

经·济·与·管·理·类·统·计·学·系·列·教·材



“十二五”普通高等教育本科国家级规划教材

教育部高等学校统计学类专业  
教学指导委员会推荐用书

# 应用多元统计分析

## (第三版)

朱建平 主编



科学出版社



经济与管理类统计学系列教材  
“十二五”普通高等教育本科国家级规划教材

# 应用多元统计分析

(第三版)

朱建平 主编



科学出版社

北京

## 内 容 简 介

本书为“十二五”普通高等教育本科国家级规划教材,同时也是教育部高等学校统计学类专业教学指导委员会推荐教材。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。主要内容包括: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.—3版.—北京:科学出版社,2016.1

经济与管理类统计学系列教材“十二五”普通高等教育本科国家级规划教材  
ISBN 978-7-03-047094-2

I. ①应... II. ①朱... III. ①多元分析-统计分析-高等学校-教材 IV.  
①0212.4

中国版本图书馆CIP数据核字(2016)第006073号

责任编辑:兰 鹏/责任校对:胡小洁

责任印制:霍 兵/责任设计:蓝正设计

**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>

三河市骏杰印刷有限公司印刷

科学出版社发行 各地新华书店经销

\*

2006年8月第 一 版 开本:787×1092 1/16

2012年6月第 二 版 印张:15 1/2

2016年1月第 三 版 2016年1月第十七次印刷

字数:367 000

**定价:36.00元**

(如有印装质量问题,我社负责调换)



# 全国经济与管理类统计学系列教材编委会

## 编委会主任:

曾五一 教育部高等学校统计学类专业教学指导委员会主任委员、厦门大学教授、博导

## 编委: (以姓氏笔画排序)

王振龙 全国统计职业教育教学指导委员会副主任委员、陕西广播电视大学教授

王艳明 教育部高等学校统计学类专业教学指导委员会委员、山东工商学院教授

刘建平 暨南大学教授、博导

刘洪 中南财经政法大学教授、博导

朱建平 教育部高等学校统计学类专业教学指导委员会委员、厦门大学教授、博导

许鹏 湖南大学教授、博导

张润楚 南开大学教授、博导

李宝瑜 山西财经大学教授、博导

李金昌 教育部高等学校经济学类专业教学指导委员会委员、浙江工商大学教授、博导

杨灿 厦门大学教授、博导

肖红叶 国家级教学名师、天津财经大学教授、博导

周恒彤 天津财经大学教授、博导

庞皓 国家级教学名师、西南财经大学教授、博导

杭斌 山西财经大学教授、博导

罗良清 教育部高等学校经济学类专业教学指导委员会委员、江西财经大学教授、博导

茆诗松 华东师范大学教授、博导

郑明 教育部高等学校统计学类专业教学指导委员会委员、复旦大学教授、博导

徐国祥 教育部高等学校统计学类专业教学指导委员会副主任委员、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。

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，我们做了必要的协调和适当的分工，尽可能

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Power Point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



# 第三版前言

《应用多元统计分析》自 2006 年 8 月由科学出版社出版发行以来,已经印刷 17 次,被许多高校采用,受到了广大教师和学生的普遍欢迎,并被遴选为“普通高等教育‘十一五’和‘十二五’国家级规划教材”,为我国经济统计专业教学改革和创新做出了贡献,同时也得到了社会的认可。

这次修订按照“十二五”国家级规划教材的要求,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。本书仍然保持了原教材的基本框架和内容体系,但是对于各章的案例分析和习题进行了较大的变动,同时对第十章多维标度法和第十一章多变量的可视化分析进行了完善修改,力求体现以下特点:

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。通过案例分析进一步加强统计理论和方法的应用,注重训练学生解决实际问题的能力,提高综合分析问题的素质。

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和方法,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的效率,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情况进行选讲。

本次修订编写工作由朱建平教授、朱平辉教授、刘云霞副教授、辛华副教授和任晓萍助理教授共同完成,并由朱建平教授担任主编统稿,刘云霞和辛华副教授协助完成总纂。

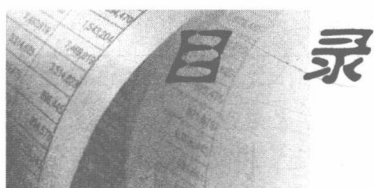
本书在修订编写和出版过程中,得到了厦门大学数据挖掘研究中心、厦门大学管理学院 MBA 中心、东北石油大学数学与统计学院、浙江工商大学现代商贸流通体系协同

