

工程设计与分析系列

SPSS

(第2版)

统计分析与数据挖掘

谢龙汉 尚 涛 蔡明京 编著

视频教学



- ★ SPSS——全球使用最广泛的统计分析软件
- ★ SPSS——金融、科研、数据挖掘领域必备之软件
- ★ 基础知识——实训实例——工程应用
- ★ 实例操作视频教学，轻松学习



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

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

工程设计与分析系列

SPSS 统计分析与数据挖掘 (第2版)

谢龙汉 尚 涛 蔡明京 编著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书基于 SPSS 21.0 编写, 每章均有大量分析案例, 结合案例对 SPSS 各模块的统计分析功能和图形功能进行详细讲解。本书具体内容为 SPSS 简介、SPSS 数据挖掘系统介绍、数据文件管理、数据预处理、基本统计分析、多重反应分析、均值的比较与检验、统计图制作、参数检验、回归分析、方差分析、相关分析、聚类分析、判别分析、因子分析、对应分析、信度分析、生存分析、对数线性模型、时间序列分析、缺失值分析, 以及 SPSS 在财务智能、数据预测、股市分析、社会经济分析、金融数据分析等方面的数据挖掘应用。

本书最大特点是抛弃了其他同类书籍中只介绍理论用法、缺乏案例分析的弊端, 全书给出大量数据挖掘分析案例, 并配有视频讲解, 为读者展示 SPSS 在数据分析、信用风险管理、直销分析、社会经济分析等实际项目中的应用。

本书适合众多领域的数据分析人员, 也可供相关专业本科生、研究生、科技人员和企事业单位工作人员, 以及从事数据挖掘、金融分析、商业咨询、财务分析的人员学习。

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 统计分析与数据挖掘 / 谢龙汉, 尚涛, 蔡明京编著. — 2 版. — 北京: 电子工业出版社, 2014.4

(工程设计与分析系列)

ISBN 978-7-121-22549-9

I. ①S… II. ①谢… ②尚… ③蔡… 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 037990 号

策划编辑: 许存权

责任编辑: 许存权 特约编辑: 马军令

印 刷: 北京市李史山胶印厂

装 订: 北京市李史山胶印厂

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: 787×1 092 1/16 印张: 31.5 字数: 810 千字

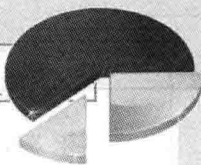
印 次: 2014 年 4 月第 1 次印刷

印 数: 4 000 册 定价: 65.00 元 (含 DVD 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。



再版前言

SPSS (Statistical Product and Service Solutions, 统计产品和服务解决方案) 是当今国际上最流行的统计分析软件之一, 具有界面友好、统计功能强大、前后处理功能完善等优点。SPSS 提供了广泛的数据收集、分类、分析和处理技术, 揭示了数据模式、异常, 以及关键变量和关系, 帮助企业深入洞察企业信息, 做出更好决策。本书从 SPSS 窗口操作出发, 用案例的形式介绍 SPSS 数据分析在各个领域的广泛应用。

本书集作者多年使用 SPSS 的工作经验, 并在改正第 1 版错误的基础上编写, 在编写过程中, 突出了以下特点。

① 直观易懂性。全书以图解实例的形式介绍基础知识和实例操作, 所有的知识模块和案例分析都尽可能详细, 模块操作采取中英文介绍的方式进行, 直观易懂, 使读者能够在最短的时间内获取最多的知识。

② 先进性。以最新的 SPSS 21.0 中文版为蓝本进行讲解, 中英文并用, 广泛吸收国内外优秀教材的成果进行内容编排, 在系统介绍基本理论和基本方法的同时, 注意介绍新的成熟的内容, 以及统计学在实际问题中的应用。

③ 实用性。全书采用了基础知识介绍和实例操作相结合的方法, 互相补充, 书中的实例大多来源于经济生活之中, 使读者在学完本书后能够快速将知识应用于实践。

④ 结构清晰, 讲解详尽。全书采用基础知识—窗口操作—综合实例分析的循序渐进的讲解方法, 一步一步地提高读者的 SPSS 操作知识, 而且每个知识点和实例都尽可能详细地讲解, 使读者学习起来轻松自如。

⑤ 全部的案例数据、程序与多媒体示范相结合。本书的配套光盘  中提供了所有实例的数据、SPSS 窗口操作视频, 读者可以在观看录像中增强对知识点的理解。

本书共 24 章, 依次介绍 SPSS 基本文件管理、基本统计分析、高级统计分析、决策树模型、神经网络模型、信用风险、社会经济评价, 以及各章节中的案例分析等内容。

