

现代外国统计学优秀著作译丛

生存数据分析的 统计方法

STATISTICAL METHODS FOR
SURVIVAL DATA ANALYSIS

[美] ELISA T. LEE 著

陈家鼎 译
戴中维
孙山泽
房祥忠 校
刘力平

中国统计出版社

现代外国统计学优秀著作译丛

生存数据分析的统计方法

(第二版)

[美] ELISA T. LEE 著

陈家鼎 戴中维 译

孙山泽 房祥忠 刘力平 校

中国统计出版社

(京) 新登字 041 号

图书在版编目 (CIP) 数据

生存数据分析的统计方法 / (美) E. T. 李著 (Lee, E. T.);
陈家鼎, 戴中维译. —北京: 中国统计出版社, 1998. 3
(现代外国统计学优秀著作译丛)
书名原文: Statistical Methods for Survival Data Analysis
ISBN 7-5037-2447-1

I. 生…

II. ①李… ②陈… ③戴…

III. 生存分析—统计—方法

IV. 0213

中国版本图书馆 CIP 数据核字 (97) 第 15251 号

著作权合同登记: 图字 01-97-0562 号

中国统计出版社出版

(北京三里河月坛南街 75 号 100826)

新华书店经销

科伦克三莱印务(北京)有限公司印刷

*

850×1168 毫米 32 开本 19.25 印张 48 万字

1998 年 4 月第 1 版 1998 年 4 月北京第 1 次印刷

印数: 1—4000 册

*

定价: 42.40 元

(版权所有 不得翻印)

版权公告:

Copyright notice:

生存数据分析的统计方法

Statistical Methods for Survival Data Analysis

(Second Edition)

[美] ELISA T. LEE

Copyright©1992 by John Wiley & Sons, Inc.

All Rights Reserved.

Authorized translation from English language edition

published by John Wiley & Sons, Inc.

本书中文版翻译、出版专有权归国家统计局

统计教育中心和中国经济出版社

现代外国统计学优秀著作译丛 专家委员会

主任:

翟立功 国家统计局副局长

副主任:

贺 铿 国家统计局副局长

王吉利 国家统计局统计教育中心主任

委员:

刁锦寰 美国芝加哥大学商学院 教授

吴建福 美国密西根大学统计系 教授

孟晓犁 美国芝加哥大学统计系 博士

张尧庭 上海财经大学数量经济研究所 教授

茆诗松 华东师范大学数理统计系 教授

陈家鼎 北京大学概率统计系 教授

郑祖康 复旦大学统计与运筹系 教授

吴喜之 南开大学数学系 教授

袁 卫 中国人民大学统计系 教授

邱 东 东北财经大学统计系 教授

郝国印 国家统计局统计教育中心副主任

谢鸿光 中国统计出版社副总编

办公室:

刘启荣 国家统计局统计教育中心教材处处长

严建辉 中国统计出版社第二书籍编辑部主任

李 毅 国家统计局统计教育中心教材处副处长

出版说明

为了加强对国外统计理论与实践的研究和了解，全面反映国外统计科研和教学的发展，促进我国统计教学改革和教材内容更新，在国家统计局领导的大力支持下，全国统计教材编审委员会组织翻译出版了这套“现代外国统计学优秀著作译丛”。

随着我国社会主义市场经济体系的逐步建立，统计教育正面临着十分严峻的挑战。一方面，在社会主义市场经济条件下，不论国家的宏观经济调控还是企业的生产经营管理，都要求准确地把握市场运行的态势，科学地分析经济中各种错综复杂的关系，因而，对统计信息的需求越来越大，对统计人才的业务素质提出了更高的要求；另一方面，我国过去的统计教育模式是按为高度集中的计划经济管理体制服务的要求建立的，培养的统计人才的知识结构比较单一，难以适应经济体制、统计体制改革的需要。为使统计人才的培养适应建立社会主义市场经济体制的需要，满足二十一世纪现代化建设的要求，缩小与国际先进水平的差距，基础在教育，关键在教材。在继续组织有关专家、学者编写一批反映国内统计科学和统计实践发展的新教材的同时，必须尽快引进并翻译出版一批外国先进统计教材。这是学习外国先进统计知识的一种直接而且十分有效的方式，对于推动国内统计教材内容更新和教学改革，造就一大批具有渊博知识和多方面

业务技能的复合型人才，具有十分重要的意义。

为了做好这套丛书的翻译出版工作，全国统计教材编审委员会成立了现代外国统计学优秀著作译丛专家委员会，对国外统计著作的出版和使用情况进行了调查研究，分析了国内对外国统计教材的需求，在此基础上制定了翻译著作选题规划。在这套丛书的翻译出版过程中，我们得到了国内外有关专家、有关院校统计系和国外有关出版公司的大力帮助和支持，在此表示衷心的感谢。

