

SPSS 19

(中文版)

统计分析实用教程

© 邓维斌 唐兴艳 编著
胡大权 周玉敏

SAS

统计分析教材



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

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

统计分析教材

SPSS 19 (中文版)

统计分析实用教程

邓维斌 唐兴艳 编著
胡大权 周玉敏

电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书基于最新 SPSS 19 中文版进行编写, 以统计分析的实际应用为主线, 在对主要统计分析方法的基本概念和统计学原理进行简明介绍的基础上, 以 54 个实例为载体对 SPSS 19 中各种分析方法的应用场合和操作过程进行了清晰说明, 并对相关领域的 23 个统计分析典型案例进行了应用方法及解决思路的详细分析。全书共有思考与练习题 63 个, 以供巩固学习效果和课后练习。

全书内容覆盖了 SPSS 19 中常用的统计分析方法, 共 12 章, 第 1 章介绍 SPSS 的基础知识; 第 2 章介绍统计数据的收集与预处理; 第 3 章介绍图表的创建与编辑; 第 4~12 章介绍 SPSS 19 的各种统计方法, 包括描述性统计分析、参数估计与假设检验(包括参数与非参数检验)、方差分析、相关分析、回归分析、聚类和判别分析、主成分分析和因子分析、时间序列分析、信度分析。与教材配套的资源包括所有实例、典型案例和习题的数据文件, 课程 PPT 教案, 以及思考与练习题的参考答案, 可登录华信教育资源网免费下载。

本书可作为高等院校经济学、管理学、教育学、心理学等相关专业本科生和研究生教材, 也可供从事统计分析和决策的各领域工作者学习参考。

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 19(中文版)统计分析实用教程/邓维斌等编著. —北京: 电子工业出版社, 2012.3

统计分析教材

ISBN 978-7-121-16132-2

I. ①S… II. ①邓… III. ①统计分析—软件包, SPSS 19—高等学校—教材 IV. ①C819

中国版本图书馆 CIP 数据核字(2012)第 034858 号

策划编辑: 秦淑灵

责任编辑: 秦淑灵

印 刷: 三河市双峰印刷装订有限公司

装 订:

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: 787×980 1/16 印张: 18.75 字数: 479 千字

印 次: 2012 年 3 月第 1 次印刷

印 数: 4000 册 定价: 38.00 元

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。

前 言

面对大量的市场调查、统计数据，必须借助分析工具才能得出科学结论，掌握数据分析方法和数据分析工具已成为当代大学生的必备技能。SPSS 统计分析软件以其易用性和强大功能已成为目前最流行的统计分析工具之一，在国内具有很大的用户群，是目前国内进行管理决策、市场分析、社会学、医学统计、金融等专业的统计分析人员应用最广泛的软件。2009 年 IBM 对 SPSS 收购后，已推出 SPSS 19 的多国语言版本（包括简体中文），在功能和操作界面上做了较大幅度的升级改动。

作者经过近 10 年的教学实践，发现学生在学习本门课程时有几个突出的问题：（1）绝大多数 SPSS 教材基于 SPSS 英文版软件编写，学生学习时有很大的语言障碍；（2）遇到具体问题不知道该用何种分析方法，即不知 SPSS 所提供的各种统计分析功能究竟适用于解决何种实际问题；（3）不知每一种分析方法的具体操作、分析步骤该如何进行；（4）不知如何组织数据，即如何将已有的数据组织成适合用 SPSS 特定分析方法的数据格式，如怎么定义变量，如何进行分组等；（5）不知如何对输出结果（包括表和图形）进行分析，如涉及假设检验的地方，分不清原假设和备选假设，不知如何根据所输出的统计量及概率值对其进行假设检验。

针对这些问题，我们在总结已有讲义、实验指导书和教材的经验和不足的基础上，编撰成了适合国内读者学习的、基于中文版 SPSS 19 的教材，该教材有如下几个突出的特点：

（1）基于 SPSS 19 中文版软件，典型案例和习题丰富。

本书以 IBM SPSS Statistics 19 简体中文版为蓝本进行编写，扫除了广大国内读者对英文版教材学习的语言障碍。全书编写以 SPSS 的实际应用为主线，组织了 54 个实例对各项统计分析方法进行介绍，并对相关领域的 23 个统计分析典型案例进行了应用方法及解决思路等的详细分析，全书共有思考与练习题 63 个，以供巩固学习效果和课后练习。

（2）全书结构清晰，体系完整，内容精简明了。

在总体内容把握上，按照“SPSS 19 概述—数据组织—图形功能—统计分析”的顺序组织，由浅入深、由基础到专业。在每章内容的安排上按“分析方法简介—统计原理与步骤—统计实例分析—典型案例—思考与练习”的顺序组织，方便读者学习。全书内容涵盖了 SPSS 中最常用的统计分析方法。同时避免了大而全的介绍，只针对最常用的统计功能进行阐述，使读者在有限的时间内学习到更多实用功能。

