顾问 王静龙 艾春荣 徐国祥 周 勇

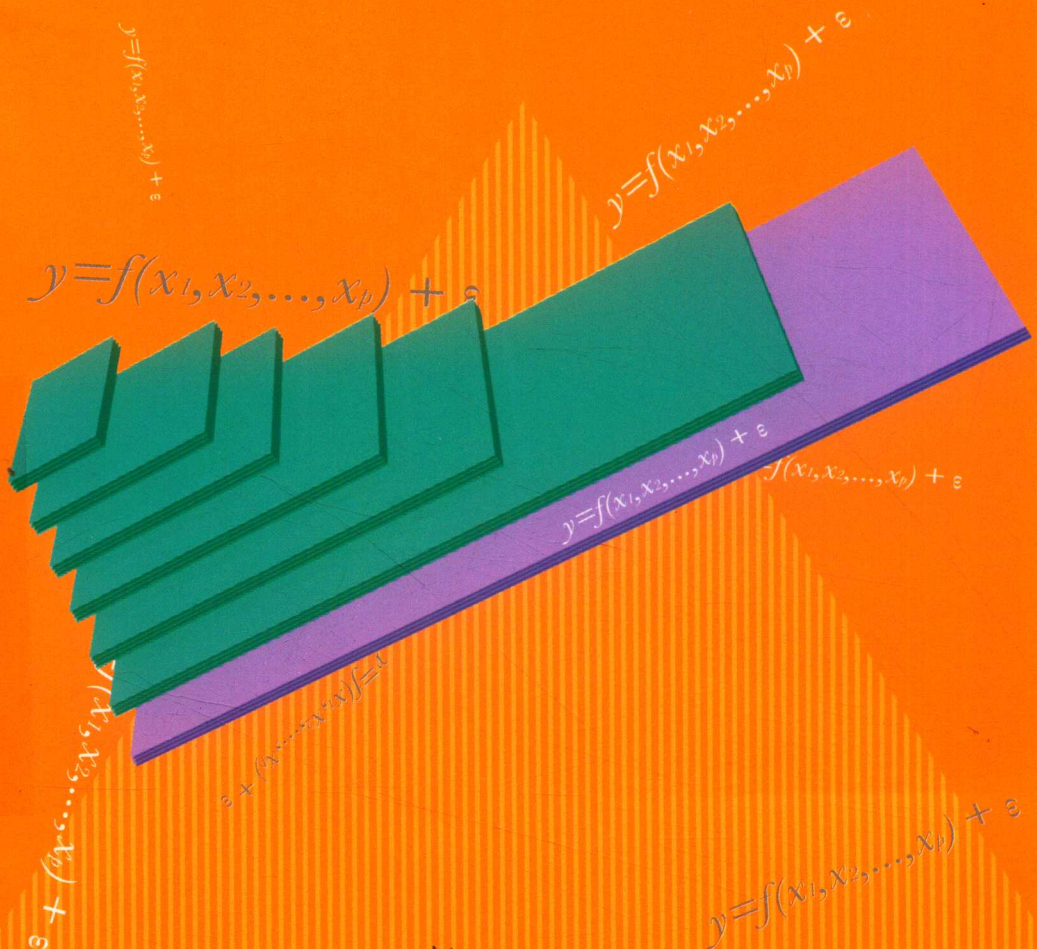
21世纪

高校统计学专业教材系列
上海市教委重点课程建设项目
上海财经大学精品课程

应用回归分析

王黎明 陈 颖 杨 楠 编著

（第二版）



非外借

复旦大学出版社

博学 21世纪

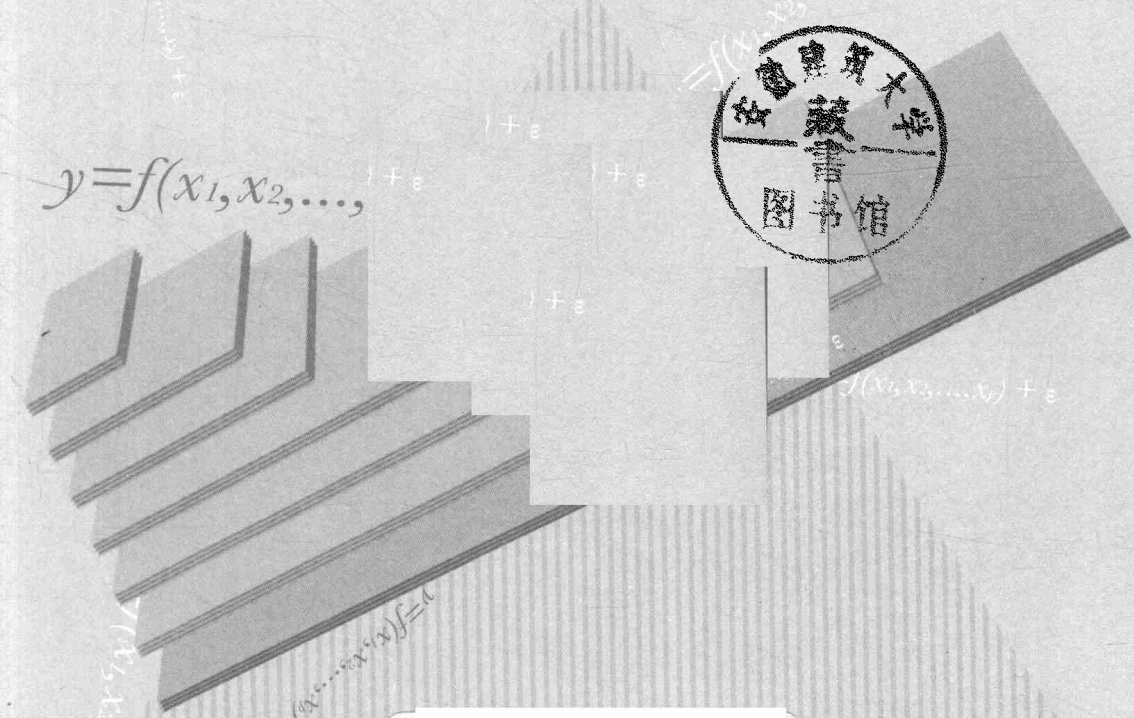
顾问 王静龙 艾春荣 徐国祥 周 勇

高校统计学专业教材系列
上海市教委重点课程建设项目
上海财经大学精品课程

应用回归分析

王黎明 陈 颖 杨 楠 编著

（第二版）



复旦大学出版社

图书在版编目(CIP)数据

应用回归分析/王黎明,陈颖,杨楠编著. —2 版. —上海: 复旦大学出版社,2018.6
(复旦博学)
21 世纪高校统计学专业教材系列　上海市教委重点课程建设项目　上海财经大学精品课程
ISBN 978-7-309-13733-0

Ⅰ. 应…　Ⅱ. ①王…②陈…③杨…　Ⅲ. 回归分析-高等学校-教材　Ⅳ. 0212.1

中国版本图书馆 CIP 数据核字(2018)第 107749 号

应用回归分析(第 2 版)
王黎明　陈　颖　杨　楠　编著
责任编辑/王联合

复旦大学出版社有限公司出版发行
上海市国权路 579 号　邮编: 200433
网址: fupnet@ fudanpress. com　http://www. fudanpress. com
门市零售: 86-21-65642857　团体订购: 86-21-65118853
