

· 高等学校计算机基础教育教材精选 · 分级教学系列教材

统计分析方法 与SPSS应用教程

杨小平 主编

刘刚 赵晋 编著



清华大学出版社

· 食尚 畅销

· 人生，得此装饰，她就是最美丽的！青春的开始，是慢慢积累的过程，而青春的结束，是突然的。人生，得此装饰，她就是最美丽的！青春的开始，是慢慢积累的过程，而青春的结束，是突然的。人生，得此装饰，她就是最美丽的！青春的开始，是慢慢积累的过程，而青春的结束，是突然的。

· 高等学校计算机基础教育教材精选 ·
分级教学系列教材

统计分析方法 与SPSS应用教程

杨小平 主编

刘刚 赵晋 编著

清华大学出版社
北京

内 容 简 介

全书介绍了常用的统计方法,内容包括参数估计、参数假设检验与非参数假设检验、方差分析、回归分析、聚类分析、判别分析等。教材体系基本按照工科及管理各专业大学统计教材大纲组织,同时增添了一个实用统计分析方法。为了满足读者需求,书中对很多内容做了技术处理,尽量回避那些较深的统计理论,强调实际应用。另外,书中附有大量的典型实例,并给出整个 SPSS 处理及分析过程,引导读者对照实践,快速掌握。

本教材可作为工科及管理学科大学本科生、研究生的教科书,同时也可供科技人员参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

统计分析方法与 SPSS 应用教程/杨小平主编; 刘刚, 赵晋编著. —北京: 清华大学出版社, 2008. 12

(高等学校计算机基础教育教材精选·分级教学系列教材)

ISBN 978-7-302-18608-3

I. 统… II. ①杨… ②刘… ③赵… 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(2008)第 146091 号

责任编辑: 索 梅 徐跃进

责任校对: 梁 毅

责任印制: 王秀菊

出版发行: 清华大学出版社 地址: 北京清华大学学研大厦 A 座

http://www.tup.com.cn 邮 编: 100084

社 总 机: 010-62770175 邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者: 北京国马印刷厂

装 订 者: 三河市新茂装订有限公司

经 销: 全国新华书店

开 本: 185×260 印 张: 17.75 字 数: 434 千字

版 次: 2008 年 12 月第 1 版 印 次: 2008 年 12 月第 1 次印刷

印 数: 1~4000

定 价: 28.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系
调换。联系电话: (010)62770177 转 3103 产品编号: 019015-01

出版说明

——高等学校计算机基础教育教材精选——

在教育部关于高等学校计算机基础教育三层次方案的指导下,我国高等学校的计算机基础教育事业蓬勃发展。经过多年的教学改革与实践,全国很多学校在计算机基础教育这一领域中积累了大量宝贵的经验,取得了许多可喜的成果。

随着科教兴国战略的实施以及社会信息化进程的加快,目前我国的高等教育事业正面临着新的发展机遇,但同时也必须面对新的挑战。这些都对高等学校的计算机基础教育提出了更高的要求。为了适应教学改革的需要,进一步推动我国高等学校计算机基础教育事业的发展,我们在全国各高等学校精心挖掘和遴选了一批经过教学实践检验的优秀教学成果,编辑出版了这套教材。教材的选题范围涵盖了计算机基础教育的三个层次,包括面向各高校开设的计算机必修课、选修课以及与各类专业相结合的计算机课程。

为了保证出版质量,同时更好地适应教学需求,本套教材将采取开放的体系和滚动出版的方式(即成熟一本,出版一本,并保持不断更新),坚持宁缺勿滥的原则,力求反映我国高等学校计算机基础教育的最新成果,使本套丛书无论在技术质量上还是文字质量上均成为真正的“精选”。

清华大学出版社一直致力于计算机教育用书的出版工作,在计算机基础教育领域出版了许多优秀的教材。本套教材的出版将进一步丰富和扩大我社在这一领域的选题范围、层次和深度,以适应高校计算机基础教育课程层次化、多样化的趋势,从而更好地满足各学校由于条件、师资和生源水平、专业领域等的差异而产生的不同需求。我们热切期望全国广大教师能够积极参与到本套丛书的编写工作中来,把自己的教学成果与全国的同行们分享;同时也欢迎广大读者对本套教材提出宝贵意见,以便我们改进工作,为读者提供更好的服务。

