

SPSS统计分析

从入门到精通

◎ 杜强 贾丽艳 严先锋 编著

(第2版)

源于实践 成就行家

SPSS

数据文件

- ▶ 全面展示新版本SPSS的功能应用。
- ▶ 包括各种数据分析应用的实例，从基本的描述性统计、均值比较、一般线性模型、相关分析、回归分析、对数线性模型、聚类分析，到高级的生存分析、时间序列分析、多重响应分析等。
- ▶ 6个完整的案例，帮助读者尽快进入实战。

人民邮电出版社
POSTS & TELECOM PRESS

SPSS统计分析 从入门到精通

● 杜强 贾丽艳 严先锋 编著

(第2版)



SPSS

人民邮电出版社
北京

图书在版编目(CIP)数据

SPSS统计分析从入门到精通 / 杜强, 贾丽艳, 严先
锋 编著. — 2版. — 北京: 人民邮电出版社, 2014. 7
ISBN 978-7-115-34720-6

I. ①S… II. ①杜… ②贾… ③严… 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036295号

内 容 提 要

本书使用 IBM SPSS Statistics 20 中文界面进行讲解和操作, 致力于使读者全面了解 SPSS, 了解和学习如何使用 SPSS 进行数据融合、数据分析、结果展示等工作, 本书介绍的是 SPSS 的窗口和对话框操作方式, 着重于 SPSS 分析软件的实际应用。

全书共分 25 章。第 1~3 章重点讲解了数据和文件的管理操作, 以及 SPSS 系统环境的设置。第 4~18 章主要介绍各种统计分析方法及其对应 SPSS 过程的操作方式, 包括描述性统计、均值比较、一般线性模型、相关分析、回归分析、对数线性模型、聚类分析、生存分析、时间序列分析、多重响应分析等几大类。第 19 章介绍各种统计图形的生成和编辑。第 20~25 章列举了用 SPSS 处理多种行业数据的案例, 包括上市公司财务数据分析、影响汇率的因素分析、多因素试验设计等多方面的应用。

本书适合自然科学和社会科学各领域、各专业的研究人员多层次的需要, 也可供相关专业本科生、研究生、专业统计分析人士以及管理人员和决策者等学习与参考。

◆ 编 著 杜 强 贾丽艳 严先锋

责任编辑 张 涛

责任印制 彭志环 焦志炜

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

大厂聚鑫印刷有限责任公司印刷

◆ 开本: 787×1092 1/16

印张: 36.75

字数: 1798 千字

2014 年 7 月第 2 版

印数: 10 881—14 380 册

2014 年 7 月河北第 1 次印刷

定价: 79.00 元 (附光盘)

读者服务热线: (010)81055410 印装质量热线: (010)81055316

反盗版热线: (010)81055315



前言

随着大数据概念的兴起，数据分析与挖掘受到越来越多行业的持续关注，如何做到大有所用、大有产出成为业内热议的话题。工欲善其事，必先利其器。在对数据及对应业务的理解方面，SPSS 始终是数据分析从业者可靠的利器之一。

IBM SPSS Statistics 20 也是与时俱进，持续提供优秀的操作体验和分析性能。其操作友好性、模型可靠性以及与其他平台的兼容性等，都有了非常大的提升。IBM SPSS Statistics 20 还增加了许多新的特点，如扩展了线性模型关于预测有序变量的功能，增加了广义线性模型（GZLMs）和广义估计方程（GEEs），可用于处理类型更广泛的统计模型问题；增加了地图可视化功能，用于创建不同类型地图的直观表示模板，包括增强在地图中按区域展示结果的功能；更快地生成表格，比现有的表格生成速度快 5 倍甚至更多，和表格相关的其他操作的速度也有显著提高；提高了排序和保存的性能；在 Server 端运行作业时，可断开与远程服务器的连接，稍后再重新连接并检索结果。

本书第 2 版采用 IBM SPSS Statistics 20（中文界面），并更新之前版中的内容，保持了与最新工具的同步性，同时对实例描述和统计结果解释等内容加以补充和拓展，使本书在使用上更加方便，进一步提高了实用价值。

我们致力于使读者全面了解 SPSS，学习如何使用 SPSS 进行数据融合、数据分析、结果展示等工作。全书分为 25 章：第 1~3 章介绍 SPSS 系统环境的设置，以及对数据和文件的基本操作；第 4~18 章是各种统计分析方法及其对应 SPSS 过程的操作详解，包括描述性统计、均值比较、一般线性模型、相关分析、回归分析、对数线性模型、分类分析、生存分析、时间序列分析、多重响应分析等；第 19 章介绍各种统计图形的生成和编辑；第 20~25 章列举了用 SPSS 处理多种行业数据的案例，包括上市公司财务数据分析、影响汇率的因素分析、多因素试验设计等多方面的应用。

