

# 迴歸分析

林建甫

Jeff Chien-Fu Lin, MD. PhD.

2015

迴歸分析概論 / 林建甫著. — 初版. — 台北市:

雙葉書廊, 2014. 07

面: 公分

參考書目: 面

含索引

ISBN XXX-XXX-XXXX-XX-X(平裝)

1. 生物統計 2. 醫學 3. 統計

306.13

97005978

©XXXX 林建甫

書名: 迴歸分析

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 XXXX 年 X 月初版一刷

ISBN: XXX-XXX-XXXX-XX-X(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  
Jeff Chien-Fu Lin, MD. PhD.  
cflin@mail.ntpu.edu.tw  
<http://web.ntpu.edu.tw/~cflin>

經歷 台北醫學大學萬芳醫院骨科主治醫師  
高雄長庚醫院骨科主治醫師  
台北榮民總醫院骨科部醫師  
智策市場研究公司統計顧問  
台北榮民總醫院生物統計顧問

學歷 美國密西根大學生物統計研究所碩士, 博士  
高雄醫學院醫學系醫學士



# 序

統計模型是使用數學方程式說明與量化變數之間的關係，迴歸模型是一種統計模型，通常用來討論一個研究資料內反應變數與解釋變數間的因果關係或關聯性。在醫學研究中，連續型反應變數為一種常見的研究資料，例如，血壓，血糖，心肺功能，肝功能，腎功能指數等等，若血壓為反應變數，醫學研究通常探討與量化影響血壓的可能因子（變數），包含年紀，性別，生活型態，共病症等等，但血壓與這些可能影響血壓的因子，科學研究未必能發現其真實的關係或用方程式呈現其真實的關係，通常需要使用統計模型的概念分析其潛在的可能關係。線性迴歸分析主要用於分析一個研究資料的一個連續型反應變數與多個解釋變數之間的線性關係，簡單線性迴歸討論資料內的一個連續型反應變數與一個解釋變數之間的關係，複迴歸或是多變數迴歸分析資料內的一個反應變數與多個解釋變數或預測變數之間的關係。若研究未特定指出迴歸模型的類別，迴歸分析通常是指線性迴歸分析，主要是假設反應變數與解釋變數之間的關係為線性關係。線性迴歸分析是所有迴歸模型的基礎，類別資料迴歸分析，存活資料迴歸分析，縱向資料迴歸分析等，都會利用線性迴歸分析的基本概念，雖然線性迴歸分析在醫學資料分析的使用頻率，少於類別資料，存活資料或縱向資料的迴歸分析，但是類別資料分析，存活分析與縱向資料分析等專業書籍不會重覆討論線性迴歸分析的基本概念，若未能對線性迴歸分析有清楚的了解，許多醫學資料研究常常誤用了迴歸模型，得到錯誤的研究結論。

本書主要寫作對象為從事臨床醫學研究的醫師與醫學研究人員，本書主要針對醫學資料，對經常使用的線性迴歸模型進行探討，本書共 17 章，全書分成兩部分，第一部分包含第 1-10 章，為線性迴歸分析的基本內容，第 1 章為迴歸分析概論，說明資料分析的概念，第 2 章說明簡單線性迴歸模型，第 3 章介紹複迴歸模型，第 4 章說明複迴歸的模型比較之變異數分析表，第 5 章討論變數轉換。第 6 章為多項式迴歸模型，第 7 章討論解釋變數為因子變數（類別變數）的操作，第 8 章說明加權最小平方法，第 9 章討論迴歸模型之診斷，第 10 章介紹變數選取與模型選擇。第二部分包含

第 11–17 章, 包含巨量資料分析與資料採礦經常使用的迴歸模型, 第 11 章為迴歸模型之預測理論, 第 12 章討論懲罰迴歸分析, 第 13 章介紹非參數估計密度函數, 第 14 章說明非參數迴歸模型, 第 15 章介紹樣條迴歸, 第 16 章介紹穩健迴歸, 第 17 章介紹分量迴歸. 本書以應用在醫學資料分析為主, 在線性迴歸模型的方法中, 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  $R$  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, SPSS 等執行運算. 本書如同其它書籍, 不可能討論所有線性迴歸分析的議題, 有些議題在本書中所割捨, 例如 類別資料迴歸模型 與 縱向資料迴歸模型等, 將另外撰寫專書討論. 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.