外埠邮购: 86-21-65109143　出版部电话: 86-21-65642845
上海同济印刷厂有限公司

开本 787×960　1/16　印张 19.5　字数 332 千
2018 年 6 月第 2 版第 1 次印刷

ISBN 978-7-309-13733-0/O·660
定价: 39.00 元

“博学而笃志，切问而近思。”

（《论语》）

博晓古今，可立一家之说；

学贯中西，或成经国之才。

内容提要

本书本以经典的最小二乘理论为基础，较全面地介绍了现代应用回归分析的基本理论和主要方法。全书共分为九章。第一章讨论了回归模型的主要任务和回归模型的建模过程；第二、三章详细地介绍了线性回归模型；第四章以残差为重要工具，讨论了回归模型的诊断问题；第五、六章讨论了多项式回归模型和含有定性变量的回归模型；第七章讨论了多元线性回归模型的有偏估计；第八章简单介绍了非线性回归模型；本书的最后一章简明介绍了SAS统计软件在回归分析中的应用。

本书可以作为统计学、数学以及经济学等专业的教材，学习本课程的学生需要熟悉概率论与数理统计的基础知识，也要具备微积分和线性代数知识。

Foreword

第二版前言

《应用回归分析》自2008年出版以来，得到了广大同行的肯定，国内数十所大学选择本书作为教材，在上海财经大学统计与管理学院本科生教学中使用了近十年。在这期间，我们团队不断改进相关内容，充实有关资料，使得本课程陆续建设完成上海市教委重点课程建设项目和上海财经大学精品课程。但是，由于作者的疏忽，本书第一版中有一些印刷错误，我们将在修订版中全部更正。另外，在使用过程中，我们感觉有些部分还需要加强。在此次修订版中进行的一些增加，具体内容如下：

