

# IBM SPSS Bootstrapping 21



注意：使用此資訊和支援的產品之前，請先閱讀 注意事項第 35 頁 包含的一般資訊。

若新版本未聲明，則此版本便適用於 IBM® SPSS® Statistics 21 以及之後發行的所有版本和修正。

Adobe 產品的擷取畫面已取得 Adobe Systems Incorporated 之翻印許可。

Microsoft 產品的擷取畫面已取得 Microsoft Corporation 之翻印許可。

授權內容：IBM 資產

**Copyright IBM Corporation 1989, 2012.**

美國政府使用者有限權利：使用、複製或披露內容皆受 IBM Corp 簽署之 GSA ADP Schedule Contract 限制約束。

---

# 序

IBM® SPSS® Statistics 為分析資料的強大系統。自助法 的選用性附加模組能提供其他本手冊所說明的分析技術。自助法 的附加模組必須與 SPSS Statistics Core 系統搭配使用，而且是完全整合到系統中。

## 關於 IBM Business Analytics

IBM Business Analytics 軟體提供完整、一致且確實的資訊，決策者可信任此資訊，並藉以改善營運績效。包括商業智慧、預測分析、財務績效和策略管理，以及分析應用程式的整合型產品組合，為目前績效提供了清晰、即時且具行動性的前瞻眼界，以及預測未來成果的能力。結合了豐富的業界解決方案、有效實證和專業服務，每種規模的組織都能引爆最高效能，確實自動化執行決策，並且交付更棒的成果。

在這項產品組合中，IBM SPSS Predictive Analytics 軟體有助於組織預測未來事件，並且針對前瞻概念提前行動，創造更棒的營運成果。全球的商業、政府和學術客戶相當倚重 IBM SPSS 技術所帶來的競爭優勢，藉此做為吸引、保有和發展更多客戶，同時降低可能的不實詐欺風險。藉由將 IBM SPSS 軟體併入每天作業，這些組織成為預測型企業 – 足以駕馭決策並使決策自動化處理，以符合營運目標，並且達到可測知的競爭優勢。如需更多資訊，或是聯絡代表人員，請造訪 <http://www.ibm.com/spss>。

## 技術支援

技術支援可提供客戶維護的服務。客戶可以電洽技術支援以取得 IBM Corp. 產品在使用上的協助，或是支援硬體環境的安裝說明。若要取得技術支援，請參閱 IBM Corp. 網站內容，網址：<http://www.ibm.com/support>。請求協助時，請準備好的您個人、組織和支援合約的相關資訊。

## 針對學生用戶的技術支援

如果您是使用任何 IBM SPSS 軟體產品之學生版、學術版或研究套件版本的學生，請參閱適用於學生的特殊線上「教育解決方案 (<http://www.ibm.com/spss/rd/students/>)」頁面。如果您是使用 IBM SPSS 軟體之大學提供副本的學生，請聯絡您大學的 IBM SPSS 產品協調人員。

## 客戶服務

如果您對於自己的貨品或帳號有任何疑問，請聯絡您的當地辦公室。請備妥您的序號以供識別。

## 訓練研討會

IBM Corp. 同時提供公開與線上訓練研討會。所有的研討會皆以傳達工作群為其特色。研討會將定期在各主要城市舉辦。如需研討會的詳細資訊，請移至 <http://www.ibm.com/software/analytics/spss/training>。

---

# 內容

## 部 I：使用手冊

1	自助法簡介	1
2	自助法	3
	支援自助法的程序	5
	BOOTSTRAP 指令的其他功能	7

## 部 II：範例

3	自助法	9
	使用自助法取得比例的信賴區間	9
	準備資料	9
	執行分析	10
	自助法規格	13
	統計	13
	次數表	14
	使用自助法取得中位數信賴區間	15
	執行分析	15
	描述性統計量(D)	17
	使用自助法選擇最佳的預測量	18
	準備資料	18
	執行分析	19
	參數估計值	26
	閱讀資料推薦	26

## 附錄

A	範例檔案	28
B	注意事項	35
	參考書目	37
	索引	39



# 部 1: 使用手冊





# 自助法簡介

收集資料時，您通常會對取樣的來源母群性質感興趣。您可以利用從樣本計算的估計值對這些母群參數進行推論。例如，若產品隨附的 Employee data.sav 資料集是較大的員工母群中的隨機樣本，則目前薪資的樣本平均數 \$34,419.57 便是員工母群平均目前薪資的估計值。再者，此估計值的樣本大小 474 的標準誤是 \$784.311，因此員工母群中平均目前薪資 95% 的信賴區間是 \$32,878.40 到 \$35,960.73，但這些估計式有多可靠？對某些「已知」的母群和規律的參數，我們大概知道樣本估計值的性質，因而能夠信賴這些結果。自助法會試圖尋找「未知」母群和不規律參數之性質的更多資訊。

圖表 1-1  
對母群平均數進行參數推論

			統計量	標準誤
目前薪資	平均數		\$34,419.57	\$784.311
	平均數的 95% 信賴區間	下限	\$32,878.40	
		上限	\$35,960.73	
	中位數		\$28,875.00	

## 自助法如何運作

簡單而言，對於樣本大小是 N 的資料集，您選定 B 「自助法」樣本大小 N 後放回原始資料集，並計算這些個別的 B 自助法樣本的估計式。這些 B 自助法估計值是大小為 B 的樣本，您可以從中推論估計式。例如，若您從 Employee data.sav 資料集選定 1,000 個自助法樣本，則「目前薪資」之樣本平均數的自助法估計標準誤 \$776.91 是估計值 \$784.311 的替代標準誤。

此外，自助法可提供沒有參數估計值之中位數的標準誤與信賴區間。

圖表 1-2  
對樣本平均數進行自助法推論

			統計量	標準誤	自助法 <sup>a</sup>			
					偏差	標準誤	95% 信賴區間	
							下界	上界
目前薪資	平均數		\$34,419.57	\$784.311	\$14.66	\$776.91	\$32,990.38	\$36,026.06
	平均數的 95% 信賴區間	下限	\$32,878.40					
		上限	\$35,960.73					
	中位數		\$28,875.00		\$-13.22	\$536.63	\$27,750.00	\$29,850.00

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

## 產品中的自助法支援

在支援自助法的程序中，自助法是以子對話方塊的形式體現。請參閱[支援自助法的程序](#)以尋找支援自助法之程序的資訊。

在對話方塊中要求自助法時，除了對話方塊所產生的一般語法以外，還會貼上個別的新 **BOOTSTRAP** 指令。**BOOTSTRAP** 指令會根據您的規格建立自助法樣本。產品於內部會將這些自助法樣本視為分割，即使它們並未明確顯示於「資料編輯程式」中也一樣。這就表示，內部有  $B*N$  個有效的觀察值，因此在自助法執行期間處理資料時，狀態列中的觀察值計數器會從 1 數到  $B*N$ 。「輸出管理系統」(OMS) 是用來收集針對每個「自助法分割」執行分析的結果。這些自助法結果合併後會連同程序產生的其餘一般輸出一併顯示於「瀏覽器」中。在某些情況下，您會看到 “bootstrap split 0” 的參照；這是原始的資料集。

# 自助法

自助法方法可獲得可靠的標準誤估計值和如下列各種估計值的信賴區間：平均數、中位數、比例、odds 比率、相關係數或迴歸係數。自助法方法也可用於建立假設檢定。在這些方法的假設有疑慮時（例如含有不等變異性殘差的迴歸模式符合小型樣本的情況下），或是無法進行參數推論或需要極複雜的公式才能計算標準誤時（例如計算中位數、四分位數及其他百分位數之信賴區間的情況下）自助法是參數估計值最有用的替代方法。

**範例。** 某家電信公司每個月損失約 27% 的客戶。為適當地將焦點放在減少顧客不忠的努力成果上，管理部門想了解此百分比在預先定義的客戶群組上是否有所變化。透過使用自助法，您可以判斷單一顧客不忠比率是否能適當描述四種主要客戶類型。[如需詳細資訊，請參閱第 3 章中的使用自助法取得比例的信賴區間中的 IBM SPSS Bootstrapping 21。](#)

在員工記錄的檢閱中，管理部門有興趣知道員工先前的工作經驗。工作經驗向右偏斜，這代表所期望的員工「典型」先前工作經驗估計值少於中位數。不過，產品中的中位數沒有參數信賴區間。[如需詳細資訊，請參閱第 3 章中的使用自助法取得中位數信賴區間中的 IBM SPSS Bootstrapping 21。](#)

管理部門也有興趣知道哪些因素與員工加薪有關聯，可以透過將線性模式套用到目前薪資與起薪間差異來觀察。當您對線性模式執行自助法時，可以使用特殊重新抽樣方法（殘差與離群自助法）來取得更精確的結果。[如需詳細資訊，請參閱第 3 章中的使用自助法選擇最佳的預測量中的 IBM SPSS Bootstrapping 21。](#)

許多程序都支援對從自助法樣本分析的結果進行自助法抽樣與合併。可指定自助法分析的控制項，已直接整合成為支援自助法之程序中的一般對話方塊。自助法對話方塊中的設定會存留在整個階段作業中，所以如果您透過對話方塊的自助法執行次數分配分析，預設會為支援自助法的其他程序開啟自助法。

