

经·济·与·管·理·类·统·计·学·系·列·教·材

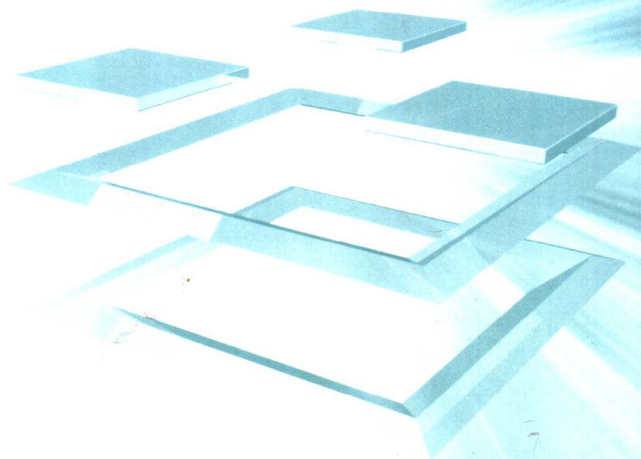


普通高等教育“十一五”国家级规划教材

教育部高等学校统计学专业
教学指导分委员会推荐用书

应用多元统计分析

朱建平 主编



 科学出版社
www.sciencep.com

经·济·与·管·理·类·统·计·学·系·列·教·材

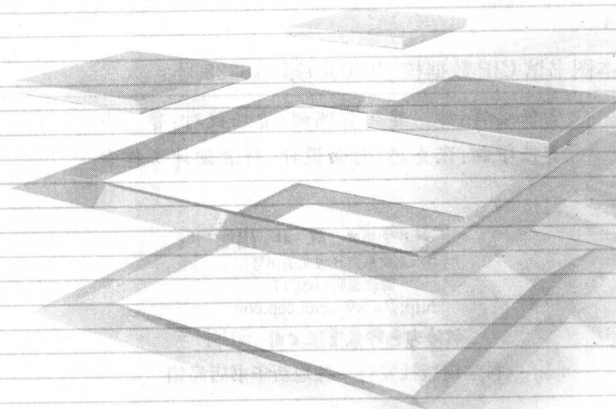


普通高等教育“十一五”国家级规划教材

教育部高等学校统计学专业
教学指导分委员会推荐用书

应用多元统计分析

朱建平 主编



科学出版社

北京

内 容 简 介

本书为普通高等教育“十一五”国家级规划教材,同时也是教育部高等学校统计学专业教学指导分委员会推荐教材。本书努力贯彻“少而精”的原则,力求以统计思想为主线,以 SPSS 软件为工具,深入浅出地介绍各种多元统计方法的理论和应用。主要包括:多元分析概述、多元正态分布的参数估计、多元正态分布均值向量和协差阵的检验、判别分析、聚类分析、主成分分析、因子分析、相应分析、典型相关分析、多维标度法、多变量的可视化分析等。特别是,本书将 SPSS 软件的学习和案例分析有机结合,体现了多元统计分析方法的应用。

本书配备多媒体教学课件,可作为经济类、管理类各专业本科生教材,同时也适合自学多元统计分析的读者阅读参考。

图书在版编目(CIP)数据

应用多元统计分析/朱建平主编. —北京:科学出版社,2006

(普通高等教育“十一五”国家级规划教材=经济与管理类统计学系列教材)

ISBN 7-03-017512-3

I. 应… II. 朱… III. 多元分析:统计分析-高等学校-教材

IV. O212.4

中国版本图书馆 CIP 数据核字(2006)第 070322 号

责任编辑:林 建 贾瑞娜/责任校对:鲁 素

责任印制:张克忠/封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

德信彩色印装有限公司印刷

科学出版社发行 各地新华书店经销

*

2006年8月第 一 版 开本:B5(720×1000)