本書所使用的數學內容僅包含加法, 乘法, 與簡單之矩陣運算, 這些內容等同於在台灣的高中與大一的基礎數學程度. 閱讀本書的臨床醫學研究的醫師, 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.

本書是作者在 2009-2014 年之間寫作, 以電子書型式, 供學生使用. 本書的起源約在 2009 年秋季, 作者在台北大學統計系講授線性迴歸分析課程時撰寫的部分教材, 包括本書第 1 章至第 10 章中的線性迴歸模型部分. 在 2010-2012 年於碩士班講授線性迴歸分析課程時, 另外增加部分之教材, 包括本書第 10 章至第 17 章中的非參數迴歸模型部分. 在 2013 年暑假期間, 前往美國 Vanderbilt University 受邀當訪問學者實, 將部份內容重新修訂. 最後在 2014 年期間整理成書. 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, 將線性迴歸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, 並共同分享線性迴歸分析的經驗與樂趣.

迴歸分析原本不是作者規劃的寫作書籍之內, 在 2009 年臨時接受大學部線性迴歸分析課程, 只好撰寫中文教材, 2010 年變動更在碩士班講授線性迴歸分析, 只好再加入更多的內容, 商學院上課時多使用商業與財務相關資料, 寫作本書則儘量改用醫學相關資料. 原本以為之前寫作存活分析的經驗, 有助於寫作本書的腳步, 但是繁重的教學時數與課程不定時的變動, 仍然延滯到 2015 年寒假期間才真正完成本書. 在寫作這段時間內的臨床工作, 受到台北市萬芳醫院骨科何為斌主任以及科內所有同仁的協助, 特此致謝. 本書經過多次改寫與修訂, 感謝歷年來協助的研究生, 包括 陳瑜希, 蔡宛蓉 與 劉亭萱等. 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  $\text{cwTeX}$  系統排版軟體所完成, 謹此致謝. 特別感謝梁文敏教授, 羅琪教授, 石瑜教授, 呂才學醫師在研究與臨床工作上的支持與協助, 本書能夠付梓, 感謝雙葉書局編輯部門的協助. 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  
2015 年 2 月

---



---

# 目錄

第 1 章 迴歸分析概論 .....	1
1.1 資料與變數 .....	1
1.2 統計模型 .....	4
1.3 迴歸分析之平均值函數與變異數函數 .....	7
1.4 迴歸分析的歷史 .....	8
1.5 線性迴歸分析: 例題資料 .....	10
1.6 資料分析練習 .....	15
第 2 章 簡單線性迴歸 .....	17
2.1 簡單線性迴歸模型 .....	18
2.2 簡單線性迴歸模型之統計假設 .....	21
2.3 參數估計: 最小平方法 .....	24
2.4 誤差變異數之估計 .....	27
2.5 最小平方法估計式的性質 .....	30
2.6 最小平方法估計式的變異數 .....	31
2.7 參數推論 .....	34
2.8 模型比較與變異數分析表 .....	39
2.9 決定係數 .....	45

2.10	配適值與預測值	47
2.11	殘差	53
2.12	線性迴歸與線性相關係數	56
2.13	反向預測	57
2.14	資料分析練習	61
第 3 章	複迴歸	67
3.1	複迴歸模型與因子控制	67
3.2	散佈圖與複迴歸	69
3.3	複迴歸模型	73
3.4	複迴歸模型與線性代數	74
3.5	參數估計	76
3.6	參數推論	84
3.7	信賴區域	86
3.8	複迴歸模型配適值與預測值	90
3.9	複迴歸模型中缺少重要自變數	97
3.10	複迴歸模型與研究設計	98
3.11	資料分析練習	100
第 4 章	變異數分析與模型比較	105
4.1	複迴歸模型與變異數分析表	106
4.2	複決定係數	111
4.3	檢定單一自變數與變異數分析表	113
4.4	貫續檢定自變數與變異數分析表	118
4.5	巢狀模型與變異數分析表	126
4.6	部分決定係數	132