本书内容全面且实例丰富，涉及基本的统计分析方法、常用的数据挖掘手段，软件操作讲解流畅，结果分析详尽透彻，是一本辅助读者使用 SPSS 进行统计分析的良好参考用书。对于读者关心的各种模型参数含义及其设置方法，书中力求讲解详尽；而且从软件操作步骤到行业案例分析，都配有图形指示和样本数据实例分析；对于数理统计和分析方法的相关背景，书中做简单铺垫后，给出算法的基本思想或关键步骤，重点介绍软件操作方法和参数设置，以期达到读者通过本书可以理解和应用相关统计方法的目的。若读者再参考有关统计分析的专业书籍或不同背景知识的书籍，再加上不断的操作演练和实践，就能做更深入的算法改进、应用创新等研究性工作。

致谢：

在此，感谢人民邮电出版社各位编辑同志和张铮老师的帮助和指导，感谢陈明、张阳、李广鹏、马惠来、卞长迪、郑琦、王命达、陈香凝、苑春苗、宁升、李鸿鹏、李明剑等参与了部分章节的编写和修改；感谢罗瑞芳博士、何力武博士、鲍青波、田旭等人参与修订和提出宝贵建议。

由于编者水平有限，统计理论和软件的发展也都非常迅速，书中纰漏和不足之处在所难免，恳请广大读者批评与指正，我们真诚希望与大家相互交流，共同进步（技术支持与完整程序下载网站：www.book95.com）。

目 录

第 1 章 SPSS 20 概述	1
1.1 SPSS 简介	1
1.2 SPSS 的安装、启动和退出	3
1.2.1 SPSS 20 的安装	3
1.2.2 SPSS 的启动	4
1.2.3 SPSS 20 的退出	6
1.3 SPSS 20 的界面及设置	6
1.3.1 常用界面	7
1.3.2 常规选项参数	10
1.3.3 查看器选项参数	12
1.3.4 文件位置选项参数	13
1.3.5 输出选项参数	14
1.3.6 图表选项参数	15
1.3.7 多重归因选项参数	17
1.3.8 枢轴表选项参数	18
1.3.9 数据选项参数	20
1.3.10 货币选项参数	21
1.3.11 脚本选项参数	22
1.3.12 语法编辑器选项参数	23
第 2 章 数据文件的建立与操作	24
2.1 数据编辑器与数据文件	24
2.1.1 数据编辑器	24
2.1.2 数据文件	27
2.2 常量、变量、操作符和表达式	28
2.2.1 常量与变量	28
2.2.2 操作符与表达式	32
2.2.3 如何定义一个变量	33
2.2.4 概率事件	38
2.3 输入数据	38
2.3.1 输入数据的方法	38
2.3.2 查看文件信息和变量信息	38
2.4 编辑数据文件	40
2.4.1 在单元格中编辑数据	40
2.4.2 插入变量与删除变量	41
2.4.3 插入观测量与删除观测量	41
2.4.4 数据的剪切、复制和粘贴	42
2.4.5 撤销操作	43
2.5 对数据文件的操作	43
2.5.1 数据文件的打开与保存	43
2.5.2 数据库文件的转换	44
第 3 章 数据文件的操作	53
3.1 数据文件的一般操作	53
3.1.1 数据排序	53
3.1.2 数据文件的拆分	54
3.1.3 数据文件的合并	56
3.1.4 数据文件的转置	59
3.1.5 变量取值的求秩	60
3.1.6 变量值的重新编码	62
3.1.7 计算新变量	66
3.2 分类汇总	69
3.2.1 数据描述	69
3.2.2 分类汇总的参数设置	69
3.2.3 分类汇总的结果	71
3.3 观测量的加权	72
3.4 数据文件的结构重组	73
3.4.1 选择数据重组方式	74
3.4.2 变量组到观测量组的重组	75
3.4.3 观测量组到变量组的重组	79
3.4.4 转置重组	82
第 4 章 基本统计分析功能	84
4.1 OLAP 在线分析过程	84
4.1.1 数据描述	84
4.1.2 OLAP 过程的操作和设置	84
4.2 个案汇总分析	88

4.2.1	个案汇总分析的参数设置	88
4.2.2	输出结果	89
4.3	按行和列的汇总分析	90
4.3.1	按行汇总过程	90
4.3.2	按列汇总过程	94
4.4	频数分析	96
4.4.1	数据描述	96
4.4.2	对分类变量的频数分析	96
4.4.3	对连续变量的频数分析	98
4.5	描述性统计分析	100
4.5.1	数据描述	100
4.5.2	描述性分析过程	100
4.6	探索性分析过程	101
4.6.1	数据描述	102
4.6.2	探索性分析实例	102
4.7	列联表分析过程	105
4.7.1	数据描述	105
4.7.2	列联表分析的参数设置	106
4.7.3	列联表分析的输出结果	109
4.8	Bootstrap 简介与设置	110
4.8.1	Bootstrap 简介	110
4.8.2	Bootstrap 参数设置	110
第 5 章 均值比较和 T 检验		113
5.1	均值分析过程	114
5.1.1	原理与方法	114
5.1.2	SPSS 实例分析	114
5.2	单样本 T 检验	116
5.2.1	原理与方法	116
5.2.2	SPSS 实例分析	117
5.3	独立样本 T 检验	118
5.3.1	原理与方法	118
5.3.2	SPSS 实例分析	119
5.4	配对样本 T 检验	120
5.4.1	原理与方法	120
5.4.2	SPSS 实例分析	121

