

SPSS and Statistical Thinking

新时代
技术
新未来

大数据方法论风靡全球
统计学应用更上一层楼

SPSS实战 与统计思维

武松 编著

附赠 全书实例完整数据集
和SPSS入门学习视频

清华大学出版社

SPSS and Statistical Thinking

新时代
技术
新未来



SPSS实战 与统计思维

武松 编著

清华大学出版社
北京

内 容 简 介

本书从统计学思维开始,由浅入深,全面系统地讲解了SPSS实战应用。本书涉及面广,从软件基本操作到高级统计分析技术,涉及SPSS目前绝大部分的应用范畴。本书涵盖SPSS概述、数据管理、统计描述分析、基本统计分析的报表制作、*t*检验、方差分析、Logistic回归、中介效应与调节效应分析以及大量的专项统计方法。本书以案例式教学为特色,书中提供了大量的应用案例,供读者实战演练。

本书不仅适合有一定统计基础的人员阅读,也适合SPSS初学者。通信、金融、制造、医药、教育科研、市场调研、连锁零售和电子商务等行业的数据分析人员,可将本书作为一本易学易练的案头参考书。信息技术、心理学、经济管理等专业的大中专院校的学生和教师,可将本书作为一本教材使用。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

SPSS实战与统计思维 / 武松编著. — 北京:清华大学出版社, 2019
(新时代·技术新未来)
ISBN 978-7-302-51322-3

I. ①S… II. ①武… III. ①统计分析—统计程序 IV. ①C819

中国版本图书馆 CIP 数据核字(2018)第 227132 号

责任编辑:刘 洋
封面设计:徐 超
版式设计:方加青
责任校对:宋玉莲
责任印制:宋 林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印装者:三河市铭诚印务有限公司

经 销:全国新华书店

开 本:187mm×235mm 印 张:26.5 字 数:510千字

版 次:2019年1月第1版 印 次:2019年1月第1次印刷

定 价:99.00元

产品编号:079974-01

作者简介

武松（松哥统计），安徽中医药大学副教授，中国疾病预防控制中心流行病与卫生统计学博士，世界中联临床科研统计学会理事，国家高级统计分析师，SPSS 高级数据分析师，微信公众号精鼎统计（data973）创始人。擅长 SPSS、SPSS-Modeler、SAS、Stata、Revman 等多种统计软件，国内多家杂志统计专业审稿人员，目前主持课题 8 项，协作子课题 11 项，出版 SPSS 专著 1 部，另外以副主编的身份参与编写了 8 部图书，参与过“十一五”“十二五”“国家自然科学基金”“卫生部专项基金”等百余项课题数据分析。在国家级刊物以第一作者或通讯作者发表文章 40 余篇，其中 SCI/Medline 收录 5 篇，获得国家发明专利 1 项，获得上海市出入境检验检疫局科技兴检三等奖 1 项。近年致力于数据分析与挖掘技术的研究与推广，举办 SPSS 数据培训近百场，培训风格幽默风趣，化复杂难懂的统计于举例与比喻之中，是业界最受欢迎的 SPSS 统计讲师之一。主编的《SPSS 统计分析大全》一书累计印刷 15 次，销量 4 万多册，雄踞多个图书销售网站统计软件类图书榜首。

前言

终于鼓起勇气，准备独自撰写一本 SPSS 实用教程。教书 18 载，看见太多的学生、教师及相关科研工作者为统计分析而发愁。我也经常纳闷，咱们本科生、研究生及博士生阶段都学习了统计，为什么一到用时就惊慌失措呢？这其中必然有教师和学生两方面的原因，教师应付教学，学生应付考试，但归根结底还是统计老师的原因。

虽然古人告诉我们“不仅要知其然，还要知其所以然”，但是松哥对此表示不同意，在科技发展如此迅猛的今天，我们能够知其然已属不易，知其所以然已不可能。比如手机已经成为我们生活的一部分，咱们会用就可以了，何必追究其工作原理，除非通信就是你的研究专业。对于统计学习，松哥一直提倡实战主义，因为需要数据分析的绝大多数用户都是非统计专业人士，过多的学习和掌握统计学的公式推导，就已将非统计专业人士的精力消耗殆尽，面对后续分析的学习，已经没有继续的勇气。所以对于初学者，如果能用简明形象的语言传授其统计的思想，然后以实际案例进行分析与解读，其效果将远胜于前者。正所谓：为伊消得人憔悴，一用就错学不会；舍得应付换应用，案例实战真英雄。

