

# IBM SPSS Missing Values 19



Note: Before using this information and the product it supports, read the general information under Notices 第 84 頁.

This document contains proprietary information of SPSS Inc, an IBM Company. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

**© Copyright SPSS Inc. 1989, 2010.**

---

# 序

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

## 關於 SPSS Inc.，是一家 IBM 公司

SPSS Inc.，是一家 IBM 公司，為全球領先的預測分析軟體和解決方案供應商。該公司完整的系列產品 — 資料收集、統計量、模型製造與部署 — 捕捉人們的態度和意見，預測客戶未來的互動結果，然後將分析融入業務程序，以依照所得見解採取行動。SPSS Inc. 解決方案藉由著重於收斂性分析、IT 架構和業務程序，以達成整個組織相互關聯的業務目標。全球商業、政府和學界客戶均仰賴 SPSS Inc. 技術為競爭優勢，以吸引、留住和增加客戶人數，同時減少欺詐並降低風險。SPSS Inc. 在 2009 年 10 月由 IBM 收購。如需詳細資訊，請造訪 <http://www.spss.com>。

## 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 SPSS Inc. 產品在使用上的協助，或是支援硬體環境的安裝說明。如果要聯絡技術支援，請參閱 SPSS Inc. 網站（網址是 <http://support.spss.com>），或是透過網站（網址是 <http://support.spss.com/default.asp?refpage=contactus.asp>）尋找當地的辦事處。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。

## 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，列示於網站上：<http://www.spss.com/worldwide>。請備妥您的序號以供識別。

## 訓練研討會

SPSS Inc. 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。如需有關這些研討會的更多資訊，請聯絡您的當地辦公室，列示於網站上：<http://www.spss.com/worldwide>。

## 其他出版品

SPSS Statistics: Guide to Data Analysis (資料分析指南)、SPSS Statistics: Statistical Procedures Companion (統計程序指南) 以及 SPSS Statistics: Advanced Statistical Procedures Companion (進階統計程序指南) 是由 Marija Norušis 撰寫，

由 Prentice Hall 發行，為推薦的輔助資料。這些出版品涵蓋 SPSS Statistics Base 模組、進階統計量模組和迴歸模組中的統計程序。不論您是資料分析的新手，還是已經準備使用高階應用程式，這些書籍都能幫助您善加利用 IBM® SPSS® Statistics 系列產品中的功能。如需其他資訊（包括出版品內容和章節樣本），請參閱作者的網站：<http://www.norusis.com>

## 部 I：使用手冊

<b>1</b>	<b>遺漏值簡介</b>	<b>1</b>
<b>2</b>	<b>遺漏值分析</b>	<b>2</b>
	顯示遺漏值形式 . . . . .	4
	顯示遺漏值的敘述統計 . . . . .	6
	估計統計量和插補遺漏值 . . . . .	7
	EM 估計選項 . . . . .	8
	迴歸方法估計選項 . . . . .	9
	預測與已預測變數 . . . . .	10
	MVA 指令的其他功能 . . . . .	11
<b>3</b>	<b>多個插補</b>	<b>12</b>
	分析形式 . . . . .	13
	插補遺漏資料值 . . . . .	15
	方法 . . . . .	17
	限制 . . . . .	19
	輸出 . . . . .	21
	MULTIPLE IMPUTATION 指令的其他功能 . . . . .	22
	使用多重插補資料 . . . . .	22
	分析多重插補資料 . . . . .	25
	多重插補選項 . . . . .	29

## 部 II：範例

<b>4</b>	<b>遺漏值分析</b>	<b>32</b>
	說明遺漏資料的形式 . . . . .	32

執行分析以顯示敘述統計 . . . . .	32
評估敘述統計 . . . . .	33
重新執行分析以顯示形式 . . . . .	39
評估形式表 . . . . .	41
針對 Little 的 MCAR 檢定重新執行分析 . . . . .	42
<b>5 多重插補</b>	<b>43</b>
使用多重插補完成並分析資料集 . . . . .	43
分析遺漏值形式 . . . . .	43
遺漏值自動插補 . . . . .	47
自定插補模式 . . . . .	54
檢查 FCS 收斂 . . . . .	62
分析完整資料 . . . . .	65
摘要 . . . . .	75
<b>附錄</b>	
<b>A 範例檔案</b>	<b>76</b>
<b>B Notices</b>	<b>84</b>
<b>索引</b>	<b>87</b>

# 部 1: 使用手冊





# 遺漏值簡介

含有遺漏值的觀察值是一項重要挑戰，因為典型的模式化程序會捨棄這些觀察值而不進行分析。當有少數遺漏值（大約少於總觀察值數目的 5%）且這些值可視為隨機遺漏時（也就是說，某個值是否遺漏和其他值並無關聯），則典型的刪除全部遺漏值方法是相對「安全」的。「遺漏值」選項可協助您判斷刪除全部遺漏值是否足夠，並能在不足時提供處理遺漏值的方法。

## 「遺漏值」分析和「多重插補」程序的比較

「遺漏值」選項提供兩組處理遺漏值的程序：

- **多重插補** 程序提供遺漏值樣式的分析，適合最後的遺漏值多重插補。換言之，會產生多種版本的資料集，每個資料集各包含一組插補值。執行統計分析時，會合併所有插補資料集的參數估計值，提供一般而言比單一插補更精確的估計值。
- **遺漏值分析** 提供有些微差異的不同敘述工具組，可分析遺漏值（尤其是 Little' s MCAR 檢定），且包含許多單一插補方法。請注意，一般認為多重插補優於單一插補。

## 遺漏值工作

您可依循下列基本步驟以開始使用遺漏值分析：

- ▶ **檢查遺漏項目。** 使用「遺漏值分析」和「分析樣式」探索您資料中的遺漏值樣式，並判斷是否需要多重插補。
- ▶ **插補遺漏值。** 使用「插補遺漏值」多重插補遺漏值。
- ▶ **分析「完整」資料。** 使用支援多重插補資料的任意程序。如需分析多重插補資料集與支援這些資料的程序清單，請參閱 [分析多重插補資料](#) 第 25 頁。

# 遺漏值分析

您可以利用「遺漏值分析」程序，來執行下列三個主要功能：

- 說明遺漏資料的形式。遺漏值位在何處？範圍有多大？成對的變數是否易於在多個觀察值中出現遺漏值？資料值是否極端？遺漏值是否隨機？
- 不同遺漏值方法的估計平均數、標準差、共變量和相關性：完全排除、成對、迴歸或 EM (expectation-maximization)。其中，成對法會將成對完成觀察值的個數，加以顯示出來。
- 使用迴歸或 EM 方法，以估計值填入（插補）遺漏值；不過，一般認為多重插補可提供更精確的結果。

當資料不完整時，您就可以利用遺漏值分析，來提出因資料不完整，所造成的種種影響。如果有遺漏值的觀察值在系統上與沒有遺漏值的觀察值不同，則結果可能令人誤解。再者，如果遺漏資料，可能會因為資訊少於原先計畫，而降低統計量的精確度。另外還有一個考量就是，許多統計程序的假設基礎，都是建構在完整的觀察值上面的，因此，如果有遺漏值的話，可能就需要使用更複雜的理論。

**範例。**在評估血友病的治療結果時，我們會同時評量數個變數。不過，並非每位患者都可以使用每一種測量。所以，我們會顯示遺漏值的形式，製成一覽表，並且發現它們是隨機的。然後再使用 EM 分析功能，來估計平均數、相關和共變異數。它也用來確定資料是隨機完全遺漏。或者，也可以用填入值來取代遺漏值，並將之存入新資料檔中，以供進一步分析。

**統計量。**就單變量統計量而言，它包括：非遺漏值個數、平均數、標準差、遺漏值個數，以及極端值個數。若使用完全排除、成對、EM 或迴歸等方法，則可得到估計平均數、共變異數矩陣，以及相關矩陣。有 EM 結果的 Little's MCAR 檢定。不同方法的平均數總和。對於用「遺漏值對非遺漏值」所定義的組別而言：t 檢定。對所有的變數：遺漏值形式顯示出變數觀察值。

## 資料考量

**資料。**您所用的資料，可以是類別或數值（比例或連續）。不過，您可估計統計量並只為數值變數插補遺漏資料。但對每個變數而言，如果遺漏值沒有被編碼成系統遺漏值的話，則必須被定義為使用者遺漏值。舉例來說，若問卷項目中的 Don't know 反應值編碼為 5，而且您又想將其視為遺漏值，則此項目就應該將 5 編碼成使用者遺漏值。

**頻率加權。**此程序允許次數（重複）加權。具有負或零重複加權值的觀察值會被忽略。系統會截斷非整數加權。

**假設。**完全排除、配對和迴歸估計是根據遺漏值形式不依賴資料值的假設。(該條件就是「**隨機完全遺漏**」或 MCAR)。所以，當資料是 MCAR 時，所有用來估計的方法(包含 EM 方法)對相關值與共變異數都能得到一致而不偏的估計值。違反 MCAR 假設會導致由完全排除、配對和迴歸方法所產生的偏差估計值。如果資料不是 MCAR，您應使用 EM 估計值。

EM 估計是依據遺漏值形式只和觀察資料有關的假設。(這種狀況稱作「**隨機遺漏**」，或 MAR。)這個假設容許以可用的資訊來調整估計值。例如，在教育程度與收入的研究中，教育程度低的受試者可能有較多的遺漏收入值。在這種情況下，資料是 MAR，而非 MCAR。換句話說，收入會被記錄的機率取決於受試者的教育程度。這個機率會因教育程度而改變，但不會因「該教育程度」的收入而變。如果收入會被記錄的機率也隨著每一個教育程度的收入值而不同(例如，有高收入的人並未申明)，則資料既非 MCAR 亦非 MAR。這不是不常發生的狀況，而且，如果發生了，沒有適用的方法。

**相關程序。**有許多程序可讓您使用完全排除估計或成對估計。像「線性迴歸」和「因子分析」，就讓您用平均值來置換遺漏值。此外，「預測附加模組」也提供數種方式，讓您置換時間數列中的遺漏值。

## 取得遺漏值分析

- ▶ 從功能表選擇：  
分析(A) > 遺漏值分析...