我们的电子邮件地址是: jiaoh@tup.tsinghua.edu.cn; 联系人: 焦虹。

清华大学出版社

2001年8月

前言

—统计分析方法与 SPSS 应用教程—

随着时代的发展,对统计分析任务的需求在飞速增加。目前在普通高校的工科、管理等专业都普遍开设有“概率统计”、“数理统计”等必修课,有的甚至还开设有“实用统计方法”、“多元分析”等选修课,以适应社会形势的要求。统计分析是一门实用性很强的学科,经常要分析处理大量的实际数据,由于课时等原因所限,现时的教材主要以讲述统计原理为主,缺少应用性指导。基于此,我们编写了这本教材,一方面简单介绍统计理论,但偏重于方法的统计背景,另一方面以经典的 SPSS 软件作为平台,介绍各种统计方法的应用实现,旨在强调理论与应用的有机结合。

为了兼顾工科、管理等专业强调理解理论,强化应用的特点,我们力求合理组织教材内容。全书大致可以分为 3 个部分。第 1 部分介绍作为应用基础平台的 SPSS 软件的一些基本功能,主要包括 SPSS 的运行环境、数据文件的编辑与管理、统计图形、基本统计分析等内容。第 2 部分主要介绍作为统计推断基本问题的假设检验,包括参数假设检验与非参数假设检验。第 3 部分为实用统计分析方法,其中包括方差分析、回归分析、聚类分析与判别分析,这些内容基本上属于多元分析的范畴。我们对后两部分内容作了针对性处理,前者读者较为熟悉,尽量介绍统计原理;后者则侧重于讲述统计方法应用的背景与思想,强调实际应用。

这本教材的编写,是编者在理论与应用结合上的一种尝试。希望通过此书的学习,加深读者对统计理论的理解,同时提高实际分析能力。鉴于水平所限,书中错误在所难免,敬请读者朋友们批评指正。

作者

2008 年 8 月

目 录

统计分析方法与 SPSS 应用教程

第 1 章 SPSS 简介	1
1.1 SPSS 概述	1
1.2 SPSS 的运行方式	3
1.3 SPSS 13.0 的安装	3
1.4 SPSS 13.0 的启动和退出	7
1.4.1 SPSS 的启动	7
1.4.2 SPSS 的退出	8
1.5 SPSS 的数据编辑窗口	8
1.5.1 SPSS 标题栏和系统状态栏	9
1.5.2 SPSS 菜单栏	10
1.5.3 SPSS 工具栏	10
1.5.4 SPSS 数据编辑窗口的移动与调整	11
1.6 SPSS 的帮助系统	11
1.6.1 主题帮助	11
1.6.2 在线指南	13
1.6.3 统计分析指导	15
1.6.4 对话框中的 Help 按钮	15
1.6.5 内容帮助	16
1.7 SPSS 运行环境设置	18
1.7.1 SPSS 状态栏的显示和隐藏	18
1.7.2 SPSS 工具栏的显示和隐藏	18
1.8 SPSS 系统参数的设置	18
1.8.1 Options 选择对话框	19
1.8.2 SPSS 通用参数设置	19
1.8.3 结果输出窗口参数设置	21
1.8.4 Currency 窗口参数设置	23
1.9 小结	24
习题 1	24

第2章 数据文件的编辑与管理	25
2.1 数据文件的建立	25
2.1.1 定义变量	25
2.1.2 数据的输入	29
2.1.3 数据的保存	30
2.2 数据文件的编辑	32
2.2.1 单元值的查找	32
2.2.2 增加或删除一个个案	32
2.2.3 观测数据的排序	33
2.2.4 观测数据的行列互换	33
2.2.5 选取观测数据的个案子集	34
2.2.6 观测数据归类分组汇总	35
2.2.7 缺失值的替代	37
2.3 数据变量的操作	37
2.3.1 增加或删除一个变量	37
2.3.2 新变量的计算生成	39
2.3.3 产生计数变量	40
2.3.4 产生分组变量	42
2.3.5 产生自动分组变量	44
2.3.6 变量集的定义和使用	46
2.4 数据文件的合并与拆分	48
2.4.1 数据文件的纵向合并	48
2.4.2 数据文件的横向合并	49
2.4.3 数据文件的拆分	50
2.5 读入其他格式数据文件	52
2.5.1 读取固定格式的文本文件	52
2.5.2 读取自由格式的文本文件	56
2.5.3 读取 Excel 文件	59
2.6 小结	60
习题 2	60
第3章 图形功能	63
3.1 统计图的创建概述	63
3.2 条形图	65
3.2.1 条形图的类型	65
3.2.2 SPSS 实践 1: 个案分组描述模式的简单条形图	66
3.2.3 SPSS 实践 2: 变量描述模式的复合条形图	69
3.2.4 SPSS 实践 3: 个案描述模式的堆栈条形图	71

