



工业和信息化普通高等教育“十二五”规划教材立项项目

21世纪高等学校计算机规划教材

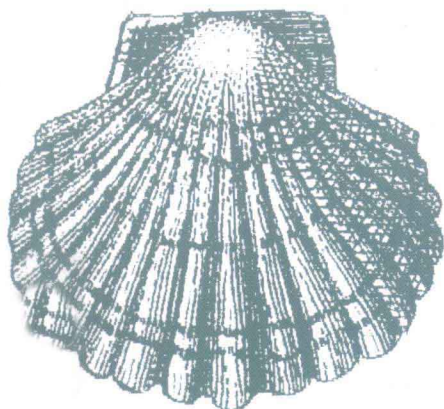
21st Century University Planned Textbooks of Computer Science

SPSS数据分析 教程

SPSS Data Analysis

李洪成 姜宏华 编著

- 采用SPSS统计分析软件的新版本—SPSS 19中文版
- 通过大量实际案例来解析数据分析的技术和技巧
- 从实际的案例入手，对相应的选项给出详细解释



高校系列

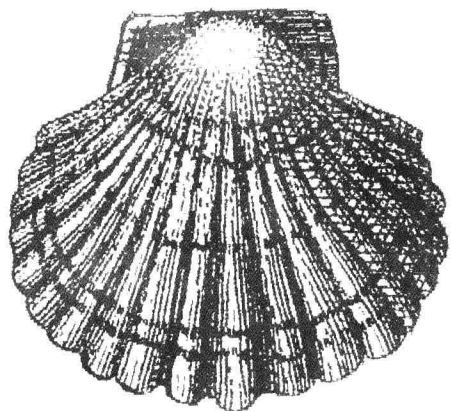
 人民邮电出版社
POSTS & TELECOM PRESS

教育“十二五”规划教材立项项目

SPSS数据分析 教程

SPSS Data Analysis

李洪成 姜宏华 编著



人民邮电出版社

北京

图书在版编目 (C I P) 数据

SPSS数据分析教程 / 李洪成, 姜宏华编著. -- 北京
人民邮电出版社, 2012. 1
21世纪高等学校计算机规划教材. 高校系列
ISBN 978-7-115-26410-7

I. ①S… II. ①李… ②姜… 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235900号

内 容 提 要

本书应用 SPSS 18 和 SPSS 19 中文版进行编写。本书首先从实用角度讲解统计分析的基本概念和理论, 通过数据仿真讲解了随机数、随机变量、分布函数、密度函数、抽样分布等基本理论, 然后从实际案例入手详细分析了描述性统计分析、均值的比较、相关分析、回归分析、方差分析、聚类分析、主成分分析、因子分析等。本书通过大量的实际案例来解析数据分析的技术和技巧, 读者通过本书可以学习和提高数据分析的技能, 掌握数据分析的技巧。

本书将统计分析的基本原理和数据分析的实践相结合, 可以作为大学统计学教学的实训教材, 也可以作为市场分析、定量分析等数据分析从业人员的参考资料。

21 世纪高等学校计算机规划教材

SPSS 数据分析教程

-
- ◆ 编 著 李洪成 姜宏华
责任编辑 蒋 亮
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
三河市海波印务有限公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 18.5 2012 年 1 月第 1 版
字数: 487 千字 2012 年 1 月河北第 1 次印刷

ISBN 978-7-115-26410-7

定价: 38.00 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223
反盗版热线: (010)67171154

释，并对分析过程给出相应的建议。具体特点如下：

- 采用 SPSS 统计分析软件的新版本——SPSS 19 中文版，介绍新的相关功能，比如非参数检验、图表构建程序、两步法聚类等；
- 从实际的案例入手，大部分数据取材于 SPSS 自带的案例数据或者 SPSS 的培训案例；
- 对相应的选项给出详细的解释；
- 每章的后面给出相关的思考与练习，读者可以对相关的内容和技巧进行练习。

感谢

本书是多人共同努力的结果，张宇客提供了第 3 章和第 5 章的初稿，周嫦琦提供了第 4 章的初稿，姜宏华编写了第 9 章、第 10 章和第 11 章，杨东梅提供了第 11 章和第 12 章的初稿，姜政毅编写了第 8 章的部分内容。徐晓琴、曾凯、曹文伟、李红军、赵鹏等也参加了本书的编写。在此一并衷心感谢！

技术支持网站

为本书提供技术支持的网站：books.minewin.com。在该网站，读者可以下载本书的案例数据、勘误信息。采用本书作为教材的教师，可以下载本书的教学课件、习题答案。

