



格致方法·定量研究系列 吴晓刚 主编

多元回归中的交互作用

[美] 詹姆斯·杰卡德 (James Jaccard) 著
罗伯特·图里西 (Robert Turrisi)

蒋勤 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

04

格致方法·定量研究系列 吴晓刚 主编

多元回归中的交互作用

[美] 詹姆斯·杰卡德(James Jaccard) 著
罗伯特·图里西(Robert Turrisi) 著
蒋勤译

SAGE Publications, Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

多元回归中的交互作用/(美)杰卡德

(Jaccard, J.), (美)图里西(Turrisi, R.)著;蒋勤

译.—上海:格致出版社;上海人民出版社,2015

(格致方法·定量研究系列)

ISBN 978-7-5432-2597-8

I. ①多… II. ①杰… ②图… ③蒋… III. ①多元回归分析-研究 IV. ①0212.1

中国版本图书馆 CIP 数据核字(2015)第 295286 号

责任编辑 顾悦

美术编辑 路静

格致方法·定量研究系列

多元回归中的交互作用

[美]詹姆斯·杰卡德 罗伯特·图里西 著

蒋勤 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 4.75
字数 90,000
版次 2016年1月第1版
印次 2016年1月第1次印刷

ISBN 978-7-5432-2597-8/C·140

定价:25.00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

经典回归模型的表达式很直接,以最常见的三变量多元回归为例, Y 是因变量, X_1 和 X_2 是自变量,可把方程 1 写成如下形式:

$$Y = a + b_1 X_1 + b_2 X_2 + e \quad [1]$$

这里的 a 、 b_1 和 b_2 都是待估计的总体参数, e 是残差。假设满足必要的假定,那么普通最小二乘法(OLS)估计就是最佳的线性无偏估计。这个 OLS 方程是加叠性的,各项共同解释 Y ,没有自变量相乘的项。假设自变量相乘产生乘积项 $X_1 X_2$,那么就可改写上面的方程为方程 2:

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 (X_1 X_2) + e \quad [2]$$

现在,我们有了既包括加叠性变量,又包括非加叠变量的方程。而乘积项或乘法的项目($X_1 X_2$)也被称为“交互作用项”,它的系数 b_3 估计交互作用。问题是,它的含义是什么呢?

通常,社会科学专业一年级的研究生会觉得这个问题挺难回答。他们这么理解加叠模型(方程 1):当 X_2 保持不变时, X_1 变化一个单位,会导致 Y 变化 b_1 个单位。但如何理解 b_3 ,

即方程 2 中非加叠项的系数? 当 X_1 对 Y 的影响取决于 X_2 的取值时, 就存在交互作用。 b_3 估计的正是这种交互作用。请考虑这个例子: Y = 个人对竞选运动的捐献额(美元), X_1 = 收入(美元), X_2 = 教育程度(0 = 未上大学, 1 = 大学及以上)。如果研究者认为对受过大学教育的人而言, 其收入对其捐献额影响更大(相比未受过大学教育者收入的作用), 那么研究者会偏向于使用包含交互作用的模型设定(方程 2), 而非方程 1。

很明显, 交互作用假设——一个变量的影响取决于另一个变量的取值——应当需要检验。回顾已发表的研究, 我们发现事实并非如此(很多研究不检验变量间的交互作用), 而我认为其原因在于交互作用这个想法较难捕捉, 因此本书非常有意义。Jaccard 博士和 Turrisi 博士使理解和估计回归模型中的交互作用变得非常清楚。比如, 他们强调, 仅仅通过调节变量在不同组之间的作用差异来检验交互作用是不够的, 如在上例中, 仅仅分别观察上大学和没上大学两组人的收入在(分别的)回归方程中对政治捐献的影响是不够的。进而, 他们提供了一个常常被忽略的见解: 真正的度量来自数据本身, 而不是测度。因此, 要用回归检验交互作用, 重要的是测度接近数据内在定距属性的程度。同时, 本书也有许多新主题, 比如聚类数据的交互作用以及随机系数模型。事实上, 仔细阅读参考文献部分, 我们能发现, 1990 年之后出版的文献有 30 多种。

研究实践者必须利用工具去假定、理解和估计交互作用。仔细学习此书能使研究者在回归分析中达到上述要求。若要在其他或更高级的研究情境下研究交互作用, 读者可参考 Jaccard 教授的其他著作。

目 录

序	1
第 1 章 导论	1
第 1 节 交互的概念	4
第 2 节 简单效应与交互比较	7
第 3 节 多元回归分析的回顾	11
第 4 节 本书概览	22
第 2 章 双向交互作用	23
第 1 节 包含乘积项的回归模型	25
第 2 节 两个连续性解释变量	28
第 3 节 一个定性解释变量与一个连续解释变量	46
第 4 节 小结	58
第 3 章 三向交互作用	59
第 1 节 连续解释变量	61
第 2 节 定性和连续解释变量	68

