普通高等教育“十一五”国家级规划教材

全国统计教材编审委员会“十一五”规划教材

多元统计分析

第二版

★ 任雪松 于秀林 编著

中国统计出版社
China Statistics Press

普通高等教育“十一五”国家级规划教材

全国统计教材编审委员会“十一五”规划教材

多元统计分析

第二版

任雪松 于秀林 编著

中国统计出版社
China Statistics Press

(京)新登字 041 号

图书在版编目(CIP)数据

多元统计分析/任雪松,于秀林编著. —北京:中国统计出版社,2010.12

ISBN 978-7-5037-6182-9

Ⅰ.①多… Ⅱ.①任…②于 Ⅲ.①多元分析:统计分析—高等学校—教材 Ⅳ.①O212.4

中国版本图书馆 CIP 数据核字(2010)第 260293 号

多元统计分析

作　　者/任雪松　于秀林
责任编辑/胡文华
装帧设计/黄　晨
出版发行/中国统计出版社
通信地址/北京市西城区月坛南街 57 号　邮政编码/100826
办公地址/北京市丰台区西三环南路甲 6 号
网　　址/www.stats.gov.cn/tjshujia
电　　话/邮购(010)63376907　书店(010)68783172
印　　刷/利兴印刷有限公司
经　　销/新华书店
开　　本/710×1000mm　1/18
字　　数/370 千字
印　　张/23.75
印　　数/1—5000 册
版　　别/2011 年 3 月第 1 版
版　　次/2011 年 3 月第 1 次印刷
书　　号/ISBN 978-7-5037-6182-9/O·77
定　　价/39.00 元

出版说明

“十一五”时期是继续深化教育改革、加强素质教育、努力建设有利于创新型科技人才生长的教育培训体系的关键时期。为了更好地培育统计创新型科技人才，适应统计教育培训的新形势，全国统计教材编审委员会制定了《“十一五”全国统计教材建设规划》(以下简称《规划》)。规划坚持“以人为本”的科学发展观，坚持统计教育与实践相结合，坚持统计教育同国际接轨，坚持培养创新型的统计人才的指导思想，编写符合国民经济发展需要和统计事业发展需要的统计教材。

这批教材是在深入分析统计教育形势和统计教材建设发展状况，总结多年来统计教材建设经验的基础上，本着以建设本科统计教材为主的方针，积极探索研究生层次的统计教材，力争使规划统计教材的编写做到层次分明，有针对性和实用性。建设精品教材，是编委会自成立以来就孜孜以求的目标。考虑到统计教材建设的实际情况，“十一五”期间，本科教材主要以修订为主，对以往规划统计教材中使用面广、得到广大教师和学生普遍认可的教材组织了修订。修订后的教材，淘汰了过时的内容和例子，增加了计算机操作和大量的案例，编写手法也做了一定的调整，在实用性、可操作性等方面有了较大的改进。

近年来，我国现代化建设快速发展，高等教育规模持续扩大，尤其是研究生教育规模的扩大，使得高等学校研究生统计教学工作面临着许多新情况、新问题，任务艰巨。因此，必须坚持科学发展观，在规模持续发展的同时，把提高研究生统计教学质量放在突出的位置，培养全面发展的创新型的统计人才。教材是统计教学的载体，建设高质量

的研究生层次的统计教材是统计教育发展的需要。因此，编委会在“十一五”期间对研究生的统计基础课教材编写做了些有益的探索。根据《规划》的要求，这批教材主要采取招标和邀请的方式组织有关院校的专家、学者编写。

值得特别提出的是，在这批教材中，有《非参数统计》、《概率论与数理统计》、《经济计量学教程》、《医学统计》、《应用时间序列分析》、《多元统计分析》、《统计学》、《现代指数理论》、《现代金融投资统计分析》9部教材入选国家教育部组织编写的“普通高等教育‘十一五’国家级规划教材”，更加充实和完善了“十一五”期间统计教材的建设。

为了便于教学和学习，这批教材里面包含了与之相配套的《学习指导与习题》，使得这批教材在编辑出版上形成了比较完整的体系。我们相信，这批教材的出版和发行，对于推动我国统计教育改革，加快我国统计教材体系和教材内容更新、改造的步伐，打造精品教材，都将起到积极的促进作用。