圖表 2-1  
遺漏值分析對話方塊



- ▶ 至少選擇一個數值(尺度)變數以估計統計量和隨意選擇插補遺漏值。  
您可以：
  - 選取類別變數(數字或字串)，然後輸入類別個數的上限(最大類別)。

- 按一下「形式」將遺漏資料的形式表列。如需詳細資訊，請參閱第 4 頁中的[顯示遺漏值形式](#)。
- 按一下「描述」來顯示遺漏值的敘述統計。如需詳細資訊，請參閱第 6 頁中的[顯示遺漏值的敘述統計](#)。
- 選取一個方法來估計統計量（平均數、共變異數與相關）與可能的插補遺漏值。如需詳細資訊，請參閱第 7 頁中的[估計統計量和插補遺漏值](#)。
- 如果您選取「EM 方法」或「迴歸方法」，請按一下「變數」，來指定估計所使用的子集。如需詳細資訊，請參閱第 10 頁中的[預測與已預測變數](#)。
- 選取一個觀察值標記變數。此變數用來在顯示個別觀察值的形式表中標示觀察值。

## 顯示遺漏值形式

圖表 2-2  
遺漏值分析形式對話方塊



您可選擇顯示不同的表格以顯示遺漏資料的形式和範圍。這些表格可以幫助您識別：

- 遺漏值位在何處
- 在個別情況下，成對的變數是否易於出現遺漏值
- 資料值是否為極端量數值

### 顯示

有三種表格可以用來顯示遺漏觀察值的形式。

**表列觀察值。**在分析變數中的遺漏值形式是以每一個形式的顯示頻率來表列。使用「以遺漏值形式來將變數排序」來指定個數和變數是否以形式的相似度來排序。使用「略過含有小於 n % 觀察值的形式」來排除不常發生的形式。

**含遺漏值的觀察值。**每一個含有遺漏值或極端值的觀察值是以每一個分析變數來表列。使用「以遺漏值形式來將變數排序」來指定個數和變數是否以形式的相似度來排序。

**全部觀察值。**每一個觀察值都表列，並為每一個變數指示遺漏值和極端值。除非在「依…排序」中指定了變數，否則觀察值會根據它們在資料檔中的出現順序，予以列示。

在顯示個別觀察值的表格中，使用下列符號：

+	極高值
-	極低值
S	系統遺漏值
A	使用者遺漏值的第一種類型
B	使用者遺漏值的第二種類型
C	使用者遺漏值的第三種類型

## 變數

對已包括在分析中的變數，您可以顯示其他資訊。您加入「其他資訊」的變數都個別地顯示在遺漏形式表中。數值（尺度）變數會顯示平均數；類別變數會顯示在每一個類別中出現形式的觀察值數目。

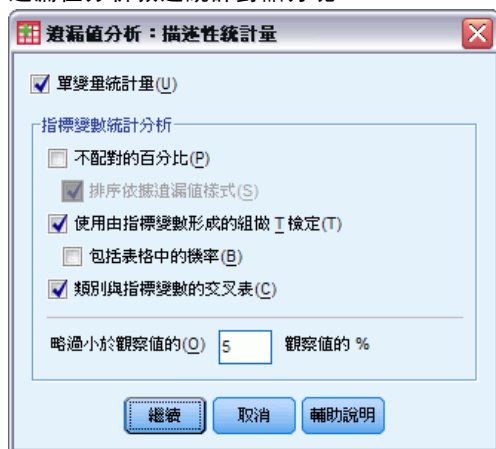
- **依?排序。**根據指定變數值的遞增或遞減順序，列出觀察值。這個選項只適用於「全部觀察值」。

## 顯示遺漏值形式

- ▶ 在主要的遺漏值分析對話方塊中，選擇您想要顯示遺漏值形式的變數。
- ▶ 按一下「形式」。
- ▶ 選取您想要顯示的形式表。

## 顯示遺漏值的敘述統計

圖表 2-3  
遺漏值分析敘述統計對話方塊



### 單變量統計量

單變量統計量有助於識別遺漏資料的一般範圍。每一個變數會顯示以下資料：

- 非遺漏值個數
- 遺漏值的數目和百分比

對於數值（尺度）變數，也會顯示下列：

- 平均數 (M)
- 標準差 (D)
- 極高值和極低值的數目

### 指標變數統計分析

為每一個變數建立一個指標變數。類別變數可表示該變數在個別觀察值中是存在或是遺漏。指標變數可以用來建立錯誤結合、t 檢定和次數分配表。

**不符合的百分比。**針對每對變數，顯示觀察值的百分比。在這些觀察值中，某個變數會含有遺漏值，而另一個變數則含有非遺漏值。在表格中，每個對角線項目，都包括單一變數的遺漏值百分比。

**使用由指標變數形成組別的 t 檢定。**您可以使用 Student's t 統計量，來比較每個數值變數中，兩個組別的平均數。組別說明變數是存在或是遺漏。此處所能顯示的統計量，包括 t 統計量、自由度、遺漏和非遺漏值個數，以及兩組別的平均數。您也可以顯示任何與 t 統計量關聯的雙尾機率。如果您的分析產生一個以上的檢定，不要拿這些機率來作顯著性檢定。這些機率僅適用於單一檢定計算。

**類別與指標變數的交叉表。**它會顯示每個類別變數的表格。對每個類別而言，此表格會顯示其他變數的非遺漏值次數及其百分比。此外，每種遺漏值的百分比，也都會被顯示出來。

**略過小於 n % 觀察值的變數遺漏。**若要縮小表格，且統計量僅計算幾個觀察值，則請加以刪除。

### 顯示敘述統計

- ▶ 在主要的遺漏值分析對話方塊中，選擇您想要顯示遺漏值敘述統計的變數。
- ▶ 按一下「描述性統計量」。
- ▶ 選擇您想要顯示的敘述性統計量。

## 估計統計量和插補遺漏值

您可以使用完全排除（只有完整的觀察值）、成對、EM (expectation-maximization) 及/或迴歸方法，來估計平均值、標準差、共變異數和相關。您也可以選擇插補遺漏值（估計置換值）。請注意解決遺漏值的問題時，一般認為**多重插補**優於單一插補。Little's MCAR 檢定對於判定是否需要插補而言，仍十分有用。

### 完全排除法

本方法只使用完整的觀察值。如果任何分析變數有遺漏值，便省略計算觀察值。

### 成對法

本方法查看成對的分析變數並只使用兩個變數都有非遺漏值的觀察值。分別為每一對變數計算次數分配值、平均數和標準差。由於在觀察值內的其他遺漏值已被忽略，兩個變數的相關和共變異數不相依於在任何其他變數所遺漏的值。

### EM 方法

本方法為部分遺漏資料假設一分配，並根據該分配之概似值進行推論。每一個疊代包含一個 E 步驟和一個 M 步驟。E 步驟可找出「遺漏」資料的條件式預期值、給定觀察值和目前參數估計值。這些預期值再用來取代「遺漏」的資料。在 M 步驟，參數的最大概似估計會當作遺漏資料已經填入的情況來計算。「遺漏」的資料會用引號括住，因為遺漏值不會直接填入。反之，會在對數概似使用其函數。

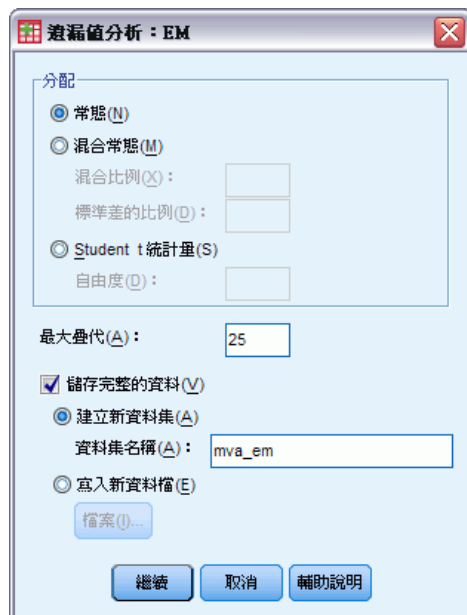
Roderick J. A. Little 用來檢定數值是否隨機完全遺漏 (MCAR) 的卡方統計量，會以 EM 矩陣註腳的方式印出。在這個檢定，虛無假設的資料是隨機完全遺漏，而 p 值在 0.05 水準時是顯著的。如果該值低於 0.05，資料便不是隨機完全遺漏。資料可能是隨機遺漏 (MAR) 或不是隨機遺漏 (NMAR)。您不可假設是兩者中的哪一個，而需分析該資料以確定資料是如何遺漏的。