第 1 章 SPSS 软件概述。包括 SPSS 软件简介、SPSS 操作入门、SPSS 各个模块, 以及 SPSS 帮助系统。

第 2 章 SPSS 数据挖掘系统。包括数据挖掘概述、SPSS 数据挖掘过程的介绍, 以便掌握数据挖掘基本概念、流程等知识。

第 3 章 数据文件、变量与函数。包括 SPSS 的变量类型、SPSS 数据文件的打开和保存, 最后介绍 SPSS 的函数。

第 4 章 数据预处理。包括最基本的数据文件的整理和数据变量的变换和计算。

第 5 章 基本统计分析。包括基本概念、频数过程、描述性统计分析过程、数据探索性分析过程, 以及列联表分析过程。

第 6 章 参数检验。包括参数估计和假设检验概述、均值 (Means) 过程、单样本 t 检验、独立样本 t 检验, 以及配对两样本 t 检验。

第 7 章 基本图形的绘制。包括统计图概述、条形图、线图、面积图、饼图、高低图、质量控制图、箱图、散点图、直方图、P-P 图和 Q-Q 图, 以及时间序列图。

第8章 非参数检验。包括非参数检验概述、 χ^2 检验、二项分布检验、游程检验、K-S检验、两独立样本分布位置检验、多个独立样本分布位置检验、两相关样本分布位置检验、多个独立样本分布位置检验。

第9章 方差分析。包括方差分析的基本原理、单因素方差分析、多因素方差分析、协方差分析。

第10章 回归分析。包括线性回归、非线性回归,以及 Logistic 过程。

第11章 相关分析。包括相关分析概述、Bivariate 过程、Partial 过程,以及 Distances (距离)过程。

第12章 聚类分析。包括聚类分析的原理、快速聚类的 SPSS 过程、系统聚类的 SPSS 过程、两阶段聚类的 SPSS 过程,以及案例分析。

第13章 判别分析。包括判别分析的基本原理、一般判别分析过程、逐步判别分析过程。

第14章 因子分析。包括因子分析,以及 SPSS 中的因子分析操作过程。

第15章 对应分析。包括对应分析的基本原理、对应分析过程、Optimal Scaling 过程。

第16章 可靠性和多维尺度分析。包括可靠性分析、多维尺度分析及案例。

第17章 生存分析。包括生存分析概述、Life Tables 过程、Kaplan-Meier 分析过程、Cox 模型回归分析。

第18章 对数线性模型。包括对数线性模型概述、General 过程、Logit 过程、模型(Model) Selection 过程。

第19章 时间序列分析。包括时间序列概述、时间序列数据的预处理、指数平滑方法、ARIMA 模型、季节分解模型。

第20章 缺失值分析。包括 SPSS 中的缺失值理论概述、SPSS 缺失值分析操作,以及缺失值实例分析。

第21章 决策树模型。包括决策树模型概述、SPSS 中参数设置,以及利用实例分析来介绍决策树模型的应用。

第22章 神经网络。包括神经网络概述、SPSS 神经网络模型的设置及实例分析。

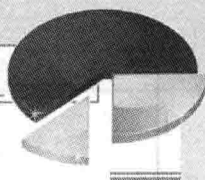
第23章 信用风险分析。包括主要信用风险概述,以及利用 SPSS 解决信用风险的各种实例分析。

第24章 SPSS 在社会经济综合评价中的应用。包括 SPSS 的各种分析案例,包括沿海省市经济综合指标的主成分分析、中国城镇居民消费结构的聚类分析研究,以及我国内地可支配收入和消费性支出之间的回归分析。

本书主要由谢龙汉、尚涛、蔡明京完成,参与编写和光盘开发的还有林伟、魏艳光、林木议、王悦阳、林伟洁、林树财、郑晓、吴苗、卢彩元、苏延全、王文娟、周金华、吕云峰、付应乾、唐长刚、王敏、杨峰、赵新宇、丁圆圆等。由于时间仓促,书中难免有疏漏之处,请读者谅解。读者可通过电子邮件 xielonghan@yahoo.com.cn 与我们交流。

