



格致方法·定量研究系列 吴晓刚 主编

Logistic回归中的交互效应

[美] 詹姆斯·杰卡德 (James Jaccard) 著
缪佳 译

- ★ 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

格致出版社 上海人民出版社

45

格致方法·定量研究系列 吴晓刚 主编

Logistic 回归中的交互效应

[美] 詹姆斯·杰卡德(James Jaccard) 著

缪 佳 译

SAGE Publications, Inc.

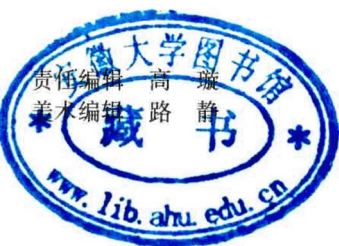
格致出版社  上海人民出版社

图书在版编目(CIP)数据

Logistic 回归中的交互效应/(美)杰卡德
(Jaccard, J.)著;缪佳译. —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 978-7-5432-2434-6

I. ①L… II. ①杰… ②缪… III. ①回归分析
IV. ①0212.1

中国版本图书馆 CIP 数据核字(2014)第 186853 号



格致方法·定量研究系列

Logistic 回归中的交互效应

[美]詹姆斯·杰卡德 著
缪佳 译

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
字数 74,000
版次 2014 年 11 月第 1 版
印次 2014 年 11 月第 1 次印刷

ISBN 978-7-5432-2434-6/C·110

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,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

分析非实验数据时,在回归模型中加入交互效应是一种常用的方法。如果 X 对 Y 的影响随着 Z 的取值不同而不同,就存在交互效应。此时正确的模型不再是 $Y=a+bX+cZ$, 而应该是 $Y=a+bX+cZ+d(XZ)$ 。其中乘积项 XZ 的估计系数 d 反映的就是交互效应。举个简单的例子,我们用 Y 表示收入, X 表示受教育年限, Z 表示性别(1 为男性, 0 为女性)。如果我们预期教育对收入的影响对男性和女性是不一样的,就需要加入交互项。对男性而言,收入的预测方程是 $Y=(a+c)+(b+d)X$, 对女性则是 $Y=a+bX$ 。如果估计系数 d 为正数且具有统计显著性,就说明相对于女性而言,教育年限的增加可以为男性带来更大的收入提高。用杰卡德博士的话说, X 的作用受到了调节因素 Z 的影响。

杰卡德博士在 1990 年就著有专著介绍多元回归中的交互效应(杰卡德、图里西:《多元回归中的交互作用》,格致方法·定量研究系列编号 4),该书推动了交互效应的应用和普及。之后,他又继续探讨了交互效应在 LISREL 和多因子方差分析(factorial ANOVA)中的应用。作为这个领域的专

家,杰卡德博士在这本专著中将交互效应的应用进一步推广到 logistic 模型中,在理论和应用层面都作出了重大贡献。虽然 logistic 模型已经被广泛应用,但是鲜有著作详细讨论如何在模型框架下使用交互效应。自 20 世纪 60 年代学者提出了交互效应的估计方法以来,在很大程度上,我们对交互效应的认识只是在等式右边加入一个乘积项,而杰卡德博士的专著大大加深了我们对交互效应的理解。

除技术细节外,本书的重点在于如何解释交互效应。作者首先回顾了如何用概率、发生比和对数发生比的形式解释 logistic 回归的结果。与普通回归相似,logistic 回归中的交互效应也用乘积项表示。最简单的例子就是两个分类变量的双向交互效应。例如,心理学家想要研究未成年人的性行为,因变量 Y 是一个两分变量(1 =有过性行为, 0 =没有性行为);解释变量有两个:一为性别(用 G 表示, 1 =男性, 0 =女性),二为母亲的就业状况(两个虚拟变量, F =全职或非全职, P =兼职或非兼职)。在 logistic 回归里,因变量是 Y 的对数发生比,自变量是 G , F , P 和乘积项 GF 和 GP 。这里的研究假设是:母亲的就业状况对男孩和女孩涉足性行为的影响是不一样的。如果 GP 的 logistic 系数(该系数的指数是一个比值,计算方法是:用母亲从事兼职工作的青少年发生性行为的性别优比除以母亲失业的青少年发生性行为的性别优比)在统计上是显著的,就说明性别和母亲的就业状态之间存在交互效应。在解释了双向交互效应之后,作者进而介绍了三向交互效应,他用到的例子是邮寄问卷的回收率研究,解释变量是三个类别变量:填答问卷有无经济奖励、问卷长度和研究问题的重要性,这是一个 $2 \times 2 \times 2$ 的多因子