第一，在第一章增加了资本资产定价模型CAPM理论作为一个例子，在关于回归名称的介绍中增加了散点图。

第二，为了理解回归模型，在第二章增加了有关的图和不同模型的描述。

第三，为了对异方差问题有更加直观的理解，在第四章增加了一个实际数据的例子。

第四，在第七章对均方误差概念给出了比较系统的说明，并对相关结论做了严格的证明。

本书修订部分都是由王黎明教授完成的，杨楠教授在教学过程中指出了很多印刷错误。上海财经大学统计与管理学院历届的本科生也对本书提出了中

肯的建议，尤其是张芮同学，他在去中国人民大学攻读研究生前，把他学习的体会发给了我们，使得我们在修订中受到很大启发，在此一并表示感谢。由于编者的水平有限，本书在取材及其结构上，或许还存在不够妥当的地方，恳请同行专家和广大读者给我们提出宝贵的批评和建议。

王黎明

2018 年 01 月

于上海财经大学

Foreword

前言

回归分析是统计学中一个非常重要的分支，是以概率论与数理统计为基础迅速发展起来的一种应用性较强的科学方法。它是由一组探求变量之间关系的技术组成，作为统计学应用最广泛的分支之一，在社会经济各部门以及各个学科领域都能得到广泛的应用。随着我国社会主义现代化建设的发展，人们越来越认识到应用定量分析技术研究问题的重要意义。特别是近些年来计算机及有关统计软件的日益普及，为在实际问题中进行大规模、快速、准确的回归分析运算提供了有力手段。

随着统计学在中国被确立为一级学科，统计专业的课程设置已有了较大变化，加强推断统计内容的学习和应用已成为中国统计界的共识。为了适应新的统计学学科体系和财经类统计学专业教学的需要，我们决定编写一套适应新时期需要的系列教材——复旦博学·21世纪高校统计学专业教材。作为系列教材之一，应用回归分析是其中较为重要的一本教材。本书写作的指导思想是：既要保持较为严谨的统计理论体系，又要努力突出实际案例的应用和统计思想的渗透，结合统计软件较全面系统地介绍回归分析的实用方法。为了贯彻这一指导思想，本书将系统介绍回归分析基本理论和方法，在理论上，本书叙述了经典的最小二乘理论，又结合应用中出现的一些问题给出对最小二乘估计的改进方法。中心主题是建立线性回归模型，评价拟合效果，并且作出结论。与此同时，本书也尽力结合中国社会、经济、自然科学等领域的研究实例，把回归分

析方法与实际应用结合起来，注意定性分析与定量分析的紧密结合，努力把同行们以及我们在实践中应用回归分析的经验和体会融入其中。全书分为九章。第一章介绍了一般回归模型的定义，讨论了回归模型的主要任务和回归模型的建模过程。第二章详细地介绍了一元线性回归模型，给出了未知参数的最小二乘估计以及极大似然估计，还讨论了一元线性回归模型的预测问题以及数据变换问题。第三章系统讨论了多元线性回归模型，详细地讨论了最小二乘估计的优良性。对于假设检验，讨论了多元回归模型的显著性检验，以及其回归系数的显著性检验。第四章以残差为重要工具，讨论了回归模型的诊断问题。第五章和第六章讨论了多项式回归模型和含有定性变量的回归模型。第七章讨论了多元线性回归模型的有偏估计。重点介绍较常用的岭估计和主成分估计，也介绍了其他的估计方法。第八章简单介绍了非线性回归模型，主要讨论了Logistic回归模型、Poisson回归和广义线性模型。本书的最后一章介绍了SAS统计软件在回归分析中的应用。

本书可以作为统计学、数学以及经济学等专业的教材。学习本课程的学生需要熟悉随机变量、参数估计、区间估计、假设检验等思想，也要熟悉正态分布及其由它导出的分布，当然，学生也要具备微积分和线性代数知识。由于本书的内容较多，教师在选用此书作教材时可以灵活选讲。本书也可以作为非统计学专业研究生回归技术的教材。根据我们多年的教学实践，本书讲授51课时较为合适，如果能有计算机和投影设备的配合，教学将会更为方便和有效。本书的写作，始终得到了复旦博学·21世纪高校统计学专业教材编委会和复旦大学出版社的支持，编写大纲和书稿都经过教材编写委员会的多次认真讨论。

本书是我们多年教学和科研工作的积累，其中的部分案例为体现其典型性引用了他人著作。在此，我们谨向对本书出版给予帮助的同行和朋友表示衷心的感谢。本书的完成也是我们多年友好合作的结果，研究生苏艳和万里同学参加了部分习题和案例的编写和整理工作，也参加了最后的统稿和校对工作。由于编者的水平有限，本书在取材及其结构上或许存在不够妥当的地方，恳请同行专家和广大读者给我们提出宝贵的批评和建议。