由于编者水平有限，书中难免有不足之处。如有偏颇之处，恳切希望读者批评指正。

编者

2011 年 12 月

目 录

第 1 章 统计学和 SPSS 统计分析软件简介

1.1 统计分析的基本概念	1
1.1.1 统计分析的步骤	2
1.1.2 数据的类型	2
1.2 常见统计分析软件简介	3
1.2.1 SPSS	3
1.2.2 SAS	4
1.2.3 Splus 或者 R	4
1.2.4 其他数据分析软件	4
1.3 SPSS 统计分析软件的发展	4
1.4 SPSS 版本和授权	5
1.5 SPSS 统计分析软件的特点	6
1.6 主要模块及功能简介	7
1.7 SPSS 的安装	9
1.8 SPSS 的几种基本运行方式	12
1.9 SPSS 的界面	14
1.10 SPSS 的图形用户界面	17
1.11 SPSS 帮助系统	19
1.12 小结	23
思考与练习	23
参考文献	25

第 2 章 数据文件的建立和管理

2.1 数据管理的特点	26
2.2 SPSS 数据编辑器简介	27
2.2.1 开始 SPSS	27
2.2.2 SPSS 的数据编辑器界面	27
2.3 新建数据文件、数据字典	31
2.4 保存文件	33
2.5 读入数据	34
2.5.1 读入 Excel 数据	35
2.5.2 读入文本数据	36
2.5.3 读入数据库数据	39

2.6 数据文件的合并	43
2.6.1 添加个案	43
2.6.2 添加变量	46
2.7 数据文件的拆分	50
附录：如何为数据库文件建立 ODBC 数据源	52
2.8 小结	53
思考与练习	54
参考文献	56

第 3 章 描述性统计分析

3.1 频率分析	57
3.2 中心趋势的描述：均值、中位数、众数、5%截尾均值	60
3.2.1 均值	60
3.2.2 中位数	60
3.2.3 众数	61
3.2.4 5%截尾均值	61
3.3 离散趋势的描述：极差、方差、标准差、均值的标准误、分位数和变异指标	62
3.3.1 极差	62
3.3.2 方差和标准差	63
3.3.3 均值的标准误	63
3.3.4 变异系数	63
3.3.5 分位数	64
3.4 分布的形状——偏度和峰度	64
3.5 SPSS 描述性统计分析	65
3.5.1 频率入口	66
3.5.2 描述子菜单	67
3.5.3 探索子菜单	68
3.5.4 表格	69
3.6 应用统计图进行描述性统计分析	71
3.6.1 定性数据的图形描述——条形图、饼图、帕累托图	71
3.6.2 定量数据的图形描述——直方图、	

茎叶图和箱图	74	6.2 单样本非参数检验	128
3.7 数据标准化	78	6.2.1 卡方检验	132
3.8 小结	79	6.2.2 二项式检验	136
思考与练习	79	6.2.3 K-S 检验	143
参考文献	80	6.2.4 Wilcoxon 符号秩检验	146
第 4 章 概率论初步	81	6.2.5 游程检验	146
4.1 离散型随机变量的仿真	81	6.3 独立样本非参数检验	147
4.1.1 均匀分布的随机数	81	6.3.1 独立样本检验简介	147
4.1.2 正态分布的随机数	84	6.3.2 独立样本检验举例	149
4.2 理论分布	87	6.4 相关样本非参数检验	151
4.2.1 二项分布的分布函数和概率	87	6.4.1 相关样本检验简介	151
4.2.2 连续分布的随机 变量——正态分布	93	6.4.2 相关样本检验举例	153
4.3 经验分布	97	6.5 小结	156
4.4 抽样分布	99	思考与练习	156
4.5 置信区间	102	参考文献	157
4.6 小结	104	第 7 章 相关分析	158
思考与练习	104	7.1 相关分析的基本概念	158
第 5 章 均值的比较	105	7.1.1 相关关系的种类	159
5.1 假设检验的思想及原理	105	7.1.2 相关分析的作用	159
5.2 均值	107	7.2 散点图	160
5.2.1 均值过程分析	107	7.2.1 散点图简介	160
5.2.2 双因素的均值过程分析	109	7.2.2 散点图——旧对话框	160
5.3 单样本 T 检验	110	7.2.3 用图表构建程序绘制散点图	163
5.3.1 数据准备	111	7.3 相关系数	165
5.3.2 单样本 T 检验	113	7.3.1 线性相关的度量——尺度 数据间的相关性的度量	166
5.3.3 置信区间和自抽样选项	114	7.3.2 Spearman 等级相关系数——定序 变量之间的相关性的度量	170
5.4 独立样本 T 检验	115	7.3.3 Kendall 的 tau-b(K)	171
5.4.1 数据初探	116	7.4 偏相关分析	171
5.4.2 T 检验	119	7.5 小结	173
5.4.3 均值差的绘图	121	思考与练习	173
5.5 配对样本 T 检验	122	参考文献	174
5.6 小结	125	第 8 章 回归分析	175
思考与练习	125	8.1 线性回归分析的基本概念	175
参考文献	126	8.2 简单线性回归	177
第 6 章 非参数检验	127	8.2.1 简单回归方程的求解	178
6.1 非参数检验简介	127	8.2.2 回归方程拟合程度检验	179

