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】戴维·诺克 (David Knoke) / 等著

列表数据分析

格致出版社 上海人民出版社

需要全本请访问 www.ertongbook.com

格致方法
定量研究系列

吴晓刚 / 主编
【美】戴维·诺克 (David Knoke) / 等著

列表数据分析

SAGE Publications ,Inc.

格致出版社 上海人民出版社

图书在版编目（CIP）数据

列表数据分析/（美）诺克（Knoke，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 978-7-5432-1902-1

Ⅰ. ①列… Ⅱ. ①诺…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 254751 号

责任编辑　顾　悦
封面装帧　人马艺术工作室·储平

格致方法·定量研究系列
列表数据分析
[美]戴维·诺克 等著
吴晓刚 主编

出　版　世纪出版集团
www.ewen.cc
格致出版社
www.hibooks.cn
上海人民出版社

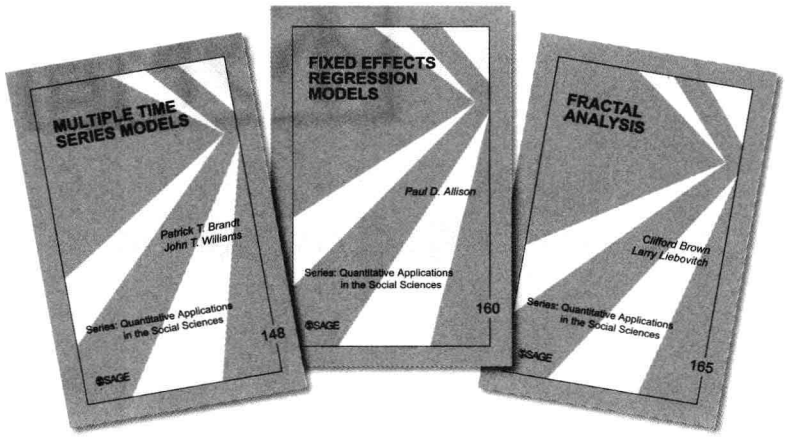
（200001　上海福建中路193号24层）

编辑部热线 021-63914988
市场部热线 021-63914081
格致出版

发　行　世纪出版集团发行中心
印　刷　上海图宇印刷有限公司
开　本　787×1092 毫米　1/16
印　张　19.5
插　页　1
字　数　284,000
版　次　2011 年 10 月第 1 版
印　次　2011 年 10 月第 1 次印刷
ISBN 978-7-5432-1902-1/C·47
定　价　49.00 元

SAGE

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 Northporth 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

本书由三种讨论交互表分析的小册子组成，分别是《对数线性模型》、《流动表分析》和《关联模型》。列表数据在分析和理解社会流动和社会分层方面具有重要的作用。本书首先介绍了用于检验变量间关系的对数线性模型及其检验和应用，然后考察了各类流动表数据分析模型的异同，最后特别介绍了双向表、三向表中的关联模型、偏关联模型和条件关联模型及其实际应用。

《对数线性模型》使用大量例子阐明了对数线性模型的具体运用，特别是用于分析多维交互表的分层对数线性模型。《流动表分析》应用一个有关父子职业问题的流动表数据来具体介绍流动表分析的概念、方法、特征等。《关联模型》分别介绍了行效应模型、列效应模型、行列效应模型、行列乘法效应模型等不同形式，并阐释了关联模型在交互表上的应用。

需要全本请访问 www.ertongbook.com

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。

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

对数线性模型

序 3

第 1 章 交互表中的关系 5

第 2 章 对数线性模型 9

第 3 章 拟合检验 26

第 4 章 实际问题的应用 38

第 5 章 对数线性模型的特殊技术 58

第 6 章 结论 70

参考文献 71

译名对照表 74

流动表分析

序 79

第 1 章 流动表分析 81

第 2 章 流动表的部分独立 93

第 3 章 社会距离模型 103

第 4 章 拓扑模型 115

第 5 章　尺度相关模型　130
第 6 章　新的发展　149
第 7 章　流动表在家庭和宗教研究中的应用　162
第 8 章　总结　170
参考文献　171
译名对照表　175