注:本书在介绍软件应用时,命令、选项等包含英文注释,有助于使用英文版软件的读者学习。

编著者



目 录

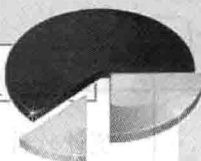
第 1 章 SPSS 软件概述 1	3.3 SPSS 函数 38
1.1 SPSS 简介 1	3.3.1 算术函数 38
1.2 SPSS 操作入门 2	3.3.2 统计函数 38
1.2.1 软件安装、启动及退出 3	3.3.3 逻辑函数 39
1.2.2 操作环境 4	3.3.4 日期和时间函数 39
1.2.3 系统参数的设置 7	3.3.5 随机变量函数 41
1.3 SPSS 的帮助系统 14	3.3.6 反分布函数 43
第 2 章 SPSS 数据挖掘系统 16	3.3.7 累计分布函数 43
2.1 数据挖掘概述 16	3.3.8 缺失值函数 45
2.1.1 数据挖掘的含义 16	3.3.9 字符串函数 46
2.1.2 数据挖掘与 OLAP 17	第 4 章 数据预处理 48
2.1.3 数据挖掘和统计学 17	4.1 数据文件的整理 48
2.1.4 数据挖掘的目的 18	4.1.1 排序个案 (Sort Case) 过程 49
2.1.5 数据挖掘应用 18	4.1.2 转置 (Transpose) 过程 49
2.1.6 数据挖掘流程 18	4.1.3 合并文件 (Merge File) 过程 50
2.2 成功的数据挖掘 19	4.1.4 分类汇总 (Aggregate) 过程 52
2.2.1 CRISP-DM 方法论 20	4.1.5 拆分文件 (Split File) 过程 54
2.2.2 选择数据挖掘工具 24	4.1.6 选择个案 (Select Cases) 过程 54
2.2.3 SPSS 数据挖掘 25	4.1.7 加权个案 (Weight Cases) 过程 55
2.3 SPSS 数据挖掘的过程 28	4.2 数据变量的变换和计算 55
2.3.1 商业理解 28	4.2.1 计算变量 (Compute Variables) 过程 56
2.3.2 数据理解 28	4.2.2 计数 (Count) 过程 58
2.3.3 数据准备 28	4.2.3 重新编码 (Recode) 过程 59
2.3.4 数据模型 29	4.2.4 个案排序 (Rank Cases) 过程 60
2.3.5 评估 29	4.2.5 自动重新编码 (Automatic Recode) 过程 62
2.3.6 部署 30	第 5 章 基本统计分析 64
第 3 章 数据文件、变量与函数 32	5.1 基本概念 64
3.1 SPSS 的变量类型 32	5.1.1 基本的统计概念 64
3.1.1 数据的输入 33	5.1.2 描述性统计分析 66
3.1.2 变量的编辑 34	5.2 频数分析 67
3.2 数据文件的打开和保存 36	5.2.1 Frequencies 过程的操作界面 67
3.2.1 打开 SPSS 数据文件 36	5.2.2 实例分析 69
3.2.2 打开其他格式的数据文件 36	
3.2.3 数据文件保存 37	

5.3	描述性统计分析过程	71	7.9	散点图	120
5.3.1	Descriptive 过程的参数设置	71	7.9.1	SPSS 参数设置	121
5.3.2	实例分析	71	7.9.2	实例分析	121
5.4	数据探索性分析过程	73	7.10	直方图	123
5.4.1	Explore 过程的参数设置	73	7.11	时间序列图	125
5.4.2	实例分析	74	7.11.1	SPSS 参数设置	125
5.5	列联表分析过程	77	7.11.2	实例分析	129
5.5.1	Crosstabs 过程的参数设置	77	第 8 章	非参数检验	132
5.5.2	实例分析	80	8.1	非参数检验概述	132
第 6 章	参数检验	83	8.2	χ^2 检验	133
6.1	参数估计和假设检验概述	83	8.2.1	χ^2 检验的参数设置	134
6.1.1	参数估计	83	8.2.2	χ^2 检验实例分析	136
6.1.2	假设检验	86	8.3	二项分布检验	138
6.2	均值 (Means) 过程	91	8.3.1	二项分布检验的参数设置	138
6.2.1	SPSS 的均值 (Means) 过程参数设置	91	8.3.2	实例分析	138
6.2.2	均值 (Means) 过程实例	92	8.4	游程检验	140
6.3	单样本 t 检验	93	8.4.1	游程检验的参数设置	141
6.3.1	单样本 t 检验过程的参数设置	93	8.4.2	实例分析	141
6.3.2	实例分析	94	8.5	单样本 K-S 检验	143
6.4	独立样本 t 检验	96	8.5.1	单样本 K-S 检验的参数设置	143
6.4.1	独立样本 t 检验过程的参数设置	96	8.5.2	实例分析	144
6.4.2	实例分析	97	8.6	两独立样本分布位置检验	146
6.5	配对两样本 t 检验	99	8.6.1	两独立样本分布位置检验的参数设置	147
6.5.1	配对两样本 t 检验过程的参数设置	99	8.6.2	实例分析	147
6.5.2	实例分析	99	8.7	多个独立样本分布位置检验	149
第 7 章	基本图形的绘制	102	8.7.1	SPSS 参数设置	149
7.1	统计图概述	102	8.7.2	实例分析	150
7.2	条形图	103	8.8	两相关样本分布位置检验	152
7.3	线图	107	8.8.1	SPSS 参数设置	152
7.4	面积图	109	8.8.2	实例分析	153
7.5	饼图	110	8.9	多个配对样本分布位置检验	154
7.5.1	SPSS 设置	110	8.9.1	SPSS 参数设置	155
7.5.2	实例分析	111	8.9.2	实例分析	155
7.6	高低图	112	第 9 章	方差分析	158
7.7	质量控制图	113	9.1	方差分析的基本原理	158
7.8	箱图	118	9.1.1	自由度与平方和分解	159
7.8.1	SPSS 参数设置	118	9.1.2	F 检验	161
7.8.2	实例分析	119	9.1.3	多重比较	162
			9.2	单因素方差分析	163

