国家“双一流”建设学科
辽宁大学应用经济学系列丛书
教材系列
财政部“十三五”规划教材

总主编◎林木西

多变量分析及R的应用

Multivariable Analysis and the Application of R

王 青 编著

中国财经出版传媒集团
经济科学出版社
Economic Science Press

辽宁大学应用经济学系列丛书·教材系列

2016年省级本科教改立项一般项目：以学生为中心、以解决实际问题为导向的《经济数学》教学模式改革研究与实践

多变量分析及R的应用

Multivariable Analysis and the Application of R

王 青 编著

中国财经出版传媒集团
经济科学出版社
Economic Science Press

图书在版编目（CIP）数据

多变量分析及R的应用/王青编著. —北京：
经济科学出版社，2018.1
（辽宁大学应用经济学系列丛书. 教材系列）
ISBN 978-7-5141-8982-7

Ⅰ.①多… Ⅱ.①王… Ⅲ.①统计分析-
应用软件-研究 Ⅳ.①C819

中国版本图书馆CIP数据核字（2018）第009589号

责任编辑：于海汛 刘战兵
责任校对：隗立娜
责任印制：李 鹏

多变量分析及R的应用
王 青 编著
经济科学出版社出版、发行 新华书店经销
社址：北京市海淀区阜成路甲28号 邮编：100142
总编部电话：010-88191217 发行部电话：010-88191522
网址：www.esp.com.cn
电子邮件：esp@esp.com.cn
天猫网店：经济科学出版社旗舰店
网址：http://jjkxcbs.tmall.com
北京季蜂印刷有限公司印装
710×1000 16开 15.25印张 230000字
2018年3月第1版 2018年3月第1次印刷
ISBN 978-7-5141-8982-7 定价：46.00元

总　序

本丛书为国家“双一流”建设学科辽宁大学应用经济学系列丛书，也是我主编的第三套系列丛书。前两套丛书出版后，总体看效果还可以：第一套是《国民经济学系列丛书》（2005年至今已出版13部），2011年被列入“十二五”国家重点图书出版物；第二套是《东北老工业基地全面振兴系列丛书》（共10部），在列入“十二五”国家重点图书出版物的同时，还被确定为2011年“十二五”规划400种精品项目（社科与人文科学155种），围绕这两套系列丛书还取得了一系列成果，获得了一些奖项。

主编系列丛书从某种意义上说是“打造概念”。比如说第一套系列丛书也是全国第一套国民经济学系列丛书，主要为辽宁大学国民经济学国家重点学科“树立形象”；第二套则是在辽宁大学连续获得国家社科基金“八五”“九五”“十五”“十一五”重大（点）项目，围绕东北（辽宁）老工业基地调整改造和全面振兴进行系统研究和滚动研究的基础上持续进行探索的结果，从而为促进我校区域经济学建设、服务地方经济不断做出新贡献。在这一过程中，既出成果也带队伍、建平台、组团队，遂使我校应用经济学学科建设不断地跃上新台阶。

主编第三套丛书旨在使辽宁大学的应用经济学一级学科建设有一个更大的发展。辽宁大学应用经济学学科的历史说长不长、说短不短。早在1958年建校伊始，便设经济系、财政系、计统系等9个系，其中经济系由原东北财经学院的工业经济、农业经济、贸易经济三系合成，财税系和计统系即原东北财经学院的财信系、计统系。后来院系调整，将

经济系留在沈阳的辽宁大学，将财政系、计统系迁到大连组建辽宁财经学院（即现东北财经大学前身），对工业经济、农业经济、贸易经济三个专业的学生培养到毕业为止。由此形成了辽宁大学重点发展理论经济学（主要是政治经济学）、辽宁财经学院重点发展应用经济学的大体格局。实际上，后来辽宁大学也发展应用经济学，东北财经大学也发展理论经济学，发展得都不错。1978 年，辽宁大学恢复招收工业经济本科生，1980 年受人民银行总行委托、经教育部批准招收国际金融本科生，1984 年辽宁大学在全国第一批成立经济管理学院，增设计划统计、会计、保险、投资经济、国际贸易等本科专业。到 20 世纪 90 年代中期，已有西方经济学、世界经济、国民经济管理、国际金融、工业经济 5 个二级学科博士点，当时在全国同类院校似不多见。1998 年建立国家重点教学基地“辽宁大学国家经济学基础人才培养基地”，同年获批建设第二批教育部人文社科重点研究基地“辽宁大学比较经济体制研究中心”（2010 年改为“转型国家经济政治研究中心”）。2000 年，辽宁大学在理论经济学一级学科博士点评审中名列全国第一；2003 年，辽宁大学在应用经济学一级学科博士点评审中并列全国第一；2010 年，新增金融、应用统计、税务、国际商务、保险等全国首批应用经济学类专业学位硕士点；2011 年，获全国第一批统计学一级学科博士点，从而成为经济学、统计学一级学科博士点“大满贯”。

