

高等学校信息管理与信息系统专业系列教材

# 数据统计与分析技术

## ——SPSS软件实用教程

黄润龙 主编



高等教育出版社  
HIGHER EDUCATION PRESS

C8119  
20071

高等学校信息管理与信息系统专业系列教材

# 数据统计与分析技术 ——SPSS 软件实用教程

黄润龙 主 编  
管于华 副主编  
何志芳 主 审



高等教育出版社

## 内容提要

SPSS 是国内外应用得十分广泛的统计软件,它可用于自然科学和社会科学领域的基本数据处理与分析,具体适用于市场营销、销售分析、市场调查、统计报告、质量控制、科学研究、社会调查、企业管理、教学及行政管理等领域。本教材以最新的 SPSS12.0 版本为范本,介绍数据预处理、统计建模、数据统计分析的基本原则、原理、技巧和操作技术。本教材注重实践,着重培养学生的实际动手能力,对大型社会调查的数据汇总、分组、整理和分析能力,对基础资料综合定量分析、研究能力等,同时关注学生对于数学模型、数据挖掘技术的深入认识和理解。

本教材可作为高等院校信息管理与信息系统、社会学、广告学、人口学、统计学、教育学、经济学等专业教材,也可供从事统计分析的人员参考。

## 图书在版编目(CIP)数据

数据统计与分析技术:SPSS 软件实用教程 / 黄润龙  
主编. —北京:高等教育出版社,2004.7 (2006重印)

ISBN 7-04-014627-4

I. 数... II. 黄... III. 统计分析—软件包, SPSS  
—高等学校—教材 IV. C819

中国版本图书馆CIP数据核字(2004)第053387号

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 100011  
总 机 010-58581000

经 销 蓝色畅想图书发行有限公司  
印 刷 北京市鑫霸印务有限公司

开 本 787×960 1/16  
印 张 23.25  
字 数 420 000

购书热线 010-58581118  
免费咨询 800-810-0598  
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>  
畅想教育 <http://www.widedu.com>

版 次 2004年7月第1版  
印 次 2006年5月第2次印刷  
定 价 29.00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 14627-00

## 前 言

信息时代,人类面临的数据和信息越来越多,面临的选择也越来越多。要从众多的信息和数据中提取有用的信息,以做出正确的决策,就必须对大量数据进行综合统计和分析。

SPSS(Statistical Product & Service Solutions,统计产品与服务解决方案)软件的实际应用范围为自然科学、社会科学、生物医学等领域。SPSS适用于市场营销、销售分析、市场调查、统计报告、质量控制、科学研究、社会调查、企业管理、教学及行政管理等领域,具有数据管理、统计分析、统计报表和图形显示等基本功能。本教材以SPSS 12.0版本为范本,介绍了大量数据预处理、数据统计分析、统计建模的基本原则、原理、技巧和计算机操作技术。

本教材首先介绍SPSS特点及主要功能、数据文件的建立与编辑、数据预处理,包括数据排序、转置、合并、分组、选择、加权和变量的组合;接着介绍描述性统计过程,包括频数分析、多维频数分析、各种描述性统计量和数据标准化处理过程;然后介绍相关分析和回归分析,包括线性回归、多元回归、逐步回归、非线性回归和曲线估计;最后介绍参数检验与非参数检验、方差分析和聚类分析等。

本教材注重实践,着重培养学生的实际动手能力,主要指对大型社会调查的数据汇总、分组、整理能力,对基础资料综合定量分析、研究能力;同时关注学生对数学模型、数据挖掘技术的深入认识和理解。通过本课程的学习,要求基本掌握数据输入、列表分析、存储技术以及数据编码、贴标签、数据转换、选择、加权、文件合并等技巧。要求学生掌握各种基本统计分析的计算方法,能对输出结果进行逻辑关系检查和输出数据准确性的检验;能根据研究目的、数据来源和数据类型选择适当的统计模型进行分析;能对输出结果运用统计学和社会学的知识进行合理解释。比如,能对各类社会调查问卷运用定量和定性分析,形成调查报告,发现被调查对象工作中存在的问题等。本教材可作为高等院校的信息管理与信息系统、社会学、广告学、人口学、统计学、经济学等专业的必修课教材,也可作为管理学、教育学、计算机、市场营销等专业的选修课教材。

目前社会上介绍数据处理的书籍不少,作为高校教材,本教材有如下特点。

(1) 本教材充分融合统计理论和统计分析实践。一方面,目前市场统计处理软件很多,统计计算非常方便,有许多现成的统计软件包,如SAS、STATA、SYSTAT等,但许多统计计算输出结果难以解释。另一方面,市场上统计学教科书很多,但讲传统理论的较多,国外统计学最新成果介绍不够,以至于用户及读者对很多统计分析过程和输出结果不熟悉。为兼顾这两方面,往往要使用两本及多本教科书。本教材充分融合统计理论和统计分析实践,从案例入手,与数理

