

第2版

赢在职场

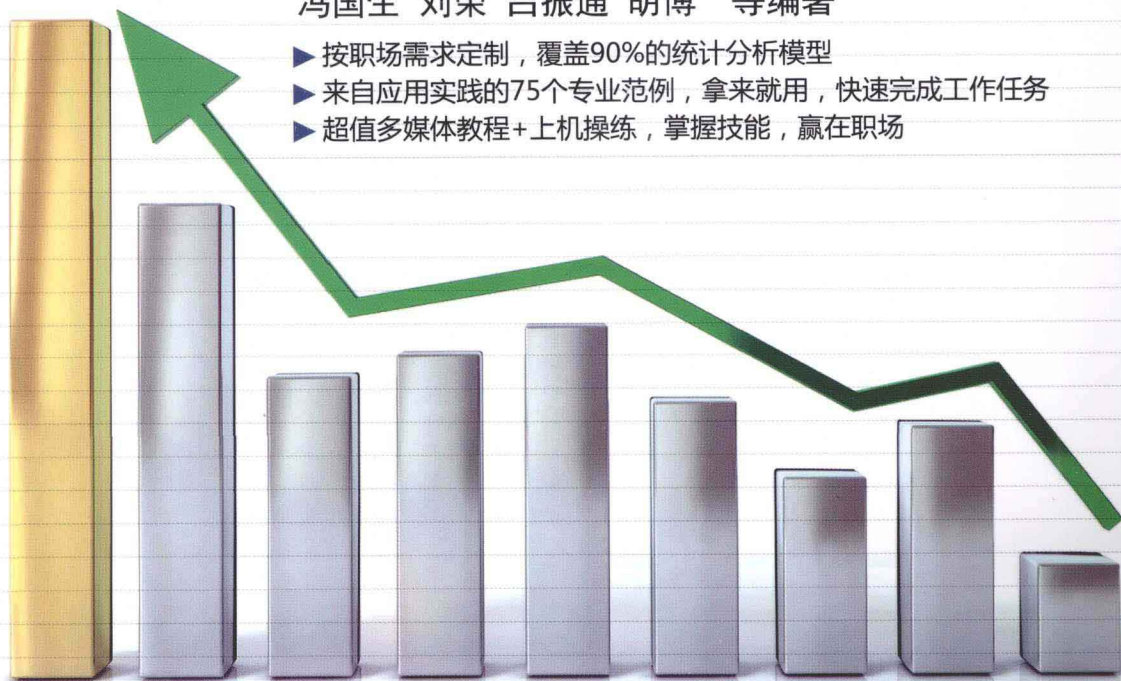
SPSS

统计分析与应用

 **920** 分钟多媒体全程实录
技术讲解+课程实验+上机操练

冯国生 刘荣 吕振通 胡博 等编著

- ▶ 按职场需求定制，覆盖90%的统计分析模型
- ▶ 来自应用实践的75个专业范例，拿来就用，快速完成工作任务
- ▶ 超值多媒体教程+上机操练，掌握技能，赢在职场



机械工业出版社
China Machine Press

SPSS

统计分析与应用

第2版

冯国生 刘荣 吕振通 胡博 等编著



机械工业出版社
China Machine Press

图书在版编目(CIP)数据

SPSS统计分析与应用 / 冯国生等编著. -- 2版. -- 北京: 机械工业出版社, 2014.2

ISBN 978-7-111-44900-3

I. ①S… II. ①冯… III. ①统计分析—软件包 IV. ①C819

中国版本图书馆CIP数据核字(2013)第282456号

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问 北京市展达律师事务所

本书精选 75 个专业范例, 覆盖 90% 以上的统计模型, 以实验教程的形式讲解如何以 SPSS 为工具, 解决各种统计分析问题。

全书共 13 章。第 1 章介绍 SPSS 基本操作及其统计分析常用功能; 第 2~13 章通过 36 个实验介绍如何在 SPSS 中高效完成以下统计工作: 描述性统计分析、均值比较与检验、方差分析、相关分析、回归分析、非参数检验、聚类分析、距离分析与判别分析、因子分析与主成分分析、生存分析以及信度分析等统计分析过程。