第3节	小结	79
第4章	其他重要问题	81
第1节	连续变量之交互作用的双线性本质	83
第2节	计算关键自变量在调节变量取不同值时的系数	86
第3节	计算偏组成项	89
第4节	变换	91
第5节	多重交互作用	92
第6节	标准化与非标准化系数	95
第7节	度量性质	97
第8节	测量误差	101
第9节	稳健分析与假定的违反	104
第10节	被试者内和重复测量设计	106
第11节	同序和非同序交互	109
第12节	显著性的区间	113
第13节	混合交互	115
第14节	最佳实验设计和统计解释力	117
第15节	协变量	119
第16节	控制实验误差	120
第17节	整体检验和交互作用	122
第18节	常见的误用	123
第19节	聚类数据和随机系数模型的交互作用	124

第 20 节 连续还是离散解释变量	126
第 21 节 调节框架的回顾	127

注释	128
----	-----

参考文献	130
------	-----

译名对照表	134
-------	-----

第 **1** 章

导 论

许多社会科学理论框架都关注因果模型。这些模型试图设定一个或多个自变量对一个或多个因变量或结果变量的影响。简单地说,因果模型存在六种可能关系,如图 1.1 所示。直接因果是一个变量 X 为另一个变量 Y 的直接原因。在理论系统语境下, X 就是 Y 的直接决定因素。间接因果是 X 对 Y 有因果作用,但 X 对 Y 产生的作用只能通过另一个变量 Z 来实现;虚假因果就是 X 和 Y 相关,但这种相关性完全是因为存在一个共同原因 Z ,即 X 和 Y 不存在真正的因果关系;互为因果是指 X 对 Y 有因果影响, Y 对 X 也有因果影响;待定因果是指 X 对 Y 有因果影响, Y 对 X 也有因果影响;

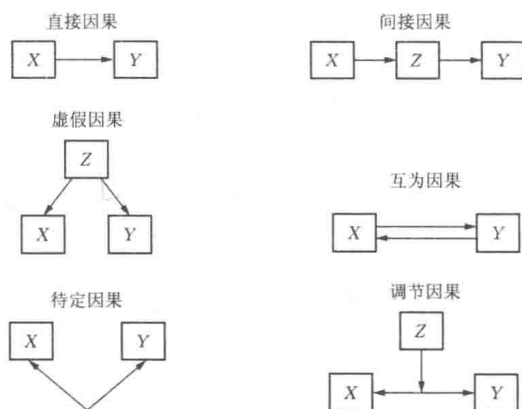


图 1.1 因果关系的类型

果是指 X 和 Y 相关,但这种关联的来源并未得到确认;最后,调节因果指 X 和 Y 之间的关系受到另外一个变量 Z 的调节作用。换句话说, X 和 Y 的关系是变动的,这种变动取决于 Z 的取值。

本书考察的是调节关联的统计分析,关注至少有一个自变量(或解释变量)是连续变量的情况。调节关系常被称为“交互作用”,尽管不同统计模型对交互作用的精确概念化有所不同。我们的关注点在于分析多元回归中的调节关联。在使用多元回归方法进行涉及连续变量的调节关联分析时,存在一些困惑需要厘清。

就如何最好地检验这些模型,现有文献充斥着大量相互矛盾的建议和忠告。本书最主要的目的就是梳理有关此主题的零散文献,说明进行包含连续变量调节关系的分析所涉及的中心议题。我们的目标是为那些具备多元回归操作知识的社会科学研究者展现一个合理的、易操作的导论。

在这一章中,我们会简单地回顾社会科学研究中交互作用分析的关键概念。首先考察一般意义上的交互作用概念,然后分析简单主效应和交互对比的概念。我们将回顾基本的多元回归程序,为下面各章节提供一个框架,包括虚拟变量的使用、简单变换对回归系数的影响以及预测值的计算。

第 1 节 | 交互的概念

如上面指出的,在社会科学中,许多方法都对交互作用进行了概念化,并且在哪种方法是最佳方法这一点上,存在很大的争议。一个流行的概念化交互作用的流派是把它当做调节关系。这一视角可用一个三变量系统进行展示,在此系统中,第一个变量被视为结果变量,第二个变量被视为自变量,第三个变量被视为调节变量,并且,结果变量被认为受到自变量的影响。而当自变量对因变量的影响取决于第三个变量的取值时(这一变量又称调节变量),就存在交互作用。例如,受教育年限对收入的影响可能因民族的不同而不同。教育对某些民族收入的影响要比另一些民族大。在调节变量框架中,收入是结果变量,受教育年限是自变量,而民族就是调节变量。再如,社会阶级对人们使用保健站频率的影响可能因性别的不同而不同。在此个案中,某人使用保健站的频率是结果变量或者因变量,社会阶级是自变量,而性别则是调节变量。性别被认为能够“调节”社会阶级对保健站使用情况的影响。

利用调节变量的方法进行交互分析要求理论家确定一个调节变量以及所谓的关键自变量。关键自变量对因变量的影响被认为根据调节自变量的某种函数关系而变化。大