2006年8月第一次印刷 印张:14 1/2

印数:1—3 000 字数:261 000

定价:22.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

经济与管理类统计学系列教材编委会

主任

曾五一 教育部高等学校统计学专业教学指导分委员会委员、厦门大学教授、博导

委员（以姓氏笔画排序）

王艳明 山东工商学院教授

王振龙 教育部高等学校统计学专业教学指导分委员会委员、西安财经学院教授

朱建平 厦门大学教授、博导

刘洪 中南财经政法大学教授

刘建平 暨南大学教授

许鹏 湖南大学教授

李金昌 浙江工商大学教授、博导

李宝瑜 山西财经大学教授、博导

杨灿 厦门大学教授、博导

肖红叶 天津财经大学教授、博导

张润楚 教育部高等学校统计学专业教学指导分委员会委员、南开大学教授、博导

茆诗松 华东师范大学教授、博导

杭斌 山西财经大学教授、博导

罗良清 江西财经大学教授

周恒彤 天津财经大学教授、博导

庞皓 教育部高等学校经济学教学指导委员会副主任、西南财经大学教授、博导

郑明 教育部高等学校统计学专业教学指导分委员会委员、复旦大学教授、博导

徐国祥 上海财经大学教授、博导

蒋萍 东北财经大学教授、博导

雷钦礼 暨南大学教授、博导

黎实 西南财经大学教授、博导



总序

统计学是有关如何测定、收集和分析反映客观总体数量的数据，以便给出正确认识的方法论科学。随着社会经济的发展和科学技术的进步，统计应用的领域越来越广，统计已经成为人们认识世界不可或缺的重要工具。

现代统计学可以分为两大类：一类是以抽象的数量为研究对象，研究一般的收集数据、分析数据方法的理论统计学；另一类是以各个不同领域的具体数量为研究对象的应用统计学。前一类统计学具有通用方法论的理学性质，其特点是计量不计质；后一类统计学则与各不同领域的实质性学科有着非常密切的联系，是有具体对象的方法论，因而具有复合性学科和边缘学科的性质。所谓应用，既包括一般统计方法的应用，也包括各自领域实质性科学理论的应用。经济与管理统计学是以社会经济数量为对象的应用统计学。要在经济和管理领域应用统计方法，必须解决如何科学地测定经济现象即如何科学地设置指标的问题，这就离不开对有关经济现象的质的研究。要对经济和管理问题进行统计分析，也必须以有关经济和管理的理论为指导。因此，经济与管理统计学的特点是在质与量的紧密联系中，研究事物的数量特征和数量表现。不仅如此，由于社会经济现象所具有的复杂性和特殊性，经济与管理统计学除了要应用一般的统计方法外，还需要研究自己独特的方法，如核算的方法、综合评价的方法等等。

从历史和现状看，我国统计学专业的办学也有两种模式：一是强调各类统计学所具有的共性。这种模式主要培养学生掌握通用的统计方法和理论。它肯定统计学的“理学性质”，按照理学类学科的特点设置课程，概率论和各种数理统计方法等通用的统计方法论在课程中占有较大分量。其培养目标是有良好的数学基础、熟练掌握统计学基本理论与各种方法，同时有一定的专门领域的知识，能够适应各个不同领域的统计工作和统计研究的统计人才。二是强调各类统计学的个

性,对于经济与管理类统计学来说,就是强调其与经济学和管理学等其他学科的密切联系,按照经济与管理类学科的特点设置课程,除统计学本身的专业课外,经济管理类的课程占相当大的比重。其培养目标是所谓的“复合型人才”,即具有坚实的经济与管理理论功底,既懂数理统计方法又懂经济统计方法,并能熟练掌握现代计算手段的经济与管理统计人才。这种人才既是统计人才又是经济管理人才,不仅能胜任基层企事业单位和政府部门的日常统计业务,而且能从事市场调查、经济预测、信息分析和其他经济管理工作。上述两种办学模式,各有特色,同时也各有一定的社会需求。从我国的国情看,现阶段后一种模式培养的人才市场需求要更大一些。应该根据“百花齐放,百家争鸣”的方针,允许多种办学模式同时并存,由各院校根据自己的特色和市场对有关人才需求的大小,自主选择合适的办学模式。