第 6 章 非参数检验		123
6.1	非参数检验简介	123
6.1.1	非参数检验与参数检验	123
6.1.2	非参数检验的优点	124
6.1.3	非参数检验的缺点	124
6.2	卡方检验	124
6.2.1	原理与方法	125
6.2.2	数据和问题描述	126
6.2.3	卡方检验实例分析	126
6.3	二项式检验	128
6.3.1	原理与方法	128
6.3.2	数据和问题描述	128
6.3.3	二项式检验实例分析	129
6.4	游程检验	130
6.4.1	原理与方法	130
6.4.2	数据和问题描述	130
6.4.3	游程检验实例分析	131
6.5	单样本 Kolmogorov-Smirnov 检验	132
6.5.1	原理与方法	132
6.5.2	数据和问题描述	132
6.5.3	单样本 K-S 检验实例分析	133
6.6	两个独立样本检验	134
6.6.1	原理与方法	134
6.6.2	数据和问题描述	134
6.6.3	两个独立样本检验实例分析	135
6.7	k 个独立样本的检验	136
6.7.1	原理与方法	136
6.7.2	数据和问题描述	137
6.7.3	k 个独立样本检验实例分析	137
6.8	两个相关样本的检验	138
6.8.1	原理与方法	138
6.8.2	数据和问题描述	140
6.8.3	两个相关样本检验的实例分析	140
6.9	k 个相关样本的检验	141
6.9.1	原理与方法	141

6.9.2	数据和问题描述	143			178
6.9.3	k 个相关样本检验的实例分析	143	8.4.2	问题描述和数据准备	179
第7章	多重响应分析	145	8.4.3	二元 Logistic 回归的参数设置	180
7.1	多重响应概述	145	8.4.4	案例的结果分析	184
7.2	多重响应变量集的定义	145	8.5	多元 Logistic 回归分析	187
7.3	多重响应变量集的频率分析	147	8.5.1	多元 Logistic 回归的原理简介	187
7.4	多重响应变量集的交叉表分析	148	8.5.2	问题描述和数据准备	187
7.5	用表过程研究多重响应变量集	151	8.5.3	多元 Logistic 回归参数设置	188
7.5.1	多重响应变量集的定义	151	8.5.4	案例的结果分析	192
7.5.2	建立包含多重响应变量集的表格	151	8.6	有序回归	194
第8章	回归分析	155	8.6.1	问题描述和数据准备	194
8.1	线性回归	155	8.6.2	有序回归的参数设置	195
8.1.1	一元线性回归的基本原理	155	8.6.3	案例的结果分析	198
8.1.2	多元线性回归的基本原理	157	8.7	概率单位回归分析	200
8.1.3	模型假设的其他检验	158	8.7.1	概率单位回归分析简介	200
8.1.4	问题描述和数据准备	159	8.7.2	问题描述和数据准备	201
8.1.5	线性回归分析的设置和操作	159	8.7.3	概率单位回归的参数设置	201
8.1.6	案例的结果分析	163	8.7.4	案例的结果分析	203
8.2	曲线回归	166	8.8	加权回归分析	204
8.2.1	曲线回归的基本原理	166	8.8.1	加权回归分析简介	204
8.2.2	问题描述和数据准备	167	8.8.2	问题描述和数据准备	205
8.2.3	曲线回归分析的设置和操作	167	8.8.3	加权回归的参数设置	206
8.2.4	案例的结果分析	169	8.8.4	案例的结果分析	206
8.3	非线性回归	170	8.9	二阶段最小二乘回归	208
8.3.1	非线性回归简介	170	8.9.1	二阶段最小二乘回归的基本原理	208
8.3.2	问题描述和数据准备	172	8.9.2	问题描述和数据准备	208
8.3.3	非线性回归的参数设置	173	8.9.3	二阶段最小二乘回归的参数设置	209
8.3.4	案例的结果分析	177	8.9.4	案例的结果分析	210
8.4	二元 Logistic 回归	177	8.10	最佳尺度回归	211
8.4.1	二元 Logistic 回归的数学原理	178	8.10.1	最佳尺度回归原理	211
			8.10.2	问题描述和数据准备	211
			8.10.3	最佳尺度回归的参数设置	212