全国统计教材编审委员会

1995年7月

译者序

生物和人的生存时间的评估和预测是生物学和医学的重要研究对象，这方面的研究涉及基础科学和技术科学的许多领域，其中数学方法特别是统计方法起着重要的作用。从数学角度来看，这种研究是对一个或多个非负随机变量（生存时间）进行统计分析。这种统计分析已形成现代数理统计学的重要分支——生存分析（*Survival Analysis*）。

生存分析就是根据试验或调查得到的数据，对生物或人的生存时间进行分析和推断。和任何其它数学方法一样，生存分析的方法和理论有广泛的应用，不限于生物学和医学领域，而且可应用于工程科学（如可靠性工程）、社会学、心理学、经济学、保险精算学等等。“生存时间”一词应作广义的理解，乃指自然界、人类社会或技术过程中某种状态的持续时间。生存分析含有许多实用的方法和丰富的理论。随着医疗实践、工程实践及其它领域的推动，不断有新的统计方法出现，应用范围越来越广。

在统计教育中如何使学生易于学习和掌握生存分析的基本方法，是一项重要任务。首先要有一本好的教材。如所周知，要写出具有现代科学水平、紧密联系我国实际又适合教学需要的教材很不容易。在我国当前的形势下，国家统计局大力挑选并组织翻译外国优秀教材是一项有深远意义的重要措施，将有力地推动我们学习和借鉴外国的工作经验。

现在摆在读者面前的这本书—E. T. Lee 著的“生存数据分析的统计方法”，就是一本优秀教材。该书内容丰富、选材恰当、文字通俗，层次分明。

从内容上看有下列特点：

1. 突出统计方法。大力介绍生存分析中各种统计方法,讲解这些方法处理什么样的问题、如何使用以及在使用时要注意的问题。基本上不进行数学推导,但介绍统计思想使人们理解所提出的统计方法是有道理的。

2. 实用性强。在正文和习题里都列举了大量的实际数据例子,以便读者学习和掌握所介绍的各种统计方法。虽然这些例子大都来自医学,相信读者会平行思考其它领域的同类问题。

3. 要求的数学知识少。只要读者知道微积分和矩阵的基本知识并对概率有初步了解就可以阅读全书,并学会运用所介绍的统计方法。

应该指出,外国的生存分析教材和专著还有多种,各有特点,在本书的第一章里有所评述。总的说来,那些书与本书不同,要求读者有较高的数学基础。此外,中文教材有“生存分析与可靠性引论”(陈家鼎著,安徽教育出版社,1993),是供概率统计专业的研究生用的,要求读者学过测度论,在内容上重点在可靠性工程与参数方法,对非参数方法介绍较少,对医学上的例子也极少论及。

总之,这本书是生存分析的优秀入门读物,适合学生和实际工作者阅读。对统计方法的数学理论感兴趣的读者,可去看下列著作:

Fleming, T. 和 Harrington, D. (1991), *Counting Processes and Survival Analysis*, John Wiley.

Andersen, P., Borgan, O., Gill, R. 和 Keiding, N. (1993), *Statistical Models Based on Counting Processes*, Springer-Verlag.

这两本书内容丰富,但要求读者具有测度论和现代随机过程论的知识并受过严格的数学训练。

本书译稿承蒙北京大学孙山泽教授、刘力平副教授和房祥忠博士进行审校。他们提出了许多修改意见,并帮助译者进行修改。

我们在此向他们表示感谢。

限于译者的水平，译文肯定还有不当之处，欢迎专家和读者批评指正。

最后要说明的是，书中出现的外文人名一律不进行翻译，保留原来的形式。外文人名之后若附有单个年份则常常表示该人在这一年发表的论文（或书）。例如，Bolin 和 Greene (1986) 是指 Bolin 和 Greene 这两个人发表于 1986 年的论文，Freireich et al. (1974) (或 Frireich 等人 (1974)) 是指 Freireich 等人发表于 1974 年的论文或书。这些论文的题目（或书的名称）和所在的刊物（或出版单位）均可从本书最后部分——参考文献目录中找到。这是读者在阅读本书时要注意的。