统计理论有机结合,注重对输出结果的合理解释,并非简单的计算机操作指南。

(2) 本教材是在实践的基础上形成并发展起来的。1992年作者率先将英国伦敦大学原版 SPSS 讲义翻译成中文,1991年—1995年先后在南京人口管理学院统计学专业本科生中广泛使用。由于 SPSS 软件的不断更新(从 SPSS 2.0、SPSS 3.0 到目前的 SPSS 12.0 版本),讲义内容也随着不断更新、调整、修改、完善。1996年后在南京人口管理学院人口经济专业和计算机专业、南京财经大学计算机专业、南京师范大学新闻传播学院广告专业、金陵女子学院社会保障专业、中北学院信息管理专业等本科生中广泛使用。

(3) 该教材支持了多项科研工作,科研又为该教材的编写提供了大量的素材和范例。SPSS 可一次处理 800 万个以上数据,各样本变量可达 300 个,数据吞吐量之大,是其他统计软件所无法比拟的。在该统计软件的帮助下,作者顺利完成了国家社科基金项目、国家统计局、国家计生委等大量课题。

(4) 本教材注重实用,通过 54~72 学时的学习,学生应能够熟练运用计算机进行基本统计运算和操作,能够进行大型调查问卷的汇总、分析和统计处理,也可进行大量数据的统计、分析和建模。SPSS 自带十多个模拟数据库,可操作性强,有助于培养学生自己动手“分析问题和解决问题”的能力。本教材还配有大量的案例和习题。

(5) 2002 年 1 月和 2003 年 5 月数据统计与分析课程和教材先后被评为南京师范大学“十五”重点课程建设项目和“十五”重点教材建设项目,学校投入了大量的人力和物力。

(6) 多所高校参与了本教材的编写。南京财经大学管于华撰写了第 9 章和第 11 章,南京人口管理干部学院宗占红撰写了第 10 章的初稿,南京师范大学黄润龙撰写了本书的其余章节,并对全书进行了统稿,南京师范大学数学与计算机科学学院何志芳对本教材进行了审阅,并提供了修改意见。此外,南京大学社会学系陈友华对本教材提供了大量建设性建议,北京大学顾大男提供了早期 DOS 版本的后 3 章初稿。本教材撰写时间跨度大、准备时间长,并参考了大量国内外的同类资料、教材和教科书,在此深表感谢!

为满足不同专业、不同层次学生学习需要,书中对习题做了必要的提示。各章既有联系,又可独立成章,学习者可有针对选择阅读。对没有统计学、社会学背景的学生,建议先阅读第 15 章。

由于水平有限,书中疏漏难免,恳望诸位学者斧正!

作者

于南京仙林大学城

2004 年 3 月

# 目 录

<b>第 1 章 概述</b> .....	(1)	1.5.3 电子教程的使用方法 .....	(24)
1.1 特点 .....	(2)	1.5.4 个案分析 .....	(25)
1.2 SPSS 的运行和运行环境 .....	(3)	1.5.5 统计教练 .....	(25)
1.2.1 对运行环境的要求 .....	(3)	1.5.6 命令语法指南 .....	(26)
1.2.2 系统的运行及主屏幕 介绍 .....	(3)	1.5.7 寻找主页 .....	(27)
1.2.3 与其他数据处理软件的 联系与区别 .....	(5)	本章小结 .....	(27)
1.3 SPSS 的主要功能 .....	(6)	习题一 .....	(28)
1.3.1 数据编辑功能 .....	(7)	<b>第 2 章 数据统计处理实例</b> .....	(29)
1.3.2 表格的生成和编辑 .....	(7)	2.1 自建数据文件的处理—— 实例 1 .....	(29)
1.3.3 图形的生成和编辑 .....	(7)	2.1.1 数据处理的构思 .....	(29)
1.3.4 与其他软件的连接 .....	(7)	2.1.2 数据录入和定义 .....	(31)
1.3.5 统计功能 .....	(7)	2.1.3 数据资料预处理 .....	(33)
1.4 系统参数设置 .....	(10)	2.1.4 数据统计分析 .....	(34)
1.4.1 总体设置 .....	(11)	2.1.5 结果输出选择 .....	(34)
1.4.2 显示输出的设置 .....	(12)	2.1.6 图形绘制 .....	(35)
1.4.3 草稿显示设置 .....	(14)	2.1.7 文件保存 .....	(35)
1.4.4 标签输出设置 .....	(14)	2.2 使用已输入的数据文件—— 实例 2 .....	(36)
1.4.5 图形设置 .....	(15)	2.2.1 变量名、变量标签的 汉化 .....	(36)
1.4.6 交互图形窗口参数设置 .....	(16)	2.2.2 变量的转换 .....	(37)
1.4.7 要点表格设置 .....	(17)	2.2.3 统计分析 .....	(38)
1.4.8 数值型变量自定义格式 设置 .....	(19)	2.2.4 图表的制作 .....	(39)
1.4.9 货币格式的设置 .....	(19)	2.3 数据文件的转换——实例 3 .....	(40)
1.4.10 脚本窗口的设置 .....	(20)	2.3.1 自由格式 ASCII 码文件的 转换 .....	(40)
1.5 帮助系统 .....	(21)	2.3.2 固定格式 ASCII 码文件的 转换 .....	(44)
1.5.1 如何获得即时提示和 帮助 .....	(21)	本章小结 .....	(44)
1.5.2 按目录、内容或字母次序 检索使用方法 .....	(22)	习题二 .....	(45)

