

● 统计数据应用丛书

基于 SPSS

的数据分析

薛薇 编著

 中国人民大学出版社

● 统计数据应用丛书


C819
48

2006

基于 SPSS

的数据分析

薛薇 编著

 中国人民大学出版社

图书在版编目 (CIP) 数据

基于 SPSS 的数据分析/薛薇编著.
北京: 中国人民大学出版社, 2006
(统计数据分析与应用丛书)
ISBN 7-300-07659-9

- I. 基…
- II. 薛…
- III. 统计分析-软件包, SPSS
- IV. C819

中国版本图书馆 CIP 数据核字 (2006) 第 115529 号

统计数据分析与应用丛书

基于 SPSS 的数据分析

薛 薇 编著

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511398 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	北京鑫丰华彩印有限公司		
规 格	170 mm×228 mm 16 开本	版 次	2006 年 10 月第 1 版
印 张	25.25 插页 1	印 次	2006 年 10 月第 1 次印刷
字 数	407 000	定 价	35.00 元

版权所有 侵权必究 印装差错 负责调换

总序

Prologue

《管子》中有一句话：不明于计数，而欲举大事，犹无舟楫而欲经于水险也。意思是说在不清楚具体数据的情况下想做大事，就如同没有桨的船只航行在激流险滩之中。

对于国家政府来说，所谓大事就是引导社会经济沿着正确的方向健康和谐地发展；对于企事业单位来说，所谓大事就是在经营管理中做出科学有效的决策从而在激烈的市场竞争中生存发展；对于科学研究来说，所谓大事就是透过事物外在的表象深入探索其内在的规律性；对于个人生活来说，所谓大事就是在人生成长和个人理财的重要关口做出明智合理的抉择。

这些都离不开数据，离不开数据分析，离不开统计应用。通过数据处理进行科学的定量分析是成大事者的基本方法和首要条件。

当前，中国面临着国际化大背景下转轨的经济体制和转型的社会环境的综合考验，瞬息万变的信息时代对国家、企事业单位和个人都提出了严峻的挑战，而作为“信息密集”的统计应用也将迎来快速发展的机遇期。

这是我们撰写这套“统计数据分析与应用丛书”的初衷，希望它能够从事统计应用的实际工作者提供及时有效的帮助，也能够为有志于进行数据分析的在校本科生和研究生打开一扇统计应用的大门。



《论语》中有一句话：工欲善其事，必先利其器。意思是说做事情要取得较好的成就，应当首先利用先进的工具或手段。

突飞猛进的信息技术已经将统计应用引领到一个崭新的水平，并渗透到统计数据处理各个环节，同时深刻影响着统计工作的全貌。计算机化和网络化是统计应用的必然趋势，它使得统计应用摆脱了复杂公式和计算的羁绊，可以为更多人所理解和使用，所以毫不夸张地说，现代信息技术是解决统计实际问题、掌握统计分析算法、建立统计应用系统的必要条件。

这是我们撰写这套丛书的另外一个初衷，就是让更多的统计应用读者能够使用计算机等信息技术实现统计数据分析与应用的目标，同时也可以让更多的计算机应用读者能够了解一些统计应用的特征和方法。同时，近年来在西方发达国家，信息技术人才和统计应用人才一直排名在就业需求榜的前列，我们认为具备统计知识和计算机知识的复合型应用人才在未来将具有巨大的发展前景和明显的从业优势。

在这套丛书的编写过程中，我们注重引进当前统计应用和计算机应用的前沿技术和理论方法，结合在统计应用、科研与教学中的丰富案例和实际经验，着力突出以下特点：

1. 针对性。针对不同的读者群，由浅入深地展开统计应用的论述，读者也可以按照《基于 EXCEL 的统计应用》、《基于 SPSS 的数据分析》和《基于信息技术的统计信息系统》所构成的应用体系，不断提高统计应用能力和统计理论水平。

