

格致方法·定量研究系列

吴晓刚 主编



社会统计的数学基础

[加] 约翰·福克斯(John Fox) 著
贺光烨 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

1

格致方法·定量研究系列 吴晓刚 主编

社会统计的数学基础

[加] 约翰·福克斯(John Fox) 著
贺光烨 译

图书在版编目(CIP)数据

社会统计的数学基础/(加)福克斯(Fox, J.)著;
贺光烨译. —上海:格致出版社;上海人民出版社,
2012

(格致方法·定量研究系列)

ISBN 978-7-5432-2109-3

I. ①社… II. ①福… ②贺… III. ①社会统计-应
用统计学-研究 IV. ①C91-03 ②0213

中国版本图书馆 CIP 数据核字(2012)第 122666 号

责任编辑 高 璇

格致方法·定量研究系列

社会统计的数学基础

[加]约翰·福克斯 著

贺光烨 译

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)



格致出版

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 7
字 数 138,000
版 次 2012 年 7 月第 1 版
印 次 2012 年 7 月第 1 次印刷
ISBN 978-7-5432-2109-3/C·64
定 价 18.00 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,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

曾经有一位社会学的博士研究生跟我说，他要去统计学系上一门基础课程，我问他为什么，他回答：“每次在我想更深入地学习高级定量方法时，总感觉有一堵无形的墙。”相对于社会科学院系，统计学系开设的课程更强调数学的基础性，因此，统计学系的学生更容易翻越这堵墙。即便“社会科学的数理基础”这套丛书考虑到了所面对的读者并没有接受足够的数学或统计学训练，然而，近期的许多话题，诸如稳健回归、潜在增长曲线模型等，均需要用到较多更深层次的数学知识，从而使许多读者望而生畏。

《社会统计的数学基础》就是为这些想进一步学习定量方法却时常感到被那堵无形的墙所阻碍的读者而编写的。这本小册子涵盖了许多数学和统计学中容易被人忽视却又至关重要的话题（如矩阵、线性代数、积分、概率理论及统计分布），这些话题经常在统计书籍和论文中出现，许多读者或许以前还接触过，但是对于大多数从事社会科学研究的读者而言，可能还比较陌生。

当得知福克斯的这个项目时，我异常兴奋并积极鼓励他

完成这本书。事实上,许多评论家包括作者本人都感叹,如果类似这样的书可以早出版几年,比如,在他们学习统计的时候,或者在他们为定量方法课程准备授课讲义的时候,那该有多好。

对于这本书,评论家一致认为:“它会是协助研究生及社会统计工作者进行研究的得力助手,也会成为大受欢迎的书籍。同时,这本书更将是对定量方法研究的一个重要补充。”

廖福挺

目 录

序	1
第 1 章 矩阵、线性代数和几何向量	1
第 1 节 矩阵	3
第 2 节 基础几何向量	23
第 3 节 向量空间与子空间	26
第 4 节 矩阵的秩及线性联立方程组的解法	34
第 5 节 特征值与特征向量	47
第 6 节 二次型及正定矩阵	52
第 7 节 推荐阅读	55
第 2 章 微积分入门	57
第 1 节 回顾	59
第 2 节 极限	66
第 3 节 函数求导	69
第 4 节 最优化	77
第 5 节 多变量和矩阵的微分学	81
第 6 节 泰勒展式	88

第 7 节	积分学的基本思想	91
第 8 节	推荐阅读	96
第 3 章	概率估计	97
第 1 节	初等概率理论	99
第 2 节	离散概率分布	116
第 3 节	连续分布	121
第 4 节	渐近分布理论:初步介绍	132
第 5 节	统计估计量的属性	138
第 6 节	最大似然估计	151
第 7 节	贝叶斯推断	167
第 8 节	推荐阅读	175
第 4 章	实际应用:线性最小二乘法回归	177
第 1 节	最小二乘法拟合	179
第 2 节	一个线性回归的统计模型	182
第 3 节	作为估计量的最小二乘法系数	184
第 4 节	回归模型的统计推断	186
第 5 节	回归模型的最大似然法估计	189
第 6 节	随机矩阵应用	191
注释		194
参考文献		199
译名对照表		201

第 **1** 章

矩阵、线性代数和几何向量

矩阵为大多数统计提供了一种自然诠释;线性代数是有关线性统计模型的代数计算;几何向量是一种非常强大的概念性工具,它在理解线性代数和标识线性模型等方面很有用。本章的目的是介绍有关矩阵、线性代数和几何向量的基本概念。这些相关话题在社会统计中应用广泛,且其编排形式相对于严格的数学表述来讲是非正式的。一方面,许多计算结果没有提供详尽的根据,而另一方面,这些根据均是提纲挈领的。对更深入的线性代数感兴趣的读者可以参看相关主题的教科书,以获得详细的解释(推荐阅读请参见本章末尾)。

第1节 | 矩阵

基本定义

矩阵是一组数字或数字变量的长方形阵列,例如,

$$\underset{(4 \times 3)}{\mathbf{X}} = \begin{bmatrix} 1 & -2 & 3 \\ 4 & -5 & -6 \\ 7 & 8 & 9 \\ 0 & 0 & 10 \end{bmatrix} \quad [1.1]$$

其更一般地表示为:

$$\underset{(m \times n)}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad [1.2]$$

像这样的 m 行 n 列矩阵可以称为 m 乘 n 阶矩阵,记做 $(m \times n)$ 。为方便起见,我有时候在矩阵的下方表示阶,如方程 1.1 和方程 1.2 所表示。矩阵的每一个元或者元素可以用它的行列下标表示,如 a_{ij} 表示矩阵 $\mathbf{A}$ 的第 i 行第 j 列元素。若矩阵为单一(实)数,则被称为“纯量”。有时为了简洁方便,我把矩阵中的典型元素放在一个括号里来表示矩阵,如 $\underset{(m \times n)}{\mathbf{A}} = \{a_{ij}\}$ 等价于方程 1.2。

一个只有一列元素的矩阵为列向量,如

$$\underset{(m \times 1)}{\mathbf{a}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

与之类似,一个只有一行元素的矩阵为行向量,

$$\mathbf{b}' = [b_1, b_2, \dots, b_n]$$

为了便于区分,我在行向量的元素间加上了逗号。

矩阵 $\mathbf{A}$ 的转置表示为 $\mathbf{A}'$, 它是将 $\mathbf{A}$ 的第 i 行转变为 $\mathbf{A}'$ 的第 i 列所构成,因此以方程 1.1 和方程 1.2 为基准,则有:

$$\underset{(3 \times 4)}{\mathbf{X}'} = \begin{bmatrix} 1 & 4 & 7 & 0 \\ -2 & -5 & 8 & 0 \\ 3 & -6 & 9 & 10 \end{bmatrix}$$

$$\underset{(n \times m)}{\mathbf{A}'} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

请注意, $(\mathbf{A}')' = \mathbf{A}$ 。通常,我所说的向量是指列向量(如上面的 $\mathbf{a}$),除非明确指出它是被转置的(如 $\mathbf{b}'$)。

N 阶矩阵,正如它的名字一样,拥有 n 行 n 列。元素 a_{ii} (例如, $a_{11}, a_{22}, \dots, a_{nn}$) 组成了方阵 $\mathbf{A}$ 的主对角线。对角线上所有元素的和叫做矩阵的“迹”:

$$\text{trace}(\mathbf{A}) \equiv \sum_{i=1}^n a_{ii}$$