



格致方法·定量研究系列

吴晓刚 主编

项目功能差异 (第二版)

[美] 史蒂文·J. 奥斯德兰 (Steven J. Osterlind) 著
霍华德·T. 埃弗森 (Howard T. Everson)

周韵 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

37

格致方法·定量研究系列 吴晓刚 主编

项目功能差异(第二版)

[美] 史蒂文·J.奥斯德兰 (Steven J.Osterlind) 著
霍华德·T.埃弗森 (Howard T.Everson)
周 韵 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

项目功能差异:第2版/(美)奥斯德兰
(Osterlind, S. J.), (美)埃弗森(Everson, H. T.)著;
周韵译. —上海:格致出版社;上海人民出版社,
2013
(格致方法·定量研究系列)
ISBN 978-7-5432-2206-9

I. ①项… II. ①奥… ②埃… ③周… III. ①社会科
学-研究方法 IV. ①C3

中国版本图书馆 CIP 数据核字(2012)第 299597 号

责任编辑 王亚丽

格致方法·定量研究系列

项目功能差异(第二版)

[美] 史蒂文·J. 奥斯德兰 著
霍华德·T. 埃弗森

周韵 译

出版 世纪出版集团 格致出版社
www.ewen.cc www.hibooks.cn
上海人民出版社
(200001 上海福建中路193号24层)



编辑部热线 021-63914988
市场部热线 021-63914081

发行 世纪出版集团发行中心
印刷 浙江临安曙光印务有限公司
开本 920×1168 毫米 1/32
印张 5.25
字数 81,000
版次 2013 年 2 月第 1 版
印次 2013 年 2 月第 1 次印刷
ISBN 978-7-5432-2206-9/C·96
定价 15.00 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中的 35 种,翻译成中文,集结成八册,于 2011 年出版。这八册书分别是:《线性回归分析基础》、《高级回归分析》、《广义线性模型》、《纵贯数据分析》、《因果关系模型》、《社会科学中的数理基础及应用》、《数据分析方法五种》和《列表数据分析》。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欢迎,他们针对丛书的内容和翻译都提出了很多中肯的建议。我们对此表示衷心的感谢。

基于读者的热烈反馈,同时也为了向广大读者提供更多的方便和选择,我们将该丛书以单行本的形式再次出版发行。在此过程中,主编和译者对已出版的书做了必要的修订和校正,还新增加了两个品种。此外,曾东林、许多多、范新光、李忠路协助主编参加了校订。今后我们将继续与 SAGE 出版社合作,陆续推出新的品种。我们希望本丛书单行本的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

往事如烟，光阴如梭。转眼间，出国已然十年有余。1996年赴美留学，最初选择的主攻方向是比较历史社会学，研究的兴趣是中国的制度变迁问题。以我以前在国内所受的学术训练，基本是看不上定量研究的。一方面，我们倾向于研究大问题，不喜欢纠缠于细枝末节。国内一位老师的话给我的印象很深，大致是说：如果你看到一堵墙就要倒了，还用得着纠缠于那堵墙的倾斜角度究竟是几度吗？所以，很多研究都是大而化之，只要说得通即可。另一方面，国内（十年前）的统计教学，总的来说与社会研究中的实际问题是相脱节的。结果是，很多原先对定量研究感兴趣的学生在学完统计之后，依旧无从下手，逐渐失去了对定量研究的兴趣。

我所就读的美国加州大学洛杉矶分校社会学系，在定量研究方面有着系统的博士训练课程。不论研究兴趣是定量还是定性的，所有的研究生第一年的头两个学期必须修两门中级统计课，最后一个学期的系列课程则是简单介绍线性回归以外的其他统计方法，是选修课。希望进一步学习定量研

究方法的可以在第二年修读另外一个三学期的系列课程,其中头两门课叫“调查数据分析”,第三门叫“研究设计”。除此以外,还有如“定类数据分析”、“人口学方法与技术”、“事件史分析”、“多层线性模型”等专门课程供学生选修。该学校的统计系、心理系、教育系、经济系也有一批蜚声国际的学者,提供不同的、更加专业化的课程供学生选修。2001年完成博士学业之后,我又受安德鲁·梅隆基金会资助,在世界定量社会科学研究的重镇密歇根大学从事两年的博士后研究,其间旁听谢宇教授为博士生讲授的统计课程,并参与该校社会研究院(Institute for Social Research)定量社会研究方法项目的一些讨论会,受益良多。