9.2.1	单因素方差分析步骤	164	11.2.3	案例——汽车销售中各变量之间的相关分析	220
9.2.2	判断与结论	165	11.3	Partial (偏相关) 过程	223
9.2.3	One-Way ANOVA 过程的参数设置	165	11.3.1	Partial 过程的参数设置	223
9.2.4	实例分析	168	11.3.2	案例——医疗门诊病人的流动情况分析	224
9.3	多因素方差分析	169	11.4	Distances (距离) 过程	226
9.3.1	只考虑主效应的多因素方差分析	169	11.4.1	Distances 过程的参数设置	226
9.3.2	存在交互效应的多因素方差分析	171	11.4.2	案例——全国各个省市自治区直辖市的农民家庭收支的分布研究	229
9.3.3	Univariate 过程参数设置	173	第 12 章	聚类分析	231
9.3.4	实例分析	178	12.1	聚类分析的原理	231
9.4	协方差分析	182	12.1.1	一般原理	232
9.4.1	协方差分析概述	182	12.1.2	聚类分析步骤	235
9.4.2	实例分析	183	12.1.3	系统聚类方法	236
第 10 章	回归分析	186	12.2	快速样本聚类过程	239
10.1	线性回归	186	12.2.1	快速聚类简介	239
10.1.1	线性回归模型	187	12.2.2	SPSS 快速聚类的设置	239
10.1.2	最小二乘估计	187	12.2.3	案例——2006 年中国主要城市空气质量的比较分析	241
10.1.3	回归方程的显著性检验	188	12.3	系统聚类过程	245
10.1.4	预测问题	190	12.3.1	系统聚类简介	245
10.1.5	SPSS 线性回归分析设置	191	12.3.2	SPSS 系统聚类设置	245
10.1.6	回归分析模型的实例分析	195	12.3.3	案例——中国经济地理区域的聚类分析	248
10.2	非线性回归	198	12.4	两阶段聚类分析	251
10.2.1	非线性回归分析的基本原理	199	12.4.1	两阶段聚类简介	251
10.2.2	非线性回归参数设置	199	12.4.2	SPSS 两阶段聚类的设置	252
10.2.3	案例——销售数量和广告投入的非线性回归分析	202	12.4.3	案例——两阶段聚类分析应用于农村居民人均收入与生活消费支出研究	253
10.3	Logistic 回归	205	第 13 章	判别分析	256
10.3.1	Logistic 回归模型概述	205	13.1	判别分析的基本原理	256
10.3.2	Binary Logistic 回归模型参数设置	206	13.1.1	判别分析简介	256
10.3.3	案例——银行贷款的信用风险分析	209	13.1.2	判别分析的数学模型与判别方法	257
第 11 章	相关分析	214	13.2	一般判别分析	263
11.1	相关分析概述	214	13.2.1	一般判别分析的参数设置	263
11.1.1	相关关系	214	13.2.2	案例——上市公司类型的比较分析	266
11.1.2	相关图形和相关系数	215	13.3	逐步判别分析	271
11.1.3	SPSS 的相关分析功能简介	217			
11.2	Bivariate (双变量) 过程	217			
11.2.1	双变量相关分析简介	217			
11.2.2	Bivariate 过程的参数设置	219			

视频教学

19.5.1	季节分解模型的参数设置	393	21.2.17	模型建立	429
19.5.2	案例——具有季节因素的服装 销售数据的预测分析	394	21.2.18	模型评估	431
第 20 章	缺失值分析	398	第 22 章	神经网络	438
20.1	缺失值理论概述	398	22.1	神经网络概述	438
20.1.1	数据缺失方式	399	22.1.1	历史及现状	439
20.1.2	缺失值处理方法	399	22.1.2	神经网络特点	440
20.2	SPSS 缺失值分析操作	403	22.1.3	神经元模型	441
20.2.1	缺失值分析的参数设置	403	22.1.4	神经网络模型	442
20.2.2	案例——数据集中缺失值的 实例分析	407	22.1.5	神经网络的学习规则	442
第 21 章	决策树模型	413	22.1.6	SPSS 神经网络模型	443
21.1	决策树模型概述	413	22.2	SPSS 神经网络模型的设置	445
21.1.1	CHAID 算法	415	22.2.1	多层感知器 (MLP) 的设置	445
21.1.2	Exhaustive CHAID 算法	416	22.2.2	径向基函数 (RBF) 的设置	453
21.1.3	CART 算法	416	22.3	实例分析	455
21.1.4	QUEST 算法	417	22.3.1	参数设置	456
21.2	决策树的参数设置	417	22.3.2	结果分析	458
21.2.1	变量设置	417	第 23 章	信用风险分析	463
21.2.2	类别 (Categories) 设置	418	23.1	信用风险概述	463
21.2.3	输出 (Output) 设置	419	23.1.1	信用风险基本概念	463
21.2.4	验证 (Validation) 设置	421	23.1.2	信用风险度量方法	464
21.2.5	保存 (Save) 设置	422	23.1.3	SPSS 中信用风险分析模块	467
21.2.6	条件 (Criteria) 设置	423	23.2	案例分析	467
21.2.7	CHAID 算法设置	424	23.2.1	Binary Logistic (二元 Logistic) 过程	467
21.2.8	CART 算法设置	424	23.2.2	Tree (决策树) 过程	472
21.2.9	QUEST 算法设置	425	23.2.3	Discriminant (判别) 过程	478
21.2.10	修剪 (Pruning) 设置	425	第 24 章	SPSS 在社会经济综合评价中的 应用	483
21.2.11	替代变量 (Surrogates) 设置	426	24.1	沿海省市经济综合指标的主成分分析	483
21.2.12	选项 (Options) 设置	426	24.2	中国内地城镇居民消费结构的聚类分析	487
21.2.13	误分类成本设置	427	24.3	我国内地可支配收入和消费性 支出之间的回归分析	491
21.2.14	利润 (Profits) 设置	427			
21.2.15	先验概率 (Prior Probabilities) 设置	428			
21.2.16	实例分析	429			