在二级学科重点学科建设方面，1984 年，外国经济思想史即后来的西方经济学、政治经济学被评为省级重点学科；1995 年，西方经济学被评为省级重点学科，国民经济管理被确定为省级重点扶持学科；1997 年，西方经济学、国际经济学、国民经济管理被评为省级重点学科和重点扶持学科；2002 年、2007 年国民经济学、世界经济学连续两届被评为国家重点学科；2007 年，金融学被评为国家重点学科。

在一级学科重点学科建设方面，2017 年 9 月，被教育部、财政部、国家发展改革委确定为“双一流”建设学科。辽宁大学确定的世界一流学科为“应用经济学”，其建设口径范围为“经济学学科群”，所对应的一级学科为应用经济学和理论经济学，遂成为东北地区唯一一个经

济学科“双一流”建设学科。这是我校继1997年成为“211工程”重点建设高校20年之后，学科建设的又一次重大跨越，也是辽宁大学经济学科三代人共同努力的结果。此前，应用经济学、理论经济学于2008年被评为第一批一级学科省级重点学科，2009年被确定为辽宁省“提升高等学校核心竞争力特色学科建设工程”高水平重点学科，2014年被确定为辽宁省一流特色学科第一层次学科，2016年被辽宁省人民政府确定为省一流学科。

在“211工程”建设方面，应用经济学一级学科在“九五”立项的重点学科建设项目是“国民经济学与城市发展”“世界经济与金融”；“十五”立项的重点学科建设项目是“辽宁城市经济”；“211工程”三期立项的重点学科建设项目是“东北老工业基地全面振兴”“金融可持续协调发展理论与政策”，基本上是围绕国家重点学科和省级重点学科而展开的。

经过多年的学科积淀与发展，辽宁大学应用经济学、理论经济学、统计学“三箭齐发”，国民经济学、金融学、世界经济三个国家重点学科“率先突破”，由长江学者特聘教授、“万人计划”领军人才、全国高校首届国家级教学名师领衔，中青年学术骨干梯次跟进，形成了一大批高水平的学术成果，培养出一批又一批优秀人才，多次获得国家级科研、教学奖励，在服务东北老工业基地全面振兴等方面做出了积极的贡献。

这套《辽宁大学应用经济学系列丛书》的编写，主要有三个目的：

一是促进“经济学学科群”一流学科全面发展。以往辽宁大学主要依托国民经济学、世界经济学和金融学三个国家重点学科和省级重点学科进行建设，取得了重要进展。这个“特色发展”的总体思路无疑是正确的。进入“十三五”时期，根据“双一流”学科建设的需要，本学科确定了区域经济学、产业经济学与东北振兴，世界经济、国际贸易学与东北亚合作，国民经济学与地方政府创新，金融学、财政学与区域发展，政治经济学与理论创新等五个学科方向。到2020年，努力将本学科建成为立足于东北经济社会发展，为东北振兴和东北亚合作做出

应有贡献的一流学科。因此，本套丛书旨在为实现这一目标提供更大的平台支持。

二是加快培养中青年骨干教师茁壮成长。目前，本学科已建成长江学者特聘教授、“万人计划”领军人才、全国高校首届国家级教学名师领衔，教育部21世纪优秀人才、教育部教指委委员、省级教学名师、校级中青年骨干教师为中坚，以老带新、新老交替的学术梯队。本丛书设学术、青年学者、教材三个子系列，重点出版中青年教师的学术著作，带动他们尽快脱颖而出，力争早日担纲学科建设。与此同时，还设立了教材系列，促进教学与科研齐头并进。

三是在经济新常态、新一轮东北老工业基地全面振兴中做出更大贡献。对新形势、新任务、新考验，提供更多具有原创性的科研成果，具有较大影响的教学改革成果，具有更高决策咨询价值的“智库”成果。

这套系列丛书的出版，得到了辽宁大学党委书记周浩波教授、校长潘一山教授和中国财经出版传媒集团副总经理、经济科学出版社社长吕萍的支持。在丛书出版之际，谨向所有关心支持辽宁大学应用经济学建设和发展的各界朋友，向辛勤付出的学科团队成员表示衷心感谢！

林木西

2017年国庆节于蕙星楼

前　言

多变量分析是近年来发展迅速的统计分析方法之一，广泛应用于自然科学、管理科学和社会、经济等各个领域。本书将在深入浅出地讲解多变量分析方法原理的基础上，侧重于结合实例介绍多变量分析方法的应用。在方法的具体实现上，本书采用了国内外广泛使用的统计软件R，详细介绍了多变量分析方法在R中的实现以及输出结果的解读。

