

高 等 学 校 教 材

# SPSS

## 统计分析基础教程 (第2版)

张文彤 邝春伟 编著



高等教育出版社  
HIGHER EDUCATION PRESS

高等学校教材

# SPSS 统计分析基础教程

SPSS Tongji Fenxi Jichu Jiaocheng

(第2版)

张文彤 邝春伟 编著



高等教育出版社·北京  
HIGHER EDUCATION PRESS BEIJING

## 内容提要

本书采用的 IBM SPSS Statistics 20 中文版, 以真实案例贯穿全书, 从统计分析实战的角度出发详细介绍 SPSS 的界面操作、数据管理、统计图表制作、统计描述和常用单因素统计分析方法的原理与实际操作, 并结合 SPSS 的强大功能进行很好地扩展。书中还提供医疗、经济、市场研究等行业的综合案例, 完全从实际案例出发讲解各类方法的综合运用, 以更好地协助读者提高实战能力。

本书对第 1 版内容进行了全面改写, 以一种全新的实战案例风格出现, 是一本难得的统计理论与 SPSS 操作相结合的参考书。

本书可作为统计学、社会学、教育学等专业本科生和研究生课程教材, 也可作为各行业中非统计专业背景、需要使用统计方法的人员以及希望从头学习 SPSS 软件使用方法的人员的参考书。

## 图书在版编目(CIP)数据

SPSS 统计分析基础教程/张文彤, 邝春伟编著. —2 版. —北京:

高等教育出版社, 2011. 11

ISBN 978-7-04-033241-4

I. ①S… II. ①张… ②邝… III. ①统计分析-软件包, SPSS-高等学校-教材 IV. ①C819

中国版本图书馆 CIP 数据核字(2011)第 211085 号

策划编辑 耿芳  
插图绘制 宗小梅

责任编辑 耿芳  
责任校对 杨凤玲

封面设计 于文燕  
责任印制 田甜

版式设计 杜微言

出版发行 高等教育出版社

社 址 北京市西城区德外大街 4 号

邮政编码 100120

印 刷 廊坊市科通印业有限公司

开 本 787mm × 1092mm 1/16

印 张 27

字 数 660 千字

购书热线 010-58581118

咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>

<http://www.hep.com.cn>

网上订购 <http://www.landaco.com>

<http://www.landaco.com.cn>

版 次 2004 年 9 月第 1 版

2011 年 11 月第 2 版

印 次 2011 年 11 月第 1 次印刷

定 价 39.00 元

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换

版权所有 侵权必究

物料号 33241-00

# 前 言

作者在2004年出版的《SPSS 统计分析基础教程》和《SPSS 统计分析高级教程》，受到许多高校教师的好评，至今仍在卓越亚马逊、当当等网站热销，在此感谢广大读者的厚爱。

7年之后，SPSS先后推出了7个版本，功能早已今非昔比，其用户群也提出了深入和复杂的阅读需求，第2版教材希望能够在以下两个方面满足读者的需求。

(1) 实战操作讲解：复杂统计模型很难用简单案例完全演示其实战应用，实战中的真实案例也往往需要用多种模型结合行业背景来分析。本书采用真实案例进行讲解，并结合不同行业进行案例分析。

(2) 软件最新功能介绍：SPSS软件近几年来增加了许多新功能，如数据挖掘方面的智能分析方法、Bootstrap抽样方法、数据管理方面的查错模块/向导等，本书均会涉及这些新功能的介绍。

严格地讲，本书和第1版相比，除章节框架类似外，内容几乎完全不同。具体而言，作者主要针对SPSS现状，有针对性地对第1版进行以下的加、减、乘、除4项工作。

(1) 加：不仅仅局限于入门内容，而是结合最新的IBM SPSS Statistics 20版的全部功能，大幅增加SPSS软件新功能、使用技巧方面的讲解，包括各种智能数据准备方法、Bootstrap抽样、蒙特卡罗方法等计算统计学算法的介绍和应用案例，从而将统计学基本原理的讲解和SPSS最新版本的功能很好地结合起来，成为兼具教学用书和参考用书双重特色的书籍。