对于每一个实验, 都从“原理、目的与要求、内容及数据来源、操作指导、结论”五个方面进行讲解, 同时提供全程语音讲解的多媒体教学文件。章后精选 39 个上机题, 在光盘中提供原始数据文件及多媒体教学动画。

本书重实践兼理论, 涉及自然科学和社会科学的各个领域, 不仅有助于读者理解统计学方法和模型的适用问题, 还为读者提供了一个即查即用的实例工具手册, 适合高等院校相关专业本科生、研究生以及从事统计分析和决策等领域的读者学习参考。

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑: 夏非彼 迟振春

中国电影出版社印刷厂印刷

2014年2月第2版第1次印刷

188mm×260mm·18.5印张

标准书号: ISBN 978-7-111-44900-3

ISBN 978-7-89405-183-7 (光盘)

定 价: 49.00元(附1DVD)

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 82728184

购书热线: (010) 68326294 88379649 68995259

投稿热线: (010) 82728184 88379604

读者信箱: booksaga@126.com

前 言

统计学的起源可以追溯至 18 世纪甚至更早，但是统计学主要的发展是在 19 世纪末和 20 世纪初才真正开始，然后到了 40 年代才逐渐成熟。统计学的萌芽产生在欧洲，17 世纪中叶至 18 世纪中叶是统计学的创立时期。18 世纪末至 19 世纪末是统计学的发展时期，20 世纪初以来，科学技术迅猛发展，社会发生了巨大变化，统计学进入了快速发展时期。

一般而言，我们可以把统计问题分成两类：描述统计和推断统计。简单地说：任何对数据（即样本）的处理导致预测或推论总体的统计称为推断统计。反之，如果我们的兴趣只限于手头现有的数据，而不准备把结果用来推断整体则称为描述统计。

从理论层面上说，统计学的理论基础是概率论和数理统计。统计学运用数学工具，记录数据产生的过程，描述概率分布，进行推定，作假定检验，形成了一个比较完整的理论体系。在应用层面，随着社会经济的发展，要求统计学提供更多的统计方法，也要求统计学能提供更有效的调查整理、分析资料的方法，这就促使了社会科学统计的大发展，现在很多社会科学统计的方法比如问卷调查等已经成为社会科学研究的重要手段。

在科学技术飞速发展的今天，统计学广泛吸收和融合相关学科的新理论，不断开发应用新技术和新方法，深化和丰富了统计学传统领域的理论与方法，并拓展了新的领域，今天的统计学已展现出更加强有力的生命力。

SPSS 是最为优秀的统计软件之一，深受各行业用户的青睐。为满足广大读者学习和掌握统计分析方法的需求，本书以 SPSS 20.0 为准，以实例为主线，以实验操作步骤和结果解释为主要内容，辅以简略明了的理论阐释，详细介绍了各种广泛应用、经典或现代的统计学模型和分析方法。

全书共分 13 章，第 1 章是关于 SPSS 的概述，介绍了 SPSS 软件操作的基本知识以及数据文件的建立和编辑以及如何使用帮助等内容。第 2 章介绍了运用 SPSS 进行描述性统计的方法和步骤，这是进行更高级的统计分析所必需的第一步。第 3 章是关于均值比较分析的内容，这是判断样本是否来自同一总体的经典方法。第 4 章介绍了相关与回归分析，这是经典统计学分析的主要手段之一，特别是回归分析几乎成为应用统计学的标准范式。第 5 章介绍了列联表分析和对数线性模型，这两者都是处理定性数据的标准工具。第 6 章是关于方差分析的内容，这是处理变量间关系问题的经典方法。第 7 章是关于因子分析和主成分分析，两者联系密切，但又有不同，都是对高维复杂数据进行降维简化的重要工具。第 8 章介绍了聚类分析的主要方法，这其实蕴含了一种从数据角度看问题的数据挖掘思想，也就是单纯从数据本身特征中提取相似性，然后将大量样本聚集成有限的便于分



