



北京高等教育精品教材
BEIJING GAODENG JIAOYU JINGPIN JIAOCAI

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

SPSS

统计分析方法及应用 (第3版)

薛薇 编著

SPSS
统计分析教材



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

含光盘



北京高等教育精品教材

BEIJING GAODENG JIAOYU JINGPIN JIAOCAI

SPSS

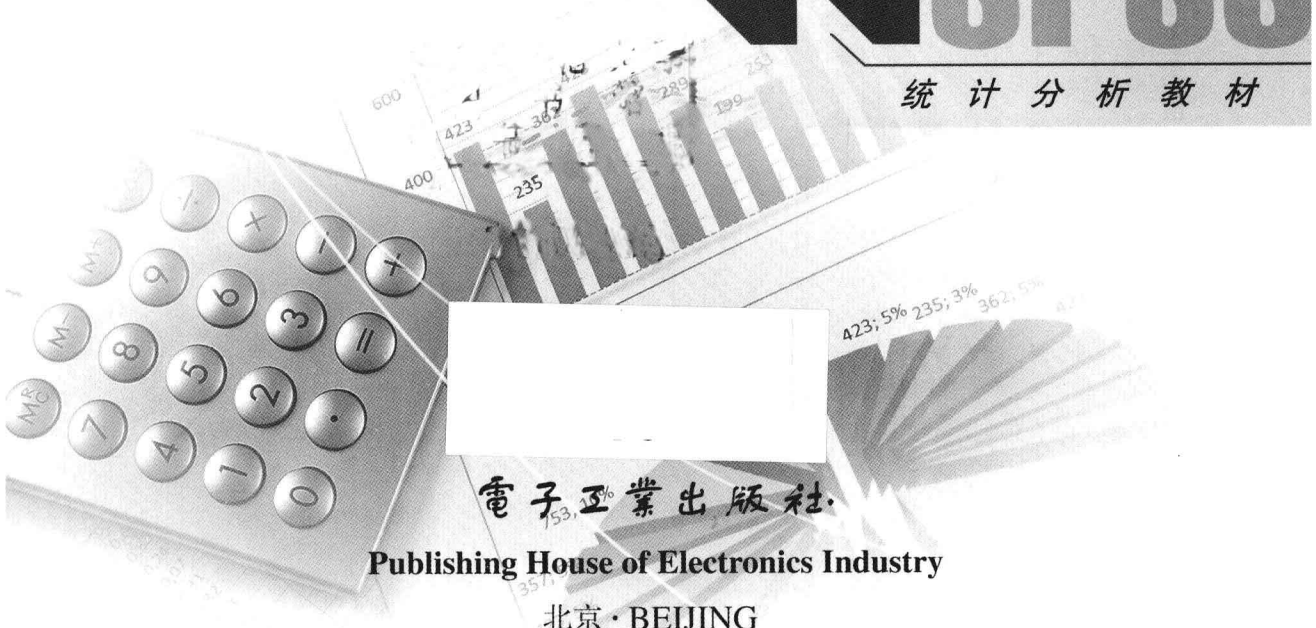
统计分析方法及应用

(第3版)

◎ 薛薇 编著



统计分析教材



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书是北京市高等教育精品教材。全书以统计分析的应用需求为主线,以通俗易懂的语言对 SPSS 中主要统计分析方法的核心思想进行系统的介绍,并对其在 SPSS 中的操作实现步骤进行详尽说明,同时配合应用案例分析,使读者能够较快领会方法的要点,掌握方法的实现操作,明确方法的适用特点。本书克服了 SPSS 手册类教材只注重操作说明而忽略原理讲解的不足,同时弥补了统计专业教材只注重原理述论而缺乏实现工具的缺憾,是一本特色鲜明、具有广泛使用价值的精品教材。配套光盘中为书中相关章节的示例数据文件,供读者练习时参考使用。

本书可作为高等院校财经类专业本科生和研究生学习 SPSS 统计分析方法和软件使用的教材,也适合于从事统计分析和决策的社会各领域读者学习使用,尤其适合于从事社会科学研究、经济管理、商业决策、人文教育、金融保险等行业的中高层管理人员学习使用。

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 统计分析方法及应用/薛薇编著. —3 版. —北京:电子工业出版社,2013.1
统计分析教材
ISBN 978-7-121-18949-4

I. ①S… II. ①薛… III. 统计分析-软件包-高等学校-教材 IV. C819

中国版本图书馆 CIP 数据核字(2012)第 271081 号

策划编辑:秦淑灵

责任编辑:秦淑灵

印 刷:三河市鑫金马印装有限公司

装 订:三河市鑫金马印装有限公司

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:787×1092 1/16 印张:24.75 字数:640 千字 黑插:1

印 次:2013 年 1 月第 1 次印刷

印 数:4000 册 定价:49.00 元 (含光盘 1 张)

凡所购买电子工业出版社图书有缺损问题的,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 zlt@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

