

需要整本电子书，联系我QQ: [2667271557](#);  
此处是样章，取的完整版的前面几页，和最后  
面几页；完整版是带书签的，样章没带书签；  
另外需要其他书，也可以找我。

# 谁说菜鸟不会 数据分析

狄松 祝迎春 张文霖 马世澎 编著

SPSS 18.0  
从入门到精通



中国工信出版集团



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
http://www.phei.com.cn



# 谁说菜鸟不会 数据分析

狄松 祝迎春 张文霖 马世澎 编著



(SPSS篇)

电子工业出版社

Publishing House of Electronics Industry

北京•BEIJING

## 内 容 简 介

作为《谁说菜鸟不会数据分析》家族的新成员，本书依然通俗地讲解数据分析的实践。

本书继续采用职场三人行的方式来构建内容，细致梳理了准专业数据分析的常见问题，并且挑选出企业实践中最容易碰到的案例，以最轻松直白的方式来讲好数据分析的故事。

本书从解决工作中的实际问题出发，从统计描述、统计推断到探索性分析，总结并提炼工作中经常用到并且非常实用的通过 SPSS 进行数据处理、数据分析实战方法与技巧。本书尽可能避免使用晦涩难懂的统计术语或模型公式，如需了解相关的统计学知识，可查阅相关的统计学书籍。

本书适合刚踏出校门，初涉职场的新人，尤其适合从事产品运营、市场营销、金融、财务、人力资源管理等工作的上班族们，本书能帮助他们提高工作效率；而从事管理、咨询、研究等工作的专业人士，也不妨阅读本书，说不定会有惊喜的发现。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

### 图书在版编目 ( CIP ) 数据

谁说菜鸟不会数据分析. SPSS 篇 / 狄松等编著. —北京: 电子工业出版社, 2016.6  
ISBN 978-7-121-28801-2

I. ①谁… II. ①狄… III. ①表处理软件②统计分析—软件包 IV. ① TP391.13 ② C819

中国版本图书馆 CIP 数据核字 ( 2016 ) 第 101045 号

策划编辑: 张月萍

责任编辑: 葛 娜

印 刷: 中国电影出版社印刷厂

装 订: 三河市皇庄路通装订厂

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 787×980 1/16 印张: 14.25 字数: 341 千字

版 次: 2016 年 6 月第 1 版

印 次: 2016 年 6 月第 1 次印刷

印 数: 10000 册 定价: 59.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: ( 010 ) 88254888, 88258888。

质量投诉请发邮件至 [zltts@phei.com.cn](mailto:zltts@phei.com.cn), 盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式: ( 010 ) 51260888-819, [faq@phei.com.cn](mailto:faq@phei.com.cn)。

# 目 录

## 第1章 SPSS 概况 / 11

- 1.1 SPSS 简介 / 12
- 1.2 SPSS 特点 / 13
- 1.3 SPSS 安装 / 15
- 1.4 SPSS 窗口 / 19
- 1.5 本章小结 / 22

## 第2章 数据处理 / 23

- 2.1 数据变量 / 24
  - 2.1.1 数据类型 / 24
  - 2.1.2 变量尺度 / 25
- 2.2 数据导入 / 27
  - 2.2.1 Excel 数据导入 / 27
  - 2.2.2 文本数据导入 / 29
- 2.3 数据清洗 / 33
- 2.4 数据抽取 / 35
  - 2.4.1 字段拆分 / 35
  - 2.4.2 随机抽样 / 38
- 2.5 数据合并 / 40
  - 2.5.1 字段合并 / 40
  - 2.5.2 记录合并 / 41
- 2.6 数据分组 / 43
  - 2.6.1 可视分箱 / 43
  - 2.6.2 重新编码 / 46
- 2.7 数据标准化 / 48
  - 2.7.1 0-1 标准化 / 48
  - 2.7.2 Z 标准化 / 50
- 2.8 本章小结 / 50