8.10.4 案例的结果分析	216	10.2.3 案例的结果分析	265
第9章 方差分析	220	10.3 偏相关分析	266
9.1 方差分析简介	220	10.3.1 偏相关分析的基本原理	266
9.1.1 t 检验与方差分析的比较	220	10.3.2 偏相关分析实例	267
9.1.2 方差分析的基本原理	221	10.4 距离分析	268
9.2 单因素方差分析	223	10.4.1 距离分析的基本概念	268
9.2.1 原理与方法	223	10.4.2 距离分析的参数设置	269
9.2.2 单因素方差分析实例	223	10.4.3 距离分析实例	272
9.3 多因素方差分析过程	228	第11章 因子分析	275
9.3.1 原理与方法	228	11.1 因子分析的原理简介	275
9.3.2 二因素方差分析实例	231	11.1.1 因子分析的基本思想	275
9.3.3 协方差分析实例	238	11.1.2 因子分析和主成分分析的联系	275
9.3.4 交互效应中随机因素的分析	241	11.1.3 因子分析的基本步骤	276
9.4 多元方差分析	245	11.2 SPSS 因子分析的应用实例	277
9.4.1 原理与方法	245	11.2.1 数据描述	277
9.4.2 多元方差分析实例	245	11.2.2 SPSS 因子分析过程的设置	278
9.5 重复测量设计的方差分析	247	11.2.3 结果分析	282
9.5.1 原理与方法	247	第12章 分类分析	288
9.5.2 SPSS 实例分析	248	12.1 聚类分析的原理简介	288
9.6 方差成分分析	253	12.1.1 聚类分析的基本概念	288
9.6.1 原理简介	253	12.1.2 聚类分析的一般原理	289
9.6.2 SPSS 实例分析	253	12.2 快速样本聚类过程	291
9.7 正交试验设计	256	12.2.1 快速聚类简介	291
9.7.1 正交试验设计简述	257	12.2.2 问题描述和数据准备	291
9.7.2 SPSS 实例分析	257	12.2.3 SPSS 快速聚类的设置	292
9.7.3 正交试验设计的方差分析	259	12.2.4 案例的结果分析	294
第10章 相关分析	261	12.3 系统聚类	295
10.1 相关分析的基本概念	261	12.3.1 系统聚类简介	295
10.1.1 相关分析的特点和应用	261	12.3.2 问题描述和数据准备	295
10.1.2 相关系数的计算	262	12.3.3 SPSS 系统聚类的设置	
10.1.3 SPSS 提供的相关分析功能	263		
10.2 两变量相关分析	263		
10.2.1 问题描述和数据准备	264		
10.2.2 相关分析的参数设置			

.....	296	348
12.3.4 案例的结果分析.....	299	13.1.4 SPSS 中的生存分析过程	348
12.3.5 对聚类结果的进一步分析	301	348
12.4 两步聚类分析.....	302	13.2 寿命表分析.....	348
12.4.1 两步聚类简介.....	302	13.2.1 寿命表分析简介.....	349
12.4.2 问题描述和数据准备	303	13.2.2 寿命表分析的基本步骤	349
.....	303	349
12.4.3 SPSS 两步聚类的设置	304	13.2.3 寿命表实例分析.....	350
.....	304	13.3 Kaplan-Meier 分析.....	352
12.4.4 案例的结果分析.....	307	13.3.1 Kaplan-Meier 分析的步骤	353
12.5 一般判别分析.....	310	353
12.5.1 判别分析的基本原理	310	13.3.2 生存曲线的比较和检验	353
.....	310	353
12.5.2 问题描述和数据准备	311	13.3.3 Kaplan-Meier 分析实例	353
.....	311	353
12.5.3 判别分析的参数设置	312	13.4 Cox 回归模型.....	357
.....	312	13.4.1 Cox 回归模型的原理简介	357
12.5.4 案例的结果分析.....	314	357
12.6 逐步判别分析实例.....	318	13.4.2 Cox 回归实例分析.....	358
12.6.1 问题描述和数据准备	318	第 14 章 信度分析.....	366
.....	318	14.1 信度分析.....	366
12.6.2 逐步判别的参数设置	319	14.1.1 信度分析的基本原理	366
.....	319	366
12.6.3 案例的结果分析.....	321	14.1.2 问题描述和数据准备	368
12.7 决策树分析.....	324	368
12.7.1 决策树分类的基本原理	324	14.1.3 信度分析的参数设置	368
.....	324	368
12.7.2 决策树过程的参数设置	326	14.1.4 案例的结果分析.....	370
.....	326	14.2 多维尺度分析.....	371
12.7.3 问题描述和数据准备	339	14.2.1 多维尺度分析简介.....	371
.....	339	14.2.2 问题描述和数据准备	371
12.7.4 案例分析.....	339	371
第 13 章 生存分析.....	346	14.2.3 ALSCAL 过程的参数设置	371
13.1 生存分析简介.....	346	371
13.1.1 生存分析的基本概念	346	14.2.4 案例的结果分析.....	374
.....	346	第 15 章 时间序列分析.....	377
13.1.2 生存分析的数据特点	348	15.1 SPSS 的时间序列分析概览.....	377
.....	348	15.1.1 创建模型的通用设置选项	378
13.1.3 生存分析的常用方法		378
		15.1.2 应用模型的通用设置选项	