第 1 章 SPSS 软件概述

回顾 SPSS 软件的发展历程,从最初的“社会科学统计软件包”(Solutions Statistical Package for the Social Sciences)到 2000 年的 SPSS (Statistical Product and Service Solutions, 统计产品与服务解决方案)软件,SPSS 软件都发生着巨大的变化。IBM 公司于 2009 年 7 月 28 日宣布将用 12 亿美元收购分析软件提供商 SPSS 公司,如今 SPSS 已发布 SPSS 21.0 版本,也标志着 SPSS 的战略方向正在做出重大调整。本章将讲述 SPSS 的发展历程,并介绍 SPSS 基本的使用方法。



本讲内容

- SPSS21.0 简介
- SPSS21.0 软件安装、启动及退出
- SPSS21.0 软件基本操作环境
- SPSS21.0 帮助系统

1.1 SPSS 简介

SPSS 是英文名称的首字母缩写,英文全称为 Statistical Package for the Social Sciences,即“社会科学统计软件包”。但是随着 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 Product and Service Solutions,即“统计产品和服务解决方案”,标志着 SPSS 的战略方向正在做出重大调整。

SPSS 是世界上最早的统计分析软件,由美国斯坦福大学三位研究生于 20 世纪 60 年代末研制,同时成立了 SPSS 公司,并于 1975 年在芝加哥组建了 SPSS 总部。1984 年,SPSS 总部首先推出了世界上第一个统计分析软件微机版本 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、技术科学、社会科学的各个领域,世界上许多有影响的报刊杂志纷纷就 SPSS 的自动统计绘图、数据的深入分析、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 30 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,是世界上应用最广泛的专业

统计软件。在国际学术界有条不成文的规定, 即在国际学术交流中, 凡是用 SPSS 软件完成的计算和统计分析, 可以不必说明算法, 由此可见其影响之大和信誉之高。

1994 年—1998 年, SPSS 公司陆续购并了 SYSTAT 公司、BMDP 软件公司、Quantime 公司、ISL 公司等, 并将各公司的主打产品收纳 SPSS 旗下, 从而使 SPSS 公司由原来的单一统计产品开发与销售转向企业、教育科研及政府机构提供全面信息统计决策支持服务, 成为了走在最新流行的“数据仓库”和“数据挖掘”领域前沿的一家综合统计软件公司。

SPSS 是世界上最早采用图形菜单驱动界面的统计软件, 它最突出的特点就是操作界面极为友好, 输出结果美观漂亮。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 的窗口方式展示各种管理和分析数据方法的功能, 对话框展示出各种功能选择项。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。是非专业统计人员的首选统计软件。在众多用户对国际常用统计软件 SAS、BMDP、GLIM、GENSTAT、EPILOG、MINITAB 的总体印象分的统计中, 其诸项功能均获得最高分。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, 能方便的从其他数据库中读入数据。其统计过程包括常用的、较为成熟的统计过程, 完全可以满足非统计专业人士的工作需要。输出结果十分美观, 存储时则是专用的 SPO 格式, 可以转存为 HTML 格式和文本格式。对于熟悉老版本编程运行方式的用户, 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。极大的方便了中、高级用户。