2. 可操作性。以计算机和网络等信息技术为基础，实现统计应用的各种需求，读者可以按照丛书指引的策略和具体的方法解决自己工作生活中的数据处理问题。

3. 通俗性。以案例说明原理，以应用解释算法，以发展脉络说明理论形成，努力体现深入浅出的结构安排和文字风格。

不断改革前进是我们这个时代的主旋律，从目前国内相关论著和教材情况看，丛书的上述特点也是许多统计应用学者和同仁所追求的共同方向，在此特别感谢中国人民大学统计学院领导的鼎力支持，感谢中国人民大学出版社各位编辑所付出的辛勤劳动。

统计作为数据处理的方法论，具有广泛的应用领域，而它根本的生命力也在于应用。我们很想为相关读者奉献一套具有一定理论高度，且具有一定



指导性和实战性的统计应用书籍，它应该以统计应用案例为主线，以计算机技术为实现工具，可以使普通经营管理人员、基层科研人员、高层决策者、一般数据处理工作者和高年级的高校学生从中获益。

薛薇

2006年10月

于中国人民大学统计学院

前言

Preface

随着市场经济的发展和竞争的日益激烈，如何运用科学的分析方法，对收集到的数据作出准确、及时的分析并制定正确的决策，已经成为企业极为关注的问题。统计学作为收集、加工和分析数据的方法科学，在企业决策中无疑起着重要作用。同时，随着现代计算机技术的不断发展和成熟，政府和企业的统计应用已逐渐呈现出数据量和数据关系复杂性快速膨胀，数据分析方法日益丰富，实施操作性强等特点。因此，统计应用软件已经成为统计应用的首要工具，而作为国内外普遍流行的 SPSS 统计分析软件在其中则扮演着不可或缺的重要角色。

本书力求从数据分析的应用需求出发，通过对实际问题 and 统计分析方法的剖析，提出解决相关问题的统计分析方案。同时，配合 SPSS 软件操作使用的详解，提出解决相关问题的具体操作步骤，从而使读者能够由浅至深地提出数据分析需求，理解常用统计分析方法的思想，并通过 SPSS 软件实现数据分析需求。

本书力求以最通俗的方式对统计分析方法的核心思想、适用范围进行讲解，避免大量的数学公式和数学证明。目的是使读者能够快速而轻松地了解方法的本质，正确运用方法；同时，方法的介绍紧紧围绕 SPSS 的输出结果开展。目的是希望读者真正理解 SPSS 为什么要输出



这些统计量及它们对于分析结论的重要性，进而正确、合理和完整地引用分析结果。

本书力求从普通数据分析人员的需求出发，对 SPSS 的操作过程给出了较为详尽的说明。但软件的使用说明并非对菜单功能的简单罗列，而是以统计分析的实践过程为主线，从 SPSS 基础和数据管理功能的说明入手，按照针对不同数据类型、分析需求从浅至深、分析方法从易到难的思路展开。这样，能使读者在较短的时间内掌握 SPSS 的核心功能和分析方法，并运用于实际工作中。随着对软件使用的不断娴熟和分析需求的不断深入，读者可继续学习和应用 SPSS 的一些高级分析方法，逐步掌握和应用 SPSS。

本书适合于从事数据分析的社会各领域相关专业的读者，尤其适合从事社会科学研究、经济管理、商业决策、人文教育等行业的中高层管理人员。同时，也能够作为高等院校财经类专业研究生和本科生掌握 SPSS 统计分析方法和软件使用的教材。

请到人大经管图书在线 <http://www.rdjg.com.cn> 中的“课件下载”栏下载本套丛书的案例数据，也可到教育部应用统计科学研究中心统计数据库研究室 www.t1soft.com 中的“服务直通车”栏下载。本书疏漏之处，恳请读者斧正。