15.2	时间序列数据的预分析	384	16.3.4	案例的结果分析	413
15.2.1	缺失值替换	384	16.4	模型选择过程	415
15.2.2	定义日期变量	385	16.4.1	模型选择过程概述	415
15.2.3	时间序列的平稳化	386	16.4.2	问题描述和数据准备	416
15.3	指数平滑模型	388	16.4.3	层次对数线性模型的操作 过程	416
15.3.1	指数平滑的基本原理	388	16.4.4	案例的结果分析	417
15.3.2	指数平滑模型的参数设置	389	第 17 章	对应分析	420
15.3.3	指数平滑模型实例分析	391	17.1	对应分析的基本原理	420
15.4	ARIMA 模型	395	17.1.1	对应分析与因子分析	420
15.4.1	ARIMA 模型的基本原理	395	17.1.2	SPSS 中的对应分析	421
15.4.2	ARIMA 模型的参数设置	396	17.1.3	使用对应分析的注意事项	421
15.4.3	ARIMA 模型实例分析	398	17.2	简单对应分析	421
15.5	季节分解模型	400	17.2.1	简单对应分析的数学原理	421
15.5.1	季节分解法概述	401	17.2.2	SPSS 简单对应分析实例	422
15.5.2	季节分解模型实例分析	401	17.3	多元对应分析	427
第 16 章	对数线性模型	406	17.3.1	多元对应分析的基本概念 及其特点	428
16.1	对数线性模型概述	406	17.3.2	多元对应分析的参数设置	428
16.1.1	简单列联表分析的不足	406	17.3.3	实例的结果分析	434
16.1.2	对数线性模型的基本形式	406	第 18 章	缺失值分析	438
16.2	常规对数线性模型过程	407	18.1	缺失值分析的概念	438
16.2.1	常规过程概述	407	18.1.1	缺失值的表现方式	438
16.2.2	问题描述和数据准备	407	18.1.2	SPSS 中的缺失值处理 方法	439
16.2.3	常规过程的参数设置	408	18.2	缺失值分析的参数设置	439
16.2.4	案例的结果分析	410	18.3	缺失值分析的实例	443
16.3	Logit 过程	411	第 19 章	统计图形	448
16.3.1	Logit 过程概述	411	19.1	概述	448
16.3.2	问题描述和数据准备	412	19.1.1	数据和变量的准备	448
16.3.3	Logit 过程的参数设置	412	19.1.2	图表构建程序的基本 操作	450
			19.1.3	旧对话框作图	451
			19.1.4	图形的编辑	452
			19.2	条形图	452
			19.2.1	数据和问题描述	452

19.2.2	用图表构建程序作条形图	452	19.11.2	用图表构建程序作散点图	481
19.2.3	用对话框创建条形图	455	19.11.3	用对话框创建散点图	484
19.3	线形图	456	19.12	直方图	486
19.3.1	数据和问题描述	457	19.12.1	数据和问题描述	486
19.3.2	用图表构建程序作线形图	457	19.12.2	用图表构建程序作直方图	486
19.3.3	用对话框创建线形图	458	19.13	P-P 概率图	487
19.4	面积图	459	19.13.1	数据和问题描述	487
19.4.1	数据和问题描述	459	19.13.2	用对话框创建 P-P 概率图	488
19.4.2	用图表构建程序作面积图	460	19.14	Q-Q 概率图	490
19.4.3	用对话框创建面积图	461	19.14.1	数据和问题描述	490
19.5	饼图	462	19.14.2	用对话框创建 Q-Q 概率图	490
19.5.1	数据和问题描述	462	19.15	时间序列图	491
19.5.2	用图表构建程序作饼图	462	19.15.1	普通序列图	491
19.5.3	用对话框创建饼图	464	19.15.2	自相关序列图和偏相关序列图	494
19.6	高低图	464	19.15.3	互相关序列图	496
19.6.1	数据和问题描述	464	19.16	双轴线图	498
19.6.2	用图表构建程序作高低图	464	19.16.1	数据和问题描述	498
19.6.3	用对话框创建高低图	466	19.16.2	用图表构建程序作双轴线图	498
19.7	帕累托图	469	第 20 章 上市公司财务危机预警分析		
19.7.1	数据和问题描述	469	500		
19.7.2	用对话框创建帕累托图	470	20.1	财务危机预警的应用简介	500
19.8	控制图	471	20.1.1	财务危机的定量定义方法	500
19.8.1	数据和问题描述	471	20.1.2	财务危机预警的模型选择	501
19.8.2	用对话框创建控制图	471	20.2	数据描述	501
19.9	箱图	477	20.2.1	数据说明	501
19.9.1	数据和问题描述	477	20.2.2	指标选择	501
19.9.2	用图表构建程序作箱图	477	20.2.3	补充说明	502
19.9.3	用对话框创建箱图	479	20.3	分析方法概述	503
19.10	误差条图	480	20.3.1	判别分析	503
19.10.1	数据和问题描述	480	20.3.2	Logistic 回归方法	503
19.10.2	用对话框创建误差条图	480	20.4	SPSS 建模过程和结论分析	504
19.11	散点图	481	20.4.1	SPSS 数据筛选操作	504
19.11.1	数据和问题描述	481			