## 第3章 描述性分析 / 53

### 3.1 频率分析 / 54

3.1.1 分类变量频率分析 / 54

3.1.2 连续变量频率分析 / 57

### 3.2 描述分析 / 61

### 3.3 交叉表分析 / 63

### 3.4 多选题定义 / 65

### 3.5 数据报表制作 / 68

3.5.1 报表类型简介 / 69

3.5.2 分类变量报表制作 / 70

3.5.3 连续变量报表制作 / 72

3.5.4 多选题报表制作 / 73

3.5.5 报表灵活运用 / 75

### 3.6 本章小结 / 80

## 第4章 相关分析 / 81

### 4.1 相关分析简介 / 82

### 4.2 相关分析实践 / 84

4.2.1 散点图绘制 / 85

4.2.2 相关分析操作 / 86

### 4.3 本章小结 / 87

## 第5章 回归分析 / 89

### 5.1 回归分析简介 / 90

5.1.1 什么是回归分析 / 90

5.1.2 线性回归分析步骤 / 91

### 5.2 简单线性回归分析 / 92

5.2.1 简单线性回归分析简介 / 92

5.2.2 简单线性回归分析实践 / 93

### 5.3 多重线性回归分析 / 99

5.3.1 多重线性回归分析简介 / 99

5.3.2 多重线性回归分析实践 / 99

## 5.4 本章小结 / 106

**第6章 自动线性建模 / 107**

## 6.1 自动建模 / 108

## 6.2 模型结果解读 / 113

## 6.3 模型预测 / 121

## 6.4 本章小结 / 122

**第7章 Logistic 回归 / 123**

## 7.1 Logistic 回归简介 / 124

## 7.2 Logistic 回归实践 / 127

## 7.2.1 Logistic 回归操作 / 128

## 7.2.2 Logistic 回归结果解读 / 129

## 7.2.3 Logistic 回归预测 / 131

## 7.3 本章小结 / 135

**第8章 时间序列分析 / 137**

## 8.1 时间序列分析简介 / 138

## 8.2 季节分解法 / 139

## 8.3 专家建模法 / 148

## 8.3.1 时间序列预测步骤 / 148

## 8.3.2 时间序列分析操作 / 149

## 8.3.3 时间序列分析结果解读 / 151

## 8.3.4 时间序列预测应用 / 153

## 8.4 本章小结 / 157

**第9章 RFM 分析 / 159**

## 9.1 RFM 分析介绍 / 160

## 9.2 RFM 分析操作 / 162

## 9.2.1 数据准备 / 162

## 9.2.2 RFM 分析实践 / 163

## 9.2.3 RFM 分析结果解读 / 167

## 9.3 RFM 分析应用 / 170

## 9.4 本章小结 / 175

## 第10章 聚类分析 / 177

### 10.1 聚类分析介绍 / 178

### 10.2 快速聚类分析 / 180

#### 10.2.1 快速聚类分析操作 / 180

#### 10.2.2 快速聚类分析结果解读 / 182

### 10.3 系统聚类分析 / 186

#### 10.3.1 系统聚类分析操作 / 186

#### 10.3.2 系统聚类分析结果解读 / 189

### 10.4 二阶聚类分析 / 193

#### 10.4.1 二阶聚类分析操作 / 193

#### 10.4.2 二阶聚类分析结果解读 / 195

### 10.5 聚类方法的对比 / 201

### 10.6 本章小结 / 202

## 第11章 因子分析 / 203

### 11.1 因子分析简介 / 204

### 11.2 因子分析实践 / 206

#### 11.2.1 因子分析操作 / 207

#### 11.2.2 因子分析结果解读 / 210

### 11.3 本章小结 / 217

## 第12章 对应分析 / 219

### 12.1 对应分析简介 / 220

### 12.2 对应分析实践 / 221

#### 12.2.1 对应分析操作 / 221

#### 12.2.2 对应分析结果解读 / 225

### 12.3 本章小结 / 228





# 谁说菜鸟不会 数据分析

狄松 祝迎春 张文霖 马世澎 编著

（第2版）

电子工业出版社  
Publishing House of Electronics Industry  
北京•BEIJING

## 内 容 简 介

作为《谁说菜鸟不会数据分析》家族的新成员，本书依然通俗地讲解数据分析的实践。

本书继续采用职场三人行的方式来构建内容，细致梳理了准专业数据分析的常见问题，并且挑选出企业实践中最容易碰到的案例，以最轻松直白的方式来讲好数据分析的故事。

本书从解决工作中的实际问题出发，从统计描述、统计推断到探索性分析，总结并提炼工作中经常用到并且非常实用的通过 SPSS 进行数据处理、数据分析实战方法与技巧。本书尽可能避免使用晦涩难懂的统计术语或模型公式，如需了解相关的统计学知识，可查阅相关的统计学书籍。

本书适合刚踏出校门，初涉职场的新人，尤其适合从事产品运营、市场营销、金融、财务、人力资源管理等工作上班族们，本书能帮助他们提高工作效率；而从事管理、咨询、研究等工作的专业人士，也不妨阅读本书，说不定会有惊喜的发现。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

### 图书在版编目 ( CIP ) 数据

谁说菜鸟不会数据分析：SPSS 篇 / 狄松等编著. —北京：电子工业出版社，2016.6  
ISBN 978-7-121-28801-2

I . ①谁… II . ①狄… III . ①表处理软件②统计分析—软件包 IV . ① TP391.13 ② C819

中国版本图书馆 CIP 数据核字 ( 2016 ) 第 101045 号

策划编辑：张月萍

责任编辑：葛 娜

印 刷：中国电影出版社印刷厂

装 订：三河市皇庄路通装订厂

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：787×980 1/16 印张：14.25 字数：341 千字

版 次：2016 年 6 月第 1 版

印 次：2016 年 6 月第 1 次印刷

印 数：10000 册 定价：59.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：( 010 ) 88254888，88258888。

质量投诉请发邮件至 [zlt@phei.com.cn](mailto:zlt@phei.com.cn)，盗版侵权举报请发邮件至 [dbqq@phei.com.cn](mailto:dbqq@phei.com.cn)。

本书咨询联系方式：( 010 ) 51260888-819，[faq@phei.com.cn](mailto:faq@phei.com.cn)。

# 前言

自《谁说菜鸟不会数据分析》系列图书上市以来，已拥有数十万读者与粉丝，口口相传，成为职场人士案头必备的参考用书，遇到问题随手翻翻，总能找到一些快意的办法，打开脑洞。同时非常荣幸地获得“出版全行业优秀畅销品”称号，这离不开广大读者的厚爱与支持。

随着数据分析在日常工作和生活中的重要性日益凸显，对于一些需要不断提升的读者来说，他们已经不满足于现状，迫切需要增强在数据分析方面的专业性。而 SPSS 因为操作简便，无须编程，分析专业，几乎是业余进阶专业的必备工具。这也促使众多读者来信催我们早日出版《谁说菜鸟不会数据分析（SPSS 篇）》。

有了上千位热心读者的不断来信咨询与支持，经过两年时间的打磨，这本书总算与读者见面了。

这本书从解决工作中的实际问题出发，总结并提炼工作中 SPSS 经常用到并且非常实用的数据处理、数据分析实战方法与技巧。本书力求通俗易懂地介绍数据分析方法与技巧，在不影响学习理解的前提下，尽可能避免使用晦涩难懂的统计术语或模型公式，如需了解相关的统计学知识，可查阅相关的统计学书籍。

本书第 1 和第 2 章由张文霖完成，第 3 章由狄松完成，第 4 和第 5 章由马世澎完成，第 6~12 章由祝迎春完成，最终由狄松统一审稿。整个写作过程是艰辛的，但是也很有成就感。我们努力讲好数据分析的故事，同时把这个故事尽量展现得美丽动人。

**如果你觉得她看起来很轻松，千万别误以为她是一本小说，她其实是一本数据分析书**

她抛开复杂的数学或者统计学原理，她只和你讲必知必会的要点，关注解决实际问题；

她不去探究科班的学术问题，她只和你耐心地分享职场中的实战案例；

她不板起脸和你讲大道理，她只和你娓娓道来切身的趣味故事；

她天生丽质，图表漂亮绝伦；

她多姿多彩，还有卡通漫画风；

可能你会觉得她肤浅……

但是，当你揭开她华丽的外衣时，你会惊艳；

也会被她通俗而不庸俗，美丽而又深刻的本质所吸引。

把她珍藏起来吧，因为：

她会循循善诱地把你领进数据分析的大门；

她会让你的简历更加具有吸引力；

## >> 谁说菜鸟不会数据分析（SPSS 篇）

她会让老板对你刮目相看；

她值得在你的书架上长期逗留，会为你的书架增添色彩。

她讲述了职场三人行的故事，她的故事还会让你偷着笑

**牛董，关键词：**私企董事、要求严格、为人苛刻。

**小白，关键词：**在职场打拼一年的伪白骨精（白领 + 骨干 + 精英）、数据分析师、单身女白领、爱臆想。

**Mr. 林，关键词：**小白现任上司、数据分析达人、成熟男士、乐于助人、做事严谨。

哪些人会对她的故事有阅读兴趣呢

- ★ 需要提升自身竞争力的职场新人。
- ★ 在市场营销、金融、财务、人力资源、产品设计等管理工作中需要进行数据分析的人士。
- ★ 经常阅读经营分析、市场研究报告的各级管理人员。
- ★ 从事咨询、研究、分析等工作的专业人士。