陈家鼎 戴中维 谨识
北京大学数学学院概率统计系
1997 年 1 月

序

生存数据一词已广泛用于涉及一定事件的时间数据，例如死亡时间、缓解时间及疾病发作时间。在过去十年里，生存数据分析的统计方法的应用已经从生物医学和可靠性研究推广到犯罪学、社会学、市场学及健康保险业务等领域。“生存数据分析的统计方法”的第二版就是满足各方面需要的一本书，书中包括了分析生存数据的各种合适的方法。这本书是为生物医学研究者、统计学家、流行病学家及其它领域需要或有兴趣分析生存数据的研究者而写的，它包括了生存分析中大多数常用的方法（参数方法和非参数方法），可以用作参考书或教科书。此外，它对医学临床试验的计划和设计提供了指导准则。这些指导准则的一部分也可用于其它类型的研究（例如流行病学研究）。本书中所叙述的大多数统计方法可应用于临床研究、流行病研究、社会科学研究及其它领域的研究。本版仍定位于应用，保持最低程度的数学水平。在某些节里需要微积分和矩阵代数的某些知识。但是，只学过大学代数的读者可以理解本书的大部分内容。除了在正文中提供了大量实际生活的例子外，我们还在习题里提供了一些规模较大的数据，以供读者用作练习。

本版在内容上比第一版有下列改进：

1. 除了临床生命表分析外，还讨论了人口生命表。
2. 介绍了标准化死亡比率(SMR)和标准化发病比率(SIR)。

3. 添加了对含有删失数据的生存数据建模的拟合优度检验。

4. 对Cox比例危险率模型一节增加了内容,包括分层及比例危险率假定的验证。

5. 讨论了优比 (*odds ratio*) 和线性罗辑斯提 (*logistic*) 模型的系数之间的关系。这种回归模型被推广到病症控制的研究中。还介绍了这种罗辑斯提回归方法的拟合优度检验。

6. 在临床试验中确定样本量的方法也包括确定试验的截止时间。

7. 介绍了重复显著性检验及分组序贯设计。

8. 修订后的文献目录包含了大量近来发表的论文。

在过去十年里,生存数据分析的很多计算程序已编制出来(包括微机上用的程序)。在本版的各章里,我们向读者指出了有关的计算机程序。由于这些程序可广泛找到,我们把第一版中收录的大部分Fortran程序都去掉了。只把产生伽玛概率纸的计算机程序保留下来,因为这个程序很符合过去十年间第一版读者的需要。

我很感谢使用这本书第一版的很多研究者、教师和学生。我对那些想得到本书但在过去几年里一直未能得到的人士表示歉意,并赞赏他们在期望得到本书时所显示出的耐心。他们推动我为第二版而努力工作。我非常感谢他们对本书的支持。很多同事和读者对第二版提出了有帮助的建议,这里不可能一一指出。特别要感谢Min Lu博士、Wansu Chen先生和J. L. Yeh博士,他们帮助我研究书中所提出的例子并查找参考文献目录中所列的那些论文。John Wiley & Sons公司的Beatrice女士在她退休之前对出本书第二版很热心。Kate Roach女士对本版不断提供支持。我高度评价并感谢他们的热情和意见。

最后,我感谢我的家庭——丈夫Sam和孩子Vivian及Jen-

nifer——所给予的关心和支持。如果没有他们经常的可靠的帮助，这本书是不可能完成的。

E. T. 李
(Elisa T. Lee)

1

引 论

1.1 前言

这本书是为生物医学研究者、流行病学家、咨询统计学家、需要生存数据分析入门教材的大学生及其他对生存时间研究有兴趣的人而写的。本书是论述生存数据的统计分析方法，生存数据来自动物的实验室研究或者来自急性病患者或慢性病患者的临床研究及流行病学研究。

生存时间可以广泛地定义为一给定的事件出现的时间。这个事件可以是疾病的发生、一种处理（治疗）的反应、病情复发或死亡。生存时间可以是无肿瘤时间，从一种治疗开始到有反应的时间，缓解时间的长度或出现死亡的时间。生存数据可以包括生存时间、对治疗的反应以及与反应、生存及疾病发生有关的病人特征。过去，生存数据的研究集中在预测反应的概率、生存概率、平均寿命以及比较试验动物的生存分布或比较病人的生存分布。近些年来，研究与反应、生存及疾病发生有关的风险和预后(*prognostic*)因素的识别已经变成生存数据分析中有同样重要性的组成部分。在本书中我们只对生物医学中的生存数据进行研究，虽然所有的方法都可应用于工业可靠性、社会科学和商业。在后面这

些领域生存数据的例子是：电子设备（元件或系统）的寿命（可靠性工程），重罪犯人的假释时间（犯罪学），首次婚姻的持续时间（社会学），报纸或杂志的篇幅和订费（市场学），工人们的补偿索赔（保险业）和他们的各种影响风险或预后因素。

