



格致方法·定量研究系列

吴晓刚 主编

应用回归导论

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

曾东林 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

46

格致方法·定量研究系列 吴晓刚 主编

应用回归导论

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

曾东林 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

应用回归导论/(美)贝克(Beck, M.S.L.)著;曾东林译.—上海:格致出版社;上海人民出版社, 2014

(格致方法·定量研究系列)

ISBN 978-7-5432-2457-5

I. ①应… II. ①贝… ②曾… III. ①回归分析-研究 IV. ①0212.1

中国版本图书馆 CIP 数据核字(2014)第 261261 号

责任编辑 高璇
美术编辑 路静

格致方法·定量研究系列

应用回归导论

[美]迈克尔·S.刘易斯-贝克 著
曾东林 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 3.75 字数 72,000 版次 2015 年 1 月第 1 版 印次 2015 年 1 月第 1 次印刷
---	--

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,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

我们对这本期待已久的关于应用回归分析的书稿终于可以付梓成书感到非常高兴。刘易斯-贝克博士用简洁、清晰的文字准确无误地完成了该书的写作。我相信那些刚入门的社会科学研究者会发现刘易斯-贝克的这本书给他们提供了一个理想的起点,以此来处理回归分析的一些入门和非技术性的内容。这本书强调的是应用回归分析,刘易斯-贝克博士提供了巧妙的例子来阐述有关正确运用和滥用回归分析的要点。他的例子包括:收入的决定因素,其中教育、资历、性别和政党立场作为自变量;影响采煤业死亡事故的因素;阿根廷大选中左右贝隆所获选票的因素;其他一些回归分析的实际应用。

刘易斯-贝克使用了很多例子来凸显其在解释回归分析所蕴含的假设方面的优势。他首先简洁地列出了这些假设,然后详尽地就每一个假设在实际使用中所要表达的意思及其实质性含义进行了非常出色的文字说明。入门者将会迅速地领会这些假设及其重要性,并知道如何在他们想要解决的实质性问题中评估这些假设。

刘易斯-贝克教授为回归分析的斜率估计和截距估计,以及对它们的解释都提供了简洁明了的处理方法。该书也展示并解释了可用于评估回归直线“拟合优度”的技术,包括对决定系数和显著性检验的讨论,后者呈现于对置信区间的更为一般的讨论之中。对显著检验的讨论包括了单尾检验和双尾检验,因而远胜于大多数初级的应用回归教材。另外一个突出的地方是在回归分析中对残差分析或者误差项的处理,它们的诊断能力在评估回归模型的假设时被清晰地展示出来。

最后,第3章介绍了多元回归分析(前面两章仅仅处理了二元回归,复杂的多元回归紧接其后)。在介绍完二元回归的基础上,该书简洁而又清晰地介绍了多元回归的原理。在多元的环境下,二元回归所涉及的每一个要点都被拓展,此外这也在一定程度上考虑了交互效应和多元共线性的复杂性问题。在总结部分,该书还关注了设定错误和测量度,其中包括虚拟变量的分析。在本书中,每一个问题都会通过大量例子来解释。

本书的重要性再怎么强调也不过分。或许回归分析比其他统计技术更能划分社会科学的学科界限。这里没必要列举其用途,因为所有社会科学的研究者,无论是那些尝试实证研究或是希望能紧跟学术前沿的人,都毫不怀疑地认同理解回归分析的必要性。

约翰·L.沙利文

目 录

出版说明	1
总序	1
序	1
第 1 章 二元回归：拟合一 条直线	1
第 1 节 精确关系与非精确关系	3
第 2 节 最小二乘法则	6
第 3 节 数据	9
第 4 节 散点图	10
第 5 节 斜率	12
第 6 节 截距	14
第 7 节 预测	15
第 8 节 评估解释效能： R^2	16
第 9 节 R^2 与 r	22

