

现代外国统计学优秀著作译丛

寿命数据中的 统计模型与方法

STATISTICAL MODELS AND METHODS
FOR LIFETIME DATA

[加拿大] J. E. Lawless 著

茆诗松

濮晓龙 译

刘 忠

葛广平 校

中国统计出版社

现代外国统计学优秀著作译丛

寿命数据中的统计模型与方法

[加拿大] J. F. Lawless 著

茆诗松

濮晓龙 译

刘 忠

~~葛广平~~ 校

中国统计出版社

版权公告:

Copyright notice:

寿命数据中的统计模型与方法
STATISTICAL MODELS AND METHODS
FOR LIFETIME DATA

[加拿大] J. F. Lawless

Copyright ©1982 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统计出版社

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制的逐步建立，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服务的要求而建立的，培养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

十分有效的方式，对于推动我国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工作中，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

在医学、生物学和保险学中很早就在研究各种各样的寿命数据的统计分析方法。到本世纪五十年代，急需提高工业产品的可靠性，吸引了很多统计学家研究各种类型的寿命数据，发表成千上万篇论文，使寿命数据分析的理论、方法和应用都发展相当迅速，短短二十余年内寿命数据分析已形成相当完善的统计分支，就在这个背景下，加拿大统计学家 J. F. 劳立斯在 1981 年摘其精华，归纳成书，畅销不衰，至今仍是世界上从事寿命数据分析的主要参考书和这个领域内研究生课程的教科书。

这本书有三个特点：第一是以讲方法为主，理论为辅，所叙理论也是为说明方法为限，且大多集中在附录里，书中所列方法至今仍是处理寿命数据的主要方法。第二是注意讲解引出每一个方法的模型和假设，这就把每一个方法的应用条件说清楚了。每当一个模型下有多种方法时，作者还做了比较，说明优劣，给出选用指南。第三是实用性强，书中几十个实际数据例子都给出详尽分析和数值计算，列出最后结果，全过程很富有启发性。我们相信实例是连接方法与应用的最佳沟通方式。

这本书是很适合工程界和医药界实际从事寿命数据分析的人们阅读，因为他们一般都具有概率和统计的基础知识。这本书也可作为这个领域研究生的教科书，本书的前五章可作为大学高年级学生的选修课教材。

这本书共 10 章和 7 个附录，第 1~2 章是刘忠所译，第 3~5 章是茆诗松所译，最后五章与附录是濮晓龙所译，最后由茆诗松统稿。我们翻译力求忠于原著而又适应我国实情，但由于经验不足，水平有限，翻译不当之处在所难免，敬请批评指正。

最后我们要感谢上海大学葛广平教授，以他精深的专业和熟练的英语校阅了全文，提出很多宝贵意见，我们都一一加以修正，这为本书译稿增色不少。

茆诗松

1996 年 10 月于华东师范大学

序 言

寿命数据和响应时间数据的统计分析已成为工程、医学和生物科学等领域中的统计学家和实际工作者们很关心的一个领域，这个领域在近几年发展迅速，在这个主题上发表的文章不仅在统计文献中，也可以在统计之外的某些学科的文献中找到。这本书归纳了寿命数据分析的结果，并综合介绍了一些重要的模型和方法。

本书的目的是对该领域进行较全面的介绍，而不将兴趣过分集中在某个应用领域。这本书中很多例子虽然是来自工程和生物科学，然而它们的方法可广泛使用。这本书含有我认为在寿命数据方法研究中最重要的一些专题。它包含各种参数模型及其有关的统计方法，非参数方法和图方法。为保持这本书有一个合理的篇幅，我不得不简述或完全省略一些在处理细节方面很有用的专题。对其中一些专题我们给出参考文献或在每章结尾处的问题与补充一节中简略地提及。

这本书可作为对寿命数据分析有兴趣的人们的参考书，也可作为这个领域中研究生课程的教科书。概率和统计推断的基本知识是应具备的，不过我仍试图精心地引出方法所依赖的模型和假设，并说明方法是怎样建立的。此外，几个附录综述了部分读者可能不熟悉的一些统计理论。对多数方法都用数值例子给以说明，这本书还有很多含有实际数据的例子。每章都以“问题与补充”这一节为结束。这些问题与补充不仅为本章内容提供练习，也为本章讨论的专题给出补充和推广。对有兴趣研究寿命数据分析方法