全书共有九章，基本覆盖了常用的多变量分析方法。第一章是绪论，是为指导全书的学习而编排的。第二章是多变量数据描述统计分析Ⅰ：表格法和图形法，介绍了利用各种图形或表格来对数据进行描述性统计分析。第三章是多变量数据描述统计分析Ⅱ：数值方法，介绍了利用概括统计量来描述定量变量的数据。第四章至第九章是有关现代多变量分析的方法，内容包括多元回归分析、广义线性模型、聚类分析、判别分析、主成分分析和因子分析。

本书的适用范围很广，可以作为数学、应用数学、金融数学、统计、经济等专业本科生以及各专业硕士和博士研究生的教科书或参考书，希望本书对教师以及各个领域的实际工作者都有参考价值。

本书参阅了许多国内外教材和资料，并引用了部分例题和习题，在此向有关作者表示衷心的感谢。该书出版得到了辽宁大学长江学者林木西教授、辽宁大学经济学院统计学系同仁们的大力支持，在此表示感谢。该书出版得到了经济科学出版社的大力支持和帮助，在此表示诚挚

的谢意。

由于水平有限，书中难免有不妥之处，敬请同行专家及广大读者批评指正。

王 青

2018年1月

书中案例所需数据文件可在经济科学出版社官网（www. esp. com. cn）的“资源下载”栏目中下载。

目　录

第一章

绪　论

多变量分析是20世纪初发展起来的统计分析方法，它是通过对多个随机变量观测数据的分析来研究多个随机变量之间的相互关系并揭示变量内在规律的分析方法[①]。多变量分析方法可应用于经济学、管理学、医学、教育学、心理学、体育科学、生态学、地质学、社会学、考古学、军事科学、环境科学、文学等很多领域。

第一节　多变量分析概述

一、多变量分析的产生与发展过程

多变量分析起源于医学和心理学。1928年威谢特（Wishert）发表论文《多元正态总体样本协方差阵的精确分布》，是多变量统计分析的开端；20世纪30年代，费希尔（Fisher）、霍特林（Hotelling）、许宝碌等奠定了多变量统计分析的理论基础；20世纪40年代，这一分析方法在心理学、教育学、生物学等方面有不少应用，但由于计算复杂且计算

① 费宇等：《多元统计分析——基于R》，中国人民大学出版社2014年版。

量大，其发展受到限制；20世纪50年代中期，随着计算机的出现和发展，多变量统计分析方法在地质、气象、医学和社会学方面得到广泛应用，多变量统计分析已渗入几乎所有的学科；到20世纪80年代后期，计算机软件包已很普遍，使用也方便，因此多变量分析方法也更为普及，在我国受到各个领域的极大关注，近40多年在理论上和应用上都取得了若干新进展。

二、多变量分析的用途

多变量分析是运用数理统计方法来研究解决多变量问题的理论和方法，它是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并揭示其内在统计规律性的数理统计学分支之一。在实际应用中，多变量分析通常用于解决以下四个方面的问题[①]：

（一）变量之间的相依性分析

分析多个或多组变量之间的相依关系，是一切科学研究尤其是经济管理研究的主要内容，简单相关分析、偏相关分析、复相关分析和典型相关分析提供了进行这类研究的必要方法。

（二）构造预测模型，进行预报控制

在自然和社会科学领域的科研与生产中，探索多元系统运行的客观规律及其与外部环境的关系，进行预测预报，以实现对系统的最优控制，是应用多变量分析技术的主要目的。在多变量分析中，用于预报控制的模型有两类：一类是预测预报模型，通常采用多元回归或逐步回归分析、非线性回归、判别分析等建模技术；另一类是描述性模型，通常采用综合评价的分析技术。

---

① 王斌会：《多元统计分析及R语言建模》，暨南大学出版社2016年版。

（三）进行数值分类，构造分类模型

在多变量分析中，往往需要将系统性质相似的事物或现象归为一类，以便找出它们之间的联系和内在规律。过去许多研究是按单因素进行定性处理，以致处理结果反映不出系统的总特征。进行数值分类，构造分类模式一般采用聚类分析和判别分析技术。