王黎明　陈颖　杨楠

2008年2月

于上海财经大学

Contents

目 录

第一章

回归分析概论

§1.1 变量间的相关关系

社会经济领域与自然科学等诸多现象之间始终存在着相互联系和相互制约的普遍规律。例如，社会经济的发展与一定的经济变量的数量变化密切联系，社会经济现象不仅同与它有关的现象构成一个普遍联系的整体，在其内部也存在着彼此关联的因素，在一定的社会环境等诸多条件的影响下，一些因素推动或制约另外一些与之关联的因素发生变化。也就是说，社会经济现象的内部和外部联系中存在一定的相关性，要认识和掌握客观经济规律，就必须探求经济现象间经济变量的变化规律，变量间的统计关系是经济变量变化规律的重要内容。

这些互相联系的经济现象和经济变量，其联系的紧密程度也是互不相同的。这中间极端的关系就是确定性关系，即一个变量的变化完全确定另外一个变量的变化。例如，一个保险公司承保汽车 5 万辆，每辆保费收入是 1 000 元，则该公司汽车承保总额为 5 000 万元。即承保总收入为 y，承保汽车数为 x，则变量 y 和 x 的关系可以表示为：$y=1\,000x$。

从这个例子可以看出，每给定一个 x，就一定可以得到一个 y，即变量 y 与 x 之间完全表现为一种确定性的关系——函数关系。在实际问题中，这样的例子还有很多。例如，银行的一年存款利率为年息 2.75%，存入的本金用 x 表示，到期的本息用 y

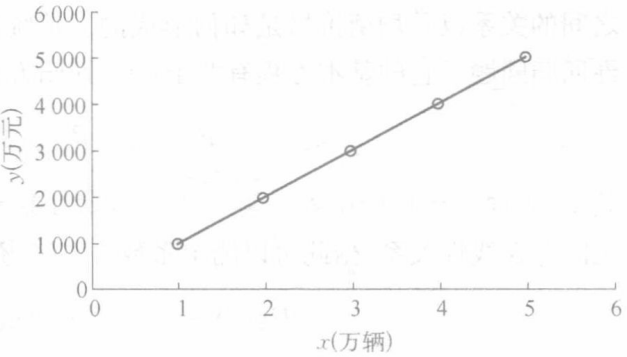


图 1.1 函数关系图

表示，则 y 与 x 有函数关系 $y=(1+0.027\,5)x$，这里的 y 与 x 仍具有线性函数关系，对于任意两个变量 y 与 x 的函数关系，可以表示为数学形式 $y=f(x)$。

一般而言，给定 p 个变量 x_1，…，x_p，就可以确定变量 y，称这种变量之间的关系为确定性关系。它往往可以用某一函数关系 $y=f(x_1,\cdots,x_p)$ 来表示。

可是，在实际问题中，变量之间存在大量非确定的关系，它们之间虽存在着密切联系，但是其密切程度不是由确切关系能够刻画的。

为此，我们再看一个例子。

根据实际生活经历，我们知道某种高档品的消费量(y)与城镇居民的收入(x)有密切关系。居民收入高了，这种消费品的销售量就大；居民收入低了，这种消费品的销售量就小。但是，居民的收入并不能完全确定该种高档品的消费量。因为商品的消费量还受到人们的消费习惯、心理因素、其他可替代商品的吸引程度以及价格的高低等因素的影响。也就是说，城镇居民的收入与该种高档品的消费量有着密切关系，且城镇居民的收入对该种高档品的消费量的大小起着主要作用，但是它并不能完全确定该种高档品的消费量。

在日常生活中，变量与变量之间表现为这种关系的有很多。例如，我们也许对银行储蓄额与居民收入、研究商品的需求量与该商品的价格、消费者的收入以及其他同类商品的价格之间的关系感兴趣，也许对研究产品的销量(如汽车)与用于产品宣传的广告费之间的关系感兴趣，也许对研究国防开支与国民生产总值(GNP)之间的关系感兴趣，也许对粮食产量与施肥量之间的关系感兴趣等。在上述各例中，或许存在某一基本理论，它规定了我们为什么期望一个变量是非独立的或说它与其他一个或几个变量有关。

在现代金融理论中，资本资产定价模型 CAPM 是在投资组合理论和资本市场理论基础上形成发展起来的，主要研究证券市场中资产的预期收益率与风险资产之间的关系以及均衡价格是如何形成的。从统计学角度来看，可以认为是一元线性回归问题。它的基本方程有两个：① 回归方程：

$$r_i=\alpha_i+\beta_i r_M+\varepsilon_i$$

其中，$E(\varepsilon_i)=0$，$\mathrm{Cov}(\varepsilon_i,r_M)=0$。假定证券 i 的收益率 r_i 与市场组合收益率 r_M 之间存在线性关系，据此可以测定系数 β_i；② 资本市场线性方程：

$$E(r_i)=r_F+\beta_i(E(r_M)-r_F)$$

它告诉我们合理的证券投资组合应选在该线上，使得风险相同的情况下能获得较