## 取得自助法分析

- ▶ 從功能表中選擇支援自助法的程序，並按一下「自助法」。

圖表 2-1  
「自助法」對話方塊

- ▶ 選取「執行自助法」。

您可以選擇性地控制下列選項：

**樣本個數。**對於產生的百分位數與 BCa 區間，建議至少使用 1000 個自助法樣本。指定一個正整數。

**設定 Mersenne Twister 的種子。**設定種子可供您複製分析。這個控制項的用途類似將 Mersenne Twister 設為作用中產生器，並在「亂數產生器」對話方塊上指定固定的起點，但重要的差異在於在此對話方塊中設定種子將保留亂數產生器的目前狀態，並在分析完成後還原該狀態。

**信賴區間。**指定大於 50 且小於 100 的信賴區間。百分位數區間只使用對應至信賴區間百分位數的排序自助法數值。。例如，95% 百分數信賴區間使用自助法值的第 2.5 個與第 97.5 的百分位數作為區間的上界與下界（會視需要內插數值）。已修正偏差與加速 (BCa) 的區間為已調整的區間，因為更為精確，所以也需要更多時間來計算。

**抽樣。**簡易方法會從原始資料集中不斷抽樣觀察值並放回。階層化方法會從原始資料集中不斷抽樣觀察值並放回，此動作是在由層變數其交叉分類所定義的層內進行的。當層之內的單位其同質性相當高，而層之間的單位又非常不同時，階層化自助法抽樣會十分有用。

## 支援自助法的程序

下列程序支援自助法。

注意：

- 自助法無法使用多重插補的資料集。如果資料集中有 Imputation\_ 變數，則「自助法」對話方塊會停用。
- 如果使用非整數加權值的話，將無法使用自助法。
- 自助法使用刪除全部遺漏值來決定觀察值基礎，亦即任何分析變數上含遺漏值的觀察值會自分析中刪除，所以當自助法生效時，刪除全部遺漏值也會生效，即使用分析程序會指定另一種形式的遺漏值處理方法。

### Statistics Base 選項

#### 次數分配表(F)

- 「統計量」表格支援平均數、標準差、變異數、中位數、偏態、峰度與百分位數的自助法估計值。
- 「次數分配」表格支援百分比的自助法估計值。

#### 描述性統計量(D)

- 「敘述統計量」表格支援平均數、標準差、變異數、偏態與峰度的自助法估計值。

#### 預檢資料

- 「描述」表格支援平均數、5% 修整平均數、標準差、變異數、中位數、偏態、峰度與四分位數範圍的自助法估計值。
- 「M 估計式」表格支援下列的自助法估計值：Huber M 估計式、Tukey' s 二權數、Hampel M 式估計值，與 Andrew' s Wave。
- 「百分位數」表格支援百分比的自助法估計值。

#### 交叉表

- 「指向性測量」表格支援下列的自助法估計值：Lambda、Goodman、Kruskal Tau、不確定係數與 Somers' d。
- 「對稱性量數」表格支援下列的自助法估計值：Phi、Cramer' s V、列聯係數、Kendall' s tau-b、Kendall' s tau-c、Gamma、Spearman 相關與 Pearson' s R。
- 「風險估計」表格支援 Odds 比率的自助法估計值。
- 「Mantel-Haenszel 常見 Odds 比率」表格支援自助法估計值與  $\ln(\text{Estimate})$  的顯著性檢定。

#### 平均數

- 「報告」表格支援平均數、中位數、分組中位數、標準差、變異數、峰度、偏態、調和平均數與幾何平均數的自助法估計值。

#### 單一樣本 T 檢定

- 「統計量」表格支援平均數與標準差的自助法估計值。
- 「檢定」表格支援平均數差異的自助法估計值與顯著性檢定。

#### **獨立樣本 T 檢定**

- 「組別統計量」表格支援平均數與標準差的自助法估計值。
- 「檢定」表格支援平均數差異的自助法估計值與顯著性檢定。

#### **成對樣本 T 檢定**

- 「統計量」表格支援平均數與標準差的自助法估計值。
- 「相關性」表格支援相關的自助法估計值。
- 「檢定」表格支援平均數的自助法估計值。

#### **單因子變異數分析**

- 「敘述統計量」表格支援平均數與標準差的自助法估計值。
- 「多重比較」表格支援平均數差異的自助法估計值。
- 「對比檢定」表格支援對比值的自助法估計值與顯著性檢定。

#### **GLM 單變量**

- 「敘述統計量」表格支援平均數與標準差的自助法估計值。
- 「參數估計值」表格支援係數 B 的自助法估計值與顯著性檢定。
- 「對比結果」表格支援差異的自助法估計值與顯著性檢定。
- 邊際平均數估計：「估計值」表格支援平均數的自助法估計值。
- 邊際平均數估計：「成對比較」表格支援平均數差異的自助法估計值。
- Post Hoc 檢定：「多重比較」表格支援平均數差異的自助法估計值。

#### **雙變數相關分析**

- 「敘述統計量」表格支援平均數與標準差的自助法估計值。
- 「相關性」表格支援相關的自助法估計值及顯著性檢定。

注意：

如果除了 Pearson 相關外，還要求無母數相關性 (Kendall 的 tau-b 或 Spearman)，則對話方塊會貼上 `CORRELATIONS` 與 `NONPAR CORR` 指令，每個指令包含個別的 `BOOTSTRAP` 指令。相同的自助法樣本會用於計算所有的相關。

合併之前，Fisher Z 轉換會套用至相關。合併之後，會套用逆 Z 轉換。

#### **偏相關**

- 「敘述統計量」表格支援平均數與標準差的自助法估計值。
- 「相關性」表格支援相關的自助法估計值。

#### **線性迴歸**

- 「敘述統計量」表格支援平均數與標準差的自助法估計值。

- 「相關性」表格支援相關的自助法估計值。
- 「模式摘要」表格支援 Durbin-Watson 的自助法估計值。
- 「係數」表格支援係數 B 的自助法估計值與顯著性檢定。
- 「相關係數」表格支援相關的自助法估計值。
- 「殘差統計量」表格支援平均數與標準差的自助法估計值。

#### 次序迴歸

- 「參數估計值」表格支援係數 B 的自助法估計值與顯著性檢定。

#### 判別分析

- 「標準典型判別函數係數」表格支援標準化係數的自助法估計值。
- 「典型判別函數係數」表格支援未標準化係數的自助法估計值。
- 「分類函數係數」表格支援係數的自助法估計值。

#### 進階統計量選項

##### GLM 多變量

- 「參數估計值」表格支援係數 B 的自助法估計值與顯著性檢定。

##### 線性混合模式

- 「固定效果估計」表格支援估計值的自助法估計值與顯著性檢定。
- 「估計共變異數參數」表格支援估計值的自助法估計值與顯著性檢定。

##### 概化線性模式

- 「參數估計值」表格支援係數 B 的自助法估計值與顯著性檢定。

##### Cox 迴歸

- 「在方程式中的變數」表格支援係數 B 的自助法估計值與顯著性檢定。

##### 迴歸選項

##### 二元 Logistic 迴歸

- 「在方程式中的變數」表格支援係數 B 的自助法估計值與顯著性檢定。

##### 多項式 Logistic 迴歸

- 「參數估計值」表格支援係數 B 的自助法估計值與顯著性檢定。

## BOOTSTRAP 指令的其他功能

指令語法語言也可以讓您：

- 執行殘差與離群自助法抽樣 (SAMPLING 次指令)

如需完整的語法資訊，請參閱《指令語法參考手冊》。

# 部 11: 範 例



# 自助法

自助法方法可獲得可靠的標準誤估計值和如下列各種估計值的信賴區間：平均數、中位數、比例、odds 比率、相關係數或迴歸係數。自助法方法也可用於建立假設檢定。在這些方法的假設有疑慮時（例如含有不等變異性殘差的迴歸模式符合小型樣本的情況下），或是無法進行參數推論或需要極複雜的公式才能計算標準誤時（例如計算中位數、四分位數及其他百分位數之信賴區間的情況下）自助法是參數估計值最有用的替代方法。

## 使用自助法取得比例的信賴區間

某家電信公司每個月損失約 27% 的客戶。為適當地將焦點放在減少顧客不忠的努力成果上，管理部門想了解此百分比在預先定義的客戶群組上是否有所變化。

此資訊收集於 telco.sav 中。[如需詳細資訊，請參閱第 28 頁附錄 A 中的範例檔案。](#) 透過使用自助法判定單一顧客不忠比率是否能適當描述四種主要客戶類型。

注意：本範例使用「次數分配表」程序，且需要 Statistics Base 選項。

## 準備資料

您必須先透過「客戶類別」來分割檔案。

- ▶ 若要分割檔案，請從「資料編輯程式」功能表中選擇：  
資料 > 分割檔案(F)...

圖表 3-1  
「分割檔案」對話方塊



- ▶ 選取比較組別。
- ▶ 選取「客戶類別」作為群組所依據的變數。
- ▶ 按一下「確定」。

## 執行分析

- ▶ 若要取得比例的自助法信賴區間，請從功能表中選擇：  
分析 (A) > 敘述統計 > 次數分配表...

