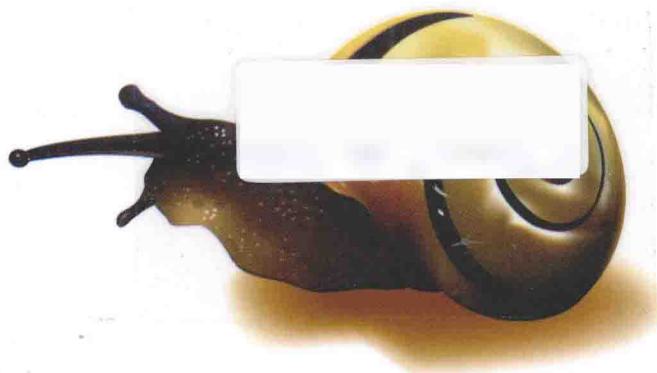


精选**143个实例**，通过SPSS软件实现统计学分析方法，给出操作提示和操作选项说明，并对SPSS输出结果**给予详尽的解释**。

SPSS 与 统计分析

(第2版)

宇传华 主编



提供以Excel格式和SPSS格式建立的实例数据文件



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

SPSS 与 统计分析

(第2版)

宇传华 主编

电子工业出版社

Publishing House of Electronics Industry

北京•BEIJING

内 容 简 介

SPSS 是世界公认的标准统计软件之一。由于其易学易用,深受广大用户,特别是非统计学专业人员的青睐。本书共 27 章,分基础篇和高级篇两部分,基础篇介绍了 SPSS 的基本知识和常用统计学方法;高级篇囊括了大量现代统计学分析方法,如决策树分析、多项分类 logistic 回归、Poisson 回归、神经网络模型、广义估计方程、ROC 分析、典型相关分析、重复测量资料分析、混合效应模型分析、时间序列分析、信度分析、结合分析、对应分析等。

本书编写特色在于:首先尽可能通俗易懂地介绍统计学方法,然后借助于 SPSS 软件实现这些方法,对于 SPSS 运行后所输出的结果给予合理的、详尽的解释。本书特别注重统计学方法的介绍,以及软件输出结果的解释。本书在附带光盘中提供了 100 余个实例数据,可供读者调用、练习。附录 C 以框架流程图形式列出了基于本书的统计学方法选择方案,供读者选择统计学分析方法时参考。

本书既可作为统计学理论学习的参考,也可作为 SPSS 数据分析的实习教材,还可作为从事统计学工作者或与数据分析有关人员的参考书。

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

SPSS 与统计分析 / 字传华主编. —2 版. —北京:电子工业出版社, 2014.7
ISBN 978-7-121-23409-5

I. ①S… II. ①字… III. ①统计分析—软件包 IV. ①C819

中国版本图书馆 CIP 数据核字(2014)第 118041 号

策划编辑:符隆美

责任编辑:葛 娜

印 刷:北京京师印务有限公司

装 订:北京京师印务有限公司

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:787×1092 1/16 印张:47 字数:1082 千字

版 次:2007 年 2 月第 1 版

2014 年 7 月第 2 版

印 次:2014 年 7 月第 1 次印刷

印 数:3000 册 定价:99.00 元(含光盘 1 张)



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。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010) 88258888。

编 委

(按章节先后为序)

曾 庆	重庆医科大学
曹 阳	Karolinska Institutet (瑞典)
宇传华	武汉大学
毛宗福	武汉大学
吕美霞	华中科技大学
方 亚	厦门大学
郭海强	中国医科大学
郝元涛	中山大学
薛富波	西安杨森制药有限公司
尹 平	华中科技大学
张菊英	四川大学
郑迎东	北京大学
张岩波	山西医科大学
洪希成	湖北省疾病预防控制中心
李春晖	武汉大学

秘书：王震坤 季洁 武汉大学

前 言

《SPSS 与统计分析》的出版已经整整 7 个年头了，承蒙读者厚爱，该书出版后进行了多次印刷，Google 学术搜索引用约 260 余次。2010 年，应出版社邀请开始组织编写第 2 版。这期间我们进行了 3 次修订，但由于想给大家呈现最好的书稿，直到这一次修订才决定让它面市。