为了更好地满足新世纪对统计人才的需要,无论是理学类统计学专业还是经济管理类统计学专业都有一个如何面向未来、面向世界、加强自身建设、更好地与国际接轨的问题。但是,这两类专业的培养目标不同,知识体系也有相当大的差异,难以完全统一或互相取代。2003年11月,教育部高等学校统计学专业教学指导分委员会在厦门召开年会,会上各方面的专家达成共识,为了促进统计学的学科建设和发展,有必要按授予学位的不同,分别制定指导性的教学规范。2004年1月,全国经济与管理类统计学专业的部分专家和学者在天津财经学院讨论了《统计学专业教学规范(授经济学学士学位)》征求意见稿,对初稿进行修改与补充,又经过教育部高等学校统计学专业教学指导分委员会研究和审定,最终形成了正式的教学规范(以下简称新规范),并已上报国家教育部。

根据新规范的设计,经济管理类统计学专业应开设的统计学专业主干课程包括以下10门:①统计学导论;②数理统计学;③应用多元统计分析;④应用时间序列分析;⑤应用抽样技术;⑥计量经济学;⑦国民经济统计学;⑧企业经营统计学;⑨证券投资分析;⑩货币与金融统计学。为了进一步提高统计教材的质量,更好地满足新世纪培养经济管理类统计人才的需要,我们成立了经济与管理类统计学系列教材编委会,组织全国高校同行分工协作,根据新规范设计的课程体系和教学内容,编辑出版一套经济与管理类统计学系列教材。本系列教材比较适合作为高等院校经济与管理类统计学专业的教材,其中《统计学导论》和《计量经济学》还可作为一般经济与管理类专业的核心课程教材。

参加本系列教材建设的有厦门大学、西南财经大学、天津财经大学、上海财经大学、浙江工商大学、山西财经大学、湖南大学、西安财经学院、南开大学、东北财经大学、中南财经政法大学、暨南大学、华东师范大学、江西财经大学、山东工商学院、复旦大学等10多所院校的教师。本系列教材实行主编负责制,担任主编和主审的老师都是曾经主编和出版过相关统计教材的国内知名教授,不

仅具有在第一线教学的经验，而且对相关学科的发展趋势和学科前沿也比较熟悉。本系列教材的编写，力求体现以下特点：

(一) 与时俱进，构建与培养目标相适应的教学内容体系。

教材建设的关键在于构建与培养目标相适应的教学内容体系。为此，要根据时代的发展，不断补充和引进新的教学内容。作为新世纪经济管理类统计专业的教材，不能只是简单地向理学类统计专业的教学内容靠拢，而应该根据自身的特点，努力贯彻“少而精”和“学以致用”的原则。在大胆吸收国外优秀教材特点的基础上，对原有的体系重新进行整理和完善，既适当增加一些在经济社会分析中有良好应用前景的数理统计理论与方法的内容，又适当增补经济社会统计方面的最新进展。同时删除过时的和不再适用的内容。尽可能做到既反映本门学科的先进水平，又比较简明易懂，便于教学。

(二) 统筹兼顾，防止低水平重复，发挥系列教材的整体功能。

适应未来需要的经济管理类统计人才必须掌握多方面的知识和能力，各种知识是相互联系的，各门课程在内容上难免有所交叉。为了提高学习的效率，更好地发挥系列教材的整体功能，在编写本系列教材的过程中，我们作了必要的协调和适当的分工，尽可能做到统筹兼顾，防止低水平重复。同时，本系列教材采用相同的版式、体例和统一规范的学术用语。

(三) 与计算机结合，培养学生的动手能力。

为了提高学生运用统计方法解决实际问题的能力，本系列教材的编写注意与计算机的紧密结合。本系列教材中统计方法类的教材均根据教材的内容，结合常用的计算机统计软件，并给出相应的案例和数据。从而使学生不仅可以从中学习统计学理论和方法，而且可以实际上机操作，培养实际动手的能力。

