

谁说菜鸟不会 数据分析

狄松 祝迎春 张文霖 马世澎 编著

SPSS
统计分析



谁说菜鸟不会 数据分析

狄松 祝迎春 张文霖 马世澎 编著

SDSS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. ①TP391.13 ②C819

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 zllts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: (010) 51260888-819, 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

SDSS
(SDSS 第 1 卷)

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, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: (010) 51260888-819, 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自己的学习机会。

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

>> 谁说菜鸟不会数据分析（SPSS 篇）

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