圖表 3-2  
「次數分配表」主對話方塊



- ▶ 選取「最後一個月份內的異動 [churn]」作為分析中的變數。
- ▶ 按一下「統計量」。

圖表 3-3  
「統計量」對話方塊



- ▶ 選取「集中趨勢」群組中的「平均數」。
- ▶ 按一下「繼續」。
- ▶ 按一下「次數分配表」對話方塊中的「自助法」。

圖表 3-4  
「自助法」對話方塊

- ▶ 選取「執行自助法」。
- ▶ 若要完全複製此範例中的結果，請選取「設定 Mersenne Twister 的種子」，並視需要輸入 9191972。
- ▶ 按一下「繼續」。
- ▶ 按一下「次數分配表」對話方塊中的「確定」。

這些選擇會產生下列指令語法：

```

SORT CASES BY custcat.
SPLIT FILE LAYERED BY custcat.
PRESERVE.
SET RNG=MT MTINDEX=9191972.
SHOW RNG.
BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES INPUT=churn
  /CRITERIA CILEVEL=95 CITYPE=PERCENTILE NSAMPLES=1000
  /MISSING USERMISSING=EXCLUDE.
FREQUENCIES VARIABLES=churn
  /STATISTICS=MEAN
  /ORDER=ANALYSIS.
RESTORE.

```

- SORT CASES 與 SPLIT FILE 指令會在變數 custcat 上分割檔案。
- PRESERVE 與 RESTORE 指令會「記住」亂數產生器的目前狀態，並在自助法結束時將系統還原至該狀態。
- SET 指令會將亂數產生器設為 Mersenne Twister，將索引設為 9191972，因此可完全複製自助法的結果。SHOW 指令會在輸出中顯示索引以供參照。
- BOOTSTRAP 指令使用簡式重新取樣方法來要求 1,000 個自助法樣本。
- 變數 churn 用於決定重新取樣的觀察值基礎。系統會將此變數上含遺漏值的記錄自分析中刪除。
- 在 BOOTSTRAP 後，會在每個自助法樣本上執行 FREQUENCIES 程序。
- STATISTICS 子指令會在原始資料上產生變數異動的平均數。此外，系統會在次數分配表中為平均數和百分比產生合併的統計量。

## 自助法規格

圖表 3-5  
自助法規格

取樣方法	簡單	
樣本個數		1000
信賴區間水準		95.0%
信賴區間類型	百分位數	

自助法規格表包含重新取樣時使用的設定，在檢查您所要的分析是否已執行時，此規格表是很有用的參考。

## 統計

圖表 3-6  
含比例自助法信賴區間的統計量表

上個月客戶流失

客戶類別	個數	有效的 遺漏值	Statistic	自助法 <sup>a</sup>			
				偏差	標準誤差	95% 信賴區間	
						下界	上界
基本服務	266	0	266	0	0	266	266
	0	0	0	0	0	0	0
	.31	.00	.03	.26	.37		
E化服務	217	0	217	0	0	217	217
	0	0	0	0	0	0	0
	.27	.00	.03	.21	.34		
加值服務	281	0	281	0	0	281	281
	0	0	0	0	0	0	0
	.16	.00	.02	.12	.20		
全方位服務	236	0	236	0	0	236	236
	0	0	0	0	0	0	0
	.37	.00	.03	.31	.44		

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

針對「客戶類別」的每一個層級，統計量表會顯示「最後一個月份內的異動」中的平均值。由於「最後一個月份內的異動」只會取值 0 和 1，因此 1 表示客戶不忠，平均數等於不忠客戶的比例。「統計量」欄會顯示通常由「次數分配表」使用原始資料集產生的數值。自助法演算法會產生「自助法」欄。

- 「偏差」是指此統計量在整個自助法樣本中的平均值，與「統計量」欄的值之間的差異。在此狀況下，系統會計算所有 1000 個自助法樣本的「最後一個月份內的異動」平均值，然後計算這些平均值的平均。
- 標準誤是整個 1000 個自助法樣本中「最後一個月份內的異動」的平均值標準誤。
- 如果 1000 個自助法樣本以遞增順序排序，則 95% 自助法信賴區間的下界是將「最後一個月份內的異動」的第 25 個與第 26 個平均值以內插法計算的結果。上界則是第 975 個與第 976 個平均值以內插法計算的結果。

表格中的結果表示不同客戶類型間的客戶不忠比率不同。具體來說，「加值服務」客戶的信賴區間不與任何其他客戶重疊，這表示平均而言，這些客戶比較不容易離開。

當使用的類別變數只有兩個值時，這些信賴區間可替代由「單一樣本無母數檢定」程序或「單一樣本 T 檢定」程序產生的信賴區間。

## 次數表

圖表 3-7  
含比例自助法信賴區間的次數分配表

客戶類別	有效的		次數	百分比	有效百分比	累積百分比	百分比的自助法 <sup>a</sup>			
							偏差	標準誤差	95% 信賴區間	
									下界	上界
基本服務	否		183	68.8	68.8	68.8	.0	2.8	63.2	74.4
	是		83	31.2	31.2	100.0	.0	2.8	25.6	36.8
	總和		266	100.0	100.0		.0	.0	100.0	100.0
E化服務	有效的	否	158	72.8	72.8	72.8	.1	3.1	66.4	78.8
	是		59	27.2	27.2	100.0	-.1	3.1	21.2	33.6
	總和		217	100.0	100.0		.0	.0	100.0	100.0
加值服務	有效的	否	237	84.3	84.3	84.3	.0	2.1	80.1	88.3
	是		44	15.7	15.7	100.0	.0	2.1	11.7	19.9
	總和		281	100.0	100.0		.0	.0	100.0	100.0
全方位服務	有效的	否	148	62.7	62.7	62.7	.0	3.2	56.4	69.1
	是		88	37.3	37.3	100.0	.0	3.2	30.9	43.6
	總和		236	100.0	100.0		.0	.0	100.0	100.0

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

「次數分配」表顯示每個類別的百分比信賴區間（比例 × 100%），因此可供所有類別變數使用。可比較的信賴區間無法在本產品的他處使用。

## 使用自助法取得中位數信賴區間

在員工記錄的檢閱中，管理部門有興趣知道員工先前的工作經驗。工作經驗向右偏斜，這代表所期望的員工「典型」先前工作經驗估計值少於中位數。但是若未進行自助法，產品的統計程序一般不會提供中位數的信賴區間。

此資訊收集於 Employee data.sav 之中。如需詳細資訊，請參閱第 28 頁附錄 A 中的範例檔案。透過使用自助法取得中位數的信賴區間。

注意：本範例使用「預檢資料」程序，且需要 Statistics Base 選項。

### 執行分析

- ▶ 若要取得中位數的自助法信賴區間，請從功能表中選擇：  
分析(A) > 敘述統計 > 預檢資料...

圖表 3-8  
「預檢資料」主對話方塊



- ▶ 選取「先前經驗（月份） [prevep]」作為依變數。
- ▶ 選取「顯示」群組中的「統計量」。
- ▶ 按一下「自助法」。

圖表 3-9  
「自助法」對話方塊

- ▶ 選取「執行自助法」。
- ▶ 若要完全複製此範例中的結果，請選取「設定 Mersenne Twister 的種子」，並視需要輸入 592004。
- ▶ 若要取得更精確的區間（但需花費更長處理時間），請選取「加速偏差修正 (BCa)」。
- ▶ 按一下「繼續」。
- ▶ 按一下「預檢資料」對話方塊中的「確定」。

這些選擇會產生下列指令語法：

```

PRESERVE.
SET RNG=MT MTINDEX=592004.
SHOW RNG.
BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES TARGET=prevexp
  /CRITERIA CILEVEL=95 CITYPE=BCA NSAMPLES=1000
  /MISSING USERMISSING=EXCLUDE.
EXAMINE VARIABLES=prevexp
  /PLOT NONE
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE

```



/NOTOTAL.  
RESTORE.

- PRESERVE 與 RESTORE 指令會「記住」亂數產生器的目前狀態，並在自助法結束時將系統還原至該狀態。
- SET 指令會將亂數產生器設為 Mersenne Twister，將索引設為 592004，因此可完全複製自助法的結果。SHOW 指令會在輸出中顯示索引以供參照。
- BOOTSTRAP 指令使用簡式重新取樣方法來要求 1000 個自助法樣本。
- VARIABLES 子指令會指定使用變數 prevexp 來判斷重新取樣的觀察值基礎。系統會將此變數上含遺漏值的記錄自分析中刪除。
- 除了要求自助法樣本數目外，CRITERIA 子指令也會要求加速偏差修正的自助法信賴區間，而非預設的百分位區間。
- 在 BOOTSTRAP 程序後，會在每個取樣法樣本上執行 EXAMINE 程序。
- PLOT 子指令會關閉圖形輸出。
- 所有其他的選項皆設為其預設值。

## 描述性統計量 (D)

圖表 3-10  
含自助法信賴區間的敘述性表格

	統計量	標準誤	自助法 <sup>a</sup>				
			偏差	標準誤	BCa 95% 信賴區間		
					下界	上界	
以前的資歷(月)	平均數	95.86	4.804	-0.01	4.86	86.39	105.20
	平均數的 95% 信賴區間						
	下限	86.42					
	上限	105.30					
	刪除兩極端各 5% 觀察值之平均數	84.64		.02	4.94	75.38	94.21
	中位數	55.00		-.11	3.66	50.00	60.00
	變異數	10938.281		18.783	977.081	8954.509	13057.229
	標準差	104.586		-.015	4.689	94.644	114.245
	最小值	0					
	最大值	476					
	範圍	476					
	四分位全距	121		-1	10	103	137
	偏態	1.510	.112	.006	.110	1.284	1.768
	峰度	1.696	.224	.040	.463	.823	2.876

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

敘述性表格包含一些統計量，以及這些統計量的自助法信賴區間。平均數 (86.39, 105.20) 的自助法信賴區間與參數信賴區間 (86.42, 105.30) 類似，並可表示「典型」員工過去約有 7-9 年的工作經驗。但是，「先前經驗 (月)」的分配偏斜，這讓中位數比平均數更適合作為「典型」目前薪資的指標。中位數 (50.00, 60.00) 的自助法信賴區間的值比平均數的信賴區間的值更低、範圍更窄，可表示「典型」員工約 4-5 年的過去工作經驗。透過使用自助法可讓數值範圍更可代表典型的先前工作經驗。

## 使用自助法選擇最佳的預測量

在審視員工記錄時，管理階層想要透過將線性模式與目前薪資和起薪之間的差異進行配適，來判斷哪些因子與員工薪資增加有關。當您對線性模式執行自助法時，可以使用特殊重新取樣方法（殘差與離群自助法）來取得更精確的結果。

此資訊收集於 Employee data.sav 之中。如需詳細資訊，請參閱第 28 頁附錄 A 中的範例檔案。

注意：本範例使用「GLM 單變量」程序，且需要 Statistics Base 選項。

### 準備資料

您必須先計算「目前薪資」與「起薪」之間的差異。

- ▶ 從功能表選擇：  
轉換(T) > 計算變數(C)...