第3章	统计图表	72
3.1	柱形图	72
3.1.1	柱形图的类型	72
3.1.2	SPSS 实践 1：个案分组描述模式的单柱形图	73
3.1.3	SPSS 实践 2：变量描述模式的单柱形图	75
3.2	条形图	76
3.2.1	条形图的类型	76
3.2.2	SPSS 实践：个案分组描述模式的简单条形图	77
3.3	线图	78
3.3.1	线图的类型	78
3.3.2	SPSS 实践 1：个案分组描述模式的单线图	79
3.3.3	SPSS 实践 2：变量描述模式的垂线图	79
3.4	面积图	80
3.4.1	面积图的类型	80
3.4.2	SPSS 实践：个案分组描述模式的简单面积图	81
3.5	饼形图	82
3.5.1	饼形图的类型	82
3.5.2	SPSS 实践：个案分组描述模式的饼形图	83
3.6	箱形图	84
3.6.1	箱形图的类型	84
3.6.2	SPSS 实践：变量描述模式的复合箱形图	85
3.7	散点图	86
3.7.1	散点图的类型	86
3.7.2	SPSS 实践：简单散点图	87
3.8	直方图	88
3.9	P-P 概率图	89
3.10	Q-Q 概率图	90
3.11	图形编辑	91
3.11.1	图形编辑窗口	91
3.11.2	图形转换	92
3.11.3	图形修饰	93
3.12	小结	94
习题 3		94
第4章	基本统计分析	96
4.1	基础知识	96
4.1.1	基本概念	96
4.1.2	常用统计量	97
4.1.3	变量的标准化变换	99
4.2	频数分析	99
4.2.1	Frequencies 对话框	99
4.2.2	频数分析实践	102
4.3	描述分析	104
4.3.1	Descriptives 对话框	104
4.3.2	描述分析实践	105
4.4	探索分析	106

4.4	4.4.1 Explore 对话框	106
	4.4.2 探索分析实践.....	109
4.5	4.5 交叉列联表分析	113
	4.5.1 交叉列联表分析.....	113
	4.5.2 Crosstabs 对话框	114
	4.5.3 交叉列联表分析实践.....	116
4.6	4.6 多选项分析	119
	4.6.1 多选项分析.....	119
	4.6.2 多选项分析实践.....	121
4.7	4.7 小结	125
	习题 4	125
第 5 章 参数假设检验 128		
5.1	5.1 假设检验概述	128
	5.1.1 基本概念.....	128
	5.1.2 原理和步骤.....	129
	5.1.3 两类错误.....	130
5.2	5.2 单样本的均值检验	131
	5.2.1 统计原理.....	131
	5.2.2 SPSS 实现	132
5.3	5.3 两个独立样本的均值差异性检验	134
	5.3.1 统计原理.....	134
	5.3.2 SPSS 实现	135
5.4	5.4 两个配对样本的均值差异性检验	138
	5.4.1 统计原理.....	138
	5.4.2 SPSS 实现	138
5.5	5.5 小结	141
	习题 5	142
第 6 章 非参数假设检验 143		
6.1	6.1 SPSS 的非参数检验	143
6.2	6.2 拟合优度 χ^2 检验	144
	6.2.1 统计原理.....	144
	6.2.2 SPSS 实现	145
	6.2.3 推广应用.....	146
6.3	6.3 二项检验	148
	6.3.1 统计原理.....	148