第3章 系统运行方式 .....	(47)	4.5.3 在给定范围内选取数据 ...	(92)
3.1 系统运行管理方式 .....	(47)	4.6 数据的加权 .....	(93)
3.1.1 窗口运行管理方式 .....	(47)	4.6.1 概述 .....	(93)
3.1.2 程序运行管理方式 .....	(48)	4.6.2 对个案加权 .....	(95)
3.1.3 混合运行管理方式 .....	(48)	4.7 数据的转换 .....	(96)
3.2 窗口形式及其功能 .....	(49)	4.7.1 用“Count”功能选项转换 数据 .....	(97)
3.2.1 数据编辑窗口 .....	(49)	4.7.2 利用“Recode”功能转换 数据 .....	(99)
3.2.2 结果输出窗口 .....	(51)	4.7.3 利用“Automatic Recode” 功能转换数据 .....	(102)
3.2.3 程序编写窗口 .....	(52)	4.7.4 缺失值的替换 .....	(103)
3.2.4 统计图形编辑窗口 .....	(53)	本章小结 .....	(105)
3.2.5 帮助窗口 .....	(54)	习题四 .....	(106)
3.3 对话框类型及其使用方法 .....	(54)	第5章 样本描述性统计 .....	(107)
3.3.1 常见对话框类型 .....	(54)	5.1 基本数学模型 .....	(107)
3.3.2 按钮与文本选择 .....	(56)	5.1.1 描述集中趋势的 统计量 .....	(107)
3.4 数据编辑窗口 .....	(56)	5.1.2 描述离中趋势的 统计量 .....	(109)
3.4.1 数据编辑窗口 .....	(57)	5.1.3 偏度和峰度 .....	(110)
3.4.2 数据文件 .....	(58)	5.2 频数分析过程 .....	(110)
3.5 常量、变量和表达式 .....	(58)	5.2.1 对话框介绍 .....	(110)
3.5.1 常量 .....	(58)	5.2.2 应用举例 .....	(113)
3.5.2 变量 .....	(60)	5.3 数据描述过程(Descriptives) ...	(115)
3.5.3 表达式 .....	(66)	5.3.1 对话框介绍 .....	(115)
3.6 构建新变量 .....	(68)	5.3.2 应用举例 .....	(116)
3.6.1 新变量构建方法 .....	(69)	5.4 数据探索过程(Explore) .....	(116)
3.6.2 新变量构建主要函数 .....	(72)	5.4.1 对话框说明 .....	(116)
本章小结 .....	(75)	5.4.2 应用实例及其解释 .....	(119)
习题三 .....	(75)	5.5 列联表分析过程 .....	(121)
第4章 分析前数据预处理 .....	(76)	5.5.1 主要功能 .....	(122)
4.1 数据的排序 .....	(76)	5.5.2 列联表分析 .....	(123)
4.2 数据的转置 .....	(79)	5.5.3 例题分析 .....	(129)
4.3 数据的拆分 .....	(80)	5.6 复选题的统计和分析 .....	(130)
4.4 数据文件的合并 .....	(81)	5.6.1 问题的提出 .....	(130)
4.4.1 样本合并 .....	(82)	5.6.2 范例分析 .....	(130)
4.4.2 变量合并 .....	(85)		
4.5 数据的选择 .....	(88)		
4.5.1 根据逻辑关系表达式 选择数据 .....	(89)		
4.5.2 随机选择数据 .....	(92)		