（3）统计分析方法、SPSS 操作和案例分析的有机结合。

从快速掌握和应用 SPSS 的角度出发，作者将 SPSS 各项功能的操作介绍与统计分析方法论述有机结合。对各种统计分析方法的原理进行了通俗易懂的介绍，但又避免了纷繁复杂的数学证明过程，使读者可以了解分析方法的核心思想，掌握方法的正确应用范围。以 54 个实例为载体，比较详细地介绍了 SPSS 中各项常用功能菜单和相应对话框的具体意义和适用情况。最后通过多领域的大量分析案例将 SPSS 的操作和统计分析方法进行有机结合。

(4) 加强对特定问题的分析, 以及对数据组织方法和分析结果的讨论。

在用 SPSS 对每个案例进行操作之前, 增加了“分析”步骤, 即针对每个具体问题, 对为什么要使用该种分析方法进行了解释和说明, 在“数据组织”步骤对特定分析方法如何组织数据做了说明, 并对每个案例的主要运行结果进行了详尽的解释和讨论。特别对易混淆的问题以注释的方式进行了说明, 以方便读者对相关概念和问题进行区别和理解。

本书可供高等院校相关专业的本科生、研究生作为教材使用, 也可作为 SPSS 统计分析培训和自学教材。另外, 在统计分析或科研中需处理数据的人士也可以参考。与教材配套的资源有所有实例、典型案例和习题的数据文件, 课程 PPT 教案, 部分思考与练习题的参考答案, 可登录华信教育资源网 <http://www.hxedu.com.cn> 免费注册下载。

本书由重庆邮电大学胡大权老师执笔编写第 1, 3, 6 章, 周玉敏老师编写第 2 章, 唐兴艳老师编写第 4, 5, 8 章, 邓维斌老师编写第 9, 10, 11 章, 费玲玲老师编写第 7, 12 章, 由邓维斌负责全书的统稿。在本书的编写过程中, 有着丰富教材编写经验的万晓榆教授、刘曜教授、杜茂康教授给予了极大的帮助和指导, 黄蜀江、刘进、陈家佳、张文彬、刘友军、陈文沛、刘小平等老师也提出了大量的建设性意见和建议, 杨双霞、刘慧兵等研究生做了大量校对工作, 在此表示衷心的感谢。

该书编写过程中借鉴了多种相关书籍, 引用了一些宝贵的资料, 在此向书籍作者表示深切谢意。本书仅就 SPSS 19 中常用的统计分析方法进行了介绍, 书中所论并不完美, 错误和疏漏之处, 恳请读者批评指正, 也欢迎切磋探讨。笔者 E-mail: dengwb@cqupt.edu.cn。

编者

2011 年 12 月

目 录

第 1 章 统计分析软件概述	1	2.3 数据文件的编辑	31
1.1 SPSS 简介	1	2.3.1 数据文件的合并	31
1.1.1 SPSS 的发展	1	2.3.2 数据文件的拆分	33
1.1.2 SPSS 统计分析软件的特点	2	2.3.3 数据的选取	35
1.1.3 SPSS 19 的新特性	4	2.3.4 数据的加权	37
1.1.4 SPSS 的模块	4	2.4 SPSS 数据加工	38
1.2 SPSS 使用基础	7	2.4.1 变量的计算	38
1.2.1 SPSS 的安装	7	2.4.2 数据可视离散化	40
1.2.2 SPSS 的界面	8	2.4.3 数据重新编码	43
1.3 SPSS 的帮助系统	11	思考与练习	45
1.3.1 主题	11	第 3 章 图表的创建与编辑	47
1.3.2 教程	12	3.1 SPSS 的图形功能概述	47
1.3.3 个案研究	12	3.1.1 SPSS 创建图形的一般过程	47
1.3.4 统计辅导	13	3.1.2 图形生成与数据文件结构	47
1.3.5 高级帮助	13	3.1.3 图形生成与数据的度量尺度	48
1.4 利用 SPSS 进行数据分析的步骤	15	3.2 图表的创建	48
1.4.1 统计学中数据分析的 一般步骤	15	3.2.1 使用图表构建程序创建	48
1.4.2 利用 SPSS 进行数据分析的 一般步骤	16	3.2.2 使用图形画板模板选择程序 创建	52
第 2 章 统计数据的收集与预处理	17	3.2.3 使用旧对话框创建	54
2.1 统计数据的收集	17	3.3 图表的编辑	56
2.1.1 问卷设计	17	3.3.1 图表编辑器布局	56
2.1.2 问卷分析	20	3.3.2 图表编辑基本方法	57
2.2 数据文件的建立	21	3.3.3 图表基本设定	58
2.2.1 统计数据的度量尺度	21	3.3.4 图表高级设定	59
2.2.2 SPSS 数据文件的特点	22	思考与练习	59
2.2.3 输入数据建立数据文件	23	第 4 章 描述性统计分析	60
2.2.4 从其他数据文件导入数据建立 数据文件	29	4.1 基本描述性统计量简介	60
		4.1.1 描述集中趋势的统计量	60
		4.1.2 描述离散程度的统计量	61