8.2.3 用回归方程预测	180	10.2.3 二分类数据	232
8.2.4 简单线性回归举例	181	10.3 类之间的距离	232
8.3 多元线性回归	183	10.4 系统聚类算法过程	233
8.3.1 多元线性回归方程简介	183	10.5 系统聚类案例	234
8.3.2 多元线性回归方程的显著性检验	183	10.6 K-均值聚类	238
8.3.3 应用举例	184	10.6.1 K-均值法简介	238
8.3.4 线性回归自变量进入的方式	187	10.6.2 K-均值法案例	239
8.4 线性回归的诊断和线性回归 过程中的其他选项	189	10.7 两步法聚类	242
8.4.1 回归分析的前提条件	189	10.7.1 两步法简介	242
8.4.2 回归分析前提条件的检验	190	10.7.2 两步法案例分析	243
8.4.3 回归诊断	192	10.8 聚类分析注意事项	247
8.5 非线性回归	197	10.9 小结	248
8.6 曲线估计	204	思考与练习	248
8.7 小结	207	参考文献	248
思考与练习	207	第 11 章 主成分分析	249
参考文献	207	11.1 主成分分析简介	249
第 9 章 方差分析	209	11.1.1 主成分分析的目的与功能	249
9.1 方差分析的术语与前提	209	11.1.2 主成分分析的数学理论	250
9.2 单因素的方差分析	210	11.2 主成分分析的应用条件	251
9.2.1 描述性数据分析	211	11.2.1 Bartlett 球形检验	251
9.2.2 单因素方差分析	211	11.2.2 KMO 统计量	252
9.3 多因素方差分析	215	11.2.3 基于相关系数矩阵还是 协方差矩阵	253
9.3.1 多因素方差分析简介	215	11.3 主成分分析案例	253
9.3.2 多因素方差分析举例	216	11.3.1 综合评价案例	253
9.4 协方差分析	220	11.3.2 主成分分析用于探索 变量间结构关系	263
9.4.1 协方差分析简介	220	11.4 小结	265
9.4.2 协方差分析案例分析	221	思考与练习	265
9.5 小结	227	参考文献	266
思考与练习	228	第 12 章 因子分析	267
参考文献	228	12.1 因子分析简介	267
第 10 章 聚类分析	229	12.2 因子分析的统计理论	268
10.1 聚类分析简介	229	12.2.1 因子分析的模型	268
10.2 个案间的距离	230	12.2.2 因子分析模型的求解方法	269
10.2.1 定距数据 (Scale Measurement) 距离定义方式	230	12.2.3 因子分析的应用前提	271
10.2.2 分类数据的频数数据 (Count) 之间的距离	231	12.2.4 因子个数的确定	271
		12.2.5 因子的解释	272

12.2.6	因子旋转	273	12.4	因子分析结果的有效性	286
12.2.7	因子得分	275	12.5	因子分析和主成分分析的比较	286
12.3	因子分析案例	275	12.6	小结	287
12.3.1	探索变量间的结构关系	275		思考与练习	287
12.3.2	因子分析在市场调查中的应用	281		参考文献	288

第 1 章

统计学和 SPSS 统计分析软件简介

本章学习目标：

- 统计分析的基本概念和有关术语；
- 统计数据的分类；
- 常见的统计分析软件以及它们的特点；
- 明确 SPSS 统计分析软件的主要应用领域；
- 明确 SPSS 主要模块及功能；
- 了解 SPSS 的安装方式和 SPSS 的界面，掌握 SPSS 运行的几种方式；
- 学习 SPSS 帮助系统。

1.1 统计分析的基本概念

统计学是研究收集数据、整理数据，总结和分析数据中所蕴含的信息，并且据此得到结论以解决特定问题或者辅助决策的一门科学。同时，统计学可以提供衡量决策的置信程度（决策的置信区间和置信度）。按照大统计学家 C. R. Rao 的说法：“今天，统计学已发展成为一门媒介科学。它研究的对象是其他学科的逻辑和方法论——作出决策的逻辑和试验这些决策的逻辑。统计学的未来依赖于向其他学习领域内的研究者正确传授统计学的观点；依赖于如何能够在其他知识领域内将其主要问题模式化。”^①

统计学研究的对象是数据，它是用于得到结论或者作出决策所依据的“事实”或者“证据”。数据的一个重要特征是它的变动性。一个班级中所有学生的身高都相同吗？当然不同！它们是因人而异的。每个人的头发的颜色相同吗？答案也是否定的。因此，不同的个体之间是有变化的。统计分析的目的之一就是描述数据的变动性，并且寻找和理解引起数据变化的原因。

统计学所研究的对象的全体所组成的集合称为总体（population）。在实际中，总体是很难获得的。有时候总体过于庞大，我们不可能得到研究对象的所有成员。例如研究上海市 18 岁的青少年的身高，那么上海市所有 18 岁的青少年的身高将是总体，我们很难得到总体的所有数据。另外一种情况下，获取总体数据的成本太大，或者道德因素等不允许获取。例如，研究某种治疗癌症的新药的效果，那么研究的对象应该是全体服用该新药的癌症病人，然而在明确该药的确有效之前，不可能把该新药应用于所有癌症病人。基于以上原因，总体数据总是难以获得的。我们一般选取总体中有代表性的一个子集合进行研究，该子集称为样本（sample）。大部分情况下，数据分析是基于样本的。

^① C. R. 劳. 统计与真理：怎样运用偶然性 [M]. 北京：科学出版社，2004.

