
6	12.2804	13.9385	19.0571	36.6827
7	21.0508	11.0130	17.1826	32.7229
8	19.5074	10.9680	16.6510	34.5668
9	26.2808	6.9112	18.8472	28.7695
10	26.1466	9.6041	12.4234	36.6952

烟草	A 尼古丁	B 尼古丁	C 尼古丁	D 尼古丁
1	10.0655	8.5977	6.7865	9.9553
2	9.4712	8.1299	10.9249	5.8495
3	9.1246	11.3401	11.3878	10.3005
4	11.3652	9.3470	9.7022	9.7140
5	11.3976	9.3049	8.0371	10.7543
6	11.2936	10.0193	10.7187	8.0262
7	10.6805	9.5843	11.2352	13.1326
8	8.1280	6.4603	7.7079	11.8559
9	10.5066	8.2589	7.5653	11.3345
10	10.6579	5.0106	9.0922	10.4763

- 作单向方差分析,检验在四种生长条件下,叶子的平均大小是否有显著差异,取  $\alpha = 0.05$ 。
- 关于生长条件对叶子大小的影响,你能得到什么结论?
- 作单向方差分析,检验在四种生长条件下,尼古丁的平均含量是否有显著

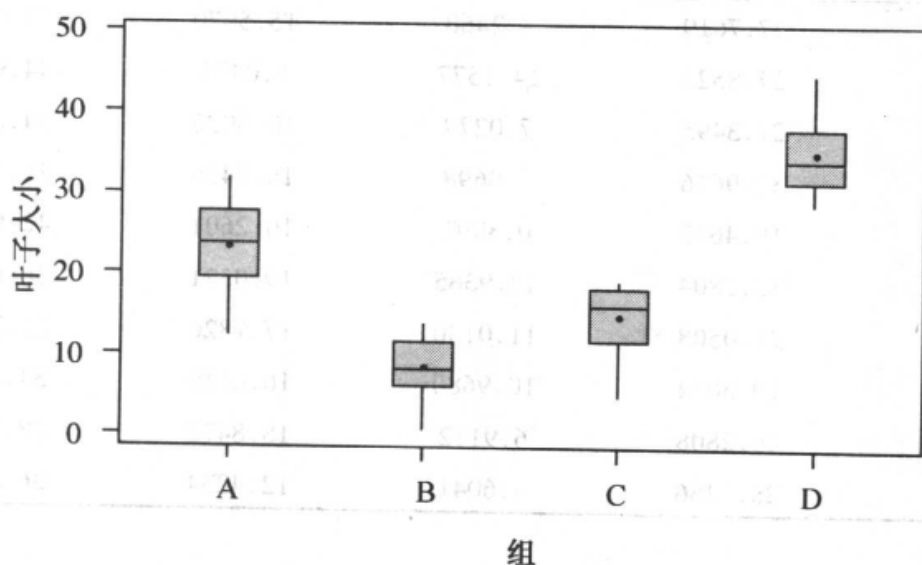


差异,取  $\alpha=0.05$ 。

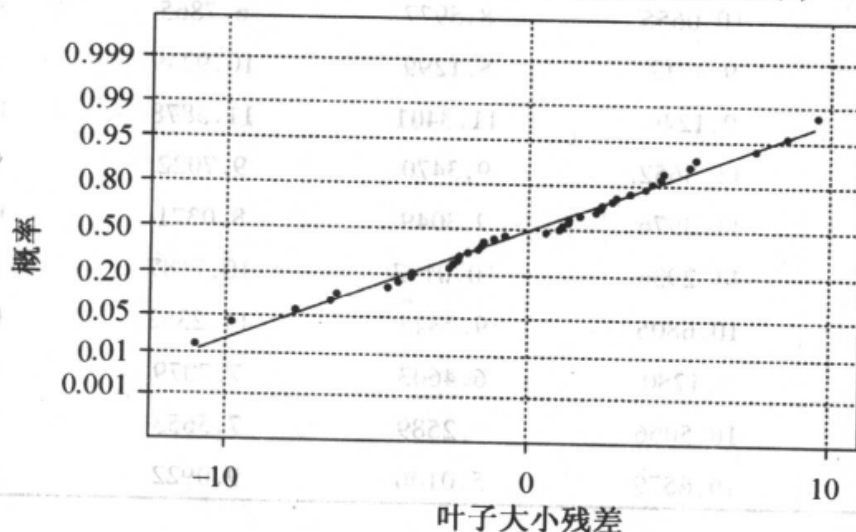
d. 关于生长条件对尼古丁含量的影响,你能得到什么结论?

e. 基于你在(b)和(d)中得到的结论,你认为烟草公司科学家的声明能被这个试验支持吗? 论证你的结论。

8.22 利用下面的图说明,练习 8.21 中尼古丁含量的数据满足方差分析的条件吗? 如果你认为条件不满足,作另一种分析并把你的结果与练习 8.21 的结果比较。



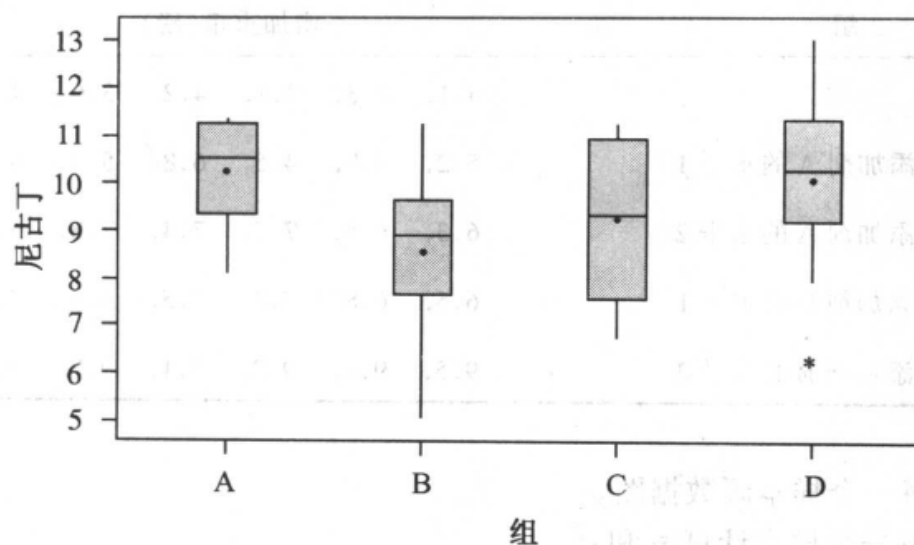
练习 8.22 中各组叶子大小的盒形图(均值用实圆点表示)



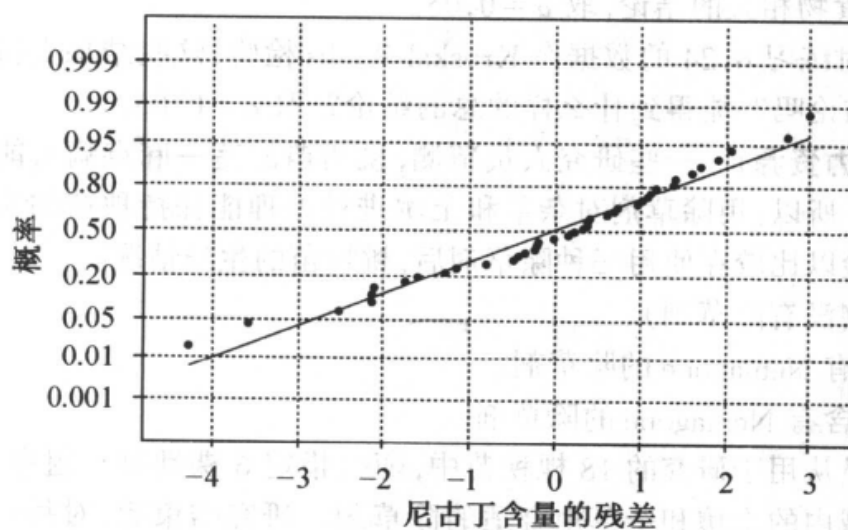
Average: 0.0000000  
StDev: 4.75535  
N: 40

Anderson-Darling Normality Test  
A-Squared: 0.205  
P-Value: 0.864

练习 8.22 中叶子大小的残差概率图



练习 8.22 中各组尼古丁含量的盒形图(均值用实圆点表示)



Average: 0.0000000

StDev: 1.62647

N: 40

Anderson-Darling Normality Test

A-Squared: 0.443

P-Value: 0.273

练习 8.22 中尼古丁含量的残差概率图

**8.23** 从你的研究领域的杂志中选一篇包含方差分析的文章, 讨论试验的设计及研究人员获得的结论类型等一些问题。

**8.24(农业)** 科学家们做了一个试验, 以检验五种不同食料对火鸡的影响, 他们为五种食料中的每一种随机指定 6 只火鸡, 喂养一段时间。

组	增加重量(磅)					
对照食料	4.1,	3.3,	3.1,	4.2,	3.6,	4.4
对照食料 + 添加剂 A 的水平 1	5.2,	4.8,	4.5,	6.8,	5.5,	6.2
对照食物 + 添加剂 A 的水平 2	6.3,	6.5,	7.2,	7.4,	7.8,	6.7
对照食物 + 添加剂 B 的水平 1	6.5,	6.8,	7.3,	7.5,	6.9,	7.0
对照食物 + 添加剂 B 的水平 2	9.5,	9.6,	9.2,	9.1,	9.8,	9.1

- 对每一个样本画数据图。
- 对每一个样本计算  $\bar{y}$  和  $s^2$ 。
- 有方差不等和非正态性的证据吗？解释之。
- 假设相对于火鸡的原重量 5 个组是可比较的，利用增加重量的数据得到与不同食物相关的结论，取  $\alpha = 0.05$ 。

**8.25** 对练习 8.24 的数据作 Kruskal-Wallis 检验，这些结果支持你从方差分析中得到的结论吗？能得到什么样的总的结论？取  $\alpha = 0.05$ 。

**8.26(人力资源)** 一些研究人员猜测，桃树苗的茎—麻点病可能与土壤中有无线虫有关。所以，用除草剂对杂草和土壤进行处理能保持秧苗的有效生长。研究人员做试验以比较在使用三种除草剂后，桃树苗的生长情况。

- 对照(没有除草剂)。
- 用含有 Nemagone 的除草剂。
- 用不含有 Nemagone 的除草剂。

研究人员从用于研究的 18 棵秧苗中，随机指定 6 棵到每一组中，对于三个组中的生长区域内的土壤和杂草用合适的除草剂。研究结束后，对每一秧苗记录它的高度(厘米)，用下列数据作方差分析以检测三组中秧苗高度的差异，取  $\alpha = 0.05$ ；给出你的结论。

除草剂 A	66	67	74	73	75	64
除草剂 B	85	84	76	82	79	86
除草剂 C	91	93	88	87	90	86

**8.27** 用练习 8.24 的数据，说明一个极端值对方差分析所得结果的影响，假定在添加剂 B 的水平 2 上，组中第 15 只火鸡的增加重量是 15.8 而不是 9.8。

- 对方差分析的假定有何影响？

b. 用 15.8 代替 9.8, 如果一个人不熟悉方差分析, 他或她能得到什么结论?

**8.28** 同练习 8.27, 如果用 15.8 代替 9.8, 对 Kruskal-Wallis 检验会发生什么影响? 有理由既作 Kruskal-Wallis 检验又作方差分析吗? 为什么?

**8.29** 为检验均值间的差异, 是否在一定情形下, Kruskal-Wallis 检验比方差分析有更高的功效? 为什么?

**8.30(工程)** 一个小公司用三种不同型号的机器生产电线的绝缘保护层, 公司想要评估由机器生产的保护层的内径。公司的一个质量工程师随机选取每个机器生产的保护层, 并且记录每个保护层的内径(毫米), 她要确定三种机器的均值和标准差之间是否有差异。

保护层	机器 A	机器 B	机器 C
1	18.1	8.7	29.7
2	2.4	56.8	18.7
3	2.7	4.4	16.5
4	7.5	8.3	63.7
5	11.0	5.8	18.9
6			107.2
7			19.7
8			93.4
9			21.6
10			17.8

a. 检验总体方差的齐性, 取  $\alpha = 0.05$ 。

b. 按照如上检验的结果, 作方差分析合适吗?

c. 如果内径的方差是不同的, 作一个能减少方差之间差异的变换, 然后进行方差分析, 确定是否平均内径有差异, 取  $\alpha = 0.05$ 。

d. 把在(c)中分析的结果与下面给出的计算机输出作比较, 哪一个原直径的方差分析?

e. 如果工程师知道机器 B 和机器 C 的方差远比机器 A 的方差大, 她会怎样设计一个不同的试验?

练习 8.30 的单向方差分析

Analysis of Variance

Source	DF	SS	MS	F	P
Factor	2	4141	2071	2.73	0.094
Error	17	12907	759		
Total	19	17048			

Individual 95 % CIs for Mean

Based on Pooled StDev

Level	N	Mean	StDev	
Machine	5	8.32	6.52	{-----*-----}
Machine	5	16.78	22.43	{-----*-----}
Machine	10	40.70	34.52	{-----*-----}

-----+-----+-----+-----

Pooled StDev = 27.55

0 25 50

**8.31** 当方差不等时, Kruskal-Wallis 检验不象方差分析那样, 受到那么大的影响。对练习 8.30 中的数据, 把 Kruskal-Wallis 检验用于原始数据及变换数据, 并且比较这两个分析的结果以证实这个结论。

## 第九章 多重比较

- 9.1 引言和案例
- 9.2 线性对照
- 9.3 控制哪个错误率
- 9.4 Fisher(费舍尔)最小显著差异法
- 9.5 Tukey 的 W 方法
- 9.6 Student-Newman-Keuls 方法
- 9.7 Dunnett 方法:处理组与对照组的比较
- 9.8 Scheffé 的 S 方法
- 9.9 小结

### 9.1 引言和案例

在第八章里,我们介绍了检验  $t$  个总体均值是否相等的方法。我们利用检验统计量  $F = \frac{s_B^2}{s_W^2}$  判断样本组间变异性相对于样本组内变异性来说是否太大。如果从样本数据计算得到的  $F$  值超过了附表 8 中的临界值,我们就拒绝零假设  $H_0: \mu_1 = \mu_2 = \cdots = \mu_t$  而接受备择假设

$H_a: t$  个总体均值中至少有一个与其余的不同

尽管拒绝零假设确实给我们提供了一些有关总体均值的信息,但我们还不知道究竟哪些值与其他值不同。例如,  $\mu_1$  与  $\mu_2$  或  $\mu_3$  相等吗?  $\mu_3$  等于  $\mu_2, \mu_4, \mu_5$  三者的均值吗? 处理均值  $\mu_1, \mu_2, \dots, \mu_t$  是否存在增长趋势? 多重比较方法及对比就是用来回答上述这些问题的。尽管存在着许多多重比较方法,但我们只集中讨论一些比较常用的方法。在研究了这些方法后,你将能应用多重比较方法评价许多出版物中的结果,或者在试验中提出合理的多重比较方法。

注意:我们只倾向于分析那些从样本数据看来有意义的比较。这种做法有时被称为数据网渔(data dredging)或数据探索(data snooping),单个比较的置信度不能反映比较的深层特征。例如,利用以前的知识我们知道利用公式

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

两个总体均值之差的区间估计的置信度是  $1 - \alpha$ 。假如我们利用方差分析检验了有关 6 个总体的假设

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$

但只是在看到最大的样本均值是  $\bar{y}_1$  及最小的样本均值是  $\bar{y}_2$  后, 我们才决定计算  $\mu_1$  与  $\mu_2$  的置信区间。这种情况下, 置信度就不是原来设想的  $1 - \alpha$  了; 这个值只适用于预先计划的比较, 即适用于未看到样本数据的情况。

在观察到样本数据后, 要进行数据探索, 一种方法是利用某个多重比较方法在一定的置信度下包含所有比较在观察到样本数据的情况下可以作出的结论, 本章中就要讨论一些这样的方法。

其他可能的选择是以数据探索性的比较为基础建立假设, 这些假设必须在接下来的试验或研究中证实。这里, 数据探索性的比较扮演着探测性的或产生零假设的角色, 并不能在数据探测的基础上得出任何结论。需要作进一步的试验证实数据探索中建立的假设。

#### 案例: 面试者的决定受不同残疾类型的影响吗?

在美国大约有 50,000,000 人报告说有身体残疾。而且, 据估计年龄在 18 到 64 岁之间的残疾人的失业率几乎是正常人失业率的两倍。因此, 身体有残疾的人在找工作时似乎比较困难。残疾人面临的问题之一可能是在受雇面谈时遭受雇主的歧视。

刊登于杂志 *Human Performance* (《人类行为学》) 1990 年第三期 157~171 页的论文 “Interviewers’ decisions related to applicant handicap type and rater empathy” (“面试官的决定与找工作的残疾人类型及其移情作用有关”) 就考察了这样的问题。研究的目的是调查是否不同类型的残疾人令评价者产生不同的移情作用及考察面试官的评价是否受到被面试者的残疾类型的影响。

**数据收集方案的设计** 研究者对五次模拟的雇佣面试作了录像。为减少不同录像带之间的偏差, 五盘录像带中使用相同的男演员(申请工作者及面试官)及包含相同九个问题的问卷。问卷直指申请者的一般资格, 因为这类申请者最易受到面试歧视的影响。这些录像带因申请者残疾的不同而不同, 但所有的申请者均被表演为永久残疾。五个条件被标记为轮椅、加拿大拐杖、听觉障碍、腿部截肢及正常人(对照)。

这项研究中的每个参与者应根据面试的录像中应试者为申请计算机销售职位而回答问题的情况对其资格进行评定。在观看录像前, 每个参与者均填写霍根认同等级表(Hogan Empathy Scale)。研究者决定让每个参与者仅观看五盘录像带中的一盘。根据先前研究中大批评定人员评定结果的变异性, 研究者们决定要对五种残疾情况中的每一种情况的平均评定结果得到准确的估计, 需要为每盘录像带配备 14 名评定人员。研究中选择了 70 名在校大学生参与。对五盘中的每一盘录像带, 随机选择 14 名学生观看, 在看完每一盘录像带后, 参与者对应试者按照两个方面进行打分: 第一个方面包含有 11 项, 用来评价评定人员对应试者的喜欢程度, 另一个方面包含有 10 项, 用来评价评定人员对应试者工作资格的评价。对于每一

个方面,各个项目的平均值就是对应试者的整体评价。研究者利用这两个变量来判断评定人员对不同的残疾类型是否有不同的反应,并判断评定人员对残疾应试者的移情效果。

下面是研究者感兴趣的一些问题:

1. 70 名评定人员移情分数的平均值是否存在差异?
2. 定级者的平均资格分数是否随着五种残疾类型的不同而不同?
3. 哪几组残疾情况导致不同的平均资格分数?
4. 对照组(正常人)的平均定级是否大于所有类型残疾者的平均定级?
5. 听觉障碍应试者的平均资格定级是否与运动障碍应试者的平均定级不同?
6. 拄拐杖应试者的平均资格定级是否与那些被截肢的或坐轮椅的应试者的平均定级不同?
7. 被截肢的应试者的平均定级是否与坐轮椅的应试者的平均定级不同?

研究者们做完试验并且从 70 名对应试者定级的参加人员中获得了如下所示的数据。移情值经概括后形成表 9.1 中的数据。

表 9.1 五种残疾条件的移情值

条件	对照(无)	听觉障碍	加拿大拐杖	截去一条腿	轮椅
平均数	21.43	22.71	20.43	20.86	19.86
标准差	3.032	3.268	3.589	3.035	3.348

表 9.2 中的数据是 70 名评定人员对具有五种残疾情况的应试者资格的打分及其描述统计量(这些数据是用上述论文中的概括统计量模拟产生的。)

表 9.2 五种残疾情况的应试者资格打分

对 照	听觉障碍	截 肢	拐 杖	轮 椅
6.1	2.1	4.1	6.7	3.0
4.6	4.8	6.1	6.7	3.9
7.7	3.7	5.9	6.5	7.9
4.2	3.5	5.0	4.6	3.0
6.1	2.2	6.1	7.6	3.5
2.9	3.4	5.7	2.9	8.1



续表				
对 照	听觉障碍	截 肢	拐 杖	轮 椅
4.6	5.5	1.1	5.2	6.4
5.4	5.2	4.0	3.5	6.4
4.1	6.8	4.7	5.2	5.8
6.4	0.4	3.0	6.6	4.6
4.0	5.8	6.6	6.9	5.8
7.2	4.5	3.2	6.1	5.5
2.4	7.0	4.5	5.9	5.0
2.9	1.8	2.1	8.8	6.2

案例中的描述统计量

riable	N	Mean	Median	TrMean	StDev	SE Mean
ntrol	14	4.900	4.600	4.875	1.638	0.438
rd of Hearing	14	4.050	4.100	4.108	1.961	0.524
putee	14	4.436	4.600	4.533	1.637	0.437
utches	14	5.914	6.300	5.925	1.537	0.411
eelchair	14	5.364	5.650	5.333	1.633	0.436

riable	Minimum	Maximum	Q1	Q3
ntrol	2.400	7.700	3.725	6.175
rd of Hearing	0.400	7.000	2.175	5.575
putee	1.100	6.600	3.150	5.950
utches	2.900	8.800	5.050	6.750
eelchair	3.000	8.100	3.800	6.400

图 9.1 中标出了资格分数。盒形图在某种程度上显示评定人员对拄拐杖的应试者给出了较高的资格分数。听力障碍及截肢者的平均资格分数比那些对照组的人及坐轮椅的情况有些低。所有五种情况中资格分数的变异性几乎是一样的。另外,这些情况表明分布中不存在极度偏态或离群值;因此,正态分布的假设是合理的。

**管理数据** 研究者们下一步将按照 2.5 节中描述的步骤为进行统计分析准备

数据。

在第九章接下来的各节中,我们将建立不同的方法以回答这儿提出的问题。

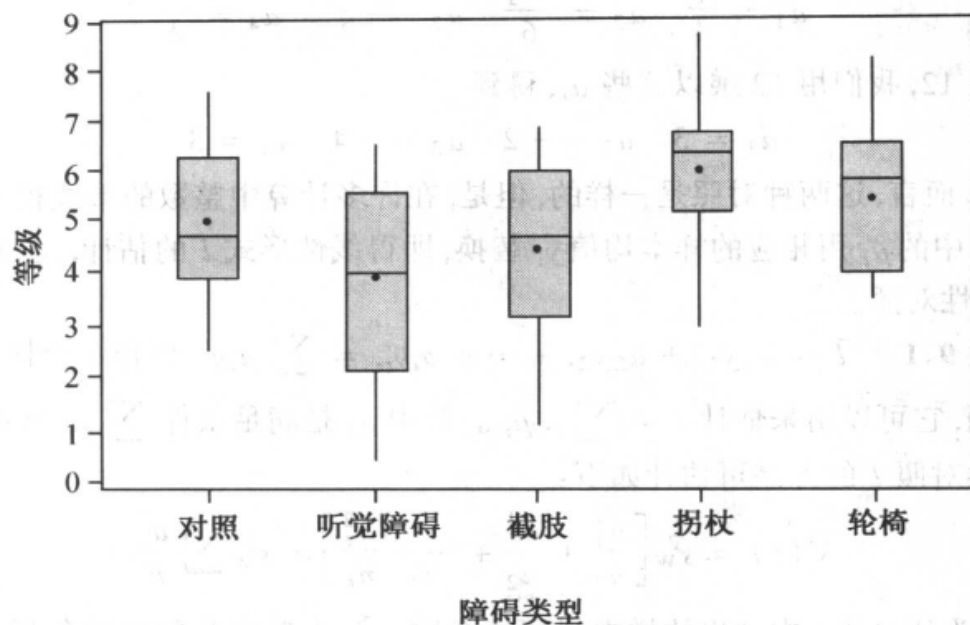


图 9.1 对不同障碍等级的盒形图(实心圆点代表均值)

## 9.2 线性对照

在建立几个不同的多重比较方法之前,我们需要以下记号及定义。考虑单向分类问题,即我们想对  $t$  个总体均值  $\mu_1, \mu_2, \dots, \mu_t$  作出比较。这  $t$  个总体均值的比较问题可以写成如下形式:

$$l = a_1\mu_1 + a_2\mu_2 + \dots + a_t\mu_t = \sum_{i=1}^t a_i\mu_i$$

这儿  $a_i$  是满足  $\sum_{i=1}^t a_i = 0$  的常数。比如,我们想比较  $\mu_1$  与  $\mu_2$ ,就写成如下线性形式

$$l = \mu_1 - \mu_2$$

注意这儿  $a_1 = 1, a_2 = -1, a_3 = a_4 = \dots = a_t = 0$ , 且  $\sum_{i=1}^t a_i = 0$ 。同理,我们要比较第一个总体均值与第二、第三个总体均值的平均数,那么  $l$  的形式如下

$$l = \mu_1 - \frac{(\mu_2 + \mu_3)}{2}$$

这里  $a_1 = 1, a_2 = a_3 = -\frac{1}{2}, a_4 = a_5 = \dots = a_t = 0$ , 且  $\sum_{i=1}^t a_i = 0$ 。

在这些对照中我们经常把所有的  $a_i$  写成整数。要这么做, 我们用一个公分母乘以这些  $a_i$ 。假如我们有四个处理均值的如下对照:

$$a_1 = \frac{1}{4} \quad a_2 = \frac{-1}{6} \quad a_3 = \frac{-1}{3} \quad a_4 = \frac{1}{4}$$

公分母是 12, 我们用 12 乘以这些  $a_i$ , 得到

$$a_1 = 3 \quad a_2 = -2 \quad a_3 = -4 \quad a_4 = 3$$

就这些  $\mu$  而言, 这两种对照是一样的, 但是, 在许多计算中整数的形式较易操作。

将  $l$  中的  $\mu_i$  用相应的样本均值  $\bar{y}_i$  替换, 即得线性形式  $l$  的估计, 记为  $\hat{l}$ 。估计  $\hat{l}$  叫做线性对照。

**定义 9.1**  $\hat{l} = a_1\bar{y}_1 + a_2\bar{y}_2 + \cdots + a_t\bar{y}_t = \sum_i a_i\bar{y}_i$  叫作  $t$  个样本均值的线性对照, 它可以用来估计  $l = \sum_i a_i\mu_i$ 。其中  $a_i$  是满足条件  $\sum_i a_i = 0$  的常数。

线性对照  $\hat{l}$  的方差可估计如下:

$$\hat{V}(\hat{l}) = s_w^2 \left[ \frac{a_1^2}{n_1} + \frac{a_2^2}{n_2} + \cdots + \frac{a_t^2}{n_t} \right] = s_w^2 \sum_i \frac{a_i^2}{n_i}$$

这里,  $n_i$  为从总体  $i$  中抽出的样本观察值个数,  $s_w^2$  为单向分类方差分析表中计算的样本组间均方。如果各样本容量一样 (也就是说, 所有的  $n_i = n$ ), 那么

$$\hat{V}(\hat{l}) = \frac{s_w^2}{n} \sum_i a_i^2$$

由  $t$  个样本均值可以组成许多不同的对照。下面定义一组特殊的对照:

**定义 9.2** 两个对照  $l_1$  及  $l_2$ , 这里

$$\hat{l}_1 = \sum_i a_i\bar{y}_i \quad \text{及} \quad \hat{l}_2 = \sum_i b_i\bar{y}_i$$

如果满足条件

$$\frac{a_1b_1}{n_1} + \frac{a_2b_2}{n_2} + \cdots + \frac{a_tb_t}{n_t} = \sum_{i=1}^t \frac{a_ib_i}{n_i} = 0$$

则称两者正交。

注意: 如果样本容量一致, 那么条件就变成

$$a_1b_1 + a_2b_2 + \cdots + a_tb_t = \sum_{i=1}^t a_ib_i = 0$$

如果一组对照中每两个都正交, 则称这组对照相互正交。

### 例 9.1

考虑比较  $t=4$  总体均值的单向分类。下列对照正交吗?

$$\hat{l}_1 = \bar{y}_1 - \bar{y}_2 \quad \hat{l}_2 = \bar{y}_3 - \bar{y}_4$$

**解答** 我们可以用如下形式改写这两个对照:

$$\hat{l}_1 = \bar{y}_{1.} - \bar{y}_{2.} + 0(\bar{y}_{3.}) + 0(\bar{y}_{4.})$$

$$\hat{l}_2 = 0(\bar{y}_{1.}) - 0(\bar{y}_{2.}) + \bar{y}_{3.} + \bar{y}_{4.}$$

这里我们看到  $a_1=1, a_2=-1, a_3=0, a_4=0$  及  $b_1=0, b_2=0, b_3=1, b_4=-1$ 。很明显

$$\sum_i a_i b_i = a_1 b_1 + a_2 b_2 + a_3 b_3 + a_4 b_4 = 0$$

因此,这两个对照正交。

### 例 9.2

参考例 9.1, 下面给出的对照正交吗?

$$\hat{l}_1 = \bar{y}_{1.} - \bar{y}_{2.} \text{ 及 } \hat{l}_2 = \bar{y}_{1.} - \bar{y}_{3.}$$

**解答** 把这两个对照改写为

$$\hat{l}_1 = \bar{y}_{1.} - \bar{y}_{2.} + 0(\bar{y}_{3.}) + 0(\bar{y}_{4.})$$

$$\hat{l}_2 = \bar{y}_{1.} + 0(\bar{y}_{2.}) - \bar{y}_{3.} + 0(\bar{y}_{4.})$$

我们看到

$$\sum_i a_i b_i = (1)(1) + (-1)0 + 0(-1) + (0)(0) = 1$$

这说明这两个对照不是正交的。

线性对照中的正交性这个概念很重要, 因为两个对照正交, 那么一个对照不包含另一个对照的任何信息。我们将要证明, 利用  $t$  个样本均值  $\bar{y}_i$ , 可以形成  $t-1$  个正交对照, 这  $t-1$  个对照相互正交。(记住  $t-1$  的一个简单办法是参考单向分类方差分析表中源于处理(样本间)的变异性的自由度的值。)另外, 很明显, 这  $t-1$  个对照的平方和加起来等于处理(样本间)平方和。相互正交将直接导致与这  $t-1$  个对照相联系的  $t-1$  个平方和独立。这样, 我们就可以把  $t-1$  个自由度与描述任何处理均值差异的处理平方和联系起来, 并将其分成  $t-1$  个相互独立的用来表明处理均值差异之所在的部分。我们将举一个例子进一步拓展这些思想并阐述这些概念。

为表明可以用某个对照来解释的各处理均值的变异程度, 我们需要计算与该处理对照相应的平方和。对每一个对照  $\hat{l} = \sum_{i=1}^t a_i \bar{y}_i$ , 我们可以如下计算这个对照的平方和(SSC):

$$\text{SSC} = \frac{\left(\sum_{i=1}^t a_i \bar{y}_i\right)^2}{\sum_{i=1}^t (a_i^2/n_i)} = \frac{(\hat{l})^2}{\sum_{i=1}^t (a_i^2/n_i)}$$

当样本容量相同时,此公式简化为

$$SSC = \frac{n(\bar{l})^2}{\sum_{i=1}^t a_i^2}$$

每个这样的平方和的自由度为 1。这样,我们就可得到  $t-1$  个正交对照,使得具有  $t-1$  个自由度的处理间平方和,等于这  $t-1$  个对照的平方和之和。下面通过例子来说明。

### 例 9.3

不同的药剂用来控制庄稼地里的杂草。我们所担心的是化学制剂的过量使用。这些药剂虽然能有效地控制杂草,但同时也流入地下水网造成健康问题。因此,为了消除化学制剂带来的污染问题,有人发明了一些新的生物除草剂。研究者们作了一项有关生物制剂的研究,借此评价与化学制剂相比其是否有效。研究包括一对照组(无任何制剂),两种生物药剂(药剂 1 与药剂 2),及两种化学制剂(化学 1 与化学 2)。把红花草种植在 30 个 1 英亩的地块上。随意指定 6 个地块接受 5 中处理中的一种。收获红花草并记录每一英亩的总产量(单位:吨)。数据如下。

制剂	1	2	3	4	5
类型	无	生物 1	生物 2	化学 1	化学 2
$\bar{y}_{i\cdot}$	1.175	1.293	1.328	1.415	1.500
$s_i$	0.1204	0.1269	0.1196	0.1249	0.1265
$n_i$	6	6	6	6	6

寻找四个正交对照并证明四个对照的各自平方和之和等于处理(样本间)的平方和。

**解答** 对这些数据进行方差分析,产生的结果用如下方差分析表表示出来。

来源	df	SS	MS	F	p 值
处理	4	0.3648	0.0912	5.96	0.0016
误差	25	0.3825	0.0153		
总和	29	0.7472			

方差分析表中,  $SS_{\text{Trt}} = 0.3648$ 。我们将在 5 个处理均值中建立 4 个正交组合

并证明  $SS_{Tr}$  可分为 4 个部分, 每一部分代表某个特定对照的 1 个自由度的平方和。表 9.3 包含了四个组合中每一个的平方和及系数。

表 9.3 杂草控制试验平方和的计算

Contrast	处理					$\sum_{i=1}^5 a_i^2$	$\bar{l}$	$SSC_i$
	1(对照组)	2(生物 1)	3(生物 2)	4(化学 1)	5(化学 2)			
对照组与药剂	4	-1	-1	-1	-1	20	-0.836	0.2097
生物与化学	0	1	1	-1	-1	4	-0.294	0.1297
生物 1 与生物 2	0	1	-1	0	0	2	-0.035	0.0037
化学 1 与化学 2	0	0	0	1	-1	2	-0.085	0.0217
$\bar{y}_i$	1.175	1.293	1.328	1.415	1.500			0.3648

要阐述表 9.3 中涉及的计算方法, 我们计算一下第一个对照的平方和, 即无药剂对有药剂。首先, 注意这个对照代表无药剂处理的产量与四种施药剂的平均产量之间的比较。把这个对照先写为

$$l = \mu_1 - \frac{(\mu_2 + \mu_3 + \mu_4 + \mu_5)}{4}$$

$$= (1)\mu_1 + \left(\frac{-1}{4}\right)\mu_2 + \left(\frac{-1}{4}\right)\mu_3 + \left(\frac{-1}{4}\right)\mu_4 + \left(\frac{-1}{4}\right)\mu_5$$

然而, 我们将每个系数乘以 4, 将系数从

$$a_1 = 1 \quad a_2 = \frac{-1}{4} \quad a_3 = \frac{-1}{4} \quad a_4 = \frac{-1}{4} \quad a_5 = \frac{-1}{4}$$

变为

$$a_1 = 4 \quad a_2 = -1 \quad a_3 = -1 \quad a_4 = -1 \quad a_5 = -1$$

下一步, 我们计算

$$\sum_{i=1}^5 a_i^2 = (4)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (-1)^2 = 20$$

及

$$\bar{l} = (4)(1.175) + (-1)(1.293) + (-1)(1.328) + (-1)(1.415) + (-1)(1.500)$$

$$= -0.836$$

最后, 我们从下式可以得到此组合的平方和

$$SSC_1 = \frac{(\bar{l})^2}{\sum_{i=1}^5 (a_i^2/n_i)} = \frac{n(\bar{l})^2}{\sum_{i=1}^5 a_i^2} = \frac{6(-0.836)^2}{20} = 0.2097$$

其余三个平方和可以类似方法得到。从表 9.3 我们得到

$$\begin{aligned} \text{SSC}_1 + \text{SSC}_2 + \text{SSC}_3 + \text{SSC}_4 &= 0.2097 + 0.1297 + 0.0037 + 0.0217 \\ &= 0.3648 = \text{SS}_{\text{Treat}} \end{aligned}$$

您将在一个练习中证明这四个对照确实相互正交。

表 9.3 阐述了我们怎么样把处理中的差异分解为不同的代表各处理均值比较的组合。在定义了对照, 获得了其估计及平方和之后, 我们需要判断哪一个对照与零有显著差异。一个对照若取零值, 则表明这个对照代表的均值间的差异不存在。例如, 如果我们的对照  $l_1$  (无药剂对药剂) 经计算得零, 那么, 我们就得出结论: 没有使用任何药剂的地块的平均产量等于所有使用四种药剂的地块的平均产量。现在我们提出假设: 某个组合  $l = \sum_{i=1}^t a_i \mu_i$  不等于零。从方差分析表的  $F$  检验变形后即可得到我们的检验。因为每个对照的平方和具有一个自由度, 其均方与其平方和一样。于是检验统计量为

$$F = \frac{\text{SSC}}{\text{MS}_{\text{Error}}} = \frac{\text{SSC}}{s_w^2}$$

检验方法总结如下。

#### 对照的 $F$ 检验

$$H_0: l = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_t \mu_t = 0$$

$$H_a: l = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_t \mu_t \neq 0$$

$$\text{T. S. : } F = \frac{\text{SSC}}{\text{MS}_{\text{Error}}}$$

R. R. : 对指定的值  $\alpha$ , 如果  $F$  大于对  $\alpha = \alpha$ ,  $df_1 = 1$ ,  $df_2 = n_T - t$  查表 8 得到的  $F$  值就拒绝  $H_0$ 。

检验假定并得到结论。

#### 例 9.4

参见例 9.3。研究者们特别关心的是生物制剂与化学制剂的效果是否一致。杂草控制中使用的四种药剂与无药处理这两者对红花草平均产量的影响是否存在显著差异? 检验四种组合中每一种的显著性。

**解答** 从例 9.3 的概括统计量表中可以看出, 样本标准差几乎相同。事实上,

$$F_{\max} = (0.1269)^2 / (0.1196)^2 = 1.13$$

从表 12 中可以得到,  $F_{\max, 0.05, 5, 5} = 16.3$ 。因此, 我们没有理由怀疑这五个总体的方差不相等。例 9.3 的方差分析表中  $p$  值为 0.0016。因此, 我们有充足的理由拒绝

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

我们由此得出结论:五个总体均值存在显著差异( $p=0.0016$ )。利用例 9.3 中我们建立的四个组合可以考察这些均值间差异的类型。由  $F_i = \text{SSC}_i / \text{MS}_{\text{Error}}$  可计算出如下四个检验统计量。

$$F_1 = \frac{0.2097}{0.0153} = 13.71 \quad F_2 = \frac{0.1297}{0.0153} = 8.48$$

$$F_3 = \frac{0.0037}{0.0153} = 0.24 \quad F_4 = \frac{0.0217}{0.0153} = 1.42$$

对  $\alpha=0.05$ ,  $\text{df}_1=1$ ,  $\text{df}_2=30-5$  查表 8, 得到  $F_{0.05,1,25}=4.24$ 。因此, 我们得到结论: 组合  $l_1$  与  $l_2$  与零有显著差异, 但是  $l_3$  与  $l_4$  与零无显著差异。利用组合  $l_1$ , 我们得出结论: 使用除草控制药剂的地块的平均亩产比没有使用任何药剂的地块的平均亩产高得多。利用组合  $l_2$ , 我们推断得到使用生物药剂控制杂草的地块的平均亩产比使用化学药剂的地块的平均亩产低一些。然而, 我们需要考察平均亩产差异的大小来决定这个差异仅仅是统计意义上的还是有经济意义的。如果这个差异在经济上是显著的, 使用生物制剂得到的生态收益将为生物制剂代替化学制剂提供正当的理由。

当我们选择对照来研究时, 目标并不是得到一组正交对照将处理的平方和分解成  $t-1$  个部分, 而是寻找的处理均值的对照, 以能够清楚地解释对研究者最有用的处理均值差异类型。对照之间的相互正交性在选择过程中有点儿象额外福利。例如, 在除草剂的分析中, 我们也有兴趣比较一下无药剂组与两种生物药剂组的平均值。这个对照与我们先前设计的许多对照相比不是正交的。我们仍旧使用这个对照并应用试验数据检验其显著性。选择哪一个对照来估计应该由试验者的整体目标来决定, 而不是根据其正交性。

我们在检验许多对照时遇到的问题称之为多重比较。当我们检验了许多对照时, 每一个对照犯第一类错误的概率为  $\alpha$ , 在这些检验中, 至少有一个第一类错误发生的概率要比  $\alpha$  大一些。在下一部分, 我们将阐述这个难点。

## 练习

### 基本技能

#### 9.1 考虑下列表达式

$$\hat{l}_1 = \bar{y}_{1\cdot} + \bar{y}_{2\cdot} - 2\bar{y}_{3\cdot}$$

$$\hat{l}_2 = \bar{y}_{1\cdot} + \bar{y}_{2\cdot} - 2\bar{y}_{4\cdot}$$

a.  $\hat{l}_1$  与  $\hat{l}_2$  是线性对照吗?

b.  $\hat{l}_1$  与  $\hat{l}_2$  正交吗?



**9.2** 在例 9.3 中,在 5 个处理均值中我们建立了四个对照。证明这四个对照相互正交。

**9.3** 在本章前面部分描述的案例中,研究者们感兴趣于回答这些问题:评定人员对不同残疾类型的反映差异有多大。针对下列每一个问题,写出五种情况平均定级的一个线性对照,借此回答研究者们提出的问题。

- a. 问题 1:对照组(无残疾)的平均定级比所有残疾类型的应试者的平均定级高吗?
- b. 问题 2:听力障碍的应试者的平均资格定级与那些移动障碍应试者的平均资格定级相同吗?
- c. 问题 3:拄拐杖应试者的平均资格定级与截肢者或坐轮椅的应试者的平均资格定级相同吗?
- d. 问题 4:截肢应试者的平均资格定级与坐轮椅的应试者的平均资格定级相同吗?

**9.4** 参考练习 9.3。针对下列每一对对照,判断它们是否正交。

- a. 问题 1 与问题 2。
- b. 问题 1 与问题 3。
- c. 问题 1 与问题 4。
- d. 问题 2 与问题 3。
- e. 问题 2 与问题 4。
- f. 问题 3 与问题 4。
- g. 这四个对照相互正交吗?

**9.5** 参考例 8.4。研究者们感兴趣于判断从密西西比河口附近抽取的样本的平均氧含量与从离河口较远的地方抽取的样本的平均氧含量相比偏低与否。写出对照回答下列每一个问题并在  $\alpha = 0.05$  下检验此对照是否等于零。清晰概括你的结果。

- a. 20 公里处的平均氧含量是否等于 1, 5, 10 公里处平均氧含量的平均数?
- b. 10 公里处的平均氧含量是否等于 1, 5 公里处平均氧含量的平均数?
- c. 5 公里处的平均氧含量是否等于 1 公里处的平均氧含量?
- d. (a), (b) 及 (c) 中定义的三个对照相互正交吗?
- e. 这三个对照的平方和等于  $SS_{Tr}$  吗?

### 9.3 控制哪个错误率

试验者希望通过  $m$  个对照来比较  $t$  个总体(处理)的均值。 $m$  个对照中的每一个都可以通过以前章节介绍过的  $F$  检验统计量来检验。假如每一个对照都用

相同的  $\alpha$  值检验, 此  $\alpha$  值表示为  $\alpha_I$ , 叫作**单个比较的第 I 类错误率**。因此,  $m$  个检验中的每一个检验犯第 I 类错误的概率是  $\alpha_I$ 。我们也需要考虑错误地拒绝  $m$  个零假设中至少一个假设的概率, 叫作**试验的第 I 类错误率**, 记为  $\alpha_E$ 。值  $\alpha_E$  考虑了我们在进行  $m$  个检验时, 每一个检验犯第 I 类错误的概率是  $\alpha_I$ 。现在, 如果  $MS_{Error}$  的自由度无限大(因此各个检验相互独立), 那么当所有的  $m$  个零假设都成立时, 错误地拒绝  $m$  个零假设中至少一个的概率可以用  $\alpha_E = 1 - (1 - \alpha_I)^m$  表示。表 9.4 包含了与不同  $m$  及  $\alpha_I$  相对应的  $\alpha_E$  值。从表 9.4 中我们观察到对给定值  $\alpha_I$ , 当检验的数量  $m$  增加时, 在  $m$  个检验中至少有一个检验拒绝  $H_0$  的概率  $\alpha_E$  变得越来越大。例如, 如果试验者想通过  $m = 10$  的正交对照来比较  $t = 20$  个总体均值, 当每一个单个检验的  $\alpha_I = 0.05$  时, 在所有的  $t$  个检验中至少有一个检验错误地拒绝  $H_0$  的概率达到 0.401。

表 9.4  $t$  个样本均值的  $m$  个独立对照时试验的第 I 类错误率的比较

$m$ , 对照个数	$\alpha_I$ , 每个检验犯第 I 类错误的概率		
	0.10	0.05	0.01
1	0.100	0.050	0.010
2	0.190	0.097	0.020
3	0.271	0.143	0.030
4	0.344	0.185	0.039
5	0.410	0.226	0.049
$\vdots$	$\vdots$	$\vdots$	$\vdots$
10	0.651	0.401	0.096

在任何实际问题中,  $MS_{Error}$  的自由度总是有限的, 因此各检验不是独立的。这样  $\alpha_E$  与  $\alpha_I$  的关系通常并不像表 9.4 描述的那样。当各检验不独立时, 很难找到相当于  $\alpha_E = 1 - (1 - \alpha_I)^m$  的表达式。但是, 可以证明, 对我们要做的总体均值间的大多数类型的比较, 试验的错误率存在如下上限:

$$\alpha_E \leq 1 - (1 - \alpha_I)^m$$

这样, 对于单个检验中的每一个, 当给定值  $\alpha_I$  时, 我们就知道了  $\alpha_E$  的最大可能值。假如, 对  $t = 20$  的总体均值的  $m = 8$  个对照, 我们希望单个检验的错误率最大为 0.05。要达到整个错误率  $\alpha_E = 0.05$ , 在这  $m$  个检验中我们应该选择怎样的  $\alpha_I$  呢? 利用前边的上限, 如果我们选择

$$\alpha_I = 1 - (1 - \alpha_E)^{\frac{1}{m}} = 1 - (1 - 0.05)^{\frac{1}{8}} = 0.0064$$

那么就可得到  $\alpha_E \leq 0.05$ 。惟一的问题是与单个检验错误率相比, 这个方法十分保

守,因此,将会增大犯第Ⅱ类错误的概率。

现在考虑一种方法,这种方法适用于任何  $m$  个检验而且较容易获得  $\alpha_E$  的上限。当我们进行许多个检验时,表 9.4 中的结果会令人烦恼。至少犯一次第一类错误的概率将会比选定的单个检验的错误率高很多。当显著的结果出现在试验结果的分析中时,我们将对此提出怀疑。如果我们控制的是试验的错误率  $\alpha_E$  而不是单个检验的错误率  $\alpha_I$  时,这个问题可稍微得到缓解。我们需要选择一个  $\alpha_I$  值使得  $\alpha_E$  可被人们接受。Bonferroni (帮佛洛尼) 不等式给我们提供了一个在  $\alpha_E$  不超过某个给定值时选择  $\alpha_I$  的方法。这个不等式阐明整个的第一类错误率  $\alpha_E$  小于或等于  $m$  个检验的单个错误率之和。因此,当  $m$  个检验中的每一个检验具有相同的单个检验错误率  $\alpha_I$  时,应用 Bonferroni 不等式得到

$$\alpha_E \leq m\alpha_I$$

如果我们想保证犯第一类错误的概率最大为  $\alpha$ ,我们对  $m$  个检验中的任何一个选择

$$\alpha_I = \frac{\alpha}{m}$$

那么,

$$\alpha_E \leq m\alpha_I = m\left(\frac{\alpha}{m}\right) = \alpha$$

这样,试验的错误率就小于等于我们指定的值。正如我们先前提到的那样,由这个方法决定的试验的错误率可能十分保守,因此将会增大犯第二类错误的概率。

### 例 9.5

参见例 9.4,我们在  $t = 5$  的处理均值间建立了  $m = 4$  个对照(比较)。如果我们想把试验的错误率控制在水平  $\alpha_E = 0.05$  上,就取

$$\alpha_I = \frac{0.05}{4} = 0.0125$$

那么  $F$  检验的临界值是  $F_{0.0125, 1, 25} = 7.24$ , 与此相对,如果我们忽略我们正在进行有关处理均值的多重检验这个事实,那么  $F_{0.05, 1, 25} = 4.24$ 。当  $SSC_i/MS_{Error} \geq 7.24$  时,我们就拒绝  $H_0$ 。从例 9.4 得到,这四个  $F$  比率值是

$$F_1 = 13.71 \quad F_2 = 8.48 \quad F_3 = 0.24 \quad F_4 = 1.42$$

利用 Bonferroni 方法,我们断言由于对照  $l_1$  与  $l_2$  的  $F$  比值比 7.24 大,因此它们与零有显著差异。应用 Bonferroni 检验方法,我们确信在这四个检验中至少有一个犯第一类错误的概率最大为 0.05。四个检验中每一个检验应用  $\alpha = 0.05$  并不能使我们估计出这四个比较中犯第一类错误的确切概率。然而,这个值要比 0.05 大得多,很可能是 0.20。

Bonferroni 方法使得我们在保持一定的试验的第一类错误概率的前提下,在观察数据前先选择一小部分对照进行分析。在许多试验中,研究者们想要对所有的处理两两进行比较或把所有的处理与对照组进行比较。这些多重比较方法的主要区别在于每一个方法要控制的错误率类型。在下面的章节中我们将会讨论几个这样的方法。

## 9.4 Fisher(费舍尔)最小显著差异法

记得在方差分析中在我们拒绝了  $t$  个总体均值不全相等的假设后,我们感兴趣于哪一个总体均值与其他的不同。R. A. Fisher(1949)建立了一个方法用来处理一组  $t$  个总体均值的两两比较。这个方法叫作 Fisher(费舍尔)最小显著差异法(LSD)。

只有当最小显著差异法应用于独立(正交)的比较或预先计划的比较时,这个比较的 Fisher 最小显著差异法的水平  $\alpha$  才是正确的。然而,许多人发现 Fisher 最小显著差异法容易计算,因此应用在所有可能的成对比较(尤其是那些完成试验后觉得有趣的比较)中;研究者们建议只有在各处理的  $F$  检验显著时才能应用 Fisher 的 LSD 方法。这种修正后的方法有时被称为 Fisher **保护最小显著差异法**。模拟研究[Cramer and Swanson(1973)]表明保护最小显著差异法的试验的错误率控制在大约等于  $F$  检验的  $\alpha$  水平上。

我们将要介绍 Fisher 保护方法,但我们仍然把它叫作费舍尔最小显著差异法。这个方法总结如下。

### Fisher 最小显著差异方法

1. 利用方差分析检验  $H_0: \mu_1 = \mu_2 = \cdots = \mu_t$ , 备择假设为至少有一个均值与其余的不同。

2. 利用  $F = \text{MSB}/\text{MSW}$ , 如果无充足证据拒绝  $H_0$ , 则要停下来。

3. 如果拒绝了  $H_0$ , 就定义**最小显著差异**为两个样本均值之间的观察差异, 这个差异是声称相应的总体均值之间有差异时所必需的。

4. 对给定的  $\alpha$  值,  $\mu_i$  对  $\mu_j$  的最小显著差异为

$$\text{LSD}_{ij} = t_{\alpha/2} \sqrt{s_w^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

这里  $n_i$  与  $n_j$  分别是总体  $i$  与  $j$  中抽取的样本的样本容量。 $t$  为当  $\alpha = \frac{\alpha}{2}$ ,  $df$  为  $s_w^2$  的自由度时的临界值  $t$  (见附表 2)。注意当  $n_1 = n_2 = \cdots = n_t = n$  时,

$$\text{LSD}_{ij} = \text{LSD} = t_{\alpha/2} \sqrt{\frac{2s_w^2}{n}}$$

5. 接着对样本均值进行两两比较。如果  $|\bar{y}_i - \bar{y}_j| \geq \text{LSD}_{ij}$ , 则表明相应的总体均值  $\mu_i$  与  $\mu_j$  不同。

6. 对每一对总体均值的比较, 犯第一类错误的概率固定在某个特定值  $\alpha$  上。

注意: 最小显著差异法类似于对任何两个总体均值  $\mu_i$  与  $\mu_j$  进行两样本的  $t$  检验, 除了我们使用从所有的  $t$  个样本中计算的总体方差  $\sigma^2$  的联合估计量  $s_w^2$ , 而不是使用从样本  $i$  和  $j$  计算的联合样本方差。另外,  $t$  值的自由度为方差分析中的  $df_w = n_T - t$ , 而不是  $n_i + n_j - 2$ 。

### 例 9.6

参见例 9.3, 我们有 5 种不同的除草剂及每种除草剂指定  $n = 6$  个地块。方差分析列在表 9.5。

表 9.5 例 9.3 中数据的方差分析表

来源	df	SS	MS	F	p 值
处理	4	0.3648	0.0912	5.96	0.0016
误差	25	0.3825	0.0153		
总和	29	0.7472			

**解答** 我们按照 LSD 方法的 5 个步骤来解决这个问题。

**步骤 1.** 我们利用方差分析表 9.5。零假设为  $H_0: \mu_1 = \mu_2 = \cdots = \mu_5$ , 其  $F$  检验的统计量为

$$F = \frac{MS_B}{MS_W} = \frac{MS_{\text{Trt}}}{MS_{\text{Error}}} = 5.9$$

$\alpha = 0.05$ ,  $df_1 = 4$ ,  $df_2 = 25$ , 如果  $F$  超过 2.76 (见附表 8) 我们拒绝  $H_0$ 。

**步骤 2, 3.** 由于  $5.96 > 2.76$ , 我们拒绝  $H_0$ , 得出结论至少有一个总体均值与其余的不同 ( $p = 0.001$ )。

**步骤 4.** 由于样本容量均为 6, 所以比较两个均值的最小显著差异为

$$\text{LSD} = t_{\alpha/2} \sqrt{\frac{2MS_{\text{Error}}}{6}} = 2.060 \sqrt{\frac{2(0.0153)}{6}} = 0.1471$$

注意从表 2, 查  $\alpha = \frac{\alpha}{2} = 0.025$ ,  $df = 25$  可得相应的  $t$  值 (2.060)。

**步骤 5.** 当样本容量相同时, 我们并不需要对所有的样本均值进行两两比较, 因为所有的比较都应用相同的 LSD, 所以采用以下方法简便一些。

a. 把样本均值从小到大排列。

药剂	1	2	3	4	5
$\bar{y}_{i\cdot}$	1.175	1.293	1.328	1.415	1.500

### b. 计算样本差值

$$\bar{y}_{\text{最大}} - \bar{y}_{\text{最小}}$$

如果这个差值大于 LSD, 我们就断言相应的总体均值相互差异显著。接下来我们计算样本差值

$$\bar{y}_{\text{第二大}} - \bar{y}_{\text{最小}}$$

把结果与 LSD 相比较。我们继续与  $\bar{y}_{\text{最小}}$  作比较。

$$\bar{y}_{\text{第三大}} - \bar{y}_{\text{最小}}$$

等等, 直到我们发现要么所有的样本均值与  $\bar{y}_{\text{最小}}$  之差大于 LSD (因此相应的总体均值不相同), 要么有一个样本均值与  $\bar{y}_{\text{最小}}$  之差小于 LSD。要是后者的话, 我们停止与  $\bar{y}_{\text{最小}}$  作对比。针对我们的数据, 与  $\bar{y}_{\text{最小}}$  对比的结果如下:

比较	结论
$\bar{y}_{\text{最大}} - \bar{y}_{\text{最小}} = \bar{y}_{5\cdot} - \bar{y}_{1\cdot} = 0.325$	> LSD; 继续
$\bar{y}_{\text{次大}} - \bar{y}_{\text{最小}} = \bar{y}_{4\cdot} - \bar{y}_{1\cdot} = 0.240$	> LSD; 继续
$\bar{y}_{\text{第三大}} - \bar{y}_{\text{最小}} = \bar{y}_{3\cdot} - \bar{y}_{1\cdot} = 0.153$	> LSD; 继续
$\bar{y}_{\text{第四大}} - \bar{y}_{\text{最小}} = \bar{y}_{2\cdot} - \bar{y}_{1\cdot} = 0.118$	< LSD; 停止

通过下面的图表来概括得到的结果:

药剂     1     2     3     4     5

那些用下划线连起来的总体, 其均值与  $\bar{y}_{1\cdot}$  的差异不显著。注意药剂 3, 4, 5 与药剂 1 的样本差值大于 LSD, 因此没有加下划线。

c. 类似地, 再与  $\bar{y}_{\text{第二小}}$  (本例中即  $\bar{y}_{2\cdot}$ ) 作比较, 对这种情况我们沿用 (b) 部分的方法。

比较	结论
$\bar{y}_{5\cdot} - \bar{y}_{2\cdot} = 0.207$	> LSD; 继续
$\bar{y}_{4\cdot} - \bar{y}_{2\cdot} = 0.122$	< LSD; 停止

药剂     1     2     3     4     5

d. 继续与  $\bar{y}_{\text{第三小}}$  (本例中即  $\bar{y}_{3.}$ ) 作比较。

比较	结论
$\bar{y}_{5.} - \bar{y}_{3.} = 0.172$	$> \text{LSD}$ ; 继续
$\bar{y}_{4.} - \bar{y}_{3.} = 0.087$	$< \text{LSD}$ ; 停止

药剂      1      2      3      4      5

e. 继续与  $\bar{y}_{\text{第四小}}$  (本例中即  $\bar{y}_{4.}$ ) 作比较。

比较	结论
$\bar{y}_{5.} - \bar{y}_{4.} = 0.085$	$< \text{LSD}$ ; 停止

药剂      1      2      3      4      5

f. 步骤(a)至(e)可概括如下:

药剂      1      2      3      4      5

—————  
—————  
—————

按照 LSD 准则, 那些没有用下划线连接起来的总体表明其均值存在显著差异。注意在(f)部分中我们可以省略从上数第三根线, 因为从上数第二根线就包含了它。显著与非显著的结果修正后总结如下

药剂      1      2      3      4      5

—————  
—————

最后, 我们知道  $\mu_1, \mu_2$  与  $\mu_3$  显著地小于  $\mu_5$ 。另外,  $\mu_3$  与  $\mu_4$  显著地大于  $\mu_1$ 。

虽然例 9.6 中讨论的 LSD 方法似乎十分费力, 但它的应用却十分简单。首先, 进行方差分析。如果我们拒绝所有总体均值相等的零假设, 我们就对样本均值两两计算 LSD。当样本容量相同时, 各组的差异都是一个数。我们应用例 9.6 中从步骤 5(a)至 5(f)的方法。我们不需要写下所有这些过程, 只要有概括线就可以了。就像步骤 5(f)那样, 最后利用 Fisher 的 LSD 给出一个简单直观用于反映各个对比的图形作总结。

在利用 LSD 方法做两两比较时有一些问题要注意。首先, 有可能出现在方差分析中整个的  $F$  检验显著但是利用 LSD 方法任何两两的比较均不显著。这种明

显异常的发生是因为  $F$  检验的零假设  $H_0: \mu_1 = \mu_2 = \cdots \mu_t$  等价于假设总体均值间所有可能的对照(成对其他)都等于零。对一组数据,显著的对照并不一定形如  $\mu_i - \mu_j$ ,而这种形式正是我们在进行两两比较时所采用的。

其次, Fisher 的最小显著差异法也可用于建立  $\mu_i - \mu_j$  的置信区间。置信度为  $100(1 - \alpha)\%$  的置信区间的形式如

$$(\bar{y}_{i.} - \bar{y}_{j.}) \pm \text{LSD}_{ij}$$

最后,当所有的样本容量相同时,所有对照的 LSD 是

$$t_{\alpha/2} \sqrt{\frac{2s_w^2}{n}}$$

## 9.5 Tukey 的 W 方法

我们知道多重比较方法的一个主要缺陷是控制每一个比较的错误率。甚至当  $\mu_1 = \mu_2 = \cdots \mu_t$  时,除非每一个比较的错误率(比如利用 Fisher 非保护的 LSD)  $\alpha$  十分小,在多重比较中得出至少有一对均值不同的概率仍然很高。为避免如此,人们建立了另外一些多重比较方法来控制不同的错误率。

Tukey(1953)提出了利用学生化极差分布的方法。当比较两个以上的样本均值时,要检验最大与最小样本均值,我们用如下检验统计量

$$\frac{\bar{y}_{\text{最大}} - \bar{y}_{\text{最小}}}{s_p \sqrt{1/n}}$$

这里  $n$  是每个样本的观察值个数,  $s_p$  是共同的总体标准差  $\sigma$  的组合估计。这个检验统计量与两个均值比较时的检验统计量十分相似,但它不服从  $t$  分布。不服从的一个原因是当我们观察到最大与最小样本均值之前不能确定到底要比较哪两个样本均值(及总体均值)。这个方法与两个均值比较的方法十分不同,两个均值比较的方法是指定检验  $H_0: \mu_1 - \mu_2 = 0$ , 观察  $\bar{y}_1$  及  $\bar{y}_2$ , 建立一个  $t$  统计量。

量

$$\frac{\bar{y}_{\text{最大}} - \bar{y}_{\text{最小}}}{s_p \sqrt{1/n}}$$

服从学生化极差分布。我们不讨论此分布的特性,而只介绍它在 Tukey 多重比较中的用途。

### Tukey 的 W 方法

1. 对  $t$  个样本均值排序。
2. 如果

$$|\bar{y}_{i.} - \bar{y}_{j.}| \geq W$$



则两个总体均值  $\mu_i$  与  $\mu_j$  不相等, 其中

$$W = q_\alpha(t, v) \sqrt{\frac{s_W^2}{n}}$$

$s_W^2$  是自由度为  $v$  的样本组内均方,  $q_\alpha(t, v)$  是比较  $t$  个不同总体时学生化极差的上侧尾部临界值,  $n$  是每个样本的观察值个数。下面将要讨论如何从附表 10 中获得  $q_\alpha(t, v)$  的值。

3. 要控制的误差率就是**试验的错误率**。即指定试验中一组或几组成对比较被错误地判断为显著的概率指定为  $\alpha$ 。

我们可以从附表 10 中查得  $q_\alpha(t, v)$  的值。 $v$  值列在表的左列,  $t$  值横跨表的第一行。对于  $\alpha = 0.05$  及  $\alpha = 0.01$ , 学生化极差的上侧尾部值分开显示。例如, 比较 10 个总体均值,  $s_W^2$  的自由度为 9, 学生化极差对应于 0.05 的上侧尾部临界值是  $q_{0.05}(10, 9) = 5.74$ 。

### 例 9.7

参见例 9.3 中的数据。利用 Tukey 的 W 方法在  $\alpha = 0.05$  时作 5 个总体均值之间的两两比较。

**解答** 第一步将样本均值从小到大排列, 产生下表。

药剂	1	2	3	4	5
$\bar{y}_i$	1.175	1.293	1.328	1.415	1.500

从例 9.6 中描述的试验, 我们得知

$t = 5$  (我们是在 5 个均值间进行两两比较的)

$v = 25$  ( $s_W^2$  的自由度等于方差分析中的  $df_{\text{Error}}$ )

$\alpha = 0.05$  (我们指定试验的错误率  $\alpha_E$  为 0.05)

$n = 6$  (对每种药剂随机指定 6 个地块)

我们从附表 10 中查得

$$q_\alpha(t, v) = q_{0.05}(5, 25) \approx 4.158$$

每一个样本均值之差的绝对值  $|\bar{y}_i - \bar{y}_j|$  要与

$$W = q_\alpha(t, v) \sqrt{\frac{s_W^2}{n}} = 4.158 \sqrt{\frac{0.0153}{6}} = 0.2100$$

作对比。把  $W$  替换为 LSD, 我们就可以使用与例 9.6 解中的第五步相同的方法来对比样本均值了。将样本均值从小到大排序后, 我们与  $\bar{y}_{\text{最小}}$  (即  $\bar{y}_1$ ) 作对比如

下:

比较	结论
$\bar{y}_{\text{最大}} - \bar{y}_{\text{最小}} = \bar{y}_{5.} - \bar{y}_{1.} = 0.325$	$> W$ ; 继续
$\bar{y}_{\text{次大}} - \bar{y}_{\text{最小}} = \bar{y}_{4.} - \bar{y}_{1.} = 0.240$	$> W$ ; 继续
$\bar{y}_{\text{第三大}} - \bar{y}_{\text{最小}} = \bar{y}_{3.} - \bar{y}_{1.} = 0.153$	$< W$ ; 停止

通过下图总结我们得到的结果:

药剂     1     2     3     4     5

与  $\bar{y}_{\text{第二小}}$  (即  $\bar{y}_{2.}$ ) 比较, 得到

比较	结论
$\bar{y}_{5.} - \bar{y}_{2.} = 0.207$	$< W$ ; 停止

药剂     1     2     3     4     5

同样,  $\bar{y}_{5.}$  与  $\bar{y}_{3.}$  及  $\bar{y}_{4.}$  比较, 得到

药剂     1     2     3     4     5

综合以上结果, 我们得到

药剂     1     2     3     4     5

简化为

药剂     1     2     3     4     5

所有没有共同下划线的总体其均值相互差异显著; 这就是说,  $\mu_4$  与  $\mu_5$  显著地大于  $\mu_1$ 。其他均值间不存在显著差异。

通过对比最小显著差异法(例 9.6)与 Tukey 的 W 方法(例 9.7)的多重比较结论, 我们发现 Tukey 方法比 LSD 方法更加保守(发现较少的显著差异)。例如, 对例 9.6 中的数据应用 Tukey 方法, 我们发现  $\mu_3$  不再显著地大于  $\mu_1$ 。同样,  $\mu_2$  与  $\mu_3$  不再显著地小于  $\mu_5$ 。之所以这样是因为, 尽管两个方法都有试验错误率, 但是我们已经指出保护 LSD 方法比 Tukey 的 W 方法每一个比较的错误率要大。

Tukey 方法的局限性在于它要求所有的样本均值来自于相同个数的数据。样本容量不相等就不能应用。如果样本容量差别不大,米勒(1981)建议用  $n_i$  的调和平均数替换  $W$  公式中的数值  $n$ :

$$n = \frac{t}{\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_t}}$$

如果各样本容量相差较大,试验者应该考虑使用 LSD 方法或 Bonferroni 检验方法。

Tukey 方法也可以用来建立两个均值之差的置信区间。然而,与 Fisher LSD 建立的置信区间不同, Tukey 方法使我们可同时建立所有对处理差异的置信区间。对某指定  $\alpha$ , 可计算出相应的  $W$ , 形如

$$(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) \pm W$$

的置信区间包含所有的差异  $\mu_i - \mu_j$  的整个概率为  $1 - \alpha$ 。这就是说,所有的区间  $(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) \pm W$  包含相应总体差异  $\mu_i - \mu_j$  的概率是  $1 - \alpha$ 。

## 9.6 Student-Newman-Keuls 方法

**Student-Newman-Keuls(SNK)**方法对 Tukey 的  $W$  方法作了修正。虽然 SNK 方法仍然使用学生化极差统计量,但依据步骤的不同使用不同的临界值来区分待检验的均值。为比较这两个方法,我们来看例 9.3。将样本均值从小到大排列如下表:

药剂	1	2	3	4	5
$\bar{y}_{i\cdot}$	1.175	1.293	1.328	1.415	1.500

Tukey 的  $W$  方法的学生化极差临界值为

$$q_{\alpha}(t, v) = q_{0.05}(5, 25) \approx 4.158$$

对五个处理均值的所有两两比较都使用这个相同的  $q$  值。

SNK 方法中,当  $t$  个样本均值从小到大排列时,距离  $r$  步的均值间的临界值为

$$W_r = q_{\alpha}(r, v) \sqrt{\frac{s_W^2}{n}}$$

我们这个例子中,  $\bar{y}_{\text{最大}}$  与  $\bar{y}_{\text{最小}}$  相差五步,要比较它们应该使用

$$W_5 = q_{\alpha}(5, v) \sqrt{\frac{s_W^2}{n}} = q_{0.05}(5, 25) \sqrt{\frac{0.0153}{6}} = 4.158 \sqrt{\frac{0.0153}{6}} = 0.2100$$

(注意:这就是 Tukey W 方法中的 W), 然而,  $\bar{y}_{\text{最大}}$  与  $\bar{y}_{\text{第二小}}$  相差四步, 与它们比较的应该是

$$\begin{aligned} W_4 &= q_\alpha(4, v) \sqrt{\frac{s_w^2}{n}} \\ &= q_{0.05}(4, 25) \sqrt{\frac{0.0153}{6}} = 3.892 \sqrt{\frac{0.0153}{6}} = 0.19650 \end{aligned}$$

表 9.6 中列出了有关例 9.3 数据的所有临界值  $W_r$ 。在附表 10 中, 用  $r$  替换  $t$ , 因为表中未给出  $v=25$ , 故取其近似就可得到值  $q_\alpha(r, v)$ 。

表 9.6 例 9.3 中  $r, q_\alpha(r, v)$  和  $W_r$  的值

$r$	2	3	4	5
$q_\alpha(r, v)$	2.915	3.523	3.892	4.158
$W_r$	0.1472	0.1779	0.1965	0.2100

Student-Newman-Keuls 方法在决定观察到的样本差异显著性时依赖这两个样本均值间排序的步宽, 它既没有试验的错误率也没有每一比较的错误率。而且, 错误率是根据均值的相应步宽而定义的。由于随着要比较的均值间的步宽的减小临界值  $W_r$  也在减小, SNK 方法就不那么保守, 因此一般比 Tukey 的 W 方法发现较多的显著差异, 这是因为 Tukey 方法不管要比较的均值相差几步都使用最大的 W 值。事实上, Tukey 的 W 的临界值是  $W_t$ , 而对所有  $r < t$  均有  $W_r < W_t$ 。另外, 当  $n_i$  都相同时, 若记 Fisher LSD 的临界值为  $W_2$ , 那么对所有  $r > 2$  均有  $LSD < W_r$ 。因此, SNK 通常比 Fisher LSD 方法发现较少具有显著差异的组。

SNK 方法总结如下。

#### SNK 方法

1. 将  $t$  个样本均值从小到大排列。
2. 对步长为  $r$  的两个均值  $\bar{y}_i$  与  $\bar{y}_j$ , 如果它们满足

$$|\bar{y}_i - \bar{y}_j| \geq W_r$$

我们就认为两者具有差异, 其中  $W_r = q_\alpha(r, v) \sqrt{s_w^2/n}$ ,  $n$  为每个样本观察值个数,  $s_w^2$  是方差分析表中样本组内均方,  $v$  为  $s_w^2$  的自由度,  $q_\alpha(r, v)$  是学生化极差的临界值。附表 10 列出了  $\alpha=0.05$  及  $\alpha=0.01$  时  $q_\alpha(r, v)$  的值。

(注意: 利用列上的  $t$  定位想要的  $r$  值。)

#### 例 9.8

参见例 9.3 中的数据。运用 SNK 方法在  $\alpha=0.05$  上作所有的两两比较。

### 解答

表 9.6 中列出了  $W_r$  的临界值。

1. 从  $\bar{y}_{\text{最大}}$  开始, 每一个样本均值都与  $\bar{y}_{\text{最小}}$  比较, 使用合适的  $W_r$  值。结果总结如下表。

比较	$W_r$	结论
$\bar{y}_5 - \bar{y}_1 = 0.325$	0.2100	$> W_5$ ; 继续
$\bar{y}_4 - \bar{y}_1 = 0.240$	0.1965	$> W_4$ ; 继续
$\bar{y}_3 - \bar{y}_1 = 0.153$	0.1779	$< W_3$ ; 停止

2. 类似的, 与  $\bar{y}_{\text{第二小}}$  对比。

比较	$W_r$	结论
$\bar{y}_5 - \bar{y}_2 = 0.207$	0.1965	$> W_4$ ; 继续
$\bar{y}_4 - \bar{y}_2 = 0.122$	0.1779	$< W_3$ ; 停止

3. 接着, 与  $\bar{y}_{\text{第三小}}$  对比。

比较	$W_r$	结论
$\bar{y}_5 - \bar{y}_3 = 0.122$	0.1779	$< W_3$ ; 停止

4. 最后, 与  $\bar{y}_{\text{第四小}}$  对比。

比较	$W_r$	结论
$\bar{y}_5 - \bar{y}_4 = 0.085$	0.1471	$< W_2$ ; 停止

利用 SNK 方法, 这些多重比较的结果展示如下:

药剂     1     2     3     4     5

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

所有不带共同下划线的总体其均值相互间存在显著差异; 这就是说,  $\mu_4$  与  $\mu_5$  显著

地大于  $\mu_1$ 。同理,  $\mu_5$  显著地大于  $\mu_2$ , 除此之外 SNK 方法未发现其他存在显著差异的均值对比。这个例子告诉我们这样一个事实: SNK 方法往往比 Tukey 的  $W$  方法发现较多的差异(因此较不保守)。

SNK 方法经修改后可以应用于样本容量不同的情形。当  $n_i \neq n_j$  时,  $r$  步均值的  $W_r$  值可以象 Tukey 方法那样做修正。这里如果  $n_i$  间差异不大的话, 将  $W_r$  公式中的  $n$  用  $n_i$  的调和平均数替换:

$$n = \frac{t}{\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_t}}$$

如果样本容量差异很大, 试验者应该考虑使用 LSD 方法或 Bonferroni 方法。

## 9.7 Dunnett 方法:处理组与对照组的比较

在许多研究与试验中, 研究者们为了便于比较, 通常包含对照(或者说控制)处理。对照的种类很多, 但通常对照只是作为一个标准以便其他处理与之对比。例如, 在许多情况下, 试验得以进行的条件对因变量产生极强的影响, 以至于在试验中有效的处理通常不能产生所希望的响应。比如, 如果昆虫数量太多的话, 大多数杀虫剂在合理用量时并不能显著地减少昆虫数量。因此, 在所喷洒区域内使用没有有效成分的喷雾剂作为对照用以显示昆虫的水平。需要使用对照的第二种情况是参与试验者不管接受任何合理的处理(治疗)都产生希望看到的响应。这被称为**安慰剂效应**。在这类研究或试验中, 完全象对待接受积极治疗的参与者一样, 随机指派一些参与者组成对照组。在大多数评价新的药物或治疗手段的临床试验中, 为了测定新药或新疗法效果的大小都要使用安慰剂处理。最后, 对照可以代表现行方法或标准方法, 这是对任何新方法进行对比的基础。

在包含对照组(或者说控制组)的试验中, 研究者们想测定接受治疗组的平均效果是否不同于控制组。Dunnett(1955)建立了一个与对照组作对比的方法, 此方法控制的是试验的第一类错误率。这个方法将每一个处理均值与对照组均值  $\bar{y}_c$  作对比并比较样本均值之差  $\bar{y}_i - \bar{y}_c$  与临界差异

$$D = d_\alpha(k, v) \sqrt{\frac{s_w^2}{n}}$$

其中,  $n_c = n_1 = \cdots = n_{t-1} = n$ 。Dunnett 方法要求样本容量相同,  $n_i = n_c$ 。表 11(附录)列出了值  $d_\alpha(k, v)$ 。Dunnett(1964)针对  $n_i$  不全相等的情况对表 11 中的数据作了调整。对比可以是单侧的也可以是双侧的, 具体总结如下。

### Dunnett 方法

1. 对某指定  $\alpha_E$  值, 比较  $\mu_i$  与  $\mu_c$ (对照组均值)的 Dunnett  $D$  值为

$$D = d_{\alpha}(k, v) \sqrt{\frac{2s_W^2}{n}}$$

其中,  $n$  为第  $i$  个处理组与对照组的共同样本容量,  $k = t - 1$  为非对照处理的个数;  $\alpha$  为想要的试验错误率;  $s_W^2$  是组内均方;  $v$  是  $s_W^2$  的自由度;  $d_{\alpha}(k, v)$  为 Dunnett 临界值(附表 11)。

2. 对双边备择假设  $H_a: \mu_i \neq \mu_c$ , 如果满足下式则认为  $\mu_i$  不等于  $\mu_c$

$$|\bar{y}_i - \bar{y}_c| \geq D$$

其中  $d_{\alpha}(k, v)$  值是表 11 中的双边值。

3. 对单边备择假设  $H_a: \mu_i > \mu_c$ , 如果满足下式则认为  $\mu_i$  大于  $\mu_c$

$$(\bar{y}_i - \bar{y}_c) \geq D$$

其中  $d_{\alpha}(k, v)$  值是表 11 中的单边值。

4. 对单边备择假设  $H_a: \mu_i < \mu_c$ , 如果满足下式则认为  $\mu_i$  小于  $\mu_c$

$$(\bar{y}_i - \bar{y}_c) \leq -D$$

其中  $d_{\alpha}(k, v)$  值是表 11 中的单边值。

5. 要控制的第一类错误率是一个试验的错误率。因此, 一个试验中一个或多个与对照处理的比较被错判为显著这种情况出现的概率指定为  $\alpha$ 。

### 例 9.9

参见例 9.3 中的数据。在  $\alpha = 0.05$  下, 分别将两种生物处理及两种化学处理与对照处理进行对比。

**解答** 我们想断定生物及化学处理是否提高了红花草的产量, 因此我们需要作对照的单边比较。

1. 从例 9.3 中, 我们知道  $s_W^2 = 0.0153$ , 其自由度为 25, 包含对照处理在内共 5 种处理。查表 11 中的单边部分可以得到 Dunnett 方法的临界值, 将下列值代入

$$\alpha = 0.05 \quad k = 5 - 1 = 4 \quad v = 25$$

得到  $d_{0.05}(4, 25) = 2.28$ 。由于  $n_c = n_2 = n_3 = n_4 = n_5 = 6$ , 故

$$\begin{aligned} D &= d_{\alpha}(k, v) \sqrt{\frac{2s_W^2}{n}} \\ &= 2.28 \sqrt{\frac{2(0.0153)}{6}} = 0.163 \end{aligned}$$

2. 如果  $(\bar{y}_i - \bar{y}_c) \geq 0.163$ , 我们就说处理均值  $\mu_i$  大于控制均值  $\mu_c$ 。总结这些对比如下

处理	$(\bar{y}_i - \bar{y}_t)$	比较	结论
生物 1	$(1.293 - 1.175) = 0.118$	$< D$	比对照组不大
生物 2	$(1.328 - 1.175) = 0.153$	$< D$	比对照组不大
化学 1	$(1.415 - 1.175) = 0.240$	$> D$	比对照组大
化学 2	$(1.500 - 1.175) = 0.325$	$> D$	比对照组大

3. 结论:在这些地块上使用任何一种生物药剂与不使用任何药剂相比红花草的平均产量并没有提高。因此,在  $\alpha = 0.05$  水平下,在红花草地块中生物药剂不能有效控制杂草。然而,使用化学药剂与不使用任何药剂相比红花草的平均产量似乎要大。

当样本容量不等时, Dunnett 方法的试验错误率不等于  $\alpha$ 。Dunnett(1964)针对样本容量不等时的情况修正了表 11 中的数据。

## 9.8 Scheffé 的 S 方法

到目前为止,我们所讨论的五种多重比较方法都是处理  $t$  个总体均值间两两比较的问题。Scheffé(1953)提出了一个更一般的方法,该方法可以用来作  $t$  个总体均值间的任何可能的比较。虽然 Scheffé 方法也可以用来作  $t$  个总体均值间的两两比较,但它比其他三个多重比较方法在检测总体均值间两两差异显著性方面要保守一些(不敏感),这是因为 Scheffé 方法与两两比较相比其包括的比较“家族”较大。

### 多重比较的 Scheffé S 方法

1. 考虑  $t$  个总体均值的任意一个线性对照

$$l = a_1\mu_1 + a_2\mu_2 + \cdots + a_t\mu_t$$

我们希望检验零假设

$$H_0: l = 0$$

相应的备择假设为

$$H_a: l \neq 0$$

2. 检验统计量为

$$\hat{l} = a_1\bar{y}_1 + a_2\bar{y}_2 + \cdots + a_t\bar{y}_t$$

3. 设

$$S = \sqrt{\hat{V}(\hat{l})} \sqrt{(t-1)F_{\alpha, df_1, df_2}}$$

其中,从 9.2 节中有

$$\hat{V}(\hat{l}) = s_w^2 \sum_i \frac{a_i^2}{n_i}$$



$t$  是总体均值的个数,  $F_{\alpha, df_1, df_2}$  是  $F$  分布的上侧临界值,  $\alpha = \alpha$ ,  $df_1 = t - 1$ ,  $df_2$  是  $s_w^2$  的自由度。

4. 对指定值  $\alpha$ , 如果  $|\hat{l}| > S$ , 则拒绝  $H_0$

5. 要控制的错误率为试验的错误率。考虑到任何可能的线性对照, 试验中一个或多个线性对照被错判为显著的概率指定为  $\alpha$ 。

### 例 9.10

参见例 9.3。我们在例 9.3 中定义了 5 个处理均值的 4 个对照, 试图考察控制(无药剂组)或四种除草剂之一使地块上红花草的平均产量的差异。利用样本数据及 Scheffé 方法确定这四个对照中哪些显著地不等于零。使用  $\alpha = 0.05$ 。

**解答** 表 9.7 给出了这四个感兴趣的对照及其估计量。为解释表 9.7 中涉及的计算, 我们计算第一个对照, 即无药剂对药剂的  $S$  值。要计算

$$S = \sqrt{\hat{V}(\bar{l})} \sqrt{(t-1)F_{\alpha, df_1, df_2}}$$

我们必须先计算  $\hat{V}(\bar{l})$ , 利用公式

$$\hat{V}(\bar{l}) = s_w^2 \sum_i \frac{a_i^2}{n_i}$$

因为所有样本均值都等于 6,  $s_w^2 = 0.0153$ , 则

$$\hat{V}(\bar{l}) = 0.0153 \left( \frac{(4)^2}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \right) = 0.0153 \frac{20}{6} = 0.0510$$

表 9.7 杂草控制试验中 Scheffé 方法的计算

对照	处理					$\Sigma a_i^2/n_i$	$\bar{l}$	$\hat{V}(\bar{l})$	$S$	结论
	控制	生物 1	生物 2	化学 1	化学 2					
	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$					
控制与药剂	4	-1	-1	1	-1	20/6	-0.836	0.0510	0.750	显著
生物与化学	0	1	1	-1	-1	4/6	-0.294	0.0102	0.336	不显著
生物 1 与生物 2	0	1	-1	0	0	2/6	-0.035	0.0051	0.237	不显著
化学 1 与化学 2	0	0	0	1	-1	2/6	-0.085	0.0051	0.237	不显著

查表 8,  $\alpha = 0.05$ ,  $df_1 = t - 1 = 4$ ,  $df_2 = 25$  ( $s_w^2$  的自由度),  $F_{0.05, 4, 25} = 2.76$ , 那么  $S$  的值为

$$S = \sqrt{0.0510} \sqrt{4(2.76)} = (0.2258)(3.323) = 0.750$$

因为  $\bar{l}$  的绝对值  $|-0.836| = 0.836$  超过了 0.750, 我们就有显著的证据 ( $\alpha = 0.05$ ) 表明地块上使用一种药剂比不使用任何药剂平均红花草产量高。其他三个对照的计算概括在表 9.7 里。注意不同的对照具有不同的 S 值, 此例中, 惟一与零存在显著差异的是第一组。其他三个对照在  $\alpha = 0.05$  下不显著。这些结论与例 9.4 中我们得到的结论有所不同, 那时我们发现第二个对照与零也存在显著差异。这两个结论不同的原因在于 Scheffé 方法在 0.05 水平上控制试验的第一类错误率, 而在例 9.4 中我们仅在 0.05 水平上控制单个比较。

Scheffé 方法可以用来建立任何可能的  $t$  个处理均值对照 (未必是成对比较) 的同时置信区间。具体地, 所有可能的形如  $l = \sum a_i \mu_i$ , 其中  $\sum a_i = 0$  的对照被包含进形如

$$\bar{l} - S < l < \bar{l} + S$$

区间的概率等于  $1 - \alpha$ 。

现在让我们回到本章开始时介绍的案例上。

**案例中数据的分析** 研究的目的是调查面试者对应试者的评价是否受到应试者身体残疾的影响。在检验假设及对五种处理作出比较前, 我们需要证明这项研究满足这些检验及多重比较方法的条件要求。

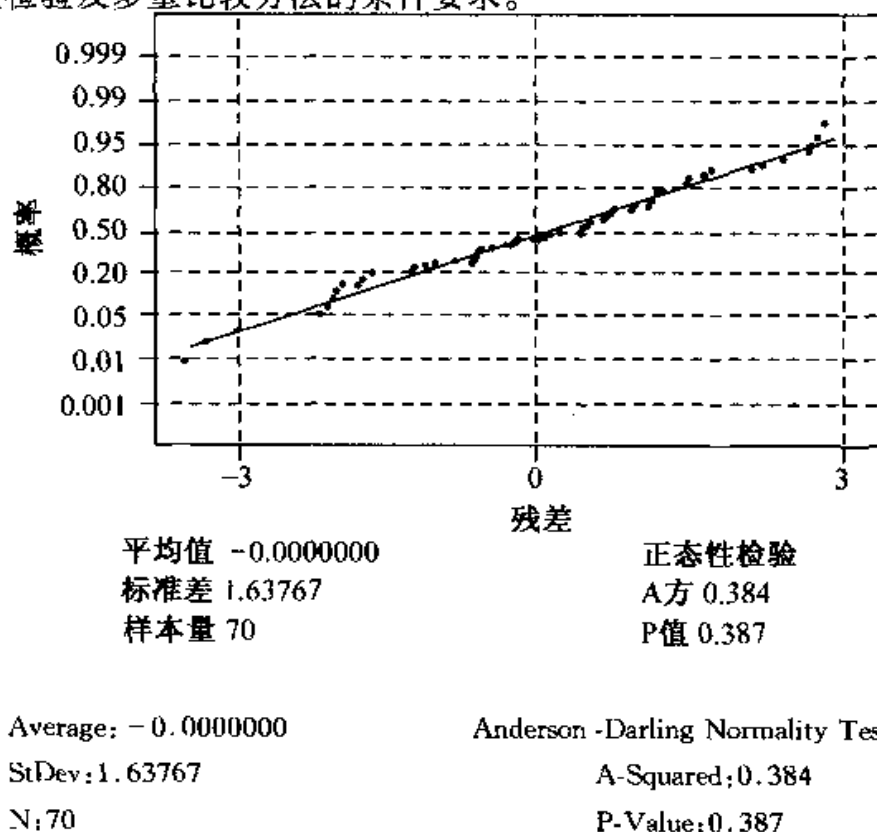


图 9.2 残差的正态概率图

从图 9.1 中可以看出, 盒形图宽度几乎相同, 没有离群值, 胡须长度几乎一样。五组应试者中的每一组平均值及中位数大致相同。因此, 方差分析的假设看起来成立。要验证这个观察, 我们计算残差并将其绘在正态概率图上。(参见图 9.2。)从此图上我们可以看到除了个别的两个值外, 所有的点几乎位于一条直线上。这样, 强有力的证据显示五个应试者资格定级的总体服从正态分布。

其次, 我们可以检验方差相等的假设。从盒形图看出, 散布大体相同。从表 9.2 提供的概括统计量来看, 我们注意到标准差范围从 1.537 到 1.961。这样, 样本标准差差异非常小。为证实这个观察, 我们利用 Levine 检验法检验方差齐性。我们要检验

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$$

$H_a$ : 方差不全相等

经计算得  $L = 0.405$ 。临界值为  $F_{0.05, 4, 25} = 2.76$ 。因此, 我们不能拒绝  $H_0$ 。更进一步, 我们计算得到  $p$  值为  $P(F_{4, 25} \geq 0.405) = 0.803$ ; 我们有自信说本研究中没有违反方差齐性的假设。

我们通过与研究者讨论试验进行的方式就可检查数据是否独立。务必确保在整个研究中, 观看录相带的房间中条件保持不变, 这样做是为了避免引起评定人员分心而影响评级效果。另外, 首先要检查移情分数的分布在五组评定人员中是相同的, 以确保在五组评定人员对应试者的资格进行评价前他们中不存在移情水平的差异。研究的假设是五组残疾情况中各平均资格定级  $\mu_i$  存在差异:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$H_a$ : 至少有一个均值不等于其他均值

计算机输出的方差分析表如下。输出中使用下列符号: 对照组(C), 听力障碍(H), 截肢(A), 拐杖(R)及轮椅(W)。

#### General Linear Models Procedure

Dependent Variable: Rating

Source	DF	Sum of Squares	F value	Pr > F
Model	4	30.47800000	2.68	0.0394
Error	65	185.05642857		
Corrected Total	69	215.53442857		

## Dunnett's One-tailed T Tests for Variable:Rating

Note: This test controls the type I experimentwise error for comparisons of all treatments against a control.

Alpha = 0.05 Confidence = 0.95 df = 65 MSE = 2.847022

Critical Value of Dunnett's T = 2.203

Minimum Significant Difference = 1.4049

Comparisons significant at the 0.05 level are indicated by ' \* \* \* '.

		Simultaneous		Simultaneous
		Lower	Difference	Upper
HC		Confidence	Between	Confidence
Comparison		Limit	Means	Limit
R	- C	- 0.3907	1.0143	2.4192
W	- C	- 0.9407	0.4643	1.8692
A	- C	- 1.8692	- 0.4643	0.9407
H	- C	- 2.2549	- 0.8500	0.5549

## Tukey's Studentized Range (HSD) Test for Variable:Rating

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha = 0.05 df = 65 MSE = 2.847022

Critical Value of Studentized Range = 3.968

Minimum Significant Difference = 1.7894

Means with the same letter are not significantly different.

Tukey Grouping		Mean	N	HC
	A	5.9143	14	R
B	A	5.3643	14	W
B	A	4.9000	14	C
B	A	4.4357	14	A
B		4.0500	14	H

从结果看,  $F$  检验的  $p$  值为 0.0394。因此, 五类残疾的平均定级存在显著差异。我们下一步要考察各组间定级存在什么类型的差异。我们对比对照组与有残疾的四组——拐杖(R), 轮椅(W), 截肢(A), 听力障碍(H), 在  $\alpha_F = 0.05$  水平下使用 Dunnett 方法。我们对四个残疾组中任何组的平均定级是否低于对照组采用单边检验:

$$H_0: \mu_i \geq \mu_c$$

$$H_a: \mu_i < \mu_c$$

我们得到结论: 对照(无残疾)组的平均定级并不比任何一个残疾组的平均定级显著地低。接着, 我们运用多重方法确定哪两个组的平均定级存在差异。分析采用 Tukey 方法,  $\alpha = 0.05$ , 结果显示在计算机输出中。具有相同 Tukey 分组字母的残疾类型的平均定级间不存在显著差异。因此, 拄拐杖应试者的平均定级显著地高于听力障碍应试者的平均定级。没有发现其他存在显著差异的组。要考察五类残疾条件下所有均值差异的大小, 我们利用 Tukey 方法, 计算十组平均差异的 95% 的同时置信区间。下面的计算机输出列出了这些区间。

Alpha = 0.05 Confidence = 0.95 df = 65 MSE = 2.847022

Critical Value of Studentized Range = 3.968

Minimum Significant Difference = 1.7894

Comparisons significant at the 0.05 level are indicated by ' \* \* \* '.

HC Comparison	Simultaneous Lower Confidence Limit	Difference Between Means	Simultaneous Upper Confidence Limit	
R - W	-1.2394	0.5500	2.3394	
R - C	-0.7751	1.0143	2.8037	
R - A	-0.3108	1.4786	3.2680	
R - H	0.0749	1.8643	3.6537	* * *
W - C	-1.3251	0.4643	2.2537	
W - A	-0.8608	0.9286	2.7180	
W - H	-0.4751	1.3143	3.1037	
C - A	-1.3251	0.4643	2.2537	

C - H	- 0.9394	0.8500	2.6394
A - H	- 1.4037	0.3857	2.1751

Contrast	DF	Contrast SS	F Value	PR > F
Control vs Handicap	1	0.01889286	0.01	0.9353
Hearing vs Mobility	1	14.82148810	5.21	0.0258
Crutches vs Amp. & Wheel	1	9.60190476	3.37	0.0709

最后,我们建立了几个对照用以评价研究者提出的其他问题。下面列出了这些问题及相应的对照。

问题	对照
无残疾定级对残疾定级	$4\mu_C - \mu_R - \mu_W - \mu_A - \mu_H$
听力定级对运动障碍定级	$0\mu_C - \mu_R - \mu_W - \mu_A + 3\mu_H$
拐杖定级对截肢及轮椅定级	$0\mu_C + 2\mu_R - \mu_W - \mu_A + 0\mu_H$

对这三个对照进行检验的零假设与备择假设形式如下:

$$H_0: l = 0$$

$$H_a: l \neq 0$$

我们使用 Bonferroni 方法在  $\alpha_E = 0.05$  下检验这三个假设。单个比较的错误率设为  $\alpha_l = \alpha_E/3 = 0.05/3 = 0.0167$ 。这样,如果这三个对照中任意一个对照的  $F$  检验的  $p$  值小于 0.0167,我们就说这个对照显著地不等于零。从计算机输出看,这三个  $p$  值分别是 0.9353, 0.0258 及 0.0709。结果,这三个对照与零均不存在显著差异。

五种平均定级中惟有一个存在显著差异的组是听力障碍的应试者与拄拐杖的应试者。研究者们在文章里详细讨论了这个差异发生的原因。

**报告结论** 总结此项研究结果的报告应该包含以下及各部分:

1. 陈述研究目的。
2. 描述研究方案,评定人员如何选择,面试如何进行。
3. 讨论研究结果的可推广性。
4. 数据集的数值及图形概括。
5. 描述所有的推断方法:

方差分析及  $F$  检验。

多重比较方法, 线性对照及置信区间。

验证满足推断方法所需要的所有条件。

6. 讨论结果及结论。

7. 解释与先前研究相比的发现。

8. 对未来研究的建议。

9. 列出数据集。

## 9.9 小 结

我们介绍了三种不同的多重比较方法(Fisher, Tukey 及 SNK)来对  $t$  个总体均值作两两比较。另一个方法, Scheffé, 可以应用于均值的任意线性对照(包括成对比较)。对每一个方法, 我们竭力指出它控制的是哪一类错误及其与其他的方法相比保守程度如何。因为所有两两及多重比较方法计算差值  $|\bar{y}_i - \bar{y}_j|$  来确定  $\mu_1$  与  $\mu_2$  的差异, 我们通过比较两个方法间各自显著所需要的差异值来感受其保守的程度。使用例 9.3 中的数据, 此信息列于表 9.8。

从表 9.8 可以看到, Scheffé 方法十分保守。其临界差异比 Tukey 的  $W$  大 13%, 比 LSD 大 61%。因此, 我们在进行处理均值的成对比较时从不建议使用 Scheffé 方法。另外, 当  $r = 2$  时, LSD 的值等于 SNK 的值; 当  $r = t = 5$  时, Tukey  $W$  的值等于 SNK 的值。因此, SNK 是界于开放的 LSD 与保守的  $W$  之间的一种折衷方法。

表 9.8 相距  $r$  步的样本均值的临界差异

方法	均值相距的步数			
	2	3	4	5
LSD	0.1471	0.1471	0.1471	0.1471
SNK	0.1471	0.1783	0.1969	0.2100
Tukey	0.2100	0.2100	0.2100	0.2100
Scheffé	0.2373	0.2373	0.2373	0.2373

我们应该使用哪种方法呢? 在比较效果时我们通常较喜欢 SNK 方法, 在与控制组(对照组)比较时较喜欢 Dunnett 方法。但是, 这些选择的原因与我们的工作环境及决策规则有很大关系。因为我们的环境也许完全不同于你们, 使用哪一个方法, 什么时候使用, 要依赖于个人。对给定的问题, 确定有关差异的决定通常是

否应该较(或较不)保守, 然后选择符合具有所希望的特点的方法。

重要公式

1. Fisher 的 LSD 方法

$$LSD = t_{\alpha/2} \sqrt{s_W^2 \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

2. Tukey 的 W 方法

$$W = q_{\alpha}(t, v) \sqrt{\frac{s_W^2}{n}}$$

3. SNK 方法

$$W_r = q_{\alpha}(r, v) \sqrt{\frac{s_W^2}{n}}$$

4. 当样本容量差异不大时, Tukey 及 SNK 方法中  $n$  的近似值

$$n = \frac{t}{\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_t}}$$

5. Dunnett 方法

$$D = d_{\alpha}(k, v) \sqrt{\frac{s_W^2}{n}}$$

6. Scheffé 方法

$$S = \sqrt{\hat{V}(l)} \sqrt{(t-1)F_{\alpha, df_1, df_2}}$$

其中

$$\hat{V}(l) = s_W^2 \sum_i \frac{a_i^2}{n_i}$$

补充练习

9.6(农业) 参见例 8.1 的数据。园艺家要考察三种不同品种苹果树叶子的磷含量。

- a. 进行方差分析。
- b. 使用 SNK 方法检验所有的两两比较, 取  $\alpha = 0.05$ 。
- c. 将(b)部分的结论与下面的 SAS 计算机输出结果进行对比。

General Linear Models Procedure  
Class Level Information



Class	Levels	Values
VARIETY	3	1 2 3

Number of observations in data set = 15

Dependent Variable: P PHOSPHORUS

Source	DF	Sum of Squares	F Value	Pr > F
Model	2	0.27664000	16.97	0.0003
Error	12	0.09780000		
Corrected Total	14	0.37444000		

R-Square	C.V.	P Mean
0.738810	13.93169	0.64800000

Source	DF	Type III SS	F Value	Pr > F
VARIETY	2	0.27664000	16.97	0.0003

Student-Newman-Keuls test for variable: PHOSPHORUS

NOTE: This test controls the type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha = 0.05 df = 12 MSE = 0.00815

Number of Means 2 3

Critical Range 0.1244028 0.1523193

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	VARIETY
A	0.77600	5	2
A	0.70800	5	3
B	0.46000	5	1

**9.7(医学)** 研究者作了一项试验,即将四种新减肥药的效果与目前已有的一种减肥药对比。研究者随机将 50 个男人的随机样本分成相等的五组, A 制剂指定给第一组, B 制剂指定给第二组, 等等。接着在正式研究前对每一个参加试验的人作体格检查并告知他超重的磅数。对比各组超重的平均磅数发现它们不存在显著差异。研究者接着开始进行试验, 每一组都服用指定的制剂一段固定的时间。在研究期结束时记录每个人的体重下降数据如下:

$A_1$	12.4	10.7	11.9	11.0	12.4	12.3	13.0	12.5	11.2	13.1
$A_2$	9.1	11.5	11.3	9.7	13.2	10.7	10.6	11.3	11.1	11.7
$A_3$	8.5	11.6	10.2	10.9	9.0	9.6	9.9	11.3	10.5	11.2
$A_4$	12.7	13.2	11.8	11.9	12.2	11.2	13.7	11.8	11.5	11.7
S	8.7	9.3	8.2	8.3	9.0	9.4	9.2	12.2	8.5	9.9

标准药剂标记为药剂 S, 四种新药剂分别标记为  $A_1$ ,  $A_2$ ,  $A_3$  及  $A_4$ 。下面列出了数据及计算机打印输出的分析结果。

# General Linear Models Procedure

## Class Level Information

Class	Levels	Values
AGENT	5	1 2 3 4 S

Number of observations in data set = 50

Dependent Variable: L WEIGHTLOSS

SOURCE	DF	Sum of Squares	F Value	Pr > F
Model	4	61.61800000	15.68	0.0001
Error	45	44.20700000		
Corrected Total	49	105.82500000		

R-Square	C. V.	L Mean
0.582263	9.035093	10.9700000

Source	DF	Type III SS	F Value	Pr > F
AGENT	4	61.61800000	15.68	0.0001

Level of	L		
A	N	Mean	SD
1	10	12.0500000	0.82898867
2	10	11.0200000	1.12130876

3	10	10.2700000	1.02637442
4	10	12.2400000	0.75601293
S	10	9.2700000	1.15859110

FISHER'S LSD for variable:WEIGHTLOSS

Alpha = 0.05 df = 45 MSE = 0.982378

Critical Value of T = 2.01

Least Significant Difference = 0.8928

Means with the same letter are not significantly different.

T Grouping	Mean	N	A
A	12.2400	10	4
A	12.0500	10	1
B	11.0200	10	2
B	10.2700	10	3
C	9.2700	10	5

Student-Newman-Keuls test for variable:L

Alpha = 0.05 df = 45 MSE = 0.982378

Number of Means	2	3	4	5
Critical Range	0.8927774	1.0742812	1.1824729	1.2594897

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	A
A	12.2400	10	4
A	12.0500	10	1
B	11.0200	10	2
B	10.2700	10	3
C	9.2700	10	5

Tukey's Studentized Range(HSD)Test for variable: L

Alpha = 0.05 df = 45 MSE = 0.982378

Critical Value of Studentized Range = 4.018

Minimum Significant Difference = 1.2595

Means with the same letter are not significantly different.

Tukey Grouping		Mean	N A		
		A	12.2400	10	4
		A	12.0500	10	1
	B	A	11.0200	10	2
	B	C	10.2700	10	3
		C	9.2700	10	5

Dunnnett's One-tailed T tests for variable;L

Alpha = 0.05 Confidence = 0.95 df = 45 MSE = 0.982378

Critical Value of Dunnnett's T = 2.222

Minimum Significant Difference = 0.9851

Comparisons significant at the 0.05 level are indicated by ' \* \* \* '.

		Simultaneous		Simultaneous	
		Lower	Difference	Upper	
A		Confidence	Between	Confidence	
Comparison		Limit	Means	Limit	
4	- s	1.9849	2.9700	3.9551	* * *
1	- s	1.7949	2.7800	3.7651	* * *
2	- s	0.7649	1.7500	2.7351	* * *
3	- s	0.0149	1.0000	1.9851	* * *

Univariate Procedure

Variable = RESIDUAL

# Moments

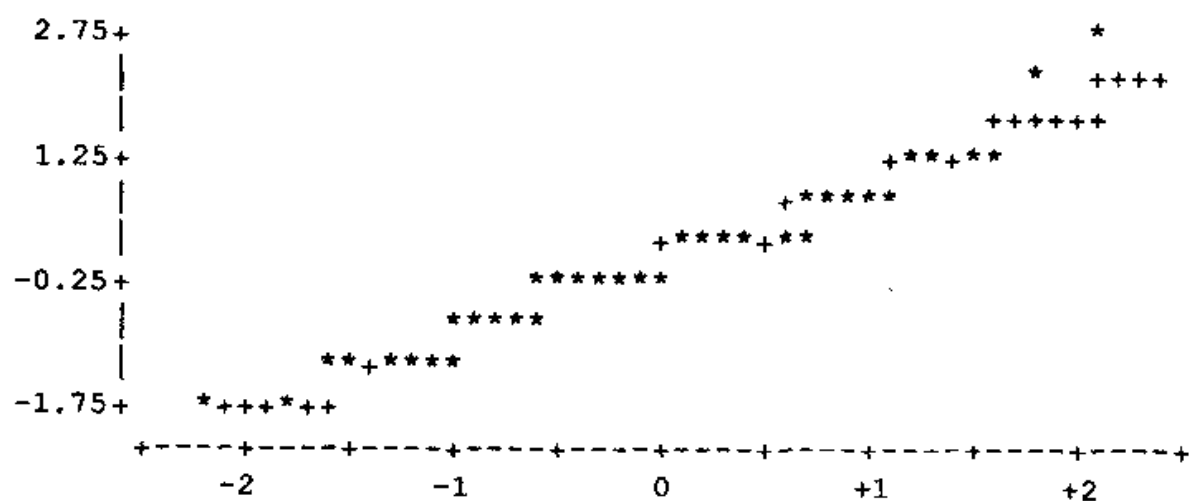
N	50	Sum Wgts	50
Mean	0	Sum	0
Std Dev	0.949833	Variance	0.902184
Skewness	0.523252	Kurtosis	0.995801
Test of Normality:		P-value	0.6737

Variable = RESIDUAL

Stem	Leaf	#	Boxplot
2	9	1	0
2	2	1	0
1	5	1	
1	00003	5	
0	556679	6	+ - - - - +
0	0112233444	10	+
-0	444433321100	12	+ - - - - +
-0	88765	5	+ - - - - +
-1	4331000	7	
-1	98	2	

- - - + - - - + - - - + - - - +

# Normal Probability Plot



进行方差分析确定五种减肥药间是否存在显著差异,用  $\alpha = 0.05$ 。有违反方差分析假定的现象吗?您将得到有关使用这五种不同的药剂导致平均体重降低的什么结论?

**9.8** 参见练习 9.7。从其中的计算机输出,利用下列方法确定哪两个组均值间存在显著差异。

a. Fisher 的 LSD,  $\alpha = 0.05$

b. Tukey 的  $W$ ,  $\alpha = 0.05$

c. SNK 方法,  $\alpha = 0.05$

**9.9** 参见练习 9.8。对下列每一种情况,确定使用哪一个多重比较方法最合适。

a. 研究者对错判任何一组药物差异显著十分关心。

b. 研究者十分关心在总体均值存在差异时,没有将这组药剂判为差异显著。

**9.10** 参见练习 9.7。研究者想知道哪一种新药与标准药相比能显著地减少平均体重。判断时使用  $\alpha = 0.05$ 。

**9.11** 参见练习 9.7。假设减肥的形式如下:

$A_1$ : 药物治疗、锻炼及社会咨询。

$A_2$ : 药物治疗、锻炼及不社会咨询。

$A_3$ : 药物治疗、社会咨询及不锻炼。

$A_4$ : 药物治疗、不锻炼及不社会咨询。

建立对照比较各药剂均值以表述如下问题:

a. 比较标准均值与四种药剂均值的平均数。

b. 比较包含社会咨询的药剂均值与不包含社会咨询的药剂均值。(忽略标准)

c. 比较包含锻炼的药剂均值与不包含锻炼的药剂均值。(忽略标准)

d. 比较包含社会咨询的药剂均值与标准药剂均值。

**9.12** 参见练习 9.11。利用某多重检验方法确定哪一个对照与零有显著差异 ( $\alpha = 0.05$ )。针对研究者提出的寻找最有效的减肥办法这个问题解释您的结论。

**9.13** 参见练习 8.2。

a. 与目前市场上四种牌子的香烟比较,新的品牌“低油”的平均焦油含量低吗?  $\alpha = 0.05$ 。

b. “低油”的平均焦油含量与四种牌子中每一种的平均焦油含量差异有多大? 使用 95% 的置信区间。

**9.14** 参见练习 8.6。品种 A 与 B 种植在得克萨斯州,品种 C 与 D 种植在佛罗里达州,品种 E 种植在加利福尼亚州。

a. 得克萨斯州两个品种的联合平均产量与佛罗里达州两个品种的联合平均

产量相同吗?  $\alpha = 0.05$ 。

- b. 得克萨斯与佛罗里达四个品种的联合平均产量与加利福尼亚品种的平均产量相同吗?  $\alpha = 0.05$ 。

9.15 参见练习 8.11。

- a. 将除草剂 1 与除草剂 2 的平均产量与对照处理对比, 取  $\alpha = 0.05$ 。

b. (a)中的方法是单侧的还是双侧的

c. 解释(a)中的结论。

9.16 参见练习 8.16。

- a. 使用合适的多重比较方法比较三个分组的平均分数, 取  $\alpha = 0.05$ 。

b. 从平均分数的差异中您能得到什么结论? 导致差异产生的分组的本质是什么?

9.17(农业) 用三个不同品种的根瘤菌接种红三叶草, 其氮含量如下:

| 3DOK1 | 3DOK5 | 3DOK7 |
|-------|-------|-------|
| 19.4  | 18.2  | 20.7  |
| 32.6  | 24.6  | 21.0  |
| 27.0  | 25.5  | 20.5  |
| 32.1  | 19.4  | 18.8  |
| 33.0  | 21.7  | 18.6  |
|       | 20.8  | 20.1  |
|       |       | 21.3  |

- a. 从平均氮含量看, 有证据显示三种处理的效果不同吗? 完整分析数据并在此基础上得出结论。  $\alpha = 0.01$ 。

b. 在(a)的分析中, 有违反所需条件的证据吗?

9.18(农业) 科学家作了一项试验用来比较对草种籽进行处理的三种不同方法: 机械划破种子表皮(ms), 酸浸(ad), 热水泡(hw)。科学家在 150 个皮氏培养皿中每一个都放入 100 粒草的种籽。在这 150 个培养皿中, 随机指定 50 个 ms, 50 个 ad, 50 个 hw。两周后, 科学家检查了每一个器皿中的发芽率。

| 方法 | 平均发芽率 | 标准差 |
|----|-------|-----|
| ms | 65.3  | 7.2 |
| ad | 82.1  | 5.4 |
| hw | 73.8  | 6.5 |

对这些数据进行单向方差分析并得出结论, 取  $\alpha = 0.05$ 。

**9.19** 参见练习 9.18。利用 SNK 方法鉴别方法间的差异, 总结您的结果。

**9.20(环境)** 为评价四种不同混合汽油的相对优点, 研究者使用 20 辆相同类型, 型号及马力大小的汽车作了一项试验, 随机指定 5 辆使用四种混合汽油中的一种。混合汽油的试验数据概括如下。

| 混合                     | 均值(英里/加仑) | 标准差 |
|------------------------|-----------|-----|
| 1(对照)                  | 26.2      | 4.3 |
| 2(对照 + 添加剂 $x$ )       | 28.1      | 5.6 |
| 3(对照 + 添加剂 $y$ )       | 29.6      | 5.1 |
| 4(对照 + 添加剂 $x$ 和 $y$ ) | 38.2      | 7.3 |

分析并得出结论, 取  $\alpha = 0.05$ 。

**9.21** 参见练习 9.20。考虑下列线性对照。指出每一种对照测量的意义。

a.  $\bar{l}_1 = \bar{y}_1 + \bar{y}_2 - \bar{y}_3 - \bar{y}_4$

b.  $\bar{l}_2 = \bar{y}_1 + \bar{y}_3 - \bar{y}_2 - \bar{y}_4$

c.  $\bar{l}_3 = \bar{y}_1 - \bar{y}_2 - \bar{y}_3 + \bar{y}_4$

**9.22** 利用 Scheffé 方法检验练习 9.21 中  $\bar{l}_3$  的显著性,  $\alpha = 0.05$ 。得出什么结论?

**9.23** 利用 Scheffé 方法建立练习 9.21 中  $\bar{l}_1$  与  $\bar{l}_2$  的置信区间。得出什么结论?

**9.24(兽医)** 研究者作了一项试验用来研究三种药物对控制拉布拉多猎狗肩部肌肉中的脂肪含量的效果。他们将 80 只狗随机分成四个处理组。A 组的狗是控制组, 在 B, C 及 D 组中每一组中的食物里加入三种新的犬恶丝虫药物之一。从四组中每一组随机选取 10 只狗接受从 4 个月到两年长度不等的处理。测量了肩部肌肉中脂肪含量的百分比并列在下边。

| 考察时间 | 处理组  |      |      |      |
|------|------|------|------|------|
|      | A    | B    | C    | D    |
| 4 个月 | 2.84 | 2.43 | 1.95 | 3.21 |
|      | 2.49 | 1.85 | 2.67 | 2.20 |
|      | 2.50 | 2.42 | 2.23 | 2.32 |



续表

| 考察时间 | 处理组   |       |       |       |
|------|-------|-------|-------|-------|
|      | A     | B     | C     | D     |
| 8 个月 | 2.42  | 2.73  | 2.31  | 2.79  |
|      | 2.61  | 2.07  | 2.53  | 2.94  |
|      | 2.23  | 2.83  | 2.32  | 2.45  |
|      | 2.48  | 2.59  | 2.36  | 2.49  |
|      | 2.48  | 2.53  | 2.46  | 2.95  |
| 1 年  | 2.23  | 2.73  | 2.04  | 2.05  |
|      | 2.65  | 2.26  | 2.30  | 2.31  |
|      | 2.30  | 2.70  | 2.85  | 2.53  |
|      | 2.30  | 2.54  | 2.75  | 2.73  |
|      | 2.38  | 2.70  | 2.62  | 2.65  |
| 2 年  | 2.05  | 2.81  | 2.50  | 2.84  |
|      | 2.13  | 2.70  | 2.69  | 2.92  |
|      | 2.64  | 3.24  | 2.90  | 2.91  |
|      | 2.56  | 3.71  | 3.02  | 2.89  |
|      | 3.30  | 2.95  | 3.78  | 3.21  |
| 平均   | 2.19  | 3.01  | 2.96  | 2.89  |
|      | 2.45  | 3.08  | 2.87  | 2.68  |
| 平均   | 2.411 | 2.694 | 2.605 | 2.698 |

假设满足方差分析的要求,研究者接着进行方差分析来评价四种处理下狗的脂肪含量的平均百分比的差异。方差分析中的计算没有考虑药物的时间长短。方差分析表如下。

| 来源 | df | SS      | MS     | F 比  | p 值    |
|----|----|---------|--------|------|--------|
| 处理 | 3  | 1.0796  | 0.3599 | 3.03 | 0.0345 |
| 误差 | 76 | 9.0372  | 0.1189 |      |        |
| 总和 | 79 | 10.1168 |        |      |        |

- a. 四个处理组中平均脂肪含量存在显著差异吗?  
 b. 三个经犬恶丝虫药物处理的组中有没有哪一组的平均脂肪含量比控制组高?

9.25 参见练习 9.24。假设研究者推断新药导致脂肪含量提高且随着对狗用药时间的增加脂肪含量也在增加。怎样利用已知数据解决这个问题?

9.26(医学) 文章“The Ames Salmonel/microsome mutagenicity assay: Issues of inference and validation”(1989, *Journal of American Statistical Association* (美国统计学会杂志), 84:651-661)讨论了化学诱因突变对人类健康的重要性及导致初生玻(基)斑岩化验诱变的生理基础, 艾姆斯氏沙门氏菌/微粒体化验。在艾姆斯氏试验中, 从单个样本中获得的变量是分布大约  $10^8$  个细菌的可见菌落的数量。艾姆斯氏试验的一般协议包括某控制剂量下的多个样本及四或五个对数等距剂量的试验化合物。每个剂量取 20 个样本进行试验得到数据如下。剂量水平为毫克/样本。

| 剂量   | 可见菌落数 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | $\bar{y}_i$ | $s_i^2$ |
|------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------------|---------|
| 控制   | 11    | 13  | 14  | 14  | 15  | 15  | 15  | 15  | 16  | 17  | 17  | 18  | 18  | 19  | 20  | 21  | 22  | 23  | 25  | 27  | 17.8        | 17.5    |
| 0.3  | 39    | 39  | 42  | 43  | 44  | 45  | 46  | 50  | 50  | 50  | 51  | 52  | 52  | 52  | 55  | 61  | 62  | 63  | 67  | 70  | 51.7        | 81.0    |
| 1.0  | 88    | 90  | 92  | 92  | 102 | 104 | 104 | 106 | 109 | 113 | 117 | 117 | 119 | 119 | 120 | 120 | 121 | 122 | 130 | 133 | 110.9       | 175.4   |
| 3.0  | 222   | 233 | 251 | 251 | 253 | 255 | 259 | 275 | 276 | 283 | 284 | 294 | 299 | 301 | 306 | 312 | 315 | 323 | 337 | 340 | 283.5       | 1131.5  |
| 10.0 | 562   | 587 | 595 | 604 | 623 | 666 | 689 | 692 | 701 | 702 | 703 | 706 | 710 | 714 | 733 | 739 | 763 | 782 | 786 | 789 | 692.3       | 4584.4  |

我们想确定随着剂量增加平均的菌落数是否也有增加的趋势。得到结论的一种方法是使用一个对照, 其常数  $a_i$  由下列方法确定: 假设处理水平是连续变量  $x$  的  $t$  个值  $x_1, x_2, \dots, x_t$ , 令  $a_i = x_i - \bar{x}$  及  $\hat{l} = \sum a_i \bar{y}_i$ 。如果  $\hat{l}$  与零有显著差异且大于零, 那么我们说  $\mu_i$  间存在正增长趋势。如果  $\hat{l}$  与零有显著差异且小于零, 那么我们说  $\mu_i$  间存在负增长趋势。此试验中, 剂量水平是处理  $x_1 = 0, x_2 = 0.3, x_3 = 1.0, x_4 = 3.0, x_5 = 10.0$ , 其均值为  $\bar{x} = 2.86$ 。因此这个对照中的系数是  $a_1 = 0 - 2.86 = -2.86, a_2 = 0.3 - 2.86 = -2.56, a_3 = 1.0 - 2.86 = -1.86, a_4 = 3.0 - 2.86 = +0.14, a_5 = 10.0 - 2.86 = +7.14$ 。我们因此需要评价下列均值间的对照

$$-2.86\bar{y}_0 - 2.56\bar{y}_{0.3} - 1.86\bar{y}_{1.0} + 0.14\bar{y}_{3.0} + 7.14\bar{y}_{10.0}$$

的显著性。如果此对照与零有显著差异且大于零, 那么我们说剂量均值间存在正增长趋势。

- a. 检验剂量均值间是否存在正增长趋势, 取  $\alpha = 0.05$ 。  
 b. 是否有违反(a)中进行检验所需必要条件的证据呢? 给出一种方法使我们能够正确检验是否存在正的趋势。

**9.27** 在评价面试者的决定与应试者残疾类型有关的案例中, 评定人员是 70 名在校大学生, 在所有工作面试的录像带中申请工作者及面试者都是同样的男演员。

- 这项研究中使用在校大学生作为计算机销售职位申请者的资格的评定人员, 试讨论其缺陷。
- 在所有的五盘面试录像带中使用相同的两个演员, 试讨论其正负面。
- 讨论不改变申请者寻求工作种类的局限性。

**9.28(医学)** 论文“The effect of an endothelin-receptor antagonist, bosentan, on blood pressure in patients with essential hypertension”(“内层神经末梢对抗剂 bosentan 对原发性高血压病人的血压的效果”)(1998, *The New England Journal of Medicine* (新英格兰医学杂志), 338:784-790)讨论了 bosentan 对调节原发性高血压病人的血压的贡献。243 名患有轻度到中度原发性高血压的病人参加了这项研究。在运用一段安慰剂后, 随机选择患者口服四种剂量(每天 100 毫克、500 毫克或 1000 毫克, 或每次 1000 毫克且每天两次)之一的 bosentan 或安慰剂。在治疗开始前及治疗 4 周后分别测量血压。研究从起点到终点反映了从治疗前的基线到经过 4 周治疗后血压水平的变化。下表中列出了数据的统计量。

|      | 血压的变化 |       |       |         |         |
|------|-------|-------|-------|---------|---------|
|      | 安慰剂   | 100mg | 500mg | 1,000mg | 2,000mg |
| 舒张压  |       |       |       |         |         |
| 均值   | -1.8  | -2.5  | -5.7  | -3.9    | -5.7    |
| 标准差  | 6.71  | 7.30  | 6.71  | 7.21    | 7.30    |
| 收缩压  |       |       |       |         |         |
| 均值   | -0.9  | -2.5  | -8.4  | -10.3   | -10.3   |
| 标准差  | 11.40 | 11.94 | 11.40 | 11.80   | 11.94   |
| 样本容量 | 45    | 44    | 45    | 43      | 44      |

- 在哪一个水平上, 与安慰剂相比, 舒张压有显著的下降? 取  $\alpha = 0.05$ 。
- 在此项研究中, 为什么使用对照(控制)处理是很重要的?
- 忽略安慰剂, 只用其余的四个剂量, 构造一个对照, 检验当剂量增加时, 收缩压的下降量是否有线性的增长的趋势。建立对照的方法见练习 9.26。
- 取  $\alpha = 0.05$ , 对于四个治疗剂量, 用 SNK 方法检验平均收缩压下降的两两差异。
- 研究人员称他们的试验为双盲试验。解释这个名词的含义。

## 第十章 类型数据

- 10.1 引言和案例
- 10.2 总体比例  $\pi$  的推断
- 10.3 两总体比例之差  $\pi_1 - \pi_2$  的推断
- 10.4 多比例的推断:卡方拟合优度检验
- 10.5 Poisson(泊松)分布
- 10.6 列联表:独立性检验和齐性检验
- 10.7 相关程度的度量
- 10.8 几率和优比
- 10.9 小结

### 10.1 引言和案例

直到现在,我们一直关心的样本数据主要是定量数据。然而,我们有时会遇到这样的情况,其中要考察的变量所取水平只能以名称或等级表示,并且我们感兴趣的是在观测中每一水平出现的次数。通过对这种类型变量观测所得到的数据被称为**类型数据**或**计数数据**。例如,装配线上的产品可以划分为三个质量等级:可接受,二等品,拒绝。同样地,研究交通问题需要知道沿着某一主干路进入城市的通勤者所使用交通工具的类型和数量。研究污染问题的人或许关心从湖里取出的样品中有多少种藻类以及每一种藻类出现的次数。消费者保护组织关心处方收费调查的结果,以比较大城市不同地区药物清单上的价格。

本章,我们将介绍基于类型数据的推断方法。

#### 案例:特许专营权的扩大以及专营店位置的选择

某汽车旅馆连锁店的总经理准备作出一项决定,要在四家集团公司中选择一家,给予他们特许专营权,以便在快速增长的“边缘城区”(办公和购物非常集中的郊区)开设一家新的汽车旅馆。总经理要求对四家集团公司新近开设的旅馆进行一次调查,以比较他们的顾客满意程度。

**数据收集方案的设计** 四家公司都在这一地区开设了其他汽车旅馆,因此,连锁店要求得到这些旅馆过去顾客的地址,以便将顾客满意问卷调查表邮寄给他们,从而收集顾客满意情况的信息。本来,可以把四家公司最近对各自的旅馆所进行的调查的分析结果作为进一步分析的基础,然而,这四家公司在调查过程中所提问题的形式和措词有很大不同。这使得数据分析变得不太可能,因为数据不一致。

问卷应被周密地设计,以调查顾客就其所住四家公司的汽车旅馆中对每个旅馆的满意情况。总经理所关心的两个关键问题应该是顾客对硬件的质量(如房间、餐厅、健身设施等)的评价和对服务质量的评价。顾客对两个问题的评价都使用5分制,1表示最差,3表示中等,5表示优秀。基于假设检验的功效和置信区间估计的精度所需的样本容量,应该从每家公司抽取至少100名顾客。但是,由于问卷的预计回收率仅为20%,因此,应从每家公司至少抽取500名顾客,并邮寄问卷。

**数据的整理和分析** 当然,总经理愿意将特许经营权授予获得最佳评价结果的公司。由于和这四家公司所达成的财务安排很相似,因此决定主要将依赖于这个评价的结果。总经理认识到,抽样调查只覆盖一小部分顾客。然而,抽样调查能够最公正地得到连锁店顾客对四家相互竞争的公司进行评价的信息。

总经理并不能肯定是强调对硬件质量的评价,还是强调对服务质量的评价。另外,是对所有等级都进行比较,还是只集中在平均中等以上的等级(4或5)。

将调查问卷的结果归纳成下面两张表,一个是对硬件质量的评价结果,一个是对服务质量的评价结果。总经理要求你分析评价结果,并对四个公司进行排序。他担心仅调查这么一小部分顾客如何能精确地估计出所有顾客对汽车旅馆的评价情况。本章所介绍的方法将能够使你回答总经理提出的问题。

#### 案例中硬件等级的频率

| 等级 | 公司  |     |     |     | 总和  |
|----|-----|-----|-----|-----|-----|
|    | G1  | G2  | G3  | G4  |     |
| 1  | 11  | 8   | 15  | 6   | 40  |
| 2  | 10  | 6   | 18  | 5   | 39  |
| 3  | 51  | 50  | 42  | 38  | 181 |
| 4  | 30  | 41  | 26  | 40  | 137 |
| 5  | 22  | 29  | 16  | 40  | 107 |
| 总和 | 124 | 134 | 117 | 129 | 504 |

#### 案例中服务等级的频率

| 等级 | 公司 |    |    |    | 总和 |
|----|----|----|----|----|----|
|    | G1 | G2 | G3 | G4 |    |
| 1  | 15 | 16 | 23 | 11 | 65 |
| 2  | 18 | 21 | 17 | 18 | 74 |

续表

| 等级 | 公司  |     |     |     | 总和  |
|----|-----|-----|-----|-----|-----|
|    | G1  | G2  | G3  | G4  |     |
| 3  | 36  | 31  | 33  | 24  | 124 |
| 4  | 29  | 35  | 21  | 33  | 118 |
| 5  | 26  | 31  | 23  | 43  | 123 |
| 总和 | 124 | 134 | 117 | 129 | 504 |

## 10.2 总体比例 $\pi$ 的推断

在第四章讨论的二项试验中, 每次试验只可能有两个结果, 记为“成功”和“失败”。我们令  $\pi$  为成功的比例,  $1 - \pi$  为失败的比例。记  $y$  为  $n$  次重复试验中成功出现的次数, 则  $y$  的概率分布为

$$P(y) = \frac{n!}{y!(n-y)!} \pi^y (1-\pi)^{n-y}$$

我们通常直观地选择二项分布参数  $\pi$  的点估计。某总体中被归类为成功的个体所占的比例为  $\pi$ , 从该总体中随机抽取容量为  $n$  的样本, 则参数  $\pi$  的最佳估计为样本中成功个体的比例。记  $y$  为  $n$  次观测中成功出现的次数, 则样本比例为

$$\hat{\pi} = \frac{y}{n}$$

在 4.12 节, 我们看到  $y$  具有丘形概率分布, 且当

$$n \geq \frac{5}{\min(\pi, 1-\pi)} \quad (\text{或等价地 } n\pi \geq 5, n(1-\pi) \geq 5)$$

时,  $y$  的概率分布可以用正态分布曲线来近似。

同样地,  $\hat{\pi} = y/n$  的分布也可用均值和标准差分别为有下式给出的正态分布来近似。

**$\hat{\pi}$  的均值和标准误差**

$$\mu_{\hat{\pi}} = \pi$$

$$\sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

用正态分布近似  $\hat{\pi}$  分布所需的条件和用正态分布近似  $y$  分布所需的条件相同。实际上, 当  $n$  很大时,  $y$  分布和  $\hat{\pi}$  分布的正态近似程度更高。因此, 本章中, 我们将假定  $\hat{\pi}$  分布能够用正态分布充分地近似, 并且将统计推断建立在以前正态分布的研究结果上。

我们能够用第五章估计  $\mu$  的方法估计  $\pi$  的置信区间, 只要用  $\hat{\pi}$  代替  $\bar{y}$ ,  $\sigma_{\hat{\pi}}$  代替  $\sigma_{\bar{y}}$  即可。这样, 二项分布参数的  $100(1-\alpha)\%$  置信区间如下。

$\pi$  的置信区间, 置信系数  $(1-\alpha)$

$$\hat{\pi} \pm z_{\alpha/2} \hat{\sigma}_{\hat{\pi}} \text{ 或 } (\hat{\pi} - z_{\alpha/2} \hat{\sigma}_{\hat{\pi}}, \hat{\pi} + z_{\alpha/2} \hat{\sigma}_{\hat{\pi}})$$

其中

$$\hat{\pi} = \frac{y}{n}, \quad \sigma_{\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

### 例 10.1

在研究新的治疗方法用以治疗癌症病人的过程中, 研究人员通常要评价新疗法的有效性, 为此, 他们要报告所有经过新方法治疗的病人中能够存活一定时间的病人所占的比例。用一种新的基因工程的方法对 870 位患有某种癌的病人进行治疗, 结果经过治疗以后, 有 330 位病人至少存活 5 年。利用 90% 置信区间, 试估计在经过基因方法治疗以后, 所有患有这种癌的病人中能够至少存活 5 年的病人所占的比例。

**解答** 根据这些数据, 得

$$\hat{\pi} = \frac{330}{870} = 0.38$$

$$\sigma_{\hat{\pi}} = \sqrt{\frac{(0.38)(0.62)}{870}} = 0.016$$

本例的置信系数是 0.90。在第五章中, 我们通过查附表 1 对应于面积  $\alpha/2$  的  $z$  值, 得  $z_{\alpha/2}$ 。由于置信系数为 0.90, 从而得对应于面积 0.05 的  $z$  值是 1.645。因此, 在经过基因疗法治疗以后, 所有患有这种癌的病人中能够至少存活 5 年的病人所占比例的 90% 置信区间为

$$0.38 \pm 1.645(0.016) \text{ 或 } 0.38 \pm 0.026$$

假定  $n$  充分大, 那么, 根据二项分布的正态近似可得到  $\pi$  的置信区间。我们规定  $n\pi$  和  $n(1-\pi)$  至少应等于 5, 但  $\pi$  是未知参数, 所以, 我们通常要求  $n\hat{\pi}$  和  $n(1-\hat{\pi})$  至少为 5。当样本容量太小, 从而违反了这一规定时, 置信区间往往太宽以至于失去意义。例如,  $n=20$ ,  $\hat{\pi}=0.2$  并不满足规定, 因为  $n\hat{\pi}=4$ 。由这些数据所得的置信区间为  $0.025 < \pi < 0.375$ , 该区间没有实际意义。如果预计的市场份额增加在 0.025 到 0.375 之间, 那么, 很少有生产经理愿意推出新产品。

当  $\pi$  非常接近于 0 或 1 时,  $\pi$  的估计会出现另外一个问题。在这种情况下, 除非样本容量非常大, 否则, 相应的总体比例通常被估计为 0 或 1。这些估计没有实际意义, 因为这将表示总体中要么没有“成功”, 要么没有“失败”。正因为如此, 在

对  $\pi$  进行估计时,不用前面所给的  $\hat{\pi}$  计算公式,而是对此作适当调整,以避免估计太极端。其中一个调整如下:

$$\text{当 } y=0 \text{ 时, } \hat{\pi}_{Adj.} = \frac{\frac{3}{8}}{\left(n + \frac{3}{4}\right)},$$

$$\text{当 } y=n \text{ 时, } \hat{\pi}_{Adj.} = \frac{\left(n + \frac{3}{8}\right)}{\left(n + \frac{3}{4}\right)}.$$

在  $y=0$  或  $y=n$  的情况下,求  $\pi$  的置信区间时,利用正态近似得到的置信区间将无效。我们用下面的置信区间,他们是根据二项分布推导的。

当  $y=0$  时,置信区间为  $(0, 1 - (\alpha/2)^{1/n})$ 。

当  $y=n$  时,置信区间为  $((\alpha/2)^{1/n}, 1)$ 。

### 例 10.2

一种 PC 操作系统被研制出来。设计者声称新系统能够和几乎所有的目前运行在 Microsoft's Windows 操作系统上的计算机程序兼容。随机抽取 50 个程序,将其在新操作系统上运行,结果无一出错。试估计  $\pi$ , 其中  $\pi$  表示所有 Microsoft's Windows 兼容程序中,在新操作系统上运行而不出差错的程序所占的比例。计算  $\pi$  的 95% 置信区间。

**解答** 如果我们用  $\pi$  的标准估计,那么,我们将得到

$$\hat{\pi} = \frac{50}{50} = 1.0$$

因此,我们得出结论:100% 的 Microsoft's Windows 兼容程序都可在新操作系统上运行而没有任何改变。这个结论正确吗? 可能不正确,因为,我们只调查了一小部分 Microsoft's Windows 兼容程序。因此,我们将使用调整的估计量和置信区间。点估计为

$$\hat{\pi}_{Adj.} = \frac{\left(n + \frac{3}{8}\right)}{\left(n + \frac{3}{4}\right)} = \frac{\left(50 + \frac{3}{8}\right)}{\left(50 + \frac{3}{4}\right)} = 0.993$$

$\pi$  的 95% 置信区间为

$$((\alpha/2)^{1/n}, 1) = ((0.05/2)^{1/50}, 1) = ((0.025)^{0.02}, 1) = (0.929, 1.0)$$

从而,我们得出结论:我们有相当的把握(95%)相信所有 Microsoft's Windows 兼容程序中,能够在新操作系统上运行而不出差错的程序所占的比例很高(92.9%到 100%)。



然而,必须记住,样本容量充分大并不能保证区间将带来多少信息。它只是判断用正态分布近似二项分布的充分性,这乃是确定置信水平的基础。

估计  $\pi$  时样本容量的计算非常接近于推断  $\mu$  时所介绍的样本容量计算方法。为了使  $\pi$  的  $100(1-\alpha)\%$  置信区间估计为  $\hat{\pi} \pm E$  (其中  $E$  是事先给定的), 所需样本容量可通过求解下面方程得到, 即  $n$  满足

$$z_{\alpha/2} \sigma_{\hat{\pi}} = E$$

解得结果如下。

$\pi$  的形如  $\hat{\pi} \pm E$  的  $100(1-\alpha)\%$  置信区间所需要的样本容量

$$n = \frac{z_{\alpha/2}^2 \pi (1-\pi)}{E^2}$$

注: 由于  $\pi$  未知, 因此, 要么根据经验猜测一个值, 要么取  $\pi = 0.5$ 。若取  $\pi = 0.5$ , 则对于宽度为  $2E$  的置信区间, 这将得到最大可能的样本容量, 因此, 产生一个样本容量的保守估计。

### 例 10.3

某一大的公众民意调查机构打算进行一项全国性的抽样调查以估计在职成年人中担心下一年会失去工作的人所占的比例。该机构必须调查多少名工人才能得到 95% 置信区间的精度在 0.02 以内?

**解答** 根据设计, 该机构想要的区间形如  $\hat{\pi} \pm 0.02$ 。为了取得这一精度, 样本容量应为

$$n = \frac{z_{\alpha/2}^2 \pi (1-\pi)}{E^2}$$

其中  $z_{\alpha/2} = 1.96$ ,  $E = 0.02$ 。如果前一次调查是最近做的, 那么, 我们就用前一次调查的样本比例代替  $\pi$ ; 否则, 取  $\pi = 0.5$ 。若取  $\pi = 0.5$ , 则所需样本容量为

$$n = \frac{(1.96)^2 (0.5)(0.5)}{(0.02)^2} = 2,401$$

即, 必须调查 2,401 名工人才能使得  $\pi$  的估计精度在 0.02 以内。

关于二项分布参数  $\pi$  的统计检验非常类似于第五章所介绍的总体均值的大样本检验。将这些检验小结如下, 其中有三个备择假设以及相应的拒绝域。记住: 对于某一具体问题, 仅选择其中一种备择假设。

$\pi$  的统计检验小结,  $\pi_0$  已知

$$\begin{array}{ll} H_0: & 1. \pi \leq \pi_0 \quad H_a: \quad 1. \pi > \pi_0 \\ & 2. \pi \geq \pi_0 \quad 2. \pi < \pi_0 \\ & 3. \pi = \pi_0 \quad 3. \pi \neq \pi_0 \end{array}$$

$$\text{T.S.: } z = \frac{\hat{\pi} - \pi_0}{\sigma_{\hat{\pi}}}$$

R.R.: 对于第一类错误概率  $\alpha$

1. 若  $z > z_{\alpha}$ , 则拒绝  $H_0$ 。
2. 若  $z < z_{\alpha}$ , 则拒绝  $H_0$ 。
3. 若  $|z| > z_{\alpha/2}$ , 则拒绝  $H_0$ 。

注: 在  $H_0$  假设下,  $\sigma_{\hat{\pi}} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$ , 同时  $n$  必须满足  $n\pi_0 \geq 5$  和  $n(1-\pi_0) \geq 5$ 。验证假定, 并作出结论。

#### 例 10.4

某小镇上的跑车车主抱怨州机动车检测站根据与家庭型轿车不同的标准来检测他们的车。以前的记录表明 30% 客车第一次不能通过检测。随机抽取 150 部跑车进行检测, 其中有 60 部未能通过检测。是否有充分证据表明跑车的首次不通过率高于客车? 给定  $\alpha = 0.05$ 。

**解答** 合适的检验方法如下:

$$H_0: \pi \leq 0.30$$

$$H_a: \pi > 0.30$$

$$\text{T.S.: } z = \frac{\hat{\pi} - \pi_0}{\sigma_{\hat{\pi}}}$$

R.R.: 对于  $\alpha = 0.05$ , 若  $z > 1.645$ , 则拒绝  $H_0$ 。

利用样本数据, 得

$$\hat{\pi} = \frac{60}{150} = 0.4 \quad \sigma_{\hat{\pi}} = \sqrt{\frac{(0.3)(0.7)}{150}} = 0.037$$

并且,  $n\pi_0 = 150(0.3) = 45 \geq 5$

$$n(1 - \pi_0) = 150(0.7) = 105 \geq 5$$

检验统计量

$$z = \frac{0.4 - 0.3}{0.037} = 2.70$$

由于  $z$  的观察值超过 1.645, 因此我们得出结论, 认为机动车检测站检测的跑车, 其第一次不通过率大于 0.3 ( $p$  值 = 0.0035)。然而, 我们必须谨慎一些, 不要把这种不同归因于跑车和家庭型轿车的检测标准不同。为了更好地找出跑车第一次不通过率较高的原因, 我们也需要进行平行的假设检验, 以考察跑车和其他类型轿车之间是否存在差异。

大多数计算机程序包并不包括这种检验。然而, 通过编码, 将成功编为 1, 失

败编为 0, 我们就能近似地进行这种检验。如果程序包中包含  $z$  检验, 那么这个方法就奏效。令  $\sigma$  等于  $\sqrt{(\pi_0)(1-\pi_0)}$ 。如果程序包中包含一个单样本的  $t$  检验, 那么结果将有所不同, 但往往有足够的近似程度。例如, 假定我们要比较两种款式的产品。我们随机抽取 100 名潜在的消费者, 他们需要在新旧款式中进行选择。只有在有充分证据表明大多数消费者宁愿选择新款式的情况下, 厂家才愿意推出新款式。因此, 我们取零假设为  $H_0: \pi \leq 0.5$ , 备择假设为  $H_a: \pi > 0.5$ 。如果 100 名消费者中有 68 名宁愿选新款式, 那么, 我们就能利用 Minitab 软件进行  $z$  检验和近似的  $t$  检验。为此, 我们在标有“Yes-No”的列中输入 68 个 1, 32 个 0。于是, 得到如下 Minitab 结果; 注意: 我们已经给定  $\sigma = \sqrt{(\pi_0)(1-\pi_0)} = \sqrt{(0.5)(1-0.5)} = 0.5$ 。

```
MTB> ZTest .5 .5 'Yes_No';
SUBC> Alternative 0.
```

Z-Test

```
Test of mu = 0.5000 vs mu not = 0.5000
The assumed sigma = 0.500
```

| Variable | N   | Mean   | StDev  | SE Mean | Z    | P-Value |
|----------|-----|--------|--------|---------|------|---------|
| Yes_No   | 100 | 0.6800 | 0.4688 | 0.0500  | 3.60 | 0.0003  |

```
MTB> TTest 0.5 'Yes_No';
SUBC> Alternative 0.
```

T-Test of the Mean

```
Test of mu = 0.500 vs mu not = 0.5000
```

| Variable | N   | Mean   | StDev  | SE Mean | T    | P-Value |
|----------|-----|--------|--------|---------|------|---------|
| Yes-No   | 100 | 0.6800 | 0.4688 | 0.0469  | 3.84 | 0.0002  |

这种  $z$  检验方法的结果和我们通过手工计算的结果一模一样。 $t$  检验方法的结果就不完全一样, 主要是因为该方法使用样本标准差而不是  $\sqrt{(\pi_0)(1-\pi_0)}$ 。

然而,结论相同,即:由于  $p$  值很小,因而强烈支持研究假设。虽然,基于 1 和 0 的  $t$  检验和  $z$  检验并不十分相同,但是,一般情况下,这两种结果是如此相同,以至于我们不必太担心它们之间的差异。

我们说  $\pi$  的  $z$  检验是近似检验,且当  $n$  很大,  $\pi_0$  并不太接近于 0 或 1 时,这种近似的效果很好。接下来,一个自然的问题就是:我们什么时候使用它? 回答这一问题有好几条规则:没有哪个规则是不可冒犯的。很多研究表明:如果  $n\pi_0$  或  $n(1-\pi_0)$  小于 2, 那么  $z$  检验的结果值得怀疑。如果  $n\pi_0$  和  $n(1-\pi_0)$  至少为 5,  $z$  检验基本上是精确的。当样本容量一定时,和  $\pi_0$  适中如 0.05 或 0.10 时的  $z$  检验相比,当  $\pi_0$  取极端值(如 0.001)时的检验精确度较差。例如,  $n\pi_0 = 1.2$  时的  $H_0: \pi = 0.0001$  检验比  $n\pi_0 = 50$  时的  $H_0: \pi = 0.10$  检验更令人怀疑。如果问题很关键,解释结果时最好有所保留,或者使用更精确的检验方法(见 Conover, 1999)。

### 练习

#### 基本技能

10.1 设想一来自于二项试验的样本,  $n = 150$ ,  $\hat{\pi} = 0.2$ 。

a. 该试验是否满足基于  $z$  的  $\pi$  置信区间对样本的要求? 给定同样的样本比例,那些样本容量值得怀疑?

b. 构造  $\pi$  的 90% 置信区间。

10.2 在什么条件下,能够用公式  $\hat{\pi} \pm z_{\alpha/2} \sigma_{\hat{\pi}}$  计算  $\pi$  的置信区间?

10.3 从二项分布总体中抽取容量为 1,500 的随机样本。如果有  $y = 1,200$  次成功。

a. 构造  $\pi$  的 95% 置信区间。

b. 构造  $\pi$  的 90% 置信区间。

10.4 参看练习 10.3。为什么在解释两置信区间时会有差异?

### 应用

10.5(社会) 专家预测今年 12 个铰接车零件中有 1 个与事故有关。其中一个理由就是 3 个铰接车零件中有 1 个有紧急危险机械问题(imminently hazardous mechanical condition),这很可能与机动车上的刹车系统有关。随机抽取 50 个铰接车零件,经过评估站测试,其中 19 个零件存在潜在的严重刹车系统问题。

a. 二项分布假定成立吗?

b. 能否用正态分布近似二项分布来得到  $\pi$  的置信区间?

c. 用这些数据给出  $\pi$  的 95% 置信区间。该区间有用吗? 假定  $\hat{\pi}$  保持不变,如何减小置信区间的宽度?

**10.6(心理学)** 在自我疗法实践的研究中,随机抽取 1,230 名成年人。下表列出了进行自我疗法的成年人的身体状况。针对每一状况,用 95% 置信区间对这部分调查作一小结。

| 身体状况      | 家庭治疗      | 响应百分比 |
|-----------|-----------|-------|
| 喉咙疼——未感冒  | 盐水或碳酸氢钠漱口 | 30    |
| 灼伤——非阳光灼伤 | 冷水或黄油     | 28    |
| 嗜酒        | 自酿饮料      | 25    |
| 超重        | 节食        | 22    |
| 因外伤疼痛     | 热敷或冷敷     | 21    |

**10.7(心理学)** 在练习 10.6 讨论的抽样调查中,441 名成年人报告最近他们有过咳嗽或感冒,260 名被调查者说他们使用过某一种直接面对顾客的药物(OTC, over-the-counter)进行治疗。数据如下。

|                    |     |
|--------------------|-----|
| 报告有过问题的人数          | 441 |
| 用过某种 OTC 治疗的病人数    | 260 |
| 用过特定种类 OTC 治疗的病人数: |     |
| 成人止疼药              | 110 |
| 成人感冒胶囊/片           | 57  |
| 咳嗽药                | 44  |
| 过敏/花粉热药物           | 9   |
| 液体冷冻剂              | 35  |
| 喷雾或吸入器             | 4   |
| 儿童止疼药              | 22  |
| 止咳药片               | 13  |
| 润喉糖/口香糖            | 9   |
| 儿童感冒胶囊/片           | 13  |
| 滴鼻药水               | 9   |
| 胸部按摩/药膏            | 9   |
| 麻醉喉咙糖              | 4   |
| 室内加湿器              | 4   |
| 其他产品               | 4   |

- a. 你如何整理和分析这些数据？可以用百分比吗？百分比相加等于 100% 吗？为什么？
- b. 根据这些数据，你能够用  $\pi$  的 95% 置信区间分析那些种类的 OTC？

**10.8(医学)** 许多 40 岁以上的人养成了对牛奶和奶制品的不耐性。奶制品厂开发了一系列不含乳糖的产品，该产品对这些人来说是可接受的。为了评价这些产品潜在的市场，奶制品厂委托一市场调研机构对其销售区域 40 岁以上的成年人进行调查。随机调查 250 名显示其中有 86 名遭受牛奶不耐性的痛苦。根据这些样本数据，计算牛奶不耐性人总体比例的 90% 置信区间，并解释之。

**10.9(社会学)** 在去年 4 月 15 日前不久，一队社会学家进行了一项调查，旨在研究他们的理论，即骗税者通常抱有某种想法以减轻他们的负罪感。调查队访问了 500 名成年人，要求他们回答，在什么情况下，他们认为所得税纳税申报表的欺骗行为是正当的。应答情况包括：

56% 同意“其他人不报告他们自己的收入。”

50% 同意“政府通常并不在意税金。”

46% 同意“如果一个人通常是遵纪守法，那么，骗税并不是一件严重的事。”

假定数据是来自于纳税人(或非纳税人)总体的一个简单随机样本，分别计算同意每一观点的总体比例的 95% 置信区间。

**10.10(民意调查学)** 国家级专栏作家最近报告了关于婚姻家庭的抽样调查结果，下面对专栏的一部分作一解释。

### 婚姻的构成要素

盖洛普提供给被访问者一个单子，列有众所周知的婚姻构成要素。比如在美国，像信仰相互尊重和相互理解被排在第一位。排在其次的有足够的钱，同样的背景，好的居住条件以及政治上的一致性。75% 的被访者投票给“和谐的性生活”，59% 给孩子，52% 给共同兴趣，48% 给“远离姻亲生活”，43% 给“分担家务”。(在西德，形成对照的是，只有 52% 投票给和谐的性生活，只有 19% 给分担家务。)

- a. 你如何用图或表展示调查结果？
- b. 你如何利用置信区间传达更多关于“对”的百分比的信息，以表达对各种婚姻构成要素的观点？为什么？
- c. 你对调查的报告方式还有什么疑虑？

**10.11(教育)** 据由国家科学基金会和宾夕法尼亚州立大学共同组织的国家科技扫盲会议上的专家指出，美国人口中有相当大一部分人是“科技文盲”。在这次会议上，一份对全国范围内的 2,000 名成年人的调查结果显示：

- 70% 不知道放射物。
- 40% 认为发射太空火箭会影响天气，且不明飞行物确实是来自其他星球的访问者。

- 80% 以上并不知道电话是如何工作的。
- 75% 并不完全了解计算机软件。
- 72% 不知道国民生产总值。

- a. 你如何用图或表展示这些数据？并画出这些图或表。
- b. 许多报纸上报道调查结果的文章在给出结论的同时并没有为读者提供详细的细节以便读者对数据进行评价，得出不同的结论。对于你为了得出自己的结论，这里遗漏了哪些细节？

**10.12(商业)** 某五金批发商的销售经理发现以前 500 次给五金店业主的电话中有 229 次导致了五金店内新的产品布局。假定 500 次电话表示一随机样本，求从长期来看，新的产品布局次数比例的 95% 置信区间。

**10.13** 就练习 10.12 所求出的置信区间，给出详细的口头解释。

**10.14(商业)** 作为市场调研的一部分，在 125 人的样本中，有 84 人知道某一产品。求总体中知道该产品的人所占比例的 95% 置信区间，并解释之。

**10.15** 练习 10.14 中的置信区间所依赖的正态近似充分吗？

**10.16(社会)** 在某大公司 40 位中层经理的样本中，我们发现其中有 8 位积极参加当地的市民组织或慈善组织。求所有中层经理中积极参加这类组织的经理所占比例的 90% 置信区间，并解释之。

**10.17(社会)** 和 1978 年相比，现在越来越多的人在外面用餐——根据全国性抽样调查的说法。下表列出了一些调查数据。

- a. 如果这些数据是建立在 1978 年和现在 1,500 名成年人的随机样本上，那么，对于每餐，我们能得出什么结论？二项分布的正态近似在这里有效吗？
- b. 根据这些数据，我们能得出现在有更多的人在外面用餐吗？为什么？

| 外面就餐 | 1978 | 现在  |
|------|------|-----|
| 早餐   | 3%   | 5%  |
| 午餐   | 18%  | 20% |
| 晚餐   | 16%  | 16% |

**10.18(医学)** 良性黏膜囊肿是上颌骨处一对瘘管上最常见的病变。在 800 名男性中，有 35 人被观察出患有良性黏膜囊肿。

- a. 在实施关于零假设  $H_0: \pi \geq 0.096$  (在以前对男性进行的研究中，这是最高的发病率) 的统计检验中，用正态近似合适吗？试解释之。
- b. 对研究假设  $H_a: \pi < 0.096$  作统计检验，给定  $\alpha = 0.05$ 。

**10.19(民意调查学)** 全国性民意调查要求随机访问至少 1,500 人，了解他们

对一个或多个问题的观点。这些访问通常都是当面进行的,因为使用邮寄方式的问卷回收率较低,而电话访问往往访问到的都是年纪较大的人,从而结果有所偏离。假定随机抽取 1,500 人,得到成年人中同意最近提出的能源保护建议的人所占的比例。

a. 如果 560 人赞成当局公布的政策,试用 95% 置信区间估计  $\pi$ ,即持赞成观点的成年人的比例。置信区间宽度的一半为多少?

b. 必须调查多少人才能使得  $\pi$  的 95% 置信区间宽度的一半等于 0.01?

**10.20(环境)** 为了估计某湖中汞含量超过 9(ppb)个单位的小龙虾所占的比例,研究人员从湖中随机捞取 20 只各种尺寸的小龙虾。在这一样本中,有 8 只小龙虾汞含量超过 9 个单位。利用这些数据估计  $\pi$  的 95% 置信区间,并解释之,其中  $\pi$  为该湖中汞含量超过 9 个单位的小龙虾所占的比例。

**10.21** 用计算机程序模拟  $n = 20, \pi = 0.4$  的二项分布。为此,需要获得  $y$  值,其中  $y$  为当样本来自于  $\pi = 0.4$  的二项分布时,20 次试验中成功的次数。再重复该模拟 39 次,得到总共 40 个模拟样本。

a. 用频率直方图将样本数据( $y$  值)作图。

b. 计算样本均值和样本标准差。将你的答案与  $y$  的真正均值和标准差进行比较。(提示:为了说明,这里给出了一个 Minitab 程序。)

40 BINOMIAL EXPERIMENTS WITH  $n = 20$  AND  $P(\text{SUCCESS}) = 0.4$

MTB>NAME C1 VALUE

MTB>RANDOM 40 C1

SUBC>BINOMIAL 20 0.4

MTB>TALLY C1

| VALUE | COUNT |
|-------|-------|
| 4     | 3     |
| 5     | 3     |
| 6     | 5     |
| 7     | 6     |
| 8     | 6     |
| 9     | 9     |
| 10    | 4     |
| 12    | 1     |



|     |    |
|-----|----|
| 13  | 2  |
| 16  | 1  |
| N = | 40 |

MTB> HISTOGRAM C1

Histogram of VALUE N = 40

| Midpoint | Count               |
|----------|---------------------|
| 4        | 3 * * *             |
| 5        | 3 * * *             |
| 6        | 5 * * * * *         |
| 7        | 6 * * * * * *       |
| 8        | 6 * * * * * *       |
| 9        | 9 * * * * * * * * * |
| 10       | 4 * * * *           |
| 11       | 0                   |
| 12       | 1 *                 |
| 13       | 2 * *               |
| 14       | 0                   |
| 15       | 0                   |
| 16       | 1 *                 |

MTB> MEAN C1

MEAN = 6.0500

MTB> STANDARD DEVIATION C1

ST. DEV = 2.5615

MTB> STOP

**10.22** 参看例 10.3。假定由最近的调查得  $\hat{\pi} = 0.15$ 。利用这一猜测出来的值代入样本容量公式进行计算。试解释你的答案和例 10.3 的答案之间的差异。

**10.23(工程)** 某制药公司一直在探究给医院职工提供一次性小药水瓶的可能性,这种药水瓶可以被用来作很多标准的实验室分析。对于某种分析,如血糖分析,技术人员往相应的小药水瓶里注入一定量的血液,当和事先放在瓶中的化学物

质充分混合时,观察其颜色。通过将混合液体的光密度(optical density)和色码图表比较,技术人员就能读出病人的血糖水平。显然,系统必须被严格控制,确保小瓶在被运到医院实验室前,将适当量的化学物质正确地密封在瓶中。从好几千瓶的一批产品中随机抽取 100 瓶,发现所有这 100 瓶中化学物质的剂量正确,且被正确地密封。

- a. 估计该批小药水瓶中次品的比例。
- b. 构造该批小药水瓶中次品比例的 99% 置信区间。
- c. 制药公司声称每批小药水瓶中次品比例不到 1%。利用(b)中的置信区间检验制药公司的说法。检验的置信水平是多少?

### 10.3 两总体比例之差 $\pi_1 - \pi_2$ 的推断

很多实际问题都牵涉到两个二项分布参数的比较。社会学家也许希望比较代表不同社会背景的两个社团中接受产前服务的妇女所占的比例。销售主管希望比较公众对其新近推出的新产品和对其竞争厂家产品的认知程度。

为了进行这种比较,我们假定独立地从两个二项分布总体中共抽取两个样本,且二项分布的参数未知,分别记为  $\pi_1$  和  $\pi_2$ 。如果由来自于总体 1 容量为  $n_1$  的随机样本观测到  $y_1$  次成功,来自于总体 2 容量为  $n_2$  的随机样本得到  $y_2$  次成功,那么,  $\pi_1$  和  $\pi_2$  的点估计分别为观测到的样本比例  $\hat{\pi}_1$  和  $\hat{\pi}_2$ 。

$$\hat{\pi}_1 = \frac{y_1}{n_1} \text{ 和 } \hat{\pi}_2 = \frac{y_2}{n_2}$$

将这种记号小结如下。

|      | 总 体                             |                                 |
|------|---------------------------------|---------------------------------|
|      | 1                               | 2                               |
| 总体比例 | $\pi_1$                         | $\pi_2$                         |
| 样本大小 | $n_1$                           | $n_2$                           |
| 成功次数 | $y_1$                           | $y_2$                           |
| 样本比例 | $\hat{\pi}_1 = \frac{y_1}{n_1}$ | $\hat{\pi}_2 = \frac{y_2}{n_2}$ |

关于两个二项分布比例的推断通常用他们的差  $\pi_1 - \pi_2$  来表示,且用样本比例的差  $\hat{\pi}_1 - \hat{\pi}_2$  进行区间估计和统计检验。 $\hat{\pi}_1 - \hat{\pi}_2$  的抽样分布近似于正态分布,其均值和标准差为

$$\mu_{\hat{\pi}_1 - \hat{\pi}_2} = \pi_1 - \pi_2$$

$$\sigma_{\hat{\pi}_1 - \hat{\pi}_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

第四章介绍了二项分布正态近似的应用条件,如果我们将同样的条件应用到上述两个二项分布,那么,这里的近似是合适的。因此,如果  $n\pi_i$  和  $n(1-\pi_i)$  都至少为 5,  $i=1,2$ , 那么,就可以用正态分布近似  $\hat{\pi}_1 - \hat{\pi}_2$  的分布。由于  $\pi_1$  和  $\pi_2$  未知,因此,这种近似的有效性通常用  $n\hat{\pi}_i$  和  $n(1-\hat{\pi}_i)$ ,  $i=1,2$  来衡量。

$\pi_1 - \pi_2$  的置信区间和统计检验比较简单,遵循  $\mu_1 - \mu_2$  的估计和检验方法。下面给出区间估计的形式;它取通常的形式,即点估计  $\pm z$ (标准差)。

$\pi_1 - \pi_2$  的  $100(1-\alpha)\%$  的置信区间

$$\hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} \sigma_{\hat{\pi}_1 - \hat{\pi}_2}$$

其中

$$\sigma_{\hat{\pi}_1 - \hat{\pi}_2} = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

注:在  $\sigma_{\hat{\pi}_1 - \hat{\pi}_2}$  公式中用  $\hat{\pi}_1$  和  $\hat{\pi}_2$  代替  $\pi_1$  和  $\pi_2$ 。当  $\hat{\pi}_1 - \hat{\pi}_2$  的正态近似有效时,这种代替所产生的误差很小。

### 例 10.5

某公司在 Michigan 州的 Grand Rapids 和 Kansas 州的 Wichita 城市地区试销一种新产品。公司在 Grand Rapids 地区几乎完全通过电视发布广告。在 Wichita 地区,公司在电视,广播,报纸和杂志上的广告投入大致相同。广告战开始的两个月后,公司进行抽样调查,以了解消费者对其产品的认知程度。求两地区知道该产品的消费者所占比例差的 95% 置信区间。

|          | Grand Rapids | Wichita |
|----------|--------------|---------|
| 调查的人数    | 608          | 527     |
| 知道该产品的人数 | 392          | 413     |

**解答** 样本中 Wichita 地区的消费者对该产品的认知程度更高,所以令 Wichita 地区为地区 1。

$$\hat{\pi}_1 = 413/527 = 0.784 \quad \hat{\pi}_2 = 392/608 = 0.645$$

估计的标准差为

$$\sqrt{\frac{(0.784)(0.216)}{527} + \frac{(0.645)(0.355)}{608}} = 0.0264$$

因此, 95% 置信区间为

$$\begin{aligned}(0.784 - 0.645) - 1.96(0.0264) &\leq \pi_1 - \pi_2 \\ &\leq (0.784 - 0.645) + 1.96(0.0264)\end{aligned}$$

即

$$0.087 \leq \pi_1 - \pi_2 \leq 0.191$$

这表明 Wichita 地区知道该产品的消费者比例比 Grand Rapids 地区知道该产品的消费者比例高 8.7% 到 19.1%。

该置信区间的构造是建立在二项分布的正态近似基础上。在第四章, 我们曾经指出, 作为一般规则, 为了应用正态近似,  $n\hat{\pi}$  和  $n(1 - \hat{\pi})$  至少应等于 5。本例中计算置信区间时, 每一样本都满足这一规则。实际中, 那些非常接近于违反这一规则的样本容量并不是很有用, 因为他们将导致太宽的置信区间。例如, 对于  $n_1 = 30$ ,  $\hat{\pi}_1 = 0.20$  和  $n_2 = 60$ ,  $\hat{\pi}_2 = 0.10$  的两个样本, 即使  $n\hat{\pi}$  和  $n(1 - \hat{\pi})$  都大于 5, 但是置信区间为  $-0.06 \leq \pi_1 - \pi_2 \leq 0.26$ ,  $\pi_1$  可能是任何一个值, 从比  $\pi_2$  低 6 个百分点到比  $\pi_2$  高 26 个百分点。

置信区间太宽以至于失去作用的原因在于每次观测只传达了很少量的信息。实际上每次观测仅传达了一个“比特”: 代表成功的一个 1 或代表失败的一个 0。例如, 公司首席执行官对赔偿问题进行调查时, 通常以年为单位列出经理的年龄。如果我们用“55 岁以上”对“55 岁以下”这样的类型数据代替真实年龄时, 我们获得的信息绝对非常有限。当每一项目只包含一点点信息时, 我们就需要大量的项目来获得足够的总量信息。无论在哪儿, 只要有可能, 最好用计量的方式而不仅仅用类型的方式来度量结果。当计量的方式不可能时, 就应该用容量相对大的样本。

两总体比例差的假设检验是建立在正态近似的  $z$  统计量基础之上。虽然零假设可以是  $\pi_1 - \pi_2$  等于某一已知的值, 但典型的零假设一般认为总体比例没有差异。该检验方法非常类似于均值差的  $t$  检验, 总结如下:

### 两总体比例之差的统计检验

$$H_0: \quad 1. \quad \pi_1 - \pi_2 \leq 0$$

$$2. \quad \pi_1 - \pi_2 \geq 0$$

$$3. \quad \pi_1 - \pi_2 = 0$$

$$H_a: \quad 1. \quad \pi_1 - \pi_2 > 0$$

$$2. \quad \pi_1 - \pi_2 < 0$$

$$3. \quad \pi_1 - \pi_2 \neq 0$$

$$\text{T.S.: } z = \frac{(\hat{\pi}_1 - \hat{\pi}_2)}{\sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}}$$

- R.R.: 1.  $z > z_\alpha$   
 2.  $z < -z_\alpha$   
 3.  $|z| > z_{\alpha/2}$

验证假定并作出结论。

注:只有当  $n_1\hat{\pi}_1, n_1(1-\hat{\pi}_1), n_2\hat{\pi}_2$  和  $n_2(1-\hat{\pi}_2)$  都大于 5 时,该检验才能被使用。

### 例 10.6

教育研究人员欲比较两种针对非英语人员进行英语教学的效果,一种是借助于计算机软件的教学方式,另一种是传统的课堂教学方式。研究人员随机地将 300 人分成两个班,一班 125 人采用计算机教学方式,另一班 175 人采用传统教学方式。在 6 个月的培训期结束之时,对所有 300 名学生进行一次测验,结果列于下表。

| 考试结果 | 计算机教学 | 传统教学 |
|------|-------|------|
| 通过   | 94    | 113  |
| 失败   | 31    | 62   |
| 总和   | 125   | 175  |

和传统教学方式相比,计算机教学方式提高了测验的通过率吗? 给定  $\alpha = 0.05$ 。

**解答**  $\pi_1$  表示采用计算机教学的测验通过率,  $\pi_2$  表示采用传统教学方式的测验通过率。我们检验假设

$$H_0: \pi_1 - \pi_2 \leq 0$$

$$H_a: \pi_1 - \pi_2 > 0$$

如果检验统计量  $z$  大于  $z_{0.05} = 1.645$ , 则拒绝  $H_0$ 。根据数据得估计量的观察值

$$\hat{\pi}_1 = \frac{94}{125} = 0.752 \quad \hat{\pi}_2 = \frac{113}{175} = 0.646$$

从而得检验统计量

$$\begin{aligned}
 z &= \frac{(\hat{\pi}_1 - \hat{\pi}_2)}{\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}} \\
 &= \frac{0.752 - 0.646}{\sqrt{\frac{0.752(1-0.752)}{125} + \frac{0.646(1-0.646)}{175}}} = 2.00
 \end{aligned}$$

由于  $z = 2.00$  大于 1.645, 因此, 我们拒绝  $H_0$ , 并得出结论: 观察支持备择假设, 即计算机教学比传统教学具有更高的通过率。查标准正态分布表, 得观察数据的  $p$  值为:  $p \text{ 值} = P(z \geq 2.00) = 0.0228$ 。有效性差异  $\pi_1 - \pi_2$  的 95% 置信区间为

$$0.752 - 0.646 \pm 1.96 \sqrt{\frac{0.752(1 - 0.752)}{125} + \frac{0.646(1 - 0.646)}{175}},$$

即  $0.106 \pm 0.104$

我们有 95% 的把握相信采用计算机教学学生的考试通过率比采用传统教学学生的考试通过率高出 0.2% 到 21%。为使我们的结论具有一定的合理性, 我们需要验证一下样本容量是否足够大。由于  $n_1 \hat{\pi}_1 = 94$ ,  $n_1(1 - \hat{\pi}_1) = 31$ ,  $n_2 \hat{\pi}_2 = 113$ ,  $n_2(1 - \hat{\pi}_2) = 62$ , 这四个数都大于 5。因此, 大样本准则满足。

### 练习

#### 基本技能

**10.24** 一容量  $n_1 = 1,000$  的随机样本来自  $\pi_1 = 0.4$  的二项分布总体, 另一与其独立的随机样本取自  $\pi_2 = 0.2$  的二项分布总体。正态近似成立吗? 试描述  $\hat{\pi}_1 - \hat{\pi}_2$  的抽样分布。

**10.25** 参看练习 10.24。从每个总体中选取多大容量的样本使得 90% 置信区间形如  $\hat{\pi}_1 - \hat{\pi}_2 \pm 0.01$ ? (提示: 假定取自两总体的样本容量相同, 解方程  $z_{\alpha/2} \sigma_{\hat{\pi}_1 - \hat{\pi}_2} = 0.01$  得共同的样本容量  $n$ , 根据练习 10.24, 利用  $\hat{\pi}_1 = 0.4$ ,  $\hat{\pi}_2 = 0.2$ )

### 应用

**10.26(民意调查学)** 一法律学生认为共和党人中赞成附加税激励的比例高于民主党人中赞成附加税激励的比例。该学生获得独立的两随机样本, 一个是 200 名共和党人, 另一个是 200 名民主党人, 发现 109 名共和党人和 86 名民主党人赞成附加税激励。试根据这些数据检验  $H_0: \pi_1 - \pi_2 \leq 0$   $H_a: \pi_1 - \pi_2 > 0$ 。给出检验的显著性水平。

**10.27(商业)** 某计算机零售商试图在两种售后服务方式之间进行选择。第一种方法强调预防性维护, 第二种强调对问题的快速反应。该零售商选取一些顾客作为样本, 其中对一部分顾客用第一种方法, 对另一部分顾客用第二种方法。六个月之后, 发现用第一种方法服务的 200 名顾客中有 171 名对服务满意, 用第二种方法服务的 200 名顾客中有 153 名对服务满意。基于 1,0 数据, 用软件 Execustat 得到的输出结果如下。

# TWO SAMPLE ANALYSIS FOR SATISFIED BY METHOD

|             |          | 1        | 2                |
|-------------|----------|----------|------------------|
| Sample size | 200      | 200      |                  |
| Mean        | 0.855    | 0.765    | diff. = 0.09     |
| Variance    | 0.124598 | 0.180678 | ratio = 0.689612 |

Std. deviation      0.352984      0.425063

95 % confidence intervals

mul - mu2:(0.0131926,0.166807)assuming equal variances

mul - mu2:(0.0131847,0.166815)not assuming equal variances

## HYPOTHESIS TEST-DIFFERENCE OF MEANS

Null hypothesis;difference of means = 0

Alternative;not equal

Equal variances assumed;no

Computed t statistic = 2.30362

P value = 0.0218

给定  $\alpha = 0.05$ , 试检验研究假设, 即总体比例不相等。仔细给出你的结论。

**10.28** 求练习 10.27 中两总体比例差的置信区间。证明根据该置信区间所得的结论与正式地检验研究假设所得的结论相同。

**10.29(商业)** 某广告社的媒体选择经理在两种杂志上为一委托银行采用同样的方式各插入一则广告。一个月以后进行市场调研, 发现 473 名第一种杂志的读者中有 226 名知道广告中提及的银行业务, 而 439 名第二种杂志的读者中有 165 名(剔除同是两种杂志的读者)。下面的 Minitab 输出结果是建立在 1,0 数据基础上的。

MTB> TWOT 'AWARE?' 'MAGAZINE'

Two

Sample T-Test and Confidence Interval

Twosample T for Aware?

| Magazine | N   | Mean  | StDev | SE Mean |
|----------|-----|-------|-------|---------|
| 1        | 473 | 0.478 | 0.500 | 0.023   |

2                      439                      0.376                      0.485                      0.023

95 % C. I. for  $\mu_1 - \mu_2$ : (0.038, 0.166)

T-Test  $\mu_1 = \mu_2$  (vs not = ): T = 3.13 P = 0.0018 DF = 908

a. 通过手工计算, 求两批读者中知道银行业务的读者所占比例之差的 95% 置信区间。并将该区间和 Minitab 输出的区间进行比较。

b. 样本容量是否足够大, 以至于可以使用正态近似?

c. 给定  $\alpha = 0.05$ , 置信区间是否显示两种比例在统计上具有显著差异?

**10.30** 利用练习 10.29 的输出结果, 给定  $\alpha = 0.05$ , 试正式地检验零假设, 即总体比例相同。

**10.31** 为了检验可靠性, 我们对两批各 30 部点阵式打印机的电动机进行严格的测试, 他们分别来自于不同的供应商。供应商 1 的电动机中有 22 部通过测试, 供应商 2 的电动机中只有 16 部通过测试。

a. 证明在  $\alpha = 0.05$  的水平下, 通过率差异在统计上并不显著(双侧)。

b. 我们可否声称我们已经证明了两供应商提供的电动机具有相同的可靠性?

**10.32** 利用练习 10.31 中的数据计算比例差的 95% 置信区间。根据两供应商的相对可靠性, 仔细解释所得结果。

**10.33(生物学)** 为了比较两种老鼠身上肿瘤的发生率, 从每种老鼠中各随机抽取 100 只(50 只公的, 50 只母的), 进行为期 1 年的观察。所有的老鼠几乎年龄相同, 并且在同等条件下饲养。利用 1 年期的样本数据, 构造两种老鼠中显现肿瘤可能性的老鼠所占比例之差的 95% 置信区间。

|             | 种类 A | 种类 B |
|-------------|------|------|
| 样本容量        | 100  | 100  |
| 表现具有潜在肿瘤的数目 | 25   | 15   |

**10.34(医学)** 有一种治疗男性斑秃的药物——至少那是数百万男性的希望, 因为 FDA 批准了 Upjohn 公司的 minoxidil 可以用于治疗这种病。在 27 个大的研究中心对 minoxidil 进行研究, 这些中心的病人被随机分配, 要么得到 minoxidil 药物, 要么得到一片外表相同的安慰剂。不考虑中心之间的差别, 假定初步结果如下:

|           | 样本容量 | 生出新发的人的百分比 |
|-----------|------|------------|
| minoxidil | 310  | 32         |
| 安慰剂       | 309  | 20         |



- a. 利用这些数据检验  $H_0: \pi_1 - \pi_2 = 0$   $H_a: \pi_1 - \pi_2 \neq 0$ , 并给出检验的  $p$ -值。
- b. 如果你正在为 FDA 工作, 那么, 在这一研究中, 你另外还需要调查什么信息?

**10.35(生物学)** 可卡因比海洛因更致命吗? *Journal of the American Medical Association* 杂志报道的一项研究发现任意服用可卡因的老鼠比任意服用海洛因的老鼠身体更差, 行为更紊乱, 死亡率更高。研究中 30 天以后的死亡率如下:

|      | 30 天内死亡的百分比 |
|------|-------------|
| 可卡因组 | 90          |
| 海洛因组 | 36          |

- a. 假定每组 100 只老鼠。试检验  $H_0: \pi_1 - \pi_2 \leq 0$   $H_a: \pi_1 - \pi_2 > 0$ , 并给出检验的  $p$  值。
- b. 对使用这两种药的人来说, 上述结果意味着什么?

## 10.4 多比例的推断: 卡方拟合优度检验

我们可以将第四章二项抽样方案推广到每次试验有  $k$  ( $k > 2$ ) 个可能结果的情形。例如, 根据政党(共和党, 民主党, 社会党, 绿党, 独立党等)将登记的选民进行分类, 或者根据病人身体条件的改善程度(重大改善, 改善, 没有变化, 更坏)对临床试验的病人进行评价。这种类型的试验或研究被称之为多项试验。下面列出其特征。

### 多项试验

1. 共做  $n$  次相同的试验。
2. 每次试验有  $k$  个可能结果, 且只有一个发生。
3. 单个试验将导致第  $i$  个结果发生的概率为  $\pi_i$ ,  $i = 1, 2, \dots, k$ , 且对于不同的试验,  $\pi_i$  保持不变。(注:  $\sum \pi_i = 1$ )。
4. 试验是独立的。
5. 我们所感兴趣的是  $n_i$ , 即导致第  $i$  个结果发生的试验次数。(注:  $\sum n_i = n$ )。

$n$  次试验中观测到每一试验结果的观测次数的概率分布称为**多项分布**, 给出如下。

$$P(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!} \pi_1^{n_1} \pi_2^{n_2} \dots \pi_k^{n_k}$$

在第四章中,我们讨论了二项概率分布,且知道

$$n! = n(n-1)\dots 1, 0! = 1$$

我们能够用多项分布计算具体事件的概率。

### 例 10.7

以前对某群牛进行的繁殖试验表明一次交配能够产下一头健康的小牛的概率为 0.83。同样地,能够产下 0 头或 2 头健康小牛的的概率分别为 0.15 和 0.02。一农场主饲养了 3 头这种母牛;求恰好得到 3 头健康小牛的的概率。

**解答** 假定 3 头母牛是随机选择的,则可以看作是  $n = 3, k = 3$  的多项式试验。其可能的结果和相应的概率列于下表。

| 结果编号 | 小牛数 | 概率 $\pi_i$ |
|------|-----|------------|
| 1    | 0   | 0.15       |
| 2    | 1   | 0.83       |
| 3    | 2   | 0.02       |

易见,结果 1, 2, 3 分别表示一头母牛产下 0, 1, 2 头健康小牛的事件。同样地,  $n_1, n_2, n_3$  分别表示产下 0, 1, 2 头健康小牛的母牛数量。为了恰好得到 3 头小牛,我们必须能够观测到下列可能的事件之一。

$$A: \begin{cases} 1 \text{ 头母牛没有产下健康的小牛: } n_1 = 1 \\ 1 \text{ 头母牛产下 1 头健康的小牛: } n_2 = 1 \\ 1 \text{ 头母牛产下 2 头健康的小牛: } n_3 = 1 \end{cases}$$

$$B: 3 \text{ 头母牛各产下 1 头小牛: } \begin{cases} n_1 = 0 \\ n_2 = 3 \\ n_3 = 0 \end{cases}$$

由于  $n = 3, k = 3$ , 于是对于 A 事件

$$P(n_1 = 1, n_2 = 1, n_3 = 1) = \frac{3!}{1!1!1!} (0.15)^1 (0.83)^1 (0.02)^1 \cong 0.015$$

同样地,对于 B 事件

$$P(n_1 = 0, n_2 = 3, n_3 = 0) = \frac{3!}{0!3!0!} (0.15)^0 (0.83)^3 (0.02)^0$$

$$= (0.83)^3 \cong 0.572$$

因此, 3 头母牛共恰好产下 3 头小牛的概率就是 A 事件与 B 事件概率之和; 即  $0.015 + 0.572 \approx 0.59$ 。

在多项分布中, 我们的主要兴趣还是关于概率  $\pi_1, \pi_2, \dots, \pi_k$  的统计检验所依赖的概率模型。我们将假定这些  $\pi$  等于某些给定的值, 然后, 考察样本数据是否支持这些假定的值。检验这一假设的其中一种方法是考察观测到的导致每一试验结果的试验次数, 并将其和期望得到的次数进行比较。例如, 在我们前面的例子中, 我们给出了对应于 0, 1, 2 头小牛的概率分别为 0.15, 0.83, 0.02。如果我们考察由 100 头交配过的母牛组成的随机样本, 那么, 我们期望将观察到 15 头母牛没有产下健康小牛。同样地, 我们也期望将观察到 83 头母牛各产下 1 头健康小牛, 2 头母牛各产下 2 头健康小牛。

**定义 10.1** 在每一次试验有  $k$  个可能结果的多项试验中,  $n$  次试验中第  $i$  个结果的期望数是  $n\pi_i$ , 其中  $\pi_i$  是单个试验中第  $i$  个结果出现的概率。

1900 年, Karl Pearson 提出了下面的检验统计量, 用以检验某些给定的概率:

$$\chi^2 = \sum_i \left[ \frac{(n_i - E_i)^2}{E_i} \right]$$

其中  $n_i$  表示具体试验中第  $i$  个结果出现的次数,  $E_i$  表示当假设的概率就是每一结果在每次试验中出现的概率时, 理论上所期望的第  $i$  个结果出现的次数。通常我们称概率  $\pi_1, \pi_2, \dots, \pi_k$  为**单元概率 (cell probability)**, 一个单元对应于  $k$  个结果中的一个结果。称对应于  $k$  个结果的观察数  $n_1, n_2, \dots, n_k$  为**观察单元数 (observed cell counts)**, 称  $E_1, E_2, \dots, E_k$  为**期望单元数 (expected cell counts)**。

假设单元概率为  $\pi_1, \pi_2, \dots, \pi_k$ 。根据定义 10.1, 我们就能计算期望的单元数, 从而考察实际观测到的数据和期望的观测数据是否拟合得很好。当然, 如果那些假设的  $\pi$  值是正确的, 那么, 观察单元数  $n_i$  不应该偏离期望单元数  $E_i$  很多, 从而  $\chi^2$  的计算值应较小。类似地, 当一个或多个假设的单元概率不正确时, 观察单元数和期望单元数将有很大差别, 从而使得  $\chi^2$  值变大。

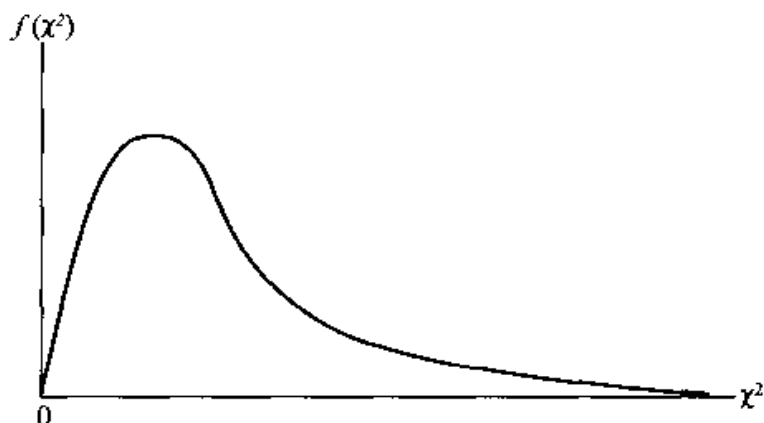
如果期望单元数  $E_i$  充分大, 那么  $\chi^2$  近似服从卡方分布。

基于  $k$  个已知单元概率的卡方拟合优度检验具有  $k-1$  个自由度。本节最后, 我们将解释为什么有  $k-1$  个自由度。查附表 7 可得到检验统计量  $\chi^2 = \sum_i \left[ \frac{(n_i - E_i)^2}{E_i} \right]$  的上分位数。图 10.1 为  $df=4$  的卡方分布。

我们将概括  $k$  个已知单元概率的卡方拟合优度检验如下。

#### 卡方拟合优度检验

零假设:  $\pi_i = \pi_{i0}, i = 1, 2, \dots, k$ , 其中  $\pi_{i0}$  为已知概率或比例。

图 10.1  $df=4$  的卡方分布

备择假设:至少有一个单元概率不等于假设的概率值。

检验统计量:  $\chi^2 = \sum_i \left[ \frac{(n_i - E_i)^2}{E_i} \right]$ , 其中  $n_i$  为类  $i$  的观察数,  $E_i = n\pi_{i0}$  是  $H_0$  成立条件下期望的观察数。

拒绝域:如果  $\chi^2$  超过对应于  $\alpha = \alpha$ ,  $df = k - 1$  的临界值。那么,拒绝  $H_0$ 。

检验假定并作出结论。

当样本容量  $n$  变大时,卡方拟合优度检验统计量的分布将越来越接近于卡方分布。其近似程度将依赖于样本容量  $n$  和单元数  $k$ 。Cochran (1954)指出,如果没有  $E_i$  小于 1,且只有 20% 的  $E_i$  小于 5,那么近似度应该是足够的。随着  $k$  增大,假定卡方拟合优度检验统计量充分近似服从于卡方分布,那么  $n/k$  趋向于减小。Agresti(1990)讨论了这样一种情形,即使期望单元数大小适中,但观察单元数较小时,卡方近似效果也是比较差的。Agresti 认为,关于如何确定合适的样本容量,我们不可能制定一个单一的规则使之适用于一切情况。然而,我们推荐使用 Cochran 的方法来考察卡方拟合优度检验统计量是否充分近似于卡方分布。当某些  $E_i$  太小时,还有其他好几种方法可供选择。可以合并类型变量的水平,从而增加观察单元数。然而,除非可以自然地重新定义类型变量,且类型变量不改变被检验假设的自然属性,否则,就不应该合并。当不可能获得充分大的观察单元数以满足卡方近似时,Agresti(1990)讨论了检验假设的精确方法。

### 例 10.8

某实验室将一种试验药品和一种有助于高血压病人稳定血压的标准药品进行比较。为此,在很多地点作了很多次临床试验,他们让可比较的高血压(根据纽约

心脏协会(NYHA)分类标准)病人服用该标准药品。实验室将这些病人对治疗的反应分成四类。表 10.1 列出了各个类以及每一类中服用标准药品病人所占的比例。

表 10.1 使用标准药品的临床结果

| 类别       | 百分比 |
|----------|-----|
| 血压明显下降   | 50  |
| 血压中等程度下降 | 25  |
| 血压略微下降   | 10  |
| 平稳或略微升高  | 15  |

实验室随机抽取 200 位高血压病人进行临床试验。正如对标准药品进行研究一样,根据 NYHA 分类标准,将所有病人进行同样的分类。利用表 10.2 检验假设:与试验药品相关的单元概率等于与标准药品相关的单元概率,给定  $\alpha = 0.05$ 。

表 10.2 样本数据

| 类别 | 观察单元数 |
|----|-------|
| 1  | 120   |
| 2  | 60    |
| 3  | 10    |
| 4  | 10    |

**解答** 该试验具有多项试验的特征,  $n = 200, k = 4$ 。

结果 1:服用过试验药品后,病人的血压明显下降。

结果 2:服用过试验药品后,病人的血压中等程度下降。

结果 3:服用过试验药品后,病人的血压略微下降。

结果 4:服用过试验药品后,病人的血压维持不变或略微上升。

零假设和备择假设为:

$$H_0: \pi_1 = 0.50, \pi_2 = 0.25, \pi_3 = 0.10, \pi_4 = 0.15$$

$H_a$ :至少有一个单元概率不等于其假设的值。

在计算检验统计量之前,我们必须求期望单元数。将这些数据列于表 10.3。

表 10.3 观察单元数和期望单元数

| 类别 | 观察单元数 $n_i$ | 期望单元数 $E_i$       |
|----|-------------|-------------------|
| 1  | 120         | $200(0.50) = 100$ |
| 2  | 60          | $200(0.25) = 50$  |
| 3  | 10          | $200(0.10) = 20$  |
| 4  | 10          | $200(0.15) = 30$  |

由于所有的期望单元数相对较大,因此,我们可以计算卡方统计量,并将其和卡方分布的临界值进行比较。

$$\begin{aligned}
 \chi^2 &= \sum_i \left[ \frac{(n_i - E_i)^2}{E_i} \right] \\
 &= \frac{(120 - 100)^2}{100} + \frac{(60 - 50)^2}{50} + \frac{(10 - 20)^2}{20} + \frac{(10 - 30)^2}{30} \\
 &= 4 + 2 + 5 + 13.33 = 24.33
 \end{aligned}$$

设定第一类错误概率  $\alpha = 0.05$ , 对于  $\alpha = 0.05$ ,  $df = k - 1 = 3$ , 我们查卡方分布的分位数。由附表 7 得临界值为 7.815。

R. R.: 若  $\chi^2 > 7.815$ , 则拒绝  $H_0$ 。

结论: 由于  $\chi^2$  的计算值大于 7.815, 因此我们拒绝零假设, 从而得出结论, 即至少有一个单元概率不等于  $H_0$  假设下的单元概率。实际上, 就试验药品来说, 有更高比例的病人落入血压中等程度下降和明显下降两类中。该检验的  $p$  值为  $p < 0.001$  (见附表 7)。

卡方拟合优度检验的前提条件和多项试验的一样, 关键是试验的独立性和单元概率的不变性。如果, 例如样本中几个病人来自同一家庭, 那么, 就违反了试验的独立性, 这是因为高血压病的遗传性很强。如果在研究期间, 医疗标准发生了变化, 采用其他医疗“标准”, 那么, 单元概率将不能保持不变。

卡方拟合优度检验的检验统计量是  $k$  项之和, 这就是为什么自由度依赖于类的个数  $k$ , 而不是总的样本容量  $n$ 。然而, 自由度为  $k - 1$ , 而不是  $k$ , 这是因为  $n_i - E_i$  之和一定等于  $n - n = 0$ ;  $k - 1$  观察单元数与期望单元数之差是自由变化的, 但是最后一个, 即第  $k$  个由条件  $n_i - E_i$  之和等于 0 所决定。

多年来, 拟合优度检验被广泛地应用于检验各种科学理论。然而, 和前面的统计检验不同, 这里, 我们感兴趣的是零假设, 而不是研究(备择)假设。不幸的是, 前面的统计检验背后的逻辑在这里并不成立。在前面所讨论的统计检验的通常情况下, 研究(备择)假设是科学家所感兴趣的; 我们建立一个适当的零假设, 收集数据, 通过拒绝  $H_0$  而支持  $H_a$ , 也就是说, 通过否定  $H_0$  而“证明” $H_a$  是正确的。

对于卡方拟合优度检验,我们并不能作同样的处理。如果某位科学家有一套理论,他希望样本数据符合他的理论,那么,他需要接受  $H_0$ 。根据我们前面的研究,若接受  $H_0$ ,则我们可能要犯第二类错误。这里,和其他检验一样,很难计算概率  $\beta$ 。一般来说,对于拟合优度检验,如果  $n$  较小,或者类的个数  $k$  较大,那么,犯第二类错误的可能性就很高。即使期望单元数  $E_i$  与我们推荐的值相一致,犯第二类错误的概率也可能很大。因此,对于卡方拟合优度检验的结果要审慎看待。即使  $H_0$  没有被拒绝,也不要盲目地去接受零假设,认为它就是事实。

### 练习

#### 基本技能

10.36 列出多项试验的特征。

10.37 二项试验与多项试验之间有什么联系?

10.38 对于多项试验,在什么条件下适合使用卡方拟合优度检验?如果样本数据并不拒绝零假设,你需要做出什么样的处理?

10.39 使用卡方拟合优度检验时,需要对样本容量施加什么样的限制条件?

### 应用

10.40 下表列出的是虚构的数据。利用这些数据作卡方拟合优度检验,零假设为  $H_0: \pi_1 = 0.2, \pi_2 = 0.15, \pi_3 = 0.40, \pi_4 = 0.15, \pi_5 = 0.10$ , 给定  $\alpha = 0.05$ 。数据与假设的概率一致吗?

| 类别 | 观察单元数 $n_i$ |
|----|-------------|
| 1  | 60          |
| 2  | 50          |
| 3  | 130         |
| 4  | 40          |
| 5  | 20          |
| 总和 | 300         |

10.41 同样给定  $\alpha = 0.05$ , 利用练习 10.40 的数据作卡方拟合优度检验,新的零假设为:  $H_0: \pi_1 = 0.15, \pi_2 = 0.20, \pi_3 = 0.45, \pi_4 = 0.15, \pi_5 = 0.05$ 。将你的结果和练习 10.40 的结果进行比较。该检验对  $H_0$  假设下单元概率的敏感性如何?如果你不能拒绝  $H_0$ , 那么,你能得出什么结论?

**10.42(商业)** 过去5年,某保险公司承接的保险单中40%为终身人寿保险,20%为普通人寿保险,25%为年度更新(ART)保险,15%为其他类型的保险。要改变保险单的这种结构,需要经过长期的努力,在佣金(commission)结构,储备金(reserve)和甚至可能需要在投资方面作出调整。从最近几个月所承接的保单中随机抽取1,000份保单,由样本数据算得结果如下。根据这些数据评价现在的比例和历史的比例是否发生了变化。试给出检验的 $p$ 值。如果有的话,哪一种保险更受欢迎?

| 类别  | 观察单元数 $n_i$ |
|-----|-------------|
| 终身  | 320         |
| 普通  | 280         |
| ART | 240         |
| 其他  | 160         |
| 总和  | 1,000       |

**10.43(社会)** 某大学和几家企业在其周边社区开展一项工作——学习计划。要求学生在3个月的见习期间和企业的社会工作者一起工作。来自该大学的同样多的学生分别被安排在化工厂,纺织厂和制药厂。根据他们见习的企业将完成计划的学生进行分类。下面的数据可以被认为是很多能够完成计划学生的一个随机样本。试检验零假设:在化工厂,纺织厂和制药厂见习并完成计划的学生各占1/3。给定 $\alpha = 0.05$ ,  $n_i$ 表示第 $i$ 组完成计划的学生人数。

| 组   | 观察单元数 $n_i$ |
|-----|-------------|
| 制药厂 | 20          |
| 化工厂 | 13          |
| 纺织厂 | 30          |

**10.44(社会)** 研究人员进行了一项试验以考察居住在一个县各社会阶层人员中精神病人的比例是否与该县社会阶层人员的分布相一致。将400位病人分类得观察单元数如下。

下层:215          中上层:60  
中下层:100        上层:25

利用这些数据检验零假设: $\pi_1 = 0.25$ ,  $\pi_2 = 0.48$ ,  $\pi_3 = 0.20$ ,  $\pi_4 = 0.07$ , 其中这些 $\pi$ 是假设的该县相应的社会阶层人员所占的比例。给定 $\alpha = 0.05$ , 试给出结论。



**10.45(民意调查学)** 在上一届总统选举时,某地区登记的选民中有 50% 为共和党,40% 为民主党,10% 为独立派人士。本次选举之前,随机抽取 200 名登记的选民,结果 90 名为共和党,80 名为民主党,30 名为独立派人士。检验研究假设,即登记的选民分布不同于上一届选举。给出检验的  $p$  值并做出结论。

**10.46(医学)** 一地方医生觉得普通感冒的发生具有一定的季节趋势。他估计每年 40% 病例发生在冬天,40% 在春天,10% 在夏天,10% 在秋天。该医生从去年普通感冒病例中随机抽取 1,000 例收集信息,基于样本数据提供的信息,你同意医生的估计吗? 给定  $\alpha = 0.05$  作统计检验并得出结论。

| 季节 | 频数  |
|----|-----|
| 冬天 | 374 |
| 春天 | 292 |
| 夏天 | 169 |
| 秋天 | 165 |

**10.47** 参看练习 10.46。如果医生声称各个季节感冒发生率没有差别,那么零假设是什么? 检验假设:普通感冒发生率没有季节趋势。给出检验的显著性水平。你对你的结论还有什么保留吗?

**10.48(医学)** 以前,关于减轻抑郁症药品的试验都是针对没有抑郁症迹象的正常成年人进行的。我们假定通过对正常人的研究得到一个大型的数据库,且为了实用,数据库能够提供正常人中对药品有不同反应的人数。参加试验的每一位成年人都要求对药品的疗效作出评价,是无效,中等有效还是有效。这些类中,被调查者的比例分别为 60%,30% 和 10%。在一项新的针对抑郁症人进行的研究中,随机抽取 85 位患有抑郁症的成年人,结果如下:

无效:30

中等有效:35

有效:20

是否有充分证据表明对药品有不同反应的患抑郁症成年人的比例不同于无抑郁症的成年人的比例。给出检验的显著性水平并做出结论。

**10.49(民意调查学)** 在一项随机调查中,调查者要求 40 位报纸编辑就最近法院判决将来可能给新闻自由带来压制谈谈他们的看法。将编辑的观点总结如下。利用这些数据检验零假设,即各个类的比例相同。给定  $\alpha = 0.05$ ,根据这些数据做出结论。如果有的话,你对你的结论还有什么保留吗?

| 压制程度 | 频数 |
|------|----|
| 没有   | 8  |
| 很少   | 8  |
| 中等   | 10 |
| 严重   | 14 |

**10.50(商业)** 某研究人员获得了由 125 位证券分析师组成的样本,他要求每位分析师从纽约股票交易所选取 4 只股票,并且期望这 4 只股票比标准和普尔指数(Poor's Index)在 3 个月期限内表现得更好。一种理论指出证券分析师的判断并不比随机抓阄更好,4 只股票中猜测正确的股票数服从多项分布,具体如下:

| 正确数          | 0      | 1      | 2      | 3      | 4      |
|--------------|--------|--------|--------|--------|--------|
| 多项概率 $\pi_i$ | 0.0625 | 0.2500 | 0.3750 | 0.2500 | 0.0625 |

如果 125 位证券分析师所作的正确猜测结果如下,那么,用这些数据作卡方拟合优度检验。给定  $\alpha = 0.05$ ,试作出结论。

| 正确数 | 0 | 1  | 2  | 3  | 4 |
|-----|---|----|----|----|---|
| 频数  | 3 | 23 | 51 | 39 | 9 |

## 10.5 Poisson (泊松) 分布

在第四章(以及在本章),我们指出,假如  $n \geq 5/\min(\pi, 1 - \pi)$ , 则正态分布为二项分布提供了很好的近似。为了确保二项分布基本上对称,这一条件是必要的。然而,很多情况下,二项概率分布严重偏态,以至于不适合使用正态近似。例如,在临床试验中,观察服用新药品的病人,有某种副作用反应的病人可能相当少。如果  $\pi$  (观察到的具有副作用的人所占的比例)是 0.001, 那么,  $\min(\pi, 1 - \pi) = 0.001$ 。对于这样的例子,为了用正态分布来近似二项分布,样本容量必须等于或大于  $5/0.001 = 5,000$ 。

1837 年, S.D. Poisson(泊松)给出了一种离散型概率分布,因此,称之为 Poisson 分布。当  $\pi$  较小,  $n$  很大,但  $n\pi$  小于 5 时,该分布为二项分布提供了很好的近似。在  $n$  次试验中,观察到  $y$  次成功的概率由下式给出

$$P(y) = \frac{\mu^y e^{-\mu}}{y!}$$

其中  $e$  是常数, 近似等于 2.71828,  $\mu$  是  $y$  的均值。附表 15 列出了对应于各个  $\mu$  的 Poisson 概率。为了用 Poisson 分布近似二项分布, 我们取  $\mu = n\pi$ 。

### 例 10.9

参看本节开始所提到的临床试验, 其中对 1,000 位病人使用新药品进行治疗。当  $\pi = 0.001$  时, 计算 1,000 位病人中没有人具有某种副作用反应(比如呕吐)的概率。

**解答** 二项分布的样本均值为  $\mu = n\pi = 1,000(0.001) = 1$ , 将  $\mu = 1$  代入 Poisson 概率分布得

$$P(y = 0) = \frac{(1)^0 e^{-1}}{0!} = e^{-1} = \frac{1}{2.71828} = 0.367879$$

(附表 15 中对应于  $y = 0$ ,  $\mu = 1$  的概率值是 0.3679)

### 例 10.10

假定在对 1,000 位病人进行新药品的临床试验以后, 没有病人呕吐。是否有理由推断所有服过这种药的病人中有呕吐反应的病人比例不到 0.001?

**解答** 当然不能这样推断。假定  $\pi = 0.001$  (即, 假定 0.1% 的人会呕吐), 则在  $n = 1,000$  次试验中, 观察到  $y = 0$  的概率为 0.368。由于这个概率相当大, 因此, 据此推断  $\pi < 0.001$  是不明智的。

虽然在一定条件下, Poisson 分布为二项分布提供了有用的近似, 但 Poisson 分布的应用并不仅限于此。特别地, 当下列假定设满足时, 可以利用 Poisson 分布求在某段时间, 某个体积, 某个空间等某随机事件出现  $y$  次的概率。

1. 事件发生在时点上; 两个或多个事件不能同时发生。
2. 一事件在某段时间里发生与否与该事件在另一不重叠的时间里发生与否是相互独立的; 也就是说, 一事件在某段时间发生或不发生都不改变该事件在其后某段时间里发生的概率。

很多情况下, 当讨论这个问题时, 我们通常作如下第三个假设:

3. 在任何一段时间里, 事件期望出现的次数都相同, 只要时间长度相同, 而不管时间的起点在哪儿。

虽然, 这些假定好象有点严格, 但是很多情况都满足这些条件。例如, 在给定的时间段(如 1 分钟)里, 来到柜台, 停车场收费亭, 检查站, 或汽车修理铺的顾客数

量都近似服从 Poisson 分布。同样地,在显微镜下可观察到的单位体积的湖水里某种藻类的块数也近似地服从 Poisson 分布。

当我们面临一组观测数据时,我们希望检查数据是否服从 Poisson 分布。为此,我们利用第 10.4 节的拟合优度检验,检验统计量为

$$\chi^2 = \sum_i \left[ \frac{(n_i - E_i)^2}{E_i} \right]$$

有两种类型的零假设。第一种假设就是:数据来自于  $\mu = \mu_0$  的 Poisson 分布;也就是说,我们希望检验  $H_0: \mu = \mu_0$  ( $\mu_0$  已知) 对备择假设  $H_a: \mu \neq \mu_0$ 。  $n_i$  表示单元  $i$  的观测数,  $E_i$  表示根据均值为  $\mu_0$  的 Poisson 分布概率计算而得的期望观测数。将检验统计量的计算值与附表 7 中对应于  $\alpha = \alpha$ ,  $df = k - 1$  的卡方值进行比较,其中  $k$  为单元个数。

我们可能感兴趣的第二种零假设并不是一个特定的分布。我们检验

$H_0$ : 观察单元数都来自同一个均值为  $\mu$  (未知) Poisson 分布。

备择假设为:并非所有的单元数都来自同一个 Poisson 分布。检验统计量为

$$\chi^2 = \sum_i \left[ \frac{(n_i - E_i)^2}{E_i} \right]$$

其中,对所有单元,  $E_i$  表示根据 Poisson 分布概率计算而得的期望观测数, Poisson 分布的均值由样本数据估计而得。拒绝域由  $\alpha = \alpha$ ,  $df = k - 2$  决定。注意两种零假设的自由度之间的差别。在后者中,我们丢失了一个自由度,因为我们必须估计 Poisson 分布参数  $\mu$ 。

### 例 10.11

环境工程人员经常要利用某些信息来检测湖的卫生情况,比如,湖中藻类种数以及每种藻类藻的块数。那些仅产有少数几种藻类但每种藻类的数量很多的湖泊被认为是营养贫乏的湖泊。在一次调查中,通过显微镜分析湖水样本,观测每个显微面积中藻类的块数。在显微镜观测了 150 个显微面积,将样本数据整理如下。 $y_i$  表示每块面积中藻类的块数,  $n_i$  表示含有  $y_i$  块藻类的显微面积数。

| $y_i$ | 0 | 1  | 2  | 3  | 4  | 5  | 6 | $\geq 7$ |
|-------|---|----|----|----|----|----|---|----------|
| $n_i$ | 6 | 23 | 29 | 31 | 27 | 13 | 8 | 13       |

给定  $\alpha = 0.05$ , 检验零假设:样本数据来自于 Poisson 概率分布。

**解答** 在计算  $\chi^2$  值之前,首先,我们必须估计 Poisson 分布参数  $\mu$ , 然后,计算期望单元数。用样本均值  $\bar{y}$  估计 Poisson 均值  $\mu$ 。由这些数据得

$$\bar{y} = \frac{\sum_i n_i y_i}{n} = \frac{486}{150} \approx 3.3$$

需要注意的是,在 13 个最大的数被压缩成最后一个单元之前,利用所有样本数据计算的样本均值为 3.3。这就是为什么这里将样本均值近似为 3.3。

当  $\mu = 3.3$  时,查附表 15 得对应于  $y = 0, 1, \dots, 7$  或 7 以上的 Poisson 概率如下。

| $y_i$                 | 0     | 1     | 2     | 3     | 4     | 5    | 6     | $\geq 7$ |
|-----------------------|-------|-------|-------|-------|-------|------|-------|----------|
| 对 $\mu = 3.3, P(y_i)$ | 0.037 | 0.122 | 0.201 | 0.221 | 0.182 | 0.12 | 0.066 | 0.051    |

利用  $E_i = nP(y_i)$  可计算任一单元的期望单元数。因此,根据我们的数据 ( $n = 150$ ) 得期望单元数如下。

| $y_i$ | 0    | 1     | 2     | 3     | 4     | 5     | 6    | $\geq 7$ |
|-------|------|-------|-------|-------|-------|-------|------|----------|
| $E_i$ | 5.55 | 18.30 | 30.15 | 33.15 | 27.30 | 18.00 | 9.90 | 7.65     |

将这些数据代入检验统计量,得

$$\begin{aligned} \chi^2 &= \sum_i \left[ \frac{(n_i - E_i)^2}{E_i} \right] \\ &= \frac{(6 - 5.55)^2}{5.55} + \frac{(23 - 18.30)^2}{18.30} + \dots + \frac{(13 - 7.65)^2}{7.65} = 6.93 \end{aligned}$$

对应于  $\alpha = 0.05$ ,  $df = k - 2 = 6$  的卡方临界值为 12.59。因为卡方计算值并没有超过 12.59,所以,我们没有充分的证据拒绝零假设:样本数据来自于 Poisson 分布。

下面就这种检验方法,我们给出一点提示。正如我们前面提到的,当我们使用卡方统计量时,所有的期望单元数应该相当大。特别地,我们要求  $E_i > 1$  且小于 5 的  $E_i$  不超过 20%。在例 10.11 中,如果大于等于 7 的  $y$  被各自看作一类时,那么这些  $E$  将不满足  $\chi^2$  的应用条件。那就是为什么我们将所有大于等于 7 的  $y$  归为一类。

## 练习

### 基本技能

**10.51** 查附表 15, 计算如下 Poisson 概率。

a.  $P(y=1)$ , 已知  $\mu = 0.5, \mu = 1.0, \mu = 3.0$ 。

b.  $P(y > 1)$ , 已知  $\mu = 1.7, \mu = 2.5, \mu = 4.2$ 。

c.  $P(y < 5)$ , 已知  $\mu = 0.2, \mu = 1.0, \mu = 2.0$ 。

### 应用

**10.52(工程)** 长期来看, 高峰期小汽车以每分钟 6 辆停在飞机场出口处。利用附表 15, 求下面的概率。(y 表示高峰期任何一分钟来到小汽车的数量)

a.  $P(y = 0)$ 。

b.  $P(y > 1)$ 。

c.  $P(y > 0.3)$ 。

**10.53(商业)** 某公司打算采用电话销售方法作为传统销售方式的补充。据估计, 每 100 个电话中只有一个能够推销成功。假定某一天打了 250 个电话:

a. 写出不到 6 次推销成功的概率表达式——不需计算。

b. 在(a)中, 你作了什么假定?

c. 用正态近似计算  $P(y < 6)$ 。

d. 用 Poisson 近似计算  $P(y < 6)$ 。

e. (c)和(d)中, 哪一个近似更好? 为什么?

**10.54(医学)** 某种出生的缺陷发生的概率为 0.0001; 即, 10,000 个婴儿中只有一个有这种缺陷。某一年某医院出生了 5,000 个婴儿, 应该用什么近似? 至少一个婴儿有这种缺陷的近似概率为多少?

**10.55(政府)** 一项政府调查研究的一部分就是要考察公共汽车专用车道的有效性, 为此, 需要调查在特定时间段, 某十字路口发生阻塞(可能导致事故的驾驶情况)的次数。在开辟公共汽车专用车道, 以前进行的研究表明上午 7:00 至 9:00 之间的高峰期, 平均每 5 分钟发生阻塞的次数非常近似于  $\mu = 2$  的 Poisson 分布。将 40 天的样本数据整理如下; 其中  $y_i$  表示阻塞的次数,  $n_i$  表示有多少个 5 分钟观察到  $y_i$ 。

| $y_i$ | 0  | 1   | 2   | 3   | 4  | 5  | $\geq 6$ |
|-------|----|-----|-----|-----|----|----|----------|
| $n_i$ | 90 | 230 | 240 | 130 | 68 | 30 | 12       |

a. Poisson 假定成立吗?

b. 利用这些数据检验研究假设, 即每 5 分钟阻塞的次数不等于 2。(提示: 用基于 Poisson 分布的卡方检验)

**10.56(工程)** 工人们记录了 90 天中, 每天尼龙纺纱工序由于断线引起停工的次数。利用下面的样本数据考察每天停工次数是否服从 Poisson 分布, 给定  $\alpha =$

0.05。在数据中,  $y_i$  表示每天停工次数,  $n_i$  表示停工  $y_i$  次的天数。

|       |    |    |    |   |   |          |
|-------|----|----|----|---|---|----------|
| $y_i$ | 0  | 1  | 2  | 3 | 4 | $\geq 5$ |
| $n_i$ | 20 | 28 | 15 | 8 | 7 | 12       |

**10.57(生物)** 昆虫学家要研究农业区昆虫的分布情况。他们将草原分成 50 平方米的小块, 然后统计每块草原上发光蚁丘(fire ant hill)的数量, 以便研究草原上发光蚁丘的分布状况。如果零假设“草原上发光蚁丘的分布是 Poisson 分布”成立, 就说蚁丘数量在草原上是随机分布的。导致拒绝这一随机性假设的有两种情况。一种可能是蚁丘的分布是均匀的, 即每块 50 平方米的草原上有同样多的蚁丘, 另一种可能是蚁丘的分布是聚群(cluster)的。如果蚁丘数量是随机分布的, 那么方差等于均值,  $\sigma^2 = \mu$ 。如果蚁丘数量分布比随机分布更均匀, 那么就说该分布是分散不足的,  $\sigma^2 < \mu$ 。如果分布比随机分布聚群性更强, 那么称分布是过度分散的,  $\sigma^2 > \mu$ 。记录了草原上 100 块 50 平方米的地块。在数据中,  $y_i$  表示每块草原上发光蚁丘的数量,  $n_i$  表示有  $y_i$  只发光蚁丘的草原块数。

|       |   |   |   |    |    |    |    |    |    |   |    |    |    |
|-------|---|---|---|----|----|----|----|----|----|---|----|----|----|
| $y_i$ | 0 | 1 | 2 | 3  | 4  | 5  | 6  | 7  | 8  | 9 | 12 | 15 | 20 |
| $n_i$ | 2 | 6 | 8 | 10 | 12 | 15 | 13 | 12 | 10 | 6 | 3  | 2  | 1  |

- 估计每 50 平方米草原上发光蚁丘的数量的均值和方差; 即用第三章的公式, 计算  $\bar{y}$  和  $s^2$ 。
- 草原上发光蚁丘的分布是随机的吗? 给定  $\alpha = 0.05$ , 利用 Poisson 分布的卡方检验拟合这些数据。
- 如果你拒绝了用卡方分布作为发光蚁丘分布模型, 那么草原上发光蚁丘的分布是更聚群, 还是更均匀?

## 10.6 列联表: 独立性检验和齐性检验

在第 10.3 节, 我们介绍了比较两个比例的检验。为此, 只要简单数一下每个样本中某一结果出现了多少次。本节, 我们将这种检验加以拓广。首先, 我们要建立一个检验统计量, 检验样本数据与理论比例的偏离是否仅是由于偶然而产生的。

我们在第四章第一次引入概率思想时, 我们首先从频数(次数)分布表开始。当时, 我们处理这些次数时好象他们代表了整个总体。实际上, 我们几乎不可能知

道总体的全部数据;通常情况下,我们仅仅知道总体的一个样本。当我们根据样本进行计数时,通常将他们处理成交叉列表或列联表的形式。本节,我们将介绍一种特殊的检验,即独立性卡方检验,它被经常用来检验列联表。

在第四章,我们引入了独立性的概念。特别地,我们讨论了变量之间的相依意味着一个变量对于另一个变量的预测比有价值的。通常样本数据之间似乎也存在着相依。本节,我们介绍一种  $\chi^2$  检验,据此可检查样本数据中所呈现的相依是否仅是一种意外,即是随机变化的结果,而不是真正的相依。

首先,频数数据被排成  $r$  行  $c$  列的交叉列表形式。一个变量可能的值决定表的行,另一变量可能的值决定表的列。我们用  $\pi_{ij}$  表示落在第  $i$  行,第  $j$  列的总体比例(或概率)。第  $i$  行总的比例为  $\pi_{i\cdot}$ ,第  $j$  列总的比例为  $\pi_{\cdot j}$ 。如果行比例(概率)和列比例(概率)是独立的,那么  $\pi_{ij} = \pi_{i\cdot} \pi_{\cdot j}$ 。例如,某大公司人事部门经理要评价在 4 个不同办公室的职员之间实行 3 种不同的弹性工作制的受欢迎程度。下表列出的一组比例  $\pi_{ij}$  显示了独立性。在第 1 个办公室喜欢第 2 种方案的职员所占比例为  $\pi_{12} = 0.03$ ,所有喜欢第 2 种方案的职员比例为  $\pi_{\cdot 2} = 0.30$ ,第 1 个办公室职员所占比例为  $\pi_{1\cdot} = 0.10$ 。于是,该单元的独立性成立,因为  $\pi_{12} = 0.03 = (\pi_{1\cdot})(\pi_{\cdot 2}) = (0.10)(0.30)$ 。其他单元的独立性也都成立。

| 喜欢的方案 | 办公室  |      |      |      | 总和   |
|-------|------|------|------|------|------|
|       | 1    | 2    | 3    | 4    |      |
| 1     | 0.05 | 0.20 | 0.15 | 0.10 | 0.50 |
| 2     | 0.03 | 0.12 | 0.09 | 0.06 | 0.30 |
| 3     | 0.02 | 0.08 | 0.06 | 0.04 | 0.20 |
| 总和    | 0.10 | 0.40 | 0.30 | 0.20 |      |

$\chi^2$  检验的零假设为:两种比例是独立的。研究假设为:存在某种相依,也就是说,并不是在每个单元,等式  $\pi_{ij} = \pi_{i\cdot} \pi_{\cdot j}$  都成立。检验统计量为对所有单元求和,其中,和式的每一项为

$$(\text{观察值} - \text{期望值})^2 / \text{期望值}$$

对于独立性检验,在零假设成立的条件下,期望值  $E_{ij}$  计算不同于拟合优度检验。独立性的零假设并没有给出行概率  $\pi_{i\cdot}$  和列概率  $\pi_{\cdot j}$  的具体数值,所以必须根据相应行、列的频率对其进行估计。如果  $n_{i\cdot}$  是第  $i$  行的实际频数,那么用  $\hat{\pi}_{i\cdot} = n_{i\cdot}/n$  来估计  $\pi_{i\cdot}$ ;同样地,用  $\hat{\pi}_{\cdot j} = n_{\cdot j}/n$  来估计  $\pi_{\cdot j}$ 。假定独立性的零假设成立,那么应有  $\hat{\pi}_{ij} = \hat{\pi}_{i\cdot} \hat{\pi}_{\cdot j} = (n_{i\cdot}/n)(n_{\cdot j}/n)$ 。



**定义 10.2** 在独立性假设下,第  $i$  行,第  $j$  列期望值的估计为

$$\hat{E}_{ij} = n\hat{\pi}_{ij} = n \frac{n_{i.}}{n} \frac{n_{.j}}{n} = \frac{(n_{i.})(n_{.j})}{n}$$

即行频数乘以列频数再除以总频数。

**例 10.12**

假定在弹性工作制例子中,根据 216 名职员的本样本数据得下面频数:

| 喜欢的方案 | 办公室 |    |    |    | 总和 |
|-------|-----|----|----|----|----|
|       | 1   | 2  | 3  | 4  |    |
| 1     | 15  | 32 | 18 | 5  | 70 |
| 2     | 8   | 29 | 23 | 18 | 78 |
| 3     | 1   | 20 | 25 | 22 | 68 |
| 总和    | 24  | 81 | 66 | 45 |    |

试计算  $\hat{E}_{ij}$  值的表。

**解答** 对于第 1 行,第 1 列,期望值的估计为

$$\hat{E}_{11} = \frac{(\text{第 1 行频数})(\text{第 1 列频数})}{\text{总频数}} = \frac{(70)(24)}{216} = 7.78$$

对其他单元作同样的计算,得下表。

| 喜欢的方案 | 办公室   |       |       |       | 总和     |
|-------|-------|-------|-------|-------|--------|
|       | 1     | 2     | 3     | 4     |        |
| 1     | 7.78  | 26.25 | 21.39 | 14.58 | 70.00  |
| 2     | 8.67  | 29.25 | 23.83 | 16.25 | 78.00  |
| 3     | 7.56  | 25.50 | 20.78 | 14.17 | 68.00  |
| 总和    | 24.01 | 81.00 | 66.00 | 45.00 | 216.00 |

易见,  $\hat{E}_{ij}$  表中行频数和列频数与观察值  $(n_{ij})$  表中相应的行频数和列频数相同(除舍入误差外)。

**独立性的  $\chi^2$  检验**

$H_0$ : 行变量与列变量相互独立。

$H_a$ : 行变量与列变量是相依的(关联的)

$$\text{T.S.: } \chi^2 = \sum_{i,j} (n_{ij} - \hat{E}_{ij})^2 / \hat{E}_{ij}$$

R.R.: 若  $\chi^2 > \chi^2_{\alpha}$ , 则拒绝  $H_0$ , 其中  $\chi^2_{\alpha}$  将自由度为  $(r-1)(c-1)$  的  $\chi^2$  分布割取面积  $\alpha$ ;  $r$  = 行数,  $c$  = 列数。

检查假定, 做出结论。

该检验统计量有时也被称之为 Pearson(皮尔逊)  $\chi^2$  统计量。

独立性  $\chi^2$  检验的自由度与双向表中固定边际不变时自由的单元数有关。例如, 在  $2 \times 2$  表(2 行, 2 列)中, 只有 1 个单元项自由变化。一旦该项固定, 我们就可通过从相应的行频数和列频数中减去该项来求出其余各单元项。在表 10.4(a), 我们任意标出了一些合计数, 标有 \* 的单元可以取任何值(不能超过相应的合计数), 但所有剩下的单元都将有合计数确定。同样地, 对于  $2 \times 3$  表(2 行, 3 列), 两个单元项可任意变化, 比如标有 \* 的两个单元。一旦这些项固定, 我们就能通过从相应的行频数和列频数[见表 10.4(b)]中减去该项来求出其余各单元项。一般来说, 对于  $r$  行  $c$  列的表, 共有  $(r-1)(c-1)$  个单元项是自由变化的。因此, 这个数表示独立性  $\chi^2$  检验的自由度。

表 10.4 (a)  $2 \times 2$  表中有 1 个自由度 (b)  $2 \times 3$  表中有 2 个自由度

| 类 B |    |    |    | 类 B |    |    |    |
|-----|----|----|----|-----|----|----|----|
| 类 A |    |    | 总和 | 类 A |    |    | 总和 |
|     | *  |    | 16 |     | *  |    | 51 |
|     |    |    | 34 |     |    |    | 40 |
| 总和  | 21 | 29 | 50 | 总和  | 28 | 41 | 91 |
| (a) |    |    |    | (b) |    |    |    |

该独立性  $\chi^2$  检验是建立在近似的基础上。一个保守的准则是, 使用这种近似时, 每个  $\hat{E}_{ij}$  必须不小于 5。一般在实践中, 如果某些  $\hat{E}_{ij}$  太小, 就将那些行(或列)合并, 直到该准则满足为止。

### 例 10.13

就例 10.8 中的数据, 完成独立性  $\chi^2$  检验。首先用  $\alpha = 0.05$ , 然后给出  $p$  值的上界。

**解答** 零假设和备择假设为:

$H_0$ : 弹性工作制的受欢迎程度与办公地点相互独立。

$H_a$ :弹性工作制的受欢迎程度依赖于办公地点。

利用例 10.12 中的  $n_{ij}$  和  $E_{ij}$  可计算检验统计量:

$$\begin{aligned} \text{T.S.: } \chi^2 &= \sum_{i,j} (n_{ij} - E_{ij})^2 / E_{ij} \\ &= (15 - 7.78)^2 / 7.78 + (32 - 26.25)^2 / 26.25 \\ &\quad + \cdots + (22 - 14.17)^2 / 14.17 \\ &= 6.70 + 1.26 + \cdots + 4.33 = 27.13 \end{aligned}$$

R.R.: 对于  $df = (3-1)(4-1) = 6$ ,  $\alpha = 0.05$ , 查附表 7 得临界值为 12.59。由于  $\chi^2 = 27.13$  超过了 12.59, 因此拒绝  $H_0$ 。事实上, 由于 27.13 大于附表 7 中对应于  $\alpha = 0.001$  的临界值, 因此,  $p$  值  $< 0.001$ 。

检查假定, 做出结论: 既然每个期望值都超过 5, 因此  $\chi^2$  近似应该是好的。从而, 我们有充分的证据表明: 弹性上班制的受欢迎程度依赖于办公地点。

另一个  $\chi^2$  统计量被称之为**似然比统计量**, 它常出现在计算机输出结果中。定义为

$$\text{似然比 } \chi^2 = \sum_{i,j} n_{ij} \ln(n_{ij} / (n_{i.} n_{.j}))$$

其中  $n_{i.}$  是第  $i$  行的总频数,  $n_{.j}$  是第  $j$  列的总频数,  $\ln$  是自然对数(以  $e = 2.71828$  为底)。该似然比的值也应该和具有同样自由度  $(r-1)(c-1)$  的  $\chi^2$  分布相比较。虽然, 不太明显, 但这种形式的  $\chi^2$  独立性检验与 Pearson 形式的检验基本相同。有理由相信, 就表中值而言, Pearson  $\chi^2$  统计量比似然比统计量更近似于服从  $\chi^2$  分布, 所以, 我们宁愿用它, 而不用似然比形式。

独立性  $\chi^2$  检验的惟一功能就是检查样本数据中表面的相依是否可能纯属偶然, 即为随机变化的结果。拒绝零假设仅表明并不能合理地将这种明显的相依归结为偶然。它并不表示任何**关联的强度**或形式。

同样的  $\chi^2$  检验统计量可应用于另一种稍微不同的抽样方法。我们围绕着独立性  $\chi^2$  检验的讨论实际上暗含着一个假定, 即数据来自于整个总体的某单一随机样本。经常有这样的情况, 即各个样本取自于列(行)变量定义的子总体。在弹性工作制例子(例 10.12)中, 数据可能来自于四个不同样本(容量分别为 24, 81, 66 和 45)而不是来自于由 216 名职员组成的单一样本, 四个样本分别取自于四个办公室。

一般地, 假定列的类别代表  $c$  个不同的子总体。容量为  $n_1, n_2, \dots, n_c$  的随机样本取自于这些子总体。然后将每个子总体的观察值归为  $r$  个类别, 他们对应于一个类型变量的  $r$  个值, 由列联表中的  $r$  行表示。研究假设为不同的子总体在  $r$  个类别上的概率分布是不同的, 其中  $r$  个类别对应于类型变量的  $r$  个水平。零假

设为,对于所有  $c$  个子总体,不同子总体的一组  $r$  个比例  $(\pi_{1j}, \pi_{2j}, \dots, \pi_{rj})$  是相同的。因此,零假设被表示为

$$H_0: (\pi_{11}, \pi_{21}, \dots, \pi_{r1}) = (\pi_{12}, \pi_{22}, \dots, \pi_{r2}) = \dots = (\pi_{1c}, \pi_{2c}, \dots, \pi_{rc})$$

该检验被称为分布的**齐性**(又译**同质性**)**检验**。齐性检验和独立性检验的机制是一致的。然而,需要注意的是,抽样方法和结论不同。对于独立性检验,我们随机地从单个总体中选取  $n$  个个体,根据两个类型变量的取值将这些个体进行分类。然后,我们要检验这两个类型变量是否相互关联。在关于比例的齐性检验中,我们从  $c$  个子总体中随机选取  $n = n_1 + n_2 + \dots + n_c$  个个体,再根据单个类型变量的取值将他们进行分类。我们要检验对于所有  $c$  个子总体,在类型变量的值上子总体的概率分布是否相同。

正如我们在第 10.4 节所讨论的,  $\chi^2$  的抽样分布对卡方分布的近似程度依赖于样本容量  $n$  和单元数  $k$ 。Cochran(1954)指出,如果没有  $E_i$  小于 1, 小于 5 的  $E_i$  不足 20%, 那么该近似程度应是足够的。Larntz(1978)和 Koehler(1986)证明,当样本容量较小时,  $\chi^2$  比似然比检验统计量更有效。针对各种不同的样本容量, Agresti(1990)比较了检验独立性的两种检验统计量的名义  $\alpha$ -水平和实际  $\alpha$ -水平。当  $n/k$  超过 1 时,  $\chi^2$  检验统计量还算可以。我们再一次推荐使用 Cochran 准则考察卡方检验统计量是否充分近似于卡方分布。当某些  $E_{ij}$  太小时,可以使用其他几种方法。调查人员将类型变量的水平进行合并以增加观察单元数。然而,除非有自然方式重新定义类的水平而不改变被检验假设的自然性,否则,不能对类进行合并。当不可能获得足够大的观察单元数以允许卡方近似时, Agresti(1990)讨论了检验假设的精确方法。例如,当两个类型变量仅有两个水平时,可以使用 Fisher 精确检验法。

#### 例 10.14

现有一随机样本,其中 200 人来自主要产油和天然气的国家,200 人来自主要产煤国,400 人来自其他国家。让他们参加一项民意测验,调查他们对 5 种可能的能源政策的态度。每个被调查者从下面选出他最赞同的政策:

1. 主要强调节约。
2. 主要强调开采国内的石油和天然气。
3. 主要强调在太阳能方面进行投资。
4. 主要强调核能的开发和安全。
5. 主要是减少环境限制,强调使用煤炭。

结果如下:

| 赞同的政策 | 石油天然气国家 | 产煤国 | 其他  | 总和  |
|-------|---------|-----|-----|-----|
| 1     | 50      | 59  | 161 | 270 |
| 2     | 88      | 20  | 40  | 148 |
| 3     | 56      | 52  | 188 | 296 |
| 4     | 4       | 3   | 5   | 12  |
| 5     | 2       | 66  | 6   | 74  |
| 总和    | 200     | 200 | 400 | 800 |

其计算见 Execustat 输出结果。每个单元的第二项是该单元数占所在列的百分比。

| Crosstabulation |            |            |             |              |
|-----------------|------------|------------|-------------|--------------|
|                 | OilGas     | Coal       | Other       | Row<br>Total |
| 1               | 50<br>25.0 | 59<br>29.5 | 161<br>40.3 | 270<br>33.75 |
| 2               | 88<br>44.0 | 20<br>10.0 | 40<br>10.0  | 148<br>18.50 |
| 3               | 56<br>28.0 | 52<br>26.0 | 188<br>47.0 | 296<br>37.00 |
| 4               | 4<br>2.0   | 3<br>1.5   | 5<br>1.3    | 12<br>1.50   |
| 5               | 2<br>1.0   | 66<br>33.0 | 6<br>1.5    | 74<br>9.25   |
| Column          | 200        | 200        | 400         | 800          |
| Total           | 25.00      | 25.00      | 50.00       | 100.00       |

Summary Statistics for Crosstabulation

| Chi-square | D. F. | P Value |
|------------|-------|---------|
| 289.22     | 8     | 0.0000  |

Warning: Some table cell counts < 5.

对三类国家, 试作分布的  $\chi^2$  齐性检验, 并给出该检验的  $p$  值。

**解答** 为了检验相应的总体分布是不同的, 利用下面的期望值表:

| 赞同的政策 | 石油天然气国家 | 产煤国  | 其他  |
|-------|---------|------|-----|
| 1     | 67.5    | 67.5 | 135 |
| 2     | 37      | 37   | 74  |
| 3     | 74      | 74   | 148 |
| 4     | 3       | 3    | 6   |
| 5     | 18.5    | 18.5 | 37  |

我们看到,期望值表中有两个  $E_{ij}$  小于 5。然而,因为只有  $2/15 = 13\%$  的  $E_{ij}$  小于 5,且所有  $E_{ij}$  都大于 1,所以,将卡方近似应用到检验统计量的准则是满足的。检验步骤如下:

$H_0$ : 列分布是齐性的。

$H_a$ : 列分布是非齐性的。

$$\begin{aligned} \text{T.S.: } \chi^2 &= \sum_{i,j} (n_{ij} - \hat{E}_{ij})^2 / \hat{E}_{ij} \\ &= (50 - 67.5)^2 / 67.5 + (88 - 37)^2 / 37 + \cdots + (6 - 37)^2 / 37 \\ &= 289.22 \end{aligned}$$

R.R.: 因为,对应于  $df = 8$ ,  $\alpha = 0.001$  的  $\chi^2$  临界值为 26.12, 所以  $p$ -值  $< 0.001$ 。

检查假定并做出结论:即使认识到  $\chi^2$  近似的精度有限,我们仍能以很小的  $p$  值拒绝齐性假设。百分比分析表明,特别是给定某一政策,不同类型的国家所占的百分比呈现出很大的差别;例如,生活在产油/气的国家,赞同政策 5 的人只占 1%,与其相比较,生活在产煤国的人赞同政策 5 的人占 33%。

本节介绍的  $\chi^2$  检验用途有限,但很重要。该检验仅仅被用来考察,从统计上数据是否显示出各个类之间具有可察觉的(显著的)关系。它并不能衡量这种显著的关系究竟有多强。在一个大的数据集中,弱相关可能是可察觉的(显著的);而在一个小的数据集中,强相关可能并不显著。

### 练习

**10.58(人力资源)** 某大型研究型公司的人事部经理将学院和大学分成 4 类,即最受欢迎的,比较好的,中等以及不受欢迎的,以在招募大学毕业生时参考。该经理收集了 156 名最近毕业的学生,让主管给每位毕业生评级。

| 学校   | 级别 |    |    |
|------|----|----|----|
|      | 优秀 | 一般 | 差  |
| 最受欢迎 | 21 | 25 | 2  |
| 较好   | 20 | 36 | 10 |
| 中等   | 4  | 14 | 7  |
| 不受欢迎 | 3  | 8  | 6  |

Execustat 计算机软件包输出结果如下:

|  | Crosstabulation |         |       | Row<br>Total |
|--|-----------------|---------|-------|--------------|
|  | Outstanding     | Average | Poor  |              |
| Most desirable                         | 21              | 25      | 2     | 48           |
|  | 43.8            | 52.1    | 4.2   | 30.77        |
| Good                                   | 20              | 36      | 10    | 66           |
|  | 30.3            | 54.5    | 15.2  | 42.31        |
| Adequate                               | 4               | 14      | 7     | 25           |
|  | 16.0            | 56.0    | 28.0  | 16.03        |
| Undesirabel                            | 3               | 8       | 6     | 17           |
|  | 17.6            | 47.1    | 35.3  | 10.90        |
| Column                                 | 48              | 83      | 25    | 156          |
| Total                                  | 30.77           | 53.21   | 16.03 | 100.00       |
| Summary Statistics for Crosstabulation |                 |         |       |              |
| Chi-square                             | D. F.           | P Value |       |              |
| 15.97                                  | 6               | 0.0139  |       |              |

Warning: Some table cell counts < 5.

- 计算  $\chi^2$  统计量的值。
- 求  $p$  值。
- 经理能否安全地得出结论, 认为学校类型和学生等级之间有关系?
- 在使用  $\chi^2$  近似时, 还存在什么问题吗?

**10.59** 行百分比(输出单元的第2项)反映了我们在练习 10.58 中发现的相关性吗?

**10.60(人力资源)** 为了进行一项有关潜在年龄歧视的研究, 需要考察一大型公司中层经理人员的提升情况。数据如下:

|     | 年龄    |       |       |        | 总和  |
|-----|-------|-------|-------|--------|-----|
|     | 30 以下 | 30~39 | 40~49 | 50 及以上 |     |
| 提升  | 9     | 29    | 32    | 10     | 80  |
| 未提升 | 41    | 41    | 48    | 40     | 170 |
| 总和  | 50    | 70    | 80    | 50     |     |

Minitab 输出结果如下:

```
MTB> Table 'promoted' 'agegroup';
```

```
SUBC> Counts;
```

```
SUBC> ColPercents;
```

```
SUBC> ChiSquare.
```

|     | ROWS: promoted |        |        |        | COLUMNS: agegroup |
|-----|----------------|--------|--------|--------|-------------------|
|     | 1              | 2      | 3      | 4      |                   |
| 1   | 9              | 29     | 32     | 10     | 80                |
|     | 18.00          | 41.43  | 40.00  | 20.00  | 32.00             |
|     | 9              | 29     | 32     | 10     | 80                |
| 2   | 41             | 41     | 48     | 40     | 170               |
|     | 82.00          | 58.57  | 60.00  | 80.00  | 68.00             |
|     | 41             | 41     | 48     | 40     | 170               |
| ALL | 50             | 70     | 80     | 50     | 250               |
|     | 100.00         | 100.00 | 100.00 | 100.00 | 100.00            |
|     | 50             | 70     | 80     | 50     | 250               |

```
CHI-SQUARE = 13.025 WITH D.F. = 3
```

```
CELL CONTENTS -
```

```
COUNT
```

```
% OF COL
```

```
COUNT
```

- 求在独立性假设下的期望数。
- 验证所输出的自由度是正确的。



c. 给定  $\alpha = 0.05$ , 年龄和提升之间是否存在统计上的显著关系?

10.61 给出练习 10.60 中  $p$  值的界限。

10.62 将练习 10.60 的数据合并如下:

|     | 年龄    |        | 总和  |
|-----|-------|--------|-----|
|     | 直到 39 | 40 及以上 |     |
| 提升  | 38    | 42     | 80  |
| 未提升 | 82    | 88     | 170 |
| 总和  | 120   | 130    |     |

Minitab 输出结果如下:

```
MTB > Table 'promoted' 'combined';
```

```
SUBC> Counts;
```

```
SUBC> ColPercents;
```

```
SUBC> ChiSquare.
```

| ROWS: | promoted | COLUMNS: combined |        |
|-------|----------|-------------------|--------|
|       | 1        | 2                 | ALL    |
| 1     | 38       | 42                | 80     |
|       | 31.67    | 32.31             | 32.00  |
|       | 38       | 42                | 80     |
| 2     | 82       | 88                | 170    |
|       | 68.33    | 67.69             | 68.00  |
|       | 82       | 88                | 170    |
| ALL   | 120      | 130               | 250    |
|       | 100.00   | 100.00            | 100.00 |
|       | 120      | 130               | 250    |

CHI SQUARE = 0.012 WITH D. F. = 1

a. 能够用一个合理的  $\alpha$ , 拒绝独立性假设吗?

b. 合并年龄类别会产生什么影响? 将该答案和练习 10.60 的答案进行比较。

## 10.7 相关程度的度量

我们在第 10.5 节讨论的  $\chi^2$  检验有一个内在的限制。通过设计, 检验仅仅回答了这样一个问题, 即从统计上类别之间是否显示出可察觉的(显著的)关系。它并不能回答这种关系是强的, 是有趣的, 或是有意义的。这并不是说该检验有什么不好; 事实上, 没有哪个检验能够回答这些问题。本节, 我们讨论如何度量交叉列表数据中所呈现的相关性的强度。

度量相关性强度最简单(通常也是最好)的方法就是简单的百分比分析。如果没有关系(即, 如果完全独立), 那么, 行或列百分比也不会呈现任何关系。例如, 假定某直接邮寄公司要检验该公司两种不同的报价, 看看他们引起的响应率是否不同。结果如下:

| 报价 | 响应  |     | 总和  |
|----|-----|-----|-----|
|    | 是   | 否   |     |
| A  | 40  | 160 | 200 |
| B  | 80  | 320 | 400 |
| 总和 | 120 | 480 | 600 |

为了检查相关性, 我们计算对每一种报价响应的百分比。我们看到,  $(40/200) = 0.20$  (即 20%) 对 A 做出反应,  $(80/400) = 0.20$  对 B 做出反应。因为百分比相同, 所以没有相关性的迹象。再者, 我们注意到, 回答“赞同”的被调查者中有三分之一, 回答“不赞同”的被调查者中也有三分之一被提供 A 报价。由于这些分数完全相同, 因此, 没有迹象表明存在统计相关。

当然, 很少有数据绝对不显示出样本中的相关性。更普遍的是, 行或列百分比不相同, 呈现出某种程度的相关。例如, 某公司计划销售一种清洁产品, 为此, 公司委托一家市场调研机构对当今的主要产品进行研究。他们所感兴趣的变量是使用频数和主要产品的的评价等级。数据如下:

| 使用 | 评价  |     |     | 总和   |
|----|-----|-----|-----|------|
|    | 尚可  | 好   | 优秀  |      |
| 很少 | 64  | 123 | 137 | 324  |
| 偶尔 | 131 | 256 | 129 | 516  |
| 经常 | 209 | 171 | 45  | 425  |
| 总和 | 404 | 550 | 311 | 1265 |

一种自然的分析方法就是利用表中的频数,将评价等级看成是使用情况的函数。要进行分析,就必须考虑给定使用因子时,等级因子的条件概率,但要知道数据仅仅是一个样本,而不是总体本身。当使用因子为很少使用时,最好的估计就是有  $64/324 = 0.1975$  (或 19.75%) 比例的人将产品评为尚可,  $123/324 = 0.3796$  评为好,  $137/324 = 0.4228$  评为优秀。对于偶尔使用者,相应的比例为  $131/516 = 0.2539$ ,  $256/516 = 0.4961$  和  $129/516 = 0.2500$ 。对于经常使用者,比例为  $0.4918$ ,  $0.4024$  和  $0.1059$ 。对于三种使用类别来讲,比例(或百分比,即比例乘以 100)差别相当大,这表明评价等级和使用频率是有关系的。再者,我们也可以计算在评价等级类别中使用类别的百分比。每种情形都显示了某种关系。因为,随着使用频率的不同,等级的百分比也存在着相当大的差异(或随着等级百分比的不同,使用频率的比例也存在着相当大的差异),所以,这显示出使用频率和评价等级之间存在着相当强的关系。

分析数据中相关性的另一种方法就是考虑可预测性。数据中显示出的相关性越强,则由其他变量值就越能预测某个变量的取值。我们可以设想一种情况,其中每位很少使用者将产品评价为优秀,每位不经常使用者将产品评价为好,每位经常使用者将产品评价为尚可。这种情况下存在着完全统计关系;从可预测性角度来说,给定使用频率,就可以精确地预测出评价等级。当然,在实际中,关系和预测是不完全的。我们需要一个相关强度的度量,定义为可预测度。

我们需要区分**因变量**——试图要预测的变量和**自变量**——被用来作预测的变量。如果给定使用频率时,要预测评价等级,那么使用频率就是自变量,评价等级就是因变量。没有内在的因果关系;因变量和自变量的选择全凭作数据分析人员的个人意愿。

最简单的预测方法就是预测因变量取其最常见值(众数);该方法是  $\lambda$  (lambda) 可预测度的基础。在使用-评价例子中,在很少使用的情况,最常见的评价是优秀,有 137 人响应;如果对每一个很少使用的个人,我们均预测他的评价是优秀,那么,我们就有  $187 = 64 + 123 = 324 - 137$  次预测错误。同样地,对于偶尔使用的情况,预测评价均为好,则将产生  $260 = 131 + 129 = 516 - 256$  次预测错误;对经常使用而言,作评价均为尚可的预测将产生  $216 = 171 + 45 = 425 - 209$  次错误。错误总数是  $187 + 260 + 216 = 663$  次。相比较而言,如果不知道使用情况,那么,我们预测因变量取其最常见值,即“好”,这样我们将犯  $715 = 404 + 311 = 1265 - 550$  次错误。因此,有足够理由认为,在预测评价结果方面,和不知道使用信息相比,当知道使用信息时,我们做得更好。

$\lambda$  的大小表明我们预测效果多么好。当使用为自变量,评价为因变量时,预测错误的差等于  $715 - 663 = 52$ ;我们用这个差作为分子,用不知道自变量情况下所犯预测错误的次数,即 715 作为分母。如果使用是自变量,那么

$$\lambda = \frac{\text{自变量未知时预测错误的次数} - \text{自变量已知时预测错误的次数}}{\text{自变量未知时预测错误的次数}}$$

$$= \frac{715 - 663}{715} = 0.073$$

$\lambda$  在 0 和 1 之间,  $\lambda = 1$  表示, 至少在样本中, 给定自变量就可精确地预测因变量。如果数据中存在独立性, 或者总是预测因变量为同一个值, 那么  $\lambda = 0$ 。在解释  $\lambda$  的其他值时, 要注意  $\lambda$  是一个错误比例的下降(PRE)。 $\lambda = 0.073$  意味着知道使用信息比不知道使用信息时预测评价所犯的误差要少 7.3%。在实际数据中, 大于 0.30 的  $\lambda$  很少见; 因此,  $\lambda = 0.073$  表明评价和使用之间存在着一定的相关性。注意,  $\lambda$  值的大小依赖于将哪个变量看作因变量。例如, 根据评价来预测使用时,  $\lambda = 0.115$ 。

### 定义 10.3 $\lambda$ 的计算

对于自变量的每个水平(每行或每列), 求因变量的众数(如果有两个或多个具有相同频数的众数时, 任选一个)。将所有非众数单元相加, 得  $K$  = 已知自变量值时预测错误的次数。

查看因变量的边际(总)频数, 将所有非众数类别相加, 得  $U$  = 自变量值未知时预测错误的次数。

$$\lambda = (U - K)/U$$

### 例 10.15

现对单位内部的办事员, 一般管理人员和基层经理进行调查, 以了解他们对弹性上班制的看法。下面是 SAS 输出结果(注: 在结果中, 办事员、一般管理人员和基层经理分别标记为 clerical、superv 和 junior manager, 四种看法分别标记为 strongly oppose(强烈反对), oppose(反对), favor(赞成), strongly favor(极力赞成)):

| TABLE OF OPINION BY LEVEL |          |        |                |       |
|---------------------------|----------|--------|----------------|-------|
| OPINION                   | LEVEL    |        |                |       |
| FREQUENCY                 |          |        |                |       |
| COL PCT                   |          |        |                |       |
|                           | clerical | superv | junior manager | TOTAL |
| strongly oppose           | 5        | 8      | 9              | 22    |
|                           | 12.50    | 26.67  | 45.00          |       |

|                |       |       |       |    |
|----------------|-------|-------|-------|----|
| oppose         | 8     | 10    | 5     | 23 |
|                | 20.00 | 33.33 | 25.00 |    |
| favor          | 14    | 7     | 4     | 25 |
|                | 35.00 | 23.33 | 20.00 |    |
| strongly favor | 13    | 5     | 2     | 20 |
|                | 32.50 | 16.67 | 10.00 |    |
| TOTAL          | 40    | 30    | 20    |    |

STATISTICS FOR TABLE OF OPINION BY LEVEL

| STATISTIC  | DF | VALUE  | PROB  |
|------------|----|--------|-------|
| CHI-SQUARE | 6  | 12.110 | 0.060 |

| STATISTIC             | VALUE | ASE   |
|-----------------------|-------|-------|
| LAMBDA ASYMMETRIC C R | 0.120 | 0.106 |
| LAMBDA ASYMMETRIC R C | 0.123 | 0.079 |

SAMPLE SIZE = 90

ASE IS THE ASYMPTOTIC STANDARD ERROR.

R|C MEANS ROW VARIABLE DEPENDENT ON COLUMN VARIABLE.

- 关于数据中呈现的相关性,  $\chi^2$  检验表明了什么?
- 通过考察列百分比, 你发现了什么?
- 将看法视为因变量, 求  $\lambda$  值, 并解释之。
- 证明  $\lambda$  的计算是正确的。

### 解答

- 在输出结果中,  $\chi^2$  统计量表示为 CHI-SQUARE = 12.110, 自由度为 6, p 值表示为 PROB = 0.060。在  $\alpha = 0.05$  下, 我们不能拒绝独立性的假设。但

差不多。另外,  $\chi^2$  检验的结果表明我们不能太相信这种表面的相关性; 这种相关性或许纯属随机变化所致。

- c. 此时行变量是因变量。结果中“R|C MEANS ROW VARIABLE DEPENDENT ON COLUMN VARIABLE”表示我们需要的是“LAMBDA ASYMMETRIC R|C”, 即把行变量是因变量时的  $\lambda$  值, 其值为 0.123。
- d. 预测的结果可以表示为是 (CLERICAL, FAVOR), (SUPERV, OPPOSE), (JUNIOR MANAGER, STRONGLY OPPOSE), 即列中最高频数的单元。将其他所有单元的频数相加, 得  $K = 5 + 8 + 13 + 8 + 7 + 5 + 5 + 4 + 2 = 57$ 。在 OPINION 的 TOTAL 列中, 我们发现最大频数对应的观点是 FAVOR。将其他类的频数相加, 得  $U = 22 + 23 + 20 = 65$ 。因此,

$$\lambda = \frac{65 - 57}{65} = 0.123$$

百分比分析和  $\lambda$  值起着和  $\chi^2$  根本不同的作用。 $\chi^2$  检验是了解有多少证据表明存在相关性, 而不管相关程度的大小。百分比分析和  $\lambda$  值是用手中的数据考察相关性究竟有多强。这两种类型的分析互为补充。

最后我们就有关计数数据和相关性的一些问题进行讨论。

1.  $\chi^2$  拟合优度检验将次数和理论概率进行比较, 其中理论概率不需要样本数据就可直接计算。相反,  $\chi^2$  独立性检验将一个子集(比如, 一行)的次数和其他行进行比较。究竟需要哪一种检验, 一种方法就是要看是否存在一组样本数据之外的理论概率。如果是这样的话, 当属拟合优度检验。

2. 正如任何显著性检验,  $\chi^2$  检验的惟一目的就在于考察样本数据的差异是否只是随机因素所致。一个检验并不能直接告诉你这种差异有多大或有多重要。

3. 特别地, 统计上可察觉的(显著的)  $\chi^2$  独立性检验并不一定意味着很强的关系, 不显著的拟合优度检验也并不一定意味着样本比例非常接近于理论概率。

4. 在决定结果是否具有实际的重要性时, 仔细地查看一下百分比至关重要。

#### 例 10.16

将总共 9,035 份被批准的抵押贷款按批准贷款的机构进行分类(银行, 银行拥有抵押权的公司或独立抵押公司), 同时还根据借款人的收入水平进行分类。分析人员想知道批准贷款的机构和获得贷款的借款人收入水平之间是否存在统计上的相关。数据见图 10.2。

根据图 10.2 所示的 Excel 输出结果, 我们能看到统计相关性的明显迹象吗? 如果能的话, 这种相关性强吗?

|    | A                        | B     | C     | D        | E     |
|----|--------------------------|-------|-------|----------|-------|
| 1  |                          | Banks | Bank  | Mortgage | Total |
| 2  |                          |       | Owned | Company  |       |
| 3  |                          |       |       |          |       |
| 4  |                          |       |       |          |       |
| 5  | < \$ 21,000              | 943   | 166   | 151      | 1260  |
| 6  | \$ 21,000 to < \$ 33,000 | 2186  | 503   | 711      | 3400  |
| 7  | \$ 33,000 to < \$ 42,000 | 708   | 211   | 589      | 1508  |
| 8  | \$ 42,000 to < \$ 50,000 | 424   | 150   | 397      | 971   |
| 9  | \$ 50,000 +              | 948   | 239   | 709      | 1896  |
| 10 |                          |       |       |          |       |
| 11 | Total                    | 5209  | 1269  | 2557     | 9035  |

图 10.2 抵押数据的 Excel 输出结果

|    | A                        | B      | C     | D        | E     | F                        | G       | H       | I       | J        | K     |
|----|--------------------------|--------|-------|----------|-------|--------------------------|---------|---------|---------|----------|-------|
| 1  |                          | Banks  | Bank  | Mortgage | Total |                          |         | Banks   | Bank    | Mortgage | Total |
| 2  |                          |        | Owned | Company  |       |                          |         |         | Owned   | Company  |       |
| 3  |                          |        |       |          |       |                          |         |         |         |          |       |
| 4  |                          |        |       |          |       |                          |         |         |         |          |       |
| 5  | < \$ 21,000              | 943    | 166   | 151      | 1260  | < \$ 21,000              | 726.435 | 176.972 | 356.593 | 1260     |       |
| 6  | \$ 21,000 to < \$ 33,000 | 2186   | 503   | 711      | 3400  | \$ 21,000 to < \$ 33,000 | 1960.22 | 477.543 | 962.236 | 3400     |       |
| 7  | \$ 33,000 to < \$ 42,000 | 708    | 211   | 589      | 1508  | \$ 33,000 to < \$ 42,000 | 869.416 | 211.804 | 426.78  | 1508     |       |
| 8  | \$ 42,000 to < \$ 50,000 | 424    | 150   | 397      | 971   | \$ 42,000 to < \$ 50,000 | 559.816 | 136.381 | 274.803 | 971      |       |
| 9  | \$ 50,000 +              | 948    | 239   | 709      | 1896  | \$ 50,000 +              | 1093.11 | 266.3   | 536.588 | 1896     |       |
| 10 |                          |        |       |          |       |                          |         |         |         |          |       |
| 11 | Total                    | 5209   | 1269  | 2557     | 9035  |                          |         |         |         |          |       |
| 12 |                          |        |       |          |       |                          |         |         |         |          |       |
| 13 | Chi-square statistic     | 534.48 |       |          |       |                          |         |         |         |          |       |
| 14 | p-value                  | 3E-110 |       |          |       |                          |         |         |         |          |       |
| 15 |                          |        |       |          |       |                          |         |         |         |          |       |
| 16 |                          | Banks  | Bank  | Mortgage | Total |                          |         |         |         |          |       |
| 17 |                          |        | Owned | Company  |       |                          |         |         |         |          |       |
| 18 | < \$ 21,000              | 78.84  | 13.17 | 11.98    | 100   |                          |         |         |         |          |       |
| 19 | \$ 21,000 to < \$ 33,000 | 64.29  | 14.79 | 20.91    | 100   |                          |         |         |         |          |       |
| 20 | \$ 33,000 to < \$ 42,000 | 46.95  | 13.99 | 39.06    | 100   |                          |         |         |         |          |       |
| 21 | \$ 42,000 to < \$ 50,000 | 43.67  | 15.45 | 40.89    | 100   |                          |         |         |         |          |       |
| 22 | \$ 50,000 +              | 50.00  | 12.61 | 37.39    | 100   |                          |         |         |         |          |       |
| 23 |                          |        |       |          |       |                          |         |         |         |          |       |
| 24 | Total                    | 57.65  | 14.05 | 28.3     | 100   |                          |         |         |         |          |       |

**解答** 首先,我们注意到抵押贷款贷方的类型是定性变量。同时,抵押数量被分成 5 组,所以,我们也应将其处理为定性变量。

为了检验统计相关性的存在,我们利用 *Excel* 输出结果中的  $\chi^2$  统计量。 $p$  值很小;由于,“E-110”表示“小数点向左移动 110 位”,因此, $p$  值为小数点后面跟 109 个 0,之后再跟一个 3。这就是说有更充分的证据显示存在相关性。

$p$  值小并不一定意味着强的相关性;它也可能是弱相关和大样本的结果。要评价相关性的强度,可查看输出左下角对应于每种贷款数量的百分比。有一个明显的模式,随着贷款数量的增加,抵押公司持有贷款份数的百分比增加,银行持有贷款份数的百分比下降。对于 50,000 美元收入水平而言,这种趋势被颠倒,但总的来说,这种趋势还保持得相当好。然而,百分比的变化不是很大,所以,我们并不能说相关性非常强。

相反,它仅仅是适度相关。

## 10.8 几率和优比

分析关于定性变量的计数数据的另一种方法就是利用几率(odds)。这种方法被广泛地应用于生物医学研究,也可能有助于分析市场调研方面的一些问题。几率的基本定义为一事件发生的概率与该事件不发生的概率之比。

**定义 10.4** 事件  $A$  的几率 =  $\frac{P(A)}{1 - P(A)}$

如果一事件发生的概率为  $\frac{2}{3}$ ,那么几率为  $\frac{2/3}{1/3} = 2$ 。通常将其表述为“事件发生的几率为 2 比 1”。几率通常被用在赛马和其他赌博项目中。赛马几率被定义为赛马不赢得比赛的几率。因此 4 比 1 的几率意味着赛马失去比赛(不赢)可能性是赢得比赛可能性的 4 倍。根据几率,比如说,和几率为 20 比 1 的赛马相比,几率为 4 比 1 的赛马是一匹更好的“赌马”。几率为 1 比 2(等价地,0.5 比 1)的赛马如何?这种马非常受欢迎,因为赢得比赛可能性是赛马失去比赛可能性的两倍(2 比 1)。

在分析几率时,只要确定感兴趣的事件是什么即可。同时,很容易将一事件的几率换算成该事件的概率。对于事件  $A$  而言

$$P(A) = \frac{\text{事件 } A \text{ 的几率}}{1 + \text{事件 } A \text{ 的几率}}$$

因此,如果一匹马(不赢)的几率为 9 比 1,那么该匹马不赢的概率为

$$\text{概率(不赢)} = \frac{9}{1 + 9} = 0.9$$



同样地, 赢的概率是 0.1。

利用几率能够方便地看出条件事件的发生如何改变一事件的概率。回想一下, 在第四章中, 给定另一事件  $B$ , 一事件  $A$  发生的条件概率为

$$P(A | B) = P(A \text{ 和 } B) / P(B)$$

经过简单的代数运算就可知道, 给定另一事件  $B$ , 有利于事件  $A$  发生的几率为

$$\frac{P(A | B)}{P(\text{非 } A | B)} = \frac{P(A)}{P(\text{非 } A)} \frac{P(B | A)}{P(B | \text{非 } A)}$$

即初始几率乘以似然比, 似然比就是  $A$  发生时条件事件的概率与  $A$  不发生时条件事件的概率之比。如果当  $A$  发生时比当  $A$  不发生时,  $B$  更有可能发生, 那么  $B$  的发生使得有利于  $A$  的几率上升。

### 例 10.17

考虑一人口总体, 其中每 1,000 人中有 1 人携带 HIV 病毒。现有一种检测方法, 携带病毒者被检测出阳性的占 95%, 不携带病毒者被(错误地)检测出阳性的占 2%。如果随机选取一人, 经检测为阳性, 那么, 携带 HIV 病毒者的几率是上升还是下降? 上升或下降多少?

**解答** 我们当然相信一个阳性检验结果会增加携带病毒的几率。如果一个阳性检验结果使得患病的机会减少, 那么这个检验一定是一个很奇怪的检验! 记事件  $A$  为“携带 HIV”, 事件  $B$  为“阳性检验结果”。

在检验之前, 随机抽取一个人, 他携带 HIV 的几率为

$$\frac{0.001}{0.999} \approx 0.001$$

阳性检验结果的发生使得几率改变为

$$\begin{aligned} \frac{P(\text{HIV} | \text{阳性})}{P(\text{不携带 HIV} | \text{阳性})} &= \frac{P(\text{HIV})}{P(\text{不携带 HIV})} \frac{P(\text{阳性} | \text{HIV})}{P(\text{阳性} | \text{不携带 HIV})} \\ &= \frac{0.001}{0.999} \frac{0.95}{0.02} = 0.0475 \end{aligned}$$

如果得到的是阳性检验结果, 那么携带 HIV 的几率从大约 0.001(比 1)上升到大约 0.0475(比 1)。

和几率密切相关, 且被广泛应用在生物医学研究中的一个概念就是**优比(odds ratio 几率比)**。顾名思义, 优比就是一个组(比如, 男人)中某事件(比如, 染上某种癌)的几率与另一组(比如, 女人)中同一事件的几率之比。优比通常用条件概率来定义, 但同样也可以用联合概率来表述。

**定义 10.5 对应于两组,某事件的优比**

如果  $A$  是任一事件,具有概率  $P(A|\text{组 1})$  和  $P(A|\text{组 2})$ ,则优比为

$$\frac{P(A|\text{组 1})/[1 - P(A|\text{组 1})]}{P(A|\text{组 2})/[1 - P(A|\text{组 2})]} = \frac{P(A \text{ 和组 1})/[1 - P(A \text{ 和组 1})]}{P(A \text{ 和组 2})/[1 - P(A \text{ 和组 2})]}$$

如果事件  $A$  在统计上独立于组,那么优比等于 1。

例如,就两种品牌的跑车而言,假定我们有下面关于购买某种悬挂装置的频数表:

|      | 是   | 否    | 总和    |
|------|-----|------|-------|
| 品牌 1 | 250 | 750  | 1,000 |
| 品牌 2 | 400 | 1600 | 2,000 |
| 总和   | 650 | 2350 | 3,000 |

我们将估计给定品牌的条件下,购买或不购买该悬挂装置的条件概率。这些概率为

|      | 是     | 否     | 总和    |
|------|-------|-------|-------|
| 品牌 1 | 0.250 | 0.750 | 1.000 |
| 品牌 2 | 0.200 | 0.800 | 1.000 |

优比  $(0.250/0.750)/(0.200/0.800) = 1.333$  表明为品牌 1 轿车购买该装置的几率较为品牌 2 轿车购买该装置的几率高 33.3%。购买悬挂装置似乎并不独立于品牌(虽然,我们不得不面对样本数据有限这样的事实)。同样,我们也可按如下方式计算优比

$$\frac{(250/3,000)/(750/3,000)}{(400/3,000)/(1,600/3,000)} = \frac{250/750}{400/1,600} = 1.333$$

通常用优比的自然对数对优比进行推断。前面曾提到过“ln”通常表示自然对数(以  $e=2.71828$  为底),且  $\ln(1)=0$ 。当根据次数数据来估计优比的自然对数时,它近似服从正态分布,期望值为总体优比的自然对数。其标准差为将表中 4 个次数的  $1/\text{次数}$  相加,然后开平方根。对于购买悬挂装置的数据,  $\ln(\text{优比}) = \ln(1.333) = 0.2874$ ;标准差估计值为

$$\sqrt{1/250 + 1/750 + 1/400 + 1/1,600} = 0.0920$$

有了这些结果,我们就能计算 95% 置信区间为

$$0.2874 - 1.96(0.0920) \leq \ln(\text{优比}) \leq 0.2874 + 1.96(0.0920)$$

$$\text{或} \quad 0.1071 \leq \ln(\text{优比}) \leq 0.4677$$

我们也可以用“反对数”——指数将优比置信区间转换为

$$e^{0.1071} \leq \text{优比} \leq e^{0.4677}$$

$$\text{或} \quad 1.113 \leq \text{优比} \leq 1.596$$

因为区间并不包含优比 1.000(或等价地,  $\ln(\text{优比})$  区间不包含 0.000), 所以, 我们可以得出结论: 数据显示了统计上可察觉的相关性。

优比是一种很有用的方法, 用以比较两总体比例  $\pi_1$  和  $\pi_2$ , 当  $\pi_1$  和  $\pi_2$  较小时, 优比方法比  $\pi_1$  和  $\pi_2$  的差  $\pi_1 - \pi_2$  更有意义。例如, 假定用复方 1 治疗的 5,000 例冠状旁通管(coronary bypass)病人, 其再梗率为  $\hat{\pi}_1 = 0.05$ , 用复方 2 治疗的 5,000 例冠状旁通管(coronary bypass)病人, 其再梗率为  $\hat{\pi}_2 = 0.02$ 。这样, 和优比相比, 它们的差  $\hat{\pi}_1 - \hat{\pi}_2 = 0.03$  就显得不太重要, 带来的信息也较少。见表 10.5。

|      | 再梗      |       |               |
|------|---------|-------|---------------|
|      | 是       | 否     |               |
| 复方 1 | 250(5%) | 4,750 | $n_1 = 5,000$ |
| 复方 2 | 100(2%) | 4,900 | $n_2 = 5,200$ |

复方 1 和复方 2 的再梗率几率如下:

$$\text{复方 1 几率} = \frac{250/5,000}{4,750/5,000} = \frac{250}{4,750} = 0.053$$

$$\text{复方 2 几率} = \frac{100/5,000}{4,900/5,000} = \frac{100}{4,900} = 0.020$$

相应的优比为  $0.053/0.020 = 2.65$ 。可见, 虽然再梗率之差仅为 0.03, 但是经过复方 1 治疗以后再梗率几率是经过复方 2 治疗以后再梗率几率的 2.65 倍。

## 10.9 小 结

本章, 我们处理的是类型数据。在很多情况下, 我们遇到的是单一变量的类型数据。首先, 我们研究单个总体比例  $\pi$  以及基于独立样本的两总体比例差  $\pi_1 - \pi_2$  的估计和检验方法。作为这些方法的推广, 为了比较总体的几个比例(两个以上)时, 我们引入了卡方拟合优度检验。

基于  $r \times c$  列联表数据, 我们利用卡方独立性检验和卡方齐性检验研究两变

量的类型数据问题。我们用  $\lambda$  度量两个类型变量之间相关性的强度。

最后,我们讨论了几率和优比,它们在涉及到二项比例的生物医学试验方面特别有用。

### 重要公式

1.  $\pi$  的置信区间  $\hat{\pi} \pm z_{\alpha/2} \sigma_{\hat{\pi}}$  其中  $\hat{\pi} = \frac{y}{n}$ ,  $\sigma_{\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$ ,

2. 形如  $\hat{\pi} \pm E$  的  $100(1-\alpha)\%$  置信区间所需样本容量

$$n = \frac{z_{\alpha/2}^2 \pi(1-\pi)}{E^2} \quad (\text{提示:如果没有合适的估计,取 } \pi = 0.5)$$

3.  $\pi$  的统计检验 T.S.:  $z = \frac{\hat{\pi} - \pi_0}{\sigma_{\hat{\pi}}}$  其中  $\sigma_{\hat{\pi}} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$

4.  $\pi_1 - \pi_2$  的置信区间  $\hat{\pi}_1 - \hat{\pi}_2 \pm z_{\alpha/2} \sigma_{\hat{\pi}_1 - \hat{\pi}_2}$

其中

$$\sigma_{\hat{\pi}_1 - \hat{\pi}_2} = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

5.  $\pi_1 - \pi_2$  的统计检验 T.S.:  $z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sigma_{\hat{\pi}_1 - \hat{\pi}_2}}$

其中

$$\sigma_{\hat{\pi}_1 - \hat{\pi}_2} = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

6. 多项分布  $P(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! n_2! \dots n_k!} \pi_1^{n_1} \pi_2^{n_2} \dots \pi_k^{n_k}$

7. 卡方拟合优度检验 T.S.:  $\chi^2 = \sum_i \left[ \frac{(n_i - E_i)^2}{E_i} \right]$

其中

$$E_i = n\pi_{i0}$$

8. 卡方独立性检验  $\chi^2 = \sum_{i,j} (n_{ij} - E_{ij})^2 / E_{ij}$

其中

$$E_{ij} = \frac{(\text{第 } i \text{ 行频数})(\text{第 } j \text{ 列频数})}{n}$$

9.  $\lambda = \frac{\text{自变量未知时预测错误的次数} - \text{自变量已知时预测错误的次数}}{\text{自变量未知时预测错误的次数}}$

10. 事件 A 的几率 =  $\frac{P(A)}{1-P(A)}$  (在二项情形, 成功的几率 =  $\frac{\pi}{1-\pi}$ )

11. 对于二项情形, 两总体优比

$$\frac{\text{总体 1 几率}}{\text{总体 2 几率}} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$$

## 补充练习

**10.63(社会)** 一位发言者建议经理如何避免工人成立工会,他声称只有25%的产业工人赞成成立工会,40%漠不关心,35%反对。另外,该发言者声称持有这些观点与他们是否为真正的工会会员无关。随机抽取600名产业工人,得到如下数据:

|     | 赞成  | 无所谓 | 反对  | 总和  |
|-----|-----|-----|-----|-----|
| 会员  | 140 | 42  | 18  | 200 |
| 非会员 | 70  | 198 | 132 | 400 |
| 总和  | 210 | 240 | 150 | 600 |

a. 数据中哪一部分与25%,40%,35%主张有关?

b. 给定  $\alpha = 0.01$ , 检验这一假设。

**10.64** 如何看练习10.63中的  $p$  值?

**10.65** 检验练习10.63中的独立性假设。它如何被最终拒绝?

**10.66** (根据练习10.63中的数据)计算赞成成立工会,漠不关心,反对的产业工人所占百分比;分别针对工会会员和非工会会员计算这些百分比。百分比揭示了会员身份和所持观点之间存在很强的相关性吗?

**10.67(民意调查学)** 三种不同的电视广告正在为同一产品作广告。将这些广告分别播放给消费者现场讨论小组;每名消费者只能看其中一种广告,然后陈述对该产品的观点。观点从1(很喜欢)到5(很不喜欢)。数据如下:

| 广告 | 年 龄 |     |     |     |     | 总和  |
|----|-----|-----|-----|-----|-----|-----|
|    | 1   | 2   | 3   | 4   | 5   |     |
| A  | 32  | 87  | 91  | 46  | 44  | 300 |
| B  | 53  | 141 | 76  | 20  | 10  | 300 |
| C  | 41  | 93  | 67  | 36  | 63  | 300 |
| 总和 | 126 | 321 | 234 | 102 | 117 | 900 |

a. 计算在独立性零假设下的期望频数。

b. 检验该假设时的自由度为多少?

c. 是否有证据表明对于不同的广告,其观点分布也不同? 给定  $\alpha = 0.01$ 。

**10.68** 给出练习10.67中  $p$  值的界。

**10.69** 根据你的判断, 练习 10.67 数据中广告类型和观点之间是否存在很强的相关性? 用百分比和  $\lambda$  值的计算支持你的判断。

**10.70(商业)** 某直邮零售商试验用三种不同的方式将定单表格合并到价目表中。在第一种类型价目表中, 定单表格位于价目表的最后; 第二种位于价目表的中间; 第三种中间和最后都有定单表格。将每份表寄给 1,000 名潜在的顾客, 他们都未曾从该零售商处买过产品。每份定单上的密码使得零售商知道它属于哪一种类型; 记录收到每种类型定单的数量。用 Excel 计算期望频数和  $\chi^2$  统计量。Excel 的 CHITEST 功能给出  $p$  值。结果如图 10.3 所示。

|    | A         | B       | C        | D         | E     | F | G                    | H |
|----|-----------|---------|----------|-----------|-------|---|----------------------|---|
| 1  | Observed  |         |          | Received? |       |   | max. freq.           |   |
| 2  |           |         | No       | Yes       | Total |   |                      |   |
| 3  |           | 1       | 944      | 56        | 1000  |   | 944                  |   |
| 4  | Type form | 2       | 961      | 39        | 1000  |   | 961                  |   |
| 5  |           | 3       | 915      | 85        | 1000  |   | 915                  |   |
| 6  |           |         | 2820     | 180       | 3000  |   | 2820                 |   |
| 7  |           |         |          |           |       |   | lambda (column dep.) |   |
| 8  | Expected  |         |          |           |       |   | 0                    |   |
| 9  |           |         | No       | Yes       | Total |   |                      |   |
| 10 |           | 1       | 940      | 60        | 1000  |   |                      |   |
| 11 | Type form | 2       | 940      | 60        | 1000  |   |                      |   |
| 12 |           | 3       | 940      | 60        | 1000  |   |                      |   |
| 13 |           |         | 2820     | 180       | 3000  |   |                      |   |
| 14 |           |         |          |           |       |   |                      |   |
| 15 |           | chi sq. | 19.18440 |           |       |   |                      |   |
| 16 |           | p-value | 0.00007  |           |       |   |                      |   |
| 17 |           |         |          |           |       |   |                      |   |

图 10.3 价目表试验结果

- 对于三种不同类型的定单, 统计独立性的零假设表明了什么?
- 在通常的  $\alpha$  水平下, 零假设能成立吗?

**10.71** 已知定单类型 TYPE FORM, 求预测收到数量 RECEIVED 的  $\lambda$  值。它捕捉到两变量之间的相关性吗?

**10.72(商业)** 某编程公司开发了一种更精心设计的、更复杂的电子制表软件程序版本。他们随机地将该程序的“beta 测试”拷贝寄给一部分当前的程序用户。根据用户提供的信息, 公司评价每位用户的复杂性; 1 表示标准, 2 表示程序的基本应用, 3 表示最复杂的应用。每位用户在当前版本和测试版本之间作出选择, 表明自己的偏好, 1 表示强烈偏好当前版本, 3 表示对两种版本没有特别偏好, 5 表示强烈偏好新的测试版本。用 JMP IN 分析数据。部分结果如下。

SOPHIST By PREFER

Crosstabs

SOPHIST PREFER

| Count | 1     | 2     | 3     | 4     | 5     | Row % |
|-------|-------|-------|-------|-------|-------|-------|
| 1     | 32    | 28    | 17    | 12    | 8     | 97    |
|       | 32.99 | 28.87 | 17.53 | 12.37 | 8.25  |       |
| 2     | 10    | 24    | 16    | 6     | 4     | 60    |
|       | 16.67 | 40.00 | 26.67 | 10.00 | 6.67  |       |
| 3     | 2     | 4     | 5     | 8     | 14    | 33    |
|       | 6.06  | 12.12 | 15.15 | 24.24 | 42.42 |       |
|       | 44    | 56    | 38    | 26    | 26    | 190   |

Tests

| Source      | DF  | - LogLikelihood | RSquare(U) |
|-------------|-----|-----------------|------------|
| Model       | 8   | 19.91046        | 0.1036     |
| Error       | 180 | 172.23173       |            |
| C Total     | 188 | 192.14219       |            |
| Total Count |     | 190             |            |

| Test             | ChiSquare | Prob> ChiSq |
|------------------|-----------|-------------|
| Likelihood Ratio | 39.821    | < 0.0001    |
| Pearson          | 44.543    | < 0.0001    |

- a. 结果标有百分数揭示了复杂性 SOPHIST 和偏好 PREFER 之间存在相关性吗? 如果数据未显示相关性, 那么, 真正的百分比应该怎样?

- b. 标有(PEARSON)CHI-SQUARE 的  $\chi^2$  值是否表明在通常的  $\alpha$  水平下存在统计上可察觉的(显著的)关系。

**10.73(商业)** 某录象机租赁商店抽取一部分顾客(大体上可视为随机样本)。商店最感兴趣的两个响应是顾客的租用频数以及顾客对商店货源的充分性所作的评价。响应数据储存在 Web 站点数据库的 EX1073.DAT 文件中,你应该将其下载到使用的软件包中。第一列包括租用频数的编码,1 为最低频数,4 为最高频数。同样地,第二列包括对商店货源充分性评价的编码,1 为最差,4 为最好。

- 对于每一个充分性评价类别,求频数。
- 检验零假设:每一类别出现的可能性相同。(很多软件包用来做拟合优度检验并不容易,所以你可能还需手工计算。)在  $\alpha = 0.10$  下,该假设能被拒绝吗?当  $\alpha = 0.01$  时怎么样?

**10.74** 利用练习 10.73 中的数据,检验零假设,即频数和货源充分性在统计上相互独立。对于该零假设, $p$  值说明了什么?

**10.75** 利用练习 10.73 中的数据,求在每一类充分性评价中顾客的百分比;分别对每个频数编码求评价的百分比。随着租用频数的增加,你能发现百分比有什么趋势吗?

**10.76(人力资源)** 某大银行主管福利的经理(从几千名职工中)抽取 353 名作为样本,以了解他们对两种医疗津贴计划的看法。感兴趣的变量为:年龄(分为 5 类,编码 1 表示最年轻的职工,5 为最年长的);看法(分为 5 类,1 为最喜欢卫生保健组织的方式,5 表示喜欢传统的医疗费一次一付的方式);根据职工是否有受抚养家属享受该计划将职工进行编码(如果没有,编为 0;如果有,编为 1)。响应数据(按顺序从第 1 到第 3 列)被列在 Web 站点数据库的 EX1076.DAT 文件中,将其下载到你的软件包中。

- 求对应于每一个年龄和看法组合的频数表。如果软件包能够做这种分析的话,对于每个年龄类别,求每种看法的百分比。对于不同的年龄类别,看法的百分比分布相同吗?
- 利用软件包检验零假设:年龄和看法相互独立。在  $\alpha = 0.05$  下,零假设能被拒绝吗?

**10.77** 练习 10.76 中主管福利的经理怀疑年龄和看法之间也许存在间接的关系:即职工的年龄与职工的受抚养家属是否享受津贴有关,而职工的受抚养家属是否享受津贴又与职工的看法有关。

- 利用软件包检验职工年龄和职工的受抚养家属是否享受津贴之间的相关性。相关性最终成立吗?
- 对于职工受抚养家属是否享受津贴与职工的看法,作同样分析。
- 利用软件包检验职工年龄和职工看法之间的相关性。对于有受抚养家属



享受津贴的职工和没有受抚养家属享受津贴的职工,分别利用软件包检验他们的年龄和看法之间的相关性。在这些检验中有一些相关性的证据吗?

**10.78(生物)** 一项致癌性研究旨在考察一种准备试用于人身上的药物是否有可能导致肿瘤。为此,总共用 300 只(150 只雄性,150 只雌性)老鼠进行为期 6 个月的试验。开始时,随机地将 100 只(50 只雄性,50 只雌性)分配到控制组,100 只分配到低剂量药物组,剩下的 100 只(50 只雄性,50 只雌性)分配到高剂量药物组。在 6 个月期间,每天给控制组注射一次惰性溶液,而给药物组注射一次掺有药物的溶液。样本数据如下表。

|     | 肿瘤数目  |    |
|-----|-------|----|
|     | 一个或更多 | 无  |
| 控制组 | 10    | 90 |
| 低剂量 | 14    | 86 |
| 高剂量 | 19    | 81 |

- 对于三个处理组中的每一组,给出患有一个或多个肿瘤的老鼠的百分比。
- 给定  $\alpha = 0.05$ , 检验对于三个处理组患有一个或多个肿瘤的老鼠的百分比分布是否存在显著差异。
- 就这种药物来说,是否存在与肿瘤有关的药物问题? 也就是说,随着药物剂量的增加,患有肿瘤的老鼠的比例是否增加?

**10.79** 下面是针对练习 10.78 中数据的 SAS 计算机输出结果。将其与你在练习 10.78 中所得的结果进行比较。

| RAT_GRP         | N_TUMORS |          |       |
|-----------------|----------|----------|-------|
| Frequency       |          |          |       |
| Expected        |          |          |       |
| Cell Chi-Square |          |          |       |
| Percent         |          |          |       |
| Row Pct         |          |          |       |
| Col Pct         | NONE     | ONE-MORE | Total |

|          |        |        |        |
|----------|--------|--------|--------|
| CONTROL  | 90     | 10     | 100    |
|          | 85.667 | 14.333 |        |
|          | 0.2192 | 1.3101 |        |
|          | 30.00  | 3.33   | 33.33  |
|          | 90.00  | 10.00  |        |
|          | 35.02  | 23.26  |        |
| HIGHDOSE | 81     | 19     | 100    |
|          | 85.667 | 14.333 |        |
|          | 0.2542 | 1.5194 |        |
|          | 27.00  | 6.33   | 33.33  |
|          | 81.00  | 19.00  |        |
|          | 31.52  | 44.19  |        |
| LOWDOSE  | 86     | 14     | 100    |
|          | 85.667 | 14.333 |        |
|          | 0.0013 | 0.0078 |        |
|          | 28.67  | 4.67   | 33.33  |
|          | 86.00  | 14.00  |        |
|          | 33.46  | 32.56  |        |
| Total    | 257    | 43     | 300    |
|          | 85.67  | 14.33  | 100.00 |

STATISTICS FOR TABLE OF RAT\_GRP BY N\_TUMORS

| Statistic                   | DF | Value | Prob  |
|-----------------------------|----|-------|-------|
| Chi-Square                  | 2  | 3.312 | 0.191 |
| Likelihood Ratio Chi-Square | 2  | 3.327 | 0.189 |
| Mantel-Haenszel Chi-Square  | 1  | 0.649 | 0.420 |
| Phi Coefficient             |    | 0.105 |       |
| Contingency Coefficient     |    | 0.104 |       |

Cramer's V

0.105

Sample Size = 300

**10.80(社会)** 一项社会研究旨在考察蓝领工人呆在第一份工作上的时间与他们所受的教育之间是否存在一定的关系。根据工会会员的记录,随机抽取一部分工人,将其分类。数据如下。

| 第一份工作上的<br>的年数 | 受教育的年数 |       |        |      |
|----------------|--------|-------|--------|------|
|                | 0~4.5  | 4.5~9 | 9~13.5 | 13.5 |
| 0~2.5          | 5      | 21    | 30     | 33   |
| 2.5~5          | 15     | 35    | 40     | 30   |
| 5~7.5          | 22     | 16    | 15     | 30   |
| 7.5            | 28     | 10    | 8      | 10   |

- 利用下面的计算机输出结果,挑出期望单元数。
- 检验研究假设,即变量“呆在第一份工作上的时间”与变量“受教育程度”相关。
- 给出该检验的显著性水平。
- 用  $\alpha = 0.05$  得出你的结论。

YRS \_ JOB

YRS \_ EDU

|                           |  |        |        |        |        |       |
|---------------------------|--|--------|--------|--------|--------|-------|
| Frequency                 |  |        |        |        |        |       |
| Expected                  |  |        |        |        |        |       |
| Cell Chi-Square           |  |        |        |        |        |       |
| Percent                   |  |        |        |        |        |       |
| Row Pct                   |  |        |        |        |        |       |
| Col Pct                   |  | 0~4.5  | 13.5   | 4.5~9  | 9~13.5 | Total |
| +-----+-----+-----+-----+ |  |        |        |        |        |       |
| 0~2.5                     |  | 5      | 33     | 21     | 30     | 89    |
|                           |  | 17.902 | 26.342 | 20.971 | 23.784 |       |

|           |        |        |        |        |        |
|-----------|--------|--------|--------|--------|--------|
|           | 9.2988 | 1.6829 | 394E-7 | 1.6243 |        |
|           | 1.44   | 9.48   | 6.03   | 8.62   | 25.57  |
|           | 5.62   | 37.08  | 23.60  | 33.71  |        |
|           | 7.14   | 32.04  | 25.61  | 32.26  |        |
| + + + + + |        |        |        |        |        |
| 2.5~5     | 15     | 30     | 35     | 40     | 120    |
|           | 24.138 | 35.517 | 28.276 | 32.069 |        |
|           | 3.4594 | 0.857  | 1.599  | 1.9614 |        |
|           | 4.31   | 8.62   | 10.06  | 11.49  | 34.48  |
|           | 12.50  | 25.00  | 29.17  | 33.33  |        |
|           | 21.43  | 29.13  | 42.68  | 43.01  |        |
| + + + + + |        |        |        |        |        |
| 5~7.5     | 22     | 30     | 16     | 15     | 83     |
|           | 16.695 | 24.566 | 19.557 | 22.181 |        |
|           | 1.6854 | 1.202  | 0.6471 | 2.3248 |        |
|           | 6.32   | 8.62   | 4.60   | 4.31   | 23.85  |
|           | 26.51  | 36.14  | 19.28  | 18.07  |        |
|           | 31.43  | 29.13  | 19.51  | 16.13  |        |
| + + + + + |        |        |        |        |        |
| 7.5       | 28     | 10     | 10     | 8      | 56     |
|           | 11.264 | 16.575 | 13.195 | 14.966 |        |
|           | 24.864 | 2.608  | 0.7738 | 3.242  |        |
|           | 8.05   | 2.87   | 2.87   | 2.30   | 16.09  |
|           | 50.00  | 17.86  | 17.86  | 14.29  |        |
|           | 40.00  | 9.71   | 12.20  | 8.60   |        |
| + + + + + |        |        |        |        |        |
| Total     | 70     | 103    | 82     | 93     | 348    |
|           | 20.11  | 29.60  | 23.56  | 26.72  | 100.00 |

STATISTICS FOR TABLE OF YRS \_ JOB BY YRS \_ EDU

|           |    |       |      |
|-----------|----|-------|------|
| Statistic | DF | Value | Prob |
|-----------|----|-------|------|

-----

|                             |   |        |       |
|-----------------------------|---|--------|-------|
| Chi-Square                  | 9 | 57.830 | 0.001 |
| Likelihood Ratio Chi-Square | 9 | 55.605 | 0.001 |
| Mantel-Haenszel Chi-Square  | 1 | 31.376 | 0.001 |
| Phi Coefficient             |   | 0.408  |       |
| Contingency Coefficient     |   | 0.377  |       |
| Cramer's V                  |   | 0.235  |       |

Sample Size = 348

**10.81(心理学)** Johns Hopkins 大学的两位研究人员研究年长者的用药情况。在最近一次研究中,病人被要求回答,就他们的药物治疗情况而言,医生给他们提供了什么样的建议。研究者发现:

- 25.4% 病人说医生并没有告诉他们药物起什么作用。
  - 91.6% 表示他们并未被告知药物可能如何“烦扰”他们。
  - 47.1% 表示医生并没有问他们在治疗开始之后,药物对他们如何“有用”或如何“烦扰”他们。
  - 87.7% 表示在和医生讨论了治疗如何有帮助或如何烦扰他们之后,药并没有改变。
- a. 假定在研究中,共接待了 500 位病人。用 95% 置信区间总结上述每一个结果。
- b. 对这些结果的有效性,你作何评论?