### 迴歸方法

本方法計算多重線性迴歸估計並具有以隨機成份擴大估計值的選項。對每一個預測值，此程序可自隨機選取的完整觀察值、隨機常態偏差，或來自 t 分配的隨機偏差（以殘差的根平均平方來調整）來加入殘差。

## EM 估計選項

圖表 2-4  
遺漏值分析 EM 方法對話方塊



透過疊代處理，EM 方法估計平均數、共變異數矩陣，以及含有遺漏值之數值（尺度）變數的相關。

**分配。**EM 根據指定分配之概似值進行推論。根據預設值，假設為常態分配。如果您知道該分配的尾部較常態分配的尾部更長，您可要求程序以學生具  $n$  自由度的  $t$  分配，來建立概似函數。混合的常態分配也提供更長尾部的分配。指定混合的常態分配之標準差的比率及此二分配的混合比例。混合的常態分配假設分配只有標準差不同。平均數必須相同。

**最大疊代。**設定最大的疊代次數以估計真正的共變異數。當到達此疊代次數時，即使估計值尚未收斂，程序也會停止。

**儲存完成的資料。**您可將插補值替代遺漏值的資料集儲存起來。但是，請注意，共變異數為基礎的統計量使用插補值將會低估它們各自的參數值。低估的程度與共同未觀察的觀察值數量成比例。

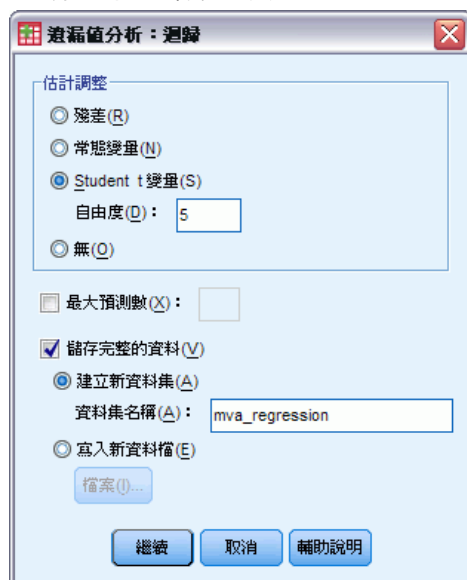
### 若要指定 EM 選項

- ▶ 在主要的遺漏值分析對話方塊中，選擇您想要採用 EM 方法來估計遺漏值的變數。
- ▶ 從「估計」組別中，選取「EM 方法」。
- ▶ 若要指定已預測變數和預測變數，請按一下「變數」。如需詳細資訊，請參閱第 10 頁中的[預測與已預測變數](#)。
- ▶ 按一下「EM 方法」。
- ▶ 選擇想要的 EM 選項。



## 迴歸方法估計選項

圖表 2-5  
遺漏值分析迴歸對話方塊



迴歸方法使用多重線性迴歸估計遺漏值。顯示平均數、共變數矩陣、以及預測變數的相關矩陣。

**估計調整。**您可以利用迴歸方法，將隨機成份加入迴歸估計值中。然後再選取殘差、常態變量、Student's t 變量，或者不予調整。

- **殘差。**從完整觀察值觀察殘差當中，隨機選擇錯誤項再加入迴歸估計值。
- **常態變量。**從含有期望值 0 的分配，以及等於迴歸的均方誤差項平方根的標準差當中，隨機取出誤差項。
- **Student's t 變量。**誤差項是以指定之自由度從 t 分配隨機抽出，並按均方誤差根 (RMSE) 縮放。

**最大預測數。**替估計程序所需使用之預測變數 (自變數) 個數，設定它的最大值。

**儲存完成的資料。**在目前作業階段中或外部 IBM® SPSS® Statistics 資料檔中寫入資料集，其遺漏值由迴歸方法估計的數值所取代。

### 指定迴歸方法選項

- ▶ 在主要的遺漏值分析對話方塊中，選擇您想要採用的迴歸方法來估計遺漏值的變數。
- ▶ 從「估計」組別中，選取「迴歸方法」。
- ▶ 若要指定已預測變數和預測變數，請按一下「變數」。如需詳細資訊，請參閱第 10 頁中的 [預測與已預測變數](#)。
- ▶ 按一下「迴歸方法」。
- ▶ 選擇想要的迴歸選項。

## 預測與已預測變數

圖表 2-6  
EM 方法和迴歸方法的遺漏值分析變數對話方塊



根據預設值，所有數值變數都是用來做 EM 和迴歸估計的。您可視需要選擇特定變數作為估計值的預測和已預測變數。給定的變數可在兩者的清單中，但有時您可能想要限制變數的運用。例如，某些分析師不喜歡對結果的變數值作估計。您可能也想要為不同的估計使用不同的變數，並多次執行此程序。例如，如果您有一組項目是護士評比，另一組是醫師評比，您可能想要使用護士項目執行一次以估計遺漏的護士項目，另外再為醫師項目的估計執行一次。

在使用迴歸方法時，產生了另一個考量。在多重迴歸中，運用大子集的自變數會比較小子集產生較差的預測值。因此，變數必須達到 F 輸入下限 4.0 方可使用。本限制可以用語法變更。

### 若要指定預測和已預測變數

- ▶ 在主要的遺漏值分析對話方塊中，選擇您想要採用的迴歸方法來估計遺漏值的變數。
- ▶ 在「估計」組別中，選取「EM 方法」或「迴歸方法」。
- ▶ 按一下「變數」。
- ▶ 如果您想要使用特定而非全部的變數作為預測和已預測變數，請選取「選取變數」並將變數移到合適的清單中。

## MVA 指令的其他功能

指令語法語言也可以讓您：

- 使用 MPATTERN、DPATTERN，或是 TPATTERN 次指令上的 DESCRIBE 關鍵字，來指定遺漏值形式、資料形式和表列形式的各個敘述變數。
- 使用 DPATTERN 次指令，為資料形式表指定多個排序變數。
- 使用 DPATTERN 次指令，為資料形式指定多個排序變數。
- 使用 EM 次指令，來指定允差和收斂。
- 使用 REGRESSION 次指令，來指定允差和 F 輸入。
- 使用 EM 和 REGRESSION 次指令，替「EM 方法」和「迴歸方法」，指定不同的變數清單。
- 為每個 TTESTS、TABULATE 和 MISMATCH，指定未列出之觀察值的不同百分比。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

# 多個插補












多重插補的目的是為遺漏值產生可能值，藉以建立數個完整的資料集。分析程序和一起使用的多重插補資料集分析程序會為每個完整的資料集產生輸出，如果原始的資料集沒有遺漏值，將會加上可預估結果的合併輸出。這些合併結果一般較單一插補方法提供的結果更為精確。

**分析變數。** 分析變數可以是：

- **名義。** 當變數數值代表實質上並未等級化的類別時（例如，有員工工作的公司部門），則此變數可視為名義。名義變數的範例包括地區、郵遞區號以及宗教團體。
- **次序。** 當變數數值代表實質上已等級化的類別時（例如，服務滿意度從非常不滿意到非常滿意分級），則此變數可視為次序。次序變數的範例包括代表滿意度或信賴程度的態度分數、以及偏好等級分數。
- **尺度。** 若一變數可視為尺度（連續），表示它的的數值代表含有實際意義矩陣的已排列順序類別，因此適合比較數值之間的距離。尺度變數的範例包括以年份表示的年齡和以千元為單位的收入。

本程序假設已指定給所有變數適當的測量水準，但您可以在來源變數清單的變數上按一下滑鼠右鍵，並選取快顯功能表上的測量水準，暫時變更變數的測量水準。

變數清單中各變數旁的圖示可識別測量水準和資料類型：

測量水準 (E)	資料類型			
	數字的	字串	日期	時間
尺度 (連續)		無		
次序				
名義				

**頻率加權。** 此程序允許次數（重複）加權。具有負或零重複加權值的觀察值會被忽略。非整數加權會捨入為最接近的整數。

**分析加權。** 分析（迴歸或取樣）加權合併於遺漏值摘要與符合插補模式中。會排除具有負或零分析加權的觀察值。

**複合樣本。**雖然「多重插補」程序能夠接受採分析加權變數形式的最終取樣加權，但其不會自行處理層、集群或其他複雜取樣結構。此外亦請注意，「複雜取樣」程序目前不會自動分析多重插補資料集。若要瞭解完整的支援合併程序清單，請參閱 [分析多重插補資料](#) 第 25 頁。

**遺漏值。**使用者和系統遺漏值會被當作無效值；也就是說，插補值時會取代這兩種遺漏值，且這兩者都會被視為插補模式中預測變數的無效值。遺漏值分析中也會將使用者與系統遺漏值視為遺漏。

**複製結果〈插補遺漏資料值〉。**如果您要精確地複製插補結果，除了使用相同的程序設定外，請為亂數產生器使用相同的初始化值、相同的資料順序和相同的變數順序。

- **亂數產生器。**此程序在計算插補值期間會使用亂數產生器。未來若要重新產生相同的隨機化結果，請先使用與亂數產生器的相同初始化值，再執行每個「插補遺漏資料值」程序。
- **觀察值順序。**採用觀察值順序進行值的插補。
- **變數順序。**完整條件規格 (FCS) 插補方法會依照「分析變數」清單中的順序插補值。

有兩種對話方塊是多重插補專用的。

- **分析形式**提供資料中遺漏值形式的描述性測量，在預檢插補之前步驟時相當實用。
- **插補遺漏資料值**用來產生多重插補。完整資料集可使用支援多重插補資料集的程序進行分析。如需分析多重插補資料集與支援這些資料的程序清單，請參閱 [分析多重插補資料](#) 第 25 頁。

## 分析形式

「分析形式」提供資料中遺漏值形式的描述性測量，在預檢插補之前步驟時相當實用。

**範例。**某電信公司想要更了解客戶資料庫中的服務使用形式。電信公司擁有客戶使用的服務之完整資料，但其收集的人口資訊內有幾個遺漏值。分析遺漏值形式有助於決定插補的下一個步驟。如需詳細資訊，請參閱第 43 頁第 5 章中的 [使用多重插補完成並分析資料集](#)。

### 若要分析遺漏資料形式

從功能表選擇：

分析 (A) > 多個插補 > 分析樣式...

