



格致方法

定量研究系列



· 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

吴晓刚 / 主编

【美】诺瓦尔·D. 格伦(Norval D. Glenn) / 等著

纵贯数据分析

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

【美】诺瓦尔·D. 格伦(Norval D. Glenn) / 等著

纵贯数据分析

SAGE Publications, Inc.

格致出版社



上海人民出版社

图书在版编目(CIP)数据

纵贯数据分析/(美)格伦(Glenn, N. D.)等著;
吴晓刚主编. —上海:格致出版社;上海人民出版社,
2011

(格致方法·定量研究系列)

ISBN 978-7-5432-1903-8

I. ①纵… II. ①格… ②吴… III. ①统计数据-统
计分析(数学) IV. ①O212.1

中国版本图书馆 CIP 数据核字(2010)第 254753 号

责任编辑 顾悦

封面装帧 人马艺术工作室·储平

纵贯数据分析

[美]诺瓦尔·D. 格伦 等著

吴晓刚 主编

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 上海图字印刷有限公司印刷
开本 787×1092 毫米 1/16
印张 21.5
插页 1
字数 309,000
版次 2011 年 8 月第 1 版
印次 2011 年 8 月第 1 次印刷
ISBN 978-7-5432-1903-8/C·48
定价 55.00 元



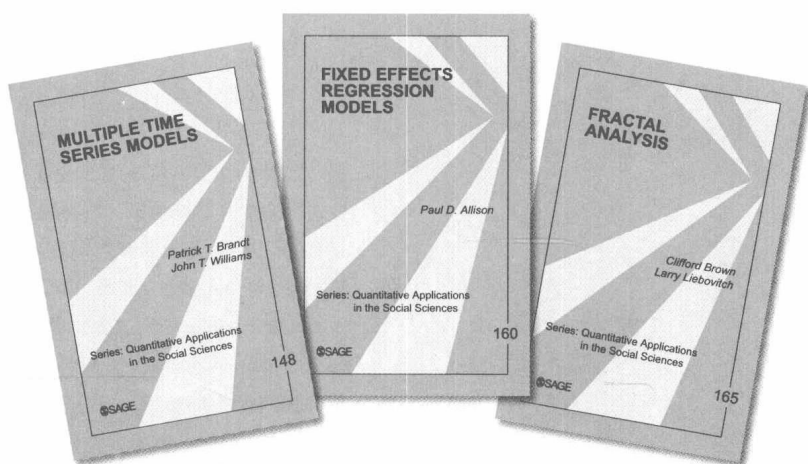
吴晓刚

美国加州大学洛杉矶分校(UCLA)社会学博士, 香港科技大学社会科学部教授、香港科技大学应用社会经济研究中心主任。主要研究领域为社会分层和流动、劳动力市场和经济社会学、教育、社会人口学、统计学和定量研究方法。多篇论文发表于 *American Journal of Sociology*、*American Sociological Review*、*Social Forces* 等国际著名专业杂志上。2006 年获美国国家教育学院 Spencer 博士后研究奖学金; 2007 年“以精辟数据研究中国社会”获美国社会学亚洲人 / 亚裔美国人“青年学人”奖。目前担任多家英文学术期刊的编委和 SSCI 期刊 *Chinese Sociological Review* (CSR) 的主编。

香港科技大学应用社会经济研究中心 (CASER) 成立于 2009 年 12 月。中心倡导以定量社会科学研究为主, 兼容多种研究方法, 结合多学科视角, 对中国面临的重大社会经济问题及其对人民生活福祉的影响进行实证研究。中心的研究项目涉及教育、就业、不平等、贫困、移民、健康等等。中心主持了“香港社会动态追踪调查”, 并与中国人民大学合作, 参与“中国教育追踪调查”。中心强调研究的科学性、独立性和社会意义, 并期望通过讨论会和工作坊以激发讨论, 将研究成果与更多的公众分享。中心网站: <http://caser.ust.hk>。



绿皮书的传奇



1976 年，当政府管理与政治学教授 Eric Uslaner 向 SAGE 出版社的创始人 Sara 建议出版关于重要定量研究方法的简明小册子时，没有人预见到这套书会取得巨大成功。

那年夏天，“社会科学定量研究方法”丛书诞生。统一的、朴素的绿色封面，上面仅仅印着书名、作者名及本书所属的系列名。Iverson 和 Northport 的《变量分析》(*Analysis of Variance*)、Nagel 的《运筹学》(*Operations Research*) 以及 Henkel 的《显著性检验》(*Tests of Significance*) 是最早出版的几本，每本售价 2.95 英镑，被形容为“像烤饼一样好卖”。