20.4.2	SPSS 判别分析建模与分析	507	22.3.2	应用因子分析进行成绩综合评价的注意事项	531
20.4.3	Logistic 回归建模与分析	511	22.4	SPSS 建模过程和结论分析	532
20.5	进一步的分析与应用	514	22.4.1	数据准备	532
20.5.1	分类结果的应用分析	515	22.4.2	SPSS 因子分析建模与分析	534
20.5.2	建模方法的改进	515	22.5	进一步的分析与应用	537
20.6	建议和推广	515	22.6	建议和推广	538
20.6.1	时间序列研究	515	22.6.1	高中生的成绩综合评价	538
20.6.2	数据的有效预警期	515	22.6.2	对缺失数据的处理	538
20.6.3	指标的简化方法	516	22.6.3	多种方法结合的综合评价模型	539
第 21 章	影响汇率的因素分析	517	第 23 章	高等教育办学条件的聚类分析	540
21.1	汇率影响因素简介	517	23.1	数据描述	540
21.2	数据描述	518	23.1.1	关于基本办学条件指标合格与否的判定	540
21.3	分析方法概述	519	23.1.2	指标选取	542
21.3.1	探索性分析	519	23.1.3	数据格式	542
21.3.2	多元回归分析	519	23.2	聚类分析法简述	542
21.4	SPSS 建模过程和结论分析	520	23.3	SPSS 建模过程和结论分析	543
21.4.1	数据准备	520	23.3.1	对专科院校进行聚类的设置操作	543
21.4.2	探索性分析	521	23.3.2	对本科院校的分析	547
21.4.3	多元回归分析	522	23.4	建议和推广	550
21.5	进一步的分析与应用	525	第 24 章	试卷信度的检验与分析	551
21.5.1	剔除存在共线性的外汇储备变量	525	24.1	试卷信度检验的背景简介	551
21.5.2	回归模型的进一步改进	526	24.1.1	测验内容的自身方面	551
21.5.3	两个回归模型比较	527	24.1.2	施测过程	551
21.6	建议和推广	528	24.1.3	被测试者的自身因素	551
21.6.1	时间序列研究	528	24.2	数据描述	552
21.6.2	汇率影响因素的定性分析	528	24.3	分析方法概述	552
第 22 章	因子分析在成绩综合评价中的应用	529	24.3.1	试卷信度的基本计算公式	552
22.1	学生成绩的综合评价简介	529	24.3.2	试卷信度的估计方法	553
22.2	数据描述	529	24.4	SPSS 建模过程和结论分析	554
22.3	分析方法概述	530	24.4.1	SPSS 信度分析的参数设置	554
22.3.1	应用因子分析进行成绩综合评价的步骤	530			

24.4.2	结果分析	554
24.5	建议和推广	556
第 25 章	多因素试验的设计与分析	557
25.1	试验设计简介	557
25.1.1	试验设计的应用	557
25.1.2	试验设计问题的解决步骤	558
25.2	数据描述	558
25.3	分析方法概述	559
25.3.1	正交设计方法	559
25.3.2	综合评分方法	560
25.4	SPSS 建模过程和结论分析	561
25.4.1	数据标准化	561
25.4.2	性能指标权重的确定	563
25.4.3	利用权重求综合指标	563
25.4.4	对综合得分的进一步分析	564
25.5	建议和推广	565

部分习题答案提示	567
习题 2	567
习题 3	567
习题 4	567
习题 5	568
习题 6	568
习题 7	569
习题 8	569
习题 9	569
习题 10	570
习题 11	570
习题 12	570
习题 13	571
习题 14	571
习题 15	571
习题 16	571
习题 17	572
习题 18	572
习题 19	572

参考文献	573
------	-----

第1章 SPSS 20 概述

SPSS 通过简单的菜单式操作,就可以方便地规范和融合搜集到的各种数据,并能实施从简单的描述性统计分析到复杂的时序分析等多种方法,对数据进行建模,返回有意义的分析结果,比如客户特征的分类、发展趋势的预测等。把这些结果应用于实际,可以帮助读者在发掘潜在客户、制定长远规划等工作上做出更加准确的判断。

本章首先来介绍 SPSS 20 软件环境的设置内容和设置方式,让用户打造适合自己的 SPSS 20 工作环境。

1.1 SPSS 简介

IBM 于 2009 年收购 SPSS 后,相继推出了 IBM SPSS Statistics 18、19 和 20。业界领先的统计分析软件 SPSS,结合 IBM 的资源优势,必将为用户提供更加方便和强大的智能分析体验。

SPSS 20 的新功能包括:地图,图形画板模板选择器新增了地图模板,例如分区图(着色地图)、带有微型图表的地图和重叠地图等;枢轴表性能,可比以前版本更快地呈现,同时保持对透视和编辑操作的完全支持;广义线性混合模型,增强对有序目标变量的分析,该功能在 Advanced Statistics 附加选项中可用。

SPSS 19 的新功能包括:增加 Bootstrapping、直销(Direct Marketing)等模块,极大地方便了分析专家和市场研究人士的应用;新模块 IBM SPSS Statistics Developer,使 SPSS 可以与 R 软件协作;更新的分析方法(广义线性模型等)和报告方式(轻量表);对于从外部来源读取的数据以及在会话中创建的新变量,改进了确定其默认测量级别的方法;增强的语法编辑器,可将语法编辑器拆分为上、下两个窗格,支持语法加亮和自动完成等功能;加入支持 salesforce.com 的数据库驱动程序等。

