格致方法·定量研究系列　吴晓刚　主编

Logistic回归入门

[美] 弗雷德·C.潘佩尔（Fred C. Pampel）　著
周穆之　译　　陈伟　校

- 革新研究理念
- 丰富研究工具
- 最权威、最前沿的定量研究方法指南

38

格致出版社　上海人民出版社

格致方法·定量研究系列　吴晓刚　主编

Logistic 回归入门

[美] 弗雷德·C.潘佩尔（Fred C. Pampel）著
周穆之　译　　陈伟　校

SAGE Publications ,Inc.

格致出版社　上海人民出版社

图书在版编目(CIP)数据

Logistic 回归入门/(美)潘佩尔(Pampel, F.C.)著;周穆之译.—上海:格致出版社:上海人民出版社,2014
(格致方法·定量研究系列)
ISBN 978-7-5432-2465-0

Ⅰ.①L… Ⅱ.①潘… ②周… Ⅲ.①线性回归-回归分析 Ⅳ.①0212.1

中国版本图书馆 CIP 数据核字(2014)第 278728 号

责任编辑 顾 悦
美术编辑 路 静

格致方法·定量研究系列

Logistic 回归入门

[美]弗雷德·C.潘佩尔 著
周穆之 译 陈伟 校

出 版 世纪出版股份有限公司 格致出版社
世纪出版集团 上海人民出版社
(200001 上海福建中路 193 号 www.ewen.co)
编辑部热线 021-63914988
市场部热线 021-63914081
www.hibooks.cn
发 行 上海世纪出版股份有限公司发行中心

印 刷 浙江临安曙光印务有限公司
开 本 920×1168 1/32
印 张 4.5
字 数 88,000
版 次 2015 年 1 月第 1 版
印 次 2015 年 1 月第 1 次印刷

ISBN 978-7-5432-2465-0/C·119 定价:20.00 元

出版说明

由香港科技大学社会科学部吴晓刚教授主编的“格致方法·定量研究系列”丛书，精选了世界著名的SAGE出版社定量社会科学研究丛书，翻译成中文，起初集结成八册，于2011年出版。这套丛书自出版以来，受到广大读者特别是年轻一代社会科学工作者的热烈欢迎。为了给广大读者提供更多的方便和选择，该丛书经过修订和校正，于2012年以单行本的形式再次出版发行，共37本。我们衷心感谢广大读者的支持和建议。

随着与SAGE出版社合作的进一步深化，我们又从丛书中精选了三十多个品种，译成中文，以飨读者。丛书新增品种涵盖了更多的定量研究方法。我们希望本丛书单行本的继续出版能为推动国内社会科学定量研究的教学和研究作出一点贡献。

总序

2003年，我赴港工作，在香港科技大学社会科学部教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来，非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”（Statistics for Social Science）为所有研究型硕士生和博士生的必修课，而第二门课“社会科学中的定量分析”为博士生的必修课（事实上，大部分硕士生在修完第一门课后都会继续选修第二门课）。我在讲授这两门课的时候，根据社会科学研究生的数理基础比较薄弱的特点，尽量避免复杂的数学公式推导，而用具体的例子，结合语言和图形，帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上，即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”，学完这些课程后，我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章；另一方面，也能在自己的研究中运用这些成熟的方法论技术。

上述两门课的内容，尽管在线性回归模型的内容上有少

量重复，但各有侧重。“社会科学里的统计学”从介绍最基本的社会研究方法论和统计学原理开始，到多元线性回归模型结束，内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型，以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法，将重点放在因变量为定类数据的分析模型上，包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型，包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生，一直鼓励和支持我将两门课的讲稿结集出版，并帮助我将原来的英文课程讲稿译成了中文。但是，由于种种原因，这两本书拖了多年还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩，每本书都写得通俗易懂，与我的教学理念是相通的。当格致出版社向我提出从这套丛书中精选一批翻译，以飨中文读者时，我非常支持这个想法，因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种语言的精准把握能力，还要有对实质内容有较深的理解能力，而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容，只有语言能力是远远不能胜任的。在短短的一年时间里，我们组织了来自中国内地及香港、台湾地区的二十几位

