

# STATISTICS FOR BUSINESS AND ECONOMICS

12th Edition

## 商务与经济统计

(原书第12版)

戴维 R. 安德森 (David R. Anderson) 辛辛那提大学  
丹尼斯 J. 斯威尼 (Dennis J. Sweeney) 辛辛那提大学  
[美] 托马斯 A. 威廉斯 (Thomas A. Williams) 罗切斯特理工学院 著  
杰弗里 D. 卡姆 (Jeffrey D. Camm) 辛辛那提大学  
詹姆斯 J. 科克伦 (James J. Cochran) 路易斯安那理工大学

张建华 王健 冯燕奇 等译

各种电子书 欢迎咨询

Q群: 75490529 (图书定制)

淘宝: <http://shop108402496.taobao.com/>



附光盘



机械工业出版社  
China Machine Press

Statistics for Business and Economics

(12th Edition)

# 商务与经济统计 (原书第12版)

这是一本广泛流行于美国高校并被誉为最经典的商务统计教材！该书在中国引进十多年来，一直被北京大学、清华大学、中国人民大学、复旦大学、上海交通大学、南开大学和中山大学等众多高校采用，深受师生推崇和喜爱。

应用性强是本书的最大特色。第12版在保留了以前版本的叙述风格和可读性的基础上，对内容进行了一定的修订，对个别章节做了更为合理的调整，并更新了一定数量的习题。作者精心设计了“方法”“应用”和“补充练习”三种题型，并设计了起提示、总结和建议作用的“注释和评论”，这些都体现出本书的实用特点。

本书适合本科生、MBA和研究生等多层次学者学习，也可作为从事工商行政管理和经济分析的各类人员手边的工具参考书。

各种电子书 欢迎咨询

Q群：75490529 (图书定制)

淘宝：<http://shop108402496.taobao.com/>

作者：戴维 R. 安德森  
(David R. Anderson)  
ISBN: 978-7-111-35029-3  
定价：109.00元

作者：道格拉斯 A. 林德  
(Douglas A. Lind)  
ISBN: 978-7-111-27083-6  
定价：82.00元

作者：道格拉斯 A. 林德  
(Douglas A. Lind)  
ISBN: 978-7-111-50710-9  
定价：89.00元



CENGAGE  
Learning  
[www.cengageasia.com](http://www.cengageasia.com)

投稿热线: (010) 88379007  
客服热线: (010) 88379210 88361066  
购书热线: (010) 68326294 88379649 68995259

华章网站: [www.hzbook.com](http://www.hzbook.com)  
网上购书: [www.china-pub.com](http://www.china-pub.com)  
数字阅读: [www.hzmedia.com.cn](http://www.hzmedia.com.cn)



CENGAGE  
Learning  
圣智学习





# STATISTICS FOR BUSINESS AND ECONOMICS

12th Edition

## 商务与经济统计

(原书第12版)

戴维 R. 安德森 (David R. Anderson) 辛辛那提大学

丹尼斯 J. 斯威尼 (Dennis J. Sweeney) 辛辛那提大学

[美] 托马斯 A. 威廉斯 (Thomas A. Williams) 罗切斯特理工学院 著

杰弗里 D. 卡姆 (Jeffrey D. Camm) 辛辛那提大学

詹姆斯 J. 科克伦 (James J. Cochran) 路易斯安那理工大学

张建华 王健 冯燕奇 等译

各种电子书 欢迎咨询

Q群: 75490529 (图书定制)

淘宝: <http://shop108402496.taobao.com/>



机械工业出版社

## 图书在版编目 (CIP) 数据

商务与经济统计 (原书第 12 版) / (美) 安德森 (Anderson, D. R.) 等著; 张建华等译. —北京: 机械工业出版社, 2015.8

(华章教材经典译丛)

书名原文: Statistics for Business & Economics

ISBN 978-7-111-51273-8

I. 商… II. ①安… ②张… III. ①商业统计—教材 ②经济统计—教材 IV. ①F712.3 ②F222

中国版本图书馆 CIP 数据核字 (2015) 第 200387 号

本书版权登记号: 图字: 01-2014-0610

David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran.  
Statistics for Business & Economics, 12th Edition

Copyright © 2014, 2012 South-Western, a part of Cengage Learning.

Original edition published by Cengage Learning. CMP Press is authorized by Cengage Learning to publish and distribute exclusively this simplified Chinese edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

All rights reserved.

本书原版由圣智学习出版公司出版。版权所有, 盗印必究。本书中文简体字翻译版由圣智学习出版公司授权机械工业出版社独家出版发行。此版本仅限在中华人民共和国境内 (不包括中国香港、澳门特别行政区及中国台湾地区) 销售。未经授权的本书出口将被视为违反版权法的行为。未经出版者预先书面许可, 不得以任何方式复制或发行本书的任何部分。

本书封底贴有

本书是美国辛  
留了以前版本的叙  
数量的习题。应用  
总结和建议作用的  
本书既可作为

各种电子书 欢迎咨询

Q群: 75490529 (图书定制)

淘宝: <http://shop108402496.taobao.com/>

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 左 萌

责任校对: 董纪丽

印 刷: 北京诚信伟业印刷有限公司

版 次: 2015 年 9 月第 1 版第 1 次印刷

开 本: 214mm×275mm 1/16

印 张: 33.25

书 号: ISBN 978-7-111-51273-8

定 价: 109.00 元 (附光盘)

ISBN 978-7-89405-830-0 (光盘)

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88379210 88361066

投稿热线: (010) 88379007

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjg@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光/邹晓东



## THE TRANSLATOR'S WORDS

统计学是研究不确定性现象数量规律性的方法论科学,在众多的专业、学科领域中,都起着重要的作用。应用统计学的概念与方法来处理商务与经济中的各种问题,就产生了商务与经济统计学,它是解决商务与经济中各种问题的有力工具之一。

戴维 R. 安德森等五位教授合著的《商务与经济统计》是一本很有特色的教材,我们曾将该书的第 7 版、第 9 版、第 10 版和第 11 版译成中文介绍给我国的读者,并得到众多读者的认可与积极的反应。2014 年《商务与经济统计》的第 12 版问世了,我们愿以最快的速度再次将新版的中文译本奉献给读者。

新版的《商务与经济统计》保留了以前版本的叙述风格与可读性,其最大特色是应用性强。本书以大量数据为基础,介绍各种统计方法在实际中的应用,每一种统计方法的介绍都联系一个案例,并配有大量的例题和练习。新版的实际数据基本来自《华尔街日报》《今日美国》《财富》等报刊,这使得学生不但对这些统计信息产生兴趣,而且还学习了统计方法及其应用。

本书的另一特色是通俗易懂。本书是为应用者准备的,全书避免了烦琐的数学推导,采用深入浅出、循序渐进的方法系统地介绍了统计学的知识。叙述严谨,基础坚实,实例与图表丰富,易于读者理解与掌握。

本书在每章附录中给出了利用 Minitab, Excel 和 StatTools 进行各种统计分析的程序步骤,并在书后的附录中详细介绍了 Excel 2010,使得学生能够很容易地利用 Minitab, Excel 和 StatTools 完成各种统计分析的运算。

本书安排了大量练习题,用来帮助学生更好地理解书中讲述的内容。其中带有“☆”号的练习题为自测题,以评价学生对书中介绍的各种统计方法的掌握程度。为方便读者自学,在附录 C 中给出了部分习题的详细解答。

作为教材,本书有较多的应用层次,既可用作研究生、MBA 和本科生的教材,也可供从事工商管理分析和经济分析的各类人员参考。读者可根据时间和需要,有选择地学习有关内容。

为了忠实于原著,我们在翻译过程中没有做任何改动。考虑到我国读者的习惯,我们将原书页边注释改为了页下注。

本书的译者如下:

第 1~3 章:张建华、张丽华

第 4~9 章:王健

第 10 章:聂巧平、叶光

第 11~12 章:张建华

第 13~14 章:冯燕奇、刘东红

第 15~16 章:冯蕾、张坤

第 17~20 章:张建华、邹华

全书由张建华负责整理和通校,黄河、陈光、王海花、杨莉、杜静、刘伟、周威至、谭欢、刘渊泽、马鑫、黄岩和李超然也参加了本书的部分翻译和校对工作,王公恕通读了全书译稿,并提出了宝贵的意见。

本书在翻译过程中得到了机械工业出版社有关领导的指导帮助,特别要感谢本书的编辑左萌、杨晓莉,他们严谨的工作作风、尽职尽责的工作态度促使本书能在最短的时间内与读者见面。

囿于译者的水平,译文中的疏漏和差错在所难免,恳请读者批评指正。

译者

2014 年 8 月于南开大学

本书是《商务与经济统计》第12版,本版我们欢迎两位著名的学者:辛辛那提大学的杰弗里 D. 卡姆和路易斯安那理工大学的詹姆斯 J. 科克伦,加入我们的作者团队。杰弗里和詹姆斯二人在统计与商务分析领域都是有造诣的教师、研究员和从业人员。詹姆斯是美国统计协会会员。你可以从随后的作者简介部分了解到他们更多的成就。我们相信杰弗里和詹姆斯的加入,将保持和提高本书想达到的效果。

本书的目的是向学生,主要是向那些工商管理 and 经济学研究领域的学生概念性地介绍统计学及其各种应用方面的知识。本书是以应用为导向,并考虑到非数学专业人员的需要而编写的,所需要的必备数学基础为代数知识。

数据分析与统计方法的应用是本书的结构及讲述内容中一个不可或缺的部分。每种方法的讨论与发展都通过应用呈现出来,并运用统计结果进行问题的决策和解答。

尽管本书以应用为导向,但我们还是谨慎地给出了合理的方法推导过程,并且运用了所涉及的论题通常会使用的符号。因此,同学们将会发现,本书为学习高级统计学打下了坚实的基础。附录中包括指导进一步学习的参考书目。

本书向学生介绍 Minitab 16 与 Microsoft® Excel 2010 软件包,并强调在统计分析应用中计算机软件包的作用。之所以介绍 Minitab,是因为在教学和统计实践中它都是最主要的统计软件包之一;Excel 虽然不是统计软件包,但它的广泛实用性和应用性,对于学生了解这个软件包的统计性能很重要。Minitab 和 Excel 的步骤在各章附录中会给出,使得在更多强调计算机运用的课堂中,教师可以灵活使用。StatTools 是由 Palisades 公司开发的一个商用 Excel 插件,对 Excel 用户扩展了统计选项的范围。在本书第1章的附录中,我们演示了如何下载和安装 StatTools。大部分章节都有一个演示使用 StatTools 完成统计程序所需步骤的附录。我们将 StatTools 变成可选的,因此对那些讲授仅仅使用 Excel 中可利用的标准统计工具的教师来说,可以这样做。

## 第12版的变化

我们对《商务与经济统计》以前版本得到的认可与积极的反应表示感谢。因此,在这次新版修订时,我们保留了以前版本的叙述风格与可读性。全书有许多变化旨在提高其教学效果。新版中的显著变化汇总如下。

### 内容修订

- **描述统计学——第2章和第3章** 我们加入了数据可视化、最佳实践等许多新资料,特别修订了这两章。第2章增加了包括复合条形图和结构条形图的新资料以及数据可视化和创建有效图形显示的最佳实践这一新的小节。第3章在位置的度量一节中增加了对几何平均数的叙述。几何平均数在金融投资的增长率、年百分比等计算中有许多应用。第3章还新增了一节:数据仪表板和如何加入汇总统计以增强其效果。
- **离散型概率分布——第5章** 为了更好地解释概率分布的作用以及第4章中提到的概率分配的资料如何用于建立离散型概率分布,我们对这一章的基础材料进行了修订。我们指出采用分配概率的相对频数法建立经验离散概率分布。应许多用户的要求,我们增加了包括二元离散分布和在金融上的应用这一新的小节(5.4节)。我们展示了如何使用这些分布,构建金融资产组合和分析。
- **多个比率的比较、独立性检验及拟合优度检验——第12章** 这一章做了很大修改。我们增加了三个或多



个总体比率相等性的检验的新的一节，这一节包括一个进行总体比率所有配对之间多重比较检验的程序。为了阐明检验涉及的两个分类变量的独立性，独立性检验这一节被重新改写，修订了包括 Minitab, Excel 和 StatTools 的分步指令。

- **新案例** 本版我们增加了 8 个新案例，使得书中的案例总数达到 31 个。在第 2 章和第 3 章增加了 3 个新的描述统计案例。在有关回归的第 14 章、第 15 章和第 16 章中增加了 5 个新案例。这些案例为学生提供了分析较大数据集并以分析结果为基础准备管理报告的机会。
- **“实践中的统计”的新应用** 每一章都以一篇“实践中的统计”的文章开始，这些文章描述了该章将要介绍的统计方法的应用。本版中新的“实践中的统计”是第 2 章描述的数据仪表板和辛辛那提动物园数据可视化的使用。我们在第 4 章还增加了一个新的“实践中的统计”，它描述 NASA 团队如何利用概率帮助营救由于矿井塌方而受困的 33 名智利矿工。
- **以实际数据为基础的新例题与练习** 我们继续努力用最近的实际数据和统计信息资料来更新本书的例题和练习。本版增加了近 180 道新例题和练习，这些例题和练习是以实际数据和原始资料为基础的。利用《华尔街日报》《今日美国》《巴伦周刊》以及一系列其他资料来源，我们进行了实际研究，以说明并创建练习来演示商务与经济统计的多种应用。我们相信，实际数据的使用可使更多的学生对统计资料产生兴趣，并使学生既学习统计方法，又学习其应用。本书第 12 版包含近 350 道以实际数据为基础的例题和练习。

## 特色与教学

作者安德森、斯威尼、威廉斯、卡姆和科克伦继续保留了以前版本中的许多特色。

### 方法练习与应用练习

每节后的练习分成两部分，即“方法”和“应用”。方法练习要求学生利用公式进行必要的计算。应用练习要求学生利用书中的实际资料。这样，学生首先把注意力集中在“基本问题”的计算上，然后再转向精巧的统计应用与解释上。

### 自测题

某些练习作为自测题（本书用“☆”标注）出现，它们的完整解答在附录 C 中给出，学生们可试着做自测题并核对答案，以便评估你对书中各章讲述的概念的理解程度。

### 页下注解与注释

关键点以及对学生提供额外知识的页下注解是本书的一个重要特色。设计这些出现在页下的注解，是为了强调和提高对本书介绍的术语和概念的理解。

在许多节的末尾提供了“注释”，以帮助更深入地了解统计方法及其应用。“注释”中包括一些注意事项，如方法的局限性、对应用的建议、对其他方法的简要描述及其他事项。

### 本书附带的数据文件

本书所附光盘有大量可供我们使用的数据文件。<sup>①</sup> 这些数据集既可以在 Minitab 中，也可以在 Excel 格式中使用。所有案例的数据集和较大练习的数据集都放在光盘上。