统计学上常常通过一些数值指标来描述总体的特征，这些指标称为总体的参数（parameter）。用来描述样本特征的指标称为统计量（statistic）。一般而言，统计量是样本所蕴含信息的概括和浓缩。

统计学可以分为描述性统计（descriptive statistics）和推断性统计（inferential statistics）。描述性统计主要进行数据的整理和汇总，这个过程可以通过描述性统计量、统计图和统计表等工具来进行。而推断性统计则是关于把从样本数据中得到的结论推广到总体，并衡量该结果应用于总体数据的可靠性的方法的集合。

1.1.1 统计分析的步骤

一般进行一次完整的统计分析，需要经过以下过程。

（1）认识研究的问题，明确研究目标。

（2）收集和研究目标有关的数据。获取总体数据一般是很困难或者比较昂贵的，典型的做法是获取总体的一个子集，即样本。数据收集在统计分析过程中是十分重要的。如果收集的数据是不正确的，那么所有的分析都是没有任何意义的，即所谓“分析的是错误数据，得到的是错误结论（garbage in, garbage out）”。

（3）对数据进行描述性统计分析。通过描述性统计分析指标或者统计图表，对数据有一个概括的认识，从中洞察可以应用的分析方法。进行统计推断，应用适当的统计方法，把从样本得到的结论推广到研究对象的总体，并且给出该结论的可信程度。

（4）应用分析结果进行决策。

进行一项统计分析项目，一般都要经过以上4个步骤。我们学校学习的统计学教材主要讲解步骤（3）和步骤（4）。步骤（2）一般在《抽样调查》课程中讲解。

上述过程可以用图1-1表示如下。

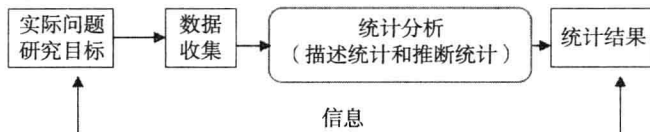


图 1-1 统计分析过程图

1.1.2 数据的类型

我们一般把分析的个体的特征称为变量。每一个个体的特征的具体取值即为该变量的观测值，即数据。根据个体特征的不同，数据（或者变量）也呈现出不同的特征。如果研究银行客户的行为特征，我们关心的用户特征有姓名、年龄、性别、受教育的年数、个人收入水平、家庭收入、工作年数、存款余额等。

这里姓名、性别两个特征是定性变量（Qualitative Variables），或者称为分类变量。性别有两个合法的取值：男和女，它们只是表示个体的特征不同，但是没有具体的数量意义。而年龄、受教育的年数、个人收入水平、家庭收入、工作年数、存款余额等则不同，它们的不同取值除了表示个体的特征不同之外，还具有具体的数量意义，这些变量被称为定量变量（Quantitative Variables）。比如年龄变量，某个体年龄 18 岁，表示该个体的年龄特征的取值为 18，并且该取值具有具体的数量意义，18 岁是小于 19 岁，并且比 17 岁的个体大 1 岁。而定性变量则没有这一层数量意义。

根据取值的特征，定量变量又分为离散变量（Discrete Variables）和连续变量（Continuous

Variables)。离散变量的所有可能的取值是有限的，或者所有可能的取值是可以一一列举的；而连续变量理论上任何两个取值之间都有无限多个可能的取值，可以精确到任意位数。例如上例中的年龄的取值一般为整数，比如 18 岁、19 岁、20 岁等，是可以列举出来的。而存款余额的取值一般不是可以列举的，理论上存款 1900.01 元和 2000 元之间可以有无限个可能的存款余额值。