创新中心和科学出版社的支持，兰鹏同志为本书的组稿、编辑做了大量的工作，在此一并表示衷心感谢！尽管我们在修订编写本书的过程中尽心竭力，书中难免有疏漏或不妥之处，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社会科学基金重大项目《大数据与统计学理论的发展研究》（13&2D148）的资助。

编 者

2016年1月于厦门



<b>第一章</b>	
	<b>多元分析概述</b> ..... 1
第一节	引言..... 1
第二节	大数据时代的多元统计分析..... 2
第三节	应用背景..... 3
第四节	计算机在统计分析中的应用..... 7
	思考与练习..... 8
<b>第二章</b>	
	<b>多元正态分布的参数估计</b> ..... 9
第一节	引言..... 9
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..... 28
<b>第三章</b>	
	<b>多元正态分布均值向量和协差阵的检验</b> ..... 30
第一节	引言..... 30
第二节	均值向量的检验..... 30
第三节	协差阵的检验..... 38
第四节	实例分析与计算机实现..... 40
	思考与练习..... 48



## 第四章

## 判别分析..... 52

- 第一节 引言..... 52
- 第二节 距离判别法..... 52
- 第三节 贝叶斯判别法..... 57
- 第四节 费希尔判别法..... 59
- 第五节 实例分析与计算机实现..... 63
- 思考与练习..... 72

## 第五章

## 聚类分析..... 74

- 第一节 引言..... 74
- 第二节 相似性的量度..... 74
- 第三节 系统聚类分析法..... 77
- 第四节  $K$  均值聚类分析..... 86
- 第五节 有序样品的聚类分析法..... 88
- 第六节 实例分析与计算机实现..... 92
- 思考与练习..... 106

## 第六章

## 主成分分析..... 109

- 第一节 引言..... 109
- 第二节 主成分的几何意义及数学推导..... 109
- 第三节 主成分的性质..... 113
- 第四节 主成分方法应用中应注意的问题..... 114
- 第五节 实例分析与计算机实现..... 116
- 思考与练习..... 124

## 第七章

## 因子分析..... 126

- 第一节 引言..... 126
- 第二节 因子分析模型..... 126
- 第三节 因子载荷矩阵求解..... 130
- 第四节 公因子重要性的分析..... 133
- 第五节 实例分析与计算机实现..... 137

思考与练习	147
-------	-----

## 第八章

<b>相应分析</b>	150
第一节 引言	150
第二节 列联表	150
第三节 相应分析的基本理论	153
第四节 相应分析中应注意的问题	155
第五节 实例分析与计算机实现	156
思考与练习	166

## 第九章

<b>典型相关分析</b>	169
第一节 引言	169
第二节 典型相关的基本理论	169
第三节 样本典型相关分析	174
第四节 典型相关分析应用中的几个问题	179
第五节 实例分析与计算机实现	183
思考与练习	190

## 第十章

<b>多维标度法</b>	193
第一节 引言	193
第二节 古典多维标度法	194
第三节 多维标度法中的几个问题	197
第四节 实例分析与计算机实现	198
思考与练习	205

## 第十一章

<b>多变量的可视化分析</b>	207
第一节 引言	207
第二节 条形图	208
第三节 面积图	212
第四节 高低图	214
第五节 雷达图分析法	216

第六节 星座图分析法·····	220
思考与练习·····	222

参考文献·····	223
-----------	-----

## 附录

常用统计表·····	224
附表 1 正态分布概率表·····	224
附表 2 $t$ 分布临界值表·····	227
附表 3 $\chi^2$ 分布临界值表·····	229
附表 4 $F$ 分布临界值表·····	230





# 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。由于计算机处理技术发生着日新月异的变化,人们处理大规模复杂数据能力日益增强;从大规模数据中提取有价值的信息能力日益提高,人们将会迅速进入大数据时代。大数据时代不仅会带来人类自然科学技术和人文社会科学的发展变革,还会给人们的生活和工作方式带来焕然一新的变化。大数据时代的到来,给多元统计分析理论的发展和方法的应用带来了发展壮大机会的同时,也使其面临着重大的挑战。