**10.82(医学)** 40 岁以上的人比较注意他们的消化系统,因为消化系统决定了他们吃什么,吃多少。一项研究旨在考察我们这个社会不同种族的人是否都有这种倾向。经随机抽样得到盎格鲁撒克逊人,德裔,拉美裔,意大利裔,西班牙裔以及非洲裔美国人的样本。抽样数据如下:

| 种族     | 响应人数<br>每组联系了 60 个人 | 消化系统不正常的人数 |
|--------|---------------------|------------|
| 盎格鲁撒克逊 | 55                  | 7          |
| 德裔     | 58                  | 6          |
| 拉美     | 52                  | 34         |
| 意大利    | 54                  | 38         |
| 西班牙    | 30                  | 20         |
| 非洲     | 49                  | 31         |

a. 是否存在因响应比率而产生的偏离?

b. 用 95% 置信区间比较盎格鲁撒克逊人和德裔美国人的比率, 即相应的  $\pi_i$ 。

**10.83** 参看练习 10.82。似乎有两种不同的比率——一种围绕着 12%, 另一种围绕着 70%。合并前两组样本数据, 同时也合并后四组样本数据。利用这些数据检验假设  $H_0: \pi_1 - \pi_2 \geq 0$   $H_a: \pi_1 - \pi_2 < 0$ , 这里  $\pi_1$  对应于第一组混合样本的比率,  $\pi_2$  对应于第二组混合样本的比率。给出检验的  $p$  值。

**10.84(商业)** 随机地从某期刊出版社未经阅读的毛条校样中抽取 40 页作为样本, 下面列出了每页错误的频数数据。

| 错误/页 | 观察频数 |
|------|------|
| 0    | 5    |
| 1    | 9    |
| 2    | 5    |
| 3    | 7    |
| 4    | 4    |
| 5    | 2    |
| 6    | 3    |
| 7    | 2    |
| 8    | 1    |
| 9    | 0    |
| 10   | 2    |

给定  $\alpha = 0.10$ , 试检验每页错误数是否服从均值为 3.2 的 Poisson 分布。

**10.85(园艺)** 一位昆虫学家对研究大批滋生在密执根果园苹果树上的成年欧洲红螨虫很感兴趣。她从果园中 10 棵相同的苹果树上分别随机地选取 50 片树叶, 对它们进行检测, 并记录下 500 片树叶中每片树叶上的螨虫数量。作为大型研究的一部分, 她要模拟该果园苹果树上螨虫数量的分布。基于所得到的数据, 对于苹果树上欧洲红螨虫的数量, Poisson 分布似乎是一个合理的模型吗?

|          |     |     |    |    |    |    |   |   |
|----------|-----|-----|----|----|----|----|---|---|
| 每片树叶的螨虫数 | 0   | 1   | 2  | 3  | 4  | 5  | 6 | 7 |
| 频率       | 233 | 127 | 57 | 33 | 30 | 10 | 7 | 3 |

**10.86** 参见案例。分析数据, 并回答总经理的问题, 应包括置信区间, 假设检验, 以及任何相关的图表。计算机输出结果如下。

# ANALYSIS OF CASE STUDY

## TABLE OF BUILDING RATINGS BY GROUP

| BUILDING RATINGS |        | GROUP  |        |        |        |
|------------------|--------|--------|--------|--------|--------|
| Frequency        |        |        |        |        |        |
| Expected         |        |        |        |        |        |
| Cell Chi-Square  |        |        |        |        |        |
| Percent          |        |        |        |        |        |
| Row Pct          |        |        |        |        |        |
| Col Pct          | G1     | G2     | G3     | G4     | Total  |
| 1                | 11     | 8      | 15     | 6      | 40     |
|                  | 9.8413 | 10.635 | 9.2857 | 10.238 |        |
|                  | 0.1364 | 0.6528 | 3.5165 | 1.7544 |        |
|                  | 2.18   | 1.59   | 2.98   | 1.19   | 7.94   |
|                  | 27.50  | 20.00  | 37.50  | 15.00  |        |
|                  | 8.87   | 5.97   | 12.82  | 4.65   |        |
| 2                | 10     | 6      | 18     | 5      | 39     |
|                  | 9.5952 | 10.369 | 9.0536 | 9.9821 |        |
|                  | 0.0171 | 1.8409 | 8.8406 | 2.4866 |        |
|                  | 1.98   | 1.19   | 3.57   | 9.99   | 7.74   |
|                  | 25.64  | 15.38  | 46.15  | 12.82  |        |
|                  | 8.06   | 4.48   | 15.38  | 3.88   |        |
| 3                | 51     | 50     | 42     | 38     | 181    |
|                  | 44.532 | 48.123 | 42.018 | 46.327 |        |
|                  | 0.9395 | 0.0732 | 759E-8 | 1.4969 |        |
|                  | 10.12  | 9.92   | 8.33   | 7.54   | 35.91  |
|                  | 28.18  | 27.62  | 23.20  | 20.99  |        |
|                  | 41.13  | 37.31  | 35.90  | 29.46  |        |
| 4                | 30     | 41     | 26     | 40     | 137    |
|                  | 33.706 | 36.425 | 31.804 | 35.065 |        |
|                  | 0.4076 | 0.5747 | 1.059  | 0.6944 |        |
|                  | 5.95   | 8.13   | 5.16   | 7.94   | 27.18  |
|                  | 21.90  | 29.93  | 18.98  | 29.20  |        |
|                  | 24.19  | 30.60  | 22.22  | 31.01  |        |
| 5                | 22     | 29     | 16     | 40     | 107    |
|                  | 26.325 | 28.448 | 24.839 | 27.387 |        |
|                  | 0.7107 | 0.0107 | 3.1455 | 5.809  |        |
|                  | 4.37   | 5.75   | 3.17   | 7.94   | 21.23  |
|                  | 20.56  | 27.10  | 14.95  | 37.38  |        |
|                  | 17.74  | 21.64  | 13.68  | 31.01  |        |
| Total            | 124    | 134    | 117    | 129    | 504    |
|                  | 24.60  | 26.59  | 23.21  | 25.60  | 100.00 |

## STATISTICS FOR TABLE OF BUILDING RATINGS BY GROUP

| Statistic                   | DF | Value  | Prob  |
|-----------------------------|----|--------|-------|
| Chi-Square                  | 12 | 34.167 | 0.001 |
| Likelihood Ratio Chi-Square | 12 | 32.737 | 0.001 |
| Mantel-Haenszel Chi-Square  | 1  | 4.139  | 0.042 |
| Phi Coefficient             |    | 0.260  |       |
| Contingency Coefficient     |    | 0.252  |       |
| Cramer's V                  |    | 0.150  |       |
| Sample Size = 504           |    |        |       |

## ANALYSIS OF CASE STUDY

TABLE OF SERVICE RATINGS BY GROUP

| SERVICE RATINGS |        | GROUP  |        |        |        |
|-----------------|--------|--------|--------|--------|--------|
| Frequency       |        |        |        |        |        |
| Expected        |        |        |        |        |        |
| Cell Chi-Square |        |        |        |        |        |
| Percent         |        |        |        |        |        |
| Row Pct         |        |        |        |        |        |
| Col Pct         | G1     | G2     | G3     | G4     | Total  |
| 1               | 11     | 8      | 15     | 6      | 40     |
|                 | 9.8413 | 10.635 | 9.2857 | 10.238 |        |
|                 | 0.1364 | 0.6528 | 3.5165 | 1.7544 |        |
|                 | 2.18   | 1.59   | 2.98   | 1.19   | 7.94   |
|                 | 27.50  | 20.00  | 37.50  | 15.00  |        |
|                 | 8.87   | 5.97   | 12.82  | 4.65   |        |
| 2               | 10     | 6      | 18     | 5      | 39     |
|                 | 9.5952 | 10.369 | 9.0536 | 9.9821 |        |
|                 | 0.0171 | 1.8409 | 8.8406 | 2.4866 |        |
|                 | 1.98   | 1.19   | 3.57   | 0.99   | 7.74   |
|                 | 25.64  | 15.38  | 46.15  | 12.82  |        |
|                 | 8.06   | 4.48   | 15.38  | 3.88   |        |
| 3               | 51     | 50     | 42     | 38     | 181    |
|                 | 44.532 | 48.123 | 42.018 | 46.327 |        |
|                 | 0.9395 | 0.0732 | 759E-8 | 1.4969 |        |
|                 | 10.12  | 9.92   | 8.33   | 7.54   | 35.91  |
|                 | 28.18  | 27.62  | 23.20  | 20.99  |        |
|                 | 41.13  | 37.31  | 35.90  | 29.46  |        |
| 4               | 30     | 41     | 26     | 40     | 137    |
|                 | 33.706 | 36.425 | 31.804 | 35.065 |        |
|                 | 0.4076 | 0.5747 | 1.059  | 0.6944 |        |
|                 | 5.95   | 8.13   | 5.16   | 7.94   | 27.18  |
|                 | 21.90  | 29.93  | 18.98  | 29.20  |        |
|                 | 24.19  | 30.60  | 22.22  | 31.01  |        |
| 5               | 22     | 29     | 16     | 40     | 107    |
|                 | 26.325 | 28.448 | 24.839 | 27.387 |        |
|                 | 0.7107 | 0.0107 | 3.1455 | 5.809  |        |
|                 | 4.37   | 5.75   | 3.17   | 7.94   | 21.23  |
|                 | 20.56  | 27.10  | 14.95  | 37.38  |        |
|                 | 17.74  | 21.64  | 13.68  | 31.01  |        |
| Total           | 124    | 134    | 117    | 129    | 504    |
|                 | 24.60  | 26.59  | 23.21  | 25.60  | 100.00 |

STATISTICS FOR TABLE OF BUILDING RATINGS BY GROUP

| Statistic                   | DF | Value  | Prob  |
|-----------------------------|----|--------|-------|
| Chi-Square                  | 12 | 34.167 | 0.001 |
| Likelihood Ratio Chi-Square | 12 | 32.737 | 0.001 |
| Mantel-Haenszel Chi-Square  | 1  | 4.139  | 0.042 |
| Phi Coefficient             |    | 0.260  |       |
| Contingency Coefficient     |    | 0.252  |       |
| Cramer's V                  |    | 0.150  |       |
| Sample Size = 504           |    |        |       |

10.87 参见案例。写一份非技术性报告,解释你的分析揭示了什么。





## 第六部分 数据分析： 回归方法和模型的建立

第十一章 线性回归和相关

第十二章 多元回归与一般线性模型

第十三章 多元回归续论



# 第十一章 线性回归和相关

- 11.1 引言和案例
- 11.2 估计模型中的参数
- 11.3 回归参数的推断
- 11.4 利用回归预测新的  $y$  值
- 11.5 线性回归中拟合不足的考察
- 11.6 逆回归问题(校准)
- 11.7 相关
- 11.8 小结

## 11.1 引言和案例

预测一个变量未来的值是重要的管理活动。财务官员必须预测未来现金流量,生产经理必须预测原材料的需求,人事经理必须预测未来职员的需求。解释过去的变化也是重要的。解释顾客数量过去的变化能够帮助经理了解对社会服务机构的的需求。找出可以解释某个汽车零部件规格限变异的变量能够帮助企业改进这个零部件的质量。回归分析的基本思想是利用一个定量的自变量的数据预测或解释一个定量的因变量。

我们可以区分预测(关于未来的值)和解释(关于现在或过去的值)。由于事后估计的原因,解释比预测更容易。但经常会用词“预测”来概括这两种情况,因此,这本书里我们有时会对预测和解释不加区别。

为了使预测(或解释)有意义,被预测的变量(因变量)和用来预测的变量(自变量)之间必须存在某种联系。毫无疑问,如果你试了足够多次,可能发现 28 种股价在一年中变化的公众股票完全能够被棒球联盟的 28 个主要球队在 7 月 4 日比赛的输赢比例来预测,但是,这样的预测是荒谬的,因为这二者之间毫无关联。预测需要一种**关联单元**,代表这两种变量之间存在关系的本质。对于时间序列数据,很简单关联单元就是时间。变量可以在同一时期观测,或为了做到真正地估计,自变量的观测比因变量的观测早一个时期。对于横断面数据(cross-sectional data),变量之间应该存在一个经济上或物理上的本质联系。假如我们打算预测不同的软饮料的市场份额的变化,那么就应该考虑到这些饮料的推销宣传活动,而不是各种品牌的实心面条调味汁的广告宣传。预测时需要关联单元似乎是显然的,但许多预测是在没有明显的关联单元的环境中进行的。

本章我们介绍简单线性回归分析,这种方法适用于一个自变量,并且预测因变

量  $y$  的方程是给定的自变量  $x$  的线性函数。例如, 假如某县公路管理部门主任想预测用于投标的重新铺路的合同的成本, 我们可以合理地预见, 这个成本是将要重新铺设的公路英里数的函数。一个合理的开始是利用线性的预测函数。令  $y$  = 这个项目总的成本(千美元),  $x$  = 需要重新铺设的公路的英里数,  $\hat{y}$  = 预测的这个项目总的成本(千美元)。预测方程  $\hat{y} = 2.0 + 3.0x$  (例如) 就是线性方程。常数项, 比如 2.0, 是截距, 可解释为当  $x=0$  时  $y$  的预测值。在公路重新铺设的例子中, 我们可以把截距解释为这个项目开始时的固定成本。 $x$  的系数, 比如 3.0, 是这条直线的斜率, 就是当  $x$  改变一个单位时  $y$  的预测变化。在公路重新铺设的例子中, 假设有两个项目相差一英里, 我们可以预测长项目的成本比短项目多 3(千美元)。一般地, 预测方程表示为

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

其中  $\hat{\beta}_0$  是截距, 而  $\hat{\beta}_1$  是斜率, 参看图 11.1。

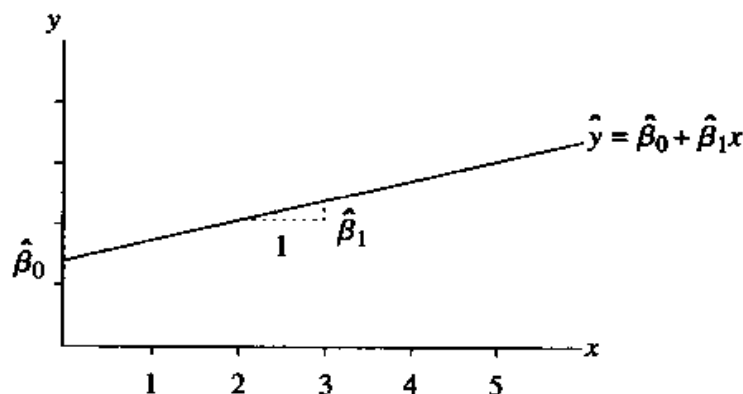


图 11.1 线性预测函数

简单线性回归的基本思想是用数据拟合与因变量  $y$  和单个自变量  $x$  相关的预测直线。简单回归的第一个假定是这两个变量的关系是线性的。按照直线性的假设, 当  $x$  变化时方程的斜率不改变。在公路重新铺设的例子中, 我们可以假定从长距离项目中不会获得(实质的)规模经济。除非直线性假定成立(至少大致), 否则利用简单线性回归就没有多大意义了。

表面上, 直线性并不总是合理的假定。例如, 我们打算用  $x$  = 某汽车商的电台广告的重复次数, 预测  $y$  = 知道这家汽车商的盛夏打折销售的司机的数量, 直线性的假定意味着商业广告的第一次播出不会比第一千零一次播出导致更多的司机知道此事。(你已经听过上述那样的商业广告)我们强烈怀疑这个假定在  $x$  的广泛范围内的有效性。我们很清楚商业广告重复的次数越多, 其影响越小, 因此一个直线的预测不很好。

假定存在直线性, 我们将  $y$  表示为  $x$  的线性函数:  $y = \beta_0 + \beta_1 x$ 。但是, 按照这

个方程,  $y$  是  $x$  的确切的线性函数; 没有留下余地考虑不可避免的误差( $y$  的实际值与预测值的差异)。为此, 对应每一个  $y$  我们引入一个随机误差项  $\epsilon_i$ , 并假定模型为

$$y = \beta_0 + \beta_1 x + \epsilon$$

我们假定随机变量  $y$  是由可预测的部分( $x$  的线性函数)和不可预测的部分(随机误差  $\epsilon_i$ )组成。系数  $\beta_0$  和  $\beta_1$  可解释为真实的截距和斜率, 而误差项  $\epsilon$  包含所有的其他已知和未知的因素的影响。在公路重新铺设的项目中, 一些不可预测的因素比如罢工、天气条件和设备故障的影响都包含在  $\epsilon$ ,  $\epsilon$  还包含一些因素比如公路的陡峭或维修前的条件——那些应该在预测中考虑的但最终没有考虑到的因素。不可预测的因素和被忽略的因素的综合影响就构成了随机误差项  $\epsilon$ 。

例如利用不同的新汽车本身的重量(自变量)预测这些新汽车所消耗汽油的平均里程(因变量), 一个方法是每辆车安排不同的司机行驶一个月时间。预测误差可能是由哪些不可预测和忽略的因素产生呢? 这个研究中不可预测(随机)因素包括司机的驾驶习惯和技术、行驶路程的类型(城市与公路)和途中所遇红灯的次数。可能被忽略的因素有发动机的规格和变速装置的类型(手动与自动)。

在回归分析的研究中, 自变量的值( $x_i$  的值)一般看作预先确定的常数, 因此随机性的惟一来源就是  $\epsilon_i$  项。尽管绝大多数经济和商务应用中  $x_i$  的值都是固定的, 但并不总是这样的。例如,  $x_i$  代表某个才能测试中申请者的得分,  $y_i$  代表申请者的生产能力, 如果数据是申请者的随机样本, 那么  $x_i$  ( $y_i$ ) 是随机变量。在回归分析研究中把  $x$  看作固定的还是随机的无关紧要, 如果  $x_i$  都是随机的, 我们可以简单地把所有概率陈述视为在  $x_i$  取定观测值的条件下的相应的概率陈述。

当我们假定所有的  $x_i$  都是常数, 那么关于  $x_i$  的模型中惟一的随机部分是随机误差项  $\epsilon_i$ 。以下是模型的正式的假设条件。

#### 定义 11.1 回归分析的正式的假定:

1. 所有误差的期望值都是零, 即对所有的  $i$ ,  $E(\epsilon_i) = 0$ 。
2. 所有误差的方差都是相同的, 即对所有的  $i$ ,  $\text{Var}(\epsilon_i) = \sigma_\epsilon^2$ 。
3. 误差是相互独立的。
4. 所有误差都服从正态分布, 即对所有的  $i$ ,  $\epsilon_i$  都是服从正态分布。

这些假定都显示在图 11.2 中。因变量的实际值服从正态分布, 其均值都落在回归线上且对自变量所有的值都具有相同的标准差。惟一未在图中显示的假定是观测值彼此的独立性。

有了这些正式的假定, 就可以导出以后的显著性检验和预测方法。我们一开始可以通过观察数据的散点图来检验模型的这些假定。散点图就是简单地画出每对  $(x_i, y_i)$  的坐标, 其中因变量作为纵轴。看看这些点是不是基本落在一条直线的周围或者是一条确定的曲线模式。除此之外, 我们还观察是否有明显地远离数据

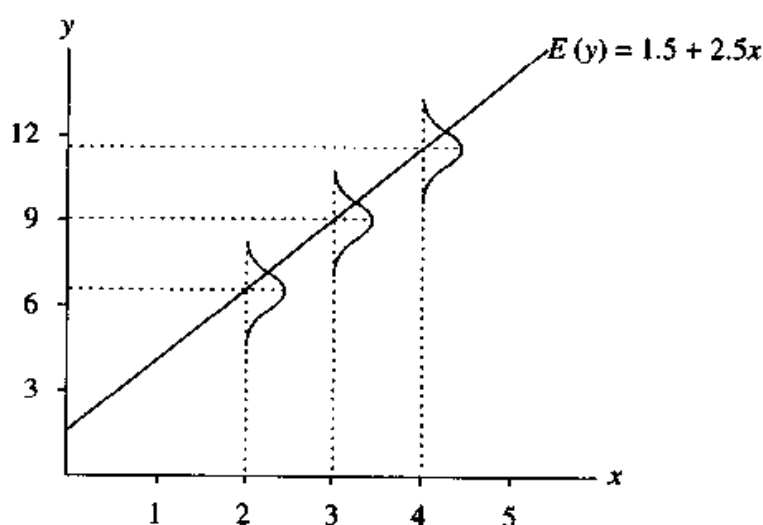


图 11.2 回归分析中  $y$  的理论分布

的一般模式的离群值。图 11.3(a)显示了一个散点图。

最近,平滑法被用来直接通过数据描绘出一条曲线而不必要假定任何特定的模型。假如这样的平滑法产生的曲线接近直线,那线性回归就是合理的。LOWESS(Locally weighted scatterplot smoother)就是一个这样的方法。粗略地讲,平滑法就是将数据沿  $x$  轴取一个非常窄小的“小段”,在小段上计算出拟合数据的直线,再将这个小段沿  $x$  轴慢慢移动,重新计算直线,如此重复下去。最后将这些所有的小直线连接成一条平滑的曲线。小段的宽度称为**带宽**;一般是由进行平滑的计算机软件控制的。简单的散点图(图 11.3(a))和通过这些数据的 LOWESS 曲线显示在图 11.3(b)中。散点图表明是曲线关系;LOWESS 曲线印证了这一点。

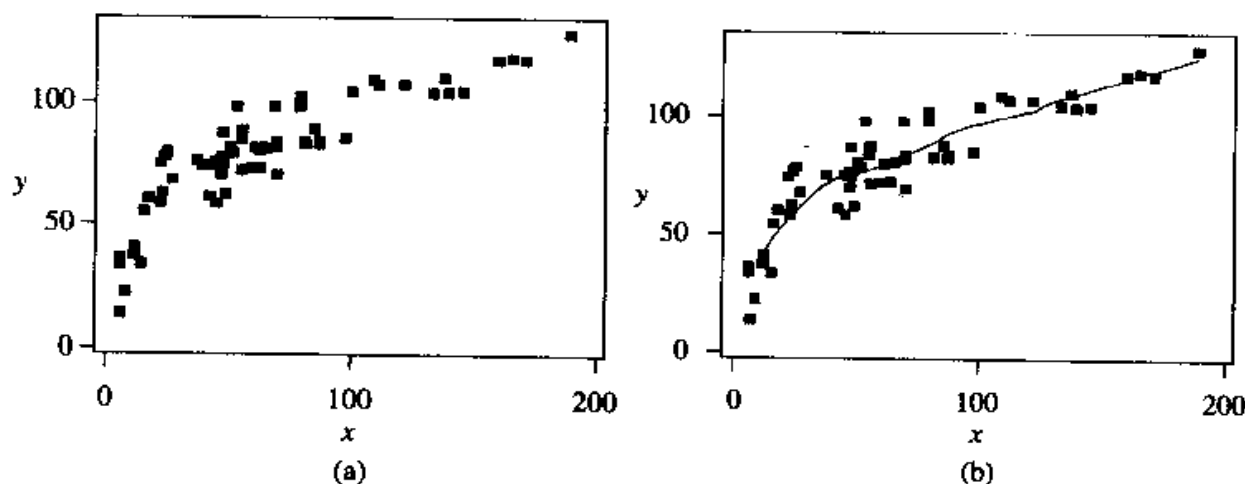


图 11.3 (a) 散点图和(b) LOWESS 曲线

另一种散点平滑是**样条拟合**。这个方法是取一个窄小的数据小段,在上面拟合出一条曲线(一般是三次方程),再移动到下一个小段拟合出另一条曲线,如此重复下去。最后将这些曲线连接成一条连续的曲线。

许多经济关系都不是线性的。例如,任何回报递减模式都趋向于产生一种增加的关系,只是其增加率是逐渐减少的。如果散点图本身或拟合的 LOWESS 曲线没有表现出线性,那么我们一般是通过自变量或因变量做变换将二者的关系直线化。一个好的统计软件或表格算法都可以计算像平方根这样的函数。变换后的变量可以简单看作另一个变量。

例如,某个大城市每个春季都要派遣职员修补街道的坑洞。每天所派遣的职员人数和修补的坑洞数都被记录下来,修补的坑洞数和职员数的散点图和带有 LOWESS 曲线的散点图显示在图 11.4 中。二者的关系是非线性的,就算没有 LOWESS 曲线,斜率下降也是明显的。这并不奇怪,因为派遣职员越多,雇佣的工人工作效率越低,这些职员不得不走得更远去发现坑洞,如此下去。所有这些原因都显示会出现回报递减现象。

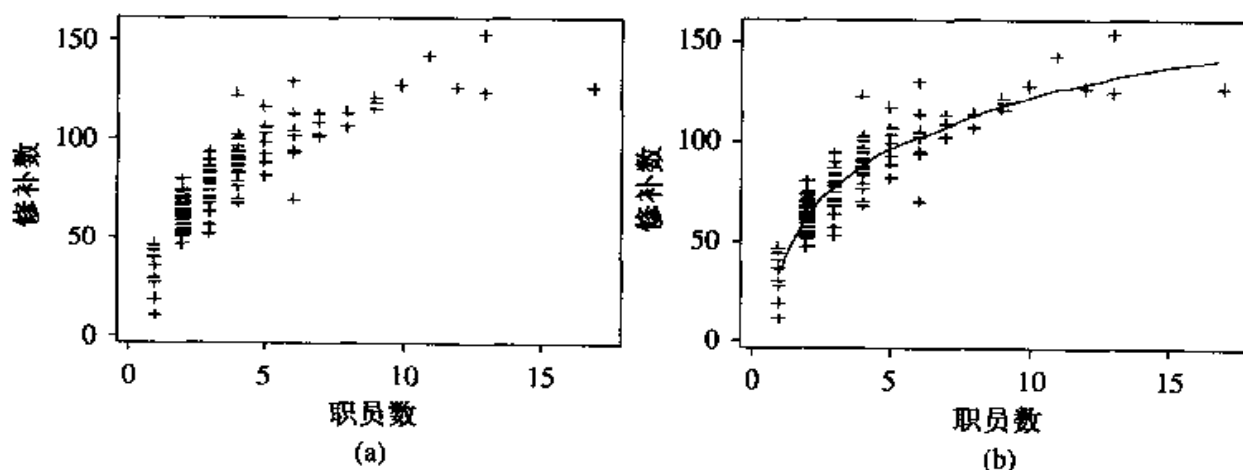


图 11.4 街道坑洞数据的散点图

我们可以通过尝试几种自变量变换的方法来寻找更接近线性的散点图。三种常用的变换是平方根变换、自然对数变换和倒数变换(1 除以变量)。将每种变换用于修补坑洞数的数据。变换后的散点图和相应的 LOWESS 曲线显示在图 11.5a—c 中。平方根变换(a)和倒数变换(c)并没有得出直线关系,而自然对数变换(b)就得出了非常好的结果。因此,我们将用  $\ln(\text{职员数})$  作为自变量。

寻找一个好的变换需要不断地尝试,也会犯不少错误。下面是一些帮助找出合适变换的建议。在散点图中寻找变换要注意两个方面。首先,二者的关系是非线性的吗?其次,沿着  $y$ (纵)轴是否有变差逐渐增大的趋势?如果有这种趋势,那



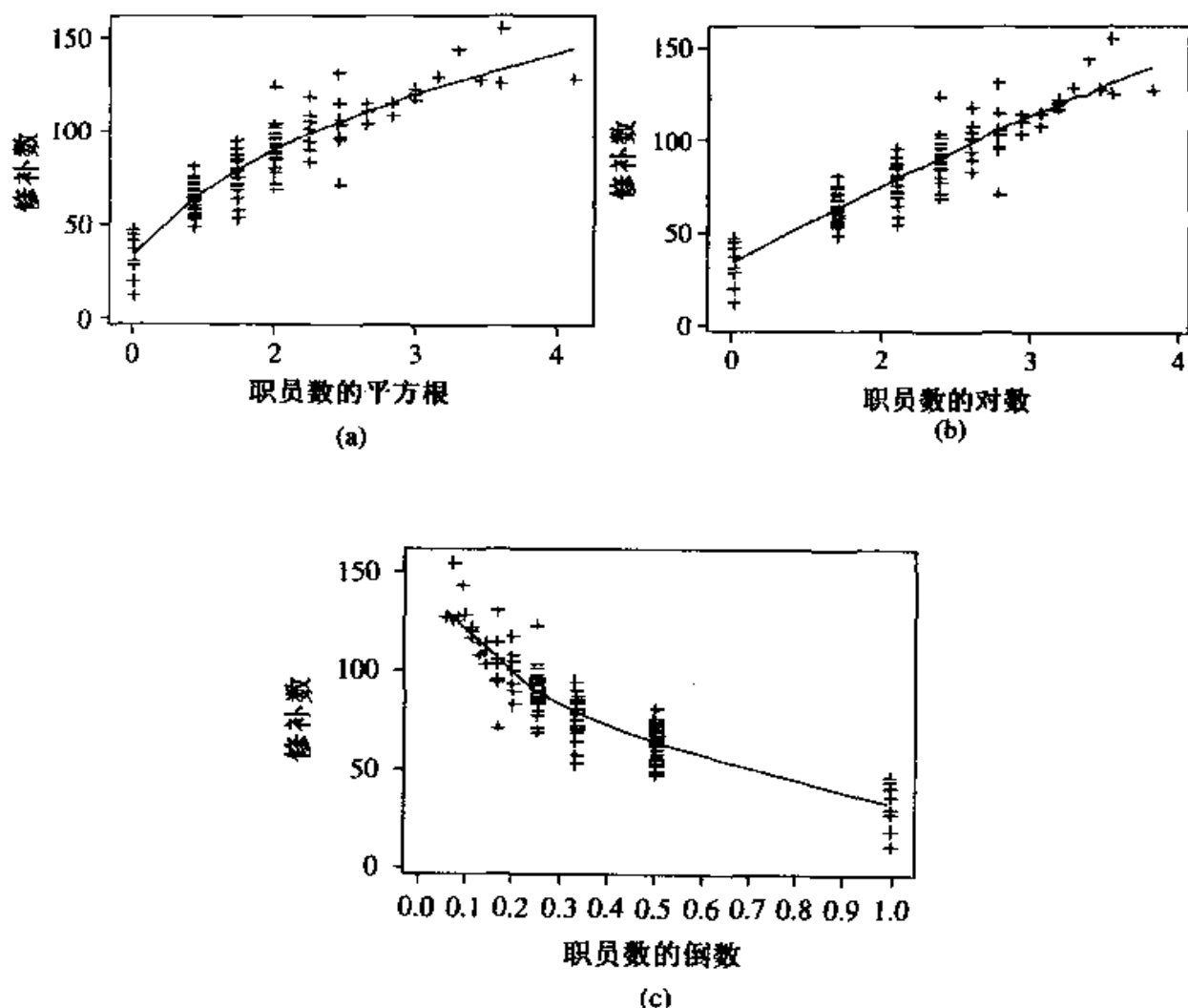


图 11.5 预测变量变换后的散点图

么常数方差的假定就有问题了。这些建议虽然不能覆盖所有的情况,但是包含了最常见的问题。

### 定义 11.2 选择变换的几个建议:

1. 假如散点图显示因变量是随自变量增加而增加,但增加的幅度是逐渐减少的;且在曲线周围的变差大致是常数,则对自变量  $x$  作平方根变换、自然对数变换或倒数变换。

2. 假如散点图显示因变量是随自变量增加而增加,但增加的幅度是逐渐增加的;且在曲线周围的变差大致是常数,用  $x$  和  $x^2$  作为预测变量。因为这个方法使用了两个变量,所以需要下两章所介绍的多元回归分析。

3. 假如散点图显示因变量是随自变量增加到最大值后下降,且在曲线周围的变差大致是常数,用  $x$  和  $x^2$  作为预测变量。

4. 假如散点图显示因变量是随自变量增加而增加, 但增加的幅度是逐渐减少的; 且在曲线周围的变差随着被预测变量  $y$  的增加而增加的, 则用  $y^2$  作为因变量。

5. 假如散点图显示因变量是随自变量增加而增加, 但增加的幅度是逐渐增加的; 且在曲线周围的变差随着被预测变量  $y$  的增加而增加的, 则用  $\ln(y)$  作为因变量。有时候也用  $\ln(x)$  作为自变量。注意原始变量的自然对数增量十分接近原始变量相应的百分比变化, 因此, 变换后的变量的斜率可以用百分比的变化很好地解释。

### 例 11.1

某航线实行飞行常客优惠项目, 该项目的参加者中得到的免费旅行大量增加。为了预测未来这些旅行数量的趋势, 项目负责人收集了最近 72 个月的数据。因变量  $y$  为免费旅行的次数; 自变量  $x$  为月数。图 11.6 给出了利用 Minitab 绘制的带有 LOWESS 平滑曲线的散点图。应该用什么变换呢?

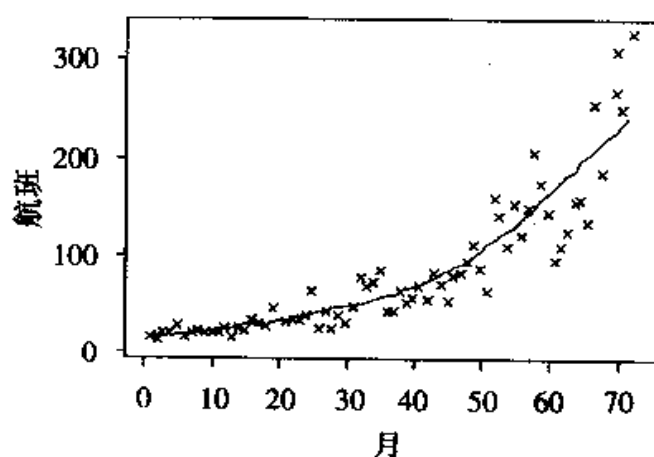


图 11.6 常客免费航班

**解答** 散点图的模式显示因变量是随自变量增加而增加, 但增加的幅度是逐渐增加的。LOWESS 曲线是明确向上的。另外, 围绕平滑曲线的变动(上下)是逐渐增加的, 曲线高端周围的点(这个例中的右端)比曲线低端周围的点更分散。变差逐渐增加建议对变量  $y$  作变换, 这时候自然对数( $\ln$ )变换常常是很好的选择。图 11.7 给出了 Minitab 对  $y$  取对数后重新绘制的散点图。散点图的模式非常接近直线, 并且围绕直线的变差十分接近常数。

我们将在第十二章详细讨论模型假定的检查。对于一个预测变量的简单回归模型, 仔细检查散点图, 最好再用平滑曲线拟合数据, 可以帮助我们避免严重的错误。

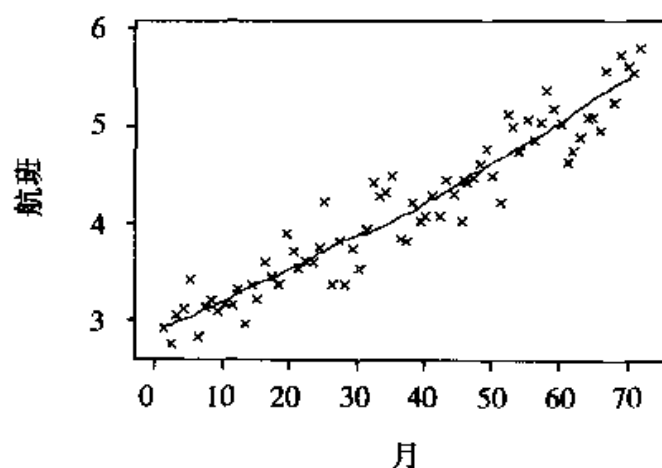


图 11.7 对数变换后的结果

一旦已经确定了任何数学变换,我们必须估计回归线的实际方程。实际上,我们惟一获得的是样本数据,总体的截距、斜率和误差的方差都必须用有限的样本数据来估计。这一节的假定使得我们能够用样本数据对总体参数做出推断。

#### 案例:比较诊断大肠杆菌的两种方法

第七章的案例分析了诊断大肠杆菌(*E. coli*)的一种新的细菌方法,即 Petrifilm HEC 试验。研究人员想要评估 HEC 试验的结果和另一个复杂的在实验室中进行的试验方法 HGMF(hydrophobic grid membrane filtration)的结果的一致性。HEC 试验的接种比传统的方法更容易、更简洁,施行也更安全。但是在使用 HEC 试验之前,必须比较在同一肉体样本上所获得的 HEC 试验的读数和 HGMF 方法的读数,以确定两种方法是否产生同样的结果。如果二者的读数不同但能够获得一个接近 HEC 读数和 HGMF 读数之间关系的方程,那么研究人员可以对 HEC 试验的读数进行校准来预测用 HGMF 方法所能获得的读数。如果 HEC 试验的结果和 HGMF 方法的结果没有关系,那么 HEC 试验不能用来诊断大肠杆菌。

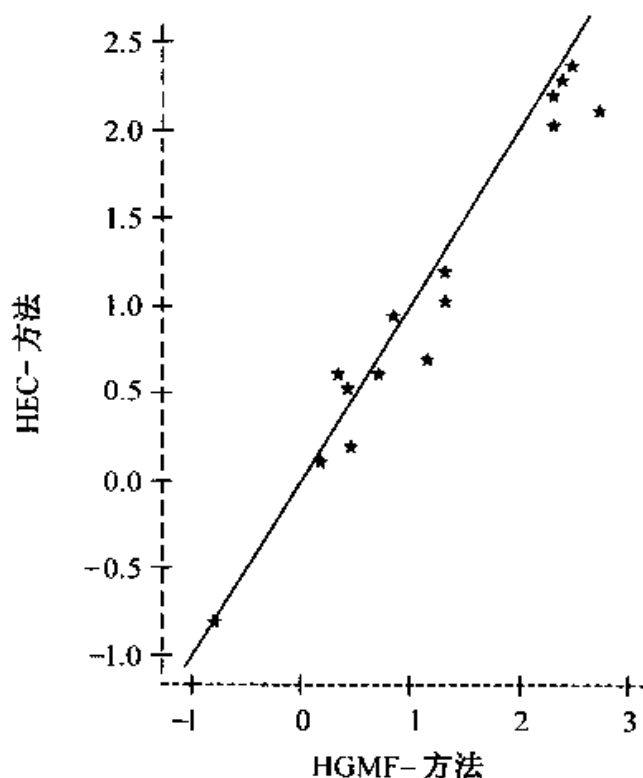
**数据收集的设计** 第七章描述了这个试验的第一个阶段。试验的第二阶段是将两种方法应用于人工污染的牛肉。来自于三头经检测是大肠杆菌阴性的 Holstein 母牛的牛肉块已经准备就绪。取出其中 18 块牛肉并用大肠杆菌污染。分别用 HEC 和 HGMF 方法检测这 18 个样本中的每一块,将这两种方法产生的大肠杆菌浓度进行必要的转换计量( $\log_{10}\text{CFU/ml}$ )。这个案例的数据是 18 对样本点,具体如下。

| RUN | HEC   | HGMF  |
|-----|-------|-------|
| 1   | 0.50  | 0.42  |
| 2   | 0.06  | 0.20  |
| 3   | 0.20  | 0.42  |
| 4   | 0.61  | 0.33  |
| 5   | 0.20  | 0.42  |
| 6   | 0.56  | 0.64  |
| 7   | -0.82 | -0.82 |
| 8   | 0.67  | 1.06  |
| 9   | 1.02  | 1.21  |
| 10  | 1.20  | 1.25  |
| 11  | 0.93  | 0.83  |
| 12  | 2.27  | 2.37  |
| 13  | 2.02  | 2.21  |
| 14  | 2.32  | 2.44  |
| 15  | 2.14  | 2.28  |
| 16  | 2.09  | 2.69  |
| 17  | 2.30  | 2.43  |
| 18  | -0.10 | 1.07  |

**数据的整理** 下一步,研究人员按照前面 2.5 节描述的步骤为以后的统计分析整理数据。他们仔细检查试验过程以确定每一对作为样本的肉都是基本一样的,这样使得 HEC 读数和 HGMF 读数的任何差异都来自于两种方法的不同。在这样的检查中,有关试验过程的问题都会被发现,有问题的观测值都会剔除出去。

**数据的分析** 研究人员感兴趣的是确定这两种方法所产生的大肠杆菌浓度的测量值是否有很强的相关性。下面是试验数据的散点图。(见下页)

散点图中的 45° 直线显示了两个方法的读数的近似的一致性。如果散点落在直线上,则这两个方法所确定的达成杆菌浓度是完全一致的。其中 17 个点都靠近这条直线但有一些偏差,因此研究人员要决定一致性的程度,并获得一个表示两个方法所得读数之间关系的方程。如果利用回归方程可以表示两个方法所得读数的准确的相关性,那么已知 HEC 读数后,研究人员可以预测 HGMF 方法的读数。这使得他们可以比较用 HEC 方法和在实验室中用 HGMF 方法所测得的大肠杆菌浓度。我们将在 11.6 节对这些数据做详细的分析。



注意: 2 个观测点被隐藏起来

## 11.2 估计模型中的参数

### 回归模型

$$y = \beta_0 + \beta_1 x + \epsilon$$

的截距  $\beta_0$  和斜率  $\beta_1$  都是总体的参数, 我们必须从样本数据中估计它们的值。误差的方差  $\sigma_\epsilon^2$  是另一个必须估计的总体参数。回归分析的第一个问题就是获得斜率、截距和方差的估计值。这一节我们讨论如何获得这些估计值。

11.1 节的公路重新铺设例子是一个很合适的说明。假设以下数据是近几年的相似项目中所获得。注意我们有关联单元: 特定的成本和英里数的联系是因为它们来自同一个项目。

|                 |     |      |      |      |      |
|-----------------|-----|------|------|------|------|
| 费用 $y_i$ (千美元): | 6.0 | 14.0 | 10.0 | 14.0 | 26.0 |
| 里程 $x_i$ (英里):  | 1.0 | 3.0  | 4.0  | 5.0  | 7.0  |

第一步是绘制数据的散点图以检查  $x$  和  $y$  的关系。记住图中每个点代表一个观测数据的坐标  $(x, y)$ , 如图 11.8 所示。散点图显示  $x$  和  $y$  之间存在不完全但大体上呈渐增的关系。似乎有可能存在一个直线关系, 但这么有限的无法看出能做什么变换。

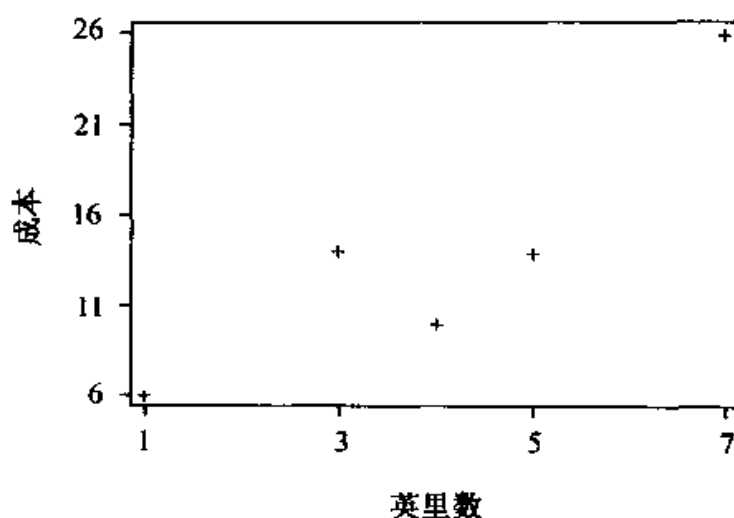


图 11.8 成本对英里数的散点图

回归分析就是找出最佳的直线预测。“最佳”的最常用的标准是根据平方预测误差。我们可以通过最小化总的平方预测误差来求出预测直线的方程——也就是，求出截距  $\hat{\beta}_0$  和斜率  $\hat{\beta}_1$ 。基于这个目的的方法称为最小二乘法，因为它通过最小化下面的量来选择  $\hat{\beta}_0$  和斜率  $\hat{\beta}_1$  的：

$$\sum (y_i - \hat{y}_i)^2 = \sum [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$$

图 11.9 中标出了预测误差，即为与预测直线之间的垂直偏差。这些偏差就是垂直距离，这是因为我们预测的是  $y$ ，误差就应该取  $y$  方向的。对于这些数据，最小二乘线就是  $\hat{y} = 2.0 + 3.0x$ ；由此方程产生的一个偏差用较小的大括号标出。

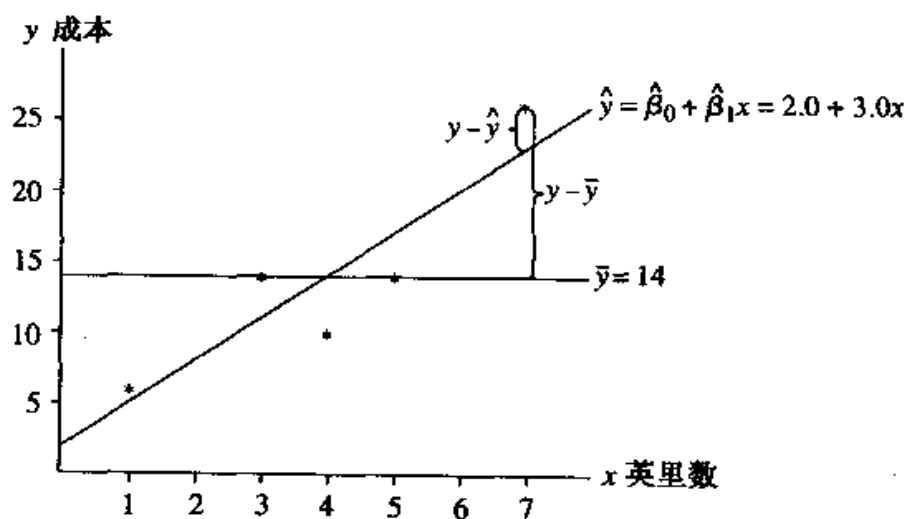


图 11.9 最小二乘线和均值的偏差

为了比较,也标出了均值  $\bar{y} = 14.0$ ; 与均值的偏差用较大的大括号标出。最小二乘法中截距和斜率的计算量有些大,因此这些计算通常都是通过计算机的。

**定义 11.3** 回归模型的截距和斜率的最小二乘估计由下式给出:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \text{ 和 } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

其中

$$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) \text{ 和 } S_{xx} = \sum_i (x_i - \bar{x})^2$$

即,  $S_{xy}$  是  $x$  的偏差乘以  $y$  的偏差的和,  $S_{xx}$  是  $x$  的偏差平方和。

关于公路重新铺设的数据,  $n = 5$  和

$$\sum x_i = 1.0 + \cdots + 7.0 = 20.0$$

因此  $\bar{x} = \frac{20.0}{5} = 4.0$ , 同样,

$$\sum y_i = 70.0, \bar{y} = \frac{70.0}{5} = 14.0$$

进一步,

$$\begin{aligned} S_{xx} &= \sum (x_i - \bar{x})^2 \\ &= (1.0 - 4.0)^2 + \cdots + (7.0 - 4.0)^2 \\ &= 20.0 \end{aligned}$$

和

$$\begin{aligned} S_{xy} &= \sum (x_i - \bar{x})(y_i - \bar{y}) \\ &= (1.0 - 4.0)(6.0 - 14.0) + \cdots + (7.0 - 4.0)(26.0 - 14.0) \\ &= 60.0 \end{aligned}$$

因此

$$\hat{\beta}_1 = \frac{60.0}{20.0} = 3.0 \text{ 和 } \hat{\beta}_0 = 14.0 - (3.0)(4.0) = 2.0$$

从  $\hat{\beta}_1 = 3$ , 我们可以得出重新铺设的公路每增加一英里, 估计成本平均增长 3000 美元。

### 例 11.2

为了检查药店处方销售额与处方配料中从供应商直接采购的比例之间的关系, 特收集 10 家药店的样本数据, 具体的数据如下:

| 药店 | 销售额(千美元), $y$ | 直接采购的配料的百分比, $x$ |
|----|---------------|------------------|
| 1  | 25            | 10               |
| 2  | 55            | 18               |
| 3  | 50            | 25               |
| 4  | 75            | 40               |
| 5  | 110           | 50               |
| 6  | 138           | 63               |
| 7  | 90            | 42               |
| 8  | 60            | 30               |
| 9  | 10            | 5                |
| 10 | 100           | 55               |

- 求出回归直线  $\hat{y} = \beta_0 + \beta_1 x$  的最小二乘估计。
- 一家药店的处方药配料中从供应商直接采购的比例为 15%，预测其处方药销售额。
- 绘制数据  $(x, y)$  的散点图和预测方程  $\hat{y} = \beta_0 + \beta_1 x$  的图像。
- 解释这个问题中的参数  $\beta_1$ 。

### 解答

- 这个方程实际上可以用任何统计计算软件计算；下面是由 Minitab 计算的结果：

MTB > Regress 'Sales' on 1 variable 'Directly'

The regression equation is

Sales = 4.70 + 1.97 Directly

| Predictor | Coef   | Stdev  | t-ratio | p     |
|-----------|--------|--------|---------|-------|
| Constant  | 4.698  | 5.952  | 0.79    | 0.453 |
| Directly  | 1.9705 | 0.1545 | 12.75   | 0.000 |

为了知道计算机是如何进行计算的，你可以从下表得出最小二乘估计：



|     | $y$  | $x$  | $y - \bar{y}$ | $x - \bar{x}$ | $(x - \bar{x})(y - \bar{y})$ | $(x - \bar{x})^2$ |
|-----|------|------|---------------|---------------|------------------------------|-------------------|
|     | 25   | 10   | -46.3         | -22.8         | 1,101.94                     | 566.44            |
|     | 55   | 18   | -16.3         | -15.8         | 257.54                       | 249.64            |
|     | 50   | 25   | -21.3         | -8.8          | 187.44                       | 77.44             |
|     | 75   | 40   | 3.7           | 6.2           | 22.94                        | 38.44             |
|     | 110  | 50   | 38.7          | 16.2          | 626.94                       | 262.44            |
|     | 138  | 63   | 66.7          | 29.2          | 1,947.64                     | 852.64            |
|     | 90   | 42   | 18.7          | 8.2           | 153.34                       | 67.24             |
|     | 60   | 30   | -11.3         | -3.8          | 42.94                        | 14.44             |
|     | 10   | 5    | -61.3         | -28.8         | 1,765.44                     | 829.44            |
|     | 100  | 55   | 28.7          | 21.2          | 608.44                       | 449.44            |
| 总和  | 713  | 338  | 0             | 0             | 6,714.60                     | 3,407.60          |
| 平均值 | 71.3 | 33.8 |               |               |                              |                   |

$$S_{xx} = \sum (x_i - \bar{x})^2 = 3,407.6$$

和

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = 6,714.6$$

代入  $\hat{\beta}_0$  和  $\hat{\beta}_1$  的计算公式,

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{6,714.6}{3,407.6} = 1.9704778 \quad \text{近似为 } 1.97$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 71.3 - 1.9704778(33.8) = 4.6978519 \quad \text{近似为 } 4.70$$

b. 当  $x = 15\%$ , 预测的销售额为  $\hat{y} = 4.70 + 1.97(15) = 34.25$  (也就是 34,250 美元)。

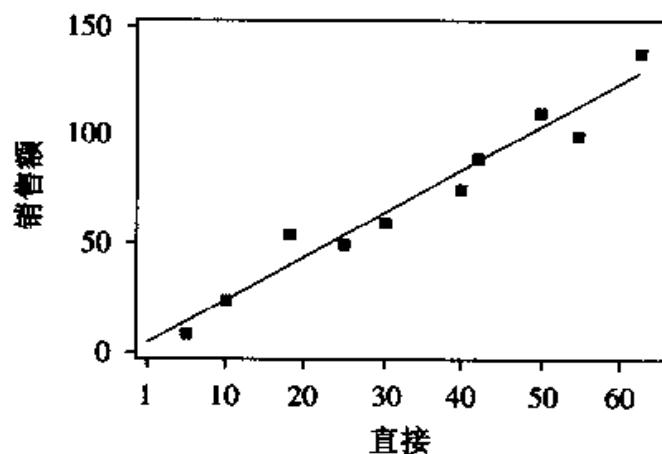


图 11.10 样本数据和最小二乘预测方程

c. 图 11.10 给出了数据( $x, y$ )和预测方程。

d. 从  $\hat{\beta}_1 = 1.97$ , 我们可以得出这样的结论, 如果一家药店医药配料中从供应商直接采购的比例增加 1%, 则预测平均销售额会增加 1,970 美元。

### 例 11.3

利用下面 Statistix 的输出结果, 确定公路重新铺设数据的最小二乘估计:

| PREDICTOR          |             |          |           |           |               |
|--------------------|-------------|----------|-----------|-----------|---------------|
| VARIABLES          | COEFFICIENT | STD      | ERROR     | STUDENT'S | T P           |
| CONSTANT           | 2.00000     | 3.82970  | 0.52      | 0.6376    |               |
| MILES              | 3.00000     | 0.85634  | 3.50      | 0.0394    |               |
| R-SQUARED          | 0.8036      | RESID.   | MEAN      | SQUARE    | (MSE) 14.6666 |
| ADJUSTED R-SQUARED | 0.7381      | STANDARD | DEVIATION |           | 3.82970       |
| SOURCE             | DF          | SS       | MS        | F         | P             |
| REGRESSION         | 1           | 180.000  | 180.000   | 12.27     | 0.0394        |
| RESIDUAL           | 3           | 44.0000  | 14.6666   |           |               |
| TOTAL              | 4           | 224.000  |           |           |               |

**解答** COEFFICIENT 列显示截距  $\hat{\beta}_0 = 2.00000$ ,  $x =$  英里的系数  $\hat{\beta}_1 = 3.0000$ 。

回归模型的斜率的估计很大程度会受到高杠杆点的潜在影响。这类点的自变量值都很大或很小, 即  $x$  方向上的离群值。它们在斜率的估计中占有很大的权重。一个正好对应的  $y$  也是离群值的高杠杆作用点称为高影响点。这样的点将改变斜率和拉弯直线。

如果删除一个数据点将导致回归线发生实质性的改变, 那么这个点就具有高影响力。要具有高影响力, 一个点首先必须是高杠杆点, 另外还必须落在其他点的模式之外。考虑图 11.11 中的两个散点图。图(a)中, 左上角的点远离其余点; 它的  $x$  值极小而具有高杠杆作用。如果我们画出通过其他点的直线, 那么这条直线将在下方远离这个点, 因而这个点在  $y$  方向上也是离群值。因此这个点也具有高影响力, 包括这个点将大大改变直线的斜率。相对的, 图(b)中,  $y$  离群值所对应的  $x$  值非常接近均值, 因而是低杠杆作用点。包括这个点将把直线向上拉, 增大截距的值, 但不会增加或减少斜率的值, 因此这个点不具有高影响力。

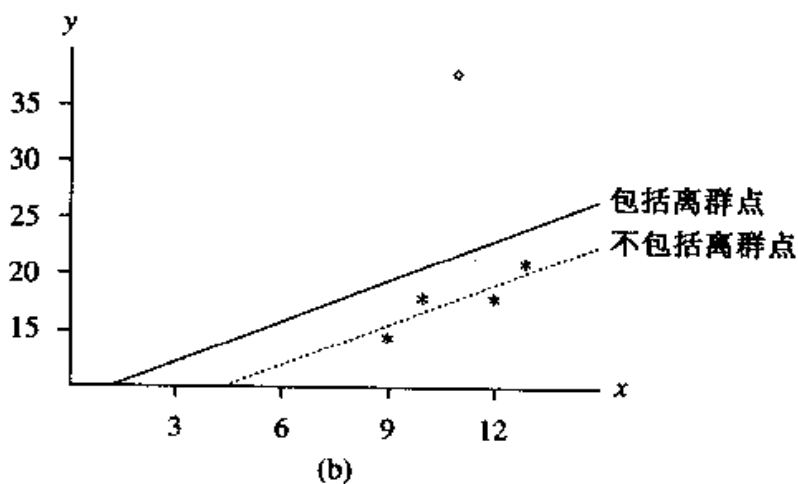
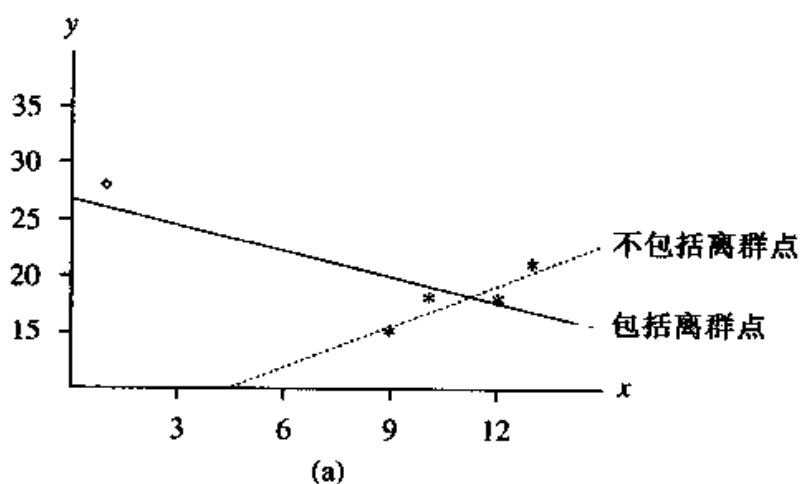


图 11.11 (a) 高影响点; (b) 低影响点

一个高杠杆点只是表示回归方程会有潜在的扭曲。包括这个点是否会“扭曲”回归方程是依赖于它的影响力(这个点是否落在靠近通过其他点的直线的地方)。必须同时具有高杠杆作用和  $y$  离群值的点才具有高影响力。

数学上, 一个点的杠杆作用可以从斜率计算中的  $S_{xy}$  项看出。计算这项值的众多方法中的一个是利用下面的公式

$$S_{xy} = \sum (x_i - \bar{x}) y_i$$

我们可以把这个方程看作  $y$  值的加权和。当  $x$  值远离均值且具有高杠杆作用时, 权重是大的正数或小的负数; 而当  $x$  值非常接近均值且有低杠杆作用时, 权重基本为零。

绝大多数进行回归分析的计算机程序都会计算几种杠杆作用或影响力的诊断度量中的某种度量。这里我们并不对所有的这些诊断度量进行总结。我们只注意那些大测量值对应高杠杆作用点或高影响点的度量。高杠杆作用( $x$  值离群)和高

影响( $x$  值离群和  $y$  值离群)之间的差别也不是完全一致的。检查程序说明以确定所使用的定义。

斜率  $\beta_1$  的标准差可由任何的统计软件计算,一般在输出结果中系数列的右边一列显示。和任何标准差一样,它表明了估计真正的总体或过程参数的准确性。 $\beta_1$  估计值的质量受到两个量的影响:误差方差  $\sigma_e^2$  和自变量的变差  $S_{xx}$ :

$$\sigma_{\beta_1} = \frac{\sigma_e}{\sqrt{S_{xx}}}$$

对于给定的  $x$ ,  $y$  的变异性  $\sigma_e$  越大,  $\sigma_{\beta_1}$  越大。很显然,如果回归线周围有大的变异,那么估计这条直线就困难了。同时,  $x$  的变差(用  $S_{xx}$  测量)越小,  $\sigma_{\beta_1}$  越大。回归方程中的斜率是  $x$  的每改变一个单位所带来的  $y$  的估计的变化;假如数据中的  $x$  改变很少时,也就是  $S_{xx}$  很小,准确估计  $y$  的变化率就困难了。如果某种品牌的食用苏打的价格几年来没有什么变化,那么估计当价格改变所带来需求量的改变明显是困难的。

$$\sigma_{\beta_0} = \sigma_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

估计的截距  $\beta_0$  的标准差受到  $n$  的影响,自然也受到样本均值的平方  $\bar{x}^2$  和  $S_{xx}$  比值的影响。斜率就是当  $x=0$  时  $y$  的预测值;例如,如果所有的  $x_i$  都是大的正数,从实际数据中预测  $x=0$  时  $y$  的值就是一个超级的外推。这样的外推会夸大误差,导致大的  $\beta_0$  的标准差。理想的估计  $\beta_0$  的情况是  $\bar{x}=0$ 。

目前为止,我们只考虑了截距和斜率的估计。我们还必须估计真的误差方差  $\sigma_e^2$ 。我们可以把这个量看作是“围绕回归线的方差”或均方预测误差。 $\sigma_e^2$  的估计是根据残差  $y_i - \hat{y}_i$ , 也就是样本的预测误差得到的。根据样本数据估计的  $\sigma_e^2$  就是残差平方和除以  $n-2$ (自由度)。估计的方差在计算机输出结果中一般显示为 MS (Error) 或 MS(Residual)。回想 MS 表示“均方”,是一个平方和除以适当的自由度:

$$s_e^2 = \frac{\sum_i (y_i - \hat{y}_i)^2}{n-2} = \frac{SS(\text{残差})}{n-2}$$

在例 11.3 的计算机输出结果中, SS(Residual)(即 SS(残差))是 44.0。

正如样本方差(第三章)定义中除以  $n-1$  而不是  $n$ , 围绕回归线的估计方差  $s_e^2$  中是除以  $n-2$ 。由  $n$  减少为  $n-2$  是因为要估计围绕回归线的变异,必须首先估计两个参数  $\beta_0$  和  $\beta_1$  以获得估计的回归线,因此用来估计  $\sigma_e^2$  的有效的样本容量为  $n-2$ 。在我们的定义中,当  $n=2$  时  $s_e^2$  是没有定义的。除以  $n-2$  的另一个理由是使  $s_e^2$  是  $\sigma_e^2$  的一个无偏估计。例 11.3 的计算机输出结果中,  $n-2=5-2=3$  是 RESIDUAL(残差)的 DF(自由度),  $s_e^2=14.6666$  是 RESIDUAL 的 MS。

样本方差的平方根  $s_e$  称为围绕回归线的样本标准差, 估计的标准误差或残差标准差。因为  $s_e$  估计的是  $y_i$  的标准差  $\sigma_e$ ,  $\sigma_e$  估计了给定自变量  $x$  时  $y$  的总体的标准差。例 11.3 的 Statistix 输出结果中,  $s_e$  显示为 STANDARD DEVIATION, 近似为 3.830。

和其他任何标准差一样, 残差标准差也可以用经验法则解释。约 95% 的预测误差在误差均值的  $\pm 2$  标准差之内; 最小二乘回归模型中误差均值总为 0。因此, 残差标准差 3.830 意味着约 95% 的预测误差小于  $\pm 2(3.830) = 7.660$ 。

估计  $\beta_0$ 、 $\beta_1$  和  $s_e$  是回归分析的基础。它们决定回归线和伴随给定  $x$  时  $y$  的误差的可能程度。下一步将是利用这些样本估计量去对真的参数进行统计推断。

#### 例 11.4

某快餐连锁店的人力资源主任研究职员的缺勤率。一旦职员打电话称生病或干脆不来上班, 餐厅经理就不得不赶紧寻找替代者或只能让工作短缺人手。人力资源主任已经收集了这家连锁店下属 10 家餐厅的数据, 包括每周每 100 名职员中缺勤的次数( $y$ )和职员在餐厅工作的平均月数( $x$ )。主任认为工作时间越长的职员越可靠, 缺勤越少。

根据下面的数据和 Minitab 的输出结果,

- 绘制散点图, 并确定直线模型是否是合理的。
- 求出模型  $y = \beta_0 + \beta_1 x + \epsilon$  中参数  $\beta_0$  和  $\beta_1$  的最小二乘估计。
- 当  $x = 19.5$  时, 预测每周每 100 名职员中缺勤的次数  $y$ 。
- 求出围绕回归线的样本标准差  $s_e$ 。
- 解释  $\beta_1$  的值。

$y$ : 31.5 33.1 27.4 24.5 27.0 27.8 23.3 24.7 16.9 18.1  
 $x$ : 18.1 20.0 20.8 21.5 22.0 22.4 22.9 24.0 25.4 27.3

MTB > Regress 'y' on 1 predictor 'x'.

The regression equation is

$y = 64.7 - 1.75 x$

| Predictor | Coef    | Stdev  | t-ratio | p     |
|-----------|---------|--------|---------|-------|
| Constant  | 64.672  | 6.762  | 9.56    | 0.000 |
| x         | -1.7487 | 0.2995 | -5.84   | 0.000 |

$s = 2.388$      $R - sq = 81.0\%$      $R - sq(adj) = 78.6\%$

Analysis of Variance

| SOURCE     | DF | SS     | MS     | F     | P     |
|------------|----|--------|--------|-------|-------|
| Regression | 1  | 194.45 | 194.45 | 34.10 | 0.000 |
| Error      | 8  | 45.61  | 5.70   |       |       |
| Total      | 9  | 240.06 |        |       |       |

**解答**

a. 图 11.12 显示了由 Statistix 软件绘制的散点图, 数据似乎都落在一条向下直线的周围, 因此没有理由使用更复杂的模型。

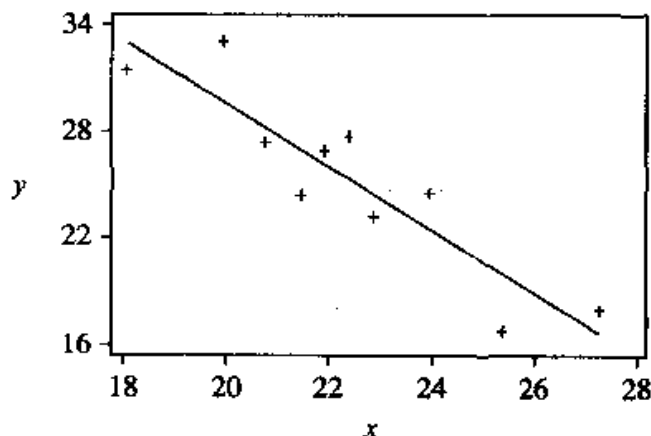


图 11.12 缺勤次数( $y$ )关于工作的平均时间( $x$ )的散点图

- b. 计算机输出结果显示了两次精度不同的系数。截距(结果中标有 Constant)是 6.4672, 斜率( $x$  的系数)是  $-1.7487$ 。注意负的斜率对应着向下的直线。
- c. 当  $x = 19.5$  时,  $y$  的预测值为
- $$\hat{y} = 6.4672 - 1.7487(19.5) = 30.57$$
- d. 围绕回归线的样本标准差(标准残差)是  $s = 2.388$ 。因此约 95% 的预测误差小于  $\pm 2(2.388) = 4.766$ 。
- e. 从  $\hat{\beta}_1 = -1.7487 \approx -1.75$ , 我们可以得出这样的结论, 如果一家餐厅职员的平均工作时间增加 1 个月, 则预测每周每 100 名职员中缺勤的平均次数会减少 1.75 次。

**练习****基本技能**

11.1 绘制下列数据的散点图和通过这些点的直线。

|   |    |    |    |    |    |    |
|---|----|----|----|----|----|----|
| X | 5  | 10 | 12 | 15 | 18 | 24 |
| Y | 10 | 19 | 21 | 28 | 34 | 40 |

11.2 利用方程  $y = 1.8 + 2.0x$

a. 预测  $x=3$  时的  $y$ .

b. 在横坐标轴刻度为 0 到 5、纵坐标轴刻度为 0 到 12 的坐标中绘制这个方程的图形。

11.3 利用以下数据求出最小二乘预测方程。

|     |   |   |   |   |   |
|-----|---|---|---|---|---|
| $x$ | 1 | 2 | 3 | 4 | 5 |
| $y$ | 2 | 4 | 6 | 7 | 9 |

11.4 利用以下数据, 回答(a)和(b)。

|     |   |   |   |   |    |
|-----|---|---|---|---|----|
| $x$ | 1 | 3 | 5 | 7 | 9  |
| $y$ | 1 | 4 | 8 | 9 | 12 |

a. 求出最小二乘预测方程。

b. 利用最小二乘预测方程预测  $x=6$  时的  $y$ 。

11.5 利用练习 11.1 的数据, 求出最小二乘预测方程, 并与徒手绘制的通过这些点的回归线进行比较。

11.6 由 SAS 软件根据数据所作的最小二乘预测方程的计算结果如下:

SAS CODE;

```
option ls = 70 ps = 55 nocenter nodate;
```

```
title 'EXERCISE 11.6';
```

```
data linreg;
```

```
input X Y;
```

```
CARDS;
```

```
10      25
```

```
18      55
```

```
25      50
```

```
40      75
```

```
50      110
```

```
63      138
```

```
42      90
```

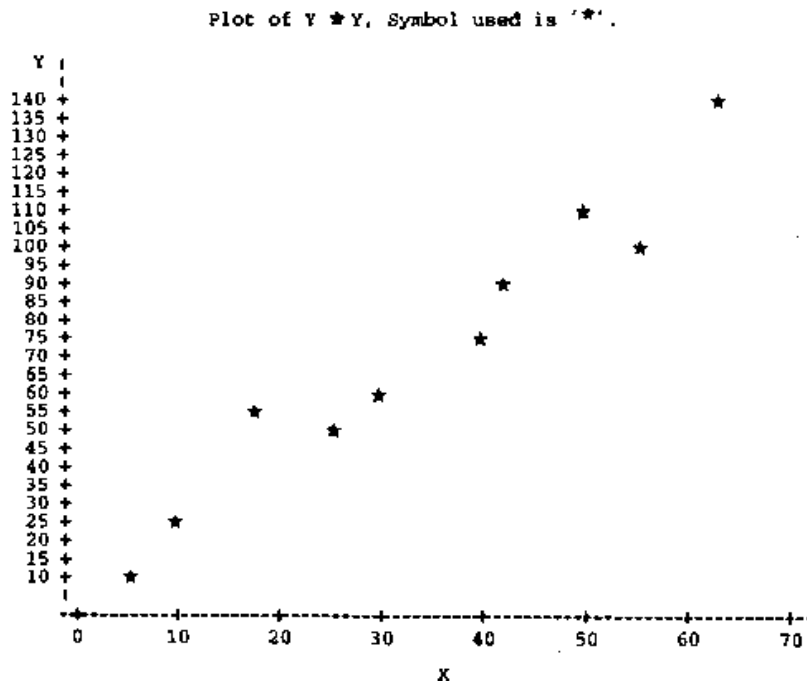
```
30      60
```

```
5       10
```

```

55      100
RUN;
PROC PLOT; PLOT Y * X = ' * ' ;
PROC REG; MODEL Y = X;
OUTPUT OUT = NEW P = PRED R = RESID;
LABEL PRED = 'PREDICTED VALUE' RESID = 'RESIDUALS';
PROC PRINT; VAR Y X PRED RESID;
RUN;
OUTPUT FOR EXERCISE 11.6

```



Dependent Variable: Y

Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 13230.96994    | 13230.96994 | 162.560 | 0.0001 |
| Error   | 8  | 651.13006      | 81.39126    |         |        |
| C Total | 9  | 13882.10000    |             |         |        |

|          |          |            |        |
|----------|----------|------------|--------|
| Root MSE | 9.02171  | R - square | 0.9531 |
| Dep Mean | 71.30000 | Adj R - sq | 0.9472 |
| C. V.    | 12.65317 |            |        |



## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|-------------------------|---------|
| INTERCEP | 1  | 4.697852           | 5.95202071     | 0.789                   | 0.4527  |
| X        | 1  | 1.970478           | 0.15454842     | 12.750                  | 0.0001  |

| OBS | Y   | X  | PREDICTED<br>VALUES | RESIDUALS |
|-----|-----|----|---------------------|-----------|
| 1   | 25  | 10 | 24.403              | 0.5974    |
| 2   | 55  | 18 | 40.166              | 14.8335   |
| 3   | 50  | 25 | 53.960              | -3.9598   |
| 4   | 75  | 40 | 83.517              | -8.5170   |
| 5   | 110 | 50 | 103.222             | 6.7783    |
| 6   | 138 | 63 | 128.838             | 9.1620    |
| 7   | 90  | 42 | 87.458              | 2.5421    |
| 8   | 60  | 30 | 63.812              | -3.8122   |
| 9   | 10  | 5  | 14.550              | -4.5502   |
| 10  | 100 | 55 | 113.074             | -13.0741  |

- 从计算机输出结果中确定最小二乘预测方程,并在数据散点图中画出回归线。
- 这个预测方程能够很好地代表这些数据吗?
- 预测  $x=35$  时的  $y$ 。

## 应用

**11.7 (农业)** 某食品加工企业进行一项试验以检验不同浓度的果胶对罐装甜薯片的硬度的影响,试验中采用三种浓度(按重量)的果胶:0%、1.5%和3%。首先准备好六罐型号为  $303 \times 406$ ,糖度(按重量)为25%的甜薯片,对于每种浓度的果胶,加入糖汁后放进随机选择的六罐甜薯片中,然后封好六罐甜薯片并在  $25^{\circ}\text{C}$  的环境中放置30天。30天后打开这六罐甜薯片并检查其中甜薯片的硬度。具体的数据如下:

| 果胶的浓度 | 0%, 0%     | 1.5%, 1.5% | 3%, 3%     |
|-------|------------|------------|------------|
| 硬度    | 50.5, 46.8 | 62.3, 67.7 | 80.1, 79.2 |

a.  $x$  代表罐头中果胶的浓度,  $y$  代表  $25^{\circ}\text{C}$  的环境中放置 30 天后甜薯片的硬度指标, 绘制样本数据的散点图。

b. 求出模型  $y = \beta_0 + \beta_1 x + \epsilon$  中参数的最小二乘估计。

11.8 参照练习 11.7, 预测果胶浓度为 1% 的罐装甜薯片在  $25^{\circ}\text{C}$  的环境中放置 30 天后, 罐中甜薯片的硬度。

11.9 (环境) 某项试验考察冰冻 7 天后鱼的质量, 10 条大小基本一致的相同种类的生鱼准备用于冰冻。其中两条是捕获后立即冰冻, 两条是捕获 3 小时后冰冻, 其余六条鱼中每两条分别于捕获 6、9 或 12 小时后冰冻。 $y$  代表冰冻 7 天后鱼的质量(以 10 分计),  $x$  代表从捕获到冰冻之间的时间。具体样本数据如下:

|   |     |     |     |     |     |     |     |     |     |     |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Y | 8.5 | 8.4 | 7.9 | 8.1 | 7.8 | 7.6 | 7.3 | 7.0 | 6.8 | 6.7 |
| X | 0   | 0   | 3   | 3   | 6   | 6   | 9   | 9   | 12  | 12  |

a. 绘制样本数据的散点图。

b. 利用最小二乘方法求出模型  $y = \beta_0 + \beta_1 x + \epsilon$  中参数的估计。

c. 解释这个问题中  $\hat{\beta}_1$  的值。

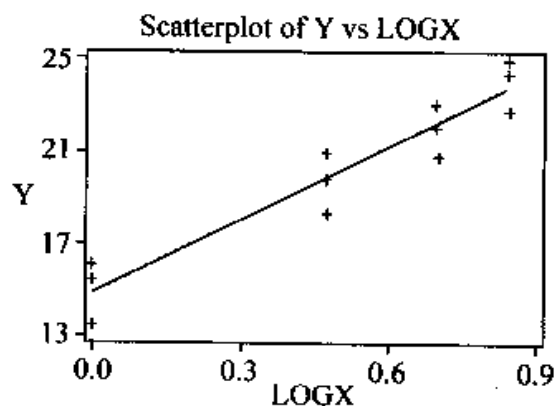
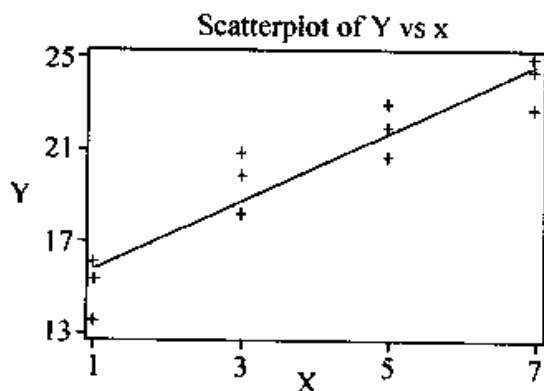
11.10 参照练习 11.9, 预测捕获 10 小时后冰冻 7 天的鱼的质量得分。你能预测捕获 18 小时后冰冻的鱼的质量得分吗?

11.11 某个回归分析研究产生以下数据和 Statistix 散点图。

|                  |        |       |       |       |       |       |       |
|------------------|--------|-------|-------|-------|-------|-------|-------|
|                  | $x$ :  | 1     | 1     | 1     | 3     | 3     | 3     |
| $x' = \log_{10}$ | $x'$ : | 0.000 | 0.000 | 0.000 | 0.477 | 0.477 | 0.477 |
|                  | $y$ :  | 13.5  | 15.4  | 16.1  | 18.3  | 19.9  | 20.9  |

|                  |        |       |       |       |       |       |       |
|------------------|--------|-------|-------|-------|-------|-------|-------|
|                  | $x$ :  | 5     | 5     | 5     | 7     | 7     | 7     |
| $x' = \log_{10}$ | $x'$ : | 0.699 | 0.699 | 0.699 | 0.845 | 0.845 | 0.845 |
|                  | $y$ :  | 20.8  | 23.1  | 22.1  | 22.8  | 24.9  | 24.5  |



- a. 在  $y$  关于  $x$  的散点图中, 近似的斜率为  $x = 7$  时的预测值与  $x = 1$  时的预测值之差除以  $x$  之差——6。
- b. 在  $y$  关于  $x'$  的散点图中, 近似求出预测线的斜率。
- c. 哪个散点图更显示出线性关系?

#### Analysis of Variance

| SOURCE     | DF | SS     | MS     | F     | p     |
|------------|----|--------|--------|-------|-------|
| Regression | 1  | 130.54 | 130.54 | 72.07 | 0.000 |
| Error      | 10 | 18.11  | 1.81   |       |       |
| Total      | 11 | 148.65 |        |       |       |

MTB > Regress 'y' 1 'logx'.

The regression equation is

$$y = 14.9 + 10.5 \log x$$

| Predictor | Coef    | Stdev  | t-ratio | p     |
|-----------|---------|--------|---------|-------|
| Constant  | 14.8755 | 0.6106 | 24.36   | 0.000 |
| logx      | 10.522  | 1.021  | 10.30   | 0.000 |

$s = 1.131$      $R - sq = 91.4\%$      $R - sq(adj) = 90.5\%$

#### Analysis of Variance

| SOURCE     | DF | SS     | MS     | F      | p     |
|------------|----|--------|--------|--------|-------|
| Regression | 1  | 135.85 | 135.85 | 106.17 | 0.000 |
| Error      | 10 | 12.80  | 1.28   |        |       |
| Total      | 11 | 148.65 |        |        |       |

**11.12** 根据练习 11.11 的数据和以下 Minitab 输出结果:

MTB > Regress 'y' 1 'x'.

The regression equation is

$$y = 14.3 + 1.48x$$

| Predictor | Coef    | Stdev  | t-ratio | p     |
|-----------|---------|--------|---------|-------|
| Constant  | 14.2917 | 0.7962 | 17.95   | 0.000 |
| x         | 1.4750  | 0.1737 | 8.49    | 0.000 |

$s = 1.346$      $R-sq = 87.8\%$      $R-sq(adj) = 86.6\%$

a. 确立最小二乘方程  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ 。

b. 求出残差标准差。

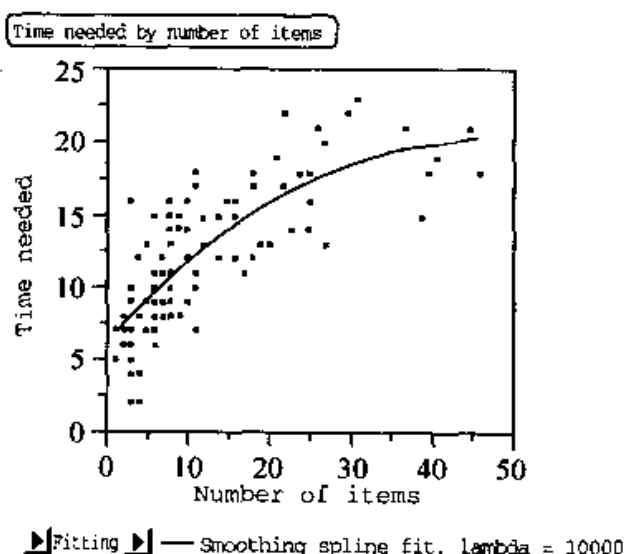
11.13 参照练习 11.11 的数据的输出结果。

a. 求出最小二乘方程  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \log x$ 。

b. 残差标准差是多少？

11.14 比较前两个练习的残差标准差，哪个更小？结果是否与你根据练习 11.11 的散点图所做出的模型选择一致？

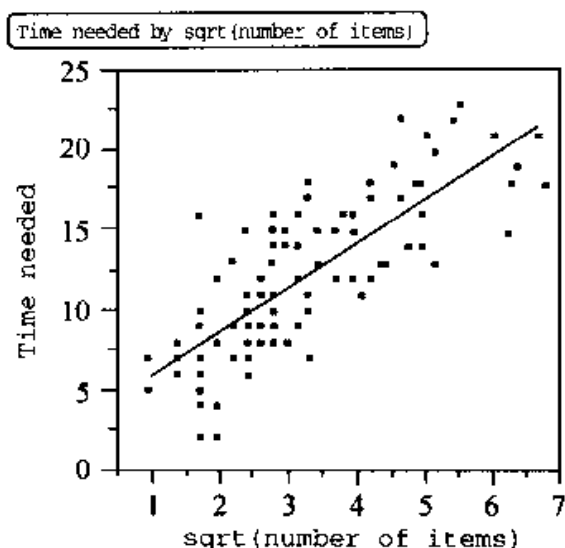
11.15 (商业) 一家邮购零售商花不少功夫在“拣”订单——选择订单上的物品并组合运出。一个小型的研究选择了 100 个订单的样本，由一个有经验的工人完成整个过程，记录下完成每个订单的时间。由 JMP 做出的散点图和样条拟合如下所示，应该使用哪类变换？



11.16 对练习 11.15 中“拣”订单时间数据进行平方根变换。数据变换后的散点图和回归分析结果如下。

a. 数据变换后的散点图更显示出线性性吗？

b. 根据变换后的数据，求出预测方程。



#### Summary of Fit

|                            |          |
|----------------------------|----------|
| RSquare                    | 0.624567 |
| RSquare Adj                | 0.620736 |
| Root Mean Square Error     | 2.923232 |
| Mean of Response           | 12.29    |
| Observations (or Sum Wgts) | 100      |

#### Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Ratio  |
|---------|----|----------------|-------------|----------|
| Model   | 1  | 1393.1522      | 1393.15     | 163.0317 |
| Error   | 98 | 837.4378       | 8.55        | Prob>F   |
| C Total | 99 | 2230.5900      |             | 0.0000   |

#### Parameter Estimates

| Term                   | Estimate  | Std Error | t Ratio | Prob> t |
|------------------------|-----------|-----------|---------|---------|
| Intercept              | 3.097869  | 0.776999  | 3.99    | 0.0001  |
| sqrt (Number of items) | 2.7633138 | 0.216418  | 12.77   | 0.0000  |

**11.17** 在练习 11.16 的 JMP 输出结果中, 残差标准差称为“Root Mean Square Error”, 找出并解释这个数。

**11.18** 在前面的练习中, 为什么变换后数据的标准残差能够与原始数据的标准残差相比?

**11.19** 作为商业银行分行研究的一部分, 收集到位于样本邮政地区的独立公司的数量( $x$ )和商业银行分行的数量( $y$ )。各城市的商业中心除外。

$x$ : 92 116 124 210 216 267 306 378 415 502 615 703  
 $y$ : 3 2 3 5 4 5 5 6 7 7 9 9

以下是数据分析的输出结果(StataQuest):

. regress Branches Business

| Source   | SS         | df | MS          | Number of obs | = | 12     |
|----------|------------|----|-------------|---------------|---|--------|
| Model    | 53.7996874 | 1  | 53.7996874  | F(1, 10)      | = | 172.60 |
| Residual | 3.11697922 | 10 | 0.311697922 | Prob > F      | = | 0.0000 |
| Total    | 56.9166667 | 11 | 5.17424242  | R-square      | = | 0.9452 |
|          |            |    |             | Adj R-square  | = | 0.9398 |
|          |            |    |             | Root MSE      | = | 0.5583 |

| Branches | Coef.     | Std. Err. | t      | p> t  | [95% Conf. Interval] |
|----------|-----------|-----------|--------|-------|----------------------|
| Business | 0.0111049 | 0.0008453 | 13.138 | 0.000 | 0.0092216 0.0129883  |
| _cons    | 1.766846  | 0.3211751 | 5.501  | 0.000 | 1.051223 2.482469    |

- 绘制数据的散点图, 说  $y$  是  $x$  的线性函数合理吗?
- 求出回归方程( $y$  作为因变量)
- 解释这个问题中  $\hat{\beta}_1$  的值。
- 求出样本残差标准差  $s_e$ 。

**11.20** 练习 11.19 的散点图中是否显示出当  $x$  增加时  $y$  的变异性增加?  
(这样会违背同方差的假定)

**11.21** 某个音箱设备架子制造商生产时需要钻孔, 方便钉金属螺钉。如果钻头坏了就必须更换, 这时不但会产生换钻头的费用, 还会耽误生产。工程师不断改变钻速并测量钻头的使用寿命  $y$  (即钻的空数, 以千计), 选用五种钻速  $x$ , 每种钻速用了四个钻头, 具体数据如下:

$x$ : 60 60 60 60 80 80 80 80 100 100  
 $y$ : 4.6 3.8 4.9 4.5 4.7 5.8 5.5 5.4 5.0 4.5

$x$ : 100 100 120 120 120 120 140 140 140 140  
 $y$ : 3.2 4.8 4.1 4.5 4.0 3.8 3.6 3.0 3.5 3.4

- 绘制这些数据的散点图, 散点图显示存在一定的关系吗? 这个关系是线性的吗?
- 是否存在离群值? 如果有, 是否是高影响?

**11.22** 练习 11.21 的数据经过 Excel 的回归分析, 得到以下输出结果:

|    | A                     | B           | C              | D      | E         | F              |
|----|-----------------------|-------------|----------------|--------|-----------|----------------|
| 1  | SUMMARY OUTPUT        |             |                |        |           |                |
| 2  |                       |             |                |        |           |                |
| 3  | Regression Statistics |             |                |        |           |                |
| 4  | Multiple R            | 0.6254      |                |        |           |                |
| 5  | R Square              | 0.3911      |                |        |           |                |
| 6  | Adjusted R Square     | 0.3573      |                |        |           |                |
| 7  | Standard Error        | 0.6324      |                |        |           |                |
| 8  | Observations          | 20          |                |        |           |                |
| 9  |                       |             |                |        |           |                |
| 10 |                       |             |                |        |           |                |
| 11 | ANOVA                 |             |                |        |           |                |
| 12 |                       | df          | SS             | MS     | F         | Significance F |
| 13 | Regression            | 1           | 4.624          | 4.624  | 11.563    | 0.0032         |
| 14 | Residual              | 18          | 7.198          | 0.400  |           |                |
| 15 | Total                 | 19          | 11.822         |        |           |                |
| 16 |                       |             |                |        |           |                |
| 17 |                       |             |                |        |           |                |
| 18 |                       | Coefficient | Standard Error | t Stat | P-value   |                |
| 19 | Intercept             | 6.03        | 0.5195         | 11.606 | 8.617E-10 |                |
| 20 | Speed                 | -0.017      | 0.005          | -3.400 | 3.188E-03 |                |

- 确定最小二乘回归线的截距和斜率。
- 斜率的符号表明钻速和钻头的使用寿命之间是什么关系？
- 确定残差标准差，解释这个数。

#### 11.23 参照练习 11.21。

- 利用练习 11.22 获得的回归线计算当  $x = 60, 80, 100, 120, 140$  时的预测值。
- 哪些  $x$  值所对应的  $y$  的实际值大多数都大于预测值？哪些  $x$  值所对应的  $y$  的实际值大多数都小于预测值？这样的模式是否表明存在线性关系？

#### 11.24 (商业) 某个地产经纪人研究 $x =$ 购房者的家庭年收入(以千美元计)

和  $y$  = 房屋的价格(以千美元计)之间的关系。这个地产经纪人从自己的基本销售地区的一季的购房抵押申请中收集了由 24 笔购房数据组成的样本。以下是具体的数据和 Stata 输出结果。

```

x:   25.0   28.5   29.2   30.0   31.0   31.5   31.9   32.0   33.0
y:   84.9   94.0   96.5   93.5  102.9   99.5  101.0  105.0   99.9

x:   33.5   34.0   35.9   36.0   39.0   39.0   40.5   40.9   42.5
y:  110.0  100.0  116.0  110.0  125.0  119.9  130.6  120.8  129.9

x:   44.0   45.0   50.0   54.6   65.0   70.0
y:  135.5  140.0  150.7  170.0  110.0  185.0

```

```
. regress Price Income
```

| Source   | SS         | df | MS         |                 |        |  |
|----------|------------|----|------------|-----------------|--------|--|
| Model    | 9432.58336 | 1  | 9432.58336 | Number of obs = | 24     |  |
| Residual | 4590.6746  | 22 | 208.667027 | F(1, 22) =      | 45.20  |  |
|          |            |    |            | Prob > F =      | 0.0000 |  |
|          |            |    |            | R-square =      | 0.6726 |  |
|          |            |    |            | Adj R-square =  | 0.6578 |  |
| Total    | 14023.258  | 23 | 609.706868 | Root MSE =      | 14.445 |  |

| Price  | Coef.    | Std. Err. | t     | p> t  | [95 % Conf. Interval] |          |
|--------|----------|-----------|-------|-------|-----------------------|----------|
| Income | 1.80264  | 0.2681147 | 6.723 | 0.000 | 1.246604              | 2.358676 |
| _cons  | 47.15048 | 10.93417  | 4.312 | 0.000 | 24.4744               | 69.82657 |

```
. drop in 23
```

```
(1 observation deleted)
```

```
. regress Price Income
```

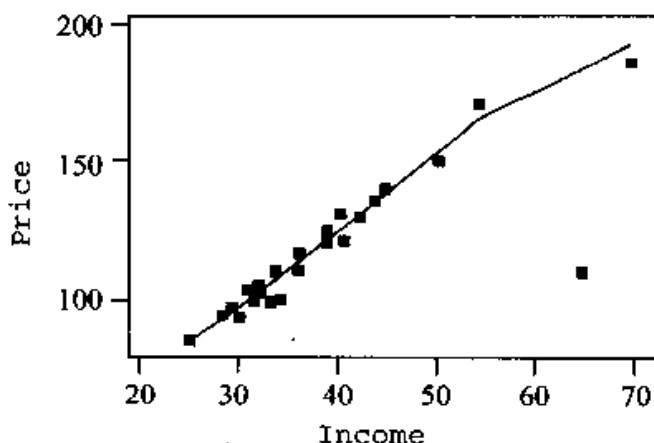
| Source   | SS         | df | MS         |                 |        |  |
|----------|------------|----|------------|-----------------|--------|--|
| Model    | 13407.5437 | 1  | 13407.5437 | Number of obs = | 23     |  |
| Residual | 549.902031 | 21 | 26.185811  | F(1, 21) =      | 512.02 |  |
|          |            |    |            | Prob>F =        | 0.0000 |  |
|          |            |    |            | R-square =      | 0.9606 |  |
|          |            |    |            | Adj R-square =  | 0.9587 |  |
| Total    | 13957.4457 | 22 | 634.429351 | Root MSE =      | 5.1172 |  |

| Price  | Coef.    | Std. Err. | t      | p> t  | [95 % Conf. Interval] |          |
|--------|----------|-----------|--------|-------|-----------------------|----------|
| Income | 2.461967 | 0.108803  | 22.628 | 0.000 | 2.235699              | 2.688236 |
| _cons  | 24.35755 | 4.286011  | 5.683  | 0.000 | 15.4443               | 33.27079 |



- a. 下面是利用 Minitab 绘制的散点图和 LOWESS 平滑线, 二者的关系基本上是线性的吗?
- b. 是否存在高杠杆点? 如果存在, 哪些似乎有高影响?



11.25 根据练习 11.24,

- a. 求出最小二乘回归方程。
- b. 解释斜率系数, 截距有什么意义。
- c. 求出标准残差。

11.26 练习 11.24 的输出结果还包含剔除点  $x = 65.0$ 、 $y = 110.0$  后的回归线。斜率有实质上的变化吗? 为什么?

### 11.3 回归参数的推断

简单回归模型中的斜率、截距和标准残差都是根据有限的数据估计的。与其他所有的统计量一样, 这几个都受到随机误差的影响。本节我们将考虑如何处理由随机误差引起的的影响。以前有关均值和比率假设检验和置信区间的概念同样可以用于这些回归概括数字。

$t$  分布可以用来构造显著性检验和斜率与截距真值的置信区间。一个自然的零假设是斜率真值  $\beta_1$  为 0。如果这个  $H_0$  为真, 那么  $x$  的变化不会对  $y$  的变化产生预测作用, 也就是说  $x$  没有任何预测  $y$  的价值。从前一节知, 样本斜率  $\hat{\beta}_1$  的期望是  $\beta_1$ , 标准差为

$$\sigma_{\hat{\beta}_1} = \sigma_e \sqrt{\frac{1}{S_{xx}}}$$

实际中,  $\sigma_e$  是未知的, 必须用标准残差  $s_e$  来估计。绝大多数回归分析的计算机输出结果中, 估计的标准误差是在系数项的旁边。 $t$  统计量给出了一个这个零假设的检验:

$$t = \frac{\hat{\beta}_1 - \beta_1}{\hat{\beta}_1 \text{ 标准误差的估计}} = \frac{\hat{\beta}_1 - \beta_1}{s_e \sqrt{1/S_{xx}}}$$

这个统计量最常见的应用小结如下。

### $\beta_1$ 的统计检验的小结

假设:

$$1. H_0: \beta_1 \leq 0, H_a: \beta_1 > 0$$

$$2. H_0: \beta_1 \geq 0, H_a: \beta_1 < 0$$

$$3. H_0: \beta_1 = 0, H_a: \beta_1 \neq 0$$

$$T.S. (\text{检验统计量}): t = \frac{\hat{\beta}_1 - 0}{s_e \sqrt{1/S_{xx}}}$$

R.R. (拒绝域): 自由度为  $df = n - 2$ , 第一类错误为  $\alpha$

1. 如果  $t > t_\alpha$  则拒绝  $H_0$ 。

2. 如果  $t < -t_\alpha$  则拒绝  $H_0$ 。

3. 如果  $|t| > t_{\alpha/2}$  则拒绝  $H_0$ 。

检查假定并得出结论。

所有的回归分析输出结果都给出了  $t$  值。

绝大多数计算机输出结果中, 这个检验都是在标准误差项之后, 标识为 T TEST 或 T STATISTIC。计算机输出结果经常也给出  $p$  值, 这样可以不用查  $t$  分布表。

### 例 11.5

利用例 11.3 的计算机输出结果(这里重新列出), 确定  $t$  统计量的值以检验公路重新铺设例子的  $H_0: \beta_1 = 0$ 。给出这个检验的显著性水平。

| PREDICTOR          |             |                          |             |         |        |
|--------------------|-------------|--------------------------|-------------|---------|--------|
| VARIABLES          | COEFFICIENT | STD ERROR                | STUDENT'S T | p       |        |
| CONSTANT           | 2.00000     | 3.82970                  | 0.52        | 0.6376  |        |
| MILES              | 3.00000     | 0.85634                  | 3.50        | 0.0394  |        |
| R-SQUARED          | 0.8036      | RESID. MEAN SQUARE (MSE) |             | 14.6666 |        |
| ADJUSTED R-SQUARED | 0.7381      | STANDARD DEVIATION       |             | 3.82970 |        |
| SOURCE             | DF          | SS                       | MS          | F       | p      |
| REGRESSION         | 1           | 180.000                  | 180.000     | 12.27   | 0.0394 |
| RESIDUAL           | 3           | 44.0000                  | 14.6666     |         |        |
| TOTAL              | 4           | 224.000                  |             |         |        |

**解答** 从输出结果中,很清楚看到检验统计量在标识为 STUDENT'S T 那一列,为  $t=3.50$ 。而双侧检验的对立假设  $H_a: \beta_1 \neq 0$  的  $p$  值,标识为  $P$ ,是 0.0394。因为这个值比较小,所以我们拒绝公路重新铺设的英里数对预测项目的成本没有影响的假设。

### 例 11.6

以下数据是食品行业的 15 家公司的高级主管的平均年龄和前一年公司股票的每股收益的年增长率。利用 Systat 的输出结果检验关于公司高级主管的平均年龄对公司股票的收益变化没有预测价值的假设。应使用单侧检验还是双侧检验?

|      |       |      |      |      |      |      |      |      |      |
|------|-------|------|------|------|------|------|------|------|------|
| 平均年龄 | $x$ : | 38.2 | 40.0 | 42.5 | 43.4 | 44.6 | 44.9 | 45.0 | 45.4 |
| 年增长率 | $y$ : | 8.9  | 13.0 | 4.7  | -2.4 | 12.5 | 18.4 | 6.6  | 13.5 |
|      | $x$ : | 46.0 | 47.3 | 47.3 | 48.0 | 49.1 | 50.5 | 51.6 |      |
|      | $y$ : | 8.5  | 15.3 | 18.9 | 6.0  | 10.4 | 15.9 | 17.1 |      |

DEP VAR:CHGEPS N:15 MULTIPLE R: 0.383 SQUARED MULTIPLE R:0.147  
STANDARD ERROR OF ESTIMATE:5.634

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | T     | P(2 TAIL) |
|----------|-------------|-----------|----------|-------|-----------|
| CONSTANT | -16.991     | 18.866    | 0.000    | 0.901 | 0.384     |
| MEANAGE  | 0.617       | 0.413     | 0.383    | 1.496 | 0.158     |

### ANALYSIS OF VARIANCE

| SOURCE     | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P     |
|------------|----------------|----|-------------|---------|-------|
| REGRESSION | 71.055         | 1  | 71.055      | 2.239   | 0.158 |
| RESIDUAL   | 412.602        | 13 | 31.739      |         |       |

**解答** 模型  $y = \beta_0 + \beta_1 x + \epsilon$  中,零假设为  $H_0: \beta_1 = 0$ 。美国商界的普遍看法是年轻的经理更积极进取和努力工作,年长的经理更富有经验而引导更好的决策。因此,选择双侧检验更好,其对立假设为  $H_a: \beta_1 \neq 0$ 。 $t$  统计量在输出结果中标识为 T 的那一列,  $t=1.496$ ,  $p$  值为 0.158。因此没有足够的证据表明公司高级主管的平均年龄和公司股票的收益变化之间存在关系。

附带说明,注意  $\beta_0$  的解释在这个例子中比较有趣,它代表一家高级主管平均年龄为 0 的公司的股票收益变化。

也可以计算斜率真值的置信区间,这是一个非常好的表示斜率估计中的不准确程度的方法。置信区间就是简单地将估计值加上或减去  $t$  分布表的值乘上标准

误差。

**斜率  $\beta_1$  的置信区间**

$$\hat{\beta}_1 - t_{\alpha/2} s_e \sqrt{\frac{1}{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2} s_e \sqrt{\frac{1}{S_{xx}}}$$

值  $t_{\alpha/2}$  自由度为  $n-2$ , 也就是误差的自由度。

### 例 11.7

利用例 11.3 的计算机输出结果, 计算斜率  $\beta_1$  的 95% 置信区间。

**解答** 计算机输出结果中,  $\hat{\beta}_1 = 3.000$ ,  $\hat{\beta}_1$  的估计的标准误差近似为 0.856。由于  $n$  是 5, 则误差的自由度是  $5-2=3$ 。查表知  $t_{\alpha/2} = t_{0.05/2} = t_{0.025} = 3.182$ 。那么  $\beta_1$  真值的相应的置信区间是

$$3.00 \pm 3.182(0.856) \quad \text{或} \quad 0.276 \text{ 到 } 5.724$$

公路重新铺设每增加一英里, 预测增加的成本可能是 276 美元与 5,724 美元之间的任何值。这个区间的长度过大主要是由于样本规模过小。

$F$  检验是另一个可以使用的检验, 也可以用于检验零假设“没有预测价值”。这个检验涉及的零假设是所有的预测变量对  $y$  都没有预测价值, 在简单线性回归分析中给出和零假设为  $H_0: \beta_1 = 0$  的双侧  $t$  检验相同的结果, 因为所有的预测变量都没有预测价值就是斜率(惟一的)为 0。  $F$  检验小结如下。

**$H_0: \beta_1 = 0$  的  $F$  检验**

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

$$\text{T. S. (检验统计量): } F = \frac{\text{SS(回归)}/1}{\text{SS(残差)}/(n-2)} = \frac{\text{MS(回归)}}{\text{MS(残差)}}$$

$$\text{R. R. (拒绝域): 分子自由度 } df_1 = 1, \text{ 分母自由度 } df_2 = n - 2,$$

如果  $F > F_\alpha$  则拒绝  $H_0$ 。

检查假定并得出结论。

SS(回归)是  $y$  的预测值与  $y$  的均值之间差异的平方和,  $\text{SS(回归)} = \sum (\hat{y}_i - \bar{y})^2$ 。SS(残差)是  $y$  的实际值与  $y$  的预测值之间差异的平方和,  $\text{SS(残差)} = \sum (y_i - \hat{y}_i)^2$ 。

几乎所有的计算机软件都计算  $F$  统计量。公路重新铺设例子中, 计算机输出结果显示  $F = 12.27$ ,  $p$  值是 0.0394, 因此再次拒绝没有预测价值的假设。简单线性回归分析问题中,  $F = t^2$  总是正确的; 这个例子中, 12.27 就近似为  $(3.50)^2$ 。  $F$  检验和  $t$  检验在简单线性回归分析是一致的, 但在多元回归分析中有不同的目的。

**例 11.8**

这里重新列出了例 11.4 的计算机输出结果, 利用  $F$  检验零假设  $H_0: \beta_1 = 0$ 。证明  $t^2 = F$ 。

The regression equation is

$$y = 64.7 - 1.75x$$

| Predictor | Coef    | Stdev  | t-ratio | p     |
|-----------|---------|--------|---------|-------|
| Constant  | 64.672  | 6.762  | 9.56    | 0.000 |
| x         | -1.7487 | 0.2995 | -5.84   | 0.000 |

s = 2.388

R-sq = 81.0 %

R-sq(adj) = 78.6 %

Analysis of Variance

| SOURCE     | DF | SS     | MS     | F     | P     |
|------------|----|--------|--------|-------|-------|
| Regression | 1  | 194.45 | 194.45 | 34.10 | 0.000 |
| Error      | 8  | 45.61  | 5.70   |       |       |
| Total      | 9  | 240.06 |        |       |       |

**解答** 计算机结果显示  $F$  统计量为 34.10,  $p$  值是 0.000 (表明实际的  $p$  值小于 0.0005)。注意  $t$  统计量 -5.84, 则  $t^2 = (-5.84)^2 = 34.11$ , 近似等于  $F$ 。

读者应该能够得出截距  $\beta_0$  的相似的假设检验和置信区间, 利用  $\hat{\beta}_0$  的估计的标准误差

$$\sigma_{\hat{\beta}_0} = s_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}$$

实际上, 人们对这个参数的兴趣不如对斜率的兴趣。特别地, 经常没有理由去假设截距真值为零 (或任何特定的值)。计算机软件几乎总是检验斜率是零的零假设, 但并不操心截距项的检验。

**练习**

**11.27** 参照练习 11.15 的数据。

- 计算  $\beta_1$  的 95% 置信区间。
- 练习 11.15 中  $H_0: \beta_1 = 0$  的解释是什么?
- 这个问题中对立假设  $H_a$  是什么?
- 数据在  $\alpha = 0.05$  水平下支持  $H_a$  吗? 说明你的假定。

**11.28** 求出前一个练习中零假设为  $H_0: \beta_1 = 0$  的检验的  $p$  值。

**11.29 (农业)** 假设一位研究人员要研究使用不同浓度 (0%、1.5% 和 3%, 按

重量)的果胶和放置在 25℃ 的环境中的罐装甜薯片的硬度之间的关系。六罐甜薯片的样本数据如下。

|          |      |      |      |      |      |      |
|----------|------|------|------|------|------|------|
| $y$ (硬度) | 50.5 | 46.8 | 62.3 | 67.7 | 80.1 | 79.2 |
| $x$ (浓度) | 0    | 0    | 1.5  | 1.5  | 3.0  | 3.0  |

a. 求出模型  $y = \beta_0 + \beta_1 x + \epsilon$  中参数的最小二乘估计。

b. 求出  $\sigma_\epsilon^2$  的估计。

c. 求出  $\beta_1$  的标准误差。

**11.30** 参照练习 11.29。设计一个零假设为果胶的浓度和放置在 25℃ 的环境中 30 天后罐装甜薯片的硬度之间没有线性关系的统计检验。算出这个检验的  $p$  值并得出结论。

**11.31 (生物)** 某种疾病传播的范围在很大程度上受空气中悬浮的感染有机体的生存能力的影响。由于所研究的这种疾病的感染性质,这些有机体的生存能力必须在密闭的房间里研究。一个方法是将准备好的包含感染有机体的溶液以雾状喷洒在房间里。任何特定时间后生物复原(biological recovery)都是用悬浮在气雾中有机体的总数中还存活的比例来衡量的。附表数据是从 13 个不同的气雾中计算的生物复原比例,每个喷雾的复原比例是在不同的时间计算得到的。

| 气雾 | 时间, $x$ (分钟) | 生物复原(%) |
|----|--------------|---------|
| 1  | 0            | 70.6    |
| 2  | 5            | 52.0    |
| 3  | 10           | 33.4    |
| 4  | 15           | 22.0    |
| 5  | 20           | 18.3    |
| 6  | 25           | 15.1    |
| 7  | 30           | 13.0    |
| 8  | 35           | 10.0    |
| 9  | 40           | 9.1     |
| 10 | 45           | 8.3     |
| 11 | 50           | 7.9     |
| 12 | 55           | 7.7     |
| 13 | 60           | 7.7     |

a. 绘制数据的散点图。

b. 由于数据图显示有些弯曲, 试试对生物复原取对数, 使数据线性化。

11.32 参照练习 11.31。

a. 拟合线性回归模型  $y = \beta_0 + \beta_1 x + \epsilon$ , 其中  $y$  是  $\log(\text{生物复原})$ 。

b. 计算  $\sigma_\epsilon$  的估计。

c. 确定  $\hat{\beta}_0$  和  $\hat{\beta}_1$  的标准误差。

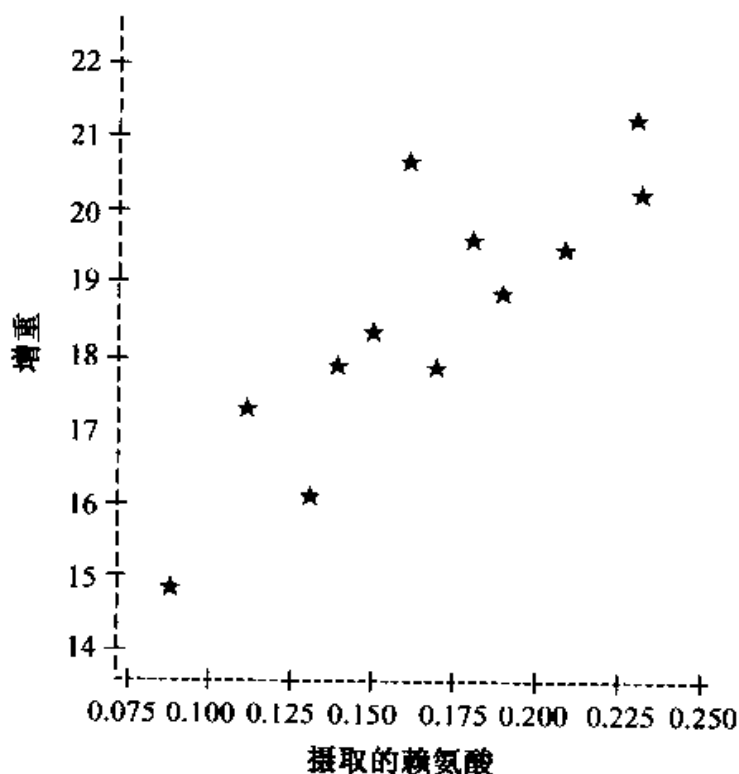
11.33 参照练习 11.31, 进行一个零假设为  $H_0: \beta_1 = 0$  的检验, 取  $\alpha = 0.05$ 。

11.34 参照练习 11.31。求出  $\beta_0$  的 95% 置信区间,  $\beta_0$  为时间为零时的  $\log(\text{生物复原比例})$  的均值。解释你的结果。(注: 当  $x = 0$  时  $E(y) = \beta_0$ )

11.35 (农业) 某个研究人员进行一项试验, 这个试验是检验用添加氨基酸赖氨酸的饲料喂养的鸡的体重的增加与所摄取的赖氨酸的量之间的关系。由于赖氨酸的比例是已知的, 又可以控制鸡的进食, 因此鸡所摄取的赖氨酸是可以确定的。这个研究中随机选取 12 只两周大的鸡为样本, 每只分别在不同的鸡笼饲养且可进食添加指定量的赖氨酸的饲料。下面给出的是试验期间鸡的体重的增加和所摄取赖氨酸的样本数据。(数据中,  $y$  代表以克计的鸡增加的体重,  $x$  代表以克计的所摄取的赖氨酸)

a. 参照计算机输出结果, 线性模型合适吗?

b. 根据计算机输出结果, 求出估计的线性回归模型  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ 。



| 鸡 | y    | x    | 鸡  | y    | x    |
|---|------|------|----|------|------|
| 1 | 14.7 | 0.09 | 7  | 17.2 | 0.11 |
| 2 | 17.8 | 0.14 | 8  | 18.7 | 0.19 |
| 3 | 19.6 | 0.18 | 9  | 20.2 | 0.23 |
| 4 | 18.4 | 0.15 | 10 | 16.0 | 0.13 |
| 5 | 20.5 | 0.16 | 11 | 17.8 | 0.17 |
| 6 | 21.1 | 0.23 | 12 | 19.4 | 0.21 |

## OUTPUT FOR EXERCISE 11.35

Dependent Variable: Y

WEIGHT GAIN

Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 28.35785       | 28.35785    | 26.522  | 0.0004 |
| Error   | 10 | 10.69215       | 1.06921     |         |        |
| C Total | 11 | 39.05000       |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 1.03403  | R-square | 0.7262 |
| Dep Mean | 18.45000 | Adj R-sq | 0.6988 |
| C.V.     | 5.60449  |          |        |

## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|----------------------------|---------|
| INTERCEP | 1  | 12.508525          | 1.19168259     | 10.497                     | 0.0001  |
| X        | 1  | 35.827989          | 6.95693918     | 5.150                      | 0.0004  |

| Variable | DF | Variable Label  |
|----------|----|-----------------|
| INTERCEP | 1  | Intercept       |
| X        | 1  | LYSINE INGESTED |



| OBS | Y    | X    | PREDICTED<br>VALUES | RESIDUALS |
|-----|------|------|---------------------|-----------|
| 1   | 14.7 | 0.09 | 15.7330             | -1.03304  |
| 2   | 17.8 | 0.14 | 17.5244             | 0.27556   |
| 3   | 19.6 | 0.18 | 18.9576             | 0.64244   |
| 4   | 18.4 | 0.15 | 17.8827             | 0.51728   |
| 5   | 20.5 | 0.16 | 18.2410             | 2.25900   |
| 6   | 21.1 | 0.23 | 20.7490             | 0.35104   |
| 7   | 17.2 | 0.11 | 16.4496             | 0.75040   |
| 8   | 18.7 | 0.19 | 19.3158             | -0.61584  |
| 9   | 20.2 | 0.23 | 20.7490             | -0.54896  |
| 10  | 16.0 | 0.13 | 17.1662             | -1.16616  |
| 11  | 17.8 | 0.17 | 18.5993             | -0.79928  |
| 12  | 19.4 | 0.21 | 20.0324             | -0.63240  |

**11.36** 参照练习 11.35 的输出结果。

a. 估计  $\sigma_e^2$ 。

b. 确定  $\hat{\beta}_1$  的标准误差。

c. 进行一个统计检验, 检验在这种饲料配方和研究时间下, 鸡所增加的体重和所摄取的赖氨酸之间存在一个直接的(正的)线性关系。

**11.37** 参照练习 11.35。

a. 这个例子中,  $\beta_0$  的物理上的解释有意义吗? (提示: 赖氨酸是混合在饲料中)

b. 考虑另一个关于鸡所增加的体重和所摄取的赖氨酸之间关系的模型

$$y = \beta_1 x + \varepsilon$$

区别这个模型和模型  $y = \beta_0 + \beta_1 x + \varepsilon$ 。

**11.38**

a. 参照练习 11.37(b), 从下面的计算机输出结果中确定模型  $y = \beta_1 x + \varepsilon$  中的  $\hat{\beta}_1$ , 其中

$$\hat{\beta}_1 = \frac{\sum xy}{\sum x^2}$$

b. 两个模型  $y = \beta_0 + \beta_1 x + \varepsilon$  和  $y = \beta_1 x + \varepsilon$  中哪个对样本数据给出了更好的拟合? (提示: 在样本观测图中检查这两个预测方程)

OUTPUT FOR EXERCISE 11.38

NOTE: NO Intercept in model. R-square is redefined.

Dependent Variable: Y      WEIGHT GAIN

## Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 3995.38497     | 3995.38497  | 342.031 | 0.0001 |
| Error   | 11 | 128.49503      | 11.68137    |         |        |
| U Total | 12 | 4123.88000     |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 3.41780  | R-square | 0.9688 |
| Dep Mean | 18.45000 | Adj R-sq | 0.9660 |
| C.V.     | 18.52467 |          |        |

## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for HO: Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|-------------------------|---------|
| X        | 1  | 106.523715         | 5.75988490     | 18.494                  | 0.0001  |

|          |      | Variable        |           |           |        |
|----------|------|-----------------|-----------|-----------|--------|
| Variable | DF   | Label           |           |           |        |
| X        | 1    | LYSINE INGESTED |           |           |        |
| OBS      | Y    | X               | PREDICTED | RESIDUALS | VALUES |
| 1        | 14.7 | 0.09            | 9.5871    | 5.11287   |        |
| 2        | 17.8 | 0.14            | 14.9133   | 2.88668   |        |
| 3        | 19.6 | 0.18            | 19.1743   | 0.42573   |        |
| 4        | 18.4 | 0.15            | 15.9786   | 2.42144   |        |
| 5        | 20.5 | 0.16            | 17.0438   | 3.45621   |        |
| 6        | 21.1 | 0.23            | 24.5005   | -3.40045  |        |
| 7        | 17.2 | 0.11            | 11.7176   | 5.48239   |        |
| 8        | 18.7 | 0.19            | 20.2395   | -1.53951  |        |
| 9        | 20.2 | 0.23            | 24.5005   | -4.30045  |        |
| 10       | 16.0 | 0.13            | 13.8481   | 2.15192   |        |
| 11       | 17.8 | 0.17            | 18.1090   | -0.30903  |        |
| 12       | 19.4 | 0.21            | 22.3700   | -2.96998  |        |

**11.39 (商业)** 某个印制汽车减震器标签的工厂研究标签的总的直接成本和印制的批量之间的关系。数据是用计算机软件 Execustat 分析的, 相关的输出结果如下所示:

Simple Regression Analysis

Linear model: TotalCost = 99.777 + 5.19179 \* Runsize

Table of Estimates

|           | Estimate | Standard Error | t Value | P Value |
|-----------|----------|----------------|---------|---------|
| Intercept | 99.777   | 2.8273         | 35.29   | 0.0000  |
| Slope     | 5.19179  | 0.0586455      | 88.53   | 0.0000  |

R-squared = 99.64 %

Correlation coeff. = 0.998

Standard error of estimation = 12.2065

Durbin-Watson statistic = 2.67999

Analysis of Variance

| Source        | Sum of Squares | D.F. | Mean Square    | F-Ratio | P Value |
|---------------|----------------|------|----------------|---------|---------|
| Model         | 1.16775e + 006 | 1    | 1.16775e + 006 | 7837.26 | 0.0000  |
| Error         | 4171.98        | 28   | 148.999        |         |         |
| Total (corr.) | 1.17192e + 006 | 29   |                |         |         |

a. 绘制数据的散点图。利用线性回归模型是否有困难? 你是否能看出有哪些地方违背了模型假设? 原始数据如下:

|     |     |     |      |     |      |     |     |     |     |     |     |     |
|-----|-----|-----|------|-----|------|-----|-----|-----|-----|-----|-----|-----|
| 批量  | 2.6 | 5.0 | 10.0 | 2.0 | 0.8  | 4.0 | 2.5 | 0.6 | 0.8 | 1.0 | 2.0 |     |
| 总成本 | 230 | 341 | 629  | 187 | 159  | 327 | 206 | 124 | 155 | 147 | 209 |     |
| 批量  | 3.0 | 0.4 | 0.5  | 5.0 | 20.0 | 5.0 | 2.0 | 1.0 | 1.5 | 0.5 | 1.0 | 1.0 |
| 总成本 | 247 | 135 | 125  | 366 | 1146 | 339 | 208 | 150 | 179 | 128 | 155 | 143 |
| 批量  | 0.6 | 2.0 | 1.5  | 3.0 | 6.5  | 2.2 | 1.0 |     |     |     |     |     |
| 总成本 | 131 | 219 | 171  | 258 | 415  | 226 | 159 |     |     |     |     |     |

b. 求出输出结果中显示的估计的回归方程。确定标准残差。

c. 计算斜率真值的 95% 置信区间。这个问题的截距和斜率的意义是什么?

**11.40** 参照练习 11.39 的计算机输出结果。

a. 确定零假设为  $H_0: \beta_1 = 0$  的检验的  $t$  统计量的值。

b. 确定这个检验的  $p$  值, 这个  $p$  值是单侧还是双侧的?

11.41 参照练习 11.39 的计算机输出结果。

a. 确定  $F$  统计量的值和相应的  $p$  值。

b. 这个  $F$  检验的  $p$  值和练习 11.40 的  $t$  检验的  $p$  值一致吗? 为什么这个关系应该保持?

## 11.4 利用回归预测新的 $y$ 值

到目前为止做的所有回归分析中, 我们已经对观测的数据之间的关系做了总结和推断, 因此已经对过去发生的做了预测。而回归分析最重要的应用之一就是对未来进行预测。在公路重新铺设的例子中, 县公路管理部门主任需要预测一个用于招标新的合同的成本。在已知销售价格预测销售数量的回归分析中, 某经理需要预测新价格下的需求。本节将讨论如何进行回归分析预测和怎样确定正的或负的误差因子。

根据  $x$  的一个给定值所预测的  $y$  值有两种可能的解释。假如公路管理部门主任将  $x = 6$  英里代入回归方程  $\hat{y} = 2.0 + 3.0x$ , 得到  $\hat{y} = 20$ , 这个预测值可以解释为以下两者之一。

“所有 6 英里长的公路重新铺设的平均成本  $E(y)$  是 20,000 美元。”

或

“某个 6 英里长的公路重新铺设的成本  $y$  是 20,000 美元。”

两种情况的最佳预测都是 20, 但是正负误差是不同的。预测平均值  $E(y)$  比预测个别  $y$  值容易, 因此预测平均值的正负误差小。我们首先讨论预测平均值的正负误差范围, 需知这是解决特定值问题的中间步骤。

平均值预测问题中, 预测变量  $x$  的值假定是已知的。过去的  $x$  值记为  $x_1, \dots, x_n$ , 新的  $x$  值记为  $x_{n+1}$ , 则  $\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$  用来预测  $E(y_{n+1})$ 。由于  $\hat{\beta}_0$  和  $\hat{\beta}_1$  都是无偏的,  $\hat{y}_{n+1}$  是  $E(y_{n+1})$  的一个无偏预测。可以证明  $\hat{y}_{n+1}$  的标准误差为

$$\sigma_e \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}}}$$

这里  $S_{xx}$  是  $x_i$  的原始的  $n$  个值的偏差平方和; 大多数计算机输出结果都是用下面的公式计算的:

$$\left( \frac{s_e}{\hat{\beta}_1 \text{ 的标准误差}} \right)^2$$

自由度为  $n - 2$  (误差自由度) 的  $t$  分布表是需要的。通常用来构造置信区间的方法一就是“估计  $\pm t$  (标准误差)”, 也可以用来获得  $E(y_{n+1})$  的置信区间。一些好

的统计计算软件可以在给定新的  $x$  值而没有确定相应的  $y$  值时计算这个置信区间。

$E(y_{n+1})$  的置信区间的计算公式

$$\hat{y}_{n+1} - t_{\alpha/2} s_e \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}}} \leq E(y_{n+1}) \leq \hat{y}_{n+1} + t_{\alpha/2} s_e \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}}}$$

其中  $t_{\alpha}$  表示在自由度为  $n-2$  的  $t$  分布中  $t_{\alpha}$  的右侧的面积为  $\alpha$  的点。

在公路重新铺设的例子中, 下面的计算机输出结果显示当  $x=6$  时  $E(y_{n+1})$  的估计值是 20, 相应的  $E(y_{n+1})$  的置信区间是 (12.29, 27.71)。

#### Resurfacing Data

##### PREDICTOR

| VARIABLES          | COEFFICIENT | STD ERROR               | STUDENT'S T | P       |
|--------------------|-------------|-------------------------|-------------|---------|
| CONSTANT           | 2.00000     | 3.82970                 | 0.52        | 0.6376  |
| MILES              | 3.00000     | 0.85634                 | 3.50        | 0.0394  |
| R-SQUARED          | 0.8036      | RESID. MEAN SQUARE(MSE) |             | 14.6666 |
| ADJUSTED R-SQUARED | 0.7381      | STANDARD DEVIATION      |             | 3.82970 |

| SOURCE     | DF | SS      | MS      | F     | P      |
|------------|----|---------|---------|-------|--------|
| REGRESSION | 1  | 180.000 | 180.000 | 12.27 | 0.0394 |
| RESIDUAL   | 3  | 44.0000 | 14.6666 |       |        |
| TOTAL      | 4  | 224.000 |         |       |        |

##### PREDICTED/FITTED VALUES OF COST

|                       |        |                    |        |
|-----------------------|--------|--------------------|--------|
| LOWER PREDICTED BOUND | 5.5791 | LOWER FITTED BOUND | 12.291 |
| PREDICTED VALUE       | 20.000 | FITTED VALUE       | 20.000 |
| UPPER PREDICTED BOUND | 34.420 | UPPER FITTED BOUND | 27.708 |
| SE (PREDICTED VALUE)  | 4.5313 | SE (FITTED VALUE)  | 2.4221 |

PREDICTOR VALUES: MILES = 6.0000

可以看到的,  $E(y_{n+1})$  的置信区间中估计的加减项依赖于样本规模  $n$  和围绕回归线的标准偏差。它也依赖于  $x_{n+1}$  到  $\bar{x}$  (以前的  $x_i$  的均值) 的平方距离与  $S_{xx}$

的比值。 $x_{n+1}$  离  $\bar{x}$  越远, 项

$$\frac{(x_{n+1} - \bar{x})^2}{S_{xx}}$$

越大。当  $x_{n+1}$  远离其他的  $x$  值时, 这项的值就大, 此时预测值就是程度很强的外推。回归线的小的估计误差会被这种外推放大,  $(x_{n+1} - \bar{x})^2 / S_{xx}$  可以称为**外推惩罚**, 这是因为它随着外推的程度而增加。

**外推**——在自变量值远离数据时的预测, 常常是吸引人而又危险的。使用它需要一个假定条件, 那就是变量之间在远离数据的情况下继续保持线性关系。一般来说, 你没有数据去检验这个假定条件。例如, 某个公司可能发现一个季度所雇职员的人数(1200 和 1400 之间)与这个季度的利润存在一个负的关系, 职员越少, 利润越大。但如果从这个事实中就得出将职员减少到 600 人会大大增加利润, 会产生惊人的风险。(你认为我们能有负数个职员吗?) 职员数量的下降迟早会对公司的业务造成负面的影响, 因此公司的利润会下降。外推惩罚往往是低估了风险, 这是因为它建立在线性关系这个假定条件上的, 而这个假定在大的外推时是不可靠的。

置信和预测区间很大程度上也受同方差假定的影响。一些回归分析情况中, 当预测值增加时围绕回归线的变异就增大, 因而违背这个假定, 此时, 置信和预测区间在变异相对小时就会太宽, 而变异相对大时就太窄。一个散点图如果显示出“扇形”就表明是异方差, 这时置信和预测区间就不准确了。

### 例 11.9

从例 11.4 的数据和下面的 Minitab 输出结果中, 求出  $x_{n+1}$  为 22.4 时  $E(y_{n+1})$  的 95% 的置信区间。比较这个区间的宽度和  $x_{n+1}$  为 30.4 时的置信区间的宽度。

```
MTB> regress 'y' on 1 variable 'x';
```

```
SUBC> predict at 22.4;
```

```
SUBC> predict at 30.4.
```

The regression equation is

$$y = 64.7 - 1.75x$$

| Predictor | Coef    | Stdev  | t-ratio | p     |
|-----------|---------|--------|---------|-------|
| Constant  | 64.672  | 6.762  | 9.56    | 0.000 |
| x         | -1.7487 | 0.2995 | -5.84   | 0.000 |

s = 2.388

R-sq = 81.0 %

R-sq(adj) = 78.6 %

## Analysis of Variance

| SOURCE     | DF | SS     | MS     | F     | p     |
|------------|----|--------|--------|-------|-------|
| Regression | 1  | 194.45 | 194.45 | 34.10 | 0.000 |
| Error      | 8  | 45.61  | 5.70   |       |       |
| Total      | 9  | 240.06 |        |       |       |

| Fit    | Stdev. Fit | 95 % C.I.        | 95 % P.I.          |
|--------|------------|------------------|--------------------|
| 25.500 | 0.755      | (23.758, 27.242) | (19.723, 31.277)   |
| 11.510 | 2.500      | ( 5.742, 17.278) | ( 3.535, 19.485)XX |

X denotes a row with X values away from the center

XX denotes a row with very extreme X values

**解答** 对于  $x_{n+1}=22.4$ , 标识为“Fit”的两个数中的第一个就显示了预测的值为 25.5, 置信区间为(23.758, 27.242)。对于  $x_{n+1}=30.4$ , 预测的值为 11.51, 置信区间为(5.742, 17.278)。第二区间的宽度大约是 11.5, 远远大于第一个区间的长度 3.5。这是因为  $x_{n+1}=30.4$  远离  $x$  的数据的范围, 外推惩罚使得区间较宽。

通常关心的预测问题是预测  $y_{n+1}$  的个体值而不是  $E(y_{n+1})$ 。大多数的计算机软件中, 预测个体值的区间称为**预测区间**。预测  $y_{n+1}$  同样是利用最佳的估计  $\hat{y}_{n+1}$ , 只是预测的误差项比预测  $E(y_{n+1})$  的误差项大。事实上, 可以证明利用  $\hat{y}_{n+1}$  预测  $y_{n+1}$  的公式如下。

 $y_{n+1}$  的预测区间

$$\hat{y}_{n+1} - t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}}} \leq y_{n+1} \leq \hat{y}_{n+1} + t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}}}$$

其中  $t_{\alpha/2}$  表示在自由度为  $n-2$  的  $t$  分布中  $t_{\alpha/2}$  的右侧的面积为  $\alpha/2$  的点。

在公路重新铺设的例子中,  $x=6$  时的  $y_{n+1}$  的相应的 95% 预测区间是(5.58, 34.42)(参看前面的计算机输出结果)。  $E(y_{n+1})$  和  $y_{n+1}$  的 95% 预测区间分别画在图 11.13, 里面的曲线是  $E(y_{n+1})$  的, 外面是  $y_{n+1}$  的。

预测  $E(y_{n+1})$  的区间和  $y_{n+1}$  的区间的惟一不同之处在于标准误差公式中的那项“+1”。有这一项表明个体值的预测准确性比平均值的预测差, 外推惩罚也是适用的, 同样也是低估了风险。如果  $n$  大而外推项小, 则项“+1”在平方根里占主要地位, 此时置信区间近似为  $\hat{y}_{n+1} - t_{\alpha/2} s_e \leq y_{n+1} \leq \hat{y}_{n+1} + t_{\alpha/2} s_e$ 。因此, 当  $n$  大时, 大约 68% 的残差(预测误差)在  $\pm 1s_e$  内, 大约 95% 是在  $\pm 2s_e$  内。考虑忽略平

方根里的其他项的原则没有多少意义。计算机输出结果中常会计算出这些项, 包含它们没有什么危害。

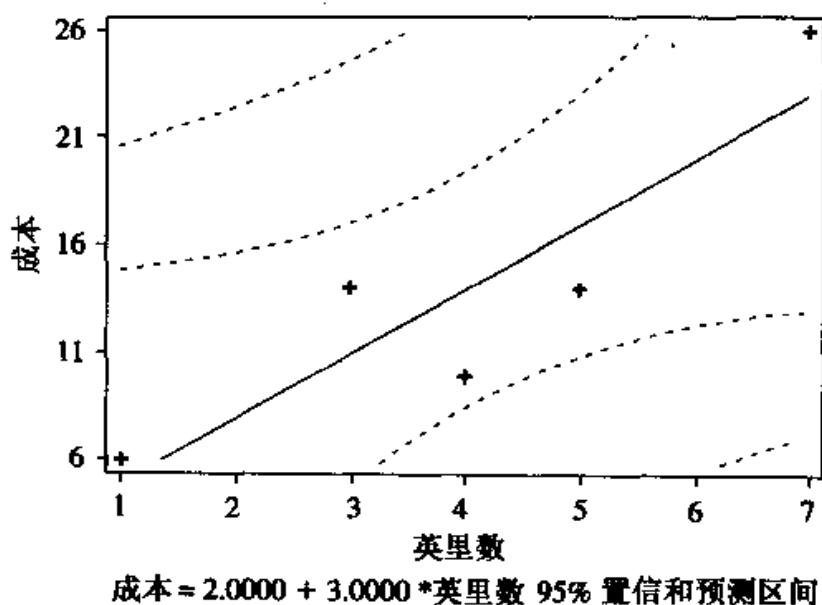


图 11.13 95% 界限的预测值与观测值

### 例 11.10

利用例 11.9 的计算机输出结果(这里重新列出), 分别求出  $x_{n+1} = 22.4$  时和  $x_{n+1} = 30.4$  时的  $y_{n+1}$  的 95% 的置信区间, 并与用  $\pm 2s_e$  估计的区间的宽度进行比较。

**解答** 在例 11.9 中, 当  $x_{n+1} = 22.4$  时  $\hat{y}_{n+1} = 25.5$ , 显示的预测区间为

$$19.72 \leq y_{n+1} \leq 31.28$$

而由  $\pm 2s_e$  确定的区间为

$$25.5 - (2)(2.388) \leq y_{n+1} \leq 25.5 + (2)(2.388) \text{ 或 } 20.72 \leq y_{n+1} \leq 30.28$$

后一个区间窄一点, 主要是因为查自由度为 8 的  $t$  分布表得出的  $t$  值稍微大于 2。

当  $x_{n+1} = 30.4$  时  $\hat{y}_{n+1} = 11.51$ , 95% 预测区间为

$$3.54 \leq y_{n+1} \leq 19.48$$

而由  $\pm 2s_e$  确定的区间为

$$11.5 - (2)(2.388) \leq y_{n+1} \leq 11.5 + (2)(2.388) \text{ 或 } 6.72 \leq y_{n+1} \leq 16.28$$

后一个区间窄很多, 不仅是因为  $t$  分布表得出的  $t$  值大于 2, 而且大的外推惩罚没有反映出来。计算机输出结果给这个预测标上“XX”, 表明所用的  $x$  值远离数据。



```
MTB>regress 'y' on 1 variable 'x';
SUBC>predict at 22.4;
SUBC>predict at 30.4.
```

The regression equation is  
 $y = 64.7 - 1.75x$

| Predictor | Coef    | Stdev  | t-ratio | p     |
|-----------|---------|--------|---------|-------|
| Constant  | 64.672  | 6.762  | 9.56    | 0.000 |
| x         | -1.7487 | 0.2995 | -5.84   | 0.000 |

s = 2.388      R-sq = 81.0 %      R-sq(adj) = 78.6 %

#### Analysis of Variance

| SOURCE     | DF | SS     | MS     | F     | p     |
|------------|----|--------|--------|-------|-------|
| Regression | 1  | 194.45 | 194.45 | 34.10 | 0.000 |
| Error      | 8  | 45.61  | 5.70   |       |       |
| Total      | 9  | 240.06 |        |       |       |

| Fit    | Stdev. Fit | 95 % C. I.       | 95 % P. I.         |
|--------|------------|------------------|--------------------|
| 25.500 | 0.755      | (23.758, 27.242) | (19.723, 31.277)   |
| 11.510 | 2.500      | ( 5.742, 17.278) | ( 3.535, 19.485)XX |

X denotes a row with X values away from the center

XX denotes a row with vary extreme X values

## 练习

### 基本技能

11.42 参照练习 11.31, 对于最小二乘方程

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

利用 95% 的置信区间估计 30 分钟时的  $\log(\text{生物复原比例})$  的均值。

11.43 利用练习 11.42 的数据, 构造 30 分钟时的  $\log(\text{生物复原比例})$  的 95% 的预测区间, 并与练习 11.42  $E(y)$  的置信区间比较。

## 应用

11.44 (工程) 某个化学家想要确定一特定的混合物暴露在空气中, 其重量的减少  $y$  与暴露时间之间的函数关系。下表给出了  $n = 12$  的样本的重量减少和

暴露时间的数据。

重量减少和暴露时间的数据

| 重量减少(磅) | 暴露时间(小时) |
|---------|----------|
| 4.3     | 4        |
| 5.5     | 5        |
| 6.8     | 6        |
| 8.0     | 7        |
| 4.0     | 4        |
| 5.2     | 5        |
| 6.6     | 6        |
| 7.5     | 7        |
| 2.0     | 4        |
| 4.0     | 5        |
| 5.7     | 6        |
| 6.5     | 7        |

a. 求出模型  $y = \beta_0 + \beta_1 x + \epsilon$  的最小二乘预测方程。

b. 检验  $H_0: \beta_1 \leq 0$ , 给出  $H_0: \beta_1 > 0$  的  $p$  值并得出结论。

11.45 参照练习 11.44, 以下给出了 SAS 的计算输出结果。

a. 确定  $4 \leq x \leq 7$  时,  $E(y)$  的 95% 的置信区间。

b. 确定  $4 \leq x \leq 7$  时,  $y$  的 95% 的预测区间。

c. 区分(a)中置信区间和(b)中预测区间的意义。

Dependent Variable: Y WEIGHT LOSS

Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 26.00417       | 26.00417    | 40.223  | 0.0001 |
| Error   | 10 | 6.46500        | 0.64650     |         |        |
| C Total | 11 | 32.46917       |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 0.80405  | R-square | 0.8009 |
| Dep Mean | 5.50833  | Adj R-sq | 0.7810 |
| C.V.     | 14.59701 |          |        |

#### Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob >  T |
|----------|----|--------------------|----------------|----------------------------|-----------|
| INTERCEP | 1  | -1.733333          | 1.16518239     | -1.488                     | 0.1677    |
| X        | 1  | 1.316667           | 0.20760539     | 6.342                      | 0.0001    |

| X | Y   | Predict Value | Std Err Predict | Lower95 % Mean | Upper95 % Mean | Lower95 % Predict | Upper95 % Predict | Residual |
|---|-----|---------------|-----------------|----------------|----------------|-------------------|-------------------|----------|
| 4 | 4.3 | 3.5333        | 0.388           | 2.6679         | 4.3987         | 1.5437            | 5.5229            | 0.7667   |
| 5 | 5.5 | 4.8500        | 0.254           | 4.2835         | 5.4165         | 2.9710            | 6.7290            | 0.6500   |
| 6 | 6.8 | 6.1667        | 0.254           | 5.6001         | 6.7332         | 4.2877            | 8.0456            | 0.6333   |
| 7 | 8.0 | 7.4833        | 0.388           | 6.6179         | 8.3487         | 5.4937            | 9.4729            | 0.5167   |
| 4 | 4.0 | 3.5333        | 0.388           | 2.6679         | 4.3987         | 1.5437            | 5.5229            | 0.4667   |
| 5 | 5.2 | 4.8500        | 0.254           | 4.2835         | 5.4165         | 2.9710            | 6.7290            | 0.3500   |
| 6 | 6.6 | 6.1667        | 0.254           | 5.6001         | 6.7332         | 4.2877            | 8.0456            | 0.4333   |
| 7 | 7.5 | 7.4833        | 0.388           | 6.6179         | 8.3487         | 5.4937            | 9.4729            | 0.0167   |
| 4 | 2.0 | 3.5333        | 0.388           | 2.6679         | 4.3987         | 1.5437            | 5.5229            | -1.5333  |
| 5 | 4.0 | 4.8500        | 0.254           | 4.2835         | 5.4165         | 2.9710            | 6.7290            | -0.8500  |
| 6 | 5.7 | 6.1667        | 0.254           | 5.6001         | 6.7332         | 4.2877            | 8.0456            | -0.4667  |
| 7 | 6.5 | 7.4833        | 0.388           | 6.6179         | 8.3487         | 5.4937            | 9.4729            | -0.9833  |

|                            |         |
|----------------------------|---------|
| Sum of Residuals           | 0       |
| Sum of Squared Residuals   | 6.4650  |
| Predicted Resid SS (Press) | 10.0309 |

11.46 以下是练习 11.39 的计算机输出结果的另一部分。

Table of Predicted Values

| Row | Runsize | Predicted TotalCost | 95.00 % Prediction Limits |         | 95.00 % Confidence Limits |         |
|-----|---------|---------------------|---------------------------|---------|---------------------------|---------|
|     |         |                     | Lower                     | Upper   | Lower                     | Upper   |
|     | 2       | 203.613             | 178.169                   | 229.057 | 198.902                   | 208.323 |

a. 预测所有印刷批量为 2,000 (即  $\text{Runsize} = 2.0$ ) 的减震器标签合同的平均总直接成本。

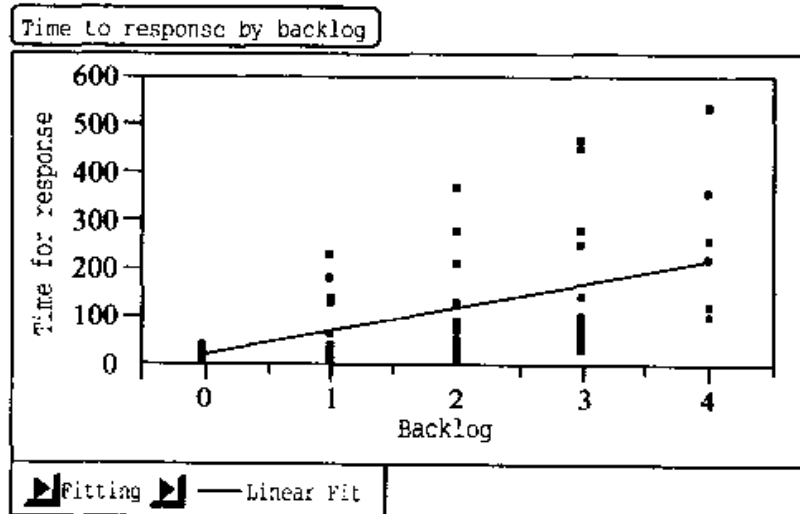
b. 确定这个均值的 95% 的置信区间。

11.47 练习 11.39 中的预测是严重的推断吗?

11.48 参照练习 11.46。

a. 预测印刷批量为 2,000 的减震器标签合同的总直接成本, 并求出 95% 的预测区间。

b. 如果这个合同的实际直接成本为 250 美元, 很奇怪吗?



| Linear Fit                 |           |                |             |         |
|----------------------------|-----------|----------------|-------------|---------|
| Summary of Fit             |           |                |             |         |
| RSquare                    |           | 0.241325       |             |         |
| RSquare Adj                |           | 0.22868        |             |         |
| Root Mean Square Error     |           | 107.3671       |             |         |
| Mean of Response           |           | 113.871        |             |         |
| Observations (or Sum Wgts) |           | 62             |             |         |
| Analysis of Variance       |           |                |             |         |
| Source                     | DF        | Sum of Squares | Mean Square | F Ratio |
| Model                      | 1         | 220008.88      | 220009      | 19.0852 |
| Error                      | 60        | 691662.09      | 11528       | Prob>F  |
| C Total                    | 61        | 911670.97      |             | 0.0001  |
| Parameter Estimates        |           |                |             |         |
| Term                       | Estimate  | Std Error      | t Ratio     | Prob> t |
| Intercept                  | 23.817935 | 24.71524       | 0.96        | 0.3391  |
| Backlog                    | 48.131793 | 11.01751       | 4.37        | 0.0001  |

练习 11.49 中等待时间数据的分析

**11.49** 某个供热承包商为出现供热问题而打电话的家庭提供维修人员。这个承包商希望能有办法估计到维修人员开始工作时为止顾客会等待多长时间。有关等待时间的分钟数和前面积压的电话数的数据已经收集到。以下是由 JMP 得到的数据的散点图和回归分析。

a. 当前面积压电话为 6 时, 计算此时电话的响应时间的预测值和近似的 95% 预测区间, 忽略外推惩罚。

b. 如果已经计算了外推惩罚, 这个惩罚会非常小吗?

**11.50** 在前一个练习中, 计算的预测区间是过窄还是过宽?

**11.51** 下面是练习 11.11 的部分输出结果。

MTB>Regress 'y' 1 'x';

SUBC>Predict 1.301.

The regression equation is

$$y = 14.3 + 1.48x$$

| Predictor | Coef    | Stdev  | t-ratio | p     |
|-----------|---------|--------|---------|-------|
| Constant  | 14.2917 | 0.7962 | 17.95   | 0.000 |
| x         | 1.4750  | 0.1737 | 8.49    | 0.000 |

s = 1.346      R-sq = 87.8 %      R-sq(adj) = 86.6 %

| Fit    | Stdev. Fit | 95 % C. I.       | 95 % P. I.          |
|--------|------------|------------------|---------------------|
| 43.792 | 2.807      | (37.536, 50.047) | (36.854, 50.729) XX |

X denotes a row with X values away from the center

XX denotes a row with very extreme X values

MTB>Regress 'y' 1 'logx';

SUBC>Predict 20.

The regression equation is

$$y = 14.9 + 10.5 \log x$$

| Predictor | Coef    | Stdev  | t-ratio | p     |
|-----------|---------|--------|---------|-------|
| Constant  | 14.8755 | 0.6106 | 24.36   | 0.000 |
| logx      | 10.522  | 1.021  | 10.30   | 0.000 |

$s = 1.131$        $R\text{-sq} = 91.4\%$        $R\text{-sq}(\text{adj}) = 90.5\%$

| Fit    | Stdev. Fit | 95 % C. I.       | 95 % P. I.          |
|--------|------------|------------------|---------------------|
| 28.565 | 0.8756     | (26.614, 30.516) | (25.378, 31.752) XX |

X denotes a row with X values away from the center

XX denotes a row with very extreme X values

- 如果采用了  $y$  为  $x$  的线性函数的模型, 当  $x = 20$  时预测的  $y$  是多少?
- 如果采用了  $y$  为  $x' = \log_{10} x$  的线性函数的模型, 当  $x = 20$  时预测的  $y$  是多少?

c. 两个预测中哪个更合理(或更少不合理性)?

11.52 给出前一个练习中你所选择的预测的 95% 预测区间。

## 11.5 线性回归中拟合不足的考察

在前面的回归分析讨论中, 我们仅仅从直观的角度考虑线性回归模型  $y = \beta_0 + \beta_1 x + \epsilon$  对数据是不是拟合得好。以前我们只是从数据的散点图检查是否有线性关系, 并检验斜率是否不为 0; 但没有方法检验是否有高阶模型更适合表示  $y$  与  $x$  之间的关系。本节将讨论线性回归模型有效性的检验。

图(或表)总是检验拟合程度的一个好的开端。首先可以利用  $y$  关于  $x$  的散点图, 其次残差  $y_i - \hat{y}_i$  关于预测的值  $\hat{y}_i$  的散点图可以显示出以下的问题:

1. 离群或错误的观测值。检查残差图, 可以很自然看出有非同寻常的大残差(指绝对值)的数据点。
2. 模型假定的违背。在  $y = \beta_0 + \beta_1 x + \epsilon$  中, 我们已经假定  $y$  与自变量  $x$  之间存在线性关系, 误差彼此独立, 服从相同方差的正态分布。

如果模型和数据没有这些明显的问题, 其残差图看起来就像图 11.14 的散点图。从这个散点图可以注意到没有特别大的残差(因此没有明显的离群值), 也没有迹象表明线性模型是不合适的。但一个高阶模型更合适时, 所观察到的散点图

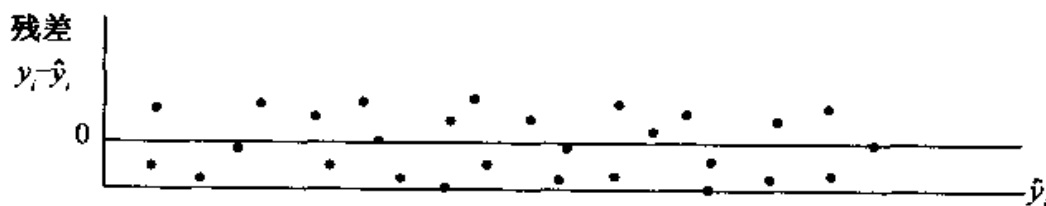


图 11.14 模型形式合适的残差散点图

就像图 11.15 所示。

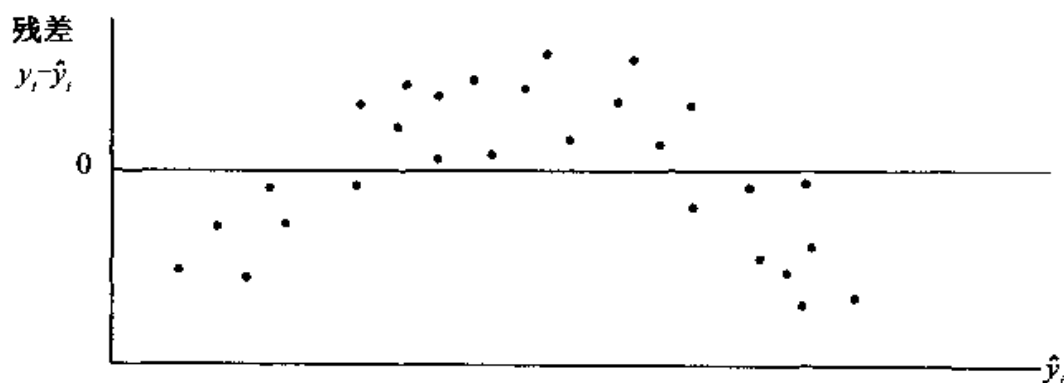


图 11.15 需要高阶模型的残差散点图

同方差假定的检查可以通过  $y$  关于  $x$  的散点图或残差  $y_i - \hat{y}_i$  关于  $x_i$  的散点图。例如, 图 11.16 所显示的残差图表明误差方差对于所有的  $x$  是齐性(相同)的; 图 11.17 的残差图表明误差方差随  $x$  的增加而增大。

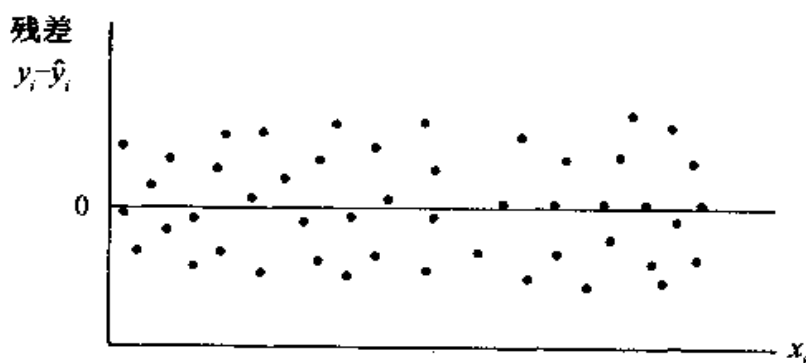


图 11.16 齐性误差方差的残差图

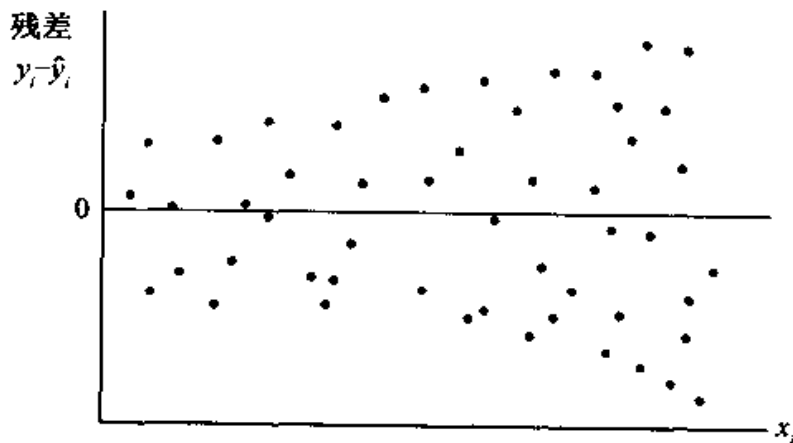


图 11.17 误差方差随  $x$  的增加而增大时的残差图

误差项的独立性和正态性问题将在以后的第十三章讨论。下面通过一个例子来说明到目前为止所学的关于残差的一些问题。

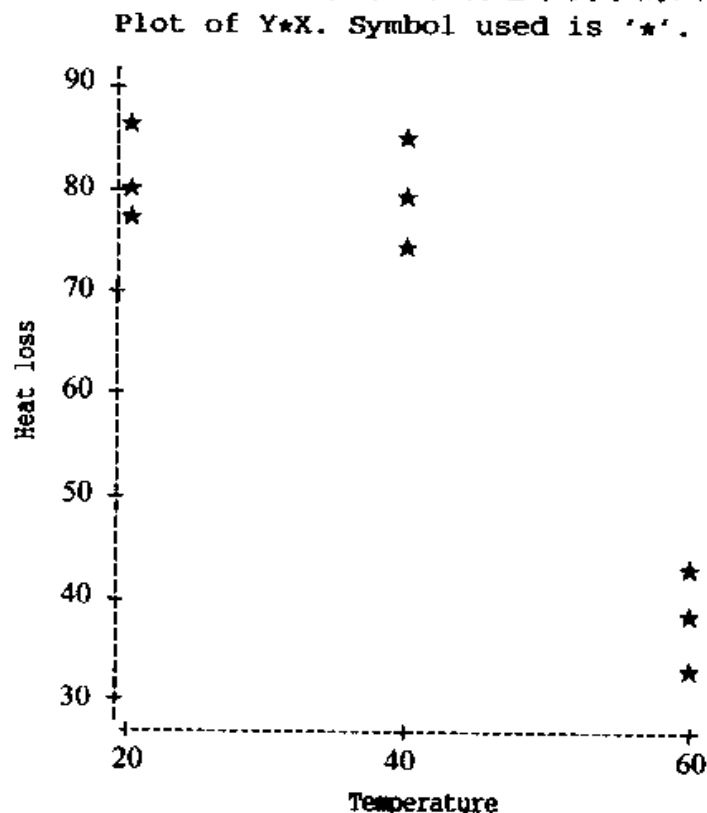
### 例 11.11

某生产新型隔热玻璃窗格的公司要检查其产品的性能, 试验设定了三种室外温度, 每种温度下随机安排了三种不同的玻璃窗格, 检查热量损失的多少。每个试验中, 窗户的温度控制在  $60^{\circ}\text{F}$ , 相对湿度 50%。

| 室外温度( $^{\circ}\text{F}$ ) | 热量损失       |
|----------------------------|------------|
| 20                         | 86, 80, 77 |
| 40                         | 78, 84, 75 |
| 60                         | 33, 38, 43 |

- 绘制数据的散点图。
- 拟合线性回归模型  $y = \beta_0 + \beta_1 x + \epsilon$ , 并检验  $H_0: \beta_1 = 0$  (确定检验的  $p$  值)。
- 计算这九个观测值的  $\hat{y}_i$  和  $y_i - \hat{y}_i$ , 并绘制  $y_i - \hat{y}_i$  关于  $\hat{y}_i$  的散点图。
- 同方差的假定合理吗?

**解答** 这里给出的计算机输出结果可以说明这个例子的四个问题。





Dependent Variable: Y HEAT LOSS

Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 2773.50000     | 2773.50000  | 21.704  | 0.0023 |
| Error   | 7  | 894.50000      | 127.78571   |         |        |
| C Total | 8  | 3668.00000     |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root SE  | 11.30423 | R-square | 0.7561 |
| Dep Mean | 66.00000 | Adj R-sq | 0.7213 |
| C. V.    | 17.12763 |          |        |

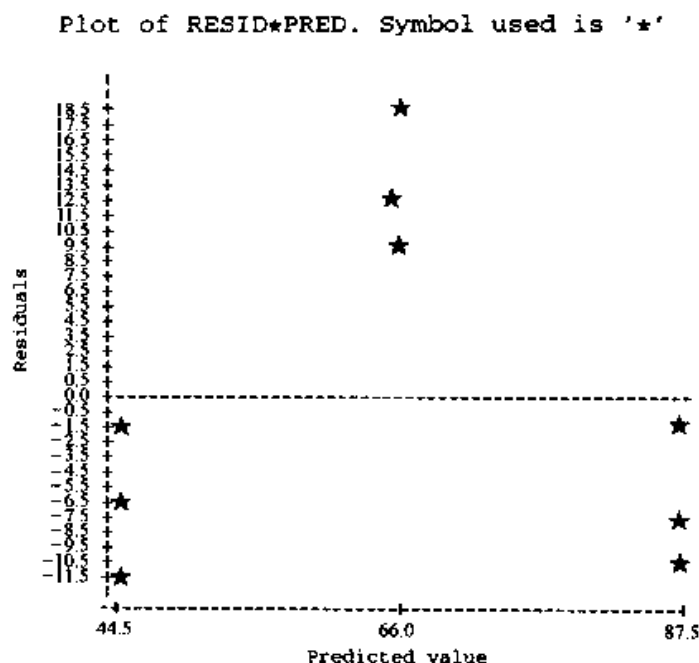
Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|-------------------------|---------|
| INTERCEP | 1  | 109.000000         | 9.96939762     | 10.933                  | 0.0001  |
| X        | 1  | -1.075000          | 0.23074672     | -4.659                  | 0.0023  |

| OBS | X  | Y  | PRED | RESID |
|-----|----|----|------|-------|
| 1   | 20 | 86 | 87.5 | -1.5  |
| 2   | 20 | 80 | 87.5 | -7.5  |
| 3   | 20 | 77 | 87.5 | -10.5 |
| 4   | 40 | 78 | 66.0 | 12.0  |
| 5   | 40 | 84 | 66.0 | 18.0  |
| 6   | 40 | 75 | 66.0 | 9.0   |
| 7   | 60 | 33 | 44.5 | -11.5 |
| 8   | 60 | 38 | 44.5 | -6.5  |
| 9   | 60 | 43 | 44.5 | -1.5  |

a.  $y$  关于  $x$  的散点图无疑显示出一种向下的线性趋势, 也有迹象表明可能存在曲线关系。

b. 线性回归模型似乎对数据拟合得较好, 检验  $H_0: \beta_1 = 0$  的  $p$  值为 0.0023,



说明二者关系是显著的。但这是最佳的模型吗?

- c. 残差  $y_i - \hat{y}_i$  关于预测的  $\hat{y}_i$  的散点图与图 11.15 相似, 说明模型中还需要另外的项。
- d. 因为在  $x = 20$  (最前的三个)、 $x = 40$  (中间的三个) 和  $x = 60$  (最后的三个) 处的残差很容易确定, 因此不需要另外的残差关于  $x$  的散点图来检验同方差的假定。原始的数据散点图和残差图清楚地表明没有这个问题。

如何检验例 11.11 中线性模型明显对数据拟合不足? 当自变量的每个水平都有多个观测值时, 我们可以就模型对数据的拟合不足进行检验, 具体方法是将  $SS$  (残差) 分成两部分: 一个是纯试验误差平方和, 另一个是拟合不足平方和。令  $y_{ij}$  表示自变量的第  $i$  个水平的第  $j$  次观测的  $y$  值, 如果自变量的第  $i$  个水平一共有  $n_i$  次观测, 那么

$$\sum_j (y_{ij} - \bar{y}_i)^2$$

提供了对纯试验误差的一个度量, 这个平方和自由度为  $n_i - 1$ 。

同样对于每个其他的  $x$  的水平, 可以计算基于试验误差的平方和。合并平方和

$$SSP_{\text{exp}} = \sum_{ij} (y_{ij} - \bar{y}_i)^2$$

称为纯试验误差平方和, 其自由度为  $\sum_i (n_i - 1)$ 。用  $SS_{\text{Lack}}$  代表  $SSE$  的剩余部分,

则有

$$SS(\text{残差}) = SSP_{\text{exp}} + SS_{\text{Lack}}$$

如果  $SS(\text{残差})$  在线性回归模型中自由度是  $n - 2$ , 那么  $SS_{\text{Lack}}$  的自由度为  $df = n - 2 - \sum_i (n_i - 1)$ 。

在模型是正确的零假设下, 我们能够获得模型误差的方差  $\sigma_e^2$  的独立的估计, 具体是将  $SSP_{\text{exp}}$  和  $SS_{\text{Lack}}$  分别除以各自的自由度, 这些都称为均方, 分别用  $MSP_{\text{exp}}$  和  $MS_{\text{Lack}}$  表示。

拟合不足的检验小结如下。

### 线性回归中拟合不足的检验

$H_0$ : 线性回归模型是适当的。

$H_a$ : 线性回归模型是不适当的

$$T.S. (\text{检验统计量}): F = \frac{MS_{\text{Lack}}}{MSP_{\text{exp}}}$$

其中

$$MSP_{\text{exp}} = \frac{SSP_{\text{exp}}}{\sum_i (n_i - 1)} = \frac{\sum_{ij} (y_{ij} - \bar{y}_i)^2}{\sum_i (n_i - 1)}$$

$$MS_{\text{Lack}} = \frac{SS(\text{残差}) - SSP_{\text{exp}}}{n - 2 - \sum_i (n_i - 1)}$$

R.R. (拒绝域): 分子自由度  $df_1 = n - 2 - \sum_i (n_i - 1)$ , 分母自由度  $df_2 = \sum_i (n_i - 1)$ , 如果  $F > F_\alpha$  则拒绝  $H_0$ 。

结论: 如果  $F$  检验是显著的, 就表明线性回归方程是不充分的。如果结果是不显著的, 则没有足够的证据表明线性回归模型是不适当的。

### 例 11.12

参照例 11.11, 对线性回归模型进行拟合检验。

**解答** 下表是  $x$  不同的水平对于纯试验误差平方和的贡献。

| $x$ 的水平 | $\bar{y}_i$ | 对于纯试验误差平方和<br>的贡献 $\sum_j (y_{ij} - \bar{y}_i)^2$ | $n_i - 1$ |
|---------|-------------|---|-----------|
| 20      | 81          | 42  | 2         |
| 40      | 79          | 42  | 2         |
| 60      | 38          | 50  | 2         |
| 总和      |             | 134   | 6         |

总结这些结果,有

$$SSP_{\text{exp}} = \sum_{ij} (y_{ij} - \bar{y}_i)^2 = 134$$

例 11.11 的输出结果显示  $SS(\text{残差}) = 894.5$ ; 因此有

$$SS_{\text{Lack}} = SS(\text{残差}) - SSP_{\text{exp}} = 894.5 - 134 = 760.5$$

试验误差平方和的自由度为  $\sum_i (n_i - 1) = 6$ ; 而  $n = 9$ , 因此  $SS_{\text{Lack}}$  的自由度为  $n - 2 - \sum_i (n_i - 1) = 1$ 。所以有

$$MS_{\text{exp}} = \frac{SSP_{\text{exp}}}{6} = \frac{134}{6} = 22.33$$

和

$$MS_{\text{Lack}} = \frac{SSP_{\text{Lack}}}{1} = 760.5$$

拟合检验的  $F$  统计量为

$$F = \frac{MS_{\text{Lack}}}{MS_{\text{exp}}} = \frac{760.5}{22.33} = 34.06$$

这里  $df_1 = 1, df_2 = 6, \alpha = 0.05$ , 如果  $F \geq 5.99$ , 就拒绝  $H_0$ 。

因为计算的  $F$  值大于 5.99, 因此拒绝  $H_0$ , 结论是线性回归模型拟合不足。例 11.11 的散点图证实了非线性。

小结: 在  $x$  的一个或多个水平有不只一个  $y$  值时, 有可能对线性回归模型的拟合不足进行检验。这个检验应该在任何利用拟合的线性回归线所做的推断之前进行。如果拟合不足的检验是显著的, 那么  $x$  的一些高阶多项式可能更适合。数据的散点图和线性回归线的残差图应该可以帮助选择合适的模型。关于选择合适模型的信息将在多元回归分析时讨论(第十二、十三章)。

如果拟合不足的检验是不显著的, 就可以进行基于拟合的线性回归线的推断。

练习

应用

**11.53 (工程)** 某洗衣机用洗涤剂制造厂, 想在新产品投放市场前对产品进行检验, 其中一个关注的领域是洗衣机中洗涤泡沫的高度与洗涤剂投放量之间的函数关系。试验是将一个标准的洗衣机注满水, 再随机放入一定量的洗涤剂, 进行检查。具体的数据如下。

| 高度, $y$    | 投放量, $x$ |
|------------|----------|
| 28.1, 27.6 | 6        |
| 32.3, 33.2 | 7        |
| 34.8, 35.0 | 8        |
| 38.2, 39.4 | 9        |
| 43.5, 46.8 | 10       |

- a. 绘制数据的散点图。
- b. 拟合线性回归模型。
- c. 利用残差图检验可能的拟合不足。

11.54 参照练习 11.53。

- a. 对线性回归模型进行拟合检验。
- b. 如果这个模型是合适的, 求出  $y$  的 95% 预测区间。

## 11.6 逆回归问题(校准)

在试验环境中, 我们常常想要估计相对于某因变量的测量值的自变量的值。这类问题将用因变量  $y$  与自变量  $x$  之间是线性关系的情况来说明。

考虑某个化学过程流动率的测量仪器的校准问题。令  $x$  表示实际的流动率,  $y$  表示在测量仪器上的读数。在校准试验中, 流动率控制在  $n$  个水平  $x_i$ , 记录相应的仪器读数  $y_i$ 。假定模型为以下形式

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

其中  $\epsilon_i$  是独立同分布的正态随机变量, 均值为零, 方差为  $\sigma_\epsilon^2$ 。那么, 利用这  $n$  个数据点  $(x_i, y_i)$ , 我们可以得到最小二乘估计  $\hat{\beta}_0$  和  $\hat{\beta}_1$ 。以后某一时间试验员会从特定的仪器读数  $y$  估计流动率  $x$ 。

$x$  的最常用的估计是用  $y$  代替最小二乘方程  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$  中的  $\hat{y}$ , 并求解这个方程:

$$\hat{x} = \frac{y - \hat{\beta}_0}{\hat{\beta}_1}$$

这里介绍两种不同的逆预测问题。第一种问题是预测  $y$  的某个观测值所对应的  $x$  值; 第二种问题是预测  $y$  的  $m > 1$  观测值(独立的回归数据)的均值所对应的  $x$  值。以下是第一种逆预测问题的求解过程。

情形 1: 根据观测的  $y$  值预测  $x$  值

$$x \text{ 的预测值: } \hat{x} = \frac{y - \hat{\beta}_0}{\hat{\beta}_1}$$

$x$  的  $100(1-\alpha)\%$  预测限:

$$\hat{x}_U = \hat{x} + \frac{1}{1-c^2}[(\hat{x} - \bar{x}) + d]$$

$$\hat{x}_L = \hat{x} - \frac{1}{1-c^2}[(\hat{x} - \bar{x}) - d]$$

其中

$$d = \frac{t_{\alpha/2} s_e}{\hat{\beta}_1} \sqrt{\frac{n+1}{n}(1-c^2) + \frac{(\hat{x} - \bar{x})^2}{S_{xx}}}, s_e^2 = \frac{SSE}{n-2}, c^2 = \frac{t_{\alpha/2}^2 s_e^2}{\hat{\beta}_1^2 S_{xx}}$$

而  $t_{\alpha/2}$  是基于自由度  $df = n-2$ 。

注意由于

$$t = \frac{\hat{\beta}_1}{s_e / \sqrt{S_{xx}}}$$

是假设  $H_0: \beta_1 = 0$  的检验统计量,  $c = t_{\alpha/2}/t$ , 我们需要  $|t| > t_{\alpha/2}$ ; 也就是说  $\beta_1$  必须是显著不为零, 于是  $c^2 < 1$  和  $0 < (1-c^2) < 1$ 。 $x$  与  $y$  之间的线性关系越强烈,  $(1-c^2)$  越大, 这样预测区间的宽度就越窄。同样要注意当  $x$  接近试验区域的中心  $\bar{x}$  时, 我们就能得到  $x$  的好的预测。如果  $x$  与  $y$  之间关系的线性性较弱, 那么在试验区域的边界点的预测就会产生非常宽的界限。

### 例 11.13

某工程师想要校准用于液体肥皂生产线上的流量检测仪器。试验中, 采用了 10 种不同的流动率, 记录下相应的仪器读数, 具体数据如下所示。利用这些数据, 确定仪器读数为 4.0 时的实际流动率  $x$  的 95% 预测区间。

**解答** 从数据中, 可以得到  $S_{xy} = 74.35$ ,  $S_{xx} = 82.5$ ,  $S_{yy} = 67.0654$ 。因此  $\hat{\beta}_1 = 74.35/82.5 = 0.9012$ ,  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 5.45 - (0.9012)(5.5) = 0.4934$ , 且  $SS(\text{残差}) = S_{yy} - \hat{\beta}_1 S_{xy} = 67.065 - (0.9012)(74.35) = 0.0608$ 。 $\sigma_e^2$  的估计基于自由度  $n-2=8$ 。

$$s_e^2 = \frac{SS(\text{残差})}{n-2} = \frac{0.0608}{8} = 0.0076$$

$$s_e = 0.0872$$

| 校准问题的数据  |           |
|----------|-----------|
| 流动率, $x$ | 仪器读数, $y$ |
| 1        | 1.4       |
| 2        | 2.3       |
| 3        | 3.1       |
| 4        | 4.2       |
| 5        | 5.1       |
| 6        | 5.8       |
| 7        | 6.8       |
| 8        | 7.6       |
| 9        | 8.7       |
| 10       | 9.5       |

对于  $\alpha = 0.05$ , 查表  $df = 8$ ,  $\alpha = 0.025$  的  $t$  值是 2.306。

$$c^2 = \frac{t_{\alpha/2, S_y}^2}{\hat{\beta}_1^2 S_{xx}} = \frac{(2.306)^2 (0.0076)}{(0.9012)^2 (82.5)} = 0.0006$$

且  $1 - c^2 = 0.9994$ 。利用  $\hat{x} = 3.8910$ , 当  $y = 4.0$  时,  $x$  的上下预测限分别是

$$\begin{aligned}\hat{x}_U &= 5.5 + \frac{1}{0.9994} \left[ -1.6090 + \frac{2.306}{0.9012} \sqrt{\frac{11}{10} (0.9994) + \frac{(-1.6090)^2}{82.5}} \right] \\ &= 5.5 + \frac{1}{0.9994} [-1.6090 + 0.2373] = 4.1274\end{aligned}$$

$$\hat{x}_L = 5.5 + \frac{1}{0.9994} [-1.6090 - 0.2373] = 3.65267$$

因此,  $x$  的 95% 预测限分别是 3.65 和 4.13。图 11.18 显示了这些界限。

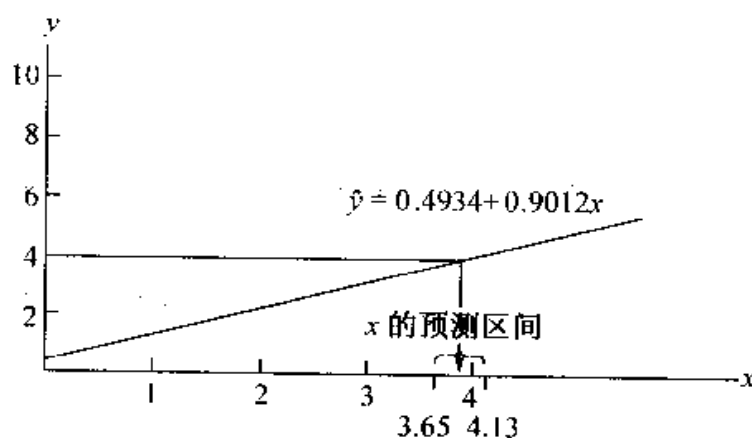


图 11.18 当  $y = 4.0$  时  $x$  的预测区间

下面接着小结第二种逆预测问题的求解过程。

### 情形 2: 根据观测的 $y$ 的 $m$ 个值预测 $x$ 值

预测相应于  $m$  个独立的  $y$  值的均值的  $100P\%$  的  $x$  值

$$x \text{ 的预测值: } \hat{x} = \frac{P\bar{y}_m - \hat{\beta}_0}{\hat{\beta}_1}$$

$$\hat{x}_U = \bar{x} + \frac{1}{1-c^2} [(\hat{x} - \bar{x}) + g]$$

$$\hat{x}_L = \bar{x} + \frac{1}{1-c^2} [(\hat{x} - \bar{x}) - g]$$

其中

$$g = \frac{t_{\alpha/2}}{\hat{\beta}_1} \sqrt{(s_y^2 P^2 + \frac{s_\varepsilon^2}{n})(1-c^2) + \frac{(\hat{x} - \bar{x})^2 s_\varepsilon^2}{S_{xx}}}$$

$\bar{y}_m$  和  $s_y$  分别是  $m$  个独立的  $y$  值的均值和标准差。

至此我们已经介绍了分析本章开始时的案例中的数据所需要的方法。

**大肠杆菌浓度案例的数据分析** 研究人员感兴趣的是评价在确定牛肉样本中大肠杆菌(*E. coli*)浓度的 HEC 和 HGMP 两种方法的一致程度。如果这两组读数存在很强的关系,那么研究人员就可以得到逆回归方程,这样可以从 HEC 读数中预测 HGMP 读数。首先我们求出 HEC 为因变量, HGMP 为自变量的回归关系,这是因为在确定大肠杆菌浓度的方法中 HGMP 具有公认的可靠性。

下面给出分析 17 对大肠杆菌浓度数据的计算机输出结果以及残差图。

Dependent Variable: HEC

HEC:METHOD

#### Analysis of Variance

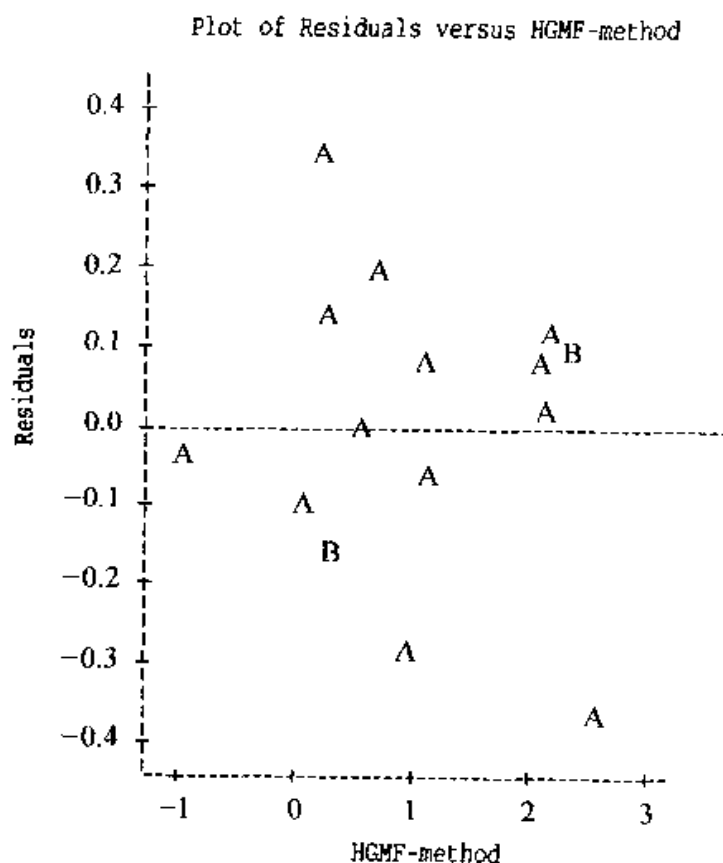
| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 14.22159       | 14.22159    | 441.816 | 0.0001 |
| Error   | 15 | 0.48283        | 0.03219     |         |        |
| C Total | 16 | 14.70442       |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 0.17941  | R-square | 0.9672 |
| Dep Mean | 1.07471  | Adj R-sq | 0.9650 |
| C. V.    | 16.69413 |          |        |



| Variable | DF | Parameter Estimate | Standard Error | T for H <sub>0</sub> : Parameter = 0 | Prob >  T |
|----------|----|--------------------|----------------|--------------------------------------|-----------|
| INTERCEP | 1  | -0.023039          | 0.06797755     | -0.339                               | 0.7394    |
| HGMF     | 1  | 0.915685           | 0.04356377     | 21.019                               | 0.0001    |



$R^2$  是 0.9672, 表明 HEC 浓度与 HGMF 浓度之间存在很强的线性关系。残差图的检查没有表明模型需要高阶项和异方差的出现。HEC 浓度关于 HGMF 浓度的最小二乘方程为

$$\widehat{\text{HEC}} = -0.023 + 0.9157 * \text{HGMF}$$

因此, 我们可以通过检验下面的假设来评价确定大肠杆菌浓度的这两种方法是否存在真正的关系:

$$H_0: \beta_0 = 0, \beta_1 = 1 \quad H_a: \beta_0 \neq 0 \text{ 或 } \beta_1 \neq 1$$

如果  $H_0$  被接受, 那么就强烈地表明关系  $\widehat{\text{HEC}} = 0 + 1 * \text{HGMF}$  是有效的, 也就是说 HEC 和 HGMF 产生相同的大肠杆菌浓度的读数。从输出结果, 检验  $H_0: \beta_0 = 0$  的  $p$  值为 0.7394, 而  $H_0: \beta_1 = 1$  的检验可以利用检验统计量:

$$t = \frac{\hat{\beta}_1 - 1}{\widehat{SE}(\hat{\beta}_1)} = \frac{0.915685 - 1}{0.04356377} = -1.935$$

这个检验统计量的  $p$  值是  $Pr(|t_{15}| \geq 1.935) = 0.0721$ 。为了获得总体显著性水平为  $\alpha = 0.05$ , 我们取假设  $H_0: \beta_0 = 0$  和  $H_0: \beta_1 = 1$  的显著性水平分别为  $\alpha = 0.025$ ; 也就是说只要两个假设检验的  $p$  值有一个小于 0.025, 我们就拒绝这两个假设。由于两个  $p$  值是 0.7394 和 0.0721, 我们不能拒绝这两个零假设, 因此得出的结论是这些数据不支持关于 HEC 和 HGMP 产生明显不同的大肠杆菌浓度的读数的假设。

由于只有 17 对 HEC 和 HGMP 的读数, 我们将建立校准曲线以确定用 HEC 浓度读数预测 HGMP 浓度读数的准确程度。利用校准方程, 有

$$\widehat{\text{HGMP}} = (\text{HEC} + 0.023)/0.9157$$

95% 预测区间为

$$\widehat{\text{HGMP}}_L = 1.1988 + 1.0104 * (\widehat{\text{HGMP}} - 1.1988 - d)$$

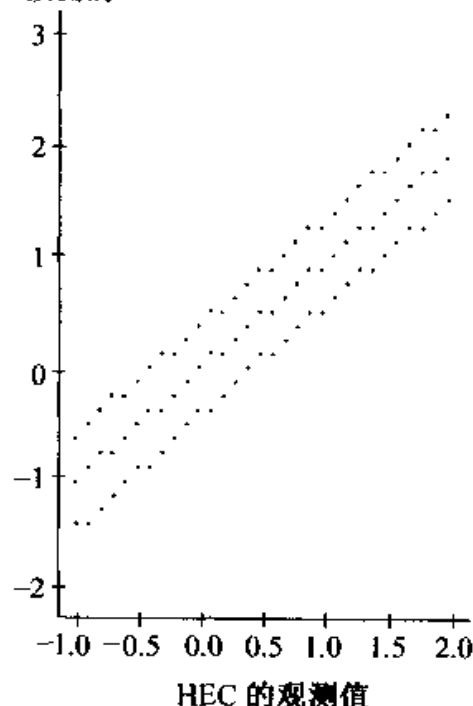
$$\widehat{\text{HGMP}}_U = 1.1988 + 1.0104 * (\widehat{\text{HGMP}} - 1.1988 + d)$$

这里

$$d = 0.4175 \sqrt{1.0479 + (\widehat{\text{HGMP}} - 1.1988)^2 / 16.9612}.$$

HEC 的观测值与相应的 HGMP 预测值的图,  
以及 95% 预测限

预测的 HGMP



下一步绘制  $\widehat{HGMF}_L$  和  $\widehat{HGMF}_U$  关于 HEC 在 -1 到 2 之间的散点图, 从图中可以看出用所观测的 HEC 读数预测 HGMF 读数的范围。

对大多数的 HEC, 95% 预测区间的宽度比一个单位稍微小一点。因此 HEC 关于大肠杆菌浓度的测定值在 -1 到 2 之间就可以导出相应的 HGMF 测定值的 95% 预测区间, 但这个预测的准确程度是不能接受的。减少预测区间的宽度的一个方法是在研究中获得更多的观测值, 超过现在的 17 个, 这个方法只要在新的研究中 HEC 读数和 HGMF 读数之间保持相同程度的关系就是可行的。

练习

应用

**11.55 (农业)** 某个林业员会在木材销售前估计某特定地点的木材量(以立方英尺计)。现在由于他的业务扩大, 需要培训一个人来帮助预测树木的立方英尺数。林业员决定校准他的助手的预测值, 于是随机选择一些很快就要砍伐的树木作为样本。对于每一棵树, 助手估计了一个立方英尺数  $y$ , 等这棵树砍伐后就可以得到实际的立方英尺数  $x$ 。从这些数据中, 林业员得到了以下模型的校准曲线

$$y = \beta_0 + \beta_1 x + \varepsilon$$

在将来, 林业员可以利用校准曲线修正他助手对木材量的估计。样本数据总结如下:

| 树   | 1  | 2  | 3 | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|-----|----|----|---|----|----|----|----|----|----|----|
| 估计值 | 12 | 14 | 8 | 12 | 17 | 16 | 14 | 14 | 15 | 17 |
| 实际值 | 13 | 14 | 9 | 15 | 19 | 20 | 16 | 15 | 17 | 18 |

用最小二乘法拟合出校准曲线。数据表明斜率是显著大于 0 吗? 其中  $\alpha = 0.05$ 。

**11.56** 参照练习 11.55。

- 当助手估计某树有 13 立方英尺时, 预测这棵树的实际体积。
- 确定(a)中树的实际体积的 95% 预测区间。

**11.57 (医学)** 某研究人员在某种利尿剂药品的研究中, 收集了 24 个病人的数据以检查剂量(药量, DOSE)与总尿量(CUMVOL)之间的关系。数据显示在下面计算机输出结果中。数据的最初拟合显示剂量 DOSE 和 CUMVOL 之间的非线性关系。研究人员决定分别作剂量的对数变换和 CUMVOL/100 的平方根的反正弦变换, 在计算机输出结果中分别表示为 LOG(DOS)和 TRANS.CUMVOL。

- 求出线性回归方程, 确定自变量和因变量。
- 利用输出结果预测当  $y$  分别为 10、14、19  $\text{cm}^3$  时的剂量。每种情况相应的 99% 预测限是多少?

## OUTPUT FOR EXERCISE 11.57

| OBS | DOSE  | LOG (DOSE) | CUMVOL | TRANS. CUMVOL |
|-----|-------|------------|--------|---------------|
| 1   | 6.00  | 1.79176    | 7.1    | 0.26972       |
| 2   | 6.00  | 1.79176    | 11.5   | 0.34598       |
| 3   | 6.00  | 1.79176    | 8.4    | 0.29405       |
| 4   | 6.00  | 1.79176    | 8.0    | 0.28676       |
| 5   | 6.00  | 1.79176    | 9.4    | 0.31161       |
| 6   | 6.00  | 1.79176    | 12.0   | 0.35374       |
| 7   | 9.00  | 2.19722    | 13.2   | 0.37183       |
| 8   | 9.00  | 2.19722    | 14.7   | 0.39348       |
| 9   | 9.00  | 2.19722    | 12.7   | 0.36438       |
| 10  | 9.00  | 2.19722    | 15.5   | 0.40465       |
| 11  | 9.00  | 2.19722    | 18.4   | 0.44333       |
| 12  | 9.00  | 2.19722    | 14.4   | 0.38923       |
| 13  | 13.50 | 2.60269    | 12.1   | 0.35528       |
| 14  | 13.50 | 2.60269    | 15.8   | 0.40878       |
| 15  | 13.50 | 2.60269    | 13.8   | 0.38061       |
| 16  | 13.50 | 2.60269    | 20.4   | 0.46863       |
| 17  | 13.50 | 2.60269    | 22.7   | 0.49661       |
| 18  | 13.50 | 2.60269    | 17.0   | 0.42499       |
| 19  | 20.25 | 3.00815    | 19.8   | 0.46114       |
| 20  | 20.25 | 3.00815    | 15.6   | 0.40603       |
| 21  | 20.25 | 3.00815    | 25.3   | 0.52705       |
| 22  | 20.25 | 3.00815    | 13.5   | 0.37624       |
| 23  | 20.25 | 3.00815    | 24.8   | 0.52129       |
| 24  | 20.25 | 3.00815    | 20.9   | 0.47481       |
| 25  | 10.00 | 2.30259    |        |               |
| 26  | 14.00 | 2.63906    |        |               |
| 27  | 19.00 | 2.94444    |        |               |

## OUTPUT FOR EXERCISE 11.57

Dependent Variable: y

TRANSFORMED CUMVOL

Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 0.06922        | 0.06922     | 32.750  | 0.0001 |
| Error   | 22 | 0.04650        | 0.00211     |         |        |
| C Total | 23 | 0.11572        |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 0.04597  | R-square | 0.5982 |
| Dep Mean | 0.39709  | Adj R-sq | 0.5799 |
| C.V.     | 11.57773 |          |        |

## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|----------------------------|---------|
| INTERCEP | 1  | 0.112770           | 0.05056109     | 2.230                      | 0.0362  |
| x        | 1  | 0.118470           | 0.02070143     | 5.723                      | 0.0001  |

| OBS | X       | Y       | PRED    | L95PRED | U95PRED | L95MEAN | U95MEAN |
|-----|---------|---------|---------|---------|---------|---------|---------|
| 1   | 1.79176 | 0.26972 | 0.32504 | 0.22429 | 0.42579 | 0.29247 | 0.35761 |
| 2   | 1.79176 | 0.34598 | 0.32504 | 0.22429 | 0.42579 | 0.29247 | 0.35761 |
| 3   | 1.79176 | 0.29405 | 0.32504 | 0.22429 | 0.42579 | 0.29247 | 0.35761 |
| 4   | 1.79176 | 0.28676 | 0.32504 | 0.22429 | 0.42579 | 0.29247 | 0.35761 |
| 5   | 1.79176 | 0.31161 | 0.32504 | 0.22429 | 0.42579 | 0.29247 | 0.35761 |
| 6   | 1.79176 | 0.35374 | 0.32504 | 0.22429 | 0.42579 | 0.29247 | 0.35761 |
| 7   | 2.19722 | 0.37183 | 0.37307 | 0.27537 | 0.47077 | 0.35175 | 0.39439 |
| 8   | 2.19722 | 0.39348 | 0.37307 | 0.27537 | 0.47077 | 0.35175 | 0.39439 |
| 9   | 2.19722 | 0.36438 | 0.37307 | 0.27537 | 0.47077 | 0.35175 | 0.39439 |
| 10  | 2.19722 | 0.40465 | 0.37307 | 0.27537 | 0.47077 | 0.35175 | 0.39439 |
| 11  | 2.19722 | 0.44333 | 0.37307 | 0.27537 | 0.47077 | 0.35175 | 0.39439 |
| 12  | 2.19722 | 0.38923 | 0.37307 | 0.27537 | 0.47077 | 0.35175 | 0.39439 |
| 13  | 2.60269 | 0.35528 | 0.42111 | 0.32341 | 0.51881 | 0.39979 | 0.44243 |
| 14  | 2.60269 | 0.40878 | 0.42111 | 0.32341 | 0.51881 | 0.39979 | 0.44243 |
| 15  | 2.60269 | 0.38061 | 0.42111 | 0.32341 | 0.51881 | 0.39979 | 0.44243 |
| 16  | 2.60269 | 0.46863 | 0.42111 | 0.32341 | 0.51881 | 0.39979 | 0.44243 |
| 17  | 2.60269 | 0.49661 | 0.42111 | 0.32341 | 0.51881 | 0.39979 | 0.44243 |
| 18  | 2.60269 | 0.42499 | 0.42111 | 0.32341 | 0.51881 | 0.39979 | 0.44243 |
| 19  | 3.00815 | 0.46114 | 0.46914 | 0.36839 | 0.56990 | 0.43658 | 0.50171 |
| 20  | 3.00815 | 0.40603 | 0.46914 | 0.36839 | 0.56990 | 0.43658 | 0.50171 |
| 21  | 3.00815 | 0.52706 | 0.46914 | 0.36839 | 0.56990 | 0.43658 | 0.50171 |
| 22  | 3.00815 | 0.37624 | 0.46914 | 0.36839 | 0.56990 | 0.43658 | 0.50171 |
| 23  | 3.00815 | 0.52129 | 0.46914 | 0.36839 | 0.56990 | 0.43658 | 0.50171 |
| 24  | 3.00815 | 0.47481 | 0.46914 | 0.36839 | 0.56990 | 0.43658 | 0.50171 |
| 25  | 2.30259 | .       | 0.38556 | 0.28816 | 0.48296 | 0.36565 | 0.40546 |
| 26  | 2.63906 | .       | 0.42542 | 0.32757 | 0.52327 | 0.40341 | 0.44742 |
| 27  | 2.94444 | .       | 0.46160 | 0.36152 | 0.56168 | 0.43118 | 0.49201 |

|                            |        |
|----------------------------|--------|
| Sum of Residuals           | 0      |
| Sum of Squared Residuals   | 0.0465 |
| Predicted Resid SS (Press) | 0.0560 |

**11.58** 参照练习 11.57。如果研究人员希望预测效果相当于四个病人服用某种知名利尿剂的效果的 50% (或 75%) 的这种利尿剂的剂量, 预测  $x$  并给出每种情形的合适的预测限。

## 11.7 相 关

一旦求出了回归线  $\hat{y} = \beta_0 + \beta_1 x$ , 我们就需要衡量它对实际值预测的好坏。一个方法是看在这个问题的背景下残差标准差的大小, 约 95% 的预测误差落在  $\pm 2s_e$  内。例如, 假如我们要预测某个化学过程的产出, 其产出范围是 0.50 到 0.94。如果一个回归模型的残差标准差为 0.01, 那么可以预见大多数预测误差在  $\pm 0.02$  内——这是比较准确的。但是, 如果残差标准差为 0.08, 那么大多数预测误差  $\pm 0.16$  内, 而产出的极差只有  $0.94 - 0.50 = 0.44$ , 因此预测结果不能令人满意。这个方法需要我们对研究的背景有很好的了解, 而更一般的方法是基于相关性的概念。

假定我们比较两种预测方法的平方预测误差, 一种是回归模型, 另一种是忽略模型而总是用  $y$  的平均值来预测。前几节的公路重新铺设例子中, 如果给出了公路的英里数  $x_i$ , 就可以利用预测方程  $\hat{y} = 2.0 + 3.0x$  预测项目的成本。实际值与预测值之间的差异, 也就是残差, 测量了预测误差。这些误差汇总成残差平方和,  $SS(\text{残差}) = \sum (y_i - \hat{y}_i)^2$ , 对于这些数据来说是 44。如果没有给出公路的英里数  $x_i$ ,  $y$  的最小平方误差预测就是均值  $\bar{y} = 14$ , 这时预测误差平方和是  $\sum (y_i - \bar{y})^2 = SS(\text{总和}) = 224$ 。那么误差减少的比例是

$$\frac{SS(\text{总和}) - SS(\text{残差})}{SS(\text{总和})} = \frac{224 - 44}{224} = 0.804$$

也就是说, 利用回归模型可以减少预测误差平方和的 80.4%, 这就说明公路重新铺设的英里数与项目的成本之间有很强的关系。

上述误差减少的比例是与  $x$  和  $y$  之间的**相关系数**紧密相联的。**相关系数**是  $x$  和  $y$  之间的线性关系的一个度量。 $x$  和  $y$  的相关性越强,  $x$  对  $y$  的预测就越好。

已知  $n$  对观测值  $(x_i, y_i)$ , 由以下公式可计算样本相关系数

$$r_{yx} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{S_{xx}S_{yy}}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

这里  $S_{xy}$  和  $S_{xx}$  的定义如前, 且

$$S_{yy} = \sum (y_i - \bar{y})^2 = SS(\text{总和})$$

这个例子中

$$r_{yx} = \frac{60}{\sqrt{(20)(224)}} = 0.896$$

一般地,如果当  $x$  增加时  $y$  也增加,那么相关系数  $r_{yx}$  是正的;如果当  $x$  增加时  $y$  减小,那么相关系数  $r_{yx}$  是负的;如果  $x$  的变化和  $y$  的变化之间没有关系,或存在非线性关系使得(当  $x$  增加时)  $y$  的增加和  $y$  的减少相互抵消,那么  $r_{yx} = 0$ 。

图 11.19 显示了  $r$  值的四种情形。图 11.19(d)中,  $y$  和  $x$  之间存在很强的关系,但  $r = 0$ ,这是近似线性关系的正负部分对称、相互抵消的结果。当  $r = 0$  时,只表示  $y$  和  $x$  之间没有“线性”关系,但高阶的(非线性)关系可能存在。这种情况说明了在散点图中画出数据的重要性。后面的第十二章中,我们将介绍一些对  $y$  和  $x$  之间的非线性关系进行建模的方法。

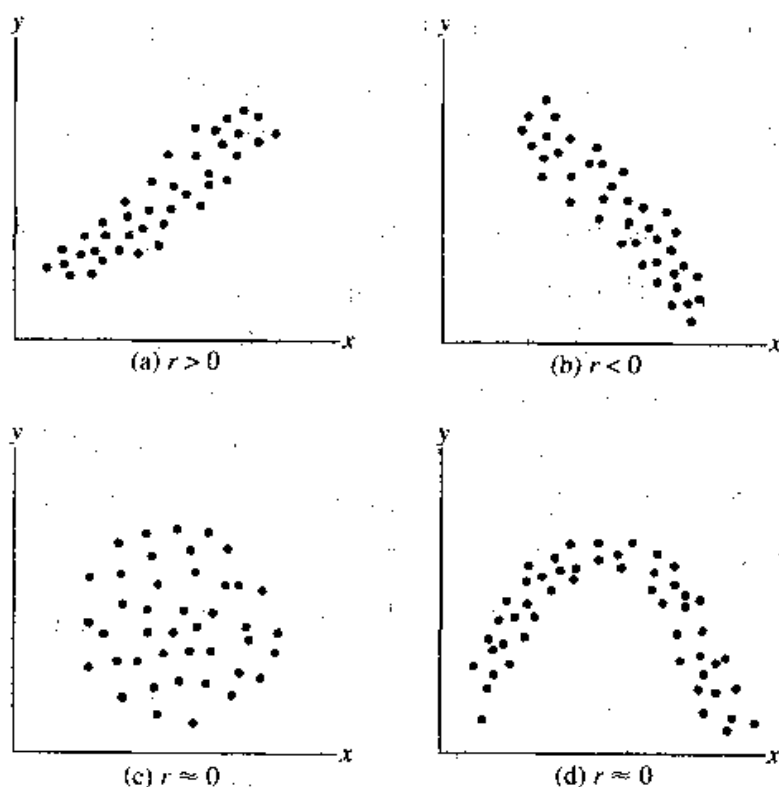


图 11.19  $r$  的解释

### 例 11.14

考虑以下数据:

|       |    |    |    |    |    |    |    |    |    |
|-------|----|----|----|----|----|----|----|----|----|
| $y$ : | 25 | 41 | 47 | 59 | 54 | 56 | 49 | 43 | 30 |
| $x$ : | 10 | 20 | 20 | 30 | 30 | 30 | 40 | 40 | 50 |

a. 二者的相关系数是正还是负?

b. 计算相关系数。

**解答**

a. 注意  $x$  在 10 与 50 之间增加时,  $y$  是先增加, 然后减低, 因此二者的相关系数应该小。又由于  $y$  的值没有减少到开始的地方, 因此二者的相关系数应该是正的。

b. 经过简单的计算, 样本均值是  $\bar{x} = 30.0000$ ,  $\bar{y} = 44.8889$ 。

$$S_{xx} = (10 - 30.0000)^2 + \cdots + (50 - 30.0000)^2 = 1,200$$

$$S_{yy} = (25 - 44.8889)^2 + \cdots + (30 - 44.8889)^2 = 1,062.8889$$

$$S_{xy} = (10 - 30.0000)(25 - 44.8889) + \cdots + (50 - 30.0000)(30 - 44.8889) \\ = 140$$

$$r_{yx} = \frac{140}{\sqrt{(1,200)(1,062.8889)}} = 0.1240$$

二者的相关系数确实是小的正数。

相关性和回归预测能力是紧密相联的。前面所定义的回归的误差减少的比例称为**决定系数**。简单地说决定系数就是相关系数的平方,

$$r_{yx}^2 = \frac{SS(\text{总和}) - SS(\text{残差})}{SS(\text{总和})}$$

也就是误差减少的比例。公路重新铺设例子中,  $r_{yx} = 0.896$ ,  $r_{yx}^2 = 0.804$ 。

相关系数为零表明方程没有预测价值; 也就是说预测  $y$  时已知  $x$  或未知  $x$  的效果都一样。相关系数为 1 或 -1 表明预测能力最好——100% 的误差减少都归于  $x$  的知识。相关系数按惯例应该用它的平方——判定系数来解释, 那么, 相关系数为 0.3 就是说只有 9% 的预测误差减少。许多书和大多数计算机程序都用方程

$$SS(\text{总和}) = SS(\text{残差}) + SS(\text{回归})$$

其中

$$SS(\text{回归}) = \sum (\hat{y}_i - \bar{y})^2$$

因为这个方程可以表示为  $SS(\text{残差}) = (1 - r_{yx}^2)SS(\text{总和})$ , 因此可以推出  $SS(\text{回归}) = r_{yx}^2 SS(\text{总和})$ , 这就再次说明了  $x$  的回归解释了  $y$  的总平方误差的比例为  $r_{yx}^2$ 。

**例 11.15**

求出例 11.14 的数据的  $SS(\text{总和})$ ,  $SS(\text{回归})$  和  $SS(\text{残差})$ 。

**解答**  $SS(\text{总和}) = S_{yy}$ , 而在例 11.14 中计算得  $S_{yy} = 1062.8889$ 。并且可得  $r_{yx} = 0.1240$ , 因此有  $r_{yx}^2 = (0.1240)^2 = 0.0154$ 。由于  $SS(\text{回归}) = r_{yx}^2 SS(\text{总和})$ , 因



此有  $SS(\text{回归}) = (0.154)(1062.8889) = 16.3685$ 。因为  $SS(\text{残差}) = SS(\text{总和}) - SS(\text{回归})$ , 所以  $SS(\text{残差}) = 1062.8889 - 16.3685 = 1046.5294$ 。

注意  $SS(\text{回归})$  和  $r^2_{yx}$  都非常小, 这就说明  $x$  不是  $y$  的好预测变量。不过事实上  $x$  和  $y$  之间的关系是很强的非线性,  $x$  的一个线性方程不能很好地预测  $y$ , 但一个非线性的方程可能会更好。

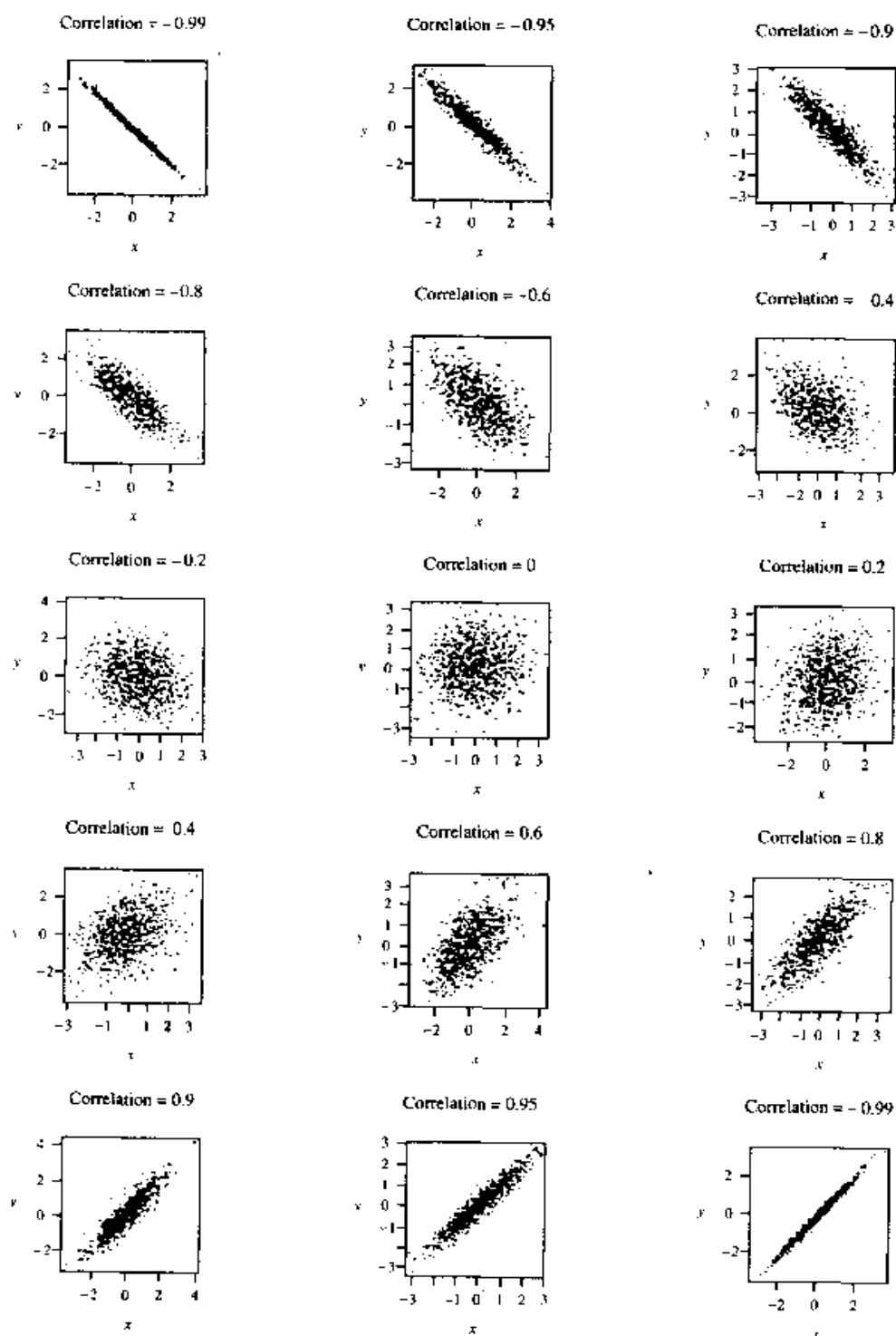


图 11.20 二元正态总体的容量为 1000 的样本

$r_{yx}$  为什么值时表明  $y$  和  $x$  之间有“很强”的关系？图 11.20 显示了 15 个散点图，分别是用从 15 个二元正态总体中随机选择的 1000 组  $(x_i, y_i)$  画出的，在这 15 个二元正态总体中两个变量的相关系数界于  $-0.99$  和  $0.99$  之间。从散点图中可以看到除非  $|r_{yx}|$  大于  $0.6$ ，否则只有很小的趋势。

样本相关系数  $r_{yx}$  是总体相关系数  $\rho_{yx}$  的估计和显著性检验的基础。统计推断总是基于一些假定的。回归分析的假定条件——即  $x$  和  $y$  之间存在线性关系，围绕回归线是同方差的——也是相关系数检验的假定条件。回归分析中， $x$  是当作已知的常数；而在相关系数检验中， $x$  是当作随机选择的（回归推断中是考虑在给定  $x$  为样本值的条件下的推断）。如果  $x$  不是当作随机抽取的，那么相关系数估计就可能是有偏的。某些书里，还假定  $x$  是来自于正态总体的。我们这里所做的统计推断并不依赖于总体正态的假定。

$\rho_{yx}$  的估计中最基本的推断问题是可能的偏差。当  $x$  的值是预先确定时，就如回归分析里一样，这个问题就会出现。 $x$  的选择会系统地增加或减少样本相关系数。一般情况下，宽区域的  $x$  值倾向于增加相关系数的大小，而窄区域的  $x$  值是减少相关系数的大小，这种影响显示在图 11.21 中。如果考虑散点图中所有的点，则  $x$  和  $y$  之间存在明显的强的关系。但是如果只考虑两条垂直虚线之间的  $x$  值，去掉外面的部分，那么样本相关系数（和决定系数）就小很多。相关系数会受  $x$  的选择的系统影响；尽管  $x$  值的区域的部分改变可能使残差标准差随机地有所改变，但残差标准差不会有系统地改变。因此，当你需要确定线性回归线预测  $y$  的准确程度时，考虑残差标准差  $s_e$  和斜率的大小是一个好的方法。

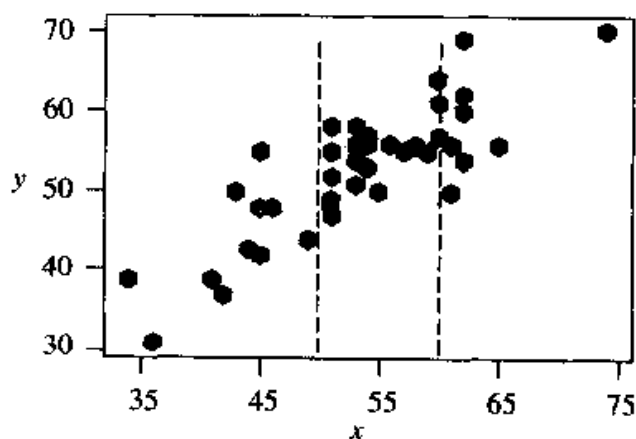


图 11.21 限制  $x$  值的区域对样本相关系数的影响

#### 例 11.16

以下数据是某个公司的 12 个数据录入员的工作能力  $y$  和才能测试得分  $x$ ：

y:41    39    47    51    43    40    57    46    50    59    61    52  
x:24    30    33    35    36    36    37    37    38    40    43    49

如果只考虑最后六个数据,二者的相关系数是大一些还是小一些?

#### Simple Regression Analysis

Linear model:  $y = 20.5394 + 0.775176 * x$

Table of Estimates

|           | Estimate | Standard Error | t Value | P Value |
|-----------|----------|----------------|---------|---------|
| Intercept | 20.5394  | 10.7251        | 1.92    | 0.0845  |
| Slope     | 0.775176 | 0.289991       | 2.67    | 0.0234  |

R-squared = 41.68 %

Correlation coeff. = 0.646

Standard error of estimation = 5.99236

File subset has been turned on, based on  $x \geq 37$ .

#### Simple Regression Analysis

Linear model:  $y = 44.7439 + 0.231707 * x$

Table of Estimates

|           | Estimate | Standard Error | t Value | P Value |
|-----------|----------|----------------|---------|---------|
| Intercept | 44.7439  | 24.8071        | 1.80    | 0.1456  |
| Slope     | 0.231707 | 0.606677       | 0.38    | 0.7219  |

R-squared = 3.52 %

Correlation coeff. = 0.188

Standard error of estimation = 6.34357

**解答** 对于所有的 12 个观测值,输出结果显示样本相关系数为 0.646,残差标准差的标识为“standard error of estimation”,显示是 5.992。关于六个最大  $x$  的得分,是  $x$  值大于 37 的子集,相关系数为 0.188,残差标准差为 6.344。从所有的 12 个观测值到六个具有最大的  $x$  值的观测值,相关系数有很大的减少,但残差标准差几乎没有多少变化。

就像对  $\beta_1$  作统计检验,我们也可以对  $\rho_{yx}$  作统计检验。

### $\rho_{yx}$ 统计检验小结

假设

$$1. H_0: \rho_{yx} \leq 0, \quad H_a: \rho_{yx} > 0$$

$$2. H_0: \rho_{yx} \geq 0, \quad H_a: \rho_{yx} < 0$$

$$3. H_0: \rho_{yx} = 0, \quad H_a: \rho_{yx} \neq 0$$

$$T.S. (\text{检验统计量}): t = r_{yx} \frac{\sqrt{n-2}}{\sqrt{1-r_{yx}^2}}$$

R.R. (拒绝域): 自由度  $df$  为  $n-2$ , 第一类错误为  $\alpha$

$$1. t > t_{\alpha}$$

$$2. t < -t_{\alpha}$$

$$3. |t| > t_{\alpha/2}$$

检查假定并得出结论。

例 11.5 中(从公路重新铺设的英里数预测项目的成本)检验了斜率真值为零的假设, 结果  $t$  统计量是 3.50。对于这些数据, 可以计算  $r_{yx} = 0.896421$ ,  $r_{yx}^2 = 0.803571$ 。因此, 相关系数的  $t$  统计量是

$$\frac{0.896\sqrt{3}}{\sqrt{1-0.803571}} = 3.50$$

由  $r$  和最小二乘方程  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$  的斜率  $\hat{\beta}_1$  的公式得出下列关系:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \sqrt{\frac{S_{yy}}{S_{xx}}} = r_{yx} \sqrt{\frac{S_{yy}}{S_{xx}}}$$

因此, 无论用什么形式, 斜率和相关系数的  $t$  检验给出一致的结果, 对于  $x$  值的任何选择  $t$  检验都是有效的。前面提到的偏差不会影响相关系数的符号。

### 例 11.17

对例 11.16(所有观测值)的数据进行零假设为零相关系数和零斜率的  $t$  检验。利用一个合适的单侧备择假设。

**解答** 首先, 合适的  $H_a$  应该是  $\rho_{yx} > 0$  (和  $\beta_1 > 0$ )。如果才能测试得分对工作能力得分有正的相关系数该多好呀! 例 11.16 中,  $n = 12$ ,  $r_{yx} = 0.646$ , 且

$$t = \frac{0.646 \sqrt{12-2}}{\sqrt{1-(0.646)^2}} = 2.68$$

由于这个值在  $df = 10$ ,  $\alpha = 0.025$  的  $t$  值 (2.2282.764) 和  $df = 10$ ,  $\alpha = 0.01$  的  $t$  值

(2.764)之间,  $p$  值就在 0.010 和 0.025 之间, 因此, 拒绝  $H_0$ 。

检验斜率  $\beta_1$  的  $t$  统计量在例 11.16 的输出结果中显示为 2.67, 这等于(近似)相关系数的  $t$  统计量 2.68。

相关系数检验提供了统计上显著和统计上重要之间差别的简洁的说明。假设某个心理学家设计了生产线上工人技术的一个测试, 并检测了 40,000 名工人的大样本。如果测试分数和实际的生产能力间的样本相关系数为 0.02, 那么

$$t = \frac{0.02 \sqrt{39,998}}{\sqrt{1 - (0.02)^2}} = 4.0$$

我们可以在任何合理的水平  $\alpha$  下拒绝零假设, 所以二者相关是“统计上显著的”。但是, 这个测试得分只解释了生产能力的平方误差的  $(0.02)^2 = 0.0004$ , 所以几乎不值得作为预测变量。记住, 统计检验中拒绝零假设, 其结论是如果零假设是真的这个样本似乎可能就不出现。检验本身并不能表明结果的实际显著性。明显地, 对于规模为 40,000 的样本, 仅仅是侥幸的抽取, 一个很小的样本相关系数如 0.02 也不可能发生的。在这个例子中测试分数与生产能力分数间没有实际意义的关系。

### 练习

**11.59** 下面重新列出了练习 11.19 的输出结果。由  $R\text{-square}(r_{yx}^2)$  值计算相关系数  $r_{yx}$ , 这个相关系数的符号应该是正还是负?

. regress Branches Business

| Source   | SS         | df        | MS          | Number of obs = 12 |                      |           |
|----------|------------|-----------|-------------|--------------------|----------------------|-----------|
| Model    | 53.7996874 | 1         | 53.7996874  | F(1,10)            | =                    | 172.60    |
| Residual | 3.11697922 | 10        | 0.311697922 | Prob > F           | =                    | 0.0000    |
| Total    | 56.9166667 | 11        | 5.17424242  | R-square           | =                    | 0.9452    |
|          |            |           |             | Adj R-square       | =                    | 0.9398    |
|          |            |           |             | Root MSE           | =                    | 0.5583    |
| Branches | Coef.      | Std. Err. | t           | P> t               | [95% Conf. Interval] |           |
| Business | 0.0111049  | 0.0008453 | 13.138      | 0.000              | 0.0092216            | 0.0129883 |
| _cons    | 1.766846   | 0.3211751 | 5.501       | 0.000              | 1.051223             | 2.482469  |

### 11.60

a. 关于练习 11.59 的数据, 检验  $x$  和  $y$  之间没有真正的相关关系的假设, 利用单侧  $H_a$  和  $\alpha = 0.01$ 。

b. 比较这个检验的结果和输出结果中斜率的  $t$  检验的结果。

11.61 参照练习 11.39 的计算机输出结果(重新列出如下)。

#### Simple Regression Analysis

Linear model: TotalCost = 99.777 + 5.19179 \* Runsize

#### Table of Estimates

|       | Estimate | Standard<br>Error | t<br>Value | P<br>Value |
|-------|----------|-------------------|------------|------------|
| Slope | 5.19179  | 0.0586455         | 88.53      | 0.0000     |

R-squared = 99.64 %

Correlation coeff. = 0.998

Standard error of estimation = 12.2065

#### Analysis of Variance

| Source        | Sum of<br>Squares | D.F. | Mean Square    | F-Ratio | P<br>Value |
|---------------|-------------------|------|----------------|---------|------------|
| Model         | 1.16775e + 006    | 1    | 1.16775e + 006 | 7837.26 | 0.0000     |
| Error         | 4171.98           | 28   | 148.999        |         |            |
| Total (corr.) | 1.17192e + 006    | 29   |                |         |            |

a. 求出  $r_{yx}^2$ , 这个非常大的值是如何在输出结果的 Sum of Squares 中反映出的?

b. 估计的  $\beta_1$  是正的, 样本相关系数的符号必然是什么?

c. 假如练习 11.39 中研究限制在 RUNSIZE 的值小于 1.8, 你预计  $r_{yx}$  的值将变大还是变小?

11.62 (商业) 假设在 10 个试验城市进行了某个新产品的广告宣传活。广告强度  $x$  是以每晚电视的黄金时间该广告的播放次数计, 它在各个城市是不同的, 广告宣传后调查获得产品认知比例  $y$ :

|       |      |      |      |      |      |      |      |      |      |      |
|-------|------|------|------|------|------|------|------|------|------|------|
| $x$ : | 4.0  | 4.5  | 5.0  | 5.5  | 6.0  | 6.5  | 7.0  | 7.5  | 8.0  | 8.5  |
| $y$ : | 10.1 | 10.3 | 10.4 | 21.7 | 36.7 | 51.5 | 67.0 | 68.5 | 68.2 | 69.3 |

MTB > Correlation 'Intensty' 'Aware'.

Correlation of Intensty and Aware = 0.956

a. 解释相关系数  $r_{yx}$ 。

b. 绘制散点图, 你认为二者关系是线性的吗? 这个关系大体上是增加的吗?

**11.63 (教育)** 某商学院对最近毕业的 MBA 的一项调查得到了工作第一年年薪和以前工作的年数的数据。以下是利用 Systat 软件得到的结果:

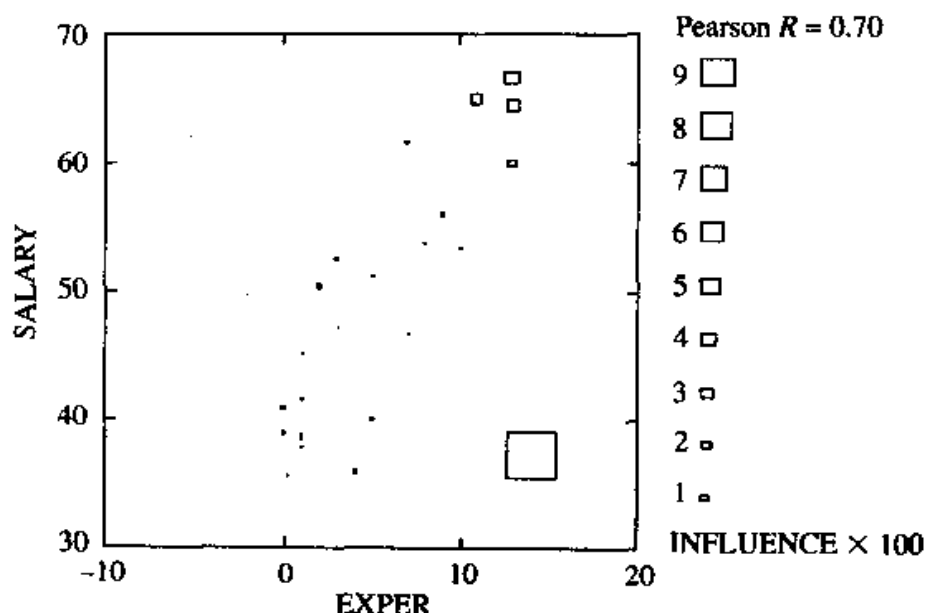
| CASE | EXPER  | SALARY | CASE | EXPER  | SALARY |
|------|--------|--------|------|--------|--------|
| 1    | 8.000  | 53.900 | 27   | 7.000  | 51.700 |
| 2    | 5.000  | 52.500 | 28   | 9.000  | 56.200 |
| 3    | 5.000  | 49.000 | 29   | 6.000  | 48.900 |
| 4    | 11.000 | 65.100 | 30   | 6.000  | 51.900 |
| 5    | 4.000  | 51.600 | 31   | 4.000  | 36.100 |
| 6    | 3.000  | 52.700 | 32   | 6.000  | 53.500 |
| 7    | 3.000  | 44.500 | 33   | 5.000  | 50.400 |
| 8    | 3.000  | 40.100 | 34   | 1.000  | 38.700 |
| 9    | 0.000  | 41.100 | 35   | 13.000 | 60.100 |
| 10   | 13.000 | 66.900 | 36   | 1.000  | 38.900 |
| 11   | 14.000 | 37.900 | 37   | 6.000  | 48.400 |
| 12   | 10.000 | 53.500 | 38   | 2.000  | 50.600 |
| 13   | 2.000  | 38.300 | 39   | 4.000  | 41.800 |
| 14   | 2.000  | 37.200 | 40   | 1.000  | 44.400 |
| 15   | 5.000  | 51.300 | 41   | 5.000  | 46.600 |
| 16   | 13.000 | 64.700 | 42   | 1.000  | 43.900 |
| 17   | 1.000  | 45.300 | 43   | 4.000  | 45.000 |
| 18   | 5.000  | 47.000 | 44   | 1.000  | 37.900 |
| 19   | 1.000  | 43.800 | 45   | 2.000  | 44.600 |
| 20   | 5.000  | 47.400 | 46   | 7.000  | 46.900 |
| 21   | 5.000  | 40.200 | 47   | 5.000  | 47.600 |
| 22   | 7.000  | 52.800 | 48   | 1.000  | 43.200 |
| 23   | 4.000  | 40.700 | 49   | 1.000  | 41.600 |
| 24   | 3.000  | 47.300 | 50   | 0.000  | 39.200 |
| 25   | 3.000  | 43.700 | 51   | 1.000  | 41.700 |
| 26   | 7.000  | 61.800 |      |        |        |

a. 浏览这些数据, 你能感觉有关系吗? 特别地, 是否显示出工作年数越少起薪越少?

b. 你是否注意到某些数据似乎不是这种模式?

**11.64** 利用 Systat 的“影响点图(influence plot)”绘制了练习 11.63 中数据

的点图。这是一个散点图,其中每个点都确定了去掉此点相关系数改变的多少。越大的点表示去掉后相关系数改变的越多。下面是这个散点图,图中显示出逐渐增大的模式吗?有没有明显不属于这种模式的点?



**11.65** 从练习 11.64 中 Systat 计算出一个回归方程,其中年薪作为因变量。部分输出结果显示如下:

| ANALYSIS OF VARIANCE |                |    |             |         |       |
|----------------------|----------------|----|-------------|---------|-------|
| SOURCE               | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P     |
| REGRESSION           | 1395.959       | 1  | 1395.959    | 47.838  | 0.000 |
| RESIDUAL             | 1429.868       | 49 | 29.181      |         |       |

- 写出预测方程,解释系数,在问题的背景下常数项(截距)的意义是什么?
- 确定残差标准差,这个数意味着什么?
- 这个明显的关系在统计上可检查出来(显著)吗?
- 年薪的变差有多少可以由工作年数的变差解释?

**11.66** 练习 11.63 中第 11 人供职于家族企业,低薪水但有企业的股票回报。从数据中去掉这个点(在影响点图中是高影响点)再用 Systat 重新分析,部分输出结果如下:

| DEP VAR: SALARY N: 50 MULTIPLE R: 0.842 SQUARED MULTIPLE R: 0.709   |             |           |          |        |           |
|---|-------------|-----------|----------|--------|-----------|
| ADJUSTED SQUARED MULTIPLE R: .703 STANDARD ERROR OF ESTIMATE: 4.071 |             |           |          |        |           |
| VARIABLE  | COEFFICIENT | STD ERROR | STD COEF | T      | P(2 TAIL) |
| CONSTANT  | 39.188      | 0.971     | 0.000    | 40.353 | 0.000     |
| EXPER   | 1.863       | 0.172     | 0.842    | 10.812 | 0.000     |



- a. 去掉散点图中这个高影响点后斜率应该增大还是减小? 是这样的吗?
- b. 去掉这个点后残差标准差应该朝哪个方向改变(增大还是减小)? 是这样的吗? 这个改变有多大?
- c. 去掉这个点后相关系数应该怎样变化? 这个改变有多大?

## 11.8 小 结

本章介绍了回归分析,主要是利用一个自变量预测一个因变量的简单回归分析。最基本的问题涉及变量之间关系的状态(线性型或曲线型),围绕预测值的变异性的变化,在预测范围内变异性是否是同样的,自变量预测因变量有多大效果,以及在多大程度上容许抽样误差。以下是本章的重要概念:

1. 应该绘制出数据的散点图。确定变量之间的关系是近似线性还是明显曲线时,如同 LOWESS 或样条曲线的平滑法是有用的。曲线型关系常常可以通过对自变量、或因变量、或二者同时作变换转换为近似线性的。
2. 估计线性回归模型的系数是用最小二乘法,即最小化残差(实际值减去预测值)平方和。由于涉及到平方误差,因此这种方法对离群值是敏感的。
3.  $x$  方向的极端观测值在拟合直线时具有高杠杆作用。如果一个高杠杆点远离直线,则它还具有高影响,也就是说去掉这个观测值会实质性改变拟合线。如果一个高影响点与其他点不是来自同一总体,就应该剔除。如果必须保留这个点,那么应该使用其他方法而不用最小二乘方法去估计模型参数。
4. 围绕回归线的变异性是通过残差的标准差来度量的。这个标准差可以用经验法则解释。标准差有时会随预测值的增加而增加,这时应该尝试对因变量做变换。
5. 回归线的斜率(和截距)的假设检验和置信区间都是基于  $t$  分布。如果变量间没有关系,则斜率为 0。如果  $x$  变量的变化范围较宽,则估计的回归线更准确。
6. 利用拟合的直线可以在新的  $x$  值处进行预测,仍然用  $t$  分布来进行。如果新的  $x$  值远离以前的  $x$  值时,预测就可能不准确。
7. 决定系数,即相关系数平方,是度量变量间关系的强弱的一个标准方法。这个度量会因为关系的非线性或人为限制  $x$  变量的变化范围有所减少。

对管理人员来说,统计最重要的应用之一是预测。一个经理可能需要在已知某项合同工作的规模的条件下预测其成本;也可能需要在已知国民生产总值目前的增长率的情况下预测某种特定商品的销售情况;还可能需要在已知劳动力规模的情况下预测生产的零部件的数量。预测的统计方法中运用最广泛的就是回归

分析。

在回归分析方法中,相关变量的过去数据被用来开发和评估预测方程。这个方程中被预测的变量称为因变量,而预测所用的变量称为自变量。本章所讨论的回归方法只涉及一个自变量。第十二章我们将把这些方法扩展为多元回归,多元回归是考虑多个自变量的情形。

以下是一个回归分析研究中所能够完成的一系列工作:

1. 利用数据求得回归方程。
2. 利用数据估计这个方程的变异性或不确定性。
3. 利用数据确定远离预测值的异常点,这些点可能代表异常的问题或机会。
4. 由于数据只是一个样本,因此能够对回归参数的(总体)真值进行推断。
5. 利用预测方程可以得出因变量未来值的一个合理的范围。
6. 利用数据估计因变量和自变量间的相关系数,这个数度量了二变量间关系的强弱程度。

重要公式

1. 斜率和截距的最小二乘估计

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \text{ 和 } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

其中

$$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) \text{ 和 } S_{xx} = \sum_i (x_i - \bar{x})^2$$

2.  $\sigma_e^2$  的估计

$$s_e^2 = \frac{\sum_i (y_i - \hat{y}_i)^2}{n - 2} = \frac{\text{SS(残差)}}{n - 2}$$

3.  $\beta_1$  的统计检验

$$H_0: \beta_1 = 0 \text{ (双侧)}$$

$$\text{T.S.: } t = \frac{\hat{\beta}_1}{s_e \sqrt{1/S_{xx}}}$$

4.  $\beta_1$  的置信区间

$$\hat{\beta}_1 \pm t_{\alpha/2} s_e \sqrt{\frac{1}{S_{xx}}}$$

5.  $H_0: \beta_1 = 0$  (双侧) 的  $F$  检验

$$\text{T.S.: } F = \frac{\text{MS(回归)}}{\text{MS(残差)}}$$

6.  $E(y_{n+1})$  的置信区间

$$\hat{y}_{n+1} \pm t_{\alpha/2} s_e \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}}}$$

7.  $y_{n+1}$  的预测区间

$$\hat{y}_{n+1} \pm t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}}}$$

8. 线性回归拟合不足的检验

$$T.S.: F = \frac{MS_{\text{Lack}}}{MS_{\text{exp}}}$$

其中

$$MS_{\text{exp}} = \frac{SSP_{\text{exp}}}{\sum (n_i - 1)} = \frac{\sum_{ij} (y_{ij} - \bar{y}_i)^2}{\sum_i (n_i - 1)}$$

$$MS_{\text{Lack}} = \frac{SS(\text{残差}) - SSP_{\text{exp}}}{n - 2 - \sum (n_i - 1)}$$

9. 基于单个  $y$  值的  $x$  的预测限

$$\hat{x} = \frac{y - \hat{\beta}_0}{\hat{\beta}_1}$$

$$\hat{x}_U = \bar{x} + \frac{1}{1 - c^2} [(\hat{x} - \bar{x}) + d]$$

$$\hat{x}_L = \bar{x} + \frac{1}{1 - c^2} [(\hat{x} - \bar{x}) - d]$$

其中

$$c^2 = \frac{t_{\alpha/2}^2 s_e^2}{\hat{\beta}_1^2 S_{xx}}$$

$$d = \frac{t_{\alpha/2} s_e}{\hat{\beta}_1} \sqrt{\frac{n+1}{n} (1 - c^2) + \frac{(\hat{x} - \bar{x})^2}{S_{xx}}}$$

10. 基于  $m$  个  $y$  值的  $x$  的预测区间

$$\hat{x}_U = \bar{x} + \frac{1}{1 - c^2} [(\hat{x} - \bar{x}) + g]$$

$$\hat{x}_L = \bar{x} + \frac{1}{1 - c^2} [(\hat{x} - \bar{x}) - g]$$

其中

$$\hat{x} = \frac{P\bar{y}_m - \hat{\beta}_0}{\hat{\beta}_1}$$

$$g = \frac{t_{\alpha/2}}{\hat{\beta}_1} \sqrt{(s_y^2 P^2 + \frac{s_e^2}{n})(1 - c^2) + \frac{(\hat{x} - \bar{x})^2 s_e^2}{S_{xx}}}$$

11. 相关系数

$$r_{yx} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \hat{\beta}_1 \sqrt{\frac{S_{xx}}{S_{yy}}}$$

12. 决定系数

$$r_{yx}^2 = \frac{SS(\text{总和}) - SS(\text{残差})}{SS(\text{总和})}$$

13.  $\rho_{yx}$  的统计检验

$H_0: \rho_{yx} = 0$  (双侧)

$$\text{T. S.: } t = r_{yx} \frac{\sqrt{n-2}}{\sqrt{1-r_{yx}^2}}$$

补充练习

11.67 考虑以下数据:

|       |    |    |    |    |    |    |    |
|-------|----|----|----|----|----|----|----|
| $x$ : | 10 | 12 | 14 | 15 | 18 | 19 | 20 |
| $y$ : | 25 | 30 | 36 | 37 | 42 | 50 | 55 |

a. 绘制散点图。

b. 利用数据, 求出模型  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  的最小二乘估计。

c. 预测当  $x = 21$  时的  $y$ 。

11.68 参照练习 11.67。

a. 计算残差标准差  $s_e$ 。

b. 计算这些数据的残差, 大多数都在 0 的  $\pm 2s_e$  内吗?

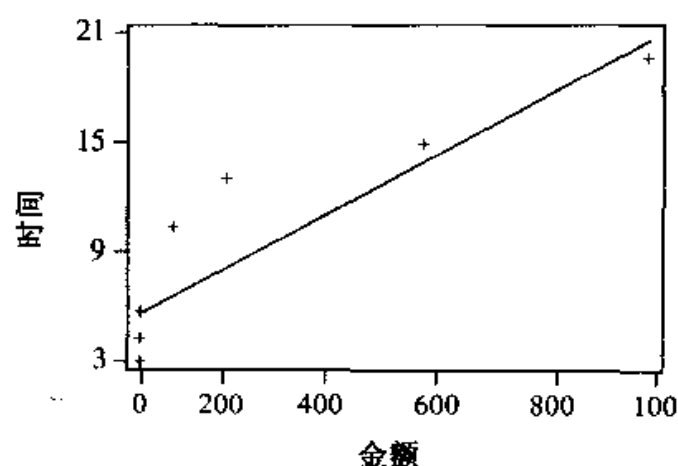
11.69 (政府) 某负责政府研究课题资助的机构是在许多私营公司的仔细检查之下工作的。一家公司研究了项目合同的金额 ( $\times 10,000$  美元) 与项目计划的提交到批准的时间的关系:

|                             |   |   |    |    |     |     |      |
|-----------------------------|---|---|----|----|-----|-----|------|
| 需要的时间 $y$ (月):              | 3 | 4 | 6  | 8  | 11  | 14  | 20   |
| 合同金额 ( $\times 10,000$ 美元): | 1 | 5 | 10 | 50 | 100 | 500 | 1000 |

$y$  关于  $x$  的散点图和 Stata 输出结果如下:

. regress Length Size

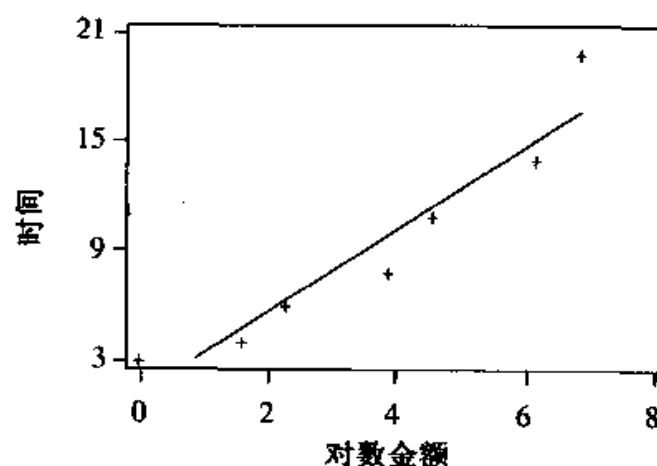
| Source   | SS         | df        | MS         | Number of obs = 7 |                      |           |
|----------|------------|-----------|------------|-------------------|----------------------|-----------|
| Model    | 191.389193 | 1         | 191.389193 | F(1,5)            | =                    | 33.78     |
| Residual | 28.3250928 | 5         | 5.66501856 | Prob > F          | =                    | 0.0021    |
| Total    | 219.714286 | 6         | 36.6190476 | R-square          | =                    | 0.8711    |
|          |            |           |            | Adj R-square      | =                    | 0.8453    |
|          |            |           |            | Root MSE          | =                    | 2.3801    |
| Length   | Coef.      | Std. Err. | t          | P> t              | [95% Conf. Interval] |           |
| Size     | 0.0148652  | 0.0025575 | 5.812      | 0.002             | 0.008291             | 0.0214394 |
| _cons    | 5.890659   | 1.086177  | 5.423      | 0.003             | 3.098553             | 8.682765  |



a. 最小二乘线是什么?

b. 进行零假设为  $H_0: \beta_1 \leq 0$  的检验, 确定此检验的  $p$  值。

11.70 参照练习 11.67 的数据。y 关于 x 的自然对数的散点图和 Stata 输出结果如下:



. regress Length lnSize

| Source   | SS         | df | MS         | Number of obs = 7 |   |        |
|----------|------------|----|------------|-------------------|---|--------|
| Model    | 199.443893 | 1  | 199.443893 | F(1, 5)           | = | 49.20  |
| Residual | 20.2703932 | 5  | 4.05407863 | Prob > F          | = | 0.0009 |
| Total    | 219.714286 | 6  | 36.6190476 | R-square          | = | 0.9077 |
|          |            |    |            | Adj R-square      | = | 0.8893 |
|          |            |    |            | Root MSE          | = | 2.0135 |

| Length | Coef.    | Std. Err. | t     | P> t  | [95 % Conf. Interval] |          |
|--------|----------|-----------|-------|-------|-----------------------|----------|
| lnSize | 2.307015 | 0.3289169 | 7.014 | 0.000 | 1.461508              | 3.152523 |
| _cons  | 1.007445 | 1.421494  | 0.709 | 0.510 | -2.646622             | 4.661511 |

a.  $\log x$  为自变量的回归线是什么?

b. 进行零假设为  $H_0: \beta_1 \leq 0$  的检验, 确定单侧对立假设  $H_a: \beta_1 > 0$  的显著性水平。

11.71 利用练习 11.69 和练习 11.70 的结果, 确定哪个回归模型提供了更好的拟合, 并说明理由。

11.72 参照前两个练习的计算机输出结果。

a. 求出线性回归线斜率  $\beta_1$  的 95% 置信区间。

b. 确定对数回归模型斜率  $\beta_1$  的 95% 置信区间。

11.73 利用你所选择的练习 11.70 的数据的模型, 预测批准一个合同金额为 750,000 美元的项目的以月计的时间, 并给出 95% 预测区间的粗略估计。

11.74 (环境) 某航空公司研究一种特定型飞机的燃料消耗, 收集了 100 次航程的数据, 这些数据包括飞行英里数  $x$  (以百英里计) 和实际的燃料消耗  $y$  (以加仑计)。Statistix 的计算结果和散点图如下。

#### UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF GALLONS

##### PREDICTOR

| VARIABLES | COEFFICIENT | STD ERROR | STUDENT'S T | P      |
|-----------|-------------|-----------|-------------|--------|
| CONSTANT  | 140.074     | 44.1293   | 3.17        | 0.0099 |
| MILES     | 0.61896     | 0.04855   | 12.75       | 0.0000 |

|                    |        |                         |         |
|--------------------|--------|-------------------------|---------|
| R-SQUARED          | 0.9420 | RESID. MEAN SQUARE(MSE) | 1182.34 |
| ADJUSTED R-SQUARED | 0.9362 | STANDARD DEVIATION      | 34.3852 |

| SOURCE     | DF | SS        | MS        | F      | P      |
|------------|----|-----------|-----------|--------|--------|
| REGRESSION | 1  | 1.921E+05 | 1.921E+05 | 162.48 | 0.0000 |
| RESIDUAL   | 10 | 11823.4   | 1182.34   |        |        |
| TOTAL      | 11 | 2.039E+05 |           |        |        |

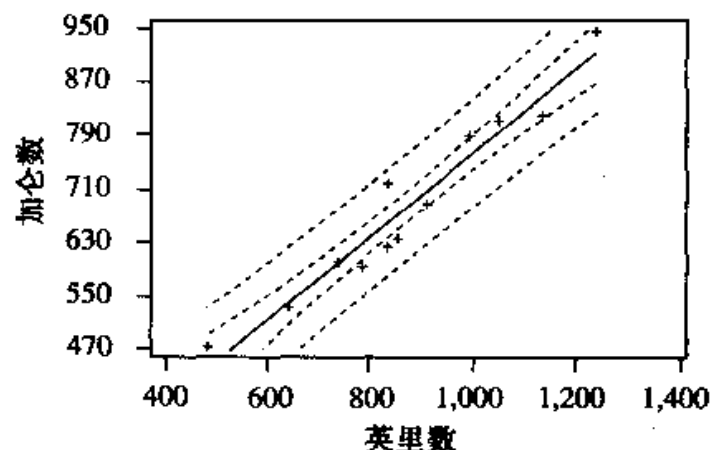
#### PREDICTED/FITTED VALUES OF GALLONS

|                       |        |                    |        |
|-----------------------|--------|--------------------|--------|
| LOWER PREDICTED BOUND | 678.33 | LOWER FITTED BOUND | 733.68 |
| PREDICTED VALUE       | 759.03 | FITTED VALUE       | 759.03 |
| UPPER PREDICTED BOUND | 839.73 | UPPER FITTED BOUND | 784.38 |
| SE (PREDICTED VALUE)  | 36.218 | SE (FITTED VALUE)  | 11.377 |

|                        |        |
|------------------------|--------|
| UNUSUALNESS (LEVERAGE) | 0.1095 |
|------------------------|--------|

PERCENT COVERAGE 95.0  
CORRESPONDING T 2.23

PREDICTOR VALUES: MILES = 1000.0



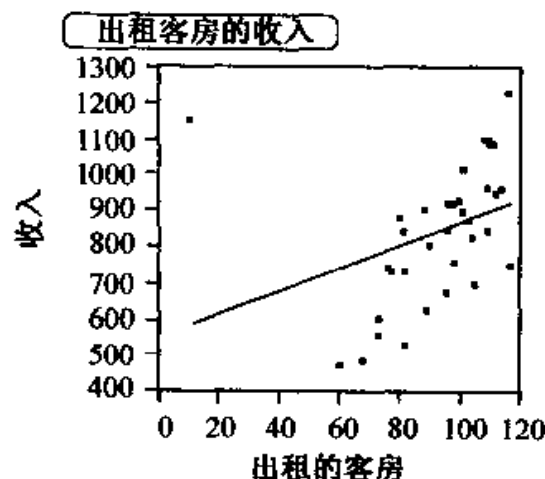
- 求出回归方程。
- 样本相关系数和决定系数是多少？并解释这两个数。
- 关于检验  $H_0: \beta_1 \leq 0$ , 有什么看法？

11.75 参照练习 11.74 的数据和输出结果。

- 预测所有 1,000 英里航程的平均燃料消耗, 并给出 95% 的置信区间。
- 预测某次 1,000 英里航程的燃料消耗, 628 加仑的燃料消耗是特别少吗？

11.76 在练习 11.74 的情况下,  $\beta_1$  的解释是什么?  $\beta_0$  有解释吗?

11.77 (商务) 某个大的郊区汽车旅馆的收入来源于客房的出租和餐厅与游戏厅的销售。客房的占有和餐厅/游戏厅的销售之间似乎应该存在一定的联系, 但旅馆的经理对这种关系的密切程度没有认识。于是收集了 36 个非假期的工作日 (星期一到星期四的晚上) 的出租客房的数量和餐厅/游戏厅的销售额。这些数据的散点图和回归分析结果如下所示。



## Linear Fit

## Summary of Fit

|                            |          |
|----------------------------|----------|
| RSquare                    | 0.118716 |
| RSquare Adj                | 0.092796 |
| Root Mean Square Error     | 182.253  |
| Mean of Response           | 854.1514 |
| Observations (or Sum Wgts) | 36       |

## Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Ratio |
|---------|----|----------------|-------------|---------|
| Model   | 1  | 152132.2       | 152132      | 4.5801  |
| Error   | 34 | 1129349.3      | 33216       | Prob>F  |
| C Total | 35 | 1281481.6      |             | 0.0396  |

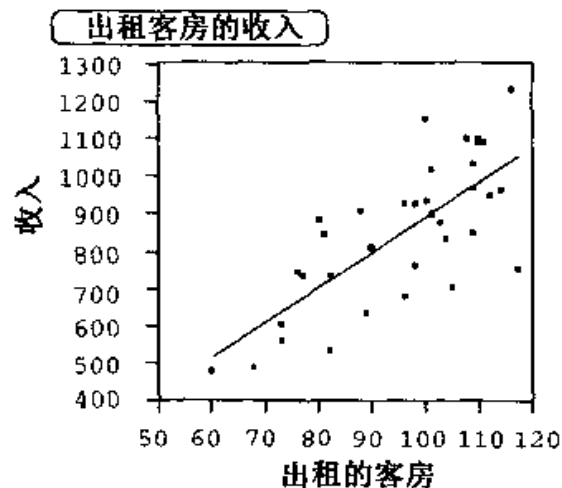
## Parameter Estimates

| Term           | Estimate  | Std Error | t Ratio | Prob>[t] |
|----------------|-----------|-----------|---------|----------|
| Intercept      | 557.72428 | 141.8019  | 3.93    | 0.0004   |
| Rooms occupied | 3.1760047 | 1.484039  | 2.14    | 0.0396   |

a. 按照输出结果, 客房的出租和收入间存在统计上显著的关系吗?

b. 如果剔除位于左上方的点, 斜率会增大还是减少? 你认为有实质性改变吗?

11.78 旅馆的数据有一个点出现输入错误, 客房出租数被记录为 10 而不是 100。改正这个错误后, 得到的输出结果如下所示。



## Linear Fit

## Summary of Fit

|                            |          |
|----------------------------|----------|
| RSquare                    | 0.552922 |
| RSquare Adj                | 0.539773 |
| Root Mean Square Error     | 129.81   |
| Mean of Response           | 854.1514 |
| Observations (or Sum Wgts) | 36       |



| Analysis of Variance |            |           |         |         |
|----------------------|------------|-----------|---------|---------|
| Parameter Estimates  |            |           |         |         |
| Term                 | Estimate   | Std Error | t Ratio | Prob> t |
| Intercept            | - 50.18525 | 141.1283  | - 0.36  | 0.7243  |
| Rooms occupied       | 9.4365563  | 1.455236  | 6.48    | 0.0000  |

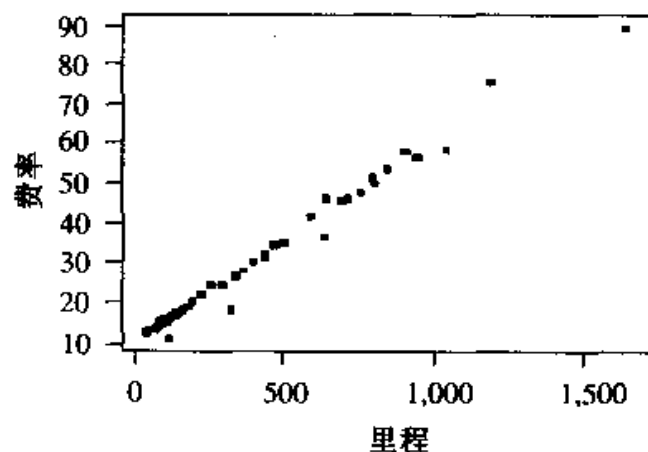
- 改正错误后斜率会有什么变化?
- 截距怎么变化?
- 离群值使得标准残差增大还是减小?
- 离群值使得  $r^2$  的值增大还是减小?

**11.79 (工程)** 某杂货店食品制造公司的管理科学部门正在为麦片产品的生产和配送开发一个线性规划模型。这个模型需要大量的起点和终点的运输成本。每个可能的组合都做详细的价格分析是不切实际的,因此选择了 50 条线路的样本。对于每条线路,找出英里数和运输费率(单位是元/100 磅)。进行回归分析后,得出的散点图和 Excel 输出结果如下:

|    | A                     | B      | C        | D       | E       | F              | G |
|----|-----------------------|--------|----------|---------|---------|----------------|---|
| 1  | SUMMARY OUTPUT        |        |          |         |         |                |   |
| 2  |                       |        |          |         |         |                |   |
| 3  | Regression Statistics |        |          |         |         |                |   |
| 4  | Multiple R            | 0.9929 |          |         |         |                |   |
| 5  | R Square              | 0.9859 |          |         |         |                |   |
| 6  | Adjusted R Square     | 0.9856 |          |         |         |                |   |
| 7  | Standard Error        | 2.2021 |          |         |         |                |   |
| 8  | Observations          | 48     |          |         |         |                |   |
| 9  |                       |        |          |         |         |                |   |
| 10 |                       |        |          |         |         |                |   |
| 11 | ANOVA                 |        |          |         |         |                |   |
| 12 |                       | df     | SS       | MS      | F       | Significance F |   |
| 13 | Regression            | 1      | 15558.63 | 15558.6 | 3208.47 | 0.00           |   |
| 14 | Residual              | 46     | 223.06   | 4.85    |         |                |   |

续表

|    | A         | B            | C              | D       | E       | F          | G          |
|----|-----------|--------------|----------------|---------|---------|------------|------------|
| 15 | Total     | 47           | 15781.7        |         |         |            |            |
| 16 |           |              |                |         |         |            |            |
| 17 |           |              |                |         |         |            |            |
| 18 |           | Coefficients | Standard Error | t Stat  | P-value | Lower 95 % | Upper 95 % |
| 19 | Intercept | 9.7709       | 0.4740         | 20.6122 | 0.0000  | 8.8167     | 10.7251    |
| 20 | Mileage   | 0.0501       | 0.0009         | 56.6434 | 0.0000  | 0.0483     | 0.0519     |



数据如下

|    |      |      |      |      |      |      |      |      |      |       |       |       |
|----|------|------|------|------|------|------|------|------|------|-------|-------|-------|
| 里程 | 50   | 60   | 80   | 80   | 90   | 90   | 100  | 100  | 100  | 110   | 110   | 110   |
| 费率 | 12.7 | 13.0 | 13.7 | 14.1 | 14.6 | 14.1 | 15.6 | 14.9 | 14.5 | 15.3  | 15.5  | 15.9  |
| 里程 | 120  | 120  | 120  | 120  | 130  | 130  | 140  | 150  | 170  | 190   | 200   | 230   |
| 费率 | 16.4 | 11.1 | 16.0 | 15.8 | 16.0 | 16.7 | 17.2 | 17.5 | 18.6 | 19.3  | 20.4  | 21.8  |
| 里程 | 260  | 300  | 330  | 340  | 370  | 400  | 440  | 440  | 480  | 510   | 540   | 600   |
| 费率 | 24.7 | 24.7 | 18.0 | 27.1 | 28.2 | 30.6 | 31.8 | 32.4 | 34.5 | 35.0  | 36.3  | 41.4  |
| 里程 | 650  | 700  | 720  | 760  | 800  | 810  | 850  | 920  | 960  | 1,050 | 1,200 | 1,650 |
| 费率 | 46.4 | 45.8 | 46.6 | 48.0 | 51.7 | 50.2 | 53.6 | 57.9 | 56.1 | 58.7  | 75.8  | 89.0  |

a. 写出回归方程和残差标准差。

b. 计算斜率真值的 95% 置信区间。

**11.80** 练习 11.79 的散点图中, 你认为这些数据有问题吗?**11.81** 关于练习 11.79, 预测一条 340 英里长的线路的运输费率, 并求出 95% 预测区间。这个练习中的外推问题严重吗?

**11.82 (社会)** 郊区城镇常常将大部分市政预算花在公众安全(警察、消防和急救)服务上。某个纳税人团体认为由于财政基数小,那些非常小的城镇花费了大量的人均预算。这个团体收集了某个大都市的 29 个郊区城镇公众安全的每笔资金支出的数据以及每个城镇的人口总数。数据经过 Minitab 软件分析,因变量为“支出(expendit)”,自变量为“城镇人口数(townpopn)”的回归模型产生下面的结果:

```
MTB>regress 'expendit' 1 'townpopn'
```

The regression equation is

expendit = 119 + 0.000532 townpopn

| Predictor | Coef      | Stdev     | t-ratio | p     |
|-----------|-----------|-----------|---------|-------|
| Constant  | 118.96    | 23.26     | 5.11    | 0.000 |
| townpopn  | 0.0005324 | 0.0006181 | 0.86    | 0.397 |

s = 43.31      R-sq = 2.7 %      R-sq(adj) = 0.0 %

Analysis of Variance

| SOURCE     | DF | SS    | MS   | F    | p     |
|------------|----|-------|------|------|-------|
| Regression | 1  | 1392  | 1392 | 0.74 | 0.397 |
| Error      | 27 | 50651 | 1876 |      |       |
| Total      | 28 | 52043 |      |      |       |

Unusual Observations

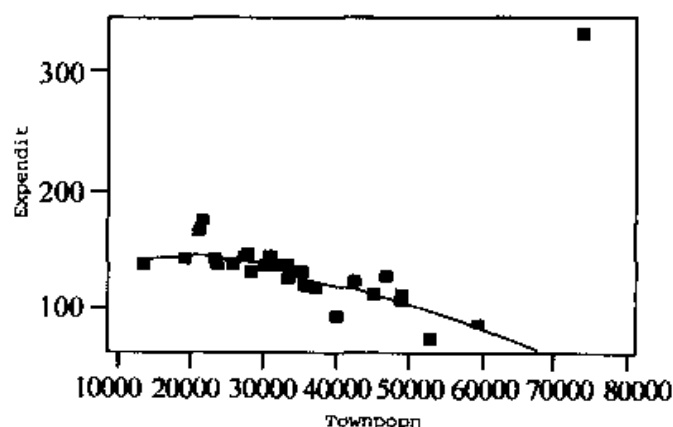
| Obs. | townpopn | expendit | Fit    | Stdev. Fit | Residual | St. Resid |
|------|----------|----------|--------|------------|----------|-----------|
| 8    | 74151    | 334.00   | 158.43 | 25.32      | 175.57   | 5.00RX    |

R denotes an obs. with a large st. resid.

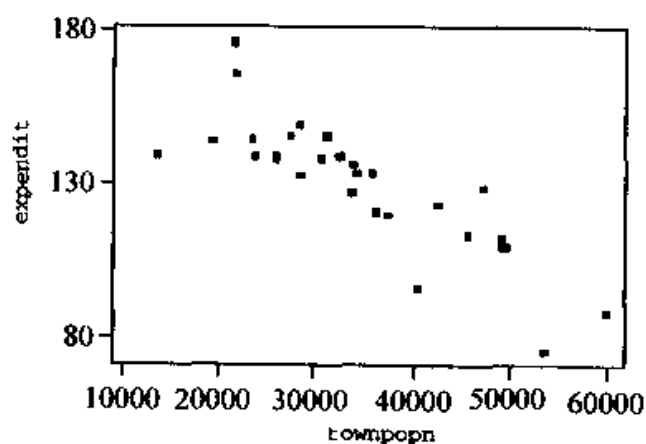
X denotes an obs. whose X value gives it large influence.

- 如果纳税人团体的想法是正确的,则回归模型的斜率的符号应该是什么?
- 输出结果中的斜率证实了这个团体的评论吗?

**11.83** 下面给出了练习 11.82 中的数据由 Minitab 绘制的散点图和 LOWESS 平滑曲线,这个散点图是否表明回归线是不合适的?为什么?



**11.84** 练习 11.82 的数据集中某个城镇有一个大的地区性购物中心。这个城镇的公众安全支出的很大部分是与这个购物中心相关的, 将这个城镇的数据从数据集中剔除, 再对剩余的数据用 Minitab 进行分析。数据的散点图如下所示。



- 解释为什么从数据集中剔除这个点后回归线会有如此实质性的变化?
- 重新分析的回归线是否证实练习 11.82 中纳税人团体的评论?

**11.85** 下面给出了练习 11.82 的数据剔除一个离群值后的回归分析结果, 所得的斜率与前面的相比有什么变化?

```
MTB>regress 'expendit' 1 'townpopn'
```

The regression equation is

$\text{expendit} = 184 - 0.00158 \text{ townpopn}$

| Predictor | Coef       | Stdev     | t-ratio | p     |
|-----------|------------|-----------|---------|-------|
| Constant  | 184.240    | 7.481     | 24.63   | 0.000 |
| townpopn  | -0.0015766 | 0.0002099 | -7.51   | 0.000 |

$s = 12.14$        $R\text{-sq} = 68.5\%$        $R\text{-sq}(\text{adj}) = 67.2\%$

#### Analysis of variance

| SOURCE     | DF | SS      | MS     | F     | p     |
|------------|----|---------|--------|-------|-------|
| Regression | 1  | 8322.7  | 8322.7 | 56.43 | 0.000 |
| Error      | 26 | 3834.5  | 147.5  |       |       |
| Total      | 27 | 12157.2 |        |       |       |

#### Unusual Observations

| Obs. | townpopn | expendit | Fit    | Stdev. Fit | Residual | St. Resid |
|------|----------|----------|--------|------------|----------|-----------|
| 5    | 40307    | 96.00    | 120.69 | 2.66       | -24.69   | -2.08R    |
| 6    | 13457    | 139.00   | 163.02 | 4.87       | -24.02   | -2.16R    |
| 13   | 59779    | 89.00    | 89.99  | 5.89       | -0.99    | -0.09 X   |
| 22   | 21701    | 176.00   | 150.03 | 3.44       | 25.97    | 2.23R     |
| 27   | 53322    | 76.00    | 100.17 | 4.67       | -24.17   | -2.16R    |

R denotes an obs. with a large st. resid.

X denotes an obs. whose X value gives it large influence.

**11.86 (生物)** 在研究对高血压治疗有效的药物时, 研究人员安排了三组试验用老鼠, 每组六只。第一组的老鼠注射了  $0.1 \text{ mg/kg}$  的某种试验药物; 而第二和第三组的分别注射了  $0.2$  和  $0.4 \text{ mg/kg}$ 。研究人员想了解注射药物后两小时后的血压比注射前的血压下降了多少。具体的数据如下所示:

|        | 剂量, $x$             | 血压下降(mm Hg), $y$ |    |    |    |    |    |
|--------|---------------------|------------------|----|----|----|----|----|
| Group1 | $0.1 \text{ mg/kg}$ | 10               | 12 | 15 | 16 | 13 | 11 |
| Group2 | $0.2 \text{ mg/kg}$ | 25               | 22 | 26 | 19 | 18 | 24 |
| Group3 | $0.4 \text{ mg/kg}$ | 30               | 32 | 35 | 27 | 26 | 29 |

#### a. 利用统计软件拟合模型

$$y = \beta_0 + \beta_1 \log_{10} x + \epsilon$$

b. 利用残差图检验(a)中模型的拟合程度。

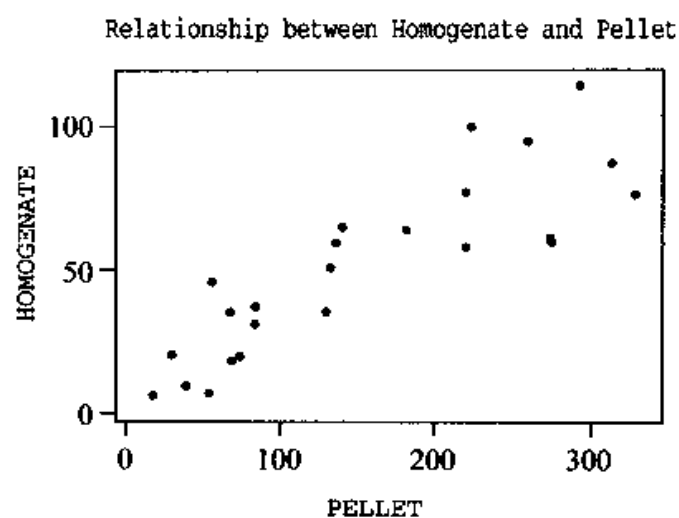
c. 进行  $H_0: \beta_1 \leq 0$ ,  $H_a: \beta_1 > 0$  的统计检验, 求出此检验的  $p$  值。

**11.87 (农业)** 某个实验室进行了一项研究以检验不同水平的氮对莴苣种植产量的影响。利用下列数据拟合线性回归方程, 并检验模型可能的拟合不足。

| 氮的代码 | 产量(每块地中露出茎的重量) |
|------|----------------|
| 1    | 21, 18, 17     |
| 2    | 24, 22, 26     |
| 3    | 34, 29, 32     |

**11.88 (医学)** 研究人员测量了从 24 个接受肠分流术的病人的一部分肠中抽取的蔗糖酶的特殊活性。抽取后, 将它们均匀后再进行酶活动的分析[Carter (1981)]。两种不同的方法用来测量蔗糖酶的活性: 匀浆法(homogenate)和糖丸法(pellet)。用这两种方法测量的 24 个病人的数据如下:

| 由匀浆法及糖丸法测得的蔗糖酶活性 |          |          |
|------------------|----------|----------|
| 病人               | 匀浆法, $y$ | 糖丸法, $x$ |
| 1                | 18.88    | 70.00    |
| 2                | 7.26     | 55.43    |
| 3                | 6.50     | 18.87    |
| 4                | 9.83     | 40.41    |
| 5                | 46.05    | 57.43    |
| 6                | 20.10    | 31.14    |
| 7                | 35.78    | 70.10    |
| 8                | 59.42    | 137.56   |
| 9                | 58.43    | 221.20   |
| 10               | 62.32    | 276.43   |
| 11               | 88.53    | 316.00   |
| 12               | 19.50    | 75.56    |
| 13               | 60.78    | 277.30   |
| 14               | 77.92    | 331.50   |
| 15               | 51.29    | 133.74   |
| 16               | 77.91    | 221.50   |
| 17               | 36.65    | 132.93   |
| 18               | 31.17    | 85.38    |
| 19               | 66.09    | 142.34   |
| 20               | 115.15   | 294.63   |
| 21               | 95.88    | 262.52   |
| 22               | 64.61    | 183.56   |
| 23               | 37.71    | 86.12    |
| 24               | 100.82   | 226.55   |



### Regression Analysis: HOMOGENATE versus PELLET

The regression equation is

$$\text{HOMOGENATE} = 10.3 + 0.267 \text{ PELLET}$$

| Predictor | Coef    | SE Coef | T    | P     |
|-----------|---------|---------|------|-------|
| Constant  | 10.335  | 5.995   | 1.72 | 0.099 |
| PELLET    | 0.26694 | 0.03251 | 8.21 | 0.000 |

$s = 15.62$        $R\text{-sq} = 75.4\%$        $R\text{-sq}(\text{adj}) = 74.3\%$

### Analysis of Variance

| Source         | DF | SS    | MS    | F     | P     |
|----------------|----|-------|-------|-------|-------|
| Regression     | 1  | 16440 | 16440 | 67.41 | 0.000 |
| Residual Error | 22 | 5366  | 244   |       |       |
| Total          | 23 | 21806 |       |       |       |

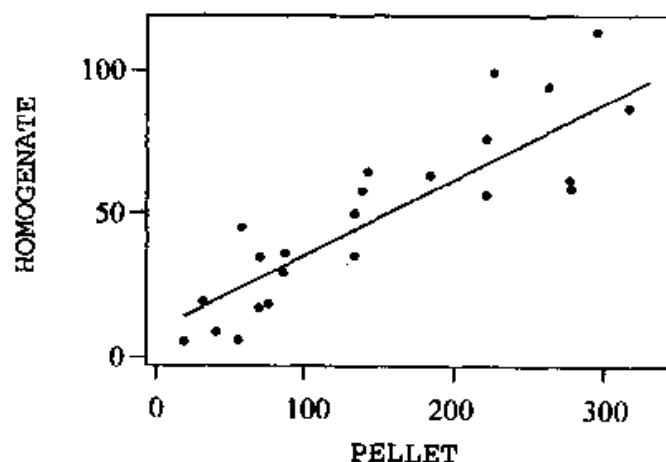
| Obs | PELLET | HOMOGENA | Fit   | SE Fit | Residual | St Resid |
|-----|--------|----------|-------|--------|----------|----------|
| 1   | 70     | 18.88    | 29.02 | 4.24   | -10.14   | -0.67    |
| 2   | 55     | 7.26     | 25.13 | 4.57   | -17.87   | -1.20    |
| 3   | 19     | 6.50     | 15.37 | 5.49   | -8.87    | -0.61    |
| 4   | 40     | 9.83     | 21.12 | 4.93   | -11.29   | -0.76    |
| 5   | 57     | 46.05    | 25.67 | 4.52   | 20.38    | 1.36     |
| 6   | 31     | 20.10    | 18.65 | 5.17   | 1.45     | 0.10     |
| 7   | 70     | 35.78    | 29.05 | 4.24   | 6.73     | 0.45     |
| 8   | 138    | 59.42    | 47.06 | 3.24   | 12.36    | 0.81     |
| 9   | 221    | 58.43    | 69.38 | 3.83   | -10.95   | -0.72    |
| 10  | 276    | 62.32    | 84.13 | 5.04   | -21.81   | -1.48    |

| Obs | PELLET | HOMOGENA | Fit   | SE Fit | Residual | St Resid |
|-----|--------|----------|-------|--------|----------|----------|
| 11  | 316    | 88.53    | 94.69 | 6.10   | -6.16    | -0.43    |
| 12  | 76     | 19.50    | 30.50 | 4.13   | -11.00   | -0.73    |
| 13  | 277    | 60.78    | 84.36 | 5.07   | -23.58   | -1.60    |
| 14  | 332    | 77.92    | 98.83 | 6.53   | -20.91   | -1.47    |
| 15  | 134    | 51.29    | 46.04 | 3.27   | 5.25     | 0.34     |
| 16  | 222    | 77.91    | 69.46 | 3.83   | 8.45     | 0.56     |
| 17  | 133    | 36.65    | 45.82 | 3.28   | -9.17    | -0.60    |
| 18  | 85     | 31.17    | 33.13 | 3.93   | -1.96    | -0.13    |
| 19  | 142    | 66.09    | 48.33 | 3.22   | 17.76    | 1.16     |
| 20  | 295    | 115.15   | 88.98 | 5.52   | 26.17    | 1.79     |
| 21  | 263    | 95.88    | 80.41 | 4.70   | 15.47    | 1.04     |
| 22  | 184    | 64.61    | 59.33 | 3.31   | 5.28     | 0.35     |
| 23  | 86     | 37.71    | 33.32 | 3.92   | 4.39     | 0.29     |
| 24  | 227    | 100.82   | 70.81 | 3.92   | 30.01    | 1.99     |

Regression Line for Homogenate versus Pellet

$$\text{HOMOGENATE} = 10.3348 + 0.266940 \text{ PELLET}$$

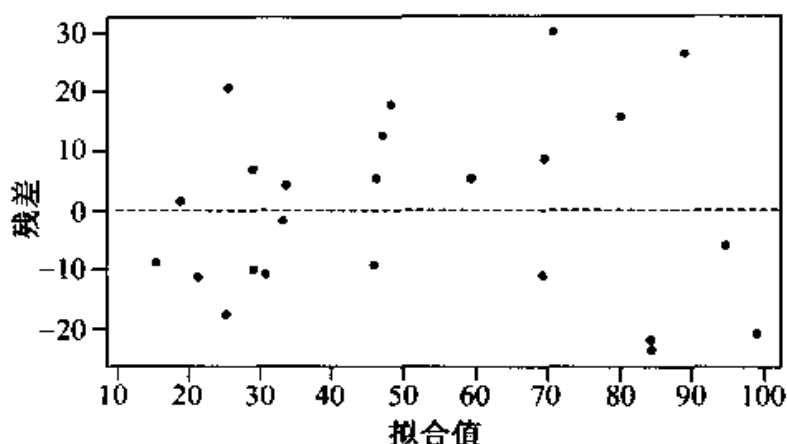
S = 15.6169    R-Sq = 75.4 %    R-Sq(adj) = 74.3 %



- 检查数据的散点图, 线性模型充分描述了这两种方法间的关系吗?
- 检查残差图; 这个散点图是否揭示出一些问题?
- 一般地, 糖丸法比匀浆法更花时间, 同时也提供了蔗糖酶活动的更准确的测量。你如何根据匀浆法的读数预测糖丸法的读数?
- 你怎样求出点估计的置信(预测)区间?

**11.89 (商业)** 某郊区的房产经纪人想仅仅根据房屋的大小预测房价。从所列的多项服务中, 这个房产经纪人获得了房屋的户型(单位为千平方英尺)和价格(单位为千美元)。所有的信息都存贮在 EX1189.DAT 文件, 读者可在出版社网站的数据集中找到, 其中第一列是价格, 第二列是户型。使用统计软件读出这个文件。





- 绘制价格对户型的散点图,这个图显示出一种增加的关系吗?
- 确定数据中明显的离群点,它是高杠杆点吗?
- 求出包括这个离群点在内的数据的回归方程。
- 剔除这个离群点,求出新的回归方程。没有这个离群点,斜率是怎样变化的?为什么?
- 确定包括离群点的模型和不包括离群点的模型的残差标准差。它们相差大吗?为什么?

**11.90** 求出练习 11.89 的剔除离群点后的回归模型。

- 解释截距(常数)项。在问题的背景下,这个数有什么意义?
- 在问题的背景下,斜率为零意味着什么?零假设为零斜率是否能被显著地拒绝?
- 计算总体的斜率真值的 95% 置信区间。计算机的输出结果应该给出估计的斜率以及标准误差,但读者可能需要自己做剩下的计算。

**11.91**

- 如果可能,根据练习 11.89 的剔除离群点的回归模型,利用计算机程序求出一个 5,000 平方英尺的房屋的价格的 95% 预测区间。如果你必须是自己做这个计算,先从计算机的输出结果中获取户型数据的均值和标准差,然后计算  $S_{xx} = (n-1)s^2$ 。根据这个数据所作的预测是明智的吗?
- 绘制价格对户型的散点图,同方差的假定是合理的吗?或者是方差随户型增大而增大?
- 关于(b)的答案对(a)的预测区间有什么影响?

**11.92 (商务)** 某个草种植公司想要预测不同邮政地区的服务需求,预测变量是邮政地区内住户密度。公司经理收集了每个邮政地区的住户数和地理位置,并计算了每千户的销售和每英亩的住户数。这些数据都存贮在出版社网站的数据集中 EX1192.DAT 文件,其中第一列是销售数据,第二列是密度(住户/英亩)。使

用统计软件读出这个文件。

- a. 求出这两个变量之间的相关系数, 这个数的符号意味什么?
- b. 求出因变量为销售, 自变量为密度的预测方程。解释截距(这个解释会有点奇怪)和斜率。
- c. 求出残差标准差, 关于预测的准确性, 这个数说明什么?

#### 11.93

- a. 求出练习 11.92 的回归模型的  $t$  统计量的值, 有没有确凿的证据表明密度是销售的预测变量?
- b. 计算斜率真值的 95% 置信区间。

**11.94** 绘制练习 11.92 的销售关于密度的散点图, 这个散点图显示出直线型的预测有意义吗?

**11.95** 参照练习 11.92。使用计算机程序计算新的变量 1/密度。

- a. 新变量的含义是什么? 特别地, 如果新变量的值是 0.50, 对于这个特殊的邮政地区有什么意义?
- b. 绘制销售关于新变量的散点图, 这里用直线型的预测合理吗?
- c. 求出销售和新变量间的相关系数, 并与练习 11.92 的销售与密度的相关系数进行比较, 二者的差别的解释是什么?

**11.96 (工程)** 某个生产公路表面标识所用油漆的制造商, 需要对开发的新配方的耐久力进行检验, 其中关心的一个问题是油漆中颜料的浓度。如果颜料浓度太低, 油漆就会很快褪色; 如果颜料浓度太高, 油漆在公路表面就不会持久。制造商将不同颜料浓度的油漆用于样本公路表面上, 并收集这些样本公路的耐久力测量值。这些数据都存贮在出版社网站的数据集中的 EX1196.DAT 文件, 其中第一列是耐久力数据, 第二列是颜料浓度。

- a. 利用计算机软件求出用浓度预测耐久力的回归方程, 并解释斜率。
- b. 求出决定系数。关于浓度的预测价值, 它说明什么?

**11.97** 练习 11.96 的回归模型中, 在  $\alpha = 0.01$  下斜率显著不为零吗?

**11.98** 绘制练习 11.96 的数据的散点图, 其中耐久力为纵轴, 浓度为横轴。

- a. 关于用直线型的预测这个散点图表明了什么?
- b. 关于练习 11.96 的相关系数这个散点图表明了什么?

**11.99 (商业)** 以前, 我们讨论过一个建筑商团体, 这个团体研究了估计建设定制房屋的成本的方法。这个问题需要你的进一步的建议。

回想建筑商用这个方法估计了 10 个“特别的”没有客户协议的房屋的成本。建筑商也获得了完成每个房屋的实际成本(用 Actual 表示)(不包括土地成本), 并与估计的成本(Estimated)比较。

“我们回到了会计那里, 他对数据进行了回归分析, 并得出了结果。会计说估

计是十分准确的,且具有 80% 的相关性和很低的  $p$  值。我们仍然怀疑这个新方法是否给出了合适的估计。由于我们只有 10% 的利润,很小的误差也会造成损害。你能介绍关于这个估计方法计算机输出结果说明什么?”

写出一个简洁的、非专业的介绍。介绍主要集中在建筑商关于估计准确性的问题上。散点图如下所示。

MTB>Regress 'Actual' on 1 variable 'Estimate'.

The regression equation is

Actual = - 34739 + 1.25 Estimate

| Predictor | Coef    | Stdev  | t-ratio | p     |
|-----------|---------|--------|---------|-------|
| Constant  | - 34739 | 60147  | - 0.58  | 0.579 |
| Estimate  | 1.2474  | 0.3293 | 3.79    | 0.005 |

s = 19313      R-sq = 64.2 %      R-sq(adj) = 59.7 %

Analysis of Variance

| SOURCE     | DF | SS         | MS         | F     | P     |
|------------|----|------------|------------|-------|-------|
| Regression | 1  | 5350811136 | 5350811136 | 14.35 | 0.005 |
| Error      | 8  | 2983948032 | 372993504  |       |       |
| Total      | 9  | 8334758912 |            |       |       |

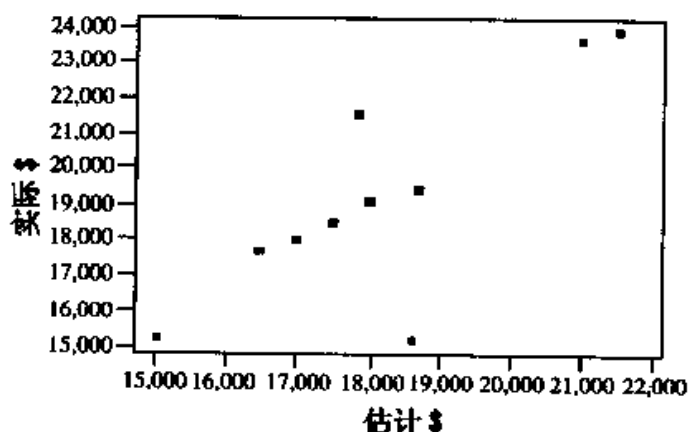
Unusual Observations

| Obs.   | Estimate | Actual  | Fit     | Stdev. Fit | Residual | St. Resid |
|--------|----------|---------|---------|------------|----------|-----------|
| 2      | 186200   | 152134  |         |            |          |           |
| 197531 | 6286     | - 45397 | - 2.49R |            |          |           |

R denotes an obs. with a large st. resid.

MTB>Correlation 'Estimate' 'Actual'.

Correlation of Estimate and Actual = 0.801



## 第十二章 多元回归与一般线性模型

- 12.1 引言和案例
- 12.2 一般线性模型
- 12.3 估计多元回归系数
- 12.4 多元回归中的推断
- 12.5 回归系数子集的检验
- 12.6 用多元回归进行的预测
- 12.7 比较几条回归线的斜率
- 12.8 Logistic 回归
- 12.9 多元回归的一些理论结果(任选)
- 12.10 小结

### 12.1 引言和案例

**案例：设计一种电钻,使其在高强度的使用下不发热**

为家用电动工具制造商工作的工程师们要设计一种电钻,这种电钻在高强度的使用下不发热。设计中三种关键的因子为:绝缘材料的厚度,马达中所使用的线的质量以及电钻钻体上排风口的大小。

**数据收集的设计** 这些工程师已经学习了一些线外质量控制的知识,所以他们用这些因子设计了一个试验。在这三个设计因子的每个水平组合上制作了十把电钻,并把它们分成两批,对这两批分别在两个(假设它们是等价的)“痛苦的检测”下进行测试。每次检测结束时测量钻体的温度,并对于每个批,计算其平均温度以及温度方差的对数。工程师们既要极小化这个均值,又要极小化方差的对数。

**数据的整理** 对于这个试验,有三个重要的因子:

IT 是电钻的绝缘层厚度(IT = 2, 3, 4, 5 或 6)

QW 是马达中所使用的线的质量(QW = 6, 7 或 8)

以及

VS, 即电钻钻体上排风口的大小(VS = 10, 11 或 12)。

这些设计因子共有  $5 \times 3 \times 3 = 45$  个不同的组合。对于因子的每个组合,生产了十把电钻,把这十把电钻分成两批,每批五把。对每把电钻都要进行高强度的检测,并在检测结束时记录其温度。下面的  $\bar{avtem}$  表示给定设计因子的组合下一批中五把电钻的平均温度。这里给出 90( $5 \times 3 \times 3 \times 2$  批)个平均温度的测量值。数据集合中还包括:

$\log v$  = 对于给定的因子组合和批(lot), 其温度的方差的对数

I2, Q2 和 V2 分别对应于三个设计因子的平方项, 计算公式为:

(设计因子 - 平均设计因子)<sup>2</sup>

| avtem | logv | IT | QW | VS | I2 | Q2 | V2 | Lot | avtem | logv | IT | QW | VS | I2 | Q2 | V2 | Lot |
|-------|------|----|----|----|----|----|----|-----|-------|------|----|----|----|----|----|----|-----|
| 185   | 3.6  | 2  | 6  | 10 | 4  | 1  | 1  | 1   | 168   | 3.4  | 4  | 7  | 11 | 0  | 0  | 0  | 2   |
| 176   | 3.7  | 2  | 6  | 10 | 4  | 1  | 1  | 2   | 160   | 2.9  | 4  | 7  | 12 | 0  | 0  | 1  | 1   |
| 177   | 3.6  | 2  | 6  | 11 | 4  | 1  | 0  | 1   | 154   | 3.1  | 4  | 7  | 12 | 0  | 0  | 1  | 2   |
| 184   | 3.7  | 2  | 6  | 11 | 4  | 1  | 0  | 2   | 169   | 2.8  | 4  | 8  | 10 | 0  | 1  | 1  | 1   |
| 178   | 3.6  | 2  | 6  | 12 | 4  | 1  | 1  | 1   | 156   | 2.9  | 4  | 8  | 10 | 0  | 1  | 1  | 2   |
| 169   | 3.4  | 2  | 6  | 12 | 4  | 1  | 1  | 2   | 168   | 2.7  | 4  | 8  | 11 | 0  | 1  | 0  | 1   |
| 185   | 3.2  | 2  | 7  | 10 | 4  | 0  | 1  | 1   | 161   | 2.7  | 4  | 8  | 11 | 0  | 1  | 0  | 2   |
| 184   | 3.2  | 2  | 7  | 10 | 4  | 0  | 1  | 2   | 156   | 2.6  | 4  | 8  | 12 | 0  | 1  | 1  | 1   |
| 180   | 3.2  | 2  | 7  | 11 | 4  | 0  | 0  | 1   | 158   | 2.7  | 4  | 8  | 12 | 0  | 1  | 1  | 2   |
| 184   | 3.5  | 2  | 7  | 11 | 4  | 0  | 0  | 2   | 164   | 3.7  | 5  | 6  | 10 | 1  | 1  | 1  | 1   |
| 179   | 3.0  | 2  | 7  | 12 | 4  | 0  | 1  | 1   | 163   | 3.7  | 5  | 6  | 10 | 1  | 1  | 1  | 2   |
| 173   | 3.2  | 2  | 7  | 12 | 4  | 0  | 1  | 2   | 161   | 3.7  | 5  | 6  | 11 | 1  | 1  | 0  | 1   |
| 179   | 2.9  | 2  | 8  | 10 | 4  | 1  | 1  | 1   | 158   | 3.4  | 5  | 6  | 11 | 1  | 1  | 0  | 2   |
| 185   | 2.7  | 2  | 8  | 10 | 4  | 1  | 1  | 2   | 154   | 3.4  | 5  | 6  | 12 | 1  | 1  | 1  | 1   |
| 180   | 2.8  | 2  | 8  | 11 | 4  | 1  | 0  | 1   | 162   | 3.7  | 5  | 6  | 12 | 1  | 1  | 1  | 2   |
| 180   | 2.7  | 2  | 8  | 11 | 4  | 1  | 0  | 2   | 163   | 2.8  | 5  | 7  | 10 | 1  | 0  | 1  | 1   |
| 169   | 2.9  | 2  | 8  | 12 | 4  | 1  | 1  | 1   | 166   | 3.0  | 5  | 7  | 10 | 1  | 0  | 1  | 2   |
| 177   | 2.8  | 2  | 8  | 12 | 4  | 1  | 1  | 2   | 159   | 3.3  | 5  | 7  | 11 | 1  | 0  | 0  | 1   |
| 172   | 3.6  | 3  | 6  | 10 | 1  | 1  | 1  | 1   | 156   | 3.3  | 5  | 7  | 11 | 1  | 0  | 0  | 2   |
| 171   | 3.9  | 3  | 6  | 10 | 1  | 1  | 1  | 2   | 152   | 3.3  | 5  | 7  | 12 | 1  | 0  | 1  | 1   |
| 172   | 3.8  | 3  | 6  | 11 | 1  | 1  | 0  | 1   | 150   | 3.3  | 5  | 7  | 12 | 1  | 0  | 1  | 2   |
| 167   | 3.6  | 3  | 6  | 11 | 1  | 1  | 0  | 2   | 165   | 2.9  | 5  | 8  | 10 | 1  | 1  | 1  | 1   |
| 165   | 3.3  | 3  | 6  | 12 | 1  | 1  | 1  | 1   | 156   | 2.7  | 5  | 8  | 10 | 1  | 1  | 1  | 2   |
| 159   | 3.4  | 3  | 6  | 12 | 1  | 1  | 1  | 2   | 155   | 2.8  | 5  | 8  | 11 | 1  | 1  | 0  | 1   |
| 169   | 3.0  | 3  | 7  | 10 | 1  | 0  | 1  | 1   | 155   | 3.2  | 5  | 8  | 11 | 1  | 1  | 0  | 2   |
| 174   | 3.3  | 3  | 7  | 10 | 1  | 0  | 1  | 2   | 149   | 2.6  | 5  | 8  | 12 | 1  | 1  | 1  | 1   |
| 163   | 3.3  | 3  | 7  | 11 | 1  | 0  | 0  | 1   | 152   | 2.9  | 5  | 8  | 12 | 1  | 1  | 1  | 2   |
| 170   | 3.3  | 3  | 7  | 11 | 1  | 0  | 0  | 2   | 165   | 3.4  | 6  | 6  | 10 | 4  | 1  | 1  | 1   |
| 169   | 3.2  | 3  | 7  | 12 | 1  | 0  | 1  | 1   | 160   | 3.7  | 6  | 6  | 10 | 4  | 1  | 1  | 2   |
| 163   | 3.2  | 3  | 7  | 12 | 1  | 0  | 1  | 2   | 157   | 3.7  | 6  | 6  | 11 | 4  | 1  | 0  | 1   |
| 178   | 2.7  | 3  | 8  | 10 | 1  | 1  | 1  | 1   | 149   | 3.7  | 6  | 6  | 11 | 4  | 1  | 0  | 2   |
| 165   | 2.7  | 3  | 8  | 10 | 1  | 1  | 1  | 2   | 149   | 3.8  | 6  | 6  | 12 | 4  | 1  | 1  | 1   |
| 167   | 2.8  | 3  | 8  | 11 | 1  | 1  | 0  | 1   | 145   | 3.7  | 6  | 6  | 12 | 4  | 1  | 1  | 2   |
| 171   | 2.8  | 3  | 8  | 11 | 1  | 1  | 0  | 2   | 154   | 3.4  | 6  | 7  | 10 | 4  | 0  | 1  | 1   |
| 166   | 2.9  | 3  | 8  | 12 | 1  | 1  | 1  | 1   | 153   | 3.2  | 6  | 7  | 10 | 4  | 0  | 1  | 2   |
| 166   | 2.7  | 3  | 8  | 12 | 1  | 1  | 1  | 2   | 150   | 3.0  | 6  | 7  | 11 | 4  | 0  | 0  | 1   |
| 161   | 3.7  | 4  | 6  | 10 | 0  | 1  | 1  | 1   | 156   | 3.1  | 6  | 7  | 11 | 4  | 0  | 0  | 2   |
| 162   | 3.7  | 4  | 6  | 10 | 0  | 1  | 1  | 2   | 146   | 3.2  | 6  | 7  | 12 | 4  | 0  | 1  | 1   |
| 169   | 3.4  | 4  | 6  | 11 | 0  | 1  | 0  | 1   | 153   | 3.3  | 6  | 7  | 12 | 4  | 0  | 1  | 2   |
| 162   | 3.7  | 4  | 6  | 11 | 0  | 1  | 0  | 2   | 161   | 2.8  | 6  | 8  | 10 | 4  | 1  | 1  | 1   |
| 159   | 3.5  | 4  | 6  | 12 | 0  | 1  | 1  | 1   | 160   | 2.9  | 6  | 8  | 10 | 4  | 1  | 1  | 2   |

续表

| avtem | logv | IT | QW | VS | I2 | Q2 | V2 | Lot | avtem | logv | IT | QW | VS | I2 | Q2 | V2 | Lot |
|-------|------|----|----|----|----|----|----|-----|-------|------|----|----|----|----|----|----|-----|
| 168   | 3.4  | 4  | 6  | 12 | 0  | 1  | 1  | 2   | 156   | 2.9  | 6  | 8  | 11 | 4  | 1  | 0  | 1   |
| 169   | 3.1  | 4  | 7  | 10 | 0  | 0  | 1  | 1   | 150   | 2.7  | 6  | 8  | 11 | 4  | 1  | 0  | 2   |
| 165   | 3.2  | 4  | 7  | 10 | 0  | 0  | 1  | 2   | 149   | 2.9  | 6  | 8  | 12 | 4  | 1  | 1  | 1   |
| 163   | 3.2  | 4  | 7  | 11 | 0  | 0  | 0  | 1   | 151   | 2.8  | 6  | 8  | 12 | 4  | 1  | 1  | 2   |

**数据的分析** 工程师们要分析这些数据,并且要极小化温度均值和温度方差的对数。他们要求你找出哪个设计因子好像对均值有影响,影响到什么程度,哪一个影响方差,哪一个平方项是重要的,最后,他们还想知道试验结果与批号(相应于检验的类型)是否有关。对这些数据将使用本章中讨论的方法,在练习 12.65~12.67 中进行分析。

联系因变量和一个定量的自变量的最简单的回归模型就是在第十一章中讨论的模型

$$y = \beta + \beta_0 x + \epsilon$$

在假定“对于一个给定的  $x$  值  $\epsilon$  的均值(又称为  $\epsilon$  的期望值)为  $E(\epsilon) = 0$ ”的前提下,这个模型表明,对于给定的  $x$  值,  $y$  的期望值由直线

$$E(y) = \beta_0 + \beta_1 x$$

给出。

表 12.1 不同施肥量下 14 块大小相同的西红柿田的产量

| 地块 | 产量(蒲式尔) | 施肥量<br>(每块地的磅数) |
|----|---------|-----------------|
| 1  | 24      | 12              |
| 2  | 18      | 5               |
| 3  | 31      | 15              |
| 4  | 33      | 17              |
| 5  | 26      | 20              |
| 6  | 30      | 14              |
| 7  | 20      | 6               |
| 8  | 25      | 23              |
| 9  | 25      | 11              |
| 10 | 27      | 13              |
| 11 | 21      | 8               |
| 12 | 29      | 18              |
| 13 | 29      | 22              |
| 14 | 26      | 25              |

并非所有的数据集都能由期望为直线的模型来充分地描述。例如,考虑表 12.1 中的数据,这些数据给出了 14 块大小相同的西红柿田在不同施肥量下的产量(蒲式尔)。由散点图 12.1 容易看出,线性方程不能充分表示产量和图中范围上的施肥量的关系,其原因在于,尽管最大的施肥量确实提高了作物的产量,但施肥过量会造成损害。

对于这种情况,一个可能的模型是

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$$

我们仍然假定  $E(\epsilon) = 0$ , 于是对于给定的  $x$  值,  $y$  的期望值由

$$E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$$

给出。图 12.1 中画出了一条这样的线。

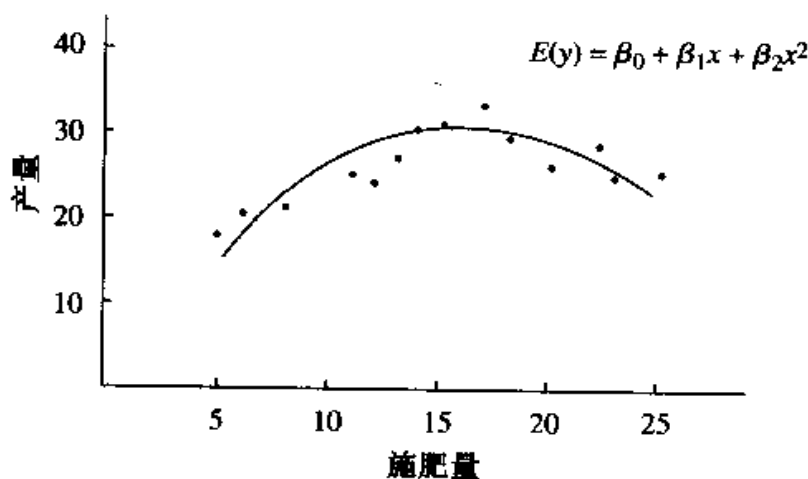


图 12.1 施肥量-产量数据的散点图

联系因变量  $y$  和一个定量的自变量  $x$  的一般的多项式回归模型为

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \epsilon$$

对于给定的  $x$  值,  $y$  的期望值为

$$E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p$$

$p$  的选取,也即合适的回归模型的选取,依赖于具体的试验。

联系因变量  $y$  和一组定量自变量的多元回归模型是一个自变量情形下多项式回归模型的直接推广。多元回归模型写为如下形式:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

这里任何自变量可能是其他自变量的幂,例如,  $x_2$  可能是  $x_1^2$ 。事实上,还有很多其他的可能性,  $x_3$  可能是一个交叉乘积  $x_1, x_2, x_4$  可能是  $\log x_1$ , 等等。惟一的限制是没有一个是其他  $x$  的完全线性函数。

多变量回归方程的最简单的形式是一阶模型,其中每一个自变量都出现,但没

有这些自变量的交叉乘积项或高次幂。例如,当有三个定量自变量时,一阶多元回归模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

对于这些一阶模型,我们可以赋予各个  $\beta$  以某种意义。参数  $b_0$  是  $y$  截距,表示当每个  $x$  均为 0 时  $y$  的期望值。对于令每个为 0 没有意义的情况,  $\beta_0$  (或其估计) 只用做预测方程的一部分,不对其做单独解释。

多元回归方程中的其他参数 ( $\beta_1, \beta_2, \dots, \beta_k$ ) 有时叫做偏斜率。在线性回归中,参数  $\beta_1$  为回归直线的斜率,它表示当  $x$  增加一个单位时,  $y$  的期望值的变化。在一阶多元回归模型中,  $\beta_1$  表示当  $x_1$  增加一个单位,而其他的各个  $x$  都保持不变时,  $y$  的期望值的变化。一般地,  $\beta_j (j \neq 0)$  表示当  $x_j$  增加一个单位,而其他的各个  $x$  都保持不变时,  $y$  的期望值的变化。对于一个多元回归模型,通常的假设如下。

**定义 12.1 多元回归的假设如下:**

1. 模型中关系的数学形式是正确的,从而对于所有的  $i$ ,  $E(\varepsilon_i) = 0$ 。
2. 对于所有  $i$ ,  $Var(\varepsilon_i) = \sigma_\varepsilon^2$ 。
3. 各个  $\varepsilon_i$  是相互独立的。
4.  $\varepsilon_i$  是正态分布的。

当我们应用一阶多元回归模型时,还有另外一个隐含的假设。由于  $x_j$  变化一个单位时,  $y$  的期望值的变化是一个常数,与其他  $x$  的取值无关,我们事实上是在假定这些自变量的效应是**可加的**。

### 例 12.1

一名商标管理人员收集了关于一个新产品的数据,其中  $y$  = 商标认可程度(能说出该商品是什么的潜在顾客的百分比),  $x_1$  = 一次电视广告中介绍该产品的时间长度(以秒为单位),  $x_2$  = 两周期间电视广告的重复次数。如果这名管理人员用一阶模型

$$\hat{y} = 0.31 + 0.042x_1 + 1.41x_2$$

来预测  $y$ , 他做了哪些假定?