---

---

4.7	資料分析練習	141
第 5 章	變數轉換	147
5.1	變數轉換	148
5.2	自變數的轉換與相關性	155
5.3	正交變數	163
5.4	自變數線性組合	165
5.5	標準化迴歸模型	166
5.6	資料分析練習	168
第 6 章	多項式迴歸與階層模型	171
6.1	多項式迴歸	172
6.2	交互作用	179
6.3	階層模型	183
6.4	資料分析練習	184
第 7 章	因子變數與共變異數分析	189
7.1	因子變數	189
7.2	二元因子變數	190
7.2.1	參照水準編碼	191
7.2.2	空格平均值編碼	192
7.2.3	因子效應編碼	193
7.3	多項式類別變數	197
7.3.1	多項式類別變數與參照水準編碼	198
7.3.2	多項式類別變數與空格平均值編碼	201
7.3.3	多項式類別變數與因子效應編碼	203
7.4	共變異數分析	209
7.5	共變異數分析模型比較	214

---

7.6	資料分析練習	229
第 8 章	加權最小平方法	235
8.1	變異數同質性假設	235
8.2	一般最小平方法	237
8.3	加權最小平方法	238
8.4	配適度檢定	246
8.5	資料分析練習	256
第 9 章	迴歸模型診斷	257
9.1	殘差與槓桿	258
9.2	殘差圖	265
9.3	離群值檢定	268
9.4	影響值	271
9.5	共線性	275
9.6	共線性與變異數膨脹因子	276
9.7	條件指數	278
9.8	誤差項常態分配假設檢定	280
9.9	檢視誤差項的相關性	286
9.10	檢定同質性變異數	286
9.11	診斷統計量: 個案資料分析	293
9.12	診斷非線性	300
9.13	反應變數的線性關係轉換: Box-Cox 轉換	301
9.14	解釋變數的線性關係轉換	308
9.15	資料分析練習	311

---

---

第 10 章	迴歸模型選擇 .....	313
10.1	完整模型與子模型 .....	314
10.2	模型選擇標準 .....	316
10.3	模型選擇程序 .....	325
10.4	資料分析練習 .....	335
第 11 章	迴歸分析與預測 .....	339
11.1	預測誤差 .....	339
11.2	偏誤與變異數分解 .....	343
11.3	模型選擇 .....	348
11.4	交叉驗證法 .....	353
11.5	模型選擇與假說檢定 .....	358
第 12 章	懲罰迴歸分析 .....	361
12.1	懲罰迴歸概論 .....	362
12.2	脊迴歸 .....	364
12.3	最小絕對壓縮與選擇 .....	370
第 13 章	非參數估計密度函數 .....	379
13.1	直方圖 .....	379
13.2	核密度估計 .....	384
13.3	信賴區間帶 .....	392
13.4	多變量機率密度估計 .....	393
第 14 章	非參數迴歸模型 .....	397
14.1	非參數迴歸 .....	397

---

14.2	局部平均估計式	398
14.3	核迴歸	402
14.4	局部加權迴歸	405
第 15 章	樣條迴歸	413
15.1	格子平滑迴歸	413
15.2	基底函數	417
15.3	分段迴歸	417
15.4	樣條迴歸	420
15.5	非參數迴歸模型變異數估計	428
15.6	信賴區間帶	430
15.7	非參數樣條複迴歸模型	431
第 16 章	穩健迴歸	441
16.1	穩健	441
16.2	穩健迴歸	442
第 17 章	分量迴歸	447
17.1	分量	447
17.2	分量迴歸	450
	參考文獻	454
	索引	466

