

存活分析

林建甫

國立台北大學統計系

存活分析 / 林建甫著. — 初版. — 台北市:

雙葉書廊, 2008. 04

面: 公分

參考書目: 面

含索引

ISBN 978-986-6672-08-8(平裝)

1. 生物統計 2. 醫學 3. 統計

306.13

97005978

©2008 林建甫

書名: 存活分析

著者: 林建甫

著作權與版權所有人: 林建甫

出版社: 雙葉書廊有限公司

發行人: 張福隆

地址: 台北市羅斯福路三段 269 巷 12 號 1 樓

<http://www.yehyeh.com.tw>

電話: 886-2-2368-4198

傳真: 886-2-2365-7990

出版日期: 西元 2008 年 4 月初版一刷

ISBN: 978-986-6672-08-8(平裝)

本書著作權與版權屬於著作者所有, 受中華民國著作權法保護, 未經著作者書面同意, 不得以任何形式轉載, 複印, 翻印或重製.

作者 林建甫
Chien-Fu Jeff Lin, MD. PhD.
台北大學統計系醫學統計學程
cflin@mail.ntpu.edu.tw
<http://web.ntpu.edu.tw/~cflin>

經歷 台北榮民總醫院生物統計顧問
智策市場研究公司統計顧問
卡方醫學統計諮詢工作室顧問
亞洲環宇醫學編譯公司統計顧問
台北大學休閒運動與管理學系兼任助理教授
輔仁大學體育研究所兼任助理教授
高雄長庚醫院骨科主治醫師
台北榮民總醫院骨科部醫師

學歷 美國密西根大學生物統計研究所碩士, 博士
高雄醫學院醫學系學士

興趣 存活分析, 臨床試驗, 生物資訊, 統計諮詢

序

在醫學研究中, 存活時間資料為一種常見的研究資料, 例如病患從診斷出癌症到死亡之間的存活時間. 本書主要目的是討論一些常用來分析存活時間的方法, 在醫學統計中, 這些方法常稱為存活分析. 一般而言, 存活分析方法是用來研究或分析樣本所觀測到的某一段時間長度之分配. 一段時間長度通常是從一特定事件起始之起始時間點算起, 計算到某一特定事件發生的時間點為止, 通常稱為事件時間或存活時間. 事件時間的資料出現在不同領域中, 譬如醫學中的癌症存活率; 公共衛生中的死亡率; 流行病學中的生命量表; 保險統計學及人口統計學中的生命統計資料; 工程學中的可靠度分析; 社會學中的事件歷史分析, 研究結婚時間到離婚時間, 到再婚時間, 人口居住時間與流動時間; 商業研究中, 市場研究之消費者對特定商品購買時間, 客戶忠誠度的時間; 公司企業的存活時間; 經濟研究中的失業, 從就業時間到失業時間, 到再就業時間等; 法學研究中, 犯罪嫌疑人從犯罪時間到被捕時間, 犯罪嫌疑人從被捕時間到起訴時間, 從起訴時間到定罪時間, 從假釋時間到再犯時間等.

在醫學研究中, 常常收集病患從診斷出癌症到死亡之間的存活時間, 但是, 並非所有進入研究的觀察對象, 其被追蹤的時間都足夠長, 所以研究者並不能夠觀察到所有研究對象最後的結果, 及所有發生死亡事件時間點. 這樣的資料通常稱為設限資料. 此外, 病患常常具有一些危險因子, 進而影響存活的機會; 但是, 許多影響事件之發生或影響存活時間長短的危險因子, 是在追蹤的時間內才發生, 稱為時間相依危險因子, 這樣的存活時間資料, 必須要有特殊的統計方法來分析. 這些特殊的統計方法, 通稱為存活分析, 這是一個籠統定義的統計名詞. 存活分析相關書籍中, 有許多可做為參考, 如 Lawless (2002) 對可靠度分析之參數模型有詳盡的討論. Cox 與 Oakes (1984) 則是以精簡的手筆, 觸及各項存活分析議題. Collett (2003) 討論單變量存活分析, 以應用為主. Hosmer 與 Lemeshow (1999) 討論單變量存活分析, 有一章討論多變量存活分析, 以應用為主. Klein 與 Moeschberger (2003) 主要討論單變量存活分析, 有一章討論多變量存活分析, 雖以應用為主, 但對存活理論有深入的介紹. Kalbfleisch 與 Prentice

(2002) 對單變量與多變量存活分析都有討論, 應用與理論並重, 但較為艱難. Fleming 與 Harrington (1991), 以及 Andersen 等 (1993) 則是以計數過程為架構, 說明存活分析之統計理論. Therneau 與 Grambsch (2000) 討論 Cox 延伸模型與多變量存活分析. Hougaard (2000) 則專門討論多變量存活分析. 但這些書籍對長期從事臨床醫學研究的醫師, 較難以一一閱讀與理解.