SAGE 选择了 20 种顶尖的定量研究工具，然后去寻找合适的作者，邀请他们围绕这些工具撰写 92 页的小书。这些薄薄的绿皮书在课堂上深受欢迎，在图书馆成套陈列。

至今，绿皮书系列中共有 160 种在初版或再版。书的主题反映了量化研究方法的发展：从基础统计知识、数据类型、测量到计算机的应用以及博弈论。这套书非常畅销，其中最畅销的一本，是 1980 年出版的 Michael Lewis-Beck 的《回归方法的应用》(*Applied Regression*)。

出版说明

本书由四种讨论纵贯研究的小册子组成,分别是《世代分析》、《纵贯研究》、《分析重复调查数据》以及《多元时间序列模型》。纵贯研究相较截面研究而言,其最大的优点是可以描述事物的发展过程和变化,并从这种变化中考察社会发展趋势。但是,纵贯研究的执行往往必须建立在多年的数据收集和分析的基础之上,而且研究的重复性较大。鉴于该研究具有较高的难度和复杂性,本书分别对纵贯研究中极具代表性的几种分析进行了介绍,旨在使研究者对纵贯研究有一定的认识,并能以此方法进行研究。

《世代分析》介绍了检验数据时常用的一项分析方法,即世代研究。这本小册子分别讨论了世代研究的目的、定义、策略等问题,并结合实例,说明了该研究的具体应用和数据选择问题。《纵贯研究》讨论了纵贯研究的目的、数据收集、研究对象与研究模型等问题,能为研究发展趋势、历史变迁或动态过程的研究者提供帮助。《分析重复调查数据》指出,重复调查是从对个人层次的微观研究转向对整体层次的宏观研究,并讨论了重复调查的基本用途,即描述、分解、解释总体趋势、对个别参数变化进行估计。《多元时间序列模型》概述了多元时间序列分析中的许多高级方法,包括同时方程模型、向量自回归模型、格兰杰因果检验、冲击反应分析等。作者通过介绍这些方法,预测并描述了变量之间内生的动态关系。

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,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。

由于所选每本书都有一篇序言,对相关方法的背景和应用作了很好的介绍,我们均予以保留,内容在此不再赘述。为了方便起见,我们将内容相似的书目集册出版,每册三至五本不等,共八册,它们分别是:《线性回归分析基础》、

《高级回归分析》、《广义线性模型》、《列表数据分析》、《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》和《数据分析方法五种》。所冠书名未必能精准涵盖其中的内容,读者可自行参阅每本书的序言或目录。

我们希望本丛书的出版,能为推动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

目 录

世代分析

序	3
第 1 章 引言	4
第 2 章 估计年龄、时期和世代效应的策略	15
第 3 章 用世代分析理解变革	38
第 4 章 数据要求与数据可用性	45
第 5 章 未来展望	54
注释	56
参考文献	57
译名对照表	60

纵贯研究

序	63
第 1 章 概述	64
第 2 章 纵贯研究的目的	67
第 3 章 纵贯数据收集设计	85
第 4 章 纵贯研究问题	94

第 5 章 纵贯分析	108
注释	135
参考文献	136
译名对照表	148

分析重复调查数据

序	153
第 1 章 导论	155
第 2 章 区分年龄、时期及世代效应	160
第 3 章 总趋势	166
第 4 章 分解总趋势	175
第 5 章 分解总变化的一般模型	190
第 6 章 发掘个人层次关系的变化	197
第 7 章 总结:分析社会变迁	219
注释	221
参考文献	223
译名对照表	228

多元时间序列模型	
序	233
第 1 章 前言	235
第 2 章 对多元时间序列模型的介绍	237
第 3 章 基本的向量自回归模型	251
第 4 章 向量自回归分析范例	294
附录 多元时间序列分析的软件	320
注释	323
参考文献	325
译名对照表	329

世 代 分 析

作者简介

诺瓦尔·D. 格伦(Norval D. Glenn)

任教于美国得克萨斯大学奥斯汀分校,是该校的阿什贝尔·史密斯社会学讲座教授和斯泰尔斯美国研究讲座教授。其主要研究兴趣是老龄化、生命历程和现代社会中的家庭关系。曾担任《当代社会学》和《家庭问题杂志》的主编,也曾是《美国社会学评论》、《民意季刊》、《婚姻与家庭杂志》、《人口学》和《社会科学研究》等杂志的编委会成员。