随着近年来 SPSS 版本的屡屡更新，其界面越来越好看，功能越来越强大，性能越来越稳定。7 年前初版中各类实例的实现基于 SPSS 13.0，在这一版中我们将结合 SPSS 20.0 进行编写。

本书特色

市面上有关 SPSS 的书籍应该说不少，也不乏优秀作品。但大多数书籍只侧重于 SPSS 相应版本的操作步骤介绍，较少阐述相应的统计学方法，使得部分读者在选择具体的统计学方法方面存在一定困难。为此，我们这本书的总体编写思路是：首先尽可能通俗易懂、详细地介绍统计学方法，然后借助于 SPSS 软件去实现。对于 SPSS 所输出的结果给予合理、详尽的解释，即统计学方法、SPSS 操作、分析结果解释齐头并进，尤其强调统计学方法的介绍与分析结果的解释。

本书内容

本书共有 27 章，分为基础篇（第 1~13 章）和高级篇（第 14~27 章）两部分。基础篇介绍了 SPSS 概况、数据类型与各种类型数据的统计学描述、概率分布与正态性检验、置信区间估计与假设检验原理、区间数据的统计推断、名义分类数据的统计推断、有序分类数据的统计推断、简单线性回归与相关、曲线回归与非线性回归、多重线性回归与相关、统计图表、诊断试验评价与 ROC 分析、以及缺失数据处理方法等。高级篇介绍了 logistic 回归、对数线性模型与 Poisson 回归、生存分析与 Cox 模型、聚类分析与判别分析、决策树分析、主成分分析与因子分析、析因分析与协方差分析、重复测量与混合效应模型、多变量方差分析与典型相关分析、信度分析、对应分析与结合分析、时间序列分析、广义线性模型、广义估计方程、神经网络模型等方法。其中后面所列 3 种方法是这一版本新增加的。每一种统计学方法均配有实例，每一实例的 SPSS 操作、输出结果解释都有较详细的说明。因此，通过本书的学习，读者不仅可以学到最新进展的统计学方法，而且也可以通过实例的学习，自己利用 SPSS 解决有关数据分析问题。

本书共提供了 100 余个实例数据，分别采用 Excel 格式和 SPSS 格式建立数据文件，

文件存放在所配光盘中,读者在学到某个例子时,只需从光盘中调出数据,按照书上给出的操作步骤点击 SPSS 软件界面上的菜单,便可获得书中所给结果。

本书除正文外,还提供了 3 个附录。附录 A 列出了 SPSS 的各类函数及其说明;附录 B 简要介绍了 SPSS 统计分析程序及其编写方法;附录 C 以框架流程图形式列出了统计学方法的选择方案,此外,在附录 C 还标出了每一种统计学方法在本书中所对应的章节号。

本书很多章节均具有其鲜明特色,如决策树分析、多项分类 logistic 回归、诊断试验的 ROC 分析、广义估计方程、神经网络模型等方法及其 SPSS 实现,在国内同类书籍中具有较高水准。

尽管本书以 SPSS 20.0 为基础编写,但本书的方法不失其普适性。所以本书亦可作为其他 SPSS 版本教学与科研的参考书。

本书编者

本书第 1,2 章由曾庆、王震坤编写,第 3,4 章由曹阳、宇传华编写,第 5 章由曹阳编写,第 6,7 章由吕美霞、毛宗福编写,第 8~10 章由张菊英编写,第 11 章由郭海强、李春晖、季洁编写,第 12,14 章由宇传华编写,第 13 章由薛富波编写,第 15 章由刘裕、郝元涛编写,第 16 章由尹平、陆芳编写,第 17 章由郑迎东、宇传华编写,第 18 章由方亚编写,第 19~21 章由张岩波编写,第 22,23 章由洪希成、宇传华编写,第 24,25 章由郝元涛编写,第 26 章由郑迎东、洪希成编写,第 27 章由李春晖、宇传华编写;此外,附录 A 由曾庆、郭海强编写,附录 B 由郑雷、宇传华编写,附录 C 由蒋丽丽、宇传华编写。