圖表 3-11  
「計算變數」對話方塊



- ▶ 輸入 diff 作為目標變數。
- ▶ 輸入 salary-salbegin 作為數值運算式。

- ▶ 按一下「確定」。

## 執行分析

若要以離群殘差自助法執行「GLM 單變量」，您必須先建立殘差。

- ▶ 從功能表選擇：  
分析(A) > 一般線性模式 > 單變量...

圖表 3-12  
「GLM 單變量」主對話方塊



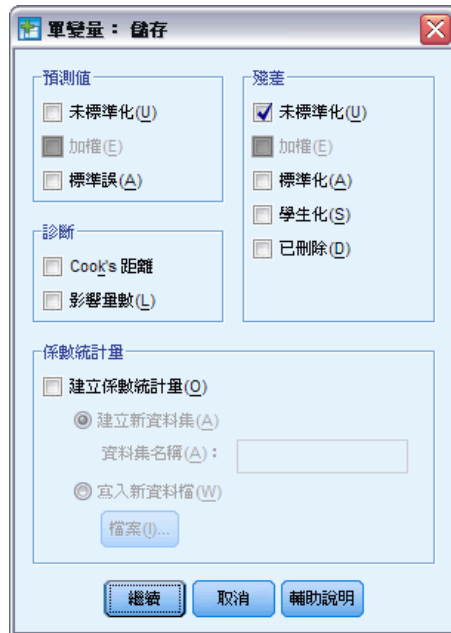
- ▶ 選取 diff 作為依變數。
- ▶ 選取「性別 [gender]」、「雇用類別 [jobcat]」與「少數類別 [minority]」作為固定因子。
- ▶ 選取「雇用月數 [jobtime]」與「先前經驗 (月) [prevexp]」作為共變量。
- ▶ 按一下「模式」。

圖表 3-13  
「模式」對話方塊



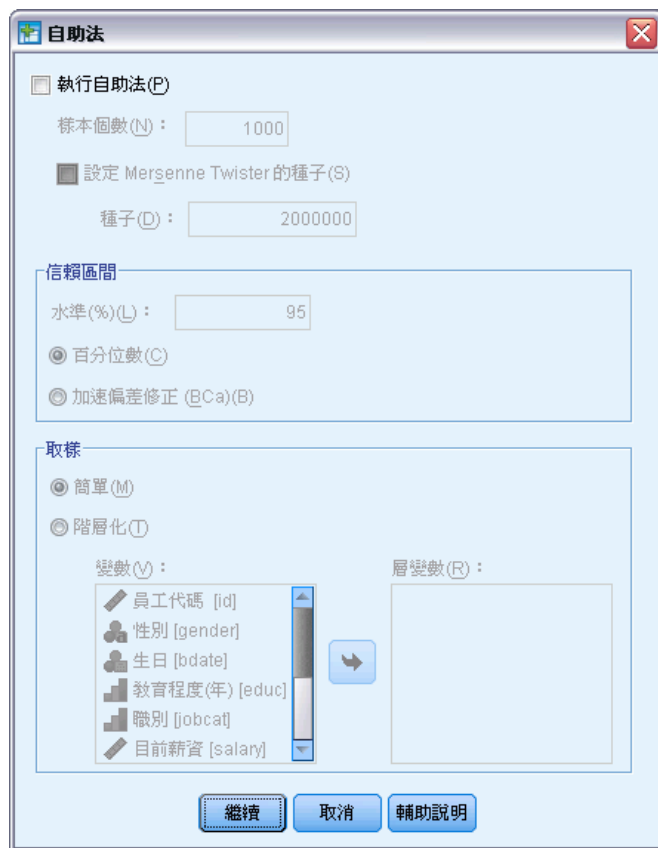
- ▶ 從「建立效果項」下拉式清單中，依序選取「自訂」與「主效果」。
- ▶ 選取「gender」到「prevexp」作為模式項目。
- ▶ 按一下「繼續」。
- ▶ 按一下「GLM 單變量」對話方塊中的「儲存」。

圖表 3-14  
「儲存」對話方塊



- ▶ 選取「殘差」組別中的「未標準化」。
- ▶ 按一下「繼續」。
- ▶ 按一下「GLM 單變量」對話方塊中的「自助法」。

圖表 3-15  
「自助法」對話方塊



- ▶ 如果有需要，請取消選取「執行自助法。」
- ▶ 按一下「GLM 單變量」對話方塊中的「確定」。資料集現在包含新變數 RES\_1，此變數包含此模式中未標準化殘差。
- ▶ 叫回「GLM 單變量」對話方塊，並按一下「儲存」。

- ▶ 取消選取「未標準化」，然後按一下「繼續」和「GLM 單變量」對話方塊中的「選項」。

圖表 3-16  
「選項」對話方塊



圖表 3-17  
「自助法」對話方塊

- ▶ 選取「執行自助法」。
- ▶ 若要完全複製此範例中的結果，請選取「設定 Mersenne Twister 的種子」，並視需要輸入 9191972。
- ▶ 此對話方塊未提供任何選項可執行離群自助法，因此請按一下「繼續，然後按一下「GLM 單變量」對話方塊中的「貼上」。

這些選擇會產生下列指令語法：

```
PRESERVE.
SET RNG=MT MTINDEX=9191972.
SHOW RNG.
BOOTSTRAP
  /SAMPLING METHOD=SIMPLE
  /VARIABLES TARGET=diff INPUT=gender jobcat minority jobtime prevexp
  /CRITERIA CILEVEL=95 CITYPE=PERCENTILE NSAMPLES=1000
  /MISSING USERMISSING=EXCLUDE.
UNIANOVA diff BY gender jobcat minority WITH jobtime prevexp
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PRINT=PARAMETER
```



```

/CRITERIA=ALPHA(.05)
/DESIGN=gender jobcat minority jobtime prevexp.
RESTORE.

```

為了執行離群自助法取樣，請編輯 **SAMPLING** 次指令的 **METHOD** 關鍵字，以讀取 **METHOD=WILD (RESIDUALS=RES\_1)**。

指令語法的「最後」集合如下所示：

```

PRESERVE.
SET RNG=MT MTINDEX=9191972.
SHOW RNG.
BOOTSTRAP
/SAMPLING METHOD=WILD(RESIDUALS=RES_1)
/VARIABLES TARGET=diff INPUT=gender jobcat minority jobtime prevexp
/CRITERIA CILEVEL=95 CITYPE=PERCENTILE NSAMPLES=1000
/MISSING USERMISSING=EXCLUDE.
UNIANOVA diff BY gender jobcat minority WITH jobtime prevexp
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/PRINT=PARAMETER
/CRITERIA=ALPHA(.05)
/DESIGN=gender jobcat minority jobtime prevexp.
RESTORE.