变量的分类关系如图 1-2 所示。

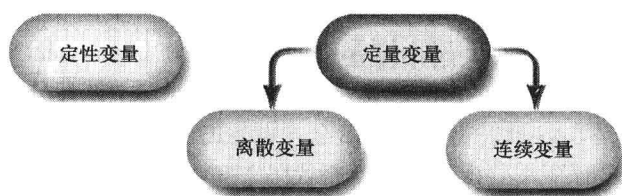


图 1-2 变量类型

根据变量的类型，统计数据也相应具有度量水平。定性变量对应的数据的度量水平为分类数据。根据取值是否具有内在的大小关系，分类数据又分为名义数据（Nominal Data）和定序数据（Ordinal Data）。定量变量对应的数据的度量水平为尺度数据（Scale Measurement）。对于尺度数据，根据 0 值实际所代表的意义，又分为区间数据和比例数据。在区间数据中，0 值不代表没有，数据的比值没有实际的意义。以温度数据为例，温度为 0 不代表没有温度。我们不能说温度 40℃ 比温度 20℃ 热两倍，因此温度为区间数据。而比例数据中，0 值是代表绝对的不存在。例如工资为 0，即代表没有工资，工资为 1000 元的人的工资则是工资为 500 元的人的工资的 2 倍。数据度量水平的分类如图 1-3 所示。

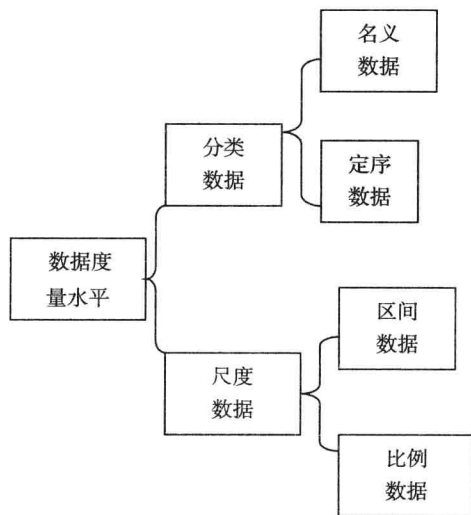


图 1-3 数据的度量水平

1.2 常见统计分析软件简介

1.2.1 SPSS

SPSS 统计分析软件是最常用的统计分析软件之一，它的具体名称为 IBM SPSS Statistics。我

们将在 1.3 节以及以后的章节中详细介绍该软件的特点和用法。

1.2.2 SAS

SAS 统计分析软件是最常用的统计分析软件之一。SAS 最早产生于美国北卡罗来纳州立大学的一个农业研究分析项目。1966 年, 美国农业部的研究项目产生了大量的数据。在美国国立卫生研究院 (National Institute of Health) 的一笔研究经费的资助下, 美国农业部资助的 8 所大学构成的“南部大学统计研究联合体 (University Statisticians Southern Experiment Stations-USSES)”开始开发一个通用统计软件包来分析农业相关的数据。该项目的成果“统计分析系统 SAS (Statistical Analysis System)”, 就是 SAS 软件名称和 SAS 公司诞生的基础。1972 年的 SAS 版本是真正意义上的发行量较大的版本。该版本共计有 37 000 行代码, 基本上是由两个人完成的, 他们是吉姆·古德奈特 (Jim Goodnight, 贡献了 32% 的代码) 和吉姆·巴尔 (Jim Barr, 贡献了 65% 的代码)^①。1976 年 7 月, 新的 SAS 版本发行, 成为统计计算和数据管理上完善的一个版本。同年古德奈特、巴尔、赫尔维格 (Helwig) 和塞尔 (Sall) 四个人离开北卡罗来纳州立大学创建 SAS 公司。

SAS 软件提供了从描述性统计到试验设计、方差分析、相关分析、回归分析、时间序列分析等多种推断性统计分析过程。SAS 分析方法的实现通过过程调用完成。许多过程同时提供了多种算法和选项。

SAS 最早是用编程方式来进行统计分析的, 相对于仅仅通过窗口对话框来操作的软件而言, 对于初学者来说学习的曲线相对较长。

1.2.3 Splus 或者 R

R 是有名的开源统计软件包, 最早由新西兰奥克兰大学统计系的 Robert Gentleman 和 Ross Ihaka 两个人所开发, 现在已经成为免费软件基金会的 GNU 项目。R 软件和最早起源于贝尔实验室的 John Chambers 开发的 S 语言类似, 可以认为是 S 语言的一种实现。S 的代码基本上不用修改就可以在 R 下运行。R 包含所有的基本统计分析功能、线性和非线性模型、经典的统计检验、时间序列分析、聚类分析, 以及高级的作图技巧等。另外, R 可以运行时调用 C 或者 Fortran 的代码来完成计算量很大的工作。R 还有用户开发的各种特殊用途的插件, 目前 (2011 年 10 月) 大约有 4 000 多个插件。SPSS 的 R 插件可以把 R 的功能和 SPSS 统计分析软件集成在一起, 充分利用 SPSS 软件的易用性和 R 的统计方法与强大的作图功能。

1.2.4 其他数据分析软件

其他常见的统计分析软件有 JMP、Minitab、Stata 等。另外, Matlab 等语言也常常成为许多统计分析工作者喜爱的数据分析软件。

1.3 SPSS 统计分析软件的发展

SPSS 是软件英文名称的首字母缩写, 最初为 Statistical Package for the Social Sciences 的缩写, 即“社会科学统计软件包”。随着 SPSS 产品服务领域的扩大和服务深度的增加, SPSS 公司已于 2000 年正式将英文全称更改为 Statistical Products and Service Solutions, 意为“统计产品与服务解决方案”, 标