本書主要寫作對象為從事資料分析的統計人員與從事臨床醫學研究的醫師, 本書所使用的數學內容僅包含加法, 乘法, 與簡單之微積分, 這些內容等同於在台灣的高中與大一的基礎數學程度. 閱讀本書的統計專業人員, 統計系研究生或大學生, 最好具備數理統計, 迴歸分析, 變異數分析等知識, 部分章節則需一些類別資料分析與縱貫資料分析等基本知識. 至於從事臨床醫學研究的醫師閱讀本書, 最好具備基本醫學統計的知識, 並建議臨床醫學研究的醫師閱讀本書時, 閱讀本書時, 以閱讀文字為主, 閱讀方程式演算過程為輔, 主要須了解書中文字與主要演算方程式之結果, 與方程式背後在資料分析實務上之意義, 對每一種分析方法的使用時機與其優缺點, 這樣則會有助於判讀醫學文獻與選擇適當的分析方式分析手中的研究資料.

本書共 14 章, 全書分成兩大部分, 第 1 章至第 9 章為單變量存活時間分析, 第 10 章至第 14 章為多變量存活時間分析. 為了每一章節之完整性, 因此少數之部份內容會有重覆之討論. 在第一部分單變量存活時間分析中, 第 1 章為存活分析概論, 簡介書中常使用的存活資料. 第 2 章說明存活函數危險函數, 設限與截略定義, 存活分析常見的基本假設, 以及建構概似函數. 第 3 章以參數模型進行存活分析, 包含加速失敗時間模型. 第 4 章以無母數方法估計存活函數, 討論生命量表, Kaplan-Meier 估計式, Nelson-Aalen 估計式. 第 5 章以無母數方法比較兩個或多個樣本之存活函數, 包含 log-rank 檢定, Wilcoxon 檢定, 最後討論分層檢定. 第 6 章討論 Cox 比例危險模型, 這是存活分析中最重要的一部份, 解釋時間固定共變數之參數估計, 基線危險函數與存活函數估計, 模型選擇, 並說明 Cox 比例危險模型在醫學上的應用. 第 7 章討論常見的 Cox 延伸模型, 如分層 Cox 模型, 時間相依共變數與時間相依 Cox 模型, 及時間相依 Cox 模型在移植

醫學上的應用。第 8 章則討論 Cox 迴歸模型之診斷, 如何利用各種殘差, 檢定比例危險之假設, 或尋找最佳函數型式。第 9 章介紹截略資料與區間設限資料的存活分析。在第二部分多變量存活時間分析中, 第 10 章為多變量存活分析簡介, 說明多變量存活時間與單變量存活時間差異。第 11 章討論群聚存活資料之邊際模型與分層模型。第 12 章則討論群聚存活資料之脆弱模型與隨機效應的觀念。第 13 章討論復發事件資料常見之 3 種廣義邊際模型。第 14 章則討論復發事件資料之脆弱模型。

本書以應用在醫學資料分析為主, 對於單變量或多變量存活分析方法中, 若現今常見之統計軟體可以執行之分析方法, 則儘可能提出討論。本書中所有的例題分析與圖表建構, 皆是以使用 \mathcal{R} 或是 SAS 統計軟體執行運算, 使用這 2 種統計軟體主要是因為作者的偏好與熟悉度, 本書中多數的例題分析與圖表建構, 也可以使用其它統計軟體, 如 STATA, S-Plus, SPSS 等執行運算。本書如同其它書籍, 不可能討論所有存活分析的議題, 在本書中所割捨的議題, 包含相加性危險模型, 多重階段模型, 競爭風險模型, 以群體癌症登記為基礎之相對存活分析, 貝氏存活分析, 以及計數過程之理論等。希望能在下次改版中加入或另外撰寫專書討論。最後, 因本人才疏學淺, 以致本書有所錯誤或疏漏, 尚請諸位先進與賢達不吝指教。

本書是作者在 2001-2007 年之間寫作, 以電子書型式, 供學生使用。本書的起源約在 2001 年秋季, 作者在台北大學統計系擔任兼任教師, 講授存活分析課程, 當時台灣並未大量進口適合作為碩士研究生之教科書, 因此自行撰寫教材, 講義原稿以書本型式寫作, 內容包括本書第 1 章至第 9 章中的存活理論部分, 隔年並將大部份內容改寫與修訂。在 2003-2005 年, 因改授其它課程而中斷寫作。在 2006 年再次講授存活分析課程, 因此也再度改寫大部分內容, 在 2006-2007 年, 對第 1 章至第 9 章大多數之分析方法, 加入實例解說。另外再加寫本書第 10 章至第 14 章理論與實例, 最後在 2008 年初重新排版, 整理成書。

作者長期從事醫學統計諮詢工作與期刊論文的統計審核, 存活分析常見之主要錯誤分別為: 研究者不知其研究資料為存活事件時間資料, 研究者不知其研究資料中的共變

數為時間相依共變數, 研究者不知其研究資料中的存活事件時間為多變量事件時間. 作者嘗試以個人臨床醫學經驗結合醫學統計諮詢經驗, 以雙方都能了解的語言與文字, 將存活分析之統計方法整理成書. 希望藉由本書之寫作, 在統計人員與臨床醫師之間, 提供一個溝通的橋樑, 並共同分享分析存活資料的經驗與樂趣.