4.1.3	描述总体分布形态的统计量	62	5.4.2	单样本 T 检验 SPSS 实例分析	93
4.2	频率分析	63	5.5	独立样本 T 检验	95
4.2.1	基本概念及统计原理	63	5.5.1	基本概念及统计原理	95
4.2.2	SPSS 实例分析	63	5.5.2	独立样本 T 检验 SPSS 实例分析	96
4.3	描述性分析	67	5.6	配对样本 T 检验	98
4.3.1	基本概念及统计原理	67	5.6.1	基本概念及统计原理	98
4.3.2	SPSS 实例分析	67	5.6.2	配对样本 T 检验 SPSS 实例分析	99
4.4	探索性分析	68	5.7	单样本的非参数检验	100
4.4.1	基本概念及统计原理	68	5.7.1	基本概念及统计原理	100
4.4.2	SPSS 实例分析	69	5.7.2	卡方检验	103
4.5	交叉表分析	75	5.7.3	二项分布检验	107
4.5.1	基本概念及统计原理	75	5.7.4	游程检验	110
4.5.2	SPSS 实例分析	76	5.7.5	单样本 K-S 检验	114
4.6	多重响应分析	81	5.8	独立样本非参数检验	117
4.6.1	基本概念及统计原理	81	5.8.1	基本概念及统计原理	117
4.6.2	多重响应分析 SPSS 实例分析	82	5.8.2	独立样本非参数检验 SPSS 实例分析	118
4.7	典型案例	84	5.9	相关样本的非参数检验	122
4.7.1	城市平均气温基本特征分析	84	5.9.1	基本概念及统计原理	122
4.7.2	商场电视品牌满意度调查	84	5.9.2	相关样本的非参数检验 SPSS 实例分析	124
	思考与练习	85	5.10	典型案例	127
第 5 章	参数估计与假设检验	86	5.10.1	蛋白饲料对小白鼠体重影响分析	127
5.1	参数估计	86	5.10.2	调控政策前后大中城市住宅销售价格指数差异性分析	127
5.1.1	点估计简介	86	5.10.3	某行业企业赢利比例判断	128
5.1.2	区间估计简介	87	5.10.4	棉条棉结杂质粒数分析	128
5.1.3	参数估计 SPSS 实例分析	87		思考与练习	129
5.2	假设检验	88	第 6 章	方差分析	131
5.2.1	基本概念及统计原理	88	6.1	方差分析简介	131
5.2.2	小概率事件原理	89	6.1.1	方差分析的概念	131
5.2.3	假设检验的一般步骤	90	6.1.2	方差分析的一般步骤	132
5.3	参数检验及非参数检验	91			
5.3.1	参数检验简介	91			
5.3.2	非参数检验简介	91			
5.3.3	参数检验及非参数检验比较	92			
5.4	单样本 T 检验	93			
5.4.1	基本概念及统计原理	93			

6.2	单因素方差分析	132	第 8 章	回归分析	174
6.2.1	基本概念及统计原理	132	8.1	回归分析简介	174
6.2.2	单因素方差分析 SPSS 实例 分析	133	8.1.1	回归分析的概念	174
6.3	多因素方差分析	139	8.1.2	回归分析的一般步骤	175
6.3.1	基本概念及统计原理	139	8.2	线性回归分析	176
6.3.2	多因素方差分析 SPSS 实例 分析	141	8.2.1	基本概念及统计原理	176
6.4	协方差分析	147	8.2.2	一元线性回归 SPSS 实例 分析	177
6.4.1	基本概念及统计原理	147	8.2.3	多元线性回归 SPSS 实例 分析	183
6.4.2	协方差分析 SPSS 实例分析	148	8.3	曲线估计	189
6.5	典型案例	151	8.3.1	基本概念及统计原理	189
6.5.1	培训材料效果分析	151	8.3.2	曲线估计 SPSS 实例分析	190
6.5.2	火箭射程影响因素分析	151	8.4	二元 Logistic 回归分析	193
6.5.3	三种治疗高血压的方法效果 分析	151	8.4.1	基本概念及统计原理	193
思考与练习		152	8.4.2	二元 Logistic 回归分析 SPSS 实例分析	194
第 7 章	相关分析	154	8.5	典型案例	199
7.1	相关分析简介	154	8.5.1	水稻产量影响因素分析	199
7.1.1	相关分析的概念	154	8.5.2	产品废品率的因素拟合	200
7.1.2	相关关系的种类	154	8.5.3	肾细胞癌转移的判断	200
7.2	两变量相关分析	155	思考与练习		201
7.2.1	基本概念及统计原理	155	第 9 章	聚类和判别分析	203
7.2.2	两变量相关分析 SPSS 实例 分析	158	9.1	聚类和判别分析简介	203
7.3	偏相关分析	160	9.1.1	基本概念	203
7.3.1	基本概念及统计原理	160	9.1.2	样本间亲疏关系的度量	204
7.3.2	偏相关分析 SPSS 实例分析	162	9.2	二阶聚类	205
7.4	距离分析	163	9.2.1	基本概念及统计原理	205
7.4.1	基本概念及统计原理	163	9.2.2	二阶聚类 SPSS 实例分析	206
7.4.2	距离分析 SPSS 实例分析	164	9.3	K-均值聚类	209
7.5	典型案例	170	9.3.1	基本概念及统计原理	209
7.5.1	有氧训练中的耗氧量研究	170	9.3.2	K-均值聚类 SPSS 实例分析	210
7.5.2	控制不良贷款	171	9.4	系统聚类	213
7.5.3	学生身体状况指标的相似性 分析	172	9.4.1	基本概念及统计原理	213
思考与练习		172	9.4.2	系统聚类 SPSS 实例分析	214
			9.5	判别分析	218