（四）简化系统结构

可采用主成分分析、因子分析、对应分析等方法，在众多因素中找出各个变量中最佳的子集合，根据子集合所包含的信息描述多元系统的结果及各个因子对系统的影响。

如何选择适当的方法来解决实际问题，需要对问题进行综合考虑。对一个问题可以综合运用多种统计方法进行分析。

三、多变量分析所包含的内容

在对社会、经济、技术系统的认识过程中，需要收集和分析大量表现系统特征和运行状态的数据信息。这类原始数据集合往往由于样本点数量巨大，用于刻画系统特征的指标变量众多，并且带有动态特性，从而形成规模宏大、复杂难辨的数据海洋。那么，我们应该如何分析和认识高维复杂数据集合中的内在规律性，简明扼要地把握系统的本质特征？如何对高维数据集合进行最佳综合，迅速将隐藏在其中的重要信息集中提取出来？如何充分发掘数据中的丰富内涵，清晰地展示系统结构，准确地认识系统元素的内在联系，以及直观地描绘系统的运动历程？利用统计学和数学方法对多维复杂数据集合进行科学分析的理论和方法，正是多变量分析研究的基本内容。

其主要范畴包括多元数据图表示法、多元相关与回归分析、聚类分析、判别分析、主成分分析、因子分析等。

第二节　多变量分析中的基本概念

一、数据的来源与类型

数据是统计工作所搜集、分析、汇总表述和解释的事实及数字。统计数据不是指单个的数字，而是所搜集的有关资料的数据集。

（一）数据的来源

从统计数据本身的来源看，统计数据最初都来源于直接的调查或试验。从使用者的角度看，统计数据主要来源于两种渠道：一是直接的调查和科学试验，这是统计数据的直接来源，我们称之为第一手或直接的统计数据；二是别人调查或试验的数据，这是统计数据的间接来源，我们称之为第二手或间接的统计数据。第二手数据主要是公开出版的或公开报道的数据，也有些是尚未公开的数据。在我国，公开出版或报道的社会经济统计数据主要来自国家和地方的统计部门以及各种报刊媒介，如公开的出版物有《中国统计年鉴》、《中国统计摘要》、《中国社会统计年鉴》、《中国工业经济统计年鉴》、《中国农村统计年鉴》、《中国人口统计年鉴》、《中国市场统计年鉴》，以及各省、市、地区的统计年鉴等。提供世界各国社会和经济数据的出版物也有许多，如《世界经济年鉴》、《国外经济统计资料》，以及世界银行各年度的《世界发展报告》等。联合国有关部门及世界各国也定期出版各种统计数据。除了公开出版的统计数据外，还可以通过其他渠道使用一些尚未公开的统计数据，以及广泛分布在各种报纸、杂志、图书、广播、电视传媒中的数据资料。现在，随着计算机网络技术的发展，也可以在网络上获取所需的各种数据资料。

（二）数据的类型

按照数据的计量尺度划分，数据可分为定类数据、定序数据、定距数据和定比数据；按照数据反映的内容划分，数据可分为数量数据与品质数据；按照数据表现形式划分，数据可分为时间数列数据、截面数据和合并（混合）数据。

数据的计量尺度有四种：定类（名义）尺度（nominal scale）是只按照事物的某种属性对其进行平行分类或分组所进行的测度，是最粗略、计量层次最低的计量尺度。如人口按照性别分为男、女两类。定序（顺序）尺度（ordinal scale）又称顺序尺度，是对事物之间等级差或顺序差别的一种测度。如将产品等级分为一等品、二等品、三等品及次品等。定距（间隔）尺度（interval scale）也称为间隔尺度，是对事物类别或次序之间间隔的测度，通常使用自然或度量衡单位作为计量尺度。如考试成绩用百分制度量、温度用摄氏度或华氏度来度量等。定距尺度的计量结果表现为数量。定比（比率）尺度（ratio scale）也称为比率尺度，它与定距尺度属于同一层次，一般可不进行区分，其计量结果也表现为数值，但其特性是可以计算两个测度值之间的比值。定距尺度与定比尺度之间的唯一差别是定比尺度有一个绝对固定的“零点”，而定距尺度中没有绝对的零点，即定距尺度计量值可以为0，“0”表示一个数值，即“0”水平，而不表示“没有”或“不存在”。如温度为0℃，表示温度的水平，并不表示没有温度。所以定距尺度中的0是一个有意义的数值。定比尺度则不同，它有一个绝对“零点”，也就是说，在定比尺度中，“0”表示“没有”或“不存在”，如产量为0，表示没有这种产品；收入为0，表示这个人没有收入。现实生活中大多数情况下使用的都是定比尺度。统计数据采用不同的计量尺度也就形成了不同的数据，即定类数据、定序数据、定距数据和定比数据。

数据可以既包括品质（定性）数据又包括数量（定量）数据两方面。定类数据和定序数据统称为品质（定性）数据；定距数据和定比数据统称为定量数据。定性数据是为了对事物进行分类而提供标签或名