人生充滿著意外與驚喜, 在偶然的機緣下, 進入到醫學與統計的世界, 先後受到醫學與統計兩個學門的師長, 悉心指導, 感激在心. 本書的完成, 要感謝美國密西根大學生物統計系給我機會, 進入到統計的世界. 首先感謝引導我進入存活分析的領域, 也是我的指導教授 Robert A. Wolfe 博士; 感謝在求學期間, 提供我助學金的教授 Richard G. Cornell 博士, Judith Bromberg 博士, M. Anthony Shorck 博士; 以及讓我有機會進入生物統計系就學的教授 Michael Boehnke 博士. 在 1995-2004 年專職從事臨床骨科工作時期, 感謝台北榮民總醫院骨科部與高雄長庚醫院骨科許多老師們的鼓勵, 以及許多同事的協助, 繼續參與醫學統計的教學與諮詢. 感謝國立台北大學統計系的前輩與同事, 讓我有機會參與統計系所的教學與研究工作, 並參與建立台北大學統計系所的醫學統計學程. 本書經過多次改寫與修訂, 感謝歷年來協助修訂的研究生, 包括亭嫻, 月麗, 志學, 惠瑜, 湘婷, 億潔, 譚景, 如茵, 百筠等. 本書的排版是利用吳聰敏教授, 吳聰慧先生, 與翁鴻翎先生開發出來的 cwTeX 系統排版軟體所完成, 謹此致謝. 本書能夠付梓, 感謝雙葉書局編輯部門的協助. 特別感謝台北榮民總醫院教學研究部主任何橈通教授, 高雄長庚醫院院長陳肇隆教授, 台北榮民總醫院骨科部主任陳天雄教授, 與台北中山醫院骨科吳濬哲主任, 給予作者長期的支持與鼓勵. 最後, 感謝在撰寫本書這段期間, 鳳凰, 湘宜, 靖庭的犧牲與支持, 讓我有機會完成本書.

林建甫 謹誌于
台北大學統計學系
2008 年 3 月

目錄

第 1 章 存活分析概論	1
1.1 存活研究與存活資料	2
1.1.1 粗死亡率及死亡發生比率	2
1.1.2 完整的觀察值	8
1.1.3 設限資料	10
1.1.4 醫學研究中典型的存活資料	11
1.2 存活資料的真實例子	13
1.2.1 Gehan 急性白血病臨床試驗	14
1.2.2 白血病患者的白血球數與死亡時間	15
1.2.3 美國退伍軍人醫院肺癌臨床試驗	16
1.2.4 非小細胞肺癌手術後存活研究	17
1.2.5 心臟移植存活時間	19
1.3 存活分析的特徵	20
1.3.1 收集存活資料	21
1.3.2 存活迴歸模型 / 邏輯斯迴歸 / 布瓦松迴歸模型的區別	23

1.4	資料分析練習	24
第 2 章	存活函數, 設限與截略	27
2.1	基本存活函數及危險函數	27
2.1.1	存活函數	28
2.1.2	危險函數	30
2.1.3	間斷型隨機變數及離散型存活函數	32
2.1.4	結合連續型及間斷型危險函數	34
2.1.5	平均餘命	35
2.2	設限和截略簡介	35
2.3	右設限觀察資料	38
2.3.1	型 I 設限 Type I Censoring	39
2.3.2	型 II 設限 Type II Censoring	40
2.4	隨機設限觀察資料	41
2.5	左設限, 雙重設限, 區間設限	42
2.6	截略觀察資料	44
2.7	存活資料的概似函數	46
2.7.1	設限與截略資料的概似函數	46
2.7.2	存活分析常用之重要統計假設	48
2.7.3	隨機設限概似函數	49