编者

2006 年 7 月

CONTENTS 目录

第 1 章 SPSS 统计分析软件概述	(1)
1.1 SPSS 使用基础	(3)
1.1.1 SPSS 软件的安装和启动	(3)
1.1.2 SPSS 的基本窗口	(4)
1.1.3 SPSS 软件的退出	(9)
1.2 SPSS 的基本运行方式	(10)
1.2.1 完全窗口菜单方式	(10)
1.2.2 程序运行方式	(12)
1.2.3 混合运行方式	(15)
1.3 利用 SPSS 进行数据分析的基本步骤	(15)
1.3.1 数据分析的一般步骤	(15)
1.3.2 利用 SPSS 进行数据分析的一般步骤	(17)
1.4 利用统计教练快速入门 SPSS	(18)
第 2 章 SPSS 数据文件的建立和管理	(21)
2.1 SPSS 数据文件	(21)
2.1.1 SPSS 数据文件的特点	(21)
2.1.2 SPSS 数据的基本组织方式	(22)
2.2 SPSS 数据的结构和定义方法	(25)
2.2.1 变量名	(26)
2.2.2 数据类型、列宽、小数位宽	(27)
2.2.3 变量名标签	(29)
2.2.4 变量值标签	(30)
2.2.5 缺失数据	(31)



2.2.6	计量尺度	(32)
2.2.7	SPSS 结构定义操作中应注意的问题	(33)
2.3	SPSS 数据的录入与编辑	(34)
2.3.1	SPSS 数据的录入	(34)
2.3.2	SPSS 数据的编辑	(36)
2.4	SPSS 数据的保存	(39)
2.4.1	SPSS 支持的数据格式	(39)
2.4.2	保存 SPSS 数据的基本操作	(40)
2.5	读取其他格式的数据文件	(41)
2.5.1	直接读入其他格式的数据文件	(42)
2.5.2	使用文本导向读入文本文件	(43)
2.5.3	使用数据库导向读入数据	(46)
2.6	SPSS 数据文件合并	(49)
2.6.1	纵向合并数据文件	(50)
2.6.2	横向合并数据文件	(54)
第 3 章	SPSS 数据的预处理	(59)
3.1	数据的排序	(60)
3.1.1	数据排序的目的	(60)
3.1.2	数据排序的应用举例	(61)
3.2	变量计算	(63)
3.2.1	变量计算的目的	(63)
3.2.2	SPSS 算术表达式	(64)
3.2.3	SPSS 条件表达式	(64)
3.2.4	SPSS 函数	(65)
3.2.5	变量计算的应用举例	(71)
3.3	数据选取	(73)
3.3.1	数据选取的目的	(73)
3.3.2	数据选取方法	(74)
3.3.3	数据选取的应用举例	(75)
3.4	计数	(77)
3.4.1	计数目的	(77)

3.4.2	计数区间	(77)
3.4.3	计数的应用举例	(78)
3.5	分类汇总	(80)
3.5.1	分类汇总的目的	(80)
3.5.2	分类汇总的应用举例	(81)
3.6	数据分组	(83)
3.6.1	数据分组的目的	(83)
3.6.2	组距分组	(83)
3.6.3	数据分组的应用举例	(84)
3.7	数据预处理的其他功能	(86)
3.7.1	数据转置	(87)
3.7.2	加权处理	(88)
3.7.3	数据拆分	(90)
3.7.4	SPSS 变量集	(91)
第 4 章	SPSS 基本统计分析	(95)
4.1	频数分析	(96)
4.1.1	频数分析的目的和基本任务	(96)
4.1.2	频数分析的应用举例	(97)
4.1.3	SPSS 频数分析的扩展功能	(101)
4.1.4	频数分析扩展功能的应用举例	(103)
4.2	计算基本描述统计量	(105)
4.2.1	基本描述统计量	(105)
4.2.2	计算基本描述统计量的应用举例	(109)
4.3	交叉分组下的频数分析	(112)
4.3.1	交叉分组下的频数分析的目的和基本任务	(112)
4.3.2	交叉列联表的主要内容	(112)
4.3.3	交叉列联表行列变量间关系的分析	(115)
4.3.4	交叉分组下的频数分析应用举例	(119)
4.4	多选项分析	(124)
4.4.1	多选项分析的目的和思路	(124)
4.4.2	多选项分析的应用举例	(129)