(四) 编写体例新颖，提高学生学习的兴趣和效率。

为了便于师生教学互动，提高学生学习的兴趣和学习效率，本系列教材在编写体例上也作了一些新的尝试。各章开篇有内容要点和教学要求提示，章末附有小结，对有关教学内容和计算公式作扼要的总结。教材中尽可能使用本国的真实数据作为案例。各教材的“思考与练习”部分不设一般常见的名词解释型的简答题和论述题，而是通过判断题、选择题、计算题和有趣味的思考题，来帮助学生掌握有关概念和计算方法。为便于学生自学，同时又为其留有独立思考和独立完成作业的余地，各教材均给出编号为奇数的习题的详解。

(五) 配套出版教学课件和光盘，便于教师组织教学。

本系列教材在出版纸质出版物的同时配套出版相应的光盘。内容包括：用PowerPoint制作的课件，教材中有关案例的数据，常用的统计表，编号为奇数的习题的详解等。

经济与管理类统计学专业系列教材的建设是一个复杂艰巨的系统工程，完成

这一工程需要全国统计教育工作者的共同努力。感谢参与本系列教材编写的全国各高校的专家和学者，感谢为本系列教材的出版提供帮助的科学出版社的领导和编辑。衷心祝愿大家的辛勤劳动能够结出丰硕的果实，能够为我国统计学的普及和提高做出更大的贡献。

曾五一

2006年1月于厦门

前 言

随着电子计算的日益普及，各行各业都开始采用计算机及相应的信息技术进行管理和决策，这使得各企事业单位生成、收集、存储和处理数据的能力大大提高，数据量与日俱增，大量复杂信息层出不穷。这样我们将面临复杂数据的处理问题，特别是研究客观事物中多个变量（或多个因素）之间相互依赖的统计规律性，它的重要理论基础之一就是多元统计分析。多元统计分析是统计学中一个非常重要的分支，它在各学科领域中得到了越来越广泛的应用，是一套非常有用的数据处理方法。

在信息爆炸的今天，人们已经意识到数据的价值。显然，大量信息在给人们带来方便的同时也带来一系列问题。例如，信息量过大，超过了人们掌握、消化的能力；一些信息真伪难辨，从而给信息的正确应用带来了困难；信息组织形式的不一致性导致人们难以对信息进行有效统一的处理，等等。为了能更好地从统计学的角度解决这些问题，我们组织编写了《应用多元统计分析》这本书，并且作为教育部统计学专业教学指导分委员会推荐的经济与管理类统计学系列教材之一。

本书的编写力求以统计思想为主线，以 SPSS 软件为工具，深入浅出地介绍各种多元统计方法的应用。其基本框架是：第一章为概述；第二章和第三章介绍多元正态总体的参数估计和假设检验；第四章至第十章介绍常用的多元统计方法，这些方法包括判别分析、聚类分析、主成分分析、因子分析、相应分析、典型相关分析、多维标度法；第十一章介绍多变量的可视化分析等。

在本书的编写过程中，我们根据经济和管理类的专业要求，突出以下特点：

第一，把握统计实质，贯穿统计思想。注重统计思想的讲述，在多元统计方法的应用上把握实质，从实际问题入手，在不失严谨的前提下，淡化统计方法本

身的数学推导,体现统计学的实用性。

第二,应用 SPSS 软件,实现统计计算。根据多元统计固有的特点,我们选用在我国广泛流行的 SPSS 软件作为计算工具。在每一章的最后,都要讲述所介绍的多元统计方法在 SPSS 软件中的实现。这样将 SPSS 软件的学习和案例分析有机结合,不仅能使得学生在实践运用中学习 SPSS 软件的操作方法,而且还能使学生对多元统计分析的意义有深入的体会。

第三,加强统计理论,完成统计实践。根据实际介绍的统计方法,我们将编写的习题分为两类:一类是继续巩固和加强统计理论及方法,包括基本概念和基本思路训练的习题;另一类是针对实际问题,培养学生结合统计方法独立解决实际问题的能力和素质的习题。