5.7 在线统计分析 .....	(135)	7.2.1 偏相关系数的含义及 计算 .....	(168)
5.7.1 基本分析过程 .....	(135)	7.2.2 偏相关系数的计算分析 方法 .....	(169)
5.7.2 例题分析 .....	(136)	7.2.3 分析实例 .....	(170)
本章小结 .....	(138)	7.3 距离分析 .....	(171)
习题五 .....	(138)	7.3.1 距离分析简介 .....	(171)
第6章 平均数比较与 T 检验 .....	(141)	7.3.2 过程 and 选择 .....	(171)
6.1 分组平均数的比较 .....	(141)	7.3.3 实例分析 .....	(173)
6.1.1 基本概念和数学模型 .....	(141)	本章小结 .....	(174)
6.1.2 Mean 过程介绍 .....	(142)	习题七 .....	(175)
6.1.3 实例分析 .....	(144)	第8章 回归分析 .....	(177)
6.2 单一样本 T 检验 .....	(145)	8.1 线性回归 .....	(177)
6.2.1 假设检验的一般理论 .....	(145)	8.1.1 回归方程的数学模型 .....	(177)
6.2.2 单一样本 T 检验的 数学模型 .....	(147)	8.1.2 回归方程的选择 .....	(180)
6.2.3 单一样本 T 检验过程 .....	(147)	8.1.3 回归方程的建立 .....	(182)
6.2.4 实例分析 .....	(148)	8.1.4 实例分析 .....	(187)
6.3 独立样本——两样本平均数 差异 T 检验 .....	(149)	8.2 曲线回归 .....	(190)
6.3.1 检验原理和一般理论 .....	(149)	8.2.1 曲线回归的计算原理 .....	(190)
6.3.2 统计过程分析 .....	(150)	8.2.2 曲线回归的计算机实现 .....	(195)
6.3.3 例题分析 .....	(150)	8.3 哑变量回归 .....	(197)
6.4 配对样本 T 检验 .....	(151)	8.4 路径分析 .....	(198)
6.4.1 数学模型 .....	(152)	8.4.1 通径图 .....	(199)
6.4.2 统计过程分析 .....	(152)	8.4.2 通径系数计算及 总作用分解 .....	(200)
6.4.3 个案分析 .....	(153)	8.4.3 实例分析 .....	(201)
6.5 多个平均数检验 .....	(155)	8.4.4 注意事项 .....	(202)
6.5.1 方差分析简介 .....	(155)	本章小结 .....	(203)
6.5.2 单因素方差分析 .....	(156)	习题八 .....	(203)
本章小结 .....	(158)	第9章 非参数检验 .....	(207)
习题六 .....	(159)	9.1 非参数检验概述 .....	(207)
第7章 相关分析 .....	(160)	9.1.1 非参数检验的特点 .....	(207)
7.1 线性相关系数 $r$ 及显著性 检验 .....	(160)	9.1.2 非参数检验方法分类 .....	(208)
7.1.1 相关模型的假设和计算 .....	(160)	9.2 单样本检验 .....	(209)
7.1.2 相关过程的实现 .....	(163)	9.2.1 卡方检验 .....	(209)
7.1.3 相关分析实例 .....	(164)	9.2.2 二项分布检验 .....	(212)
7.2 偏相关系数 .....	(168)	9.2.3 游程检验 .....	(214)



9.2.4 单样本 K-S 检验 .....	(216)	基本过程 .....	(248)
9.2.5 单样本非参数检验方法		10.4.3 主要分析结果 .....	(249)
比较 .....	(218)	10.5 多因变量线性模型的	
9.3 独立样本差异的显著性		方差分析 .....	(250)
检验 .....	(218)	10.5.1 完全窗口分析 .....	(250)
9.3.1 两个独立样本差异的		10.5.2 结果输出及其解释 .....	(252)
显著性检验 .....	(218)	本章小结 .....	(253)
9.3.2 多个独立样本差异的		习题十 .....	(253)
显著性检验 .....	(221)	<b>第 11 章 聚类分析与判别分析</b> .....	(256)
9.4 相关样本差异的显著性		11.1 聚类分析与判别分析过程	
检验 .....	(223)	概述 .....	(256)
9.4.1 两个相关样本差异的		11.1.1 聚类分析 .....	(257)
显著性检验 .....	(223)	11.1.2 判别分析 .....	(257)
9.4.2 多个相关样本差异的		11.1.3 Classify 的功能 .....	(258)
显著性检验 .....	(226)	11.2 快速聚类过程 .....	(258)
本章小结 .....	(228)	11.2.1 快速聚类的概念及	
习题九 .....	(228)	分析步骤 .....	(258)
<b>第 10 章 方差分析</b> .....	(232)	11.2.2 快速聚类的应用实例 .....	(259)
10.1 方差分析概述 .....	(232)	11.3 分层聚类分析 .....	(265)
10.1.1 方差分析的基本内容 .....	(234)	11.3.1 分层聚类的概念及	
10.1.2 方差分析的原理 .....	(235)	分析步骤 .....	(265)
10.1.3 F 分布 .....	(236)	11.3.2 分层聚类的应用实例 .....	(265)
10.2 单因变量单因素方差分析 .....	(236)	11.4 判别分析 .....	(271)
10.2.1 单因素方差分析的		11.4.1 判别分析的概念及分析	
基本原理 .....	(237)	步骤 .....	(271)
10.2.2 单因素方差分析的		11.4.2 判别分析的常用方法及	
基本过程 .....	(238)	判别函数效果的验证	
10.2.3 主要分析结果 .....	(242)	方法 .....	(272)
10.3 单因变量多因素方差分析 .....	(244)	11.4.3 判别分析的应用举例 .....	(273)
10.3.1 多因素方差分析的		本章小结 .....	(282)
基本原理 .....	(244)	习题十一 .....	(282)
10.3.2 多因素方差分析的		<b>第 12 章 因子分析和主成分分析</b> .....	(285)
基本过程 .....	(245)	12.1 基本原理 .....	(285)
10.4 协方差分析 .....	(247)	12.1.1 因子分析 .....	(285)
10.4.1 协方差分析的		12.1.2 主成分分析法 .....	(288)
基本原理 .....	(247)	12.2 基本分析过程 .....	(292)
10.4.2 协方差分析的		12.2.1 分析过程选择 .....	(292)