(2) 减：配合软件界面的中文化，在确信读者能够理解含义并正确使用的前提下，大幅简化了书中对软件操作界面的重复介绍和解释。

(3) 乘：本书均采用真实的数据案例进行结构安排和讲解，并且以案例教程的方式出现，使本书更贴近统计分析实战。同时在每一部分后面增加“统计实战案例集锦”一章，提供医疗、经济、市场研究等各行业的复杂真实案例，完全从实际案例出发讲解各类方法的综合运用，将原理、操作和实战应用有机结合起来，真正帮助读者做到学以致用。

(4) 除：对过于艰深、在国内基本上没有实用价值的方法和操作，或直接省略，或仅进行简介，以突出实用性。另一方面，对于统计理论，也完全出于实用性的考虑，只介绍必要内容，保留对统计理论深入浅出讲解的风格，以大幅降低初学者的入门难度，最大程度地提高本书含金量。

本书第1~12章、16~19章、附录由张文彤编写，第13、14、15章由张文彤和邝春伟共同编写。本书内容完整覆盖目前国内大部分专业本科统计课程的教学范围，并结合SPSS的强大功能进行了很好地扩展。各章后均附有思考与练习题，涉及统计理论的章节还提供了小结，可以用做各专业本科生和研究生的统计学教材。本书同时也是一本SPSS 15以上版本的通用入门教材，因此完全可以作为各行业中非统计专业背景、需要使用统计方法的人员，以及希望从头学习SPSS软件使用方法人员的参考书。

本书在编写过程中得到了多方协助，承蒙INTAGE China的谢巍总经理允许作者使用中国消费者信心调研项目的真实数据作为本书案例，SPSS China(析数软件)则在技术资料方面提供了

诸多帮助,并同意将本书作为 SPSS China 官方推荐教材使用,特在此一并致谢。

为便于读者交流和使用本书,这里特公布相关网址如下:

作者微博:<http://weibo.com/wintone>

读者交流微群:<http://q.weibo.com/749521>

本书案例数据、内容更新下载:<http://www.MedStatStar.com>

希望本书能够帮助读者更好地了解统计分析方法,从而进一步促进统计分析方法在国内的普及。也希望广大读者能一如既往地踊跃提出自己使用中的宝贵意见和建议,使得本书推出第 3 版时能够更上一层楼,更好地满足大家的学习和工作需求。

编 者

2011 年 8 月

# 目 录

## 第一部分 数据管理与软件入门

<b>第 1 章 SPSS 入门</b> .....	3	1.6.2 半试验研究支持下的统计 方法论 .....	23
1.1 SPSS 概述 .....	3	1.6.3 偏智能化、自动化分析的数据 挖掘应用方法论.....	24
1.1.1 SPSS 发展简史与版本选择.....	3	思考与练习 .....	25
1.1.2 SPSS 的产品定位 .....	5	<b>第 2 章 数据录入与数据获取</b> .....	26
1.1.3 SPSS 的基本特点 .....	6	2.1 CCSS 案例项目背景 .....	26
1.1.4 SPSS 的客户机/服务器结构 与模块化结构 .....	6	2.1.1 项目背景 .....	26
1.2 SPSS 操作入门.....	7	2.1.2 项目问卷 .....	27
1.2.1 SPSS 的安装与激活 .....	7	2.2 数据格式概述 .....	28
1.2.2 SPSS 的启动与退出 .....	8	2.2.1 统计软件中数据的录入 格式 .....	28
1.2.3 SPSS 的操作方式 .....	9	2.2.2 变量属性 .....	29
1.2.4 SPSS 对话框操作基本规范.....	9	2.3 数据的直接录入 .....	33
1.3 SPSS 的窗口、菜单和结果输出 .....	11	2.3.1 操作界面说明 .....	33
1.3.1 SPSS 的 4 种窗口 .....	11	2.3.2 开放题和简单单选题的 录入 .....	34
1.3.2 SPSS 的菜单 .....	12	2.3.3 多选题的录入 .....	37
1.3.3 SPSS 的 4 种结果输出.....	14	2.4 外部数据的获取 .....	39
1.3.4 分析结果的保存和导出 .....	16	2.4.1 读取电子表格数据文件 .....	39
1.4 SPSS 的系统选项、中文化设置与 附加安装包 .....	17	2.4.2 读取文本数据文件 .....	41
1.4.1 SPSS 的系统选项与中文化 设置 .....	17	2.4.3 用 ODBC 接口读取各种数据 库文件 .....	43
1.4.2 SPSS 网站提供的附加安 装包 .....	18	2.5 数据的保存 .....	44
1.5 SPSS 的帮助系统 .....	19	2.6 数据编辑窗口常用操作技巧集锦 .....	45
1.5.1 学习向导 .....	19	思考与练习 .....	48
1.5.2 帮助菜单 .....	21	<b>第 3 章 变量级别的数据管理</b> .....	49
1.5.3 针对高级用户的帮助功能 .....	22	3.1 变量赋值 .....	50
1.6 数据分析方法论概述 .....	22	3.1.1 常用基本概念 .....	50
1.6.1 严格设计支持下的统计 方法论 .....	23		