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”(Statistics for Social Science)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,这两本书拖了四年多还没有完成。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。中山大学马骏教授向格致出版社何元龙社长推荐了这套书,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港台地区的二十几位研究生参与了这项工程,他们目前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。

关于每一位译者的学术背景,书中相关部分都有简单的介绍。尽管每本书因本身内容和译者的行文风格有所差异,校对也未免挂一漏万,术语的标准译法方面还有很大的改进空间,但所有的参与者都做了最大的努力,在繁忙的学习和研究之余,在不到一年的时间内,完成了三十五本书、超过百万字的翻译任务。李骏、叶华、张卓妮、贺光烨、宋曦、於嘉、郑冰岛和林宗弘除了承担自己的翻译任务之外,还在初稿校对方面付出了大量的劳动。香港科技大学霍英东南沙研究院的工作人员曾东林,协助我通读了全稿,在此

我也致以诚挚的谢意。有些作者,如香港科技大学黄善国教授、美国约翰·霍普金斯大学郝令昕教授,也参与了审校工作。

我们希望本丛书的出版,能为建设国内社会科学定量研究的扎实学风作出一点贡献。

吴晓刚

于香港九龙清水湾

序

从广义上说,社会科学中的定量研究有两大任务。首先,社会科学研究者希望能够用特定测量尺度捕捉并描摹社会现象,而这一测量尺度必须对应该社会现象的概念化。其次,研究者通常希望能够用某种解释性统计模型描摹社会现象,以便从模型中更深刻地理解现象的成因。

本系列中绝大部分丛书可被归于测量和建模这两个类别中(其中,建模类在数量上稍占上风)。此外,仅有一小部分丛书是有关统计与数学方面的,或是探讨一个截然不同的研究模式(例如个体为本模型——然而即便对于个体为本建模而言,测量和建模方面的知识也是至关重要的)。

测量的历史源远流长。远古时代,人们为了生存即有了测量的需要,如建造房屋、缝制衣物、交换生存所需的食品货物等。然而,除基本生计外,人们也需要测量:若是没有足够好的测量,我们就不会有斯通亨治巨石阵。然而,直到近代,我们才有了近乎社会科学意义上的测量。约翰·格兰特(John Graunt)于1662年所作的《关于死亡表的自然与政治观察》可被看作测量人类生存的第一次严肃尝试。而对于教

育和心理学研究中如何建构有用的测量,这样的思考则是更近的事情了。直到1904年,爱德华·桑代克(Edward Thorndike)才意识到,在拼写能力测试中使用拼对的词数为标准存在无法避免的模糊性。

无需赘言,对于社会科学研究而言,测量(或教育与教育心理学中的测试)极其重要。我们也需要明白,测量为何对本系列丛书同样至关重要。继奥斯德兰(Osterlind)于1983年出版《测试项目偏差》后,相关研究欣欣向荣。如今,测试项目这一术语本身也不足以描述这一领域中的问题了。也正因为如此,奥斯德兰与埃弗森(Everson)合著了这本新书《项目功能差异》。本书并不仅仅是对奥斯德兰1983年著作的简单延伸——项目功能差异这一术语,其含义就已经超过了测试项目偏差。简言之,项目功能差异指的是,测试项目在经过能力配对后的不同社会群体(如种族、性别和阶级)间存在功能性差别。因此,尽管这一方法主要立足于教育与心理学,却对广义的社会科学研究具有重要意义。

本书是对旧作的及时更新,更为本丛书带来了领域内的新发展。在本书中,奥斯德兰和埃弗森回顾了一系列项目功能差异统计方法,包括传统的 Mantel-Haenszel 检验、似然比检验,以及 logistic 回归应用(这一方法在本系列其他丛书中,常被作为一种解释性模型——而不是对测量的评估——加以讨论)。尽管本书着重关注教育与心理学,然而从测量的角度说,其他社会科学领域的读者也将大获裨益。

廖福挺

目 录

序	1
第 1 章 导 言	1
第 1 节 读者与背景知识	4
第 2 节 项目功能差异研究的角色	6
第 3 节 公平性与项目功能差异	7
第 4 节 偏差与项目功能差异	8
第 5 节 项目功能差异与歧视	9
第 6 节 理解“项目功能差异”之概念	10
第 7 节 本书使用语言	11
第 8 节 本书所用示例	12
第 2 章 项目功能差异描述	13
第 1 节 项目功能差异正式定义	16
第 2 节 项目功能差异:统一与否	20
第 3 章 项目功能差异的统计侧面	25
第 1 节 并非均值区别	26
第 2 节 测量误差	27
第 3 节 系统性的区别	29