本书第 2 版的修订过程中,武汉大学的王震坤、孙敏、李春晖、桑舒平、杨子娟、姚聪、夏欣、李鹏、连肖、季洁、魏娜娜、李妍、鲍俊哲以及山西医科大学的曹红艳承担了 SPSS 操作图片和步骤的改写以及文字编辑处理等修订工作。

致谢

在本书的编写过程中,特别值得一提的是魏娜娜同学,她做事认真踏实,发现了很多书稿中存在的潜在错误。此外,还有部分武汉大学公共卫生学院的其他师生也参与了本书的编写和修订工作,在此一并表示衷心感谢!

由于作者水平有限,本书可能还存在部分不尽如人意的地方,恳请各位读者通过 E-mail 等通信方式给予指正。

宇传华 (yuchua@163.com) 于武汉大学

2014 年 2 月 18 日

目 录

基 础 篇

第 1 章 概 述	2
1.1 SPSS 简介	2
1.2 使用 SPSS 进行数据分析的基本步骤	3
1.3 主要窗口和功能	3
1.3.1 数据编辑窗口	4
1.3.2 结果浏览窗口	6
1.3.3 程序编辑窗口	13
1.4 通过数据编辑窗口输入数据	14
1.4.1 使用数据窗口输入数据	14
1.4.2 定义变量	15
1.4.3 数据输入实例	20
1.5 SPSS 数据文件的存取	24
1.5.1 存取保存的 SPSS 文件	24
1.5.2 读取保存的数据文件	25
1.5.3 读取 Excel 电子表格数据文件	25
1.5.4 读取 Access 数据库 (ODBC 数据接口)	26
1.5.5 保存 SPSS 数据文件	29
1.6 数据的编辑与整理	30
1.6.1 发现重复数据	30
1.6.2 选择数据	32
1.6.3 定义权重	35
1.6.4 数据排序	36
1.6.5 数据转置	37
1.6.6 数据合并	38
1.6.7 数据拆分	40
1.6.8 数据汇总	41
1.6.9 查找数据	43
1.7 数据转换	45

1.7.1	公式计算	45
1.7.2	数据编码	48
1.7.3	替代缺失数据	50
1.7.4	数据例编秩	51
1.7.5	频数分组	53
1.8	帮助的获取	53
1.8.1	按专题组织的帮助	53
1.8.2	通过对话框内的 Help 按钮使用帮助	54
1.8.3	使用统计教练	54
1.8.4	使用联机帮助和网络讨论组	54
第 2 章	数据类型与统计学描述	55
2.1	数据分类	55
2.2	制作频数表	56
2.2.1	区间数据频数分段	56
2.2.2	用 Frequencies 编制频数表	62
2.3	用 Descriptives 进行区间数据的统计描述	67
2.3.1	操作过程	67
2.3.2	结果解释	68
2.4	用 Explore 进行区间数据的统计描述	69
2.4.1	操作过程	70
2.4.2	结果解释	72
2.5	用 Bivariate 进行变量间的相关与协方差分析	76
2.5.1	操作过程	76
2.5.2	结果解释	77
2.5.3	描述性统计分析过程的比较	78
2.6	名义数据的统计描述	80
2.6.1	单个名义变量的描述分析	80
2.6.2	多指标的描述分析	82
第 3 章	概率分布与正态性检验	87
3.1	概率分布	87
3.1.1	正态分布	87
3.1.2	二项分布	90
3.1.3	Poisson 分布	94
3.2	抽样分布	96
3.2.1	t 分布	96

