当代科学前沿论丛

NEW FRONTIERS OF SCIENCES

Logistic 回归模型
——方法与应用

王济川　郭志刚　WANG JICHUAN　GUO ZHIGANG

LOGISTIC REGRESSION MODELS: METHODS AND APPLICATION

高等教育出版社
HIGHER EDUCATION PRESS

当代科学前沿论丛

Logistic 回归模型
——方法与应用

王济川　郭志刚

高等教育出版社

内容简介

在社会科学诸如社会学、心理学、人口学、政治学、经济学以及公共卫生学当中，大量的观测因变量是二分类测量(即 $y=1$ 或 $y=0$)。本书专题介绍了在分析二分类因变量时最常使用的统计分析模型之一——logistic 回归模型。本书深入浅出，理论联系实际，通过例题分析，并结合计算机统计软件的应用，详细介绍、阐述了该模型及其应用。同时，还介绍了如何将 logistic 回归模型扩展到序次 logistic 回归模型和多项 logit 模型，以分析序次变量和多分类名义变量为因变量的数据。本书提供用 SAS 和 SPSS 进行具体例题分析的计算机程序及相关数据，并对这两种软件的模型估计结果进行详尽的解释和对比分析。本书的读者对象为社会科学各专业的教师及研究生，以及社会科学专业研究人员。

图书在版编目(CIP)数据

Logistic 回归模型——方法与应用/王济川，郭志刚.
北京：高等教育出版社，2001.9
ISBN 7-04-009910-1

Ⅰ.L… Ⅱ.①王…②郭… Ⅲ.回归分析—统计模型 Ⅳ.0212.1

中国版本图书馆 CIP 数据核定(2001)第 023720 号

Logistic 回归模型——方法与应用
王济川 郭志刚

出版发行	高等教育出版社		
社　　址	北京市东城区沙滩后街 55 号	邮政编码	100009
电　　话	010-64054588	传　　真	010-64014048
网　　址	http://www.hep.edu.cn http://www.hep.com.cn		
经　　销	新华书店北京发行所		
印　　刷	国防工业出版社印刷厂		
开　　本	787×960　1/16	版　　次	2001 年 9 月第 1 版
印　　张	18.25	印　　次	2001 年 9 月第 1 次印刷
字　　数	260 000	定　　价	38.20 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

《当代科学前沿论丛》专家委员会

（按姓氏笔画为序）

（国内部分）

王　夔　冯　端　师昌绪　曲钦岳　朱清时
孙　枢　李三立　李大潜　李国杰　杨芙清
吴建屏　邹承鲁　张尧庭　陈　竺　陈佳洱
陈希孺　陈宜瑜　周秀骥　姜伯驹　袁亚湘
钱　易　徐光宪　徐端夫　徐冠华　翟中和
戴立信　戴汝为

（海外部分）

王中林　文小刚　邓兴旺　田　刚　丛京生
刘　钧　汤　超　许　田　危　岩　严晓海
李　凯　李　明　邱子强　佘振苏　范剑青
周午纵　郑元芳　宫　鹏　俞陆平　袁钧瑛
徐希平　程正迪　鄂维南

王济川

1947年出生。1982年四川大学经济系毕业。1986年于美国康乃尔大学获社会学硕士学位，1990年获该校博士学位。1989年9月—1991年8月于美国密执安大学人口研究中心作博士后研究。现任美国俄亥俄州怀特州立大学医学院社区卫生系教授．

王济川博士的主要研究领域为社会科学定量分析方法、人口分析方法、毒品滥用及疾病预防项目的评估。

郭志刚

1954年出生。1982年于中国人民大学工业经济系获经济学学士，1985年于加拿大西安大略大学获社会学硕士，1990年于中国人民大学人口研究所获法学博士。1992年10月至1994年1月于美国布朗大学人口研究中心作博士后研究。1985年至1999年在中国人民大学人口研究所工作。现任北京大学社会学系教授。郭志刚博士的主要研究领域为人口统计技术、社会科学定量分析方法，以及人口、婚姻、家庭、老年等社会问题的分析。曾编著《社会科学研究的量化方法》、《社会统计分析方法——SPSS软件应用》，并撰写、编著、翻译多部人口学研究著作，发表了大量学术论文。

出版者的话

人类创造了科学技术，科学技术推动了人类的文明进程。两者的互动影响，今天已达到了前所未有的程度：人类的经济发展和社会进步的需要，为科学技术迅猛的创新，提供了强大的动力；科学技术的发展，在急剧地改变着人类的思维方式、学习方式、工作方式、生活方式、娱乐方式。科学技术已成为强大的社会生产力和巨大的社会资本。现在，每个国家，每个地区，甚至每个单位，都把科学技术创新、科学技术转化为生产力作为头等大事，抢占科学技术制高点，以此来提高自己的综合实力。

新中国成立50多年特别是改革开放20多年来，随着经济的蓬勃发展，科学技术得到了长足的进步，两弹一星、载人飞船、生物工程、信息技术等正在大步追赶国际先进水平。科学技术转化成的强大生产力，对国民经济发展和社会进步、对增强综合国力产生了重大的影响。

改革开放以来，在中国共产党的“科教兴国”方针的鼓舞下，举国上下，尊重科技，学习科技，普及科技，创新科技，应用科技，发展科技，已蔚然成风。科技结硕果、神州尽彩虹的绚丽画面，正在展示于世人面前。自16世纪中叶中国科学技术失去世界领先地位后所形成的中西科学技术的差距，现在正在缩小。重振中华科学技术雄风的序幕已经拉开。