故事作者的致谢

感谢广大读者的支持，让作者下定决心写这本书。在此要衷心感谢成都道然科技有限责任公司 的姚新军先生，感谢他的提议和在写作过程中的支持。感谢参与本书优化的朋友：王斌、李伟、张强林、万雷、李平、王晓、景小燕、余松。非常感谢本书的插画师王馨和张雅文的辛苦劳动，您的作品也让本书增色了不少。

感谢邓凯、黄成明、石军、沈浩、郑来轶、马广斌等书评作者，感谢他们在百忙之中抽空阅读书稿，撰写书评，并提出宝贵意见。

最后，感谢四位作者的家人，感谢他们默默无闻的付出，没有他们的理解与支持，同样也没有本书。

尽管我们对书稿进行了多次修改，仍然不可避免地会有疏漏和不足之处，敬请广大读者批评指正，我们会在适当的时间进行修订，以满足更多人的需要。

本书配套案例数据下载方式：

（1）<http://blog.sina.com.cn/xiaowenzi22>

（2）关注微信订阅号：小蚊子数据分析，回复“1”或“SPSS 篇”获取下载链接

（3）<http://read.zhiliaobang.com/pages/article/43>



## 业内人士的推荐（排名不分先后，以姓氏拼音排序）

本书将看似“浮云”的数据分析领域，蕴于商业化的场景之中，生动形象地让读者了解到“给力”的数据分析师是如何炼成的！引导非专业人士从数据的角度，认识、剖析、解决商业问题；对专业人士而言，亦能提供一次梳理和提高的学习机会。

邓凯

数据挖掘与数据分析博主，资深数据分析师

这是一本适合普通大众的“专业”数据分析书，由浅入深，富有体系。既有一口气读完的冲动，又想马上找一台电脑试一试这些“新奇”的分析方法，更想拿一些数据来分析找找其中的规律。

读完本书，你会发现数据分析的乐趣，它并不是那么枯燥，数据背后的故事简直是太有意思了。从此你将发现：无论是新闻媒体，还是企业报表中的数字，都将不再孤独，因为它们在那里，在和你说着话！

祝愿大家早日练就一颗数据分析的“心”！

黄成明

数据化管理顾问及培训师，零售及服装企业数据化管理咨询顾问

SPSS 等统计软件的应用是以统计学知识为基础的，而现实是我们的“数据分析人员”，往往不具备统计学基础知识和系统的研究训练。因此大家在应用统计软件解决问题时，哪怕是一个小问题，也会觉得无从入手，并在具体的数据处理和统计分析过程中，处处一头雾水，心里没底。

随着大数据时代的到来，我们最迫切需要的倒不是 IT 行业所说的“大数据”，而是在利用好现有数据的条件下，能够掌握统计分析利器进行敏捷深刻的研究思考。

我非常喜欢《谁说菜鸟不会数据分析》系列书籍，“菜鸟”系列的长篇“小说”我都是一口气读完的，享受了在阅读过程中和作者的思路同步的趣味盎然，这本书同样如此。强烈推荐这本 SPSS 统计分析软件的入门应用书籍，祝愿大家都和小白一起学有所成。

马广斌，博士

北京数海时代分析技术有限公司 总经理

原 析数软件（SPSS China）统计服务事业部 总经理



当谈到用数据解决问题时，我经常用这样的语言去诠释：“如果你不能量化它，你就不能理解它，如果不理解就不能控制它，不能控制也就不能改变它”。数据无处不在，信息时代最主要的特征就是“数据处理”，数据分析正以我们从未想象过的方式影响着日常生活。

在知识与信息技术时代，每个人都面临着如何有效地吸收、理解和利用信息的挑战。那些能够有效利用工具从数据中提炼信息、发现知识的人，最终往往成为各行各业的强者！

这本书向我们清晰又友好地介绍了数据分析方法、技巧与工具，强烈推荐读一读这本书，或许会给你带来更大的惊喜！

沈浩

中国传媒大学电视与新闻学院，教授

调查统计研究所，副所长

数据挖掘研发中心，主任

IPSOS 公司，首席技术顾问

数据分析理论、公式和方法对部分初学者来说是枯燥、乏味的，或陷入云山雾罩中不得其道。本书最大的特点是使用幽默风趣的语言，结合工作中典型案例加以分析、解读，是一本数据分析工作者值得一读的好书。

石军

安徽同微信息技术有限公司，总经理

数据分析是一种能力，更是一种思想。此书结构有层次、内容全面、通俗易懂，通过 SPSS 工具一步步带你走进数据分析的世界，探索数据分析的价值，让数据分析变得既简单又有趣。

郑来秋

数据分析网创始人，某知名互联网公司数据分析专家

# 目 录

## 第1章 SPSS 概况 / 11

- 1.1 SPSS 简介 / 12
- 1.2 SPSS 特点 / 13
- 1.3 SPSS 安装 / 15
- 1.4 SPSS 窗口 / 19
- 1.5 本章小结 / 22

## 第2章 数据处理 / 23

- 2.1 数据变量 / 24
  - 2.1.1 数据类型 / 24
  - 2.1.2 变量尺度 / 25
- 2.2 数据导入 / 27
  - 2.2.1 Excel 数据导入 / 27
  - 2.2.2 文本数据导入 / 29
- 2.3 数据清洗 / 33
- 2.4 数据抽取 / 35
  - 2.4.1 字段拆分 / 35
  - 2.4.2 随机抽样 / 38
- 2.5 数据合并 / 40
  - 2.5.1 字段合并 / 40
  - 2.5.2 记录合并 / 41
- 2.6 数据分组 / 43
  - 2.6.1 可视分箱 / 43
  - 2.6.2 重新编码 / 46
- 2.7 数据标准化 / 48
  - 2.7.1 0-1 标准化 / 48
  - 2.7.2 Z 标准化 / 50
- 2.8 本章小结 / 50