前 言

《SPSS 统计分析方法及应用》出版以来,得到了广大读者的支持和喜爱。在第 3 版中,我们将仍然保持前版的写作风格,努力以深入浅出的方式,有条理地全面介绍统计分析方法的核心原理,并以实际数据为纽带说明 SPSS 的操作过程,同时以应用案例为背景剖析统计分析的应用解决方案,光盘中提供书中相关章节的示例数据文件,供读者练习时参考使用。

随着 SPSS 软件版本的不断升级,前版对软件操作部分的讲解内容已显得陈旧。在听取读者各方面意见的基础上,我们对前版进行了修订。修订后的《SPSS 统计分析方法及应用(第 3 版)》(以下简称《第 3 版》),对以下方面进行了调整和增补:

第一,以 SPSS 18.0 中文版为主要讲解对象。在操作方面,SPSS 的最新版本与我们所选用的版本之间无大差异。

第二,为适应软件新版本的需要,重新更换了前版中的所有操作窗口截图,并调整了部分应用案例。

第三,对章节安排进行了重新调整,使内容更具逻辑性和连贯性。

第四,增加了多项 Logistic 回归分析。随着回归分析方法应用面的不断拓展,如何处理和分析多分类型被解释变量,已成为极为普遍的分析需求。为此,《第 3 版》增加了相关内容。

第五,增加了多分类有序回归分析。虽然该部分属于广义线性模型的范畴,相对较难,但出于应用以及对 Logistic 回归内容延伸的考虑,《第 3 版》增加了相关章节。

第六,增加了判别分析。判别分析作为经典的多元统计分析方法,由于其对分类型变量的分类预测有独到之处,因此《第 3 版》增加了相关内容。

第七,对前版中时间序列分析的应用案例部分进行了必要精简。

统计分析需求日益旺盛已是一个不争事实。在此背景下,掌握权威统计分析软件 SPSS,是高端管理人才必备的职业技能。真诚希望读者能够通过阅读本书,逐步领会统计分析方法的精髓,掌握 SPSS 软件的操作,举一反三灵活应用于统计分析的实践中。

本书不妥和错误之处,敬请各位读者指正。

编 者

于中国人民大学统计学院

第2版前言

一、统计应用的特点

在计算机技术迅猛发展的今天,政府和企业的统计应用已逐渐呈现以下特征:

第一,数据量和数据关系复杂性快速膨胀

数据量和数据关系复杂性快速膨胀是目前统计应用中遇到的主要问题。通常,一般规模的统计分析项目中,数据量可达到百万条甚至近千万条。统计应用也不再以统计年鉴和统计报表的纸面方式获得数据,而要求直接从电子化多媒体化的基础业务系统、管理信息系统和决策支持系统的数据库(数据仓库)中提取数据。利用计算机和网络进行有效的数据组织和数据管理已经成为统计应用的基本环节和基础方式。

第二,数据分析方法日益丰富

在数据处理和数据分析中,以坚实的数学理论为基础的统计分析方法已获得广泛应用。近年来以计算机技术为动力的数据挖掘技术也为数据分析增添了新的方法和思路。机器学习、神经网络和生物遗传算法等已成为处理海量数据,探索数据内在规律性,提取数据中未知知识的重要手段。这些数据分析方法的成果不断地体现在统计分析软件和数据挖掘软件体系中,并通过强大的计算机处理能力和网络分布处理能力进行模型的高速计算。

第三,统计应用需要具有可操作性的整体解决方案

传统的统计业务处理流程是报表上报、人工汇总、撰写分析报告,最终实现为有关管理决策部门提供信息监督和咨询服务的目标。统计应用的科研长期以来也沿用了这种套路,即由统计业务方提供数据,由统计方法专家进行数据分析、撰写分析报告和提供咨询建议。这样的应用方式存在两个主要问题:首先,许多企事业单位统计应用起点较低,基础较薄弱,无法提供完善的报表支持,出现或资料不完整或没有电子化的数据来源等问题,使得统计分析常常成为无米之炊;其次,不定期的统计报告方式难以满足统计业务的长期发展需要。目前,越来越多的统计应用要求研究人员提供能够与统计业务绑定的、可长期运行的、具有可操作性的统计应用整体解决方案。这个整体方案将统计业务处理功能、统计数据、统计分析方法完整地集成起来。因此,如果没有一个符合企事业单位统计业务自身特点的统计信息系统,统计分析方法的应用将很难实现健康的、落地生根的、滚动式的发展。

