



格致方法·定量研究系列

吴晓刚 主编

事件史和生存分析 (第二版)

[美] 保罗·D.埃里森 (Paul D. Allison) 著

范新光 译 刘孟宇 校

- ★ 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

62

格致方法·定量研究系列 吴晓刚 主编

事件史和生存分析 (第二版)

[美] 保罗·D.埃里森 (Paul D. Allison) 著
范新光 译 刘孟宇 校

SAGE Publications, Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

事件史和生存分析:第二版/(美)保罗·D.埃里森(Paul D.Allison)著;范新光译.—上海:格致出版社:上海人民出版社,2017.6
(格致方法·定量研究系列)
ISBN 978-7-5432-2096-6

I. ①事… II. ①保… ②范… III. ①统计分析-研究 IV. ①C812

中国版本图书馆 CIP 数据核字(2017)第 103351 号

责任编辑 贺俊逸

格致方法·定量研究系列
事件史和生存分析(第二版)

[美]保罗·D.埃里森 著
范新光 译 刘孟宇 校

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 5.75
字数 94,000
版次 2017 年 6 月第 1 版
印次 2017 年 6 月第 1 次印刷

ISBN 978-7-5432-2096-6/C·179

定价:32.00 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。研修课程设计方案体现“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

社会科学家分析许多感兴趣的现象时,关注的是事件的时间性:生命期望、在失业之后重新找到工作所需要的时间、婚姻存续的时长、累犯的间隔时间,等等。几乎所有关于事件时间的数据的一个关键特征是删截(censoring):例如,在一个关于累犯的研究中,研究者想记录犯人从监狱里被释放出来的一年内再次被捕的情况。尽管某些人最后有可能在接下来直到研究结束的时期内再次被捕,但也有一些初犯在这一时间段里并没有再次被捕。

研究此类事件发生时间的方法在许多学科领域都有进展,包括社会学中的事件史分析、工程领域的失效分析以及更广泛意义上的生物统计领域的生存分析。这些术语反映了不同学科的关注,但它们在本质上是相通的。一旦认识到它们基本的研究单位,我们便发现了研究事件时间性的一种共同方法。而这一方法就

是本书的主题。

有很多术语可以去称呼它,生存分析在社会科学领域里使用得最为广泛。在保罗·埃里森第二版重新命名的关于生存分析的小册子中,他向我们展示了一个关于这一主题涵盖甚广的介绍,同时将文笔集中在生存回归模型上,这一模型将事件发生时间和解释变量联系到了一起。生存回归模型——更深入地说,Cox比例机会模型——已经被应用于从生存数据得出因果推断的研究之中,这在社会科学中变得更为普遍,同时也应用到了对基于特殊设计的预测研究之中,例如对信用卡拖欠的研究中。

尽管埃里森教授阐述了很多种生存分析的方法,而且其中的一些方法本身还很复杂,但他对生存分析的解释是清楚明了与通俗易懂相结合的典范。本书的突出之处在于,它讨论了离散时间数据的方法,这在以往的讨论中往往被忽略;并且,它不仅关注单次独特事件(典型的例子如死亡),而且也关注了多重事件(包括“竞争性”事件,例如婚姻终结是由离婚或者死亡所导致),以及多次发生事件(例如失业的重复发生)。

埃里森教授这本书的第一版有一大批忠实的读者。我相信本书的第二版将同样会对新一代把生存分析应用到他们研究之中的社会科学家们有所裨益。

约翰·福克斯

第二版前言

两版间隔的 30 年是很长的一段时间。第一版(出版于 1984 年)已经有些过时,尤其是对于电脑软件和程序命令的描述。但是上一版的基本结构和大部分内容依然得到保留。我采用了一种不一样的策略:首先介绍的是简单地通过 logistic 回归就可以执行的离散时间方法,然后转向参数连续时间方法,紧接着是半参数 Cox 回归。