由上面的叙述可知, SPSS 具有以下特点。

- 操作简便: 以对话框方式操作, 绝大多数操作过程可通过单击鼠标完成。
- 在线帮助方便: 用户可在 SPSS 的任一过程中获得帮助, 查询主题和索引, 根据帮助框中的指导进行操作。
- 数据转换功能较强: 可存取和转换多种数据类型, 如 dBase, Lotus, Excel, ASCII 文件等。
- 数据管理功能强大: 集数据录入、转换、检索、管理、统计分析、作图、制表及编辑功能于一身。
- 程序生成简化: 系统能将对话框指定的命令、子命令和选择项等内容自动编写成 SPSS 命令语句, 并可以编辑, 继而形成 SPSS 环境下的可执行程序文件。
- 统计分析方法全面丰富: 含有最新的统计方法, 如对应分析 (Correspondence Analysis), 联合分析 (Conjoint Analysis), 多分类变量的 Logistic 回归分析等, 且所用方法具有权威性。
- 结果输出规范: 输出结果主要为图形方式, 规范而简洁, 还可根据个人要求编辑输出方式。

1.2 SPSS 操作入门

本节将详细讲述 SPSS21.0 软件的安装、系统环境和基本操作问题, 初步展示 SPSS21.0 系统的各种特点和用法。

1.2.1 软件安装、启动及退出

1. SPSS21.0 的安装

首先,从 SPSS 的官方网站上下载 SPSS21.0 软件的安装程序,然后解压到 D 盘中,双击“setup.exe”安装文件;或者把装有 SPSS21.0 软件的安装盘,放入计算机的光驱中,系统会自动的弹出 SPSS21.0 的安装对话框,然后根据对话框的提示即可完成 SPSS21.0 软件的安装。

2. SPSS21.0 的启动

SPSS21.0 软件的启动有两种方式,第一种是双击桌面的快捷方式;第二种是在开始菜单中选择“程序→IBM SPSS Statistics→SPSS21.0→IBM SPSS Statistics 21”,单击即可,如图 1-1 所示。

双击桌面 SPSS21.0 的快捷方式,系统进入如图 1-2 所示的界面,则 SPSS 正常启动。进入 SPSS21.0 界面后系统弹出如图 1-3 所示的 SPSS21.0 的启动选项,图 1-3 中各个选项的含义如下。

- 运行图形化教程 (Run the Tutorial)。
- 输入数据 (Type in Data)。
- 运行存在的查询文件 (Run an Existing Query)。
- 新建数据库查询 (Create New Query using Database Wizard)。

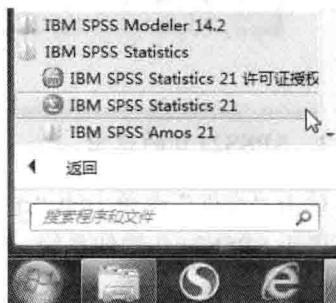


图 1-1 开始菜单中 SPSS 的选项

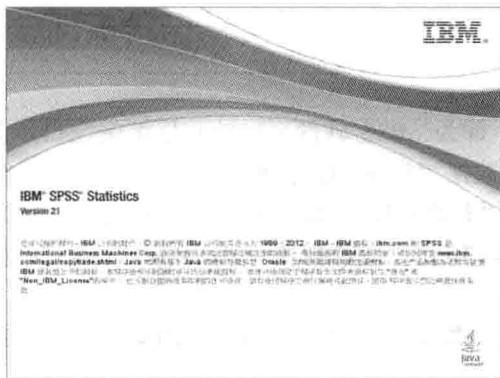


图 1-2 SPSS21.0 启动界面

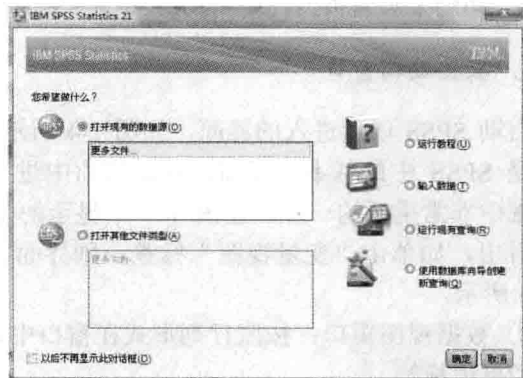


图 1-3 SPSS21.0 的启动选项

- 打开数据文件 (Open an Existing Data Source)。
- 打开类型的数据文件 (Open another Type of File)。
- 不要显示此对话框 (Don't Show this Dialog in the Future)。