第 4 节	能力配对	30
第 5 节	控制条件:外部依据与内部依据	31
第 6 节	数据清洁	33
第 4 章	重要考量	35
第 1 节	项目功能差异:统计判断与经验判断	37
第 2 节	统计偏向并非不公平	38
第 3 节	完整测试与单个项目	39
第 4 节	数字:项目与样本	41
第 5 节	理性视角	43
第 5 章	项目偏差与项目功能差异的历史	45
	测试公平性的标准	48
第 6 章	快速而不全面的方法	51
第 1 节	项目排序法	53
第 2 节	能力组法	54
第 3 节	过时的 ANOVA 方法	56
第 7 章	Mantel-Haenszel 步骤	57
第 1 节	卡方列联表	60
第 2 节	M-H 比值比	62
第 8 章	非参数方法	65
第 1 节	使用 SIBTEST 的项目功能差异	67
第 2 节	Dorans 标准化	70
第 9 章	依托于项目反应理论的方法	73
第 1 节	项目反应理论的框架	76
第 2 节	项目反应曲线	79

第 3 节	单参数模型	83
第 4 节	双参数模型	85
第 5 节	三参数模型	88
第 6 节	依托于项目反应理论的项目功能差异方法	90
第 7 节	项目参数中的区别	91
第 8 节	似然比检验	93
第 9 节	区域测量	99
第 10 节	多分类项目中检验项目功能差异的项目 反应理论方法	102
第 10 章	logistic 回归	105
第 1 节	logistic 回归的项目功能差异表达	108
第 2 节	项目功能差异 logistic 回归示例	110
第 11 章	特定的项目功能差异研究方法	115
第 1 节	多分类评分项目	117
第 2 节	机考	120
第 3 节	计算机自适应测验(CAT)	121
第 4 节	翻译的测试	123
第 12 章	未来研究方向	125
第 1 节	效度论题	127
第 2 节	原假设检验	129
第 3 节	统计模型(HLM 模型及其他)	130
第 4 节	等同性检验	132
第 5 节	DFIT 与 CDIF 检验	133
第 13 章	总结	135
参考文献		137
译名对照表		146

第 **1** 章

导 言

本书旨在取代奥斯德兰于 1983 年所著的《项目偏差检验》一书。《项目偏差检验》为本系列(即《社会科学中的量性应用》)的第 30 本书。在过去的 25 年里,对于组间在检验项目中存在的可识别的差异表现,这方面的心理测量研究经历了重大的发展,因而仅仅对过往的著述进行更新已远远不够。项目调查的具体技术有所改变(一些旧方法被抛弃,新的技术取而代之),不仅如此,更关键的是,“项目偏差”这一概念本身也被重新界定。与局限于单个项目的早期研究相比,当下的相关研究大多立足于更加宽广的视角。如今,我们不仅着眼于结构表现(construct representation)中的测试公平(test fairness)与组内无差性(group invariance),还关注项目功能差异(Differential Item Functioning, DIF)。此外,一些研究者正试图从实证角度,识别不同组间表现差异的原因(Allalouf、Hambleton & Sireci, 1999; Roussos & Stout, 1996)。在这一背景下,“项目功能差异”并非仅是对“项目偏差”的更确切表述,而是评估中广义上的现象。它涉及心理评价中一系列不同方面,对于测试的开发和检验有着重要的影响。它是项目和测试功能方面尚未有定论的研究的一部分。

立足于“项目功能差异”这一更宽广的视角,本书旨在向读者呈现项目功能差异及相关领域的前沿理论与研究,特别是项目功能差异的识别。通过讨论一系列相关共发问题,进而介绍一系列可用于研究项目功能偏差现象的统计方法。本书希望能够帮助读者广泛了解有关项目功能差异的方法技术。然而本书涉及的话题深广,并非薄薄一本书所能涵盖。因而贯穿本书,我们也将引述其他对进一步研究测量偏差、项目偏差和项目功能检验有助益的资料。这是一个充满活力的广阔领域,相关文献众多,并在不断发展。