的读者本书给出了十分广泛的近代的参考文献和尚未解决的问题。

第1章含有寿命分布的入门资料和几个最重要的参数模型的归纳。截尾被引入，并考虑了对截尾的几种统计推断方式。在第2章中，讨论单变量寿命数据的某些分析方法以及如何获得分布特征的非参数估计；寿命表和图方法发挥关键作用。第3、4和5章分别对几个重要的参数模型给出推断，包括指数、威布尔、伽玛、对数正态和广义伽玛分布。在第6章中通过基于这些分布的回归模型把推断延伸到含有伴随变量的问题上去。第7和第8章提出无分布方法，第7章对比例危险回归模型给出分析方法，第8章对单个和多个样本问题给出无分布方法。在第9章中对寿命分布模型考虑其拟合优度检验。第10章含有两个重要专题：多变量模型和随机过程模型的简短讨论，对这二个专题还不可能给出广泛的处理方法。这本书中有几节是带星号的(*)，这些是技术性的讨论，初次阅读可以略去。

关于本书提及的方法最后有一个说明，正如现代统计中经常用到的那样，计算机如果不是必须的，也是很有用的工具。对某些问题，可以不要求使用计算机的方法，但更多场合计算机是不可少的。只要可能，我都会指出方法的计算特点，并在附录里补充计算方面的一些材料。

这本书的部分工作是在1978年到1979年的休假期间在伦敦Imperial College和University of Reading做的。对他们的款待深表感谢。我同样要对两位非常优秀的打字员，Annemarie Nittel和Lynda Hohner表示我的感谢，她们在整理我的手稿中作了长时间的辛勤劳动。

J. F. 劳立斯

Waterloo, Ontario

1981年6月

目 录

序言

1 基本概念和模型	1
1.1 引言	1
1.2 寿命分布的基本概念	8
1.2.1 连续模型	8
1.2.2 离散模型	10
1.2.3 关于危险函数的几点注释	11
1.3 一些重要的模型	13
1.3.1 指数分布	14
1.3.2 威布尔分布	16
1.3.3 伽玛分布	20
1.3.4 对数正态分布	24
1.3.5 广义伽玛分布	25
1.3.6 其它寿命分布模型	27
1.3.7 混合模型	28
1.3.8 回归模型	28
1.3.9 多元寿命分布	29
1.3.10 关于选择寿命分布模型的几点注释	30
1.4 截尾和统计方法	31
1.4.1 截尾类型	31
a. I型截尾, b. II型截尾, c. 随机截尾, d. 更一般的截尾过程*	
1.4.2 基于截尾数据的统计推断	42

1.5	问题与补充	44
2	寿命表、图及有关方法	52
2.1	引言	52
2.2	寿命表	53
2.2.1	标准寿命表方法	54
2.2.2	标准寿命表方法: 理论*	59
2.2.3	寿命表估计的其它方法*	68
2.3	生存函数的非参数估计	70
2.3.1	乘积限估计	70
2.3.2	乘积限估计的性质*	75
2.3.3	经验累积危险函数	78
2.4	图法	80
2.4.1	生存函数或危险函数的估计用图	80
2.4.2	概率图和危险图	84
2.5	参数的最小二乘估计	91
2.6	问题与补充	93
3	指数分布的统计方法	100
3.1	单样本: 完全数据或Ⅰ型截尾数据	101
3.1.1	检验和置信区间	103
3.2	单样本: Ⅰ型截尾数据	105
3.2.1	有关 θ 的推断	105
3.2.2	近似方法的适用性*	110
3.2.3	讨论	111
3.3	多个指数分布的比较	111
3.3.1	Ⅰ型截尾数据	111
3.3.2	Ⅰ型截尾数据	115
3.4	试验方案与寿命试验方法	117
3.4.1	Ⅰ型截尾(无替换)寿命检验方案	118
3.4.2	某些其它的寿命检验方案	121
	a. 有替换Ⅰ型截尾, b. 有替换Ⅰ型截尾, c. 无替换	

	I 型截尾, d. 序贯方案	
3.4.3	试验方案的总注释	125
3.5	双参数指数分布	126
3.5.1	II 型截尾数据	126
3.5.2	I 型截尾数据	131
3.6	重要注释	133
3.7	问题与补充	134
4	威布尔和极值分布的推断方法	141
4.1	单样本: 完全或 I 型截尾数据	142
4.1.1	点估计	142
4.1.2	置信区间和检验	147
	a. 位置和尺度参数的置信区间, b. 用条件方法获得的置信区间, c. 获得精确置信区间的其它方法, d. 枢轴量分布的近似, e. 生存(可靠度)函数的估计, f. 选择方法的评论	
4.2	单样本: I 型截尾数据	170
4.2.1	极大似然估计	171
4.2.2	基于 m. l. e. 的大样本方法	172
4.2.3	似然比方法	175
4.2.4	关于大样本方法的比较和注释	178
4.3	多个威布尔或极值分布的比较	180
4.3.1	威布尔形状(极值尺度)参数的比较	181
	a. II 型截尾样本下二个形状参数的比较, b. 任意截尾样本: 似然比方法, c. 全是 II 型截尾时的简单检验	
4.3.2	多个威布尔尺度参数或分位数的比较	188
4.4	其它一些问题	192
4.4.1	三参数威布尔分布的推断	192
4.4.2	威布尔模型下的寿命试验方案	197
4.5	问题与补充	199
5	几个其它模型的推断方法	206