## 第3章 描述性分析 / 53

### 3.1 频率分析 / 54

#### 3.1.1 分类变量频率分析 / 54

#### 3.1.2 连续变量频率分析 / 57

### 3.2 描述分析 / 61

### 3.3 交叉表分析 / 63

### 3.4 多选题定义 / 65

### 3.5 数据报表制作 / 68

#### 3.5.1 报表类型简介 / 69

#### 3.5.2 分类变量报表制作 / 70

#### 3.5.3 连续变量报表制作 / 72

#### 3.5.4 多选题报表制作 / 73

#### 3.5.5 报表灵活运用 / 75

### 3.6 本章小结 / 80

## 第4章 相关分析 / 81

### 4.1 相关分析简介 / 82

### 4.2 相关分析实践 / 84

#### 4.2.1 散点图绘制 / 85

#### 4.2.2 相关分析操作 / 86

### 4.3 本章小结 / 87

## 第5章 回归分析 / 89

### 5.1 回归分析简介 / 90

#### 5.1.1 什么是回归分析 / 90

#### 5.1.2 线性回归分析步骤 / 91

### 5.2 简单线性回归分析 / 92

#### 5.2.1 简单线性回归分析简介 / 92

#### 5.2.2 简单线性回归分析实践 / 93

### 5.3 多重线性回归分析 / 99

#### 5.3.1 多重线性回归分析简介 / 99

#### 5.3.2 多重线性回归分析实践 / 99

## 5.4 本章小结 / 106

## 第6章 自动线性建模 / 107

### 6.1 自动建模 / 108

### 6.2 模型结果解读 / 113

### 6.3 模型预测 / 121

### 6.4 本章小结 / 122

## 第7章 Logistic 回归 / 123

### 7.1 Logistic 回归简介 / 124

### 7.2 Logistic 回归实践 / 127

#### 7.2.1 Logistic 回归操作 / 128

#### 7.2.2 Logistic 回归结果解读 / 129

#### 7.2.3 Logistic 回归预测 / 131

### 7.3 本章小结 / 135

## 第8章 时间序列分析 / 137

### 8.1 时间序列分析简介 / 138

### 8.2 季节分解法 / 139

### 8.3 专家建模法 / 148

#### 8.3.1 时间序列预测步骤 / 148

#### 8.3.2 时间序列分析操作 / 149

#### 8.3.3 时间序列分析结果解读 / 151

#### 8.3.4 时间序列预测应用 / 153

### 8.4 本章小结 / 157

## 第9章 RFM 分析 / 159

### 9.1 RFM 分析介绍 / 160

### 9.2 RFM 分析操作 / 162

#### 9.2.1 数据准备 / 162

#### 9.2.2 RFM 分析实践 / 163

#### 9.2.3 RFM 分析结果解读 / 167

### 9.3 RFM 分析应用 / 170

## 9.4 本章小结 / 175

### 第10章 聚类分析 / 177

#### 10.1 聚类分析介绍 / 178

#### 10.2 快速聚类分析 / 180

##### 10.2.1 快速聚类分析操作 / 180

##### 10.2.2 快速聚类分析结果解读 / 182

#### 10.3 系统聚类分析 / 186

##### 10.3.1 系统聚类分析操作 / 186

##### 10.3.2 系统聚类分析结果解读 / 189

#### 10.4 二阶聚类分析 / 193

##### 10.4.1 二阶聚类分析操作 / 193

##### 10.4.2 二阶聚类分析结果解读 / 195

#### 10.5 聚类方法的对比 / 201

#### 10.6 本章小结 / 202

### 第11章 因子分析 / 203

#### 11.1 因子分析简介 / 204

#### 11.2 因子分析实践 / 206

##### 11.2.1 因子分析操作 / 207

##### 11.2.2 因子分析结果解读 / 210

#### 11.3 本章小结 / 217

### 第12章 对应分析 / 219

#### 12.1 对应分析简介 / 220

#### 12.2 对应分析实践 / 221

##### 12.2.1 对应分析操作 / 221

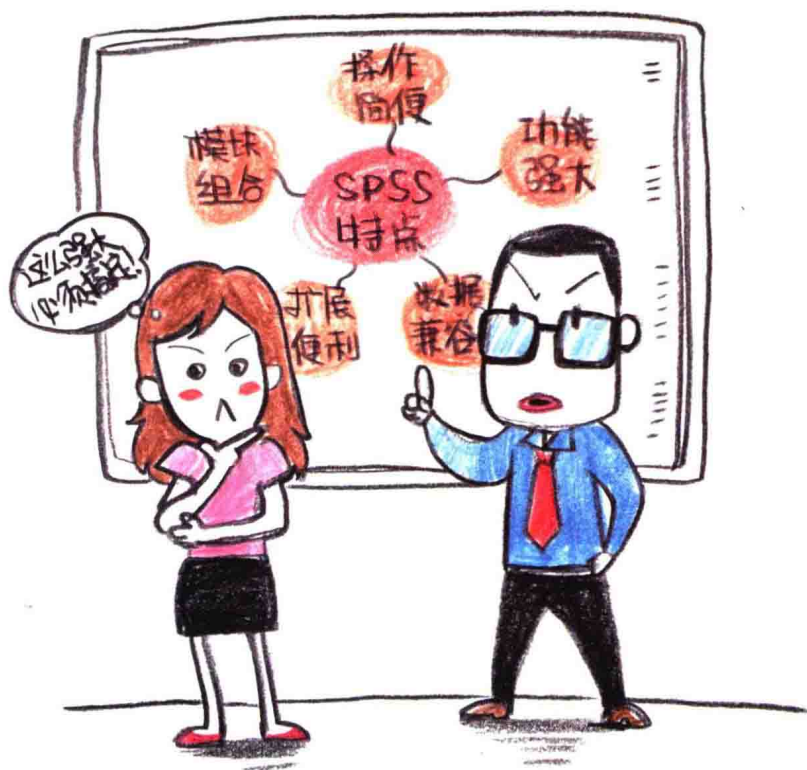
##### 12.2.2 对应分析结果解读 / 225

#### 12.3 本章小结 / 228



# 第1章

## SPSS 概况



**Mr. 林一大早来到办公室，就对小白说：**小白，现在公司业务不断在发展扩大，需要更多的数据分析支持，牛董已经批准了我们的招聘计划，你现在拟个招聘要求，然后发给 HR 同事，让他们协助招聘 3 名数据分析师。

**小白应声道：**好嘞！有什么要求？

**Mr. 林想了想：**要求主要有以下几点：

- （1）统计学、数学或计算机等相关专业，本科及以上学历；
- （2）熟练使用 SQL、Excel、PPT 等常用工具；
- （3）拥有良好的逻辑分析与独立思考能力；
- （4）熟练使用 SPSS 等统计分析工具者优先。