3.2.2	χ^2 分布	98
3.2.3	F 分布	100
3.3	正态性检验	101
3.3.1	P-P 图法	102
3.3.2	Q-Q 图法	104
3.3.3	直方图、箱式图与茎叶图	105
3.3.4	计算法	111
第4章	区间估计与假设检验	114
4.1	均数的区间估计	114
4.1.1	σ 已知时总体均数的置信区间	115
4.1.2	σ 未知时总体均数的置信区间	116
4.1.3	两总体均数间差值的置信区间	118
4.2	总体方差、总体标准差的置信区间	120
4.3	率的区间估计	121
4.3.1	总体率的置信区间	121
4.3.2	两总体率差值的置信区间	121
4.4	假设检验与两类错误	121
4.4.1	假设检验的概念与原理	122
4.4.2	假设检验的两类错误	123
4.4.3	假设检验的基本步骤	124
4.5	样本含量的估计与检验效能	125
4.5.1	影响样本量大小的因素	125
4.5.2	总体均数区间估计的样本含量	126
4.5.3	样本均数与总体均数比较样本含量估计	126
4.5.4	完全随机设计两样本均数比较的样本含量估计	127
4.5.5	完全随机设计多个样本均数比较的样本含量估计	128
4.5.6	估计总体率时的样本含量估计	129
4.5.7	样本率与总体率比较的样本含量估计	129
4.5.8	两样本率比较的样本含量估计	130
4.5.9	多个样本率比较的样本含量估计	130
4.5.10	直线相关分析的样本含量估计	131
4.5.11	检验效能	131
第5章	区间数据的统计推断	134
5.1	t 检验	134
5.1.1	单个总体均数的 t 检验	134

5.1.2	独立样本成组 t 检验	136
5.1.3	成对样本 t 检验	138
5.2	单因素方差分析	140
5.2.1	两组资料的单因素方差分析	140
5.2.2	多组资料的单因素方差分析	140
5.3	双因素方差分析	142
5.3.1	基本分析步骤	142
5.3.2	关于 Univariate 过程对话框的说明	145
5.4	对比与事后检验	148
5.4.1	对比	148
5.4.2	事后检验	150
5.4.3	Bootstrap	153
5.5	方差齐性检验	155
第 6 章	名义分类数据的统计推断	157
6.1	四格表数据的卡方检验	157
6.1.1	一般四格表卡方检验	157
6.1.2	连续性校正卡方检验	165
6.2	$R \times C$ 无序列联表的卡方检验	169
6.2.1	多个样本率的卡方检验	169
6.2.2	多个样本构成的卡方检验	171
6.3	Fisher's 精确检验	173
6.3.1	四格表的精确概率法	173
6.3.2	$R \times C$ 列联表精确概率	175
第 7 章	有序数据的统计推断	179
7.1	$R \times C$ 单向有序序列联表的检验	179
7.1.1	Wilcoxon 秩和检验	179
7.1.2	趋势 χ^2 检验	182
7.1.3	Kruskal-Wallis 检验	184
7.1.4	实例与操作	185
7.2	双向有序序列联表的检验	187
7.2.1	Spearman 等级相关	187
7.2.2	Jonckheere-Terpstra 检验	189
7.2.3	Cochran-Mantel-Haenszel 统计分析	191
7.3	几个相关有序样本的非参数检验	194
7.3.1	2 相关样本的秩检验	194

7.3.2	多组相关样本检验	198
第 8 章	简单线性回归与相关	201
8.1	一般的简单线性回归	201
8.1.1	线性回归的概念	201
8.1.2	建立线性回归方程	202
8.1.3	回归系数的假设检验	203
8.1.4	实例与操作	204
8.2	加权的简单线性回归	213
8.2.1	加权最小二乘估计	214
8.2.2	加权线性回归方程的假设检验	214
8.2.3	实例与操作	215
8.3	简单线性相关	218
8.3.1	概念	218
8.3.2	线性相关系数的意义和计算	219
8.3.3	相关系数的假设检验	219
8.3.4	实例与操作	220
第 9 章	曲线回归与非线性回归	224
9.1	曲线直线化变换方法	224
9.1.1	变量的变换	224
9.1.2	变量变换后实现线性回归的步骤	225
9.1.3	实例与操作	225
9.2	曲线回归	227
9.2.1	一般步骤	227
9.2.2	SPSS 操作提示	228
9.2.3	实例与操作	230
9.3	非线性回归	233
9.3.1	基本原理	233
9.3.2	SPSS 操作提示	233
9.3.3	实例与操作	237
第 10 章	多重线性回归与相关	241
10.1	多项式回归	241
10.2	多重回归分析方法	242
10.2.1	多重回归模型	242
10.2.2	参数估计	242