在新版中,最明显的变化是标题,从《事件史分析:纵贯数据的回归》(*Event History Analysis: Regression for Longitudinal Event Data*)到简单的《事件史和生存分析》(*Event History and Survival Analysis*)。尽管“事件史分析”这一术语出色地将这些方法广泛地应用到了所有类型的事件数据之上,但大多数研究者如今将其称为生存分析,这是因为他们的生物统计学背景使他们一直关注死亡事件发生的建模。

大部分数据集也得到了更新,主要因为原始数据在过去 30 年里已经丢失了。新的数据集可以在 <http://www.statisticalhorizons.com/resources/data-sets> 上下载,包括了 Stata 和 SAS 数据集。

以下是每章的主要变动和增补的概述:

第 2 章(“离散时间方法”)现在提供了关于删截更为详细的讨论,包括了如何对信息删截进行敏感性检验的例子。

第 3 章(“连续时间数据的参数法”)对加速失效时间模型给予了更多的关注。同时,本章更加强了对结果的解释和评估模型拟合度的方法。

第 4 章(“Cox 回归”)借助于程序陈述法和分割区间法,对时变解释变量进行了更加详细的介绍。同时,本书包含了对检验比例机会假定和处理连接数据(tied data)方法的更具扩展性的讨论。最后,有一个简短的部分介绍了如何用 Cox 模型做预测。

第 5 章(“多种类事件”)现在增加了检验多种类型事件的系数差别的例子。这里也增加了累计发生方程的新版块。这一方程是处理竞争性风险的一种替代性且越来越流行的方法。

第 6 章(“重复事件”)描述了一些在 1984 年还没有出现的新方法,包括处理事件计次数据的负二项模型、稳健标准误和共享异质性(随机效应)模型。基于间隔次数的方法被从基于发生次数的方法中区分开来。

所有例子的计算命令(使用 SAS 和 Stata)现在可以通过网上获得(<http://www.statisticalhorizons.com/resources.books> 和 www.sagepub.com/allisonevent)。我会尽可能地更新它们。

目 录

序	1
第二版前言	1
第 1 章 导言	1
第 1 节 事件史分析的难题	4
第 2 节 事件史方法综述	7
第 3 节 计算	13
第 2 章 离散时间方法	15
第 1 节 一个离散时间的例子	17
第 2 节 离散时间机会	19
第 3 节 logistic 回归模型	21
第 4 节 模型估计	23
第 5 节 生物化学例子的估计值	25
第 6 节 似然比卡方检验	27
第 7 节 离散时间的 logistic 方法存在的问题	29
第 8 节 删截	32
第 9 节 离散时间 vs. 连续时间	36

第 3 章	连续时间数据的参数法	37
第 1 节	连续时间的机会	39
第 2 节	参数比例机会模型	41
第 3 节	极大似然估计	44
第 4 节	一个实证案例	45
第 5 节	加速失效时间模型	50
第 6 节	评估模型拟合度	53
第 7 节	异质性的隐性来源	57
第 8 节	为什么选择参数模型?	59
第 4 章	Cox 回归	61
第 1 节	比例机会模型	63
第 2 节	部分似然	65
第 3 节	部分似然应用于累犯数据	67
第 4 节	时变解释变量	69
第 5 节	应用包含时变解释变量的模型	72
第 6 节	检验和放松比例机会假设	80
第 7 节	时间尺度原点的选择	89
第 8 节	离散时间数据的 Cox 回归	91
第 9 节	基于 Cox 模型的预测	94

第 5 章 多种类事件	97
第 1 节 多种类事件的分类	99
第 2 节 平行过程的估计	103
第 3 节 竞争性风险模型	105
第 4 节 竞争性风险的实例	107
第 5 节 不同种类事件间的依赖	113
第 6 节 累计发生函数	114
第 6 章 重复事件	119
第 1 节 重复事件的计数分析	122
第 2 节 基于间隔时间的方法	125
第 3 节 基于起点时间的方法	131
第 4 节 扩展	134
第 7 章 结论	137
附录	141
参考文献	151
译名对照表	154

第 **1** 章

导 言