① 本书配套光盘中除了提到的数据文件，还有各章的单章习题、各章附录。

## 感谢

我们要感谢编审的工作，他们提供的意见和建议有助于继续完善本书。他们是：

AbouEl-Makarim	Reidar Hagtvedt	J. G. Pitt
Aboueissa, University of Southern Maine	University of Alberta School of Business	University of Toronto
Kathleen Arano	Clifford B. Hawley	Scott A. Redenius Brandeis University
Fort Hays State University	West Virginia University	Sandra Robertson
Musa Ayar	Vance A. Hughey	Thomas Nelson
Uw-baraboo/Sauk County	Western Nevada College	Community College
Kathleen Burke	Tony Hunnicutt	Sunil Sapra
SUNY Cortland	Ouachita Technical College	California State University, Los Angeles
YC Chang	Stacey M. Jones	Kyle Vann Scott
University of Notre Dame	Albers School of Business and Economics, Seattle University	Snead State Community College
David Chen	Dukpa Kim	Rodney E. Stanley
Rosemont College and Saint Joseph's University	University of Virginia	Tennessee State University
Margaret E. Cochran	Rajaram Krishnan	Jennifer Strehler
Northwestern State University of Louisiana	Earlham College	Oakton Community College
Thomas A. Dahlstrom	Robert J. Lenke	Ronald Stunda
Eastern University	Lake Forest College	Valdosta State University
Anne Drougas	Philip J. Mizzi	Cindy van Es
Dominican University	Arizona State University	Cornell University
Fesseha Gebremikael	Mehdi Mohaghegh	Jennifer VanGilder
Strayer University/ Calhoun Community College	Norwich University	Ursinus College
Malcolm C. Gold	Mihail Motzev	Jacqueline Wroughton
University of Wisconsin— Marshfield/Wood County	Walla Walla University	Northern Kentucky University
Joel Goldstein	Somnath Mukhopadhyay	Dmitry Yarushkin
Western Connecticut State University	The University of Texas at El Paso	Grand View University
Jim Grant	Kenneth E. Murphy	David Zimmer
Lewis & Clark College	Chapman University	Western Kentucky University
	Ogbonnaya John Nwoha	
	Grambling State University	
	Claudiney Pereira	
	Tulane University	



对本书和前几版的构思提供有益意见和建议的许多同事和朋友，我们继续表示感谢。他们是：

Mohammad Ahmadi	Gopal Dorai	John Leschke
University of Tennessee	William Patterson	University of Virginia
at Chattanooga	University	
Lari Arjomand	Nicholas Farnum	Martin S. Levy
Clayton College and State	California State	University of Cincinnati
University	University—Fullerton	
Robert Balough	Donald Gren	John S. Loucks
Clarion University	Salt Lake Community	St. Edward's University
	College	
Philip Boudreaux	Paul Guy	David Lucking-Reiley
University of Louisiana	California State	Vanderbilt University
Mike Bourke	University—Chico	
Houston Baptist	Clifford Hawley	Bala Maniam
University	West Virginia University	Sam Houston State
James Brannon	Jim Hightower	University
University of Wisconsin—	California State	
Oshkosh	University, Fullerton	Don Marx
John Bryant	Alan Humphrey	University of Alaska,
University of Pittsburgh	University of Rhode Island	Anchorage
Peter Bryant	Ann Hussein	Tom McCullough
University of Colorado	Philadelphia College of	University of California—
	Textiles and Science	Berkeley
Terri L. Byczkowski	C. Thomas Innis	Ronald W. Michener
University of Cincinnati	University of Cincinnati	University of Virginia
Robert Carver	Ben Isselhardt	Glenn Milligan
Stonehill College	Rochester Institute of	Ohio State University
	Technology	
Richard Claycombe	Jeffery Jarrett	Mitchell Muesham
McDaniel College	University of Rhode	Sam Houston State
Robert Cochran	Island	University
University of Wyoming		
Robert Collins	Ronald Klimberg	Roger Myerson
Marquette University	St. Joseph's University	Northwestern University
David W. Cravens	David A. Kravitz	
Texas Christian	George Mason	Richard O'Connell
University	University	Miami University of Ohio
Tom Dahlstrom	David Krueger	Alan Olinsky
Eastern College	St. Cloud State University	Bryant College
		Ceyhun Ozgur
		Valparaiso University
		Tom Pray
		Rochester Institute of
		Technology

Harold Rahmlow	California	Paso
St. Joseph's University	Hroki Tsurumi	Andrew Welki
H. V. Ramakrishna	Rutgers University	John Carroll University
Penn State University at	David Tufte	Ari Wijetunga
Great Valley	University of New	Morehead State University
Tom Ryan	Orleans	J. E. Willis
Case Western Reserve	Victor Ukpolo	Louisiana State University
University	Austin Peay State	Mustafa Yilmaz
Bill Seaver	University	Northeastern University
University of Tennessee	Ebenge Usip	Gary Yoshimoto
Alan Smith	Youngstown State	St. Cloud State University
Robert Morris College	University	Yan Yu
Willbann Terpening	Cindy Van Es	University of Cincinnati
Gonzaga University	Cornell University	Charles Zimmerman
Ted Tsukahara	Jack Vaughn	Robert Morris College
St. Mary's College of	University of Texas-El	

我们感谢为本书提供“实践中的统计”特色文章的工商业界的合作者，我们通过每一篇文章的来源附加认识了他们中的每个人。我们还要感谢高级策划编辑 Charles McCormick Jr.，初级编辑 Maggie Kubale，内容规划经理 Tamborah Moore，MPS Limited 的项目经理 Lynn Lustberg，媒体编辑 Chris Valentine 以及在本书筹备期间给予支持和编辑建议的 Cengage South-Western 的所有人。

戴维 R. 安德森  
丹尼斯 J. 斯威尼  
托马斯 A. 威廉斯  
杰弗里 D. 卡姆  
詹姆斯 J. 科克伦



## ABOUT THE AUTHOR

## 戴维 R. 安德森 (David R. Anderson)

戴维 R. 安德森是辛辛那提大学工商管理学院数量分析系教授。他出生在北达科他州大福克斯市，在普度大学获得学士、硕士和博士学位。安德森教授担任数量分析与运作管理系主任，并担任工商管理学院的副院长。此外，他还是学院首届教学大纲的协调人。

在辛辛那提大学，安德森教授不但为商科专业的学生讲授基础统计学，而且还讲授研究生水平的回归分析、多元分析和科学管理课程。他也在华盛顿特区的美国劳工部讲授统计学课程。他因在教学上和对学生的组织服务方面的突出成就而荣获提名与奖励。

安德森教授已在统计学、管理科学、线性规划以及生产与运作管理领域与他人合作出版了 10 部著作。他是一位活跃在抽样和统计方法领域的咨询顾问。

## 丹尼斯 J. 斯威尼 (Dennis J. Sweeney)

丹尼斯 J. 斯威尼是辛辛那提大学数量分析系教授和生产率提高中心主任。他出生在衣阿华州得梅因市，在德雷克大学获得工商管理学士学位，在印第安纳大学获得工商管理硕士和工商管理博士学位，并成为 NDEA 会员。斯威尼教授曾在宝洁公司管理科学小组工作，并在杜克大学作了一年的客座教授。斯威尼教授担任辛辛那提大学数量分析系主任和工商管理学院的副院长。

斯威尼教授已在管理科学与统计学领域发表和出版了 30 多篇论文和专著。国家自然科学基金、IBM 公司、宝洁公司、美国联合百货 (Federated Department Stores)、美国克罗格公司 (Kroger) 和辛辛那提天然气和电气公司等都曾对他的研究给予资助，这些研究的成果在《管理科学》《运筹学》、*Mathematical Programming*、《决策科学》等杂志上发表。

斯威尼教授在统计学、管理科学、线性规划、生产与运作管理等领域已与他人合作出版了 10 部专著。

## 托马斯 A. 威廉斯 (Thomas A. Williams)

托马斯 A. 威廉斯是罗切斯特理工学院商学院的管理科学教授。他出生在纽约州埃尔迈拉市，在克拉克森大学获得学士学位，在伦斯勒理工学院完成研究生学业并获得硕士和博士学位。

在进入罗切斯特理工学院商学院之前，威廉斯教授在辛辛那提大学工商管理学院从事了 7 年教学工作，他在那里制定了信息系统专业的本科教学计划，并且担任协调人。在罗切斯特理工学院，他是决策科学系的第一任主席。他不但讲授本科生的管理科学与统计学课程，而且还讲授研究生的回归与决策分析课程。

威廉斯教授在管理科学、统计学、生产与运作管理和数学领域与他人合作出版了 11 部专著。他为《财富》500 强中多家公司提供咨询服务，从数据分析的使用到大型回归模型的开发，都在他的工作范围之内。

## 杰弗里 D. 卡姆 (Jeffrey D. Camm)

杰弗里 D. 卡姆是辛辛那提大学数量分析教授，运营、商业分析与信息系统系主任，卡尔 H. 林德纳商学院商

业研究院研究员。他出生在俄亥俄州辛辛那提市，在泽维尔大学获得学士学位，在克莱姆森大学获得博士学位。从1984年起他在辛辛那提大学工作，是斯坦福大学的访问学者和达特茅斯学院商学院工商管理客座教授。

卡姆博士在运营管理中优化问题应用领域发表了30多篇论文。他的研究成果发表在《科学》《管理科学》《运筹学》、*Interfaces* 和其他专业杂志上。在辛辛那提大学工作期间，他被任命为教学优秀的 Dornoff 成员，并在2006年因运筹学实践教学获得了 INFORMS 奖。他是宣扬实践的坚定信徒。作为运筹学顾问，他服务于多家公司和政府机构。2005~2010年，他担任 *Interfaces* 总编辑，目前是 INFORMS 教育学报编委。

## 詹姆斯 J. 科克伦 (James J. Cochran)

詹姆斯 J. 科克伦是路易斯安那理工大学拉斯顿银行授衔的数量分析教授。他出生在俄亥俄州的代顿市，在莱特州立大学获得学士、硕士和工商管理硕士学位，在辛辛那提大学获得博士学位。从2000年起在路易斯安那理工大学工作，是斯坦福大学、塔尔卡大学和南非大学的访问学者。

科克伦教授在运筹学和统计方法的发展和应用方面发表了超过24篇论文。他的研究发表在《管理科学》《美国统计》《统计通讯——理论与方法》《运营研究（欧洲版）》，*Journal of Combinatorial Optimization* 和其他专业杂志上。在2008年他因运筹学实践教学获得了 INFORMS 奖，在2010年获 Mu Sigma Rho 统计教育奖。科克伦教授2005年入选国际统计学会，2010年成为美国统计协会会员。他以提高对实际问题应用质量的方法，强烈主张运筹学和统计学教育的有效性。科克伦教授在乌拉圭的蒙得维的亚、南非的开普敦、哥伦比亚的卡塔赫纳、印度的斋浦尔、阿根廷的布宜诺斯艾利斯、肯尼亚的内罗毕组织和主持教学效果研讨班。作为运筹学顾问，他服务于多家公司和非营利组织。目前他是 INFORMS 教育学报总编辑及 *Interfaces*, *The Journal of the Chilean of Operations Research* 及 *ORION* 的编委。

## 托马斯 A. 威廉斯 (Thomas A. Williams)

托马斯 A. 威廉斯是圣路易大学工商管理学院的教授。他出生在俄亥俄州的代顿市，在莱特州立大学获得学士、硕士和工商管理硕士学位，在辛辛那提大学获得博士学位。从2000年起在圣路易大学工作，是斯坦福大学、塔尔卡大学和南非大学的访问学者。威廉斯教授在运营管理中优化问题应用领域发表了30多篇论文。他的研究成果发表在《科学》《管理科学》《运筹学》、*Interfaces* 和其他专业杂志上。在辛辛那提大学工作期间，他被任命为教学优秀的 Dornoff 成员，并在2006年因运筹学实践教学获得了 INFORMS 奖。他是宣扬实践的坚定信徒。作为运筹学顾问，他服务于多家公司和政府机构。2005~2010年，他担任 *Interfaces* 总编辑，目前是 INFORMS 教育学报编委。

## 杰弗里 D. 坎宁 (Jeffrey D. Canning)

杰弗里 D. 坎宁是圣路易大学工商管理学院的教授。他出生在俄亥俄州的代顿市，在莱特州立大学获得学士、硕士和工商管理硕士学位，在辛辛那提大学获得博士学位。从2000年起在圣路易大学工作，是斯坦福大学、塔尔卡大学和南非大学的访问学者。



# 数理统计及相关课程

课程名称	书号	书名、作者及出版时间	版别	定价
统计学	978-7-111-33687-7	统计学(强森)(2011年)	外版	49
数据、模型与决策	978-7-111-38280-5	数据、模型与决策:管理科学篇(第13版)(安德森)(2012年)	外版	75
数据、模型与决策	978-7-111-49612-0	数据、模型与决策:基于电子表格的建模和案例研究方法(第5版)(希利尔)(2015年)	外版	89
数据、模型与决策	978-7-111-48099-0	数据、模型与决策:基于电子表格的建模和案例研究方法(英文版·第4版)	外版	85
时间序列分析	978-7-111-33864-2	时间序列分析:预测与控制(第4版)(博克斯)(2011年)	外版	59
时间序列分析	978-7-111-38801-2	应用计量经济学:时间序列分析(第3版)(恩德斯)(2012年)	外版	69
商务与经济统计	978-7-111-37641-5	商务与经济统计(第11版)(安德森)(2012年)	外版	108
商务与经济统计	即将出版	商务与经济统计(第12版)(安德森)(2015年)	外版	109
商务与经济统计	978-7-111-24366-3	商务与经济统计(第6版)(组博尔德)(2008年)	外版	90
商务与经济统计	978-7-111-35029-3	商务与经济统计(英文版·第11版)(安德森)(2011年)	外版	109
商务与经济统计	978-7-111-38666-7	商务与经济统计精要(第6版)(安德森)(2012年)	外版	59
商务与经济统计	978-7-111-40479-8	商务与经济统计精要(英文版·第6版)(安德森)(2012年)	外版	69
经济决策模型	978-7-111-26846-8	经济决策的概率模型(迈尔森)(2009年)	外版	48
计量经济学学习指导	978-7-111-31370-0	经济计量学精要(第4版)习题集(古扎拉蒂)(2010年)	外版	29
计量经济学	978-7-111-30817-1	经济计量学精要(第4版)(古扎拉蒂)(2010年)	外版	49
计量经济学	978-7-111-35537-3	应用计量经济学(第6版)(施图德蒙德)(2011年)	外版	59
运筹学	978-7-111-44029-1	运筹学(第3版)(精品课)(熊伟)(2014年)	本版	35
运筹学	978-7-111-44298-1	运筹学(李峰)(2014年)	本版	39
统计学学习指导	978-7-111-22168-5	应用统计学习指导(精品课)(孙炎)(2007年)	本版	19
统计学	978-7-111-48630-5	统计学(第2版)(张兆丰)(2014年)	本版	35
统计学	978-7-111-47889-8	统计学(第4版)(精品课)(李金昌)(“十二五”普通高等教育本科国家级规划教材)(2014年)	本版	39
统计学	978-7-111-21720-9	统计学(精品课)(郑珍远)(2007年)	本版	32
统计学	978-7-111-42075-0	统计学(卢小广)(2013年)	本版	35
统计学	978-7-111-42504-5	统计学(向蓉美)(2013年)	本版	39
统计学	978-7-111-31321-2	统计学(曾五一)(2010年)	本版	35
统计学	978-7-111-45966-8	统计学原理(宫春子)(2014年)	本版	35
统计学	978-7-111-29041-4	应用统计基础(精品课)(曹艳英)(2009年)	本版	38
统计学	978-7-111-47018-2	应用统计学(第2版)(精品课)(“十二五”职业教育国家规划教材)(孙炎)(2014年)	本版	35
统计学	978-7-111-44677-4	应用统计学(精品课)(谢忠秋)(2014年)	本版	35
统计、计量分析软件	978-7-111-20747-4	Eviews使用指南与案例(张晓峒)(2007年)	本版	35
数量经济学	978-7-111-26575-7	应用数量经济学(“十一五”国家级规划教材)(张晓峒)(2009年)	本版	45
数据、模型与决策	即将出版	数据、模型与决策(梁樑)(2015年)	本版	40
计量经济学	978-7-111-48459-2	计量经济学(第2版)(李宝仁)(2014年)	本版	39
计量经济学	978-7-111-42076-7	计量经济学基础(张兆丰)(2013年)	本版	35
计量经济学	即将出版	计量经济学及其应用(第2版)(杜江)(2015年)	本版	39
计量经济学	978-7-111-29842-7	计量经济学及其应用(杜江)(2010年)	本版	29
高级运筹学	978-7-111-24349-6	高级运筹学(马良)(2008年)	本版	30
概率论和数理统计	978-7-111-26974-8	应用概率统计(彭美云)(2009年)	本版	27
概率论和数理统计	978-7-111-28975-3	应用概率统计学习指导与习题选解(彭美云)(2009年)	本版	18