1. SPSS 的特点

(1) 界面友好,操作简单。SPSS 的命令语句、子命令及各种选项绝大部分都包含在各种菜单和对话框中,因此用户无须花大量时间记忆繁杂的命令、过程和选项。在 SPSS 中,大多数操作可以通过菜单和对话框来完成,因此操作便捷,易于学习和使用。

(2) 适用性好,因人而异。虽然大部分统计分析方法可以通过菜单和对话框来完成,但是,对于熟悉 SPSS 编程语法的用户,也可以在语法窗口中直接编写程序语句,从而灵活地完成各种复杂的统计分析任务。另外,用对话框指定命令、子命令和选项之后,通过单击界面上的“粘贴”按钮,可将当前对话框设置对应的语法命令,自动粘贴到语法编辑窗口中,还可将其保存为文件,以便重复使用和实现自动化。

(3) 算法隐藏。具有第四代语言的特点,只需通过菜单的选择以及对话框的操作告诉系统要做什么,而无须告诉系统怎样做。用户只需了解统计分析原理,无须通晓统计分析的各种算法细节,即可得到想要的结果。

(4) 接口完善。具有完善的数据转换接口,其他软件生成的数据文件(例如 Excel 文件、Access

文件、关系数据库生成的 DBF 文件、文本编辑软件生成的 ASCII 码数据文件、SAS 数据文件、salesforce.com 提供的数据等) 均可方便地转换成可供 SPSS 分析的数据文件。

SPSS 支持 OLE 技术和 ActiveX 技术, 生成的交互表格和图形对象可以与同类软件进行自动嵌入与链接。而且, SPSS 20 还支持 Sax Basic 和 Python 语言。

(5) 功能强大。SPSS 的核心部分是统计功能, 可以完成数理统计分析任务, 提供了从简单的单变量分析到复杂的多变量分析的多种方法, 既包括常规的相关分析、回归分析、方差分析、卡方检验、t 检验和非参数检验; 也包括多元回归分析、聚类分析、判别分析、主成分分析和因子分析, 还包括时间序列分析、生存分析和可靠性分析等。

(6) 表格和图形化功能。SPSS 可以直接生成数十种风格的表格 (OLAP cubes), 伴随其他分析过程又可生成一般表、多响应表和频数表等表格。利用专门的编辑窗口或结果查看窗口, 还能编辑所生成的表格, 如表 1-1 所示。

SPSS 拥有强大的图形功能, 能生成数十种统计图形, 包括条形图、线图、面积图、饼图、高低图、帕累托图、控制图、箱图、误差条图、散点图、直方图、P-P 概率图、Q-Q 概率图和时间序列图等。通过直观、漂亮的统计图形, 能更加形象地显示分析结果, 如图 1-1 所示。

SPSS 20 的模型浏览器能够交互式地改变部分图形的内容和结构, 使得以前输出的多个图表在一个界面中就能以交互的方式进行分析, 再加上鼠标提示的功能, 可提供给用户更好的分析体验。

表 1-1 SPSS 生成表格

OLAP Cubes						
时段: 6						
	名 称					
	四川长虹		招商银行		总 计	
	收盘价	交易量	收盘价	交易量	收盘价	交易量
合计	26.23	1 586 805	71	3 999 344	97.23	5 586 149
N	4	4	4	4	8	8
均值	6.557 5	396 701.3	17.75	999 836	12.153 8	698 268.6
标准差	0.085 61	13 123.4	0.5	556 136.7	5.991 83	493 885.5
总和的%	0.04	0.038	0.109	0.097	0.149	0.135
合计 N 的%	0.069	0.069	0.069	0.069	0.138	0.138

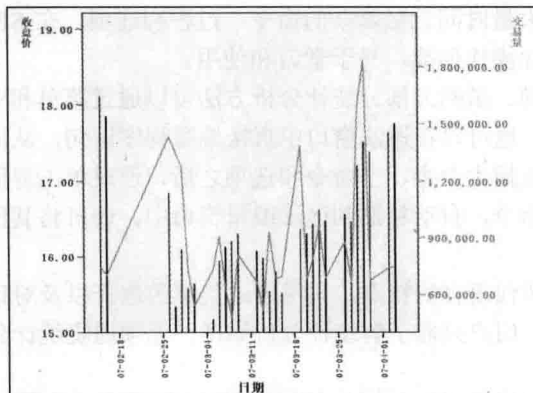


图 1-1 SPSS 图形样例

2. SPSS 20 的不同版本

SPSS 软件是分模块销售的，用户可根据自己的应用需要和用户数量，确定所需要使用的配置。下面所指的 SPSS Statistics 客户端即大家熟悉的 SPSS 统计分析软件。