圖表 3-1  
「分析形式」對話方塊



- ▶ 選擇至少兩個分析變數。此程序會分析這些變數的遺漏值。

### 選用設定

**分析加權。** 此變數包含分析（迴歸或取樣）加權。程序會合併遺漏值摘要中的分析權重。會排除具有負或零分析加權的觀察值。

**輸出。** 以下選擇性輸出可供使用：

- **遺漏值摘要。** 這顯示了面板化的圓餅圖，其中對於具備一或多遺漏值的分析變數、觀察值或個別資料值顯示了的數量與百分比。
- **遺漏值形式。** 以表列形式顯示遺漏值。各形式會對應到分析變數之中具有相同形式、不完整與完整資料的一組觀察值。您可使用此輸出，判斷可否在您的資料上使用單調插補方法，以及若不行的話，可得知您的資料與單調形式的近似程度。程序會將分析變數排序為顯露或趨近單調形式。若重新排序之後沒有出現單調形式，您就可推斷出，該資料在分析變數如此排序的時候具有單調形式。
- **具備最高頻率遺漏值的變數。** 這會顯示一個表格，其中分析變數是依照遺漏值百分比遞減排列。表格包含了尺度變數的敘述統計（平均數與標準差）。

您可以控制要顯示的變數數量上限，以及要顯示出的變數遺漏百分比下限。則會顯示符合這兩種條件的變數集合。例如，將變數數量上限設為 50，並將遺漏百分比下限設為 25，即可要求表格顯示多達 50 個遺漏值至少 25% 的變數。若分析變數有 60 個，但只有 15 個具備 25% 以上的遺漏值，輸出只會包含 15 個變數。

## 插補遺漏資料值

「插補遺漏資料值」可用來產生多重插補。完整資料集可使用支援多重插補資料集的程序進行分析。如需分析多重插補資料集與支援這些資料的程序清單，請參閱[分析多重插補資料](#)第 25 頁。

**範例。**某電信公司想要更了解客戶資料庫中的服務使用形式。電信公司擁有客戶使用的服務之完整資料，但其收集的人口資訊內有幾個遺漏值。而且這些值並不是隨機遺漏值，因此可以在此使用多重插補完成資料集。如需詳細資訊，請參閱第 43 頁第 5 章中的[使用多重插補完成並分析資料集](#)。

### 若要插補遺漏資料值

從功能表選擇：

分析 (A) > 多個插補 > 插補遺漏資料值...

圖表 3-2  
插補遺漏資料值變數索引標籤



- ▶ 在插補模式中選擇至少兩個變數。程序會插補這些變數的遺漏資料多重值。
- ▶ 指定要計算的插補數目。此值預設為 5。

- 指定要寫入插補資料的資料集或 IBM® SPSS® Statistics-格式資料檔案。

輸出資料集包含原始觀察值資料和遺漏資料，以及針對每個插補含有插補值的一組觀察值。例如，若原始資料集有 100 個觀察值，而您有 5 個插補，則輸出資料集將有 600 個觀察值。輸入資料集中的所有變數都包含在輸出資料集中。現有變數的字典內容（名稱、標記等）會複製到新資料集。檔案亦包含新變數、Imputation\_、指示插補的數值變數（0 為原始資料，或 1..n 為具備已插補數值的觀察值）。

建立輸出資料集時，程序會自動將 Imputation\_ 變數定義為分割變數。若執行程序時分割起了作用，則輸出資料集中，分割變數值的各種組合都會包含一組插補。

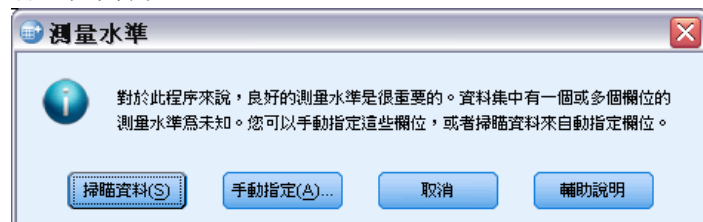
### 選用設定

**分析加權。** 此變數包含分析（迴歸或取樣）加權。程序會於插補遺漏值所用的迴歸與分類模式中，合併分析加權。插補摘要中也會使用分析加權；例如，平均數、標準差、標準誤。會排除具有負或零分析加權的觀察值。

### 具有未知測量水準的欄位

若在資料集中出現一或多個未知的變數（欄位）測量水準，就會顯示「測量水準」警示。由於測量水準會影響此程序的結果計算，因此所有變數皆必須具有已定義的測量水準。

圖表 3-3  
測量水準警示



- **掃描資料。** 讀取作用中資料集的資料，並且針對目前具有未知測量水準的任何欄位指派預設的測量水準。若為大型資料集，則讀取時可能需要一些時間。
- **手動指派。** 開啟對話方塊，以列出具有未知測量水準所有欄位。您可以使用此對話方塊，來指派上述欄位的測量水準。您也可以在此「資料編輯程式」的「變數檢視」中指派測量水準。

由於測量水準是此程序的重要項目，因此您在所有欄位皆擁有已定義的測量水準之前，無法存取對話方塊來執行此程序。



## 方法

圖表 3-4  
插補遺漏資料值方法索引標籤

「方法」索引標籤會指定遺漏值的插補方式，包括使用的模式類型。類別預測變數為經過編碼的指標（虛擬）。

**插補方法。**「自動化」方法會掃描資料，若資料顯示遺漏值的單調形式，則使用單調方法；否則就會使用完整條件規格。若您確定要使用的方法，可將此方法指定為「自訂」方法。

- **完整條件規格。**這是 Markov chain Monte Carlo (MCMC) 法的疊代，當遺漏資料為任意值時（單調或非單調）可使用。

對於採用變數清單中指定順序的每個疊代和變數，完整條件規格 (FCS) 法會使用模式中的其他所有可用變數作為預測變數，以適用於單變量（單一依變數）模式，之後會插補適合變數的遺漏值。此方法在達到最大疊代數目前會持續進行，並將位於最大疊代的已插補值儲存為已插補資料集。

**最大疊代。** 這指定疊代的數量，或 FCS 方法使用 Markov chain 時要進行的「步驟」數。若自動選擇 FCS 方法，則會使用 10 個疊代的預設數量。若您明確的選擇 FCS，可指定自訂的疊代數量。若 Markov chain 未收斂的話。您可新增疊代數量。在「輸出」索引標籤中，您可儲存 FCS 疊代歷程資料，並繪圖評估收斂。

- **單調。** 這是非疊代方法，僅能在資料具有遺漏值的單調形式時使用。若您可排列變數，讓某變數具有非遺漏值時，先前所有變數亦具有非遺漏值，那麼就表示單調形式存在。若指定其作為「自訂」方法，請務必以單調樣式的顯示順序來指定清單中的變數。

針對單調順序中的每個變數，單調方法會使用模式中的所有先前變數作為預測變數，以適用於單變量（單一依變數）模式，之後會插補適合變數的遺漏值。這些已插補值會儲存為已插補資料集。

**包含二因子交互作用。** 自動選擇插補方法時，各變數的插補模式會包含預測變數的常數項與主效應。選擇特定方法時，您可選擇性的在類別預測變數間包含所有可能的二因子交互作用。

**尺度變數的模式類型。** 自動選擇插補方法時，會使用線性迴歸作為尺度變數的單變量方法。若選用特定方法，您可改為選擇預測平均數配對 (PMM) 作為尺度變數的方法。PMM 是線性迴歸的變形，其中會將迴歸模式計算出的插補值與最接近的觀察值進行比對。

Logistic 迴歸一向用來當作類別變數的單變量模式。不論模式類型為何，皆會使用指標（虛擬）編碼來處理類別預測變數。

**奇異性容忍值。** 奇異（或非可逆）矩陣具有線性相依行，其會導致嚴重的估計演算法問題。事實上，近似奇異矩陣會導致不良的結果，因此，此程序將會處理其行列式小於奇異容忍值的矩陣。指定一個正值。

## 限制

圖表 3-5  
插補遺漏資料值限制索引標籤

**插補遺漏資料值**

變數 方法 限制 輸出

變數摘要的資料掃描

限制掃描的觀察值數目(L) 觀察值(C): 5000

變數摘要(V):