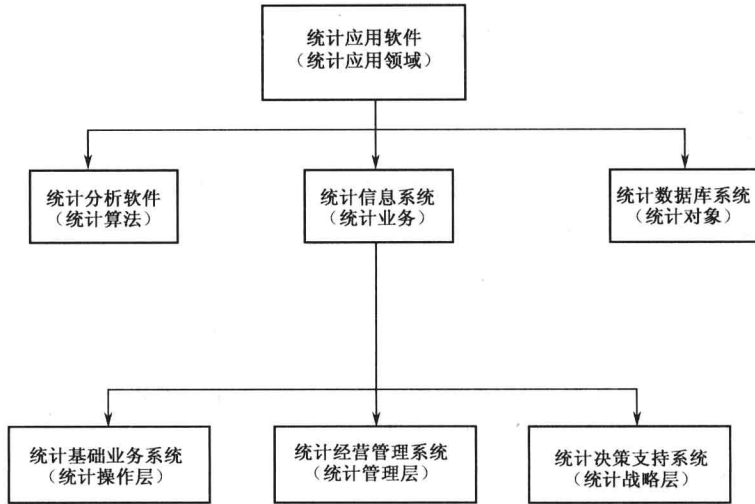
总之,如果说数学是统计方法的首要工具,那么,统计应用软件已经成为统计应用的首要工具。

二、统计应用软件的分类

长期以来,统计无论在国内外政府和企业都有着广泛的应用,但统计应用软件在企事业信息化建设中的认可度却相对较低。产生这个矛盾的一个重要原因就是统计应用软件概念混淆,分类不清。目前,人们在提及统计应用软件时会涉及统计分析软件、统计信息系统、统计数据库系统等方面。规范统计应用软件分类标准对统计应用软件发展和统计信息化建设有着深远的和现实的意义。我们经过对国内外大量资料的分析研究,结合自身统计应用软件开发应用的经验,提出了统计应用软件分类标准,希望能抛砖引玉,引起大家对统计应用软件的关注和研究。

首先,统计应用软件是应用软件的一种分类,是应用软件从应用领域或应用行业划分出的

一个分支。其次,根据应用性质,统计应用软件可划分为统计分析软件、统计信息系统和统计数据库系统三大类。



- 统计分析软件是依据统计分析方法开发的支持数据分析的工具型软件,如 SPSS 和 SAS 等。
- 统计数据库系统是基于统计数据库管理系统而建立的包含统计数据资源的数据库系统。其中,统计数据库管理系统是根据统计处理对象(统计数据)的特征和一般处理功能而研制的数据库管理系统(DBMS)。
- 统计信息系统是采用软件开发平台,结合统计数据库系统和统计分析软件等核心技术开发的服务于统计业务处理和统计数据分析的信息系统。根据统计业务的不同处理层次和服务对象,统计信息系统一般又可分为统计基础业务系统、统计经营管理系统和统计决策支持系统。统计基础业务系统主要用于统计数据的采集处理和统计报表的采集处理,包括统计专项调查与普查处理系统等;统计经营管理系统主要用于统计数据和报表的汇总、查询、传输、基本分析和信息发布等;统计决策支持系统主要用于统计数据综合分析预测和深入的数据挖掘处理等。

因此,本书将重点讨论的 SPSS 软件是一种专业的统计分析软件,用于数据的各种分析,进而最终为企事业的科学决策服务。

三、关于本书

“数据”是科学决策的重要依据。“让数据说话”是科学决策的正确途径,掌握数据分析方法和数据分析软件工具是科学决策的有效手段。本书旨在通过对 SPSS 统计分析方法、软件及应用案例的介绍分析,使读者由浅入深地了解和掌握统计分析方法,快速掌握 SPSS 软件的使用,并灵活运用于数据分析和科学决策中。

- 以统计分析过程为主线介绍 SPSS

目前,关于 SPSS 软件的书籍比较多,也受到广大读者的普遍青睐。作者参考分析了这些书籍,并在多年教学经验和实践的基础上,总结形成了一套独特的 SPSS 软件教学方式。

由于 SPSS 软件是一种专业性较强的统计分析软件,因此作者以为该软件的介绍方法应有别于其他非专业性的、大众化的计算机软件。对于大众化软件,一般可以按照软件中菜单的

顺序来依次讲解；而对 SPSS 这样的专业软件来说，该方法则存在许多不利于快速掌握和应用 SPSS 的问题。

很多已经具备本专业领域知识的读者学习 SPSS 软件过程中遇到的问题是：首先，拿到一批数据后不知道如何用 SPSS 组织它们，不知道如何利用 SPSS 对它们进行基本加工和整理；其次，不知道应从何处入手分析，应采用 SPSS 中的哪些分析方法和功能实现对各类数据的由浅入深的分析，不知道如何理解和合理解释分析结果等。

因此，作者以为，对 SPSS 软件的学习介绍应从实际应用出发，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、分析需求由浅至深、分析方法从易到难的思路展开，而非软件菜单的逐一介绍。这样，能使读者在非常短的时间内掌握 SPSS 的核心功能和分析方法，并可很快运用于实际工作中。另外，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，进而逐步实现对 SPSS 功能的全面掌握和应用。