第 3 章 參數模型存活分析	51
3.1 存活分析常見的參數分配	52
3.1.1 指數分配	52
3.1.2 韋伯分配	55
3.1.3 廣義 Gamma 分配	57
3.1.4 對數常態分配	58
3.1.5 對數邏輯斯分配	59
3.1.6 極值分配 – Gumbel 分配	60
3.1.7 標準極值 (最小值) 分配	60
3.2 估計	61
3.2.1 概似函數	61
3.2.2 最大概似估計式	63
3.2.3 Newton-Raphson 演算法	64
3.2.4 參數假設檢定	66
3.2.5 參數信賴區間	69
3.2.6 Delta 法估計變異數近似函數	71
3.3 配適簡單參數模型	72
3.3.1 配適指數分配	73
3.3.2 配適韋伯分配	78
3.3.3 參數族之選擇	81
3.4 存活分析之迴歸模型	82
3.4.1 比率危險模型	83
3.4.2 可加性危險模型	84

3.4.3	加速失敗時間模型	85
3.4.4	ATF 模型與危險函數	86
3.4.5	加速失敗時間模型: 韋伯分配	88
3.4.6	韋伯分配與比例危險模型	90
3.5	資料分析練習	91
第 4 章	無母數方法估計存活函數	95
4.1	生命量表	96
4.1.1	生命量表法或精算法	99
4.1.2	Delta 法估計變異數近似函數	101
4.1.3	生命量表中存活函數的近似變異數	102
4.1.4	生命量表的種類	104
4.2	Kaplan-Meier (Product-Limit) 估計式	109
4.2.1	Kaplan-Meier 存活函數	109
4.2.2	Kaplan-Meier 存活函數之變異數估計	114
4.2.3	危險函數估計式	118
4.2.4	存活函數的逐點信賴區間	121
4.2.5	存活函數的信賴區間帶	124
4.3	存活時間平均值, 中位數及百分位數	127
4.3.1	存活時間平均數估計式	127
4.3.2	存活時間中位數及百分位數估計式	128
4.4	資料分析練習	129

第 5 章 無母數方法比較存活函數	133
5.1 兩個或多個樣本檢定	134
5.1.1 兩組樣本的假設檢定	135
5.1.2 無母數方法比較兩組樣本之存活分配	136
5.1.3 Log-Rank 檢定	141
5.1.4 (Gehan) Wilcoxon 檢定	142
5.1.5 Tarone-Ware 檢定	142
5.1.6 Peto-Wilcoxon 檢定	143
5.1.7 檢定之比較與選擇	145
5.2 比較多組樣本之存活函數	146
5.3 分層分析	152
5.3.1 比較配對研究設計之存活時間	156
5.4 資料分析練習	157
第 6 章 Cox 比例危險模型	159
6.1 Cox 比例危險迴歸簡介	160
6.1.1 Cox PH 簡單迴歸分析	160
6.1.2 Cox 比例危險複迴歸模型	163
6.2 共變數與存活時間關聯性之測量	165
6.2.1 共變數之相對風險	165
6.2.2 共變數影響之效力調整與交互作用	167
6.2.3 多項式迴歸	168

6.3	概似函數	169
6.3.1	概似函數簡介	169
6.3.2	完全概似函數	169
6.3.3	Cox 的概似函數: 在沒有相同事件時間的偏概似函數	170
6.4	參數估計	172
6.4.1	分數函數與資訊函數	172
6.4.2	以牛頓法計算最大偏概似函數估計式	175
6.5	參數推論	176
6.5.1	參數假設檢定	176
6.5.2	參數信賴區間	178
6.6	Cox PH 模型實例	181
6.7	Cox 概似函數: 多個相同事件時間	187
6.8	建立模型	195
6.8.1	調整潛在的干擾變數	195
6.8.2	Akaike 資訊準則和 Schwartz 的貝氏資訊準則	196
6.8.3	逐步選取法	197
6.9	基線存活曲線之估計	201
6.9.1	基線存活曲線的概似函數	201
6.9.2	基線危險函數之估計	206
6.9.3	估計基線存活函數: 無相同事件時間	207
6.9.4	估計基線存活函數: 相同事件時間	208
6.9.5	統計軟體中的存活函數估計	212

6.10 Cox 概似函數的討論	218
6.10.1 Cox 的偏概似函數如同 Profile 概似函數	218
6.10.2 Cox 對排序時間的偏概似函數與邊際概似函數	220
6.10.3 從全概似函數看偏概似函數	222
6.11 資料分析練習	224
第 7 章 Cox 延伸模型	231
7.1 分層分析	231
7.1.1 分層 Cox 比例危險模型	231
7.1.2 檢定分層分析中平行線假設	233
7.2 時間相依共變數 Cox 模型	238
7.2.1 內在與外在時間相依共變數	238
7.2.2 研究追蹤中治療更動或改變: 時間相依共變數	239
7.2.3 治療改變後才測量之時間相依共變數	240
7.2.4 在比例危險模型中定義時間相依共變數	242
7.2.5 時間相依共變數的 Cox 非比例危險模型	245
7.2.6 時間相依共變數的偏概似函數	246
7.2.7 時間相依共變數的概似函數與估計	248
7.2.8 計數過程與時間相依共變數	249
7.2.9 史丹佛心臟移植研究資料	252
7.3 資料分析練習	263