研究生参与了这项工程，他们当时大部分是香港科技大学的硕士和博士研究生，受过严格的社会科学统计方法的训练，也有来自美国等地对定量研究感兴趣的博士研究生。他们是香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛，硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊，应用社会经济研究中心研究员李俊秀；香港大学教育学院博士研究生洪岩璧；北京大学社会学系博士研究生李丁、赵亮员；中国人民大学人口学系讲师巫锡炜；中国台湾"中央"研究院社会学所助理研究员林宗弘；南京师范大学心理学系副教授陈陈；美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛；美国加州大学洛杉矶分校社会学系博士研究生宋曦；哈佛大学社会学系博士研究生郭茂灿和周韵。

参与这项工作的许多译者目前都已经毕业，大多成为中国内地以及香港、台湾等地区高校和研究机构定量社会科学方法教学和研究的骨干。不少译者反映，翻译工作本身也是他们学习相关定量方法的有效途径。鉴于此，当格致出版社和 SAGE 出版社决定在"格致方法·定量研究系列"丛书中推出另外一批新品种时，香港科技大学社会科学部的研究生仍然是主要力量。特别值得一提的是，香港科技大学应用社会经济研究中心与上海大学社会学院自 2012 年夏季开始，在上海(夏季)和广州南沙(冬季)联合举办《应用社会科学研究方法研修班》，至今已经成功举办三届。研修课程设计体现"化整为零、循序渐进、中文教学、学以致用"的方针，吸引了一大批有志于从事定量社会科学研究的博士生和青年学者。他们中的不少人也参与了翻译和校对的工作。他们在

繁忙的学习和研究之余，历经近两年的时间，完成了三十多本新书的翻译任务，使得“格致方法·定量研究系列”丛书更加丰富和完善。他们是：东南大学社会学系副教授洪岩璧，香港科技大学社会科学部博士研究生贺光烨、李忠路、王佳、王彦蓉、许多多，硕士研究生范新光、缪佳、武玲蔚、臧晓露、曾东林，原硕士研究生李兰，密歇根大学社会学系博士研究生王骁，纽约大学社会学系博士研究生温芳琪，牛津大学社会学系研究生周穆之，上海大学社会学院博士研究生陈伟等。

陈伟、范新光、贺光烨、洪岩璧、李忠路、缪佳、王佳、武玲蔚、许多多、曾东林、周穆之，以及香港科技大学社会科学部硕士研究生陈佳莹，上海大学社会学院硕士研究生梁海祥还协助主编做了大量的审校工作。格致出版社编辑高璇不遗余力地推动本丛书的继续出版，并且在这个过程中表现出极大的耐心和高度的专业精神。对他们付出的劳动，我在此致以诚挚的谢意。当然，每本书因本身内容和译者的行文风格有所差异，校对未免挂一漏万，术语的标准译法方面还有很大的改进空间。我们欢迎广大读者提出建设性的批评和建议，以便再版时修订。

我们希望本丛书的持续出版，能为进一步提升国内社会科学定量教学和研究水平作出一点贡献。

吴晓刚

于香港九龙清水湾

序

用方程式估计一个二分因变量的时候，在可选择的数据分析工具中，logisitc 回归已经基本取代了普通最小二乘法(OLS)。即便是初学者也知道当 Y 是一个二分变量时，OLS 的结果也不会出现在最后发表的文章里。这种实践方法上的进步部分来自启发性的论文和书稿在过去 20 年来的积累。这一系列图书主要致力于教育读者。

基于有许多研究人员和专家对 logit 模型表示关注，依然有人可能会问额外的处理是否还有必要。回答是肯定的，一如我们手上这本书。对于新手来说，与普通最小二乘法相比，logistic 回归还是很难对付。所有统计软件的程序里面已经包含了 logistic 回归，执行一个 logit 程序其实很简单。可是，为什么要执行这个程序？此外，所得出的结果是什么意思？由于这些问题解释起来还是非常复杂的，任何一个尽责的教方法的老师都会给予非常认真的思量。最近刚刚上手掌握普通最小二乘法的新手需要的是一本入门性的参考书。这就是潘佩尔教授所著此书的用心所在。