最近出版的作品涉及的主题主要包括向政策制定者和公众传播社会科学的研究发现、婚姻选择中的体制机制的变化以及婚姻中年龄和婚姻稳定之间的关系。

译者简介

於 嘉

2009年毕业于中国人民大学,获法学学士学位,现为香港科技大学社会科学部在读硕士研究生。研究方向为社会分层、性别不平等。

序

许多社会科学学科都以“世代”研究为教学重心,例如人口统计学、流行病学、心理学、政治科学和社会学。格伦·埃尔德(Glen Elder)曾在其著作《大萧条中的孩子们——社会变迁对人生的影响》(*Children of Great Depression——Social Change in Life Experience*)一书中对“世代”这一概念进行了深入研究。该书讲述了一群孩子在成长过程中种种坎坷的经历,对他们而言,现代社会许多司空见惯的东西都是遥不可及、需要拼命争取才能拥有的。在本书的讨论中,我们以“迷惘的一代”为例,讲述他们怎样在战后努力向社会证明自己的能力。正是这些独特的经历造就了关于世代的研究。

一般而言,世代分析属于定量研究的范畴,研究者利用很多世代概念的估测方法对一些结果变量的效应进行研究。例如,在大萧条时期中,经历过战争的社会群体身上的某些行为特征。从这种意义上来讲,世代的成员都必须在特定时间段有同样的经历。另一方面,同一年出生的人(或特定时期出生的人群)经历了同样的社会变迁,或多或少都会表现出相似的性格特征。不同年龄的人有时也表现出这种群体特征。例如,同一年开始实习的毕业生们。所以在大多数情况下,“世代”指的是某一特定的出生人群。

20世纪70年代出版的《大萧条中的孩子们——社会变迁对人生的影响》点燃了研究者的研究热情,他们开始寻找估测特定社会行为或社会现象以及估计世代效应的方法。这可并不简单,因为估测对象不仅仅包括世代(出生年份)的效应,还有年龄(出生后)以及特定时期(现有年份)的效应。众所周知,如果无法识别限制,那么APC线性模型和叠加模型也就无法被识别,因为年龄、时期和世代三个因素恰好构成一个线性函数式,其中 $\text{年龄} = \text{时期} - \text{出生年}$,为一个恒等式。

诺瓦尔·D. 格伦发表的《世代分析》一书中提供了几种应对识别问题的新方法。作者在书中对世代分析进行了更深层次的讨论,内容涉及方法论基础——包括数据问题——并详细讲述了几种可供研究者选择的APC估测法。社会科学各个学科的研究者都会在此书中获得很多收获,那些研究年龄变化以及社会文化变迁的学者们将是最大的受益人。

廖福挺(Tim Futing Liao)

第1章 | 引言

世代分析并不是一项统计技术，而是检验数据时经常采用的策略。在过去的几十年中，随着可用数据量的不断增加，越来越多的社会科学研究者开始关注世代分析。虽然世代分析在与年龄增长及社会文化变革相关的重大问题中显示出有效性，但显然更多研究者是被其异常神秘的方法论难题所吸引的。毫无疑问，这种兴趣并非完全有益。许多针对世代分析的研究只是练兵，研究者花了许多精力去证明的问题其实在逻辑上不可行。这样的研究导致了许多过于自信，却很可能不正确的结论。

因此，我写这本书的主要目的，一方面是阻止研究者怀着这种错误的观念进行机械化的世代研究，另一方面是为了鼓励正确的研究方式。能更深入地了解世代分析的方法自然很好，但前提是研究者必须谨慎，且拥有足够的专业经验，这样才能正确破解其中的奥妙。

世代研究的目的

一般来说，研究者会在两种情况下使用世代分析，虽然这两种情况的目的都非常明确，但还是引发不少人对研究的目的进行一些区分。

在多数情况下，世代研究主要用于估测人类年龄变化所产生的效应，有时也估测婚姻产生的效应。显然，随机实验最适合用来估测诱因，但不能用于研究年龄变化产生的效应。我们无法将一定数量的个体随机分成实验组和对照组，无法观测实验组个体年龄变化的过程，并保证对照组不受与年龄变化相关的刺激因素的影响。因此，研究年龄变化效应的学生只能在一些非实验性方法