---

---

# 表目錄

1.1	飲用啤酒的數量與血中酒精濃度 . . . . .	7
1.2	血壓, 年紀與壓力指數關係變數說明 . . . . .	11
1.3	血壓, 年紀與壓力指數關係資料 . . . . .	11
1.4	鱒魚的年紀與鱒魚體內多氯聯苯殘餘量: 資料 . . . . .	14
1.5	股骨頸骨折成本與效果分析: 變數名稱與定義 . . . . .	16
2.1	飲用啤酒的數量與血中酒精濃度 . . . . .	18
2.2	符號定義 . . . . .	25
2.3	簡單線性迴歸與變異數分析表 . . . . .	43
2.4	飲用啤酒的數量與血中酒精濃度變異數分析表 . . . . .	44
2.5	飲用啤酒的數量與血中酒精濃度: 配適值與殘差值 . . . . .	55
2.6	四組假設性資料 (Anscombe, 1973) . . . . .	57
2.7	血壓與身體質量指數: 資料 . . . . .	62
2.8	跑步機上心肺功能測試 與 10-公里 長跑比賽成績 (單位時間: 分) . . . . .	65
3.1	學童智商與孩童出生排序以及母親年紀關係 . . . . .	69
3.2	學童智商與孩童出生排序以及母親年紀: 迴歸分析 . . . . .	71
3.3	學童智商與孩童出生排序以及母親年紀: 參數推論 . . . . .	89

3.4	學童智商與孩童出生排序以及母親年紀: 配適值與預測值 . . . . .	95
3.5	血壓, 年紀 與 壓力指數: 複迴歸分析之參數估計結果 . . . . .	97
3.6	成年人血壓調查研究: 變數說明 . . . . .	101
3.7	起司熟成風味研究 . . . . .	103
4.1	複迴歸模型與變異數分析表 . . . . .	109
4.2	學童智商與孩童出生排序以及母親年紀: 變異數分析表 . . . . .	110
4.3	檢定 完整模型 與 刪除單一自變數 $X_j$ 模型 的變異數分析表 (在完整模型中包含 $p$ 個解釋變數) . . . . .	115
4.4	合併檢定完整模型中的單一自變數與變異數分析表 (SAS Type III)	115
4.5	學童智商與孩童出生排序以及母親年紀: 檢定完整模型中單一自變數 $X_2$ (age) 的變異數分析表 . . . . .	117
4.6	學童智商與孩童出生排序以及母親年紀: 檢定完整模型中單一自變數 $X_1$ (birth) 的變異數分析表 . . . . .	118
4.7	學童智商與孩童出生排序以及母親年紀: SAS Type III 變異數分析表	118
4.8	貫序檢定自變數與變異數分析表 (SAS Type I) . . . . .	122
4.9	學童智商與孩童出生排序以及母親年紀: SAS Type I 變異數分析表 (1) . . . . .	124
4.10	學童智商與孩童出生排序以及母親年紀: SAS Type I 變異數分析表 (2) . . . . .	125
4.11	學童智商與孩童出生排序以及母親年紀: 比較 SAS Type I 變異數分析表 . . . . .	126
4.12	比較巢狀模型與變異數分析表 . . . . .	129
4.13	血壓, 年紀與壓力指數: 迴歸分析參數估計 . . . . .	131
4.14	血壓, 年紀與壓力指數: 巢狀模型比較之變異數分析表 . . . . .	132
4.15	$X, Y$ 和 $Z$ 之間四種可能的偏相關係數 . . . . .	134