课程名称	书号	书名、作者及出版时间	版别	定价
预测理论与方法	978-7-111-38130-3	经济预测基础教程(第4版)(迪博尔德)(2012年)	外版	59
国际经济学	978-7-111-34033-1	国际经济学(第8版)(赫斯特德)(2011年)	外版	62
国际经济学	978-7-111-35280-8	国际经济学(英文版·第8版)(赫斯特德)(2011年)	外版	69
国际经济学	978-7-111-49021-0	国际经济学:国际金融分册(第8版)(阿普尔亚德)(2015年)	外版	39
国际经济学	978-7-111-47783-9	国际经济学:国际贸易分册(第8版)(阿普尔亚德)(2014年)	外版	59
管理经济学	978-7-111-48424-0	管理经济学(第11版)(莫瑞斯)(2014年)	外版	75
管理经济学	978-7-111-39670-3	管理经济学(第12版)(麦圭根)(2012年)	外版	89
管理经济学	978-7-111-41881-8	管理经济学(第12版)(中国版)(麦圭根)(2013年)	外版	55
管理经济学	978-7-111-31879-8	管理经济学(第9版)(中国版)(托马斯、刘延平)(2010年)	外版	48
管理经济学	978-7-111-26569-6	管理经济学(英文版·第11版)(麦圭根)(2009年)	外版	68
管理经济学	978-7-111-48567-4	管理经济学(英文版·第11版)(莫瑞斯)(2014年)	外版	89
发展经济学	978-7-111-45840-1	发展经济学(第11版)(托达罗)(2014年)	外版	89
财政学	即将出版	财政学(第4版)(格鲁伯)(2015年)	外版	79
经贸英语	978-7-111-40348-7	经贸英语阅读教程(第3版)(潘忠)(2012年)	本版	25
经贸英语	978-7-111-44666-8	商务英语阅读(沈素萍)(2013年)	本版	35
经贸英语	978-7-111-46218-7	商务英语阅读教程(杨伶俐)(2014年)	本版	29
经济法	978-7-111-47546-0	经济法(第2版)(葛恒云)(2014年)	本版	35
经济法	978-7-111-44659-0	经济法(第2版)(郭懿美)(2014年)	本版	39
经济法	978-7-111-41779-8	经济法(刘大洪)(2013年)	本版	39
经济法	978-7-111-13974-4	经济法基础与实务(黄瑞)(2008年)	本版	32
国际经济学	978-7-111-25578-9	国际经济学(精品课)(赵英军)(2009年)	本版	32
国际经济关系学	978-7-111-27371-4	国际经济关系学概论(周林)(2009年)	本版	30
管理经济学	978-7-111-39608-6	管理经济学(毛蕴诗)(“十二五”普通高等教育本科国家级规划教材)(2012年)	本版	45
产业经济学	978-7-111-49568-0	产业经济学(刘志彪)(2015年)	本版	39
财政学	978-7-111-27276-2	财政学(第2版)(李友元)(2009年)	本版	36
财经应用文写作	978-7-111-42715-5	财经应用文写作(刘常宝)(2013年)	本版	30
博弈论	978-7-111-30394-7	博弈论及其应用(李帮义)(2010年)	本版	25

# 华章教材经典译丛（清明上河图）系列



课程名称	书号	书名、作者及出版时间	定价
财务会计	即将出版	财务会计：概念、方法与应用（第14版）（威尔）（2015年）	95
财务会计	978-7-111-39244-6	财务会计教程（第10版）（亨格瑞）（2012年）	79
财务管理（公司理财）学习指导	978-7-111-32466-9	公司理财（第8版）习题集（沃森）（2010年）	42
财务管理（公司理财）	978-7-111-36751-2	公司理财（第9版）（罗斯）（2012年）	88
财务管理（公司理财）	978-7-111-47887-4	公司理财（精要版）（第10版）（罗斯）（2014年）	75
电子商务	978-7-111-45187-7	电子商务：管理与社会网络的视角（第7版）（特班）（2014年）	79
战略管理	978-7-111-39138-8	战略管理：概念与案例（第8版）（希尔）（2012年）	69
战略管理	978-7-111-43844-1	战略管理：获取持续的竞争优势（第4版）（巴尼）（2013年）	69
商业伦理学	978-7-111-37513-5	企业伦理学（第7版）（乔治）（2012年）	79
领导学	978-7-111-47356-5	领导学（全球版·第8版）（尤克尔）（2014年）	65
管理学	978-7-111-46255-2	管理学（诺里奇）（2014年）	69
管理学	978-7-111-41449-0	管理学：原理与实践（第8版）（罗宾斯）（2013年）	59
管理学	即将出版	管理学：原理与实践（第9版）（罗宾斯）（2015年）	59
管理技能	978-7-111-37591-3	管理技能开发（第8版）（基顿）（2012年）	98
创业管理	即将出版	百森创业教学法：基于实践的视角（奈克）（2015年）	49
创业管理	978-7-111-40258-9	公司创新与创业（第3版）（库拉特科）（2012年）	49
项目管理	978-7-111-39774-8	项目管理：基于团队的方法（布朗）（2012年）	19
数据、模型与决策	978-7-111-49612-0	数据、模型与决策：基于电子表格的建模和案例研究方法（第5版）（希利尔）（2015年）	89
管理会计	978-7-111-39512-6	管理会计教程（第15版）（亨格瑞）（2012年）	88
投资银行学	978-7-111-41476-6	投资银行：对冲基金和私募股权投资（斯托瓦克）（2013年）	99
金融中介学	978-7-111-43694-2	金融市场与金融机构（第7版）（米什金）（2013年）	99
金融学（货币银行学）指导或案例	978-7-111-44311-7	货币金融学（第2版）学习指导（米什金）（2013年）	15
金融学（货币银行学）	978-7-111-34261-8	货币金融学（第2版）（米什金）（2011年）	75
金融市场学	978-7-111-26674-7	金融市场学（第10版）（罗斯）（2009年）	79
金融工程学习指导	978-7-111-30014-4	期权、期货及其他衍生产品习题集（第7版）（赫尔）（2010年）	42
金融工程	978-7-111-48437-0	期权、期货及其他衍生产品（第9版）（赫尔）（2014年）	109
（证券）投资学学习指导	978-7-111-42662-2	投资学习题集（第9版）（博迪）（2013年）	49
（证券）投资学	978-7-111-39028-2	投资学（第9版）（博迪）（2012年）	98
（证券）投资学	978-7-111-44485-5	投资学（第9版）（专业版）（博迪）（2013年）	199
中级宏观经济学	978-7-111-43155-8	宏观经济学（第5版）（布兰查德）（2013年）	75
西方经济学学习指导	978-7-111-33099-8	哈伯德《经济学》学习指南（第3版）（斯卡希尔）（2011年）	45
西方经济学学习指导	978-7-111-31352-6	经济学精要（精要版）（第4版）学习指南（拉什）（2010年）	39
西方经济学（微观）	978-7-111-32767-7	经济学（微观）（第3版）（哈伯德）（2011年）	59
西方经济学（微观）	978-7-111-42810-7	经济学（微观部分）（第2版）（斯通）（2013年）	55
西方经济学（宏观）	978-7-111-32768-4	经济学（宏观）（第3版）（哈伯德）（2011年）	49
西方经济学（宏观）	978-7-111-42849-7	经济学（宏观部分）（第2版）（斯通）（2013年）	49
西方经济学	978-7-111-28688-0	经济学：私人与公共选择（第12版）（格瓦特尼）（2009年）	78
西方经济学	978-7-111-27481-0	经济学原理（精要版）（第4版）（帕金）（2009年）	62
商务与经济统计	978-7-111-37641-5	商务与经济统计（第11版）（安德森）（2012年）	108
商务与经济统计	即将出版	商务与经济统计（第12版）（安德森）（2015年）	109
财政学	即将出版	财政学（第4版）（格鲁伯）（2015年）	79
组织行为学	978-7-111-44814-3	组织行为学精要（第12版）（罗宾斯）（2014年）	45
人力资源管理	978-7-111-40189-6	人力资源管理（亚洲版·第2版）（德斯勒）（2012年）	65
消费者行为学	978-7-111-47509-5	消费者行为学（第12版）（霍金斯）（2014年）	79
市场营销学（营销管理）	978-7-111-43017-9	市场营销学（第11版）（阿姆斯特朗、科特勒）（2013年）	75
市场营销学（营销管理）	978-7-111-43202-9	市场营销原理（亚洲版·第3版）（科特勒）（2013年）	79
服务营销学	978-7-111-48495-0	服务营销（第6版）（泽丝曼尔）（2014年）	75
供应链（物流）管理	978-7-111-45565-3	供应链管理（第4版）（鲍尔索克斯）（2014年）	59
管理信息系统	978-7-111-34151-2	管理信息系统（第11版）（劳顿）（2011年）	55

## 译者序

## 前 言

## 作者简介

## 第1章 数据与统计资料 ..... 1

## 实践中的统计: 彭博商业周刊 ..... 2

## 1.1 统计学在商务和经济中的应用 ..... 3

## 1.2 数据 ..... 4

## 1.3 数据来源 ..... 8

## 1.4 描述统计 ..... 11

## 1.5 统计推断 ..... 12

## 1.6 计算机与统计分析 ..... 13

## 1.7 数据挖掘 ..... 13

## 1.8 统计实践的道德准则 ..... 14

## 小结 ..... 15

## 关键术语 ..... 15

## 补充练习 ..... 16

## 第2章 描述统计学 I: 表格法和图形法 ..... 19

## 实践中的统计: 高露洁-棕榄公司 ..... 20

## 2.1 汇总分类变量的数据 ..... 21

## 2.2 数量型数据汇总 ..... 24

## 2.3 用表格方式汇总两个变量的数据 ..... 32

## 2.4 用图形显示方式汇总两个变量的数据 ..... 37

## 2.5 数据可视化: 创建有效图形显示的

## 最佳实践 ..... 39

## 小结 ..... 44

## 关键术语 ..... 45

## 重要公式 ..... 45

## 补充练习 ..... 45

## 案例 2-1 Pelican 商店 ..... 47

## 案例 2-2 电影业 ..... 48

## 第3章 描述统计学 II: 数值方法 ..... 49

## 实践中的统计: Small Fry 设计公司 ..... 50

## 3.1 位置的度量 ..... 50

## 3.2 变异程度的度量 ..... 59

3.3 分布形态、相对位置的度量以及异常值的  
检测 ..... 63

## 3.4 五数概括法和箱形图 ..... 67

## 3.5 两变量间关系的度量 ..... 71

3.6 数据仪表盘: 增加数值度量以提高  
有效性 ..... 75

## 小结 ..... 78

## 关键术语 ..... 78

## 重要公式 ..... 79

## 补充练习 ..... 79

## 案例 3-1 Pelican 商店 ..... 81

## 案例 3-2 电影业 ..... 82

## 案例 3-3 亚太地区的商学院 ..... 83

## 案例 3-4 天使巧克力的网络交易 ..... 84

## 案例 3-5 非洲象数量 ..... 84

## 第4章 概率 ..... 86

## 实践中的统计: 概率在救援中的应用 ..... 87

## 4.1 试验、计数法则和概率分配 ..... 88

## 4.2 事件及其概率 ..... 94

## 4.3 概率的基本性质 ..... 96

## 4.4 条件概率 ..... 99

## 4.5 贝叶斯定理 ..... 103

## 小结 ..... 107

## 关键术语 ..... 107

## 重要公式 ..... 108

## 补充练习 ..... 108

## 案例 Hamilton 县的法官们 ..... 110

## 第5章 离散型概率分布 ..... 112

## 实践中的统计: 花旗银行 ..... 113

## 5.1 随机变量 ..... 113

## 5.2 离散型概率分布 ..... 115

## 5.3 数学期望与方差 ..... 117

## 5.4 二元分布、协方差和金融资产组合 ..... 119



5.5 二项概率分布 .....	125	重要公式 .....	195
5.6 泊松概率分布 .....	131	补充练习 .....	195
5.7 超几何概率分布 .....	134	案例 8-1 YOUNG PROFESSIONAL 杂志 .....	196
小结 .....	136	案例 8-2 Gulf Real Estate Properties 公司 .....	197
关键术语 .....	137	案例 8-3 Metropolitan Research 有限公司 .....	198
重要公式 .....	137	<b>第 9 章 假设检验</b> .....	200
补充练习 .....	138	实践中的统计: John Morrell 有限公司 .....	201
<b>第 6 章 连续型概率分布</b> .....	140	9.1 原假设和备择假设的建立 .....	201
实践中的统计: 宝洁公司 .....	141	9.2 第一类错误和第二类错误 .....	204
6.1 均匀概率分布 .....	141	9.3 总体均值的检验: $\sigma$ 已知情形 .....	205
6.2 正态概率分布 .....	144	9.4 总体均值的检验: $\sigma$ 未知情形 .....	214
6.3 二项概率的正态近似 .....	150	9.5 总体比率 .....	217
6.4 指数概率分布 .....	151	9.6 假设检验与决策 .....	220
小结 .....	154	9.7 计算第二类错误的概率 .....	220
关键术语 .....	154	9.8 对总体均值进行假设检验时样本 容量的确定 .....	222
重要公式 .....	154	小结 .....	224
补充练习 .....	154	关键术语 .....	225
案例 Specialty 玩具公司 .....	155	重要公式 .....	225
<b>第 7 章 抽样和抽样分布</b> .....	157	补充练习 .....	226
实践中的统计: MeadWestvaco 有限公司 .....	158	案例 9-1 Quality Associates 有限公司 .....	227
7.1 EAI 的抽样问题 .....	159	案例 9-2 Bayview 大学商科学生的道德行为 .....	228
7.2 抽样 .....	159	<b>第 10 章 两总体均值和比例的推断</b> .....	230
7.3 点估计 .....	163	实践中的统计: 美国食品与药物管理局 .....	231
7.4 抽样分布简介 .....	165	10.1 两总体均值之差的推断: $\sigma_1$ 和 $\sigma_2$ 已知 .....	231
7.5 $\bar{x}$ 的抽样分布 .....	166	10.2 两总体均值之差的推断: $\sigma_1$ 和 $\sigma_2$ 未知 .....	235
7.6 $\bar{p}$ 的抽样分布 .....	171	10.3 两总体均值之差的推断: 匹配样本 .....	240
7.7 点估计的性质 .....	174	10.4 两总体比例之差的推断 .....	243
7.8 其他抽样方法 .....	175	小结 .....	247
小结 .....	177	关键术语 .....	247
关键术语 .....	177	重要公式 .....	247
重要公式 .....	178	补充练习 .....	248
补充练习 .....	178	案例 10-1 Par 公司 .....	249
<b>第 8 章 区间估计</b> .....	180	<b>第 11 章 总体方差的统计推断</b> .....	250
实践中的统计: Food Lion .....	181	实践中的统计: 美国审计总署 .....	251
8.1 总体均值的区间估计: $\sigma$ 已知情形 .....	181	11.1 一个总体方差的统计推断 .....	251
8.2 总体均值的区间估计: $\sigma$ 未知情形 .....	184	11.2 两个总体方差的统计推断 .....	257
8.3 样本容量的确定 .....	190	小结 .....	261
8.4 总体比率 .....	191		
小结 .....	194		
关键术语 .....	194		