3.1.2	“计算变量”过程对话框	51	4.4.3	复制数据属性	80
3.1.3	案例:年龄变量 S3 的分组	51	4.4.4	新建自定义属性和设置未知 测量属性	81
3.2	已有变量值的分组合并	52	4.5	与数据准备有关的功能	82
3.2.1	对连续性变量进行分组 合并	52	4.5.1	SPSS 中与数据准备相关的 功能	82
3.2.2	分类变量类别的合并	53	4.5.2	数据验证模块	83
3.3	连续性变量的离散化	54	4.5.3	标识重复个案	85
3.3.1	可视离散化过程	54	4.5.4	标识异常个案	87
3.3.2	最优离散化过程	55	思考与练习	89	
3.4	变量的自动重编码与数值移动	57	<b>第 5 章 SPSS 编程与扩展</b>	<b>90</b>	
3.4.1	变量的自动重编码	57	5.1	SPSS 编程入门	90
3.4.2	变量值的移动	58	5.1.1	基本语法规则	90
3.5	转换菜单中的其他功能	59	5.1.2	SPSS 程序的创建方式	92
3.5.1	指定数值的查找与计数	59	5.1.3	结构化语句简介*	93
3.5.2	变量的编秩	59	5.1.4	一个简单程序示例	95
3.5.3	自动准备建模数据	60	5.2	语法编辑窗口操作入门	96
3.5.4	随机数字生成器	61	5.2.1	语法编辑窗口界面	96
思考与练习		62	5.2.2	程序的运行与调试	97
<b>第 4 章 文件级别的数据管理</b>	<b>63</b>		5.3	INCLUDE 命令与宏程序	98
4.1	几个常用过程	63	5.3.1	INCLUDE 命令	98
4.1.1	排序个案	63	5.3.2	宏程序	99
4.1.2	分割文件	65	5.4	OMS 系统与程序自动化	100
4.1.3	选择个案	65	5.4.1	OMS 系统	100
4.1.4	加权个案	67	5.4.2	程序自动化	103
4.1.5	分类汇总	68	思考与练习	104	
4.2	数据文件的重组与转置	70	<b>第 6 章 统计实战案例集锦(一)</b>	<b>105</b>	
4.2.1	数据的长型与宽型格式	70	6.1	数据异常值的自动核查与报告	105
4.2.2	长型格式转换为宽型格式	71	6.1.1	项目背景	105
4.2.3	宽型格式转换为长型格式	73	6.1.2	分析思路	106
4.2.4	数据转置	74	6.1.3	利用数据验证模块实现 查错	106
4.3	多个数据文件的合并	75	6.1.4	利用函数功能实现查错	108
4.3.1	一些基本概念	75	6.1.5	项目总结与讨论	110
4.3.2	数据文件的纵向拼接	75	6.2	CCSS 项目数据的自动计算与 处理	110
4.3.3	数据文件的横向合并	76			
4.4	与数据字典有关的功能	78			
4.4.1	数据字典的基本概念	78			
4.4.2	定义变量属性	79			