模式中的變數	百分比遺漏	最小觀察值	最大觀察值
tenure	3.20	1	72
age	2.50	18	77
marital	11.50	0	1

已掃描的觀察值: 1000

定義限制(D):

模式中的變數	角色	最小值	最大值	捨入
gender	插補做為預測			
reside	插補做為預測	1		1
lninc	插補做為預測	0		

排除含有大量遺漏值的變數(E)

最大百分比遺漏(M):

最大觀察值取樣(A): 50

最大參數取樣(X): 2

增加最大參數取樣可明顯增加分析時間。

「限制」索引標籤可讓您在插補過程中限制變數角色，以及限制尺度變數的合理插補值範圍。此外，您也可以將分析限制為含有小於最大遺漏值百分比的變數。

**掃描資料以取得變數摘要。** 按一下「掃描資料」會使清單顯示分析變數，以及各變數觀察到的遺漏百分比、最小值與最大值。摘要可根據所有觀察值，或限定掃描「觀察值」文字方塊中所指定的前 n 個觀察值。按一下「重新掃描資料」，更新分配摘要。

### 定義限制

- 角色。** 這可讓您自訂要插補的變數和/或要當作預測值處理的變數集合。一般而言，各分析變數會同時被當作插補模式中的依變數與預測變數。「角色」可用來為您希望「僅當作預測值使用」的變數關閉插補，或排除變數使其無法被當作預測變數使用（僅插補），這樣可讓預測模式更精簡。這是唯一能指定類別變數（或僅當作預測值使用的變數）的限制。

- **最小值與最大值。** 這幾欄可讓您指定尺度變數可用插補值的最小值與最大值。若插補值落在範圍外，程序會抽出其他值，直到找到範圍內插補值為止，或是抽出數量達到上限為止（請參閱下方的**最大抽取**）。唯有在「方法」索引標籤中選取了「**線性迴歸**」尺度變數模式類別，才能使用這幾欄。
- **四捨五入。** 部分變數可做為尺度使用，但會使值自然受到進一步限制；例如，一個家庭中的人數必須是整數，去雜貨店花費的金額則不能有分數。此行可讓您指定可接受的最小面額。例如，若要獲得整數值，您要指定 1 做為捨入面額；若要獲得捨入為最接近分的值，您要指定 0.01。一般而言，值會捨入為最近的整數乘以捨入面額。下表顯示當插補值為 6.64823（捨入前）時，不同的捨入值如何運作。

捨入面額	捨入 6.64832 的值
10	10
1	7
0.25	6.75
0.1	6.6
0.01	6.65

**排除有大量遺漏值的變數。** 一般而言，分析變數若具備可估計插補模式的足夠數值，無論有多少遺漏值，都會被插補且當作預測值使用。您可選擇排除所含遺漏值比例較高的變數。例如，若您把遺漏百分比上限定為 50，則遺漏值含量超過 50% 的分析變數就不會被插補，也不會被當作插補模式中的預測值。

**抽出上限。** 若尺度變數指定了最小值或最大值（請參閱上述的**最小值與最大值**），則程序會嘗試抽出觀察值，直到數值組合位於指定範圍內。若對每個觀察值抽取特定次數之後，還是未獲得數值組合，程序會嘗試抽取其他模式參數組合，並重複觀察值抽取程序。若抽取觀察值與參數達特定次數之後，還是未獲得範圍內的數值組合，則會發生錯誤。

請注意，增加這些數值會增加處理時間。若程序花費太多時間，或無法找到適當抽取值，請檢查指定的最小值和最大值，並確保這些數值正確。

## 輸出

圖表 3-6  
插補遺漏資料值輸出索引標籤



**顯示。** 控制輸出顯示。整體插補摘要會顯示出來，其中包含插補規格、疊代（針對完整條件規格法）、已插補的依變數、排除不予插補的依變數、插補序列等相關表格。若有指定的話，也會顯示分析變數的限制。

- **插補模式。** 這裡會顯示依變數與預測值的插補模式，並包含單變量類型、模式效應，以及已插補數值數量。
- **敘述統計。** 這裡會顯示已插補數值的依變數之敘述統計。對於尺度變數，敘述統計包含原始輸入資料（插補前）、插補值（已插補）與完整資料（原始值加上插補值插補後的值一再進行插補）的平均數、個數、標準差、最小值、最大值。對於類別變數，敘述統計包含原始輸入資料（插補前）、插補值（已插補）與完整資料（原始值加上插補值插補後的值一再進行插補）的數量與類別百分比。

**疊代歷程。** 使用完整條件規格插補法時，您可要求包含疊代歷程資料的資料集，以進行 FCS 插補。資料集會為已插補的數值疊代和插補各尺度依變數，以包含平均數和標準差。您可繪製資料圖，協助評估模式收斂。如需詳細資訊，請參閱第 62 頁第 5 章中的 [檢查 FCS 收斂](#)。

## MULTIPLE IMPUTATION 指令的其他功能

指令語法語言也可以讓您：

- 指定會顯示敘述統計的變數子集（`IMPUTATIONSUMMARIES` 次命令）。
- 在單一次程序執行中指定遺漏形式與插補的分析。
- 指定插補變數時所允許的模式參數最大數目（關鍵字 `MAXMODELPARAM`）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

## 使用多重插補資料

建立多重插補 (MI) 資料集時，會新增新變數，也就是變數標籤為插補數的 `Imputation_`，且會用來以遞增順序排序資料集。原始資料集的觀察值具有數值 0。已插補數值的觀察值會被編號 1 到 M，其中 M 是插補數。

開啟資料集時，有 `Imputation_` 的話代表資料集有可能是 MI 資料集。

### 啟動多重插補插補資料集進行分析

資料集必須使用比較組別選項分割，其中 `Imputation_` 需做為分組變數，才能在分析中當作 MI 資料集使用。亦可定義其他變數上的分割。

從功能表選擇：

資料 > 分割檔案 (F)...

圖表 3-7  
分割檔案對話方塊



- ▶ 選取比較組別。
- ▶ 選擇插補數 [`Imputation_`] 做為分組觀察值依據的變數。

或者，若您開啟標示（請參閱下方），會依據插補數 [`Imputation_`] 分割檔案。

### 區分已插補數值與觀察值

您可根據儲存格背景顏色、字型與粗體字（用於已插補數值），區分已插補數值與觀察值。如需使標示生效的詳細資訊，請參閱[多重插補選項](#)第 29 頁。當您在目前的階段作業中建立含有「插補遺漏值」的新資料集時，預設會開啟標示功能。當您開啟包含插補的已儲存資料檔案時，會關閉標示功能。

圖表 3-8  
資料編輯器，插補標示關閉

	Imputation	tenure	age	marital	address	income
1034	1	11	27	1	16	55
1035	1	60	46	0	13	163
1036	1	20	35	1	4	52
1037	1	66	60	0	38	211
1038	1	44	57	1	1	186
1039	1	11	41	1	0	39

若要開啟標示，請從「資料編輯程式」功能表選擇：  
檢視 > 標示插補資料...

圖表 3-9  
資料編輯器，插補標示開啟

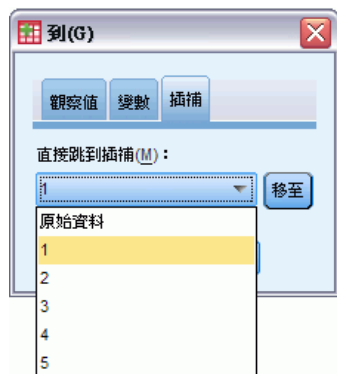
	Imputation	tenure	age	marital	address	income
1034	1	11	27	1	16	55
1035	1	60	46	0	13	163
1036	1	20	35	1	4	52
1037	1	66	60	0	38	211
1038	1	44	57	1	1	186
1039	1	11	41	1	0	39

或可按一下「資料編輯程式」中「資料檢視」其編輯列右緣的插補標示按鈕，開啟標示。

### 在插補之間移動

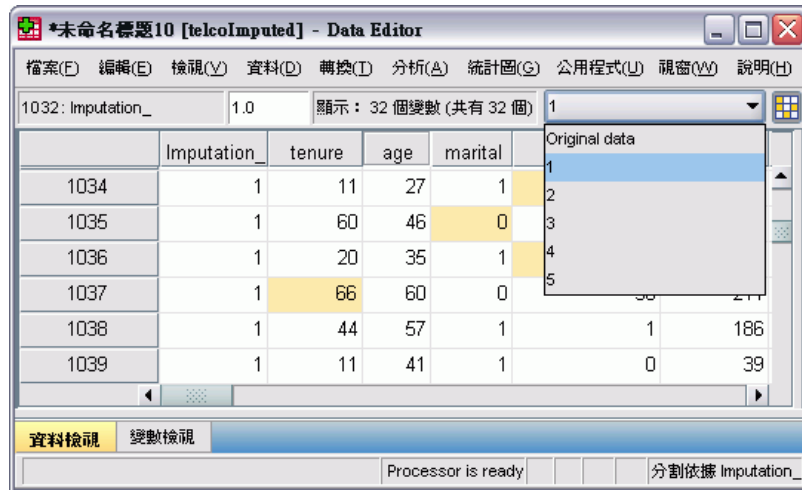
- ▶ 從功能表選擇：  
編輯 > 直接跳到插補 (M)...
- ▶ 從下拉式清單中選取插補 (或原始資料)。