**小白问：**SPSS？目前我们还没怎么用到这个工具呀！

**Mr. 林笑着说：**目前我们的数据分析基础工作已经理顺，Excel 能满足我们的大部分分析工作。随着公司业务不断发展扩大，牛董也对我们的分析工作提出了更高的要求，有的问题需要用 SPSS 等更为专业的统计软件才能解决，所以我们要不断提升专业能力，以满足业务分析的需求。

**小白：**那我现在是不是也要开始学习 SPSS 了？如何学呢？

**Mr. 林：**这样吧，每天下班后，抽些时间向你介绍些 SPSS 相关知识与使用。

**小白兴奋的叫道：**太好了！Mr. 林你好帅啊！

**Mr. 林笑道：**哈哈！赶紧去处理招聘数据分析师的工作吧。另外，今晚就开始学习 SPSS。

**小白：**Yes Sir！

## 1.1 SPSS 简介

下班后，小白如约来到 Mr. 林办公桌前，迫不及待地**说：**Mr. 林，可以开始了吗？

**Mr. 林：**小白真准时，现在我们就先来了解一下什么是 SPSS。

SPSS 是广大统计爱好者和数据分析师最熟悉的名字，它是一款在市场研究、医学统计、政府和企业的数据分析应用中久享盛名的统计分析工具。

小白，知道什么叫 SPSS 吗？

**小白摇了摇头：**不知道，为什么呢？

**Mr. 林：**SPSS 是由美国斯坦福大学三位研究生于 1968 年一起开发的一个统计软件包，SPSS 是该软件英文名称的首字母缩写，原意为“Statistical Package for the Social Sciences”，即“社会科学统计软件包”。

2000 年，随着 SPSS 公司产品服务领域的扩大和服务深度的增加，SPSS 公司整个产品线的名称都进行了调整，现在 SPSS 软件的名称全称为“Statistical Product and Service

Solutions”，意为“统计产品与服务解决方案”，而英文缩写 SPSS 没有改变。

2009 年，SPSS 公司宣布重新包装旗下的 SPSS 产品线，定位为预测统计分析软件 PASW（Predictive Analytics Software），用户对这个名称难以接受。

2010 年，随着 SPSS 公司被 IBM 公司并购，软件也相应地更名为 IBM SPSS Statistics。

现在，SPSS 旗下主要有 4 个产品。

- ★ IBM SPSS Statistics（原 SPSS）：统计分析产品；
- ★ IBM SPSS Modeler（原 Clementine）：数据挖掘产品；
- ★ IBM SPSS Data Collection（原 Dimensions）：数据采集产品；
- ★ IBM SPSS Decision Management（原 Predictive Enterprise Services）：企业应用服务。

我们常说的 SPSS，指的是 IBM SPSS Statistics，后续的介绍同样采用简称 SPSS。

小白：好的，原来 SPSS 还有如此曲折的故事呀。

## 1.2 SPSS 特点

小白继续问道：那 SPSS 有何过人之处让它拥有这么多粉丝？

Mr. 林：问得好，我们一起来看一下 SPSS 的五大特点，如图 1-1 所示。



图 1-1 SPSS 五大特点

### 1. 操作简便

SPSS 的操作界面友好、简便，类似于熟悉的 Windows 风格界面，数据视图也类似于 Excel 布局。对于各种统计方法的使用，只要了解统计分析的基本原理，无须通晓统计方法的各种算法，无须编程，大多数分析可通过“菜单”、“对话框”操作来完成，即可得到所需要的统计分析结果，非统计专业人士也能快速上手。

### 2. 功能强大

SPSS 非常全面地涵盖了数据分析主要操作流程，提供了数据获取、数据处理、数据分析、

数据展现等数据分析操作。其中 SPSS 涵盖了各种统计方法与模型，从简单的描述统计分析方法到复杂的多因素统计分析方法，例如数据的描述性分析、相关分析、方差分析、回归分析、Logistic 回归、聚类分析、判别分析、因子分析、对应分析等，应有尽有。

3. 数据兼容

SPSS 能够导入及导出多种格式的数据文件或结果。例如，SPSS 可导入文本、Excel、Access、SAS、Stata 等数据文件，SPSS 还能够把其表格、图形结果直接导出为 Word、Excel、PowerPoint、txt 文本、pdf、html 等格式文件。

4. 扩展便利

SPSS 可以调用 R 语言的各种统计包或 Python 的功能模块，实现最新统计方法的调用，增强 SPSS 的扩展性。

5. 模块组合

SPSS 是一个综合性的产品家族，它为各分析阶段提供了丰富的模块功能。SPSS Statistics Base 是基础的软件平台，具备强大的数据管理能力、输入输出界面管理能力，以及完备的常见统计分析功能。其他每个独立扩充功能模块均在 SPSS Statistics Base 的基础上，为其增加某方面的分析功能。用户可以根据自己的分析需要及计算机配置灵活选择组合使用。

根据 SPSS 模块功能的不同，可以将 SPSS 常用模块大致划分为四个分析阶段：数据处理、描述性分析、推断性分析和探索性分析，各分析阶段对应的具体模块如图 1-2 所示。

分析阶段	模块	功能
数据处理	Data Preparation	提供数据校验、清理等数据处理工具
	Missing Values	提供缺失数据的处理与分析
	Complex Samples	提供多阶段复杂抽样技术
描述性分析	Statistics Base	提供最常用的数据处理、统计分析
	Custom Tables	提供创建交互式分析报表功能
推断性分析	Advanced Statistics	提供强大且复杂的单变量和多变量分析技术
	Regression	提供线性、非线性回归分析技术
	Forecasting	提供ARIMA、指数平滑等时间序列模型
探索性分析	Categories	提供针对分类数据的分析工具
	Conjoint	提供联合分析市场研究工具
	Direct Marketing	提供直销活动效果分析工具
	Decision Trees	提供分类决策树模型分析方法
	Neural Networks	提供神经网络模型分析方法