---

---

4.16	血壓, 年紀與壓力指數: 巢狀模型比較之變異數分析表 II . . . . .	139
4.17	身體脂肪與皮褶厚度: 變數說明 . . . . .	141
4.18	身體脂肪與皮褶厚度: 部分資料 . . . . .	142
4.19	汽車: 變數說明 . . . . .	143
4.20	Alumni Donations: 變數名稱與定義 . . . . .	145
5.1	老鼠腦組織與血液中甲苯濃度 . . . . .	150
5.2	老鼠腦組織與血液中甲苯濃度之變數轉換: 6 種模型比較 . . . . .	153
5.3	糖尿病 (DM) 患者手術血糖: 部分資料 . . . . .	157
5.4	糖尿病 (DM) 患者手術血糖: 5 種模型比較 . . . . .	161
5.5	身體脂肪與身體質量指數 . . . . .	169
5.6	變數轉換: 4 組原始資料說明 . . . . .	170
6.1	年紀與類固醇: 資料 . . . . .	174
6.2	年紀與類固醇: 多項式模型之貫續檢定變異數分析表 (Type I 變異數分析表) . . . . .	176
6.3	年紀與類固醇: 4 種多項式模型比較的參數估計 . . . . .	177
6.4	血壓, 年紀與壓力指數: 交互作用 . . . . .	182
6.5	年紀與肌肉質量 . . . . .	184
6.6	番茄產量: 溫度與濕度 . . . . .	187
7.1	止吐劑臨床試驗 . . . . .	194
7.2	止吐劑臨床試驗: 二組獨立樣本 $T$ 檢定 . . . . .	195
7.3	止吐劑臨床試驗: 簡單線性迴歸模型 . . . . .	197
7.4	非類固醇消炎藥物與出血 . . . . .	206
7.5	非類固醇消炎藥物與顯微出血: ANOVA 表 . . . . .	206
7.6	非類固醇消炎藥物 與 顯微出血: 簡單線性迴歸模型 . . . . .	208

---

7.7	臨床試驗: 比較 Reg 與 Syn 對降低血糖的效應: 部分資料 . . . . .	220
7.8	臨床試驗: 比較 Reg 與 Syn 對降低血糖的效應: 變數與定義 . . . . .	221
7.9	合成胰島素與一般胰島素降低血糖的效應 . . . . .	223
7.10	合成胰島素與一般胰島素對降低血糖的效應: $T$ 檢定 . . . . .	223
7.11	合成胰島素與一般胰島素對降低血糖的效應: 8 種模型比較 . . . . .	228
7.12	基礎代謝率與身體肌肉質量 . . . . .	230
7.13	解熱複合藥: 臨床試驗 . . . . .	232
7.14	鈣與高血壓 . . . . .	233
7.15	糖尿病治療臨床試驗: 變數說明 . . . . .	234
8.1	老鼠腦組織與血液中藥物濃度: 部分資料 . . . . .	243
8.2	老鼠腦組織與血液中藥物濃度: 加權最小平方法參數估計 . . . . .	246
8.3	迴歸模型檢定缺少配適度與變異數分析表 . . . . .	251
8.4	老鼠靜脈注射藥物與血中濃度: 資料 . . . . .	252
8.5	老鼠藥物劑量與血液中藥物濃度: 參數估計 . . . . .	254
8.6	: 老鼠藥物劑量與血液中藥物濃度: 缺少配適度檢定 (Test for Lack of Fit) 與變異數分析表 . . . . .	255
9.1	藥物劑量與老鼠平衡反應測試 . . . . .	283
9.2	藥物劑量與老鼠平衡反應測試: 簡單線性迴歸模型的參數估計 . . . . .	284
9.3	藥物劑量與老鼠腦組織濃度: 簡單線性迴歸的參數估計 . . . . .	292
9.4	最大呼氣流速與年紀, 身高, 體重關係: 資料 . . . . .	295
9.5	最大呼氣流速與年紀, 身高, 體重關係: 參數估計 . . . . .	297
9.6	最大呼氣流速與年紀, 身高, 體重關係: 模型診斷統計量 . . . . .	297
9.7	動物睡眠時間: 部分資料 . . . . .	304
9.8	動物睡眠時間: 變數說明 . . . . .	305