关联模型

序　181
第 1 章　简介　183
第 2 章　双向表中的关联模型　186
第 3 章　分析三向表的偏关联模型　222
第 4 章　条件关联模型在三向交互表上的应用　241
第 5 章　关联模型的实际应用　274
第 6 章　结论　288
注释　290
参考文献　292
译名对照表　300

需要全本请访问 www.ertongbook.com

对数线性模型

作 者 简 介

戴维·诺克(David Knoke)

美国印第安纳大学社会学系副教授。因致力于研究自愿性社团问题而被美国国家精神健康研究院授予“研究型科学家发展奖”。

彼得·J. 伯克(Peter J. Burke)

美国印第安纳大学社会学系教授兼系主任。目前的研究关注两个问题:理解自我的结构和小群体互动中的谈话结构。

译 者 简 介

盛智明

2008 年和 2010 年先后获得北京大学社会学硕士学位和香港科技大学社会科学部硕士学位,目前是香港科技大学社会科学部的博士研究生,研究方向是当代中国的社会运动、政治参与和基层政治改革。

序

通过撰写这本《对数线性模型》，戴维·诺克和彼得·J.伯克帮了我们一个大忙。在社会科学研究的大多数领域，最近几年，涌现出大量依靠各种技术对类别或名义变量进行多元分析的文章。但是，绝大多数从事实践工作的社会科学家都对这些技术感到困惑，因为这些术语通常并不为人所熟悉，而且看似与相关和回归分析中那些众所周知的概念无甚关联。如果你正被那些含有诸如优比、边际发生比或条件发生比、一般对数线性模型、饱和或非饱和模型、效应参数等类似术语的文章所困扰，阅读此书将对你大有裨益。诺克和伯克从头开始，为我们介绍、界定、讨论，并用大量的例子来澄清这些术语的含义。在这一过程中，他们将这些神秘的概念变得易于理解，即使对于大多数门外汉而言也是如此。

诺克和伯克讨论了一般对数线性模型，这一模型并不区分自变量和因变量，而是通过分析单元格的期望频数来检验类别变量间的关系；他们也讨论了 logit 模型，这一模型通过分析作为自变量函数的因变量的期望发生比来检验自变量与因变量之间的关系。他们从处理二分变量的方法开始讨论，然后逐步形成一个处理多类别变量的方法。

本书使用了大量的例子，大部分来源于政治社会学。本书中被扩展的例子包括讨论在控制了种族和教育后，自愿性社团组织成员身份与选举投票率之间的关系分析；对美国大众的公民自由态度的人口决定因素的因果分析；对党派认同与 1972 年和 1976 年总统选举之间关系的截面分析的比较研究；对从 1956 年到 1960 年之间的固定样本追踪研究中党派认同与宗教关系的检验；对宗教与堕胎态度之间关系的分析；对代际职业流动的考察以及一些其他例子。每个例子都阐明了对数线性模型的具体运用，比如，用作因果模型的模拟，进行时间序列分析，同步检验多个类别自变量对一个类别因变量的影响等等。由于诺克和伯克毫不吝啬地使用了许多不同数据库的例子，读者们不仅可以对有关对数线性模型的设计和检验的基本概念有所了解，同时也可以培养一种对这些模型的广泛适用范围的良好判断力。很显然，对数线性模型有着更为宽广的应用范围，虽然近些年对数线性模型或许已经在最大程度上在社会学中“流行”起来，但毫无疑问在接下来的 10 年里，它将在政治学、经济学、人类学、大众传播和其他领域中成为一个更加重要的

工具。它甚至可能使心理学和教育测试中的方差分析技术得到长足发展。

诺克和伯克很明显是对数线性模型的热情拥趸，虽然他们希望不仅可以阐释这个模型，而且可以将它发扬光大，但是他们也承认这一模型的一些不足，并在本书中涵盖了一些与这些建模技术的应用有关的具体问题以及不太明确的实质性问题。他们在合适的阶段对自己的陈述做了一个很好的总结。在总结中，他们讨论了对数线性模型运用中的具体问题，这些问题是所有希望有效使用这些模型的人所必须面对的。

尽管一些材料比较难懂，需要认真细致地学习，尤其是对那些统计初学者而言更是如此，但我完全相信，诺克和伯克在本书中的清晰阐述将使此书能够被广泛接受。对数线性模型是一个很难阐述清楚的题目，而本书则是最好的对数线性模型教学材料之一。