6.2.1 项目背景 .....	110
6.2.2 分析思路 .....	111
6.2.3 具体操作 .....	112

6.2.4 项目总结与讨论 .....	114
思考与练习 .....	114

## 第二部分 统计描述与统计图表

### 第7章 连续变量的统计描述与参数估计 .....

7.1 连续变量的统计描述指标体系 .....	117
7.1.1 集中趋势的描述指标 .....	118
7.1.2 离散趋势的描述指标 .....	119
7.1.3 分布特征、其他趋势的描述指标 .....	120
7.1.4 SPSS 中的相应功能 .....	121
7.2 连续变量的参数估计指标体系 .....	122
7.2.1 正态分布 .....	122
7.2.2 参数的点估计 .....	123
7.2.3 参数的区间估计 .....	124
7.2.4 SPSS 中的相应功能 .....	125
7.3 案例:信心指数的统计描述 .....	125
7.3.1 使用频率过程进行分析 .....	125
7.3.2 使用描述过程进行分析 .....	127
7.3.3 使用探索过程进行分析 .....	128
7.4 Bootstrap 方法 .....	131
7.4.1 模型 .....	131
7.4.2 案例:对总指数进行 Bootstrap 估计 .....	132
思考与练习 .....	134

### 第8章 分类变量的统计描述与参数估计 .....

8.1 指标体系概述 .....	135
8.1.1 单个分类变量的统计描述 .....	135
8.1.2 多个分类变量的联合描述 .....	136
8.1.3 多选题的统计描述 .....	136
8.1.4 分类变量的参数估计 .....	137

8.1.5 SPSS 中的相应功能 .....	137
8.2 案例:对学历等背景变量进行描述 .....	138
8.2.1 使用频率过程进行描述 .....	138
8.2.2 使用交叉表过程进行描述 .....	138
8.3 案例:对多选题 C0 还贷状况进行描述 .....	140
8.3.1 多选题的频数列表 .....	140
8.3.2 多选题的列联表分析 .....	141
思考与练习 .....	143

### 第9章 数据的报表呈现 .....

9.1 统计表入门 .....	144
9.1.1 统计表的基本框架 .....	144
9.1.2 表头、数据区与汇总项 .....	145
9.1.3 单元格的数据类型 .....	146
9.1.4 几种基本表格类型 .....	146
9.1.5 SPSS 中的报表功能 .....	148
9.1.6 SPSS 中统计表的基本绘制步骤 .....	149
9.2 简单案例:题目 A3 的标准统计报表制作 .....	149
9.2.1 案例简介 .....	149
9.2.2 绘制表格基本框架 .....	150
9.2.3 设置摘要统计量及格式 .....	152
9.2.4 调整各种显示细节 .....	153
9.3 复杂案例:题目 A3a 的标准统计报表制作 .....	154
9.3.1 案例简介 .....	154
9.3.2 多选题、表格基本框架及汇总项的设定 .....	155
9.3.3 设定分类变量小结和汇	



总项 .....	155	10.5.3 分段条图与百分条图案例: 比较不同月份的 A3a 选项比例分布 .....	192
9.3.4 对话框的其他选项卡 .....	157	10.5.4 条图的编辑 .....	194
9.4 表格的编辑 .....	158	10.5.5 带误差线的条图与误 差图 .....	194
9.4.1 基本编辑操作 .....	159	10.6 线图、面积图、点图与垂线图 .....	197
9.4.2 主要编辑菜单功能 .....	160	10.6.1 多重线图案例:分城市比较 信心指数随时间的变化 趋势 .....	197
9.4.3 表格属性的详细设置 .....	161	10.6.2 线图的编辑 .....	198
9.5 表格模板技术 .....	163	10.6.3 面积图、点图与垂线图 .....	199
9.5.1 模板技术简介 .....	163	10.7 散点图 .....	200
9.5.2 表格的中文兼容问题的 解决 .....	165	10.7.1 简单散点图案例:年龄 S3 与 消费者信心指数间的关系 .....	200
思考与练习 .....	165	10.7.2 散点图的编辑 .....	201
<b>第 10 章 数据的图形展示</b> .....	<b>166</b>	10.7.3 分组散点图案例:分性别 考察年龄对信心指数值 的影响 .....	203
10.1 统计图概述 .....	166	10.7.4 散点图矩阵案例:年龄 S3 与现状指数、预期指数 的关系 .....	204
10.1.1 统计图的基本框架 .....	166	10.7.5 三维散点图 .....	205
10.1.2 统计图的种类 .....	168	10.8 P-P 图和 Q-Q 图 .....	206
10.1.3 SPSS 的统计绘图功能 .....	171	10.8.1 P-P 图 .....	206
10.2 直方图与茎叶图 .....	171	10.8.2 Q-Q 图 .....	208
10.2.1 案例:绘制消费者信心值 的直方图 .....	172	10.9 控制图与 Pareto 图 .....	208
10.2.2 图形的基本编辑操作 .....	174	10.9.1 控制图 .....	208
10.2.3 直方图图形框架的修改 .....	178	10.9.2 Pareto 图 .....	211
10.2.4 直方图的衍生图形 .....	180	10.10 其他统计图 .....	212
10.2.5 茎叶图 .....	182	10.10.1 高低图 .....	212
10.3 箱图 .....	183	10.10.2 ROC 曲线 .....	213
10.3.1 案例:用箱图分月份考察消 费者信心的分布 .....	183	10.10.3 时间序列分析中使用的 图形 .....	215
10.3.2 箱图的编辑 .....	184	思考与练习 .....	216
10.4 饼图 .....	186	<b>第 11 章 统计实战案例集锦(二)</b> .....	<b>217</b>
10.4.1 案例:分城市、月份考察 样本性别比例 .....	186	11.1 探索消费者信心指数随背景资料的	
10.4.2 饼图的编辑 .....	187		
10.5 条图与误差图 .....	188		
10.5.1 简单条图案例:比较不同职 业人群的消费信心值 .....	189		
10.5.2 复式条图案例:分职业进 一步比较不同人群的 现状和预期指数 .....	190		