---

---

9.9	大腦重量與體重: 參數估計 . . . . .	306
10.1	高血壓, 年紀與壓力指數: 最佳子模型選擇 . . . . .	321
10.2	高血壓, 年紀與壓力指數: 最佳子模型參數估計 . . . . .	322
10.3	高血壓, 年紀與壓力指數: 所有子模型選擇 I . . . . .	324
10.4	高血壓, 年紀與壓力指數: 所有子模型選擇 II . . . . .	325
10.5	高血壓, 年紀與壓力指數: 向前選取程序 . . . . .	333
10.6	高血壓, 年紀與壓力指數: 逐步選取程序 . . . . .	334
10.7	Cystic Fibrosis: 變數名稱與定義 . . . . .	336
10.8	末期腎病 (ESRD) 生活品質: 4 種治療方式比較 . . . . .	338
12.1	身體脂肪與皮褶厚度: 部分資料 . . . . .	366
12.2	身體脂肪與皮褶厚度: 變數說明 . . . . .	367
12.3	脊迴歸 與 一般線性迴歸 之參數估計 . . . . .	368
12.4	攝護腺癌 與 PSA: 變數說明 . . . . .	375
12.5	攝護腺癌 與 PSA: OLS, Ridge 與 Lasso 參數估計比較 . . . . .	376
15.1	一般線性迴歸 與 相加性非參數迴歸: 模型比較 . . . . .	440
16.1	一般線性迴歸 與 穩健迴歸: 參數估計 . . . . .	445
17.1	新生兒體重研究: 變數說明 . . . . .	451
17.2	分量迴歸: 參數估計 . . . . .	453

---



---

## 圖目錄

1.1	飲用啤酒的數量與血中酒精濃度: (1) 觀測資料散佈圖, (2) 假設 $Y$ 與 $X$ 的線性迴歸線, 和 探索 $Y$ 與 $X$ 的真正關係的近似平滑曲線 .	7
1.2	血壓, 年紀與壓力指數配對散佈圖 . . . . .	12
1.3	鱒魚的年紀與鱒魚體內多氯聯苯殘餘量: (1) 觀測資料散佈圖, (2) 假設 $Y$ (pbc) 與 $X$ (age) 的線性迴歸線, 和 $Y$ 與 $X$ 真正關係的近似曲線 . . . . .	13
2.1	飲用啤酒的數量與血中酒精濃度: 資料觀測散佈圖, 簡單線性迴歸線與平滑曲線 . . . . .	19
2.2	簡單線性迴歸線 . . . . .	20
2.3	變異數分析表: 模型比較 . . . . .	40
2.4	總平方合分解: $SSY = RSS + SSReg$ . . . . .	42
2.5	飲用啤酒的數量與血中酒精濃度: 簡單線性迴歸線, 信賴區間帶與預測區間帶 . . . . .	52
2.6	常見的殘差圖 . . . . .	54
2.7	飲用啤酒的數量與血中酒精濃度: 殘差圖 . . . . .	55
2.8	Anscombe (1973) 的 4 個假設性資料: 散佈圖與迴歸線 . . . . .	58
2.9	原始資料散佈圖: $Y = \beta_0 + \beta_1 X + \varepsilon$ 相對 $X = \alpha_0 + \alpha_1 Y + \varepsilon$ . . .	59