重要公式 .....	261	关键词 .....	347
补充练习 .....	261	重要公式 .....	348
案例 空军训练计划 .....	262	补充练习 .....	348
<b>第 12 章 多个比率的比较、独立性及拟合</b> .....		案例 14-1 测量股票市场风险 .....	351
<b>优度检验</b> .....	263	案例 14-2 美国交通部 .....	351
实践中的统计: 联合劝募协会 .....	264	案例 14-3 挑选一台傻瓜型 (point-and-shoot)	
12.1 三个或多个总体比率的相等性的检验 .....	264	数码相机 .....	352
12.2 独立性检验 .....	271	案例 14-4 找到最合适的汽车价值 .....	353
12.3 拟合优度检验 .....	275	<b>第 15 章 多元回归</b> .....	355
小结 .....	280	实践中的统计: dunnhumby .....	356
关键词 .....	281	15.1 多元回归模型 .....	356
重要公式 .....	281	15.2 最小二乘法 .....	358
补充练习 .....	281	15.3 多元判定系数 .....	362
案例 两党议程变更 .....	282	15.4 模型的假定 .....	364
<b>第 13 章 实验设计与方差分析</b> .....	283	15.5 显著性检验 .....	365
实践中的统计: Burke 市场营销服务公司 .....	284	15.6 应用估计的回归方程进行估计和预测 .....	369
13.1 实验设计和方差分析简介 .....	285	15.7 分类自变量 .....	370
13.2 方差分析和完全随机化实验设计 .....	288	15.8 残差分析 .....	374
13.3 多重比较方法 .....	294	15.9 logistic 回归 .....	378
13.4 随机化区组设计 .....	299	小结 .....	386
13.5 析因实验 .....	302	关键词 .....	386
小结 .....	306	重要公式 .....	387
关键词 .....	307	补充练习 .....	388
重要公式 .....	307	案例 15-1 消费者调查股份有限公司 .....	389
补充练习 .....	308	案例 15-2 预测 NASCAR 车手的奖金 .....	390
案例 13-1 Wentworth 医疗中心 .....	309	案例 15-3 找到最合适的汽车价值 .....	391
案例 13-2 销售人员的报酬 .....	310	<b>第 16 章 回归分析: 建立模型</b> .....	392
<b>第 14 章 简单线性回归</b> .....	312	实践中的统计: Monsanto 公司 .....	393
实践中的统计: 联盟数据系统 (ALLIANCE DATA SYSTEMS) .....	313	16.1 一般线性模型 .....	393
14.1 简单线性回归模型 .....	314	16.2 确定什么时候增加或者删除变量 .....	402
14.2 最小二乘法 .....	316	16.3 大型问题的分析 .....	406
14.3 判定系数 .....	321	16.4 变量选择方法 .....	409
14.4 模型的假定 .....	325	16.5 实验设计的多元回归方法 .....	413
14.5 显著性检验 .....	326	16.6 自相关性和杜宾-瓦特森检验 .....	415
14.6 应用估计的回归方程进行估计和预测 .....	332	小结 .....	418
14.7 计算机解法 .....	336	关键词 .....	418
14.8 残差分析: 证实模型假定 .....	338	重要公式 .....	418
14.9 残差分析: 异常值和有影响的观测值 .....	342	补充练习 .....	419
小结 .....	347	案例 16-1 职业高尔夫球协会巡回赛的统计	
		分析 .....	419

案例 16-2 产自意大利皮埃蒙特地区的葡萄酒 评级 .....	420
<b>第 17 章 时间序列分析及预测</b> .....	422
实践中的统计：内华达职业健康诊所 .....	423
17.1 时间序列的模式 .....	424
17.2 预测精度 .....	428
17.3 移动平均法和指数平滑法 .....	431
17.4 趋势推测法 .....	437
17.5 季节性和趋势 .....	445
17.6 时间序列分解法 .....	449
小结 .....	456
关键术语 .....	457
重要公式 .....	457
补充练习 .....	458
案例 17-1 预测食品和饮料的销售 .....	459
案例 17-2 预测损失的销售额 .....	460
<b>第 18 章 非参数方法</b> .....	461
实践中的统计：West Shell Realtors 公司 .....	462
18.1 符号检验 .....	462
18.2 威尔科克森符号秩检验 .....	467
18.3 曼-惠特尼-威尔科克森检验 .....	470
18.4 克鲁斯卡尔-沃利斯检验 .....	475
18.5 秩相关 .....	477
小结 .....	480
关键术语 .....	480
重要公式 .....	480
补充练习 .....	481

<b>第 19 章 质量管理的统计方法</b> .....	482
实践中的统计：陶氏化学公司 .....	483
19.1 理念和框架 .....	484
19.2 统计过程控制 .....	486
19.3 接受抽样 .....	495
小结 .....	501
关键术语 .....	501
重要公式 .....	502
补充练习 .....	502
<b>第 20 章 指数</b> .....	503
实践中的统计：美国劳工部，劳工统计局 .....	504
20.1 价比 .....	504
20.2 综合物价指数 .....	505
20.3 根据价比计算综合物价指数 .....	507
20.4 一些重要的价格指数 .....	508
20.5 根据物价指数减缩一个数列 .....	510
20.6 物价指数：其他注意事项 .....	511
20.7 物量指数 .....	512
小结 .....	513
关键术语 .....	514
重要公式 .....	514
补充练习 .....	514
<b>附录 A 参考文献</b> .....	
<b>附录 B 统计表格</b> .....	
<b>附录 C 自测题解答与偶数题答案</b> .....	
<b>附录 D Microsoft Excel 2010 和统计分析工具</b> .....	
<b>附录 E 利用 Minitab 和 Excel 计算 <math>p</math>-值</b> .....	

# 第 1 章

## 数据与统计资料

### CHAPTER

# 1

实践中的统计：彭博商业周刊

1.1 统计学在商务和经济中的应用

1.2 数据

1.3 数据来源

1.4 描述统计

1.5 统计推断

1.6 计算机与统计分析

1.7 数据挖掘

1.8 统计实践的道德准则



## 实践中的统计

彭博商业周刊<sup>②</sup>

纽约州, 纽约市

《彭博商业周刊》(Bloomberg Businessweek) 是世界上拥有最广泛读者的商业类杂志, 其全球发行量超过 100 万份。该杂志在全球的 145 个记者站中有 1 700 多名忠于职守的记者和编辑, 主要登载各类有关全球经济和贸易活动的饶有兴趣的文章。除了刊登当前热门话题的特色文章外, 还有一些关于国际贸易、经济分析、信息处理和科技进展等常设栏目。特色文章和常设栏目传递的信息能帮助读者及时透视时事风云变幻, 以便评估其对商务和经济状况产生的影响。

几乎每一期《彭博商业周刊》(更名前叫《商业周刊》) 都发表一篇有关当前热门话题的深度报告, 通常这些深度报告包含统计论据和概述, 以帮助读者了解商务和经济信息。例如, 2011 年 3 月 3 日那一期的封面故事, 讨论了工商业逐渐把其主要业务处理向云计算转移; 2011 年 5 月 30 日那一期, 有一篇濒临崩溃的美国邮政服务业的报告; 2011 年 8 月 1 日那一期, 发表了一篇为什么债务危机比想象的更为严重的文章。另外, 每一期《彭博商业周刊》都发布关于经济状况的统计资料, 包括产品指数、股票价格、共同基金和

利率等。

《彭博商业周刊》还利用统计报告和统计信息来帮助管理其自身的经营。例如, 一份关于订阅者的年度综述报告可以帮助公司了解订阅者的人口特征、阅读习惯、购物偏好和生活方式等。《彭博商业周刊》的管理人员利用从调查中得出的统计结论更好地为订阅者和广告客户提供服务。最近对《彭博商业周刊》北美订阅者的调查数据表明, 90% 的《彭博商业周刊》订阅者在家使用计算机, 并且 64% 的订阅者因工作需要而购买了计算机。这一统计结果提醒《彭博商业周刊》的管理者: 订阅者会对个人计算机发展趋势方面的文章感兴趣, 而且, 调查的结论也会引起潜在的广告客户的兴趣。订阅者中在家使用计算机和因工作使用计算机有如此之高的百分比, 将会刺激个人计算机厂商考虑在《彭博商业周刊》上做广告。

本章我们讨论可用于统计分析的数据类型, 并说明如何取得这些数据。我们将介绍描述统计和统计推断, 它们是将数据转为有意义的且易于解释的统计信息的方法。

我们经常在报纸和杂志上的文章中看到下列各种报道:

- 美国劳工部报道失业率降至过去 3 年的最低点 8.2% (《华盛顿邮报》, 2012 年 4 月 6 日)。
- 每年每一位美国人平均消费 23.2 夸脱的冰淇淋、牛奶冻、果汁牛奶冻、冰冻甜食和其他大量生产的冷冻乳制品 (makeicecream.com website, 2012 年 4 月 2 日)。
- 度假屋销售价格的中位数为 121 300 美元 (@CNNMoney, 2012 年 3 月 29 日)。
- 位于田纳西州 Pigeon Forge 的多来坞主题公园的野鹰宽翼过山车最大时速为每小时 98 千米 (USA Today website, 2012 年 4 月 5 日)。
- 钉板方式的社交图片分享网站 Pinterest 的注册用户数量从 1 月中旬到 2 月中旬增加了 85% (CNBC, 2012 年 3 月 29 日)。
- Pew 研究中心报道美国新娘初婚年龄的中位数创纪录高达 26.5 岁 (Significance, 2012 年 2 月)。
- 2011 年第 4 季度加拿大人上网时间平均达到 45 小时 (CBC News, 2012 年 3 月 2 日)。
- 美联储报道每个人的平均信用卡债务为 5 204 美元 (PRWeb website, 2012 年 4 月 5 日)。

前面报道中的数据事实 (8.2%, 23.2, 121 300 美元, 61, 85%, 26.5, 45, 5 204 美元) 称为统计资料 (statistics)。在这种用法中, 术语统计资料是指数据值事实, 如平均数、中位数、百分数和最大值等, 它们可以帮助我们了解各种商务和经济状况。然而, 你将看到, 统计学的研究领域或对象的内涵更加丰富。广义上讲, 统

② 作者感谢为“实践中的统计”提供了案例的调研管理人员 Charlene Trentham。

计学是搜集、分析、表述和解释数据的科学与艺术。特别是在商务和经济活动中,搜集、分析、表述和解释数据旨在帮助管理者和决策者更好地理解商务和经济环境的变化,并作出科学、正确的决策。在本书中,我们强调的是统计学在商务和经济决策方面的用途。

本章以统计学在商务和经济中的一些应用实例为开端,在第1.2节我们定义了数据这一术语的含义,并给出了数据集的概念。在这一节中,还介绍了变量和观测值等一些重要的术语,讨论了数量型数据和分类型数据之间的区别,并且说明了截面数据和时间序列数据的应用。在第1.3节中,我们探讨了如何从现有的资料来源中获得数据,或通过调查和设计的实验研究获得新数据。对如今在数据获取中起着重要作用的互联网,也给予了高度的重视。在第1.4节和第1.5节中,论述了数据在进行描述统计和统计推断中的用途。在本章的最后3节还给出了计算机在统计分析中的角色,简单介绍了数据挖掘,并且讨论了统计实践的道德准则。本章末尾的附录中对 Excel 的插件 StatTools 做了介绍,它扩展了 Microsoft Excel 用户的统计选项功能。

## 1.1 统计学在商务和经济中的应用

如今在贸易和经济全球化的背景下,人们可以获取大量的统计信息。最成功的管理者和决策者是那些能够理解信息并有效利用信息的人。本节我们举例说明统计学在商务和经济中的应用。

### 1.1.1 会计

会计师事务所在对其客户进行审计时需要使用统计抽样程序。例如,假设一个事务所想确定列示在客户资产负债表上的应收账款金额是否真实地反映了应收账款的实际金额。通常应收账款的数量是如此之大,以致查看和验证每一账户将花费大量的时间和费用。在这种情况下,一般的做法是:审计人员从账户中选择一个子集作为样本,在查看样本账户的准确性后,审计师得出有关列示在客户资产负债表上的应收账款金额是不是可以接受的结论。

### 1.1.2 财务

财务顾问们利用各种各样的统计信息指导投资。在股票市场中,财务顾问们综合了包括市盈率和股息等方面的财务数据,通过对比单只股票和股票市场平均状况的信息,就可以得出某一只股票是否具有投资价格的结论。例如,《华尔街日报》(2012年3月19日)报道说,标准普尔500公司的平均股息收益率是2.2%,微软的股息收益是2.42%。这种情况下,关于股息收益率的统计信息表明,微软的股息收益率比标准普尔500公司的平均收益率高。这一信息和其他有关微软的信息将帮助财务顾问作出是买、是卖还是持股的建议。

### 1.1.3 市场营销

零售结账柜台的电子扫描仪正用于搜集各种市场调研用的数据。例如,数据供应商 ACNielsen 公司和 Information Resources 公司,从商店购买 POS 扫描数据,经过加工处理,作出统计汇总后再出售给制造商。制造商为取得这些扫描数据,每一类产品要花费数十万美元。制造商也购买特价销售和利用店内陈列品等促销活动的数据及统计研究报告。产品品牌经理可以查看扫描资料和促销活动统计资料,从而能更好地理解促销活动和销售额之间的关系。这样的分析对制定各种产品未来的市场营销战略大有裨益。

### 1.1.4 生产

由于当今进入了重视质量的年代,因此质量管理是统计学在生产中的一项重要应用。各种统计质量管理图用于监测生产过程的产出。特别地,用  $\bar{x}$  控制图可以监测平均产出。例如,假定有一台 340 克的软饮料罐装机,定期地从产品中选择一些罐装饮料作为样本,计算出样本罐装量的平均值。这一平均值或  $\bar{x}$  的值标在一张  $\bar{x}$  控制图上。当该数值位于控制上限以上时,则表明产品罐装量过高;当该数值位于控制下限以下时,则表明产品罐装量不足;当  $\bar{x}$  的值位于控制图的控制上限和控制下限之间时,这一过程称为处于“受控”状态,并且只要  $\bar{x}$  的值落

在控制图的控制上限和控制下限之间,就允许连续生产。更贴切的解释是, $\bar{x}$ 控制图能帮助确定何时必须调整和修正生产过程。

### 1.1.5 经济

人们经常要求经济学家对未来的经济或某一方面的发展作出预测。他们在进行预测时需要用到各种统计信息。例如,在预测通货膨胀率时,经济学家利用诸如生产者价格指数、失业率和制造业开工率等指标的统计信息。这些统计指标往往要输入到预测通货膨胀率的计算机预测模型中。

本节介绍的统计学的这些应用是本书的组成部分。这些例子可使我们对统计学的广泛应用有一个大致的了解。为了补充这些例子,我们请求商务和经济领域的专家提供每章开篇的“实践中的统计”,以便引出每章所要讲解的内容。“实践中的统计”充分显示统计学在各种各样的商务和经济问题中的重要性。