**解答** 首先,他假定了一个线性模型,从而  $y$  随其中任何一个变量的变化速率是一致的。他假定电视广告每增加一秒,认可程度就增加 0.042 个百分点,而不论是从 10 秒增加到 11 秒还是从 59 秒增加到 60 秒。同样,广告每增加一次重复,认可程度就增加 1.41 个百分点,而不论是第二次重复还是第二十二次。

第二,他假定了没有交互作用。他使用的这个一阶模型假定了对于给定的广告持续时间(即  $x_1$  保持不变),增加一次广告重复(即  $x_2$  有一个增量)的效果不依赖于给定的广告持续时间是多久(10 秒, 27 秒, 60 秒还是任何其他什么值)。



什么时候可加性的假设是可以得到满足的呢? 图 12.2(a)显示了  $y$  对  $x_1$  的散点图, 而图 12.2(b)显示的是同样的数据散点图, 其中各点上有一个标识, 标出了第二个自变量  $x_2$  的不同的水平( $x_2$  取值 1, 2 或 3)。从图 12.2(a)我们看出,  $y$  关于  $x_1$  近似成线性趋势; 图 12.2(b)中相应于自变量  $x_2$  的三个水平的三条线平行, 这表明, 不论  $x_2$  在什么水平上, 对于  $x_1$  的一个单位的变化, 我们可以预期的  $y$  的变化是一样的。这样的数据就表明了  $x_1$  与  $x_2$  的效应是可加的, 因而形如  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$  的一阶模型是合适的。

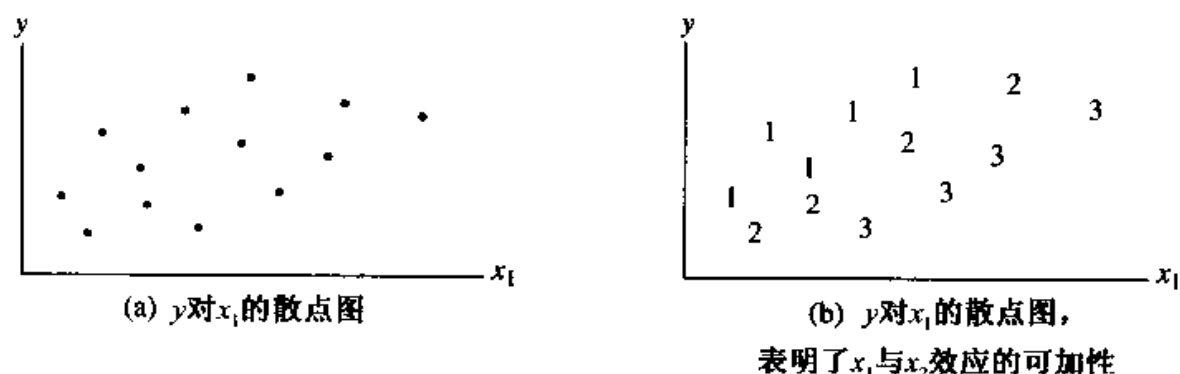


图 12.2

图 12.3 表明了一种在  $x_1$  与  $x_2$  间有交互效应存在的情形。虽然  $y$  对  $x_1$  的散点图与图 12.2(a)类似, 但图 12.3 中的三条线不平行, 这表明, 对于  $x_1$  的一个单位的变化, 可以预期的  $y$  的变化与  $x_2$  的水平有关。当这种现象发生时, 我们就说自变量  $x_1$  与  $x_2$  是交互的。此时, 假定效应具有可加性的一阶模型是不适用的。最粗浅的做法是在模型中加入交叉乘积项  $x_1 x_2$ 。

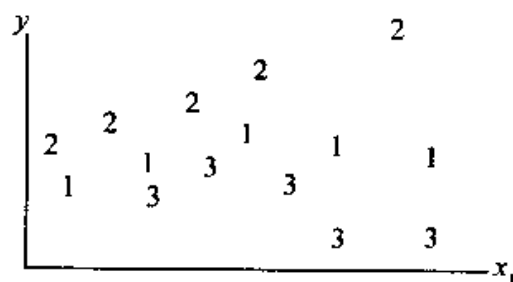


图 12.3  $y$  对  $x_1$  的散点图, 表明了  $x_1$  与  $x_2$  效应的不可加性(交互效应)

允许  $x_1$  与  $x_2$  间的交互效应的最简单的模型是

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

注意, 对于一个给定的  $x_2$  的值(比如说  $x_2 = 2$ ),  $y$  的期望值为

$$\begin{aligned} E(y) &= \beta_0 + \beta_1 x_1 + \beta_2(2) + \beta_3 x_1(2) \\ &= (\beta_0 + 2\beta_2) + (\beta_1 + 2\beta_3) x_1 \end{aligned}$$

此处,截距和斜率分别为 $(\beta_0 + 2\beta_2)$ 和 $(\beta_1 + 2\beta_3)$ 。对于 $x_2 = 3$ ,相应的截距和斜率则分别为 $(\beta_0 + 3\beta_2)$ 和 $(\beta_1 + 3\beta_3)$ 。显然,两条回归线的斜率是不同的,因而两条线是不平行的。

并非所有的试验都可以用一阶多元回归模型来刻画。在这样的情况下,高阶的多元回归模型可能是适用的,此时,因为模型中包含自变量的交叉乘积项或高阶项,各 $\beta$ 的解释也更加困难。我们将着重于找出对于样本数据拟合良好的多元回归模型,并把各个 $\beta$ 放入整个模型进行考虑,而不把注意力集中于单个 $\beta$ 的解释。

我们刚刚讨论过的模型,是针对这样的回归问题的,即试验者要建立一个模型,来反映一个响应变量与一个或多个定量自变量之间的关系。为一个试验建立模型的问题不限于定量自变量的情形。

考虑这样的问题:建立响应变量 $y$ 与一组定性自变量或既有定量自变量也有定性自变量的一组变量之间的关系模型。对于第一种情形(联系 $y$ 于一个或一组定性自变量),假定我们要比较相距30英里的两个观测站所观测到的一场暴风雨当中每分钟的平均雷电放电次数。用 $y$ 表示在一个示波镜上记录到的一分钟期间的放电次数,我们可以写出以下两个模型:

$$\text{观测站 1: } y = \mu_1 + \epsilon$$

$$\text{观测站 2: } y = \mu_2 + \epsilon$$

这就是说,我们假定观测站1上的观察值在总体均值 $\mu_1$ 附近随机“波动”,同样,观测站2上的观察值与总体均值 $\mu_2$ 相差一个随机的量 $\epsilon$ 。这两个模型不是新的,在第六章中比较两个总体的均值时已经用来描述从这些总体得到的观察值。这里我们所要讲到的新的内容是,我们可以把这两个模型结合成一个形如

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

的模型,其中 $\beta_0$ 和 $\beta_1$ 为未知参数, $\epsilon$ 为随机误差, $x_1$ 是一个虚拟变量,其意义如下。我们令

$x_1 = 1$  如果观察值是从观测站2得到的

$x_1 = 0$  如果观察值是从观测站1得到的

对于从观测站1得到的观察值,我们把 $x_1 = 0$ 代入到模型,得到

$$y = \beta_0 + \beta_1(0) + \epsilon = \beta_0 + \epsilon$$

因此, $\beta_0 = \mu_1$ ,即观测站1上观测值的总体平均。类似地,把 $x_1 = 1$ 代入到我们的模型,得到

$$y = \beta_0 + \beta_1(1) + \epsilon = \beta_0 + \beta_1 + \epsilon$$

由于 $\beta_0 = \mu_1$ 以及 $\beta_0 + \beta_1$ 必等于 $\mu_2$ ,我们有 $\beta_1 = \mu_2 - \mu_1$ ,即观测站2与观测站1

上观测值的平均差异。

联系  $y$  与定性自变量“观测站”的模型  $y = \beta_0 + \beta_1 x_1 + \epsilon$  可以扩充至定性变量有多于两个水平的情形。我们可以使用多于一个的虚拟变量来做到这一点。考虑一个这样的试验,其中定量变量有四个水平。我们称这些水平为处理。我们可以写出模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

其中

$x_1 = 1$  如果是处理 2,  $x_1 = 0$  其他

$x_2 = 1$  如果是处理 3,  $x_2 = 0$  其他

$x_3 = 1$  如果是处理 4,  $x_3 = 0$  其他

为了解释这个方程中的各个  $\beta$ , 构造一个期望值的表很方便。由于  $\epsilon$  的期望值为 0,  $y$  的期望值的一般表达式为

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

处理 1 上观察值的期望值由在上式中令  $x_1 = 0, x_2 = 0, x_3 = 0$  得到, 代入后得  $E(y) = \beta_0$ 。处理 2 上观察值的期望值由在上式中令  $x_1 = 1, x_2 = 0, x_3 = 0$  得到, 代入后得  $E(y) = \beta_0 + \beta_1$ 。代入  $x_1 = 0, x_2 = 1, x_3 = 0$  和  $x_1 = 0, x_2 = 0, x_3 = 1$ , 分别得到处理 3 和处理 4 上的期望值。这些期望值总结在表 12.2 中。

表 12.2 有四个水平的试验的期望值

| 处 理              |                            |                            |                            |
|------------------|----------------------------|----------------------------|----------------------------|
| 1                | 2                          | 3                          | 4                          |
| $E(y) = \beta_0$ | $E(y) = \beta_0 + \beta_1$ | $E(y) = \beta_0 + \beta_2$ | $E(y) = \beta_0 + \beta_3$ |

如果我们用  $\mu_1$  表示处理 1 上的均值,  $\mu_2$  表示处理 2 上的均值, 等等, 那么从表 12.2 得到

$$\mu_1 = \beta_0 \quad \mu_2 = \beta_0 + \beta_1 \quad \mu_3 = \beta_0 + \beta_2 \quad \mu_4 = \beta_0 + \beta_3$$

关于其中的  $\beta$  解方程, 得到

$$\beta_0 = \mu_1 \quad \beta_1 = \mu_2 - \mu_1 \quad \beta_2 = \mu_3 - \mu_1 \quad \beta_3 = \mu_4 - \mu_1$$

这些处理均值之间的任何比较都可以用  $\beta$  来表示。例如, 比较  $\mu_4 - \mu_3$  可以写为  $\beta_3 - \beta_2$ ,  $\mu_3 - \mu_2$  可以写为  $\beta_2 - \beta_1$ 。

### 例 12.2

考虑一个假想的试验, 其中有四个处理 ( $t = 4$ ), 并且这四个处理上的均值已

知。如果  $\mu_1 = 7$ ,  $\mu_2 = 9$ ,  $\mu_3 = 6$ ,  $\mu_4 = 15$ , 试确定模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

中  $\beta_0, \beta_1, \beta_2$  和  $\beta_3$  的值, 其中

$x_1 = 1$  如果是处理 2,  $x_1 = 0$  其他

$x_2 = 1$  如果是处理 3,  $x_2 = 0$  其他

$x_3 = 1$  如果是处理 4,  $x_3 = 0$  其他

**解答** 基于我们从表 12.2 中看到的, 得

$$\beta_0 = \mu_1 \quad \beta_1 = \mu_2 - \mu_1 \quad \beta_2 = \mu_3 - \mu_1 \quad \beta_3 = \mu_4 - \mu_1$$

把  $\mu_1, \mu_2, \mu_3$  和  $\mu_4$  的已知值代入得

$$\beta_0 = 7 \quad \beta_1 = 9 - 7 = 2 \quad \beta_2 = 6 - 7 = -1 \quad \beta_3 = 15 - 7 = 8$$

### 例 12.3

参见例 12.2。用  $\beta$  表示  $\mu_3 - \mu_2$  和  $\mu_3 - \mu_4$ , 并做  $\beta$  之间的减法来检查你的答案。

**解答** 利用  $\beta$  和  $\mu$  之间的关系, 我们可以看出

$$\beta_2 - \beta_1 = (\mu_3 - \mu_1) - (\mu_2 - \mu_1) = \mu_3 - \mu_2$$

以及

$$\beta_2 - \beta_3 = (\mu_3 - \mu_1) - (\mu_4 - \mu_1) = \mu_3 - \mu_4$$

用  $\beta$  的值计算相应的差, 得

$$\beta_2 - \beta_1 = -1 - (2) = -3$$

和

$$\beta_2 - \beta_3 = -1 - (8) = -9$$

这些计算结果分别与  $\mu_3 - \mu_2$  和  $\mu_3 - \mu_4$  的已知结果一致。

### 例 12.4

用虚拟变量为具有  $t$  个处理的试验写出一个模型, 并识别其中的  $\beta$ 。

**解答** 我们可以写出如下形式的模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{t-1} x_{t-1} + \epsilon$$

其中

$x_1 = 1$  如果是处理 2,  $x_1 = 0$  其他

$x_2 = 1$  如果是处理 3,  $x_2 = 0$  其他

$\vdots$

$x_{t-1} = 1$  如果是处理  $t$ ,  $x_{t-1} = 0$  其他

该模型的期望值表为

| 处 理              |                            |     |                                |
|------------------|----------------------------|-----|--------------------------------|
| 1                | 2                          | ... | $t$                            |
| $E(y) = \beta_0$ | $E(y) = \beta_0 + \beta_1$ | ... | $E(y) = \beta_0 + \beta_{t-1}$ |

由此, 我们得到

$$\beta_0 = \mu_1$$

$$\beta_1 = \mu_2 - \mu_1$$

$\vdots$

$$\beta_{t-1} = \mu_t - \mu_1$$

在上述给出的描述方法中, 响应变量与定性变量“处理”相联系, 对于处理的  $t$  个水平, 我们在模型中用虚拟变量加入了  $t-1$  个  $\beta$ 。关于有多个定性自变量的模型, 我们将在第十五和十六章中进行进一步的讨论, 在那里, 我们将考虑几个不同的试验设计的方差分析。

## 12.2 一般线性模型

现在, 重要的是我们可以看到, 对于联系一个响应变量和一组定量自变量的多元回归模型以及联系  $y$  和一组定性自变量的模型, 都可以用一个一般的模型来表示。这个模型称为 **一般线性模型**, 具有如下形式:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

对于多元回归模型, 其中的  $x$  表示定量自变量(如水的重量或体积)、自变量的幂和自变量的交叉乘积。我们已经在 12.1 节中讨论过几个回归模型; 本章后面以及第十三章我们还将对一般线性模型在回归中的应用做进一步的讨论。

当  $y$  联系于一组定性变量时, 一般线性模型中的  $x$  表示虚拟变量(编码为 0 和 1)或虚拟变量的乘积。我们在 12.1 节中讨论了在有一个定性变量时如何用虚拟变量来表示  $y$ 。同样的方法也可以建立  $y$  与多个定性变量之间的关系模型, 我们将在第十五章中讨论这样的模型, 在那里, 我们将更详细地讨论方差分析。

当  $y$  既联系于定性变量也联系于定量变量时, 也可以用一般线性模型。12.7

节中将给出一个这种情况的特殊例子, 其他的应用将在第十六章中讨论。

为什么这个模型称为一般线性模型呢, 况且该模型还可以用来表示多项式模型? 一般线性模型中的“线性”一词是指诸  $\beta$  进入模型的方式, 而不是指自变量在模型中出现的方式。一个一般线性模型关于诸  $\beta$  (在通常的代数意义下) 是线性的。

为什么我们现在讨论一般线性模型呢? 我们在这一章中将要给出的关于多元回归模型中单个参数  $\beta$ , 一组  $\beta$  以及  $E(y)$  的推断方法适用于任何一般线性模型。也就是说, 使用一般线性模型的方法, 我们得到关于多元回归的推断 (第十二和第十三章) 以及方差分析 (第十五章至第十九章) 的一条共同的线索。当你学习这六章的时候, 尽可能随时回过头来与一般线性模型相联系, 我们将帮助你建立这种联系。从这一章的 12.3 节到 12.10 节, 我们将集中讨论一般线性模型的特殊情形, 即多元回归模型。

## 练习

### 基本技能

#### 12.1

- 写出一个联系响应变量  $y$  与三个定性自变量的一阶多元回归模型。
- 说明这个模型可以写成一个一般线性模型。

**12.2** 写出一个联系响应变量  $y$  与三个定量自变量的二阶多元回归模型。(提示: 一阶模型包含有  $x_j$  的项, 二阶模型除包含有这些项外, 还含有二阶项和交叉项。)

**12.3** 参见练习 12.2。说明你给出的模型可以写成一个一般线性模型的形式, 并说明一般线性模型中各项的意义。

#### 12.4 考虑模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

其中

$$x_1 = \begin{cases} 1 & \text{对于处理 2} \\ 0 & \text{其他} \end{cases} \quad x_2 = \begin{cases} 1 & \text{对于处理 3} \\ 0 & \text{其他} \end{cases}$$

- 解释模型中的各个  $\beta$ 。
- 用该模型表示处理 2 与处理 3 上平均响应的差。

**12.5 (选择题)** 参见练习 12.4。假定在模型中增加一项  $\beta_3 x_3$ , 其中  $x_3$  表示由定性变量“位置”所确定的虚拟变量:

$$x_3 = \begin{cases} 1 & \text{对于位置 2} \\ 0 & \text{其他} \end{cases}$$

- 解释模型中的各个  $\beta$ 。(提示: 考虑三个处理和两个位置的所有组合。)

b. 写出对于位置 2, 处理 2 与处理 3 上平均响应的差。对于位置 1, 这个差还一样吗?

c. 给出一个试验的例子, 使得对于这个试验, 该模型可能是一个合理的近似。

**12.6 (选择题)** 在一项研究中, 考察了一个定量自变量(年龄)对于反应时间(用刹车时间来测量)的影响。试验对象中有男性也有女性。有人提出了两个模型:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \text{ 和 } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

其中

$$x_1 = \text{年龄(岁)} \quad \text{和} \quad x_2 = \begin{cases} 1 & \text{女性} \\ 0 & \text{男性} \end{cases}$$

给出这两个模型中诸  $\beta$  的解释, 并说明这两个模型的实际区别。

### 12.3 估计多元回归系数

多元回归模型建立了  $y$  与一组定量自变量之间的联系。对于一个  $n$  次测量的随机样本, 我们可以把第  $i$  个观测值写成

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad (i = 1, 2, \cdots, n, n > k)$$

其中  $x_{i1}, x_{i2}, \cdots, x_{ik}$  为相应于观测值  $y_i$  的各定量自变量的取值。

为了得到多元回归模型中  $\beta_0, \beta_1, \cdots$  和  $\beta_k$  的最小二乘估计, 我们应用第十一章中对于线性回归模型所使用的方法。我们有  $n$  次观测的随机样本, 通过选取  $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$  使之极小化  $SS(\text{残差}) = \sum_i (y_i - \hat{y}_i)^2$ , 找到最小二乘预测方程

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k.$$

然而, 虽然对于线性回归方程

$$y = \beta_0 + \beta_1 x + \varepsilon$$

容易写出  $\beta_0$  和  $\beta_1$  的解, 但现在我们必须对于  $\beta_0, \beta_1, \cdots$  和  $\beta_k$  同时解一组方程才能得到它们的估计, 这组方程叫做**正规方程**, 如下所示。

|          | $y_i$            |   | $\hat{\beta}_0$            |   | $x_{i1}\hat{\beta}_1$            |   | $\cdots$ |   | $x_{ik}\hat{\beta}_k$            |
|----------|------------------|---|----------------------------|---|----------------------------------|---|----------|---|----------------------------------|
| 1        | $\sum y_i$       | = | $n\hat{\beta}_0$           | + | $\sum x_{i1}\hat{\beta}_1$       | + | $\cdots$ | + | $\sum x_{ik}\hat{\beta}_k$       |
| $x_{i1}$ | $\sum x_{i1}y_i$ | = | $\sum x_{i1}\hat{\beta}_0$ | + | $\sum x_{i1}^2\hat{\beta}_1$     | + | $\cdots$ | + | $\sum x_{i1}x_{ik}\hat{\beta}_k$ |
| $\vdots$ | $\vdots$         |   |                            |   |                                  |   |          |   |                                  |
| $x_{ik}$ | $\sum x_{ik}y_i$ | = | $\sum x_{ik}\hat{\beta}_0$ | + | $\sum x_{ik}x_{i1}\hat{\beta}_1$ | + | $\cdots$ | + | $\sum x_{ik}^2\hat{\beta}_k$     |

注意这组方程的特点:在象上表中所做的那样标出行和列以后,我们可以通过行列相乘再求和来得到正规方程中的任何一项。例如,第二个方程中的最后一项可以这样得到:行元素  $x_{i1}$  乘以列元素  $(x_{ik}\hat{\beta}_k)$ , 再求和, 结果是  $\sum x_{i1}x_{ik}\hat{\beta}_k$ 。由于所有正规方程中的项都可以如是得到, 要写出这些用以求解最小二乘估计  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$  的方程容易得很。这些方程的解不一定是显而易见的, 这也就是为什么我们将借助于各种各样的软件包来求解的原因所在。

### 例 12.5

在练习 11.44 中, 我们给出了在暴露于空气中不同的时间后一种化合物的重量损失。另外还有可以利用的信息, 即暴露过程中环境的湿度。完全的数据在表 12.3 中给出。

表 12.3 重量损失、暴露时间和相对湿度数据

| 重量损失 $y$ (磅) | 暴露时间 $x_1$ (小时) | 相对湿度, $x_2$ |
|--------------|-----------------|-------------|
| 4.3          | 4               | 0.20        |
| 5.5          | 5               | 0.20        |
| 6.8          | 6               | 0.20        |
| 8.0          | 7               | 0.20        |
| 4.0          | 4               | 0.30        |
| 5.2          | 5               | 0.30        |
| 6.6          | 6               | 0.30        |
| 7.5          | 7               | 0.30        |
| 2.0          | 4               | 0.40        |
| 4.0          | 5               | 0.40        |
| 5.7          | 6               | 0.40        |
| 6.5          | 7               | 0.40        |

a. 如果假定的模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

其中  $x_1$  为暴露时间,  $x_2$  为相对湿度, 试对这一回归问题建立正规方程。

b. 从下面给出的计算机输出结果确定  $\beta_0, \beta_1$  和  $\beta_2$  的最小二乘估计, 预测当暴露时间为 6.5 小时, 相对湿度为 0.35 时的重量损失。

(注: 在下面结果中, OBS 表示观察值序列号, WT\_LOSS 表示重量损失, TIME 表示时间, HUMID 表示相对湿度)



OUTPUT FOR EXAMPLE 12.5

| OBS | WT _ LOSS | TIME | HUMID |
|-----|-----------|------|-------|
| 1   | 4.3       | 4.0  | 0.20  |
| 2   | 5.5       | 5.0  | 0.20  |
| 3   | 6.8       | 6.0  | 0.20  |
| 4   | 8.0       | 7.0  | 0.20  |
| 5   | 4.0       | 4.0  | 0.30  |
| 6   | 5.2       | 5.0  | 0.30  |
| 7   | 6.6       | 6.0  | 0.30  |
| 8   | 7.5       | 7.0  | 0.30  |
| 9   | 2.0       | 4.0  | 0.40  |
| 10  | 4.0       | 5.0  | 0.40  |
| 11  | 5.7       | 6.0  | 0.40  |
| 12  | 6.5       | 7.0  | 0.40  |
| 13  | .         | 6.5  | 0.35  |

Dependent Variable: WT \_ LOSS WEIGHT LOSS

Analysis of Variance

| Source  | DF       | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----------|----------------|-------------|---------|--------|
| Model   | 2        | 31.12417       | 15.56208    | 104.133 | 0.0001 |
| Error   | 9        | 1.34500        | 0.14944     |         |        |
| C Total | 11       | 32.46917       |             |         |        |
|         |          |                |             |         |        |
|         | Root MSE | 0.38658        | R-square    | 0.9586  |        |
|         | Dep Mean | 5.50833        | Adj R-sq    | 0.9494  |        |
|         | C. V.    | 7.01810        |             |         |        |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter = 0 | Prob>  T |
|----------|----|--------------------|----------------|-------------------------|----------|
| INTERCEP | 1  | 0.666667           | 0.69423219     | 0.960                   | 0.3620   |
| TIME     | 1  | 1.316667           | 0.09981464     | 13.191                  | 0.0001   |
| HUMID    | 1  | 8.000000           | 1.36676829     | -5.853                  | 0.0002   |

| OBS | WT_LOSS | PRED    | RESID    | L95MEAN | U95MEAN |
|-----|---------|---------|----------|---------|---------|
| 1   | 4.3     | 4.33333 | -0.03333 | 3.80985 | 4.85682 |
| 2   | 5.5     | 5.65000 | -0.15000 | 5.23519 | 6.06481 |
| 3   | 6.8     | 6.96667 | -0.16667 | 6.55185 | 7.38148 |
| 4   | 8.0     | 8.28333 | -0.28333 | 7.75985 | 8.80682 |
| 5   | 4.0     | 3.53333 | 0.46667  | 3.11091 | 3.95576 |
| 6   | 5.2     | 4.85000 | 0.35000  | 4.57346 | 5.12654 |
| 7   | 6.6     | 6.16667 | 0.43333  | 5.89012 | 6.44321 |
| 8   | 7.5     | 7.48333 | 0.01667  | 7.06091 | 7.90576 |
| 9   | 2.0     | 2.73333 | -0.73333 | 2.20985 | 3.25682 |
| 10  | 4.0     | 4.05000 | -0.05000 | 3.63519 | 4.46481 |
| 11  | 5.7     | 5.36667 | 0.33333  | 4.95185 | 5.78148 |
| 12  | 6.5     | 6.68333 | -0.18333 | 6.15985 | 7.20682 |
| 13  | .       | 6.42500 | .        | 6.05269 | 6.79731 |

Sum of Residuals 0  
Sum of Squared Residuals 1.3450  
Predicted Resid SS (Press) 2.6123

### 解答

a. 对于这个模型,三个正规方程如下。

|          | $y_i$            |   | $\hat{\beta}_0$            |   | $x_{i1}\hat{\beta}_1$            |   | $x_{i2}\hat{\beta}_2$            |
|----------|------------------|---|----------------------------|---|----------------------------------|---|----------------------------------|
| 1        | $\sum y_i$       | = | $n\hat{\beta}_0$           | + | $\sum x_{i1}\hat{\beta}_1$       | + | $\sum x_{i2}\hat{\beta}_2$       |
| $x_{i1}$ | $\sum x_{i1}y_i$ | = | $\sum x_{i1}\hat{\beta}_0$ | + | $\sum x_{i1}^2\hat{\beta}_1$     | + | $\sum x_{i1}x_{i2}\hat{\beta}_2$ |
| $\vdots$ | $\vdots$         |   |                            |   |                                  |   |                                  |
| $x_{i2}$ | $\sum x_{i2}y_i$ | = | $\sum x_{i2}\hat{\beta}_0$ | + | $\sum x_{i2}x_{i1}\hat{\beta}_1$ | + | $\sum x_{i2}^2\hat{\beta}_2$     |

对于这些数据,我们有

$$\begin{aligned} \sum y_i &= 66.10 & \sum x_{i1} &= 66 & \sum x_{i2} &= 3.60 \\ \sum x_{i1}y_i &= 383.3 & \sum x_{i2}y_i &= 19.19 & \sum x_{i1}x_{i2} &= 19.8 \\ \sum x_{i1}^2 &= 378 & \sum x_{i2}^2 &= 1.16 \end{aligned}$$

把这些数代入正规方程,得到如下结果:

$$\begin{aligned} 66.1 &= 12\hat{\beta}_0 + 66\hat{\beta}_1 + 3.6\hat{\beta}_2 \\ 383.3 &= 66\hat{\beta}_0 + 378\hat{\beta}_1 + 19.8\hat{\beta}_2 \\ 19.19 &= 3.6\hat{\beta}_0 + 19.8\hat{\beta}_1 + 1.16\hat{\beta}_2 \end{aligned}$$

- b. 解(a)中的正规方程可得  $\hat{\beta}_0, \hat{\beta}_1$  和  $\hat{\beta}_2$ 。这个解当与这里给出的输出结果的数值一致。最小二乘预测方程为

$$\hat{y} = 0.667 + 1.317x_1 - 8.000x_2$$

其中  $x_1$  为暴露时间,  $x_2$  为相对湿度。把  $x_1 = 6.5$  及  $x_2 = 0.35$  代入上述方程, 我们有

$$\hat{y} = 0.667 + 1.317(6.5) - 8.000(0.35) = 6.428$$

该值除了舍入误差以外, 与计算机输出结果中第 13 个观测值处的预测值一致。

有很多软件程序可以用来计算一般线性模型中参数的最小二乘估计。这些程序的输出中, 一般都含有变量名的一个列表, 并附有所估计的偏斜率, 这些偏斜率被标以“COEFFICIENTS”(系数), 也有的标以“ESTIMATES”(估计)或“PARAMETERS”(参数)。截距  $\hat{\beta}_0$  通常叫做“INTERCEPT”(截距)或“CONSTANT”(常数), 有时, 它也被与斜率一起列出, 只是不带变量名。

### 例 12.6

下面给出的三个变量的数据用 Excel 中的 spreadsheet 程序进行了分析。找出偏斜率和截距的估计。

|         |     |     |     |    |    |    |    |    |    |
|---------|-----|-----|-----|----|----|----|----|----|----|
| y:      | 25  | 34  | 28  | 40 | 36 | 42 | 44 | 53 | 49 |
| $x_1$ : | -10 | -10 | -10 | 0  | 0  | 0  | 10 | 10 | 10 |
| $x_2$ : | -5  | 0   | 5   | -5 | 0  | 5  | -5 | 0  | 5  |

|           | Coefficients | Standard Error | t Stat | P-value |
|-----------|--------------|----------------|--------|---------|
| Intercept | 39.0         | 1.256          | 31.055 | 7.4E-08 |
| X1        | 0.983        | 0.154          | 6.393  | 0.0007  |
| X2        | 0.333        | 0.308          | 1.084  | 0.3202  |

**解答** 如在输出结果中所标出的, 截距的值为 39.0。与  $x_1$  和  $x_2$  相应的偏斜率分别为 0.983 和 0.333。很多程序都用类似的方法在一列中标出回归系数。

在多元回归方程中, 自变量  $x_j$  的系数一般不等于对该自变量做简单线性回归时它的系数。在多元回归中,  $x_j$  的系数表示当其他自变量保持不变时, 该自变量变化的效应。在简单线性回归中, 所有其他潜在的自变量都被省略。如果其他自变量与  $x_j$  是相关的(因此当  $x_j$  变化时有不能保持常数的趋势), 用  $x_j$  作为惟一自变量的简单线性回归不仅捕捉了  $x_j$  变化时的直接效应, 而且也反映了与之相关的其他  $x$  的变化的间接效应。在多元回归中, 通过令其他变量保持不变, 我们摆脱了这种间接效应。

**例 12.7**

比较下面的 StataQuest 程序的输出结果中,多元回归模型与简单(一个预测因子)回归模型中  $x_1$  的系数。解释这两个系数不同的原因。

```
. regress y x1 x2
```

| y     | Coef. | Std. Err. | t     | P> t  | [95% Conf. Interval] |          |
|-------|-------|-----------|-------|-------|----------------------|----------|
| x1    | 1     | 1.870829  | 0.535 | 0.646 | -7.049526            | 9.049526 |
| x2    | 3     | 4.1833    | 0.717 | 0.548 | -14.99929            | 20.99929 |
| _cons | 10    | 1.183216  | 8.452 | 0.014 | 4.909033             | 15.09097 |

```
. regress y x1
```

| y     | Coef. | Std. Err. | t     | P> t  | [95% Conf. Interval] |          |
|-------|-------|-----------|-------|-------|----------------------|----------|
| x1    | 2.2   | 0.7659417 | 2.872 | 0.064 | -0.2375683           | 4.637568 |
| _cons | 10    | 1.083205  | 9.232 | 0.003 | 6.552758             | 13.44724 |

```
. correlate y x1 x2
```

|    | y      | x1     | x2     |
|----|--------|--------|--------|
| y  | 1.0000 |        |        |
| x1 | 0.8563 | 1.0000 |        |
| x2 | 0.8704 | 0.8944 | 1.0000 |

**解答** 在多元回归模型中,  $x_1$  的系数为 1, 但在简单回归模型中, 其系数为 2.2。之所以有如此差异, 是因为两个  $x$  是相关的(结果中显示二者相关系数为 0.8944)。在多元回归模型中, 我们考虑固定  $x_2$ , 而  $x_1$  变化; 在简单回归模型中, 我们考虑  $x_1$  变化而让  $x_2$  自由。

除了估计截距和偏斜率外, 估计**残差标准差**  $s_e$  也很重要。残差标准差也称为估计的标准误差。残差的定义仍然如前, 即  $y$  的观测值与预测值之间的差:

$$y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik})$$

残差平方和  $SS(\text{残差})$ , 又叫做  $SS(\text{误差})$ , 正如其名, 定义为各个预测误差的平方的和:

$$\begin{aligned} SS(\text{残差}) &= \sum (y_i - \hat{y}_i)^2 \\ &= \sum [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik})]^2 \end{aligned}$$

该平方和的自由度是  $n - (k + 1)$ 。一个自由度是由于截距而减去的, 而对应于  $k$

个偏斜率中的每一个也减去一个自由度。均方残差 MS(残差), 也叫做 MS(误差), 定义为残差平方和除以  $n - (k + 1)$ 。最后, 残差标准差  $s_e$  是 MS(残差) 的平方根。

残差标准差也叫做“std dev”, “估计的标准误差”, 或“MSE 方根”。如果输出结果不清楚, 你可以手工对 MS(残差) 开平方。一如既往, 我们总是用经验准则来解释标准差, 即预测误差的大约 95% 落在均值的  $\pm 2$  残差标准差的范围内, 预测误差的平均值自然是 0。

$$s_e = \sqrt{\text{MS(残差)}} = \sqrt{\frac{\text{SS(残差)}}{n - (k + 1)}}$$

### 例 12.8

在下面给出的例 12.6 中数据的输出结果中找出 SS(残差) 以及  $s_e$ 。

|    | A   | B             | C                  | D       | E     | F     |
|----|---|---------------|--------------------|---------|-------|-------|
| 1  |   |               |                    |         |       |       |
| 2  | The regression equation is $y = 39.0 + 0.983 x_1 + 0.333 x_2$ |               |                    |         |       |       |
| 3  |   |               |                    |         |       |       |
| 4  | Predictor   | Coef          | Stdev              | t-ratio | p     |       |
| 5  |   |               |                    |         |       |       |
| 6  | Constant  | 39.000        | 1.256              | 31.05   | 0.000 |       |
| 7  | x1  | 0.9833        | 0.1538             | 6.39    | 0.001 |       |
| 8  | x2  | 0.3333        | 0.3076             | 1.08    | 0.320 |       |
| 9  |   |               |                    |         |       |       |
| 10 | $s = 3.768$   | R-sq = 87.5 % | R-sq(adj) = 83.3 % |         |       |       |
| 11 |   |               |                    |         |       |       |
| 12 |   |               |                    |         |       |       |
| 13 | Analysis of Variance  |               |                    |         |       |       |
| 14 |   |               |                    |         |       |       |
| 15 |   |               |                    |         |       |       |
| 16 | SOURCE  | DF            | SS                 | MS      | F     | P     |
| 17 | Regression  | 2             | 596.83             | 298.420 | 21.02 | 0.002 |
| 18 | Error   | 6             | 85.17              | 14.190  |       |       |
| 19 | Total   | 8             | 682.00             |         |       |       |

**解答** 在标有“Analysis of Variance”(方差分析)的一段输出结果中, SS(残差)显示为  $SS(\text{Error}) = 85.17$ ,  $df$ (即自由度)为 6,  $MS$ (误差)为 14.19。残差标准差由  $s = 3.768$  标示出来。注意, 在不计舍入误差时, 有  $3.768 = \sqrt{14.19}$ 。

残差标准差在决定用回归方程作出的预测所可能有的误差时是很关键的。预测单个  $y$  值时使用的精确的标准误差将在 12.4 节中给出。若忽略外推和自由度效应, 粗略地说, 可能的误差近似为  $\pm 2s_e$ 。这个近似可以作为衡量一个回归模型的预测质量的一个粗略的指标。

### 例 12.9

一个商业学校的招生办公室建立了一个回归模型, 以用能力测验成绩和班级排名来预测等级分平均值( $4.00 = A$ ;  $2.00 = C$  平均, 最小的毕业平均值;  $0.00 = F$ )。残差标准差为  $s_e = 0.46$ 。问该值能表明用这个回归方程可给出很精确的预测吗?

**解答** 预测的可能误差的一种度量是  $2s_e = 0.92$ 。例如, 如果预测到的平均值为 2.80, 那么一个人的等级分大约在  $2.80 - 0.92 = 1.88$ (成绩不好不能毕业)到  $2.80 + 0.92 = 3.72$ (可以优异成绩毕业)! 这不是精确的预测。

练习

应用

**12.7 (医药)** 一家制药公司希望了解一种药物的剂量与效果之间的关系。为此, 在 15 个试管中接种一种病毒, 并在  $30^\circ\text{C}$  的条件下培养 5 天。要研究 5 个不同的剂量水平(2, 4, 8, 16 和 32 mg), 为每个剂量水平随机指定三个试管, 给每个试管只注射一个剂量水平的药物, 然后测量其响应(药物对培养病毒的作用强度的一种度量)。所得数据如下。

| 剂量水平 | 响应         |
|------|------------|
| 2    | 5, 7, 3    |
| 4    | 10, 12, 14 |
| 8    | 15, 17, 18 |
| 16   | 20, 21, 19 |
| 32   | 23, 24, 29 |

- 画出数据的图。
- 对这些数据拟合一个线性回归模型。
- 还有什么样的模型可能是合适的?

- d. 下面给出用 SAS 软件对线性和二次回归方程算得的输出结果。哪一个回归方程看起来拟合得更好？为什么？

OUTPUT FOR EXERCISE 12.7

| OBS | DOSE | RESPONSE |
|-----|------|----------|
| 1   | 2    | 5        |
| 2   | 2    | 7        |
| 3   | 2    | 3        |
| 4   | 4    | 10       |
| 5   | 4    | 12       |
| 6   | 4    | 14       |
| 7   | 8    | 15       |
| 8   | 8    | 17       |
| 9   | 8    | 18       |
| 10  | 16   | 20       |
| 11  | 16   | 21       |
| 12  | 16   | 19       |
| 13  | 32   | 23       |
| 14  | 32   | 24       |
| 15  | 32   | 29       |

Dependent Variable: RESPONSE PROTECTIVE STRENGTH

Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 590.91613      | 590.91613   | 44.280  | 0.0001 |
| Error   | 13 | 173.48387      | 13.34491    |         |        |
| C Total | 14 | 764.40000      |             |         |        |

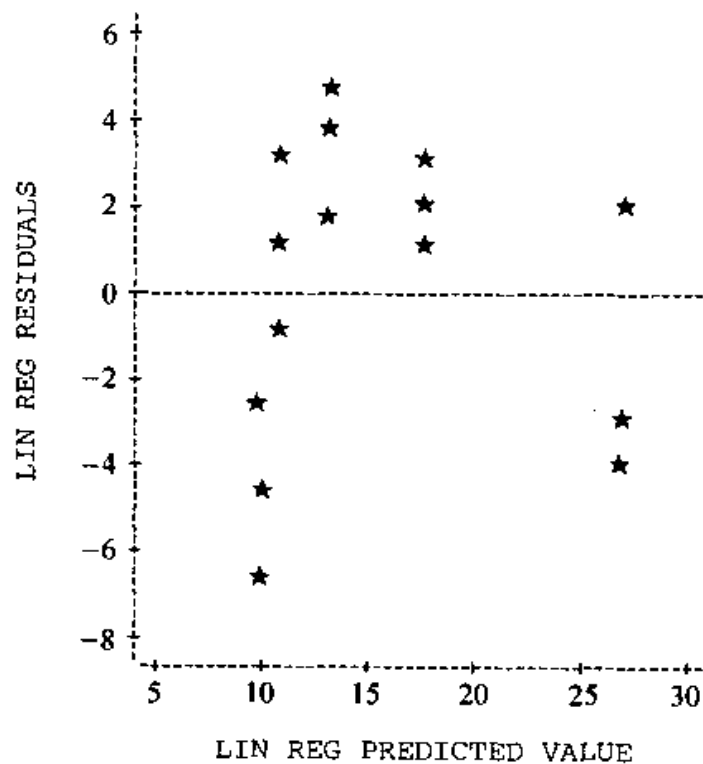
|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 3.65307  | R-square | 0.7730 |
| Dep Mean | 15.80000 | Adj R-sq | 0.7556 |
| C.V.     | 23.12069 |          |        |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|-------------------------|---------|
| INTERCEP | 1  | 8.666667           | 1.42786770     | 6.070                   | 0.0001  |
| DOSE     | 1  | 0.575269           | 0.08645016     | 6.654                   | 0.0001  |

| OBS | DOSE | RESPONSE | PRED    | RESID    |
|-----|------|----------|---------|----------|
| 1   | 2    | 5        | 9.8172  | -4.81720 |
| 2   | 2    | 7        | 9.8172  | -2.81720 |
| 3   | 2    | 3        | 9.8172  | -6.81720 |
| 4   | 4    | 10       | 10.9677 | -0.96774 |
| 5   | 4    | 12       | 10.9677 | 1.03226  |
| 6   | 4    | 14       | 10.9677 | 3.03226  |
| 7   | 8    | 15       | 13.2688 | 1.73118  |
| 8   | 8    | 17       | 13.2688 | 3.73118  |
| 9   | 8    | 18       | 13.2688 | 4.73118  |
| 10  | 16   | 20       | 17.8710 | 2.12903  |
| 11  | 16   | 21       | 17.8710 | 3.12903  |
| 12  | 16   | 19       | 17.8710 | 1.12903  |
| 13  | 32   | 23       | 27.0753 | -4.07527 |
| 14  | 32   | 24       | 27.0753 | -3.07527 |
| 15  | 32   | 29       | 27.0753 | 1.92473  |

Sum of Residuals                      0  
Sum of Squared Residuals            173.4839  
Predicted Resid SS (Press)        238.0013





# QUADRATIC REGRESSION ANALYSIS

Dependent Variable: RESPONSE PROTECTIVE STRENGTH

## Analysis of Variance

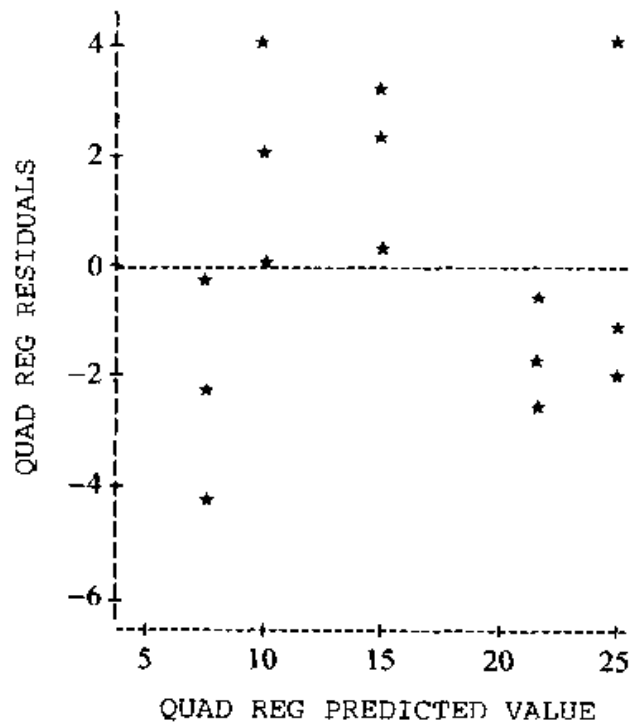
| Source  | DF | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|---------|----|-------------------|----------------|---------|--------|
| Model   | 2  | 673.82062         | 336.91031      | 44.634  | 0.0001 |
| Error   | 12 | 90.57938          | 7.54828        |         |        |
| C Total | 14 | 764.40000         |                |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 2.74741  | R-square | 0.8815 |
| Dep Mean | 15.80000 | Adj R-sq | 0.8618 |
| C.V.     | 17.38869 |          |        |

## Parameter Estimates

| Variable | DF | Parameter<br>Estimate | Standard<br>Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|-----------------------|-------------------|----------------------------|---------|
| INTERCEP | 1  | 4.483660              | 1.65720388        | 2.706                      | 0.0191  |
| DOSE     | 1  | 1.506325              | 0.28836373        | 5.224                      | 0.0002  |
| DOSE2    | 1  | -0.026987             | 0.00814314        | -3.314                     | 0.0062  |

| OBS | DOSE | RESPONSE | PREDICTED | RESIDUAL |
|-----|------|----------|-----------|----------|
| 1   | 2    | 5        | 7.3884    | -2.38836 |
| 2   | 2    | 7        | 7.3884    | -0.38836 |
| 3   | 2    | 3        | 7.3884    | -4.38836 |
| 4   | 4    | 10       | 10.0772   | -0.07717 |
| 5   | 4    | 12       | 10.0772   | 1.92283  |
| 6   | 4    | 14       | 10.0772   | 3.92283  |
| 7   | 8    | 15       | 14.8071   | 0.19292  |
| 8   | 8    | 17       | 14.8071   | 2.19292  |
| 9   | 8    | 18       | 14.8071   | 3.19292  |
| 10  | 16   | 20       | 21.6762   | -1.67615 |
| 11  | 16   | 21       | 21.6762   | -0.67615 |
| 12  | 16   | 19       | 21.6762   | -2.67615 |
| 13  | 32   | 23       | 25.0512   | -2.05123 |
| 14  | 32   | 24       | 25.0512   | -1.05123 |
| 15  | 32   | 29       | 25.0512   | 3.94877  |



**12.8** 参见练习 12.7。为了线性化响应与自变量之间的关系,常对剂量水平做对数变换。

- 查对数表或用计算器求出 5 个剂量水平的对数。
- 如果以  $x_i$  表示对数剂量,拟合模型

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

下面给出了拟合后的残差图。

- 对(b)中的结果与练习 12.7 中的结果进行比较。对数变换提供了比练习 12.7 中更好的线性拟合吗?

#### REGRESSION ANALYSIS USING NATURAL LOGARITHM OF DOSE

Dependent Variable: RESPONSE PROTECTIVE STRENGTH

Analysis of Variance

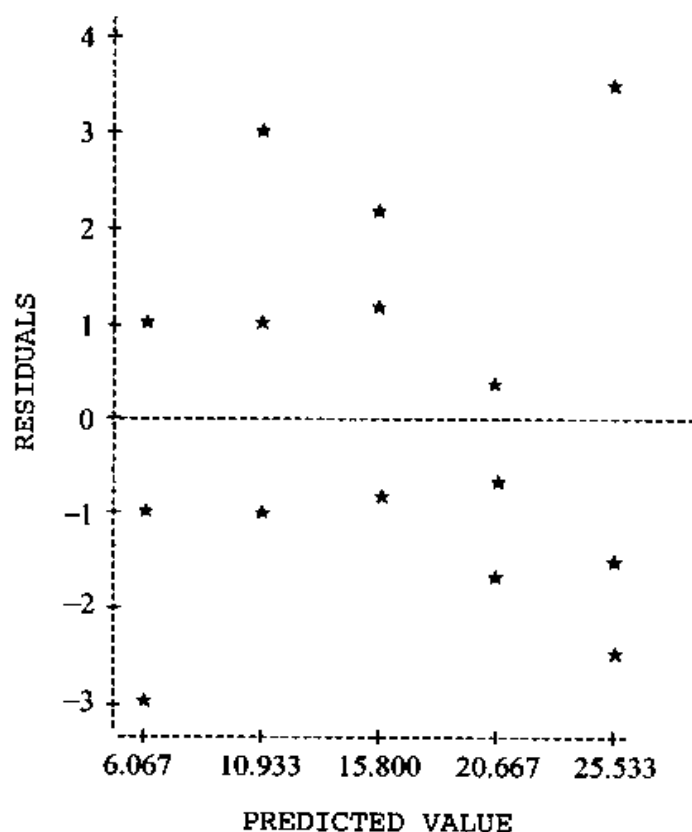
| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 710.53333      | 710.53333   | 171.478 | 0.0001 |
| Error   | 13 | 53.86667       | 4.14359     |         |        |
| C Total | 14 | 764.40000      |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 2.03558  | R-square | 0.9295 |
| Dep Mean | 15.80000 | Adj R-sq | 0.9241 |
| C.V.     | 12.88342 |          |        |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|----------------------------|---------|
| INTERCEP | 1  | 1.200000           | 1.23260547     | 0.974                      | 0.3480  |
| LOG_DOSE | 1  | 7.021116           | 0.53616972     | 13.095                     | 0.0001  |



**12.9 (商业)** 一家化工工业公司研究其促销活动(基本的直接接触和贸易展览)、直接开发费用以及短期研究投入对销售额的影响。他们收集了 24 个季度(6 年)的数据,并用 Stata 多元回归程序进行了分析,结果如下(每个季度 100,000 美元):

```
.regress Sale Promo Devel Research
```

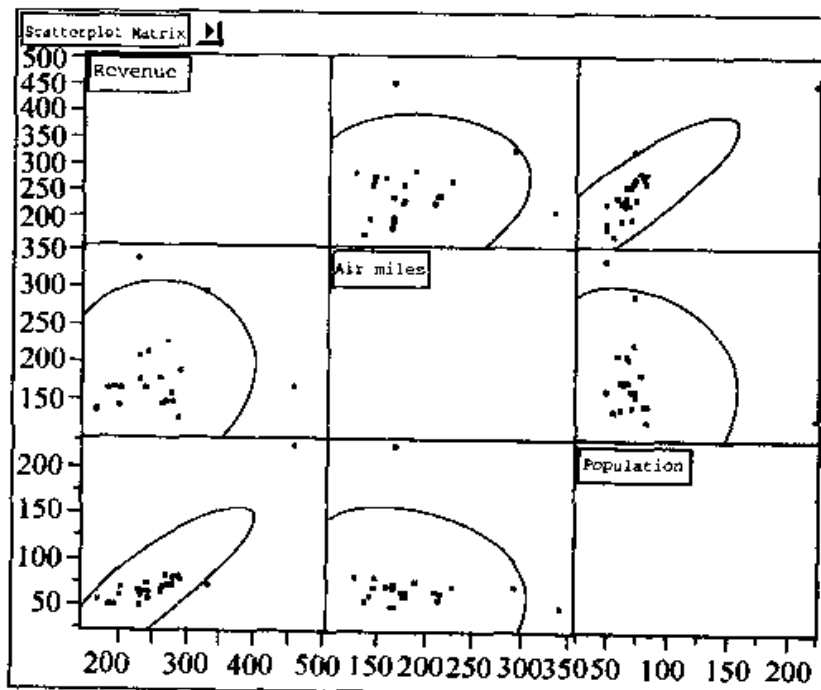
|          |            |    |            |                 |        |
|----------|------------|----|------------|-----------------|--------|
| Source   | SS         | df | MS         | Number of obs = | 24     |
| Model    | 43901.7677 | 3  | 14633.9226 | F(3, 20) =      | 22.28  |
| Residual | 13136.2323 | 20 | 656.811614 | Prob > F =      | 0.0000 |
|          |            |    |            | R-square =      | 0.7697 |
|          |            |    |            | Adj R-square =  | 0.7351 |
| Total    | 57038.00   | 23 | 2479.91304 | Root MSE =      | 25.628 |

| Sales    | Coef.     | Std. Err. | t      | p> T  | [95 % Conf. Interval] |          |
|----------|-----------|-----------|--------|-------|-----------------------|----------|
| Promo    | 136.0983  | 28.10759  | 4.842  | 0.000 | 77.46689              | 194.7297 |
| Devel    | -61.17526 | 50.94102  | -1.201 | 0.244 | -167.4364             | 45.08585 |
| Research | -43.69508 | 48.32298  | -0.904 | 0.377 | -144.495              | 57.10489 |
| _cons    | 326.3893  | 241.6129  | 1.351  | 0.192 | -177.6063             | 830.3849 |

- 写出估计后的回归方程。
- 找出 MS(残差) 以及其平方根, 残差标准差。
- 给出促销费用的系数  $\hat{\beta}_1$  的解释。

**12.10 (商业)** 在航空支线上, 航空公司把顾客从一些小城市运送到一个大的空中枢纽。为了基于每个小城市到该枢纽的距离(英里)和小城市的人口对通向 22 个小城市的支线收入进行预测, 做了一项回归研究。得到的相关系数和散点图如下。

| Correlations |         |           |            |
|--------------|---------|-----------|------------|
| Variable     | Revenue | Air miles | Population |
| Revenue      | 1.0000  | 0.1133    | 0.8632     |
| Air miles    | 0.1133  | 1.0000    | -0.1502    |
| Population   | 0.8632  | -0.1502   | 1.0000     |



- a. 这两个自变量是严重相关的吗?
- b. 这些散点图是否表明可能有杠杆点的问题?

12.11 用 JMP 软件对航空支线的数据进行了多元回归分析, 部分结果如下。

Response: Revenue

Summary of Fit

|                            |          |
|----------------------------|----------|
| RSquare                    | 0.805479 |
| RSquare Adj                | 0.785003 |
| Root Mean Square Error     | 27.7539  |
| Mean of Response           | 252.5    |
| Observations (or Sum Wgts) | 22       |

Parameter Estimates

| Term       | Estimate  | Std Error | t Ratio | Prob> t |
|------------|-----------|-----------|---------|---------|
| Intercept  | 86.736862 | 27.74907  | 3.13    | 0.0056  |
| Air miles  | 0.2922116 | 0.120336  | 2.43    | 0.0253  |
| Population | 1.5310653 | 0.174004  | 8.80    | 0.0000  |

Whole-Model Test

Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Ratio |
|---------|----|----------------|-------------|---------|
| Model   | 2  | 60602.202      | 30301.1     | 39.3378 |
| Error   | 19 | 14635.298      | 770.3       | Prob>F  |
| C Total | 21 | 75237.500      |             | 0.0000  |

- a. 不考虑这些输出结果, 你能否期望空中里程与人口的斜率的符号(正或负)吗?
- b. 上述输出中的斜率的符号是否与你所期待的符号一致?
- c. 给出空中里程的系数的意义。

12.12 (工程) 一家制造商要检验一台用于对试验纤维进行检验的磨损检测仪的研磨效果。研磨时机器有 6 种转速: 100, 120, 140, 160, 180 和 200 转每分钟。剪裁了 48 块同样质地的 5 英寸见方的纤维织品, 为每种转速随机指定其中 8 块。把这些织品指定到机器的顺序是随机的, 每一块织品在指定的转速下试验 3 分钟。测量并记录下每一块织品的磨损量, 数据如下。

| 转速(Speed)(转/分) | 磨损量(Wear)                                      |
|----------------|--|
| 100            | 23.0, 23.5, 24.4, 25.2, 25.6, 26.1, 24.8, 25.6 |
| 120            | 26.7, 26.1, 25.8, 26.3, 27.2, 27.9, 28.3, 27.4 |
| 140            | 28.0, 28.4, 27.0, 28.8, 29.8, 29.4, 28.7, 29.3 |
| 160            | 32.7, 32.1, 31.9, 33.0, 33.5, 33.7, 34.0, 32.5 |
| 180            | 43.1, 41.7, 42.4, 42.1, 43.5, 43.8, 44.2, 43.6 |
| 200            | 54.2, 43.7, 53.1, 53.8, 55.6, 55.9, 54.7, 54.5 |

- 画出该数据集的图(对于所有转速而言,磨损量的变异性都是一样的,所以你可以对每个转速画出其平均值,这样可以节省时间)。
- 什么样的回归模型看上去是合适的呢?
- 对于线性、二次和三次模型的输出结果如下面几页。哪一个回归方程的拟合效果更好些?为什么?
- 数据中有什么特殊的现象吗?可能发生了什么?

#### LINEAR REGRESSION ANALYSIS FOR WEAR TESTER DATA

Dependent Variable: FABRIC WEAR

##### Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 4326.79207     | 4326.79207  | 291.474 | 0.0001 |
| Error   | 46 | 682.84710      | 14.84450    |         |        |
| C Total | 47 | 5009.63917     |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 3.85286  | R-square | 0.8637 |
| Dep Mean | 34.92917 | Adj R-sq | 0.8607 |
| C.V.     | 11.03048 |          |        |

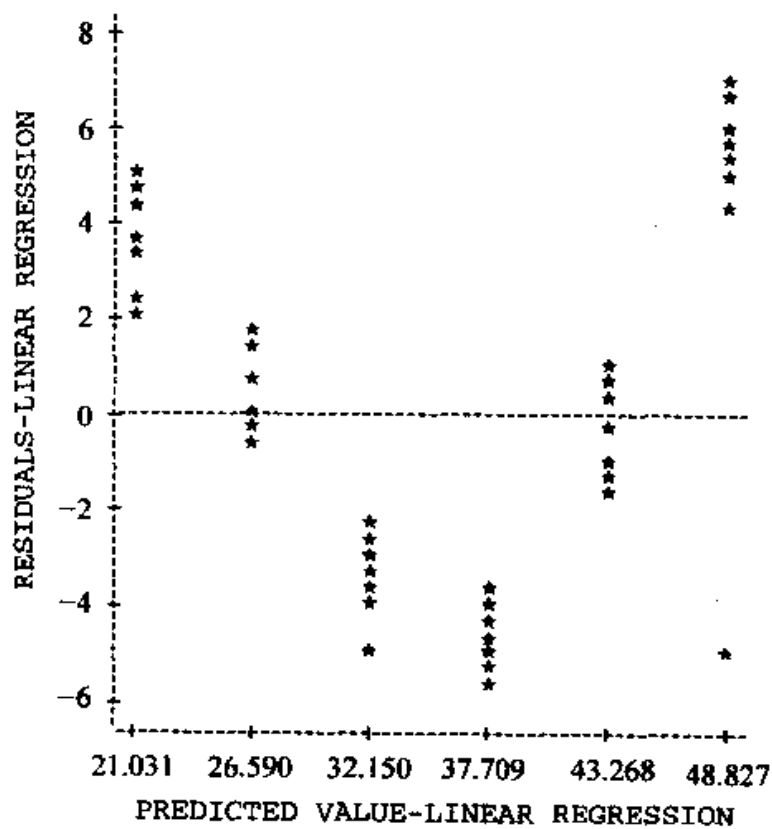
##### Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|-------------------------|---------|
| INTERCEP | 1  | -6.765476          | 2.50470943     | -2.701                  | 0.0096  |
| SPEED    | 1  | 0.277964           | 0.01628129     | 17.073                  | 0.0001  |

| Variable |    |               |
|----------|----|---------------|
| Variable | DF | Label         |
| INTERCEP | 1  | Intercept     |
| SPEED    | 1  | MACHINE SPEED |

| OBS | SPEED | WEAR | PRED1   | RESID1   |
|-----|-------|------|---------|----------|
| 1   | 100   | 23.0 | 21.0310 | 1.96905  |
| 2   | 100   | 23.5 | 21.0310 | 2.46905  |
| 3   | 100   | 24.4 | 21.0310 | 3.36905  |
| 4   | 100   | 25.2 | 21.0310 | 4.16905  |
| 5   | 100   | 25.6 | 21.0310 | 4.56905  |
| 6   | 100   | 26.1 | 21.0310 | 5.06905  |
| 7   | 100   | 24.8 | 21.0310 | 3.76905  |
| 8   | 100   | 25.6 | 21.0310 | 4.56905  |
| 9   | 120   | 26.7 | 26.5902 | 0.10976  |
| 10  | 120   | 26.1 | 26.5902 | -0.49024 |
| 11  | 120   | 25.8 | 26.5902 | -0.79024 |
| 12  | 120   | 26.3 | 26.5902 | -0.29024 |
| 13  | 120   | 27.2 | 26.5902 | 0.60976  |
| 14  | 120   | 27.9 | 26.5902 | 1.30976  |
| 15  | 120   | 28.3 | 26.5902 | 1.70976  |
| 16  | 120   | 27.4 | 26.5902 | 0.80976  |
| 17  | 140   | 28.0 | 32.1495 | -4.14952 |
| 18  | 140   | 28.4 | 32.1495 | -3.74952 |
| 19  | 140   | 27.0 | 32.1495 | -5.14952 |
| 20  | 140   | 28.8 | 32.1495 | -3.34952 |
| 21  | 140   | 29.8 | 32.1495 | -2.34952 |
| 22  | 140   | 29.4 | 32.1495 | -2.74952 |
| 23  | 140   | 28.7 | 32.1495 | -3.44952 |
| 24  | 140   | 29.3 | 32.1495 | -2.84952 |
| 25  | 160   | 32.7 | 37.7088 | -5.00881 |
| 26  | 160   | 32.1 | 37.7088 | -5.60881 |
| 27  | 160   | 31.9 | 37.7088 | -5.80881 |
| 28  | 160   | 33.0 | 37.7088 | -4.70881 |
| 29  | 160   | 33.5 | 37.7088 | -4.20881 |
| 30  | 160   | 33.7 | 37.7088 | -4.00881 |
| 31  | 160   | 34.0 | 37.7088 | -3.70881 |

| OBS | SPEED | WEAR | PRED1   | RESID1   |
|-----|-------|------|---------|----------|
| 32  | 160   | 32.5 | 37.7088 | -5.20881 |
| 33  | 180   | 43.1 | 43.2681 | -0.16810 |
| 34  | 180   | 41.7 | 43.2681 | -1.56810 |
| 35  | 180   | 42.4 | 43.2681 | -0.86810 |
| 36  | 180   | 42.1 | 43.2681 | -1.16810 |
| 37  | 180   | 43.5 | 43.2681 | 0.23190  |
| 38  | 180   | 43.8 | 43.2681 | 0.53190  |
| 39  | 180   | 44.2 | 43.2681 | 0.93190  |
| 40  | 180   | 43.6 | 43.2681 | 0.33190  |
| 41  | 200   | 54.2 | 48.8274 | 5.37262  |
| 42  | 200   | 43.7 | 48.8274 | -5.12738 |
| 43  | 200   | 53.1 | 48.8274 | 4.27262  |
| 44  | 200   | 53.8 | 48.8274 | 4.97262  |
| 45  | 200   | 55.6 | 48.8274 | 6.77262  |
| 46  | 200   | 55.9 | 48.8274 | 7.07262  |
| 47  | 200   | 54.7 | 48.8274 | 5.87262  |
| 48  | 200   | 54.5 | 48.8274 | 5.67262  |





NOTE: 6 obs hidden

QUADRATIC REGRESSION ANALYSIS FOR WEAR TESTER DATA

Dependent Variable: FABRIC WEAR

Analysis of Variance

| Source  | DF | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|---------|----|-------------------|----------------|---------|--------|
| Model   | 2  | 4839.89302        | 2419.94651     | 641.532 | 0.0001 |
| Error   | 45 | 169.74614         | 3.77214        |         |        |
| C Total | 47 | 5009.63917        |                |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 1.94220  | R-square | 0.9661 |
| Dep Mean | 34.92917 | Adj R-sq | 0.9646 |
| C.V.     | 5.56039  |          |        |

Parameter Estimates

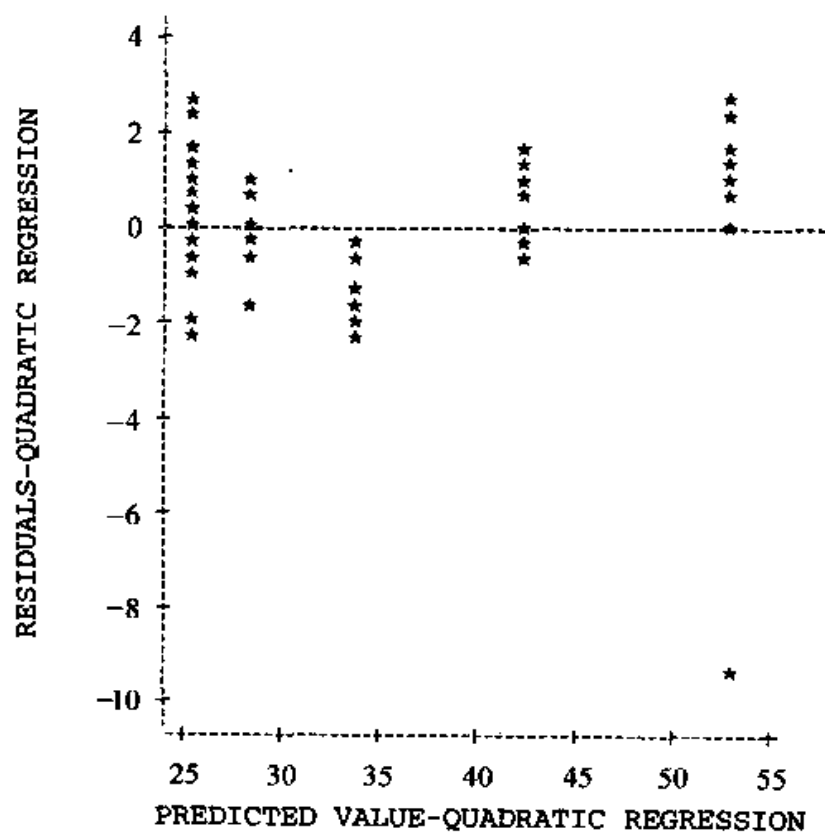
| Variable | DF | Parameter<br>Estimate | Standard<br>Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|-----------------------|-------------------|----------------------------|---------|
| INTERCEP | 1  | 63.139286             | 6.12529888        | 10.308                     | 0.0001  |
| SPEED    | 1  | -0.705071             | 0.08468583        | -8.326                     | 0.0001  |
| SPEED2   | 1  | 0.003277              | 0.00028096        | 11.663                     | 0.0001  |

| Variable | DF | Variable<br>Label |
|----------|----|-------------------|
| INTERCEP | 1  | Intercept         |
| SPEED    | 1  | MACHINE SPEED     |
| SPEED2   | 1  | SPEED SQUARED     |

| OBS | SPEED | WEAR | PRED2   | RESID2   |
|-----|-------|------|---------|----------|
| 1   | 100   | 23.0 | 25.4000 | -2.40000 |
| 2   | 100   | 23.5 | 25.4000 | -1.90000 |
| 3   | 100   | 24.4 | 25.4000 | -1.00000 |
| 4   | 100   | 25.2 | 25.4000 | -0.20000 |
| 5   | 100   | 25.6 | 25.4000 | 0.20000  |
| 6   | 100   | 26.1 | 25.4000 | 0.70000  |

| OBS | SPEED | WEAR | PRED2   | RESID2   |
|-----|-------|------|---------|----------|
| 7   | 100   | 24.8 | 25.4000 | -0.60000 |
| 8   | 100   | 25.6 | 25.4000 | 0.20000  |
| 9   | 120   | 26.7 | 25.7164 | 0.98357  |
| 10  | 120   | 26.1 | 25.7164 | 0.38357  |
| 11  | 120   | 25.8 | 25.7164 | 0.08357  |
| 12  | 120   | 26.3 | 25.7164 | 0.58357  |
| 13  | 120   | 27.2 | 25.7164 | 1.48357  |
| 14  | 120   | 27.9 | 25.7164 | 2.18357  |
| 15  | 120   | 28.3 | 25.7164 | 2.58357  |
| 16  | 120   | 27.4 | 25.7164 | 1.68357  |
| 17  | 140   | 28.0 | 28.6543 | -0.65429 |
| 18  | 140   | 28.4 | 28.6543 | -0.25429 |
| 19  | 140   | 27.0 | 28.6543 | -1.65429 |
| 20  | 140   | 28.8 | 28.6543 | 0.14571  |
| 21  | 140   | 29.8 | 28.6543 | 1.14571  |
| 22  | 140   | 29.4 | 28.6543 | 0.74571  |
| 23  | 140   | 28.7 | 28.6543 | 0.04571  |
| 24  | 140   | 29.3 | 28.6543 | 0.64571  |
| 25  | 160   | 32.7 | 34.2136 | -1.51357 |
| 26  | 160   | 32.1 | 34.2136 | -2.11357 |
| 27  | 160   | 31.9 | 34.2136 | -2.31357 |
| 28  | 160   | 33.0 | 34.2136 | -1.21357 |
| 29  | 160   | 33.5 | 34.2136 | -0.71357 |
| 30  | 160   | 33.7 | 34.2136 | -0.51357 |
| 31  | 160   | 34.0 | 34.2136 | -0.21357 |
| 32  | 160   | 32.5 | 34.2136 | -1.71357 |
| 33  | 180   | 43.1 | 42.3943 | 0.70571  |
| 34  | 180   | 41.7 | 42.3943 | -0.69429 |
| 35  | 180   | 42.4 | 42.3943 | 0.00571  |
| 36  | 180   | 42.1 | 42.3943 | -0.29429 |
| 37  | 180   | 43.5 | 42.3943 | 1.10571  |
| 38  | 180   | 43.8 | 42.3943 | 1.40571  |
| 39  | 180   | 44.2 | 42.3943 | 1.80571  |
| 40  | 180   | 43.6 | 42.3943 | 1.20571  |
| 41  | 200   | 54.2 | 53.1964 | 1.00357  |
| 42  | 200   | 43.7 | 53.1964 | -9.49643 |
| 43  | 200   | 53.1 | 53.1964 | -0.09643 |

| OBS | SPEED | WEAR | PRED2   | RESID2  |
|-----|-------|------|---------|---------|
| 44  | 200   | 53.8 | 53.1964 | 0.60357 |
| 45  | 200   | 55.6 | 53.1964 | 2.40357 |
| 46  | 200   | 55.9 | 53.1964 | 2.70357 |
| 47  | 200   | 54.7 | 53.1964 | 1.50357 |
| 48  | 200   | 54.5 | 53.1964 | 1.30357 |



#### CUBIC REGRESSION ANALYSIS FOR WEAR TESTER DATA

Dependent Variable: FABRIC WEAR

#### Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 3  | 4846.78202     | 1615.59401  | 436.494 | 0.0001 |
| Error   | 44 | 162.85714      | 3.70130     |         |        |
| C Total | 47 | 5009.63917     |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 1.92388  | R-square | 0.9675 |
| Dep Mean | 34.92917 | Adj R-sq | 0.9653 |
| C.V.     | 5.50794  |          |        |

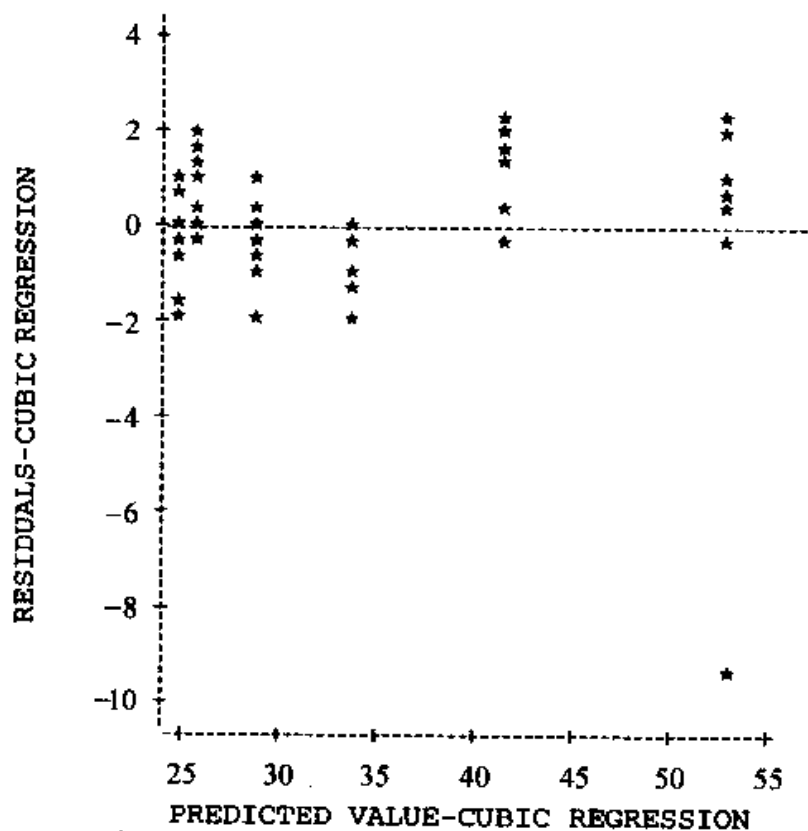
## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H <sub>0</sub> :<br>Parameter = 0 | Prob >  T |
|----------|----|--------------------|----------------|---|-----------|
| INTERCEP | 1  | 18.872619          | 33.00952220    | 0.572                                   | 0.5704    |
| SPEED    | 1  | 0.238477           | 0.69668199     | 0.342                                   | 0.7338    |
| SPEED2   | 1  | -0.003208          | 0.00476113     | -0.674                                  | 0.5040    |
| SPEED3   | 1  | 0.000014410        | 0.00001056     | 1.364                                   | 0.1794    |

| Variable | DF | Variable Label |
|----------|----|----------------|
| INTERCEP | 1  | Intercept      |
| SPEED    | 1  | MACHINE SPEED  |
| SPEED2   | 1  | SPEED SQUARED  |
| SPEED3   | 1  | SPEED CUBED    |

| OBS | SPEED | WEAR | PRED3   | RESID3   |
|-----|-------|------|---------|----------|
| 1   | 100   | 23.0 | 25.0542 | -2.05417 |
| 2   | 100   | 23.5 | 25.0542 | -1.55417 |
| 3   | 100   | 24.4 | 25.0542 | -0.65417 |
| 4   | 100   | 25.2 | 25.0542 | 0.14583  |
| 5   | 100   | 25.6 | 25.0542 | 0.54583  |
| 6   | 100   | 26.1 | 25.0542 | 1.04583  |
| 7   | 100   | 24.8 | 25.0542 | -0.25417 |
| 8   | 100   | 25.6 | 25.0542 | 0.54583  |
| 9   | 120   | 26.7 | 26.2006 | 0.49940  |
| 10  | 120   | 26.1 | 26.2006 | -0.10060 |
| 11  | 120   | 25.8 | 26.2006 | -0.40060 |
| 12  | 120   | 26.3 | 26.2006 | 0.09940  |
| 13  | 120   | 27.2 | 26.2006 | 0.99940  |
| 14  | 120   | 27.9 | 26.2006 | 1.69940  |
| 15  | 120   | 28.3 | 26.2006 | 2.09940  |
| 16  | 120   | 27.4 | 26.2006 | 1.19940  |
| 17  | 140   | 28.0 | 28.9310 | -0.93095 |
| 18  | 140   | 28.4 | 28.9310 | -0.53095 |
| 19  | 140   | 27.0 | 28.9310 | -1.93095 |
| 20  | 140   | 28.8 | 28.9310 | -0.13095 |
| 21  | 140   | 29.8 | 28.9310 | 0.86905  |
| 22  | 140   | 29.4 | 28.9310 | 0.46905  |
| 23  | 140   | 28.7 | 28.9310 | -0.23095 |
| 24  | 140   | 29.3 | 28.9310 | 0.36905  |
| 25  | 160   | 32.7 | 33.9369 | -1.23690 |
| 26  | 160   | 32.1 | 33.9369 | -1.83690 |
| 27  | 160   | 31.9 | 33.9369 | -2.03690 |
| 28  | 160   | 33.0 | 33.9369 | -0.93690 |

| OBS | SPEED | WEAR | PRED3   | RESID3   |
|-----|-------|------|---------|----------|
| 29  | 160   | 33.5 | 33.9369 | -0.43690 |
| 30  | 160   | 33.7 | 33.9369 | -0.23690 |
| 31  | 160   | 34.0 | 33.9369 | 0.06310  |
| 32  | 160   | 32.5 | 33.9369 | -1.43690 |
| 33  | 180   | 43.1 | 41.9101 | 1.18988  |
| 34  | 180   | 41.7 | 41.9101 | -0.21012 |
| 35  | 180   | 42.4 | 41.9101 | 0.48988  |
| 36  | 180   | 42.1 | 41.9101 | 0.18988  |
| 37  | 180   | 43.5 | 41.9101 | 1.58988  |
| 38  | 180   | 43.8 | 41.9101 | 1.88988  |
| 39  | 180   | 44.2 | 41.9101 | 2.28988  |
| 40  | 180   | 43.6 | 41.9101 | 1.68988  |
| 41  | 200   | 54.2 | 53.5423 | 0.65774  |
| 42  | 200   | 43.7 | 53.5423 | -9.84226 |
| 43  | 200   | 53.1 | 53.5423 | -0.44226 |
| 44  | 200   | 53.8 | 53.5423 | 0.25774  |
| 45  | 200   | 55.6 | 53.5423 | 2.05774  |
| 46  | 200   | 55.9 | 53.5423 | 2.35774  |
| 47  | 200   | 54.7 | 53.5423 | 1.15774  |
| 48  | 200   | 54.5 | 53.5423 | 0.95774  |



**12.13** 参见练习 12.12 中的数据。假定试验中另外一个变量是受到控制的,并且对于每个转速,前四块织品用 0.2 的浓度加了保护涂层,而后四块织品用 0.4 的浓度加了保护涂层。 $x_1$  表示机器转速, $x_2$  表示涂层的浓度。用手头的统计软件拟合下面的模型。哪个模型看起来能提供更好的拟合?为什么?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \epsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_1 x_2 + \beta_5 x_1^2 x_2 + \epsilon$$

## 12.4 多元回归中的推断

在线性回归模型  $y = \beta_0 + \beta_1 x + \epsilon$  中,我们对  $\beta_0$  和  $\beta_1$  进行推断。同样,我们对于一般线性模型(当然也包括多元回归),也考虑对于其中任何参数的推断。

然而,在此之前,我们必须首先介绍决定系数  $R^2$ 。决定系数  $R^2$  的定义和解释与第十一章中的  $r^2$  非常相似。(习惯上对于多元回归用  $R^2$ ,而对于简单线性回归则用  $r^2$ 。)像在第十一章中一样,我们把决定系数定义为在已知  $x_1, x_2, \dots, x_k$  时, $y$  的平方误差的缩减比例。例如,如果我们有一个三个自变量  $x_1, x_2, x_3$  的多元回归模型,并且  $R^2_{y \cdot x_1 x_2 x_3} = 0.736$ ,那么,我们由  $x_1, x_2$  和  $x_3$  的变异性可以解释  $y$  的变异性的 73.6%。确切地说,

$$R^2_{y \cdot x_1 \dots x_k} = \frac{SS(\text{总和}) - SS(\text{残差})}{SS \text{ 总和}}$$

其中

$$SS(\text{总和}) = \sum (y_i - \bar{y})^2$$

### 例 12.10

在例 12.8 的计算机输出结果中找出  $R^2_{y \cdot x_1 x_2}$ 。

**解答** 我们所求的是  $R - sq = 87.5\%$ ,而不是标有 adj 的那个值。换一种方法,输出结果表明,  $SS(\text{总和}) = 682.00$ ,  $SS(\text{残差}) = 85.17$ ,由此我们可以得出  $R^2_{y \cdot x_1 x_2} = (682.00 - 85.17)/682.00 = 0.875$ 。

在一个多元回归方程中,复决定系数  $R^2$  不小于任何一个单个变量的决定系数,除此以外, $R^2$  与单决定系数  $r^2_{yx_1}, r^2_{yx_2}, \dots, r^2_{yx_k}$  之间没有一般的关系。如果所有自变量相互之间是完全不相关的,那么  $R^2$  恰为所有单决定系数的和。换言之,如果所有  $x$  是互不相关的,则用所有预测因子建立的模型的  $SS(\text{回归})$  等于用每一个  $x$  所建立的简单回归模型的  $SS(\text{回归})$  的和。如果  $x$  之间是相关的,则很难把  $x_1, x_2, \dots, x_k$  的总预测价值  $R^2_{y \cdot x_1 \dots x_k}$  分解成单独的  $x_1$ ,单独的  $x_2, \dots$ ,单独的  $x_k$

的相应值。

当自变量之间相关时,有**共线性**(有时称为复共线性)存在。在多元回归中,我们试图分离各个预测因子的预测价值。当预测因子高度相关时,这样做非常困难。例如,假若我们要解释地方房产销售随时间的变异,用国民生产总值(GDP)和国家储蓄收入(DI)作为预测因子。DI几乎恰好与GDP成比例,所以这两个预测因子的相关性极高。现在你说,房产销售中的变异性是更多地产生于GDP的变异性还是产生于DI的变异性呢?如果你能把它们分开,那真算你走运!它们中的任何一个几乎与两个一起同样地能解释房产销售中的变异。

共线性通常在某种程度上存在于多元回归的研究中。当各个 $x$ 轻度相关时问题不大,但若相关很强,问题就严重了。因此,如果在一个回归研究中有共线性——事实上它通常在某种程度上存在——那就不容易把综合的 $R^2_{y \cdot x_1 \dots x_k}$ 分解成与各个 $x$ 对应的分量。相关的 $x$ 在解释 $y$ 的变异性时常常相互覆盖,所以常常(但不总是)有

$$R^2_{y \cdot x_1 \dots x_k} < r^2_{yx_1} + r^2_{yx_2} + \dots + r^2_{yx_k}$$

许多统计计算机程序输出**序列平方和**。这些平方和,是当各个自变量按照你指定给程序的变量次序依次进入回归模型时,SS(回归)的增量,即新进入的自变量对回归平方和的贡献。序列平方和非常依赖于变量进入模型的具体顺序。当然,这个麻烦的制造者也是共线性。例如,如果回归研究中的所有变量是高度正相关的(就像在经济数据中那样),则无论第一个进入模型的自变量是哪一个,通常它都能解释大部分 $y$ 的变异,而其余的自变量的贡献则很少。此时,对于任何 $x$ ,当给定所有其他的自变量时,其解释能力(有时称为那个 $x$ 的单独预测价值)是小的。当数据表现出严重的共线性时,确实很难分离出各个自变量的预测价值。

### 例 12.11

下面是例 12.6 中数据的输出结果,解释其中的序列平方和。如果用 $x_2$ 和 $x_1$ 依次作为预测因子,我们会得到相同的序列平方和吗?

```
MTB>Correlation 'y' 'x1' 'x2'.
```

|    | y     | x1    |
|----|-------|-------|
| x1 | 0.922 |       |
| x2 | 0.156 | 0.000 |

```
MTB>Regress 'y' 2 'x1' 'x2'.
```

The regression equation is

$$Y = 39.0 + 0.983 x_1 + 0.333 x_2$$

| Predictor | Coef   | Stdev  | t-ratio | p     |
|-----------|--------|--------|---------|-------|
| Constant  | 39.000 | 1.256  | 31.05   | 0.000 |
| x1        | 0.9833 | 0.1538 | 6.39    | 0.001 |
| x2        | 0.3333 | 0.3076 | 1.08    | 0.320 |

s = 3.768      R-sq = 87.5 %      R-sq(adj) = 83.3 %

#### Analysis of Variance

| SOURCE     | DF | SS     | MS     | F     | p     |
|------------|----|--------|--------|-------|-------|
| Regression | 2  | 596.83 | 298.42 | 21.02 | 0.002 |
| Error      | 6  | 85.17  | 14.19  |       |       |
| Total      | 8  | 682.00 |        |       |       |

| SOURCE | DF | SEQ SS |
|--------|----|--------|
| x1     | 1  | 580.17 |
| x2     | 1  | 16.67  |

**解答** 结果中标有 SEQ SS 的一列显示, 一个  $x_1$  就解释了  $y$  的总变差中的 580.17, 在  $x_1$  后增加  $x_2$  则解释了  $y$  的总变差中的另外的 16.67。在这个例子中, 两个自变量是完全不相关的, 这是一个少见的情况。在这种情况下, 把预测因子加入模型的次序无关紧要。

MTB>Regress 'y' 2 'x2' 'x1'.

The regression equation is

$$Y = 39.0 + 0.333 x_2 + 0.983 x_1$$

| Predictor | Coef   | Stdev  | t-ratio | p     |
|-----------|--------|--------|---------|-------|
| Constant  | 39.000 | 1.256  | 31.05   | 0.000 |
| x2        | 0.3333 | 0.3076 | 1.08    | 0.320 |
| x1        | 0.9833 | 0.1538 | 6.39    | 0.001 |

s = 3.768      R-sq = 87.5 %      R-sq(adj) = 83.3 %

#### Analysis of Variance



| SOURCE     | DF | SS     | MS     | F     | P     |
|------------|----|--------|--------|-------|-------|
| Regression | 2  | 596.83 | 298.42 | 21.02 | 0.002 |
| Error      | 6  | 85.17  | 14.19  |       |       |
| Total      | 8  | 682.00 |        |       |       |

| SOURCE | DF | SEQ SS |
|--------|----|--------|
| x2     | 1  | 16.67  |
| x1     | 1  | 580.17 |

12.4 节的思想在于回归系数和标准差  $s_e$  的(最佳猜测)点估计。由于这些估计基于 样本数据,它们会在某种程度上有误差,管理人员在解释模型时也应该考虑到这种误差。我们现在给出多元回归模型中偏斜率的检验。

首先,我们考虑多元回归模型中诸偏斜率( $\beta_1, \beta_2, \dots, \beta_k$ )的一个总的零假设(原假设)的检验。该零假设即  $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 。由这个假设,不包含任何自变量的回归模型具有完全的预测价值。它告诉我们,所有那些精心选择的预测因子是绝对无用的。研究假设是很一般化的,即  $H_a$ :至少一个  $\beta_i \neq 0$ 。该假设只不过是说,在所有预测因子中,有某一个有一些预测价值。

检验统计量是第十一章中的  $F$  统计量。为了给出这个检验,我们首先定义  $y$  关于  $x_1, x_2, \dots, x_k$  的回归平方和。我们再次记该平方和为  $SS(\text{回归})$ ,也称为  $SS(\text{模型})$ 或解释了的平方和,它是  $y$  的预测值与平均值之间差的平方和。

### 定义 12.2

$$\begin{aligned} SS(\text{回归}) &= \sum (\hat{y}_i - \bar{y})^2 \\ SS(\text{总和}) &= \sum (y_i - \bar{y})^2 \\ &= SS(\text{回归}) + SS(\text{残差}) \end{aligned}$$

不像  $SS(\text{总和})$ 和  $SS(\text{残差})$ ,我们不把  $SS(\text{回归})$ 解释成预测误差项,而是把它解释成为预测值  $\hat{y}_i$  随  $x$  的变化而变化的程度。如果  $SS(\text{回归}) = 0$ ,则预测的  $y(\hat{y})$ 值都相同。此时,关于  $x$  的信息对于  $y$  的预测是无用的。如果  $SS(\text{回归})$ 相对于  $SS(\text{残差})$ 比较大,这就表明自变量  $x_1, x_2, \dots, x_k$  有实际的预测价值。我们用均方而不是用平方和来表示检验统计量。同往常一样,均方是平方和除以其自由度。

### $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ 的 $F$ 检验

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_a: \text{至少有一个 } \beta \neq 0$$

$$\text{T.S.: } F = \frac{SS(\text{回归})/k}{SS(\text{残差})/[n - (k + 1)]} = \frac{MS(\text{回归})}{MS(\text{残差})}$$

$$\text{R.R.: } df_1 = k, df_2 = n - (k + 1), \text{当 } F > F_\alpha \text{ 时,拒绝 } H_0.$$

检查假定,得出结论。

**例 12.12**

- a. 在下面重新列出的例 12.11 的计算机输出结果中找出 SS(回归)。  
 b. 找出  $F$  统计量。  
 c. 我们是否可以放心地说自变量  $x_1$  和  $x_2$  一起, 至少有一定的预测能力?

MTB>regress c1 on 2 vars c2 c3

The regression equation is

$$y = 39.0 + 0.983 x_1 + 0.333 x_2$$

| Predictor | Coef   | Stdev  | t-ratio | p     |
|-----------|--------|--------|---------|-------|
| Constant  | 39.000 | 1.256  | 31.05   | 0.000 |
| x1        | 0.9833 | 0.1538 | 6.39    | 0.001 |
| x2        | 0.3333 | 0.3076 | 1.08    | 0.320 |

s = 3.768      R-sq = 87.5 %      R-sq(adj) = 83.3 %

Analysis of Variance

| SOURCE     | DF | SS     | MS     | F     | p     |
|------------|----|--------|--------|-------|-------|
| Regression | 2  | 596.83 | 298.42 | 21.02 | 0.002 |
| Error      | 6  | 85.17  | 14.19  |       |       |
| Total      | 8  | 682.00 |        |       |       |

**解答**

- a. 输出结果的方差分析表明, SS(回归)为 596.83。  
 b. 那里也给出了 MS(回归)和 MS(残差)的值。MS(残差)被标记为 MS(Error)(MS(误差)), 这是常见到的另一个名字。

$$F = \frac{\text{MS(回归)}}{\text{MS(残差)}} = \frac{298.42}{14.19} = 21.02$$

- c. 对于  $df_1 = 2$ ,  $df_2 = 6$ ,  $\alpha = 0.01$ , 表中的  $F$  值为 10.92。因此, 我们有很强的依据( $p$  值远小于 0.01, 为 0.002)拒绝零假设, 认为这些  $x$  合在一起至少有一定的预测价值。

这个  $F$  检验也可以用  $R^2$  来表示。回忆  $R^2_{y \cdot x_1 x_2 \cdots x_k}$  衡量了由于所有  $x$  预测因子的知识而带来的  $y$  的误差平方和的减少。因为  $y$  关于  $x$  的回归解释了  $y$  的总误差平方和中的  $100R^2_{y \cdot x_1 x_2 \cdots x_k} \%$ ,

$$SS(\text{回归}) = R^2_{y \cdot x_1 x_2 \cdots x_k} SS(\text{总和})$$

剩下的  $1 - R^2$  部分与残差平方和有关:

$$SS(\text{残差}) = (1 - R^2_{y \cdot x_1 x_2 \cdots x_k}) SS(\text{总和})$$

$F$  检验统计量可以重新写为

$$F = \frac{MS(\text{回归})}{MS(\text{残差})} = \frac{R^2_{y \cdot x_1 x_2 \cdots x_k} / k}{(1 - R^2_{y \cdot x_1 x_2 \cdots x_k}) / [n - (k + 1)]}$$

该统计量要与自由度为  $df_1 = k$  和  $df_2 = n - (k + 1)$  的  $F$  分布的表值相比较。

### 例 12.13

一家大城市银行研究其各支行中平均每笔账目的大小(ACCTSIZE)与支行所在邮政编码地区的人均收入(INCOME), 商业账号的数量(BUSIN)以及与之竞争的银行支行数量(COMPET)之间的关系。数据用 Staistix 进行了分析, 结果如下:

#### CORRELATIONS (PEARSON)

|        | ACCTSIZE | BUSIN   | COMPET |
|--------|----------|---------|--------|
| BUSIN  | -0.6934  |         |        |
| COMPET | 0.8196   | -0.6527 |        |
| INCOME | 0.4526   | 0.1492  | 0.5571 |

#### UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF ACCTSIZE

##### PREDICTOR

| VARIABLES | COEFFICIENT | STD ERROR | STUDENT'S T | P      | VIF |
|-----------|-------------|-----------|-------------|--------|-----|
| CONSTANT  | 0.15085     | 0.73776   | 0.20        | 0.8404 |     |
| BUSIN     | -0.00288    | 8.894E-04 | -3.24       | 0.0048 | 5.2 |
| COMPET    | -0.00759    | 0.05810   | -0.13       | 0.8975 | 7.4 |
| INCOME    | 0.26528     | 0.10127   | 2.62        | 0.0179 | 4.3 |

|                    |        |                          |         |
|--------------------|--------|--------------------------|---------|
| R-SQUARED          | 0.7973 | RESID. MEAN SQUARE (MSE) | 0.03968 |
| ADJUSTED R-SQUARED | 0.7615 | STANDARD DEVIATION       | 0.19920 |

| SOURCE     | DF | SS      | MS      | F     | P      |
|------------|----|---------|---------|-------|--------|
| REGRESSION | 3  | 2.65376 | 0.88458 | 22.29 | 0.0000 |

|          |    |         |         |
|----------|----|---------|---------|
| RESIDUAL | 17 | 0.67461 | 0.03968 |
| TOTAL    | 20 | 3.32838 |         |

- a. 给出多元回归预测方程。  
b. 用  $R^2$  的值检验假设  $H_0: \beta_1 = \beta_2 = \beta_3 = 0$  (注意:  $n = 21$ )

**解答**

- a. 从输出结果知, 多元回归预测方程为

$$\hat{y} = 0.15085 - 0.00288x_1 - 0.00759x_2 + 0.26528x_3$$

- b. 基于  $R^2$  的检验过程为

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_a: \text{至少有一个 } \beta_j \text{ 不为 } 0$$

$$\text{T.S.: } F = \frac{R^2_{y \cdot x_1 x_2 x_3} / 3}{(1 - R^2_{y \cdot x_1 x_2 x_3}) / (21 - 4)} = \frac{0.7973 / 3}{0.2027 / 17} = 22.29$$

R.R.: 对于  $df_1 = 3$  和  $df_2 = 17$ ,  $F$  分布的 0.05 临界值为 3.20。

由于  $F$  统计量的计算值 22.29 大于 3.20, 我们拒绝  $H_0$ , 并认为有一个或更多的  $x$  有预测能力。这一点从  $p$  值也可以得到:  $p$  值为 0.0000, 比 0.05 小得多。注意, 我们算得的  $F$  值与输出结果中给出的相同。

在这个  $F$  检验中拒绝零假设并不等于完事大吉。拒绝零假设只不过是说, 确有依据表明在自变量中某些变量具有一定的预测价值, 但这既不表明自变量与因变量的关系有多强, 也不表明哪些自变量是有用的。因此, 接下来我们的任务就是作出关于每个偏斜率的推断。

为了作出这些推断, 我们需要估计每个偏斜率的标准误差。与往常一样, 对于任何基于样本的估计, 其标准误差都表示这个估计有多么准确。大多数回归的计算机程序都计算并显示参数估计的标准误差。偏斜率的标准误差依赖于三个方面: 残差标准差, 预测因子变差的大小, 以及对应的自变量与其他自变量的相关程度。我们在下面给出的标准误差的表达式对于考察共线性是有益的, 但不太适合于用作计算。就把计算交给计算机做吧!

**定义 12.3 多元回归中  $\beta_j$  的标准误差的估计:**

$$s_{\beta_j} = s_e \sqrt{\frac{1}{\sum (x_{ij} - \bar{x}_j)^2 (1 - R^2_{x_j \cdot x_1 \dots x_{j-1} x_{j+1} \dots x_k})}}$$

其中  $R^2_{x_j \cdot x_1 \dots x_{j-1} x_{j+1} \dots x_k}$  是以  $x_j$  为因变量, 所有其他  $x$  变量为自变量的多元回归的  $R^2$  的值。注意,  $s_e$  为  $y$  关于  $x_1, x_2, \dots, x_k$  的多元回归中的残差标准差。

与简单回归中一样,残差标准差越大,系数估计的不确定性也越大。此外,预测因子的变异性越小,系数估计的标准误差越大。上述标准误差的最重要的应用在于用来说明 **共线性效应**。如果一个自变量  $x_j$  与另一个或几个自变量高度共线性,由定义  $R^2_{x_j \cdot x_1 \cdots x_{j-1} x_{j+1} \cdots x_k}$  就会很大,从而  $1 - R^2_{x_j \cdot x_1 \cdots x_{j-1} x_{j+1} \cdots x_k}$  接近于 0。除以一个接近于 0 的数使标准误差变得很大。所以,严重的共线性的一个重要的效应是导致偏斜率的估计有很大的标准误差,即这些估计不很精确。

$1 - R^2_{x_j \cdot x_1 \cdots x_{j-1} x_{j+1} \cdots x_k}$  称为**方差膨胀因子(VIF)**。它衡量了一个系数的方差(标准误差的平方)由于共线性而增加的程度。某些计算机软件包输出这个因子的值,由此我们可以看到其中的共线性问题的严重程度。如果 VIF 是 1,说明根本就没有共线性。如果该值很大,比如 10 或更大,那么共线性就是一个严重的问题。

对于任何偏斜率的估计,如果它有大的标准误差,表示其可能有大的误差。 $x_j$  的偏斜率  $\hat{\beta}_j$  估计了当其他  $x$  都保持不变,而  $x_j$  增加一个单位时的效应。如果  $x_j$  与其他  $x$  有很强共线性,那么当  $x_j$  增加时,其他的  $x$  也会变化而不能保持常量,因此,当  $x_j$  与其他自变量严重共线时,估计  $\hat{\beta}_j$  就很困难,就可能会有大的误差。

每个偏斜率估计  $\hat{\beta}_j$  的标准误差可以用于求  $\beta_j$  的置信区间和对其作假设检验。置信区间有熟悉的形式,即“估计加上或减去(表值)·(标准误差的估计)”,其中的表值是从自由度为  $n - (k + 1)$  的  $t$  分布表中查得的。

**定义 12.4**  $\beta_j$  的置信区间为

$$\hat{\beta}_j - t_{\alpha/2} s_{\hat{\beta}_j} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2} s_{\hat{\beta}_j}$$

其中  $t_{\alpha/2}$  在自由度为  $n - (k + 1)$  的  $t$  分布的右侧尾部切下的面积为  $\alpha/2$ 。

### 例 12.14

对于例 12.7 中的数据,计算两因子模型中的  $\hat{\beta}_1$  的 95% 置信区间。有关的输出结果如下:

```
.regress y x1 x2
```

| y     | Coef. | Std. Err. | t     | P> t  | [95 % Conf. Interval] |          |
|-------|-------|-----------|-------|-------|-----------------------|----------|
| x1    | 1     | 1.870829  | 0.535 | 0.646 | -7.049526             | 9.049526 |
| x2    | 3     | 4.1833    | 0.717 | 0.548 | -14.99929             | 20.99929 |
| _cons | 10    | 1.183216  | 8.452 | 0.014 | 4.909033              | 15.09097 |

**解答** 由上述结果  $\hat{\beta}_1$  为 1.00, 其标准误差为 1.870829。在自由度为  $df = n$

$-(k+1)=5-(2+1)=2$  的  $t$  分布的尾部切下面积 0.025 的  $t$  值为 4.303。置信区间为  $1.00 - 4.303(1.870829) \leq \beta_1 \leq 1.00 + 4.303(1.870829)$ , 或  $-7.050 \leq \beta_1 \leq 9.050$ 。输出结果中的置信区间保留了更多的小数位。

### 例 12.15

在例 12.12 的输出结果中找出  $x_2$  的偏斜率的估计以及它的标准误差。计算  $\hat{\beta}_2$  的 90% 的置信区间。

MTB>Regress 'y' on 2 vars 'x1' 'x2'

The regression equation is

$Y = 39.0 + 0.983 x_1 + 0.333 x_2$

| Predictor | Coef   | Stdev  | t-ratio | p     |
|-----------|--------|--------|---------|-------|
| Constant  | 39.000 | 1.256  | 31.05   | 0.000 |
| x1        | 0.9833 | 0.1538 | 6.39    | 0.001 |
| x2        | 0.3333 | 0.3076 | 1.08    | 0.320 |

**解答**  $\hat{\beta}_2$  为 0.3333, 标准误差(标有 Stdev)为 0.3076; 查表得  $t$  值为 1.943 [尾部面积为 0.05, 自由度为  $9 - (2 + 1) = 6df$ ]。所求的置信区间为  $0.3333 - 1.943(0.3076) \leq \beta_2 \leq 0.3333 + 1.943(0.3076)$  或  $-0.2644 \leq \beta_2 \leq 0.9310$ 。

### $H_0: \beta_j = 0$ 的解释

关于  $\beta_j$  的通常的零假设为  $H_0: \beta_j = 0$ 。这个假设并不是说  $x_j$  本身对预测没有价值, 而是说该自变量在其他自变量的基础上对预测没有附加的价值, 即如果其他变量已经用在回归模型中, 最后加入  $x_j$  不会改进预测。 $H_0: \beta_j = 0$  的检验衡量  $x_j$  是否对于预测有附加的价值。 $H_0$  的  $t$  检验总结如下。

### 关于 $\beta_j$ 的检验的小结

$H_0: 1. \beta_j \leq 0$

2.  $\beta_j \geq 0$

3.  $\beta_j = 0$

$H_a: 1. \beta_j > 0$

2.  $\beta_j < 0$

3.  $\beta_j \neq 0$

T.S.:  $t = \hat{\beta}_j / s_{\hat{\beta}_j}$

R.R.: 1.  $t > t_\alpha$

$$2. t < t_{\alpha}$$

$$3. |t| > t_{\alpha/2}$$

其中  $t_{\alpha}$  在自由度为  $df = n - (k + 1)$  的  $t$  分布右侧尾部切下的面积为  $\alpha$ 。  
检查假定并得出结论。

所有多元回归的程序都输出这个检验统计量的值。

### 例 12.16

- a. 用例 12.14 中的信息检验  $H_0: \beta_1 = 0$ , 取  $\alpha = 0.05$ , 备择假设为双边假设。
- b. 这个检验的结果与前面置信区间的结果是一致的吗?

### 解答

- a. 假设  $H_0: \beta = 0$   $H_a: \beta$  的检验统计量为  $\hat{t} = \hat{\beta}_1 / s_{\hat{\beta}_1} = 1.00 / 1.871 = 0.535$ 。

由于自由度为  $5 - (2 + 1) = 2$  的  $t$  分布的 0.025 分位数为 4.303, 我们不能拒绝  $H_0$ , 也就是说,  $x_1$  没有表现出在另一个自变量  $x_2$  存在时有附加的预测能力。

- b. 95% 的置信区间包含 0, 这也意味着对于该双边检验问题, 在  $\alpha = 0.05$  的水平上不能拒绝  $H_0: \beta = 0$ 。

### 例 12.17

在例 12.15 的输出结果中, 找出检验  $H_0: \beta_2 = 0$  的  $t$  统计量。在常见的水平上, 能支持  $H_a: \beta_2 > 0$  吗?

**解答**  $t$  统计量的上方标有“t-ratio”。对于  $x_2$ ,  $t$  统计量为 1.08。与自由度 6 和  $\alpha = 0.10$  对应的  $t$  表值为 1.440, 所以在  $\alpha = 0.10$  的水平上不能拒绝  $H_0$ 。从另一方面来看,  $p$  值为 0.320, 比  $\alpha = 0.10$  大, 同样也不能拒绝  $H_0$ 。

本章我们所讨论的  $F$  检验和  $t$  检验是针对不同的零假设的。有时会有这样的情况:  $F$  检验导致了拒绝  $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ , 但没有一个对于  $H_0: \beta_j = 0$  的  $t$  检验是显著的。在这种情况下, 我们可以说, 回归方程作为一个整体具有预测能力, 但我们找不出哪个变量具有预测能力。记住每一个  $t$  检验是检验单独的预测价值的。给定所有其他变量, 对于预测来说这个变量附加了价值吗? 当两个或多个预测因子高度相关的时候, 经常发生这样的事, 即虽然所有  $x$  一起被证实是有用的, 但没有  $x_j$  有显著的、单独的预测能力。如果我们试图用国民生产总值和国家储蓄收入来预测房产的销售, 我们就不大可能在给定 DI 时找到 GDP 的附加价值, 或者在给定 GDP 时找到 DI 的附加价值。

## 练习

**12.14** 参见练习 12.9 的计算机输出结果。重新列出如下。

.regress Sale Promo Devel Research

|          |            |    |            |               |   |        |
|----------|------------|----|------------|---------------|---|--------|
| SOURCE   | SS         | df | MS         | Number of obs | = | 24     |
| MODEL    | 43901.7677 | 3  | 14633.9226 | F(3, 20)      | = | 22.28  |
| Residual | 13136.2323 | 20 | 656.811614 | Prob > F      | = | 0.0000 |
| Total    | 57038.00   | 23 | 2479.91304 | R-square      | = | 0.7697 |
|          |            |    |            | Adj R-square  | = | 0.7351 |
|          |            |    |            | Root MSE      | = | 25.628 |

|          |           |           |        |       |                      |          |
|----------|-----------|-----------|--------|-------|----------------------|----------|
| SALES    | Coef.     | Std. Err. | t      | P> T  | [95% Conf. Interval] |          |
| Promo    | 136.0983  | 28.10759  | 4.842  | 0.000 | 77.46689             | 194.7297 |
| Devel    | -61.17526 | 50.94102  | -1.201 | 0.244 | -167.4364            | 45.08585 |
| Research | -43.69508 | 48.32298  | -0.904 | 0.377 | -144.495             | 57.10489 |
| _cons    | 326.3893  | 241.6129  | 1.351  | 0.192 | -177.6063            | 830.3849 |

- 找到  $F$  统计量。
- 没有总预测价值的假设在水平  $\alpha = 0.05$  上能被拒绝吗？
- 找出对于促销(PROMO)的系数  $\beta_1$  的  $t$  统计量。
- 检验研究假设  $\beta_1 \neq 0$ , 取  $\alpha = 0.05$ 。
- 给出(d)中检验的结论。

**12.15** 找出前面练习(d)中检验的  $p$  值, 检验是单侧的还是双侧的？

**12.16** 总结练习 12.14 中的  $t$  检验。你检验的零假设是什么？

**12.17** 下面的数据是人工构造出来的, 用来说明相关和不相关的自变量的效应。

|    |    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|----|
| y: | 17 | 21 | 26 | 22 | 27 | 25 | 28 | 34 | 29 | 37 | 38 | 38 |
| x: | 1  | 1  | 1  | 1  | 2  | 2  | 2  | 2  | 3  | 3  | 3  | 3  |
| w: | 1  | 2  | 3  | 4  | 1  | 2  | 3  | 4  | 1  | 2  | 3  | 4  |
| v: | 1  | 1  | 2  | 2  | 3  | 3  | 4  | 4  | 5  | 5  | 6  | 6  |

这些数据的输出结果如下:

MTB>Regress 'Y' 3 'X' 'W' 'V'.

The regression equation is

$$Y = 10.0 + 5.00X + 2.00W + 1.00V$$



| Predictor | Coef   | Stdev | t-ratio | p     |
|-----------|--------|-------|---------|-------|
| Constant  | 10.000 | 5.766 | 1.73    | 0.121 |
| X         | 5.000  | 6.895 | 0.73    | 0.489 |
| W         | 2.000  | 1.528 | 1.31    | 0.227 |
| V         | 1.000  | 3.416 | 0.29    | 0.777 |

s = 2.646      R-sq = 89.5 %      R-sq(adj) = 85.6 %

Analysis of Variance

| SOURCE     | DF | SS     | MS     | F     | p     |
|------------|----|--------|--------|-------|-------|
| Regression | 3  | 479.00 | 159.67 | 22.81 | 0.000 |
| Error      | 8  | 56.00  | 7.00   |       |       |
| Total      | 11 | 535.00 |        |       |       |

| SOURCE | DF | SEQ SS |
|--------|----|--------|
| X      | 1  | 392.00 |
| W      | 1  | 86.40  |
| V      | 1  | 0.60   |

- 找出 MS(回归)和 MS(残差)。
- F 检验统计量的值是什么?
- 确定该 F 检验的 p 值。
- 从 F 检验可以得到什么结论?
- 计算  $x$  的真实系数的 95% 置信区间。

**12.18** 在练习 12.11 中, 我们曾经把每条支线上的收入作为小城市人口和它到空中枢纽的航空距离的函数进行预测。下面重新列出由 JMP 软件计算的结果。是否有证据清楚地表明两个自变量一起对于预测收入至少有一些价值?

| Response: Revenue          |          |  |  |  |
|----------------------------|----------|--|--|--|
| Summary of Fit             |          |  |  |  |
| RSquare                    | 0.805479 |  |  |  |
| RSquare Adj                | 0.785003 |  |  |  |
| Root Mean Square Error     | 27.7539  |  |  |  |
| Mean of Response           | 252.5    |  |  |  |
| Observations (or Sum Wgts) | 22       |  |  |  |

| Parameter Estimates |           |           |         |          |
|---------------------|-----------|-----------|---------|----------|
| Term                | Estimate  | Std Error | t Ratio | Prob>[t] |
| Intercept           | 86.736862 | 27.74907  | 3.13    | 0.0056   |
| Air miles           | 0.2922116 | 0.120336  | 2.43    | 0.0253   |
| Population          | 1.5310653 | 0.174004  | 8.80    | 0.0000   |

| Whole-Model Test     |    |                |             |         |
|----------------------|----|----------------|-------------|---------|
| Analysis of Variance |    |                |             |         |
| Source               | DF | Sum of Squares | Mean Square | F Ratio |
| Model                | 2  | 60602.202      | 30301.1     | 39.3378 |
| Error                | 19 | 14635.298      | 770.3       | Prob>F  |
| C Total              | 21 | 75237.500      |             | 0.0000  |

**12.19** 在航空支线的回归问题中,是否有很强的依据表明每一个自变量在给定另一个自变量时,都有附加的预测价值?

**12.20** 用练习 12.18 中航空支线问题的输出结果,计算两个偏斜率的 90% 的置信区间。有关的自由度在输出结果中标有“Error”的地方给出。

**12.21** 一家金属加工厂用多元回归方法进行了一项能源的研究。因变量是  $y$  = 每天的能源耗费(以千美元为单位),自变量为  $x_1$  = 一天中加工的金属的吨数,  $x_2$  = 平均外部温度 - 60°F(与工会的合同要求当外界温度达到 60°F 时,要开冷气),  $x_3$  = 所使用的机械的瓦特数,  $x_4 = x_1 x_2$ 。用 Statistix 对数据进行了分析,部分输出结果如下( $x_1$  在输出结果中表示为 METAL,  $x_2$  表示为 TEMP,  $x_3$  表示为 WATTS,  $x_4$  表示为 METXTEMP):

#### CORRELATIONS (PEARSON)

|          | ENERGY | METAL   | METXTEMP | TEMP   |
|----------|--------|---------|----------|--------|
| METAL    | 0.6128 |         |          |        |
| METXTEMP | 0.4929 | 0.1094  |          |        |
| TEMP     | 0.4007 | -0.0606 | 0.9831   |        |
| WATTS    | 0.5775 | 0.2239  | 0.3630   | 0.3529 |

#### UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF ENERGY

##### PREDICTOR

| VARIABLES | COEFFICIENT | STD ERROR | STUDENT'S T | P      | VIF   |
|-----------|-------------|-----------|-------------|--------|-------|
| CONSTANT  | 7.20439     | 17.5322   | 0.41        | 0.6855 |       |
| METAL     | 1.36291     | 0.92438   | 1.47        | 0.1559 | 8.8   |
| TEMP      | 0.30588     | 1.62104   | 0.19        | 0.8522 | 250.0 |
| WATTS     | 0.01024     | 0.00473   | 2.16        | 0.0427 | 1.5   |
| METXTEMP  | -0.00277    | 0.07722   | -0.04       | 0.9717 | 246.4 |

|           |        |                          |         |
|-----------|--------|--------------------------|---------|
| R-SQUARED | 0.6636 | RESID. MEAN SQUARE (MSE) | 6.51555 |
|-----------|--------|--------------------------|---------|

ADJUSTED R-SQUARED      0.5963      STANDARD DEVIATION      2.55255

| SOURCE     | DF | SS      | MS      | F    | P      |
|------------|----|---------|---------|------|--------|
| REGRESSION | 4  | 257.048 | 64.2622 | 9.86 | 0.0001 |
| RESIDUAL   | 20 | 130.311 | 6.51555 |      |        |
| TOTAL      | 24 | 387.360 |         |      |        |

CASES INCLUDED 25 MISSING CASES 0

- 写出估计后的模型。
- 总结各个  $t$  检验的结果。
- 对于  $X_4$ , 即 METXTEMP 的系数, 计算其 95% 的置信区间。
- 输出中 VIF 列表明了什么样的共线性问题?

## 12.5 回归系数子集的检验

上一节中, 我们给出了检验一个回归模型中所有回归系数的  $F$  检验以及检验一个系数的  $t$  检验。还有一个  $F$  检验, 用来检验真实系数中的某几个为 0 这样的零假设, 即给定其他的预测因子后, 这几个预测因子没有价值。例如, 如果我们要基于国家的最低工资、国家的通货膨胀率、地区的人口密度和地区中中档房屋租价来预测各个地区当前的工资额度, 那么就可能要检验一下在给定有关国家的两个变量后, 与地区有关的两个变量(密度和房价)是否增加了预测的精度。

此时的零假设为“密度和房价的真实系数为 0”。根据这个零假设, 一旦最低工资和通货膨胀率作为预测因子包括在模型中, 密度和房价这两个自变量一起对与预测就没有价值了。

检验的思想是, 比较当密度和房价包含或不包含在预测方程中时的  $SS$ (回归)或  $R^2$ 。当它们包含在模型中时,  $R^2$  将至少与不包含它们时的值一样大。类似地, 完全模型的  $SS$ (回归)比较大。上述零假设的  $F$  检验是要检验回归平方和的这一增量是否比在零假设下所能预期的大一些。一般地, 以  $k$  记预测因子的总数,  $g$  为其系数没有被假设为 0 的预测因子的个数( $g < k$ ), 则  $k - g$  表示其系数被假设为 0 的预测因子的个数。关键是要求出用所有预测因子(完全模型)和只用零假设中没有出现的  $g$  个预测因子(简略模型)时的  $SS$ (回归)。一旦算得它们的值, 检验可按下面的步骤进行。为了记号简便, 我们假定简略模型中包含  $\beta_1, \beta_2, \dots, \beta_g$ , 从而包含在零假设中的是后面的那些变量。

### 预测因子子集的 $F$ 检验

$$H_0: \beta_{g+1} = \beta_{g+2} = \dots = \beta_k = 0$$

$H_a: H_0$  不真

$$\text{T.S.: } F = \frac{[\text{SS}(\text{回归, 完全}) - \text{SS}(\text{回归, 简略})]/(k - g)}{\text{SS}(\text{回归, 完全})/[n - (k + 1)]}$$

R.R.:  $F > F_\alpha$ , 这里,  $F_\alpha$  在自由度为  $df_1 = (k - g)$  和  $df_2 = [n - (k + 1)]$  的  $F$  分布右侧尾部切下面积  $\alpha$ 。

检查假定, 得到结论。

### 例 12.18

某州渔业委员会为了在湖中重新放养适当数量的鱼苗, 要估计从给定的湖中一个季度内鲈鱼捕捞的数量。委员会可以通过在季度前后采用大规模“拉网式”的方法很准确地得到季度内的捕捞数量, 但这个方法耗费巨大, 日常无法实施。因此, 委员会抽取了一些湖的样本, 并记录下  $y$ : 季度内的捕捞量(千条/每平方英里湖面),  $x_1$ : 每平方英里湖面岸边居民数,  $x_2$ : 按平方英里计算的湖的大小,  $x_3 = 1$  如果该湖对公众开放, 否则取值为 0,  $x_4$ : 一个结构指标(结构有杂草湖床(weed beds), 沉降树(sunken trees), 地面陷落(drop-offs)以及其他鲈鱼栖息之地), 数据如下。

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|-----|-------|-------|-------|-------|
| 3.6 | 92.2  | 0.21  | 0     | 81    |
| 0.8 | 86.7  | 0.30  | 0     | 26    |
| 2.5 | 80.2  | 0.31  | 0     | 52    |
| 2.9 | 87.2  | 0.40  | 0     | 64    |
| 1.4 | 64.9  | 0.44  | 0     | 40    |
| 0.9 | 90.1  | 0.56  | 0     | 22    |
| 3.2 | 60.7  | 0.78  | 0     | 80    |
| 2.7 | 50.9  | 1.21  | 0     | 60    |
| 2.2 | 86.1  | 0.34  | 1     | 30    |
| 5.9 | 90.0  | 0.40  | 1     | 90    |
| 3.3 | 80.4  | 0.52  | 1     | 74    |
| 2.9 | 75.0  | 0.66  | 1     | 50    |
| 3.6 | 70.0  | 0.78  | 1     | 61    |
| 2.4 | 64.6  | 0.91  | 1     | 40    |
| 0.9 | 50.0  | 1.10  | 1     | 22    |
| 2.0 | 50.0  | 1.24  | 1     | 50    |
| 1.9 | 51.2  | 1.47  | 1     | 37    |
| 3.1 | 40.1  | 2.21  | 1     | 61    |
| 2.6 | 45.0  | 2.46  | 1     | 39    |
| 3.4 | 50.0  | 2.80  | 1     | 53    |

委员会确信  $x_1$  和  $x_2$  对于预测  $y$  都是重要变量, 因为这两个变量反应了湖上捕捞的强度大小。这就有了一个问题:  $x_3$  和  $x_4$  作为附加的预测因子是否有用? 因此, 拟合了包含和不包含  $x_3$  和  $x_4$  的回归模型(只含各个变量的线性项)。部分 Minitab 的输出结果如下:

MTB>regress 'catch' on 4 variables 'residenc' 'size' 'access' 'structur'

The regression equation is

catch = -1.94 + 0.0193 residenc + 0.332 size + 0.836 access + 0.0477 structur

| Predictor | Coef     | Stdev    | t-ratio | p     |
|-----------|----------|----------|---------|-------|
| Constant  | -1.9378  | 0.9081   | -2.13   | 0.050 |
| residenc  | 0.01929  | 0.01018  | 1.90    | 0.077 |
| size      | 0.3323   | 0.2458   | 1.35    | 0.196 |
| access    | 0.8355   | 0.2250   | 3.71    | 0.002 |
| structur  | 0.047714 | 0.005056 | 9.44    | 0.000 |

s = 0.4336      R-sq = 88.2 %      R-sq(adj) = 85.0 %

Analysis of Variance

| SOURCE     | DF | SS      | MS     | F     | p     |
|------------|----|---------|--------|-------|-------|
| Regression | 4  | 21.0474 | 5.2619 | 27.98 | 0.000 |
| Error      | 15 | 2.8206  | 0.1880 |       |       |
| Total      | 19 | 23.8680 |        |       |       |

| SOURCE   | DF | SEQ SS  |
|----------|----|---------|
| residenc | 1  | 0.2780  |
| size     | 1  | 1.5667  |
| access   | 1  | 2.4579  |
| structur | 1  | 16.7448 |

MTB>regress 'catch' on 2 vars 'residenc' 'size'

The regression equation is

catch = -0.11 + 0.0310 residenc + 0.679 size

| Predictor | Coef    | Stdev   | t-ratio | p     |
|-----------|---------|---------|---------|-------|
| Constant  | -0.107  | 2.336   | -0.05   | 0.964 |
| residenc  | 0.03102 | 0.02650 | 1.17    | 0.258 |

size            0.6794      0.6178          1.10      0.287

s = 1.138          R-sq = 7.7 %          R-sq(adj) = 0.0 %

#### Analysis of Variance

| SOURCE     | DF | SS     | MS    | F    | p     |
|------------|----|--------|-------|------|-------|
| Regression | 2  | 1.845  | 0.922 | 0.71 | 0.505 |
| Error      | 17 | 22.023 | 1.295 |      |       |
| Total      | 19 | 23.868 |       |      |       |

- 写出完全和简略模型。
- 写出检验“省略的变量没有附加的预测价值”时的零假设。
- 对于上述零假设进行  $F$  检验。

#### 解答

- 完全和简略模型分别为

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$$

和

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

基于样本数据的相应的多元回归预测方程为

$$\text{完全: } \hat{y} = -1.94 + 0.0193x_1 + 0.332x_2 + 0.836x_3 + 0.0477x_4$$

$$\text{简略: } \hat{y} = -0.11 + 0.0310x_1 + 0.679x_2$$

- $x_3$  和  $x_4$  没有附加预测能力的零假设为  $H_0: \beta_3 = \beta_4 = 0$ .
- 对于(b)中  $H_0$  的检验统计量要用到  $SS(\text{回归, 完全}) = 21.0474$ ,  $SS(\text{回归, 简略}) = 1.845$ ,  $SS(\text{残差, 完全}) = 2.8206$ ,  $k = 4$ ,  $g = 2$  以及  $n = 20$ :

$$\begin{aligned} \text{T.S.: } F &= \frac{[SS(\text{回归, 完全}) - SS(\text{回归, 简略})]/(4 - 2)}{SS(\text{回归, 完全})/[20 - 5]} \\ &= \frac{(21.0474 - 1.845)/2}{2.8206/15} = 51.059 \end{aligned}$$

自由度为 2 和 15 的表值  $F_{0.01}$  为 6.36, 检验统计量的值比表值大出很多, 所以我们有充分的依据断言开放和结构两个变量具有附加的预测价值 ( $p < 0.001$ )。

#### 练习

**12.22** 练习 12.9 的输出结果如下。

- 找出  $R^2$  的值,并用它来确认  $F$  统计量的计算结果。
- 你能否断言这些自变量至少有一些预测能力?

```
.regress Sales Promo Devel Research
```

| SOURCE   | SS         | df        | MS         | Number of obs = 24    |                       |          |
|----------|------------|-----------|------------|-----------------------|-----------------------|----------|
|          |            |           |            | F(3,20) = 22.28       |                       |          |
| MODEL    | 43901.7677 | 3         | 14633.9226 | Prob > F = 0.0000     |                       |          |
| Residual | 13136.2323 | 20        | 656.811614 | R-square = 0.7697     |                       |          |
|          |            |           |            | Adj R-square = 0.7351 |                       |          |
| Total    | 57038.00   | 23        | 2479.91304 | Root MSE = 25.628     |                       |          |
| SALES    | Coef.      | Std. Err. | t          | P> t                  | [95 % Conf. Interval] |          |
| Promo    | 136.0983   | 28.10759  | 4.842      | 0.000                 | 77.46689              | 194.7297 |
| Devel    | -61.17526  | 50.94102  | -1.201     | 0.244                 | -167.4364             | 45.08585 |
| Research | -43.69508  | 48.32298  | -0.904     | 0.377                 | -144.495              | 57.10489 |
| _cons    | 326.3893   | 241.6129  | 1.351      | 0.192                 | -177.6063             | 830.3849 |

**12.23** 练习 12.22 中数据的另一个回归分析只用了开发费用作为自变量。输出如下:

```
.regress Sales Promo
```

| SOURCE   | SS         | df        | MS         | Number of obs = 24    |                       |          |
|----------|------------|-----------|------------|-----------------------|-----------------------|----------|
|          |            |           |            | F(1,22) = 50.80       |                       |          |
| MODEL    | 39800.7248 | 1         | 39800.7248 | Prob > F = 0.0000     |                       |          |
| Residual | 17237.2752 | 22        | 783.512509 | R-square = 0.6978     |                       |          |
|          |            |           |            | Adj R-square = 0.6841 |                       |          |
| Total    | 57038.00   | 23        | 2479.91304 | Root MSE = 27.991     |                       |          |
| SALES    | Coef.      | Std. Err. | t          | P> t                  | [95 % Conf. Interval] |          |
| Promo    | 78.24931   | 10.97888  | 7.127      | 0.000                 | 55.48051              | 101.0181 |
| _cons    | -0.6490769 | 44.58506  | -0.015     | 0.989                 | -93.11283             | 91.81468 |

- 找出该简略模型的  $R^2$  的值。
- 用  $\alpha = 0.01$  实施  $F$  检验的各个步骤。
- 我们是否能得到这样的结论,即省去的自变量对于预测至少有一些价值?

**12.24** 基于例 12.13 中的数据建立了两个模型,结果如下,其中的变量

BUSIN, COMPET 和 INCOME 分别表示商业账号数量, 竞争银行数以及人均收入。

## CORRELATIONS (PEARSON)

|        | ACCTSIZE BUSIN COMPET |          |        |
|--------|-----------------------|----------|--------|
| BUSIN  | - 0.6934              |          |        |
| COMPET | 0.8196                | - 0.6527 |        |
| INCOME | 0.4526                | 0.1492   | 0.5571 |

CASES INCLUDED 21 MISSING CASES 0

(Model 1)

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF ACCTSIZE

## PREDICTOR

| VARIABLES | COEFFICIENT | STD ERROR | STUDENT'S T | P      | VIF |
|-----------|-------------|-----------|-------------|--------|-----|
| CONSTANT  | 0.15085     | 0.73776   | 0.20        | 0.8404 |     |
| BUSIN     | - 0.00288   | 8.894E-04 | - 3.24      | 0.0048 | 5.2 |
| COMPET    | - 0.00759   | 0.05810   | - 0.13      | 0.8975 | 7.4 |
| INCOME    | 0.26528     | 0.10127   | 2.62        | 0.0179 | 4.3 |

|                      |        |                          |         |
|----------------------|--------|--------------------------|---------|
| R-SQUARED            | 0.7973 | RESID. MEAN SQUARE (MSE) | 0.03968 |
| ADJUSTED R - SQUARED | 0.7615 | STANDARD DEVIATION       | 0.19920 |

| SOURCE     | DF | SS      | MS      | F     | P      |
|------------|----|---------|---------|-------|--------|
| REGRESSION | 3  | 2.65376 | 0.88458 | 22.29 | 0.0000 |
| RESIDUAL   | 17 | 0.67461 | 0.03968 |       |        |
| TOTAL      | 20 | 3.32838 |         |       |        |

(Model 2)

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF ACCTSIZE

## PREDICTOR

| VARIABLES | COEFFICIENT | STD ERROR | STUDENT'S T | P      |
|-----------|-------------|-----------|-------------|--------|
| CONSTANT  | 0.12407     | 0.96768   | 0.13        | 0.8993 |
| INCOME    | 0.20191     | 0.09125   | 2.21        | 0.0394 |

|             |        |                          |         |
|-------------|--------|--------------------------|---------|
| R - SQUARED | 0.2049 | RESID. MEAN SQUARE (MSE) | 0.13928 |
|-------------|--------|--------------------------|---------|



ADJUSTED R - SQUARED      0.1630      STANDARD DEVIATION      0.37321

| SOURCE     | DF | SS      | MS      | F    | P      |
|------------|----|---------|---------|------|--------|
| REGRESSION | 1  | 0.68192 | 0.68192 | 4.90 | 0.0394 |
| RESIDUAL   | 19 | 2.64645 | 0.13928 |      |        |
| TOTAL      | 20 | 3.32838 |         |      |        |

CASES INCLUDED 21 MISSING CASES 0

- 找出只用 INCOME(收入)作为预测因子的简略模型的  $R^2$  值。
- 找出完全模型的  $R^2$  值。
- 比较(a)(b)中的两个值。只用变量 INCOME 能给出一个充分的拟合吗?

**12.25** 基于输出中的平方和计算前一练习中的  $F$  统计量,并解释  $F$  检验的结果。

**12.26 (社会)** 一家汽车筹资公司使用一个相当复杂的汽车贷款信用积分系统。填写所需的问卷很费时间,这不仅花费销售工作人员的时间,也冒着疏远顾客的风险。公司决定看一看是否可以用三个变量(年龄、月家庭收入以及偿还债务所占收入的比例)能够对信用积分进行合理地近似。收集了由 500 份申请组成的一个样本的数据(没有明显的偏差),计算了各份申请的复杂系统下的积分作为多元回归中的因变量。利用 JMP 进行计算,部分结果如下。

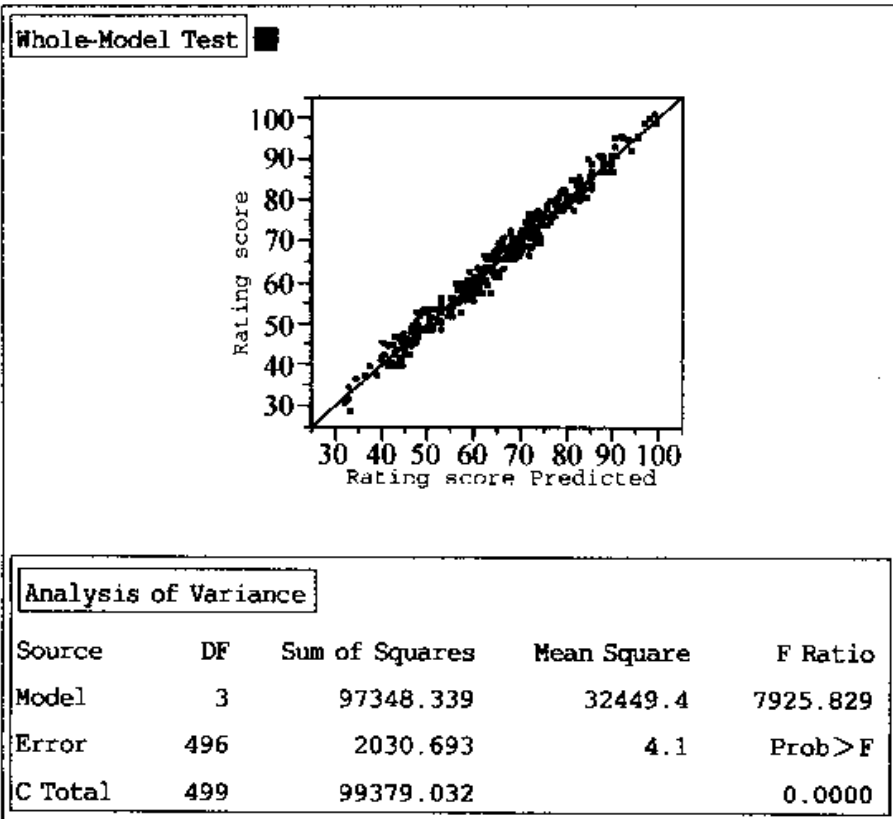
- 这三个预测因子解释了积分变差中的多大部分?
- 用这个数验证总体  $F$  统计量的计算。
- 这个  $F$  检验能否清楚地表明这三个预测因子对于积分的预测是有价值的?

Response: Rating score

| Summary of Fit             |          |
|----------------------------|----------|
| RSquare                    | 0.979566 |
| RSquare Adj                | 0.979443 |
| Root Mean Square Error     | 2.023398 |
| Mean of Response           | 65.044   |
| Observations (or Sum Wgts) | 500      |

| Parameter Estimates |           |           |         |         |
|---------------------|-----------|-----------|---------|---------|
| Term                | Estimate  | Std Error | t Ratio | Prob> t |
| Intercept           | 54.657197 | 0.634791  | 86.10   | 0.0000  |
| Age                 | 0.0056098 | 0.011586  | 0.48    | 0.6285  |
| Monthly income      | 0.0100597 | 0.000157  | 64.13   | 0.0000  |
| Debt fraction       | -39.95239 | 0.883684  | -45.21  | 0.0000  |

| Effect Test    |       |    |                |          |        |
|----------------|-------|----|----------------|----------|--------|
| Source         | Nparm | DF | Sum of Squares | F Ratio  | Prob>F |
| Age            | 1     | 1  | 0.960          | 0.2344   | 0.6285 |
| Monthly income | 1     | 1  | 16835.195      | 4112.023 | 0.0000 |
| Debt fraction  | 1     | 1  | 8368.627       | 2044.05  | 0.0000 |



12.27 只用月收入变量对信用积分数据重新进行了分析。JMP 的输出结果如下所示。

Response: Rating score

| Summary of Fit             |          |
|----------------------------|----------|
| RSquare                    | 0.895261 |
| RSquare Adj                | 0.895051 |
| Root Mean Square Error     | 4.571792 |
| Mean of Response           | 65.044   |
| Observations (or Sum Wgts) | 500      |

Lack of Fit

Parameter Estimates

| Term           | Estimate  | Std Error | t Ratio | Prob> t |
|----------------|-----------|-----------|---------|---------|
| Intercept      | 30.152827 | 0.572537  | 52.67   | 0.0000  |
| Monthly income | 0.0135544 | 0.000208  | 65.24   | 0.0000  |

- 在去掉年龄和债务比例这两个预测因子后,回归平方和减少了多少?
- 一旦给定收入变量,年龄和债务比例这两个变量增加的预测价值在通常的水平上是显著的吗?

**12.28(工程)** 一家化学公司试验在两种催化剂的不同用量下的产量。在催化剂 1 的五个水平和催化剂 2 的四个水平的搭配上测量了产量。拟合一个二阶模型来近似描述预想的非线性关系,其中各变量的定义为: $y$  = 产量,  $x_1$  = 催化剂 1 的用量,  $x_2$  = 催化剂 2 的用量,  $x_3 = x_1^2$ ,  $x_4 = x_1x_2$ ,  $x_5 = x_2^2$ 。回归分析的部分结果如下(结果中  $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_4$ 、 $x_5$  分别表示为 Cat1、Cat2、@Cat1Sq、@Cat1Cat2、@Cat2sq)。

Multiple Regression Analysis

Dependent variable: Yield

Table of Estimates

|           | Estimate | Standard Error | t Value | P Value |
|-----------|----------|----------------|---------|---------|
| Constant  | 50.0195  | 4.3905         | 11.39   | 0.0000  |
| Cat1      | 6.64357  | 2.01212        | 3.30    | 0.0052  |
| Cat2      | 7.3145   | 2.73977        | 2.67    | 0.0183  |
| @Cat1Sq   | -1.23143 | 0.301968       | -4.08   | 0.0011  |
| @Cat1Cat2 | -0.7724  | 0.319573       | -2.42   | 0.0299  |
| @Cat2Sq   | -1.1755  | 0.50529        | -2.33   | 0.0355  |

R-squared = 86.24 %

Adjusted R-squared = 81.33 %

Standard error of estimation = 2.25973

Analysis of Variance

| Source | Sum of Squares | D.F. | Mean Square | F-Ratio | P Value |
|--------|----------------|------|-------------|---------|---------|
| Model  | 448.193        | 5    | 89.6386     | 17.55   | 0.0000  |
| Error  | 71.489         | 14   | 5.10636     |         |         |

Total(corr.)    519.682    19

#### Conditional Sums of Squares

| Source    | Sum of<br>Squares | D.F. | Mean Square | F - Ratio | P<br>Value |
|-----------|-------------------|------|-------------|-----------|------------|
| Cat1      | 286.439           | 1    | 286.439     | 56.09     | 0.0000     |
| Cat2      | 19.3688           | 1    | 19.3688     | 3.79      | 0.0718     |
| @Cat1Sq   | 84.9193           | 1    | 84.9193     | 16.63     | 0.0011     |
| @Cat1Cat2 | 29.8301           | 1    | 29.8301     | 5.84      | 0.0299     |
| @Cat2Sq   | 27.636            | 1    | 27.636      | 5.41      | 0.0355     |
| Model     | 448.193           | 5    |             |           |            |

#### Multiple Regression Analysis

Dependent variable: Yield

#### Table of Estimates

|          | Estimate | Standard<br>Error | t<br>Value | P<br>Value |
|----------|----------|-------------------|------------|------------|
| Constant | 70.31    | 2.57001           | 27.36      | 0.0000     |
| Cat1     | -2.676   | 0.560822          | -4.77      | 0.0002     |
| Cat2     | -0.8802  | 0.70939           | -1.24      | 0.2315     |

R-squared = 58.85 %

Adjusted R-squared = 54.00 %

Standard error of estimation = 3.54695

#### Analysis of Variance

| Source       | Sum of<br>Squares | D.F. | Mean Square | F - Ratio | P<br>Value |
|--------------|-------------------|------|-------------|-----------|------------|
| Model        | 305.808           | 2    | 152.904     | 12.15     | 0.0005     |
| Error        | 213.874           | 17   | 12.5808     |           |            |
| Total(corr.) | 519.682           | 19   |             |           |            |

a. 写出估计后的完全模型。

- b. 写出估计后的简略模型。
- c. 找出完全模型和简略模型的  $R^2$  的值。
- d. 有无令人信服的证据表明模型中的二阶项改进了模型的预测能力?

## 12.6 用多元回归进行的预测

多元回归模型的一个重要应用是在给定  $x$  变量的值以后, 预测相应的  $y$  值。最优猜测预测很简单: 只要把  $x$  的值代入估计后的回归方程即可。在这一节中, 我们讨论有关的标准误差。

与简单回归一样, 对于给定的  $x$  值,  $y$  的预测值可以用两种方式解释。首先, 该值可以认为是对  $E(y)$  的一个最佳猜测, 这里  $E(y)$  为  $x$  取给定的值时,  $y$  的无穷多个观测值的平均值。另一种解释, 通常也是更感兴趣的解释是, 该值是对于  $x$  有给定的值的个体的  $y$  的预测值。不论哪种解释, 标准误差的计算都需要矩阵代数, 我们在本书中对此不做要求。

计算机程序通常给出个体  $y$  值预测的标准误差。虽然这个信息也可以用来求  $E(y)$  的估计的标准误差, 但往往更关心的是个体  $y$  值的预测。在多数计算机程序的输出结果中, 均值的预测区间称为置信区间; 而把个体值的预测区间称为预测区间。预测中加上或减去的项可以由标准误差乘以自由度为  $df = n - (k + 1)$  的一个  $t$  表值得到。实际上, 许多计算机程序直接给出加上或减去的项。作为一个粗略的近似, 我们可以用  $\pm 2 s_e$  来表示一个个体值的预测误差限。

### 例 12.19

一家食品厂的广告部经理要建立一个方程, 以基于儿童电视节目中的广告费用(c), 白天电视节目中的广告费用(d), 报纸中的广告费用(n)预测销售量(s)。收集了前 30 个月的月份数据(除以价格指数以控制通货膨胀的影响), 拟合了一个多元回归模型, 计算机输出结果如下。

```
MTB>regress c1 3 c2-c4;
SUBC>predict at 31 5 12.
```

The regression equation is

$$s = 0.053 + 0.00562 c + 0.0184 d - 0.00600 n$$

| Predictor | Coef     | Stdev    | t-ratio | p     |
|-----------|----------|----------|---------|-------|
| Constant  | 0.0526   | 0.1374   | 0.38    | 0.705 |
| c         | 0.005618 | 0.002930 | 1.92    | 0.066 |

|   |           |          |       |       |
|---|-----------|----------|-------|-------|
| d | 0.01841   | 0.01211  | 1.52  | 0.141 |
| n | -0.005996 | 0.004362 | -1.37 | 0.181 |

s + 0.04736      R-sq + 30.8 %      R-sq(adj) = 22.9 %

#### Analysis of Variance

| SOURCE     | DF | SS       | MS       | F    | p     |
|------------|----|----------|----------|------|-------|
| Regression | 3  | 0.026003 | 0.008668 | 3.86 | 0.021 |
| Error      | 26 | 0.058317 | 0.002243 |      |       |
| Total      | 29 | 0.084320 |          |      |       |

| SOURCE | DF | SEQ SS   |
|--------|----|----------|
| c      | 1  | 0.000330 |
| d      | 1  | 0.021434 |
| n      | 1  | 0.004238 |

| Fit     | Stdev.Fit | 95 % C.I.          | 95 % P.I.          |
|---------|-----------|--------------------|--------------------|
| 0.24686 | 0.01998   | (0.20579, 0.28794) | (0.14118, 0.35255) |

- 写出回归方程。
- 找出  $c = 31$ ,  $d = 5$ ,  $n = 12$  时  $y$  的预测值。找出  $E(y)$  的 95% 置信区间的上限和下限, 以及个体  $y$  值 95% 的预测上限和下限。

#### 解答

- 由标有“Coef”的那一列给出回归方程为

$$\hat{y} = 0.0526 + 0.005618c + 0.01841d - 0.005996n$$

- $y$  的预测值上面标有“Fit”。如同把  $c = 31$ ,  $d = 5$ ,  $n = 12$  代入方程中所能得到的那样,  $y$  的预测值为 0.24686。均值  $E(y)$  的 95% 置信区间在结果中标有“95% C.I.”的地方给出, 即 (0.20579, 0.28794), 而个体  $y$  值 95% 的预测下限和上限分别为 0.14118 和 0.35255。

多元回归中的外推比线性回归中更为微妙。在简单回归中, 外推发生在对数据范围以外的  $x$  值预测其  $y$  值的情形。在多元回归中, 我们不仅要关心各个单个预测因子的范围, 而且还要考虑几个预测因子一起的值的集合。用一个多元回归模型预测一位 30 岁的中层管理人员的工资, 或者预测一位有 25 年工作经验的中层管理人员的工资可能是相当合理的, 但用回归来预测 30 岁的有 25 年工作经验的中层管理人员的工资就不合理了! 外推不仅与用于建立模型的各单个  $x_j$  预测

因子的范围有关,而且还与  $x_j$  的相关性有关。在工资预测的例子中,显而易见,年龄和经验是正相关的,所以小年龄与多年经验的组合在数据中不会出现。当用多元回归进行预测时,我们必须考虑,每个单个变量的值是否是合理的,而且还要考虑所选择的预测因子值的组合是否是合理的。

### 例 12.20

州渔业委员会希望用例 12.18 中的数据来预测每平方英里 8 个居民,0.7 平方英里大小,对外开放变量为 1,结构指标为 55 的一个湖中的捕捞量,以及各变量值分别为 55,1.0,1 和 40 的另一个湖中的捕捞量。Minitab 的输出结果如下:

```
MTB>regress 'catch' on 4 variables 'residenc' 'size' 'access' 'structur';
SUBC>predict at 8 .7 1 55;
SUBC>predict at 55 1.0 1 40.
```

The regression equation is

catch = -1.94 + 0.0193 residenc + 0.332 size + 0.836 access + 0.0477 structur

| Predictor | Coef     | Stdev    | t-ratio | p     |
|-----------|----------|----------|---------|-------|
| Constant  | -1.9378  | 0.9081   | -2.13   | 0.050 |
| residenc  | 0.01929  | 0.01018  | 1.90    | 0.077 |
| size      | 0.3323   | 0.2458   | 1.35    | 0.196 |
| access    | 0.8355   | 0.2250   | 3.71    | 0.002 |
| structur  | 0.047714 | 0.005056 | 9.44    | 0.000 |

s = 0.4336      R-sq = 88.2 %      R-sq(adj) = 85.0 %

| Fit    | Stdev.Fit | 95 % C.I.        | 95 % P.I.           |
|--------|-----------|------------------|---------------------|
| 1.9090 | 0.6812    | (0.4567, 3.3613) | (0.1874, 3.6306) XX |
| 2.1998 | 0.1850    | (1.8054, 2.5941) | (1.1947, 3.2049)    |

X denotes a row with X values away from the center

XX denotes a row with very extreme X values

给出两个湖的捕捞量的 95% 的预测区间。为什么第一个预测区间比第二个宽那么多?

**解答** 预测区间上面标有“95 % P.I.”,对于第一个湖,预测区间为(0.1874,

3.6306), 对于第二个湖, 为(1.1947, 3.2049)。第一个区间带有一个警告信息, 说是  $x$  的值很极端(该信息是由计算机软件给出的, 标有 XX 记号, 译者注)。如果我们回过头来检查一下数据, 就会发现没有一个湖每平方英里的居民数接近 8。因此, 预测是一个严重的外推, 这使得预测区间很宽。在本例中, 问题出自一个预测因子, 其余的  $x$  值都在数据的范围内。

### 练习

**12.29** 基于练习 12.17 中的数据进行了一项预测。回忆  $x$  从 1 变到 3,  $w$  从 1 变到 4,  $v$  从 1 变到 6。下面是 Minitab 的输出结果:

```
MTB>Correlation 'y' 'x' 'w' 'v'.
```

|   | y     | x     | w     |
|---|-------|-------|-------|
| x | 0.856 |       |       |
| w | 0.402 | 0.000 |       |
| v | 0.928 | 0.956 | 0.262 |

```
MTB>Regress 'y' 3 'x' 'w' 'v';
```

```
SUBC>Predict at x 3 w 1 v 6.
```

The regression equation is

$$y = 10.0 + 5.00x + 2.00w + 1.00v$$

s = 2.646      R-sq = 89.5 %      R-sq(adj) = 85.6 %

| Fit    | Stdev.Fit | 95 % C.I.        | 95 % P.I.           |
|--------|-----------|------------------|---------------------|
| 33.000 | 4.077     | (23.595, 42.405) | (21.789, 44.212) XX |

X denotes a row with X values away from the center

XX denotes a row with very extreme X values

找出 95% 的预测区间。解释为什么 Minitab 给出“很极端的 X 值”的警告。

**12.30** 参见练习 12.28 中化学公司的数据。基于有和没有二阶项的模型对  $x_1 = 3.5$ ,  $x_2 = 0.35$ (观察点 21) 以及  $x_1 = 3.5$ ,  $x_2 = 2.5$ (观察点 22) 时的产量预测值进行了计算。Execustat 的输出结果如下:

### Multiple Regression Analysis

Dependent variable: Yield



Table of Estimates

|           | Estimate | Standard Error | t Value | P Value |
|-----------|----------|----------------|---------|---------|
| Constant  | 50.0195  | 4.3905         | 11.39   | 0.0000  |
| Cat1      | 6.64357  | 2.01212        | 3.30    | 0.0052  |
| Cat2      | 7.3145   | 2.73977        | 2.67    | 0.0183  |
| @Cat1Sq   | -1.23143 | 0.301968       | -4.08   | 0.0011  |
| @Cat1Cat2 | -0.7724  | 0.319573       | -2.42   | 0.0299  |
| @Cat2Sq   | -1.1755  | 0.50529        | -2.33   | 0.0355  |

R-squared = 86.24 %

Adjusted R-squared = 81.33 %

Standard error of estimation = 2.25973

Table of Predicted Values (Missing Data Only)

| Row | Predicted Yield | 95.00 % Prediction Limits |         | 95.00 % Confidence Limits |         |
|-----|-----------------|---------------------------|---------|---------------------------|---------|
|     |                 | Lower                     | Upper   | Lower                     | Upper   |
| 21  | 59.926          | 54.7081                   | 65.1439 | 57.993                    | 61.8589 |
| 22  | 62.3679         | 57.0829                   | 67.6529 | 60.2605                   | 64.4753 |

### Multiple Regression Analysis

Dependent variable: Yield

Table of Estimates

|          | Estimate | Standard Error | t Value | P Value |
|----------|----------|----------------|---------|---------|
| Constant | 70.31    | 2.57001        | 27.36   | 0.0000  |
| Cat1     | -2.676   | 0.560822       | -4.77   | 0.0002  |
| Cat2     | -0.8802  | 0.70939        | -1.24   | 0.2315  |

R-squared = 58.85 %

Adjusted R-squared = 54.00 %

Standard error of estimation = 3.54695

Table of Predicted Values (Missing Data Only)

| Row | Yield   | 95.00 %           |         | 95.00 %           |         |
|-----|---------|-------------------|---------|-------------------|---------|
|     |         | Prediction Limits |         | Confidence Limits |         |
|     |         | Lower             | Upper   | Lower             | Upper   |
| 21  | 57.8633 | 50.028            | 65.6986 | 55.5416           | 60.185  |
| 22  | 58.7435 | 51.0525           | 66.4345 | 56.9687           | 60.5183 |

- 找出在回归模型  $\hat{y} = 50.0195 + 6.6436x_1 + 7.3145x_2 - 1.2314x_1^2 - 0.7724x_1x_2 - 1.1755x_2^2$  下个体预测的 95% 预测区间。
- 找出在回归模型  $\hat{y} = 70.3100 - 2.6760x_1 - 0.8802x_2$  下个体预测的 95% 预测区间。
- 在(a) 中模型下的预测区间比(b)中模型下的预测区间是否窄的多?

## 12.7 比较几条回归线的斜率

这个问题是在既有定性自变量又有定量自变量时建立一个多元回归模型这样一个一般问题的特例。说明这个特殊问题的最好的办法是举一个例子。

### 例 12.21

研究人员要比较两种药品(A 和 B)的不同剂量对于老鼠响应的影响。研究采用了一个由 60 只老鼠构成的样本, 这些老鼠处于特殊的紧张状态。把它们随机地分为两个大小相等的组, 第一组的老鼠接受药品 A, A 有三个水平(5, 10 和 20mg), 每个水平随机地分配十只老鼠。同样地, 第二组中的 30 只老鼠接受药品 B, B 有三个水平即 5, 10 和 20mg, 每个水平随机地分配十只老鼠。每只老鼠都接受了给它指定的药剂, 30 分钟观察期过后, 给每只老鼠的焦急表症打分, 分数从 0 到 30。假定焦急得分是药物剂量的线性函数, 写出联系老鼠得分与两个自变量“药品”和“剂量”的模型。解释的其中的各个  $\beta$ 。

### 解答

对于这个试验, 我们有一个定性变量(药品)和一个定量变量(剂量)。令  $x_1$  表示剂量, 则我们有模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

其中

$$x_1 = \text{剂量}$$

$$x_2 = 1 \text{ 如果为产品 B, } x_2 = 0 \text{ 否则}$$

在此模型下,  $y$  的期望值为

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

对于药品 A 和药品 B, 分别代入  $x_2 = 0$  和  $x_2 = 1$ , 我们得到对于给定的剂量, 老鼠焦急得分的期望值为

$$\text{药品 A: } E(y) = \beta_0 + \beta_1 x_1$$

$$\text{药品 B: } E(y) = \beta_0 + \beta_1 x_1 + \beta_2 + \beta_3 x_1 = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x_1$$

这两个期望值表现为线性回归直线。模型的参数可以用这些直线的截距与斜率来解释。具体说来,

$\beta_0$ : 药品 A 回归直线的  $y$  截距,

$\beta_1$ : 药品 A 回归直线的斜率,

$\beta_2$ : 药品 B 和 A 的回归直线的  $y$  截距的差,

$\beta_3$ : 药品 B 和 A 的回归直线的斜率的差。

图 12.4(a) 显示了  $\beta_3 \neq 0$  的情形(也就是说, 在两个变量“药品”和“剂量”之间存在交互效应)。因此, 回归直线不平行。图 12.4(b) 显示了  $\beta_3 = 0$  (无交互效应)的情形, 此时, 两条直线平行。

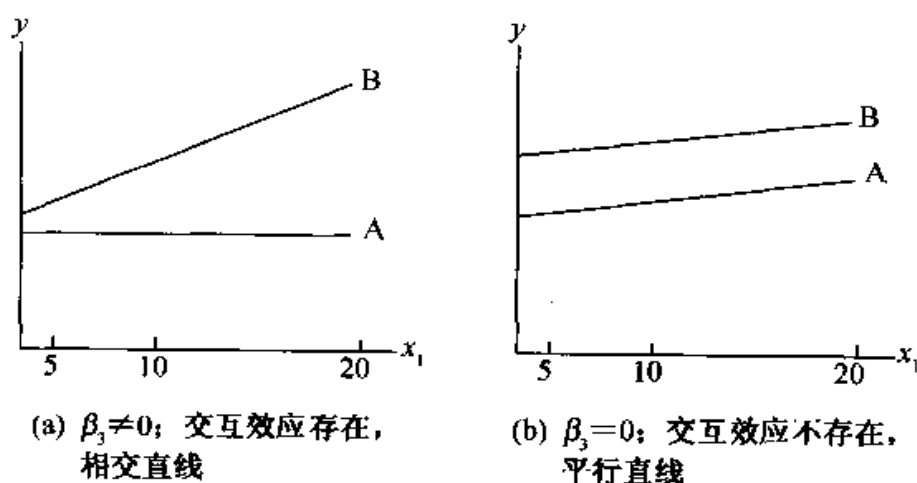


图 12.4 两条回归直线的比较

### 例 12.22

例 12.21 中讨论的试验的样本数据列于表 12.4。感兴趣的响应是由训练有素的研究人员给出的焦急得分。用这些数据拟合一般线性模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

表 12.4 老鼠的焦急得分

| 药品 | 剂量(mg)       |              |              |
|----|--------------|--------------|--------------|
|    | 5            | 10           | 20           |
| A  | 15 16        | 18 16        | 20 17        |
|    | 16 15        | 17 15        | 19 18        |
|    | 18 16        | 18 19        | 21 21        |
|    | 13 17        | 19 18        | 18 20        |
|    | <u>19 15</u> | <u>20 16</u> | <u>19 17</u> |
|    | av = 16      | av = 17.6    | av = 19.0    |
| B  | 16 15        | 19 18        | 24 23        |
|    | 17 15        | 21 20        | 25 24        |
|    | 18 18        | 22 21        | 23 22        |
|    | 17 17        | 23 22        | 25 26        |
|    | <u>15 16</u> | <u>20 19</u> | <u>25 24</u> |
|    | av = 16.4    | av = 20.5    | av = 24.1    |

试验人员特别感兴趣的是比较回归直线的斜率。斜率的差异表明药品对于老鼠的焦急程度有不同的影响。构造一个统计检验,以检验两个斜率是否相等,取  $\alpha = 0.05$ 。

**解答** 用完全模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

我们得到最小二乘拟合为

$$\hat{y} = 15.30 + 0.19x_1 - 0.70x_2 + 0.30x_1x_2$$

SS(回归, 完全) = 442.10, SS(残差, 完全) = 133.63 (参见下面的计算机输出)。

相应于  $H_0: \beta_3 = 0$  (也就是说, 斜率相同) 的简略模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

REGRESSION ANALYSIS OF ANXIETY TREATMENTS - COMPLETE MODEL

Model: MODEL1

Dependent Variable: SCORE

Analysis of Variance

| Source   | DF       | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|----------|----------|-------------------|----------------|---------|--------|
| Model    | 3        | 442.10476         | 147.36825      | 61.758  | 0.0001 |
| Error    | 56       | 133.62857         | 2.38622        |         |        |
| C Total  | 59       | 575.73333         |                |         |        |
| Root MSE | 1.54474  | R - square        | 0.7679         |         |        |
| Dep Mean | 18.93333 | Adj R - sq        | 0.7555         |         |        |
| C.V.     | 8.15884  |                   |                |         |        |

#### Parameter Estimates

| Variable   | DF | Parameter<br>Estimate | Standard<br>Error | T for H0:<br>Parameter = 0 | Prob> T |
|------------|----|-----------------------|-------------------|----------------------------|---------|
| INTERCEP   | 1  | 15.300000             | 0.59827558        | 25.573                     | 0.0001  |
| DOSE       | 1  | 0.191429              | 0.04522538        | 4.233                      | 0.0001  |
| PRODUCT    | 1  | -0.700000             | 0.84608944        | -0.827                     | 0.4116  |
| PRD - DOSE | 1  | 0.300000              | 0.06395835        | 4.691                      | 0.0001  |

| Variable   | DF | Variable<br>Label  |
|------------|----|--------------------|
| INTERCEP   | 1  | Intercept          |
| DOSE       | 1  | DRUG DOSE LEVEL    |
| PRODUCT    | 1  | DRUG PRODUCT       |
| PRD - DOSE | 1  | PRODUCT TIMES DOSE |

#### REGRESSION ANALYSIS OF ANXIETY TREATMENTS - REDUCED MODEL

Model: MODEL1

Dependent Variable: SCORE

#### Analysis of Variance

| Source | DF | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|--------|----|-------------------|----------------|---------|--------|
|--------|----|-------------------|----------------|---------|--------|

|         |    |           |           |        |        |
|---------|----|-----------|-----------|--------|--------|
| Model   | 2  | 389.60476 | 194.80238 | 59.656 | 0.0001 |
| Error   | 57 | 186.12857 | 3.26541   |        |        |
| C Total | 59 | 575.73333 |           |        |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 1.80705  | R-square | 0.6767 |
| Dep Mean | 18.93333 | Adj R-sq | 0.6654 |
| C.V.     | 9.54425  |          |        |

## Parameter Estimates

| Variable | DF | Parameter | Standard   | T for H0:     |           |
|----------|----|-----------|------------|---------------|-----------|
|          |    | Estimate  | Error      | Parameter = 0 | Prob >  T |
| INTERCEP | 1  | 13.550000 | 0.54711020 | 24.766        | 0.0001    |
| DOSE     | 1  | 0.341429  | 0.03740940 | 9.127         | 0.0001    |
| PRODUCT  | 1  | 2.800000  | 0.46657715 | 6.001         | 0.0001    |

| Variable | DF | Variable        |
|----------|----|-----------------|
|          |    | Label           |
| INTERCEP | 1  | Intercept       |
| DOSE     | 1  | DRUG DOSE LEVEL |
| PRODUCT  | 1  | DRUG PRODUCT    |

由此我们得到

$$\hat{y} = 13.55 + 0.34x_1 - 2.80x_2$$

SS(回归, 简略) = 389.60。归因于  $x_1, x_2$  的残差平方和的下降为

$$\begin{aligned} SS_{drop} &= SS(\text{回归, 完全}) - SS(\text{回归, 简略}) \\ &= 442.10 - 389.60 = 52.50 \end{aligned}$$

从而

$$\begin{aligned} F &= \frac{[SS(\text{回归, 完全}) - SS(\text{回归, 简略})]/(k - g)}{SS(\text{回归, 完全})/[n - (k + 1)]} \\ &= \frac{52.50/1}{133.63/56} = 22.00 \end{aligned}$$

由于观测到的  $F$  值超过了附表 8 中与  $df_1 = 1, df_2 = 56$  (实际上为 60) 和  $\alpha = 0.05$  对应的值 4.00, 我们拒绝  $H_0$ , 认为两组的斜率是不同的。注意, 我们可以用  $t$  检验来检验  $H_0: \beta_3 = 0$ , 并得到相同的结果。从计算机输出结果,  $t$  统计量的值为

4.69, 在 0.0001 的水平上是显著的。对于这种类型的检验问题,  $t$  统计量和  $F$  统计量的值是有关系的:  $t^2 = F$  (这里  $4.69^2 \approx 22$ )。

上面给出的关于比较两条回归直线的斜率的结果很容易推广到三条或更多条回归直线的比较的情形, 只要增加模型中的虚拟变量以及所有定量变量和虚拟变量之间的交互作用项就行了。例如, 在比较三条回归直线的斜率时, 模型应当包含定量变量  $x_1$ , 两个虚拟变量  $x_2$  和  $x_3$ , 以及两个交互作用项  $x_1x_2$  和  $x_1x_3$ 。

### 练习

### 应用

**12.31(医学)** 试验人员想比较三种不同的药品的效力。为此, 往 12 支试管中接种了要研究的病毒培养物, 在  $35^\circ\text{C}$  的条件下孵化两天。三种药品各取 4 个水平(0.2, 0.4, 0.8 和 1.6), 12 支试管培养物中每支试管接受剂量-药品的一个组合。比较这些药品的方法之一是考察它们的(对于剂量)斜率。

a. 写出联系响应  $y$  与自变量“剂量”和“药品”的一般线性模型, 使得期望响应为  $\log$  剂量( $x_1$ )的线性函数。解释模型中的参数。

b. 假定三条回归直线有相同的截距从直观上看来是合理的, 因为这相应于药品的 0 剂量水平。对(a)中的模型作出变化, 以适应这一假定。

**12.32** 参见练习 12.32。

a. 用下面的数据作出三个斜率的比较。为了你的检验拟合完全模型和简略模型。取  $\alpha = 0.05$ 。

| 剂量  | 药品  |     |     |
|-----|-----|-----|-----|
|     | A   | B   | C   |
| 0.2 | 2.0 | 1.8 | 1.3 |
| 0.4 | 4.3 | 4.1 | 2.0 |
| 0.8 | 6.5 | 4.9 | 2.8 |
| 1.6 | 8.9 | 5.7 | 3.4 |

b. 是否有依据表明斜率是相等的?

c. 就如何检验零假设“截距全为 0”提出建议。

## 12.8 Logistic 回归

在许多研究中,许多响应变量表现为两个可能的值之一。比如,响应变量是一个取值于 0, 1 的二值随机变量。例如,黄曲霉素  $B_1$  被怀疑为致癌物质。在关于它的一项研究中,用了不同剂量的黄曲霉素  $B_1$  喂养试验动物。在一段时间以后,解剖这些动物,记录下患肝脏肿瘤的患者数目。如果一个动物有肿瘤,其响应值  $y=1$ , 否则  $y=0$ 。类似地,银行要确定哪些顾客最像是能够归还贷款的人。为此,他们要记录若干自变量以描述这些顾客的可靠程度,然后,决定这些变量是否与二值变量  $y$  有关,这里,  $y=1$  如果顾客偿还贷款,  $y=0$  如果顾客不偿还贷款。下面介绍一个联系二值变量  $y$  与解释变量的模型。

当响应变量  $y$  为二值变量时,  $y$  的分布归结为一个值,即  $p = Pr(y=1)$ 。我们要把  $p$  与自变量的一个线性组合联系起来,困难是  $p$  在 0, 1 之间变化,而解释变量的线性组合可以在  $(-\infty, +\infty)$  内变化。在第十章中,我们介绍了把概率变为几率(odds, 原文误为 odds ratio——译者注)的变换。由于概率从 0 变到 1, 几率从 0 变到无穷。取几率的对数,则当概率从 0 变到 1 时,得到一个取值于  $(-\infty, +\infty)$  的变量。用来研究二值变量与一组解释变量之间联系的常用模型为 logistic 回归分析模型。在这个模型中,几率的自然对数与解释变量通过一个线性模型联系起来。我们这里只考虑单个自变量的情形,但是,这个模型可以推广到多个自变量的模型。令  $p(x)$  为自变量等于  $x$  时,  $y$  等于 1 的概率。我们用  $x$  的线性函数来建立对数几率的模型,即简单 logistic 回归模型:

$$\ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x$$

这个变换可以直接用  $p(x)$  表示成

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

例如,一只暴露于  $x$  个单位的黄曲霉素  $B_1$  的老鼠患癌症的概率就可能用上述方程中的  $p(x)$  来表示。 $\beta_0$  和  $\beta_1$  的值可以从观测数据用极大似然估计的方法估计出来。

我们可以用  $p(x)$  来解释 logistic 模型中的参数  $\beta_0$  和  $\beta_1$ 。从截距参数  $\beta_0$  可以得到自变量  $x=0$  时,与  $y=1$  对应的事件的发生概率。例如,当动物没有暴露于黄曲霉素  $B_1$  时,患癌症的概率相应于  $x=0$  时  $y=1$  的概率,即  $p(0)$ 。从 logistic 回归模型可以得到

$$p(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$$



斜率参数  $\beta_1$  衡量了事件发生概率与自变量的取值之间相关联的程度。当  $\beta_1 = 0$  时, 事件发生的概率与  $x$  取值的大小没有关系。在我们的例子中, 这相当于说动物患肝癌的概率与暴露的黄曲霉素  $B_1$  的量的大小无关。图 12.5 中给出了两个 logistic 回归函数的曲线。如果  $\beta_1 > 0$ , 事件发生的概率随自变量的值的增加而增加; 如果  $\beta_1 < 0$ , 则事件发生的概率随自变量的值的增加而减少。

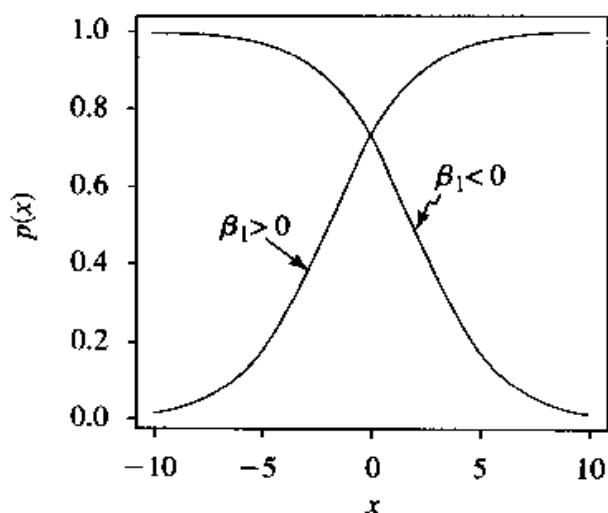


图 12.5 logistic 回归函数

当  $\beta_0$  和  $\beta_1$  均为 0 时, 事件的发生和不发生有同样的可能性, 这是因为

$$p(x) = \frac{e^0}{1 + e^0} = \frac{1}{1 + 1} = \frac{1}{2}$$

这表明与  $y = 1$  相应的事件发生的概率与变量  $x$  无关。由此, 模型对于确定事件发生的概率没有提供信息, 所以不论自变量  $x$  的值是什么, 事件的发生与不发生的机会相等。

不论我们是应用简单 logistic 回归模型还是多元 logistic 回归模型, 用于估计模型参数的计算方法都需要用计算机软件来实现。下面我们用一个例子来说明 logistic 回归模型的应用。

### 例 12.23

A. F. Smith(1967)的一个研究报告(*Lancet* 杂志, 第 2 卷, 178)记录了被怀疑有过心脏病发作的患者的一种酶, 即肌氨酸酐酶(CK)的水平。这项研究的目的是判断患者入住医院时 CK 量的测量, 能否有助于识别被诊断为有心脏病发作的患者是否真有过心脏病发作。测量了 360 名患者入院时的 CK 的水平, 在经过一段时间之后, 医生浏览了这些患者的记录, 确定 360 名患者中哪些患者实际上真有过心脏病发作。下面的表中给出了试验的数据。

| CK 值 | 有心脏病患者数 | 无心脏病患者数 |
|------|---------|---------|
| 20   | 2       | 88      |
| 60   | 13      | 26      |
| 100  | 30      | 8       |
| 140  | 30      | 5       |
| 180  | 21      | 0       |
| 220  | 19      | 1       |
| 260  | 18      | 1       |
| 300  | 13      | 1       |
| 340  | 19      | 0       |
| 380  | 15      | 0       |
| 420  | 7       | 0       |
| 460  | 8       | 0       |
| 500  | 35      | 0       |

用计算机软件计算回归方程曲线的估计以及心脏病发作概率的 95% 置信区间, 输出结果如下。

#### LOGISTIC REGRESSION ANALYSIS EXAMPLE

The LOGISTIC Procedure

Data Set: WORK. LOGREG

Response Variable (Events): R

Response Variable (Trials): N

Number of Observations: 13

Link Function: Logit

| Response Profile |          |       |
|------------------|----------|-------|
| Ordered          | Binary   |       |
| Value            | Outcome  | Count |
| 1                | EVENT    | 230   |
| 2                | NO EVENT | 130   |

Model Fitting Information and Testing Global Null Hypothesis BETA = 0

| Intercept     |      |            |                           |
|---------------|------|------------|---------------------------|
| Intercept and |      |            |                           |
| Criterion     | Only | Covariates | Chi-Square for Covariates |

|           |         |         |                                |
|-----------|---------|---------|--------------------------------|
| AIC       | 472.919 | 191.773 | .                              |
| SC        | 476.806 | 199.545 | .                              |
| - 2 LOG L | 470.919 | 187.773 | 283.147 with 1 DF (p = 0.0001) |
| Score     | .       | .       | 159.142 with 1 DF (p = 0.0001) |

## Analysis of Maximum Likelihood Estimates

| Variable | DF | Parameter Estimate | Standard Error | wald Chi-Square | Pr > Chi-Square | Standardized Estimate |
|----------|----|--------------------|----------------|-----------------|-----------------|-----------------------|
| INTERCPT | 1  | - 3.0284           | 0.3670         | 68.0948         | 0.0001          |                       |
| CK       | 1  | 0.0351             | 0.00408        | 73.9842         | 0.0001          | 3.100511              |

## LOGISTIC REGRESSION ANALYSIS EXAMPLE

| OBS | CK  | PRED    | LCL     | UCL     |
|-----|-----|---------|---------|---------|
| 1   | 20  | 0.08897 | 0.05151 | 0.14937 |
| 2   | 60  | 0.28453 | 0.21224 | 0.36988 |
| 3   | 100 | 0.61824 | 0.51935 | 0.70821 |
| 4   | 140 | 0.86833 | 0.78063 | 0.92436 |
| 5   | 180 | 0.96410 | 0.91643 | 0.98502 |
| 6   | 220 | 0.99094 | 0.97067 | 0.99724 |
| 7   | 260 | 0.99776 | 0.99000 | 0.99950 |
| 8   | 300 | 0.99945 | 0.99662 | 0.99991 |
| 9   | 340 | 0.99986 | 0.99886 | 0.99998 |
| 10  | 380 | 0.99997 | 0.99962 | 1.00000 |
| 11  | 420 | 0.99999 | 0.99987 | 1.00000 |
| 12  | 460 | 1.00000 | 0.99996 | 1.00000 |
| 13  | 500 | 1.00000 | 0.99999 | 1.00000 |

- CK 水平与心脏病发作概率通过 logistic 回归模型的联系是显著的吗?
- 从计算机的输出结果, 得到  $\beta_0$  与  $\beta_1$  的估计。
- 作为 CK 水平的函数, 构造心脏病发作概率的估计。特别地, 估计 CK 水平为 140 的患者的发作概率。

## 解答

- 从输出结果, 我们得到, 在 logistic 回归中假设  $H_0: \beta_1 = 0$   $H_a: \beta_1 \neq 0$  的  $p$  值为 0.0001, 由此, CK 与心脏病发作的概率是显著的。
- 从计算机输出结果, 我们得到  $\hat{\beta}_0 = -3.0284$  和  $\hat{\beta}_1 = 0.0351$ 。注意  $\hat{\beta}_1$  是正的, 这表明, 有高 CK 水平的患者已经发生心脏病发作的概率大。

c. 作为 CK 水平的函数, 心脏病发作概率的估计由

$$p(\hat{CK}) = \frac{e^{-3.0284+0.0351 * CK}}{1 + e^{-3.0284+0.0351 * CK}}$$

我们用此公式来计算一个 CK 水平为 140 的患者受到心脏病发作的概率。这个概率为

$$p(\hat{CK}) = \frac{e^{-3.0284+0.0351 * 140}}{1 + e^{-3.0284+0.0351 * 140}} = \frac{e^{1.886}}{1 + e^{1.886}} = 0.868$$

由上面的计算机输出, 我们得到这个概率的 95% 置信区间为从 0.781 到 0.924。因此, 我们有 95% 的自信说, 水平为 140 的患者中, 78.1% 到 92.4% 的人有过心脏病发作。心脏病发作的概率的估计以及这些概率的 95% 的置信区间由图 12.6 表示出来。我们注意, 心脏病发作的概率的估计随 CK 水平的增加而迅速增长。这表明 CK 水平是心脏病发作的一个有用的指标。

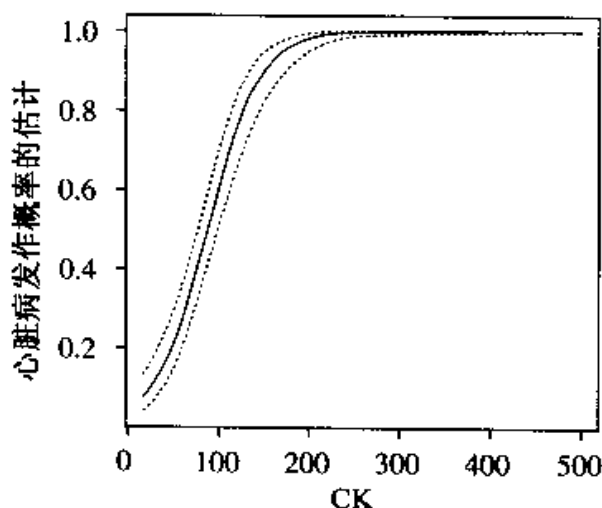


图 12.6 心脏病发作概率的估计及其 95% 置信区间

练习

应用

**12.33(工程)** 一位质量控制工程师研究了系统控制工程师的经验和他们在给定时间内完成包括跟踪所有计算机程序和控制设备这样的复杂设计的能力之间的关系。对于 25 位经验(按月计算)大相径庭的工程师, 给了他们同样的控制设计项目。研究的结果在下表中给出, 其中, 如果一名工程师在规定的时间内成功地完成了该设计项目, 则对应的  $y = 1$ , 否则, 如果完不成项目, 则对应的  $y = 0$ 。

a. 用下面的计算机输出确定经验与完成任务的概率是否有联系。

| 经验月数 | 项目完成 | 经验月数 | 项目完成 |
|------|------|------|------|
| 2    | 0    | 15   | 1    |
| 4    | 0    | 16   | 1    |
| 5    | 0    | 17   | 0    |
| 6    | 0    | 19   | 1    |
| 7    | 0    | 20   | 1    |
| 8    | 1    | 22   | 0    |
| 8    | 1    | 23   | 1    |
| 9    | 0    | 24   | 1    |
| 10   | 0    | 27   | 1    |
| 10   | 0    | 30   | 0    |
| 11   | 1    | 31   | 1    |
| 12   | 1    | 32   | 1    |
| 13   | 0    |      |      |

b. 计算有 24 个月经验的工程师成功地完成任务的概率。给出这个概率的 95% 置信区间。

SAS Code for Logistic Regression

```
option ls = 70 ps = 55 nocenter nodate;
data logreg;
input x y @@;
label x = 'MONTHS EXPERIENCE' y = 'SUCCESS INDICATOR'; cards;
2 0 4 0 5 0 6 0 7 0 8 1 8 1 9 0 10 0 10 0 11 1 12 1
13 0 15 1 16 1 17 0 19 1 20 1 22 0 23 1 24 1 27 1 30 0 31 1 32 1
run;
proc print;
proc logistic descending;
model y = x;
output out = new p = pred lower = lcl upper = ucl;
proc sort; by x;
proc print; var x pred lcl ucl;
run;
```

-----

The LOGISTIC Procedure

Data Set: WORK. LOGREG

Response Variable: Y                      SUCCESS INDICATOR

Response Levels: 2

Number of Observations: 25

Link Function: Logit

Response Profile

Ordered

| Value | Y | Count |
|-------|---|-------|
| 1     | 1 | 13    |
| 2     | 0 | 12    |

## Analysis of Maximum Likelihood Estimates

| Variable | DF | Parameter Estimate | Standard Error | Wald Chi-Square | Pr > Chi-Square | Standardized Estimate |
|----------|----|--------------------|----------------|-----------------|-----------------|-----------------------|
| INTERCPT | 1  | -1.6842            | 0.9451         | 3.1759          | 0.0747          | .                     |
| X        | 1  | 0.1194             | 0.0589         | 4.1091          | 0.0427          | 0.585706              |

The LOGISTIC Procedure

## Association of Predicted Probabilities and Observed Responses

|             |   |        |           |   |       |
|-------------|---|--------|-----------|---|-------|
| Concordant  | = | 77.6 % | Somers' D | = | 0.551 |
| Discordant  | = | 22.4 % | Gamma     | = | 0.551 |
| Tied        | = | 0.0 %  | Tau-a     | = | 0.287 |
| (156 pairs) |   |        | c         | = | 0.776 |

## The LOGISTIC Procedure

| OBS | X  | PRED    | 95 % Lower Limit | 95 % Upper Limit |
|-----|----|---------|------------------|------------------|
| 1   | 2  | 0.19070 | 0.04320          | 0.55155          |
| 2   | 4  | 0.23029 | 0.06487          | 0.56339          |
| 3   | 5  | 0.25213 | 0.07884          | 0.57042          |
| 4   | 6  | 0.27530 | 0.09518          | 0.57839          |
| 5   | 7  | 0.29974 | 0.11399          | 0.58749          |
| 6   | 8  | 0.32538 | 0.13526          | 0.59794          |
| 7   | 8  | 0.32538 | 0.13526          | 0.59794          |
| 8   | 9  | 0.35211 | 0.15884          | 0.61001          |
| 9   | 10 | 0.37980 | 0.18434          | 0.62397          |
| 10  | 10 | 0.37980 | 0.18434          | 0.62397          |
| 11  | 11 | 0.40830 | 0.21117          | 0.64011          |
| 12  | 12 | 0.43742 | 0.23858          | 0.65863          |

| OBS | X  | PRED    | Limit   | Limit   |
|-----|----|---------|---------|---------|
| 13  | 13 | 0.46698 | 0.26568 | 0.67964 |
| 14  | 15 | 0.52660 | 0.31574 | 0.72839 |
| 15  | 16 | 0.55623 | 0.33753 | 0.75512 |
| 16  | 17 | 0.58547 | 0.35684 | 0.78239 |
| 17  | 19 | 0.64199 | 0.38830 | 0.83514 |
| 18  | 20 | 0.66894 | 0.40092 | 0.85917 |
| 19  | 22 | 0.71954 | 0.42133 | 0.90040 |
| 20  | 23 | 0.74299 | 0.42962 | 0.91732 |
| 21  | 24 | 0.76512 | 0.43691 | 0.93186 |
| 22  | 27 | 0.82333 | 0.45436 | 0.96307 |
| 23  | 30 | 0.86958 | 0.46732 | 0.98065 |
| 24  | 31 | 0.88253 | 0.47097 | 0.98447 |
| 25  | 32 | 0.89435 | 0.47436 | 0.98756 |

**12.34** 最近开发出一种内粉刷用涂料的添加剂,这种添加剂可以大大增加涂料抵抗玷污的能力。为调查这种添加剂对于儿童是否安全,进行了一项研究。把不同剂量的这种添加剂喂给试验动物,记录了动物患肝癌的个数。数据有下表给出。

|         |    |    |    |    |     |     |
|---------|----|----|----|----|-----|-----|
| 剂量(ppm) | 0  | 10 | 25 | 50 | 100 | 200 |
| 实验动物数   | 30 | 20 | 20 | 30 | 30  | 30  |
| 患癌症的动物数 | 0  | 2  | 2  | 7  | 25  | 30  |

- 用下面的计算机输出确定喂给试验动物的添加剂的剂量是否与动物患肝癌的概率有关。
- 计算添加剂暴露量为 100ppm 的试验动物患肝癌的概率;给出这个概率的 95% 置信区间。

SAS Code for Exercise

```
option ls = 70 ps = 55 nocenter nodate;
TITLE 'OUTPUT FOR EXERCISE';
data logreg;
input x R N @@;
label x = 'AMOUNT (PPM)'; cards;
0 0 30 10 2 20 25 2 20 50 7 30 100 25 30 200 30 30
run;
proc print;
```

```

proc logistic descending;
model R/N = x;
output out = new p = pred lower = lc1 upper = uc1;
proc sort; by x;
proc print; var x pred lc1 uc1;
run;

```

-----

OUTPUT FOR EXERCISE

| OBS | X   | R  | N  |
|-----|-----|----|----|
| 1   | 0   | 0  | 30 |
| 2   | 10  | 2  | 20 |
| 3   | 25  | 2  | 20 |
| 4   | 50  | 7  | 30 |
| 5   | 100 | 25 | 30 |
| 6   | 200 | 30 | 30 |

The LOGISTIC Procedure

Response Profile

| Ordered<br>Value | Binary<br>Outcome | Count |
|------------------|-------------------|-------|
| 1                | EVENT             | 66    |
| 2                | NO EVENT          | 94    |

Analysis of Maximum Likelihood Estimates

| Variable | DF | Parameter<br>Estimate | Standard<br>Error | Wald<br>Chi-Square | Pr><br>Chi-Square | Standardized<br>Estimate |
|----------|----|-----------------------|-------------------|--------------------|-------------------|--------------------------|
| INTERCPT | 1  | -3.6429               | 0.5530            | 43.3998            | 0.0001            |                          |
| X        | 1  | 0.0521                | 0.00824           | 39.9911            | 0.0001            | 2.044518                 |

Analysis of Maximum  
Likelihood Estimates

| Variable | Odds<br>Ratio | Variable<br>Label |
|----------|---------------|-------------------|
| INTERCPT | .             | Intercept         |
| X        | 1.053         | AMOUNT(PPM)       |



| OBS | X   | PRED    | 95 % Lower<br>Limit | 95 % Upper<br>Limit |
|-----|-----|---------|---------------------|---------------------|
| 1   | 0   | 0.02551 | 0.00878             | 0.07182             |
| 2   | 10  | 0.04221 | 0.01681             | 0.10203             |
| 3   | 25  | 0.08783 | 0.04308             | 0.17077             |
| 4   | 50  | 0.26156 | 0.16907             | 0.38142             |
| 5   | 100 | 0.82738 | 0.66925             | 0.91905             |
| 6   | 200 | 0.99886 | 0.98818             | 0.99989             |

## 12.9 多元回归的一些理论结果(任选)

在本节中,我们用矩阵的符号来给出多元回归的一些理论结果,目的是看一看多元回归的计算,不论是手工计算还是用计算机计算,在实际上是如何进行的。我们不给出其中多数结果的证明,这些证明可在许多专门的教材中找到,如 Draper and Smith(1998)。

矩阵记号的使用在于是多元回归本身。回忆把响应变量与一组自变量联系在一起的形如

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

的模型,该模型称为一般线性模型。在这个模型中,截距参数和偏斜率参数的最小二乘估计  $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$  可以用矩阵得到。

定义  $n \times 1$  矩阵  $\mathbf{Y}$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

为观察值矩阵,  $n \times (k+1)$  矩阵  $\mathbf{X}$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

为增加了一列 1 的自变量取值矩阵。矩阵的第一行包含一个 1 和对应于第一个  $y$  的观测值的  $k$  个自变量的取值,第二行包含一个 1 和对应于  $y_2$  的  $k$  个自变量的取值,同理,其他的行包含其余的各个  $y$  的观测值对应的自变量的取值。

接下来我们来看多元回归模型中截距参数和偏斜率参数的最小二乘估计  $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$ 。回忆最小二乘原理是通过极小化残差平方和来选取估计。熟悉微积分的读者会看到,求  $SS(\text{残差})$  关于  $\beta_0, \beta_1, \cdots, \beta_k$  的微分并令其为 0, 即可得到这

个问题的解。所得到的正规方程,用矩阵表示出来,即为

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}$$

其中

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

为要求的系数估计的向量。假定矩阵  $\mathbf{X}'\mathbf{X}$  有逆矩阵(只要没有一个  $x_j$  与其他  $x$  完全共线性即可满足这一假定),则该方程的解为

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

#### 例 12.24

假设在某个试验中,

$$\mathbf{Y} = \begin{bmatrix} 25 \\ 19 \\ 33 \\ 23 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & -2 & 5 \\ 1 & -2 & -5 \\ 1 & 2 & 5 \\ 1 & 2 & -5 \end{bmatrix}$$

求预测方程

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

的最小二乘估计。

**解答** 由这些数据,

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 100 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 100 \\ 24 \\ 80 \end{bmatrix}$$

矩阵  $\mathbf{X}'\mathbf{X}$  是一个对角矩阵,其逆矩阵很容易求。正规方程的解为

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \\ &= \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.0625 & 0 \\ 0 & 0 & 0.01 \end{bmatrix} \begin{bmatrix} 100 \\ 24 \\ 80 \end{bmatrix} = \begin{bmatrix} 25 \\ 1.5 \\ 0.8 \end{bmatrix} \end{aligned}$$

预测方程为

$$\hat{y} = 25 + 1.5x_1 + 0.8x_2$$

在多元回归的计算中最困难的部分是计算  $\mathbf{X}'\mathbf{X}$  的逆矩阵。对于大多数实际的多元回归问题来说,手工计算逆矩阵要花数小时,计算机计算要花不到一秒的时间。这就是为什么多数多元回归问题要用计算机软件来算的原因。

一旦求出  $\mathbf{X}'\mathbf{X}$  的逆矩阵,并算出向量  $\hat{\beta}$ ,接下来就要计算残差标准差。这个艰巨的任务就是要计算  $SS(\text{残差}) = \sum (y_i - \hat{y}_i)^2$ ,而这个和可以写为  $SS(\text{残差}) = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'(\mathbf{X}'\mathbf{Y})$ 。

### 例 12.25

用例 12.24 中的数据计算  $SS(\text{残差})$ 。

解答

计算出  $\hat{\beta}$  和  $\mathbf{X}'\mathbf{Y}$  分别为  $\begin{bmatrix} 25 \\ 1.5 \\ 0.8 \end{bmatrix}$  和  $\begin{bmatrix} 100 \\ 24 \\ 80 \end{bmatrix}$

$$\mathbf{Y}'\mathbf{Y} = \begin{bmatrix} 25 & 19 & 33 & 23 \end{bmatrix} \begin{bmatrix} 25 \\ 19 \\ 33 \\ 23 \end{bmatrix} = 2,604$$

由残差平方和的矩阵表达式得

$$SS(\text{残差}) = 2,604 - \mathbf{Y}'\mathbf{Y} = \begin{bmatrix} 25 & 1.5 & 0.8 \end{bmatrix} \begin{bmatrix} 100 \\ 24 \\ 80 \end{bmatrix} = 4$$

用类似的方法可得到  $SS(\text{回归})$  和  $SS(\text{总和})$ 。虽然这些和的公式也可以用纯矩阵的符号表示出来,但混合使用矩阵和代数的符号表示起来更简单:

$$SS(\text{回归}) = \hat{\beta}'(\mathbf{X}'\mathbf{Y}) - \frac{(\sum y_i)^2}{n}$$

$$SS(\text{总和}) = \mathbf{Y}'\mathbf{Y} - \frac{(\sum y_i)^2}{n}$$

定义 12.5  $\hat{\beta}_j$  的标准误差的估计为

$$S_{\hat{\beta}_j} = s_e \sqrt{v_{jj}}$$

其中  $s_e$  是回归方程的标准差,  $v_{jj}$  是矩阵

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} v_{00} & & & \\ & v_{11} & & \\ & & \ddots & \\ & & & v_{kk} \end{bmatrix}$$

$j+1$  行、 $j+1$  列的元素。由于要得到诸  $\hat{\beta}_j$ , 必须要计算出  $(\mathbf{X}'\mathbf{X})^{-1}$ , 所以容易得到这个标准误差的估计。

## 12.10 小 结

本章强化了把响应  $y$  表示成一个或多个自变量的函数的内容, 讨论了多元回归模型(所有自变量都是定量的)以及结合了定性变量信息的模型。这些模型可以表示为如下一般线性模型的形式:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

在讨论了各种模型以及这些模型中  $\beta$  的解释后, 我们给出了用于得到最小二乘估计  $\hat{\beta}$  的正规方程。

用  $\hat{\beta}_j$  和  $\hat{\beta}_j$  的标准差建立了单个参数  $\hat{\beta}$  的置信区间和统计检验。我们还考虑了一组  $\hat{\beta}$  的统计检验, 基于一组  $x$  值的  $E(y)$  的置信区间以及给定  $x$  的值后的预测区间。

所有这些推断都需要进行一定量的数字计算, 除非有计算机软件可用。有时, 如果读者熟悉矩阵运算, 这些计算可以用手工进行(见 12.9 节)。然而, 即使用矩阵的方法, 随着自变量个数的增加, 也会无能为力。因此, 十分清楚, 一般线性模型的推断应当使用计算机软件, 以利于分析和极小化计算误差。此时, 我们的工作就是浏览和解释这些软件的输出结果。

除了少数几个练习为了提高你对这些计算的理解而包含一些计算外, 本章中其余的练习和下一章中的回归问题都大量地利用了窘盘的输出结果。

关于多元回归的概念另外还有一些内容:

1. 一阶模型(即不包含变换了的项, 如一个变量的平方项或乘积项)中的回归系数应解释为偏斜率, 即一个自变量增加一个单位, 而其余的自变量保持不变时因变量的变化的预期值。
2. 不仅自变量与因变量之间的相关是重要的, 而且自变量之间的相关也是重要的。共线性——自变量之间的相关性——意味着当变量被加入回归模型或从模型中去除时, 回归系数将发生变化。
3. 回归模型的效果不仅可以由  $R^2$  表示出来, 也可以由残差标准差来表示。用残差标准差看一下要预测的值大体落在预测值的什么范围内常是有益

的。

4. 一如既往, 回归模型中各种统计检验只不过表明, 认为所考虑的模式的信息比随机出现的来得多时, 你有多强的依据。这些检验并不直接表明一个预测模型有多么好。特别地, 在大样本的情形下, 一个大的总体  $F$  统计量的值可能仅仅表明回归模型有较弱的预测能力。
5. 多元回归中的  $t$  检验表明, 把一个自变量作为预测因子加入到模型中时, 是否附加了特有的预测价值。很可能出现这样的情况, 即几个自变量中没有哪个附加了具有统计显著性的特有的预测价值, 但把它们从模型都删除会引起模型预测能力的严重的下降。尤其在有严重的共线性时, 这种情况更可能发生。
6. 方差膨胀因子(VIF)是衡量共线性在估计自变量的系数时的作用的一个有用的指标。VIF 的值越高, 共线性对于斜率估计的精确性的影响也越严重。
7. 多元回归中的外推可能是件微妙的事情。一组新的  $x$  值, 其每个分量单独考虑时, 可能没有什么不合理, 但某些分量一起考虑时, 就可能远远地落在原来的数据范围之外。

#### 重要公式

$$1. R^2_{y \cdot x_1 \cdots x_k} = \frac{SS(\text{总和}) - SS(\text{残差})}{\text{总和}}$$

其中

$$SS(\text{总和}) = \sum (y_i - \bar{y})^2$$

$$SS(\text{残差}) = \sum (y_i - \hat{y}_i)^2$$

$$2. SS(\text{回归}) = \sum (\hat{y}_i - \bar{y})^2 \text{ 和}$$

$$SS(\text{总和}) = \sum (y_i - \bar{y})^2 = SS(\text{回归}) + SS(\text{残差})$$

$$3. H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0 \text{ 的 } F \text{ 检验}$$

$$F = \frac{SS(\text{回归})/k}{SS(\text{残差})/[n - (k + 1)]}$$

4.

$$s_{\hat{\beta}_j} = s_e \sqrt{\frac{1}{\sum (x_{ij} - \bar{x}_j)(1 - R^2_{x_j \cdot x_1 \cdots x_{j-1} x_{j+1} \cdots x_k})}}$$

和

$$s_e = \frac{MS(\text{残差})}{n - (k + 1)}$$

5.  $\beta_j$  的置信区间为

$$\hat{\beta}_j \pm t_{\alpha/2} s_{\hat{\beta}_j} \leq \beta_j \leq \hat{\beta}_j \pm t_{\alpha/2} s_{\hat{\beta}_j}$$

6.  $\beta_i$  的统计检验

$$\text{T.S.: } t = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$$

7. 预测因子子集的  $F$  检验

$$H_0: \beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0$$

$$\text{T.S.: } F = \frac{[\text{SS}(\text{回归, 完全}) - \text{SS}(\text{回归, 简略})]/(k - g)}{\text{SS}(\text{回归, 完全})/[n - (k + 1)]}$$

### 补充练习

**12.35 (商业)** 在一项进口微型轿车需求的研究中, 收集了 12 个大都市区域内的数据, 其中的变量为:

需求(Demand): 进口微型轿车的销售量占总销售量的百分比。

教育(Educ): 成人上学年数的平均数。

收入(Income): 人均收入。

人口(Popn): 地区人口。

家庭大小(Famsize): 完整家庭的平均人数。

Minitab 输出如下:

```
MTB> Regress 'Demand' 4 'Educ' 'Income' 'Popn' 'Famsize'.
```

The regression equation is

```
Demand = - 1.3 + 5.55 Educ + 0.89 Income + 1.92 Popn - 11.4 Famsize
```

| Predictor | Coef     | Stdev | t-ratio | p     |
|-----------|----------|-------|---------|-------|
| Constant  | - 1.32   | 57.98 | - 0.02  | 0.982 |
| Educ      | 5.550    | 2.702 | 2.05    | 0.079 |
| Income    | 0.885    | 1.308 | 0.68    | 0.520 |
| Popn      | 1.925    | 1.371 | 1.40    | 0.203 |
| Famsize   | - 11.389 | 6.669 | - 1.71  | 0.131 |

```
s = 2.686    R-sq = 96.2 %    R-sq(adj) = 94.1 %
```

Analysis of Variance

| SOURCE     | DF | SS      | MS     | F     | p     |
|------------|----|---------|--------|-------|-------|
| Regression | 4  | 1295.70 | 323.93 | 44.89 | 0.000 |
| Error      | 7  | 50.51   | 7.22   |       |       |
| Total      | 11 | 1346.22 |        |       |       |

| SOURCE  | DF | SEQ SS  |
|---------|----|---------|
| Educ    | 1  | 1239.95 |
| Income  | 1  | 32.85   |
| Popn    | 1  | 1.86    |
| Famsize | 1  | 21.04   |

#### Unusual Observations

| Obs. | Educ | Demand | Fit   | Stdev.Fit | Residual | St.Resid |
|------|------|--------|-------|-----------|----------|----------|
| 9    | 9.3  | 13.100 | 9.760 | 2.149     | 3.340    | 2.07R    |

R denotes an obs. with a large st. resid.

- 写出回归方程, 并把每个系数的标准误差写在该系数的下面, 可以加上括号。
- 找出  $R^2$  和残差标准差。
- 结果中写有“Unusual Observations”的部分表明第 9 个观测值远离预测值, 距离为 2.07 倍标准差。这是否说明第 9 个观察值是一个严重的离群值?

**12.36** 总结练习 12.35 的输出结果中关于  $F$  检验和各种  $t$  的结论。

**12.37** 练习 12.35 中数据的另一种分析是只用变量 Educ 和 Famsize 来预测需求。结果如下:

```
MTB > Regress 'Demand' 2 'Educ' 'Famsize'.
```

The regression equation is

Demand = - 19.2 + 7.79 Educ + 9.46 Famsize

| Predictor | Coef    | Stdev | t-ratio | p     |
|-----------|---------|-------|---------|-------|
| Constant  | - 19.17 | 45.87 | - 0.42  | 0.686 |
| Educ      | 7.793   | 2.490 | 3.13    | 0.012 |
| Famsize   | - 9.464 | 5.207 | - 1.82  | 0.103 |

$s = 2.939$        $R\text{-sq} = 94.2\%$        $R\text{-sq}(\text{adj}) = 92.9\%$

Analysis of Variance

| SOURCE     | DF | SS      | MS     | F     | p     |
|------------|----|---------|--------|-------|-------|
| Regression | 2  | 1268.48 | 634.24 | 73.43 | 0.000 |
| Error      | 9  | 77.73   | 8.64   |       |       |
| Total      | 11 | 1346.22 |        |       |       |

a. 找出这个简略模型的  $R^2$  的值。

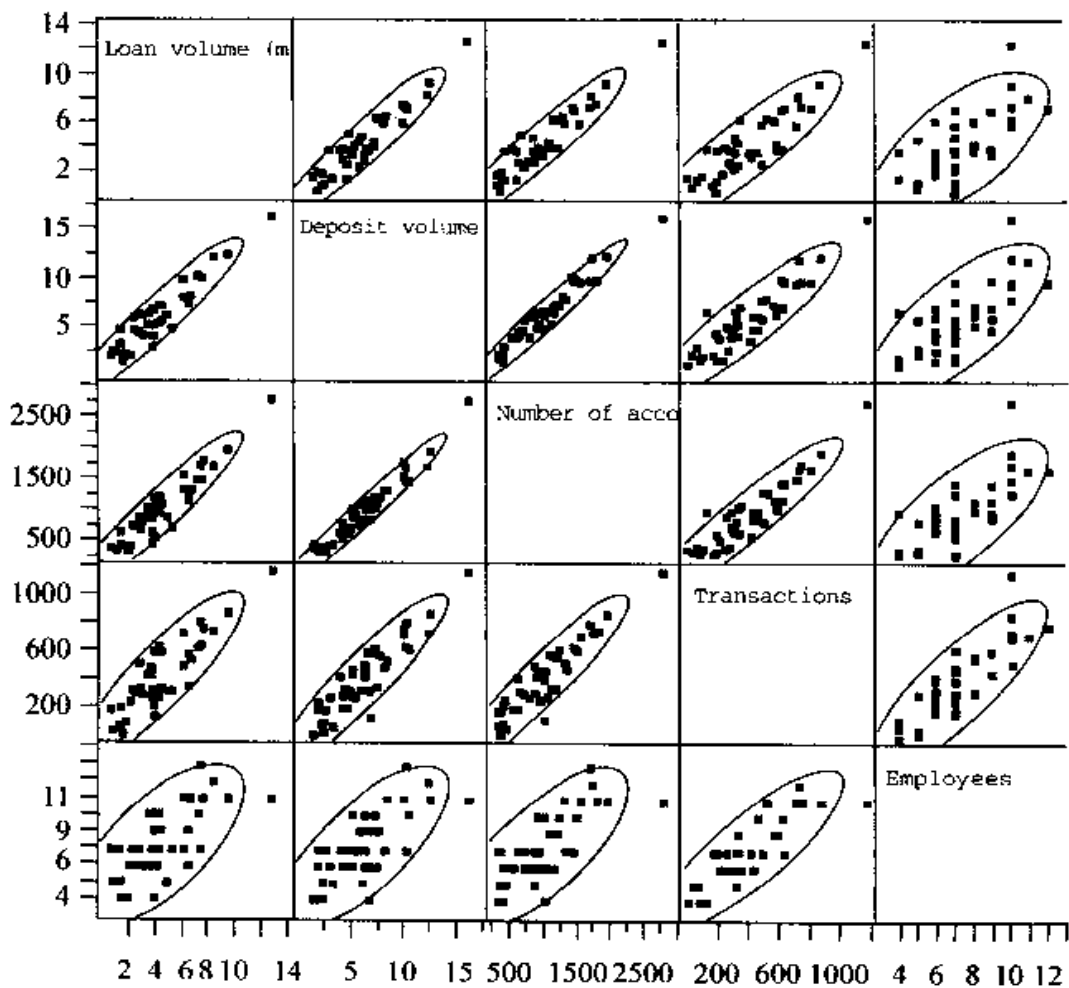
b. 取  $\alpha = 0.05$  检验假设“Income 和 Popn 的真实系数为 0”。结论是什么？

**12.38 (商业)** 银行支行办公室的业务之一是为小企业和个人提供有息借贷。作为考核支行经理业绩研究的一部分,一家银行从一些样本支行收集了关于当前借贷总量(Loan volume 因变量)、该支行开户的账号的储蓄总额(Deposit volume)以及这些账号的个数(accounts)、日平均业务数量(Transactions)和该支行雇员的总数(Employees)的数据。相关和散点阵如图中所示。

a. 哪一个自变量是借贷总量的最好的预测因子？

b. 是否有严重的共线性问题？

c. 是否存在有极端影响的点？





| Correlations              |                        |                           |                    |              |           |
|---------------------------|------------------------|---------------------------|--------------------|--------------|-----------|
| Variable                  | Loan volume (millions) | Deposit volume (millions) | Number of accounts | Transactions | Employees |
| Loan volume (millions)    | 1.0000                 | 0.9369                    | 0.9403             | 0.8766       | 0.6810    |
| Deposit volume (millions) | 0.9369                 | 1.0000                    | 0.9755             | 0.9144       | 0.7377    |
| Number of accounts        | 0.9403                 | 0.9755                    | 1.0000             | 0.9299       | 0.7487    |
| Transactions              | 0.8766                 | 0.9144                    | 0.9299             | 1.0000       | 0.8463    |
| Employees                 | 0.6810                 | 0.7377                    | 0.7487             | 0.8463       | 1.0000    |

**12.39** 用 JMP 对银行支行的数据建立了一个回归模型,一些结果如下。

- 用  $R^2$  的值计算总体  $F$  统计量。是否有清楚的证据表明该模型对于预测是有价值的? 取  $\alpha = 0.05$ 。
- 哪个因子已经被证实是具有独有预测价值的? 仍取  $\alpha = 0.05$ 。
- 解释你对前两部分的答案中的矛盾。

Response: Loan volume (millions)

| Summary of Fit            |          |
|---------------------------|----------|
| RSquare                   | 0.894477 |
| RSquare Adj               | 0.883369 |
| Root Mean Square Error    | 0.870612 |
| Mean of Response          | 4.383395 |
| Observations(or Sum Wgts) | 43       |

| Parameter Estimates       |           |           |         |         |
|---------------------------|-----------|-----------|---------|---------|
| Term                      | Estimate  | Std Error | t Ratio | Prob> t |
| Intercept                 | 0.2284381 | 0.6752    | 0.34    | 0.7370  |
| Deposit volume (millions) | 0.3222099 | 0.191048  | 1.69    | 0.0999  |
| Number of accounts        | 0.0025812 | 0.001314  | 1.96    | 0.0569  |
| Transactions              | 0.0010058 | 0.001878  | 0.54    | 0.5954  |
| Employees                 | -0.119898 | 0.130721  | -0.92   | 0.3648  |

**12.40** 另一个多元回归模型只用了储蓄总额和账号的数量作为自变量。结果如下。

- 省却业务数量和雇员数量两个变量是否使得  $R^2$  有严重的下降?
- 用  $R^2$  的值检验零假设“业务数量和雇员数量的系数为 0”。结论是什么?

Response: Loan volume (millions)

| Summary of Fit            |          |
|---------------------------|----------|
| RSquare                   | 0.892138 |
| RSquare Adj               | 0.886744 |
| Root Mean Square Error    | 0.857923 |
| Mean of Response          | 4.383395 |
| Observations(or Sum Wgts) | 43       |

| Parameter Estimates       |           |           |         |         |
|---------------------------|-----------|-----------|---------|---------|
| Term                      | Estimate  | Std Error | t Ratio | Prob> t |
| Intercept                 | -0.324812 | 0.290321  | -1.12   | 0.2699  |
| Deposit volume (millions) | 0.3227636 | 0.187509  | 1.72    | 0.0929  |
| Number of accounts        | 0.002684  | 0.001166  | 2.30    | 0.0266  |

**12.41 (工程)** 一家计算机软件公司的档案部经理要预测存档中等大小的计算机程序所需要的时间。他记录了 26 个程序的有关数据, 变量为:  $y$  = 所需要的天数,  $x_1$  = 子程序的个数,  $x_2$  = 每个子程序的平均行数,  $x_3 = x_1 x_2$ ,  $x_4 = x_2^2$ ,  $x_5 = x_1 x_2^2$ 。这些数据的部分输出结果如下:

#### Multiple Regression Analysis

Dependent variable: Y

| Table of Estimates |              |                |         |         |
|--------------------|--------------|----------------|---------|---------|
|                    | Estimate     | Standard Error | t Value | P Value |
| Constant           | -16.8198     | 11.631         | -1.45   | 0.1636  |
| X1                 | 1.47019      | 0.365944       | 4.02    | 0.0007  |
| X2                 | 0.994778     | 0.611441       | 1.63    | 0.1194  |
| @X1X2              | -0.0240071   | 0.0237565      | -1.01   | 0.3243  |
| @X2Sq              | -0.01031     | 0.007374       | -1.40   | 0.1774  |
| @X1X2Sq            | -0.000249574 | 0.000351779    | 0.71    | 0.4862  |

R-squared = 91.72 %

Adjusted R-squared = 89.65 %

Standard error of estimation = 3.39011

Durbin-Watson statistic = 2.12676

Mean absolute error = 2.4127

#### Analysis of Variance

| Source        | Sum of Squares | D.F. | Mean Square | F-Ratio | P Value |
|---------------|----------------|------|-------------|---------|---------|
| Model         | 2546.03        | 5    | 509.205     | 44.31   | 0.0000  |
| Error         | 229.857        | 20   | 11.4929     |         |         |
| Total (corr.) | 2775.88        | 25   |             |         |         |

a. 写出多元回归模型, 并找出残差标准差。

b.  $x_3$  有显著的、独有的预测价值吗?

**12.42** 对于练习 12.41 中的数据拟合了模型  $y = \beta_0 + \beta_1 x_1 + \epsilon$ 。部分输出为:

| Multiple Regression Analysis |           |                |         |         |
|------------------------------|-----------|----------------|---------|---------|
| Dependent variable: Y        |           |                |         |         |
| Table of Estimates           |           |                |         |         |
|                              | Estimate  | Standard Error | t Value | P Value |
| Constant                     | 0.840085  | 3.43375        | 0.24    | 0.8089  |
| X1                           | 1.01583   | 0.0792925      | 12.81   | 0.0000  |
| X2                           | 0.0558262 | 0.0515066      | 1.08    | 0.2897  |

R-squared = 90.64 %

Adjusted R-squared = 89.83 %

Standard error of estimation = 3.36066

Durbin-Watson statistic = 2.2053

Mean absolute error = 2.57584

| Analysis of Variance |                |      |             |         |         |
|----------------------|----------------|------|-------------|---------|---------|
| Source               | Sum of Squares | D.F. | Mean Square | F-Ratio | P Value |
| Model                | 2516.12        | 2    | 1258.06     | 111.39  | 0.0000  |
| Error                | 259.763        | 23   | 11.294      |         |         |
| Total (corr.)        | 2775.88        | 25   |             |         |         |

a. 写出估计后的完全模型和简略模型。

b. 在水平  $\alpha = 0.05$  上, 增加  $x_3, x_4$  和  $x_5$  对  $R^2$  有显著的改进吗? 近似求出这个检验的  $p$  值。

**12.43 (商业)** 一家小型方便食品连锁店做了一项回归分析, 以解释 16 个分店的销售总额的变异。研究中的变量如下:

销售额(Sales): 分店的日平均销售额, 以千美元为单位。

面积(Size): 分店面积, 以千平方英尺为单位。

停车位(Parking): 分店旁边的自由停车位个数。

收入(Income): 分店所在邮政编码地区户平均收入的估计。

从一个回归的程序(stataQuest)得到的输出如下:

```
.regress Sale Size Parking Income
```

| Source   | SS         | df | MS         | Number of obs | = | 16     |
|----------|------------|----|------------|---------------|---|--------|
| Model    | 27.1296056 | 3  | 9.04320188 | F(3,12)       | = | 15.16  |
| Residual | 7.15923792 | 12 | 0.59660316 | Prob > F      | = | 0.0002 |
| Total    | 34.2888436 | 15 | 2.2859229  | R-square      | = | 0.7912 |
|          |            |    |            | Adj R-square  | = | 0.7390 |
|          |            |    |            | Root MSE      | = | 0.7724 |

| Sales   | Coef.     | Std. Err. | t     | P> T  | [95% Conf. Interval] |
|---------|-----------|-----------|-------|-------|----------------------|
| Size    | 2.547936  | 1.200827  | 2.122 | 0.055 | -0.0684405 5.164313  |
| Parking | 0.2202793 | 0.1553877 | 1.418 | 0.182 | -0.1182814 0.5588401 |
| Income  | 0.5893221 | 0.1780576 | 3.310 | 0.006 | 0.2013679 0.9772763  |
| _cons   | 0.872716  | 1.945615  | 0.449 | 0.662 | -3.366415 5.111847   |

```
.correlate Sales Size Parking Income
(obs = 16)
```

|         | Sales  | Size   | Parking | Income |
|---------|--------|--------|---------|--------|
| Sales   | 1.0000 |        |         |        |
| Size    | 0.7415 | 1.0000 |         |        |
| Parking | 0.6568 | 0.6565 | 1.0000  |        |
| Income  | 0.7148 | 0.4033 | 0.3241  | 1.0000 |

- 写出回归方程,指出各系数的标准误差。
- 小心解释每一个系数。
- 找出  $R^2$  和残差标准差。
- 在这项研究中,有严重的共线性问题吗?

**12.44** 总结练习 12.43 中  $F$  检验和  $t$  检验的结果。

**12.45 (农业)** 一家牲畜饲料添加剂的生产商要研究饲养肉牛使其达到上市重量所需要的天数。他买了 18 头年龄、重量基本一致的阉割公牛,并进行饲养。每头牛按照一个指定的蛋白质含量、抗生素浓度和饲料添加剂的百分比的组合来饲养。数据如下:

|               |    |    |    |    |    |    |    |    |    |
|---------------|----|----|----|----|----|----|----|----|----|
| 牛:            | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
| 蛋白质(Protein): | 10 | 10 | 10 | 10 | 10 | 10 | 15 | 15 | 15 |
| 抗生素(Antibio): | 1  | 1  | 1  | 2  | 2  | 2  | 1  | 1  | 1  |

|               |    |    |    |    |    |    |    |    |    |
|---------------|----|----|----|----|----|----|----|----|----|
| 添加剂(Supplem): | 3  | 5  | 7  | 3  | 5  | 7  | 3  | 5  | 7  |
| 时间(Time):     | 88 | 82 | 81 | 82 | 83 | 75 | 80 | 80 | 75 |
| 牛:            | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 蛋白质(Protein): | 15 | 15 | 15 | 20 | 20 | 20 | 20 | 20 | 20 |
| 抗生素(Antibio): | 2  | 2  | 2  | 1  | 1  | 1  | 2  | 2  | 2  |
| 添加剂(Supplem): | 3  | 5  | 7  | 3  | 5  | 7  | 3  | 5  | 7  |
| 时间(Time):     | 77 | 76 | 72 | 79 | 74 | 75 | 74 | 70 | 69 |

从 Systat 得到的计算机输出如下:

#### CORRELATIONS (PEARSON)

|         | TIME    | PROTEIN | ANTIBIO |
|---------|---------|---------|---------|
| PROTEIN | -0.7111 |         |         |
| ANTIBIO | -0.4180 | 0.0000  |         |
| SUPPLEM | -0.4693 | 0.0000  | -0.0000 |

CASES INCLUDED 18 MISSING CASES 0

#### UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF TIME

##### PREDICTOR

| VARIABLES | COEFFICIENT | STD ERROR | STUDENT'S T | P      | VIF |
|-----------|-------------|-----------|-------------|--------|-----|
| CONSTANT  | 102.708     | 2.31037   | 44.46       | 0.0000 |     |
| PROTEIN   | -0.83333    | 0.09870   | -8.44       | 0.0000 | 1.0 |
| ANTIBIO   | -4.00000    | 0.80589   | -4.96       | 0.0002 | 1.0 |
| SUPPLEM   | -1.37500    | 0.24675   | -5.57       | 0.0001 | 1.0 |

|                    |        |                          |         |
|--------------------|--------|--------------------------|---------|
| R-SQUARED          | 0.9007 | RESID. MEAN SQUARE (MSE) | 2.92261 |
| ADJUSTED R-SQUARED | 0.8794 | STANDARD DEVIATION       | 1.70956 |

| SOURCE     | DF | SS      | MS      | F     | P      |
|------------|----|---------|---------|-------|--------|
| REGRESSION | 3  | 371.083 | 123.694 | 42.32 | 0.0000 |
| RESIDUAL   | 14 | 40.9166 | 2.92261 |       |        |

TOTAL            17        412.000

PREDICTED/FITTED VALUES OF TIME

|                       |        |                    |        |
|-----------------------|--------|--------------------|--------|
| LOWER PREDICTED BOUND | 73.566 | LOWER FITTED BOUND | 76.469 |
| PREDICTED VALUE       | 77.333 | FITTED VALUE       | 77.333 |
| UPPER PREDICTED BOUND | 81.100 | UPPER FITTED BOUND | 78.197 |
| SE (PREDICTED VALUE)  | 1.7564 | SE (FITTED VALUE)  | 0.4029 |

UNUSUALNESS (LEVERAGE)    0.0556  
 PERCENT COVERAGE            95.0  
 CORRESPONDING T            2.14

PREDICTOR VALUES: PROTEIN = 15.000, ANTIBIO = 1.5000, SUPPLEM = 5.0000

- 写出回归方程。
- 找出标准差。
- 找出  $R^2$  的值。
- 在这些数据中有严重的共线性问题吗？

12.46 参见练习 12.45。

- 预测按 15% 蛋白质、1.5% 抗生素浓度和 5% 添加剂饲养的牛的饲养时间。
- 这些自变量的值代表了一个从数据进行的外推吗？
- 给出(a)中预测的平均时间的 95% 置信区间。

12.47 对于练习 12.45 中的数据, 还用只含有蛋白质含量作为自变量的回归模型进行了分析, 结果如下:

UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF TIME

PREDICTOR

| VARIABLES | COEFFICIENT | STD ERROR | STUDENT'S T | P      |
|-----------|-------------|-----------|-------------|--------|
| CONSTANT  | 89.8333     | 3.20219   | 28.05       | 0.0000 |
| PROTEIN   | - 0.83333   | 0.20598   | - 4.05      | 0.0009 |

|                    |        |                          |         |
|--------------------|--------|--------------------------|---------|
| R-SQUARED          | 0.5057 | RESID. MEAN SQUARE (MSE) | 12.7291 |
| ADJUSTED R-SQUARED | 0.4748 | STANDARD DEVIATION       | 3.56779 |

| SOURCE     | DF | SS      | MS      | F     | P      |
|------------|----|---------|---------|-------|--------|
| REGRESSION | 1  | 208.333 | 208.333 | 16.37 | 0.0009 |
| RESIDUAL   | 16 | 203.666 | 12.7291 |       |        |
| TOTAL      | 17 | 412.000 |         |       |        |

a. 写出回归方程。

b. 找出  $R^2$ 。

c. 在  $\alpha = 0.05$  的水平上检验零假设“ANTIBIO 和 SUPPLEM 的系数为 0”。

**12.48 (人力资源)** 一起性别歧视的诉讼宣称某个小型学院在工资上对女性员工有歧视。一项回归研究中考虑了如下的变量:

Salary: 每年的基本工资(以千美元为单位)。

Senior: 在学院的资历(年)。

Sex: 男性为 1, 女性为 0。

RankD1: 正教授为 1, 其他为 0。

RankD2: 副教授为 1, 其他为 0。

RankD3: 助理教授为 1, 其他为 0。

Doct: 有博士学位者为 1, 其他为 0。

注意, 对于讲师和助教, 以上三个 RankD 变量的值均为 0。用 Excel 得到的计算机输出结果如下。

|    | A                     | B      | C | D | E | F |
|----|-----------------------|--------|---|---|---|---|
| 1  |                       |        |   |   |   |   |
| 2  | Regression Statistics |        |   |   |   |   |
| 3  |                       |        |   |   |   |   |
| 4  | Multiple R            | 0.9716 |   |   |   |   |
| 5  | R Square              | 0.944  |   |   |   |   |
| 6  | Adjusted R Square     | 0.9294 |   |   |   |   |
| 7  | Standard Error        | 2.3375 |   |   |   |   |
| 8  | Observations          | 30     |   |   |   |   |
| 9  |                       |        |   |   |   |   |
| 10 |                       |        |   |   |   |   |

续表

|    | A          | B            | C              | D       | E       | F              |
|----|------------|--------------|----------------|---------|---------|----------------|
| 11 | ANOVA      |              |                |         |         |                |
| 12 |            | df           | SS             | MS      | F       | Significance F |
| 13 | Regression | 6            | 2119.347       | 353.225 | 64.646  | 0.000          |
| 14 | Residual   | 23           | 125.672        | 5.464   |         |                |
| 15 | Total      | 29           | 2245.019       |         |         |                |
| 16 |            |              |                |         |         |                |
| 17 |            |              |                |         |         |                |
| 18 |            | Coefficients | Standard Error | t Stat  | P-value |                |
| 19 | Intercept  | 18.6784      | 1.3788         | 13.5470 | 0.0000  |                |
| 20 | Senior     | 0.5420       | 0.0762         | 7.1176  | 0.0000  |                |
| 21 | Sex        | 1.2074       | 0.0649         | 1.1339  | 0.2685  |                |
| 22 | RankD1     | 8.7779       | 1.9380         | 4.5293  | 0.0002  |                |
| 23 | RankD2     | 4.4211       | 1.7797         | 2.4842  | 0.0207  |                |
| 24 | RankD3     | 2.7165       | 1.4239         | 1.9079  | 0.0690  |                |
| 25 | Doct       | 0.9225       | 1.2589         | 0.7328  | 0.4711  |                |

- 写出回归方程。
- 变量 Sex 的系数的解释是什么?
- 变量 RankD1 的系数的解释是什么?

12.49 参见练习 12.48。

- 检验假设“Sex 的真实系数大于 0”，取  $\alpha = 0.05$ 。
- 以上检验的结论对于这项性别歧视的诉讼表明了什么?

12.50

- 找出练习 12.48 中 F 统计量的值。
- 这个统计量检验什么样的零假设?
- 在  $\alpha = 0.01$  的水平上拒绝这个零假设吗? 这个检验的  $p$  值有多大?

12.51 练习 12.48 中数据的另一个回归模型在自变量表中省却了 Sex 和 Doct, 结果如下:

Regression Statistics

Multiple R 0.9697

R Square 0.9403



|                   |        |
|-------------------|--------|
| Adjusted R Square | 0.9307 |
| Standard Error    | 2.3160 |
| Observations      | 30     |

## ANOVA

|            | df | SS       | MS      | F      | Significance F |
|------------|----|----------|---------|--------|----------------|
| Regression | 4  | 2110.925 | 527.731 | 98.389 | 0.0000         |
| Residual   | 25 | 134.093  | 5.364   |        |                |
| Total      | 29 | 2245.019 |         |        |                |

|           | Coefficients | Standard Error | t Stat  | P-value |
|-----------|--------------|----------------|---------|---------|
| Intercept | 19.7113      | 1.0776         | 18.2913 | 0.0000  |
| Senior    | 0.5572       | 0.0744         | 7.4893  | 0.0000  |
| RankD1    | 9.2414       | 1.8214         | 5.0738  | 0.0000  |
| RankD2    | 5.1050       | 1.5875         | 3.2158  | 0.0036  |
| RankD3    | 3.2243       | 1.3204         | 2.4418  | 0.0220  |

a. 找出简略模型的  $R^2$  的值。

b. 检验零假设“Sex 和 Doct 的真实系数为 0”，取  $\alpha = 0.01$ 。

**12.52** 对信息系统管理员做了一项调查,以预测某个大城市区域内初级程序员和分析员的年薪。管理员写明了他们公司里付给初级程序员和分析员的标准年薪,公司信息处理部门中的雇员数,公司的毛利润率(以美分/每美元销售为单位),以及公司的信息处理费用占总管理费用的百分比。数据存放在 Web 站点数据集的名为“EX1252.DAT”的文件中,第一列为年薪,第二列为雇员数,第三列为利润率,第四列为信息处理费用。

a. 以年薪作为因变量,其他三个变量作为预测因子建立一个多元回归方程。

b. 有无依据表明这三个预测因子至少对于年薪有一定的预测价值? 求出适当的检验的  $p$  值。

c. 如果有的话,哪一个自变量作为最后一个进入方程的预测因子有显著的预测作用? 取  $\alpha = 0.05$ 。

**12.53**

a. 求出练习 12.52 中的回归模型的决定系数  $R^2$ 。

b. 只用雇员数作为自变量建立回归模型,并求出这个回归模型的决定系数。

c. 手工检验零假设“在已经给定雇员数的前提下,加入利润率和信息处理费用这两个变量并不产生附加的预测价值”,取  $\alpha = 0.05$ 。从这个检验你能得到什么结论? 预测作用? 取  $\alpha = 0.05$ 。

**12.54** 求练习 12.52 中每两个预测因子的相关系数。数据中有严重的共线性问题吗？

**12.55 (政府)** 一个政府机构除了付给研究项目的直接费用外,还付给研究承包人一些间接的费用。虽然间接费用随研究合同而有所差异,但都占总合同经费的一个不小的比例。该机构的一个特别工作组收集了关于间接费用(与直接费用的比值)、承包人雇员的个数、合同的大小与承包人年收入的百分比,以及个人消费与直接费用的百分比的数据。这四个变量按照上述的顺序存放在 Web 站点数据集的名为“EX1255.DAT”的文件中。

- 求每两个变量的相关系数,数据中有严重的共线性问题吗？
- 画出间接费用与其他每一个变量的散点图,找出可能的强影响点。
- 用所有数据,包括任何潜在的离群值建立一个回归方程(间接费用作为因变量)。
- 删去潜在的离群值,并修改回归方程,各个斜率变化了多少？

**12.56** 考虑练习 12.55 中删去了离群值的回归模型。

- 求  $F$  统计量。该统计量检验什么零假设？基于  $F$  检验我们能得到什么结论？
- 对于每个自变量,求出  $t$  统计量的值。基于这些  $t$  检验,我们能得到什么结论？

**12.57** 用练习 12.55 中删去了离群值的回归模型,预测当承包人有 500 个雇员,合同是承包人收入的 2.5%,个人消费是直接费用的 55% 时的间接费用。求 95% 的预测区间。在这种情形下,间接费用等于直接费用的 88.9% 是不合理的吗？

**12.58 (商业)** 一家快速增长的计算机商店的老板想用四个解释变量即显示的字幕的条数,显示的总长度,当前 IBM 兼容机的用户群体大小,以及当前苹果兼容机的用户群体大小来解释每两周计算机软件的销售量的增长。数据按时间顺序存放在 Web 站点数据集的名为“EX1258.DAT”的文件中,第一列为销售量,第二列为字幕数,第三列为总长度,第四列为 IBM 群体大小,第五列为苹果群体大小。

- 在进行计算以前,考虑当时的经济状况,并说明你期望每个偏斜率是什么符号？
- 用销售量作为因变量,其他变量作为自变量建立一个多元回归方程。每个偏斜率都有你期望的符号吗？
- 构造字幕这个变量的系数的 95% 的置信区间。计算机输出结果应当包含该系数的标准误差。这个置信区间包含 0 吗？

**12.59**

- 在练习 12.58 的回归模型中,在通常的  $\alpha$  水平上,零假设“没有一个变量对

于预测是有价值的”会被拒绝吗?

- b. 根据  $t$  检验的结果, 如果有的话, 在给定所有其他预测因子后, 哪个预测因子在  $\alpha = 0.05$  的水平上增加了显著的预测价值?

**12.60** 用练习 12.58 中的数据求所有两个变量之间的相关系数。数据中的共线性问题严重吗?

**12.61** 比较练习 12.28 中回归模型的决定系数  $R^2$  与销售量和字幕条数相关系数的平方。为检验零假设“给定字幕条数以后, 总长度、IBM 群体大小和苹果群体大小不增加模型的预测价值”, 计算相应的  $F$  统计量。在  $\alpha = 0.01$  的水平上会拒绝这个假设吗?

**12.62 (商业)** 一商家的市场研究部经理做了一项调查, 以研究是哪些因子决定从顾客手中接到的定单的大小。从公司计算机储存的销售数据中, 经理得到了 180 个邮政编码地区的平均定单的大小。一位业余助手找到了最近普查的信息, 得到了每个地区人均收入(Income)、接受正式教育的平均年数(Educn)以及现房的中位价格(HousePr)的数据。(这位助手未能找到其中两个邮政编码地区的房价数据, 以 0 记之。)这位经理还想知道气候是否也对定单大小有一定影响, 所以也收集了关于冬季和夏季日均最高温度(用 Wint Temp 和 Summ Temp 表示)的数据。

这位经理求助于你, 要求你帮助分析数据。下面的结果只是一次初步尝试。经理想知道是否有证据表明温度变量不是无足轻重, 也想知道还有哪些变量是有用的。还有, 把缺失数据以 0 代替是对还是错, 或者说是否歪曲了拟合的结果? 请对这些输出结果以及你所做的其他分析做出基本的、不太艰涩的解释。

```
MTB > name c1 'AvgOrder' c2 'Income' c3 'Educn' &
CONT> c4 'HousePr' c5 'WintTemp' c6 'SummTemp'
MTB > correlations of c1-c6
```

|          | AvgOrder | Income | Educn | HousePr | WintTemp |  |
|----------|----------|--------|-------|---------|----------|--|
| Income   | 0.205    |        |       |         |          |  |
| Educn    | 0.171    | 0.913  |       |         |          |  |
| HousePr  | 0.269    | 0.616  | 0.561 |         |          |  |
| WintTemp | -0.134   | -0.098 | 0.014 | 0.066   |          |  |
| SummTemp | -0.068   | -0.115 | 0.005 | 0.018   | 0.481    |  |

```
MTB > regress c1 on 5 variables in c2-c6
```

The regression equation is

$$\begin{aligned}\text{AvgOrder} = & 36.2 + 0.078 \text{ Income} - 0.019 \text{ Educn} \\ & + 0.0605 \text{ HousePr} - 0.223 \text{ WintTemp} + 0.006 \text{ SummTemp}\end{aligned}$$

| Predictor | Coef    | Stdev   | t-ratio | p     |
|-----------|---------|---------|---------|-------|
| Constant  | 36.18   | 12.37   | 2.92    | 0.004 |
| Income    | 0.0780  | 0.4190  | 0.19    | 0.853 |
| Educn     | -0.0189 | 0.5180  | -0.04   | 0.971 |
| HousePr   | 0.06049 | 0.02161 | 2.80    | 0.006 |
| WintTemp  | -0.2231 | 0.1259  | -1.77   | 0.078 |
| SummTemp  | 0.0063  | 0.1646  | 0.04    | 0.969 |

s = 4.747      R-sq = 9.6 %      R-sq(adj) = 7.0 %

#### Analysis of Variance

| SOURCE     | DF  | SS      | MS    | F    | p     |
|------------|-----|---------|-------|------|-------|
| Regression | 5   | 417.63  | 83.53 | 3.71 | 0.003 |
| Error      | 174 | 3920.31 | 22.53 |      |       |
| Total      | 179 | 4337.94 |       |      |       |

| SOURCE   | DF | SEQ SS |
|----------|----|--------|
| Income   | 1  | 182.94 |
| Educn    | 1  | 7.18   |
| HousePr  | 1  | 142.63 |
| WintTemp | 1  | 84.84  |
| SummTemp | 1  | 0.03   |

#### Unusual Observations

| Obs. | Income | AvgOrder | Fit    | Stdev.Fit | Residual | St.Resid |
|------|--------|----------|--------|-----------|----------|----------|
| 25   | 17.1   | 23.570   | 36.555 | 0.632     | -12.985  | -2.76R   |
| 78   | 11.9   | 24.990   | 34.950 | 0.793     | -9.960   | -2.13R   |
| 83   | 13.4   | 36.750   | 29.136 | 2.610     | 7.614    | 1.92X    |
| 87   | 14.3   | 45.970   | 35.918 | 0.463     | 10.052   | 2.13R    |
| 111  | 11.1   | 21.720   | 33.570 | 0.802     | -11.850  | -2.53R   |
| 113  | 10.4   | 43.500   | 33.469 | 0.817     | 10.031   | 2.15R    |
| 143  | 16.1   | 20.350   | 27.915 | 3.000     | -7.565   | -2.06RX  |
| 149  | 13.2   | 44.970   | 35.369 | 0.604     | 9.601    | 2.04R    |
| 169  | 13.5   | 44.650   | 34.361 | 0.660     | 10.289   | 2.19R    |
| 180  | 13.7   | 23.050   | 34.929 | 0.469     | -11.879  | -2.51R   |

R denotes an obs. with a large st.resid.

X denotes an obs. whose X value gives it large influence.

**12.63** 下面的表给出了一项试验中 12 名男性充血性心力衰竭患者的人口统计的数据。

| 心力衰竭患者的人口统计数据<br>(NYHA Ⅲ类或Ⅳ类) |           |      |            |            |                                 |                    |
|-------------------------------|-----------|------|------------|------------|---------------------------------|--------------------|
| 患者                            | 年龄<br>(岁) | 疾病历史 | 身高<br>(cm) | 体重<br>(kg) | 基本                              |                    |
|                               |           |      |            |            | 心脏指数<br>(L/min/m <sup>2</sup> ) | 肺毛细血管楔压<br>(mm Hg) |
| 01                            | 67        | 5yr  | 172.0      | 57.0       | 1.6                             | 40                 |
| 02                            | 45        | 2yr  | 170.0      | 67.0       | 2.4                             | 25                 |
| 03                            | 59        | 8yr  | 172.7      | 102.0      | 2.2                             | 39                 |
| 04                            | 63        | 1yr  | 175.3      | 74.9       | 1.7                             | 39                 |
| 05                            | 55        | 1yr  | 172.7      | 92.0       | 2.3                             | 34                 |
| 06                            | 65        | 1yr  | 178.0      | 90.0       | 1.6                             | 36                 |
| 07                            | 62        | 2yr  | 163.0      | 67.0       | 1.4                             | 36                 |
| 08                            | 60        | 1yr  | 182.5      | 72.0       | 2.2                             | 17                 |
| 09                            | 72        | 2yr  | 168.0      | 71.0       | 1.3                             | 37                 |
| 10                            | 44        | 3mo  | 163.0      | 68.0       | 2.4                             | 28                 |
| 11                            | 63        | 5yr  | 172.0      | 82.0       | 2.1                             | 38                 |
| 12                            | 63        | 1yr  | 163.0      | 64.0       | 1.1                             | 36                 |

\* :yr 表示年,mo 表示月

a. 对每一个变量用箱线图概括这些数据。

b. 用散点图显示(1)(心脏指数(Cardiac Index, CI), 年龄)以及(肺毛细血管楔压(pulmonary capillary wedge pressure, PCWP), 年龄)(2)(CI, 疾病历史(disease duation)), 以及(PCWP, 疾病历史)。是否有证据表明年龄和 CI 或 PCWP 是相关的? 疾病历史与 CI 或 PCWP 之间的相关性如何?

**12.64** 用练习 12.63 中的数据拟合了几个多元回归模型;  $y_1 = \text{CI}$ ,  $y_2 = \text{PCWP}$ ,  $x_1 = \text{年龄}$ ,  $x_2 = \text{疾病历史}$ 。

a.  $y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

b.  $y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$

c.  $y_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

d.  $y_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$

#### REGRESSION ANALYSIS, MODEL I

Dependent Variable: CI

## Analysis of Variance

| Source  | DF | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|---------|----|-------------------|----------------|---------|--------|
| Model   | 2  | 1.56955           | 0.78478        | 9.298   | 0.0065 |
| Error   | 9  | 0.75961           | 0.08440        |         |        |
| C Total | 11 | 2.32917           |                |         |        |

|          |          |            |        |
|----------|----------|------------|--------|
| Root MSE | 0.29052  | R - square | 0.6739 |
| Dep Mean | 1.85833  | Adj R - sq | 0.6014 |
| C.V.     | 15.63333 |            |        |

## Parameter Estimates

| Variable | DF | Parameter<br>Estimate | Standard<br>Error | T for H0:<br>Parameter = 0 | Prob >  T |
|----------|----|-----------------------|-------------------|----------------------------|-----------|
| INTERCEP | 1  | 4.475622              | 0.63976685        | 6.996                      | 0.0001    |
| AGE      | 1  | -0.046203             | 0.01083529        | -4.264                     | 0.0021    |
| DURATION | 1  | 0.060395              | 0.03852829        | 1.568                      | 0.1514    |

## REGRESSION ANALYSIS, MODEL II

Dependent Variable: CI

## Analysis of Variance

| Source  | DF | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|---------|----|-------------------|----------------|---------|--------|
| Model   | 3  | 1.57161           | 0.52387        | 5.532   | 0.0237 |
| Error   | 8  | 0.75755           | 0.09469        |         |        |
| C Total | 11 | 2.32917           |                |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 0.30772  | R-square | 0.6748 |
| Dep Mean | 1.85833  | Adj R-sq | 0.5528 |
| C.V.     | 16.55915 |          |        |

## Parameter Estimates

| Variable | DF | Parameter<br>Estimate | Standard<br>Error | T for H0:<br>Parameter = 0 | Prob >  T |
|----------|----|-----------------------|-------------------|----------------------------|-----------|
| INTERCEP | 1  | 4.599307              | 1.07814691        | 4.266                      | 0.0027    |

|           |   |            |            |         |        |
|-----------|---|------------|------------|---------|--------|
| AGE       | 1 | - 0.048340 | 0.01848097 | - 2.616 | 0.0309 |
| DURATION  | 1 | - 0.022410 | 0.56287924 | - 0.040 | 0.9692 |
| AGE _ DUR | 1 | 0.001376   | 0.00932590 | 0.147   | 0.8864 |

| Variable  |    |           | Variable  |
|-----------|----|-----------|-----------|
| Variable  | DF |           | Label     |
| INTERCEP  | 1  |           | Intercept |
| AGE       | 1  |           |           |
| DURATION  | 1  |           |           |
| AGE _ DUR | 1  | AGE TIMES | DURATION  |

#### REGRESSION ANALYSIS, MODEL III

Dependent Variable: PCWP

#### Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 2  | 221.88101      | 110.94051   | 3.259   | 0.0862 |
| Error   | 9  | 306.36899      | 34.04100    |         |        |
| C Total | 11 | 528.25000      |             |         |        |

|          |          |            |        |
|----------|----------|------------|--------|
| Root MSE | 5.83447  | R-square   | 0.4200 |
| Dep Mean | 33.75000 | Adj R - sq | 0.2911 |
| C.V.     | 17.28731 |            |        |

#### Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob>  T |
|----------|----|--------------------|----------------|----------------------------|----------|
| INTERCEP | 1  | 7.298786           | 12.84835977    | 0.568                      | 0.5839   |
| AGE      | 1  | 0.400475           | 0.21760372     | 1.840                      | 0.0989   |
| DURATION | 1  | 1.021327           | 0.77375900     | 1.320                      | 0.2194   |

#### REGRESSION ANALYSIS, MODEL IV

Dependent Variable: PCWP

#### Analysis of Variance

| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 3  | 228.56515      | 76.18838    | 2.034   | 0.1878 |
| Error   | 8  | 299.68485      | 37.46061    |         |        |
| C Total | 11 | 528.25000      |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 6.12051  | R-square | 0.4327 |
| Dep Mean | 33.75000 | Adj R-sq | 0.2199 |
| C.V.     | 18.13484 |          |        |

## Parameter Estimates

| Variable  | DF | Parameter Estimate | Standard Error | T for H0: Parameter = 0 | Prob> T |
|-----------|----|--------------------|----------------|-------------------------|---------|
| INTERCEP  | 1  | 14.344026          | 21.44389171    | 0.669                   | 0.5224  |
| AGE       | 1  | 0.278775           | 0.36757883     | 0.758                   | 0.4700  |
| DURATION  | 1  | -3.695301          | 11.19543293    | -0.330                  | 0.7498  |
| AGE _ DUR | 1  | 0.078352           | 0.18548824     | 0.422                   | 0.6838  |

| Variable  | DF | Variable Label     |
|-----------|----|--------------------|
| INTERCEP  | 1  | Intercept          |
| AGE       | 1  |                    |
| DURATION  | 1  |                    |
| AGE _ DUR | 1  | AGE TIMES DURATION |

哪一个模型提供了对心脏指数数据的最好的拟合？哪一个模型提供了对肺毛细血管楔压数据的最好的拟合？这些分析是否证实了你在练习 12.63 中的结论？为什么？

下面的所有练习参见本章开始所介绍的案例。

**12.65** 数据的原始分析包括画出设计因子对表现变量  $avtem$  和  $logv$  的图，以便看看哪一个设计因子与表现变量有关。更进一步，这些图还表明表现变量和设计因子之间的关系是线性的还是需要更高阶的模型。

a. 作各个设计因子对表现变量的图，即画出  $(IT, avtem)$ ,  $(QW, avtem)$  和  $(VS, avtem)$  的图，以及  $(IT, logv)$ ,  $(QW, logv)$  和  $(VS, logv)$  的图。

b. 考察  $avtem$  的三张图，如果它与某个设计因子有关系的话，描述这个关系。

c. 考察  $logv$  的三张图，如果它与某个设计因子有关系的话，描述这个关系。

**12.66** 考察散点图以后，我们要考察几个模型，以确定哪个模型对于  $avtem$



数据提供了总体上最好的拟合, 并且没有拟合过头, 即没有在模型中放入太多的项。对 avtem 的数据拟合以下模型。

$$\text{模型 1: } \text{avtem} = \beta_0 + \beta_1 \text{IT} + \beta_2 \text{QW} + \beta_3 \text{VS} + \epsilon$$

$$\text{模型 2: } \text{avtem} = \beta_0 + \beta_1 \text{IT} + \beta_2 \text{QW} + \beta_3 \text{VS} + \beta_4 \text{I2} + \beta_5 \text{Q2} + \beta_6 \text{V2} + \epsilon$$

$$\text{模型 3: } \text{avtem} = \beta_0 + \beta_1 \text{IT} + \beta_2 \text{QW} + \beta_3 \text{VS} + \beta_4 \text{IT} * \text{QW} + \beta_5 \text{IT} * \text{VS} + \beta_6 \text{QW} * \text{VS} + \epsilon$$

$$\text{模型 4: } \text{avtem} = \beta_0 + \beta_1 \text{IT} + \beta_2 \text{QW} + \beta_3 \text{VS} + \beta_4 \text{I2} + \beta_5 \text{Q2} + \beta_6 \text{V2} + \beta_7 \text{IT} * \text{QW} + \beta_8 \text{IT} * \text{VS} + \beta_9 \text{QW} * \text{VS} + \epsilon$$

- a. 基于以上 4 个模型的  $R^2$  的值, 你会挑选哪一个作为对数据的“最好”的拟合?
- b. 用  $\alpha = 0.05$  检验假设“模型 2 与模型 1 没有显著差别”。
- c. 用  $\alpha = 0.05$  检验假设“模型 3 与模型 1 没有显著差别”。
- d. 用  $\alpha = 0.05$  检验假设“模型 4 与模型 3 没有显著差别”。
- e. 用  $\alpha = 0.05$  检验假设“模型 4 与模型 2 没有显著差别”。
- f. 通过散点图和(a)~(e)的结果, 你向工程师推荐哪个模型? 解释你挑选该模型的理由。

**12.67** 考察散点图以后, 我们要考察几个模型, 以确定哪个模型对于 logv 数据提供了总体上最好的拟合, 并且没有拟合过头, 即没有在模型中放入太多的项。对 logv 的数据拟合以下模型。

$$\text{模型 1: } \log v = \beta_0 + \beta_1 \text{IT} + \beta_2 \text{QW} + \beta_3 \text{VS} + \epsilon$$

$$\text{模型 2: } \log v = \beta_0 + \beta_1 \text{IT} + \beta_2 \text{QW} + \beta_3 \text{VS} + \beta_4 \text{I2} + \beta_5 \text{Q2} + \beta_6 \text{V2} + \epsilon$$

$$\text{模型 3: } \log v = \beta_0 + \beta_1 \text{IT} + \beta_2 \text{QW} + \beta_3 \text{VS} + \beta_4 \text{IT} * \text{QW} + \beta_5 \text{IT} * \text{VS} + \beta_6 \text{QW} * \text{VS} + \epsilon$$

$$\text{模型 4: } \log v = \beta_0 + \beta_1 \text{IT} + \beta_2 \text{QW} + \beta_3 \text{VS} + \beta_4 \text{I2} + \beta_5 \text{Q2} + \beta_6 \text{V2} + \beta_7 \text{IT} * \text{QW} + \beta_8 \text{IT} * \text{VS} + \beta_9 \text{QW} * \text{VS} + \epsilon$$

- a. 基于以上 4 个模型的  $R^2$  的值, 你会挑选哪一个作为对数据的“最好”的拟合?
- b. 用  $\alpha = 0.05$  检验假设“模型 2 与模型 1 没有显著差别”。
- c. 用  $\alpha = 0.05$  检验假设“模型 3 与模型 1 没有显著差别”。
- d. 用  $\alpha = 0.05$  检验假设“模型 4 与模型 3 没有显著差别”。
- e. 用  $\alpha = 0.05$  检验假设“模型 4 与模型 2 没有显著差别”。
- f. 通过散点图和(a)~(e)的结果, 你向工程师推荐哪个模型? 解释你挑选该模型的理由。

## 第十三章 多元回归续论

- 13.1 引言和案例
- 13.2 变量的挑选(第一步)
- 13.3 模型形式的确定(第二步)
- 13.4 模型假设的检查(第三步)
- 13.5 小结

### 13.1 引言和案例

在第十二章,我们给出了需要应用多元回归模型的背景,讨论了一般线性模型及其在多元回归模型中的应用,介绍了正规方程(组),即用来获得多元回归模型中的各参数  $\beta$  的最小二乘估计的一组方程。接下来,我们给出了各个  $\hat{\beta}_j$  的标准误差,以及它们在单个参数  $\beta_j$ , 一组  $\beta$ ,  $E(y)$  和  $y$  的未来值的推断中的应用。我们还考虑了几条回归线斜率的比较和 logistic 回归这些特殊情况。最后,我们用简练的矩阵表示给出了这些推断方法。

本章将讨论多元回归模型的实际应用。对于一个给定的问题,如何建立一个合适的多元回归模型呢?虽然没有铁的规则,但我们可以给出一些提示。

首先,对于每一个问题,你必须给回归方程确定一个因变量和几个候选的自变量。这个挑选自变量的过程将在 13.2 节中加以介绍。在 13.3 节中,我们将讨论如何挑选多元回归方程的形式。建立一个多元回归模型的最后一步,是检查当前的假定是否没有被违反。检查模型假定的有效性的方法将在 13.4 节中给出。

对于一个给定的问题,使用一次这些步骤你未必能够得到一个合适的模型。对于有些问题,你可能要反复利用这些步骤才能建立起一个模型。例如,我们在已经考虑了候选的自变量(第一步),并挑选了包含其中的一些自变量的模型形式(第二步)以后,可能发现违反了模型中某些假定(第三步)。这可能意味着我们要返回第一步或者第二步,但在此之前,我们希望已经从前面的思考中变得聪明,可以修改所考虑的变量以及(或者)挑选出来的模型。最后,形成一个满足试验者需要的回归模型。由此模型出发,就可以利用第十二章的分析技术对模型参数、 $E(y)$  和  $y$  进行推断了。

**案例:**建立一个回归模型以帮助一家办公用品公司的销售部经理评估流动推销员的表现

一家办公用品公司的销售部管理人员评估流动推销员的表现时遇到了困难。这些推销员四处奔走,运送该公司的产品,创办展销活动,努力增加市场份额,介绍新的产品,并发现与该公司产品有关的各种问题。在这种工作中,旅行花费了很多的时间。

**数据收集的设计** 销售部管理人员相信,推销员表现优劣的一个重要因子是在路上花费大量时间的深层动机,另外一些变量也有作用。有些销售区域由于人口的差异或零售渠道的差异而比别的区域更有潜力。在大的区域内更加困难,因为跑路的时间更长。

一个重要的变量是补助费。一些推销员拿到工资和一份销售的佣金,而另外一些人则只拿到一份更大的佣金。虽然有些管理人员认为重要的因子是佣金状态和销路的组合,销售部管理人员怀疑这两组之间有差异。尤其是,他们怀疑有多条有效销路并且只拿佣金的推销员是非常高产的。此外,管理人员还怀疑对于有许多销路的推销员,利润可能被夸大了,他们希望知道每条销路上的利润情况。

**数据的整理** 收集了 51 个推销员的数据,数据中包括区号 DIST,该推销员促成的所有定单的纯利润率 PROFIT(这是感兴趣的因变量),区的面积 AREA(以千平方英里为单位),区内人口数 POPN(以百万为单位)区内销路的数量 OUTLETS,补助费的示性变量 COMMIS(对于只拿佣金的人取 1,拿一部分工资一部分佣金的人取 0)。数据如下:

| DIST | PROFIT | AREA  | POPN  | OUTLETS | COMMIS |
|------|--------|-------|-------|---------|--------|
| 1    | 1011   | 16.96 | 3.881 | 213     | 1      |
| 2    | 1318   | 7.31  | 3.141 | 158     | 1      |
| 3    | 1556   | 7.81  | 3.766 | 203     | 1      |
| 4    | 1521   | 7.31  | 4.587 | 170     | 1      |
| 5    | 979    | 19.84 | 3.648 | 142     | 1      |
| 6    | 1290   | 12.37 | 3.456 | 159     | 1      |
| 7    | 1596   | 6.15  | 3.695 | 178     | 1      |
| 8    | 1155   | 14.21 | 3.609 | 182     | 1      |
| 9    | 1412   | 7.45  | 3.801 | 181     | 1      |
| 10   | 1194   | 14.43 | 3.322 | 148     | 1      |
| 11   | 1054   | 6.12  | 5.124 | 227     | 0      |
| 12   | 1157   | 11.71 | 4.158 | 139     | 1      |
| 13   | 1001   | 9.36  | 3.887 | 179     | 0      |
| 14   | 831    | 19.14 | 2.230 | 124     | 1      |
| 15   | 857    | 11.75 | 4.468 | 205     | 0      |
| 16   | 188    | 40.34 | 0.297 | 85      | 1      |
| 17   | 1030   | 7.16  | 4.224 | 211     | 0      |
| 18   | 1331   | 9.37  | 3.427 | 145     | 1      |

| DIST | PROFIT | AREA  | POP   | OUTLETS | COMMIT |
|------|--------|-------|-------|---------|--------|
| 19   | 643    | 7.62  | 4.031 | 205     | 1      |
| 20   | 992    | 27.54 | 2.370 | 166     | 1      |
| 21   | 795    | 15.97 | 3.903 | 149     | 1      |
| 22   | 1340   | 12.97 | 3.423 | 186     | 1      |
| 23   | 689    | 17.36 | 2.390 | 141     | 0      |
| 24   | 1726   | 6.24  | 4.947 | 223     | 1      |
| 25   | 1056   | 11.20 | 4.166 | 176     | 0      |
| 26   | 989    | 18.09 | 4.063 | 187     | 1      |
| 27   | 895    | 13.32 | 3.105 | 131     | 1      |
| 28   | 1028   | 14.97 | 4.116 | 170     | 0      |
| 29   | 771    | 21.92 | 1.510 | 144     | 1      |
| 30   | 484    | 34.91 | 0.741 | 126     | 1      |
| 31   | 917    | 8.46  | 5.260 | 234     | 0      |
| 32   | 1786   | 7.52  | 5.744 | 210     | 0      |
| 33   | 1063   | 14.43 | 2.703 | 141     | 1      |
| 34   | 1001   | 15.37 | 3.583 | 158     | 0      |
| 35   | 1052   | 11.20 | 4.469 | 167     | 1      |
| 36   | 1610   | 7.20  | 4.951 | 174     | 1      |
| 37   | 1486   | 13.49 | 3.474 | 211     | 1      |
| 38   | 1576   | 6.56  | 4.637 | 172     | 1      |
| 39   | 1665   | 9.35  | 3.900 | 185     | 1      |
| 40   | 878    | 11.12 | 3.766 | 166     | 0      |
| 41   | 849    | 10.58 | 3.876 | 189     | 0      |
| 42   | 775    | 17.82 | 2.753 | 164     | 0      |
| 43   | 1012   | 10.03 | 4.449 | 193     | 0      |
| 44   | 1436   | 10.01 | 4.680 | 157     | 1      |
| 45   | 798    | 10.70 | 4.806 | 200     | 0      |
| 46   | 519    | 24.38 | 2.367 | 142     | 0      |
| 47   | 1701   | 6.57  | 5.563 | 199     | 0      |
| 48   | 1387   | 6.64  | 4.357 | 166     | 1      |
| 49   | 1717   | 9.24  | 4.670 | 221     | 1      |
| 50   | 1032   | 11.62 | 3.993 | 180     | 0      |
| 51   | 973    | 12.85 | 3.923 | 193     | 0      |

**数据的分析** 用这些数据进行多元回归分析,看看被管理人员怀疑对 PROFIT 有影响的变量是否真的有影响;尤其是要弄清 COMMIT 和 OUTLETS 是否有组合效应。考虑是用 PROFIT 本身作为因变量好还是用 PROFIT 被 OUTLETS 除了以后作为因变量好;省略那些对预测没有多少价值的变量。如果有的话,找出并纠正

所有对于模型假定的违背。给这些管理人员写一份简短的非技术性的报告,并解释你的发现。这个数据集合的分析将按照本章中所讨论的方法进行。

## 13.2 变量的挑选(第一步)

在构建一个多元回归模型时,最关键的决定或许是自变量的挑选。在本章后面的几节中,我们将考虑许多精化多元回归分析的方法,但是,首先我们必须就考虑把哪些自变量包含进去作出决断,从而决定出去收集哪些自变量的数据。如果我们没有有用的数据,也就不可能建立有用的预测模型。

虽然原本的最优策略可能是构造一个带有许多变量的巨大的多元回归模型,但这样的模型难以解释,并且收集和分析数据都很花时间。那么,一个管理人员如何在原始的变量中合理地挑选出包含在回归分析中的变量呢?

在数据的原始挑选中,问题领域的知识至关重要。首先找出要研究的因变量。那些通过观察、预测过这个因变量并解释其变化的人,通常对于哪些因子(自变量)影响该变量有深入的见解。因此,首先要咨询对于要研究的因变量有最多经验的人。例如,假定问题是要预测在 40 个地区的每一个中某种便宜的计算机打印机在下个季度的销售量。那么因变量  $y$  就是地区销售量。某些自变量,如在每个地区的广告预算以及销路的数量等都是候选的变量。一个出色的地区销售经理无疑会给出其他的自变量。

挑选自变量时一个重要的考虑是共线性问题,即严重相关的自变量。多元回归中的一个偏斜率估计了改变一个自变量而其他自变量保持不变时的预测效应。然而,如果某些或全部自变量一起变动,就很难把每个自变量的预测效应分离出来。当自变量高度相关时,一个常见的结果是:总体  $F$  检验是很显著的,但单个的  $t$  检验都不显著。 $F$  检验结果的显著性只不过表明了自变量中存在可以探测出来的预测价值,而不显著的  $t$  值则表明,在给定其他所有变量的条件下,没有一个变量的附加的预测价值是可以探测到的。其原因在于,高度相关的预测因子相互作为代表,它们中的任何一个或许是有用的,但加上其他的预测因子不起什么作用。当在多元回归模型中用到了所有严重共线性的自变量时,就不可能确定到底是哪一个因子与因变量相关联。

评估一组自变量中的共线性有若干种方法。最简单的方法是看(Pearson)相关矩阵,这个矩阵可以由几乎所有的计算机软件包算出。其中的相关系数越大,共线性问题也越严重。在多数情况下,任何接近或超过 0.9 的相关系数都表示复相关问题的严重性。

一些计算机软件包可以作出散点图阵,即对每一对变量做其相应的散点图而产生的一组散点图。在这样的图阵中,共线性表现为两个自变量有近似线性的关

系。例如,抽取了一群从事有关汽车内容写作的作家,请他们为一辆新型小汽车就汽车的性能(perform)、舒适程度(comfort)、外观(appear)和综合质量打分(overall),分数从0到100。做这项研究的推销经理想知道哪些变量能最好地预测作家对于汽车的打分。用 Minitab 给出的散点图阵如图 13.1 所示。在性能、舒适程度和外观的得分之间有清楚的线性关系,表明了有明显的共线性。下面的相关矩阵证实了这一事实:

```
MTB>correlations c1-c4
```

```
Correlations (pearson)
overall perform comfort
perform 0.698
comfort 0.769 0.801
appear 0.630 0.479 0.693
```

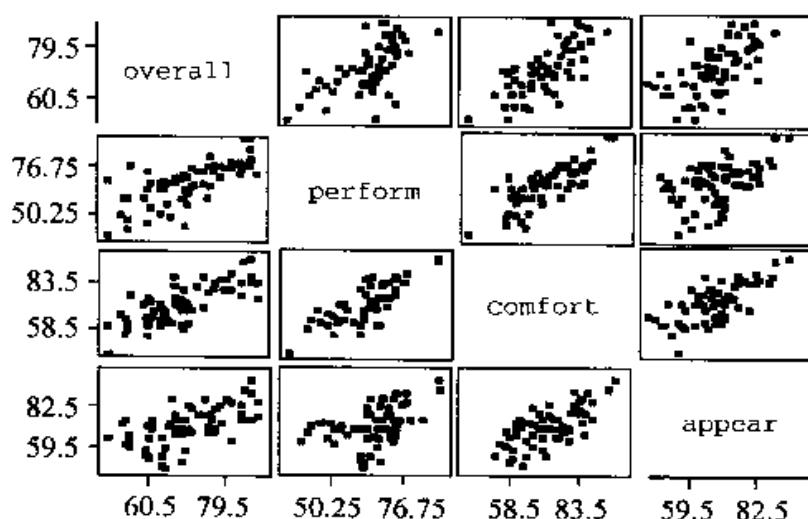


图 13.1 汽车作家数据的散点图阵

散点图阵也可用于探索非线性关系或者离群值。图阵中包含了因变量对各个自变量的散点图。有时,图阵中会清楚地显示出一条曲线或者一个离群值。有时,另一些自变量的效应可能隐含某个问题。本章后面讨论的残差分析是另一种检查是否有违背假定的现象的良好手段。

相关阵和散点图阵并不总能把共线性问题暴露得一览无余。有时,尽管两个预测因子分别与第三个只有不强的相关,但两个因子一起却可以把第三个预测得很好。(直接劳动时间和间接劳动时间一起把总劳动时间预测得相当好,而其中的每一个都只能对总的劳动时间给出不完全的预测。)在各种各样的计算机软件中,已经编入了多种更为复杂的诊断共线性的方法,其中之一是在第十二章中讨论的

方差膨胀因子(VIF), 即  $1/(1 - R^2)$ , 这里  $R^2$  指一个自变量的变差中有多少被其他变量所解释。这个 VIF 考虑了预测因子中的所有关系, 因而比简单的相关系数更完全。Cook and Weisberg (1982) 和 Belsley, Kuh, and Welsch (1980) 中定义了几种诊断共线性的度量。大多数统计学计算机程序的说明书中都指明程序中能计算出来的是这些量中的哪一个, 结果表明了什么。

### 例 13.1

一家连锁超市推出一项某种品牌的冰淇淋的促销活动。从收款台的扫描数据得到了促销活动的周末一天实际销售量(Sales)的数据。三个可能的解释变量分别为: 商店规模的大小(以千平方英尺计算, SqFeet), 商店接待的顾客数(以百人计算, NumCusts)和购物量的平均值(AvgSize, 也是从扫描数据中得到的)。图 13.2

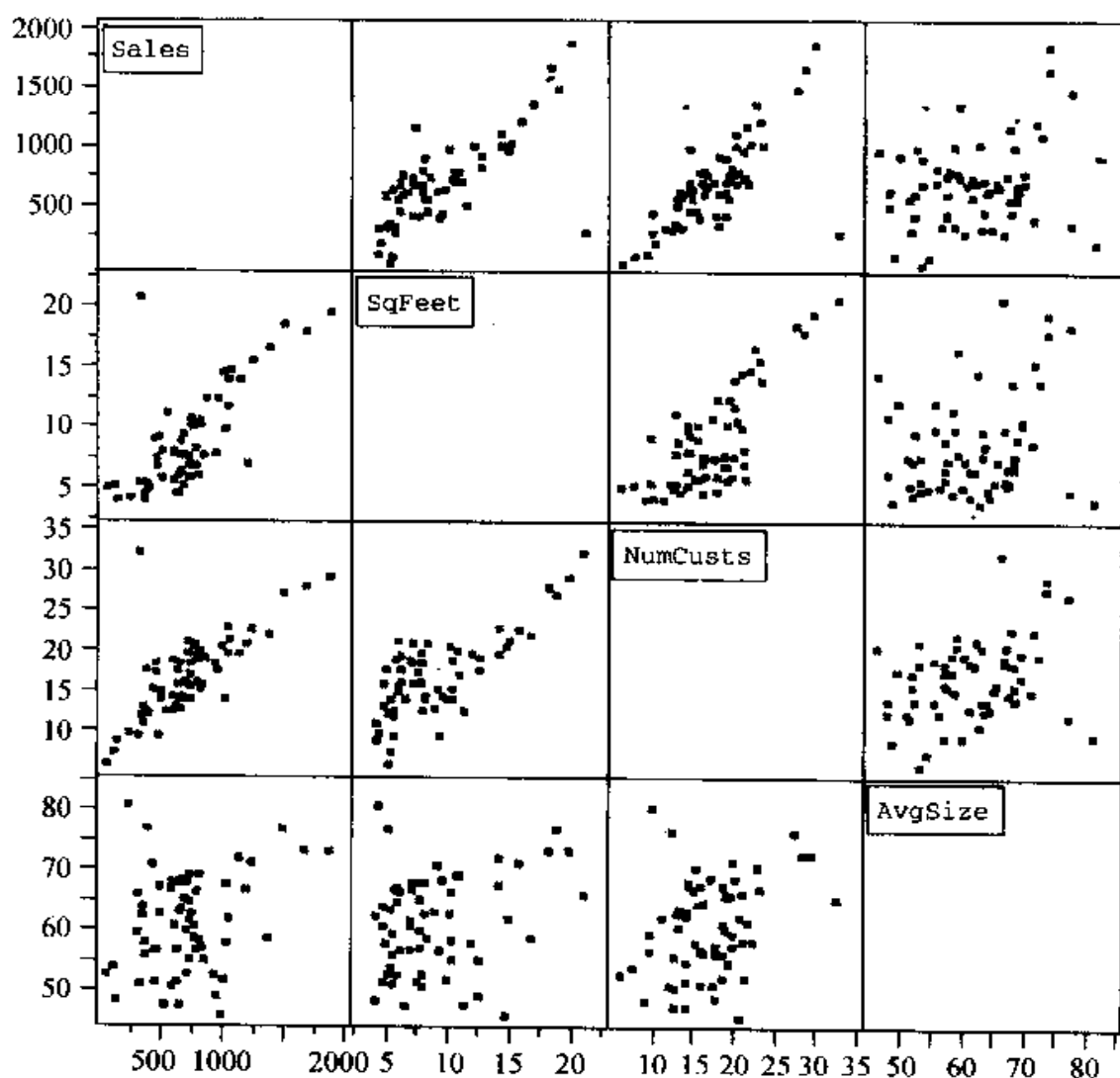


图 13.2 冰淇淋数据的散点图阵

为由这些数据画出的散点图阵。有共线性的现象吗?由任何其他的问题吗?

**解答** 看图阵右上方的六个散点图。这些图表明,在 SqFeet 即商店规模的大小和 NumCusts 即顾客数之间有清楚的增长关系,但并不特别强。而这两个自变量与平均购物量之间相关性不大。因此,有中等程度的共线性问题。该图阵没有严重的非线性问题的迹象,但有一个离群值,在 Sales 对 SqFeet 的图中,一个商店有最大的面积而其销售量几乎最小。这个离群值有相当强的杠杆作用,该商店的两个自变量值都是极端的(大小和顾客数)。该店可能有显著的影响,因为对于这两个预测因子中的任何一个,与该店相应的点都落在回归线下方很远的地方。进一步的检查发现,41 号商店的冰淇淋展台在周末失去了吸引力,商店经理不得已从展台上撤下了冰淇淋。回归分析当中把该店略去了。

避免共线性问题的最好的方法之一是在一项回归研究开始之前,恰当地选取预测变量。要注意选取那些与因变量确实相关、但相互之间又没有相关关系的自变量。如果可能的话,要找出能够反映因变量的各种分量的自变量。例如,假定我们要预测在 40 个销售区内,与个人计算机配套使用的低价位打印机的销售量。总的销售量由几种类型的买主的购买量构成。我们可以找出几个重要的买主类型:大学生,家庭用户,小型公司和计算机网络工作站。因此,我们可以努力得到大学的新生数量、家庭收入、小公司的成立时间以及新网络安装的数据,并用这几个因子作为自变量。其中每一个变量对于打印机销售数量的预测都是有益的,并且在这些自变量中没有明显的相关。懂得你要预测的这个变量的人常常能够识别出该量的分量,并就每个分量给出合理的预测。

### 例 13.2

一家微机销售和服务公司关心服务呼叫的数量。公司在每个销售区域内有几个服务网点,需要服务的计算机用户会打电话给距离最近的服务网点。这些网点由在总部受过训练的技师组成。关键的问题是:应该为这些技师在总部分派责任,还是把他们派到服务网点上;每个月必须做一次分派工作的决策。服务网点上所需要的技师数几乎总是随需要服务的呼叫次数成比例地增加。与服务经理讨论得知,决定服务呼叫次数大小的关键变量可能是:在用计算机数量,新安装的数量,最近是否有了型号的变化以及平均温度(高温,或许还有与之相关的高湿度会导致更频繁的计算机故障,尤其在没有完善的空调设备的房间内)。在这些变量中,哪些变量与其他变量是相关的?

**解答** 很难想像温度会与其他变量相关。在用的计算机数与新安装的计算机数当有某种程度的相关,这只是因为每一台新安装的计算机都是一台在用的计算



机。除非该公司刚刚经历了一次快速增长,我们认为二者之间不会有严重的相关(然而,我们要看数据)。型号的变化与在用数量以及新安装数量之间的相关不明显。我们应当收集数据,并作出相关分析。

经理在开始做回归研究时,可能会把太多的自变量放入到回归模型中去。因此,我们需要某些指导性的原则,以帮助从候选的变量中挑选出包含在最后的回归模型中的自变量。

在与专家讨论确定了一个自变量的清单后,要从该清单挑选出应该包含在回归模型中的自变量,有若干方法可供选用。本书中我们将讨论几个这样的方法,进一步的细节可以参看 Neter, Kutner, Nachtsheim, Wasserman(1996)。

第一种挑选方法是用因变量和候选变量表中的一个或多个自变量进行所有可能的回归。显然,除非分析可以使用适当的软件用计算机来完成,并且有充分好的硬件设备来高效地运行大量的回归模型的计算,否则,不应选用这种方法。

作为一个说明,我们考虑由 20 个独立的药品公司组成的样本,并为每个公司假想一个月处方销售量 VOLUME 数据。这些数据,以及总营业面积(FLOOR-SP)、处方部占总营业面积的百分比(PRESC-RX)、公司可用的停车位(PARKING)、公司是否位于一个购物中心内(SHOPCNTR)以及周围社区的人均收入(INCOME)的数据记录于表 13.1。

表 13.1 20 家药品公司的数据

| 序号     | 公司 | VOLUME | FLOOR-SP | PRESC-RX | PARKING | SHOPCNTR | INCOME |
|--------|----|--------|----------|----------|---------|----------|--------|
| 1      | 1  | 22     | 4900     | 9        | 40      | 1        | 18     |
| 2      | 2  | 19     | 5800     | 10       | 50      | 1        | 20     |
| 3      | 3  | 24     | 5000     | 11       | 55      | 1        | 17     |
| 4      | 4  | 28     | 4400     | 12       | 30      | 0        | 19     |
| 5      | 5  | 18     | 3850     | 13       | 42      | 0        | 10     |
| 6      | 6  | 21     | 5300     | 15       | 20      | 1        | 22     |
| 7      | 7  | 29     | 4100     | 20       | 25      | 0        | 8      |
| 8      | 8  | 15     | 4700     | 22       | 60      | 1        | 15     |
| 9      | 9  | 12     | 5600     | 24       | 45      | 1        | 16     |
| 10     | 10 | 14     | 4900     | 27       | 82      | 1        | 14     |
| 11     | 11 | 18     | 3700     | 28       | 56      | 0        | 12     |
| 12     | 12 | 19     | 3800     | 31       | 38      | 0        | 8      |
| 13     | 13 | 15     | 2400     | 36       | 35      | 0        | 6      |
| 14     | 14 | 22     | 1800     | 37       | 28      | 0        | 4      |
| 15     | 15 | 13     | 3100     | 40       | 43      | 0        | 6      |
| 16     | 16 | 16     | 2300     | 41       | 20      | 0        | 5      |
| 17     | 17 | 8      | 4400     | 42       | 46      | 1        | 7      |
| 18     | 18 | 6      | 3300     | 42       | 15      | 0        | 4      |
| 19     | 19 | 7      | 2900     | 45       | 30      | 1        | 9      |
| 20     | 20 | 17     | 2400     | 46       | 16      | 0        | 3      |
| N = 20 |    |        |          |          |         |          |        |

在进行表 13.1 中数据的所有可能的回归之前,我们需要考虑用什么样的准则从所有可能的回归中挑选拟合得最好的方程。从所有可能的回归方程中挑选最优回归方程的第一个,也许是最简单的准则,是计算用来估计  $\sigma_e^2$  的  $s_e^2 = MS(\text{残差}) = SS(\text{残差})/[n - (k + 1)]$ 。由于该量用于大多数关于模型参数和  $E(y)$  的推断(统计检验和置信区间),选择使得  $s_e^2$  最小的模型是合理的。第二个准则使用每个模型的决定系数  $R^2$ 。通过考察有最高  $R^2$  值的模型,我们可以看到是否有某个适当的模式对于包含在模型中的变量的个数给出一个建议,并识别出这些变量。

### 例 13.3

考虑表 13.1 中的数据。用  $R^2$  准则在含有 1 个,2 个,3 个及 4 个自变量的回归方程中决定最优的回归方程。

**解答** 这里给出 SAS 的输出结果,有最高  $R^2$  值的方程见表 13.2。

SAS OUTPUT FOR PROC RSQUARE

ALL POSSIBLE SUBSETS ANALYSIS

N = 20 Regression Models for Dependent Variable: VOLUME

NUMBER

| IN MODEL | R-square  | C(p)    | Variables in Model |          |
|----------|-----------|---------|--------------------|----------|
| 1        | 0.4393318 | 10.1709 | PRESC_RX           |          |
| 1        | 0.1479899 | 23.7702 | INCOME             |          |
| 1        | 0.0410534 | 28.7618 | SHOPCNTR           |          |
| 1        | 0.0335317 | 29.1129 | FLOOR_SP           |          |
| 1        | 0.0048042 | 30.4539 | PARKING            |          |
| -----    |           |         |                    |          |
| 2        | 0.6656627 | 1.6062  | FLOOR_SP           | PRESC_RX |
| 2        | 0.6470647 | 2.4744  | PRESC_RX           | SHOPCNTR |
| 2        | 0.5474878 | 7.1224  | PRESC_RX           | INCOME   |
| 2        | 0.5314244 | 7.8722  | PRESC_RX           | PARKING  |
| 2        | 0.4957679 | 9.5366  | SHOPCNTR           | INCOME   |
| 2        | 0.2565364 | 20.7035 | FLOOR_SP           | SHOPCNTR |
| 2        | 0.2348733 | 21.7147 | FLOOR_SP           | INCOME   |
| 2        | 0.2054310 | 23.0890 | PARKING            | INCOME   |
| 2        | 0.0685567 | 29.4780 | FLOOR_SP           | PARKING  |
| 2        | 0.0421078 | 30.7126 | PARKING            | SHOPCNTR |
| -----    |           |         |                    |          |

|       |           |         |          |          |          |                 |
|-------|-----------|---------|----------|----------|----------|-----------------|
| 3     | 0.6907243 | 2.4364  | FLOOR_SP | PRESC_RX | SHOPCNTR |                 |
| 3     | 0.6794331 | 2.9635  | FLOOR_SP | PRESC_RX | PARKING  |                 |
| 3     | 0.6664115 | 3.5713  | FLOOR_SP | PRESC_RX | INCOME   |                 |
| 3     | 0.6625912 | 3.7496  | PRESC_RX | PARKING  | SHOPCNTR |                 |
| 3     | 0.6471156 | 4.4720  | PRESC_RX | SHOPCNTR | INCOME   |                 |
| 3     | 0.6024323 | 6.5577  | PRESC_RX | PARKING  | INCOME   |                 |
| 3     | 0.5001258 | 11.3332 | FLOOR_SP | SHOPCNTR | INCOME   |                 |
| 3     | 0.4982807 | 11.4193 | PARKING  | SHOPCNTR | INCOME   |                 |
| 3     | 0.2650711 | 22.3051 | FLOOR_SP | PARKING  | SHOPCNTR |                 |
| 3     | 0.2556961 | 22.7427 | FLOOR_SP | PARKING  | INCOME   |                 |
| ----- |           |         |          |          |          |                 |
| 4     | 0.6987395 | 4.0623  | FLOOR_SP | PRESC_RX | PARKING  | SHOPCNTR        |
| 4     | 0.6932666 | 4.3177  | FLOOR_SP | PRESC_RX | SHOPCNTR | INCOME          |
| 4     | 0.6805857 | 4.9097  | FLOOR_SP | PRESC_RX | PARKING  | INCOME          |
| 4     | 0.6630086 | 5.7301  | PRESC_RX | PARKING  | SHOPCNTR | INCOME          |
| 4     | 0.5012890 | 13.2789 | FLOOR_SP | PARKING  | SHOPCNTR | INCOME          |
| ----- |           |         |          |          |          |                 |
| 5     | 0.7000737 | 6.0000  | FLOOR_SP | PRESC_RX | PARKING  | SHOPCNTR INCOME |
| ----- |           |         |          |          |          |                 |

表 13.2  $R^2$  准则下的最优拟合模型

| 自变量的个数 | $R^2$ | 变量                  |
|--------|-------|---------------------|
| 1      | 0.439 | 处方部面积百分比            |
| 2      | 0.666 | 总营业面积和处方部面积百分比      |
| 3      | 0.691 | 总营业面积、处方部面积百分比和购物中心 |
| 4      | 0.699 | 除人均收入外的其他变量         |

虽然从一个变到两个自变量时,  $R^2$  的值有一个跳跃,但继续增加变量时改进不大。因此,基于  $R^2$  准则的最优拟合模型包含总营业面积和处方部面积百分比。

用  $R^2$  作为挑选最优回归方程的准则的问题之一是,随着自变量个数的增加,  $R^2$  是增加的,哪怕新的自变量只有很小的预测能力。另一个当增加新变量时不必增加的挑选最优回归的方法如下。

记住我们搜寻的目标是对将来的值  $y$  产生最优预测的回归方程中所包含的自变量子集。然而,不幸的是,由于并不知道将来的值是什么,我们将努力寻求最优回归方程,使得对于已知的样本  $y$  值有最好的拟合。在强调对于已知的样本  $y$  值的最优拟合与强调选取对于未来  $y$  值的最优预测因子之间,一座可能的桥梁是把样本数据分为两部分,一部分用于拟合各种各样的回归方程,而另一部分用来校

验预测方程对于“未来”的值的预测有多好。至于应把多少数据用来拟合,把多少数据用来校验预测,虽然并无普遍接受的规则,然而,当样本大小  $n$  大于  $2p + 20$  时,把数据分为两半是合理的,这里,  $p$  为最大的候选模型中的参数个数。最优预测方程的一个可能的准则是,对于用来校验预测的那部分数据,极小化  $\sum (y_i - \hat{y}_i)^2$ 。

一旦用这个分离数据的方法挑选出了回归模型,就要用全体的样本数据来得到最后的预测方程。因此,虽然看起来好像我们只用了一部分数据,但其实最后的预测方程是用所有数据得到的。

然而,观察是要花钱的。要得到足够的数据以应用分离数据的方法挑选拟合最优的回归方程,有时是不现实的。在这种情况下,可以用 PRESS 统计量来完成校验。对于一个  $y$  值的样本和一个联系  $y$  于一组  $x$  的回归模型,我们首先去掉第一个观察值,用其余的  $n - 1$  个观察值拟合该模型。基于所得到的拟合方程,估计第一个观察值(记为  $\hat{y}_1^*$ )并计算出残差  $y_1 - \hat{y}_1^*$ 。重复该过程  $n - 1$  次,依次除去第二、三、 $\dots$ ,第  $n$  个观察值,每次都对除去的观察值计算残差,然后得到 PRESS 统计量为

$$\text{PRESS} = \sum_{i=1}^n (y_i - \hat{y}_i^*)^2$$

选取使得 PRESS 统计量最小的模型作为最优拟合模型。

#### 例 13.4

对于表 13.1 中的数据,计算 PRESS 统计量,以确定最优拟合回归方程。

**解答** 这里给出 SAS 的输出结果。基于 PRESS 统计量有最小值的最优拟合模型含有自变量总营业面积和处方部面积百分比。

```
SAS OUTPUT FOR ALL POSSIBLE SUBSET ANALYSIS
PRESS STATISTIC
N = 20 Regression Models for Dependent Variable: VOLUME
NUMBER IN PRESS STATISTIC VARIABLES IN MODEL
MODEL
1 516.391 PRESC_RX
1 772.163 INCOME
1 869.668 FLOOR_SP
1 887.636 SHOPCNTR
1 907.636 PARKING
```

|       |         |          |          |          |                 |
|-------|---------|----------|----------|----------|-----------------|
| 2     | 347.007 | FLOOR_SP | PRESC_RX |          |                 |
| 2     | 368.757 | PRESC_RX | SHOPCNTR |          |                 |
| 2     | 479.976 | PRESC_RX | PARKING  |          |                 |
| 2     | 485.820 | SHOPCNTR | INCOME   |          |                 |
| 2     | 547.150 | PRESC_RX | INCOME   |          |                 |
| 2     | 762.507 | FLOOR_SP | SHOPCNTR |          |                 |
| 2     | 787.578 | PARKING  | INCOME   |          |                 |
| 2     | 797.404 | FLOOR_SP | INCOME   |          |                 |
| 2     | 916.644 | FLOOR_SP | PARKING  |          |                 |
| 2     | 975.912 | PARKING  | SHOPCNTR |          |                 |
| ----- |         |          |          |          |                 |
| 3     | 370.843 | FLOOR_SP | PRESC_RX | SHOPCNTR |                 |
| 3     | 371.671 | FLOOR_SP | PRESC_RX | PARKING  |                 |
| 3     | 378.166 | PRESC_RX | PARKING  | SHOPCNTR |                 |
| 3     | 455.424 | PRESC_RX | SHOPCNTR | INCOME   |                 |
| 3     | 482.387 | FLOOR_SP | PRESC_RX | INCOME   |                 |
| 3     | 513.246 | PRESC_RX | PARKING  | INCOME   |                 |
| 3     | 523.006 | PARKING  | SHOPCNTR | INCOME   |                 |
| 3     | 602.214 | FLOOR_SP | SHOPCNTR | INCOME   |                 |
| 3     | 819.792 | FLOOR_SP | PARKING  | SHOPCNTR |                 |
| 3     | 890.550 | FLOOR_SP | PARKING  | INCOME   |                 |
| ----- |         |          |          |          |                 |
| 4     | 405.832 | FLOOR_SP | PRESC_RX | PARKING  | SHOPCNTR        |
| 4     | 458.014 | PRESC_RX | PARKING  | SHOPCNTR | INCOME          |
| 4     | 471.086 | FLOOR_SP | PRESC_RX | SHOPCNTR | INCOME          |
| 4     | 513.468 | FLOOR_SP | PRESC_RX | PARKING  | INCOME          |
| 4     | 684.190 | FLOOR_SP | PARKING  | SHOPCNTR | INCOME          |
| ----- |         |          |          |          |                 |
| 5     | 513.915 | FLOOR_SP | PRESC_RX | PARKING  | SHOPCNTR INCOME |
| ----- |         |          |          |          |                 |

到此为止,我们已经讨论了从自变量的一个子集中挑选最优拟合回归方程的准则。一般来说,如果我们选择了一个模型,其中丢掉了个或多个重要的预测因子,就说这个模型是诠释不充分的,此时本应当用丢掉的这些自变量解释的  $y$  值的变异性就会附加到估计的误差方差中。另一方面,如果我们选择了一个模型,其中包含了一个或多个“额外”的预测因子,就说这个模型是诠释过头的,此时可能有复共线性问题,我们以后再处理这一问题。防止诠释不充分和诠释过头的关键是基于  $C_p$  统计量的准则,该准则针对过头和不充分诠释的问题平衡了前面的挑选准则的得失,从而达到挑选最优子集回归方程的目的。 $C_p$  统计量[参见 Mallows

(1973)]为

$$C_p = \frac{SS(\text{残差})_p}{s_e^2} - (n - 2p)$$

其中  $SS(\text{残差})_p$  表示从有  $p$  个参数(包括  $\beta_0$ )的模型得到的误差平方和,  $s_e^2$  为包含最多个数的自变量的回归方程的均方误差。在一个模型挑选问题中,对于每一个拟合的回归方程计算  $C_p$  的值。理论的结果表明,对于最优拟合模型应有  $C_p \approx p$ 。

### 例 13.5

参见例 13.3 中的输出。确定含有 1, 2, 3, 4 和 5 个变量的所有可能的回归的  $C_p$  值。挑选含有 1, 2, 3, 4 个自变量的最优拟合方程。基于  $C_p$  统计量,哪个回归方程给出最优的总体拟合?

**解答** 最优拟合模型总结在表 13.3 中。基于  $C_p$  准则,含有 2, 3 和 4 个自变量的最优模型区别不是很大。最重要的预测变量应为总营业面积和处方部所占的百分比,因为它们出现在所有含 2, 3 和 4 个自变量的最优拟合模型中。注意,这两个变量也是在例 13.3 中找出的重要变量。

表 13.3  $C_p$  准则下的最优拟合模型

| 自变量的个数 | $p$ | $C_p$ | 变量                         |
|--------|-----|-------|----------------------------|
| 1      | 2   | 10.17 | 处方部面积百分比                   |
| 2      | 3   | 1.61  | 总营业面积, 处方部面积百分比            |
|        |     | 2.47  | 处方部面积百分比, 购物中心             |
| 3      | 4   | 2.96  | 总营业面积, 处方部面积百分比, 停车位       |
| 4      | 5   | 4.06  | 总营业面积, 处方部面积百分比, 停车位, 购物中心 |

**最优子集回归** 提供了另外一种从含有  $k$  个自变量的回归方程中寻找最优拟合模型的方法。该方法使用了一个可以避免进行所有可能的回归的算法,相应的计算机程序给出  $M$  ( $M$  由用户给定)个含有 1 个自变量、 $M$  个含有 2 个自变量、...,直到含有所有的  $K$  个自变量的回归方程。某些程序允许用户给定最优准则(如  $C_p$  和最大  $R^2$ ),而另一些程序则用固定的准则。例如, Minitab 使用最大  $R^2$  准则对于每个子集的大小来挑选  $M$  个最优子集,对于每个  $k = 1, 2, \dots, K$  算出有最大  $R^2$  的  $M$  个回归方程。下面我们用法 13.1 中的数据来说明这个方法。

Response is VOLUME

## VARIABLES INCLUDED

Indicated by X

F P S

L R P H

O E A O I

O S R P N

R C K C C

- - I N O

S R N T M

P X G R E

No.

Vars

In

## Model

R-S $\alpha$ 

R-Sq(adj)

C-D

2

|   |      |      |      |        |   |
|---|------|------|------|--------|---|
| 1 | 43.9 | 40.8 | 10.2 | 4.8351 | x |
|---|------|------|------|--------|---|

|   |      |      |      |        |   |
|---|------|------|------|--------|---|
| 1 | 14.8 | 10.1 | 23.8 | 5.9604 | X |
|---|------|------|------|--------|---|

|   |      |      |     |        |   |   |
|---|------|------|-----|--------|---|---|
| 2 | 66.6 | 62.6 | 1.6 | 3.8420 | X | X |
|---|------|------|-----|--------|---|---|

|   |      |      |     |        |   |   |
|---|------|------|-----|--------|---|---|
| 2 | 64.7 | 60.6 | 2.5 | 3.9474 | X | X |
|---|------|------|-----|--------|---|---|

|   |      |      |     |        |   |   |   |
|---|------|------|-----|--------|---|---|---|
| 3 | 69.1 | 63.3 | 2.4 | 3.0089 | X | X | X |
|---|------|------|-----|--------|---|---|---|

|   |      |      |     |        |   |   |   |
|---|------|------|-----|--------|---|---|---|
| 3 | 67.9 | 61.9 | 3.0 | 3.8778 | X | X | X |
|---|------|------|-----|--------|---|---|---|

|   |      |      |     |        |   |   |   |   |
|---|------|------|-----|--------|---|---|---|---|
| 4 | 69.9 | 61.8 | 4.1 | 3.8825 | X | X | X | X |
|---|------|------|-----|--------|---|---|---|---|

|   |      |      |     |        |   |   |   |   |
|---|------|------|-----|--------|---|---|---|---|
| 4 | 69.3 | 61.1 | 4.3 | 3.9176 | X | X | X | X |
|---|------|------|-----|--------|---|---|---|---|

|   |      |      |     |        |   |   |   |   |   |
|---|------|------|-----|--------|---|---|---|---|---|
| 5 | 70.0 | 59.3 | 6.0 | 4.0099 | X | X | X | X | X |
|---|------|------|-----|--------|---|---|---|---|---|

还有许多其他的方法可以用来挑选最优回归。虽然我们打算再花费大量的篇幅来讨论这一主题,但要简单介绍一下**向后剔除方法**和**逐步回归方法**。

向后剔除方法从拟合包含所有候选自变量的回归模型出发,对于每一个自变量  $x_j$ , 我们计算

$$F_j = \frac{SS_{\text{drop}_j}}{MS(\text{残差})}, \quad j = 1, 2, \dots$$

其中  $SS_{\text{drop}_j}$  为从含有所有自变量的模型中去掉  $x_j$  后回归平方和的下降量,  $MS$  (残差) 为从完全模型得到的均方误差。以  $\min F_j$  记最小的  $F_j$  的值。如果  $\min F_j < F_\alpha$ , 其中  $\alpha$  为事先选好的显著性水平, 那么就从方程中除去相应于  $\min F_j$  的自变量。然后, 从这个除去了一个自变量的模型出发, 重复向后挑选的上述步骤, 再从自变量的候选清单中除去一个自变量。

向后剔除的方法从包含所有自变量的完全模型出发, 每次除去一个自变量, 直到找出一个适当的回归模型未为止, 此时, 有  $\min F_j > F_\alpha$ 。所得到的模型就是最优的回归方程。另一方面, 逐步回归是从另一个方向出发, 即从模型  $y = \beta_0 + \epsilon$ , 每次增加一个变量, 直到满足停止条件为止。开始的时候, 进入到方程的第一个变量是有回归  $F$  检验的最大  $F$  值的那一个。第二阶段, 把使得含有两个自变量的回归方程的  $F$  检验统计量有最大值的那两个自变量放入模型。注意, 第一阶段选进模型的那个自变量在第二阶段中未必能包含进来, 即最好的单个自变量未必是最好的两个自变量之一。由于这一点, 有人对逐步回归做了简化(有时称为向前挑选), 其中一旦某个自变量进入到回归方程中, 此后就不会被拿出来。

### 例 13.7

用向后剔除的方法对于例 13.3 中的数据找出包括在回归方程中的变量, 并对所得结论作出评述。

**解答** 把向后剔除的方法应用于表 13.1 中的数据, SAS 的输出结果如下。如前所指出, 向后剔除开始于含有所有(5 个)候选变量的回归方程。这一步在向后剔除的过程中称为第 0 步。然后, 排除一个又一个自变量, 直到  $\min F_j > F_\alpha$ 。注意, 在第 1 步, 变量“income”(收入)被除去, 在第 2 步变量“parking”(停车位)被除去, 在此例中第 3 步是最后一步, 变量“shopcntr”被除去。输出结果中给出, 其余的变量构成了基于向后剔除方法的最优拟合回归方程。即

$$\hat{y} = 48.21 - 0.004(\text{营业面积}) - 0.582(\text{处方部百分比})$$

这恰与我们从别的挑选方法得到的结果相同。



REGRESSION ANALYSIS, USING BACKWARD ELIMINATION

Backward Elimination Procedure for Dependent Variable VOLUME

Step 0 All Variables Entered R-square = 0.70007369 C(p) = 6.00000000

|            | DF | Sum of Squares | Mean Square  | F    | Prob>F |
|------------|----|----------------|--------------|------|--------|
| Regression | 5  | 525.44030541   | 105.08806108 | 6.54 | 0.0025 |
| Error      | 14 | 225.10969459   | 16.07926390  |      |        |
| Total      | 19 | 750.55000000   |              |      |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | 42.08710826        | 10.43775070    | 261.42703544           | 16.26 | 0.0012 |
| FLOOR_SP | -0.00241878        | 0.00183889     | 27.81923726            | 1.73  | 0.2095 |
| PRESC_RX | -0.50046955        | 0.16429694     | 149.19783807           | 9.28  | 0.0087 |
| PARKING  | -0.03690284        | 0.06546687     | 5.10907792             | 0.32  | 0.5819 |
| SHOPCNTR | -3.09957355        | 3.24983522     | 14.62673442            | 0.91  | 0.3564 |
| INCOME   | 0.10666360         | 0.42742012     | 1.00135642             | 0.06  | 0.8066 |

Bounds on condition number: 7.823107, 117.1991

Step 1 Variable INCOME Removed R-square = 0.69873952 C(p) = 4.06227626

|            | DF | Sum of Squares | Mean Square  | F    | Prob>F |
|------------|----|----------------|--------------|------|--------|
| Regression | 4  | 524.43894899   | 131.10973725 | 8.70 | 0.0008 |
| Error      | 15 | 226.11105101   | 15.07407007  |      |        |
| Total      | 19 | 750.55000000   |              |      |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | 43.46782063        | 8.56960161     | 387.83321233           | 25.73 | 0.0001 |
| FLOOR_SP | -0.00228513        | 0.00170330     | 27.13112543            | 1.80  | 0.1997 |
| PRESC_RX | -0.52910174        | 0.11386382     | 325.48983690           | 21.59 | 0.0003 |
| PARKING  | -0.03952477        | 0.06256589     | 6.01580808             | 0.40  | 0.5371 |
| SHOPCNTR | -2.71387948        | 2.76799605     | 14.49041122            | 0.96  | 0.3424 |

Bounds on condition number: 5.071729, 46.98862

Step 2 Variable PARKING Removed R-square = 0.69072432 C(p) = 2.43641080

|            | DF | Sum of Squares | Mean Square  | F     | Prob>F |
|------------|----|----------------|--------------|-------|--------|
| Regression | 3  | 518.42314091   | 172.80771364 | 11.91 | 0.0002 |
| Error      | 16 | 232.12685909   | 14.50792869  |       |        |
| Total      | 19 | 750.55000000   |              |       |        |

| Parameter | Standard Error | Type II |
|-----------|----------------|---------|
|-----------|----------------|---------|

| Variable | Estimate    | Error      | Sum of Squares | F     | Prob>F |
|----------|-------------|------------|----------------|-------|--------|
| INTERCEP | 42.82702645 | 8.34803435 | 381.83242065   | 26.32 | 0.0001 |
| FLOOR_SP | -0.00247284 | 0.00164539 | 32.76871130    | 2.26  | 0.1523 |
| PRESC_RX | -0.52941361 | 0.11170410 | 325.87978038   | 22.46 | 0.0002 |
| SHOPCNTR | -3.03834296 | 2.66836223 | 18.81002755    | 1.30  | 0.2716 |

Bounds on condition number; 4.917388, 30.31995

Step 3 Variable SHOPCNTR Removed R-square = 0.66566267 C(p) = 1.60624219

|            | DF | Sum of Squares | Mean Square  | F     | Prob>F |
|------------|----|----------------|--------------|-------|--------|
| Regression | 2  | 499.61311336   | 249.80655668 | 16.92 | 0.0001 |
| Error      | 17 | 250.93688664   | 14.76099333  |       |        |
| Total      | 19 | 750.55000000   |              |       |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | 48.29085530        | 6.89043477     | 725.02357305           | 49.12 | 0.0001 |
| FLOOR_SP | -0.00384228        | 0.00113262     | 169.87259933           | 11.51 | 0.0035 |
| PRESC_RX | -0.58189034        | 0.10263739     | 474.44587802           | 32.14 | 0.0001 |

Bounds on condition number; 2.290122, 9.160487

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination Procedure for Dependent Variable VOLUME

| Step | Variable Removed | Number In | Partial R <sup>2</sup> | Model R <sup>2</sup> | C(p)   | F      | Prob>F |
|------|------------------|-----------|------------------------|----------------------|--------|--------|--------|
| 1    | INCOME           | 4         | 0.0013                 | 0.6987               | 4.0623 | 0.0623 | 0.8066 |
| 2    | PARKING          | 3         | 0.0080                 | 0.6907               | 2.4364 | 0.3991 | 0.5371 |
| 3    | SHOPCNTR         | 2         | 0.0251                 | 0.6657               | 1.6062 | 1.2965 | 0.2716 |

### 例 13.8

叙述把逐步回归应用于表 13.1 中的数据所得到的结果。

**解答** 下面是表 13.1 中数据的 SAS 输出结果。逐步回归开始于模型  $y = \beta_0 + \epsilon$ , 每次增加一个变量。对于这些数据, 在逐步回归的第 1 步中, 变量“处方部百分比”加到回归方程中, 在第 2 步中, 变量“营业面积”进入回归方程, 第 3 步变量“购物中心”添加到回归方程。没有其他变量能够满足  $p = 0.5$  时进入回归方程的条件。如果挑选变量的准则更严格, 此时对于每个新的自变量需要更小的  $p$  值 (比如 0.15 或更小), 那么, 在逐步回归的第 3 步中, 变量“购物中心” ( $p$  值为 0.2716) 则不能进入回归方程, 从而得到的结果与用其他方法得到的相同。

REGRESSION ANALYSIS, USING FORWARD ELIMINATION

Forward Selection Procedure for Dependent Variable VOLUME

Step 1 Variable PRESC\_RX Entered R-square = 0.43933184 C(p) = 10.17094219

|            | DF | Sum of Squares | Mean Square  | F     | Prob>F |
|------------|----|----------------|--------------|-------|--------|
| Regression | 1  | 329.74051403   | 329.74051403 | 14.10 | 0.0014 |
| Error      | 18 | 420.80948597   | 23.37830478  |       |        |
| Total      | 19 | 750.55000000   |              |       |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F      | Prob>F |
|----------|--------------------|----------------|------------------------|--------|--------|
| INTERCEP | 25.98133346        | 2.58814791     | 2355.90463660          | 100.77 | 0.0001 |
| PRESC_RX | -0.32055657        | 0.08535423     | 329.74051403           | 14.10  | 0.0014 |

Bounds on condition number; 1, 1

Step 2 Variable FLOOR\_SP Entered R-square = 0.66566267 C(p) = 1.60624219

|            | DF | Sum of Squares | Mean Square  | F     | Prob>F |
|------------|----|----------------|--------------|-------|--------|
| Regression | 2  | 499.61311336   | 249.80655668 | 16.92 | 0.0001 |
| Error      | 17 | 250.93688664   | 14.76099333  |       |        |
| Total      | 19 | 750.55000000   |              |       |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | 48.29085530        | 6.89043477     | 725.02357305           | 49.12 | 0.0001 |
| FLOOR_SP | -0.00384228        | 0.00113262     | 169.87259933           | 11.51 | 0.0035 |
| PRESC_RX | -0.58189034        | 0.10263739     | 474.44587802           | 32.14 | 0.0001 |

Bounds on condition number; 2.290122, 9.160487

Step 3 Variable SHOPCNTR Entered R-square = 0.69072432 C(p) = 2.43641080

|            | DF | Sum of Squares | Mean Square  | F     | Prob>F |
|------------|----|----------------|--------------|-------|--------|
| Regression | 3  | 518.42314091   | 172.80771364 | 11.91 | 0.0002 |
| Error      | 16 | 232.12685909   | 14.50792869  |       |        |
| Total      | 19 | 750.55000000   |              |       |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | 42.82702645        | 8.34803435     | 381.83242065           | 26.32 | 0.0001 |
| FLOOR_SP | -0.00247284        | 0.00164539     | 32.76871130            | 2.26  | 0.1523 |
| PRESC_RX | -0.52941361        | 0.11170410     | 325.87978038           | 22.46 | 0.0002 |
| SHOPCNTR | -3.03834296        | 2.66836223     | 18.81002755            | 1.30  | 0.2716 |

Bounds on condition number; 4.917388, 30.31995

NO other variable met the 0.5000 significance level for entry into the model.

Summary of Forward Selection Procedure for Dependent Variable VOLUME

| Step | Variable Entered | Number In | Partial $R^2$ | Model $R^2$ | C(p)    | F       | Prob>F |
|------|------------------|-----------|---------------|-------------|---------|---------|--------|
| 1    | PRESC_RX         | 1         | 0.4393        | 0.4393      | 10.1709 | 14.1046 | 0.0014 |
| 2    | FLOOR_SP         | 2         | 0.2263        | 0.6657      | 1.6062  | 11.5082 | 0.0035 |
| 3    | SHOPCNTR         | 3         | 0.0251        | 0.6907      | 2.4364  | 1.2965  | 0.2716 |

在典型的回归问题中,你通过与专家讨论或者用任何一种可能的挑选方法来确定进入回归方程的变量。例如,我们可以计算所有可能的回归,或应用最优子集回归方法,或使用逐步回归(向后剔除)方法。还有很多其他的方法。有时,不同的准则得到的最优(或近似最优)模型一致,有时则不同。哪个方法最好呢?我们应该相信和使用哪一个呢?

对于这些问题的最重要的回答是,当可以使用计算机并有合用的软件时,这些挑选方法中的任何一个都可能是有效的;没有一个方法被公认是比其他方法更好的。因此,不要试图用多个或所有挑选方法,你应当从一种方法(或许就是由于你手边的计算机中的某个软件中有这样一个方法)开始,通过反复不断的使用,逐渐积累关于这个方法的知识。这样,你就可以用这个方法武装起来,能够解决几乎所有你遇到的回归问题了。

## 练习

### 应用

**13.1(教育)** (班级作业)你学院或学校的招生办主任想建立一个回归方程,以基于一个学生的高中的学习成绩来预测一年以后他或她的年级点平均(GPA)。与你的同学讨论这个项目,并请教其他专家,列出包含在回归模型中的候选自变量。是应该建立一个回归模型了事,还是应该考虑得更多些?虚拟变量有用吗?

**13.2(班级作业)** 见练习 13.1。从招生办获得数据,并应用某个挑选方法识别可能的回归模型。

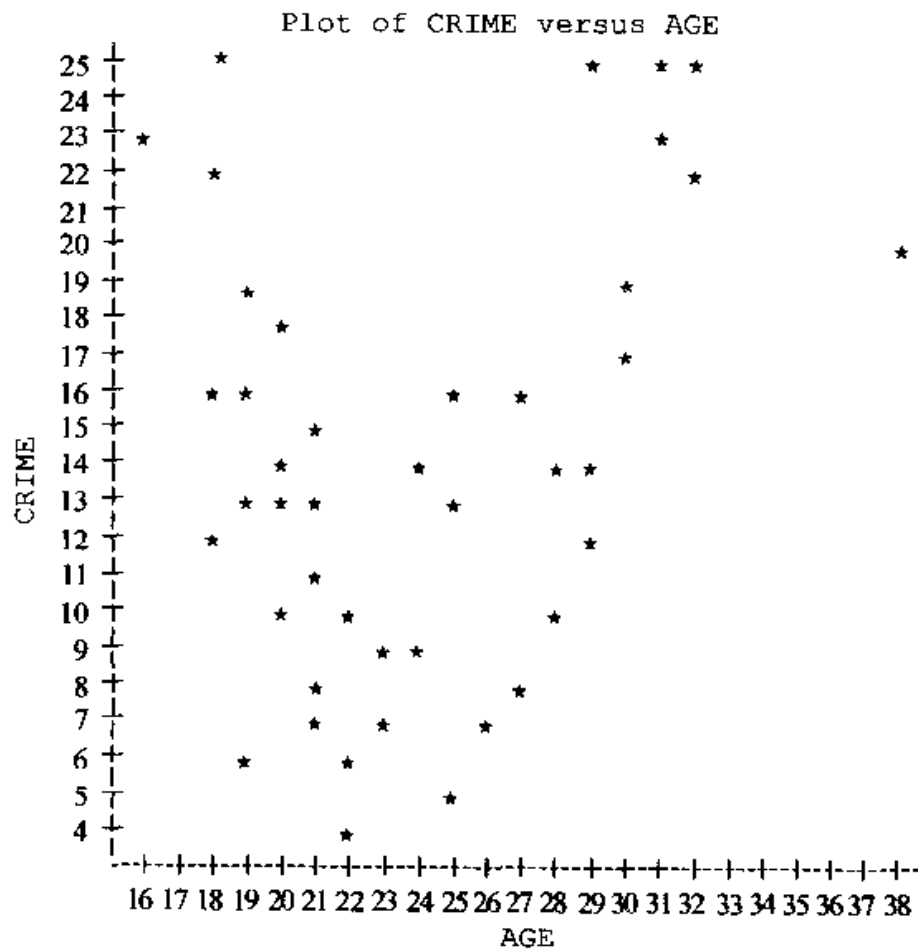
**13.3(社会)** 一位调查人员在一个州立大学访问了 45 名学生,问他们下列行动中的每一个是否为犯罪行为:恶性袭击,持械抢劫,纵火,无神论,盗车,入室盗窃,非暴力反抗,共产主义,吸毒成瘾,贪污,强奸,赌博,同性恋,诈骗,纳粹主义,贿赂,价格垄断,卖淫,猥亵儿童,性别歧视,入店行窃,罢工,露天采矿,叛国,故意破坏。对于每一个学生,调查人员记录下该生认为是犯罪的行为的数量 Crime,以及该生个人的其他信息(接受大学教育的年数 college,年龄 age,父母的收入 income,

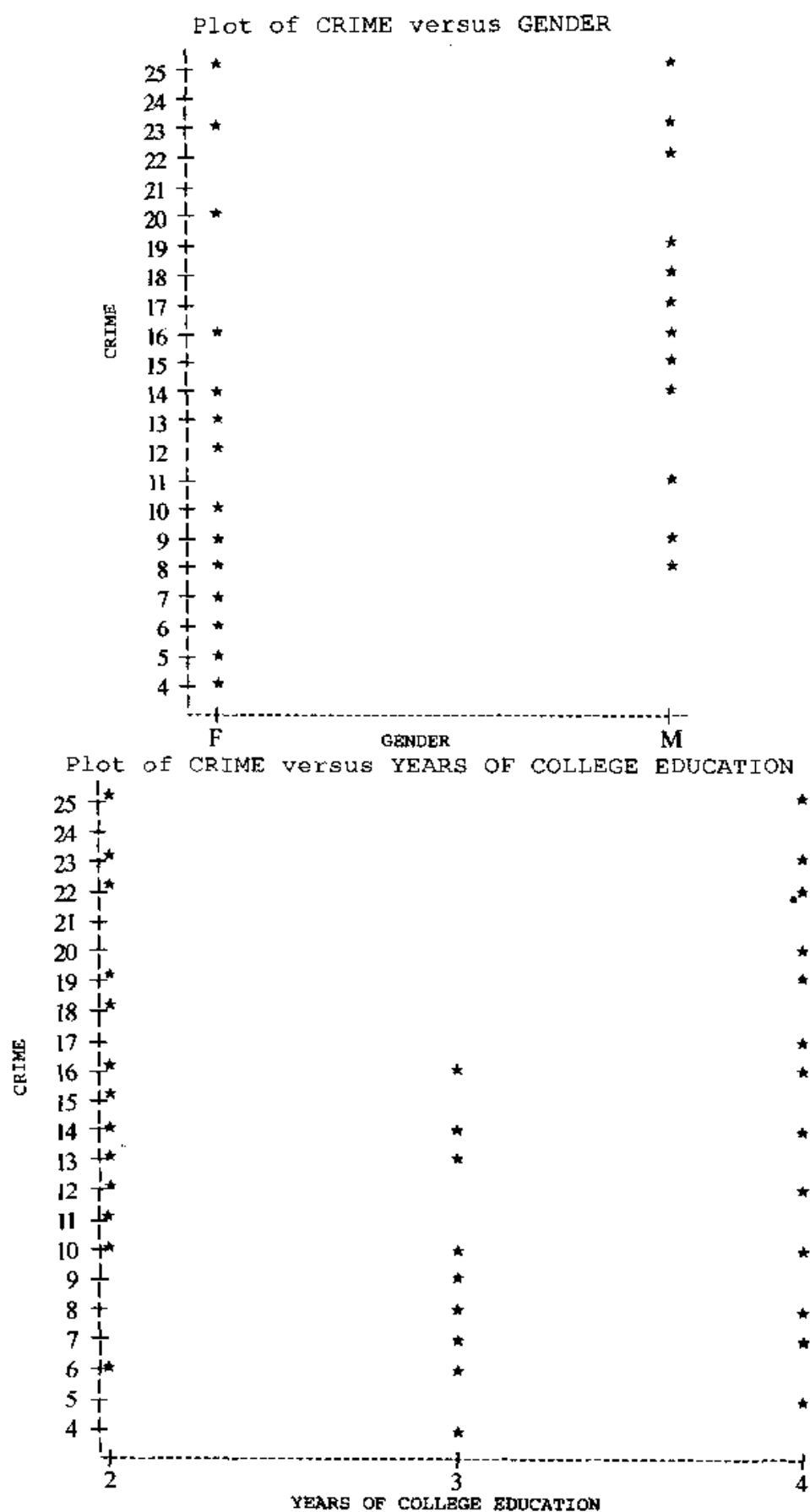
和性别 sex)。数据如下。参考给出的输出结果,识别潜在的共线性问题。用最优子集回归程序的输出结果确定哪些变量应该包含在回归模型中。你能就在访问中应该收集的其他变量给出建议吗?