● 统计分析方法、SPSS 操作及案例分析的有机结合

目前，虽然介绍 SPSS 软件的书籍比较多，但将统计分析方法、SPSS 操作及应用案例分析有机结合的书籍并不十分多见。有些书籍比较侧重对 SPSS 操作的手册性介绍，对数据分析方法讲解很少，给大量有统计分析需求、但又不很了解统计分析方法的读者带来诸多学习上的困难；有些书籍则比较侧重统计分析方法的论述，但却缺乏对 SPSS 操作实现的必要说明，这样又会给方法的实际运用带来障碍。从快速掌握和应用 SPSS 的角度看，作者以为，较为合理的方法是将上述两方面有机结合，并配合大量的多领域的分析案例，使读者一方面可以了解分析方法的核心思想，掌握方法的正确应用范围，不至于滥用和误用方法，另一方面也使读者能够快速熟悉和掌握 SPSS，并在理解分析方法的基础上给分析结果以正确合理的解释。

● 通俗的统计分析方法讲解，详尽的 SPSS 操作过程说明

针对读者群的特点，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，方法的介绍是紧紧围绕 SPSS 的输出结果展开的，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分析结论的重要性，进而会正确、合理和完整地引用分析结果。另外，本书对 SPSS 的操作过程也给出了较为详尽的说明，但并非是对菜单功能的简单罗列，而是将其穿插于分析案例的实现过程中。因此，读者通过案例的学习，能够获得了解方法、掌握 SPSS 操作步骤、合理解释分析结果的多方面收获。

本书适合于从事统计分析和决策的社会各领域各相关专业的读者，尤其适合于从事社会科学、经济管理、商业决策、人文教育、金融保险等行业的中高层管理人员。同时，也可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。书中配套案例数据光盘供读者学习和练习。

由于水平所限，书中的问题与错误，请读者批评指正。

编 者
于中国人民大学统计学院

目 录

第 1 章 SPSS 统计分析软件概述	(1)
1.1 SPSS 的发展及特点	(1)
1.2 SPSS 使用基础	(2)
1.2.1 SPSS 软件的安装和启动	(2)
1.2.2 SPSS 的基本操作环境	(3)
1.2.3 SPSS 软件的退出	(6)
1.2.4 SPSS 软件的三种基本使用方式	(6)
1.3 利用 SPSS 进行数据分析的基本步骤	(9)
1.3.1 数据分析的一般步骤	(10)
1.3.2 利用 SPSS 进行数据分析的一般步骤	(11)
第 2 章 SPSS 数据文件的建立和管理	(12)
2.1 SPSS 数据文件	(12)
2.1.1 SPSS 数据文件的特点	(12)
2.1.2 SPSS 数据的基本组织方式	(12)
2.2 SPSS 数据的结构和定义方法	(13)
2.2.1 变量名(Name)	(14)
2.2.2 数据类型(Type)、宽度(Width)、列宽度(Columns)	(14)
2.2.3 变量名标签(Label)	(15)
2.2.4 变量值标签(Values)	(15)
2.2.5 缺失值(Missing)	(16)
2.2.6 计量标准(Measure)	(17)
2.2.7 结构定义的基本操作	(18)
2.3 SPSS 结构定义的应用案例	(18)
2.4 SPSS 数据的录入与编辑	(20)
2.4.1 SPSS 数据的录入	(20)
2.4.2 SPSS 数据的编辑	(21)
2.5 SPSS 数据的保存	(23)
2.5.1 SPSS 支持的数据格式	(23)
2.5.2 保存 SPSS 数据的基本操作	(23)
2.6 读取其他格式的数据文件	(24)
2.6.1 直接读入其他格式的数据文件	(25)
2.6.2 使用文本向导读入文本文件	(25)
2.6.3 使用数据库向导读入数据	(28)
2.7 SPSS 数据文件合并	(31)
2.7.1 纵向合并数据文件	(31)
2.7.2 横向合并数据文件	(34)

第 3 章	SPSS 数据的预处理	(37)
3.1	数据的排序	(37)
3.1.1	数据排序的目的	(37)
3.1.2	数据排序的基本操作	(38)
3.1.3	数据排序的应用举例	(39)
3.2	变量计算	(39)
3.2.1	变量计算的目的	(39)
3.2.2	SPSS 算术表达式	(40)
3.2.3	SPSS 条件表达式	(40)
3.2.4	SPSS 函数	(41)
3.2.5	变量计算的基本操作	(44)
3.2.6	变量计算的应用举例	(46)
3.3	数据选取	(46)
3.3.1	数据选取的目的	(46)
3.3.2	数据选取	(46)
3.3.3	数据选取的基本操作	(47)
3.3.4	数据抽样的应用举例	(48)
3.4	计数	(48)
3.4.1	计数目的	(48)
3.4.2	计数区间	(49)
3.4.3	计数的基本操作	(49)
3.4.4	计数的应用举例	(50)
3.5	分类汇总	(50)
3.5.1	分类汇总的目的	(50)
3.5.2	分类汇总的基本操作	(51)
3.5.3	分类汇总的应用举例	(52)
3.6	数据分组	(53)
3.6.1	数据分组的目的	(53)
3.6.2	SPSS 的单变量值分组	(53)
3.6.3	SPSS 的组距分组	(55)
3.6.4	SPSS 的分位数分组	(57)
3.7	数据预处理的其他功能	(60)
3.7.1	数据转置	(60)
3.7.2	加权处理	(61)
3.7.3	数据拆分	(61)
第 4 章	SPSS 基本统计分析	(63)
4.1	频数分析	(63)
4.1.1	频数分析的目的和基本任务	(63)
4.1.2	频数分析的基本操作	(64)
4.1.3	SPSS 频数分析的扩展功能	(64)