### 1.1.6 信息系统

信息系统管理员负责维护某个组织的计算机网络日复一日地运行。大量的统计信息帮助管理员评价计算机网络[包括局域网(LANs)、广域网(WANs)、网络段、内部网和其他数据通信信息]的表现。统计诸如系统中的平均用户数、系统中组件下载次数的比例和日常宽带使用的比例都是统计信息的例子。这些统计信息可以帮助系统管理员更好地了解和管理计算机网络。

## 1.2 数据

数据(data)是为了描述和解释所搜集、分析、汇总的事实和数字。将用于特定研究而搜集的所有数据称为研究的数据集(data set)。表1-1是一个包含参加世界贸易组织的60个国家信息的数据集。世界贸易组织支持国际贸易自由化,并提供了解决贸易争端的场所。

### 1.2.1 个体、变量和观测值

个体(element)是指搜集数据的实体。在表1-1的数据集中,每一个国家是一个个体,国家或个体的名称列在表1-1中的第1列。有60个国家,数据集中就有60个个体。

变量(variable)是个体中所感兴趣的那些特征。表1-1的数据集中有下列5个变量:

- WTO身份:在世界贸易组织中国家的成员身份,它可以是成员或观察员。
- 人均GDP(美元):国家的总产出除以该国的总人口数,它通常用于比较国家的经济生产率。
- 贸易逆差(1000美元):国家的进口总额与出口总额之差。
- 惠誉评级:由惠誉国际组织评价的国家主权信用评级<sup>①</sup>;信用评级从最高的AAA到最低的F,并且可以用+或-来微调。
- 惠誉评级展望:未来2年内信用评级可能变动的方向性指标,展望可以是正面、稳定或负面。

表1-1 世界贸易组织60个国家的数据集

国家	WTO 身份	人均 GDP (美元)	贸易逆差 (1000 美元)	惠誉评级	惠誉评级展望
亚美尼亚	成员	5 400	2 673 359	BB -	稳定
澳大利亚	成员	40 800	-33 304 157	AAA	稳定
奥地利	成员	41 700	12 796 558	AAA	稳定
阿塞拜疆	观察员	5 400	-16 747 320	BBB -	正面
巴林	成员	27 300	3 102 665	BBB	稳定
比利时	成员	37 600	-14 930 833	AA +	负面

① 惠誉国际组织是国际公认的三大统计评价机构之一,由美国证券交易监督委员会设计。另外两大机构是标准普尔和穆迪投资者服务公司。

(续)

国家	WTO 身份	人均 GDP (美元)	贸易逆差 (1000 美元)	惠誉评级	惠誉评级展望
巴西	成员	11 600	-29 796 166	BBB	稳定
保加利亚	成员	13 500	4 049 237	BBB -	正面
加拿大	成员	40 300	-1611 380	AAA	稳定
佛得角	成员	4 000	874 459	B +	稳定
智利	成员	16 100	-14 558 218	A +	稳定
中国	成员	8 400	-156 705 311	A +	稳定
哥伦比亚	成员	10 100	-1 561 199	BBB -	稳定
哥斯达黎加	成员	11 500	5 807 509	BB +	稳定
克罗地亚	成员	18 300	8 108 103	BBB -	负面
塞浦路斯	成员	29 100	6 623 337	BBB	负面
捷克共和国	成员	25 900	-10 749 467	A +	正面
丹麦	成员	40 200	-15 057 343	AAA	稳定
厄瓜多尔	成员	8 300	1 993 819	B -	稳定
埃及	成员	6 500	28 486 933	BB	负面
萨尔瓦多	成员	7 600	5 019 363	BB	稳定
爱沙尼亚	成员	20 200	802 234	A +	稳定
法国	成员	35 000	118 841 542	AAA	稳定
格鲁吉亚	成员	5 400	4 398 153	B +	正面
德国	成员	37 900	-213 367 685	AAA	稳定
匈牙利	成员	19 600	-9 421 301	BBB -	负面
冰岛	成员	38 000	-504 939	BB +	稳定
爱尔兰	成员	39 500	-59 093 323	BBB +	负面
以色列	成员	31 000	6 722 291	A	稳定
意大利	成员	30 100	33 568 668	A +	负面
日本	成员	34 300	31 675 424	AA	负面
哈萨克斯坦	观察员	13 000	-33 220 437	BBB	正面
肯尼亚	成员	1 700	9 174 198	B +	稳定
拉脱维亚	成员	15 400	2 448 053	BBB -	正面
黎巴嫩	观察员	15 600	13 715 550	B	稳定
立陶宛	成员	18 700	3 359 641	BBB	正面
马来西亚	成员	15 600	-39 420 064	A -	稳定
墨西哥	成员	15 100	1 288 122	BBB	稳定
秘鲁	成员	10 000	-7 888 993	BBB	稳定
菲律宾	成员	4 100	15 667 209	BB +	稳定
波兰	成员	20 100	19 552 976	A -	稳定
葡萄牙	成员	23 200	21 060 508	BBB -	负面
韩国	成员	31 700	-37 509 141	A +	稳定
罗马尼亚	成员	12 300	13 323 709	BBB -	稳定
俄罗斯	观察员	16 700	-151 400 000	BBB	正面
卢旺达	成员	1 300	939 222	B	稳定
塞尔维亚	观察员	10 700	8 275 693	BB -	稳定
塞舌尔	观察员	24 700	666 026	B	稳定
新加坡	成员	59 900	-27 110 421	AAA	稳定
斯洛伐克	成员	23 400	-2 110 626	A +	稳定
斯洛文尼亚	成员	29 100	2 310 617	AA -	负面
南非	成员	11 000	3 321 801	BBB +	稳定
瑞典	成员	40 600	-10 903 251	AAA	稳定
瑞士	成员	43 400	-27 197 873	AAA	稳定
泰国	成员	9 700	2 049 669	BBB	稳定
土耳其	成员	14 600	71 612 947	BB +	正面
英国	成员	35 900	162 316 831	AAA	负面
乌拉圭	成员	15 400	2 662 628	BB	正面
美国	成员	48 100	784 438 559	AAA	稳定
赞比亚	成员	1 600	-1 805 198	B +	稳定



在一项研究中,对每个个体的每一变量收集测量值,从而得到了数据。对某一特定个体得到的测量值集合称为一个观测值(observation)。如表1-1所示,我们看到第一个观测值(亚美尼亚)的测量值集合是成员、5 400、2 673 359、BB-和稳定。第二个观测值(澳大利亚)的测量值集合是成员、40 800、-33 304 157、AAA和稳定。60个个体的数据集有60个观测值。

### 1.2.2 测量尺度

搜集数据需要按下列几种测量尺度之一来度量:名义尺度、顺序尺度、间隔尺度和比率尺度。测量尺度决定了数据中蕴涵的信息量,并表明最适合的数据汇总和统计分析方法。

当一个变量的数据包含了用来识别个体属性的标记或名称时,测量尺度称为**名义尺度**(nominal scale)。例如,表1-1中的数据,我们看到WTO身份变量的测量尺度是名义尺度,因为数据“成员”和“观测员”是用来识别国家身份类型的标记。在测量尺度是名义尺度的情形中,可以使用数值代码及非数字的标记。例如,为了便于数据搜集,并做好将数据录入计算机数据库的准备,对WTO身份变量我们可以使用数值代码:用1代表世界贸易组织中的成员国;用2代表观测员国。即使数据是数值型的,测量尺度仍是名义尺度的。

如果数据具有名义数据的性质,并且数据的顺序或等级的意义明确,那么这种变量的测量尺度是**顺序尺度**(ordinal scale)。如表1-1中所示,惠誉评级的测量尺度是顺序尺度,因为从AAA到F的评级分类可以按从最佳的信用评级AAA到最差的信用评级F排序。评级字母提供了类似名义数据的标记,但是数据还可以根据信用评级排序,这使得测量尺度是顺序的。顺序尺度也可以用数值代码,例如你在学校中的班级排名。

如果数据具有顺序数据的所有性质,并且可以按某一固定度量单位表示数值间的间隔,则这种变量的测量尺度是**间隔尺度**(interval scale)。间隔数据永远是数值型的。学生能力测验(SAT)的分数是间隔尺度数据的一个例子。例如,3个学生的SAT分数分别为620、550和470,它们能够按最好到最差进行排序。另外,分数之差是有一定意义的。比如,学生1的分数比学生2的分数多70分( $=620-550$ ),学生2的分数比学生3的分数多80分( $=550-470$ )。

如果数据具有间隔数据的所有性质,并且两个数值之比是有意义的,则这种变量的测量尺度是**比率尺度**(ratio scale)。像距离、高度、重量和时间等变量都用比率尺度来度量。比率尺度需要有一个零值,变量取零值时表示什么也不存在。例如,让我们来考虑汽车的成本,零值意味着汽车没有成本或是免费的。另外,如果我们比较成本为30 000美元的汽车与成本为15 000美元的汽车,比率值 $30\,000/15\,000=2$ ,表示第一辆汽车的成本是第二辆汽车成本的2倍。

### 1.2.3 分类型数据和数量型数据

数据还可以进一步划分为分类型和数值型。归属于某一类别的数据称为**分类型数据**(categorical data),分类型数据既可以用名义尺度度量也可以用顺序尺度度量。用于表示大小或多少的数值称为**数量型数据**(quantitative data)。数值型数据既可以用间隔尺度度量也可以用比率尺度度量。

**分类变量**(categorical variable)是用分类型数据表示的变量,**数量变量**(quantitative variable)是用数量型数据表示的变量。统计分析方法是否适合一个特定变量,取决于变量是分类变量还是数量变量<sup>①</sup>。如果变量是分类变量,则统计分析方法极其有限。我们通过记录每一类别中观测值的数目,或计算每一类别中观测值的比例来汇总分类型数据。但是,即使分类型数据用数值代码表示,对其进行加、减、乘和除等数学运算也是没有意义的。第2.1节将介绍汇总分类型数据的方法。

另一方面,对数量变量进行数学运算,可以得到有意义的结果。例如,对于数量变量,可以先求和,然后除以观测值的个数,可以计算出平均数。这个平均数通常是有意义的并且易于解释。一般地,当数据是数值型时,有更多的统计方法可供选择。第2.2节和第3章会介绍汇总数量型数据的方法。

① 适合汇总数据的统计方法取决于数据是分类型数据还是数值型数据。

### 1.2.4 截面数据和时间序列数据

为了便于统计分析,有必要对截面数据和时间序列数据进行区分。**截面数据**(cross-sectional data)是在相同或近似相同的同一时点上搜集的数据。表1-1中数据是截面数据,因为它们描述了60个世界贸易组织国家的5个变量在同一时点上的情况。**时间序列数据**(time series data)是在几个时期内搜集的数据。例如,图1-1的时间序列是2007~2012年美国常规普通汽油每加仑的平均价格曲线。我们注意到,汽油价格在2008年夏季达到顶点后,2008年下半年汽油价格急剧下降。从2008年后,每加仑汽油平均价格重新开始稳步攀升,到2012年再次接近历史新高。



图1-1 美国常规普通汽油每加仑的平均价格

资料来源: Energy Information Administration, U. S. Department of Energy (美国能源部能源情报署), 2012年3月。

在商务与经济出版物中经常可以看到时间序列数据的图形,这些图形可以帮助分析家们了解过去发生的情况,确定现象随时间变动的趋势,推测时间序列的未来水平。时间序列数据的图形有各种形态,如图1-2所示。稍加研究,这些图形通常易于理解和解释。例如,图1-2a是2002~2012年道琼斯工业平均价格指数的图形。在2002年4月普通股票市场价格接近10 000点,在随后的5年中股指不断攀升,到2007年10月超过14 000点。然而,在2007年创历史新高后,股指急速下降。到2009年3月,糟糕的经济条件迫使道琼斯工业平均价格指数重新回到了7 000点的水平。对投资者来说,这是一个令人惊慌、失去信心的时期。然而,到2009年年末,股指有所复苏,达到10 000点。到2012年年初,股指稳步攀升,已超过13 000点。

图1-2b是2005~2011年麦当劳公司的净收益图形,2008年和2009年的经济衰退,实际上对麦当劳公司而言是受益的,期间其净收益创了历史新高。麦当劳公司的净收益的增长表明公司的繁荣是由于在经济衰落期人们减少了去相对昂贵的餐厅就餐,转而寻求麦当劳提供更方便的替代食品。在2010~2011年,麦当劳公司的净收益继续创历史新高。

图1-2c是一年期间佛罗里达州南部旅店入住率的一个时间序列图形。在2~3月,当佛罗里达州南部的气候对游客有吸引力时,出现95%~98%的极高入住率。事实上,每年的1~4月是佛罗里达州南部旅店典型高入住率的季节。另外,8~10月入住率较低,最低的50%入住率出现在9月,高温以及飓风季节是这段时间入住率下降的主要原因。

#### 注释和评论

1. 观测值是数据集中每个个体的测量值的集合。因此,观测值的个数总是与个体的个数一致。每个个体的测量值的个数等于变量的个数。因此,数据项的总数是个体的个数乘以变量的个数。
2. 数量型数据可以是离散的也可以是连续的。度量可数事物多少的数量型数据是离散的(例如,5分钟内接到电话的个数),度量不可数事物的多少的数量型数据是连续的(例如,体重或时间)。

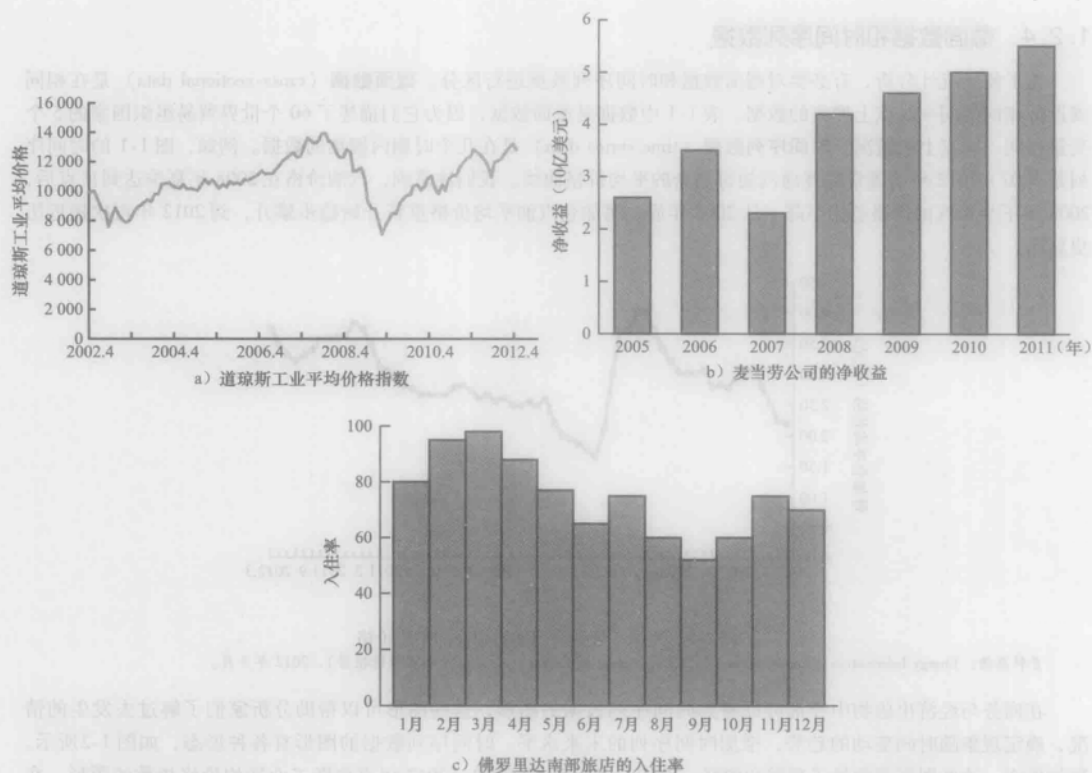


图 1-2 时间序列数据的各种图形

### 1.3 数据来源

数据可以从现有来源中搜集,或通过调查获得,亦可通过为了得到新数据而设计的实验性研究获得。

#### 1.3.1 现有来源

在某些情况下,为某一特定应用所需要的数据已经存在。公司保持有关其雇员、客户和义务运作等各种数据库。关于雇员工薪、年龄和服务年限的数据通常可以从内部个人记录中获得;关于销售收入、广告支出、配送成本、库存水平以及产量的数据通常可以从其他内部记录中获得;绝大多数公司还保存其客户的详细资料。表 1-2 是一些从公司内部记录得到的常用的数据。