为了能使我国的科学技术水平在不久的将来赶上并达到世界先进水平，我们不仅要自己进行科学技术创新，也要学习世界上一切国家的先进科学技术；不仅要靠国内的科技工作者发展我国的科学技术，还要借助海外学者特别是华人学者的力量。在这种思想的指导下，我们萌生了组织海外学者编写科技前沿丛书的想法。这一想法在海内外学者中引起了强烈的反响：在他们中，有的出谋献策，有的出资开会，有的撰稿，有的审稿，有的愿把稿酬作为基金，……海内外学者的诚言乐行，极大地感染着我们，鼓舞着我们；这一想法得到了教育部陈至立部长和分管我社的周远清副部长的肯定和支持，这增加了我们开展此项工作的决心和信心。根据各方面意见，经过反复研究，最后将丛书定名为《当代科学前沿论丛》。《论丛》是我们献给祖国母亲的21世纪的圣礼，企盼我国能在21世纪夺回三四百年前失去的科学技术领先的地位。《论丛》如能在推动我国科学技术进步和“科教兴国”中有所作用，将是我们的最大欣慰。为了

做好本《论丛》的出版工作，我们邀请了国内一些著名科学家和在海外工作的部分优秀学者组成《论丛》的专家委员会，帮助筹划、组织和评议《论丛》的出版。随着学科的发展，专家委员会的成员可能会有所变化。我们向一切关心和支持《论丛》出版工作的人士，表示衷心的感谢。由于缺乏经验，《论丛》出版后，编辑出版方面的不足，在所难免，诚望各方指正。

高等教育出版社

2000 年 6 月

前　言

在过去的20年中，由于计算机技术和相应统计软件的迅猛发展，量化分析已经成为社会科学各个学科领域中广为应用的技术方法．在社会科学诸如社会学、心理学、人口学、政治学、经济学以及公共卫生学当中，logistic回归模型是对二分类因变量(dichotomous dependent variable)(即 $y=1$ 或 $y=0$)进行回归分析时最为普遍应用的多元量化分析方法。根据Hosmer及其同事的统计(1991)，在1985—1989年间，国际知名刊物《美国公共卫生杂志》上发表的文章中约有20%(579篇文章中的113篇)应用了logistic回归模型．虽然logistic回归已经达到了如此流行的程度，但不少使用这一模型的人对于该模型的性质和原理的理解仍不很充分，在实际应用中常有困惑和这样那样的问题，对于模型结果的阐释还存在许多不一致．并且，在很多应用该方法的研究中连模型拟合优度的评价也被忽略了．比如，在上述所统计的113篇文章中，只有5%(6篇文章)涉及到模型拟合优度的评价①．

在现有的统计教科书中，一般都有logistic回归模型的内容．然而，在这些教科书中，logistic回归往往不是作为中心内容，缺乏关于这种方法的详尽讨论．有关专著在国外很少，国内尚无．国外的一些专著中对于logistic回归模型的实际应用，特别是结合统计软件运行模型并对模型结果进行解释方面较为欠缺．本书的主要目的是提供对于logistic回归模型的深入专题介绍，专注于这一方法本身的讨论，以及模型结果的详细阐释．作者尽量以深入浅出的手法，旁证博引，理论联系实际，大量运用例题并结合计算机统计软件的使用，介绍和讨论该模型的原理及运用．本书的服务对象是社会科学各专业及公共卫生学专业的教师及研究生，以及社会科学专业研究人员．读者在学习本书内容之前应对多元回归和统计推断的基础知识有所了解．

本书将采用国际上广泛使用的统计软件SAS (Statistics Analysis System)和SPSS (Statistics Package for Social Sciences)来分析书中的例题．本书将提供用这两种软件进行具体例题分析的计算机程序，并对于这两种软件的模型估计结果进行详细的解释和对比分析．本书中例题的主要数据是由作者模拟制作

① Hosmer. Taber, and Lemeshow. 1991.

的，其原始数据可从下列网址下载：

http：//www.hep.com.cn；http：//www.wright.edu/~jwang；

http：//www.disa.pku.edu.cn/课程.

本书共由8章组成.

在第1章中，我们将首先讨论分析二分类因变量时所产生的问题，并讨论经典的线性概率模型(linear probability model，简标为LPM)及其局限性. 然后介绍logistic回归模型.

在第2章中，我们将介绍logistic回归模型估计所用的最大似然估计法(maximum likelihood estimation，简标为MLE)、模型估计的假设条件，以及最大似然估计的性质. 此外，还将介绍对分组数据进行logit分析的加权最小二乘法(weighted least squares，简标为WLS).

第3章介绍logistic回归模型的评价，讨论各种拟合优度(goodness of fit)，预测准确性(predictive accuracy)和模型卡方统计(model chi-square statistic).

第4章关注于logistic模型回归系数意义的阐释. 除了讨论发生比率(odds ratio)、预测概率(predicted probability)和互动影响(interactions)外，这一章还要讨论使用各种不同编码时分类自变量回归系数的意义和解释.

第5章讨论logistic回归系数的统计推断(statistical inference).

第6章的内容涉及模型的选择，讨论建立模型过程中的策略.

第7章关于模型的诊断，讨论多元共线性(multicollinearity)、有问题的数据架构(problematic data configuration)、极端值(outliers)、特异影响案例(influential observations)和过离散分布(overdispersion)等问题，以及这些问题的补救对策.

在最后一章中，我们将介绍与logistic回归类似的另外一种分析二分类因变量的备选模型——probit模型. 然后，将logistic回归模型扩展到序次logistic回归模型(ordered logistic regression model)和多项式logit模型(multinomial logit model)，这些模型分别用以解决序次变量和多分类名义变量为因变量的问题.

责任编辑　林云裳
封面设计　刘晓翔
责任绘图　朱　静
版式设计　周顺银
责任校对　陈　荣
责任印制　杨　明

目　录

TABLE OF CONTENTS

二分类因变量与 logistic 回归模型

第一章