6.3.2 SPSS 实现	149
6.4 单样本 K-S 检验	151
6.4.1 统计原理	151
6.4.2 SPSS 实现	152
6.5 随机性检验	154
6.5.1 统计原理	154
6.5.2 SPSS 实现	155
6.6 两个独立样本的非参数检验	157
6.6.1 问题的提出	157
6.6.2 常用检验方法	158
6.6.3 SPSS 实现	160
6.7 多个独立样本的非参数检验	163
6.7.1 常用统计方法	163
6.7.2 SPSS 实现	165
6.8 两个相关样本的非参数检验	168
6.8.1 常用统计方法	168
6.8.2 SPSS 实现	170
6.9 多个相关样本的非参数检验	173
6.9.1 常用统计方法	173
6.9.2 SPSS 实现	177
6.10 小结	181
习题 6	181
第 7 章 方差分析	182
7.1 统计背景	182
7.2 单因素方差分析	183
7.2.1 统计思想与原理	183
7.2.2 SPSS 实现	184
7.2.3 结果分析	187
7.3 多重比较	188
7.4 双因素方差分析	189
7.4.1 交互作用	189
7.4.2 无交互作用的双因素方差分析	190
7.4.3 有交互作用的双因素方差分析	191
7.4.4 SPSS 实现	192
7.4.5 结果分析	196
7.5 小结	199
习题 7	199

第8章 回归分析	200
8.1 一元线性回归	200
8.1.1 回归模型及参数估计	200
8.1.2 回归方程的显著性检验	202
8.1.3 预测	203
8.1.4 SPSS 实现	203
8.2 多元线性回归	207
8.2.1 回归模型	207
8.2.2 参数估计	208
8.2.3 显著性检验	209
8.2.4 预测	210
8.2.5 SPSS 实现	211
8.2.6 逐步回归分析	213
8.3 残差分析	216
8.3.1 基本思想	216
8.3.2 各种形式的残差图	216
8.3.3 SPSS 实现	218
8.4 曲线拟合	220
8.4.1 统计背景	220
8.4.2 SPSS 实现	222
8.5 小结	225
习题 8	225
第9章 聚类分析	226
9.1 概述	226
9.2 系统聚类法	227
9.2.1 样品间的距离	227
9.2.2 类间的距离	229
9.2.3 SPSS 实现	230
9.2.4 结果分析	234
9.3 快速聚类法	235
9.3.1 基本思想	235
9.3.2 SPSS 实现	236
9.3.3 结果分析	239
9.4 小结	240
习题 9	240

第 10 章 判别分析	241
10.1 概述	241
10.2 距离判别、Bayes 判别和 Fisher 判别	243
10.2.1 距离判别	243
10.2.2 Bayes 判别	244
10.2.3 Fisher 判别	245
10.2.4 SPSS 实现	246
10.2.5 结果分析	250
10.3 逐步判别	255
10.4 小结	258
习题 10	258
附录 A 应用范文	259
参考文献	265

SPSS(Statistical Package for Social Science,社会科学统计软件包) for Windows 是目前国内最为流行的统计软件之一,它在各个领域中都得到了广泛的普及和应用。本章介绍 SPSS 的统计特色与基本功能。

1.1 SPSS 概述

1. 智能操作,易学易用

在大多数情况下,SPSS 的操作是通过选择菜单、图标按钮和对话框选项来完成的,用户只要掌握了主窗口图标按钮的功能,就可以选择功能按钮来完成相应的统计分析任务。通过选择菜单和所需对话框选项的操作,命令 SPSS 系统去做该做的工作,而不必告诉系统应该怎样去做。