为了提高学生的学习兴趣和学习效率,考虑到不同的使用对象和教学特点,对部分内容可根据实际需要进行选讲。

本书第一、二、三、四、八章由厦门大学朱建平教授编写,第五、六、七章由厦门大学朱平辉副教授编写,第九、十、十一章由山西财经大学范霄文副教授编写。每章最后计算机实现的内容由厦门大学博士研究生殷瑞飞编写,每章习题的编写工作由河南财经学院靳刘蕊讲师完成,PowerPoint 的制作由厦门大学袁加军工程师完成。本书由朱建平教授担任主编并进行修改和总纂。暨南大学雷钦礼教授担任本书主审,认真审阅了全书。

本书在编写过程中参考了一些国内外相关文献资料,书后列出了主要参考文献。本书的出版得到了科学出版社的大力支持,在此一并表示衷心感谢!编写一本好的书并不容易,尽管我们努力想奉献一本令读者满意的书,但书中难免有疏漏或错误之处,恳请读者多提宝贵意见,以便今后进一步修改与完善。

本书的编写得到了国家教育部“新世纪优秀人才支持计划”(Program for New Century Excellent Talents in University, NCET)的资助。

编者

2006年5月

目 录

总序 前言

第一章

多元分析概述 1

第一节 引言 1

第二节 应用背景 2

第三节 计算机在统计分析中的应用 6

第二章

多元正态分布的参数估计 8

第一节 引言 8

第二节 基本概念 9

第三节 多元正态分布 14

第四节 多元正态分布的参数估计 17

第五节 实例分析与计算机实现 23

思考与练习 26

第三章	
	多元正态分布均值向量和协差阵的检验 28
第一节	引言 28
第二节	均值向量的检验 28
第三节	协差阵的检验 37
	思考与练习 38
第四章	
	判别分析 42
第一节	引言 42
第二节	距离判别法 43
第三节	贝叶斯判别法 47
第四节	费希尔判别法 50
第五节	实例分析与计算机实现 54
	思考与练习 60
第五章	
	聚类分析 62
第一节	引言 62
第二节	相似性的量度 63
第三节	系统聚类分析法 65
第四节	K 均值聚类分析 75
第五节	有序样品的聚类分析法 77
第六节	实例分析与计算机实现 82
	思考与练习 90

第六章

主成分分析	93
第一节 引言	93
第二节 主成分的几何意义及数学推导	94
第三节 主成分的性质	97
第四节 主成分方法应用中应注意的问题	99
第五节 实例分析与计算机实现	100
思考与练习	106

第七章

因子分析	109
第一节 引言	109
第二节 因子分析模型	110
第三节 因子载荷矩阵求解	113
第四节 公因子重要性的分析	117
第五节 实例分析与计算机实现	121
思考与练习	128

第八章

相应分析	131
第一节 引言	131
第二节 列联表	132
第三节 相应分析的基本理论	134
第四节 相应分析中应注意的问题	136
第五节 实例分析与计算机实现	137
思考与练习	146

第九章	
	典型相关分析 149
第一节	引言..... 149
第二节	典型相关的基本理论..... 150
第三节	样本典型相关分析..... 154
第四节	典型相关分析应用中的几个问题..... 160
第五节	实例分析与计算机实现..... 164
	思考与练习..... 170
第十章	
	多维标度法 173
第一节	引言..... 173
第二节	古典多维标度法..... 174
第三节	权重多维标度..... 183
第四节	实例分析与计算机实现..... 184
	思考与练习..... 190
第十一章	
	多变量的可视化分析 193
第一节	引言..... 193
第二节	折线图分析法..... 194
第三节	条形图分析法..... 196
第四节	散点图分析法..... 197
第五节	雷达图分析法..... 199
第六节	星座图分析法..... 202
	思考与练习..... 205
	参考文献 206

附录

	常用统计表	207
附表 1	正态分布概率表	207
附表 2	t 分布临界值表	210
附表 3	χ^2 分布临界值表	211
附表 4	F 分布临界值表	212