12.2.2 个例分析过程 .....	(297)	个例 .....	(328)
本章小结 .....	(302)	14.4 市场调查资料综合分析	
习题十二 .....	(302)	个例 .....	(332)
<b>第 13 章 时间序列分析</b> .....	(304)	本章小结 .....	(336)
13.1 指数平滑法 .....	(304)	习题十四 .....	(336)
13.1.1 基本原理 .....	(305)	<b>第 15 章 数据分析的基本原理</b> .....	(339)
13.1.2 基本过程 .....	(306)	15.1 数据的定义 .....	(339)
13.2 季节分解法 .....	(309)	15.2 数据的类型 .....	(340)
13.2.1 基本原理 .....	(309)	15.3 数据变量的属性 .....	(344)
13.2.2 基本过程 .....	(310)	15.3.1 尺度的意义及种类 .....	(344)
13.3 自回归法 .....	(311)	15.3.2 定类尺度 .....	(344)
13.3.1 基本原理 .....	(312)	15.3.3 定序尺度 .....	(345)
13.3.2 基本过程 .....	(312)	15.3.4 定距尺度 .....	(345)
13.4 自回归综合移动平均		15.3.5 定比尺度 .....	(346)
模型 (ARIMA) .....	(316)	15.3.6 市场调查的分析方法 .....	(348)
13.4.1 基本模型 .....	(316)	15.4 数据变换 .....	(349)
13.4.2 基本过程选择 .....	(316)	15.4.1 数据变换基本原理 .....	(349)
本章小结 .....	(318)	15.4.2 正态分布数据评价标准	
习题十三 .....	(318)	—Q-Q 图原理 .....	(351)
<b>第 14 章 个例综合分析</b> .....	(320)	15.4.3 数据正态化评价个例 .....	(352)
14.1 社会调查样本综合分析		15.5 数据和数据处理 .....	(356)
个例 .....	(320)	本章小结 .....	(357)
14.2 人事工资档案综合分析		习题十五 .....	(358)
个例 .....	(326)	<b>参考文献</b> .....	(360)
14.3 地区经济发展综合分析			

# 第1章 概 述

## 本章提要

本章介绍 SPSS 软件的特点、SPSS 软件对系统的基本要求。通过本章的学习,要求了解 SPSS 软件的基本特点;熟悉 SPSS 主要统计分析过程以及主要菜单、子菜单的英文名及其中文译名;能够对 SPSS 系统进行参数设置;能够获得即时提示和帮助。

随着人们生活水平的提高和市场经济的发展,计算机普及程度越来越高,大量的数据越来越倾向于用计算机来处理。目前市场上数据处理的软件很多,各有各的特色。有以资料查询为目的的 Access 数据库软件,有以数学分析为目的的 Mathematica 软件,有数据资料处理软件 Excel,也有调查(数据)资料统计分析软件 SPSS。Excel 是微软公司捆绑销售的办公软件,是一种非常直观的多页面电子表格,其特点是对软件、硬件要求低,应用灵活,适用于对样本量偏少的数据做一般性统计处理;SPSS 软件全面地吸取了电子表格的优点,虽然是单页面,但统计功能强,数据处理量大,是调查数据的专业统计分析、处理软件。本书主要介绍 SPSS 软件,由于其内容十分丰富,这里不可能介绍所有的内容,力求用 54 ~ 72 课时介绍 SPSS 基本内容。

20 世纪 60 年代后期,美国斯坦福大学 Norman H. Nie 等 3 位研究生研制了最初的 SPSS 统计分析系统,并成立了 SPSS 公司。随着微机的问世,20 世纪 70 年代该公司率先推出了 SPSS 的微机版本,占领了微机市场。1975 年组建了 SPSS 有限公司,总部设在芝加哥。1984 年 SPSS 公司推出第一个微机版——SPSS/PC + 2.0,1987 年推出了 SPSS 3.0 版本,而后又推出适用于 UNIX、Macintosh 和 OS/2 等操作系统的软件产品。1992 年 SPSS 5.0 版提供可视化窗口,而后,该公司推出了最新 SPSS 6.0、SPSS 7.5 版本,1999 年 9 月推出 SPSS 10.0 版本,2002 年 9 月推出 SPSS 11.5 版本,2003 年 9 月推出 SPSS 12.0 版本,更加丰富了统计绘图和高级统计功能,现在很多单位和个人使用的是 SPSS 10.0 ~ SPSS 12.0。SPSS 原是指“Statistics Package for Social Science”,即社会科学统计软件包,但其实际应用范围已远远超出社会科学范畴,在很多自然科学、生物医学科学领域也得到广泛的应用。SPSS 与 SAS (Statistics Analysis System)、