单击图 1-3 中的“确定 (OK)”按钮,则系统进入 SPSS21.0 的主界面,如图 1-4 所示。

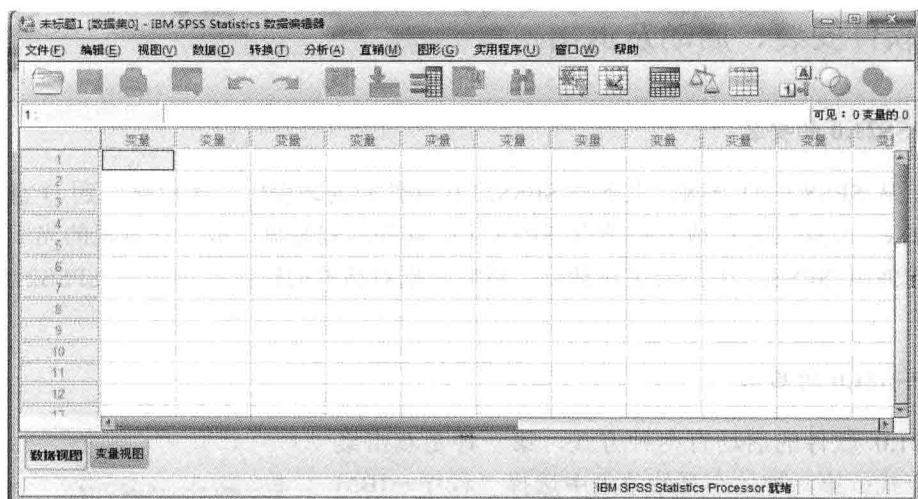


图 1-4 SPSS21.0 的主界面

3. SPSS21.0 的退出

单击“文件”菜单中的“退出”选项，或者直接单击图 1-4 窗口右上角的“关闭”按钮，即可退出 SPSS21.0 软件系统。

1.2.2 操作环境

熟悉 SPSS 软件的操作环境是进行数据分析挖掘的基础，SPSS 的操作环境主要有数据编辑窗口（SPSS Data Editor）、结果浏览窗口（SPSS Viewer）、程序编辑窗口（SPSS Syntax Editor）和脚本编辑窗口（Script）。

1. 数据编辑窗口

启动 SPSS 以后进入的界面，如图 1-4 所示，即为 SPSS 软件系统的数据编辑窗口，此窗口是 SPSS 中最基本的界面，在此窗口中进行数据挖掘前的数据整理编辑工作，也是数据挖掘中非常重要的一步，在图 1-4 中显示的是“数据视图（Data View）”，由此图的底部可以看出，如单击“变量视图”标签，则界面会转换到“变量视图（Variable View）”，如图 1-5 所示。

① 数据视图窗口：按照行列形式在窗口中显示数据，可以在此窗口中浏览、修改数据值和数据值标签。

- 行：表示观察个体，由观察对象的所有属性组成。
- 列：表示变量，一个变量是所有观察对象的某个属性的集合。
- 数据格：表示对应观察对象的某个属性的观察值或者标签。

② 变量视图窗口：创建、显示和修改数据视图窗口中变量属性的窗口。

- 行：表示变量。
- 列：表示变量的属性。



图 1-5 数据变量视图

2. 结果浏览窗口

输出窗口的显示是在选择了一定的变量和选择统计方法或输入了 SPSS 程序命令后才自动生成的。如果运行程序发生错误，则系统给出出错的信息并停止运行。输出窗口主要显示统计结果，包括各种图表等，如图 1-6 所示，图中的具体内容如下。

- 输出窗口包括两部分：左边为大纲视图，右边为显示统计结果。
- 此结果可以作为输出文件进行保存。
- 输出窗口有自己的菜单栏，其大部分菜单与主菜单相同，输出窗口的菜单也可以执行所有的统计分析功能，对数据文件进行分析，分析结果直接显示在输出窗口。
- 程序中打开多个输出窗口，新开的输出窗口按先后顺序分别标记为输出 1 (output1)，输出 2 (output2) 等。
- 双击输出窗口的生成图形可以进一步对其进行编辑或修改。



图 1-6 SPSS 的输出窗口

3. 程序编辑窗口

单击图 1-4 中的菜单“文件 (File) → 新建 (New) → 语法 (Syntax)”，即可打开“语法编辑 (Syntax)”窗口，如图 1-7 所示。语法编辑窗口就是编写、调试和运行 SPSS 程序的出口，大部分的 SPSS 功能可以利用窗口操作来完成，通过 SPSS 程序，用户可以获得想要的 数据分析过程。