变化规律 .....	217	生产 .....	225
11.1.1 项目背景 .....	217	11.2.1 项目背景 .....	225
11.1.2 分析思路 .....	217	11.2.2 分析思路 .....	225
11.1.3 具体操作 .....	218	11.2.3 具体操作 .....	227
11.1.4 项目总结与讨论 .....	224	11.2.4 项目总结与讨论 .....	230
11.2 CCSS 项目分析报告的自动化		思考与练习 .....	230

### 第三部分 常用假设检验方法

<b>第 12 章 分布类型的检验</b> .....	233	<b>第 13 章 连续变量的统计推断(一)—— <math>t</math> 检验</b> .....	251
12.1 假设检验的基本思想 .....	233	13.1 $t$ 检验概述 .....	251
12.1.1 问题的提出 .....	233	13.1.1 $t$ 检验的基本原理 .....	251
12.1.2 假设检验的标准步骤 .....	234	13.1.2 SPSS 中的相应功能 .....	253
12.1.3 假设检验的两类错误 .....	235	13.2 样本均数与总体均数的比较 .....	253
12.1.4 假设检验中的其他问题 .....	235	13.2.1 单样本案例: 星期一— 城市信心指数与基准 值的比较 .....	253
12.2 正态分布检验 .....	236	13.2.2 单样本 $t$ 检验中的其他 问题 .....	255
12.2.1 K-S 检验的原理 .....	236	13.3 成组设计两样本均数的比较 .....	256
12.2.2 案例: 考察信心指数分布是 否服从正态分布 .....	236	13.3.1 方法原理 .....	256
12.2.3 使用旧对话框分析案例 .....	240	13.3.2 案例: 不同收入水平家庭的 信心指数比较 .....	257
12.3 二项分布检验 .....	241	13.3.3 适用条件与方差齐性 检验 .....	259
12.3.1 二项分布检验的原理 .....	241	13.4 配对设计样本均数的比较 .....	260
12.3.2 案例: 考察抽样数据的性别 分布是否平衡 .....	241	13.4.1 方法原理 .....	261
12.3.3 使用旧对话框分析案例 .....	242	13.4.2 案例: 治疗前后舒张压均 数的比较 .....	261
12.4 游程检验 .....	243	13.5 本章小结 .....	263
12.4.1 游程检验的原理 .....	243	思考与练习 .....	263
12.4.2 案例: 考察 CCSS 抽样数据 是否随机 .....	244	<b>第 14 章 连续变量的统计推断(二)—— 单因素方差分析</b> .....	265
12.4.3 使用旧对话框分析案例 .....	245	14.1 方差分析简介 .....	265
12.5 蒙特卡罗方法 .....	247		
12.5.1 蒙特卡罗方法简介 .....	247		
12.5.2 蒙特卡罗方法的 SPSS 实现 .....	247		
12.6 本章小结 .....	249		
思考与练习 .....	250		