析的某几类,这是现代多元统计分析的主要工具之一。第9章是关于判别分析和距离分析的,判别分析作为一种样本分类方法具有十分重要的应用,而距离分析则更是多元统计分析的核心,因为多元统计分析中根据距离概念的不同定义形成了多种截然不同的分析方法。第10章介绍信度分析和尺度分析,两者主要应用在对统计调查问卷的评估方面,是实务操作中很重要的工具。第11章介绍了SPSS时间序列分析的主要内容,时间序列是信号处理、系统分析、生物统计、医学统计、计量经济等领域的主要分析对象,因此这一章介绍的方法在各个领域都有着广泛的应用。第12章是关于生存分析的内容,这是现代统计学中十分重要的一种分析方法,虽然诞生于对生存数据的分析,但已广泛应用于多个领域。最后一章介绍非参数统计分析的内容,非参模型是相对于参数模型来说的,非参数方法具有计算简便、“稳健性”等特点,因此早已成为统计分析的标准范式之一。

本书作者从需要进行数据分析的读者的角度出发,结合自身多年的SPSS使用经验,在详细介绍软件操作的同时,注意将相应的基本统计学知识融入其中。另外,本书中的各章不仅详细介绍了实例的具体操作和操作结果分析,而且还配有一定数量的上机操作练习题供读者学习使用。读者只需按照书中介绍的步骤进行实际操作,相信很快就能完全掌握本书的内容。

本书内容充实、实用,可作为大专院校统计、经济管理、数学、教育、生物、医学、心理学等专业的实用统计分析实验教材,也可作为统计工作者、科技工作者、工程技术人员以及经济管理人员参考用书。

全书由冯国生、刘荣、吕振通、胡博主持编写,参加本书编写工作的还有吕平、王坚宁、高克臻、张云霞、王魁、许小荣、王东、王龙、张银芳、周新国、王松年、张凤琴、陈可汤、陈作聪、聂阳、沈毅、张华杰、彭一明、朱丽云、张秀梅、张玉兰、李爽、田伟、王文婷等。本书在编写过程中吸收了前人的研究成果,在此一并表示感谢。

由于作者水平有限,书中的错误或不当之处在所难免,诚恳地欢迎各位同行专家和广大读者批评指正,并提出宝贵的意见。

编 者

2013年10月

目 录

前言

第 1 章 SPSS 概述	1
1.1 SPSS 的启动与退出	2
1.2 SPSS 窗口介绍	4
1.2.1 数据编辑窗口	4
1.2.2 语法窗口	5
1.2.3 结果输出窗口	7
1.3 SPSS 菜单操作简介	9
1.4 数据文件的建立	11
1.4.1 SPSS 的数据文件	11
1.4.2 变量与观测值	13
1.4.3 在 SPSS 中定义变量	15
1.5 数据的录入和编辑	19
1.5.1 数据的录入	19
1.5.2 数据的编辑	19
1.6 数据文件的操作	22
1.6.1 分类整理	22
1.6.2 数据转置	23
1.6.3 数据合并	24
1.7 变量转换	27
1.7.1 变量计算	27
1.7.2 生成新的时间序列	29
1.7.3 缺失值的替换	30
1.8 统计分析报告	32
1.8.1 在线分析报告	32
1.8.2 观测值概述	34
1.8.3 按行概述观测值	36
1.8.4 按列概述观测值	41
1.9 统计分析功能概述	43
1.10 SPSS 帮助系统	43
1.10.1 帮助菜单的帮助系统	43