图 1-2 SPSS 常用模块及功能说明



**小白好奇地说：**那 SPSS 就没有一些不足之处吗？

**Mr. 林：**当然有，每个工具都有自己的优势与不足，SPSS 也不例外，它最主要的不足之处是其输出结果不方便直接用于我们的数据分析报告中。虽然 SPSS 可以直接导出为 txt、doc、ppt、xls 等文档格式，但通常与我们的数据分析报告风格、要求不符，需要我们再次进行加工整理。

**小白追问道：**这是为何呢？

**Mr. 林：**因为我们写的报告需要风格统一，包括字体、颜色等，虽然现在 SPSS 的输出结果相对美观，但是与我们的报告风格不统一，把它们放在一起，会给人东拼西凑的感觉。

**小白附和着说道：**用现在流行的话说，就是一点也不高端、大气、上档次。

**Mr. 林：**没错，我们可以将 SPSS 分析结果复制、粘贴至 Excel 中，重新绘制与分析报告风格统一的图表，再用于分析报告中，所以后面也不再单独介绍 SPSS 图表功能的使用了。

**小白：**好的，Excel 绘制图表就比较灵活、方便。

**Mr. 林：**所以我们应该根据分析需求及各种分析工具的特点选择合适的分析工具，只要能高效解决问题就是好工具，不用刻意追求使用 SPSS、SAS 等高级分析工具，只在需要的时候使用，避免出现杀鸡用牛刀的情况。

**小白点了点头：**明白。

## 1.3 SPSS 安装

**Mr. 林：**SPSS 支持 Windows、Linux 和 Mac OS 操作系统，后续我们将以 Windows 7 + IBM SPSS Statistics 23 组合进行 SPSS 安装及使用的介绍。

**小白：**那对计算机的配置有要求吗？

**Mr. 林：**随着 Windows 版本的不断升级，操作系统运行所需要的内存容量不断增加。在 SPSS 运行的同时，计算机也在运行一些其他应用软件，为保证计算机的运行速度和各应用软件的功能正常实现，建议计算机 CPU 处理频率大于等于 1GHz，内存大于等于 1GB。

现在我们就一起来安装 SPSS。

**STEP 01** 打开 SPSS 安装程序所在文件夹，用鼠标双击 SPSS 安装程序，打开“安装向导”界面，系统自动进行 SPSS 安装文件的解压缩，解压缩完成后进入 SPSS 安装向导授权说明界面，单击【下一步】按钮，如图 1-3 所示。

**STEP 02** 根据自身情况选择许可证类型，此处默认选择【单个用户许可证】，单击【下一步】按钮，如图 1-4 所示。

**STEP 03** 在软件许可协议对话框中，选择【我接受许可协议中的全部条款】，单击【下一步】按钮，如图 1-5 所示。



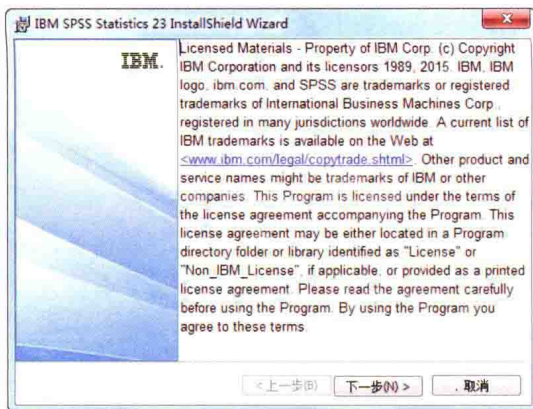


图 1-3 SPSS 安装向导 1

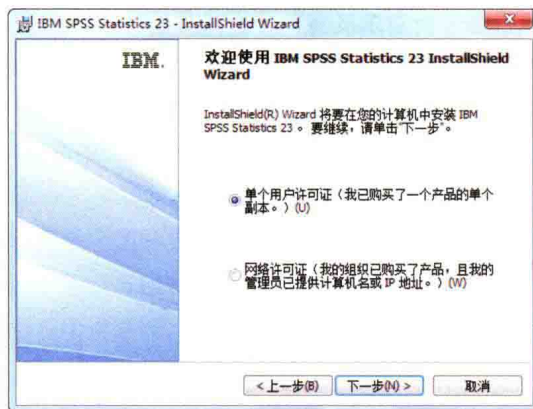


图 1-4 SPSS 安装向导 2

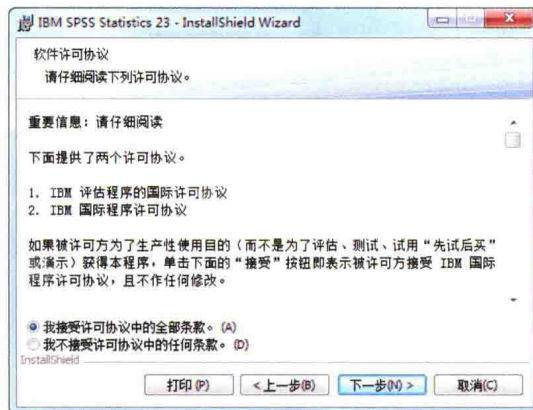


图 1-5 SPSS 安装向导 3

**STEP 04** 根据自身情况填写用户姓名、单位信息，单击【下一步】按钮，如图 1-6 所示。



图 1-6 SPSS 安装向导 4

**STEP 05** 系统会自动安装英语帮助，也可以选择其他帮助语言，此处我们选择【简体中文】，单击【下一步】按钮，如图 1-7 所示。

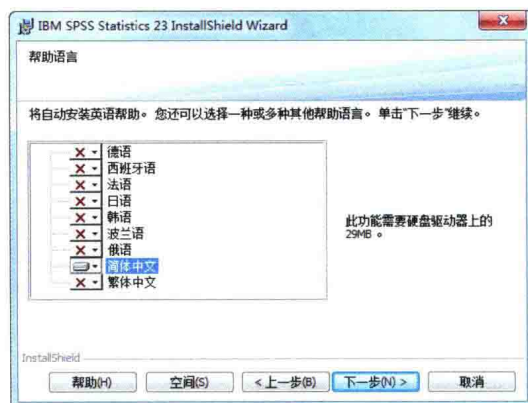


图 1-7 SPSS 安装向导 5

**STEP 06** 在辅助技术对话框中，可根据自己的需要选择是否启用 JAWS for Windows 屏幕阅读软件，此处我们选择【否】，单击【下一步】按钮，如图 1-8 所示。

**STEP 07** 在 IBM SPSS Statistics - Essentials for Python 对话框中，可根据自己的需要选择是否安装 IBM SPSS Statistics - Essentials for Python，此处我们选择【否】，单击【下一步】按钮，如图 1-9 所示。