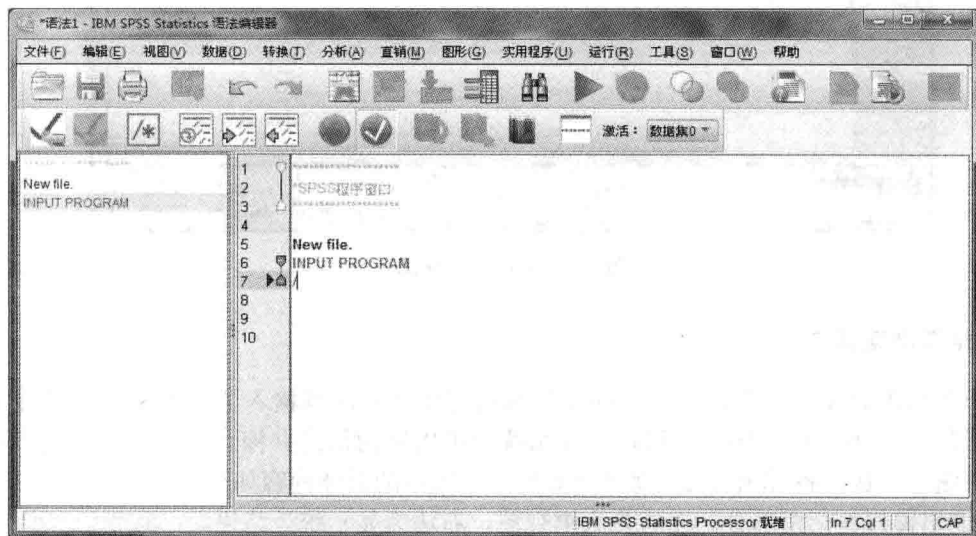


图 1-7 “语法编辑”窗口

语法编辑窗口详细介绍如下。

- 语法编辑窗口按照 SPSS 规则编写 SPSS 程序语句，是一个非激活窗口。只有调动了 一个具体的统计分析程序，并通过单击“粘贴 (Paste)”按钮后，此窗口才会 打开。
- 在窗口中可以对其内容进行修改、保存，从主菜单中单击“运行 (Run)”按钮可以 提交系统运行。
- 其中大部分菜单与主菜单相同，且窗口的菜单也可以执行所有的统计分析功能，对 数据文件进行分析，分析结果直接显示在输出窗口。
- 程序中 可以打开多个语句窗口，新开的语句窗口按先后顺序分别标记为语法 1 (Syntax1)，语法 2 (Syntax2) 等。

4. 脚本编辑窗口 (Script)

单击图 1-7 中的菜单“文件 (File) → 新建 (New) → 脚本 (Script)”，即可打开“脚本 编辑窗口 (Script)”，如图 1-8 所示。脚本编辑窗口是一个非常 有特色的窗口，其使用 Sax BASIC 语言的编程环境。脚本编辑窗口功能如下。

- 定制输出特征：显示、操作对话框；使用命令语句执行数据转换和统计分析；将图 表输出为多种图表格式文件等。

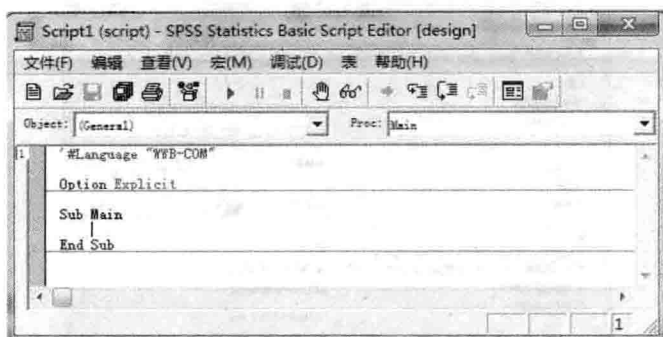


图 1-8 “脚本编辑”窗口

- 通过 Sax BASIC 语言，用户编写自己所需要的程序。
- 在脚本文件夹中安装有较多示范性脚本文件，可以直接调用这些文件来实现某些功能，也可以用这些现存的脚本文件为基础，通过编辑，以实现某些其他功能。
- 程序中打开多个窗口，新开的脚本编辑窗口按先后顺序分别标记为脚本 1 (Script1)，脚本 2 (Script2) 等。

1.2.3 系统参数的设置

系统参数设置通过选择主菜单的“编辑”→“选项”，然后打开对话框完成，如图 1-9 所示为“系统参数设置”对话框，有些参数设置在设置完成后立即生效，有些要在 SPSS 重启后生效。系统参数设置包括常规 (General)、查看器 (Viewer)、数据 (Data)、货币 (Currency)、输出标签 (Output Labels)、图表 (Charts)、枢轴表 (Pivot Tables)、文件位置 (File Locations)、脚本 (Scripts)、多重归因、语法编辑器的参数设置。

1. 常规

常规面板可以设置系统中各种通用参数，所有设置的参数可以自动保存，如图 1-9 所示，各项参数的具体含义如下。

- 变量列表 (Variable List): 设置显示变量顺序的方式。下面的单项选择可以设定变量在变量表中的显示方式和显示顺序。显示方式可选变量标签或变量名。显示顺序可选按变量的字母顺序排列或按在文件中出现的先后顺序排序。
- Windows: 启动 SPSS 时语句窗口状态。
- 数据和语法的特征编码 (Character Encoding for Data and Syntax)。
- 输出选项 (Output): 包括系统度量 (Measurement System)、显示语言 (Language)、结果通知方式 (Notification) 等。

2. 查看器

单击图 1-9 中上方的“查看器”选项，则打开如图 1-10 所示的“观察窗口设置”对话框。在改变了参数以后，再次运行 SPSS 后才能生效。如图 1-10 所示，各项参数的具体含义如下。