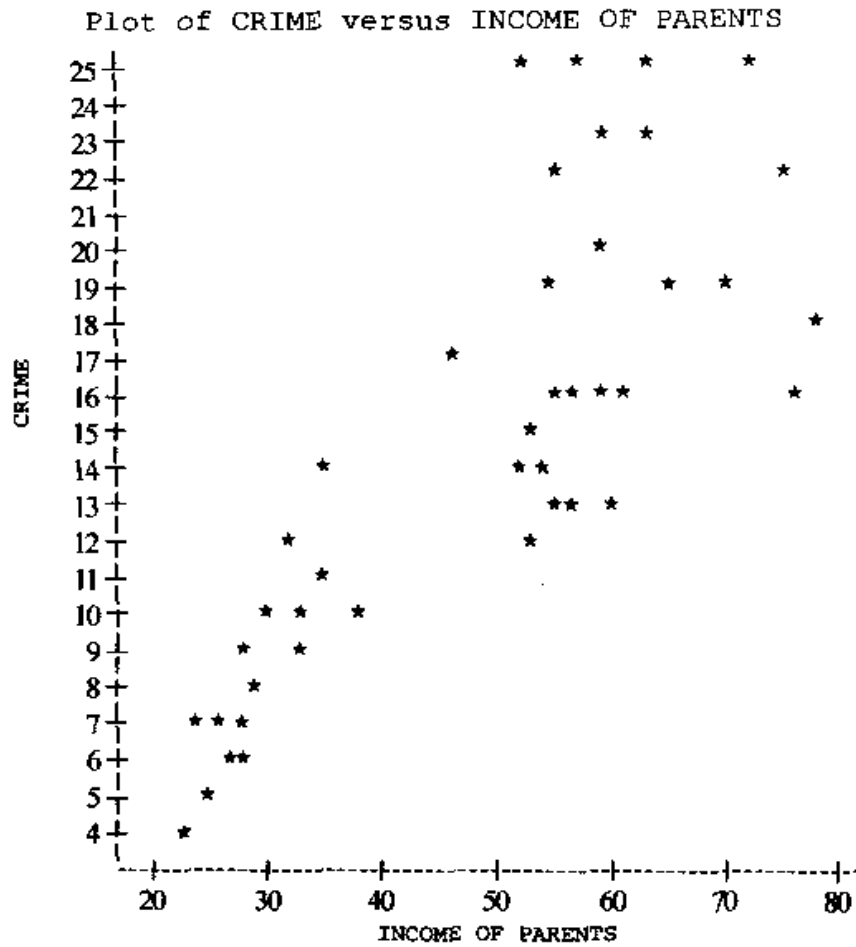
LISTING OF DATA

| OBS | CRIME | AGE | COLLEGE | INCOME | SEX |
|-----|-------|-----|---------|--------|-----|
| 1   | 23    | 16  | 2       | 63     | 1   |
| 2   | 25    | 18  | 2       | 72     | 1   |
| 3   | 22    | 18  | 2       | 75     | 1   |
| 4   | 16    | 18  | 2       | 61     | 0   |
| 5   | 19    | 19  | 2       | 65     | 1   |
| 6   | 19    | 19  | 2       | 70     | 1   |
| 7   | 18    | 20  | 2       | 78     | 1   |
| 8   | 16    | 19  | 2       | 76     | 0   |
| 9   | 12    | 18  | 2       | 53     | 0   |
| 10  | 13    | 19  | 2       | 56     | 0   |
| 11  | 16    | 19  | 2       | 59     | 1   |
| 12  | 13    | 20  | 2       | 55     | 0   |
| 13  | 13    | 21  | 2       | 60     | 0   |
| 14  | 14    | 20  | 2       | 52     | 0   |
| 15  | 14    | 24  | 3       | 54     | 0   |
| 16  | 13    | 25  | 3       | 55     | 0   |
| 17  | 16    | 25  | 3       | 55     | 0   |
| 18  | 16    | 27  | 4       | 56     | 1   |
| 19  | 14    | 28  | 4       | 52     | 1   |
| 20  | 20    | 38  | 4       | 59     | 0   |
| 21  | 25    | 29  | 4       | 63     | 1   |
| 22  | 19    | 30  | 4       | 55     | 1   |
| 23  | 23    | 31  | 4       | 59     | 0   |
| 24  | 25    | 32  | 4       | 52     | 1   |
| 25  | 22    | 32  | 4       | 55     | 1   |
| 26  | 25    | 31  | 4       | 57     | 0   |
| 27  | 17    | 30  | 4       | 46     | 1   |
| 28  | 14    | 29  | 4       | 35     | 0   |
| 29  | 12    | 29  | 4       | 32     | 0   |
| 30  | 10    | 28  | 4       | 30     | 0   |
| 31  | 8     | 27  | 4       | 29     | 0   |
| 32  | 7     | 26  | 4       | 28     | 0   |

| OBS | CRIME | AGE | COLLEGE | INCOME | SEX |
|-----|-------|-----|---------|--------|-----|
| 33  | 5     | 25  | 4       | 25     | 0   |
| 34  | 9     | 24  | 3       | 33     | 0   |
| 35  | 7     | 23  | 3       | 26     | 0   |
| 36  | 9     | 23  | 3       | 28     | 1   |
| 37  | 10    | 22  | 3       | 38     | 0   |
| 38  | 4     | 22  | 3       | 24     | 0   |
| 39  | 6     | 22  | 3       | 28     | 0   |
| 40  | 8     | 21  | 3       | 29     | 1   |
| 41  | 11    | 21  | 2       | 35     | 1   |
| 42  | 10    | 20  | 2       | 33     | 0   |
| 43  | 6     | 19  | 2       | 27     | 0   |
| 44  | 7     | 21  | 3       | 24     | 0   |
| 45  | 15    | 21  | 2       | 53     | 1   |







## Backward Elimination Procedure for Dependent Variable CRIME

Step 0 All Variables Entered R-square = 0.82783940 C(p) = 5.00000000

|            | DF | Sum of Squares | Mean Square  | F     | Prob>F |
|------------|----|----------------|--------------|-------|--------|
| Regression | 4  | 1301.62108953  | 325.40527238 | 48.09 | 0.0001 |
| Error      | 40 | 270.69002158   | 6.76725054   |       |        |
| Total      | 44 | 1572.31111111  |              |       |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | -10.82338752       | 2.39210442     | 138.54102767           | 20.47 | 0.0001 |
| AGE      | 0.43238152         | 0.20236447     | 30.89427247            | 4.57  | 0.0388 |
| COLLEGE  | -0.02399594        | 1.22148794     | 0.00261162             | 0.00  | 0.9844 |
| INCOME   | 0.29025487         | 0.03141812     | 577.57817022           | 85.35 | 0.0001 |
| SEX      | 2.45416550         | 0.87466592     | 53.27648156            | 7.87  | 0.0077 |

Bounds on condition number: 7.476669, 68.21544

Step 1 Variable COLLEGE Removed R-square = 0.82783774 C(p) = 3.00038592



|  | DF                 | Sum of Squares | Mean Square            | F      | Prob>F |
|--|--------------------|----------------|------------------------|--------|--------|
| Regression                                     | 3                  | 1301.61847791  | 433.87282597           | 65.72  | 0.0001 |
| Error  | 41                 | 270.69263320   | 6.60225935             |        |        |
| Total  | 44                 | 1572.31111111  |                        |        |        |
| Variable                                       | Parameter Estimate | Standard Error | Type II Sum of Squares | F      | Prob>F |
| INTERCEP                                       | -10.82193315       | 2.36163189     | 138.63658941           | 21.00  | 0.0001 |
| AGE  | 0.42872187         | 0.07806990     | 199.10244384           | 30.16  | 0.0001 |
| INCOME   | 0.29058236         | 0.02630415     | 805.71727230           | 122.04 | 0.0001 |
| SEX  | 2.45108843         | 0.84997169     | 54.90370062            | 8.32   | 0.0062 |
| Bounds on condition number: 1.202437, 10.21103 |                    |                |                        |        |        |

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination Procedure for Dependent Variable CRIME

| Step | Variable Removed | Number In | Partial R <sup>2</sup> | Model R <sup>2</sup> | C(p)   | F      | Prob>F |
|------|------------------|-----------|------------------------|----------------------|--------|--------|--------|
| 1    | COLLEGE          | 3         | 0.0000                 | 0.8278               | 3.0004 | 0.0004 | 0.9844 |

13.4 参见练习 13.3。从一个逐步回归程序得到如下的输出。把这些结果与练习 13.3 中的结果相比较。

#### REGRESSION ANALYSIS, FORWARD SELECTION

Forward Selection Procedure for Dependent Variable CRIME

Step 1 Variable INCOME Entered R-square = 0.66453936 C(p) = 36.94132731

|                                  | DF                 | Sum of Squares | Mean Square            | F     | Prob>F |
|----------------------------------|--------------------|----------------|------------------------|-------|--------|
| Regression                       | 1                  | 1044.86262180  | 1044.86262180          | 85.18 | 0.0001 |
| Error                            | 43                 | 527.44848931   | 12.26624394            |       |        |
| Total                            | 44                 | 1572.31111111  |                        |       |        |
| Variable                         | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
| INTERCEP                         | -0.19647505        | 1.66089569     | 0.17164917             | 0.01  | 0.9064 |
| INCOME                           | 0.30177022         | 0.03269660     | 1044.86262180          | 85.18 | 0.0001 |
| Bounds on condition number: 1, 1 |                    |                |                        |       |        |

Step 2 Variable AGE Entered R-square = 0.79291863 C(p) = 9.11353325

|            | DF | Sum of Squares | Mean Square  | F     | Prob>F |
|------------|----|----------------|--------------|-------|--------|
| Regression | 2  | 1246.71477730  | 623.35738865 | 80.41 | 0.0001 |
| Error      | 42 | 325.59633381   | 7.75229366   |       |        |
| Total      | 44 | 1572.31111111  |              |       |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F      | Prob>F |
|----------|--------------------|----------------|------------------------|--------|--------|
| INTERCEP | -11.33832496       | 2.55169607     | 153.06296650           | 19.74  | 0.0001 |
| AGE      | 0.43163600         | 0.08458942     | 201.85215549           | 26.04  | 0.0001 |
| INCOME   | 0.32018698         | 0.02624270     | 1154.03879316          | 148.86 | 0.0001 |

Bounds on condition number: 1.01928, 4.077119

Step 3 Variable SEX Entered R-square = 0.82783774 C(p) = 3.00038592

|            | DF | Sum of Squares | Mean Square  | F     | Prob>F |
|------------|----|----------------|--------------|-------|--------|
| Regression | 3  | 1301.61847791  | 433.87282597 | 65.72 | 0.0001 |
| Error      | 41 | 270.69263320   | 6.60225935   |       |        |
| Total      | 44 | 1572.31111111  |              |       |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F      | Prob>F |
|----------|--------------------|----------------|------------------------|--------|--------|
| INTERCEP | -10.82193315       | 2.36163189     | 138.63658941           | 21.00  | 0.0001 |
| AGE      | 0.42872187         | 0.07806990     | 199.10244384           | 30.16  | 0.0001 |
| INCOME   | 0.29058236         | 0.02630415     | 805.71727230           | 122.04 | 0.0001 |
| SEX      | 2.45108843         | 0.84997169     | 54.90370062            | 8.32   | 0.0062 |

Bounds on condition number: 1.202437, 10.21103

No other variable met the 0.5000 significance level for entry into the model.

Summary of Forward Selection Procedure for Dependent Variable CRIME

| Step | Variable Entered | Number In | Partial R <sup>2</sup> | Model R <sup>2</sup> | C(p)    | F       | Prob>F |
|------|------------------|-----------|------------------------|----------------------|---------|---------|--------|
| 1    | INCOME           | 1         | 0.6645                 | 0.6645               | 36.9413 | 85.1820 | 0.0001 |
| 2    | AGE              | 2         | 0.1284                 | 0.7929               | 9.1135  | 26.0377 | 0.0001 |
| 3    | SEX              | 3         | 0.0349                 | 0.8278               | 3.0004  | 8.3159  | 0.0062 |

**13.5(农业)** 一个公司对食物添加剂(蛋白质和抗生素)对于牲畜达到上市的重量所花的喂养时间的影响感兴趣。试讨论为了得到一个预测达到上市重量所花时间的多元回归方程,需要考虑哪些变量?

### 13.3 模型形式的确定(第二步)

在前一节中,我们介绍了几种为给定的回归问题选择候选自变量的方法。通过向所研究问题的专家请教,我们可以得到一个起始点,并使用某一种挑选方法来得到一个可能的多元回归模型。本节主要介绍如何精选第一步中得到的信息,以得到一个有用的多元回归模型。

虽然在多元回归中已经有了包含  $k$  个候选自变量的子集和因变量  $y$ , 我们还不清楚因变量和自变量之间的真正关系。设在假定的回归模型中,  $y$  对  $x_1, x_2, \dots, x_k$  比真正的模型有更低的阶。那么,只要在每个自变量的因子水平组合上有多于一个的观测值,我们就可以用第十一章讨论过的方程  $F = MS_{\text{Lack}}/MS(\text{残差})$  来检验拟合的多项式模型的不充分性。

另一种检验所假定的(拟合)模型的方法是看残差  $(y_i - \hat{y}_i)$  对  $x_j$  的散点图。例如,设第一步选出变量  $x_1, x_2$ , 和  $x_3$  为响应值  $y$  的一个合理的回归方程的自变量子集,但不知道是一个什么样的多项式函数,我们首先使用多元线性回归模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

来得到最小二乘方程  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$ , 从残差  $(y_i - \hat{y}_i)$  对每个  $x$  的残差图可以看出哪一个变量取更高的次数可能比较合适。我们先指出对一个自变量的回归问题如何使用这一方法,再推广到多元回归的情况。

#### 例 13.9

在放射免疫测定中,一种作了放射性标记的激素被加入到一个装有专门对这一激素的抗体的试管里。这两种物质会化合成一种抗原-抗体的复合体。为了测量激素同抗体的反应程度,我们测量被抗体制约的激素的量与游离状态激素的量的比值。一般,实验者在试管中放入不同剂量的激素  $x$  (下面结果用 DOSE 表示),并对每个试管中受限与游离的放射性标记的比例  $y$  (BOUND/FREE COUNT) 进行测定。通常情况下,  $y$  与  $x$  是接近线性的。表 13.4 为放射免疫测定实验中,11 支试管的数据。

表 13.4 放射免疫测定实验数据

| 受限/游离  | 剂量(浓度) | 受限/游离   | 剂量(浓度) |
|--------|--------|---------|--------|
| 9.900  | 0.00   | 62.045  | 1.50   |
| 10.465 | 0.25   |         |        |
| 10.312 | 0.50   | 78.327  | 1.75   |
| 13.633 | 0.75   | 90.307  | 2.00   |
| 20.784 | 1.00   | 97.348  | 2.25   |
| 36.164 | 1.25   | 102.686 | 2.50   |

a. 做样本的散点图并拟合线性回归模型

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

b. 画出残差对标记和对  $\hat{y}$  的图。线性模型是否充分地拟合了数据?

c. 提出一个更好的模型。

解答 计算机的输出如下:

| Data | Display    |         |      |        |
|------|------------|---------|------|--------|
| ROW  | BOUND/FREE | COUNT   | DOSE | DOSE_2 |
| 1    |            | 9.900   | 0.00 | 0.0000 |
| 2    |            | 10.465  | 0.25 | 0.0625 |
| 3    |            | 10.312  | 0.50 | 0.2500 |
| 4    |            | 13.633  | 0.75 | 0.5625 |
| 5    |            | 20.784  | 1.00 | 1.0000 |
| 6    |            | 36.164  | 1.25 | 1.5625 |
| 7    |            | 62.045  | 1.50 | 2.2500 |
| 8    |            | 78.327  | 1.75 | 3.0625 |
| 9    |            | 90.307  | 2.00 | 4.0000 |
| 10   |            | 97.348  | 2.25 | 5.0625 |
| 11   |            | 102.686 | 2.50 | 6.2500 |

Regression Analysis: BOUND/FREE COUNT versus DOSE

The regression equation is

BOUND/FREE COUNT = -7.19 + 44.4 DOSE

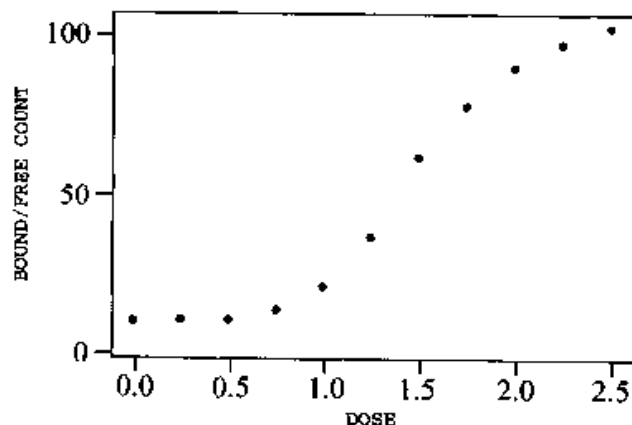
| Predictor | Coef   | SE Coef | T     | P     |
|-----------|--------|---------|-------|-------|
| Constant  | -7.189 | 6.226   | -1.15 | 0.278 |
| DOSE      | 44.440 | 4.210   | 10.56 | 0.000 |

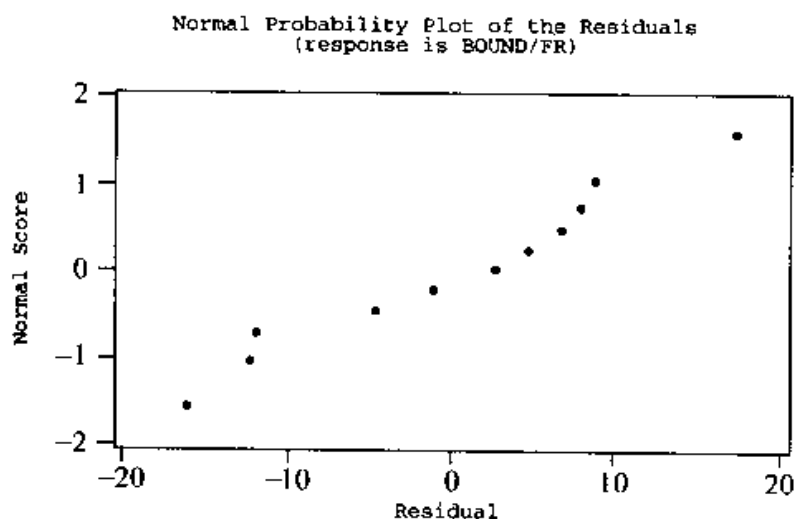
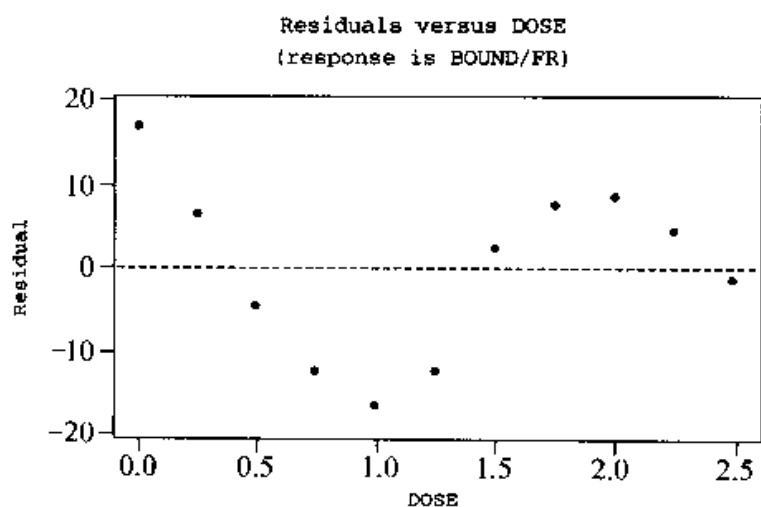
S = 11.04      R-Sq = 92.5 %      R-Sq(adj) = 91.7 %

Analysis of Variance

| Source         | DF | SS    | MS    | F      | P     |
|----------------|----|-------|-------|--------|-------|
| Regression     | 1  | 13577 | 13577 | 111.44 | 0.000 |
| Residual Error | 9  | 1097  | 122   |        |       |
| Total          | 10 | 14674 |       |        |       |

Plot of BOUND/FREE COUNT versus DOSE





a, b 线性拟合方程为

$$\hat{y} = -7.189 + 44.440x$$

$y$ (比值)对  $x$ (含量百分比)的图清楚地表明线性回归模型的不足,残差图也证实了这种不足。线性回归在含量百分比刻度的上、下两端低估了比率,而在含量百分比的中部高估了比率。

c. 一个可能的改进模型为二次模型

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

在本章的后面会对这一例子作进一步讨论。

除了两个自变量的情况外,散点图对查出自变量间的交互作用不是很有效,因为对大多数实际问题,有太多的变量,以至于很难表示出变量间的关系,并把它们对响应值  $y$  的联合作用二维散点图表示出来。也许最合理的建议是使用前一节介绍的最优子集回归方法和用候选自变量进行一些“尝试-改正”的拟合过程,再加上

一些常识来决定在多元回归模型中使用哪些交叉项。

虚拟变量(定性自变量)的出现在对多项式模型的拟合的判断中不起主要作用。需要记住的重要的事是当回归模型中既有定量变量,又有虚拟变量时,对每一组虚拟变量的取值,我们都对定量变量得到一个回归模型。所以,查找拟合不足的画图法必须分别对每一组虚拟变量的取值做。通过仔细分析这些图,我们也能分辨出不同虚拟变量给多项式模型带来的可能的不同。

### 例 13.10

一个公司的分析员希望建立一个回归模型来预测某一地区某一品牌汽车普通型和豪华型的销售情况。理论讨论和公司以前在这一地区销售情况的调查表明,汽车贷款利率(interest rate)和每加仑汽油的价格(price/gallon gasoline)是关键的预测变量。如下是汽油动力的普通型(standard car)和豪华型汽车(luxury car)在前 18 个月中每月的销售量(sales)(以 1,000 为单位)。拟合一个线性回归模型,并用残差图来决定(如果有的话)哪一变量要有更高幕次的项。对普通型和豪华型的结论一样吗?为多元回归方程提出可能要增加的项。

**解答** 回归模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

其中

$y$  = 每月销量(以 1,000 为单位)

$x_1$  = 每加仑油价

$x_2$  = 利率

$x_3 = \begin{cases} 1, & \text{普通型} \\ 0, & \text{豪华型} \end{cases}$  (下面数据表中用 TYP CAR 表示)

从输出中知,回归方程为

$$\hat{y} = 56.074 - 16.144x_1 - 2.332x_2 + 14.422x_3$$

把  $x_3 = 0$  和 1 代入方程,我们得到普通型和豪华型的回归方程分别为;

$x_3 = 0$  (豪华型)

$$\hat{y} = 56.074 - 16.144x_1 - 2.332x_2$$

$x_3 = 1$  (普通型)

$$\hat{y} = 56.074 - 16.144x_1 - 2.332x_2 + 14.422$$

$$= 70.496 - 16.144x_1 - 2.332x_2$$

在两个模型  $y$  对  $x_1$  和  $x_2$  的图中可以看出,销量与每加仑汽油价格和利率间存在非线性关系,但对普通型模型的偏差更大,这从两个模型的残差图中可以看出。

## MULTIPLE REGRESSION ANALYSIS

## DATA LISTING

| MONTH | MONTHLY<br>SALES<br>(1000) | PRICE/<br>GALLON<br>GASOLINE | INTEREST<br>RATE( % ) | TYPECAR |
|-------|----------------------------|------------------------------|-----------------------|---------|
| 1     | 22.1                       | 1.89                         | 6.1                   | 1       |
| 1     | 7.2                        | 1.89                         | 6.1                   | 0       |
| 2     | 15.4                       | 1.94                         | 6.2                   | 1       |
| 2     | 5.4                        | 1.94                         | 6.2                   | 0       |
| 3     | 11.7                       | 1.95                         | 6.3                   | 1       |
| 3     | 7.6                        | 1.95                         | 6.1                   | 0       |
| 4     | 10.3                       | 1.82                         | 8.2                   | 1       |
| 4     | 2.5                        | 1.82                         | 8.2                   | 0       |
| 5     | 11.4                       | 1.85                         | 9.8                   | 1       |
| 5     | 2.4                        | 1.85                         | 9.8                   | 0       |
| 6     | 7.5                        | 1.78                         | 10.3                  | 1       |
| 6     | 1.7                        | 1.78                         | 10.3                  | 0       |
| 7     | 13.0                       | 1.76                         | 10.5                  | 1       |
| 7     | 4.3                        | 1.76                         | 10.5                  | 0       |
| 8     | 12.8                       | 1.76                         | 8.7                   | 1       |
| 8     | 3.7                        | 1.76                         | 8.7                   | 0       |
| 9     | 14.6                       | 1.75                         | 7.4                   | 1       |
| 9     | 3.9                        | 1.75                         | 7.4                   | 0       |
| 10    | 18.9                       | 1.74                         | 6.9                   | 1       |
| 10    | 7.0                        | 1.74                         | 6.9                   | 0       |
| 11    | 19.3                       | 1.70                         | 5.2                   | 1       |
| 11    | 6.8                        | 1.70                         | 5.2                   | 0       |
| 12    | 30.1                       | 1.70                         | 4.9                   | 1       |
| 12    | 10.1                       | 1.70                         | 4.9                   | 0       |
| 13    | 28.2                       | 1.68                         | 4.3                   | 1       |
| 13    | 9.4                        | 1.68                         | 4.3                   | 0       |
| 14    | 25.6                       | 1.60                         | 3.7                   | 1       |
| 14    | 7.9                        | 1.60                         | 3.7                   | 0       |
| 15    | 37.5                       | 1.61                         | 3.6                   | 1       |
| 15    | 14.1                       | 1.61                         | 3.6                   | 0       |
| 16    | 36.1                       | 1.64                         | 3.1                   | 1       |
| 16    | 14.5                       | 1.64                         | 3.1                   | 0       |
| 17    | 39.8                       | 1.67                         | 1.8                   | 1       |
| 17    | 14.9                       | 1.67                         | 1.8                   | 0       |
| 18    | 44.3                       | 1.68                         | 2.3                   | 1       |
| 18    | 15.6                       | 1.68                         | 2.3                   | 0       |

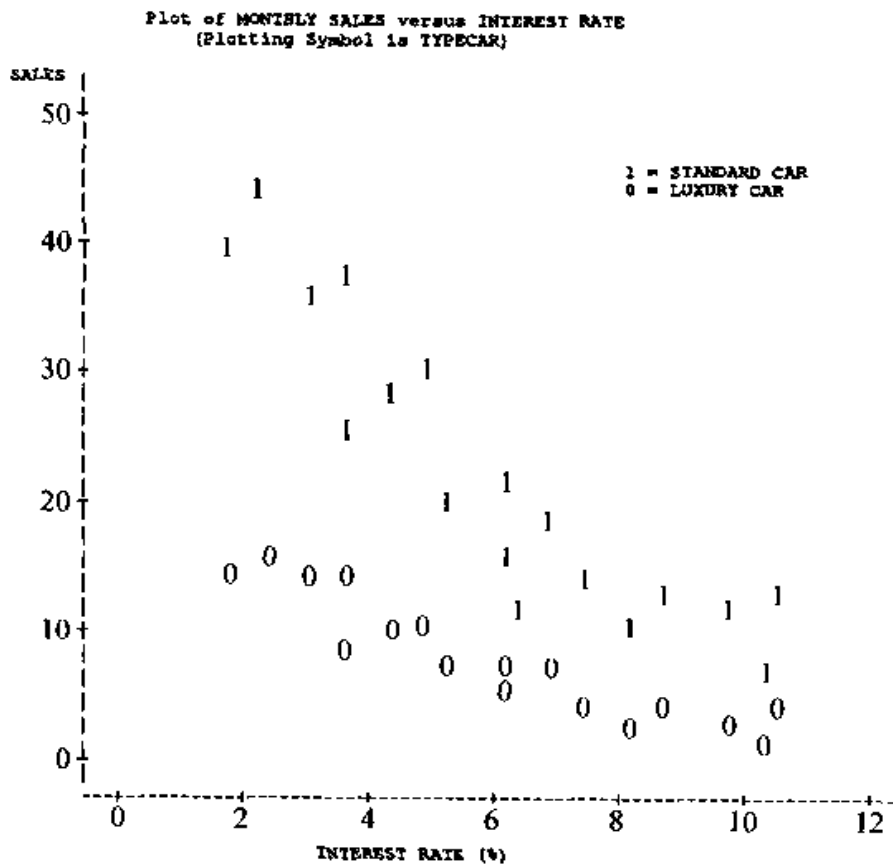
Dependent Variable;SALES

## Analysis of Variance

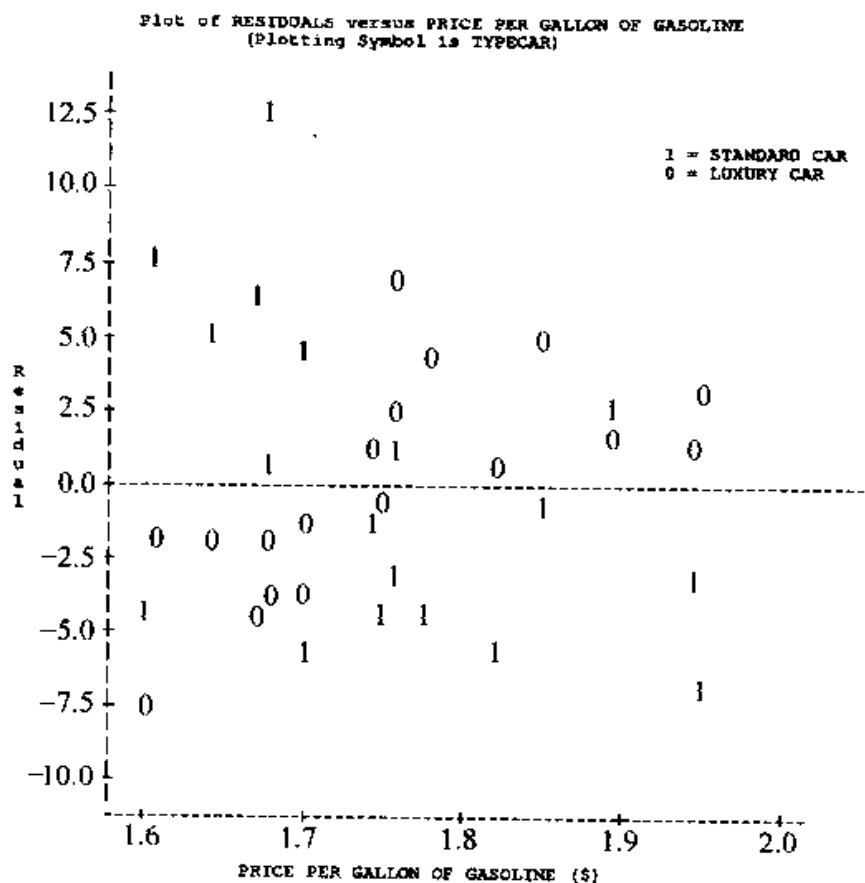
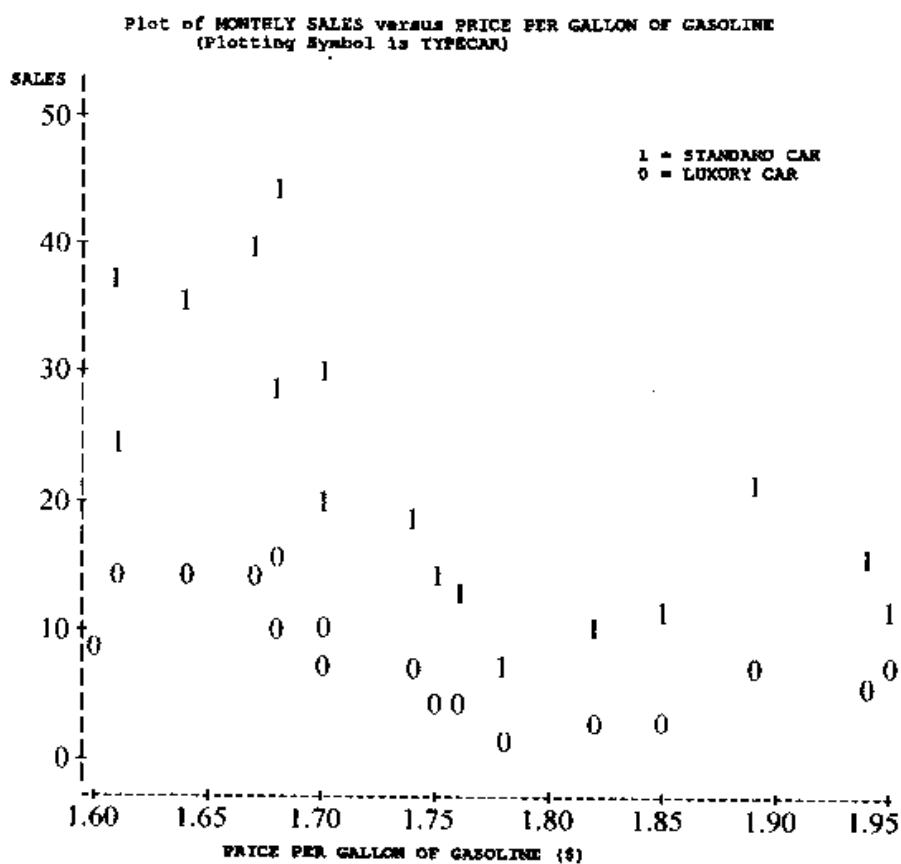
| Source   | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|----------|----|----------------|-------------|---------|--------|
| Model    | 3  | 3716.34235     | 1238.78078  | 54.900  | 0.0001 |
| Error    | 32 | 722.05765      | 22.56430    |         |        |
| C Total  | 35 | 4438.40000     |             |         |        |
| Root MSE |    | 4.75019        | R-square    | 0.8373  |        |
| Dep Mean |    | 14.93333       | Adj R-sq    | 0.8221  |        |
| C. V.    |    | 31.80931       |             |         |        |

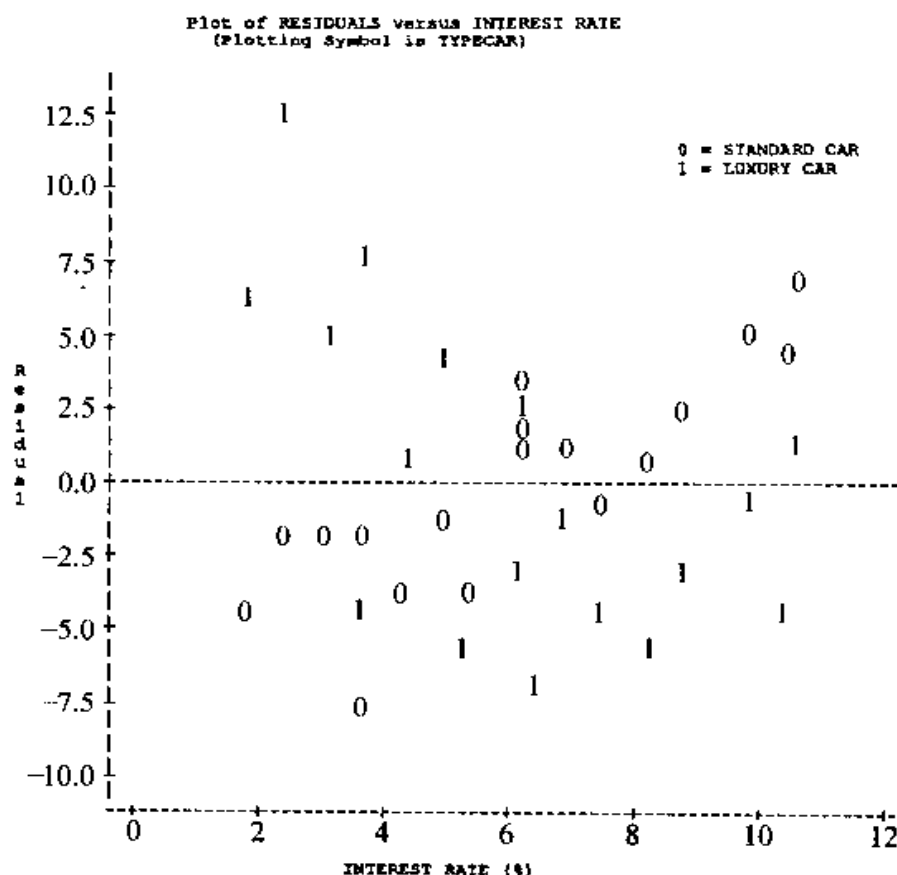
## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|----------------------------|---------|
| INTERCEP | 1  | 50.620740          | 15.18488648    | 3.334                      | 0.0022  |
| GASPRICE | 1  | -16.436642         | 9.26565678     | -1.774                     | 0.0856  |
| INTEREST | 1  | -2.328968          | 0.36053082     | -6.460                     | 0.0001  |
| TYPECAR  | 1  | 14.448100          | 1.58340161     | 9.125                      | 0.0001  |









对于豪华型,每加仑价格和利率的残差图表明,当  $x_1$  和  $x_2$  较小时,有低估现象,当  $x_1$  和  $x_2$  较大时,则有高估现象;相应的,普通型车的残差图则表明对数据拟合得很好,虽然也存在一些弯曲表明回归模型中需要加入  $x_1$  和  $x_2$  的更高次项。一个具有形式

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \beta_5 x_3 + \beta_6 x_1 x_3 + \beta_7 x_1^2 x_3 \\ + \beta_8 x_2 x_3 + \beta_9 x_2^2 x_3 + \epsilon$$

的回归模型就考虑到  $y$  与  $x_1$  和  $x_2$  关系中的弯曲。这个模型对不同的车型也有不同的回归系数。有些读者也许还会考虑到加入两个定量自变量的交叉项。这个模型的部分输出如下:

MULTIPLE REGRESSION ANALYSIS

Dependent Variable: MONTHLY SALES

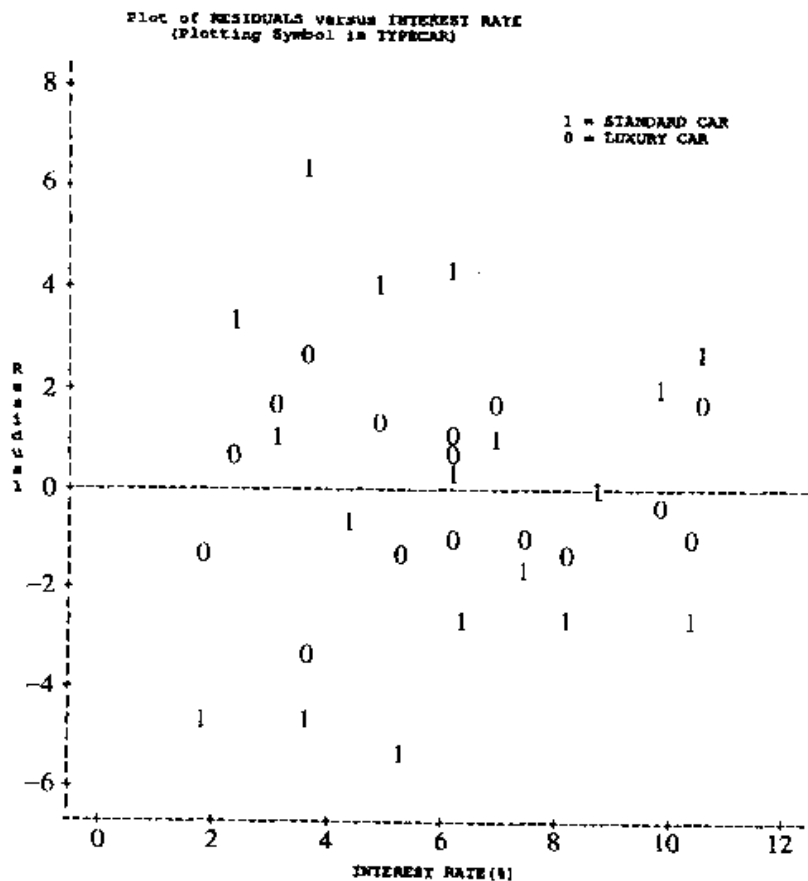
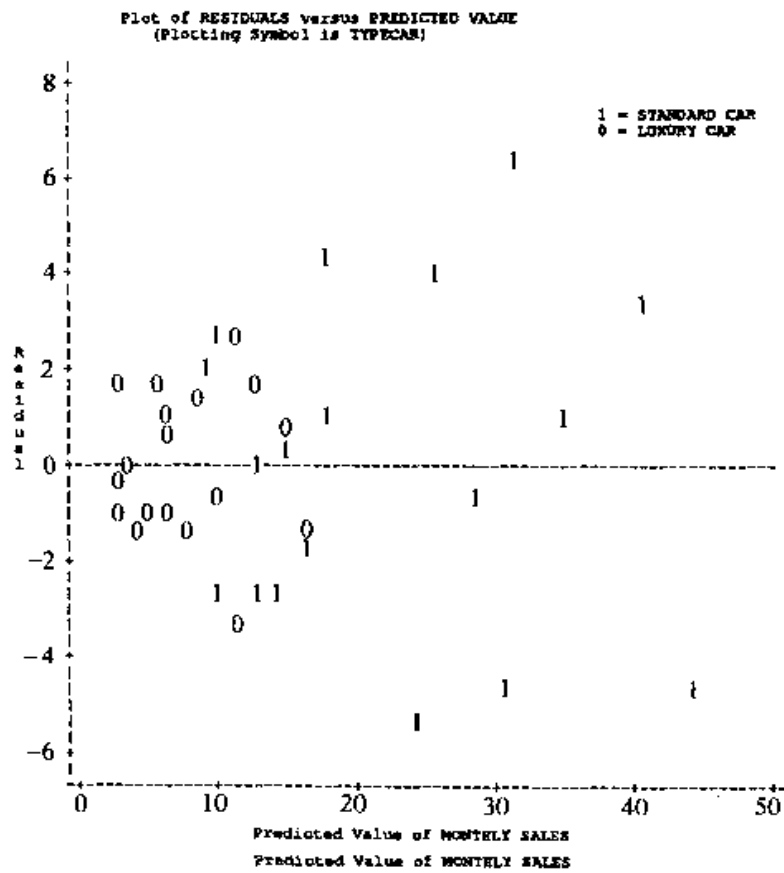
Analysis of Variance

| Source   | DF       | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|----------|----------|-------------------|----------------|---------|--------|
| Model    | 9        | 4203.84688        | 467.09410      | 51.777  | 0.0001 |
| Error    | 26       | 234.55312         | 9.02127        |         |        |
| C Total  | 35       | 4438.40000        |                |         |        |
| Root MSE | 3.00354  | R-square          | 0.9472         |         |        |
| Dep Mean | 14.93333 | Adj R-sq          | 0.9289         |         |        |
| C.V.     | 20.11301 |                   |                |         |        |

Parameter Estimates

| Variable   | DF | Parameter<br>Estimate | Standard<br>Error | T for H0:<br>Parameter = 0 | Prob> T |
|------------|----|-----------------------|-------------------|----------------------------|---------|
| INTERCEP   | 1  | - 3.217629            | 276.34988059      | - 0.012                    | 0.9908  |
| PRICE      | 1  | 28.494610             | 310.03429094      | 0.092                      | 0.9275  |
| INTEREST   | 1  | - 3.490166            | 1.58442091        | - 2.203                    | 0.0367  |
| TYPECAR    | 1  | - 300.022301          | 389.84323007      | - 0.770                    | 0.4485  |
| PRICE _ 2  | 1  | - 7.966301            | 86.26524998       | - 0.092                    | 0.9271  |
| INTER _ 2  | 1  | 0.153829              | 0.11736084        | 1.311                      | 0.2014  |
| PR _ TYPE  | 1  | 389.508123            | 437.32299428      | 0.891                      | 0.3813  |
| PR2 _ TYPE | 1  | - 112.642229          | 121.67435792      | - 0.926                    | 0.3631  |
| IN _ TYPE  | 1  | - 5.030243            | 2.24082386        | - 2.245                    | 0.0335  |
| IN2 _ TYPE | 1  | 0.217441              | 0.16597436        | 1.310                      | 0.2016  |

| Variable   | DF | Variable<br>Label                  |
|------------|----|------------------------------------|
| INTERCEP   | 1  | Intercept                          |
| PRICE      | 1  | PRICE OF GALLON OF GASOLINE        |
| INTEREST   | 1  | INTEREST RATE( % )                 |
| TYPECAR    | 1  | TYPE OF CAR                        |
| PRICE _ 2  | 1  | PRICE SQUARED                      |
| INTER _ 2  | 1  | INTEREST SQUARED                   |
| PR _ TYPE  | 1  | PRICE TIMES TYPE OF CAR            |
| PR2 _ TYPE | 1  | PRICE SQUARED TIMES TYPE OF CAR    |
| IN _ TYPE  | 1  | INTEREST TIMES TYPE OF CAR         |
| IN2 _ TYPE | 1  | INTEREST SQUARED TIMES TYPE OF CAR |



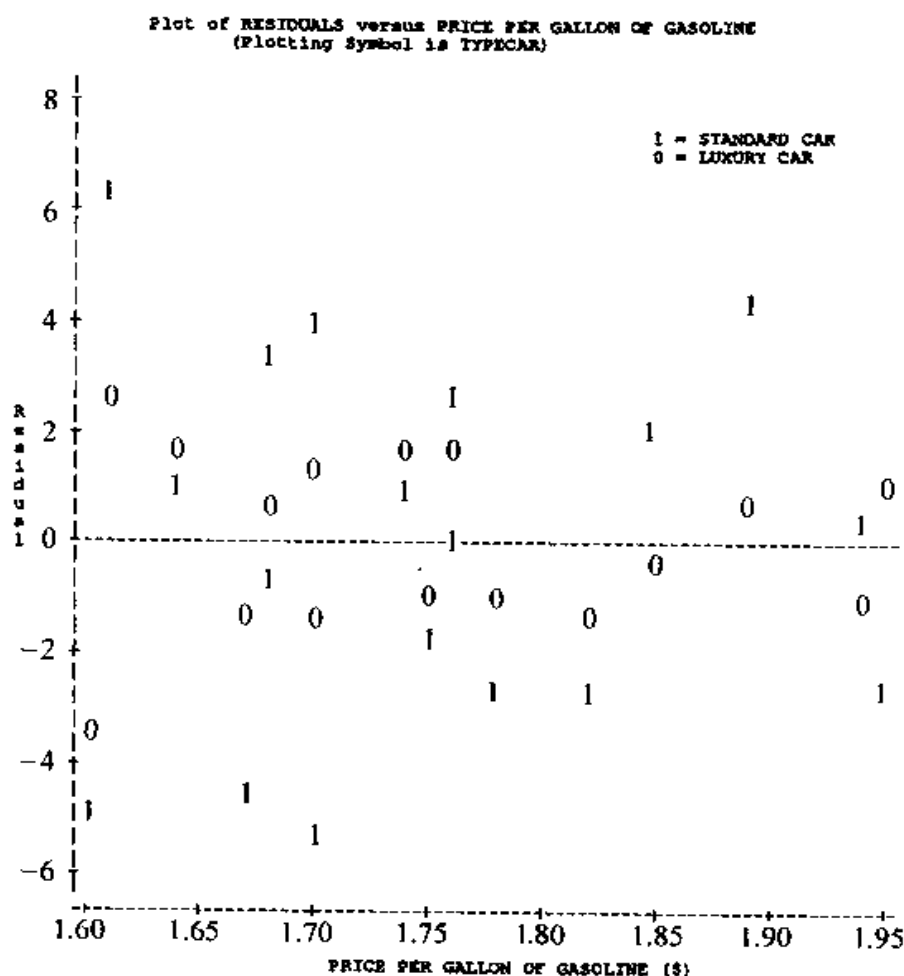
到目前为止,我们对拟合不足的原因只考虑了多项式系数和交叉项问题。而实际上,有时拟合不足是由于  $y$  不能表示为自变量子集元素的多项式的形式,这时,  $\beta$  的非线性模型可能更合适。

大多数基础经济学书中都指出,如果数量  $y$  在每单位时间以  $r$  的速率增长,在时刻  $t$ ,  $y$  的值为

$$y_t = y_0 e^{rt}$$

其中  $y_0$  为初始值。这一关系可以通过对数变换转化成线性关系:

$$\log y_t = \log y_0 + rt$$



记  $\beta_0 = \log y_0$ ,  $\beta_1 = r$ , 使用第十一章讲述的简单线性模型即可对数据进行拟合。当  $y$  是一个经济学变量,如总销量时,对数变换常被用于多元回归模型:

$$\log y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon$$

Cobb-Douglas 产量函数是另一类可以转化为回归方程的例子:

$$y = cl^{\alpha} k^{\beta}$$

其中,  $y$  是产量,  $l$  是劳动力投入,  $k$  是资金投入,  $\alpha$  和  $\beta$  是未知常数。同样, 我们对方程作对数变换, 得:

$$\begin{aligned}\log y &= (\log c) + \alpha(\log l) + \beta(\log k) \\ &= \beta_0 + \beta_1(\log l) + \beta_2(\log k)\end{aligned}$$

它表示对数产量对对数劳动量和对数资金的回归是线性的。

### 例 13.11

经济学的一个重要概念是需求的价格弹性, 即价格每变化一个百分点, 需求数量变化的百分点的负值。研究表明价格弹性为 1 表示一个(小的)价格变动不会引起总收入的变化。一个非弹性的需求(弹性小于 1)表示价格的微小增量会引起收入的增加, 弹性需求则相反。

以下是每天对莴苣的需求量  $y$  (每 100 人消费的包数), 对不同价格  $x$  (每包) 的数据。其他一些影响需求量的因素都不变, 如, 商店都位于一般居住地区, 没有竞争者进行莴苣促销等等。数据是

|        |        |        |        |        |        |        |        |        |        |        |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| $x$ :  | 0.79   | 0.79   | 0.84   | 0.84   | 0.89   | 0.89   | 0.94   | 0.94   | 0.99   | 0.99   |
| $y$ :  | 40.2   | 37.1   | 37.4   | 34.9   | 32.8   | 35.5   | 30.6   | 34.2   | 31.2   | 29.8   |
| $xy$ : | 31.758 | 29.309 | 31.416 | 29.316 | 29.192 | 31.595 | 28.764 | 32.148 | 30.888 | 29.502 |

- $xy$  表示什么经济量?
- 当  $x$  增加时,  $xy$  有没有变化的趋势?
- 如果  $xy$  是常数,  $\log x + \log y$  是否也为常数?
- 如果一种产品的价格弹性为 1,  $\log y$  对  $\log x$  的回归方程是什么样子?

### 解答

- $xy$  = 每包价格  $\times$  每 100 个人消费的包数, 所以, 它表示每 100 人的消费额。
- 数据图没有明显的趋势, 收入 ( $xy$ ) 接近于常数。
- 因为  $\log xy = \log \text{常数} = \log x + \log y$ , 所以  $\log x + \log y$  也为常数。
- 价格弹性为 1 表示  $\log y = \text{常数} - \log x$ , 回归方程以  $\log y$  为因变量,  $\log x$  为自变量, 其斜率为  $-1$  (加或减一个随机变量  $\epsilon$ )。因此, 在弹性研究中,  $\log y$  和  $\log x$  的回归模型是很有用的。

对数变换只是可能的变换中的一种, 也是很有用的一种, 因为, 对数变换把乘的关系转换为加的关系。自然对数(底为  $e = 2.7182818$ ), 也记为  $\ln(y)$ , 是特别有用的, 因为结果大体上表示了百分率变化。例如, 如果对中学教师收入的预测为  $\ln(\text{收入}) = \text{常数} + 0.042(\text{工作时间}) + \text{其他}$ , 那么, 增加一年工作时间, “其他”项保持不变, 收入的预测值约增加 4.2%, 这种对应并不精确, 但当值小于或等于 0.2

时,是很接近的。

另一种常用的变换是倒数变换,  $1/y$ , 例如, 如果  $y$  是按米每秒计算的速度,  $1/y$  就是每米需要的时间。这个变换对变化剧烈的曲线很有用, 对数变换对平缓的曲线较有用。使用计算机, 可以很容易的同时得到它们。当因变量先升至最高, 又下降时, 一个很有用的变换是平方变换,  $x^2$ , 在这个变换中, 不是把  $x$  用  $x^2$  代替, 而是把它们都当作自变量。当因变量先达到最小, 再上升时, 同时使用  $x$  和  $x^2$  也是很有效的。关于可能的变换的更广泛的讨论见 Tukey(1977)。

### 例 13.12

重新考虑例 13.2, 温度的作用也许不是线性的, 使用从那个例子中得到的自变量可以计算出一个回归模型。图 13.3 画出了残差对平均温度的图, 图中也画出了一条 LOWESS 平滑线。这个图是否表明需要加入平方项?

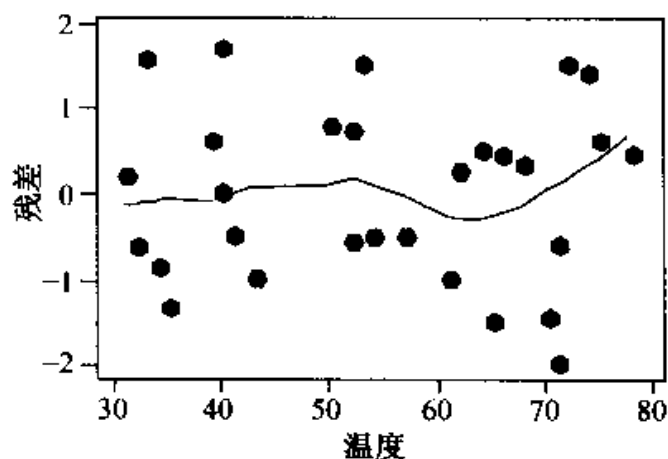


图 13.3 微型计算机数据的残差图

**解答** 图中并没有什么很明显的問題, LOWESS 曲线并不像抛物线(那是平方项的形状)。在右侧有一些点看上去会产生一条曲线。我们不好说此图是使用二次项带来了什么附加的依据。

(本节以下部分为选学内容)

我们用计算机软件和输出来说明非线性模型的拟合问题。我们要做的推理同一般线性模型的最小二乘类似, 事实上, 这个过程有时也被称为非线性最小二乘。同以前一样, 定义残差平方和:

$$SS(\text{残差}) = \sum_i (y_i - \hat{y}_i)^2$$

问题是, 找一个方法来得到估计值  $\hat{a}_1, \hat{a}_2, \dots$ , 使得  $SS(\text{残差})$  最小。用来寻找估计

的方程组也称为正规方程组,但与一般线性模型的最小二乘法不同的是,正规方程的形式依赖于所使用的非线性模型,而且,因为正规方程包括参数的非线性函数,它们的解也是很复杂的。由于这些困难,人们研究了很多迭代算法来求正规方程的解。

有微积分基础的读者可以知道,非线性模型的正规方程涉及对非线性方程的各个参数  $\alpha_i$  分别求偏导。当然,大多数软件(如 SAS、NONLIN、Splus)都可以求导数,因此无需给出正规方程的形式,而只需给出非线性方程的形式。我们用前面的一个例子的数据来说明。

回想在例 13.9 中,我们用放射免疫测定数据拟合了线性回归模型,模型的残差图表示加入一个二次项可能更合适:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$$

修改后的模型计算结果如下,注意到,在残差图中周期性依然很明显,因此二次模型也拟合得不很好。

Regression Analysis: BOUND/FREE COUNT versus DOSE, DOSE\_2

The regression equation is

BOUND/FREE COUNT = 2.88 + 17.6 DOSE + 10.7 DOSE\_2

| Predictor | Coef   | SE Coef | T    | P     |
|-----------|--------|---------|------|-------|
| Constant  | 2.884  | 7.175   | 0.40 | 0.698 |
| DOSE      | 17.58  | 13.35   | 1.32 | 0.225 |
| DOSE_2    | 10.745 | 5.144   | 2.09 | 0.070 |

S = 9.418      R-Sq = 95.2 %      R-Sq(adj) = 94.0 %

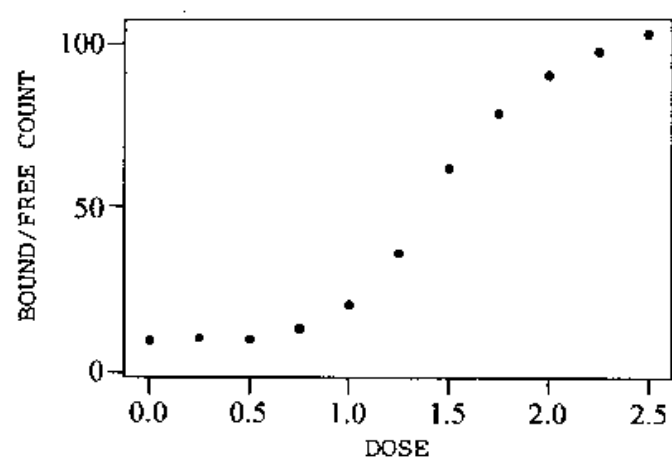
Analysis of Variance

| Source         | DF | SS      | MS     | F     | P     |
|----------------|----|---------|--------|-------|-------|
| Regression     | 2  | 13964.4 | 6982.2 | 78.72 | 0.000 |
| Residual Error | 8  | 709.6   | 88.7   |       |       |
| Total          | 10 | 14674.0 |        |       |       |

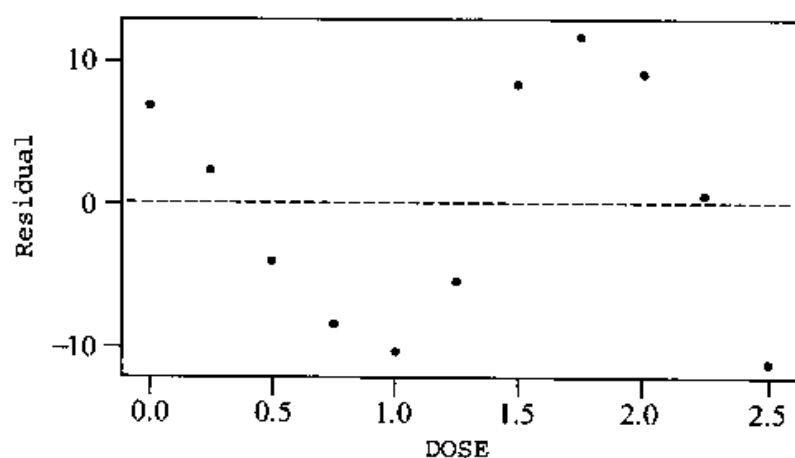
| Source | DF | Seq SS  |
|--------|----|---------|
| DOSE   | 1  | 13577.4 |
| DOSE_2 | 1  | 386.9   |



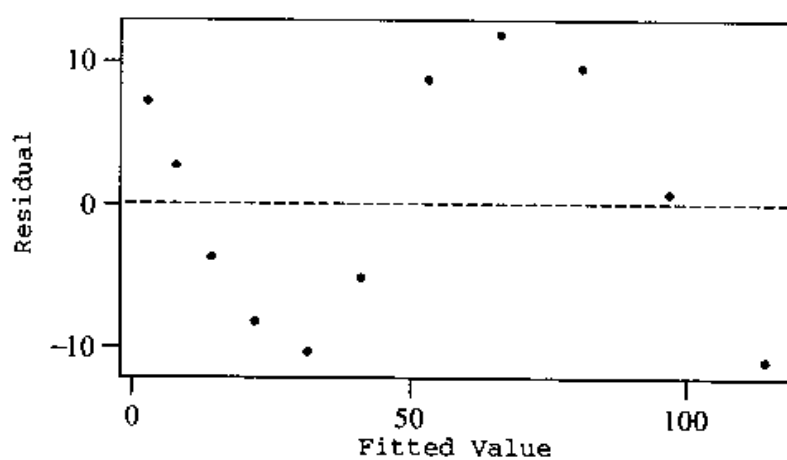
Plot of BOUND/FREE COUNT versus DOSE



Residuals versus DOSE  
(response is BOUND/FR)



Residuals versus the Fitted Values  
(response is BOUND/FR)



一个也许可以帮助平整 S 型数据的非线性模型为:

$$y = \frac{\beta_0 - \beta_3}{1 + (x/\beta_2)^{\beta_1}} + \beta_3$$

### 例 13.13

使用非线性估计方法来拟合之前面的放射免疫测定数据。

**解答** 用 SAS 来拟合同本数据,正如我们从残差图看出的,这个非线性模型对样本数据的拟合比线性或二次模型都好。

参数解释如下:

$\beta_0$ :  $y$  在曲线下端的值;

$\beta_3$ :  $y$  在曲线上端的值;

$\beta_1$ : 斜率的度量;

$\beta_2$ : 相应于  $y$  在  $\beta_0$  与  $\beta_3$  之间的浓度百分比( $x$ )。

(原文中  $\beta_1$  和  $\beta_2$  的解释颠倒了——译者注)

#### NONLINEAR REGRESSION ANALYSIS

##### DATA LISTING

| BOUND/FREE |         |      |
|------------|---------|------|
| OBS        | COUNT   | DOSE |
| 1          | 9.900   | 0.00 |
| 2          | 10.465  | 0.25 |
| 3          | 10.312  | 0.50 |
| 4          | 13.633  | 0.75 |
| 5          | 20.784  | 1.00 |
| 6          | 36.164  | 1.25 |
| 7          | 62.045  | 1.50 |
| 8          | 78.327  | 1.75 |
| 9          | 90.307  | 2.00 |
| 10         | 97.348  | 2.25 |
| 11         | 102.686 | 2.50 |

#### Non-Linear Least Squares Summary Statistics Dependent Variable COUNT

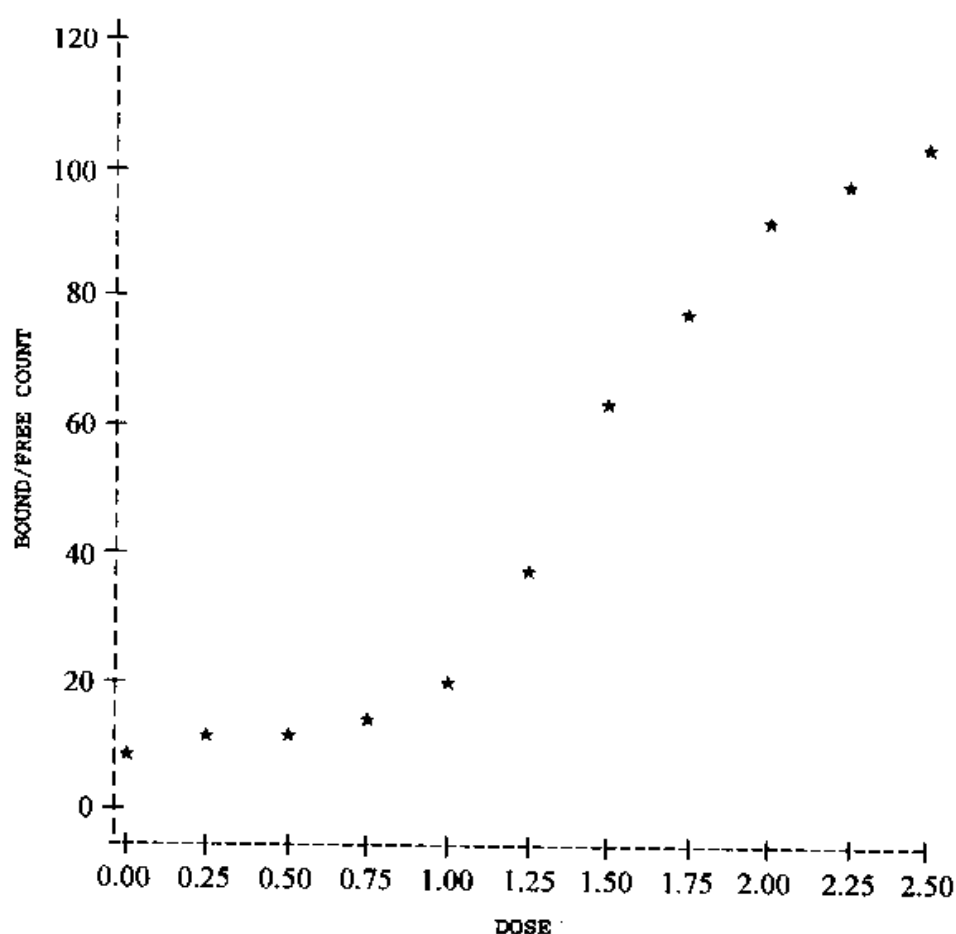
| Source            | DF | Sum of Squares | Mean Square  |
|-------------------|----|----------------|--------------|
| Regression        | 4  | 40390.959650   | 10097.739913 |
| Residual          | 7  | 9.675063       | 1.382152     |
| Uncorrected Total | 11 | 40400.634713   |              |
| (Corrected Total) | 10 | 14673.985182   |              |

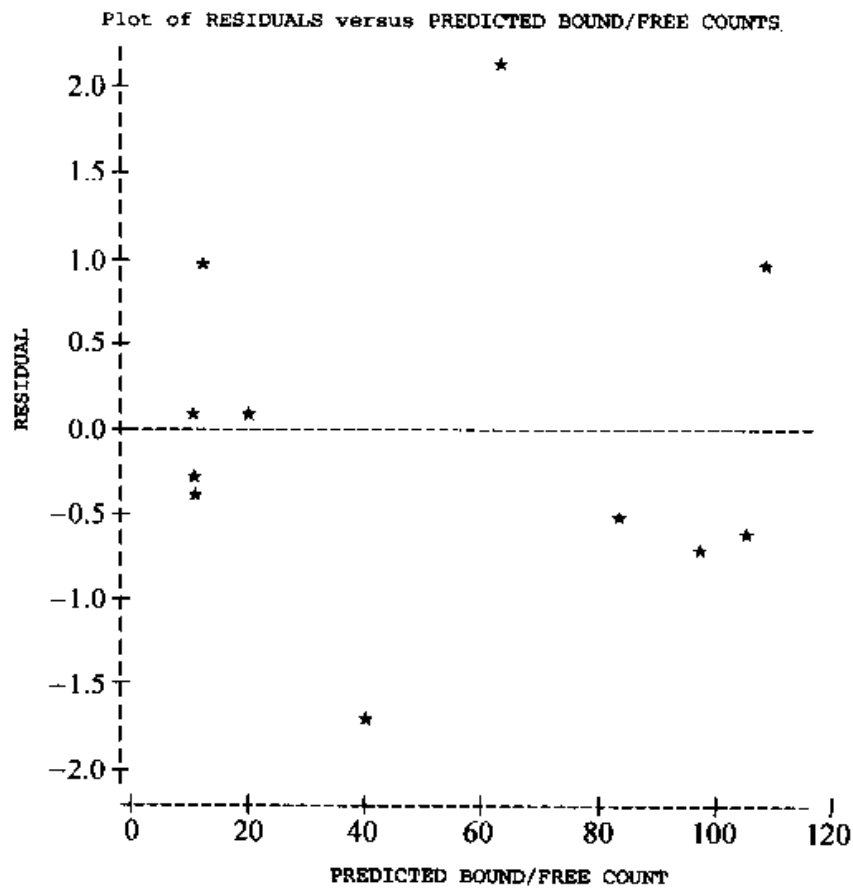
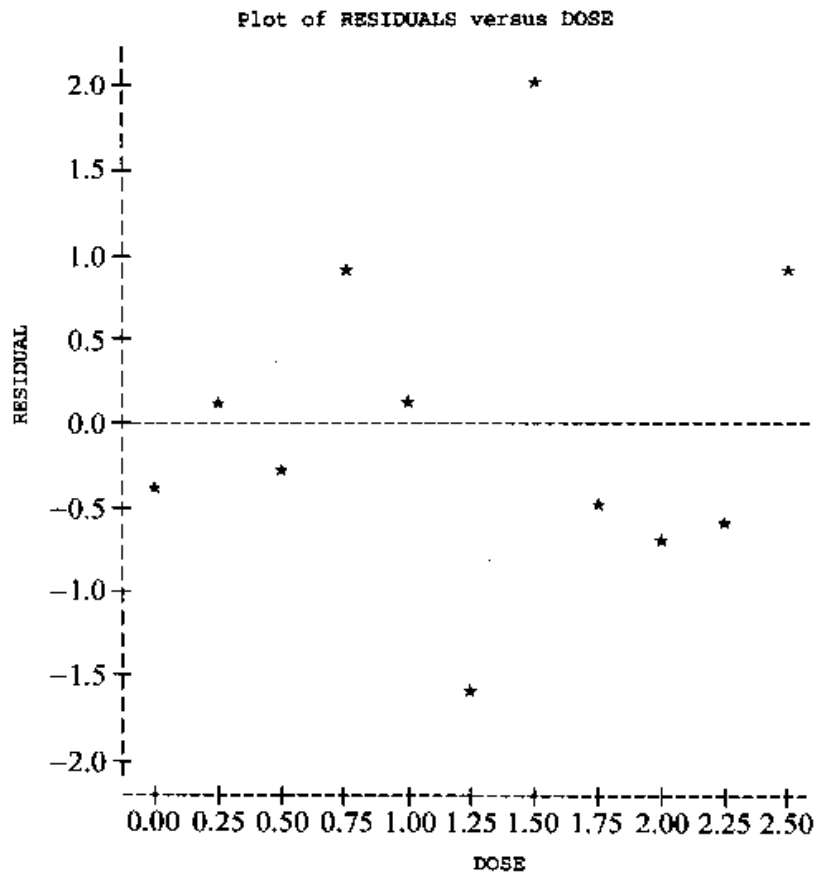
| Parameter                     | Estimate     | Asymptotic<br>Std. Error | Asymptotic 95 %<br>Confidence Interval |              |
|-------------------------------|--------------|--------------------------|--|--------------|
|                               |              |                          | Lower                                  | Upper        |
| B0                            | 10.3172019   | 0.6302496017             | 8.82688647                             | 11.80751738  |
| B1                            | 5.3700868    | 0.2558475371             | 4.76509868                             | 5.97507498   |
| B2                            | 1.4863334    | 0.0154121366             | 1.44988919                             | 1.52277759   |
| B3                            | 107.3777343  | 1.7277534567             | 103.29221381                           | 111.46325486 |
| Asymptotic Correlation Matrix |              |                          |  |              |
| Corr                          | B0           | B1                       | B2                                     | B3           |
| B0                            | 1            | 0.4317133357             | 0.1141723596                           | -0.255171767 |
| B1                            | 0.4317133357 | 1                        | -0.514768068                           | -0.808689153 |
| B2                            | 0.1141723596 | -0.514768068             | 1                                      | 0.7939083509 |
| B3                            | -0.255171767 | -0.808689153             | 0.7939083509                           | 1            |

NOTE: Missing values were generated as a result of performing an operation on missing values.

Each place is given by (number of times) AT (statement)/(line);(column).4 AT 1/815;16

Plot of BOUND/FREE COUNT versus DOSE





我们还可以使用拟合方程依据浓度百分比来预测  $y$ (比率)。

练习

应用

**13.6(医药)** 以下是 20 个病人在一剂药量下的血液含药量达到的最高值(BLOOD, mg/ml), 以及病人的体重(WEIGHT, lb)和药量(DOSE, mg)的数据, 使用下面的输出拟合一个线性回归直线, 并使用残差图来分析在回归模型中可能增加的项。

SAS OUTPUT FOR EXERCISE 13.6

DATA LISTING

| OBS | BLOOD | DOSE | WEIGHT |
|-----|-------|------|--------|
| 1   | 300   | 1    | 120    |
| 2   | 250   | 1    | 135    |
| 3   | 210   | 1    | 150    |
| 4   | 150   | 1    | 128    |
| 5   | 210   | 2    | 150    |
| 6   | 230   | 2    | 160    |
| 7   | 350   | 2    | 135    |
| 8   | 270   | 2    | 180    |
| 9   | 380   | 4    | 132    |
| 10  | 330   | 4    | 148    |
| 11  | 270   | 4    | 190    |
| 12  | 240   | 4    | 195    |
| 13  | 340   | 8    | 150    |
| 14  | 330   | 8    | 160    |
| 15  | 180   | 8    | 200    |
| 16  | 320   | 8    | 140    |
| 17  | 270   | 16   | 195    |
| 18  | 290   | 16   | 170    |
| 19  | 315   | 16   | 161    |
| 20  | 350   | 16   | 145    |

Dependent Variable: PEAK BLOOD LEVEL

## Analysis of Variance

| Source   | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|----------|----|----------------|-------------|---------|--------|
| Model    | 2  | 22290.44079    | 11145.22040 | 3.684   | 0.0468 |
| Error    | 17 | 51423.30921    | 3024.90054  |         |        |
| C Total  | 19 | 73713.75000    |             |         |        |
| Root MSE |    | 54.99910       | R-square    | 0.3024  |        |
| Dep Mean |    | 279.25000      | Adj R-sq    | 0.2203  |        |
| C. V.    |    | 19.69529       |             |         |        |

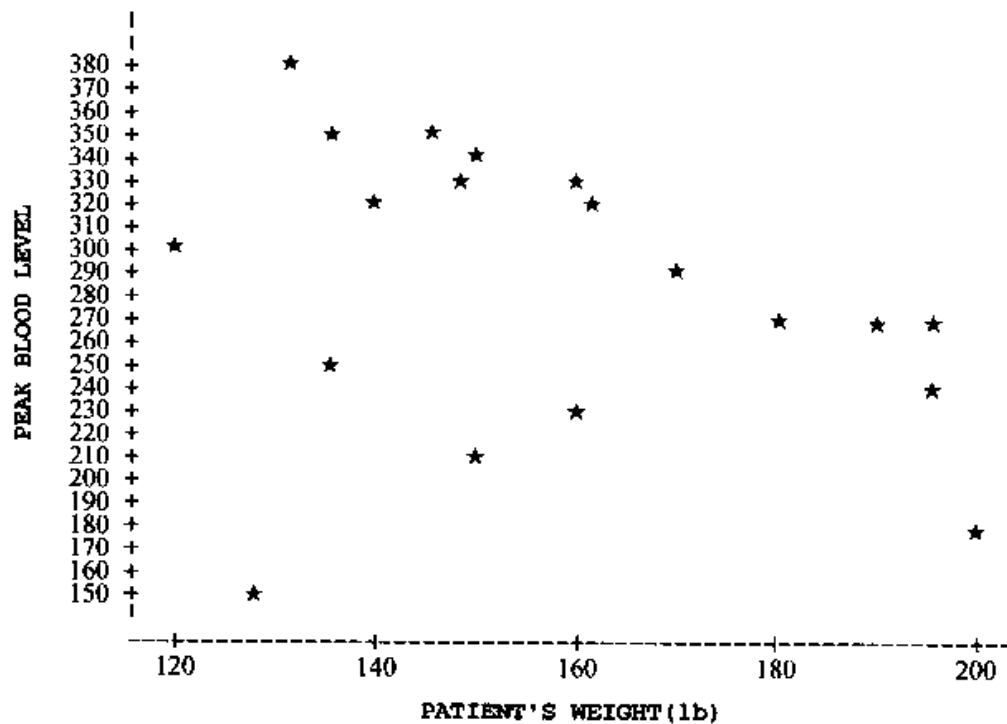
## Parameter Estimates

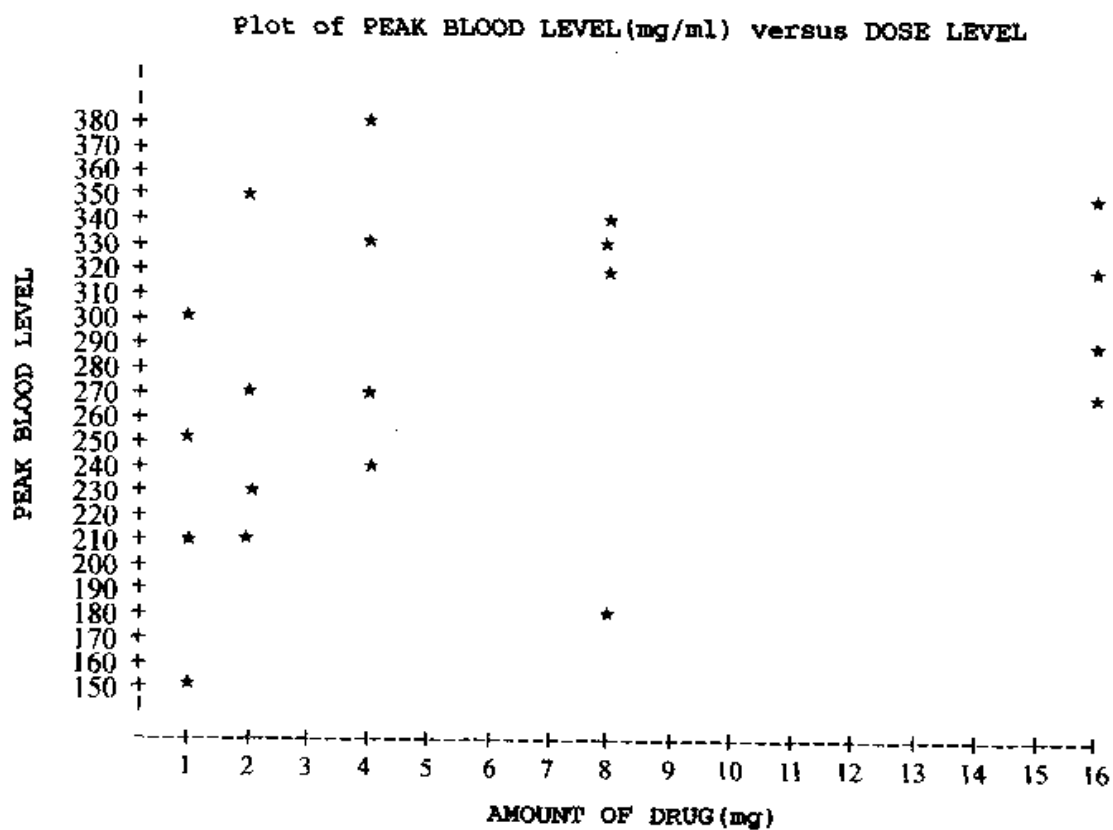
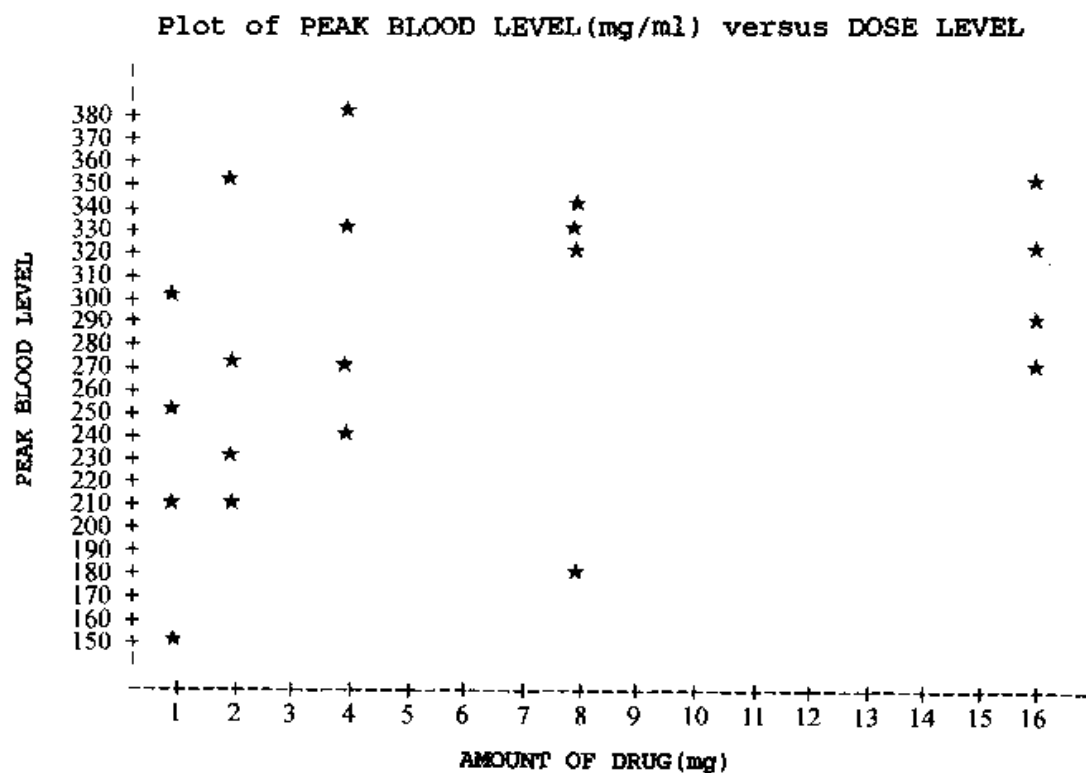
| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|----------------------------|---------|
| INTERCEP | 1  | 432.602294         | 84.69454320    | 5.108                      | 0.0001  |
| DOSE     | 1  | 5.546666           | 2.40278001     | 2.308                      | 0.0338  |
| WEIGHT   | 1  | -1.194285          | 0.55838151     | -2.139                     | 0.0473  |

## Variable

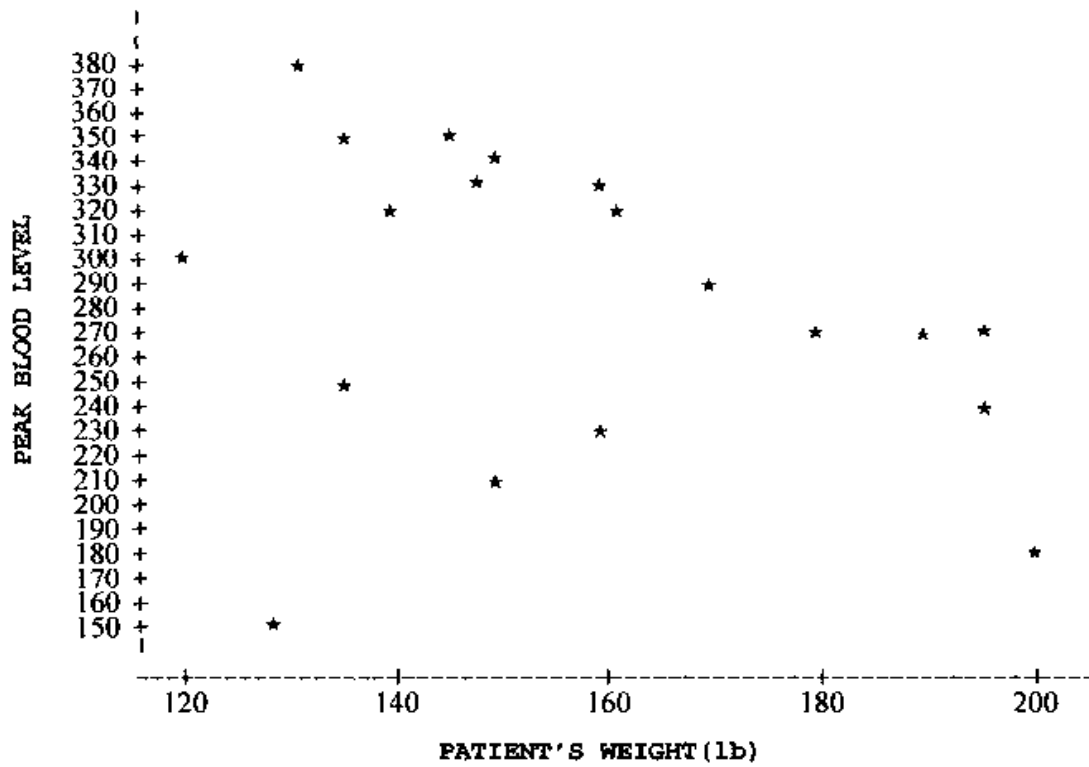
| Variable | DF | Label                 |
|----------|----|-----------------------|
| INTERCEP | 1  | Intercept             |
| DOSE     | 1  | AMOUNT OF DRUG (mg)   |
| WEIGHT   | 1  | PATIENT'S WEIGHT (lb) |

Plot of PEAK BLOOD LEVEL(mg/ml) versus PATIENT'S WEIGHT(lb)

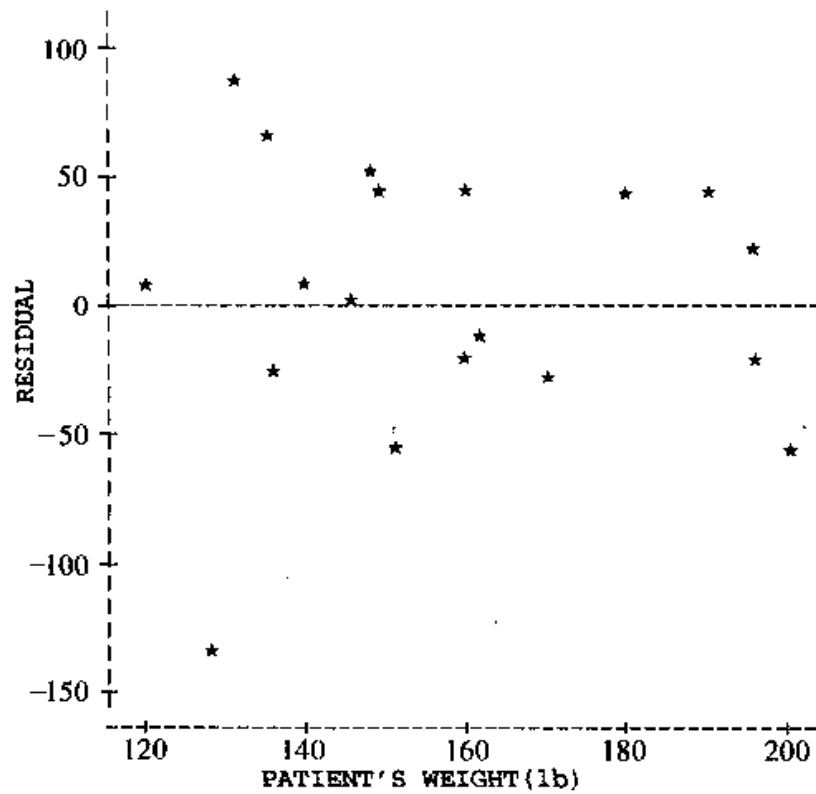




Plot of PEAK BLOOD LEVEL(mg/ml) versus PATIENT'S WEIGHT(lb)

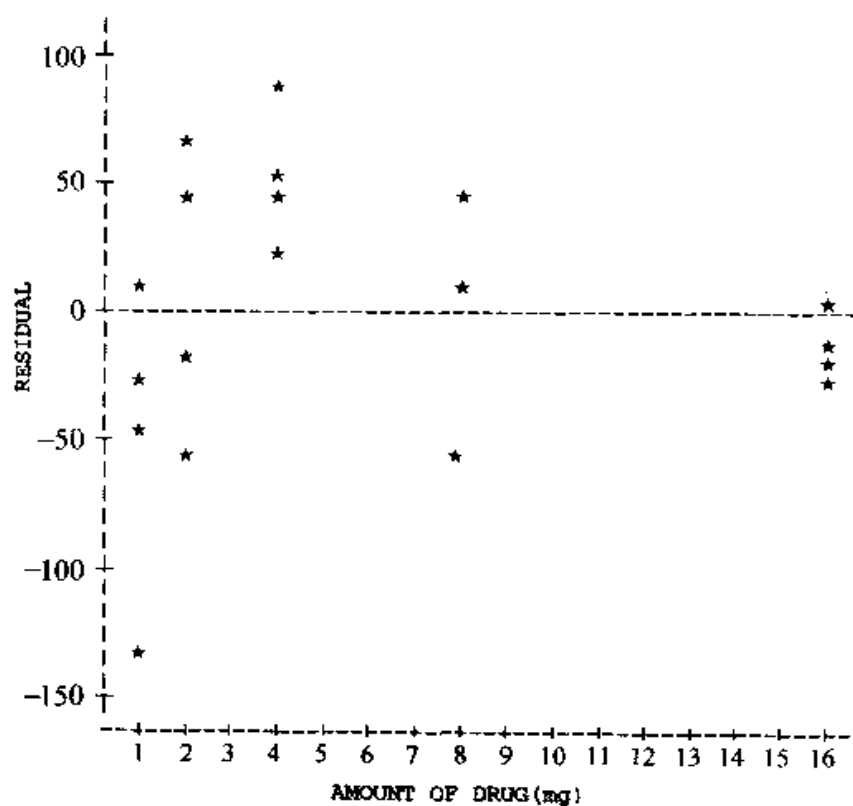


Plot of RESIDUAL versus PATIENT'S WEIGHT

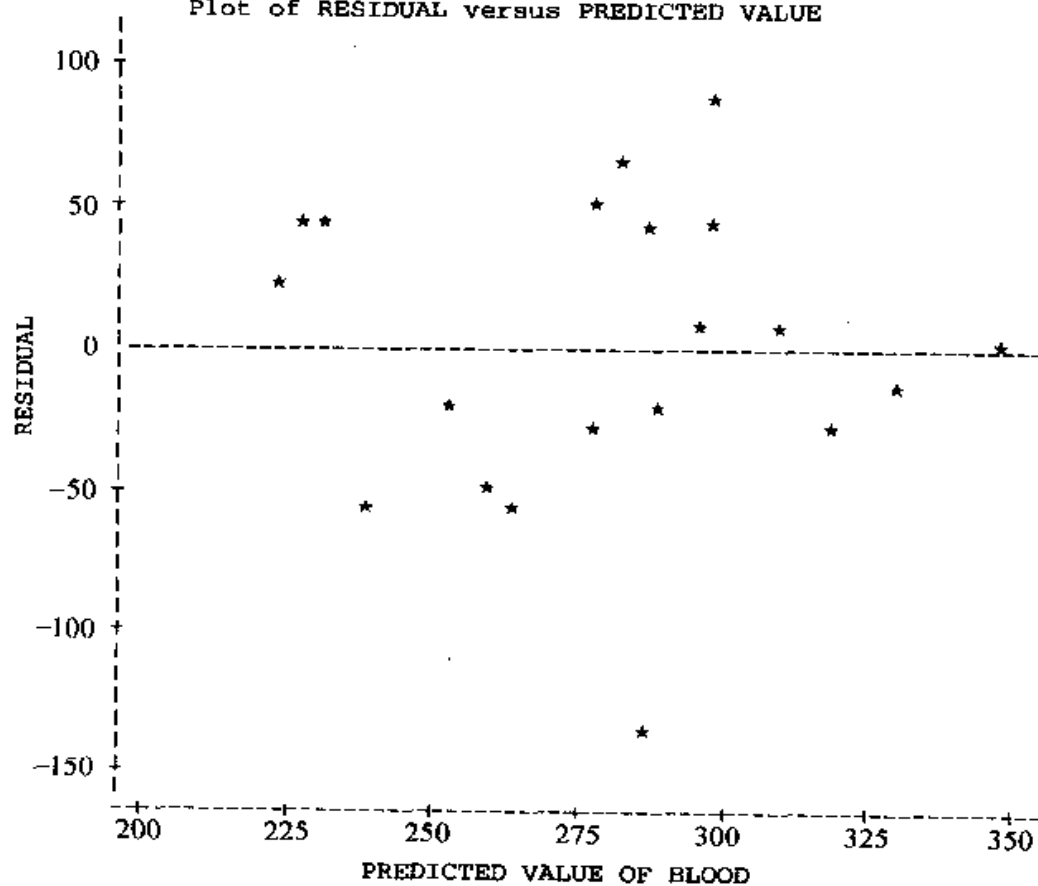




Plot of RESIDUAL versus AMOUNT OF DRUG



Plot of RESIDUAL versus PREDICTED VALUE



## 13.7 练习 13.6 续, 指出并讨论如下输出中拟合的模型。

SAS OUTPUT FOR EXERCISE 13.7

## REGRESSION ANALYSIS

Dependent Variable: BLOOD      PEAK BLOOD LEVEL

## Analysis of Variance

| Source   | DF | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|----------|----|-------------------|----------------|---------|--------|
| Model    | 3  | 41167.19623       | 13722.39874    | 6.746   | 0.0030 |
| Error    | 16 | 32546.55377       | 2034.15961     |         |        |
| C Total  | 19 | 73713.75000       |                |         |        |
| Root MSE |    | 45.10166          | R-square       | 0.5585  |        |
| Dep Mean |    | 279.25000         | Adj R-sq       | 0.4757  |        |
| C.V.     |    | 16.15100          |                |         |        |

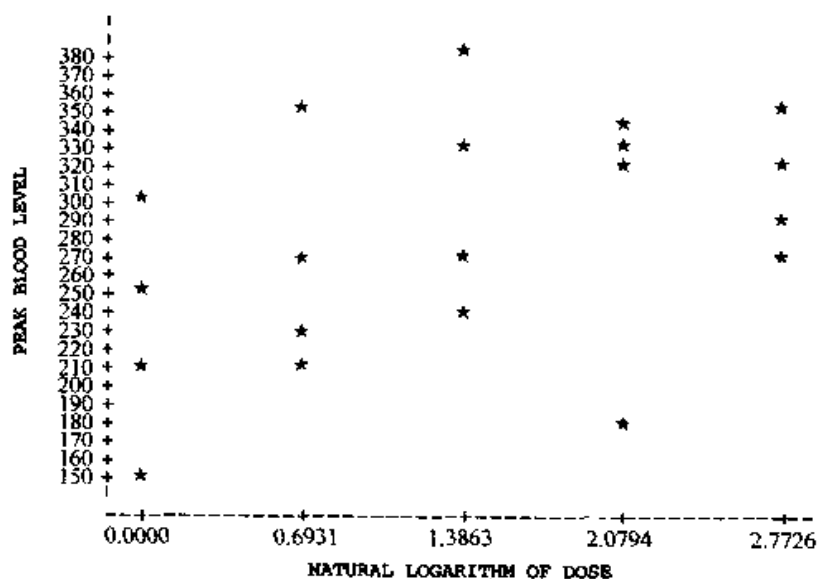
## Parameter Estimates

| Variable   | DF | Parameter<br>Estimate | Standard<br>Error | T for H0:<br>Parameter = 0 | Prob> T |
|------------|----|-----------------------|-------------------|----------------------------|---------|
| INTERCEP   | 1  | 288.062394            | 128.09498236      | 2.249                      | 0.0390  |
| LOG _ DOSE | 1  | 174.815459            | 79.24455964       | 2.206                      | 0.0423  |
| WEIGHT     | 1  | -0.344162             | 0.88171355        | -0.390                     | 0.7014  |
| LGDS _ WGT | 1  | -0.862927             | 0.51644880        | -1.671                     | 0.1142  |

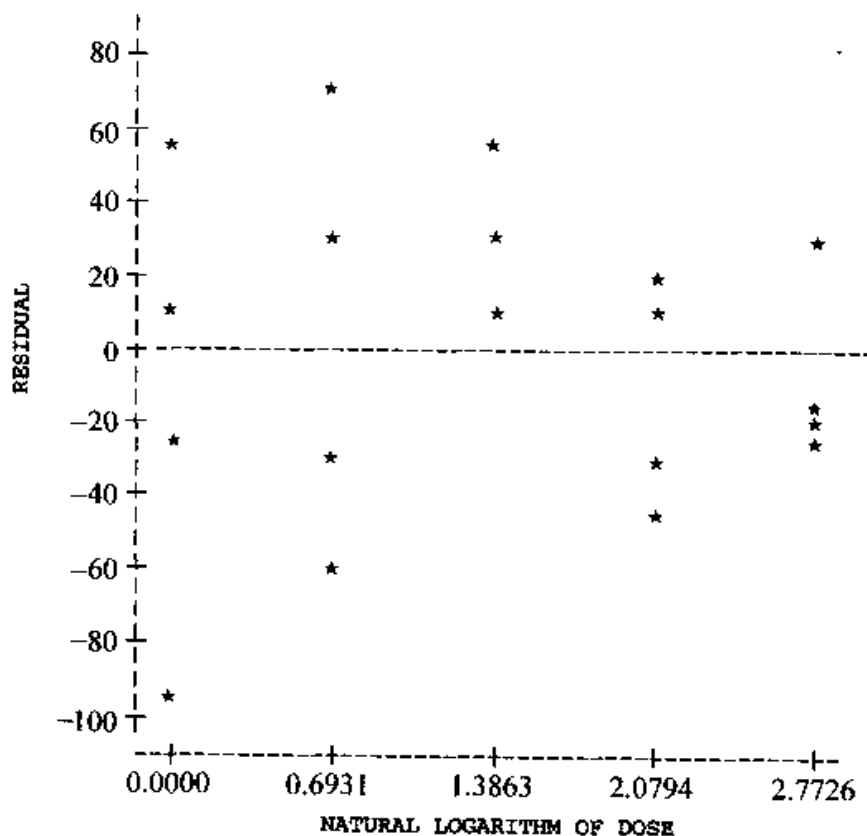
## Variable

| Variable   | DF | Label                          |
|------------|----|--------------------------------|
| INTERCEP   | 1  | Intercept                      |
| LOG _ DOSE | 1  | NATURAL LOGARITHM OF DOSE      |
| WEIGHT     | 1  | PATIENT'S WEIGHT(lb)           |
| LGDS _ WGT | 1  | LOGARITHM OF DOSE TIMES WEIGHT |

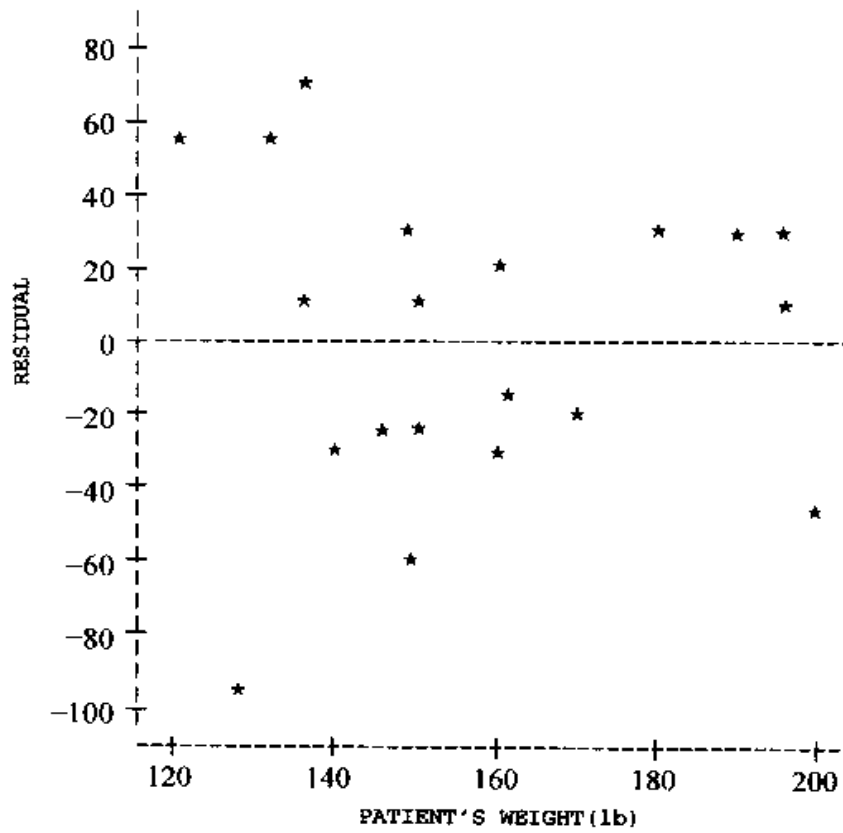
Plot of PEAK BLOOD LEVEL versus NATURAL LOGARITHM OF DOSE



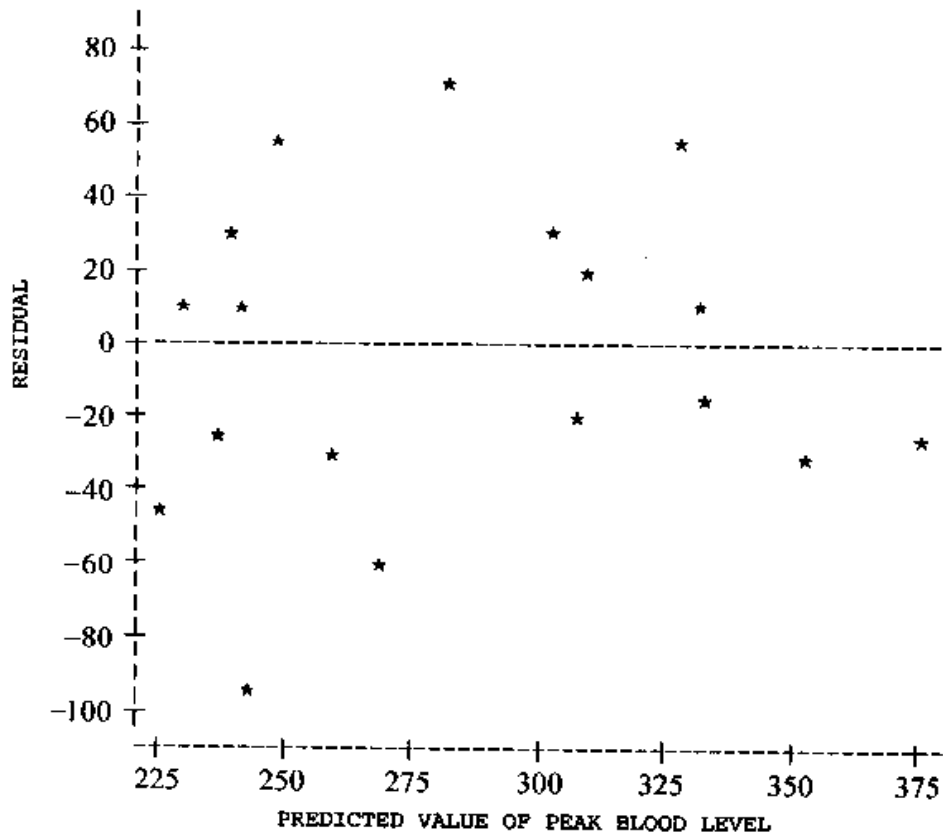
Plot of RESIDUALS versus NATURAL LOGARITHM OF DOSE



Plot of RESIDUALS versus PATIENT'S WEIGHT



Plot of RESIDUALS versus PREDICTED VALUE



**13.8(商业)** 一个专门从事公司对慈善事业、学校、文化机构等捐赠研究的顾问常被要求建议一个合适的捐赠数量。顾问试着根据已有的 38 家公司的捐助信息,用回归分析来预测公司对大学的捐助,其中商业信息来自公司每年的报表,而其他信息来自经济学杂志等。凭经验,顾问认为捐赠的水平受以下几个因素的影响:公司的利润,公司的规模,公司是否属于高科技产业(如信息处理、电子、化学),公司管理者的受教育水平,公司是否同员工的捐助相一致。利润可以通过税前或税后收入来确定,公司规模由雇员数或销售总额确定,教育水平由平均受教育时间或管理层人员受高等教育的比例确定。

- a. 税前和税后收入是不是高度相关? 雇员人数和销售总额呢?
- b. 讨论如何定义利润、规模和受教育水平使得这些变量间的相关系数不太大。

**13.9** 练习 13.8 中的顾问建议如下建立一个产业类型的变量:

$$\text{产业} = \begin{cases} 3, & \text{如果公司主要属于电子产业;} \\ 2, & \text{如果公司主要属于信息处理产业;} \\ 1, & \text{如果公司主要属于化学产业;} \\ 0, & \text{其他。} \end{cases}$$

- a. 解释这为什么不是个好方法。
- b. 提出一个更好的方法来表示这些产业。
- c. 如何把“公司是否同员工的捐助相一致”这一因子加入回归模型中?

**13.10** 练习 13.8 中的顾问收集了如下变量的信息:

CONTRIB:捐助数(百万美元计)。

INCOME:税前收入(百万美元计)。

SIZE:员工人数(千人计)。

DPDUMMY:1, 如果公司主要经营内容为信息处理。

0, 如果不是。

ELDUMMY:1, 如果公司主要经营内容为电子。

0, 如果不是。

CHDUMMY:1, 如果公司主要经营内容为化学。

0, 如果不是。

EDLEVEL:管理人员中受过高等教育的比例。

MATCHING:1, 公司能同员工的捐助相一致。

0, 如果不能。

- a. 把 CONTRIB 当作因变量,其他变量作为自变量是否合适? 特别是,这个做法是否会引起复共线性?
- b. 变量 CONTRIB/INCOME 代表什么含义?

**13.11** 参照练习 13.10, 顾问怀疑, 不同的产业间规模对 CONTRIB/INCOME 的影响有很大的不同。如何对回归模型作一些修改以检验这一怀疑?

Response: Sales

#### Summary of Fit

|                            |          |
|----------------------------|----------|
| RSquare                    | 0.434158 |
| RSquare Adj                | 0.417182 |
| Root Mean Square Error     | 17.9913  |
| Mean of Response           | 356.7692 |
| Observations (or Sum Wgts) | 104      |

#### Lack of Fit

#### Parameter Estimates

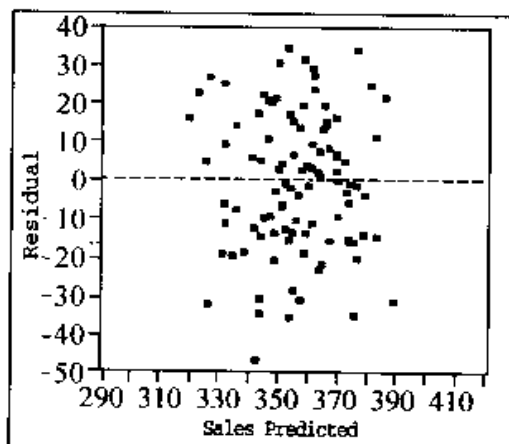
| Term                | Estimate  | Std Error | t Ratio | Prob> t |
|---------------------|-----------|-----------|---------|---------|
| Intercept           | 129.85375 | 80.66628  | 1.61    | 0.1106  |
| Price               | 44.849952 | 39.93534  | 1.12    | 0.2641  |
| Category sales      | 0.1214871 | 0.018249  | 6.66    | 0.0000  |
| Promotion by other? | -19.95964 | 3.702304  | -5.39   | 0.0000  |

#### Effect Test

#### Whole-Model Test

#### Analysis of Variance

| Source  | DF  | Sum of Squares | Mean Square | F Ratio |
|---------|-----|----------------|-------------|---------|
| Model   | 3   | 24835.761      | 8278.59     | 25.5759 |
| Error   | 100 | 32368.701      | 323.69      | Prob>F  |
| C Total | 103 | 57204.462      |             | 0.0000  |



**13.12** 参照练习 13.10, 顾问怀疑增长的 EDLEVEL 的影响是自增长的, 即, 其他都相等时, 公司的 EDLEVEL = 0.2 增至 0.3 时 CONTRIB/INCOME 几乎没什么变化, 而当 EDLEVEL = 0.4 增至 0.5 时, 变化有所增加, 当 EDLEVEL = 0.6 增至 0.7 时, 变化增加得更大。

- 如何建立一个回归模型来检验这个怀疑?
- 如果顾问的怀疑是正确的, 而且如果画出残差对 EDLEVEL 的散点图, 残差图会是什么样子?

**13.13(商业)** 一个连锁超市统计了一种品牌的薄脆饼干在一周时间内在 104 家连锁店内的销售情况。分析员想要基于所有品牌的薄脆饼干的销售总额, 所考虑的品牌的价格, 及在是否存在着其他竞争品牌的促销活动(如果有促销 promotion = 1, 否则 = 0), (所考虑的品牌无促销)来预测一个店中的销售量。用 JMP 软件的多元回归得到输出如下。

- 解释 promotion 变量的系数。
- 其他品牌的促销是否会增加或减少题目中品牌的销量? 系数是否表明了这一点?
- 在通常的  $\alpha$  下, 系数与 0 有显著差异吗?

**13.14** 在前面的问题中, 如何能以 95% 的置信区间预测某一周的销量?

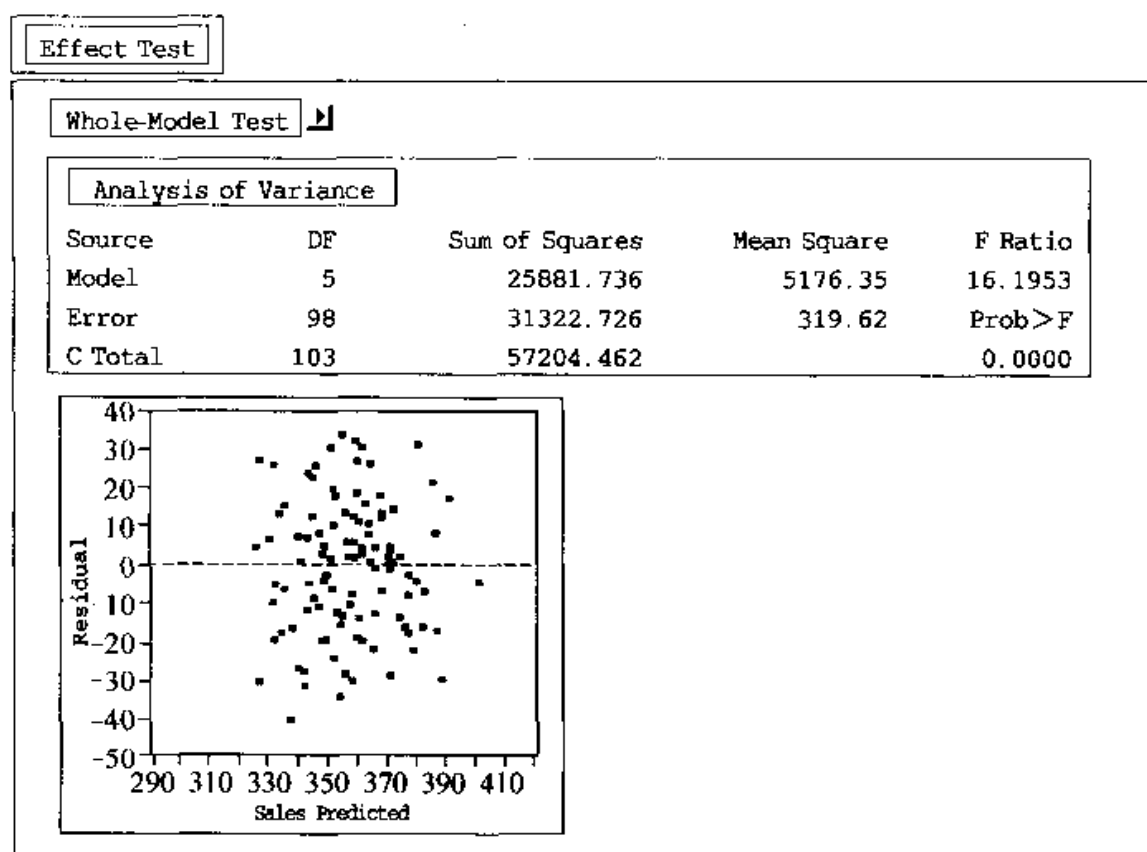
**13.15(商业)** 下面是对薄脆饼干数据的另一个回归模型, 其中包含了促销变量与价格的乘积项和促销变量与薄脆类饼干的销售额的乘积项。这个模型的输出如下图。当竞争品牌存在促销时, 乘积项的系数对预测销量有何作用? 特别的, 这些系数是否会影响模型的截距或斜率?

Response: Sales

| Summary of Fit             |          |
|----------------------------|----------|
| RSquare                    | 0.452443 |
| RSquare Adj                | 0.424506 |
| Root Mean Square Error     | 17.87791 |
| Mean of Response           | 356.7692 |
| Observations (or Sum Wgts) | 104      |

Lack of Fit

| Parameter Estimates |           |           |         |         |
|---------------------|-----------|-----------|---------|---------|
| Term                | Estimate  | Std Error | t Ratio | Prob> t |
| Intercept           | 26.806609 | 98.33649  | 0.27    | 0.7857  |
| Price               | 90.233085 | 47.75194  | 1.89    | 0.0618  |
| Category sales      | 0.1335274 | 0.023854  | 5.60    | 0.0000  |
| Promotion by other? | 287.6092  | 172.2049  | 1.67    | 0.0981  |
| Price*Promotio      | -142.4326 | 86.15011  | -1.65   | 0.1015  |
| Category*Promotio   | -0.024087 | 0.036816  | -0.65   | 0.5145  |



**13.16(工程)** 一个生产用来包装食品(如饼干和点心)的薄膜的公司在薄膜的硬度方面碰到一些问题。要使用现代化包装机,硬度(在某种度量下)就需要很高。硬度被认为是生产过程中一些变量影响的结果。回归分析试图通过这些变量的不同组合来预测薄膜的硬度。一共做了 32 次试验,使用如下的变量记录数据:

STIFF:硬度

MELT:溶解温度( $^{\circ}\text{F}$ )

CHILL:冷却温度( $^{\circ}\text{F}$ )

REPEL:使用回收颗粒的百分比

SPEED:生产线上的生产速度(英尺/分钟)

KNIFE:真空刀的位置

公司的化学工程师对这些变量间的关系有很多不确定的地方。如下是一阶模型的输出:

PEARSON CORRELATION MATRIX

|       | CHILL  | KNIFE | MELT  | REPEL | SPEED | CHILL |
|-------|--------|-------|-------|-------|-------|-------|
|       | 1.000  |       |       |       |       |       |
| KNIFE | -0.000 | 1.000 |       |       |       |       |
| MELT  | -0.000 | 0.000 | 1.000 |       |       |       |



|       |       |        |        |        |       |
|-------|-------|--------|--------|--------|-------|
| REPEL | 0.000 | 0.000  | 0.000  | 1.000  |       |
| SPEED | 0.000 | -0.000 | -0.000 | -0.000 | 1.000 |
| STIFF | 0.138 | -0.308 | 0.059  | -0.886 | 0.030 |

NUMBER OF OBSERVATIONS: 32

a. 数据中的共线性如何?

b. 32 个观测是在如下因子水平的组合中各观测一次, MELT = 510, 530, 550, 570, CHILL = 70, 80, 90, 100, REPEL = 20, 30, 在 MELT 和 CHILL 之间, MELT 和 REPEL 之间的相关程度如何?

13.17 用一阶模型来拟合练习 13.16 中的数据, 得到如下的输出和图, 可否直接看出非线性? RESSTIFF 表示残差。

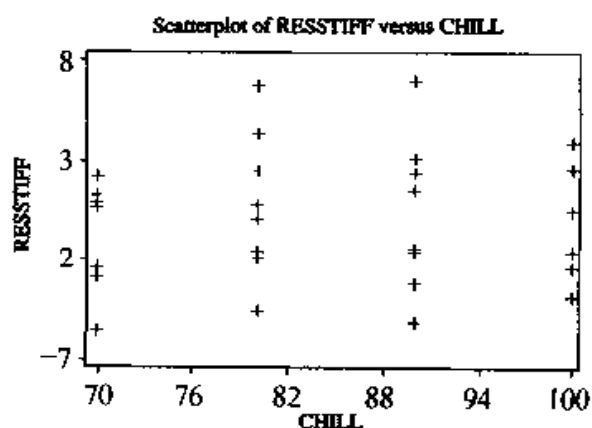
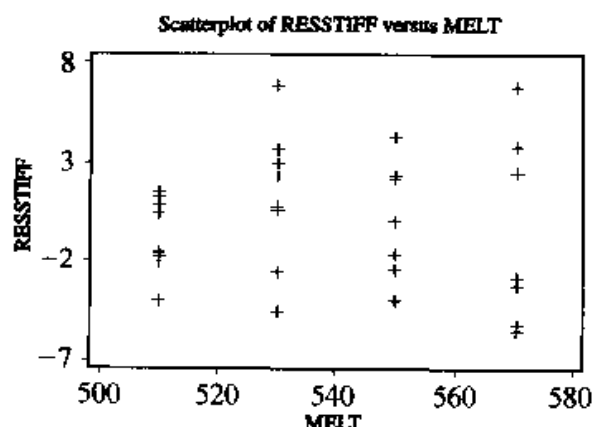
DEP VAR: STIFF N: 32 MULTIPLE R: 0.951 SQUARED MULTIPLE R: 0.904

ADJUSTED SQUARED MULTIPLE R: .885 STANDARD ERROR OF ESTIMATE: 3.571

| VARIABLE | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T       | P(2 TAIL) |
|----------|-------------|-----------|----------|-----------|---------|-----------|
| CONSTANT | 170.962     | 20.509    | 0.000    | .         | 8.336   | 0.000     |
| MELT     | 0.028       | 0.028     | 0.059    | 1.000     | 0.974   | 0.339     |
| CHILL    | 0.128       | 0.056     | 0.138    | 1.000     | 2.258   | 0.033     |
| REPEL    | -1.838      | 0.126     | -0.886   | 1.000     | -14.556 | 0.000     |
| SPEED    | 0.007       | 0.014     | 0.030    | 1.000     | 0.487   | 0.630     |
| KNIFE    | -0.319      | 0.063     | -0.308   | 1.000     | -5.050  | 0.000     |

# ANALYSIS OF VARIANCE

| SOURCE     | SUM-OF-SQUARES | DF | MEAN-SQUARE | F-RATIO | P     |
|------------|----------------|----|-------------|---------|-------|
| REGRESSION | 3106.400       | 5  | 621.280     | 48.732  | 0.000 |
| RESIDUAL   | 331.475        | 26 | 12.749      |         |       |



**13.18** 为了检测练习 13.16 中是否存在非线性,使用二阶模型(包括 MELT 和 CHILL 项的平方项),得到如下的输出:

DEP VAR: STIFF N: 32 MULTIPLE R:0.956 SQUARED MULTIPLE R: 0.914

ADJUSTED SQUARED MULTIPLE R:0.889 STANDARD ERROR OF ESTIMATE:3.513

| VARIABLE      | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | T        | P(2 TAIL) |
|---------------|-------------|-----------|----------|-----------|----------|-----------|
| CONSTANT      | - 308.000   | 459.031   | 0.000    | .         | - 0.671  | 0.509     |
| MELT          | 1.377       | 1.677     | 2.972    | 0.000     | 0.821    | 0.420     |
| CHILL         | 3.634       | 2.362     | 3.920    | 0.001     | 1.539    | 0.137     |
| REPEL         | - 1.838     | 0.124     | - 0.886  | 1.000     | - 14.793 | 0.000     |
| SPEED         | 0.034       | 0.031     | - 0.148  | 0.200     | - 1.107  | 0.279     |
| KNIFE         | - 0.319     | 0.062     | - 0.308  | 1.000     | - 5.132  | 0.000     |
| MELT * MELT   | - 0.001     | 0.002     | - 2.913  | 0.000     | - 0.805  | 0.429     |
| CHILL * CHILL | - 0.021     | 0.014     | - 3.787  | 0.001     | - 1.485  | 0.151     |

#### ANALYSIS OF VARIANCE

| SOURCE     | SUM - OF - SQUARES | DF | MEAN - SQUARE | F - RATIO | P     |
|------------|--------------------|----|---------------|-----------|-------|
| REGRESSION | 3141.625           | 7  | 448.804       | 36.359    | 0.000 |
| RESIDUAL   | 296.250            | 24 | 12.344        |           |       |

- 这个模型的  $R^2$  比练习 13.17 中的一阶模型的  $R^2$  大多少?
- 对全模型和简约的模型,使用 F 检验来检验零假设(即增加平方项没有增加预测价值)。取  $\alpha = 0.05$
- $t$  统计量是否表明平方项同前面的因子一样显著( $\alpha = 0.05$ )?

**13.19** 对练习 13.16 的数据使用一阶模型的向前逐步回归,如下是输出:

#### PEARSON CORRELATION MATRIX

|       | CHILL   | KNIFE   | MELT    | REPEL   | SPEED |
|-------|---------|---------|---------|---------|-------|
| CHILL | 1.000   |         |         |         |       |
| KNIFE | - 0.000 | 1.000   |         |         |       |
| MELT  | - 0.000 | 0.000   | 1.000   |         |       |
| REPEL | 0.000   | 0.000   | 0.000   | 1.000   |       |
| SPEED | 0.000   | - 0.000 | - 0.000 | - 0.000 | 1.000 |
| STIFF | 0.138   | - 0.308 | 0.059   | - 0.886 | 0.030 |

MINIMUM TOLERANCE FOR ENTRY INTO MODEL = .010000

STEP # 1 R = 0.886 RSQUARE = 0.786

TERM ENTERED: REPEL

| VARIABLE   | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F | P |
|------------|-------------|-----------|----------|-----------|---|---|
| 1 CONSTANT |             |           |          |           |   |   |

|                                    |             |           |          |           |         |       |
|------------------------------------|-------------|-----------|----------|-----------|---------|-------|
| 4 REPEL                            | - 1.838     | - 0.175   | - 0.886  | .1E + 01  | 109.988 | 0.000 |
| STEP # 2 R = 0.938 RSQUARE = 0.880 |             |           |          |           |         |       |
| TERM ENTERED: KNIFE                |             |           |          |           |         |       |
| VARIABLE                           | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F       | 'P'   |
| 1 CONSTANT                         |             |           |          |           |         |       |
| 4 REPEL                            | - 1.838     | 0.133     | - 0.886  | .1E + 01  | 190.301 | 0.000 |
| 6 KNIFE                            | - 0.319     | - 0.067   | - 0.308  | .1E + 01  | 22.906  | 0.000 |
| STEP # 3 R = 0.948 RSQUARE = 0.899 |             |           |          |           |         |       |
| TERM ENTERED: CHILL                |             |           |          |           |         |       |
| VARIABLE                           | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F       | 'P'   |
| 1 CONSTANT                         |             |           |          |           |         |       |
| 3 CHILL                            | 0.128       | 0.056     | 0.138    | .1E + 01  | 5.253   | 0.030 |
| 4 REPEL                            | - 1.838     | - 0.124   | - 0.886  | .1E + 01  | 218.210 | 0.000 |
| 6 KNIFE                            | - 0.319     | - 0.062   | - 0.308  | .1E + 01  | 26.265  | 0.000 |
| STEP # 4 R = .950 RSQUARE = .903   |             |           |          |           |         |       |
| TERE ENTERED: MELT                 |             |           |          |           |         |       |
| VARIABLE                           | COEFFICIENT | STD ERROR | STD COEF | TOLERANCE | F       | 'P'   |
| 1 CONSTANT                         |             |           |          |           |         |       |
| 2 MELT                             | 0.028       | 0.028     | 0.059    | .1E + 01  | 0.977   | 0.332 |
| 3 CHILL                            | 0.128       | 0.056     | 0.138    | .1E + 01  | 5.249   | 0.030 |
| 4 REPEL                            | - 1.838     | - 0.124   | - 0.886  | .1E + 01  | 218.028 | 0.000 |
| 6 KNIFE                            | - 0.319     | - 0.062   | - 0.308  | .1E + 01  | 26.243  | 0.000 |

- 列出自变量进入模型的顺序;
- 按最大(绝对值)到最小的顺序列出自变量与 STIFF 的相关系数;
- 比较两个排序的不同。

13.20 参照练习 13.19, 对全模型和简约模型使用 12.5 节介绍的  $F$  检验法来检验假设: 最后两个进入回归模型的变量没有预测价值。

13.21 练习 13.10 中的顾问把 CONTRIB/INCOME 作为因变量建立回归模型。

MTB>regress c9 on 7 vars in 'Income'-'Matching'

The regression equation is

Cont/Inc = 0.0211 - 0.000093 Income + 0.00153 Size + 0.00168 DPDummy  
 + 0.00713 ELDummy + 0.00281 CHDummy - 0.0144 EdLevel  
 + 0.00138 Matching

|           |          |          |           |       |
|-----------|----------|----------|-----------|-------|
| Predictor | Coef     | Stdev    | t - ratio | p     |
| Constant  | 0.021085 | 0.003497 | 6.03      | 0.000 |

|          |            |           |       |       |
|----------|------------|-----------|-------|-------|
| Income   | -0.0000933 | 0.0001033 | -0.90 | 0.372 |
| Size     | 0.0015301  | 0.0006699 | 2.28  | 0.028 |
| DPDummy  | 0.001684   | 0.004720  | 0.36  | 0.723 |
| ELDummy  | 0.007132   | 0.006355  | 1.12  | 0.269 |
| CHDummy  | 0.002808   | 0.003810  | 0.74  | 0.466 |
| EdLevel  | -0.01436   | 0.01597   | -0.90 | 0.374 |
| Matching | 0.001381   | 0.002092  | 0.66  | 0.513 |

s = 0.006057    R-sq = 21.3%    R-sq(adj) = 6.8%

#### Analysis of Variance

| SOURCE     | DF | SS         | MS         | F    | p     |
|------------|----|------------|------------|------|-------|
| Regression | 7  | 0.00037598 | 0.00005385 | 1.47 | 0.208 |
| Error      | 38 | 0.00139418 | 0.00003669 |      |       |
| Total      | 45 | 0.00177115 |            |      |       |

a. 是否拒绝“没有自变量有显著作用”的假设(取一个合理的  $\alpha$ )?

b. 哪个变量同最后一个进入的预测因子有相同的显著性(如取  $\alpha = 0.05$ )?

**13.22** 如下是一个比练习 13.21 中简单一点的回归模型,它是因变量关于自变量 DPDummy、ELDummy、EdLevel 和 Matching 的回归模型。输出如下:

MTB> regress c9 on 4 vars 'DPDummy' 'ELDummy' 'EdLevel' 'Matching'

The regression equation is

Cont/Inc = 0.0202 - 0.00378 DPDummy - 0.00099 ELDummy  
+ 0.0097 EdLevel + 0.00197 Matching

| Predictor | Coef      | Stdev    | t-ratio | p     |
|-----------|-----------|----------|---------|-------|
| Constant  | 0.020233  | 0.002304 | 8.78    | 0.000 |
| DPDummy   | -0.003775 | 0.003912 | -0.97   | 0.340 |
| ELDummy   | -0.000995 | 0.005131 | -0.19   | 0.847 |
| EdLevel   | 0.00969   | 0.01276  | 0.76    | 0.452 |
| Matching  | 0.001974  | 0.001995 | 0.99    | 0.328 |

s = 0.006343    R-sq = 6.9%    R-sq(adj) = 0.0%

#### Analysis of Variance

| SOURCE     | DF | SS         | MS         | F    | p     |
|------------|----|------------|------------|------|-------|
| Regression | 4  | 0.00012167 | 0.00003042 | 0.76 | 0.560 |
| Error      | 41 | 0.00164949 | 0.00004023 |      |       |
| Total      | 45 | 0.00177115 |            |      |       |

a. 相对于这里考虑的模型,练习 13.21 中的模型  $R^2$  的增量为多少?

b. 在  $\alpha = 0.05$  的 F 检验中,这个增量显著吗?

- c. 把前面的模型当作“全系数”模型, 计算这个模型的  $C_p$ , 从得到的信息, 你认为哪个模型更合理?

### 13.4 模型假设的检查(第三步)

到目前为止, 我们已经找出了可能的自变量(第一步)并考虑了多元回归模型的形式(第二步)。下面我们应检验选择的模型是否有效。回想在第十一章, 我们对如下形式的回归模型:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i$$

所做的基本假设如下:

1. 零期望: 对于所有的  $i$ ,  $E(\epsilon_i) = 0$ 。
2. 等方差: 对于所有的  $i$ ,  $V(\epsilon_i) = \sigma_\epsilon^2$ 。
3. 正态性:  $\epsilon_i$  服从正态分布。
4. 独立性:  $\epsilon_i$  是相互独立的。

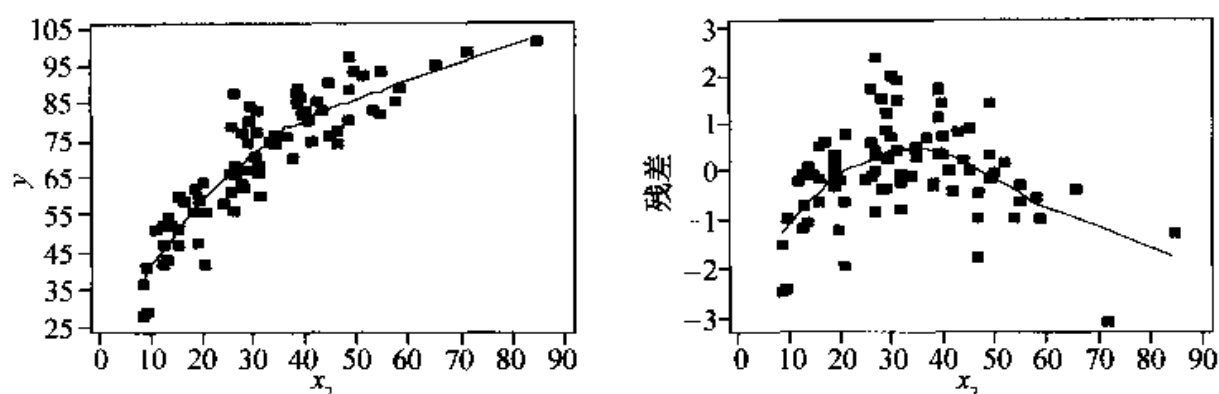
注意到因为多元回归的假设写成了随机误差的形式, 所以检验假设的一个合理的方法是使用  $\epsilon_i$  的估计值——残差  $y_i - \hat{y}_i$ 。

第一个假设(0 期望)是关于模型选择和是否需要在模型中增加自变量的。如果我们在第一步和第二步中已经做了这个工作, 假设 1 就满足了。第十一章和 13.3 节都讨论了用残差图来检验模型拟合不足的问题, 如果我们在第一步和第二步中未做这一工作, 残差图会有助于查出这一问题。

回想一下残差是指实际的  $y$  值和使用回归模型预测的  $y$  值之差。在画残差图时, 我们常使用**标准化残差**。标准化残差是用标准差为单位表示的, 所以, 如果标准化残差 = -3.00, 表示点距离回归直线 3 个标准差。通常, 减去数据的预测部分会更清楚的显示出其他结构。特别地, 在一阶模型(只含线性项)中, 点对各自变量的残差图, 通常可以显示出数据中的进一步的结构, 从而可改进回归模型。

一种可能是非线性。我们在本章的前一部分讨论了非线性及其变换。残差的明显的曲线性反应了数据的曲线关系, 这表示回归方程采用不同的数学形式会改进模型的预测值。残差对每一自变量  $x$  的残差图常可显示出这一问题。散点图的平滑, 如 LOWESS, 可以用来寻找残差的曲线, 例如, 图 13.4 画出了  $y$  对  $x_2$  和残差对  $x_2$  的散点图, 我们觉得在残差图中曲线关系更明显。LOWESS 曲线在两个图中都很有助益。

当发现存在非线性时, 尝试对自变量或因变量做变换。一个标准做法是对除虚拟变量外的所有变量做(自然)对数变换。这个模型在一个自变量有微小变化而其他自变量不变时, 对因变量的估计也有一个微小的变化。其他一些有用的变换包括只对一个或几个自变量取对数, 取自变量的平方根, 或对自变量及因变量取倒

图 13.4  $y$  对  $x_2$  和残差对  $x_2$  的散点图

数等。使用好的计算机软件,这些变换很容易得到。

假设 2,即等方差的性质可以使用残差图来考察。检验非等方差的一个最简单的残差图是画残差对预测值  $\hat{y}_i$  的图。大多数统计软件都把这些图作为回归分析的一部分。

#### 例 13.14

使用 SAS,表 13.5 的数据可以用模型  $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$  拟合。考察残差对  $\hat{y}_i$  的图以探索是否存在非等方差性。你能否看出方差不等?

表 13.5 例 13.14 的数据

|     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $y$ | 11  | 10  | 2   | 14  | 22  | 10  | 20  | 19  | 32  | 23  | 40  | 37  |
| $x$ | 0.5 | 1   | 1.2 | 1.4 | 1.7 | 1.8 | 2   | 2.3 | 2.5 | 2.8 | 3   | 3.1 |
| $y$ | 30  | 43  | 55  | 29  | 45  | 60  | 53  | 30  | 42  | 25  | 63  | 51  |
| $x$ | 3.5 | 3.6 | 3.8 | 4.2 | 4.4 | 5.1 | 5.2 | 5.4 | 5.5 | 6   | 6.2 | 6.3 |

**解答** 从 SAS 输出的残差图,我们可以看出,残差随着  $y$  的预测值的增加而增加,表示存在方差不等的情况。同时,因为  $y$  通过回归模型与  $x$  直接有关(即  $y$  随  $x$  增加),所以,残差随  $x$  的增加而增加。残差的这种特点表示  $\epsilon_i$  的方差随  $x$  的增加而增加。 $y$  对  $x$  的附图也表明了这一点。

DATA LISTING

| OBS | Y  | X   | X2   |
|-----|----|-----|------|
| 1   | 11 | 0.5 | 0.25 |

|    |    |     |       |
|----|----|-----|-------|
| 2  | 10 | 1.0 | 1.00  |
| 3  | 2  | 1.2 | 1.44  |
| 4  | 14 | 1.4 | 1.96  |
| 5  | 22 | 1.7 | 2.89  |
| 6  | 10 | 1.8 | 3.24  |
| 7  | 20 | 2.0 | 4.00  |
| 8  | 19 | 2.3 | 5.29  |
| 9  | 32 | 2.5 | 6.25  |
| 10 | 23 | 2.8 | 7.84  |
| 11 | 40 | 3.0 | 9.00  |
| 12 | 37 | 3.1 | 9.61  |
| 13 | 30 | 3.5 | 12.25 |
| 14 | 43 | 3.6 | 12.96 |
| 15 | 55 | 3.8 | 14.44 |
| 16 | 29 | 4.2 | 17.64 |
| 17 | 45 | 4.4 | 19.36 |
| 18 | 60 | 5.1 | 26.01 |
| 19 | 53 | 5.2 | 27.04 |
| 20 | 30 | 5.4 | 29.16 |
| 21 | 42 | 5.5 | 30.25 |
| 22 | 25 | 6.0 | 36.00 |
| 23 | 63 | 6.2 | 38.44 |
| 24 | 51 | 6.3 | 39.69 |

Dependent Variable: Y

Analysis of Variance

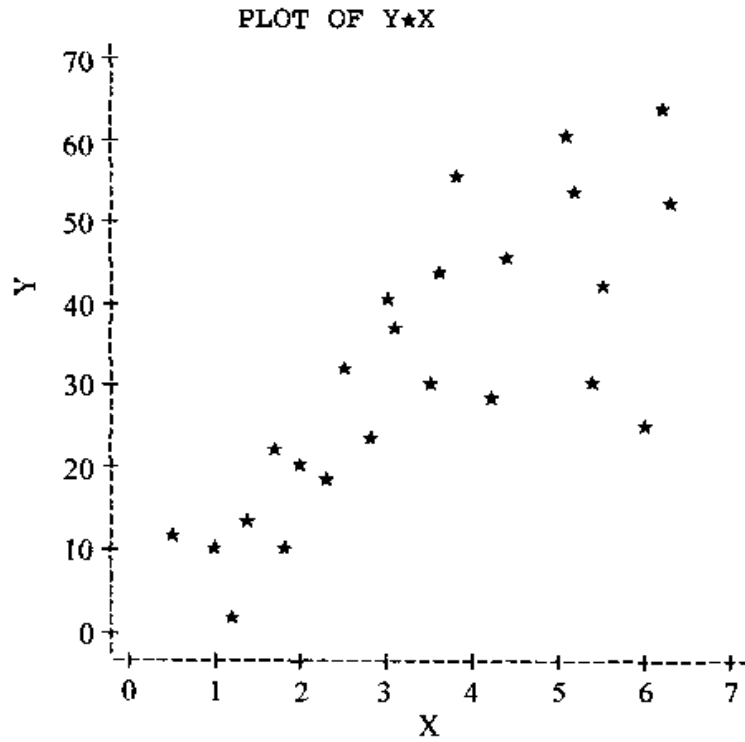
| Source   | DF | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|----------|----|-------------------|----------------|---------|--------|
| Model    | 2  | 4458.40552        | 2229.20276     | 20.448  | 0.0001 |
| Error    | 21 | 2289.42782        | 109.02037      |         |        |
| C Total  | 23 | 6747.83333        |                |         |        |
| Root MSE |    | 10.44128          | R-square       | 0.6607  |        |
| Dep Mean |    | 31.91667          | Adj R-sq       | 0.6284  |        |
| C.V.     |    | 32.71420          |                |         |        |

Parameter Estimates

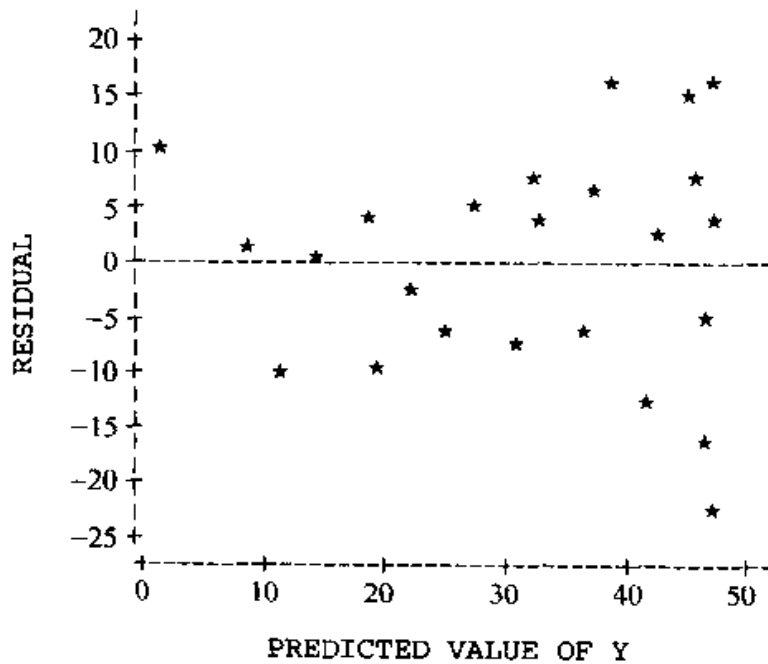
| Variable | DF | Parameter<br>Estimate | Standard<br>Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|-----------------------|-------------------|----------------------------|---------|
| INTERCEP | 1  | -6.871747             | 8.87156858        | -0.775                     | 0.4472  |
| X        | 1  | 17.105361             | 5.73414408        | 2.983                      | 0.0071  |
| X2       | 1  | -1.349036             | 0.79186923        | -1.704                     | 0.1032  |

Variable

| Variable  | DF | Label       |
|-----------|----|-------------|
| INTERCEPT | 1  | Intercept   |
| X         | 1  |             |
| X2        | 1  | X - SQUARED |



PLOT OF RESIDUALS VERSUS PREDICTED VALUE OF Y





回归模型中方差不等会有哪些影响呢?首先,如果回归直线的方差不是常数,最小二乘估计就不是尽可能准确的估计。这时,使用一种称为**加权最小二乘**的方法可能更准确[见 Draper 和 Smith(1998)]。更重要的是,加权最小二乘改进了模型参数的检验统计量( $F$  和  $t$  检验)和参数的区间估计,因为它们通常基于更小的标准误差。

当方差不等时,受影响更重的是  $E(y)$  的估计和  $y$  的预测。在这些方面, $y$  的点估计可能是合理的,但区间的长度可能太大或太小,这依赖于我们是否过大或过小地估计了试验地区的方差。

要消除不等方差的影响可以使用加权最小二乘,但在本书中我们不涉及这个方法。当不等方差与  $y$  之间存在某种关系时,对  $y$  的变换也可以解决问题。在第十一章讨论了几种  $y$  的变换,在第八章方差分析中讨论了有助于稳定方差的方法,这些都可以用于某些回归的情形。

### 例 13.15

参照例 13.14 的数据,我们已经检查了方差不等的问题。因为沿着回归直线的方差有随着  $x$  增加的趋势,对  $y$  的平方根变换可以用来稳定方差。根据计算机的输出和残差图来判断方差不等问题是否已解决。SAS 产生如下输出:

#### DATA LISTING

| OBS | Y  | X   | SQRT_Y  | X_SQUARED |
|-----|----|-----|---------|-----------|
| 1   | 11 | 0.5 | 3.31662 | 0.25      |
| 2   | 10 | 1.0 | 3.16228 | 1.00      |
| 3   | 2  | 1.2 | 1.41421 | 1.44      |
| 4   | 14 | 1.4 | 3.74166 | 1.96      |
| 5   | 22 | 1.7 | 4.69042 | 2.89      |
| 6   | 10 | 1.8 | 3.16228 | 3.24      |
| 7   | 20 | 2.0 | 4.47214 | 4.00      |
| 8   | 19 | 2.3 | 4.35890 | 5.29      |
| 9   | 32 | 2.5 | 5.65685 | 6.25      |
| 10  | 23 | 2.8 | 4.79583 | 7.84      |
| 11  | 40 | 3.0 | 6.32456 | 9.00      |
| 12  | 37 | 3.1 | 6.08276 | 9.61      |
| 13  | 30 | 3.5 | 5.47723 | 12.25     |
| 14  | 43 | 3.6 | 6.55744 | 12.96     |
| 15  | 55 | 3.8 | 7.41620 | 14.44     |

| OBS | Y  | X   | SQRT_Y  | X_SQUARED |
|-----|----|-----|---------|-----------|
| 16  | 29 | 4.2 | 5.38516 | 17.64     |
| 17  | 45 | 4.4 | 6.70820 | 19.36     |
| 18  | 60 | 5.1 | 7.74597 | 26.01     |
| 19  | 53 | 5.2 | 7.28011 | 27.04     |
| 20  | 30 | 5.4 | 5.47723 | 29.16     |
| 21  | 42 | 5.5 | 6.48074 | 30.25     |
| 22  | 25 | 6.0 | 5.00000 | 36.00     |
| 23  | 63 | 6.2 | 7.93725 | 38.44     |
| 24  | 51 | 6.3 | 7.14143 | 39.69     |

Dependent Variable: SQRT\_Y      SQUARE ROOT OF Y

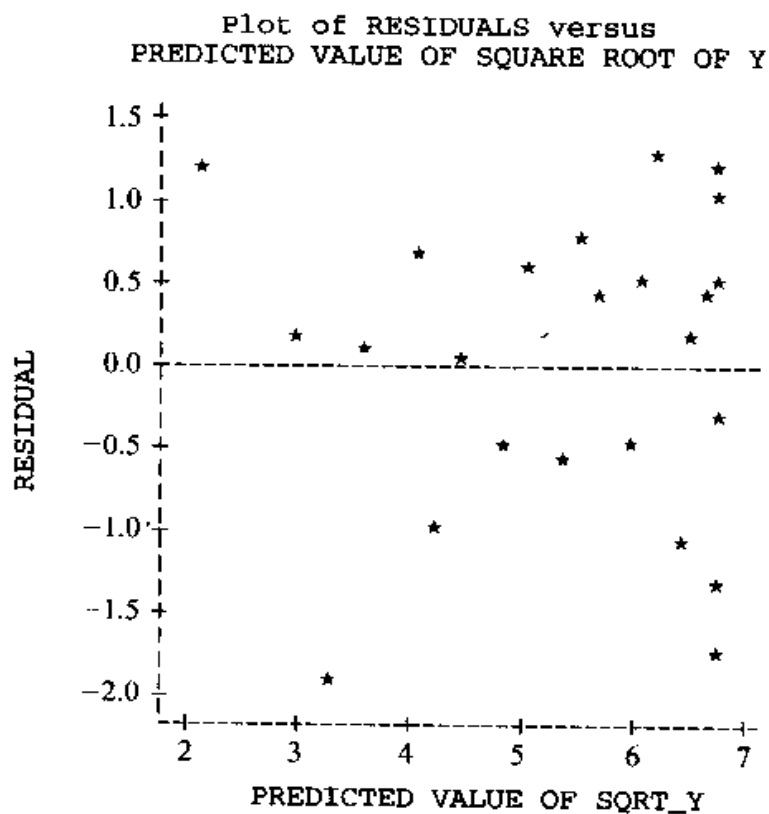
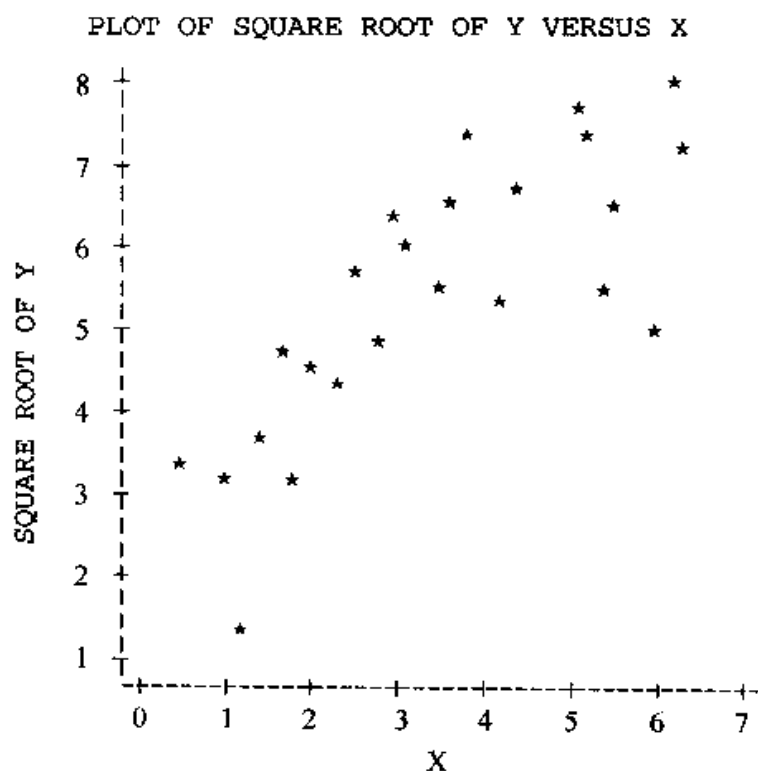
#### Analysis of Variance

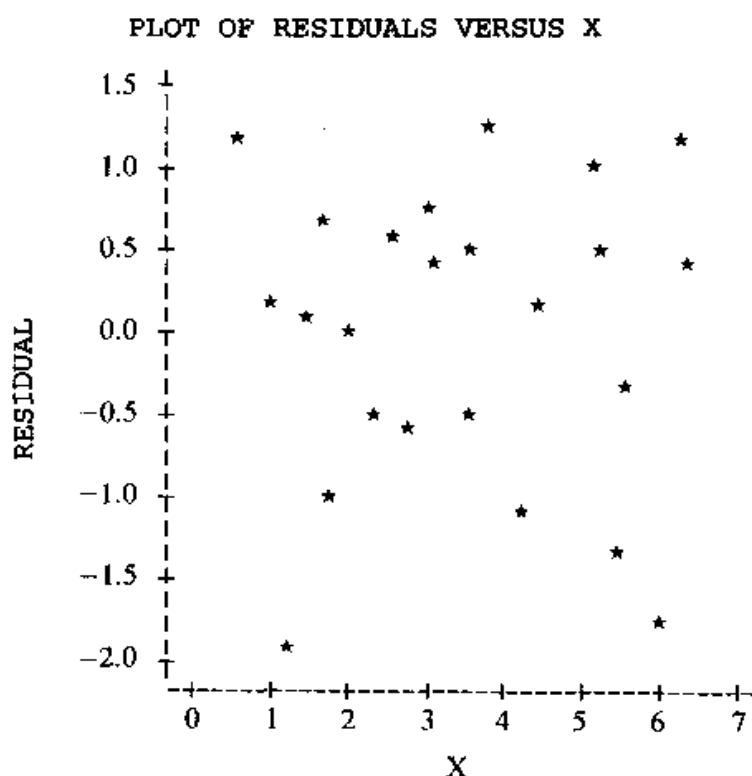
| Source   | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|----------|----|----------------|-------------|---------|--------|
| Model    | 2  | 45.00905       | 22.50453    | 24.683  | 0.0001 |
| Error    | 21 | 19.14653       | 0.91174     |         |        |
| C Total  | 23 | 64.15558       |             |         |        |
| Root MSE |    | 0.95485        | R-square    | 0.7016  |        |
| Dep Mean |    | 5.40773        | Adj R-sq    | 0.6731  |        |
| C.V.     |    | 17.65715       |             |         |        |

#### Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|----------------------------|---------|
| INTERCEP | 1  | 1.189795           | 0.81130083     | 1.467                      | 0.1573  |
| X        | 1  | 1.990218           | 0.52438482     | 3.795                      | 0.0011  |
| X2       | 1  | -0.176856          | 0.07241607     | -2.442                     | 0.0235  |

| Variable | DF | Variable Label |
|----------|----|----------------|
| INTERCEP | 1  | Intercept      |
| X        | 1  |                |
| X2       | 1  | X - SQUARED    |





**解答** 这里的输出表明这个模型对样本数据的拟合好得多,特别是从残差图上可以看出来。

多元回归的第三个假设是  $\epsilon_i$  的正态性, 使用某种散点图和残差图, 我们可以发现偏倚或离群值这些非正态性的表现。

直方图或茎叶图可以用来判断偏态。由假设,  $\epsilon_i$  服从均值为 0 的正态分布。如果残差的直方图不关于 0 对称, 就表明存在一定的偏倚。例如, 图 13.5(a) 的残差图关于 0 对称, 表明其无偏倚。相反, 图 13.5(b) 的残差图有右偏。

(a) 区间的中点

观察值的个数

|      |    |                                     |
|------|----|-------------------------------------|
| -2.0 | 3  | x x x                               |
| -1.5 | 10 | x x x x x x x x x x                 |
| -1.0 | 16 | x x x x x x x x x x x x x x x       |
| -0.5 | 15 | x x x x x x x x x x x x x x         |
| 0.0  | 20 | x x x x x x x x x x x x x x x x x x |
| 0.5  | 15 | x x x x x x x x x x x x x x         |
| 1.0  | 11 | x x x x x x x x x x                 |
| 1.5  | 6  | x x x x x                           |
| 2.0  | 3  | x x x                               |
| 2.5  | 0  |                                     |
| 3.0  | 1  | x                                   |

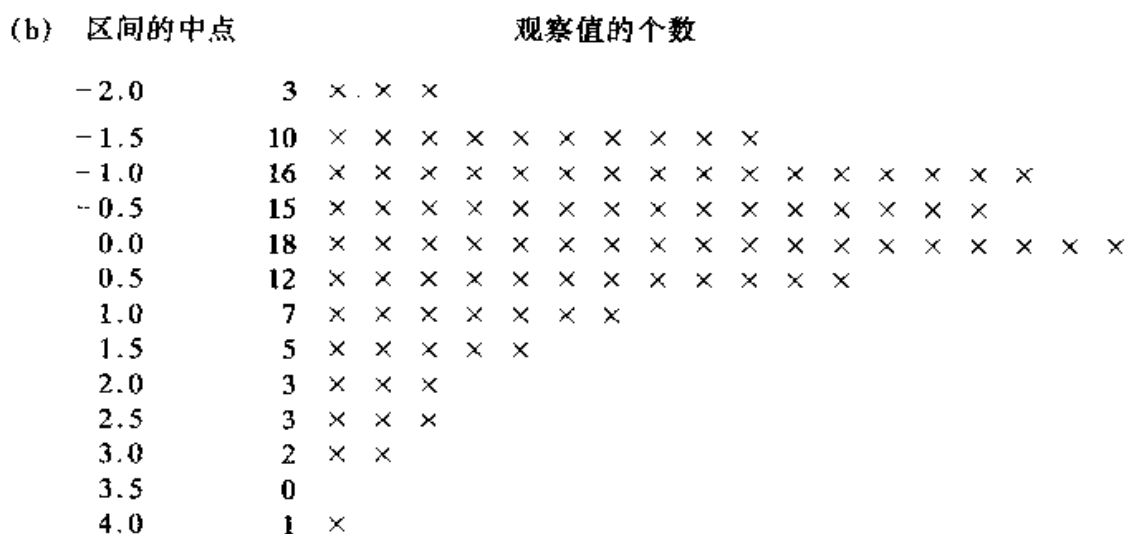
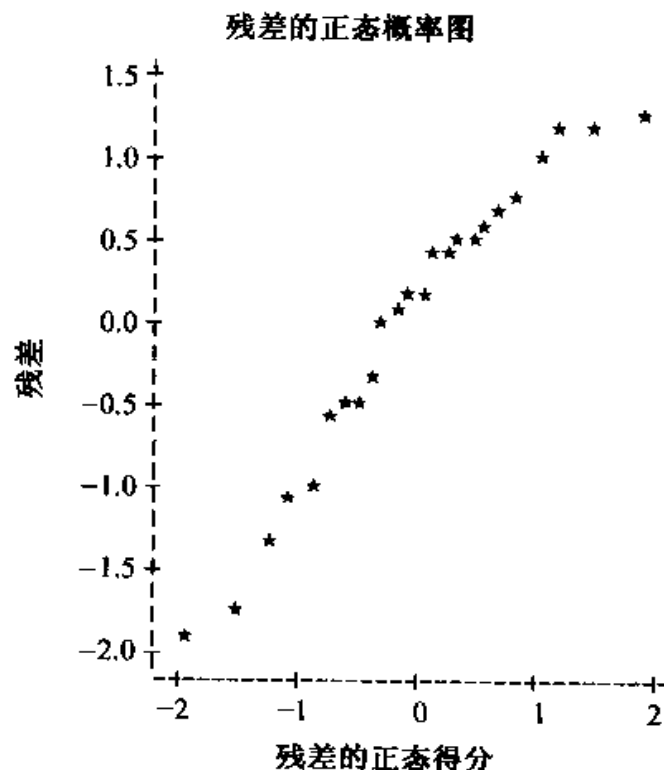


图 13.5 (a)残差以 0 为中心 (b)残差右偏

另一种检验非正态性的方法是使用残差的正态**概率图**。这个方法的思路是, 如果残差服从正态分布, 它的正态概率图应近似为一条直线。大多数统计软件都提供一个选项来做这种概率图。如果需要, 我们可以做出这个图。

### 例 13.16

参见例 13.15, 使用计算机输出结果来判断残差中是否存在非正态性的依据。



**解答** 正态概率图是近似于直线的,所以我们有理由认为残差服从正态分布。

离群点的出现是非正态性的更强的表现,可以用散点图或其他残差图来发现它。离群点是一个离其他数据较远的点。在第十一章,我们讲到过要注意杠杆点( $x$  离群点)和影响点( $x$  和  $y$  都离群)。一个强影响点可能会使回归方程变形。此外,某些离群点也表明对管理行为的要求,如果回归分析发现一块土地的价格比预想的低得多,那么这块土地很可能是一个好买卖;一个销售部门如果有比回归模型预测结果好得多的业绩,证明它的员工都很出色。相反的,比预测结果差的部门可能存在一些问题。有时可以把产生离群点的原因分离出来,有时则不能。离群点的产生可能是由于记录或录入错误,也可能是由在不同条件下进行观测产生的。如果发现是这种原因,可以纠正记录数据或者从分析中删去这个数据。如果没有明显的原因来纠正或删除离群点,就对包含和不包含离群数据两种情况分别作回归分析,看哪个模型更合理。无论离群点的来源或起因是什么,如果没能发现它们,就可能对回归方程产生严重的影响。

从线性回归模型  $y = \beta_0 + \beta_1 x + \varepsilon$  中,  $y$  对  $x$  的散点图可以用来判断离群点,见表 13.6 和图 13.6, 易见画圈的点为离群点。对表 13.6 的线性拟合的输出如下(带有残差图和正态概率图)。在每个图中,相应于可能的离群点(62, 125)的点也被画了圈。Minitab 程序产生了如下分析。

表 13.6 数据

| Obs      | $x$ | $y$ |
|----------|-----|-----|
| 1        | 10  | 120 |
| 2        | 20  | 115 |
| 3        | 21  | 250 |
| 4        | 27  | 210 |
| 5        | 29  | 300 |
| 6        | 33  | 330 |
| 7        | 40  | 295 |
| 8        | 44  | 400 |
| 9        | 52  | 380 |
| 10       | 56  | 460 |
| 11       | 62  | 125 |
| 12       | 68  | 510 |
| $N = 12$ |     |     |

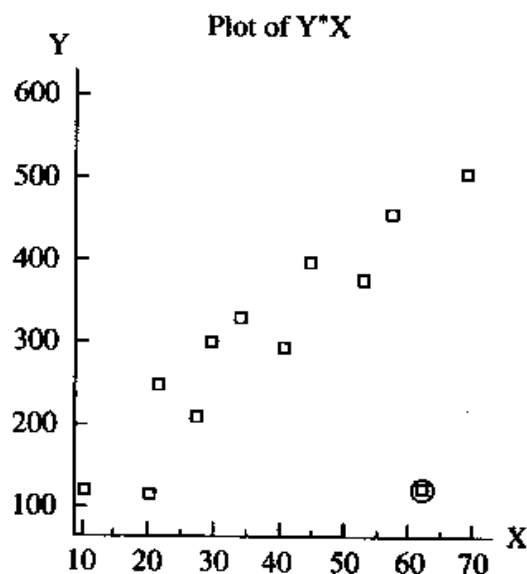
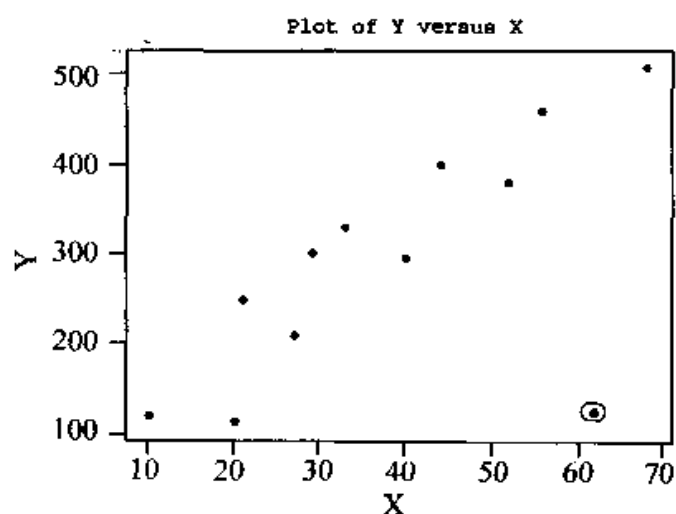


图 13.6 表 13.6 的散点图



Regression Analysis: y versus x

The regression equation is

$$y = 114 + 4.59 x$$

| Predictor | Coef   | SE Coef | T    | P     |
|-----------|--------|---------|------|-------|
| Constant  | 114.36 | 75.53   | 1.51 | 0.161 |
| x         | 4.595  | 1.787   | 2.57 | 0.028 |

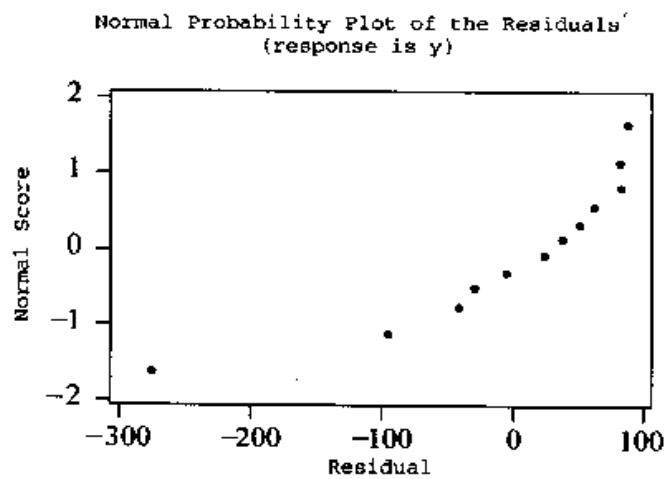
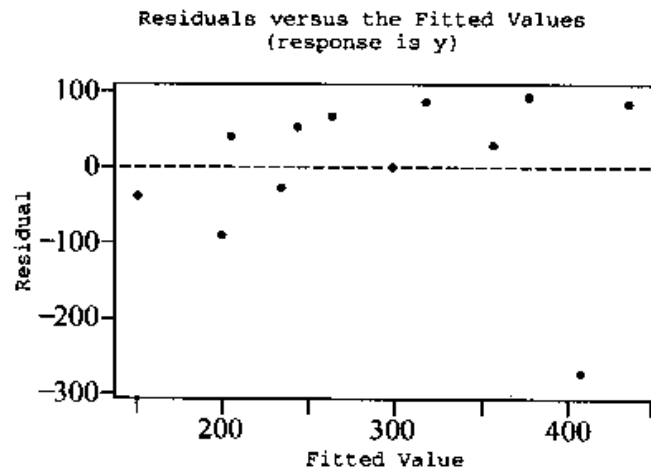
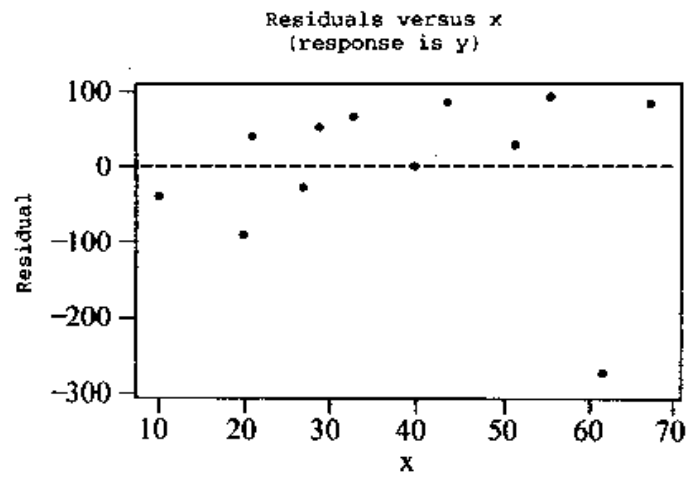
$$s = 108.1 \quad R-Sq = 39.8\% \quad R-Sq(adj) = 33.8\%$$

Analysis of Variance

| Source         | DF | SS     | MS    | F    | P     |
|----------------|----|--------|-------|------|-------|
| Regression     | 1  | 77201  | 77201 | 6.61 | 0.028 |
| Residual Error | 10 | 116755 | 11676 |      |       |
| Total          | 11 | 193956 |       |      |       |

| Obs | x    | y     | Fit   | SE Fit | Residual | Standardized Residual |
|-----|------|-------|-------|--------|----------|-----------------------|
| 1   | 10.0 | 120.0 | 160.3 | 59.7   | -40.3    | -0.45                 |
| 2   | 20.0 | 115.0 | 206.2 | 45.4   | -91.2    | -0.93                 |
| 3   | 21.0 | 250.0 | 210.8 | 44.2   | 39.2     | 0.40                  |
| 4   | 27.0 | 210.0 | 238.4 | 37.4   | -28.4    | -0.28                 |
| 5   | 29.0 | 300.0 | 247.6 | 35.5   | 52.4     | 0.51                  |
| 6   | 33.0 | 330.0 | 266.0 | 32.7   | 64.0     | 0.62                  |
| 7   | 40.0 | 295.0 | 298.1 | 31.3   | -3.1     | -0.03                 |
| 8   | 44.0 | 400.0 | 316.5 | 32.7   | 83.5     | 0.81                  |
| 9   | 52.0 | 380.0 | 353.3 | 39.4   | 26.7     | 0.27                  |
| 10  | 56.0 | 460.0 | 371.7 | 44.2   | 88.3     | 0.90                  |
| 11  | 62.0 | 125.0 | 399.2 | 52.3   | -274.2   | -2.90R                |
| 12  | 68.0 | 510.0 | 426.8 | 61.2   | 83.2     | 0.93                  |

R denotes an observation with a large standardized residual





这个数据集有助于显示查找离群点时出现的问题。有时,一个图是不够的。例如,散点图和正态概率图清楚的标识出离群点,而残差图就不太明显,因为离群点通过把拟合直线向自己方拉近而大大影响了拟合直线。这使得一些残差比没有离群点时变大。因此,在没有使用多个不同的方法进行检验前,不要忙于下结论。在多元回归中问题变得更困难了,因为在多元回归中无法画出简单的散点图。

另一个查出离群点的方法是刀切(jackknife)法。这种方法每次排除一个数据点,然后对剩余的数据计算回归系数。当离群值被排除时,回归系数有很大变化。这个过程也可以用于排除多个离群值,如一次排除2个或3个点,但寻找合适的回归方程的时间就变得很长。因此,虽然一次一个点的刀切法并不总能找出多个离群点,但也常被用于查找离群点。

实际上,把多种方法结合使用对检验离群值是比较有用的。首先, $x, y$ 的散点图可以查出一些离群点。对残差的检验可以证实这些可疑点。如果离群点和残差图都没有找出离群点,检验就可以结束了。然而一些特殊的离群点的识别可能需要额外的工作,如使用刀切法。

如果发现了离群点,该做什么呢?当然记录或抄写错误很容易纠正,有时,一个离群点显然来自另一总体。例如,财富杂志500强的公司肯定不属于小公司的研究范围,这种情况下,离群点就可以从数据中剔除。除非有绝对的理由,否则都不应轻易剔除数据。

### 例 13.17

设一组回归研究的数据如下:

$x$ : 10 13 16 18 20 22 24 27 30

$y$ : 31 35 42 45 51 53 59 31 70

画出数据散点图,表示出离群点,对包含和不包含离群点分别拟合一元回归方程。

**解答** 数据散点图(图 13.7)表明,除了  $x=27, y=31$  的点,斜率=2,截距=10 的直线很好的拟合了数据。如果包含那个点( $x=27, y=31$ ),最小二乘方程为

$$\hat{y} = 19.94 + 1.32x$$

如果不包括那个点,回归方程为:

$$\hat{y} = 9.93 + 2.00x$$

散点图清楚地表明观测(27, 31)是一个强影响点,在包含这个点时回归方程被扭曲。

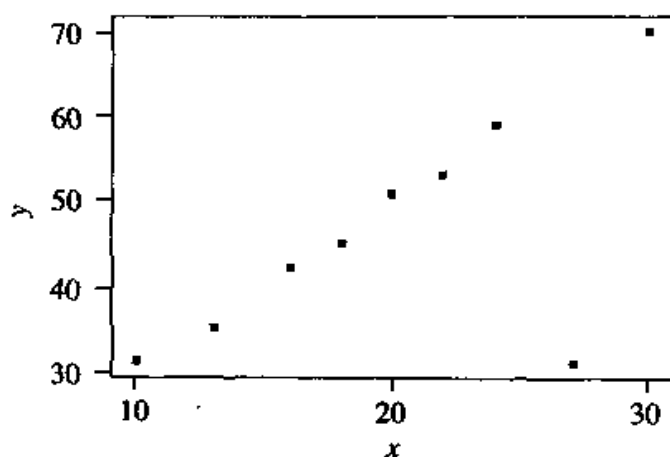


图 13.7 离群点的影响

**例 13.18**

对例 13.17 中的数据,使用刀切方法,每次剔除一个点,检查估计的斜率和截距来发现可能的离群点。

**解答** 我们使用计算机反复作回归分析,每次去掉一个点,如下是估计的斜率和截距。注意到最后两个点可能为离群点,因为去掉它们引起方程很大变化。

| 去掉的数据点 | 斜率      | 截距       |
|--------|---------|----------|
| 10.31  | 1.21286 | 22.47672 |
| 13.35  | 1.26116 | 21.42333 |
| 16.42  | 1.33281 | 19.55234 |
| 18.45  | 1.32834 | 19.60120 |
| 20.51  | 1.31953 | 19.35947 |
| 22.53  | 1.29235 | 19.97601 |
| 24.59  | 1.21563 | 21.04531 |
| 27.31  | 2.00354 | 9.93239  |
| 30.70  | 0.79712 | 28.42905 |

因此,虽然散点图 13.7 只标识出一个可能的离群点(点(27, 31)),使用刀切方法对残差的检查找出了第二个可能的离群点(点(30, 70))。但去掉点(27, 31)后所作的回归表明,点(30, 70)并不是离群点。

Regression Analysis: Y versus X (With the point (27, 31) Removed)

The regression equation is

$$Y = 9.93 + 2.00 X$$

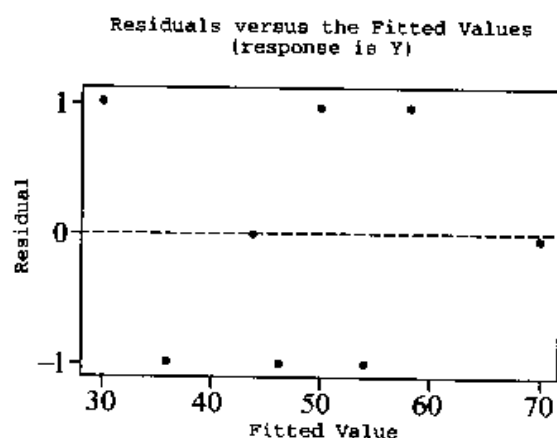
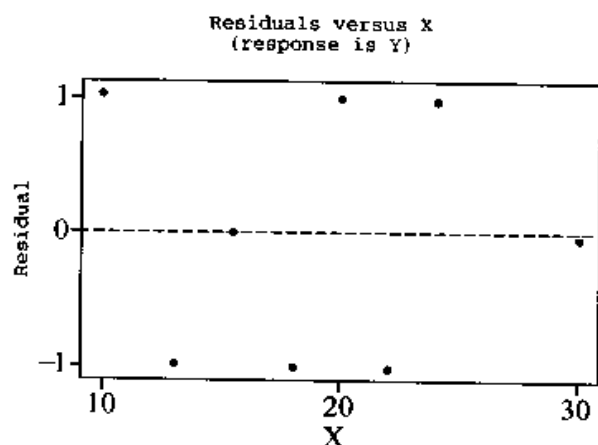
| Predictor | Coef    | SE Coef | T     | P     |
|-----------|---------|---------|-------|-------|
| Constant  | 9.932   | 1.190   | 8.34  | 0.000 |
| X         | 2.00354 | 0.05944 | 33.71 | 0.000 |

S = 0.9997 R-Sq = 99.5 % R-Sq(adj) = 99.4 %

Analysis of Variance

| Source         | DF | SS     | MS     | F       | P     |
|----------------|----|--------|--------|---------|-------|
| Regression     | 1  | 1135.5 | 1135.5 | 1136.17 | 0.000 |
| Residual Error | 6  | 6.0    | 1.0    |         |       |
| Total          | 7  | 1141.5 |        |         |       |

| Obs | X    | Y      | Fit    | SE Fit | Residual | St Resid |
|-----|------|--------|--------|--------|----------|----------|
| 1   | 10.0 | 31.000 | 29.968 | 0.647  | 1.032    | 1.36     |
| 2   | 13.0 | 35.000 | 35.978 | 0.507  | -0.978   | -1.14    |
| 3   | 16.0 | 42.000 | 41.989 | 0.399  | 0.011    | 0.01     |
| 4   | 18.0 | 45.000 | 45.996 | 0.360  | -0.996   | -1.07    |
| 5   | 20.0 | 51.000 | 50.003 | 0.357  | 0.997    | 1.07     |
| 6   | 22.0 | 53.000 | 54.010 | 0.393  | -1.010   | -1.10    |
| 7   | 24.0 | 59.000 | 58.017 | 0.457  | 0.983    | 1.11     |
| 8   | 30.0 | 70.000 | 70.038 | 0.737  | -0.038   | -0.06    |



最后一个假设是  $\epsilon_i$  相互独立, 因而不相关。当观测的时间次序已知时, 作为时间序列数据(即观测点是沿时间轴取得的), 可以构造残差对时间的图来观察残差是否序列相关。例如, 如果存在正序列相关, 在时间上相近的残差倾向于差不多; 负序列相关表明相近的残差是不同的。这种正、负序列相关的示意图分别如图

13.8(a)和图 13.8(b), 图 13.8(c)为无明显序列相关性的残差图。

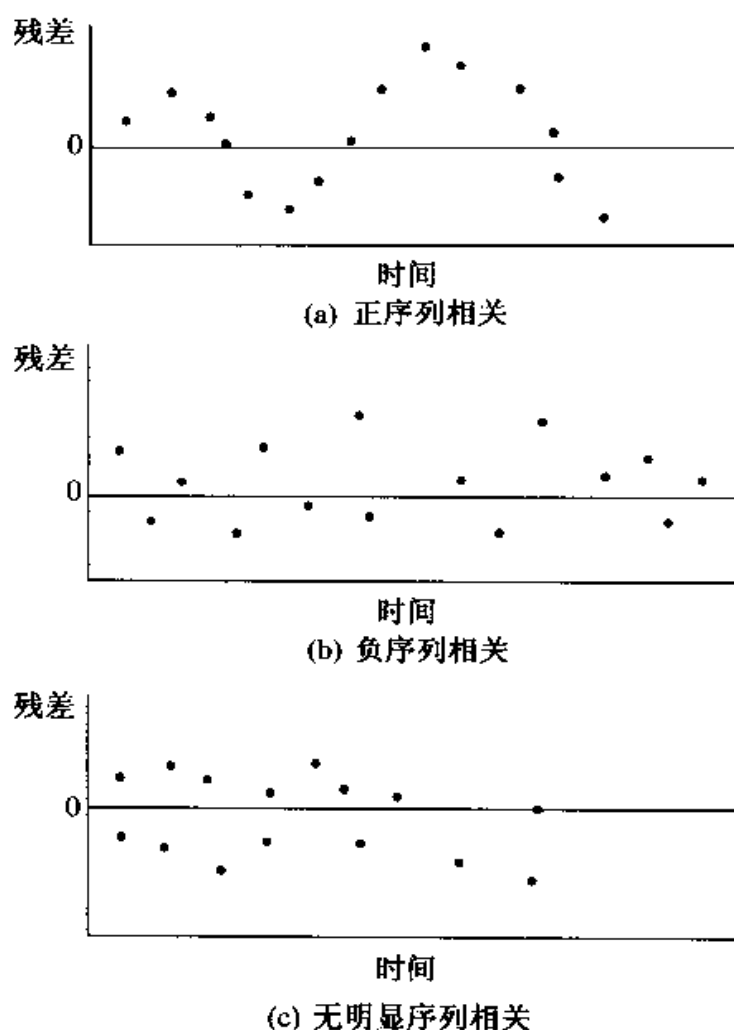


图 13.8

对序列相关的统计检验基于 Durbin-Watson 统计量。设  $\epsilon_t$  为在时刻  $t$  的残差,  $n$  为时间点的总数。那么 **Durbin-Watson 检验统计量**为

$$d = \frac{\sum_{t=1}^{n-1} (\epsilon_{t+1} - \epsilon_t)^2}{\sum_t \epsilon_t^2}$$

建立这个统计量的思路如下: 如果存在正序列相关, 相近的残差应相近, 它们差的平方  $(\epsilon_{t+1} - \epsilon_t)^2$  也应比残差不相关时小一些; 同样的, 如果残差间存在负相关, 残差间差的平方应比无相关时大。

如果不存在序列相关, Durbin-Watson 检验统计量  $d$  的值接近于 2.0, **正序列相关**时,  $d < 2.0$ , **负序列相关**时,  $d > 2.0$ 。J. Durbin 和 G. S. Watson(1951)年做出了  $d$  的临界值表,  $d$  小于 1.5(或大于 2.5)时, 人们就怀疑存在正(负)相关。

**例 13.19**

下面是某种个人电脑的月销售额。

| 月份, $x$ | 销售额(百万美元) $y$ |
|---------|---------------|
| 1       | 6.0           |
| 2       | 6.3           |
| 3       | 6.1           |
| 4       | 6.8           |
| 5       | 7.5           |
| 6       | 8.0           |
| 7       | 8.1           |
| 8       | 8.5           |
| 9       | 9.0           |
| 10      | 8.7           |
| 11      | 7.9           |
| 12      | 8.2           |
| 13      | 8.4           |
| 14      | 9.0           |

画出数据图, 根据线性回归方程画出残差对时间的图。从图中看得出序列相关吗?

**解答** 样本散点图和线性回归的残差图清楚地表明数据中存在序列相关。

| OBS | MONTH | COMPUTER SALES        |
|-----|-------|-----------------------|
|     | SALE  | (MILLIONS OF DOLLARS) |
| 1   | 1     | 6.0                   |
| 2   | 2     | 6.3                   |
| 3   | 3     | 6.1                   |
| 4   | 4     | 6.8                   |
| 5   | 5     | 7.5                   |
| 6   | 6     | 8.0                   |
| 7   | 7     | 8.1                   |
| 8   | 8     | 8.5                   |
| 9   | 9     | 9.0                   |

| OBS | MONTH | COMPUTER SALES<br>SALE (MILLIONS OF DOLLARS) |
|-----|-------|--|
| 10  | 10    | 8.7  |
| 11  | 11    | 7.9  |
| 12  | 12    | 8.2  |
| 13  | 13    | 8.4  |
| 14  | 14    | 9.0  |

Dependent Variables: Y      SALES (MILLIONS OF DOLLARS)

#### Analysis of Variance

| Source   | DF | Sum of<br>Squares | Mean<br>Square | F Value | Prob>F |
|----------|----|-------------------|----------------|---------|--------|
| Model    | 1  | 10.57540          | 10.57540       | 34.302  | 0.0001 |
| Error    | 12 | 3.69960           | 0.30830        |         |        |
| C Total  | 13 | 14.27500          |                |         |        |
| Root MSE |    | 0.55525           | R-square       | 0.7408  |        |
| Dep Mean |    | 7.75000           | Adj R - sq     | 0.7192  |        |
| C.V.     |    | 7.16449           |                |         |        |

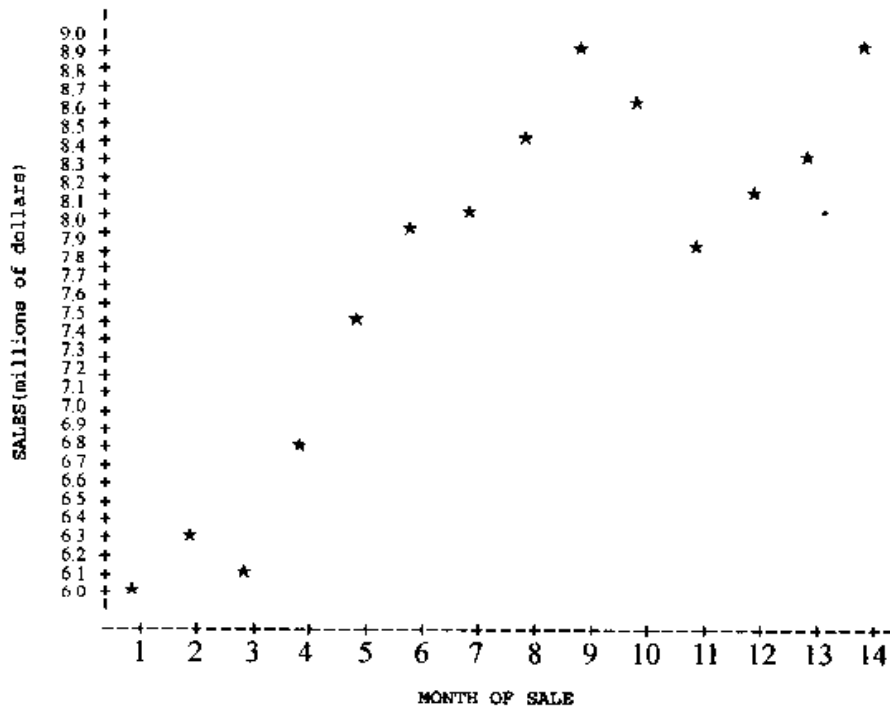
#### Parameter Estimates

| Variable | DF | Parameter<br>Estimate | Standard<br>Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|-----------------------|-------------------|----------------------------|---------|
| INTERCEP | 1  | 6.132967              | 0.31344787        | 19.566                     | 0.0001  |
| X        | 1  | 0.215604              | 0.03681259        | 5.857                      | 0.0001  |

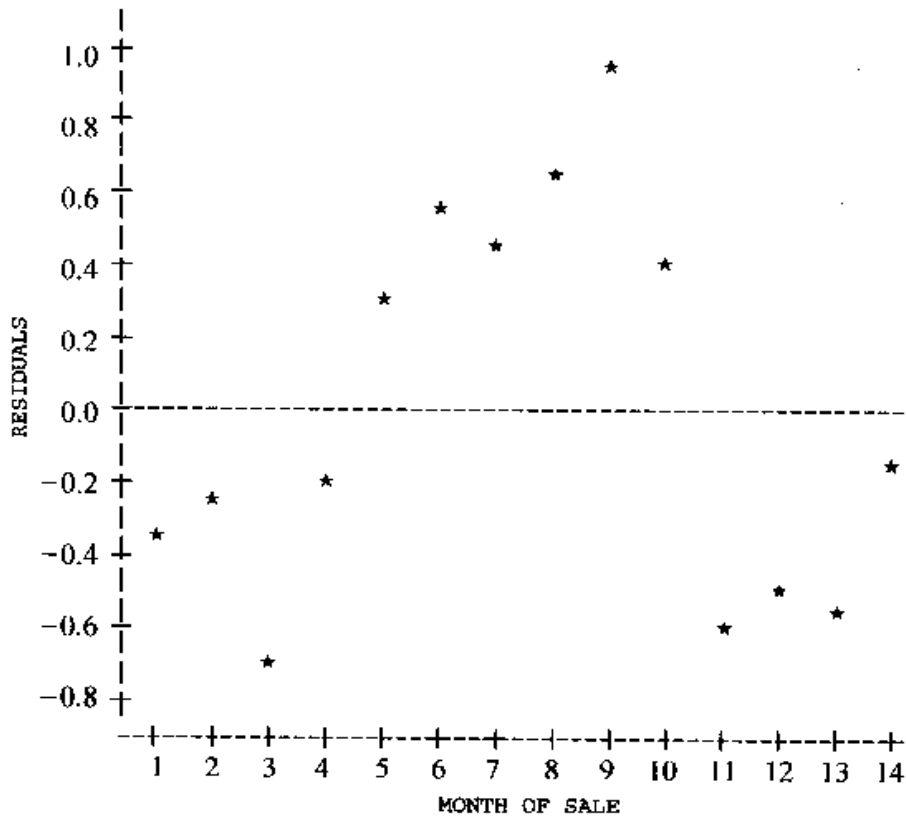
| Variable | DF | Variable<br>Label |
|----------|----|-------------------|
| INTERCEP | 1  | Intercept         |
| X        | 1  | MONTH             |

Durbin - Watson D      0.625  
(For Number of Obs.)      14  
1st Order Autocorrelation      0.668

Plot of SALES versus MONTH OF SALE



Plot of RESIDUALS versus MONTH OF SALE



**例 13.20**

计算例 13.19 的 Durbin-Watson 统计量的值。它是否证明了从图中看到的情况?

**解答** 根据例 13.19 的输出,  $d = 0.625$ , 因为这个值比 1.5 小得多, 我们认为存在正序列相关; 残差图也表明了这一点。

如果怀疑存在序列相关, 那么原本的多元回归模型就不太适合, 而需要另选一个。对当误差不独立时, 时间序列数据的分析研究进行了很多年, 本书中不准备一一介绍。在此, 只讨论一种基于一阶差分的简单回归方法。

基于一阶差分的回归很易使用, 同时, 它也是解决序列相关的一种粗糙的方法。作为  $y$  对  $x$  的简单线性回归, 我们计算差分  $y_t - y_{t-1}$  和  $x_t - x_{t-1}$ , 对  $n-1$  个  $y$  的差分 and  $n-1$  个  $x$  的差分作回归, 这可以帮助消除序列相关。如果没有消除, 可以参考其他时间序列的书来做更深入的分析。

我们讨论过的残差图对于回归模型拟合中的诊断问题是很有用的。然而, 由于残差也是随机变量, 它有时也会造成错误。一些研究人士认为, 使用“标准化”残差会好一点。一种特殊类型的标准化残差, 被称为“学生化残差”, 已经成为大多数软件(如 SAS)的输出。

如果你使用的软件使用标准化残差, 那么你就可以用标准化残差代替普通的残差来对回归模型的拟合程度进行评估。从理论上讲, 这些标准化残差的均值为 0, 方差为 1。大的残差是指那些绝对值较大(如大于 3)的残差。

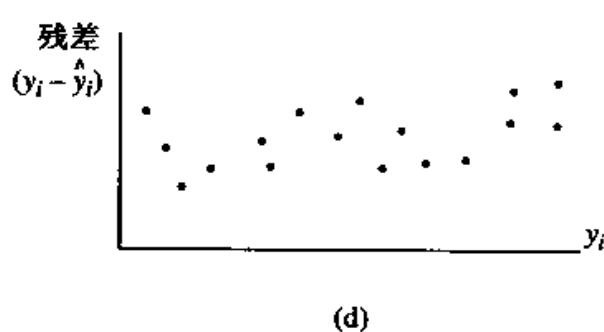
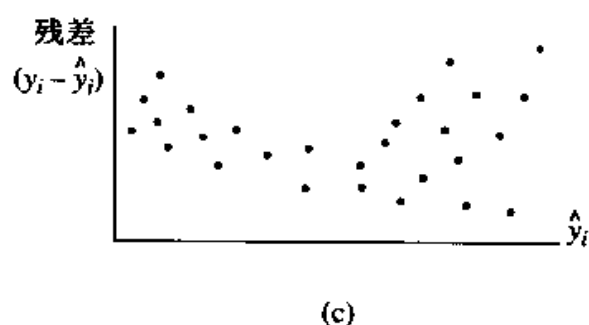
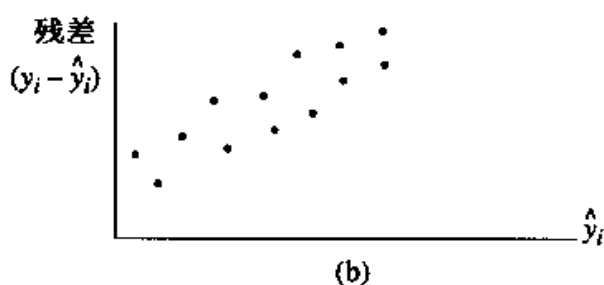
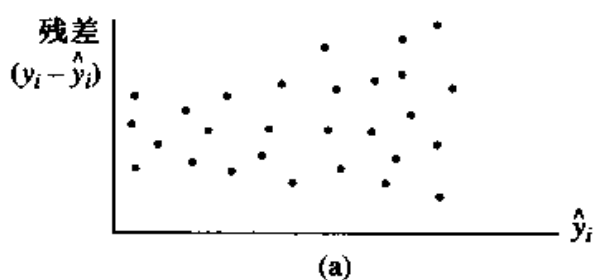
**练习****基本技能**

**13.23** 下面的残差图中显示了几种不同的模式。指出这些图是否存在问题, 如果存在, 指出问题和可能的解决方法。

**13.24** 参考例 13.14 的数据, 计算一阶差分并用  $y$  的差分 and  $x$  的差分进行回归。使用差分模型可否看出序列相关? 你使用什么图来得出结论?

**13.25** 练习 13.24 中数据的 Durbin-Watson 统计量的值为多少? 它同你之前的结论一致吗?





## 应用

**13.26(社会学)** 一位社会学学者研究每 100 人中非暴力犯罪的比率( $y$ )与 5 年前这一比率( $x_1$ )和当前失业率( $x_2$ )之间的关系。下面是 20 个不同城市的数据。

| CITY | PRESENT<br>RATE | RATE<br>5 YEARS<br>AGO | PRESENT<br>UNEMPLOYMENT<br>RATE |
|------|-----------------|------------------------|---------------------------------|
| 1    | 13              | 14                     | 5.1                             |
| 2    | 8               | 10                     | 2.7                             |
| 3    | 14              | 16                     | 4.0                             |
| 4    | 10              | 10                     | 3.4                             |
| 5    | 12              | 16                     | 3.1                             |
| 6    | 11              | 12                     | 4.3                             |
| 7    | 7               | 8                      | 3.8                             |
| 8    | 6               | 7                      | 3.2                             |
| 9    | 10              | 12                     | 3.2                             |
| 10   | 16              | 20                     | 4.1                             |
| 11   | 16              | 14                     | 5.9                             |
| 12   | 9               | 10                     | 4.0                             |
| 13   | 11              | 10                     | 4.1                             |
| 14   | 18              | 20                     | 5.0                             |
| 15   | 9               | 13                     | 3.1                             |

| CITY | PRESENT | RATE           | PRESENT              |
|------|---------|----------------|----------------------|
|      | RATE    | 5 YEARS<br>AGO | UNEMPLOYMENT<br>RATE |
| 16   | 10      | 6              | 6.3                  |
| 17   | 15      | 10             | 5.7                  |
| 18   | 14      | 14             | 5.2                  |
| 19   | 17      | 16             | 4.9                  |
| 20   | 6       | 8              | 3.0                  |

使用这里的输出:

**a. 拟合模型**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

**b. 考察回归模型的假设, 讨论假设是否成立; 如果不成立, 给出可能的改进措施。**

SAS OUTPUT FOR EXERCISES 13.26

DATA LISTING

| OBS | RATE | RATE_5 | UNEMPLOY | RT5_UNEP |
|-----|------|--------|----------|----------|
| 1   | 13   | 14     | 5.1      | 71.4     |
| 2   | 8    | 10     | 2.7      | 27.0     |
| 3   | 14   | 16     | 4.0      | 64.0     |
| 4   | 10   | 10     | 3.4      | 34.0     |
| 5   | 12   | 16     | 3.1      | 49.6     |
| 6   | 11   | 12     | 4.3      | 51.6     |
| 7   | 7    | 8      | 3.8      | 30.4     |
| 8   | 6    | 7      | 3.2      | 22.4     |
| 9   | 10   | 12     | 3.2      | 38.4     |
| 10  | 16   | 20     | 4.1      | 82.0     |
| 11  | 16   | 14     | 5.9      | 82.6     |
| 12  | 9    | 10     | 4.0      | 40.0     |
| 13  | 11   | 10     | 4.1      | 41.0     |
| 14  | 18   | 20     | 5.0      | 100.0    |
| 15  | 9    | 13     | 3.1      | 40.3     |
| 16  | 10   | 6      | 6.3      | 37.8     |
| 17  | 15   | 10     | 5.7      | 57.0     |
| 18  | 14   | 14     | 5.2      | 72.8     |

| OBS | RATE | RATE _ 5 | UNEMPLOY | RT5 _ UNEP |
|-----|------|----------|----------|------------|
| 19  | 17   | 16       | 4.9      | 78.4       |
| 20  | 6    | 8        | 3.0      | 24.0       |

#### MULTIPLE REGRESSION ANALYSIS

Dependent Variable: RATE NONVIOLENT CRIME RATE PER 1000

#### Analysis of Variance

| Source   | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|----------|----|----------------|-------------|---------|--------|
| Model    | 3  | 234.27348      | 78.09116    | 67.442  | 0.0001 |
| Error    | 16 | 18.52652       | 1.15791     |         |        |
| C Total  | 19 | 252.80000      |             |         |        |
| Root MSE |    | 1.07606        | R-square    | 0.9267  |        |
| Dep Mean |    | 11.60000       | AdjR-sq     | 0.9130  |        |
| C.V.     |    | 9.27639        |             |         |        |

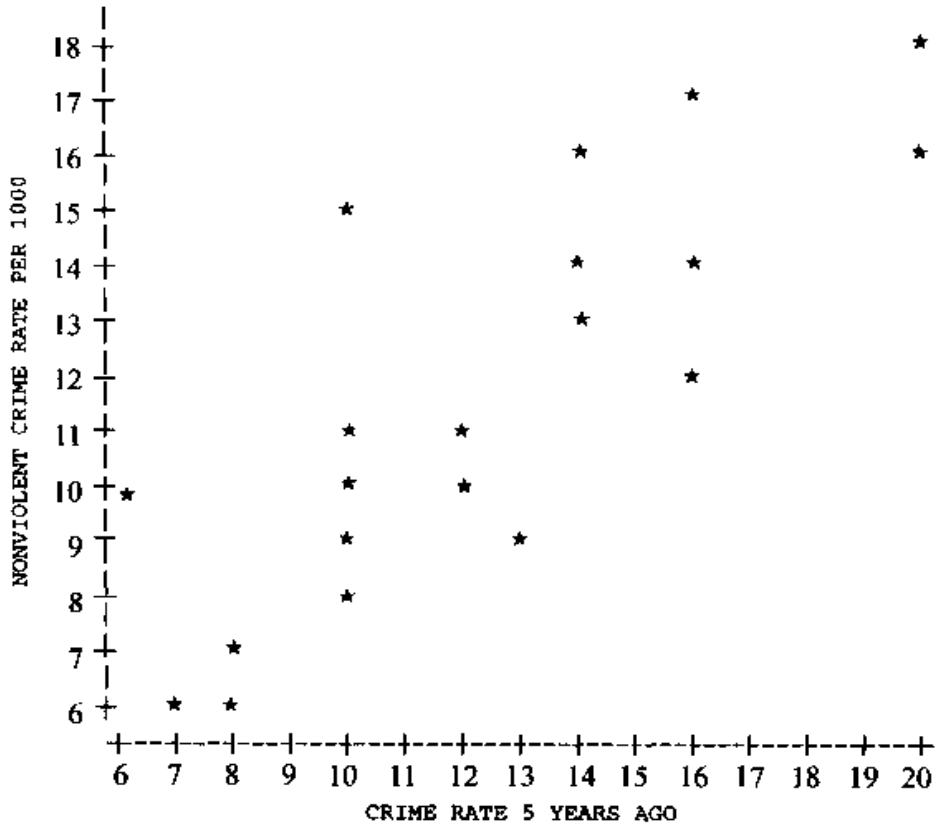
#### Parameter Estimates

| Variable   | DF | Parameter Estimate | Standard Error | T for H0: Parameter = 0 | Prob> T |
|------------|----|--------------------|----------------|-------------------------|---------|
| INTERCEP   | 1  | -2.704052          | 3.37622689     | -0.801                  | 0.4349  |
| RATE _ 5   | 1  | 0.517215           | 0.30264512     | 1.709                   | 0.1068  |
| UNEMPLOY   | 1  | 1.449811           | 0.74635173     | 1.943                   | 0.0699  |
| RT5 _ UNEP | 1  | 0.035338           | 0.06631783     | 0.533                   | 0.6015  |

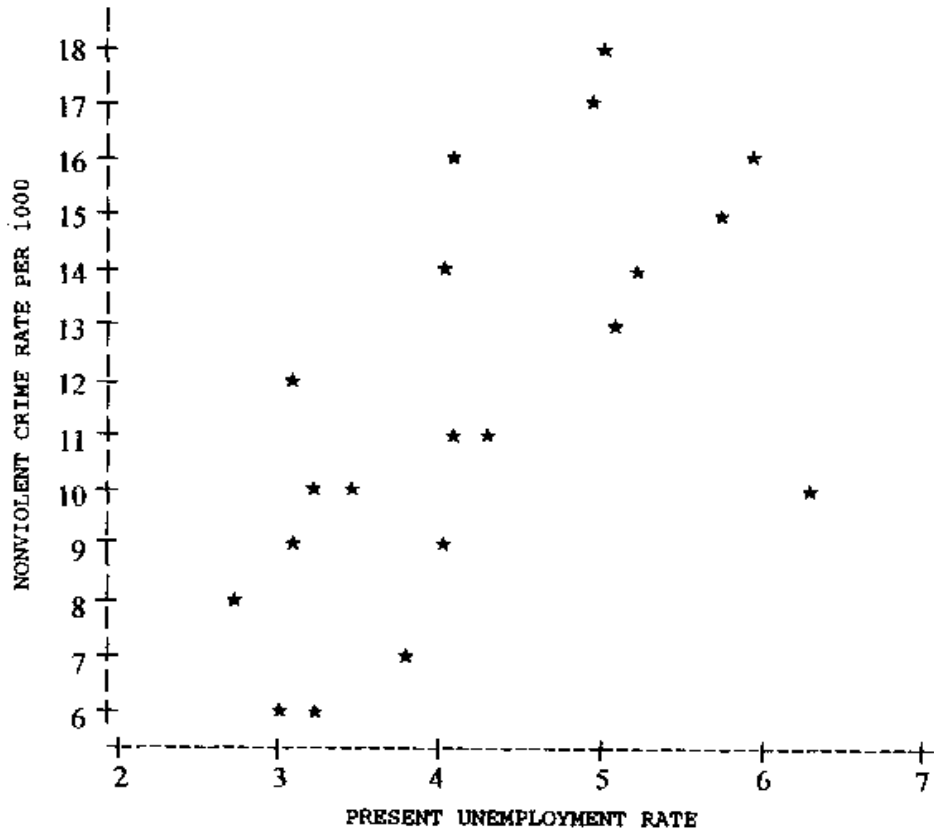
| Variable   | DF | Variable Label            |
|------------|----|---------------------------|
| INTERCEP   | 1  | Intercept                 |
| RATE _ 5   | 1  | CRIME RATE 5 YEARS AGO    |
| UNEMPLOY   | 1  | PRESENT UNEMPLOYMENT RATE |
| RT5 _ UNEP | 1  | RATE _ 5 TIMES UNEMPLOY   |

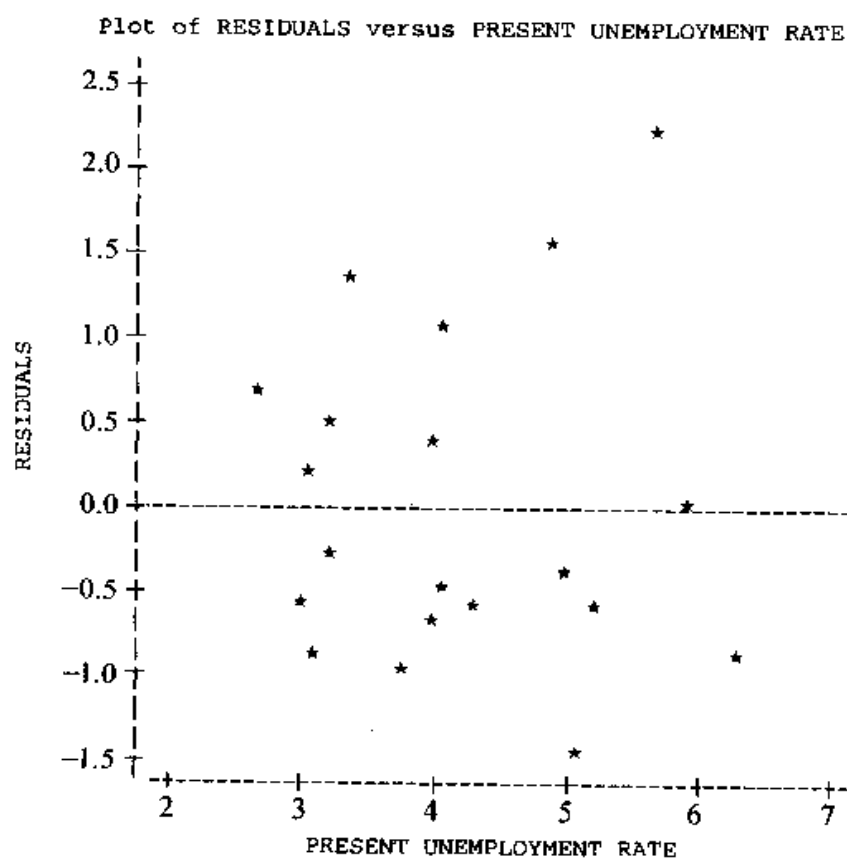
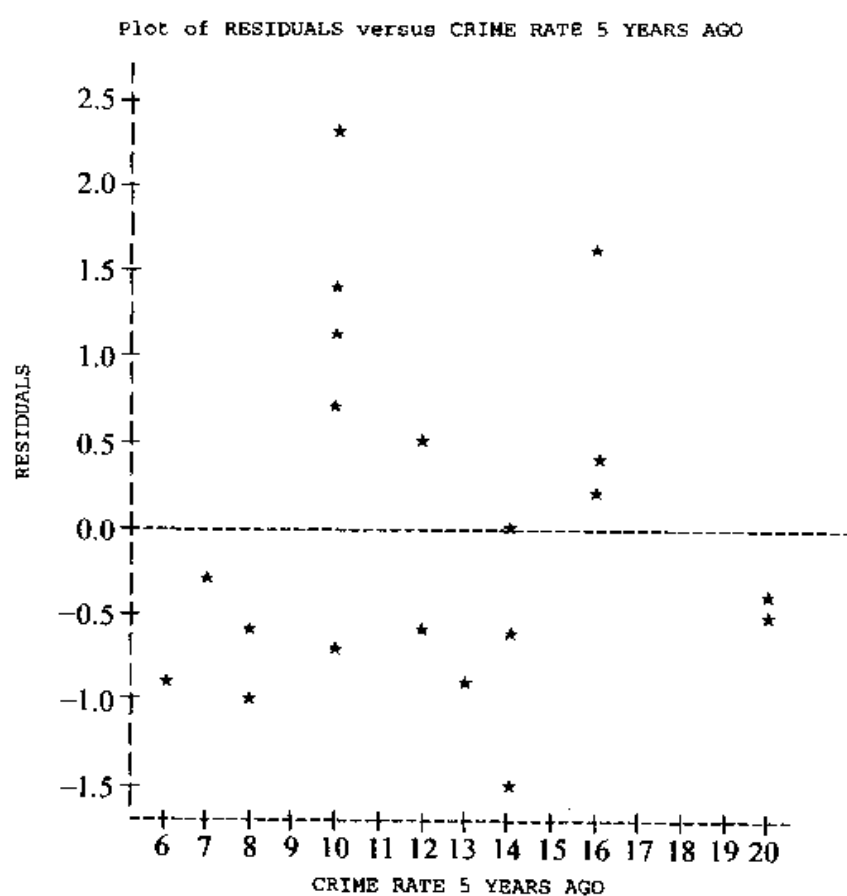
Durbin-Watson D 2.403  
 (For Number of Obs.) 20  
 1st Order Autocorrelation -0.269

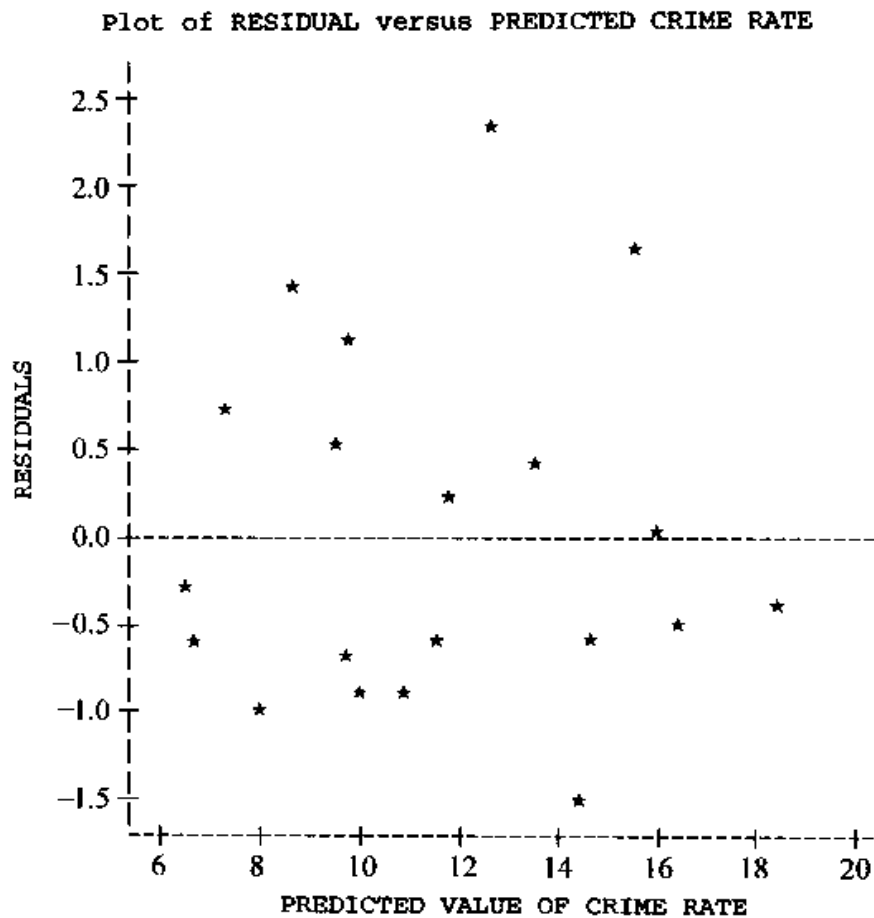
Plot of PRESENT NONVIOLENT CRIME RATE versus CRIME RATE 5 YEARS AGO



Plot of PRESENT NONVIOLENT CRIME RATE versus PRESENT UNEMPLOYMENT RATE



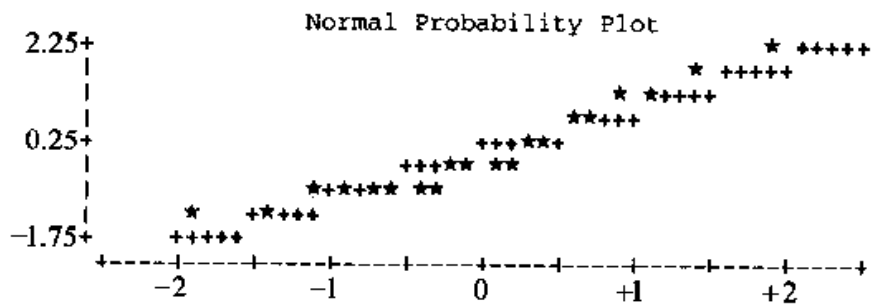




## SAS UNIVARIATE PROCEDURE FOR RESIDUAL ANALYSIS

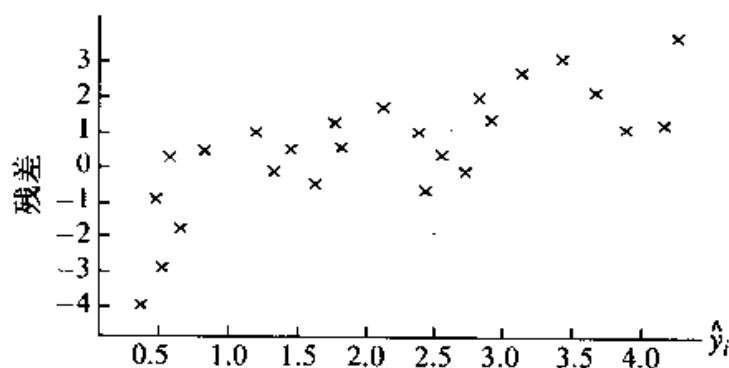
Variable=Residual

| Stem | Leaf    | # | Boxplot |
|------|---------|---|---------|
| 2    | 3       | 1 |         |
| 1    | 6       | 1 |         |
| 1    | 14      | 2 |         |
| 0    | 57      | 2 | +-----+ |
| 0    | 24      | 2 | +       |
| -0   | 430     | 3 | *-----* |
| -0   | 9976665 | 7 | +-----+ |
| -1   | 0       | 1 |         |
| -1   | 5       | 1 |         |



**13.27** 参照练习 13.26, 为一个 5 年前犯罪率为 9%、当前失业率为 16% 的城市预测现在的犯罪率。这个预测存在问题吗? 为什么?

**13.28** 一个证券公司对 25 家不同的高科技企业每股收益作了回归分析, 下图画出了估计值  $\hat{y}$  和残差, 是否存在迹象表明假设不成立? 有没有其他的检验或图表?



## 13.5 小 结

这一章主要涉及了多元回归中的一些实际问题。做法的第一步是决定因变量和包含在模型中的备选自变量集。我们讨论了从相关领域专家中得到信息的重要性, 和最优子集回归方法在挑选自变量中的用处。

在第二步中讨论了多元回归方程的一种特殊形式: 多项式回归。特别的, 要关注一个假定模型的拟合不足问题。如果自变量集上有重复观测, 就可以对多项式模型的拟合不足问题进行检验。拟合不足也可以用残差图来考察。

如同我们所讨论的那样, 按照第一步和第二步去做, 有时也会有问题, 这与所得到的数据有关。例如, 如果可以同时得到很多变量的数据, 就可以建立多元回归模型, 然后可以请教专家或使用最优子集回归来挑选(第一步)得到的自变量。基于第一步的讨论和分析, 在第二步中可以对模型进行进一步的修正。然而, 有时并不能得到很多可能的自变量的值, 在这种情况下, 第一步中专家的建议就很重要了。根据专家指出的重要因子来收集数据, 在得到这些备选变量的数据后, 再使用子集回归和第二步中建模的方法来改进模型。

多元回归问题的最后一步是检验多元回归的假设: 零期望、等方差、正态性、独立性。虽然也提出了一些正式的检验, 但检查假设的最好的方法是使用散点图和概率残差图。一般情况下, 一个人在使用散点图考察和解释数据的经验越丰富, 所得到的回归方程就越好。

## 重要公式

1.  $C_p$  统计量

$$C_p = \frac{SS(\text{残差})_p}{s_\epsilon^2} - (n - 2p)$$

## 2. Durbin-Watson 统计量

$$d = \frac{\sum_{t=1}^{n-1} (\epsilon_{t+1} - \epsilon_t)^2}{\sum_t \epsilon_t^2}$$

## 3. 向后剔除

$$F_d = \frac{SS_{\text{drop}_j}}{MS(\text{残差})}, \quad j = 1, 2, \dots$$

## 补充练习

**13.29** 使用如下数据拟合一个模型,画出数据图并给出一个多项式模型。

|     |    |    |    |    |    |    |    |    |    |    |    |    |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|
| $y$ | 7  | 8  | 6  | 12 | 15 | 13 | 7  | 10 | 11 | 14 | 16 | 17 |
| $x$ | 10 | 10 | 10 | 15 | 15 | 15 | 20 | 20 | 20 | 25 | 25 | 25 |

**13.30** 参照练习 13.29 的数据。

a. 拟合模型

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon$$

b. 取  $\alpha = 0.05$ , 检验拟合不足。

c. 检验残差图来验证回归假设是否正确。

**13.31** 参照练习 13.29, 设第 3、5、6、10 个数据缺失。

a. 拟合一个三次模型。

b. 检查残差并比较练习 13.30 和 13.31 的拟合程度。

**13.32(医药)** 一个制药公司希望得到药量与效用间关系的信息,为此,把 15 个被注入病毒的试管在  $30^\circ\text{C}$  下培养了 5 天,3 组试管随机地分配给 5 个不同的待测剂量(2, 4, 8, 16, 32mg),每个试管只注入一剂,并得到其响应(一种产品对病毒环境抵抗力的度量),数据如下:



| 剂量 | 响应         |
|----|------------|
| 2  | 5, 7, 3    |
| 4  | 10, 12, 14 |
| 8  | 15, 17, 18 |
| 16 | 20, 21, 19 |
| 32 | 23, 24, 29 |

- 画出数据图。
- 对数据分别拟合线性和二次模型。
- 哪个模型更准确？
- 把在(b)中得到的结果与 SAS 的输出结果作比较。

SAS OUTPUT FOR EXERCISE 13.32

DATA LISTING

| OBS | DOSE | RESPONSE | DOSE2 |
|-----|------|----------|-------|
| 1   | 2    | 5        | 4     |
| 2   | 2    | 7        | 4     |
| 3   | 2    | 3        | 4     |
| 4   | 4    | 10       | 16    |
| 5   | 4    | 12       | 16    |
| 6   | 4    | 14       | 16    |
| 7   | 8    | 15       | 64    |
| 8   | 8    | 17       | 64    |
| 9   | 8    | 18       | 64    |
| 10  | 16   | 20       | 256   |
| 11  | 16   | 21       | 256   |
| 12  | 16   | 19       | 256   |
| 13  | 32   | 23       | 1024  |
| 14  | 32   | 24       | 1024  |
| 15  | 32   | 29       | 1024  |

REGRESSION ANALYSIS WITH LINEAR DOSE TERM IN MODEL

Dependent Variable: RESPONSE POTENCY OF DRUG

Analysis of Variance

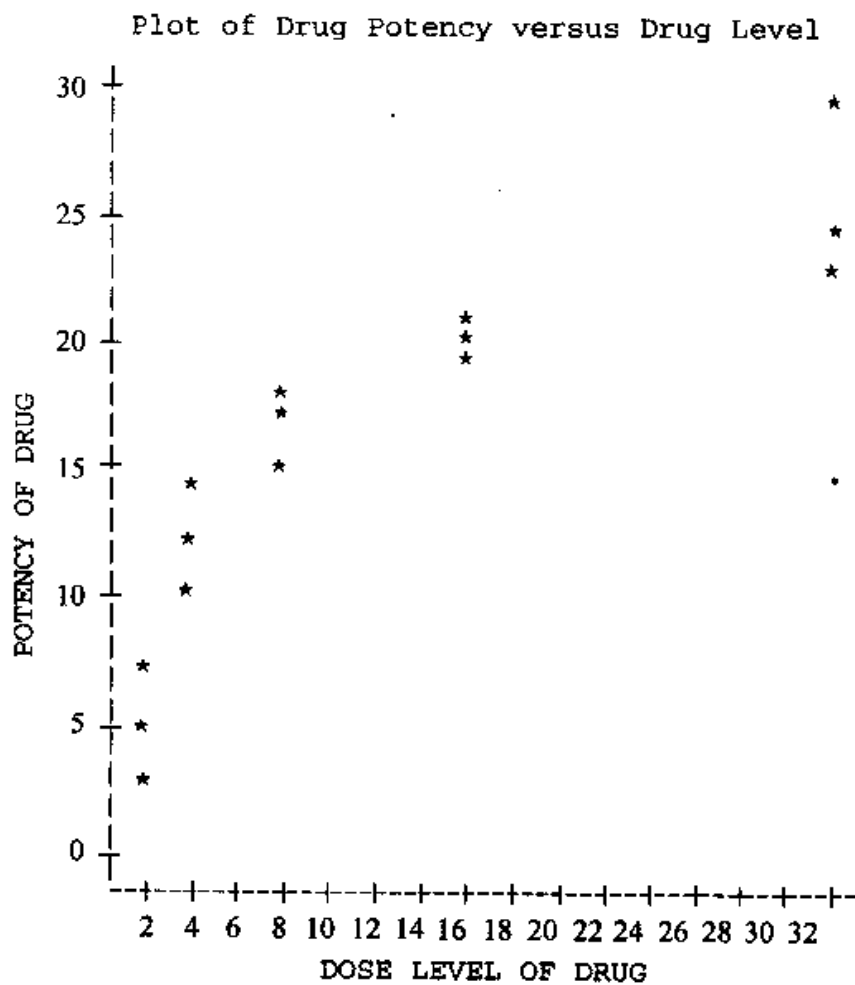
| Source  | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|----|----------------|-------------|---------|--------|
| Model   | 1  | 590.91613      | 590.91613   | 44.280  | 0.0001 |
| Error   | 13 | 173.48387      | 13.34491    |         |        |
| C Total | 14 | 764.40000      |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| Root MSE | 3.65307  | R-square | 0.7730 |
| Dep Mean | 15.80000 | Adj R-sq | 0.7556 |
| C. V.    | 23.12069 |          |        |

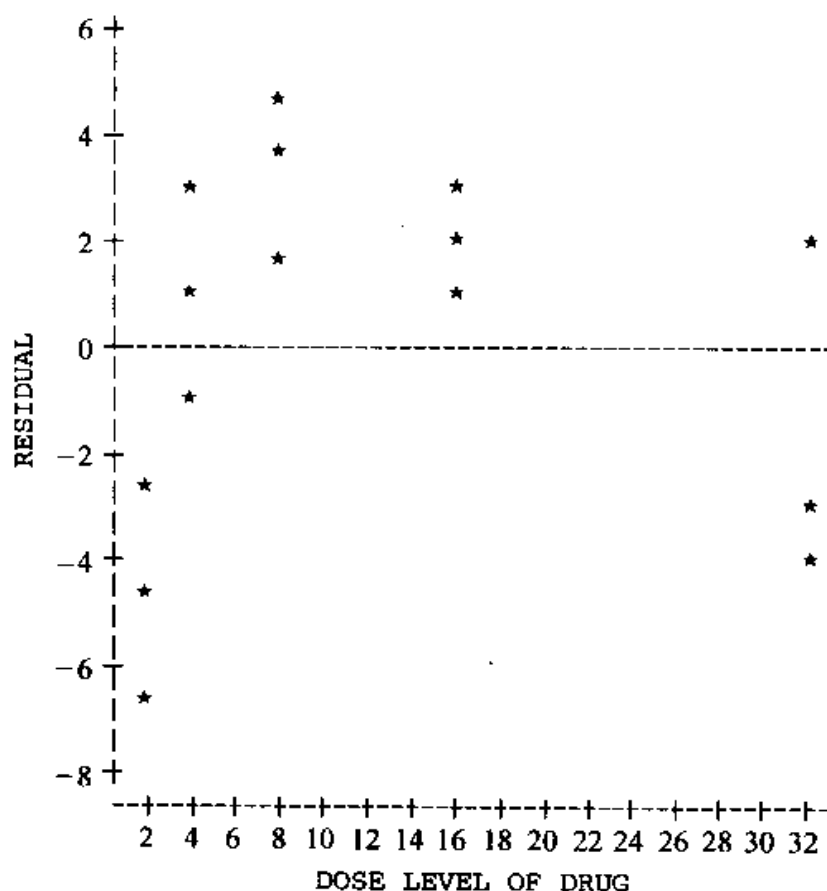
## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H <sub>0</sub> :<br>Parameter = 0 | Prob >  T |
|----------|----|--------------------|----------------|---|-----------|
| INTERCEP | 1  | 8.666667           | 1.42786770     | 6.070                                   | 0.0001    |
| DOSE     | 1  | 0.575269           | 0.08645016     | 6.654                                   | 0.0001    |

| Variable | DF | Variable Label     |
|----------|----|--------------------|
| INTERCEP | 1  | Intercept          |
| DOSE     | 1  | DOSE LEVEL OF DRUG |



Plot of Residuals (linear model) versus Dose Level



# REGRESSION ANALYSIS WITH QUADRATIC TERM IN DOSE

Dependent Variable: RESPONSE POTENCY OF DRUG

## Analysis of Variance

| Source   | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|----------|----|----------------|-------------|---------|--------|
| Model    | 2  | 673.82062      | 336.91031   | 44.634  | 0.0001 |
| Error    | 12 | 90.57938       | 7.54828     |         |        |
| C Total  | 14 | 764.40000      |             |         |        |
| Root MSE |    | 2.74741        | R - square  | 0.8815  |        |
| Dep Mean |    | 15.80000       | Adj R - sq  | 0.8618  |        |
| C. V.    |    | 17.38869       |             |         |        |

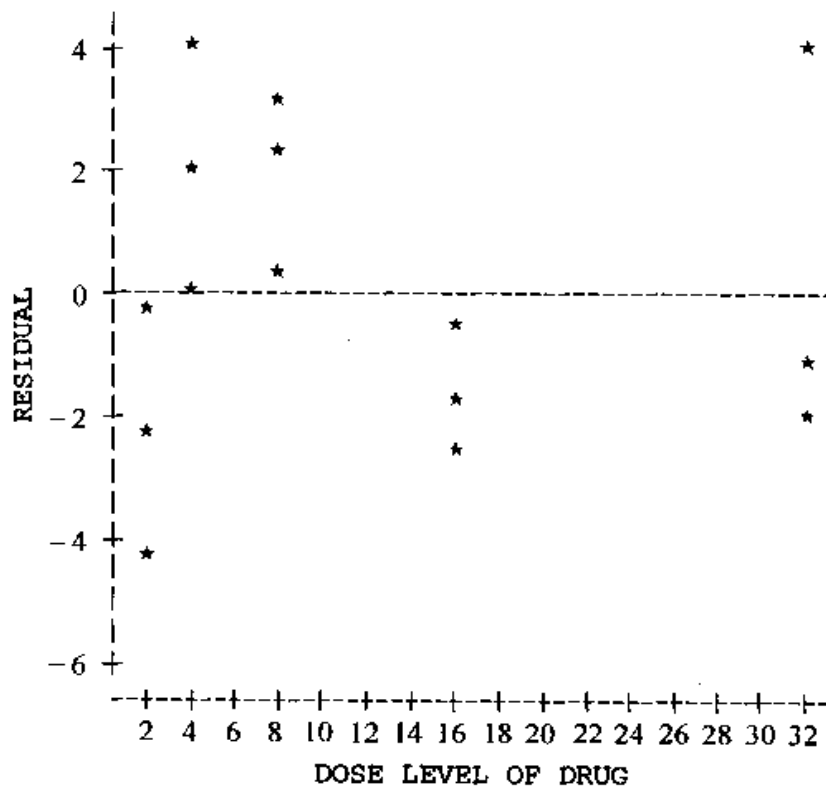
## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter = 0 | Prob>  T |
|----------|----|--------------------|----------------|-------------------------|----------|
| INTERCEP | 1  | 4.483660           | 1.65720388     | 2.706                   | 0.0191   |
| DOSE     | 1  | 1.506325           | 0.28836373     | 5.224                   | 0.0002   |

|       |   |           |            |        |        |
|-------|---|-----------|------------|--------|--------|
| DOSE2 | 1 | -0.026987 | 0.00814314 | -3.314 | 0.0062 |
|-------|---|-----------|------------|--------|--------|

| Variable | DF | Label              |
|----------|----|--------------------|
| INTERCEP | 1  | Intercept          |
| DOSE     | 1  | DOSE LEVEL OF DRUG |
| DOSE2    | 1  | DOSE SQUARED       |

Plot of Residuals (quadratic model) versus Dose Level



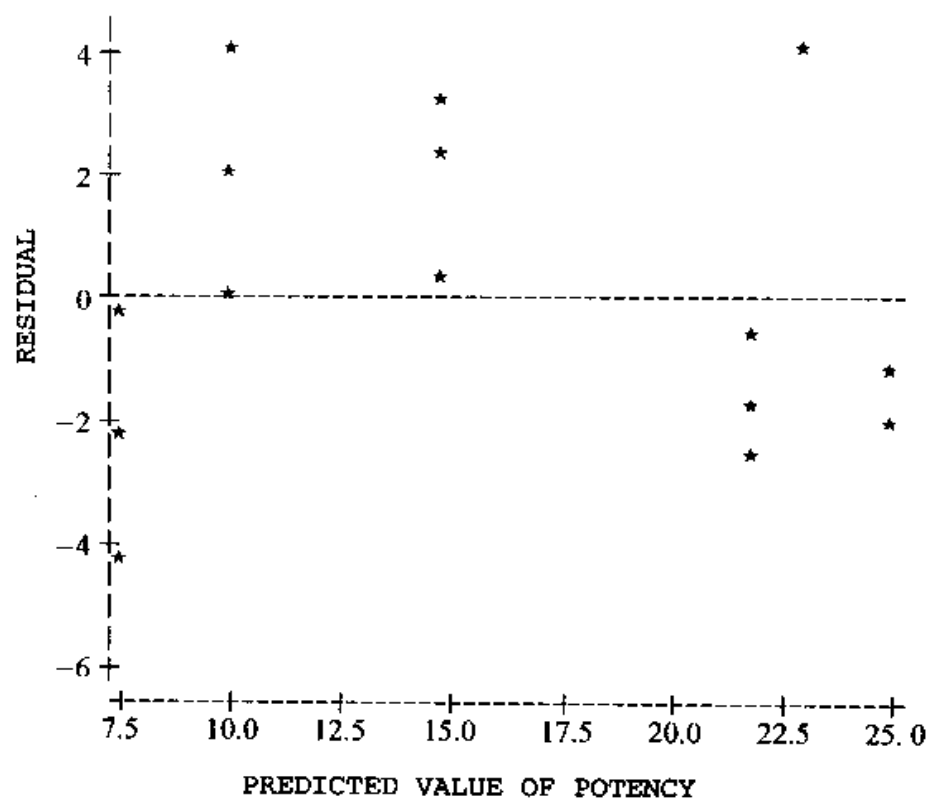
**13.33** 参见练习 13.32 中的数据。许多时候,可以用剂量水平的对数变换来线性化响应与自变量之间的关系。

- 查对数表或用电子计算器得到 5 个剂量水平的对数。
- 用  $x_1$  记对数剂量,拟合模型

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

- 把(b)中的结果与下面的计算机输出做比较。
- 这三个模型中,哪一个看起来更合适?为什么?

Plot of Residuals (quadratic model) versus Predicted Potency



SAS OUTPUT FOR EXERCISE 13.33

DATA LISTING

| OBS | DOSE | RESPONSE | LOG _ DOSE |
|-----|------|----------|------------|
| 1   | 2    | 5        | 0.69315    |
| 2   | 2    | 7        | 0.69315    |
| 3   | 2    | 3        | 0.69315    |
| 4   | 4    | 10       | 1.38629    |
| 5   | 4    | 12       | 1.38629    |
| 6   | 4    | 14       | 1.38629    |
| 7   | 8    | 15       | 2.07944    |
| 8   | 8    | 17       | 2.07944    |
| 9   | 8    | 18       | 2.07944    |
| 10  | 16   | 20       | 2.77259    |
| 11  | 16   | 21       | 2.77259    |
| 12  | 16   | 19       | 2.77259    |
| 13  | 32   | 23       | 3.46574    |
| 14  | 32   | 24       | 3.46574    |
| 15  | 32   | 29       | 3.46574    |

Dependent Variable: RESPONSE POTENCY OF DRUG

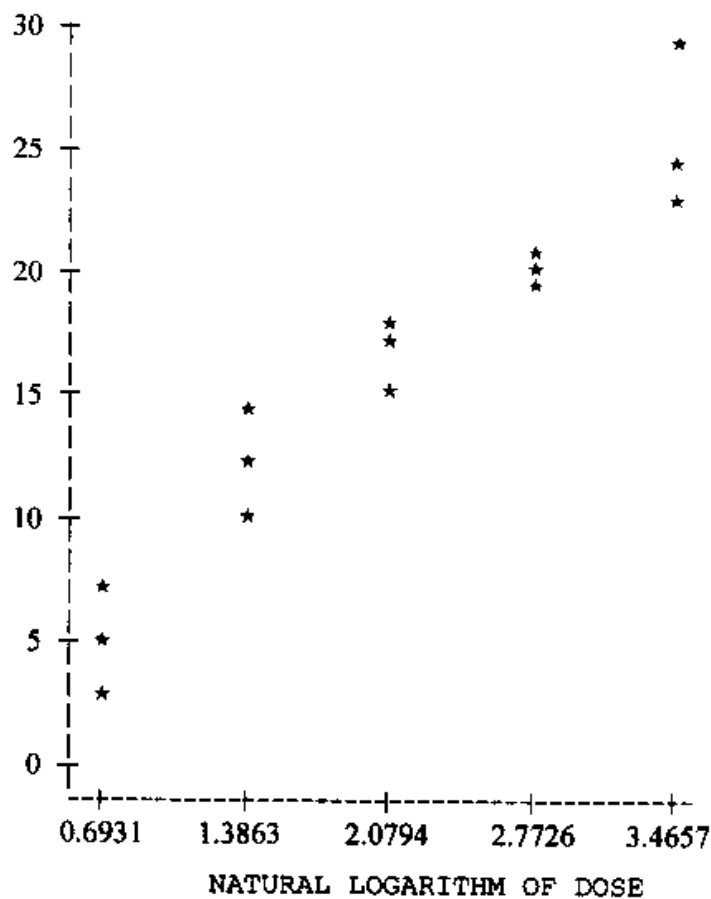
Analysis of Variance

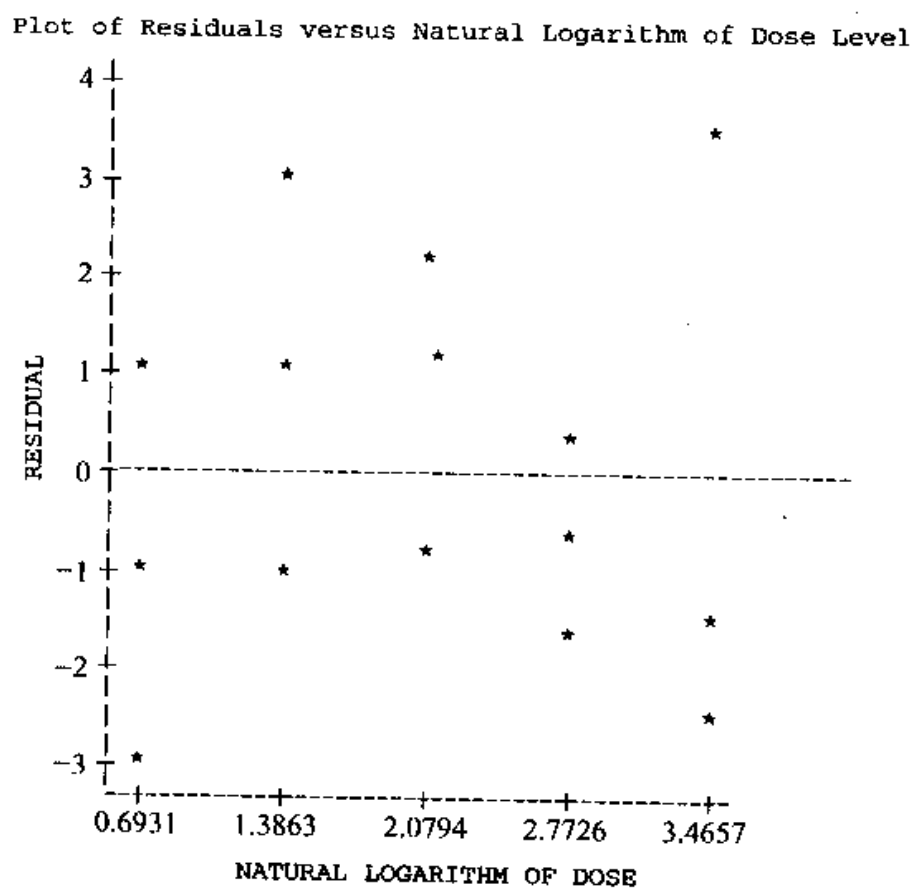
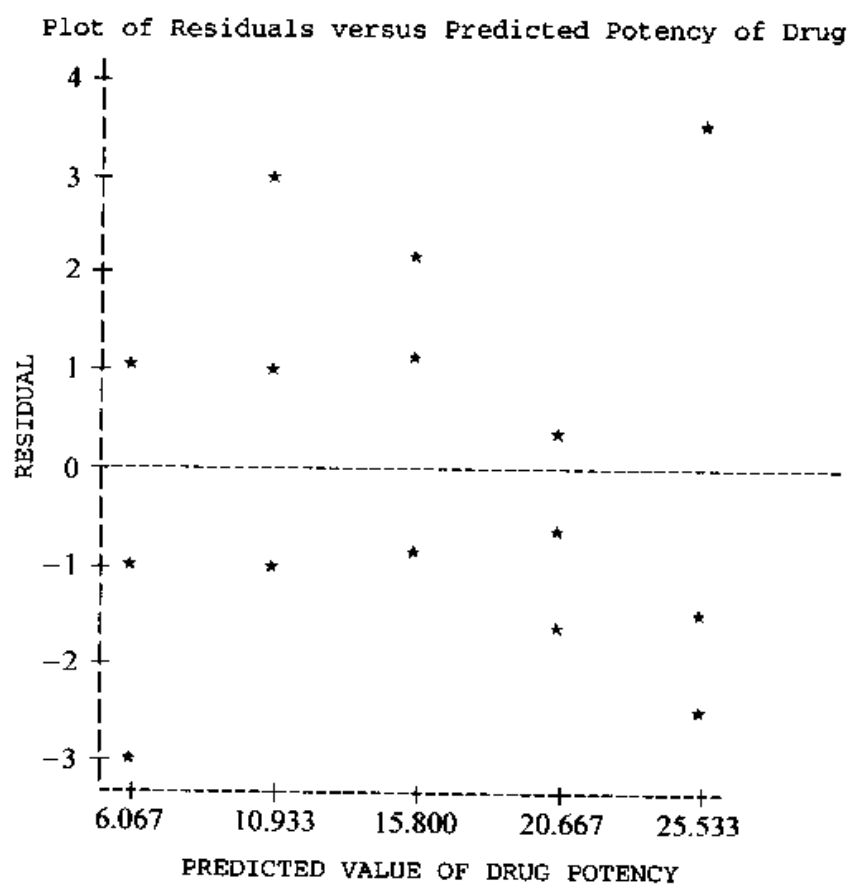
| Source   | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|----------|----|----------------|-------------|---------|--------|
| Model    | 1  | 710.53333      | 710.53333   | 171.478 | 0.0001 |
| Error    | 13 | 53.86667       | 4.14359     |         |        |
| C Total  | 14 | 764.40000      |             |         |        |
| Root MSE |    | 2.03558        | R-square    | 0.9295  |        |
| Dep Mean |    | 15.80000       | Adj R-sq    | 0.9241  |        |
| C.V.     |    | 12.88342       |             |         |        |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|----------------------------|---------|
| INTERCEP | 1  | 1.200000           | 1.23260547     | 0.974                      | 0.3480  |
| LOG_DOSE | 1  | 7.021116           | 0.53616972     | 13.095                     | 0.0001  |

| Variable | DF | Variable Label            |
|----------|----|---------------------------|
| INTERCEP | 1  | Intercept                 |
| LOG_DOSE | 1  | NATURAL LOGARITHM OF DOSE |

Plot of Drug Potency versus Natural Logarithm of Dose Level  
DRUG POTENCY



**13.34(工程)** 做一项试验以研究一种新型广告涂料抵抗天气的能力,自变量有两个:温度  $x_1$  和暴露时间  $x_2$ 。样本数据列出如下:

|                         |     |     |     |     |    |     |
|-------------------------|-----|-----|-----|-----|----|-----|
| $y$                     | 120 | 101 | 110 | 105 | 92 | 130 |
| $x_1(^{\circ}\text{C})$ | -10 | -10 | 0   | 0   | 10 | 10  |
| $x_2(\text{月})$         | 1   | 3   | 2   | 2   | 1  | 3   |

a. 拟合模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

b. 考察残差,并对你的发现做出评注。

**13.35** 参考练习 13.34。

a. 我们能够拟合下列模型吗?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \beta_6 x_1 x_2^2 + \beta_7 x_1^2 x_2 + \beta_8 x_1^2 x_2^2 + \epsilon$$

b. 检查练习 13.34 中模型的拟合不足,给出建议。

**13.36(工程)** 对于试验纤维用的磨损测试机的研磨效果,用一种特殊纤维在六个不同的机器转速下进行了测试。剪下 48 块 5 英寸见方的这种纤维织物,为六个转速 100, 120, 140, 160, 180 和 200(转/分钟)中的每一个随机指定 8 块布料。布料指定给机器的顺序是随机的,每块布料在指定的机器速度上研磨 3 分钟,测量并记下磨损的程度。数据见下面的表。

a. 画出每个转速下的平均值,建议一个模型。

b. 用数据拟合建议的模型。

c. 要检查这个模型的假定,可以用哪种残差图?

| 机器速度(转/分钟) | 磨损程度   |
|------------|--|
| 100        | 23.0, 23.5, 24.4, 25.2, 25.6, 26.1, 24.8, 25.6 |
| 120        | 26.7, 26.1, 25.8, 26.3, 27.2, 27.9, 28.3, 27.4 |
| 140        | 28.0, 28.4, 27.0, 28.8, 29.8, 29.4, 28.7, 29.3 |
| 160        | 32.7, 32.1, 31.9, 33.0, 33.5, 33.7, 34.0, 32.5 |
| 180        | 43.1, 41.7, 42.4, 42.1, 43.5, 43.8, 44.2, 43.6 |
| 200        | 54.2, 43.7, 53.1, 53.8, 55.6, 55.9, 54.7, 54.5 |

**13.37** 参见练习 13.36 中的数据。假定有另外一个受控变量,即每个速度下的前 4 块布料做浓度为 0.2 的保护层处理,而后 4 块布料做浓度为 0.4 的保护层



处理。用  $x_1$  表示机器的转速,  $x_2$  表示保护层处理时的浓度, 拟合模型:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \epsilon$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_1 x_2 + \beta_5 x_1^2 x_2 + \epsilon$$

**13.38(工程)** 一家洗衣粉制造公司希望在投放市场前对某种新产品进行检验。所关心的是洗衣机中洗衣粉泡沫的高度与放入的洗衣粉量的多少以及洗涤中的搅动程度的关系。对于标准型号的洗衣机的洗涤筒, 放满水以后, 随机指定不同的搅动程度(按分钟测量)和洗衣粉的用量, 并对洗衣机进行测试。数据见下表。

- 画出数据的图, 并建议一个模型。
- 正态性假定看上去能满足吗?
- 拟合一个适当的模型。
- 用残差图来探索是否可能有违反模型假定的现象。

| 高度, $y$ | 搅动程度, $x_1$ | 用量, $x_2$ |
|---------|-------------|-----------|
| 28.1    | 1           | 6         |
| 32.3    | 1           | 7         |
| 34.8    | 1           | 8         |
| 38.2    | 1           | 9         |
| 43.5    | 1           | 10        |
| 60.3    | 2           | 6         |
| 63.7    | 2           | 7         |
| 65.4    | 2           | 8         |
| 69.2    | 2           | 9         |
| 72.9    | 2           | 10        |
| 88.2    | 3           | 6         |
| 89.3    | 3           | 7         |
| 94.1    | 3           | 8         |
| 95.7    | 3           | 9         |
| 100.6   | 3           | 10        |

**13.39** 参见练习 13.38。下面的模型是否更合适? 为什么?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \beta_6 x_1 x_2^2 + \beta_7 x_1^2 x_2 + \beta_8 x_1^2 x_2^2 + \epsilon$$

**13.40** 参见练习 13.38 中的数据。

- 我们可以检验下列模型的拟合不足吗?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

$$+ \beta_6 x_1 x_2^2 + \beta_7 x_1^2 x_2 + \beta_8 x_1^2 x_2^2 + \epsilon$$

b. 对于这组样本数据, 写出相应的完全模型。注意, 如果在一个或多个设计点上有重复,  $SS_{\text{lack}}$  的自由度等于完全模型中参数的个数和(a)中模型的参数个数的差。

13.41 参见例 13.10。

a. 解释模型中的参数。

b. 拟合“完全”模型。

c. 从标准模型和有多余变量的模型给出结论。

13.42(化学) 在六个温度上考察了一种溶液的溶解性, 见下表:

| $y$ , 溶解性(按重量算) | $x$ , 温度( $^{\circ}\text{C}$ ) |
|-----------------|--------------------------------|
| 43, 45, 42      | 0                              |
| 32, 33, 37      | 25                             |
| 21, 28, 29      | 50                             |
| 15, 14, 9       | 75                             |
| 12, 10, 8       | 100                            |
| 7, 6, 2         | 125                            |

a. 画出这些数据的图, 拟合一个合适的模型。

b. 如果可能, 检验拟合不足, 用  $\alpha = 0.05$ 。

c. 考察残差, 得出结论。

13.43 参见练习 13.42。假设我们丢失了观测值 6, 8 和 14。

a. 拟合模型  $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$ 。

b. 用  $\alpha = 0.05$  检验拟合不足。

d. 再一次考察残差。

13.44 参见练习 13.37 中的数据。

a. 检验模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_1 x_2 + \beta_5 x_1^2 x_2 + \epsilon$$

的拟合不足。

b. 写出该试验的完全模型。

13.45 参见练习 13.32 中的数据。检验二次模型的拟合不足。

13.46(心理学) 一位心理学家要考察睡眠深度对于一个人进行简单算术计算的能力的影响。为此, 对候选人进行了筛选, 选出的人日常的睡眠模式很相近。

从这些人中,挑选了 20 个人。把每个选中的人随机地分到五个组之一,每组 4 个人。

组 1: 0 小时睡眠

组 2: 2 小时睡眠

组 3: 4 小时睡眠

组 4: 6 小时睡眠

组 5: 8 小时睡眠

然后,让所有人员在接下来的 24 小时内按日常的方式生活。

第二天吃过早饭以后,对每个人进行测试,记录下 10 分钟内答对的算术加法的个数。当天晚上,每个人所睡的时间决定于他或她所在的组。第二天早饭后,对每个人再进行一次测试,测试的题目不同但难度相同。

设响应变量为第一次答对的个数减去第二次答对的个数,数据如下。

| 组 | 响应, $y$        |
|---|----------------|
| 1 | 39, 33, 41, 40 |
| 2 | 25, 29, 34, 26 |
| 3 | 10, 18, 14, 17 |
| 4 | 4, 6, -1, 9    |
| 5 | -5, 0, -3, -8  |

a. 画出这些数据的图,并从该图给出一个建议模型。

b. 拟合所建议的模型。

c. 就模型假定进行考察,看有无违反模型假定的现象。

**13.47(工程)** 为确定一种金属的弯曲程度(WARPING)  $y$  和温度( $^{\circ}\text{C}$ , TEMP)之间的关系进行一项试验。所得样本数据如下表。注意,在每个温度上观察了三次。用下面的计算机输出完成(a)~(d)。

| 弯曲程度       | 温度( $^{\circ}\text{C}$ ) |
|------------|--------------------------|
| 10, 13, 12 | 15                       |
| 14, 12, 11 | 20                       |
| 14, 12, 16 | 25                       |
| 18, 19, 22 | 30                       |
| 25, 21, 20 | 35                       |
| 23, 25, 26 | 40                       |
| 30, 31, 34 | 45                       |
| 35, 33, 38 | 50                       |

- a. 画出数据图, 并确定线性或二次模型哪个看起来更合适?
- b. 如果拟合一个线性模型, 试给出预测方程, 并把预测方程的曲线加在(a)中画出的散点图上。
- c. 如果拟合一个二次模型, 试给出预测方程, 并把预测方程的曲线加在(a)中画出的散点图上。哪个模型看起来好些, 是线性还是二次?
- d. 用线性和二次预测方程分别预测 27℃ 时的弯曲程度。

## SAS OUTPUT FOR EXERCISE 13.47

## DATA LISTING

| OBS | WARPING | TEMP | TEMP2 |
|-----|---------|------|-------|
| 1   | 10      | 15   | 225   |
| 2   | 13      | 15   | 225   |
| 3   | 12      | 15   | 225   |
| 4   | 14      | 20   | 400   |
| 5   | 12      | 20   | 400   |
| 6   | 11      | 20   | 400   |
| 7   | 14      | 25   | 625   |
| 8   | 12      | 25   | 625   |
| 9   | 16      | 25   | 625   |
| 10  | 18      | 30   | 900   |
| 11  | 19      | 30   | 900   |
| 12  | 22      | 30   | 900   |
| 13  | 25      | 35   | 1225  |
| 14  | 21      | 35   | 1225  |
| 15  | 20      | 35   | 1225  |
| 16  | 23      | 40   | 1600  |
| 17  | 25      | 40   | 1600  |
| 18  | 26      | 40   | 1600  |
| 19  | 30      | 45   | 2025  |
| 20  | 31      | 45   | 2025  |
| 21  | 34      | 45   | 2025  |
| 22  | 35      | 50   | 2500  |
| 23  | 33      | 50   | 2500  |
| 24  | 38      | 50   | 2500  |

## LINEAR REGRESSION OF WARPING ON TEMPERATURE

Dependent Variable: AMOUNT OF WARPING

# Analysis of Variance

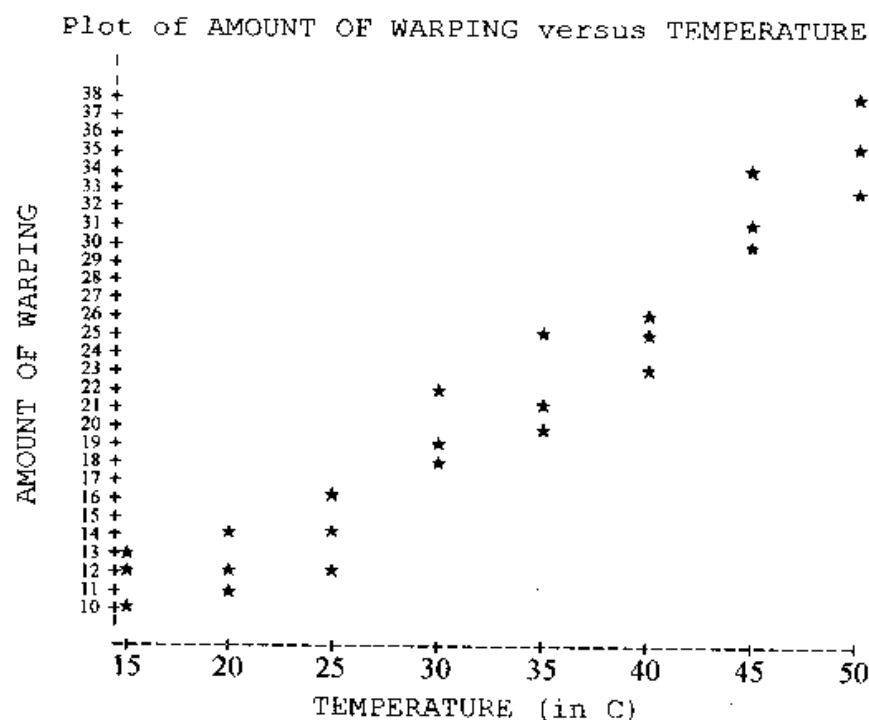
| Source   | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|----------|----|----------------|-------------|---------|--------|
| Model    | 1  | 1571.62698     | 1571.62698  | 265.546 | 0.0001 |
| Error    | 22 | 130.20635      | 5.91847     |         |        |
| C Total  | 23 | 1701.83333     |             |         |        |
| Root MSE |    | 2.43279        | R-square    | 0.9235  |        |
| Dep Mean |    | 21.41667       | Adj R-sq    | 0.9200  |        |
| C.V.     |    | 11.35933       |             |         |        |

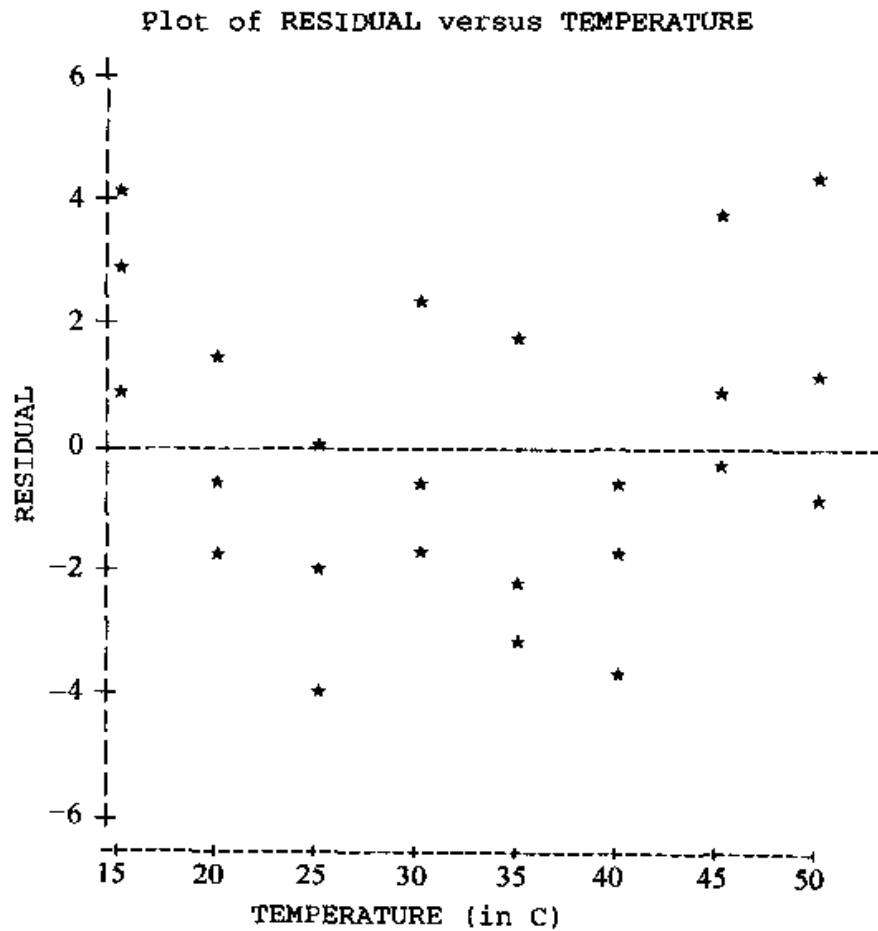
# Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|----------------------------|---------|
| INTERCEP | 1  | -1.539683          | 1.49370995     | -1.031                     | 0.3138  |
| TEMP     | 1  | 0.706349           | 0.04334604     | 16.296                     | 0.0001  |

| Variable | DF | Label              |
|----------|----|--------------------|
| INTERCEP | 1  | Intercept          |
| LOG_DOSE | 1  | TEMPERATURE (in C) |

Durbin-Watson D 0.908  
(For Number of Obs. ) 24  
1st Order Autocorrelation 0.474





#### QUADRATIC REGRESSION OF WARPING ON TEMPERATURE

Dependent Variable: AMOUNT OF WARPING

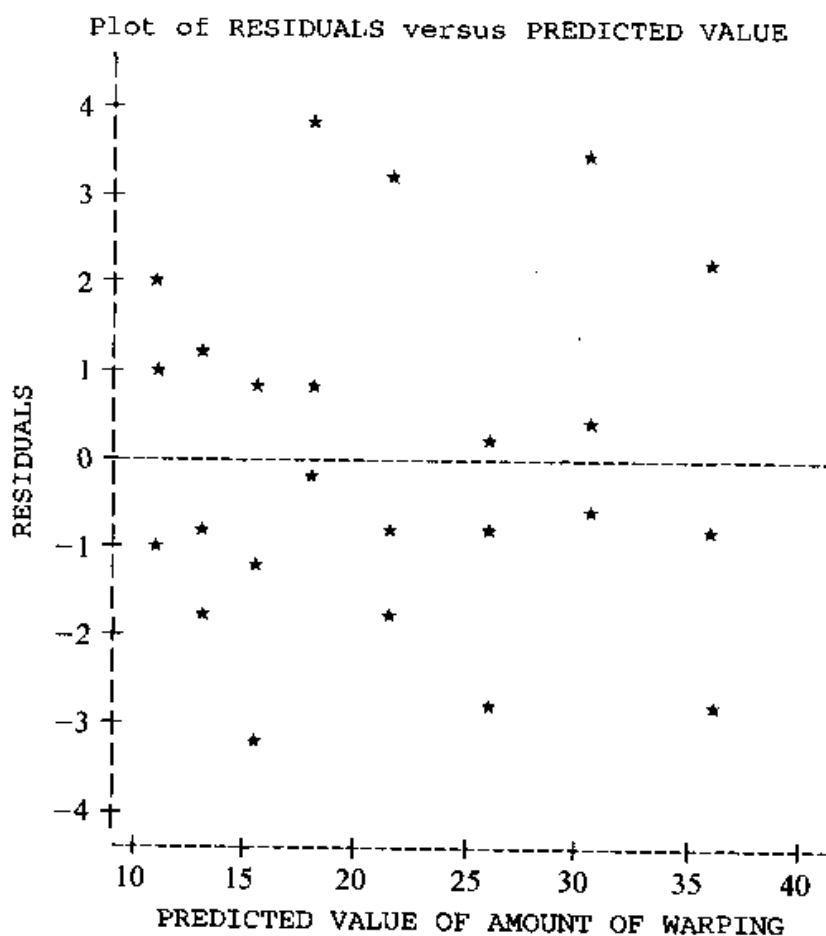
Analysis of Variance

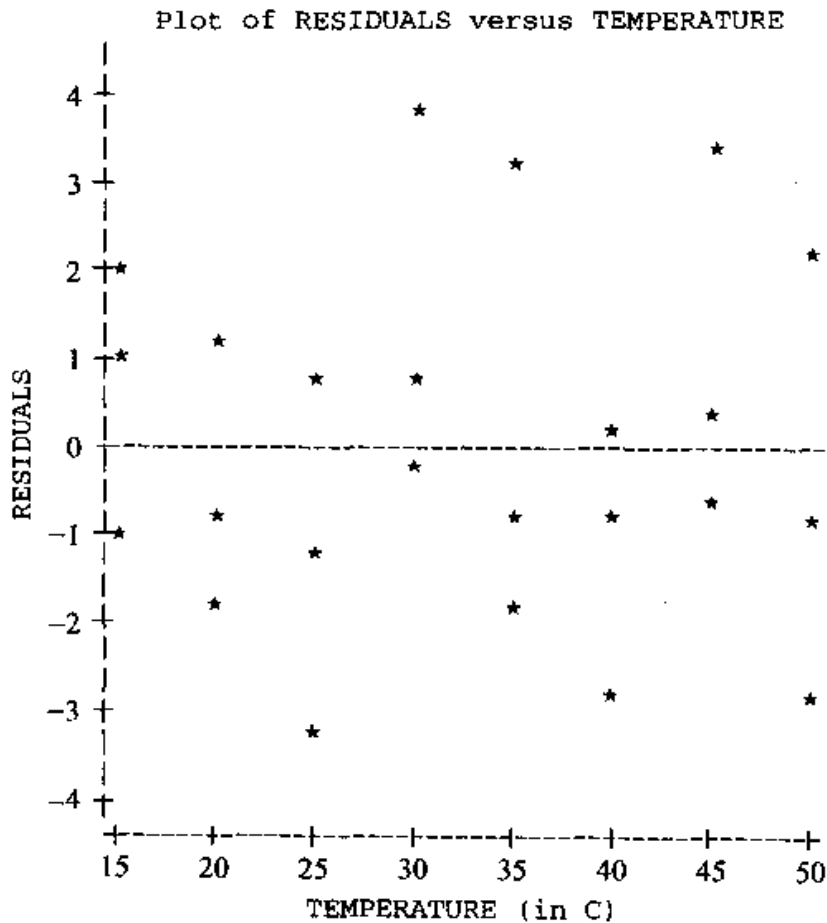
| Source   | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|----------|----|----------------|-------------|---------|--------|
| Model    | 2  | 1613.92063     | 806.96032   | 192.761 | 0.0001 |
| Error    | 21 | 87.91270       | 4.18632     |         |        |
| C Total  | 23 | 1701.83333     |             |         |        |
| Root MSE |    | 2.04605        | R - square  | 0.9483  |        |
| Dep Mean |    | 21.41667       | Adj R - sq  | 0.9434  |        |
| C. V.    |    | 9.55354        |             |         |        |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H <sub>0</sub> :<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|---|---------|
| INTERCEP | 1  | 9.178571           | 3.59852022     | 2.551                                   | 0.0186  |
| TEMP     | 1  | -0.046825          | 0.23974742     | -0.195                                  | 0.8470  |
| TEMP2    | 1  | 0.011587           | 0.00364553     | 3.178                                   | 0.0045  |

| Variable                  |    |                     |
|---------------------------|----|---------------------|
| Variable                  | DF | Label               |
| INTERCEP                  | 1  | Intercept           |
| TEMP                      | 1  | TEMPERATURE (in C)  |
| TEMP2                     | 1  | TEMPERATURE SQUARED |
| Durbin-Watson D           |    | 1.451               |
| (For Number of Obs.)      |    | 24                  |
| 1st Order Autocorrelation |    | 0.240               |





**13.48(商业)** 多元回归的应用之一是用来制定表现的衡量标准。换句话说,回归方程可以用来预测当某些条件得到满足时,一个人应该表现有多好。在某个这种类型的研究中,为了解一个推销员的销售量,从国内的四个部分(东北、东南、中西和西部)抽取了 50 个销售区域组成的样本,收集了有关这些区域的数据。这些数据包括个体销售表现,以及几个潜在的预测变量。这些变量如下。

$y$  = 销售区域表现,按区域推销员的合计销售量计算,

$x_1$  = 在公司工作的时间(月),

$x_2$  = 广告,或公司的努力(在区内所花的广告费用),

$x_3$  = 市场份额(前 4 年市场份额的加权平均),

$x_4$  = 国家中地区的指示变量(1 = 东北, 0 = 其他),

$x_5$  = 国家中地区的指示变量(1 = 东南, 0 = 其他),

$x_6$  = 国家中地区的指示变量(1 = 中西, 0 = 其他),

$x_7$  = 指示变量(1 = 男性推销员, 0 = 女性推销员)。

这些数据用 Minitab 分析的结果如下:



MTB> DESCRIBE C1-C10

|    | N  | MEAN   | MEDIAN | TRMEAN | STDEV  | SEMEAN |
|----|----|--------|--------|--------|--------|--------|
| Y  | 50 | 3335   | 3396   | 3277   | 1579   | 223    |
| X1 | 50 | 96.62  | 85.00  | 93.86  | 66.33  | 9.38   |
| X2 | 50 | 5002   | 5069   | 4915   | 2370   | 335    |
| X3 | 50 | 7.335  | 7.305  | 7.297  | 1.668  | 0.236  |
| C5 | 50 | 2.460  | 2.000  | 2.455  | 1.129  | 0.160  |
| X4 | 50 | 0.8200 | 1.0000 | 0.8636 | 0.3881 | 0.0549 |
| X5 | 50 | 0.2600 | 0.0000 | 0.2273 | 0.4431 | 0.0627 |
| X6 | 50 | 0.2600 | 0.0000 | 0.2273 | 0.4431 | 0.0627 |
| X7 | 50 | 0.2400 | 0.0000 | 0.2045 | 0.4314 | 0.0610 |

|    | MIN    | MAX    | Q1     | Q3     |
|----|--------|--------|--------|--------|
| Y  | 131    | 7205   | 2033   | 4367   |
| X1 | 000    | 237.00 | 40.00  | 144.25 |
| X2 | 222    | 10832  | 3038   | 6564   |
| X3 | 4.131  | 11.205 | 5.987  | 8.569  |
| C5 | 1.000  | 4.000  | 1.000  | 3.250  |
| X4 | 0.0000 | 1.0000 | 1.0000 | 1.0000 |
| X5 | 0.0000 | 1.0000 | 0.0000 | 1.0000 |
| X6 | 0.0000 | 1.0000 | 0.0000 | 1.0000 |
| X7 | 0.0000 | 1.0000 | 0.0000 | 0.2500 |

MTB> REGRESS 'Y' ON 7 'X1' 'X2' 'X3' 'X4' 'X5' 'X6' 'X7'

The regression equation is

$$Y = 16.4 - 0.000546X1 + 0.667X2 + 0.0302X3 - 0.116X4 - 0.041X5 - 33.3X6 - 33.6X7$$

| Predictor | Coef       | Stdev     | t-ratio   |
|-----------|------------|-----------|-----------|
| Constant  | 16.3944    | 0.2931    | 55.94     |
| X1        | -0.0005463 | 0.0007607 | -0.72     |
| X2        | 0.666689   | 0.000047  | 14315.675 |
| X3        | 0.03024    | 0.06467   | 0.47      |
| X4        | -0.1163    | 0.1128    | -1.03     |
| X5        | -0.0412    | 0.1201    | -0.34     |
| X6        | -33.3155   | 0.1204    | -276.81   |
| X7        | -33.6118   | 0.1185    | -283.70   |

S = 0.2864 R-Sq = 100.0 % R-Sq(adj) = 100.0 %

## Analysis of Variance

| SOURCE     | DF | SS        | MS       |
|------------|----|-----------|----------|
| Regression | 7  | 122189056 | 17455576 |
| Error      | 42 | 3         | 0        |
| Total      | 49 | 122189056 |          |

| SOURCE | DF | SEQ SS   |
|--------|----|----------|
| X1     | 1  | 33243924 |
| X2     | 1  | 88931584 |
| X3     | 1  | 1        |
| X4     | 1  | 80       |
| X5     | 1  | 4972     |
| X6     | 1  | 1880     |
| X7     | 1  | 6602     |

| Obs. | X1  | Y       | Fit-Stdev. | Fit  | Residual | St. Resid. |
|------|-----|---------|------------|------|----------|------------|
| 1    | 62  | 3407.00 | 3406.54    | 0.09 | 0.46     | 1.68       |
| 2    | 70  | 131.00  | 131.17     | 0.14 | -0.17    | -0.69      |
| 3    | 186 | 4650.00 | 4649.93    | 0.09 | 0.07     | 0.27       |
| 4    | 13  | 1971.00 | 1970.91    | 0.11 | 0.09     | 0.35       |
| 5    | 20  | 4168.00 | 4167.94    | 0.11 | 0.06     | 0.21       |
| 6    | 0   | 3047.00 | 3047.28    | 0.10 | -0.28    | -1.03      |
| 7    | 31  | 1196.00 | 1195.91    | 0.13 | 0.09     | 0.36       |
| 8    | 61  | 2415.00 | 2414.91    | 0.10 | 0.09     | 0.34       |
| 9    | 48  | 1987.00 | 1987.12    | 0.09 | -0.12    | -0.46      |
| 10   | 101 | 2214.00 | 2213.84    | 0.10 | 0.16     | 0.61       |
| 11   | 145 | 4333.00 | 4333.14    | 0.27 | -0.14    | -1.36X     |
| 12   | 200 | 6253.00 | 6253.08    | 0.12 | -0.08    | -0.29      |
| 13   | 81  | 1714.00 | 1713.87    | 0.12 | 0.13     | 0.49       |
| 14   | 124 | 5146.00 | 5146.01    | 0.09 | -0.01    | -0.04      |
| 15   | 24  | 3469.00 | 3469.27    | 0.11 | -0.27    | -1.04      |
| 16   | 216 | 4124.00 | 4123.60    | 0.11 | 0.40     | 1.53       |
| 17   | 232 | 3851.00 | 3851.17    | 0.14 | -0.17    | -0.69      |
| 18   | 109 | 2172.00 | 2171.83    | 0.10 | 0.17     | 0.64       |
| 19   | 75  | 1743.00 | 1743.25    | 0.12 | -0.25    | -0.97      |
| 20   | 5   | 2269.00 | 2268.93    | 0.11 | 0.07     | 0.27       |
| 21   | 12  | 3429.00 | 3429.24    | 0.10 | -0.24    | -0.88      |
| 22   | 90  | 1986.00 | 1985.83    | 0.10 | 0.17     | 0.64       |
| 23   | 209 | 3623.00 | 3623.21    | 0.12 | -0.21    | -0.82      |
| 24   | 167 | 5429.00 | 5429.16    | 0.15 | -0.16    | -0.64      |
| 25   | 170 | 4511.00 | 4511.22    | 0.10 | -0.22    | -0.81      |
| 26   | 42  | 1478.00 | 1477.94    | 0.12 | 0.06     | 0.24       |
| 27   | 167 | 3385.00 | 3385.22    | 0.11 | -0.22    | -0.84      |
| 28   | 98  | 1660.00 | 1660.84    | 0.11 | -0.84    | -3.16R     |

| Obs. | X1  | Y       | Fit Stdev. | Fit  | Residual | St. Resid. |
|------|-----|---------|------------|------|----------|------------|
| 29   | 144 | 1212.00 | 1211.69    | 0.12 | 0.31     | 1.20       |
| 30   | 78  | 4592.00 | 4592.00    | 0.09 | 0.00     | 0.00       |
| 31   | 116 | 2876.00 | 2875.85    | 0.09 | 0.15     | 0.55       |
| 32   | 89  | 4349.00 | 4349.02    | 0.09 | -0.02    | -0.06      |
| 33   | 37  | 2096.00 | 2095.80    | 0.09 | 0.20     | 0.72       |
| 34   | 34  | 5308.00 | 5308.07    | 0.11 | -0.07    | -0.26      |
| 35   | 165 | 5731.00 | 5730.01    | 0.10 | 0.99     | 3.70R      |
| 36   | 41  | 1121.00 | 1120.84    | 0.11 | 0.16     | 0.62       |
| 37   | 80  | 2356.00 | 2355.91    | 0.12 | 0.09     | 0.34       |
| 38   | 140 | 7205.00 | 7204.80    | 0.13 | 0.20     | 0.79       |
| 39   | 48  | 3562.00 | 3561.96    | 0.13 | 0.04     | 0.15       |
| 40   | 203 | 4133.00 | 4132.94    | 0.11 | 0.06     | 0.23       |
| 41   | 71  | 2049.00 | 2049.12    | 0.09 | -0.12    | -0.42      |
| 42   | 13  | 2512.00 | 2511.90    | 0.09 | 0.10     | 0.36       |
| 43   | 144 | 3722.00 | 3721.89    | 0.09 | 0.11     | 0.40       |
| 44   | 11  | 2806.00 | 2805.74    | 0.13 | 0.26     | 1.01       |
| 45   | 34  | 1477.00 | 1477.10    | 0.09 | -0.10    | -0.37      |
| 46   | 94  | 4040.00 | 4039.96    | 0.08 | 0.04     | 0.16       |
| 47   | 237 | 6633.00 | 6633.36    | 0.12 | -0.36    | -1.37      |
| 48   | 115 | 3203.00 | 3203.04    | 0.12 | -0.04    | -0.17      |
| 49   | 66  | 4423.00 | 4423.27    | 0.10 | -0.27    | -1.00      |
| 50   | 113 | 5563.00 | 5563.38    | 0.10 | -0.38    | -1.40      |

R denotes an obs. with a large st. resid.

X denotes an obs. whose X value gives it large influence.

做一个检验,看一看西部的推销员是否比东北部的推销员的工作更有成效(其他情况都一样)。给出零假设和备择假设,检验统计量的计算值和临界值,以及你的结论,取  $\alpha = 0.05$ 。

**13.49** 参见练习 13.48。当在销售区域内的广告费用增加 1,000 美元时,推销员的销售量的平均增量估计为多少?

**13.50** 参见练习 13.48。做一个检验,看男性推销员的平均销售量是否比女性推销员高 200 个单位(其他情况都一样),取  $\alpha = 0.05$ 。

**13.51** 参见练习 13.48。某公司的销售部经理特别关心地区的不同对于男、女推销员的影响不一致。

a. 建议一个回归模型,使得其中能够包含国内四个地区和推销员性别的交互作用。

b. 解释该模型中各个新  $\beta$  的意义。

**13.52(商业)** 在对房产价格关于九个自变量进行的回归分析中,用了 22 处房产的随机样本的数据。该项研究中的变量如下:

PRICE = 售价(美元)

BATHS = 浴室的数量(化妆间 = 1/2 个浴室)

BEDA = 卧室数量的虚拟变量(1 = 2 个卧室, 0 = 其他)

BEDB = 卧室数量的虚拟变量(1 = 3 个卧室, 0 = 其他)

BEDC = 卧室数量的虚拟变量(1 = 4 个卧室, 0 = 其他)

CARA = 车库类型的虚拟变量(1 = 无车库, 0 = 其他)

CARB = 车库类型的虚拟变量(1 = 单车库, 0 = 其他)

AGE = 房龄(年)

LOT = 占地面积(平方码)

DOM = 在市场上待售的时间(天)

在这项研究中,房子有两个、三个、四个或五个卧室,可能无车库,或单车车库,或双车车库。如上,我们用两个虚拟变量来对车库的这三种情况进行编码。

用 Minitab 对这些数据进行了分析,结果见下面。用完全模型(九个自变量),估计下列两方的平均差别:

- 无车库的房产和有一个单车车库的房产。
- 有一个单车车库的房产和有一个双车车库的房产。
- 无车库的房产和有一个双车车库的房产。

#### MINITAB OUTPUT FOR EXERCISE 13.52

##### DATA DISPLAY

| Row | PRICE | BATHS | BEDA | BEDB | BEDC | CARA | CARB | AGE | LOT  | DOM |
|-----|-------|-------|------|------|------|------|------|-----|------|-----|
| 1   | 25750 | 1.0   | 1    | 0    | 0    | 1    | 0    | 23  | 9680 | 164 |
| 2   | 37950 | 1.0   | 0    | 1    | 0    | 0    | 1    | 7   | 1889 | 67  |
| 3   | 46450 | 2.5   | 0    | 1    | 0    | 0    | 0    | 9   | 1941 | 315 |
| 4   | 46550 | 2.5   | 0    | 0    | 1    | 1    | 0    | 18  | 1813 | 61  |
| 5   | 47950 | 1.5   | 1    | 0    | 0    | 0    | 1    | 2   | 1583 | 234 |
| 6   | 49950 | 1.5   | 0    | 1    | 0    | 0    | 0    | 10  | 1533 | 116 |
| 7   | 52450 | 2.5   | 0    | 0    | 1    | 0    | 0    | 4   | 1667 | 162 |
| 8   | 54050 | 2.0   | 0    | 1    | 0    | 0    | 1    | 5   | 3450 | 80  |
| 9   | 54850 | 2.0   | 0    | 1    | 0    | 0    | 0    | 5   | 1733 | 63  |
| 10  | 52050 | 2.5   | 0    | 1    | 0    | 0    | 0    | 5   | 3727 | 102 |

| Row | PRICE  | BATHS | BEDA | BEDB | BEDC | CARA | CARB | AGE | LOT  | DOM |
|-----|--------|-------|------|------|------|------|------|-----|------|-----|
| 11  | 54392  | 2.5   | 0    | 1    | 0    | 0    | 0    | 7   | 1725 | 48  |
| 12  | 53450  | 2.5   | 0    | 1    | 0    | 0    | 0    | 3   | 2811 | 423 |
| 13  | 59510  | 2.5   | 0    | 1    | 0    | 0    | 1    | 11  | 5653 | 130 |
| 14  | 60102  | 2.5   | 0    | 1    | 0    | 0    | 0    | 7   | 2333 | 159 |
| 15  | 63850  | 2.5   | 0    | 0    | 1    | 0    | 0    | 6   | 2022 | 314 |
| 16  | 62050  | 2.5   | 0    | 0    | 0    | 0    | 0    | 5   | 2166 | 135 |
| 17  | 69450  | 2.0   | 0    | 1    | 0    | 0    | 0    | 15  | 1836 | 71  |
| 18  | 82304  | 2.5   | 0    | 0    | 1    | 0    | 0    | 8   | 5066 | 338 |
| 19  | 81850  | 2.0   | 0    | 1    | 0    | 0    | 0    | 0   | 2333 | 147 |
| 20  | 70050  | 2.0   | 0    | 1    | 0    | 0    | 0    | 4   | 2904 | 115 |
| 21  | 112450 | 2.5   | 0    | 0    | 1    | 0    | 0    | 1   | 2930 | 11  |
| 22  | 127050 | 3.0   | 0    | 0    | 1    | 0    | 0    | 9   | 2904 | 36  |

Descriptive Statistics: PRICE, BATHS, BEDA, BEDB, BEDC, CARA, CARB, AGE, LOT, DO

| Variable | N  | Mean   | Median | TrMean | StDev  | SE Mean |
|----------|----|--------|--------|--------|--------|---------|
| PRICE    | 22 | 62023  | 54621  | 60585  | 22749  | 4850    |
| BATHS    | 22 | 2.182  | 2.500  | 2.200  | 0.524  | 0.112   |
| BEDA     | 22 | 0.0909 | 0.0000 | 0.0500 | 0.2942 | 0.0627  |
| BEDB     | 22 | 0.591  | 1.000  | 0.600  | 0.503  | 0.107   |
| BEDC     | 22 | 0.2727 | 0.0000 | 0.2500 | 0.4558 | 0.0972  |
| CARA     | 22 | 0.0909 | 0.0000 | 0.0500 | 0.2942 | 0.0627  |
| CARB     | 22 | 0.1818 | 0.0000 | 0.1500 | 0.3948 | 0.0842  |
| AGE      | 22 | 7.45   | 6.50   | 7.05   | 5.48   | 1.17    |
| LOT      | 22 | 2895   | 2250   | 2624   | 1868   | 398     |
| DOM      | 22 | 149.6  | 123.0  | 142.9  | 109.8  | 23.4    |

| Variable | Minimum | Maximum | Q1     | Q3     |
|----------|---------|---------|--------|--------|
| PRICE    | 25750   | 127050  | 49450  | 69600  |
| BATHS    | 1.000   | 3.000   | 2.000  | 2.500  |
| BEDA     | 0.0000  | 1.0000  | 0.0000 | 0.0000 |
| BEDB     | 0.000   | 1.000   | 0.000  | 1.000  |
| BEDC     | 0.0000  | 1.0000  | 0.0000 | 1.0000 |
| CARA     | 0.0000  | 1.0000  | 0.0000 | 0.0000 |
| CARB     | 0.0000  | 1.0000  | 0.0000 | 0.0000 |
| AGE      | 0.00    | 23.00   | 4.00   | 9.25   |
| LOT      | 1533    | 9680    | 1793   | 3060   |
| DOM      | 11.0    | 423.0   | 66.0   | 181.5  |

Regression Analysis: PRICE versus BATHS, BEDA, BEDB, BEDC, CARA, CARB, AGE, LOT, DOM

The regression equation is

$$\text{PRICE} = 39617 + 11686 \text{ BATHS} + 15128 \text{ BEDA} + 2477 \text{ BEDB} + 26114 \text{ BEDC} - 44023 \text{ CARA} \\ - 12375 \text{ CARB} - 506 \text{ AGE} + 3.40 \text{ LOT} - 86.0 \text{ DOM}$$

| Predictor | Coef    | SE Coef | T      | P     |
|-----------|---------|---------|--------|-------|
| Constant  | 39617   | 30942   | 1.28   | 0.225 |
| BATHS     | 11686   | 10428   | 1.12   | 0.284 |
| BEDA      | 15128   | 26254   | 0.58   | 0.575 |
| BEDB      | 2477    | 17783   | 0.14   | 0.892 |
| BEDC      | 26114   | 18118   | 1.44   | 0.175 |
| CARA      | - 44023 | 22775   | - 1.93 | 0.077 |
| CARB      | - 12375 | 10759   | - 1.15 | 0.272 |
| AGE       | - 506   | 1111    | - 0.46 | 0.657 |
| LOT       | 3.399   | 2.504   | 1.36   | 0.200 |
| DOM       | - 86.05 | 35.72   | - 2.41 | 0.033 |

S = 16531      R - Sq = 69.8 %      R - Sq(adj) = 47.2 %

#### Analysis of Variance

| Source         | DF | SS          | MS        | F    | P     |
|----------------|----|-------------|-----------|------|-------|
| Regression     | 9  | 7588195915  | 843132879 | 3.09 | 0.036 |
| Residual Error | 12 | 3279393939  | 273282828 |      |       |
| Total          | 21 | 10867589854 |           |      |       |

| Source | DF | Seq SS     |
|--------|----|------------|
| BATHS  | 1  | 3352323167 |
| BEDA   | 1  | 24291496   |
| BEDB   | 1  | 668205893  |
| BEDC   | 1  | 261898228  |
| CARA   | 1  | 1261090278 |
| CARB   | 1  | 133807628  |
| AGE    | 1  | 5848       |
| LOT    | 1  | 300736097  |
| DOM    | 1  | 1585837280 |

#### Unusual Observations

| Obs | BATHS | PRICE | Fit   | SE Fit | Residual | St Resid |
|-----|-------|-------|-------|--------|----------|----------|
| 7   | 2.50  | 52450 | 84651 | 7506   | - 32201  | - 2.19R  |
| 16  | 2.50  | 62050 | 62050 | 16531  | - 0      | * X      |

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

Regression Analysis: PRICE versus BATHS, BEDA, BEDC, CARA, CARB, LOT, DOM

The regression equation is

$$\text{PRICE} = 39091 + 11712 \text{ BATHS} + 14183 \text{ BEDA} + 24531 \text{ BEDC} - 50962 \text{ CARA} - 12121 \text{ CARB} \\ + 3.08 \text{ LOT} - 84.8 \text{ DOM}$$

| Predictor | Coef    | SE Coef | T      | P     |
|-----------|---------|---------|--------|-------|
| Constant  | 39091   | 21445   | 1.82   | 0.090 |
| BATHS     | 11712   | 9531    | 1.23   | 0.239 |
| BEDA      | 14183   | 16759   | 0.85   | 0.412 |
| BEDC      | 24531   | 9021    | 2.72   | 0.017 |
| CARA      | - 50962 | 15878   | - 3.21 | 0.006 |
| CARB      | - 12121 | 10010   | - 1.21 | 0.246 |
| LOT       | 3.082   | 2.231   | 1.38   | 0.189 |
| DOM       | - 84.81 | 33.24   | - 2.55 | 0.023 |

S = 15443      R - Sq = 69.3 %      R - Sq(adj) = 53.9 %

Analysis of Variance

| Source         | DF | SS          | MS         | F    | P     |
|----------------|----|-------------|------------|------|-------|
| Regression     | 7  | 7528777484  | 1075539641 | 4.51 | 0.008 |
| Residual Error | 14 | 3338812370  | 238486598  |      |       |
| Total          | 21 | 10867589854 |            |      |       |

| Source | DF | Seq SS     |
|--------|----|------------|
| BATHS  | 1  | 3352323167 |
| BEDA   | 1  | 24291496   |
| BEDC   | 1  | 929454598  |
| CARA   | 1  | 1261501483 |
| CARB   | 1  | 133856231  |
| LOT    | 1  | 274447991  |
| DOM    | 1  | 1552902518 |

## Unusual Observations

| Obs | BATHS | PRICE | Fit   | SE Fit | Residual | St Resid |
|-----|-------|-------|-------|--------|----------|----------|
| 7   | 2.50  | 52450 | 84299 | 6973   | - 31849  | - 2.31R  |

R denotes an observation with a large standardized residual

Regression Analysis: PRICE versus BATHS, BEDC, CARA, CARB, LOT, DOM

The regression equation is

$$\text{PRICE} = 44534 + 8336 \text{ BATHS} + 24649 \text{ BEDC} - 47007 \text{ CARA} - 10588 \text{ CARB} + 3.54 \text{ LOT} - 76.7 \text{ DOM}$$

| Predictor | Coef    | SE Coef | T      | P     |
|-----------|---------|---------|--------|-------|
| Constant  | 44534   | 20264   | 2.20   | 0.044 |
| BATHS     | 8336    | 8574    | 0.97   | 0.346 |
| BEDC      | 24649   | 8934    | 2.76   | 0.015 |
| CARA      | - 47007 | 15030   | - 3.13 | 0.007 |
| CARB      | - 10588 | 9751    | - 1.09 | 0.295 |
| LOT       | 3.539   | 2.144   | 1.65   | 0.120 |
| DOM       | - 76.67 | 31.51   | - 2.43 | 0.028 |

S = 15296      R - Sq = 67.7 %      R - Sq(adj) = 54.8 %

## Analysis of Variance

| Source         | DF | SS          | MS         | F    | P     |
|----------------|----|-------------|------------|------|-------|
| Regression     | 6  | 7357974702  | 1226329117 | 5.24 | 0.004 |
| Residual Error | 15 | 3509615152  | 233974343  |      |       |
| Total          | 21 | 10867589854 |            |      |       |

| Source | DF | Seq SS     |
|--------|----|------------|
| BATHS  | 1  | 3352323167 |
| BEDC   | 1  | 883193335  |
| CARA   | 1  | 1307168140 |
| CARB   | 1  | 111305152  |
| LOT    | 1  | 318872879  |
| DOM    | 1  | 1385112029 |



#### Unusual Observations

| Obs | BATHS | PRICE | FIT   | SE Fit | Residual | St Resid |
|-----|-------|-------|-------|--------|----------|----------|
| 7   | 2.50  | 52450 | 83502 | 6843   | - 31052  | - 2.27R  |

R denotes an observation with a large standardized residual

Regression Analysis: PRICE versus BEDC, CARA, CARB, LOT, DOM

The regression equation is

$$\text{PRICE} = 62606 + 28939 \text{ BEDC} - 52659 \text{ CARA} - 14153 \text{ CARB} + 3.52 \text{ LOT} - 75.6 \text{ DOM}$$

| Predictor | Coef    | SE Coef | T      | P     |
|-----------|---------|---------|--------|-------|
| Constant  | 62606   | 8056    | 7.77   | 0.000 |
| BEDC      | 28939   | 7755    | 3.73   | 0.002 |
| CARA      | - 52659 | 13837   | - 3.81 | 0.002 |
| CARB      | - 14153 | 9019    | - 1.57 | 0.136 |
| LOT       | 3.523   | 2.140   | 1.65   | 0.119 |
| DOM       | - 75.64 | 31.44   | - 2.41 | 0.029 |

S = 15270

R - Sq = 65.7 %

R - Sq(adj) = 54.9 %

#### Analysis of Variance

| Source         | DF | SS          | MS         | F    | P     |
|----------------|----|-------------|------------|------|-------|
| Regression     | 5  | 7136792581  | 1427358516 | 6.12 | 0.002 |
| Residual Error | 16 | 3730797273  | 233174830  |      |       |
| Total          | 21 | 10867589854 |            |      |       |

| Source | DF | Seq SS     |
|--------|----|------------|
| BEDC   | 1  | 2901187555 |
| CARA   | 1  | 2274636373 |
| CARB   | 1  | 292810426  |
| LOT    | 1  | 318495206  |
| DOM    | 1  | 1349663021 |

#### Unusual Observations

| Obs | BEDC | PRICE  | Fit   | SE Fit | Residual | St Resid |
|-----|------|--------|-------|--------|----------|----------|
| 1   | 0.00 | 25750  | 31641 | 13849  | - 5891   | - 0.92 X |
| 4   | 1.00 | 46550  | 40659 | 13849  | 5891     | 0.92 X   |
| 7   | 1.00 | 52450  | 85164 | 6614   | - 32714  | - 2.38 R |
| 22  | 1.00 | 127050 | 99052 | 7948   | 27998    | 2.15 R   |

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

Regression Analysis: PRICE versus BEDC, CARA, CARB, LOT, DOM

The regression equation is

$$\text{PRICE} = 59313 + 31921 \text{ BEDC} - 48742 \text{ CARA} + 3.02 \text{ LOT} - 69.0 \text{ DOM}$$

| Predictor | Coef   | SE Coef | T     | P     |
|-----------|--------|---------|-------|-------|
| Constant  | 59313  | 8105    | 7.32  | 0.000 |
| BEDC      | 31921  | 7836    | 4.07  | 0.001 |
| CARA      | -48742 | 14183   | -3.44 | 0.003 |
| LOT       | 3.025  | 2.206   | 1.37  | 0.188 |
| DOM       | -69.00 | 32.46   | -2.13 | 0.049 |

S = 15913      R - Sq = 60.4 %      R - Sq(adj) = 51.1 %

Analysis of Variance

| Source         | DF | SS          | MS         | F    | P     |
|----------------|----|-------------|------------|------|-------|
| Regression     | 4  | 6562672180  | 1640668045 | 6.48 | 0.002 |
| Residual Error | 17 | 4304917674  | 253230451  |      |       |
| Total          | 21 | 10867589854 |            |      |       |

| Source | DF | Seq SS     |
|--------|----|------------|
| BEDC   | 1  | 2901187555 |
| CARA   | 1  | 2274636373 |
| LOT    | 1  | 242949284  |
| DOM    | 1  | 1143898968 |

Unusual Observations

| Obs | BEDC | PRICE  | Fit   | SE Fit | Residual | St Resid |   |
|-----|------|--------|-------|--------|----------|----------|---|
| 1   | 0.00 | 25750  | 28533 | 14284  | -2783    | -0.40    | X |
| 4   | 1.00 | 46550  | 43767 | 14284  | 2783     | 0.40     | X |
| 7   | 1.00 | 52450  | 85098 | 6893   | -32648   | -2.28    | R |
| 22  | 1.00 | 127050 | 97533 | 8221   | 29517    | 2.17     | R |

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

Regression Analysis: PRICE versus BEDC, CARA, DOM

The regression equation is

$$\text{PRICE} = 66338 + 30129 \text{ BEDC} - 38457 \text{ CARA} - 60.4 \text{ DOM}$$

| Predictor | Coef   | SE Coef | T     | P     |
|-----------|--------|---------|-------|-------|
| Constant  | 66338  | 6433    | 10.31 | 0.000 |
| BEDC      | 30129  | 7913    | 3.81  | 0.001 |
| CARA      | -38457 | 12329   | -3.12 | 0.006 |
| DOM       | -60.41 | 32.62   | -1.85 | 0.081 |

S = 16298

R - Sq = 56.0 %

R - Sq(adj) = 48.7 %

Analysis of Variance

| Source         | DF | SS          | MS         | F    | P     |
|----------------|----|-------------|------------|------|-------|
| Regression     | 3  | 6086432104  | 2028810701 | 7.64 | 0.002 |
| Residual Error | 18 | 4781157750  | 265619875  |      |       |
| Total          | 21 | 10867589854 |            |      |       |

| Source | DF | Seq SS     |
|--------|----|------------|
| BEDC   | 1  | 2901187555 |
| CARA   | 1  | 2274636373 |
| DOM    | 1  | 910608176  |

Unusual Observations

| Obs | BEDC | PRICE  | Fit   | SE Fit | Residual | St Resid |   |
|-----|------|--------|-------|--------|----------|----------|---|
| 1   | 0.00 | 25750  | 17975 | 12322  | 7775     | 0.73     | X |
| 4   | 1.00 | 46550  | 54325 | 12322  | -7775    | -0.73    | X |
| 7   | 1.00 | 52450  | 86682 | 6960   | -34232   | -2.32    | R |
| 22  | 1.00 | 127050 | 94293 | 8065   | 32757    | 2.31     | R |

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

Regression Analysis: PRICE versus BEDC, CARA

The regression equation is

$$\text{PRICE} = 57231 + 29518 \text{ BEDC} - 35840 \text{ CARA}$$

| Predictor | Coef   | SE Coef | T     | P     |
|-----------|--------|---------|-------|-------|
| Constant  | 57231  | 4403    | 13.00 | 0.000 |
| BEDC      | 29518  | 8396    | 3.52  | 0.002 |
| CARA      | -35840 | 13006   | -2.76 | 0.013 |

S = 17308      R - Sq = 47.6 %      R - Sq(adj) = 42.1 %

#### Analysis of Variance

| Source         | DF | SS          | MS         | F    | P     |
|----------------|----|-------------|------------|------|-------|
| Regression     | 2  | 5175823928  | 2587911964 | 8.64 | 0.002 |
| Residual Error | 19 | 5691765926  | 299566628  |      |       |
| Total          | 21 | 10867589854 |            |      |       |

| Source | DF | Seq SS     |
|--------|----|------------|
| BEDC   | 1  | 2901187555 |
| CARA   | 1  | 2274636373 |

#### Unusual Observations

| Obs | BEDC | PRICE  | Fit   | SE Fit | Residual | St Resid |
|-----|------|--------|-------|--------|----------|----------|
| 1   | 0.00 | 25750  | 21391 | 12939  | 4359     | 0.38 X   |
| 4   | 1.00 | 46550  | 50909 | 12939  | -4359    | -0.38 X  |
| 7   | 1.00 | 52450  | 86749 | 7391   | -34299   | -2.19 R  |
| 22  | 1.00 | 127050 | 86749 | 7391   | 40301    | 2.58 R   |

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

#### Regression Analysis: PRICE versus BEDC

The regression equation is

$$\text{PRICE} = 54991 + 25785 \text{ BEDC}$$

| Predictor | Coef  | SE Coef | T     | P     |
|-----------|-------|---------|-------|-------|
| Constant  | 54991 | 4989    | 11.02 | 0.000 |
| BEDC      | 25785 | 9554    | 2.70  | 0.014 |

S = 19958      R - Sq = 26.7 %      R - Sq(adj) = 23.0 %

#### Analysis of Variance

| Source         | DF | SS          | MS         | F    | P     |
|----------------|----|-------------|------------|------|-------|
| Regression     | 1  | 2901187555  | 2901187555 | 7.28 | 0.014 |
| Residual Error | 20 | 7966402299  | 398320115  |      |       |
| Total          | 21 | 10867589854 |            |      |       |

## Unusual Observations

| Obs | BEDC | PRICE  | Fit   | SE Fit | Residual | St Resid |
|-----|------|--------|-------|--------|----------|----------|
| 22  | 1.00 | 127050 | 80776 | 8148   | 46274    | 2.54R    |

R denotes an observation with a large standardized residual

**13.53** 参见练习 13.52。用完全回归模型做一个检验,以确定房龄每长一年房产价格的下降幅度是否小于 2,500 美元。给出零假设和  $p$  值,并得出你的结论,取  $\alpha = 0.05$ 。

**13.54** 参见练习 13.52。假定我们希望修改上面的九个变量的模型,使得能够反映出价格与房龄的关系因卧室的个数而异。

a. 写出一个这样的模型的公式。

b. 模型参数的什么组合表现了五个卧室、单车车库的房产与两卧室、双车车库的房产的差别?