^① 参见 http://www.barrsystems.com/about_us/the_company/sas_history.asp。

志着 SPSS 的战略方向作出了重大调整。2009 年 3 月 19 日, SPSS 公司将 SPSS 四大系列产品 (Statistics Family、Modeling Family、Data Collection Family、Deployment Family) 整合到一个综合分析平台, 把四类产品统一加上 PASW (为 Predictive Analysis Software 的首字母) 前缀, 喻义 SPSS 产品的发展方向为预测分析领域。此后 SPSS 把正在发行的 SPSS 17 统计分析软件正式更名为 PASW Statistics 17, 此后发行的版本 18 的官方名称为 PASW Statistics 18。同年的 10 月 2 日, IBM 宣布完成收购 SPSS 公司, 随后 SPSS 统计分析产品更名为 IBM SPSS Statistics。本书写作时用到的软件为 PASW Statistics 19。现在已经更新为 IBM SPSS Statistics 20, 也是目前的最新版本。虽然产品名称历经变迁, 但是软件自身的统计分析功能变化不大。在本书中, 还是采用简单的名称: SPSS 或者 SPSS Statistics。

SPSS 是世界上最早的统计分析软件, 是美国斯坦福大学的三位研究生于 20 世纪 60 年代末开发出来的。Norman Nie——SPSS 软件的三位创始者之一, 当时是斯坦福大学政治学的博士研究生, 为了分析从多个国家收集到的几千份调查问卷, 与 Bent (斯坦福大学运筹学方向研究生)、Hull 一起开发了一套自动化处理数据和输出统计分析结果的程序。他们开发的第一个版本于 1968 年正式发布。一开始, SPSS 就以其丰富的、高质量的文档而被广泛传播和应用。随着 SPSS 销售的迅速增长, SPSS 软件的两位创始人——Norman Nie 和 Hull 于 1975 年在芝加哥成立了 SPSS 公司。

1984 年 SPSS 公司首先推出了世界上第一个基于个人电脑的统计分析软件 SPSS/PC+, 开创了 SPSS 微机系列产品的开发方向, 极大地扩充了它的应用范围, 并使其能很快地应用于自然科学、社会科学等各个领域。世界上许多有影响的报刊纷纷就 SPSS 的自动统计绘图、深入的数据分析、使用方便、功能齐全等方面给予了高度的评价与称赞。迄今, SPSS 软件已有 40 余年的成长历史。全球约有 25 万家产品用户, 它们分布于通信、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等多个领域和行业。SPSS 是世界上应用最广泛的专业统计分析软件之一。

1.4 SPSS 版本和授权

截至目前 (2010 年 3 月), SPSS 的最新版本为 PASW Statistics 18 (尽管官方已经更名为 IBM SPSS Statistics, 软件发行早于新名称, 因此仍然沿用名称 PASW Statistics)。SPSS 软件的升级相对比较有规律, 基本上每年发行一个新版本。SPSS 软件的最近发行历史为:

SPSS 11.0.1——2001 年 11 月发布;

SPSS 12.0——2003 年发布;

SPSS 13.0——2004 年发布;

SPSS 14.0——2005 年发布;

SPSS 15.0.1——2006 年 11 月发布;

SPSS 16.0.2——2008 年 4 月发布;

SPSS Statistics 17.0.1——2008 年 12 月发布;

PASW Statistics 17.0.2——2009 年 3 月发布;

PASW Statistics 18.0——2009 年 8 月发布;

IBM SPSS Statistics 19.0——2010 年 8 月发布;

IBM SPSS Statistics 20.0——2011 年 8 月发布。

其中, SPSS 版本 14 和版本 16 都有官方的中文版发行。从版本 17 开始, SPSS 把所有支持的语言集成到一起, 可以在选项中选择 11 种语言的任何一种版本。

SPSS 安装完成之后需要有授权号码才能正常运行。SPSS 17 有自带的试用授权，试用期为 1 个月。SPSS 18 的试用期限为 3 周，而 SPSS 19 的试用期限为 2 周。SPSS 程序安装完成后，会要求输入授权号码或者许可证。许可证授权向导允许您获取一个 PASW Statistics 许可证。如果您没有立刻获取许可证，可以启用临时试用许可证，临时许可证可以启用所有 SPSS 的高级模块。产品试用期从首次应用许可证开始，到期后 PASW Statistics 将不再运行，您必须获取许可证才能继续使用。许可证通过锁定代码（Lock Code）绑定计算机硬件。如果更换了计算机或其硬件，则需要新锁定代码并再次进行授权过程，才能继续使用 SPSS 软件。如果超出了许可证协议中规定的可允许授权数量，授权将失败。另外，许可证对时间变化敏感。如果更改了系统时间，将会导致 SPSS 软件运行失败。