第 2 章	二元回归：假设与推断	23
第 1 节	回归假设	25
第 2 节	置信区间与显著性检验	31
第 3 节	单尾检验	35
第 4 节	显著性检验：一个经验法则	37
第 5 节	参数估计不显著的原因	39
第 6 节	Y 的预测误差	42
第 7 节	残差分析	44
第 8 节	对采煤业死亡事故的安全执法的效果：一个二元回归案例	50
第 3 章	多元回归	55
第 1 节	一般方程	57
第 2 节	解释参数估计	59
第 3 节	置信区间和显著性检验	63
第 4 节	R^2	65
第 5 节	预测 Y 值	66
第 6 节	交互效应的可能性	68
第 7 节	四变量模型：修正设定错误	71
第 8 节	多重共线性问题	74
第 9 节	高度多重共线性：一个例子	80
第 10 节	自变量的相对重要性	83

第 11 节	回归模型拓展:虚拟变量	87
第 12 节	采煤业死亡事故的决定因素:一个多元 回归案例	93
第 13 节	下一步?	96
注释		99
参考文献		100
译名对照表		101

第 **1** 章

二元回归：拟合一 条直线

社会研究学者经常询问两个变量之间的关系,这样的例子非常多。譬如,男性是否比女性更多地参与政治? 工人阶级是否比中产阶级更倾向于自由主义? 民主党的国会议员是否比共和党的国会议员花费更多的纳税人款项? 失业率的变化是否与总统民意支持率的变化相关联? 关于这种常见疑问的具体实例是,“变量 X 与变量 Y 的关系是什么”? 答案来自二元回归——一种用一条直线来拟合散点的简单技术。

第1节 | 精确关系与非精确关系

两个变量 X 与 Y 可能精确或非精确地关联。在自然科学中，变量之间通常具有精确的关系，最简单的关系是自变量（“原因”）——标记为“ X ”，以及因变量（“结果”）——标记为“ Y ”，两者的关系是一条直线，用方程表示为：

$$Y = a + bX$$

其中，系数的值 a 和 b 为这条直线限定了精确的高度和倾斜度。从而，系数 a 被视为截距或者常数，系数 b 被视为斜率。例如，表 1.1 使用一组假设的数据来显示 Y 与 X 是线性的关系，用方程表示如下：

$$Y = 5 + 2X$$

表 1.1 X 与 Y 的完全线性关系

$Y = 5 + 2X$	
X	Y
0	5
1	7
2	9
3	11
4	13
5	15

图 1.1a 显示了这条拟合了表 1.1 数据的直线。我们注意到对于每一个 X 的观测值,有且仅有一个可能的 Y 值。例如,对应于 X 的值为 1, Y 一定是等于 7。如果 X 的值增加一个单位, Y 就会精确地增加 2 个单位。因此,知道了 X 的值, Y 值则能精确地被预测。我们都熟知的一个例子是:

$$Y = 32 + 9/5X$$

其中,华氏温度(Y)是摄氏温度(X)的精确线性函数。

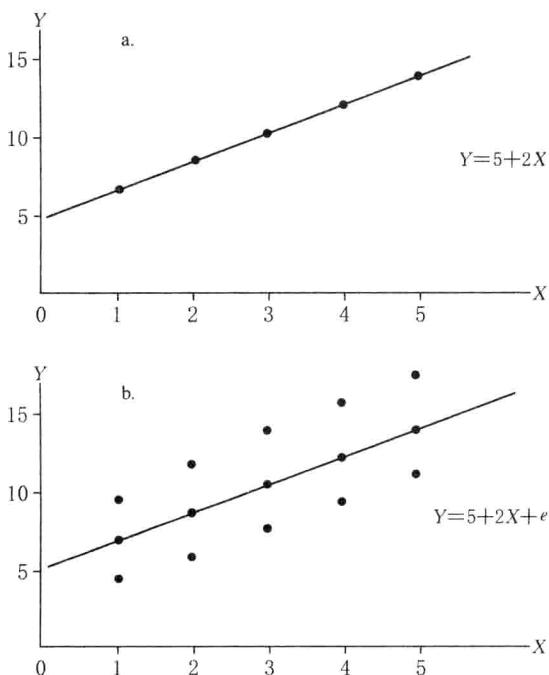


图 1.1 X 与 Y 的非精确线性关系

与上述例子相反,在社会科学中变量之间的关系几乎都是非精确的。更为真实的是,两个社会科学变量之间的线性