

IBM SPSS Bootstrapping 22

IBM

附註

使用本資訊及其支援產品之前，請先閱讀第 7 頁的『注意事項』中的資訊。

產品資訊

此版本適用於 IBM® SPSS® Statistics 22.0.0 版及所有後續版本與修訂版，除非新版本中另有指示。

目錄

第 1 章 重複取樣簡介	1	注意事項.	7
第 2 章 重複取樣	3	商標	8
支援重複取樣的程序	4	索引	9
BOOTSTRAP 指令的其他功能.	6		

第 1 章 重複取樣簡介

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，但這些估計值有多可靠？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。重複取樣會試圖尋找「未知」母群和不規律參數之性質的更多資訊。

重複取樣如何運作

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

此外，重複取樣可提供沒有參數估計值之中位數的標準誤與信賴區間。

產品中的重複取樣支援

在支援重複取樣的程序中，重複取樣是以子對話框的形式體現。如需支援引導之程序的相關資訊，請參閱第 4 頁的『支援重複取樣的程序』。

在對話框中要求重複取樣時，除了對話框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重複取樣樣本。產品於內部會將這些重複取樣樣本視為分割，即使它們並未明確顯示於「資料編輯器」中也一樣。這就表示，內部有 $B*N$ 個有效的觀察值，因此在重複取樣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「重複取樣分割」執行分析的結果。這些重複取樣結果合併後會連同程序產生的其餘一般輸出一併顯示於「檢視器」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。

第 2 章 重複取樣

重複取樣方法可獲得穩健性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、勝算比、相關係數或迴歸係數。重複取樣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型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) 重複取樣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流失約 27% 的客戶。為適當地將焦點放在減少客戶流失的努力成果上，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。透過使用重複取樣，您可以判斷單一客戶流失比率是否能適當描述四種主要客戶類型。

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。

管理部門也有興趣知道哪些因素與員工加薪有關聯，可以透過將線性模型套用到目前薪資與起薪間差異來觀察。當您對線性模型執行重複取樣時，可以使用特殊重新抽樣方法 (殘差與離群重複取樣) 來取得更精確的結果。

許多程序都支援對從重複取樣樣本分析的結果進行重複取樣抽樣與合併。可指定重複取樣分析的控制項，已直接整合成爲支援重複取樣之程序中的一般對話框。重複取樣對話框中的設定會存留在整個階段作業中，所以如果您透過對話框的重複取樣執行次數分配分析，預設會爲支援重複取樣的其他程序開啓重複取樣。

取得重複取樣分析

1. 從功能表中選擇支援重複取樣的程序，並按一下「**重複取樣**」。
2. 選取「**執行重複取樣**」。

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 個重複取樣樣本。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 設爲作用中產生器，並在「亂數產生器」對話框上指定固定的起點，但重要的差異在於在此對話框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 但小於 100 的信任層次。百分位數間隔只使用對應至信任間隔百分位數的依序引導值。例如，95% 百分數信賴區間使用重複取樣值的第 2.5 個與第 97.5 的百分位數作爲區間的上界與下界 (會視需要內插數值)。已修正偏差與加速 (BCa) 的區間爲已調整的區間，因爲更爲精確，所以也需要更多時間來計算。

抽樣。簡式方法會從原始資料集中重複取樣觀察值並放回。**階層化**方法會從原始資料集中不斷抽樣觀察值並放回，此動作是在由分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，階層化重複取樣抽樣會十分有用。

支援重複取樣的程序

下列程序支援重複取樣。

附註：

- 重複取樣無法使用多重插補的資料集。如果資料集中有 *Imputation_* 變數，則「重複取樣」對話框會停用。
- 如果使用非整數加權值的話，將無法使用重複取樣。
- 重複取樣使用完全刪除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，所以當重複取樣生效時，完全刪除遺漏值也會生效，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。支援下列特性：

- 「統計量」表格支援平均數、標準差、變異數、中位數、偏斜度、峰度與百分位數的重複取樣估計。
- 「次數分配」表格支援百分比的重複取樣估計。

描述性。支援下列特性：

- 「描述性統計量」表格支援平均數、標準差、變異數、偏斜度與峰度的重複取樣估計。

探索。支援下列特性：

- 「描述性統計量」表格支援平均數、5% 修整平均數、標準差、變異數、中位數、偏斜度、峰度與四分位距的重複取樣估計。
- 「M 估計值」表格支援下列的重複取樣估計：Huber M 估計值、Tukey's 二權數、Hampel M 式估計值，與 Andrew's Wave。
- 「百分位數」表格支援百分比的重複取樣估計。

交叉表。支援下列特性：

- 「指向性測量」表格支援下列的重複取樣估計：Lambda (λ)、Goodman、Kruskal Tau、不確定性係數與 Somers' d。
- 「對稱性量數」表格支援下列的重複取樣估計：Phi (ϕ) 值、克瑞瑪 V (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」表格支援勝算比的重複取樣估計。
- 「Mantel-Haenszel 共同勝算比」表格支援重複取樣估計與 $\ln(\text{Estimate})$ 的顯著性檢定。

平均值。支援下列特性：

- 「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、偏斜度、調和平均數與幾何平均數的重複取樣估計。

單一樣本 T 檢定。支援下列特性：

- 「統計量」表格支援平均數與標準差的重複取樣估計。
- 「檢定」表格支援平均數差異的重複取樣估計與顯著性檢定。

獨立樣本 T 檢定。支援下列特性：

- 「組別統計量」表格支援平均數與標準差的重複取樣估計。
- 「檢定」表格支援平均數差異的重複取樣估計與顯著性檢定。

成對樣本 T 檢定。支援下列特性：

- 「統計量」表格支援平均數與標準差的重複取樣估計。
- 「相關性」表格支援相關的重複取樣估計。
- 「檢定」表格支援平均數的重複取樣估計。

單向變異數分析。支援下列特性：

- 「描述性統計量」表格支援平均數與標準差的重複取樣估計。
- 「多重比較」表格支援平均數差異的重複取樣估計。
- 「對比檢定」表格支援對比值的重複取樣估計與顯著性檢定。