圖表 3-10  
移至對話方塊



或者，可以從「資料編輯器」其「資料檢視」中編輯列的下拉式清單選取插補。

圖表 3-11  
資料編輯器，插補標示開啟



選取插補時，會保存相對觀察值位置。例如，若原始的資料集中有 1000 個觀察值，則觀察值 1034 (第一個插補中的第 34 個觀察值) 會顯示於網格頂端。若您在下拉式清單中選取插補 2、2034 觀察值，則插補 2 中的第 34 個觀察值會顯示於網格頂端。若您在下拉式清單中選取原始資料，觀察值 34 會顯示於網格頂端。在插補之間瀏覽時，也會保存行位置，如此就能輕鬆比較插補之間的值。



### 轉換與編輯插補值

有時必須執行已插補資料的轉換。例如，您可取得薪資變數全部數值的記錄，並將結果另存為新變數。使用插補資料所計算出的值，若與使用原始資料計算出的值不同，會被視為已插補。

若您在「資料編輯程式」中的某儲存格編輯插補值，該儲存格仍會被視為已插補的。不建議以這種方式編輯插補值。

## 分析多重插補資料

許多程序都支援從多重插補資料集的分析中合併結果。開啟插補標記時，支持合併的程序旁會顯示一個小圖示。例如，在「分析」功能表的「敘述統計」子功能表中，「次數分配表」、「描述性統計量」、「預檢資料」、「交叉表」都支援合併，而「比率」、「P-P 圖」和「Q-Q 圖」則否。

圖表 3-12  
分析功能表，插補標示開啟



表格輸出和模式 PMML 都可以合併。要求合併輸出並不需要新程序；而是「選項」對話方塊上會有個新索引標籤，讓您全面控制多重插補輸出。

- **表格輸出的合併。** 依照預設，當您在多重插補 (MI) 資料集執行支援的程序時，會自動產生包含了各插補、原始 (未插補) 資料、已合併 (最終) 結果的結果，這些在插補間變化時都會被納入考量。合併的統計量會依程序而不同。
- **PMML 的合併。** 您亦可從匯出 PMML 的支援程序取得合併的 PMML。合併的 PMML 也是以相同的方式要求，但與未合併的 PMML 不同的是，會被儲存起來。

不支援的程序所產生的結果不是已合併輸出，也不是已合併 PMML 檔案。

### 合併等級

輸出會使用下列兩種等級之一合併：

- **單純合併。** 僅可用於合併的參數。
- **單變量合併。** 已合併的參數、其標準誤、檢定統計量與有效自由度、p-值、信賴區間以及合併診斷 (遺漏資訊的分數、相對效能、變異數相對增量) 等，若可使用都會顯示出來。

通常會合併係數 (迴歸與相關)、平均數 (與平均數差異) 及個數。當有可用的統計量標準錯誤時，就會使用單變量合併，否則會使用單純合併。

### 支援合併的程序

下列程序在針對各輸出指定的合併程度上支援 MI 資料集。

#### 次數分配表 (F)

- 「統計量」表格在單變量合併中支援平均數 (如果也要求平均數的標準誤的話)，在單純合併中支援有效 N 和遺漏 N。
- 「次數分配」表在單純合併中支援次數。

#### 描述性統計量 (D)

- 「敘述統計」表格在單變量合併中支援平均數 (如果也要求平均數的標準誤的話)，在單純合併中支援 N。

#### 交叉表

- 「交叉表列」表格在單純合併中支援個數。

#### 平均數

- 「報告」表在單變量合併中支援平均數 (如果也要求平均數的標準誤的話)，在單純合併中支援 N。

#### 單一樣本 T 檢定

- 「統計量」表格在單變量合併中支援平均數，在單純合併中支援個數。
- 「檢定」表格在單純合併中支援平均數差異。

#### 獨立樣本 T 檢定

- 「組別統計量」表格在單變量合併中支援平均數，在單純合併中支援個數。
- 「檢定」表在單變量合併中支援平均數差異。

### 成對樣本 T 檢定

- 「統計量」表格在單變量合併中支援平均數，在單純合併中支援個數。
- 「相關性」表格在單純合併中支援相關與 N。
- 「檢定」表格在單變量合併中支援平均數。

### 單因子變異數分析

- 「敘述統計」表格在單變量合併中支援平均數，在單純合併中支援個數。
- 「對比檢定」表格在單變量合併中支援對比值。

### GLM 單變量、GLM 多變量與重複的 GLM

- 「受試者間因子」表在單純合併中支援個數。
- 「敘述統計」表格在單純合併中支援平均數與個數。
- 「參數估計值」表格在單變量合併中支援係數 B。
- 邊際平均數估計：「估計值」表格在單變量合併中支援平均數。
- 邊際平均數估計：「成對比較」表格在單變量合併中支援平均數差異。

### 線性混合模式

- 「敘述統計」表格在單純合併中支援平均數與個數。
- 「固定效果估計」在單變量合併中支援估計。
- 「估計共變異數參數」表在單變量合併中支援估計。
- 邊際平均數估計：「估計值」表格在單變量合併中支援平均數。
- 邊際平均數估計：「成對比較」表格在單變量合併中支援平均數差異。

「**概化線性模式**」與「**概化估計方程式**」。這些程序支援合併的 PMML。

- 「類別變數資訊」表格在單純合併中支援個數與百分比。
- 「連續變數資訊」在單純合併中支援個數與平均數。
- 「參數估計值」表格在單變量合併中支援係數 B。
- 邊際平均數估計：「估計係數」表格在單純合併中支援平均數。
- 邊際平均數估計：「估計值」表格在單變量合併中支援平均數。
- 邊際平均數估計：「成對比較」表格在單變量合併中支援平均數差異。

### 雙變數相關分析

- 「敘述統計」表格在單純合併中支援平均數與個數。
- 「相關性」表格在單純合併中支援相關與 N。

### 偏相關

- 「敘述統計」表格在單純合併中支援平均數與個數。
- 「相關性」表格在單純合併中支援相關性。

**線性迴歸**。此程序支援合併的 PMML。

- 「敘述統計」表格在單純合併中支援平均數與個數。

- 「相關性」表格在單純合併中支援相關與 N。
- 「係數」表格在單變量合併中支援 B，在單純合併中支援相關性。
- 「相關係數」表格在單純合併中支援相關性。
- 「殘差統計量」表格在單純合併中支援平均數與個數。

**二元 Logistic 迴歸。** 此程序支援合併的 PMML。

- 「方程式中變數」表格在單變量合併中支援 B。

**多項式 Logistic 迴歸。** 此程序支援合併的 PMML。

- 「參數估計值」表格在單變量合併中支援係數 B。

#### **次序迴歸**

- 「參數估計值」表格在單變量合併中支援係數 B。

**判別分析。** 此程序支援合併的模式 XML。

- 「組別統計」表格在單純合併中支援平均數與有效個數。
- 「合併的組別內矩陣」表格在單純合併中支援相關性。
- 「典型判別函數係數」表格在單純合併中支援未標準化係數。。
- 「各組重心上函數」表格在單純合併中支援未標準化係數。
- 「分類函數係數」表格在單純合併中支援係數。