第一章

多元分析概述

第一节 引言

多元统计分析是运用数理统计方法来研究解决多指标问题的理论和方法。近30年来,随着计算机应用技术的发展和科研生产的迫切需要,多元统计分析技术被广泛地应用于地质、气象、水文、医学、工业、农业和经济等许多领域,已经成为解决实际问题的有效方法。然而,随着电子计算的日益普及,各行各业都开始采用计算机及相应的信息技术进行管理和决策,这使得各企事业单位生成、收集、存储和处理数据的能力大大提高,数据量与日俱增,大量复杂信息层出不穷。在信息爆炸的今天,人们已经意识到数据最值钱的时代已经到来。显然,大量信息在给人们带来方便的同时也带来了一系列问题。例如,信息量过大,超过了人们掌握、消化的能力;一些信息真伪难辨,从而给信息的正确应用带来困难;信息组织形式的不一致性导致难以对信息进行有效统一的处理,等等,这种变化使传统的数据库技术和数据处理手段已经不能满足要求。Internet的迅猛发展,也使得网络上的各种资源信息异常丰富,在其中进行信息的查找真如大海捞针。这样又对多元统计分析理论的发展和方法的应用提出了新的挑战。

多元统计分析起源于20世纪初,1928年Wishart发表论文《多元正态总体样本协差阵的精确分布》,可以说是多元分析的开端。20世纪30年代R. A. Fisher、H. Hotelling、S. N. Roy、许宝騄等人做了一系列的奠基性工作,使多元分析在理论上得到了迅速发展。40年代多元分析方法在心理、教育、生物等方面有不少应用,但由于计算量大,使其发展受到影响,甚至停滞了相当长的时间。50年代中期,随着电子计算机的出现和发展,多元分析方法在地质、气象、医学、社会学等方面得到

广泛的应用。60年代通过应用和实践又完善、发展了理论,由于新的理论、新的方法不断涌现,又促使它的应用范围更加扩大。70年代初期多元统计分析在我国才受到各个领域的极大关注,并在理论研究和应用上也取得了许多显著成绩,有些研究工作已达到国际水平,并已形成一支科技队伍,活跃在各条战线上。在20世纪末与21世纪初,人们获得的数据正在以前所未有的速度急剧增加,产生了很多超大型数据库,遍及超级市场销售、银行存款、天文学、粒子物理、化学、医学以及政府统计等领域,多元统计与人工智能和数据库技术相结合,在经济、商业、金融、天文等领域得到了成功的应用。

为了让读者更好、更系统地掌握多元统计分析的理论与方法,本书重点介绍多元正态总体的参数估计和假设检验以及常用的统计方法。这些方法包括判别分析、聚类分析、主成分分析、因子分析、对应分析、典型相关分析、多维标度法以及多变量的可视化分析等。与此同时,我们将利用在我国广泛流行的SPSS统计软件来实现实证分析,做到在理论的学习中体会应用,在应用的分析中加深理论。

第二节 应用背景

统计方法是科学研究的一种重要工具,其应用颇为广泛。特别地,多元统计分析方法常常被应用于自然科学、社会科学等领域的问题中。为了进一步体现多元统计分析方法的应用,我们首先从宏观的角度认识统计学应用的背景,然后从微观的角度显示多元统计分析应用的广泛性。

一、统计学的生命力在于应用

(一) 统计学产生于应用

统计学的发展过程可以看出统计学产生于应用,在应用的过程中发展,它的生命力在于应用。

300年前,威廉·配第(1623~1687)写的《政治算术》,从其研究方法看,被认为是一本统计学著作。政治算术学派的统计学家将统计方法应用于各自熟悉和感兴趣的研究领域,都还是把其应用对象当作肯定性事物之间的联系来进行研究的。他们确信,事物现象存在着简单明了的数量关系,需要用定性与定量的方法将这种关系(规律)揭示或描述出来,使人们能够更具体、真切地认识世界。

数理统计学派的奠基人凯特勒在统计学中引入了概率论,把它应用于自然界和社会的许多方面,从而为人们认识和说明不确定现象及其相互之间的联系开辟出了一条道路。在自然科学和社会科学的许多领域,都留下凯特勒应用统计学研究的烙印。自从凯特勒把概率论引入了应用中的统计学,人们对客观世界的认识