SYSTAT(SYSem for the STATistics)是并列的当今世界功能强大、应用广泛的 3 大统计软件。而且 SPSS 软件是世界上问世最早、应用最广、最易于操作的统计软件,同时也是我国最流行的社会调查资料分析软件。20 世纪 90 年代 SPSS 改称为“Statistical Product & Service Solutions”,即统计产品与服务解决方案,在 Windows 环境下执行,适用于 Windows 95、Windows 97、Windows 98 和 Windows 2000。SPSS 适用于市场营销、销售分析、市场调查、统计报告、质量控制、科学研究、社会调查、企业管理、教学及行政管理等领域,具有数据管理、统计分析、统计报表和图形展示打印等功能。SPSS 软件一般包括 7 个模块:数据录入、编辑和管理;调查资料汇总,统计绘图与制表;多变量统计;线性模型;广义线性模型;统计预测和响应分析;因子分析和路径分析等。本书介绍的是 SPSS 12.0 版本,它能满足资料分析的绝大部分需要。

## 1.1 特 点

SPSS 具有 Windows 软件的一般特点,如命令语句、子命令及选择项的选择,绝大部分由对话框的操作完成。因此,用户无需花各种时间记忆各种命令、过程、选择项。除数据输入工作需要使用键盘完成外,大多数操作都是通过“菜单”、“图标按钮”、“对话框”来完成的,因此操作简便,易于学习、使用。

此外,为完成对数据的统计分析工作,本软件还有以下特点。

① 对于常见的统计分析方法完全可以通过对“菜单”、“对话框”的操作完成,因此无需编写程序。对于熟悉 SPSS/PC 语言的老用户也可以在语句窗中直接输入程序语句,然后单击“Run”按钮提交运行,或者可用对话框指定命令、子命令和选择项后,单击“Paste”按钮将与选择对应的语句自动置于“Syntax”窗中,并以文件形式保存。这样,在数据量很大的情况下,通过执行批处理命令,提高计算机运行效率。SPSS 适用于新、老用户。

② 具有第 4 代语言的特点,只要通过菜单的选择、对话框的操作,即可告诉系统要做什么,无需告之具体怎样做。只要了解统计分析原理,即可得到统计分析结果。

③ 与其他软件有数据转换接口。其他软件生成的数据文件,例如关系数据库生成的 DBF 文件或用文本编辑软件生成的 ASCII 码数据文件,均可方便地转换成可供分析的 SPSS 的数据文件。该软件生成的表格、图形等文件很容易复制或剪贴到 Excel、Word 文件中。

④ 分析方法丰富,本软件提供了从简单描述数据统计分析到多因素数据分

析等大多数的统计分析技术,兼有很强的统计图表的生成、编辑功能。本软件功能全、运行速度快,既可编程又可通过按钮运行,且内存可调整。可同时运行 300 多个变量、100 万个以上样本的数据资料文件。

## 1.2 SPSS 的运行和运行环境

### 1.2.1 对运行环境的要求

SPSS 软件包适应性强,对计算机硬件要求条件低,SPSS 8.0 版本占 3 个文件夹、434 个文件和 56 MB 硬盘空间,SPSS 10.0 版本占 5 个文件夹、446 个文件和 74 MB 硬盘空间,SPSS 11.5 版本占 12 个文件夹和 142 MB 硬盘空间,SPSS 12.0 版本占 17 个文件夹和 225 MB 硬盘空间。运行 SPSS 软件最低要求 16 MB 内存容量。由于统计分析软件的数据量往往很大,矩阵运算要求的内存空间就比较大。如要求进行多因素分析,最好配备 32 MB 以上内存。

根据经验,硬盘容量在装配了软件之后,自由空间应该占总容量的 30% 以上;否则,存取速度会有较明显的下降。用户可以根据这一经验和所要装配的模块的大小检查自己的硬盘空间是否合适。

SPSS 汉化版本不够成熟,大部分版本屏幕、提示、窗口为英文环境。SPSS 的变量、变量值、变量标签(说明)、变量值说明(标签)及其输出是能够汉化的,其可在中文版 Windows 98、Windows 2000 及 Windows XP 下直接运行。如在英文版 Windows 环境下,则往往需要借助“中文之星”软件才能设计中文标签。但是,能够汉化的仅是变量、变量值、变量标签、值标签等,对于大量的菜单、提示、输出格式等都不能汉化,即使勉强汉化,翻译得也不是很准确,容易造成大量的理解困难等问题。这是初学者遇到的主要困难之一。

### 1.2.2 系统的运行及主屏幕介绍

打开 SPSS 软件,主屏幕就会出现 6 项选择的屏幕提示,如图 1.1 所示。

- ① “Run the tutorial”是运行 SPSS 电子教程,即进入图解帮助程序。
- ② “Type in data”是在 SPSS 提供的电子表格中输入资料。
- ③ “Run an existing query”是执行一个已经存在的 SPSS 数据库文件。
- ④ “Create new query using Database Wizard”是在资料收集系统引导下,一步一步创建新文件。