1.2 删失数据

很多研究者认为生存数据分析只不过是两种统计方法对一个特殊问题的应用：若已知生存时间的分布符合正态分布则用参数统计方法，若生存分布未知则用非参数方法。如果所有个体的生存时间均准确知道，则这个看法有道理。但是，某些生存时间不能准确知道，而且生存分布常常与正态分布相差很远。因此，需要新的统计技巧。生命科学中的生存数据有一个最重要的特点：在研究期间结束时某些个体身上还没有出现我们关心的事件。例如，在研究周期结束时某些病人仍然活着或处于缓解状态，这些个体的确切生存时间是不知道的。这些被称为删失观测或删失时间 (*censored time*)。在研究过程中的某个时候对某些病人失去跟踪观测时也可发生这种情况。

删失有三种类型。

1. I 型删失

动物研究通常是以有固定数目的动物接受一种或多种处理 (*treatment*) 开始。由于时间和费用受到限制，研究者常常不能等到所有动物死亡。一种选择是在一个固定时间周期内观测（比如六个月），在截止时间之后仍可能有些动物活着，但不继续观测了，这些动物的生存时间是不知道的只知其不小于研究期间的长度，这些称为删失观测值。对于在试验期间死亡的动物，所记录的生存时间是从试验开始到其死亡的时间，这些数叫做准确的或非删失观测值。某些动物可能意外失踪或死亡，它们的生存时间——从试验开始到意外失踪或死亡——也是删失观测。在 I 型删失里，

没有意外的失踪，所有的删失观测值均等于研究期间的长度。

例如，假设有六只老鼠（*rat*）放在致癌物质的环境中——在其脚趾注入肿瘤细胞。我们观测一定大小的肿瘤的出现时间。研究者决定 30 周后停止试验。图 1.1 描述了老鼠 A, B, D 分别在第 10 周, 第 15 周, 第 25 周出现肿瘤；老鼠 C 和 E 到研究结束时仍没有肿瘤，它们的无肿瘤时间至少是 30 周；老鼠 F 在 19 周后意外死亡，身上并无肿瘤。这样，生存数据（无肿瘤时间）是 10, 15, 30+, 25, 30+, 19+（周）、这里 + 号表示是删失观测。

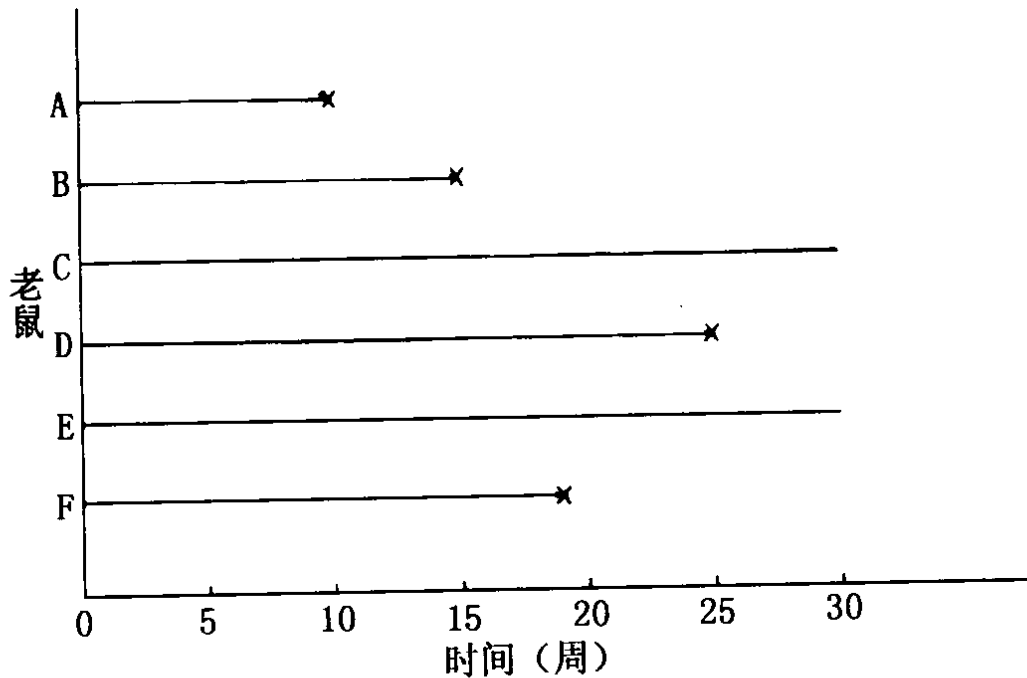


图 1.1 I 型删失数据的例子

2. II 型删失

在动物研究中的另一种选择是试验进行到有一固定数目的动物死亡为止。在这种情况下，如果没有意外的失踪，删失观测值等于最大的非删失观测值。例如，在六只老鼠的试验里（见图 1.2），研究者决定有四只出现肿瘤就停止试验。那么，生存时间（或无肿瘤时间）是 10, 15, 35+, 25, 35, 19+（周）。