表 1-2 从公司内部记录的可用数据的例子

数据来源	一些典型数据的可用数据
雇员记录	姓名、地址、社会保障号码、工资、休假天数、病假天数和奖金
生产记录	部件或产品号、生产数量、直接人工成本和原材料成本
存货记录	部件或产品号、库存单位数、重复订货水平、经济订货批量和折扣表
销售记录	产品号、销售量、区域销售量和分类客户销售量
信用记录	客户名称、地址、电话号码、信贷额度和应收账款余额
客户概况	年龄、性别、收入水平、家庭规模、地址和偏好

大量的商务和经济数据可以从专门从事搜集和保存数据的机构那里获得。公司可以通过租赁或购买的方式使用这些来自外部的数据。邓白氏、彭博社和道琼斯公司是三家能够为客户提供大量商业数据服务的公司。AC-Nielsen 和 Information Resources 公司也成功地地为广告商和制造商提供商务数据搜集和加工服务。

数据也可从各类行业协会和专门营利的机构中获得。美国旅游业协会保存各州与旅游相关的信息,如游客数量和旅游费用等,旅游行业的公司和个人会对这些数据感兴趣。研究生管理咨询委员会保存了关于考试成绩、学生特长和研究生培养计划等的信息。大多数来自这些类型的数据用户使用时需要支付适当的费用。

互联网的持续发展,已使其成为数据和统计信息的一个重要来源。几乎所有的公司都会提供公司的一般信息,如销售额、雇员数、产量、产品价格和产品说明的网站。另外,大多数公司专门在互联网获取有用的市场信息资料。因此,人们能从网上查阅到股票报价、餐馆菜价、工资数据等几乎无限的各种信息。

政府机构是现有数据的另一个重要来源。例如,美国劳工部保存大量的有关就业率、工资率、劳动力规模和工会会员等数据。表 1-3 列示了部分政府机构和它们能提供的一些数据。大部分政府机构也通过网站向社会提供其搜集和加工处理的数据。图 1-3 显示的是美国劳工统计局的主页。

表 1-3 来自选取政府机构的可用数据的例子

政府机构	一些可用的数据
人口普查局	人口数、家庭数和家庭收入的数据
联邦储备委员会	货币供应、信誉记录、汇率和贴现率的数据
管理和预算办公室	财政收入、支出和联邦政府债务的数据
商务部	商务活动数据、分行业总产值、分行业利润水平以及行业增长和下降的数据
劳工统计局	消费支出、小时工资、失业率、安全记录和国际统计资料

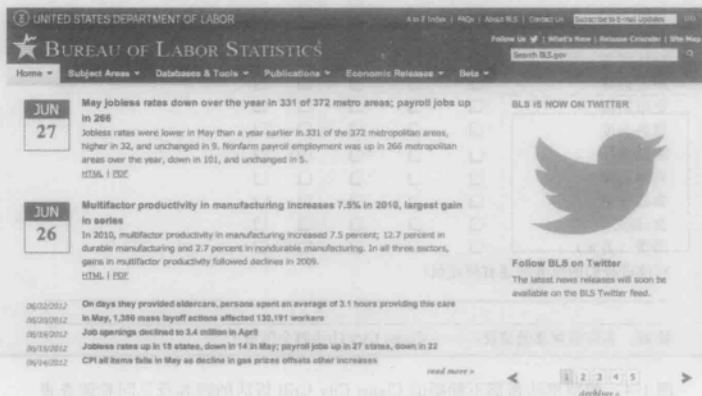


图 1-3 美国劳工统计局主页

### 1.3.2 统计研究

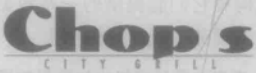
有时一些特殊需要的数据并非总能从现有来源中获得。在这种情况下,常常通过统计研究的方式获得数据。统计研究可划分为实验性的或观测性的两类。

在一项实验性的研究<sup>①</sup>中,首先要确定感兴趣的主要变量,然后控制一个或多个其他变量,以便获得它们如何影响主要变量的数据。例如,一家制药公司可能会进行一项实验,以获得一种新药如何影响血压的情况。在研究中,血压是被关注的主要变量,新药的剂量是影响血压的另一个变量,实验希望能找出新药的剂量与血压之间

① 曾经进行过的最大规模的实验性统计研究是 1954 年美国公共卫生署的沙克脊髓灰质炎疫苗实验,该研究从全美小学 1~3 年级的学生中选出了近 200 万名儿童。

的因果关系。为了获得有关新药效果的数据,研究人员选择一些个体组成样本。控制新药的剂量,不同的个体组给予不同的药剂水平,然后搜集每组服药前后的血压数据。实验数据的统计分析将会有助于了解新药如何影响血压。

在非实验性或观测性统计研究中,并不是有目的地去控制感兴趣的变量。调查也许是观测性研究中最常用的方法。例如,在一个征求顾客意见的调查中,首先要明确研究的问题,然后设计调查问卷,选择一些人作为样本。一些饭店利用观测研究来获取顾客对饭店的食品质量、服务和就餐环境等方面的数据。佛罗里达州那不勒斯的 Chops City Grill 饭店使用的顾客意见调查表如图 1-4 所示。注意,请顾客根据自己的感受对综合经验、接待问候、管理(餐桌巡视)和整体服务等 12 个变量按优秀、良好、平均、普通和较差的等级顺序打分,得到的数据能帮助 Chops City Grill 的管理人员保持高水平的饭店食品和服务质量。



---

日期: \_\_\_\_\_ 服务员姓名: \_\_\_\_\_

顾客是我们应予以最优先考虑的。请占用您一点点宝贵时间填写这张调查表,这将使我们更好地为您提供服务。您可将此表交到前台或邮寄给我们,谢谢!

调查项目	优秀	良好	平均	普通	较差
综合经验	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
接待问候	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
管理(餐桌巡视)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
整体服务	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
职业素养	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
菜单知识	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
服务态度	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
酒品选择	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
菜单选择	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
食品质量	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
食品展示	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
消费(美元)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

对改进我们的服务您还有什么建议?

\_\_\_\_\_  
 谢谢,非常感谢您的建议。——Chops City Grill 的全体员工

图 1-4 佛罗里达州那不勒斯的 Chops City Grill 饭店的顾客意见问卷调查表

将数据和统计分析结果用于管理决策,管理人员必须清楚获得数据所需要的时间和成本。当必须在一个相对较短的时期内获得数据时,利用现有数据来源是一个比较理想的方法,如果重要的数据不容易从现有来源中获取,则必须考虑获取数据所要占用的时间和成本。在所有的情形中,决策制定者必须意识到统计分析对决策制定过程中的贡献。数据搜集和统计分析所付出的成本是不会超过利用这些信息制定一个好的决策所节省的费用的。

### 1.3.3 数据采集误差

管理人员任何时候都应牢记:统计研究中的数据是可能产生误差的。使用错误的数据也许会比不使用任何数据更糟糕。每当取得的数据值不等于真值或实际值时,就会产生数据搜集误差(真值或实际值是通过一个正确的程序获得),这类错误可能在许多数据搜集方法中发生。例如,采访者可能会发生记录错误,如将 24 岁写为

③ 对吸烟者和不吸烟者的研究是观测性研究,因为调查者并不确知或控制那些人会将吸烟,那些人不会吸烟。



42岁,或者被采访者回答问题时可能曲解了问题并作出不正确回答。

实验数据分析人员在搜集和记录数据时要特别小心,以确保不发生错误。可采取一些特殊措施以检查数据的内在一致性。例如,这种措施会提醒分析人员查看一下有关数据的准确性:如一个年龄为22岁的人却说有20年的工作经验。数据分析人员也应注意查看一些过大或过小的被称为异常值的数值,它们可能是有错误的数据。在第3章我们将介绍一些统计学家用来识别异常值的方法。

在数据搜集过程中经常会发生误差。盲目地利用一切可以得到的或不认真地搜集到的数据,将可能产生误差并导致作出错误的决策。因此,按照正确的步骤搜集准确的数据可以确保决策信息的可信度,提高数据的利用价值。

## 1.4 描述统计

绝大多数的报纸、杂志、公司报告和其他出版物上的统计信息以读者易于理解的方式汇总和披露。将数据以表格、图形或数值形式汇总的统计方法被称为描述统计(descriptive statistics)。

再来参见表1-1中参加世界贸易组织的60个国家的数据集。用描述统计的方法对这一数据集中的信息进行汇总。例如,考虑未来2年内国家信用评级可能变动的方向性指标——惠誉评级展望变量,该变量用负面、稳定和正面来记录。表1-4给出了每一个成员国惠誉评级展望数据的表格汇总,图1-5给出了相同数据的图形汇总,该类图形称为条形图。这些表格和图形汇总一般能使数据更容易解释。我们能很容易地从表1-4和图1-5中看到,绝大多数信用评级的惠誉展望是稳定,有65%的国家具有这个评级。信用评级的负面展望和正面展望相似,具有负面展望的国家(18.3%)比具有正面展望的国家(16.7%)略微多一些。

图1-6是被称为直方图的表1-1中人均GDP这个数量变量数据的图形汇总。从直方图上很容易看出,60个国家的人均GDP的变动范围为0~60 000美元,主要集中在10 000~20 000美元,只有1个国家的人均GDP超过50 000美元。

表1-4 60个国家惠誉信用评级展望类型的频数和百分数

惠誉评级展望	频数	百分数(%)
正面	10	16.7
稳定	39	65.0
负面	11	18.3

除了表格和图形表示方式以外,数值型描述统计也用于汇总数据。最常用的数值型描述统计是平均数或均值。利用表1-1中60个国家人均GDP数据,我们可以将所有60个国家的人均GDP相加再除以60,计算出平均数。这样可得到人均GDP的平均数为21 387美元。这个平均数是数据的中心趋势或中心位置的一个度量。

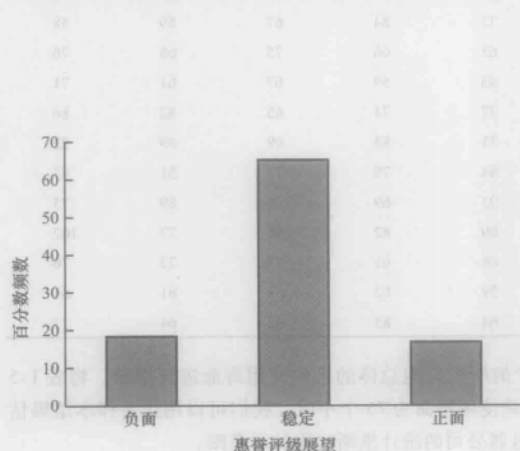


图1-5 60个国家惠誉信用评级展望的条形图

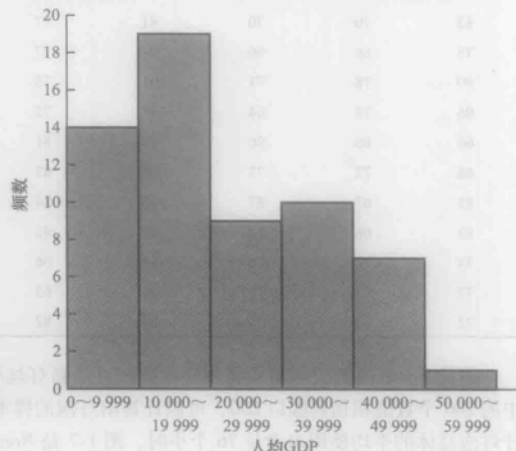


图1-6 60个国家人均GDP的直方图

近年来,人们关注描述统计的拓展和表述,对这类统计方法的兴趣在不断提高。在第2章和第3章中将集中讨论描述统计的表格、图形和数值方法。

## 1.5 统计推断

在很多情形下,我们需要搜集有关大的个体组(个人、公司、选民、家庭、产品和客户等)的信息。出于对时间、费用和其他因素的考虑,只能从大的个体组的一小部分中搜集数据。在特定研究中,较大的个体组称为总体(population),较小的个体组称为样本(sample)。我们一般使用下列定义。

**总体** 总体是在一个特定研究中所有个体组成的集合。

**样本** 样本是总体的一个子集。

搜集总体全部数据的调查过程称为普查(census)<sup>①</sup>,搜集样本数据的调查过程称为抽样调查(sample survey)。统计学的一个主要贡献就是利用样本数据对总体特征进行估计和假设检验。这一过程被称为统计推断(statistical inference)。

作为统计推断的一个例子,让我们考察由 Norris 电器公司所做的一项研究。Norris 电器公司生产一种能在各种电器产品上使用的高亮度灯泡。为了提高灯泡的使用寿命,产品设计部门开发出一种使用新型灯丝的灯泡。在这个例子中,总体定义为使用新灯丝生产的所有灯泡。为了估计新灯丝的优点,用生产出的 200 只新灯丝灯泡组成样本,并进行测试,记录每只灯泡在灯丝被烧断之前使用的小时数,样本数据见表 1-5。

表 1-5 Norris 电器公司 200 只灯泡的使用寿命

(单位:小时)

107	73	68	97	76	79	94	59	98	57
54	65	71	70	84	88	62	61	79	98
66	62	79	86	68	74	61	82	65	98
62	116	65	88	64	79	78	79	77	86
74	85	73	80	68	78	89	72	58	69
92	78	88	77	103	88	63	68	88	81
75	90	62	89	71	71	74	70	74	70
65	81	75	62	94	71	85	84	83	63
81	62	79	83	93	61	65	62	92	65
83	70	70	81	77	72	84	67	59	58
78	66	66	94	77	63	66	75	68	76
90	78	71	101	78	43	59	67	61	71
96	75	64	76	72	77	74	65	82	86
66	86	96	89	81	71	85	99	59	92
68	72	77	60	87	84	75	77	51	45
85	67	87	80	84	93	69	76	89	75
83	68	72	67	92	89	82	96	77	102
74	91	76	83	66	68	61	73	72	76
73	77	79	94	63	59	62	71	81	65
73	63	63	89	82	64	85	92	64	73

假设 Norris 电器公司希望利用样本数据对用新灯丝生产的所有灯泡总体的平均使用寿命进行推断。将表 1-5 中的 200 个数值相加再除以 200, 可以计算出灯泡的样本平均使用寿命为 76 个小时。我们可以用这一样本结果估计灯泡总体的平均使用寿命是 76 个小时。图 1-7 是 Norris 电器公司的统计推断过程的示意图。

① 美国政府每 10 年进行一次人口普查,而市场调查机构每天进行抽样调查。

每当统计学家利用样本对我们感兴趣的总体的特征进行估计时,他们通常都要注明估计的质量或精确度。以 Norris 电器公司为例,统计学家可能指出,新灯泡总体的平均使用寿命的点估计值是 76 个小时,边际误差为  $\pm 4$  小时。这样,新灯泡总体的平均使用寿命的区间估计是 72 ~ 80 个小时。统计学家也可以指出,在 72 ~ 80 个小时的区间内有多大的把握程度包含的总体平均数。