1.10.2 右键帮助系统	45
1.11 上机练习	46
Exercise 1-1 定义变量练习	46
Exercise 1-2 转置练习	46
Exercise 1-3 概述报告练习	46
第 2 章 描述性统计分析	47
实验 2-1 频数分析	47
实验基本原理	47
实验目的与要求	47
实验内容及数据来源	48
实验操作指导	48
实验结论	50
实验 2-2 描述统计量	51
实验基本原理	51
实验目的与要求	51
实验内容及数据来源	51
实验操作指导	52
实验结论	53
实验 2-3 数据探索	53
实验基本原理	53
实验目的与要求	54
实验内容及数据来源	54
实验操作指导	54
实验结论	57
实验 2-4 比率统计分析	61
实验基本原理	61
实验目的与要求	61
实验内容及数据来源	61
实验操作指导	61
实验结论	64
上机练习	64
Exercise 2-1 血清胆固醇频数分布分析	64
Exercise 2-2 血清胆固醇常规统计分析	64
Exercise 2-3 心血管病分组变量探索性分析	65
Exercise 2-4 磁疗效果显著性差异分析	65
Exercise 2-5 检验产品纯度	65



第 3 章 均值比较分析	66
实验 3-1 单样本 T 检验	66
实验基本原理	66
实验目的与要求	67
实验内容及数据来源	67
实验操作指导	67
实验结论	68
实验 3-2 独立样本 T 检验	69
实验基本原理	69
实验目的与要求	69
实验内容及数据来源	70
实验操作指导	70
实验结论	71
实验 3-3 配对样本 T 检验	72
实验基本原理	72
实验目的与要求	72
实验内容及数据来源	72
实验操作指导	73
实验结论	73
上机练习	74
Exercise 3-1 检验产品是否符合质量要求	74
Exercise 3-2 检验两台仪器的测量结果有无显著差异	75
Exercise 3-3 检验两种轮胎耐磨性的差异	75
第 4 章 相关分析与回归分析	76
实验 4-1 两变量相关分析	76
实验基本原理	76
实验目的与要求	77
实验内容及数据来源	77
实验操作指导	77
实验结论	79
实验 4-2 偏相关分析	81
实验基本原理	81
实验目的与要求	81
实验内容及数据来源	81
实验操作指导	82
实验结论	83
实验 4-3 线性回归分析	84



实验基本原理	84
实验目的与要求	85
实验内容及数据来源	85
实验操作指导	86
实验结论	92
实验 4-4 非线性回归分析	96
实验基本原理	96
实验目的与要求	97
实验内容及数据来源	97
实验操作指导	97
实验结论	102
实验 4-5 Logistic 回归分析	104
实验基本原理	104
实验目的与要求	104
实验内容及数据来源	105
实验操作指导	105
实验结论	109
上机练习	112
Exercise 4-1 产妇与婴儿体重相关分析	112
Exercise 4-2 高血压病因线性回归分析	113
Exercise 4-3 预测研究所的净收益	113
Exercise 4-4 癌细胞转移的 Logistic 回归分析	113
第 5 章 列联表分析与对数线性模型	115
实验 5-1 列联表分析	115
实验基本原理	115
实验目的与要求	116
实验内容及数据来源	116
实验操作指导	116
实验结论	120
实验 5-2 对数线性模型	123
实验基本原理	123
实验目的与要求	124
实验内容及数据来源	124
实验操作指导	124
实验结论	128
上机练习	132
Exercise 5-1 病毒抗体反应情况的差异检验	132