10.2.3	回归方程的假设检验与拟合优度评价	243
10.2.4	自变量的选择	243
10.2.5	SPSS 操作提示	244
10.2.6	实例与操作	247
10.3	共线性解决方案与校正	252
10.3.1	多重共线性的诊断	252
10.3.2	共线性解决方案	253
10.4	残差分析与回归诊断	254
10.5	交互作用与哑变量问题	254
10.5.1	交互作用	254
10.5.2	哑变量的设置	255
10.6	复相关系数与偏相关系数	256
10.6.1	复相关系数、决定系数与调整决定系数	257
10.6.2	偏相关系数	257
第 11 章	统计图的制作	261
11.1	条图	262
11.2	3-D 条图	268
11.3	线图	269
11.4	面积图	273
11.5	圆图	274
11.6	高低图	275
11.7	帕累托图	277
11.8	质量控制图	279
11.9	箱图	282
11.10	误差条图	284
11.11	分群金字塔图	286
11.12	散点图	287
11.13	直方图	291
11.14	P-P 概率图	292
11.15	Q-Q 概率图	294
11.16	序列图	295
11.17	统计图形的编辑加工	297
11.17.1	图形编辑窗口简介	297
11.17.2	图形特征的编辑	298
11.17.3	坐标轴编辑	304

11.17.4 图例的编辑	306
11.17.5 添加和显示/隐藏图形元素	306
第 12 章 诊断试验评价与 ROC 分析	308
12.1 常用的诊断试验评价指标	308
12.1.1 正确率	309
12.1.2 灵敏度	309
12.1.3 特异度	310
12.1.4 Youden 指数	311
12.1.5 阳性似然比	311
12.1.6 阴性似然比	312
12.1.7 阳性预测价值	312
12.1.8 阴性预测价值	313
12.1.9 优势比及其有关指标	314
12.1.10 Kappa	316
12.2 ROC 曲线	317
12.2.1 ROC 分析的基本原理	318
12.2.2 SPSS 操作说明	320
12.2.3 实例与结果解释	322
第 13 章 缺失值分析	331
13.1 缺失值分析简介	331
13.1.1 基本概念	331
13.1.2 缺失机制	332
13.1.3 缺失值的常用处理方法	335
13.2 SPSS 操作提示	340
13.2.1 SPSS 的缺失值处理方法	340
13.2.2 缺失值处理的 SPSS 操作	341
13.3 结果解释	345

高 级 篇

第 14 章 logistic 回归	356
14.1 二项分类 logistic 回归	356
14.1.1 方法介绍	357
14.1.2 SPSS 操作选项说明	366
14.1.3 实例与结果解释	371

14.2	条件 logistic 回归	386
14.2.1	方法介绍	386
14.2.2	SPSS 操作选项说明	387
14.2.3	实例与结果解释	387
14.3	有序 logistic 回归	393
14.3.1	方法介绍	393
14.3.2	SPSS 操作选项说明	395
14.3.3	实例与结果解释	398
14.4	多项分类 logistic 回归	404
14.4.1	方法介绍	404
14.4.2	SPSS 操作选项说明	406
14.4.3	实例与结果解释	409
第 15 章	对数线性模型与 Poisson 回归	414
15.1	列联表的对数线性模型	414
15.1.1	方法介绍	414
15.1.2	实例与操作	416
15.2	Poisson 回归	430
15.2.1	基本原理	430
15.2.2	实例与操作	431
第 16 章	生存分析与 Cox 模型	435
16.1	常用术语	435
16.2	非参数分析	437
16.2.1	寿命表法	437
16.2.2	Kaplan-Meier 法	442
16.3	Cox 回归模型	447
16.3.1	方法介绍	447
16.3.2	实例与操作	449
16.4	时间依存变量的处理方法	455
16.4.1	时间依存变量 Cox 模型	455
16.4.2	Cox w/Time-Dep Cov 过程操作说明	457
第 17 章	聚类、判别与决策树分析	460
17.1	概述	460
17.1.1	聚类分析基础知识	460
17.1.2	判别分析基础知识	461