第 8 章 Cox 迴歸模型之診斷	269
8.1 殘差	270
8.1.1 Cox-Snell 殘差	270
8.1.2 平賭殘差	271
8.1.3 Deviance 殘差	276
8.1.4 分數殘差	276
8.1.5 Schoenfeld (分數) 殘差	277
8.2 評估 Cox 模型的適合度	279
8.3 決定一個共變數的最佳函數形式: 平賭殘差	280
8.4 檢查成比例模型假設	281
8.5 檢查 Cox 模型離群值: Deviance 殘差	285
8.6 個體個別共變數之影響值	285
8.7 Cox 模型診斷實例	287
8.8 資料分析練習	298
第 9 章 截略與區間設限資料存活分析	299
9.1 左截略資料的存活函數	300
9.1.1 左截略	300
9.1.2 左截略及延遲進入的資料	306
9.1.3 左截略及與時間相依的分層	308
9.2 右截略資料的存活函數	309

9.3	左設限與區間設限資料的存活函數	314
9.3.1	左設限資料的存活函數	314
9.3.2	區間設限資料的存活函數	316
第 10 章	多變量存活分析簡介	321
10.1	多重事件與多變量存活時間實例	322
10.1.1	糖尿病視網膜病變臨床試驗 (DRS)	322
10.1.2	膀胱癌臨床試驗 (Bladder)	324
10.2	群聚存活時間	325
10.3	多重復發事件時間	326
10.4	穩健的變異數估計與計算	327
10.5	Cox 比例危險模型與多變量存活分析	329
10.6	資料建構與模型選擇	331
10.7	無序多重存活時間	332
10.8	有序多重存活時間分析	333
10.9	資料分析練習	335
第 11 章	群聚存活資料: 邊際模型	339
11.1	邊際模型簡介	339
11.2	邊際模型之統計操作	341
11.3	邊際模型的限制	343

11.4 資料分析練習	345
第 12 章 群聚存活資料: 脆弱模型	347
12.1 隨機效應與脆弱	348
12.2 群聚資料: 分層 Cox 模型	349
12.3 隨機效應與脆弱模型	353
12.3.1 脆弱模型之條件危險函數	353
12.3.2 群聚事件時間之聯合機率模型	354
12.3.3 脆弱模型之邊際與條件危險函數	355
12.4 相關性之測量	359
12.5 群聚資料隨機效應脆弱模型之檢定	360
12.6 常見之隨機效應脆弱模型	362
12.6.1 Gamma 脆弱模型	363
12.6.2 Positive Stable 脆弱模型	364
12.6.3 對數常態脆弱模型	365
12.7 脆弱模型之估計與推論	366
12.7.1 估計 Gamma 脆弱模型之 EM 演算法	366
12.7.2 PPL 演算法	369
12.7.3 Positive Stable 脆弱模型之估計與推論	371
12.8 DRS 資料脆弱模型實例	372
12.9 脆弱模型之延伸	374
12.9.1 邊際比例危險之脆弱模型	374

12.9.2	指定錯誤的脆弱模型	375
12.9.3	檢測脆弱模型之假設	375
12.9.4	基因與脆弱模型	376
12.10	群聚脆弱模型結論	376
12.11	資料分析練習	378
第 13 章	復發事件資料: 邊際模型	379
13.1	復發事件與布瓦松過程	380
13.1.1	時間軸	381
13.1.2	布瓦松過程與計數過程	382
13.1.3	布瓦松迴歸模型	383
13.1.4	邊際事件發生率模型	384
13.2	復發事件資料: 常見邊際模型	385
13.2.1	Andersen 與 Gill 邊際模型	386
13.2.2	Wei, Lin, 與 Wiessfeld (WLW) 邊際模型	389
13.2.3	Prentice, Williams 與 Petersen (PWP) 條件模型	391
13.3	模型配適與資料型式	392
13.4	膀胱癌臨床試驗: 廣義邊際模型	395
13.5	復發事件資料邊際模型摘要	400
13.6	資料分析練習	402

第 14 章 復發事件資料: 脆弱模型	405
14.1 隨機效應與脆弱模型	406
14.1.1 復發事件資料: 過度分散布瓦松迴歸	407
14.1.2 概似函數	408
14.1.3 Gamma 脆弱模型	408
14.1.4 Positive Stable 脆弱模型	409
14.1.5 估計與推論	409
14.2 膀胱癌臨床試驗: 脆弱模型 (I)	410
14.3 膀胱癌臨床試驗: 脆弱模型 (I)	413
14.4 脆弱模型的選擇	418
14.5 資料分析練習	419
參考文獻	421
索引	432