```

- **PRESERVE** 與 **RESTORE** 指令會「記住」亂數產生器的目前狀態，並在自助法結束時將系統還原至該狀態。
- **SET** 指令會將亂數產生器設為 Mersenne Twister，將索引設為 9191972，因此可完全複製自助法的結果。**SHOW** 指令會在輸出中顯示索引以供參照。
- **BOOTSTRAP** 指令使用離群重新取樣與 **RES\_1** 作為包含殘差的變數，要求 1000 個自助法樣本。
- **VARIABLES** 子指令指定 **diff** 是線性模式中的目標變數，該子指令與變數 **gender**、**jobcat**、**minority**、**jobtime** 和 **prevexp** 用於決定重新取樣的觀察值基礎。系統會將這些變數上含遺漏值的記錄自分析中刪除。
- 除了要求自助法樣本數目外，**CRITERIA** 子指令也會要求加速偏差修正的自助法信賴區間，而非預設的百分位區間。
- **BOOTSTRAP** 之後，會在每個自助法樣本上執行 **UNIANOVA** 程序，此程序會產生原始資料的參數估計量。此外，系統會為模式係數產生合併的統計量。

## 參數估計值

圖表 3-18  
參數估計值 (P)

依變數: diff

參數	B 之估計值	標準誤差	t	顯著性	95% 信賴區間	
					下界	上界
截距	22789.014	2920.700	7.803	.000	17049.673	28528.355
[gender=f]	-4085.253	726.416	-5.624	.000	-5512.701	-2657.804
[gender=m]	0 <sup>a</sup>	.	.	.	.	.
[jobcat=1]	-17717.706	939.798	-18.853	.000	-19564.463	-15870.949
[jobcat=2]	-13101.918	1780.683	-7.358	.000	-16601.061	-9602.776
[jobcat=3]	0 <sup>a</sup>	.	.	.	.	.
[minority=0]	1332.363	819.349	1.626	.105	-277.705	2942.431
[minority=1]	0 <sup>a</sup>	.	.	.	.	.
jobtime	145.539	32.586	4.466	.000	81.505	209.572
prevexp	-21.423	3.575	-5.993	.000	-28.447	-14.398

a.由於這個參數重複，所以把它設成零。

「參數估計值」表顯示模式項目的一般非自助法參數估計值。[minority=0] 的顯著值 0.105 大於 0.05，表示「少數類別」對於薪資的增加沒有影響。

圖表 3-19  
自助法參數估計量

依變數: diff

參數	B 之估計值	自助法 <sup>a</sup>				
		偏差	標準誤差	顯著性(雙尾)	95% 信賴區間	
					下界	上界
截距	22789.014	-95.084	3280.762	.001	16079.630	28835.063
[gender=f]	-4085.253	32.480	622.971	.001	-5365.321	-2892.131
[gender=m]	0	0	0	.	0	0
[jobcat=1]	-17717.706	46.324	1454.230	.001	-20671.451	-14889.507
[jobcat=2]	-13101.918	47.958	1753.311	.001	-16658.596	-9671.891
[jobcat=3]	0	0	0	.	0	0
[minority=0]	1332.363	-10.592	651.144	.012	57.831	2642.534
[minority=1]	0	0	0	.	0	0
jobtime	145.539	.707	35.285	.001	79.081	217.761
prevexp	-21.423	-.065	2.859	.001	-27.533	-16.055

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

現在請看一下「參數估計值自助法」表格。在「標準誤」欄中，您會看到一些係數（例如截距）的參數標準誤過小，而無法與自助法估計量比較，因此信賴區間更廣。對於某些係數，例如 [minority=0]，參數標準誤過大，而由自助法結果報告的顯著值 0.006（小於 0.05）表示所觀察到少數員工與非少數員工之間薪資增加的差異不是偶然。管理階層現在知道這個差異是值得進一步調查，以判定可能的原因。

## 閱讀資料推薦

如需自助法的詳細資訊，請參閱下列文字：

Davison, A. C., 和 D. V. Hinkley. 2006. *Bootstrap Methods and their Application* (自助法方法和其應用程式). : Cambridge University Press (劍橋大學出版部).

Shao, J., 和 D. Tu. 1995. *The Jackknife and Bootstrap* (Jackknife 和自助法). 紐約: Springer.

# 範例檔案

與產品同時安裝的範例檔存放在安裝目錄的範例子目錄中。在下列每種語言的「範例」子目錄中存有個別資料夾：英文、法文、德文、義大利文、日文、韓文、波蘭文、俄文、簡體中文、西班牙文和繁體中文。

並非所有範例檔案皆提供各種語言。如果範例檔案沒提供您需要的語言，語言資料夾有英文版的範例檔案。

## 說明

以下是使用於本文件中不同範例的範例檔之簡要描述。

- **accidents.sav**。這是有關某保險公司研究年齡和性別風險因子對給定地區汽車意外事件的假設資料檔。每一個觀察值對應至一個年齡類別和性別的交叉分類。
- **adl.sav**。這是有關致力於確定一個建議中風病患治療類型之效益的假設資料檔。醫師隨機指定女性中風病患至兩個組別之一。第一組接受標準的物理治療，而第二組則接受額外的情緒治療。在治療了三個月後，將每一個病患進行日常活動的能力記分為次序變數。
- **advert.sav**。這是有關一家零售商致力於調查廣告費與廣告後銷售情形之間的關係的假設資料檔。為了這個目的，他們收集了過往銷售數字和相關的廣告費用。
- **aflatoxin.sav**。這是有關檢定玉米作物是否有黃麴毒素（一種毒物，其濃度在介於和處於作物產量中都有很大的差異）的假設資料檔。一名穀物加工者收到來自 8 個作物產量各 16 個樣本，並以十億當量（PPB）來測量黃麴毒素的水準。
- **anorectic.sav**。在將厭食/暴食行為症狀學標準化的過程中，研究人員（Van der Ham, Meulman, Van Strien, 和 Van Engeland, 1997）研究了 55 個飲食失調的青少年。每個病患在四年之中被訪問四個回合，所以得到總數為 220 的觀察值。在每次觀察中，為病患在 16 種症狀上逐一評分。目前遺漏了第二次訪察的病患 71，第二次訪察的病患 76，以及第三次訪察的病患 47 的症狀分數，因此只剩下 217 個有效觀察值。
- **bankloan.sav**。這是有關一家銀行致力於減少放款利率預設值的假設資料檔。本檔包含 850 位以前的客戶與現在的準客戶的財務和人口資料。前 700 個觀察值為以前有借貸的客戶。最後 150 個觀察值是銀行需要作信用風險優良與不良分類的準客戶。
- **bankloan\_binning.sav**。這是包含 500 位以前客戶的財務和人口資料的假設資料檔。
- **behavior.sav**。在典型範例（Price 和 Bouffard, 1974）中，52 名學生被要求為 15 種情境與 15 種行為組合評等，等級共分為 10 點，從 0 = 「非常適當」到 9 = 「非常不適當」。平均值超過個別值，值會被視為相異性。
- **behavior\_ini.sav**。本資料檔包含 behavior.sav 之二維解的起始組態。
- **brakes.sav**。這是有關一間生產高性能汽車碟型煞車片工廠中品質管制的假設資料檔。資料檔包含由 8 個生產機器分別取得 16 個碟片的直徑測量。煞車的目標直徑是 322 公釐。

- **breakfast.sav.** 在經典研究中(Green 和 Rao, 1972), 21 名 Wharton 學院 MBA 學生及其配偶被要求為 15 項早餐食品按喜愛程度分出等級: 從 1 = 「最喜愛」到 15 = 「最不喜愛」。他們的喜愛程度分六種不同情況記錄, 從「整體喜愛」到「點心, 僅配飲料」。
- **breakfast-overall.sav.** 本資料檔只包含第一種情況—「整體喜愛」—所喜愛的早餐項目。
- **broadband\_1.sav.** 這是包含全國性寬頻服務地區用戶數目的假設資料檔。本資料檔包含四年期間 85 個地區每月的用戶數目。
- **broadband\_2.sav.** 本資料檔與 broadband\_1.sav 相同, 但多了三個月的資料。
- **car\_insurance\_claims.sav.** 一個在別處 (McCullagh 和 Nelder, 1989) 出現和分析過, 有關汽車損害理賠的資料集。理賠金額的平均數可建立模式為具有 gamma 分配, 使用反連結函數將依變數的平均數相關至一被保險人年齡、車輛類型、和車齡的線性組合。提出理賠的數量可以用作尺度權重。
- **car\_sales.sav.** 本資料檔包含假設性的銷售估計、定價、和不同的品牌與車輛型式的實體規格。定價和實體規格是由 edmunds.com 和製造商處輪流取得。
- **car\_sales\_uprepared.sav.** 這是 car\_sales.sav 的修改版本, 其中不包含任何欄位的轉換版本。
- **carpet.sav.** 在一個普遍的範例 (Green 和 Wind, 1973) 中, 計劃銷售全新地毯清潔機的公司想要檢驗影響消費者偏好的五個因子—包裝設計、品牌名稱、價格、「優秀家用品」獎章及退費保證。包裝設計有三個因子水準, 每個水準中的清潔刷位置都不相同; 三個品牌名稱 (K2R、Glory、及 Bissell); 三個價格水準; 且最後兩個因子各有兩個水準 (無論無或有)。十名消費者將這些因子所定義的 22 種組合分級。「偏好」變數包含每個組合平均排名的等級。排名數值較小者會對應高偏好程度。這個變數反映每個組合偏好的整體量數。
- **carpet\_prefs.sav.** 本資料檔是根據 carpet.sav 所描述的相同範例, 但它包含 10 個消費者每一個人的實際等級。消費者被要求將 22 個產品組合從最喜歡排列到最不喜歡。變數「PREF1」到「PREF22」包含相關組合的識別碼, 如 carpet\_plan.sav 中所定義。
- **catalog.sav.** 本資料檔包含郵購公司銷售三項產品的每月假設銷售數字。也包含五個可能預測變數的資料。
- **catalog\_seasfac.sav.** 本資料檔與 catalog.sav 相同, 不過多了一組由「週期性分解」程序所計算的週期性因子以及隨附的資料變數。
- **cellular.sav.** 這是有關一家手機公司致力於減少顧客不忠的假設資料檔。顧客不忠傾向分數套用於帳戶, 範圍由 0 至 100。帳戶分數 50 或以上有可能正尋求變更供應商。
- **ceramics.sav.** 這是有關一家製造商致力於確定一種新的優良合金是否較標準的合金有較大的耐熱性的假設資料檔。每一個觀察值代表對合金之一的不同檢定; 記錄了讓軸承失效的溫度。
- **cereal.sav.** 這是有關對 880 人的早餐喜好進行訪談的假設資料檔, 也記下他們的年齡、性別、婚姻狀況、和是否有活躍的生活型態 (根據他們是否一週運動兩次)。每一個觀察值代表一位不同的應答者。
- **clothing\_defects.sav.** 這是有關一家服裝工廠品質管制過程的假設資料檔。由該工廠所生產的每一批產品中, 檢查員取出一件服裝的樣本並計算不合格的服裝個數。