GLM 單變數。支援下列特性：

- 「描述性統計量」表格支援平均數與標準差的重複取樣估計。
- 「參數估計值」表格支援係數 B 的重複取樣估計與顯著性檢定。
- 「對比結果」表格支援差異的重複取樣估計與顯著性檢定。
- 「預估邊際平均值」：「預估」表格支援引導預估平均值。
- 「預估邊際平均值」：「成對比較」表格支援引導預估平均值差異。
- 「事後檢定」：「多重比較」表格支援引導預估「平均值差異」。

雙變數相關性。支援下列特性：

- 「描述性統計量」表格支援平均數與標準差的重複取樣估計。
- 「相關性」表格支援相關的重複取樣估計及顯著性檢定。

附註：

如果除了皮爾森 (Pearson) 相關外，還要求無母數相關性 (Kendall 的 tau-b 或 Spearman)，則對話框會貼上 CORRELATIONS 與 NONPAR CORR 指令，每個指令包含個別的 BOOTSTRAP 指令。相同的重複取樣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。

局部相關性。支援下列特性：

- 「描述性統計量」表格支援平均數與標準差的重複取樣估計。
- 「相關性」表格支援相關的重複取樣估計。

線性回歸。支援下列特性：

- 「描述性統計量」表格支援平均數與標準差的重複取樣估計。
- 「相關性」表格支援相關的重複取樣估計。
- 「模型摘要」表格支援 Durbin-Watson 的重複取樣估計。
- 「係數」表格支援係數 B 的重複取樣估計與顯著性檢定。
- 「相關係數」表格支援相關的重複取樣估計。
- 「殘差統計量」表格支援平均數與標準差的重複取樣估計。

序數回歸。支援下列特性：

- 「參數估計值」表格支援係數 B 的重複取樣估計與顯著性檢定。

區別分析。支援下列特性：

- 「標準典型區別函數係數」表格支援標準化係數的重複取樣估計。
- 「典型區別函數係數」表格支援未標準化係數的重複取樣估計。
- 「分類函數係數」表格支援係數的重複取樣估計。

進階統計量選項

GLM 多變數。支援下列特性：

- 「參數估計值」表格支援係數 **B** 的重複取樣估計與顯著性檢定。

線性混合模型。支援下列特性：

- 「固定效果估計」表格支援估計值的重複取樣估計與顯著性檢定。
- 「估計共變異數參數」表格支援估計值的重複取樣估計與顯著性檢定。

一般性線性模型。支援下列特性：

- 「參數估計值」表格支援係數 **B** 的重複取樣估計與顯著性檢定。

Cox 回歸。支援下列特性：

- 「在方程式中的變數」表格支援係數 **B** 的重複取樣估計與顯著性檢定。

迴歸選項

二元羅吉斯回歸。支援下列特性：

- 「在方程式中的變數」表格支援係數 **B** 的重複取樣估計與顯著性檢定。

多項式羅吉斯迴歸。支援下列特性：

- 「參數估計值」表格支援係數 **B** 的重複取樣估計與顯著性檢定。

BOOTSTRAP 指令的其他功能

指令語法語言也可以讓您：

- 執行殘差與離群重複取樣抽樣 (SAMPLING 次指令)

如需完整的語法資訊，請參閱《指令語法參考手冊》。

注意事項

本資訊適用於美國提供的產品與服務。

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 (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 CORPORATION 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。

本資訊中任何對非 IBM 網站的敘述僅供參考，IBM 對該網站並不提供任何保證。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
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

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等效合約中提供。

此處所包含的任何效能資料皆於受控制的環境中取得。因此，其他作業環境中所取得的結果可能有顯著的不同。某些測量可能是於開發階段的系統中執行，因此無法保證在一般所用的系統中進行測量結果會相同。此外，某些測量可能已經透過外插法進行預測。實際測量結果可能不同。此文件的使用者應確認資料是否適用於特定環境中。

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。有關非 IBM 產品的性能問題應直接洽詢該產品供應商。

關於 IBM 未來方針或意圖的所有聲明僅代表目標或目的，得依規定未另行通知即變更或撤銷。

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。

這些範例程式或任何衍生成果的每份複本或任何部分，都必須依照下列方式併入著作權聲明：

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。

這些範例程式或任何衍生成果的每份複本或任何部分，都必須依照下列方式併入著作權聲明：

© (貴客戶公司名稱) (年度) 。本程式之若干部分係衍生自 IBM 公司的範例程式。

© Copyright IBM Corp. _輸入一個以上年度_。All rights reserved.

商標

IBM、IBM 標誌及 ibm.com 是 International Business Machines Corp. 在世界許多管轄區註冊的商標或註冊商標。其他產品及服務名稱可能是 IBM 或其他公司的商標。IBM 商標的最新清單可在 Web 的『Copyright and trademark information』中找到，網址為 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。

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。

Java 和所有以 Java 為基礎的商標及標誌是 Oracle 及 (或) 其子公司的商標或註冊商標。

索引

索引順序以中文字，英文字，及特殊符號之次序排列。

〔九劃〕

重複取樣 3
支援的程序 4



Printed in Taiwan