1. 必须获取许可证或启用临时试用许可证才能使用 PASW Statistics。
2. 安装正式授权软件之后，不要轻易更改系统时间，否则会导致软件运行失败。

1.5 SPSS 统计分析软件的特点

SPSS 是世界上最早采用图形菜单驱动界面的统计软件，最突出的特点就是操作界面友好，输出结果美观。它将几乎所有的功能都以统一、规范的界面展现出来，使用 Windows 窗口展示各种管理和分析数据的功能，以对话框方式展示各种功能选择项。用户只要掌握一定的 Windows 操作技能，粗通统计分析原理，就可以使用该软件为特定的科研工作服务，或者进行企业级的数据分析。

SPSS 采用类似 Excel 表格的方式输入与管理数据，数据接口十分通用，能方便地从任何类型的数据文件或者数据库中读入数据。SPSS 统计过程既包括了常用的、成熟的统计过程，也包含了一些高级的统计分析方法，例如 Bootstrapping 方法、神经网络、市场直销方法等，完全可以满足专业人士的工作需求。SPSS 输出结果十分美观，采用 SPSS 专有的 SPO 格式存储结果，也可以直接将结果转存为 HTML 格式、文本格式、PDF 格式或者 PPT。对于熟悉老版本编程方式运行 SPSS 的用户，SPSS 还特别设计了语法生成窗口，用户只需在菜单中选好各个选项，然后单击【粘贴】按钮，就可以自动生成标准的 SPSS 程序。极大地方便了中、高级用户的使用。

SPSS 的主要特点如下。

(1) 操作简单：除了数据录入及部分语法命令程序需要键盘键入外，大多数操作可通过“菜单”、“按钮”和“对话框”来完成。使用者只需要掌握简单的 Windows 操作技巧，便可应用 SPSS 软件进行统计分析。

(2) 无须编程：具有第四代语言的特点，只需告诉系统要做什么，无需说明要怎样做。只要了解统计分析的原理，而无需通晓各种统计算法，便可得到需要的统计分析结果。对于常见的统计方法，SPSS 的命令语句、子命令及选择项的选取大多可通过“对话框”操作完成。因此，用户无需花大量时间记忆大量的命令、过程、选择项，从而避免了漫长的学习过程。同时，熟悉或精通编程者，如果喜欢，可以通过编程来实现窗口和对话框分析的所有功能。

(3) 功能强大：具有完整的数据输入、编辑、统计分析、报表、图形制作等功能。自带 11 种类型共计 136 个函数。SPSS 提供了从简单的统计描述到复杂的多元统计分析方法，比如：数据的探索性分析、统计描述、列联表分析、二维相关、秩相关、偏相关、方差分析、非参数检验、多元回归、生存分析、协方差分析、判别分析、因子分析、聚类分析、非线性回归、

Logistic 回归等。随着版本的更新, SPSS 的功能不断完善, 比如: 在 SPSS 16 版本中增加了神经网络模块, 在 SPSS 17 版本中增加了 EZ RFM 分析, 在 SPSS 18 版本中新增了 Bootstrapping 分析。

(4) 方便的数据接口: 能够读取及输出多种格式的文件。比如: 由 dBASE、FoxBASE、FoxPRO 产生的*.dbf 文件, 文本编辑器软件生成的 ASC II 数据文件, Excel 的*.xls 文件等均可转换成可供分析的 SPSS 数据文件。同样地, 能够把 SPSS 的图形转换为 7 种不同格式的图形文件。SPSS 的输出结果可保存为 txt、PDF、Word、PPT 或者 HTML 格式的文件。

(5) 灵活的功能模块组合: SPSS for Windows 软件分为若干功能模块。用户可以根据自己的分析需要和计算机的实际配置情况灵活选择。

(6) 与其他程序的无缝结合: SPSS 新版本(版本 16 或者更高的版本)可以调用开源统计分析软件 R 或者开源高级程序语言 Python 的功能模块, 实现诸如支持向量机(SVM)、关联分析、偏最小二乘回归分析等功能。



SPSS 统计软件版本 17 和版本 18 更名为 PASW Statistics。IBM 购买了 SPSS 公司之后, 统计分析产品的名称为 IBM SPSS Statistics。从版本 19 开始, 产品正式更名为 IBM SPSS Statistics。SPSS 公司的官方网站从 2011 年初也正式迁移到 IBM 平台。

1.6 主要模块及功能简介

SPSS Statistics 主要包含 17 个模块, 这些模块的组合丰富了 SPSS Statistics 的分析功能。这 17 个模块分别介绍如下。

1. IBM SPSS Statistics

SPSS Statistics 是 SPSS Statistics 软件运行的平台, 确保您能综合使用 SPSS Statistics 软件的其他功能模块和 Statistics 家族的产品。在这个平台上, 您可以完成任意需求的数据分析。该模块是从 SPSS 18 版本才开始有的, 在以前版本中, 它是和 SPSS Statistics Base 模块在一起的。