第 1 章介绍了当因变量是二分变量时 logistic 回归的逻

辑。在那种情况下，普通的回归会遭遇各种问题，比如，非线性、无意义的估计、非正态、方差不齐，这些都会导致无效估计。将二分因变量转化成 logit 可以消除这些问题。潘佩尔教授解释了 logit 的概念（Y 的比数取了对数）以及它的运作原理，并提供了一个非常有用的说明对数的附录（我已经发现学生们第一次接触这些时都需要复习一下对数。现在他们已经有了方便的材料）。

第 2 章涉及对结果的解释，也就是本书的正文。许多教科书在这方面都呈一片混乱的景象。中心议题就是，X 的影响是什么？在 OLS 里，这是根据回归的斜率来概括的。在 logistic 回归里并不会如此直接。关键是有三个可能。首先，X 上每变化一个单位，可以直接将斜率的估计解释为在 logit 上期待的改变。可是这种用法没有什么直观的意义。第二种用法就是 X 上每变化一个单位，将系数转化为在比数上的变化（而非比数取了对数）。这种方法看上去绝对比第一种更有意义。第三种是用概率来描述 X 带来的影响。如果 X 从一个基线值增加了一个标准差，例如它的平均值增加了一个标准差，那么就可以计算出发生 Y 的概率增加的量。这种解释的困难之处就在于 X 必须是一系列指定好的值，而非在任何一个 X 上变化一个单位都适用。这些以及其他一些解释上的困难都在本书中进行了评价。

普通回归的估计方法是最小二乘法。然而当 Y 是一个二分变量时，由于本身非线性的关系，最小二乘法再也不是一个有效的估计了。因此，这里使用的是最大似然估计（MLE），作者在第 3 章里有详细的解说。尽管最终得出了很好的模型拟合，但这个领域依然有争议。第 4 章剩余部分进

一步讲述了争议内容，也就是是否要用 probit 而不用 logit。在阐明了二者的相似和不同之后，本书用了一个有说服力的例子说明了 logit 更加合适。总体而言，对那些正寻找一本介绍流行的 logistic 模型书籍的研究人员来说，潘佩尔的书就是他们所需的。

迈克尔·S.刘易斯-贝克

前　言

我称此书为“入门”，因为它将 logistic 回归里被认为是理所当然的内容进行了清晰的阐述。有些著作假设读者对比数、对数、最大似然估计以及非线性函数已经有了相当的熟悉，因而对概念的解释很抽象。另外的一些著作跳过了 logistic 回归逐步推理的逻辑框架而直接给了例子和对实际系数的解释。因此，学生有时就无法理解 logistic 回归背后的逻辑。这本书就是用基本的语言和最简单的例子来介绍这个逻辑。

第 1 章在简要介绍了用线性回归分析二分因变量所带来的问题后，提供了一个非技术的解释，然后更细致地介绍了 logit 转换。第 2 章介绍了核心内容——logistic 回归系数的解释。第 3 章涉及最大似然估计的含义以及 logistic 回归中模型的解释力。第 4 章回顾了 probit 分析，这是一个类似 logistic 回归的分析二分因变量的方法。第 5 章简要介绍了 logistic 回归的原理如何应用于三个或者更多个名义因变量的分析。因为 logistic 回归的基本逻辑同样适用于最后一章的延伸，后面的章节没有再如第 1 章到第 3 章一般对 logistic

回归给予那么深入细致的讨论。最后，附录回顾了对数的含义，也许能够帮助一些学生来理解对数在 logistic 回归以及普通回归中的应用。

目 录

第1章

Logistic 回归的逻辑

许多社会现象本质上是离散的或定性的，而不是连续的或定量的，比如某个事件是否发生、个人做出某种而非另一种选择、个人或集体由一种状态到另一种状态。人们会经历生产、去世、迁移（国内和国际）、结婚、离婚、加入或者退出就业市场、领取社会福利、收入跌破贫困线、投票给某候选人、支持或者反对某议题、犯罪、遭到逮捕、辍学、上大学、参加某个组织、生病、皈依某种宗教等情况，或者其他涉及某种特性、事件或者选择的情况。同样，大型社会机构，如社团、组织或者国家也会经历成立、分裂、消失等情形，或者由一个阶段过渡到另外一个阶段。

二分离散现象通常采取二分指示或者虚拟变量的形式。尽管这两个值可以用任何数字代表，但用 0 和 1 来代表有其优势。1 表示该事件发生，它所占的比例其实就是虚拟变量的均值，也可以用概率来阐述。