表目錄

1.1	一個社區中肺癌的盛行率及累計發生率	4
1.2	樣本數為 20 的兒童血癌存活時間完整追蹤研究 (單位: 月)	9
1.3	設限資料: 觀察 50 位個體之存活時間 (單位: 月)	10
1.4	Gehan 急性白血病臨床試驗: 比較 6-MP 組與對照組的緩解時間 . .	14
1.5	白血病患者的白血球數與死亡時間	15
1.6	美國退伍軍人醫院肺癌臨床試驗	16
1.7	VGH 第一期非小細胞肺癌手術後存活研究: 7 位個體部分研究結果 .	17
1.8	VGH 第一期非小細胞肺癌手術後存活研究: 變數說明	18
1.9	心臟移植存活時間: 5 位病患部分研究結果	19
1.10	心臟移植存活時間	19
1.11	資料收集中重要的時間及日期	22
1.12	存活分析中常見的統計差異	22
1.13	胃賁門癌存活分析: 變數說明	25

2.1	人類存活率 (1990)	29
2.2	設限與截略之概似函數	47
3.1	DMBA 致癌物之老鼠存活試驗	73
3.2	DMBA 致癌物之老鼠存活試驗: 配適指數分配結果	77
3.3	DMBA 致癌物之老鼠存活試驗: 配適韋伯分配結果	81
3.4	美國退伍軍人醫院肺癌臨床試驗: 韋伯分配加速失敗時間模型	89
3.5	胸膜惡性間皮瘤第二期臨床試驗: 變數說明	92
3.6	胸膜惡性間皮瘤第二期臨床試驗: 部分個體資料	93
4.1	存活分析: 生命量表	98
4.2	VGH 第一期非小細胞肺癌手術後存活研究的生命量表 (I)	105
4.3	VGH 第一期非小細胞肺癌手術後存活研究的生命量表 (II)	106
4.4	VGH 第一期非小細胞肺癌手術後存活研究: 在生命量表中的區間中點 之危險率	107
4.5	存活分析: Kaplan-Meier 法	113
4.6	急性白血病臨床試驗中 21 位孩童接受 6-MP 治療: Kaplan-Meier 存 活率函數與標準誤差估計直	117
4.7	復發惡性腦星狀細胞瘤研究 (20 位成人以週為單位的死亡或設限時間)	123
4.8	復發惡性腦星狀細胞瘤: Kaplan-Meier 存活率和累積危險率	123
4.9	威爾母士氏瘤資料: 變數說明	129
4.10	多發性骨髓瘤存活分析: 變數說明	131

5.1	一系列 2×2 列聯表分析: 階層 1	138
5.2	一系列 2×2 列聯表分析: 階層 i	138
5.3	一系列 2×2 列聯表分析: 階層 k	138
5.4	Mantel-Haenszel 統計量應用在一序列之 2×2 列聯表: 在時間點 t_i (階層 i) 兩組樣本存活之結果	139
5.5	Gehan 急性白血病臨床試驗: 比較 6-MP 與安慰劑治療後存活時間 .	143
5.6	一序列 $G \times 2$ 的列聯表: 群組 G 在時間點 t_i 時情況	147
5.7	F98-神經膠質瘤老鼠動物實驗: 三組治療之存活時間	150
5.8	美國退伍軍人醫院肺癌臨床試驗: 分層 log-rank 檢定	156
5.9	慢性肉芽腫病第 1 次嚴重感染時間分析 I: 變數說明	158
6.1	白血病患者的白血球數與死亡時間資料: 單一變數的 Cox 模型 . .	184
6.2	白血病患者的白血球數與死亡時間資料: 2 個變數的 Cox 模型 . .	185
6.3	白血病患者的白血球數與死亡時間資料: 有交互作用項的 Cox 模型 .	186
6.4	Gehan 急性白血病臨床試驗研究: 比較 Cox 模型在處理相同事件時 間之不同計算方法	194
6.5	美國退伍軍人醫院肺癌臨床試驗: 以逐步選取法建立模型之結果 . . .	200
6.6	Gehan 急性白血病臨床試驗研究: 型 (1) 存活函數估計	214
6.7	Gehan 的臨床試驗: 型 (2) 與型 (3) 存活函數估計	215
6.8	惡性黑色素瘤存活分析: 變數說明	225
6.9	惡性腦神經膠質瘤臨床試驗: 變數說明	227