约翰·L. 沙利文(John L. Sullivan)

第 1 章 | 交互表中的关系

我们将通过对自愿性社团成员身份与投票参与率之间关系的一个详细分析来展现对数线性模型的基本原理。这个真实问题来源于研究民主参与的政治社会学。多年以来，研究者已经认识到，自愿性社团组织的成员更有可能参与各种各样的政治活动，诸如就社区问题与公共官员沟通，为候选人开展竞选活动和参与选举投票(Verba & Nie, 1972; Olsen, 1972)。但是，一些问题依然存在：自愿性社团成员身份与参与投票这个相关关系是不是社会地位的虚假结果？因为社会地位与这两个变量都正相关。当控制了社团参与程度与社会地位之后，白人和黑人在政治行为上是否仍然存在差异(Thomson & Knoke, 1980)？

为了检验这些假设，我们选用了 1977 年综合社会调查的数据，这个非机构性调查有 1530 个成年人(18 岁及以上)的全国样本，是在国家科学基金的资助下，由詹姆士·A. 戴维斯(James A. Davis)领导的芝加哥全国民意研究中心每年进行一次(现在为每两年一次)。“选举参与”(V)指的是调查对象是否参加了 1976 年的选举(不具备选举资格者被排除)。“自愿性社团组织的成员身份”(M)就是调查对象属于所列出的 16 种社团组织的数量，但排除了教会成员身份(Knoke & Thomson, 1977，该书讨论了教会成员身份如何与其他种类的成员身份不同)。我们将那些不属于任何组织的人与那些具有一个或更多组织成员身份的人做对比。种族(R)也是一个二分变量，分为白人和非白人(大多数为黑人)。最后，将教育(E)重新编码为 3 个大的类别：高中毕业以下、高中毕业、有大学经历并包含大学毕业及以上。最终，我们将分析这个四维交互表中的关系，但在刚开始时，我们集中分析成员身份—投票参与间的关系，将后一个变量

界定为依赖于或取决于前一个变量。后面的例子将会处理多类别变量以及讨论将变量合并为更少类别时的风险。

确定两个类别变量间关系或相关的传统方法是计算自变量中各类别的百分比，并比较这些百分比。如果变量各类别间的百分比差异显著(使用一般的卡方独立性检验)，那就表明存在相关关系。相关关系的形式——单调、线性或非线性的——取决于交互表单元格中百分比的分布模式(Reynolds，1977)。在表 1.1 中，没有组织成员身份的人中有 54%参加了投票，而在有一个或更多成员身份的人中则有 75%参加。在自愿性社团组织中，有成员身份的人比没有成员身份的人在投票参与率上高了 21 个百分点。这张表的卡方检验值为 67.7，表明在这两个变量之间存在着统计上显著的 ($p < 0.001$) 相关关系。

表 1.1　选举参与和组织成员身份交互表

		成员身份(M)		
		一个或更多	没　有	合　计
投票参与 (V)	投票	$f_{11} = 689$	$f_{12} = 298$	$f_{1.} = 987$
	没有投票	$f_{21} = 232$	$f_{22} = 254$	$f_{2.} = 486$
	合计	$f_{.1} = 921$	$f_{.2} = 552$	$f_{..} = 1473$

为了运用对数线性模型，我们首先必须重新概念化因变量。发生比将代替比例——单元格频数除以类别总数——成为有待解释的方差的基本形式。在日常生活中，我们最熟悉的关于发生比的例子来自赛马和其他形式的赌博。发生比就是属于某一类别的频数与不属于这一类别频数的比率。它可以被解释为一个随机选取的个体被观察到属于一个我们关注的类别而不是另一个类别的机会。例如，在表 1.1 中，一个人在 1976 年总统选举中投票的发生比是 $987/486 = 2.03$，或者大约为 2∶1(注意，这里可能出现了一些自我报告时的夸大，因为潜在投票者的实际投票率大概为 55%，只有 1.22 的发生比)。

刚才计算出的发生比是一个边际发生比，只应用了表中一边的总频数而没有考虑任何其他变量的影响。我们也可以计算表中的条件发生比，它对应于条件频数。条件发生比就是在所给出的一个组织成员身份的特定层次上，参加投

需要全本请访问 www.ertongbook.com