9.5.1	基本概念及统计原理	218	11.1.3	创建时间序列	253
9.5.2	判别分析 SPSS 实例分析	219	11.2	指数平滑法	255
9.6	典型案例	225	11.2.1	基本概念及统计原理	255
9.6.1	美国 22 家企业类型划分	225	11.2.2	指数平滑法 SPSS 实例 分析	256
9.6.2	销售地区的选择	226	11.3	ARIMA 模型	264
9.6.3	地区降水量区域类型判别	227	11.3.1	基本概念及统计原理	264
思考与练习		228	11.3.2	ARIMA 实例分析	266
第 10 章	主成分分析和因子分析	230	11.4	时序序列的季节性分解	273
10.1	主成分分析和因子分析简介	230	11.4.1	基本概念及统计原理	273
10.1.1	基本概念和主要用途	230	11.4.2	季节性分解的实例分析	273
10.1.2	主成分和公因子数量 的确定	232	11.5	典型案例	276
10.1.3	主成分分析与因子分析 的区别与联系	232	11.5.1	中国社会消费品零售总额 分析	276
10.2	主成分分析	233	11.5.2	中国彩电出口数据分析	277
10.2.1	基本概念及统计原理	233	11.5.3	城市温度的季节性分解	278
10.2.2	主成分分析 SPSS 实例 分析	234	思考与练习		278
10.3	因子分析	242	第 12 章	信度分析	280
10.3.1	基本概念及统计原理	242	12.1	内在信度分析	280
10.3.2	因子分析 SPSS 实例分析	243	12.1.1	基本概念及统计原理	280
10.4	典型案例	247	12.1.2	内在信度实例分析	281
10.4.1	医院工作质量评价分析	247	12.2	再测信度分析	286
10.4.2	各省、市、自治区城市市政 设施建设状况分析	248	12.2.1	基本概念及统计原理	286
思考与练习		250	12.2.2	再测信度实例分析	287
第 11 章	时间序列分析	251	12.3	Kendall 和谐系数	289
11.1	时间序列的建立和平稳化	251	12.3.1	基本概念及统计原理	289
11.1.1	填补缺失值	251	12.3.2	Kendall 和谐系数实例分析	290
11.1.2	定义日期变量	252	思考与练习		291
			参考文献		292

第 1 章 统计分析软件概述

在科学研究中，常常需要对大量的数据进行统计处理，这是一项细致而烦琐的工作，如果完全依靠手工来进行，工作量较大，且难以保证准确性，也达不到高精度。为了减轻整理和计算大量数据的负担，提高工作效率，我们必须充分利用现代化的技术手段。随着计算机软件技术的发展，计算机在分析数据方面发挥了相当大的作用，它功能多，速度快，计算精确，较易利用，并且计算机统计软件可以完成更为精确系统的数据分析与统计计算。

在资料统计处理中常采用的统计软件有 SPSS Inc.公司的 SPSS 统计软件系统、SAS 统计分析系统和 Microsoft 公司的 Excel 软件等。SPSS 是 Statistics Package for Social Sciences（社会科学统计软件包）的缩写，它是社会科学研究人员首选的统计软件，也是目前世界上最流行的统计软件之一。现随着服务领域的扩大和服务深度的增加，英文已更改为“Statistics Product and Service Solution”，意为“统计产品与服务解决方案”。SPSS 是在经济学、生物学、心理学、医疗卫生、体育、农业、林业、商业、金融等各个领域广泛应用的软件。它不仅能够实现统计功能，还能将分析结果用多种清晰简练的表格和数十种生动形象的二维、三维图形来表达，真正做到实用与美观的统一。

作为全书的开篇，本章介绍 SPSS 的基础知识，主要包括 SPSS 19 的新特性、主要功能模块、常用窗口、帮助系统的使用，以及利用 SPSS 进行数据分析的基本步骤。

1.1 SPSS 简介

1.1.1 SPSS 的发展

SPSS 统计软件系统最早是 1968 年由美国斯坦福大学的 3 位学生开发的一个软件包，基于这一系统，于 1975 年在芝加哥成立了 SPSS 公司。1984 年，SPSS 首先推出了世界上第一个统计分析软件微机版本 SPSS/PC+，它迅速占领了微机市场，扩大了自己的用户量，开创了 SPSS 微机系列产品的开发方向。