Exercise 5-2 冠心病病因对数线性分析	132
第6章 方差分析	133
实验 6-1 单因素方差分析	133
实验基本原理	133
实验目的与要求	134
实验内容及数据来源	134
实验操作指导	134
实验结论	139
实验 6-2 单因变量多因素方差分析	142
实验基本原理	142
实验目的与要求	143
实验内容及数据来源	143
实验操作指导	143
实验结论	151
实验 6-3 多变量方差分析	153
实验基本原理	153
实验目的与要求	153
实验内容及数据来源	153
实验操作指导	154
实验结论	157
上机练习	160
Exercise 6-1 检验各行业的服务质量差异	160
Exercise 6-2 对不同工厂的同型号电池质量进行评估	160
Exercise 6-3 三种麻醉方法的方差分析	161
第7章 因子分析和主成分分析	162
实验 7-1 因子分析	162
实验基本原理	162
实验目的与要求	163
实验内容及数据来源	163
实验操作指导	164
实验结论	169
实验 7-2 主成分分析	172
实验基本原理	172
实验目的与要求	173
实验内容及数据来源	173
实验操作指导	173
实验结论	176



上机练习	177
Exercise 7-1 用因子分析法研究产量指标数据	177
Exercise 7-2 对体检指标进行主成分分析	177
第8章 聚类分析	178
实验 8-1 K-均值聚类	178
实验基本原理	178
实验目的与要求	179
实验内容及数据来源	179
实验操作指导	179
实验结论	182
实验 8-2 分层聚类	184
实验基本原理	184
实验目的与要求	184
实验内容及数据来源	184
实验操作指导	185
实验结论	189
上机练习	191
Exercise 8-1 用聚类分析法分类工厂周围大气污染区域	191
Exercise 8-2 分析不同国家的人口出生、死亡数据	192
Exercise 8-3 对各省学生的体质进行评估	192
第9章 判别分析和距离分析	194
实验 9-1 判别分析	194
实验基本原理	194
实验目的与要求	195
实验内容及数据来源	195
实验操作指导	195
实验结论	200
实验 9-2 距离分析	204
实验基本原理	204
实验目的与要求	205
实验内容及数据来源	205
实验操作指导	206
实验结论	210
上机练习	211
Exercise 9-1 运动员等级的判别分析	211
Exercise 9-2 地区收入水平的距离分析	212



第 10 章 信度分析和尺度分析	213
实验 10-1 信度分析	213
实验基本原理	213
实验目的与要求	214
实验内容及数据来源	214
实验操作指导	215
实验结论	218
实验 10-2 尺度分析	218
实验基本原理	218
实验目的与要求	218
实验内容及数据来源	218
实验操作指导	219
实验结论	223
上机练习	224
Exercise 10-1 产品评价的尺度分析	224
Exercise 10-2 对问卷调查结果进行信度分析	225
第 11 章 时间序列分析	226
实验 11-1 指数平滑模型	226
实验基本原理	226
实验目的与要求	227
实验内容及数据来源	227
实验操作指导	227
实验 11-2 自回归集成移动平均模型	231
实验基本原理	231
实验目的与要求	231
实验内容及数据来源	232
实验操作指导	232
实验结论	234
实验 11-3 季节分解方法	235
实验基本原理	235
实验目的与要求	235
实验内容及数据来源	235
实验操作指导	236
实验结论	238
上机练习	238
Exercise 11-1 Winters 线性平滑方法应用	238
Exercise 11-2 ARIMA 模型应用	239



Exercise 11-3 对季度数据进行分析	239
第 12 章 生存分析	240
实验 12-1 生命表分析	240
实验基本原理	240
实验目的与要求	241
实验内容及数据来源	241
实验操作指导	242
实验结论	245
实验 12-2 Kaplan-Meier 分析	246
实验基本原理	246
实验目的与要求	246
实验内容及数据来源	246
实验操作指导	247
实验结论	250
实验 12-3 风险比例模型	252
实验基本原理	252
实验目的与要求	253
实验内容及数据来源	253
实验操作指导	253
实验结论	259
上机练习	261
Exercise 12-1 用生命表方法计算生产率	261
Exercise 12-2 Kaplan-Meier 方法分析新药的治疗效果	261
Exercise 12-3 Cox 回归方法分析服务生命长度	262
第 13 章 非参数检验方法	263
实验 13-1 单样本的 Kolmogorov-Smirnov 检验	263
实验基本原理	263
实验目的与要求	264
实验内容及数据来源	264
实验操作指导	264
实验结论	265
实验 13-2 两个独立样本的检验	266
实验基本原理	266
实验目的与要求	267
实验内容及数据来源	267
实验操作指导	268
实验结论	269