14.1.1 进行方差分析的原因	265	15.3 两个独立样本的非参数检验	291
14.1.2 方差分析的基本思想	265	15.3.1 方法原理	291
14.1.3 单因素方差分析的应用 条件	267	15.3.2 案例:不同收入家庭经济 现状感受值的比较	293
14.2 案例:不同时点消费者信心 指数的比较	269	15.3.3 使用旧对话框分析案例	294
14.3 均数间的多重比较	272	15.4 多个独立样本的非参数检验	295
14.3.1 直接校正检验水准	272	15.4.1 方法原理	296
14.3.2 专用的两两比较方法	273	15.4.2 案例:不同时点上的家庭经 济现状感受值比较	296
14.3.3 两两比较方法的选择 策略	274	15.4.3 使用旧对话框分析案例	299
14.3.4 多重比较结果出现矛盾 时的解释	275	15.5 多个相关样本的非参数检验	300
14.3.5 案例:不同时点信心指数 的两两比较	275	15.5.1 Friedman 检验	300
14.4 各组均数的精细比较	277	15.5.2 案例:不同时段世博会入 园人数比较	301
14.4.1 方法原理*	277	15.5.3 使用旧对话框分析案例	303
14.4.2 案例:事先计划的两时点 均数比较	278	15.5.4 Kendall 协和系数检验与 Cochran 检验	303
14.5 组间均数的趋势检验	279	15.6 秩变换分析方法	306
14.5.1 方法原理	279	15.6.1 秩变换分析原理简介	306
14.5.2 案例:前3个时点的信心 指数线性趋势检验	280	15.6.2 案例:用秩变换来比较不同 时点的家庭经济感受值	306
14.6 本章小结	281	15.7 本章小结	307
思考与练习	281	思考与练习	308
<b>第15章 有序分类变量的统计推断—— 非参数检验</b>	<b>283</b>	<b>第16章 无序分类变量的统计推断—— 卡方检验</b>	<b>310</b>
15.1 非参数检验概述	283	16.1 卡方检验概述	310
15.1.1 非参数检验的意义	283	16.1.1 卡方检验的基本原理	310
15.1.2 非参数检验预备知识	284	16.1.2 卡方检验的用途	311
15.2 两个配对样本的非参数检验	285	16.1.3 SPSS 中的相应功能	311
15.2.1 方法原理	285	16.2 单样本案例:考察抽样数据的 性别分布	312
15.2.2 案例:北京大学与清华 大学2002年高考录取 分数比较	286	16.2.1 用新对话框界面分析本 案例	312
15.2.3 使用旧对话框分析案例	289	16.2.2 使用旧对话框分析案例	314
		16.3 两样本案例:不同收入级别家庭的 轿车拥有率比较	315

16.4	两分类变量间关联程度的度量	318	18.1.1	相关分析与回归分析的 联系与区别	343
16.4.1	相对危险度与优势比	318	18.1.2	简单回归分析的原理和 要求	344
16.4.2	案例:计算家庭收入级别 和轿车拥有情况的关联 程度	319	18.2	案例:建立用年龄预测总信心 指数值的回归方程	346
16.5	一致性检验与配对卡方检验	320	18.3	多重线性回归模型入门	350
16.5.1	Kappa 一致性检验	320	18.3.1	模型简介	350
16.5.2	配对卡方检验	322	18.3.2	多重线性回归模型的标准 分析步骤	350
16.6	分层卡方检验	322	18.3.3	回归方程中的自变量筛选 方法	353
16.7	本章小结	325	18.3.4	SPSS 中与多重线性回归 模型相关的功能	354
	思考与练习	325	18.3.5	案例:建立自变量包括年龄、 家庭收入的信心指数回归 方程	355
<b>第 17 章</b>	<b>相关分析</b>	327	18.4	本章小结	359
17.1	相关分析简介	327		思考与练习	359
17.1.1	相关分析的指标体系	327	<b>第 19 章</b>	<b>统计实战案例集锦(三)</b>	360
17.1.2	SPSS 中的相应功能	329	19.1	X 药物对原发性高血压治疗的 临床试验研究	360
17.2	简单相关分析	331	19.1.1	项目背景	360
17.2.1	方法原理	331	19.1.2	研究方法	360
17.2.2	案例:考察信心指数值和 年龄的相关性	333	19.1.3	数据准备	361
17.2.3	秩相关系数	335	19.1.4	基线情况比较	363
17.2.4	Kendall 等级相关系数	336	19.1.5	疗效比较	366
17.3	偏相关分析	336	19.1.6	安全性评价	367
17.3.1	方法原理	336	19.1.7	分析结论与总结	369
17.3.2	案例:控制家庭收入的影响 之后考察年龄的作用	337	19.2	咖啡屋需求调查案例	370
17.4	Distance 过程	338	19.2.1	项目背景	370
17.4.1	距离测量与相似性测量的 指标体系	339	19.2.2	数据预分析	372
17.4.2	案例:基因间距离的 计算	340	19.2.3	主体问卷分析	374
17.5	本章小结	342	19.2.4	项目总结与讨论	379
	思考与练习	342	19.3	牙膏新品购买倾向研究案例	379
<b>第 18 章</b>	<b>线性回归模型入门</b>	343	19.3.1	研究背景	379
18.1	线性回归模型简介	343			