#### **「卡方檢定」**

- 「描述性統計量」表格在單純合併中支援平均數與個數。
- 「次數分配」表格在單純合併中支援觀察次數。

#### **二項式檢定**

- 「描述性統計量」表格在單純合併中支援平均數與個數。
- 「檢定」表格在單純合併中支援個數、觀察比例和檢定比例。

#### **連檢定**

- 「描述性統計量」表格在單純合併中支援平均數與個數。

#### **單一樣本 Kolmogorov-Smirnov 檢定**

- 「描述性統計量」表格在單純合併中支援平均數與個數。

#### **兩個獨立樣本檢定**

- 「等級」表格在單純合併中支援平均等級與個數。
- 「次數分配」表格在單純合併中支援個數。

#### **多個獨立樣本的檢定**

- 「等級」表格在單純合併中支援平均等級與個數。
- 「次數分配」表格在單純合併中支援個數。

### 兩個相關樣本檢定

- 「等級」表格在單純合併中支援平均等級與個數。
- 「次數分配」表格在單純合併中支援個數。

### 多個相關樣本的檢定

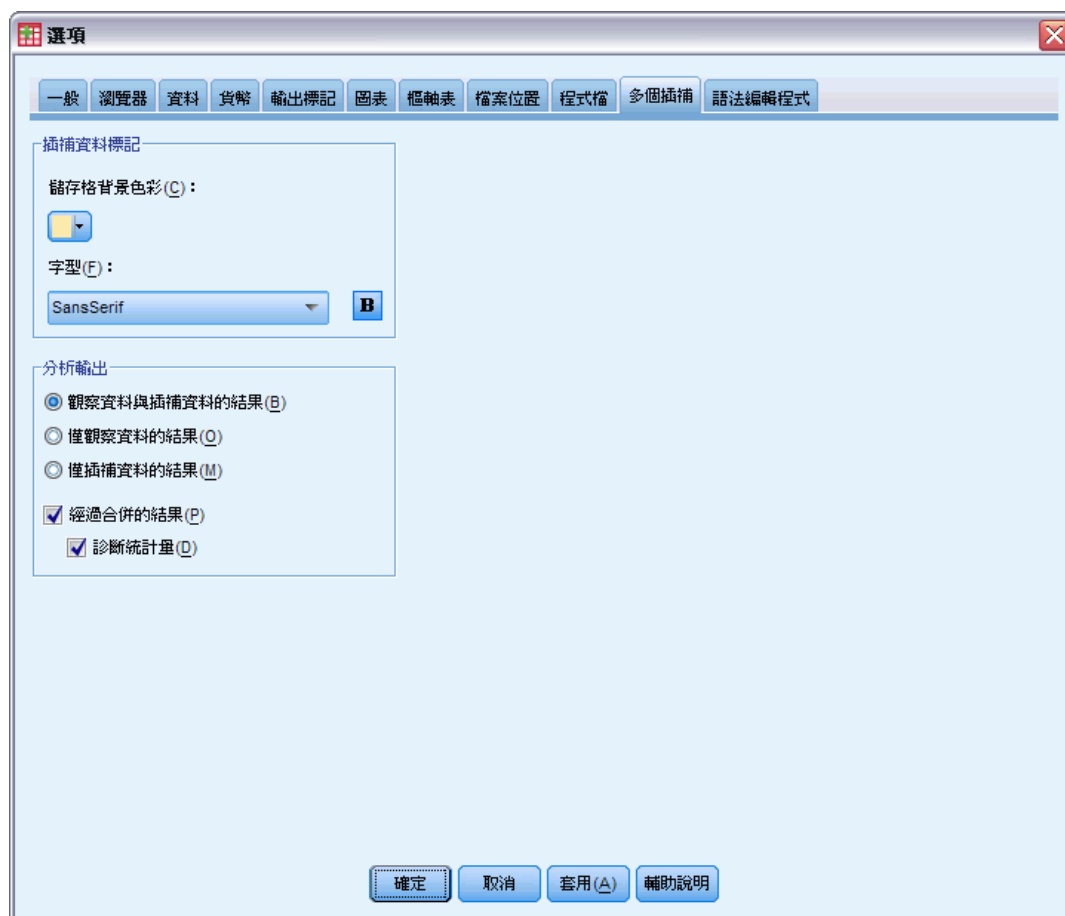
- 「等級」表格在單純合併中支援平均等級。

**Cox 迴歸。** 此程序支援合併的 PMML。

- 「方程式中變數」表格在單變量合併中支援 B。
- 「共變量平均數」表格在單純合併中支援平均數。

## 多重插補選項

圖表 3-13  
「選項」對話方塊：「多重插補」索引標籤



「多重插補」索引標籤控制兩種和「多重插補」相關的偏好設定：

**「插補資料」的外觀。** 根據預設，含有插補資料的儲存格與含有非插補資料的儲存格背景顏色將不同。不同外觀的插補資料會讓捲動資料集和尋找這些儲存格變得容易。您可以變更預設的儲存格背景顏色、字型，並讓插補資料以粗體類型顯示。

**分析輸出。** 此群組控制分析多重插補資料集時，所產生的瀏覽器輸出類型。根據預設，將會針對原始的（插補前）資料集和每個插補資料集產生輸出。此外，對於支援插補資料合併的程序，將會產生最終的合併結果。執行單變量合併時，也將顯示合併診斷。不過，您可以隱藏任何您不想看見的輸出。

### **若要設定多重插補選項**

在功能表上，選擇：

編輯 > 選項

按一下「多重插補」索引標籤。

# 部 11: 範 例

# 遺漏值分析

## 說明遺漏資料的形式

某電信公司想要更了解客戶資料庫中的服務使用形式。在進行進一步分析前，該公司想要確認資料完全是隨機遺漏。

取自客戶資料庫的隨機樣本包含在 telco\_missing.sav 中。如需詳細資訊，請參閱第 76 頁附錄 A 中的[範例檔案](#)。

## 執行分析以顯示敘述統計

- ▶ 若要執行「遺漏值分析」，請從功能表選擇：  
分析 > 遺漏值分析...

圖表 4-1  
遺漏值分析對話方塊



- ▶ 選擇「婚姻狀況 [marital]」、「教育程度 [ed]」、「退休 [retire]」和「性別 [gender]」作為類別變數。

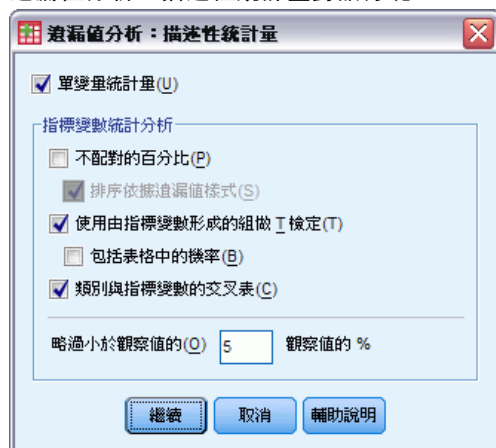


- ▶ 選擇「服務期間 (以月為單位) [tenure]」到「家庭人數 [reside]」作為數值 (尺度) 變數。

在這個時候，您可以執程序，並取得單變量統計量，但是我們將選擇其他敘述統計。

- ▶ 按一下「描述性統計量」。

圖表 4-2  
遺漏值分析：描述性統計量對話方塊



在「描述性統計量」對話方塊中，您可以指定要在輸出中顯示的各種敘述統計。預設的單變量統計量可協助您判斷遺漏資料的一般範圍，而指標變數統計量可提供一個變數的遺漏資料形式如何影響其他變數值的詳細資訊。

- ▶ 選擇「使用由指標變數形成組別的 t 檢定」。
- ▶ 選擇「類別與指標變數的交叉表」。
- ▶ 按一下「繼續」。
- ▶ 在「遺漏值分析」主對話方塊中，按一下「確定」。

## 評估敘述統計

對於此範例，輸出內容包括：

- 單變量統計量
- 個別變異數 t 檢定表格，包括存在或遺漏其他變數時的次組別平均數
- 各類別變數的表格，顯示各類別中各數值 (尺度) 變數的遺漏資料頻率

圖表 4-3  
單變量統計量表

	個數	平均數	標準差	遺漏值		極值個數 <sup>a</sup>	
				個數	百分比	極小值	極大值
MonthsWithService	968	35.56	21.268	32	3.2	0	0
Age	975	41.75	12.573	25	2.5	0	0
YearsAtAddress	850	11.47	9.965	150	15.0	0	9
Income	821	71.1462	83.14424	179	17.9	0	71
YearsWithEmployer	904	11.00	10.113	96	9.6	0	15
PeopleInHousehold	966	2.32	1.431	34	3.4	0	33
MaritalStatus	885			115	11.5		
EducationalLevel	965			35	3.5		
RetirementStatus	916			84	8.4		
Gender	958			42	4.2		

a. 超出範圍 (Q1 - 1.5\*IQR, Q3 + 1.5\*IQR) 的觀察值個數。

單變量統計量可供您先行檢視逐一變數的遺漏資料範圍。各變數的非遺漏值個數會顯示在 N 欄，而遺漏值個數會顯示在「遺漏個數」欄。「遺漏百分比」欄會顯示含遺漏值的觀察值百分比，並提供比對變數中遺漏資料範圍的適當測量方式。「收入（以千為單位）」擁有最多包含遺漏值的觀察值（17.9%），而「年齡（以年為單位）」擁有的數量最少（2.5%）。「收入」也含有最多極端值。

圖表 4-4  
個別變異數 t 檢定表格

	MonthsWithService	Age	YearsAtAddress	Income	YearsWithEmployer	PeopleInHousehold
t	.4	.3	.	3.5	1.4	1.0
自由度	202.2	192.5	.	313.6	191.1	199.5
# 出現數	819	832	850	693	766	824
# 遺漏值個數	149	143	0	128	138	142
平均數(非遺漏值)	35.68	41.79	11.47	74.0779	11.20	2.34
平均數(遺漏值)	34.91	41.49	.	55.2734	9.86	2.21
t	-5.0	-8.3	-3.9	.	-5.9	3.6
自由度	249.5	222.8	191.1	.	203.3	315.2
# 出現數	793	801	693	821	741	792
# 遺漏值個數	175	174	157	0	163	174
平均數(非遺漏值)	33.93	40.01	10.67	71.1462	9.91	2.39
平均數(遺漏值)	42.97	49.73	14.97	.	15.93	2.02
t	-1.0	-.4	-.7	.5	.	-.3
自由度	110.5	110.2	97.6	114.9	.	110.9
# 出現數	877	881	766	741	904	874
# 遺漏值個數	91	94	84	80	0	92
平均數(非遺漏值)	35.34	41.69	11.37	71.4953	11.00	2.31
平均數(遺漏值)	37.70	42.27	12.32	67.9125	.	2.37
t	.0	1.8	1.2	-.8	.9	-2.2
自由度	148.1	149.5	138.8	121.2	128.3	134.2
# 出現數	856	862	748	728	805	857
# 遺漏值個數	112	113	102	93	99	109
平均數(非遺漏值)	35.56	42.00	11.61	70.3887	11.10	2.28
平均數(遺漏值)	35.57	39.85	10.43	77.0753	10.17	2.61
t	-.6	-.4	-.4	.3	.	.2
自由度	95.4	94.4	84.0	93.2	.	99.0
# 出現數	888	893	777	751	904	885
# 遺漏值個數	80	82	73	70	0	81
平均數(非遺漏值)	35.44	41.70	11.42	71.3356	11.00	2.32
平均數(遺漏值)	36.89	42.29	11.96	69.1143	.	2.30