(1) 教育版 SPSS Statistics 的客户端：SPSS 学生版，适用于中级学生的标准版包含 Base、Advanced、Regression、Resource Center 几个模块，适用于高级学生的专业版包含 Forecasting、Bootstrapping 等全部功能；SPSS 教师专业版，支持所有非商业的研究和教学应用模块；SPSS 校园版，一个满足全校所有学习和研究用户的 12 个月的授权；IBM SPSS 还支持自定义的模块组合购买，以适应多样化的需求。

(2) 商业版 SPSS Statistics 的客户端，有以下三种：SPSS 标准版，主要含有 Linear、Nonlinear、Simulation modeling、Custom Tables 四个模块，能够进行数据探索及常见的多元统计分析；SPSS 专业版，在标准版的基础上增加了几个模块（Data preparation、Decision trees、Forecasting 等），适合处理更多任务的高级用户；SPSS 白金版，含有 SPSS 客户端的所有十多个模块，增加了 Structural equation modeling、Bootstrapping、Direct marketing 等功能。

(3) 对于企业级应用，IBM SPSS 推荐采用企业级的三层体系架构，即 DB Server + SPSS Server + SPSS Clients。该方案具备有力的海量数据运算能力：充分利用了数据库系统的性能、多线程分析和多 CPU 分布式计算的优势、优化的变量挑选算法等。

1.2 SPSS 的安装、启动和退出

本节介绍 SPSS 的安装、启动和退出操作，从这里我们开始踏上 SPSS 应用的旅程。

1.2.1 SPSS 20 的安装

运行 SPSS 20 的安装程序后，首先见到的是如图 1-2 所示的准备界面，稍等片刻即进入如图 1-3 所示的安装向导界面，此后为标准的 Windows 安装过程，一直单击“下一步”按钮，直到完成安装。建议在安装的过程中不要运行其他程序。



图 1-2 SPSS 安装 1

安装的最后一步界面如图 1-4 所示，单击“确定”按钮后，进入授权和注册界面，要求用户选择授权类型，也可以单击“取消”按钮稍后注册。

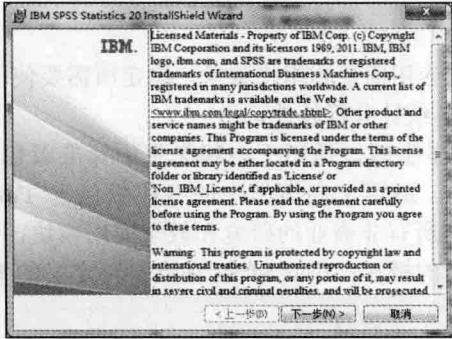


图 1-3 SPSS 安装 2

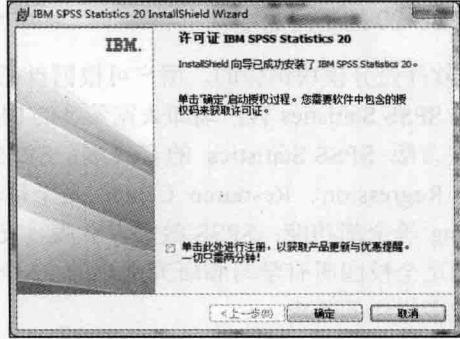


图 1-4 SPSS 安装 3

1.2.2 SPSS 的启动

本小节主要介绍 SPSS 的窗口、菜单运行方式，通过选择窗口、菜单与对话框来完成各种分析过程。

1. 启动

启动 SPSS 程序，可以双击如图 1-5 所示的桌面图标，也可以通过在“开始”菜单中依次单击“程序→IBM SPSS Statistics→IBM SPSS Statistics 20”启动程序。启动后，首先出现如图 1-6 所示的启动界面，显示该软件的版本信息。首次启动后，会先打开如图 1-7 所示的选择界面，表明 SPSS 已正常启动，在此可以设置“以后不再显示此对话框”。

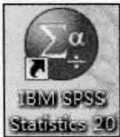


图 1-5 SPSS 的桌面图标

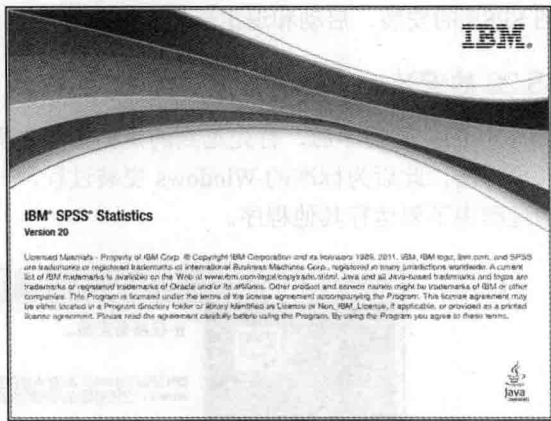


图 1-6 SPSS 的启动界面

对于初学者，一个好的开始是查看软件自身的帮助教程。SPSS 20 提供了较为完善和实用的帮助系统，并支持中文。

在图 1-7 所示中，单击选中“运行教程”单选框，再单击“确定”按钮，就打开如图 1-8 所示的教程界面（通过单击主菜单“帮助→教程”也可打开此教程），这是一个循序渐进的实践指南。另外，SPSS 20 的帮助系统是通过网页的方式访问的。