限于水平和经验，这批教材的编审、出版工作还会有缺点和不足，诚恳欢迎教材的使用单位、广大教师和同学们提出批评和建议。

全国统计教材编审委员会

2006年6月

再版前言

多元统计分析简称多元分析，是数理统计学的一个重要分支。回想11年前这本书出版时，市面上此类书不多，而近年来这类书的出版量增加很快，看到多元统计分析方法在诸多领域中的广泛应用，深感统计方法的重要性以及影响的巨大。由于多元统计分析方法处理的是多维数据，往往数据量很大，而计算机技术的普及和发展为处理多维数据提供了强有力的平台，多元统计分析之所以如此快速的发展，主要基于两点：一是实用性强，由于定量分析的重要性使多元统计分析的主要方法成为处理多维数据不可缺少的重要工具（什么是多元统计分析，它能解决哪些问题可见本书第一章内容）；二是计算机的普及和发展，为处理多维数据提供了强有力的平台。这种多维数据分析方法和计算机技术的结合是相辅相成的。各种统计方法的广泛应用，使计算机的功能得到普及和发展，产生出一些相当复杂的统计软件包，如SPSS、SAS、Minitab和R等，反过来，有了计算机软件的开发，使大量复杂的各种统计运算变得简单可行，因此也促进了很多统计理论和方法的产生和发展。实践证明，多元统计方法是处理多维数据不可缺少的重要工具，并日益显示出无比的魅力。

为了适应国民经济诸多领域做定量分析的需要以及当前教学改革不断深入发展的需要，目前国内很多高等院校相继给研究生和本科生开设了该课程。

作者在原出版《多元统计分析》一书的基础上，根据全国统计教材编委会专家评审组通过的《编写大纲》要求，对原书进行了修改、充实。借助再版的机会补充了：路径分

析和多维标度法等，力争写出一本适合财经、统计、管理等专业的教材，同时也想给对多元分析方法感兴趣的广大科技工作者提供一本较系统掌握这一方法的良好参考书。

本书特点：1. 概念清晰，方法明了，强调实际应用。2. 在一元统计分析的基础上深入浅出地介绍多元分析的内容，并着重介绍多元分析中常用的各种方法，讲清各种方法的实际背景和统计思想，同时每种方法都给出具体实例。3. 为了适合不同层次读者的需要和加深对各种方法的理解以及期望读者能灵活地运用这些方法，作者对多元分析中的一些理论也给出适当的论证和说明，但大多数理论，只是叙述结果，而有关理论证明，可查看本书后面列出的参考书。4. 本书介绍的各种统计方法可使用国内外通用的 SPSS 和 SAS 软件去实现，不再附计算程序。5. 本书对主要章节给出附注，目的有两个：一是对本章节所介绍的内容进一步引申。二是扩展，即进一步补充所介绍的内容，因此，附注内容根据学生情况可选讲或不讲，对实际工作者可选读或不读，对全书主要内容的掌握不受影响。

学习本书之前应具备以下三方面的知识：1. 由于向量和矩阵是研究多元数据的重要工具，所以要求读者具有一定的线性代数知识，本书附录中针对本书的需要复习了有关这方面的基本知识。2. 多元统计分析是建立在一元统计分析的基础上的，因此要求读者具有初等数理统计知识。3. 多元统计分析是依赖于计算机的发展而发展的，为了做到学以致用，要求读者会调用国内外通用的某一种统计软件包能上机操作即可，并不要求自编程序去实现各种方法的计算。

本书在编写过程中得到全国统计教材编委会专家们的关心和帮助，以及北京大学数学科学院陈家鼎教授给予的热心指导。书稿完成后，又经过全国统计教材编委会召开的专家审稿会作了评审，他们对书稿提出了许多很好的建议。最后由中国科学院系统科学研究所吴启光教授审稿。为了便于读者阅读本书时，更好地理解和掌握有关内