19.3.2	分析思路 .....	380	19.4.1	项目背景 .....	388
19.3.3	数据预分析 .....	381	19.4.2	数据的采集 .....	388
19.3.4	数据建模 .....	384	19.4.3	数据预分析 .....	389
19.3.5	项目总结与讨论 .....	387	19.4.4	数据建模 .....	390
19.4	证券业市场绩效与市场结构关系 的实证分析 .....	388	19.4.5	项目总结与讨论 .....	392
				思考与练习 .....	393
<b>附录</b> .....					394
附录1 SPSS 函数一览表 .....					394
附录2 各种情形下最常用统计检验方法索引 .....					405
附录3 统计术语英汉名词对照表 .....					407
附录4 IBM SPSS Statistics 19/20 介绍 .....					413
<b>参考文献</b> .....					416

# 第一部分

## 数据管理与软件入门



# 第1章 SPSS 入门

## 1.1 SPSS 概述

SPSS 软件是世界上应用最广泛的专业统计软件之一,在全球约有 25 万用户,分布于通信、医疗、银行、证券、保险、制造、商业、市场研究和科研教育等多个领域和行业,全球 500 强中约有 80% 的公司使用 SPSS,而在市场研究和市场调查领域则拥有超过 80% 的市场占有率,和 SAS 被并称为当今最权威的两大统计软件。



SPSS 实际上是该软件的简称,其全称则发生过几次变化,最早为 Statistical Package for Social Sciences,意为“社会科学统计软件包”;后来随着 SPSS 产品服务领域的扩大和服务深度的增加,SPSS 公司于 2002 年将英文全称更改为 Statistical Product and Service Solutions,意为“统计产品与服务解决方案”,以反映市场的新趋势;但是在 2009 年 4 月,基于一系列原因,SPSS 公司做出了一个令广大用户无法接受的决定:将 SPSS 软件更名为 PASW (Predictive Analytics Software) Statistics! 幸好在当年 9 月,SPSS 公司被 IBM 收购,而新东家则立即终止了更名计划,重新将软件命名为 IBM SPSS Statistics,算是给这一事件画上了句号。但无论名称如何更改,SPSS 软件的风格和基本定位始终未变,用户都喜欢称其为 SPSS,它也一直是广大用户所喜爱的强大统计工具。

### 1.1.1 SPSS 发展简史与版本选择

SPSS 的历史开始于 1968 年,斯坦福大学的 3 位不同专业的研究生(两位博士生、一位硕士生)编制出了世界上最早的统计软件系统,并将其命名为 SPSS。随后,该软件和相应成立的 SPSS 公司走上了持续发展的创新之路。

#### 1. 公司与软件简史

(1) 1968—1975 年:SPSS 成为真正的产品。从一个雏形开始,经过不断的代码积累和修改,SPSS 终于成为成熟的、可销售的产品。

(2) 1975—1984 年:SPSS 公司成为真正的公司。在一系列探索之后,SPSS 公司终于确立了以统计软件和统计分析服务为主业的方向。

(3) 1984—1992 年:PC 时代。SPSS 公司在全球首家推出了 PC 版的统计分析软件 SPSS/PC + 4,该版本为全球第一套以图形菜单为驱动界面的统计软件,也是 DOS 时代的统计软件经典之作。