实验 13-3 多个独立样本的检验	270
实验基本原理	270
实验目的与要求	271
实验内容及数据来源	271
实验操作指导	272
实验结论	274
实验 13-4 两个相关样本的检验	275
实验基本原理	275
实验目的与要求	275
实验内容及数据来源	275
实验操作指导	276
实验结论	277
实验 13-5 多个相关样本的检验	278
实验基本原理	278
实验目的与要求	278
实验内容及数据来源	278
实验操作指导	279
实验结论	280
上机练习	281
Exercise 13-1 检验两种材料的硬度有无显著差异	281
Exercise 13-2 检验不同操作方法对产品检验的影响	281
Exercise 13-3 用 McNemar 方法检验商场促销活动效果	282
Exercise 13-4 用 Friedman 方法检验计划受偏好程度	282

第 1 章 SPSS 概述

SPSS 是 Statistical Package for the Social Science (社会科学统计软件包) 的简称, 是一种集成化的计算机数据处理应用软件。1968 年, 美国斯坦福大学 H. Nie 等三位大学生开发了最早的 SPSS 统计软件, 并于 1975 年在芝加哥成立了 SPSS 公司。迄今为止, SPSS 已有 30 余年的成长历史, 全球约有 25 万家产品用户, 广泛分布于通信、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。1994~1998 年间, SPSS 公司陆续购并了 SYSTAT、BMDP 等公司, 由原来单一的统计产品开发转向为企业、教育科研及政府机构提供全面的信息统计决策支持服务。伴随 SPSS 服务领域的扩大和深度的增加, 2000 年 SPSS 公司决定将其全称更改为 Statistical Product and Service solutions (统计产品与服务解决方案)。IBM 公司于 2009 年 7 月 28 日宣布用 12 亿美元收购 SPSS。如今 SPSS 已升级至版本 19.0, 而且更名为 PASW Statistics。

目前, 世界上最著名的数据分析软件是 SAS 和 SPSS。SAS 是为专业统计分析人员设计的, 功能强大, 灵活多样, 为专业人士所喜爱。而 SPSS 是为广大的非专业人士设计的, 操作简便, 好学易懂, 简单实用, 因而很受非专业人士的青睐。此外, 与 SAS 软件相比, SPSS 主要面向社会科学研究领域, 因而更广泛应用于教育科学研究, 是国外教育科研人员必备的科研工具。1988 年, 中国高教学会首次推广了这种软件, SPSS 已经成为国内教育科研人员最常用的工具。SPSS 软件具有以下特点:

- 集数据录入、数据编辑、数据管理、统计分析、报表制作以及图形绘制为一体, 功能非常强大, 可针对整体的大型统计项目提供完善的解决方案。
- 工作界面友好完善、布局合理且操作简便, 大部分统计分析过程可以借助鼠标, 通过菜单命令的选择、对话框的参数设置, 点击功能按钮来完成, 不需要用户记忆大量的操作命令。菜单分类合理, 并且可以灵活编辑菜单以及设置工具栏。
- 具有完善的数据转换接口, 可以方便地与 Windows 的其他应用程序进行数据共享和交换。既可以读取 Excel、FoxPro、Lotus 等电子表格和数据软件产生的数据文件, 也可以读取 ASCII 数据文件。
- 提供强大的程序编辑能力和二次开发能力, 可满足高级用户完成更为复杂的统计分析任务的需要, 具有丰富的内部函数和统计功能。
- 具有强大的统计图表绘制和编辑功能, 尤其是三维统计图绘制功能十分突出, 图形美



观大方，输出报告形式灵活、编辑方便易行。

- 附带丰富的数据资料实例和完善的实用指南，为用户学习掌握软件的使用方法提供更多的方便。软件启动后，用户可直接上网访问 SPSS 公司的主页获得更多的帮助和信息。