6.10 骨惡性肉瘤的存活分析: 變數說明	229
7.1 白血病患者的白血球數與死亡時間: 標準的 Cox PH 模型與分層 Cox PH 模型分析的參數估計	236
7.2 白血病患者的白血球數與死亡時間: 分層分析與非分層分析的估計存活函數	237
7.3 AIDS 與時間相依共變數的結果	241
7.4 血壓與時間相依共變數的分析結果	241
7.5 計數過程與時間相依共變數資料輸入	250
7.6 計數過程與時間相依共變數的資料編碼	251
7.7 計數過程與多個時間點的重覆測量值之資料編碼	252
7.8 史丹佛心臟移植研究資料: 變數定義	253
7.9 史丹佛心臟移植資料: 10 位病患詳細資料	254
7.10 史丹佛心臟移植資料; 時間相依 Cox 模型結果	260
7.11 UIS 藥物濫用和預防 AIDS 臨床試驗: 變數說明	265
7.12 末期腎臟病腎臟移植存活研究: 變數說明	267
7.13 末期腎臟病腎臟移植存活研究: 10 位個體部分資料	267
8.1 原發性膽汁肝硬化	288
9.1 Channing 安養中心中老年居民死亡時間: 變數說明	301
9.2 Channing 安養中心中老年居民死亡時間: 部分個體資料	301

9.3	AIDS 潛伏期時間	310
9.4	AIDS 潛伏期時間: 部分個體資料	310
9.5	所有個體以右截略資料為基礎的 AIDS 潛伏期時間分配估計值 . . .	313
9.6	乳癌病患乳房外貌惡化變縮的時間	317
9.7	以所有乳癌病患乳房外貌美觀惡化變縮之區間設限的資料為根據, 估計其存活機率	319
10.1	糖尿病視網膜病變臨床試驗 (DRS): 部份資料	323
10.2	糖尿病視網膜病變臨床試驗 (DRS): 變數說明	323
10.3	膀胱癌臨床試驗 (Bladder): 部份資料	324
10.4	膀胱癌臨床試驗 (Bladder): 變數說明	325
10.5	群聚共變數同義詞	326
10.6	Gehan 急性白血病臨床試驗: 配適一般 Cox 模型, 比較獨立的參數變異數估計, 與考慮配對樣本之三明治估計式	329
10.7	多變量存活分析可能之組合	330
10.8	有序, 無序資料與分層邊際模型組合	334
10.9	大腸癌臨床試驗存活分析: 變數說明 I	336
10.10	大腸癌臨床試驗存活分析: 變數說明 II	337
11.1	DRS 資料之群聚邊際模型: 比較一般 Cox 模型與加入變異數三明治估計式	343
11.2	血液透析導管感染時間分析: 變數說明	346

12.1 DRS 資料: 比較分層 Cox 模型與其他邊際模型	351
12.2 DRS 資料: 比較脆弱模型	373
12.3 配對小老鼠存活實驗: 變數說明	378
13.1 AG, PWP 邊際模型與計數過程資料輸入	388
13.2 有序, 無序資料與分層邊際模型組合	393
13.3 三種常用的有序多重事件存活分析之資料建構	394
13.4 Bladder 原始資料: AG 與 PWP 資料輸入配適一般模型	396
13.5 Bladder 資料: 比較 AG, WLW, PWP 廣義邊際模型	397
13.6 Bladder 資料: 比較 WLW, PWP 模型之治療與分層交互作用分析 .	398
13.7 Bladder 資料: 比較考慮第一次復發與多重復發事件	399
13.8 囊狀纖維化臨床試驗: 肺功能惡化次數表	403
13.9 囊狀纖維化臨床試驗: 變數說明	404
14.1 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 邊際模型與脆弱模型 (I) .	411
14.2 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 邊際模型與脆弱模型 (II)	411
14.3 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 邊際模型與脆弱模型 (III)	412
14.4 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 邊際模型與脆弱模型 (IV)	412
14.5 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 脆弱模型	415
14.6 膀胱癌臨床試驗 (Bladder) 資料: 比較 WLW 脆弱模型	416
14.7 膀胱癌臨床試驗 (Bladder) 資料: 比較 PWP 脆弱模型	417

14.8 慢性肉芽腫病嚴重感染時間分析 II: 變數說明 420

圖目錄

1.1	發生率與盛行率 (Incidence and Prevalence): ● initial follow-up, × death, ○ alive.	5
1.2	發生率與人-年 (Incidence and Person-Year): ● initial follow-up, × death, ○ alive.	7
1.3	樣本數為 20 的兒童血癌存活時間完整追蹤研究: (a) 莖葉圖; (b) 機率 密度估計圖.	9
1.4	日曆時間 (Calender Time): 從 1980 到 1990 追蹤 7 位有肺癌的個 體情況. (● initial follow-up, × death, ○ alive.)	12
1.5	病患時間: 從 1980 到 1990 追蹤 7 位有肺癌的個體情況. (● initial follow-up, × death, ○ alive.)	13
2.1	不同型式的危險函數	32
2.2	常見的設限時間形式: ● initial follow-up, × death, ○ alive. . .	38
2.3	同時存在左截略與右截略觀察資料	45
3.1	不同危險率的指數型存活函數	53