SPSS 的智能操作过程简单易学,用户只需掌握统计分析的基本原理,便可迅速对统计分析的输出结果做出适当的解释和判断,从而完成统计分析任务。

2. 一般情况下无须编写程序

SPSS 软件实际上有自己独立的程序语言。低版本的 SPSS 软件由于受到功能的限制,各级菜单中没有相应统计方法的按钮,因此就需要在语句窗口中利用 SPSS 程序语言编写相应的程序语句,然后提交软件运行。目前的高版本软件集成了高智能化设计,使得常用的统计分析方法均可在菜单栏中找到相应的选项或命令,用户可以通过选择按钮来完成所作的统计计算。高版本软件都是向下兼容的,对于熟悉 SPSS 程序语言的用户仍可根据习惯来操作 SPSS 系统;但对于新用户来讲,利用屏幕显示的 SPSS 智能化对话框中的命令语句、子命令以及选择项,就可以完成所需的绝大部分统计分析需求,再也不必花大量时间来记忆繁杂的程序语言中的命令语句、过程语句以及选择项等。用户只须掌握基本的统计分析思想和统计分析方法,明确自己要做什么,再通过对话框找出所需的统计分析选项,就可以命令 SPSS 系统完成所需要的统计分析任务。

3. 数据转换接口有很好的兼容性

SPSS 可以方便地将其他软件生成的数据文件转换成可供分析的 SPSS 的数据文件,例如 dBase 数据文件、文本编辑生成的 ASCII 码数据文件、Office 中的 Excel 数据文件等。

4. 具有丰富的统计分析功能和完善的分析报告功能

SPSS 的统计分析功能是 SPSS 系统的核心部分,它包含基本统计分析和实用统计分析两大类,同时 SPSS 还提供了全新的常规图文操作界面,以帮助输出完善的统计分析报告。

1) 基本统计分析

SPSS 的基本统计分析功能相当丰富,其中主要包括观测数据的频数分析、描述分析、探索分析、交叉列联表分析、多选项分析、线性组合测量分析、多重响应分析等。这些功能可以完成对观测数据总体特征的分析,例如求观测数据的平均数、中位数、众数、方差和标准差、偏度和峰度等反映分布特征的描述统计量,还可以进行观测数据的变化趋势分析以及输出基本统计报表等。

2) 实用统计分析

新版 SPSS 的实用统计分析方法几乎涵盖了所有常用数据统计方法,如参数估计、参数假设检验、非参数假设检验、单因素方差分析与多因素方差分析、回归分析(包括 Logistic 回归、线性回归、Probit 回归、Cox 回归、曲线拟合等)与相关分析、聚类分析、判别分析、因子分析、信度分析、可靠性分析、生存分析及时间序列分析等。

3) 全新的常规图文操作界面

SPSS 的图文界面操作非常便捷,内容十分丰富,完善的统计分析报告可以提供强大的图表生成、编辑与输出等功能。系统可以输出多种清晰的统计表格和各种形象而直观的二维、三维图形,例如条形图、线图、饼形图、直方图、箱形图、散点图、帕雷托图、P-P 概率图、Q-Q 概率图、曲线图、时间序列图等。新版的 SPSS 在常规图中引入了更多的交互图功能,例如图组(paneled charts)、带误差线的分类图形、误差条形图和线图、三维效果的简单图、堆积图和分段数据的饼形图等等。系统还增加了诸如人口金字塔和点密度图等几种新的图形,用户还可以根据需要将不同的图形在同一报表中输出。

新版 SPSS 的输出过程几乎都不再使用文本,而是用更美观的枢轴表。这些枢轴表可以被直接导出到 PowerPoint 中,极大地方便了用户的使用。用户可以对统计量进行排序,在输出表格中根据需要进行合并或更改等操作。

5. Complex Samples 模块增加了统计建模的功能

SPSS 中的 Complex Samples 模块通常用于实现复杂抽样的设计方案,以及对相应的分类数据进行描述。新版 SPSS 对该模块进行了改进,完善了统计建模功能,使得一般线性模型可以被完整地引入到复杂抽样模块中,再对市场调研中的客户满意度数据进行建模,实现对复杂抽样研究中各种连续性变量的建模预测,从而使得复杂抽样的统计推断模型方法得到了推广。

从抽样设计、统计描述到复杂统计建模分析过程,系统都会自动使用 Logistic 回归进行复杂抽样来寻找影响因素,使多阶段分层整群抽样或者复杂的 PPS 抽样等方法都可以