设计。

之后的章节进一步讨论了更复杂情况下交互效应的解释方法,包括分类和连续变量的交互效应、两个连续变量的交互效应,以及当因变量是多类别变量时的交互效应。为了解释第三种情况,杰卡德博士列举了一项儿童心理学的研究。研究者将儿童对照料者的情感依附分为三种类型,解释变量是家庭环境和母亲的影响。研究者使用了 SPSS 统计软件对数据进行了多项式 logistic 回归。正如作者指出的那样,目前的统计软件已经能够处理 logistic 回归中的交互效应问题,真正的困难在于如何进行解释,本书的目的即在于一步步地指导读者去克服这一困难。

迈克尔·S.刘易斯-贝克

前言

本书介绍了如何在 logistic 回归中用乘积项来进行交互效应分析。我们主要关注在实际研究中可能遇到的多种情况,以及对 logistic 模型系数的解释。我们假设读者已经具备了 logistic 回归的基本知识,并对多层次 logistic 回归(hierarchical logistic regression)有一定了解。本书的重点不是从技术层面探讨这些复杂专题,相反,我们希望为读者提供一种非技术性的概论式指导,让他们知道在用乘积项来表示交互效应时,如何解释模型的 logistic 系数。虽然很多关于 logistic 回归的书都讨论过检验是否存在交互效应的一般方法,但是很少有专著指导读者去解释和理解方程的估计系数,本书正好填补了这一空白。我们旨在为应用型的研究者提供有关多元回归和 logistic 回归的基本知识,因此,我们避开了那些令人望而生畏的复杂的计算公式,取而代之的是易于理解(但稍微有点繁琐)的、基于统计软件所估计的参数和标准误进行计算的方法。为了减少四舍五入造成的差别,在所有例子中我们都保存了四位小数,不过一些例子中计算的结果还是有细微差别。

我要感谢以下审稿人富有建设性的评论和建议:美国博林格林州立大学(Bowling Green State University)社会学系的艾尔弗雷德·德马里斯(Alfred DeMaris),荷兰蒂尔堡大学(Tilburg University)社会和行为科学部的雅克·哈根纳斯(Jacques Hagenaars),行为科学研究所的斯科特·梅纳德(Scott Menard),以及迪米特里·里亚哈夫斯基(Dimitri Likhovitski),保罗·戈伦(Paul Goren),格伦·迪恩(Glenn Deane),理查德·阿尔巴(Richard Alba)。此外,我还要感谢迈克尔·刘易斯-贝克的大力支持。他们为本书的出版付出了大量的时间和精力。

目 录

序	1
前言	1
第 1 章 概述	1
第 1 节 概率和发生比	3
第 2 节 Logistic 回归模型	5
第 3 节 类别型解释变量和虚拟变量	7
第 4 节 Logistic 回归的预测值	9
第 5 节 系数解释	11
第 6 节 概率、发生比和对数发生比的换算	14
第 7 节 自变量的转换	16
第 8 节 交互效应的定义	18
第 9 节 多层次完全模型	22
第 10 节 交互项分析和分开进行的 logistic 回归	25
第 2 章 定性变量间的交互效应	27
第 1 节 双向交互效应	29

第 2 节	三向交互效应	35
第 3 章	定性和定量/连续变量的交互效应	41
第 1 节	调节变量是定性变量的双向交互效应	43
第 2 节	调节变量是定量变量的双向交互效应	47
第 3 节	三向交互效应	51
第 4 章	两个定量/连续变量的交互效应	57
第 1 节	双向交互效应	59
第 2 节	三向交互效应	62
第 5 章	多类别模型	65
第 1 节	定序回归模型	67
第 2 节	多类别名义变量	71
第 6 章	与交互效应相关的其他问题	75
第 1 节	展示交互效应的方法	77
第 2 节	计算置信区间	82
第 3 节	当调节变量取不同值时,计算关键自变量的系数	83
第 4 节	定量/连续变量交互项的双线性	86
第 5 节	分离成分项	89
第 6 节	多项的交互效应	90
第 7 节	多重共线性	93

第 8 节 模型选择和简化	94
第 9 节 转换	96
第 10 节 混杂的交互效应	97
第 11 节 电脑软件	98
注释	99
参考文献	100
译名对照表	102

第 **1** 章

概 述