



格致方法·定量研究系列 吴晓刚 主编

数据分析概论

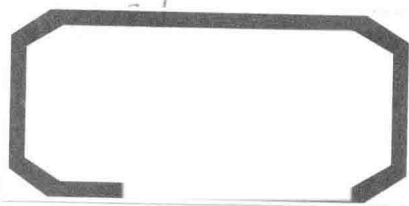
[美] 迈克尔·S.刘易斯-贝克 (Michael S. Lewis-Beck) 著
洪岩璧 译

- ★ 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

49

格致方法·定量研究系列 吴晓刚 主编



数据分析概论

[美] 迈克尔·S.刘易斯-贝克 (Michael S.Lewis-Beck) 著

洪岩璧 译

REID

SAGE Publications, Inc.

格致出版社  上海人民出版社

图书在版编目(CIP)数据

数据分析概论/(美)迈克尔·S.刘易斯-贝克著;
洪岩璧译.—上海:格致出版社;上海人民出版社,
2019.8

(格致方法·定量研究系列)

ISBN 978-7-5432-3038-5

I. ①数… II. ①迈… ②洪… III. ①统计数据-统
计分析 IV. ①0212.1

中国版本图书馆 CIP 数据核字(2019)第 159453 号

责任编辑 顾悦

美术编辑 路静

格致方法·定量研究系列

数据分析概论

[美]迈克尔·S.刘易斯-贝克 著

洪岩璧 译

出版 格致出版社

上海人民出版社

(200001 上海福建中路 193 号)

发行 上海人民出版社发行中心

印刷 浙江临安曙光印务有限公司

开本 920×1168 1/32

印张 4.5

字数 89,000

版次 2019 年 8 月第 1 版

印次 2019 年 8 月第 1 次印刷

ISBN 978-7-5432-3038-5/C·222

定价 35.00 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、有关关节数据的模型、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研究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

《韦氏新大学词典》(*Webster's New Collegiate Dictionary*)把“数据”定义为“一组事实”，因此社会科学数据(经验观察)是关于人类行为领域的事实。与流行的看法相反，事实并不会自己说话。数据分析的任务是试图赋予这些事实以意义。我之所以说“试图”赋予意义，是因为如果数据不好，它们就不能产生阐释，或者产生一个虚假的阐释。假设数据很好，那么分析就能为所研究的社会现象提供一个合理的描述和解释。

数据分析涉及对统计工具的系统运用。我们如何能够获得这些工具并恰当地使用它们？为了有效地学习分析技巧，我们需要从最简单的部分开始，将其作为构建更复杂技巧的基石。譬如，初学者经常犯的一个错误就是跳过基础知识，直接进入多元回归分析。为了学好多元回归，你必须首先扎实地掌握一元和二元统计知识。通过学习例如两个变量间的相关系数(皮尔逊相关系数)，你就能熟悉相关、强度、线性、测量层次、推论和标准化等概念。这为理解二元回归提供了基础，虽然有别于上述概念，但并非完全不同。一旦

熟悉掌握了二元回归,扩展到多元回归就比较容易,读者对后者也会有更深的理解。

本书旨在为定量研究数据分析的每一步提供统计学基础。在简述数据搜集之后,作者会讨论一元统计(测量集中趋势和离散),之后又讲解了对相关的测量(皮尔逊相关系数、tau 和 lambda 系数)和显著性检验,最后讨论了简单回归和多元回归。本书给出了必要的数学公式,但更注重平实的解释。同时,作者也利用图表来帮助阐释。本书非常注重实践的运用,并且以一个运用大学生样本数据来探讨影响学业能力因素的例子贯穿全书。

迈克尔·S.刘易斯-贝克

目 录

序	1
第 1 章 导论	1
第 2 章 数据搜集	5
第 1 节 研究问题	8
第 2 节 样本	10
第 3 节 测量	11
第 4 节 数据编码、输入和检查	15
第 3 章 一元统计	17
第 1 节 集中趋势	19
第 2 节 离散	23
第 3 节 集中趋势、离散和异常值	30

第 4 章	相关测量	33
第 1 节	相关	35
第 2 节	定序数据:tau 相关测量	39
第 3 节	名义数据:Goodman-Kruskal lambda 系数	45
第 4 节	二分变量:灵活选择	48
第 5 节	小结	49
第 5 章	显著性检验	51
第 1 节	显著性检验的逻辑:一个简单的例子	54
第 2 节	运用同一逻辑:二元相关测度	60
第 3 节	几个重要问题	65
第 4 节	小结	69
第 6 章	简单回归	71
第 1 节	Y 是关于 X 的方程	73
第 2 节	最小二乘法法则	76
第 3 节	截距和斜率	78
第 4 节	预测和拟合优度	81
第 5 节	显著性检验和置信区间	84
第 6 节	报告回归结果	88

第 7 章 多元回归	89
第 1 节 例子	91
第 2 节 统计控制	93
第 3 节 模型设定错误	96
第 4 节 虚拟变量	99
第 5 节 共线性	102
第 6 节 交互效应	107
第 7 节 非线性	110
第 8 节 归纳和结论	115
第 8 章 建议	117
附录	119
参考文献	123
译名对照表	125

第 **1** 章

导 论

社会科学研究始自有关人类行为的问题。不同学科提出的问题不尽相同。我们可以举一些例子：为什么女性比男性更喜欢投票？什么使人对工作感到满意？较高的失业率是否会引发犯罪率的上升？为了提高学生成绩，是否应该把讲座式授课改成讨论式授课？如果某社区接受安置一套化学废弃物处理设备，那么这个社区的居民罹患癌症的风险是否会增加？这些问题来自政治学、心理学、社会学、教育学和健康政策领域。基于一定的方法论，任何定量社会科学家都可以回答这些问题。这一方法论统一了貌似各不相同的多个学科，即完全基于系统性的经验观察，然后运用统计检验。这些检验结果（数据分析）有助于回答这些研究问题。分析做得越好，得到的结论就越扎实。

虽然不同的定量研究者可能对同一个问题感兴趣，并运用相同的研究工具，但这并不意味着他们一定会使用相同的分析策略。有些研究者的研究路径非常形式化，先设计具体的假设和测量工具，然后审慎地运用所选择的检验。其他一些研究路径则没那么形式化，研究者对想法和数据进行自由探索，在寻找“正确”模型的过程中使用不同的测算。优秀的研究者知晓这两种策略的优缺点，知道何时运用形式化的程

式,何时需要去探索发现。判断必须最终来自对统计结果的恰当解释,尤其是当这些结果来自非实验的社会研究时,而这正是本书所关注的。

毫无疑问,一位证明自己掌握了技术的研究者会在当今的学术界占上风。本书的目的是让初出茅庐的社会研究者具备一定的能力来为手头的问题选择恰当的统计检验工具。第2章论及初始的数据搜集,并提供了一个数据,本书将使用这个数据来说明各种分析技术。第3章介绍一元统计,即一次描述一个变量。第4章讨论相关测量,分析两个变量之间的关系。第5章探讨两个变量间关系的统计显著性问题。第6章关注简单回归,即一个因变量只受一个自变量影响。第7章阐述多元回归,即一个因变量受多个自变量影响。第8章进行总结,并给出一些分析建议。