2. IBM SPSS Statistics Base

SPSS Statistics Base 模块是 SPSS Statistics 软件的基础模块, 提供数据访问、数据管理和准备、数据分析和报告、统计图表等功能。含有基本的统计分析过程, 例如计数、交叉列表分析、描述统计、探索分析、均值比较、方差分析、相关性分析、非参数检验、多重响应分析、因子分析、线性回归、曲线估计、聚类分析、判别分析以及尺度分析等。

3. IBM SPSS Regression

SPSS Regression 是非线性建模分析程序, 使您能够应用高级、成熟的方法分析数据。当您预测行为和事件, 而数据不满足线性回归假设时, 可利用多项/二项 Logistic 回归、非线性回归、加权最小二乘、两阶段最小二乘、Probit 等回归方法实现。

4. IBM SPSS Advanced Statistics

SPSS Advanced Statistics 含有专门用以描述、拟合数据间内在复杂关系的统计方法, 可以使分析更为准确, 并得出更为可靠的结论。SPSS Advanced Statistics 提供了一组功能强大的单变量及多变量的高级分析技巧。SPSS Advanced Statistics 模块的多变量分析技术包括: 广义线性模型

(GZLMS)、混合模型、一般线性模型 (GLM)、方差成分估计、MANOVA、Kaplan-Meire 估计、Cox 回归、层次对数线性模型、对数线性模型、生存分析等。

5. IBM SPSS Custom Tables

SPSS Custom Tables 在创建表格的同时，能够实时更新，帮助分析人员在较少时间里，做出美观、精确的表格。利用 SPSS Custom Tables 可以展现调查、客户满意度、投票选举等的结果分析。灵活的交互功能、创建表格时的可预览性，及其统计推断和数据管理的能力，可以帮助用户方便清楚地了解结果。

6. IBM SPSS Categories

SPSS Categories 可以图形化展示数据中的潜在关系，通过启发性的概念映射、最优尺度、偏好尺度和数据降维技术，揭示数据中全部潜在的关系。SPSS Categories 为分析人员提供了深入分析复杂分类数据和高维数据的全部工具。通过 SPSS Categories 可以直观地解释数据，了解大型表中的计数、分级和排序中的关联情况。

7. IBM SPSS Exact Tests

为了确定变量之间的关系，研究人员往往首先查看交叉表或非参数检验中的 P 值。如果数据满足潜在的假设，用传统的计算方法是可行的。但是，如果数据属于小样本，或数据变量中很高的比例集中于某一类别，或者不得不将数据细分为几个类别，传统的检验方法将不能得出正确结论。SPSS Exact Tests 可以得到更为准确的结果。

8. IBM SPSS Missing Values

利用 SPSS Missing Values 填充缺失数据，建立更佳模型，得到更有效的结论。缺失值可能会严重影响分析结果，如果把它们忽略或者计算时排除它们，很可能会得出不正确的结论。SPSS Missing Values 是每位关心数据有效性的分析人员的有力工具，该模块提供 6 种缺失值诊断报告，使分析人员可以从多个角度检查数据，发现数据缺失模式。然后，分析人员可以估计摘要统计量，并利用统计方法填充缺失值。

9. IBM SPSS Conjoint

SPSS Conjoint 是用来模拟消费者决策过程的研究工具。SPSS Conjoint 能加强使用者对消费者偏好的理解，更有效地进行产品设计、定价和营销；帮助使用者衡量产品或服务在消费者心中的位置。具备了这些知识，公司在设计产品时，就能把对于目标市场最重要的属性特征包含进来，根据这些属性值进行定价，并专注于最有可能吸引目标购买者的点。即使市场上的竞争者、产品和价格随时间发生改变，依然可以继续利用由 SPSS Conjoint 得出的结果来模拟情况发生变动后的市场。这样在投入大量资源进行产品开发和营销活动前，就能够提前预测市场的响应。

10. IBM SPSS Complex Samples

SPSS Complex Samples 提供了专门的统计工具，帮助使用者计算出复杂抽样设计的统计量和标准误差。绝大多数常规的统计软件都假定数据是通过简单随机抽样取得的，而简单随机抽样在大多数大规模调查中既不现实也不经济。此外，用常规统计分析方法分析此类样本数据有得到错误结果的风险。例如，统计数据的估计标准误差经常太小，易导致对精度产生错误的认识。SPSS Complex Samples 将抽样设计融合到调查分析中，因此能在由复杂抽样得到的总体中获得更加有效的统计推论。

11. IBM SPSS Decision Trees

SPSS Decision Trees 能识别群体及预测结果。该模块能够直接创建分类决策树，帮助快速并