⑤ “Open an existing data source”为默认选项,打开已存在的 SPSS 文件。

⑥ “Open another type of file”为打开已存在的 SPSS 其他文件,如输出文件。

若选中“Don't show this dialog in the future”选项,那么以后开机时将不出现此显示屏幕。

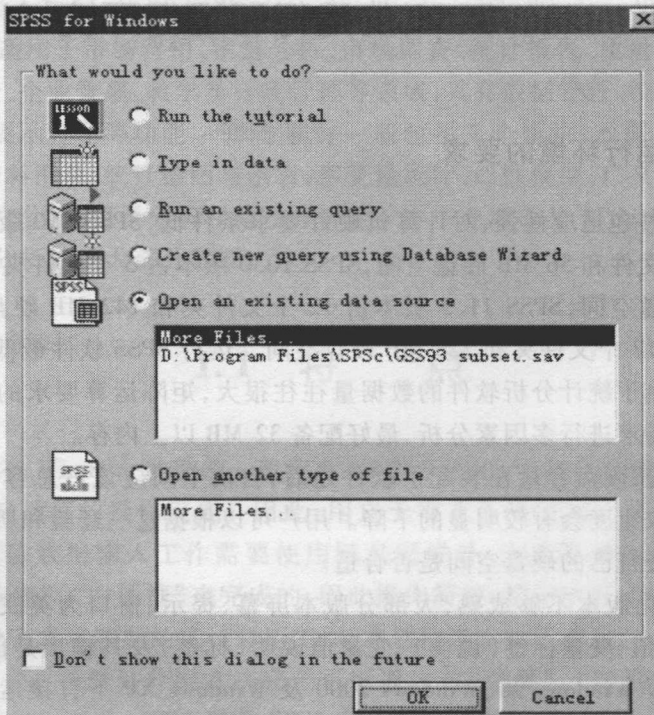


图 1.1 SPSS 起始画面

如果选择⑤,则打开已存在的 SPSS 数据文件;也可单击“Cancel”按钮,自行输入资料,接着打开文件夹(Open File),选择计算机自带文件“World95. sav”,则屏幕上出现一电子表格,如图 1.2 所示。

图 1.2 中第一行是标题栏,第二行、第三行和第四行分别是菜单栏、标准工具栏和编辑栏。再下面是电子表格,横栏表示列标题栏(变量名),纵列表示行标签,通常表示被调查的个案数。

标题栏位于 SPSS 表格的第一行,显示正在操作的 SPSS 文件名,如果是一个新文件且没有保存,则显示为默认文件名“Untitled”。

菜单栏位于 SPSS 表的第二行,SPSS 共有“File”(文件)、“Edit”(编辑)、“View”(视图)、“Data”(数据)、“Transform”(转换)、“Analyze”(分析)、

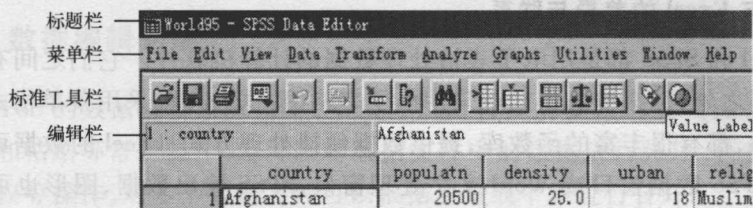


图 1.2 SPSS 数据文件电子表格

“Graphs”(图形)、“Utilities”(通用)、“Windows”(窗口)和“Help”(帮助)10 个菜单。

标准工具栏列出可供使用的快捷命令按钮,位于 SPSS 表的第三行,单击后可迅速执行这些命令。标准工具栏列出的快捷命令按钮共有 16 个,其依次分别是打开文件、文件存盘、文件打印、对话回忆、退回/重做、转到图形、指定样本、寻找指定变量、寻找某变量值、插入某样本、插入某变量、样本分类、样本加权、样本选择、值变量标签的显示、集的设置等。

编辑栏位于 SPSS 表格的第四行,用于编辑、输入公式或数据的区域。

最重要、最富有特色的是菜单栏中“Data”(数据)、“Transform”(转换)、“Analyze”(分析)菜单。“Data”是指变量定义,具体包括数据的长度,小数点的定义,给变量贴标签,数据或变量的插入,数据合并、加权,数据选择、分组等。“Transform”是指变量转换,通过计算使之形成新变量或重新编码使之形成新变量。“Analyze”是指变量统计分析,具体包括变量总计、相关和回归、平均数比较、聚类分析等。

SPSS 的窗口共分七种:资料编辑窗口(开始窗口)、资料输出窗口、句法编辑窗口、快速表格编辑窗口、图形编辑窗口、文件输出编辑窗口、句法编辑窗口和手稿编辑窗口。常用的为前 3 个。

① “Untitled”资料编辑窗口:可定义变量,输入、编辑数据文件。

② “Output”资料输出窗口:显示程序或过程的运行信息、输出结果。

③ “Syntax”句法编辑窗口:可写入程序,并逐行执行。

SPSS 启动后的主屏幕可直接出现某一窗口。启动后立即显示相应窗口是在系统初始状态中设置的。

### 1.2.3 与其他数据处理软件的联系与区别

随着计算技术的发展,社会上数据处理软件越来越多,那么与其他软件相比,SPSS 软件有什么特色?在何时应选用 SPSS 软件呢?