**13.55** 参见练习 13.52。在自变量为从原始的九个变量中选取出来的模型中,你选择哪个模型为“最优”模型?为什么选择该模型?

**13.56** 参见练习 13.52。在另一项研究中,也使用了相同的那 22 处房产的数据,并把 PRICE 关于 LIST,即房产的报价(以千美元为单位)进行了回归分析。

Best Subsets Regression: PRICE versus BATHS, BEDA, BEDB, BEDC, CARA, CARB, AGE, LOT, DOM

Response is PRICE

| Vars | R - Sq | R - Sq(adj) | C-p  | S     | B |   |   |   |   |   |   |   |   |   |
|------|--------|-------------|------|-------|---|---|---|---|---|---|---|---|---|---|
|      |        |             |      |       | A | B | B | B | C | C |   |   |   |   |
|      |        |             |      |       | T | E | E | E | A | A | A | L | D |   |
|      |        |             |      |       | H | D | D | D | R | R | G | O | O |   |
|      |        |             |      |       | S | S | A | B | C | A | B | E | T | M |
| 1    | 30.8   | 27.4        | 9.5  | 19385 | X |   |   |   |   |   |   |   |   |   |
| 1    | 26.7   | 23.0        | 11.2 | 19958 |   |   |   |   | X |   |   |   |   |   |
| 2    | 47.6   | 42.1        | 4.8  | 17308 |   |   |   |   | X | X |   |   |   |   |
| 2    | 39.4   | 33.1        | 8.1  | 18612 |   |   |   |   | X |   |   | X |   |   |
| 3    | 56.0   | 48.7        | 3.5  | 16298 |   |   |   |   | X | X |   |   |   | X |
| 3    | 51.0   | 42.8        | 5.5  | 17200 | X |   |   |   | X | X |   |   |   |   |

|   |      |      |      |       |   |   |   |   |   |   |
|---|------|------|------|-------|---|---|---|---|---|---|
| 4 | 60.4 | 51.1 | 3.8  | 15913 |   | X | X |   | X | X |
| 4 | 60.2 | 50.8 | 3.8  | 15950 | X | X | X |   |   | X |
| 5 | 65.7 | 54.9 | 3.7  | 15270 |   | X | X | X | X | X |
| 5 | 65.2 | 54.3 | 3.9  | 15382 | X | X | X |   | X | X |
| 6 | 67.7 | 54.8 | 4.8  | 15296 | X | X | X | X | X | X |
| 6 | 66.5 | 53.1 | 5.3  | 15576 |   | X | X | X | X | X |
| 7 | 69.3 | 53.9 | 6.2  | 15443 | X | X | X | X | X | X |
| 7 | 68.6 | 52.9 | 6.5  | 15611 | X | X | X | X | X | X |
| 8 | 69.8 | 51.2 | 8.0  | 15896 | X | X | X | X | X | X |
| 8 | 69.3 | 50.4 | 8.2  | 16019 | X | X | X | X | X | X |
| 9 | 69.8 | 47.2 | 10.0 | 16531 | X | X | X | X | X | X |

## Data Display

| Row | PRICE  | LIST   |
|-----|--------|--------|
| 1   | 25750  | 29900  |
| 2   | 37950  | 39900  |
| 3   | 46450  | 44900  |
| 4   | 46550  | 47500  |
| 5   | 47950  | 49900  |
| 6   | 49950  | 49900  |
| 7   | 52450  | 53000  |
| 8   | 54050  | 54900  |
| 9   | 54850  | 54900  |
| 10  | 52050  | 55900  |
| 11  | 54392  | 55900  |
| 12  | 53450  | 56000  |
| 13  | 59510  | 62000  |
| 14  | 60102  | 62500  |
| 15  | 63850  | 63900  |
| 16  | 62050  | 66900  |
| 17  | 69450  | 72500  |
| 18  | 82304  | 82254  |
| 19  | 81850  | 82900  |
| 20  | 70050  | 99900  |
| 21  | 112450 | 117000 |
| 22  | 127050 | 139000 |

## Descriptive Statistics: PRICE, LTST

| Variable | N | Mean | Median | TrMean | StDev | SE Mean |
|----------|---|------|--------|--------|-------|---------|
|----------|---|------|--------|--------|-------|---------|

|       |    |       |       |       |       |      |
|-------|----|-------|-------|-------|-------|------|
| PRICE | 22 | 62023 | 54621 | 60585 | 22749 | 4850 |
| LIST  | 22 | 65521 | 55950 | 63628 | 25551 | 5447 |

| Variable | Minimum | Maximum | Q1    | Q3    |
|----------|---------|---------|-------|-------|
| PRICE    | 25750   | 127050  | 49450 | 69600 |
| LIST     | 29900   | 139000  | 49900 | 74939 |

Regression Analysis: PRICE versus LIST

The regression equation is

$$\text{PRICE} = 5406 + 0.864 \text{ LIST}$$

| Predictor | Coef    | SE Coef | T     | P     |
|-----------|---------|---------|-------|-------|
| Constant  | 5406    | 3363    | 1.61  | 0.124 |
| LIST      | 0.86411 | 0.04797 | 18.01 | 0.000 |

S = 5616      R - Sq = 94.2 %      R - Sq(adj) = 93.9 %

Analysis of Variance

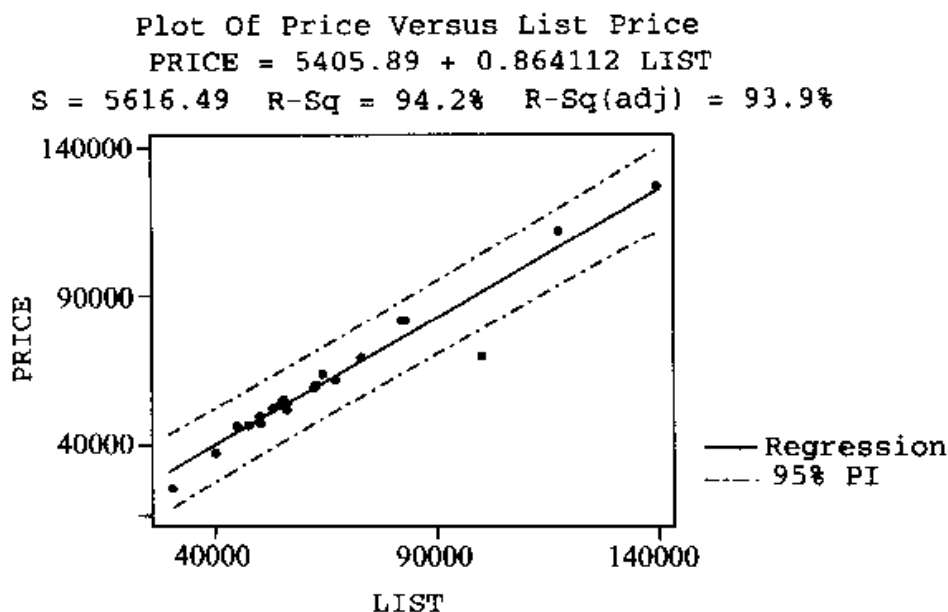
| Source         | DF | SS          | MS          | F      | P     |
|----------------|----|-------------|-------------|--------|-------|
| Regression     | 1  | 10236690015 | 10236690015 | 324.51 | 0.000 |
| Residual Error | 20 | 630899838   | 31544992    |        |       |
| Total          | 21 | 10867589854 |             |        |       |

Unusual Observations

| Obs | LIST   | PRICE  | Fit    | SE Fit | Residual | St Resid |
|-----|--------|--------|--------|--------|----------|----------|
| 20  | 99900  | 70050  | 91731  | 2038   | - 21681  | - 4.14 R |
| 22  | 139000 | 127050 | 125518 | 3723   | 1532     | 0.36 X   |

R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.



a. 用这个回归的结果, 预测报价为 70,000 美元的房产的售价。

b. 你的预测低了 3,000 美元或更多的机会有多大?

**13.57(社会学)** 使用练习 13.52 中售价的数据, 考察售价与两个自变量, 即房间的数量(ROOMS)与平方英尺数(SQFT)之间的关系。用下列数据。

| 行           | 售价    | 房间数 | 平方英尺数 |
|-------------|-------|-----|-------|
| 1           | 25.75 | 5   | 986   |
| 2           | 37.95 | 5   | 998   |
| 3           | 46.45 | 7   | 1,690 |
| 4           | 46.55 | 8   | 1,829 |
| 5           | 47.95 | 6   | 1,186 |
| 6           | 49.95 | 6   | 1,734 |
| 7           | 52.45 | 7   | 1,684 |
| 8           | 54.05 | 7   | 1,846 |
| 9           | 54.85 | 7   | 1,690 |
| (continues) |       |     |       |
| 10          | 52.05 | 7   | 1,910 |
| 11          | 54.39 | 7   | 1,784 |
| 12          | 53.45 | 6   | 1,690 |
| 13          | 59.51 | 7   | 1,590 |
| 14          | 60.10 | 8   | 1,855 |
| 15          | 63.85 | 8   | 2,212 |
| 16          | 62.05 | 10  | 2,784 |



续表

| 行  | 售价     | 房间数 | 平方英尺数 |
|----|--------|-----|-------|
| 17 | 69.45  | 7   | 2,190 |
| 18 | 82.30  | 8   | 2,259 |
| 19 | 81.85  | 7   | 1,919 |
| 20 | 70.05  | 7   | 1,685 |
| 21 | 112.45 | 10  | 2,654 |
| 22 | 127.05 | 10  | 2,756 |

用下面的计算机输出结果回答(a)(b)和(c)

#### MULTIPLE REGRESSION ANALYSIS

Dependent Variable: PRICE      SELLING PRICE(1000 \$ )

#### Analysis of Variance

| Source   | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|----------|----|----------------|-------------|---------|--------|
| Model    | 2  | 6816.77693     | 3408.38847  | 15.987  | 0.0001 |
| Error    | 19 | 4050.68890     | 213.19415   |         |        |
| C Total  | 21 | 10867.46584    |             |         |        |
| Root MSE |    | 14.60117       | R-square    | 0.6273  |        |
| Dep Mean |    | 62.02273       | Adj R-sq    | 0.5880  |        |
| C. V.    |    | 23.54164       |             |         |        |

#### Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0:<br>Parameter = 0 | Prob> T |
|----------|----|--------------------|----------------|----------------------------|---------|
| INTERCEP | 1  | - 16.975979        | 18.94658431    | - 0.896                    | 0.3815  |
| ROOMS    | 1  | 4.336062           | 6.04912439     | 0.717                      | 0.4822  |
| SQFT     | 1  | 0.025511           | 0.01737891     | 1.468                      | 0.1585  |

| Variable | DF | Variable Label  |
|----------|----|-----------------|
| INTERCEP | 1  | Intercept       |
| ROOMS    | 1  | NUMBER OF ROOMS |
| SQFT     | 1  | SQUARE FEET     |

- a. 做一个检验,看两个变量 ROOMS(房间数)和 SQFT(平方英尺数)一起,是否含有 PRICE(售价)的信息,取  $\alpha = 0.05$ 。
- b. 做一个检验,看 ROOMS 的系数是否为 0,取  $\alpha = 0.05$ 。
- c. 做一个检验,看 SQFT 的系数是否为 0,取  $\alpha = 0.05$ 。

13.58 参见练习 13.57。

- a. 解释(a)中结果和(b)、(c)中结果表面上的不一致性。
- b. 如果把 ROOMS 从模型中拿掉, SQFT 的  $t$  值会怎么样?

13.59(医药) 为了确定在美国的医院中传染监视和控制措施是否减低了同院传染的发病率而进行了一项研究。从参加一项大型研究的 338 家医院中挑选了由 28 所医院组成的随机样本,并从这些医院得到了数据。数据集的每一行提供了一个医院的各个变量的信息。这些变量如下:

RISK = 输出变量,医院中同院传染的平均发生率的估计值(百分数)

STAY = 输入变量,所有患者住院的平均长度(天)

AGE = 输入变量,患者的平均年龄(岁)

RCR = 输入变量,培养物的数量与没有同院传染症状的患者的人数的比值(乘以 100)

SCHOOL = 虚拟变量,是否与医学院有从属关系,1 = 有,0 = 无

DV<sub>1</sub> = 虚拟变量,表示地区,1 = 东北,0 = 其他

DV<sub>2</sub> = 虚拟变量,表示地区,1 = 北中,0 = 其他

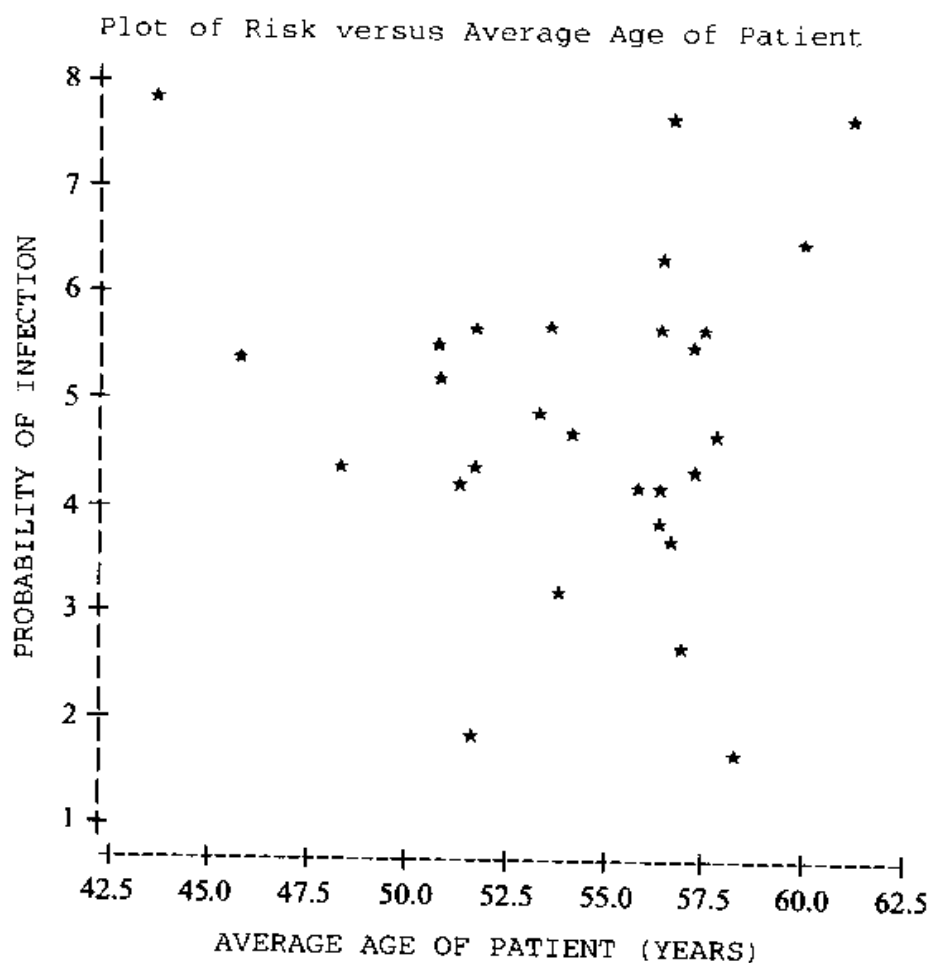
DV<sub>3</sub> = 虚拟变量,表示地区,1 = 南部,0 = 其他

(注意,国家有 4 个地理区域,即东北,北中,南部和西部。国家的这 4 个区域只需要三个虚拟变量。)用 SAS 分析这些数据,得到以下结果。

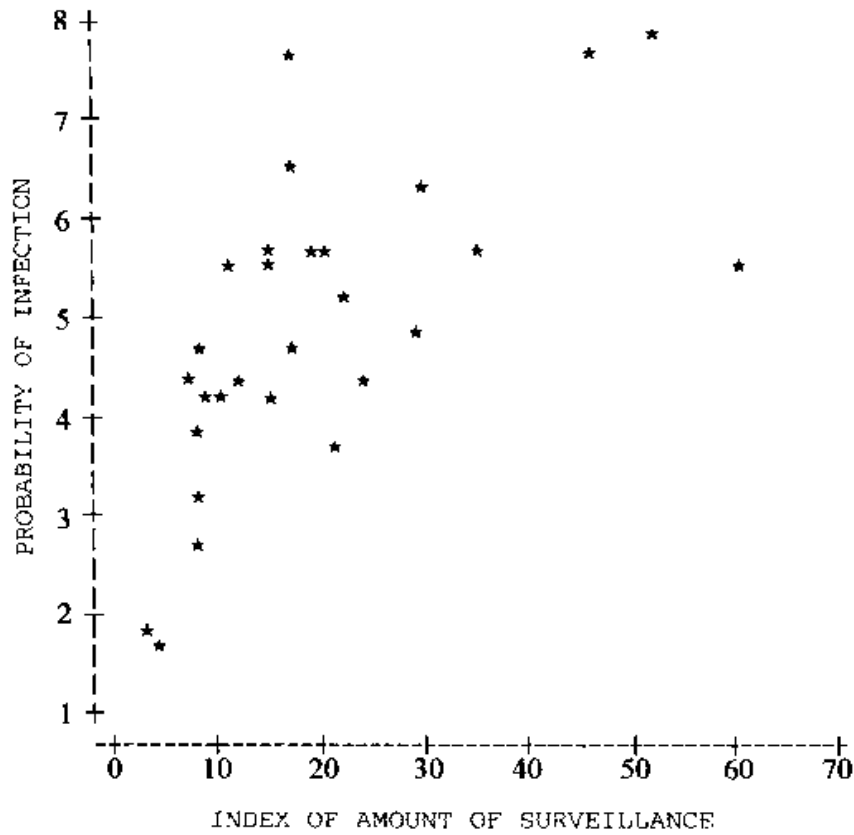
#### DATA LISTING

| Obs | RISK | STAY  | AGE  | INS  | SCHOOL | RC1 | RC2 | RC3 |
|-----|------|-------|------|------|--------|-----|-----|-----|
| 1   | 4.1  | 7.13  | 55.7 | 9.0  | 0      | 0   | 0   | 1   |
| 2   | 1.6  | 8.82  | 58.2 | 3.8  | 0      | 1   | 0   | 0   |
| 3   | 2.7  | 8.34  | 56.9 | 8.1  | 0      | 0   | 1   | 0   |
| 4   | 5.6  | 8.95  | 53.7 | 18.9 | 0      | 0   | 0   | 1   |
| 5   | 5.7  | 11.20 | 56.5 | 34.5 | 0      | 0   | 0   | 0   |
| 6   | 5.1  | 9.76  | 50.9 | 21.9 | 0      | 1   | 0   | 0   |
| 7   | 4.6  | 9.68  | 57.8 | 16.7 | 0      | 0   | 1   | 0   |
| 8   | 5.4  | 11.18 | 45.7 | 60.5 | 1      | 1   | 0   | 0   |
| 9   | 4.3  | 8.67  | 48.2 | 24.4 | 0      | 0   | 1   | 0   |
| 10  | 6.3  | 8.84  | 56.3 | 29.6 | 0      | 0   | 0   | 0   |
| 11  | 4.9  | 11.07 | 53.2 | 28.5 | 1      | 0   | 0   | 0   |

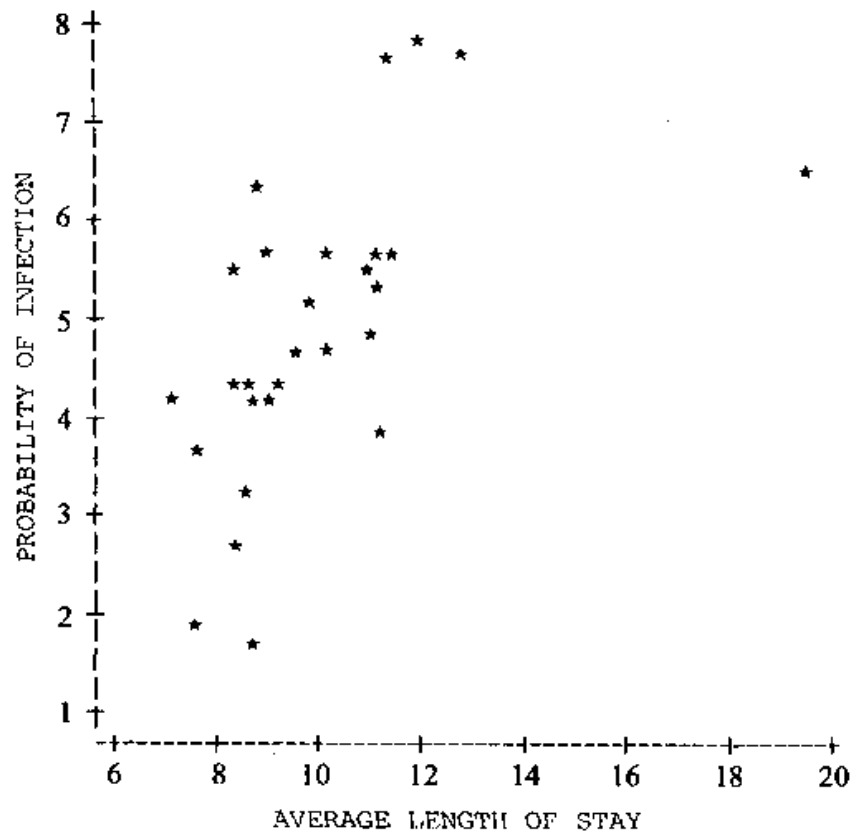
| Obs | RISK | STAY  | AGE  | INS  | SCHOOL | RC1 | RC2 | RC3 |
|-----|------|-------|------|------|--------|-----|-----|-----|
| 12  | 4.3  | 8.30  | 57.2 | 6.8  | 0      | 0   | 1   | 0   |
| 13  | 7.7  | 12.78 | 56.8 | 46.0 | 1      | 0   | 0   | 0   |
| 14  | 3.7  | 7.58  | 56.7 | 20.8 | 0      | 1   | 0   | 0   |
| 15  | 4.2  | 9.00  | 56.3 | 14.6 | 0      | 0   | 1   | 0   |
| 16  | 5.6  | 10.12 | 51.7 | 14.9 | 1      | 0   | 1   | 0   |
| 17  | 5.5  | 8.37  | 50.7 | 15.1 | 0      | 1   | 0   | 0   |
| 18  | 4.6  | 10.16 | 54.2 | 8.4  | 1      | 0   | 0   | 1   |
| 19  | 6.5  | 19.56 | 59.9 | 17.2 | 0      | 0   | 0   | 0   |
| 20  | 5.5  | 10.90 | 57.2 | 10.6 | 0      | 1   | 0   | 0   |
| 21  | 1.8  | 7.67  | 51.7 | 2.5  | 0      | 0   | 1   | 0   |
| 22  | 4.2  | 8.88  | 51.5 | 10.1 | 0      | 0   | 1   | 0   |
| 23  | 5.6  | 11.48 | 57.6 | 20.3 | 0      | 0   | 0   | 0   |
| 24  | 4.3  | 9.23  | 51.6 | 11.6 | 0      | 1   | 0   | 0   |
| 25  | 7.6  | 11.41 | 61.1 | 16.6 | 0      | 0   | 0   | 0   |
| 26  | 7.8  | 12.07 | 43.7 | 52.4 | 0      | 1   | 0   | 0   |
| 27  | 3.1  | 8.63  | 54.0 | 8.4  | 0      | 0   | 0   | 0   |
| 28  | 3.9  | 11.15 | 56.5 | 7.7  | 0      | 0   | 0   | 0   |



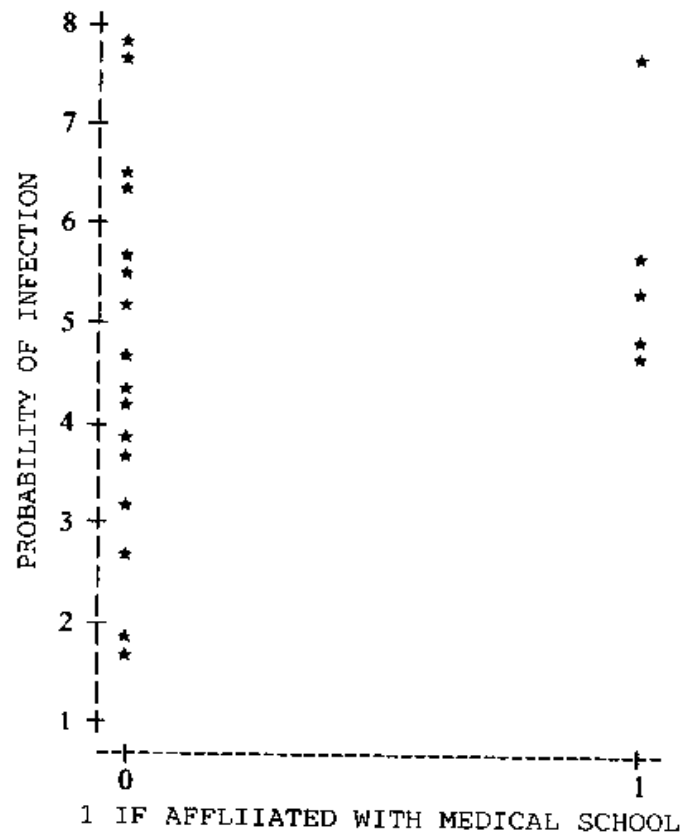
Plot of Risk versus Index of Surveillance



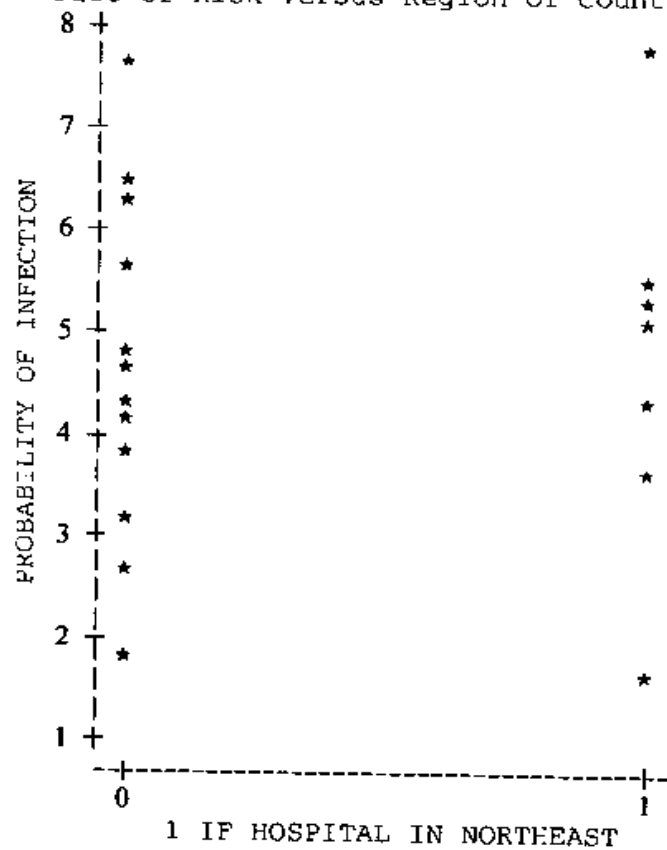
Plot of Risk versus Length of Stay in Hospital



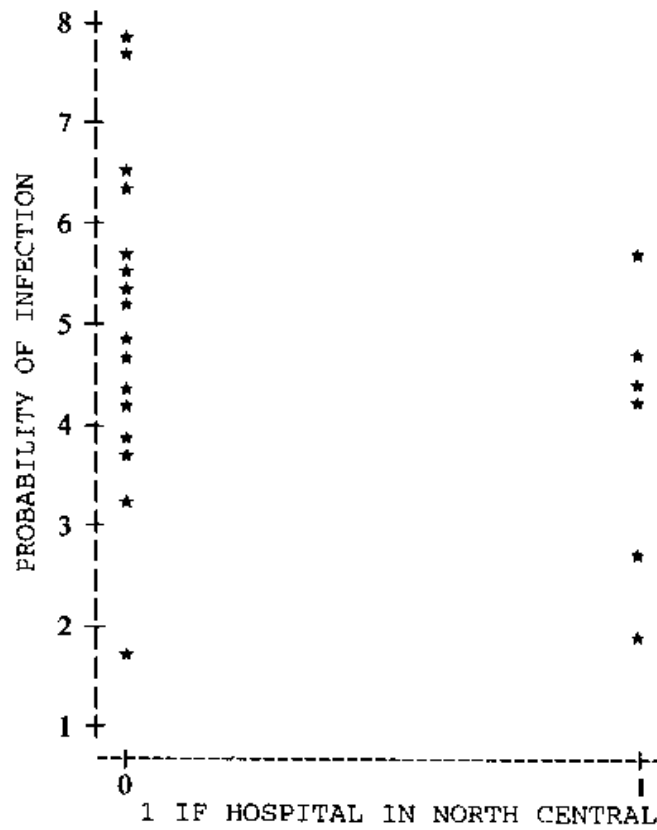
Plot of Risk versus Medical School Affiliation



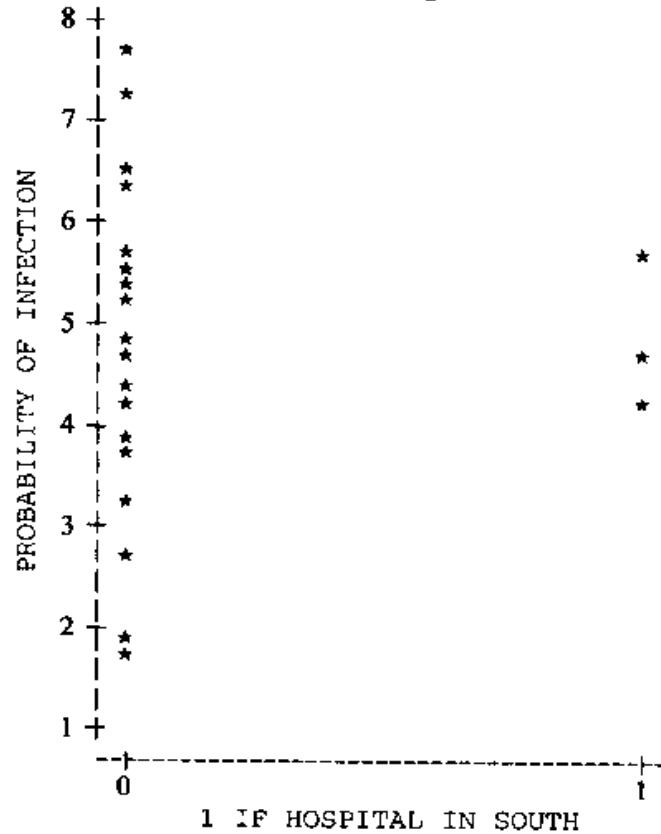
Plot of Risk versus Region of Country



Plot of Risk versus Region of Country



Plot of Risk versus Region of Country



Correlation Analysis of the Independent Variables:

7 VAR Variables: STAY AGE INS SCHOOL RC1 RC2 RC3

Simple Statistics

| Variable | N  | Mean    | Std Dev | Sum      | Minimum | Maximum |
|----------|----|---------|---------|----------|---------|---------|
| STAY     | 28 | 10.0332 | 2.3729  | 280.9300 | 7.1300  | 19.5600 |
| AGE      | 28 | 54.3393 | 4.0802  | 1522     | 43.7000 | 61.1000 |
| INS      | 28 | 19.2821 | 14.3288 | 539.9000 | 2.5000  | 60.5000 |
| SCHOOL   | 28 | 0.1786  | 0.3900  | 5.0000   | 0       | 1.0000  |
| RC1      | 28 | 0.2857  | 0.4600  | 8.0000   | 0       | 1.0000  |
| RC2      | 28 | 0.2857  | 0.4600  | 8.0000   | 0       | 1.0000  |
| RC3      | 28 | 0.1071  | 0.3150  | 3.0000   | 0       | 1.0000  |

Pearson Correlation Coefficients/Prob> |R| under Ho; Rho = 0/N = 28

|        | STAY     | AGE      | INS      | SCHOOL   | RC1      | RC2      | RC3      |
|--------|----------|----------|----------|----------|----------|----------|----------|
| STA    | 1.00000  | 0.18019  | 0.35014  | 0.20586  | -0.07993 | -0.32591 | -0.19127 |
|        | 0.0      | 0.3589   | 0.0678   | 0.2933   | 0.6860   | 0.0906   | 0.3296   |
| AGE    | 0.18019  | 1.00000  | -0.47243 | 0.23498  | -0.39490 | -0.06737 | 0.01678  |
|        | 0.3589   | 0.0      | 0.0111   | 0.2287   | 0.0375   | 0.7334   | 0.9325   |
| INS    | 0.35014  | -0.47243 | 1.00000  | 0.41016  | 0.23847  | -0.31552 | -0.17682 |
|        | 0.0678   | 0.0111   | 0.0      | 0.0302   | 0.2217   | 0.1019   | 0.3681   |
| SCHOOL | 0.20586  | -0.23498 | 0.41016  | 1.00000  | -0.08847 | -0.08847 | 0.13998  |
|        | 0.2933   | 0.2287   | 0.0302   | 0.0      | 0.6544   | 0.6544   | 0.4774   |
| RC1    | -0.07993 | -0.39490 | 0.23847  | -0.08847 | 1.00000  | -0.40000 | -0.21909 |
|        | 0.6860   | 0.0375   | 0.2217   | 0.6544   | 0.0      | 0.0349   | 0.2627   |
| RC2    | -0.32591 | -0.06737 | -0.31552 | -0.08847 | -0.40000 | 1.00000  | -0.21909 |
|        | 0.0906   | 0.7334   | 0.1019   | 0.6544   | 0.0349   | 0.0      | 0.2627   |
| RC3    | -0.19127 | 0.01678  | -0.17682 | 0.13998  | -0.21909 | -0.21909 | 1.00000  |
|        | 0.3296   | 0.9325   | 0.3681   | 0.4774   | 0.2627   | 0.2627   | 0.0      |

Backward Elimination Procedure for Dependent Variable RISK

Step 0 All Variables Entered R-square = 0.60724861 C(p) = 8.00000000

|            | DF | Sum of Squares | Mean Square | F    | Prob>F |
|------------|----|----------------|-------------|------|--------|
| Regression | 7  | 39.49805177    | 5.64257882  | 4.42 | 0.0041 |
| Error      | 20 | 25.54623394    | 1.27731170  |      |        |
| Total      | 27 | 65.04428571    |             |      |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F    | Prob>F |
|----------|--------------------|----------------|------------------------|------|--------|
| INTERCEP | -1.07800774        | 4.69134824     | 0.06744431             | 0.05 | 0.8206 |
| STAY     | 0.23613428         | 0.11569116     | 5.32126218             | 4.17 | 0.0547 |
| AGE      | 0.04359681         | 0.07810854     | 0.39793239             | 0.31 | 0.5829 |
| INS      | 0.06923673         | 0.02278287     | 11.79650358            | 9.24 | 0.0065 |
| SCHOOL   | -0.41516871        | 0.64822732     | 0.52395194             | 0.41 | 0.5291 |
| RC1      | -0.26955673        | 0.68941266     | 0.19527144             | 0.15 | 0.6999 |
| RC2      | -0.19268071        | 0.71943459     | 0.09162010             | 0.07 | 0.7916 |
| RC3      | 0.70243224         | 0.88962481     | 0.79632801             | 0.62 | 0.4390 |

Bounds on condition number: 2.315515, 94.11721

Step 1 Variable RC2 Removed R-square = 0.60584002 C(p) = 6.07172885

|            | DF | Sum of Squares | Mean Square | F    | Prob>F |
|------------|----|----------------|-------------|------|--------|
| Regression | 6  | 39.40643167    | 6.56773861  | 5.38 | 0.0017 |
| Error      | 21 | 25.63785404    | 1.22085019  |      |        |
| Total      | 27 | 65.04428571    |             |      |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | -1.81224950        | 3.72184905     | 0.28945486             | 0.24  | 0.6314 |
| STAY     | 0.24597088         | 0.10725430     | 6.42096620             | 5.26  | 0.0322 |
| AGE      | 0.05262498         | 0.06888511     | 0.71251762             | 0.58  | 0.4534 |
| INS      | 0.07154787         | 0.02061408     | 14.70713325            | 12.05 | 0.0023 |
| SCHOOL   | -0.42280540        | 0.63312506     | 0.54445805             | 0.45  | 0.5115 |
| RC1      | -0.15497958        | 0.52853481     | 0.10496975             | 0.09  | 0.7722 |
| RC3      | 0.83288104         | 0.72780215     | 1.59882767             | 1.31  | 0.2653 |

Bounds on condition number: 1.929521, 53.56369

#### QUADRATIC REGRESSION OF WARPING ON TEMPERATURE

173

Step 2 Variable RC1 Removed R-square = 0.60422621 C(p) = 4.15390906

|            | DF | Sum of Squares | Mean Square | F    | Prob>F |
|------------|----|----------------|-------------|------|--------|
| Regression | 5  | 39.30146193    | 7.86029239  | 6.72 | 0.0006 |
| Error      | 22 | 25.74282379    | 1.17012835  |      |        |
| Total      | 27 | 65.04428571    |             |      |        |



| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | -2.21637907        | 3.38468174     | 0.50174830             | 0.43  | 0.5194 |
| STAY     | 0.24760767         | 0.10486035     | 6.52437780             | 5.58  | 0.0275 |
| AGE      | 0.05898907         | 0.06400415     | 0.99394033             | 0.85  | 0.3667 |
| INS      | 0.07087867         | 0.02005725     | 14.61240661            | 12.49 | 0.0019 |
| SCHOOL   | -0.38736862        | 0.60843670     | 0.47429829             | 0.41  | 0.5309 |
| RC3      | 0.87192445         | 0.70049715     | 1.81291925             | 1.55  | 0.2263 |

Bounds on condition number: 1.905871, 36,65382

Step 3 Variable SCHOOL Removed R-square = 0.59693428 C(p) = 2.52523447

|            | DF | Sum of Squares | Mean Square | F    | Prob>F |
|------------|----|----------------|-------------|------|--------|
| Regression | 4  | 38.82716364    | 9.70679091  | 8.52 | 0.0002 |
| Error      | 23 | 26.21712207    | 1.13987487  |      |        |
| Total      | 27 | 65.04428571    |             |      |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | -2.30479519        | 3.33782686     | 0.54349337             | 0.48  | 0.4968 |
| STAY     | 0.23848508         | 0.10252510     | 6.16764346             | 5.41  | 0.0292 |
| AGE      | 0.06257589         | 0.06292612     | 1.12722159             | 0.99  | 0.3304 |
| INS      | 0.06713326         | 0.01892561     | 14.34276871            | 12.58 | 0.0017 |
| RC3      | 0.76072793         | 0.66954727     | 1.47147677             | 1.29  | 0.2676 |

Bounds on condition number: 1.741914, 23.03492

Step 4 Variable AGE Removed R-square = 0.57960421 C(p) = 1.40772979

|            | DF | Sum of Squares | Mean Square | F     | Prob>F |
|------------|----|----------------|-------------|-------|--------|
| Regression | 3  | 37.69994205    | 12.56664735 | 11.03 | 0.0001 |
| Error      | 24 | 27.34434367    | 1.13934765  |       |        |
| Total      | 27 | 65.04428571    |             |       |        |

| Variable                                       | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|--|--------------------|----------------|------------------------|-------|--------|
| INTERCEP                                       | 0.88480344         | 0.92355510     | 1.04574126             | 0.92  | 0.3476 |
| QUADRATIC REGRESSION OF WARPING ON TEMPERATURE |                    |                |                        |       |        |
| STAY   | 0.28060533         | 0.09334523     | 10.29588785            | 9.04  | 0.0061 |
| INS  | 0.05622030         | 0.01541554     | 15.15391450            | 13.30 | 0.0013 |

|     |            |            |            |      |        |
|-----|------------|------------|------------|------|--------|
| RC3 | 0.74723631 | 0.66925498 | 1.42032908 | 1.25 | 0.2753 |
|-----|------------|------------|------------|------|--------|

Bounds on condition number: 1.162616, 10.11556

Step 5 Variable RC3 Removed R-square = 0.55776787 C(p) = 0.51969728

|            | DF | Sum of Squares | Mean Square | F     | Prob>F |
|------------|----|----------------|-------------|-------|--------|
| Regression | 2  | 36.27961297    | 18.13980648 | 15.77 | 0.0001 |
| Error      | 25 | 28.76467275    | 1.15058691  |       |        |
| Total      | 27 | 65.04428571    |             |       |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | 1.15123509         | 0.89658440     | 1.89699030             | 1.65  | 0.2109 |
| STAY     | 0.26598212         | 0.09287658     | 9.43651980             | 8.20  | 0.0084 |
| INS      | 0.05416385         | 0.01538042     | 14.26927648            | 12.40 | 0.0017 |

Bounds on condition number: 1.139728, 4.558912

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination Procedure for Dependent Variable RISK

| Step | Variable Removed                                    | Number In | Partial R <sup>2</sup> | Model R <sup>2</sup> | C(p)   | F      | Prob>F |
|------|---|-----------|------------------------|----------------------|--------|--------|--------|
| 1    | RC2<br>1 IF HOSPITAL IN NORTH CENTRAL               | 6         | 0.0014                 | 0.6058               | 6.0717 | 0.0717 | 0.7916 |
| 2    | RC1<br>1 IF HOSPITAL IN NORTHEAST                   | 5         | 0.0016                 | 0.6042               | 4.1539 | 0.0860 | 0.7722 |
| 3    | SCHOOL<br>1 IF AFFILIATED WITH MEDICAL SCHOOL, 0 IF | 4         | 0.0073                 | 0.5969               | 2.5252 | 0.4053 | 0.5309 |
| 4    | AGE<br>AVERAGE AGE OF PATIENT (YEARS)               | 3         | 0.0173                 | 0.5796               | 1.4077 | 0.9889 | 0.3304 |
| 5    | RC3<br>1 IF HOSPITAL IN SOUTH                       | 2         | 0.0218                 | 0.5578               | 0.5197 | 1.2466 | 0.2753 |

这七个输入变量包含有输出变量 RISK 的信息吗? 给出你检验的  $p$  值。

基于完全回归模型(七个输入变量), 我们能否有至少 95% 的把握说, 在其他

完全一样的条件下,南方医院比西方医院的传染发生率至少高 5%?

**13.60** 参见练习 13.59。

a. 考虑下列两种说法:

在医院所在地区和医院是否属一个医学院之间有复共线性。

在医院所在地区和医院是否属一个医学院之间有交互作用。

这两种说法的区别是什么?要确定这两个说法是对的还是错的,各需要什么根据?在 SAS 的输出结果中,有这样的根据吗?如果有的话,那么你认为这些说法是对呢还是错呢?

b. 构造一个模型,使其能够反映所在地区和是否属某个医学院这两个变量之间的交互作用。对于该模型,东北的、且属于某个医学院的医院与一个西部的、不属于任何一个医学院的截距之间有什么差别?

**13.61** 参见练习 13.59。假定我们要从完全模型中去掉某些我们认为对输出变量贡献不大的变量,你最后选择的模型是什么?为什么你选择这个模型?

**13.62** 参见练习 13.59。预测东北部附属某个医学院的医院中的传染发生率,其中患者的平均住院时间为 10 天,平均年龄为 64 岁,日常培养比为 20%。这个预测是内插还是外推?你是怎么知道的呢?

**13.63(生物)** 三十名志愿者参加了如下的试验。这些人测量了自己的脉搏速率(这很容易做到:用一只手的拇指和食指按在脖子两边的两条动脉上即可得到)。然后,要他们掷一枚硬币。如果正面向上,他们就在原地跑 1 分钟。接下来,再请他们测量自己的脉搏,记录下前后两次脉搏速率的差,以及这些学生的其他特征。用这些变量作为自变量,做脉搏速率的回归分析,以解释脉搏速率的差。这些变量是:

PULSE = 前后两次脉率的差

RUN = 虚拟变量, 1 = 未做原地跑, 0 = 做了原地跑

SMOKE = 虚拟变量, 1 = 不吸烟, 0 = 吸烟

HEIGHT = 身高, 以英寸为单位

WEIGHT = 体重, 以磅为单位

PHYS1 = 虚拟变量, 1 = 大量体育锻炼, 0 = 其他

PHYS2 = 虚拟变量, 1 = 中量体育锻炼, 0 = 其他

a. 进行适当的检验,以确定所有这些自变量一起,是否显著地解释了脉率的变化,用  $\alpha = 0.01$  得出你的结论。

b. 此处是否有复共线性的问题?你的依据是什么?复共线性如何影响你用回归做出预测的能力?

c. 基于完全回归模型(六个自变量),计算那些参加了大量体育活动的人相对于那些没有多少体育活动的人的 PULSE 变量的平均增量的估计。我们是

否能有 95% 的把握认为这个平均增量的实际值大于 0?

LISTING OF DATA FOR EXERCISE 13.63

| Obs | PULSE | RUN | SMOKE | HEIGHT | WEIGHT | PHYS1 | PHYS2 |
|-----|-------|-----|-------|--------|--------|-------|-------|
| 1   | -29   | 0   | 1     | 66     | 140    | 0     | 1     |
| 2   | -17   | 0   | 1     | 72     | 145    | 0     | 1     |
| 3   | -14   | 0   | 0     | 73     | 160    | 1     | 0     |
| 4   | -22   | 0   | 0     | 73     | 190    | 0     | 0     |
| 5   | -21   | 0   | 1     | 69     | 155    | 0     | 1     |
| 6   | -25   | 0   | 1     | 73     | 165    | 0     | 0     |
| 7   | -5    | 0   | 1     | 72     | 150    | 1     | 0     |
| 8   | -9    | 0   | 1     | 74     | 190    | 0     | 1     |
| 9   | -18   | 0   | 1     | 72     | 195    | 0     | 1     |
| 10  | -23   | 0   | 1     | 71     | 138    | 0     | 1     |
| 11  | -14   | 0   | 0     | 74     | 160    | 0     | 0     |
| 12  | -21   | 0   | 1     | 72     | 155    | 0     | 1     |
| 13  | 8     | 0   | 0     | 70     | 153    | 1     | 0     |
| 14  | -13   | 0   | 1     | 67     | 145    | 0     | 1     |
| 15  | -21   | 0   | 1     | 71     | 170    | 1     | 0     |
| 16  | -1    | 0   | 1     | 72     | 175    | 1     | 0     |
| 17  | -16   | 0   | 0     | 69     | 175    | 0     | 1     |
| 18  | -15   | 1   | 1     | 68     | 145    | 0     | 0     |
| 19  | 4     | 1   | 0     | 75     | 190    | 0     | 1     |
| 20  | -3    | 1   | 1     | 72     | 180    | 1     | 0     |
| 21  | 2     | 1   | 0     | 67     | 140    | 0     | 1     |
| 22  | -5    | 1   | 1     | 70     | 150    | 0     | 1     |
| 23  | -1    | 1   | 1     | 73     | 155    | 0     | 1     |
| 24  | -5    | 1   | 1     | 74     | 148    | 1     | 0     |
| 25  | -6    | 1   | 0     | 68     | 150    | 0     | 1     |
| 26  | -6    | 1   | 0     | 73     | 155    | 0     | 1     |
| 27  | 8     | 1   | 0     | 66     | 130    | 0     | 1     |
| 28  | -1    | 1   | 1     | 69     | 160    | 0     | 1     |
| 29  | -5    | 1   | 1     | 66     | 135    | 1     | 0     |
| 30  | -3    | 1   | 1     | 75     | 160    | 1     | 0     |

Correlation Analysis

6 'VAR' Variables: RUN SMOKE HEIGHT WEIGHT PHYS1 PHYS2

Simple Statistics

| Variable | N  | Mean     | Std Dev | Sum     | Minimum  | Maximum  |
|----------|----|----------|---------|---------|----------|----------|
| RUN      | 30 | 0.4333   | 0.5040  | 13.0000 | 0        | 1.0000   |
| SMOKE    | 30 | 0.6667   | 0.4795  | 20.0000 | 0        | 1.0000   |
| HEIGHT   | 30 | 70.8667  | 2.7759  | 2126    | 66.0000  | 75.0000  |
| WEIGHT   | 30 | 158.6333 | 17.5391 | 4759    | 130.0000 | 195.0000 |
| PHYS1    | 30 | 0.3000   | 0.4661  | 9.0000  | 0        | 1.0000   |
| PHYS2    | 30 | 0.5667   | 0.5040  | 17.0000 | 0        | 1.0000   |

Pearson Correlation Coefficients/Prob>|R| under H0:Rho = 0/N = 30

|        | RUN      | SMOKE    | HEIGHT   | WEIGHT   | PHYS1    | PHYS2    |
|--------|----------|----------|----------|----------|----------|----------|
| RUN    | 1.00000  | -0.09513 | -0.12981 | -0.25056 | 0.01468  | 0.08597  |
|        | 0.0      | 0.6170   | 0.4942   | 0.1817   | 0.9386   | 0.6515   |
| SMOKE  | -0.09513 | 1.00000  | 0.01727  | -0.06834 | 0.15430  | -0.04757 |
|        | 0.6170   | 0.0      | 0.9278   | 0.7197   | 0.4156   | 0.8029   |
| HEIGHT | -0.12981 | 0.01727  | 1.00000  | 0.59885  | 0.19189  | -0.28919 |
|        | 0.4942   | 0.9278   | 0.0      | 0.0005   | 0.3097   | 0.1211   |
| WEIGHT | -0.25056 | -0.06834 | 0.59885  | 1.00000  | 0.01392  | -0.11221 |
|        | 0.1817   | 0.7197   | 0.0005   | 0.0      | 0.9418   | 0.5549   |
| PHYS1  | 0.01468  | 0.15430  | 0.19189  | 0.01392  | 1.00000  | -0.74863 |
|        | 0.9386   | 0.4156   | 0.3097   | 0.9418   | 0.0      | 0.0001   |
| PHYS2  | 0.08597  | -0.04757 | -0.28919 | -0.11221 | -0.74863 | 1.00000  |
|        | 0.6515   | 0.8029   | 0.1211   | 0.5549   | 0.0001   | 0.0      |

Backward Elimination Procedure for Dependent Variable PULSE

Step 0 All Variables Entered R-square = 0.62973045 C(p) = 7.00000000

|            | DF | Sum of Squares | Mean Square  | F    | Prob>F |
|------------|----|----------------|--------------|------|--------|
| Regression | 6  | 1850.58887109  | 308.43147852 | 6.52 | 0.0004 |
| Error      | 23 | 1088.11112891  | 47.30917952  |      |        |
| Total      | 29 | 2938.70000000  |              |      |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | -31.68830679       | 36.42360015    | 35.80780871            | 0.76  | 0.3933 |
| RUN      | 11.40166481        | 2.66171908     | 868.07553823           | 18.35 | 0.0003 |
| SMOKE    | -6.89029281        | 2.74454278     | 298.18154585           | 6.30  | 0.0195 |
| HEIGHT   | 0.13169561         | 0.60021947     | 2.27754970             | 0.05  | 0.8283 |
| WEIGHT   | 0.02303608         | 0.09440380     | 2.81697901             | 0.06  | 0.8094 |

|       |             |            |              |      |        |
|-------|-------------|------------|--------------|------|--------|
| PHYS1 | 13.43465041 | 4.25117641 | 472.47616161 | 9.99 | 0.0044 |
| PHYS2 | 7.80635269  | 3.97815470 | 182.17065424 | 3.85 | 0.0619 |

Bounds on condition number: 2.464274, 62.50691

Step 1 Variable HEIGHT Removed R-square = 0.62895543 C(p) = 5.04814181

|            | DF | Sum of Squares | Mean Square  | F    | Prob>F |
|------------|----|----------------|--------------|------|--------|
| Regression | 5  | 1848.31132139  | 369.66226428 | 8.14 | 0.0001 |
| Error      | 24 | 1090.38867861  | 45.43286161  |      |        |
| Total      | 29 | 2938.70000000  |              |      |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | -24.25519127       | 13.11118684    | 155.48750216           | 3.42  | 0.0767 |
| RUN      | 11.43076116        | 2.60516294     | 874.68284765           | 19.25 | 0.0002 |
| SMOKE    | -6.85327902        | 2.68448142     | 296.10525519           | 6.52  | 0.0175 |
| WEIGHT   | 0.03529782         | 0.07456145     | 10.18209732            | 0.22  | 0.6402 |
| PHYS1    | 13.44838310        | 4.16556957     | 473.54521380           | 10.42 | 0.0036 |
| PHYS2    | 7.65315557         | 3.83795325     | 180.65576063           | 3.98  | 0.0576 |

Bounds on condition number: 2.406131, 40.22006

Step 2 Variable WEIGHT Removed R-square = 0.62549060 C(p) = 3.26336637

|            | DF | Sum of Squares | Mean Square  | F     | Prob>F |
|------------|----|----------------|--------------|-------|--------|
| Regression | 4  | 1838.12922407  | 459.53230602 | 10.44 | 0.0001 |
| Error      | 25 | 1100.57077593  | 44.02283104  |       |        |
| Total      | 29 | 2938.70000000  |              |       |        |

| Variable | Parameter Estimate | Standard Error | Type II Sum of Squares | F     | Prob>F |
|----------|--------------------|----------------|------------------------|-------|--------|
| INTERCEP | -18.30152045       | 3.64892257     | 1107.44716129          | 25.16 | 0.0001 |
| RUN      | 11.13212935        | 2.48810400     | 881.24648295           | 20.02 | 0.0001 |
| SMOKE    | -6.96302377        | 2.63262467     | 307.96107626           | 7.00  | 0.0139 |
| PHYS1    | 13.32514812        | 4.09240540     | 466.72897076           | 10.60 | 0.0032 |
| PHYS2    | 7.45071026         | 3.75440264     | 173.37705597           | 3.94  | 0.0583 |

Bounds on condition number: 2.396734, 27.36375

All variables left in the model are significant at the 0.1000 level.

Summary of Backward Elimination Procedure for Dependent Variable PULSE

| Step | Variable<br>Removed<br>Label | Number<br>In | Partial<br>$R^2$ | Model<br>$R^2$ | $C(p)$ | F      | Prob>F |
|------|------------------------------|--------------|------------------|----------------|--------|--------|--------|
| 1    | HEIGHT<br>HEIGHT(INCHES)     | 5            | 0.0008           | 0.6290         | 5.0481 | 0.0481 | 0.8283 |
| 2    | WEIGHT<br>WEIGHT(POUNDS)     | 4            | 0.0035           | 0.6255         | 3.2634 | 0.2241 | 0.6402 |

13.64 参见练习 13.63。

- 对于未做原地跑、很少参加体育活动的人,给出其脉率差关于身高和体重的回归线。
- 考虑下面两种说法:
  - 在吸烟变量和体育活动虚拟变量之间有复共线性。
  - 在吸烟变量和体育活动虚拟变量之间有交互作用。

这两种说法之间有什么差别?如果这两种说法都是对的,解释数据集中存在的关系。

13.65 参见练习 13.63。

- 你所选择的好的预测方程是什么?为什么你选该方程?
- 上述模型中不包含任何交互作用。构造一个模型,使其能够包括每两个定量变量之间的交互作用。

13.66(化学) 这个练习中的数据取自 Brown, Healy and Kearns(1981)关于钙的化学化验的讨论。准备好一组标准的溶液,把这些标准的以及一些未知的溶液用分光光度计按某个单位测量出来( $y$ )。用一个线性模型拟合标准的溶液数据,然后把未知溶液的测量值代入,得到未知溶液的值( $x$ )。这些标准溶液和未知的溶液的准备需要许多实验室操作,标准溶液的实际浓度与它们的目标值之间会有些差异。这些差异可以用这种很精确的仪器测量出来。目标值分别为 2.0, 2.0, 2.5, 3.0, 3.0mmol/l(毫摩尔/升),“重复”测量独立地进行。对标准和未知的溶液重复测量四次,在每次化验中,每种未知溶液有两个样品,每个样品都有四个读数,其中一个样品用仪器中最优选择的条件进行测量,而另一个样品则用稍弱一点的条件进行测量。以  $y$  记分光光度计读数,  $x$  记实际的 mmol/l 数。

下面的数据表是由四个实验室按照上述方法进行化验得到的六个未知溶液样品的数据。标准溶液用 2.0A, 2.0B, 2.5, 3.0A, 3.0B 表示,未知溶液则以 U1, U2, W1, W2, Y1, Y2 表示。

| 实验室/溶液 |      | 测量值  |      |      |  | 实验室/溶液 |      | 测量值  |      |      |  |
|--------|------|------|------|------|--|--------|------|------|------|------|--|
| 1 W1   | 1206 | 1202 | 1202 | 1201 |  | 3 W1   | 1090 | 1098 | 1090 | 1100 |  |
| 1 2.0A | 1068 | 1071 | 1067 | 1066 |  | 3 2.0A | 969  | 975  | 969  | 972  |  |
| 1 W2   | 1194 | 1193 | 1189 | 1185 |  | 3 U2   | 1088 | 1092 | 1087 | 1085 |  |
| 1 2.0B | 1072 | 1068 | 1064 | 1067 |  | 3 2.0B | 969  | 960  | 960  | 966  |  |
| 1 U1   | 1387 | 1387 | 1384 | 1380 |  | 3 U1   | 1270 | 1261 | 1261 | 1269 |  |
| 1 2.5  | 1333 | 1321 | 1326 | 1317 |  | 3 2.5  | 1196 | 1196 | 1209 | 1200 |  |
| 1 U2   | 1394 | 1390 | 1383 | 1376 |  | 3 W2   | 1261 | 1268 | 1270 | 1273 |  |
| 1 3.0A | 1579 | 1576 | 1578 | 1572 |  | 3 3.0A | 1451 | 1440 | 1439 | 1449 |  |
| 1 Y1   | 1478 | 1480 | 1473 | 1466 |  | 3 Y1   | 1352 | 1349 | 1353 | 1343 |  |
| 1 3.0B | 1579 | 1571 | 1579 | 1567 |  | 3 3.0B | 1439 | 1433 | 1433 | 1445 |  |
| 1 Y2   | 1483 | 1477 | 1482 | 1472 |  | 3 Y2   | 1349 | 1353 | 1349 | 1355 |  |
| 2 W1   | 1017 | 1017 | 1012 | 1020 |  | 4 2.0A | 1122 | 1117 | 1119 | 1120 |  |
| 2 2.0A | 910  | 916  | 915  | 915  |  | 4 W2   | 1256 | 1254 | 1256 | 1263 |  |
| 2 W2   | 1012 | 1018 | 1015 | 1023 |  | 4 W1   | 1260 | 1251 | 1252 | 1264 |  |
| 2 2.0B | 913  | 923  | 914  | 921  |  | 4 2.0B | 1122 | 1110 | 1111 | 1116 |  |
| 2 U1   | 1188 | 1199 | 1197 | 1202 |  | 4 U2   | 1453 | 1447 | 1451 | 1455 |  |
| 2 2.5  | 1129 | 1148 | 1136 | 1147 |  | 4 2.5  | 1386 | 1381 | 1381 | 1387 |  |
| 2 U2   | 1186 | 1196 | 1193 | 1199 |  | 4 U1   | 1450 | 1446 | 1448 | 1457 |  |
| 2 3.0A | 1359 | 1378 | 1370 | 1373 |  | 4 3.0A | 1656 | 1663 | 1659 | 1665 |  |
| 2 Y1   | 1263 | 1280 | 1280 | 1279 |  | 4 Y2   | 1543 | 1548 | 1543 | 1545 |  |
| 2 3.0B | 1349 | 1361 | 1359 | 1363 |  | 4 3.0B | 1658 | 1658 | 1661 | 1660 |  |
| 2 Y2   | 1259 | 1269 | 1259 | 1265 |  | 4 Y1   | 1545 | 1546 | 1548 | 1544 |  |

a. 画出标准溶液的  $y$  对  $x$  的图, 每个实验室单独画一张。

b. 对于每个实验室拟合线性回归方程  $y = \beta_0 + \beta_1 x + \varepsilon$ , 由此对于每种未知溶液的测量值  $y$  预测相应的值  $x$ 。基于对每种未知溶液的四个  $x$  的预测值, 计算  $x$  的预测值的标准差。

c. 哪一个实验室看起来能做出更好的  $x$  的预测? 为什么?

**13.67** 参见练习 13.66。假定对每种未知的溶液算出两个样品的  $y$  的平均值, 并用之拟合练习 13.66 中的线性模型。

a. 对于每个实验室, 线性回归线有变化吗?

b. 对于这四个实验室, 用这些新的回归方程预测  $x$ , 预测值有变化吗? 为什么?

**13.68** 参见练习 13.66。使用自变量  $x$ , 建议一个一般线性模型, 使其可以用来拟合来自所有四个实验室的数据。解释这个一般线性模型中的参数。



**13.69** 参见练习 13.68。

a. 拟合练习 13.68 中的模型。

b. 分别给出每个实验室的回归模型。

c. 这些回归模型与前面得到的回归方程有区别吗？

d. 拟合一个模型与分别对各个实验室拟合模型比较起来，有哪些优点？

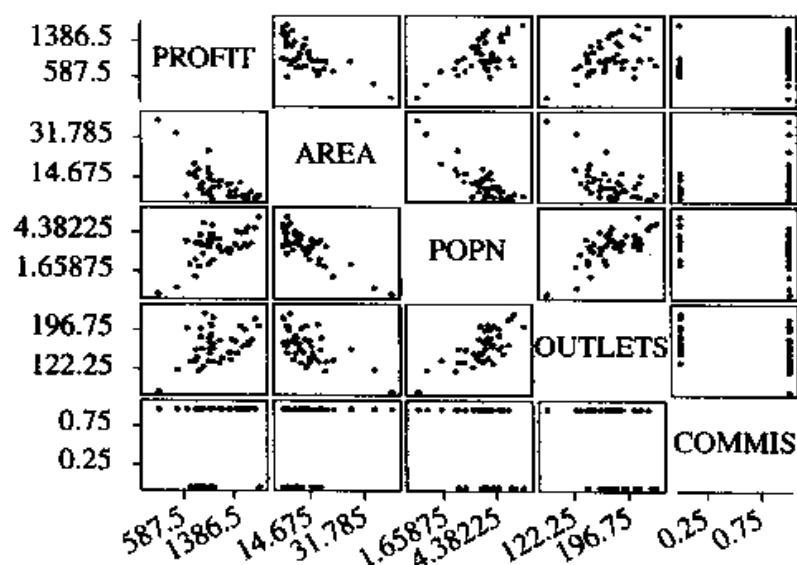
**13.70** 参见本章开始时介绍的案例。用下面给出的计算机输出回答销售部经理的问题。特别地，确定经理认为对利润有影响的那些变量是否真有影响。考虑单个的变量以及这些变量的可能的函数。例如，利润和销路数量的关系会否由于两种不同的佣金发放方式而有所不同。省去任何没有多少预测价值的变量。考察回归模型所需要的条件，看是否有违反这些条件的现象。如果有任何违反的现象，作出必要的矫正，使得回归分析的结果可用。对于你通过回归分析得到的结果，给出非技术性的解释，并把你所得到的关于销售部经理的问题的结论写成一份报告。

Data Display

| Row | DIST | PROFIT | AREA  | POP   | OUTLETS | COMMIS | AREA_2  | POP_2   |
|-----|------|--------|-------|-------|---------|--------|---------|---------|
| 1   | 1    | 1011   | 16.96 | 3.881 | 213     | 1      | 287.64  | 15.0622 |
| 2   | 2    | 1318   | 7.31  | 3.141 | 158     | 1      | 53.44   | 9.8659  |
| 3   | 3    | 1556   | 7.81  | 3.766 | 203     | 1      | 61.00   | 14.1828 |
| 4   | 4    | 1521   | 7.31  | 4.587 | 170     | 1      | 53.44   | 21.0406 |
| 5   | 5    | 979    | 19.84 | 3.648 | 142     | 1      | 393.63  | 13.3079 |
| 6   | 6    | 1290   | 12.37 | 3.456 | 159     | 1      | 153.02  | 11.9439 |
| 7   | 7    | 1596   | 6.15  | 3.695 | 178     | 1      | 37.82   | 13.6530 |
| 8   | 8    | 1155   | 14.21 | 3.609 | 182     | 1      | 201.92  | 13.0249 |
| 9   | 9    | 1412   | 7.45  | 3.801 | 181     | 1      | 55.50   | 14.4476 |
| 10  | 10   | 1194   | 14.43 | 3.322 | 148     | 1      | 208.22  | 11.0357 |
| 11  | 11   | 1054   | 6.12  | 5.124 | 227     | 0      | 37.45   | 26.2554 |
| 12  | 12   | 1157   | 11.71 | 4.158 | 139     | 1      | 137.12  | 17.2890 |
| 13  | 13   | 1001   | 9.36  | 3.887 | 179     | 0      | 87.61   | 15.1088 |
| 14  | 14   | 831    | 19.41 | 2.230 | 124     | 1      | 376.75  | 4.9729  |
| 15  | 15   | 857    | 11.75 | 4.468 | 205     | 0      | 138.06  | 19.9630 |
| 16  | 16   | 188    | 40.34 | 0.297 | 85      | 1      | 1627.32 | 0.0882  |
| 17  | 17   | 1030   | 7.16  | 4.224 | 211     | 0      | 51.27   | 17.8422 |
| 18  | 18   | 1331   | 9.37  | 3.427 | 145     | 1      | 87.80   | 11.7443 |
| 19  | 19   | 643    | 7.62  | 4.031 | 205     | 1      | 58.06   | 16.2490 |
| 20  | 20   | 992    | 27.54 | 2.370 | 166     | 1      | 758.45  | 5.6169  |
| 21  | 21   | 795    | 15.97 | 3.903 | 149     | 1      | 255.04  | 15.2334 |
| 22  | 22   | 1340   | 12.97 | 3.423 | 186     | 1      | 168.22  | 11.7169 |

| Row | DIST | PROFIT | AREA  | POPN  | OUTLETS | COMMIS | AREA_2  | POPN_2  |
|-----|------|--------|-------|-------|---------|--------|---------|---------|
| 23  | 23   | 689    | 17.36 | 2.390 | 141     | 0      | 301.37  | 5.7121  |
| 24  | 24   | 1726   | 6.24  | 4.947 | 223     | 1      | 38.94   | 24.4728 |
| 25  | 25   | 1056   | 11.20 | 4.166 | 176     | 0      | 125.44  | 17.3556 |
| 26  | 26   | 989    | 18.09 | 4.063 | 187     | 1      | 327.25  | 16.5080 |
| 27  | 27   | 895    | 13.32 | 3.105 | 131     | 1      | 177.42  | 9.6410  |
| 28  | 28   | 1028   | 14.97 | 4.116 | 170     | 0      | 224.10  | 16.9415 |
| 29  | 29   | 771    | 21.92 | 1.510 | 144     | 1      | 480.49  | 2.2801  |
| 30  | 30   | 484    | 34.91 | 0.741 | 126     | 1      | 1218.71 | 0.5491  |
| 31  | 31   | 917    | 8.46  | 5.260 | 234     | 0      | 71.57   | 27.6676 |
| 32  | 32   | 1786   | 7.52  | 5.744 | 210     | 0      | 56.55   | 32.9935 |
| 33  | 33   | 1063   | 14.43 | 2.703 | 141     | 1      | 208.22  | 7.3062  |
| 34  | 34   | 1001   | 15.37 | 3.583 | 158     | 0      | 236.24  | 12.8379 |
| 35  | 35   | 1052   | 11.20 | 4.469 | 167     | 1      | 125.44  | 19.9720 |
| 36  | 36   | 1610   | 7.20  | 4.951 | 174     | 1      | 51.84   | 24.5124 |
| 37  | 37   | 1486   | 13.49 | 3.474 | 211     | 1      | 181.98  | 12.0687 |
| 38  | 38   | 1576   | 6.56  | 4.637 | 172     | 1      | 43.03   | 21.5018 |
| 39  | 39   | 1665   | 9.35  | 3.900 | 185     | 1      | 87.42   | 15.2100 |
| 40  | 40   | 878    | 11.12 | 3.766 | 166     | 0      | 123.65  | 14.1828 |
| 41  | 41   | 849    | 10.58 | 3.876 | 189     | 0      | 111.94  | 15.0234 |
| 42  | 42   | 775    | 17.82 | 2.753 | 164     | 0      | 317.55  | 7.5790  |
| 43  | 43   | 1012   | 10.03 | 4.449 | 193     | 0      | 100.60  | 19.7936 |
| 44  | 44   | 1436   | 10.01 | 4.680 | 157     | 1      | 100.20  | 21.9024 |

案例中的散点图阵



CORRELATION ANALYSIS FOR CASE STUDY

5 VAR VARIABLES: PROFIT AREA POPN OUTLETS COMMIS

SIMPLE STATISTICS

| VARIABLE | N  | MEAN     | STD DEV  | SUM      | MINIMUM  | MAXIMUM  |
|----------|----|----------|----------|----------|----------|----------|
| PROFIT   | 44 | 1114     | 346.6525 | 48995    | 188.0000 | 1786     |
| AREA     | 44 | 13.2798  | 7.2163   | 584.3100 | 6.1200   | 40.3400  |
| POP      | 44 | 3.6757   | 1.0910   | 161.7310 | 0.2970   | 5.7440   |
| OUTLETS  | 44 | 172.1364 | 31.0795  | 7574     | 85.0000  | 234.0000 |
| COMMIS   | 44 | 0.6818   | 0.4712   | 30.0000  | 0        | 1.0000   |

PEARSON CORRELATION COEFFICIENTS/PROB> |R| UNDER HO:RHO = 0/N = 44

|         | PROFIT   | AREA     | POP      | OUTLETS  | COMMIS   |
|---------|----------|----------|----------|----------|----------|
| PROFIT  | 1.00000  | -0.68852 | 0.60822  | 0.43273  | 0.23584  |
|         | 0.0      | 0.0001   | 0.0001   | 0.0033   | 0.1233   |
| AREA    | -0.68852 | 1.00000  | -0.83383 | -0.63678 | 0.18534  |
|         | 0.0001   | 0.0      | 0.0001   | 0.0001   | 0.2284   |
| POP     | 0.60822  | -0.83383 | 1.00000  | 0.73194  | -0.28712 |
|         | 0.0001   | 0.0001   | 0.0      | 0.0001   | 0.0588   |
| OUTLETS | 0.43273  | -0.63678 | 0.73194  | 1.00000  | -0.33842 |
|         | 0.0033   | 0.0001   | 0.0001   | 0.0      | 0.0246   |
| COMMIS  | 0.23584  | 0.18534  | -0.28712 | -0.33842 | 1.00000  |
|         | 0.1233   | 0.2284   | 0.0588   | 0.0246   | 0.0      |

Best Subsets Regression:PROFIT versus AREA,POP,...

Response is PROFIT

| Vars | R-Sq | R-Sq(adj) | C-p  | S      | A | N | S | S | 2 | 2 | E | P | S |
|------|------|-----------|------|--------|---|---|---|---|---|---|---|---|---|
| 1    | 47.4 | 46.2      | 23.4 | 254.37 | X |   |   |   |   |   |   |   |   |
| 1    | 40.9 | 39.4      | 31.3 | 269.75 |   |   |   |   | X |   |   |   |   |
| 2    | 62.8 | 61.0      | 6.9  | 216.62 | X |   |   |   |   |   |   |   | X |
| 2    | 61.4 | 59.5      | 8.5  | 220.52 | X |   |   |   |   |   | X |   |   |
| 3    | 65.7 | 63.1      | 5.4  | 210.61 | X |   |   |   |   | X |   |   | X |
| 3    | 65.0 | 62.4      | 6.2  | 212.67 |   | X |   |   | X | X | X |   |   |

[illegible]

Regression Analysis: PROFIT versus AREA, POPN, ...

The regression equation is

$$\begin{aligned} \text{PROFIT} = & 1710 - 18.5 \text{ AREA} + 196 \text{ POPN} - 9.46 \text{ OUTLETS} - 267 \text{ COMMIS} + 0.140 \text{ AREA\_2} \\ & + 24.7 \text{ POPN\_2} - 20.1 \text{ COMM\_AREA} - 337 \text{ COMM\_POPN} \\ & + 11.9 \text{ COMM\_DISTRICTS} \end{aligned}$$

| Predictor  | Coef    | SE Coef | T      | P     |
|------------|---------|---------|--------|-------|
| Constant   | 1710    | 1254    | 1.36   | 0.182 |
| AREA       | - 18.51 | 33.11   | - 0.56 | 0.580 |
| POPN       | 195.5   | 349.6   | 0.56   | 0.580 |
| OUTLETS    | - 9.459 | 4.715   | - 2.01 | 0.053 |
| COMMIS     | - 267   | 1176    | - 0.23 | 0.822 |
| AREA _ 2   | 0.1402  | 0.6870  | 0.20   | 0.839 |
| POPN _ 2   | 24.74   | 37.79   | 0.65   | 0.517 |
| COMM _ ARE | - 20.10 | 33.24   | - 0.60 | 0.549 |
| COMM _ POP | - 336.8 | 162.6   | - 2.07 | 0.046 |
| COMM _ DIS | 11.874  | 5.023   | 2.36   | 0.024 |

$$S = 207.0 \quad R - Sq = 71.8 \% \quad R - Sq(\text{adj}) = 64.3 \%$$

### Analysis of Variance

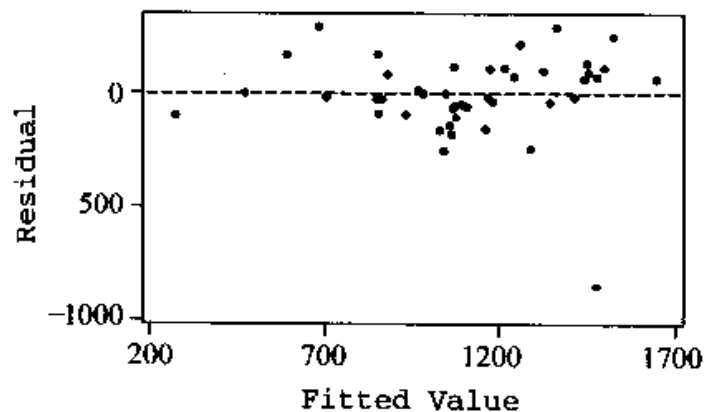
| Source         | DF | SS      | MS     | F    | P     |
|----------------|----|---------|--------|------|-------|
| Regression     | 9  | 3709689 | 412188 | 9.62 | 0.000 |
| Residual Error | 34 | 1457534 | 42869  |      |       |
| Total          | 43 | 5167223 |        |      |       |

|                      |      |         |        |        |          |          |
|----------------------|------|---------|--------|--------|----------|----------|
| Source               | DF   | Seq SS  |        |        |          |          |
| AREA                 | 1    | 2449586 |        |        |          |          |
| POPN                 | 1    | 19727   |        |        |          |          |
| OUTLETS              | 1    | 8896    |        |        |          |          |
| COMMIS               | 1    | 814136  |        |        |          |          |
| AREA_2               | 1    | 19932   |        |        |          |          |
| POPN_2               | 1    | 25074   |        |        |          |          |
| COMM_ARE             | 1    | 78282   |        |        |          |          |
| COMM_POP             | 1    | 54524   |        |        |          |          |
| COMM_DIS             | 1    | 239533  |        |        |          |          |
| Unusual Observations |      |         |        |        |          |          |
| Obs                  | AREA | PROFIT  | Fit    | SE Fit | Residual | St Resid |
| 16                   | 40.3 | 188.0   | 279.2  | 175.0  | - 91.2   | - 0.82X  |
| 19                   | 7.6  | 643.0   | 1484.3 | 74.9   | - 841.3  | - 4.36R  |

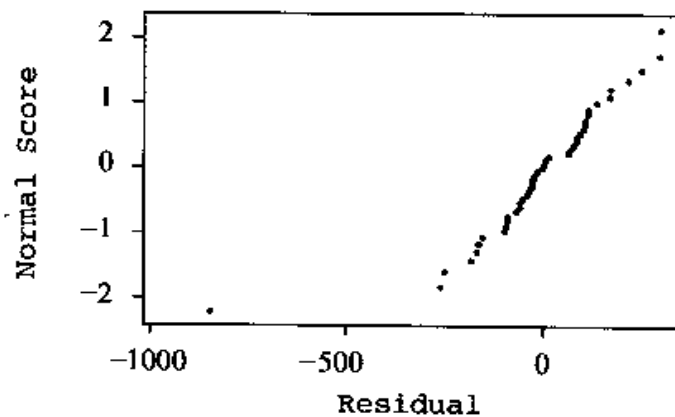
R denotes an observation with a large standardized residual

X denotes an observation whose X value gives it large influence.

Residuals versus the Fitted Values  
(response is PROFIT)



Normal Probability Plot of the Residuals  
(response is PROFIT)



## 第七部分 试验设计与方差分析

第十四章 试验和研究的设计概念

第十五章 标准设计的方差分析

第十六章 协方差分析

第十七章 一些固定效应、随机效应和混合效应模型的方差分析

第十八章 重复测量与交叉设计

第十九章 一些非平衡设计的方差分析

第二十章 分析结果的传达和备案



## 第十四章 试验和研究的设计概念

- 14.1 引言
- 14.2 研究的类型
- 14.3 设计的试验;术语
- 14.4 控制试验误差
- 14.5 试验单元对处理的随机化
- 14.6 确定重复试验的次数
- 14.7 小结

### 14.1 引言

一个试验的设计是建立一种方案的过程,通过该方案的实施,我们可以根据已记录下的响应对不同的处理组作出比较。我们可以用许多不同的方法来进行一项科学研究。在某些研究中,研究人员从未受干扰的自然过程或环境来收集信息——例如,一项有关公立、教会及私立学校的二年级学生阅读得分差异的研究。在另一些研究中,科学家在环境条件受到高度控制的实验室中工作,而这样的环境条件,可能是为了进行该项研究完全由人为设置的。有关温度和湿度对壁虱生命周期影响的研究就是在实验室中进行的,因为在壁虱生活的自然环境中,温度和湿度是无法人为控制的。对研究中因素的这种控制,使得昆虫学家更易于用温度和湿度水平的不同来解释所得到的结果,因为在整个试验过程中几乎所有的其他条件是保持不变的。然而,在自然环境下,许多其他因素的变化也会造成壁虱生命周期的变化。这样,在这些人为的环境下,人为控制越多,试验能刻划的自然界真实情况就越少。为了使试验有实用价值,必须在条件的人为控制和反映现实之间保持某种切实的平衡。在这一章,我们将给出一些标准的试验设计方案,并介绍用来分析从这些试验得来的数据的方法。

14.2 节描述几种类型的研究,并且给出与设计的试验有关的几个名词术语。在 14.3 节中,我们将讨论在一项研究中,实施试验之前需要考虑的若干重要元素。在 14.4 和 14.5 节中,讨论在设计的试验中,随机化对于控制试验误差和建立有效的推断机制所起的作用。14.6 节给出了几种确定为从试验数据得到所关心的总体和处理的适当推断所需要的重复次数的方法。



## 14.2 研究的类型

在第二章,我们讨论了观察研究和设计的研究之间的区别。各种形式的观察研究,如民意测验、抽样调查、流行病学研究等等常被用于许多不同的背景以回答所关心的问题。抽样调查被用于了解国民对有关如控制枪支、利率、税款、福利、社会安全和国债等问题的观点的变动。同样,我们从每日的报纸、杂志、电视、广播及互联网收到其他有关政治、社会、教育和健康问题的民意测验的结果。

在观察研究中,所关心的因子(或处理)在我们进行测量时不受控制。例如,一个引起极大关注的公共健康问题是心脏病与饮食中的脂肪含量之间的联系。随机地为志愿者指定几种不同脂肪含量的食谱之一,然后在一段时间内监视他们以观察是否患上心脏病,这样做既不道德而且要花费太多的金钱和时间。由于不能控制所关心的因子(食物中的脂肪含量),科学家必须用观察研究来解决这一问题。可以这样做:在心脏病患者和没有心脏病的人中各抽取一个样本,然后比较这两组人的饮食。此外,还必须注意记录下其他的相关因子,例如,每个人家族的心脏病史、吸烟习惯、日常锻炼、年龄和性别等,用以对组间的差异进行调整,排除食物脂肪含量外的其他因子的影响。即使进行了这些调整,我们也很难指称在食物的高脂肪含量与心脏病之间具有因果关系。事实上,如果在对于其他因素进行了相应的调节后,发现患心脏病组的食物脂肪含量趋向于比不患心脏病组的高,我们将报告说,食物的高脂肪含量与心脏病之间有关联,而不是说二者之间具有因果关系。

换言之,在观察试验中,我们从因子(或处理)已经存在的总体中抽取样本,然后,我们就关心的因子(或处理)对样本组进行比较。与之相反,在试验环境受到控制的试验研究中,我们可以随机地为作为研究对象的人指定因子(或处理),然后观察所关心的响应。对于我们上面心脏病的例子,两者的区别如下:

**观察研究:**我们分别从心脏病患者的总体和不患心脏病的人的总体中抽取样本,然后比较这两组样本中人的食物的脂肪含量。

**试验研究:**忽略道德因素,我们将给志愿参与者分配几种不同脂肪含量的食谱(处理)之一,然后在一段时期后,就所关心的响应(心脏病)比较不同的处理。

在这一章,我们集中讨论设计的试验。在这类研究中,研究人员通过以下两种方法中的一种来控制关键的因子。

**方法 1:**把试验中的研究对象随机地分配给处理。例如,给 10 只老鼠随机地分配所研究的试验药物的 4 种剂量水平中的一种。

**方法 2:**从所关心的不同总体中随机地选择研究对象。例如,从大城市和小城市的动物收容所中随机地选出 50 只母狗和 50 只公狗,然后检测它们体内是否有

犬恶丝虫。

在方法 1 中,研究人员从一个由具有一致性的试验单元构成的总体中抽取试验单元,并且完全控制随后进行的把试验单元随机地指定给处理这一过程。在方法 2 中,研究人员能控制从处理总体的随机抽样,但不能控制试验单元对处理的分配。

在设计的试验中,很关键的一点是科学家遵从试验开始之前已经建立好的系统方案。这一方案包括如何实施所有的随机化,如何把试验单元分配给处理或如何从处理总体中选出试验单元。外部的因素可能会出现,从而影响试验单元。这些因素可能是试验单元中轻微的差别或在试验进行期间外部环境中的细微差别。随机化过程保证,平均来说,任何观察到的不同处理组中试验单元响应间的较大差别,可被归因于所关心的因素间的差别,而不是归因于试验过程中未受控的因素。试验方案还应当包括许多其他有关如何进行试验的内容,其中一些应当包含的事项列如下表:

1. 试验的研究目标。
2. 选择要变化的因子(处理)。
3. 识别试验单元或试验环境中可能出现的外部因素(区组因子)。
4. 要测量的试验单元的特征(响应变量)。
5. 随机化的方法,即从处理总体中随机抽样或者是随机地给处理分配试验单元。
6. 记录试验单元的响应所用的方法。
7. 选择分配给每一处理的试验单元的个数,这可能需要指定显著性水平和检验的功效或置信区间的精度和可靠性。
8. 一张完整的可利用的资源 and 材料的清单。

本书第十五章到第十九章讲述设计的试验的分析方法。

### 14.3 设计的试验:术语

我们将对若干词语的普通用法和在设计的试验的框架内为这些词语和概念所赋予的意义做出区分。更简洁地说,设计的试验是一种这样的研究,它为就一个指定的响应变量对多组个体进行观察、测量和评价提供了一种特定的方案。研究人员在试验的过程中,控制该方案中的要素以得到数据,并且通过对该数据的统计推断对所关心的组作出有效的比较。

我们用下面例子来说明此处将要定义的概念和术语。

## 例 14.1

研究人员正在研究在何种条件下,商业养殖的虾能达到最大重量的增加。研究选择了3种水温( $25^{\circ}$ ,  $30^{\circ}$ ,  $35^{\circ}$ )和4种水的含盐量(10%, 20%, 30%, 40%)。虾放养在指定水温和盐分的容器中。在6周以后,记录下每一个容器中虾重量的增加。有许多其他的因素会影响虾的重量增加,如容器中虾的密集程度、虾的种类、虾的大小及食物类型等。试验如下进行:研究人员有24只容器可以利用;为研究选定了虾的种类和苗虾的大小;容器中虾的放养密度固定在一个给定的水平上不变;给24只容器中的每一只分配3种水温中的一种和4种盐分中的一种;规定在研究期间所有其他可识别的条件对于24只容器都相同。事实上,这些量的水平会有一些变化。在容器中养殖6周后,把虾捞出并称得重量。

在设计的试验中有两类变量。被称为**因子**的受控制的变量是被研究人员选出来用来作比较的。响应变量是被研究人员记录到的但未受控制的**测量值或观测值**。受控的变量形成研究假设所定义的比较组。在我们的例子中,水温和盐分是受控变量或因子,虾的重量是响应变量。

设计的试验中的**处理**是从因子构造出来的条件。因子是通过考察研究假设所提出的问题选择出来的。在某些试验中,可能只有单一的一个因子,因而因子的水平和处理是相同的。这类处理设计称为**单向分类**。在多数情况下我们有几个因子,处理是由各因子的水平组合而成的。这类处理设计称为**因子处理设计**。对于14.1中描述的研究虾的例子,有两个因子:具有3个水平( $25^{\circ}$ ,  $30^{\circ}$ ,  $35^{\circ}$ )的水温和具有4个水平(10%, 20%, 30%, 40%)的盐分。这样我们可以用两因子的各水平组合生成 $3(4) = 12$ 个处理。表示这12种处理的因子水平组合如下:

( $25^{\circ}$ , 10%) ( $25^{\circ}$ , 20%) ( $25^{\circ}$ , 30%) ( $25^{\circ}$ , 40%)

( $30^{\circ}$ , 10%) ( $30^{\circ}$ , 20%) ( $30^{\circ}$ , 30%) ( $30^{\circ}$ , 40%)

( $35^{\circ}$ , 10%) ( $35^{\circ}$ , 20%) ( $35^{\circ}$ , 30%) ( $35^{\circ}$ , 40%)

在例14.1中使用了24只容器,并且它们被随机地分配给12种处理,每两只容器分到一种处理。

在其他情况下,因子的个数可能很大,因此处理的个数也会变得很大,这样试验只能考察所有可能处理的一个子集。例如,假定我们正考虑研究下列因素对大豆亩产量的影响。因子有:因子1——5种大豆品种;因子2——3种植植密度;因子3——4种施肥水平;因子4——得克萨斯州的6个地方;因子5——3种灌溉等级。从这5个因子我们能形成 $(5)(3)(4)(6)(3) = 1080$ 个不同的处理。这将使得试验非常巨大而且费用昂贵。在这种情形下,将从1080种可能的处理中选择一个处理子集来研究五个因子与大豆产量之间的关系。这类试验称为**部分析因试验**,因为在试验中实际上只使用了部分可能的处理。为了尽可能多的回答研究人员关心的问题,必须要对选择哪些处理进行试验给予极大的重视。

一类特殊的处理被称为**对照处理**(control treatment, 又译控制处理), 这类处理构成了比较其余处理的效应的基准。有三种情况特别需要对照处理。首先, 试验进行的条件可能会阻止有效的处理展现他们的效用。在这种情况下, 由没有处理构成的对照处理可能有助于揭示; 试验条件不允许处理显示它们效应间的差异。例如, 做一个试验, 以确定种植西红柿的菜园中最有效的施氮肥量。如果在向试验用的土壤中施加氮肥之前, 土壤的肥力已经很高, 那么所有施加的氮肥量的不同水平将呈现出相等的效应。然而如果在试验中使用了由不施氮肥组成的处理, 即对照处理, 那么土壤的高肥力就会因为对照处理与施加氮肥的处理有同样效果而被发现。

第二类对照是其他所有的处理都与之相比的标准处理方法。提出几种新的方法以取代现有的、完善的方法, 就属于这种情况。最后, 安慰剂对照被用于试验期间通过控制研究对象就可以从研究对象得到一响应值的情形。例如, 人们会因为看了医生并得到处方的治疗而表现出疼痛程度的暂时缓解。这样, 在评价减轻病人疼痛的几种不同方法时, 把没有有效成分的处理即安慰剂在病人毫不知情的条件下分配给一组病人, 然后把有有效成分的处理与安慰剂作出比较以确定它们的真实疗效。

**试验单元**是接受我们随机分配的处理的物理整体, 或从一个处理总体中随机选出的个体。对于例 14.1 有关虾的研究中, 试验单元是容器。

考虑另一个试验, 其中研究人员要用试验用鼠来检验一种新药各种不同剂量(处理)的差别。如果研究人员随机地给每只老鼠指定一个药的剂量, 那么试验单元就是单只老鼠。一旦一个处理分配给了一个试验单元, 就说该处理发生了一次**重复**。一般来说, 我们会对每一处理随机地分配几个试验单元, 这样, 对于每一个特定的处理我们将得到若干独立的观察值, 因而该处理就有几次重复。在例 14.1 中, 每一处理我们有 2 次重复。

**测量单元**与试验单元不同, 它是我们对其进行测量的物理整体。在许多试验中, 试验单元和测量单元是一致的。例 14.1 中, 与试验单元一样测量单元也是容器。然而, 如果称量了每个容器中的每只虾的重量, 那么由于我们是将处理应用于容器的, 故试验单元将是容器, 而测量单元是虾。

### 例 14.2

考虑下面的试验。要评估用于煎炸用锅的四种防护涂料, 对四种涂料中的每一种涂料, 随机地指定五只煎锅。对于这 20 只煎锅, 在每只煎锅的 3 个指定区域测量了涂料的一种抗磨性能指标。确定这个研究中的下述各项: 试验的设计, 处理, 重复次数, 试验单元, 测量单元和总的测量次数。

### 解答

试验的设计:单向分类。

处理:四种防护涂料。

重复:每一处理有五只煎锅(即每一处理有五次独立重复实现)。

试验单元:煎锅,因为涂料(处理)是随机地指定给煎锅的。

测量单元:煎锅的指定区域。

总的测量次数: $4(5)(3) = 60$ 。

**试验误差**用来描述被分配给相同处理,并在“相同”的试验条件下被观测的各试验单元响应之间的变异。试验误差不为零的原因包括:(a)在接受处理之前试验单元间固有的差别;(b)记录测量设施的变异;(c)处理条件设置方面的变异;(d)除了处理因子以外的所有外部因子对响应变量的影响。

### 例 14.3

回到前面所讨论的实验室中的试验问题,其中研究人员随机地把单一剂量的药物分配给每一只老鼠,2 小时后,测量老鼠血液中药物的浓度。对于这项试验,试验单元和测量单元是相同的,即老鼠。给出该试验中四种可能的试验误差来源。[参见上面的(a)~(d)]

**解答** 我们可以写出如下这些试验误差的来源:

- a. 得到处理之前试验单元间固有的差别。老鼠之间存在细微的生理差别,因此即使对于获得完全相同剂量水平的两只老鼠,2 小时后他们血液中的药物含量也会稍有差别。
- b. 记录测量设施的变异。因实验技术人员确定老鼠体内药物含量时所用的方法的不同,响应间也会有差异。即使对同一老鼠做几次血液药物等级的测量,所测出来的药物含量也可能会因仪器间的变异,技术人员的不同,或实验室条件的变动而不尽相同。
- c. 处理条件设置方面的偏差。如果每一处理有一次以上的重复,那么处理可能随老鼠不同而不同。例如,假定对于每一剂量(处理)有十次重复。我们不大可能做到十只老鼠得到的药物剂量都与处理所规定的完全相等。注射器中的药量可能会稍有不同,注射并进入血流中的量也会稍有不同。
- d. 除处理因子外的所有外部因子对响应变量(血液水平)的影响。假定老鼠被放在笼子里,且在测定血液中药物含量之前给以等量的食物和水。然而,笼子的温度、湿度、外部刺激及其他条件可能有些不同,这都会对老鼠的响应产生影响。

这样,在试验期间实验室内外部条件的这些差别与变异都会对试验中试验误差的大小有所贡献。

为了检验研究假设和建立有关处理总体均值函数的置信区间,我们需要对**试验误差方差**进行估计。我们将在后续的章节中详细说明获得这种估计所用的方法,其中试验的设计,处理的设计及重复次数都影响到这一方差估计。在大多数基本试验中,我们只有一个具有  $t$  水平的因子,而且试验单元和测量单元是完全一致的,此时试验误差方差的估计是由来自获得相同处理的试验单元之响应的方差汇总而成的。在老鼠血液中药物水平的试验研究中,我们把老鼠随机地分配给四种处理,每种处理十只老鼠。通过计算每一剂量水平上十个血液药物水平响应的方差,再把四个单独的方差估计进行汇总,得到试验误差方差的一个估计。

#### 例 14.4

回到例 14.1,假定每一处理被分配给 2 个容器,且每个容器中放入 40 只虾。6 周以后,每只虾都称了重。确定试验单元、测量单元、因子、处理、重复观测数及可能的试验误差来源。

**解答** 这是一个因子处理设计,有两个因子,即是温度和盐分。分配给容器的处理由一个温度水平和一个盐分水平组合而成。试验中我们一共有  $(3)(4) = 12$  种可能的处理。这 12 种处理如下:

(25°, 10%) (25°, 20%) (25°, 30%) (25°, 40%)

(30°, 10%) (30°, 20%) (30°, 30%) (30°, 40%)

(35°, 10%) (35°, 20%) (35°, 30%) (35°, 40%)

我们然后随机地把 12 种处理中的每一种分配给 2 个容器,这导致每一处理有两次重复。试验单元是容器,因为处理是被随机地分配给单个容器的。40 只虾被放养在一个容器中,6 周以后,记录下每只虾的重量。测量单元是单个的虾,因为它是被测量的物理实体。由此,在该试验中,试验单元和测量单元是不同的。几个可能的误差来源是,放入容器之前虾重量间的差异;在 6 周的试验期间温度和盐分水平维持的准确性;在试验结束时,称量个虾重量的精度;喂虾的食物量的一致性(6 周期间每只虾都给以完全等量的食物了吗?)及可能影响虾生长的其他条件的差异。

## 14.4 控制试验误差

正如我们在例 14.3 和例 14.4 中看到的那样,试验中有许多潜在的试验误差

来源。当试验误差的方差较大时,我们推断的精确性将大打折扣,处理均值的估计将有较大的标准差,这将导致不精确(或较宽)的置信区间,同时也使得假设检验犯第Ⅱ类错误的概率较大。因而,能被用于减小试验误差的任何技巧都将大大改进试验,提高推断的精度。

研究人员能够控制许多种潜在的试验误差来源,这些误差来源包括:(1)进行试验的程序方法;(2)试验单元和测量单元的选择;(3)进行测量和记录所遵循的方法;(4)试验单元的区组划分;(5)试验设计的类型;(6)协变量的使用。下面我们来讨论每一误差来源是怎么影响试验误差的,研究人员如何使得这些误差来源对试验误差方差大小的影响达到最小。

### 试验方法

如果进行试验所必需的各程序步骤没有被仔细、精确地完成,结果会造成响应变量方差的增加。这儿不仅涉及到实施试验和测量响应变量的工作人员,也包括他们在工作中所使用的设备。全体工作人员必须在构造处理和进行试验方面得到良好的培训。他们的表现对试验的成功与否有直接的关系,这一点应予强调。研究人员应该在预算的范围内尽可能为技术人员提供能够得到最精确测量的设备。在整个试验过程中,对设备频繁、定期地进行维护和校准至关重要。整个试验期间必须尽可能保持试验的条件不变。否则,响应间的差异可能是因为试验条件的变化而不是处理间的差异造成的。当试验方案的质量不是很高时,响应变量的方差可能会膨胀。进行测量时所用的方法不恰当,仪器的刻度不准确或实验室内的条件未受控制,可能会导致出现不能真实反映处理对响应变量效应的极端观察值。极端观察值也可能因实验技术人员或数据管理人员的记录误差而产生。不管哪种情况,研究人员必须围绕极端观察值进行调查,然后决定是否从分析中删去该观察值。如果要删除这样的值,那么在最终报告的附录部分应给出为何不包括该数据的说明。

当在整个研究过程中,试验方案没有被一致地执行时,会造成两个可能的后果。一个是响应变量方差的膨胀,另一个是处理均值的估计偏差。例如,假定我们测量注射了四种可能的药物剂量之一的老鼠体内血液药物的含量。用于测量药物注射准确数量的仪器工作不正常。对给定的药物剂量,给第一只老鼠注射的药量少于规定的剂量,而给最后一只老鼠注射的药量超过规定的剂量。这样,当血液中药物的总量被测定后,这些测量值的方差将有所增加,但是处理均值的估计可能是无偏的,因为超量和不足量也许会相互抵消。另一方面,如果规定得到最低剂量水平的所有老鼠都被注射了过多的药物,而规定得到最高剂量水平的所有老鼠都未注射足够的药物,那么处理均值的估计将是有偏的。对于低剂量组处理的均值可能估计过高,而高剂量组处理的均值会被估计过低。因此,保证对于所有的试验单

元试验方案执行的一致性,对于试验的成败是至关重要的。对于实验室研究或现场研究,保证环境条件的一致性同样重要。外部因素如温度、湿度、光照量、暴露于空气中的污染物及其他未被控制的因素,如果没有一致地运用于试验单元,则会导致研究中方差的膨胀和处理均值估计的偏差。

### 选择试验单元和测量单元

如果试验中所用的试验单元在那些可能影响响应变量的特征方面不相似,则试验误差方差会增加。研究的一个目标是判定获得不同处理的试验单元的平均响应是否存在差异。研究人员必须确定所关心的试验单元总体。试验单元随机地从该总体中抽出,然后被随机地分配给处理,这当然是一种理想的状态。实际上,研究人员由于出于费用、可用性及道德方面的考虑,他对试验单元所能作的选择是受到限制的。因此,从这些试验数据可能得出的推断也只能是有限的。如果我们考察潜在的试验单元的群体,特征更接近的单元会产生更精确的处理均值的比较。然而,如果试验单元过分一致,那么能被作出适当推断的总体将受到很大的限制。试考虑下面的例子。

#### 例 14.5

经营儿童用品的一项营销活动要利用电视广告作为其主要的营销策略。一家市场调查公司受雇来确定孩子们对广告的关注时间的长度是否因广告中的产品类型不同而有所差异。该调查公司决定考察四类产品:体育器械,健康小食品,鞋和光盘游戏。该公司从一所纽约市的公立学校抽取了 100 名四年级的学生参与调查。每 25 名学生被随机地分配去看四种产品中的一种商业广告,记录下这 100 个孩子对广告的关注时间。市场调查公司认为从同样的学校体制及同一年级水平抽取参与者,所得到的研究对象组具有一致性。这一抽样方法存在什么问题?

**解答** 市场调查公司认为从同一年级水平和同样的学校选择学生,比更一般的抽样方法更能得到一致的试验单元集,这一点或许是正确的。然而,这一方法严重地限制了从该研究所能做出的推断:调查的结果可能仅适用于居住在非常大的城市中的四年级学生。包括其他年级和来自较小城市的孩子的抽样方法可能会提供更为实用的研究结果。

### 通过区组划分来减小试验误差方差

当我们认为可以利用的试验单元间在重要特征方面有较大差异时,利用区组划分来减小试验误差的方差被证明是非常有效的。根据它们在可能影响响应变量的特征上的相似性把试验单元分到各组,这样做不仅使得各试验单元的集合或区



组内部的各单元是相似的,而且当考虑所有的单元时,又不乏对重要特征的广泛包涵。分别在每一区组内随机地分配处理,然后在由相似的试验单元构成的组内进行处理的比较,由此得到的处理的比较不会因原先试验单元集间存在很大差异而被掩盖。区组设计使得我们能够把与用来划分区组的特征相联系的变异从试验误差中分离出来。

有许多准则可以用来对试验单元进行区组划分。包括下面几种:

1. 根据研究对象的物理特征,如年龄,体重,性别,健康状况,受教育程度等。
2. 密切相关的单元,如双胞胎或同一窝动物。
3. 试验单元的空间位置,如相邻的地块或同一试验台上植物的位置等。
4. 进行试验的时间,如一周中的每一天,因为环境条件可能天天在变。

5. 做试验的人,因为如果几个操作人员或技术人员进行试验,他们之间在如何进行测量或处理试验单元方面可能会有些差别。

在所有这些例子中,我们试图在每一区组中观察所有的处理。例如,如果我们研究三条装配线每一条上生产出有重大缺陷的汽车的数量,可能要用每周的工作日作为区组变量,并一定在每周的所有5个工作日内比较每一条装配线。

### 利用协变量来减少变异性

协变量是指与响应变量有关的变量。我们利用试验单元的物理特征来产生区组。例如,在比较一个新食谱和对照食谱对减少狗的体重的效果的研究中,假定被用来做研究的狗的年龄在1岁到12岁之间。我们可以把狗分成三个区组: $B_1$ —3岁以下的, $B_2$ —3岁到8岁的, $B_3$ —8岁以上的。一个更精确的方法是记录狗的年龄,然后在比较食谱的效果时把它结合到模型中去。在比较新食谱和对照食谱之前,根据狗的年龄对响应变量进行调整,这样我们会得到一个更精确的比较。不像在区组中那样使用年龄段,我们在这儿使用狗的确切的年龄,这就减小了试验误差的方差。

在一项试验中,选用哪些变量作为协变量取决于该试验本身。协变量应与响应变量有联系,它必须可以测量,且不受处理的影响。在大多数情况下,协变量在处理分配给试验单元之前就在试验单元上测量好了。协变量的例子有土壤的肥力、原料中的杂质含量、试验单元的重量、学生SAT(教育性向测验)的得分、研究对象的胆固醇含量、田间害虫的密度。下例将举例说明协变量的运用。

#### 例 14.6

在该研究中,我们将把补充光照(SL)和部分遮蔽(PS)对大豆产量的影响与正常光照(NL)作比较。正常光照将作为对照。每类光照都随机地被指定给15棵大豆植株,这些大豆被种植在温室里。当组织试验时,研究人员认识到这些植物的大

小和成熟程度有差别。因此,在试验开始时,测量出每个植株的高度这个植物活力的可测特征,并把它作为协变量。这就使得研究人员可以根据植株起初的大小来调整单株大豆的产量。对每一植株,我们记录两个量( $x, y$ ),其中  $x$  是试验开始时植株的高度,  $y$  是试验结束时大豆的产量。为了确定协变量是否对响应变量有影响,我们画出了关于这两个变量的图以认识他们间可能的联系。如果二者之间不存在任何联系,那么在分析中就不必使用协变量。如果两个变量有关联,那么必须在比较三种处理的均值之前用协方差分析的方法对响应变量作出适当的调整。初步评价关联存在性的方法是,对每一处理分别用不同记号画出响应变量对协变量的图。对于大豆数据,这种图如图 14.1 所示。

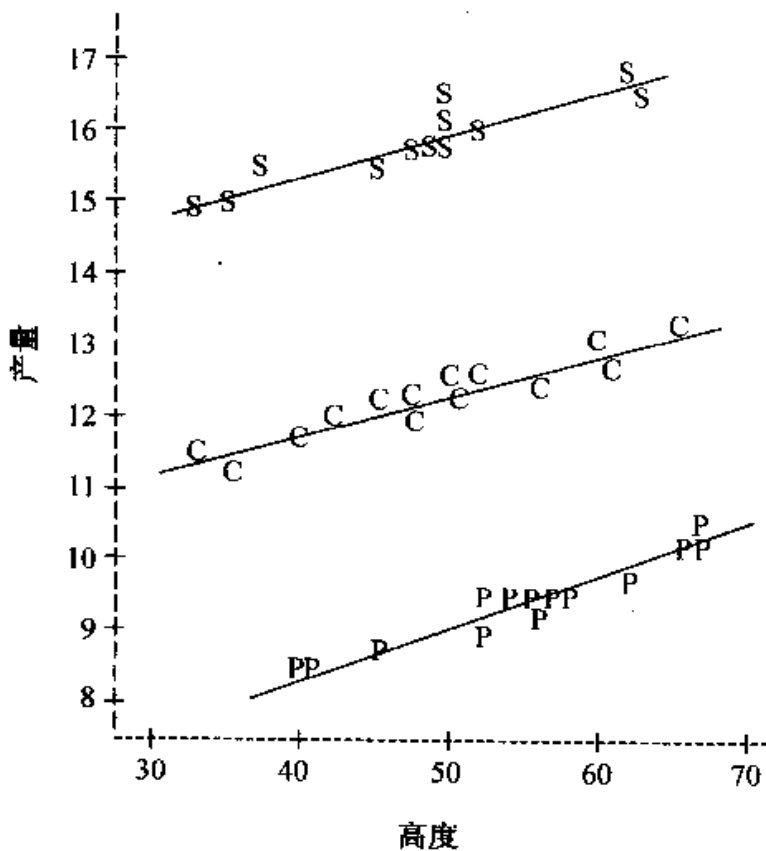


图 14.1 植株高度对产量的图;S 表示 SL,  
C 表示 NL,P 表示 PS

从图 14.1 我们看出,协变量即起初的植株高度和响应变量即产量之间表现出增函数的关系;而且,三种处理的产量看起来也不相同;响应变量间的一些变异既与植株起初的高度有关又与植物获得的光照量有关。这样,我们在检验三种处理平均产量间的差别之前,必须确定跟高度有关的方差的量。我们可以利用方差分析的方法达到这一目的。协方差分析的方法将在第十六章详细讨论。

## 14.5 试验单元对处理的随机化

与前面章节中的讨论一样,统计的方法基于这样的条件,即从试验中得到的数据等同于从正态分布的响应总体中抽取的随机样本。当试验是从已经确立的处理总体中随机地抽取试验单元时,我们实际上可以验证这一条件是否有效。然而,在那些其中我们要选取试验单元以满足特定的准则,或者试验单元是研究用农场中可利用的小块地的试验中,响应值构成来自某个总体的随机样本的想法是值得疑问的。

当我们随机地把处理分配给试验单元,然后观察响应变量时,要求这些观察值是独立的。本书用来估计参数,建立置信区间和进行假设检验的统计理论都基于观察互相独立这一条件。在更高等的书中,可以看到处理相依数据如时间序列数据或空间相关数据的方法。为了确信得到有效的结果,我们必须保证观察值是独立的。如果试验单元之间存在物理上的关系,那么这些观察值就会不独立。考虑下面的例子。

### 例 14.7

假定我们要比较四种根激素( $S_1, S_2, S_3, S_4$ )对云杉树苗生根密度的影响。32棵云杉树苗被栽种在盆中,这些盆如图 14.2 所示的那样被放在温室中的试验台上。在温室试验中,如果温室中的温度稍有不同或光照量有些变化,那么互相靠近的植物可能比互相远离的植物更能表现出相似的响应。这样的响应是正相关的。在其他的背景中,我们可能得到相反的结果,即互相临近的试验单元由于对资源的竞争,它们的响应倾向于朝向偏离总均值的两个相反方向。例如,如果动物从同一个容器里进食,那它们将会为获得食物而争抢。如果一些动物吃的比他们应得的

|       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
| II    |       |       |       | I     |       |       |       |
| $S_2$ | $S_2$ | $S_2$ | $S_2$ | $S_1$ | $S_1$ | $S_1$ | $S_1$ |
| $S_2$ | $S_2$ | $S_2$ | $S_2$ | $S_1$ | $S_1$ | $S_1$ | $S_1$ |
| III   |       |       |       | IV    |       |       |       |
| $S_3$ | $S_3$ | $S_3$ | $S_3$ | $S_4$ | $S_4$ | $S_4$ | $S_4$ |
| $S_3$ | $S_3$ | $S_3$ | $S_3$ | $S_4$ | $S_4$ | $S_4$ | $S_4$ |

图 14.2 试验台上的盆的排放

份额多,那它们长得较快,而其余的动物则倾向于较小。

为了以最有效的方式进行试验,技术人员分配激素如下:

将激素  $S_1$  施给位于第 I 象限的八棵树苗;

将激素  $S_2$  施给位于第 II 象限的八棵树苗;

将激素  $S_3$  施给位于第 III 象限的八棵树苗;

将激素  $S_4$  施给位于第 IV 象限的八棵树苗。

假定空调吹遍整个温室,并且台子上不同处的温度有些差别。具体地说, I 象限中的盆得到稍凉的空气, III 象限的盆得到稍暖的空气。在起初设计试验时,这些差别是不可识别的。这样,在此种安排之下,获得激素  $S_1$  的植物与获得其他激素的那些植物相比,可能处于竞争劣势,因为他们在相对较低的温度下生长。

试验单元物理上的接近或技术人员做试验的次序会使得来自这些试验单元的响应相关。为了避免这类情况的发生,我们需要随机地把处理分配给试验单元。随机分配为数据是独立的观察值这一模型条件提供了依据。然而,处理对试验单元的随机分配并不完全消除数据值相依的问题,数据之间的相关性也会来源于试验期间出现的其他情况。因此,试验人员始终必须要清楚任何能进入试验环境且导致来自某些试验单元的响应对其他试验单元的响应有影响的物理机制。

假定我们有  $N$  个齐性的试验单元和  $t$  个处理,我们想把  $r_i$  个试验单元随机地分配给第  $i$  个处理,此处,  $r_1 + r_2 + \cdots + r_t = N$ 。该随机分配包括下面几个步骤:

1. 对试验单元从 1 到  $N$  编号;
2. 用随机数表或计算机软件编程以得到  $1 \sim N$  这  $N$  个数的一个随机排列;
3. 把 1 号处理分给随机数列中前  $r_1$  个数对应的试验单元,把 2 号处理分给随机数列中接下来的  $r_2$  个数对应的试验单元。依此类推,直至把  $t$  号处理分给随机数列中最后  $r_t$  个数对应的试验单元。

我们将在下面举例说明这一方法的实施。

#### 例 14.8

假定我们考察例 14.7 中描述的研究,其中我们想要比较四种根激素( $S_1, S_2, S_3, S_4$ )对云杉树苗生根密度的影响。可用于研究的树苗有 32 棵,而且每类激素被施给 8 棵树苗。按照上面概括的步骤,我们对 32 棵树苗从 1 到 32 编号。其次,我们得到数 1 到 32 的一个随机排列。利用软件包,我们得到下面一组这样的随机排列数。

|    |    |    |   |    |    |   |    |    |    |    |    |    |    |    |    |
|----|----|----|---|----|----|---|----|----|----|----|----|----|----|----|----|
| 31 | 21 | 4  | 7 | 28 | 3  | 1 | 24 | 17 | 15 | 13 | 29 | 9  | 22 | 20 | 27 |
| 25 | 12 | 14 | 5 | 18 | 26 | 8 | 30 | 23 | 32 | 11 | 19 | 16 | 6  | 10 | 2  |

这样我们就得到下面所示的树苗对激素的安排:

|    |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 树苗 | 31    | 21    | 4     | 7     | 28    | 3     | 1     | 24    | 17    | 15    | 13    | 29    | 9     | 22    | 20    | 27    |
| 激素 | $S_1$ | $S_1$ | $S_1$ | $S_1$ | $S_1$ | $S_1$ | $S_1$ | $S_1$ | $S_2$ | $S_2$ | $S_2$ | $S_2$ | $S_2$ | $S_2$ | $S_2$ | $S_2$ |
| 树苗 | 25    | 12    | 14    | 5     | 18    | 26    | 8     | 30    | 23    | 32    | 11    | 19    | 16    | 6     | 10    | 2     |
| 激素 | $S_3$ | $S_3$ | $S_3$ | $S_3$ | $S_3$ | $S_3$ | $S_3$ | $S_3$ | $S_4$ | $S_4$ | $S_4$ | $S_4$ | $S_4$ | $S_4$ | $S_4$ | $S_4$ |

图 14.3 显示了试验台上激素对树苗的安排。编号为 1 到 8 的树苗在试验台的第 I 象限, 编号为 9 到 16 的树苗在试验台的第 II 象限, 编号为 17 到 24 的树苗在试验台的第 III 象限, 编号为 25 到 32 的树苗在试验台的第 IV 象限。

|       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
| II    |       |       |       | I     |       |       |       |
| $S_2$ | $S_4$ | $S_4$ | $S_3$ | $S_1$ | $S_4$ | $S_1$ | $S_1$ |
| $S_2$ | $S_3$ | $S_2$ | $S_4$ | $S_3$ | $S_4$ | $S_1$ | $S_3$ |
| III   |       |       |       | IV    |       |       |       |
| $S_2$ | $S_3$ | $S_4$ | $S_2$ | $S_3$ | $S_3$ | $S_2$ | $S_1$ |
| $S_1$ | $S_2$ | $S_4$ | $S_1$ | $S_2$ | $S_3$ | $S_1$ | $S_4$ |

图 14.3 试验台上处理的随机安排

假定 32 个盆的这种布置是放在四张不同的台子上, 而不是简单地放在同一张台子的四个不同部位。台与台之间离得相当远, 研究人员看出四张台子所处的温度和光照条件不同。她考察了如图 14.3 中的安排, 并注意到 I 号桌子上的 8 棵树苗中, 有 4 棵被随机地分配激素 1。如果 I 号台所处的条件非常有利于根的生长, 那么这将使得试验产生偏差。另一个可以换用的给处理分配试验单元的方法称为**随机化完全区组设计**。在这种设计中, 试验单元不是齐性的。研究人员首先必须检查试验单元, 然后把它们放入各组, 使得同一组内的试验单元比不同组中的试验单元要相似些。如果每一组中有足够多的试验单元, 处理将在每一组范围内随机化。这种分配方法与在每一组内分别进行**完全随机化设计**相一致。

这种把处理分配给试验单元的方法, 其步骤如下。假定我们有  $t$  个处理和  $N$  个试验单元。进一步假定试验单元可以分成  $b$  组, 每组含有  $t$  个试验单元。这要求试验单元的个数  $N$  满足  $N = bt$ 。

#### 随机化完全区组设计的随机化步骤:

1. 把试验单元分成  $b$  组, 每组含  $t$  个齐性的试验单元。

2. 在第 1 组中,对试验单元从 1 到  $t$  编号。
3. 使用随机数表或计算机软件获得数 1 到  $t$  的一个随机排列。
4. 在第 1 组中,相应于随机排列中第一个数字的试验单元获得处理 1,相应于随机排列中第二个数字的试验单元获得处理 2,依此类推。
5. 对余下的各组(区组)重复 2~4 步的内容。

在以上的随机化完全区组设计中,每个区组只有每一处理的一次重复。我们可以把它推广到每个区组中每一处理有  $r$  次重复的情况。上述随机化步骤稍作修正同样适用于具有  $t$  个处理,  $b$  个区组,  $r$  次重复的推广了的随机化完全区组设计。这些试验单元将被分成  $b$  组,每组含有  $rt$  个齐性的试验单元。(注意  $N = brt$ )

#### 推广的随机完全区组设计的随机化步骤:

1. 把试验单元分成  $b$  组,每组含  $rt$  个齐性的试验单元。
  2. 在第 1 组中,对试验单元从 1 到  $rt$  编号。
  3. 使用随机数表或计算机软件获得数 1 到  $rt$  的一个随机排列。
  4. 在第 1 组中,相应于随机排列中前  $r$  个数字的试验单元获得处理 1,相应于随机排列中接下来的  $r$  个数字的试验单元获得处理 2。如此继续试验单元对处理的分配,直到处理  $t$  获得  $r$  个试验单元为止。
  5. 对余下的试验单元组重复 2~4 步的内容。
- 我们用下面的例子说明这一步骤。

#### 例 14.9

假定四张试验台在所处温度和阳光下的暴露量方面有相当大的差别。栽有树苗的 32 个盆中每 8 个盆被放在一个台上。我们面临的情况是,同一台上的两盆树苗所处的条件比不同台上的两盆树苗所处的条件更一致。这样,我们要进行一个随机化完全区组设计,其中 4 种激素中每一种都被随机地分配给每一台子上的其中两盆。注意这样我们就有一个  $N = 32$ ,  $b = 4$ ,  $r = 2$  的推广的随机化完全区组设计。

#### 解答

1. 32 个盆中每 8 盆被随机地放到一个台上。
2. 对每一台上的 8 个盆从 1~8 编号。
3. 我们得到如下数字 1~8 的一个随机排列:6, 7, 1, 4, 3, 8, 5, 2。
4. I 号台上的编号为 6 和 7 的盆被分以  $S_1$ , 1 号和 4 号被分以  $S_2$ , 3 号和 8 号被分以  $S_3$ , 5 号和 2 号被分以  $S_4$ 。
5. II 号台的随机排列为:1, 8, 2, 5, 4, 3, 7, 6;

Ⅲ号台的随机排列为:5, 7, 3, 8, 2, 4, 1, 6;

Ⅳ号台的随机排列为:1, 4, 5, 7, 3, 2, 8, 6。

6. Ⅱ号台:  $S_1$  分给编号为 1, 8 的盆;  $S_2$  分给编号为 2, 5 的盆;

$S_3$  分给编号为 4, 3 的盆;  $S_4$  分给编号为 7, 6 的盆。

Ⅲ号台:  $S_1$  分给编号为 5, 7 的盆;  $S_2$  分给编号为 3, 8 的盆;

$S_3$  分给编号为 2, 4 的盆;  $S_4$  分给编号为 1, 6 的盆。

Ⅳ号台:  $S_1$  分给编号为 1, 4 的盆;  $S_2$  分给编号为 5, 7 的盆;

$S_3$  分给编号为 3, 2 的盆;  $S_4$  分给编号为 8, 6 的盆。

我们得到如图 14.4 给出的随机安排。

|       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
| Ⅱ     |       |       |       | Ⅰ     |       |       |       |
| $S_3$ | $S_3$ | $S_2$ | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_2$ |
| $S_2$ | $S_4$ | $S_4$ | $S_1$ | $S_4$ | $S_1$ | $S_1$ | $S_3$ |
| Ⅲ     |       |       |       | Ⅳ     |       |       |       |
| $S_3$ | $S_2$ | $S_3$ | $S_4$ | $S_1$ | $S_3$ | $S_3$ | $S_1$ |
| $S_1$ | $S_4$ | $S_1$ | $S_2$ | $S_2$ | $S_4$ | $S_2$ | $S_4$ |

图 14.4 四张台子上处理的随机区组排列

在这一排列中, 四种处理的每一种被随机地分配给温室中每一台上的两个盆。现在, 如果四张台子上盆所处的条件之间存在较大的差异, 那么这一安排使得由台子所处环境造成的差异在 4 种处理上有更均衡的分布。此外, 每一台子上处理的随机化将有助于分散在施加处理之前树苗间存在的任何差异及每一张台子上盆间条件的差异。

## 14.6 确定重复试验的次数

试验中重复试验的次数在决定处理均值估计的精度和关于处理均值间差异的假设检验的功效方面起着关键的作用。在大多数情况下, 重复次数越大, 估计的精度越高, 处理均值的置信区间越精确, 假设检验的功效也越大。研究人员在使用非常大的重复观测次数时受到以下条件的约束: 进行试验的费用、处理大量试验单元所需要的时间及试验单元的可用性。因此, 研究人员必须确定为达到对估计精度

或假设检验功效的合理要求所需的最少重复观测次数。

### 通过给定估计精度来确定重复观测次数

我们可以通过指定处理均值的置信度为  $100(1 - \alpha)\%$  的置信区间想要达到的长度来确定重复观测的次数。在第五章中, 我们已经给出了一个公式, 用来确定保证样本估计落在真实处理均值  $\pm E$  个单位的范围内的概率不小于  $100(1 - \alpha)\%$  所需要的样本量。如果令  $r$  为重复观测的次数,  $\sigma$  是试验的标准差,  $E$  是希望达到的估计精度, 那么我们可以用下面的公式来计算  $r$  的近似值。

有  $100(1 - \alpha)\%$  的把握保证估计落在处理均值  $\mu$  的  $\pm E$  个单位内所需的样本量为

$$r = \frac{(z_{\alpha/2})^2 \hat{\sigma}^2}{E^2}$$

使用此公式时, 试验人员必须确定

1. 想得到的置信水平,  $100(1 - \alpha)\%$ 。
2. 精度水平,  $E$ 。
3.  $\sigma$  的一个估计。 $\sigma$  的估计可以从初步研究、过去类似的试验、有关类似的试验的文献或使用一个较粗糙的估计  $\hat{\sigma} = (\text{最大值} - \text{最小值})/4$  得到。

我们下面举例说明这些计算。

#### 例 14.10

某研究人员正在设计一项试验, 以研究在施用四种量的氮肥后山核桃的产量。她想得到处理均值  $\mu_1, \mu_2, \mu_3$  和  $\mu_4$  的估计, 并使其满足: 有 95% 的把握确信估计在真实产量均值  $\pm 4$  磅的范围内。她要确定为达到这一目标所必需的重复观测次数。

**解答** 从以往的试验知道, 产量的范围是 40 磅到 70 磅(原文误为 60 磅, 译者注)。这样, 可给出  $\sigma$  的一个估计

$$\hat{\sigma} = \frac{70 - 40}{4} = 7.5$$

从正态分布表查出  $z_{0.025} = 1.96$ 。如该研究人员所指定的那样,  $E$  的值为 4 磅, 由此我们确定重复观测的次数为

$$r = \frac{(z_{\alpha/2})^2 \hat{\sigma}^2}{E^2} = \frac{(1.96)^2 (7.5)^2}{(4)^2} = 13.51$$

因此, 为达到所希望的精度, 研究人员应在每一处理进行 14 次重复观测。



这个用来确定重复观测数的方法,没有考虑第八章中讨论的用来检测处理均值间的差异的  $F$  检验的功效。因而,在多数研究中,人们更喜欢使用下述确定重复观测次数的方法。

### 通过指定 $F$ 检验的功效来确定重复观测的次数

在包含  $t$  个处理的研究中,其中一个目标是检验假设

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_t$$

$H_a$ : 并非所有的  $\mu$  都相等

回想第八章我们给出的检验方法:

$$F = \frac{MST}{MSE}$$

如果  $F \geq F_{\alpha, t-1, N-t}$  则拒绝  $H_0$

此处  $MST$  和  $MSE$  是方差分析表中的均方。重复观测次数  $r_1 = r_2 = \cdots = r_t$  通过指定以下关于检验统计量的参数来确定:

1. 显著性水平,  $\alpha$ 。
2. 两个处理均值间差异  $D = |\mu_i - \mu_j|$  的大小,这个差异在实际中是显著的。
3. 当有任何一对处理均值间的差异大于  $D = |\mu_i - \mu_j|$  时,犯第 II 类错误的概率。
4. 方差  $\sigma^2$ 。

犯第 II 类错误的概率  $\beta(\lambda)$  由自由度为  $\nu_1, \nu_2$ , 非中心参数为

$$\lambda = \frac{r \sum_{i=1}^t (\mu_i - \mu_{..})^2}{\sigma^2}$$

的非中心  $F$  分布确定。对于至少有一对处理的均值相差  $D$  个单位或更大的情况,  $\lambda$  的最小值如下给出:

$$\lambda = \frac{rD^2}{2\sigma^2}$$

附录中的表 14 包含  $F$  检验的功效,它等于  $1 - \beta(\lambda)$ 。该表使用参数  $\phi = \sqrt{\lambda/t}$  而不用  $\mu_i$  的值。利用该表我们能确定为满足给定的具体要求所需要的重复观测次数。下例说明了需要进行的计算。

#### 例 14.11

回到例 14.10, 其中某研究人员正设计一项试验来研究在施用四种量的氮肥后山核桃的产量。研究人员知道,如果山核桃的平均产量间的差别超过 15 磅,那么选用相应于高产量的处理具有一定的经济意义。因此,该研究人员想确定必需的重复观测次数以保证:只要有任何一对氮肥施肥量对应的山核桃平均产量之间

相差 15 磅以上, 就有 90% 的把握确信  $F$  检验会拒绝  $H_0$ , 从而检验出平均产量间存在差异。该检验必须要求  $\alpha = 0.05$ 。

**解答** 从以往的试验知道, 山核桃的平均产量在 40 磅到 70 磅(原文误为 60 磅, 译者注)之间。因此, 给出  $\sigma$  的一个估计

$$\hat{\sigma} = \frac{70 - 40}{4} = 7.5$$

我们有

$$\alpha = 0.05, t = 4, \nu_1 = t - 1 = 4 - 1 = 3,$$

$$\nu_2 = N - t = rt - t = t(r - 1) = 4(r - 1),$$

此处  $r$  是所需的重复观测次数。另外,  $D$  的值为 15。所以

$$\phi = \sqrt{\frac{rD^2}{2t\hat{\sigma}^2}} = \sqrt{\frac{r(15)^2}{2(4)(7.5)^2}} = 0.707\sqrt{r}$$

图 14.5 包含求解这个问题所需要的功效曲线。注意  $\nu_1 = 3, \alpha = 0.05$ , 且曲线以  $\nu_2$  的值为标记。我们要确定  $r$  的值, 使得当  $\phi = 0.707\sqrt{r}$  时检验的功效至少是 0.9。

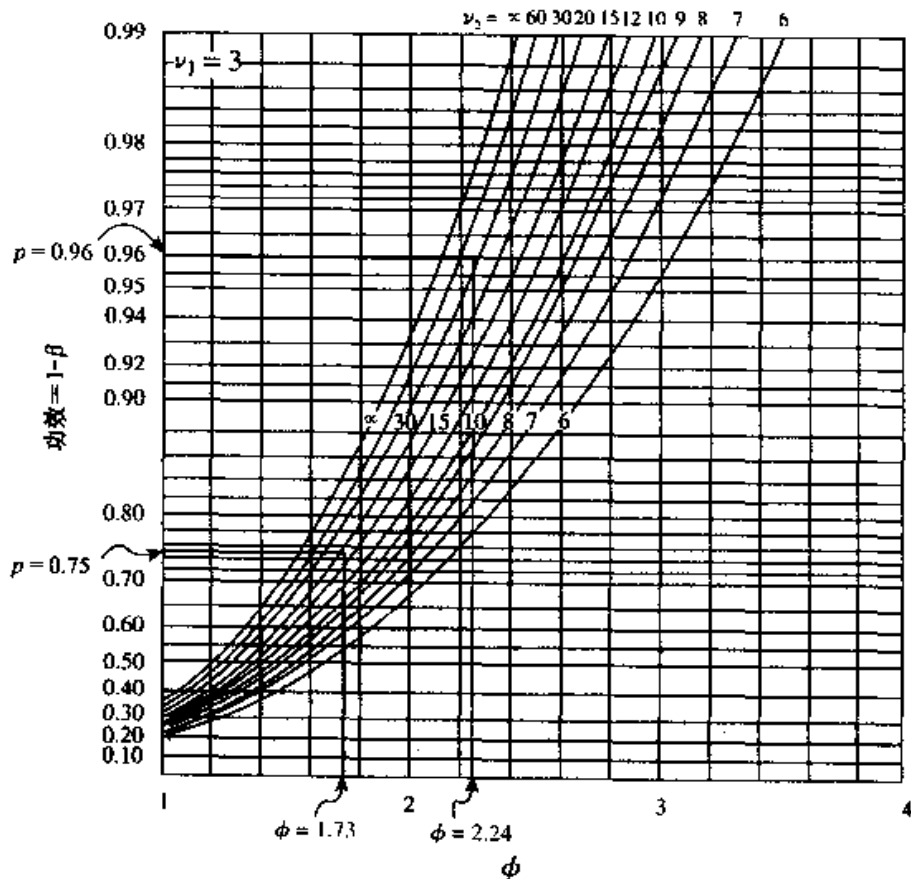


图 14.5 方差分析检验的功效( $\alpha = 0.05, t = 4$ )

为此,我们要不断地选择  $r$  的值,直至达到这个必需的阈值。

这种确定  $r$  的的方法是通过反复试验实现的。首先我们猜想  $r=6$ 。其次,我们计算得  $\nu_2=4(6-1)=3$ ,  $\phi=0.707\sqrt{6}=1.73$ 。在图 14.5 中,我们在  $\phi$  轴上确定  $\phi=1.73$  的点,然后从 1.73 这点向标记为 20 的曲线画一条竖直线,我们再在它们的交点处画一条水平线与标记为“功效  $=1-\beta$ ”的轴相交,读出交点处相应的值为 0.75。这样,如果我们在试验中进行 6 次重复观测,在  $D=15$  时,功效只有 0.75,这太低了。然后我们试用  $r=10$ ,发现功效为 0.96,该值可以接受;然而,或许再小一点的  $r$  值也可能会达到我们的要求。于是,我们试试  $r=8$ ,发现功效等于 0.89,此值稍稍小了一点点。最后,我们发现当  $r=9$  时,功效为 0.93。因此,为满足功效的要求,试验的每一处理需要 9 次重复观测。上述的计算过程概括如下:

在确定完重复观测次数后,实际情况可能不允许在同一时间或同一地点对这么多的试验单元进行完整的试验。在这样的情况下,我们可以利用随机化完全区组设计的思想,按时间或地点划分区组。在例 14.11 中,我们确定出需要对 4 个处理做 9 次重复观测,即共需要 36 个试验单元。假定我们在农业研究中心的一个地点只有 12 个试验单元,而该中心有三处这样的地点,每处都有 12 块试验地。于是,我们可以在每一个地点对每一处理进行三次重复观察,把地点作为该试验设计中的区组。

表 14.1 重复观测次数的确定

| $r$ | $\nu_2=4(r-1)$ | $\phi=0.707\sqrt{r}$ | 功效   |
|-----|----------------|----------------------|------|
| 6   | 20             | 1.73                 | 0.75 |
| 10  | 36             | 2.24                 | 0.96 |
| 8   | 28             | 2.00                 | 0.89 |
| 9   | 32             | 2.12                 | 0.93 |

## 14.7 小 结

在这一章中,我们考察了为研究人员提供信息的各类研究。特别地,我们介绍了几种用于受控试验的设计方法,讨论了可能影响试验单元的响应之变异的各种条件和变量。这是极端重要的问题,因为它会影响我们处理均值估计的精度和检验统计量的功效。通过运用适当的试验设计,试验误差的方差可以大大减小。在设计的试验中,重要的一点是随机化过程。试验单元被随机地分配给处理这一点至关重要。不使用随机化手段,我们将失去估计、置信区间及检验统计量的许多统计性质。本章最后部分介绍了在设计的试验中如何确定重复观测次数。

## 练习

**14.1(工程)** 研究人员做质量控制试验以评估灌溉用塑料管的质量。在这个试验设计中,共有 24 根管子,是从两个生产厂家每家随机抽取 12 根得到的。记录下每根管子五个位置处相应的抗压强度。管子是在两种水温之一和三类硬化剂之一的条件组合下生产出来的。试验条件如下:

| 管子号 | 生产厂家 | 温度(*F) | 硬化剂   | 管子号 | 生产厂家 | 温度(*F) | 硬化剂   |
|-----|------|--------|-------|-----|------|--------|-------|
| 1   | 1    | 200    | $H_1$ | 13  | 1    | 200    | $H_3$ |
| 2   | 1    | 175    | $H_2$ | 14  | 1    | 175    | $H_3$ |
| 3   | 2    | 200    | $H_1$ | 15  | 2    | 200    | $H_3$ |
| 4   | 2    | 175    | $H_2$ | 16  | 2    | 175    | $H_3$ |
| 5   | 1    | 200    | $H_1$ | 17  | 1    | 200    | $H_2$ |
| 6   | 1    | 175    | $H_2$ | 18  | 1    | 175    | $H_1$ |
| 7   | 2    | 200    | $H_1$ | 19  | 2    | 200    | $H_2$ |
| 8   | 2    | 175    | $H_2$ | 20  | 2    | 175    | $H_1$ |
| 9   | 1    | 200    | $H_3$ | 21  | 1    | 200    | $H_2$ |
| 10  | 1    | 175    | $H_3$ | 22  | 1    | 175    | $H_1$ |
| 11  | 2    | 200    | $H_3$ | 23  | 2    | 200    | $H_2$ |
| 12  | 2    | 175    | $H_3$ | 24  | 2    | 175    | $H_1$ |

确定该试验设计中的如下内容。

- 因子。
- 因子水平。
- 区组。
- 试验单元。
- 测量单元。
- 重复观测次数。
- 协变量。
- 处理。

**14.2** 考虑一个你可能要做的研究或试验,该试验用于回答你所研究的领域中的研究问题。陈述这一研究问题并说明它为什么重要,叙述你将如何进行试验以获得数据来回答你的问题。在你的描述中要包含以下内容:

- 研究问题。
- 对试验单元的描述。
- 对测量单元的描述。
- 处理设计。

- 设计结构。
- 随机化方法。
- 重复观测次数。
- 可能的区组划分。

**14.3** 对在下面描述的每一试验,确定设计的重要特征,像在练习 14.1 中那样,尽可能多地包含各项内容以充分地描述该项设计。

- 一位园艺学家要测量位于南德克萨斯的某研究农场的一个果园中的橙子内维生素 C 的含量。他对果园中不同地点、不同的产果月份以及每棵树上的橙子之间维生素 C 含量的变异感兴趣。他把果园分成 8 个部分,并随机地从每一部分选了一棵树。从 10 月到来年 5 月期间,正好是果树结果的时候,他每月从 8 棵树的三个不同位置,即顶部、中间、下边分别摘取 10 个橙子。园艺学家想监测产果季节中橙子维生素 C 的含量,并确定长在树的不同位置的橙子维生素 C 的含量是否存在比较大的差异。
- 某医学专家想比较两种治疗某种疾病的不同疗法( $T_1$ ,  $T_2$ )。她要在八个医院进行试验,并认为响应变量在医院之间可能有差异。每个医院有 4 个监护区,她在每个监护区随机选出 4 个患者参与研究。在每一医院内,随机地把  $T_1$  分配给某两个监护区,其余两个接受疗法  $T_2$ 。一个监护区中的所有患者得到相同的治疗。对每一个病人测量了单个的响应变量。
- 在(b)描述的设计中,作如下变动。在每一医院中,两种疗法随机地分配给患者,即要求每一监护区中的四个患者中,有两个获得  $T_1$ ,另两个获得  $T_2$ 。
- 计划一个试验,用来比较三类学校,即公立、私立且非教区附属的、教区附属的学校六年级学生的阅读能力。研究人员在美国五个地区中的每个地区选取两个大城市进行研究。在每一城市,对每种类型的学校她都随机地选择了一所,并在选取的学校中随机选择一个六年级的班,记录下每一班级 20 名学生标准化考试的成绩。研究人员还关注 30 个学校间家庭收入水平的差异可能对阅读能力的影响,因此她还得到了参与研究的每一学生的家庭收入数据。

**14.4(兽医)** 一试验被设计用来评价不同水平的训练对狗的健康的影响。两个训练水平是  $L_1$ —每天跑 2 英里,  $L_2$ —每两天跑 1 英里。在 3 个月的研究期结束时,将对每只狗进行呼吸和心血管的健康状况测试,从中计算出一个健康指数。有 16 只狗可用来做研究,它们的健康状况良好且个头基本相同,就饲养情况来说这个个头属于正常范围。下表记录了 16 只狗的有关性别、年龄的信息。

| 狗 | 性别 | 年龄 | 狗  | 性别 | 年龄 |
|---|----|----|----|----|----|
| 1 | F  | 5  | 9  | F  | 8  |
| 2 | F  | 3  | 10 | F  | 9  |
| 3 | M  | 4  | 11 | F  | 6  |
| 4 | M  | 7  | 12 | M  | 8  |
| 5 | M  | 2  | 13 | F  | 2  |
| 6 | M  | 3  | 14 | F  | 1  |
| 7 | F  | 5  | 15 | M  | 6  |
| 8 | M  | 9  | 16 | M  | 3  |

- 为了使研究的误差方差尽可能小,在分配处理之前你如何对这些狗进行分组? 列出每一组中相应的狗。
- 使用随机数发生器或随机数表得到一组随机数。描述你给每个狗分配处理的方法。

**14.5(商业)** 某计算机杂志要对用来编制联邦年度个人所得税表的四种软件,根据他们完成表格所用的时间来进行排名。该研究将选取收入少于 100,000 美元的个人,并详细列举他们的个人所得税的扣除额。试确定:为使当任意一对均值之间的差异大于 30 分钟时,以显著水平  $\alpha = 0.05$ 、检验功效为 0.9 检验出平均完成时间存在差异,每一软件程序需要多少个人。从以往对使用类似软件的研究中知道,完成时间的标准差可以认为在 12.25 分钟左右。

**14.6** 要做一个试验来分析六种处理的差异。试确定:为使当任意一对处理均值之间的差异超过 20 个单位时,在显著水平 0.05、检验功效为 0.8 下检验出处理均值之间的差异,需要有多少试验单元。从以往的研究知道,响应变量的标准差近似为 9 个单位。

**14.7(生物)** 一家大海产食品公司的研究专家计划调查在三种不同的储藏温度下,牡蛎和贻贝的细菌生长情况。研究可利用 9 个冷藏单元。她计划给每 3 个冷藏单元分配一种冷藏温度,在每一冷藏单元中放一袋牡蛎和一袋贻贝,储藏两周时间。在两周的储藏期结束时,将把袋取出并从每一袋中抽取两个样品测量它们的细菌数。我们所关心的“处理”因子是温度(水平为:0, 5℃, 10℃)和海产品(水平为:牡蛎,贻贝)。她还会记录下在放入冷藏柜之前每袋海产品含有的细菌数。确定下面有关试验设计的每一项内容。

- 因子。
- 因子的水平。
- 区组。

- d. 试验单元。
- e. 测量单元。
- f. 重复测量的次数。
- g. 协变量。
- h. 处理。

**14.8** 考虑练习 14.7。如果仅使用三个冷藏单元, 每个冷藏温度一个冷藏单元, 在每一温度下的那个储藏单元中, 对每一种海产品抽取三个样本, 那么试验人员仍得到同样数目的观察值。以这样的方式进行试验, 其潜在的困难是什么?

**14.9(商业)** 比较四种蛋糕配方的嫩润程度。研究人员在进行试验时, 先配制用料, 然后烘烤蛋糕。对于每一种配方, 一次配制只做一个蛋糕。对所有的配方使用同样的烘烤温度和烘烤时间。电烤箱足够大, 在任何一个烘烤周期内, 都可以在烤箱的  $P_1$  到  $P_4$  的位置同时烘烤四个蛋糕。四个位置如下所示:

|       |       |
|-------|-------|
| $P_1$ | $P_2$ |
| $P_3$ | $P_4$ |

- a. 如果对每一种配方制作  $r$  个蛋糕, 讨论适当的试验设计和随机化方案。
- b. 假定试验人员担心烤箱中四个不同的位置(前面对后面, 左边对右边)会使蛋糕有显著的差别, 那么你的设计还合适吗? 如果不合适, 给出一个适当的设计。
- c. 如果要对五种配方进行考察, 但每一次只能同时烤四个蛋糕, 对(b)中所叙述的设计提出修改意见。

**14.10** 对下面的每种情况, 确定设计是完全随机化设计、随机化完全区组设计还是拉丁方设计。如果有处理的因子结构, 指出它是两因子结构还是三因子结构。如果试验单元不同于测量单元, 把两者都指出来。

- a. 由 N(氮)肥的 3 个水平、P(磷)肥的 4 个水平、K(钾)肥的 4 个水平的所有可能的水平组合组成 48 个处理。随机地抽选了五个种植花生的农场, 然后在每一个农场把 48 个处理随机地分配给其中的 48 块花生试验田。
- b. 十个不同的软件包被随机地分配给 30 名研究生, 确定出了他们完成指定任务所花的时间。
- c. 四种不同的釉料以两种不同的厚度被涂到陶罐上。用来上釉的炉能同时放置八个罐, 上釉需要一天的时间。试验人员想得到试验的八次重复测量。因为窑中的条件每天会有一些变化, 试验时间用了八天。在每一天中, 涂层厚度和釉料种类的每一组合被随机地用于窑中的一个陶罐。

**14.11** 你的一个同事就她正在做的一个试验来求助于你。在该试验中, 请

一些消费者来品尝五种不同配方的肉食品。当一个顾客品尝了一个样品后,他或她会就一些特征为该样品打分,然后这些特征分将合并一个总分。因而,一个消费者对每一种配方都给出一个总评分。有关文献指出在这类试验中,一些消费者倾向于对所有的样品都给低分,其他的消费者倾向于给所有的样品以高分。

- a. 这儿有两种可能的试验设计方法。设计 A 将抽取 100 个顾客组成样本,在该组中每 20 位顾客被随机地分配五种配方中的任意一种,这样每位顾客只品尝一种配方。设计 B 将使用由 100 个顾客组成的样本,其中的每一位顾客品尝所有的五种配方,五种配方的食品将按随机的次序放在每一位顾客面前。你将推荐使用哪一种?说明作出这种回答的理由。
- b. 当研究人员被问及试验进展如何时,她回答说有一个配方闻起来是如此的糟糕,以至于把它给排除在分析之外了。如果选用设计 B,那这对于分析来说成问题吗?为什么?如果选用设计 A 呢?为什么?



## 第十五章 标准设计的方差分析

- 15.1 引言和案例
- 15.2 单因子的完全随机化设计
- 15.3 随机化完全区组设计
- 15.4 拉丁方设计
- 15.5 完全随机化设计中的因子处理结构
- 15.6 随机化完全区组设计中的因子处理结构
- 15.7 处理差异的估计和处理均值的比较
- 15.8 小结

### 15.1 引言和案例

在第十四章中,我们介绍了试验设计中的一些概念,也谈到进行科学研究和试验的过程。在这个科学过程中,研究人员整理提出假设,对试验(研究)进行计划,收集并分析数据,然后得到结论,而这些结论转而又导致新的假设的提出。试验设计中的这些概念对于这样的科学过程是基本的。毋庸置疑,为了从试验(研究)得到符合逻辑的结论,必须精确清楚地提出假设,认真仔细地设计试验、实施试验,恰当地分析试验。对一个设计的试验的分析,需要针对该试验建立一个模型,并清楚地说明这个模型适用的条件。最后,试验结果的科学报告应当包含数据的图表,模型条件的验证,统计分析的总结和关于所提出的研究假设的结论。在这一章,我们讨论一些标准的试验设计及其分析。

15.2 节回顾了第八章中完全随机化设计的方差分析。这里主要讨论均值的比较。15.3 节和 15.4 节对完全随机化设计进行扩充,重点仍然放在处理均值的比较上,但在这些设计中,必须对“讨厌”变量进行控制。对于这种类型的设计,我们将考虑其处理的安排,讨论其优缺点,并给出一个模型和对所得数据的方差分析。15.5 节介绍因子设计。因子设计主要考虑两个或更多自变量(因子)对于响应的影响,而不像 15.2 节至 15.4 节中的设计那样去比较处理均值;尤其注意评估每个因子单独的效应,以及在与其它因子的组合中的作用。并非所有的设计都比较处理均值或考察因子对响应的效应。在 15.6 节中,我们讨论的设计结合了 15.3 节和 15.4 节中的区组设计和 15.5 节中的因子设计的特点。本章其余的内容包括这些试验设计中处理均值的估计和比较,验证模型条件有效性的方法,以及

当标准模型的条件不满足时,可供选择的其它方法。

### 案例:低脂肪大红肠的配制

消费者对饮食健康的关心和对低脂肪食品的要求已经推动了肉制品公司开发各种低脂肪的肉食制品。为了在保证维持产量,极小化制作成本并保持味道鲜美的条件下寻找脂肪的替代品,已经对很多成分进行了考察。发表在 *Meat Science*, 53:47~57 的论文“Utilization of soy protein isolate and konjac blends in a low-fat bologna (model system)”(1999)中描述的试验考虑了此类问题。研究者认为,为了在不影响低脂肪肉制品质量的前提下降低生产成本,要用非肉质配料,如大豆蛋白分离物(SPI)代替其中的一部分碎肉块。以前的试验虽然已经表明了 SPI 在碎肉特征上的效果,但关于在低脂肪肉制品中的应用效果的研究不多。另外,为了在降低脂肪含量的同时改善凝胶和容水性能,已经有人把魔芋粉用于肉制品生产中。因此,当把肉换为 SPI 时,有必要掺入魔芋粉以保持高脂肪制品的特点。

**数据收集的设计** 在该项研究中,要考察的因子有三个:在肉制品中掺入魔芋混合物的类型,魔芋的掺入量以及 SPI 替换的百分比。还有许多其他感兴趣的可能的因子,包括烹调时间、温度、肉制品的类型、熏制长度等。然而,研究人员选取了这些因子在大红肠的商业制作中的常用水平,从而把研究限制在以上三个最重要的因子。从而得到一个具有表 15.1 给出的 12 个处理的试验。

表 15.1 低脂肪大红肠研究的处理设计

| 处理 | 混合物水平<br>(%) | 魔芋混合物 | SPI<br>(%) |
|----|--------------|-------|------------|
| 1  | 0.5          | KSS   | 1.1        |
| 2  | 0.5          | KSS   | 2.2        |
| 3  | 0.5          | KSS   | 4.4        |
| 4  | 0.5          | KNC   | 1.1        |
| 5  | 0.5          | KNC   | 2.2        |
| 6  | 0.5          | KNC   | 4.4        |
| 7  | 1            | KSS   | 1.1        |
| 8  | 1            | KSS   | 2.2        |
| 9  | 1            | KSS   | 4.4        |
| 10 | 1            | KNC   | 1.1        |
| 11 | 1            | KNC   | 2.2        |
| 12 | 1            | KNC   | 4.4        |

此项研究的目的是评估各种作为部分瘦肉替代品的魔芋混合物,并用一个低脂肪大红肠模型刻画这些混合物的作用。评估时使用了两种魔芋混合物(KSS=魔芋粉/淀粉和KNC=魔芋粉/角叉胶/淀粉),它们的水平取为0.5%和1%,还使用了SPI替换肉蛋白的三个水平(1.1%,2.2%和4.4%,DWB)。

该试验是一个完全随机化设计,具有一个 $2 \times 2 \times 3$ 的三因子处理结构,12个处理,每个处理3次重复。在这36次试验中,测量了许多响应变量,但我们仅讨论对最后制品的组织进行测量得到的结果。这些结果是由一台Instron通用检测仪(一种可用来测量组织力学特性的仪器——译者注)测量得来的。平均响应如表15.2所示。

**数据的整理** 研究人员接下来按照2.5节中给出的步骤为统计分析准备数据。他们需要验证产品结构的质地指标是否正确地得到了记录,所有计算机文件是否与现场数据相吻合。(表15.2中的数据是用论文中给出的概括统计量模拟出来的)

表 15.2 低脂肪大红肠研究中的肉质地指标的均值

| 魔芋水平<br>(%) | 魔芋混合物 | SPI<br>(%) | 组织指标                | 组织指标<br>平均值 |
|-------------|-------|------------|---------------------|-------------|
| 0.5         | KSS   | 1.1        | 107.3, 110.1, 112.6 | 110.0       |
| 0.5         | KSS   | 2.2        | 97.9, 100.1, 102.0  | 100.0       |
| 0.5         | KSS   | 4.4        | 86.8, 88.1, 89.1    | 88.0        |
| 0.5         | KNC   | 1.1        | 108.1, 110.1, 111.8 | 110.0       |
| 0.5         | KNC   | 2.2        | 108.6, 110.2, 111.2 | 110.0       |
| 0.5         | KNC   | 4.4        | 95.0, 95.4, 95.5    | 95.3        |
| 1           | KSS   | 1.1        | 97.3, 99.1, 100.6   | 99.0        |
| 1           | KSS   | 2.2        | 92.8, 94.6, 96.7    | 94.7        |
| 1           | KSS   | 4.4        | 86.8, 88.1, 89.1    | 88.0        |
| 1           | KNC   | 1.1        | 94.1, 96.1, 97.8    | 96.0        |
| 1           | KNC   | 2.2        | 95.7, 97.6, 99.8    | 97.7        |
| 1           | KNC   | 4.4        | 90.2, 92.1, 93.7    | 92.0        |

**数据的分析** 研究人员的兴趣在于评估低脂肪大红肠的组织指标与SPI的百分比增长之间的关系,并对两种魔芋混合物在两种水平下的这种关系进行比较。我们将在15.5节末尾讨论这个例子中数据的分析。

## 15.2 单因子的完全随机化设计

我们记得,完全随机化设计用来比较  $t$  个总体(处理)均值  $\mu_1, \mu_2, \dots, \mu_t$ 。假定有  $t$  个不同的总体,我们从中分别抽取容量为  $n_1, n_2, \dots, n_t$  的样本。用试验设计的术语来说,我们假定有  $n_1 + n_2 + \dots + n_t$  个齐性的试验单元(对之进行测量的人或物体),把处理随机地分配到试验单元,使得有  $n_1$  个单元接受处理 1,  $n_2$  个单元接受处理 2, 等等。试验的目标是对这些相应的处理(总体)均值进行推断。

考虑由表 15.3 给出的完全随机化设计的数据。

表 15.3 完全随机化设计

| 处理    |          |          |       |            | 均值          |
|-------|----------|----------|-------|------------|-------------|
| 1     | $y_{11}$ | $y_{12}$ | ..... | $y_{1n_1}$ | $\bar{y}_1$ |
| 2     | $y_{21}$ | $y_{22}$ | ..... | $y_{2n_2}$ | $\bar{y}_2$ |
| ..... | .....    | .....    | ..... | .....      | .....       |
| $t$   | $y_{t1}$ | $y_{t2}$ | ..... | $y_{tn_t}$ | $\bar{y}_t$ |

对于有  $t$  个处理, 每个处理有  $n_i$  个观测值的完全随机化设计, 其模型可以写为如下形式

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

此处, 模型中各项的定义为:

$y_{ij}$ : 接受处理  $i$  的第  $j$  个试验单元上的观测值;

$\mu$ : 总的处理均值, 未知常数;

$\alpha_i$ : 处理  $i$  的效应, 未知常数;

$\epsilon_{ij}$ : 接受处理  $i$  的第  $j$  个试验单元的响应的随机误差。我们要求各个  $\epsilon_{ij}$  服从均值为 0, 共同方差为  $\sigma_\epsilon^2$  的正态分布。此外, 误差之间必须是相互独立的。

从模型的上述条件可以看出, 由第  $i$  个处理记录得到的第  $j$  个响应  $y_{ij}$  服从均值为  $\mu + \alpha_i$ 、方差为  $\sigma_\epsilon^2$  的正态分布。各处理均值之间的差别在于  $\alpha_i$ , 即它们所对应的处理效应。因此, 检验

$$H_0: \mu_1 = \mu_2 = \dots = \mu_t \quad H_a: \text{并非所有 } \mu_i \text{ 都相等}$$

等价于检验

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_t = 0 \quad H_a: \text{并非所有 } \alpha_i \text{ 都为 0}$$

检验统计量是用对测量值关于其平均  $\bar{y} \dots = \sum_{ij} y_{ij}$  的总平方和进行分解的思想

得到的。在第八章, 我们曾给出这一总平方和的定义:

$$\text{TSS} = \sum_{ij} (y_{ij} - \bar{y} \dots)^2$$

我们把总平方和分解为两个不同的变差来源: 一个是处理间的变异, 一个是每个处理内部各个  $y_{ij}$  之间的变异。这第二个变异的来源称为“误差”, 因为它表示不能被处理之间的差别所解释的那部分变异。

可以证明, TSS 的分解具有形式

$$\sum_{ij} (y_{ij} - \bar{y} \dots)^2 = \sum_i n_i (\bar{y}_{i.} - \bar{y} \dots)^2 + \sum_{ij} (y_{ij} - \bar{y}_{i.})^2$$

当各个处理的重复数相同, 即  $n_1 = n_2 = \dots = n_t = n$  时, 上述分解成为

$$\sum_{ij} (y_{ij} - \bar{y} \dots)^2 = n \sum_i (\bar{y}_{i.} - \bar{y} \dots)^2 + \sum_{ij} (y_{ij} - \bar{y}_{i.})^2$$

等号右边的第一项表示各处理均值  $\bar{y}_{i.}$  关于总均值  $\bar{y} \dots$  的变异性的 大小, 因此称为**处理间平方和**(SST), 它是各  $y_{ij}$  间源于处理各均值  $\mu_i$  之间的差异的变异性的一个度量。处理间平方和由

$$\text{SST} = n \sum_i (\bar{y}_{i.} - \bar{y} \dots)^2$$

给出。第二项称为**误差平方和**, 它表示  $y_{ij}$  之间未被处理均值间的差异所解释的变异性。这个变异性表示在实施处理之前各试验单元之间的差异以及试验期间各试验单元所暴露的条件之间的差异。误差平方和由

$$\text{SSE} = \sum_{ij} (y_{ij} - \bar{y}_{i.})^2$$

给出。回想第八章, 我们曾经把这些分析总结在如下的方差分析表中。

表 15.4 完全随机化设计的方差分析表

| 来源 | SS  | df      | MS                                  | F                         |
|----|-----|---------|-------------------------------------|---------------------------|
| 处理 | SST | $t - 1$ | $\text{MST} = \text{SST} / (t - 1)$ | $\text{MST} / \text{MSE}$ |
| 误差 | SSE | $N - t$ | $\text{MSE} = \text{SSE} / (N - t)$ |                           |
| 总和 | TSS | $N - 1$ |                                     |                           |

当  $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_t = 0$  为真时, MST 和 MSE 均为试验误差方差  $\sigma_e^2$  的无偏估计, 也就是说, 当  $H_0$  成立时, MST 和 MSE 在大量重复抽样时的均值都等于  $\sigma_e^2$ 。这一均值称为**期望均方**。这一事实表示为

$$E(\text{MST}) = \sigma_e^2 \quad \text{和} \quad E(\text{MSE}) = \sigma_e^2$$

由此, 我们可以期望当  $H_0$  为真时,  $F = \text{MST} / \text{MSE}$  的值接近于 1。当  $H_a$  为真, 并且各处理均值之间存在差异时, MSE 仍然是  $\sigma_e^2$  的一个无偏估计

$$E(\text{MSE}) = \sigma_e^2$$

然而, MST 此时不再是  $\sigma_e^2$  的无偏估计。事实上, 可以证明处理的期望均方为

$$E(\text{MST}) = \sigma_e^2 + n\theta_T$$

其中  $\theta_T = 1/(t-1) \sum_i \alpha_i^2$ 。当  $H_a$  为真时, 某些  $\alpha_i$  不为 0, 从而  $\theta_T$  大于 0。因此, MST 有过高估计  $\sigma_e^2$  的倾向。这样, 比值  $F = \text{MST}/\text{MSE}$  将倾向于比 1 大, 于是我们在  $F$  分布的右侧尾部拒绝  $H_0$ 。

具体说来, 对于选定的犯第一类错误的概率  $\alpha$ , 当计算出来的  $F$  值大于  $F_{\alpha, t-1, N-t}$ , 这里  $F_{\alpha, t-1, N-t}$  为用  $\alpha = \alpha$ ,  $df_1 = t-1$ ,  $df_2 = N-t$  在附录表 8 中查出的  $F$  分布的临界值。注意  $df_1$  和  $df_2$  分别相应于 AOV 表中 MST 和 MSE 的自由度。

当完全随机化设计用作比较  $t$  个处理均值的试验设计时, 有如下的长处和不足。

#### 完全随机化设计的长处和不足

##### 长处

1. 该设计非常容易构造;
2. 该设计易于分析, 即使在各个处理上样本容量不同时也是如此;
3. 该设计可被用于有任意个处理的试验。

##### 不足

1. 虽然完全随机化设计可以用于任意个处理的情形, 但最好处理个数比较小;
2. 对之实施处理的各个试验单元必须尽可能一致。任何外部的变异来源都会引起误差项的膨胀, 从而使得对处理均值的差异的探测更加困难。

#### 例 15.1

乡间道路路况安全的一个重要因子是在公路上用反射涂料标出行车线。这些行车线为司机在很少或没有夜间照明的路上行车提供了参照线路。目前使用的涂料的问题是长时间维持其反射性能。一位研究人员进行一项试验, 把三种新型涂料( $P_2, P_3, P_4$ )与现在使用的涂料( $P_1$ )进行比较。这些涂料被涂在 6 英尺长的路段上, 响应变量是使用 6 个月以后这些标记的反射性能下降的百分比。试验中共有 16 个路段, 每种涂料被随机地涂在 4 个路段上。这些涂料在公路上使用 6 个月以后, 计算出每个路段反射性能的下降幅度, 所得测量值如表 15.5 所示。

表 15.5 反射性能测量值

| 路段       | 1  | 2  | 3  | 4  | 均值    |
|----------|----|----|----|----|-------|
| 涂料 $P_1$ | 28 | 35 | 27 | 21 | 27.75 |
| $P_2$    | 21 | 36 | 25 | 18 | 25    |
| $P_3$    | 26 | 38 | 27 | 17 | 27    |
| $P_4$    | 16 | 25 | 22 | 18 | 20.25 |

看起来  $P_4$  比其他 3 种涂料能在更长的时间内维持其反射性能, 因为它的反射性能下降最少。我们现在通过检验假设

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad H_a: \text{并非所有的 } \mu_i \text{ 都相等}$$

来确认这一点。我们要用前面给出的公式计算各个平方和, 并由此建立方差分析表:

$$\bar{y}_{..} = \frac{y_{..}}{N} = \frac{400}{16} = 25$$

$$\begin{aligned} \text{TSS} &= \sum_{ij} (y_{ij} - \bar{y}_{..})^2 \\ &= (28 - 25)^2 + (35 - 25)^2 + \cdots + (22 - 25)^2 + (18 - 25)^2 = 692 \end{aligned}$$

$$\begin{aligned} \text{SST} &= n \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &= 4[(27.75 - 25)^2 + (25 - 25)^2 + (27 - 25)^2 + (20.25 - 25)^2] \\ &= 136.5 \end{aligned}$$

$$\text{SSE} = \text{TSS} - \text{SST} = 692 - 136.5 = 555.5$$

现在, 我们可以得到如下的方差分析表:

| 来源 | SS    | df | MS     | F    | p 值    |
|----|-------|----|--------|------|--------|
| 处理 | 136.5 | 3  | 45.5   | 0.98 | 0.4346 |
| 误差 | 555.5 | 12 | 46.292 |      |        |
| 总和 | 692   | 15 |        |      |        |

由于  $p \text{ 值} = 0.4346 > 0.05 = \alpha$ , 我们不能拒绝  $H_0$ 。四种涂料在反射性能下降的平均值上没有显著差异。

研究人员对例 15.5 中给出的研究结果不大放心, 因为他确信至少有一种涂料比目前使用的涂料有所改进。他考察了在这项研究中所使用的 16 个路段的路况

以及交通流量,发现这些道路在该项研究期间的交通流量很低。他决定重新设计这个试验,以使研究结果更具有一般性。新一轮试验选取了4个不同的地点,这几个地点有不同的交通流量。在15.3节我们将讨论如何进行这项试验,届时我们会发现出现了第二个变异的来源,即路段的位置。

### 15.3 随机化完全区组设计

我们现在修改上面反射涂料的试验,把四个不同的位置考虑进去。研究人员在每个位置上选定4个6英尺长的路段。如果我们随机地把4种涂料分配给这16个路段,可能得到如表15.6所列出的随机化方案。

表 15.6 4种涂料对16个路段的随机分配

| 位 置   |       |       |       |
|-------|-------|-------|-------|
| 1     | 2     | 3     | 4     |
| $P_1$ | $P_2$ | $P_3$ | $P_4$ |
| $P_1$ | $P_2$ | $P_3$ | $P_4$ |
| $P_1$ | $P_2$ | $P_3$ | $P_4$ |
| $P_1$ | $P_2$ | $P_3$ | $P_4$ |

虽然在该设计中对每一个处理仍然有4个观测值,但我们所能看到的涂有这4种涂料的道路反射性之间的任何差别,都有可能完全归结于四个位置上道路条件和交通流量的差别。由于位置和涂料类型这两个因子是相互混杂的,我们无法确定道路标记的反射性能下降的差异是由于标记的位置不同造成的还是由于做这些标记所使用的涂料不同所造成的。这说明16个道路标记受到一个外部变异来源的影响,即道路标记的位置。如果四个位置有不同的环境条件或不同的交通流量,那么这16个试验单元就不是齐性的,从而也就不能基于他们来评估4个处理,即4种涂料的效应。

上述所说的完全随机化设计不适用于这里的试验问题。为了考虑在指定处理之前试验单元之间的差别,我们需要用随机化完全区组设计。在第十四章,我们谈到如何对为试验单元分配处理的随机化进行限制,使得在每个位置上,对于4种涂料的每一种,都有一个路段使用这种涂料。表15.7给出了这样的随机化结果。注意,每个位置包含4个路段,这4个路段分别涂有不同的涂料。因而,由于四个位置上的道路条件不同而产生的变异就可以得到评估和控制。这样,我们就可以用摆脱了位置之间的变异性的样本均值来进行4种涂料之间的两两比较。例



如,如果我们做检验

$$H_0: \mu_{P_1} - \mu_{P_2} = 0$$

$$H_a: \mu_{P_1} - \mu_{P_2} \neq 0$$

并且拒绝了  $H_0$ , 则  $\mu_{P_1}, \mu_{P_2}$  之间的差异就是由两种涂料的反射性之间的差异造成的, 而不是由于位置的差异造成的, 因为在四个位置中的每一个上, 都有一个路段使用了  $P_1$  和  $P_2$ 。

表 15.7 把 4 种涂料指定给 16 个路段的随机化完全区组分配

| 位 置   |       |       |       |
|-------|-------|-------|-------|
| 1     | 2     | 3     | 4     |
| $P_2$ | $P_2$ | $P_1$ | $P_1$ |
| $P_1$ | $P_4$ | $P_3$ | $P_2$ |
| $P_3$ | $P_1$ | $P_4$ | $P_4$ |
| $P_4$ | $P_3$ | $P_2$ | $P_3$ |

在第十四章, 我们曾经讨论了如何分别在每个区组中把处理随机地分配给试验单元。在本例中, 区组就是道路的位置。在同一个位置上的 4 个路段, 比起在不同位置上的路段来, 在环境条件和交通流量方面可能更相似些。由此, 我们实质上是在进行四个独立的完全随机化试验, 每个位置一个。应用随机化完全区组设计, 我们已经有效地滤除了位置之间的变异, 从而能够在处理均值  $\mu_{P_1}, \mu_{P_2}, \mu_{P_3}, \mu_{P_4}$  之间作出更精确的比较。

一般说来, 当有外部的变异来源(区组)存在时, 我们可以使用随机化完全区组设计来比较  $t$  个处理均值。如果有  $b$  个不同的区组, 为了滤除区组之间的变异性, 我们在每个区组内为  $t$  个处理中的每个处理分配一个试验单元。在上述例子中, 我们有  $t=4$  处理(涂料的类型)和  $b=4$  个区组(位置)。

我们可以给出随机化完全区组设计的定义如下:

**定义 15.1** 随机化完全区组设计是用来在  $b$  个区组中比较  $t$  个处理的试验设计, 其中每个区组包含  $t$  个齐性的试验单元。在区组内随机地把处理分配给试验单元, 在每个区组中, 每个处理只出现一次。

随机化完全区组设计具有一些长处和不足, 列出如下。

#### 随机化完全区组设计的长处与不足

##### 长处

1. 该设计对于有单个外部变异来源的情形, 可用于  $t$  个处理均值的比较;
2. 其统计分析简单;

3. 设计容易构造;
4. 可以用在有任何个区组和任意个处理的情况。

#### 不足

1. 由于区组内的单元必须是齐性的,设计最好用在处理个数相当小的情形;
2. 该设计只控制一个外部变异来源(即源于区组的变异)。另外的外部变异来源将倾向于增加误差项,使得对处理差异的探索变得更加困难;
3. 每个处理对于响应的效应在各个区组中必须是相同的。

考虑表 15.8 中来自于随机化完全区组设计的数据。注意,虽然这些数据看上去与完全随机化设计中的数据(见表 15.3)相似,但与试验单元分配处理的方法是有差别的。

表 15.8 随机化完全区组设计的数据

| 处理  | 区组             |                |     |                | 均值             |
|-----|----------------|----------------|-----|----------------|----------------|
|     | 1              | 2              | ... | $b$            |                |
| 1   | $y_{11}$       | $y_{12}$       | ... | $y_{1b}$       | $\bar{y}_{.1}$ |
| 2   | $y_{21}$       | $y_{22}$       | ... | $y_{2b}$       | $\bar{y}_{.2}$ |
| ... | ...            | ...            | ... | ...            | ...            |
| $t$ | $y_{t1}$       | $y_{t2}$       | ... | $y_{tb}$       | $\bar{y}_{.t}$ |
| 均值  | $\bar{y}_{.1}$ | $\bar{y}_{.2}$ | ... | $\bar{y}_{.b}$ | $\bar{y}_{..}$ |

对于来自于随机化完全区组设计的一个观测值,其模型可以写为形式

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

其中模型的各项定义如下:

$y_{ij}$ : 第  $j$  个区组中接受处理  $i$  的试验单元上的响应值;

$\mu$ : 总的处理均值,未知常数;

$\alpha_i$ : 处理  $i$  的效应,未知常数;

$\beta_j$ : 区组  $j$  的效应,未知常数;

$\epsilon_{ij}$ : 第  $j$  个区组中接受处理  $i$  的试验单元的响应的随机误差。我们要求各个  $\epsilon_{ij}$  服从均值为 0, 共同方差为  $\sigma_e^2$  的正态分布。另外,这些误差之间相互独立。

由我们对模型要求的这些条件可以得出,记录到的第  $j$  个区组中处理  $i$  的响应值服从均值为

$$E(y_{ij}) = \mu + \alpha_i + \beta_j$$

方差为  $\sigma_e^2$  的正态分布。表 15.9 给出了表 15.8 中的数据的总体均值(期望值)。

表 15.9 随机化完全区组设计中  $y_{ij}$  的期望值

| 处 理 | 区 组                                    |  |     |  |
|-----|--|--|-----|--|
|     | 1                                      | 2                                      | ... | b                                      |
| 1   | $E(y_{11}) = \mu + \alpha_1 + \beta_1$ | $E(y_{12}) = \mu + \alpha_1 + \beta_2$ | ... | $E(y_{1b}) = \mu + \alpha_1 + \beta_b$ |
| 2   | $E(y_{21}) = \mu + \alpha_2 + \beta_1$ | $E(y_{22}) = \mu + \alpha_2 + \beta_2$ | ... | $E(y_{2b}) = \mu + \alpha_2 + \beta_b$ |
| ... | ...                                    | ...                                    | ... | ...                                    |
| t   | $E(y_{t1}) = \mu + \alpha_t + \beta_1$ | $E(y_{t2}) = \mu + \alpha_t + \beta_2$ | ... | $E(y_{tb}) = \mu + \alpha_t + \beta_b$ |

关于这个期望值的表做以下几点说明。第一,任何一对接受相同处理的观测值(在表 15.9 中出现在同一行)的均值之间只有区组效应  $\beta_j$  不同。例如,  $y_{11}$  和  $y_{12}$  (接受处理 1 的两个观测值)的期望值为

$$E(y_{11}) = \mu + \alpha_1 + \beta_1 \quad E(y_{12}) = \mu + \alpha_1 + \beta_2$$

它们的差为

$$E(y_{11}) - E(y_{12}) = (\mu + \alpha_1 + \beta_1) - (\mu + \alpha_1 + \beta_2) = \beta_1 - \beta_2$$

这说明了这样一个事实:即  $y_{11}$  是在区组 1 中记录到的,  $y_{12}$  是在区组 2 中记录到的,但二者都是来自于接受处理 1 的试验单元的,因而,没有处理效应的差异存在,只可能有区组效应的差异。第二,在同一区组中的两个观测值(在表 15.9 中出现在同一列)的均值之间只有处理效应不同。例如,  $y_{11}$  和  $y_{21}$  都出现在区组 1。由表 15.9, 它们的均值的差为

$$E(y_{11}) - E(y_{21}) = (\mu + \alpha_1 + \beta_1) - (\mu + \alpha_2 + \beta_1) = \alpha_1 - \alpha_2$$

它说明了这样一个事实:即这些试验单元接受了不同处理,但是在同一区组中被观测的,因而,二者之间只可能有处理效应的差异存在,没有区组效应的差异。第三,对于接受不同处理,并在不同的区组中被观测的两个试验单元,它们的响应值的期望值之间既有处理差异的效应,也有区组差异的效应。比如,观测值  $y_{11}$  和  $y_{22}$  的期望值之差为

$$E(y_{11}) - E(y_{22}) = (\mu + \alpha_1 + \beta_1) - (\mu + \alpha_2 + \beta_2) = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)$$

利用这些有关随机化区组设计的模型的知识,我们可以说明过滤的概念,并看出随机化区组设计如何滤除了由区组带来的变异。考虑一个有  $t=3$  个处理(1, 2 和 3),  $b=3$  个区组的随机化区组设计,如表 15.10 所示。

表 15.10 有  $t=3$  个处理,  $b=3$  个区组的随机化完全区组设计

| 区 组 | 处 理 |   |   |
|-----|-----|---|---|
| 1   | 1   | 2 | 3 |
| 2   | 1   | 3 | 2 |
| 3   | 3   | 1 | 2 |

这个随机化区组设计的模型是

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad (i = 1, 2, 3 \quad j = 1, 2, 3)$$

假设我们希望估计处理 2 和处理 1 的平均响应的差, 即  $\alpha_2 - \alpha_1$ 。相应的样本均值的差  $\bar{y}_{2.} - \bar{y}_{1.}$  可以作为  $\alpha_2 - \alpha_1$  的一个点估计。代入上面的模型, 得到

$$\begin{aligned} \bar{y}_{1.} &= \frac{1}{3} \sum_j y_{1j} \\ &= \frac{1}{3} [(\mu + \alpha_1 + \beta_1 + \varepsilon_{11}) + (\mu + \alpha_1 + \beta_2 + \varepsilon_{12}) + (\mu + \alpha_1 + \beta_3 + \varepsilon_{13})] \\ &= \mu + \alpha_1 + \bar{\beta} + \bar{\varepsilon}_1 \end{aligned}$$

其中  $\bar{\beta}$  表示三个区组效应  $\beta_1, \beta_2$  和  $\beta_3$  的平均值,  $\bar{\varepsilon}_1$  表示三个随机误差  $\varepsilon_{11}, \varepsilon_{12}$  和  $\varepsilon_{13}$  的平均值。同样, 容易得到

$$\bar{y}_{2.} = \mu + \alpha_2 + \bar{\beta} + \bar{\varepsilon}_2$$

由此

$$\bar{y}_{2.} - \bar{y}_{1.} = (\alpha_2 - \alpha_1) + (\bar{\varepsilon}_2 - \bar{\varepsilon}_1)$$

注意区组效应是如何消去, 从而留下  $(\bar{\varepsilon}_2 - \bar{\varepsilon}_1)$  作为用  $\bar{y}_{2.} - \bar{y}_{1.}$  估计  $\alpha_2 - \alpha_1$  的误差的。

如果用一个完全随机设计代替了随机化完全区组设计, 处理就会被随机地分配给试验单元, 不大可能出现每一个处理在每个区组中都出现的情况。当一个处理在一个区组中出现多于一次时, 我们用  $\bar{y}_{2.} - \bar{y}_{1.}$  作为  $\alpha_2 - \alpha_1$  的估计, 就不能像上面那样所有区组效应都消去。这样, 估计的误差就不仅包含  $\bar{\varepsilon}_2 - \bar{\varepsilon}_1$  而且还包含消不去的区组效应, 也就是说,

$$\bar{y}_{2.} - \bar{y}_{1.} = (\alpha_2 - \alpha_1) + [(\bar{\varepsilon}_2 - \bar{\varepsilon}_1) + (\text{未消除的区组效应})]$$

因此, 随机化区组设计在比较处理均值时, 滤除了源于区组的变异性, 从而减少了估计的误差。

图 15.1 是一张期望值  $\mu_{ij}$  的图。它表明, 不同区组 (比如说  $j$  和  $j'$ ) 中接受相同处理的观测值的均值之间的差, 对于所有处理是一致的, 即

$$\mu_{ij} - \mu_{ij'} = \beta_j - \beta_{j'} \quad \text{对 } i = 1, \dots, t$$

这意味着具有相同处理的均值之间的连线形成一组平行线。

使用随机化完全区组设计的主要目标是考察  $t$  个处理均值  $\mu_1, \mu_2, \dots, \mu_t$  之间的差异, 这里  $\mu_i$  为处理  $i$  的平均响应。零假设是处理均值之间没有差异, 研究假设是处理均值之间有差异, 即

$$H_0: \mu_1 = \mu_2 = \dots = \mu_t \quad H_a: \text{至少有一个 } \mu_i \text{ 与别的不同}$$

这组假设等价于检验

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_t = 0 \quad H_a: \text{至少有一个 } \alpha_i \text{ 不为 } 0$$

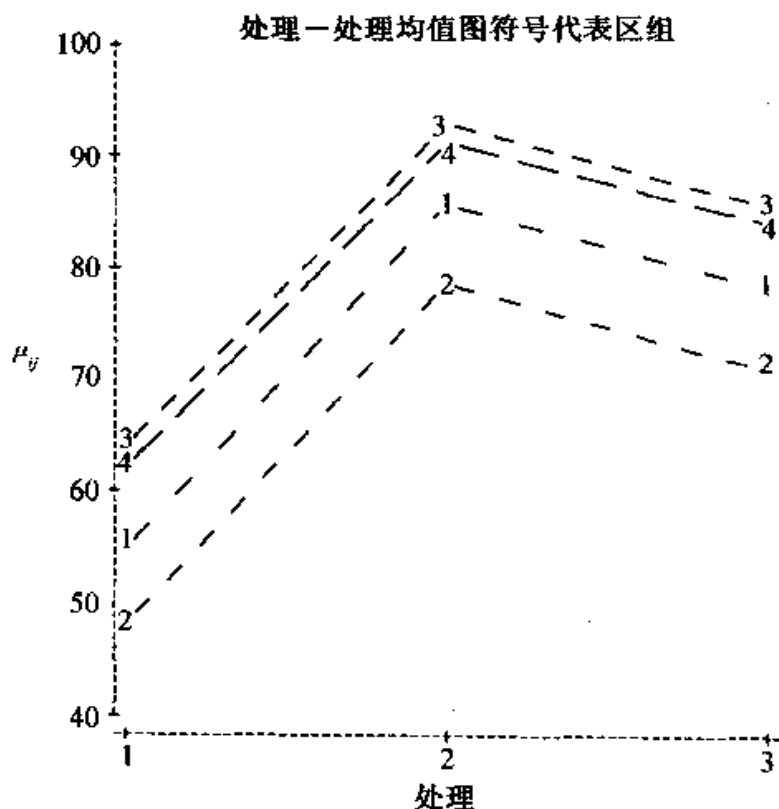


图 15.1 随机化区组设计中的处理均值

这两组假设之所以是等价的,是因为正如我们在表 15.9 中所看到的那样,当比较同一个区组中的两个不同处理的平均响应时,这两个平均响应的差为

$$\mu_{i.} - \mu_{i'.} = \alpha_i - \alpha_{i'}$$

所以,在  $H_0$  下,对于给定的区组,这两个处理有相同的平均响应。通过考察随机化区组设计的模型,并把总平方和分解为处理效应、区组效应和随机效应三个部分,我们可以得到上述假设的检验统计量。利用表 15.8,我们可以引入分解总平方和所需要的记号。这些记号如下。

$y_{ij}$ : 区组  $j$  中处理  $i$  上的观测值

$t$ : 处理个数

$b$ : 区组个数

$\bar{y}_{i.}$ : 处理  $i$  的样本均值,  $\bar{y}_{i.} = \frac{1}{b} \sum_{j=1}^b y_{ij}$

$\bar{y}_{.j}$ : 区组  $j$  的样本均值,  $\bar{y}_{.j} = \frac{1}{t} \sum_{i=1}^t y_{ij}$

$\bar{y}_{..}$ : 总样本均值,  $\bar{y}_{..} = \frac{1}{tb} \sum_{ij} y_{ij}$

各观测值关于它们的平均值的总平方和,与以前一样,定义为

$$\text{TSS} = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$$

这个平方和将被分解成三个不同的变异来源:一个是源于处理之间的变异,一个是源于区组之间的变异,另一个则是源于不能被处理差异和区组差异所解释的所有其他的变异,称之为**误差**。**TSS 的分解**可以通过对随机化完全区组设计的模型

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

的考察得到。模型中的参数有样本估计:

$$\hat{\mu} = \bar{y}_{..} \quad \hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..} \quad \text{和} \quad \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$$

用一些代数计算可以证明 TSS 具有如下形式:

$$\begin{aligned} \sum_{ij} (y_{ij} - \bar{y}_{..})^2 &= b \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 + t \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 \\ &\quad + \sum_{ij} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \end{aligned}$$

我们用上面的参数估计来解释这个分解中的各项。等号右边第一项度量了各处理均值  $\bar{y}_{i.}$  相对于总平均值  $\bar{y}_{..}$  的变异性。因此

$$\text{SST} = b \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = b \sum_i (\hat{\alpha}_i)^2$$

称为**处理间平方和**,它是各个  $y_{ij}$  中由于处理均值之间的差异而产生的变异性的一个度量。类似地,第二项

$$\text{SSB} = t \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 = t \sum_j (\hat{\beta}_j)^2$$

度量了各个区组均值相对于总平均值的变异性,称之为**区组间平方和**。第三项,称为**误差平方和**,SSE,表示各  $y_{ij}$  中没有被处理和区组差异所解释的变异性。这一项有几种不同的表示形式:

$$\text{SSE} = \sum_{ij} (e_{ij})^2 = \sum_{ij} (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2 = \text{TSS} - \text{SST} - \text{SSB}$$

其中  $e_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$  为用来验证模型条件的残差。我们可以把这些计算总结成如表 15.11 的 AOV 表。

表 15.11 随机化完全区组设计的方差分析表

| 来源 | SS  | df               | MS   | F                         |
|----|-----|------------------|--|---------------------------|
| 处理 | SST | $t - 1$          | $\text{MST} = \text{SST} / (t - 1)$        | $\text{MST} / \text{MSE}$ |
| 区组 | SSB | $b - 1$          | $\text{MSB} = \text{SSB} / (b - 1)$        | $\text{MSB} / \text{MSE}$ |
| 误差 | SSE | $(b - 1)(t - 1)$ | $\text{MSE} = \text{SSE} / (b - 1)(t - 1)$ |                           |
| 总和 | TSS | $bt - 1$         |  |                           |

对

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0 \quad H_a: \text{至少有一个 } \alpha_i \text{ 不为 } 0$$

的检验统计量为比值

$$F = \frac{MST}{MSE}$$

当  $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0$  为真时, MST 和 MSE 都是试验误差方差  $\sigma_e^2$  的无偏估计, 也就是说, 当  $H_0$  成立时, MST 和 MSE 在大量重复抽样时的均值都等于  $\sigma_e^2$ 。这一均值称为期望均方。这一事实表示为

$$E(MST) = \sigma_e^2 \quad E(MSE) = \sigma_e^2$$

由此, 我们可以期望当  $H_0$  为真时,  $F = MST/MSE$  的值接近于 1。

当  $H_a$  为真时, MSE 的期望值仍然是  $\sigma_e^2$ 。然而, MST 此时不再是  $\sigma_e^2$  的无偏估计。事实上, 可以证明处理的期望均方为

$$E(MST) = \sigma_e^2 + b\theta_T \quad \text{其中 } \theta_T = \frac{1}{t-1} \sum_i \alpha_i^2$$

因此, 处理均值中大的差异会导致  $\theta_T$  有大的取值, MST 的期望值将比 MSE 的期望值来得大。这样, 比值  $F = MST/MSE$  将倾向于比 1 大。于是, 当我们观测到的  $F$  值大于一个  $F$  分布的右侧尾部的值时, 就拒绝  $H_0$ 。

对于一个取定的犯第一类错误的概率, 上述讨论给出了下面的判决规则:

拒绝  $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0$  当  $F = MST/MSE$  大于  $F_{\alpha, df_1, df_2}$

这里  $F_{\alpha, df_1, df_2}$  是用  $\alpha =$  取定的犯第一类错误的概率,  $df_1 = df_{MST} = t - 1$  和  $df_2 = df_{MSE} = (b - 1)(t - 1)$  从附录的表 8 中查出的。另一种方法是利用

$$p \text{ 值} = P(F_{df_1, df_2} > F_{obs})$$

计算检验统计量的  $p$  值, 其中的概率是用自由度  $df_1 = t - 1$  和  $df_2 = (b - 1)(t - 1)$  的  $F$  分布计算的。计算出  $p$  值后, 我们比较  $p$  值与取定的犯第一类错误的概率, 小的  $p$  值支持研究假设, 而对大的  $p$  值, 我们则不能拒绝  $H_0$ 。

一般来说, 仅当考察区组划分是否有效地减少了试验单元的变异性时, 才讨论区组效应, 因而不检验关于区组效应的假设。然而, 我们仍然可以问, 在一个试验中, 区组划分是否提高了处理均值比较的精度。以  $MSE_{RCB}$  和  $MSE_{CR}$  分别记一个随机化完全区组设计和一个完全随机化设计的均方误差。两个设计的精度的一种度量是第  $i$  个处理均值估计的方差,  $\hat{\mu}_i = \bar{y}_{i.}$  ( $i = 1, 2, \dots, t$ )。对于一个随机化完全区组设计,  $\bar{y}_{i.}$  的估计方差为  $MSE_{RCB}/b$ 。对于完全随机化设计,  $\bar{y}_{i.}$  的估计方差为  $MSE_{CR}/r$ , 这里  $r$  是每个处理的重复观测数, 满足关系式

$$\frac{MSE_{CR}}{r} = \frac{MSE_{RCB}}{b} \quad \text{或} \quad \frac{MSE_{CR}}{MSE_{RCB}} = \frac{r}{b}$$

数  $\frac{r}{b}$  称为随机化完全区组设计与完全随机化设计相比较的相对效率, 记为  $RE(RCB, CR)$ 。  $MSE_{CR}$  比  $MSE_{RCB}$  大得越多, 要想在完全随机化设计中得到的处理均

值估计的精度与在随机化完全区组设计中相同,所需要的  $r$  就越大。因此,如果区组划分是有效的,我们可以期望,随机化完全区组设计中试验单元中的变异性比完全随机化设计中试验单元中的变异性要小,比值  $MSE_{CR}/MSE_{RCB}$  会较大,从而  $r$  比  $b$  大许多。这样,要在估计  $\mu_i$  时达到同样的估计精度,在完全随机化设计中所需的数据量比在随机化完全区组设计中所需要的要大。当区组划分没有效果时,比值  $MSE_{CR}/MSE_{RCB}$  将接近于 1,  $r$  和  $b$  相近。

对随机化完全区组设计相对于完全随机化设计的效率的评估,不能用实践的方法来进行,因为我们并不实施完全随机化设计。然而,我们可以使用随机化完全区组设计中的均方,即  $MSB$  和  $MSE$ ,通过公式

$$RE(RCB, CR) = \frac{MSE_{CR}}{MSE_{RCB}} = \frac{(b-1)MSB + b(t-1)MSE}{(bt-1)MSE}$$

来得到相对效率  $RE(RCB, CR)$ 。当  $RE(RCB, CR)$  远比 1 大时,  $r$  比  $b$  大,我们就可以说区组划分是有效的,因为在完全随机化设计中比在随机化完全区组设计中需要更多的观测值。

### 例 15.2

某研究人员进行了一个试验,来比较三种不同的杀虫剂在一种菜豆上的使用效果。为了得到充分的数据,有必要利用 4 块不同的土地做试验。由于这 4 块地的土壤肥力、灌溉条件和避风情况有些差别,该研究人员决定以地块作为区组,进行随机化完全区组试验。每个地块分为三行,各行之间留有一定距离,以便杀虫剂能够限制在一行中使用。每行种植 100 粒种子,并使用指定给该行的杀虫剂。在每一地块内,杀虫剂被随机地分配给块中的行,使得在该地块中每一种杀虫剂有一行使用。关心的响应变量是每行中的出苗数,数据和均值由表 15.12 给出。

表 15.12 例 15.2 中按杀虫剂和地块给出的出苗数

| 杀虫剂  | 地 块 |    |    |    | 杀虫剂均值 |
|------|-----|----|----|----|-------|
|      | 1   | 2  | 3  | 4  |       |
| 1    | 56  | 48 | 66 | 62 | 58    |
| 2    | 83  | 78 | 94 | 93 | 87    |
| 3    | 80  | 72 | 83 | 85 | 80    |
| 地块均值 | 73  | 66 | 81 | 80 | 75    |

- 对这个试验,写出适当的统计模型。
- 进行比较三种杀虫剂效果的方差分析,取  $\alpha = 0.05$ 。
- 把你的结果总结在 AOV 表中。



d. 计算随机化完全区组设计相对于完全随机化设计的相对效率。

**解答** 我们看出,该试验是一个有  $b=4$  个区组,每个区组有  $t=3$  个处理的随机化完全区组设计。其统计模型为

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \quad i = 1, 2, 3 \quad j = 1, 2, 3, 4$$

从表 15.12 可知,我们可以用  $\hat{\mu}_{i.} = \bar{y}_{i.}$  来估计处理均值  $\mu_{i.}$ , 计算结果为

$$\hat{\mu}_{1.} = 58 \quad \hat{\mu}_{2.} = 87 \quad \hat{\mu}_{3.} = 80$$

看来使用杀虫剂 1 的行与使用其他两种杀虫剂的行比起来出苗最少。下面我们估计模型中的参数,并构造 AOV 表。回想  $\hat{\mu} = \bar{y}_{..}$ ,  $\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$  和  $\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$ 。用  $\hat{\mu} = \bar{y}_{..} = 75$ , 我们得到

杀虫剂效应

$$\hat{\alpha}_1 = 58 - 75 = -17$$

$$\hat{\alpha}_2 = 87 - 75 = 12$$

$$\hat{\alpha}_3 = 80 - 75 = 5$$

区组效应

$$\hat{\beta}_1 = 73 - 75 = -2$$

$$\hat{\beta}_2 = 66 - 75 = -9$$

$$\hat{\beta}_3 = 81 - 75 = 6$$

$$\hat{\beta}_4 = 80 - 75 = 5$$

把这些式子代入平方和,得到

$$\text{TSS} = \sum_{ij} (y_{ij} - \bar{y}_{..})^2 = (56 - 75)^2 + (48 - 75)^2 + \cdots + (85 - 75)^2 = 2,296$$

$$\text{SST} = b \sum_i (\hat{\alpha}_i)^2 = 4[(-17)^2 + (12)^2 + (5)^2] = 1,832$$

$$\text{SSB} = t \sum_j (\hat{\beta}_j)^2 = 3[(-2)^2 + (-9)^2 + (6)^2 + (5)^2] = 438$$

相减得

$$\text{SSE} = \text{TSS} - \text{SST} - \text{SSB} = 2,296 - 1,832 - 438 = 26$$

表 15.13 中的方差分析表总结了上述结果。注意,表中某一来源的均方由该来源上的平方和除以其自由度得到。

表 15.13 例 15.2 中数据的 AOV 表

| 来源 | SS    | df | MS     | F      | p 值    |
|----|-------|----|--------|--------|--------|
| 处理 | 1,832 | 2  | 916    | 211.38 | 0.0001 |
| 区组 | 438   | 3  | 146    | 33.69  | 0.0004 |
| 误差 | 26    | 6  | 4.3333 |        |        |
| 总和 | 2,296 | 11 |        |        |        |

对处理均值间差异的  $F$  检验, 即

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0 \quad H_a: \text{至少一个 } \alpha_i \text{ 不为 } 0$$

使用比值 MST/MSE。由于计算所得的  $F$  值 211.38 大于由  $df_1 = 2, df_2 = 6$  以及  $\alpha = 0.05$  查得的  $F$  分布的表值 5.14, 我们拒绝  $H_0$ , 认为在三种杀虫剂下的出苗数有显著差异 ( $p$  值  $< 0.0001$ )。

接下来我们讨论区组划分相对于完全随机化设计是否提高了分析的精度。从 AOV 表, 我们得到  $MSB = 146, MSE = 4.3333$ 。因而, 该随机化完全区组设计相对于完全随机化设计的相对效率为

$$\begin{aligned} RE(RCB, CR) &= \frac{(b-1)MSB + b(t-1)MSE}{(bt-1)MSE} \\ &= \frac{(4-1)146 + 4(3-1)(4.3333)}{[(4)(3)-1](4.3333)} = 9.92 \end{aligned}$$

即在完全随机化设计中, 为了得到与随机化完全区组设计中处理均值估计的同样精度, 每个处理需要大约 10 倍数目的观测值。这些地块在物理特征上有相当大的差异, 因此, 在这个试验中, 区组划分起到了关键的作用。

例 15.2 中的结果仅当我们能够确认模型条件与观测到的数据的表现相一致时, 才是有效的。因此, 我们用残差  $e_{ij} = y_{ij} - \bar{\mu} - \hat{\alpha}_i - \hat{\beta}_j$  来看一看模型的正态性条件, 等方差条件, 观察相互独立的条件在数据中是否看上去得到了满足。下面的例子中包含了这种分析的计算机输出结果。

### 例 15.3

这里给出例 15.2 中试验的计算机输出结果。对这些结果与用平方和的定义计算的结果进行比较, 并说明模型条件是否看上去得到了满足。

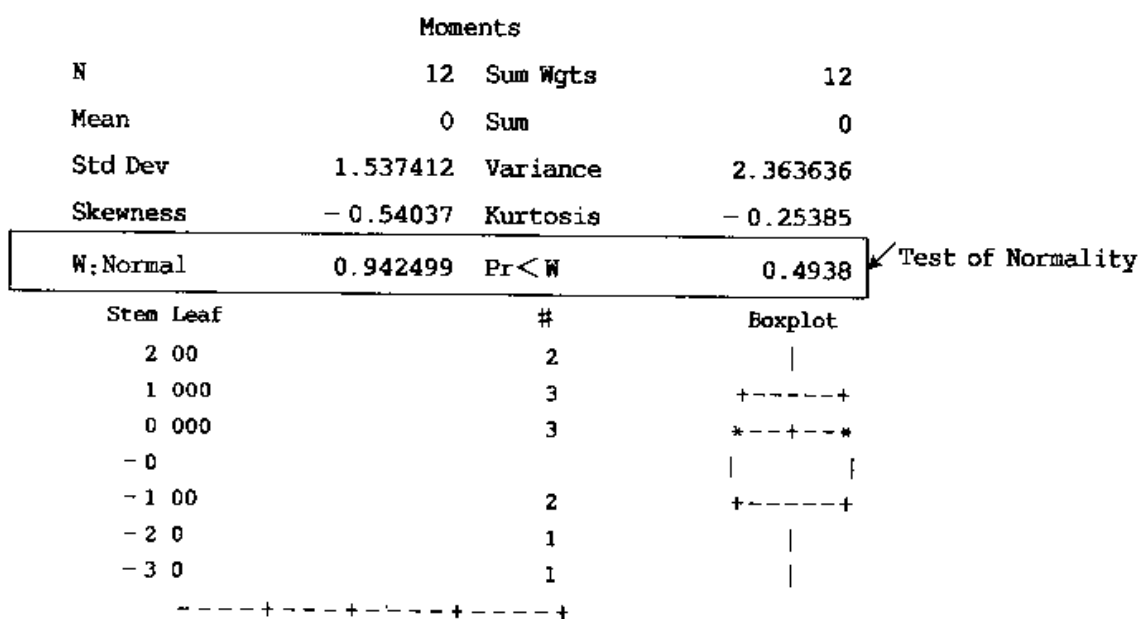
Dependent Variable: NUMBER OF SEEDLINGS

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 5  | 2270.0000      | 454.0000    | 104.77  | 0.0001 |
| Error           | 6  | 26.0000        | 4.3333      |         |        |
| Corrected Total | 11 | 2296.0000      |             |         |        |

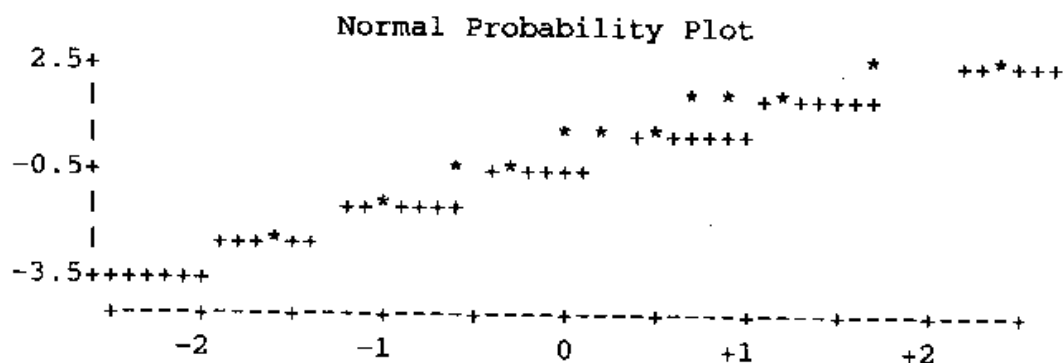
| Source       | DF | Type I SS | Mean Square | F Value | Pr > F |
|--------------|----|-----------|-------------|---------|--------|
| INSECTICIDES | 2  | 1832.0000 | 916.0000    | 211.38  | 0.0001 |
| PLOTS        | 3  | 438.0000  | 146.0000    | 33.69   | 0.0004 |

RESIDUAL ANALYSIS

Variable = RESIDUALS



Variable = RESIDUALS



**解答** 注意我们手算的结果与这里给出的计算机输出结果一致。一般说来, 我们的手算结果有一些舍入误差, 从而得到的结果与计算机输出结果有一些差别。我们强烈建议, 在方差分析中使用计算机软件程序进行计算, 以避免这些潜在的舍入误差。至于说到模型条件是否得到满足, 我们首先指出, 对于假设  $H_0$ : 残差服从正态分布, Shapiro-Wilks 检验的  $p$  值为 0.4389, 因而我们不能拒绝  $H_0$ , 正态性条件看来得到满足。另外, 茎叶图、盒型图和正态概率图也都与残差服从正态分布的条件相吻合。图 15.2 是一张残差—估计处理均值图。从该图可以看出, 残差中的变异性不随处理而变化。

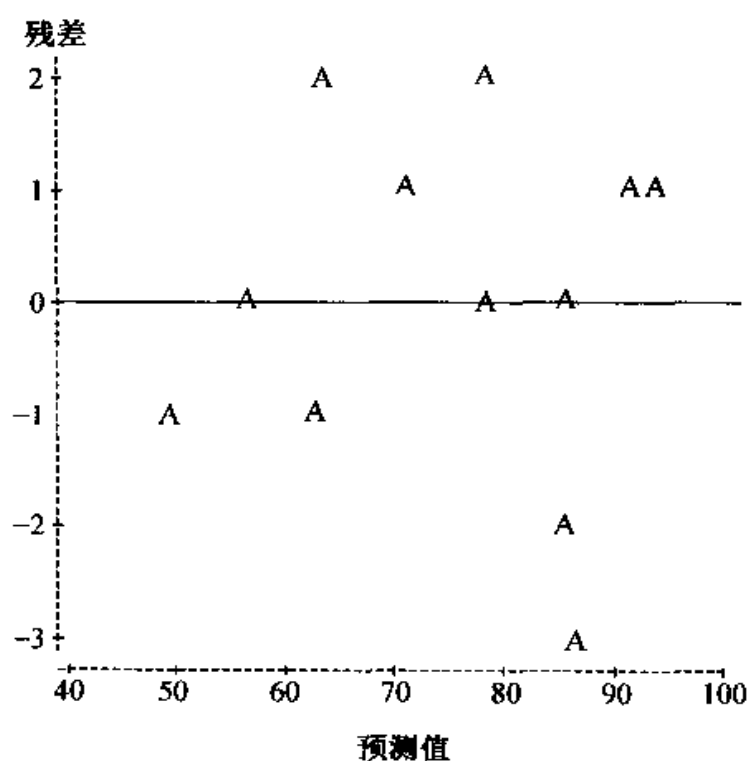


图 15.2 残差—处理均值

## 练习

**15.1(教育)** 某研究人员要确定让来自于低收入社区的一年级学生参加一个“良好的开端”的活动是否有助于他们改善在学校的表现。该研究人员得到一个由六个参加了这个活动的孩子和六个没参加该活动的孩子组成的随机样本。这些孩子在家里从父母那里得到的支持的类型大不相同,因此,在了解了这十二个孩子的家庭环境后,研究人员基于家庭环境的相似性对这些孩子进行了配对。在一年级完成以后,对这些学生进行了综合智力测验,结果如下。

| 配对 | 测验得分          |               |
|----|---------------|---------------|
|    | 参加了“良好的开端”的活动 | 未参加“良好的开端”的活动 |
| 1  | 58            | 47            |
| 2  | 73            | 67            |
| 3  | 85            | 69            |
| 4  | 76            | 62            |
| 5  | 88            | 77            |
| 6  | 90            | 77            |

a. 参加了“良好的开端”活动的学生是否比不参加该活动的学生有更高的平均得分?  $\alpha = 0.05$ 。

b. 给出随机化完全区组设计相对于完全随机化设计的效率。解释你的结果。

**15.2** 参考练习 15.1. 用配对  $t$  检验来分析上述数据, 并把结果与练习 15.1 中的结果进行比较。(在随机化完全区组设计中处理差异的  $F$  检验, 在只有两个处理的情形, 等价于第六章的配对  $t$  检验)

**15.3(工程)** 在火箭推进剂的研制中进行一项试验, 比较由氧化物、黏合剂和燃料作为成分的四种混合物。这四种混合物, 按照放入的氧化物的混合比例列出为:

| 混合物 | 氧化物 | 黏合剂 | 燃料  |
|-----|-----|-----|-----|
| 1   | 0.4 | 0.4 | 0.2 |
| 2   | 0.4 | 0.2 | 0.4 |
| 3   | 0.6 | 0.2 | 0.2 |
| 4   | 0.5 | 0.3 | 0.2 |

为了比较这四种混合物, 为每种混合物准备了 5 个不同的样品用来做试验。有 5 位实验员。随机地为每位实验员在每种混合物中指定一个样品, 要求其测量这些推进剂样品的推力。数据总结如下:

| 混合物 | 实验员   |       |       |       |       |
|-----|-------|-------|-------|-------|-------|
|     | 1     | 2     | 3     | 4     | 5     |
| 1   | 2,340 | 2,355 | 2,362 | 2,350 | 2,348 |
| 2   | 2,658 | 2,650 | 2,665 | 2,640 | 2,653 |
| 3   | 2,449 | 2,458 | 2,432 | 2,437 | 2,445 |
| 4   | 2,403 | 2,410 | 2,418 | 2,397 | 2,405 |

a. 找出该试验中的区组和处理。

b. 给出随机化的方法。

c. 为什么对于该试验, 在完全随机化设计和随机化完全区组设计中更喜欢用后者?

**15.4** 参见 15.3。

a. 对这个试验写出一个模型。

b. 估计模型中的参数。

c. 用下面给出的计算机输出结果做方差分析, 其中  $\alpha = 0.05$ 。

- d. 关于被试验的四种混合物中最好的混合物,你能给出什么结论?(注:响应值越高,火箭推进剂的推力越好)
- e. 计算随机化区组设计相对于完全随机化设计的相对效率。对所得值给出解释。区组在减少试验单元间的变异性方面效果如何?给出解释。

## General Linear Models Procedure For Data in Exercise 15.3

Dependent Variable: THRUST

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 7  | 261713.45      | 37387.64    | 542.96  | 0.0001 |
| Error           | 12 | 826.30         | 68.86       |         |        |
| Corrected Total | 19 | 262539.75      |             |         |        |

| R-Square | C. V.    | Root MSE | Y Mean |
|----------|----------|----------|--------|
| 0.996853 | 0.336807 | 8.2981   | 2463.8 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|--------|----|-----------|-------------|---------|--------|
| M      | 3  | 261260.95 | 87086.98    | 1264.73 | 0.0001 |
| I      | 4  | 452.50    | 113.12      | 1.64    | 0.2273 |

**15.5(心理学)** 一位在一家大公司工作的工业心理学家做一项试验,以评价背景音乐对于秘书打字效率的影响。他从众多的秘书中挑选了一个由七个人组成的随机样本。给每个秘书放三种背景音乐:无音乐,古典音乐和激烈的摇滚乐,然后对她们进行标准的打字测验,以测验他们的打字速度并在出现打字错误时给以惩罚。对于每一个参加试验的秘书,随机化指定放听音乐的顺序。在下面给出的结果中,得分越高,表现越好。这是一个特殊类型的随机化完全区组试验,其中单个的试验单元用作一个区组,接受所有的处理。

| 音乐类型 | 秘 书 |    |    |    |    |    |    |
|------|-----|----|----|----|----|----|----|
|      | 1   | 2  | 3  | 4  | 5  | 6  | 7  |
| 无音乐  | 20  | 17 | 24 | 20 | 22 | 25 | 18 |
| 激烈摇滚 | 20  | 18 | 23 | 18 | 21 | 22 | 19 |
| 古典   | 24  | 20 | 27 | 22 | 24 | 28 | 16 |

- a. 对这个试验写出一个统计模型,并估计你的模型中的参数。
- b. 对于这三种音乐,平均打字的效率之间有无差异?取  $\alpha = 0.05$ 。
- c. 对于该随机化完全区组设计,可加模型合适吗?(提示:像图 15.1 那样,画

出数据的图)

- d. 计算该随机化区组设计相对于完全随机化设计的相对效率, 并对该值进行解释。区组在减少试验单元的变异性方面有效吗? 为什么?

15.6 参见练习 15.5。练习 15.5 中数据的计算机输出结果如下。比较这里的结果和你计算的结果。模型的条件是否看上去得到了满足?

General Linear Models Procedure For Exercise 15.5

Dependent Variable: TYPING EFFICIENCY

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 8  | 180.28571      | 22.53571    | 9.53    | 0.0004 |
| Error           | 12 | 28.38095       | 2.36508     |         |        |
| Corrected Total | 20 | 208.66667      |             |         |        |

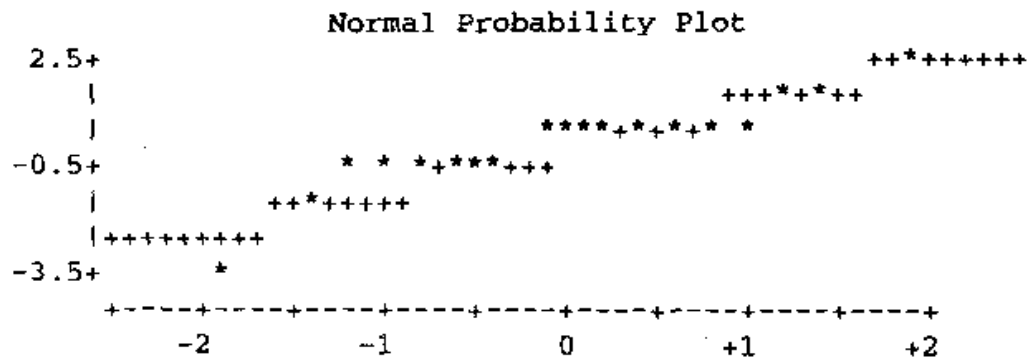
| R-Square | C. V.    | Root MSE | Y Mean |
|----------|----------|----------|--------|
| 0.863989 | 7.208819 | 1.5379   | 21.333 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|--------|----|-----------|-------------|---------|--------|
| M      | 2  | 30.95238  | 15.47619    | 6.54    | 0.0120 |
| S      | 6  | 149.33333 | 24.88889    | 10.52   | 0.0003 |

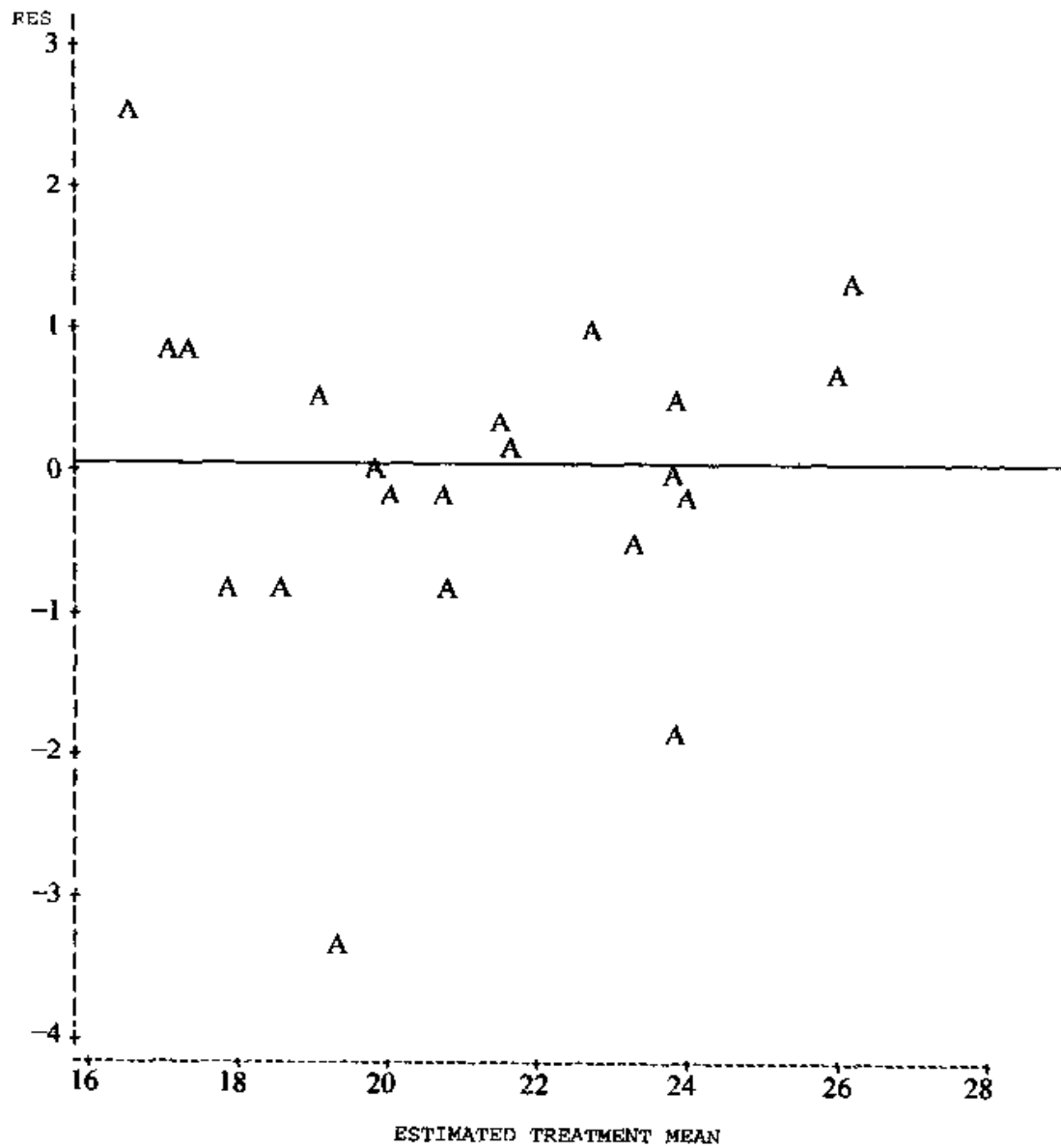
RESIDUAL ANALYSIS:

Variable = RESIDUAL

| Moments                         |           |          |               |
|---------------------------------|-----------|----------|---------------|
| N                               | 21        | Sum Wgts | 21            |
| Mean                            | 0         | Sum      | 0             |
| Std Dev                         | 1.191238  | Variance | 1.419048      |
| Skewness                        | -0.77527  | Kurtosis | 2.587721      |
| W: Normal                       | 0.936418  | Pr < W   | 0.1813        |
| Stem                            | Leaf      | #        | Boxplot       |
| 2                               | 5         | 1        | 0             |
| 1                               | 03        | 2        |               |
| 0                               | 001355789 | 9        | + - - + - - + |
| -0                              | 9985211   | 7        | + - - - - +   |
| -1                              | 8         | 1        |               |
| -2                              |           |          |               |
| -3                              | 3         | 1        | 0             |
| - - - + - - - + - - - + - - - + |           |          |               |



Plot of RESIDUAL VERSUS ESTIMATED TREATMENT MEAN





**15.7(心理学)** 一位质量控制工程师想办一个训练班,对工人进行全面质量管理(TQM)原理的培训。要在全公司进行这样的培训费用巨大,因而该工程师想看一看四种类型的培训班中哪一种最有效。响应变量是在参加培训班后工人生产产值的增加。由于培训的效果可能与工人事先对于 TQM 的态度有关,在参加培训班以前,就工人对 TQM 的态度进行了测试,并依据测试结果把他们分成了五组,每组有四个人。在每个组中,随机地把这些工人分派到四种类型的班中。工人的产值增量如下。

| 训练班类型 | 态 度  |       |      |       |       | 均值    |
|-------|------|-------|------|-------|-------|-------|
|       | 1    | 2     | 3    | 4     | 5     |       |
| A     | 33   | 38    | 39   | 42    | 62    | 42.8  |
| B     | 35   | 37    | 43   | 47    | 71    | 46.6  |
| C     | 40   | 42    | 45   | 52    | 74    | 50.6  |
| D     | 54   | 50    | 55   | 62    | 84    | 61.0  |
| 均值    | 40.5 | 41.75 | 45.5 | 50.75 | 72.75 | 50.25 |

- 对这个试验写出一个统计模型,并估计你的模型中的参数。
- 对于这四种训练班,平均产值增量之间有无差异? 取  $\alpha = 0.05$ 。
- 对于该随机化完全区组设计,可加模型合适吗?(提示:像图 15.1 那样,画出数据的图)
- 计算该随机化区组设计相对于完全随机化设计的相对效率,并对该值进行解释。区组在减少试验单元的变异性方面有效吗? 为什么?

**15.8** 参见练习 15.7。练习 15.7 中数据的计算机输出结果如下。比较这里的结果和你计算的结果。模型的条件是否看上去得到了满足?

General Linear Models Procedure For Exercise 15.7

Dependent Variable: INCREASE IN PRODUCTIVITY

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 7  | 3708.0500      | 529.7214    | 114.12  | 0.0001 |
| Error           | 12 | 55.7000        | 4.6417      |         |        |
| Corrected Total | 19 | 3763.7500      |             |         |        |

|          |          |          |        |
|----------|----------|----------|--------|
| R-Square | C. V.    | Root MSE | Y Mean |
| 0.985201 | 4.287468 | 2.1545   | 50.250 |

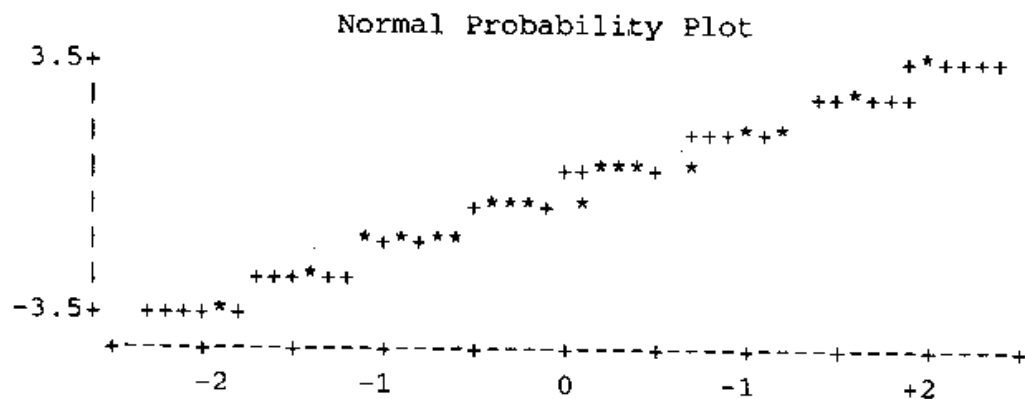
| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|--------|----|-----------|-------------|---------|--------|
| W      | 3  | 922.5500  | 307.5167    | 66.25   | 0.0001 |
| A      | 4  | 2785.5000 | 696.3750    | 150.03  | 0.0001 |

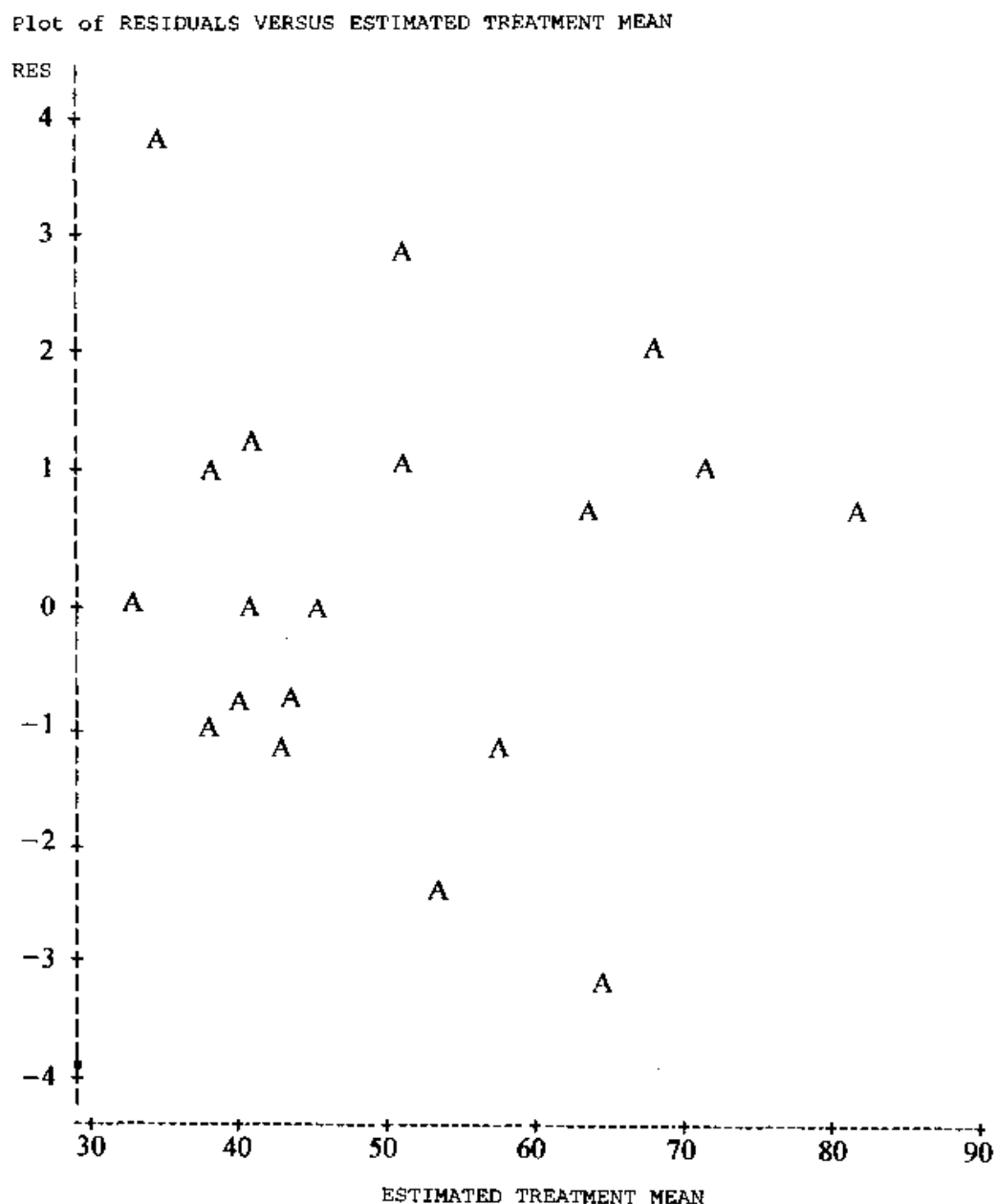
## RESIDUAL ANALYSIS:

| Moments  |          |          |          |
|----------|----------|----------|----------|
| N        | 20       | Sum Wgts | 20       |
| Mean     | 0        | Sum      | 0        |
| Std Dev  | 1.712185 | Variance | 2.931579 |
| Skewness | 0.207266 | Kurtosis | 0.149301 |
| W:Normal | 0.985267 | Pr < W   | 0.9725   |

| Stem | Leaf  | # | Boxplot |
|------|-------|---|---------|
| 3    | 7     | 1 |         |
| 2    | 8     | 1 |         |
| 1    | 029   | 3 |         |
| 0    | 5599  | 4 | +--+--+ |
| -0   | 88110 | 5 | *-----* |
| -1   | 8321  | 4 | +-----+ |
| -2   | 5     | 1 |         |
| -3   | 3     | 1 |         |

-----+-----+-----+-----+





## 15.4 拉丁方设计

随机化完全区组设计用于有一个因子,并且实验员要控制一个外部变异因子的情况。对于有多个可能的外部变异来源的情形,拉丁方设计是适当的设计。考虑下面的例子。

### 例 15.4

一家大的律师事务所正在研究在四种电子表格软件中哪一种对他们的秘书班

子更合适。所有这四种软件都来自于国内知名的软件公司,因而这几个软件的输出质量都是可以接受的。对这些软件的最后的选择取决于哪一种软件更容易学。事物所的软件顾问注意到,学习软件中各种子程序所用的时间与秘书个人以及所处理问题的类型有关。由于参加到研究中的秘书不能正常工作,事物所决定从秘书班子中只选四人参加该试验,并限制他们离开日常岗位的时间。因此,顾问选定了四种类型的问题,并决定让每个秘书完成一项属于其中一种类型的任务。该研究中所要考虑的因子为

电子表格软件:A, B, C, D

秘书: 1, 2, 3, 4

问题: I (账目), II (数据表), III (概括统计量), IV (图表)

对于顾问来说,秘书和问题类型这两个因子为必须要考虑的外部变异来源,但不是最主要的。响应变量是秘书完成指定的任务所需要的时间,每个秘书都要完成四个问题。该顾问首先考虑用表 15.14 给出的随机化完全区组设计。

表 15.14 电子表格软件研究中的一个随机化完全区组设计

| 问题  | 秘 书 |   |   |   |
|-----|-----|---|---|---|
|     | 1   | 2 | 3 | 4 |
| I   | A   | A | C | A |
| II  | B   | D | A | D |
| III | D   | B | D | B |
| IV  | C   | C | B | C |

在这个设计中,对于每一位秘书,分别把她所要使用的四种电子表格软件随机地分配给她所要完成的四个问题。假设问题类型对完成该项任务所需时间有强的影响,比如,第 I 类问题费时远比其他问题多,而第 IV 类问题费时最少。那么,该设计对于软件 A 将产生一个很大的负偏差,因为该软件三次用于解决第 I 类型的问题,同时对于软件 C 会产生一个正偏差,因为它三次被用于解决第 IV 类型的问题。这样,如果用软件 C 完成四项任务的平均时间最短,我们不能确定到底是由于软件 C 是好程序还是由于四项任务中有三项为第 IV 类型的问题。

此例说明了这样一种情形,即试验单元(例中的问题)受到问题的类型和解决问题的秘书这两个外部变异来源的影响。为了在滤除第一个变异来源,也就是秘书的影响的基础上,再滤除第二个变异来源,即问题的类型的影响,我们可以修改上述的随机化完全区组设计。为此,我们对随机化进行限制,使得每个处理出现在每一行(问题类型)和每一列(秘书)。表 15.15 就是这样的—个随机化。注意,电

子表格软件已经这样被分配到问题的类型和秘书:每个电子表格软件在每个类型的问题上使用一次,每个秘书使用一次。因此,这种调整使得利用样本均值进行的电子表格软件的成对比较,免受由问题类型和秘书的不同而产生的变异的影响。

表 15.15 电子表格软件研究中的一个拉丁方设计

| 问题  | 秘 书 |   |   |   |
|-----|-----|---|---|---|
|     | 1   | 2 | 3 | 4 |
| I   | A   | B | C | D |
| II  | B   | C | D | A |
| III | C   | D | A | B |
| IV  | D   | A | B | C |

这个试验设计称为**拉丁方设计**。一般地,一个拉丁方设计可以用于在有两个外部变异来源存在的情况下比较  $t$  个处理均值,其中我们要把这两个外部因子分别划分成为  $t$  行和  $t$  列,然后把这  $t$  个处理随机地指定给这些行和列,使得每个处理在每行中出现一次,在每列中出现一次(见表 15.15)。

拉丁方设计的长处和不足列出如下。

### 拉丁方设计的长处与不足

#### 长处

1. 设计尤其适合于在有两个外部变异来源,每个变异来源有  $t$  个水平的情况下,比较  $t$  个处理的均值。

2. 分析简单。

#### 不足

1. 虽然对于任何  $t$  的值都可以构造出拉丁方,但当  $5 \leq t \leq 10$  时,拉丁方设计最适合于用来比较  $t$  个处理。

2. 任何另外的外部变异来源都会使误差项膨胀,从而使得对处理均值间的差别的探索变得更加困难。

3. 每一个处理对于响应值的效应必须在每一行上近似相同,在每一列上也近似相同。

下面给出拉丁方设计的定义。

**定义 15.2** 一个  $t \times t$  拉丁方设计包含  $t$  行和  $t$  列。 $t$  个处理被随机地分配给在行和列中的试验单元,使得每个处理出现在每一行和每一列中。

拉丁方设计中响应值的模型可以写为

$$y_{ij} = \mu + \alpha_k + \beta_i + \gamma_j + \varepsilon_{ij}$$

其中模型中的项定义如下:

$y_{ij}$ : 第  $i$  行第  $j$  列试验单元上的观测值;

$\mu$ : 总平均值, 未知常数;

$\alpha_k$ : 处理  $k$  的效应, 未知常数;

$\beta_i$ : 第  $i$  行的效应, 未知常数;

$\gamma_j$ : 第  $j$  列的效应, 未知常数;

$\varepsilon_{ij}$ : 第  $i$  行第  $j$  列上试验单元上的响应值的随机误差。我们要求各个  $\varepsilon_{ij}$  服从均值为 0, 共同方差为  $\sigma_e^2$  的正态分布。进一步, 误差必须是相互独立的。

可以得出, 在上述模型条件下, 在第  $i$  行第  $j$  列试验单元上记录到的响应值服从均值为

$$E(y_{ij}) = \mu + \alpha_k + \beta_i + \gamma_j$$

方差为  $\sigma_e^2$  的正态分布。注意, 这里我们不必指明试验单元所接受的处理, 因为一旦我们知道了试验单元处于哪行哪列, 它所接受的处理也就知道了。

我们可以用这个模型来说明拉丁方设计如何滤除了由行和列中的变异来源产生的变异。为此, 我们考虑一个有  $t=4$  处理(I, II, III, IV)的拉丁方设计, 其中两个外部变异来源各有 4 个水平。该设计在表 15.16 中给出。

表 15.16 一个  $4 \times 4$  拉丁方设计

| 行 | 列   |     |     |     |
|---|-----|-----|-----|-----|
|   | 1   | 2   | 3   | 4   |
| 1 | I   | II  | III | IV  |
| 2 | II  | III | IV  | I   |
| 3 | III | IV  | I   | II  |
| 4 | IV  | I   | II  | III |

如果我们希望用样本均值的差  $\bar{y}_3 - \bar{y}_1$  来估计  $\alpha_3 - \alpha_1$ , 即处理 III 和处理 I 的平均响应的差, 那么, 可以从模型得到  $\bar{y}_3$  和  $\bar{y}_1$  的表达式, 计算时留心这些处理出现在哪一行哪一列。以  $y_{ij}$  记第  $i$  行第  $j$  列的观测值, 由表 15.16 知

$$\begin{aligned} \bar{y}_1 &= \frac{1}{4}(y_{11} + y_{24} + y_{33} + y_{42}) \\ &= \mu + \alpha_1 + \frac{1}{4}(\beta_1 + \beta_2 + \beta_3 + \beta_4) + \frac{1}{4}(\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4) + \bar{\varepsilon}_1 \end{aligned}$$

其中  $\bar{\epsilon}_1$  是处理 I 上四个观测值的随机误差的平均值。同样,

$$\begin{aligned} y_3 &= \frac{1}{4}(y_{13} + y_{22} + y_{31} + y_{44}) \\ &= \mu + \alpha_3 + \frac{1}{4}(\beta_1 + \beta_2 + \beta_3 + \beta_4) + \frac{1}{4}(\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4) + \bar{\epsilon}_3 \end{aligned}$$

样本均值的差为

$$\bar{y}_3 - \bar{y}_1 = \alpha_3 - \alpha_1 + (\bar{\epsilon}_3 - \bar{\epsilon}_1)$$

$\alpha_3 - \alpha_1$  的估计误差是  $\bar{\epsilon}_3 - \bar{\epsilon}_1$ 。

如果使用了随机化区组设计,并且在其中以行作为区组,那么处理仅在各行内被随机化,这就很可能有某些处理在同一列中出现多于一次,样本均值的差也就成为

$$\bar{y}_3 - \bar{y}_1 = \alpha_3 - \alpha_1 + [(\bar{\epsilon}_3 - \bar{\epsilon}_1) + (\text{消不去的列效应})]$$

由此,估计的误差会由于这些消不去的列效应而膨胀。由于同样的原因,如果在拉丁方设计适用的场合使用了完全随机化设计,那么估计的误差也会由于消不去的行列效应而膨胀。

我们可以对给定的关于模型中参数的假设进行检验。特别地,我们可能希望检验  $t$  个处理均值之间没有差异这样一个假设。这个假设可以叙述为

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0 \quad (\text{即 } t \text{ 个处理均值是相等的})$$

备择假设为

$$H_a: \text{至少有一个 } \alpha_i \text{ 不等于 } 0 (\text{即至少有一个处理均值与其他均值不同})$$

通过考察拉丁方设计的模型,并把总平方和分解为处理效应、行效应、列效应和随机效应几项,我们可以得到该假设的检验统计量。

各观测值关于它们的平均值的总平方和,与以前一样,定义为

$$\text{TSS} = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$$

这个平方和将被分解成四个不同的变异来源:一个是源于处理之间的变异,一个是源于行之间的变异,一个是源于列之间的变异,另一个则是源于既不能被处理差异也不能被行列差异所解释的所有其他的变异,称之为误差。TSS 的分解可以通过对拉丁方设计的模型

$$y_{ij} = \mu + \alpha_k + \beta_i + \gamma_j + \epsilon_{ij}$$

的考察得到。模型中的参数有样本估计:

$$\hat{\mu} = \bar{y}_{..} \quad \hat{\alpha}_k = \bar{y}_{.k} - \bar{y}_{..} \quad \hat{\beta}_i = \bar{y}_{i.} - \bar{y}_{..} \quad \hat{\gamma}_j = \bar{y}_{.j} - \bar{y}_{..}$$

$$\text{TSS} = t \sum_k (\bar{y}_{.k} - \bar{y}_{..})^2 + t \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 + t \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 + \text{SSE}$$

我们用上面的参数估计来解释这个分解中的各项。等号右边第一项度量了各处理均值  $\bar{y}_{.k}$  相对于总平均值  $\bar{y}_{..}$  的变异性。因此

$$SST = t \sum_k (\bar{y}_{.k} - \bar{y}_{..})^2 = t \sum_k (\hat{\alpha}_k)^2$$

称为**处理间平方和**,它是各个  $y_{ij}$  中由于处理均值之间的差异而产生的变异性的一个度量。第二项

$$SSR = t \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = t \sum_i (\hat{\beta}_i)^2$$

度量了各个行均值  $\bar{y}_{i.}$  相对于总平均值的变异性,称之为**行间平方和**。第三个变异来源,称为**列间平方和**,度量了各个列均值  $\bar{y}_{.j}$  相对于总平均值的变异性,它由

$$SSC = t \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 = t \sum_j (\hat{\gamma}_j)^2$$

给出。最后的变异来源称为**误差平方和**,SSE 表示各  $y_{ij}$  中没有被处理和行、列差异所解释的变异性。这一项由

$$SSE = TSS - SST - SSR - SSC$$

给出。我们可以把这些计算总结成如表 15.17 的 AOV 表。

表 15.17  $t \times t$  拉丁方设计的方差分析表

| 来源 | SS  | df               | MS                         | F         |
|----|-----|------------------|----------------------------|-----------|
| 处理 | SST | $t - 1$          | $MST = SST/(t - 1)$        | $MST/MSE$ |
| 行  | SSR | $t - 1$          | $MSR = SSR/(t - 1)$        | $MSR/MSE$ |
| 列  | SSC | $t - 1$          | $MSC = SSC/(t - 1)$        | $MSC/MSE$ |
| 误差 | SSE | $(t - 1)(t - 2)$ | $MSE = SSE/(t - 1)(t - 2)$ |           |
| 总和 | TSS | $t^2 - 1$        |                            |           |

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0 \quad H_a: \text{至少有一个 } \alpha_k \text{ 不等于 } 0$$

的检验统计量为比值

$$F = \frac{MST}{MSE}$$

对于我们的模型,有

$$E(MSE) = \sigma_e^2 \quad \text{和} \quad E(MST) = \sigma_e^2 + t\theta_T$$

其中  $\theta_T = [1/(t - 1)] \sum_k \alpha_k^2$ 。当  $H_0$  为真时,  $\alpha_k$  等于 0 对于一切  $k = 1, \dots, t$  成立,从而  $\theta_T = 0$ 。由此,当  $H_0$  为真时,我们可以期望  $MST/MSE$  的值接近于 1。然而,在研究假设  $H_a$  下,  $\theta_T$  大于 0,因为至少有一个  $\alpha_k$  不为 0。因此,处理均值中大的差异会导致  $\theta_T$  有大的取值,  $MST$  的期望值将比  $MSE$  的期望值来得大。这样,我们可以期望  $F = MST/MSE$  比 1 大。于是,当我们观测到的  $F$  值大于一个  $F$  分布的上尾部的值时,就拒绝  $H_0$ 。



对于一个取定的犯第一类错误的概率, 上述讨论给出了下面的判决规则:

拒绝  $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0$  当  $F = \text{MST}/\text{MSE}$  大于  $F_{\alpha, df_1, df_2}$

这里,  $F_{\alpha, df_1, df_2}$  是用  $\alpha$  = 取定的犯第一类错误的概率,  $df_1 = df_{\text{MST}} = t - 1$  和  $df_2 = df_{\text{MSE}} = (t - 1)(t - 2)$  从附录的表 8 中查出的。另一种方法是利用

$$p \text{ 值} = P(F_{df_1, df_2} > F_{\text{obs}})$$

计算检验统计量的  $p$  值, 其中的概率是用自由度为  $df_1 = t - 1$  和  $df_2 = (t - 1)(t - 2)$  的  $F$  分布计算的。计算出  $p$  值后, 我们比较  $p$  值与取定的犯第一类错误的概率, 小的  $p$  值支持研究假设, 而对大的  $p$  值, 我们则不能拒绝  $H_0$ 。

一般来说, 仅当考察行列划分是否有效地减少了试验单元的变异性时, 才讨论行和列的效应, 因而一般不检验关于行列效应的假设。像在随机化区组设计中一样, 我们可以与完全随机化设计比较效率。我们想知道, 在一个试验中, 考虑行和列的变异来源是否提高了处理均值比较的精度。以  $\text{MSE}_{\text{LS}}$  和  $\text{MSE}_{\text{CR}}$  分别记一个拉丁方设计和一个完全随机化设计的均方误差, 以  $\text{RE}(\text{LS}, \text{CR})$  记与完全随机化设计进行比较时拉丁方设计的相对效率。我们可以用拉丁方设计中的均方  $\text{MSR}$ ,  $\text{MSC}$  和  $\text{MSE}$ , 通过下列公式来计算相对效率  $\text{RE}(\text{LS}, \text{CR})$ :

$$\text{RE}(\text{LS}, \text{CR}) = \frac{\text{MSE}_{\text{LS}}}{\text{MSE}_{\text{CR}}} = \frac{\text{MSR} + \text{MSC} + (t - 1)\text{MSE}}{(t + 1)(\text{MSE})}$$

当  $\text{RE}(\text{LS}, \text{CR})$  远比 1 大时, 我们就说行和(或)列的变异来源是有效的, 因为要想在估计处理均值时获得同样的精度, 在完全随机化设计中比在拉丁方设计中需要更多的观测值。

### 例 15.5

例 15.4 中那家律师事物所进行了评价电子表格软件的试验, 所得数据见表 15.18。利用这些数据回答下列问题。

- 对该试验, 写出一个适当的模型。
- 为比较用这些电子表格软件完成任务所需的平均时间, 做方差分析。取  $\alpha = 0.05$ 。
- 把你的结果总结在 AOV 表中。
- 计算拉丁方设计相对于完全随机化设计的相对效率。

**解答** 我们看出, 该试验设计是一个有  $t = 4$  行(问题),  $t = 4$  列(秘书)和  $t = 4$  处理(电子表格软件)的拉丁方设计。一个适当的统计模型为

$$y_{ij} = \mu + \alpha_k + \beta_i + \gamma_j + \varepsilon_{ij} \quad i, j, k = 1, 2, 3, 4$$

从表 15.18 可知, 我们可以用  $\hat{\mu}_k = \bar{y}_k$  来估计处理效应  $\mu_k$ , 所得估计值为

表 15.18 完成任务所需要的时间(小时):拉丁方设计

| 问题  | 秘书     |        |        |        | 行均值  | 电子表格<br>软件平均 |
|-----|--------|--------|--------|--------|------|--------------|
|     | 1      | 2      | 3      | 4      |      |              |
| I   | A(0.3) | B(1.8) | C(0.7) | D(1.2) | 1.0  | A:0.45       |
| II  | B(1.4) | C(1.4) | D(1.1) | A(0.5) | 1.1  | B:1.5        |
| III | C(0.5) | D(1.5) | A(0.5) | B(1.1) | 0.9  | C:1.05       |
| IV  | D(1.0) | A(0.5) | B(1.7) | C(1.6) | 1.2  | D:1.2        |
| 列均值 | 0.8    | 1.3    | 1.0    | 1.1    | 1.05 |              |

$$\hat{\mu}_k = 0.45 \quad \hat{\mu}_2 = 1.5 \quad \hat{\mu}_3 = 1.05 \quad \hat{\mu}_4 = 1.2$$

看起来用电子表格软件 A 完成任务所花的平均时间较其他三种软件短一些。现在我们来估计模型参数,并构造方差分析表。回忆  $\hat{\mu} = \bar{y}_{...}$ ,  $\hat{\alpha}_k = \bar{y}_{.k} - \bar{y}_{...}$ ,  $\hat{\beta}_i = \bar{y}_{i.} - \bar{y}_{...}$  和  $\hat{\gamma}_j = \bar{y}_{.j} - \bar{y}_{...}$ , 由  $\hat{\mu} = \bar{y}_{...} = 1.05$ , 我们得到

电子表格软件效应

问题效应

秘书效应

$$\hat{\alpha}_1 = 0.45 - 1.05 = -0.6 \quad \hat{\beta}_1 = 1 - 1.05 = -0.05 \quad \hat{\gamma}_1 = 0.8 - 1.05 = -0.25$$

$$\hat{\alpha}_2 = 1.5 - 1.05 = 0.45 \quad \hat{\beta}_2 = 1.1 - 1.05 = 0.05 \quad \hat{\gamma}_2 = 1.3 - 1.05 = 0.25$$

$$\hat{\alpha}_3 = 1.05 - 1.05 = 0 \quad \hat{\beta}_3 = 0.9 - 1.05 = -0.15 \quad \hat{\gamma}_3 = 1 - 1.05 = -0.05$$

$$\hat{\alpha}_4 = 1.2 - 1.05 = 0.15 \quad \hat{\beta}_4 = 1.2 - 1.05 = 0.15 \quad \hat{\gamma}_4 = 1.1 - 1.05 = 0.05$$

代入到平方和公式,有

$$\begin{aligned} \text{TSS} &= \sum_{ij} (y_{ij} - \bar{y}_{...})^2 = (0.3 - 1.05)^2 + (1.8 - 1.05)^2 + \cdots + (1.6 - 1.05)^2 \\ &= 3.66 \end{aligned}$$

$$\text{SST} = t \sum_k (\hat{\alpha}_k)^2 = 4[(-0.6)^2 + (0.45)^2 + (0)^2 + (0.15)^2] = 2.34$$

$$\text{SSR} = t \sum_i (\hat{\beta}_i)^2 = 4[(-0.05)^2 + (0.05)^2 + (-0.15)^2 + (0.15)^2] = 0.2$$

$$\text{SSC} = t \sum_j (\hat{\gamma}_j)^2 = 4[(-0.25)^2 + (0.25)^2 + (-0.05)^2 + (-0.05)^2] = 0.52$$

相减得

$$\text{SSE} = \text{TSS} - \text{SST} - \text{SSB} - \text{SSC} = 3.66 - 2.34 - 0.2 - 0.52 = 0.6$$

表 15.19 中的方差分析表总结了上述结果。对于方差分析表中的每一种来源,其均方由该来源上的平方和除以其自由度得到。

表 15.19 例 15.5 中电子表格软件研究的 AOV 表

| 来源     | SS   | df | MS    | F    | p 值    |
|--------|------|----|-------|------|--------|
| 电子表格软件 | 2.34 | 3  | 0.78  | 7.8  | 0.0171 |
| 问题     | 0.2  | 3  | 0.067 | 0.67 | 0.6025 |
| 秘书     | 0.52 | 3  | 0.173 | 1.73 | 0.2592 |
| 误差     | 0.6  | 6  | 0.1   |      |        |
| 总和     | 3.66 | 15 |       |      |        |

对处理均值差异的  $F$  检验,即假设

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0 \quad H_a: \text{至少有一个 } \alpha_k \text{ 不等于 } 0$$

的  $F$  检验使用  $F$  统计量  $MST/MSE$ 。由于计算所得  $F$  值 7.8 大于根据  $df_1 = 3, df_2 = 6$  和  $\alpha = 0.05$  从  $F$  分布表查得的值,我们拒绝  $H_0$ ,认为四种电子表格软件的平均完成时间之间有显著差异( $p$  值 = 0.0171)。看来使用软件 A 比起其他三种软件平均完成任务的时间要短得多。

下面我们看一看考虑了这两个外部变异来源,相对于完全随机化设计是否有效地提高了分析的精度。由上面的 AOV 表,我们有  $MSR = 0.067, MSC = 0.173$ , 以及  $MSE = 0.1$ 。因此,该拉丁方设计相对于完全随机化设计的相对效率为

$$\begin{aligned} RE(LS, CR) &= \frac{MSR + MSC + (t-1)MSE}{(t+1)MSE} \\ &= \frac{0.067 + 0.173 + (4-1)(0.1)}{(4+1)(0.1)} = 1.08 \end{aligned}$$

这就是说,要在完全随机化设计中使得处理均值的估计与在拉丁方设计中具有同样的精度,每个处理需要增加 8% 的观测值。该拉丁方设计对于精度的改进没有那么高,是由于不同类型的问题的完成时间差别很小。不同秘书的平均完成时间之间的差别大一些,但这些差别对总的试验结果来说没有产生大的变异。

例 15.5 中的结果,仅在我们能够确信模型条件和观测到的数据相一致时才是有效的。因此,我们用残差  $e_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_k - \hat{\beta}_i - \hat{\gamma}_j$  来检查对于这些观测到的数据而言,正态性条件、等方差条件和独立性条件是否近似地得到了满足。下面的例子给出了这样一个分析的计算机输出结果。

### 例 15.6

这里给出例 15.5 中描述的试验的计算机输出结果。把这些结果与用平方和定义算得的结果进行比较,并说明模型条件是否是有效的。

**解答** 注意我们手算的结果与计算机输出的结果基本上相同, 差别是由舍入误差造成的。在我们的手工计算中, 可能会有大的舍入误差, 而这些舍入误差可能导致所得结果与计算机输出的结果不相一致。我们强烈推荐在方差分析的计算中使用计算机软件程序, 以避免潜在的舍入误差。为了检查是否满足模型条件, 我们首先注意, 对于正态性条件, 假设  $H_0$ : “残差服从正态分布”的 Shapiro-Wilks 检验的  $p$  值为  $p$  值 = 0.9098. 因而, 我们不能拒绝  $H_0$ , 正态性条件看上去得到了满足。此外, 茎叶图、盒型图和正态概率图也显示出与“残差服从正态分布”这个条件的一致性。图 15.3 是一张残差-估计处理均值图。从这张图上看, 残差的变异随样本均值变化不大。

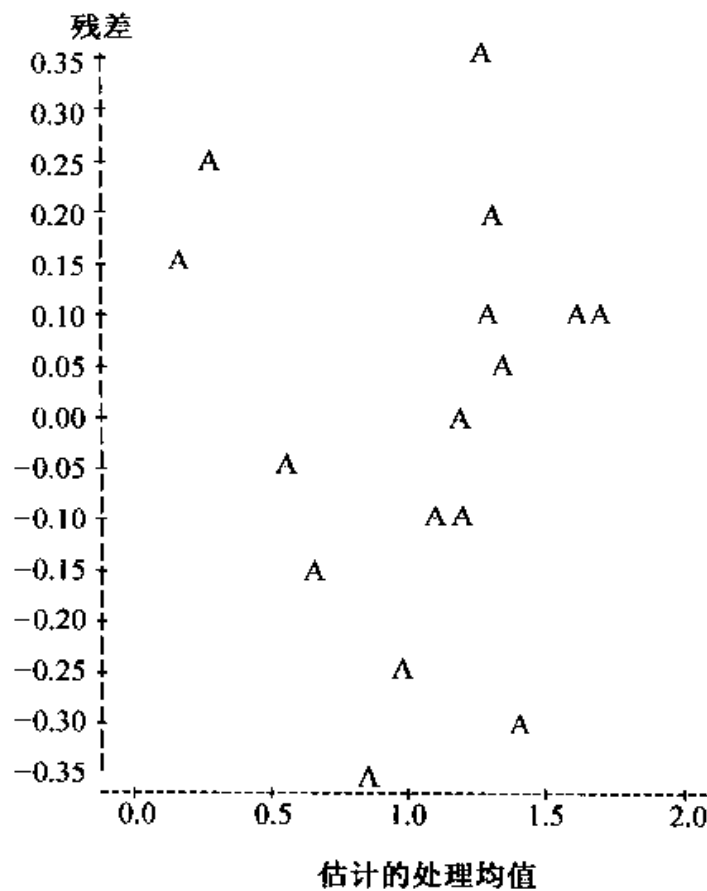


图 15.3 残差-估计处理均值图

General Linear Models Procedure For Example 15.5

Dependent Variable: TIME TO COMPLETION

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|--------|----|----------------|-------------|---------|--------|
| Model  | 9  | 3.0600000      | 0.3400000   | 3.40    | 0.0751 |
| Error  | 6  | 0.6000000      | 0.1000000   |         |        |

|                 |          |           |             |         |        |
|-----------------|----------|-----------|-------------|---------|--------|
| Corrected Total | 15       | 3.6600000 |             |         |        |
|                 | R-Square | C. V.     | Root MSE    |         | Y Mean |
|                 | 0.836066 | 30.11693  | 0.3162      |         | 1.0500 |
| Source          | DF       | Type I SS | Mean Square | F Value | Pr > F |
| R               | 3        | 0.2000000 | 0.0666667   | 0.67    | 0.6025 |
| C               | 3        | 0.5200000 | 0.1733333   | 1.73    | 0.2592 |
| T               | 3        | 2.3400000 | 0.7800000   | 7.80    | 0.0171 |

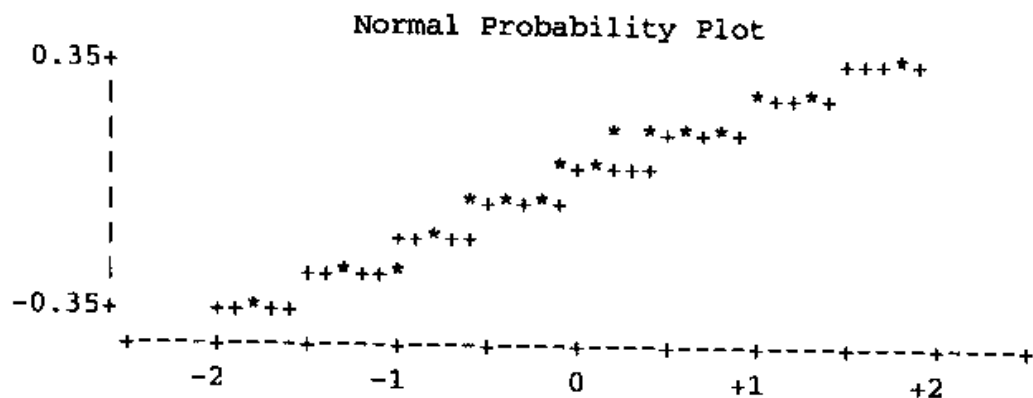
# RESIDUAL ANALYSIS:

Variable=RESIDUAL

|          |          |          |  |         |
|----------|----------|----------|--|---------|
|          | Moments  |          |  |         |
| N        | 16       | Sum Wgts |  | 16      |
| Mean     | 0        | Sum      |  | 0       |
| Std Dev  | 0.2      | Variance |  | 0.04    |
| Skewness | -0.17143 | Kurtosis |  | -0.6544 |
| W:Normal | 0.97711  | Pr < W   |  | 0.9098  |

Variable = RESIDUALS

| Stem                         | Leaf | # | Boxplot   |
|------------------------------|------|---|-----------|
| 3                            | 5    | 1 |           |
| 2                            | 05   | 2 |           |
| 1                            | 0005 | 4 | +-----+   |
| 0                            | 05   | 2 | *---+---* |
| -0                           | 5    | 1 |           |
| -1                           | 500  | 3 | +-----+   |
| -2                           | 5    | 1 |           |
| -3                           | 50   | 2 |           |
| -----+-----+-----+-----+     |      |   |           |
| Multiply Stem.Leaf by 10**-1 |      |   |           |



## 练习

## 应用

**15.9(农业)** 在西瓜田中做一项试验,以比较两种施肥区位(广播法和带状法)和两种不同的施肥量。最近的研究表明,在西瓜田中广播施肥方法(在要施肥的区域上广泛播撒)比施肥于种子附近的带状施肥法好。在该试验中,研究人员希望在每英亩施以 160-70-135 磅的施肥量上,比较两种氮-磷-钾肥的施肥区位(广播法和带状法)以及两种品牌(A 和 B)的微量元素。对两种区位和两种微量元素的四种组合用拉丁方田间试验。

把一块大面积的瓜田分成行和列,并在这块田上根据一个拉丁方设计把这些处理随机地进行分配。对于每一行列的组合,在西瓜出苗 30 天后,测量一棵西瓜秧苗的晾干重量。数据如下。

| 行 | 列 |      |   |      |   |      |   |      |
|---|---|------|---|------|---|------|---|------|
|   | 1 |      | 2 |      | 3 |      | 4 |      |
| 1 | 1 | 1.75 | 3 | 1.43 | 4 | 1.28 | 2 | 1.66 |
| 2 | 2 | 1.70 | 1 | 1.78 | 3 | 1.40 | 4 | 1.31 |
| 3 | 4 | 1.35 | 2 | 1.73 | 1 | 1.69 | 3 | 1.41 |
| 4 | 3 | 1.45 | 4 | 1.36 | 2 | 1.65 | 1 | 1.73 |

处理 1: 广播, A

处理 3: 带状, A

处理 2: 广播, B

处理 4: 带状, B

a. 对于该试验, 写出一个适当的模型。

b. 用这些数据做方差分析, 给出每一个检验的  $p$  值和所得到的结论。

**15.10(工程)** 一家石油公司对比较四种不同汽油(A, B, C 和 D)每加仑所能跑的英里数感兴趣。由于在驾驶特征和车辆品牌上可能存在很大的差异, 这两个外部变异来源被用做研究中的“区组”变量。研究人员挑选了四种不同品牌的汽车和四位司机。按照下面表中给出的方式, 把汽油分配给这些司机和汽车。在每次试验后记录下每加仑汽油的平均里程(单位: mpg(英里/加仑))如下。

| 司机 | 汽车品牌    |         |         |         |
|----|---------|---------|---------|---------|
|    | 1       | 2       | 3       | 4       |
| 1  | A(15.5) | B(33.8) | C(13.7) | D(29.2) |
| 2  | B(16.3) | C(26.4) | D(19.1) | A(22.5) |
| 3  | C(10.5) | D(31.5) | A(17.5) | B(30.1) |
| 4  | D(14.0) | A(34.5) | B(19.7) | C(21.6) |

- a. 对于该试验, 写出一个模型。
- b. 估计模型中的参数。
- c. 做方差分析, 取  $\alpha = 0.05$ 。
- d. 关于最好的汽油, 你能得到什么结论?
- e. 计算拉丁方设计相对于完全随机化设计的相对效率, 并解释所得到的值。  
在减少试验单元的变异性方面, 这些区组变量是有效的吗? 为什么?
- f. 在将来的试验中, 你是否仍然推荐把汽车品牌 and 司机都作为区组变量? 为什么?

15.11 下面给出了练习 15.10 中的数据的计算机输出结果。

- a. 比较这里给出的结果和你在练习 15.10 中得到的结果。
- b. 对于这些数据, 模型条件看上去得到了满足吗? 为什么?

| OBS | R | C | T | Y    |
|-----|---|---|---|------|
| 1   | 1 | 1 | 1 | 15.5 |
| 2   | 1 | 2 | 2 | 33.8 |
| 3   | 1 | 3 | 3 | 13.7 |
| 4   | 1 | 4 | 4 | 29.2 |
| 5   | 2 | 1 | 2 | 16.3 |
| 6   | 2 | 2 | 3 | 26.4 |
| 7   | 2 | 3 | 4 | 19.1 |
| 8   | 2 | 4 | 1 | 22.5 |
| 9   | 3 | 1 | 3 | 10.5 |
| 10  | 3 | 2 | 4 | 31.5 |
| 11  | 3 | 3 | 1 | 17.5 |
| 12  | 3 | 4 | 2 | 30.1 |
| 13  | 4 | 1 | 4 | 14.0 |
| 14  | 4 | 2 | 1 | 34.5 |
| 15  | 4 | 3 | 2 | 19.7 |
| 16  | 4 | 4 | 3 | 21.6 |

General Linear Models Procedure

Dependent Variable: MILES PER GALLON

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|--------|----|----------------|-------------|---------|--------|
| Model  | 9  | 869.97563      | 96.66396    | 22.42   | 0.0006 |

|                 |    |           |         |
|-----------------|----|-----------|---------|
| Error           | 6  | 25.86375  | 4.31062 |
| Corrected Total | 15 | 895.83937 |         |

|          |          |          |        |
|----------|----------|----------|--------|
| R-Square | C. V.    | Root MSE | Y Mean |
| 0.971129 | 9.333878 | 2.0762   | 22.244 |

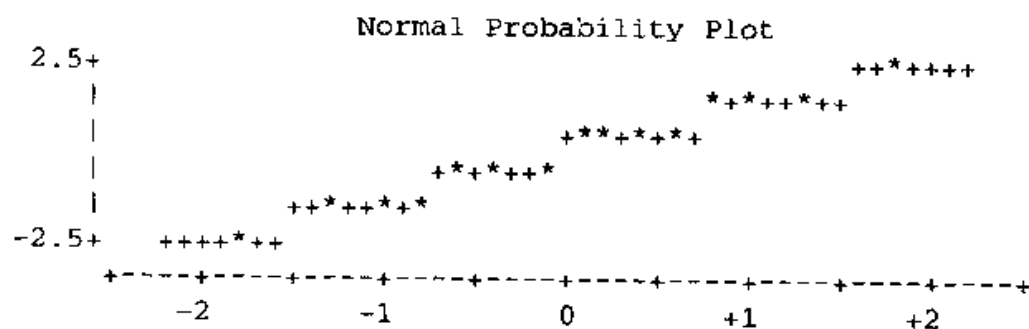
| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|--------|----|-----------|-------------|---------|--------|
| R      | 3  | 8.33187   | 2.77729     | 0.64    | 0.6143 |
| C      | 3  | 755.37188 | 251.79063   | 58.41   | 0.0001 |
| T      | 3  | 106.27188 | 35.42396    | 8.22    | 0.0151 |

## RESIDUALS ANALYSIS:

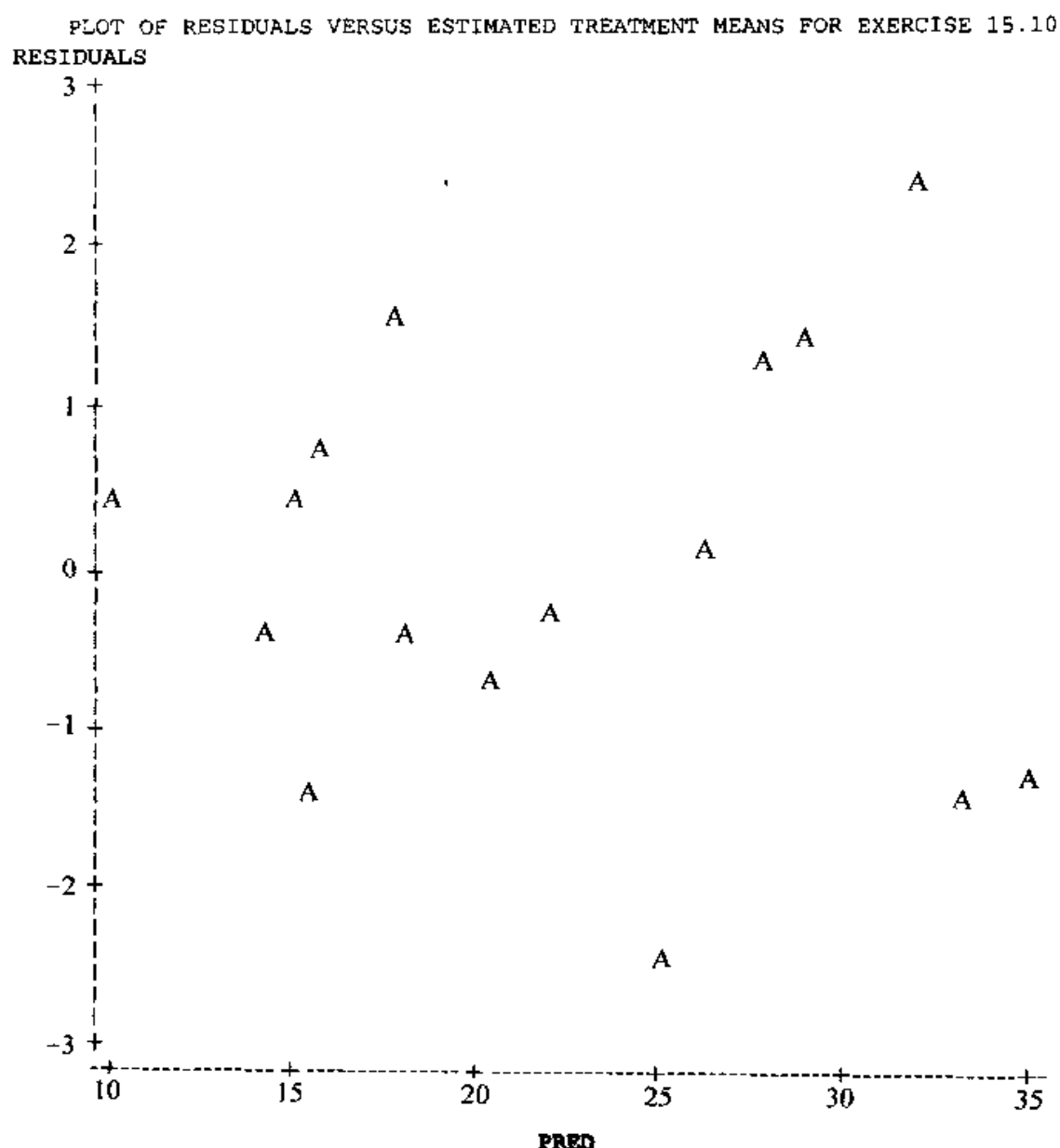
Variable = RESIDUALS

| Moments   |          |          |         |
|-----------|----------|----------|---------|
| N         | 16       | Sum Wgts | 16      |
| Mean      | 0        | Sum      | 0       |
| Std Dev   | 1.313107 | Variance | 1.72425 |
| Skewness  | 0.043408 | Kurtosis | -0.41   |
| W; Normal | 0.985867 | Pr < W   | 0.9840  |

| Stem Leaf                | # | Boxplot |
|--------------------------|---|---------|
| 2 5                      | 1 |         |
| 1 346                    | 3 | +-----+ |
| 0 2457                   | 4 | +       |
| -0 7443                  | 4 | *-----* |
| -1 543                   | 3 | +-----+ |
| -2 4                     | 1 |         |
| -----+-----+-----+-----+ |   |         |







## 15.5 完全随机化设计中的因子处理结构

在第十四章,我们介绍了试验设计中的两个要素,即随机化方法和处理结构。在15.2节,15.3节和15.4节中,我们考虑了随机化技术以及通过使用区组变量实现对外部变异来源的控制。完全随机化设计被用来比较 $t$ 个处理,其中的试验单元是齐性的,因而处理对试验单元的随机化没有任何约束。15.3节和15.4节讨论了随机化区组设计和拉丁方设计。在这两个设计中,试验单元不是齐性的,必须在随机分配处理之前把它们分成由齐性试验单元构成的组(区组)。这就导致了一个在完全随机化设计中不曾出现的加之于随机化上的约束。在随机化完全区组

设计中,涉及到识别试验单元的一个特征(区组变量),而在拉丁方设计中,则允许试验者在建立齐性的试验单元组时使用试验单元的两个不同的特征。

在这一节,我们来讨论处理设计,即如何从几个因子,而不是仅从单个因子的几个水平来构造处理。这些类型的试验,出现在要对两个或更多自变量对响应变量  $y$  的影响进行考察的情形。例如,假定一家公司开发了一种新型家用胶黏剂,要考察温度和湿度对于这种胶黏剂的黏结力的影响。在任何一项研究中,都会出现几个与处理设计有关的问题。首先,我们必须考虑什么样的因子(自变量)是最感兴趣的。其次,对于每一个因子,必须决定其水平数以及各个水平的实际设置。第三,挑选出各个因子的水平后,我们必须选择在试验中使用的各个因子的水平组合。

因子的选择以及对每一个因子其适当的水平的选择,与预算、完成研究的时间有关,最重要的,与试验者对于当前研究的实际情况的了解有关。在很多时候,需要对文献进行详细的查阅,以确定在当前这个研究所属的领域内知识的现状。当试验者选定了每个自变量的水平以后,他或她必须决定哪些因子水平组合是最感兴趣的并且是可行的。在某些情况下,有的因子水平组合不产生能够在试验单元上给出合理的响应的试验设置,一些组合可能由于毒性或实际操作困难而不适宜用于试验中。

如第二章所讨论的,考察两个或更多因子对于响应的效应的一种方法是一次一因子方法。为了考察某个变量的效应,试验者变化该变量的水平,而把其余自变量的水平固定。在保持其余自变量不变的情况下,对每个变量都进行这个过程。假设试验者要考察两个自变量氮和磷对于某种作物产量的影响。为简单起见,假定在该试验中为每个变量挑选了两个水平:氮的施肥量为每块地 40 磅和 60 磅,磷的施肥量为每块地 10 磅和 20 磅。研究中的试验单元是把某个农场中的一块地划分为小块得来的,面积不大,也相当一致。对于这个试验,选择的因子水平组合可能如表 15.20。图 15.4 是这些因子水平组合的一个图示。

表 15.20 一次一因子方法的因子水平组合

| 组合 | 氮  | 磷  |
|----|----|----|
| 1  | 60 | 10 |
| 2  | 40 | 10 |
| 3  | 40 | 20 |

从图 15.4 可以看出,分别有一个差可以用来估计氮的效应和磷的效应。组合 1 和组合 2 的响应值的差可以估计氮的效应,组合 2 和组合 3 的响应值的差可以

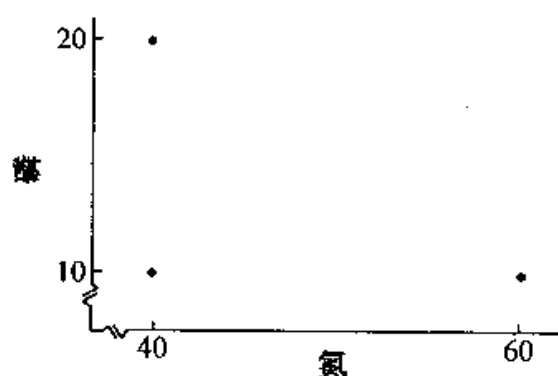


图 15.4 一次一因子方法的因子水平组合

估计磷的效应。

表 15.21 给出的是该试验中相应于这三个因子水平组合的虚拟的产量。假定试验者要用样本信息确定给出最大产量的因子水平组合。从表中我们看出,当氮的用量从 40 增加到 60(保持磷肥用量为 10)时,该作物的产量增加了。而当磷肥用量从 10 增加到 20(保持氮肥用量为 40)时,作物的产量也增加了。由此看来,说“同时增加氮肥和磷肥施用量会得到更高的作物产量”似乎不无道理。问题在于这个断言基于这样一个假设,即一个因子的效应在另一个因子的两个水平上是相同的。

表 15.21 三个因子水平组合的产量

| 观测值<br>(产量) | 氮  | 磷  |
|-------------|----|----|
| 145         | 60 | 10 |
| 125         | 40 | 10 |
| 160         | 40 | 20 |
| ?           | 60 | 20 |

上面我们已经看到,当磷肥的用量为 10 时,把氮肥的用量从 40 变到 60 产量发生了什么变化。但当磷肥的用量为 20 时,把氮肥的用量从 40 变到 60 产量也会增产大约 20 个单位吗?