- **coffee.sav.** 本資料檔是關於六種冰咖啡品牌的感覺印象(Kennedy, Riquier, 和 Sharp, 1996)。對 23 種冰咖啡中每一種的印象屬性, 由群眾來選取依其屬性描述的所有品牌。該六種品牌已標示為 AA、BB、CC、DD、EE、和 FF, 以保持機密。
- **contacts.sav.** 這是有關一群公司電腦銷售代表聯絡清單的假設資料檔。每一個聯絡人依他們在公司所服務的部門及其公司的等級而分類。最後一次銷售的金額、到最後一次銷售的時間、和該聯絡人公司的規模也都被列入記錄。
- **creditemo.sav.** 這是有關一家百貨公司致力於評估近期信用卡促銷活動效果的假設資料檔。為達此目標, 隨機選取了 500 位持卡人。有半數收到廣告, 促銷在未來三個月購買將獲得降低利率的優惠。半數收到標準的週期性廣告。
- **customer\_dbase.sav.** 這是有關一家公司致力於使用其資料倉庫的資訊來對最有可能回應的客戶提供優惠的假設資料檔。隨機選取客戶庫的子集, 提供優惠, 再將他們的回應記錄下來。
- **customer\_information.sav.** 本檔案是包含客戶郵寄資訊的假設資料檔, 例如姓名和地址。
- **customer\_subset.sav.** 80 個 customer\_dbase.sav 的觀察值子集。
- **debate.sav.** 這是有關一項政治辯論會參與者辯論前和辯論後接受調查之成對反應的假設資料檔。每一個觀察值對應至一位不同的應答者。
- **debate\_aggregate.sav.** 這是將 debate.sav 中之反應作整合的假設資料檔。每一個觀察值對應至辯論前和辯論後對偏好之交叉分類的反應。
- **demo.sav.** 這是有關提供郵寄每月優惠之購買客戶資料庫的假設資料檔。記錄了客戶是否對該優惠回應, 以及各種的人口資訊。
- **demo\_cs\_1.sav.** 這是有關一家公司致力於匯編調查資訊資料庫之第一步的假設資料檔。每一個觀察值對應至一個不同的城市, 也記錄了其地區、省、區、和城市識別。
- **demo\_cs\_2.sav.** 這是有關一家公司致力於匯編調查資訊資料庫之第二步的假設資料檔。每一個觀察值對應至在第一步中選取的城市中的一個不同的家庭單位, 也記錄了其地區、省、區、分區、和單位識別。也納入了由該設計的前兩階段所得之取樣資訊。
- **demo\_cs.sav.** 這是包含以複合取樣設計所收集之調查資訊的假設資料檔。每一個觀察值對應至一個不同的家庭單位, 也記錄了各種的人口和取樣資訊。
- **dmdata.sav.** 這是包含直效行銷公司之人口和購買資訊的假設資料檔。dmdata2.sav 包含收到測試郵件的連絡人子集資訊, 而 dmdata3.sav 則包含剩下未收到測試郵件的連絡人資訊。
- **dietstudy.sav.** 本假設資料檔包含對「Stillman 飲食法」(Rickman, Mitchell, Dingman, 和 Dalen, 1974) 研究的結果。每一個觀察值對應至一個不同的受試者, 並記錄下他或她飲食法前、後之體重(磅)和三酸甘油酯水準(毫克/100 毫升)。
- **dvdplayer.sav.** 這是有關新 DVD 播放器開發的假設資料檔。市場行銷團隊使用原型收集了焦點組別資料。每一個觀察值對應至不同調查到的使用者, 並記錄下一些有關他們的人口資訊和他們對有關原型問題的回應。
- **german\_credit.sav.** 本資料檔取自(Blake 和 Merz, 1998) 艾文(Irvine) 在加州大學機器學習資料庫儲存器的「德國信用」資料集。
- **grocery\_1month.sav.** 本假設資料檔是將 grocery\_coupons.sav 資料檔和每週購買的「彙總」, 因此每一個觀察值對應至一個不同的客戶。結果部份每週變更的變數消失了, 而目前所記錄的銷售量是在研究的四週期間銷售量之總和。

- **grocery\_coupons.sav**。這是包含某連鎖雜貨店想要知道他們客戶購買習慣所收集之調查資料的假設資料檔。每一個客戶被追蹤了四週，每一個觀察值對應至一個不同的客戶一週，並記錄有關客戶在何處及如何購物的資訊，包含那一週在雜貨店花了多少錢。
- **guttman.sav**。Bell (Bell, 1961) 以此表說明可能的社會團體。Guttman (Guttman 值, 1968) 過去曾使用此表的一部分，在這部分中有 5 個變數，分別說明 7 個理論社會團體的社會互動、團體歸屬感、成員實際接觸和關係正式性，而這 7 個群組包括：群眾（例如，足球場上的人）、觀眾（例如在戲院中和課堂上的人）、公眾（例如，報紙讀者和電視觀眾）、暴民（和群眾相似，但互動較為激烈）、原級團體（親密性）、次級團體（自願性）和現代社群（因親密的身體接近而導致鬆散的結盟和特殊服務的需求）。
- **health\_funding.sav**。這是包含醫療保健基金（每 100 個人口的金額）、疾病率（每 10,000 個人口的比率）、造訪醫療保健機構的比例（每 10,000 個人口的比率）的假設資料檔。每一個觀察值代表一個不同的城市。
- **hivassay.sav**。這是有關一家製藥實驗室致力於開發一種偵測 HIV 感染快速檢驗的假設資料檔。檢驗結果是八個紅色加深的陰影，陰影愈深表示感染的可能性愈大。進行了一項實驗室的試驗，在 2,000 個血液樣本中，有半數遭到 HIV 的感染，而半數則未感染。
- **hourlywagedata.sav**。這是有關在辦公室和醫院任職的護士依經驗水準不同之鐘點費的假設資料檔。
- **insurance\_claims.sav**。這是有關一家保險公司想要建立模式來標示可疑及可能的詐欺理賠之假設資料檔。每一個觀察值代表個不同的理賠。
- **insure.sav**。這是有關一家保險公司正在研究表示客戶是否必定理賠 10 年壽險合約之風險因子的假設資料檔。在資料檔中的每一個觀察值代表二份合約，其一記錄了理賠而另一則否，二者的年齡和性別相符。
- **judges.sav**。這是有關受過訓練的裁判（加上一位熱心人士）為 300 個體操表演評分的假設資料檔。每一列代表一個不同的表演；裁判們觀看相同的表演。
- **kinship\_dat.sav**。Rosenberg 與 Kim (Rosenberg 和 Kim, 1975) 致力於分析 15 個親屬關係稱呼（姑/姨、兄弟、堂/表兄弟姐妹、女兒、父親、孫女、祖父、祖母、孫子、母親、姪子/外甥、姪女/外甥女、姐妹、兒子、叔/舅父）。他們請四組大學生（兩組女性、兩組男性）根據其相似性來分類整理這些稱謂。他們請其中兩組（一組女性、一組男性）作兩次分類整理，第二次要根據與第一次不同的準則進行分類整理。因此，總共得到六個「來源」。每一個來源對應至一個  $15 \times 15$  的相似性矩陣，其儲存格等於來源中人數減去物件在該來源中分為同組的次數。
- **kinship\_ini.sav**。本資料檔包含 kinship\_dat.sav 之三維解的起始組態。
- **kinship\_var.sav**。本資料檔包含自變數「性別」、「世代」、和可用來解讀 kinship\_dat.sav 解答維度的（分離）「度」。尤其，它們可用來將解答空間限制為這些變數的線性組合。
- **marketvalues.sav**。本資料檔有關於一項在伊立諾州阿爾岡京 (Algonquin, Ill.) 的新屋開發案自 1999 年至 2000 年之房屋銷售情況。這些銷售與公共記錄有關。
- **nhis2000\_subset.sav**。「國民健康訪問調查 (NHIS)」為美國民間人口的一大型民眾調查。其以具全國代表性的家庭為樣本，面對面的完成訪問。而取得各家庭中成員的人口統計學資訊及健康行為、健康狀態方面等觀察報告。本資料檔包含一個 2000 年調查資訊的子集。國家衛生統

計中心。2000 年「國民健康訪問調查 (NHIS)」。公用資料檔案和文件。  
[ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NHIS/2000/](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/)。2003 年曾存取。