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, 许宝騷等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。20 世纪 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。20 世纪 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。20 世纪 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。20 世纪 70 年代初期多元统计分析才在我国受到各个领域的极大关注,并在理论研究和应用上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。20 世纪 80 年代初期数据在不同信息管理系统之间的共享使数据接口的标准化越来越得到强调,为数据的共享和交流提供了捷径,80 年代后期,互联网概念的兴起、“普适计算”(Ubiquitous Computing)理论的实现以及传感器对信息自动采集、传递和计算成为现实,为数据爆炸式增长提供了平台,为多元统计理论和方法的应用开辟了新的领域。20 世纪 90 年代,由于数据驱动,数据呈指数增长,企业界和学术界也不断对此现象及其意义进行探讨,为大数据概念的广泛传播提供了途径。进入 21 世纪以来,世界上许多国家开始关注大数据的发展和应用,一些学者和专家发起了关于大数据研究和应用的深入探讨,如 M.S. Viktor 和 C. Kenneth 所著的《大数据时代》等,对大数据促进人们生活、工作与思维的变革奠定了基础。在此期间,多

元统计与人工智能和数据库技术相结合,将通过互联网和物联网在经济、商业、金融、天文等行业得到更广泛的应用。

为了让读者更好地较为系统地掌握多元统计分析与方法,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 大数据时代的多元统计分析

大数据是信息科技高速发展的产物,如果利用多元统计分析与理论方法处理大数据问题,必须要全面深入理解大数据的概念,必须理解大数据产生的时代背景,然后根据大数据时代背景理解大数据概念。

### 一、“大数据时代”背景介绍

格雷布林克(Grobelink. M)在《纽约时报》2012年2月的一篇专栏中称,“大数据时代”已经降临,在商业、经济及其他领域中,管理者决策越来越依靠数据分析,而不是依靠经验和直觉。“大数据”概念之所以被炒得如火如荼,是因为大数据时代已经到来。

如果说19世纪以蒸汽机为主导的产业革命时代终结了传统的以手工劳动为主的生产方式,并从而推动了人类社会生产力的变革;那么20世纪以计算机为主导的技术革命则方便了人们的生活,并推动人类生活方式发生翻天覆地的变化。我们认为,随着计算机互联网、移动互联网、物联网、车联网的大众化和博客、论坛、微信等网络交流方式的日益红火,数据资料的增长正发生着“秒新分异”的变化,大数据时代已经到来毋庸置疑。据不完全统计,一天之中,互联网产生的全部数据可以刻满1.68亿张DVD。国际数据公司(IDC)的研究结果表明,2008年全球产生的数据量为0.49ZB(1024EB=1ZB, 1024PB=1EB, 1024TB=1PB, 1024GB=1TB),2009年的数据量为0.8ZB,2010年增长为1.2ZB,2011年的数量高达1.82ZB,相当于全球每人产生200GB以上的数据,而到2012年,人类生产的所有印刷材料的数据量是200PB,全人类历史上所有语言资料积累的数据量大约是5EB。哈佛大学社会学教授加里·金说,“大数据这是一场革命,庞大的数据资源使得各个领域开始了量化进程,无论学术界、商界还是政府,所有领域都将开始这种进程”。在大数据时代,因为等同于数据的知识随处可寻,对数据的处理和分析才显得难能可贵,所以在大数据时代,如何能从纷繁芜杂的数据中提取有价值的知识是多元统计分析方法面对的首要问题。

### 二、大数据对多元统计分析的要求

在大数据时代,数据引领人们生活,引导商业变革和技术创新。从大数据的时代背景来看,我们可以把大数据作为研究对象,从数据本身和处理数据的技术两个思路理解

大数据，这样理解大数据就有狭义和广义之分：狭义的大数据是指数据的结构形式和规模，是从数据的字面意义理解；广义的大数据不仅包括数据的结构形式和数据的规模，还包括处理数据的技术。

狭义角度的大数据，是指计量起始单位至少是 PB，EB 或 ZB 的数据规模，其不仅包括结构化数据，还包括半结构化数据和非结构化数据。我们应该从横向和纵向两个维度解读大数据：横向是指数据规模，从这个角度来讲，大数据等同于海量数据，指大数据包含的数据规模巨大；纵向是指数据结构形式，从这个角度来说，大数据不仅包含结构化数据，更多的是指半结构化的数据和非结构化数据，指大数据包含的数据形式多样。广义角度的大数据，不仅包含大数据结构形式和规模，还泛指大数据的处理技术。大数据的处理技术是指能够从不断更新增长、有价值信息转瞬即逝的大数据中抓取有价值信息的能力。在大数据时代，传统针对小数据处理的技术可能不再适用。这样，就产生了专门针对大数据的处理技术，大数据的处理技术也衍生为大数据的代名词。不管从广义的角度，还是从狭义的角度来看，大数据的核心是数据，而数据是统计研究的对象，从大数据中寻找有价值的信息关键在于对数据进行正确的统计分析。因此，鉴定“大数据”应该在现有数据处理技术水平的基础上引入统计学的思想。