为了回答这个问题,我们应当把氮肥 60 磷肥 20 的因子水平组合应用到另一块试验用地并观察其作物产量。如果这个产量是 180,那么从以上三个因子水平组合得到的信息就是正确的,从而可以用来预测具有最大产量的因子水平组合。现在,假定按氮肥和磷肥的高用量施肥,结果得到的产量为 110。如果真是这样,

我们称两个因子氮和磷有交互效应。也就是说,一个因子对于响应变量的效应在另一个因子的不同水平上是不同的,此时,用一次一因子法得到的信息将给出一个错误的预报。

上面的两个在 60-20 因子水平组合上的作物产量与三个原始设计点上的产量一起在图 15.5 中标示出来。图 15.5(a)说明了两个因子之间没有交互效应的情况。氮对于产量的效应在两个磷的水平上是相同的。相反,图 15.5(b)则显示了一种两个因子交互效应的情况。

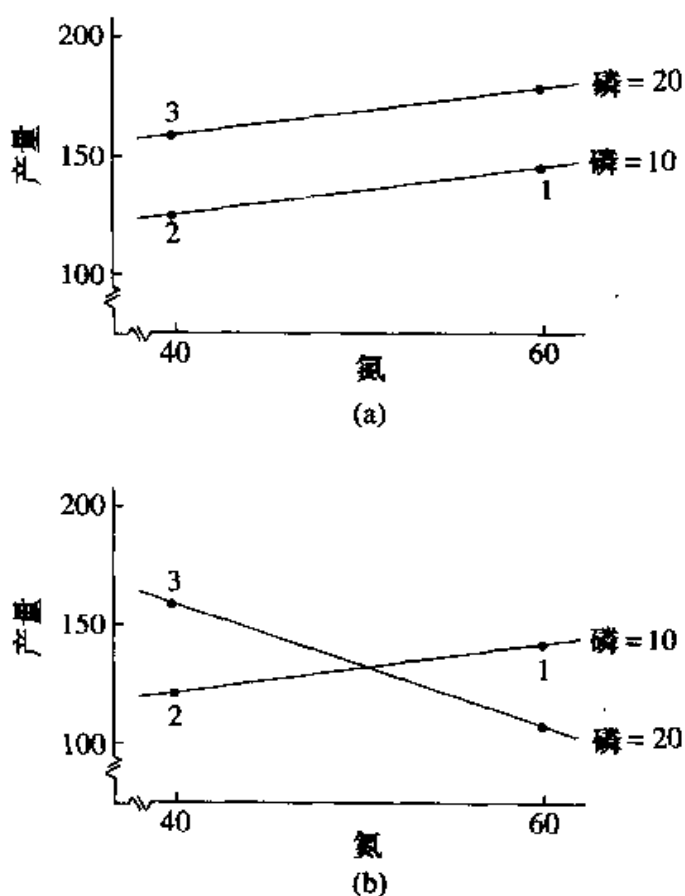


图 15.5 三个设计点上的产量和第四个设计点上可能的产量

我们看到,调查两个因子对于响应变量的效应的一次一因子法仅适用于这两个因子没有交互效应的情况。虽然这一点是通过一个有两个因子,每个因子各有两个水平的简单例子说明的,但一次一因子法的这一不足,在讨论两个以上因子对于响应变量的效应时更加突出。

不论交互效应存在与否,析因试验对于考察两个或更多因子在响应值  $y$  上的效应都是有用的。如前所述,每个变量的水平个数的选择以及这些变量的实际设置是很重要的。但现在,我们假定在一位熟知该研究领域的研究人员的帮助下,已

经做出了这些选择,那么我们必须决定在哪些因子水平组合上对  $y$  进行观测。

按照经典的说法,析因试验不被称为设计,因为析因试验讲究因子水平的选择和因子水平组合的挑选,但不讨论如何把处理分配给试验单元。除非另有说明,我们假定处理是随机地分配给试验单元的,因而因子水平组合相当于完全随机化设计中的“处理”。

**定义 15.3** 析因试验是指在自变量的所有因子水平组合上都对响应变量  $y$  进行了观测的试验。

看我们前面的例子。如果我们有兴趣考察每块地施用 40 和 60 磅这两个氮肥( $x_1$ )水平以及每块地施用 10 和 20 磅这两个磷肥( $x_2$ )水平对于作物产量的效应,就使用一个完全随机化设计,把表 15.22 给出的四个因子水平组合(处理)随机地分配给试验单元。

表 15.22 作物产量的  $2 \times 2$  析因试验

| 因子水   | 平组合   |
|-------|-------|
| $x_1$ | $x_2$ |
| 40    | 10    |
| 40    | 20    |
| 60    | 10    |
| 60    | 20    |

同样地,如果我们希望在两个水平 40 和 60 上考察  $x_1$ ,在三个水平 10, 15 和 20 上考察  $x_2$ ,我们就在完全随机化设计中使用表 15.23 中的六个因子水平组合。

表 15.23 作物产量的  $2 \times 3$  析因试验

| 因子水平  | 组合    |
|-------|-------|
| $x_1$ | $x_2$ |
| 40    | 10    |
| 40    | 15    |
| 40    | 20    |
| 60    | 10    |
| 60    | 15    |
| 60    | 20    |

**例 15.7**

某汽车制造厂要考察发动机转速  $x_1$  (以每分钟的转数来度量) 和地面上的行驶速度  $x_2$  (以每小时的英里数来度量) 对于每加仑汽油所能行驶里程的影响。研究人员在咨询了公司的机械师和其他人员后, 决定考察  $x_1$  的三个水平 800, 1 000 和 1 200 以及  $x_2$  的三个水平 30, 50 和 70。试给出在一个  $3 \times 3$  析因试验中所使用的因子水平组合。

**解答** 由析因试验的定义, 我们将在下列  $x_1$  和  $x_2$  的组合上观测每加仑汽油所能行驶的里程:

|       |     |     |     |       |       |       |       |       |       |
|-------|-----|-----|-----|-------|-------|-------|-------|-------|-------|
| $x_1$ | 800 | 800 | 800 | 1 000 | 1 000 | 1 000 | 1 200 | 1 200 | 1 200 |
| $x_2$ | 30  | 50  | 70  | 30    | 50    | 70    | 30    | 50    | 70    |

本节中所给出的析因试验的例子都有两个自变量。然而, 这里的方法适用于任意个数的因子和因子的任意个水平。比如, 假若我们有四个不同的因子  $x_1, x_2, x_3$  和  $x_4$ , 它们的水平个数分别为 2, 3, 3 和 4, 那么就可以形成一个  $2 \times 3 \times 3 \times 4$  析因试验, 其中有  $2 \cdot 3 \cdot 3 \cdot 4 = 72$  个因子水平组合。

对于一次一因子方法和析因试验方法应该做一个最后的比较。用析因试验不仅可以得到因子交互效应的信息, 而且在没有交互效应时, 我们可以用更少的观测值得到关于单个因子效应的至少同样多的信息。为了说明这一点, 我们考虑前面氮和磷的  $2 \times 2$  析因试验。如果因子之间没有交互效应, 数据就会像图 15.6(a) 那样。为了方便起见, 我们把数据重新作成表 15.24, 并在其中把四个因子水平组合编号为 1 至 4。如果使用了  $2 \times 2$  析因试验, 并且在两因子之间没有交互效应存在, 我们可以得到两个独立的差用来考察两个因子在响应变量上的效应。从表 15.24 可以看出, 观测值 1 和 4 之间的差及观测值 2 和 3 之间的差可以用来度量磷的效应, 类似地, 观测值 4 和 3 之间的差及观测值 2 和 1 之间的差可以用来度量氮的两个水平在地块产量上的效应。

在这同一个试验中, 如果我们使用了一次一因子方法, 为了在无交互效应的情况下考察氮和磷各自的效应时得到同样个数的独立的差, 需要六个观测值(在表 15.24 中前三个因子水平组合处各两个观测值)。

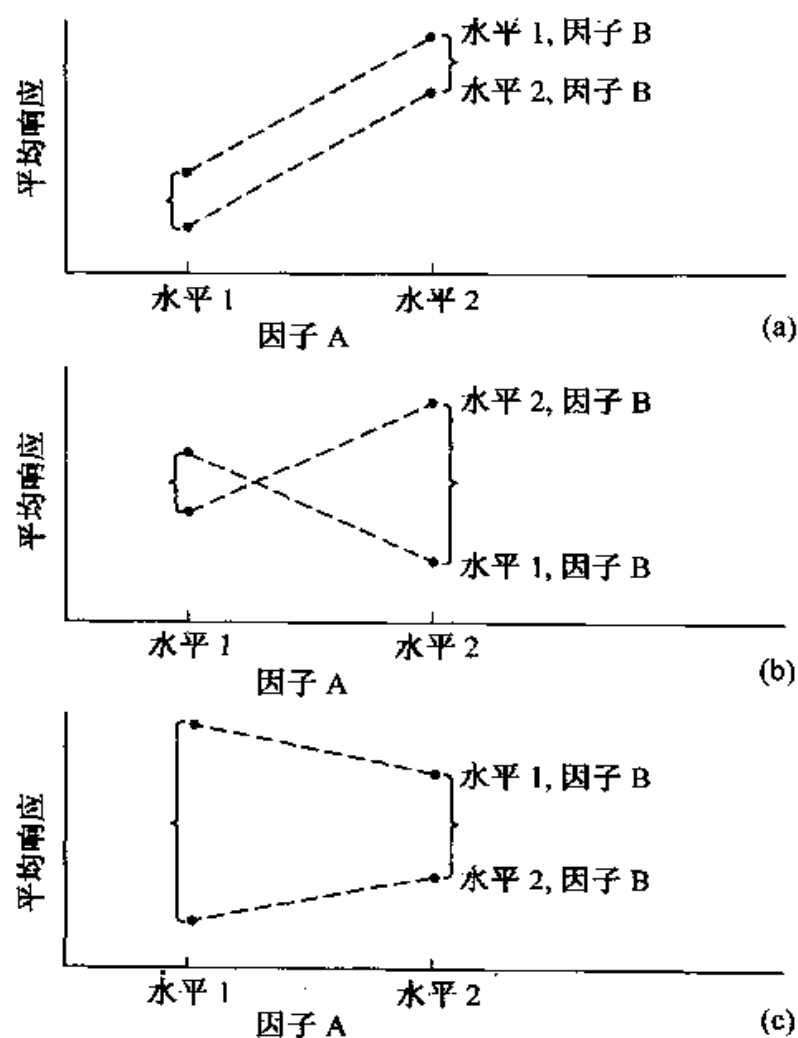


图 15.6  $2 \times 2$  析因试验中有、无交互效应的说明: (a) 因子 A 和 B 间无交互效应; (b) 因子 A 和 B 间有交互效应; (c) 因子 A 和 B 间有交互效应

表 15.24  $2 \times 2$  析因试验的因子水平组合

| 组合 | 产量  | 氮  | 磷  |
|----|-----|----|----|
| 1  | 145 | 60 | 10 |
| 2  | 125 | 40 | 10 |
| 3  | 165 | 40 | 20 |
| 4  | 180 | 60 | 20 |

在具有两因子因子处理结构, 重复次数  $n > 1$  的完全随机化试验中, 观测值的

模型可以写为如下形式

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

其中模型中的各项定义如下:

$y_{ijk}$ : 因子 A 第  $i$  个水平和因子 B 第  $j$  个水平上的第  $k$  个试验单元上的响应;

$\mu$ : 总平均值, 未知常数;

$\alpha_i$ : 因子 A 第  $i$  个水平的效应, 未知常数;

$\beta_j$ : 因子 B 第  $j$  个水平的效应, 未知常数;

$\alpha\beta_{ij}$ : 因子 A 第  $i$  个水平与因子 B 第  $j$  个水平的交互效应, 未知常数;

$\epsilon_{ijk}$ : 接受因子 A 第  $i$  个水平和因子 B 第  $j$  个水平的第  $k$  个试验单元上响应的随机误差。我们要求这些随机误差服从均值为 0, 共同方差为  $\sigma_e^2$  的正态分布, 并且这些误差相互独立。

由上述模型的条件可以证明, 从接受因子 A 第  $i$  个水平和因子 B 第  $j$  个水平的第  $k$  个试验单元上所观测到的响应服从均值为

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j$$

方差为  $\sigma_e^2$  的正态分布。

为了说明这个模型, 考虑一个无交互作用的两因子析因试验, 例如氮和磷的  $2 \times 2$  析因试验:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

$2 \times 2$  析因试验的期望值如表 15.25 所示。

表 15.25  $2 \times 2$  析因试验中的期望值

| 因子 A | 因子 B                       |                            |
|------|----------------------------|----------------------------|
|      | 水平 1                       | 水平 2                       |
| 水平 1 | $\mu + \alpha_1 + \beta_1$ | $\mu + \alpha_1 + \beta_2$ |
| 水平 2 | $\mu + \alpha_2 + \beta_1$ | $\mu + \alpha_2 + \beta_2$ |

这个模型假定对于因子 A 的任何两个水平, 总体均值(期望值)的差不随所考虑的因子 B 的水平而变化。同样的性质在比较因子 B 的两个水平时也成立。例如, 不论我们所考虑的因子 B 的水平是什么, 因子 A 的水平 1 和水平 2 上的平均响应的差都是  $\alpha_1 - \alpha_2$ 。因而, 对因子 A 的两个水平之间没有差异的检验可以写为  $H_0: \alpha_1 - \alpha_2 = 0$  的形式。同样, 对于因子 A 的任何一个水平, 因子 B 的两个水平上平均响应的差均为  $\beta_1 - \beta_2$ , 因子 B 的两个水平无差异的检验是  $H_0: \beta_1 - \beta_2 = 0$ 。这种现象在随机化区组设计中也看到过。



如果模型中项的可加性假设不成立,那么我们就需要一个带有解释交互效应的项的模型。

在每个处理上有  $n$  次观测的  $2 \times 2$  析因试验的期望值在表 15.26 中给出。

表 15.26  $2 \times 2$  析因试验中的期望值,有重复观测

| 因子 A | 因子 B  |   |
|------|---|---|
|      | 水平 1  | 水平 2  |
| 水平 1 | $\mu + \alpha_1 + \beta_1 + \alpha\beta_{11}$ | $\mu + \alpha_1 + \beta_2 + \alpha\beta_{12}$ |
| 水平 2 | $\mu + \alpha_2 + \beta_1 + \alpha\beta_{21}$ | $\mu + \alpha_2 + \beta_2 + \alpha\beta_{22}$ |

从表 15.26 中可以看到,在因子 B 的水平 1 上,因子 A 水平 1 和水平 2 的平均响应的差为

$$(\alpha_1 - \alpha_2) + (\alpha\beta_{11} - \alpha\beta_{21})$$

但对于因子 B 的水平 2,相应的差为

$$(\alpha_1 - \alpha_2) + (\alpha\beta_{12} - \alpha\beta_{22})$$

由于对于因子 B 的不同水平,因子 A 的水平 1 和水平 2 的平均响应的差也不同,模型不再是可加的,此时我们说这两个因子有交互效应。

**定义 15.4** 称两个因子 A 和 B 是有交互效应的,如果一个因子的两水平上的平均响应的差在第二个因子的不同水平上不保持常数。

在汽油辛烷值的测量中,当结合两种混合成分形成汽油混合物时,可能发生交互效应。混合物的辛烷性质可能与由考察其成分而形成的期待大相径庭。在这种情况下,交互作用可能对混合物的性态起正面或负面作用,我们分别说这些成分加强,或削弱了另一个成分。

我们可以用前面图 15.6 中的侧面图(profile plot)来详细解释交互效应的概念。正如我们在图 15.6(a)中所看到的那样,当没有交互效应的时候,对于因子 A 的各个水平,因子 B 的两个水平上平均响应值的差是一致的。然而,在图 15.6(b)和图 15.6(c)中,我们看到当因子 A 从水平 1 变到水平 2 时,因子 B 的两个水平上平均响应值的差也随之而变。

值得注意的是,交互效应不限于两个因子。对于三个因子 A, B, C,我们可能有因子 A 与 B 之间、A 与 C 之间、B 与 C 之间的交互效应,这些两个因子之间的交互效应可以立即从定义 15.4 中得到解释。如, AC 交互效应的存在表明, A 的水平上平均响应值的差随 C 的水平而变化。因子 A、B、C 之间的一个三因子交互效应可能意味着 C 的水平上平均响应值的差随因子 A 与 B 的水平组合而变化。

为了对带有因子间交互效应的析因试验进行方差分析,我们需要在每个处理(因子水平组合)处有  $n > 1$  个观测值。假定在一个完全随机化两因子试验中,因子 A 有  $a$  个水平,因子 B 有  $b$  个水平,在  $ab$  个处理中的每一个处理中,有  $n$  个观测值。我们来构造该试验的方差分析。在分解总平方和之前,我们先需要定义以下记号。

$y_{ijk}$ : 接受因子 A 的第  $i$  个水平和因子 B 的第  $j$  个水平的第  $k$  个试验单元上的观测值;

$\bar{y}_{i..}$ : 因子 A 的第  $i$  个水平上观测值的样本平均值,  $\bar{y}_{i..} = \frac{1}{bn} \sum_{jk} y_{ijk}$ ;

$\bar{y}_{.j.}$ : 因子 B 的第  $j$  个水平上观测值的样本平均值,  $\bar{y}_{.j.} = \frac{1}{an} \sum_{ik} y_{ijk}$ ;

$\bar{y}_{ij.}$ : 因子 A 的第  $i$  个水平和因子 B 的  $j$  个水平上观测值的样本平均值,  $\bar{y}_{ij.} = \frac{1}{n} \sum_k y_{ijk}$ ;

$\bar{y}_{...}$ : 总的样本平均值,  $\bar{y}_{...} = \frac{1}{abn} \sum_{ijk} y_{ijk}$ 。

各观测值关于它们的平均值的总平方和,与以前一样,定义为

$$\text{TSS} = \sum_{ijk} (y_{ijk} - \bar{y}_{...})^2$$

这个平方和将被分解成四个不同的变异来源:两个来自于因子 A 和 B 的主效应,一个是来自于因子 A 和 B 之间的交互效应,另一个则是源于不能由主效应和交互效应所解释的所有其他的变异,称之为误差。TSS 的分解可以通过对模型

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

的考察得到。模型中的参数有样本估计:

$$\hat{\mu} = \bar{y}_{...} \quad \hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...} \quad \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$$

$$\hat{\alpha}\hat{\beta}_{ij} = (\bar{y}_{ij.} - \bar{y}_{...}) - \hat{\alpha}_i - \hat{\beta}_j = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

通过一些代数运算,我们可以证明, TSS 具有下列形式:

$$\begin{aligned} \sum_{ijk} (y_{ijk} - \bar{y}_{...})^2 &= bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 + an \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &\quad + n \sum_{ij} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{ijk} (y_{ijk} - \bar{y}_{ij.})^2 \end{aligned}$$

我们用上面的参数估计来解释这个分解中的各项。等号右边第一项度量了因子 A 的主效应,该项可以写为

$$\text{SSA} = bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = bn \sum_i (\hat{\alpha}_i)^2$$

类似地,等号右边第二项度量了因子 B 的主效应,可以写为

$$\text{SSB} = an \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 = an \sum_j (\hat{\beta}_j)^2$$

第三项度量了因子 A 和因子 B 之间的交互效应, 可以写为

$$SSAB = n \sum_{ij} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 = n \sum_{ij} (\hat{\alpha}\hat{\beta}_{ij})^2$$

最后的那一项称为误差平方和, SSE, 表示各  $y_{ijk}$  中没有被主效应和交互效应所解释的变异性。这一项有几种表示形式。像以前一样, 利用模型定义残差  $e_{ijk} = y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\alpha}\hat{\beta}_{ij} = y_{ijk} - \bar{y}_{ij.}$ 。于是

$$SSE = \sum_{ijk} (y_{ijk} - \bar{y}_{ij.})^2 = \sum_{ijk} (e_{ijk})^2$$

另一种形式是  $SSE = TSS - SSA - SSB - SSAB$ 。平方和的分解总结成 AOV 表, 如表 15.27 所示。

表 15.27 完全随机化两因子析因试验的 AOV 表

| 来源   | SS   | df               | MS                             | F            |
|------|------|------------------|--------------------------------|--------------|
| 主效应  |      |                  |                                |              |
| A    | SSA  | $a - 1$          | $MSA = SSA / (a - 1)$          | $MSA / MSE$  |
| B    | SSB  | $b - 1$          | $MSB = SSB / (b - 1)$          | $MSB / MSE$  |
| 交互效应 |      |                  |                                |              |
| AB   | SSAB | $(a - 1)(b - 1)$ | $MSAB = SSAB / (a - 1)(b - 1)$ | $MSAB / MSE$ |
| 误差   |      |                  |                                |              |
| 误差   | SSE  | $ab(n - 1)$      | $MSE = SSE / ab(n - 1)$        |              |
| 总和   | TSS  | $abn - 1$        |                                |              |

从 AOV 表我们可以看出, 如果在每一个处理上仅有一个观测值, 即  $n = 1$ , 那么误差有 0 个自由度。因此, 如果因子 A 和 B 之间有交互效应, 并且  $n = 1$ , 则对于交互效应和主效应就没有有效的检验方法。然而, 如果因子间没有交互效应, 那么交互效应项可以当作误差项使用, 把 SSE 换为 SSAB。但  $n = 1$  的试验非常少见, 因为在大多数情形, 研究人员在进行试验之前并不知道因子 A 和 B 之间是否存在交互效应。因此, 为了保证对主效应和交互效应的检验是有效的, 我们需要  $n > 1$ 。

### 例 15.8

做一项试验, 以确定四种不同的杀虫剂对于三种柑橘( $B_1, B_2, B_3$ )产量的影响。对于每一种柑橘, 随机地从果园里挑选了 8 棵树, 把每一种杀虫剂随机地分配给其中的两棵, 并按照指定的水平使用。试验期结束时, 记录下柑橘的产量(蒲式耳/每棵树), 所得数据见表 15.28。

表 15.28 果树产量的  $3 \times 4$  析因试验数据, 每个处理有  $n = 2$  个观测值

| 品种, B | 杀虫剂, A |    |    |    |
|-------|--------|----|----|----|
|       | 1      | 2  | 3  | 4  |
| 1     | 49     | 50 | 43 | 53 |
|       | 39     | 55 | 38 | 48 |
| 2     | 55     | 67 | 53 | 85 |
|       | 41     | 58 | 42 | 73 |
| 3     | 66     | 85 | 69 | 85 |
|       | 68     | 92 | 62 | 99 |

- 对于该试验, 写出一个适当的模型。
- 建立方差分析表, 并用  $\alpha = 0.05$  进行主效应和交互效应的  $F$  检验。
- 画出处理均值的侧面图。

**解答** 题中所述试验为一个完全随机化  $3 \times 4$  析因试验, 其中因子 A, 即杀虫剂有  $a = 3$  个水平, 因子 B, 即品种有  $b = 4$  个水平。在两个因子的 12 个水平组合上的重复试验次数为  $n = 2$ 。

- 对于带有两因子间交互效应的  $3 \times 4$  析因试验, 模型为

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}, i = 1, 2, 3, 4; j = 1, 2, 3; k = 1, 2$$

其中  $\mu$  为每棵树产量的总平均,  $\alpha_i$  为杀虫剂第  $i$  个水平的效应,  $\beta_j$  为柑橘品种第  $j$  个水平的效应,  $\alpha\beta_{ij}$  为杀虫剂第  $i$  个水平和柑橘品种第  $j$  个水平的交互效应。

- 在多数试验中, 我们极力推荐使用计算机软件程序来得到 AOV 表。但为了说明 AOV 表的计算, 我们在此例中使用各个平方和的定义来计算。首先, 我们估计模型中的参数。为此, 我们使用表 15.29 中给出的处理均值。

表 15.29 A 和 B 的因子组合(处理)的样本均值

| 品种, B  | 杀虫剂, A |       |       |       | 品种平均值  |
|--------|--------|-------|-------|-------|--------|
|        | 1      | 2     | 3     | 4     |        |
| 1      | 44     | 52.5  | 40.5  | 50.5  | 46.875 |
| 2      | 48     | 62.5  | 47.5  | 79    | 59.25  |
| 3      | 67     | 88.5  | 65.5  | 92    | 78.25  |
| 杀虫剂平均值 | 53     | 67.83 | 51.17 | 78.83 | 61.46  |

算得的参数估计如下:

主效应

因子 A

$$\hat{\alpha}_1 = \bar{y}_{1..} - \bar{y}_{...} = 53 - 61.46 = -8.46$$

$$\hat{\alpha}_2 = \bar{y}_{2..} - \bar{y}_{...} = 67.83 - 61.46 = 6.37$$

$$\hat{\alpha}_3 = \bar{y}_{3..} - \bar{y}_{...} = 51.17 - 61.46 = -10.29$$

$$\hat{\alpha}_4 = \bar{y}_{4..} - \bar{y}_{...} = 73.83 - 61.46 = 12.37$$

因子 B

$$\hat{\beta}_1 = \bar{y}_{.1.} - \bar{y}_{...} = 46.875 - 61.46 = -14.585$$

$$\hat{\beta}_2 = \bar{y}_{.2.} - \bar{y}_{...} = 59.25 - 61.46 = -2.21$$

$$\hat{\beta}_3 = \bar{y}_{.3.} - \bar{y}_{...} = 78.25 - 61.46 = 16.79$$

交互效应

$$\hat{\alpha\beta}_{11} = \bar{y}_{11.} - \bar{y}_{1..} - \bar{y}_{.1.} + \bar{y}_{...} = 44 - 53 - 46.875 + 61.46 = 5.585$$

$$\hat{\alpha\beta}_{12} = \bar{y}_{12.} - \bar{y}_{1..} - \bar{y}_{.2.} + \bar{y}_{...} = 48 - 53 - 59.25 + 61.46 = -2.79$$

$$\hat{\alpha\beta}_{13} = \bar{y}_{13.} - \bar{y}_{1..} - \bar{y}_{.3.} + \bar{y}_{...} = 67 - 53 - 78.25 + 61.46 = -2.79$$

$$\hat{\alpha\beta}_{21} = \bar{y}_{21.} - \bar{y}_{2..} - \bar{y}_{.1.} + \bar{y}_{...} = 52.5 - 67.83 - 46.875 + 61.46 = -0.745$$

$$\hat{\alpha\beta}_{22} = \bar{y}_{22.} - \bar{y}_{2..} - \bar{y}_{.2.} + \bar{y}_{...} = 62.5 - 67.83 - 59.25 + 61.46 = -3.12$$

$$\hat{\alpha\beta}_{23} = \bar{y}_{23.} - \bar{y}_{2..} - \bar{y}_{.3.} + \bar{y}_{...} = 88.5 - 67.83 - 78.25 + 61.46 = 3.88$$

$$\hat{\alpha\beta}_{31} = \bar{y}_{31.} - \bar{y}_{3..} - \bar{y}_{.1.} + \bar{y}_{...} = 40.5 - 51.17 - 46.875 + 61.46 = 3.915$$

$$\hat{\alpha\beta}_{32} = \bar{y}_{32.} - \bar{y}_{3..} - \bar{y}_{.2.} + \bar{y}_{...} = 47.5 - 51.17 - 59.25 + 61.46 = -1.46$$

$$\hat{\alpha\beta}_{33} = \bar{y}_{33.} - \bar{y}_{3..} - \bar{y}_{.3.} + \bar{y}_{...} = 65.5 - 51.17 - 78.25 + 61.46 = -2.46$$

$$\hat{\alpha\beta}_{41} = \bar{y}_{41.} - \bar{y}_{4..} - \bar{y}_{.1.} + \bar{y}_{...} = 50.5 - 73.83 - 46.875 + 61.46 = -8.745$$

$$\hat{\alpha\beta}_{42} = \bar{y}_{42.} - \bar{y}_{4..} - \bar{y}_{.2.} + \bar{y}_{...} = 79 - 73.83 - 59.25 + 61.46 = 7.38$$

$$\hat{\alpha\beta}_{43} = \bar{y}_{43.} - \bar{y}_{4..} - \bar{y}_{.3.} + \bar{y}_{...} = 92 - 73.83 - 78.25 + 61.46 = 1.38$$

接下来, 我们计算总平方和。

$$\begin{aligned} \text{TSS} &= \sum_{ijk} (y_{ijk} - \bar{y}_{...})^2 = (46 - 61.46)^2 + (50 - 61.46)^2 + \cdots + (99 - 61.46)^2 \\ &= 7,187.96 \end{aligned}$$

主效应平方和为

$$\text{SSA} = bn \sum_i (\hat{\alpha}_i)^2 = (3)(2)[(-8.46)^2 + (6.37)^2 + (-10.29)^2 + (12.37)^2]$$

$$= 2,226.29$$

$$SSB = an \sum_i (\hat{\beta}_i)^2 = (4)(2)[(-14.585)^2 + (-2.21)^2 + (16.79)^2]$$

$$= 3,996.08$$

交互效应平方和为

$$SSAB = n \sum_{ij} (\bar{\alpha}\beta_{ij})^2 = (2)[(5.585)^2 + (-2.79)^2 + \cdots + (1.38)^2] = 456.92$$

误差平方和为

$$SSE = TSS - SSA - SSB - SSAB$$

$$= 7,187.96 - 2,226.29 - 3,996.08 - 456.92 = 508.67$$

这个在每个处理上重复试验次数为 2 的完全随机化  $3 \times 4$  析因试验的方差分析表由表 15.30 给出。

表 15.30 例 15.8 中水果产量试验的 AOV 表

| 来源       | SS       | df | MS       | F     |
|----------|----------|----|----------|-------|
| 杀虫剂, A   | 2,226.29 | 3  | 742.10   | 17.51 |
| 品种, B    | 3,996.08 | 2  | 1,998.04 | 47.13 |
| 交互效应, AB | 456.92   | 6  | 76.15    | 1.80  |
| 误差       | 508.67   | 12 | 42.39    |       |
| 总和       | 7,187.96 | 23 |          |       |

第一个要做的检验必须是因子 A 与 B 交互效应的检验, 因为如果交互效应显著, 就可能难以得到主效应的合理解释。F 统计量为

$$F = \frac{MSAB}{MSE} = \frac{76.15}{42.39} = 1.80$$

计算所得的这个 F 值没有超过 F 表中与  $\alpha = 0.05$ ,  $df_1 = 6$ ,  $df_2 = 12$  对应的表值 3.00。因此, 没有充分的依据表明杀虫剂水平与树的品种水平之间存在交互效应。通过侧面图我们可以看出交互效应不明显。图 15.7 中画出了该试验中各个处理的样本平均值。

从该图我们可以看出, 对于 4 种杀虫剂, 3 种柑橘树的平均产量之间的差大体上相同, 也就是说, 三个品种所对应的三条线大体上是平行的, 因而, 品种的水平与杀虫剂的水平之间的交互效应是不显著的。由于交互效应不显著, 下一步我们便可以检验两个因子的主效应。这些检验分别考察品种之间的差异以及杀虫剂之间的差异。对于杀虫剂, F 统计量为

$$F = \frac{MSA}{MSE} = \frac{742.10}{42.39} = 17.51$$

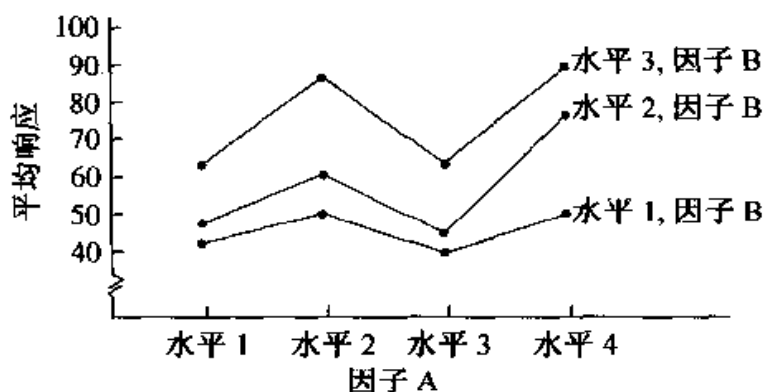


图 15.7 例 15.8 中水果产量试验的侧面图

计算所得的这个  $F$  值大于  $F$  表中与  $\alpha = 0.05$ ,  $df_1 = 3$ ,  $df_2 = 12$  对应的表值 3.49。因此, 我们有充分的依据表明在 4 种杀虫剂水平上的平均产量之间有显著的差别。对于品种,  $F$  统计量为

$$F = \frac{MSB}{MSE} = \frac{1,998.04}{42.39} = 47.13$$

这个  $F$  值大于  $F$  表中与  $\alpha = 0.05$ ,  $df_1 = 2$ ,  $df_2 = 12$  对应的表值 3.89。因此, 我们有充分的依据表明在 3 种柑橘树的平均产量之间有显著的差别。在 15.7 节, 我们将讨论如何把因子 A 和因子 B 的两个有差异的水平找出来。

在交互效应显著の場合, 解释主效应  $F$  检验的结果时必须慎重行事。首先, 我们应当用样本均值  $\bar{y}_{ij}$  构造侧面图。如果例 15.8 的侧面图像图 15.8 那样, 则说明有在因子 A 与因子 B (杀虫剂与品种) 之间存在交互效应的表征。毋庸置疑, 如果  $MSE$  相对于  $MSAB$  不是很大, 那么交互效应的  $F$  检验将是显著的。

对于图 15.8 中的侧面图, 主效应的  $F$  检验适用吗? 回答是“适用”, 因为其中

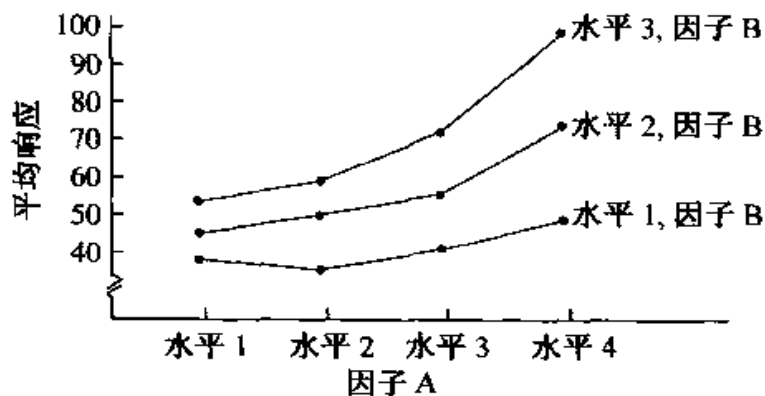


图 15.8 有显著交互效应存在时的侧面图, 其中交互效应是有序的

的交互效应是有序的,即虽然随着因子 A 的水平变化,因子 B 的水平之间差异的大小也在变化,但因子 B 的各个水平上的均值的顺序是不变的。很显然,在图 15.8 中,因子 B 的水平 3 上的均值总是比水平 1 和水平 2 上的均值大;同样,不论在因子 A 的哪个水平上,因子 B 的水平 2 上的均值总是比水平 1 上的均值大。然而,在此情况下,我们在给出结论时必须十分小心。如果我们发现,对于因子 A 的所有水平,因子 B 的水平 3 上的平均响应比水平 1 和水平 2 上的平均响应大,并且因子 B 的各水平间有显著差异,我们可能断言因子 B 的水平 3 上的平均值明显地比其他两个水平上的平均值高。但是,注意在因子 A 的水平 1 上,因子 B 的三个水平上的平均响应相差无几。因此,如果我们使用 A 的水平 1,那么因子 B 的三个水平将会产生等效的平均响应。于是,我们得到的关于因子 B 的各个水平的平均响应之间的差异的结论,在因子 A 的不同水平上是不一致的;并且可能在因子 A 的某些水平上,因子 B 的主效应的检验会相互矛盾。

当交互效应有序时,主效应的检验可能是有意义的。然而,并非所有的交互效应都是有序的。图 15.9 中的侧面图显示了一个有显著的交互效应,并且主效应的检验可能导致误解的情况。无序的交互效应,就像图 15.9 中那样,可能把主效应遮挡起来。不是说主效应的检验在统计上是不对的,而是这些检验可能导致对试验结果的不恰当的解释。在因子 A 的水平 1 上,因子 B 的三个水平上的平均响应之间相差甚少;在因子 A 的水平 3 上,因子 B 的水平 3 产生的响应比水平 2 产生的响应大得多;相反,在因子 A 的水平 4 上,因子 B 的水平 2 产生的平均响应比水平 3 产生的平均响应大得多。因而,当两个因子有显著的交互效应时,因子 B 的不同水平上的平均响应之间的比较必须在因子 A 的各个水平上分别进行。这就是说,此时关于因子 B 的水平的一个结论并不对因子 A 的所有水平都成立。

当我们的试验中有三个因子时,计算要复杂得多,但关于主效应和交互效应的解释与只有两个因子时相仿。对于 A、B、C 三个因子,我们可能有 A 与 B 之间的

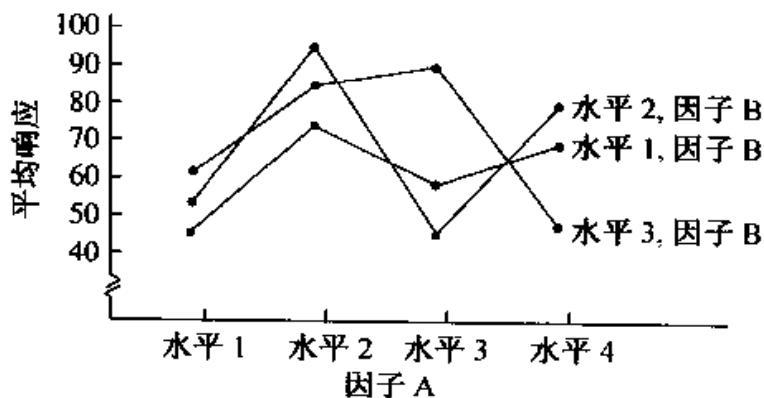


图 15.9 有显著交互效应存在时的侧面图,  
其中交互效应是无序的



交互效应, A 与 C 之间的交互效应以及 B 与 C 之间的交互效应。对于这些两因子交互效应的解释, 可以从定义 15.4 立即得出。如 AC 交互效应的存在表明因子 A 的各水平上平均响应之间的差异随因子 C 的水平而变。与前面所讨论的一样, 在解释主效应时也必须谨慎。三个因子 A、B 和 C 之间的交互效应可能表明因子 C 的各个水平上平均响应之间的差异随因子 A 与因子 B 的水平组合的不同而不同。三因子交互效应的第二种解释是因子 A 与 B 的交互效应的模式随因子 C 的水平而变化。这样, 当存在三因子交互效应, 并且我们已经在因子 C 的每个水平上, 分别画出因子 A 与因子 B 的两因子交互效应的侧面图时, 一定能够在这几张侧面图上看到不同的模式。

对于具有三因子因子处理结构和  $n > 1$  次重复的完全随机化试验, 模型可以写为

$$y_{ijkm} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijkm}$$

其中模型的各项定义如下:

$y_{ijkm}$ : 接受因子 A 第  $i$  个水平, 因子 B 第  $j$  个水平和因子 C 第  $k$  个水平的第  $m$  个试验单元上的响应;

$\mu$ : 总平均, 未知常数;

$\alpha_i$ : 因子 A 第  $i$  个水平的效应, 未知常数;

$\beta_j$ : 因子 B 第  $j$  个水平的效应, 未知常数;

$\gamma_k$ : 因子 C 第  $k$  个水平的效应, 未知常数;

$\alpha\beta_{ij}$ : 因子 A 第  $i$  个水平与因子 B 第  $j$  个水平的两因子交互效应, 未知常数;

$\alpha\gamma_{ik}$ : 因子 A 第  $i$  个水平与因子 C 第  $k$  个水平的两因子交互效应, 未知常数;

$\beta\gamma_{jk}$ : 因子 B 第  $j$  个水平与因子 C 第  $k$  个水平的两因子交互效应, 未知常数;

$\alpha\beta\gamma_{ijk}$ : 因子 A 第  $i$  个水平, 因子 B 第  $j$  个水平和因子 C 第  $k$  个水平的三因子交互效应, 未知常数;

$\varepsilon_{ijkm}$ : 接受因子 A 第  $i$  个水平, 因子 B 第  $j$  个水平和因子 C 第  $k$  个水平的第  $m$  个试验单元上响应的随机误差。我们要求这些随机误差服从均值为 0, 共同方差为  $\sigma_\varepsilon^2$  的正态分布, 并且这些误差相互独立。

由上述模型的条件可以证明, 从接受因子 A 第  $i$  个水平, 因子 B 第  $j$  个水平和因子 C 第  $k$  个水平的第  $m$  个试验单元上所观测到的响应服从均值为

$$E(y_{ijkm}) = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk},$$

方差为  $\sigma_\varepsilon^2$  的正态分布。

下列记号用于把总平方和分解为主效应分量, 交互效应分量和误差分量的过程之中。

$y_{ijkm}$ : 接受因子 A 第  $i$  个水平, 因子 B 第  $j$  个水平和因子 C 第  $k$  个水平的第

$m$  个试验单元上的观测值;

$\bar{y}_{i...}$ : 在因子 A 的第  $i$  个水平上观测值的样本平均值,

$$\bar{y}_{i...} = \frac{1}{bcn} \sum_{jkm} y_{ijkm}$$

$\bar{y}_{.j..}$ : 在因子 B 的第  $j$  个水平上观测值的样本平均值,

$$\bar{y}_{.j..} = \frac{1}{acn} \sum_{ikm} y_{ijkm}$$

$\bar{y}_{...k}$ : 在因子 C 的第  $k$  个水平上观测值的样本平均值,

$$\bar{y}_{...k} = \frac{1}{abn} \sum_{ijm} y_{ijkm}$$

$\bar{y}_{ij..}$ : 在因子 A 的第  $i$  个水平与因子 B 的第  $j$  个水平上观测值的样本平均值,

$$\bar{y}_{ij..} = \frac{1}{cn} \sum_{km} y_{ijkm}$$

$\bar{y}_{i..k}$ : 在因子 A 的第  $i$  个水平与因子 C 的第  $k$  个水平上观测值的样本平均值,

$$\bar{y}_{i..k} = \frac{1}{bn} \sum_{jm} y_{ijkm}$$

$\bar{y}_{.jk.}$ : 在因子 B 的第  $j$  个水平与因子 C 的第  $k$  个水平上观测值的样本平均值,

$$\bar{y}_{.jk.} = \frac{1}{an} \sum_{im} y_{ijkm}$$

$\bar{y}_{ijh.}$ : 在因子 A 第  $i$  个水平, 因子 B 第  $j$  个水平和因子 C 第  $k$  个水平上观测值的样本平均值,

$$\bar{y}_{ijh.} = \frac{1}{n} \sum_m y_{ijkm}$$

$\bar{y}_{....}$ : 总样本平均值,

$$\bar{y}_{....} = \frac{1}{abcn} \sum_{ijkm} y_{ijkm}$$

模型中的参数有样本估计:

$$\hat{\mu} = \bar{y}_{....} \quad \hat{\alpha}_i = \bar{y}_{i...} - \bar{y}_{....} \quad \hat{\beta}_j = \bar{y}_{.j..} - \bar{y}_{....} \quad \hat{\gamma}_k = \bar{y}_{...k} - \bar{y}_{....}$$

$$\hat{\alpha\beta}_{ij} = (\bar{y}_{ij..} - \bar{y}_{....}) - \hat{\alpha}_i - \hat{\beta}_j = \bar{y}_{ij..} - \bar{y}_{i...} - \bar{y}_{.j..} + \bar{y}_{....}$$

$$\hat{\alpha\beta}_{ik} = (\bar{y}_{i..k} - \bar{y}_{....}) - \hat{\alpha}_i - \hat{\gamma}_k = \bar{y}_{i..k} - \bar{y}_{i...} - \bar{y}_{...k} + \bar{y}_{....}$$

$$\hat{\beta\gamma}_{jk} = (\bar{y}_{.jk.} - \bar{y}_{....}) - \hat{\beta}_j - \hat{\gamma}_k = \bar{y}_{.jk.} - \bar{y}_{.j..} - \bar{y}_{...k} + \bar{y}_{....}$$

$$\hat{\alpha\beta\gamma}_{ijk} = (\bar{y}_{ijh.} - \bar{y}_{....}) - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k - \hat{\alpha\beta}_{ij} - \hat{\alpha\gamma}_{ik} - \hat{\beta\gamma}_{jk}$$

$$= \bar{y}_{ijk.} - \bar{y}_{ij..} - \bar{y}_{ik..} - \bar{y}_{.jk.} + \bar{y}_{i...} + \bar{y}_{.j..} + \bar{y}_{...k.} - \bar{y}_{....}$$

由此,模型拟合的残差为

$$\begin{aligned} e_{ijk.} &= y_{ijk.} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k - \hat{\alpha}\hat{\beta}_{ij} - \hat{\alpha}\hat{\gamma}_{ik} - \hat{\beta}\hat{\gamma}_{jk} - \hat{\alpha}\hat{\beta}\hat{\gamma}_{ijk} \\ &= y_{ijk.} - \bar{y}_{ijk.} \end{aligned}$$

由以上各式,我们可以把因子 A 有  $a$  个水平,因子 B 有  $b$  个水平,因子 C 有  $c$  个水平,并且每个因子水平组合(处理)处有  $n$  个观测值的三因子析因试验的总平方和分解为主效应(单个因子水平间的变异),两因子交互作用,三因子交互作用和误差平方和。

主效应的平方和为

$$\begin{aligned} \text{SSA} &= bcn \sum_i (\hat{\alpha}_i)^2 = bcn \sum_i (\bar{y}_{i...} - \bar{y}_{....})^2 \\ \text{SSB} &= acn \sum_j (\hat{\beta}_j)^2 = acn \sum_j (\bar{y}_{.j..} - \bar{y}_{....})^2 \\ \text{SSC} &= abn \sum_k (\hat{\gamma}_k)^2 = abn \sum_k (\bar{y}_{...k.} - \bar{y}_{....})^2 \end{aligned}$$

两因子交互效应平方和为

$$\begin{aligned} \text{SSAB} &= cn \sum_{ij} (\hat{\alpha}\hat{\beta}_{ij})^2 = cn \sum_{ij} (\bar{y}_{ij..} - \bar{y}_{....})^2 - \text{SSA} - \text{SSB} \\ \text{SSAC} &= bn \sum_{ik} (\hat{\alpha}\hat{\gamma}_{ik})^2 = bn \sum_{ik} (\bar{y}_{i.k.} - \bar{y}_{....})^2 - \text{SSA} - \text{SSC} \\ \text{SSBC} &= an \sum_{jk} (\hat{\beta}\hat{\gamma}_{jk})^2 = an \sum_{jk} (\bar{y}_{.jk.} - \bar{y}_{....})^2 - \text{SSB} - \text{SSC} \end{aligned}$$

三因子交互效应平方和为

$$\begin{aligned} \text{SSABC} &= n \sum_{ijk} (\hat{\alpha}\hat{\beta}\hat{\gamma}_{ijk})^2 \\ &= n \sum_{ijk} (\bar{y}_{ijk.} - \bar{y}_{....})^2 - \text{SSAB} - \text{SSAC} - \text{SSBC} - \text{SSA} - \text{SSB} - \text{SSC} \end{aligned}$$

误差平方和由

$$\begin{aligned} \text{SSE} &= \sum_{ijk.} (e_{ijk.})^2 \\ &= \sum_{ijk.} (y_{ijk.} - \bar{y}_{ijk.})^2 \\ &= \text{TSS} - \text{SSA} - \text{SSB} - \text{SSC} - \text{SSAB} - \text{SSAC} - \text{SSBC} - \text{SSABC} \end{aligned}$$

给出,其中  $\text{TSS} = \sum_{ijk.} (y_{ijk.} - \bar{y}_{....})^2$ 。

对于因子 A 有  $a$  个水平,因子 B 有  $b$  个水平,因子 C 有  $c$  个水平,并且  $abc$  个处理中每个处理上有  $n$  个观测值的完全随机化三因子析因试验, AOV 表由表 15.31 给出。

表 15.31 完全随机化三因子析因试验的 AOV 表

| 来源   | SS    | df                      | MS                                      | F             |
|------|-------|-------------------------|---|---------------|
| 主效应  |       |                         |   |               |
| A    | SSA   | $a - 1$                 | $MSA = SSA / (a - 1)$                   | $MSA / MSE$   |
| B    | SSB   | $b - 1$                 | $MSB = SSB / (b - 1)$                   | $MSB / MSE$   |
| C    | SSC   | $c - 1$                 | $MSC = SSC / (c - 1)$                   | $MSC / MSE$   |
| 交互效应 |       |                         |   |               |
| AB   | SSAB  | $(a - 1)(b - 1)$        | $MSAB = SSAB / (a - 1)(b - 1)$          | $MSAB / MSE$  |
| AC   | SSAC  | $(a - 1)(c - 1)$        | $MSAC = SSAC / (a - 1)(c - 1)$          | $MSAC / MSE$  |
| BC   | SSBC  | $(b - 1)(c - 1)$        | $MSBC = SSBC / (b - 1)(c - 1)$          | $MSBC / MSE$  |
| ABC  | SSABC | $(a - 1)(b - 1)(c - 1)$ | $MSABC = SSABC / (a - 1)(b - 1)(c - 1)$ | $MSABC / MSE$ |
| 误差   | SSE   | $abc(n - 1)$            | $MSE = SSE / abc(n - 1)$                |               |
| 总和   | TSS   | $abcn - 1$              |   |               |

从该 AOV 表我们可以看出,如果在每个处理上只有一次观测,即  $n = 1$ ,那么误差平方和的自由度为 0。因此,如果模型中含交互效应项,并且  $n = 1$ ,那么对于交互效应和主效应就没有有效的检验。然而,如果已知某些交互效应为 0,那么就可以把这些交互效应项结合起来作为误差项,用于检验模型中的其他项。但  $n = 1$  的试验很少用,因为在多数情况下,研究人员在实施试验之前不知道哪些交互效应会是 0。因而,为了能对主效应和交互效应进行有效的检验,我们需要  $n > 1$ 。

当三因子交互效应显著时,我们必须对两因子交互效应和主效应做出与三因子交互效应不显著时不同的处理,这使得三因子试验的分析有一些复杂。下面的图 15.10 给出了三因子试验分析的一般方法。该图来自于 G. Milliken 和 D. Johnson 的 *Analysis of Messy Data* 一书。

我们将通过分析案例中的数据来说明如何分析三因子试验。

#### 案例, 统计分析

由于在三因子试验中平方和所需要的计算很多,手算可能会导致大的舍入误差,我们用软件程序计算,所得结果见表 15.32。

这个 AOV 表中的记号意义如下:L 表示魔芋水平,B 表示魔芋混合物的类型,P 表示 SPI 的水平。由于三因子交互效应不显著( $L * B * P, p = 0.3106$ ),我们接下来考察两因子交互效应。三组两因子交互效应都具有显著的水平, ( $L * B, p = 0.0008$ ), ( $L * P, p < 0.0001$ ), ( $B * P, p < 0.0001$ ),从而,三个都是高度显著的。为了考察三个因子间可能存在的关系,我们需要计算样本平均值  $\bar{y}_{ij..}$ ,  $\bar{y}_{i.k.}$  及  $\bar{y}_{.jk.}$ 。这些值见表 15.33(表中称这些平均值为“两种方式平均值”)。

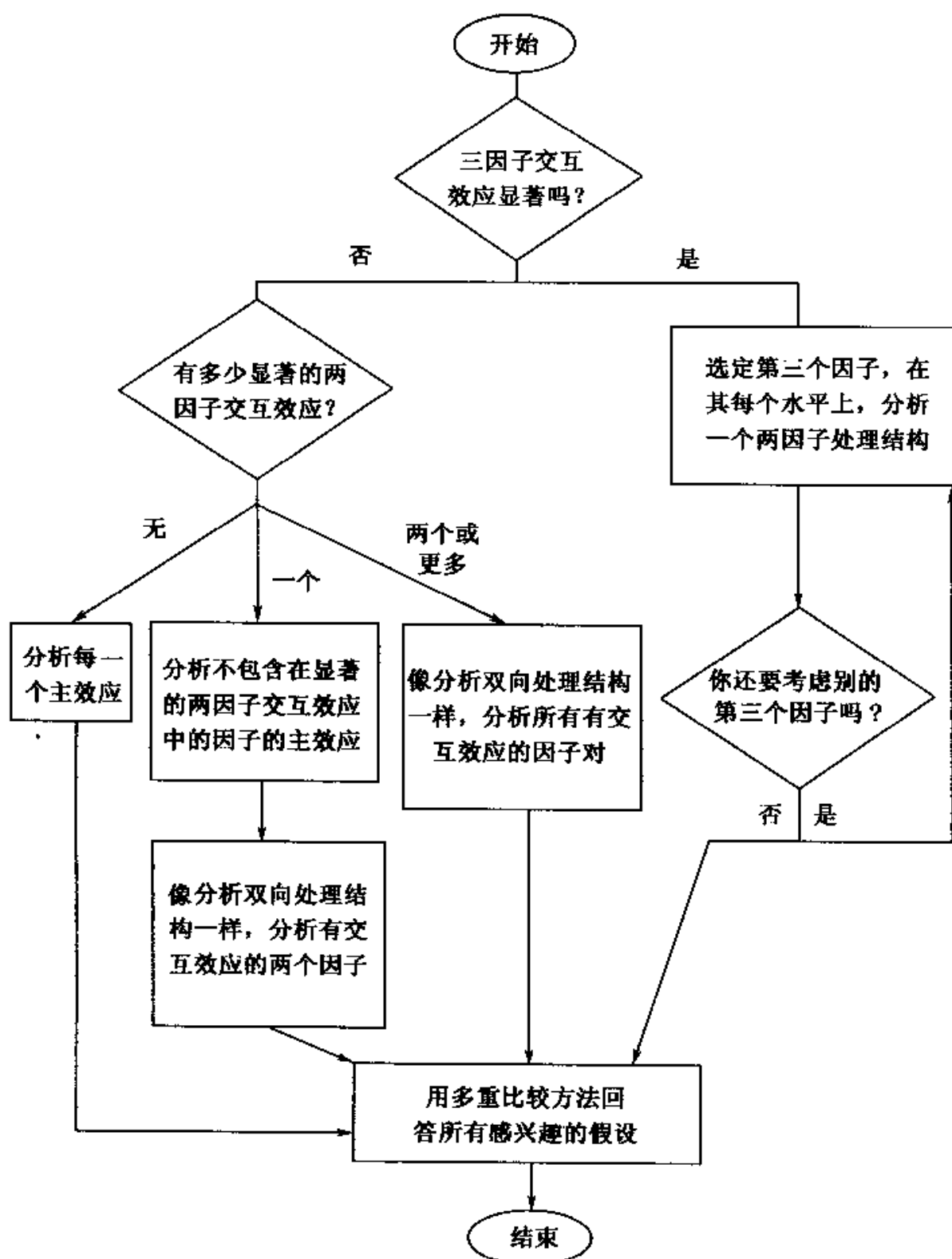


图 15.10 三因子试验的分析方法

表 15.32 三因子析因试验案例中数据的 AOV 表

General Linear Models Procedure

Dependent Variable: Texture of Meat:

| Source          | DF       | Sum of Squares | Mean Square | F Value | Pr > F  |
|-----------------|----------|----------------|-------------|---------|---------|
| Model           | 11       | 2080.28750     | 189.11705   | 62.40   | 0.0001  |
| Error           | 24       | 72.74000       | 3.03083     |         |         |
| Corrected Total | 35       | 2153.02750     |             |         |         |
|                 | R-Square | C. V.          | Root MSE    |         | Y Mean  |
|                 | 0.966215 | 1.769387       | 1.74093     |         | 98.3917 |
| Source          | DF       | Type I SS      | Mean Square | F Value | Pr > F  |
| Main Effects:   |          |                |             |         |         |
| L               | 1        | 526.70250      | 526.70250   | 173.78  | 0.0001  |
| B               | 1        | 113.42250      | 113.42250   | 37.42   | 0.0001  |
| P               | 2        | 1090.11500     | 545.05750   | 179.84  | 0.0001  |
| Interactions:   |          |                |             |         |         |
| L * B           | 1        | 44.22250       | 44.22250    | 14.59   | 0.0008  |
| L * P           | 2        | 182.53500      | 91.26750    | 30.11   | 0.0001  |
| B * P           | 2        | 115.84500      | 57.92250    | 19.11   | 0.0001  |
| L * B * P       | 2        | 7.44500        | 3.72250     | 1.23    | 0.3106  |

表 15.33 案例中数据的平均值表

| 水平(%) | 混合物 | SPI(%) | 两种方式平均值 |
|-------|-----|--------|---------|
| 0.5   | KSS | *      | 99.3    |
| 0.5   | KNC | *      | 105.1   |
| 1     | KSS | *      | 93.9    |
| 1     | KNC | *      | 95.2    |
| 0.5   | *   | 1.1    | 110.0   |
| 0.5   | *   | 2.2    | 105.0   |
| 0.5   | *   | 4.4    | 91.7    |
| 1     | *   | 1.1    | 97.5    |
| 1     | *   | 2.2    | 96.2    |
| 1     | *   | 4.4    | 90.0    |
| *     | KSS | 1.1    | 104.5   |
| *     | KSS | 2.2    | 97.4    |
| *     | KSS | 4.4    | 88.0    |
| *     | KNC | 1.1    | 103.0   |
| *     | KNC | 2.2    | 103.9   |
| *     | KNC | 4.4    | 93.7    |

利用表中的平均值可以画出魔芋的水平与魔芋的类型, 魔芋的水平与 SPI 的水平以及魔芋的类型与 SPI 的水平之间的交互效应的侧面图, 如图 15.11 所示。

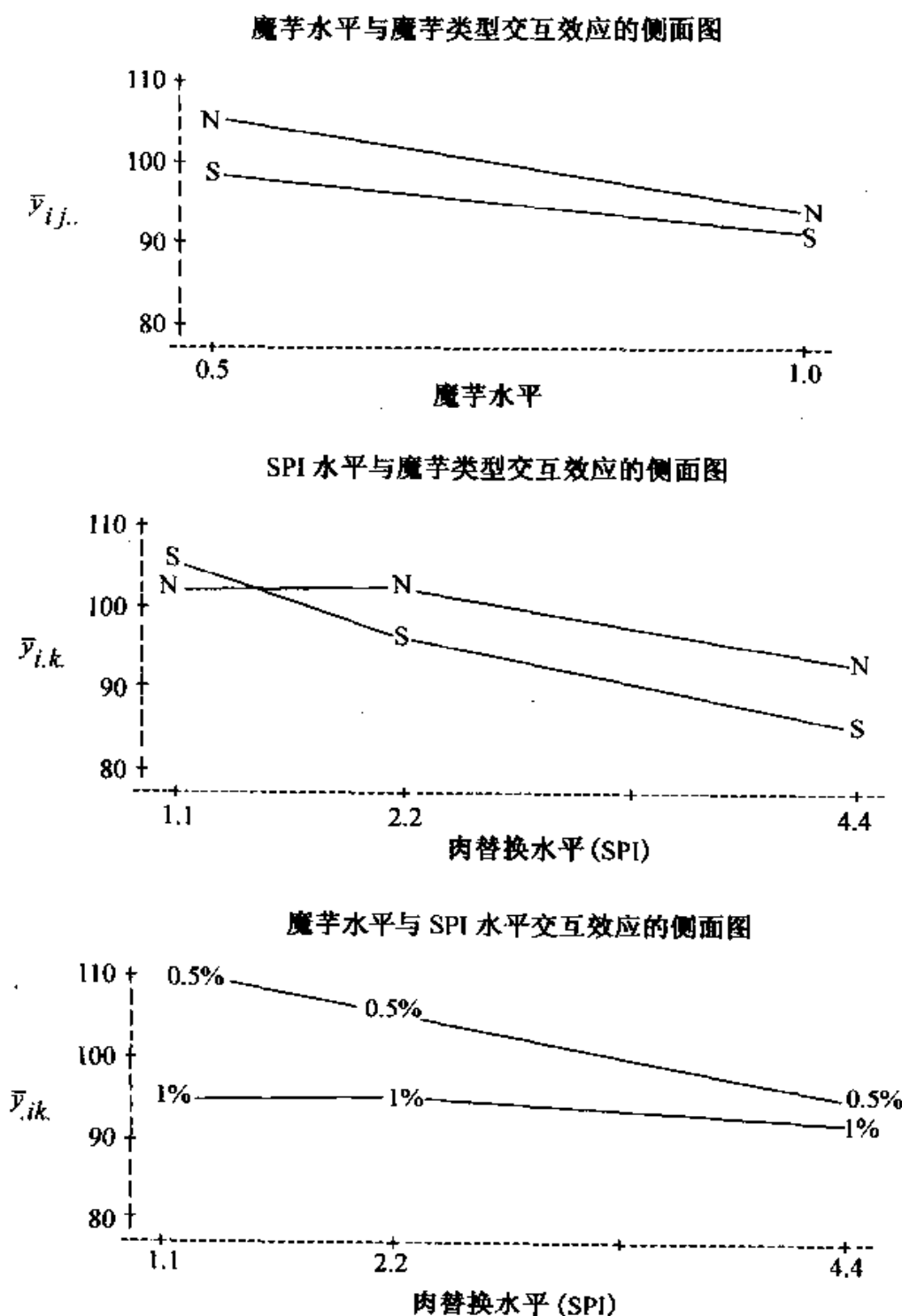


图 15.11 两因子交互效应的侧面图

从图 15.11 我们可以看出,肉制品的组织指标平均值随着肉制品中魔芋类型,魔芋水平以及 SPI 的水平不同而有相当的差异。当魔芋的水平是 1% 时,在肉制品的平均组织指标之间相差无几。然而,在 5% 的水平上,KNC 魔芋所得到的产品的平均组织指标比魔芋混合物 KSS 的相应值高一些。当考虑 SPI 的水平在大红肠平均组织指标上的效应时,我们可以看到,在 SPI 的 1% 的水平上,用 0.5% 魔芋和 1% 魔芋产生不小的差异。随着 SPI 水平的增加,差异明显地减小。还有,在 SPI 的 1% 的水平上,两种魔芋混合物之间没有本质的差异,但随着 SPI 水平的增加,用 KNC 混合物比用 KSS 混合物生产出的肉制品有更高的组织指标。所有关于三个因子和肉制品平均组织指标的关系的这些观察结果都需要用多重比较的方法进行确认。我们将在 15.7 节中讨论这种方法。

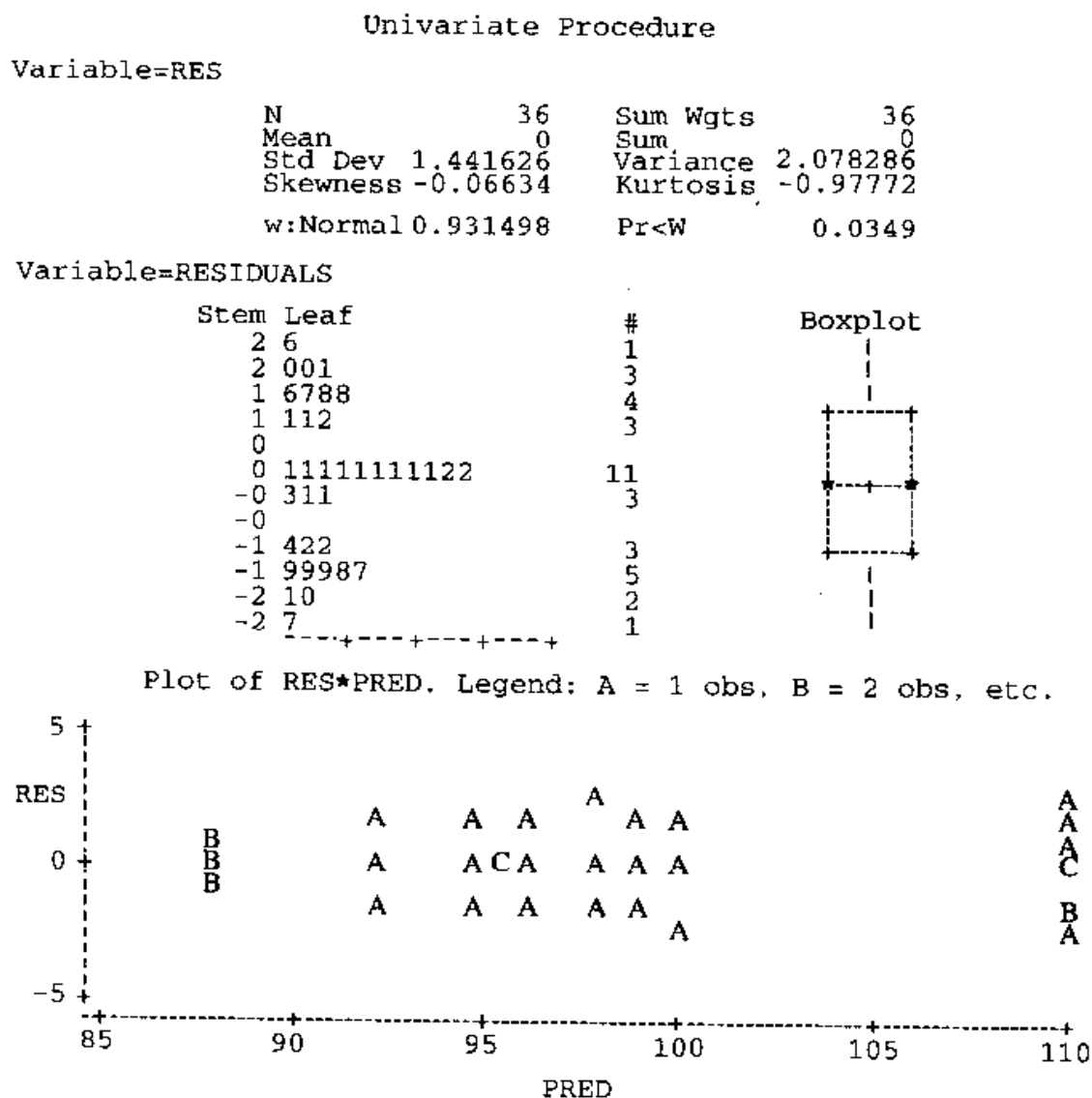


图 15.12 案例的残差分析



图 15.12 包含了这些组织指标数据的残差分析,其中的残差是用公式

$$e_{ijkm} = y_{ijkm} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k - \hat{\alpha\beta}_{ij} - \hat{\alpha\gamma}_{ik} - \hat{\beta\gamma}_{jk} - \hat{\alpha\beta\gamma}_{ijk} = y_{ijkm} - \bar{y}_{ijk}.$$

得到的。

残差  $e_{ijkm}$  的概括统计量表明,偏度近似为 0,但峰度近似为 -1。从茎叶图和盒型图可以看出,这些残差大体对称,但在 0 附近有一个很高的峰。Shapiro-Wilks 正态性检验的  $p$  值为 0.0349,这表明残差有一定的非正态特性。然而,由于没有离群值,并且也没有几个残差的值落在两个极端的附近,正态性的假设大体上满足。残差对估计的处理均值  $\bar{y}_{ijk}$  的图表明,随着平均组织指标的增加,变异性有轻微的增加。然而,这种增加不足以破坏  $F$  检验所具有的对于少许模型条件偏离的稳健性。由此,可以认为正态性和等方差条件都能得到满足,AOV 表中的  $F$  检验将是有效的。

练习

应用

**15.12 (心理学)** 一家大的广告公司专门制作儿童用品的商业广告。该公司想设计一项研究以调查影响儿童对广告注意时间的因子。初步的调查表明,两个可能的重要因子分别是儿童的年龄和广告产品的种类。公司要确定广告引起儿童注意的平均时间长度是否因这两个因子水平的不同而有大的差异。如果确有差异,公司想根据产品类型和目标年龄组来确定商业广告的新类型。使用的三个年龄组为:

$A_1$ : 5~6 岁       $A_2$ : 7~8 岁       $A_3$ : 9~10 岁

挑选的产品类型为:

$P_1$ : 早餐用谷类食品       $P_2$ : 视频游戏

在每个年龄组募集了 30 名儿童。对于三种产品的每一种,在每一年龄组随机指定 10 名儿童看 60 秒该种产品的广告。研究人员记录下他们在观看广告的过程中的注意时间,得到如下数据。

| 儿童 | $A_1 - P_1$ | $A_2 - P_1$ | $A_3 - P_1$ | $A_1 - P_2$ | $A_2 - P_2$ | $A_3 - P_2$ |
|----|-------------|-------------|-------------|-------------|-------------|-------------|
| 1  | 19          | 19          | 37          | 39          | 30          | 51          |
| 2  | 36          | 35          | 6           | 18          | 47          | 52          |
| 3  | 40          | 22          | 28          | 32          | 6           | 43          |
| 4  | 30          | 28          | 4           | 22          | 27          | 48          |
| 5  | 4           | 1           | 32          | 16          | 44          | 39          |

续表

| 儿童 | $A_1 - P_1$ | $A_2 - P_1$ | $A_3 - P_1$ | $A_1 - P_2$ | $A_2 - P_2$ | $A_3 - P_2$ |
|----|-------------|-------------|-------------|-------------|-------------|-------------|
| 6  | 10          | 27          | 16          | 2           | 26          | 33          |
| 7  | 30          | 27          | 8           | 36          | 33          | 56          |
| 8  | 5           | 16          | 41          | 43          | 48          | 43          |
| 9  | 34          | 3           | 29          | 7           | 23          | 40          |
| 10 | 21          | 18          | 18          | 16          | 21          | 51          |

平均值    22.8    19.5    22.0    23.2    30.6    45.6

年龄组均值:  $A_1$      $A_2$      $A_3$     产品类型均值:  $P_1$      $P_2$   
                     23.0    25.1    33.8                      21.4    33.1

- 给出这个设计。
- 为该设计写出一个模型,解释模型中的所有项。
- 估计模型中的参数。
- 计算这些数据的平方和,并把所得结果总结在 AOV 表中。

15.13 参见练习 15.12。

- 对于两个因子,即年龄和产品类型,画出侧面图。
- 进行适当的  $F$  检验,并从中得出关于年龄和产品类型对于儿童平均注意时间的效应的结论。

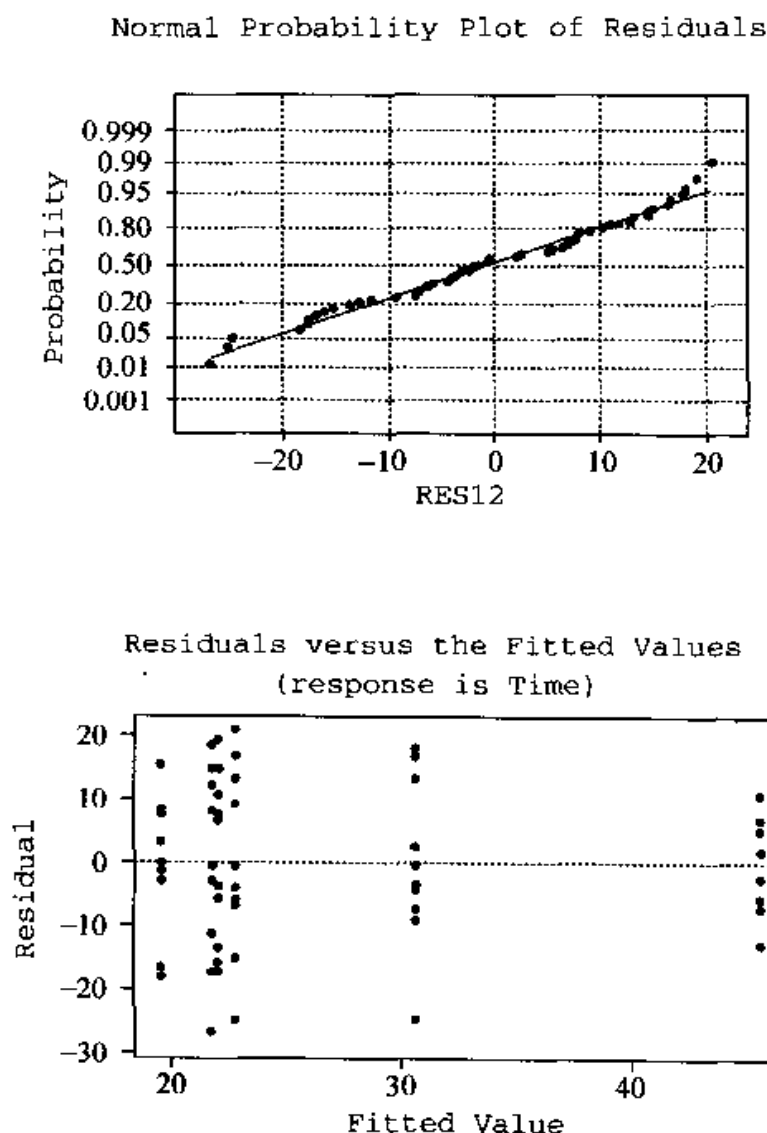
15.14 对于练习 15.12 中的数据,用 Minitab 软件包得到的输出结果如下。

- 比较从 Minitab 得到的 AOV 表与你在练习 15.13 中得到的结果。
- 用残差图检查在为使  $F$  检验有效所需要的条件中,是否有的得不到满足。

#### Two-Way Analysis of Variance

##### Analysis of Variance for Time

| Source      | DF | SS    | MS   | F     | P     |
|-------------|----|-------|------|-------|-------|
| Age         | 2  | 1320  | 660  | 4.49  | 0.016 |
| Product     | 1  | 2052  | 2052 | 13.97 | 0.000 |
| Interaction | 2  | 1341  | 671  | 4.57  | 0.015 |
| Error       | 54 | 7932  | 147  |       |       |
| Total       | 59 | 12645 |      |       |       |



## 15.6 随机化完全区组设计中的因子处理结构

在前面的一节中,我们把析因试验定义为在所有自变量的因子水平组合上都对响应变量  $y$  进行观测的试验。在析因试验中自变量的因子水平组合被随机地指定给试验单元,因而我们可以用完全随机化设计来调查这些因子对响应变量的效应。

有时,研究的目标使得我们希望调查某些因子对于响应变量的效应,而对于某些其他的外部变异来源要用划分区组的方法予以滤除。这种情况需要使用带有来自于一个析因试验的处理的区组设计。为了有效地滤除外部变异来源,从而着重于要调查的这些因子对于响应变量的效应,我们将利用已经学习的区组设计(随机

化区组设计和拉丁方设计)的知识。我们通过下面的例子予以说明。

### 例 15.9

一位营养学家要研究使用三种新型面粉做成并在三种不同温度下烤制出来的面包中蛋白质含量的百分比。她决定对于九种面粉-温度组合,在每一组合上烤制三个面包,这样共有 27 个面包。对这些面包测量其蛋白质的百分比。然而,她在任何给定的一天内只能烤九个面包。提出一个适当的试验设计。

**解答** 由于在一天内可以烤九个面包,为了得到所希望的那么多个观测值,可以在这三天内重复进行  $3 \times 3$  析因试验。设计如下。

| 面粉类型 | 第一天<br>温度 |   |   | 第二天<br>温度 |   |   | 第三天<br>温度 |   |   |
|------|-----------|---|---|-----------|---|---|-----------|---|---|
|      | 1         | 2 | 3 | 1         | 2 | 3 | 1         | 2 | 3 |
| A    | y         | y | y | y         | y | y | y         | y | y |
| B    | y         | y | y | y         | y | y | y         | y | y |
| C    | y         | y | y | y         | y | y | y         | y | y |

注意,这个设计实际上是一个随机化区组设计,其中区组为烤制时间,处理是  $3 \times 3$  析因试验的九个因子水平组合。这样,利用这个随机化区组设计,我们能够在比较这些处理时滤除由于讨厌变量烤制时间而产生的变异性。由于这些处理是来自于一个析因试验的因子水平组合,我们可以在滤除由于烤制时间不同而产生的变异的基础上,考察两个因子(因子 A 和因子 B)对于响应变量的效应。

### 例 15.10

对于例 15.9 中的按  $b = 3$  的随机化区组设计进行的  $3 \times 3$  析因试验,构造一个方差分析表,写出各个变异来源和自由度。

**解答**  $t = 9, b = 3$  的随机化区组设计的方差分析表如下:

| 来源 | SS  | df |
|----|-----|----|
| 处理 | SST | 8  |
| 区组 | SSB | 2  |
| 误差 | SSE | 16 |
| 总和 | TSS | 26 |

由于这个随机化区组设计的处理是一个  $3 \times 3$  析因试验的九个因子水平组合, 我们可以按 15.5 节中的方法把处理平方和进一步划分为  $3 \times 3$  析因试验中的各个变异来源。修改后的 AOV 表为:

| 来源 | SS   | df |
|----|------|----|
| 处理 | SST  | 8  |
| A  | SSA  | 2  |
| B  | SSB  | 2  |
| AB | SSAB | 4  |
| 区组 | SSB  | 2  |
| 误差 | SSE  | 16 |
| 总和 | TSS  | 26 |

在此我们不用  $F = MST/MSE$  做比较处理均值的综合检验, 而是对因子试验进行方差分析以考察交互作用和主效应。这些  $F$  检验将使用适当的均方(MSAB, MSA 和 MSB)作分子, 用从表中计算出来的 MSE 作分母。(译者注: 注意表中第四行和第六行的两个 SSB 意义不同, 它们分别表示因子 B 的效应的平方和及区组效应的平方和)

### 练习

#### 基本技能

**15.15** 用图表给出一个有  $b = 3$  个区组, 每个区组上有一个  $3 \times 5$  析因试验的设计, 并写出完整的方差分析表(包括来源, 平方和, 自由度)。

**15.16** 用图表给出一个有  $b = 2$  个区组, 每个区组上有一个  $2 \times 4 \times 3$  析因试验的设计, 并写出该设计的完整的方差分析表。

## 15.7 处理差异的估计和处理均值的比较

我们已经强调了随机化区组设计、拉丁方设计和析因试验的方差分析的重要性。然而, 有时我们可能对估计两个处理(同一因子的不同水平或者不同的因子水平组合)上的平均响应的差更感兴趣。例如, 一位环境工程师可能对估计修复工作前后湖中溶解氧的平均含量的差更感兴趣, 而不是要检验是否存在差异。因而, 这位工程师会提出这样的问题: “溶解氧的平均含量的差是什么”, 而不是问“清理工

作前后溶解氧的平均含量有差异吗?”

Fisher 的 LSD 方法可以用来评估随机化区组设计、拉丁方设计以及各种设计下的  $k$  因子析因试验中处理均值的差异。以  $\bar{y}_i$  记处理  $i$  的平均响应,  $\bar{y}_{i'}$  记处理  $i'$  的平均响应,  $n_i$  记每个处理下观测的次数。这两个处理的平均响应的差异  $\mu_i - \mu_{i'}$  的一个  $100(1 - \alpha)\%$  的置信区间由下式给出。

**处理均值差异的  $100(1 - \alpha)\%$  的置信区间**

$$(\bar{y}_i - \bar{y}_{i'}) \pm t_{\alpha/2} s_e \sqrt{\frac{2}{n_i}}$$

其中  $s_e$  为 AOV 表中 MSE 的平方根,  $t_{\alpha/2}$  可以用  $\alpha = \alpha/2$  和 MSE 的自由度从附录中的表 2 查得。

### 例 15.11

一家公司要比较空中交通管理人员使用的三种显示屏幕。对于每种显示屏幕,在五种不同的模拟紧急情况下进行考察。三十名受过良好训练并具有相似工作经历的空中交通管理人员参加了这项研究。把这些空中交通管理人员随机地分配给显示屏幕与紧急情况的组合,使得每一种屏幕 - 紧急情况的组合有两名空中交通管理人员。对于每一名管理员,记录他在每一种紧急情况下稳定下来所需要的时间(秒)。这些数据在表 15.34 中给出。

**表 15.34 显示屏幕数据(时间以秒计)**

| 显示屏幕, B | 紧急情况, A |    |    |    |    |
|---------|---------|----|----|----|----|
|         | 1       | 2  | 3  | 4  | 5  |
| 1       | 18      | 31 | 22 | 39 | 15 |
|         | 16      | 35 | 27 | 36 | 12 |
| 2       | 13      | 33 | 24 | 35 | 10 |
|         | 15      | 30 | 21 | 38 | 16 |
| 3       | 24      | 42 | 40 | 52 | 28 |
|         | 28      | 46 | 37 | 57 | 24 |

- 构造一个侧面图。
- 作出包括检验交互效应在内的方差分析。

### 解答

a. 样本均值在表 15.35 中给出, 由这些样本均值画出的侧面图见图 15.13。由该侧面图我们可以看出, 对于任意一对显示屏幕, 这些管理员的平均反应时间的差在所有五种紧急情况下基本保持不变。显示屏幕 1 和显示屏幕 2 在这五种紧急情况下都产生了基本相同的平均反应时间, 而显示屏幕 3 产生的平均反应时间一致地比其他两种显示屏幕要长。我们下面将要通过考虑反应时间相对于观测平均时间的变异性, 进行假设检验来确认这些判断。

表 15.35 显示屏幕-紧急情况下的平均反应时间

| 显示屏幕, B             | 紧急情况, A |      |      |      |      | 平均值 $\bar{y}_{.j.}$    |
|---------------------|---------|------|------|------|------|------------------------|
|                     | 1       | 2    | 3    | 4    | 5    |                        |
| 1                   | 17      | 33   | 24.5 | 37.5 | 13.5 | 25.1                   |
| 2                   | 14      | 31.5 | 22.5 | 36.5 | 13   | 23.5                   |
| 3                   | 26      | 44   | 38.5 | 54.5 | 26   | 37.8                   |
| 平均值 $\bar{y}_{i..}$ | 19.0    | 36.2 | 28.5 | 42.8 | 17.5 | $\bar{y}_{...} = 28.8$ |

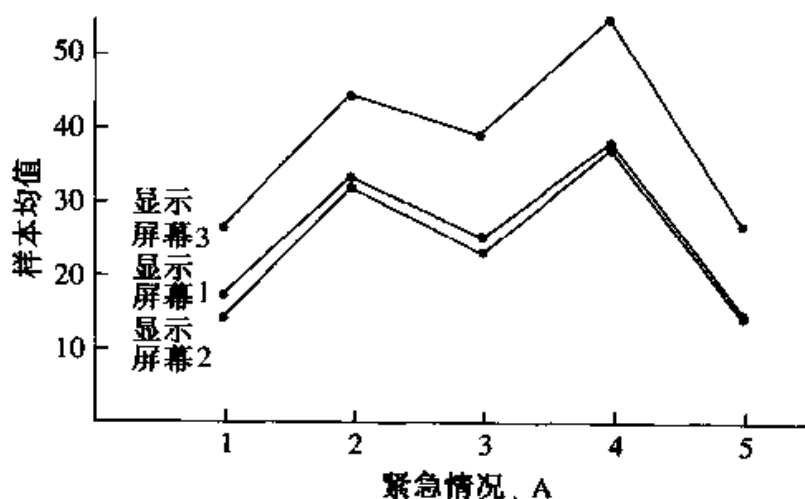


图 15.13 每种紧急情况下屏幕均值的图

表 15.36 显示屏幕-紧急情况研究的 AOV 表

| General Linear Models Procedure           |          |                |             |         |        |
|---|----------|----------------|-------------|---------|--------|
| Dependent Variable: y, Stabilization Time |          |                |             |         |        |
| Source                                    | DF       | Sum of Squares | Mean Square | F Value | Pr > F |
| Model                                     | 14       | 4122.8000      | 294.4857    | 41.67   | 0.0001 |
| Error                                     | 15       | 106.0000       | 7.0667      |         |        |
| Corrected Total                           | 29       | 4228.8000      |             |         |        |
|   | R-Square | C. V.          | Root MSE    |         | Y Mean |
|   | 0.974934 | 9.230279       | 2.6583      |         | 28.800 |
| Source                                    | DF       | Type I SS      | Mean Square | F Value | Pr > F |
| D   | 2        | 1227.8000      | 613.9000    | 86.87   | 0.0001 |
| E   | 4        | 2850.1333      | 712.5333    | 100.83  | 0.0001 |
| D * E                                     | 8        | 44.8667        | 5.6083      | 0.79    | 0.6167 |

b. 计算机输出的方差分析表见表 15.36.

第一个假设检验是关于两个因子,即紧急情况和显示屏幕类型之间的交互效应的检验。由于计算所得  $F$  值 0.79 小于  $F$  的临界值 2.64(相应于  $\alpha = 0.05$ ,  $df_1 = 8$  和  $df_2 = 15$ ),没有充分根据( $p$  值 = 0.6167)表明在紧急情况和显示屏幕类型之间有交互效应。这就确认了我们通过侧面图所作出的判断。由于交互效应不显著,我们接下来检验显示屏幕类型的主效应。计算所得  $F$  值为 86.87,大于  $F$  的临界值 3.68(相应于  $\alpha = 0.05$ ,  $df_1 = 2$  和  $df_2 = 15$ ),所以我们有充分的依据( $p$  值 < 0.0001)表明,对于三种显示屏幕,平均反应时间之间有显著的差异。

### 例 15.12

参见例 15.11。研究人员对于三种类型的显示屏幕所产生的平均反应时间差异的大小很感兴趣。用 95% 的置信区间估计这些差异。

**解答** 由于在显示屏幕类型和紧急情况类型之间没有显著的交互效应,不同类型的显示屏幕所产生的平均反应时间差异的大小,在所有五种紧急情况下应大体一致。于是,我们可以在这五种紧急情况上求平均,来考察显示屏幕类型的主效应:  $\hat{\mu}_j = \bar{y}_{.j.}$ ,  $j = 1, 2, 3$ 。从表 15.35,我们得到

$$\bar{y}_{.1.} = 25.1 \quad \bar{y}_{.2.} = 23.5 \quad \bar{y}_{.3.} = 37.8$$

$\alpha = 0.025$ ,  $df = 15$  所对应的  $t$  值是 2.131,  $\sigma_e$  的估计是



$$s_e = \sqrt{\text{MSE}} = \sqrt{7.0667} = 2.66$$

两种显示屏所产生的平均反应时间的差, 即  $\mu_{.j.} - \mu_{.j' .}$  的 95% 的置信区间的公式为

$$\bar{y}_{.j.} - \bar{y}_{.j' .} \pm t_{\alpha/2} s_e \sqrt{\frac{2}{n_t}}$$

对于显示屏 2 和 3, 每种显示屏我们有  $n_t = 10$  个观测值, 于是得到

$$37.8 - 23.5 \pm (2.131)(2.66) \sqrt{\frac{2}{10}}$$

$$14.3 \pm 2.54$$

即, 从 11.76 到 16.84。因此, 我们有 95% 的把握断言, 显示屏 2 和显示屏 3 的平均反应时间的差在 11.76 到 16.84 之间。同样地, 我们可以计算显示屏 1 和 3, 以及显示屏 1 和 2 的平均反应时间的差的置信区间。

在用  $F$  检验确定了有显著的主效应以后, 我们可以继续做出两种进一步的推断。首先, 我们可以利用例 15.12 中的方法求出任意一对因子水平上的均值差的置信区间: 对于因子 A, 这些差具有形式  $\mu_{i.} - \mu_{i' .}$ , 而对于因子 B, 它们有形式  $\mu_{.j.} - \mu_{.j' .}$ 。由此也就估计出了这两个因子效应的大小。其次, 我们可以决定这些因子的哪些水平不同于其余的水平。

如在第九章中所讨论的那样, 为了在对因子水平做两两比较时控制试验误差的大小, 我们使用**多重比较方法**。对于因子 A, 共有  $a(a-1)/2$  对, 而对于因子 B, 共有  $b(b-1)/2$  对。与第九章中一样, 选取使用哪一种多重比较的方法也依赖于所做的试验。所有在第九章讨论的方法, 如 LSD 法, Tukey 方法, SNK 法或 Scheffé 方法都可以应用在随机化完全区组设计, 拉丁方设计或  $k$  因子的因子试验的场合。第九章有关这些方法的公式中的量  $s_w^2$  用 MSE 代替, MSE 的自由度可以从 AOV 表中得到, 样本量  $n$  在这里指所比较的每个平均值中观测值的个数, 以  $\bar{y}_{i.}$  为例, 即指计算  $\bar{y}_{i.}$  时所用的数据值的个数。

### 例 15.13

参见例 15.11 和表 15.34 以及表 15.35 中的数据。用 Tukey 的  $W$  方法找出显示屏中的显著差异。

**解答** 对于 Tukey 的  $W$  方法, 我们使用第九章给出的公式

$$W = q_\alpha(t, v) \sqrt{\frac{s_w^2}{n}}$$

其中  $s_w^2$  为 AOV 表中的 MSE, 自由度  $v = 15$ ,  $q_\alpha(t, v)$  是用于比较  $t$  个不同总体均

值的学生化极差( $\alpha = \alpha$ )的上侧临界值。由附录中的表 10, 比较三种显示屏幕, 每个样本均值有十个观测值时,  $q_\alpha(t, v)$  的值为

$$q_{.05}(3, 15) = 3.67$$

由于每个均值有十个观测值,  $W$  的值为

$$W = q_\alpha(t, v) \sqrt{\frac{s_w^2}{n}} = 3.67 \sqrt{\frac{7.07}{10}} = 3.09$$

由表 15.35, 各个显示屏幕上的均值为

$$\bar{y}_{.1.} = 25.1 \quad \bar{y}_{.2.} = 23.5 \quad \bar{y}_{.3.} = 37.8$$

首先, 我们把样本均值从小到大进行排列:

| 显示屏幕 | 2    | 1    | 3    |
|------|------|------|------|
| 均值   | 23.5 | 25.1 | 37.8 |

对于差异(绝对值)超过  $W = 3.09$  的两个均值, 我们认为它们两个彼此差异显著。这个多重比较的结果总结如下:

显示屏幕     2     1     3

这就是说, 显示屏幕 1 和 2 比显示屏幕 3 有明显短的平均反应时间, 但在显示屏幕 1 和 2 之间, 我们不能探索出任何区别。

### 案例, 处理差异的比较

由于三因子交互效应  $L * S * P$  是不显著的( $p$  值 = 0.3106), 我们来考察研究人员感兴趣的两因子交互效应。他们希望了解在肉制品中, SPI 的百分比的增加对平均组织指标的影响。这样, 我们需要把平均组织指标的差作为 SPI 的百分比的函数来考察。由于在 SPI 和魔芋的水平之间有显著的交互效应( $p$  值 < 0.0001), 在 SPI 和魔芋的类型之间也存在显著的交互效应( $p$  值 < 0.0001), 我们需要进行四个 SPI 百分比的水平上的均值的分离。研究人员担心不能把各对均值的差异找出来, 所以我们使用 Tukey 的方法。

首先, 我们分别对于魔芋的两个水平 0.5% 和 1.0%, 比较 SPI 的各个百分比所对应的平均组织指标。Tukey 的  $W$  统计量的值为

$$W = q_\alpha(t, df_{\text{error}}) \sqrt{\frac{s_e^2}{n_i}}$$

其中  $t = 3$ , 即 SPI 的百分比的水平数; 从 AOV 表知  $df_{\text{error}} = 24$ ,  $s_e^2 = 3.0308$ ; 由于  $\bar{y}_{i.k}$  基于六个值算出,  $n_i = 6$  为在每个魔芋的水平上, SPI 的各个百分比所对应的平均组织指标中的观测值个数。由此, 从附录中的表 10, 我们得到  $q_\alpha(t, df_{\text{error}}) =$

$q_{0.05}(3, 24) = 3.53$ , 从而

$$W = 3.53 \sqrt{\frac{3.0308}{6}} = 2.51$$

于是, 在  $\alpha = 0.05$  的显著性水平上, 任何其差别大于 2.51 的一对均值  $\bar{y}_{i.k} - \bar{y}_{i.l.k}$  都认为是具有显著差异的。结果总结在表 15.37 中。

表 15.37 在每个魔芋的水平上, SPI 的百分比所对应的平均组织指标

| 魔芋水平 | SPI(%) |       |      |
|------|--------|-------|------|
|      | 1.1    | 2.2   | 4.4  |
| 0.5% | 110.0  | 105.0 | 91.7 |
|      | a      | b     | c    |
| 1.0% | 97.5   | 96.2  | 90.0 |
|      | a      | a     | b    |

在魔芋的水平 0.5% 上, 所有 SPI 的三个百分比产生的平均组织指标之间都是具有显著差别的。SPI 百分比的水平越高, 产生的平均组织指标值越小。对于魔芋的水平 1.0%, SPI 百分比 1.1 和 2.2 所对应的平均组织指标之间无显著差异, 而 SPI 百分比 4.4 所对应的平均组织指标比起 1.1 和 2.2 所对应的平均组织指标来, 有显著的下降。因此, SPI 的百分比与平均组织指标之间的关系, 在不同的魔芋的水平上是不同的。类似地, 在两个魔芋的类型上, 对于 SPI 的百分比与平均组织指标之间的关系, 我们得到如下结果(表 15.38)。W 表达式中的所有量与前面一样, 因为在每一个魔芋类型 - SPI 的百分比所对应的均值  $\bar{y}_{ik}$  上, 观测值的个数为  $n_i = 6$ 。因此,  $W = 2.51$ 。

表 15.38 在每个魔芋的类型上, SPI 的百分比所对应的平均组织指标

| 魔芋类型 | SPI(%) |       |      |
|------|--------|-------|------|
|      | 1.1    | 2.2   | 4.4  |
| KSS  | 104.5  | 97.4  | 88.0 |
|      | a      | b     | c    |
| KNC  | 103.0  | 103.9 | 93.7 |
|      | a      | a     | b    |

对于 KSS, 所有 SPI 的三个百分比产生的平均组织指标之间都是具有显著差别的。而对于 KNC, SPI 百分比 1.1 和 2.2 所对应的平均组织指标之间无显著差异, 而 SPI 百分比 4.4 所对应的平均组织指标比起 1.1 和 2.2 所对应的平均组织指标来, 有显著的下降。因此, SPI 的百分比与平均组织指标之间的关系, 对于不同的魔芋的类型是不同的。

## 15.8 小 结

在这一章中, 我们讨论了几种不同的试验设计和处理结构的方差分析。所考虑过的设计有: 完全随机化设计, 随机化完全区组设计和拉丁方设计。通过这些设计, 我们说明了如何极小化由外部变量带来的、我们所不感兴趣的变异性的效应, 从而对均值做出更精确的比较。因子处理结构可用于调查一个或多个因子对于试验响应的效应。它既可用于完全随机化设计, 随机化完全区组设计, 也可用于拉丁方设计。比如, 一个试验人员可能希望在用区组划分方法排除一个或多个外部变异来源之后, 考察一个或多个因子对于响应的效应。

对于本章中讨论的每个设计, 我们给出了设计的方案(包括处理的安排), 指出了该设计潜在的长处和不足, 建立了该设计的一个模型, 并讲解了相应的方差分析方法。最后, 对于每一个设计, 我们讨论了如何在处理均值间进行多重比较。

值得指出的是, 本章所给出的设计仅仅是最基本的设计, 而且, 我们仅仅考虑了平衡设计的情况, 也就是说, 对于每一个处理(因子水平组合), 所指定给的试验单元的数目是一样的。比如, 在两因子析因试验中, 如果在  $t = ab$  个处理中的每一个上, 都观测了完全相同个数的试验单元, 那么这个设计就是一个平衡设计。平衡设计的公式和分析比非平衡设计要简单一些。在第十九章, 我们将把这一章的结果推广到一些非平衡设计上。

### 重要公式

#### 1. 完全随机化设计中的单因子模型

模型:  $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$

平方和:

总和  $TSS = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$

处理  $SST = n \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = n \sum_i (\hat{\alpha}_i)^2$

误差  $SSE = \sum_{ij} (e_{ij})^2 = \sum_{ij} (y_{ij} - \bar{\mu} - \hat{\alpha}_i)^2 = TSS - SST$

#### 2. 随机化完全区组设计中的单因子模型

模型:  $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$

平方和:

$$\text{总和 TSS} = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$$

$$\text{处理 SST} = b \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = b \sum_i (\hat{\alpha}_i)^2$$

$$\text{区组 SSB} = t \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 = t \sum_j (\hat{\beta}_j)^2$$

$$\text{误差 SSE} = \sum_{ij} (e_{ij})^2 = \sum_{ij} (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2 = \text{TSS} - \text{SST} - \text{SSB}$$

### 3. 拉丁方设计中的单因子模型

$$\text{模型: } y_{ij} = \mu + \alpha_k + \beta_i + \gamma_j + \epsilon_{ij}$$

平方和:

$$\text{总和 TSS} = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$$

$$\text{处理 SST} = t \sum_k (\bar{y}_{.k} - \bar{y}_{..})^2 = t \sum_k (\hat{\alpha}_k)^2$$

$$\text{行 SSR} = t \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 = t \sum_i (\hat{\beta}_i)^2$$

$$\text{列 SSC} = t \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 = t \sum_j (\hat{\gamma}_j)^2$$

$$\text{误差 SSE} = \text{TSS} - \text{SST} - \text{SSR} - \text{SSC}$$

### 4. 完全随机化设计中两因子析因试验模型

$$\text{模型: } y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

平方和:

$$\text{总和 TSS} = \sum_{ijk} (y_{ijk} - \bar{y}_{...})^2$$

$$\text{因子 A SSA} = bn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = bn \sum_i (\hat{\alpha}_i)^2$$

$$\text{因子 B SSB} = an \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 = an \sum_j (\hat{\beta}_j)^2$$

$$\text{交互效应 SSAB} = n \sum_{ij} (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 = n \sum_{ij} (\hat{\alpha\beta}_{ij})^2$$

$$\text{误差 SSE} = \sum_{ijk} (y_{ijk} - \bar{y}_{ij.})^2 = \sum_{ijk} (e_{ijk})^2$$

### 5. 处理均值差的 $100(1 - \alpha)\%$ 置信区间

$$(\bar{y}_i - \bar{y}_{i'}) \pm t_{\alpha/2, df_{\text{error}}} s_e \sqrt{2/n_t}$$

### 补充练习

**15.17(工程)** 一位试验人员要考察一种新型黏结产品的黏结力。考察时的条件设定为三个不同的温度(280°F, 300°F 和 320°F)和四个不同的压力(100, 150, 200 和 250 磅/平方英寸)。这位试验人员准备了足够多的这种新型黏结产品, 从而可以在每个温度-压力条件组合上测试三个这种产品的样品。假定试验人员每天只能测试 12 个样品, 并且实验室中的条件日有所变。试给出一个试验设计, 其中考虑到实验室中不同日期的条件变化, 并给出一个图表, 指明在每一个温度-压力组合上的黏结剂样品。

**15.18(教育)** 要研究儿童虐待对儿童在校表现的影响。儿童虐待分为三种程度, 定义如下:

“受到虐待的儿童”——受到体罚的儿童;

“被忽略的儿童”——没有受到充分关怀的儿童；

“无虐待”——受到正常关怀并且没有受到体罚的儿童。

试验人员从三种程度的儿童中各随机选取了 30 名男孩和 30 名女孩，其中受到虐待的孩子和被忽略的孩子是从州立儿童福利机构的记录中抽选的，未受虐待的儿童是从当地的一所学校中抽选的。对于所有选中的孩子，记录了他们参加阅读、数学和一般科学三门功课的标准分级测验的成绩。

a. 假若所有学生都参加第七级测验，试给出一个设计。

b. 假若把这些孩子等分成三组，分别参加第三、五、七级测验，试给出一个设计。

**15.19(政府)** 一个中西部大城市的管理人员正在与分别代表警察、消防和建筑监察行业的三个协会就这些行业中雇员的工资进行谈判。这三家协会称，该市这三个行业的起点工资有巨大的差别，而在大多数城市中，这三种行业的起点工资之间没有显著差别。为了得到全国范围内的起点工资的信息，城市管理人员决定在八个地理分区中的每一个地区随机抽取一个城市，并由此得到该地区这三种行业的起点年薪(单位：千美元)。数据如下。

| 地区   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 均值    |
|------|------|------|------|------|------|------|------|------|-------|
| 警察   | 32.3 | 33.2 | 30.8 | 30.5 | 30.1 | 30.2 | 28.4 | 27.9 | 30.42 |
| 消防   | 31.9 | 32.8 | 31.6 | 31.2 | 30.8 | 30.6 | 28.7 | 27.5 | 30.64 |
| 监察   | 27.9 | 27.8 | 26.5 | 26.8 | 26.4 | 26.8 | 25.3 | 25.9 | 26.68 |
| 地区平均 | 30.7 | 31.3 | 29.6 | 29.5 | 29.1 | 29.2 | 27.5 | 27.1 | 29.25 |

a. 给这一研究写出一个模型，并解释模型中的所有各项。

b. 下面是用 Minitab 软件得到的方差分析结果。这些结果是否意味着在三个行业的平均起点工资之间存在着差异？取  $\alpha = 0.05$ 。

c. 给出你的检验的显著性水平。

d. 哪两种工作的起点工资之间有显著差异？

Two-Way Analysis of Variance for Salary

| Source | DF | SS      | MS      | F      | P     |
|--------|----|---------|---------|--------|-------|
| Job    | 2  | 62.8712 | 31.4356 | 955.23 | 0.000 |
| Region | 7  | 6.5213  | 0.9316  | 28.31  | 0.000 |
| Error  | 14 | 0.4607  | 0.0329  |        |       |
| Total  | 23 | 69.8532 |         |        |       |

15.20 参见练习 15.19。

- 以工作类型和地区作为因子,画出该数据的侧面图。图上是否显示出两个因子之间有交互效应?如果看出有交互效应,你能用这些数据对交互效应进行检验吗?如果不能,为什么?
- 地理分区变量的使用,比起不分地理区域而直接随机地抽取八个城市的完全随机化试验来,是否提高了效率?
- 找出在未来的研究中,可能需要考虑的其他变异来源。

15.21(农业) 为了比较四种锰的施用量(利用  $\text{MnSO}_4$ )与四种铜的施用量(利用  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ )对大豆产量的影响,进行一项研究。把一大块地分成了 32 个田块。随机地为 16 个因子水平组合(处理)中的每一个指定两个田块,然后,在整块地里种上大豆,每行间距离 3 英尺。32 个田块产量给出如下(千克/公顷)。

| 铜    | 锰       |         |         |         | 按铜平均    |
|------|---------|---------|---------|---------|---------|
|      | 20      | 50      | 80      | 110     |         |
| 1    | 1 558   | 2 003   | 2 490   | 2 830   | 2 221.5 |
|      | 1 578   | 2 033   | 2 470   | 2 810   |         |
| 3    | 1 590   | 2 020   | 2 620   | 2 860   | 2 278.0 |
|      | 1 610   | 2 051   | 2 632   | 2 841   |         |
| 5    | 1 558   | 2 003   | 2 490   | 2 830   | 2 255.1 |
|      | 1 550   | 2 010   | 2 690   | 2 910   |         |
| 7    | 1 328   | 2 010   | 2 887   | 2 960   | 2 302.0 |
|      | 1 427   | 2 031   | 2 832   | 2 941   |         |
| 按锰平均 | 1 524.9 | 2 020.1 | 2 638.9 | 2 872.8 | 2 264.2 |

- 指出这一试验的设计。
- 对于这一试验,写出一个适当的统计模型。
- 作出侧面图;这张图就锰和铜对于大豆产量的效应说明了什么?

15.22 参见练习 15.21。

- 用下面给出的计算机输出结果,检验锰和铜对于大豆产量的效应之间的交互效应,取  $\alpha = 0.05$ 。

## General Linear Models Procedure for Exercise 15.22

Dependent Variable: SOYBEAN YIELD

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 15 | 9167706.7      | 611180.4    | 305.08  | 0.0001 |
| Error           | 16 | 32053.5        | 2003.3      |         |        |
| Corrected Total | 31 | 9199760.2      |             |         |        |

| R-Square | C. V.    | Root MSE | Y Mean |
|----------|----------|----------|--------|
| 0.996516 | 1.976839 | 44.759   | 2264.2 |

| Source  | DF | Type III SS | Mean Square | F Value | Pr > F |
|---------|----|-------------|-------------|---------|--------|
| CU      | 3  | 28199.3     | 9399.8      | 4.69    | 0.0155 |
| MN      | 3  | 8935108.1   | 2978369.4   | 1486.70 | 0.0001 |
| CU * MN | 9  | 204399.3    | 22711.0     | 11.34   | 0.0001 |

- 锰的哪个水平显示出有最高的产量?
- 铜的哪个水平显示出有最高的产量?
- 锰和铜的哪个组合显示出有最高的产量?

**15.23** 假定我们有一个完全随机化三因子析因试验, 水平为  $3 \times 4 \times 6$ , 在这 72 个处理的每一个上有三次重复试验。假设三因子交互效应不显著。

- 写出一个模型来描述这种试验中的响应  $y_{ijkm}$ 。
- 给出一个这种试验的完整的 AOV 表。
- 草拟能够描述下列三种两因子交互效应的侧面图:  $F_1 * F_2$  显著但有序,  $F_2 * F_3$  不显著,  $F_1 * F_3$  显著但无序。

**15.24(农业)** 为了比较不同水平的土壤 pH 值和钙添加剂对于柑橘树树干直径增加的影响, 做一项试验。通过每年施用一次自然硫磺(elemental sulfur)、石膏肥料(gypsum)、苏打灰(soda ash)以及其他成分把 pH 值保持在 4, 5, 6 和 7, 钙添加剂使用了三个水平(100, 200 和 300 磅/英亩)。2 年以后, 在每个因子水平组合上考察了三棵树直径的增加量。数据如下。



| pH 值 | 钙   |     |     |
|------|-----|-----|-----|
|      | 100 | 200 | 300 |
| 4.0  | 5.2 | 7.4 | 6.3 |
|      | 5.9 | 7.0 | 6.7 |
|      | 6.3 | 7.6 | 6.1 |
| 5.0  | 7.1 | 7.4 | 7.3 |
|      | 7.4 | 7.3 | 7.5 |
|      | 7.5 | 7.1 | 7.2 |
| 6.0  | 7.6 | 7.6 | 7.2 |
|      | 7.2 | 7.5 | 7.3 |
|      | 7.4 | 7.8 | 7.0 |
| 7.0  | 7.2 | 7.4 | 6.8 |
|      | 7.5 | 7.0 | 6.6 |
|      | 7.2 | 6.9 | 6.4 |

- 构造一张侧面图。这些数据表明了什麼？
- 写出一个适当的统计模型。
- 进行方差分析,并指明该试验设计是什麼设计,取  $\alpha = 0.05$ 。

**15.25** 参见练习 15.24。

- 用下面给出的计算机输出结果检验交互效应和主效应,取  $\alpha = 0.05$ 。
- 关于土壤 pH 值和钙添加剂对于柑橘树的平均树干直径增加的效应,你得到什麼结论？

General Linear Models Procedure for Exercise 15.25

Dependent Variable: Increase in Tree Diameter

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 11 | 9.1830556      | 0.8348232   | 12.32   | 0.0001 |
| Error           | 24 | 1.6266667      | 0.0677778   |         |        |
| Corrected Total | 35 | 10.8097222     |             |         |        |

| R-Square | C. V.    | Root MSE | Y Mean |
|----------|----------|----------|--------|
| 0.849518 | 3.691335 | 0.2603   | 7.0528 |

| Source  | DF | Type III SS | Mean Square | F Value | Pr > F |
|---------|----|-------------|-------------|---------|--------|
| PH      | 3  | 4.4608333   | 1.4869444   | 21.94   | 0.0001 |
| CA      | 2  | 1.4672222   | 0.7336111   | 10.82   | 0.0004 |
| PH * CA | 6  | 3.2550000   | 0.5425000   | 8.00    | 0.0001 |

**15.26** 参见练习 15.24。

- 用 Tukey 的 W 方法决定三种钙(CA)的用量所对应的树干直径的平均增量之间的差异, 取  $\alpha = 0.05$ 。
- 你的关于三种钙的用量所对应的树干直径的平均增量之间的差异的结论, 是否对于所有四个 pH 值的水平都是一样的?

| Level of |   |            | -----Y----- |
|----------|---|------------|-------------|
| PH       | N | Mean       | SD          |
| 4        | 9 | 6.50000000 | 0.75828754  |
| 5        | 9 | 7.31111111 | 0.15365907  |
| 6        | 9 | 7.40000000 | 0.25000000  |
| 7        | 9 | 7.00000000 | 0.36400549  |

| Level of |    |            | -----Y----- |
|----------|----|------------|-------------|
| CA       | N  | Mean       | SD          |
| 100      | 12 | 6.95833333 | 0.75252102  |
| 200      | 12 | 7.33333333 | 0.28069179  |
| 300      | 12 | 6.86666667 | 0.45193188  |

| Level of | Level of |   | -----Y----- |            |
|----------|----------|---|-------------|------------|
| PH       | CA       | N | Mean        | SD         |
| 4        | 100      | 3 | 5.80000000  | 0.55677644 |
| 4        | 200      | 3 | 7.33333333  | 0.30550505 |
| 4        | 300      | 3 | 6.36666667  | 0.30550505 |
| 5        | 100      | 3 | 7.33333333  | 0.20816660 |
| 5        | 200      | 3 | 7.26666667  | 0.15275252 |
| 5        | 300      | 3 | 7.33333333  | 0.15275252 |
| 6        | 100      | 3 | 7.40000000  | 0.20000000 |
| 6        | 200      | 3 | 7.63333333  | 0.15275252 |
| 6        | 300      | 3 | 7.16666667  | 0.15275252 |
| 7        | 100      | 3 | 7.30000000  | 0.17320508 |

|   |     |   |            |            |
|---|-----|---|------------|------------|
| 7 | 200 | 3 | 7.10000000 | 0.26457513 |
| 7 | 300 | 3 | 6.60000000 | 0.20000000 |

15.27 参见练习 15.24。

- 用包含在下述计算机输出结果中的残差分析确定在做适当的  $F$  检验所需要的条件中, 是否有的条件得不到满足。
- 如果有任何条件得不到满足, 试给出一种办法来克服这一困难。

Variable-RESIDUALS

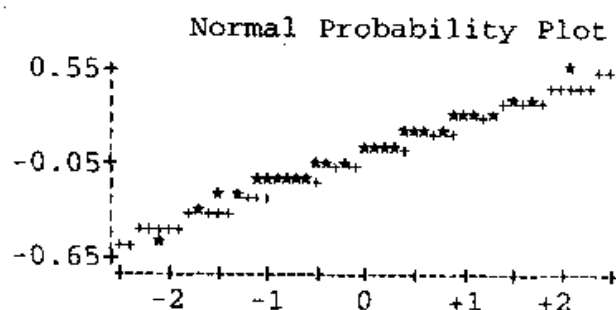
Moments

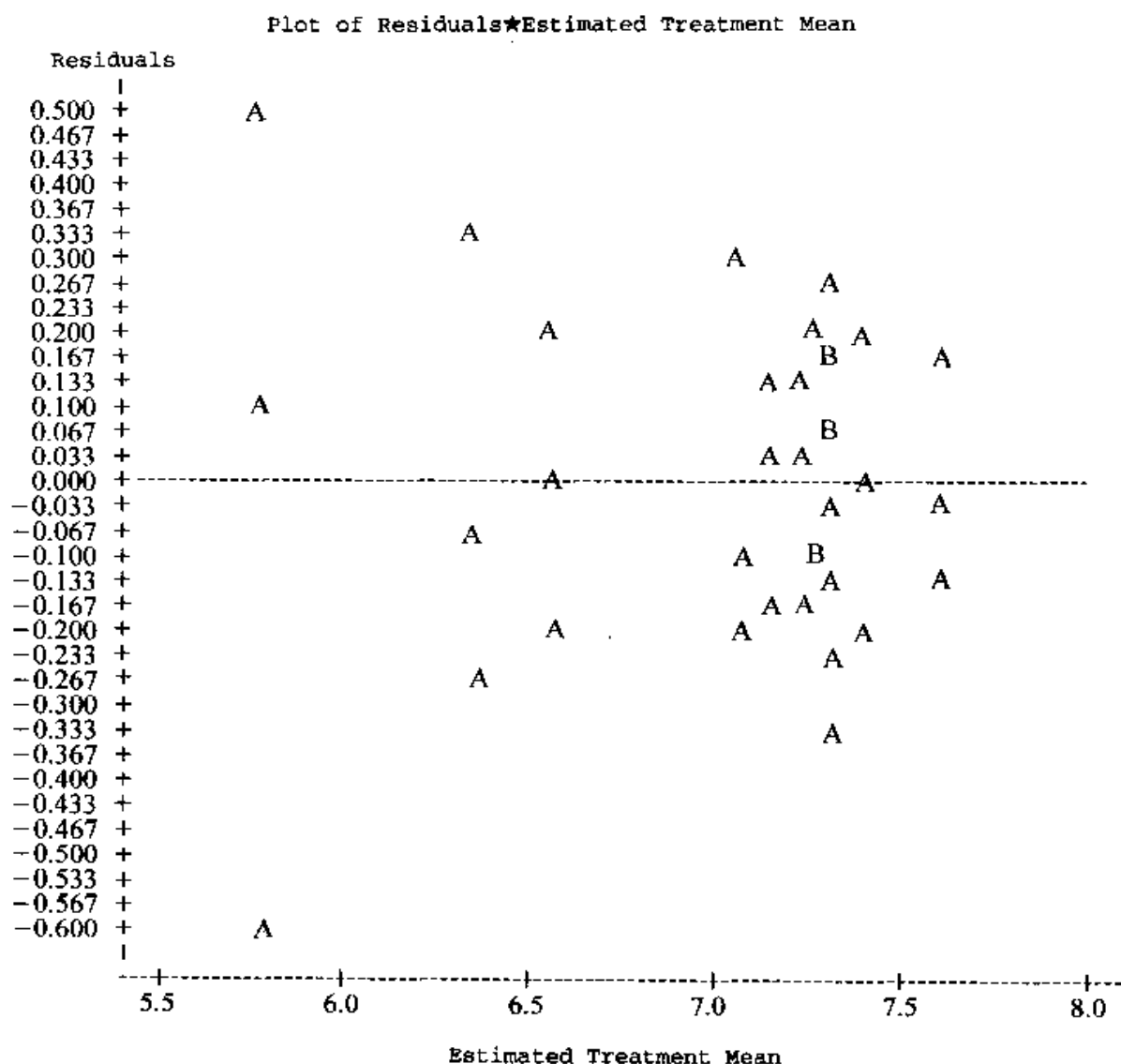
|          |          |          |          |
|----------|----------|----------|----------|
| N        | 36       | Sum Wgts | 36       |
| Mean     | 0        | Sum      | 0        |
| Std Dev  | 0.215583 | Variance | 0.046476 |
| Skewness | -0.22276 | Kurtosis | 0.699897 |
| W:Normal | 0.986851 | Pr<W     | 0.9562   |

Variable-RESIDUALS

| Stem | Leaf    | # | Boxplot |
|------|---------|---|---------|
| 5    | 0       | 1 |         |
| 4    |         |   |         |
| 3    | 03      | 2 |         |
| 2    | 0007    | 4 |         |
| 1    | 033777  | 6 |         |
| 0    | 003377  | 6 |         |
| -0   | 733     | 3 |         |
| -1   | 7733000 | 7 |         |
| -2   | 73000   | 5 |         |
| -3   | 3       | 1 |         |
| -4   |         |   |         |
| -5   |         |   |         |
| -6   | 0       | 1 |         |

-----+-----+-----+  
Multiply Stem.Leaf by 10\*\*-1





**15.28(医药)** 研究人员进行了一项试验,以比较患者的平均口腔体温,这些患者在九种高血压常见处方药物治疗方案中接受了其中的一种。研究人员关心药物的疗效是否与患者的高血压紊乱的严重程度有关。把满足这项研究的准入条件的高血压患者按照血压紊乱的严重程度分为三个水平。然后,为这些患者随机地指定九种药物疗法之一。研究中的每一位患者,在指定的研究期间,每天早晨6:00点服用给定的药物。从早晨8:00开始,每一个小时量一次体温,持续10个小时。在此期间,患者不允许做任何体力活动,并且必须躺在床上。为了摆脱日间体温变化的影响,记录下每个患者10次体温记录的平均值。这些数据在下面的表中给出。

- a. 指明这个试验的设计是什么设计。
- b. 写出一个适当的统计模型,并说明模型中参数的含义。

| 严重<br>程度 | 药 物  |      |      |      |      |      |      |      |      |
|----------|------|------|------|------|------|------|------|------|------|
|          | A    | B    | C    | D    | E    | F    | G    | H    | I    |
| 1        | 97.8 | 98.1 | 98.0 | 97.3 | 97.9 | 97.9 | 97.1 | 98.0 | 97.8 |
|          | 97.2 | 98.1 | 97.8 | 97.3 | 97.8 | 97.9 | 97.6 | 97.8 | 98.0 |
|          | 97.6 | 98.0 | 98.1 | 97.5 | 97.8 | 97.8 | 97.3 | 98.0 | 97.7 |
|          | 97.2 | 97.7 | 97.8 | 97.5 | 97.7 | 97.8 | 97.7 | 97.9 | 97.9 |
|          | 97.6 | 97.7 | 97.9 | 97.6 | 97.8 | 97.6 | 97.5 | 98.0 | 97.8 |
| 2        | 97.6 | 97.8 | 97.9 | 97.5 | 97.8 | 98.0 | 97.6 | 97.9 | 98.0 |
|          | 97.4 | 97.7 | 98.1 | 97.4 | 97.8 | 97.7 | 97.5 | 98.0 | 97.6 |
|          | 97.3 | 97.6 | 97.8 | 97.5 | 97.7 | 97.8 | 97.6 | 97.9 | 98.0 |
|          | 97.5 | 97.7 | 97.8 | 97.6 | 97.7 | 97.9 | 97.5 | 97.9 | 97.9 |
|          | 97.5 | 97.7 | 97.6 | 97.7 | 97.8 | 97.8 | 97.3 | 97.8 | 97.9 |
| 3        | 97.5 | 97.6 | 98.0 | 97.9 | 97.7 | 97.9 | 97.4 | 97.8 | 98.0 |
|          | 97.9 | 97.7 | 97.8 | 97.8 | 97.8 | 98.0 | 97.8 | 97.8 | 98.1 |
|          | 97.6 | 97.9 | 98.1 | 97.8 | 97.9 | 97.7 | 97.4 | 98.0 | 97.9 |
|          | 97.6 | 97.9 | 97.7 | 97.8 | 98.0 | 97.9 | 97.6 | 97.9 | 98.1 |
|          | 97.7 | 97.8 | 98.7 | 97.6 | 98.1 | 97.9 | 97.6 | 97.8 | 97.9 |

15.29 参见练习 15.28。

a. 对该试验,完成下面的方差分析表。

| 来源   | SS     | df | MS | F |
|------|--------|----|----|---|
| 严重程度 | 0.3628 |    |    |   |
| 药物   | 3.5117 |    |    |   |
| 交互效应 | 0.5012 |    |    |   |
| 误差   | 2.6520 |    |    |   |
| 总和   | 7.0277 |    |    |   |

- b. 九种药物所对应的平均体温之间的差异,是否对于所有的三种血压紊乱的严重程度来说都一样? 取  $\alpha = 0.05$ 。
- c. 对于不同的药物和不同的严重程度,平均体温之间是否存在显著差异? 取  $\alpha = 0.05$ 。
- d. 使用侧面图来帮助你讨论关于药物和严重程度对患者的平均体温的效应。

**15.30(医药)** 一位医生要考察人们在锻炼极限测试中的表现与其超出正常的体重(由标准体重-身高表决定)之间的关系。为此,他抽取了由 28 位健康成年女性组成的样本,这些女性是从到社区诊所做日常体检的人当中随机挑选的,她们的年龄从 25 岁到 40 岁。随机挑选的过程使得在下述四种体重的分类中各抽选七人。

正常体重(比标准体重-身高表中的体重值轻不超过 10 %)

1% ~ 10% 超重

11% ~ 20% 超重

超重 20% 以上

作为体检的一部分,要求每个人在带有测定仪的自行车上做运动,直到疲劳为止。记录下每个人到疲劳的时间(按分钟计算)。数据如下。

| 分类           | 疲劳时间                       |
|--------------|----------------------------|
| 正常           | 25, 28, 19, 27, 23, 30, 35 |
| 1% ~ 10% 超重  | 24, 26, 18, 16, 14, 12, 17 |
| 11% ~ 20% 超重 | 15, 18, 17, 25, 12, 10, 23 |
| 超重 20% 以上    | 10, 9, 18, 14, 6, 4, 15    |

a. 指出该试验设计是什么设计,并写出一个适当的统计模型。

b. 取  $\alpha = 0.05$ , 进行方差分析。

**15.31** 参见练习 15.30。

a. 为了调查年龄、性别和超出正常的体重对于疲劳时间的影响,你如何设计一个试验?

b. 假定该医生要研究三个定量变量,即超出正常体重的百分比、年龄和疲劳时间之间的关系,写出一个可能的模型。

**15.32(环境)** 为了调查五种不同设计的商用保温玻璃窗格的热量损失,进行一项试验。研究人员为了得到适用于国家大部分地区的结果,决定在五个温度下对这些玻璃窗格进行评估。这五个温度是  $0^{\circ}\text{F}$ ,  $20^{\circ}\text{F}$ ,  $40^{\circ}\text{F}$ ,  $60^{\circ}\text{F}$  和  $80^{\circ}\text{F}$ 。从每种设计类型的玻璃窗格中抽取 10 个样品,把其中每两个玻璃窗格样品随机地指定给

五个外部温度设置之一。对于所有的五个外部温度,内部温度都控制在 70°F。五种玻璃窗格设计的热量损失给出如下。

- 指出该试验设计是什么设计,并写出一个适当的统计模型。
- 五种玻璃窗格设计的平均热量损失之间存在显著差异吗? 取  $\alpha = 0.05$ 。这里给出这些数据的 AOV 表。
- 在所有五个温度上,这五种设计之间的差异是否保持一致? 取  $\alpha = 0.05$ ,并画出侧面图来说明你的结论。
- 取  $\alpha = 0.05$ ,用 Tukey 的 W 方法比较五种设计的平均热量损失。

| 外部温度<br>设置(°F) | 玻璃窗格设计     |            |            |            |          |
|----------------|------------|------------|------------|------------|----------|
|                | A          | B          | C          | D          | E        |
| 80             | 7.2, 7.8   | 7.1, 7.9   | 8.1, 8.8   | 8.3, 8.9   | 9.3, 9.8 |
| 60             | 8.1, 8.1   | 8.0, 8.9   | 8.2, 8.9   | 8.1, 8.8   | 9.2, 9.9 |
| 40             | 9.0, 9.9   | 9.2, 9.8   | 10.0, 10.8 | 10.2, 10.7 | 9.9, 9.0 |
| 20             | 9.2, 9.8   | 9.1, 9.9   | 10.1, 10.8 | 10.3, 10.9 | 9.3, 9.8 |
| 0              | 10.2, 10.8 | 10.1, 10.9 | 11.1, 11.8 | 11.3, 11.9 | 9.3, 9.9 |

General Linear Models Procedure for Exercise 15.32

Dependent Variable: HEAT LOSS

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 24 | 58.07280000    | 10.47   | 0.0001 |
| Error           | 25 | 5.78000000     |         |        |
| Corrected Total | 49 | 63.85280000    |         |        |

|          |          |            |
|----------|----------|------------|
| R-Square | C. V.    | Y Mean     |
| 0.909479 | 5.067797 | 9.48800000 |

| Source | DF | Type III SS | F Value | Pr > F |
|--------|----|-------------|---------|--------|
| T      | 4  | 39.77880000 | 43.01   | 0.0001 |
| D      | 4  | 7.32280000  | 7.92    | 0.0003 |
| T * D  | 16 | 10.97120000 | 2.97    | 0.0073 |

| T  | D | N | Mean       | T  | N  | Mean       | D | N  | Mean       |
|----|---|---|------------|----|----|------------|---|----|------------|
| 0  | a | 2 | 10.5000000 |    |    |            | a | 10 | 9.01000000 |
| 0  | b | 2 | 10.5000000 | 0  | 10 | 10.7300000 | b | 10 | 9.09000000 |
| 0  | c | 2 | 11.4500000 | 20 | 10 | 9.9200000  | c | 10 | 9.86000000 |
| 0  | d | 2 | 11.6000000 | 40 | 10 | 9.8500000  | d | 10 | 9.94000000 |
| 0  | e | 2 | 9.5000000  | 60 | 10 | 8.6200000  | e | 10 | 9.54000000 |
| 20 | a | 2 | 9.5000000  | 80 | 10 | 8.3200000  |   |    |            |
| 20 | b | 2 | 9.5000000  |    |    |            |   |    |            |
| 20 | c | 2 | 10.4500000 |    |    |            |   |    |            |
| 20 | d | 2 | 10.6000000 |    |    |            |   |    |            |

**15.33** 参见练习 15.24。在这个试验中,研究人员没有注意到试验实际上是在四个不同的柑橘园中进行的,这些柑橘园位于不同的州。柑橘园 1 中的土壤 pH 值为 4.0,柑橘园 2 中的土壤 pH 值为 5.0,柑橘园 3 中的土壤 pH 值为 6.0,柑橘园 4 中的土壤 pH 值为 7.0。在每个柑橘园中,为三棵树随机地指定三个钙添加剂水平 100, 200, 300(磅/英亩)之一。数据如下。

| 柑橘园 | pH 值 | 钙             |               |               |
|-----|------|---------------|---------------|---------------|
|     |      | 100           | 200           | 300           |
| 1   | 4.0  | 5.2, 5.9, 6.3 | 7.4, 7.0, 7.6 | 6.3, 6.7, 6.1 |
| 2   | 5.0  | 7.1, 7.4, 7.5 | 7.4, 7.3, 7.1 | 7.3, 7.5, 7.2 |
| 3   | 6.0  | 7.6, 7.2, 7.4 | 7.6, 7.5, 7.8 | 7.2, 7.3, 7.0 |
| 4   | 7.0  | 7.2, 7.5, 7.2 | 7.4, 7.0, 6.9 | 6.8, 6.6, 6.4 |

- 练习 15.24 中所得到的关于土壤 pH 值和钙对于树干直径平均增量的效应的结论,在有了这些新的信息以后,有何变化?
- 设计一个新的试验,以使土壤 pH 值和钙对于树干直径平均增量的效应能够得到有效的评估。所有四个柑橘园必须与 pH 值的四个水平和钙的三个水平一起用在你的设计中。

**15.34(心理学)** 为了解不同的强化水平和隔离水平对于儿童回忆能力的影响,进行一项试验。分析人员从一个相当一致的四年级班组中挑选了由 36 个儿童组成的随机样本。试验中使用了两个水平的强化(无强化和口头强化)和三个水平的隔离(20, 40, 60 分钟)。把这些学生随机地分配到六个处理组,每个处理组有六个学生。

每个学生要与分析人员共同度过 30 分钟。在此期间,该生要记忆一段指定的内容,并按照为他(她)指定的处理组的规定进行相应的强化。这 30 分钟过后,按



照他(她)所在的处理组的规定,把该学生隔离相应的时间,然后检验他(她)对所记忆的那段内容的回忆情况。数据如下。

| 强化水平 | 隔离时间(分钟) |    |    |    |    |    |
|------|----------|----|----|----|----|----|
|      | 20       |    | 40 |    | 60 |    |
| 无强化  | 26       | 19 | 30 | 36 | 6  | 10 |
|      | 23       | 18 | 25 | 28 | 11 | 14 |
|      | 28       | 25 | 27 | 24 | 17 | 19 |
| 口头强化 | 15       | 16 | 24 | 26 | 31 | 38 |
|      | 24       | 22 | 29 | 27 | 29 | 34 |
|      | 25       | 21 | 23 | 21 | 35 | 30 |

用这里给出的计算机输出结果,得出你的结论。

General Linear Models Procedure for Exercise 15.34

Dependent Variable: TEST SCORE

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr>F   |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 5  | 1410.8889      | 282.1778    | 17.88   | 0.0001 |
| Error           | 30 | 473.3333       | 15.7778     |         |        |
| Corrected Total | 35 | 1884.2222      |             |         |        |

| R-Square | C. V.    | Root MSE | Y Mean |
|----------|----------|----------|--------|
| 0.748791 | 16.70520 | 3.9721   | 23.778 |

| Source      | DF | Type III SS | Mean Square | F Value | Pr>F   |
|-------------|----|-------------|-------------|---------|--------|
| REINFORCE   | 1  | 196.0000    | 196.0000    | 12.42   | 0.0014 |
| TIME        | 2  | 156.2222    | 78.1111     | 4.95    | 0.0139 |
| INTERACTION | 2  | 1058.6667   | 529.3333    | 33.55   | 0.0001 |

**15.35(商业)** 一家食品加工厂测试了一种新型早餐饮料的不同配方。共有 12 种配方,这些配方是由三个甜度水平,两个卡路里含量水平和两种颜色搭配而成的。六个评审小组为这 12 种饮料配方打了分。

- 指出该设计是什么设计。
- 写出一个适当的模型。
- 给出该设计的方差分析表。

| 甜度水平 | 颜 色   |      |       |      |
|------|-------|------|-------|------|
|      | 1     |      | 2     |      |
|      | 卡路里水平 |      | 卡路里水平 |      |
|      | 1     | 2    | 1     | 2    |
| 1    | 59.5  | 42.5 | 54.5  | 40.1 |
| 2    | 66.8  | 49.6 | 64.7  | 50.1 |
| 3    | 52.0  | 39.3 | 35.1  | 30.2 |

**15.36** 对于练习 15.35 中的试验设计,算出的方差分析表如下。这个表中缺少了什么?

| 来源   | SS          | df | MS          | F 值    | Pr>F   |
|------|-------------|----|-------------|--------|--------|
| 主效应  |             |    |             |        |        |
| A    | 4,149.55556 | 2  | 2,074.76389 | 75.51  | 0.0001 |
| B    | 624.22222   | 1  | 624.22222   | 22.72  | 0.0001 |
| C    | 3,200.00000 | 1  | 3,200.00000 | 116.46 | 0.0001 |
| 交互效应 |             |    |             |        |        |
| AB   | 488.52778   | 2  | 244.26389   | 8.89   | 0.0004 |
| AC   | 203.08333   | 2  | 101.54167   | 3.70   | 0.0307 |
| BC   | 80.22222    | 1  | 80.22222    | 2.92   | 0.0927 |
| ABC  | 24.19444    | 2  | 12.09722    | 0.44   | 0.6459 |
| 误差   | 1,648.66667 | 60 | 27.47778    |        |        |

**15.37** 参见练习 15.36。假定评审小组之间没有变异(从而 MSE 可以作为误差的度量),试提取关于这些配方的结论。基于这里给定的处理均值,哪个配方看起来与众不同? 所看到的一切能够从一系列侧面图中得到解释吗? 为什么?

**15.38 (人力资源)** 从一家公司的三个部门分别抽取了 12, 9, 18 个雇员组成的随机样本,并对他们进行数值打分以评价他们的工作表现。概括数据如下。

| 部门 | $n$ | $\bar{y}$ | $s$ |
|----|-----|-----------|-----|
| 研究 | 12  | 21.2      | 8.3 |
| 开发 | 9   | 15.4      | 7.3 |
| 营销 | 18  | 27.4      | 8.2 |

- a. 指出该设计的类型。  
b. 写出一个适当的模型。

**15.39** 参见练习 15.38。取  $\alpha = 0.10$  进行方差分析, 并得到有关结论。  
(注: 分数越高表现越好)

**15.40(医药)** 研究人员对一种药品在四个长度不同的储存期内的稳定性感兴趣。四个储存期为 1, 3, 6 和 9 个月。这种药是由 30mg/mL 某种药品的活性成分制成的。储存期过后, 测量药品中活性成分的含量。这些药品储存在 30°C 的恒温条件下。研究中使用了两个实验室, 每三个 2mL 装药瓶被随机地指定给四个储存期之一。储存期终结时, 测定每瓶中活性成分的含量。对于每一瓶药品, 还测量并记录了其 pH 值。数据如下。

- a. 写出反映每一瓶上的响应值与其因子, 即储存时间长度和实验室之间的关系模型。  
b. 列出 a 中模型的方差分析表, 但不必计算平方和。

| 时间(在<br>30°C 下保<br>存的月<br>数) | 实验室 | 活性成分<br>(mg/mL) | pH   | 时间(在<br>30°C 下保<br>存的月<br>数) | 实验室 | 活性成分<br>(mg/mL) | pH   |
|------------------------------|-----|-----------------|------|------------------------------|-----|-----------------|------|
| 1                            | 1   | 30.03           | 3.61 | 1                            | 2   | 30.12           | 3.87 |
| 1                            | 1   | 30.10           | 3.60 | 1                            | 2   | 30.10           | 3.80 |
| 1                            | 1   | 30.14           | 3.57 | 1                            | 2   | 30.02           | 3.84 |
| 3                            | 1   | 30.10           | 3.50 | 3                            | 2   | 29.90           | 3.70 |
| 3                            | 1   | 30.18           | 3.45 | 3                            | 2   | 29.95           | 3.80 |
| 3                            | 1   | 30.23           | 3.48 | 3                            | 2   | 29.85           | 3.75 |
| 6                            | 1   | 30.03           | 3.56 | 6                            | 2   | 29.75           | 3.90 |
| 6                            | 1   | 30.03           | 3.74 | 6                            | 2   | 29.85           | 3.90 |
| 6                            | 1   | 29.96           | 3.81 | 6                            | 2   | 29.80           | 3.90 |
| 9                            | 1   | 29.81           | 3.60 | 9                            | 2   | 29.75           | 3.77 |
| 9                            | 1   | 29.79           | 3.55 | 9                            | 2   | 29.85           | 3.74 |
| 9                            | 1   | 29.82           | 3.59 | 9                            | 2   | 29.80           | 3.76 |

**15.41** 参见练习 15.40。下面给出了对于两个因变量(即  $y_1 = \text{mg/mL}$  活性成分和  $y_2 = \text{pH}$ )所进行的方差分析的计算机输出结果。基于这些分析, 提取关于这些 2mL 瓶装药品的稳定性的结论, 取  $\alpha = 0.05$ 。

## General Linear Models Procedure for Exercise 15.41

Dependent Variable:MG/ML

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 7  | 0.4674000      | 0.0667714   | 27.30   | 0.0001 |
| Error           | 16 | 0.0391333      | 0.0024458   |         |        |
| Corrected Total | 23 | 0.5065333      |             |         |        |

| R-Square | C. V.    | Root MSE | Y1 Mean |
|----------|----------|----------|---------|
| 0.922743 | 0.165090 | 0.0495   | 29.957  |

| Source     | DF | Type III SS | Mean Square | F Value | Pr > F |
|------------|----|-------------|-------------|---------|--------|
| TIME       | 3  | 0.2937667   | 0.0979222   | 40.04   | 0.0001 |
| LAB        | 1  | 0.0912667   | 0.0912667   | 37.32   | 0.0001 |
| TIME * LAB | 3  | 0.0823667   | 0.0274556   | 11.23   | 0.0003 |

## General Linear Models Procedure

Dependent Variable:pH

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 7  | 0.4201625      | 0.0600232   | 21.47   | 0.0001 |
| Error           | 16 | 0.0447333      | 0.0027958   |         |        |
| Corrected Total | 23 | 0.4648958      |             |         |        |

| R-Square | C. V.    | Root MSE | Y2 Mean |
|----------|----------|----------|---------|
| 0.903778 | 1.429232 | 0.0529   | 3.6996  |

| Source     | DF | Type III SS | Mean Square | F Value | Pr > F |
|------------|----|-------------|-------------|---------|--------|
| TIME       | 3  | 0.1144458   | 0.0381486   | 13.64   | 0.0001 |
| LAB        | 1  | 0.2970375   | 0.2970375   | 106.24  | 0.0001 |
| TIME * LAB | 3  | 0.0086792   | 0.0028931   | 1.03    | 0.4038 |

**15.42** 参见练习 15.40。与在练习 15.40 中一样,同种类型的关于 mg/mL 和 pH 的数据在 40°C 上产生。数据给出如下。

- 下面分别对响应变量  $y_1 = \text{mg/mL}$  和  $y_2 = \text{pH}$  给出了相应于练习 15.40(a) 中模型的方差分析的计算机输出结果。三个因子对于药品的平均活性成分的效应如何? 取  $\alpha = 0.05$ 。
- 下面分别对响应变量  $y_1 = \text{mg/mL}$  和  $y_2 = \text{pH}$  给出了相应于练习 15.40(a)

中模型的方差分析的计算机输出结果。三个因子对于药品的平均 pH 值的效应如何？取  $\alpha = 0.05$ ，并画出侧面图来说明你的结论。

- c. 你的关于储存期和温度对于平均 pH 值的影响的结论，就两个实验室来说，是否是一样的？
- d. 你的关于储存期和温度对于药品的平均活性成分的影响的结论，就两个实验室来说，是否是一样的？

| 时间(在<br>40°C 下保<br>存的月<br>数) | 实验室 | 活性成分<br>(mg/mL) | pH   | 时间(在<br>40°C 下保<br>存的月<br>数) | 实验室 | 活性成分<br>(mg/mL) | pH   |
|------------------------------|-----|-----------------|------|------------------------------|-----|-----------------|------|
| 1                            | 1   | 30.08           | 3.61 | 1                            | 2   | 30.12           | 3.80 |
| 1                            | 1   | 30.10           | 3.60 | 1                            | 2   | 30.10           | 3.70 |
| 1                            | 1   | 30.14           | 3.59 | 1                            | 2   | 30.02           | 3.81 |
| 3                            | 1   | 30.03           | 3.39 | 3                            | 2   | 29.90           | 3.70 |
| 3                            | 1   | 30.18           | 3.45 | 3                            | 2   | 29.85           | 3.80 |
| 3                            | 1   | 30.26           | 3.29 | 3                            | 2   | 29.80           | 3.75 |
| 6                            | 1   | 29.90           | 3.63 | 6                            | 2   | 29.75           | 3.80 |
| 6                            | 1   | 29.90           | 3.71 | 6                            | 2   | 29.70           | 3.70 |
| 6                            | 1   | 29.96           | 3.65 | 6                            | 2   | 29.75           | 3.70 |
| 9                            | 1   | 29.81           | 3.51 | 9                            | 2   | 29.65           | 3.64 |
| 9                            | 1   | 29.85           | 3.38 | 9                            | 2   | 29.75           | 3.68 |
| 9                            | 1   | 29.72           | 3.32 | 9                            | 2   | 29.70           | 3.60 |

General Linear Models Procedure for Exercise 15.42

Dependent Variable: MG/ML

| Source          | DF | Sum of<br>Squares | Mean<br>Square | F Value | Pr > F |
|-----------------|----|-------------------|----------------|---------|--------|
| Model           | 15 | 1.1475917         | 0.0765061      | 25.29   | 0.0001 |
| Error           | 32 | 0.0968000         | 0.0030250      |         |        |
| Corrected Total | 47 | 1.2443917         |                |         |        |

| R-Square | C. V.    | Root MSE | Y1 Mean |
|----------|----------|----------|---------|
| 0.922211 | 0.183719 | 0.0550   | 29.937  |

| Source            | DF | Type III SS | Mean Square | F Value | Pr > F |
|-------------------|----|-------------|-------------|---------|--------|
| TIME              | 3  | 0.7360083   | 0.2453361   | 81.10   | 0.0001 |
| LAB               | 1  | 0.22963333  | 0.2296333   | 75.91   | 0.0001 |
| TIME * LAB        | 3  | 0.1443167   | 0.0481056   | 15.90   | 0.0001 |
| TEMP              | 1  | 0.0184083   | 0.0184083   | 6.09    | 0.0192 |
| TIME * TEMP       | 3  | 0.0120750   | 0.0040250   | 1.33    | 0.2817 |
| LAB * TEMP        | 1  | 0.0027000   | 0.0027000   | 0.89    | 0.3519 |
| TIME * LAB * TEMP | 3  | 0.0044500   | 0.0014833   | 0.49    | 0.6914 |

General Linear Models Procedure

Dependent Variable: pH

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 15 | 0.9848667      | 0.0656578   | 20.48   | 0.0001 |
| Error           | 32 | 0.1026000      | 0.0032063   |         |        |
| Corrected Total | 47 | 1.0874667      |             |         |        |

| R-Square | C. V.    | Root MSE | Y2 Mean |
|----------|----------|----------|---------|
| 0.905652 | 1.547802 | 0.0566   | 3.6583  |

| Source            | DF | Type III SS | Mean Square | F Value | Pr > F |
|-------------------|----|-------------|-------------|---------|--------|
| TIME              | 3  | 0.2286000   | 0.0762000   | 23.77   | 0.0001 |
| LAB               | 1  | 0.5676750   | 0.5676750   | 177.05  | 0.0001 |
| TIME * LAB        | 3  | 0.0555583   | 0.0185194   | 5.78    | 0.0028 |
| TEMP              | 1  | 0.0816750   | 0.0816750   | 25.47   | 0.0001 |
| TIME * TEMP       | 3  | 0.0250917   | 0.0083639   | 2.61    | 0.0686 |
| LAB * TEMP        | 1  | 0.0003000   | 0.0003000   | 0.09    | 0.7617 |
| TIME * LAB * TEMP | 3  | 0.0259667   | 0.0086556   | 2.70    | 0.0621 |

**15.43(商业)** 某制造商的日常原材料供应是不定的,并且受到一定的限制。该制造商可以用这些原材料生产两种不同的产品,并且可以调整这两种产品的生产比例。生产每一种产品时,从单位原材料所获得的利润依赖于生产该种产品的持续时间,从而依赖于指定生产这种产品的原材料数量。其他因子,如工人生产,机械故障,等等,也影响每单位原材料的利润,但它们对于利润的效应是随机的、不可控制的。制造商已经进行了一个试验,用来考察原材料供应水平,  $S_1$  和指定给

两种产品生产线的原材料的比例,  $R$ , 对于每单位原材料利润的影响。最终的目标是对于每天的原材料供应量  $S$ , 选择最好的比例  $R$ 。试验中选用的原材料的供应水平为 15, 18 和 21 吨, 指定给两种产品生产线的比例为 1/2, 1 和 2。响应值是一天之中单位原材料的利润(以分为单位)。在每个组合上按随机的次序重复了三次。这 27 天的数据列在下面的表中。

| 原材料分配的<br>比例( $R$ ) | 原材料供应(吨)   |            |            |
|---------------------|------------|------------|------------|
|                     | 15         | 18         | 21         |
| 1/2                 | 22, 20, 21 | 21, 19, 20 | 19, 18, 20 |
| 1                   | 21, 20, 19 | 23, 24, 22 | 20, 19, 21 |
| 2                   | 17, 18, 16 | 21, 11, 20 | 20, 22, 24 |

- 基于这里给出的方差分析结果, 给出有关结论, 取  $\alpha = 0.05$ 。
- 找出  $S$  和  $R$  的两个最好的组合。这两个组合有显著的差别吗? 限定所有成对组合比较的犯错误的概率不超过 0.05。

General Linear Models Procedure for Exercise 15.43

Dependent Variable: PROFIT

| Source          | DF | Sum of<br>Squares | Mean<br>Square | F Value | Pr > F |
|-----------------|----|-------------------|----------------|---------|--------|
| Model           | 8  | 93.185185         | 11.648148      |         |        |
| Error           | 18 | 82.666667         | 4.592593       | 2.54    | 0.0482 |
| Corrected Total | 26 | 175.851852        |                |         |        |

| R-Square | C. V.    | Root MSE | Y Mean |
|----------|----------|----------|--------|
| 0.529907 | 10.75500 | 2.1430   | 19.926 |

| Source         | DF | Type III SS | Mean Square | F Value | Pr > F |
|----------------|----|-------------|-------------|---------|--------|
| RATIO          | 2  | 22.296296   | 11.148148   | 2.43    | 0.1166 |
| SUPPLY         | 2  | 4.962963    | 2.481481    | 0.54    | 0.5917 |
| RATIO * SUPPLY | 4  | 65.925926   | 16.481481   | 3.59    | 0.0255 |

PROFIT MEANS

| Level of<br>RATIO | Level of<br>SUPPLY | MEAN  | RATIO | MEANS | SUPPLY | MEANS |
|-------------------|--------------------|-------|-------|-------|--------|-------|
| 0.5               | 15                 | 21.00 | 0.5   | 20.00 | 15     | 19.33 |

|     |    |       |     |       |    |       |
|-----|----|-------|-----|-------|----|-------|
| 0.5 | 18 | 20.00 | 1.0 | 21.00 | 18 | 20.11 |
| 0.5 | 21 | 19.00 | 2.0 | 18.78 | 21 | 20.33 |
| 1.0 | 15 | 20.00 |     |       |    |       |
| 1.0 | 18 | 23.00 |     |       |    |       |
| 1.0 | 21 | 20.00 |     |       |    |       |
| 2.0 | 15 | 17.00 |     |       |    |       |
| 2.0 | 18 | 17.33 |     |       |    |       |
| 2.0 | 21 | 22.00 |     |       |    |       |

**15.44(商业)** 某制造商经常给居住在另一城市的顾客通过航空邮递寄送小型包裹,并且在很多情况下,寄出的包裹尽快到达顾客手中这一点是很重要的。有三家不同的公司提供航空邮递服务,包括接收邮件和在 24 小时内投递给收件人。制造厂家邮送部门的负责人希望知道这几家公司的邮送速度有无差异,以及邮寄的时间是否有影响。为了了解这些问题,设计了一个试验。包裹在随机选定的时刻送出,投送每个包裹的公司也是随机选择的。顾客记录下每个包裹到达的时间,从而可以计算出每个包裹在邮送途中的时间。这些时间被舍入到最接近的小时数。共 54 个包裹的试验结果由下表给出。

| 时间 | 公 司                   |                        |                     |
|----|-----------------------|------------------------|---------------------|
|    | 1                     | 2                      | 3                   |
| 上午 | 8, 6, 6, 12, 7, 8     | 11, 11, 9, 10, 8, 11   | 7, 4, 6, 4, 9, 7    |
| 下午 | 7, 10, 8, 11, 9, 11   | 10, 13, 10, 12, 11, 10 | 10, 8, 6, 5, 8, 6   |
| 晚上 | 13, 11, 14, 11, 9, 12 | 12, 6, 9, 9, 10, 6     | 8, 11, 9, 9, 10, 12 |

- 如果先用虚拟变量的方法代替 AOV 方法进行分析,为了在模型中既包含各个主效应也包含各个交互效应,需要几个虚拟变量? 对于这个回归模型,  $R^2$  是什么?
- 什么样的依据能够说明,在一天中的不同时间,最好的公司也不同? 用 5% 的显著性水平,你能得到什么结论? 构造一幅图,以表明这样的性质,即邮递公司之间的任何差异随送出包裹的时间不同而不同。
- 有无一家公司看上去比其他两家好些? 你如何用置信区间来比较最好的和次好的公司?

General Linear Models Procedure for Exercise 15.44

Dependent Variable: SPEED



| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 8  | 154.37037      | 19.29630    | 6.06    | 0.0001 |
| Error           | 45 | 143.33333      | 3.18519     |         |        |
| Corrected Total | 53 | 297.70370      |             |         |        |

| R-Square | C. V.    | Root MSE | Y Mean |
|----------|----------|----------|--------|
| 0.518537 | 19.66822 | 1.7847   | 9.0741 |

| Source      | DF | Type III SS | Mean Square | F Value | Pr > F |
|-------------|----|-------------|-------------|---------|--------|
| TIME        | 2  | 38.259259   | 19.129630   | 6.01    | 0.0049 |
| FIRM        | 2  | 50.037037   | 25.018519   | 7.85    | 0.0012 |
| TIME * FIRM | 4  | 66.074074   | 16.518519   | 5.19    | 0.0016 |

MEAN SPEED FOR TREATMENTS AND EACH FACTOR

| TIME | FIRM | MEAN  | TIME | MEAN   | FIRM | MEAN  |
|------|------|-------|------|--------|------|-------|
| A    | 1    | 9.33  | A    | 9.167  | 1    | 9.611 |
| A    | 2    | 11.00 | M    | 8.000  | 2    | 9.888 |
| A    | 3    | 7.16  | N    | 10.056 | 3    | 7.722 |
| M    | 1    | 7.83  |      |        |      |       |
| M    | 2    | 10.00 |      |        |      |       |
| M    | 3    | 6.16  |      |        |      |       |
| N    | 1    | 11.66 |      |        |      |       |
| N    | 2    | 8.66  |      |        |      |       |
| N    | 3    | 9.83  |      |        |      |       |

**15.45(工程)** 某纺织公司正在考虑一种合成纤维的三种染色方法。他们希望知道这三种染色方法在质量上是否确有差别。为了帮助判断,该公司进行了一项试验。把五种布料各剪为三块,把这三块随机地指定给这三种染色方法,每种染色方法一块。然后,给这些布块分级评分,以表明染色的质量。结果如下。

| 染色 | 布料品种 |    |    |    |    |
|----|------|----|----|----|----|
|    | 1    | 2  | 3  | 4  | 5  |
| A  | 74   | 78 | 76 | 82 | 77 |
| B  | 81   | 86 | 90 | 93 | 73 |
| C  | 95   | 99 | 90 | 87 | 93 |

- a. 指明这个设计的类型。
- b. 进行方差分析,并得出关于染色方法的结论,取  $\alpha = 0.05$ 。
- c. 给出该设计相对于不按布料品种进行区组划分的设计的效率。

## General Linear Models Procedure for Exercise 15.45

Dependent Variable: QUALITY

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 6  | 688.00000      | 114.66667   | 3.34    | 0.0596 |
| Error           | 8  | 274.93333      | 34.36667    |         |        |
| Corrected Total | 14 | 962.93333      |             |         |        |

| R-Square | C. V.    | ROOT MSE | Y Mean |
|----------|----------|----------|--------|
| 0.714484 | 6.902240 | 5.8623   | 84.933 |

| Source   | DF | Type III SS | Mean Square | F Value | Pr > F |
|----------|----|-------------|-------------|---------|--------|
| DYE      | 2  | 593.73333   | 296.86667   | 8.64    | 0.0100 |
| SPECIMEN | 4  | 94.26667    | 23.56667    | 0.69    | 0.6216 |

| DYE | Mean  | SPECIMEN | Mean  |
|-----|-------|----------|-------|
| 1   | 77.40 | 1        | 83.33 |
| 2   | 84.60 | 2        | 87.67 |
| 3   | 92.80 | 3        | 85.33 |
|     |       | 4        | 87.33 |
|     |       | 5        | 81.00 |

**15.46(心理学)** 一个试验测试了工厂中播放的音乐对于工人生产的影响。试验中对四种音乐节目(A, B, C, D)和无音乐(E)进行了比较。每种节目一次播放一整天,重复了五次。因而整个试验持续了5个星期。为了控制各个星期之间,以及一个星期内各日之间的变异,对这个试验的25天采用了拉丁方设计。每种节目在某个星期一播放一次,某个星期二播放一次, ..., 某个星期五播放一次;一个星期中每种节目只播放一次。

- a. 在这五种音乐下,平均生产量看起来有差别吗? 取  $\alpha = 0.05$ 。
- b. 如果工人的产量之间有差别,哪一种音乐节目比起没有音乐来,有更高的产量?

| 星期 | 星期一    | 星期二    | 星期三    | 星期四    | 星期五    |
|----|--------|--------|--------|--------|--------|
| 1  | 133(E) | 139(B) | 140(C) | 140(D) | 145(A) |
| 2  | 139(A) | 136(E) | 141(B) | 143(C) | 146(D) |
| 3  | 138(B) | 139(D) | 140(E) | 139(A) | 142(C) |
| 4  | 137(C) | 140(A) | 136(D) | 129(E) | 132(B) |
| 5  | 142(D) | 143(C) | 142(A) | 144(B) | 132(E) |

General Linear Models Procedure for Exercise 15.45

Dependent Variable: Y OUTPUT

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 12 | 313.12000      | 26.09333    | 2.59    | 0.0561 |
| Error           | 12 | 120.72000      | 10.06000    |         |        |
| Corrected Total | 24 | 433.84000      |             |         |        |

| R-Square | C. V.    | Root MSE | Y Mean |
|----------|----------|----------|--------|
| 0.721741 | 2.280522 | 3.1718   | 139.08 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|--------|----|-------------|-------------|---------|--------|
| WEEK   | 4  | 123.44000   | 30.86000    | 3.07    | 0.0589 |
| DAY    | 4  | 11.84000    | 2.96000     | 0.29    | 0.8761 |
| MUSIC  | 4  | 177.84000   | 44.46000    | 4.42    | 0.0200 |

| WEEK | Mean  | DAY | Mean  | MUSIC | Mean  |
|------|-------|-----|-------|-------|-------|
| 1    | 139.4 | 1   | 137.8 | A     | 141.0 |
| 2    | 141.0 | 2   | 139.4 | B     | 138.8 |
| 3    | 139.6 | 3   | 139.8 | C     | 141.0 |
| 4    | 134.8 | 4   | 139.0 | D     | 140.6 |
| 5    | 140.6 | 5   | 139.4 | E     | 134.0 |

15.47(农业) 下面是5个农场中小麦的产量(单位:磅)。从每个农场按照土壤的肥力挑选五块地,并按土地的肥沃程度从高到低依次编号为1、2、3、4、5。每块地所接受的处理(施肥)由括号中的字母表示。

a. 指出这个设计是什么设计。

b. 进行方差分析,并得出关于五种施肥方法的结论,取  $\alpha = 0.01$ 。

| 农场 | 肥 力     |        |        |        |        |
|----|---------|--------|--------|--------|--------|
|    | 1       | 2      | 3      | 4      | 5      |
| 1  | (D)10.3 | (E)8.6 | (A)6.7 | (C)7.6 | (B)5.8 |
| 2  | (E) 8.8 | (B)6.7 | (C)6.7 | (A)4.8 | (D)6.0 |
| 3  | (A) 6.3 | (C)8.3 | (B)6.8 | (D)8.0 | (E)8.8 |
| 4  | (C) 8.9 | (D)7.4 | (E)8.2 | (B)6.2 | (A)4.4 |
| 5  | (B)7.3  | (A)4.4 | (D)7.7 | (E)6.8 | (C)6.7 |

**15.48** 参见练习 15.47。用多重比较方法对处理均值做两两比较。指出哪个错误的概率受到了控制。

General Linear Models Procedure for Exercises 15.47 and 15.48

Dependent Variable: YIELD

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 12 | 46.067200      | 3.838933    | 9.88    | 0.0002 |
| Error           | 12 | 4.663200       | 0.388600    |         |        |
| Corrected Total | 24 | 50.730400      |             |         |        |

| R-Square | C. V.    | Root MSE | Y Mean |
|----------|----------|----------|--------|
| 0.908079 | 8.745481 | 0.6234   | 7.1280 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|--------|----|-------------|-------------|---------|--------|
| FARM   | 4  | 6.522400    | 1.630600    | 4.20    | 0.0236 |
| PLOT   | 4  | 11.266400   | 2.816600    | 7.25    | 0.0033 |
| FERT   | 4  | 28.278400   | 7.069600    | 18.19   | 0.0001 |

| FARM | Mean | PLOT | Mean | FERTILIZER | Mean |
|------|------|------|------|------------|------|
| 1    | 7.80 | 1    | 8.32 | A          | 5.32 |
| 2    | 6.60 | 2    | 7.08 | B          | 6.56 |
| 3    | 7.64 | 3    | 7.22 | C          | 7.64 |
| 4    | 7.02 | 4    | 6.68 | D          | 7.88 |
| 5    | 6.58 | 5    | 6.34 | E          | 8.24 |

## 第十六章 协方差分析

- 16.1 引言和案例
- 16.2 具有一个协变量的完全随机化设计
- 16.3 外推问题
- 16.4 多维协变量和更复杂的设计
- 16.5 小结

### 16.1 引言和案例

在一些试验中, 试验单元并不是齐性的, 或者在试验条件中存在一些并非由处理决定的变异。例如, 要设计一项研究来评估几种针对八岁儿童的阅读教学方法。响应变量是儿童在参加完这个阅读训练项目后的测试分数。但是, 参加研究的儿童在参加之前阅读能力上也有差异, 而且很多学校之外的因素也会影响儿童的阅读分数, 如与儿童家庭有关的社会经济变量。用来描述试验单元之间或试验条件之间差异的变量称为**协变量**。协方差分析是可以减少协变量对于处理均值影响的一种方法。它可以提高参数估计的精度和假设检验的功效。

在第十五章中, 我们曾利用随机化完全区组和拉丁方设计来解决上述问题。试验单元划分为区组, 从而, 与把所有的试验单元作为一个整体相比, 每个区组内的试验单元具有更好的齐性。这样我们就可以使响应变量的方差中由处理之外的因素造成的那部分变小。

在很多试验中很难或不可能将试验单元分为区组。在进行试验之前可能并不知道哪些特征能用于区分试验单元, 或者影响响应值的变量直到试验开始后才可以看出来。在某些情况下, 可能每个区组内的试验单元很少, 以至不能用于考察所有的处理。包含这些情况的试验可以从以下的例子中看到。

- 要进行一项临床试验来评估治疗慢性疼痛的几种传统方法和一些新的替代方法。试验中病人疼痛的水平与他们有这种症状的时间长短, 他们的年龄, 健康状况, 以及其他一些会影响治疗的因素有关。研究人员可以按这些因素中的若干个分成区组, 但是其他一些协变量可能会对试验结果造成不良影响。
- 有人建议在 Texas 的大牧区内用空中撒播杀虫剂的方法来控制火蚁。将杀虫剂喷洒到牧场上有若干种可能的方法。由于 EPA(美国环保署)关注

喷洒中相对于目标区域的漂移,所以要设计一项研究来衡量喷洒技术的准确性。对于四种撒播杀虫剂的方法,记录其在目标区域内着附的杀虫剂量  $y$ 。试验应只在无风或风很小的日子里进行。然而在 Texas 经常会有阵风影响喷洒的准确性。因此,一个重要的协变量就是在喷洒过程中目标区域上空的风速。

- 光纤电缆的生产商要调查三种用于做光缆涂层的机器。所感兴趣的响应变量是电缆在涂层后的拉伸强度。尽管涂层的厚度是一致的,都是 1.5mm,在 100 米长的电缆上其图层还是会有一些厚度上的不同。这些不同很可能会影响电缆的拉伸强度。检验在恒温的实验室中进行,试验进行了五天。因为天与天之间存在着环境和技术人员的差异,研究者决定按天划分区组,并记录下在电缆断裂处的涂层的厚度。这样,试验中既用到了区组变量又用到了协变量。

在下面案例分析所描述的试验中,其响应变量不仅与指定的处理有关,而且还与研究期间在试验单元上测量得到的协变量有关。

#### 案例:凉爽季节草坪的评价

高尔夫球场的草坪维护者所面临的一个问题是滤过性毒菌疾病(viral diseases)的流行,这种疾病能够破坏草坪。早春时节,天气凉爽潮湿,草坪上的草还未完全从冬眠中恢复,此时这些疾病更为危险。几种用于高尔夫球场的新的草皮品种已经培植出来。这些品种可以抵抗维护者所关心的那种滤过性毒菌疾病。在采用这些新品绿化高尔夫草场之前,有必要对它们对于相应地面的适应性做出评价。

**数据收集的设计** 在设计一个适当的试验评价这些品种时,研究者考虑了下列因素:

1. 评价这些品种时需要使用哪些性能指标?
2. 本国的地理区域特征会否影响草皮品种的性能?
3. 在高尔夫季节中不同的时间内这些品种表现是否有差别?
4. 土壤的哪些因素会影响草皮品种的性能特征?
5. 为获得对草皮品种性能的可靠估计,对每个品种需作多少次重复观测?
6. 在试验期间,哪些环境因素会影响草皮的性能?
7. 哪些统计方法可以用来对这些品种的差异进行有效的评估?

8. 在最终的报告中,需要给出哪些信息来证明这些品种在高尔夫草场绿化适用性上的差别?

在以前的研究中,已发现三种草皮( $C_1$ 、 $C_2$  和  $C_3$ )对早春的滤过性毒菌疾病的抵抗力最强。然后,研究者通过与高尔夫球场的管理者讨论,确定最主要的衡量草皮性能的指标是球被球杆击中后在草坪上的滚动速度。美国高尔夫联合会(UGSA)已经开发出了一种叫做高尔夫球测速仪(Stimpmeter)的装置来测定球在

草上的速度。高尔夫球测速仪是一根长 36 英寸的铝杆,在它的一侧有一条刻好的跑道。跑道上有一个槽口,用来撑住球直到高尔夫球测速仪的一端被抬高大约 20 度角。高尔夫球沿高尔夫球测速仪相反两方向滚动后的平均距离作为在草上的速度。球滚的越远,速度就越快。影响速度的重要因素是草的长度,地面的硬度及地面的坡度。

研究者决定研究考察八个不同的地区。在每个地区,他们选择了一个高尔夫球场并铺了三块草坪。三块草地间具有相同的土壤结构和坡度。每个区域内三种草随机地分到一块草地上。这样与地理位置有关的影响草上的速度的因素通过划分区组得到了控制。一个非常重要而研究者又难以控制的因素是试验期间的湿度。因此就决定记录下湿度,把它作为协变量。在八个地区测得的草上速度和湿度数据如表 16.1。

表 16.1 三种草皮上的球速

| 地区 | $C_1$ |      | $C_2$ |      | $C_3$ |      |
|----|-------|------|-------|------|-------|------|
|    | 湿度    | 速度   | 湿度    | 速度   | 湿度    | 速度   |
| 1  | 31.60 | 7.56 | 29.42 | 8.88 | 89.60 | 8.20 |
| 2  | 54.12 | 7.41 | 44.44 | 8.20 | 37.17 | 9.15 |
| 3  | 42.34 | 7.64 | 84.38 | 7.20 | 37.32 | 9.24 |
| 4  | 53.82 | 6.81 | 88.42 | 7.12 | 89.21 | 8.31 |
| 5  | 86.70 | 6.86 | 71.33 | 8.16 | 58.57 | 9.42 |
| 6  | 76.27 | 6.86 | 45.50 | 8.68 | 66.68 | 9.26 |
| 7  | 68.66 | 7.22 | 66.79 | 8.25 | 82.78 | 8.93 |
| 8  | 47.27 | 7.64 | 58.34 | 8.22 | 29.52 | 9.89 |

试验期间的每块草坪上的速度测量值与相应的湿度的读数由图 16.1 描画出来。画出的数据点表明,对这三种草,在速度和湿度水平间存在相类似的负相关。然而  $C_3$  总体上显得比其他两种具有更快的速度。在本章的后续部分我们将建立模型和分析方法,以使我们能对各个品种的速度读数按照试验球场所在的地理区域位置以及试验期间的湿度数据进行调整。由于协方差分析既有方差分析的特点,也有回归分析的特点,我们将利用一般的线性模型来对此类数据进行分析。通过参考前面章节的知识,并基于这些一般线性模型的理论,我们可以更容易地理解方差分析与回归模型的结合。我们从只含有一个协变量的完全随机化设计开始。

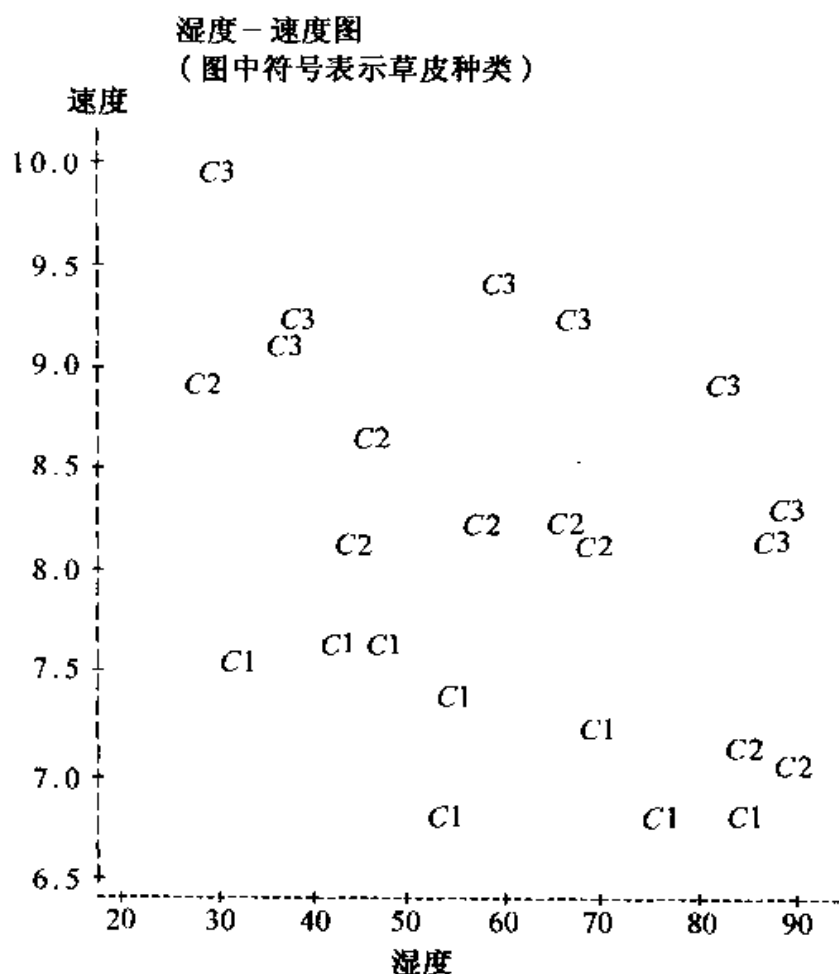


图 16.1 湿度与三种草皮上的高尔夫球速

## 16.2 具有一个协变量的完全随机化设计

完全随机化设计可用于比较  $t$  个总体的均值。为达到此目的,我们必须获得响应变量  $y$  在第  $i$  个总体处的  $n_i$  个观测值的随机样本 ( $i = 1, 2, \dots, t$ )。而现在,除了要测量每个试验单元上响应变量  $y$  的值,我们还要测量另一个变量  $x$ , 经常称之为**协变量**。比如,在研究几种不同的用于提高 8 岁儿童阅读水平的方法的效应时,我们不但要考查每个儿童最终的成绩,还要考查他们在参加研究之前的阅读水平。最后,我们要用  $y$  和  $x$  共同提供的信息来比较这几种不同的方法。

注意,  $x$  可以看为一个自变量,但与前些章里讨论的情况不同,这里我们在观测  $x$  之前无法控制它的取值(像我们控制温度或压力那样)。尽管如此,我们仍可以为完全随机化设计建立一个模型,并将协变量作为一个自变量来处理。



我们将用一个有  $t = 3$  个处理, 一个协变量的完全随机化设计来举例说明协方差分析的步骤。

### 例 16.1

本项研究中, 我们比较两个处理——即缓慢释放的肥料(S)和快速释放的肥料(F), 与对照(C)——一种标准的肥料对于花生米产量(克)的效应。每个处理有 10 次重复, 试验在温室中进行。在开始试验时, 研究者认识到这 30 株花生的健康水平与成长水平并不完全相同。因此, 研究者在试验开始前记录下了每株的高度(厘米)作为衡量植物健康与成长水平的标准, 如下表所示。

| 对照(C) |    | 缓慢释放(S) |    | 快速释放(F) |    |
|-------|----|---------|----|---------|----|
| 产量    | 高度 | 产量      | 高度 | 产量      | 高度 |
| 12.2  | 45 | 16.6    | 63 | 9.5     | 52 |
| 12.4  | 52 | 15.8    | 50 | 9.5     | 54 |
| 11.9  | 42 | 16.5    | 63 | 9.6     | 58 |
| 11.3  | 35 | 15.0    | 33 | 8.8     | 45 |
| 11.8  | 40 | 15.4    | 38 | 9.5     | 57 |
| 12.1  | 48 | 15.6    | 45 | 9.8     | 62 |
| 13.1  | 60 | 15.8    | 50 | 9.1     | 52 |
| 12.7  | 61 | 15.8    | 48 | 10.3    | 67 |
| 12.4  | 50 | 16.0    | 50 | 9.5     | 55 |
| 11.4  | 33 | 15.8    | 49 | 8.5     | 40 |

图 16.2 以协变量植株高度为横轴变量, 给出了每个处理处产量的散点图。

例 16.1 中描述的试验是一个具有 3 个处理, 一个协变量的完全随机化试验。如果我们假设产量  $y_{ij}$  与协变量  $x_{ij}$  之间是直线关系, 则这个只有一个协变量的完全随机化设计的模型为:

$$y_{ij} = \mu_i + \beta_1(x_{ij} - \bar{x}_{..}) + \epsilon_{ij}$$

或

$$y_{ij} = \beta_0 + \alpha_i + \beta_1 x_{ij} + \epsilon_{ij}$$

其中  $i = 1, 2, \dots, t, j = 1, 2, \dots, n$ ,  $\mu_i$  是第  $i$  个处理的均值,  $\beta_1$  是  $y_{ij}$  对  $x_{ij}$  回归的斜率,  $\beta_0$  是  $y_{ij}$  对  $x_{ij}$  回归的截距,  $\alpha_i$  是第  $i$  个处理的效应,  $\epsilon_{ij}$  是试验误差, 为服从均值为 0, 方差为  $\sigma_\epsilon^2$  的正态分布的随机变量。协方差分析中加之于模型上的主要条件还有:

1. 响应变量  $y$  与协变量  $x$  的关系是线性的。

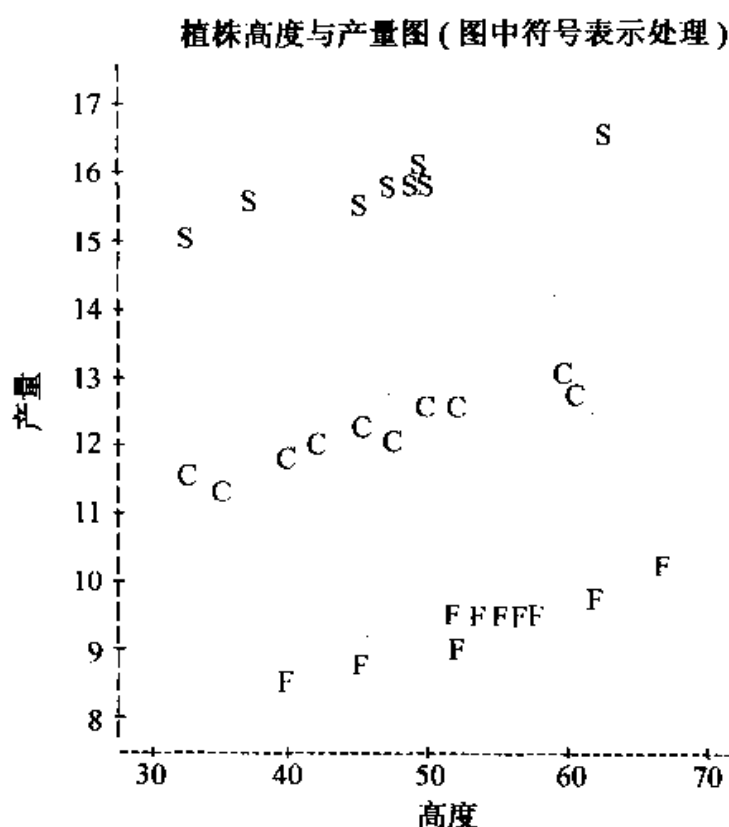


图 16.2 协变量植株高度与三个处理下的花生米产量

2. 对于所有处理, 回归的系数  $\beta_1$  相同。

3. 处理不影响协变量  $x_{ij}$ 。

协方差分析要对响应变量  $y$  拟合许多模型。首先, 我们需要确定协变量  $x$  是否使试验误差有很显著的减少 (这里意指模型中协变量的引入使得模型中的误差部分有显著的减少, 也就是说, 协变量对于响应值的变异有显著的影响——译者注)。如果减少是显著的, 那我们就将观测到的处理均值  $\bar{y}_i$  用调整后的处理均值  $\hat{\mu}_{Adj, i}$  来替代, 其中  $\hat{\mu}_{Adj, i}$  已根据协变量对响应变量的效应作了调整。这样, 处理差异的显著性就取决于调整后的均值而不是观察到的均值。

现在我们把协方差分析所需要的模型用公式表达出来。 $y_{ij}$  与  $t$  个处理及协变量之间关系的模型可写为方差分析模型的形式, 并可进一步表达为回归的形式。

全模型:  $y_{ij} = \beta_0 + \alpha_i + \beta_1 x_{ij} + \epsilon_{ij}$

其次, 我们给出两个简略模型的公式, 一个不包括协变量, 另一个忽略处理之间的差异但包括协变量。

简略模型 I:  $y_{ij} = \beta_0 + \alpha_i + \epsilon_{ij}$

简略模型 II:  $y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$

这三个模型也可写为第十二章的回归 (一般线性) 模型的形式。我们之所以将

之转换为回归模型的形式,是为了使之更适于用各种统计软件包作分析。

$$\text{全模型: } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_t x_t + \epsilon$$

其中

$x_1 =$  协变量

$x_2 = 1$  若处理 2 被使用       $x_2 = 0$  否则

$x_3 = 1$  若处理 3 被使用       $x_3 = 0$  否则

.....

$x_t = 1$  若处理  $t$  被使用       $x_t = 0$  否则

用上述模型做出期望值的表是很有用的,表 16.2 是基于全模型给出的。注意各处理具有相同的斜率( $\beta_1$ ),但截距是不同的。

表 16.2 全模型中的期望值

| 处理       | 期望值                                 |
|----------|-------------------------------------|
| 1        | $\beta_0 + \beta_1 x_1$             |
| 2        | $(\beta_0 + \beta_2) + \beta_1 x_1$ |
| $\vdots$ | $\vdots$                            |
| $t$      | $(\beta_0 + \beta_t) + \beta_1 x_1$ |

为了确定协变量的作用,接下来我们来拟合不包括协变量的简略模型。

$$\text{简略模型 I: } y = \beta_0 + \beta_2 x_2 + \cdots + \beta_t x_t + \epsilon$$

然后,拟合第二个简略模型,其中忽略了处理的差异而保留了协变量。

$$\text{简略模型 II: } y = \beta_0 + \beta_1 x_1 + \epsilon$$

对每个模型,我们可算得它的误差平方和,并以下列符号表示:

$SSE_F$  = 全模型的误差平方和

$SSE_{RI}$  = 简略模型 I 的误差平方和

$SSE_{RII}$  = 简略模型 II 的误差平方和

协变量对于响应变量作用的显著性,可通过检验假设“处理的回归直线斜率为 0”来确定。对于全模型,此假设为:

$$H_0: \beta_1 = 0 \quad H_a: \beta_1 \neq 0$$

检验统计量基于因模型中引入协变量而带来的平方和的减少,表示为:

$$SS_{Cov} = SSE_{RI} - SSE_F$$

由此,我们得到  $F$  检验统计量为:

$$F = \frac{SS_{Cov}}{SSE_F / (N - t - 1)}$$

其中  $N$  是试验中的总观测次数。判决规则由下式给出:

拒绝  $H_0: \beta_1 = 0$  当  $F \geq F_{\alpha, 1, N-t-1}$

如果我们断定协变量与响应变量间有显著的线性关系, 我们下一步就要用调整后的处理均值来检验处理的显著效应。也就是说, 我们要检验如下假设:

$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_t = 0$      $H_a$ : 并非所有的  $\alpha_i$  均为 0

这等同于在回归模型中检验所有的回归直线具有相同的截距( $\beta_0$ )。从表 16.2 可知, 我们是要检验:

$H_0: \beta_2 = \beta_3 = \cdots = \beta_t = 0$      $H_a$ : 并非所有的  $\beta_2, \beta_3, \cdots, \beta_t$  均为 0

检验统计量基于因模型中引入处理均值的差而带来的平方和的减少, 表示为:

$$SS_{T\pi} = SSE_{R\pi} - SSE_F$$

$F$  检验统计量为:

$$F = \frac{SS_{T\pi}/(t-1)}{SSE_F/(N-t-1)}$$

判决规则由下式给出:

拒绝  $H_0: \beta_2 = \beta_3 = \cdots = \beta_t = 0$  当  $F \geq F_{\alpha, t-1, N-t-1}$  时

如果拒绝了  $H_0$ , 我们可以用由如下公式估计的调整后的处理均值来衡量处理间的差异:

$$\hat{\mu}_{Adj, i} = \bar{y}_{i.} - \hat{\beta}_1(\bar{x}_{i.} - \bar{x}_{..})$$

这个值是通过观测得到的处理均值根据协变量的效应进行调整后得到的, 其中协变量的效应又是通过考虑第  $i$  个处理的各试验单元协变量值的平均值与所有处理的平均值的差的大小来估计的。

我们也可以用回归模型来估计调整后的处理均值。从表 16.2, 对处理  $i = 2, 3, \cdots, t$  我们有

$$\mu_i = E(y) = \beta_0 + \beta_i + \beta_1 x_1$$

对  $i = 1$ ,

$$\mu_1 = E(y) = \beta_0 + \beta_1 x_1$$

调整后的处理均值是通过估计在每个处理上相应于协变量的总平均值  $x_1 = \bar{x}_{..}$  的  $y$  的均值而得到的。对处理  $i = 2, 3, \cdots, t$ , 有:

$$\hat{\mu}_{Adj, i} = \hat{\beta}_0 + \hat{\beta}_i + \hat{\beta}_1 \bar{x}_{..}$$

对处理 1 有

$$\hat{\mu}_{Adj, 1} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_{..}$$

第  $i$  个处理均值的估计值  $\hat{\mu}_{Adj, i}$  的标准误差的估计为:

$$SE(\hat{\mu}_{Adj, i}) = \sqrt{MSE_F \left( \frac{1}{n} + \frac{(\bar{x}_{i.} - \bar{x}_{..})^2}{E_{xx}} \right)}$$

其中  $E_{xx} = \sum \sum_{ij} (x_{ij} - \bar{x}_{i.})^2$ 。两个调整后的处理均值差  $\hat{\mu}_{\text{Adj}, i} - \hat{\mu}_{\text{Adj}, h}$  的标准误差的估计为:

$$SE(\hat{\mu}_{\text{Adj}, i} - \hat{\mu}_{\text{Adj}, h}) = \sqrt{MSE_F \left( \frac{2}{n} + \frac{(\bar{x}_{i.} - \bar{x}_{h.})^2}{E_{xx}} \right)}$$

其中  $MSE_F$  是全模型的均方。这些估计的标准误差可用于确定调整后的处理均值和它们的差的置信区间。

我们用下例来说明协方差分析的思想。

### 例 16.2

在例 16.1 中, 我们三个处理——对照肥料(C), 缓慢释放的肥料(S)和快速释放的肥料(F), 并且我们用研究开始时的植株高度作为协变量; 响应变量是花生米产量, 有十次重复。

- 写出协方差分析的模型。
  - 利用下面给出的计算机输出结果检验协变量的引入是否使试验误差有显著的减少。
  - 给出三个处理上的线性回归方程。
  - 分别计算三个处理上的观测和调整后的处理均值。
  - 在根据协变量调整后, 三个处理之间有显著的差异吗?
- 计算机输出结果如下。

FULL MODEL

General Linear Models Procedure

Dependent Variable: Y YIELD

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 3  | 214.37595      | 71.45865    | 4447.85 | 0.0001 |
| Error           | 26 | 0.41771        | 0.01607     |         |        |
| Corrected Total | 29 | 214.79367      |             |         |        |

|           |              | T for H0: | Pr >  T | Std Error of |
|-----------|--------------|-----------|---------|--------------|
| INTERCEPT | 9.529256364  | 71.34     | 0.0001  | 0.13357349   |
| X1(Cov)   | 0.055809949  | 20.41     | 0.0001  | 0.00273429   |
| X2(S)     | 3.571637117  | 62.62     | 0.0001  | 0.05703267   |
| X3(F)     | -3.144155615 | -52.08    | 0.0001  | 0.06037390   |

## REDUCED MODEL I

## General Linear Models Procedure

Dependent Variable: Y YIELD

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 2  | 207.68267      | 103.84133   | 394.28  | 0.0001 |
| Error           | 27 | 7.11100        | 0.26337     |         |        |
| Corrected Total | 29 | 214.79367      |             |         |        |

| Parameter | Estimate    | T for H0:<br>Parameter = 0 | Pr >  T | Std Error of<br>Estimate |
|-----------|-------------|----------------------------|---------|--------------------------|
| INTERCEPT | 12.13000000 | 74.74                      | 0.0001  | 0.16228690               |
| X2(S)     | 3.70000000  | 16.12                      | 0.0001  | 0.22950833               |
| X3(F)     | -2.72000000 | -11.85                     | 0.0001  | 0.22950833               |

## REDUCED MODEL II

## General Linear Models Procedure

Dependent Variable: Y YIELD

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 1  | 0.4721494      | 0.4721494   | 0.06    | 0.8057 |
| Error           | 28 | 214.3215172    | 7.6543399   |         |        |
| Corrected Total | 29 | 214.7936667    |             |         |        |

| Parameter | Estimate    | T for H0:<br>Parameter = 0 | Pr >  T | Std Error of<br>Estimate |
|-----------|-------------|----------------------------|---------|--------------------------|
| INTERCEPT | 13.14900450 | 4.64                       | 0.0001  | 2.83300563               |
| X1(Cov)   | -0.01387451 | -0.25                      | 0.8057  | 0.05586395               |

## 解答

- a. 这是一个完全随机化设计, 有三个处理, 每个处理处有 10 次重复, 有一个协变量。因此模型为  $y_{ij} = \mu_i + \beta_1(x_{ij} - \bar{x}_{..}) + \epsilon_{ij}$ ,  $i = 1, 2, 3, j = 1, \dots, 10$ . 全模型用回归的记号表示如下。

全模型(各回归直线有相同的斜率和不同的截距):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

其中  $y$  = 产量

$x_1$  = 植株高度

$x_2 = 1$  若处理为 S

$x_2 = 0$  否则

$x_3 = 1$  若处理为 F       $x_3 = 0$  否则

与三个处理相应的期望的值为:

| 处理 | 期望的值                                |
|----|-------------------------------------|
| C  | $\beta_0 + \beta_1 x_1$             |
| S  | $(\beta_0 + \beta_2) + \beta_1 x_1$ |
| F  | $(\beta_0 + \beta_3) + \beta_1 x_1$ |

相应的简略模型为:

简略模型 I (回归直线的斜率为 0; 即, 响应变量与协变量无关):

$$y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

和

简略模型 II (回归直线有相同的截距  $\beta_0$  和相同的斜率  $\beta_1$ )

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

- b. 我们要检验协变量是否会使试验误差减少; 即, 我们需要检验共同斜率为 0。

$$H_0: \beta_1 = 0 \quad H_a: \beta_1 \neq 0$$

由计算机的输出结果,

$$SSE_F = 0.41771 \quad SSE_{R1} = 7.11100$$

这样, 我们有

$$SS_{Cov} = SSE_{R1} - SSE_F = 7.111 - 0.41771 = 6.69329$$

F 检验为

$$F = \frac{6.69329}{0.41771/(30-3-1)} = 416.62 \text{ 和 } F_{0.05, 1, 26} = 4.23$$

因为 416.62 大于 4.23, 我们拒绝  $H_0$  并给出结论: 植物种子的产量与植株高度(协变量)显著相关(即斜率  $\beta_1$  不为 0)。

- c. 由全模型的输出结果我们可以得到最小二乘估计为:

$$\hat{\beta}_0 = 9.53 \quad \hat{\beta}_1 = 0.0558 \quad \hat{\beta}_2 = 3.57 \quad \hat{\beta}_3 = -3.14$$

三种处理下根据初始的植株高度调整后的种子产量的估计值为:

$$\text{对照: } \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 = 9.53 + 0.0558 x_1$$

$$\begin{aligned} \text{缓慢释放: } \hat{y} &= (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 x_1 = (9.53 + 3.57) + 0.0558 x_1 \\ &= 13.1 + 0.0558 x_1 \end{aligned}$$

$$\begin{aligned} \text{快速释放: } \hat{y} &= (\hat{\beta}_0 + \hat{\beta}_3) + \hat{\beta}_1 x_1 = (9.53 - 3.14) + 0.0558 x_1 \\ &= 6.39 + 0.0558 x_1 \end{aligned}$$

- d. 观测到的样本均值如下。

|     | 对照肥料  | 缓慢释放肥料 | 快速释放肥料 | 总平均    |
|-----|-------|--------|--------|--------|
| $y$ | 12.13 | 15.83  | 9.41   | 12.457 |
| $x$ | 46.60 | 48.90  | 54.20  | 49.900 |

在每个回归方程中,我们可以用植株高度的总平均值代替  $x_1$  来估计调整后的均值

$$\text{对照: } \hat{\mu}_{\text{Adj},1} = 9.53 + 0.0558(49.90) = 12.31$$

$$\text{缓慢释放: } \hat{\mu}_{\text{Adj},2} = 13.1 + 0.0558(49.90) = 15.88$$

$$\text{快速释放: } \hat{\mu}_{\text{Adj},3} = 6.39 + 0.0558(49.90) = 9.17$$

或者我们可以用下面的公式获得上述估计:

$$\hat{\mu}_{\text{Adj},i} = \bar{y}_{i\cdot} - \hat{\beta}_1(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})$$

$$\text{对照: } \hat{\mu}_{\text{Adj},1} = 12.13 - 0.0558(46.60 - 49.90) = 12.31$$

$$\text{缓慢释放: } \hat{\mu}_{\text{Adj},2} = 15.83 - 0.0558(48.9 - 49.90) = 15.88$$

$$\text{快速释放: } \hat{\mu}_{\text{Adj},3} = 9.41 - 0.0558(54.20 - 49.90) = 9.17$$

因为用缓慢释放肥料的植物的平均植株高度低于总平均高度,所以观测的平均种子产量 15.83 调整为 15.88,而快速释放肥料的产量则由 9.41 减少为 9.17。

- e. 我们可以通过简略模型 II 的误差平方和来检验三种处理下平均产量间的差异。我们需检验如下假设:

$$H_0: \mu_{\text{Adj},1} = \mu_{\text{Adj},2} = \cdots = \mu_{\text{Adj},t} \quad H_a: \text{并非所有 } \mu'_{\text{Adj},i} \text{ 都相等}$$

这相当于检验回归直线具有相同的截距这样一个零假设,即检验:

$$H_0: \beta_2 = \beta_3 = 0 \quad H_a: \beta_2 \neq 0 \text{ 或 } \beta_3 \neq 0$$

由计算机的输出结果,

$$\text{SSE}_F = 0.41771 \quad \text{SSE}_{\text{RII}} = 214.3215$$

所以我们有

$$\text{SS}_{\text{TII}} = \text{SSE}_{\text{RII}} - \text{SSE}_F = 214.3215 - 0.41771 = 213.90$$

F 检验为

$$F = \frac{213.90/(3-1)}{0.41771/(30-3-1)} = 6,657.13 \quad \text{及} \quad F_{0.05,2,26} = 3.37$$

因为 6,657.13 大于 3.37,我们拒绝  $H_0$ ,认为各截距不等,因此对于三种肥料,调整后的产量间有显著差异。

例 16.2 中得到的结论依赖于模型条件是否满足。我们可以通过考察模型拟合的残差



$$e_{ij} = y_{ij} - \hat{\beta}_0 - \hat{\beta}_1 x_{1ij} - \hat{\beta}_2 x_{2ij} - \cdots - \hat{\beta}_t x_{tij}$$

来检查误差项独立且服从相同的正态分布的条件是否成立。我们可以用图和正态性检验的方法检验  $e_{ij}$  的等方差和正态性条件是否满足。

我们将通过下面的方式检验协方差分析的三个另外的条件。

**响应变量与协变量之间的关系是线性的**我们可以像在回归分析中用图和假设检验的方法那样检验这个条件。我们可以分别对每个处理以  $x, y$  为坐标画图, 判定画出的点是否呈直线。用第十一章的方法可以对每个处理分别拟合一回归直线。然后, 我们可以考察这  $t$  条拟合直线的残差, 进行拟合不足的检验, 以决定拟合时是否需要协变量  $x_{ij}$  的更高阶的项。

**对于所有  $t$  个处理, 回归系数(斜率)是相同的**考虑如下模型:

模型 A:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_t x_t \\ + \beta_{t+1} x_1 x_2 + \beta_{t+2} x_1 x_3 + \cdots + \beta_{2t-1} x_1 x_t + \epsilon$$

其中  $x_2, \dots, x_t$  表示处理的指示变量,  $x_1$  表示协变量。这个回归模型对于每个处理会产生不同的回归直线, 这些直线可能具有不同的斜率和不同的截距。(模型 A 的期望的值见表 16.3)

表 16.3 模型 A 的期望值

| 处理       | 期望值  |
|----------|--|
| 1        | $\beta_0 + \beta_1 x_1$                              |
| 2        | $(\beta_0 + \beta_2) + (\beta_1 + \beta_{t+1}) x_1$  |
| 3        | $(\beta_0 + \beta_3) + (\beta_1 + \beta_{t+2}) x_1$  |
| $\vdots$ | $\vdots$   |
| $t$      | $(\beta_0 + \beta_t) + (\beta_1 + \beta_{2t-1}) x_1$ |

下面我们考虑一个简略模型, 其中对于各个处理, 斜率是一致的, 但允许截距之间有差别:

$$\text{模型 B: } y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_t x_t + \epsilon$$

斜率相等的检验, 包含于下列假设的检验之中:

$$H_0: \beta_{t+1} = \beta_{t+2} = \cdots = \beta_{2t-1} = 0$$

$$H_a: \beta_{t+1}, \beta_{t+2}, \dots, \beta_{2t-1} \text{ 不全为 } 0$$

通过拟合模型 A 和 B, 得到下述检验统计量:

$$F = \frac{(\text{SSE}_B - \text{SSE}_A)/(t-1)}{\text{SSE}_A/(N-2t)} \quad \text{自由度 } df_1 = t-1, df_2 = N-2t$$

由此可以确定联系响应变量与协变量的一组回归直线是否有相同的斜率。这是一个关键的问题,因为如果各条直线的斜率是不同的,那么,调整后的均值将高度依赖于调整时所选用的协变量的水平。这种情况类似于在两个因子的试验中,当交互效应显著时,关于一个因子的推断依赖于第二个因子。图 16.3 中显示了联系响应变量与协变量的回归直线具有不同斜率的情形。由该图可以看出,调整量的大小随处理和协变量的水平不同而有很大差异。

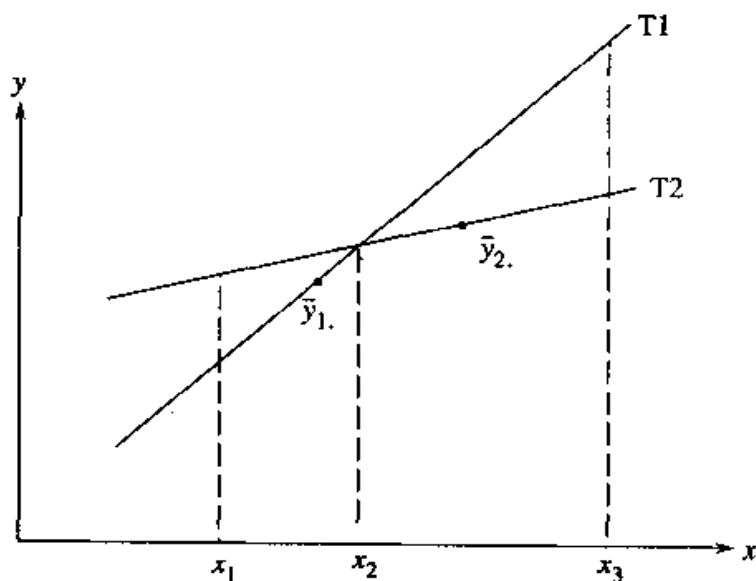


图 16.3 具有不同斜率的联系响应变量与协变量的回归直线

**处理不影响协变量**  $x_{ij}$  当试验中的协变量  $x$  与响应变量  $y$  都受到处理的影响时,我们就不能有效地运用协方差分析。此时适当的分析方法当属多元分析,即将响应变量看为二元变量  $(x, y)$  处理。如果在把处理随机分配到试验单元之前,已经对协变量进行了测量,这时就可以使用协方差分析的模型,因为此时处理就不会影响协变量。如果协变量是对试验中某些条件的测量——即,协变量需要在试验进行过程中测量——那么试验者就必须确定处理是否会影响协变量。只有在试验者确定处理不会影响协变量后,我们才能正确地根据协变量调整处理的均值。

### 例 16.3

见例 16.1。利用下面的计算机输出结果评价协方差分析模型中的必要条件是否得到满足。

MODEL A; DIFFERENT SLOPES FOR EACH TREATMENT

General Linear Models Procedure.

Number of observations in data set = 30

Dependent Variable: Y YIELD

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 5  | 214.43722      | 42.88744    | 2887.70 | 0.0001 |
| Error           | 24 | 0.35644        | 0.01485     |         |        |
| Corrected Total | 29 | 214.79367      |             |         |        |

| Source  | DF | Type III SS | Mean Square | F Value | Pr > F |
|---------|----|-------------|-------------|---------|--------|
| X1      | 1  | 2.6167178   | 2.6167178   | 176.19  | 0.0001 |
| X2      | 1  | 2.5905994   | 2.5905994   | 174.43  | 0.0001 |
| X3      | 1  | 1.4990044   | 1.4990044   | 100.93  | 0.0001 |
| X2 * X1 | 1  | 0.0190292   | 0.0190292   | 1.28    | 0.2688 |
| X3 * X1 | 1  | 0.0151538   | 0.0151538   | 1.02    | 0.3225 |

| Parameter | Estimate     | T for H0:<br>Parameter = 0 | Pr >  T | Std Error of<br>Estimate |
|-----------|--------------|----------------------------|---------|--------------------------|
| INTERCEPT | 9.491768741  | 46.88                      | 0.0001  | 0.20245904               |
| X1        | 0.056614405  | 13.27                      | 0.0001  | 0.00426518               |
| X2        | 3.906558043  | 13.21                      | 0.0001  | 0.29578964               |
| X3        | -3.519620102 | -10.05                     | 0.0001  | 0.35033468               |
| X2 * X1   | -0.006886936 | -1.13                      | 0.2688  | 0.00608421               |
| X3 * X1   | 0.006814587  | 1.01                       | 0.3225  | 0.00674632               |

MODEL B; SAME SLOPE FOR ALL TREATMENTS

General Linear Models Procedure

Number of observations in data set = 30

Dependent Variable: Y YIELD

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 3  | 214.37595      | 71.45865    | 4447.85 | 0.0001 |
| Error           | 26 | 0.41771        | 0.01607     |         |        |
| Corrected Total | 29 | 214.79367      |             |         |        |
| Source          | DF | Type III SS    | Mean Square | F Value | Pr > F |
| X1              | 1  | 6.693287       | 6.693287    | 416.62  | 0.0001 |

|    |   |           |           |         |        |
|----|---|-----------|-----------|---------|--------|
| X2 | 1 | 63.007424 | 63.007424 | 3921.82 | 0.0001 |
| X3 | 1 | 43.572654 | 43.572654 | 2712.12 | 0.0001 |

| Parameter | Estimate     | T for H <sub>0</sub> :<br>Parameter = 0 | Pr >  T | Std Error of<br>Estimate |
|-----------|--------------|---|---------|--------------------------|
| INTERCEPT | 9.529256364  | 71.34                                   | 0.0001  | 0.13357349               |
| X1        | 0.055809949  | 20.41                                   | 0.0001  | 0.00273429               |
| X2        | 3.571637117  | 62.62                                   | 0.0001  | 0.05703267               |
| X3        | -3.144155615 | -52.08                                  | 0.0001  | 0.06037390               |

**解答** 由图 16.2 我们可以看到,以植株高度为横轴,种子产量为纵轴得到的三个处理的曲线都可以很好地用直线来拟合,且它们的斜率大体相同;也就是说,我们得到了截距不同的三条平行直线。前面的计算机输出结果是通过用模型 A (不同斜率不同截距)和模型 B(相同斜率不同截距)对种子产量的数据拟合后得到的。

从输出结果可以计算得

$$F = \frac{(SSE_B - SSE_A)/(t - 1)}{SSE_A/(N - 2t)} = \frac{(0.41771 - 0.35644)/(3 - 1)}{0.35644/(30 - 6)} = 2.06$$

其中  $df_1 = 2$ ,  $df_2 = 24$ 。因为  $F_{0.05, 2, 24} = 3.40$ , 我们不能拒绝  $H_0$ , 由此, 我们得出结论: 三条直线的斜率没有显著不同。因为协变量植株高度是在将化肥种类分配给植物之前测量的, 所以处理不会对协变量有影响。等方差和正态性等条件可以通过残差分析来评定。

## 练习

### 基本技能

**16.1** 考虑一个完全随机化设计,  $t = 5$  个处理, 一个协变量  $x_1$ , 每个处理处有六个观测值。假设每个处理处响应变量  $y$  与协变量  $x_1$  间有线性关系, 写出完全的一般线性模型。给出模型中各参数的意义。

**16.2** 根据练习 16.1, 给出如下情况下模型中参数的关系, 并对每种情况作出示意图。

- 直线不平行。
- 直线平行但不重合。
- 直线重合。

**16.3** 在练习 16.1 中, 对于这五个处理, 你将如何检验这些直线是否平行? 你是如何获得检验统计量的? 检验统计量的自由度是多少?

**16.4** 在练习 16.1 中, 假设直线是平行的. 给出调整后处理均值的检验. 对  $x_1 = 5$ , 你是如何估计处理 1 的响应变量的均值的?

**16.5** 对下面的试验做方差分析. 一名研究者想衡量三种不同加工方法下硅片表面涂膜厚度平均值的差异. 给六块硅片随机地分配加工方法. 对每块硅片都记录下涂膜厚度( $Y$ )和加工过程中试验室的温度( $X$ ). 研究者注意到了温度的波动可能会对涂膜厚度有影响. 检验三种加工方法下的涂膜厚度的均值是否有差异.

| 加工过程 |     |     |     |     |     |
|------|-----|-----|-----|-----|-----|
| 1    |     | 2   |     | 3   |     |
| $x$  | $y$ | $x$ | $y$ | $x$ | $y$ |
| 26   | 100 | 24  | 118 | 37  | 124 |
| 35   | 150 | 28  | 134 | 31  | 95  |
| 28   | 106 | 29  | 138 | 34  | 120 |
| 31   | 95  | 32  | 147 | 27  | 86  |
| 29   | 113 | 36  | 165 | 28  | 98  |
| 34   | 144 | 35  | 159 | 25  | 81  |

### 16.3 外推问题

在上一节中, 我们讨论了在只有一个协变量时如何比较完全随机化设计中的两个(或多个)处理. 我们曾经谈到, 如果对于各个处理, 回归方程关于协变量均是线性的, 并且各条直线相互平行, 就可以用调整后的处理均值来比较各处理. 然而, 像大多数方法一样, 我们不能盲目地使用协方差分析方法. 即使线性和平行的假设都能满足, 如果各处理下协变量的取值范围没有充分的叠合, 我们也会遇到问题. 我们将以下例来说明这一点.

假设我们要比较酒精嗜好者和吸毒者的自尊分数. 我们收集了一个包括九个酒精嗜好者和九个吸毒者的样本, 对每个人, 我们记录下他(她)的自尊分数和年龄. 数据见表 16.4.

如果我们不看数据就盲目地进行协方差分析, 我们可以发现对酒精嗜好者和吸毒者, 回归方程都呈线性且平行. 从图 16.4 的计算机输出结果来看, 酒精嗜好者(A)的数据对应的点近似呈直线, 同样, 吸毒者(D)的数据也如此. 根据两个模型的误差平方和, 可算得

表 16.4 酒精嗜好者和吸毒者样本的自尊分数

| 酒精嗜好者 |    | 吸毒者 |    |
|-------|----|-----|----|
| 自尊    | 年龄 | 自尊  | 年龄 |
| 25    | 15 | 20  | 30 |
| 22    | 17 | 17  | 31 |
| 24    | 18 | 18  | 33 |
| 20    | 19 | 15  | 35 |
| 21    | 21 | 14  | 36 |
| 17    | 22 | 15  | 37 |
| 14    | 23 | 12  | 38 |
| 16    | 24 | 10  | 40 |
| 15    | 25 | 11  | 41 |

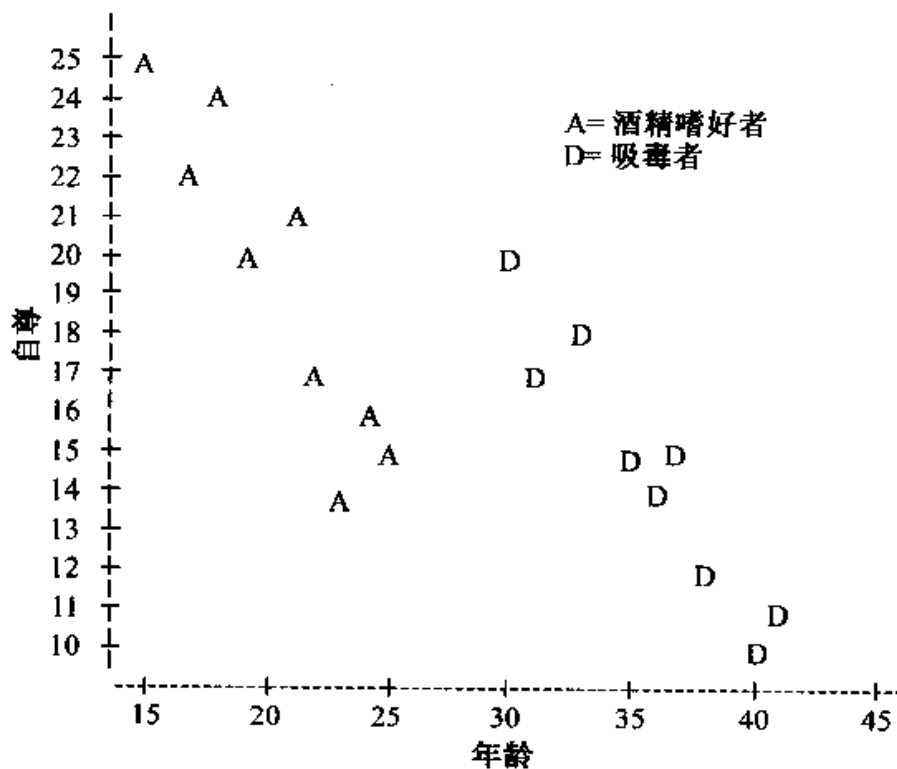


图 16.4 年龄-自尊分数图

$$F = \frac{(30.88 - 27.39)/(2 - 1)}{1.9567} = 1.78$$

其中  $df_1 = 1, df_2 = 14$ 。观测到的  $F$  值的  $p$  值为  $\Pr(F \geq 1.78) = 0.2035$ 。这样我们将接受原假设, 认为酒精嗜好者和吸毒者的自尊分数关于年龄的回归直线斜率相同。此外, 由模型 B 的输出结果, 得到关于调整后自尊分数的均值是否有差异的检验的  $p$  值为  $\Pr(F \geq 34.14) < 0.0001$ 。这意味着两种嗜好者调整后的自尊的分数有显著差异。

#### MODEL A: DIFFERENT SLOPES AND TREATMENT DIFFERENCES

Dependent Variable: Y SELF-ESTEEM

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 3  | 286.60611      | 95.53537    | 48.82   | 0.0001 |
| Error           | 14 | 27.39389       | 1.95671     |         |        |
| Corrected Total | 17 | 314.00000      |             |         |        |

| Source  | DF | Type III SS | Mean Square | F Value | Pr > F |
|---------|----|-------------|-------------|---------|--------|
| X1      | 1  | 188.51593   | 188.51593   | 96.34   | 0.0001 |
| X2      | 1  | 0.43265     | 0.43265     | 0.22    | 0.6454 |
| X2 * X1 | 1  | 3.48284     | 3.48284     | 1.78    | 0.2035 |

| Parameter | Estimate      | T for H0:<br>Parameter = 0 | Pr >  T | Std Error of<br>Estimate |
|-----------|---------------|----------------------------|---------|--------------------------|
| INTERCEPT | 44.18390805 B | 9.49                       | 0.0001  | 4.65570471               |
| X1        | -0.82758621 B | -6.37                      | 0.0001  | 0.12987748               |
| X2        | -2.60800443 B | -0.47                      | 0.6454  | 5.54628759               |
| X2 * X1   | -0.26036560 B | -1.33                      | 0.2035  | 0.19515497               |

#### MODEL B: SAME SLOPES AND TREATMENT DIFFERENCES

Dependent Variable: Y SELF-ESTEEM

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 2  | 283.12327      | 141.56163   | 68.77   | 0.0001 |
| Error           | 15 | 30.87673       | 2.05845     |         |        |
| Corrected Total | 17 | 314.00000      |             |         |        |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|--------|----|-------------|-------------|---------|--------|
| X1     | 1  | 185.12327   | 185.12327   | 89.93   | 0.0001 |

|    |   |          |          |       |        |
|----|---|----------|----------|-------|--------|
| X2 | 1 | 70.27928 | 70.27928 | 34.14 | 0.0001 |
|----|---|----------|----------|-------|--------|

| Parameter | Estimate      | T for H0:<br>Parameter = 0 | Pr >  T | Std Error of<br>Estimate |
|-----------|---------------|----------------------------|---------|--------------------------|
| INTERCEPT | 48.29686944 B | 13.50                      | 0.0001  | 3.57834982               |
| X1        | -0.94290288   | -9.48                      | 0.0001  | 0.09942750               |
| X2        | -9.68641053 B | -5.84                      | 0.0001  | 1.65775088               |

## REDUCED MODEL I: TREATMENT DIFFERENCES WITH NO COVARIATE

## General Linear Models Procedure

Dependent Variable: Y SELF-ESTEEM

| Source          | DF | Sum of<br>Squares | Mean<br>Square | F Value | Pr > F |
|-----------------|----|-------------------|----------------|---------|--------|
| Model           | 1  | 98.000000         | 98.000000      | 7.26    | 0.0160 |
| Error           | 16 | 216.000000        | 13.500000      |         |        |
| Corrected Total | 17 | 314.000000        |                |         |        |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|--------|----|-------------|-------------|---------|--------|
| X2     | 1  | 98.000000   | 98.000000   | 7.26    | 0.0160 |

## REDUCED MODEL II: COVARIATE BUT NO TREATMENT DIFFERENCES

51

## General Linear Models Procedure

Dependent Variable: Y SELF-ESTEEM

| Source          | DF | Sum of<br>Squares | Mean<br>Square | F Value | Pr > F |
|-----------------|----|-------------------|----------------|---------|--------|
| Model           | 1  | 212.84398         | 212.84398      | 33.67   | 0.0001 |
| Error           | 16 | 101.15602         | 6.32225        |         |        |
| Corrected Total | 17 | 314.00000         |                |         |        |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|--------|----|-------------|-------------|---------|--------|
| X1     | 1  | 212.84398   | 212.84398   | 33.67   | 0.0001 |

| Parameter | Estimate    | T for H0:<br>Parameter = 0 | Pr >  T | Std Error of<br>Estimate |
|-----------|-------------|----------------------------|---------|--------------------------|
| INTERCEPT | 28.57258960 | 13.73                      | 0.0001  | 2.08069635               |
| X1        | -0.41248834 | -5.80                      | 0.0001  | 0.07109137               |

那么酒精嗜好者和吸毒者真的有不同的自尊分数吗？对分数差异的一个可能解释就是我们所处理的是两个不同年龄组的数据；样本中酒精嗜好者年龄介于 15



到 25 岁,而吸毒者则介于 30 到 41 岁。上面输出结果中的散点图证实了这两组间所存在的年龄的差异。

两组的平均年龄分别为 20.44 和 35.67,而总平均年龄为 28.06。注意,两个样本组的年龄范围内都不包含总平均年龄。我们没有关于 30 岁以下吸毒者自尊分数的信息,也没有 25 岁以上酒精嗜好者的信息。因此,对于两个由“调整”年龄(28.06)得到的调整后的自尊分数做比较就不太合适了,因为其中包括了对所观测到的样本的年龄范围之外的外推。在此例中,我们很难对不同年龄或存在其他差异(未测量的)的两组酒精嗜好者和吸毒者做出什么比较。

所以,千万别忘了看看你的数据。外推时潜在的问题,尽管可能不会像上面例子那样显而易见,但在数据的图中可能会明显地看出来。这样,当作出的调整(或做任何比较)本身就不妥当时,可以避免使用协方差分析对调整后的处理均值做比较。当将这些分析方法扩展到分析不止一个协变量或更复杂的试验设计时,也可能会遇到相同的问题——但此时将更难于检测到这些问题。

## 16.4 多维协变量和更复杂的设计

16.2 节中讨论的方法也可用于含有一个或多个协变量的完全随机化设计。模型中包含有多个协变量只意味着我们的模型中含有不止一个定量的自变量。例如,我们可能希望比较几个不同职业团体的社会地位  $y$ ,同时又要考虑每个群体中每个人在中学后受到正式教育的年数  $x_1$ ,每个人的收入水平  $x_2$  等信息。如前面提到的,我们不要把自己局限于协变量的一次方。这样我们可以用下面的模型将响应变量与两个协变量( $x_1$  和  $x_2$ )及三个处理的联系表示出来:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_3 + \beta_5 x_4 + \beta_6 x_1 x_3 + \beta_7 x_1 x_4 + \beta_8 x_1^2 x_3 \\ + \beta_9 x_1^2 x_4 + \beta_{10} x_2 x_3 + \beta_{11} x_2 x_4 + \epsilon$$

其中

$$x_3 = 1 \text{ 当使用处理 2} \quad x_3 = 0 \text{ 否则}$$

$$x_4 = 1 \text{ 当使用处理 3} \quad x_4 = 0 \text{ 否则}$$

用像表 16.1 那样的期望值表我们可以很容易地给出各个  $\beta$  的解释。

对更复杂设计的协方差分析也可以通过一般线性模型的方法解决。在随机化完全区组设计和拉丁方设计中,根据协变量作调整时的技巧类似于我们在完全随机化设计中讨论过的方法。我们用下例来说明对于随机化区组设计模型是如何建立的。

**例 16.4**

假设我们有一随机化完全区组设计,两个区组,三个处理,一个协变量  $x$ , 每个区组内每个处理处有  $n > 1$  次观测。假设对每个处理,响应变量与协变量都有着线性关系,写出这个试验的模型。给出模型中参数的意义。

**解答** 模型如下

$$y_{ijk} = \beta_0 + \alpha_i + \gamma_j + \beta_1 x_{ijk} + \varepsilon_{ijk}$$

其中  $i = 1, 2, 3; j = 1, 2$ ; 及  $k = 1, \dots, n$ 。参数的意义如下:  $\beta_0$  是  $y$  关于  $x$  的回归直线的截距,  $\alpha_i$  是第  $i$  个处理的效应,  $\gamma_j$  是第  $j$  个区组的效应,  $\beta_1$  是  $y$  关于  $x$  的回归直线的斜率,  $\varepsilon_{ij}$  是随机误差项。我们可以用一般线性模型写做:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 x_2 + \beta_6 x_1 x_3 + \beta_7 x_1 x_4 + \varepsilon$$

其中

$x_1$  = 协变量

$x_2 = 1$  对于区组 2                       $x_2 = 0$  否则

$x_3 = 1$  对于处理 2                       $x_3 = 0$  否则

$x_4 = 1$  对于处理 3                       $x_4 = 0$  否则

由此我们立即可以看出,该模型将响应变量  $y$  和一个定量变量  $x_1$  及两个定性变量即区组和处理联系起来。模型中各个  $\beta$  的解释可从表 16.5 中的期望值得到。

**表 16.5 有一个协变量的随机区组设计的期望取值**

| 区组 | 处理 | 期望取值  |
|----|----|---|
| 1  | 1  | $\beta_0 + \beta_1 x_1$   |
|    | 2  | $(\beta_0 + \beta_3) + (\beta_1 + \beta_6) x_1$                     |
|    | 3  | $(\beta_0 + \beta_4) + (\beta_1 + \beta_7) x_1$                     |
| 2  | 1  | $(\beta_0 + \beta_2) + (\beta_1 + \beta_5) x_1$                     |
|    | 2  | $(\beta_0 + \beta_2 + \beta_3) + (\beta_1 + \beta_5 + \beta_6) x_1$ |
|    | 3  | $(\beta_0 + \beta_2 + \beta_4) + (\beta_1 + \beta_5 + \beta_7) x_1$ |

例 16.4 中建立的模型除了可以表示出各区组内每个处理处  $y$  与  $x_1$  的线性关系,它还可以表示出截距间以及斜率间的差异。如果我们想检验处理间和区组间斜率是否相等,我们可以检验如下零假设:

$$H_0: \beta_5 = \beta_6 = \beta_7 = 0$$

如没有足够的依据拒绝  $H_0$ ,我们可以再考虑简化模型(由在上述模型中令  $\beta_5 = \beta_6$

$= \beta_7 = 0$  得到)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon$$

检验对于协变量调整后的处理间是否有差异, 可以通过拟合全模型和对于零假设

$$H_0: \beta_3 = \beta_4 = 0$$

的简化模型来进行。

### 案例的数据分析: 高尔夫球场草坪上的球速

此研究的目的在于比较击中后的高尔夫球在用三种不同品种的草皮铺成的高尔夫球场上的平均速度。对应数据的图由图 16.1 给出。由图上看, 响应变量——击中后球的速度——对三种草皮以相似的斜率系数与相对湿度呈线性关系。我们将根据数据建立模型, 评价模型条件是否满足, 并检验对于三种草皮调整后的速度是否有差异。因为土壤性质、天气条件等方面存在地域差异, 我们选取了八个不同的地区进行试验。在每个试验地点, 给每种草皮都分配一块草地。协变量——相对湿度, 在得到每块草地上速度测量值的同时被记录下来。这样, 我们就有一个随机化完全区组设计, 其中八个区组(即试验地区), 三个处理(草皮)以及一个协变量(相对湿度)。我们通过建立模型, 将响应变量(草地上的球速)与区组、处理和协变量联系起来, 并假设每个地区内对三个处理(草皮)可以有不同的斜率, 但假设同一处理在不同地区具有相同的斜率。

模型 I: 以一般线性模型的符号, 上述情况下的全模型表示如下:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_9 x_1 + \beta_{12} x_{10} x_1 + \epsilon$$

其中

$x_1$  = 相对湿度(协变量)

$x_2 = 1$  如果是地区 1

$x_2 = 0$  否则

$x_3 = 1$  如果是地区 2

$x_3 = 0$  否则

$x_4 = 1$  如果是地区 3

$x_4 = 0$  否则

$x_5 = 1$  如果是地区 4

$x_5 = 0$  否则

$x_6 = 1$  如果是地区 5

$x_6 = 0$  否则

$x_7 = 1$  如果是地区 6

$x_7 = 0$  否则

$x_8 = 1$  如果是地区 7

$x_8 = 0$  否则

$x_9 = 1$  如果是草皮 1

$x_9 = 0$  否则

$x_{10} = 1$  如果是草皮 2

$x_{10} = 0$  否则

模型 I 中的各期望值见表 16.6。

注意草皮品种间具有不同的斜率, 但每种草皮在不同地区的斜率却相同。

要检验速度与相对湿度间的线性关系对于三种草皮是否一样——即三条直线是否斜率相等,我们将用以下模型拟合数据,其中三条直线有相同的斜率,但有不同的截距。

模型 II:协变量斜率相同时地区和草皮的差异

表 16.6 案例中模型 I 中的期望值

| 地区 | 草皮品种  |   |                                     |
|----|---|---|-------------------------------------|
|    | 1   | 2   | 3                                   |
| 1  | $(\beta_0 + \beta_2) + (\beta_1 + \beta_{11})x_1$ | $(\beta_0 + \beta_2) + (\beta_1 + \beta_{12})x_1$ | $(\beta_0 + \beta_2) + \beta_1 x_1$ |
| 2  | $(\beta_0 + \beta_3) + (\beta_1 + \beta_{11})x_1$ | $(\beta_0 + \beta_3) + (\beta_1 + \beta_{12})x_1$ | $(\beta_0 + \beta_3) + \beta_1 x_1$ |
| ⋮  | ⋮   | ⋮   | ⋮                                   |
| 7  | $(\beta_0 + \beta_8) + (\beta_1 + \beta_{11})x_1$ | $(\beta_0 + \beta_8) + (\beta_1 + \beta_{12})x_1$ | $(\beta_0 + \beta_8) + \beta_1 x_1$ |
| 8  | $\beta_0 + (\beta_1 + \beta_{11})x_1$             | $\beta_0 + (\beta_1 + \beta_{12})x_1$             | $\beta_0 + \beta_1 x_1$             |

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 \\ + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \epsilon$$

根据上两个模型拟合得到的计算机输出结果如下。

MODEL I: REGION AND TREATMENT DIFFERENCES WITH COVARIATE HAVING UNEQUAL SLOPES

General Linear Models Procedure

Dependent Variable: S SPEED

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 12 | 18.592940      | 1.549412    | 54.78   | 0.0001 |
| Error           | 11 | 0.311125       | 0.028284    |         |        |
| Corrected Total | 23 | 18.904065      |             |         |        |
| Source          | DF | Type III SS    | Mean Square | F Value | Pr > F |
| X1              | 1  | 0.8535690      | 0.8535690   | 30.18   | 0.0002 |
| X2              | 1  | 0.2203698      | 0.2203698   | 7.79    | 0.0175 |
| X3              | 1  | 0.1929255      | 0.1929255   | 6.82    | 0.0242 |
| X4              | 1  | 0.1384441      | 0.1384441   | 4.89    | 0.0490 |
| X5              | 1  | 0.2766251      | 0.2766251   | 9.78    | 0.0096 |

|          |   |           |           |       |        |
|----------|---|-----------|-----------|-------|--------|
| X6       | 1 | 0.0476899 | 0.0476899 | 1.69  | 0.2207 |
| X7       | 1 | 0.0011902 | 0.0011902 | 0.04  | 0.8412 |
| X8       | 1 | 0.0189954 | 0.0189954 | 0.67  | 0.4299 |
| X9       | 1 | 0.4863322 | 0.4863322 | 17.19 | 0.0016 |
| X10      | 1 | 0.0252386 | 0.0252386 | 0.89  | 0.3651 |
| X1 * X9  | 1 | 0.0902496 | 0.0902496 | 3.19  | 0.1016 |
| X1 * X10 | 1 | 0.1332566 | 0.1332566 | 4.71  | 0.0527 |

MODEL II: REGION AND TREATMENT DIFFERENCES WITH COVARIATE HAVING EQUAL SLOPES

General Linear Models Procedure

Dependent Variable: S SPEED

| Source          | DF | Sum of      | Mean        | F Value | Pr > F |
|-----------------|----|-------------|-------------|---------|--------|
|                 |    | Squares     | Square      |         |        |
| Model           | 10 | 18.435323   | 1.843532    | 51.13   | 0.0001 |
| Error           | 13 | 0.468741    | 0.036057    |         |        |
| Corrected Total | 23 | 18.904065   |             |         |        |
| Source          | DF | Type III SS | Mean Square | F Value | Pr > F |
| X1              | 1  | 3.135616    | 3.135616    | 86.96   | 0.0001 |
| X2              | 1  | 0.099813    | 0.099813    | 2.77    | 0.1201 |
| X3              | 1  | 0.166114    | 0.166114    | 4.61    | 0.0513 |
| X4              | 1  | 0.169050    | 0.169050    | 4.69    | 0.0496 |
| X5              | 1  | 0.234314    | 0.234314    | 6.50    | 0.0242 |
| X6              | 1  | 0.039421    | 0.039421    | 1.09    | 0.3148 |
| X7              | 1  | 0.009653    | 0.009653    | 0.27    | 0.6136 |
| X8              | 1  | 0.039862    | 0.039862    | 1.11    | 0.3122 |
| X9              | 1  | 14.089314   | 14.089314   | 390.75  | 0.0001 |
| X10             | 1  | 3.730729    | 3.730729    | 103.47  | 0.0001 |

要检验斜率是否相等,可检验模型 I 中的如下假设:

$$H_0: \beta_{11} = \beta_{12} = 0 \quad H_a: \beta_{11} \neq 0 \text{ 且 / 或 } \beta_{12} \neq 0$$

检验统计量为

$$F = \frac{(SSE_{II} - SSE_I) / (df_{EII} - df_{EI})}{MSE_I} = \frac{(.4687 - 0.3111) / (13 - 11)}{0.0283} = 2.79$$

$p$  值为  $\Pr(F_{2,11} \geq 2.79) = 0.1050$ 。可见,此数据支持三种草皮具有相同斜率的假设。其次,我们可以检验这三种草皮调整后的均值是否有差异。我们可建立一个模型,其中对这三种草皮协变量的斜率相同,但不再考虑草皮间的差异,同时保留由于区组变量——地区所造成的差异。

模型 III:协变量的斜率相同,地区间有差异而草皮品种间无差异

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \epsilon$$

此拟合此模型的计算机输出结果如下。

MODEL III: COVARIATE WITH EQUAL SLOPES, REGION BUT NO  
CULTIVAR DIFFERENCES

General Linear Models Procedure

Dependent Variable: S SPEED

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 8  | 4.3410695      | 0.5426337   | 0.56    | 0.7950 |
| Error           | 15 | 14.5629951     | 0.9708663   |         |        |
| Corrected Total | 23 | 18.9040646     |             |         |        |
| Source          | DF | Type III SS    | Mean Square | F Value | Pr > F |
| X1              | 1  | 2.0172033      | 2.0172033   | 2.08    | 0.1700 |
| X2              | 1  | 0.1190733      | 0.1190733   | 0.12    | 0.7311 |
| X3              | 1  | 0.1670196      | 0.1670196   | 0.17    | 0.6842 |
| X4              | 1  | 0.2163489      | 0.2163489   | 0.22    | 0.6437 |
| X5              | 1  | 0.4184864      | 0.4184864   | 0.43    | 0.5214 |
| X6              | 1  | 0.0032885      | 0.0032885   | 0.00    | 0.9544 |
| X7              | 1  | 0.0000020      | 0.0000020   | 0.00    | 0.9989 |
| X8              | 1  | 0.0031418      | 0.0031418   | 0.00    | 0.9554 |

为考察调整后各草皮品种所对应的均值是否有差异,我们作如下检验

$$H_0: \mu_{\text{Adj}, C1} = \mu_{\text{Adj}, C2} = \mu_{\text{Adj}, C3} \quad H_a: \mu_{\text{Adj}, C} \text{ 不全相等}$$

这组假设相当于检验模型 II 中的假设:

$$H_0: \beta_9 = \beta_{10} = 0 \quad H_a: \beta_9 \neq 0 \text{ 且 / 或 } \beta_{10} \neq 0$$

检验统计量为:

$$F = \frac{(\text{SSE}_{\text{III}} - \text{SSE}_{\text{II}}) / (\text{df}_{\text{EIII}} - \text{df}_{\text{EII}})}{\text{MSE}_{\text{II}}} = \frac{(14.5640 - 0.4687) / (15 - 13)}{0.0361} = 195.21$$

$p$  值为  $\Pr(F_{2,13} \geq 195.21) < 0.0001$ 。可见,数据显著支持研究假设,即三种草皮上的调整后速度间有显著差异。我们可以通过检查图 16.5 的速度——相对湿度

图来进一步调查在这三种草皮间存在什么样的差异, 其中的直线是通过模型 II 做参数估计后得到的。我们可以观察到草皮 C3 所对应的速度总是比另两种草皮上的速度要快, 而 C2 又要比 C1 快。

图中符号为草皮的种类

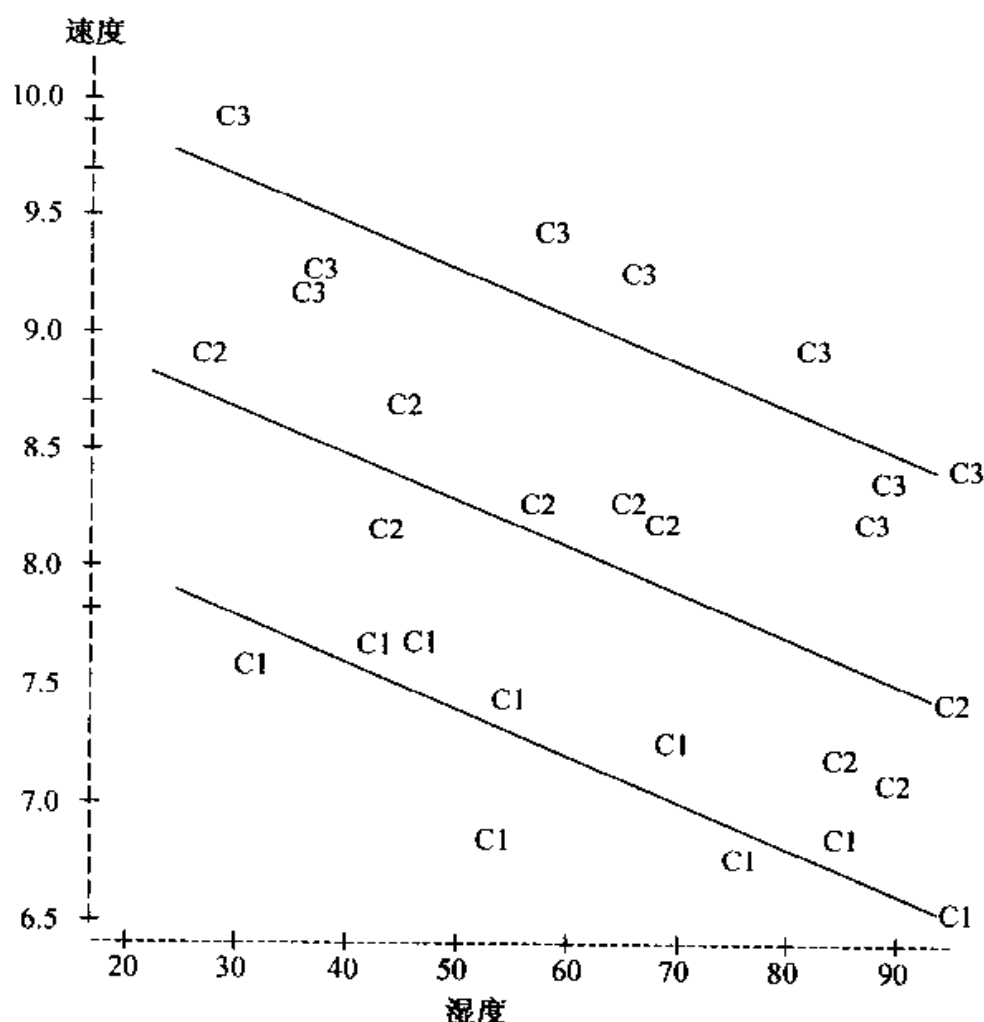


图 16.5 速度-相对湿度图以及拟合的回归直线

调整后平均速度的估计值由表 16.7 给出, 同时给出了它们的标准差, 以用来构造平均速度的 95% 的置信区间。从表 16.7 可见, 草皮 C3 上的调整后的速度平均值比 C2 要快约 1 个单位, 而 C2 上的速度要比 C1 上的速度快约 1 个单位。这样大小的速度均值差异被认为是很有实际意义的差异, 从而将会大大帮助高尔夫球场的设计者们去为他们的球场选择合适的草皮品种。

在应用上面得到的结果之前, 研究者必须检查协方差分析模型中设定的条件在此试验中是否满足。对下面的残差图和观测数据的图进行研究将有助于检验模型假设的有效性。计算机得出的模型 II 的残差分析结果见图 16.6。

表 16.7 调整后平均速度的估计值及其 95% 的置信区间

| 草皮 | $\hat{\mu}_{Adj}$ | $SE(\hat{\mu}_{Adj})$ | 95% 置信区间     |
|----|-------------------|-----------------------|--------------|
| C1 | 7.20              | 0.0674                | (7.05, 7.35) |
| C2 | 8.12              | 0.0672                | (7.98, 8.72) |
| C3 | 9.08              | 0.0672                | (8.94, 9.23) |

残差的茎叶图和盒形图并没有显示有极端值的存在。正态概率图显示有少数残差值偏离了拟合直线。然而正态性检验得到的  $p$  值为 0.3405, 显著支持残差具有正态性。残差—预测值图上, 各残差值总稳定地分布在预测值附近, 说明并未违背残差等方差的假设。我们对等斜率的假设也进行了检验, 结果是此条件可以满足。从图 16.5 中标出的数值我们可看到在速度与相对湿度间存在着线性关系。这样, 在本试验进行中并没有违反协方差分析的前提条件。

**报告结论** 我们需要写一份报告来总结在此试验中的发现, 它应该包括下述内容。

1. 对此项研究目的的阐述;
2. 对试验设计的描述, 地区是如何选的, 每个地区的三块草坪是如何得到的, 草皮品种是如何分配到草坪上的;
3. 对此项研究结果的可推广性的讨论;
4. 数据集的数值和图形概括;
5. 所用的各种推断方法的描述:
  - 给出所有研究假设及其  $F$  检验值和  $p$  值,
  - 列出调整后处理均值的估计值和置信区间,
  - 证明得到推论所需要的必要条件已得到满足;
6. 对试验结果和所得结论的讨论;
7. 相对于以前的研究, 对此项研究的发现进行解释;
8. 对将来研究的建议;
9. 列出数据集。

Univariate Procedure variable=RESIDUALS

| Moments  |          |          |          |
|----------|----------|----------|----------|
| N        | 24       | Sum Wgts | 24       |
| Mean     | 0        | Sum      | 0        |
| Std Dev  | 0.142759 | Variance | 0.02038  |
| Skewness | 0.522974 | Kurtosis | -0.22996 |
| W:Normal | 0.954191 | Pr<W     | 0.3405   |



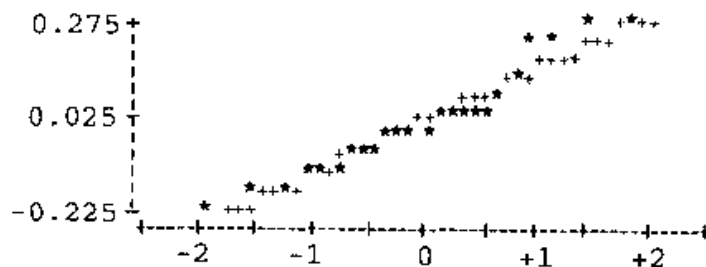
Variable-RESIDUALS

| Stem | Leaf  | # | Boxplot |
|------|-------|---|---------|
| 2    | 79    | 2 |         |
| 2    | 12    | 2 |         |
| 1    |       |   |         |
| 1    | 3     | 1 |         |
| 0    | 8     | 1 |         |
| 0    | 11134 | 5 |         |
| -0   | 220   | 3 |         |
| -0   | 8755  | 4 |         |
| -1   | 433   | 3 |         |
| -1   | 6     | 1 |         |
| -2   | 40    | 2 |         |

-----+-----+-----+-----+

Multiply Stem.Leaf by 10\*\*-1

Normal Probability Plot



Plot of RESIDUALS versus PREDICTED

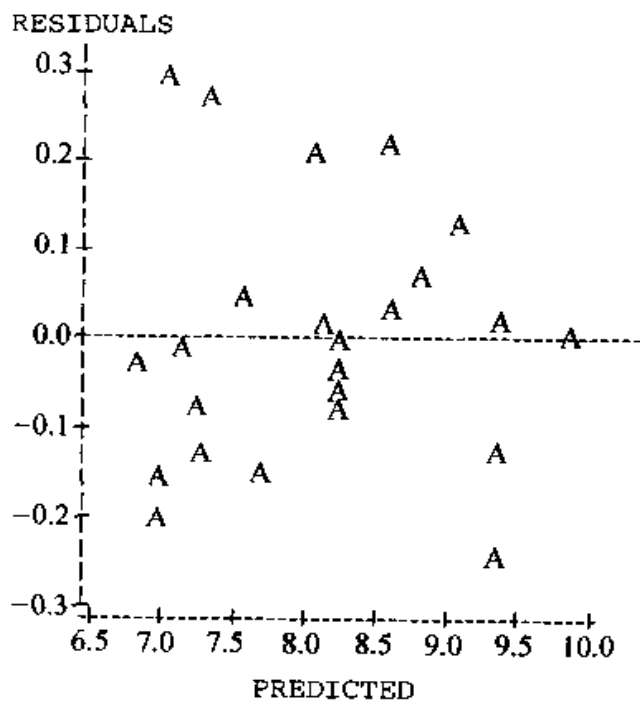


图 16.6

## 练习

## 基本技能

**16.6** 写出含一个协变量  $x_1$  的  $4 \times 4$  的拉丁方设计的模型。假设响应变量与协变量间的关系是线性的。解释模型中的参数的意义。

**16.7** 参见 16.6。

- a. 指出你将如何检验不同的直线是否平行?  $F$  检验的自由度是如何确定的?
- b. 指出你将如何检验根据协变量调整后的处理效应。

**16.8** 参见 16.8。假设响应变量是协变量  $x_1$  的二次函数, 写出它的完全模型。你能给出模型参数的解释吗? 在二次模型中你如何检验是否平行?

## 16.5 小 结

本章中, 我们给出了协方差分析的方法。在这里, 对  $y$  的每个取值, 我们同时观测到一个伴随变量  $x$  的取值。这第二个变量, 即协变量, 可看为不受控制的定量的自变量。正是由于这一点, 我们可以用前面章节中的一般线性模型的方法来建立模型。

很多情况下, 涉及到协方差分析时, 总是假设响应变量与协变量  $x$  间的关系是线性的, 且对于各个处理组, 直线的斜率是相等的。然后可以对根据协变量调整后的处理均值进行检验。事实上, 很多人在进行协方差分析时并没有检验关于平行的假设是否成立。与其给某一试验的情形强加一个特殊模型, 不如通过设计去假定一个合理(不一定是线性的)模型来处理响应变量  $y$  和协变量  $x$  的关系。然后, 在知道模型中参数的含义后, 我们可以提出关于参数的一些假设并通过完全模型和简略模型来检验它们。

## 补充练习

**16.9(医药)** 一名研究者要研究三种不同的抗抑郁药(A, B 和 C)对抑郁症病人的效用。为此, 把病人根据年龄-性别的组合分成了六组。每组随机选出三个病人, 试验者将三种药随机地分配给他们。在试验开始的当天, 首先记录下了每个病人的基数即治疗前的抑郁水平读数。然后开始已分配好的治疗方案, 并保持一周。此时, 再记录下每个病人的第二个读数即治疗后的水平。治疗前后的水平读数见下(分数越高说明抑郁程度越深)。

- a. 确定试验的设计方法。
- b. 写出治疗后的响应变量  $y$  与治疗前的读数  $x_1$  的一阶关系模型。

| 区组 | 性别 | 年龄<br>(年) | 治疗前 |    |    | 治疗后 |    |    |
|----|----|-----------|-----|----|----|-----|----|----|
|    |    |           | A   | B  | C  | A   | B  | C  |
| 1  | F  | <20       | 48  | 36 | 31 | 21  | 25 | 17 |
| 2  | F  | 20—40     | 43  | 31 | 28 | 22  | 21 | 19 |
| 3  | F  | >40       | 44  | 35 | 29 | 18  | 24 | 18 |
| 4  | M  | <20       | 42  | 38 | 29 | 26  | 20 | 17 |
| 5  | M  | 20—40     | 37  | 34 | 28 | 21  | 24 | 15 |
| 6  | M  | >40       | 41  | 36 | 26 | 18  | 24 | 19 |

16.10 参见 16.9。

a. 用计算机程序完成练习 16.9(b)部分的模型拟合, 取  $\alpha = 0.05$ 。

b. 检验直线是否平行。

c. 假设直线是平行的, 检验根据协变量调整后的处理均值是否有差异, 取  $\alpha = 0.05$ 。

16.11 参见 16.9 和 16.10。

a. 假设这些响应的线是平行的, 检验根据协变量调整后的区组是否有差异, 取  $\alpha = 0.05$ 。

b. 你如何将区组平方和划分为五个自由度为 1 的平方和?

c. 写出一个模型并进行(b)中的检验, 取  $\alpha = 0.05$ 。

**16.12(社会)** 要进行一项研究来确定社会经济因素会否影响幼儿的语言表达能力。首先定义了四个社会经济阶层, 并选了 20 个六岁以下的孩子参加试验。研究假设为这四个阶层的平均表达能力是不同的。研究者认为, 对幼儿来说, 他们的表达能力能在几个月内有显著的提高。因此, 他们决定记录下来每个孩子的确切的年龄(到月)和语言表达能力(通过测验来衡量)。数据如下。

| 社 会 经 济 阶 层 |      |     |      |     |      |     |      |
|-------------|------|-----|------|-----|------|-----|------|
| 1           |      | 2   |      | 3   |      | 4   |      |
| 年龄          | 语言表达 | 年龄  | 语言表达 | 年龄  | 语言表达 | 年龄  | 语言表达 |
| (月)         | 水平   | (月) | 水平   | (月) | 水平   | (月) | 水平   |
| 40          | 26.2 | 20  | 20.8 | 54  | 34.3 | 27  | 33.1 |
| 37          | 27.5 | 65  | 39.0 | 27  | 25.1 | 36  | 37.1 |
| 30          | 19.6 | 51  | 34.3 | 25  | 27.0 | 23  | 47.3 |

续表

| 社 会 经 济 阶 层 |      |     |      |     |      |     |      |
|-------------|------|-----|------|-----|------|-----|------|
| 1           |      | 2   |      | 3   |      | 4   |      |
| 年龄          | 语言表达 | 年龄  | 语言表达 | 年龄  | 语言表达 | 年龄  | 语言表达 |
| (月)         | 水平   | (月) | 水平   | (月) | 水平   | (月) | 水平   |
| 61          | 43.2 | 56  | 39.4 | 44  | 29.1 | 31  | 47.3 |
| 41          | 32.4 | 16  | 23.7 | 31  | 33.3 | 48  | 53.7 |
| 21          | 23.5 | 29  | 23.8 | 39  | 38.4 | 48  | 59.6 |
| 18          | 15.6 | 20  | 37.2 | 25  | 14.9 | 16  | 36.0 |
| 36          | 18.5 | 20  | 33.0 | 18  | 38.7 | 32  | 41.2 |
| 16          | 23.6 | 17  | 21.9 | 17  | 32.7 | 31  | 44.2 |
| 41          | 21.0 | 35  | 36.1 | 22  | 34.0 | 24  | 48.9 |
| 19          | 11.9 | 25  | 31.7 | 24  | 23.8 | 20  | 53.0 |
| 30          | 10.2 | 21  | 27.6 | 28  | 13.3 | 26  | 42.8 |
| 26          | 29.8 | 27  | 26.0 | 23  | 32.4 | 24  | 50.8 |
| 28          | 20.6 | 25  | 20.3 | 17  | 36.2 | 33  | 42.1 |
| 16          | 13.5 | 25  | 32.6 | 26  | 33.7 | 21  | 42.6 |
| 28          | 17.2 | 28  | 25.8 | 23  | 29.2 | 25  | 45.0 |
| 19          | 29.3 | 33  | 21.2 | 26  | 33.2 | 37  | 59.8 |
| 34          | 25.6 | 16  | 36.3 | 35  | 28.5 | 36  | 37.9 |
| 20          | 25.6 | 22  | 34.2 | 31  | 31.4 | 19  | 38.9 |
| 18          | 18.4 | 23  | 17.7 | 37  | 36.2 | 34  | 45.0 |

- a. 画出数据的散点图。对每个组来说,表达能力和年龄呈线性关系吗?  
 b. 写出表达能力与年龄是一次关系时的模型,其中各社会经济阶层都有其对应的直线。

16.13 参见 16.12。

- a. 根据下面的计算机输出结果,检验这四个社会经济阶层中表达能力与年龄间的关系方程是否是平行的直线。  
 b. 这四组的平均语言表达能力分数有显著差别吗? 检验此假设。取  $\alpha = 0.05$   
 c. 给出每组中调整后表达能力的平均分数的 95% 的置信区间。

MODEL I; DIFFERENT SLOPES AND TREATMENT DIFFERENCES

General Linear Models Procedure

Dependent Variable: Y VSKILL

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 7  | 6380.2373      | 911.4625    | 20.91   | 0.0001 |
| Error           | 72 | 3138.4195      | 43.5892     |         |        |
| Corrected Total | 79 | 9518.6569      |             |         |        |
| Source          | DF | Type III SS    | Mean Square | F Value | Pr > F |
| X1(AGE)         | 1  | 188.81711      | 188.81711   | 4.33    | 0.0410 |
| X2(C1)          | 1  | 577.36595      | 577.36595   | 13.25   | 0.0005 |
| X3(C2)          | 1  | 170.50366      | 170.50366   | 3.91    | 0.0518 |
| X4(C3)          | 1  | 29.75438       | 29.75438    | 0.68    | 0.4114 |
| X5(X1 * X2)     | 1  | 5.46901        | 5.46901     | 0.13    | 0.7242 |
| X6(X1 * X3)     | 1  | 13.40413       | 13.40413    | 0.31    | 0.5809 |
| X7(X1 * X4)     | 1  | 72.26787       | 72.26787    | 1.66    | 0.2020 |

MODEL II; SAME SLOPES BUT TREATMENT DIFFERENCES

General Linear Models Procedure

Dependent Variable: Y VSKILL

| Source          | DF           | Sum of Squares          | Mean Square | F Value               | Pr > F |
|-----------------|--------------|-------------------------|-------------|-----------------------|--------|
| Model           | 4            | 6222.8867               | 1555.7217   | 35.40                 | 0.0001 |
| Error           | 75           | 3295.7702               | 43.9436     |                       |        |
| Corrected Total | 79           | 9518.6569               |             |                       |        |
| Parameter       | Estimate     | T for H0; Parameter = 0 | Pr >  T     | Std Error of Estimate |        |
| INTERCEPT       | 37.03395378  | 14.59                   | 0.0001      | 2.53830797            |        |
| X1(AGE)         | 0.28023845   | 4.02                    | 0.0001      | 0.06973064            |        |
| X2(C1)          | -22.17859463 | -10.56                  | 0.0001      | 2.09962198            |        |
| X3(C2)          | -15.94679732 | -7.60                   | 0.0001      | 2.09711070            |        |
| X4(C3)          | -14.91889270 | -7.12                   | 0.0001      | 2.09650778            |        |

MODEL III; SAME SLOPES AND NO TREATMENT DIFFERENCES

General Linear Models Procedure

Number of observations in data set = 80

General Linear Models Procedure

Dependent Variable: Y VSKILL

| Source          | DF          | Sum of        | Mean        | F Value      | Pr > F |
|-----------------|-------------|---------------|-------------|--------------|--------|
|                 |             | Squares       | Square      |              |        |
| Model           | 1           | 930.68400     | 930.68400   | 8.45         | 0.0047 |
| Error           | 78          | 8587.97288    | 110.10222   |              |        |
| Corrected Total | 79          | 9518.65687    |             |              |        |
| Source          | DF          | Type III SS   | Mean Square | F Value      | Pr > F |
| X1              | 1           | 930.68400     | 930.68400   | 8.45         | 0.0047 |
|                 |             | T for H0:     | Pr >  T     | Std Error of |        |
| Parameter       | Estimate    | Parameter = 0 |             | Estimate     |        |
| INTERCEPT       | 22.61763850 | 6.69          | 0.0001      | 3.38324077   |        |
| X1              | 0.32035109  | 2.91          | 0.0047      | 0.11018516   |        |

**16.14(商业)** 一个主食发售商想评价三种不同的促销方式对微波比萨饼销量的作用。公司的市场部确定了三种用于杂货店的促销方法。

促销方法 1 分发产品样品给商店中的顾客,同时不在商店增设货架。

促销方法 2 在商店中通常的位置增设货架。

促销方法 3 除在通常位置设立货架外,在过道尽头设立专门的展示架用完全随机化设计,为每种促销方法随机地分配五个商店。市场部希望尽量控制除促销方法外的其他可能会影响产品销量的因素。例如,这 15 个商店的价格以及对此商品的广告宣传都相同。促销期间的产品销量( $y$ )被记录下来。因为各商店可能会有不同的顾客种类,即顾客是否更倾向于购买该产品,所以每个商店在促销前的该产品销量( $x$ )也被记录下来。控制住这个因素是很关键的,因为可以想象在促销前销量最高的商店应该在促销进行时销量也很高。这将会遮蔽促销方法作用之间的差异。数据如下。

| 促销方法 |     |     |     |     |     |
|------|-----|-----|-----|-----|-----|
| 1    |     | 2   |     | 3   |     |
| $y$  | $x$ | $y$ | $x$ | $y$ | $x$ |
| 47   | 31  | 52  | 44  | 33  | 33  |
| 48   | 36  | 47  | 36  | 41  | 39  |
| 45   | 32  | 47  | 39  | 40  | 40  |
| 54   | 38  | 36  | 28  | 30  | 26  |
| 42   | 29  | 43  | 35  | 37  | 39  |

- a. 画出数据的散点图。对每种促销方法,促销时的销量和促销前的销量是否是线性关系?

- b. 写出用促销前的销量进行调整的、促销时销量和促销方法间关系的模型。
- c. 用计算机程序作(b)中模型的拟合。
- d. 检验促销时销量和促销前销量间的方程是否是平行线。
- e. 三种促销方法的平均销量有显著差异吗? 检验此假设, 取  $\alpha = 0.05$ 。
- f. 对三种促销方法, 给出各自促销时的调整后平均销量的 95% 的置信区间。

# 第十七章 一些固定效应、随机效应 和混合效应模型的方差分析

## 17.1 引言和案例

## 17.2 具有随机处理效应的单因子试验:随机效应模型

## 17.3 随机效应模型的扩充

## 17.4 混合效应模型

## 17.5 计算期望均方的规则

## 17.6 套抽样和裂区设计

## 17.7 小结

## 17.1 引言和案例

在前面章节所遇到的试验和研究中,都含有试验因子和处理,研究中处理的水平是由研究人员选定的。从试验数据所进行的推断仅是针对这些水平而言的。第十六章的案例中包含三个新的草皮品种,并且这是研究者感兴趣的仅有的草皮品种。在这个试验中,要研究的总体仅仅是这三个草皮品种所对应的球速的三个总体。

如果高尔夫联合会决定,为了验证在原始试验中得到的平均速度,要重复进行试验。这三种草皮可能会种植在另外一组草坪上并重复这个试验。在各个因子均有一组事先选定的水平时,用于考察响应变量变异性的模型称为固定效应模型。从这样的模型所得到的推断仅限于研究中所用到的那组特定的处理水平。

**定义 17.1** 在一个试验的**固定效应模型**中,试验中的所有因子都有一组事先决定的水平,并且推断仅对该试验中实际用到的因子水平进行。

在某些研究中,主要的兴趣在于识别作为响应变量变异来源的因子。在产品改进研究中,质量控制工程师要确定生产工序中的哪些因子是变异的主要来源,并估计这些变异来源中的每一个对于产品的总变异的贡献。称这些因子为**方差分量**。当试验中采用的这些因子的水平是从所有可能的水平总体中随机挑选出来的时候,联系响应变量和这些因子之间的模型就是**随机效应模型**。而从这样的模型所作出的推断,也从这些随机挑选出来的试验中用到的因子水平推广至因子水平的总体。在产品改进研究中,通常的变异来源之一是工序的操作。公司可能有上百名操作人员,而被随机选中参加到研究中的可能只有五、六人。然而,质量工



程师感兴趣的是所有操作人员的表现,而不只是参加研究的那几名。

**定义 17.2** 在一个试验的**随机效应模型**中,试验中采用的因子水平是从所有可能的水平总体中随机选取的。从这样的试验数据所进行的推断,不只是针对试验中用到的那些因子水平,而是针对从中挑选这些水平的因子水平总体而言的。

在许多研究中,既包含带有事先确定的一组水平的因子,又包含其水平是从一个水平总体中随机挑选出来的因子。一个随机化试验中的区组,可能代表从某个农业研究机构的地块的总体中抽取的一个容量为  $b$  的随机样本。此时,就认为区组的效应是随机效应。假定处理是为抵抗某种病毒而开发出来的四个大豆新品种,这个处理的水平是固定的,因为研究人员只对这四个品种感兴趣;而地块的水平是随机的,因为研究人员不只对这些地块感兴趣,而是了解一片宽广的土地上这些处理的效应。当试验中用到的一些因子具有从一个因子水平总体中随机挑选出来的水平,而另一些因子具有事先确定的水平时,联系响应变量与因子水平之间关系的模型就是**混合效应模型**。

**定义 17.3** 在一个试验的**混合效应模型**中,试验中采用的某些因子的水平是从所有可能的水平总体中随机选取的,而另一些因子的水平是事先决定的。从这样的试验数据所进行的关于具有固定水平因子的推断,仅是针对试验中用到的那些水平的,而关于具有随机挑选的水平的那些因子所进行的推断,是针对从中挑选那些水平的因子水平总体中的所有水平而言的。

在这一章,我们考虑各种随机效应和混合效应模型。对于每一个模型,我们将给出适当的方差分析方法,并说明如何估计各个方差分量。下面的案例描述了一个混合效应试验。

#### **案例:扩张接缝处压降的分析**

发电站中的一个主要问题是涡轮机中扩张接缝处的压降。设计工程师想要设计一项研究来确定哪些是最可能影响压降读数的因子。一旦这些因子被确定并根据对扩张接缝处压降的影响大小决定出那些最关键的因子,工程师就可以进行生产过程中的设计改进或改变生产过程中操作者的训练方式。这些改变很可能花费昂贵或者耗费时间,所以工程师需要确定哪些因子对于减少压降效果最大。

**数据收集的设计** 为设计一个适当的试验来分析压降,设计工程师考虑到以下一些方面:

1. 哪些因子应该用于研究?
2. 对因子的哪些水平感兴趣?
3. 多少水平才足以识别主要的变异性来源?
4. 为获得可靠的方差分量的估计需要在每个因子—水平组合处作多少次重复?
5. 在试验期间哪些环境因素可能影响压力计的准确性?
6. 在分析扩张接缝处压降的变异性的来源时有有效的统计方法是什么?

7. 为了说明所有重要的变异性来源都已经被识别出来, 在最终的报告中应包括哪些信息?

本研究中选定的因子为接缝入口处的气体温度和操作者使用的压力计种类。工程师认为需要进行因子试验来决定哪个因子对于压降的效应最大。代表涡轮机操作时可能温度范围的三个温度是 15℃, 25℃ 和 35℃。从数百种用于监测线上压力的压力计中, 随机选取了四种用于研究。为获得 12 个因子水平组合中每个组合处压降期望的精确估计值, 决定在每个处理处进行 6 次重复。这 72 次试验得到的数据由表 17.1 给出。

表 17.1 扩张接缝处的压降

| 15℃ |       |       |       | 25℃   |       |       |       | 35℃   |       |       |       |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| G1  | G2    | G3    | G4    | G1    | G2    | G3    | G4    | G1    | G2    | G3    | G4    |       |
| 40  | 43    | 42    | 47    | 57    | 49    | 44    | 36    | 35    | 41    | 42    | 41    |       |
| 40  | 34    | 35    | 47    | 57    | 43    | 45    | 49    | 35    | 43    | 41    | 44    |       |
| 37  | 38    | 35    | 40    | 65    | 51    | 49    | 38    | 35    | 44    | 34    | 35    |       |
| 47  | 42    | 41    | 36    | 67    | 49    | 45    | 45    | 46    | 36    | 35    | 46    |       |
| 42  | 39    | 43    | 41    | 63    | 45    | 46    | 38    | 41    | 42    | 39    | 44    |       |
| 41  | 35    | 36    | 47    | 59    | 43    | 43    | 42    | 42    | 41    | 36    | 46    |       |
| 均值  | 41.17 | 38.50 | 38.67 | 43.00 | 61.33 | 46.67 | 45.33 | 41.33 | 39.00 | 41.17 | 37.83 | 42.67 |

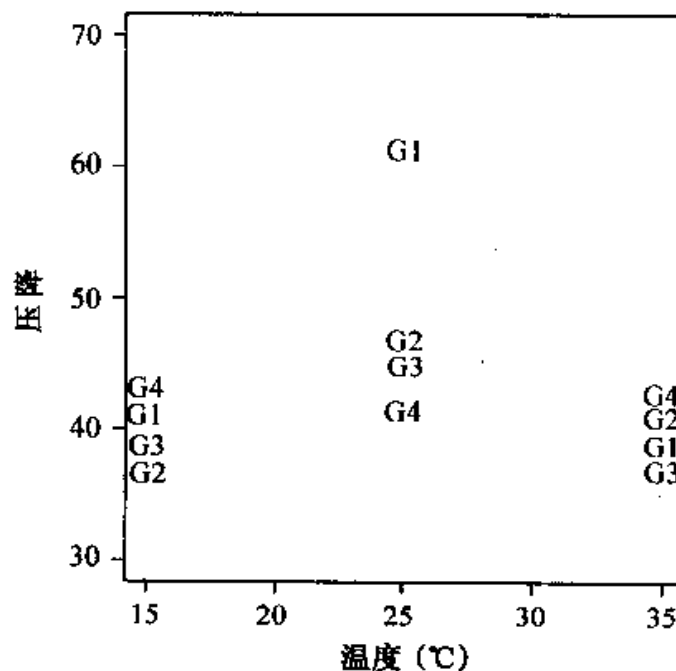


图 17.1 12 个处理上平均压降的侧面图

图 17.1 给出了 12 个处理上样本均值的侧面图。从图中我们可以看到, 压力计 G1 对应的压降比其他三种压力计对应的变异性更大。为了判定这个观测到的差异是否仅是由于随机变化, 我们将在本章接下来的部分中给出模型和分析的技巧, 以便能确定到底哪些因子对压降的总的变异贡献最大。

## 17.2 具有随机处理效应的单因子试验: 随机效应模型

对于单因子试验来说, 说明固定效应和随机效应模型差异的最好方法是举例子。假设我们想要对三个跟踪站处闪电时的静电释放强度进行比较, 这三个跟踪站是在距一个大学内的中央计算系统 20 英里半径内的。如果试验中这三个跟踪站是惟一可用的跟踪站并且要得到的就是这几个跟踪站的推断, 那么我们可以写出如下的**固定效应模型**:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

其中  $y_{ij}$  是跟踪站  $i$  ( $i = 1, 2, 3$ ) 处的第  $j$  次观测值,  $\mu$  是总均值,  $\alpha_i$  是第  $i$  个跟踪站的固定效应。我们假定  $\epsilon$  服从均值为 0 方差为  $\sigma^2$  的正态分布。

然而, 假如我们关心的不仅仅是这三个跟踪站, 而认为它们是从很多跟踪站的可能位置中选出的三个随机样本, 那么推断就不仅与发生在样本位置的事情有关, 也会与跟踪站其他可能位置处可能发生的事情有关。可以解释这个差异的模型就是**随机效应模型**:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

尽管这个模型与前面的固定效应模型看上去完全相同, 但一些假定却是不同的:

1.  $\mu$  仍是总均值, 未知常数。
2.  $\alpha_i$  是第  $i$  个跟踪站的随机效应。我们假设  $\alpha_i$  服从均值为 0 方差为  $\sigma_a^2$  的正态分布。
3.  $\alpha_i$  间相互独立。
4. 如前,  $\epsilon_{ij}$  服从均值为 0 方差为  $\sigma_e^2$  的正态分布。
5.  $\epsilon_{ij}$  间相互独立。
6. 随机分量  $\alpha_i$  与  $\epsilon_{ij}$  间相互独立。

要说明固定效应模型与随机效应模型间的差别, 我们可以假设是在重复做这个试验。对固定模型, 我们总是使用这相同的三个跟踪站, 所以关于这三个位置处期望强度或期望强度间差异的推断将是有意义的。然而对于随机效应模型, 我们将使用由另外三个跟踪站构成的随机样本 (即由另三个  $\alpha$  组成的样本)。现在不再是集中在某一次试验中的三个  $\alpha$  组成的特定集合的效应上, 而是需要检查所有可能  $\alpha$  值构成的总体的变异性。这将用如下表 17.2 的方差分析表来说明。

表 17.2 单因子试验的 AOV 表:固定或随机模型

| 来源 | SS  | df         | EMS |                          |                                 |
|----|-----|------------|-----|--------------------------|---------------------------------|
|    |     |            | MS  | 固定效应                     | 随机效应                            |
| 处理 | SST | $t - 1$    | MST | $\sigma_e^2 + n\theta_T$ | $\sigma_e^2 + n\sigma_\alpha^2$ |
| 误差 | SSE | $t(n - 1)$ | MSE | $\sigma_e^2$             | $\sigma_e^2$                    |
| 总和 | TSS | $tn - 1$   |     |                          |                                 |

可以看到,固定效应与随机效应模型的方差分析表相同,除了期望均方(EMS)这一列不同。大家可以回想起在第十三章和第十五章的表中并没有这一列,因为在备择假设下所有均方(除了 MSE)的期望取值都等于  $\sigma_e^2$  加上一个依赖于所检验的参数的正常数。一般情况下, $t$  个处理(跟踪站),每个处理处有  $r$  次观测值时的 AOV 表与表 17.2 是相同的。对固定效应模型,  $\theta_T$  是常数  $\alpha_i$  的正函数,而随机效应模型中的  $\sigma_\alpha^2$  代表了  $\alpha_i$  总体的方差。对于上面的例子,固定效应模型下三个跟踪站的强度均值相等的检验如下(同第十五章):

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0 \text{ (即,三个均值相等)}$$

$$H_a: \text{至少有一个 } \alpha \text{ 不等于 } 0$$

$$\text{T.S.: } F = \text{MST/MSE, 其中 } df_1 = t - 1, df_2 = t(n - 1)$$

随机效应模型下对  $\alpha$  总体的变异性的检验也使用这个检验统计量。零假设和备择假设为:

$$H_0: \sigma_\alpha^2 = 0$$

$$H_a: \sigma_\alpha^2 > 0$$

$$\text{T.S.: } F = \text{MST/MSE, 其中 } df_1 = t - 1, df_2 = t(n - 1)$$

因为我们假设  $\alpha$  是从均值为 0 方差为  $\sigma_\alpha^2$  的正态总体中选取的,所以零假设是说  $\alpha$  来自于均值为 0 方差也为 0 的正态总体,即该总体中所有  $\alpha$  的值都等于 0。

这样,尽管两个模型的零假设形式不同,但它们表示的意义却是很相似的。对固定效应模型,我们假设样本中的那些  $\alpha$  (也是仅有的  $\alpha$ ) 都等于 0,而在随机效应模型中,零假设是指不仅样本中的  $\alpha$ ,同时还有总体中其他的  $\alpha$  都等于 0。

同样备择假设也是如此。在固定效应模型中,我们假设至少有一个  $\alpha$  与其余的不同,即  $\alpha$  集合中有一些变异性。而对于随机效应模型来说,备择假设是  $\sigma_\alpha^2 > 0$ ;也就是说,总体中的  $\alpha$  值并不全相等。

在单因子的随机效应模型中,响应变量的均值和方差是:

$$E(y_{ij}) = \mu, \sigma_y^2 = \text{Var}(y_{ij}) = \sigma_\alpha^2 + \sigma_e^2$$

因此,在大多数随机效应模型中,我们需要确定  $\sigma_a^2$  相对于  $\sigma_e^2$  的大小,从而能估计出处理对于响应变量总的变异性的效应的大小。因为  $\sigma_a^2$  或  $\sigma_e^2$  是未知的,我们可以借助于方差分析中矩匹配估计的思想建立对这些量的估计。从下面的 AOV 表,我们可以看到 MST 的期望取值为  $\sigma_a^2 + n\sigma_e^2$ , MSE 的期望取值为  $\sigma_e^2$ 。

| 来源 | MS  | 随机效应                       |
|----|-----|----------------------------|
| 处理 | MST | $\sigma_a^2 + n\sigma_e^2$ |
| 误差 | MSE | $\sigma_e^2$               |

我们令样本均方等于其估计值,取代总体方差,得到:

$$\hat{\sigma}_e^2 = \text{MSE}, \hat{\sigma}_a^2 = (\text{MST} - \text{MSE})/n$$

这样有,  $\hat{\sigma}_y^2 = \hat{\sigma}_a^2 + \hat{\sigma}_e^2$ 。响应变量的方差由此分成两种来源,即处理和试验误差,并由此得到两种来源所占的比例,见表 17.3。

表 17.3 响应变量方差的按比例分配

| 方差来源 | 估计值  | 占总方差的比例                             |
|------|--|-------------------------------------|
| 处理   | $\hat{\sigma}_a^2 = (\text{MST} - \text{MSE})/n$         | $\hat{\sigma}_a^2/\hat{\sigma}_y^2$ |
| 误差   | $\hat{\sigma}_e^2 = \text{MSE}$                          | $\hat{\sigma}_e^2/\hat{\sigma}_y^2$ |
| 总和   | $\hat{\sigma}_y^2 = \hat{\sigma}_a^2 + \hat{\sigma}_e^2$ | 1.0                                 |

研究者也许还想要对响应变量的均值  $\mu$  进行估计。下面我们给出了  $\mu$  的点估计和估计的标准误差:

$$\hat{\mu} = \bar{y}_{..} \text{ 和 } \text{SE}(\hat{\mu}) = \sqrt{\text{MST}/tn}$$

由此我们可以得到  $\mu$  的  $100(1-\alpha)\%$  置信区间:

$$\hat{\mu} \pm t_{\alpha/2, \text{df}_{\text{ERROR}}} \text{SE}(\hat{\mu}) \text{ 或 } \bar{y}_{..} \pm t_{\alpha/2, t(n-1)} \sqrt{\text{MST}/tn}$$

### 例 17.1

考虑一下我们刚刚用来说明处理效应随机的单因子试验的那个例子。两个研究生在为—个电子工程的教授工作,他们得到资助去记录三个跟踪站处闪电时的电荷释放强度(即电场强度)。因为夏季发生雷暴的频率很高(Florida 平均每年中有 80 或更多天有暴风雨),研究生们要在地图上从 20 英里半径的区域内随机地选取一个地点,并将他们的跟踪仪器安放在那里(当然要获得主人的允许)。在五天里,每天上

午 8 点到下午 5 点间, 他们要监测他们的仪器, 直到记录下了强度的最大值。然后他们在另外两个随机选取的地点重复这个过程。得到的样本数据(伏特每米)如下表 17.4。

表 17.4 闪电时电荷释放强度(伏特每米)

| 跟踪站  | 强度    |       |       |       |       | 平均值      |
|------|-------|-------|-------|-------|-------|----------|
| 1    | 20    | 1,050 | 3,200 | 5,600 | 50    | 1,984    |
| 2    | 4,300 | 70    | 2,560 | 3,650 | 80    | 2,132    |
| 3    | 100   | 7,700 | 8,500 | 2,960 | 3,340 | 4,520    |
| 总平均值 |       |       |       |       |       | 2,878.67 |

- 写出适当的统计模型, 说明每个项的意义。
- 进行方差分析并解释你的结果, 取  $\alpha = .05$ 。
- 估计方差分量和它们相对总变异的比值。
- 估计日闪电电荷释放强度最大值的均值, 置信度 95%。

**解答** 因为跟踪站是随机选取的, 所以我们可以用单因子随机效应模型把日闪电电荷释放强度最大值  $y_{ij}$  和第  $i$  个跟踪站和第  $j$  天联系起来:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (i = 1, 2, 3; j = 1, 2, \dots, 5)$$

其中  $\mu$  是日闪电释放强度最大值的均值,  $\alpha_i$  是第  $i$  个被随机选到的跟踪站的随机效应,  $\epsilon_{ij}$  是所有其他变异性的来源所带来的随机效应。随机效应方差分析中平方和公式的计算与固定效应中的计算相同。因此我们有:

$$\begin{aligned} SST &= n \sum_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = 5[(1,984 - 2,878.67)^2 + (2,132 - 2,878.67)^2 \\ &\quad + (4,520 - 2,878.67)^2] = 20,259,573.3 \\ TSS &= \sum_{ij} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = (20 - 2,878.67)^2 + (1,050 - 2,878.67)^2 \\ &\quad + \dots + (3,340 - 2,878.67)^2 = 108,249,173.3 \end{aligned}$$

做减法,

$$SSE = TSS - SST = 108,249,173.3 - 20,259,573.3 = 87,989,600$$

我们可以由这些计算结果建立 AOV 表, 见表 17.5。

表 17.5 例 17.1 数据的 AOV 表

| 来源  | SS            | df | MS            | EMS                        | F    |
|-----|---------------|----|---------------|----------------------------|------|
| 跟踪站 | 20,269,573.3  | 2  | 10,129,786.65 | $\sigma_e^2 + 5\sigma_a^2$ | 1.38 |
| 误差  | 87,989,600.0  | 12 | 7,332,466.67  | $\sigma_e^2$               |      |
| 总和  | 108,249,173.3 | 14 |               |                            |      |

对  $H_0: \sigma_a^2 = 0$  的  $F$  检验的自由度  $df_1 = 2, df_2 = 12$ 。因为算得  $F$  值为 1.38, 没有超过 3.89(附录表 8 中取  $\alpha = .05, df_1 = 2$  及  $df_2 = 12$  得到的值), 所以我们没有足够证据说跟踪站与跟踪站间表示强度差异的随机分量之间有显著的差异。就像电力工程师假定的那样, 最好仅在一个跟踪站处工作, 因为通常强度的变异性与跟踪站与电荷释放地点的距离有关, 而我们很难控制这个来源。实际上我们可计算方差分量的估计值:

$$\sigma_e^2 = 7,332,466.67$$

$$\sigma_a^2 = (10,129,786.65 - 7,332,466.67)/5 = 559,464$$

由此得到

$$\sigma_y^2 = 7,332,466.67 + 559,464 = 7,891,930.67$$

跟踪站差异相对于总变异的比例是  $559,464/7,891,930.67 = .0709$ 。由跟踪站间差异带来的日最大闪电强度的变异性仅占总变异性的 7.1%。我们可得到日最大闪电强度均值的 95% 置信区间:

$$\bar{y} \pm t_{0.25, 12} SE(\bar{y})$$

$$2878.67 \pm (2.179) \sqrt{10,129,786.65/15} \text{ 或 } 2878.67 \pm 1790.65$$

由此, 我们有 95% 的把握说日最大闪电强度的均值在区间(1088, 4669)内。

练习

应用

**17.1(医药)** 一个医药公司想要检验一种在桶中混合所得到的液体药品的效力。为此, 从一个月内的产品中随机取了 5 桶作为样本, 每个桶中分别取四次样品。

a. 写出本试验的随机效应模型, 确定其中各项的意义。

b. 根据下面的样本数据进行方差分析, 取  $\alpha = .05$ 。

| 桶 1 | 桶 2 | 桶 3 | 桶 4 | 桶 5 |
|-----|-----|-----|-----|-----|
| 3.2 | 2.6 | 3.4 | 4.2 | 1.8 |
| 3.8 | 2.9 | 3.9 | 4.4 | 2.3 |
| 3.5 | 2.8 | 3.3 | 4.3 | 1.9 |
| 3.0 | 2.0 | 3.1 | 4.2 | 2.1 |

**17.2** 假设练习 17.1 中的医药公司想要估计从一个月生产的那种液体药品中随机选出的一桶的预期效力。

- a. 用练习 17.1 的数据, 给出随机选出的桶中药品平均效力的点估计。
- b. 给出随机选出的桶中药品平均效力的 95% 的置信区间。

### 17.3 随机效应模型的扩充

单因子试验中随机效应模型的思想可以扩充应用到第十五章提到的任何区组设计和因子试验。我们没有时间一一讨论每种情况, 但我们首先要分析一个区组效应和处理效应都随机的随机化区组设计。

要进行一项试验以检验在分析齿菌斑的 DNA 成分时不同的分析员和受试者的效应。三个女性受试者(18~20 岁)被选入这项研究。每个受试者可以保持她们的正常饮食, 同时每天补充 30mg(15 片)蔗糖。研究过程中不允许刷牙和漱口。在一周结束时, 从每个受试者的完整齿列上都刮下牙菌斑, 并分为三份。随机地选出三个分析员, 每一个都拿到那三个受试者的没有标名字的齿菌斑样本, 并要对它们的 DNA 成分作分析(以微克计)。这个两因子试验的样本数据见下表 17.6。

表 17.6 齿菌斑样本的 DNA 成分

| 分析员 | 受试者  |       |      | 平均值   |
|-----|------|-------|------|-------|
|     | 1    | 2     | 3    |       |
| 1   | 13.2 | 10.6  | 8.5  | 10.77 |
| 2   | 12.5 | 9.6   | 7.9  | 10.00 |
| 3   | 13.0 | 9.9   | 8.3  | 10.40 |
| 平均值 | 12.9 | 10.03 | 8.23 | 10.39 |

此试验是个随机化区组设计, 其中受试者代表区组, 分析员代表处理。试验单元是从受试者齿列上刮下的牙菌斑的样本。如果我们假设这三个受试者代表从一大群可能的受试者中随机选取的样本, 同样三个分析员代表从很多可能的分析员中随机选取的样本, 那么我们可以写出如下**随机效应模型**, 从而将 DNA 成分与两个因子——“分析员”和“受试者”联系起来。

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

假定如下:

1.  $\mu$  是未知的总含量均值。
2.  $\alpha_i$  是第  $i$  个分析员的随机效应, 它服从均值为 0 方差  $\sigma_\alpha^2$  的正态分布。
3.  $\alpha_i$  间相互独立。



4.  $\beta_j$  是第  $j$  个受试者的随机效应, 它是个服从均值为 0 方差为  $\sigma_\beta^2$  的正态分布的随机变量。

5.  $\beta_j$  间相互独立。

6.  $\alpha_i$ ,  $\beta_j$  和  $\epsilon_{ij}$  间相互独立。

一定要注意在假设处理效应和区组的效应是随机而不是固定时的差别。举例来说, 如果研究中的三个分析员是我们惟一感兴趣的分析员, 我们关注的就只是这几个特定分析员对 DNA 成分均值效应的差别。而现在分析员的效应是作为随机变量处理的, 我们得到的推断也就是关于研究员总体的效应的。因为已假设这个正态总体的均值为 0, 所以我们要判断方差是否大于 0。

一般的两因子完全随机化设计的 AOV 表可见表 17.7, 其中因子 A 有  $a$  个水平, B 有  $b$  个水平, 且没有重复。这个 AOV 表也可应用在随机化完全区组设计中, A 代表处理, B 代表区组。就像单因子试验的随机效应模型那样, 两因子试验中固定效应和随机效应模型的方差分析表除了期望均方项外其他都是相同的。

表 17.7 双因子试验的 AOV 表, 因子 A 有  $a$  个水平, 因子 B 有  $b$  个水平

| 来源 | SS  | df               | MS  | EMS                             |  |
|----|-----|------------------|-----|---------------------------------|--|
|    |     |                  |     | 固定效应                            | 随机效应                                   |
| A  | SSA | $a - 1$          | MSA | $\sigma_\epsilon^2 + b\theta_A$ | $\sigma_\epsilon^2 + b\sigma_\alpha^2$ |
| B  | SSB | $b - 1$          | MSB | $\sigma_\epsilon^2 + a\theta_B$ | $\sigma_\epsilon^2 + a\sigma_\beta^2$  |
| 误差 | SSE | $(a - 1)(b - 1)$ | MSE | $\sigma_\epsilon^2$             | $\sigma_\epsilon^2$                    |
| 总和 | TSS | $ab - 1$         |     |                                 |  |

平方和以及均方的计算过程与第十五章的完全相同。检验方法的差别在表 17.8 中以因子 A 为例作出了说明。同样的结果也适用于因子 B。

表 17.8 对因子 A 检验方法的差别

| 固定效应模型  | 随机效应模型                      |
|---|-----------------------------|
| $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$ | $H_0: \sigma_\alpha^2 = 0$  |
| $H_a$ : 至少有一个 $\alpha$ 与其他的不同                     | $H_a: \sigma_\alpha^2 > 0$  |
| T.S.: $F = \frac{MSA}{MSE}$                       | T.S.: $F = \frac{MSA}{MSE}$ |
| R.R.: 基于 $df_1 = a - 1, df_2 = (a - 1)(b - 1)$    | R.R.: 同左                    |

下面我们不再继续讨论例子,我们将讨论在每个因子水平组合处有  $n > 1$  次观察的因子试验的随机效应模型,然后说明它的检验方法。

在第十五章中,我们考虑过每个单元处有  $n > 1$  次观察的完全随机化设计中  $a \times b$  析因试验的固定效应模型。 $a \times b$  析因试验的随机效应模型和与之相对应的固定效应模型具有相同的形式,只是假定不同:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

其中  $y_{ijk}$  是在因子 A 的第  $i$  个水平因子 B 的第  $j$  个水平处的第  $k$  次观测的响应,  $\mu$ ,  $\alpha_i$ ,  $\beta_j$  和  $\varepsilon_{ijk}$  的定义同没有重复的随机效应模型中一样。此外我们还有如下假定:

1.  $\alpha\beta_{ij}$  是因子 A 的第  $i$  个水平因子 B 的第  $j$  个水平的随机效应,  $\alpha\beta_{ij}$  服从均值为 0 方差  $\sigma_{\alpha\beta}^2$  的正态分布。

2.  $\alpha\beta_{ij}$  间相互独立。

3.  $\alpha_i$ ,  $\beta_j$ ,  $\alpha\beta_{ij}$  和  $\varepsilon_{ijk}$  之间相互独立。

固定效应和随机效应模型相应的 AOV 表见表 17.9。

表 17.9 每个试验单元处有  $n$  次观测的  $a \times b$  析因试验的 AOV 表

| 来源 | SS   | df               | MS   | EMS                         |  |
|----|------|------------------|------|-----------------------------|--|
|    |      |                  |      | 固定效应                        | 随机效应   |
| A  | SSA  | $a - 1$          | MSA  | $\sigma_e^2 + nb\theta_A$   | $\sigma_e^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_\alpha^2$ |
| B  | SSB  | $b - 1$          | MSB  | $\sigma_e^2 + na\theta_B$   | $\sigma_e^2 + n\sigma_{\alpha\beta}^2 + an\sigma_\beta^2$  |
| AB | SSAB | $(a - 1)(b - 1)$ | MSAB | $\sigma_e^2 + n\theta_{AB}$ | $\sigma_e^2 + n\sigma_{\alpha\beta}^2$                     |
| 误差 | SSE  | $ab(n - 1)$      | MSE  | $\sigma_e^2$                | $\sigma_e^2$   |
| 总和 | TSS  | $abn - 1$        |      |                             |  |

表 17.10 给出了这两种模型中 AB 交互效应平方和的适当的检验。

表 17.10 固定效应、随机效应模型的交互效应检验的比较

| 固定效应模型   | 随机效应模型                            |
|--|-----------------------------------|
| $H_0: \alpha\beta_{11} = \alpha\beta_{12} = \cdots = \alpha\beta_{ab} = 0$ | $H_0: \sigma_{\alpha\beta}^2 = 0$ |
| $H_a$ : 至少有一个 $\alpha\beta_{ij}$ 与其他的不同                                    | $H_a: \sigma_{\alpha\beta}^2 > 0$ |
| T. S.: $F = \frac{MSAB}{MSE}$  | T. S.: $F = \frac{MSAB}{MSE}$     |
| R. R.: 根据 $df_1 = (a - 1)$ , $df_2 = ab(n - 1)$                            | R. R.: 同左                         |

不像单因子试验和没有重复的两因子试验,此处固定效应与随机效应模型中检验主效应的统计量并不相同。而且,对随机效应模型来说,即使 AB 交互效应( $\sigma_{\alpha\beta}^2$ )的检验结果是显著的,对于  $\sigma_\alpha^2$  和  $\sigma_\beta^2$  的检验也可继续进行。我们可以回想前面讲到的固定效应模型中,当交互效应显著时,只有在轮廓图显示交互效应是“有序的”时,检验主效应才是有意义的。对随机效应模型,我们感兴趣的是确定响应变量  $y$  变异性的各种来源(如  $\sigma_{\alpha\beta}^2$ ,  $\sigma_\alpha^2$  和  $\sigma_\beta^2$ )。即使当  $\sigma_{\alpha\beta}^2$  看来比 0 大时,检验  $\sigma_\alpha^2$  和  $\sigma_\beta^2$  也是有意义的。在固定效应模型中,当 AB 的交互效应不显著时,我们可用如下统计量分别检验因子 A 和因子 B 的主效应:

$$F = \frac{MSA}{MSE} \text{ 和 } F = \frac{MSB}{MSE}$$

就像从表 17.9 看到的期望均方列那样,无论对  $H_0: \sigma_{\alpha\beta}^2 = 0$  检验结果如何,我们可以利用表 17.11 中的检验步骤对  $\sigma_\alpha^2$  分量和  $\sigma_\beta^2$  分量做  $F$  检验,要注意检验统计量与固定效应情况的不同,固定效应时所有  $F$  统计量的分母都是 MSE。

表 17.11 有重复的  $a \times b$  析因试验的检验:随机效应模型

| 因子 A                                    | 因子 B                                    |
|---|---|
| $H_0: \sigma_\alpha^2 = 0$              | $H_0: \sigma_\beta^2 = 0$               |
| $H_a: \sigma_\alpha^2 > 0$              | $H_a: \sigma_\beta^2 > 0$               |
| T.S.: $F = \frac{MSA}{MSAB}$            | T.S.: $F = \frac{MSB}{MSAB}$            |
| R.R.: 基于                                | R.R.: 基于                                |
| $df_1 = (a - 1), df_2 = (a - 1)(b - 1)$ | $df_1 = (b - 1), df_2 = (a - 1)(b - 1)$ |

在很多因子具有随机效应的试验中,我们需要估计方差分量  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$ ,  $\sigma_{\alpha\beta}^2$  及  $\sigma_\epsilon^2$ 。我们还可用 AOV 矩匹配估计,即将样本均方用 AOV 中的期望均方做匹配得到每个方差分量的估计。利用表 17.9 中的各个 MS 和 EMS 我们得到:

$$\hat{\sigma}_\epsilon^2 = MSE$$

$$\hat{\sigma}_{\alpha\beta}^2 = (MSAB - MSE)/n$$

$$\hat{\sigma}_\beta^2 = (MSB - MSAB)/an$$

及

$$\hat{\sigma}_\alpha^2 = (MSA - MSAB)/bn$$

从两个因子的水平都是随机选取的随机效应模型,我们有:

$$E(y_{ijk}) = \mu \text{ 和 } \sigma_y^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma_\epsilon^2$$

于是有,  $\hat{\sigma}_y^2 = \hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}_\epsilon^2$ 。然后我们可以确定四种变异来源在总变异性  $\hat{\sigma}_y^2$  中

所占的比例:因子 A, 因子 B, 交互效应和试验误差。见表 17.12:

表 17.12 响应变量的总变异性的按比例分配

| 方差来源    | 估计值  | 占总方差的比例   |
|---------|--|---|
| 因子 A    | $\hat{\sigma}_\alpha^2 = (\text{MSA} - \text{MSAB})/bn$  | $\hat{\sigma}_\alpha^2/\hat{\sigma}_y^2$        |
| 因子 B    | $\hat{\sigma}_\beta^2 = (\text{MSB} - \text{MSAB})/an$   | $\hat{\sigma}_\beta^2/\hat{\sigma}_y^2$         |
| AB 交互效应 | $\hat{\sigma}_{\alpha\beta}^2 = (\text{MSAB} - \text{MSE})/n$  | $\hat{\sigma}_{\alpha\beta}^2/\hat{\sigma}_y^2$ |
| 误差      | $\hat{\sigma}_\epsilon^2 = \text{MSE}$   | $\hat{\sigma}_\epsilon^2/\hat{\sigma}_y^2$      |
| 总计      | $\hat{\sigma}_y^2 = \hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + \hat{\sigma}_\epsilon^2$ |   |

研究者可能还想估计响应变量的均值  $\mu$ 。我们有  $\mu$  的点估计和估计的标准误差如下:

$$\hat{\mu} = \bar{y}_{...} \text{ 和 } \text{SE}(\hat{\mu}) = \sqrt{(\text{MSA} + \text{MSB} - \text{MSAB})/abn}$$

我们可建立如下  $\mu$  的  $100(1 - \alpha)\%$  置信区间:

$$\bar{y}_{...} \pm t_{\alpha/2, \text{df}_{\text{Approx.}}} \sqrt{(\text{MSA} + \text{MSB} - \text{MSAB})/abn}$$

其中  $t$  检验统计量的自由度是用 Satterthwaith 近似得到的,

$$\text{df}_{\text{Approx.}} = \frac{(\text{MSA} + \text{MSB} - \text{MSAB})^2}{(\text{MSA})^2/(a-1) + (\text{MSB})^2/(b-1) + (\text{MSAB})^2/(a-1)(b-1)}$$

很多情况下这个值并不是整数, 我们取小于或等于  $\text{df}_{\text{Approx.}}$  的最大整数。

在一些试验中, 某些方差分量的估计值可能是负数。当然定义中的方差分量必须是个非负数, 所以当样本估计值是负数时我们必须另想办法。

**A1.** 我们可令估计值等于 0 并用 0 作为方差分量的估计值。然而此时估计值将不再是方差分量的无偏估计。

**A2.** 方差分量的估计值为负可能说明模型中某些因素不适合于此试验。试验可能需要更复杂的模型。

**A3.** 方差分量还有别的估计方法, 但在数学知识上超过了本书的范围。像 REML 和 MINIQUE 等方法现在可用 SAS 软件完成。然而我们还是应该仔细检查数据, 因为负的方差分量估计值通常说明模型不合适。

### 例 17.2

一个消费者用品代理商行想要评测一种补品中钙水平的测定是否准确。现在有很多可参加试验的实验室及很多种对钙的化学化验方法。代理商行随机选出了三个实验室和三种化验用于研究, 每个实验室都要使用全这三种化验方法。现在准备

好了十八个含 10mg 钙的样品, 每个化验 - 实验室组合处都随机分配到两个样品。钙含量的测定结果见下(括号中的数是化验 - 实验室组合处的平均值):

| 化验方法   | 实验室    |         |         | 化验方法均值 |
|--------|--------|---------|---------|--------|
|        | 1      | 2       | 3       |        |
| 1      | 10.9   | 10.5    | 9.7     | 10.3   |
|        | 10.9   | 9.8     | 10.0    |        |
|        | (10.9) | (10.15) | (9.85)  |        |
| 2      | 11.3   | 9.4     | 8.8     | 10.1   |
|        | 11.7   | 10.2    | 9.2     |        |
|        | (11.5) | (9.8)   | (9.0)   |        |
| 3      | 11.8   | 10.0    | 10.4    | 10.8   |
|        | 11.2   | 10.7    | 10.7    |        |
|        | (11.5) | (10.35) | (10.55) |        |
| 实验室平均值 | 11.3   | 10.1    | 9.8     | 10.4   |

- 对此试验进行方差分析, 所有检验取  $\alpha = 0.05$ 。
- 估计所有的方差分量并确定它们占总变异性的比例。
- 估计所有实验室和化验方法对应的平均含钙水平。

**解答** 利用第十五章的公式, 我们算得平方和如下:

$$\begin{aligned} \text{TSS} &= \sum_{ijk} (y_{ijk} - \bar{y}_{...})^2 = (10.9 - 10.4)^2 + (10.9 - 10.4)^2 + \cdots + (10.7 - 10.4)^2 \\ &= 12.00 \end{aligned}$$

$$\begin{aligned} \text{SSA} &= \sum_i 6(\bar{y}_{i..} - \bar{y}_{...})^2 = 6\{(10.3 - 10.4)^2 + (10.1 - 10.4)^2 + (10.8 - 10.4)^2\} \\ &= 1.56 \end{aligned}$$

$$\begin{aligned} \text{SSL} &= \sum_j 6(\bar{y}_{.j.} - \bar{y}_{...})^2 = 6\{(11.3 - 10.4)^2 + (10.1 - 10.4)^2 + (9.8 - 10.4)^2\} \\ &= 7.56 \end{aligned}$$

$$\begin{aligned} \text{SSAL} &= \sum_{ij} 2(\bar{y}_{ij.} - \bar{y}_{...})^2 = 2\{(10.9 - 10.4)^2 + (10.15 - 10.4)^2 + (9.85 - 10.4)^2 \\ &\quad + \cdots + (10.55 - 10.4)^2\} = 1.64 \end{aligned}$$

$$\text{SSE} = \text{TSS} - \text{SSA} - \text{SSL} - \text{SSAL} = 12.00 - 1.56 - 7.56 - 1.64 = 1.24$$

用方差分析表将结果总结如下:

| 来源       | SS    | df | MS     | EMS   |
|----------|-------|----|--------|---|
| 化验       | 1.56  | 2  | 0.78   | $\sigma_e^2 + 2\sigma_{\alpha\beta}^2 + 6\sigma_\alpha^2$ |
| 实验室      | 7.56  | 2  | 3.78   | $\sigma_e^2 + 2\sigma_{\alpha\beta}^2 + 6\sigma_\beta^2$  |
| 化验 * 实验室 | 1.64  | 4  | 0.41   | $\sigma_e^2 + 2\sigma_{\alpha\beta}^2$                    |
| 误差       | 1.24  | 9  | 0.1378 | $\sigma_e^2$  |
| 总计       | 12.00 | 17 |        |   |

利用上面 AOV 表提供的结果,可进行适当的统计检验。对 AB 的交互效应我们有

$$H_0: \sigma_{\alpha\beta}^2 = 0$$

$$H_a: \sigma_{\alpha\beta}^2 > 0$$

$$\text{T.S.: } F = \frac{MS_{AB}}{MSE} = \frac{0.41}{0.1378} = 2.98$$

R.R.: 对  $\alpha = 0.05$ , 如果  $F$  值超过 3.63 就拒绝  $H_0$ , 临界值取  $\alpha = 0.05$ ,  $df_1 = 4$  和  $df_2 = 9$ 。

结论: 没有足够理由拒绝  $H_0$ 。因子 A、B 的水平之间看来没有显著的交互效应。

对因子 B 我们有

$$H_0: \sigma_\beta^2 = 0$$

$$H_a: \sigma_\beta^2 > 0$$

$$\text{T.S.: } F = \frac{MS_B}{MS_{AB}} = \frac{3.78}{0.41} = 9.22$$

R.R.: 对  $\alpha = 0.05$ , 如果  $F$  值超过 6.94 就拒绝  $H_0$ , 临界值取  $\alpha = 0.05$ ,  $df_1 = 2$  和  $df_2 = 4$ 。

结论: 因为  $F$  的观测值远大于 6.94, 我们拒绝  $H_0$  并认为实验室与实验室之间对钙含量的测定有显著差别。

对因子 A 的检验如下:

$$H_0: \sigma_\alpha^2 = 0$$

$$H_a: \sigma_\alpha^2 > 0$$

$$\text{T.S.: } F = \frac{MS_A}{MS_{AB}} = \frac{0.78}{0.41} = 1.90$$

R.R.: 对  $\alpha = 0.05$ , 如果  $F$  值超过 3.63 就拒绝  $H_0$ , 临界值取  $\alpha = 0.05$ ,  $df_1 = 2$  和  $df_2 = 4$ 。

结论: 没有足够理由认为化验方法与化验方法间有显著的差异。

下面我们估计方差分量。利用方差分析表中的 MS 和 EMS, 我们有

$$\hat{\sigma}_e^2 = MSE = 0.1378$$

$$\hat{\sigma}_{\alpha\beta}^2 = (\text{MSAB} - \text{MSE})/n = (0.41 - 0.1378)/2 = 0.1361$$

$$\hat{\sigma}_{\beta}^2 = (\text{MSB} - \text{MSAB})/an = (3.78 - 0.41)/6 = 0.5617$$

及

$$\hat{\sigma}_{\alpha}^2 = (\text{MSA} - \text{MSAB})/bn = (0.78 - 0.41)/6 = 0.0617$$

同样,对于有两个因子水平是随机选取的随机效应模型我们有

$$E(y_{ijk}) = \mu \text{ 和 } \sigma_y^2 = \sigma_{\alpha}^2 + \sigma_{\beta}^2 + \sigma_{\alpha\beta}^2 + \sigma_{\epsilon}^2$$

所以

$$\sigma_y^2 = 0.0617 + 0.5617 + 0.1361 + 0.1378 = 0.8973$$

我们可以把总变异性  $\sigma_y^2$  按比例分配给变异性的四种来源:化验,实验室,交互效应和试验误差。

| 方差来源 | 估计值    | 占总方差的比例                 |
|------|--------|-------------------------|
| 化验   | 0.0617 | $0.0617/0.8973 = 0.069$ |
| 实验室  | 0.5617 | $0.5617/0.8973 = 0.626$ |
| 交互效应 | 0.1361 | $0.1361/0.8973 = 0.152$ |
| 误差   | 0.1378 | $0.1378/0.8973 = 0.156$ |
| 总计   | 0.8973 | 1.0                     |

因为样品钙水平的测定结果有显著的差异,所以研究者就不再关心总均值水平  $\mu$  的估计。然而为说明这个方法,我们将继续进行。得到  $\mu$  的点估计和标准误差为:

$$\hat{\mu} = \bar{y}_{...} = 10.4 \text{ 和 } \text{SE}(\hat{\mu}) = \sqrt{(\text{MSA} + \text{MSB} - \text{MSAB})/abn} = 0.4802$$

$\mu$  的  $100(1-\alpha)\%$  置信区间为:

$$\bar{y}_{...} \pm t_{\alpha/2, \text{df}_{\text{Approx.}}} \sqrt{(\text{MSA} + \text{MSB} - \text{MSAB})/abn}$$

或

$$10.4 \pm (t_{\alpha/2, \text{df}_{\text{Approx.}}})(0.4802)$$

其中  $t$  检验统计量的自由度是用 Satterthwaith 近似得到的,

$$\begin{aligned} \text{df}_{\text{Approx.}} &= \frac{(\text{MSA} + \text{MSB} - \text{MSAB})^2}{(\text{MSA})^2/(a-1) + (\text{MSB})^2/(b-1) + (\text{MSAB})^2/((a-1)(b-1))} \\ &= \frac{(4.15)^2}{(0.78)^2/2 + (3.78)^2/2 + (0.41)^2/4} = 2.3 \end{aligned}$$

取小于或等于  $\text{df}_{\text{Approx.}}$  的最大整数,则  $\text{df}_{\text{Approx.}} = 2$ 。因为  $t_{.025,2} = 4.303$ , 我们得到所有的实验室和化验方法对应的含钙水平平均值的 95% 置信区间为:

$$10.4 \pm (4.303)(0.4802)$$

$$10.4 \pm 2.1 \text{ 即 } (8.3, 12.5)$$

本节中,我们比较了完全随机化设计与每个单元处有  $n$  次观测值的  $a \times b$  析因试验的随机效应模型和固定效应模型。这个研究并不够彻底,但已经说明了除固定效应模型外还有别的模型。对随机效应模型更详细的研究应包括含更多因子(多于两个)的析因试验和 17.6 节中的**套抽样试验**。对于后一种试验设计,因子 B 的水平是嵌套在因子 A 的水平中(而不是与之相互交叉)的。比如,要考察一种化学制品的效力,我们先从不同的生产厂取样,再从厂内产品批中取,再在每批产品中取。注意因子“批”与因子“生产厂”并不是相互交叉的,因为比方说,厂 1 的第一批和厂 2 的第一批是不同的。

在第 17.4 节中,我们将把本节的结果扩充到  $a \times b$  析因试验的混合效应模型。

### 练习

### 应用

**17.3(化学)** 现在要进行一项研究来检验对齿菌斑 DNA 含量的化学分析的变异性。研究中选的两个主要来源是不同的分析员和不同的受试者。研究人员从一大群可以参加研究的分析员中随机选了三个,同样选了三个女性受试者(18~20 岁)。从每个受试者的完整齿列上刮下齿菌斑并分为三份。每个分析员都拿到每个受试者的一份没有标名字的样品并对它们进行 DNA 含量的分析(以微克计)。数据如下。

| 分析员 | 受试者  |       |      | 平均值   |
|-----|------|-------|------|-------|
|     | 1    | 2     | 3    |       |
| 1   | 13.2 | 10.6  | 8.5  | 10.77 |
| 2   | 12.5 | 9.6   | 7.9  | 10.00 |
| 3   | 13.0 | 9.9   | 8.3  | 10.40 |
| 平均值 | 12.9 | 10.03 | 8.23 | 10.39 |

- 写出一个合适的统计模型,并辨别模型中所有项的意义。
- 写出期望均方。

**17.4** 根据例 17.3,做方差分析,所有检验取  $\alpha = 0.05$ 。

**17.5(商业)** 一个市场调查公司想要研究一种新的促销活动对于提高 D-cell 电池品牌的效应。研究在随机选取的四个标准的大城市的统计区域(SMASs)进行,每个城市内随机选取了三个连锁销售店(从一大堆杂货店、药店和百货公司的名单中选取),记录下指定区域在促销活动后随机选的两周内的销量(以美元计),数据如下。



| 连锁销售店 | SMSA |     |     |     |
|-------|------|-----|-----|-----|
|       | 1    | 2   | 3   | 4   |
| 1     | 98   | 149 | 79  | 340 |
|       | 112  | 126 | 61  | 302 |
| 2     | 87   | 96  | 119 | 125 |
|       | 75   | 138 | 104 | 133 |
| 3     | 140  | 159 | 169 | 460 |
|       | 190  | 185 | 150 | 420 |

- 写出适当的线性统计模型, 列出假定并指出各项的含义。
- 进行方差分析, 写出期望均方, 取  $\alpha = 0.05$ 。
- 将销量的总变异性分解为各个来源带来的变异性。

## 17.4 混合效应模型

在 17.3 节中, 我们对随机化区组设计和完全随机化设计中一般  $a \times b$  吸引试验的固定效应模型和随机效应模型的方差分析表做了比较。假设对于同样的这些试验设计, 我们的模型是一个因子效应固定另一个随机的混合效应模型。例如, 17.3 节中我们考虑了一个试验, 这个试验是检验不同的受试者和不同的分析员对齿菌斑 DNA 成分的效应。如果三个受试者是随机选择而三个分析员是所关心的仅有的分析员, 那么我们就有了一个随机化区组设计的混合模型, 其中分析员是固定的, 受试者是随机的。

让我们考虑完全随机化设计中一般  $a \times b$  析因试验的混合模型。除假定不同外, 模型与 17.3 中的一样:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

其中因子 A 的水平固定, 因子 B 的水平随机选取; 假定如下:

- $\mu$  是未知的总响应均值。
- $\alpha_i$  是因子 A 的第  $i$  个水平对应的固定效应,  $\sum_i \alpha_i = 0$ 。
- $\beta_j$  是因子 B 的第  $j$  个水平对应的随机效应。 $\beta_j$  独立同服从均值为 0 方差  $\sigma_\beta^2$  的正态分布。

4.  $\alpha\beta_{ij}$  是因子 A 的第  $i$  个水平和因子 B 的第  $j$  个水平的交互效应,  $\sum_i \alpha\beta_{ij} = 0$ , 对任意  $j = 1, \dots, b$ ,  $\alpha\beta_{ij}$  服从均值为 0 方差  $\sigma_{\alpha\beta}^2$  的正态分布。对所有  $j \neq j'$ ,  $\alpha\beta_{ij}$  和  $\alpha\beta_{ij'}$  是独立的。

5.  $\beta_j$ ,  $\alpha\beta_{ij}$  和  $\varepsilon_{ijk}$  间独立。

因为  $\sum_i \alpha\beta_{ij} = 0$ , 对任意  $j = 1, \dots, b$ , 所以对每个  $j$ ,  $\alpha\beta_{1j}, \alpha\beta_{2j}, \dots, \alpha\beta_{aj}$  是相关的, 而对所有的  $j \neq j'$ ,  $\alpha\beta_{ij}$  和  $\alpha\beta_{ij'}$  是独立的。即, 因子 B 的同一个水平对应的交互效应是相关的, 而因子 B 不同水平对应的交互效应之间却是独立的。

根据假定, 有重复的两因子试验的固定、随机和混合模型的方差分析表如下表 17.13。

表 17.13  $a \times b$  析因试验的 AOV 表, 每个单元处有  $n$  次观测

| 来源 | SS   | df               | MS   | EMS                                   |  |   |
|----|------|------------------|------|---------------------------------------|--|---|
|    |      |                  |      | 固定效应                                  | 随机效应   | 混合效应<br>A 固定, B 随机  |
| A  | SSA  | $a - 1$          | MSA  | $\sigma_\varepsilon^2 + bn\theta_A$   | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2 + bn\sigma_\alpha^2$ | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2 + bn\theta_A$ |
| B  | SSB  | $b - 1$          | MSB  | $\sigma_\varepsilon^2 + an\theta_B$   | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2 + an\sigma_\beta^2$  | $\sigma_\varepsilon^2 + an\sigma_\beta^2$                     |
| AB | SSAB | $(a - 1)(b - 1)$ | MSAB | $\sigma_\varepsilon^2 + n\theta_{AB}$ | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2$                     | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2$              |
| 误差 | SSE  | $ab(n - 1)$      | MSE  | $\sigma_\varepsilon^2$                | $\sigma_\varepsilon^2$   | $\sigma_\varepsilon^2$  |
| 总计 | TSS  | $nab - 1$        |      |                                       |  |   |

表 17.13 中的期望均方哪一列对于进行适当的显著性检验是很有帮助的。混合模型中对  $\sigma_{\alpha\beta}^2$  的检验和随机效应模型中一样。

对  $\sigma_{\alpha\beta}^2$  的检验:

$$H_0: \sigma_{\alpha\beta}^2 = 0$$

$$H_a: \sigma_{\alpha\beta}^2 > 0$$

$$\text{T.S.: } F = \frac{\text{MSAB}}{\text{MSE}}$$

$$\text{R.R.: 基于 } df_1 = (a - 1)(b - 1), df_2 = ab(n - 1)$$

无论对  $\sigma_{\alpha\beta}^2$  的检验结果如何, 我们都可以继续对因子 A 和 B 作如下检验, 检验也要用到表 17.13 中期望均方列中的项。对因子 A 的检验如下:

$$H_0: \alpha_1 = \alpha_2 = \dots = 0$$

$H_a$ : 至少有一个  $\alpha$  与别不同

$$\text{T.S.: } F = \frac{\text{MSA}}{\text{MSAB}}$$

$$\text{R.R.: 取 } df_1 = (a - 1), df_2 = (a - 1)(b - 1)$$

对因子 B 的检验为:

$$H_0: \sigma_{\beta}^2 = 0$$

$$H_a: \sigma_{\beta}^2 > 0$$

$$\text{T.S.: } F = \frac{\text{MSB}}{\text{MSE}}$$

$$\text{R.R.: 取 } df_1 = (b - 1), df_2 = ab(n - 1)$$

$a \times b$  因子试验的混合效应模型的方差分析过程同样也适用于随机化区组设计, 其中处理是固定的, 区组则假设为随机的, 并且每个区组和处理处有  $n$  次观测值。我们将用下面的例子说明随机模型。

### 例 17.3

现在要设计一项研究来评测两种不同的防晒剂( $s_1$  和  $s_2$ ) 在保护人们肌肤暴露在阳光中时不受到灼伤或晒黑方面的效果。由 40 个受试者(20~25 岁)组成的随机样本同意参加这项研究。在每个受试者背部肩下方腰背上方处划分出 1 平方英寸的范围。各二十名受试者被随机分到一种防晒剂。在使用一定量的指定的防晒剂之前, 先记录下指定范围内肌肤的颜色读数; 同样在使用防晒剂并暴露在阳光中 2 小时后又记录下读数。公司注意到肤色的测量结果差异极大, 并想要估计一下由于读取读数的技术员造成的变异性, 所以公司从它的全世界范围的员工中随机选了 10 名技术员参加这项研究。四个受试者, 其中两个用  $s_1$  两个用  $s_2$ , 被随机分配给一个技术员作考察。下面表 17.14 记录的数据是受试者的两次读数的差值(晒后减去晒前)。响应变量值越高说明灼伤程度越深。

表 17.14 例 17.13 防晒剂试验的数据

| 防晒剂(A) | 技术员     |         |        |         |         |
|--------|---------|---------|--------|---------|---------|
|        | 1       | 2       | 3      | 4       | 5       |
| $s_1$  | 8.2     | 3.6     | 10.7   | 3.9     | 12.9    |
|        | 7.6     | 3.5     | 10.3   | 4.4     | 12.1    |
| 平均值    | (7.9)   | (3.55)  | (10.5) | (4.15)  | (12.5)  |
| $s_2$  | 6.1     | 4.3     | 9.6    | 2.3     | 12.4    |
|        | 6.8     | 4.7     | 9.2    | 2.5     | 12.8    |
| 平均值    | (6.45)  | (4.5)   | (9.4)  | (2.4)   | (12.6)  |
| 平均值    | (7.175) | (4.025) | (9.95) | (3.275) | (12.55) |

续表

| 防晒剂(A) | 技术员    |         |         |         |         | 平均值   |
|--------|--------|---------|---------|---------|---------|-------|
|        | 6      | 7       | 8       | 9       | 10      |       |
| $s_1$  | 5.5    | 9.1     | 13.7    | 8.1     | 2.5     | 7.82  |
|        | 5.9    | 9.7     | 13.2    | 8.7     | 2.8     |       |
| 平均值    | (5.7)  | (9.4)   | (13.45) | (8.4)   | (2.65)  |       |
| $s_2$  | 4.8    | 8.3     | 12.9    | 8.0     | 2.1     | 7.15  |
|        | 4.0    | 8.6     | 13.6    | 7.5     | 2.5     |       |
| 平均值    | (4.4)  | (8.45)  | (13.25) | (7.75)  | (2.3)   |       |
| 平均值    | (5.05) | (8.925) | (13.35) | (8.075) | (2.475) | 7.485 |

这是个完全随机化设计, 有两个因子——有两个固定水平的防晒剂种类(A)和随机选取了 10 个水平的技术员, 每个防晒剂-技术员组合处有两个受试者。分析数据以确定防晒剂之间和技术员之间的差异。

**解答** 我们可用下面的公式计算 AOV 表中各种变异性来源对应的平方和。

$$\begin{aligned} \text{TSS} &= \sum_{ijk} (y_{ijk} - \bar{y}_{...})^2 = (8.2 - 7.485)^2 + (7.6 - 7.485)^2 + \dots \\ &\quad + (2.5 - 7.485)^2 = 530.59 \end{aligned}$$

$$\text{SSA} = \sum_i 20(\bar{y}_{i..} - \bar{y}_{...})^2 = 20\{(7.82 - 7.485)^2 + (7.15 - 7.485)^2\} = 4.49$$

$$\begin{aligned} \text{SSB} &= \sum_j 4(\bar{y}_{.j.} - \bar{y}_{...})^2 = 4\{(7.175 - 7.485)^2 + (4.025 - 7.485)^2 + \dots \\ &\quad + (2.475 - 7.485)^2\} = 517.49 \end{aligned}$$

$$\begin{aligned} \text{SSAB} &= \sum_{ij} 2(\bar{y}_{ij.} - \bar{y}_{...})^2 - \text{SSA} - \text{SSB} = 2\{(7.9 - 7.485)^2 + (3.55 - 7.485)^2 \\ &\quad + (10.5 - 7.485)^2 + \dots + (2.3 - 7.485)^2\} - 4.49 - 517.49 = 5.97 \end{aligned}$$

$$\text{SSE} = \text{TSS} - \text{SSA} - \text{SSB} - \text{SSAB} = 530.59 - 4.49 - 517.49 - 5.97 = 2.64$$

将  $a=2$ ,  $b=10$  和  $n=2$  代入与表 17.13 相似的 AOV 表中就得到了表 17.15 的结果。

表 17.15 例 17.3 数据的 AOV 表

| 来源 | SS     | df | MS    | EMS<br>固定模型   |
|----|--------|----|-------|---|
| A  | 4.49   | 1  | 4.49  | $\sigma_e^2 + 2\sigma_{\alpha\beta}^2 + 20\theta_A$ |
| B  | 517.49 | 9  | 57.50 | $\sigma_e^2 + 4\sigma_\beta^2$                      |
| AB | 5.97   | 9  | 0.66  | $\sigma_e^2 + 2\sigma_{\alpha\beta}^2$              |
| 误差 | 2.64   | 20 | 0.13  | $\sigma_e^2$  |
| 总计 | 530.59 | 39 |       |   |

对随机分量  $\alpha\beta_{ij}$  的检验如下:

$$H_0: \sigma_{\alpha\beta}^2 = 0$$

$$H_a: \sigma_{\alpha\beta}^2 > 0$$

$$\text{T.S.: } F = \frac{\text{MSAB}}{\text{MSE}} = \frac{0.66}{0.13} = 5.08$$

R.R.: 对  $\alpha = 0.05$ , 如果计算出的  $F$  值大于 2.39 (附表 8 中  $\alpha = 0.05$ ,  $df_1 = 9$ ,  $df_2 = 20$  对应的值), 我们就拒绝  $H_0$ 。

结论: 因为 5.08 大于 2.39, 我们拒绝  $H_0$ , 结论为  $\sigma_{\alpha\beta}^2 > 0$ ; 即, A (遮光剂) 的第  $i$  个水平和 B (技术员) 的第  $j$  个水平组合是随机变异很重要的来源。从中我们可推断出, 由技术员造成的肤色测定的变异对于这两种遮光剂是不同的。

接下来评估技术员的效应。

$$H_0: \sigma_\beta^2 = 0$$

$$H_a: \sigma_\beta^2 > 0$$

$$\text{T.S.: } F = \frac{\text{MSB}}{\text{MSE}} = \frac{57.50}{0.13} = 442.31$$

R.R.: 对  $\alpha = 0.05$ , 如果计算出的  $F$  值大于 2.39 (附表 8 中  $\alpha = 0.05$ ,  $df_1 = 9$ ,  $df_2 = 20$  对应的值), 我们就拒绝  $H_0$ 。

结论: 因为 442.31 大于 2.39, 拒绝  $H_0$  并认为  $\sigma_\beta^2 > 0$ 。所以技术员与技术员之间的差异是随机变异的重要来源。

对因子 A 有:

$$H_0: \alpha_1 = \alpha_2 = 0$$

$$H_a: \alpha_1 \neq \alpha_2$$

$$\text{T.S.: } F = \frac{\text{MSA}}{\text{MSAB}} = \frac{4.49}{0.66} = 6.80$$

R.R.: 对  $\alpha = 0.05$ , 如果计算出的  $F$  值大于 5.12 (附表 8 中  $\alpha = 0.05$ ,  $df_1 = 9$ ,  $df_2 = 20$  对应的值), 我们就拒绝  $H_0$ 。

结论:因为  $6.80 > 5.12$ , 拒绝  $H_0$  并认为两种遮光剂的响应均值(晒后减晒前)不同。因为  $\bar{y}_{s1} = 7.82$ ,  $\bar{y}_{s2} = 7.15$ , 我们认为平均说来  $s_2$  比  $s_1$  能提供更多保护。但是就像前面表明的, 技术员和技术员与遮光剂的组合也是变异性的重要来源。下面我们分析案例中的数据。

#### 案例数据分析: 扩张接缝处的压降

本研究的目的是确定电涡轮机上扩张接缝处的压降是否与气体温度有关。而且研究者还想估计一下由不同种类的测压计造成的读数差异, 以及确定读数的变化是否与不同的气体温度相符合。在图 17.1 中, 我们观察到当温度从  $15^\circ\text{C}$  升到  $25^\circ\text{C}$  时压降有轻微的升高; 而在温度进一步由  $25^\circ\text{C}$  升到  $35^\circ\text{C}$  时压降却相应下降。四种测压计测得的压降与这三个温度相当的一致, 只除了在  $25^\circ\text{C}$  时测压计 G1 测得的压降平均值比另三种测得的更高。下面 12 个温度-测压计组合处的均值和标准差表显示标准差是相当稳定的, 除了测压计 G1 在  $25^\circ\text{C}$  时压降的均值比其他 11 种温度-测压计处理的压降均值要高。

| 压降读数的均值和标准差 |       |       |       |       |       |      |      |      |
|-------------|-------|-------|-------|-------|-------|------|------|------|
| 温度          | 均 值   |       |       |       | 标 准 差 |      |      |      |
|             | G1    | G2    | G3    | G4    | G1    | G2   | G3   | G4   |
| 15          | 41.17 | 38.50 | 38.67 | 43.00 | 3.31  | 3.62 | 3.72 | 4.69 |
| 25          | 61.33 | 46.67 | 45.33 | 41.33 | 4.27  | 3.44 | 2.07 | 4.97 |
| 35          | 39.00 | 41.17 | 37.83 | 42.67 | 4.69  | 2.79 | 3.31 | 4.18 |

因为这四种测压计是从这个公司使用的所有测压计中随机选出的, 所以我们要估计上面表中和图 17.1 中观察到的模式相对于从中选出测压计的总体来说是否是显著的差异。此外, 我们还要确定  $15^\circ\text{C}$  到  $35^\circ\text{C}$  的各温度范围之间的压降均值是否有显著差异。此处温度因子是固定的, 测压计因子是随机的。用下面的模型来拟合数据:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$$

其中  $y_{ijk}$  是温度  $i$  下用测压计  $j$  时第  $k$  次重复测量的压降。在进行假设检验和构建置信区间之前, 我们要考察推断成立所必须满足的试验条件是否恰当。检查下面的残差图能帮助我们考查模型条件的正确性。下面是拟合模型残差分析的计算机输出结果。

# Residual Analysis for Case Study

Univariate Procedure

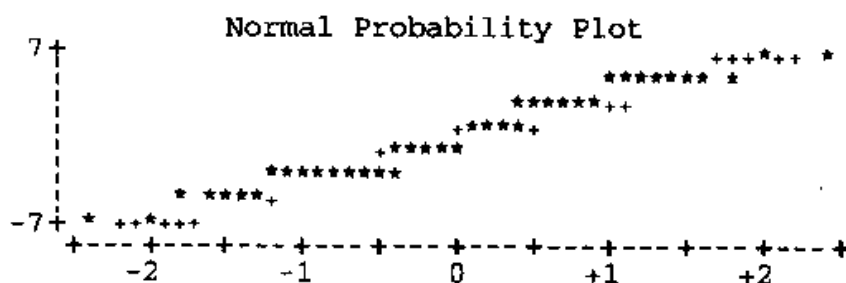
Variable = RESIDUALS

| Moments  |          |          |          |
|----------|----------|----------|----------|
| N        | 72       | Sum Wgts | 72       |
| Mean     | 0        | Sum      | 0        |
| Std Dev  | 3.532545 | Variance | 12.47887 |
| Skewness | 0.014971 | Kurtosis | -0.87963 |
| W;Normal | 0.959963 | Pr<W     | 0.0655   |

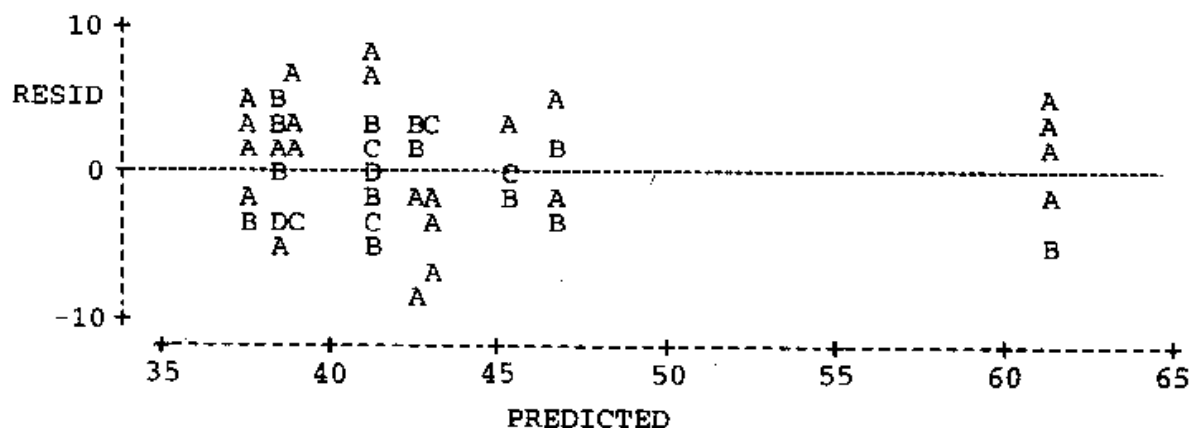
Variable=RESIDUALS

| Stem         | Leaf           | #  | Boxplot |
|--------------|----------------|----|---------|
| 6            | 07             | 2  |         |
| 4            | 000233578      | 9  |         |
| 2            | 03338023335777 | 14 |         |
| 0            | 5778823378     | 10 |         |
| -0           | 877322533222   | 12 |         |
| -2           | 87777533087330 | 14 |         |
| -4           | 325332000      | 9  |         |
| -6           | 70             | 2  |         |
| -----+-----+ |                |    |         |

Variable=RESIDUALS



Plot of RESIDUALS★PREDICTED Legend: A = 1 obs, B = 2 obs, etc.