### 1. 与 Excel 的差异与联系

Excel 和 SPSS 都是用电子表格处理数据的计算机软件。它们之间有同亦有异。相同之处,数据输入格式都采用电子表格,数据处理都采用菜单格式的统计分析软件,都有很丰富的函数库,提供数据便捷处理方式;Excel 的数据可以直接复制到 SPSS 数据窗口或 Word 文字处理窗口,SPSS 输出数据、图形也可直接复制到 Excel 窗口或 Word 文字处理窗口。

其差异为:Excel 是普及型数据处理软件,常包含在办公系统内,Excel 软件非常适合于各种数学运算,统计功能相对较弱,有中文版本,使用人数较多,Excel 适合处理小数量数据,样本数常为几十到几千个,变量一般为几十个,Excel 对计算机软件、硬件要求比较低,占用计算机空间小;SPSS 则是专业性数据处理软件,适合于各种统计运算,不适合进行灵活的数学运算,没有中文版本,大多是专业人员使用,SPSS 适合于处理大规模数据,样本可达几千万个,变量也可达数百个,SPSS 对计算机软件、硬件要求相对高,占用计算机空间较大。

此外,从形式上讲,Excel 可有多张电子表格,数据的排列随意性比较大,数据处理比较灵活;而 SPSS 仅有一张电子表格,对数据的排列有严格要求,“行”为样本,“列”为变量,数据处理格式比较固定。从功能上讲,Excel 仅能借助函数完成简单的统计分析;而 SPSS 能够独立完成统计分析过程,能够借助这些统计功能完成所有的统计任务。

Excel 发展多年,从格式到模型比较完善、成熟,近年改动较少;SPSS 每次改版变化都比较大,如 SPSS 11.0 的版本在 SPSS 10.0 版本基础上删除了不少统计功能,而在最新的 SPSS 12.0 版本又将统计分析功能增加到 16 个,并增加完善了电子地图功能,所以 SPSS 尚在不断完善之中。

### 2. 与 FoxBASE、Access 的差异与联系

FoxBASE、Access 是数据间查询数据库,注重的是数据外部逻辑关系、外部结构问题,不涉及数据内部统计、分析、结构问题;SPSS 研究数据内部的数理关系或逻辑结构,具体是数据的加工处理、数据挖掘处理。

## 1.3 SPSS 的主要功能

SPSS 可实现数据处理一般过程,如数据编辑、表格生成、图形生成和编辑、与其他软件的连接和各种统计分析、计算过程,下面分别予以介绍。



### 1.3.1 数据编辑功能

在 SPSS 的数据编辑器窗口中,不仅可以对打开的变量进行增加、删除、复制、剪切和粘贴等常规操作,还可以对数据文件中的数据进行排序、转置、拆分、聚合、加权等操作,对多个数据文件可以根据变量或个案进行合并。

### 1.3.2 表格的生成和编辑

利用 SPSS 可生成数十种不同风格的表格,按功能分为一般表、多响应表和频数表等。利用专门的编辑窗口或直接在查看器中可以编辑所生成的表格。

### 1.3.3 图形的生成和编辑

利用 SPSS 可以生成数十种基本图和交互图。基本图包括条形图、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、序列图和时间序列图等,有的基本图中又可进一步细分。交互图比基本图更漂亮,包括不同风格的二维图、三维图。交互图也包括条形交互图、点形交互图、线形交互图、带形交互图、饼形交互图、箱形交互图、误差条形交互图、直方交互图和散点交互图等。图形生成以后,可以进行编辑。SPSS 提供了强大的图形编辑功能,然而,很多图形同样可以在 Excel 软件中获得,所以本文仅以较少的笔墨予以描述。

### 1.3.4 与其他软件的连接

SPSS 能直接打开 Excel、dBase、FoxBASE、Lotus 1 - 2 - 3、Access 等文件和文本编辑器等生成的数据文件。SPSS 生成的图形可以保存为多种图形格式。高版本的 SPSS 软件支持 OLE 技术和 ActiveX 技术,使生成的表格或交互图对象可以与其他同样支持该技术的软件进行自动嵌入与链接。SPSS 还有内置的 VBA 客户语言,可以通过 Visual Basic 编程来控制 SPSS。

### 1.3.5 统计功能

统计功能是 SPSS 的核心,利用该软件,可以又快又好地完成所有的数理统计任务。各版本的 SPSS 统计功能是不同的。这里以 SPSS 12.0 版本为例。SPSS 的基本统计功能包括:样本数据的描述性分析;统计检验(包括参数检验及非参数检验);方差分析;列联表;相关分析;回归分析;聚类分析;判别分析;因