20 世纪 80 年代末，Microsoft 公司发布 Windows 操作系统后，SPSS 迅速向 Windows 移植。至 1993 年 6 月，正式推出 SPSS for Windows 6.0 版本。随后几乎每年推出一个更新版本，2009 年，SPSS 公司将 4 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 4 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，并且，从此开始成为多国语言版本，有了官方的中文界面及使用手册。随后，SPSS 公司被 IBM 收购成其子公司，将 SPSS 统计分析产品更名为 IBM SPSS Statistics。目前最新版本为 IBM SPSS Statistics 19。本书以 IBM SPSS Statistics 19 for Windows 为蓝本，结合统计学知识，对各领域常见统计分析案例进行分析讲述。

1.1.2 SPSS 统计分析软件的特点

随着 SPSS 的版本不断更新, 软件功能不断完善, 操作越来越简便, 与其他软件的接口也越来越多。SPSS Statistics 19 for Windows 具有以下特点。

1. Windows 风格, 界面友好

SPSS Statistics for Windows 最突出的特点就是操作界面友好, 输出结果美观。SPSS 是第一个采用人机交互界面的统计软件, 非常容易学习和使用。自从 1995 年 SPSS 公司与微软公司合作开发 SPSS 界面后, SPSS 界面变得越来越友好, 操作也越来越简单, 这就使熟悉微软公司产品的用户学习 SPSS Statistics 操作时, 很容易上手。

SPSS 界面非常友好, 熟悉的 Windows 风格界面, 数据视图也类似 Excel 布局。具有第四代语言的特点, 告诉系统要做什么, 无须告诉怎样做。只要了解统计分析的原理, 无须通晓统计方法的各种算法, 即可得到需要的统计分析结果。除了数据录入及部分命令程序等少数输入工作需要键盘输入外, 大多数操作可通过鼠标拖曳、单击“菜单”、“按钮”和“对话框”来完成。

SPSS 功能强大, 界面友好, 易学易用。SPSS 界面完全是菜单式, 使用下拉菜单来选择所需要执行的复杂的统计命令, 使用 Windows 的窗口方式展示各种管理和分析数据方法的功能, 使用对话框展示出各种功能选择项, 只要掌握一定的 Windows 操作技能, 粗略了解统计分析原理, 就可以使用该软件为特定的科研工作服务。开放式的命令语句窗口, 可以通过复制和粘贴的方法来学习和使用其“统计程序”句法语言, 同时也适合数据分析专家和研究员使用。

2. 易学易用

SPSS 易于操作, 易于入门, 结果易于阅读, 对统计软件的学习不会冲淡统计的主题, 这样研究人员就可以将精力集中在社会研究方法、市场研究方法、营销的业务问题上, 而不是忙于编程和统计。一般稍有统计基础的人经过几天的培训即可用 SPSS 做简单的数据分析, 包括绘制图表、简单回归、相关分析等。当然, 真正应用好 SPSS 软件的关键在于如何科学地设计研究方案、严谨地收集数据、严密深入地对数据进行统计分析及解释, 以及适度保守地下研究结论和进行决策。这一方面要求研究者掌握数理统计的基本知识, 另一方面也要求研究者多进行实践, 在实践中了解各种统计结果的实际意义。

从某种意义上讲, SPSS 软件还可以帮助数学功底不够扎实的用户学习运用现代统计技术。用户仅需要关心某个问题应该采用何种统计方法, 并初步掌握对计算结果的解释方法, 而不需要了解其具体运算过程, 就可以在使用手册的帮助下完成对数据的定量分析。现在很多用户只需要适当练习, 就能掌握简单的操作分析, 因此 SPSS 特别受非统计专业数据分析人员的青睐。

SPSS 采用类似 Excel 表格的方式输入与管理数据, 数据接口较为通用, 能方便地从其他数据库中读入数据, 包括常用的、较为成熟的统计方法, 完全可以满足非统计专业人士的工作需要, 是非统计专业人员的首选统计软件。

3. 功能全面

SPSS 针对初学者、熟练者及精通者都比较适用, 提供了数据获取、数据管理与准备、数据分

析、结果报告这样一个数据分析的完整过程，因此非常全面地涵盖了数据分析的整个流程，特别适合设计调查方案、对数据进行统计分析，以及制作研究报告中的相关图表。

此外，SPSS 具有完整的数据输入、编辑、统计分析、报表、图形制作等功能。仅 SPSS Base 模块就提供了从简单的统计描述到复杂的多因素统计分析方法，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、一元方差分析、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、聚类分析等常见的分析方法。

4. 强大的编程功能，支持二次开发

对于常见的统计方法，SPSS 的命令语句、子命令及选择项的选择等绝大部分采用“对话框”的操作方式。因此，用户无须花大量时间记忆大量的命令、过程、选择项。

由于 SPSS 19 具备强大的 Syntax 编程功能，且包括了 SPSS Programmability Extension 模块的编程扩展功能，那些熟练或精通者也较喜欢 SPSS，因为他们可以通过编程，在 SPSS 命令语法语言的基础上提供与其他编程语言的结合功能，来实现更强大的功能。例如，用其他语言编写的程序代码，如 Python 和 R，可以管理使用 SPSS 语法编写的任务流。使用 SPSS 19 提供的扩展编程功能和特性，让 SPSS for Windows 成为了最强大的统计开发平台之一。