鉴于此，松哥立意写一本没有统计公式的统计分析与 SPSS 实战教程，一些统计思想与理论尽量用一些白话方式进行表达。思想阐明以后，就以实际案例进行案例分析，SPSS 实战操作与结果的详细解读。其实松哥在 2014 年曾出过一本《SPSS 统计分析大全》，销量很好，出版社已经多次加印，但回头看还有很多地方不太完善，甚至今天松哥对当时的写作思路都开始否定。于是决定，重新撰写一本《SPSS 实战与统计思维》，这是一个疯狂的、自虐式的决定，松哥给自己一年的时间，将自己的经验与心得呈现在读者的面前，松哥相信这会是一本值得收藏的好书！

本书分为六个篇章进行讲述，第一篇：统计思维；第二篇：SPSS 数据库构建与数据管理；第三篇：初级统计说一说（描述性统计）；第四篇：中级统计比一比（差异性分析）；
试读结束，需要全本请在线购买：www.ertongbook.com

第五篇：高级找关系（统计模型）；第六篇：专项统计。本书定位是一本初学者的入门宝典，中级用户的提升指南，高级用户的拓展手册。本书发行后会继续征集读者意见，修订完善，最终成为一本 SPSS 学习的红宝书！

本书得到安徽省教育厅重点研究项目（2015jyxm186）、安徽中医药大学教学研究课题（NO.YB201012）和（2013xjzc012）资助，特此感谢。

写书是无比艰辛的，给自己一个目标，在岁月的河流中，留下活过的印记！立字为据！

松哥统计

2017年5月21日

于安徽合肥

目录

第一篇 统计思维

第 1 章 核心统计概念	002
1.1 总体与样本 (population and sample)	002
1.1.1 总体	002
1.1.2 样本	002
1.2 参数与统计量 (parameter and statistics)	002
1.3 概率与频率 (probability and frequency)	003
1.3.1 概率	003
1.3.2 频率	004
1.4 误差 (error)	004
1.5 同质与变异 (homogeneity and variation)	006
1.5.1 同质	006
1.5.2 变异	006
1.6 随机化原则 (random principle)	006
1.7 因素与水平 (factor and level)	007
1.7.1 因素	007
1.7.2 水平	007
1.8 变量 (variable)	007
1.8.1 计量变量	008
1.8.2 计数变量	008
1.8.3 等级变量	008

1.8.4 变量之间相互转换	008
第 2 章 常见统计设计	009
2.1 成组设计	009
2.2 配比设计	009
2.2.1 配对设计	010
2.2.2 配伍组设计	010
2.3 析因设计	010
2.4 重复测量设计	011
2.5 其他设计	012
第 3 章 统计思想	013
3.1 抽样的思想	013
3.2 总体推断思想	013
3.3 反证法思想	014
3.4 小概率思想	014
3.5 误差控制思想	015

第二篇 SPSS 数据库构建与数据管理

第 4 章 SPSS 简介与数据库构建	018
4.1 SPSS 简介	018
4.2 SPSS 24.0 安装	018
4.2.1 确定计算机系统	019
4.2.2 安装步骤	019
4.3 SPSS 启动与退出	022
4.4 SPSS 24.0 窗口简介	022
4.4.1 数据编辑窗口	022
4.4.2 变量编辑窗口	023
4.4.3 结果输出窗口	024

4.5	SPSS 数据库构建	024
4.5.1	间接法数据库构建	025
4.5.2	直接法数据库构建	027
4.5.3	数据编辑	029
4.6	SPSS 分析过程三级窗口	030
4.6.1	一级窗口 (数据窗口)	030
4.6.2	二级窗口 (功能窗口)	031
4.6.3	三级窗口 (参数窗口)	031
第 5 章 SPSS 24.0 数据管理		032
5.1	文件、编辑、查看菜单介绍	032
5.1.1	文件 (常用5个)	033
5.1.2	编辑 (常用5个)	034
5.1.3	查看	035
5.2	10 项常用数据管理功能	036
5.2.1	排序个案 (sort cases) (☆)	036
5.2.2	转置文件 (transpose)	037
5.2.3	合并文件 (merge files)	039
5.2.4	拆分文件 (split files)	043
5.2.5	选择个案 (select cases) (☆)	045
5.2.6	加权个案 (weight cases) (☆)	049
5.2.7	计算变量 (compute)	051
5.2.8	重新编码 (recode)	051
5.2.9	自动重新编码	053
5.2.10	缺失值替换 (replace missing value)	055