17.1.3 SPSS 聚类和判别分析模块	463
17.2 聚类分析	463
17.2.1 二阶段聚类	463
17.2.2 K 中心聚类	468
17.2.3 层次聚类	470
17.3 判别分析	474
17.4 决策树分析	481
17.4.1 基本原理	481
17.4.2 决策树	490
17.4.3 操作提示	491
17.4.4 结果解释	492
第 18 章 主成分分析与因子分析	496
18.1 主成分分析	496
18.1.1 概述	496
18.1.2 实例与操作	498
18.2 因子分析	511
18.2.1 概述	511
18.2.2 实例与操作	512
18.3 主成分分析与因子分析的联系及区别	518
第 19 章 多因素方差分析	520
19.1 随机区组设计及其方差分析	520
19.1.1 概述	520
19.1.2 实例与操作	521
19.2 析因设计及其方差分析	525
19.2.1 概述	525
19.2.2 实例与操作	525
19.3 嵌套设计及其方差分析	528
19.3.1 概述	528
19.3.2 实例与操作	528
19.4 交叉设计及其方差分析	530
19.4.1 概述	530
19.4.2 实例与操作	530
第 20 章 重复测量与混合效应模型	533
20.1 重复测量方差分析	533

20.1.1	分层随机抽样重复测量数据	534
20.1.2	重复测量设计临床试验数据	546
20.2	线性混合效应模型	549
20.2.1	分层随机抽样调查数据的混合效应模型分析	549
20.2.2	重复测量数据的混合效应模型分析	555
第 21 章	多变量方差分析	560
21.1	单因素设计资料的多元方差分析	561
21.1.1	单样本分析	561
21.1.2	两样本单因素设计资料	564
21.2	多因素资料的多元方差分析	566
21.2.1	两因素设计	566
21.2.2	配对设计资料的多元方差分析	574
21.2.3	重复测量设计资料的多元方差分析	576
21.3	典型相关分析	577
第 22 章	广义线性模型	583
22.1	概述	583
22.1.1	模型的组成	583
22.1.2	常见的几种广义线性模型	585
22.1.3	广义线性模型的三种估计方程及参数估计	585
22.1.4	模型诊断	585
22.2	logistic 回归	586
22.3	Poisson 对数线性模型	594
第 23 章	广义估计方程	600
23.1	概述	600
23.1.1	广义估计方程的基本理论	600
23.1.2	作业相关矩阵	602
23.1.3	广义估计方程的应用	603
23.2	实例与操作	604
23.2.1	数据的一般情况	604
23.2.2	SPSS 操作提示与选项说明	604
23.2.3	SPSS 输出结果及其解释	609
第 24 章	对应分析与结合分析	613
24.1	对应分析	613

24.1.1	方法介绍	613
24.1.2	SPSS 操作选项说明	616
24.1.3	实例分析	617
24.1.4	多重对应分析	620
24.2	结合分析	620
24.2.1	方法介绍	620
24.2.2	SPSS 操作选项说明	625
24.2.3	实例分析	625
第 25 章	信度分析	630
25.1	重复测量法与分半信度法	631
25.1.1	方法介绍	631
25.1.2	实例与操作	632
25.2	Cronbach α 系数	635
25.2.1	方法介绍	635
25.2.2	SPSS 操作选项说明	635
25.2.3	实例描述	637
25.3	Kappa 系数	638
25.3.1	方法介绍	638
25.3.2	实例描述	639
25.3.3	操作选项说明	640
25.3.4	结果解释	641
25.4	Kendall 和谐系数	642
25.4.1	方法介绍	642
25.4.2	实例描述	642
25.4.3	SPSS 操作选项说明	643
25.4.4	主要结果	644
第 26 章	时间序列分析	645
26.1	概述	645
26.1.1	时间序列数据及其分析方法	645
26.1.2	时间序列分析的模型、公式和记号	646
26.1.3	SPSS 时间序列分析功能	650
26.2	时间序列数据的预处理	651
26.2.1	定义日期变量	651
26.2.2	创建时间序列	653
26.2.3	填补缺失数据	658