5. 支持丰富的数据源，具备强大的数据访问和管理能力

SPSS 可以同时打开多个数据集，方便研究时对不同数据库进行比较分析和数据库转换处理。软件提供了更强大的数据管理功能，可帮助用户通过 SPSS 使用其他应用程序和数据库。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 ASCII 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。支持 Excel、文本、dBase、Access、SAS 等格式的数据文件。能够把 SPSS 的图形转换为 7 种图形文件。结果可保存为*.txt、word、PPT 及 html 格式的文件。

此外，通过使用 ODBC 的数据接口，可以直接访问以结构化查询语言（SQL）为数据访问标准的数据库管理系统，通过数据库导出向导功能可以方便地将数据写入到数据库中，等等。

在 SPSS Statistics 19 中，新增加了 Salesforce.com 的数据库驱动程序，允许分析人员访问 Salesforce.com 中的数据，就像访问 SQL 数据库中的数据一样。

6. 灵活的配置方案

对于商业用户而言，SPSS Statistics 是一种按照模块进行配置的软件产品，主要包括 SPSS Statistics Base 模块和其他一系列扩充功能模块。SPSS Statistics Base 是基础的软件平台，具备强大的数据管理能力和输入输出界面管理能力，并具备完备的常见统计分析功能，而其他每个独立扩充功能模块均在 SPSS Statistics Base 的基础上，为其增加某方面的分析功能。

7. 支持多种操作系统

客户端 SPSS 支持 Windows XP (32 位)、Windows Vista and 7 (32 位和 64 位)、Linux 和 Mac OS。服务器端 SPSS 支持 Windows Server 2003 (32 位和 64 位)、Windows Server 2008 (32 位和 64 位)、AIX、HP-UX、Solaris。

1.1.3 SPSS 19 的新特性

IBM SPSS 公司最新发行的 SPSS Statistics 19, 保留了 SPSS Statistics 软件一贯的界面风格, 并在原先的 18.0 版本基础之上进一步完善了分析功能, 提高了分析性能和交互能力, 新版本软件有如下增强和改进。

1. 广义线性混合模型

广义线性混合模型涵盖了从简单线性回归到复杂非正态纵向数据多变量模型的各种模型。在因变量与自变量呈非线性关系的情况下, 对因变量的预测更为精确。

2. 自动线性模型

自动线性模型使得用户无须深厚的统计知识即可自动创建功能强大的线性模型, 预测结果输出简单、准确。包括自动数据准备功能、增强模型稳定性和准确性的技术、提供分类预测的支持, 以及对大数据集的性能增强等。

3. 语法编辑器增强

新版本可以将编辑器窗格拆分为上、下两个窗格; 可以增加或减少语法块的缩进, 或者采用与所粘贴语法类似的格式自动缩进选定内容; 新的工具栏按钮允许取消注释, 并且新的选项设置允许将语法粘贴在光标位置; 还可以导航到下一个或上一个语法错误(例如不匹配的引号), 这样很容易在运行语法之前找到这些错误。

4. 自动设置测量级别

分析前可以根据数据特征自动设置测量级别。

5. 改进的直销模块

新的评分向导可以轻松应用预测模型对数据评分, 并且评分不再需要 IBM SPSS Statistics 服务器。更容易获得预先建立的模型, 以更好的方式给数据评分, 并可以从 [Salesforce.com](https://www.salesforce.com) 中获得客户信息。

提供更智能、更直观的结果输出。直销过程提供“智能”输出——帮助用户评估结果的简单、非技术性说明。

6. 更快的执行性能

建立数据透视表的输出速度比以前提升了两倍。因此, 当涉及大型报表输出或大量运算的报表时, 可大大节省做报告时间。此外, 输出报表所占用的内存空间也将更少。

1.1.4 SPSS 的模块

SPSS 统计分析软件是一款模块化设计的产品, 用户可以根据需要选择功能模块进行配置购买, 以节省费用。它主要包括 SPSS Statistics Base 模块和其他一系列扩充功能模块, 共 16 个, 每个独立扩充功能模块均可在 SPSS Statistics Base 模块基础上, 为其增加某方面的分析功能。下面简要介绍 16 个模块的功能。

1. Statistics Base 模块

Statistics Base 模块管理整个软件平台，以及数据访问、数据处理和输出，并能进行很多种常见的基本统计分析和报告，其中包括计数、交叉表和描述统计、OLAP 立方和码本报告。它还提供了多种降维、分类和细分方法，例如因子分析、聚类分析、最近邻元素分析和判别函数分析。此外，SPSS Statistics Base 模块还提供了广泛的平均值比较算法和预测方法，例如 T 检验、方差分析、线性回归和序数回归。

2. Advanced Statistics 模块

Advanced Statistics 模块为顺序结果建立更灵活、更成熟的模型，在处理嵌套数据时可得到更精确的预测模型，可以分析事件历史和持续时间数据。它包括广义线性模型 (GZLMS)、广义估计方程 (GEES)、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、多因子系统模式的对数线性模型、对数线性模型、生存分析。