**STEP 08** 根据自己的计算机情况或需求，选择 SPSS 安装路径，如果 C 盘空间有限，可更改安装到其他空间相对充足的硬盘上，此处默认采用 C 盘安装路径，单击【下一步】按钮，如图 1-10 所示。

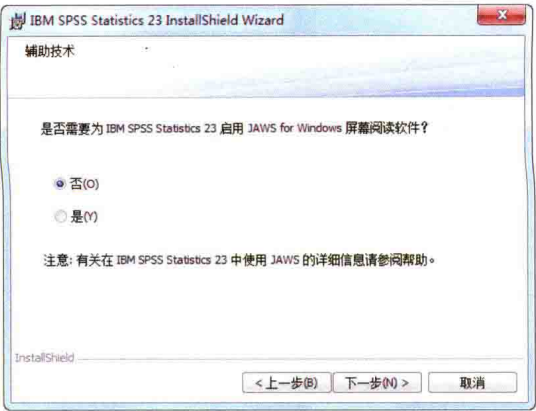


图 1-8 SPSS 安装向导 6



图 1-9 SPSS 安装向导 7



图 1-10 SPSS 安装向导 8

## 12.2.2 对应分析结果解读

Mr. 林：SPSS 的对应分析提供了以下输出结果。

第一个输出结果是“对应表”，如图 12-8 所示。该表再现交叉表结果，其数据和 Excel 里面的交叉表数据完全一致。另外，最后一行和列的“活动边际”是分别计算了每行和每列的总和。

对应表								
形象	品牌							活动边际
	L	H	D	A	F	B	T	
技术领先、经常创新	85	79	97	52	38	42	94	487
性价比高	85	88	76	76	72	68	48	513
质量好、返修率低	88	72	70	54	38	33	72	427
外观时尚	62	74	63	60	47	57	57	420
知名度高	115	101	86	58	58	49	75	542
外观稳重大气	52	65	51	52	57	46	60	383
维修响应及时	59	58	56	42	56	37	38	346
交货及时	72	70	65	56	45	62	47	417
一次性解决故障率高	51	47	47	35	29	25	43	277
店面分布广泛	110	67	56	29	45	32	37	376
口碑好	84	79	65	37	27	17	64	373
活动边际	863	800	732	551	512	468	635	4561

图 12-8 对应分析输出结果（1）：对应表

第二个输出结果是“摘要表”，如图 12-9 所示。该表输出了对应分析的统计量结果以及累积百分比。

摘要								
维	奇异值	惯量	卡方	显著性	惯量比例		置信度奇异值	
					占	累积	标准差	相关性
								2
1	.145	.021			.575	.575	.014	.025
2	.102	.010			.289	.864	.015	
3	.051	.003			.071	.935		
4	.034	.001			.032	.967		
5	.027	.001			.020	.987		
6	.022	.000			.013	1.000		
总计		.036	165.607	.000 <sup>a</sup>	1.000	1.000		

a. 60 自由度

图 12-9 对应分析输出结果（2）：摘要表

小白：我需要重点关注哪个指标呢？

Mr. 林：只需要重点关注对应分析图的解释能力，也就是累积惯量比例就行。之前也介绍过，对应分析实际上是一种“降维”方法，所以，必然会存在信息的损失。对于一个二维平面图形来说，其解释能力的大小在很大程度上影响了结果的解读。所以，我们需要重点关注前两个维度的累积惯量比例。

在通常情况下，前两个维度的累积惯量比例达到 80% 及以上，就说明对应分析图的效果非常好，是具有可读性的。在这个案例中，对应分析图的累积惯量比例达到了 86.4%，说明对应分析图的效果非常好。

第三个输出结果是“行 / 列点总览表”，如图 12-10 所示。这个结果主要提供了各类别在各维度上的得分，它们也是对应分析图中横、纵轴坐标对应的数值，以及行列对维度互相之间的贡献大小。所以，本例中的对应分析图就是由行 / 列点总览表两个维度的得分绘制出来的。

行点总览 <sup>a</sup>									
形象	数量	维得分		惯量	贡献				
		1	2		点对维的惯量		维对点的惯量		总计
					1	2	1	2	
技术领先·经常创新	.107	.368	.468	.005	.100	.229	.417	.479	.896
性价比高	.112	-.503	-.151	.005	.197	.025	.894	.057	.951
质量好·返修率低	.094	.306	.153	.002	.061	.021	.700	.123	.823
外观时尚	.092	-.342	.221	.002	.074	.044	.686	.202	.888
知名度高	.119	.173	-.141	.001	.025	.023	.639	.299	.938
外观稳重大气	.084	-.378	.247	.003	.083	.050	.531	.161	.691
维修响应及时	.076	-.319	-.192	.002	.054	.027	.483	.124	.607
交货及时	.091	-.361	-.010	.003	.082	.000	.642	.000	.642
一次性解决故障率高	.061	.107	.153	.000	.005	.014	.333	.486	.819
店面干净整洁	.082	.319	-.838	.007	.058	.565	.167	.811	.978
口碑好	.082	.680	.044	.006	.262	.002	.917	.003	.920
活动总计	1.000			.036	1.000	1.000			

a. 对称正态化

列点总览 <sup>a</sup>									
品牌	数量	维得分		惯量	贡献				
		1	2		点对维的惯量		维对点的惯量		总计
					1	2	1	2	
L	.189	.380	-.475	.009	.189	.417	.464	.512	.976
H	.175	.094	-.054	.001	.011	.005	.242	.057	.299
D	.160	.177	.144	.002	.035	.032	.411	.191	.602
A	.121	-.353	.232	.003	.104	.064	.644	.197	.841
F	.112	-.492	-.272	.006	.188	.081	.625	.135	.760
B	.103	-.691	.032	.008	.339	.001	.876	.001	.877
T	.139	.372	.542	.007	.134	.400	.381	.572	.953
活动总分	1.000			.036	1.000	1.000			

a. 对称正态化

图 12-10 对应分析输出结果（3）：行 / 列点总览表



第四个输出结果是最重要的——“对应分析图”，如图 12-11 所示，图中各类别散点在图形中的距离和位置反映了它们各自之间的关系：

- ★ 在同一维度上，例如横轴上，同一变量的类别距离越近，说明在这个维度上差异越小。例如，“外观稳重大气”和“外观时尚”对于参与品牌形象评价的用户来说，差异相对较小；
- ★ 在对应分析图中，不同变量散点之间的距离越近，说明它们的相关性越大。例如，“品牌 T”和“技术领先，经常创新”距离很靠近，说明用户认为品牌 T 属于技术领先、经常创新的品牌。