3.1	加入自變數圖 (add-variable plot) . . . . .	70
3.2	信賴區域與信賴信賴區間 . . . . .	90
4.1	血壓, 年紀與壓力指數配對散佈圖 . . . . .	130
5.1	老鼠腦組織與血液中甲苯濃度之變數轉換: 6 種模型散佈圖於迴歸線	154
5.2	糖尿病 (DM) 患者手術血糖: 變數間的散佈圖 . . . . .	162
6.1	年紀 (age), $X$ , 與類固醇 (steroid), $Y$ : 原始資料散佈圖與多項模型 比較 . . . . .	176
7.1	二元因子解釋變數與迴歸模型 . . . . .	195
7.2	非類固醇消炎藥物 與 顯微出血: (a) 盒狀圖 與 (b) 散佈圖 . . . . .	207
7.3	共變異數分析模型 . . . . .	216
7.4	交互作用: 量化與質化 . . . . .	218
7.5	(左) 2 組在治療 12 個月之後的空腹血糖 FBS2 對治療前的空腹血糖 FBS1 散佈圖; (右) 2 組在治療前與治療後空腹血糖降低值, $FBSD =$ $FBS2 - FBS1$ , 對治療前的空腹血糖 FBS1 散佈圖. . . . .	221
7.6	(左) 2 組治療 12 個月後的空腹血糖 FBS2 之盒狀圖 (box-plot), (右) 2 組治療前與治療 12 個月之後, 空腹血糖降低值 $FBSD = FBS2 -$ $FBS1$ 之盒狀圖 (box-plot). . . . .	222
7.7	合成胰島素與一般胰島素對降低血糖的效應: 共變數分析的 4 種模 型 (註: 未顯示原點 $(0,0)$ ) . . . . .	225
8.1	(左) 老鼠腦組織與血液中藥物濃度散佈圖; (右) 血液中藥物濃度樣 平均值的平方與腦組織濃度樣平變異數 . . . . .	244
8.2	原始觀測資料與線性迴歸缺少配適度 . . . . .	246

---

8.3	(左上: A) 老鼠藥物劑量 $X$ 與血液中藥物濃度 $Y$ 的散佈圖並加上迴歸線; (右上: B) 藥物劑量樣本平均值 $\bar{X}$ 與血液中藥物濃度樣本平均值 $\bar{Y}$ 的散佈圖並加上迴歸線; (左下: C) 左上圖中迴歸線的殘差 (residual) 與配適值 (fitted value) 的散佈圖; (右下: D) 右上圖中迴歸線的殘差與配適值的散佈圖. . . . .	253
9.1	常見的殘差圖 . . . . .	267
9.2	藥物劑量與老鼠平衡反應測試的劑量反應分析: (1) 劑量 (Dose, $X$ ) 對反應 (Response, $Y$ ) 的散佈圖與簡單線性迴歸線; (2) 殘差對配適值的散佈圖; (3) 殘差對劑量 (Dose, $X$ ) 的散佈圖; (4) 內部標準化殘差 $r_{i,\text{stud}}$ 的 QQ Plot. . . . .	285
9.3	藥物劑量與老鼠平衡反應測試的劑量反應分析: (1) 與 (2) $M_0$ 最小平方法: (1) $M_0$ 殘差對配適值的散佈圖; (2) $M_0$ 內部標準化殘差的 QQ Plot. (3) 與 (4) $M_1$ 加權最小平方法: (3) $M_1$ 殘差對配適值的散佈圖; (4) $M_1$ 內部標準化殘差的 QQ Plot. . . . .	294
9.4	殘差分析 . . . . .	298
9.5	影響值分析 . . . . .	299
9.6	變異數變化分析 . . . . .	300
9.7	大腦重量與體重: (1) 原始資料 brain 對 wt 散佈圖; (2) 對數轉換, $\log(\text{brain})$ 對 $\log(\text{wt})$ 散佈圖; (3) brain $\sim$ wt 的 Box-Cox 指數轉換對數概似函數值; (4) $\log(\text{brain}) \sim \log(\text{wt})$ 的 Box-Cox 指數轉換放大顯示對數概似函數值. . . . .	307
10.1	(1) $C_p$ 對 $(p + 1)$ 散佈圖; (2) $C_p - (p + 1)$ 對 $(p + 1)$ 散佈圖. . . . .	319
10.2	模型選擇標準統計量與參數數目 . . . . .	323
11.1	偏誤-變異數的權衡交易 (Bias-Variance Trade-Off) . . . . .	344