容，在个别章节选取了经典著作中的典型例子，在此一并向他们表示衷心的感谢！特别要感谢中国统计出版社的同志们，没有他们的关心和支持，这本书不可能面世。

希望通过这本书的出版，为多元分析的普及和发展起到一定的促进作用，也为高等院校的教学改革和培养更多全面发展的人才做些有意义的工作，使这一有效的数学工具更好地为社会主义市场经济服务。

由于水平有限，书中难免有不足之处，欢迎读者批评指正。

编者

2010 年底于中国人民大学

目录

目录

目录

目录

目录

目录

目录

第一章

绪　论

1.1　什么是多元统计分析及发展简史

在工业、农业、医学、气象、环境以及经济、管理等诸多领域中，常常需要同时观测多个指标。例如，要了解一个国家经济发展的类型需要观测的指标有：人均国民收入，人均工农业产值、人均消费水平等等。要衡量一个地区的经济发展，需观测的指标有：总产值、利润、效益、劳动生率、万元生产值耗、固定资产、流动资金周转率、物价、信贷、税收等等；要了解一种岩石，需观测或化验的指标也很多，如：颜色、硬度、含碳量、含硫量等等；在医学诊断中，要判断某人是有病还是无病，也需要做多项指标的体检，如：血压、心脏脉搏跳动的次数、白血球、体温等等。总之，在科研、生产和日常生活中，受多种指标共同作用和影响的现象是大量存在的，举不胜举。上述指标，在数学上通常称为变量，由于每次观测的指标值是不能预先确定的，因此每个指标可用随机变量来表示。

如何同时对多个随机变量的观测数据进行有效地分析和研究呢？一种做法是把多个随机变量分开分析，一次处理一个去分析研究；另一种做法是同时进行分析研究。显然前者做法有时是有效的，但一般来说，由于变量多，避免不了变量之间有相关性，如果分开处理不仅会丢失很多信息，往往也不容易取得好的研究结果。而后一种做法通常可以用多元统计分析方法来解决，通过对多个随机变量观测数据的分析，来研究变量之间的相互关系以及揭示这些变量内在的变化规律，如果说一元统计分析是研究一个随机变量统计规律的学科，那么多元统计分析则是研究多个随机变量之间相互依赖关系以及内在统计规律性的一门统计学科。其内容既包括一元统计分析中某些理论和方法的直接推广，也包括多元随机变量特有的一些问题，如：利用多元统计分析中某些方法，

对研究的对象进行分类（指标分类或样品分类）和简化（把相互依赖的变量变成独立的或降低复杂集合的维数等等）。在当前科技和经济迅速发展的今天，在国民经济许多领域特别对社会经济现象的分析，只停留在定性分析上往往是不够的。为提高科学性、可靠性、通常需要定性与定量分析相结合。实践证明，多元统计分析是实现做定量分析的有效工具。

多元统计分析简称为多元分析起源于本世纪初，1982 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、许宝碌等人作了一系列的奠基性工作，使多元分析在理论上得到了迅速的发展。40 年代在心理、教育、生物等方面有不少的应用，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，使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和发展了理论，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 年代初期在我国才受到各个领域的极大关注，30 余年来我国在多元分析的理论研究和应用上也取得了很多显著成绩，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。

1.2 多元分析能解决哪些类型的实际问题

下面例举一些实际问题，从中不仅可以看到多元分析能解决哪些不同类型的问题，而且还可以看到多元分析应用的广度和深度，它将会引起学习者们的浓厚兴趣。

经济学：

1. 对我国 30 个省市自治区的社会情况进行分析，一般不是逐个省市自治区去分析，而较好地做法是选取能反映社会情况的代表性指标，如：人口密度、城市和农村的平均每人每月收入和支出情况、居住面积、城市绿化覆盖率等等，根据这些指标对 30 个省市自治区进行分类，然后根据分类结果对社会情况进行综合评价。又如要考察北京、天津等几所大城市的企业情况，首先要选取企业方面有代表性指标，如：企业个数、工业总产值、平均人数、固定资产净值、资金利税率、资金利润率、全员劳动生产率等等。由于要考虑的指标多，通常先对指标进行分类，按分类结果对指标进行综合分析给出企业的评价。如何分类？可用 Q 型和 R 型聚类分析法。

2. 在经济学中，可根据人均国民收入、人均农业产值、人均消费水平等多