在该模块中轻松地完成。方差分析模型、线性回归模型、Logistic 回归模型等复杂的统计模型也可以在该模块中得到广泛使用。

1.2 SPSS 的运行方式

SPSS 主要使用如下 3 种运行方式。

1. 完全窗口菜单运行方式

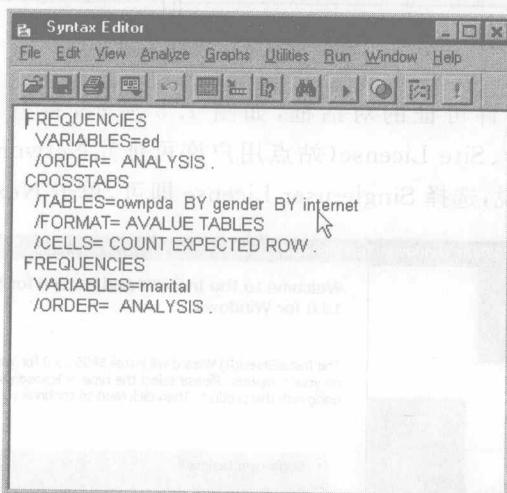
通过选择窗口菜单和对话框完成各种操作,由于不需要编程,且简单易用,成为大多数 SPSS 用户的常用方式。

2. 批处理方式

将编写好的 SPSS 语句程序存为一个文件,在 SPSS 的 Production Facility 程序中打开并运行。

3. 程序运行方式

这种方式是在命令(Syntax)窗口中直接运行编写好的 SPSS 语言程序或者在脚本(Script)窗口中运行脚本程序,如图 1.1 所示。



```
FREQUENCIES
  VARIABLES=ed
  /ORDER= ANALYSIS .
CROSSTABS
  /TABLES=ownpda BY gender BY internet
  /FORMAT=AVALUE TABLES
  /CELLS= COUNT EXPECTED ROW .
FREQUENCIES
  VARIABLES=marital
  /ORDER= ANALYSIS .
```

图 1.1 Syntax 窗口 SPSS 的语法

1.3 SPSS 13.0 的安装

具有计算机常识的用户都会有应用软件的安装经验,下面以 SPSS 13.0 为例介绍 SPSS 系统的安装。在整个安装过程中,用户可以自由选择要安装的组件,以满足将来的

统计分析工作需要。SPSS 13.0 的安装步骤如下。

- (1) 启动 Windows 后,进入 SPSS 13.0 安装文件所在的文件夹,双击打开 setup.exe 文件,即开始运行安装程序。
- (2) 程序运行后,出现安装程序的 SPSS 13.0 for Windows-Install Shield Wizard 对话框,如图 1.2 所示,单击 Next 按钮继续安装。

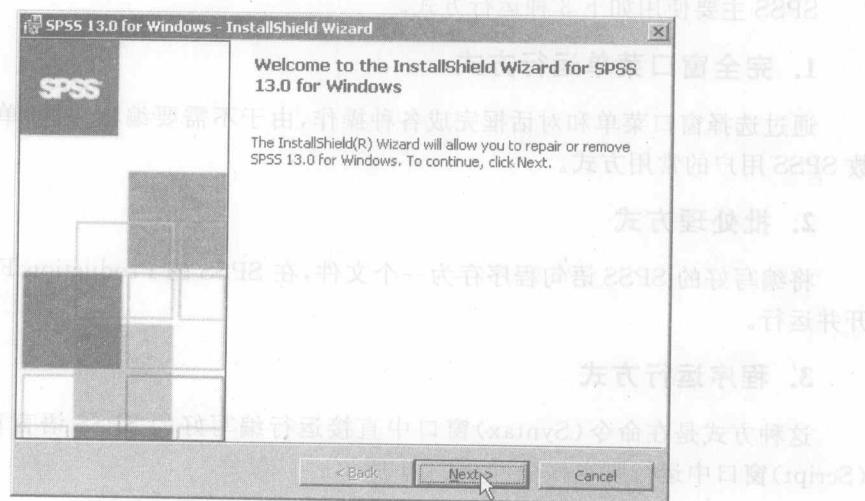


图 1.2 Install Shield Wizard 对话框

- (3) 出现选择软件许可证的对话框,如图 1.3 所示,将有 3 种选择: Single-user License(单用户许可证)、Site License(站点用户许可证)、Network License(网络用户许可证)。对于单用户来说,选择 Single-user License 即可,单击 Next 按钮继续安装。