4.1.4	频数分析的应用举例	(65)
4.2	计算基本描述统计量	(68)
4.2.1	基本描述统计量	(68)
4.2.2	计算基本描述统计量的基本操作	(70)
4.2.3	计算基本描述统计量的应用举例	(71)
4.3	交叉分组下的频数分析	(73)
4.3.1	交叉分组下频数分析的目的和基本任务	(73)
4.3.2	交叉列联表的主要内容	(73)
4.3.3	交叉列联表行列变量间关系的分析	(75)
4.3.4	交叉分组下的频数分析基本操作	(78)
4.3.5	交叉分组下的频数分析应用举例	(79)
4.3.6	SPSS 中列联表分析的其他方法	(82)
4.4	多选项分析	(85)
4.4.1	多选项分析的目的	(85)
4.4.2	多选项分析的基本操作	(88)
4.4.3	多选项分析的应用举例	(90)
4.5	比率分析	(91)
4.5.1	比率分析的目的和主要指标	(91)
4.5.2	比率分析的基本步骤	(92)
4.5.3	比率分析的应用举例	(92)
第 5 章	SPSS 的参数检验	(94)
5.1	参数检验概述	(94)
5.1.1	推断统计与参数检验	(94)
5.1.2	假设检验的基本思想	(94)
5.1.3	假设检验的基本步骤	(95)
5.2	单样本 t 检验	(96)
5.2.1	单样本 t 检验的目的	(96)
5.2.2	单样本 t 检验的基本步骤	(96)
5.2.3	单样本 t 检验的基本操作	(97)
5.2.4	单样本 t 检验的应用举例	(98)
5.3	两独立样本 t 检验	(100)
5.3.1	两独立样本 t 检验的目的	(100)
5.3.2	两独立样本 t 检验的基本步骤	(101)
5.3.3	两独立样本 t 检验的基本操作	(102)
5.3.4	两独立样本 t 检验的应用举例	(103)
5.4	两配对样本 t 检验	(106)
5.4.1	两配对样本 t 检验的目的	(106)
5.4.2	两配对样本 t 检验的基本步骤	(107)
5.4.3	两配对样本 t 检验的基本操作	(108)
5.4.4	两配对样本 t 检验的应用举例	(108)

第 6 章	SPSS 的方差分析	(110)
6.1	方差分析概述	(110)
6.2	单因素方差分析	(111)
6.2.1	单因素方差分析的基本思想	(111)
6.2.2	单因素方差分析的数学模型	(112)
6.2.3	单因素方差分析的基本步骤	(112)
6.2.4	单因素方差分析的基本操作	(113)
6.2.5	单因素方差的应用举例	(113)
6.2.6	单因素方差分析的进一步分析	(114)
6.2.7	单因素方差应用举例的进一步分析	(118)
6.3	多因素方差分析	(124)
6.3.1	多因素方差分析的基本思想	(124)
6.3.2	多因素方差分析的数学模型	(126)
6.3.3	多因素方差分析的基本步骤	(127)
6.3.4	多因素方差分析的基本操作	(128)
6.3.5	多因素方差分析的应用举例	(128)
6.3.6	多因素方差分析的进一步分析	(129)
6.3.7	多因素方差分析应用举例的进一步分析	(133)
6.4	协方差分析	(135)
6.4.1	协方差分析的基本思路	(135)
6.4.2	协方差分析的数学模型	(136)
6.4.3	协方差分析的基本操作	(136)
6.4.4	协方差分析的应用举例	(136)
第 7 章	SPSS 的非参数检验	(141)
7.1	单样本的非参数检验	(141)
7.1.1	总体分布的卡方检验	(141)
7.1.2	二项分布检验	(143)
7.1.3	单样本 K-S 检验	(145)
7.1.4	变量值随机性检验	(147)
7.2	两独立样本的非参数检验	(148)
7.2.1	两独立样本的曼-惠特尼 U 检验(Mann-Whitney U)	(149)
7.2.2	两独立样本的 K-S 检验	(150)
7.2.3	两独立样本的游程检验(Wald-Wolfwitz Runs)	(151)
7.2.4	极端反应检验(Moses Extreme Reactions)	(152)
7.2.5	两独立样本非参数检验的基本操作	(153)
7.2.6	两独立样本非参数检验的应用举例	(154)
7.3	多独立样本的非参数检验	(155)
7.3.1	中位数检验	(156)
7.3.2	多独立样本的 Kruskal-Wallis 检验	(157)
7.3.3	多独立样本的 Jonckheere-Terpstra 检验	(158)