残差的盒形图和茎叶图显示没有极端值。正态概率图显示一些残差稍微偏离拟合的线条。但是正态性检验的  $p$  值为 .0655, 所以没有显著证据说残差不服从正态分布。残差对预测值的图显示关于残差的等方差假定并没有被违背, 因为在预测值附近残差的分布相当稳定。所以由数据看来, 正态性和等方差等条件是满足的。测压计是从测压计总体中随机选出这个条件以及试验进行总遵循响应之间独立的方式这个条件将通过与和此试验有关的研究者进行讨论来检查。现在我们给出 AOV 表和有关的假设检验。

案例的 AOV 表

| 来源     | SS       | df | MS     | EMS   | $F$   | $p$ 值  |
|--------|----------|----|--------|---|-------|--------|
| 温度(T)  | 1,133.78 | 2  | 556.89 | $\sigma_e^2 + 6\sigma_{\alpha\beta}^2 + 24\theta_A$ | 3.02  | .1238  |
| 测压计(G) | 437.22   | 3  | 145.74 | $\sigma_e^2 + 18\sigma_{\beta}^2$                   | 9.87  | <.0001 |
| T * G  | 1,106.78 | 6  | 184.46 | $\sigma_e^2 + 6\sigma_{\alpha\beta}^2$              | 12.49 | <.0001 |
| 误差     | 886.00   | 60 | 14.77  | $\sigma_e^2$  |       |        |
| 总计     | 3,563.78 | 71 |        |   |       |        |

对数据拟合上述模型的计算机输出结果如下。

## General Linear Models Procedure

Dependent Variable: Y DROP

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|----------------|-------------|---------|--------|
| Model           | 11 | 2677.77778     | 243.43434   | 16.49   | 0.0001 |
| Error           | 60 | 886.00000      | 14.76667    |         |        |
| Corrected Total | 71 | 3563.77778     |             |         |        |

| R-Square | C.V.     | Root MSE | Y Mean  |
|----------|----------|----------|---------|
| 0.751387 | 8.925078 | 3.84274  | 43.0556 |

| Source       | DF | Type III SS | Mean Square | F Value | Pr > F |
|--------------|----|-------------|-------------|---------|--------|
| TEMP         | 2  | 1133.77778  | 566.88889   | 38.39   | 0.0001 |
| GAUGE        | 3  | 437.22222   | 145.74074   | 9.87    | 0.0001 |
| TEMP * GAUGE | 6  | 1106.77778  | 184.46296   | 12.49   | 0.0001 |

Dependent Variable: Y

Source: TEMP



Error; MS(TEMP \* GAUGE)

| DF | Type III MS  | Denominator |              | F Value | Pr > F |
|----|--------------|-------------|--------------|---------|--------|
|    |              | DF          | MS           |         |        |
| 2  | 556.88888888 | 6           | 184.46296296 | 3.01897 | 0.1238 |

根据 AOV 表, 我们确认气体温度和测压计的种类之间有显著的交互效应( $p$  值  $< 0.0001$ )。所以, 在  $15^{\circ}\text{C}$  到  $35^{\circ}\text{C}$  的温度范围内压降均值和气体温度间的关系对于所有测压计种类来说并不是相同的。这个结论确认了我们在图 17.1 的侧面图中观察到的关系。此外, 测压计种类引起的压降变化也是很显著的( $p$  值  $< 0.0001$ )。所以测量压降用的测压计在研究中的各温度范围下测得的结果并不一致。然而三个温度下的压降均值并无显著差异( $p$  值  $= 0.1238$ )。这样看来, 涡轮机中扩张接缝处压降变异的主要来源是用于测量压降的测压计种类而不是气体温度。

**结果报告** 我们需要写份报告来总结我们在此试验中的发现。报告中应包括:

1. 对研究目的的陈述。

2. 对研究设计的描述, 研究中的因子是如何选取的, 测压计是如何从测压计总体中选出的, 使用已选出的温度范围的可行性, 何时读取压降读数的时间选择以及其他因子——比如操作者的差异、环境差异、涡轮机运行条件——是如何控制的。

3. 对结果从此研究应用到涡轮机总体的可推广性的讨论。

4. 数据集的数值和图表概括。

5. 所有推断方法的描述:

- 叙述所有研究假设及其  $F$  检验和  $p$  值,
- 证实推断方法用到的必要条件得到了满足。

6. 对结果和结论的讨论。

7. 对与以前研究相关的发现的讨论。

8. 对以后研究的建议。

9. 列出数据集。

这个讨论只为混合模型的学习提供了一个简单的入门介绍。事实上, 我们可以在研究生水平的学习中花一个或更多章节讨论适合于混合模型的话题。对于更复杂的工作, 我们可以分析有三个或更多因子(一些随机, 另一些固定)的析因试验。此外, 当研究两个因子(都是固定效应)对响应变量的效应同时以第三个因子(随机的)为区组变量时, 裂区设计就成为随机化区组设计中析因试验的重要替代

方法。裂区设计和属于随机化区组设计的析因试验之间的差别在于将处理应用于试验单元的方法上。裂区设计中,对每个区组,因子1的水平是随机地分配到试验单元中的,然后第二个因子的水平再被随机地分到因子1每个水平内的子单元中。这种随机化与随机化区组设计中的析因试验的随机化是非常不同的。有关此话题的讨论将在17.6节中进行。

## 练习

## 应用

**17.6(环境)** 下面的研究是要分析培育出的四种用来控制火蚁的化学制剂的效果。化学制剂放置的环境条件可能会对它杀死火蚁的效力产生影响,所以研究者从一大堆可选的位置中随机选了五个位置,每个位置代表一个随机选取的环境。为减少火蚁群体不同和它们居住的土堤不同带来的影响,研究者建立了40个人造的火蚁居住的土丘并让有相似祖先的50,000只火蚁居住在里面。研究者随机地给20个化学剂-位置组合的每一个分配了两个土丘。一周内杀死的火蚁数量被记录下来。火蚁死亡的数目(以千记)如下。

a. 为本研究写出合适的线性模型,辨明模型中所有项的意义。

b. 计算试验的平方和并在AOV表中报告出来,注意AOV表中要包括期望均方的列。

| 地点 | 化学制剂 |     |     |     |
|----|------|-----|-----|-----|
|    | 1    | 2   | 3   | 4   |
| 1  | 7.2  | 4.2 | 9.5 | 5.4 |
|    | 9.6  | 3.5 | 9.3 | 3.9 |
| 2  | 8.5  | 2.9 | 8.8 | 6.3 |
|    | 9.6  | 3.3 | 9.2 | 6.0 |
| 3  | 9.1  | 1.8 | 7.6 | 6.1 |
|    | 8.6  | 2.4 | 7.1 | 5.6 |
| 4  | 8.2  | 3.6 | 7.3 | 5.0 |
|    | 9.0  | 4.4 | 7.0 | 5.4 |
| 5  | 7.8  | 3.7 | 9.2 | 6.5 |
|    | 8.0  | 3.9 | 8.3 | 6.9 |

**17.7** 根据例17.6,进行方差分析,给出结论,取 $\alpha=0.05$ 。

## 17.5 计算期望均方的规则

在第十五章我们讨论了固定效应模型下单因子和两因子试验的方差分析;本章前面我们又讨论了随机和混合模型。本节中我们将看到对任何在每个因子水平组合处有  $n$  次观测的  $k$ -因子试验的数据,利用一些相当简单的规则就可能写出固定、随机或混合模型所有主效应和交互效应的期望均方。*这些准则的重要性在于,在写出某个不熟悉的试验设计的期望均方后,我们常常就可以建立适当的  $F$  检验。* 固定和随机模型的假设将与在前面几节中描述固定、随机和混合模型时的一样。

在我们继续讨论计算期望均方的准则之前,先要给出两条判定交互效应是固定效应还是随机效应的准则。

**交互效应的判定准则:**

1. 如果一个固定效应与另一个固定效应交互作用,则所得的交互效应是固定效应。
2. 如果一个随机效应与另一个效应(固定或随机)交互作用,则所得的交互效应是个随机的分量。

### 例 17.4

考虑一个  $3 \times 6$  的析因试验,每个因子水平组合处有两个观测值。判定如下情况下 AB 的交互效应是随机的还是固定的。

- a. A 和 B 都是固定效应。
- b. A 是固定 B 是随机的。
- c. A 和 B 都是随机的。

**解答** 我们应用上面的准则来判定交互效应。

- a. AB 是固定效应,因为是 A(固定)与 B(固定)交互作用的结果。
- b. AB 是随机分量,因为是 A(固定)与 B(随机)交互作用的结果。
- c. AB 是随机的,因为是 A(随机)与 B(随机)交互作用的结果。

### 例 17.5

考虑因子为 A、B 和 C 的析因试验。判定 AB, AC, BC 和 ABC 的交互效应是固定还是随机的,其中 A 和 B 是固定的而 C 是随机的。

**解答** 我们应用上述的判定准则。

AB 是固定的, A(固定)与 B(固定)交互作用。

AC 是随机的, A(固定)与 C(随机)交互作用。

BC 是随机的, B(固定)与 C(随机)交互作用。

ABC 是随机的, A(固定)与 BC(随机)交互作用。

在叙述计算期望均方的规则前, 建立一个均方表很方便。步骤总结如下并以一个  $a \times b$  的析因试验来举例说明, 其中因子 A 随机、B 固定, A 和 B 的每个因子水平组合处有  $n$  次观测。

**构造均方表的步骤:**

1. 写出试验的模型。对  $a \times b$  析因试验, 模型为

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{k[ij]}$$

注意:  $\epsilon$  项中使用了括弧是为了说明 A 和 B 每个因子水平组合处(即, 对  $i, j$  的每个选择)有  $k = 1, 2, \dots, n$  次观测。

2. 用模型中每个量(除了  $\mu$ )作为一行的标题构造一个表, 表的形式如下。

---

 $\alpha_i$ 
 $\beta_j$ 
 $\alpha\beta_{ij}$ 
 $\epsilon_{k[ij]}$ 


---

3. 模型中每个下标形成一列。

---

 $i$ 
 $j$ 
 $k$ 


---

 $\alpha_i$ 
 $\beta_j$ 
 $\alpha\beta_{ij}$ 
 $\epsilon_{k[ij]}$ 


---

4. 在每列标题上方说明下标对应的是固定(F)效应还是随机效应(R)。

5. 还要标示出每个下标的水平数。对我们的例子由第 4 步和第 5 步得到的表如下。

|                    | $a$ | $b$ | $n$ |
|--------------------|-----|-----|-----|
|                    | $R$ | $F$ | $R$ |
|                    | $i$ | $j$ | $k$ |
| $\alpha_i$         |     |     |     |
| $\beta_j$          |     |     |     |
| $\alpha\beta_{ij}$ |     |     |     |
| $\epsilon_{k[ij]}$ |     |     |     |

6. 对给定一系列的每一行, 写入列下标对应的水平数, 除非行中量的下标含有列下标。

|                    | $a$ | $b$ | $n$ |
|--------------------|-----|-----|-----|
|                    | $R$ | $F$ | $R$ |
|                    | $i$ | $j$ | $k$ |
| $\alpha_i$         |     | $b$ | $n$ |
| $\beta_j$          | $a$ |     | $n$ |
| $\alpha\beta_{ij}$ |     |     | $n$ |
| $\epsilon_{k[ij]}$ |     |     |     |

7. 查看已经列在表中第一列的模型中的各量, 对下标中含有括号的量, 在含有括号里涉及到的下标的列下面写上 1。

|                    | $a$ | $b$ | $n$ |
|--------------------|-----|-----|-----|
|                    | $R$ | $F$ | $R$ |
|                    | $i$ | $j$ | $k$ |
| $\alpha_i$         |     | $b$ | $n$ |
| $\beta_j$          | $a$ |     | $n$ |
| $\alpha\beta_{ij}$ |     |     | $n$ |
| $\epsilon_{k[ij]}$ | 1   | 1   |     |

8. 对于一系列中剩下的位置, 如果此列的头上是  $F$  就填上 0; 如果是  $R$  就填上 1。

|                    | $a$ | $b$ | $n$ |
|--------------------|-----|-----|-----|
|                    | $R$ | $F$ | $R$ |
|                    | $i$ | $j$ | $k$ |
| $\alpha_i$         | 1   | $b$ | $n$ |
| $\beta_j$          | $a$ | 0   | $n$ |
| $\alpha\beta_{ij}$ | 1   | 0   | $n$ |
| $\epsilon_{k[ij]}$ | 1   | 1   | 1   |

这就是两因子试验计算期望均方用的均方表, 其中因子 A 是随机的, 因子 B 是固定的, 在 A 和 B 的每个因子水平处有  $n$  次观测。

当你有均方表之后计算期望均方就是很简单的了。下面列出了规则。

#### 利用均方表计算 EMS 的规则

1. 察看要计算均方的那一项的下标。
2. 删去均方表中那些不含此量下标的行。
3. 遮上表内那些以此项中未被括起的下标打头的列。
4. 将每一行余下的未被遮起的量相乘, 就得到了期望均方中各项的协系数。

#### 例 17.6

计算两因子试验中的  $E(MSA)$ , 因子 A(随机)有  $a$  个水平, 因子 B(固定)有  $b$  个水平, 每个因子水平组合处有  $n$  次观测。

**解答** 根据刚刚给出的均方表, 我们注意到  $\alpha_i$  的下为  $i$ ; 所以我们删去表中第二行, 遮去第 1 列(以  $i$  顶头), 并将余下的项相乘。表 17.16 显示了这些步骤、乘法和 EMS 的各量。

表 17.16 计算  $E(MSA)$  的均方表

|                    | $a$ | $b$ | $n$ |      |                |
|--------------------|-----|-----|-----|------|----------------|
|                    | $R$ | $F$ | $R$ | 余下项的 | EMS            |
|                    | $i$ | $j$ | $k$ | 乘积   |                |
| $\alpha_i$         | 1   | $b$ | $n$ | $bn$ | $bn\sigma_a^2$ |
| $\beta_j$          | $a$ | 0   | $n$ | —    |                |
| $\alpha\beta_{ij}$ | 1   | 0   | $n$ | 0    |                |
| $\epsilon_{k[ij]}$ | 1   | 1   | 1   | 1    | $\sigma_e^2$   |

利用表 17.16 的最后一列, 我们有

$$E(\text{MSA}) = \sigma_{\epsilon}^2 + bn\sigma_{\alpha}^2$$

$E(\text{MSB})$  的计算遵从类似的方法。 $\beta_j$  的下标为  $j$ , 所以我们删去均方表中第二列(以  $j$  顶头)和第一行(不含  $j$ )。将余下项相乘得到期望均方中各项系数(见表 17.17)。

表 17.17 计算  $E(\text{MSB})$  的均方表

|                    | $a$ | $b$ | $n$ |      |                           |
|--------------------|-----|-----|-----|------|---------------------------|
|                    | $R$ | $F$ | $R$ | 余下项的 | EMS                       |
|                    | $i$ | $j$ | $k$ | 乘积   |                           |
| $\alpha_i$         | 1   | b   | n   | —    |                           |
| $\beta_j$          | a   | 0   | n   | an   | $an\theta_{\beta}$        |
| $\alpha\beta_{ij}$ | 1   | 0   | n   | n    | $n\sigma_{\alpha\beta}^2$ |
| $\epsilon_{k[ij]}$ | 1   | 1   | 1   | 1    | $\sigma_{\epsilon}^2$     |

所以,

$$E(\text{MSB}) = \sigma_{\epsilon}^2 + n\sigma_{\alpha\beta}^2 + an\theta_{\beta}$$

其中  $\theta_{\beta}$  是形如下式的常数

$$\theta_{\beta} = \frac{\sum_i \beta_j^2}{b-1}$$

### 例 17.7

- 建立因子 A 固定、因子 B 随机的两因子试验的期望均方计算用表, A 和 B 的每个因子水平组合处有  $n$  次观测。
- 计算  $E(\text{MSA})$ 。

### 解答

- 此试验的模型为

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{k[ij]}$$

相应的均方表见表 17.18。

表 17.18 例 17.7 的均方表

|                    | $a$ | $b$ | $n$ |
|--------------------|-----|-----|-----|
|                    | $F$ | $R$ | $R$ |
|                    | $i$ | $j$ | $k$ |
| $\alpha_i$         | 0   | $b$ | $n$ |
| $\beta_j$          | $a$ | 1   | $n$ |
| $\alpha\beta_{ij}$ | 0   | 1   | $n$ |
| $\epsilon_{k[ij]}$ | 1   | 1   | 1   |

b. 从表 17.19 中可以找到  $E(\text{MSA})$ 。

表 17.19 计算  $E(\text{MSA})$ 

|                    | $a$ | $b$ | $n$ |      |                           |
|--------------------|-----|-----|-----|------|---------------------------|
|                    | $F$ | $F$ | $R$ | 余下项的 | EMS                       |
|                    | $i$ | $j$ | $k$ | 乘积   |                           |
| $\alpha_i$         | 0   | $b$ | $n$ | $bn$ | $bn\theta_\alpha$         |
| $\beta_j$          | $a$ | 1   | $n$ | —    |                           |
| $\alpha\beta_{ij}$ | 0   | 1   | $n$ | $n$  | $n\sigma_{\alpha\beta}^2$ |
| $\epsilon_{k[ij]}$ | 1   | 1   | 1   | 1    | $\sigma_\epsilon^2$       |

所以  $E(\text{MSA}) = \sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 + bn\theta_\alpha$ , 其中  $\theta_\alpha = \sum_i \alpha_i^2 / (a - 1)$ 。做适当的符号变换, 就会发现这与我们在前面的因子 B 固定的例子中得到的结果相同。

在前面我们只着重关注了固定效应模型。对这种模型, 检验统计量都是用相应的均方作分子被 MSE 除得到的。然而对于随机和混合效应模型, 检验统计量并不总是这样。交互效应的检验统计量, 即  $F$  等于  $\text{MSAB}/\text{MSE}$ , 对固定、随机和混合模型都是相同的; 但对因子 A 和 B 的  $F$  检验根据对  $\alpha_i$  和  $\beta_j$  的假设而变。例如, 因子 A 的  $F$  检验是  $\text{MSA}/\text{MSE}$  当因子 A 固定且 B 固定或因子 A 随机 B 固定时。相反, 当因子 A 固定 B 随机或 A 随机 B 也随机时, 因子 A 的  $F$  检验是  $\text{MSA}/\text{MSAB}$ 。由此可见对于随机模型和混合模型, 辨识期望均方是很重要的。

对两因子试验有一种特殊的情况需要提到: 当 A 和 B 每个因子水平组合处只有一次观测 ( $n = 1$ ) 时, 我们的模型为



$$y_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ij}$$

你可以从表 17.20 缩减的 AOV 表中看到,误差项没有自由度,所以没有对交互效应的检验,并且视模型情况有时对主效应也可能没有有效的检验。惟一可能补救的情况就是当我们能假设因子 A 和 B 间没有交互效应时,所有的主效应可以利用模型

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

的均方误差来检验,无论是固定、随机还是混合模型。当无交互效应的假设不合理时,为了获得交互效应和主效应的有效检验,试验必须考虑在因子 A 和 B 的因子水平组合处进行重复( $n > 1$ )。

表 17.20 缩减的 AOV 表,两因子试验( $n = 1$ )

| 来源 | df               |
|----|------------------|
| A  | $a - 1$          |
| B  | $b - 1$          |
| AB | $(a - 1)(b - 1)$ |
| 误差 | —                |
| 总计 | $ab - 1$         |

对两因子试验用的规则同样可用于更复杂的模型,虽然它们看来有些麻烦,但在实践中它们简单易用。我们将给出一个三因子试验的例子。对于有关假设、推导和更复杂应用的更多细节,可看 Hicks, Turner(1999)和 Kuehl(1999)。

### 例 17.8

给出每个因子水平组合处有  $n = 4$  次观测的  $3 \times 5 \times 2$  析因试验的期望均方。将因子 A 和 B 视为固定,因子 C 视为随机。

**解答** 下面给出了此试验的全模型及相应的均方表:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{l[ijk]}$$

我们将建立针对  $a, b, c$  和  $n$  一般取值的均方表,然后再代入具体值。利用前面讨论的规则得到的均方表见表 17.21。对此均方表应用 EMS 规则就可得到模型中各量的期望均方。比如对  $E(MSA)$ ,未被遮去的项见表 17.22。

表 17.21 例 17.8 的均方表

|                           | $a$ | $b$ | $c$ | $n$ |
|---------------------------|-----|-----|-----|-----|
|                           | $F$ | $F$ | $R$ | $R$ |
|                           | $i$ | $j$ | $k$ | $l$ |
| $\alpha_i$                | 0   | $b$ | $c$ | $n$ |
| $\beta_j$                 | $a$ | 0   | $c$ | $n$ |
| $\gamma_k$                | $a$ | $b$ | 1   | $n$ |
| $\alpha\beta_{ij}$        | 0   | 0   | $c$ | $n$ |
| $\alpha\gamma_{ik}$       | 0   | $b$ | 1   | $n$ |
| $\beta\gamma_{jk}$        | $a$ | 0   | 1   | $n$ |
| $\alpha\beta\gamma_{ijk}$ | 0   | 0   | 1   | $n$ |
| $\epsilon_l[ijk]$         | 1   | 1   | 1   | 1   |

表 17.22 例 17.8  $E(\text{MSA})$  的计算

|                           | $a$ | $b$ | $c$ | $n$ |       |                             |
|---------------------------|-----|-----|-----|-----|-------|-----------------------------|
|                           | $F$ | $F$ | $R$ | $R$ | 余下项的  | EMS                         |
|                           | $i$ | $j$ | $k$ | $l$ | 乘积    |                             |
| $\alpha_i$                | 0   | $b$ | $c$ | $n$ | $bcn$ | $bcn\theta_A$               |
| $\beta_j$                 | $a$ | 0   | $c$ | $n$ | —     |                             |
| $\gamma_k$                | $a$ | $b$ | 1   | $n$ | —     |                             |
| $\alpha\beta_{ij}$        | 0   | 0   | $c$ | $n$ | 0     |                             |
| $\alpha\gamma_{ik}$       | 0   | $b$ | 1   | $n$ | $bn$  | $bn\sigma_{\alpha\gamma}^2$ |
| $\beta\gamma_{jk}$        | $a$ | 0   | 1   | $n$ | —     |                             |
| $\alpha\beta\gamma_{ijk}$ | 0   | 0   | 1   | $n$ | 0     |                             |
| $\epsilon_l[ijk]$         | 1   | 1   | 1   | 1   | 1     | $\sigma_\epsilon^2$         |

从表中最后一列我们有

$$E(\text{MSA}) = \sigma_\epsilon^2 + bn\sigma_{\alpha\gamma}^2 + bcn\theta_A$$

代入  $a=3, b=5, c=2$  和  $n=4$ , 上式变为

$$E(\text{MSA}) = \sigma_\epsilon^2 + 20\sigma_{\alpha\gamma}^2 + 40\theta_A$$

其中

$$\theta_A = \sum_i \alpha_i^2 / 2$$

类似可以得到因子 B 和 C 的期望均方如下

$$E(\text{MSB}) = \sigma_\epsilon^2 + an\sigma_{\beta\gamma}^2 + acn\theta_B$$

$$= \sigma_e^2 + 12\sigma_{\beta\gamma}^2 + 24\theta_B$$

和

$$\begin{aligned} E(\text{MSC}) &= \sigma_e^2 + abn\sigma_\gamma^2 \\ &= \sigma_e^2 + 60\sigma_\gamma^2 \end{aligned}$$

计算  $AB$  交互效应的期望均方的表格见表 17.23。MSAB 的期望均方为

$$\begin{aligned} E(\text{MSAB}) &= \sigma_e^2 + n\sigma_{a\beta\gamma}^2 + cn\theta_{AB} \\ &= \sigma_e^2 + 4\sigma_{a\beta\gamma}^2 + 8\theta_{AB} \end{aligned}$$

其中

$$\theta_{a\beta} = \sum_{ij} \frac{a\beta_{ij}^2}{8}$$

表 17.23 例 17.8  $E(\text{MSAB})$  的计算

|                           | $a$ | $b$ | $c$ | $n$ |      |                            |
|---------------------------|-----|-----|-----|-----|------|----------------------------|
|                           | $F$ | $F$ | $R$ | $R$ | 余下项的 | EMS                        |
|                           | $i$ | $j$ | $k$ | $l$ | 乘积   |                            |
| $\alpha_i$                | 0   | $b$ | $c$ | $n$ | —    |                            |
| $\beta_j$                 | $a$ | 0   | $c$ | $n$ | —    |                            |
| $\gamma_k$                | $a$ | $b$ | 1   | $n$ | —    |                            |
| $\alpha\beta_{ij}$        | 0   | 0   | $c$ | $n$ | $cn$ | $cn\theta_{AB}$            |
| $\alpha\gamma_{ik}$       | 0   | $b$ | 1   | $n$ | —    |                            |
| $\beta\gamma_{jk}$        | $a$ | 0   | 1   | $n$ | —    |                            |
| $\alpha\beta\gamma_{ijk}$ | 0   | 0   | 1   | $n$ | $n$  | $n\sigma_{a\beta\gamma}^2$ |
| $\epsilon_{l[ijk]}$       | 1   | 1   | 1   | 1   | 1    | $\sigma_e^2$               |

在对  $AC$  和  $BC$  的交互效应应用 EMS 规则后, 得到

$$\begin{aligned} E(\text{MSAC}) &= \sigma_e^2 + bn\sigma_{a\gamma}^2 \\ &= \sigma_e^2 + 20\sigma_{a\gamma}^2 \end{aligned}$$

和

$$\begin{aligned} E(\text{MSBC}) &= \sigma_e^2 + an\sigma_{\beta\gamma}^2 \\ &= \sigma_e^2 + 12\sigma_{\beta\gamma}^2 \end{aligned}$$

同样可以得到 MSABC 和 MSE 的期望

$$\begin{aligned} E(\text{MSABC}) &= \sigma_e^2 + n\sigma_{a\beta\gamma}^2 \\ &= \sigma_e^2 + 4\sigma_{a\beta\gamma}^2 \end{aligned}$$

和

$$E(\text{MSE}) = \sigma_e^2$$

对这个每个单元处有  $n=4$  次观测, 因子  $A$  和  $B$  固定而  $C$  随机的  $3 \times 5 \times 2$  析因试验, 表 17.24 是对我们根据本章讲到的 EMS 规则算得的期望均方的总结, 其中包括了检验某变异来源是否显著时有有效的  $F$  检验的分母。

表 17.24 例 17.8 的部分 AOV 表

| 来源  | EMS   | $F$ 检验的分子 |
|-----|---|-----------|
| A   | $\sigma_e^2 + 20\sigma_{\alpha\gamma}^2 + 40\theta_A$       | MSAC      |
| B   | $\sigma_e^2 + 12\sigma_{\beta\gamma}^2 + 24\theta_B$        | MSBC      |
| C   | $\sigma_e^2 + 60\sigma_{\alpha\gamma}^2$                    | MSE       |
| AB  | $\sigma_e^2 + 4\sigma_{\alpha\beta\gamma}^2 + 8\theta_{AB}$ | MSABC     |
| AC  | $\sigma_e^2 + 20\sigma_{\alpha\gamma}^2$                    | MSE       |
| BC  | $\sigma_e^2 + 12\sigma_{\beta\gamma}^2$                     | MSE       |
| ABC | $\sigma_e^2 + 4\sigma_{\alpha\beta\gamma}^2$                | MSE       |
| 误差  | $\sigma_e^2$  | *         |

**例 17.9**

根据例 17.8, 对于

$$H_0: \theta_A = 0 \text{ 和 } H_0: \sigma_{\beta\gamma}^2 = 0$$

给出合适的  $F$  检验统计量

**解答** 根据表 17.24 中的期望均方, 很明显  $H_0: \theta_A = 0$  的检验统计量是  $F = \text{MSA}/\text{MSAC}$ ;  $H_0: \sigma_{\beta\gamma}^2 = 0$  的检验统计量是  $F = \text{MSBC}/\text{MSE}$ 。

对固定效应模型我们总可以获得所有变异性来源的有效检验, 但这一点对于某些随机效应模型和混合效应模型却并不一定成立。表 17.25, 17.26, 17.27 和表 17.28 显示了几个三因子试验的 EMS。在这些表中我们给出了那些可以进行  $F$  检验的方差分量在进行  $F$  检验时的分母, \* 号说明此方差分量没有有效的  $F$  检验。对随机效应和混合效应模型中无法进行有效  $F$  检验的变异性来源可以建立近似的  $F$  检验。这些检验可以借助一些计算机软件实现——如 SAS 和 SPSS。对这些检验的论述可见 Hicks, Turner(1999)和 Kuehl(1999)。

表 17.25 所有因子固定、有  $n$  次重复的三因子  $a \times b \times c$  设计

| 来源  | 所有因子固定                                |         |
|-----|---------------------------------------|---------|
|     | EMS                                   | F 检验的分母 |
| A   | $\sigma_{\epsilon}^2 + bcn\theta_A$   | MSE     |
| B   | $\sigma_{\epsilon}^2 + acn\theta_B$   | MSE     |
| C   | $\sigma_{\epsilon}^2 + abn\theta_C$   | MSE     |
| AB  | $\sigma_{\epsilon}^2 + cn\theta_{AB}$ | MSE     |
| AC  | $\sigma_{\epsilon}^2 + bn\theta_{AC}$ | MSE     |
| BC  | $\sigma_{\epsilon}^2 + an\theta_{BC}$ | MSE     |
| ABC | $\sigma_{\epsilon}^2 + n\theta_{ABC}$ | MSE     |
| 误差  | $\sigma_{\epsilon}^2$                 | *       |

表 17.26 所有因子随机、有  $n$  次重复的三因子  $a \times b \times c$  设计

| 来源  | 所有因子随机  |         |
|-----|---|---------|
|     | EMS   | F 检验的分母 |
| A   | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2 + cn\sigma_{a\beta}^2 + bn\sigma_{a\gamma}^2 + bcn\sigma_a^2$             | *       |
| B   | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2 + cn\sigma_{a\beta}^2 + an\sigma_{\beta\gamma}^2 + acn\sigma_{\beta}^2$   | *       |
| C   | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2 + bn\sigma_{a\gamma}^2 + an\sigma_{\beta\gamma}^2 + abn\sigma_{\gamma}^2$ | *       |
| AB  | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2 + cn\sigma_{a\beta}^2$  | MSABC   |
| AC  | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2 + bn\sigma_{a\gamma}^2$   | MSABC   |
| BC  | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2 + an\sigma_{\beta\gamma}^2$   | MSABC   |
| ABC | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2$  | MSE     |
| 误差  | $\sigma_{\epsilon}^2$   | *       |

表 17.27 因子 A、B 随机, 因子 C 固定, 有  $n$  次重复的三因子  $a \times b \times c$  设计

| 来源  | A、B 随机, C 固定   |         |
|-----|--|---------|
|     | EMS  | F 检验的分母 |
| A   | $\sigma_{\epsilon}^2 + cn\sigma_{a\beta}^2 + bcn\sigma_a^2$  | MSAB    |
| B   | $\sigma_{\epsilon}^2 + cn\sigma_{a\beta}^2 + acn\sigma_{\beta}^2$  | MSAB    |
| C   | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2 + bn\sigma_{a\gamma}^2 + an\sigma_{\beta\gamma}^2 + abn\theta_C$ | *       |
| AB  | $\sigma_{\epsilon}^2 + cn\sigma_{a\beta}^2$  | MSE     |
| AC  | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2 + bn\sigma_{a\gamma}^2$  | MSABC   |
| BC  | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2 + an\sigma_{\beta\gamma}^2$                                      | MSABC   |
| ABC | $\sigma_{\epsilon}^2 + n\sigma_{a\beta\gamma}^2$   | MSE     |
| 误差  | $\sigma_{\epsilon}^2$  | *       |

表 17.28 因子 A 随机, 因子 B、C 固定, 有  $n$  次重复的三因子  $a \times b \times c$  设计

| 来源  | A 随机, B、C 固定  |         |
|-----|---|---------|
|     | EMS   | F 检验的分母 |
| A   | $\sigma_e^2 + bc n \sigma_a^2$                            | MSE     |
| B   | $\sigma_e^2 + cn \sigma_{a\beta}^2 + ac n \theta_B$       | MSAB    |
| C   | $\sigma_e^2 + bn \sigma_{a\gamma}^2 + ab n \theta_C$      | MSAC    |
| AB  | $\sigma_e^2 + cn \sigma_{a\beta}^2$                       | MSE     |
| AC  | $\sigma_e^2 + bn \sigma_{a\gamma}^2$                      | MSE     |
| BC  | $\sigma_e^2 + n \sigma_{a\beta\gamma}^2 + an \theta_{BC}$ | MSABC   |
| ABC | $\sigma_e^2 + n \sigma_{a\beta\gamma}^2$                  | MSE     |
| 误差  | $\sigma_e^2$  | *       |

17.2 节和 17.3 节说明了方差分量的估计。继前面部分中介绍的步骤后, 令均方等于期望均方, 我们可得到平衡设计随机效应和混合效应模型方差分量的估计。很多计算机软件的程序可以实现这些计算—如 SAS 和 SPSS。但是对某些程序, 它所施加在模型上的条件与我们用的条件, 即所谓的经典条件是不同的。在这些软件程序强加给模型的条件, 对有些固定模型得到的期望均方是不一样的。所以方差分量的估计值和适当的  $F$  检验可能会和经典条件下得到的结果不同。不平衡设计方差分量的估计是个复杂的问题, 超出了本书的范围。对这一问题的详细讨论可见 Searle, Casella 和 McCulloch(1992)。

## 17.6 套抽样和裂区设计

有时在试验中一个因子是嵌套在另一个内的, 可以用下例来说明这种情况。一个医药公司进行检验以测定它们的产品在一个特定时间点的稳定性(在室温条件下)。用了两个制造地点。在每个地点, 随机获取三批产品, 并从每批产品中再随机抽取十个药片。这个设计可用图 17.2 表示如下。

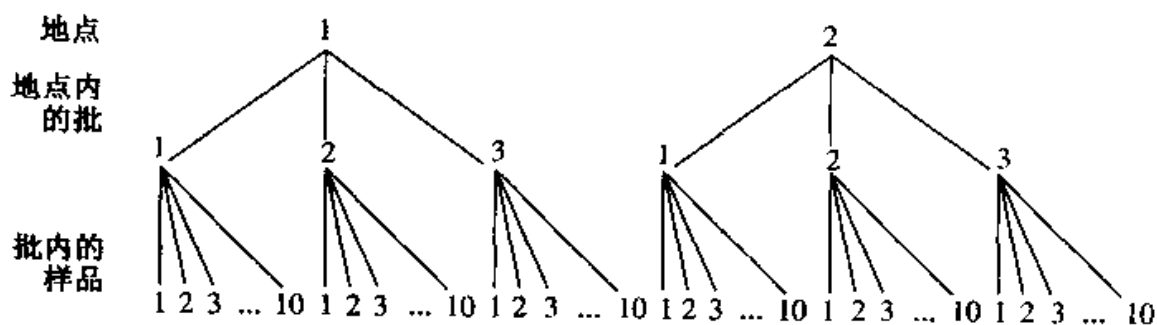


图 17.2 产品批嵌套在地点内的两因子试验

虽然这看起来像通常的两因子试验,因子  $A$  是地点,因子  $B$  是产品批,但是要注意到从地点 1 取的三批产品和从地点 2 取的三批产品是不同的。从这个意义上说因子  $B$  (产品批)是嵌套在因子  $A$  (地点)里的。在这种试验情况下不可能衡量因子  $B$  与因子  $A$  的交互效应,因为不像析因试验中对因子  $A$  和  $B$  的安排那样,因子  $B$  的每个水平并没有和因子  $A$  的每个水平都同时存在。这里,一个地点内的三个产品批对于这个地点来说是惟一的。

因子  $B$  嵌套在因子  $A$  中的两因子试验(每个单元处有  $n$  次观测)模型一般可写为

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk} \quad \begin{array}{l} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, n \end{array}$$

注意这个模型和 17.3 节中两因子试验的模型相似,除了没有交互效应项  $\alpha\beta_{ij}$  并且因子  $B$  的项是  $\beta_{j(i)}$ ,下标表示因子  $B$  的第  $j$  个水平是嵌套在因子  $A$  的第  $i$  个水平里的。此设计的方差分析表见表 17.29。

表 17.29 因子  $B$  嵌套在因子  $A$  中的两因子试验(每个单元有  $n$  次观测)的 AOV 表

| 来源     | SS     | df          | MS     | EMS                       |   |  |
|--------|--------|-------------|--------|---------------------------|---|--|
|        |        |             |        | 固定                        | 混合( $A$ 固定)                                 | 随机   |
| $A$    | SSA    | $a - 1$     | MSA    | $\sigma_e^2 + bn\theta_A$ | $\sigma_e^2 + n\sigma_\beta^2 + bn\theta_A$ | $\sigma_e^2 + n\sigma_\beta^2 + bn\sigma_\alpha^2$ |
| $B(A)$ | SSB(A) | $a(b - 1)$  | MSB(A) | $\sigma_e^2 + n\theta_B$  | $\sigma_e^2 + n\sigma_\beta^2$              | $\sigma_e^2 + n\sigma_\beta^2$                     |
| 误差     | SSE    | $ab(n - 1)$ | MSE    | $\sigma_e^2$              | $\sigma_e^2$                                | $\sigma_e^2$                                       |
| 总计     | TSS    | $abn - 1$   |        |                           |   |  |

AOV 中的平方和用下面的公式计算。

$$\begin{aligned} \text{TSS} &= \sum_{ijk} (y_{ijk} - \bar{y} \dots)^2 \\ \text{SSA} &= \sum_i bn (\bar{y}_{i..} - \bar{y} \dots)^2 \\ \text{SSB}(A) &= \sum_i \sum_j n (\bar{y}_{ij.} - \bar{y}_{i..})^2 \\ \text{SSE} &= \text{TSS} - \text{SSA} - \text{SSB}(A) \end{aligned}$$

表 17.29 中还给出了三种常见情况的期望均方,尤其要注意下面的问题:

1. 因子  $B$  的  $F$  检验总是

$$F = \frac{\text{MSB}(A)}{\text{MSE}}$$

2. 固定效应模型中因子 A 的 F 检验是

$$F = \frac{MSA}{MSE}$$

但在随机和混合效应模型中因子 A 的相应检验是

$$F = \frac{MSA}{MSB(A)}$$

3.  $n=1$  时, 没有因子 B 的检验, 但我们可以用下式检验随机和混合效应模型中的因子 A

$$F = \frac{MSA}{MSB(A)}$$

### 例 17.10

研究者进行试验以测定降低血压用的心血管药物中一种加膜药片成分的一致性。他们从两个调制地点中每一个获取三批产品作为随机样本; 从每批产品中随机取五个药片作样本进行化验以测定成分的一致性。数据如下:

|         |      |      |      |      |      |      |
|---------|------|------|------|------|------|------|
| 地点      | 1    |      |      | 2    |      |      |
| 每个地点内的批 | 1    | 2    | 3    | 1    | 2    | 3    |
| 每个批内的药片 | 5.03 | 4.64 | 5.10 | 5.05 | 5.46 | 4.90 |
|         | 5.10 | 4.73 | 5.15 | 4.96 | 5.15 | 4.95 |
|         | 5.25 | 4.82 | 5.20 | 5.12 | 5.18 | 4.86 |
|         | 4.98 | 4.95 | 5.08 | 5.12 | 5.18 | 4.86 |
|         | 5.05 | 5.06 | 5.14 | 5.05 | 5.11 | 5.07 |

a. 进行方差分析, 取  $\alpha=0.05$ 。

b. 有证据说明在成分一致性上产品批与产品批之间有变化吗? F 检验的进行依赖于我们对产品批固定还是随机的假设吗?

c. 对产品批做出结论。

### 解答

a. 对这些数据我们有  $a=2$  个调制地点, 每个地点有  $b=3$  个产品批以及每批中有  $n=5$  个药片。由数据我们算得样本均值如下:



| 地点 | 产品批   |       |       | 地点平均值   |
|----|-------|-------|-------|---------|
|    | 1     | 2     | 3     |         |
| 1  | 5.082 | 4.84  | 5.134 | 5.01867 |
| 2  | 5.06  | 5.216 | 4.928 | 5.068   |
| 总  |       |       |       | 5.04333 |

算得平方和如下:

$$TSS = (5.03 - 5.04333)^2 + (5.10 - 5.04333)^2 + \cdots + (5.07 - 5.04333)^2 - 0.76348$$

$$SSA = 15[(5.01867 - 5.04333)^2 + (5.069 - 5.04333)^2] = 0.01824$$

$$SSB(A) = 5[(5.082 - 5.01867)^2 + (4.84 - 5.01867)^2 + (5.134 - 5.01867)^2 + (5.06 - 5.068)^2 + (5.216 - 5.068)^2 + (4.928 - 5.068)^2] = 0.45401$$

$$SSE = TSS - SSA - SSB(A) = 0.76348 - 0.01824 - 0.45401 = 0.29123$$

这个数据集的计算机输出结果在下面给出。注意这些平方和与我们的计算结果稍有出入,这是舍入误差造成的微小偏离。此试验我们将使用计算机输出的平方和作方差分析表,见下。

| 来源   | SS      | df | MS      | F    |
|------|---------|----|---------|------|
| A    | 0.01825 | 1  | 0.01825 | 0.16 |
| B(A) | 0.45401 | 4  | 0.11350 | 9.39 |
| 误差   | 0.29020 | 24 | 0.01209 |      |
| 总计   | 0.76246 | 29 |         |      |

#### CONTENT UNIFORMITY OF FILM-COATED TABLETS

General Linear Models Procedure

Dependent Variable: Y      CONTENT

| Source          | DF | Sum of     | Mean       | F Value | Pr > F |
|-----------------|----|------------|------------|---------|--------|
| Model           | 5  | 0.47226667 | 0.09445333 | 7.81    | 0.0002 |
| Error           | 24 | 0.29020000 | 0.01209167 |         |        |
| Corrected Total | 29 | 0.76246667 |            |         |        |

| R-Square | C. V. | Root MSE | Y Mean |
|----------|-------|----------|--------|
|----------|-------|----------|--------|

|            |          |             |             |         |        |
|------------|----------|-------------|-------------|---------|--------|
|            | 0.619393 | 2.180346    | 0.10996     | 5.04333 |        |
| Source     | DF       | Type III SS | Mean Square | F Value | Pr>F   |
| SITE       | 1        | 0.01825333  | 0.01825333  | 1.51    | 0.2311 |
| BATH(SITE) | 4        | 0.45401333  | 0.11350333  | 9.39    | 0.0001 |

## Tests of Hypotheses for Mixed Model Analysis of Variance

Dependent Variable: Y      CONTENT

Source: SITE

Error: MS (BATCH(SITE))

|    |              | Denominator |              |         | Denominator |  |
|----|--------------|-------------|--------------|---------|-------------|--|
| DF | Type III MS  | DF          | MS           | F Value | Pr>F        |  |
| 1  | 0.0182533333 | 4           | 0.1135033333 | 0.1608  | 0.7089      |  |

Source: BATCH(SITE)

Error: MS(Error)

|    |              | Denominator |              |         | Denominator |  |
|----|--------------|-------------|--------------|---------|-------------|--|
| DF | Type III MS  | DF          | MS           | F Value | Pr>F        |  |
| 4  | 0.1135033333 | 24          | 0.0120916667 | 9.3869  | 0.0001      |  |

b., c. 产品批的  $F$  检验为

$$F = \frac{MSB(A)}{MSE} = 9.39$$

其中自由度  $df_1 = 4$ ,  $df_2 = 24$ 。因为  $F$  的观测值 9.39 超过了  $\alpha = .05$  水平下  $F$  表中的值, 所以结论是在药片成分一致性上产品批与产品批之间存在相当大的变化。这个检验与产品批是否随机无关。

到现在为止你应该已经意识到随着嵌套效应的引入, 一系列全新的试验设计已经展开。除两因子设计外, 你还可以设想一般的多因子设计, 如因子  $B$  嵌套在因子  $A$  的水平中, 因子  $C$  嵌套在因子  $A$  和  $B$  中等等。表 17.30 是所有因子都随机的三因子嵌套设计的方差分析表。

也有对这些设计的其他扩展。例如你可以进行一个三因子试验, 因子  $A$  和  $B$  是交叉分类的而因子  $C$  是嵌套在因子  $A$  和  $B$  的水平内的。这是一个部分嵌套设计的例子。

表 17.30 三因子嵌套设计的 AOV 表——所有因子都是随机的  
(每单元有  $n$  次观测)

| 来源      | SS        | df           | MS        | EMS  |
|---------|-----------|--------------|-----------|--|
| A       | SSA       | $a - 1$      | MSA       | $\sigma_e^2 + n\sigma_\gamma^2 + cn\sigma_\beta^2 + bc n\sigma_\alpha^2$ |
| B(A)    | SSB(A)    | $a(b - 1)$   | MSB(A)    | $\sigma_e^2 + n\sigma_\gamma^2 + cn\sigma_\beta^2$                       |
| C(A, B) | SSC(A, B) | $ab(c - 1)$  | MSC(A, B) | $\sigma_e^2 + n\sigma_\gamma^2$  |
| 误差      | SSE       | $abc(n - 1)$ | MSE       | $\sigma_e^2$   |
| 总计      | TSS       | $abcn - 1$   |           |  |

假设某市场调研公司要在国内四个地理区域 ( $B_1, \dots, B_4$ ) 内从潜在的客户群中抽样以获得他们对两种产品 ( $A_1$  和  $A_2$ ) 的意见。在每个地区随机抽取六个销售产品  $A_i$  的商店作为样本。在地区  $B_j$  为产品  $A_i$  选定的每个商店里, 采访了十个人对产品  $i$  的意见。这个设计中, 因子 C(商店) 是嵌套在因子 A(产品) 和因子 B(地区) 的水平内的, 并且嵌套在因子 A、B 水平内的因子 C 每个水平处都有  $n=10$  次观测(意见)。

遇到嵌套和部分嵌套设计的可能性是无穷尽的, 但是遗憾在此处我们没有机会分析它们。感兴趣的读者可以参考 Kuehl(1999) 和 Montgomery(1997) 以获得对这个问题的更多处理方法。但是我们要考虑一个与部分嵌套设计类似的常用设计。它被称为裂区设计, 这是因为它起源于农业试验。我们用一个例子说明它的用途。

要比较两种不同的肥料水平下三种大豆的产量。如果我们想要在肥料和大豆品种的每个组合处得到  $n=2$  次观测, 就需要有 12 块相同尺寸的地。将肥料作为因子 A, 大豆品种作为因子 T, 一种可能的设计方法是完全随机化设计中的标准  $2 \times 3$  析因试验, 每个因子水平组合有  $n=2$  次观测。但是因为向地里施肥是在准备播种土地时进行的, 所以(逻辑上)很难在每块地里种上要求的大豆品种之前就把肥料  $A_1$  先施加到由因子 A 和 T 决定的六块土地上再把肥料  $A_2$  施加到另六块土地上。

一个更易于执行的设计是将每种肥料施加到两“整块”土地上, 然后在每整块地内将三种大豆种在三“小块”地上(与前面设计中每块地的尺寸相同)。这种设计如图 17.3 所示。

此设计称为裂区设计, 设计中有两阶段随机化。首先, 因子 A(肥料) 的水平是随机分到整区中的; 其次, 因子 T(大豆) 的水平是随机分到整区中的小区上的(见图 17.4)。利用这个设计更容易准备土地, 将适当的肥料施加到整区上, 然后再将大豆种到小区上; 而不是准备肥料, 先将肥料施加到小区上然后再将大豆种到小区上, 就像标准的  $2 \times 3$  析因试验中那样。

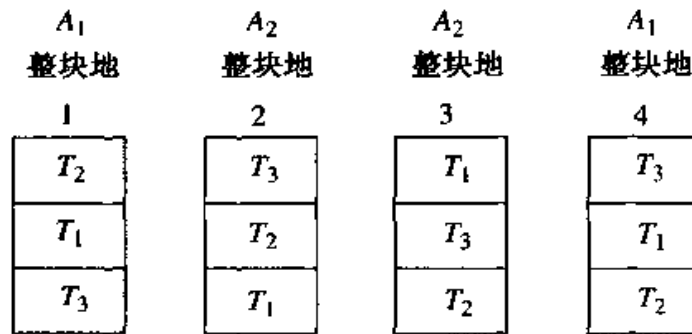


图 17.3 裂区设计

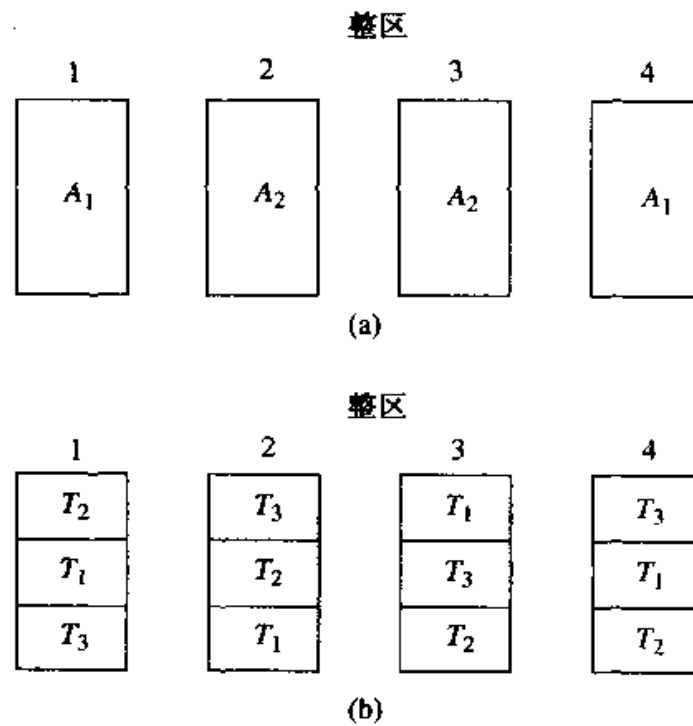


图 17.4 完全随机化裂区设计的两阶段随机化

因为在整区水平上和小区水平上的随机化都是依据完全随机化设计进行的, 所以此设计通常被称为完全随机化裂区设计。

考虑因子  $A$  有  $a$  个水平, 因子  $T$  有  $t$  个水平并且因子  $A$  的第  $i$  个水平处有  $n$  次重复的完全随机化裂区设计的模型。若  $y_{ijk}$  表示因子  $A$  的第  $i$  个水平因子  $T$  的第  $j$  个水平处的第  $k$  次的响应, 则

$$y_{ijk} = \mu + \alpha_i + \tau_j + \alpha\tau_{ij} + \delta_{ik} + \varepsilon_{ijk}$$

其中

$\alpha_i$ : 因子 A 第  $i$  个水平的固定效应。

$\tau_j$ : 因子 T 第  $j$  个水平的固定效应。

$\alpha\tau_{ij}$ : 因子 A 第  $i$  个水平因子 T 第  $j$  个水平的固定效应。

$\delta_{ik}$ : 接受 A 的第  $i$  个水平的第  $k$  整区的随机效应。 $\delta_{ik}$  独立, 服从均值为 0 方差为  $\sigma_\delta^2$  的正态分布。

$\epsilon_{ijk}$ : 随机误差。 $\epsilon_{ijk}$  独立, 服从均值为 0 方差为  $\sigma_\epsilon^2$  的正态分布。 $\delta_{ik}$  和  $\epsilon_{ijk}$  是相互独立的。

此设计和模型的 AOV 表见表 17.31。

表 17.31 完全随机化裂区设计的 AOV

| 来源   | SS    | df                | EMS   |
|------|-------|-------------------|---|
| 整区之间 |       |                   |   |
| A    | SSA   | $a - 1$           | $\sigma_\epsilon^2 + t\sigma_\delta^2 + tn\theta_A$ |
| 整区误差 | SS(A) | $a(n - 1)$        | $\sigma_\epsilon^2 + t\sigma_\delta^2$              |
| 整区内  |       |                   |   |
| T    | SST   | $(t - 1)$         | $\sigma_\epsilon^2 + an\theta_T$                    |
| AT   | SSAT  | $(a - 1)(t - 1)$  | $\sigma_\epsilon^2 + n\theta_{AT}$                  |
| 小区误差 | SSE   | $a(n - 1)(t - 1)$ | $\sigma_\epsilon^2$                                 |
| 总计   | TSS   | $atn - 1$         |   |

当然可以用标准的公式计算 AOV 中的平方和, 但是我们建议利用计算机输出结果得到它们。根据期望均方我们有下面的分析。

#### 整区分析

$$H_0: \theta_A = 0 \text{ (或等价于 } H_0: \text{所有 } \alpha_i = 0), F = \frac{MSA}{MS(A)}$$

#### 小区分析

$$H_0: \theta_{AT} = 0 \text{ (或等价于 } H_0: \text{所有 } \alpha\tau_{ij} = 0), F = \frac{MSAT}{MSE}$$

$$H_0: \theta_T = 0 \text{ (或等价于 } H_0: \text{所有 } \tau_j = 0), F = \frac{MST}{MSE}$$

此设计的一个变化是引入一个区组因子(比如农场)。对我们的这个例子, 可能会有  $b = 2$  个农场, 每个农场有  $a = 2$  块整块地, 每整块地分为  $t = 3$  块小块地; 见图 17.5。因为整块地的随机化遵从随机化区组设计, 而每整块地内的小块地的随机化遵从完全随机化设计, 所以这个设计常被称为随机化区组裂区设计。

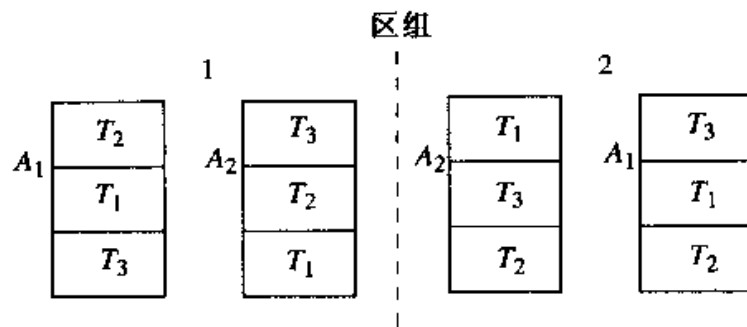


图 17.5 随机化区组裂区设计

这种更一般的有  $b$  个区组的两因子裂区设计的模型是：

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \tau_k + \alpha\tau_{ik} + \varepsilon_{ijk}$$

其中  $y_{ijk}$  表示在第  $j$  个区组内因子  $A$  第  $i$  个水平因子  $T$  第  $k$  个水平处的测量值。参数  $\alpha_i$ ,  $\tau_k$  和  $\alpha\tau_{ik}$  就是通常两因子试验中主效应和交互效应的参数；而  $\beta_j$  是区组  $j$  带来的效应,  $\alpha\beta_{ij}$  是第  $j$  个区组和因子  $A$  的第  $i$  个水平之间的交互效应。此模型的分析见表 17.32。此处我们假设因子  $A$  和  $T$  是固定效应而区组是随机的。

表 17.32 随机化区组裂区设计的 AOV( $A, T$  固定; 区组随机)

| 来源       | SS   | df                | EMS   |
|----------|------|-------------------|---|
| 整区之间     |      |                   |   |
| 区组       | SSB  | $b - 1$           | $\sigma_e^2 + a\sigma_\beta^2$                      |
| A        | SSA  | $a - 1$           | $\sigma_e^2 + t\sigma_{\alpha\beta}^2 + bt\theta_A$ |
| AB(整区误差) | SSAB | $(a - 1)(b - 1)$  | $\sigma_e^2 + t\sigma_{\alpha\beta}^2$              |
| 整区内      |      |                   |   |
| T        | SST  | $(t - 1)$         | $\sigma_e^2 + ab\theta_T$                           |
| AT       | SSAT | $(a - 1)(t - 1)$  | $\sigma_e^2 + b\theta_{AT}$                         |
| 小区误差     | SSE  | $a(b - 1)(t - 1)$ | $\sigma_e^2$  |
| 总计       | TSS  | $abt - 1$         |   |

表 17.32 中列出的各变异性来源的平方和是利用析因试验中主效应和交互效应的一般公式或者合适的软件包得到的。通过期望均方, 我们可以得到分析中的大块地部分里对因子  $A$  的有效  $F$  检验以及小块地部分对因子  $T$  和  $AT$  交互效应的  $F$  检验, 如下所示。注意没有对区组带来的变异性的检验。

### 整区分析

$$H_0: \theta_A = 0 \text{ (或等价于 } H_0: \text{所有 } \alpha_i = 0), F = \frac{MSA}{MSAB}$$

### 小区分析

$$H_0: \theta_{AT} = 0 \text{ (或等价于 } H_0: \text{所有 } \alpha\tau_{ik} = 0), F = \frac{MSAT}{MSE}$$

$$H_0: \theta_T = 0 \text{ (或等价于 } H_0: \text{所有 } \tau_k = 0), F = \frac{MST}{MSE}$$

#### 例 17.11

$b = 3$  个区组的两因子裂区设计中的大豆产量如下(以每小块地为单位,按蒲式耳计)。肥料(因子 A)随机地施加到每个农场内的整块地上,然后将大豆种类(因子 T)随机地分配到每整块地内的小块地上。利用下面的样本数据进行方差分析。给出每个检验的大约的  $p$  值。

| 1  |      |      | 2  |      |      | 3  |     |     |
|----|------|------|----|------|------|----|-----|-----|
| 肥料 |      |      | 肥料 |      |      | 肥料 |     |     |
| 种类 | 1    | 2    | 种类 | 1    | 2    | 种类 | 1   | 2   |
| 1  | 10.6 | 10.9 | 2  | 11.9 | 11.5 | 3  | 9.5 | 9.8 |
| 2  | 11.4 | 11.7 | 3  | 12.6 | 12.1 | 1  | 8.1 | 8.2 |
| 3  | 11.8 | 12.4 | 1  | 11.6 | 10.8 | 2  | 8.7 | 9.3 |

**解答** 对这些数据有  $a = 2, b = 3, t = 3$  和  $n = 1$ , 平方和如下(根据下面计算机输出结果中的 Type III SS 列):

$$SSA = 0.845 \quad SSAT = 0.00333$$

$$SSB = 28.863 \quad SSE = 0.227$$

$$SSAB = 0.0433 \quad TSS = 35.325$$

$$SST = 5.343$$

SPLIT-PLOT DESIGN: WHOLE PLOT TRT-FERT AND SPLIT-PLOT TRT-VARIETY

General Linear Models Procedure

Dependent Variable: Y YIELD

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|--------|----|----------------|-------------|---------|--------|
|--------|----|----------------|-------------|---------|--------|

|                 |    |            |           |        |        |
|-----------------|----|------------|-----------|--------|--------|
| Model           | 9  | 35.0983333 | 3.8998148 | 137.64 | 0.0001 |
| Error           | 8  | 0.2266667  | 0.0283333 |        |        |
| Corrected Total | 17 | 35.3250000 |           |        |        |

|          |          |          |         |
|----------|----------|----------|---------|
| R-Square | C.V.     | Root MSE | Y Mean  |
| 0.993583 | 1.570685 | 0.16833  | 10.7167 |

| Source | DF | Type III SS | Mean Square | F Value | Pr>F   |
|--------|----|-------------|-------------|---------|--------|
| F      | 1  | 0.8450000   | 0.8450000   | 29.82   | 0.0006 |
| B      | 2  | 28.8633333  | 14.4316667  | 509.35  | 0.0001 |
| B * F  | 2  | 0.0433333   | 0.0216667   | 0.76    | 0.4967 |
| V      | 2  | 5.3433333   | 2.6716667   | 94.29   | 0.0001 |
| V * F  | 2  | 0.0033333   | 0.0016667   | 0.06    | 0.9433 |

Dependent Variable: Y YIELD

Tests of Hypotheses using the Type III SS for B \* F as an error term

| Source | DF | Type III SS | Mean Square | F Value | Pr>F   |
|--------|----|-------------|-------------|---------|--------|
| F      | 1  | 0.84500000  | 0.84500000  | 39.00   | 0.0247 |

方差分析表为:

| 来源        | SS     | df | MS     | F     | p 值   |
|-----------|--------|----|--------|-------|-------|
| 整块地之间     |        |    |        |       |       |
| 区组        | 28.863 | 2  | 14.431 | —     | —     |
| A         | .845   | 1  | 0.845  | 39.00 | .0247 |
| AB(整块地误差) | .0433  | 2  | .0217  | —     | —     |
| 整块地内      |        |    |        |       |       |
| T         | 5.343  | 2  | 2.672  | 94.29 | .0001 |
| AT        | .00333 | 2  | .00167 | .060  | .9433 |
| 小块地误差     | .227   | 8  | .0283  | —     | —     |
| 总计        | 35.325 | 17 |        |       |       |

此处的两因子裂区设计和第十五章讨论的标准两因子设计的区别在于随机化的过程。在裂区设计中,随机化的过程分为两个阶段:首先在每个区组内将因子 A 的水平随机分配到整区中,然后再将因子 B 的水平随机分配到每区组内每整区



里的小区单元中。相反,在随机化区组设计的两因子试验(见 15.3 节)中,随机化仅是一步过程;在每个区组中直接将处理(即两个因子的因子水平组合)随机分配到试验单元中。包括有均值分解、比较,处理均值的估计和置信区间在内的 AOV 后的分析对于裂区设计来说比我们前面讨论过的设计更为复杂。关于这方面, Kuehl(1999), Snedecor 与 Cochran(1980)和 Lentner 与 Bishop(1993)都是很好的参考书。

## 17.7 小 结

就线性模型来说固定、随机和混合效应模型是很好区别的。固定效应模型是将响应变量和  $k \geq 1$  个独立变量及一个随机分量联系起来;而随机效应模型则是  $k = 0$  且有不只一个随机分量的一般线性模型。至于混合模型,它是固定效应和随机效应模型的组合体,将  $k \geq 1$  个独立变量以及多于一个的随机分量联系起来。

我们举例说明了随机效应模型在完全随机化设计和  $a \times b$  析因试验等情况下的应用,指明了随机效应模型和其相应的固定效应模型在进行方差分析中显著性检验时的类似之处。我们利用  $a \times b$  析因试验举例说明了从固定模型的方差分析中得到的推断。

遗憾的是,在入门课程中只能对随机和混合效应模型进行有限的讨论。为了扩充本课中的讨论,第 17.5 节给出的结论在推导平衡设计方差分析表中各变异性来源的期望均方时是很有用的。利用这些期望值,我们可以尝试构建适当的检验统计量来检验模型中任何固定效应或是随机效应的显著性。

关于随机或混合效应模型的各种问题中,最难的部分在于试图估计  $E(y)$ , 以及给出在随机效应模型时给出其适当的置信区间,或给出在混合模型时对于固定效应的某些水平或水平组合处  $y$  的平均值。我们举例说明了如何得到随机效应模型中  $E(y)$  的估计以及如何构建适当的置信区间。在混合模型中这个问题则变得更为复杂。

本章最后涉及到的题目是套设计和裂区设计,简单介绍了第十五章中讨论的和本章前面几节讨论的基本析因试验的几种变化。当在多因子试验的框架中需要考虑嵌套效应时,有很多更为常见的设计,而此处提供的设计只是其中的几种。感兴趣的读者可以查阅本书后面的参考资料,更详细地继续这个主题;特别 Kuehl(1999)是很好的参考书。

### 补充练习

17.8 辨别  $\theta_A$ (当因子  $A$  固定)的意义和  $\sigma_a^2$ (当因子  $A$  随机)的意义。

17.9 考虑一个  $n = 5$  次观测的  $2 \times 3 \times 4$  析因试验。假设因子  $A$  的水平是

固定的而因子  $B$  和  $C$  的水平是从水平的总体中随机选出的。

- 写出此试验的模型。确定模型中各量的意义并说明加在各量上的条件。
- 构建部分 AOV 表, 要包括自由度和所有变异性来源的期望均方。
- 给出对变异性来源显著性所做的所有适当的  $F$  检验的均方比值。

**17.10** 考虑一个  $n = 6$  次观测的  $3 \times 4 \times 2$  析因试验。假设因子  $A$  和  $B$  的水平是固定的而因子  $C$  的水平是从水平总体中随机选出的。

- 写出此试验的模型。确定模型中各量的意义并说明加在各量上的条件。
- 构建部分 AOV 表, 要包括  $df$  和所有变异性来源的期望均方。
- 给出对变异性来源显著性所做的所有适当的  $F$  检验的均方比值。

**17.11** 考虑一个  $n = 4$  次观测的  $3 \times 5 \times 5$  析因试验。假设因子  $B$  的水平是固定的而因子  $A$  和  $C$  的水平是从水平总体中随机选出的。

- 写出此试验的模型。确定模型中各量的意义并说明加在各量上的条件。
- 构建部分 AOV 表, 要包括  $df$  和所有变异性来源的期望均方。
- 给出对变异性来源显著性所做的所有适当的  $F$  检验的均方比值。

**17.12(环境)** 参见练习 17.6。假设这四种化学制剂是从数百种用于控制火蚁的化学剂中随机选出的。研究者想要确定一种化学制剂控制火蚁的效力是否随不同的环境变化。

a. 写出此种情况的适当模型。说明施加给此模型中各量的条件与当这几种化学制剂是仅有的研究者关心的化学制剂时的模型中的条件有何不同。

- 建立 AOV 表, 并检验所有相关的假设。
- 将得到的结论和推断与练习 17.6 中得到的做比较。

**17.13** 参见练习 17.12。

- 哪个模型和分析更合适? 解释你的答案。
- 哪种情况下固定效应模型更合适?
- 给出那些可以用精确  $F$  检验进行检验的变异性来源的检验统计量。

**17.14(工程)** 某大学的土木工程系得到一大笔资助研究校园内的交通问题, 并且要提出可供选择的解决方案。此研究的一小部分是要对每日穿过校园而又不使用学校设施的汽车计数。为此, 在每个入口处安置一组志愿者, 以记录每辆通过检查点汽车的牌照号和它进入或离开的时间。通过比较所有检查点的数据清单, 并给汽车穿过校园留出一段合理的时间, 组员们就可以确定在上午 8:00 到下午 5:00 间穿过校园而又不使用校园设施的汽车数量。从一学年中随机抽取 6 周做样本, 每周内取中间的两天的交通流量的数据如下。

| 周 1 | 周 2 | 周 3 | 周 4 | 周 5 | 周 6 |
|-----|-----|-----|-----|-----|-----|
| 680 | 438 | 539 | 264 | 693 | 530 |
| 618 | 520 | 600 | 198 | 646 | 575 |

a. 写出适当的线性统计模型, 确定模型中各项的意义。

b. 进行方差分析, 标明期望均方, 取  $\alpha = 0.05$ 。

**17.15** 参见练习 17.14, 估计随机选出的一周中间的一天内穿过但不使用学校设施的汽车的平均数目, 并给出大概的置信区间。(提示: 参见例 17.1)

**17.16(医药)** 设计了一项研究来考察新的治疗方法在降低高血压病人的收缩压时的效力。选出三种药品(D1, D2, D3)进行考察; 还有很多种用于降低血压的非药物治疗方法, 包括节食、锻炼、生物反馈等方法的各种组合, 研究者随机地选出了三种非药物治疗方法(ND1, ND2, ND3)用于研究。病人的年龄通常会阻碍各种治疗的效力, 所以, 高血压病人被分成了两个年龄组(A1, A2)。54 个病人被分到这两个年龄组并被随机指定一种药物和一种非药物治疗的组合。在参加研究 2 个月后, 记录下了每个病人收缩压比研究开始时血压读数的减少。下表给出了这些值。

|    | 年龄组 A1 |     |     | 年龄组 A2 |     |     |
|----|--------|-----|-----|--------|-----|-----|
|    | 非药物    |     |     | 非药物    |     |     |
|    | ND1    | ND2 | ND3 | ND1    | ND2 | ND3 |
| 药物 | 33     | 37  | 41  | 34     | 48  | 44  |
| D1 | 34     | 38  | 42  | 33     | 46  | 46  |
|    | 35     | 36  | 39  | 38     | 45  | 49  |
| 药物 | 46     | 44  | 43  | 47     | 44  | 44  |
| D2 | 45     | 48  | 44  | 49     | 48  | 46  |
|    | 46     | 49  | 45  | 45     | 46  | 41  |
| 药物 | 38     | 45  | 36  | 36     | 46  | 38  |
| D3 | 34     | 45  | 37  | 39     | 47  | 36  |
|    | 37     | 44  | 35  | 35     | 44  | 35  |

- a. 写出此研究的模型, 确定模型各项的意义并阐明各项必须满足的条件。
- b. 建立 AOV 表, 要包括期望均方。
- c. 检验所有有关的变异性来源的显著性, 取  $\alpha = 0.05$ 。
- d. 关于药物和非药物治疗方法的组合在治疗高血压时的效果是否有差异, 你能做出什么结论?

**17.17** 参见练习 15.3。假设我们认为这五个研究者是从可以从事火箭推进试验的所有研究者总体中随机选出的。

- a. 写出适当的线性统计模型, 确定各量的含义并列出假设。
- b. 进行方差分析, 在方差分析表中要包括期望均方列。

**17.18** 参见练习 17.17。对固定效应和随机效应模型, 指出检验的假设之间的区别和结论之间的区别。

**17.19** 参见练习 15.42。假设这两个实验室是从参加研究的实验室总体中随机选出的, 同时时间和温度也是变异性的可能来源。

- a. 计算所有变异性来源的期望均方。
- b. 检验所有变异性来源的显著性, 取  $\alpha = 0.05$ 。
- c. 比较此处得到的结果和练习 15.42 中得到的结果。
- d. 把实验室的效应看为随机效应比把它看为固定效应更合适吗? 解释你的答案。

**17.20** 参见练习 15.32。假设这五种窗格设计是从研究中窗格设计的总体中随机选出的。

- a. 计算所有变异性来源的期望均方。
- b. 检验所有变异性来源的显著性, 取  $\alpha = 0.05$ 。
- c. 比较此处得到的结果和练习 15.32 中得到的结果。
- e. 把窗格设计的效应看为随机效应比把它看为固定效应更合适吗? 解释你的答案。

**17.21** 参见练习 15.28 中描述的研究。

- a. 认为这九种药物是从可能使用的药物总体中随机选出的, 写出此时研究的模型。
- b. 给出所有变异性来源的期望均方。
- c. 说明你的分析和结论相对于练习 15.28 中的会如何变化?

**17.22(工程)** 在将计算机芯片焊牢在飞机导航系统的主板上时, 影响焊接的两个最关键的因素是插入焊料使用的机器以及操纵机器的操作员。从公司工厂的很多种机器和很多操作员中随机选出了四种焊接机器和三个操作员。每个操作员都用四种机器的每一种做两次焊接, 得到的焊接强度数据如下。

| 操作员 | 机器  |     |     |     |
|-----|-----|-----|-----|-----|
|     | 1   | 2   | 3   | 4   |
| 1   | 204 | 205 | 203 | 205 |
|     | 205 | 210 | 204 | 203 |
| 2   | 205 | 205 | 206 | 209 |
|     | 207 | 206 | 204 | 207 |
| 3   | 211 | 207 | 209 | 215 |
|     | 209 | 210 | 214 | 212 |

- 写出此研究的模型,要包括模型中的各量以及各量需满足的条件。
- 给出 AOV 表,要包括期望均方。
- 对于机器和操作员对焊接强度变异性的效应,你能得到什么结论?

**17.23** 参见练习 17.22。

- 估计此研究中的方差分量。
- 对焊接强度的总变异性,按比例分配给各变异性来源。
- 给出平均焊接强度的 95% 置信区间。

**17.24(环境)** 对一个地域能否进行烟煤的表层开采做调查,从这个地域的六个地点分别取得了岩心土壤的样本。将每份岩心样本再分为四个子样本以进行样本中硫含量的独立的分析。

- 该设计是什么设计? 给出这个试验的模型。
- 给出 AOV 中的变异性来源和各量的自由度。

**17.25** 练习 17.24 的样本数据如下。进行方差分析并给出结论,取  $\alpha = .05$ 。

| 地点 | 分析   |      |      |      |
|----|------|------|------|------|
|    | 1    | 2    | 3    | 4    |
| 1  | 15.2 | 16.8 | 17.5 | 16.2 |
| 2  | 13.1 | 13.8 | 12.6 | 12.9 |
| 3  | 17.5 | 17.1 | 16.7 | 16.5 |
| 4  | 18.3 | 18.4 | 18.6 | 17.9 |
| 5  | 12.8 | 13.6 | 14.2 | 14.0 |
| 6  | 13.5 | 13.9 | 13.6 | 14.1 |

**17.26(工程)** 药片硬度是同一种药品的不同配方之间的一个比较指标;一个配方中成分的某些组合(除去能起作用的药物外的成分)可能会导致药片比的别的

组合时硬度更高。假设随机地选出一个配方下的三批产品用于研究。从每批产品中都随机选出三份 1 - kg 重的药片的样本,再从每份 1 - kg 样本中随机选出 7 片药片做检验。硬度的读数见下。

| 样本 | 产品批 1 |    |     | 产品批 2 |     |     | 产品批 3 |    |    |
|----|-------|----|-----|-------|-----|-----|-------|----|----|
|    | 1     | 2  | 3   | 1     | 2   | 3   | 1     | 2  | 3  |
|    | 85    | 76 | 95  | 108   | 117 | 101 | 71    | 81 | 72 |
|    | 94    | 87 | 98  | 100   | 106 | 108 | 85    | 70 | 68 |
|    | 91    | 90 | 94  | 105   | 103 | 100 | 78    | 84 | 80 |
|    | 98    | 91 | 96  | 109   | 109 | 99  | 68    | 83 | 72 |
|    | 85    | 88 | 99  | 104   | 100 | 117 | 85    | 72 | 75 |
|    | 96    | 94 | 100 | 102   | 104 | 109 | 67    | 81 | 79 |
|    | 93    | 96 | 93  | 108   | 102 | 105 | 76    | 78 | 74 |

- 指出该设计是什么设计。
- 给出适当的模型和假设。
- 给出 AOV 中的变异性来源和自由度。
- 进行方差分析并对此研究中的配方的药片硬度数据做出结论。取  $\alpha = .05$ 。

**17.27(心理学)** 一名研究者正在设计一个试验来比较一种肉制品的九种不同配方。一个因子 F 是肉中的脂肪含量(10%, 15%, 20%);另一个因子 C 是烹制方法(电烤, 火烤, 油煎)。她准备好九种组合中每种样品,让品尝者根据各种标准给这些样品打分。研究中有四个品尝者;每个品尝者要尝九个样品。品尝者和品尝者之间有差别,但是样本被品尝的顺序对得分不会有影响。样品将用如下方法来准备以保证可以得到肉制品的样品同时可以为品尝者保温。将脂肪含量为 10% 的肉制品分为相同的三份,然后每种烹制方法被随机地分配到一份上。对于脂肪含量为 15% 和 20% 的肉制品重复这个过程。然后由一个品尝者品尝这九个样品并给它们打分。对其他三个品尝者重复这整个过程。得到的分数如下。

| 品尝者   | 脂肪含量 10% |    |    | 脂肪含量 15% |    |    | 脂肪含量 20% |    |    |
|-------|----------|----|----|----------|----|----|----------|----|----|
|       | 电烤       | 火烤 | 油煎 | 电烤       | 火烤 | 油煎 | 电烤       | 火烤 | 油煎 |
| 品尝者 1 | 75       | 79 | 82 | 78       | 82 | 81 | 81       | 85 | 87 |
| 品尝者 2 | 61       | 65 | 74 | 65       | 73 | 80 | 75       | 81 | 86 |
| 品尝者 3 | 75       | 78 | 79 | 80       | 82 | 83 | 87       | 88 | 92 |
| 品尝者 4 | 73       | 76 | 78 | 75       | 81 | 90 | 86       | 92 | 93 |

- 指出该设计的类型。
- 给出适当的模型和假设。
- 给出 AOV 中的变异性来源和自由度。
- 进行方差分析并对脂肪含量和烹制方法对肉制品味道的效应做出结论, 取  $\alpha = 0.05$ 。

下面给出了此数据集的计算机输出结果。

General Linear Models Procedure

|                      |                 | Class          | Levels      | Values         |                   |       |
|----------------------|-----------------|----------------|-------------|----------------|-------------------|-------|
|                      |                 | TASTER         | 4           | 1 2 3 4        |                   |       |
|                      |                 | FAT            | 3           | 10 15 20       |                   |       |
|                      |                 | METHOD         | 3           | BA BR F        |                   |       |
| Dependent Variable;Y |                 | SCORE          |             |                |                   |       |
| Source               | DF              | Sum of Squares | Mean Square | F Value        | Pr>F              |       |
| Model                | 17              | 931.666667     | 54.803922   | 5.97           | 0.0002            |       |
| Error                | 18              | 165.333333     | 9.185185    |                |                   |       |
| Corrected Total      | 35              | 1097.000000    |             |                |                   |       |
|                      |                 | R-Square       | C.V.        | Root MSE       | Y Mean            |       |
|                      |                 | 0.849286       | 3.780508    | 3.03071        | 80.1667           |       |
| Source               | DF              | Type III SS    | Mean Square | F Value        | Pr>F              |       |
| T                    | 3               | 291.888889     | 97.296296   | 10.59          | 0.0003            |       |
| F                    | 2               | 146.000000     | 73.000000   | 7.95           | 0.0034            |       |
| T * F                | 6               | 451.777778     | 75.296296   | 8.20           | 0.0002            |       |
| M                    | 2               | 32.166667      | 16.083333   | 1.75           | 0.2019            |       |
| F * M                | 4               | 9.833333       | 2.458333    | 0.27           | 0.8949            |       |
| Level of FAT         | Level of METHOD | N              | MEAN SCORE  | FAT MEAN SCORE | METHOD MEAN SCORE |       |
| 10                   | BA              | 4              | 80.75       | 10 80.25       | BAKE              | 81.75 |
| 10                   | BR              | 4              | 78.75       | 15 79.75       | BROIL             | 80.33 |
| 10                   | F               | 4              | 81.25       | 20 84.25       | FRY               | 82.17 |
| 15                   | BA              | 4              | 80.25       |                |                   |       |
| 15                   | BR              | 4              | 79.00       |                |                   |       |
| 15                   | F               | 4              | 80.00       |                |                   |       |
| 20                   | BA              | 4              | 84.25       |                |                   |       |
| 20                   | BR              | 4              | 83.25       |                |                   |       |
| 20                   | F               | 4              | 85.25       |                |                   |       |

## 第十八章 重复测量与交叉设计

### 18.1 引言和案例

### 18.2 有重复观测的单因子试验

### 18.3 一个因子有重复观测的两因子试验

### 18.4 交叉设计

### 18.5 小结

## 18.1 引言和案例

本书中到目前为止所讨论的所有试验(配对差异试验除外)中,我们都假定在每一个试验单元上只有一个观察值。例如,在一个比较三种不同的心血管药物对于血压影响的试验中,我们可以使用一个单因子试验,其中  $n_1$  个患者指定使用药物 1,  $n_2$  个患者指定使用药物 2,  $n_3$  个患者指定使用药物 3。此时,模型为

$$y = \mu + \alpha_i + \varepsilon_{ij}$$

其中  $\alpha_i$  是药物  $i$  的(固定或随机)效应,  $\varepsilon_{ij}$  表示用药物  $i$  治疗的第  $j$  个患者的随机效应。对于这个设计,从每个患者我们只能得到一个观测值。

许多从事应用研究的实际工作人员从费用和效率的角度出发,认为应该从每个试验单元获取多于一个的观测值。比如,在进行临床研究时,寻找满足研究要求并且愿意参加临床试验的患者往往很困难。因此,一旦找到了适当数量的患者,尽可能多地从这些患者获取信息这一点是很重要的。在这一章中,我们将考虑几种不同的包含一个或多个因子以及重复观测的试验。

在交叉设计的试验中,每个研究对象接受所有的处理。这种研究中的单个对象用做区组,从而降低了试验的误差。与每个对象只接受一个处理的设计相比,这种设计提高了处理比较的精度。在重复观测设计的试验中,随着指定的处理的执行,在  $t$  个不同的时间点上,我们获得  $t$  次不同的观测的值。这样的试验由表 18.1 表示,其中  $y_{ijk}$  表示第  $j$  个患者用第  $i$  种药物治疗时第  $k$  次的观测值。注意,对于每个患者,我们有  $t > 1$  个,而不是 1 个观测值。

对同一个研究对象在不同时间进行多次观察,往往比每一次都对不同的对象进行观察能更有效地利用试验资源。比如,需要较少的研究对象,从而降低试验费用。再如,可以得到很精确的时间趋势的估计。本章中所给出的方法,既可以分析从交叉研究中得到的数据,也可以分析从重复测量研究所得到的数据。这两种设计的应用背景都非常宽阔,其应用广泛存在于制药工业、研究开发和多数工业的制造过程中。药物研究者、生态学研究者以及其他众多研究领域中的工作者都要



评估时间的趋势,因此会发现重复测量设计是有用的。在研究中,当数据中的时间趋势代之以空间关系时,这些设计的扩充版本可能会是适宜的。这样的例子有带状煤田的勘测,石油溢漏效应的评估和工业设施周围空气污染的评估。对于在空间中有重复测量的研究,其模型一般来说比本章中所研究的时间趋势更加复杂。关于空间数据的模型的进一步的内容,可以在 Ripley(1976), Haining(1990) 和 Cressie(1993)中找到。

表 18.1 每个患者重复的时间点

| 药物 | 时 间         |             |     |             |
|----|-------------|-------------|-----|-------------|
|    | 1           | 2           | ... | $t$         |
| 1  | $y_{111}$   | $y_{112}$   | ... | $y_{11t}$   |
|    | $\vdots$    | $\vdots$    |     | $\vdots$    |
|    | $y_{1n_11}$ | $y_{1n_12}$ | ... | $y_{1n_1t}$ |
|    |             |             |     |             |
| 2  | $y_{211}$   | $y_{212}$   | ... | $y_{21t}$   |
|    | $\vdots$    | $\vdots$    |     | $\vdots$    |
|    | $y_{2n_21}$ | $y_{2n_22}$ | ... | $y_{2n_2t}$ |
|    |             |             |     |             |
| 3  | $y_{311}$   | $y_{312}$   | ... | $y_{31t}$   |
|    | $\vdots$    | $\vdots$    |     | $\vdots$    |
|    | $y_{3n_31}$ | $y_{3n_32}$ | ... | $y_{3n_3t}$ |
|    |             |             |     |             |

下面的案例说明了如何在有重复测量的设计中评估时间趋势。

#### 案例:石油泄漏对植物生长的影响

我们在第六章中考察了这个案例的一小部分。1992年1月7日,一条地下输油管道破裂,引起得克萨斯州 San Patricio 县 Chiltipin Creek 地区一片湿地的污染。清除工作包括燃烧这片湿地上的污染区域。为了评估溢油对这片湿地上植物的影响,研究人员设计了一项关于在按照某种方案燃烧后植物生长情况的研究。在一篇没有公开发表的得克萨斯 A & M 大学的博士论文中, Newman(1997)描述了研究者们就 *Distichlis spicata*, 一种对该溢油地区非常重要的植物的发现。

**数据收集的设计** 对于研究人员来说,两个重要的问题是:

1. 溢油和燃烧以后,有油的地区是否恢复了?
2. 实现恢复花了多长时间?

为了回答这些问题,研究人员需要有一个基点,以作为在燃烧数月后溢油地区的 *Distichlis spicata* 密度比较的参照。这种植物的密度依赖于土壤的特征,土地的斜度,环境条件,天气以及许多其他因素。研究人员挑选了附近一片土地,以作为对照。这块土地没有受到溢油的影响,但又与溢油地点有相似的土壤和环境条件。在溢油地点和对照地点,各随机选择了 20 块土地,在经过了 9 个月以后,大约每三个月观测一次,共观测了 8 次。在这期间,分别记录了这 40 块土地上 *Distichlis spicata* 的数量。

这个试验的设计是重复观测设计,有两个处理,即溢油地点和对照地点,每块土地有 8 次观测,历时两年。为了回答研究人员的那两个问题,我们将利用 *Distichlis spicata* 的数量。因此,我们的研究假设可叙述如下。

**研究假设 1** 研究期间,溢油土地和对照土地上 *Distichlis spicata* 的平均密度有无差别?

**研究假设 2** 研究期间,*Distichlis spicata* 的平均密度有无显著的趋势?

**研究假设 3** 溢油土地和对照土地上的趋势是否一致?

**数据的整理** 数据包括各块对照土地和燃烧后的(溢油)土地上 8 次观察的 *Distichlis spicata* 的植株数量,总共有 320 个数据值。研究人员为确定在现场记录的数据中是否含有明显的错误,对数据进行了检查。起初,研究人员决定删除第 1 和第 20 块对照土地,因为在 1992 年 10 月期间观察到植株数量很少。然而,为了不使抽样计划产生偏差,他们最终还是保留了这两块土地。然后,按照 2.5 节中谈到的步骤,数据被转化成计算机文件,为分析数据做好准备。数据在表 18.2 中给出。

表 18.2 两种处理下 *Distichlis spicata* 的数量

| 处理 | 地块 | 1992 年<br>10 月 | 1993 年<br>7 月 | 1993 年<br>10 月 | 1994 年<br>1 月 | 1994 年<br>4 月 | 1994 年<br>7 月 | 1994 年<br>10 月 | 1995 年<br>1 月 |
|----|----|----------------|---------------|----------------|---------------|---------------|---------------|----------------|---------------|
| 燃烧 | 1  | 27             | 25            | 18             | 21            | 26            | 22            | 20             | 27            |
|    | 2  | 5              | 15            | 10             | 12            | 10            | 11            | 12             | 9             |
|    | 3  | 17             | 26            | 26             | 25            | 15            | 10            | 14             | 17            |
|    | 4  | 41             | 41            | 42             | 38            | 34            | 26            | 26             | 25            |
|    | 5  | 25             | 28            | 22             | 27            | 24            | 16            | 18             | 23            |
|    | 6  | 11             | 24            | 13             | 20            | 16            | 13            | 10             | 14            |
|    | 7  | 37             | 40            | 33             | 31            | 32            | 30            | 25             | 31            |
|    | 8  | 38             | 38            | 33             | 38            | 39            | 35            | 32             | 38            |
|    | 9  | 31             | 33            | 25             | 30            | 28            | 21            | 17             | 19            |

续表

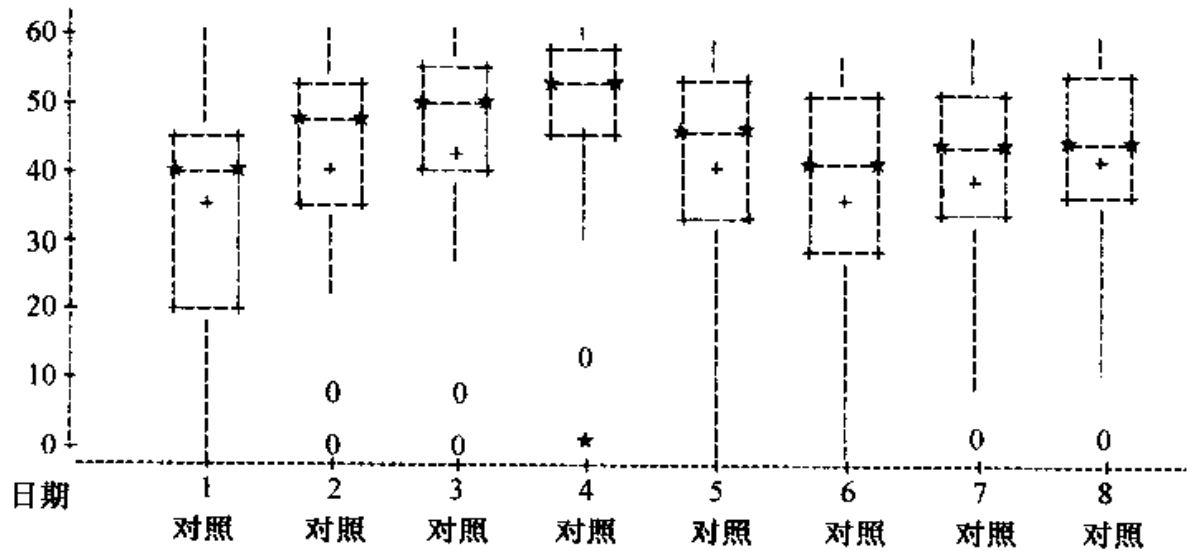
| 处理 | 地块 | 1992 年<br>10 月 | 1993 年<br>7 月 | 1993 年<br>10 月 | 1994 年<br>1 月 | 1994 年<br>4 月 | 1994 年<br>7 月 | 1994 年<br>10 月 | 1995 年<br>1 月 |
|----|----|----------------|---------------|----------------|---------------|---------------|---------------|----------------|---------------|
| 燃烧 | 10 | 24             | 25            | 21             | 24            | 24            | 19            | 17             | 22            |
|    | 11 | 22             | 27            | 31             | 30            | 32            | 30            | 25             | 34            |
|    | 12 | 26             | 45            | 39             | 35            | 35            | 36            | 30             | 27            |
|    | 13 | 32             | 38            | 34             | 45            | 41            | 28            | 31             | 31            |
|    | 14 | 35             | 37            | 35             | 42            | 35            | 32            | 27             | 29            |
|    | 15 | 26             | 23            | 19             | 18            | 21            | 13            | 11             | 19            |
|    | 16 | 22             | 29            | 24             | 24            | 20            | 16            | 18             | 24            |
|    | 17 | 50             | 54            | 56             | 60            | 51            | 52            | 49             | 52            |
|    | 18 | 17             | 29            | 23             | 39            | 31            | 24            | 26             | 34            |
|    | 19 | 25             | 37            | 29             | 32            | 28            | 14            | 13             | 24            |
|    | 20 | 33             | 39            | 39             | 48            | 36            | 34            | 30             | 34            |
| 处理 | 地块 | 1992 年<br>10 月 | 1993 年<br>7 月 | 1993 年<br>10 月 | 1994 年<br>1 月 | 1994 年<br>4 月 | 1994 年<br>7 月 | 1994 年<br>10 月 | 1995 年<br>1 月 |
| 对照 | 1  | 7              | 0             | 0              | 1             | 0             | 0             | 0              | 0             |
|    | 2  | 57             | 46            | 49             | 51            | 48            | 43            | 40             | 40            |
|    | 3  | 43             | 59            | 59             | 60            | 58            | 53            | 55             | 58            |
|    | 4  | 43             | 53            | 52             | 53            | 53            | 53            | 52             | 54            |
|    | 5  | 59             | 55            | 59             | 60            | 54            | 47            | 54             | 53            |
|    | 6  | 42             | 48            | 50             | 48            | 43            | 37            | 38             | 38            |
|    | 7  | 35             | 42            | 50             | 55            | 41            | 40            | 44             | 45            |
|    | 8  | 40             | 51            | 53             | 57            | 53            | 38            | 43             | 36            |
|    | 9  | 24             | 52            | 54             | 59            | 57            | 55            | 57             | 39            |
|    | 10 | 42             | 49            | 50             | 54            | 51            | 44            | 39             | 41            |
|    | 11 | 16             | 31            | 39             | 47            | 24            | 22            | 33             | 35            |
|    | 12 | 54             | 58            | 60             | 60            | 54            | 51            | 48             | 51            |
|    | 13 | 30             | 43            | 43             | 47            | 39            | 36            | 49             | 56            |
|    | 14 | 47             | 50            | 60             | 60            | 54            | 52            | 57             | 57            |
|    | 15 | 40             | 40            | 47             | 49            | 43            | 41            | 48             | 52            |
|    | 16 | 11             | 23            | 27             | 31            | 17            | 19            | 24             | 29            |
|    | 17 | 41             | 45            | 42             | 44            | 41            | 33            | 31             | 42            |
|    | 18 | 50             | 52            | 55             | 53            | 45            | 42            | 35             | 51            |
|    | 19 | 8              | 8             | 7              | 12            | 6             | 5             | 8              | 10            |
|    | 20 | 0              | 0             | 0              | 1             | 0             | 0             | 0              | 0             |

**数据的分析** 研究的下一步是概括数据并画出数据的图。按处理和日期计算的平均植物数量见表 18.3。

表 18.3 按处理和日期计算的平均植物数量

| 处理  | 1992 年<br>10 月 | 1993 年<br>7 月 | 1993 年<br>10 月 | 1994 年<br>1 月 | 1994 年<br>4 月 | 1994 年<br>7 月 | 1994 年<br>10 月 | 1995 年<br>1 月 |
|-----|----------------|---------------|----------------|---------------|---------------|---------------|----------------|---------------|
| 燃烧后 | 27.20          | 32.65         | 28.60          | 31.95         | 28.90         | 24.10         | 22.55          | 26.65         |
| 对照  | 34.45          | 40.25         | 42.80          | 45.10         | 39.05         | 35.55         | 37.75          | 39.35         |

植物数量



植物数量

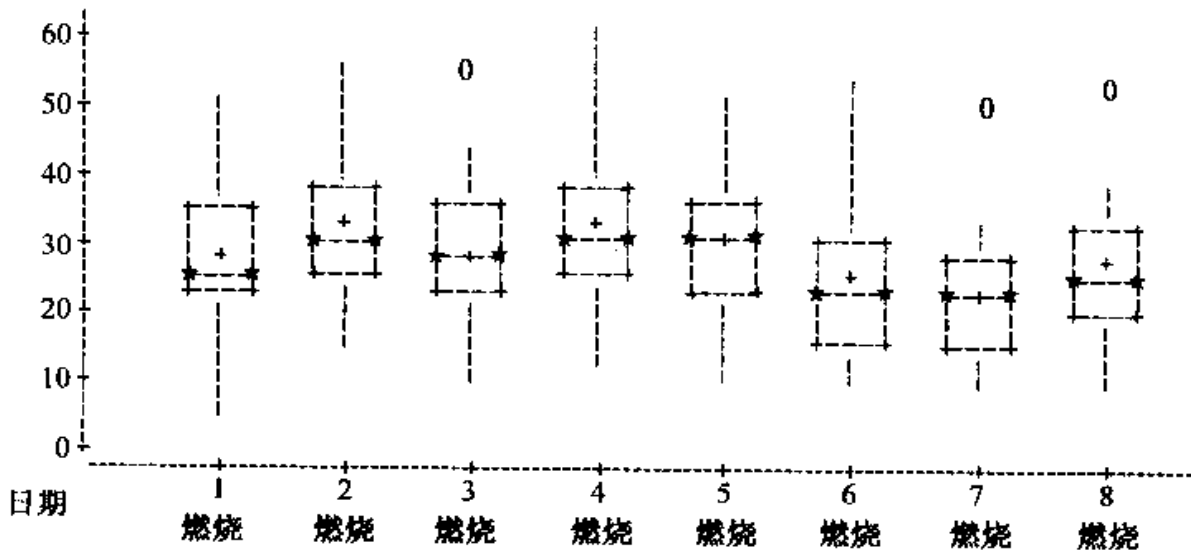


图 18.1 按处理和日期画出的植物数量的盒形图

对于每一个日期和每一个处理,植物数量的盒形图如图 18.1 所示。这些盒形图表明,对照地块的植物数量的中位数比溢油地块的相应值来得大。然而,对照地块之间的差别比溢油地块之间的差别大一些。这可能是由于在溢油地块上进行燃烧处理的缘故,燃烧通常使得各地块的条件比燃烧前更加一致。要知道这一点能否扩展至地块的总体而不仅仅是对于在研究中观测到的地块,需要对数据拟合模型,并检验相应的统计研究假设。我们将在讨论一般重复观测设计的分析方法之后,在本章的末尾再对这些数据进行分析。

## 18.2 有重复观测的单因子试验

在上一节,我们讨论了为什么会需要对每个患者有一个以上的观察值。对每个患者需要有一个以上观察值的另一个原因是不同患者之间的变异性常常比同一个患者上的变异性大得多。我们在 6.5 节配对  $t$  检验的例子中曾经看到过这一点。如果是这种情况,那么把患者作为区组,给每一个患者每一种治疗可能要好些。这样,药物之间的比较是患者内的比较,而不是像给药物  $i$  指定  $n_i$  个不同的患者的单因子试验中那样,比较是在患者间的。反映这种患者内比较的特性的单因子试验可由表 18.4 表示。

表 18.4 药物 1, 2, 3 的患者内的比较

| 药物 | 患 者      |          |     |          |
|----|----------|----------|-----|----------|
|    | 1        | 2        | ... | $n$      |
| 1  | $y_{11}$ | $y_{12}$ | ... | $y_{1n}$ |
| 2  | $y_{21}$ | $y_{22}$ | ... | $y_{2n}$ |
| 3  | $y_{31}$ | $y_{32}$ | ... | $y_{3n}$ |

在这个设计中,对于  $n$  个患者中的每一个,都依次安排了这三种药物。在一个给定的治疗时期,一种药物被安排给一名患者。在经过一个充分长的“清洗”时间后,再安排另外一种给该患者。重复这个过程,直到这名患者接受了所有三种药物的治疗。三种药物治疗的顺序是随机安排的。该设计中的关键是两种药物治疗之间的清洗时间要充分长,使得一种药物的结果不影响另一种药物的结果。

由此,我们可从每一位患者得到多于一个的观测值,并设想对于研究中的三种药物能得到更多有用的信息。对于这样的试验,一个模型为

$$y_{ij} = \mu + \alpha_i + \pi_j + \varepsilon_{ij}$$

其中  $\mu$  为总平均响应,  $\alpha_i$  为第  $i$  种药物的效应,  $\pi_j$  为第  $j$  个患者的效应,  $\varepsilon_{ij}$  为接受第  $i$  种药物治疗的第  $j$  个患者响应的试验误差。

注意, 这个模型看上去与任何其他有  $a$  种药物和  $n$  个患者的单因子试验的模型没什么两样。然而, 模型假设是不同的, 因为我们从每个患者得到多于一个的观测值。对于这个模型, 我们做如下假设。

1.  $\alpha_i$  是一个常数, 并且  $\sum \alpha_i = 0$ 。
2. 诸  $\pi_j$  是相互独立的, 并且服从正态分布  $(0, \sigma_\pi^2)$ 。
3. 诸  $\varepsilon_{ij}$  与诸  $\pi_j$  是相互独立的。
4. 诸  $\varepsilon_{ij}$  服从正态分布  $(0, \sigma_\varepsilon^2)$ 。
5. 各个  $\varepsilon_{ij}$  间的相关关系如下:

对于  $i \neq i'$ ,  $\varepsilon_{ij}$  和  $\varepsilon_{i'j}$  是相关的;

对于  $j \neq j'$ ,  $\varepsilon_{ij}$  和  $\varepsilon_{ij'}$  是独立的。

也就是说, 从同一名患者观测到的两个观察值是相关的, 而从不同患者观测到的观察值是相互独立的。由这些假设可以得出,  $y_{ij}$  的方差为  $\sigma_\pi^2 + \sigma_\varepsilon^2$ 。一个更进一步的假设是从患者  $j$  观测到的任何两个观察值  $y_{ij}$  和  $y_{i'j}$  的协方差是一个常数。这些假设可以视为对观察值的方差-协方差矩阵的要求, 此时的协方差矩阵具有复合对称性。有关相关观察值的讨论超出了本书的范围, 有兴趣的读者请参考 Kuehl (1999) 以及 Vonesh and Chinchilli (1997)。

表 18.5 表 18.4 中给出的试验的方差分析表

| 来源  | SS  | df               | EMS(A 固定, 患者随机)                        |
|-----|-----|------------------|--|
| 患者间 | SSP | $n - 1$          | $\sigma_\varepsilon^2 + a\sigma_\pi^2$ |
| 患者内 |     |                  |  |
| A   | SSA | $a - 1$          | $\sigma_\varepsilon^2 + n\theta_A$     |
| 误差  | SSE | $(a - 1)(n - 1)$ | $\sigma_\varepsilon^2$                 |
| 总和  | TSS | $an - 1$         |  |

这种试验设计的方差分析和上述这一组假设由表 18.5 给出。读者当熟记在心。当假设成立, 从而符合对称性成立时, 可以对因子 A 的效应进行统计检验 ( $F = \text{MSA}/\text{MSE}$ )。然而, 要使得用  $F = \text{MSA}/\text{MSE}$  检验因子 A 的效应的 F 检验有效, 还有更为一般的条件。那么, 上述假设有多么苛刻呢? 怎样才能知道何时检验是适宜的呢?

对这些问题没有简单的答案,因为没有简单的方法可用来检查复合对称性。上述提到的使得因子  $A$  的  $F$  检验有效的更为一般的条件(称为 Huynh-Feldt 条件)常常由于从同一患者观测到的时间相距较近的观察值之间的相关性比相距较远的观察值之间的相关性来得强而不能得到满足。所以,使用时需当心这一点。一般来说,当方差-协方差矩阵没有复合对称性时,因子  $A$  的  $F$  检验会有一个正偏差,这将导致拒绝  $H_0$ :所有  $\alpha_i = 0$  的可能性比决定临界  $F$ -值的检验水平高一些。

从实用的角度看,在一个给定的试验环境下,最好的办法是确保在实施两个不同的治疗之间留出充分的时间,以清洗(或摆脱)前次治疗的影响,同时确保这种设计仅仅应用于疾病相对稳定的情形,以保证在接受治疗和清洗以后,每个患者(或试验单元)基本上和接受治疗前一样。例如,即使在研究降血压药的效果时,我们也期望高血压症足够稳定,患者在接受了第一种药物治疗并经过清洗阶段以后,在继续接受第二种指定的药物治疗前,其血压能够恢复到用药前的水平,如此等等。

在 18.3 中,我们还将继续讨论如何判断使检验有效的假设是否成立,以及当这些假设不成立时,我们应该怎么办。关于这方面的进一步的内容,请参阅含有对重复观察试验进行细致讨论的更高水平的教科书(如 Kuehl(1999)以及 Vonesh and Chinchilli(1997))。

### 18.3 一个因子有重复观测的两因子试验

我们可以把有重复观测试验的讨论推广至两个因子的情形。例如,在比较心血管药物的降血压的效果时,我们可能随机化患者,使得  $n$  个不同的患者接受三种药物的每一种。然而,不是使每一个患者接受每一种药物,而是我们可能对于每一个患者在一定时间段内测量多次。例如,我们可能希望在患者服用指定的药物之前的一刻测量他的血压,然后在服药后的第一个小时内每 15 分钟测量一次,在接下来的 6 个小时内每一小时测量一次。这种类型的试验的数据如表 18.1 所示。注意,这是一个有重复测量的两因子(药物和时间)试验,重复测量是对两个因子中的一个(时间)进行的。

对于一个比较因子  $A$ (药物)的  $a$  个水平的两因子试验,其中因子  $A$  的每个水平有  $n$  个患者,因子  $B$ (时间)有  $b$  个水平,我们将使用如下模型:

$$y_{ijk} = \mu + \alpha_i + \pi_{j(i)} + \beta_k + \alpha\beta_{ik} + \varepsilon_{ijk}$$

其中  $\alpha_i, \beta_k, \alpha\beta_{ik}$  均为固定效应,分别相应于因子  $A$ (药物)的主效应,因子  $B$ (时间)的主效应和二者的交互效应。 $\pi_{j(i)}$  为因子  $A$  的第  $i$  个水平上第  $j$  个患者的随机效应。我们假定各个  $\pi_{j(i)}$  相互独立,并且都服从正态分布  $(0, \sigma_\pi^2)$ 。

表 18.6 其中一个因子有重复测量的两因子试验的方差分析

| 来源     | SS     | df                | EMS(A, B 固定, 患者随机)                        |
|--------|--------|-------------------|---|
| 患者间    |        |                   |   |
| A      | SSA    | $a - 1$           | $\sigma_e^2 + b\sigma_\pi^2 + bn\theta_A$ |
| A 中的患者 | SSP(A) | $a(n - 1)$        | $\sigma_e^2 + b\sigma_\pi^2$              |
| 患者内    |        |                   |   |
| B      | SSB    | $b - 1$           | $\sigma_e^2 + an\theta_B$                 |
| AB     | SSAB   | $(a - 1)(b - 1)$  | $\sigma_e^2 + n\theta_{AB}$               |
| 误差     | SSE    | $a(b - 1)(n - 1)$ | $\sigma_e^2$                              |
| 总和     | TSS    | $abn - 1$         |   |

基于这些假设, 我们得到如表 18.6 给出的方差分析表。从表 18.6 可以清楚地看出, 我们可以进行如下的检验:

$$1. H_0: \theta_{AB} = 0$$

$$F = \frac{MSAB}{MSE}$$

$$2. H_0: \theta_B = 0$$

$$F = \frac{MSB}{MSE}$$

$$3. H_0: \theta_A = 0$$

$$F = \frac{MSA}{MSP(A)}$$

### 例 18.1

有 10 名志愿者同意参加一项研究, 以确定同一种药品的同一剂量在两种不同的药剂形式(胶囊和片剂)下血流中药物的浓度。假定在一定的范围内, 浓度越高药品越有效。其中 5 个人被随机地指定服用胶囊(C), 另外 5 人则服用片剂(T)。他们从白天晚上 8 点开始禁食, 第二天早晨 8 点服药, 服药 4 小时以后才能饮食。在服药后 0.5, 1, 2, 3 和 4 小时时分别取了(15mL)血样, 经分析得到血流中药物的浓度。数据(ng/mL)给出如下。

- 对每一种药剂形式画出平均样本数据(时间 - 响应)的图。在这段时间内, 两种药剂形式显示出了不同的效用吗?
- 进行有重复测量的方差分析。



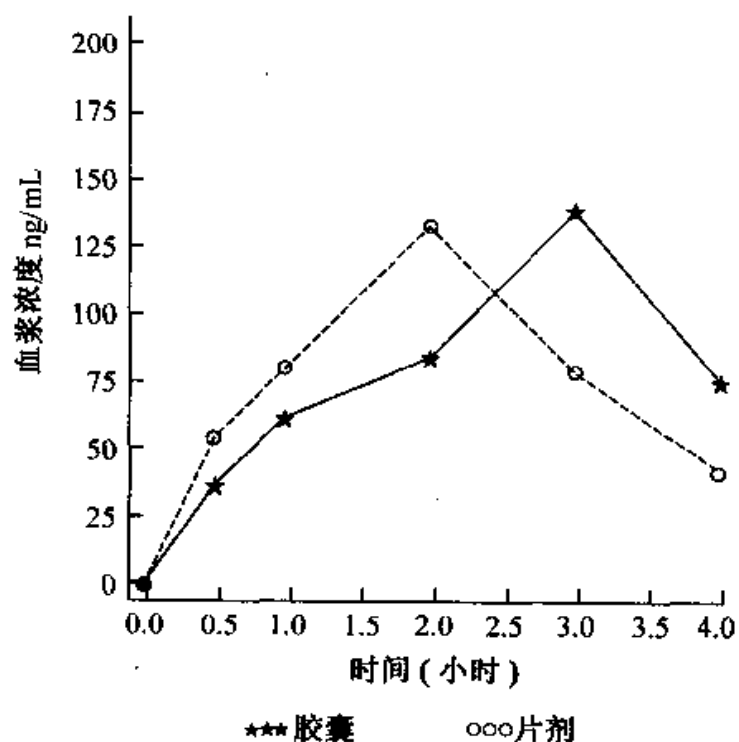
| 患者 | 片剂时间 |    |     |    |    | 患者 | 胶囊时间 |    |    |     |    |
|----|------|----|-----|----|----|----|------|----|----|-----|----|
|    | 0.5  | 1  | 2   | 3  | 4  |    | 0.5  | 1  | 2  | 3   | 4  |
|    | 50   | 75 | 120 | 60 | 30 | 1  | 30   | 55 | 80 | 130 | 65 |
|    | 40   | 80 | 135 | 70 | 40 | 2  | 25   | 50 | 75 | 125 | 60 |
|    | 55   | 75 | 125 | 85 | 50 | 3  | 35   | 65 | 85 | 140 | 85 |
|    | 70   | 85 | 140 | 90 | 40 | 4  | 45   | 70 | 90 | 145 | 80 |
|    | 60   | 90 | 150 | 95 | 50 | 5  | 50   | 75 | 95 | 160 | 90 |

### 解答

a. 按药剂形式和时间计算的平均血浆浓度如下。

| 形式 | 时 间 |    |     |     |    |
|----|-----|----|-----|-----|----|
|    | 0.5 | 1  | 2   | 3   | 4  |
| C  | 37  | 63 | 85  | 140 | 76 |
| T  | 55  | 81 | 134 | 80  | 42 |

血浆浓度的图给出如下。



b.

| 来源     | SS     | df | MS    | F      | p 值    |
|--------|--------|----|-------|--------|--------|
| 患者间    |        |    |       |        |        |
| 形式     | 40.5   | 1  | 40.5  | 0.083  | 0.7810 |
| 形式中的患者 | 3 920  | 8  | 490   |        |        |
| 患者内    |        |    |       |        |        |
| 时间     | 34 288 | 4  | 8 572 | 279.90 | 0.0001 |
| 时间×形式  | 19 472 | 4  | 4 868 | 159.00 | 0.0001 |
| 误差     | 980    | 32 | 21.01 |        |        |

结论:有根据(基于时间×药剂形式交互效应的显著性)( $p < 0.0001$ )表明两种药剂形式随时间变化有不同的浓度模式。

因子 A 的  $F$  检验基于研究对象(患者)间的效应,因而不受因子 B 上重复测量的影响。然而,患者内效应的  $F$  比值是受到影响的,并且像在有重复测量的单因子试验中一样,我们必须关心在什么样的条件下  $F$  检验才是适当的。如果  $y_{ijk}$  的方差-协方差矩阵的复合对称性成立,那么我们可以应用这些  $F$  检验。有人建议(Greenhouse and Geisser(1959), Huynh and Feldt(1970)),当对  $F$  检验成立的条件有所偏离时,用调整后的  $F$  值来决定一个重复测量  $F$  检验的显著性。许多作者提出了各种各样的调整方法,这些方法的做法大同小异。对所考虑的  $F$  检验的分子和分母的自由度定义一个乘积调整因子 Epsilon,记之为  $e$ ,以区别于上述模型中的误差  $\epsilon$ 。对于大多数调整方法,调整因子  $e$  介于 0,1 之间,当使得  $F$  检验有效的条件得到满足时,取 1,否则,随着与有效条件偏离程度的增加  $e$  的值将越小。对于一个给定的情形,在决定了  $e$  的值以后,把计算所得的  $F$  统计量与将分子分母的自由度都乘以  $e$  以后作为两个自由度的  $F$  分布的临界值相比较。

我们以表 18.6 中的试验设计为例来说明调整的基本思想。此时,我们有一个两因子试验,关于第二个因子 B 有重复测量。当 Huynh-Feldt 条件成立时,对于患者内效应,即表 18.6 中的 B 和 AB 的  $F$  检验是有效的。

对于一个给定的试验,我们计算出  $e$  的值,把  $df_1$  和  $df_2$  都乘以  $e$  来调整  $F$  检验的自由度。所以,为了检验  $H_0: \theta_{AB} = 0$ ,利用样本数据计算出一个  $e$  的值,然后计算  $F$  统计量

$$F = \frac{MS_{AB}}{MSE}$$

并将其与基于自由度  $df_1 = e(a-1)(b-1)$  和  $df_2 = ea(b-1)(n-1)$  的临界值  $F_\alpha$  相比较。注意,当  $e = 1$  时,条件成立,我们得到原来建议使用的自由度  $df_1 = (a-1)(b-1)$  和  $df_2 = a(b-1)(n-1)$ 。

当分析有重复测量数据的试验时,如果有 SAS 软件可用,那么可以使用 PROC GLM 来计算两种不同的自由度调整方法下检验的  $p$  值。SAS 中使用的第一种调整方法是由 Greenhouse 和 Geisser(1959)提出的,这种方法用样本计算  $e$  的一个估计值。在 SAS 输出中,这种方法被标以“G-G”。模拟研究表明,这种调整过于保守,因为实际的  $p$  值比 G-G 调整给出的  $p$  值可能小很多。第二种调整因子(由 Huynh 和 Feldt(1970)提出)则基于  $e$  的一个不同的公式。然而,也需要从样本计算调整因子的一个估计。然后, $F$  统计量的临界值的自由度用这个  $e$  的估计进行调整。这种调整在 PROC GLM 的输出中被标以“H-F”。虽然 Greenhouse-Geisser 的  $e$  和 Huynh-Feldt 的  $e$  均必落入区间  $0 < e \leq 1$ ,但 H-F 方法中  $e$  的估计有时会大于 1。此时,用  $e = 1$  来确定  $F$  检验的自由度。

#### General Linear Models Procedure

Dependent Variable: CONC

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 17 | 57720.5000000  | 110.87  | 0.0001 |
| Error           | 32 | 980.0000000    |         |        |
| Corrected Total | 49 | 58700.5000000  |         |        |

| Source        | DF | Type I SS     | F Value | Pr > F |
|---------------|----|---------------|---------|--------|
| FORM          | 1  | 40.5000000    | 1.32    | 0.2587 |
| PATIENT(FORM) | 8  | 3920.0000000  | 16.00   | 0.0001 |
| TIME          | 4  | 34288.0000000 | 279.90  | 0.0001 |
| FORM * TIME   | 4  | 19472.0000000 | 158.96  | 0.0001 |

Source: FORM

Error: MS(PATIENT(FORM))

|    |             | Denominator | Denominator |         |        |
|----|-------------|-------------|-------------|---------|--------|
| DF | Type III MS | DF          | MS          | F Value | Pr > F |
| 1  | 40.5        | 8           | 490         | 0.0827  | 0.7810 |

#### Repeated Measures Analysis of Variance

Tests of Hypotheses for Between Subjects Effects

| Source | DF | Type III SS  | F Value | Pr > F |
|--------|----|--------------|---------|--------|
| FORM   | 1  | 40.5000000   | 0.08    | 0.7810 |
| Error  | 8  | 3920.0000000 |         |        |

## Repeated Measures Analysis of Variance

## Univariate Tests of Hypotheses for Within Subject Effects

Source: TIME

| DF | Type III SS | Mean Square | F Value | Pr > F | Adj<br>G-G | Pr > F<br>H-F |
|----|-------------|-------------|---------|--------|------------|---------------|
| 4  | 34288.00000 | 8572.00000  | 279.90  | 0.0001 | 0.0001     | 0.0001        |

Source: TIME\* FORM

| DF | Type III SS | Mean Square | F Value | Pr > F | Adj<br>G-G | Pr > F<br>H-F |
|----|-------------|-------------|---------|--------|------------|---------------|
| 4  | 19472.00000 | 4868.00000  | 158.96  | 0.0001 | 0.0001     | 0.0001        |

Source: Error(TIME)

| DF | Type III SS | Mean Square |
|----|-------------|-------------|
| 32 | 980.00000   | 30.62500    |

Greenhouse-Geisser Epsilon = 0.7374

Huynh-Feldt Epsilon = 1.3610

**例 18.2**

参考例 18.1 中的计算机输出。

- 找出 Greenhouse-Geisser 的调整因子  $e$  和 Huynh-Feldt 的调整因子  $e$  的估计值;
- 关于时间效应和时间药剂形式交互效应的检验的结论受这些调整的影响吗?

**解答**

- $e$  的 Greenhouse-Geisser 估计值为 0.7374, Huynh-Feldt 估计值为 1.3610。
- 时间效应: 基于 G-G 调整和 H-F 调整的  $F$  检验的  $p$  值分别为 0.0001 和 0.0001, 它们与原来的  $F$  检验一样, 因此, 调整没有改变原有的结论。

时间  $\times$  药剂形式交互效应: 与时间效应的检验一样, 调整没有改变检验的  $p$  值, 从而也没有改变检验的结论。

我们指出, 如果在进行有重复观测试验的方差分析时, 使用了 SAS 系统, 最好检查一下对于有关有重复测量的因子的效应的  $F$  检验, 调整的效果如何。如果原来的  $F$  检验(即  $e=1$ )的结果与基于 H-F 调整或 G-G 调整的结果不同, 那么, 我们建议采用基于不太保守的 H-F 调整的结论。

下面,我们来完成案例中数据的分析。

### 案例数据的分析:石油泄漏对植物生长的影响

此项研究的目的在于考察石油溢漏和接下来在溢油地块上的燃烧对植物 *Distichlis spicata* 生长密度的影响。由于没有溢油地域溢油和燃烧前植物密度的基础数据,只能与没有受到溢油影响的地域进行比较。在图 18.2 中,给出了八个观测时点处对照和燃烧地块的植物密度的侧面图。对照地块(C)上的平均密度一致地比燃烧地块(B)上的平均密度高。除了两个观测时点(D3 和 D7)以外,对照地块和燃烧地块上的平均密度的变化也类似。在 D3 和 D7 这两个时点上,燃烧后的地块上平均密度比前一个时点下降了,而对照地块上的平均密度却是上升的。我们接下来构造有重复测量的方差分析,来确认上述的观察结果。

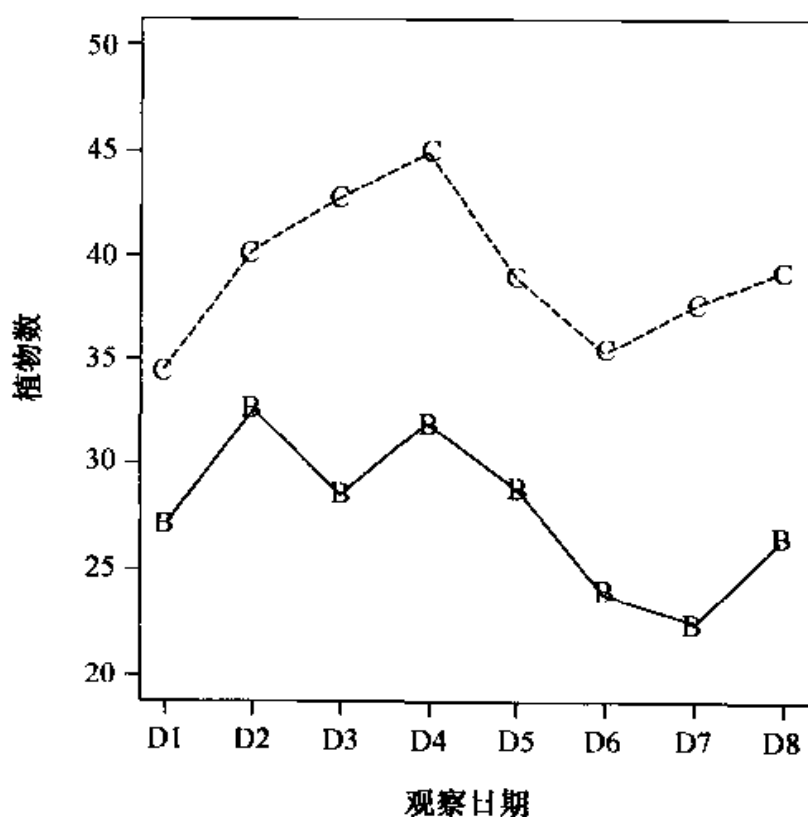


图 18.2 植物密度的侧面图

植物密度数据的方差分析表给出如下。

| 来源             | 调整后的 $p$ 值 |     |           |       |        |        |        |
|----------------|------------|-----|-----------|-------|--------|--------|--------|
|                | SS         | df  | MS        | F     | $p$ 值  | G-G    | H-F    |
| 地块间            |            |     |           |       |        |        |        |
| 处理             | 10 511.11  | 1   | 10 511.11 | 6.56  | 0.0145 |        |        |
| 处理中的地块         | 60 844.63  | 38  | 1 601.17  |       |        |        |        |
| 地块内            |            |     |           |       |        |        |        |
| 日期             | 2 845.09   | 7   | 406.44    | 19.35 | 0.0001 | 0.0001 | 0.0001 |
| 日期 $\times$ 处理 | 602.29     | 7   | 86.04     | 4.10  | 0.0003 | 0.0046 | 0.0029 |
| 误差             | 5,587.88   | 266 | 21.01     |       |        |        |        |

Greenhouse-Geisser 因子 = 0.5269

Huynh-Feldt 因子 = 0.5355

日期与处理的交互效应是显著的,这确证了我们从侧面图中看到的结果。更进一步,在燃烧后的地块与对照地块之间平均密度有显著的差异。对照地块比燃烧后的地块有更高的平均植物密度,见表 18.3。在第一个观测日期,二者之间的差为 7.25,到最后,即到两年多一点以后,差异扩大到 12.70。因此,溢油地块上的平均植物密度表明,在植物密度上没有恢复,1992 年 10 月为 27.20,至 1995 年 1 月为 26.65。由于对照地块在研究期间与溢油地块有相似的土壤和环境条件,而对照地块上的植物密度从 34.45 变到 39.35,有所增加,我们可以断言,溢油和接下来的燃烧导致这些地域上植物密度的下降。

**报告结论** 总结这项研究的结论时我们需要包括以下内容。

1. 阐述研究的目标。
2. 描述研究设计,地块是如何挑选的,植物密度是如何测量出来的,测量日期是如何选择出来的,以及所选择的地块在环境和土壤条件方面相似性的论证。
3. 讨论该研究的结论能否推广至其他溢油地域。
4. 数据的数值和图表示。
5. 所有推断方法的描述:
  - 明确陈述研究假设,
  - 表示试验条件的模型,
  - 模型条件的验证,
  - 包含  $p$  值的 AOV 表。
6. 讨论研究的结果和结论。
7. 相对于以前的研究,解释此项研究中的发现。
8. 相对于石油溢漏的类似处理所做的推断。

## 9. 数据表。

## 练习

## 应用

**18.1 (环境)** 卡晏花蜱(cayenne tick)被认为是对野生动植物、牲畜和人类有害的害虫。它分布于西半球北纬  $30^{\circ}$  到南纬  $30^{\circ}$  的区域内。已经知道这种壁虱是若干种疾病的潜在携带者,但对它们的生态知之甚少。为此,进行了下面的研究,以考察这种壁虱的生存潜力及其与环境饱和差(SD, saturation deficit)的关系。饱和差是一项环境指标,它与温度和相对湿度有关,当温度增加时 SD 增加,但当相对湿度增加时,SD 下降。因此,SD 的高值联系于高温和低相对湿度,这会引引起壁虱经历最强的脱水。研究中挑选使用了 SD 的 5 个值(2.98, 4.83, 5.80, 8.88, 13.38mm/Hg)。人工建立起与这些值相应的环境,5 只壁虱被随机地分配到 5 种环境之一。在大约 3 周的研究期间内,每隔两天记录一次壁虱整个身体的水分损失。这些壁虱的水分损失(mg)给出如下。

| SD   | 壁虱 | 暴露的天数 |      |      |      |      |      |      |      |      |      |      |
|------|----|-------|------|------|------|------|------|------|------|------|------|------|
|      |    | 1     | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   |
| 2.98 | 1  | 0.54  | 0.59 | 0.64 | 0.73 | 0.76 | 0.89 | 0.93 | 1.01 | 1.08 | 1.15 | 1.23 |
|      | 2  | 0.69  | 0.75 | 0.81 | 0.90 | 0.97 | 1.20 | 1.14 | 1.19 | 1.26 | 1.38 | 1.43 |
|      | 3  | 0.77  | 0.80 | 0.87 | 0.94 | 1.01 | 1.10 | 1.17 | 1.24 | 1.34 | 1.41 | 1.51 |
|      | 4  | 0.64  | 0.69 | 0.77 | 0.83 | 0.88 | 0.96 | 1.04 | 1.09 | 1.20 | 1.23 | 1.31 |
|      | 5  | 0.51  | 0.58 | 0.62 | 0.71 | 0.74 | 0.81 | 0.88 | 0.93 | 0.99 | 1.03 | 1.13 |
| 4.83 | 1  | 0.64  | 0.71 | 0.77 | 0.89 | 0.90 | 1.00 | 1.06 | 1.14 | 1.22 | 1.34 | 1.39 |
|      | 2  | 0.80  | 0.91 | 0.97 | 1.01 | 1.11 | 1.19 | 1.29 | 1.31 | 1.37 | 1.47 | 1.54 |
|      | 3  | 0.79  | 0.85 | 0.89 | 0.99 | 1.04 | 1.05 | 1.16 | 1.21 | 1.32 | 1.39 | 1.47 |
|      | 4  | 0.77  | 0.82 | 0.88 | 0.92 | 1.01 | 1.09 | 1.19 | 1.27 | 1.35 | 1.44 | 1.58 |
|      | 5  | 0.79  | 0.84 | 0.91 | 0.98 | 1.07 | 1.14 | 1.19 | 1.31 | 1.37 | 1.46 | 1.55 |
| 5.80 | 1  | 0.72  | 0.79 | 0.83 | 0.94 | 0.98 | 1.09 | 1.12 | 1.21 | 1.28 | 1.34 | 1.41 |
|      | 2  | 0.89  | 0.94 | 1.01 | 1.21 | 1.27 | 1.40 | 1.44 | 1.49 | 1.49 | 1.58 | 1.63 |
|      | 3  | 0.97  | 0.99 | 1.07 | 1.09 | 1.21 | 1.30 | 1.37 | 1.44 | 1.54 | 1.61 | 1.73 |
|      | 4  | 0.85  | 0.88 | 0.97 | 1.05 | 1.09 | 1.17 | 1.24 | 1.29 | 1.30 | 1.23 | 1.51 |
|      | 5  | 0.71  | 0.78 | 0.82 | 0.91 | 0.94 | 1.11 | 1.19 | 1.23 | 1.29 | 1.33 | 1.43 |

续表

| SD    | 壁虱 | 暴露的天数 |      |      |      |      |      |      |      |      |      |      |
|-------|----|-------|------|------|------|------|------|------|------|------|------|------|
|       |    | 1     | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   |
| 8.88  | 1  | 0.93  | 0.99 | 1.03 | 1.14 | 1.18 | 1.29 | 1.33 | 1.36 | 1.38 | 1.54 | 1.62 |
|       | 2  | 1.09  | 1.14 | 1.21 | 1.41 | 1.47 | 1.55 | 1.64 | 1.69 | 1.71 | 1.78 | 1.83 |
|       | 3  | 1.19  | 1.20 | 1.07 | 1.29 | 1.31 | 1.50 | 1.57 | 1.64 | 1.74 | 1.81 | 1.93 |
|       | 4  | 1.05  | 1.08 | 1.17 | 1.25 | 1.29 | 1.37 | 1.44 | 1.49 | 1.50 | 1.53 | 1.71 |
|       | 5  | 1.01  | 1.09 | 1.18 | 1.21 | 1.29 | 1.31 | 1.39 | 1.43 | 1.49 | 1.53 | 1.63 |
| 13.38 | 1  | 1.05  | 1.09 | 1.13 | 1.24 | 1.28 | 1.39 | 1.43 | 1.56 | 1.68 | 1.74 | 1.82 |
|       | 2  | 1.29  | 1.34 | 1.41 | 1.51 | 1.57 | 1.65 | 1.74 | 1.79 | 1.83 | 1.88 | 1.93 |
|       | 3  | 1.38  | 1.40 | 1.47 | 1.49 | 1.51 | 1.60 | 1.69 | 1.74 | 1.79 | 1.87 | 2.03 |
|       | 4  | 1.23  | 1.28 | 1.37 | 1.45 | 1.49 | 1.57 | 1.64 | 1.69 | 1.70 | 1.73 | 1.81 |
|       | 5  | 1.23  | 1.29 | 1.38 | 1.41 | 1.49 | 1.52 | 1.48 | 1.53 | 1.59 | 1.63 | 1.78 |

- 画出这些数据的侧面图, 显示对于每个 SD 的值, 平均体重的损失随时间的变化;
- 对于卡晏花蜱, 是否随着饱和差的增加其损失的体重也在增加?

### 18.2 参考练习 18.1。

- 对这一设计, 给出一个模型;
- 对于此项研究, 构造一个方差分析表;
- 对于卡晏花蜱, 饱和差的增加是否引起其损失的体重也在增加? 取  $\alpha = 0.05$ ;
- 对于每一 SD 的水平, 卡晏花蜱的损失体重在研究期间是否增加了?

**18.3 (医学)** 为了研究抗组胺剂在对由已知的变态反应源引起的阳性皮肤反应的抑制效果(与一个安慰剂相比较), 常常使用一个模型。考虑下列情形。为了寻找用于研究中的 20 个对象, 对许多个人进行了筛选, 找出其中对于变态反应源过敏的人。然后, 把这 20 个人随机地分配到两个处理组(已知的抗组胺剂和一种外表完全一样的安慰剂), 每组 10 人。研究开始(药前)时, 测量一下过敏读数作为基准。然后, 每个患者服用指定的药剂三天。在首次服药后 1 小时, 2 小时, 3 小时, 4 小时和 8 小时后分别测量其过敏读数。20 名患者的皮肤过敏反应抑制的百分比(比起基准读数来, 使用变态反应源引起的肿胀面积的减少)数据给出如下。



| 处理 | 患者 | 抑制时间(小时) |       |      |      |      |
|----|----|----------|-------|------|------|------|
|    |    | 1        | 2     | 3    | 4    | 8    |
| 1  | 1  | 10.5     | 28.2  | 15.3 | 43.0 | 29.0 |
|    | 2  | 41.2     | 25.3  | 27.8 | 28.0 | 53.2 |
|    | 3  | 43.0     | 20.8  | 29.3 | 5.2  | 26.5 |
|    | 4  | 61.4     | 61.6  | 62.8 | 43.8 | 19.6 |
|    | 5  | 5.0      | 28.2  | 31.6 | 19.5 | 2.3  |
|    | 6  | -10.2    | 27.2  | 38.1 | 35.5 | 18.0 |
|    | 7  | -12.9    | 22.1  | 34.0 | 43.4 | 34.2 |
|    | 8  | 27.1     | 26.5  | 38.8 | 28.5 | 17.4 |
|    | 9  | 13.0     | 19.7  | 23.5 | 29.4 | 39.6 |
|    | 10 | 28.9     | 26.1  | 11.2 | 18.1 | 16.5 |
| 2  | 1  | 3.0      | 9.3   | 1.0  | 15.0 | 3.0  |
|    | 2  | -1.5     | -10.1 | 20.2 | 18.3 | 13.5 |
|    | 3  | 10.8     | 20.6  | 28.3 | 25.2 | 15.8 |
|    | 4  | 15.3     | 19.8  | 25.4 | 31.3 | 21.7 |
|    | 5  | 8.7      | 8.0   | 17.5 | 26.6 | 16.4 |
|    | 6  | -4.6     | 5.8   | 12.7 | 15.6 | 29.6 |
|    | 7  | -16.6    | 28.4  | 32.7 | 34.4 | 15.8 |
|    | 8  | 9.4      | 15.7  | 22.7 | 29.8 | 23.2 |
|    | 9  | -19.3    | 15.7  | 21.7 | 30.4 | 26.1 |
|    | 10 | -12.8    | 12.3  | 0.1  | 21.3 | 10.6 |

(负值表示与基准数据比较起来, 肿胀的面积有所扩大。)

a. 对于每一组, 比较不同时间的均值和标准差;

b. 对于每一组, 以时间为横轴, 平均抑制百分比为纵轴画出数据的图。抗组胺剂组与安慰剂组看上去是不同的吗?

**18.4** 参考练习 18.3 中的数据。对该设计给出一个模型, 并作有重复测量的方差分析来比较两个组。方差分析的结果与在练习 18.3 中从图中得出的判断一致吗?

**18.5** 关于抗组胺剂的作用, 我们还可以提出另一个问题: 如何定义抗组胺剂作用的起始? 对每个处理组, 用  $t$  检验决定一个时间, 在此时间处, 过敏读数比基准读数有显著的减少。这会导致什么样的建议?

## 18.4 交叉设计

我们现在考虑 18.2 节中所讨论的单因子试验的一种扩充。回忆表 18.4, 我们曾经给出这样一个试验的数据: 在这个试验中,  $n$  个患者中的每一个按照一种随机的顺序接受了同样的三种药物的治疗。对于药物的一个拉丁方安排形成一个试验设计, 当患者或条件不随时间而变时, 它与有重复测量的单因子试验(即对每位患者做多次观测并做处理的患者内比较)有相同的优点。

对于这种试验的情形, 一个  $3 \times 3$  拉丁方设计由表 18.7 给出。该设计称为一个三阶段(three-period)交叉设计。

表 18.7 一个  $3 \times 3$  拉丁方设计

| 序列 | 患者  | 因子 B(阶段) |       |       |
|----|-----|----------|-------|-------|
|    |     | 1        | 2     | 3     |
| 1  | $n$ | $A_1$    | $A_2$ | $A_3$ |
| 2  | $n$ | $A_2$    | $A_3$ | $A_1$ |
| 3  | $n$ | $A_3$    | $A_1$ | $A_2$ |

在这个设计中,  $3n$  个患者被随机地指定到设计的序列(行), 每个序列  $n$  个。阶段相应于服用药物的顺序。该设计的模型为

$$y_{ijkl} = \mu + \delta_k + \pi_{l(k)} + \alpha_i + \beta_j + \alpha\beta_{ij}^* + \epsilon_{ijkl}$$

其中  $\delta_k$  为第  $k$  个序列的固定效应,  $\pi_{l(k)}$  为序列  $k$  中第  $l$  个患者的随机效应,  $\alpha_i$  和  $\beta_j$  分别为药物(因子 A)和阶段(因子 B)的固定效应。在交互作用项上加星号的原因以后再做解释。这个设计(三阶段交叉设计)的方差分析由表 18.8 给出。

表 18.8 三阶段交叉设计的方差分析

| 来源     | SS       | df          | EMS(A, B 固定, 患者随机)                            |
|--------|----------|-------------|---|
| 患者间    |          |             |   |
| 序列     | SSSeq    | 2           | $\sigma_e^2 + 3\sigma_\pi^2 + 3n\theta_{Seq}$ |
| 序列中的患者 | SSP(Seq) | $3(n-1)$    | $\sigma_e^2 + 3\sigma_x^2$                    |
| 患者内    |          |             |   |
| A      | SSA      | 2           | $\sigma_e^2 + 3n\theta_A$                     |
| B      | SSB      | 2           | $\sigma_e^2 + 3n\theta_B$                     |
| AB*    | SSAB*    | 2           | $\sigma_e^2 + n\theta_{AB}$                   |
| 误差     | SSE      | $3(2)(n-1)$ | $\sigma_e^2$                                  |
| 总和     | TSS      | $9n-1$      |   |

方差分析中患者间部分的效应的平方和、自由度和期望均方,以及患者内部分的因子  $A$  和  $B$  的效应的平方和、自由度和期望均方都是直截了当的。但注意,在  $AB$  的交互效应项上加了一个 \* 号,并且这个项的自由度只有 2,而不是 4。实际上,损失的这两个自由度在序列的平方和上。事实上,容易证明,  $AB$  交互效应平方和等于

$$SSAB = SS_{Seq} + SSAB^*$$

以后我们就用该式来计算  $SSAB^*$ 。

### 例 18.3

12 位男性志愿参加一项研究,以比较一种药品的三种不同的药剂形式的效用持续时间。形式 1 是 50mg 的片剂,形式 2 是 100mg 的片剂,形式 3 为缓释胶囊。使用三阶段交叉设计,对于三个处理阶段中的每一个,指定 4 位志愿者。在每一个处理日,给以志愿者指定形式的药剂,并观察其药效(血压下降)的持续时间。试验中每一个处理阶段有一周的清洗时间。样本数据如下。

| 序列 | 患者 | 阶 段 |     |     |
|----|----|-----|-----|-----|
|    |    | 1   | 2   | 3   |
| 1  | 1  | 1.5 | 2.2 | 3.4 |
|    | 2  | 2.0 | 2.6 | 3.1 |
|    | 3  | 1.6 | 2.7 | 3.2 |
|    | 4  | 1.1 | 2.3 | 2.9 |
| 2  | 1  | 2.5 | 3.5 | 1.9 |
|    | 2  | 2.8 | 3.1 | 1.5 |
|    | 3  | 2.7 | 2.9 | 2.4 |
|    | 4  | 2.4 | 2.6 | 2.3 |
| 3  | 1  | 3.3 | 1.9 | 2.7 |
|    | 2  | 3.1 | 1.6 | 2.5 |
|    | 3  | 3.6 | 2.3 | 2.2 |
|    | 4  | 3.0 | 2.5 | 2.0 |

| 序列 | 患者    | 阶 段   |       |       |
|----|-------|-------|-------|-------|
|    |       | 1     | 2     | 3     |
| 1  | $n=4$ | $A_1$ | $A_2$ | $A_3$ |
| 2  | $n=4$ | $A_2$ | $A_3$ | $A_1$ |
| 3  | $n=4$ | $A_3$ | $A_1$ | $A_2$ |

**解答** 下面给出了方差分析的计算机输出结果。基于这些结果,有阶段  $\times$  处理交互效应的迹象( $p=0.0853$ ),而由于处理效应非常显著( $p=0.0001$ ),这一交互效应看来可以忽略不计。这一点可以从下面给出的图中得到证实,该图中对每个序列沿个阶段画出了其平均持续时间。由图可以看出,最长的平均持续时间是

## General Linear Models Procedure

Dependent Variable: DURATION

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 17 | 11.08638889    | 5.69    | 0.0003 |
| Error           | 18 | 2.06333333     |         |        |
| Corrected Total | 35 | 13.14972222    |         |        |

| Source         | DF | Type I SS  | F Value | Pr > F |
|----------------|----|------------|---------|--------|
| SEQ            | 2  | 0.23388889 | 1.02    | 0.3804 |
| PATIENT(SEQ)   | 9  | 0.66916667 | 0.65    | 0.7425 |
| TREAT          | 2  | 9.51722222 | 41.51   | 0.0001 |
| PERIOD         | 2  | 0.01722222 | 0.08    | 0.9279 |
| TREAT * PERIOD | 2  | 0.64888889 | 2.83    | 0.0853 |

## Tests of Hypotheses for Mixed Model Analysis of Variance

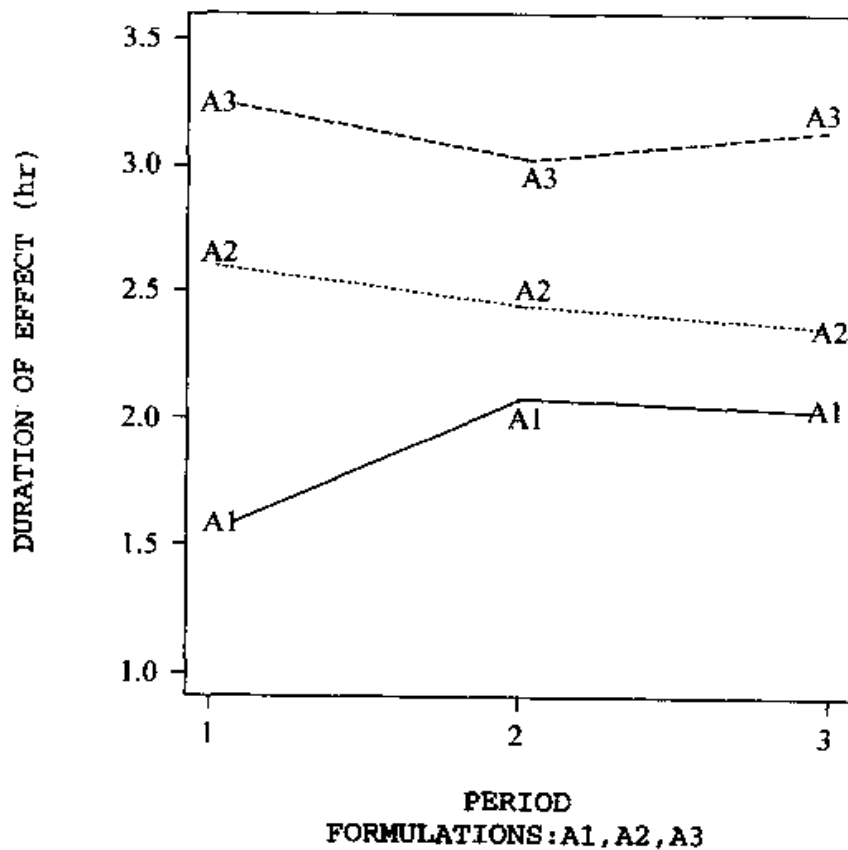
Dependent Variable: DURATION

Source: SEQ

Error: MS(PATIENT(SEQ))

| DF | Type I MS    | Denominator<br>DF | Denominator<br>MS | F Value | Pr > F |
|----|--------------|-------------------|-------------------|---------|--------|
| 2  | 0.1169444444 | 9                 | 0.0743518519      | 1.5729  | 0.2595 |

Plot of MEAN DURATION versus PERIOD OF STUDY



当只有两种药物需要考察的时候,相应的拉丁方设计称为两阶段交叉设计,其中将有  $2n$  个患者,这些患者被随机地指定到两个序列,每个序列中  $n$  个患者。这个两阶段交叉设计如表 18.9 所示。

表 18.9 两阶段交叉设计中的安排

| 序列 | 患者  | 因子 B(阶段) |       |
|----|-----|----------|-------|
|    |     | 1        | 2     |
| 1  | $n$ | $A_1$    | $A_2$ |
| 2  | $n$ | $A_2$    | $A_1$ |

相应的方差分析模型为

$$y_{ijkl} = \mu + \delta_k + \pi_{l(k)} + \alpha_i + \beta_j + \epsilon_{ijkl}$$

其中  $\delta_k$  为第  $k$  个序列的固定效应,  $\alpha_i$  和  $\beta_j$  分别为处理  $i$  和阶段  $j$  的固定效应,  $\pi_{l(k)}$  仍如前,为序列  $k$  中第  $l$  个人的随机效应。

注意,在这个模型中,没有  $AB$  交互效应。我们必须假定这个交互效应是可以忽略的,否则,这个设计将是不适当的,因为没有检验  $AB$  交互效应的显著性所需要的自由度。两阶段交叉设计的方差分析表由表 18.10 给出。

表 18.10 两阶段交叉设计的 AOV 表

| 来源     | SS       | df       | EMS(A, B 固定, 患者随机)                            |
|--------|----------|----------|---|
| 患者间    |          |          |   |
| 序列     | SSSeq    | 1        | $\sigma_e^2 + 2\sigma_\pi^2 + 2n\theta_{Seq}$ |
| 序列中的患者 | SSP(Seq) | $2(n-1)$ | $\sigma_e^2 + 2\sigma_\pi^2$                  |
| 患者内    |          |          |   |
| A      | SSA      | 1        | $\sigma_e^2 + 2n\theta_A$                     |
| B      | SSB      | 1        | $\sigma_e^2 + 2n\theta_B$                     |
| 误差     | SSE      | $2(n-1)$ | $\sigma_e^2$                                  |
| 总和     | TSS      | $4n-1$   |   |

对于本章讨论的有重复测量的设计,还有许多其他的扩充。例如,可以把关于在表 18.4 中说明的那个因子的重复测量的概念与交叉设计结合起来。表 18.11 给出了这种设计的一个例子,其中不是在每个阶段内对每个患者只观测一次,而是在  $t$  个不同的时间点进行观测。例如,我们可以在用第  $i$  种药剂治疗后的第一个小时内每隔 15 分钟测量一次血压,然后,在接下来的 7 个小时内每隔一小时测量一次。在每个阶段内都这样做,因而,在每个阶段对每个患者将有总共 10 个血压

测量值。

表 18.11 有重复测量的两阶段交叉设计

| 序列 | 阶 段   |   |     |     |       |   |     |     |
|----|-------|---|-----|-----|-------|---|-----|-----|
|    | 1     |   |     |     | 2     |   |     |     |
|    | 时间    |   |     |     | 时间    |   |     |     |
|    | 1     | 2 | ... | $t$ | 1     | 2 | ... | $t$ |
| 1  | $A_1$ |   |     |     | $A_2$ |   |     |     |
| 2  | $A_2$ |   |     |     | $A_1$ |   |     |     |

虽然我们不打算给出本章所讨论的有重复观测试验的这样一种扩充的方差分析,也不再讨论更为复杂的有重复观测的设计,但如果读者要对每个试验单元做多次观测,那么我们希望读者能够对所有可能的设计有所了解。感兴趣的读者请参考 Vonesh and Chinchilli(1997), Crowder and Hand(1990), Jones and Kenward(1994), Diggle, Liang and Zeger(1996)。

## 18.5 小 结

在本章中,我们讨论了一些与有重复测量的试验有关的基本概念和设计。我们介绍了单因子试验,两因子试验,这些试验的分析,以及两阶段、三阶段交叉设计的特殊情形。然而,这些方法仅仅是一个开始。对于这些内容,我们没有给出详尽的讨论,而是通过简单介绍这些内容来看一看有重复观察的设计和方法的适用性和益处。至于这类试验的设计和分析,可以从其他含有有关重复测量内容的详细介绍的材料或教科书中得到。

### 补充练习

**18.6(心理学)** 对某种尚在研究阶段的药品在睡眠实验室中进行试验,以确定这种药品对于睡眠持续时间的效应。有 16 名患者志愿参加此项研究。把他们随机分配到两个药物序列,其中 8 人在第一阶段接受这种药品的治疗,并在第二阶段接受外表一模一样的安慰剂治疗,另外 8 人则以相反的次序接受治疗。

- 指出该设计是什么设计。
- 给出这个设计的一个模型。
- 叙述可能影响该设计的适用性的假设。

**18.7** 练习 18.6 中患者的睡眠持续数据(小时/夜)如下。

| 序列 | 患者 | 阶 段 |     |
|----|----|-----|-----|
|    |    | 1   | 2   |
| 1  | 1  | 8.6 | 8.0 |
|    | 2  | 7.5 | 7.1 |
|    | 3  | 8.3 | 7.4 |
|    | 4  | 8.4 | 7.3 |
|    | 5  | 6.4 | 6.4 |
|    | 6  | 6.9 | 6.8 |
|    | 7  | 6.5 | 6.1 |
|    | 8  | 6.0 | 5.7 |
| 2  | 9  | 7.3 | 7.9 |
|    | 10 | 7.5 | 7.6 |
|    | 11 | 6.4 | 6.3 |
|    | 12 | 6.8 | 7.5 |
|    | 13 | 7.1 | 7.7 |
|    | 14 | 8.2 | 8.6 |
|    | 15 | 7.2 | 7.8 |
|    | 16 | 6.7 | 6.9 |

序列 1 先接受试验药品, 然后接受安慰剂; 序列 2 中治疗的次序相反。

a. 计算每个序列每个阶段的均值和标准差。

b. 画出这些数据的图, 显示在这项研究中发生了什么。试验药品会影响睡眠的持续时间吗? 是怎样影响的?

c. 对这个设计做有重复测量的方差分析, 并从中得出你的结论, 取  $\alpha = 0.05$ 。方差分析证实了你在(b)中的印象吗?

**18.8** 参见练习 18.6。假定我们忽略患者接受治疗的次序。数一下服用试验药品时比服用安慰剂时有更长的睡眠持续时间的人的个数。

a. 给出一个简单的检验方法, 以检验试验药品的效用。

b. 对于(a)中的检验, 给出其  $p$  值。

**18.9** 参见练习 18.6。假设序列 1 第二阶段中的睡眠持续时间为:

8.5   7.6   8.5   8.3   7.2   7.0   6.4   6.1

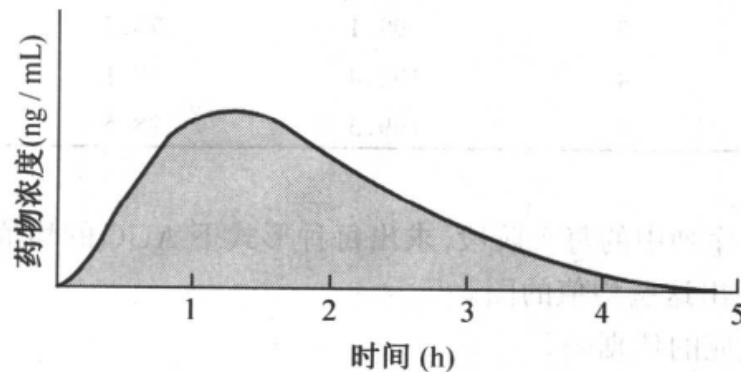
a. 画出这两个序列的数据图。

b. 这个设计是否看起来仍然合适? 对于所发生的现象有没有一种可能的解释?

**18.10** 参见练习 18.9。尽管有了从第二阶段得到的结果, 如果我们只利用

第一阶段的结果,还是可以进行处理组的患者间的比较。给出一个适当的检验方法,并实施这个检验,给出该检验的  $p$  值,并由此得到你的结论。

**18.11(医药)** 我们中的许多人经常看到与同一种药品的一般形式和品牌形式的生物有效度(bioavailability)有关的广告。比较一种药品的两种形式的生物有效度的方法之一是比较用这两种形式治疗的患者的浓度曲线下的面积(AUC, areas under the concentration curve)。例如,下图中阴影部分的面积表示用一种药品的某个剂量治疗的患者的 AUC。



为了比较两种品牌减肥药( $A_1, A_2$ )和一种一般减肥药( $A_3$ )的生物有效度,使用了三阶段交叉设计。研究中药品治疗安排的三个序列如下:

序列 1:  $A_1, A_2, A_3$

序列 2:  $A_2, A_3, A_1$

序列 3:  $A_3, A_1, A_2$

对于三个序列中的每一个,指定一个由 5 个患者构成的随机样本。这 15 个患者的 AUC 值如下。

| 序列 | 患者 | 阶 段   |       |      |
|----|----|-------|-------|------|
|    |    | 1     | 2     | 3    |
| 1  | 1  | 80.2  | 40.4  | 38.4 |
|    | 2  | 79.1  | 38.5  | 36.1 |
|    | 3  | 108.4 | 78.3  | 56.5 |
|    | 4  | 41.2  | 38.2  | 26.2 |
|    | 5  | 72.7  | 58.5  | 36.3 |
| 2  | 1  | 74.6  | 51.2  | 48.6 |
|    | 2  | 125.3 | 100.5 | 86.4 |
|    | 3  | 145.5 | 108.5 | 96.4 |



续表

| 序列 | 患者 | 阶 段   |      |      |
|----|----|-------|------|------|
|    |    | 1     | 2    | 3    |
| 2  | 4  | 86.7  | 68.8 | 58.2 |
|    | 5  | 107.8 | 78.5 | 53.1 |
| 3  | 1  | 79.7  | 40.4 | 37.2 |
|    | 2  | 89.2  | 68.8 | 56.2 |
|    | 3  | 99.1  | 76.5 | 43.9 |
|    | 4  | 102.4 | 88.1 | 53.4 |
|    | 5  | 109.3 | 98.5 | 76.8 |

a. 对于每个序列中的每个阶段, 求出每种形式下 AUC 的均值, 并以阶段为横轴, 均值为纵轴画出这些均值的图。

b. 有阶段效应的依据吗?

c. 各种形式的 AUC 有区别吗?

18.12 参见练习 18.11。进行三阶段交叉设计的方差分析, 取  $\alpha=0.05$ 。所得结果证实了你在练习 18.11 中给出的直观推测吗?

18.13 参见练习 18.11。只用第一阶段中的数据比较三种形式下的平均 AUC 值。这种分析是否确认了练习 18.12 中的结论? 练习 18.12 中的分析是否比本练习中的“平行”分析更合适? 为什么?

18.14(医药) 为了证实某种正在研究中的药品对于用标准方法治疗无效的癫痫症患者的发病次数的抑制效果, 进行一项研究。有 30 名患者参加了研究, 其中 15 人被随机地指定给这种药品的治疗组, 另外 15 人指定给安慰剂组。患者的一般情况的数据如下。

|        |                | 分 组                  |                     |
|--------|----------------|----------------------|---------------------|
|        |                | 试验药品<br>( $n_1=15$ ) | 安慰剂<br>( $n_2=15$ ) |
| 年龄(岁)  | 均值( $\pm$ 标准差) | 37.2( $\pm 10.5$ )   | 39.5( $\pm 9.6$ )   |
|        | 极差             | 19~68                | 21~65               |
| 性别     | 男              | 20                   | 16                  |
|        | 女              | 10                   | 14                  |
| 患病持续时间 | 均值( $\pm$ 标准差) | 10.7( $\pm 6.5$ )    | 11.5                |
|        | 极差             | 1~18                 | 1~26                |

- a. 这两个组关于表中这些变量是可比较的吗?  
 b. 平均年龄或平均患病持续时间有差别吗? 你如何做这些比较?  
 c. 如何比较这两组中性别的分布?

**18.15** 下面给出练习 18.14 的研究中癫痫发作的数据。注意, 我们有发作的基础记录, 同时也有治疗期间 5 个月的发作记录。

- a. 画出两个组的月份 - 平均发作次数的图。试验药品看起来有效吗?  
 b. 基于  $\alpha = 0.01$  做有重复测量的方差分析, 并给出你的结论。

| 组   | 患者 | 基础记录 | 时间(月) |    |    |    |    |
|-----|----|------|-------|----|----|----|----|
|     |    |      | 1     | 2  | 3  | 4  | 5  |
| 药剂  | 1  | 15   | 11    | 10 | 6  | 5  | 3  |
|     | 2  | 13   | 6     | 5  | 1  | 2  | 1  |
|     | 3  | 12   | 8     | 3  | 0  | 3  | 0  |
|     | 4  | 18   | 4     | 2  | 3  | 1  | 2  |
|     | 5  | 30   | 15    | 14 | 10 | 8  | 20 |
|     | 6  | 14   | 7     | 9  | 3  | 4  | 1  |
|     | 7  | 25   | 12    | 18 | 13 | 10 | 6  |
|     | 8  | 22   | 21    | 18 | 16 | 17 | 25 |
|     | 9  | 23   | 17    | 14 | 10 | 7  | 1  |
|     | 10 | 14   | 2     | 1  | 0  | 0  | 0  |
|     | 11 | 15   | 4     | 5  | 6  | 3  | 2  |
|     | 12 | 17   | 8     | 7  | 8  | 2  | 6  |
|     | 13 | 26   | 13    | 10 | 9  | 7  | 4  |
|     | 14 | 28   | 2     | 1  | 3  | 1  | 3  |
|     | 15 | 29   | 27    | 29 | 25 | 24 | 22 |
| 安慰剂 | 1  | 16   | 15    | 18 | 14 | 13 | 12 |
|     | 2  | 18   | 14    | 13 | 12 | 10 | 15 |
|     | 3  | 14   | 10    | 5  | 4  | 6  | 7  |
|     | 4  | 19   | 15    | 16 | 9  | 12 | 15 |
|     | 5  | 12   | 10    | 14 | 16 | 17 | 12 |
|     | 6  | 11   | 13    | 8  | 7  | 6  | 11 |
|     | 7  | 31   | 32    | 30 | 21 | 24 | 20 |
|     | 8  | 32   | 35    | 34 | 31 | 20 | 24 |
|     | 9  | 21   | 20    | 18 | 15 | 16 | 18 |
|     | 10 | 26   | 22    | 23 | 21 | 15 | 14 |
|     | 11 | 13   | 10    | 14 | 12 | 8  | 6  |
|     | 12 | 17   | 15    | 10 | 3  | 2  | 3  |
|     | 13 | 18   | 16    | 12 | 14 | 13 | 11 |
|     | 14 | 23   | 15    | 14 | 18 | 19 | 20 |
|     | 15 | 10   | 8     | 11 | 10 | 9  | 6  |

**18.16** 参见练习 18.15 中的数据。

a. 考虑从基础发作次数到第 5 个月的发作次数的变化。用这些数据比较这两个组,能得到相似的结论吗?

b. 由于发作次数有很强的变异性,有人会比较两组中各患者中的最大变化。这个方法支持你前面得到的结论吗?

**18.17 (环境)** 用 12 辆汽车组成的一个随机样本得到两种车型(MODEL)汽油燃烧效率的测量值,每种车型 6 辆车。对于每辆车,在 5 个不同的时间测量其燃烧效率。

a. 在每个时间点,计算每种车型的平均效率,并作出这些均值的图。

b. 取  $\alpha=0.05$  做方差分析,并给出你的结论。

c. 在下面给出的方差分析中,对于车型内的比较,调整因子有无作用?如果有,是什么样的作用?

| 车型 | 车 | 时间 1 | 时间 2 | 时间 3 | 时间 4 | 时间 5 |
|----|---|------|------|------|------|------|
| 1  | 1 | 1.43 | 1.47 | 1.39 | 1.40 | 1.44 |
| 1  | 2 | 1.50 | 1.41 | 1.51 | 1.53 | 1.41 |
| 1  | 3 | 1.79 | 1.88 | 1.89 | 2.00 | 1.90 |
| 1  | 4 | 1.87 | 1.78 | 2.00 | 2.00 | 2.11 |
| 1  | 5 | 1.85 | 1.89 | 1.93 | 1.86 | 1.81 |
| 1  | 6 | 1.89 | 1.66 | 1.78 | 1.77 | 1.67 |
| 2  | 1 | 1.63 | 1.62 | 1.64 | 1.63 | 1.53 |
| 2  | 2 | 1.81 | 1.83 | 1.84 | 1.83 | 1.86 |
| 2  | 3 | 2.25 | 2.10 | 2.34 | 2.27 | 2.32 |
| 2  | 4 | 1.79 | 1.80 | 1.92 | 2.03 | 2.02 |
| 2  | 5 | 2.11 | 2.00 | 2.33 | 2.46 | 2.35 |
| 2  | 6 | 2.10 | 2.03 | 2.00 | 2.09 | 1.87 |

General Linear Models Procedure

Repeated Measures Analysis of Variance

Tests of Hypotheses for Between Subjects Effects

| Source | DF | Type III SS | F Value | Pr > F |
|--------|----|-------------|---------|--------|
| MODEL  | 1  | 0.95760667  | 3.38    | 0.0960 |
| Error  | 10 | 2.83722667  |         |        |

General Linear Models Procedure

## Repeated Measures Analysis of Variance

## Tests of Hypotheses for within Subject Effects

Source: TIME

| DF | Type III SS | Mean Square | F Value | Pr > F | Adj<br>G-G | Pr > F<br>H-F |
|----|-------------|-------------|---------|--------|------------|---------------|
| 4  | 0.09579333  | 0.02394833  | 3.03    | 0.0285 | 0.0719     | 0.0512        |

Source: TIME \* MODEL

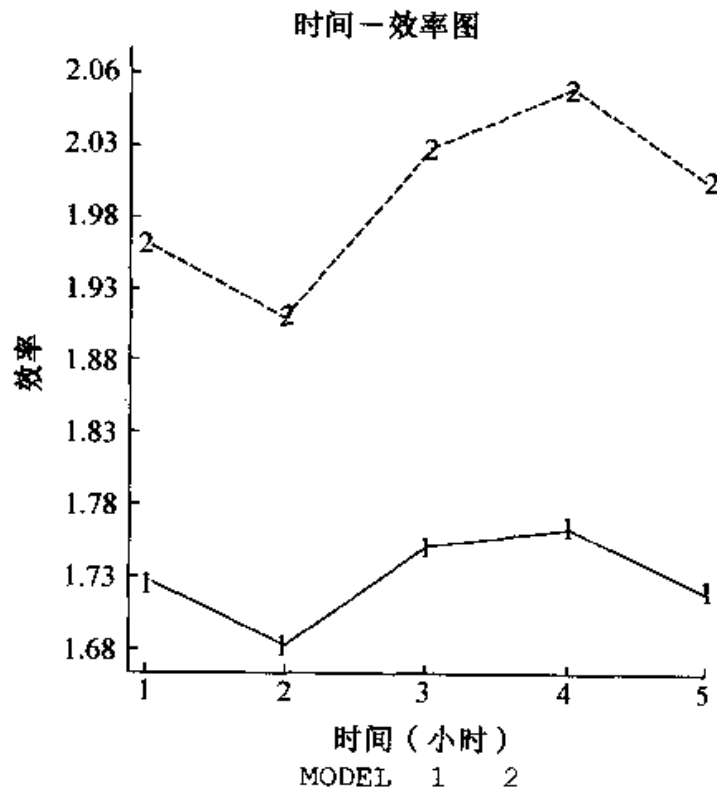
| DF | Type III SS | Mean Square | F Value | Pr > F | Adj<br>G-G | Pr > F<br>H-F |
|----|-------------|-------------|---------|--------|------------|---------------|
| 4  | 0.01182667  | 0.00295667  | 0.37    | 0.8260 | 0.6906     | 0.7528        |

Source: Error (TIME)

| DF | Type III SS | Mean Square |
|----|-------------|-------------|
| 40 | 0.31654000  | 0.00791350  |

Greenhouse-Geisser Epsilon = 0.4943

Huynh-Feldt Epsilon = 0.6770



## 第十九章 一些非平衡设计的方差分析

### 19.1 引言和案例

### 19.2 有一个或多个缺失观察值的随机化区组设计

### 19.3 有缺失数据的拉丁方设计

### 19.4 平衡不完全区组(BIB)设计

### 19.5 小结

## 19.1 引言和案例

在第八章和第十五章,我们考察了平衡设计的方差分析,其中我们利用适当的公式(以及相应的计算机程序)构造了 AOV 表,并建立了假设检验;我们还考虑了进行方差分析的另外一种方式。我们看到,方差分析表中与一种变异来源相应的平方和,可以通过计算拟合简略模型和完全模型的误差平方和的减少来得到。虽然没有提倡用拟合完全模型和简略模型的方法来计算平衡设计中各个变异来源上的平方和,但我们确实曾指出,这个方法是一般性的,对于任何试验设计都适用。特别地,在本章中,我们将应用完全模型和简略模型来得到非平衡设计的分析中的各个平方和。对于这些非平衡设计中的平方和,不再有简单易用的公式。

你可能会问,既然非平衡设计分析起来困难得多,那么试验人员为什么还要用非平衡设计进行研究呢?事实上,大多数研究都以平衡设计开始,但是,由于很多不同原因中的任何一种,试验人员不能像正在使用的平衡设计中所规定的那样,在每个试验单元上都得到相同个数的观察值。考虑比较三种不同的减肥药的一项研究。有五家不同的诊所(区组)参加了研究,根据某个随机化区组设计的方案,患者被随机地分配到三个治疗组。尽管试验人员计划给每个诊所中的每种治疗指定 5 个体重超标的人,但几乎可以肯定,最后记数的结果在分配给每个组的人数上会有一些不平衡。几乎所有的诊所都会预期有少数人完不成这项研究。某些人可能从做试验的社区中搬走了,另外一些人可能由于这次试验中减肥效果不明显而放弃了。还有,试验者也可能发现,不能在每个诊所中都找到 15 个愿意参加这项试验的体重超标的人。由于研究结尾时的非平衡设计是经常发生的,我们必须学会分析从非平衡设计得到的数据。

下面我们考虑一个案例,其中我们在实施试验以前,了解该设计的非平衡性,因此,我们可以设计这项研究,使得其能够部分地适应这种非平衡性,极小化处理效应估计中的偏差。

[illegible]

续表

| 房产 | 评 估 员 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | 1     | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  |
| 9  |       |     | 156 |     | 155 |     |     |     |     |     | 150 |     |     | 138 | 124 | 156 |
| 10 |       |     | 156 |     |     | 128 |     | 155 |     |     |     | 153 | 155 |     |     | 122 |
| 11 | 155   |     | 158 | 157 |     |     |     |     | 142 |     |     |     | 123 | 155 |     |     |
| 12 |       | 118 |     |     |     |     | 110 |     |     | 113 |     |     | 118 | 125 |     | 111 |
| 13 |       |     | 152 |     |     | 111 | 150 |     | 112 | 128 |     |     |     |     |     | 130 |
| 14 | 115   |     |     |     |     |     | 112 | 110 |     |     | 135 |     | 130 |     |     | 128 |
| 15 |       | 115 |     |     |     |     |     | 110 | 145 |     |     | 135 |     | 124 | 120 |     |
| 16 |       |     |     | 157 | 120 |     | 150 | 135 | 120 |     |     |     |     |     |     | 132 |

**数据的整理** 县里的官员们接下来要按照 2.5 节中给出的方法为统计分析准备好数据。官员们需要确认这些估价员们在评定每一所房产时都是独立完成的, 而没有与任何其他估价员进行过商讨。然后, 要检查这些数据以避免数据传送时的错误, 并建立一个计算机文件, 为数据的统计分析做好准备。

**数据的分析** 由于设计不是一个完全的区组设计——在 256 个可能的区组 - 处理组合中只有 96 个被观测到——我们不能应用第十五章当中的模型和分析方法。该案例的分析将在 19.4 节中给出。

## 19.2 有一个或多个缺失观察值的随机化区组设计

对于任何设计, 只要观察值的个数不是对于所有因子水平的组合都一样, 我们就称这个设计是**非平衡**的。因此, 有一个或多个缺失观察值的随机化区组设计或拉丁方设计都是非平衡设计。我们以一个最简单的例子, 即有一个缺失观察值的随机化区组设计开始讨论。

对于有一个缺失观察值的随机化区组设计, 其方差分析很容易。首先, 我们估计那个缺失观察值的取值, 并校正估计的偏差, 然后利用随机化完全区组设计的公式即可。估计缺失观察值  $M$  的公式为

$$\hat{M} = \frac{ty_{i.} + by_{.j} - y_{..}}{(t-1)(b-1)}$$

其中  $t$  为处理的个数,  $b$  为区组的个数,  $y_{i.}$  为有缺失观察值的处理上所有观察值的和,  $y_{.j}$  为有缺失观察值的区组上所有观察值的和,  $y_{..}$  为所有观察值的和。用上述缺失观察值的估计值代替缺失的值, 代入平衡设计的公式, 计算出处理的平方和 SST。然后, 我们必须在 SST 上减去量

$$\text{Bias} = \frac{\{y_{\cdot j} - (t-1)\hat{M}\}^2}{t(t-1)}$$

来校正由于使用缺失值的估计而带来的偏差。也就是说,校正后的处理平方和为  $\text{SST}_C = \text{SST} - \text{Bias}$ 。处理的校正均方为  $\text{MST}_C = \text{SST}_C/(t-1)$ ,其期望值为  $\sigma_e^2 + b\theta_T$ ,与完全区组设计的情形相同。由此,可以对处理效应进行准确的  $F$  检验。

我们用以下例子来说明这个设计的方差分析。

### 例 19.1

为了确定辅以乳水的奶牛食料的营养价值,做一个试验。有五家奶牛场参加了这项研究。从每个奶牛场抽取由 5 头奶牛组成的随机样本,把其中的每一头奶牛随机地指定给四种处理中的一种,因而每个处理组有 5 头牛。

处理 1:只饮水。

处理 2:乳水加 30.2 升水/天。

处理 3:乳水加 15.1 升水/天。

处理 4:只饮乳水。

每个处理组中牛的食料除了以上列出的液体部分外,每头牛每天还喂 7.5 公斤谷物。

感兴趣的响应变量之一是每天干草的消费量。观测到的干草消费量数据(公斤/每头牛)列在表 19.1 中。不幸的是,如在表中所看到的,第二个奶牛场中按第四种方法喂养的牛从这项研究中撤出了,并且没有用别的牛来替代。那头牛得了一种传染病(这种病与处理无关),为了安全起见,把它从研究中撤了出来。

表 19.1 牛的干草消费量

| 奶牛场 | 处 理  |     |     |     |
|-----|------|-----|-----|-----|
|     | 1    | 2   | 3   | 4   |
| 1   | 15.4 | 9.6 | 9.5 | 8.4 |
| 2   | 14.8 | 9.3 | 9.4 | M   |
| 3   | 15.9 | 9.8 | 9.7 | 9.3 |
| 4   | 15.5 | 9.4 | 9.2 | 8.1 |
| 5   | 14.7 | 9.2 | 9.0 | 7.9 |

估计缺失值,并进行方差分析,取  $\alpha = 0.01$ 。

**解答** 对于这个  $b = 5, t = 4$  随机化区组设计,统计量  $y_{i\cdot}, y_{\cdot j}$  和  $y_{\cdot\cdot}$  定义如下:



$y_{i.}$  = 处理 4 上所有观察值的和

$$= 8.4 + 9.3 + 8.1 + 7.9 = 33.7$$

$y_{.j}$  = 区组 2 内所有观察值的和

$$= 14.8 + 9.3 + 9.4 = 33.5$$

$y_{..}$  = 所有观测值的和

$$= 15.4 + 9.6 + \cdots + 7.9 = 204.1$$

缺失值的估计为

$$\begin{aligned}\hat{M} &= \frac{ty_{i.} + by_{.j} - y_{..}}{(t-1)(b-1)} = \frac{4(33.7) + 5(33.5) - 204.1}{3(4)} \\ &= \frac{98.2}{12} = 8.183\end{aligned}$$

在估计了缺失值以后,我们可以用第十五章的公式来计算方差分析用的平方和。处理均值和区组均值为

$$\begin{array}{ll}\bar{y}_{1.} = 15.26 & \bar{y}_{.1} = 10.725 \\ \bar{y}_{2.} = 9.46 & \bar{y}_{.2} = 10.425 \\ \bar{y}_{3.} = 9.36 & \bar{y}_{.3} = 11.175 \\ \bar{y}_{4.} = 8.377 & \bar{y}_{.4} = 10.55 \\ & \bar{y}_{.5} = 10.20 \\ \bar{y}_{..} = 10.615\end{array}$$

注意,处理 4 上和区组 2 内的新的均值中使用了缺失值的估计值。类似地,所有观测值的平均值中也使用了缺失值的估计值。

$$TSS = (15.4 - 10.615)^2 + (9.6 - 10.615)^2 + \cdots + (7.9 - 10.615)^2 = 150.21$$

$$\begin{aligned}SSB &= 4\{(10.725 - 10.615)^2 + (10.425 - 10.615)^2 + (11.175 - 10.615)^2 \\ &\quad + (10.55 - 10.615)^2 + (10.2 - 10.615)^2\} = 2.16\end{aligned}$$

$$\begin{aligned}SST &= 5\{(15.26 - 10.615)^2 + (9.46 - 10.615)^2 + (9.36 - 10.615)^2 \\ &\quad + (8.377 - 10.615)^2\} = 147.48\end{aligned}$$

$$SSE = 150.21 - 2.16 - 147.48 = 0.57$$

$$\text{偏差校正量} = \frac{[33.7 - (4-1)8.183]^2}{4(4-1)} = 6.98$$

$$\text{校正后的处理平方和} = 147.48 - 6.98 = 140.50$$

非平衡和平衡设计的方差分析表的惟一区别在于,由于  $n$  表示实际的观测值的个数,非平衡设计的误差与平衡设计的误差相比,每缺失一个值就损失一个自由度。对于我们这个例子而言,其 AOV 表见表 19.2。

表 19.2 例 19.1 中数据的 AOV 表

| 来源  | SS     | df | MS     | F     | p 值    |
|-----|--------|----|--------|-------|--------|
| 奶牛场 | 2.16   | 4  | 0.54   |       |        |
| 处理  | 140.50 | 3  | 46.83  | 904.0 | 0.0001 |
| 误差  | 0.57   | 11 | 0.0518 |       |        |
| 总和  | 150.21 | 18 |        |       |        |

在  $\alpha=0.01$  下, 处理和区组的  $F$  检验都是显著的( $F$  检验的临界值分别为 6.22 和 5.67)。如我们从数据可以看到的, 处理一(只饮水)下的牛比其他任何辅以乳水饲养的牛吃了更多的干草。

看到方差分析的结果后, 我们可能希望在这些处理均值中做一些比较。我们将用 Fisher 的最小显著差异方法做两两比较。有缺失观察值的处理与任何其他处理之间的最小显著差异为

$$\text{LSD} = t_{\alpha/2} \sqrt{\text{MSE} \left( \frac{2}{b} + \frac{t}{b(b-1)(t-1)} \right)}$$

对于任何两个没有缺失值的处理, 最小显著差异如前, 即

$$\text{LSD} = t_{\alpha/2} \sqrt{\frac{2\text{MSE}}{b}}$$

在有多个缺失值的随机化区组设计中, 缺失值的估计公式变得更为复杂, 最小显著差异的公式也是如此。为此, 我们将考虑用拟合完全模型和简略模型来分析非平衡设计。下面我们首先以一个非平衡随机化区组设计为例来说明这种方法。

由于在第十二章中带有虚拟变量的一般线性模型的形式下用计算机进行分析需要更多的数据输入, 为了对处理进行检验, 我们把完全模型和简略模型表示如下:

完全模型(模型 1):  $y_{ij} = \mu + \beta_j + \alpha_i + \epsilon_{ij}$

简略模型(模型 2):  $y_{ij} = \mu + \beta_j + \epsilon_{ij}$

其中  $\beta_j$  是第  $j$  个区组效应,  $\alpha_i$  是第  $i$  个处理效应。

通过拟合模型 1(用 SAS 或其他计算机软件), 我们得到  $\text{SSE}_1$ 。同样, 拟合模型 2 得到  $\text{SSE}_2$ 。这两个误差平方和的差, 即  $\text{SSE}_2 - \text{SSE}_1$  给出由于处理而引起的平方和的下降。由于这是一个非平衡设计, 当比较处理均值时, 不能像平衡设计中那样把区组效应消掉(参见第十五章)。平方和的差  $\text{SSE}_2 - \text{SSE}_1$  则已经把由设计的非平衡性引起的任何区组效应消除了。这个差我们称之为针对区组调整后的处理平方和。

$$\text{SSE}_2 - \text{SSE}_1 = \text{SST}_{\text{adj}}$$

未对任何处理差异进行调整的区组平方和由差

$$SSB = TSS - SST_{adj} - SSE$$

给出, 这里 SSE 和 TSS 为来自于完全模型的平方和(注意: 我们也可以用 15.3 节中的公式得到未调整的区组平方和)。

检验处理效应的方差分析表由表 19.3 给出。表中,  $n$  为实际观察值的数目。

表 19.3 非平衡随机化区组设计中处理效应检验的 AOV 表

| 来源                | SS          | df              | MS          | F               |
|-------------------|-------------|-----------------|-------------|-----------------|
| 区组                | SSB         | $b - 1$         | —           | —               |
| 处理 <sub>adj</sub> | $SST_{adj}$ | $t - 1$         | $MST_{adj}$ | $MST_{adj}/MSE$ |
| 误差                | SSE         | $n - b - t + 1$ | MSE         |                 |
| 总和                | TSS         | $n - 1$         |             |                 |

为了计算检验区组效应的平方和, 完全模型(模型 1)仍然如前, 简略模型(模型 2)为

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

平方和的下降  $SSE_2 - SSE_1$  为针对处理效应调整后的区组平方和, 记为  $SSB_{adj}$ 。由此得到

$$SST = TSS - SSB_{adj} - SSE$$

AOV 表由表 19.4 给出。

表 19.4 非平衡随机化区组设计中区组效应检验的 AOV 表

| 来源                | SS          | df              | MS          | F               |
|-------------------|-------------|-----------------|-------------|-----------------|
| 区组 <sub>adj</sub> | $SSB_{adj}$ | $b - 1$         | $MSB_{adj}$ | $MSB_{adj}/MSE$ |
| 处理                | SST         | $t - 1$         | —           | —               |
| 误差                | SSE         | $n - t - b + 1$ | MSE         | —               |
| 总和                | TSS         | $n - 1$         |             |                 |

注意,  $SST$  和  $SST_{adj}$  在非平衡设计中不是同一个量。对于平衡设计, 它们是一样的。同样地,  $SSB$  和  $SSB_{adj}$  在非平衡设计中也是不同的量。对于一个非平衡设计, 我们有下列的等式:

$$TSS = SST_{adj} + SSB + SSE = SST + SSB_{adj} + SSE$$

但是

$$TSS \neq SST_{adj} + SSB_{adj} + SSE$$

## 练习

## 基本技能

19.1 参见例 19.1 中的数据以及下面用计算机得到的 SAS 输出结果。

General Linear Models Procedure: FULL MODEL

Dependent Variable: CONS

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 7  | 143.41548246   | 394.80  | 0.0001 |
| Error           | 11 | 0.57083333     |         |        |
| Corrected Total | 18 | 143.98631579   |         |        |

| Source | DF | Type III SS  | F Value | Pr > F |
|--------|----|--------------|---------|--------|
| DAIRY  | 4  | 2.11266667   | 10.18   | 0.0011 |
| TREAT  | 3  | 140.80083333 | 904.41  | 0.0001 |

General Linear Models Procedure: REDUCED MODEL WITHOUT DAIRY

Dependent Variable: CONS

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 3  | 141.30281579   | 263.28  | 0.0001 |
| Error           | 15 | 2.68350000     |         |        |
| Corrected Total | 18 | 143.98631579   |         |        |

| Source | DF | Type III SS  | F Value | Pr > F |
|--------|----|--------------|---------|--------|
| TREAT  | 3  | 141.30281579 | 263.28  | 0.0001 |

General Linear Models Procedure: REDUCED MODEL WITHOUT TREAT

Dependent Variable: CONS

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 4  | 2.61464912     | 0.06    | 0.9914 |
| Error           | 14 | 141.37166667   |         |        |
| Corrected Total | 18 | 143.98631579   |         |        |

| Source | DF | Type III SS | F Value | Pr > F |
|--------|----|-------------|---------|--------|
| DAIRY  | 4  | 2.61464912  | 0.06    | 0.9914 |

a. 给出检验处理效应的完全模型和简略模型。

b. 构造检验处理效应的方差分析表。对你的检验给定一个显著性水平,并给出你的结论。

19.2 参见例 19.1。用最小显著差异准则,找出哪些处理与其他处理不同,取  $\alpha = 0.05$ 。

19.3 参见例 15.1。假定第一个区组(地块 1)中的第一个观察值缺失。通过估计这个缺失的观察值进行数据分析,然后取  $\alpha = 0.05$  作方差分析。

19.4 参见练习 19.3。通过拟合完全模型和简略模型进行相应的分析,并把你的结论与在练习 19.3 中得出的结论相比较。

19.5 参见练习 19.1。拟合简略模型  $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$  以得到  $SSE_2$ 。平方和的下降是对处理效应调整后的区组平方和。验证通过这个方法计算得到的  $SSB_{adj}$  与练习 19.1 的计算机输出结果中第三个模型的平方和一系列中给出的值一样。

19.6 参见练习 15.3 中的数据。假定在火箭推进剂的试验中,3 号实验员在分析混合物 2 时,一台设备发生了故障。他们没有返回到实验室去准备再做一次,而是继续其余的试验,得到推进剂的推力数据。

a. 估计这个缺失值。

b. 用  $\alpha = 0.05$  作方差分析。

19.7 参见练习 19.6。

a. 用完全模型和简略模型进行方差分析,并把你的结果与练习 19.6 中所得结果进行比较。

b. 如果混合物 4 和实验员 1 处的响应值也缺失,你将如何分析这些数据?

### 19.3 有缺失数据的拉丁方设计

回想  $t \times t$  拉丁方设计可以用来在滤除另外两个变异来源(行和列)的同时,比较  $t$  个处理均值。这些处理被随机地分配,使得每个处理在每一行和每一列内都出现。在这一节中,我们来说明在只有一个缺失观察值的情况下拉丁方试验的方差分析方法。对于更复杂的设计,我们将使用拟合带有缺失观测值的完全模型和简略模型的一般方法,就像 19.2 节对随机化区组设计所谈到的那样。

在拉丁方设计中,估计单个缺失值的公式为

$$\hat{M} = \frac{t(y_{i.} + y_{.j} + y_k) - 2y_{..}}{(t-1)(t-2)}$$

其中  $y_{i.}$ ,  $y_{.j}$  和  $y_k$  分别表示对应于确实观测值的行、列和处理总和,  $y_{..}$  为试验中所有观测值的总和,  $t$  为拉丁方设计中的处理个数。

总平方和、行平方和、列平方和、处理平方和以及误差平方和分别用第十五章中完全的拉丁方设计中的公式计算,其中把缺失的观测值用其估计  $\hat{M}$  代替。处

理的均方为平衡拉丁方设计中处理的期望均方的有偏估计。这个偏差由

$$\text{Bias} = \left( \frac{y_{..} - y_{i.} - y_{.j} - (t-1)y_k}{(t-1)(t-2)} \right)^2$$

估计。校正处理平方和由

$$\text{SST}_C = \text{SST} - \text{Bias}$$

给出。 $\text{MST}_C = \text{SST}_C / (t-1)$  为  $\sigma_e^2 + t\theta_T$  的无偏估计。令  $n = t^2 - 1$  为该拉丁方设计中观测到的数据值的个数, 我们得到如下当缺失的那个数据用  $\hat{M}$  估计时拉丁方设计的方差分析表。

表 19.5 有一个缺失值的拉丁方设计的 AOV 表

| 来源 | SS             | df           | MS             | F                           |
|----|----------------|--------------|----------------|-----------------------------|
| 行  | SSR            | $t-1$        | MSR            | —                           |
| 列  | SSC            | $t-1$        | MSC            | —                           |
| 处理 | $\text{SST}_C$ | $t-1$        | $\text{MST}_C$ | $\text{MST}_C / \text{MSE}$ |
| 误差 | SSE            | $n - 3t + 2$ | MSE            |                             |
| 总和 | TSS            | $n - 1$      |                |                             |

### 例 19.2

一家公司为了挑选投放于一项大规模消费者接受性调查中的长筒袜子, 考虑了很多种不同类型的尼龙袜子的性质(如强度, 伸缩性等等)。

有五种袜子(A, B, C, D 和 E)通过了基本的筛选, 进入下一轮更严格的检验。作为这个检验的一部分, 从每种袜子中挑选 5 个样品, 5 个测试人员分别在 5 天中测试 5 种袜子, 以比较这些袜子的伸缩性, 测试时所用的拉力相同。分配测试样品时按一个拉丁方的随机分配进行。拉伸数据(cm)在表 19.6 中给出。

表 19.6 例 19.2 中的拉伸数据

| 测试员 | 时 间    |        |        |        |        |  |  |  |  |  |
|-----|--------|--------|--------|--------|--------|--|--|--|--|--|
|     | 1      | 2      | 3      | 4      | 5      |  |  |  |  |  |
| 1   | B 22.1 | A 18.6 | C 23.0 | E 24.3 | D 17.1 |  |  |  |  |  |
| 2   | C 23.5 | D 16.5 | A 18.7 | B 22.0 | E M    |  |  |  |  |  |
| 3   | D 17.4 | E 23.8 | B 22.8 | C 23.9 | A 20.0 |  |  |  |  |  |
| 4   | A 20.3 | B 23.4 | E 25.9 | D 18.7 | C 24.2 |  |  |  |  |  |
| 5   | E 25.7 | C 24.8 | D 18.9 | A 20.6 | B 24.6 |  |  |  |  |  |

注意,第2号测试人员E类袜子的观测值丢失了,同时试验中也没有为得到该值重新去做试验。用这节中的方法估计这个缺失的值。

**解答** 对于例中的数据,对应于缺失观测值的处理、行和列的总和为

$$y_{5.} = 99.70 \quad y_{2.} = 80.70 \quad y_{.5} = 85.90 \quad y_{..} = 520.80$$

由于  $t = r = c = 5$ , 我们有

$$\hat{M} = \frac{5(80.70 + 85.90 + 99.70) - 2(520.80)}{(5-1)(5-2)} = 24.1583$$

下面我们把缺失观测值用其最小二乘估计  $\hat{M}$  代替,用第十五章的公式计算平方和。

| 测试员                     | 时间                        | 类型                     | 总和                        |
|-------------------------|---------------------------|------------------------|---------------------------|
| $\bar{y}_{1.} = 21.020$ | $\bar{y}_{.1} = 21.800$   | $\bar{y}_1 = 19.640$   | $\bar{y}_{..} = 21.79833$ |
| $y_{2.} = 20.97166$     | $\bar{y}_{.2} = 21.420$   | $\bar{y}_2 = 22.980$   |                           |
| $\bar{y}_{3.} = 21.580$ | $\bar{y}_{.3} = 21.860$   | $y_3 = 23.880$         |                           |
| $\bar{y}_{4.} = 22.500$ | $\bar{y}_{.4} = 21.900$   | $\bar{y}_4 = 17.720$   |                           |
| $\bar{y}_{5.} = 22.920$ | $\bar{y}_{.5} = 22.01166$ | $\bar{y}_5 = 24.77166$ |                           |

$$\begin{aligned} \text{TSS} &= (22.1 - 21.79833)^2 + (18.6 - 21.79833)^2 + \cdots + (24.6 - 21.79833)^2 \\ &= 197.20 \end{aligned}$$

$$\begin{aligned} \text{SSR} &= 5\{(21.020 - 21.79833)^2 + (20.97166 - 21.79833)^2 + \cdots + (22.920 \\ &\quad - 21.79833)^2\} = 15.44 \end{aligned}$$

$$\begin{aligned} \text{SSC} &= 5\{(21.8 - 21.79833)^2 + (21.42 - 21.79833)^2 + \cdots + (22.01166 \\ &\quad - 21.79833)^2\} = 1.01 \end{aligned}$$

$$\begin{aligned} \text{SST} &= 5\{(19.64 - 21.79833)^2 + (22.98 - 21.79833)^2 + \cdots + (24.77166 \\ &\quad - 21.79833)^2\} = 179.31 \end{aligned}$$

$$\text{SSE} = 197.20 - 15.44 - 1.01 - 179.31 = 1.44$$

$$\text{Bias} = \left( \frac{520.80 - 80.70 - 85.90 - (5-1)99.70}{(5-1)(5-2)} \right)^2 = 13.82$$

$$\text{校正的处理} = \text{SST}_C = 179.31 - 13.82 = 165.49$$

这项研究的方差分析表如下。

| 来源  | SS     | df | MS    | F      |
|-----|--------|----|-------|--------|
| 测试员 | 15.44  | 4  | 3.86  | —      |
| 时间  | 1.01   | 4  | 0.25  |        |
| 种类  | 165.49 | 4  | 41.37 | 316.04 |
| 误差  | 1.44   | 11 | 0.13  |        |
| 总和  | 197.20 | 23 |       |        |

当处理效应显著时,我们可以用下面的公式做两两比较。有缺失观察值的处理与任何其他处理之间的最小显著差异为

$$\text{LSD} = t_{\alpha/2} \sqrt{\text{MSE} \left( \frac{2}{t} + \frac{1}{(t-1)(t-2)} \right)}$$

对于任何两个没有缺失值的处理,最小显著差异如前,即

$$\text{LSD} = t_{\alpha/2} \sqrt{\frac{2\text{MSE}}{t}}$$

MSE 的值从方差分析表中得到。

对于有多个缺失值的拉丁方设计,用拟合完全模型和简略模型的方法来调整由于缺失数据而造成的非平衡设计中的处理平方和更容易些。完全模型如下:

$$\text{模型 1: } y_{ij} = \mu + \alpha_k + \beta_i + \gamma_j + \epsilon_{ij}$$

其中  $y_{ij}$  为第  $i$  行第  $j$  列处理  $k$  上的观察值。不用估计缺失值,用观测数据拟合这个模型,我们得到误差平方和,记之为  $\text{SSE}_1$ 。然后,不用估计缺失值,我们用观测数据拟合没有处理效应的简略模型

$$\text{模型 2: } y_{ij} = \mu + \beta_i + \gamma_j + \epsilon_{ij}$$

又得到一个误差平方和,记之为  $\text{SSE}_2$ 。这两个误差平方和的差就是处理的校正平方和:

$$\text{SST}_C = \text{SSE}_2 - \text{SSE}_1$$

对于处理效应的检验为表 19.6 中给出的  $F$  检验

$$F = \frac{\text{SST}_C / (t-1)}{\text{SSE}_1 / (n-3t+2)}$$

其中  $n$  为观测到的数据个数。我们可以用类似的方法得到行和列的校正平方和。通过拟合包含处理效应和行效应,但不包含列效应的简略模型,我们可以得到用来构造调整后的列平方和的误差平方和。同样,我们也可以得到调整后的行平方和。在多数情况下,对列效应和行效应的显著性检验是不感兴趣的。

## 练习

### 基本技能

**19.8** 参见例 19.2 和下面给出的处理平方和和误差平方和计算的计算机输出结果。比较计算机输出的结果与用估计的缺失值算出的结果。



General Linear Models Procedure: FULL MODEL FOR EXAMPLE 19.2

Dependent Variable: ELONG

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 12 | 189.95683333   | 120.66  | 0.0001 |
| Error           | 11 | 1.44316667     |         |        |
| Corrected Total | 23 | 191.40000000   |         |        |

| Source  | DF | Type I SS    | F Value | Pr > F |
|---------|----|--------------|---------|--------|
| INVEST  | 4  | 22.32850000  | 42.55   | 0.0001 |
| DAY     | 4  | 2.13400000   | 4.07    | 0.0291 |
| VERSION | 4  | 165.49433333 | 315.35  | 0.0001 |

| Source  | DF | Type III SS  | F Value | Pr > F |
|---------|----|--------------|---------|--------|
| INVEST  | 4  | 14.36883333  | 27.38   | 0.0001 |
| DAY     | 4  | 0.94283333   | 1.80    | 0.1998 |
| VERSION | 4  | 165.49433333 | 315.35  | 0.0001 |

General Linear Models Procedure: REDUCED MODEL WITHOUT TREATMENT

Dependent Variable: ELONG

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 8  | 24.46250000    | 0.27    | 0.9646 |
| Error           | 15 | 166.93750000   |         |        |
| Corrected Total | 23 | 191.40000000   |         |        |

| Source | DF | Type I SS   | F Value | Pr > F |
|--------|----|-------------|---------|--------|
| INVEST | 4  | 22.32850000 | 0.50    | 0.7352 |
| DAY    | 4  | 2.13400000  | 0.05    | 0.9952 |

| Source | DF | Type III SS | F Value | Pr > F |
|--------|----|-------------|---------|--------|
| INVEST | 4  | 23.49000000 | 0.53    | 0.7172 |
| DAY    | 4  | 2.13400000  | 0.05    | 0.9952 |

**19.9** 参见练习 19.8。

- 检验各种长筒袜子的平均伸缩性之间的差异是否显著。
- 用最小显著差异准则确定哪两种类型的袜子之间有显著差异。

## 应用

**19.10(环境)** 一家石油公司要比较四种不同品牌(I, II, III, IV)的汽油每加仑所能行驶的英里数。考虑到由于司机和车型的不同,所能行驶的里程会有很大的差异,在下面的拉丁方设计中,这两种外部变异来源被用做“区组”变量。按照一个拉丁方设计,每个司机用指定品牌的汽油,在标准的跑道上驾驶每一种型号的汽车。然而,当3号司机用II型汽油驾驶第4种型号的汽车时,汽车的化油器发生了故障。这种故障直到研究完成后才被发现,因而无法得到替代的数据。每加仑的英里数如下。

| 司机 | 车 型 |      |     |      |     |      |     |      |
|----|-----|------|-----|------|-----|------|-----|------|
|    | 1   |      | 2   |      | 3   |      | 4   |      |
| 1  | IV  | 15.5 | II  | 33.9 | III | 13.2 | I   | 29.1 |
| 2  | II  | 16.3 | III | 26.6 | I   | 19.4 | IV  | 22.8 |
| 3  | III | 10.8 | I   | 31.1 | IV  | 17.1 | II  | —    |
| 4  | I   | 14.7 | IV  | 34.0 | II  | 19.7 | III | 21.6 |

a. 估计缺失值,做方差分析,取  $\alpha = 0.05$ 。

b. 取  $\alpha = 0.05$ ,用 Fisher 的最小显著差异方法进行处理之间的比较。

**19.11** 对于练习 19.10 中的数据,用拟合完全模型和简略模型的方法进行方差分析。

## 19.4 平衡不完全区组(BIB)设计

本章中到目前为止所讨论的设计中的不平衡,都是由于在试验或数据处理过程中发生了某些事件,从而引起了不可预见的情形而造成的。然而,有时为了实现做试验的目的,我们不得不牺牲一定的平衡而设计一个试验。这经常发生在每个区组上的试验单元个数少于所考虑的处理的个数的情形。考虑下面的例子。

### 例 19.3

一家化学公司的质量控制实验室需要评定一种涂料的 5 种不同的形式(A, B, C, D, E)所产生的颜色是否一致。按惯例每种形式取 4 个样品。实验室有 5 名技术员来做这项试验,每个技术员每天最多可以评估 4 个样品。因此,不可能进行随机化完全区组试验,因为不可能所有形式都被每一位技术员评估。然而,在设计上有可能达到一种部分平衡,即每两种形式的涂料有相同个数的技术员来评估。表 19.7 中给出这样的—一个设计。

表 19.7 指定涂料给质量控制技术员

| 技术员 | 形 式 |   |   |   |
|-----|-----|---|---|---|
| 1   | D   | B | A | E |
| 2   | E   | A | A | D |
| 3   | A   | C | D | B |
| 4   | C   | E | B | A |
| 5   | B   | D | E | C |

注意,每对形式的涂料由三个技术员来评定。

任何随机化区组设计,如果要考察的处理的个数  $t$  比每个区组内可用的试验单元的个数来得大,就称之为**不完全区组设计**。比如说,当有若干含有  $k(<t)$  个试验单元的大小一致的区组存在或者可以构造出来时,不完全的区组设计就是不可避免的。然而,在一个设计中,有可能达到部分平衡。这样的不完全区组设计之一定义如下。

**定义 19.1** 一个平衡不完全区组(BIB, balanced incomplete block design)设计中有  $t$  个处理,分配到  $b$  个区组,满足如下条件:

1. 每个区组含有  $k(<t)$  个试验单元;
2. 每个处理在每个区组中最多出现一次;
3. 每个区组包含  $k$  个处理;
4. 每个处理恰好在  $r$  个区组中出现;
5. 每对处理一起在  $\lambda$  个区组中出现。

从定义 19.1,我们可以得出,一个设计之为 BIB 设计,必有

- 在一个区组中,每一对处理一起出现的次数相同。
- 每个处理被观测  $r$  次。
- 观测值的个数  $n$  必然满足  $n = rt = kb$ 。
- $\lambda < r < b$ 。
- $\lambda = r(k-1)/(t-1)$  必为整数。

#### 例 19.4

参见例 19.3。我们有  $b=5$  个区组(技术员)和  $t=5$  个处理(形式)。每个区组有  $k=4$  个处理,从而  $k=4<5=t$ ,这导致不完全区组设计,其中每种形式出现在  $r=4$  个区组中。为使该设计是一个 BIB 设计,需要每对形式都由  $\lambda = r(k-1)/(t-1) = 4(4-1)/(5-1) = 3$  个技术员来评估。考察表 19.7 中把各种形式分

配给技术员的方案,我们看到每对形式由 3 个技术员来评估。因此,表 19.7 中给出的设计为 BIB 设计。

在许多情况下,我们设计一个试验时不能完全按照理想来进行,因为并非对所有  $t, k, b, r$  值的组合都存在 BIB 设计。例如,假定我们要考察  $t = 6$  个处理,并且有  $b = 4$  个区组,每个区组包含  $k = 3$  个试验单元,这样,每个处理可以被观测  $r = 2$  次。然而,要使设计是一个 BIB 设计,  $\lambda = r(k-1)/(t-1)$  应为整数。但事实上,  $\lambda = 2(3-1)/(6-1) = 4/5$ , 显然不是整数。因而,对于这组处理和区组,我们不能构造出 BIB 设计。BIB 设计有一套方法,同时也有更复杂的不完全区组设计。Cochran and Cox (1957), Lentner and Bishop (1993) 和 Kuehl (1999) 的书中有 BIB 设计的表格以及构造这些设计的方法。有些统计软件(如 SAS 和 Minitab)对于给定的  $t, k, b$  和  $r$  值可以构造出 BIB 设计。

一个平衡不完全设计的方差分析,可以通过针对该设计提出的公式进行,也可以像非平衡设计那样,用拟合完全模型和简略模型的方法进行。对于表 19.8 中的方差分析表,我们将给出其简单的计算公式。

表 19.8 一个平衡不完全区组设计的方差分析表

| 来源 | SS          | df              | MS          | F               |
|----|-------------|-----------------|-------------|-----------------|
| 区组 | SSB         | $b - 1$         | —           | —               |
| 处理 | $SST_{adj}$ | $t - 1$         | $MST_{adj}$ | $MST_{adj}/MSE$ |
| 误差 | SSE         | $n - t - b + 1$ | MSE         |                 |
| 总和 | TSS         | $n - 1$         |             |                 |

BIB 设计的模型如下:

$$y_{ijg} = \mu + \alpha_i + \beta_j + \epsilon_{ijg} \quad i = 1, \dots, t; \quad j = 1, \dots, b; \quad g = n_{ij}$$

其中当第  $i$  个处理在第  $j$  个区组中出现时,  $n_{ij} = 1$ , 否则等于 0。出现在模型中的项  $\mu$  为总均值,  $\alpha_i$  为第  $i$  个处理的效应,  $\beta_j$  为第  $j$  个区组效应,  $\epsilon_{ijg}$  相互独立并且服从均值为 0, 方差为  $\sigma_e^2$  的正态分布。与前面一样,从这个模型中,我们按照下面的公式计算出 SSB(相对于处理进行了调整的区组平方和)以及总平方和:

$$TSS = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$$

其中  $n = rt = bk$  为实际数据值的个数,

$$SSB = k \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2$$

其中  $\bar{y}_{.j}$  为第  $j$  个区组中观测值的平方和,  $\bar{y}_{..}$  为总平均。然后,我们定义

$y_{i.} =$  第  $i$  个处理上所有观测值的和

$B_{(i)}$  = 包含第  $i$  个处理的各个区组上所有观测值的和  
 相对于区组调整后的处理平方和为

$$SST_{\text{adj}} = \frac{t-1}{nk(k-1)} \sum_i (ky_{i\cdot} - B_{(i)})^2$$

误差平方和为

$$SSE = TSS - SSB - SST_{\text{adj}}$$

如方差分析表中所示, 对于“各个处理的均值没有差异”的假设, 检验统计量为  $MST_{\text{adj}} = MSE$ 。

### 例 19.5

一家大公司在某个地区选取了一个由 12 名潜在消费者组成的随机样本, 以在他们的帮助下比较 8 种试验枕头和一种已经投放市场的枕头的特征(例如坚实性和弹性)。从以前的研究中该公司知道, 在一段给定的时间内, 多数人可以利用其注意力集中的时间评估 3 个枕头, 因而, 公司决定使用表 19.9 中给出的设计。

表 19.9 例 19.5 中的舒适得分

| 区组<br>(消费者) |   | 处理<br>(枕头) |   |    |   |    | 区组总和  | 区组均值 |
|-------------|---|------------|---|----|---|----|-------|------|
| 1           | A | 59         | B | 26 | C | 38 | 123   | 41   |
| 2           | D | 85         | E | 92 | F | 69 | 246   | 82   |
| 3           | G | 74         | H | 52 | I | 27 | 153   | 51   |
| 4           | A | 63         | D | 70 | G | 68 | 201   | 67   |
| 5           | B | 26         | E | 98 | H | 59 | 183   | 61   |
| 6           | C | 31         | F | 60 | I | 35 | 126   | 42   |
| 7           | A | 62         | E | 85 | I | 30 | 177   | 59   |
| 8           | B | 23         | F | 73 | G | 75 | 171   | 57   |
| 9           | C | 49         | D | 74 | H | 51 | 174   | 58   |
| 10          | A | 52         | F | 76 | H | 43 | 171   | 57   |
| 11          | B | 18         | D | 79 | I | 41 | 138   | 46   |
| 12          | C | 42         | E | 84 | G | 81 | 207   | 69   |
| 总和          |   |            |   |    |   |    | 2 070 | 57.5 |

首先为各种枕头随机地指定一个从 A 到 I 的字母, 然后准备一些台子, 把 3 种不同种类的枕头按照表 19.9 中的安排放到每个台子上。所有枕头都用相同的白色枕套封起来, 从外表颜色上相互之间没有区别。枕套上的惟一标记是一个四位

数字的号码,作为研究人员辨别这些枕头的标识。所有这些准备停当以后,把 12 名潜在的消费者随机地指定到一个台子,以比较该台子上的 3 种枕头。这些消费者按舒适程度为每种枕头打一个分,分值从 1 到 100(分值越高表示枕头越舒适)。这些枕头的得分记录在表 19.9 中(其中字母表示枕头类型, A 代表已经投放市场的枕头)。

验证此中使用的设计为 BIB 设计。用本节中的公式进行方差分析,取  $\alpha = 0.05$  检验 9 种枕头的平均舒适度得分之间有无显著差异。

**解答** 我们需要验证上述设计满足 BIB 设计的所有条件。我们注意到,在此设计中,有 9 个处理(枕头), 12 个区组(消费者), 每个区组有三个观察值(每个消费者比较 3 种枕头), 并且每种枕头由 4 名消费者为其打分, 每个消费者对每一种枕头至多打分一次。也就是说,  $t = 9, b = 12, k = 3, r = 4$ , 由此,  $n = (9)(4) = (12)(3) = 36$ 。

下面计算  $\lambda$ :  $\lambda = r(k-1)/(t-1) = 4(3-1)/(9-1) = 1$ , 这就是说, 每对枕头的类型恰有一名消费者打分。我们可以从表 19.9 直接验证这一点。因此我们得出, 该研究中的设计为一个 BIB 设计。为便于使用本节中给出的公式进行分析, 我们把各个总和和均值作成表, 见表 19.10。

表 19.10 表 19.9 中数据的总和

| 处理 | $y_{i\cdot}$ | $B_{(i)}$ | $ky_{i\cdot} - B_{(i)}$ |
|----|--------------|-----------|-------------------------|
| A  | 236          | 672       | 36                      |
| B  | 93           | 615       | -336                    |
| C  | 160          | 630       | -150                    |
| D  | 308          | 759       | 165                     |
| E  | 359          | 813       | 264                     |
| F  | 278          | 714       | 120                     |
| G  | 298          | 732       | 162                     |
| H  | 205          | 681       | -66                     |
| I  | 133          | 594       | -195                    |
| 总和 |              | 2 070     | 0                       |

为了说明表 19.10 中各个值的来历, 我们考虑与处理 A 有关的各值:

$$y_{i\cdot} = \text{处理 A 上所有值的和} = 59 + 63 + 62 + 52 = 236$$

$$B_{(i)} = \text{包含 A 的各个区组上总和的和} = 123 + 201 + 177 + 171 = 672$$

$$ky_{i\cdot} - B_{(i)} = (3)236 - 672 = 36$$

由表 19.9 和表 19.10 中的值计算平方和, 我们得到

$$SST_{\text{adj}} = \frac{t-1}{nk(k-1)} \sum_i (ky_{i\cdot} - B_{(i)})^2 = \frac{(9-1)(316\ 638)}{(36)(3)(3-1)} = 11\ 727.33$$

类似地, 用表 19.9 中的区组均值, 我们有

$$SSB = k \sum_j (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 = 3[(41 - 57.5)^2 + \cdots + (69 - 57.5)^2] = 4\ 575$$

用表 19.9 中的值, 我们得到总平方和

$$TSS = \sum_{ij} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = [(59 - 57.5)^2 + \cdots + (81 - 57.5)^2] = 16\ 861$$

以及误差平方和

$$SSE = TSS - SSB - SST_{\text{adj}} = 16\ 861 - 4\ 575 - 11\ 727.33 = 558.67$$

检验 9 种枕头的平均舒适度得分之间有无显著差异的方差分析表由表 19.11 给出。由于计算所得的  $F$  值 41.98 大于与  $df_1 = 8, df_2 = 16$  和  $\alpha = 0.05$  对应的表值 2.59, 我们说 9 种枕头的平均舒适度得分之间有显著差异 ( $p$  值  $< 0.0001$ )。

表 19.11 例 19.5 中数据的 AOV 表

| 来源  | SS        | df | MS       | $F$   | $p$ 值  |
|-----|-----------|----|----------|-------|--------|
| 消费者 | 4 575     | 11 | 415.91   | —     | —      |
| 处理  | 11 727.33 | 8  | 1 465.92 | 41.98 | 0.0001 |
| 误差  | 558.67    | 16 | 34.92    | —     | —      |
| 总和  | 16 861    | 35 | —        | —     | —      |

鉴于处理均值差异的  $F$  检验是显著的, 我们很自然地想看一看哪些处理均值与其他均值之间有显著的差异。为此, 我们引入如下记号: 以  $\hat{\mu}_i$  记处理  $i$  的均值的估计

$$\hat{\mu}_i = \bar{y}_{\cdot\cdot} + \frac{ky_{i\cdot} - B_{(i)}}{t\lambda}$$

其中  $\bar{y}_{\cdot\cdot}$  为总的样本均值。于是, 处理  $i$  和处理  $i'$  的均值之间的差的估计为

$$\hat{\mu}_i - \hat{\mu}_{i'} = \frac{[ky_{i\cdot} - B_{(i)}] - [ky_{i'\cdot} - B_{(i')}]}{t\lambda}$$

任何一对处理均值之间的最小显著差异为

$$LSD = t_{\alpha/2} \sqrt{\frac{2kMSE}{t\lambda}}$$

### 例 19.6

计算处理均值的估计, 并用  $\alpha = 0.05$  给出所有处理均值对之间的差。

**解答** 对于例 19.5 中的 BIB 设计, 我们有  $\bar{y}_{..} = 57.5$ ,  $t = 9$ ,  $\lambda = 1$ 。因此, 用表 19.10 中  $ky_{i.} - B_{(i)}$  的一列, 我们以

$$\hat{\mu}_i = \bar{y}_{..} + \frac{ky_{i.} - B_{(i)}}{t\lambda} = 57.5 + \frac{ky_{i.} - B_{(i)}}{(9)(1)}$$

来计算估计的处理均值。

| 处理 | $\bar{y}_{i.}$ | $ky_{i.} - B_{(i)}$ | $\hat{\mu}_i$ |
|----|----------------|---------------------|---------------|
| A  | 59.00          | 36                  | 61.50         |
| B  | 23.25          | -336                | 20.17         |
| C  | 40.00          | -150                | 40.83         |
| D  | 77.00          | 165                 | 75.83         |
| E  | 89.75          | 264                 | 86.83         |
| F  | 69.50          | 120                 | 70.83         |
| G  | 74.50          | 162                 | 75.50         |
| H  | 51.25          | -66                 | 50.17         |
| I  | 33.25          | -195                | 35.83         |

注意, 比较原来的处理均值  $\bar{y}_{i.}$  与最小二乘均值估计  $\hat{\mu}_i$ , 可以看出, 有的均值变大了, 而有的均值变小了, 这与处理所在的各个区组总和的相对大小有关。

由于  $MSE = 34.92$ ,  $df_{Error} = 16$ , 我们得到

$$LSD = t_{\alpha/2} \sqrt{\frac{2kMSE}{t\lambda}} = 2.12 \sqrt{\frac{2(3)(34.92)}{(9)(1)}} = 10.23$$

下面把这 9 个最小二乘均值估计按升序列出, 并附以用 LSD 比较各对均值差异显著性的结果, 其中在同一条线上面的两个处理彼此之间的差异是不显著的。

|       |       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| B     | I     | C     | H     | A     | F     | G     | D     | E     |
| 20.17 | 35.83 | 40.83 | 50.17 | 61.50 | 70.83 | 75.50 | 75.83 | 86.83 |

另外, 处理的调整后的平方和以及相应的  $F$  检验的计算也可以通过拟合两个



模型进行。首先,拟合一个既包含区组效应也包含处理效应的完全模型,得到  $SSE_1$ , 然后,拟合一个没有区组效应的简略模型得到  $SSE_2$ 。处理的调整平方和  $SST_{Adj}$  由下式计算

$$SST_{Adj} = SSE_2 - SSE_1$$

其自由度为  $df_{T-t} = df_{E_2} - df_{E_1}$ 。处理效应的  $F$  检验统计量为  $F = MST_{Adj} / MSE_1$ 。

**财产估价案例的分析** 此项研究的目的在于确定县里的估价员们对于居民财产价值的评价是否一致。研究中的因子有区组因子,即 16 种居民财产,以及处理因子,即 16 名县财产估价员。处理效应是随机的,因为这 16 名估价员是从县内估价员的总体中随机挑选出来的,并且县里的官员感兴趣的不仅是这参加到研究中的 16 名估价员,而且还对有关全体估价员的结果感兴趣。设计是一个不完全区组设计,因为每个处理(估价员)在 16 个区组(财产)中只被观测 6 次。下面我们验证该设计是一个 BIB 设计。

首先,我们找出 BIB 设计中的参数:

$$t = 16 \quad b = 16 \quad k = 6 \quad r = 6$$

BIB 设计要求有  $n = (16)(6) = 96$  个观察值,并且  $\lambda = 6(6-1)/(16-1) = 2$ 。由此,我们得出,要使该研究是 BIB 设计,必须满足如下条件:每一对估价员同时评估 2 种财产,每一名估价员必须评估 16 种财产中的 6 种,我们共得到 96 个估价值。考察给出的数据可以看出,这些条件都得到了满足。下面,为了分析这些数据,我们来拟合所需要的模型。联系价值的变异与估价员效应、财产效应以及其他变异来源的模型由

$$\text{完全模型 I: } y_{ijg} = \mu + \alpha_i + \beta_j + \epsilon_{ijg}$$

其中  $\mu$  为所有估价员所估价值的总平均值,  $\alpha_i$  为估价员  $i$  对价值的随机效应,  $\beta_j$  为财产  $j$  对价值的随机效应,  $\epsilon_{ijg}$  表示估价的所有其他变异来源的随机效应。其次,我们拟合简略模型。第一个简略模型为没有估价员效应的模型。

$$\text{简略模型 I: } y_{ijg} = \mu + \beta_j + \epsilon_{ijg}$$

从这个模型,我们可以得到来源于估价员的调整后的平方和。接下来,我们拟合没有财产效应的模型。

$$\text{简略模型 II: } y_{ijg} = \mu + \alpha_i + \epsilon_{ijg}$$

从这个模型,我们可以得到来源于财产的调整后的平方和。

下面给出计算机输出结果。这些结果包含通过拟合上述三个模型得到的误差平方和,即  $SSE_{Full}$ ,  $SSE_{Red I}$  和  $SSE_{Red II}$ 。

General Linear Models: FULL MODEL

Dependent Variable: VALUATION

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 30 | 16976.0919219  | 4.51    | 0.0001 |
| Error           | 65 | 8161.2414114   |         |        |
| Corrected Total | 95 | 25137.3333333  |         |        |

| Source | DF | Type III SS   | F Value | Pr > F |
|--------|----|---------------|---------|--------|
| ASR    | 15 | 3759.0919219  | 2.00    | 0.0291 |
| P      | 15 | 10343.8800172 | 5.49    | 0.0001 |

General Linear Models: REDUCED MODEL WITHOUT TREATMENT VARIABLE(ASSESSOR)

Dependent Variable: VAL

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 15 | 13217.0000000  | 5.91    | 0.0001 |
| Error           | 80 | 11920.3333333  |         |        |
| Corrected Total | 95 | 25137.3333333  |         |        |

| Source | DF | Type III SS   | F Value | Pr > F |
|--------|----|---------------|---------|--------|
| P      | 15 | 13217.0000000 | 5.91    | 0.0001 |

General Linear Models: REDUCED MODEL WITHOUT BLOCK VARIABLE(PROPERTY)

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 15 | 6632.21190476  | 1.91    | 0.0339 |
| Error           | 80 | 18505.12142857 |         |        |
| Corrected Total | 95 | 25137.3333333  |         |        |

| Source | DF | Type III SS   | F Value | Pr > F |
|--------|----|---------------|---------|--------|
| ASR    | 15 | 6632.21190476 | 1.91    | 0.0339 |

对于由各个估价员之间的差异引起的平均估价的差异的统计显著性的检验,按如下方法进行。

$$SST_{Adj} = SSE_{Red I} - SSE_{Full} = 11\,920.33 - 8\,161.24 = 3\,759.09$$

其自由度为  $df_{Trt} = df_{ERed I} - df_{EFull} = 80 - 65 = 15$ 。现在,我们来检验是否由于估价员之间的差异造成所估价值具有显著的变异。由于估价员是变异的一个具有随机性的来源,我们是要检验

$$H_0 : \sigma_a^2 = 0 \quad H_a : \sigma_a^2 \neq 0$$

计算出检验统计量的值

$$F = \frac{SST_{Adj}/df_{Trr}}{SSE_{Full}/df_{EFull}} = \frac{3\,759.09/15}{8\,161.24/65} = 2.00$$

其  $p$  值 = 0.0291。我们可以将该  $F$  值与自由度为  $df_1 = 15$ ,  $df_2 = 65$  的  $F$  分布表中的 0.05 百分位数 1.82 相比较, 得到这样的结果: 由于估价员之间的差异导致了显著的变异。类似地, 我们得到由于财产之间的差异而产生的调整后的平方和。

$$SSR_{Adj} = SSE_{Red II} - SSE_{Full} = 18\,505.12 - 8\,161.24 = 10\,343.88$$

其自由度  $df_{Block} = df_{ERed II} - df_{EFull} = 80 - 65 = 15$ 。我们可以把所得到的结果总结成如下的方差分析表。

| 来源  | df | SS        | EMS                              | $F$  | $p$ 值  |
|-----|----|-----------|----------------------------------|------|--------|
| 财产  | 15 | 10 343.88 | $\sigma_e^2 + 5.33\sigma_{财产}^2$ | —    | —      |
| 估价员 | 15 | 3 759.09  | $\sigma_e^2 + 5.33\sigma_{估}^2$  | 2.00 | 0.0291 |
| 误差  | 65 | 8 161.24  | $\sigma_e^2$                     | —    | —      |

注意, 财产方差和估价员方差的乘数不像随机化完全去组设计中那样为 16。由于设计的不完全性, 我们得到如下的期望均方的值:

$$\text{区组的期望均方: } EMS_{Block} = \sigma_e^2 + \frac{bk - t}{b - 1} \sigma_{Block}^2$$

以及

$$\text{处理的期望均方: } EMS_{Trr} = \sigma_e^2 + \frac{\lambda t}{k} \sigma_{Trr}^2$$

从上述 AOV 表, 我们可以得到如下的方差分量的估计值。

$$\hat{\sigma}_e^2 = 8\,161.24/65 = 125.56$$

$$\hat{\sigma}_{财产}^2 = (10\,343.88/15 - 125.56)/5.33 = 105.82$$

$$\hat{\sigma}_{估}^2 = (3\,759.09/15 - 125.56)/5.33 = 23.46$$

由此, 我们得到如下的估价中变异性的比例分配。

| 变差来源 | 方差的估计  | 占总变差的比例(%) |
|------|--------|------------|
| 财产   | 105.82 | 41.5       |
| 评估员  | 23.46  | 9.2        |
| 误差   | 125.56 | 49.3       |
| 总和   | 254.84 | 100        |

虽然我们发现由于估价员而产生的变异性是显著的( $p$  值 = 0.0291), 财产价

值的变异性中只有不超过 10% 的部分是归结于估价员的。因此, 我们可以确定这些估价员对于该县内中等价位的财产具有一定的一致性。

**结论的报告** 县里负责该项研究的官员写给县长的报告中应包含以下内容。

1. 陈述该研究的目标。
2. 描述研究的设计, 该项研究中所评价的财产是如何选取出来的, 以及估价是如何进行的。
3. 讨论该研究所得到的结论对于全县范围内的财产评价的适用性。
4. 数据的数值描述和图表示。
5. 描述所有的推断方法。
  - 叙述研究假设,
  - 表示试验条件的模型,
  - 模型条件的验证,
  - AOV 表, 包括  $p$  值。
6. 结果和结论的讨论。
7. 关于居民就财产评价中的偏差的抱怨, 研究结论的相应的解释。
8. 数据的列表。

## 19.5 小 结

在这一章中, 我们从具有缺失观察值的随机化区组设计的分析入手, 讨论了一些非平衡设计的方差分析, 提出了两种分析方法。在第一种方法中, 我们先估计缺失观察值, 然后按照第十五章中给出的通常的方法进行分析。虽然估计单个缺失值容易做到, 但当有更多的缺失值时, 这个方法变得更加困难。第二种方法中, 通过拟合完全的和简略的模型来得到调整平方和, 此种方法可以用于有一个或多个缺失值的情形。

对于拉丁方设计, 我们又一次指出如何估计单个缺失观察值并继而用通常的分析。然而, 像随机化区组设计那样, 当有多于一个的观察值缺失时, 用拟合完全模型和简略模型的方法更适当些。

最后, 我们考虑了另一类非平衡设计, 即不完全区组设计。我们讨论的特殊的设计有不完全随机化区组设计, 其中并非所有的处理都出现在每一个区组中。这些不完全区组设计具有一定的平衡, 因为各对处理中的双方一起出现在同一区组中的次数, 对于所有的处理对都一样。我们用适当的公式说明了平衡不完全区组设计的分析。BIB 设计的方差分析可以像有缺失观察值的随机化区组设计以及拉丁方设计中那样, 用拟合完全模型和简略模型的方法进行。

### 重要公式

#### 1. 缺失观察值, 随机化区组设计

$$a. \hat{M} = \frac{ty_{i.} + by_{.j} - y_{..}}{(t-1)(b-1)}$$

#### b. 处理平方和的偏差校正

$$\text{Bias} = \frac{\{y_{.j} - (t-1)\hat{M}\}^2}{t(t-1)}$$

校正后的处理平方和为  $\text{SST}_C = \text{SST} - \text{Bias}$ 。

#### 2. 随机化区组设计下的 Fisher LSD

##### a. 对于任何一对没有缺失值的处理

$$\text{LSD} = t_{\alpha/2} \sqrt{\frac{2\text{MSE}}{b}}$$

##### b. 对于有缺失值的那个处理和任何其他处理

$$\text{LSD} = t_{\alpha/2} \sqrt{\text{MSE} \left( \frac{2}{b} + \frac{t}{b(b-1)(t-1)} \right)}$$

#### 3. 随机化区组设计下的两个等式

$$\text{SSB} = \text{TSS} - \text{SST}_{\text{Adj}} - \text{SSE}$$

$$\text{SST} = \text{TSS} - \text{SSB}_{\text{Adj}} - \text{SSE}$$

#### 4. 缺失观察值, 拉丁方设计

$$a. \hat{M} = \frac{t(y_{i..} + y_{.j.} + y_{..k}) - 2y_{..}}{(t-1)(t-2)}$$

#### b. 处理平方和的偏差校正

$$\text{Bias} = \frac{\{y_{.j} - (t-1)\hat{M}\}^2}{t(t-1)}$$

校正后的处理平方和为  $\text{SST}_C = \text{SST} - \text{Bias}$ 。

#### 5. 拉丁方设计下的 Fisher LSD

##### a. 对于任何一对没有缺失值的处理

$$\text{LSD} = t_{\alpha/2} \sqrt{\frac{2\text{MSE}}{t}}$$

##### b. 于有缺失值的那个处理和任何其他处理

$$\text{LSD} = t_{\alpha/2} \sqrt{\text{MSE} \left( \frac{2}{t} + \frac{1}{(t-1)(t-2)} \right)}$$

#### 6. 不完全区组设计的平方和

$$\text{SST}_{\text{Adj}} = \frac{t-1}{nk(k-1)} \sum_i (ky_{i.} - B_{(i)})^2$$

$$SSE = TSS - SSB - SST_{Adj}$$

### 7. 不完全区组设计下处理均值的两两比较

$$\hat{\mu}_i - \hat{\mu}_{i'} = \frac{[ky_{i\cdot} - B_{(i)}] - [ky_{i'\cdot} - B_{(i')}]}{t\lambda}$$

$$LSD = t_{\alpha/2} \sqrt{\frac{2kMSE}{t\lambda}}$$

#### 补充练习

**19.12(医学)** 一位医生要用对于豚草皮肤变态反应检验极度敏感的患者做试验,以比较六种抗组胺药物。为此,这位医生从自己接受的患者中抽取了由 10 位变态反应患者组成的随机样本,并按照下表中给出的试验设计,为这些患者指定了处理(抗组胺药)。每个患者在不同的试验阶段右臂接受了指定的抗组胺药物的注射。在注射以后的指定时间测量出患者臂上围绕注射点的发红的面积。数据由下表给出。

| 患者 |   |    | 处 理 |    |   |    |
|----|---|----|-----|----|---|----|
| 1  | B | 25 | A   | 41 | F | 40 |
| 2  | E | 37 | B   | 46 | A | 42 |
| 3  | C | 45 | D   | 33 | B | 37 |
| 4  | E | 34 | D   | 35 | A | 46 |
| 5  | B | 31 | F   | 42 | D | 34 |
| 6  | C | 56 | E   | 36 | F | 65 |
| 7  | D | 33 | A   | 42 | C | 67 |
| 8  | F | 49 | D   | 37 | E | 30 |
| 9  | C | 59 | A   | 40 | F | 55 |
| 10 | B | 36 | C   | 57 | E | 34 |

a. 指出该设计是什么设计。

b. 说明该设计的特点。

**19.13** 参见练习 19.12。这些数据反应出处理均值之间的差异吗? 取  $\alpha = 0.05$ 。

**19.14** 参见练习 19.13。用最小显著差异准则确定处理差异, 取  $\alpha = 0.05$ 。

**19.15** 用一个计算机程序做与练习 19.13 中相同的分析, 比较这两个不同的结果。

**19.16** 参见练习 19.15。用一个计算机程序进行方差分析。你所得到的结果与例题中的结果一致吗?

**19.17** 说明在一个平衡不完全设计中, 你如何检验区组效应的显著性。

**19.18(心理学)** 一家公司的市场调查部为了考察公众对一种新电视游戏卡的介绍的反应, 在四个地区分别选取了三家不同的连锁店, 并比较了它们的周销售额(以下美元为单位)。

| 地区 |    | 连锁店 |    |    |
|----|----|-----|----|----|
|    |    | 1   | 2  | 3  |
| N  | W1 | 35  | 17 | 7  |
|    | W2 | 30  | 22 | 12 |
| S  | W1 | 42  | 30 | 22 |
|    | W2 | 48  | 28 | 19 |
| E  | W1 | 35  | 35 | 15 |
|    | W2 | 38  | 40 | 20 |
| W  | W1 | 22  | 43 | 28 |
|    | W2 | 26  | 48 | 23 |

- 写出一个适当的模型(其中包括周的效应), 以及方差分析表中的变异来源。
- 如果我们分析两周的总销售额的数据, 你的模型会有什么变化?
- 取  $\alpha=0.05$ , 利用第十五章的公式, 对两周的销售数据作方差分析。

**19.19** 参见练习 19.18。对连锁店的平均值用 Tukey 方法来比较几个不同的地区, 取  $\alpha=0.05$ 。

**19.20** 参见练习 19.18。假定由于后勤的问题, 使得对该产品的介绍在北部和东部的连锁店 1 推迟了一周, 从而, 北部和东部的连锁店 1 在第一周的数据不可用。

- 写出一个适当的模型。
- 提出一个用手头备有的软件分析此中数据的方法。
- 对于在(b)中所提出的方法, 写出其模型。

**19.21(人力资源)** 一家国外的汽车制造商正在花费数百万美元在美国建设一个很大的制造厂(一个屋顶下覆盖大约 70 英亩)。他们的目标之一是雇佣美国的工人在美国生产高质量的汽车。这项庞大的计划中的一部分是送新工人中的大约 20% 去该公司的母公司所在的国家接受附加的训练。这种附加训练的价值的衡量方法之一是看有 20% 的工人接受过这种训练的装配线上的产品是否变得更好了。下面是六条装配线上的数据(其中有三条装配线上有接受过训练的工人)。两名不同的检验员都各自对从每条装配线上随机选取的两辆车进行了检查。用这

些数据回答下列问题。

| 附加训练 |     |   | 无附加训练 |     |    |
|------|-----|---|-------|-----|----|
| 装配线  | 检验员 |   | 装配线   | 检验员 |    |
|      | 1   | 2 |       | 1   | 2  |
| 1    | 6   | 6 | 4     | 8   | 7  |
|      | 3   | 4 |       | 5   | 5  |
| 2    | 4   | 3 | 5     | 10  | 9  |
|      | 2   | 2 |       | 4   | 4  |
| 3    | 2   | 3 | 6     | 15  | 13 |
|      | 1   | 1 |       | 7   | 6  |

- 建议一个适当的因变量。
- 对于该试验写出一个适当的模型,并解释其中各项的意义。
- 写出方差分析表中的变异来源和各平方和的自由度。

19.22 参见练习 19.21。

- 建议一个分析这些数据的工具。
- 用下面给出的输出结果得出结论。
- 你能给出用来解释这些数据的图吗?

General Linear Models Procedure; MODEL FOR EXERCISE 19.22

Dependent Variable: DEFECTS

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 11 | 190.83333333   | 1.98    | 0.1275 |
| Error           | 12 | 105.00000000   |         |        |
| Corrected Total | 23 | 295.83333333   |         |        |

| Source                | DF | Type III SS  | F Value | Pr > F |
|-----------------------|----|--------------|---------|--------|
| INSPECT               | 1  | 0.66666667   | 0.08    | 0.7872 |
| TRAIN                 | 1  | 130.66666667 | 14.93   | 0.0023 |
| INSPECT * TRAIN       | 1  | 1.50000000   | 0.17    | 0.6861 |
| LINE(TRAIN)           | 4  | 56.66666667  | 1.62    | 0.2329 |
| LINE * INSPECT(TRAIN) | 4  | 1.33333333   | 0.04    | 0.9968 |

19.23 参见练习 19.21。假设 2 号检验员未能评估从 4 号装配线抽取的第二辆车,1 号检验员漏检了从 3 号装配线抽取的第一辆车。



- a. 模型有变化吗？建议一个分析这些数据的方法。
- b. 用下面的计算机输出结果给出结论。

General Linear Models Procedure: FULL MODEL FOR EXERCISE 19.23

Dependent Variable: DEFECTS

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 11 | 180.81818182   | 1.60    | 0.2326 |
| Error           | 10 | 102.50000000   |         |        |
| Corrected Total | 21 | 283.31818182   |         |        |

| Source                | DF | Type III SS  | F Value | Pr > F |
|-----------------------|----|--------------|---------|--------|
| INSPECT               | 1  | 0.03571429   | 0.00    | 0.9541 |
| TRAIN                 | 1  | 124.32142857 | 12.13   | 0.0059 |
| INSPECT * TRAIN       | 1  | 0.89285714   | 0.09    | 0.7739 |
| LINE(TRAIN)           | 4  | 48.75000000  | 1.19    | 0.3733 |
| LINE * INSPECT(TRAIN) | 4  | 2.62500000   | 0.06    | 0.9913 |

General Linear Models Procedure: REDUCED MODEL I FOR EXERCISE 19.23

Dependent Variable: DEFECTS

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 11 | 180.81818182   | 1.60    | 0.2326 |
| Error           | 10 | 102.50000000   |         |        |
| Corrected Total | 21 | 283.31818182   |         |        |

| Source                | DF | Type III SS | F Value | Pr > F |
|-----------------------|----|-------------|---------|--------|
| INSPECT               | 1  | 0.03571429  | 0.00    | 0.9541 |
| INSPECT * TRAIN       | 1  | 0.89285714  | 0.09    | 0.7739 |
| LINE(TRAIN)           | 4  | 48.75000000 | 1.19    | 0.3733 |
| LINE * INSPECT(TRAIN) | 4  | 2.62500000  | 0.06    | 0.9913 |

General Linear Models Procedure: REDUCED MODEL II FOR EXERCISE 19.23

Dependent Variable: DEFECTS

| Source          | DF | Sum of Squares | F Value | Pr > F |
|-----------------|----|----------------|---------|--------|
| Model           | 11 | 180.81818182   | 1.60    | 0.2326 |
| Error           | 10 | 102.50000000   |         |        |
| Corrected Total | 21 | 283.31818182   |         |        |

---

| Source              | DF | Type III SS  | F Value | Pr>F   |
|---------------------|----|--------------|---------|--------|
| INSPECT             | 1  | 0.03571429   | 0.00    | 0.9541 |
| TRAIN               | 1  | 124.32142857 | 12.13   | 0.0059 |
| INSPECT*TRAIN       | 1  | 0.89285714   | 0.09    | 0.7739 |
| LINE*INSPECT(TRAIN) | 8  | 52.76666667  | 0.64    | 0.7279 |

## 第二十章 分析结果的传达和备案

- 20.1 引言
- 20.2 做好传达沟通工作所面临的困难
- 20.3 传达的障碍:图形的歪曲
- 20.4 传达的障碍:有偏抽样
- 20.5 传达的障碍:样本容量
- 20.6 为统计分析准备数据
- 20.7 统计分析的指导原则和报告
- 20.8 文档和结果的保存
- 20.9 小结

### 20.1 引言

在第一章我们曾指出,统计学是从数据中学习的科学。从数据中学习的四个步骤形成了本书的主线,这四个步骤即(1)数据收集的设计,(2)为分析和概括数据而做的准备,(3)数据的分析,以及(4)结论的报告。我们在本书中,尤其是在各个案例研究中,按照这四个步骤进行分析,从而突出了这些步骤。

在前面各章中,我们讨论了多种统计方法,如何把这些方法应用于特定的数据集合,以及如何解释在从计算机输出结果中所得到的发现。我们没有集中讨论在接收到数据以后到为分析准备好计算机可读的数据文件为止这一过程中的数据处理的各个步骤,也没有讨论概括统计分析结果的报告的形式和内容。在这一章,我们来考虑这些数据处理的步骤以及报告的写法。本章不是一个包含各种所用的工具的手册,而是概要谈一谈管理人员或研究工作者所应该知晓的若干步骤。作为一个例子,本章讨论已经高度正规化了的制药业中的做法。其他工业和组织中的做法多少会有些不同。我们首先介绍在要进行有效地传达和沟通时所可能遇到的困难以及经常出现的不足。

### 20.2 做好传达沟通工作所面临的困难

在本书中我们一直在讨论如何理解数据。在理解数据的过程中,其最后一步是结果的传达和沟通。你如何传达一项研究或调查的结果呢?可能的方法,包括

各种口头的和书面的,数不胜数。有很多种口头的和书面的传达方法。例如,在一家公司或组织内部,书面的传达方法从非正式的简短的便条或记录到正式的项目报告,多种多样(图 20.1)。

要用简练的、毫不含糊的语言来传达一项统计分析的结果是很困难的。事实上,大多数事物的描述都是困难的。例如,不妨试一试精确地描述坐在你旁边一个人,使得一个陌生人能够从具有类似体貌特征的一群人中把这个人找出来。这不是一件容易的事情。指纹、声波纹和照片——所有图形描述——都是识别人的最精确的方法。一组测量值的描述也是一项很困难的工作。然而,就像对人的描述一样,通过图形或图片的方式更容易做到。

岩洞壁画传达给我们许多有关史前人类生活的信息。类似地,从绘画和雕刻中我们也得到大量的关于古代巴比伦人、埃及人、希腊人和罗马人的生活和文化的信息。艺术已经被用来研究各个时代各种各样的生活方式、历史和文化的概貌。毫不奇怪,附有文字说明的图形和图表可以帮助我们理解和传达统计分析的意义。

| 口头传达  |      |
|-------|------|
| 非正式谈话 | 正式提呈 |

| 书面传达            |            |
|-----------------|------------|
| 内部(如公司内部,大学内部等) | 外部         |
| 备忘录             | 信函(或给编辑的信) |
| 正式项目报告          | 科学期刊论文     |

图 20.1 书面的和口头的传达方式

在读统计分析的结果或传达我们自己的统计分析结果时,我们必须小心,以免由于我们列出数据或给出这些结果的方式而发生误解。我们经常听到这样的说法,“用统计很容易撒谎。”这种说法并不新鲜。英国著名政治家 Disraeli 就曾说过,“有三种谎言:谎言,令人憎恶的谎言和统计。”问题出在哪儿呢?

首先,对于真相的歪曲只可能发生在我们进行传达的时候。我们可以利用图

形、图表、声音、气味、味道、语汇、数字等任何其他工具来表现我们的感觉,从而实现传达;在使用其中的任何方法或方法的组合进行传达时都可能发生歪曲。

如此看来,我们所作出的陈述有可能使他人产生误解,因为我们可能在进行数据收集过程的解释或所做分析的解釋时省略了某些东西。比如,我们可能由于不经意而没有解释清楚某个数值陈述的意义,或我们可能漏掉了解释清楚分析结果所必须的某些背景信息。即便是一个正确的陈述,也可能由于读者缺乏基本的统计知识而被误解。比如,95%的置信区间这样一个清楚的推断表达,对于一个从来没有接触过统计概念的人来说是毫无意义的。

现在,我们来看一些潜在的影响传达效果的障碍。我们在给出统计分析的结果或试图解释别人提供给我们的结果时,必须仔细考虑这些障碍。

### 20.3 传达的障碍:图形的歪曲

图片很容易歪曲真相。在很多产品,包括软饮料、啤酒、化妆品、衣服、汽车等的推销中,都用有吸引力的、年轻的模特。这种推销留给我们印象并不淡薄,这就是:用了这些产品,我们也会变得看起来像这些模特的样子。你可曾从这些广告的某个中醒来,想一想这些广告中的信息何以联系于产品的质量或用途?你可曾想到你是如何被广告误导的吗?推销商品时性感的使用非常突出,而我们也似乎接受了这种歪曲。下面是一份关于啤酒广告的文章,刊载于 *USA Today*, 2000 年 3 月 15 日,它说明了我们是如何被这种类型的广告所“操纵”的。

---

#### 性意味内容在啤酒广告中卷土重来 “淫秽”电视节目导致更自由的促销方式

Michael McCarthy  
*USA TODAY*

纽约——所谓“啤酒加姑娘”广告中的“姑娘”又回来了。

大约十年以前,在 Anita Hill 论战中引起的举国关于性骚扰问题的反省迫使推销商们把传统的、被视为有性别歧视色彩的内容,如“瑞典比基尼组”从广告中清除了。

现在,啤酒制造商们,包括 Miller Brewing, Anheuser-Busch 以及 Heineken,都把性感的内容又拣回来,放进了电视广告中。

“我们正处在一个新的十年,大家的口味也正在改变”,营销部顾问 Laura Ries

说。

随着电视愈加“淫秽”,啤酒广告反映出了“更加自由的环境”,她说,“曾几何时,用狗、猫和企鹅来做广告,如今,有性感内容的推销又复活了,它抢夺了人们的注意力。”

另一些人仍旧把这些推销视为性别歧视的叫卖。“我们绝对看到了一种趋势”,全国妇女组织(NOW, National Organization for Women)媒体监督委员会主席 Sonia Ossorio 说。她还警告说这些新的勾当更加阴险狡诈,因为他们采取的不是淫荡的方式,而是诙谐的方式。

最近的广告说明了这样一种趋势。

**Miller 公司** 该公司将在本周开始的 NCAA 篮球锦标赛上展开 Miller Lite 啤酒的推销活动。他们的广告中又重提他们的传统口号:“现在是 Miller 时间”,并重播过去的影像。在其中一个广告中,三个小伙子雨天来到一家荒废的酒吧,他们使劲地擦着自己的眼睛,但却完全忽略了正要擦干身体而未来得及躲藏的、赤身裸体的、湿漉漉的两个大美人。

在另一个广告中,一个不走运的人找到了一份专职司机的工作,这使得他在真实的奖赏面前却无动于衷;此时,三个美丽的留学生正在寻找寄宿之处,而他的伙伴们却露出嫉妒的神情。

“这些广告是以幽默的形式作成的,”Miller 公司销售部高级副总裁 Bob Mikulay 说,它们是对小伙子们在没有啤酒喝时会做什么的联想。

对 Miller Lite 啤酒要做一次牵引启动,它在美国市场所占的份额从十年前的 10% 下降到了现在的 8%,*Beer Marketer's Insights* 杂志执行主编 Eric Shepard 如是说。

**Heineken 公司** 该公司“啤酒关乎一切”的展买活动明显地使用了性别比喻。

在一个叫做“不成熟的斟酒”的黄金电视节目的广告中,一位小伙子在酒吧内瞥见对面有一位漂亮的女性。她以优雅的、诱人的方式把 Heineken 啤酒倒进她的高脚杯子。当小伙子试图同样做时,他动作潦草,灌得太快,弄得满地白沫。

在一个叫做“错误的酒吧车厢”的广告中,一个乘坐火车的家伙正坐在酒吧车厢喝着一种别的啤酒,酒吧车厢内坐满了喝着大路啤酒的各色人等。透过窗户,他看到相邻轨道的火车上一位美丽的金发碧眼的女性坐在酒吧车厢里,车厢里正在举行聚会,大家喝着 Heineken 啤酒。令他懊恼的是,当他的火车开动时,那位女郎向他挥了挥手。

**Anheuser-Busch 公司** 你不必成为 Fellini 的影迷就能在其最近为 Michelob Light 啤酒所做的广告“火车上的陌生人”中得到其象征主义的信息。

一名妇女走进了一节火车车厢,看到两个面对面坐着的小伙子,就坐在了其中一个人的旁边。火车走进了隧道。当火车钻出隧道的时候,这名妇女已经更换了

座位,搂抱着正在喝着 Michelob Light 啤酒的那个家伙。他得意洋洋地冲着他对面的人咧嘴笑着。

Ries 认为,这种新的广告已经带来了“啤酒面包”的广为流传。“这个广告显示了男女机会的平等,片中妇女处于主动地位,而那两个男性只是这个笑话中的笑柄而已,”她说。

Cheryl Berman, Leo Burnett 的创意主管并不认为这是老式啤酒广告的卷土重来。“这些广告只不过是多了男人的一些笑料,”她说。

她还认为,这个广告是对近年来显得无聊的啤酒推销方法的改进。这些老的方法包括 Miller 的虚构广告人“Dick”。

“有很多种奖励秀广告,”她说,“它们确实在(工业)创意秀中占据了很高的地位,但公众并不买它们的账。”

在 Miller Lite 啤酒广告机构 Ogilvy & Mather 工作的 Rick Bokyo 说,展销活动也向妇女朋友展示了快乐的“Miller 时间”。“它起缘于今天的消费者,我有一个 21 岁的女儿,她常说‘现在是 Miller 时间’。”

但是,性别歧视啤酒广告的广告人掌握其中问题的第一手材料。

广告机构代理商 Patrick Scullin 在 1991 年电视上播出他为 Old Milwaukee 啤酒所制作的并不著名的广告“瑞典比基尼组”以后不久,就丢了他的工作甚至于差点儿丢了他的前程。他所在的公司中的一组女工控告说,这个促销广告帮助制造了一种鼓励性骚扰的氛围。之后,公司草草收场了事。

“我的邮政编码变成了西伯利亚”,现在任亚特兰大 Ames Scullin O’Haire 公司创意主管的 Scullin 说,“他们旨在蒙骗,但他们却进入了媒体行业。”

Scullin 不认为这些新广告中有很大的改变。“只不过把片中穿比基尼的妇女换成了穿紧身牛仔裤的妇女,”他说,“人们不喜欢承认这一点。但只要有生物,啤酒广告中就会有性别歧视。”

---

电视、互联网以及产品的介绍图片通常比实物更具有吸引力,但我们一般默认这种歪曲。

统计的图形有第三章中介绍的直方图、频率多边形、饼图和条形图。这些数值结果的描绘或演示很难与可爱的女孩或英俊的小伙儿的素描联系起来,因而,也与这种最常见的图片歪曲无关。但是,可能出现其他形式的歪曲。你可以压缩或拉伸坐标轴而歪曲实际结果。由此造成误解的原因是,人们通常认为平缓的斜线表示小的增加,而陡峭的斜线意味着大的增加。

例如,假定某年的前 6 个月的消费者价格指数的值分别为 160, 165, 178, 189, 196 和 210。我们可以用图 20.2 中的频率多边形来显示这个消费者价格指数指标向上走的趋势。在这张图中,这个指数的上升是很明显的,但看起来上升的幅度却

不大。另一方面,我们可以用一种很不相同的方式来给出这组样本数据,如图 20.3。在该图中,纵轴被拉伸,已经不包含 0 点。注意,图中陡峭的斜线给人以大幅上升的印象。另外一种达到同样效果的方式——降低或增加陡峭程度——是拉伸或压缩横轴。

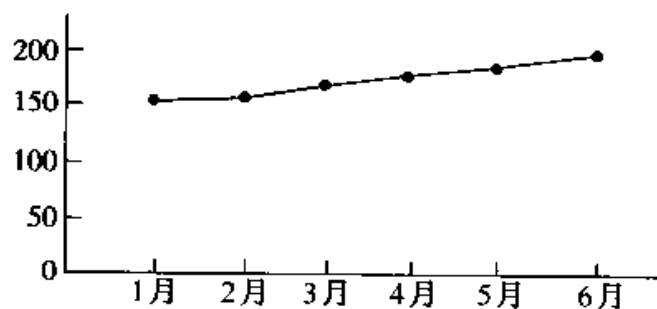


图 20.2 消费者价格指数的变化

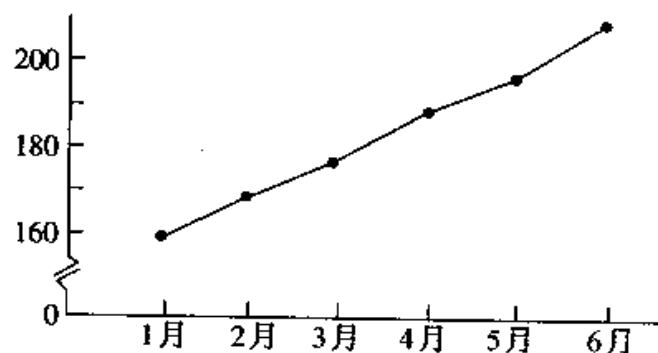


图 20.3 消费者价格指数的变化

当我们用条形图、直方图、频率多边形或其他图形来表示数据的时候,必须注意,不要压缩或拉伸坐标轴,因为这样做会给大多数读者造成麻烦。响应变量的增加或减少是大还是小,依赖于对于这些变化的观察者来说所具有的重要性,而不是在图形显示中所看到的斜度。在实际当中,多数人只看图形中的斜度。

## 20.4 传达的障碍:有偏抽样

最常见的由于统计造成的歪曲之一发生在实验者在不知情(有时是知道的)的情况下,从不适当的总体中抽取了样本。也就是说,他或她从不是真正要研究的总体的测量值集合中抽取了样本。

例如,假定我们要看看纳税人对某个儿童公园和娱乐中心的意见。挑选了由



若干家庭组成的随机样本,面访者被送到这些家庭。不幸的是,样本中 40% 的家庭中无人,于是,我们又在市内的其他家庭中随机抽取若干户代替这些无人的家庭,以弥补上述的不足。最后所得到的样本就是从不适当的总体中抽取的样本,因而我们说该样本是有偏的。

在上述的家庭调查中,要研究的总体是来自这个城市的所有家庭的意见的全体。而抽样时却不是这样,仅仅是从一个小得多的总体,或者说所有家庭意见集合的一个子集——抽取样本时恰好在家的户主的意见集合。有可能在这两个集合中喜爱这个公园的户主的比例是相等的,此时,把抽样限制在在家的户主上不造成损失。但更可能的情况是,在家的那些人小孩,在这批人中喜欢这个公园的比例比整个城市更高一些。这样,我们就有了一个有偏的样本,因为得到的结果更能反映有小孩的家庭的意见。也许,看待这个问题的一种更好的方式是注意到我们无意仅从所要研究的总体的一个子集中抽取样本。

利用邮寄问卷调查的方法往往会导致有偏的样本。在某种意义上讲,调查者往往使挑选和抽样单位的数量依赖于接收这些调查的人的兴趣,可以利用的时间,以及各种其他的个人特征。非常忙的人和能量大的人可能随手把这些问卷扔进废纸篓;你也很少能从那些不感兴趣或专注于其他活动的不活跃的人那里听到什么意见。最常见的情况是,回答者是些活跃分子——那些非常赞成的人,那些反对的人或者那些从这项调查的某种结果中能得到点什么的人。

虽然众多的媒体评论员和分析家都把选举的结果作为对重要问题的公众意见的一种表示,但众所周知,投票的结果代表了公众意见的一个有偏的样本。投票的人仅代表了有资格投票人的一少部分,远远不到一半;他们是那些希望体验自己作为居民的权利和责任的人,或者是受到某种特别启发而来参加的人。所导致的投票人的子集对于所要研究的总体和整个国家有资格投票的选民的意见来说,不具有代表性。

当人们试图把从一个总体所得到的试验结果外推到另一个总体时,也会发生从不适当的总体中抽样的情况。关于各种产品(例如糖精)在鼯鼠、老鼠、小猎犬的乳腺等引发癌症的效果的试验结果已经被大量地发表了。这些结果常常被用来暗指人在经常地或大量地暴露于这些产品后,也会有患癌症的高风险。这样的推断并不总是可信的,因为这些试验结果不是从人的试验中得到的。很可能人比老鼠有高得多的抵抗力,或许由于某种原因,人会有完全的抵抗能力。在小型哺乳动物身上癌症的药物诱发确实表明需要人们注意和小心,但并不证明这种药物对人一定有害。注意,我们不是在批评在各种动物上的试验,因为对于一些可能对人体有毒害的药物,这常是获得信息的惟一的方法。我们只是指出,实验者知道,他们其实是从一个仅仅与要研究的总体相类似(很可能是并不太像)的总体中抽样的。

工程师也检验“老鼠”以代替“人”。这里所说的老鼠,是指微缩模型或一个新

的工程系统的试验工厂。在这些模型上做试验所产生的结果偶尔也会与大的、真实的系统所产生的结果有重大的差异。因而我们再一次看到了从不适当的总体中抽样的问题,但由于经济原因这是工程师所能做到的最好的试验。通常没有基金可以用来在生产之前进行大量的等规模模型的试验。

还有很多有偏抽样或从不适当的总体中抽样的例子。关键是当我们传达一项研究或调查的结果时,我们必须清楚样本是如何抽取的,以及样本是否是从要研究的总体中~~随机~~抽取的。如果在已经公开发表的研究或试验的结果中没有给出这方面的信息,那么读者当对这些推断有所保留。

## 20.5 传达的障碍:样本容量

误解也可能发生在忽略了样本容量的讨论的情形。例如,假定你读到一份报告,说一个样本中大约 75% 的人喜欢一幢新建的摩天大楼。进一步的调查表明,调查人员只抽取了 4 个人。当 4 个人中有 3 个人喜欢这个项目时,调查人员决定停止这项调查。当然,我们在此例中有些夸张,但我们基于容量为 25 的样本也得到这种不具说服力的结果,尽管许多主顾认为这个样本容量已经足够大了。众所周知,在从二项分布中抽样时,要得到充分的信息需要很大的样本量。

有幸的是,现在许多发表的报告都提供了更多关于样本容量以及公众调查是如何进行的等方面的信息。十年以前,很少有报告说明样本中包含多少人,说明样本的抽取方法的报告更是凤毛麟角。现在情况不同了。事实上,媒体有时在试图解释清楚调查是如何进行的这一方面走得太远了。一个例子是下面发表在杂志 *Wall Street Journal* (华尔街杂志) 上的文章。除了样本中所包含的人数以及近似加或减(置信区间)外,我们中有多少人能理解得更多呢? 要解释清楚已经做了些什么,需要一个有良好的统计学和抽样调查训练的人。这个故事的教训同样是:试图用不含糊的术语来沟通。

### 公众意见调查是如何进行的

这项华尔街杂志/美国全国广播公司新闻调查,自上星期五至星期一,在全国范围内电话采访了 18 岁或 18 岁以上的成年人共 4 159 人。有 2 630 人回答了问题。

样本是从电话交换台的完整列表中抽取的,这些列表的选取使得全国范围内的每个地区按照其人口的比例有其代表。家庭则是这样挑选的:给所有的电话号码——包括列出来的和未列出来的——一个被选中的概率。调查的结果通过加权对样本中由于教育、年龄、种族、性别和宗教不同造成的变异进行了调整。

如果对所有美国的成年人用同样的问题进行调查,那么所得到的结果与本次调查的结果相差不超过上下两个百分点的机会是二十分之十九。如果在更小的组内调查,犯错误的概率会更大。

## 20.6 为统计分析准备数据

我们先来讨论一项研究中的数据处理所包含的步骤。在实际中,这些步骤所占的工作量,会达到从收到原始数据到提交分析结果为止全部工作量的 75%。这些步骤是什么呢?为什么它们这么重要,这么耗时呢?