本章主要介绍 SPSS 的启动与退出、SPSS 中的窗口及其组成以及 SPSS 如何对数据进行管理等，使读者对 SPSS 软件包有一个初步的认识，方便后面的学习。

1.1 SPSS 的启动与退出

启动 SPSS（本书采用的是 SPSS 20.0）后，将弹出如图 1-1 所示的启动对话框，该对话框提供了选择进入 SPSS 的各种方式。

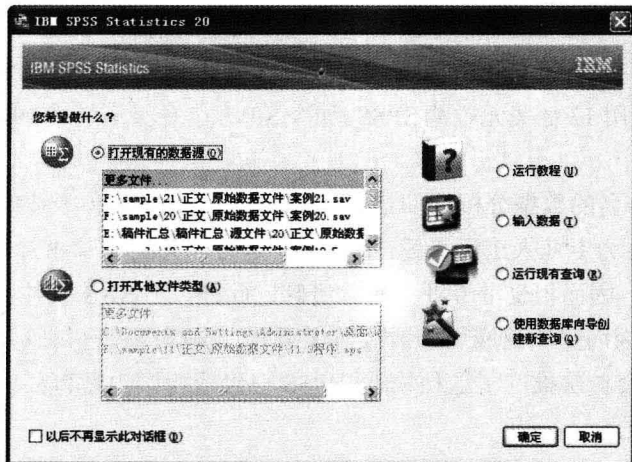


图 1-1 启动对话框

该对话框中包含以下几个选项。

- 运行教程：选择此项将打开操作指导，可以根据主题单击所需要的指导的图标，查看基本操作的指导信息。
- 输入数据：选择此项将显示数据编辑窗口，输入数据，建立新数据集。
- 运行现有查询：选择此项将显示打开文件窗口，让用户选择一个 SPQ 文件。
- 使用数据库向导创建新查询：选择此项将打开数据库处理工具，可将诸如 DBF 格式文件、XLS 格式的 Excel 文件以及 SQL 等数据库文件转换成 SPSS 数据文件。
- 打开现有的数据源：选择此项可让用户在打开的窗口中选择一个 SAV 格式的文件，即 SPSS 数据文件。
- 打开其他文件类型：选择此项可让用户在打开的窗口中选择一个其他格式的文件。

注 意

在启动对话框底部有一个复选框“以后不再显示此对话框”，若勾选该复选框，以后启动 SPSS 时将不再显示该对话框，直接进入 SPSS 数据编辑窗口。

关于 SPSS 的退出，有以下几种方法：

- 选择“文件 | 退出”命令，如图 1-2 所示。



图 1-2 退出 SPSS

- 双击 SPSS 窗口左上角的  图标，或者右键单击标题栏的任何位置，从弹出的快捷菜单中单击“关闭”选项，如图 1-3 所示。

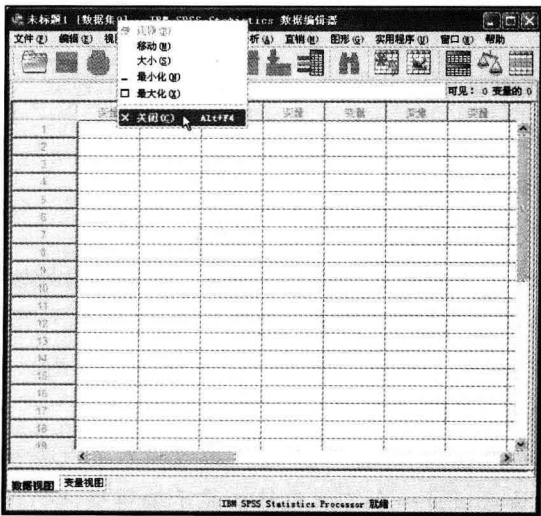


图 1-3 关闭 SPSS