3. III 型删失

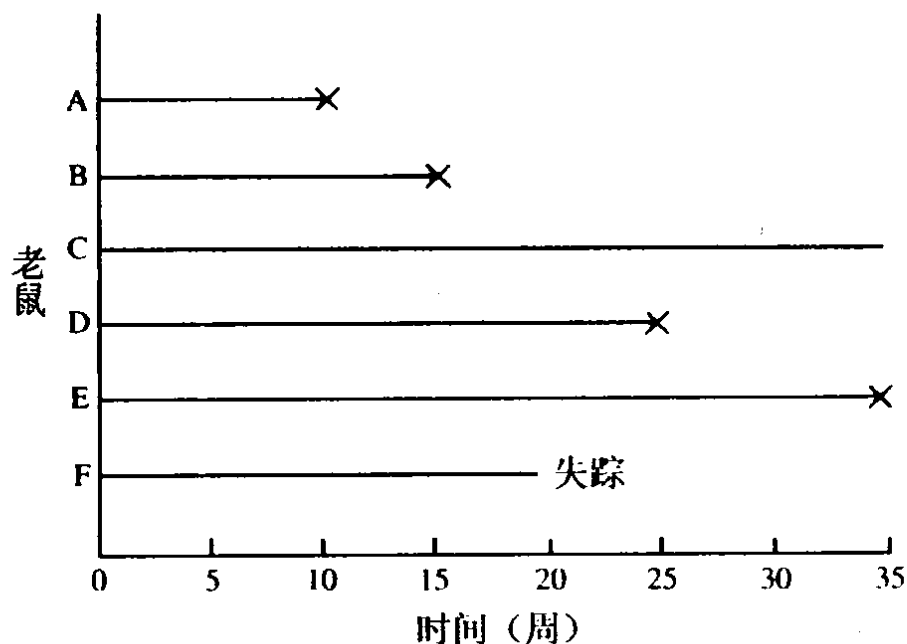


图 1.2 ■ 型删失数据的例子

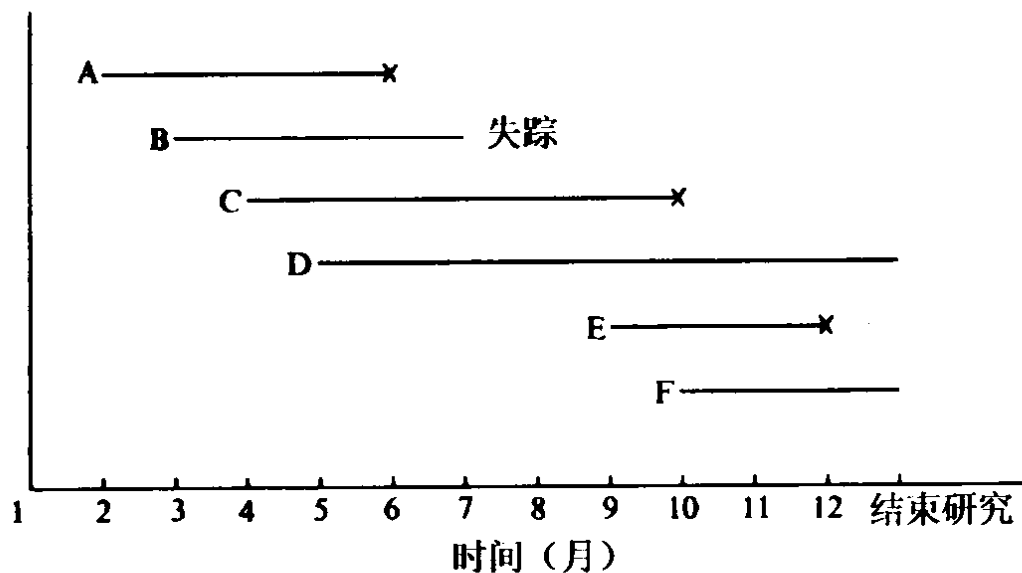


图 1.3 ■ 型删失数据的例子

在大多数临床研究中，研究期间是固定的，病人在此期间的不同时间进入研究，某些人可能在研究期间结束之前死亡，他们