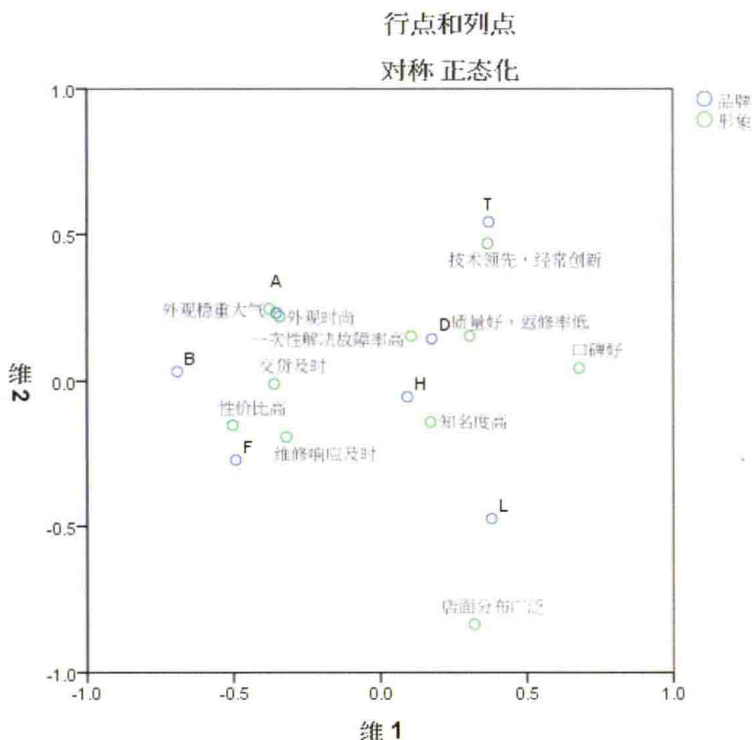


图 12-11 对应分析输出结果（4）：对应分析图

**小白：**现在我明白怎么解读这个对应分析图了。

首先要评估对应分析图的解释能力，如果前两个维度的累积惯量比例不低于 80%，则说明该图的效果非常好，然后从不同方面来解读变量之间的关系。

**Mr. 林：**没错，看来你已经掌握了。那么，我再问你一个问题，如果需要自己在 Excel 中绘制这个对应分析图，要怎么做呢？

**小白：**我刚才也想问这个问题，SPSS 的图形输出不便于直接粘贴到报告或演示文档中，如

果需要自己在 Excel 中绘制图形，我会把图 12-10 所示的行 / 列点总览表复制到 Excel 中，然后使用其中的“维得分”作为每个变量的横、纵轴数据绘制散点图。

**Mr. 林：**孺子可教也，对应分析就介绍到这里。

## 12.3 本章小结

**Mr. 林：**现在一起来回顾一下今天所学的内容：

- ★ 了解什么是对应分析，以及其原理与步骤、优劣势、应用领域；
- ★ 了解 SPSS 中如何进行对应分析操作；
- ★ 了解如何解读 SPSS 对应分析结果。

SPSS 的主要功能使用就介绍到这里，SPSS 的强大功能想必你已经深有体会，方法与工具再高级，目的都是为了解决业务问题，而不是为了追求高大上才使用。还是那句话，只要能快速、有效地解决业务问题，就是好方法、好工具，希望你在未来能够正确、灵活地运用这些数据分析方法与工具，快速、有效地解决实际业务问题。

**小白：**感谢恩师教诲，小白在此鞠躬以示谢意，等一下，我请你去吃大餐，就当是我的 SPSS 谢师宴啦。



## 专家评论

当谈到用数据解决问题时，我经常用这样的语言去诠释：“如果你不能量化它，你就不能理解它，如果不理解就不能控制它，不能控制也就不能改变它”。数据无处不在，信息时代的最主要特征就是“数据处理”，数据分析正以我们从未想象过的方式影响着日常生活。

在知识经济与信息技术时代，每个人都面临着如何有效地吸收、理解和利用信息的挑战。那些能够有效利用工具从数据中提炼信息、发现知识的人，最终往往成为各行各业的强者！

这本书向我们清晰又友好地介绍了数据分析方法、技巧与工具，强烈推荐读一读这本书，它或许会给你带来意想不到的收获！

沈浩

中国传媒大学电视与新闻学院，教授；调查统计研究所，副所长  
数据挖掘研发中心，主任；SPSSOS公司，首席技术顾问

SPSS等统计软件的应用是以统计学知识为基础的，而现实是我们的“数据分析人员”，往往不具备统计学基础知识和系统的研究训练。因此大家在应用统计软件解决问题时，哪怕是一个小问题，都会觉得无从入手，并在具体的数据处理和统计分析过程中，处处一头雾水，心里没底。

随着大数据时代的到来，我们最迫切需要的倒不是IT行业所说的“大数据”，而是在利用好现有数据的条件下，能够掌握统计分析利器进行敏捷、深刻的研究思考。

我非常喜欢《谁说菜鸟不会数据分析》系列书籍，“菜鸟”系列的长篇“小说”我都是一口气读完的，享受了在阅读过程中和作者的思路同步的趣味盎然，这本书同样如此。强烈推荐这本SPSS统计分析软件的入门应用书籍，祝愿大家都和小白一起学有所成。

马广斌 博士

北京数海时代分析技术有限公司 总经理  
析数软件（SPSS China）统计服务事业部 原总经理



博文视点Broadview



@博文视点Broadview



策划编辑：张月萍  
责任编辑：葛娜  
封面设计：李玲

上架建议：计算机 / 数据分析

ISBN 978-7-121-28801-2



9 787121 288012 >

定价：59.00元