为了回答这些问题,我们列出主要的数据处理步骤,这些步骤始于数据的接收,终于数据统计分析的开始。然后,我们分别讨论每个步骤。

### 定义 20.1 为分析准备数据的步骤

1. 接收原始数据资源。
2. 从原始数据资源建立数据库。
3. 编辑数据库。
4. 校正和澄清原始数据资源。
5. 最后确定数据库。
6. 从数据库建立数据文件。

**1. 接收原始数据资源** 对于将要进行概括和分析的每项研究,数据以某种形式到达我们手中,我们称这种形式为原始数据资源。对于一项临床试验,原始数据资源通常是用 8.5 英寸宽 11 英寸长的纸作成的病例报告表、报告单,这些纸用来记录进入到研究中的患者的数据。对于其他类型的研究,原始数据资源则可能是用做实验记录的记录纸,磁带(或任何其他形式的机读数据),手制表格,等等。

保留原始数据资源是很重要的,因为它是“数据轨迹”的开始,这里所说的数据轨迹,是指从原始数据得到研究的结论的过程。在许多包含分析和概括的不同研究的咨询中,都要保存一份记录单,其中记有研究中关键的信息和原始数据资料。研究记录单中所包含的一般信息如下。

### 定义 20.2 研究数据记录单

1. 收到的数据,数据从何处来。
2. 研究的负责人。
3. 指定的和其他的统计工作者。
4. 研究的简短描述。

5. 研究过的处理(复合,配制等等)。
6. 原始数据资源。
7. 测量到的响应。
8. 研究的基准编号。
9. 估计到的(或实际的)完成日期。
10. 其他相关信息。

之后,当进行完研究的分析,传达完研究的结果时,可以把诸如研究结果是如何传达的,这些结果记录在何处,保存了什么数据文件以及这些文件保存在什么地方等信息添加在记录单上。

**2. 从原始数据资源建立数据库** 对于计划要做统计分析的多数研究来说,都要做一个机器可读的数据库。建立这个数据库的步骤和形成的数据库的最后形式多种多样,要看统计分析所使用的软件系统。然而,我们可以基于输入系统的形式给出几条指导原则。

当数据是在终端从键盘输入的时候,首先检查原始数据的易读性。任何不清晰的数值、字母或其他问题都要提请研究同仁的注意。然后,应订立一套编码规则,用来为数据指定列数和变量名字。此时,还应为缺失数据(例如,那些不能利用的数据)指定编码。此外,对于每一个变量,给出一个简单描述也是有益的。从终端键入的这个数据文件叫做**机器可读数据库**。应该打印一份数据库内容的清单并对照原始数据进行仔细检查,从终端校正发现的任何错误,然后再校验新得到的清单。

有时,数据收到时就是机器可读的形式。在这种情况下,磁带或磁盘文件就被视为数据库。然而,为了读出这个数据库,你必须有一组编码规则。用这组编码规则,得到数据库内容的清单,然后,仔细检查一下,看看是否所有数值和字母看起来都是合理的,以及创建这个数据文件所用的格式是否合适。其中出现的任何问题的继续下一步之前都必须解决。

某些数据集合太小,不必要从原始数据建立一个机器可读的数据文件。此时,计算可以手工进行,或者把数据输入电子计算器。在此情况下,检查任何计算,看看是否有错。不要相信你看到的一切,重新做一次计算不是一个坏主意。

**3. 编辑数据库** 编辑的形式和编辑过程的完整性高度依赖于研究的类型和在分析之前你要求达到什么样的精度和数据的完整性。例如,使用 SAS 的时候,最好检查一下每个变量的最小值、最大值和频率分布,以确保没有看上去不合理的信息。

还应该做些其他的检查。画出数据的图,看看是否有问题;按照数据的结构做某种**逻辑检查**。例如,如果数据是在对患者几次不同的访问期间记录下来的,那

么,在第二次访问时记录的数据就不会早于第一次访问时的数据;类似地,如果一个患者在第二次访问后就失去跟踪了,那么,我们就不可能有这个患者的第三次以及以后的访问的数据。

对于小的数据集,我们可以手工编辑这些数据,但是对于大的数据集,这项工作可能耗时良多、单调乏味。如果需要机器编辑,就要找一个可以允许用户进行特定的数据编辑的软件系统。即便如此,对于更复杂的编辑和逻辑检查,为了用机器编辑数据,可能须要编写一个适合于自己使用的程序。这种编程工作可能是一件纷繁耗时的事,在接到数据之前应预先做好计划。

**4. 校正和澄清原始数据资源** 从接收到原始数据时起,到传达统计分析结果为止,常常会出现关于原始数据的可识别和准确性问题。我们发现,为了确定研究的数据轨迹,保留一份这些问题或出入的清单是有益的。如果需要对原始数据资源进行一项修正或澄清,这种修正或澄清应该在这个清单上记录下来,然后再对原始数据资料进行相应的修改。如果不需要修正,也应该在清单上注明。记住,对于原始数据资源所作出的任何改变,机器可读数据库中也应该进行相应的改变。

**5. 最后确定数据库** 可能有人一直告诉你说,一项研究的所有数据是一次到齐的。当然并非总是如此。例如,在一项市场调查中,可能对不同的地方在不同的时间进行调查,因而,负责数据处理的人并不是一次收到所有数据。每个数据子集必须经过创建、编辑和校对数据库这样一个数据处理的过程,直至最后该项调查完成并把这些数据进行编辑,写进数据库。此时,应该对数据库再进行一次检查,在分析数据之前做一次最后的校对。这是由于,对于大型的数据集,分析和概括要花费相当的人力和计算机时间。与其到后来改变数据库并重复所有分析,还不如慎重地从最后的数据库进行分析。

**6. 从数据库建立数据文件** 一般说来,要从机器可读数据库创建一两组数据文件。第一组,称为**原始文件**,它们反映数据库的基本结构。对照数据库的清单,检查这些文件的清单,以确认数据中的变量是否按照正确的形式被读入,以及缺失数据的代码是否已经保留下来。对于某些研究,原始文件实际上用来编辑数据库。

第二组数据文件称为**工作文件**,可以从原始文件中产生。工作文件是指定为分析之用的。它们可能是从原始文件进行重新构造得来的,也可能是从中挑选出来的一个重要变量,或者在原始文件中用插入、计算或变换的方法产生或增加了一个新的变量。这些工作文件的清单应与原始文件的清单相对照,以确保所做的重新构造或变量挑选是对的。对于计算或变换得来的变量,用手工计算来进行核对,以验证程序编码的正确无误。

如果原始数据文件和工作文件都是 SAS 数据集,你应该利用 SAS 提供的文件管理特性。当创建 SAS 数据集时,应当为该数据集指定一个至多包含 40 个字符的描述标识(descriptive label)。这个标识可以与数据集存储在一起,当调用名为

contents 的例程来显示数据集合的内容时,这个标识可显示出来。可以给所有变量起一个描述性名字,最长 8 个字符,这些名字应具有与项目中有关变量相应的含义。还有,最长 40 个字符的变量标识可以用来提供附加的信息。SAS 编码中可以加上题目以识别各个项目和描述各项工作。对每一文件可保留一份清单(用 proc print)和分析字典(用 proc contents)。

对于用其他软件包创建文件的情况,你应该使用该计算机程序中可用的标识和文件管理特性。

即使对数据运用了适当的统计方法,从研究中得出的结论也不能僭越分析所使用的数据,因此你得做出判断。在分析之前,你在这些数据处理的零碎工作上所花费的时间总量,真切地依赖于研究的特性、原始数据资源的质量及你对数据的精度和完全性所要达到的信度。

## 20.7 统计分析的指导原则和报告

在这一节中,我们简单讨论进行统计分析的几个指导原则,并列出用来传达结果的统计报告的一些重要内容。一个大研究项目的统计分析,通常可以分为三种类型的分析:(1)初步分析;基本分析;(3)备用分析。

**初步分析**,通常用描述性统计方法或图形的方法。通过初步分析,统计工作者熟悉所面对的数据,并为所有后面的分析打下基础。这样的分析可以包含频率分布,直方图,描述统计量,对于各处理组的可比较性的考察,相关性,一个变量和两个变量的图。

**基本分析**则瞄准研究和分析的目标,从中得到结论。**备用分析**包括用其他的方法证实从基本分析所得到的结论,也可以包括与标准的方法比起来不容易被接受的新的统计方法。分析遵循以下几条原则。

### 定义 20.3 初步分析,基本分析和备用分析

1. 用已经经过大量检验的软件实施分析
2. 在计算机输出结果上进行标记,说明分析的是什么研究项目,分析中讨论的是什么对象(动物,患者等等),以及所喜好的分析的简短描述。例如,在 SAS 软件中, TITLE 语句就很有用。
3. 在输出中使用变量标识和值标识(如, 0 = 无, 1 = 中等)。
4. 提供在每一个分析中所使用的数据的清单。
5. 对于所有的分析,仔细检查输出结果。计算机成功运行完成了这项工作吗? 样本容量,均值和自由度对吗? 也许还需要进行其他的检查。
6. 保存所有初步分析,基本分析和备用分析的结果,这些结果为得到研究结论提供了基本的信息。

在完成统计分析以后,必须从中抽出结论,并且把结果传达给关心这项研究结论的人。有时,需要用正式的书面统计报告的方式来传达统计分析的结果。我们认为,按如下的大纲写出的统计分析报告是有用的和充满信息的。

**定义 20.4 统计报告的一般大纲**

1. 总结。
2. 引言。
3. 试验设计和研究方法。
4. 描述统计量。
5. 统计方法。
6. 结果和结论。
7. 讨论。
8. 数据清单。

## 20.8 文档和结果的保存

数据的处理、分析和总结这个过程的最后部分是建立文档和保存结果。对于由别人来进行仔细审查的正式的统计分析,非常重要的一点是,提供对所有的数据处理和统计分析的详细文档,从而使得数据轨迹是清晰的,数据库或工作文件是容易读的。由此,审查者可以按照所做过的一切重做一遍,或者将这个分析进行扩展。文档和存储文件中的内容,依你工作的特定环境而定。一般的文档存储文件的内容如下。

**定义 20.5 研究文档和存储文件**

1. 统计报告。
2. 研究描述。
3. 随机编码(用来把研究对象指定到各处理组)。
4. 重要的信件。
5. 文件创建信息。
6. 初步、基本和备用分析。
7. 原始数据资源。
8. 数据管理单,其中包括日志,以及有关数据文件储存的信息。

建立文档和储存文件的主要动机在于,为了我们自己或其他人的使用而提供一个清楚的数据和分析“试验”,因为使用时需要重新访问这些数据。在任何特定的情况下,首先想一想这些文档是否必要,如果是的话,需要多么详细。检验你的

文档的完整性和明白易懂的程度的一个好的方法是,要求一位对你的项目不熟悉,但却懂得你的研究领域的同事重新构造甚至重做一遍你所做过的基本分析。如果他或她可以按照你的文档从头做到尾,那么你就做的不错。

## 20.9 小 结

在这一章中,我们讨论了如何提呈数据的统计分析结果,以及把这些结果有效地传达给有关的人员时所可能出现的一些问题。提呈统计分析结果以及进行有效地传达不是一件轻而易举的事情。挡在有效地传达统计分析结果的路上的障碍包括图形的歪曲,有偏抽样和忽略对于样本容量和抽样方法的讨论。在对这些障碍有了一定的认识之后,我们可以更好地批评和理解面向我们的传达,同时也可以在我们的分析传达给别人时做得更好。

在本章的最后一个专题中,我们讨论了文档和储存文件。当完成了你的分析,得到了你的结论并把这些结果传达给有关的人以后,恨不能推迟甚至放弃文档备案和储存文件。然而,值得花一些时间来想一想将来重温你的分析的可能性,并确定为了对这一重温的过程(你或他人)有所助益你应该怎么办。

最后,我们应该把我们的统计分析放在我们正在考虑的实际问题当中。统计分析的报告未必就是一个重要问题的答案,它可能只是答案的一部分。例如,我们可能揭示出片剂比胶囊释放药物来得快些,但在决定是在市场上投放片剂还是胶囊时,这不是惟一的考虑。诸如成本,可口性以及稳定性也是需要考虑的。进行这些考虑时一些相关的分析不属于统计的范畴。

### 补充练习

### 班级项目

班上的每个人选择近来自己看到或听到的一则广告,并对其中可能的歪曲提出批评,同时提出能澄清信息等方面的改进意见(注意:我们不要求你从商业的角度改进它)。就此提交一份报告。





## 附录 统 计 表

- 表 1 标准正态曲线下的面积
- 表 2 学生  $t$  分布的百分位数
- 表 3  $t$  检验的第二类错误概率曲线
- 表 4 用于中位数的置信区间和符号检验的百分位数:  $C_{\alpha, n}$
- 表 5 Wilcoxon 秩和检验的临界值  $T_L$  和  $T_U$ : 独立样本
- 表 6 Wilcoxon 符号秩检验的临界值
- 表 7 卡方分布的百分位数:  $\chi^2_{\alpha}$
- 表 8  $F$  分布的百分位数
- 表 9  $2\arcsin \sqrt{\hat{\pi}}$  的值
- 表 10 学生化极差分布的百分位数
- 表 11 Dunnett 检验的百分位数  $d_{\alpha}(k, v)$
- 表 12  $F_{\max} = S_{\max}^2 / S_{\min}^2$  的百分位数
- 表 13 随机数
- 表 14 方差分析表中检验的功效
- 表 15 泊松概率( $\mu$  在 0.1 与 4.0 之间)



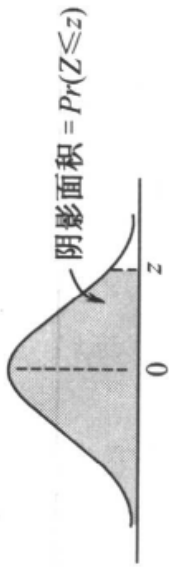


表 1 标准正态曲线下的面积

| $z$  | 0.00   | 0.01   | 0.02   | 0.03   | 0.04   | 0.05   | 0.06   | 0.07   | 0.08   | 0.09   |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| -3.4 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0002 |
| -3.3 | 0.0005 | 0.0005 | 0.0005 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0003 |
| -3.2 | 0.0007 | 0.0007 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0005 | 0.0005 | 0.0005 |
| -3.1 | 0.0010 | 0.0009 | 0.0009 | 0.0009 | 0.0008 | 0.0008 | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| -3.0 | 0.0013 | 0.0013 | 0.0013 | 0.0012 | 0.0012 | 0.0011 | 0.0011 | 0.0011 | 0.0010 | 0.0010 |
| -2.9 | 0.0019 | 0.0018 | 0.0018 | 0.0017 | 0.0016 | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |
| -2.8 | 0.0026 | 0.0025 | 0.0024 | 0.0023 | 0.0023 | 0.0022 | 0.0021 | 0.0021 | 0.0020 | 0.0019 |
| -2.7 | 0.0035 | 0.0034 | 0.0033 | 0.0032 | 0.0031 | 0.0030 | 0.0029 | 0.0028 | 0.0027 | 0.0026 |
| -2.6 | 0.0047 | 0.0045 | 0.0044 | 0.0043 | 0.0041 | 0.0040 | 0.0039 | 0.0038 | 0.0037 | 0.0036 |
| -2.5 | 0.0062 | 0.0060 | 0.0059 | 0.0057 | 0.0055 | 0.0054 | 0.0052 | 0.0051 | 0.0049 | 0.0048 |
| -2.4 | 0.0082 | 0.0080 | 0.0078 | 0.0075 | 0.0073 | 0.0071 | 0.0069 | 0.0068 | 0.0066 | 0.0064 |
| -2.3 | 0.0107 | 0.0104 | 0.0102 | 0.0099 | 0.0096 | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| -2.2 | 0.0139 | 0.0136 | 0.0132 | 0.0129 | 0.0125 | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| -2.1 | 0.0179 | 0.0174 | 0.0170 | 0.0166 | 0.0162 | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| -2.0 | 0.0228 | 0.0222 | 0.0217 | 0.0212 | 0.0207 | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| -1.9 | 0.0287 | 0.0281 | 0.0274 | 0.0268 | 0.0262 | 0.0256 | 0.0250 | 0.0244 | 0.0239 | 0.0233 |
| -1.8 | 0.0359 | 0.0351 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0301 | 0.0294 |
| -1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0384 | 0.0375 | 0.0367 |

续表

| <i>z</i> | 0.00   | 0.01   | 0.02   | 0.03   | 0.04   | 0.05   | 0.06   | 0.07   | 0.08   | 0.09   |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| -1.6     | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| -1.5     | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0571 | 0.0559 |
| -1.4     | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 | 0.0681 |
| -1.3     | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| -1.2     | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0985 |
| -1.1     | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| -1.0     | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| -0.9     | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |
| -0.8     | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| -0.7     | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2296 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| -0.6     | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| -0.5     | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2810 | 0.2776 |
| -0.4     | 0.3446 | 0.3409 | 0.3372 | 0.3336 | 0.3300 | 0.3264 | 0.3228 | 0.3192 | 0.3156 | 0.3121 |
| -0.3     | 0.3821 | 0.3783 | 0.3745 | 0.3707 | 0.3669 | 0.3632 | 0.3594 | 0.3557 | 0.3520 | 0.3483 |
| -0.2     | 0.4207 | 0.4168 | 0.4129 | 0.4090 | 0.4052 | 0.4013 | 0.3974 | 0.3936 | 0.3897 | 0.3859 |
| -0.1     | 0.4602 | 0.4602 | 0.4602 | 0.4602 | 0.4602 | 0.4602 | 0.4602 | 0.4602 | 0.4602 | 0.4602 |
| -0.0     | 0.5000 | 0.4960 | 0.4920 | 0.4880 | 0.4840 | 0.4801 | 0.4761 | 0.4721 | 0.4681 | 0.4641 |

面积

|       |            |
|-------|------------|
| -3.50 | 0.00023263 |
| -4.00 | 0.00003167 |
| -4.50 | 0.00000340 |
| -5.00 | 0.00000029 |

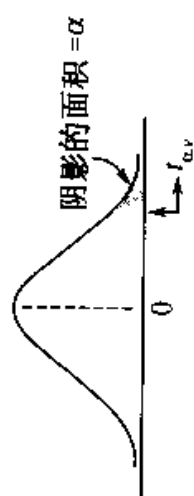
来源:由 M. Longnecker 用 Splus 计算得到。

续表

| $z$ | 0.00   | 0.01   | 0.02   | 0.03   | 0.04   | 0.05   | 0.06   | 0.07   | 0.08   | 0.09   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5398 | 0.5398 | 0.5398 | 0.5398 | 0.5398 | 0.5398 | 0.5398 | 0.5398 | 0.5398 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |

表  
续

[illegible]

表 2 学生  $t$  分布的百分位数

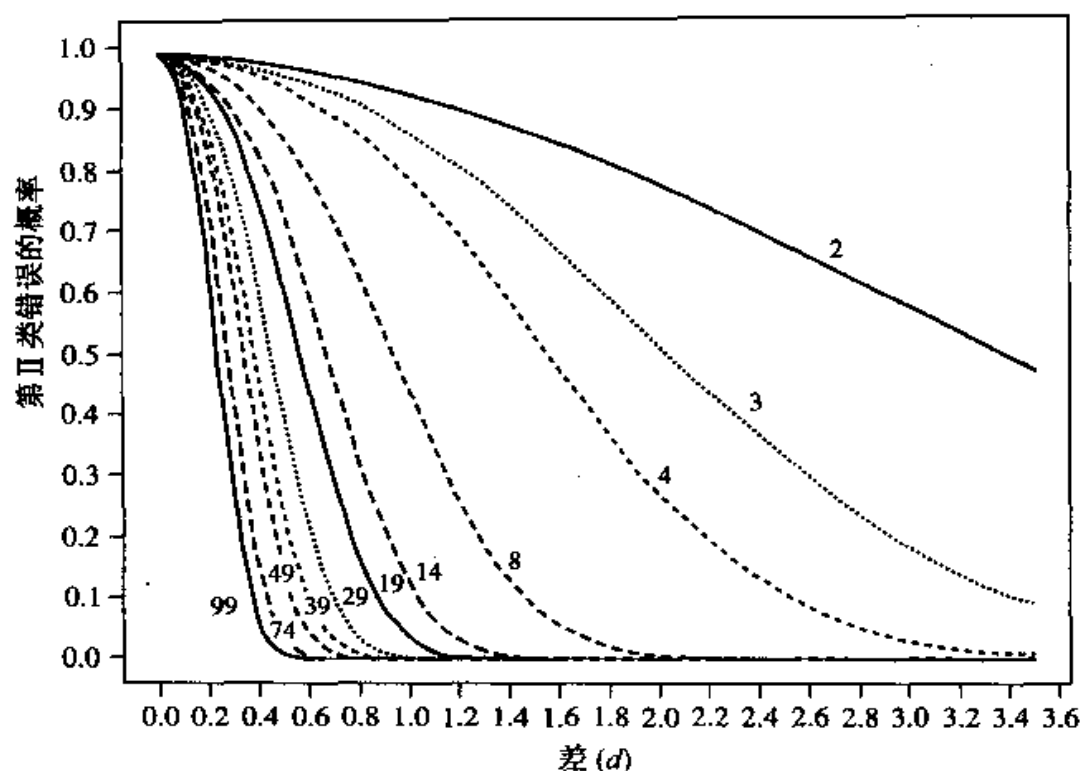
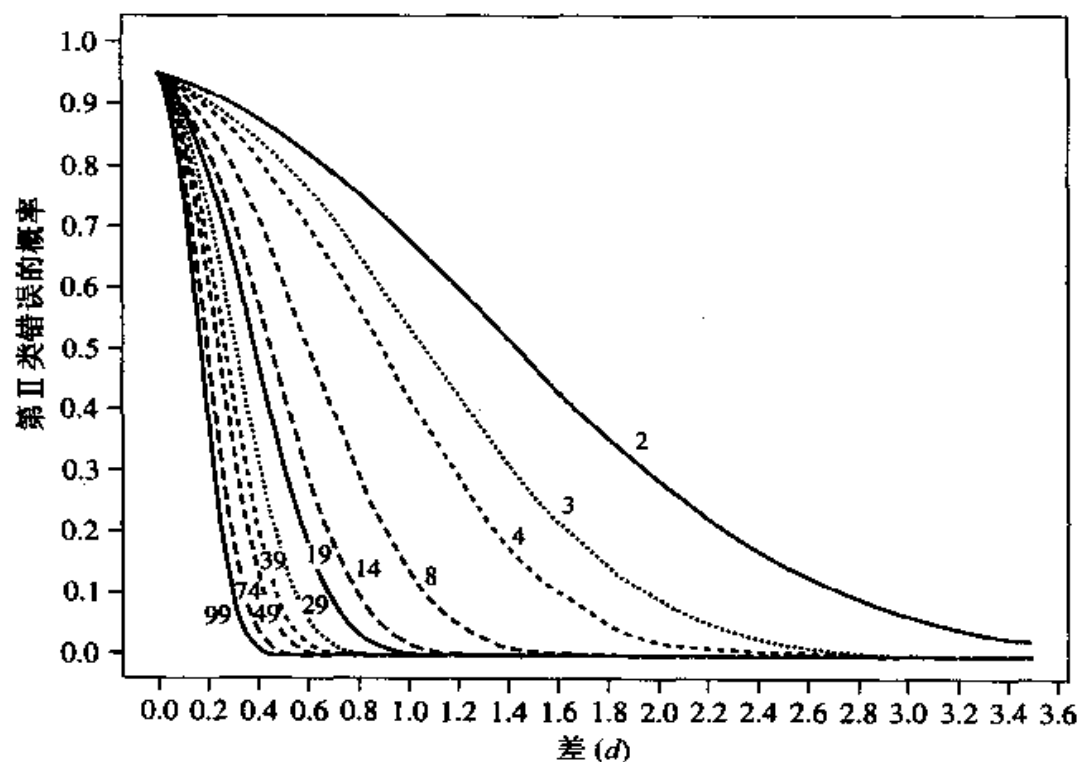
| $df/\alpha =$ | 0.40  | 0.25  | 0.10  | 0.05  | 0.025  | 0.01   | 0.005  | 0.001   | 0.0005  |
|---------------|-------|-------|-------|-------|--------|--------|--------|---------|---------|
| 1             | 0.325 | 1.000 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 318.309 | 636.619 |
| 2             | 0.289 | 0.816 | 1.886 | 2.920 | 4.303  | 6.965  | 9.925  | 22.327  | 31.599  |
| 3             | 0.277 | 0.765 | 1.638 | 2.353 | 3.182  | 4.541  | 5.841  | 10.215  | 12.924  |
| 4             | 0.271 | 0.741 | 1.533 | 2.132 | 2.776  | 3.747  | 4.604  | 7.173   | 8.610   |
| 5             | 0.267 | 0.727 | 1.476 | 2.015 | 2.571  | 3.365  | 4.032  | 5.893   | 6.869   |
| 6             | 0.265 | 0.718 | 1.440 | 1.943 | 2.447  | 3.143  | 3.707  | 5.208   | 5.959   |
| 7             | 0.263 | 0.711 | 1.415 | 1.895 | 2.365  | 2.998  | 3.499  | 4.785   | 5.408   |
| 8             | 0.262 | 0.706 | 1.397 | 1.860 | 2.306  | 2.896  | 3.355  | 4.501   | 5.041   |
| 9             | 0.261 | 0.703 | 1.383 | 1.833 | 2.262  | 2.821  | 3.250  | 4.297   | 4.781   |
| 10            | 0.260 | 0.700 | 1.372 | 1.812 | 2.228  | 2.764  | 3.169  | 4.144   | 4.587   |
| 11            | 0.260 | 0.697 | 1.363 | 1.796 | 2.201  | 2.718  | 3.106  | 4.025   | 4.437   |
| 12            | 0.259 | 0.695 | 1.356 | 1.782 | 2.179  | 2.681  | 3.055  | 3.930   | 4.318   |
| 13            | 0.259 | 0.694 | 1.350 | 1.771 | 2.160  | 2.650  | 3.012  | 3.852   | 4.221   |
| 14            | 0.258 | 0.692 | 1.345 | 1.761 | 2.145  | 2.624  | 2.977  | 3.787   | 4.140   |
| 15            | 0.258 | 0.691 | 1.341 | 1.753 | 2.131  | 2.602  | 2.947  | 3.733   | 4.073   |



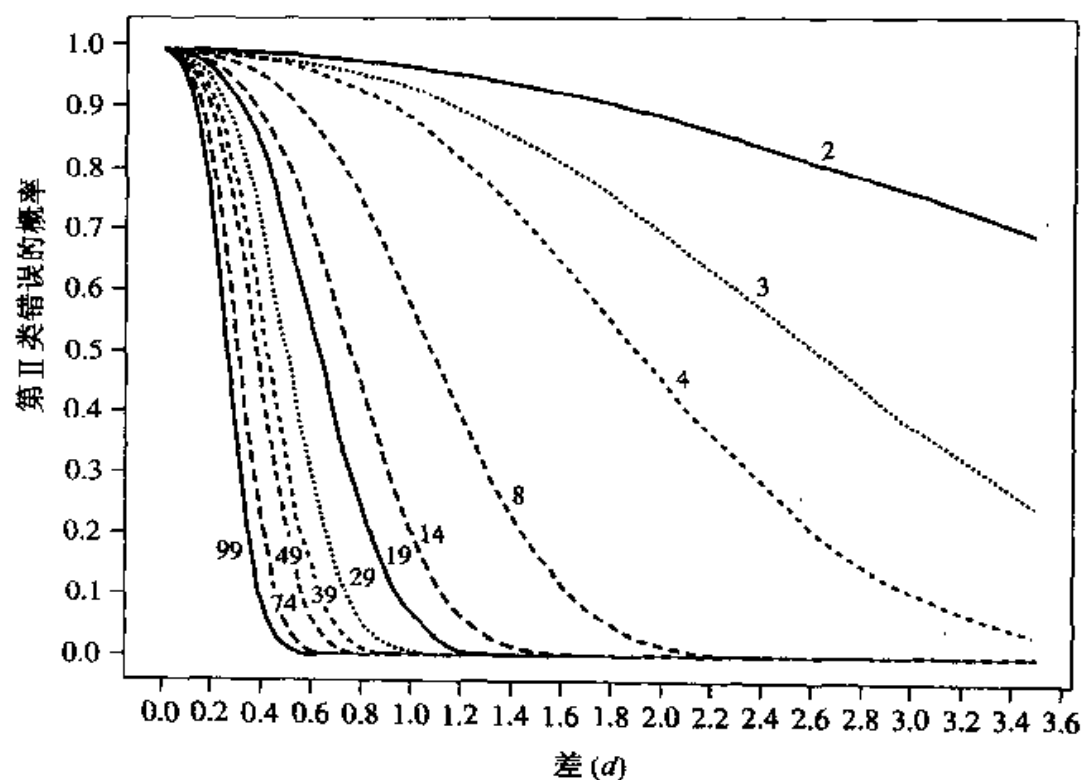
续表

| df/ $\alpha =$ | 0.40  | 0.25  | 0.10  | 0.05  | 0.025 | 0.01  | 0.005 | 0.001 | 0.0005 |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| 16             | 0.258 | 0.690 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015  |
| 17             | 0.257 | 0.689 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965  |
| 18             | 0.257 | 0.688 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922  |
| 19             | 0.257 | 0.688 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883  |
| 20             | 0.257 | 0.687 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850  |
| 21             | 0.257 | 0.686 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 | 3.819  |
| 22             | 0.256 | 0.686 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792  |
| 23             | 0.256 | 0.685 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 | 3.768  |
| 24             | 0.256 | 0.685 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745  |
| 25             | 0.256 | 0.684 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 | 3.725  |
| 26             | 0.256 | 0.684 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707  |
| 27             | 0.256 | 0.684 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.690  |
| 28             | 0.256 | 0.683 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674  |
| 29             | 0.256 | 0.683 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 | 3.659  |
| 30             | 0.256 | 0.683 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 | 3.646  |
| 35             | 0.255 | 0.682 | 1.306 | 1.690 | 2.030 | 2.438 | 2.724 | 3.340 | 3.591  |
| 40             | 0.255 | 0.681 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 | 3.551  |
| 50             | 0.255 | 0.679 | 1.299 | 1.676 | 2.009 | 2.403 | 2.678 | 3.261 | 3.496  |
| 60             | 0.254 | 0.679 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 | 3.460  |
| 120            | 0.254 | 0.677 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 | 3.373  |
| inf.           | 0.253 | 0.674 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291  |

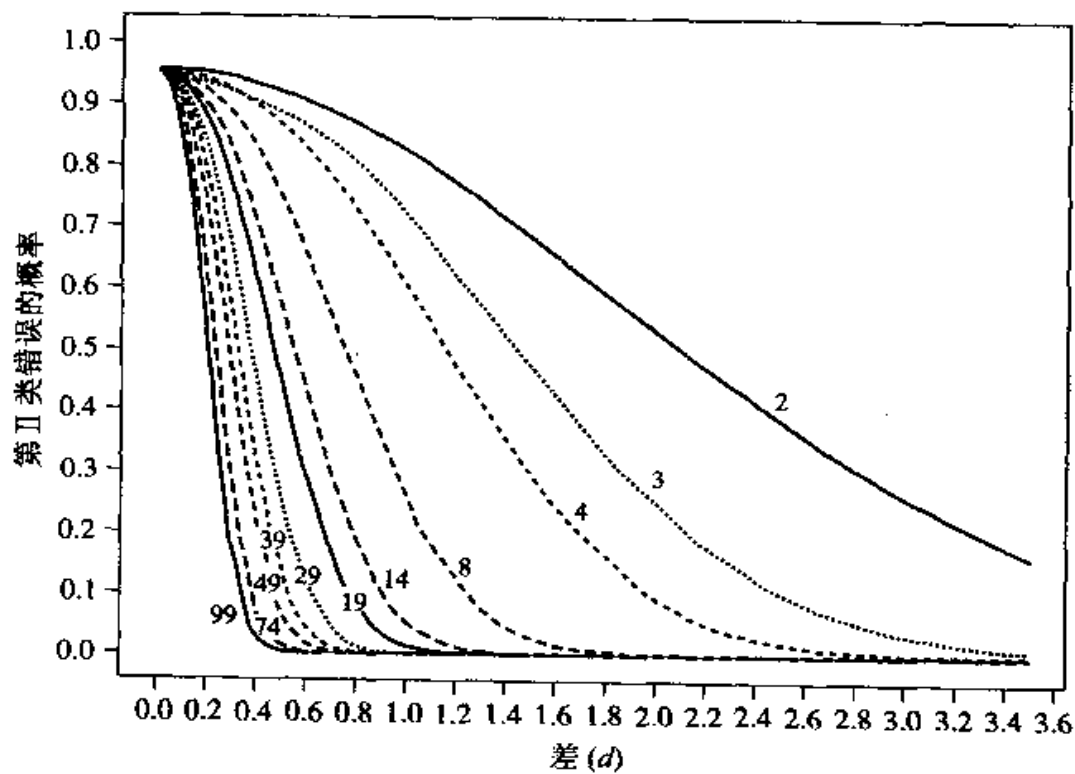
来源:由 M. Longnecker 用 Splus 计算得到。

表 3  $t$  检验的第二类错误概率曲线 $\alpha = 0.01$  时第二类错误概率曲线(单侧) $\alpha = 0.05$  时第二类错误概率曲线(单侧)

来源:由 M. Longnecker 用 SAS 计算得到。



$\alpha = 0.01$  时第二类错误的概率曲线(双侧)



$\alpha = 0.05$  时第二类错误的概率曲线(双侧)

来源:由 M. Longnecker 用 SAS 计算得到。

表4 用于中位数的置信区间和符号检验的百分位数:  $C_{\alpha, n}$ 

| $\alpha(2)$ | 0.20 | 0.10 | 0.05  | 0.025 | 0.01  | 0.005  | 0.0025 | 0.002 | $\alpha(1)$ | 0.20 | 0.10 | 0.05  | 0.025 | 0.01  | 0.005  | 0.0025 | 0.002 |
|-------------|------|------|-------|-------|-------|--------|--------|-------|-------------|------|------|-------|-------|-------|--------|--------|-------|
| $\alpha(1)$ | 0.10 | 0.05 | 0.025 | 0.01  | 0.005 | 0.0025 | 0.001  |       | $\alpha(2)$ | 0.10 | 0.05 | 0.025 | 0.01  | 0.005 | 0.0025 | 0.001  |       |
| $n$         |      |      |       |       |       |        |        |       |             |      |      |       |       |       |        |        |       |
| 1           | *    | *    | *     | *     | *     | *      | *      | *     | 26          | 9    | 8    | 7     | 6     | 6     | 5      | 4      |       |
| 2           | *    | *    | *     | *     | *     | *      | *      | *     | 27          | 9    | 8    | 7     | 7     | 6     | 5      | 5      |       |
| 3           | *    | *    | *     | *     | *     | *      | *      | *     | 28          | 10   | 9    | 8     | 7     | 6     | 6      | 5      |       |
| 4           | 0    | *    | *     | *     | *     | *      | *      | *     | 29          | 10   | 9    | 8     | 7     | 7     | 6      | 5      |       |
| 5           | 0    | 0    | *     | *     | *     | *      | *      | *     | 30          | 10   | 10   | 9     | 8     | 7     | 6      | 6      |       |
| 6           | 0    | 0    | 0     | *     | *     | *      | *      | *     | 31          | 11   | 10   | 9     | 8     | 7     | 7      | 6      |       |
| 7           | 1    | 0    | 0     | 0     | *     | *      | *      | *     | 32          | 11   | 10   | 9     | 8     | 8     | 7      | 6      |       |
| 8           | 1    | 1    | 0     | 0     | 0     | *      | *      | *     | 33          | 12   | 11   | 10    | 9     | 8     | 8      | 7      |       |
| 9           | 2    | 1    | 1     | 0     | 0     | 0      | *      | *     | 34          | 12   | 11   | 10    | 9     | 9     | 8      | 7      |       |
| 10          | 2    | 1    | 1     | 0     | 0     | 0      | 0      | 0     | 35          | 13   | 12   | 11    | 10    | 9     | 8      | 8      |       |
| 11          | 2    | 2    | 1     | 1     | 0     | 0      | 0      | 0     | 36          | 13   | 12   | 11    | 10    | 9     | 9      | 8      |       |
| 12          | 3    | 2    | 2     | 1     | 1     | 0      | 0      | 0     | 37          | 14   | 13   | 12    | 10    | 10    | 9      | 8      |       |
| 13          | 3    | 3    | 2     | 1     | 1     | 1      | 0      | 0     | 38          | 14   | 13   | 12    | 11    | 10    | 9      | 9      |       |
| 14          | 4    | 3    | 2     | 2     | 1     | 1      | 1      | 1     | 39          | 15   | 13   | 12    | 11    | 11    | 10     | 9      |       |
| 15          | 4    | 3    | 3     | 2     | 2     | 1      | 1      | 1     | 40          | 15   | 14   | 13    | 12    | 11    | 10     | 9      |       |
| 16          | 4    | 4    | 3     | 2     | 2     | 2      | 1      | 1     | 41          | 15   | 14   | 13    | 12    | 11    | 11     | 10     |       |
| 17          | 5    | 4    | 4     | 3     | 2     | 2      | 2      | 1     | 42          | 16   | 15   | 14    | 13    | 12    | 11     | 10     |       |
| 18          | 5    | 5    | 4     | 3     | 3     | 2      | 2      | 2     | 43          | 16   | 15   | 14    | 13    | 12    | 11     | 11     |       |
| 19          | 6    | 5    | 4     | 4     | 3     | 3      | 2      | 2     | 44          | 17   | 16   | 15    | 13    | 13    | 12     | 11     |       |
| 20          | 6    | 5    | 5     | 4     | 4     | 3      | 3      | 2     | 45          | 17   | 16   | 15    | 14    | 13    | 12     | 11     |       |
| 21          | 7    | 6    | 5     | 4     | 4     | 4      | 3      | 3     | 46          | 18   | 16   | 15    | 14    | 13    | 13     | 12     |       |
| 22          | 7    | 6    | 5     | 5     | 4     | 4      | 4      | 3     | 47          | 18   | 17   | 16    | 15    | 14    | 13     | 12     |       |
| 23          | 7    | 7    | 6     | 5     | 5     | 4      | 4      | 3     | 48          | 19   | 17   | 16    | 15    | 14    | 13     | 12     |       |
| 24          | 8    | 7    | 6     | 5     | 5     | 5      | 4      | 4     | 49          | 19   | 18   | 17    | 15    | 15    | 14     | 13     |       |
| 25          | 8    | 7    | 7     | 6     | 5     | 5      | 5      | 4     | 50          | 19   | 18   | 17    | 16    | 15    | 14     | 13     |       |

注: \*号表示没有该水平的检验或置信区间。

来源: 由 M. Longnecker 用 Splus 计算得到。

表 5 Wilcoxon 秩和检验的临界值  $T_L$  和  $T_U$ ; 独立样本

检验统计量是关于容量较小的样本的秩的和 (如果样本容量相等, 两个样本的秩和都可以用)。

| a. $\alpha = 0.025$ 单侧 |   | $\alpha = 0.05$ 双侧 |       |       |       |       |       |       |       |       |       |
|------------------------|---|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $n_1 \backslash n_2$   |   | 3                  | 4     | 5     | 6     | 7     | 8     | 9     | 10    |       |       |
|                        |   | $T_L$              | $T_U$ | $T_L$ | $T_U$ | $T_L$ | $T_U$ | $T_L$ | $T_U$ | $T_L$ | $T_U$ |
| 3                      | 5 | 16                 | 6     | 18    | 6     | 21    | 7     | 23    | 7     | 26    | 8     |
| 4                      | 6 | 18                 | 11    | 25    | 12    | 28    | 12    | 32    | 13    | 35    | 14    |
| 5                      | 6 | 21                 | 12    | 28    | 18    | 37    | 19    | 41    | 20    | 45    | 21    |
| 6                      | 7 | 23                 | 12    | 32    | 19    | 41    | 26    | 52    | 28    | 56    | 29    |
| 7                      | 7 | 26                 | 13    | 35    | 20    | 45    | 28    | 56    | 37    | 68    | 39    |
| 8                      | 8 | 28                 | 14    | 38    | 21    | 49    | 29    | 61    | 39    | 73    | 49    |
| 9                      | 8 | 31                 | 15    | 41    | 22    | 53    | 31    | 65    | 41    | 78    | 51    |
| 10                     | 9 | 33                 | 16    | 44    | 24    | 56    | 32    | 70    | 43    | 83    | 54    |

| b. $\alpha = .05$ 单侧 |    | $\alpha = 0.10$ 双侧 |       |       |       |       |       |       |       |       |       |
|----------------------|----|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $n_1 \backslash n_2$ |    | 3                  | 4     | 5     | 6     | 7     | 8     | 9     | 10    |       |       |
|                      |    | $T_L$              | $T_U$ | $T_L$ | $T_U$ | $T_L$ | $T_U$ | $T_L$ | $T_U$ | $T_L$ | $T_U$ |
| 3                    | 6  | 15                 | 7     | 17    | 7     | 20    | 8     | 22    | 9     | 24    | 9     |
| 4                    | 7  | 17                 | 12    | 24    | 13    | 27    | 14    | 30    | 15    | 33    | 16    |
| 5                    | 7  | 20                 | 13    | 27    | 19    | 36    | 20    | 40    | 22    | 43    | 24    |
| 6                    | 8  | 22                 | 14    | 30    | 20    | 40    | 28    | 50    | 30    | 54    | 32    |
| 7                    | 9  | 24                 | 15    | 33    | 22    | 43    | 30    | 54    | 39    | 66    | 41    |
| 8                    | 9  | 27                 | 16    | 36    | 24    | 46    | 32    | 58    | 41    | 71    | 52    |
| 9                    | 10 | 29                 | 17    | 39    | 25    | 50    | 33    | 63    | 43    | 76    | 54    |
| 10                   | 11 | 31                 | 18    | 42    | 26    | 54    | 35    | 67    | 46    | 80    | 57    |

来源: F. Wilcoxon and R. A. Wilcoxon, *Some Rapid Approximate Statistical Procedures* (Pearl River, N. Y. Lederle Laboratories, 1964), pp. 20 - 23. 经 American Cyanamid Company 允许重印。

表 6 Wilcoxon 符号秩检验的临界值 [ $n = 5(1)54$ ]

| 单侧           | 双侧          | $n = 5$  | $n = 6$  | $n = 7$  | $n = 8$  | $n = 9$  |
|--------------|-------------|----------|----------|----------|----------|----------|
| $p = 0.1$    | $p = 0.2$   | 2        | 3        | 5        | 8        | 10       |
| $p = 0.05$   | $p = 0.1$   | 0        | 2        | 3        | 5        | 8        |
| $p = 0.025$  | $p = 0.05$  |          | 0        | 2        | 3        | 5        |
| $p = 0.01$   | $p = 0.02$  |          |          | 0        | 1        | 3        |
| $p = 0.005$  | $p = 0.01$  |          |          |          | 0        | 1        |
| $p = 0.0025$ | $p = 0.005$ |          |          |          |          | 0        |
| $p = 0.001$  | $p = 0.002$ |          |          |          |          |          |
| 单侧           | 双侧          | $n = 15$ | $n = 16$ | $n = 17$ | $n = 18$ | $n = 19$ |
| $p = 0.1$    | $p = 0.2$   | 36       | 42       | 48       | 55       | 62       |
| $p = 0.05$   | $p = 0.1$   | 30       | 35       | 41       | 47       | 53       |
| $p = 0.025$  | $p = 0.05$  | 25       | 29       | 34       | 40       | 46       |
| $p = 0.01$   | $p = 0.02$  | 19       | 23       | 27       | 32       | 37       |
| $p = 0.005$  | $p = 0.01$  | 15       | 19       | 23       | 27       | 32       |
| $p = 0.0025$ | $p = 0.005$ | 12       | 15       | 19       | 23       | 27       |
| $p = 0.001$  | $p = 0.002$ | 8        | 11       | 14       | 18       | 21       |
| 单侧           | 双侧          | $n = 25$ | $n = 26$ | $n = 27$ | $n = 28$ | $n = 29$ |
| $p = 0.1$    | $p = 0.2$   | 113      | 124      | 134      | 145      | 157      |
| $p = 0.05$   | $p = 0.1$   | 100      | 110      | 119      | 130      | 140      |
| $p = 0.025$  | $p = 0.05$  | 89       | 98       | 107      | 116      | 126      |
| $p = 0.01$   | $p = 0.02$  | 76       | 84       | 92       | 101      | 110      |
| $p = 0.005$  | $p = 0.01$  | 68       | 75       | 83       | 91       | 100      |
| $p = 0.0025$ | $p = 0.005$ | 60       | 67       | 74       | 82       | 90       |
| $p = 0.001$  | $p = 0.002$ | 51       | 58       | 64       | 71       | 79       |

续表

| 单侧           | 双侧          | $n = 35$ | $n = 36$ | $n = 37$ | $n = 38$ | $n = 39$ |
|--------------|-------------|----------|----------|----------|----------|----------|
| $p = 0.1$    | $p = 0.2$   | 235      | 250      | 265      | 281      | 297      |
| $p = 0.05$   | $p = 0.1$   | 213      | 227      | 241      | 256      | 271      |
| $p = 0.025$  | $p = 0.05$  | 195      | 208      | 221      | 235      | 249      |
| $p = 0.01$   | $p = 0.02$  | 173      | 185      | 198      | 211      | 224      |
| $p = 0.005$  | $p = 0.01$  | 159      | 171      | 182      | 194      | 207      |
| $p = 0.0025$ | $p = 0.005$ | 146      | 157      | 168      | 180      | 192      |
| $p = 0.001$  | $p = 0.002$ | 131      | 141      | 151      | 162      | 173      |
| 单侧           | 双侧          | $n = 45$ | $n = 46$ | $n = 47$ | $n = 48$ | $n = 49$ |
| $p = 0.1$    | $p = 0.2$   | 402      | 422      | 441      | 462      | 482      |
| $p = 0.05$   | $p = 0.1$   | 371      | 389      | 407      | 426      | 446      |
| $p = 0.025$  | $p = 0.05$  | 343      | 361      | 378      | 396      | 415      |
| $p = 0.01$   | $p = 0.02$  | 312      | 328      | 345      | 362      | 379      |
| $p = 0.005$  | $p = 0.01$  | 291      | 307      | 322      | 339      | 355      |
| $p = 0.0025$ | $p = 0.005$ | 272      | 287      | 302      | 318      | 334      |
| $p = 0.001$  | $p = 0.002$ | 249      | 263      | 277      | 292      | 307      |
| 单侧           | 双侧          | $n = 10$ | $n = 11$ | $n = 12$ | $n = 13$ | $n = 14$ |
| $p = 0.1$    | $p = 0.2$   | 14       | 17       | 21       | 26       | 31       |
| $p = 0.05$   | $p = 0.1$   | 10       | 13       | 17       | 21       | 25       |
| $p = 0.025$  | $p = 0.05$  | 8        | 10       | 13       | 17       | 21       |
| $p = 0.01$   | $p = 0.02$  | 5        | 7        | 9        | 12       | 15       |
| $p = 0.005$  | $p = 0.01$  | 3        | 5        | 7        | 9        | 12       |
| $p = 0.0025$ | $p = 0.005$ | 1        | 3        | 5        | 7        | 9        |
| $p = 0.001$  | $p = 0.002$ | 0        | 1        | 2        | 4        | 6        |

续表

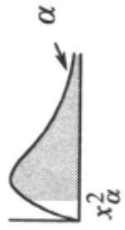
| 单侧           | 双侧          | $n = 20$ | $n = 21$ | $n = 22$ | $n = 23$ | $n = 24$ |
|--------------|-------------|----------|----------|----------|----------|----------|
| $p = 0.1$    | $p = 0.2$   | 69       | 77       | 86       | 94       | 104      |
| $p = 0.05$   | $p = 0.1$   | 60       | 67       | 75       | 83       | 91       |
| $p = 0.025$  | $p = 0.05$  | 52       | 58       | 65       | 73       | 81       |
| $p = 0.01$   | $p = 0.02$  | 43       | 49       | 55       | 62       | 69       |
| $p = 0.005$  | $p = 0.01$  | 37       | 42       | 48       | 54       | 61       |
| $p = 0.0025$ | $p = 0.005$ | 32       | 37       | 42       | 48       | 54       |
| $p = 0.001$  | $p = 0.002$ | 26       | 30       | 35       | 40       | 45       |
| 单侧           | 双侧          | $n = 30$ | $n = 31$ | $n = 32$ | $n = 33$ | $n = 34$ |
| $p = 0.1$    | $p = 0.2$   | 169      | 181      | 194      | 207      | 221      |
| $p = 0.05$   | $p = 0.1$   | 151      | 163      | 175      | 187      | 200      |
| $p = 0.025$  | $p = 0.05$  | 137      | 147      | 159      | 170      | 182      |
| $p = 0.01$   | $p = 0.02$  | 120      | 130      | 140      | 151      | 162      |
| $p = 0.005$  | $p = 0.01$  | 109      | 118      | 128      | 138      | 148      |
| $p = 0.0025$ | $p = 0.005$ | 98       | 107      | 116      | 126      | 136      |
| $p = .001$   | $p = .002$  | 86       | 94       | 103      | 112      | 121      |



续表

| 单侧           | 双侧          | $n = 40$ | $n = 41$ | $n = 42$ | $n = 43$ | $n = 44$ |
|--------------|-------------|----------|----------|----------|----------|----------|
| $p = 0.1$    | $p = 0.2$   | 313      | 330      | 348      | 365      | 384      |
| $p = 0.05$   | $p = 0.1$   | 286      | 302      | 319      | 336      | 353      |
| $p = 0.025$  | $p = 0.05$  | 264      | 279      | 294      | 310      | 327      |
| $p = 0.01$   | $p = 0.02$  | 238      | 252      | 266      | 281      | 296      |
| $p = 0.005$  | $p = 0.01$  | 220      | 233      | 247      | 261      | 276      |
| $p = 0.0025$ | $p = 0.005$ | 204      | 217      | 230      | 244      | 258      |
| $p = .001$   | $p = .002$  | 185      | 197      | 209      | 222      | 235      |
| 单侧           | 双侧          | $n = 50$ | $n = 51$ | $n = 52$ | $n = 53$ | $n = 54$ |
| $p = 0.1$    | $p = 0.2$   | 503      | 525      | 547      | 569      | 592      |
| $p = 0.05$   | $p = 0.1$   | 466      | 486      | 507      | 529      | 550      |
| $p = 0.025$  | $p = 0.05$  | 434      | 453      | 473      | 494      | 514      |
| $p = 0.01$   | $p = 0.02$  | 397      | 416      | 434      | 454      | 473      |
| $p = 0.005$  | $p = 0.01$  | 373      | 390      | 408      | 427      | 445      |
| $p = 0.0025$ | $p = 0.005$ | 350      | 367      | 384      | 402      | 420      |
| $p = 0.001$  | $p = 0.002$ | 323      | 339      | 355      | 372      | 389      |

来源:由 P.J.Hildebrand 计算。

表 7 卡方分布的百分位数:  $\chi_{\alpha}^2$ 

| $df/\alpha =$ | 0.999    | 0.995    | 0.99     | 0.975    | 0.95     | 0.90    |
|---------------|----------|----------|----------|----------|----------|---------|
| 1             | 0.000002 | 0.000039 | 0.000157 | 0.000982 | 0.003932 | 0.01579 |
| 2             | 0.002001 | 0.01003  | 0.02010  | 0.05064  | 0.1026   | 0.2107  |
| 3             | 0.02430  | 0.07172  | 0.1148   | 0.2158   | 0.3518   | 0.5844  |
| 4             | 0.09080  | 0.2070   | 0.2971   | 0.4844   | 0.7107   | 1.064   |
| 5             | 0.2102   | 0.4117   | 0.5543   | 0.8312   | 1.145    | 1.610   |
| 6             | 0.3811   | 0.6757   | 0.8721   | 1.237    | 1.635    | 2.204   |
| 7             | 0.5985   | 0.9893   | 1.239    | 1.690    | 2.167    | 2.833   |
| 8             | 0.8571   | 1.344    | 1.646    | 2.180    | 2.733    | 3.490   |
| 9             | 1.152    | 1.735    | 2.088    | 2.700    | 3.325    | 4.168   |
| 10            | 1.479    | 2.156    | 2.558    | 3.247    | 3.940    | 4.865   |
| 11            | 1.834    | 2.603    | 3.053    | 3.816    | 4.575    | 5.578   |
| 12            | 2.214    | 3.074    | 3.571    | 4.404    | 5.226    | 6.304   |
| 13            | 2.617    | 3.565    | 4.107    | 5.009    | 5.892    | 7.042   |
| 14            | 3.041    | 4.075    | 4.660    | 5.629    | 6.571    | 7.790   |
| 15            | 3.483    | 4.601    | 5.229    | 6.262    | 7.261    | 8.547   |
| 16            | 3.942    | 5.142    | 5.812    | 6.908    | 7.962    | 9.312   |
| 17            | 4.416    | 5.697    | 6.408    | 7.564    | 8.672    | 10.09   |
| 18            | 4.905    | 6.265    | 7.015    | 8.231    | 9.390    | 10.86   |
| 19            | 5.407    | 6.844    | 7.633    | 8.907    | 10.12    | 11.65   |
| 20            | 5.921    | 7.434    | 8.260    | 9.591    | 10.85    | 12.44   |

续表

| $df/\alpha =$ | 0.999  | 0.995  | 0.99   | 0.975  | 0.95   | 0.90   |
|---------------|--------|--------|--------|--------|--------|--------|
| 21            | 6.447  | 8.034  | 8.897  | 10.28  | 11.59  | 13.24  |
| 22            | 6.983  | 8.643  | 9.542  | 10.98  | 12.34  | 14.04  |
| 23            | 7.529  | 9.260  | 10.20  | 11.69  | 13.09  | 14.85  |
| 24            | 8.085  | 9.886  | 10.86  | 12.40  | 13.85  | 15.66  |
| 25            | 8.649  | 10.52  | 11.52  | 13.12  | 14.61  | 16.47  |
| 26            | 9.222  | 11.16  | 12.20  | 13.84  | 15.38  | 17.29  |
| 27            | 9.803  | 11.81  | 12.88  | 14.57  | 16.15  | 18.11  |
| 28            | 10.39  | 12.46  | 13.56  | 15.31  | 16.93  | 18.94  |
| 29            | 10.99  | 13.12  | 14.26  | 16.06  | 17.71  | 19.77  |
| 30            | 11.59  | 13.79  | 14.95  | 16.79  | 18.49  | 20.60  |
| 40            | 17.92  | 20.71  | 22.16  | 24.43  | 26.51  | 29.05  |
| 50            | 24.67  | 27.99  | 29.71  | 32.36  | 34.76  | 37.69  |
| 60            | 31.74  | 35.53  | 37.48  | 40.48  | 43.19  | 46.46  |
| 70            | 39.04  | 43.28  | 45.44  | 48.76  | 51.74  | 55.33  |
| 80            | 46.52  | 51.17  | 53.54  | 57.15  | 60.39  | 64.28  |
| 90            | 54.16  | 59.20  | 61.75  | 65.65  | 69.13  | 73.29  |
| 100           | 61.92  | 67.33  | 70.06  | 74.22  | 77.93  | 82.36  |
| 120           | 77.76  | 83.85  | 86.92  | 91.57  | 95.70  | 100.62 |
| 240           | 177.95 | 187.32 | 191.99 | 198.98 | 205.14 | 212.39 |

续表

| $\alpha=0.10$ | 0.05  | 0.025 | 0.01  | 0.005 | 0.001 | df |
|---------------|-------|-------|-------|-------|-------|----|
| 2.706         | 3.841 | 5.024 | 6.635 | 7.879 | 10.83 | 1  |
| 4.605         | 5.991 | 7.378 | 9.210 | 10.60 | 13.82 | 2  |
| 6.251         | 7.815 | 9.348 | 11.34 | 12.84 | 16.27 | 3  |
| 7.779         | 9.488 | 11.14 | 13.28 | 14.86 | 18.47 | 4  |
| 9.236         | 11.07 | 12.83 | 15.09 | 16.75 | 20.52 | 5  |
| 10.64         | 12.59 | 14.45 | 16.81 | 18.55 | 22.46 | 6  |
| 12.02         | 14.07 | 16.01 | 18.48 | 20.28 | 24.32 | 7  |
| 13.36         | 15.51 | 17.53 | 20.09 | 21.95 | 26.12 | 8  |
| 14.68         | 16.92 | 19.02 | 21.67 | 23.59 | 27.88 | 9  |
| 15.99         | 18.31 | 20.48 | 23.21 | 25.19 | 29.59 | 10 |
| 17.28         | 19.68 | 21.92 | 24.72 | 26.76 | 31.27 | 11 |
| 18.55         | 21.03 | 23.34 | 26.22 | 28.30 | 32.91 | 12 |
| 19.81         | 22.36 | 24.74 | 27.69 | 29.82 | 34.53 | 13 |
| 21.06         | 23.68 | 26.12 | 29.14 | 31.32 | 36.12 | 14 |
| 22.31         | 25.00 | 27.49 | 30.58 | 32.80 | 37.70 | 15 |
| 23.54         | 26.30 | 28.85 | 32.00 | 34.27 | 39.25 | 16 |
| 24.77         | 27.59 | 30.19 | 33.41 | 35.72 | 40.79 | 17 |
| 25.99         | 28.87 | 31.53 | 34.81 | 37.16 | 42.31 | 18 |
| 27.20         | 30.14 | 32.85 | 36.19 | 38.58 | 43.82 | 19 |
| 28.41         | 31.41 | 34.17 | 37.57 | 40.00 | 45.31 | 20 |
| 29.62         | 32.67 | 35.48 | 38.93 | 41.40 | 46.80 | 21 |
| 30.81         | 33.92 | 36.78 | 40.29 | 42.80 | 48.27 | 22 |
| 32.01         | 35.17 | 38.08 | 41.64 | 44.18 | 49.73 | 23 |

续表

| $\alpha = 0.10$ | 0.05   | 0.025  | 0.01   | 0.005  | 0.001  | df  |
|-----------------|--------|--------|--------|--------|--------|-----|
| 33.20           | 36.42  | 39.36  | 42.98  | 45.56  | 51.18  | 24  |
| 34.38           | 37.65  | 40.65  | 44.31  | 46.93  | 52.62  | 25  |
| 35.56           | 38.89  | 41.92  | 45.64  | 48.29  | 54.05  | 26  |
| 36.74           | 40.11  | 43.19  | 46.96  | 49.65  | 55.48  | 27  |
| 37.92           | 41.34  | 44.46  | 48.28  | 50.99  | 56.89  | 28  |
| 39.09           | 42.56  | 45.72  | 49.59  | 52.34  | 58.30  | 29  |
| 40.26           | 43.77  | 46.98  | 50.89  | 53.67  | 59.70  | 30  |
| 51.81           | 55.76  | 59.34  | 63.69  | 66.77  | 73.40  | 40  |
| 63.17           | 67.50  | 71.42  | 76.15  | 79.49  | 86.66  | 50  |
| 74.40           | 79.08  | 83.30  | 88.38  | 91.95  | 99.61  | 60  |
| 85.53           | 90.53  | 95.02  | 100.43 | 104.21 | 112.32 | 70  |
| 96.58           | 101.88 | 106.63 | 112.33 | 116.32 | 124.84 | 80  |
| 107.57          | 113.15 | 118.14 | 124.12 | 128.30 | 137.21 | 90  |
| 118.50          | 124.34 | 129.56 | 135.81 | 140.17 | 149.45 | 100 |
| 140.23          | 146.57 | 152.21 | 158.95 | 163.65 | 173.62 | 120 |
| 268.47          | 277.14 | 284.80 | 293.89 | 300.18 | 313.44 | 240 |

来源：由 P. J. Hildebrand 计算。

表 8 F 分布的百分位数 ( $df_2$  从 1 到 6)

| $df_2$ | $\alpha$ | $df_1$ |       |       |       |       |       |       |       |       |       |
|--------|----------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|        |          | 1      | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
| 1      | 0.25     | 5.83   | 7.50  | 8.20  | 8.58  | 8.82  | 8.98  | 9.10  | 9.19  | 9.26  | 9.32  |
|        | 0.10     | 39.86  | 49.50 | 53.59 | 55.83 | 57.24 | 58.20 | 58.91 | 59.44 | 59.86 | 60.19 |
|        | 0.05     | 161.4  | 199.5 | 215.7 | 224.6 | 230.2 | 234.0 | 236.8 | 238.9 | 240.5 | 241.9 |
|        | 0.025    | 647.8  | 799.5 | 864.2 | 899.6 | 921.8 | 937.1 | 948.2 | 956.7 | 963.3 | 968.6 |
|        | 0.01     | 4052   | 5000  | 5403  | 5625  | 5764  | 5859  | 5928  | 5981  | 6022  | 6056  |
| 2      | 0.25     | 2.57   | 3.00  | 3.15  | 3.23  | 3.28  | 3.31  | 3.34  | 3.35  | 3.37  | 3.38  |
|        | 0.10     | 8.53   | 9.00  | 9.16  | 9.24  | 9.29  | 9.33  | 9.35  | 9.37  | 9.38  | 9.39  |
|        | 0.05     | 18.51  | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 | 19.40 |
|        | 0.025    | 38.51  | 39.00 | 39.17 | 39.25 | 39.30 | 39.33 | 39.36 | 39.37 | 39.39 | 39.40 |
|        | 0.01     | 98.50  | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.37 | 99.39 | 99.40 |
| 3      | 0.005    | 198.5  | 199.0 | 199.2 | 199.2 | 199.3 | 199.3 | 199.4 | 199.4 | 199.4 | 199.4 |
|        | 0.001    | 998.5  | 999.0 | 999.2 | 999.2 | 999.3 | 999.3 | 999.4 | 999.4 | 999.4 | 999.4 |
|        | 0.25     | 2.02   | 2.28  | 2.36  | 2.39  | 2.41  | 2.42  | 2.43  | 2.44  | 2.44  | 2.44  |
|        | 0.10     | 5.54   | 5.46  | 5.39  | 5.34  | 5.31  | 5.28  | 5.27  | 5.25  | 5.24  | 5.23  |
|        | 0.05     | 10.13  | 9.55  | 9.28  | 9.12  | 9.01  | 8.94  | 8.89  | 8.85  | 8.81  | 8.79  |
|        | 0.025    | 17.44  | 16.04 | 15.44 | 15.10 | 14.88 | 14.73 | 14.62 | 14.54 | 14.47 | 14.42 |
|        | 0.01     | 34.12  | 30.82 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 | 27.35 | 27.23 |
|        | 0.005    | 55.55  | 49.80 | 47.47 | 46.19 | 45.39 | 44.84 | 44.43 | 44.13 | 43.88 | 43.69 |
|        | 0.001    | 167.0  | 148.5 | 141.1 | 137.1 | 134.6 | 132.8 | 131.6 | 130.6 | 129.9 | 129.2 |

续表

| df <sub>2</sub> | $\alpha$ | df <sub>1</sub> |       |       |       |       |       |       |       |       |       |
|-----------------|----------|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                 |          | 1               | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
| 4               | 0.25     | 1.81            | 2.00  | 2.05  | 2.06  | 2.07  | 2.08  | 2.08  | 2.08  | 2.08  | 2.08  |
|                 | 0.10     | 4.54            | 4.32  | 4.19  | 4.11  | 4.05  | 4.01  | 3.98  | 3.95  | 3.94  | 3.92  |
|                 | 0.05     | 7.71            | 6.94  | 6.59  | 6.39  | 6.26  | 6.16  | 6.09  | 6.04  | 6.00  | 5.96  |
|                 | 0.025    | 12.22           | 10.65 | 9.98  | 9.60  | 9.36  | 9.20  | 9.07  | 8.98  | 8.90  | 8.84  |
|                 | 0.01     | 21.20           | 18.00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.98 | 14.80 | 14.66 | 14.55 |
| 5               | 0.005    | 31.33           | 26.28 | 24.26 | 23.15 | 22.46 | 21.97 | 21.62 | 21.35 | 21.14 | 20.97 |
|                 | 0.001    | 74.14           | 61.25 | 56.18 | 53.44 | 51.71 | 50.53 | 49.66 | 49.00 | 48.47 | 48.05 |
|                 | 0.25     | 1.69            | 1.85  | 1.88  | 1.89  | 1.89  | 1.89  | 1.89  | 1.89  | 1.89  | 1.89  |
|                 | 0.10     | 4.06            | 3.78  | 3.62  | 3.52  | 3.45  | 3.40  | 3.37  | 3.34  | 3.32  | 3.30  |
|                 | 0.05     | 6.61            | 5.79  | 5.41  | 5.19  | 5.05  | 4.95  | 4.88  | 4.82  | 4.77  | 4.74  |
| 6               | 0.025    | 10.01           | 8.43  | 7.76  | 7.39  | 7.15  | 6.98  | 6.85  | 6.76  | 6.68  | 6.62  |
|                 | 0.01     | 16.26           | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.46 | 10.29 | 10.16 | 10.05 |
|                 | 0.005    | 22.78           | 18.31 | 16.53 | 15.56 | 14.94 | 14.51 | 14.20 | 13.96 | 13.77 | 13.62 |
|                 | 0.001    | 47.18           | 37.12 | 33.20 | 31.09 | 29.75 | 28.83 | 28.16 | 27.65 | 27.24 | 26.92 |
|                 | 0.25     | 1.62            | 1.76  | 1.78  | 1.79  | 1.79  | 1.78  | 1.78  | 1.78  | 1.77  | 1.77  |
| 7               | 0.10     | 3.78            | 3.46  | 3.29  | 3.18  | 3.11  | 3.05  | 3.01  | 2.98  | 2.96  | 2.94  |
|                 | 0.05     | 5.99            | 5.14  | 4.76  | 4.53  | 4.39  | 4.28  | 4.21  | 4.15  | 4.10  | 4.06  |
|                 | 0.025    | 8.81            | 7.26  | 6.60  | 6.23  | 5.99  | 5.82  | 5.70  | 5.60  | 5.52  | 5.46  |
|                 | 0.01     | 13.75           | 10.92 | 9.78  | 9.15  | 8.76  | 8.47  | 8.26  | 8.10  | 7.98  | 7.87  |
|                 | 0.005    | 18.63           | 14.54 | 12.92 | 12.03 | 11.46 | 11.07 | 10.79 | 10.57 | 10.39 | 10.25 |
| 8               | 0.001    | 35.51           | 27.00 | 23.70 | 21.92 | 20.80 | 20.03 | 19.46 | 19.03 | 18.69 | 18.41 |

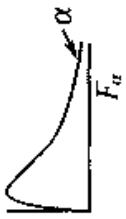
续表

| df <sub>1</sub> |       |       |       |       |       |       |       |       |       |       |                 | α | inf. | df <sub>2</sub> |
|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----------------|---|------|-----------------|
| 12              | 15    | 20    | 24    | 30    | 40    | 60    | 120   | 240   | inf.  | α     | df <sub>2</sub> |   |      |                 |
| 9.41            | 9.49  | 9.58  | 9.63  | 9.67  | 9.71  | 9.76  | 9.80  | 9.83  | 9.85  | 0.25  | 1               |   |      |                 |
| 60.71           | 61.22 | 61.74 | 62.00 | 62.26 | 62.53 | 62.79 | 63.06 | 63.19 | 63.33 | 0.10  |                 |   |      |                 |
| 243.9           | 245.9 | 248.0 | 249.1 | 250.1 | 251.1 | 252.2 | 253.3 | 253.8 | 254.3 | 0.05  |                 |   |      |                 |
| 976.7           | 984.9 | 993.1 | 997.2 | 1001  | 1006  | 1010  | 1014  | 1016  | 1018  | 0.025 |                 |   |      |                 |
| 6106            | 6157  | 6209  | 6235  | 6261  | 6287  | 6313  | 6339  | 6353  | 6366  | 0.01  |                 |   |      |                 |
| 3.39            | 3.41  | 3.43  | 3.43  | 3.44  | 3.45  | 3.46  | 3.47  | 3.47  | 3.48  | 0.25  | 2               |   |      |                 |
| 9.41            | 9.42  | 9.44  | 9.45  | 9.46  | 9.47  | 9.47  | 9.48  | 9.49  | 9.49  | 0.10  |                 |   |      |                 |
| 19.41           | 19.43 | 19.45 | 19.45 | 19.46 | 19.47 | 19.48 | 19.49 | 19.49 | 19.50 | 0.05  |                 |   |      |                 |
| 39.41           | 39.43 | 39.45 | 39.46 | 39.46 | 39.47 | 39.48 | 39.49 | 39.49 | 39.50 | 0.025 |                 |   |      |                 |
| 99.42           | 99.43 | 99.45 | 99.46 | 99.47 | 99.47 | 99.48 | 99.49 | 99.50 | 99.50 | 0.01  |                 |   |      |                 |
| 199.4           | 199.4 | 199.4 | 199.5 | 199.5 | 199.5 | 199.5 | 199.5 | 199.5 | 199.5 | 0.005 |                 |   |      |                 |
| 999.4           | 999.4 | 999.4 | 999.5 | 999.5 | 999.5 | 999.5 | 999.5 | 999.5 | 999.5 | 0.001 |                 |   |      |                 |
| 2.45            | 2.46  | 2.46  | 2.46  | 2.47  | 2.47  | 2.47  | 2.47  | 2.47  | 2.47  | 0.25  | 3               |   |      |                 |
| 5.22            | 5.20  | 5.18  | 5.18  | 5.17  | 5.16  | 5.15  | 5.14  | 5.14  | 5.13  | 0.10  |                 |   |      |                 |
| 8.74            | 8.70  | 8.66  | 8.64  | 8.62  | 8.59  | 8.57  | 8.55  | 8.54  | 8.53  | 0.05  |                 |   |      |                 |
| 14.34           | 14.25 | 14.17 | 14.12 | 14.08 | 14.04 | 13.99 | 13.95 | 13.92 | 13.90 | 0.025 |                 |   |      |                 |
| 27.05           | 26.87 | 26.69 | 26.60 | 26.50 | 26.41 | 26.32 | 26.22 | 26.17 | 26.13 | 0.01  |                 |   |      |                 |
| 43.39           | 43.08 | 42.78 | 42.62 | 42.47 | 42.31 | 42.15 | 41.99 | 41.91 | 41.83 | 0.005 |                 |   |      |                 |
| 128.3           | 127.4 | 126.4 | 125.9 | 125.4 | 125.0 | 124.5 | 124.0 | 123.7 | 123.5 | 0.001 |                 |   |      |                 |



续表

|       | df <sub>1</sub> |       |       |       |       |       |       |       |       |       | inf.  | $\alpha$ | df <sub>2</sub> |
|-------|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|-----------------|
|       | 12              | 15    | 20    | 24    | 30    | 40    | 60    | 120   | 240   |       |       |          |                 |
| 2.08  | 2.08            | 2.08  | 2.08  | 2.08  | 2.08  | 2.08  | 2.08  | 2.08  | 2.08  | 2.08  | 2.08  | 0.25     | 4               |
| 3.90  | 3.87            | 3.84  | 3.83  | 3.83  | 3.82  | 3.80  | 3.79  | 3.78  | 3.77  | 3.76  | 3.76  | 0.10     |                 |
| 5.91  | 5.86            | 5.80  | 5.77  | 5.77  | 5.75  | 5.72  | 5.69  | 5.66  | 5.64  | 5.63  | 5.63  | 0.05     |                 |
| 8.75  | 8.66            | 8.56  | 8.51  | 8.51  | 8.46  | 8.41  | 8.36  | 8.31  | 8.28  | 8.26  | 8.26  | 0.025    |                 |
| 14.37 | 14.20           | 14.02 | 13.93 | 13.93 | 13.84 | 13.75 | 13.65 | 13.56 | 13.51 | 13.46 | 13.46 | 0.01     |                 |
| 20.70 | 20.44           | 20.17 | 20.03 | 20.03 | 19.89 | 19.75 | 19.61 | 19.47 | 19.40 | 19.32 | 19.32 | 0.005    |                 |
| 47.41 | 46.76           | 46.10 | 45.77 | 45.77 | 45.43 | 45.09 | 44.75 | 44.40 | 44.23 | 44.05 | 44.05 | 0.001    |                 |
| 1.89  | 1.89            | 1.88  | 1.88  | 1.88  | 1.88  | 1.88  | 1.87  | 1.87  | 1.87  | 1.87  | 1.87  | 0.25     | 5               |
| 3.27  | 3.24            | 3.21  | 3.19  | 3.19  | 3.17  | 3.16  | 3.14  | 3.12  | 3.11  | 3.10  | 3.10  | 0.10     |                 |
| 4.68  | 4.62            | 4.56  | 4.53  | 4.53  | 4.50  | 4.46  | 4.43  | 4.40  | 4.38  | 4.36  | 4.36  | 0.05     |                 |
| 6.52  | 6.43            | 6.33  | 6.28  | 6.28  | 6.23  | 6.18  | 6.12  | 6.07  | 6.04  | 6.02  | 6.02  | 0.025    |                 |
| 9.89  | 9.72            | 9.55  | 9.47  | 9.47  | 9.38  | 9.29  | 9.20  | 9.11  | 9.07  | 9.02  | 9.02  | 0.01     |                 |
| 13.38 | 13.15           | 12.90 | 12.78 | 12.78 | 12.66 | 12.53 | 12.40 | 12.27 | 12.21 | 12.14 | 12.14 | 0.005    |                 |
| 26.42 | 25.91           | 25.39 | 25.13 | 25.13 | 24.87 | 24.60 | 24.33 | 24.06 | 23.92 | 23.79 | 23.79 | 0.001    |                 |
| 1.77  | 1.76            | 1.76  | 1.75  | 1.75  | 1.75  | 1.75  | 1.74  | 1.74  | 1.74  | 1.74  | 1.74  | 0.25     | 6               |
| 2.90  | 2.87            | 2.84  | 2.82  | 2.82  | 2.80  | 2.78  | 2.76  | 2.74  | 2.73  | 2.72  | 2.72  | 0.10     |                 |
| 4.00  | 3.94            | 3.87  | 3.84  | 3.84  | 3.81  | 3.77  | 3.74  | 3.70  | 3.69  | 3.67  | 3.67  | 0.05     |                 |
| 5.37  | 5.27            | 5.17  | 5.12  | 5.12  | 5.07  | 5.01  | 4.96  | 4.90  | 4.88  | 4.85  | 4.85  | 0.025    |                 |
| 7.72  | 7.56            | 7.40  | 7.31  | 7.31  | 7.23  | 7.14  | 7.06  | 6.97  | 6.92  | 6.88  | 6.88  | 0.01     |                 |
| 10.03 | 9.81            | 9.59  | 9.47  | 9.47  | 9.36  | 9.24  | 9.12  | 9.00  | 8.94  | 8.88  | 8.88  | 0.005    |                 |
| 17.99 | 17.56           | 17.12 | 16.90 | 16.90 | 16.67 | 16.44 | 16.21 | 15.98 | 15.86 | 15.75 | 15.75 | 0.001    |                 |

表 8 F 分布的百分位数 ( $df_2$  从 7 到 12)

| $df_2$ | $\alpha$ | $df_1$ |       |       |       |       |       |       |       |       |       |
|--------|----------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|        |          | 1      | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
| 7      | 0.25     | 1.57   | 1.70  | 1.72  | 1.72  | 1.71  | 1.71  | 1.70  | 1.70  | 1.69  | 1.69  |
|        | 0.10     | 3.59   | 3.26  | 3.07  | 2.96  | 2.88  | 2.83  | 2.78  | 2.75  | 2.72  | 2.70  |
|        | 0.05     | 5.59   | 4.74  | 4.35  | 4.12  | 3.97  | 3.87  | 3.79  | 3.73  | 3.68  | 3.64  |
|        | 0.025    | 8.07   | 6.54  | 5.89  | 5.52  | 5.29  | 5.12  | 4.99  | 4.90  | 4.82  | 4.76  |
|        | 0.01     | 12.25  | 9.55  | 8.45  | 7.85  | 7.46  | 7.19  | 6.99  | 6.84  | 6.72  | 6.62  |
|        | 0.005    | 16.24  | 12.40 | 10.88 | 10.05 | 9.52  | 9.16  | 8.89  | 8.68  | 8.51  | 8.38  |
| 8      | 0.001    | 29.25  | 21.69 | 18.77 | 17.20 | 16.21 | 15.52 | 15.02 | 14.63 | 14.33 | 14.08 |
|        | 0.25     | 1.54   | 1.66  | 1.67  | 1.66  | 1.66  | 1.65  | 1.64  | 1.64  | 1.63  | 1.63  |
|        | 0.10     | 3.46   | 3.11  | 2.92  | 2.81  | 2.73  | 2.67  | 2.62  | 2.59  | 2.56  | 2.54  |
|        | 0.05     | 5.32   | 4.46  | 4.07  | 3.84  | 3.69  | 3.58  | 3.50  | 3.44  | 3.39  | 3.35  |
|        | 0.025    | 7.57   | 6.06  | 5.42  | 5.05  | 4.82  | 4.65  | 4.53  | 4.43  | 4.36  | 4.30  |
|        | 0.01     | 11.26  | 8.65  | 7.59  | 7.01  | 6.63  | 6.37  | 6.18  | 6.03  | 5.91  | 5.81  |
| 9      | 0.005    | 14.69  | 11.04 | 9.60  | 8.81  | 8.30  | 7.95  | 7.69  | 7.50  | 7.34  | 7.21  |
|        | 0.001    | 25.41  | 18.49 | 15.83 | 14.39 | 13.48 | 12.86 | 12.40 | 12.05 | 11.77 | 11.54 |
|        | 0.25     | 1.51   | 1.62  | 1.63  | 1.63  | 1.62  | 1.61  | 1.60  | 1.60  | 1.59  | 1.59  |
|        | 0.10     | 3.36   | 3.01  | 2.81  | 2.69  | 2.61  | 2.55  | 2.51  | 2.47  | 2.44  | 2.42  |
|        | 0.05     | 5.12   | 4.26  | 3.86  | 3.63  | 3.48  | 3.37  | 3.29  | 3.23  | 3.18  | 3.14  |
|        | 0.025    | 7.21   | 5.71  | 5.08  | 4.72  | 4.48  | 4.32  | 4.20  | 4.10  | 4.03  | 3.96  |
| 10     | 0.01     | 10.56  | 8.02  | 6.99  | 6.42  | 6.06  | 5.80  | 5.61  | 5.47  | 5.35  | 5.26  |
|        | 0.005    | 13.61  | 10.11 | 8.72  | 7.96  | 7.47  | 7.13  | 6.88  | 6.69  | 6.54  | 6.42  |
|        | 0.001    | 22.86  | 16.39 | 13.90 | 12.56 | 11.71 | 11.13 | 10.70 | 10.37 | 10.11 | 9.89  |

续表

| df <sub>2</sub> | $\alpha$ | df <sub>1</sub> |       |       |       |       |      |      |      |      |      |
|-----------------|----------|-----------------|-------|-------|-------|-------|------|------|------|------|------|
|                 |          | 1               | 2     | 3     | 4     | 5     | 6    | 7    | 8    | 9    | 10   |
| 10              | 0.25     | 1.49            | 1.60  | 1.60  | 1.59  | 1.59  | 1.58 | 1.57 | 1.56 | 1.56 | 1.55 |
|                 | 0.10     | 3.29            | 2.92  | 2.73  | 2.61  | 2.52  | 2.46 | 2.41 | 2.38 | 2.35 | 2.32 |
|                 | 0.05     | 4.96            | 4.10  | 3.71  | 3.48  | 3.33  | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 |
|                 | 0.025    | 6.94            | 5.46  | 4.83  | 4.47  | 4.24  | 4.07 | 3.95 | 3.85 | 3.78 | 3.72 |
|                 | 0.01     | 10.04           | 7.56  | 6.55  | 5.99  | 5.64  | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 |
| 11              | 0.005    | 12.83           | 9.43  | 8.08  | 7.34  | 6.87  | 6.54 | 6.30 | 6.12 | 5.97 | 5.85 |
|                 | 0.001    | 21.04           | 14.91 | 12.55 | 11.28 | 10.48 | 9.93 | 9.52 | 9.20 | 8.96 | 8.75 |
|                 | 0.25     | 1.47            | 1.58  | 1.58  | 1.57  | 1.56  | 1.55 | 1.54 | 1.53 | 1.53 | 1.52 |
|                 | 0.10     | 3.23            | 2.86  | 2.66  | 2.54  | 2.45  | 2.39 | 2.34 | 2.30 | 2.27 | 2.25 |
|                 | 0.05     | 4.84            | 3.98  | 3.59  | 3.36  | 3.20  | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 |
| 12              | 0.025    | 6.72            | 5.26  | 4.63  | 4.28  | 4.04  | 3.88 | 3.76 | 3.66 | 3.59 | 3.53 |
|                 | 0.01     | 9.65            | 7.21  | 6.22  | 5.67  | 5.32  | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 |
|                 | 0.005    | 12.23           | 8.91  | 7.60  | 6.88  | 6.42  | 6.10 | 5.86 | 5.68 | 5.54 | 5.42 |
|                 | 0.001    | 19.69           | 13.81 | 11.56 | 10.35 | 9.58  | 9.05 | 8.66 | 8.35 | 8.12 | 7.92 |
|                 | 0.25     | 1.46            | 1.56  | 1.56  | 1.55  | 1.54  | 1.53 | 1.52 | 1.51 | 1.51 | 1.50 |
| 12              | 0.10     | 3.18            | 2.81  | 2.61  | 2.48  | 2.39  | 2.33 | 2.28 | 2.24 | 2.21 | 2.19 |
|                 | 0.05     | 4.75            | 3.89  | 3.49  | 3.26  | 3.11  | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 |
|                 | 0.025    | 6.55            | 5.10  | 4.47  | 4.12  | 3.89  | 3.73 | 3.61 | 3.51 | 3.44 | 3.37 |
|                 | 0.01     | 9.33            | 6.93  | 5.95  | 5.41  | 5.06  | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 |
|                 | 0.005    | 11.75           | 8.51  | 7.23  | 6.52  | 6.07  | 5.76 | 5.52 | 5.35 | 5.20 | 5.09 |
|                 | 0.001    | 18.64           | 12.97 | 10.80 | 9.63  | 8.89  | 8.38 | 8.00 | 7.71 | 7.48 | 7.29 |

续表

| df <sub>1</sub> |       |       |       |       |       |       |       |       |       | inf.  | $\alpha$ | df <sub>2</sub> |
|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|-----------------|
| 12              | 15    | 20    | 24    | 30    | 40    | 60    | 120   | 240   | inf.  |       |          |                 |
| 1.68            | 1.68  | 1.67  | 1.67  | 1.66  | 1.66  | 1.65  | 1.65  | 1.65  | 1.65  | 0.25  | 0.25     | 7               |
| 2.67            | 2.63  | 2.59  | 2.58  | 2.56  | 2.54  | 2.51  | 2.49  | 2.48  | 2.47  | 0.10  | 0.10     |                 |
| 3.57            | 3.51  | 3.44  | 3.41  | 3.38  | 3.34  | 3.30  | 3.27  | 3.25  | 3.23  | 0.05  | 0.05     |                 |
| 4.67            | 4.57  | 4.47  | 4.41  | 4.36  | 4.31  | 4.25  | 4.20  | 4.17  | 4.14  | 0.025 | 0.025    |                 |
| 6.47            | 6.31  | 6.16  | 6.07  | 5.99  | 5.91  | 5.82  | 5.74  | 5.69  | 5.65  | 0.01  | 0.01     |                 |
| 8.18            | 7.97  | 7.75  | 7.64  | 7.53  | 7.42  | 7.31  | 7.19  | 7.13  | 7.08  | 0.005 | 0.005    |                 |
| 13.71           | 13.32 | 12.93 | 12.73 | 12.53 | 12.33 | 12.12 | 11.91 | 11.80 | 11.70 | 0.001 | 0.001    |                 |
| 1.62            | 1.62  | 1.61  | 1.60  | 1.60  | 1.59  | 1.59  | 1.58  | 1.58  | 1.58  | 0.25  | 0.25     | 8               |
| 2.50            | 2.46  | 2.42  | 2.40  | 2.38  | 2.36  | 2.34  | 2.32  | 2.30  | 2.29  | 0.10  | 0.10     |                 |
| 3.28            | 3.22  | 3.15  | 3.12  | 3.08  | 3.04  | 3.01  | 2.97  | 2.95  | 2.93  | 0.05  | 0.05     |                 |
| 4.20            | 4.10  | 4.00  | 3.95  | 3.89  | 3.84  | 3.78  | 3.73  | 3.70  | 3.67  | 0.025 | 0.025    |                 |
| 5.67            | 5.52  | 5.36  | 5.28  | 5.20  | 5.12  | 5.03  | 4.95  | 4.90  | 4.86  | 0.01  | 0.01     |                 |
| 7.01            | 6.81  | 6.61  | 6.50  | 6.40  | 6.29  | 6.18  | 6.06  | 6.01  | 5.95  | 0.005 | 0.005    |                 |
| 11.19           | 10.84 | 10.48 | 10.30 | 10.11 | 9.92  | 9.73  | 9.53  | 9.43  | 9.33  | 0.001 | 0.001    |                 |
| 1.58            | 1.57  | 1.56  | 1.56  | 1.55  | 1.54  | 1.64  | 1.53  | 1.53  | 1.53  | 0.25  | 0.25     | 9               |
| 2.38            | 2.34  | 2.30  | 2.28  | 2.25  | 2.23  | 2.21  | 2.18  | 2.17  | 2.16  | 0.10  | 0.10     |                 |
| 3.07            | 3.01  | 2.94  | 2.90  | 2.86  | 2.83  | 2.79  | 2.75  | 2.73  | 2.71  | 0.05  | 0.05     |                 |
| 3.87            | 3.77  | 3.67  | 3.61  | 3.56  | 3.51  | 3.45  | 3.39  | 3.36  | 3.33  | 0.025 | 0.025    |                 |
| 5.11            | 4.96  | 4.81  | 4.73  | 4.65  | 4.57  | 4.48  | 4.40  | 4.35  | 4.31  | 0.01  | 0.01     |                 |
| 6.23            | 6.03  | 5.83  | 5.73  | 5.62  | 5.52  | 5.41  | 5.30  | 5.24  | 5.19  | 0.005 | 0.005    |                 |
| 9.57            | 9.24  | 8.90  | 8.72  | 8.55  | 8.37  | 8.19  | 8.00  | 7.91  | 7.81  | 0.001 | 0.001    |                 |

续表

| df <sub>1</sub> |      |      |      |      |      |      |      |      |      |       |    | inf. | $\alpha$ | df <sub>2</sub> |
|-----------------|------|------|------|------|------|------|------|------|------|-------|----|------|----------|-----------------|
| 12              | 15   | 20   | 24   | 30   | 40   | 60   | 120  | 240  |      |       |    |      |          |                 |
| 1.54            | 1.53 | 1.52 | 1.52 | 1.51 | 1.51 | 1.50 | 1.49 | 1.49 | 1.48 | 0.25  | 10 |      |          |                 |
| 2.28            | 2.24 | 2.20 | 2.18 | 2.16 | 2.13 | 2.11 | 2.08 | 2.07 | 2.06 | 0.10  |    |      |          |                 |
| 2.91            | 2.85 | 2.77 | 2.74 | 2.70 | 2.66 | 2.62 | 2.58 | 2.56 | 2.54 | 0.05  |    |      |          |                 |
| 3.62            | 3.52 | 3.42 | 3.37 | 3.31 | 3.26 | 3.20 | 3.14 | 3.11 | 3.08 | 0.025 |    |      |          |                 |
| 4.71            | 4.56 | 4.41 | 4.33 | 4.25 | 4.17 | 4.08 | 4.00 | 3.95 | 3.91 | 0.01  |    |      |          |                 |
| 5.66            | 5.47 | 5.27 | 5.17 | 5.07 | 4.97 | 4.86 | 4.75 | 4.69 | 4.64 | 0.005 |    |      |          |                 |
| 8.45            | 8.13 | 7.80 | 7.64 | 7.47 | 7.30 | 7.12 | 6.94 | 6.85 | 6.76 | 0.001 |    |      |          |                 |
| 1.51            | 1.50 | 1.49 | 1.49 | 1.48 | 1.47 | 1.47 | 1.46 | 1.45 | 1.45 | 0.25  | 11 |      |          |                 |
| 2.21            | 2.17 | 2.12 | 2.10 | 2.08 | 2.05 | 2.03 | 2.00 | 1.99 | 1.97 | 0.10  |    |      |          |                 |
| 2.79            | 2.72 | 2.65 | 2.61 | 2.57 | 2.53 | 2.49 | 2.45 | 2.43 | 2.40 | 0.05  |    |      |          |                 |
| 3.43            | 3.33 | 3.23 | 3.17 | 3.12 | 3.06 | 3.00 | 2.94 | 2.91 | 2.88 | 0.025 |    |      |          |                 |
| 4.40            | 4.25 | 4.10 | 4.02 | 3.94 | 3.86 | 3.78 | 3.69 | 3.65 | 3.60 | 0.01  |    |      |          |                 |
| 5.24            | 5.05 | 4.86 | 4.76 | 4.65 | 4.55 | 4.45 | 4.34 | 4.28 | 4.23 | 0.005 |    |      |          |                 |
| 7.63            | 7.32 | 7.01 | 6.85 | 6.68 | 6.52 | 6.35 | 6.18 | 6.09 | 6.00 | 0.001 |    |      |          |                 |
| 1.49            | 1.48 | 1.47 | 1.46 | 1.45 | 1.45 | 1.44 | 1.43 | 1.43 | 1.42 | 0.25  | 12 |      |          |                 |
| 2.15            | 2.10 | 2.06 | 2.04 | 2.01 | 1.99 | 1.96 | 1.93 | 1.92 | 1.90 | 0.10  |    |      |          |                 |
| 2.69            | 2.62 | 2.54 | 2.51 | 2.47 | 2.43 | 2.38 | 2.34 | 2.32 | 2.30 | 0.05  |    |      |          |                 |
| 3.28            | 3.18 | 3.07 | 3.02 | 2.96 | 2.91 | 2.85 | 2.79 | 2.76 | 2.72 | 0.025 |    |      |          |                 |
| 4.16            | 4.01 | 3.86 | 3.78 | 3.70 | 3.62 | 3.54 | 3.45 | 3.41 | 3.36 | 0.01  |    |      |          |                 |
| 4.91            | 4.72 | 4.53 | 4.43 | 4.33 | 4.23 | 4.12 | 4.01 | 3.96 | 3.90 | 0.005 |    |      |          |                 |
| 7.00            | 6.71 | 6.40 | 6.25 | 6.09 | 5.93 | 5.76 | 5.59 | 5.51 | 5.42 | 0.001 |    |      |          |                 |

表 8 F 分布的百分位数( $df_2$  从 13 到 18)

| $df_2$ | $\alpha$ | $df_1$ |       |       |      |      |      |      |      |      |      |
|--------|----------|--------|-------|-------|------|------|------|------|------|------|------|
|        |          | 1      | 2     | 3     | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| 13     | 0.25     | 1.45   | 1.55  | 1.55  | 1.53 | 1.52 | 1.51 | 1.50 | 1.49 | 1.49 | 1.48 |
|        | 0.10     | 3.14   | 2.76  | 2.56  | 2.43 | 2.35 | 2.28 | 2.23 | 2.20 | 2.16 | 2.14 |
|        | 0.05     | 4.67   | 3.81  | 3.41  | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 |
|        | 0.025    | 6.41   | 4.97  | 4.35  | 4.00 | 3.77 | 3.60 | 3.48 | 3.39 | 3.31 | 3.25 |
|        | 0.01     | 9.07   | 6.70  | 5.74  | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 | 4.19 | 4.10 |
| 14     | 0.005    | 11.37  | 8.19  | 6.93  | 6.23 | 5.79 | 5.48 | 5.25 | 5.08 | 4.94 | 4.82 |
|        | 0.001    | 17.82  | 12.31 | 10.21 | 9.07 | 8.35 | 7.86 | 7.49 | 7.21 | 6.98 | 6.80 |
|        | 0.25     | 1.44   | 1.53  | 1.53  | 1.52 | 1.51 | 1.50 | 1.49 | 1.48 | 1.47 | 1.46 |
|        | 0.10     | 3.10   | 2.73  | 2.52  | 2.39 | 2.31 | 2.24 | 2.19 | 2.15 | 2.12 | 2.10 |
|        | 0.05     | 4.60   | 3.74  | 3.34  | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 |
| 15     | 0.025    | 6.30   | 4.86  | 4.24  | 3.89 | 3.66 | 3.50 | 3.38 | 3.29 | 3.21 | 3.15 |
|        | 0.01     | 8.86   | 6.51  | 5.56  | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 | 4.03 | 3.94 |
|        | 0.005    | 11.06  | 7.92  | 6.68  | 6.00 | 5.56 | 5.26 | 5.03 | 4.86 | 4.72 | 4.60 |
|        | 0.001    | 17.14  | 11.78 | 9.73  | 8.62 | 7.92 | 7.44 | 7.08 | 6.80 | 6.58 | 6.40 |
|        | 0.25     | 1.43   | 1.52  | 1.52  | 1.51 | 1.49 | 1.48 | 1.47 | 1.46 | 1.46 | 1.45 |
|        | 0.10     | 3.07   | 2.70  | 2.49  | 2.36 | 2.27 | 2.21 | 2.16 | 2.12 | 2.09 | 2.06 |
|        | 0.05     | 4.54   | 3.68  | 3.29  | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 |
|        | 0.025    | 6.20   | 4.77  | 4.15  | 3.80 | 3.58 | 3.41 | 3.29 | 3.20 | 3.12 | 3.06 |
|        | 0.01     | 8.68   | 6.36  | 5.42  | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 |
|        | 0.005    | 10.80  | 7.70  | 6.48  | 5.80 | 5.37 | 5.07 | 4.85 | 4.67 | 4.54 | 4.42 |
|        | 0.001    | 16.59  | 11.34 | 9.34  | 8.25 | 7.57 | 7.09 | 6.74 | 6.47 | 6.26 | 6.08 |

续表

| df <sub>2</sub> | $\alpha$ | df <sub>1</sub> |       |      |      |      |      |      |      |      |      |
|-----------------|----------|-----------------|-------|------|------|------|------|------|------|------|------|
|                 |          | 1               | 2     | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| 16              | 0.25     | 1.42            | 1.51  | 1.51 | 1.50 | 1.48 | 1.47 | 1.46 | 1.45 | 1.44 | 1.44 |
|                 | 0.10     | 3.05            | 2.67  | 2.46 | 2.33 | 2.24 | 2.18 | 2.13 | 2.09 | 2.06 | 2.03 |
|                 | 0.05     | 4.49            | 3.63  | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 |
|                 | 0.025    | 6.12            | 4.69  | 4.08 | 3.73 | 3.50 | 3.34 | 3.22 | 3.12 | 3.05 | 2.99 |
|                 | 0.01     | 8.53            | 6.23  | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 | 3.69 |
| 17              | 0.005    | 10.58           | 7.51  | 6.30 | 5.64 | 5.21 | 4.91 | 4.69 | 4.52 | 4.38 | 4.27 |
|                 | 0.001    | 16.12           | 10.97 | 9.01 | 7.94 | 7.27 | 6.80 | 6.46 | 6.19 | 5.98 | 5.81 |
|                 | 0.25     | 1.42            | 1.51  | 1.50 | 1.49 | 1.47 | 1.46 | 1.45 | 1.44 | 1.43 | 1.43 |
|                 | 0.10     | 3.03            | 2.64  | 2.44 | 2.31 | 2.22 | 2.15 | 2.10 | 2.06 | 2.03 | 2.00 |
|                 | 0.05     | 4.45            | 3.59  | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 |
| 18              | 0.025    | 6.04            | 4.62  | 4.01 | 3.66 | 3.44 | 3.28 | 3.16 | 3.06 | 2.98 | 2.92 |
|                 | 0.01     | 8.40            | 6.11  | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 | 3.59 |
|                 | 0.005    | 10.38           | 7.35  | 6.16 | 5.50 | 5.07 | 4.78 | 4.56 | 4.39 | 4.25 | 4.14 |
|                 | 0.001    | 15.72           | 10.66 | 8.73 | 7.68 | 7.02 | 6.56 | 6.22 | 5.96 | 5.75 | 5.58 |
|                 | 0.25     | 1.41            | 1.50  | 1.49 | 1.48 | 1.46 | 1.45 | 1.44 | 1.43 | 1.42 | 1.42 |
|                 | 0.10     | 3.01            | 2.62  | 2.42 | 2.29 | 2.20 | 2.13 | 2.08 | 2.04 | 2.00 | 1.98 |
|                 | 0.05     | 4.41            | 3.55  | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 |
|                 | 0.025    | 5.98            | 4.56  | 3.95 | 3.61 | 3.38 | 3.22 | 3.10 | 3.01 | 2.93 | 2.87 |
|                 | 0.01     | 8.29            | 6.01  | 5.09 | 4.58 | 4.25 | 4.01 | 3.84 | 3.71 | 3.60 | 3.51 |
|                 | 0.005    | 10.22           | 7.21  | 6.03 | 5.37 | 4.96 | 4.66 | 4.44 | 4.28 | 4.14 | 4.03 |
|                 | 0.001    | 15.38           | 10.39 | 8.49 | 7.46 | 6.81 | 6.35 | 6.02 | 5.76 | 5.56 | 5.39 |

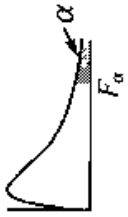
续表

| df <sub>1</sub> |      |      |      |      |      |      |      |      |      |          |                 |  |
|-----------------|------|------|------|------|------|------|------|------|------|----------|-----------------|--|
| 12              | 15   | 20   | 24   | 30   | 40   | 60   | 120  | 240  | inf. | $\alpha$ | df <sub>2</sub> |  |
| 1.47            | 1.46 | 1.45 | 1.44 | 1.43 | 1.42 | 1.42 | 1.41 | 1.40 | 1.40 | 0.25     | 13              |  |
| 2.10            | 2.05 | 2.01 | 1.98 | 1.96 | 1.93 | 1.90 | 1.88 | 1.86 | 1.85 | 0.10     |                 |  |
| 2.60            | 2.53 | 2.46 | 2.42 | 2.38 | 2.34 | 2.30 | 2.25 | 2.23 | 2.21 | 0.05     |                 |  |
| 3.15            | 3.05 | 2.95 | 2.89 | 2.84 | 2.78 | 2.72 | 2.66 | 2.63 | 2.60 | 0.025    |                 |  |
| 3.96            | 3.82 | 3.66 | 3.59 | 3.51 | 3.43 | 3.34 | 3.25 | 3.21 | 3.17 | 0.01     |                 |  |
| 4.64            | 4.46 | 4.27 | 4.17 | 4.07 | 3.97 | 3.87 | 3.76 | 3.70 | 3.65 | 0.005    |                 |  |
| 6.52            | 6.23 | 5.93 | 5.78 | 5.63 | 5.47 | 5.30 | 5.14 | 5.05 | 4.97 | 0.001    |                 |  |
| 1.45            | 1.44 | 1.43 | 1.42 | 1.41 | 1.41 | 1.40 | 1.39 | 1.38 | 1.38 | 0.25     | 14              |  |
| 2.05            | 2.01 | 1.96 | 1.94 | 1.91 | 1.89 | 1.86 | 1.83 | 1.81 | 1.80 | 0.10     |                 |  |
| 2.53            | 2.46 | 2.39 | 2.35 | 2.31 | 2.27 | 2.22 | 2.18 | 2.15 | 2.13 | 0.05     |                 |  |
| 3.05            | 2.95 | 2.84 | 2.79 | 2.73 | 2.67 | 2.61 | 2.55 | 2.52 | 2.49 | 0.025    |                 |  |
| 3.80            | 3.66 | 3.51 | 3.43 | 3.35 | 3.27 | 3.18 | 3.09 | 3.05 | 3.00 | 0.01     |                 |  |
| 4.43            | 4.25 | 4.06 | 3.96 | 3.86 | 3.76 | 3.66 | 3.55 | 3.49 | 3.44 | 0.005    |                 |  |
| 6.13            | 5.85 | 5.56 | 5.41 | 5.25 | 5.10 | 4.94 | 4.77 | 4.69 | 4.60 | 0.001    |                 |  |
| 1.44            | 1.43 | 1.41 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.36 | 0.25     | 15              |  |
| 2.02            | 1.97 | 1.92 | 1.90 | 1.87 | 1.85 | 1.82 | 1.79 | 1.77 | 1.76 | 0.10     |                 |  |
| 2.48            | 2.40 | 2.33 | 2.29 | 2.25 | 2.20 | 2.16 | 2.11 | 2.09 | 2.07 | 0.05     |                 |  |
| 2.96            | 2.86 | 2.76 | 2.70 | 2.64 | 2.59 | 2.52 | 2.46 | 2.43 | 2.40 | 0.025    |                 |  |
| 3.67            | 3.52 | 3.37 | 3.29 | 3.21 | 3.13 | 3.05 | 2.96 | 2.91 | 2.87 | 0.01     |                 |  |
| 4.25            | 4.07 | 3.88 | 3.79 | 3.69 | 3.58 | 3.48 | 3.37 | 3.32 | 3.26 | 0.005    |                 |  |
| 5.81            | 5.54 | 5.25 | 5.10 | 4.95 | 4.80 | 4.64 | 4.47 | 4.39 | 4.31 | 0.001    |                 |  |



续表

| df <sub>1</sub> |      |      |      |      |      |      |      |      |      |       |    | inf. | $\alpha$ | df <sub>2</sub> |
|-----------------|------|------|------|------|------|------|------|------|------|-------|----|------|----------|-----------------|
| 12              | 15   | 20   | 24   | 30   | 40   | 60   | 120  | 240  |      |       |    |      |          |                 |
| 1.43            | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.35 | 1.34 | 0.25  | 16 |      |          |                 |
| 1.99            | 1.94 | 1.89 | 1.87 | 1.84 | 1.81 | 1.78 | 1.75 | 1.73 | 1.72 | 0.10  |    |      |          |                 |
| 2.42            | 2.35 | 2.28 | 2.24 | 2.19 | 2.15 | 2.11 | 2.06 | 2.03 | 2.01 | 0.05  |    |      |          |                 |
| 2.89            | 2.79 | 2.68 | 2.63 | 2.57 | 2.51 | 2.45 | 2.38 | 2.35 | 2.32 | 0.025 |    |      |          |                 |
| 3.55            | 3.41 | 3.26 | 3.18 | 3.10 | 3.02 | 2.93 | 2.84 | 2.80 | 2.75 | 0.01  |    |      |          |                 |
| 4.10            | 3.92 | 3.73 | 3.64 | 3.54 | 3.44 | 3.33 | 3.22 | 3.17 | 3.11 | 0.005 |    |      |          |                 |
| 5.55            | 5.27 | 4.99 | 4.85 | 4.70 | 4.54 | 4.39 | 4.23 | 4.14 | 4.06 | 0.001 |    |      |          |                 |
| 1.41            | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.33 | 0.25  | 17 |      |          |                 |
| 1.96            | 1.91 | 1.86 | 1.84 | 1.81 | 1.78 | 1.75 | 1.72 | 1.70 | 1.69 | 0.10  |    |      |          |                 |
| 2.38            | 2.31 | 2.23 | 2.19 | 2.15 | 2.10 | 2.06 | 2.01 | 1.99 | 1.96 | 0.0   |    |      |          |                 |
| 2.82            | 2.72 | 2.62 | 2.56 | 2.50 | 2.44 | 2.38 | 2.32 | 2.28 | 2.25 | 0.025 |    |      |          |                 |
| 3.46            | 3.31 | 3.16 | 3.08 | 3.00 | 2.92 | 2.83 | 2.75 | 2.70 | 2.65 | 0.01  |    |      |          |                 |
| 3.97            | 3.79 | 3.61 | 3.51 | 3.41 | 3.31 | 3.21 | 3.10 | 3.04 | 2.98 | 0.005 |    |      |          |                 |
| 5.32            | 5.05 | 4.78 | 4.63 | 4.48 | 4.33 | 4.18 | 4.02 | 3.93 | 3.85 | 0.001 |    |      |          |                 |
| 1.40            | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.32 | 1.32 | 0.25  | 18 |      |          |                 |
| 1.93            | 1.89 | 1.84 | 1.81 | 1.78 | 1.75 | 1.72 | 1.69 | 1.67 | 1.66 | 0.10  |    |      |          |                 |
| 2.34            | 2.27 | 2.19 | 2.15 | 2.11 | 2.06 | 2.02 | 1.97 | 1.94 | 1.92 | 0.05  |    |      |          |                 |
| 2.77            | 2.67 | 2.56 | 2.50 | 2.44 | 2.38 | 2.32 | 2.26 | 2.22 | 2.19 | 0.025 |    |      |          |                 |
| 3.37            | 3.23 | 3.08 | 3.00 | 2.92 | 2.84 | 2.75 | 2.66 | 2.61 | 2.57 | 0.01  |    |      |          |                 |
| 3.86            | 3.68 | 3.50 | 3.40 | 3.30 | 3.20 | 3.10 | 2.99 | 2.93 | 2.87 | 0.005 |    |      |          |                 |
| 5.13            | 4.87 | 4.59 | 4.45 | 4.30 | 4.15 | 4.00 | 3.84 | 3.75 | 3.67 | 0.001 |    |      |          |                 |

表 8 F 分布的百分位数( $df_2$  从 19 到 24)

| $df_2$ | $\alpha$ | $df_1$ |       |      |      |      |      |      |      |      |      |
|--------|----------|--------|-------|------|------|------|------|------|------|------|------|
|        |          | 1      | 2     | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| 19     | 0.25     | 1.41   | 1.49  | 1.49 | 1.47 | 1.46 | 1.44 | 1.43 | 1.42 | 1.41 | 1.41 |
|        | 0.10     | 2.99   | 2.61  | 2.40 | 2.27 | 2.18 | 2.11 | 2.06 | 2.02 | 1.98 | 1.96 |
|        | 0.05     | 4.38   | 3.52  | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 |
|        | 0.025    | 5.92   | 4.51  | 3.90 | 3.56 | 3.33 | 3.17 | 3.05 | 2.96 | 2.88 | 2.82 |
|        | 0.01     | 8.18   | 5.93  | 5.01 | 4.50 | 4.17 | 3.94 | 3.77 | 3.63 | 3.52 | 3.43 |
| 20     | 0.005    | 10.07  | 7.09  | 5.92 | 5.27 | 4.85 | 4.56 | 4.34 | 4.18 | 4.04 | 3.93 |
|        | 0.001    | 15.08  | 10.16 | 8.28 | 7.27 | 6.62 | 6.18 | 5.85 | 5.59 | 5.39 | 5.22 |
|        | 0.25     | 1.40   | 1.49  | 1.48 | 1.47 | 1.45 | 1.44 | 1.43 | 1.42 | 1.41 | 1.40 |
|        | 0.10     | 2.97   | 2.59  | 2.38 | 2.25 | 2.16 | 2.09 | 2.04 | 2.00 | 1.96 | 1.94 |
|        | 0.05     | 4.35   | 3.49  | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 |
| 21     | 0.025    | 5.87   | 4.46  | 3.86 | 3.51 | 3.29 | 3.13 | 3.01 | 2.91 | 2.84 | 2.77 |
|        | 0.01     | 8.10   | 5.85  | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 | 3.37 |
|        | 0.005    | 9.94   | 6.99  | 5.82 | 5.17 | 4.76 | 4.47 | 4.26 | 4.09 | 3.96 | 3.85 |
|        | 0.001    | 14.82  | 9.95  | 8.10 | 7.10 | 6.46 | 6.02 | 5.69 | 5.44 | 5.24 | 5.08 |
|        | 0.25     | 1.40   | 1.48  | 1.48 | 1.46 | 1.44 | 1.43 | 1.42 | 1.41 | 1.40 | 1.39 |
|        | 0.10     | 2.96   | 2.57  | 2.36 | 2.23 | 2.14 | 2.08 | 2.02 | 1.98 | 1.95 | 1.92 |
|        | 0.05     | 4.32   | 3.47  | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 | 2.42 | 2.37 | 2.32 |
|        | 0.025    | 5.83   | 4.42  | 3.82 | 3.48 | 3.25 | 3.09 | 2.97 | 2.87 | 2.80 | 2.73 |
|        | 0.01     | 8.02   | 5.78  | 4.87 | 4.37 | 4.04 | 3.81 | 3.64 | 3.51 | 3.40 | 3.31 |
|        | 0.005    | 9.83   | 6.89  | 5.73 | 5.09 | 4.68 | 4.39 | 4.18 | 4.01 | 3.88 | 3.77 |
|        | 0.001    | 14.59  | 9.77  | 7.94 | 6.95 | 6.32 | 5.88 | 5.56 | 5.31 | 5.11 | 4.95 |

续表

| df <sub>2</sub> | $\alpha$ | df <sub>1</sub> |      |      |      |      |      |      |      |      |      |
|-----------------|----------|-----------------|------|------|------|------|------|------|------|------|------|
|                 |          | 1               | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| 22              | 0.25     | 1.40            | 1.48 | 1.47 | 1.45 | 1.44 | 1.42 | 1.41 | 1.40 | 1.39 | 1.39 |
|                 | 0.10     | 2.95            | 2.56 | 2.35 | 2.22 | 2.13 | 2.06 | 2.01 | 1.97 | 1.93 | 1.90 |
|                 | 0.05     | 4.30            | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 |
|                 | 0.025    | 5.79            | 4.38 | 3.78 | 3.44 | 3.22 | 3.05 | 2.93 | 2.84 | 2.76 | 2.70 |
|                 | 0.01     | 7.95            | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.59 | 3.45 | 3.35 | 3.26 |
| 23              | 0.005    | 9.73            | 6.81 | 5.65 | 5.02 | 4.61 | 4.32 | 4.11 | 3.94 | 3.81 | 3.70 |
|                 | 0.001    | 14.38           | 9.61 | 7.80 | 6.81 | 6.19 | 5.76 | 5.44 | 5.19 | 4.99 | 4.83 |
|                 | 0.25     | 1.39            | 1.47 | 1.47 | 1.45 | 1.43 | 1.42 | 1.41 | 1.40 | 1.39 | 1.38 |
|                 | 0.10     | 2.94            | 2.55 | 2.34 | 2.21 | 2.11 | 2.05 | 1.99 | 1.95 | 1.92 | 1.89 |
|                 | 0.05     | 4.28            | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 | 2.37 | 2.32 | 2.27 |
| 24              | 0.025    | 5.75            | 4.35 | 3.75 | 3.41 | 3.18 | 3.02 | 2.90 | 2.81 | 2.73 | 2.67 |
|                 | 0.01     | 7.88            | 5.66 | 4.76 | 4.26 | 3.94 | 3.71 | 3.54 | 3.41 | 3.30 | 3.21 |
|                 | 0.005    | 9.63            | 6.73 | 5.58 | 4.95 | 4.54 | 4.26 | 4.05 | 3.88 | 3.75 | 3.64 |
|                 | 0.001    | 14.20           | 9.47 | 7.67 | 6.70 | 6.08 | 5.65 | 5.33 | 5.09 | 4.89 | 4.73 |
|                 | 0.25     | 1.39            | 1.47 | 1.46 | 1.44 | 1.43 | 1.41 | 1.40 | 1.39 | 1.38 | 1.38 |
| 24              | 0.10     | 2.93            | 2.54 | 2.33 | 2.19 | 2.10 | 2.04 | 1.98 | 1.94 | 1.91 | 1.88 |
|                 | 0.05     | 4.26            | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 |
|                 | 0.025    | 5.72            | 4.32 | 3.72 | 3.38 | 3.15 | 2.99 | 2.87 | 2.78 | 2.70 | 2.64 |
|                 | 0.01     | 7.82            | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 | 3.17 |
|                 | 0.005    | 9.55            | 6.66 | 5.52 | 4.89 | 4.49 | 4.20 | 3.99 | 3.83 | 3.69 | 3.59 |
|                 | 0.001    | 14.03           | 9.34 | 7.55 | 6.59 | 5.98 | 5.55 | 5.23 | 4.99 | 4.80 | 4.64 |

续表

| df <sub>1</sub> |      |      |      |      |      |      |      |      |      |          | df <sub>2</sub> |
|-----------------|------|------|------|------|------|------|------|------|------|----------|-----------------|
| 12              | 15   | 20   | 24   | 30   | 40   | 60   | 120  | 240  | inf. | $\alpha$ |                 |
| 1.40            | 1.38 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.32 | 1.31 | 1.30 | 0.25     | 19              |
| 1.91            | 1.86 | 1.81 | 1.79 | 1.76 | 1.73 | 1.70 | 1.67 | 1.65 | 1.63 | 0.10     |                 |
| 2.31            | 2.23 | 2.16 | 2.11 | 2.07 | 2.03 | 1.98 | 1.93 | 1.90 | 1.88 | 0.05     |                 |
| 2.72            | 2.62 | 2.51 | 2.45 | 2.39 | 2.33 | 2.27 | 2.20 | 2.17 | 2.13 | 0.025    |                 |
| 3.30            | 3.15 | 3.00 | 2.92 | 2.84 | 2.76 | 2.67 | 2.58 | 2.54 | 2.49 | 0.01     |                 |
| 3.76            | 3.59 | 3.40 | 3.31 | 3.21 | 3.11 | 3.00 | 2.89 | 2.83 | 2.78 | 0.005    |                 |
| 4.97            | 4.70 | 4.43 | 4.29 | 4.14 | 3.99 | 3.84 | 3.68 | 3.60 | 3.51 | 0.001    |                 |
| 1.39            | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 0.25     | 20              |
| 1.89            | 1.84 | 1.79 | 1.77 | 1.74 | 1.71 | 1.68 | 1.64 | 1.63 | 1.61 | 0.10     |                 |
| 2.28            | 2.20 | 2.12 | 2.08 | 2.04 | 1.99 | 1.95 | 1.90 | 1.87 | 1.84 | 0.05     |                 |
| 2.68            | 2.57 | 2.46 | 2.41 | 2.35 | 2.29 | 2.22 | 2.16 | 2.12 | 2.09 | 0.025    |                 |
| 3.23            | 3.09 | 2.94 | 2.86 | 2.78 | 2.69 | 2.61 | 2.52 | 2.47 | 2.42 | 0.01     |                 |
| 3.68            | 3.50 | 3.32 | 3.22 | 3.12 | 3.02 | 2.92 | 2.81 | 2.75 | 2.69 | 0.005    |                 |
| 4.82            | 4.56 | 4.29 | 4.15 | 4.00 | 3.86 | 3.70 | 3.54 | 3.46 | 3.38 | 0.001    |                 |
| 1.38            | 1.37 | 1.35 | 1.34 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.28 | 0.25     | 21              |
| 1.87            | 1.83 | 1.78 | 1.75 | 1.72 | 1.69 | 1.66 | 1.62 | 1.60 | 1.59 | 0.10     |                 |
| 2.25            | 2.18 | 2.10 | 2.05 | 2.01 | 1.96 | 1.92 | 1.87 | 1.84 | 1.81 | 0.05     |                 |
| 2.64            | 2.53 | 2.42 | 2.37 | 2.31 | 2.25 | 2.18 | 2.11 | 2.08 | 2.04 | 0.025    |                 |
| 3.17            | 3.03 | 2.88 | 2.80 | 2.72 | 2.64 | 2.55 | 2.46 | 2.41 | 2.36 | 0.01     |                 |
| 3.60            | 3.43 | 3.24 | 3.15 | 3.05 | 2.95 | 2.84 | 2.73 | 2.67 | 2.61 | 0.005    |                 |
| 4.70            | 4.44 | 4.17 | 4.03 | 3.88 | 3.74 | 3.58 | 3.42 | 3.34 | 3.26 | 0.001    |                 |

续表

| df <sub>1</sub> |      |      |      |      |      |      |      |      |      | inf. | $\alpha$ | df <sub>2</sub> |
|-----------------|------|------|------|------|------|------|------|------|------|------|----------|-----------------|
| 12              | 15   | 20   | 24   | 30   | 40   | 60   | 120  | 240  |      |      |          |                 |
| 1.37            | 1.36 | 1.34 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.28 | 1.28 | 1.28 | 0.25     | 22              |
| 1.86            | 1.81 | 1.76 | 1.73 | 1.70 | 1.67 | 1.64 | 1.60 | 1.59 | 1.57 | 1.57 | 0.10     |                 |
| 2.23            | 2.15 | 2.07 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 | 1.81 | 1.78 | 1.78 | 0.05     |                 |
| 2.60            | 2.50 | 2.39 | 2.33 | 2.27 | 2.21 | 2.14 | 2.08 | 2.04 | 2.00 | 2.00 | 0.025    |                 |
| 3.12            | 2.98 | 2.83 | 2.75 | 2.67 | 2.58 | 2.50 | 2.40 | 2.35 | 2.31 | 2.31 | 0.01     |                 |
| 3.54            | 3.36 | 3.18 | 3.08 | 2.98 | 2.88 | 2.77 | 2.66 | 2.60 | 2.55 | 2.55 | 0.005    |                 |
| 4.58            | 4.33 | 4.06 | 3.92 | 3.78 | 3.63 | 3.48 | 3.32 | 3.23 | 3.15 | 3.15 | 0.001    |                 |
| 1.37            | 1.35 | 1.34 | 1.33 | 1.32 | 1.31 | 1.30 | 1.28 | 1.28 | 1.27 | 1.27 | 0.25     | 23              |
| 1.84            | 1.80 | 1.74 | 1.72 | 1.69 | 1.66 | 1.62 | 1.59 | 1.57 | 1.55 | 1.55 | 0.10     |                 |
| 2.20            | 2.13 | 2.05 | 2.01 | 1.96 | 1.91 | 1.86 | 1.81 | 1.79 | 1.76 | 1.76 | 0.05     |                 |
| 2.57            | 2.47 | 2.36 | 2.30 | 2.24 | 2.18 | 2.11 | 2.04 | 2.01 | 1.97 | 1.97 | 0.025    |                 |
| 3.07            | 2.93 | 2.78 | 2.70 | 2.62 | 2.54 | 2.45 | 2.35 | 2.31 | 2.26 | 2.26 | 0.01     |                 |
| 3.47            | 3.30 | 3.12 | 3.02 | 2.92 | 2.82 | 2.71 | 2.60 | 2.54 | 2.48 | 2.48 | 0.005    |                 |
| 4.48            | 4.23 | 3.96 | 3.82 | 3.68 | 3.53 | 3.38 | 3.22 | 3.14 | 3.05 | 3.05 | 0.001    |                 |
| 1.36            | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 | 1.28 | 1.27 | 1.26 | 1.26 | 0.25     | 24              |
| 1.83            | 1.78 | 1.73 | 1.70 | 1.67 | 1.64 | 1.61 | 1.57 | 1.55 | 1.53 | 1.53 | 0.10     |                 |
| 2.18            | 2.11 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 | 1.79 | 1.76 | 1.73 | 1.73 | 0.05     |                 |
| 2.54            | 2.44 | 2.33 | 2.27 | 2.21 | 2.15 | 2.08 | 2.01 | 1.97 | 1.94 | 1.94 | 0.025    |                 |
| 3.03            | 2.89 | 2.74 | 2.66 | 2.58 | 2.49 | 2.40 | 2.31 | 2.26 | 2.21 | 2.21 | 0.01     |                 |
| 3.42            | 3.25 | 3.06 | 2.97 | 2.87 | 2.77 | 2.66 | 2.55 | 2.49 | 2.43 | 2.43 | 0.005    |                 |
| 4.39            | 4.14 | 3.87 | 3.74 | 3.59 | 3.45 | 3.29 | 3.14 | 3.05 | 2.97 | 2.97 | 0.001    |                 |

表 8 F 分布的百分位数 ( $df_2$  从 25 到 30)

| $df_2$ | $\alpha$ | $df_1$ |      |      |      |      |      |      |      |      |      |
|--------|----------|--------|------|------|------|------|------|------|------|------|------|
|        |          | 1      | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| 25     | 0.25     | 1.39   | 1.47 | 1.46 | 1.44 | 1.42 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 |
|        | 0.10     | 2.92   | 2.53 | 2.32 | 2.18 | 2.09 | 2.02 | 1.97 | 1.93 | 1.89 | 1.87 |
|        | 0.05     | 4.24   | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 | 2.24 |
|        | 0.025    | 5.69   | 4.29 | 3.69 | 3.35 | 3.13 | 2.97 | 2.85 | 2.75 | 2.68 | 2.61 |
|        | 0.01     | 7.77   | 5.57 | 4.68 | 4.18 | 3.85 | 3.63 | 3.46 | 3.32 | 3.22 | 3.13 |
| 26     | 0.005    | 9.48   | 6.60 | 5.46 | 4.84 | 4.43 | 4.15 | 3.94 | 3.78 | 3.64 | 3.54 |
|        | 0.001    | 13.88  | 9.22 | 7.45 | 6.49 | 5.89 | 5.46 | 5.15 | 4.91 | 4.71 | 4.56 |
|        | 0.25     | 1.38   | 1.46 | 1.45 | 1.44 | 1.42 | 1.41 | 1.39 | 1.38 | 1.37 | 1.37 |
|        | 0.10     | 2.91   | 2.52 | 2.31 | 2.17 | 2.08 | 2.01 | 1.96 | 1.92 | 1.88 | 1.86 |
|        | 0.05     | 4.23   | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 | 2.22 |
| 27     | 0.025    | 5.66   | 4.27 | 3.67 | 3.33 | 3.10 | 2.94 | 2.82 | 2.73 | 2.65 | 2.59 |
|        | 0.01     | 7.72   | 5.53 | 4.64 | 4.14 | 3.82 | 3.59 | 3.42 | 3.29 | 3.18 | 3.09 |
|        | 0.005    | 9.41   | 6.54 | 5.41 | 4.79 | 4.38 | 4.10 | 3.89 | 3.73 | 3.60 | 3.49 |
|        | 0.001    | 13.74  | 9.12 | 7.36 | 6.41 | 5.80 | 5.38 | 5.07 | 4.83 | 4.64 | 4.48 |
|        | 0.25     | 1.38   | 1.46 | 1.45 | 1.43 | 1.42 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 |
|        | 0.10     | 2.90   | 2.51 | 2.30 | 2.17 | 2.07 | 2.00 | 1.95 | 1.91 | 1.87 | 1.85 |
|        | 0.05     | 4.21   | 3.35 | 2.96 | 2.73 | 2.57 | 2.46 | 2.37 | 2.31 | 2.25 | 2.20 |
|        | 0.025    | 5.63   | 4.24 | 3.65 | 3.31 | 3.08 | 2.92 | 2.80 | 2.71 | 2.63 | 2.57 |
|        | 0.01     | 7.68   | 5.49 | 4.60 | 4.11 | 3.78 | 3.56 | 3.39 | 3.26 | 3.15 | 3.06 |
|        | 0.005    | 9.34   | 6.49 | 5.36 | 4.74 | 4.34 | 4.06 | 3.85 | 3.69 | 3.56 | 3.45 |
|        | 0.001    | 13.61  | 9.02 | 7.27 | 6.33 | 5.73 | 5.31 | 5.00 | 4.76 | 4.57 | 4.41 |

续表

| df <sub>2</sub> | $\alpha$ | df <sub>1</sub> |      |      |      |      |      |      |      |      |      |
|-----------------|----------|-----------------|------|------|------|------|------|------|------|------|------|
|                 |          | 1               | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| 28              | 0.25     | 1.38            | 1.46 | 1.45 | 1.43 | 1.41 | 1.40 | 1.39 | 1.38 | 1.37 | 1.36 |
|                 | 0.10     | 2.89            | 2.50 | 2.29 | 2.16 | 2.06 | 2.00 | 1.94 | 1.90 | 1.87 | 1.84 |
|                 | 0.05     | 4.20            | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 | 2.19 |
|                 | 0.025    | 5.61            | 4.22 | 3.63 | 3.29 | 3.06 | 2.90 | 2.78 | 2.69 | 2.61 | 2.55 |
|                 | 0.01     | 7.64            | 5.45 | 4.57 | 4.07 | 3.75 | 3.53 | 3.36 | 3.23 | 3.12 | 3.03 |
|                 | 0.005    | 9.28            | 6.44 | 5.32 | 4.70 | 4.30 | 4.02 | 3.81 | 3.65 | 3.52 | 3.41 |
| 29              | 0.001    | 13.50           | 8.93 | 7.19 | 6.25 | 5.66 | 5.24 | 4.93 | 4.69 | 4.50 | 4.35 |
|                 | 0.25     | 1.38            | 1.45 | 1.45 | 1.43 | 1.41 | 1.40 | 1.38 | 1.37 | 1.36 | 1.35 |
|                 | 0.10     | 2.89            | 2.50 | 2.28 | 2.15 | 2.06 | 1.99 | 1.93 | 1.89 | 1.86 | 1.83 |
|                 | 0.05     | 4.18            | 3.33 | 2.93 | 2.70 | 2.55 | 2.43 | 2.35 | 2.28 | 2.22 | 2.18 |
|                 | 0.025    | 5.59            | 4.20 | 3.61 | 3.27 | 3.04 | 2.88 | 2.76 | 2.67 | 2.59 | 2.53 |
|                 | 0.01     | 7.60            | 5.42 | 4.54 | 4.04 | 3.73 | 3.50 | 3.33 | 3.20 | 3.09 | 3.00 |
| 30              | 0.005    | 9.23            | 6.40 | 5.28 | 4.66 | 4.26 | 3.98 | 3.77 | 3.61 | 3.48 | 3.38 |
|                 | 0.001    | 13.39           | 8.85 | 7.12 | 6.19 | 5.59 | 5.18 | 4.87 | 4.64 | 4.45 | 4.29 |
|                 | 0.25     | 1.38            | 1.45 | 1.44 | 1.42 | 1.41 | 1.39 | 1.38 | 1.37 | 1.36 | 1.35 |
|                 | 0.10     | 2.88            | 2.49 | 2.28 | 2.14 | 2.05 | 1.98 | 1.93 | 1.88 | 1.85 | 1.82 |
|                 | 0.05     | 4.17            | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 |
|                 | 0.025    | 5.57            | 4.18 | 3.59 | 3.25 | 3.03 | 2.87 | 2.75 | 2.65 | 2.57 | 2.51 |
|                 | 0.01     | 7.56            | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 |
|                 | 0.005    | 9.18            | 6.35 | 5.24 | 4.62 | 4.23 | 3.95 | 3.74 | 3.58 | 3.45 | 3.34 |
|                 | 0.001    | 13.29           | 8.77 | 7.05 | 6.12 | 5.53 | 5.12 | 4.82 | 4.58 | 4.39 | 4.24 |


续表

| df <sub>1</sub> |      |      |      |      |      |      |      |      |      |      |          |                 |
|-----------------|------|------|------|------|------|------|------|------|------|------|----------|-----------------|
|                 | 12   | 15   | 20   | 24   | 30   | 40   | 60   | 120  | 240  | inf. | $\alpha$ | df <sub>2</sub> |
|                 | 1.36 | 1.34 | 1.33 | 1.32 | 1.31 | 1.29 | 1.28 | 1.27 | 1.26 | 1.25 | 0.25     | 25              |
|                 | 1.82 | 1.77 | 1.72 | 1.69 | 1.66 | 1.63 | 1.59 | 1.56 | 1.54 | 1.52 | 0.10     |                 |
|                 | 2.16 | 2.09 | 2.01 | 1.96 | 1.92 | 1.87 | 1.82 | 1.77 | 1.74 | 1.71 | 0.05     |                 |
|                 | 2.51 | 2.41 | 2.30 | 2.24 | 2.18 | 2.12 | 2.05 | 1.98 | 1.94 | 1.91 | 0.025    |                 |
|                 | 2.99 | 2.85 | 2.70 | 2.62 | 2.54 | 2.45 | 2.36 | 2.27 | 2.22 | 2.17 | 0.01     |                 |
|                 | 3.37 | 3.20 | 3.01 | 2.92 | 2.82 | 2.72 | 2.61 | 2.50 | 2.44 | 2.38 | 0.005    |                 |
|                 | 4.31 | 4.06 | 3.79 | 3.66 | 3.52 | 3.37 | 3.22 | 3.06 | 2.98 | 2.89 | 0.001    |                 |
|                 | 1.35 | 1.34 | 1.32 | 1.31 | 1.30 | 1.29 | 1.28 | 1.26 | 1.26 | 1.25 | 0.25     | 26              |
|                 | 1.81 | 1.76 | 1.71 | 1.68 | 1.65 | 1.61 | 1.58 | 1.54 | 1.52 | 1.50 | 0.10     |                 |
|                 | 2.15 | 2.07 | 1.99 | 1.95 | 1.90 | 1.85 | 1.80 | 1.75 | 1.72 | 1.69 | 0.05     |                 |
|                 | 2.49 | 2.39 | 2.28 | 2.22 | 2.16 | 2.09 | 2.03 | 1.95 | 1.92 | 1.88 | 0.025    |                 |
|                 | 2.96 | 2.81 | 2.66 | 2.58 | 2.50 | 2.42 | 2.33 | 2.23 | 2.18 | 2.13 | 0.01     |                 |
|                 | 3.33 | 3.15 | 2.97 | 2.87 | 2.77 | 2.67 | 2.56 | 2.45 | 2.39 | 2.33 | 0.005    |                 |
|                 | 4.24 | 3.99 | 3.72 | 3.59 | 3.44 | 3.30 | 3.15 | 2.99 | 2.90 | 2.82 | 0.001    |                 |
|                 | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.28 | 1.27 | 1.26 | 1.25 | 1.24 | 0.25     | 27              |
|                 | 1.80 | 1.75 | 1.70 | 1.67 | 1.64 | 1.60 | 1.57 | 1.53 | 1.51 | 1.49 | 0.10     |                 |
|                 | 2.13 | 2.06 | 1.97 | 1.93 | 1.88 | 1.84 | 1.79 | 1.73 | 1.70 | 1.67 | 0.05     |                 |
|                 | 2.47 | 2.36 | 2.25 | 2.19 | 2.13 | 2.07 | 2.00 | 1.93 | 1.89 | 1.85 | 0.025    |                 |
|                 | 2.93 | 2.78 | 2.63 | 2.55 | 2.47 | 2.38 | 2.29 | 2.20 | 2.15 | 2.10 | 0.01     |                 |
|                 | 3.28 | 3.11 | 2.93 | 2.83 | 2.73 | 2.63 | 2.52 | 2.41 | 2.35 | 2.29 | 0.005    |                 |
|                 | 4.17 | 3.92 | 3.66 | 3.52 | 3.38 | 3.23 | 3.08 | 2.92 | 2.84 | 2.75 | 0.001    |                 |



续表

| df <sub>1</sub> |      |      |      |      |      |      |      |      |      |          | df <sub>2</sub> |
|-----------------|------|------|------|------|------|------|------|------|------|----------|-----------------|
| 12              | 15   | 20   | 24   | 30   | 40   | 60   | 120  | 240  | inf. | $\alpha$ |                 |
| 1.34            | 1.33 | 1.31 | 1.30 | 1.29 | 1.28 | 1.27 | 1.25 | 1.24 | 1.24 | 0.25     | 28              |
| 1.79            | 1.74 | 1.69 | 1.66 | 1.63 | 1.59 | 1.56 | 1.52 | 1.50 | 1.48 | 0.10     |                 |
| 2.12            | 2.04 | 1.96 | 1.91 | 1.87 | 1.82 | 1.77 | 1.71 | 1.68 | 1.65 | 0.05     |                 |
| 2.45            | 2.34 | 2.23 | 2.17 | 2.11 | 2.05 | 1.98 | 1.91 | 1.87 | 1.83 | 0.025    |                 |
| 2.90            | 2.75 | 2.60 | 2.52 | 2.44 | 2.35 | 2.26 | 2.17 | 2.12 | 2.06 | 0.01     |                 |
| 3.25            | 3.07 | 2.89 | 2.79 | 2.69 | 2.59 | 2.48 | 2.37 | 2.31 | 2.25 | 0.005    |                 |
| 4.11            | 3.86 | 3.60 | 3.46 | 3.32 | 3.18 | 3.02 | 2.86 | 2.78 | 2.69 | 0.001    |                 |
| 1.34            | 1.32 | 1.31 | 1.30 | 1.29 | 1.27 | 1.26 | 1.25 | 1.24 | 1.23 | 0.25     | 29              |
| 1.78            | 1.73 | 1.68 | 1.65 | 1.62 | 1.58 | 1.55 | 1.51 | 1.49 | 1.47 | 0.10     |                 |
| 2.10            | 2.03 | 1.94 | 1.90 | 1.85 | 1.81 | 1.75 | 1.70 | 1.67 | 1.64 | 0.05     |                 |
| 2.43            | 2.32 | 2.21 | 2.15 | 2.09 | 2.03 | 1.96 | 1.89 | 1.85 | 1.81 | 0.025    |                 |
| 2.87            | 2.73 | 2.57 | 2.49 | 2.41 | 2.33 | 2.23 | 2.14 | 2.09 | 2.03 | 0.01     |                 |
| 3.21            | 3.04 | 2.86 | 2.76 | 2.66 | 2.56 | 2.45 | 2.33 | 2.27 | 2.21 | 0.005    |                 |
| 4.05            | 3.80 | 3.54 | 3.41 | 3.27 | 3.12 | 2.97 | 2.81 | 2.73 | 2.64 | 0.001    |                 |
| 1.34            | 1.32 | 1.30 | 1.29 | 1.28 | 1.27 | 1.26 | 1.24 | 1.23 | 1.23 | 0.25     | 30              |
| 1.77            | 1.72 | 1.67 | 1.64 | 1.61 | 1.57 | 1.54 | 1.50 | 1.48 | 1.46 | 0.10     |                 |
| 2.09            | 2.01 | 1.93 | 1.89 | 1.84 | 1.79 | 1.74 | 1.68 | 1.65 | 1.62 | 0.05     |                 |
| 2.41            | 2.31 | 2.20 | 2.14 | 2.07 | 2.01 | 1.94 | 1.87 | 1.83 | 1.79 | 0.025    |                 |
| 2.84            | 2.70 | 2.55 | 2.47 | 2.39 | 2.30 | 2.21 | 2.11 | 2.06 | 2.01 | 0.01     |                 |
| 3.18            | 3.01 | 2.82 | 2.73 | 2.63 | 2.52 | 2.42 | 2.30 | 2.24 | 2.18 | 0.005    |                 |
| 4.00            | 3.75 | 3.49 | 3.36 | 3.22 | 3.07 | 2.92 | 2.76 | 2.68 | 2.59 | 0.001    |                 |

表 8 F 分布的百分位数 ( $df_2$  至少为 40)


| $df_2$ | $\alpha$ | $df_1$ |      |      |      |      |      |      |      |      |      |
|--------|----------|--------|------|------|------|------|------|------|------|------|------|
|        |          | 1      | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| 40     | 0.25     | 1.36   | 1.44 | 1.42 | 1.40 | 1.39 | 1.37 | 1.36 | 1.35 | 1.34 | 1.33 |
|        | 0.10     | 2.84   | 2.44 | 2.23 | 2.09 | 2.00 | 1.93 | 1.87 | 1.83 | 1.79 | 1.76 |
|        | 0.05     | 4.08   | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 |
|        | 0.025    | 5.42   | 4.05 | 3.46 | 3.13 | 2.90 | 2.74 | 2.62 | 2.53 | 2.45 | 2.39 |
|        | 0.01     | 7.31   | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 | 2.80 |
| 60     | 0.005    | 8.83   | 6.07 | 4.98 | 4.37 | 3.99 | 3.71 | 3.51 | 3.35 | 3.22 | 3.12 |
|        | 0.001    | 12.61  | 8.25 | 6.59 | 5.70 | 5.13 | 4.73 | 4.44 | 4.21 | 4.02 | 3.87 |
|        | 0.25     | 1.35   | 1.42 | 1.41 | 1.38 | 1.37 | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 |
|        | 0.10     | 2.79   | 2.39 | 2.18 | 2.04 | 1.95 | 1.87 | 1.82 | 1.77 | 1.74 | 1.71 |
|        | 0.05     | 4.00   | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 |
| 90     | 0.025    | 5.29   | 3.93 | 3.34 | 3.01 | 2.79 | 2.63 | 2.51 | 2.41 | 2.33 | 2.27 |
|        | 0.01     | 7.08   | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 | 2.63 |
|        | 0.005    | 8.49   | 5.79 | 4.73 | 4.14 | 3.76 | 3.49 | 3.29 | 3.13 | 3.01 | 2.90 |
|        | 0.001    | 11.97  | 7.77 | 6.17 | 5.31 | 4.76 | 4.37 | 4.09 | 3.86 | 3.69 | 3.54 |
|        | 0.25     | 1.34   | 1.41 | 1.39 | 1.37 | 1.35 | 1.33 | 1.32 | 1.31 | 1.30 | 1.29 |
|        | 0.10     | 2.76   | 2.36 | 2.15 | 2.01 | 1.91 | 1.84 | 1.78 | 1.74 | 1.70 | 1.67 |
|        | 0.05     | 3.95   | 3.10 | 2.71 | 2.47 | 2.32 | 2.20 | 2.11 | 2.04 | 1.99 | 1.94 |
|        | 0.025    | 5.20   | 3.84 | 3.26 | 2.93 | 2.71 | 2.55 | 2.43 | 2.34 | 2.26 | 2.19 |
|        | 0.01     | 6.93   | 4.85 | 4.01 | 3.53 | 3.23 | 3.01 | 2.84 | 2.72 | 2.61 | 2.52 |
|        | 0.005    | 8.28   | 5.62 | 4.57 | 3.99 | 3.62 | 3.35 | 3.15 | 3.00 | 2.87 | 2.77 |
|        | 0.001    | 11.57  | 7.47 | 5.91 | 5.06 | 4.53 | 4.15 | 3.87 | 3.65 | 3.48 | 3.34 |

续表

| $df_2$ | $\alpha$ | $df_1$ |      |      |      |      |      |      |      |      |      |
|--------|----------|--------|------|------|------|------|------|------|------|------|------|
|        |          | 1      | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| 120    | 0.25     | 1.34   | 1.40 | 1.39 | 1.37 | 1.35 | 1.33 | 1.31 | 1.30 | 1.29 | 1.28 |
|        | 0.10     | 2.75   | 2.35 | 2.13 | 1.99 | 1.90 | 1.82 | 1.77 | 1.72 | 1.68 | 1.65 |
|        | 0.05     | 3.92   | 3.07 | 2.68 | 2.45 | 2.29 | 2.18 | 2.09 | 2.02 | 1.96 | 1.91 |
|        | 0.025    | 5.15   | 3.80 | 3.23 | 2.89 | 2.67 | 2.52 | 2.39 | 2.30 | 2.22 | 2.16 |
|        | 0.01     | 6.85   | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.79 | 2.66 | 2.56 | 2.47 |
| 240    | 0.005    | 8.18   | 5.54 | 4.50 | 3.92 | 3.55 | 3.28 | 3.09 | 2.93 | 2.81 | 2.71 |
|        | 0.001    | 11.38  | 7.32 | 5.78 | 4.95 | 4.42 | 4.04 | 3.77 | 3.55 | 3.38 | 3.24 |
|        | 0.25     | 1.33   | 1.39 | 1.38 | 1.36 | 1.34 | 1.32 | 1.30 | 1.29 | 1.27 | 1.27 |
|        | 0.10     | 2.73   | 2.32 | 2.10 | 1.97 | 1.87 | 1.80 | 1.74 | 1.70 | 1.65 | 1.63 |
|        | 0.05     | 3.88   | 3.03 | 2.64 | 2.41 | 2.25 | 2.14 | 2.04 | 1.98 | 1.92 | 1.87 |
| inf.   | 0.025    | 5.09   | 3.75 | 3.17 | 2.84 | 2.62 | 2.46 | 2.34 | 2.25 | 2.17 | 2.10 |
|        | 0.01     | 6.74   | 4.69 | 3.86 | 3.40 | 3.09 | 2.88 | 2.71 | 2.59 | 2.48 | 2.40 |
|        | 0.005    | 8.03   | 5.42 | 4.38 | 3.82 | 3.45 | 3.19 | 2.99 | 2.84 | 2.71 | 2.61 |
|        | 0.001    | 11.10  | 7.11 | 5.60 | 4.78 | 4.25 | 3.89 | 3.62 | 3.41 | 3.24 | 3.09 |
|        | 0.25     | 1.32   | 1.39 | 1.37 | 1.35 | 1.33 | 1.31 | 1.29 | 1.28 | 1.27 | 1.25 |
|        | 0.10     | 2.71   | 2.30 | 2.08 | 1.94 | 1.85 | 1.77 | 1.72 | 1.67 | 1.63 | 1.60 |
|        | 0.05     | 3.84   | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 | 1.83 |
|        | 0.025    | 5.02   | 3.69 | 3.12 | 2.79 | 2.57 | 2.41 | 2.29 | 2.19 | 2.11 | 2.05 |
|        | 0.01     | 6.63   | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 | 2.51 | 2.41 | 2.32 |
|        | 0.005    | 7.88   | 5.30 | 4.28 | 3.72 | 3.35 | 3.09 | 2.90 | 2.74 | 2.62 | 2.52 |
|        | 0.001    | 10.83  | 6.91 | 5.42 | 4.62 | 4.10 | 3.74 | 3.47 | 3.27 | 3.10 | 2.96 |

续表

| df <sub>1</sub> |      |      |      |      |      |      |      |      |      |      |          |                 |
|-----------------|------|------|------|------|------|------|------|------|------|------|----------|-----------------|
|                 | 12   | 15   | 20   | 24   | 30   | 40   | 60   | 120  | 240  | inf. | $\alpha$ | df <sub>2</sub> |
|                 | 1.31 | 1.30 | 1.28 | 1.26 | 1.25 | 1.24 | 1.22 | 1.21 | 1.20 | 1.19 | 0.25     | 40              |
|                 | 1.71 | 1.66 | 1.61 | 1.57 | 1.54 | 1.51 | 1.47 | 1.42 | 1.40 | 1.38 | 0.10     |                 |
|                 | 2.00 | 1.92 | 1.84 | 1.79 | 1.74 | 1.69 | 1.64 | 1.58 | 1.54 | 1.51 | 0.05     |                 |
|                 | 2.29 | 2.18 | 2.07 | 2.01 | 1.94 | 1.88 | 1.80 | 1.72 | 1.68 | 1.64 | 0.025    |                 |
|                 | 2.66 | 2.52 | 2.37 | 2.29 | 2.20 | 2.11 | 2.02 | 1.92 | 1.86 | 1.80 | 0.01     |                 |
|                 | 2.95 | 2.78 | 2.60 | 2.50 | 2.40 | 2.30 | 2.18 | 2.06 | 2.00 | 1.93 | 0.005    |                 |
|                 | 3.64 | 3.40 | 3.14 | 3.01 | 2.87 | 2.73 | 2.57 | 2.41 | 2.32 | 2.23 | 0.001    |                 |
|                 | 1.29 | 1.27 | 1.25 | 1.24 | 1.22 | 1.21 | 1.19 | 1.17 | 1.16 | 1.15 | 0.25     | 60              |
|                 | 1.66 | 1.60 | 1.54 | 1.51 | 1.48 | 1.44 | 1.40 | 1.35 | 1.32 | 1.29 | 0.10     |                 |
|                 | 1.92 | 1.84 | 1.75 | 1.70 | 1.65 | 1.59 | 1.53 | 1.47 | 1.43 | 1.39 | 0.05     |                 |
|                 | 2.17 | 2.06 | 1.94 | 1.88 | 1.82 | 1.74 | 1.67 | 1.58 | 1.53 | 1.48 | 0.025    |                 |
|                 | 2.50 | 2.35 | 2.20 | 2.12 | 2.03 | 1.94 | 1.84 | 1.73 | 1.67 | 1.60 | 0.01     |                 |
|                 | 2.74 | 2.57 | 2.39 | 2.29 | 2.19 | 2.08 | 1.96 | 1.83 | 1.76 | 1.69 | 0.005    |                 |
|                 | 3.32 | 3.08 | 2.83 | 2.69 | 2.55 | 2.41 | 2.25 | 2.08 | 1.99 | 1.89 | 0.001    |                 |
|                 | 1.27 | 1.25 | 1.23 | 1.22 | 1.20 | 1.19 | 1.17 | 1.15 | 1.13 | 1.12 | 0.25     | 90              |
|                 | 1.62 | 1.56 | 1.50 | 1.47 | 1.43 | 1.39 | 1.35 | 1.29 | 1.26 | 1.23 | 0.10     |                 |
|                 | 1.86 | 1.78 | 1.69 | 1.64 | 1.59 | 1.53 | 1.46 | 1.39 | 1.35 | 1.30 | 0.05     |                 |
|                 | 2.09 | 1.98 | 1.86 | 1.80 | 1.73 | 1.66 | 1.58 | 1.48 | 1.43 | 1.37 | 0.025    |                 |
|                 | 2.39 | 2.24 | 2.09 | 2.00 | 1.92 | 1.82 | 1.72 | 1.60 | 1.53 | 1.46 | 0.01     |                 |
|                 | 2.61 | 2.44 | 2.25 | 2.15 | 2.05 | 1.94 | 1.82 | 1.68 | 1.61 | 1.52 | 0.005    |                 |
|                 | 3.11 | 2.88 | 2.63 | 2.50 | 2.36 | 2.21 | 2.05 | 1.87 | 1.77 | 1.66 | 0.001    |                 |

续表

| df <sub>1</sub> |      |      |      |      |      |      |      |      |      |       |                 | α    | inf. | 240  | 120  | 240  | inf. | α     | df <sub>2</sub> |
|-----------------|------|------|------|------|------|------|------|------|------|-------|-----------------|------|------|------|------|------|------|-------|-----------------|
| 12              | 15   | 20   | 24   | 30   | 40   | 60   | 120  | 240  | inf. | α     | df <sub>2</sub> |      |      |      |      |      |      |       |                 |
| 1.26            | 1.24 | 1.22 | 1.21 | 1.19 | 1.18 | 1.16 | 1.13 | 1.12 | 1.10 | 0.25  | 120             | 1.10 | 1.10 | 1.12 | 1.13 | 1.12 | 1.10 | 0.25  | 120             |
| 1.60            | 1.55 | 1.48 | 1.45 | 1.41 | 1.37 | 1.32 | 1.26 | 1.23 | 1.19 | 10    |                 | 1.19 | 1.19 | 1.23 | 1.26 | 1.23 | 1.19 | 10    |                 |
| 1.83            | 1.75 | 1.66 | 1.61 | 1.55 | 1.50 | 1.43 | 1.35 | 1.31 | 1.25 | 0.05  |                 | 1.25 | 1.25 | 1.31 | 1.35 | 1.31 | 1.25 | 0.05  |                 |
| 2.05            | 1.94 | 1.82 | 1.76 | 1.69 | 1.61 | 1.53 | 1.43 | 1.38 | 1.31 | 0.025 |                 | 1.31 | 1.31 | 1.38 | 1.43 | 1.38 | 1.31 | 0.025 |                 |
| 2.34            | 2.19 | 2.03 | 1.95 | 1.86 | 1.76 | 1.66 | 1.53 | 1.46 | 1.38 | 0.01  |                 | 1.38 | 1.38 | 1.46 | 1.53 | 1.46 | 1.38 | 0.01  |                 |
| 2.54            | 2.37 | 2.19 | 2.09 | 1.98 | 1.87 | 1.75 | 1.61 | 1.52 | 1.43 | 0.005 |                 | 1.43 | 1.43 | 1.52 | 1.61 | 1.52 | 1.43 | 0.005 |                 |
| 3.02            | 2.78 | 2.53 | 2.40 | 2.26 | 2.11 | 1.95 | 1.77 | 1.66 | 1.54 | 0.001 |                 | 1.54 | 1.54 | 1.66 | 1.77 | 1.66 | 1.54 | 0.001 |                 |
| 1.25            | 1.23 | 1.21 | 1.19 | 1.18 | 1.16 | 1.14 | 1.11 | 1.09 | 1.07 | 0.25  | 240             | 1.07 | 1.07 | 1.09 | 1.11 | 1.09 | 1.07 | 0.25  | 240             |
| 1.57            | 1.52 | 1.45 | 1.42 | 1.38 | 1.33 | 1.28 | 1.22 | 1.18 | 1.13 | 0.10  |                 | 1.13 | 1.13 | 1.18 | 1.22 | 1.18 | 1.13 | 0.10  |                 |
| 1.79            | 1.71 | 1.61 | 1.56 | 1.51 | 1.44 | 1.37 | 1.29 | 1.24 | 1.17 | 0.05  |                 | 1.17 | 1.17 | 1.24 | 1.29 | 1.24 | 1.17 | 0.05  |                 |
| 2.00            | 1.89 | 1.77 | 1.70 | 1.63 | 1.55 | 1.46 | 1.35 | 1.29 | 1.21 | 0.025 |                 | 1.21 | 1.21 | 1.29 | 1.35 | 1.29 | 1.21 | 0.025 |                 |
| 2.26            | 2.11 | 1.96 | 1.87 | 1.78 | 1.68 | 1.57 | 1.43 | 1.35 | 1.25 | 0.01  |                 | 1.25 | 1.25 | 1.35 | 1.43 | 1.35 | 1.25 | 0.01  |                 |
| 2.45            | 2.28 | 2.09 | 1.99 | 1.89 | 1.77 | 1.64 | 1.49 | 1.40 | 1.28 | 0.005 |                 | 1.28 | 1.28 | 1.40 | 1.49 | 1.40 | 1.28 | 0.005 |                 |
| 2.88            | 2.65 | 2.40 | 2.26 | 2.12 | 1.97 | 1.80 | 1.61 | 1.49 | 1.35 | 0.001 |                 | 1.35 | 1.35 | 1.49 | 1.61 | 1.49 | 1.35 | 0.001 |                 |
| 1.24            | 1.22 | 1.19 | 1.18 | 1.16 | 1.14 | 1.12 | 1.08 | 1.06 | 1.00 | 0.25  | inf.            | 1.00 | 1.00 | 1.06 | 1.08 | 1.06 | 1.00 | 0.25  | inf.            |
| 1.55            | 1.49 | 1.42 | 1.38 | 1.34 | 1.30 | 1.24 | 1.17 | 1.12 | 1.00 | 0.10  |                 | 1.00 | 1.00 | 1.12 | 1.17 | 1.12 | 1.00 | 0.10  |                 |
| 1.75            | 1.67 | 1.57 | 1.52 | 1.46 | 1.39 | 1.32 | 1.22 | 1.15 | 1.00 | 0.05  |                 | 1.00 | 1.00 | 1.15 | 1.22 | 1.15 | 1.00 | 0.05  |                 |
| 1.94            | 1.83 | 1.71 | 1.64 | 1.57 | 1.48 | 1.39 | 1.27 | 1.19 | 1.00 | 0.025 |                 | 1.00 | 1.00 | 1.19 | 1.27 | 1.19 | 1.00 | 0.025 |                 |
| 2.18            | 2.04 | 1.88 | 1.79 | 1.70 | 1.59 | 1.47 | 1.32 | 1.22 | 1.00 | 0.01  |                 | 1.00 | 1.00 | 1.22 | 1.32 | 1.22 | 1.00 | 0.01  |                 |
| 2.36            | 2.19 | 2.00 | 1.90 | 1.79 | 1.67 | 1.53 | 1.36 | 1.25 | 1.00 | 0.005 |                 | 1.00 | 1.00 | 1.25 | 1.36 | 1.25 | 1.00 | 0.005 |                 |
| 2.74            | 2.51 | 2.27 | 2.13 | 1.99 | 1.84 | 1.66 | 1.45 | 1.31 | 1.00 | 0.001 |                 | 1.00 | 1.00 | 1.31 | 1.45 | 1.31 | 1.00 | 0.001 |                 |

来源:由 P.J. Hildebrand 计算。

表 9  $2\arcsin\sqrt{x}$  的值

| $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 0.001       | 0.0633      | 0.041       | 0.4078      | 0.36        | 1.2870      | 0.76        | 2.1177      | 0.971       | 2.7993      |
| 0.002       | 0.0895      | 0.042       | 0.4128      | 0.37        | 1.3078      | .77         | 2.1412      | 0.972       | 2.8053      |
| 0.003       | 0.1096      | 0.043       | 0.4178      | 0.38        | 1.3284      | 0.78        | 2.1652      | 0.973       | 2.8115      |
| 0.004       | 0.1266      | 0.044       | 0.4227      | 0.39        | 1.3490      | 0.79        | 2.1895      | 0.974       | 2.8177      |
| 0.005       | 0.1415      | 0.045       | 0.4275      | 0.40        | 1.3694      | 0.80        | 2.2143      | 0.975       | 2.8240      |
| 0.006       | 0.1551      | 0.046       | 0.4323      | 0.41        | 1.3898      | 0.81        | 2.2395      | 0.976       | 2.8305      |
| 0.007       | 0.1675      | 0.047       | 0.4371      | 0.42        | 1.4101      | 0.82        | 2.2653      | 0.977       | 2.8371      |
| 0.008       | 0.1791      | 0.048       | 0.4418      | 0.43        | 1.4303      | 0.83        | 2.2916      | 0.978       | 2.8438      |
| 0.009       | 0.1900      | 0.049       | 0.4464      | 0.44        | 1.4505      | 0.84        | 2.3186      | 0.979       | 2.8507      |
| 0.010       | 0.2003      | 0.050       | 0.4510      | 0.45        | 1.4706      | 0.85        | 2.3462      | 0.980       | 2.8578      |
| 0.011       | 0.2101      | 0.06        | 0.4949      | 0.46        | 1.4907      | 0.86        | 2.3746      | 0.981       | 2.8650      |
| 0.012       | 0.2195      | 0.07        | 0.5355      | 0.47        | 1.5108      | 0.87        | 2.4039      | 0.982       | 2.8725      |
| 0.013       | 0.2285      | 0.08        | 0.5735      | 0.48        | 1.5308      | 0.88        | 2.4341      | 0.983       | 2.8801      |
| 0.014       | 0.2372      | 0.09        | 0.6094      | 0.49        | 1.5508      | 0.89        | 2.4655      | 0.984       | 2.8879      |
| 0.015       | 0.2456      | 0.10        | 0.6435      | 0.50        | 1.5708      | 0.90        | 2.4981      | 0.985       | 2.8960      |
| 0.016       | 0.2537      | 0.11        | 0.6761      | 0.51        | 1.5908      | 0.91        | 2.5322      | 0.986       | 2.9044      |
| 0.017       | 0.2615      | 0.12        | 0.7075      | 0.52        | 1.6108      | 0.92        | 2.5681      | 0.987       | 2.9131      |
| 0.018       | 0.2691      | 0.13        | 0.7377      | 0.53        | 1.6308      | 0.93        | 2.6062      | 0.988       | 2.9221      |
| 0.019       | 0.2766      | 0.14        | 0.7670      | 0.54        | 1.6509      | 0.94        | 2.6467      | 0.989       | 2.9315      |
| 0.020       | 0.2838      | 0.15        | 0.7954      | 0.55        | 1.6710      | 0.95        | 2.6906      | 0.990       | 2.9413      |

续表

| $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ | $\hat{\pi}$ |       |        |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------|--------|
| 0.021       | 0.2909      | 0.16        | 0.8230      | 0.56        | 1.6911      | 0.951       | 2.6952      | 0.991 | 2.9516 |
| 0.022       | 0.2978      | 0.17        | 0.8500      | 0.57        | 1.7113      | 0.952       | 2.6998      | 0.992 | 2.9625 |
| 0.023       | 0.3045      | 0.18        | 0.8763      | 0.58        | 1.7315      | 0.953       | 2.7045      | 0.993 | 2.9741 |
| 0.024       | 0.3111      | 0.19        | 0.9021      | 0.59        | 1.7518      | 0.954       | 2.7093      | 0.994 | 2.9865 |
| 0.025       | 0.3176      | 0.20        | 0.9273      | 0.60        | 1.7722      | 0.955       | 2.7141      | 0.995 | 3.0001 |
| 0.026       | 0.3239      | 0.21        | 0.9521      | 0.61        | 1.7926      | 0.956       | 2.7189      | 0.996 | 3.0150 |
| 0.027       | 0.3301      | 0.22        | 0.9764      | 0.62        | 1.8132      | 0.957       | 2.7238      | 0.997 | 3.0320 |
| 0.028       | 0.3363      | 0.23        | 1.0004      | 0.63        | 1.8338      | 0.958       | 2.7288      | 0.998 | 3.0521 |
| 0.029       | 0.3423      | 0.24        | 1.0239      | 0.64        | 1.8338      | 0.959       | 2.7338      | 0.999 | 3.0783 |
| 0.030       | 0.3482      | 0.25        | 1.0472      | 0.65        | 1.8546      | 0.960       | 2.7389      |       |        |
| 0.031       | 0.3540      | 0.26        | 1.0701      | 0.66        | 1.8965      | 0.961       | 2.7440      |       |        |
| 0.032       | 0.3597      | 0.27        | 1.0928      | 0.67        | 1.9177      | 0.962       | 2.7492      |       |        |
| 0.033       | 0.3654      | 0.28        | 1.1152      | 0.68        | 1.9391      | 0.963       | 2.7545      |       |        |
| 0.034       | 0.3709      | 0.29        | 1.1374      | 0.69        | 1.9606      | 0.964       | 2.7598      |       |        |
| 0.035       | 0.3764      | 0.30        | 1.1593      | 0.70        | 1.9823      | 0.965       | 2.7652      |       |        |
| 0.036       | 0.3818      | 0.31        | 1.1810      | 0.71        | 2.0042      | 0.966       | 2.7707      |       |        |
| 0.037       | 0.3871      | 0.32        | 1.2025      | 0.72        | 2.0264      | 0.967       | 2.7762      |       |        |
| 0.038       | 0.3924      | 0.33        | 1.2239      | 0.73        | 2.0488      | 0.968       | 2.7819      |       |        |
| 0.039       | 0.3976      | 0.34        | 1.2451      | 0.74        | 2.0715      | 0.969       | 2.7876      |       |        |
| 0.040       | 0.4027      | 0.35        | 1.2661      | 0.75        | 2.0944      | 0.970       | 2.7934      |       |        |

来自: "Experimental Design: Procedures for the Behavioral Sciences", Roger E. Kirk 著, 版权所有; Wadsworth Publishing Company, Inc. 1968, 经该公司许可重印。

表 10 学生化极差分布的百分位数

| 误差<br>df | $\alpha$ | $t$ - 处理均值的个数 |      |      |      |      |      |      |      |       |       |
|----------|----------|---------------|------|------|------|------|------|------|------|-------|-------|
|          |          | 2             | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10    | 11    |
| 5        | 0.05     | 3.64          | 4.60 | 5.22 | 5.67 | 6.03 | 6.33 | 6.58 | 6.80 | 6.99  | 7.17  |
|          | 0.01     | 5.70          | 6.98 | 7.80 | 8.42 | 8.91 | 9.32 | 9.67 | 9.97 | 10.24 | 10.48 |
| 6        | 0.05     | 3.46          | 4.34 | 4.90 | 5.30 | 5.63 | 5.90 | 6.12 | 6.32 | 6.49  | 6.65  |
|          | 0.01     | 5.24          | 6.33 | 7.03 | 7.56 | 7.97 | 8.32 | 8.61 | 8.87 | 9.10  | 9.30  |
| 7        | 0.05     | 3.34          | 4.16 | 4.68 | 5.06 | 5.36 | 5.61 | 5.82 | 6.00 | 6.16  | 6.30  |
|          | 0.01     | 4.95          | 5.92 | 6.54 | 7.01 | 7.37 | 7.68 | 7.94 | 8.17 | 8.37  | 8.55  |
| 8        | 0.05     | 3.26          | 4.04 | 4.53 | 4.89 | 5.17 | 5.40 | 5.60 | 5.77 | 5.92  | 6.05  |
|          | 0.01     | 4.75          | 5.64 | 6.20 | 6.62 | 6.96 | 7.24 | 7.47 | 7.68 | 7.86  | 8.03  |
| 9        | 0.05     | 3.20          | 3.95 | 4.41 | 4.76 | 5.02 | 5.24 | 5.43 | 5.59 | 5.74  | 5.87  |
|          | 0.01     | 4.60          | 5.43 | 5.96 | 6.35 | 6.66 | 6.91 | 7.13 | 7.33 | 7.49  | 7.65  |
| 10       | 0.05     | 3.15          | 3.88 | 4.33 | 4.65 | 4.91 | 5.12 | 5.30 | 5.46 | 5.60  | 5.72  |
|          | 0.01     | 4.48          | 5.27 | 5.77 | 6.14 | 6.43 | 6.67 | 6.87 | 7.05 | 7.21  | 7.36  |
| 11       | 0.05     | 3.11          | 3.82 | 4.26 | 4.57 | 4.82 | 5.03 | 5.30 | 5.35 | 5.49  | 5.61  |
|          | 0.01     | 4.39          | 5.15 | 5.62 | 5.97 | 6.25 | 6.48 | 6.67 | 6.84 | 6.99  | 7.13  |
| 12       | 0.05     | 3.08          | 3.77 | 4.20 | 4.52 | 4.75 | 4.95 | 5.12 | 5.27 | 5.39  | 5.51  |
|          | 0.01     | 4.32          | 5.05 | 5.50 | 5.84 | 6.10 | 6.32 | 6.51 | 6.67 | 6.81  | 6.94  |
| 13       | 0.05     | 3.06          | 3.73 | 4.15 | 4.45 | 4.69 | 4.88 | 5.05 | 5.19 | 5.32  | 5.43  |
|          | 0.01     | 4.26          | 4.96 | 5.40 | 5.73 | 5.98 | 6.19 | 6.37 | 6.53 | 6.67  | 6.79  |
| 14       | 0.05     | 3.03          | 3.70 | 4.11 | 4.41 | 4.64 | 4.83 | 4.99 | 5.13 | 5.25  | 5.36  |
|          | 0.01     | 4.21          | 4.89 | 5.32 | 5.63 | 5.88 | 6.08 | 6.26 | 6.41 | 6.54  | 6.66  |
| 15       | 0.05     | 3.01          | 3.67 | 4.08 | 4.37 | 4.59 | 4.78 | 4.94 | 5.08 | 5.20  | 5.31  |
|          | 0.01     | 4.17          | 4.84 | 5.25 | 5.56 | 5.80 | 5.99 | 6.16 | 6.31 | 6.44  | 6.55  |



续表

| 误差       |          | $t = \text{处理均值的个数}$ |      |      |      |      |      |      |      |      |      |
|----------|----------|----------------------|------|------|------|------|------|------|------|------|------|
| df       | $\alpha$ | 2                    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   |
| 16       | 0.05     | 3.00                 | 3.65 | 4.05 | 4.33 | 4.56 | 4.74 | 4.90 | 5.03 | 5.15 | 5.26 |
|          | 0.01     | 4.13                 | 4.79 | 5.19 | 5.49 | 5.72 | 5.92 | 6.08 | 6.22 | 6.35 | 6.46 |
| 17       | 0.05     | 2.98                 | 3.63 | 4.02 | 4.30 | 4.52 | 4.70 | 4.86 | 4.99 | 5.11 | 5.21 |
|          | 0.01     | 4.10                 | 4.74 | 5.14 | 5.43 | 5.66 | 5.85 | 6.01 | 6.15 | 6.27 | 6.38 |
| 18       | 0.05     | 2.97                 | 3.61 | 4.00 | 4.28 | 4.49 | 4.67 | 4.82 | 4.96 | 5.07 | 5.17 |
|          | 0.01     | 4.07                 | 4.70 | 5.09 | 5.38 | 5.60 | 5.79 | 5.94 | 6.08 | 6.20 | 6.31 |
| 19       | 0.05     | 2.96                 | 3.59 | 3.98 | 4.25 | 4.47 | 4.65 | 4.79 | 4.92 | 5.04 | 5.14 |
|          | 0.01     | 4.05                 | 4.67 | 5.05 | 5.33 | 5.55 | 5.73 | 5.89 | 6.02 | 6.14 | 6.25 |
| 20       | 0.05     | 2.95                 | 3.58 | 3.96 | 4.23 | 4.45 | 4.62 | 4.77 | 4.90 | 5.01 | 5.11 |
|          | 0.01     | 4.02                 | 4.64 | 5.02 | 5.29 | 5.51 | 5.69 | 5.84 | 5.97 | 6.09 | 6.19 |
| 24       | 0.05     | 2.92                 | 3.53 | 3.90 | 4.17 | 4.37 | 4.54 | 4.68 | 4.81 | 3.92 | 5.01 |
|          | 0.01     | 3.96                 | 4.55 | 4.91 | 5.17 | 5.37 | 5.54 | 5.69 | 5.81 | 5.92 | 6.02 |
| 30       | 0.05     | 2.89                 | 3.49 | 3.85 | 4.10 | 4.30 | 4.46 | 4.60 | 4.72 | 4.82 | 4.92 |
|          | 0.01     | 3.89                 | 4.45 | 4.80 | 5.05 | 5.24 | 5.40 | 5.54 | 5.65 | 5.76 | 5.85 |
| 40       | 0.05     | 2.86                 | 3.44 | 3.79 | 4.04 | 4.23 | 4.39 | 4.52 | 4.63 | 4.73 | 4.82 |
|          | 0.01     | 3.82                 | 4.37 | 4.70 | 4.93 | 5.11 | 5.26 | 5.39 | 5.50 | 5.60 | 5.69 |
| 60       | 0.05     | 2.83                 | 3.40 | 3.74 | 3.98 | 4.16 | 4.31 | 4.44 | 4.55 | 4.65 | 4.73 |
|          | 0.01     | 3.76                 | 4.28 | 4.59 | 4.82 | 4.99 | 5.13 | 5.25 | 5.36 | 5.45 | 5.53 |
| 120      | 0.05     | 2.80                 | 3.36 | 3.68 | 3.92 | 4.10 | 4.24 | 4.36 | 4.47 | 4.56 | 4.64 |
|          | 0.01     | 3.70                 | 4.20 | 4.50 | 4.71 | 4.87 | 5.01 | 5.12 | 5.21 | 5.30 | 5.37 |
| $\infty$ | 0.05     | 2.77                 | 3.31 | 3.63 | 3.86 | 4.03 | 4.17 | 4.29 | 4.39 | 4.47 | 4.55 |
|          | 0.01     | 3.64                 | 4.12 | 4.40 | 4.60 | 4.76 | 4.88 | 4.99 | 5.08 | 5.16 | 5.23 |

此表选自 E. S. Pearson and H. O. Hartley 主编, *Biometrika Tables for Statisticians*, 2d ed., Vol 1 (New York: Cambridge University Press, 1958), Table 29. 经编辑和 Biometrika 主管同意重印。

续表

| 误差 | $t =$ 处理均值的个数 |       |       |       |       |       |       |       |       |       |          |
|----|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
|    | df            | 12    | 13    | 14    | 15    | 16    | 17    | 18    | 19    | 20    | $\alpha$ |
| 5  | 5             | 7.32  | 7.47  | 7.60  | 7.72  | 7.83  | 7.93  | 8.03  | 8.12  | 8.21  | 0.05     |
|    |               | 10.70 | 10.89 | 11.08 | 11.24 | 11.40 | 11.55 | 11.68 | 11.81 | 11.93 | 0.01     |
| 6  | 6             | 6.79  | 6.92  | 7.03  | 7.14  | 7.24  | 7.34  | 7.43  | 7.51  | 7.59  | 0.05     |
|    |               | 9.48  | 9.65  | 9.81  | 9.95  | 10.08 | 10.21 | 10.32 | 10.43 | 10.54 | 0.01     |
| 7  | 7             | 6.43  | 6.55  | 6.66  | 6.76  | 6.85  | 6.94  | 7.02  | 7.10  | 7.17  | 0.05     |
|    |               | 8.71  | 8.86  | 9.00  | 9.12  | 9.24  | 9.35  | 9.46  | 9.55  | 9.65  | 0.01     |
| 8  | 8             | 6.18  | 6.29  | 6.39  | 6.48  | 6.57  | 6.65  | 6.73  | 6.80  | 6.87  | 0.05     |
|    |               | 8.18  | 8.31  | 8.44  | 8.55  | 8.66  | 8.76  | 8.85  | 8.94  | 9.03  | 0.01     |
| 9  | 9             | 5.98  | 6.09  | 6.19  | 6.28  | 6.36  | 6.44  | 6.51  | 6.58  | 6.64  | 0.05     |
|    |               | 7.78  | 7.91  | 8.03  | 8.13  | 8.23  | 8.33  | 8.41  | 8.49  | 8.57  | 0.01     |
| 10 | 10            | 5.83  | 5.93  | 6.03  | 6.11  | 6.19  | 6.27  | 6.34  | 6.40  | 6.47  | 0.05     |
|    |               | 7.49  | 7.60  | 7.71  | 7.81  | 7.91  | 7.99  | 8.08  | 8.15  | 8.23  | 0.01     |
| 11 | 11            | 5.71  | 5.81  | 5.90  | 5.98  | 6.06  | 6.13  | 6.20  | 6.27  | 6.33  | 0.05     |
|    |               | 7.25  | 7.36  | 7.46  | 7.56  | 7.65  | 7.73  | 7.81  | 7.88  | 7.95  | 0.01     |
| 12 | 12            | 5.61  | 5.71  | 5.80  | 5.86  | 5.95  | 6.02  | 6.09  | 6.15  | 6.21  | 0.05     |
|    |               | 7.06  | 7.17  | 7.26  | 7.36  | 7.44  | 7.52  | 7.59  | 7.66  | 7.73  | 0.01     |
| 13 | 13            | 5.53  | 5.63  | 5.71  | 5.79  | 5.86  | 5.93  | 5.99  | 6.05  | 6.11  | 0.05     |
|    |               | 6.90  | 7.01  | 7.10  | 7.19  | 7.27  | 7.35  | 7.42  | 7.48  | 7.55  | 0.01     |
| 14 | 14            | 5.46  | 5.55  | 5.64  | 5.71  | 5.79  | 5.85  | 5.91  | 5.97  | 6.03  | 0.05     |
|    |               | 6.77  | 6.87  | 6.96  | 7.05  | 7.13  | 7.20  | 7.27  | 7.33  | 7.39  | 0.01     |
| 15 | 15            | 5.40  | 5.49  | 5.57  | 5.65  | 5.72  | 5.78  | 5.85  | 5.90  | 5.96  | 0.05     |
|    |               | 6.66  | 6.76  | 6.84  | 6.93  | 7.00  | 7.07  | 7.14  | 7.20  | 7.26  | 0.01     |

续表

| 误差<br>df | t = 处理均值的个数 |      |      |      |      |      |      |      |      |      |  | $\alpha$ |
|----------|-------------|------|------|------|------|------|------|------|------|------|--|----------|
|          | 12          | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   |      |  |          |
| 16       | 5.35        | 5.44 | 5.52 | 5.59 | 5.66 | 5.73 | 5.79 | 5.84 | 5.90 | 0.05 |  |          |
|          | 6.56        | 6.66 | 6.74 | 6.82 | 6.90 | 6.97 | 7.03 | 7.09 | 7.15 | 0.01 |  |          |
| 17       | 5.31        | 5.39 | 5.47 | 5.54 | 5.61 | 5.67 | 5.73 | 5.79 | 5.84 | 0.05 |  |          |
|          | 6.48        | 6.57 | 6.66 | 6.73 | 6.81 | 6.87 | 6.94 | 7.00 | 7.05 | 0.01 |  |          |
| 18       | 5.27        | 5.35 | 5.43 | 5.50 | 5.57 | 5.63 | 5.69 | 5.74 | 5.79 | 0.05 |  |          |
|          | 6.41        | 6.50 | 6.58 | 6.65 | 6.73 | 6.79 | 6.85 | 6.91 | 6.97 | 0.01 |  |          |
| 19       | 5.23        | 5.31 | 5.39 | 5.46 | 5.53 | 5.59 | 5.65 | 5.70 | 5.75 | 0.05 |  |          |
|          | 6.34        | 6.43 | 6.51 | 6.58 | 6.65 | 6.72 | 6.78 | 6.84 | 6.89 | 0.01 |  |          |
| 20       | 5.20        | 5.28 | 5.36 | 5.43 | 5.49 | 5.55 | 5.61 | 5.66 | 5.71 | 0.05 |  |          |
|          | 6.28        | 6.37 | 6.45 | 6.52 | 6.59 | 6.65 | 6.71 | 6.77 | 6.82 | 0.01 |  |          |
| 24       | 5.10        | 5.18 | 5.25 | 5.32 | 5.38 | 5.44 | 5.49 | 5.55 | 5.59 | 0.05 |  |          |
|          | 6.11        | 6.19 | 6.26 | 6.33 | 6.39 | 6.45 | 6.51 | 6.56 | 6.61 | 0.01 |  |          |
| 30       | 5.00        | 5.08 | 5.15 | 5.21 | 5.27 | 5.33 | 5.38 | 5.43 | 5.47 | 0.05 |  |          |
|          | 5.93        | 6.01 | 6.08 | 6.14 | 6.20 | 6.26 | 6.31 | 6.36 | 6.41 | 0.01 |  |          |
| 40       | 4.90        | 4.98 | 5.04 | 5.11 | 5.16 | 5.22 | 5.27 | 5.31 | 5.36 | 0.05 |  |          |
|          | 5.76        | 5.83 | 5.90 | 5.96 | 6.02 | 6.07 | 6.12 | 6.16 | 6.21 | 0.01 |  |          |
| 60       | 4.81        | 4.88 | 4.94 | 5.00 | 5.06 | 5.11 | 5.15 | 5.20 | 5.24 | 0.05 |  |          |
|          | 5.60        | 5.67 | 5.73 | 5.78 | 5.84 | 5.89 | 5.93 | 5.97 | 6.01 | 0.01 |  |          |
| 120      | 4.71        | 4.78 | 4.84 | 4.90 | 4.95 | 5.00 | 5.04 | 5.09 | 5.13 | 0.05 |  |          |
|          | 5.44        | 5.50 | 5.56 | 5.61 | 5.66 | 5.71 | 5.75 | 5.79 | 5.83 | 0.01 |  |          |
| $\infty$ | 4.62        | 4.68 | 4.74 | 4.80 | 4.85 | 4.89 | 4.93 | 4.97 | 5.01 | 0.05 |  |          |
|          | 5.29        | 5.35 | 5.40 | 5.45 | 5.49 | 5.54 | 5.57 | 5.61 | 5.65 | 0.01 |  |          |

表 11 Dunnett 检验的百分位数  $d_{\alpha}(k, v)$ 

| $v$      | $\alpha = 0.05$ (单侧) |      |      |      |      |      |      |      |      |      |      |      |      |  |  |  |  |  |  |
|----------|----------------------|------|------|------|------|------|------|------|------|------|------|------|------|--|--|--|--|--|--|
|          | $k=2$                | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 15   | 20   |  |  |  |  |  |  |
| 5        | 2.44                 | 2.68 | 2.85 | 2.98 | 3.08 | 3.16 | 3.24 | 3.30 | 3.36 | 3.41 | 3.45 | 3.57 | 3.72 |  |  |  |  |  |  |
| 6        | 2.34                 | 2.56 | 2.71 | 2.83 | 2.92 | 3.00 | 3.07 | 3.12 | 3.17 | 3.22 | 3.26 | 3.37 | 3.50 |  |  |  |  |  |  |
| 7        | 2.27                 | 2.48 | 2.62 | 2.73 | 2.82 | 2.89 | 2.95 | 3.01 | 3.05 | 3.10 | 3.13 | 3.23 | 3.36 |  |  |  |  |  |  |
| 8        | 2.22                 | 2.42 | 2.55 | 2.66 | 2.74 | 2.81 | 2.87 | 2.92 | 2.96 | 3.01 | 3.04 | 3.14 | 3.25 |  |  |  |  |  |  |
| 9        | 2.18                 | 2.37 | 2.50 | 2.60 | 2.68 | 2.75 | 2.81 | 2.86 | 2.90 | 2.94 | 2.97 | 3.06 | 3.18 |  |  |  |  |  |  |
| 10       | 2.15                 | 2.34 | 2.47 | 2.56 | 2.64 | 2.70 | 2.76 | 2.81 | 2.85 | 2.89 | 2.92 | 3.01 | 3.12 |  |  |  |  |  |  |
| 11       | 2.13                 | 2.31 | 2.44 | 2.53 | 2.60 | 2.67 | 2.72 | 2.77 | 2.81 | 2.85 | 2.88 | 2.96 | 3.07 |  |  |  |  |  |  |
| 12       | 2.11                 | 2.29 | 2.41 | 2.50 | 2.58 | 2.64 | 2.69 | 2.74 | 2.78 | 2.81 | 2.84 | 2.93 | 3.03 |  |  |  |  |  |  |
| 13       | 2.09                 | 2.27 | 2.39 | 2.48 | 2.55 | 2.61 | 2.66 | 2.71 | 2.75 | 2.78 | 2.82 | 2.90 | 3.00 |  |  |  |  |  |  |
| 14       | 2.08                 | 2.25 | 2.37 | 2.46 | 2.53 | 2.59 | 2.64 | 2.69 | 2.72 | 2.76 | 2.79 | 2.87 | 2.97 |  |  |  |  |  |  |
| 15       | 2.07                 | 2.24 | 2.36 | 2.44 | 2.51 | 2.57 | 2.62 | 2.67 | 2.70 | 2.74 | 2.77 | 2.85 | 2.95 |  |  |  |  |  |  |
| 16       | 2.06                 | 2.23 | 2.34 | 2.43 | 2.50 | 2.56 | 2.61 | 2.65 | 2.69 | 2.72 | 2.75 | 2.83 | 2.93 |  |  |  |  |  |  |
| 17       | 2.05                 | 2.22 | 2.33 | 2.42 | 2.49 | 2.54 | 2.59 | 2.64 | 2.67 | 2.71 | 2.74 | 2.81 | 2.91 |  |  |  |  |  |  |
| 18       | 2.04                 | 2.21 | 2.32 | 2.41 | 2.48 | 2.53 | 2.58 | 2.62 | 2.66 | 2.69 | 2.72 | 2.80 | 2.89 |  |  |  |  |  |  |
| 19       | 2.03                 | 2.20 | 2.31 | 2.40 | 2.47 | 2.52 | 2.57 | 2.61 | 2.65 | 2.68 | 2.71 | 2.79 | 2.88 |  |  |  |  |  |  |
| 20       | 2.03                 | 2.19 | 2.30 | 2.39 | 2.46 | 2.51 | 2.56 | 2.60 | 2.64 | 2.67 | 2.70 | 2.77 | 2.87 |  |  |  |  |  |  |
| 24       | 2.01                 | 2.17 | 2.28 | 2.36 | 2.43 | 2.48 | 2.53 | 2.57 | 2.60 | 2.64 | 2.66 | 2.74 | 2.83 |  |  |  |  |  |  |
| 30       | 1.99                 | 2.15 | 2.25 | 2.33 | 2.40 | 2.45 | 2.50 | 2.54 | 2.57 | 2.60 | 2.63 | 2.70 | 2.79 |  |  |  |  |  |  |
| 40       | 1.97                 | 2.13 | 2.23 | 2.31 | 2.37 | 2.42 | 2.47 | 2.51 | 2.54 | 2.57 | 2.60 | 2.67 | 2.75 |  |  |  |  |  |  |
| 60       | 1.95                 | 2.10 | 2.21 | 2.28 | 2.35 | 2.39 | 2.44 | 2.48 | 2.51 | 2.54 | 2.56 | 2.63 | 2.72 |  |  |  |  |  |  |
| 120      | 1.93                 | 2.08 | 2.18 | 2.26 | 2.32 | 2.37 | 2.41 | 2.45 | 2.48 | 2.51 | 2.53 | 2.60 | 2.68 |  |  |  |  |  |  |
| $\infty$ | 1.92                 | 2.06 | 2.16 | 2.23 | 2.29 | 2.34 | 2.38 | 2.42 | 2.45 | 2.48 | 2.50 | 2.56 | 2.64 |  |  |  |  |  |  |

来自: C. W. Dunnett (1955), "A Multiple Comparison Procedure for Comparing Several Treatments with a Control," *Journal of the American Statistical Association* 50, 1112-1118. 经 *Journal of the American Statistical Association* 允许重印。版权所有, 1955 American Statistical Association 保留所有权利。C. W. Dunnett (1964), "New Tables for Multiple Comparisons with a Control", *Biometrics* 20, 482-491. 1980 年 C. W. Dunnett 还给出了另外的表。

续表

| $\alpha = 0.01$ (单侧) |       |      |      |      |      |      |      |      |      |      |      |      |      |  |
|----------------------|-------|------|------|------|------|------|------|------|------|------|------|------|------|--|
| $v$                  | $k=2$ | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 15   | 20   |  |
| 5                    | 3.90  | 4.21 | 4.43 | 4.60 | 4.73 | 4.85 | 4.94 | 5.03 | 5.11 | 5.17 | 5.24 | 5.39 | 5.59 |  |
| 6                    | 3.61  | 3.88 | 4.07 | 4.21 | 4.33 | 4.43 | 4.51 | 4.59 | 4.64 | 4.70 | 4.76 | 4.89 | 5.06 |  |
| 7                    | 3.42  | 3.66 | 3.83 | 3.96 | 4.07 | 4.15 | 4.23 | 4.30 | 4.35 | 4.40 | 4.45 | 4.57 | 4.72 |  |
| 8                    | 3.29  | 3.51 | 3.67 | 3.79 | 3.88 | 3.96 | 4.03 | 4.09 | 4.14 | 4.19 | 4.23 | 4.34 | 4.48 |  |
| 9                    | 3.19  | 3.40 | 3.55 | 3.66 | 3.75 | 3.82 | 3.89 | 3.94 | 3.99 | 4.04 | 4.08 | 4.18 | 4.31 |  |
| 10                   | 3.11  | 3.31 | 3.45 | 3.56 | 3.64 | 3.71 | 3.78 | 3.83 | 3.88 | 3.92 | 3.96 | 4.06 | 4.18 |  |
| 11                   | 3.06  | 3.25 | 3.38 | 3.48 | 3.56 | 3.63 | 3.69 | 3.74 | 3.79 | 3.83 | 3.86 | 3.96 | 4.08 |  |
| 12                   | 3.01  | 3.19 | 3.32 | 3.42 | 3.50 | 3.56 | 3.62 | 3.67 | 3.71 | 3.75 | 3.79 | 3.88 | 3.99 |  |
| 13                   | 2.97  | 3.15 | 3.27 | 3.37 | 3.44 | 3.51 | 3.56 | 3.61 | 3.65 | 3.69 | 3.73 | 3.81 | 3.92 |  |
| 14                   | 2.94  | 3.11 | 3.23 | 3.32 | 3.40 | 3.46 | 3.51 | 3.56 | 3.60 | 3.64 | 3.67 | 3.76 | 3.87 |  |
| 15                   | 2.91  | 3.08 | 3.20 | 3.29 | 3.36 | 3.42 | 3.47 | 3.52 | 3.56 | 3.60 | 3.63 | 3.71 | 3.82 |  |
| 16                   | 2.88  | 3.05 | 3.17 | 3.26 | 3.33 | 3.39 | 3.44 | 3.48 | 3.52 | 3.56 | 3.59 | 3.67 | 3.78 |  |
| 17                   | 2.86  | 3.03 | 3.14 | 3.23 | 3.30 | 3.36 | 3.41 | 3.45 | 3.49 | 3.53 | 3.56 | 3.64 | 3.74 |  |
| 18                   | 2.84  | 3.01 | 3.12 | 3.21 | 3.27 | 3.33 | 3.38 | 3.42 | 3.46 | 3.50 | 3.53 | 3.61 | 3.71 |  |
| 19                   | 2.83  | 2.99 | 3.10 | 3.18 | 3.25 | 3.31 | 3.36 | 3.40 | 3.44 | 3.47 | 3.50 | 3.58 | 3.68 |  |
| 20                   | 2.81  | 2.97 | 3.08 | 3.17 | 3.23 | 3.29 | 3.34 | 3.38 | 3.42 | 3.45 | 3.48 | 3.56 | 3.65 |  |
| 24                   | 2.77  | 2.92 | 3.03 | 3.11 | 3.17 | 3.22 | 3.27 | 3.31 | 3.35 | 3.38 | 3.41 | 3.48 | 3.57 |  |
| 30                   | 2.72  | 2.87 | 2.97 | 3.05 | 3.11 | 3.16 | 3.21 | 3.24 | 3.28 | 3.31 | 3.34 | 3.41 | 3.50 |  |
| 40                   | 2.68  | 2.82 | 2.92 | 2.99 | 3.05 | 3.10 | 3.14 | 3.18 | 3.21 | 3.24 | 3.27 | 3.34 | 3.42 |  |
| 60                   | 2.64  | 2.78 | 2.87 | 2.94 | 3.00 | 3.04 | 3.08 | 3.12 | 3.15 | 3.18 | 3.20 | 3.27 | 3.35 |  |
| 120                  | 2.60  | 2.73 | 2.82 | 2.89 | 2.94 | 2.99 | 3.03 | 3.06 | 3.09 | 3.12 | 3.14 | 3.20 | 3.28 |  |
| $\infty$             | 2.56  | 2.68 | 2.77 | 2.84 | 2.89 | 2.93 | 2.97 | 3.00 | 3.03 | 3.06 | 3.08 | 3.14 | 3.21 |  |

续表

| $\alpha = 0.05$ (双侧) |       |      |      |      |      |      |      |      |      |      |      |      |      |
|----------------------|-------|------|------|------|------|------|------|------|------|------|------|------|------|
| $v$                  | $k=2$ | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 15   | 20   |
| 5                    | 3.03  | 3.29 | 3.48 | 3.62 | 3.73 | 3.82 | 3.90 | 3.97 | 4.03 | 4.09 | 4.14 | 4.26 | 4.42 |
| 6                    | 2.86  | 3.10 | 3.26 | 3.39 | 3.49 | 3.57 | 3.64 | 3.71 | 3.76 | 3.81 | 3.86 | 3.97 | 4.11 |
| 7                    | 2.75  | 2.97 | 3.12 | 3.24 | 3.33 | 3.41 | 3.47 | 3.53 | 3.58 | 3.63 | 3.67 | 3.78 | 3.91 |
| 8                    | 2.67  | 2.88 | 3.02 | 3.13 | 3.22 | 3.29 | 3.35 | 3.41 | 3.46 | 3.50 | 3.54 | 3.64 | 3.76 |
| 9                    | 2.61  | 2.81 | 2.95 | 3.05 | 3.14 | 3.20 | 3.26 | 3.32 | 3.36 | 3.40 | 3.44 | 3.53 | 3.65 |
| 10                   | 2.57  | 2.76 | 2.89 | 2.99 | 3.07 | 3.14 | 3.19 | 3.24 | 3.29 | 3.33 | 3.36 | 3.45 | 3.57 |
| 11                   | 2.53  | 2.72 | 2.84 | 2.94 | 3.02 | 3.08 | 3.14 | 3.19 | 3.23 | 3.27 | 3.30 | 3.39 | 3.50 |
| 12                   | 2.50  | 2.68 | 2.81 | 2.90 | 2.98 | 3.04 | 3.09 | 3.14 | 3.18 | 3.22 | 3.25 | 3.34 | 3.45 |
| 13                   | 2.48  | 2.65 | 2.78 | 2.87 | 2.94 | 3.00 | 3.06 | 3.10 | 3.14 | 3.18 | 3.21 | 3.29 | 3.40 |
| 14                   | 2.46  | 2.63 | 2.75 | 2.84 | 2.91 | 2.97 | 3.02 | 3.07 | 3.11 | 3.14 | 3.18 | 3.26 | 3.36 |
| 15                   | 2.44  | 2.61 | 2.73 | 2.82 | 2.89 | 2.95 | 3.00 | 3.04 | 3.08 | 3.12 | 3.15 | 3.23 | 3.33 |
| 16                   | 2.42  | 2.59 | 2.71 | 2.80 | 2.87 | 2.92 | 2.97 | 3.02 | 3.06 | 3.09 | 3.12 | 3.20 | 3.30 |
| 17                   | 2.41  | 2.58 | 2.69 | 2.78 | 2.85 | 2.90 | 2.95 | 3.00 | 3.03 | 3.07 | 3.10 | 3.18 | 3.27 |
| 18                   | 2.40  | 2.56 | 2.68 | 2.76 | 2.83 | 2.89 | 2.94 | 2.98 | 3.01 | 3.05 | 3.08 | 3.16 | 3.25 |
| 19                   | 2.39  | 2.55 | 2.66 | 2.75 | 2.81 | 2.87 | 2.92 | 2.96 | 3.00 | 3.03 | 3.06 | 3.14 | 3.23 |
| 20                   | 2.38  | 2.54 | 2.65 | 2.73 | 2.80 | 2.86 | 2.90 | 2.95 | 2.98 | 3.02 | 3.05 | 3.12 | 3.22 |
| 24                   | 2.35  | 2.51 | 2.61 | 2.70 | 2.76 | 2.81 | 2.86 | 2.90 | 2.94 | 2.97 | 3.00 | 3.07 | 3.16 |
| 30                   | 2.32  | 2.47 | 2.58 | 2.66 | 2.72 | 2.77 | 2.82 | 2.86 | 2.89 | 2.92 | 2.95 | 3.02 | 3.11 |
| 40                   | 2.29  | 2.44 | 2.54 | 2.62 | 2.68 | 2.73 | 2.77 | 2.81 | 2.85 | 2.87 | 2.90 | 2.97 | 3.06 |
| 60                   | 2.27  | 2.41 | 2.51 | 2.58 | 2.64 | 2.69 | 2.73 | 2.77 | 2.80 | 2.83 | 2.86 | 2.92 | 3.00 |
| 120                  | 2.24  | 2.38 | 2.47 | 2.55 | 2.60 | 2.65 | 2.69 | 2.73 | 2.76 | 2.79 | 2.81 | 2.87 | 2.95 |
| $\infty$             | 2.21  | 2.35 | 2.44 | 2.51 | 2.57 | 2.61 | 2.65 | 2.69 | 2.72 | 2.74 | 2.77 | 2.83 | 2.91 |

续表

| $\alpha = 0.01$ (双侧) |       |      |      |      |      |      |      |      |      |      |      |      |      |  |
|----------------------|-------|------|------|------|------|------|------|------|------|------|------|------|------|--|
| $v$                  | $k=2$ | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 15   | 20   |  |
| 5                    | 4.63  | 4.98 | 5.22 | 5.41 | 5.56 | 5.69 | 5.80 | 5.89 | 5.98 | 6.05 | 6.12 | 6.30 | 6.52 |  |
| 6                    | 4.21  | 4.51 | 4.71 | 4.87 | 5.00 | 5.10 | 5.20 | 5.28 | 5.35 | 5.41 | 5.47 | 5.62 | 5.81 |  |
| 7                    | 3.95  | 4.21 | 4.39 | 4.53 | 4.64 | 4.74 | 4.82 | 4.89 | 4.95 | 5.01 | 5.06 | 5.19 | 5.36 |  |
| 8                    | 3.77  | 4.00 | 4.17 | 4.29 | 4.40 | 4.48 | 4.56 | 4.62 | 4.68 | 4.73 | 4.78 | 4.90 | 5.05 |  |
| 9                    | 3.63  | 3.85 | 4.01 | 4.12 | 4.22 | 4.30 | 4.37 | 4.43 | 4.48 | 4.53 | 4.57 | 4.68 | 4.82 |  |
| 10                   | 3.53  | 3.74 | 3.88 | 3.99 | 4.08 | 4.16 | 4.22 | 4.28 | 4.33 | 4.37 | 4.42 | 4.52 | 4.65 |  |
| 11                   | 3.45  | 3.65 | 3.79 | 3.89 | 3.98 | 4.05 | 4.11 | 4.16 | 4.21 | 4.25 | 4.29 | 4.39 | 4.52 |  |
| 12                   | 3.39  | 3.58 | 3.71 | 3.81 | 3.89 | 3.96 | 4.02 | 4.07 | 4.12 | 4.16 | 4.19 | 4.29 | 4.41 |  |
| 13                   | 3.33  | 3.52 | 3.65 | 3.74 | 3.82 | 3.89 | 3.94 | 3.99 | 4.04 | 4.08 | 4.11 | 4.20 | 4.32 |  |
| 14                   | 3.29  | 3.47 | 3.59 | 3.69 | 3.76 | 3.83 | 3.88 | 3.93 | 3.97 | 4.01 | 4.05 | 4.13 | 4.24 |  |
| 15                   | 3.25  | 3.43 | 3.55 | 3.64 | 3.71 | 3.78 | 3.83 | 3.88 | 3.92 | 3.95 | 3.99 | 4.07 | 4.18 |  |
| 16                   | 3.22  | 3.39 | 3.51 | 3.60 | 3.67 | 3.73 | 3.78 | 3.83 | 3.87 | 3.91 | 3.94 | 4.02 | 4.13 |  |
| 17                   | 3.19  | 3.36 | 3.47 | 3.56 | 3.63 | 3.69 | 3.74 | 3.79 | 3.83 | 3.86 | 3.90 | 3.98 | 4.08 |  |
| 18                   | 3.17  | 3.33 | 3.44 | 3.53 | 3.60 | 3.66 | 3.71 | 3.75 | 3.79 | 3.83 | 3.86 | 3.94 | 4.04 |  |
| 19                   | 3.15  | 3.31 | 3.42 | 3.50 | 3.57 | 3.63 | 3.68 | 3.72 | 3.76 | 3.79 | 3.83 | 3.90 | 4.00 |  |
| 20                   | 3.13  | 3.29 | 3.40 | 3.48 | 3.55 | 3.60 | 3.65 | 3.69 | 3.73 | 3.77 | 3.80 | 3.87 | 3.97 |  |
| 24                   | 3.07  | 3.22 | 3.32 | 3.40 | 3.47 | 3.52 | 3.57 | 3.61 | 3.64 | 3.68 | 3.70 | 3.78 | 3.87 |  |
| 30                   | 3.01  | 3.15 | 3.25 | 3.33 | 3.39 | 3.44 | 3.49 | 3.52 | 3.56 | 3.59 | 3.62 | 3.69 | 3.78 |  |
| 40                   | 2.95  | 3.09 | 3.19 | 3.26 | 3.32 | 3.37 | 3.41 | 3.44 | 3.48 | 3.51 | 3.53 | 3.60 | 3.68 |  |
| 60                   | 2.90  | 3.03 | 3.12 | 3.19 | 3.25 | 3.29 | 3.33 | 3.37 | 3.40 | 3.42 | 3.45 | 3.51 | 3.59 |  |
| 120                  | 2.85  | 2.97 | 3.06 | 3.12 | 3.18 | 3.22 | 3.26 | 3.29 | 3.32 | 3.35 | 3.37 | 3.43 | 3.51 |  |
| $\infty$             | 2.79  | 2.92 | 3.00 | 3.06 | 3.11 | 3.15 | 3.19 | 3.22 | 3.25 | 3.27 | 3.29 | 3.35 | 3.42 |  |





续表

|          |      | 上 1% 数 |      |      |      |       |       |       |       |       |       |    |  |
|----------|------|--------|------|------|------|-------|-------|-------|-------|-------|-------|----|--|
| $df_2$   | $t$  | 2      | 3    | 4    | 5    | 6     | 7     | 8     | 9     | 10    | 11    | 12 |  |
| 2        | 199  | 448    | 729  | 1036 | 1362 | 1705  | 2063  | 2432  | 2813  | 3204  | 3605  |    |  |
| 3        | 47.5 | 85     | 120  | 151  | 184  | 21(6) | 24(9) | 28(1) | 31(0) | 33(7) | 36(1) |    |  |
| 4        | 23.2 | 37     | 49   | 59   | 69   | 79    | 89    | 97    | 106   | 113   | 120   |    |  |
| 5        | 14.9 | 22     | 28   | 33   | 38   | 42    | 46    | 50    | 54    | 57    | 60    |    |  |
| 6        | 11.1 | 15.5   | 19.1 | 22   | 25   | 27    | 30    | 32    | 34    | 36    | 37    |    |  |
| 7        | 8.89 | 12.1   | 14.5 | 16.5 | 18.4 | 20    | 22    | 23    | 24    | 26    | 27    |    |  |
| 8        | 7.50 | 9.9    | 11.7 | 13.2 | 14.5 | 15.8  | 16.6  | 17.9  | 18.9  | 19.8  | 21    |    |  |
| 9        | 6.54 | 8.5    | 9.9  | 11.1 | 12.1 | 13.1  | 13.9  | 14.7  | 15.3  | 16.0  | 16.6  |    |  |
| 10       | 5.85 | 7.4    | 8.6  | 9.6  | 10.4 | 11.1  | 11.8  | 12.4  | 12.9  | 13.4  | 13.9  |    |  |
| 12       | 4.91 | 6.1    | 6.9  | 7.6  | 8.2  | 8.7   | 9.1   | 9.5   | 9.9   | 10.2  | 10.6  |    |  |
| 15       | 4.07 | 4.9    | 5.5  | 6.0  | 6.4  | 6.7   | 7.1   | 7.3   | 7.5   | 7.8   | 8.0   |    |  |
| 20       | 3.32 | 3.8    | 4.3  | 4.6  | 4.9  | 5.1   | 5.3   | 5.5   | 5.6   | 5.8   | 5.9   |    |  |
| 30       | 2.63 | 3.0    | 3.3  | 3.4  | 3.6  | 3.7   | 3.8   | 3.9   | 4.0   | 4.1   | 4.2   |    |  |
| 60       | 1.96 | 2.2    | 2.3  | 2.4  | 2.4  | 2.5   | 2.5   | 2.6   | 2.6   | 2.7   | 2.7   |    |  |
| $\infty$ | 1.00 | 1.0    | 1.0  | 1.0  | 1.0  | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   | 1.0   |    |  |

$s_{\max}^2$  和  $s_{\min}^2$  分别为自由度为  $df_2 = n - 1$  的  $t$  个独立的均方中的最大和最小值。在  $t = 2$  的列中,  $df_2 = 2$  和  $\infty$  的行的值是准确的。其他地方, 5% 分位数的第三位数字可能有一两个单位的误差, 而 1% 分位数的第三位数字可能有几个单位的误差。  $df_2 = 3$  行中括号内的第三位数字是最不精确的。来自 *Biometrika Tables for Statisticians*, 3rd ed., Vol. 1, E. S. Pearson and H. O. Hartley 主编 (New York: Cambridge University Press, 1966), Table, p. 202. 经 *Biometrika* 主管允许重印。

表 13 随机数

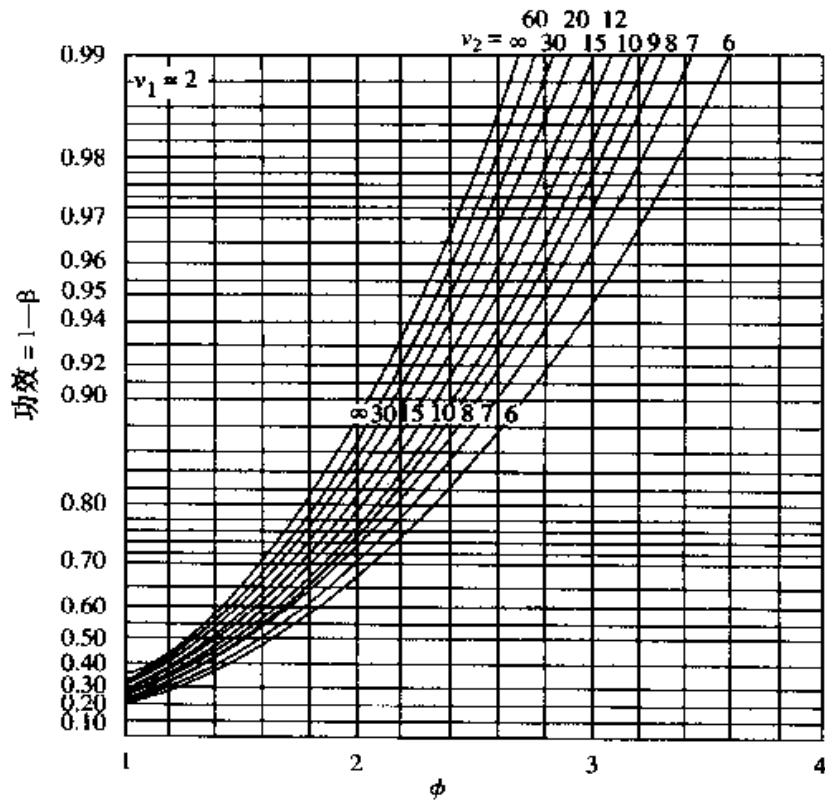
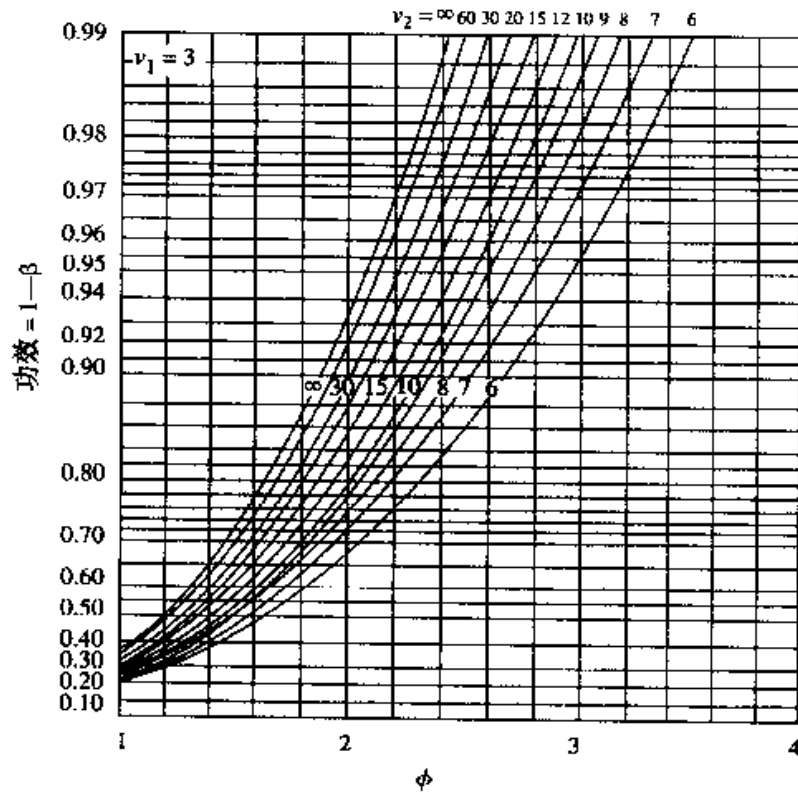
| 行<br>列 | (1)   | (2)   | (3)   | (4)   | (5)   | (6)   | (7)   | (8)   | (9)   | (10)  | (11)  | (12)  | (13)  | (14)  |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1      | 10480 | 15011 | 01536 | 02011 | 81647 | 91646 | 69179 | 14194 | 62590 | 36207 | 20969 | 99570 | 91291 | 90700 |
| 2      | 22368 | 46573 | 25595 | 85393 | 30995 | 89198 | 27982 | 53402 | 93965 | 34095 | 52666 | 19174 | 39615 | 99505 |
| 3      | 24130 | 48360 | 22527 | 97265 | 76393 | 64809 | 15179 | 24830 | 49340 | 32081 | 30680 | 19655 | 63348 | 58629 |
| 4      | 42167 | 93093 | 06243 | 61680 | 07856 | 16376 | 39440 | 53537 | 71341 | 57004 | 00849 | 74917 | 97758 | 16379 |
| 5      | 37570 | 39975 | 81837 | 16656 | 06121 | 91782 | 60468 | 81305 | 49684 | 60672 | 14110 | 06927 | 01263 | 54613 |
| 6      | 77921 | 06907 | 11008 | 42751 | 27756 | 53498 | 18602 | 70659 | 90655 | 15053 | 21916 | 81825 | 44394 | 42880 |
| 7      | 99562 | 72905 | 56420 | 69994 | 98872 | 31016 | 71194 | 18738 | 44013 | 48840 | 63213 | 21069 | 10634 | 12952 |
| 8      | 96301 | 91977 | 05463 | 07972 | 18876 | 20922 | 94595 | 56869 | 69014 | 60045 | 18425 | 84903 | 42508 | 32307 |
| 9      | 89579 | 14342 | 63661 | 10281 | 17453 | 18103 | 57740 | 84378 | 25331 | 12566 | 58678 | 44947 | 05585 | 56941 |
| 10     | 85475 | 36857 | 53342 | 53988 | 53060 | 59533 | 38867 | 62300 | 08158 | 17983 | 16439 | 11458 | 18593 | 64952 |
| 11     | 28918 | 69578 | 88231 | 33276 | 70997 | 79936 | 56865 | 05859 | 90106 | 31595 | 01547 | 85590 | 91610 | 78188 |
| 12     | 63553 | 40961 | 48235 | 03427 | 49626 | 69445 | 18663 | 72695 | 52180 | 20847 | 12234 | 90511 | 33703 | 90322 |
| 13     | 09429 | 93969 | 52636 | 92737 | 88974 | 33488 | 36320 | 17617 | 30015 | 08272 | 84115 | 27156 | 30613 | 74952 |
| 14     | 10365 | 61129 | 87529 | 85689 | 48237 | 52267 | 67689 | 93394 | 01511 | 26358 | 85104 | 20285 | 29975 | 89868 |
| 15     | 07119 | 97336 | 71048 | 08178 | 77233 | 13916 | 47564 | 81056 | 97735 | 85977 | 29372 | 74461 | 28551 | 90707 |

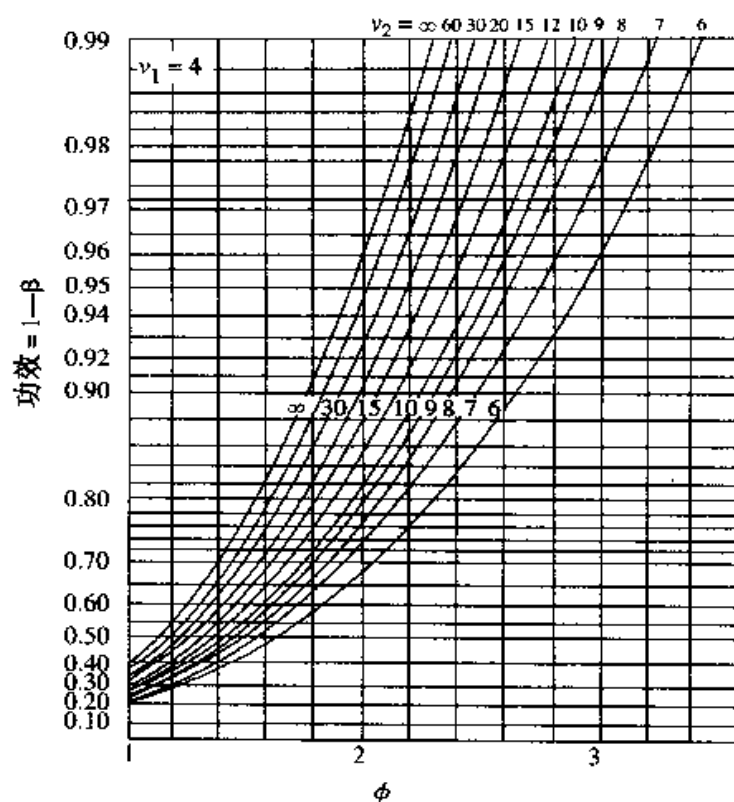
续表

| 行<br>列 | (1)   | (2)   | (3)   | (4)   | (5)   | (6)   | (7)   | (8)   | (9)   | (10)  | (11)  | (12)  | (13)  | (14)  |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 16     | 51085 | 12765 | 51821 | 51259 | 77452 | 16308 | 60756 | 92144 | 49442 | 53900 | 70960 | 63990 | 75601 | 40719 |
| 17     | 02368 | 21382 | 52404 | 60268 | 89368 | 19885 | 55322 | 44819 | 01188 | 65255 | 64835 | 44919 | 05944 | 55157 |
| 18     | 01011 | 54092 | 33362 | 94904 | 31273 | 04146 | 18594 | 29852 | 71585 | 85030 | 51132 | 01915 | 92747 | 64951 |
| 19     | 52162 | 53916 | 46369 | 58586 | 23216 | 14513 | 83149 | 98736 | 23495 | 64350 | 94738 | 17752 | 35156 | 35749 |
| 20     | 07056 | 97628 | 33787 | 09998 | 42698 | 06691 | 76988 | 13602 | 51851 | 46104 | 88916 | 19509 | 25625 | 58104 |
| 21     | 48663 | 91245 | 85828 | 14346 | 09172 | 30168 | 90229 | 04734 | 59193 | 22178 | 30421 | 61666 | 99904 | 32812 |
| 22     | 54164 | 58492 | 22421 | 74103 | 47070 | 25306 | 76468 | 26384 | 58151 | 06646 | 21524 | 15227 | 96909 | 44592 |
| 23     | 32639 | 32363 | 05597 | 24200 | 13363 | 38005 | 94342 | 28728 | 35806 | 06912 | 17012 | 64161 | 18296 | 22851 |
| 24     | 29334 | 27001 | 87637 | 87308 | 58731 | 00256 | 45834 | 15398 | 46557 | 41135 | 10367 | 07684 | 36188 | 18510 |
| 25     | 02488 | 33062 | 28834 | 07351 | 19731 | 92420 | 60952 | 61280 | 50001 | 67658 | 32586 | 86679 | 50720 | 94953 |

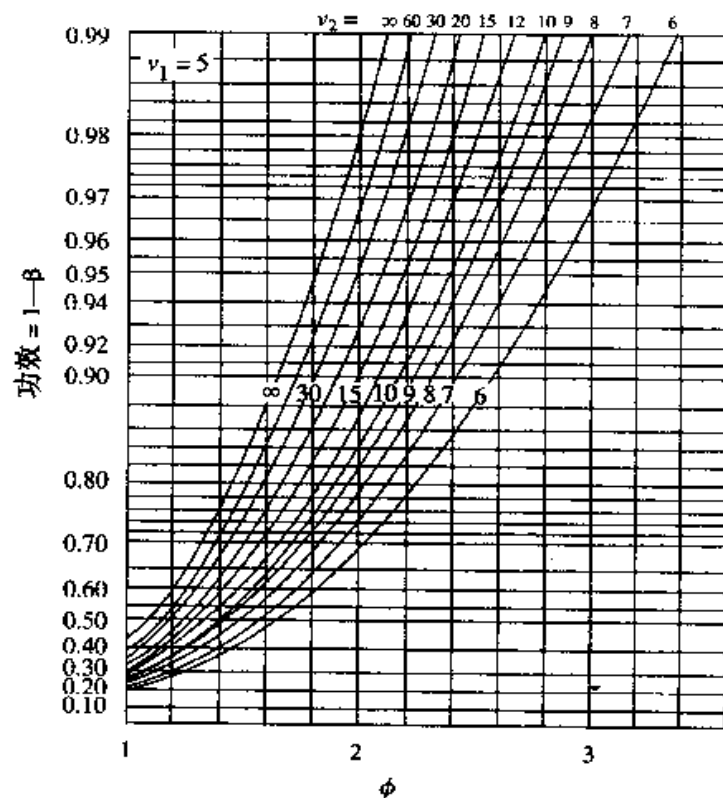
选自: William H. Beyer, 主编: *Handbook of Tables for Probability and Statistics*, 第2版, The Chemical Rubber Co., 1968. 经 CRC 出版公司的许可使用。

表 14 方差分析表中检验的功效

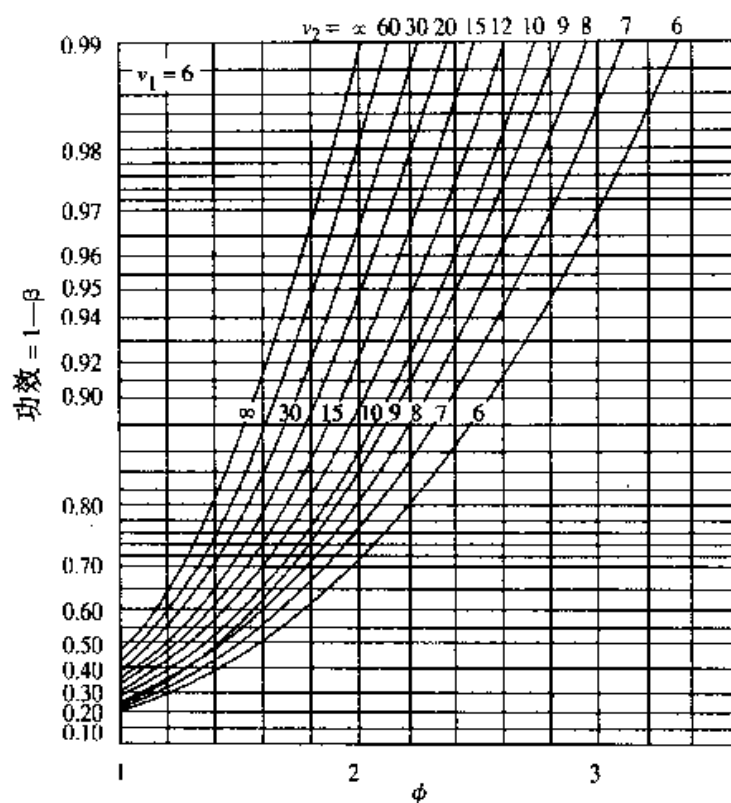
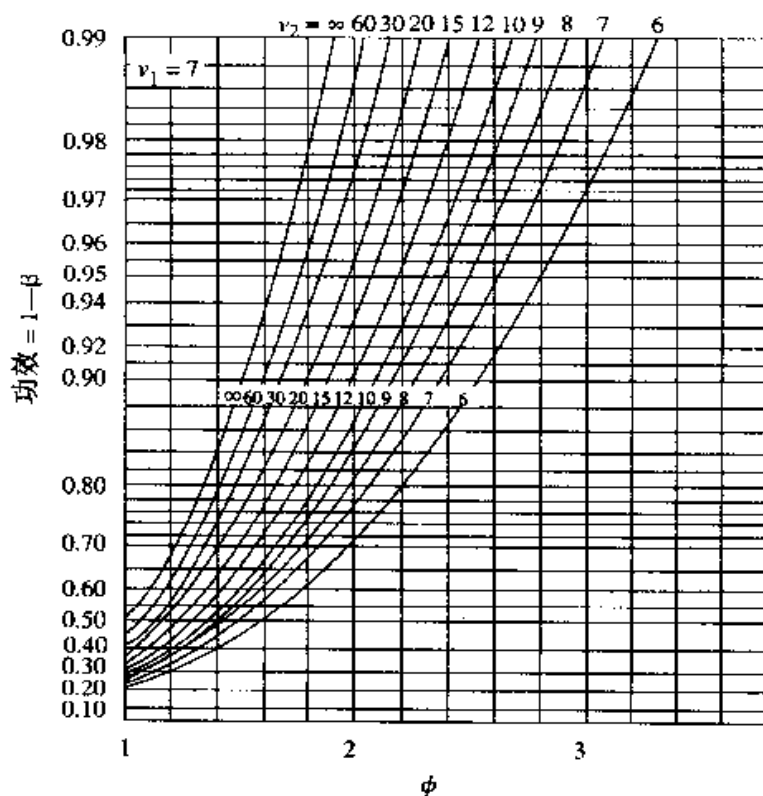
方差分析中检验的功效( $\alpha = 0.05, t = 3$ )方差分析中检验的功效( $\alpha = 0.05, t = 4$ )

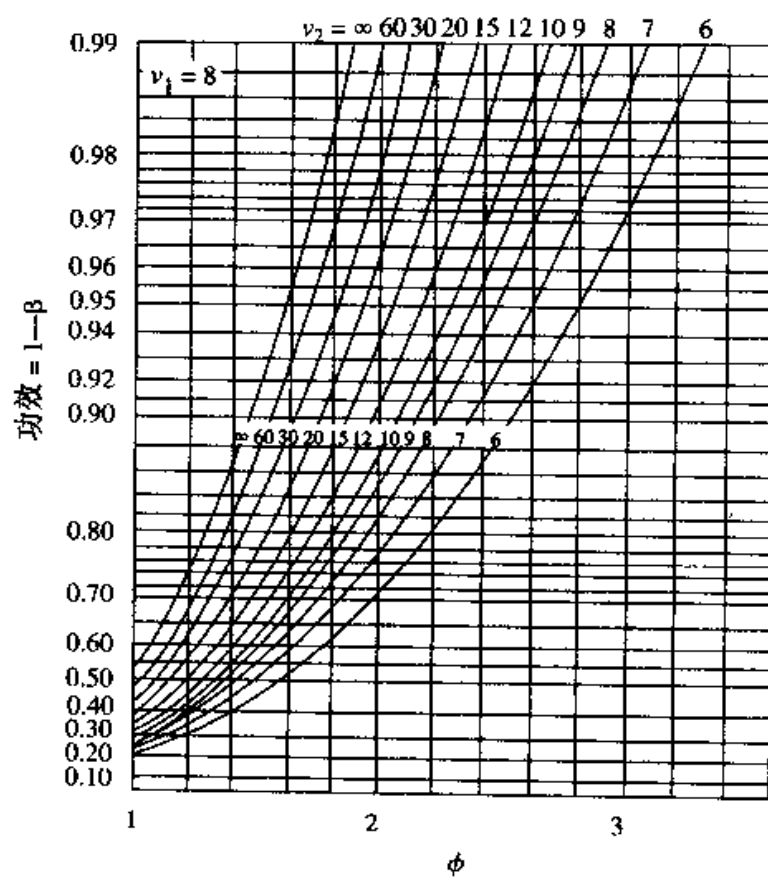


方差分析中检验的功效( $\alpha = 0.05, t = 5$ )



方差分析中检验的功效( $\alpha = 0.05, t = 6$ )

方差分析中检验的功效( $\alpha = 0.05, t = 7$ )方差分析中检验的功效( $\alpha = 0.05, t = 8$ )



方差分析中检验的功效( $\alpha = 0.05, t = 9$ )

表 15 泊松概率( $\mu$  在 0.1 与 4.0 之间)

| $y$ | $\mu$  |        |        |        |        |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|     | 0.1    | 0.2    | 0.3    | 0.4    | 0.5    | 0.6    | 0.7    | 0.8    | 0.9    | 1.0    |
| 0   | 0.9048 | 0.8187 | 0.7408 | 0.6703 | 0.6065 | 0.5488 | 0.4966 | 0.4493 | 0.4066 | 0.3679 |
| 1   | 0.0905 | 0.1637 | 0.2222 | 0.2681 | 0.3033 | 0.3293 | 0.3476 | 0.3595 | 0.3659 | 0.3679 |
| 2   | 0.0045 | 0.0164 | 0.0333 | 0.0536 | 0.0758 | 0.0988 | 0.1217 | 0.1438 | 0.1647 | 0.1839 |
| 3   | 0.0002 | 0.0011 | 0.0033 | 0.0072 | 0.0126 | 0.0198 | 0.0284 | 0.0383 | 0.0494 | 0.0613 |
| 4   | 0.0000 | 0.0001 | 0.0003 | 0.0007 | 0.0016 | 0.0030 | 0.0050 | 0.0077 | 0.0111 | 0.0153 |
| 5   | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0004 | 0.0007 | 0.0012 | 0.0020 | 0.0031 |
| 6   | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0003 | 0.0005 |

| $y$ | $\mu$  |        |        |        |        |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|     | 1.1    | 1.2    | 1.3    | 1.4    | 1.5    | 1.6    | 1.7    | 1.8    | 1.9    | 2.0    |
| 0   | 0.3329 | 0.3012 | 0.2725 | 0.2466 | 0.2231 | 0.2019 | 0.1827 | 0.1653 | 0.1496 | 0.1353 |
| 1   | 0.3662 | 0.3614 | 0.3543 | 0.3452 | 0.3347 | 0.3230 | 0.3106 | 0.2975 | 0.2842 | 0.2707 |
| 2   | 0.2014 | 0.2169 | 0.2303 | 0.2417 | 0.2510 | 0.2584 | 0.2640 | 0.2678 | 0.2700 | 0.2707 |
| 3   | 0.0738 | 0.0867 | 0.0998 | 0.1128 | 0.1255 | 0.1378 | 0.1496 | 0.1607 | 0.1710 | 0.1804 |
| 4   | 0.0203 | 0.0260 | 0.0324 | 0.0395 | 0.0471 | 0.0551 | 0.0636 | 0.0723 | 0.0812 | 0.0902 |
| 5   | 0.0045 | 0.0062 | 0.0084 | 0.0111 | 0.0141 | 0.0176 | 0.0216 | 0.0260 | 0.0309 | 0.0361 |
| 6   | 0.0008 | 0.0012 | 0.0018 | 0.0026 | 0.0035 | 0.0047 | 0.0061 | 0.0078 | 0.0098 | 0.0120 |
| 7   | 0.0001 | 0.0002 | 0.0003 | 0.0005 | 0.0008 | 0.0011 | 0.0015 | 0.0020 | 0.0027 | 0.0034 |
| 8   | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0003 | 0.0005 | 0.0006 | 0.0009 |



续表

| y  | $\mu$  |        |        |        |        |        |        |        |        |        |  |
|----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--|
|    | 2.1    | 2.2    | 2.3    | 2.4    | 2.5    | 2.6    | 2.7    | 2.8    | 2.9    | 3.0    |  |
| 0  | 0.1225 | 0.1108 | 0.1003 | 0.0907 | 0.0821 | 0.0743 | 0.0672 | 0.0608 | 0.0550 | 0.0498 |  |
| 1  | 0.2572 | 0.2438 | 0.2306 | 0.2177 | 0.2052 | 0.1931 | 0.1815 | 0.1703 | 0.1596 | 0.1494 |  |
| 2  | 0.2700 | 0.2681 | 0.2652 | 0.2613 | 0.2565 | 0.2510 | 0.2450 | 0.2384 | 0.2314 | 0.2240 |  |
| 3  | 0.1890 | 0.1966 | 0.2033 | 0.2090 | 0.2138 | 0.2176 | 0.2205 | 0.2225 | 0.2237 | 0.2240 |  |
| 4  | 0.0992 | 0.1082 | 0.1169 | 0.1254 | 0.1336 | 0.1414 | 0.1488 | 0.1557 | 0.1622 | 0.1680 |  |
| 5  | 0.0417 | 0.0476 | 0.0538 | 0.0602 | 0.0668 | 0.0735 | 0.0804 | 0.0872 | 0.0940 | 0.1008 |  |
| 6  | 0.0146 | 0.0174 | 0.0206 | 0.0241 | 0.0278 | 0.0319 | 0.0362 | 0.0407 | 0.0455 | 0.0504 |  |
| 7  | 0.0044 | 0.0055 | 0.0068 | 0.0083 | 0.0099 | 0.0118 | 0.0139 | 0.0163 | 0.0188 | 0.0216 |  |
| 8  | 0.0011 | 0.0015 | 0.0019 | 0.0025 | 0.0031 | 0.0038 | 0.0047 | 0.0057 | 0.0068 | 0.0081 |  |
| 9  | 0.0003 | 0.0004 | 0.0005 | 0.0007 | 0.0009 | 0.0011 | 0.0014 | 0.0018 | 0.0022 | 0.0027 |  |
| 10 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 | 0.0003 | 0.0004 | 0.0005 | 0.0006 | 0.0008 |  |
| 11 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 |  |

| y | $\mu$  |        |        |        |        |        |        |        |        |        |
|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|   | 3.1    | 3.2    | 3.3    | 3.4    | 3.5    | 3.6    | 3.7    | 3.8    | 3.9    | 4.0    |
| 0 | 0.0450 | 0.0408 | 0.0369 | 0.0334 | 0.0302 | 0.0273 | 0.0247 | 0.0224 | 0.0202 | 0.0183 |
| 1 | 0.1397 | 0.1304 | 0.1217 | 0.1135 | 0.1057 | 0.0984 | 0.0915 | 0.0850 | 0.0789 | 0.0733 |
| 2 | 0.2165 | 0.2087 | 0.2008 | 0.1929 | 0.1850 | 0.1771 | 0.1692 | 0.1615 | 0.1539 | 0.1465 |
| 3 | 0.2237 | 0.2226 | 0.2209 | 0.2186 | 0.2158 | 0.2125 | 0.2087 | 0.2046 | 0.2001 | 0.1954 |
| 4 | 0.1733 | 0.1781 | 0.1823 | 0.1858 | 0.1888 | 0.1912 | 0.1931 | 0.1944 | 0.1951 | 0.1954 |
| 5 | 0.1075 | 0.1140 | 0.1203 | 0.1264 | 0.1322 | 0.1377 | 0.1429 | 0.1477 | 0.1522 | 0.1563 |
| 6 | 0.0555 | 0.0608 | 0.0662 | 0.0716 | 0.0771 | 0.0826 | 0.0881 | 0.0936 | 0.0989 | 0.1042 |

表 15 Poisson 概率( $\mu$  在 3.1 与 10.0 之间)

| $y$ | $\mu$  |        |        |        |        |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|     | 3.1    | 3.2    | 3.3    | 3.4    | 3.5    | 3.6    | 3.7    | 3.8    | 3.9    | 4.0    |
| 7   | 0.0246 | 0.0278 | 0.0312 | 0.0348 | 0.0385 | 0.0425 | 0.0466 | 0.0508 | 0.0551 | 0.0595 |
| 8   | 0.0095 | 0.0111 | 0.0129 | 0.0148 | 0.0169 | 0.0191 | 0.0215 | 0.0241 | 0.0269 | 0.0298 |
| 9   | 0.0033 | 0.0040 | 0.0047 | 0.0056 | 0.0066 | 0.0076 | 0.0089 | 0.0102 | 0.0116 | 0.0132 |
| 10  | 0.0010 | 0.0013 | 0.0016 | 0.0019 | 0.0023 | 0.0028 | 0.0033 | 0.0039 | 0.0045 | 0.0053 |
| 11  | 0.0003 | 0.0004 | 0.0005 | 0.0006 | 0.0007 | 0.0009 | 0.0011 | 0.0013 | 0.0016 | 0.0019 |
| 12  | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 | 0.0003 | 0.0003 | 0.0004 | 0.0005 | 0.0006 |
| 13  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 |

| $y$ | $\mu$  |        |        |        |        |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|     | 4.1    | 4.2    | 4.3    | 4.4    | 4.5    | 4.6    | 4.7    | 4.8    | 4.9    | 5.0    |
| 0   | 0.0166 | 0.0150 | 0.0136 | 0.0123 | 0.0111 | 0.0101 | 0.0091 | 0.0082 | 0.0074 | 0.0067 |
| 1   | 0.0679 | 0.0630 | 0.0583 | 0.0540 | 0.0500 | 0.0462 | 0.0427 | 0.0395 | 0.0365 | 0.0337 |
| 2   | 0.1393 | 0.1323 | 0.1254 | 0.1188 | 0.1125 | 0.1063 | 0.1005 | 0.0948 | 0.0894 | 0.0842 |
| 3   | 0.1904 | 0.1852 | 0.1798 | 0.1743 | 0.1687 | 0.1631 | 0.1574 | 0.1517 | 0.1460 | 0.1404 |
| 4   | 0.1951 | 0.1944 | 0.1933 | 0.1917 | 0.1898 | 0.1875 | 0.1849 | 0.1820 | 0.1789 | 0.1755 |
| 5   | 0.1600 | 0.1633 | 0.1662 | 0.1687 | 0.1708 | 0.1725 | 0.1738 | 0.1747 | 0.1753 | 0.1755 |
| 6   | 0.1093 | 0.1143 | 0.1191 | 0.1237 | 0.1281 | 0.1323 | 0.1362 | 0.1398 | 0.1432 | 0.1462 |
| 7   | 0.0640 | 0.0686 | 0.0732 | 0.0778 | 0.0824 | 0.0869 | 0.0914 | 0.0959 | 0.1002 | 0.1044 |
| 8   | 0.0328 | 0.0360 | 0.0393 | 0.0428 | 0.0463 | 0.0500 | 0.0537 | 0.0575 | 0.0614 | 0.0653 |
| 9   | 0.0150 | 0.0168 | 0.0188 | 0.0209 | 0.0232 | 0.0255 | 0.0281 | 0.0307 | 0.0334 | 0.0363 |
| 10  | 0.0061 | 0.0071 | 0.0081 | 0.0092 | 0.0104 | 0.0118 | 0.0132 | 0.0147 | 0.0164 | 0.0181 |
| 11  | 0.0023 | 0.0027 | 0.0032 | 0.0037 | 0.0043 | 0.0049 | 0.0056 | 0.0064 | 0.0073 | 0.0082 |
| 12  | 0.0008 | 0.0009 | 0.0011 | 0.0013 | 0.0016 | 0.0019 | 0.0022 | 0.0026 | 0.0030 | 0.0034 |
| 13  | 0.0002 | 0.0003 | 0.0004 | 0.0005 | 0.0006 | 0.0007 | 0.0008 | 0.0009 | 0.0011 | 0.0013 |
| 14  | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0002 | 0.0002 | 0.0003 | 0.0003 | 0.0004 | 0.0005 |
| 15  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0001 | 0.0002 |

续表

| $y$ | 5.5    | 6.0    | 6.5    | 7.0    | 7.5    | 8.0    | 8.5    | 9.0    | 9.5    | 10.0   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0   | 0.0041 | 0.0025 | 0.0015 | 0.0009 | 0.0006 | 0.0003 | 0.0002 | 0.0001 | 0.0001 | 0.0000 |
| 1   | 0.0225 | 0.0149 | 0.0098 | 0.0064 | 0.0041 | 0.0027 | 0.0017 | 0.0011 | 0.0007 | 0.0005 |
| 2   | 0.0618 | 0.0446 | 0.0318 | 0.0223 | 0.0156 | 0.0107 | 0.0074 | 0.0050 | 0.0034 | 0.0023 |
| 3   | 0.1133 | 0.0892 | 0.0688 | 0.0521 | 0.0389 | 0.0286 | 0.0208 | 0.0150 | 0.0107 | 0.0076 |
| 4   | 0.1558 | 0.1339 | 0.1118 | 0.0912 | 0.0729 | 0.0573 | 0.0443 | 0.0337 | 0.0254 | 0.0189 |
| 5   | 0.1714 | 0.1606 | 0.1454 | 0.1277 | 0.1094 | 0.0916 | 0.0752 | 0.0607 | 0.0483 | 0.0378 |
| 6   | 0.1571 | 0.1606 | 0.1575 | 0.1490 | 0.1367 | 0.1221 | 0.1066 | 0.0911 | 0.0764 | 0.0631 |
| 7   | 0.1234 | 0.1377 | 0.1462 | 0.1490 | 0.1465 | 0.1396 | 0.1294 | 0.1171 | 0.1037 | 0.0901 |
| 8   | 0.0849 | 0.1033 | 0.1188 | 0.1304 | 0.1373 | 0.1396 | 0.1375 | 0.1318 | 0.1232 | 0.1126 |
| 9   | 0.0519 | 0.0688 | 0.0858 | 0.1014 | 0.1144 | 0.1241 | 0.1299 | 0.1318 | 0.1300 | 0.1251 |
| 10  | 0.0285 | 0.0413 | 0.0558 | 0.0710 | 0.0858 | 0.0993 | 0.1104 | 0.1186 | 0.1235 | 0.1251 |
| 11  | 0.0143 | 0.0225 | 0.0330 | 0.0452 | 0.0585 | 0.0722 | 0.0853 | 0.0970 | 0.1067 | 0.1137 |
| 12  | 0.0065 | 0.0113 | 0.0179 | 0.0263 | 0.0366 | 0.0481 | 0.0604 | 0.0728 | 0.0844 | 0.0948 |
| 13  | 0.0028 | 0.0052 | 0.0089 | 0.0142 | 0.0211 | 0.0296 | 0.0395 | 0.0504 | 0.0617 | 0.0729 |
| 14  | 0.0011 | 0.0022 | 0.0041 | 0.0071 | 0.0113 | 0.0169 | 0.0240 | 0.0324 | 0.0419 | 0.0521 |
| 15  | 0.0004 | 0.0009 | 0.0018 | 0.0033 | 0.0057 | 0.0090 | 0.0136 | 0.0194 | 0.0265 | 0.0347 |

表 15 Poisson 概率 ( $\mu$  在 5.5 与 20.0 之间)

| $y$ | $\mu$  |        |        |        |        |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|     | 5.5    | 6.0    | 6.5    | 7.0    | 7.5    | 8.0    | 8.5    | 9.0    | 9.5    | 10.0   |
| 16  | 0.0001 | 0.0003 | 0.0007 | 0.0014 | 0.0026 | 0.0045 | 0.0072 | 0.0109 | 0.0157 | 0.0217 |
| 17  | 0.0000 | 0.0001 | 0.0003 | 0.0006 | 0.0012 | 0.0021 | 0.0036 | 0.0058 | 0.0088 | 0.0128 |
| 18  | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0005 | 0.0009 | 0.0017 | 0.0029 | 0.0046 | 0.0071 |
| 19  | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0004 | 0.0008 | 0.0014 | 0.0023 | 0.0037 |
| 20  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0003 | 0.0006 | 0.0011 | 0.0019 |
| 21  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0003 | 0.0005 | 0.0009 |
| 22  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0002 | 0.0004 |
| 23  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 |

| $y$ | $\mu$  |        |        |        |        |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|     | 11.0   | 12.0   | 13.0   | 14.0   | 15.0   | 16.0   | 17.0   | 18.0   | 19.0   | 20.0   |
| 0   | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 1   | 0.0002 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2   | 0.0010 | 0.0004 | 0.0002 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 3   | 0.0037 | 0.0018 | 0.0008 | 0.0004 | 0.0002 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 4   | 0.0102 | 0.0053 | 0.0027 | 0.0013 | 0.0006 | 0.0003 | 0.0001 | 0.0001 | 0.0000 | 0.0000 |
| 5   | 0.0224 | 0.0127 | 0.0070 | 0.0037 | 0.0019 | 0.0010 | 0.0005 | 0.0002 | 0.0001 | 0.0001 |
| 6   | 0.0411 | 0.0255 | 0.0152 | 0.0087 | 0.0048 | 0.0026 | 0.0014 | 0.0007 | 0.0004 | 0.0002 |
| 7   | 0.0646 | 0.0437 | 0.0281 | 0.0174 | 0.0104 | 0.0060 | 0.0034 | 0.0019 | 0.0010 | 0.0005 |
| 8   | 0.0888 | 0.0655 | 0.0457 | 0.0304 | 0.0194 | 0.0120 | 0.0072 | 0.0042 | 0.0024 | 0.0013 |
| 9   | 0.1085 | 0.0874 | 0.0661 | 0.0473 | 0.0324 | 0.0213 | 0.0135 | 0.0083 | 0.0050 | 0.0029 |
| 10  | 0.1194 | 0.1048 | 0.0859 | 0.0663 | 0.0486 | 0.0341 | 0.0230 | 0.0150 | 0.0095 | 0.0058 |

续表

| $y$ | 11.0   | 12.0   | 13.0   | 14.0   | 15.0   | 16.0   | 17.0   | 18.0   | 19.0   | 20.0   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 11  | 0.1194 | 0.1144 | 0.1015 | 0.0844 | 0.0663 | 0.0496 | 0.0355 | 0.0245 | 0.0164 | 0.0106 |
| 12  | 0.1094 | 0.1144 | 0.1099 | 0.0984 | 0.0829 | 0.0661 | 0.0504 | 0.0368 | 0.0259 | 0.0176 |
| 13  | 0.0926 | 0.1056 | 0.1099 | 0.1060 | 0.0956 | 0.0814 | 0.0658 | 0.0509 | 0.0378 | 0.0271 |
| 14  | 0.0728 | 0.0905 | 0.1021 | 0.1060 | 0.1024 | 0.0930 | 0.0800 | 0.0655 | 0.0514 | 0.0387 |
| 15  | 0.0534 | 0.0724 | 0.0885 | 0.0989 | 0.1024 | 0.0992 | 0.0906 | 0.0786 | 0.0650 | 0.0516 |
| 16  | 0.0367 | 0.0543 | 0.0719 | 0.0866 | 0.0960 | 0.0992 | 0.0963 | 0.0884 | 0.0772 | 0.0646 |
| 17  | 0.0237 | 0.0383 | 0.0550 | 0.0713 | 0.0847 | 0.0934 | 0.0963 | 0.0936 | 0.0863 | 0.0760 |
| 18  | 0.0145 | 0.0255 | 0.0397 | 0.0554 | 0.0706 | 0.0830 | 0.0909 | 0.0936 | 0.0911 | 0.0844 |
| 19  | 0.0084 | 0.0161 | 0.0272 | 0.0409 | 0.0557 | 0.0699 | 0.0814 | 0.0887 | 0.0911 | 0.0888 |
| 20  | 0.0046 | 0.0097 | 0.0177 | 0.0286 | 0.0418 | 0.0559 | 0.0692 | 0.0798 | 0.0866 | 0.0888 |
| 21  | 0.0024 | 0.0055 | 0.0109 | 0.0191 | 0.0299 | 0.0426 | 0.0560 | 0.0684 | 0.0783 | 0.0846 |
| 22  | 0.0012 | 0.0030 | 0.0065 | 0.0121 | 0.0204 | 0.0310 | 0.0433 | 0.0560 | 0.0676 | 0.0769 |
| 23  | 0.0006 | 0.0016 | 0.0037 | 0.0074 | 0.0133 | 0.0216 | 0.0320 | 0.0438 | 0.0559 | 0.0669 |
| 24  | 0.0003 | 0.0008 | 0.0020 | 0.0043 | 0.0083 | 0.0144 | 0.0226 | 0.0328 | 0.0442 | 0.0557 |
| 25  | 0.0001 | 0.0004 | 0.0010 | 0.0024 | 0.0050 | 0.0092 | 0.0154 | 0.0237 | 0.0336 | 0.0446 |
| 26  | 0.0000 | 0.0002 | 0.0005 | 0.0013 | 0.0029 | 0.0057 | 0.0101 | 0.0164 | 0.0246 | 0.0343 |
| 27  | 0.0000 | 0.0001 | 0.0002 | 0.0007 | 0.0016 | 0.0034 | 0.0063 | 0.0109 | 0.0173 | 0.0254 |
| 28  | 0.0000 | 0.0000 | 0.0001 | 0.0003 | 0.0009 | 0.0019 | 0.0038 | 0.0070 | 0.0117 | 0.0181 |
| 29  | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0004 | 0.0011 | 0.0023 | 0.0044 | 0.0077 | 0.0125 |
| 30  | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0006 | 0.0013 | 0.0026 | 0.0049 | 0.0083 |
| 31  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0003 | 0.0007 | 0.0015 | 0.0030 | 0.0054 |
| 32  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0004 | 0.0009 | 0.0018 | 0.0034 |
| 33  | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0005 | 0.0010 | 0.0020 |

来源:由 D. K. Hildebrand 计算。

## 参 考 文 献

- Agresti, A. (1990), *Categorical Data Analysis*. New York: Wiley.
- Brown, S., M. Healy, and M. Kearns (1981), "Report on the interlaboratory trial of the reference method for the determination of total calcium in serum." *Journal of Clinical Chemistry and Clinical Biochemistry* 19, 395-426.
- Carmer, S., and M. Swanson (1973), "An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods." *Journal of the American Statistical Association* 68, 66-74.
- Carroll, R., R. Chen, E. George, T. Li, H. Newton, H. Schmiedliche, and N. Wang (1997), "Ozone exposure and population density in Harris County, Texas." *Journal of the American Statistical Association* 92, 392-415.
- Carter, R. (1981), "Restricted maximum likelihood estimation of bias and reliability in the comparison of several measuring methods." *Biometrics* 37, 733-741.
- Cochran, W. (1954), "Some methods for strengthening the common  $\chi^2$  test." *Biometrics* 10, 417-451.
- Cochran, W., and G. Cox (1957), *Experimental Design*. 2nd ed. New York: Wiley.
- Conover, J. (1998), *Practical Nonparametric Statistics*. 3rd ed. New York: Wiley.
- Cressie, N. (1993), *Statistics for Spatial Data*. New York: Wiley.
- Crowder, M., and D. Hand (1990), *Analysis of Repeated Measures*. London: Chapman and Hall.
- Cryer, J., and R. Miller (1991), *Statistics for Business: Data Analysis and Modeling*. Boston, PWS-Kent.
- Deming, W. (1981), *Quality, Productivity, and Competitive Position*. Cambridge, Mass.: MIT-CAES.
- Devore, J. (2000), *Probability and Statistics for Engineering and the Sciences*. 5th ed. Pacific Grove, Cal.: Duxbury Press.
- Diggle, P., K. Liang, and S. Zeger (1996), *Analysis of Longitudinal Data*. New York: Oxford University Press.
- Dunnnett, C. (1955), "A multiple comparison procedure for comparing several treatments with a control." *Journal of the American Statistical Association* 50, 1096-1121.
- Dunnnett, C. (1964), "New tables for multiple comparisons with a control." *Biometrics* 20, 482-491.
- Durbin, J., and G. Watson (1951), "Testing for serial correlation in least squares, II." *Biometrika* 38, 159-178.
- Effron, B., and R. Tibshirani (1993), *An Introduction to the Bootstrap*. London: Chapman and Hall.

- Fisher, R. A. (1949), *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- Greenhouse, S., and S. Geisser (1959), "On methods in the analysis of profile data." *Psychometrika* 24, 95-112.
- Haining, R. (1990), *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.
- Hammer, M. (1996), *Beyond Reengineering*. New York: Harper Collins.
- Hammer, M., and J. Champy (1993), *Reengineering the Corporation: A Manifesto for Business Revolution*. New York: Harper Collins.
- Hicks, C., and K. Turner (1999), *Fundamental Concepts in the Design of Experiments*. 5th ed. New York: Oxford University Press.
- Hildebrand, D., and L. Ott (1998), *Statistical Thinking for Managers*. 4th ed. Pacific Grove, Cal.: Duxbury Press.
- Huynh, H., and L. Feldt (1970), "Conditions under which mean square ratios in repeated measurement designs have fixed *F*-distributions." *Journal of the American Statistical Association* 65, 1582-1589.
- Jones, B., and M. Kenard (1989), *Design and Analysis of Cross-Over Trials*. London: Chapman and Hall.
- Koehler, K. (1986), "Goodness-of-fit tests for log-linear models in sparse contingency tables." *Journal of the American Statistical Association* 81, 483-493.
- Kuehl, R. (1999), *Design of Experiments: Statistical Principles of Research Design and Analysis*. 2nd ed. Pacific Grove, Cal.: Duxbury Press.
- Larntz, K. (1978), "Small-sample comparison of exact levels for chi-squared goodness-of-fit statistics." *Journal of the American Statistical Association* 73, 253-263.
- Lentner, M., and T. Bishop (1993), *Experimental Design and Analysis*. 2nd ed. Blacksburg, Va.: Valley Book Company.
- Mallows, C. (1973), "Some comments on  $C_p$ ." *Technometrics* 15, 661-675.
- Manly, B. (1998), *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 2nd ed. London: Chapman and Hall.
- Meyer, C. (1993), *Fast Cycle Time*. New York: The Free Press.
- Miller, R. (1981), *Simultaneous Statistical Inference*. 2nd ed. New York: Springer-Verlag.
- Montgomery, D. (1997), *Design and Analysis of Experiments*. 5th ed. New York: Wiley.
- Neter, J., M. Kutner, C. Nachtsheim, and W. Wasserman (1996), *Applied Linear Statistical Models*. 4th ed. Boston: WCB McGraw-Hill.
- Newman, R. (1998), "Testing parallelism among the profiles after a certain time period." Unpublished PhD dissertation. Texas A&M University.
- Raftery, A., and J. Zeh (1998), "Estimating bowhead whale population size and

rate of increase from the 1993 census." *Journal of the American Statistical Association* 93, 451-462.

Randles, R., and D. Wolfe (1979), *Introduction to the Theory of Nonparametric Statistics*. New York: Wiley.

Ripley, B. (1988), *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.

Rowe, N., R. Anderson, and L. Wanninger (1974), "Effects of ready-to-eat breakfast cereals on dental caries experience in adolescent children: A three-year study." *Journal of Dental Research* 53, 33.

Scheaffer, R. L., W. Mendenhall, and L. Ott (1996), *Elementary Survey Sampling*, 5th ed. Pacific Grove, Cal.: Duxbury Press.

Scheffé, H. (1953), "A method for judging all contrasts in the analysis of variance." *Biometrika* 40, 87-104.

Searle, S., G. Casella, and C. McCulloch (1992), *Variance Components*. New York: Wiley.

Snedecor, G., and W. Cochran (1980), *Statistical Methods*, 7th ed. Ames, Iowa: Iowa State University Press.

Tukey, J. (1977), *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.

U.S. Bureau of Labor Statistics, *Handbook of Methods*, Vols. I and II (1982). Washington, D.C.: U.S. Department of Labor.

Vonesh, E., and V. Chinchilli (1997), *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker.

Welch, B. (1938), "The significance of the difference between two means when the population variances are unequal," *Biometrika* 29, 350-362.



# 索引

- additive effects, 621
- adjusted treatment means, 948, 950
- American Statistical Association (ASA), 22
- analysis of covariance, 943–974
  - case study, 944–946
  - completely randomized design, multiple covariates, 962–970
    - checking results with plots of residuals, 968–970
    - example of, 963
    - exercises for, 970
    - models for, 964–968
    - overview of, 962–963
    - reporting conclusions of, 970
  - completely randomized design, one covariate, 946–958
    - adjusted treatment means and, 948, 950
    - examples of, 947, 951–952, 957–958
    - exercises for, 958–959
    - model conditions, 948, 955–957
    - models and tests of hypotheses, 949–951
    - overview of, 946–947
    - exercises for, 971–974
  - extrapolation problems, 959–962
  - introduction, 943–944
    - covariates defined, 943
    - examples of covariate applications, 943–944
- analysis of variance (AOV), 383–384
  - alternatives to
    - Kruskal–Wallis test, 410–414
    - transformations of data, 403–410
  - checking conditions for, 396–403
    - examples of, 397–403
    - overview of, 396–397
    - reporting conclusions of, 403
    - using residuals analysis for, 397
  - crossover designs and, 1041
  - model of, 394–396
  - overview of, 379
  - power of, 848
  - sample variation and, 383
  - statistical tests for, 384–393
    - AOV table summarizing, 389
    - completely randomized design, 387–388
    - developing test statistic, 386–387
    - examples of, 389–392
    - exercises for, 392–393
    - mean square and, 388
    - multiple  $t$  tests, 385
    - sum of squares and, 388
    - for two-factor experiment, repeated measures, 1032
  - analysis of variance (AOV), fixed-, random-, and mixed-effects, 975–1024
    - case study, 976–978
    - exercises for, 1020–1024
    - fixed-effects model
      - defined, 975
      - vs. random-effects model, 978–979
    - introduction, 975–976
    - mixed-effects model, 992–1000
      - assumptions for, 992
      - defined, 976
      - example of, 993–996
      - exercises for, 1000
      - tests of significance for, 992–993
  - nested sampling and split-plot design, 1010–1020
    - AOV table for, 1014
    - examples of, 1011–1014, 1018–1019
    - overview of, 1010–1011
    - split-plot design, 1014–1018
  - random-effects model, 978–983
    - AOV moment matching and, 980
    - assumptions of, 978–979
    - defined, 976
    - example of, 981–982
    - exercises for, 982–983
    - expected mean squares and, 979
    - extensions of, 983–992
    - hypothesis tests for, 979–980
    - vs. fixed-effects model, 978–979
  - random-effects model, block design
    - assumptions for, 983–984
    - estimating variance components, 986–987
    - example of, 987–989
    - exercises for, 991–992
    - hypothesis testing for, 989–990
    - model and AOV tables for, 985
    - nested sampling and, 990–991
    - rules for obtaining expected mean squares, 1000–1010
    - classifying interactions, 1000–1001
    - examples of, 1001, 1003–1005, 1006–1010
    - mean square tables, 1001–1003
  - analysis of variance (AOV), for standard designs, 853–942
    - case study, 854–855
    - comparing treatment means, 916–922
      - applying multiple comparison procedures, 920
      - case study, 921–922
      - examples of, 917–921
      - using Fisher's LSD procedure, 916–917
  - completely randomized design, single factor, 855–859
    - advantages/disadvantage of, 858
    - example of, 858–859
    - model for, 856
    - overview of, 855–856
    - using total sum of squares as test statistic, 856–857
  - exercises for, 924–942
  - factorial treatment in completely randomized design, 891–914
    - accounting for interaction in, 898–899
    - AOV table for, 908
    - constructing AOV table for, 899–900
    - example of, 901
    - exercises for, 913–914
    - first test of significance for, 902–903
    - interaction of factors and, 894

- analysis of variance (AOV) (*contd.*)  
 interpreting *F* test for main effects, 904–905  
 models for, 897–898, 905–907  
 one-at-a-time approach, 892  
 overview of factorial experiments, 894–897  
 sum of squares calculations for, 907–908  
 summary of method of analysis, 909  
 factorial treatment in randomized complete block design, 914–916  
 AOV table for, 916  
 example of, 915–916  
 exercise for, 916  
 overview of, 914–915  
 introduction, 853–854  
 key formulas for, 923  
 Latin square design, 879–891  
 advantages/disadvantages of, 880  
 defined, 881  
 examples of, 879–880, 884–888  
 exercises for, 888–891  
 filtering with, 881–882  
 model for, 881  
 overview of, 879–880  
 relative efficiency of, 884  
 testing for treatment effects, 882–884  
 randomized complete block design, 859–878  
 advantages/disadvantages of, 861  
 defined, 861  
 examples of, 868–871  
 exercises for, 872–878  
 filtering and, 863–864  
 model for, 863–864  
 overview of, 859–861  
 relative efficiency of, 867–868  
 testing difference among treatment means, 865–867  
 analysis of variance (AOV), unbalanced designs, 1051–1076  
 balanced incomplete block designs, 1063–1072  
 AOV table for, 1065  
 comparing treatment means, 1067–1068  
 defined, 1064  
 examples of, 1063–1066, 1068  
 model for, 1065  
 overview of, 1063  
 reporting conclusions of, 1071–1072  
 sum of squares for, 1066  
 case study, 1052–1053  
 exercises for, 1074–1076  
 introduction, 1051–1052  
 key formulas for, 1072–1073  
 Latin square design, missing data, 1059–1063  
 comparing treatment means, 1061  
 estimating missing value, 1059  
 example of, 1059–1060  
 exercises for, 1062–1063  
 fitting complete and reduced models for, 1061  
 randomized block design, one or more missing observations, 1053–1058  
 comparing treatment means, 1056  
 estimating value of missing observation, 1053  
 estimation bias and, 1053–1054  
 examples of, 1054–1055  
 exercises for, 1057–1058  
 fitting complete and reduced models for, 1056–1057  
 testing effect of treatments, 1057  
 AOV moment matching, 980  
 AOV tables  
 balanced incomplete block designs, 1065  
 factorial experiment ( $a \times b$ ), 985  
 factorial treatment in completely randomized design, 899–900, 908, 916  
 fixed or random-effects models, 979  
 mixed-effects model, 998  
 nested sampling and split-plot design, 1014  
 random-effects model, block design, 985  
 split-plot design, 1016, 1019  
 statistical tests for AOV, 389  
 two-factor experiment, 984  
 two-period crossover design, 1044  
 arithmetic mean (*See* mean)  
 associations  
 strength of association  
 contingency tables, 505  
 measuring, 510–516  
 unit of association, 532  
 vs. causal relationships, 830  
 backup analyses, 1087  
 backward elimination, 717–720  
 balanced incomplete block (BIB) designs, 1063–1072  
 AOV table for, 1065  
 comparing treatment means, 1067–1068  
 defined, 1064  
 examples of, 1063–1066, 1068  
 model for, 1065  
 overview of, 1063  
 sum of squares for, 1066  
 bar charts  
 accuracy of data presentation with, 1081  
 cluster bar graphs, 102–103  
 description of, 45  
 examples of, 46  
 guideline for constructing, 46  
 stacked bar graphs, 102  
 uses of, 50  
 Bayes' formula, 136–140  
 equation for, 138  
 false positive/false negative test results, 136  
 likelihoods in, 138  
 observable events in, 138  
 prior/posterior probabilities in, 138  
 sensitivity/specificity of diagnostic tests, 137  
 states of nature in, 138  
 benchmarks (*See* control treatments)  
 best-guess estimation (*See* point estimation)  
 best subset regression  
 multiple regression, 730–731  
 selecting multiple regression variables, 716–717  
 biased samples, 1082–1083  
 binomials, 144–154  
 defining binomial experiments, 144  
 formula for computation of, 146–149  
 mean and standard deviation of, 149–151  
 normal approximation to, 182–185  
 properties of, 145  
 public opinion polls as examples of, 144  
 block design (*See* randomized complete block design)

- blocking variables (*See also* randomized complete block design)
  - split-plot design and, 1016
  - use of, 891
- Bonferroni procedure
  - error rate control and, 439–440
  - testing hypothesis with, 457
- bootstrap techniques, for sampling distribution, 349–350
- box-and-whiskers plot, 97
- boxplots, 96–101
  - examples of, 99, 106–107
  - overview of, 96–97
  - skeletal boxplot, 96–97
  - steps in construction of, 98
  - strengths of, 99–100
  - using for residuals analysis, 968
  - using to check normality, 396–397
  - vs. stem-and-leaf plot, 96
- calculators, data analysis with, 41–43
- calibration, inverse regression and, 582
- categorical data, 469–530
  - case study, 470–471
  - contingency tables and, 501–510
    - cross tabulations with, 502
    - dependence and, 502
    - estimated expected value, 503
    - examples of, 503–508
    - exercises for, 508–510
    - likelihood ratio statistic, 505
    - strength of association, 505
    - tests of homogeneity, 505–506
    - tests of independence, 504
  - exercises for, 521–530
  - inferences about a population proportion, 471–482
    - confidence interval for, 472–474
    - examples of, 472–476
    - exercises for, 477–482
    - mean and standard error, 472
    - sample size requirement, 477
    - statistical test for, 475
  - inferences about difference between two population proportions, 482–488
    - confidence interval for, 483
    - examples of, 483–485
    - exercises for, 486
    - notation for comparing, 482
    - rule for sample size, 484
    - statistical test for, 484–485
  - inferences about several proportions, 488–497
    - examples of, 489, 491–493
    - exercises for, 494–497
    - expected number of outcomes, 490
    - multinomial distribution, 488–489
    - multinomial experiments, 488
    - statistical test for, 490–494
  - introduction, 469
  - key formulas for, 520–521
  - measuring strength of relation, 510–516
    - dependent and independent variables and, 512
    - examples of, 513–516
    - percentage analysis, 510–511
    - predictability analysis, 511–512
  - odds and odds ratios, 516–519
    - example of, 517
    - formula for, 516–517
    - odds ratio defined, 518–519
  - Poisson distribution, 497–501
    - assumptions for, 498
    - defined, 497
    - examples of, 497–500
    - formula for, 497
    - tests using, 498–499
- causal relationships, compared with associations, 830
- cell probabilities, 490
- census polls, 19–20
- Central Limit Theorem, 175, 179–180, 349
- central tendency, measures of
  - data skew and, 76–77
  - exercises for, 77–81
  - mean, 73–76, 77
  - median, 71–73, 77
  - mode, 70, 77
  - overview of, 69–70
- chi-square distribution, 344–346
  - approximating, 490
  - critical values of, 345
  - data normality and, 349, 351
  - densities of, 345
  - overview of, 344–345
  - statistical table of, 1100–1101
  - upper-tail and lower-tail values of, 346
- chi-square goodness-of-fit test, 490–497
- class frequency, 48
- class intervals, constructing, 47–48
- classical interpretation of probability
  - defined, 124
  - determining probability with, 126
- clinical trials, 7
- cluster bar graphs, 102–103
- cluster effect, of data, 274
- cluster sampling, 20
- Cobb–Douglas production function, 738
- coefficient of determination
  - for multiple regression, 646, 712
  - for simple linear regression, 592–593
- coefficient of variation
  - as measure of variability, 96
  - transformations of data and, 406
- coefficients, estimating mean, 197–200
- coefficients, linear regression, 591
- coefficients, multiple regression, 627–646
  - examples for, 628–633
  - exercises for, 633–646
  - least-squares estimates, 630
  - normal equations for, 627–628
  - residual standard deviation and, 632–633
  - testing subset of regression coefficients, 657–665
- collinearity
  - assessing, 708–710
  - avoiding, 710
  - effect of, 652
  - overspecification and, 715
  - problems of, 646–647
- communication and documentation, 1077–1089
  - communication hurdles
    - biased samples, 1082–1083
    - graphical distortions, 1079–1082
    - sample size, 1083–1084
  - data preparation, steps in, 1084–1087
  - difficulty of good communication, 1078–1079
  - documenting and storing results, 1088–1089
  - introduction, 1077
  - report guidelines, 1087–1088
  - types of, 1079
- complements of events, 129–130 (*See also* events)

- completely randomized design  
 analysis of variance for, 387–388  
 experimental designs and, 28–29  
 model for observations in,  
 394–396  
 assumptions for, 395  
 conditions for, 394  
 sum of terms and, 394–395
- completely randomized design, multi-  
 ple covariates, 962–970  
 checking results with plots of resid-  
 uals, 968–970  
 example of, 963  
 exercises for, 970  
 models for, 964–968  
 overview of, 962–963  
 reporting conclusions of, 970  
 testing hypotheses, 964
- completely randomized design, one  
 covariate, 947  
 adjusted treatment means and,  
 948, 950  
 examples of, 947, 951–952,  
 957–958  
 exercises for, 958–959  
 model assumptions  
 reviewed, 955–957  
 stated, 948  
 models and sum of squares for,  
 949–951  
 overview of, 946–947
- completely randomized design, single  
 factor, 855–859  
 advantages/disadvantages of, 858  
 example of, 858–859  
 model for, 856  
 overview of, 855–856  
 using total sum of squares as test  
 statistic, 856–857
- compound symmetry of observations,  
 1030–1031
- computers, data analysis with, 41–43
- conditional probability, 131–136  
 defined, 132  
 independent vs. dependent events,  
 133–134  
 multiplication law and, 132  
 vs. unconditional probability, 132
- confidence coefficient  
 estimating population mean,  
 197–200  
 estimating population variance,  
 346  
 population proportion inferences  
 and, 472
- confidence interval  
 defined, 666  
 estimating standard error in multi-  
 ple regression, 653  
 Fisher's least significant difference  
 and, 444  
 linear regression forecasting and,  
 567–568  
 mean and, 196, 200–201  
 mean, unknown variance and, 234  
 median and, 243–245  
 population proportions and,  
 472–474, 483  
 population variance and  
 comparing two, 359  
 single, 346  
 sample size and, 205, 316  
 Scheffé's *S* method and, 453–454  
 slope and, 559–560  
 Tukey's *W* procedure and, 447
- confounded factors, 860
- constant variance  
 regression model assumptions,  
 759, 760–765  
 using weighted least-squares,  
 762–763
- contingency tables, 501–510  
 constructing, 101–102  
 cross tabulations with, 502  
 dependence and, 502  
 examples of, 503–508  
 exercises for, 508–510  
 expected value and, 503  
 likelihood ratio and, 505  
 strength of association in, 505  
 tests of homogeneity, 505–506  
 tests of independence, 504
- continuity correction, normal approxi-  
 mation and, 184
- continuous random variables,  
 154–157  
 defined, 142  
 probability distribution of,  
 155–157  
 vs. discrete random variables, 154
- contrasts  
 linear contrasts, 431–438  
 defined, 432  
 examples of, 432–437  
 exercises for, 437–438  
*F* test for, 436  
 notation for, 431  
 orthogonal contrasts  
 defined, 432  
 overview of, 432–433
- control treatments  
 defined, 833  
 use in experiments, 450
- correlation, 590–600  
 assumptions for, 593–595  
 coefficient of determination and,  
 592–593  
 correlation coefficient and, 591  
 examples of, 592–597  
 exercises for, 597–600  
 measuring predictions with,  
 590–591
- correlation coefficient, 591
- correlation matrix, 708
- count data (*See* categorical data)
- covariance, analysis of (*See* analysis  
 of covariance)
- covariates (*See also* analysis of covar-  
 iance)  
 defined, 947  
 using to reduce variability,  
 839–840
- $C_p$  statistic, 716
- cross-product term, in multiple regres-  
 sion, 619
- cross tabulations, with contingency ta-  
 bles, 502
- crossover designs, 1040–1044  
 analysis of variance for, 1041  
 defined, 1025  
 example of, 1041–1043  
 layout for two-period crossover de-  
 sign, 1043  
 model for, 1040  
 using Latin square design, 1040
- data  
 analysis, 2  
 cluster effect of, 274  
 collection  
 design for, 18  
 processes included in, 19  
 role in statistics, 2  
 statisticians and, 10  
 collection techniques, 24–25  
 direct observation, 25  
 personal interviews, 24  
 self-administered question-  
 naire, 25  
 telephone interviews, 24–25  
 evaluating normality of, 351  
 management  
 procedures for, 35–38  
 steps in process of, 35–37  
 preparation, role in statistics, 2

- preparation, steps in
  - process of learning from, 2-5
  - spatial correlation of, 274
- data description, 40-120
  - boxplot, 96-101
    - construction of, 98
    - overview of, 96-97
    - skeletal boxplot, 96-97
    - strengths of, 99-100
  - exercises for, 110-120
  - graphical methods, 43-69
    - bar charts, 45-46
    - data organization guidelines, 43
    - exercises for, 63-69
    - frequency histogram, 46-54
    - general guidelines, 62
    - pie charts, 43-45
    - stem-and-leaf plots, 54-57
    - time series, 57-62
  - introduction, 40-41
  - key formulas, 110
  - measures of central tendency, 69-81, 81-96
    - data skew and, 76-77
    - exercises for, 77-81
    - mean, 73-76, 77
    - median, 71-73, 77
    - mode, 70, 77
    - overview of, 69-70
  - measures of variability
    - coefficient of variation, 96
    - deviation, 86-87
    - Empirical Rule for, 89-93
    - exercises for, 93-96
    - interquartile range, 86
    - overview of, 81-82
    - percentiles, 82-86
    - range, 82
    - standard deviation, 88-89
    - variance, 87-88
  - multiple variables, 101-109
    - cluster bar graphs and, 102-103
    - contingency tables and, 101-102
    - exercises for, 107-109
    - scatterplots and, 102-103
    - side-by-side boxplots and, 104-107
    - stacked bar graphs and, 102
    - using calculators and computers for, 41-43
- data dredging, 427-428
- data snooping, 427-428
- data-splitting approach, multiple regression, 714
- data trail
  - defined, 1084
  - overview of, 36
- data variation (*See* variability, measures of)
- databases
  - creating, 1085
  - creating data files from, 1086-1087
  - editing, 1085-1086
  - finalizing, 1086
- dependent events, 133-134
- dependent variables
  - contingency tables and, 502
  - measuring strength of relations in, 512
  - vs. independent variables, 512
- descriptive statistics, 40
- designed experiments, 829-852 (*See also* experimental design)
  - comparing observation studies with, 830-831
  - controlling experimental error, 836-840
    - experimental procedures and, 836-837
    - selecting experimental and measurement units, 837-838
    - using blocking, 838-839
    - using covariates, 839-840
  - determining number of replications, 845-848
    - accuracy of estimator specifications and, 845-846
    - examples of, 846-847
    - using *F* test for, 846-848
  - elements of plan for, 831
  - exercises for, 848-852
  - introduction, 829
  - randomization of experimental units to treatments, 840-844
    - examples of, 841-842, 844
    - overview of, 840-841
    - randomized complete block design and, 843-845
    - steps in random assignment process, 842-843
- terminology, 831-836
  - control treatment, 833
  - experimental error, 834-835
  - experimental unit, 833
  - factorial treatment design, 832
  - factors, 832
  - fractional factorial experiments, 833
  - measurement unit, 834
  - measurements or observations, 832
  - one-way classification, 832
  - replication, 833-834
  - treatment design, 832
  - treatments, 832
  - variance of experimental error, 835-836
  - unbalanced designs, 1051
- deviation, 86-87 (*See also* residual standard deviation; standard deviation)
- df (degree of freedom), 504
- diagnostic measures, leverage and influence, 546
- direct observation, data collection, 25
- discrete random variables (*See also* binomials)
  - probability distribution for, 142
  - properties of, 143
- documentation, 1088-1089
- dummy variables
  - defined, 623
  - lack of fit and, 730
- Dunnett's procedure, 450-452
  - example of, 451-452
  - statistical table of, 1117-1120
  - steps in, 450-451
  - use of controls, 450
- Empirical Rule, 89-93
  - approximating standard deviation, 91-92, 632
  - statement of, 89
  - utility of, 90-91
- equal variance condition, evaluating, 355
- error rate control, 438-440
  - Bonferroni inequality and, 439-440
  - examples of, 440
  - Type I error rate, 438-440
- error terms, 395
- error variance (*See also* residual standard deviation)
  - estimating based on residuals, 546
  - estimating for linear regression, 540
  - estimating for multiple regression, 712
  - mean square estimates of, 580
- estimated expected value, 503

- estimation
  - comparing with hypothesis testing, 193
  - estimation bias
    - in Latin square design, 1059
    - in randomized block design, 1053–1054
  - inferences and, 193
  - of mean, 196–204
  - ratio estimation, 20
- event relations, probability, 128–131
  - complements of events, 129–130
  - intersection of events, 130
  - mutually exclusive events, 128–129
  - union of events, 130
- events, defined, 124
- expected cell counts, 490
- expected mean squares (EMS), 1000–1010 (*See also* mean squares)
  - classifying interactions, 1000–1001
  - comparing fixed- and random-effects model, 979
  - in completely randomized design with single factor, 857
  - constructing mean square tables, 1001–1003
  - examples of, 1001, 1003–1005, 1006–1010
  - in randomized complete block design, 866
  - rules for obtaining, 1003
- expected number of outcomes, in multinomial experiments, 490
- expected values, defined, 619
- experimental design, 28–30 (*See also* designed experiments)
  - completely randomized design, 28–29
  - Latin square design, 29
  - randomized block design, 29
- experimental error
  - causes of, 834–835
  - controlling, 836–840
    - experimental procedures and, 836–837
    - selecting experimental and measurement units, 837–838
    - using blocking, 838–839
    - using covariates, 839–840
  - example of, 834–835
  - lack of fit and, 579–580
  - sources of, 836
  - variance of, 835–836
- experimental units
  - defined, 833
  - randomization of, 840–841
  - selecting, 837–838
- experiments (*See* designed experiments)
- experimentwise error rate
  - Bonferroni inequality and, 439
  - controlling, 438–440
  - Dunnett's procedure for controlling, 450
  - in Scheffé's method, 450
  - in Tukey's *W* procedure, 445
- explained sum of squares (*See* sum of squares regression-SS(Regression))
- explanation, compared with prediction, 531–532
- exploratory data analysis (EDA)
  - graphical methods of, 54
  - quartiles of distribution in, 96
- extrapolation
  - analysis of covariance and, 959–962
  - extrapolation penalty, 569
  - linear regression forecasting and, 568–569
  - multiple regression and, 667
- F* distribution, 356–357
  - critical values of, 357
  - defined, 356
  - densities of, 356
  - properties of, 356
  - statistical table of, 1102–1113
- F* tests
  - analysis of covariance and, 950
  - analysis of variance and, 386–387
  - contrasts and, 436
  - correlated variables and, 708
  - determining number of replications, 846–848
  - expected mean square rules and, 1000
  - factorial treatment, completely randomized design, 904–905
  - null hypothesis and, 560
  - one-factor experiments, repeated measures, 1031
  - power of, 846–848
  - statistical table of power curves for AOV, 1123–1126
  - testing regression coefficients, 648–651, 657–658
  - two-factor experiments, repeated measures, 1034
- factorial experiments, 30–33
  - applications of, 32
  - defined, 32, 832, 895
  - extending random-effects model to, 983
  - interaction of factors in, 31
  - one-at-a-time approach and, 30
    - 31, 892–894, 896
  - overview of, 894–897
  - random-effects model for, 985–986
  - use of factors in, 30
  - vs. split-plot design, 999
- factorial treatment, completely randomized design, 891–914
  - accounting for interaction in, 898–899
  - AOV table for, 908
  - constructing AOV table for, 899–900
  - example of, 901
  - exercises for, 913–914
  - first test of significance for, 902–903
  - interaction of factors and, 894
  - interpreting *F* test for main effects, 904–905
  - models for, 897–898
  - models for three factors, 905–907
  - one-at-a-time approach, 892
  - sum of squares calculations for, 907–908
  - summary of method of analysis, 909
- factorial treatment, randomized complete block design, 914–916
  - AOV table for, 916
  - example of, 915–916
  - exercise for, 916
  - overview of, 914–915
- factors (*See also* variables)
  - confounded factors, 860
  - in designed experiments, 30, 832
  - measuring main effects of, 900
  - nested sampling and, 1010
  - in observational studies, 830
  - three-factor experiment measuring, 909
  - random and fixed effects, 1008–1009
  - treatment design and, 892
- false negative/false positive, in Bayes' formula, 136
- filtering
  - Latin square design, 881–882

- randomized complete block design, 863–864
- first-order models, multiple regression, 619
- Fisher's least significant difference (LSD) (*See* least significant difference (LSD))
- fixed-effects model
  - defined, 975
  - formula for, 978
  - vs. random-effects model, 978–979, 1005
- forecasting
  - linear regression, 567–576
    - confidence interval for, 567–568
    - dangers of extrapolation, 568–569
    - examples of, 569–572
    - exercises for, 572–576
    - overview of, 567
    - prediction interval in, 570–571
    - prediction vs. explanation, 531–532
  - multiple regression, 666–670
    - examples of, 666–668
    - exercises for, 668–670
    - extrapolation and, 667
    - overview of, 666
- formulas
  - analysis of variance (AOV), 923
  - analysis of variance (AOV), unbalanced designs, 1072–1073
  - categorical data, 520–521
  - data description, 110
  - linear regression, 602–603
  - multiple-comparison procedures, 459
  - multiple regression, 620, 688, 783
  - population central values, 251–253
  - population variances, 373–374, 415
  - probability distributions, 187
- forward selection (*See* stepwise regression)
- fractional factorial experiments, 833
- frequency histogram (*See also* histograms)
  - accuracy of data presentation with, 1081
  - examples of, 49
- frequency tables, 47
- general linear model
  - matrix notation for, 684
- multiple regression and, 625–627
  - exercises using, 626–627
  - formula for, 626
- graphical methods
  - bar charts, 45–46
  - communication distortions with, 1079–1082
  - exercises for, 63–69
  - frequency histogram, 46–54
  - guidelines for, 43, 62
  - pie charts, 43–45
  - representing lack of fit with, 576–577
  - stem-and-leaf plots, 54–57
  - time series, 57–62
- grouped data
  - median for, 72
  - range for, 82
- Hartley's  $F_{\max}$  test
  - examples, 366–371
  - overview of, 366
  - recommended use of, 396
  - statistical table of, 1121
  - when to use, 368
- heavy-tailed distributions, 236
- high influence point
  - effect on regression equations, 767
  - estimating regression slope and, 545
- high leverage point, 545–546
- histograms (*See also* frequency histogram)
  - accuracy of data presentation with, 1081
  - creating frequency table for, 47–48
  - Empirical Rule and, 89
  - examples of, 49, 51–54
  - measuring probability in, 50
  - skewed right or left, 54
  - symmetric, 54
  - uniform, 54
  - unimodal and bimodal, 52–54
  - uses of, 50
  - variability in, 81
  - vs. bar and pie charts, 50
- homogeneity, tests of, 505–506
- hypothesis generation, data snooping and, 428
- hypothesis testing
  - comparing with estimation, 193
  - decision rule for, 224–225
  - inferences and, 193
  - null hypotheses and, 227
  - research hypotheses and, 207
- incomplete block design, 1064 (*See also* balanced incomplete block (BIB) design)
- independence
  - conditional probability and, 131–136
  - multiple regression and, 773–781
    - assumptions for, 759
    - diagnosing problems with residual plots, 776
    - Durbin-Watson test statistic, 773
    - first differences alternative, 776
    - serial correlation and, 773–776
    - time series data and, 773
  - tests of, 503, 504
- independent events, 133–134
- independent samples, 134
- independent variables
  - measuring strength of relations in, 512
  - selecting, 708
  - too many, 711
  - vs. dependent variables, 512
- individual error rate
  - Bonferroni inequality and, 439
  - Type I errors, 438
- inferential statistics (*See* statistical inference)
- influence (outliers), 767
- interaction of factors, 622
  - accounting for, 898–899
  - classifying as fixed or random effect, 1000–1001
  - defined, 898
  - disorderly interactions, 905
  - in factorial experiments, 31
  - measuring, 900
  - more than two factors, 899
  - one-at-a-time approach to, 894
  - orderly interactions, 904
- intercept
  - comparing slopes and, 671
  - defined, 532
  - least-squares estimates of, 542
  - in linear regression, 540
  - in multiple regression, 630, 632
  - random error and, 557
- interquartile range (IQR), 86
- intersecting lines, compared with parallel lines, 671
- intersection of events, probability laws and, 130
- interval estimate, 196

inverse regression problems, 582–590  
     calibration as, 582  
     examples of, 583–584  
     exercises for, 588–590  
     predictions, 583, 585  
 inverse transformations, 739

jackknife method, for outlier detection, 769–770

key formulas (See formulas)  
 Kruskal-Wallis test, 410–414  
     example of, 410–412  
     exercises for, 413–414  
     testing more than two populations, 410

lack of fit  
     checking with residual plots, 759  
     dummy variables and, 730  
     in linear regression, 576–582  
         examples of, 577–579, 581–582  
         exercises for, 582  
         mean square estimates and, 580  
         partitioning residuals, 579–580  
         process for, 580  
         using graphs for, 576–577  
     in multiple regression, 727  
     nonlinearity and, 737

large samples, median, 245–246, 247–248

Latin square design, 879–891  
     advantages/disadvantages of, 880  
     crossover designs and, 1040  
     defined, 881  
     examples of, 879–880, 884–888  
     exercises for, 888–891  
     filtering with, 881–882  
     model for, 881  
     overview of, 879–880  
     relative efficiency of, 884  
     testing for treatment effects, 882–884

    types of experimental designs, 29

Latin square design, missing data, 1059–1063  
     comparing treatment means, 1061  
     estimating missing value, 1059  
     example of, 1059–1060  
     exercises for, 1062–1063  
     fitting complete and reduced models for, 1061

least significant difference (LSD), 440–444

AOV for standard designs and, 920  
 balanced incomplete block designs, 1068  
 comparing treatment means  
     standard designs, 916–917  
     unbalanced designs, 1056  
 comparing with Tukey's *W* procedure, 446  
 confidence interval for, 444  
 examples of, 441–444  
 Fisher's protected LSD and, 440–441  
 sample size and, 444  
 steps in, 441, 442–443  
 using with Latin square design, 1061

least-squares method  
     estimating multiple regression coefficients, 627, 630  
     linear regression, 541–544  
     multiple regression, 728  
 level of confidence, estimating mean and, 196, 204–205  
 level of significance (*p*-value), 224–228  
     defined, 224  
     hypothesis testing using, 224–225  
     null hypotheses and, 227

Levine's test  
     confidence interval and, 454  
     examples of, 368–371  
     exercises for, 371–372  
     overview of, 368  
     use of, 368, 396

leverage (outliers), 767

likelihoods  
     defined, 138  
     ratio statistic for, 505

linear contrasts, 431–438  
     defined, 432  
     examples of, 432–437  
     exercises for, 437–438  
     *F* test for contrasts, 436  
     mutually orthogonal contrasts, 432–433

    notation for, 431  
     orthogonal contrasts, 432–433

linear regression, 531–616  
     analyzing, 532  
     case study, 538–540  
     comparing slopes, 671  
     correlation and, 590–600  
         assumptions for correlation inference, 593–595

    coefficient of determination and, 592–593  
     correlation coefficient and, 591  
     examples of, 592–597  
     exercises for, 597–600  
     measuring accuracy of predictions with, 590–591

estimating parameters, 540–557  
     creating scatterplot for, 540–541  
     examples of, 542–544, 547–548  
     exercises for, 548–557  
     high leverage points and, 545–546  
     measures of leverage and influence, 546  
     using least-squares method, 541–542  
     using residuals analysis, 546–547

exercises for, 603–616

forecasting, 567–576  
     confidence interval for, 567–568

dangers of extrapolation, 568–569

examples of, 569–572  
 exercises for, 572–576  
 overview of, 567

prediction interval in, 570–571  
 inferences about parameters, 557–566

    accounting for random error, 557

    confidence interval for slope, 559–560

    examples of, 558–561  
     exercises for, 561–566

    using *F* test for null hypothesis, 560

    using *t* test for slope, 557–558

introduction  
     analyzing simple regression, 532

    assumptions, 532–535

    checking assumptions, 534–537

    choosing transformations, 537–538

    comparing prediction and explanation, 531–532

    use of random error term, 533

inverse regression problems, 582–590

    calibration as, 582  
     examples of, 583–584



- exercises for, 588–590
- predictions, 583, 585
- key formulas for, 602–603
- lack of fit, 576–582
  - examples of, 577–579, 581–582
  - exercises for, 582
  - mean square estimates and, 580
  - partitioning residuals and, 579–580
  - process for, 580
  - using graphs for, 576–577
- linear regression variables vs. multiple regression variables, 631
- linearity, assumption of, 532–533
- logarithmic transformation, 738–739, 759–760
- logic checks, raw data source, 37, 1086
- logistic regression, 675–683
  - example of, 677–679
  - exercises for, 679–683
  - functions of, 676
  - modeling associations, 675
  - multiple logistic regression model, 677
  - simple logistic regression model, 675–677
- LOWESS curve
  - as a smoother, 534–536
  - using with regression models, 739
  - using with residual plots, 759
- machine-readable database
  - defined, 36
  - key-entered data and, 1085
- Mann-Whitney *U* test, 289 (*See also* Wilcoxon rank sum test)
- mean ( $\mu$ )
  - binomial probability distribution and, 149–151
  - characteristics of, 77
  - comparing with variability, 81
  - defined, 73
  - estimating, 196–204
    - confidence coefficient, 197–200
    - confidence interval, 200–201
    - interval estimate and level of confidence, 196
    - sampling distribution, 196
    - standard deviation, 201–202
  - measures of central tendency and, 73–76
  - population central values and, 193
  - population proportion inferences and, 472
  - sample size for estimating, 204–207
    - formula for computation of, 205
    - level of confidence, 204–205
    - tolerable error, 204–205
  - sample size for testing, 219–224
    - for two-sided test, 221
    - Type I/Type II errors and, 219–220
  - statistical test for
    - computing probability of Type II error, 214, 214–216, 218
    - guidelines for, 208
    - null hypothesis, 208, 209
    - one-tailed tests, 211, 216
    - rejection region, 208–210
    - research hypothesis, 207
    - summary of, 212–213
    - test statistic and, 208
    - two-tailed tests, 211, 216
    - Type I/Type II errors, 209, 218–219
  - statistical tests for, 207–219
  - transformations of data and, 407
- mean ( $\mu$ ), multiple comparisons (*See* multiple-comparison procedures)
- mean ( $\mu$ ), unknown variance, 228–243
  - accuracy considerations for *t* procedures, 237–238
  - adjusting for nonnormal distributions, 236–237
  - case study, 238–239
  - confidence interval for, 234
  - robust methods for, 238
  - Student's *t* distribution and
    - properties of, 230
    - vs. normal distribution, 229
  - summary of statistical test for, 231
- mean ( $\mu_1$ ) – mean ( $\mu_2$ )
  - independent samples, 267–287
    - cluster effect of data and, 274
    - confidence interval for, 267–268
    - confidence interval for (when variance is unequal), 276
    - effect of unequal population variances, 274–275
    - exercises for, 280–287
    - sample size for, 314–316
  - spatial correlation and, 274
  - statistical test for, 271–274
  - t* test for, 275–276
  - Type I errors and, 279
  - weighted average of sampling variances, 268
- paired data
  - confidence interval for, 303
  - estimating, 299–308
  - overview of, 299–300
- mean squares (*See also* expected mean squares (EMS))
  - constructing mean square tables, 1001–1003
  - defined, 388
  - estimating error variance with, 580
- measurements
  - measurement units
    - defined, 834
    - example of, 834
    - selecting, 837–838
  - remedies for measurement problems, 23
  - as response variables, 832
- median (*M*)
  - characteristics of, 77
  - defined, 71
  - for grouped data, 72
  - inferences about, 243–250
    - approximating for large samples, 245–248
    - confidence interval for, 243–245
    - sign test vs. *t* test, 248
    - summarizing statistical test for, 246–247
    - using sign test, 246
  - population central values and, 193
- Minitab
  - probability and, 185–186
    - calculating binomial probabilities, 185
    - calculating normal probabilities, 185
    - generating random numbers, 185
    - generating sampling distribution, 186
  - scatterplot matrix, 708
  - statistics software and, 42
- mixed-effects model, 992–1000
  - assumptions for, 992
  - defined, 976
  - example of, 993–996

- mixed-effects model (*contd.*)  
 exercises for, 1000  
 reporting conclusions of, 999  
 residual analysis for, 997  
 tests of significance for, 992–993
- mode  
 characteristics of, 77  
 defined, 70
- models  
 balanced incomplete block designs, 1065  
 completely randomized design, 394–396  
   assumptions for, 395  
   conditions for, 394  
   multiple covariates, 964–968  
   one covariate, 949–951  
   single factor, 856  
   sum of terms and, 394–395  
 crossover designs, 1040  
 factorial treatment, completely randomized design, 897–898, 905–907  
 Latin square design, 881  
 linear regression, 540–557  
   creating scatterplot for, 540–541  
   examples of, 542–544, 547–548  
   exercises for, 548–557  
   high leverage points and, 545–546  
   measures of leverage and influence, 546  
   using least-squares method, 541–542  
   using residuals analysis, 546–547  
 multiple regression, 727–782  
   checking assumptions, 731–737, 758–759  
   checking constant variance, 760–765  
   checking independence, 773–781  
   checking lack of fit, 727–730  
   checking normality, 765–773  
   checking zero expectation, 759–760  
   exercises for, 745–758  
   logarithmic transformation and, 738–739  
   nonlinear least-squares, 740–744  
   using best subset regression, 730–731  
   using nonlinear models, 737–744  
   using scatterplots, 727–730  
 randomized complete block design, 862–864  
 two-factor experiments, repeated measures, 1031  
 multicollinearity (See collinearity)  
 multinomial distribution, 488–489  
 multinomial experiments, 488  
 multiple-comparison procedures, 427–468  
   case study, 428–431  
   comparing treatment means, 920  
   Dunnett's procedure, 450–452  
   example of, 451–452  
   steps in, 450–451  
   use of controls, 450  
 error rate control and, 438–440  
   Bonferroni inequality and, 439–440  
   examples of, 440  
   experimentwise Type I error rate, 438–440  
   individual comparison Type I error rate, 438  
 exercises for, 459–468  
 Fisher's least significant difference and, 440–444  
   confidence interval for, 444  
   for equal sample sizes, 444  
   examples of, 441–444  
   Fisher's protected LSD and, 440–441  
   steps in, 441  
 introduction, 427–428  
 key formulas for, 459  
 linear contrasts and, 431–438  
   defined, 432  
   examples of, 432–437  
   exercises for, 437–438  
   F test for, 436  
   notation for, 431  
   orthogonal contrasts and, 432–433  
   reporting conclusions of, 458  
 Scheffé's *S* method, 452–458  
   confidence interval for, 453–454  
   example of, 453  
   overview of, 452  
   steps in, 452  
 Student–Newman–Keuls procedure, 447–450  
   comparing with Tukey's procedure, 447–448  
   example of, 448–449  
   steps in, 448  
 Tukey's *W* procedure, 444–447  
   comparing with LSD procedure, 446  
   confidence interval for, 447  
   example of, 445–446  
   experimentwise error rate in, 445  
   steps in, 445  
   use of Studentized range distribution in, 444
- multiple logistic regression models, 677
- multiple regression, 617–709  
 case study, 617–620  
 comparing slopes, 670–675  
 examples of, 670–674  
 exercises for, 674–675  
 extending to three or more lines, 674  
 intersecting lines vs. parallel lines, 671  
 estimating coefficients, 627–646  
 examples for, 628–633  
 exercises for, 633–646  
 least-squares estimates, 630  
 linear regression variables vs. multiple regression variables, 631  
 normal equations for, 627–628  
 residual standard deviation and, 632–633  
 exercises for, 689–709  
 forecasting, 666–670  
 examples of, 666–668  
 exercises for, 668–670  
 extrapolation and, 667  
 overview of, 666  
 general linear model for  
   exercises using, 626–627  
   formula for, 626  
 inferences in, 646–657  
   coefficient of determination and, 646  
   collinearity and, 646–647  
   examples of, 647–651, 653–655  
   exercises for, 655–657  
   F test of all coefficients, 648–651, 655  
   *t* test of individual coefficients, 652–655  
 introduction, 620–625  
   assumptions for, 621  
   examples of, 621–622, 624–625

- first-order models, 620–622
- formula for multiple regression model, 620
- higher-order models, 622–625
- parameters of, 620–621
- key formulas for, 688
- logistic regression and, 675–683
  - example of, 677–679
  - exercises for, 679–683
  - modeling associations, 675
  - multiple logistic regression model, 677
  - simple logistic regression model, 675–677
- testing coefficients, 657–665
  - complete and reduced models for, 658
  - example of, 658–661
  - exercises for, 661–665
  - F* test of predictors, 657–658
- theory, 683–686
  - computing estimated standard error, 686
  - computing inverse of the  $X'X$  matrix, 685
  - computing  $SS(\text{Regression})$  and  $SS(\text{Total})$ , 686
  - example of, 684–686
  - general linear model, 684
  - normal equations in matrix notation, 684
- using general linear model for, 625–627
- multiple regression, applying, 705–828
  - case study, 706–707
  - checking model assumptions, 758–782
    - checking constant variance, 760–765
    - checking independence, 773–781
    - checking normality, 765–773
    - checking zero expectation, 759–760
    - overview of, 758–759
  - exercises for, 783–828
  - introduction, 705
  - key formulas for, 783
  - model formation, 727–758
    - exercises for, 745–758
    - logarithmic transformation and, 738–739
    - nonlinear least squares, 740–744
    - testing lack of fit, 727–730
    - testing with best subset regression, 730–731
    - testing with residual plots, 731–737
    - using nonlinear models, 737–744
    - using scatterplots, 727–730
  - selecting variables, 707–727
    - backward elimination, 717–720
    - best subset regression, 716–717
    - collinearity and, 708–710
    - data-splitting approach, 714
    - examples for, 709–722
    - exercises for, 722–727
    - overview of, 707–708
    - performing all possible regressions, 711–714
    - PRESS statistic and, 714–715
    - stepwise regression, 718, 720–722
    - underspecification/over-specification, 715
- multiple regression coefficients, 627–646
  - examples for, 628–633
  - exercises for, 633–646
  - least-squares estimates, 630
  - linear regression variables vs. multiple regression variables, 631
  - normal equations for, 627–628
  - residual standard deviation, 632–633
  - testing, 657–665
- multiple variables (See variables, multiple)
- multiplication law, conditional probability and, 132
- mutually exclusive events, probability laws and, 128–129
- mutually orthogonal contrasts, 432–433
- natural logarithms, 407, 739
- nested sampling, 1010–1020
  - AOV table for, 1014
  - examples of, 1011–1014, 1018–1019
  - F* tests for, 1012–1014
  - factors of, 1010
  - overview of, 1010–1011
  - partially nested design, 1014
  - random-effects model, 990–991
  - split-plot design, 1014–1018
- nonlinearity
  - least squares and, 740–744
  - multiple regression and, 737–744
  - using transformations with, 759–760
- nonnormal distributions, adjusting, 236–237
- normal approximation, 292–295
- normal curve, 157–166, 765
  - 100<sup>th</sup> percentile of distribution, 162–166
  - approximating binomial, 182–185
  - area under, 158–162
  - assumptions
    - checking, 765–773
    - stating, 759
  - calculating normal probabilities, 185
  - continuity correction and, 184
  - overview of, 157–158
  - standard normal curve and, 1091–1092
  - using probability plots with, 766
  - using scatterplots and residual plots with, 765
- normal equations
  - estimating regression coefficients, 627–628
  - for multiple regression, 684
- normal population (See population mean)
- normal probability distribution (See normal curve)
- null hypothesis
  - accepting/rejecting, 227
  - estimating standard error, 653–654
  - multiple *t* tests and, 385
  - randomized complete block design and, 865
  - statistical tests for, 208, 209
  - testing regression coefficients, 658
- numerical descriptive measures
  - comparing parameters and statistics, 69–70
  - overview of, 69
  - value of, 109
- observational studies
  - vs. designed experiments, 830–831
  - vs. scientific studies, 34–35
- observations
  - compound symmetry of, 1030–1031
  - costs of, 714

- observations (*contd.*)
  - defined, 138
  - direct observation, 25
  - estimating value of missing observations, 1053
  - reasons for multiple observations, 1027-1029
  - as response variables, 832
- observed cell counts, 490
- OC curve, 214
- odds, 516-519 (*See also* probability)
  - example of, 517
  - formula for, 516-517
  - odds ratio defined, 518-519
- one-at-a-time approach
  - factorial experiments and, 30-31, 892, 896
  - problems of interaction and, 894
- one-tailed tests
  - computing probability for, 216
  - overview of, 211
- one-way classification (*See also* completely randomized design)
  - defined, 394
  - population mean and, 431
  - treatment design and, 832
- opinion polls
  - applying statistics to, 9-10
  - Gallup and Harris polls, 20
- original files
  - machine-readable database and, 37-38
  - sets of, 1086
- orthogonal contrasts, 432-433
- outcome, defined, 124
- outliers
  - detecting
    - jackknife method, 769-770
    - scatterplots or residual plots, 765
  - how to deal with, 771
  - leverage and influence types, 767
  - problems in identifying, 769
- overspecification, multiple regression variables, 715
- p*-value (*See* level of significance (*p*-value))
- paired data (*See* mean ( $\mu_1$ ) - mean ( $\mu_2$ ), paired data)
- parallel lines vs. intersecting lines, 671
- parameters
  - estimating with scatterplots, 540-541
- linear regression
  - estimating model parameters, 540-557
  - inferences about parameters, 557-566
  - multiple regression, 620-621
  - numerical descriptive measures and, 69-70
  - population parameters, 193, 341
  - regression parameters
    - accounting for random error, 557
    - confidence interval for slope, 559-560
    - examples of, 558-561
    - exercises for, 561-566
    - inferences about, 557-566
    - using *F* test for null hypothesis, 560
    - using *t* test for slope, 557-558
  - partial slopes, 621
  - partitions of TSS
    - completely randomized design, single factor, 857
    - factorial treatment, completely randomized design, 899
    - Latin square design, 882
    - randomized complete block design, 865
- Pearson correlation matrix (*See* correlation matrix)
- percentiles
  - defined, 82
  - examples of, 82-86
  - percentage analysis, 510-511
  - percentile of distribution (100<sup>th</sup>), 162-166
  - quartiles of distribution, 83
- personal interviews, data collection, 24
- pi* (*See* proportion)
- pie charts
  - constructing, 45
  - description of, 43
  - examples of, 44-45
  - uses of, 50
- placebo effect
  - controls and, 450, 833
  - use of, 7
- plots (*See* by individual type)
- point estimation, 194, 648
- Poisson distribution, 497-501
  - assumptions for, 498
  - defined, 497
  - examples of, 497-500
  - formula for, 497
  - tests using, 498-499
- Poisson probabilities, statistical table, 1127-1129
- Poisson random variable, 403-404
- pooled estimate, population variance, 384
- population central values, more than two populations, 379-425
  - case study, 379-384
  - checking AOV conditions, 396-403
  - examples of, 397-403
  - overview of, 396-397
  - reporting conclusions of, 403
  - using residuals analysis for, 397
- exercises for, 416-425
- introduction, 379
- key formulas for, 415
- Kruskal-Wallis test for, 410-414
  - example of, 410-412
  - exercises for, 413-414
  - testing more than two populations, 410
- observation model, completely randomized design, 394-396
  - assumptions for, 395
  - conditions for, 394
  - sum of terms and, 394-395
- testing equality, 384-393
  - AOV table of results, 389
  - completely randomized design, 387-388
  - developing test statistic for, 386-387
  - examples of, 389-392
  - exercises for, 392-393
  - mean square and, 388
  - sum of squares and, 388
  - using multiple *t* tests, 385
- transformations, 403-409
  - defined, 403
  - examples of, 404-409
  - exercises for, 409
  - selecting new variables, 403-404
- population central values, single population
  - case study, 193-195
  - estimating mean, 196-204
    - confidence coefficient, 197-200
    - confidence interval, 200-201
    - interval estimate and level of confidence, 196

sampling distribution, 196  
 standard deviation, 201–202  
 estimating mean, unknown variance, 228–243  
   accuracy of  $t$  procedures, 237–238  
   adjusting for nonnormal distributions, 236–237  
   applying to case study, 238–239  
   confidence interval for, 234  
   robust methods for, 238  
   Student's  $t$  distribution vs. normal distribution, 229  
   Student's  $t$  properties, 230  
   summary of statistical test for, 231  
 estimating median, 243–250  
   approximating for large samples, 245–248  
   confidence interval, 243–245  
   sign test vs.  $t$  test, 248  
   statistical test for, 246–247  
   using sign test, 246  
 exercises for, 253–262  
 introduction, 192–193  
   estimation vs. hypothesis testing, 193  
   parameters, 193  
 key formulas, 251–253  
 level of significance ( $p$ -value), 224–228  
   accepting/rejecting null hypotheses, 227  
   decision rule for hypothesis testing, 224–225  
 sample size, 204–207, 219–224  
   formula for computation of, 205  
   level of confidence and, 204–205  
   tolerable error and, 204–205  
   for two-sided test, 221  
   Type I/Type II errors and, 219–220  
 statistical test for mean, 207–219  
   computing probability of Type II error, 214–216, 218  
   guidelines for, 208  
   null hypothesis, 208, 209  
   one-tailed tests, 211, 216  
   rejection region and, 208–210  
   research hypothesis and, 207  
   summary of, 212–213  
   test statistic and, 208  
   two-tailed tests, 211, 216

Type I/Type II errors, 209, 218–219  
 population central values, two populations, 263–339  
   estimating mean ( $\mu_1$ ) – mean ( $\mu_2$ )  
     paired data, 299–308  
     confidence interval, 303  
     examples of, 300–304  
     exercises for, 305–308  
     overview of, 299–300  
   estimating mean ( $\mu_1$ ) – mean ( $\mu_2$ )  
     independent samples, 267–287  
     cluster effect of data and, 274  
     confidence interval, 267–268  
     confidence interval, variance unequal, 276  
     effect of unequal variances, 274–275  
     examples of, 268–274  
     exercises for, 280–287  
     spatial correlation and, 274  
     statistical test for, 271–274  
      $t$  test for, 275–276  
     Type I errors and, 279  
     weighted average of sampling variances, 268  
   exercises for, 319–339  
   introduction to, 263–267  
     sampling distribution properties, 266–267  
     theorem for sampling distribution, 266  
   sample size, 314–316  
     confidence interval, independent samples, 314  
     confidence interval, paired samples, 316  
     examples of, 315  
     testing independent samples, 315  
     testing paired samples, 316  
   Wilcoxon rank sum test, 287–299  
     calculating rank sum statistics, 288–289  
     examples of, 289–296  
     exercises for, 296–299  
     normal approximation and, 292–295  
     overview of, 288  
     summary of, 289  
     vs.  $t$  test, 295–296  
   Wilcoxon signed-rank test, 308–314  
     computing, 308  
     example of, 309–312

exercises for, 312–314  
 $g$  groups and, 308–309  
 overview of, 309  
 population mean (See mean)  
 population mean, multiple comparisons (See multiple-comparison procedures)  
 population median (See median)  
 population parameters  
   central values and, 193  
   statistical inference and, 341  
 population proportion, inferences about  
   confidence interval for, 472–474  
   examples of, 472–476  
   exercises for, 477–482  
   mean and standard error, 472  
   sample size requirement, 477  
   statistical test for, 475  
 population proportion, inferences about difference between two populations  
   confidence interval for, 483  
   examples of, 483–485  
   exercises for, 486  
   notation for comparing, 482  
   rule for sample size, 484  
   statistical test for, 484–485  
 population proportion, inferences about several proportions  
   examples of, 489, 491–493  
   exercises for, 494–497  
   expected number of outcomes, 490  
   multinomial distribution, 488–489  
   multinomial experiments, 488  
   statistical test for, 490–494  
 population variances, 341–378 (See also variance)  
   case study, 341–344  
   introduction, 341  
   key formulas for, 373–374  
   for single population, 344–355  
     chi-square distribution and, 344–346  
     confidence interval, 346  
     examples of, 346–351  
     exercises for, 351–355  
     statistical test for, 348  
     Type I errors and, 350–351  
     unbiased estimators and, 344  
   for two populations, 365–372  
     examples of, 366–371  
     exercises for, 371–372  
     Hartley's  $F_{max}$  test, 366–368  
     Levine's test, 368–371

- population variances (*contd.*)  
 for two populations, 355–365  
 confidence interval, 359  
 evaluating equal variance condition, 355  
 examples of, 357–363  
 exercises for, 363–365  
*F* distribution and, 356–357  
 reporting conclusions of, 362–363  
 statistical test for, 358
- posterior probabilities, 138
- power curves, 214
- power values  
*t* test and, 237  
*t* test vs. sign test, 248
- predictability analysis  
 measuring strength of relation, 511–512
- prediction (*See* forecasting, linear regression)
- prediction interval  
 defined, 666  
 for linear regression forecasting, 570–571
- preliminary analyses, 1087
- PRESS statistic, 714–715
- primary analyses, 1087
- probability (*See also* odds)  
 basing inferences on, 122–125  
 Bayes' formula, 136–140  
 false positive/false negative test results, 136  
 likelihoods in, 138  
 observable events in, 138  
 prior/posterior probabilities in, 138  
 sensitivity/specificity of diagnostic tests, 137  
 statement of, 138  
 states of nature in, 138
- classical interpretation  
 defined, 124  
 determining probability with, 126
- computation of probability of  
 Type II error, 214, 216
- conditional probability, 131–136  
 conditional vs. unconditional probability, 132  
 defined, 132  
 independent vs. dependent events, 133–134  
 multiplication law and, 132  
 probability of the intersection, 132–133
- definitions of  
 classical interpretation, 124  
 relative frequency interpretation, 124  
 subjective interpretation, 124–125
- determining, 125–128  
 with classical interpretation, 126  
 with relative frequency interpretation, 126–128
- event relations and, 128–131  
 complements of events, 129–130  
 intersection of events, 130  
 mutually exclusive events, 128–129  
 union of events, 130
- exercises for, 187–190. *See also* probability distributions
- measuring with histograms, 50
- Minitab instructions for, 185–186
- calculating binomial probabilities, 185  
 calculating normal probabilities, 185  
 generating random numbers, 185  
 generating sampling distribution, 186
- probability of the intersection, 132–133
- random variables and, 141–142
- probability distributions  
 binomials and, 144–154  
 defining binomial experiments, 144  
 formula for computation of, 146–149  
 mean and standard deviation of, 149–151  
 properties of, 145  
 chi-square distribution, 344–346  
 for continuous random variables, 154–157  
 for discrete random variables, 142–143  
*F* distribution and, 356–357  
 key formulas for, 187  
 normal curve and, 157–166  
 100<sup>th</sup> percentile of distribution, 162–166  
 approximating binomial, 182–185  
 area under, 158–162  
 overview of, 157–158
- random sampling and, 166–171  
 defined, 167  
 random number tables and, 168
- sampling distribution and, 171–182  
 Central Limit Theorem, 175, 179–180  
 contrasting sample histogram with, 180  
 defined, 171  
 examples of, 171–173  
 interpretations of, 180  
 standard of error, 175
- probability plots  
 checking normality, 766  
 identifying outliers, 770
- profile plots  
 constructing, 901  
 two-way interactions, 911  
 using with interactions, 898–899
- proportions  
 defined, 193  
 transformations of data and, 409
- pure experimental error, lack of fit and, 579–580
- quadratic transformations, 739
- qualitative random variables, 141, 623
- quality improvement process, 13–14
- quantitative random variables, 623  
 (*See also* random variables)
- quartiles of distribution, 96
- random-effects model, 978–983  
 AOV moment matching and, 980  
 assumptions of, 978–979  
 defined, 976  
 example of, 981–982  
 exercises for, 982–983  
 expected mean squares and, 979  
 formula for, 978  
 hypothesis tests for, 979–980  
 vs. fixed-effects model, 978–979, 1005
- random-effects model, block design, 983–992  
 assumptions for, 983–984  
 estimating variance components, 986–987  
 example of, 987–989  
 exercises for, 991–992  
 hypothesis testing for, 989–990  
 model and AOV tables for, 985  
 nested sampling and, 990–991

- random error
    - accounting for in linear regression, 557
    - use of random error term, 533
  - random numbers
    - generating, 185
    - statistical table of, 1122
    - tables of, 168
  - random sampling, 166–171
    - defined, 167
    - random number tables and, 168
  - random variables, 141–142, 623
  - randomization, experimental units to treatments, 840–844
    - examples of, 841–842, 844
    - overview of, 840–841
    - randomized complete block design and, 843–845
    - steps in random assignment process, 842–843
  - randomized complete block design, 859–878
    - advantages/disadvantages of, 861
    - analysis of covariance and, 943
    - comparing treatment means, 865–867
    - criteria for, 838–839
    - defined, 83, 861
    - examples of, 868–871
    - exercises for, 872–878
    - experimental randomization and, 843–845
    - extending random-effects model to, 983
    - filtering and, 863–864
    - model for, 863–864
    - overview of, 859–861
    - reducing experimental error with, 834–835
    - relative efficiency of, 867–868
    - types of experimental designs, 29
  - randomized complete block design, one or more missing observations, 1053–1058
    - comparing treatment means, 1056
    - estimating value of missing observation, 1053
    - estimation bias and, 1053–1054
    - examples of, 1054–1055
    - exercises for, 1057–1058
    - fitting complete and reduced models for, 1056–1057
    - testing effect of treatments, 1057
  - range
    - defined, 82
    - interquartile range, 86
  - Studentized range distribution, 444, 1115–1116
  - rank sum statistics
    - calculating, 288–289
    - for more than two population variances, 410
  - ranks, defined, 288
  - ratio estimation, 20
  - raw data source
    - correcting and clarifying, 37, 1086
    - creating database from, 36–37
    - data preparation and, 1084–1085
    - defined, 1084
    - receiving, 36
  - regression analysis (*See also* linear regression; multiple regression)
    - analysis of covariance and, 949–950
    - diagnostic measures of leverage and influence, 546
    - formal assumptions of, 534–535
    - logistic regression and, 675–677
    - straight-line prediction and, 541
  - regression parameters
    - accounting for random error, 557
    - confidence interval for slope, 559–560
    - examples of, 558–561
    - exercises for, 561–566
    - inferences about, 557–566
    - using *F* test for null hypothesis, 560
    - using *t* test for slope, 557–558
  - rejection region, normal curve, 208–210
  - relative efficiency
    - Latin square design, 884
    - randomized complete block design, 867–868
  - relative frequency, defined, 48
  - relative frequency histogram (*See also* histograms)
    - different variability, same mean in, 81
    - examples of, 49–51
  - relative frequency interpretation of probability
    - defined, 124
    - determining probability with, 126–128
  - repeated measures, 1025–1050
    - case study, 1026–1027
    - crossover designs and, 1040–1044
      - analysis of variance for, 1041
      - example of, 1041–1043
      - model for, 1040
    - using Latin square design, 1040
  - defined, 1025
  - exercises for, 1045–1050
  - introduction, 1025–1026
    - crossover designs defined, 1025
    - repeated measure designs defined, 1025
  - single-factor experiments, 1027–1031
    - assumptions for, 1029–1030
    - compound symmetry of observations, 1030–1031
    - reasons for multiple observations, 1027–1029
  - two-factor experiments, 1031–1040
    - examples of, 1032–1034, 1036
    - exercises for, 1038–1040
    - F* tests for, 1034
    - model for, 1031
    - reporting conclusions of, 1038
    - tests for, 1032
- replications
  - defined, 833–834
  - determining number of, 845–848
    - accuracy of estimator specifications and, 845–846
    - examples of, 846–847
    - using *F* test for, 846–848
- report guidelines, 1087–1088
- research hypotheses, 207
- residual plots
  - checking analysis of covariance with, 968–970
  - checking constant variance with, 760
  - checking lack of fit with, 759
  - checking multiple regression assumptions with, 731–737
  - comparing with scatterplots and probability plots, 770
  - detecting outliers with, 765, 770
  - diagnosing collinearity with, 776
- residual standard deviation
  - analysis of covariance and, 951
  - defined, 547
  - measuring prediction accuracy, 590, 632–633
  - random error and, 557
- residuals analysis
  - checking AOV conditions with, 397
  - estimating true error variance with, 546–547
  - using boxplots for, 968

- response variables, 832, 839–840  
robust methods, 238  
root MSE (See residual standard deviation)  
round-off error, 909
- sample histogram vs. sampling distribution, 180
- sample size  
communication hurdles and, 1083–1084  
for mean, 204–207  
computation of, 205  
level of confidence, 204–205  
tolerable error, 204–205  
for mean ( $\mu_1$ ) – mean ( $\mu_2$ )  
confidence interval, independent samples, 314  
confidence interval, paired samples, 316  
examples of, 315  
testing independent samples, 315  
testing paired samples, 316  
population proportion inferences and, 477, 484  
as statistical problem, 7  
testing mean and, 219–224  
two-sided test and, 221  
Type I/Type II errors and, 219–220
- samples  
biased samples, 1082–1083  
standard deviation, 547  
variation, 383–384
- sampling distribution, 171–182  
bootstrap techniques and, 349–350  
Central Limit Theorem, 175, 179–180  
defined, 171  
estimating mean, 196  
estimating mean ( $\mu_1$ ) – mean ( $\mu_2$ ), 266–267  
examples of, 171–173  
generating, 186  
interpretations of, 180  
standard of error, 175  
theorem for, 266  
vs. sample histogram, 180
- sampling techniques, 21–22  
cluster sampling, 20  
ratio estimation, 20  
single random sampling, 21  
stratified random sampling, 21–22  
systematic sampling, 22
- SAS  
checking models with, 742  
generating residual plots with, 760  
statistics software types and, 42
- scatterplots  
checking regression analysis assumptions, 534–537  
detecting outliers (nonlinear relations), 709, 765, 770  
displaying data with multiple variables, 102–103  
estimating model parameters with, 540–541  
key features of, 536–537  
limitations of, 730  
testing multiple regression models for lack of fit, 727–730  
testing variables with, 708–709  
use of smoothers with, 534–535  
using transformation variables with, 535  
vs. residual plots and probability plots, 770
- Scheffé's  $S$  method, 452–458  
AOV for standard designs and, 920  
confidence interval for, 453–454  
example of, 453  
overview of, 452  
steps in, 452
- scientific studies, 27–34  
complicated designs, 30–33  
data collection and, 19  
experimental designs for, 28–30  
completely randomized design, 28–29  
Latin square design, 29  
randomized block design, 29  
factorial experiments, 30–33  
applications of, 32  
defined, 32  
interaction of factors in, 31  
one-at-a-time approach, 30–31  
use of factors, 30  
overview of, 27–28  
vs. observational studies, 34–35
- self-administered questionnaires, data collection, 25
- sensitivity, of diagnostic tests, 137
- separate-variance  $t$  tests, 276
- sequential sums of squares (SS), 647
- serial correlation  
of data, 274  
Durbin-Watson test statistic for, 773
- first differences approach and, 776  
positive/negative correlation, 773  
time series data and, 773
- side-by-side boxplots, 104–107
- sign test  
comparing with  $t$  test, 248  
statistical table of, 1096  
testing one-sided hypotheses with, 246
- simple linear regression (See linear regression)
- simple logistic regression model, 675–677 (See also logistic regression)
- single-factor experiments, repeated measures, 1027–1031  
assumptions for, 1029–1030  
compound symmetry of observations, 1030–1031  
reasons for multiple observations, 1027–1029
- single random sampling, 21
- skeletal boxplot, 96–97
- skewness (See also normal curve)  
checking normality, 765  
measures of central tendency and, 76–77  
skewed distributions, 236  
skewed right or left histograms defined, 54  
example of, 53
- slope  
comparing, 670–675  
examples of, 670–674  
exercises for, 674–675  
extending comparison to three or more lines, 674  
intersecting lines vs. parallel lines, 671  
confidence interval for, 559–560  
defined, 532  
least-squares estimates of, 542  
in linear regression, 540  
in multiple regression, 630, 632  
random error and, 557
- smoothers, 535, 759
- software packages, for statistics, 42–43
- spatial correlation, of data, 274
- specificity of diagnostic tests, in Bayes' formula, 137
- spline fit, 535
- split-plot design  
nested sampling and split-plot design, 1014–1018  
subplot analysis, 1018



- vs. factorial experiments, 999
- vs. standard two-factor experiments, 1019
- wholeplot analysis, 1017
- SPSS, statistics software, 42
- squared prediction error, 541
- SS(Error) (See sum of squares residuals-SS(Residual))
- SS(Model) (See sum of squares regression-SS(Regression))
- SS(Regression) (See sum of squares regression-SS(Regression))
- SS(Residual) (See sum of squares residuals-SS(Residual))
- SS(Total) (See sum of squares, total (TSS))
- stacked bar graphs, 102
- standard deviation, 88–89 (See also population variances)
  - applications of, 341
  - approximating, 91–92
  - binomial probability distribution and, 149–151
  - defined, 88
  - estimating mean and, 201–202
  - in multiple regression, 652, 686
  - multiple regression forecasting and, 666
  - population parameters, 193
  - transformations of data and, 407
- standard error, 175, 472 (See also residual standard deviation)
- standard normal curve, statistical table of, 1091–1092
- standardized residuals, zero expectation and, 759
- states of nature, 138
- statistical inference
  - based on probability, 122–125
  - data description and, 40–41
  - guidelines for, 1087–1088
  - hypothesis testing and, 193
  - linear regression, 557–566
    - accounting for random error, 557
    - confidence interval for slope, 559–560
    - examples of, 558–561
    - exercises for, 561–566
    - using  $F$  test for null hypothesis, 560
    - using  $t$  test for slope, 557–558
  - median ( $M$ ), 243–250
    - approximating for large samples, 245–248
    - confidence interval for, 243–245
    - sign test vs.  $t$  test, 248
    - summarizing statistical test for, 246–247
    - using sign test, 246
  - multiple regression
    - coefficient of determination and, 646
    - colinearity and, 646–647
    - examples of, 647–651, 653–655
    - exercises for, 655–657
    - $F$  test of all coefficients, 648–651, 655
    - $t$  test of individual coefficient, 652–655
  - population central values, single population, 251–253
  - population central values, two populations
    - calculating rank sum statistics, 288–289
    - computing, 308
    - examples of, 289–292, 309–312
    - exercises for, 296–299, 312–314
    - $g$  groups and, 308–309
    - normal approximation and, 292–295
    - overview of, 288, 309
    - summary of, 289
    - using  $t$  test for, 295–296
  - population parameters and, 341
  - population proportions, difference between two, 482–488
    - confidence interval for, 483
    - examples of, 483–485
    - exercises for, 486
    - notation for comparing, 482
    - rule for sample size, 484
    - statistical test for, 484–485
  - population proportions, several, 488–497
    - examples of, 489, 491–493
    - exercises for, 494–497
    - expected number of outcomes, 490
    - multinomial distribution, 488–489
    - multinomial experiments, 488
    - statistical test for, 490–494
  - population proportions, single, 471–482
    - confidence interval for, 472–474
    - examples of, 472–476
    - exercises for, 477–482
    - mean and standard error, 472
    - sample size requirement, 477
    - statistical test for, 475
- statistical reports, 1088
- statistical tables, 1091–1129
  - $F$  test power curves for AOV, 1123–1126
  - percentage points for Dunnett's test, 1117–1120
  - percentage points for Hartley's  $F_{max}$  test, 1121
  - percentage points of chi-square distribution, 1100–1101
  - percentage points of  $F$  distribution, 1102–1113
  - percentage points of sign test, 1096
  - percentage points of Student's  $t$  distribution, 1093
  - percentage points of the Studentized range, 1115–1116
  - percentage points of Wilcoxon rank sum test, 1097
  - percentage points of Wilcoxon signed-rank test, 1098–1099
  - Poisson probabilities, 1127–1129
  - random numbers, 1122
  - standard normal curve, 1091–1092
  - values of 2 arcsin, 1114
- statistical tests (See also hypothesis testing;  $t$  test)
  - five parts of, 208
  - for mean ( $\mu$ ), 231
  - for mean ( $\mu_1$ ) – mean ( $\mu_2$ ) independent samples, 271–274
  - for median ( $M$ ), 246–247
  - for more than two population variances, 384–393
    - completely randomized design for, 387–388
    - developing test statistic for, 386–387
    - examples of, 389–392
    - exercises for, 392–393
    - mean square and, 388
    - sum of squares and, 388
    - summarizing results in AOV table, 389
    - using multiple  $t$  tests, 385
  - population proportion inferences and, 475
  - for single population variance, 348
  - $t$  test power curves and, 1094–1095

symmetry of observations, com-  
pound, 1030–1031  
systematic sampling, 22

*t* statistic (See *t* tests)

*t* tests

accuracy of, 237–238  
evaluating equal variance, 355  
for mean ( $\mu$ ), 208  
for mean ( $\mu_1$ ) – mean ( $\mu_2$ )  
independent samples, 275–276  
paired data, 303  
for more than two population vari-  
ances, 385  
for slope, 557–558  
statistical table of power curves  
and, 1094–1095  
vs. sign test, 248  
vs. Wilcoxon rank sum test,  
295–296  
vs. Wilcoxon signed-rank test, 311  
telephone interviews, data collection,  
24–25  
terms, in completely randomized de-  
sign, 394–395

test statistic (See *t* tests)

tests

Poisson distribution, 498–499  
regression coefficients  
complete and reduced models  
for, 658  
example of, 658–661  
exercises for, 661–665  
*F* test of predictors, 657–658  
tests of significance  
factorial treatment in com-  
pletely randomized de-  
sign, 902–903  
mixed-effects model, 992–993

theorems

Central Limit Theorem, 175,  
179–180  
sampling distribution, 266

time series

overview of, 57–58  
serial correlation and, 773  
uses of, 59

tolerable error, 204–205

total sum of squares (TSS) (See sum  
of squares, total (TSS))

transformations

for central values, more than two  
populations  
defined, 403  
examples of, 404–409

exercises for, 409

selecting new variables,  
403–404

choosing, 537–538

inverse transformations and, 739

logarithmic, 538, 738

for more than two population vari-  
ances, 403–409

quadratic, 739

straightening scatterplot graphs  
with, 535–536

treatment means

adjusted treatment means, 948,  
950–951

balanced incomplete block de-  
signs, 1067–1068

comparing, 916–922

case study, 921–922

examples of, 917–921

multiple comparison proce-  
dures and, 920

using Fisher's LSD procedure,  
916–917

for Latin square design, 1061

for randomized block design,  
1056

treatments

control treatments, 833

defined, 623, 832

design of, 892

factorial treatment design, 832

observational studies and, 830

randomly assigning experimental  
units to, 840–841

trial-and-error, selecting multiple re-  
gression models, 730

true error variance, linear regression,  
546–547

Tukey's *W* procedure, 444–447

AOV for standard designs and,  
920

confidence interval for, 447

example of, 445–446

experimentwise error rate in, 445  
steps in, 445

use of Studentized range distribu-  
tion in, 444

vs. LSD procedure, 446

vs. SNK procedure, 447–448

two-factor experiments, repeated  
measures, 1031–1040

examples of, 1032–1034, 1036

exercises for, 1038–1040

*F* tests for, 1034

model for, 1031

reporting conclusions of, 1038  
tests for, 1032

two-tailed tests

computing probability for, 216

overview of, 211

sample size for, 221

two-way decision process, 209

Type I errors

Bonferroni inequality and, 438

defined, 209

mean ( $\mu$ ) and, 219–220

mean ( $\mu_1$ ) – mean ( $\mu_2$ ) and, 279

multiple *t* tests and, 385

probability of, 218–219

for single population variance,  
349–351

for two population variances, 363

Type II errors

Bonferroni inequality and, 438

defined, 209

mean ( $\mu$ ) and, 219–220

multiple *t* tests and, 385

probability of, 214–216, 218–219,  
234

for two population variances, 363

unbalanced designs, 1053

unbiased estimators

for completely randomized design,  
857

population variances and, 344

for randomized complete block de-  
sign, 866

variance and, 87

unconditional probability, 132

underspecification, 715

unequal population variances,  
274–275

uniform histogram

defined, 54

example of, 52

unimodal/bimodal histograms, 52–54

defined, 52

examples of, 52–54

union of events, 130

unique predictive value, 647

unit of association, in prediction, 532

variability, measures of

deviation, 86–87

Empirical Rule for, 89–93

exercises for, 93–96

interquartile range, 86

overview of, 81–82

percentiles, 82–86

- variability, measures of (*contd.*)
  - range, 82
  - standard deviation, 88–89
  - variance, 87–88
  - vs. measures of central tendency, 81
- variables (*See also* factors)
  - covariates and, 839–840
  - in designed experiments, 832
  - multiple, 101–109
    - cluster bar graphs, 102–103
    - contingency tables and, 101–102
    - exercises for, 107–109
    - scatterplots, 102–103
    - side-by-side boxplots, 104–107
    - stacked bar graphs, 102
  - selecting for multiple regression, 707–727
    - backward elimination, 717–720
    - best subset regression, 716–717
    - collinearity and, 708–710
    - data-splitting approach, 714
    - examples for, 709–722
    - exercises for, 722–727
    - overview of, 707–708
    - performing all possible regressions, 711–714
  - PRESS statistic and, 714–715
  - stepwise regression, 718, 720–722
  - underspecification/over-specification, 715
  - too many independent variables, 711
  - types, 141–142
    - continuous random variables, 142
    - discrete random variables, 142
    - qualitative random variables, 141
    - quantitative random variables, 141–142
- variance (*See also* analysis of variance (AOV); population variances)
  - components
    - defined, 975
    - estimating, 986–989
  - defined, 87–88
  - in designed experiments, 835–836
  - reducing through blocking, 838–839
- variance inflation factor (VIF)
  - defined, 652
  - diagnosing collinearity with, 709
- weighted average, sampling variances, 268
- weighted least-squares, 762–763
- Wilcoxon rank sum test, 287–299
  - calculating rank sum statistics, 288–289
  - normal approximation and, 292–295
  - overview of, 288
  - statistical table of, 1097
  - summary of, 289
  - vs. *t* test, 295–296
- Wilcoxon signed-rank test, 308–314
  - computing, 308
  - example of, 309–312
  - exercises for, 312–314
  - g* groups and, 308–309
  - overview of, 309
  - statistical table of, 1098–1099
  - vs. paired *t* test, 311
- work files
  - from machine-readable database, 38
  - sets of data files and, 1086
- zero expectation, multiple regression, 759–760

## 译 后 记

《统计方法与数据分析引论》一书终于与读者见面了,作为译者由衷地高兴。本书是独具特点的统计学教科书:第一,本书内容丰富,既包括了初等统计教科书中常见的内容,也包括了诸如试验设计这些统计的入门书中不太常见的内容;第二,书中概念和方法讲解直观清楚,尤其是结合大量的实际问题,把统计思想分析得很透彻,有利于读者通过阅读领会统计概念和方法的本质;第三,阅读本书只需要高中数学知识,不需要高等数学,因而有利于工程技术人员和社会科学等其他方面的实际工作者作为学习统计学的起步教材;第四,书中包含大量来自于实际的问题作为案例、例题和习题,如大部分章节都有案例,全书共有 1022 个练习题,涉及工程、医药、农业、生物、化学、教育、商业、调查科学、人力资源、心理学等几十个方面,这对于应用统计工作者和从事统计教学的教师都是难得的素材。本书虽看起来篇幅较长,但读起来容易。由此,译者认为,不论对于研究生水平的统计学教学还是对于本科生水平的教学,不论对于统计学或概率统计学专业还是对于其他开设统计学课程的专业,都是一本难得的教学参考书。

由于本书的性质,原作者在写作和语言组织方面并不追求学术上的严格,语言生动活泼,词语变化灵活,不拘一格,译者努力保持原书的风格,但也避免辞藻过繁而造成歧义,个别地方做了注解;另一方面,译者也遇到众多科学领域中的术语的翻译问题,尽管已经做出很大努力(例如译者曾就医学方面的词汇请教过协和医院的教授),但毕竟涉及面太广而时间紧迫,难免有未尽人意之处:个别药名、植物名尚无译名,仍用原文;书中列有大量统计软件的输出结果,这些结果保持软件输出的原始形式,尽管熟悉统计软件的读者很容易读懂这些输出,但为了阅读方便,我们在一些地方增加了适当的阅读说明。另外,为了读者方便,在译者认为有必要的地方(如国内尚无统一翻译的名词)在括号内加注了原文。

王建稳翻译了第五到第十章,王强翻译了第十一章,杨中华翻译了第十二章部分,其余由张忠占主译,最后由张忠占统一了全书译稿。在翻译和统稿过程中,改正了原书中的一些错误。

译者首先感谢科学出版社杨波先生和卢秀娟女士和刘晓炜女士在本书的翻译和出版过程中给予的支持和鼓励,也感谢他们慧眼识书,选择了这样一本很有特点的统计学教科书;同时感谢陈希孺院士和北京工业大学、北方工业大学和对外经济贸易大学各位同事的帮助;译者也非常感谢北京工业大学概率论与数理统计学科的研究生谢田法、李春红、吕颖、孔祥荣、王莉丽以及其他各位的帮助,这些帮助加

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识



速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识

速了翻译的进程;最后,译者也借此感谢他们的家人给予的支持。

由于水平所限,错讹之处难免,欢迎读者不吝赐教。

译者谨识