7.3.4	多独立样本非参数检验的基本操作	(159)
7.3.5	多独立样本非参数检验的应用举例	(160)
7.4	两配对样本的非参数检验	(161)
7.4.1	两配对样本的 McNemar 检验	(162)
7.4.2	两配对样本的符号检验	(163)
7.4.3	两配对样本 Wilcoxon 符号秩检验	(164)
7.4.4	两配对样本非参数检验的基本操作	(165)
7.4.5	两配对样本非参数检验的应用举例	(165)
7.5	多配对样本的非参数检验	(167)
7.5.1	多配对样本的 Friedman 检验	(167)
7.5.2	多配对样本的 Cochran Q 检验	(169)
7.5.3	多配对样本的 Kendall 协同系数检验	(171)
7.5.4	多配对样本非参数检验的基本操作	(172)
7.5.5	多配对样本非参数检验的应用举例	(172)
第 8 章	SPSS 的相关分析和线性回归分析	(175)
8.1	相关分析和回归分析概述	(175)
8.2	相关分析	(175)
8.2.1	散点图	(175)
8.2.2	相关系数	(177)
8.2.3	相关分析应用举例	(180)
8.3	偏相关分析	(182)
8.3.1	偏相关分析和偏相关系数	(182)
8.3.2	偏相关分析的基本操作	(183)
8.3.3	偏相关分析的应用举例	(183)
8.4	回归分析	(184)
8.4.1	回归分析概述	(184)
8.4.2	线性回归模型	(185)
8.4.3	回归参数的普通最小二乘估计	(187)
8.4.4	回归方程的统计检验	(187)
8.4.5	多元回归分析中的其他问题	(195)
8.4.6	线性回归分析的基本操作	(197)
8.4.7	线性回归分析的其他操作	(198)
8.4.8	线性回归分析的应用举例	(201)
8.5	曲线估计	(207)
8.5.1	曲线估计概述	(207)
8.5.2	曲线估计的基本操作	(208)
8.5.3	曲线估计的应用举例	(209)
第 9 章	SPSS 的 Logistic 回归分析	(212)
9.1	Logistic 回归分析概述	(212)
9.2	二项 Logistic 回归分析	(213)

9.2.1	二项 Logistic 回归方程	(213)
9.2.2	二项 Logistic 回归方程系数的含义	(215)
9.2.3	二项 Logistic 回归方程的检验	(216)
9.2.4	二项 Logistic 回归分析中的虚拟自变量	(219)
9.3	二项 Logistic 回归分析的应用	(220)
9.3.1	二项 Logistic 回归分析的基本操作	(220)
9.3.2	二项 Logistic 回归分析的其他操作	(222)
9.3.3	二项 Logistic 回归应用示例	(223)
9.4	多项 Logistic 回归分析	(229)
9.4.1	多项 Logistic 回归分析概述	(229)
9.4.2	多项 Logistic 回归分析的基本操作和应用	(230)
9.4.3	多项 Logistic 回归分析的其他操作和应用	(230)
9.5	多项有序回归分析	(235)
9.5.1	多项有序回归分析概述	(235)
9.5.2	多项有序回归分析的基本操作和应用	(236)
第 10 章	SPSS 的聚类分析	(242)
10.1	聚类分析的一般问题	(242)
10.1.1	聚类分析的意义	(242)
10.1.2	聚类分析中“亲疏程度”的度量方法	(243)
10.1.3	聚类分析几点说明	(246)
10.2	层次聚类	(247)
10.2.1	层次聚类的两种类型和两种方式	(247)
10.2.2	个体与小类、小类与小类间“亲疏程度”的度量方法	(248)
10.2.3	层次聚类的基本操作	(250)
10.2.4	层次聚类的应用举例	(254)
10.3	K-Means 聚类	(257)
10.3.1	K-Means 聚类分析的核心步骤	(257)
10.3.2	K-Means 聚类分析的基本操作	(258)
10.3.3	K-Means 聚类分析的应用举例	(259)
第 11 章	SPSS 的因子分析	(262)
11.1	因子分析概述	(262)
11.1.1	因子分析的意义	(262)
11.1.2	因子分析的数学模型和相关概念	(263)
11.2	因子分析的基本内容	(264)
11.2.1	因子分析的基本步骤	(264)
11.2.2	因子分析的前提条件	(265)
11.2.3	因子提取和因子载荷矩阵的求解	(266)
11.2.4	因子的命名	(269)
11.2.5	计算因子得分	(270)
11.3	因子分析的基本操作及案例	(271)