3.2	不同形狀的韋伯存活函數及危險函數	56
3.3	DMBA 資料的非參數之 Kaplan-Meier 存活曲線與指數存活曲線	77
4.1	時間軸與分割之時間區間	96
4.2	VGH 第一期非小細胞肺癌手術後存活研究的生命量表: (a) 存活函數 $S(t)$, (b) $-\log(\text{存活函數})$, (c) $\log(-\log(\text{存活函數}))$, (d) 機率密度函數, 以及 (e) 危險函數	108
4.3	Kaplan-Meier 存活曲線	114
4.4	急性白血病臨床試驗中 21 位孩童接受 6-MP 治療: Kaplan-Meier 存活曲線; (左) 存活曲線, (右) 包含對稱信賴區間的存活曲線.	118
4.5	復發惡性腦星狀細胞瘤的 Kaplan-Meier 存活曲線	124
4.6	VGH 第一期非小細胞肺癌手術後存活研究: Kaplan-Meier 存活曲線及各種信賴區間帶	126
5.1	Gehan 的急性白血病研究: 比較兩個樣本之存活曲線	134
5.2	Gehan 急性白血病研究的存活曲線及信賴帶	144
5.3	F98-神經膠質瘤老鼠動物實驗: 比較三組治療之的存活曲線及累積危險函數曲線	151
5.4	美國退伍軍人醫院肺癌臨床試驗之存活曲線: 左圖為比較兩種治療存活率; 右圖為比較四種細胞存活率	154
5.5	美國退伍軍人醫院肺癌臨床試驗: 以 4 細胞型態作為分層下的 2 組治療方式之存活曲線	155

6.1	白血病患者的白血球數與死亡時間資料的存活函數: (左) 以 Kaplan-Meier 方法為基礎; 與 (右) $\log(-\log(\text{survival}))$	183
6.2	Gehan 急性白血病臨床試驗研究: 比較 Kaplan-Meier 與 Cox PH 模型下的形式 (3) 估計存活曲線	216
6.3	美國退伍軍人醫院肺癌臨床試驗: 基於 Cox PH 模型, Karnofsky 分數 (kps) 在 60 且在 4 種細胞型態之下的母體估計存活曲線.	217
7.1	白血病患者的白血球數與死亡時間: 比較分層分析與非分層分析的之估計存活曲線	236
7.2	移植效應: 移植視為時間相依共變數的危險函數 (Cox, 1984)	244
7.3	史丹佛心臟移植資料, 使用時間相依共變數 Cox 模型之存活曲線估計: (左): 兩個模擬世代; (右): 時間相依共變數個體.	262
8.1	Cox-Snell 殘差對殘差的估計累積危險率圖: (左) 整體; 和 (右) edema 的三個群組.	289
8.2	Cox-Snell 殘差對來自 edema 分層 Cox 模型殘差的估計累積危險率圖: (左) 整體; 和 (右) edema 的三個群組.	290
8.3	虛無模型的平賭殘差對共變數圖形.	291
8.4	加入共變數的平賭殘差對共變數圖形.	293
8.5	第 k 層 Kaplan-Meier 存活曲線的分層估計式 $\hat{S}_k(t), \log[-\log(\hat{S}_k(t))] = \log[\hat{H}(t Z = k)]$ 對 t 的圖形.	294
8.6	scaled Schoenfeld 殘差對時間 futime 繪製平滑曲線圖形.	295
8.7	離群值的 Deviance 殘差分析.	296

8.8	影響點的 Dfbeta 分析.	297
9.1	(左) 危險集合數量為年齡函數; (右) 給定進入年齡 > 68 , 估計之條件存活機率.	304
9.2	忽略截略時間 (從出生後或從進入安養中心後計算存活時間) 及考慮截略時間 (全部, 男性及女性) 的存活機率, 時間軸從 0 開始.	305
9.3	忽略截略時間 (從出生後計算存活時間) 及考慮截略時間 (全部, 男性及女性) 的存活機率, 時間軸從 68 歲開始.	306
9.4	給定感染時間小於 8^+ 年, AIDS 潛伏期時間時間之條件機率, $\hat{F}_X(x) = P[X \leq x, (x = \tau^+ - r) X \leq \tau = 8^+]$, 的估計曲線	312
9.5	以所有乳癌病患乳房外貌美觀惡化之區間設限的資料為根據, 估計其存活曲線	318
12.1	比較脆弱模型之條件危險函數與邊際危險函數	357
12.2	在脆弱模型之下, 比較條件危險函數與邊際危險函數的風險比. 隨著時間增加, 邊際危險函數之危險比值趨近於 1, $\lim_{t \rightarrow \infty} HR^M(t) \rightarrow 1$	358
13.1	三種常見之有序多重存活分析的邊際模型 (Therneau 與 Grambsch, 2000)	386