4.5	比率分析	(135)
4.5.1	比率分析的目的和主要指标	(135)
4.5.2	比率分析的应用举例	(136)
第5章	SPSS 的参数检验	(141)
5.1	参数检验概述	(141)
5.1.1	推断统计与参数检验	(141)
5.1.2	假设检验的基本思想	(142)
5.1.3	假设检验的基本步骤	(143)
5.2	单样本 t 检验	(145)
5.2.1	单样本 t 检验的目的	(145)
5.2.2	单样本 t 检验的基本步骤	(146)
5.2.3	单样本 t 检验的应用举例	(147)
5.3	两独立样本 t 检验	(151)
5.3.1	两独立样本 t 检验的目的	(151)
5.3.2	两独立样本 t 检验的基本步骤	(152)
5.3.3	两独立样本 t 检验的应用举例	(154)
5.4	两配对样本 t 检验	(158)
5.4.1	两配对样本 t 检验的目的	(158)
5.4.2	两配对样本 t 检验的基本步骤	(159)
5.4.3	两配对样本 t 检验的应用举例	(160)
第6章	SPSS 的方差分析	(163)
6.1	方差分析概述	(163)
6.2	单因素方差分析	(165)
6.2.1	单因素方差分析的基本思想	(165)
6.2.2	单因素方差分析的数学模型	(167)
6.2.3	单因素方差分析的基本步骤	(168)
6.2.4	单因素方差分析的应用举例	(169)
6.2.5	单因素方差分析的进一步分析	(171)
6.2.6	单因素方差分析应用举例的进一步分析	(175)
6.3	多因素方差分析	(185)
6.3.1	多因素方差分析的基本思想	(185)



6.3.2	多因素方差分析的数学模型	(188)
6.3.3	多因素方差分析的基本步骤	(188)
6.3.4	多因素方差分析的应用举例	(190)
6.3.5	多因素方差分析的进一步分析	(193)
6.3.6	多因素方差分析应用举例的进一步分析	(195)
6.4	协方差分析	(200)
6.4.1	协方差分析的基本思路	(200)
6.4.2	协方差分析的数学模型	(202)
6.4.3	协方差分析的应用举例	(202)
第7章	SPSS 的非参数检验	(211)
7.1	单样本的非参数检验	(212)
7.1.1	总体分布的卡方检验	(212)
7.1.2	二项分布检验	(215)
7.1.3	单样本 K-S 检验	(217)
7.1.4	变量值随机性检验	(221)
7.2	两独立样本的非参数检验	(223)
7.2.1	两独立样本的曼-惠特尼 U 检验	(224)
7.2.2	两独立样本的 K-S 检验	(226)
7.2.3	两独立样本的游程检验	(227)
7.2.4	极端反应检验	(228)
7.2.5	两独立样本非参数检验的应用举例	(230)
7.3	多独立样本的非参数检验	(233)
7.3.1	中位数检验	(234)
7.3.2	多独立样本的 Kruskal-Wallis 检验	(236)
7.3.3	多独立样本的 Jonckheere-Terpstra 检验	(238)
7.3.4	多独立样本非参数检验的应用举例	(239)
7.4	两配对样本的非参数检验	(242)
7.4.1	两配对样本的 McNemar 检验	(243)
7.4.2	两配对样本的符号检验	(245)
7.4.3	两配对样本 Wilcoxon 符号秩检验	(246)
7.4.4	两配对样本非参数检验的应用举例	(248)



7.5	多配对样本的非参数检验	(251)
7.5.1	多配对样本的 Friedman 检验	(251)
7.5.2	多配对样本的 Cochran Q 检验	(254)
7.5.3	多配对样本的 Kendall 协同系数检验	(256)
7.5.4	多配对样本非参数检验的应用举例	(258)
第 8 章	SPSS 的相关分析	(261)
8.1	相关分析	(261)
8.2	绘制散点图	(262)
8.2.1	散点图的特点	(262)
8.2.2	散点图的应用举例	(264)
8.3	计算相关系数	(270)
8.3.1	相关系数的特点	(270)
8.3.2	相关系数的种类	(271)
8.3.3	计算相关系数的应用举例	(273)
8.4	偏相关分析	(276)
8.4.1	偏相关分析和偏相关系数	(276)
8.4.2	偏相关分析的应用举例	(278)
第 9 章	SPSS 的线性回归分析	(281)
9.1	回归分析概述	(281)
9.1.1	什么是回归分析	(281)
9.1.2	如何得到回归线	(282)
9.1.3	回归分析的一般步骤	(283)
9.2	线性回归分析和线性回归模型	(284)
9.2.1	一元线性回归模型	(284)
9.2.2	多元线性回归模型	(285)
9.2.3	回归参数的普通最小二乘估计	(286)
9.3	回归方程的统计检验	(287)
9.3.1	回归方程的拟合优度检验	(287)
9.3.2	回归方程的显著性检验	(290)
9.3.3	回归系数的显著性检验	(292)
9.3.4	残差分析	(294)