11.2	迴歸模型複雜程度相對於樣本資料與預測資料的預測誤差. . . . .	347
12.1	正規參數 $\lambda$ 與 GCOV score, 迴歸係數 $\beta$ 之相對關係 . . . . .	368
12.2	效應自由度 $df(\lambda)$ 與 迴歸係數 $\beta$ , 正規參數 $\lambda$ , 模型選擇 $AIC, BIC,$ GCV 之相對關係 . . . . .	369
12.3	左側為 lasso (L1) 與 右側為 ridge (L2) 迴歸模型: 參數估計之限制 與誤差函數 (Hastie et al., 2009) . . . . .	371
12.4	lasso (L1) 與ridge (L2) 迴歸模型: 參數先驗分配之差異 . . . . .	371
12.5	左小圖: Lasso 的迴歸係數與 L1 Norm; 右小圖: Lasso 的 $-\log(\lambda)$ 與 CV-MSE. . . . .	377
12.6	Ridge, Lasso 的迴歸係數與 L2 Norm, L1 Norm 之間的差異 . . . . .	377
13.1	老忠實噴泉等待時間 (waiting): 直方圖的機率密度估計. . . . .	381
13.2	插入式估計式 (plug-in estimator) 與 留一交叉驗證法 (LOO-CV) 選取最佳平滑參數 $h$ 對估計風險 $\hat{\mathcal{J}}(h)$ 的影響 . . . . .	385
13.3	核函數估計: 長方盒形核函數中, 3 組不同 $h$ 的估計呈現鋸齒狀. . . . .	386
13.4	核函數: 左上為 boxcar, 右上為 Gaussian, 左下為 Epanechnikov, 右下為 tricube. . . . .	387
13.5	核機率密度估計式 $\hat{f}_n(x_0)$ : 垂直線表示 點 $x_0$ . . . . .	388
13.6	核密度估計: 比較 4 種不同核函數差異. . . . .	389
13.7	Gaussian 核密度估計: 比較 4 種不同平滑參數 $h$ 差異. . . . .	390
13.8	機率密度估計: 左側為直方圖的機率密度估計 右側為Gaussian 核 機率密度估計. . . . .	392
13.9	“等待時間” (waiting) 與 “噴發時間” (eruptions): 散佈圖 與 等高 圖. . . . .	395

---

13.10 “等待時間” (waiting) 與 “噴發時間” (eruptions): 等高密度圖 與 核多變量機率密度估計圖 . . . . .	395
14.1 年齡 (age) 與 脊椎骨質密度 (spbmd) 的關聯性: 局部平均估計法	402
14.2 核函數: 左上為 boxcar, 右上為 Gaussian, 左下為 Epanechnikov, 右下為 tricube. . . . .	404
14.3 年齡 (age) 與 脊椎骨質密度 (spbmd) 的關聯性: 比較 4 種寬帶平 滑參數 . . . . .	406
14.4 局部加權迴歸: 左圖比較不同階的多項式迴歸; 右圖比較男性與女性 的年齡 (age) 與 脊椎骨質密度 (spbmd) 的關聯性. . . . .	411
15.1 青少年脊椎骨質密度: 格子平滑迴歸 . . . . .	416
15.2 青少年脊椎骨質密度: 分段線性迴歸 . . . . .	419
15.3 青少年脊椎骨質密度: 分段連續線性迴歸 . . . . .	420
15.4 立方次樣條基底函數 (cubic B-spline basis): 在 $[0, 1]$ 之間有 9 節點	424
15.5 青少年脊椎骨質密度: Cubic Spline . . . . .	427
15.6 紐約市臭氧濃度資料散佈圖 . . . . .	439
15.7 模型配適顯示 $\log(\widehat{\text{ozone}})$ 與 radiation, wind, temp 之間的關係.	440
16.1 蘇格蘭高地路跑競賽資料散佈圖 . . . . .	446
17.1 $Y$ 的累積機率密度函數 $F_Y(y)$ 與 $Q_Y(\tau)$ 的關係. . . . .	448
17.2 $Y$ 的機率密度函數 $f(y)$ 與 分量 $\tau$ . . . . .	448
17.3 凹向上函數 $\rho_\tau(u)$ 與 $\Psi(u)$ . . . . .	449
17.4 分量迴歸: 個別變數的影響力 . . . . .	452