- **ozone.sav**。本資料包含對六個氣象變數所作的 330 個觀察值，以自其餘的變數中預測臭氧濃度。先前研究人員中，(Breiman 和 Friedman 檢定(F), 1985)、(Hastie 和 Tibshirani, 1990) 在這些會阻礙標準迴歸方式的變數中發現非線性。
- **pain\_medication.sav**。本假設資料檔包含治療慢性關節炎疼痛之消炎藥物臨床試驗的結果。特別關注於藥物發生作用的時間以及它是如何與現用藥物作比較。
- **patient\_los.sav**。本假設資料檔包含對因可能為心肌梗塞 (MI, 或「心臟病」) 入院病患的治療記錄。每一個觀察值對應至一個不同的病患並記錄許多與其留院期間有關的變數。
- **patlos\_sample.sav**。本假設資料檔包含病患在為心肌梗塞 (MI, 或「心臟病」) 治療期間接受血栓溶解治療的治療記錄樣本。每一個觀察值對應至一個不同的病患並記錄許多與其留院期間有關的變數。
- **poll\_cs.sav**。這是有關民意測驗專家致力於確定交付立法之前公眾對法案支持水準的假設資料檔。觀察值對應至登記選民。每一個觀察值記錄下選民的郡、鎮、和他居住的鄰近範圍。
- **poll\_cs\_sample.sav**。本假設資料檔包含列於 poll\_cs.sav 中的選民樣本。樣本是根據在 poll\_csplan 計劃檔中指定的設計來取得，而本資料檔記錄了包含機率和樣本權重。不過，請注意，由於取樣計劃採用到機率 - 比例 - 大小 (PPS) 方法，也用到一個包含聯合選擇機率的檔案 (poll\_jointprob.sav)。其他與選民人口及其對提議法案之意見有關的變數都在取樣後收集並加入資料檔中。
- **property\_assess.sav**。這是有關郡財產估價人員致力於對限定資源保持財產價值評估維持最新的假設資料檔。觀察值對應至郡內過去一年銷售的財產。資料檔中的每一個觀察值記錄了財產所在的鎮、上次訪查該財產的估價人員、自那次評估後經過的時間、當時定的估價、和該財產銷售價值。
- **property\_assess\_cs.sav**。這是有關州財產估價人員致力於對限定資源保持財產價值評估維持最新的假設資料檔。觀察值對應至州中的財產。資料檔中的每一個觀察值記錄了郡、鎮、和財產所在的鄰近範圍、自最後一次評估後經過的時間、和當時定的估價。
- **property\_assess\_cs\_sample.sav**。本假設資料檔包含列於 property\_assess\_cs.sav 中的財產樣本。樣本是根據在 property\_assess\_csplan 計劃檔中指定的設計來取得，而本資料檔記錄了包含機率和樣本權重。另外的變數「目前價值」是在取樣後收集並加入資料檔中。
- **recidivism.sav**。這是有關政府法令執行機構致力於瞭解其轄區內之再犯率的假設資料檔。每一個觀察值對應至一個先前的違法者並記錄其人口資訊、第一次犯罪的一些細節、然後是直到第二次被捕的時間 (如果它發生在第一次被捕的兩年之內)。
- **recidivism\_cs\_sample.sav**。這是有關政府法令執行機構致力於瞭解其轄區內之再犯率的假設資料檔。每一個觀察值對應到一個先前的違法者，在 2003 年六月第一次被捕後釋放，並記錄其人口資訊、第一次犯罪的一些細節、和第二次被捕日期 (如果它發生在 2006 年六月之前)。違法者是根據在 recidivism\_cs\_csplan 中所指定的取樣計劃之樣本部門來選取；由於取樣計劃採用到機率 - 比例 - 大小 (PPS) 方法，也用到一個包含聯合選擇機率的檔案 (recidivism\_cs\_jointprob.sav)。



- **rfm\_transactions.sav**。本檔案是包含購買交易資料的假設資料檔，包括購買日期、購買項目及每一項交易的金額。
- **salesperformance.sav**。這是有關評估兩個新售貨員訓練課程的假設資料檔。六十個員工，分成三個組別，全部接受標準訓練。此外，組別二得到技術訓練；組別三則是實務輔導簡介。每一個員工在訓練課程結束時接受測驗並記錄他們的分數。在資料檔中每一個觀察值代表一個不同的訓員，並記錄他們所分派的組別和他們在測驗中得到的分數。
- **satisf.sav**。這是有關一家零售公司在 4 個商店位置所作之滿意度調查的假設資料檔。總共有 582 位客戶接受調查，每一個觀察值代表一位客戶的反應。
- **screws.sav**。這個資料檔包含螺絲釘、螺栓、螺帽和圖釘之特色的資訊 (Hartigan, 1975)。
- **shampoo\_ph.sav**。這是有關一家美髮產品工廠品質管制過程的假設資料檔。在固定的時間間隔，記錄下六個不同輸出批次的測量和它們的 pH 值。目標範圍是 4.5 - 5.5。
- **ships.sav**。一個在別處 (McCullagh et al., 1989) 出現和分析過，有關商船因風浪所造成損壞的資料集。事件次數可建立模式為以 Poisson 率發生，給定船型、建造期間、和服務期間。以因子交叉分類所形成的表格的每一個儲存格服務月數的整合，提供了暴露於風險之值。
- **site.sav**。這是有關一家公司致力於為事業擴展選擇新地點的假設資料檔。他們僱請兩位顧問分別評估該地點，除了一份廣泛的報告之外，他們還要將每個地點摘要為前景「佳」、「可」、或「差」。
- **smokers.sav**。本資料檔是由「1998 年全國家庭毒品濫用調查」中摘錄，且是美國家庭的機率樣本。(<http://dx.doi.org/10.3886/ICPSR02934>) 因此，在分析本資料檔的第一步應該是將資料加權以反映母群體傾向。
- **stocks.sav** 本假設資料檔包含一年的股票價格和數量。
- **stroke\_clean.sav**。本假設資料檔包含一個醫療資料庫，其在以「資料準備」選項中的程序清理之後的狀態。
- **stroke\_invalid.sav**。本假設資料檔包含一個醫療資料庫的起始狀態並包含幾個資料輸入錯誤。
- **stroke\_survival**。本假設資料檔是有關缺血性中風的病患，其在結束康復計畫後存活時間方面，面臨許多挑戰。中風後，記載了心肌梗塞、缺血性中風、或出血性中風的發生，以及事件記錄的時間。由於它只包含在康復計劃所管制的中風存活的病患，此樣本的左側被截斷。
- **stroke\_valid.sav**。本假設資料檔包含一個醫療資料庫，在其值以「驗證資料」程序檢查之後的狀態。它仍包含可能的異常觀察值。
- **survey\_sample.sav**。本資料檔包含調查資料，包括人口資料和各種態度測量。雖然已修改一些資料數值，且為人口資料之目的新增了一些額外的虛構變數，但是資料仍是以「1998 NORC 基本社會調查」的變數子集為基礎。
- **telco.sav**。這是有關一家電信公司致力於在客戶庫中減少顧客不忠的假設資料檔。每一個觀察值對應至一位不同的客戶並記錄不同的人口資料和服務使用方式資訊。
- **telco\_extra.sav**。本資料檔類似於 telco.sav 資料檔，但「任期」的對數轉換客戶花費變數已予刪除，並更換為標準的對數轉換客戶花費變數。
- **telco\_missing.sav**。本資料檔是 telco.sav 資料檔的子集，不過某些人口資料值已更換為遺漏值。

- **testmarket.sav**。本假設資料檔有關於一家速食連鎖店計劃在菜單中加入新的項目。有三個可能的活動來促銷此新產品，所以該新項目在幾個隨機選取市場中的地點作介紹。在每一個地點使用不同的促銷，並記錄該新項目前四週的每週銷售量。每一個觀察值對應至一個不同的地點-週。
- **testmarket\_1month.sav**。本假設資料檔是將 testmarket.sav 資料檔和每週購買的「彙總」，因此每一個觀察值對應至一個不同的客戶。結果部份每週變更的變數消失了，而目前所記錄的銷售量是在研究的四週期間銷售量之總和。
- **tree\_car.sav**。這是包含人口資料和車輛購買價格資料的假設資料檔。
- **tree\_credit.sav**。這是包含人口資料和銀行放款歷史資料的假設資料檔。
- **tree\_missing\_data.sav** 這是包含有大量遺漏值的人口資料和銀行放款歷史資料的假設資料檔。
- **tree\_score\_car.sav**。這是包含人口資料和車輛購買價格資料的假設資料檔。
- **tree\_textdata.sav**。一個只有兩個變數的簡單資料檔，主要目的在顯示變數預設狀態（在指定量測水準和數值標記之前）。
- **tv-survey.sav**。這是有關一家電視製片廠考量是否要延長一個成功節目的播送所作之調查的假設資料檔。有 906 位應答者被問到在不同的狀況下他們是否願意觀看這個節目。每一列代表一個不同的應答者；每一行為一個不同的狀況。
- **ulcer\_recurrence.sav**。本檔案包含一項用來比較兩種防止潰瘍復發治療法功效之研究的部分資訊。它是很好的區間受限資料範例，且已在別處 (Collett, 2003) 出現和分析過。
- **ulcer\_recurrence\_recoded.sav**。本檔案是將 ulcer\_recurrence.sav 的資訊重新組織，以讓您為此研究的每一個區間事件機率而非只是研究目的事件機率建立模式。它已在別處 (Collett et al., 2003) 出現和分析過。
- **verd1985.sav**。本資料檔有關於一項調查 (Verdegaal, 1985)。在調查中記錄了來自 15 個受訪者對 8 個變數的回應。所需的變數被分成三組。集 1 包括 age 和 marital，集 2 包括 pet 和 news，集 3 包括 music 和 live。Pet 調整為多重名義量數，age 調整為次序量數，其他的變數調整為單一名義量數。
- **virus.sav**。這是有關一家網際網路服務提供者致力於在其網路上判斷病毒之影響的假設資料檔。他們在其網路上追蹤從發現病毒直到控制威脅的這段時間，被病毒感染之電子郵件的流量（約略）百分比。
- **wheeze\_steubenville.sav**。這是空氣污染對兒童健康之影響 (Ware, Dockery, Spiro III, Speizer, 和 Ferris Jr., 1984) 縱向研究的子集。本資料包含來自俄亥俄州 Steubenville, 年齡 7、8、9 和 10 歲兒童的氣喘聲狀態之重複二元測量，以及其母親在本研究的第一年是否抽煙的固定記錄。
- **workprog.sav**。這是有關一項政府職業計劃，設法將弱勢民眾安置到較好之工作的假設資料檔。一個樣本的可能計劃參與者被追蹤，他們之中某些被選取加入本計劃，而其他的則否。每一個觀察值代表一位不同的計劃參與者。
- **worldsales.sav** 本假設資料檔包含依洲和產品分類之銷貨收益。