从统计学科与计算机学科性质出发，我们可以这样来定义“大数据”：大数据指那些超过传统数据系统处理能力、超越经典统计思想研究范围、不借用网络无法用主流软件工具及技术进行单机分析的复杂数据的集合，对于这一数据集合，在一定的条件下和合理的时间内，我们可以通过现代计算机技术和创新统计方法，有目的地进行设计、获取、管理、分析，揭示隐藏在其中的有价值的模式和知识。

毫无疑问，由于计算机处理技术日新月异，人们能处理大规模复杂数据能力日益增强，从大规模数据中提取有价值的信息能力日益提高，人们将会迅速进入大数据时代。大数据时代，不仅会带来人类自然科学技术和人文社会科学的发展变革，同时也对数据的处理方法提出了更高的要求。

统计学是一门古老的学科，是处理数据的重要方法之一，已经有三百多年的历史，在自然科学和人文社会科学的发展中起到了举足轻重的作用；统计学又是一门生命力极其旺盛的学科，它海纳百川又博采众长，随着各门具体学科的发展不断壮大自己。毫不例外，大数据时代的到来，给统计学科带来了发展壮大机会的同时，也使得统计学科面临着重大的挑战。怎样深刻地认识和把握这一发展契机，怎样更好地理解 and 应对这一重大挑战，这就迫使我们在对多变量统计分析的理论和方法进行学习和研究的基础上，重新审视并提出适合现代数据分析的思想、理念与方法。

### ■ 第三节 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

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

300 多年前，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成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，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及描述更全面、更接近于实际了。他在广泛应用拉普拉斯等概率论中的正态曲线、误差法则、大数法则等成果的过程中，为统计学增添了数理统计方法，进而又扩展了统计学的应用范围。

在应用中对发展统计方法贡献显著的当推生物统计学派的戈尔登(1822—1921)、皮尔逊(1857—1936)和农业实验学派的孟德尔(1822—1884)、戈塞特(1876—1937)等。戈尔登六年中测量了近万人的“身高、体重、阔度、呼吸力、拉力和压力、手击的速率、听力、视力、色觉及个人的其他资料”。在探究这些数据内在联系的过程中提出了今天在自然科学和社会科学领域中广泛应用的“相关”思想。将大量数据加以综合描述和比较，从而能使他的遗传理论建立在比较精确的基础上，为统计学引入了中位数、四分位数、分布、回归等极为重要的概念和方法。皮尔逊在检验他老师戈尔登的“祖先遗传法则”和自然选择中“淘汰”对器官的相关及变异的影响中，导入了复相关的概念和方法。在讨论生物退化、反祖、遗传、随机交配等问题中，展开了回归与相关的研究，并提出以 $\chi^2$ 检验作为曲线配合适合度的一种量度的思想。

农业实验学派的孟德尔和戈塞特同样是在实验回答各自应用领域中出现的新要求、新课题，发展了统计思想和统计分析方法。孟德尔及其后继者贝特森等创建的遗传试验手段，比通过记录生命外部联系曲折反映事物内在本质的描述统计更加深刻。他们运用推断的理论与实验的方法，通常只用小样本来处理。戈塞特的  $t$  分布与小样本思想更是在由于“有些实验不能多次地进行”，从而“必须根据极少数的事例(小样本)来判断实验结果的正确性”的情况下产生的。今天，这些统计思想和分析推断方法已经成为了科学家不可缺少的基本研究工具。

近现代，统计学已经空前广泛应用于最高级的运动形式——社会。其结果便是出现了一系列与其应用对象指导理论和其他相关学科交织在一起的边缘学科。如在社会经济方面的投入产出经济学、经济计量学、统计预测学、统计决策学等。在这些边缘学科中，统计学与其应用对象结合更紧密、更自然。这些学科的专家学者至少在两个或两个以上