第三篇 初级统计说一说 (描述性统计)

第 6 章	统计描述指标	058
6.1	计量变量	058

6.1.1 集中趋势	059
6.1.2 离散趋势	060
6.2 等级变量	062
6.3 计数变量	062
6.3.1 率 (rate)	063
6.3.2 构成比 (constituent ratio)	063
6.3.3 相对比 (relative ratio)	063
第 7 章 统计表	064
7.1 统计表的结构	064
7.2 制表原则	065
7.3 统计表分类	066
7.4 SPSS 定制专业统计表	066
7.5 SPSS 统计结果展示方式	070
第 8 章 统计图	072
8.1 SPSS 24.0 绘图功能简介	072
8.1.1 图表构建器简介	073
8.1.2 图形画板模板选择程序	077
8.2 条形图 (Bar)	079
8.2.1 统计图的结构	081
8.2.2 统计图的绘图原则	081
8.2.3 统计图形的选择	081
8.2.4 模块解读	082
8.2.5 统计图编辑	086
8.3 3-D 条形图 (3-D Bar)	088
8.4 线图 (Line)	089
8.5 面积图 (Area)	091
8.5.1 简单面积图	091
8.5.2 堆积面积图	092
8.6 饼图 (Pie)	093

8.7	高低图 (High-Low Charts)	094
8.8	箱图 (Boxplot)	096
8.9	误差条图 (Error bar)	097
8.10	人口金字塔图 (population Pyramid)	099
8.11	散点图 (Scatter)	100
8.11.1	简单分布散点图	101
8.11.2	矩阵分布散点图	101
8.11.3	简单点图	102
8.11.4	重叠分布散点图	103
8.11.5	3-D分布散点图	103
8.12	直方图 (Histogram)	104

第四篇 中级统计比一比 (差异性分析)

第9章	t 检验	108
9.1	基本思想与类型	108
9.2	单样本 t 检验	109
9.2.1	设计思想	109
9.2.2	案例实战	109
9.2.3	案例解析	109
9.2.4	实战步骤	110
9.2.5	结果解读	110
9.2.6	拓展理解	111
9.3	两独立样本 t 检验	111
9.3.1	设计思想	111
9.3.2	案例实战	112
9.3.3	案例解析	112
9.3.4	实战步骤	112
9.3.5	结果解读	113
9.3.6	拓展理解	114

9.4	配对样本 t 检验	114
9.4.1	设计思想	114
9.4.2	案例实战	115
9.4.3	案例解析	115
9.4.4	实战步骤	116
9.4.5	结果解读	116
9.4.6	拓展理解	117
9.5	t 检验小结	118
第 10 章 方差分析		119
10.1	单因素设计方差分析	119
10.2	随机区组设计方差分析	124
10.3	析因设计方差分析	129
10.4	重复测量设计方差分析	135
10.5	协方差分析	159
10.6	交叉设计方差分析	163
10.7	拉丁方设计方差分析	165
10.8	嵌套设计方差分析	168
10.8.1	两因素嵌套	169
10.8.2	三因素嵌套	172
10.9	正交设计方差分析	177
10.10	裂区设计方差分析	181
第 11 章 卡方检验		185
11.1	成组四格表卡方	186
11.2	成组 $R \times C$ 表	189
11.3	成组 $R \times C$ 表效应指标比较	191
11.4	$R \times C$ 表确切概率法	192
11.5	线性趋势卡方	193
11.6	配对设计方表	194

11.7	分层卡方	196
11.8	卡方分割	199
第 12 章 等级资料比较		201
12.1	$R \times C$ 表 (单向有序)	201
12.1.1	列有序	201
12.1.2	行有序	201
12.2	$R \times C$ 表 (双向有序)	202
12.2.1	属性相同	202
12.2.2	属性不同	203
第 13 章 非参数检验		205
13.1	非参数卡方	205
13.2	二项检验	207
13.3	游程检验	208
13.4	单样本 $K-S$ 检验	209
13.5	2 独立样本检验	210
13.6	K 独立样本检验	212
13.7	2 个相关样本检验	214
13.8	K 个相关样本检验	216
13.9	非参数检验和参数检验	218
13.9.1	非参数检验的优点	219
13.9.2	非参数检验的缺点	219
13.9.3	两种检验的选择与效度	219
第 14 章 多元方差分析		220
14.1	单组资料	220
14.2	两组比较	222
14.3	多组比较	223
14.4	轮廓分析	226