# 注意事項

本資訊適用於全球提供之產品與服務。

IBM 可能並未在其他國家提供在本文件中討論到的產品、服務或功能。有關目前在貴地區可供使用的產品與服務相關資訊，請洽您當地的 IBM 服務代表。對於 IBM 產品、程式或服務的任何參考，目的並不是要陳述或暗示只能使用 IBM 產品、程式或服務。任何功能相等且未侵犯 IBM 智慧財產權的產品、程式或服務皆可使用。但是，評估及確認任何非 IBM 產品、程式或服務的操作之責任應由使用者承擔。

IBM 可能有一些擁有專利或專利申請中的項目包含本文件所描述的內容。本文件的提供並不表示授與您對於這些專利的權利。您可以將書面的授權查詢寄至：

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U. S. A.

對於與雙位元組字元集 (DBCS) 資訊相關的授權查詢，請與貴國的 IBM 智慧財產部門聯絡，或將查詢郵寄至：

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

**下列條款不適用於英國，或其他任何當地法律規定與下列條款不一致的國家：**

INTERNATIONAL BUSINESS MACHINES 聲明係以「現狀」提供，沒有任何保固；不作任何明示或默示的保證，包括但不限於不侵權、適銷性或適合某一特定用途之保證。某些州不允許特定交易中明示或默示的保固聲明，因此，此聲明或許對您不適用。

此資訊內容可能包含技術失準或排版印刷錯誤。此處資訊會定期變更，這些變更將會納入新版的聲明中。IBM 可能會隨時改善和 / 或變更此聲明中所述的產品和 / 或程式，恕不另行通知。

此資訊中對於非 IBM 網站之任何參考僅為查閱方便而設，而且在任何情況中均不得作為那些網站之背書。該「網站」的內容並非此 IBM 產品的部分內容，使用該「網站」需自行承擔風險。

IBM 可能會以任何其認為適當的方式使用或散佈您提供的任何資訊，無需對您負責。

意欲針對達成以下目的而擁有本程式相關資訊之程式被授權人：(i) 在獨立建立的程式與其他程式（包括本程式）之間交換資訊及 (ii) 共用已交換的資訊，應聯絡：

IBM Software Group, 收件人：Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA。

在適當條款與條件之下，包括某些情況下（支付費用），或可使用此類資訊。

在本文件中描述的授權程式及其適用之所有授權材料皆由 IBM 在與我方簽訂之 IBM 客戶合約、IBM 國際程式授權合約或任何相等效力合約中提供。

有關非 IBM 產品的資訊來自於那些產品的供應商、其公佈內容或其他可公開取得的來源。IBM 尚未測試那些產品，且無法確認效能準確度、相容性或任何其他與非 IBM 產品相關的理賠。對於非 IBM 產品之功能有任何問題，應由那些產品之供應商加以解決。

此資訊包含用於日常企業運作的資料和報表範例。為了儘可能提供完整說明，範例中包含了人名、公司名稱、品牌名稱和產品名稱。所有的名稱皆為虛構，使用的名稱或地址和實際的企業如有雷同純屬巧合。

如果您正在螢幕上檢視此資訊，則圖片和彩色說明可能不會顯示。

## 商標

IBM、IBM 標誌、ibm.com 和 SPSS 為 IBM Corporation 之註冊商標，已經於世界各地許多法律管轄區域註冊。IBM 註冊商標的清單目前可於「網站」上取得，網址為：<http://www.ibm.com/legal/copytrade.shtml>。

Adobe、Adobe 標誌、PostScript 以及 PostScript 標誌為 Adobe Systems Incorporated 於美國和 / 或其他國家的註冊商標或商標。

Intel、Intel 標誌、Intel Inside、Intel Inside 標誌、Intel Centrino、Intel Centrino 標誌、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 為 Intel Corporation 或其分公司於美國和其他國家的商標或註冊商標。

Java 和所有 Java 商標和標誌皆為 Sun Microsystems, Inc. 於美國和 / 或其他國家的商標。

Linux 為 Linus Torvalds 於美國和 / 或其他國家的註冊商標。

Microsoft、Windows、Windows NT 和 Windows 標誌為 Microsoft Corporation 於美國和 / 或其他國家的商標。

UNIX 為 The Open Group 於美國和其他國家的註冊商標。

本產品使用 WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>。

其他產品和服務名稱可能為 IBM 或其他公司的商標。

Adobe 產品的擷取畫面已取得 Adobe Systems Incorporated 之翻印許可。

Microsoft 產品的擷取畫面已取得 Microsoft Corporation 之翻印許可。



---

# 參考書目

- Bell, E. H. 1961. Social foundations of human behavior: (人類行為之社會基礎) Introduction to the study of sociology (社會學研究簡介). 紐約: Harper & Row.
- Blake, C. L., 和 C. J. Merz. 1998. "UCI Repository of machine learning databases (機器學習資料庫 UCI 儲存器)." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Breiman, L., 和 J. H. Friedman 檢定(F). 1985. Estimating optimal transformations for multiple regression and correlation (估計多重迴歸與相關之最適轉換). Journal of the American Statistical Association (美國統計協會彙報), 80, .
- Collett, D. 2003. Modelling survival data in medical research (模式化醫學研究中的存活資料), 2 ed. Boca Raton: Chapman & Hall/CRC.
- Davison, A. C., 和 D. V. Hinkley. 2006. Bootstrap Methods and their Application (自助法方法和其應用程式). : Cambridge University Press (劍橋大學出版部).
- Green, P. E., 和 V. Rao. 1972. Applied multidimensional scaling (應用多元尺度方法). Hinsdale, Ill.: Dryden Press (Dryden 出版社).
- Green, P. E., 和 Y. Wind. 1973. Multiattribute decisions in marketing: (市場行銷之多重屬性決策) A measurement approach (測量方法). Hinsdale, Ill.: Dryden Press (Dryden 出版社).
- Guttman 值, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points (尋找組態點最小座標空間之一般非計量技巧). Psychometrika (心理學計量報導), 33, .
- Hartigan, J. A. 1975. Clustering algorithms (集群演算法). 紐約: John Wiley and Sons.
- Hastie, T., 和 R. Tibshirani. 1990. Generalized additive models (概化附加模式). 倫敦: Chapman and Hall.
- Kennedy, R., C. Riquier, 和 B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research (市場研究類別資料之對應分析實際應用). Journal of Targeting, Measurement, and Analysis for Marketing (市場行銷之目標訂定、測量與分析雜誌), 5, .
- McCullagh, P., 和 J. A. Nelder. 1989. 概化線性模式, 第二版 ed. 倫敦: Chapman & Hall.
- Price, R. H., 和 D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior (行為適切性與情境限制作為社會行為維度). Journal of Personality and Social Psychology (人格與社會心理學雜誌), 30, .
- Rickman, R., N. Mitchell, J. Dingman, 和 J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet (實行 Stillman 飲食法期間血膽固醇改變情形). Journal of the American Medical Association (美國醫學協會彙報), 228, .
- Rosenberg, S., 和 M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research (多變量研究中作為資料收集程序之排序方法). Multivariate Behavioral Research (多變量行為研究), 10, .
- Shao, J., 和 D. Tu. 1995. The Jackknife and Bootstrap (Jackknife 和自助法). 紐約: Springer.

---

參考書目

Van der Ham, T., J. J. Meulman, D. C. Van Strien, 和 H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: (依經驗法將青少年飲食異常次組別化) A longitudinal perspective (縱向觀點). *British Journal of Psychiatry* (英國心理學雜誌), 170, .

Verdegaal, R. 1985. Meer sets analyse voor kwalitatieve gegevens (in Dutch) (更多性質資料的集合分析 (荷蘭文)). Leiden (萊頓): Department of Data Theory, University of Leiden (萊頓大學資料理論系).

Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, 和 B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities (六個城市中的兒童二手煙、氣體煮食與呼吸道健康). *American Review of Respiratory Diseases* (美國呼吸道疾病評論), 129, .

---

# 索引

中位數信賴區間  
在自助法中, 17

參數估計值  
在自助法中, 26

商標, 36

比例的信賴區間  
在自助法中, 13 - 14

法律注意事項, 35

範例檔案  
位置, 28

自助法, 3, 9  
中位數信賴區間, 17  
參數估計值, 26  
支援的程序, 5  
比例的信賴區間, 13 - 14  
自助法規格, 13  
自助法規格  
在自助法中, 13