	11.3.1 因子分析的基本操作	(271)
	11.3.2 因子分析的应用举例	(273)
第 12 章	SPSS 的对应分析	(280)
12.1	对应分析概述	(280)
	12.1.1 对应分析的提出	(280)
	12.1.2 对应分析的基本思想	(280)
12.2	对应分析的基本步骤	(281)
12.3	对应分析的基本操作及案例	(283)
	12.3.1 对应分析的基本操作	(283)
	12.3.2 对应分析的应用举例	(285)
第 13 章	SPSS 的判别分析	(293)
13.1	判别分析概述	(293)
13.2	距离判别法	(293)
	13.2.1 距离判别的基本思路	(293)
	13.2.2 判别函数的计算	(294)
13.3	Fisher 判别法	(295)
	13.3.1 Fisher 判别的基本思路	(295)
	13.3.2 Fisher 判别的计算	(297)
13.4	贝叶斯判别法	(298)
	13.4.1 贝叶斯判别的基本思路	(298)
	13.4.2 贝叶斯判别的计算	(298)
13.5	判别分析的基本操作及案例	(299)
	13.5.1 判别分析的基本操作	(299)
	13.5.2 判别分析的准备工作:均值检验和协方差阵齐性检验	(300)
	13.5.3 解读判别结果	(303)
第 14 章	SPSS 的信度分析	(309)
14.1	信度分析概述	(309)
	14.1.1 信度分析的提出	(309)
	14.1.2 信度分析的基本原理	(310)
14.2	信度分析的基本操作及案例	(311)
	14.2.1 信度分析的基本操作	(311)
	14.2.2 信度分析的应用举例	(312)
第 15 章	SPSS 的对数线性模型	(316)
15.1	对数线性模型概述	(316)
	15.1.1 模型的提出	(316)
	15.1.2 基本概念和基本思路	(317)
15.2	饱和模型和非饱和层次对数线性模型	(318)
	15.2.1 饱和模型和参数估计	(318)
	15.2.2 饱和模型检验	(320)
	15.2.3 非饱和和层次模型	(325)

15.2.4	建立饱和模型和非饱和层次对数线性模型的基本操作	(325)
15.2.5	饱和模型和非饱和层次对数线性模型的应用举例	(326)
15.3	广义对数线性	(328)
15.3.1	广义对数线性的概述	(328)
15.3.2	建立广义对数线性的基本操作	(328)
15.3.3	广义对数线性模型的应用举例	(330)
15.4	Logit 对数线性模型	(332)
15.4.1	Logit 对数线性模型的概述	(332)
15.4.2	Logit 对数线性模型的应用举例	(333)
第 16 章	SPSS 的时间序列分析	(336)
16.1	时间序列分析概述	(336)
16.1.1	时间序列的相关概念	(336)
16.1.2	时间序列分析的一般步骤	(338)
16.1.3	SPSS 时间序列分析的特点	(340)
16.2	数据准备	(340)
16.3	时间序列的图形化观察及检验	(341)
16.3.1	时间序列的图形化观察及检验目的	(341)
16.3.2	时间序列的图形化观察工具	(342)
16.3.3	时间序列的检验方法	(349)
16.3.4	时间序列的图形化观察的基本操作和应用举例	(349)
16.4	时间序列的预处理	(353)
16.4.1	时间序列预处理的目的是主要方法	(353)
16.4.2	时间序列预处理的基本操作	(355)
16.5	时间序列的简单回归分析法和趋势外推法	(357)
16.5.1	简单回归分析法和趋势外推法概述	(357)
16.5.2	简单回归分析法和趋势外推法应用举例	(358)
16.6	指数平滑法	(361)
16.6.1	指数平滑法的基本思想	(361)
16.6.2	指数平滑法的模型	(362)
16.6.3	指数平滑法的基本操作	(364)
16.6.4	指数平滑法的应用举例	(367)
16.7	ARIMA 模型分析	(368)
16.7.1	ARIMA 分析的基本思想和模型	(368)
16.7.2	ARIMA 分析的基本操作	(370)
16.7.3	ARIMA 分析的应用举例	(371)
16.8	季节调整法	(377)
16.8.1	季节调整法的基本思想和模型	(377)
16.8.2	季节调整法的基本操作	(378)
16.8.3	季节调整法的应用举例	(379)

第 1 章 SPSS 统计分析软件概述

中国企业面临着市场化的挑战。在过去改革开放的三十多年里,我国的经济体制由计划经济步入了社会主义市场经济的运行轨道,这一客观现实要求中国企业必须能够独立地面对市场经济的大风大浪。企业只有利用一切可能的条件和手段,及时获取各种千变万化的市场信息,准确理解信息中蕴涵的深层含义,并迅速制定出符合企业自身特点的经营战略,在动态的市场环境中不断快速调整竞争策略和方向,才有可能在残酷的市场竞争中处于不败之地并不断发展。所以,吸收先进的管理理论,利用先进的信息技术和数据分析方法,“多快好省”地获取市场信息,了解市场变化,对市场信息进行科学分析和科学决策,已经成为中国企业能否成为竞争“赢家”的关键。在加入 WTO 后,中国企业在面临着更广泛的客户选择的同时,也面临着更强大对手的竞争。因此,在这种大趋势下,如果说工业化时代企业信奉“时间是金钱,效率是生命”的企业精神,那么,进入 21 世纪的信息化时代,企业则应树立“信息是金钱,决策是生命”的科学理念。