第五篇 高级找关系 (统计模型)

第 15 章 关联与相关分析	232
15.1 相关与关联简介	232
15.1.1 基本概念	232
15.1.2 关联与相关的类型	233
15.2 Pearson 相关及偏相关	236
15.2.1 Pearson 相关	236
15.2.2 偏相关分析	239
15.3 Spearman 相关及偏相关	243
15.3.1 Spearman 相关	243
15.3.2 Spearman 偏相关	244
15.4 典型相关	246
15.5 拓展: 相关的校正	249
第 16 章 线性回归	251
16.1 简单线性回归	251
16.1.1 线性回归条件	251
16.1.2 线性回归建模策略	252
16.2 多重线性回归	257
第 17 章 Logistic 回归	262
17.1 二项 Logistic 回归	262
17.2 有序 Logistic 回归	267
17.3 多项 Logistic 回归	269
17.4 条件 Logistic 回归	272
第 18 章 生存分析	277
18.1 生存分析概述	277
18.1.1 基本概念	277
18.1.2 生存分析方法	278

18.2	寿命表法	278
18.3	Kaplan-Meier 法	282
18.4	Cox 回归	285
18.5	时间依赖 Cox 回归	288
第 19 章 聚类与判别分析		290
19.1	系统聚类	290
19.1.1	简介	290
19.1.2	基本思想	290
19.1.3	案例实战	290
19.1.4	案例解析	291
19.1.5	实战步骤	291
19.1.6	结果解读	292
19.1.7	拓展理解	294
19.2	快速聚类	295
19.2.1	简介	295
19.2.2	基本思想	295
19.2.3	案例实战	295
19.2.4	案例解析	295
19.2.5	案例实战	296
19.2.6	结果解读	297
19.2.7	拓展理解	299
19.3	两步聚类	299
19.3.1	简介	299
19.3.2	基本思想	299
19.3.3	案例实战	300
19.3.4	案例解析	300
19.3.5	案例实战	300
19.3.6	结果解读	300
19.3.7	拓展理解	302
19.4	Fisher 判别与 Bayes 判别	303

19.4.1	简介	303
19.4.2	基本思想	303
19.4.3	案例实战	303
19.4.4	案例解析	303
19.4.5	案例实战	303
19.4.6	主要结果解读	304
19.4.7	知识小结	306
第 20 章 主成分与因子分析		308
20.1	主成分分析	308
20.1.1	主成分思想	308
20.1.2	主成分分析	310
20.1.3	主成分回归	315
20.1.4	主成分评价	318
20.2	因子分析	321
20.2.1	因子分析思想	321
20.2.2	因子分析实战	321

第六篇 专项统计

第 21 章 信度与效度分析		326
21.1	信度分析	326
21.1.1	信度分类	326
21.1.2	信度优化方法	328
21.2	效度分析	331
21.2.1	效度分类	331
21.2.2	效度分析实战	332
第 22 章 ROC 曲线		337
22.1	诊断试验与 ROC 概述	337

22.2	连续性计量资料 ROC	340
22.2.1	问题 (1) ROC 实战	340
22.2.2	问题 (2) ROC 实战	342
22.2.3	问题 (3) 多指标联合诊断	348
第 23 章 中介效应与调节效应		350
23.1	中介效应与调节效应简介	350
23.2	中介效应 SPSS 实现	352
23.2.1	案例实战	352
23.2.2	案例解读	352
23.2.3	实战步骤	352
23.2.4	结果解读	353
23.3	调节效应 SPSS 实现 (X 与 M 均为分类变量)	354
23.3.1	案例实战	355
23.3.2	案例实战	355
23.3.3	结果解读	355
23.4	调节效应 SPSS 实现 (M 为分类变量, X 为连续变量)	356
23.4.1	案例实战	356
23.4.2	案例实战	356
23.4.3	结果解读	356
23.5	调节效应 SPSS 实现 (M 为连续变量, X 为分类或者连续变量)	357
23.5.1	案例实战	357
23.5.2	案例实战	357
23.5.3	结果解读	358
23.6	Process 插件安装	359
23.7	Process 分析中介效应	361
23.7.1	实战步骤	361
23.7.2	结果解读	362
23.8	Process 分析调节效应	363
23.8.1	实战步骤	363
23.8.2	结果解读	364