(4) 1992—1996 年:Windows 时代。在 1992 年,SPSS 在全球首家推出了 Windows 版的统计分析软件 SPSS 6,随着这一软件的成功,公司也走上了快速发展之路,并收购了诸如 SYSTAT (1994) 和 Jandel(1996) 等一系列同行企业。



(5) 1997—2002年:向大企业进化。SPSS 软件在不断推陈出新,经典的 11 版就在这一期间推出,更重要的是并购行动在继续,诸如 Quantime(市场研究应用软件)、ISL(数据挖掘软件)、ShowCase(商务智能中间件)、NetGenesis(网络数据分析应用)、LexiQuest(文本挖掘软件)和 netExs(OLAP 网络接口及界面)等一系列具有战略价值的公司被并购,这也意味着公司开始形成完整的产品线,SPSS 这一产品的定位及重要性开始下降。

(6) 2003—2008年:向预测分析转型。在完成上述并购后,SPSS 开始重新整合产品线,并开始统一向商务智能与预测分析转型。SPSS 软件被定位为产品线中的普及类工具,和其余产品形成高低端搭配。但这一过程并不顺利,显然市场的成熟速度落后于预期,但公司坚持了下来,并走完了这一段路。SPSS 软件也仍然在不断更新,13 版堪称又一经典,从 17 版开始则提供了基本成熟的中文界面与结果输出。

(7) 2009—现在:新的一页。随着 IBM 的收购,SPSS 产品揭开了新的一页,已站在了更高的平台之上,未来表现令人期待。而最新的 SPSS 19 及 SPSS 20 则为其并购之后的作品,其软件界面已经彻底改变为 IBM 的蓝色风格。

## 2. 软件版本比较

由于 SPSS 大约 1 年时间就会推出一个新版本,导致用户使用的版本可能很多,并不一定都是最新版。为了便于读者选择,这里列出从 11 版至今各版本的基本情况与笔者的评价。

(1) 11 版:重新设计了软件界面,加入了混合线性模型等新方法,随后的 11.5 版又新增了 Custom Tables 模块,并提供多语言输出,并首次能将结果直接导出为 XLS 文件。

(2) 12 版:图形输出更改为目前使用的系统,加入了复杂抽样模块,也是首个提供简体中文版(单独销售)的版本。由于 12 版是软件开发转向 Java 之后的第一个作品,产品质量实在不能算成功,当时 SAS 的新版本也因为采用了 Java 语言而未能成功,两种软件相当。

(3) 13 版:真正的经典之作,很多用户目前仍在用。开始加入树模型等智能统计分析方法,复杂抽样模块等有了较大更新,输出接口加入了 OMS 系统。对于配置较低(内存 1 GB 以下,CPU 仍在迅驰之前的级别)的旧机器,笔者强烈推荐使用该版本。

(4) 14 版:首次可以同时打开多个数据文件,提供了现在已成为主要绘图界面的新的“图表生成器”界面,新增了 Data Validation 模块,此外还提供了很多算法、数据管理、统计方法等细节的更新。

(5) 15 版:提供了 Programmability Extension 功能,可直接调用 Python 等语言编写的代码。加入了 GEE 等统计模型。该版本也较为经典,但代价是和 13 版相比,对系统的硬件要求明显提高。

(6) 16 版:用 Java 重写了整个用户界面,操作更加灵活。加入了神经网络和 PLS(偏最小二乘法),提供了对 R 语言的支持。结果文件也改成了新的 spv 格式。

(7) 17 版:首次提供了包括简体中文的多语言界面,至此 SPSS 中文版才开始在国内得到广泛使用。引进了 SPSS EZ RFM 模块、最近邻分析等一系列比较特殊的分析方法/模块,对语法窗口做了大幅升级,增加了最受 SPSS Statistics 专业用户欢迎的新功能,例如,自动完成、Syntax 代码字符颜色标记、代码行数和断点展示等功能。

(8) 18 版:增加了 Bootstrapping 和 Direct Marketing 两大模块,以及一系列统计方法、数据管理、用户界面、第三方接口等方面的更新,而且每个模块均可独立存在并运行,不再依赖于 Base