中国企业面临着客户需求个性化的挑战。一方面,随着企业生产技术和服务水平的普遍提高,产品的“同质化”特征越来越明显,市场竞争也已呈白热化趋势。此时,企业拓展生存空间、提升竞争力的有效途径是为客户提供满足客户需要的、竞争对手无法“效仿”的个性化服务;另一方面,随着现代社会人们自我意识的不断加强和对个性化需求的不懈追求,对企业提供个性化服务的呼声也愈来愈高。所有这些都强烈地冲击着传统的企业经营模式,企业的经营战略正逐步由“以产品为中心”的卖方市场向“以客户为中心”的买方市场转移,从关心重量、数量、质量的工业化思路向满足个性化服务的方向转移。

市场化和个性化对企业的挑战,要求企业承担起信息化和定量分析的重任,只有信息化才能够迅速全面地感知市场化和个性化的要求,只有定量分析才能对感知信息做出科学分析和正确决策。

以数据库技术等为代表的企业信息化技术,为收集、存储和管理大量的市场和客户数据,建立企业级的数据库管理信息系统打下了坚实的基础。而统计分析软件则是基于数据库基础之上实现深层次定量分析并辅助企业级决策的有效工具。

SPSS 统计分析软件以其强大的统计分析功能、方便易用的用户操作方式、灵活的表格式分析报告和精美的图形展现形式,赢得了各领域广大数据分析人员的喜爱,并得到了广泛应用。同时,SPSS 统计分析软件也已成为许多大专院校统计学专业和其他财经类、管理类专业本科学生和研究生的必修课程。

1.1 SPSS 的发展及特点

SPSS(Statistical Package for the Social Science)社会科学统计软件包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 软件主要在中小型计算机上运行,统称为 SPSS_x 版,主要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,并确立了微机个人用户市场第一的地位。20 世纪 90 年代,随着微机 Windows 图形化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 第七版至第二十版又相继诞生,统称为 SPSS for Windows 版。

1994 年至 1998 年间,SPSS 公司陆续购并了 SYSTAT、BMDP、Quantime、ISL、ShowCase 等公司,并将其各自的主打产品收纳于 SPSS 麾下,从而使 SPSS 由原来单一的统计分析软件向为企业、教育科研、政府机构等统计决策服务的综合性产品发展。为此 SPSS 公司已将原英文名称更改为 Statistical Product and Service Solutions,即统计产品与服务解决方案。目前,SPSS 在全球约有 26 万家产品用户,分布于通信、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等多个行业和领域,已成为世界最为流行、应用最为广泛的专业统计分析软件之一。

2009 年 7 月,IBM 公司斥资 12 亿美元收购了 SPSS 软件公司,SPSS 软件更名为 IBM SPSS Statistics,也称 PASW(Predictive Analysis SoftWare) Statistics。

从 SPSS/PC+版本到 SPSS for Windows 版本,SPSS 在用户操作和分析结果的展现方面做了非常大的改进。SPSS/PC+版本是通过命令行方式(即用户输入 SPSS 命令和参数)完成数据的管理和统计分析工作的,统计分析文字结果和图形结果均以文本字符方式展现。SPSS for Windows 的第五版、第六版,在保留以前版本的手工输入命令、参数操作方式的同时,还为用户提供了直观的图形化菜单界面。用户的数据管理和统计分析工作可以非常方便地通过鼠标单击菜单或按钮,并配合简单的对话框输入来实现,从而免去了记忆命令和参数的负担,也不需要任何计算机编程。同时,对于图形分析结果,能够以图形点阵的方式显示,并允许用户对其进行编辑。SPSS for Windows 第七版以后的版本,进一步完善了用户的操作界面,增加了用户熟悉的与 Windows 95 风格一致的工具栏按钮。在分析结果的输出处理方面,改变了原来版本的输出方式,以崭新的树形结构方式管理分析结果,以文字、表格、图形混排的形式展现分析结果,以强大、灵活的编辑功能随心所欲地编辑分析结果,以内容丰富的“统计教练”联机帮助方式帮助用户理解分析结果。同时,还实现了对互联网的全面适应,并支持数据的动态收集、分析和 HTML 格式报告等。

“易学易用易普及”已成为 SPSS 软件最大的竞争优势之一,也是广大数据分析人员对其偏爱有加的主要原因。而大量成熟的统计分析方法、完善的数据定义操作管理、开放的数据接口以及灵活的统计表格和统计图形,更是 SPSS 常胜不衰的重要法宝。

1.2 SPSS 使用基础

快速掌握 SPSS for Windows,需要首先了解 SPSS 安装的基本步骤,熟悉 SPSS 的基本操作环境,掌握启动和退出 SPSS 的方法。

1.2.1 SPSS 软件的安装和启动

作为 Windows 操作系统的应用软件产品,SPSS for Windows 安装的基本步骤与其他常用软件是基本相同的。