## 1.6 计算机与统计分析

由于涉及大量的数据,因此统计人员经常使用计算机软件进行分析。例如,在 Norris 电器公司的例子中(见表 1-5),如果没有计算机,那么计算 200 只灯泡的平均使用寿命将相当乏味。为了方便计算机的使用,教材中较大的数据集可以从本书所附光盘获得,下载的数据文件适合于 Minitab 和 Excel 格式。本章末的附录还有使用 Minitab, Excel 和 Excel 的插件 StatTools 来完成本章所介绍的统计方法的详细步骤。

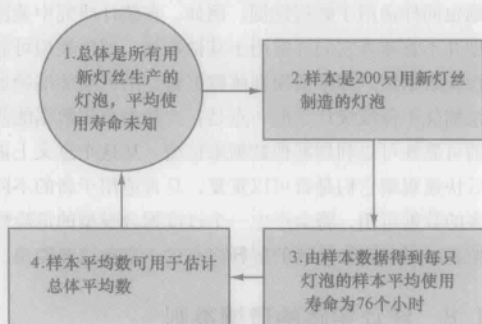


图 1-7 Norris 电器公司统计推断的过程

## 1.7 数据挖据

由于有磁卡片输入机、条形码扫描仪和 POS 终端,许多机构每天可以获取大量的基础数据。即使一个当地的使用触摸屏显示器来登录订单和处理账单的小饭店,搜集大量的数据也是有意义的。对于大型零售公司,搜集数据的真实值仅仅是概念性的,并且弄清楚如何有效地使用这些数据来提高效率也是一个难题。例如,大型零售商(如沃尔玛)每天获取 2 000 万 ~ 3 000 万个交易数据,电信公司(如法国电信和美国电话电报公司)每天有超过 3 亿个电话记录,Visa 每秒处理 6 800 次支付交易,或每天处理近 6 亿个交易。储存和管理这些交易数据是一件有意义的事情。

术语数据仓库是指获取、储存和维护数据的过程。计算能力和数据搜集工具可以在几秒内实现储存和检索非常庞大的数据。分析仓库中的数据可以对机构作出新的战略和较高收益的决策。

**数据挖据**(data mining)问题是研究从非常大的数据库中开发有用的决策信息的方法。利用统计学、数学和计算机科学的综合方法,分析人员在仓库中“挖据数据”,并将其转化为有用的信息,取名为数据挖据。数据挖据领域的先驱 Kurt Thearling 博士定义数据挖据为“从(庞大的)数据库中自动提取预测性的信息”。Thearling 博士的定义中的两个关键词是“自动”和“预测性的”。数据挖据系统是最有效的使用自动程序从数据中提取信息,而其用户仅仅需要使用最一般的甚至是模糊的查询。同时数据挖据软件使揭示隐藏的预测信息的过程自动化,而这些隐藏的信息过去是需要亲自动手进行分析。

拥有大型客服中心的公司(如零售商、金融机构和通信公司)成为数据挖据的主要应用者。数据挖据成果用于帮助零售商(如亚马逊和巴诺)确定已经购买了一件特殊商品的消费者,是否还有可能购买一件或多件相关的商品。然后,当消费者登录公司网站并且购买一件商品时,网站利用突然跳出窗口来促使消费者有可能追加购买额外商品。在另一个应用中,数据挖据用于确定在某个特殊的购物旅行中消费有可能超过 20 美元的消费者,然后将这些消费者划分为收到特别电子邮件赠券和定期电子邮件赠券的类型,从而在赠券到期之前促使他们进行下一次购物旅行。

数据挖据是一门非常依赖于统计方法的技术,这些统计方法有多元回归和逻辑回归等。它创造性地综合了所有这些方法以及包括人工智能和机器学习等计算机科学技术,从而使数据挖据是有效的。由诸如甲骨文、Teradata 和 SAS 等公司开发的用以实施商用数据挖据的软件包,需要大量的时间和金钱上的投入。本书介绍的统计概念将有助于理解数据挖据软件包中使用的统计方法,并使你更好地理解其所产生的统计信息。

由于统计模型在建立数据挖掘预测模型中扮演着重要的角色<sup>①</sup>，许多统计学家在建立统计模型中所关注的问题也同样适用于数据挖掘。例如，在统计研究中关注的问题是模型的可靠性，适用于一个特殊样本数据的统计模型并不意味着它同样适用于其他数据。评估模型可靠性的一个常见方法是将样本数据分为两部分：训练数据集和检验数据集。如果使用训练数据建立的模型能准确地预测检验数据的数值，则我们认为这个模型是可靠的。数据挖掘优于传统统计学的一点是：大量的可用数据能使数据挖掘软件划分数据集，因此由训练数据集所建立的模型的可靠性可以利用其他数据来检测。从这个意义上讲，数据集的划分允许用数据挖掘来建立模型和相互关系，然后快速观察它们是否可以重复，是否适用于新的不同的数据。另一方面，对数据挖掘应用的一个忠告是，有这么多的数据可用，将会产生一个过度拟合模型的危险性，在这一点似乎存在误导性的联系和因果关系的结论。数据挖掘结果和附加检测的解释将有助于避免这些隐患。

## 1.8 统计实践的道德准则

道德行为是我们一切活动中的追求。由于统计在数据的搜集、分析、表述和说明过程中的重要作用，道德问题继而在统计中产生。在统计研究中，不道德行为形式很多，其中包括不正确的抽样、不恰当的数据分析、误导性的图表绘制、不恰当地使用统计摘要和统计结果的片面解释，等等。

你开始做统计工作时，我们就鼓励你在搜集数据、分析、口头陈述以及在书面报告中，要保持公平、谨慎、客观和中立的态度。另外，作为数据的使用者，你也应该意识到他人在统计过程中不道德行为的可能性。当你在报纸、电视、互联网或其他媒介上见到统计时，要持几分怀疑态度，关注这些统计的来源、目的和客观性。

美国统计协会——美国最高统计和统计工作者的专业组织，发布报告“统计实践的道德准则”<sup>②</sup>，旨在帮助统计工作者作出符合道德准则的决定，并交流心得，帮助学生学会如何负责任地进行统计工作。报告中包括67条准则、8个主要领域，分别为：职业精神；对投资者、客户和雇主的责任；对出版物和证明的责任；对研究课题的责任；对研究团队的同事的责任；对其他统计学家或统计从业人员的责任；关于不当行为的指控的责任；机构、个人、代理人的雇主或其他聘用的统计从业人员的客户的责任。

职业精神领域的一条准则解决的问题是多次检测直到获得理想结果的行为。例如，在1.5节中我们讨论了Norris电器公司的一个200只新灯丝高亮度灯泡的样本的统计研究，样本的平均使用寿命为76个小时，由此推断出所有新灯丝灯泡的使用寿命。然而，思考一下，由于Norris只选取了一个样本，那么另外一个样本可能会提供不同的平均使用寿命，这么想不是没有道理的。

假定Norris的管理层希望抽样结果能支持他们声称的新灯泡的平均使用寿命至少是80个小时。再假定Norris的管理层决定继续研究，不断生产并抽样检测，每次都抽200只新灯丝灯泡，直到有一个样本的平均使用寿命达到至少80个小时。如果研究重复进行许多次，最终会有一个样本（仅是偶然）给出期望的结果，从而支持Norris作出的声明。在这种情形下，消费者将受到误导，认为新产品会比现在的要好。显然，这种行为是不道德的，它代表统计在实践中出现了重大的使用错误。

对出版物和证明领域里有几条道德准则涉及数据的处理问题。例如，一位统计学家需要说明一项研究中考虑的所有数据，并对实际上使用的样本进行解释。Norris电器公司最初样本中的200只灯泡的使用寿命是76个小时；这比管理层希望得到的至少80个小时的结果要小很多。现在假设在回顾了平均使用寿命为76个小时的结果之后，Norris剔除所有使用寿命小于或等于70个小时的观察值，并宣称生产过程中的启动问题造成了这些灯泡的瑕疵。之后，样本中剩余灯泡的平均寿命就成了82个小时。你对Norris所宣布的灯泡的平均寿命是82个小时产

① 在揭示数据中的关系和预测未来结果时，统计方法在数据挖掘中扮演着重要的角色。然而，数据挖掘的全部领域以及统计在数据挖掘中的应用已超出本书的范围。

② American Statistical Association “Ethical Guidelines for Statistical Practice”，1999。

生过怀疑吗?

如果 Norris 把使用寿命小于或等于 70 个小时的灯泡剔除,只是为了提供 82 个小时的平均使用寿命,那么毫无疑问这种做法就是不道德的。但是,如果这些灯泡真是由于在生产过程中的启动问题造成的瑕疵,因而不应该包含在分析中,进行研究的统计人员应该考虑所有的数据并对抽样的具体操作进行解释。否则,就可能产生误导,对公司和统计人员都构成不道德行为。

在美国统计协会报告的共享价值部分有一条准则提出,统计从业人员应当避免任何对预测结果有倾向性的统计工作。当使用不具有代表性的样本而作出声明时,这种不道德的行为就会发生。例如,在美国的许多地区,饭店内是不允许吸烟的。假如烟草业的说客在允许吸烟的饭店进行调查,目的是估计赞成在饭店内允许吸烟的比例。抽样结果表明 90% 的受访者同意在饭店内允许吸烟。根据这个抽样结果,说客声称饭店所有就餐者中有 90% 的人赞成在饭店内允许吸烟。在这个案例中,我们应该指出只在允许吸烟的饭店所做的抽样会歪曲调查结果。如果只报告该研究的这个最后结果,而不熟悉研究的细节(如样本是在允许吸烟的饭店搜集的),读者就会被误导。

美国统计协会报告的涵盖范围很广,道德准则不仅适用于统计工作者,同样对统计信息的使用者也有效。当你从事统计研究工作时,我们建议你读一读这个报告,对道德相关问题有更好的了解,并获得一些背景知识,以确保当你开始在实践中使用统计工作时,你的工作符合道德标准。

## 小结

统计是搜集、分析、表述和解释数据的艺术和科学。几乎每个商务或经济专业的大学生都要学习统计学课程。我们以描写商务和经济中典型的统计应用开始本章的论述。

数据是指搜集和分析的事实和数字。我们可以用四种测量尺度来取得一个特定变量的数据,它们是名义尺度、顺序尺度、间隔尺度和比率尺度。当数据用于确认个体属性的标记或名称时,相应变量的测量尺度称为名义尺度。如果数据显示名义数据的性质,并且数据的顺序或等级是有意义的,则测量尺度是顺序尺度。如果数据具有顺序数据的性质,并且可以按某一固定度量单位表示数值间的间隔,则测量尺度是间隔尺度。最后,如果数据具有间隔数据的所有性质,并且两个数值之比是有意义的,则测量尺度是比率尺度。

为了统计分析之目的,数据可划分为分类型数据和数量型数据。分类型数据是用标记或名称来识别每一个体属性。分类型数据既可以用名义尺度度量也可以用顺序尺度度量,既可以是非数值型的也可以是数值型的。数量型数据是表示大小或多少的数值。数量型数据既可以用间隔尺度度量也可以用比率尺度度量。只有当数据是数量型的,普通的算术运算才有意义。因而,适合于数量型数据的统计计算并非总适用于分类型数据。

在第 1.4 和第 1.5 节我们介绍了描述统计和统计推断。描述统计是用表格、图形和数值方法来汇总数据。统计推断是利用样本数据估计总体特征并进行假设检验的过程。在本章最后三节,给出了计算机在统计分析中作用的信息,介绍了数据挖掘这个比较新的领域,并且简要描述统计实践的道德准则。

## 关键术语

**statistics 统计学** 搜集、分析、表述和解释数据的艺术和科学。

**data 数据** 是描述和解释所搜集、分析和汇总的事实和数字。

**data set 数据集** 在特定研究中搜集到的所有数据。

**elements 个体** 搜集数据的实体。

**variable 变量** 个体中所感兴趣的特征。

**observation 观测值** 对某一特定个体搜集的测量值

的集合。

**nominal scale 名义尺度** 是指用数据的标记或名称来识别个体属性的一种变量测量尺度,名义数据既可以是非数值型的,也可以是数值型的。

**ordinal scale 顺序尺度** 具有名义数据的性质,并能显示数据的顺序或等级有明确意义的一种变量测量尺度。顺序数据既可以是非数值型的,也可以是数值型的。

**interval scale 间隔尺度** 具有顺序数据的性质,并可  
以按某一固定度量单位表示数值间的间隔的一种变  
量测量尺度。间隔数据永远是数值型的。

**ratio scale 比率尺度** 具有所有间隔数据的性质,并  
且两个数值之比是有意义的一种变量测量尺度。比  
率数据永远是数值型的。

**categorical data 分类型数据** 用于识别每一个体属  
性的标记或名称。分类型数据既可以用名义尺度度  
量也可以用顺序尺度度量,既可以是非数值型的也  
可以是数值型的。

**quantitative data 数量型数据** 表示事物大小或多  
少的数值。数量型数据既可以用间隔尺度度量也  
可以用比率尺度度量。

**categorical variable 分类变量** 用分类型数据表示  
的变量。

**quantitative variable 数量变量** 用数量型数据表示

变量。

**cross-sectional data 截面数据** 在相同或近似相  
同的同一时点上搜集的数据。

**time series data 时间序列数据** 在几个时期内搜  
集的数据。

**descriptive statistics 描述统计** 数据的表格、图形和  
数值汇总的方法。

**population 总体** 在一个特定研究中所有感兴趣的个  
体组成的集合。

**sample 样本** 总体的一个子集。

**census 普查** 搜集总体全部数据的调查。

**sample survey 抽样调查** 搜集样本数据的调查。

**statistical inference 统计推断** 利用样本数据估计或  
对总体特征进行假设检验的过程。

**data mining 数据挖掘** 利用统计和计算机科学的方法  
从非常大的数据库中获取有用信息的过程。

## 补充练习

2. 美国能源部提供各种汽车燃料燃烧效率的信息。

表 1-6 是由 10 辆汽车组成的一个样本 (Fuel Economy website, 2008 年 2 月 22 日)。数据有汽车类型 (小型、中型或大型)、发动机的汽缸数、每加仑城市行驶里程 (单位:英里)、每加仑公路行驶里程以及推荐使用的燃料 (柴油、优质汽油或普通汽油)。

- 这个数据集中有多少个体?
- 这个数据集中有多少变量?
- 哪些变量是分类变量,哪些变量是数量变量?

d. 每个变量使用哪种类型的测量尺度?

4. 表 1-7 中给出 7 所大学的数据,其中有捐赠额 (单位:10 亿美元)和申请者被接受的比例 (《今日美国》,2008 年 2 月 3 日)。每所学校所在的州、校园所处位置以及学校代表队在 NCAA 的分级可以从美国教育统计中心网站 (2008 年 2 月 22 日)上得到。

- 这个数据集中有多少个体?
- 这个数据集中有多少个变量?
- 哪些变量是分类变量,哪些是数量变量?