5.1	伽玛分布	206
5.1.1	点估计	207
5.1.2	参数的区间估计和检验: 无截尾数据	213
5.1.3	截尾数据或无截尾数据的似然比方法	218
5.2	对数正态和正态分布	224
5.2.1	点估计	224
5.2.2	区间估计和检验	231
	a. 完全样本的精确方法, b. II 型截尾样本的精确方法, c. 截尾样本的近似方法	
5.3	广义伽玛分布	242
5.3.1	极大似然估计和大样本方法	246
5.3.2	完全样本的精确方法	254
5.4	多项式危险函数和其它模型	258
5.4.1	具有多项式危险函数的模型	259
5.4.2	混合模型和其它模型	262
5.5	来自连续模型的分组数据	265
5.5.1	极大似然	265
5.5.2	最小二乘法	267
5.6	问题与补充	271
6	参数回归模型	278
6.1	引言	278
6.1.1	比例危险模型	281
6.1.2	$\log T$ 的位置-尺度模型	282
6.2	残差分析和其它模型检查	283
6.2.1	寻找模型	284
6.2.2	残差检查	285
6.2.3	讨论	288
6.3	指数回归模型	288
6.3.1	极大似然方法	289
6.3.2	精确条件方法	296
6.3.3	在每个 x 处有多个观测数据	302

6.4	威布尔和极值回归模型	304
6.4.1	极大似然方法	305
6.4.2	在每个 x 处有多个观测数据	313
6.5	正态和对数正态回归模型	320
6.5.1	极大似然估计 (截尾数据)	320
6.5.2	区间估计和假设检验	323
6.5.3	其它方法	328
6.6	伽玛和对数伽玛回归模型	328
6.7	最小二乘估计	335
6.7.1	普通最小二乘估计	335
6.7.2	在每个 x 处有多个观测数据	340
6.8	问题与补充	342
7	比例危险及有关的回归模型的无分布方法	350
7.1	引言	350
7.2	连续 PH 模型的统计方法	351
7.2.1	β 的估计与检验	352
7.2.2	两个或多个寿命分布的比较	355
7.2.3	似然函数 (7.2.3) 的合理性*	361
7.2.4	生存函数的估计	366
7.2.5	使用比例危险模型的数据分析	370
7.3	分组数据的回归方法	379
7.3.1	有回归变量的寿命表模型	379
7.3.2	分组数据的分布相等的检验	388
7.4	其它一些问题	397
7.4.1	基于偏似然方法的效*	397
7.4.2	与时间有关的回归变量	399
7.4.3	偏离比例危险的评定	401
7.5	问题与补充	402
8	非参数方法和无分布方法	408

8.1	生存函数和分位数的非参数估计	408
8.1.1	生存函数的区间估计	408
8.1.2	分布分位数的区间估计	415
8.2	分布比较的秩检验	418
8.2.1	m 个样本问题的线性秩检验	418
8.2.2	截尾数据下的指数顺序记分检验 (对数秩检验) ...	426
8.2.3	截尾数据下的广义 Wilcoxon 检验	429
8.2.4	讨论	431
8.3	问题与补充	434
9	拟合优度检验	437
9.1	拟合检验的一些通用方法	438
9.1.1	基于 EDF 的检验	438
9.1.2	平滑拟合优度检验	445
9.1.3	分组数据的拟合检验	447
9.2	特定分布下的拟合检验	451
9.2.1	指数分布的拟合检验	451
9.2.2	威布尔分布或极值分布的拟合检验	459
9.2.3	正态分布或对数正态分布的拟合检验	463
9.2.4	其它分布	467
9.2.5	回归模型的拟合检验	469
9.2.6	模型识别和假设检验	471
9.3	模型选择和统计方法	476
9.4	问题与补充	477
10	多变量模型和随机过程模型	483
10.1	多变量模型	484
10.1.1	多变量寿命时间分布	486
10.1.2	具竞争死亡原因的模型	492
10.2	某些随机过程模型	499
10.2.1	泊松过程	502
10.2.2	非齐次 (时间相依的) 泊松过程	503

10.2.3	更新过程	504
10.2.4	一个例子 (例 10.2.1 再考察)	505
10.2.5	Markov 过程	509
10.2.6	半 Markov 过程	510
10.3	问题与补充	512

附 录

A.	记号、缩写及其它概念	517
B.	伽玛函数及有关的一些函数	519
C.	渐近方差公式	522
D.	次序统计量	524
E.	极大似然估计的大样本理论	528
F.	极大似然的优化方法	533
G.	位置—尺度参数模型中的推断	538

参考文献	544
主题词索引	579