3. Bootstrapping 模块

Bootstrapping 模块可以更有效地使用小样本量的数据，通过数据自身重采用的功能，可以模拟大样本情况下的采样结果，从而对数据结构特征和偏差有更直接的认识。该方法可以导出稳健的标准误估计值，并能为诸如均值、中位数、比例、比率、相关系数或回归系数等估计值导出置信区间。

4. Categories 模块

Categories 模块用启发性的二维图和感知图清晰地表现数据中的关系，可以更完整、更方便地分析数据。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。

5. Complex Samples 模块

如果使用了特别复杂的抽样方案，该模块可以计算复杂样本的统计数据，得到更精确的结果。它拥有专门的规划工具和统计方法，提供各种向导来制定取样方案或详细定义样本，并提供专门的技术来解决样本设计及相伴标准误差，能够减少得出错误或误导性推论的风险。Complex Samples 模块将抽样设计融入调查分析之中，对复杂抽样数据的总体得到更加有效的统计推论，对于调查、市场、民意研究人员或者社会科学家来说是必不可少的统计工具。

6. Conjoint 模块

Conjoint 模块提供一种实际的方式，用以度量单个产品属性如何影响消费者和市民偏好。帮助市场研究人员和新产品开发部门了解在消费者心目中什么产品属性是重要的，了解最偏爱的属性水平是什么，从而进行定价研究，以及品牌价格研究。在产品投入大批量生产之前进行这些研究，以避免可能的失误。

7. Custom Tables 模块

Custom Tables 模块可帮助分析人员在较少的时间里创建各种美观、精确的表格，包括复杂的行列表和多重响应数据的显示。

8. Data Preparation 模块

Data Preparation 模块可以简化数据准备过程，在预处理数据时轻易地识别虚假的和无效的观测、变量、和数据值，确认可疑的或者残缺的案例，查看数据缺失模式，描述变量分布以备分析，更准确地应用针对于分类变量的算法，还可以用为分类变量设计的运算法则来做更多精确的工作。使用 Data Preparation 模块，可以迅速找到多元的极端值，执行数据检验，为建模预处理数据。

9. Decision Trees 模块

Decision Trees 模块基于数据挖掘中发展起来的树结构模型对分类变量或连续变量进行预测，可以方便、快速地对样本进行细分。此过程为探索性和证实性分类分析提供验证工具。

10. Direct Marketing 模块

Direct Marketing 模块主要用来处理市场直销中的一些分析需求。目前提供 RFM 客户评分、客户分群、目标客户轮廓概括、客户响应评分、不同营销行为响应测量等模型，使其营销计划尽可能地发挥效力。

11. Exact Tests 模块

Exact Tests 模块可在小样本或分布非常不均匀的样本可能导致常规检验不准确的情况下计算统计检验的精确 P 值。此选项只在 Windows 操作系统中可用。

12. Forecasting 模块

Forecasting 模块通过使用多种曲线拟合模型、平滑模型和用于估计自回归函数的方法，执行综合的预测和时间序列分析。

13. Missing Values 模块

缺失数据会带来偏差或错误的分析结果，简单代入法或者简单的回归法都不能正确地填补缺失值，Missing Values 模块描述了缺失数据的模式、估计均值和其他统计量，并利用统计方法填充缺失值。

14. Neural Networks 模块

Neural Networks 模块可以通过将产品需求预测为价格函数以及其他变量的函数，或根据购买习惯和人口统计特征对客户进行分类来制定经营决策。是非线性数据建模工具。它们可以用来建立输入与输出之间的复杂关系模型，也可用来查找数据中的模式。

15. Regression 模块

Regression 模块提供了用于分析不能拟合传统线性统计模型的数据的方法。它包括一些用于 probit 分析、logistic 回归、权重估计、两阶段最小平方回归和常规非线性回归的过程。

16. Programmability Extension 模块

Programmability Extension 模块可以使用外部语言来执行 SPSS 一连串的分析动作，以达到自动化的目的。通过撰写内嵌在 SPSS 18 里的 Python 程序来控制 SPSS 的各式语法工作执行，如执行设定变量属性、观察程序输出、错误码或条件状态等。运用外部程序结合新的后端 API 处理，可扩大语法执行工作的弹性。

1.2 SPSS 使用基础

1.2.1 SPSS 的安装

IBM SPSS Statistics 19 for Windows 的安装同其他 Windows 应用软件一样,非常简单,下面简要介绍安装 SPSS 的步骤。

1. 启动安装

将 SPSS 软件安装盘放入光驱,如果系统设置为自动运行光盘状态,则光盘自动执行 setup 应用程序,出现安装界面,如图 1-1 所示。若不能自动运行,则运行资源管理器,打开光盘中 Windows\setup.exe 文件,出现安装界面(注:如果是 SPSS 官方网站下载的试用版,则直接运行安装文件,进入安装过程)。

2. 设置安装选项

单击 Install IBM SPSS Statistics 19 选项,进入 SPSS 的安装,按照安装向导,根据提示设置安装信息:

- 正版 SPSS 需输入 SN 序列号,试用版不需要。
- 安装过程中会询问许可证的不同类型,即用户的不同类型,如图 1-2 所示。根据实际购买情况,如果是单机用户,选择选项“单个用户许可证”;如果是企业购买了这款软件,同时给个人一个使用号码,选择选项“站点许可证”;如果企业购买的软件是网络版,则选择选项“网络许可证”。
- 需要接受软件使用协议。
- 用户需要填写用户名、单位名称,如果是网络版,需设置许可证服务器名称或地址。
- 选择帮助语言,默认是“英文”+“中文”,试用版需单独下载帮助语言包。
- 系统默认的安装路径是 C:\Program Files\IBM\SPSS\Statistics\19\,如用户需改变安装路径,可以单击“更改”按钮来自定义安装位置。
- 单击“安装”按钮,开始安装软件。

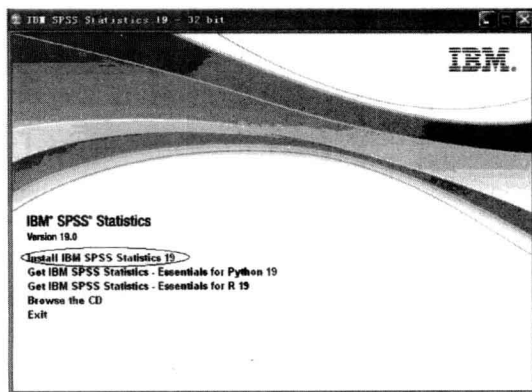


图 1-1 安装前的选择画面

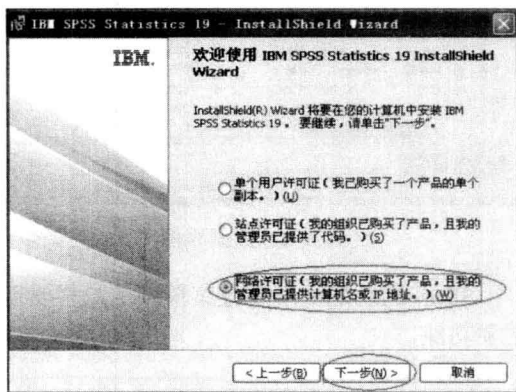


图 1-2 许可证类型选择界面

3. 软件授权

安装完毕,启动 SPSS 授权过程,根据软件的授权码,连接 SPSS 公司的许可证管理服务器获取许可证。成功授权之后,软件方可正常使用。试用版有临时许可证供短期试用。

1.2.2 SPSS 的界面

SPSS 主要有 5 类窗口,分别为数据编辑窗口、结果输出窗口、结果编辑窗口、语法编辑窗口和脚本编辑窗口。

1. 数据编辑窗口

数据编辑窗口是 SPSS 软件中最常用的窗口,这个窗口主要用来处理数据和定义数据字典,它分为两个视图:一个是用于显示和处理数据的数据视图(Data View),另一个是用于变量定义和查看的变量视图(Variable View)。

数据视图如图 1-3 所示,提供类似 Excel 电子表格的编辑窗口,在该窗口中可以创建、编辑、浏览数据文件。其操作和 Excel 非常相似。在 SPSS 中允许打开多个数据文件名进行编辑、浏览,正在编辑的数据文件称为活动数据文件,只有活动数据文件的数据才能被分析处理。SPSS 的数据以表格的形式呈现,表的每一行表示一个观察个案(Case),每一列表示一个变量(Variable),表的大小由变量数和观察个案数确定。一般情况下,分析的数据应以 SPSS 数据文件的形式保存,最常用的 SPSS 数据文件扩展名为“*.sav”,保存数据文件的同时也保存了变量属性和变量值。

变量视图的功能是定义数据集的数据字典,它用来定义、显示和修改数据集中的变量信息,变量视图如图 1-4 所示。

	姓名	日期	产品	数量	单价	金额	变量
1	李汉青	01-Jan-2003	彩电	42	3200	134400	
2	张三中	02-Jan-2003	彩电	40	3200	128000	
3	李开	03-Jan-2003	空调	3	3200	9600	
4	张国华	04-Jan-2003	微波炉	24	2100	50400	
5	王三	05-Jan-2003	热水器	24	2300	55200	
6	刘利国	06-Jan-2003	彩电	12	3200	38400	
7	杜为	07-Jan-2003	洗衣机	5	2200	11000	
8	吴兵	08-Jan-2003	洗衣机	48	2200	105600	

图 1-3 数据视图

	名称	类型	宽度	小数	标签	值
1	姓名	字符串	6	0		无
2	日期	日期	11	0		无
3	产品	字符串	6	0		无
4	数量	数值(N)	11	0		无
5	单价	数值(N)	11	0		无
6	金额	数值(N)	11	0		无

图 1-4 变量视图

SPSS 的功能主要通过菜单和工具栏实现,工具栏是常用菜单项的快捷方式,下面介绍菜单的主要功能。

(1) 文件

“文件”菜单负责新建各种类型的文件、读入文件或数据库的内容、保存文件、将数据输出到