表 1-6 10 辆汽车燃料燃烧效率信息

汽车	类型	汽缸数	城市行驶 (英里/加仑)	公路行驶 (英里/加仑)	燃料
奥迪 A8	大型	12	13	19	优质汽油
宝马 328Xi	小型	6	17	25	优质汽油
凯迪拉克 CTS	中型	6	16	25	普通汽油
克莱斯勒 300	大型	8	13	18	优质汽油
福特福克斯	小型	4	24	33	普通汽油
现代伊兰特	中型	4	25	33	普通汽油
吉普大切诺基	中型	6	17	26	柴油
庞蒂亚克 G6	小型	6	15	22	普通汽油
丰田凯美瑞	中型	4	21	31	普通汽油
大众捷达	小型	5	21	29	普通汽油

⊖ 本书所有章节的习题部分中单数题请参见本书配套光盘。

⊖ 1 加仑 (美制) ≈ 3.8 升。



表 1-7 7 所大学的数据

学校	州	校园位置	捐赠 (10 亿美元)	申请者被接受的比例 (%)	NCAA 的分级
安默特斯学院	马萨诸塞	城镇: 边缘	1.7	18	III
杜克大学	北卡罗来纳	城市: 中型	5.9	21	I-A
哈佛大学	马萨诸塞	城市: 中型	34.6	9	I-AA
斯沃斯莫尔学院	宾夕法尼亚	市郊: 大型	1.4	18	III
宾夕法尼亚大学	宾夕法尼亚	城市: 大型	6.6	18	I-AA
威廉姆斯学院	马萨诸塞	城镇: 边缘	1.9	18	III
耶鲁大学	康涅狄格	城市: 中型	22.5	9	I-AA

6. 《外交事务》杂志进行了一项调查, 以了解其订阅者的简况 (《外交事务》, 2008 年 2 月 23 日)。提出的问题如下:

- 在过去的 12 个月内您有几天住在旅店?
- 您在哪儿购书? 列出了 3 个选项: 书店、互联网和读书俱乐部
- 您是否拥有或租赁过豪华车? (是或否)
- 您多大年龄?
- 在过去 3 年的国外旅行中, 您的目的地是哪里? 列出了 7 个国外目的地

对每一个问题的数据是分类型的还是数量型的发表意见。

8. 金融时报/哈里斯民意调查对 6 个欧洲国家和美国的成年人每月进行一次民意在线调查。美国有 1 015 名成年人参加了 1 月份的民意调查, 其中一个问题是: “您如何评价联邦银行在处理金融市场信用问题时的表现?” 可能的答案有: 优秀、良好、中等、较差和极差 (Harris Interactive website, 2008 年 1 月)

- 这项调查的样本容量是多大?
- 数据是分类型的还是数量型的?
- 对这一问题, 用平均数或百分比数据汇总, 哪种方法更合适?
- 在美国的回答者中, 10% 的人说联邦银行做得很好, 有多少人给出这种答复?

10. 运输统计局综合住房调查是一项年度调查, 并作为美国交通运输部的信息来源。在部分调查中, 采访了调查者, 并询问如下问题: “机动车驾驶员在开车时是否可以手持电话接听。” 可能的答案有: 非常同意、稍微同意、稍微反对和非常反对。对这个问题有 44 人表示非常同意、130 人表示稍微同意、165 人表示稍微反对、741 人表示非常反对 (Bureau of Transportation website, 2010 年 8 月)。

a. 这个问题的答案提供的是分类型数据还是数量型数据?

b. 对这一问题的答案, 用平均数或百分比数据汇总, 哪种方法更合适?

c. 表示非常同意允许机动车驾驶员在开车时可以手持电话接听的答案占多大比例?

d. 该结果预示普遍支持还是反对允许机动车驾驶员在开车时可以手持电话接听?

12. 夏威夷旅游局搜集在夏威夷的游客数据。下列问题是在飞往夏威夷的航班上, 向乘客分发的问卷中所征询的 16 个问题中的一部分。

- 此行是我的第几次夏威夷之旅: 第 1, 2, 3, 4... 次
- 此行的主要理由是: (包括度假、惯例和蜜月等 10 种)
- 我计划住在: (包括旅店、公寓、亲戚和宿舍等 11 类)
- 在夏威夷逗留的总天数

a. 被研究的总体是什么?

b. 对飞往夏威夷航班的乘客总体, 用问卷调查是一个好方法吗?

c. 对 4 个问题中的每一个所提供的分类型数据还是数量型数据进行评论。

14. 下面是 2007~2010 年 Hertz, Avis 和 Dollar 这 3 家提供租车服务的公司的租车数量数据 (单位: 1 000 辆)。

公司	2007	2008	2009	2010
Hertz	327	311	286	290
Dollar	167	140	106	108
Avis	204	220	300	270

a. 绘制 2007~2010 年每个租车公司租车数量的时间序列图, 在同一张图上显示 3 家公司的时间

序列。

- b. 市场占有率最大的租车公司是哪家？对市场占有率如何随时间变动加以评价。
  - c. 用 2010 年的租车数据绘制条形图。这张图是依据截面数据还是时间序列数据绘制的？
16. 美国能源部能源信息局提供了 2007 年 1 月至 2012 年 3 月之间美国常规普通汽油每加仑平均价格的时间序列数据（Energy Information Administration website, 2012 年 4 月）。利用互联网获取 2012 年 3 月以来的常规普通汽油每加仑的平均价格。
- a. 扩展图 1-1 的时间序列图形。
  - b. 你如何解释从 2012 年 3 月以来常规普通汽油每加仑的平均价格？
  - c. 时间序列是否继续呈现每加仑的平均价格在夏季上涨？请解释。
18. 关于完税、在线纳税筹划和电子申报服务，舆论研究公司对 1021 名成年人（18 岁及以上）进行电话随机调查。调查结果表明其中有 684 人计划使用电子申报他们的税务（CompleteTax Tax Prep Survey, 2010 年）。
- a. 确定一种描述统计量，能用于估计所有使用电子申报的纳税人的百分数。
  - b. 调查报告说，人们最常用的报税方法是雇佣会计师或职业报税人。如果有 60% 的调查者使用了这种方式，那么有多少人会雇佣会计师或职业报税人？
  - c. 人们申报纳税常用的其他方法还有手工填报、使用在线税务服务和税务软件。申报纳税方法的数据是分类型数据还是数量型数据？
20. 《巴伦周刊》Big Money 对 131 名投资经理进行了一次民意调查，结果如下：
- 43% 的经理认为他们自己在股票市场操作极佳。
  - 在接下来的 12 个月里，股票的平均预期回报率为 11.2%。
  - 21% 的经理认为医疗健康类股票极有可能是接下来 12 个月中股票市场的主导板块。
  - 当问及预计科技股和电信股估计大约需要多长时间才恢复上涨时，他们的回答都是 2.5 年。
- a. 列举两种描述统计。
  - b. 对接下来 12 个月里投资经理总体的股票平均预期回报率作出推断。
  - c. 对科技股和电信股恢复上涨的时间长度进行推断。
22. 尼尔森公司从欧洲、亚太、美洲和中东的 47 个市场对消费者进行调查，以确定影响他们购买食品地点的最主要的因素。用等级 1（最低）到 5（最高）表示。等级最高的因素是物有所值，平均分为 4.32；等级次高的因素是更好地选择优质品牌和产品，平均分为 3.78；等级最低的因素是使用可循环购物袋和包装，平均分为 2.71（Nielsen website, 2008 年 2 月 24 日）。假设你受雇于一个食品连锁店进行类似的调查，以确定在北卡罗来纳州夏洛特市连锁店，消费者认为哪个因素是影响他们确定购买食品地点的最主要因素。
- a. 你进行的这项研究的总体是什么？
  - b. 你如何搜集这项研究的数据？
24. 5 名学生期中考试成绩的样本数据如下：72, 65, 82, 90, 76。下列表述中哪一个是正确的，哪一个因太空泛而受到怀疑？
- a. 5 名学生的平均期中成绩是 77。
  - b. 参加考试的所有学生的平均期中成绩是 77。
  - c. 参加考试的所有学生的平均期中成绩的估计值是 77。
  - d. 一半以上参加考试的学生的成绩在 70~85。
  - e. 如果这个样本中还包含其他 5 名学生，他们的成绩将在 65~90。

Q105	Q005	Q008	Q009	Q010
100	100	100	100	100
101	101	101	101	101
102	102	102	102	102

## 第2章

# 描述统计学 I：表格法和图形法

## CHAPTER

# 2

实践中的统计：高露洁-棕榄公司

- 2.1 汇总分类变量的数据
- 2.2 数量型数据汇总
- 2.3 用表格方式汇总两个变量的数据
- 2.4 用图形显示方式汇总两个变量的数据
- 2.5 数据可视化：创建有效图形显示的最佳实践

## 实践中的统计

高露洁-棕榄公司<sup>①</sup>

纽约州, 纽约市

高露洁-棕榄 (Colgate-Palmolive) 公司于 1806 年在纽约开业时, 还只是一家经营香皂和蜡烛的小商店。但今天, 高露洁-棕榄公司已是一家遍布世界 200 多个国家和地区, 拥有超过 40 000 名员工的大公司。除了最著名的品牌高露洁、棕榄、Ajax 和 Fab 外, 同时还兼营 Mennen、希尔斯科学食品 (Hill's Science Diet) 和希尔斯处方食品 (Hill's Prescription Diet) 等产品。

高露洁-棕榄公司在其家用洗衣粉产品的生产过程中质量保证计划中使用统计学方法。一个关键问题是顾客对盒装洗衣粉数量的满意度。相同尺寸的每一个盒子里都装入相同重量的洗衣粉, 但是洗衣粉所占的体积受洗衣粉密度的影响。例如, 如果洗衣粉的密度较大时, 要达到所规定的包装重量, 就只需要较少体积的洗衣粉。这样, 当顾客打开包装盒时, 盒子看上去显然没有装满。

为了控制洗衣粉密度过大这个问题, 需要对洗衣粉密度的可接受范围加以限制。定期抽取统计样本, 并测量每一个洗衣粉样本的密度。然后将汇总数据提供给操作人员, 以便在需要将密度保持在规定的质量规格内时, 操作人员可以采取正确的措施。

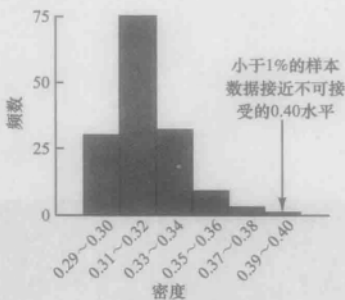
根据在一周内采集的一个容量为 150 的样本, 得到密度的频数分布和直方图如右图所示。密度水平超过 0.40 是不可以接受的上限。频数分布和直方图显示, 所有产品的密度小于或等于 0.40 时, 操作符合质

量标准。检查这些统计汇总结果的管理人员对洗衣粉产品的质量感到满意。

在本章中, 我们将学习有关描述统计的表格法和图形法, 包括频数分布、条形图、直方图、茎叶显示、交叉分组表等内容。这些方法的目的是汇总数据, 以便使统计数据易于理解和解释。

密度数据的频数分布

密度	频数
0.29 ~ 0.30	30
0.31 ~ 0.32	75
0.33 ~ 0.34	32
0.35 ~ 0.36	9
0.37 ~ 0.38	3
0.39 ~ 0.40	1
总计	150



密度数据的直方图

正如第 1 章所述, 数据可以分为分类型或数量型, 分类型数据 (categorical data) 是用标签或名称来识别项目的类型。数量型数据 (quantitative data) 是表示多少或大小的数值。本章介绍的是用于汇总分类型数据和数量型数据的表格法和图形法。人们在年度报告、报纸文章和研究报告中常可以看到数据的表格和图形汇总, 因此, 理解它们是怎样形成的以及如何解释它们至关重要。

我们首先讨论汇总单变量数据的表格法和图形法, 接下来讨论汇总两个变量数据的表格法和图形法, 它们在某种程度上揭示了两个变量之间的关系。术语数据可视化 (data visualization) 常常用于描述汇总和表述一个数据集信息的图形显示的效用。最后一节给出了数据可视化的简介和创建有效的图形显示的指导方针。

现代统计软件包提供了汇总数据和编制图形的强大功能, Minitab 和 Excel 是其中应用最广泛的两个软件包。在本章的附录中, 我们将演示它们的部分功能。

① 作者感谢为“实践中的统计”提供了案例的高露洁-棕榄公司质量保证部经理 William R. Fowle。

## 2.1 汇总分类变量的数据

### 2.1.1 频数分布

我们从频数分布 (frequency distribution) 的定义开始, 讨论如何使用表格法和图形法汇总分类类型数据。

#### 频数分布

频数分布是一种数据的表格汇总, 表示在几个互不重叠组别中的每一组项目的个数 (即频数)。

我们用下面的例子来说明如何编制和解释分类类型数据的频率分布。可口可乐、健怡可乐、胡椒博士、百事可乐和雪碧是 5 种受欢迎的软饮料。假设表 2-1 中的数据是 50 次购买软饮料的样本数据。

表 2-1 50 次软饮料购买的样本数据

可口可乐	可口可乐	可口可乐	雪碧	可口可乐
健怡可乐	胡椒博士	健怡可乐	胡椒博士	健怡可乐
百事可乐	雪碧	可口可乐	百事可乐	百事可乐
健怡可乐	可口可乐	雪碧	健怡可乐	百事可乐
可口可乐	健怡可乐	百事可乐	百事可乐	百事可乐
可口可乐	可口可乐	可口可乐	可口可乐	百事可乐
胡椒博士	可口可乐	可口可乐	可口可乐	可口可乐
健怡可乐	雪碧	可口可乐	可口可乐	胡椒博士
百事可乐	可口可乐	百事可乐	百事可乐	百事可乐
百事可乐	健怡可乐	可口可乐	胡椒博士	雪碧

为了编制这些数据的频数分布, 我们计算表 2-1 中每一种软饮料的出现次数。可口可乐出现 19 次, 健怡可乐出现 8 次, 胡椒博士出现 5 次, 百事可乐出现 13 次, 雪碧出现 5 次。这些数字汇总在表 2-2 的频数分布中。

这个频数分布汇总说明了在 50 次购买软饮料的样本中, 5 种软饮料是如何分配的。它提供了比表 2-1 中的原始数据更多的内涵。观察这个频数分布, 我们看到可口可乐是排在首位, 百事可乐居第二位, 第三位是健怡可乐, 雪碧和胡椒博士并列第四。频数分布还揭示了这 5 种销售量最高的软饮料受欢迎程度的信息。

表 2-2 购买软饮料的频数分布

软饮料	频数
可口可乐	19
健怡可乐	8
胡椒博士	5
百事可乐	13
雪碧	5
总计	50

### 2.1.2 相对频数分布和百分数频数分布

频数分布表明在几个互不重叠的组别中每一组项目的个数 (即频数)。然而, 我们往往对每一组的项目所占的比例或百分比更感兴趣。一组的相对频数是属于该组别的项目个数占总数的比例。对一个有  $n$  个观测值的数据集, 每一组的相对频数由下式给出。

相对频数

$$\text{组的相对频数} = \frac{\text{组的频数}}{n} \quad (2-1)$$

一组的百分数频数是相对频数乘以 100。

相对频数分布 (relative frequency distribution) 是每一组相对频数数据的表格汇总。百分数频数分布 (percent frequency distribution) 是每一组百分数频数数据的表格汇总。表 2-3 是软饮料数据的相对频数分布的百分数频数分布。在表 2-3 中, 我们看到可口可乐的相对频数为  $19/50 = 0.38$ , 健怡可乐的相对频数为  $8/50 = 0.16$ , 依此类推。从百分数频数分布中, 我们看到在购买的软饮料中, 有 38% 是可口可乐, 有 16% 是健怡可乐, 等等。我们还注意到前三类软饮料占购买量的 80% ( $= 38\% + 26\% + 16\%$ )。