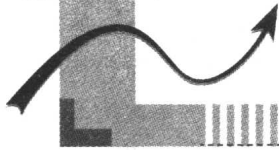


9.4	多元回归分析中的其他问题	(300)
9.4.1	解释变量的筛选问题	(300)
9.4.2	变量的多重共线性问题	(301)
9.5	线性回归分析的基本操作	(303)
9.5.1	线性回归分析的基本操作	(303)
9.5.2	线性回归分析的其他操作	(304)
9.6	线性回归分析的应用举例	(309)
9.6.1	强制进入策略模型	(310)
9.6.2	向后筛选策略模型	(312)
9.7	曲线估计	(318)
9.7.1	曲线估计概述	(318)
9.7.2	曲线估计的应用举例	(320)
第 10 章	SPSS 的聚类分析	(327)
10.1	聚类分析的一般问题	(327)
10.1.1	聚类分析的意义	(327)
10.1.2	聚类分析中“亲疏程度”的度量方法	(329)
10.1.3	聚类分析的几点说明	(334)
10.2	层次聚类	(336)
10.2.1	层次聚类的两种类型和两种方式	(336)
10.2.2	个体与小类、小类与小类间“亲疏程度” 的度量方法	(337)
10.2.3	层次聚类的基本操作	(339)
10.2.4	层次聚类的应用举例	(345)
10.3	K-Means 聚类	(350)
10.3.1	K-Means 聚类分析的核心步骤	(350)
10.3.2	K-Means 聚类分析的应用举例	(352)
第 11 章	SPSS 的因子分析	(359)
11.1	因子分析概述	(359)
11.1.1	因子分析的意义	(359)
11.1.2	因子分析的数学模型和相关概念	(361)
11.2	因子分析的基本内容	(363)



11.2.1	因子分析的基本步骤	(363)
11.2.2	因子分析的前提条件	(364)
11.2.3	因子提取和因子载荷矩阵的求解	(366)
11.2.4	因子的命名	(370)
11.2.5	计算因子得分	(372)
11.3	因子分析的基本操作及案例	(373)
11.3.1	因子分析的基本操作	(373)
11.3.2	因子分析的应用举例	(377)

第 1 章



SPSS统计分析软件概述

SPSS 是 Statistical Package for the Social Science 的英文缩写，意思是社会科学统计软件包。

SPSS 也是 Statistical Product and Service Solutions 的英文缩写，意思是统计产品与服务解决方案。

社会科学统计软件包 (SPSS) 是世界著名的统计分析软件之一。20 世纪 60 年代末，美国斯坦福大学的三位研究生研制开发了最早的统计分析软件 SPSS，并于 1975 年在芝加哥成立了专门研发和经营 SPSS 软件的 SPSS 公司。此时的 SPSS 软件主要运行在中小型计算机上，统称为 SPSSx 版，面向企事业单位用户。20 世纪 80 年代初微型计算机出现，SPSS 公司以其敏锐的市场洞察力和雄厚的技术实力，于 1984 年推出了运行在 DOS 操作系统上的 SPSS 微机版第一版，随后又相继推出了第二版、第三版等，统称为 SPSS/PC+ 版，并确立了个人用户市场第一的地位。90 年代，随着微机 Windows 图形操作系统的出现和盛行，SPSS 公司又研制出了以 Windows 为运行平台的 SPSS 第五版、第六版。90 年代中后期，为适应用户在 Windows 95 操作系统环境下工作的习惯，并迎合 Internet 的广泛使用，SPSS 第七版至第十四版又相继诞