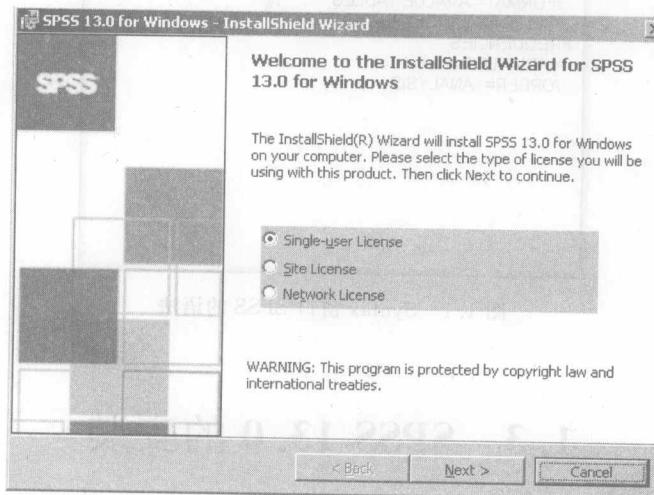


图 1.3 许可证类型对话框

(4) 出现 License Agreement 许可证协议对话框,如图 1.4 所示,选择 I accept the terms in the license agreement 接受协议,单击 Next 按钮。

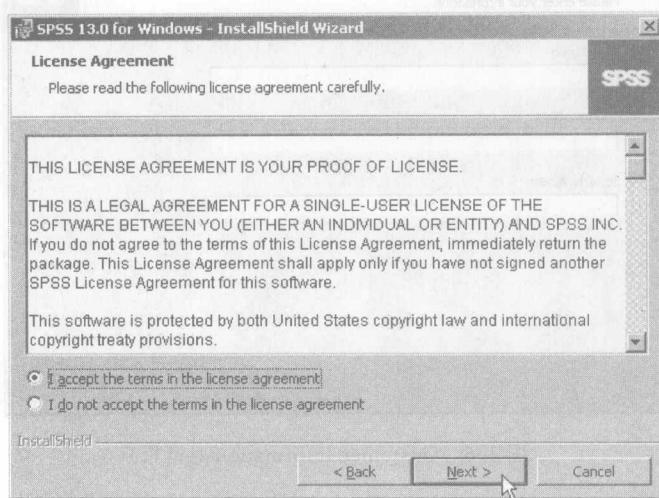


图 1.4 License Agreement 对话框

(5) 系统出现 ReadMe Information 对话框,如图 1.5 所示,用户可以通过对话框了解 SPSS 的各种信息,单击 Next 按钮。

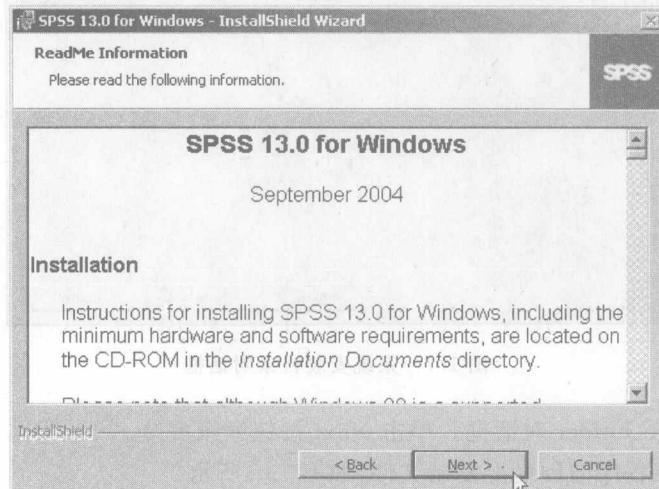


图 1.5 ReadMe Information 对话框

(6) 出现用户信息 Customer Information 对话框,如图 1.6 所示,此时应在相应位置输入本人的名字、所在组织的名字、软件的序列号,输入完成后单击 Next 按钮继续安装。

(7) 出现选择安装目录 Destination Folder 对话框,如图 1.7 所示。系统默认的安装目录为 C:\Program Files\SPSS\,如果需要改变,单击 Change 按钮,打开选择目录 Choose Directory 子对话框,可选择希望安装的目标文件夹的位置,再单击 OK 按钮,即