個別變異數 t 檢定表格可協助識別遺漏值形式可能影響數量（尺度）變數的變數。t 檢定是使用指定個別觀察值的變數是否存在或遺漏的指標變數進行計算。指標變數的次組別平均數也會製成表格。請注意，只有在變數至少有 5% 觀察值的遺漏值時，才會建立指標變數。

較年長的應答者似乎不太可能回報收入水準。當收入遺漏時，平均年齡為 49.73，而收入未遺漏時，則為 40.01。事實上，遺漏收入似乎會影響多個數量（尺度）變數的平均數。這表示資料可能不是完全隨機遺漏。

圖表 4-5  
婚姻狀況 [marital] 交叉表

			總和	Unmarried	Married	遺漏
						系統遺漏值
YearsAtAddress	目前	個數	850	390	358	102
		百分比	85.0	85.5	83.4	88.7
	遺漏	% 系統遺漏值	15.0	14.5	16.6	11.3
Income	目前	個數	821	380	348	93
		百分比	82.1	83.3	81.1	80.9
	遺漏	% 系統遺漏值	17.9	16.7	18.9	19.1
YearsWithEmployer	目前	個數	904	418	387	99
		百分比	90.4	91.7	90.2	86.1
	遺漏	% 系統遺漏值	9.6	8.3	9.8	13.9
RetirementStatus	目前	個數	916	423	392	101
		百分比	91.6	92.8	91.4	87.8
	遺漏	% 系統遺漏值	8.4	7.2	8.6	12.2

類別變數與指標變數的交叉表顯示的資訊，類似個別變異數 t 檢定表格中出現的資訊。指標變數會再次建立，但這次不包括用於針對各類別變數計算各類別頻率的狀況。這些值可協助您判斷類別之間的遺漏值是否不同。

檢視表格中的marital (婚姻狀況)，婚姻類別之間的指標變數遺漏值個數似乎差異不大。是否已婚或未婚似乎不影響任何數量 (尺度) 變數的資料是否遺漏。例如，未婚者回報地址 (現址居住年數) 的機率为 85.5%，而已婚者回報同一變數的機率为 83.4%。其中差距很小，很可能是隨機發生。

圖表 4-6  
教育程度 [ed] 交叉表

			總和	Did not complete high school	High school degree	Some college	College degree	Post-undergraduate degree	遺漏 系統遺漏值
YearsAtAddress	目前	個數	850	163	240	175	186	56	30
		百分比	85.0	83.2	85.7	88.4	81.9	87.5	85.7
	遺漏	% 系統遺漏值	15.0	16.8	14.3	11.6	18.1	12.5	14.3
Income	目前	個數	821	155	229	165	193	50	29
		百分比	82.1	79.1	81.8	83.3	85.0	78.1	82.9
	遺漏	% 系統遺漏值	17.9	20.9	18.2	16.7	15.0	21.9	17.1
YearsWithEmployer	目前	個數	904	178	254	178	204	60	30
		百分比	90.4	90.8	90.7	89.9	89.9	93.8	85.7
	遺漏	% 系統遺漏值	9.6	9.2	9.3	10.1	10.1	6.3	14.3
MaritalStatus	目前	個數	885	193	278	148	184	52	30
		百分比	88.5	98.5	99.3	74.7	81.1	81.3	85.7
	遺漏	% 系統遺漏值	11.5	1.5	.7	25.3	18.9	18.8	14.3
RetirementStatus	目前	個數	916	180	259	180	207	60	30
		百分比	91.6	91.8	92.5	90.9	91.2	93.8	85.7
	遺漏	% 系統遺漏值	8.4	8.2	7.5	9.1	8.8	6.3	14.3

現在考慮 ed (教育程度) 交叉表。如果應答者至少有大學教育程度，應答者的婚姻狀況比較可能會遺漏。至少有 98.5% 非大學教育程度的應答者回報婚姻狀況。另一方面，只有 81.1% 有大學學位的應答者回報婚姻狀況。對於大學肄業的應答者，這個數字更低。

圖表 4-7  
退休 [retire] 交叉表

			總和	No	Yes	遺漏 系統遺漏值
YearsAtAddress	目前	個數	850	744	33	73
		百分比	85.0	85.0	80.5	86.9
Income	目前	遺漏 % 系統遺漏值	15.0	15.0	19.5	13.1
		個數	821	732	19	70
YearsWithEmployer	目前	百分比	82.1	83.7	46.3	83.3
		遺漏 % 系統遺漏值	17.9	16.3	53.7	16.7
MaritalStatus	目前	個數	904	864	40	0
		百分比	90.4	98.7	97.6	.0
	目前	遺漏 % 系統遺漏值	9.6	1.3	2.4	100.0
		個數	885	777	38	70
	目前	百分比	88.5	88.8	92.7	83.3
		遺漏 % 系統遺漏值	11.5	11.2	7.3	16.7

retire (退休) 出現更明顯的差異。相較於未退休者，退休者比較不可能回報收入所得。只有 46.3% 的已退休客戶回報收入水準，而非退休者回報收入水準的比例為 83.7。

圖表 4-8  
性別 [gender] 交叉表

			總和	Male	Female	遺漏 系統遺漏值
YearsAtAddress	目前	個數	850	363	456	31
		百分比	85.0	78.6	91.9	73.8
Income	目前	遺漏 % 系統遺漏值	15.0	21.4	8.1	26.2
		個數	821	381	406	34
YearsWithEmployer	目前	百分比	82.1	82.5	81.9	81.0
		遺漏 % 系統遺漏值	17.9	17.5	18.1	19.0
MaritalStatus	目前	個數	904	412	457	35
		百分比	90.4	89.2	92.1	83.3
	目前	遺漏 % 系統遺漏值	9.6	10.8	7.9	16.7
		個數	885	400	445	40
RetirementStatus	目前	百分比	88.5	86.6	89.7	95.2
		遺漏 % 系統遺漏值	11.5	13.4	10.3	4.8
	目前	個數	916	420	461	35
		百分比	91.6	90.9	92.9	83.3
	目前	遺漏 % 系統遺漏值	8.4	9.1	7.1	16.7

gender (性別) 出現另一項差異。男性遺漏地址資訊的狀況多於女性。雖然這些差異可能是隨機發生所致，但是情況似乎並非如此。資料似乎不是完全隨機遺漏。

我們將檢視遺漏資料的形式，以進一步探討這個問題。

## 重新執行分析以顯示形式

圖表 4-9  
遺漏值分析對話方塊



- ▶ 叫回「遺漏值分析」對話方塊。對話方塊會記得先前分析時使用的變數。請勿變更這些變數。
- ▶ 按一下「形式」。

圖表 4-10  
遺漏值分析形式對話方塊



在「形式」對話方塊中，您可以選擇各種形式表。我們將顯示按照遺漏值形式分組的表列形式。由於 ed（教育程度）、retire（退休）和 gender（性別）中的遺漏形式似乎會影響資料，因此我們選擇顯示這些變數的其他資訊。由於遺漏值數量相當多，因此我們也將包含收入（以千為單位的家庭收入）的其他資訊。

- ▶ 選擇「按照遺漏值形式分組的表列觀察值」。
- ▶ 選擇 Income、EducationalLevel、RetirementStatus 和 Gender，並加入至「額外的訊息」清單。
- ▶ 按一下「繼續」。
- ▶ 在「遺漏值分析」主對話方塊中，按一下「確定」。



## 評估形式表

圖表 4-11  
表列形式表

觀察值個數	遺漏形式 <sup>a</sup>									數 值 併 計	Income <sup>e</sup>	EducationalLevel <sup>d</sup>				RetirementStatus <sup>d</sup>		Gender <sup>d</sup>			
	Age	PeopleInHousehold	MonthsWithService	EducationalLevel	Gender	RetirementStatus	YearsWithEmployer	MaritalStatus	YearsAtAddress			Income	Did not complete high school	High school degree	Some college	College degree	Post-undergraduate degree	No	Yes	Male	Female
475										475	76.5853	99	157	87	101	31	463	12	201	274	
109									X	584	.	27	35	19	17	11	95	14	47	62	
16									X	687	.	5	9	0	1	1	12	4	12	4	
87									X	562	54.4368	21	27	9	24	6	85	2	66	21	
13		X								488	56.0000	4	3	2	3	1	13	0	4	9	
60				X					X	535	77.2167	1	2	27	24	6	59	1	35	25	
16					X					491	47.8125	0	0	0	0	0	16	0	6	10	
17			X							492	76.2353	2	7	3	4	1	17	0	7	10	
18					X					493	54.1111	3	7	4	4	0	17	1	0	0	
16						X			X	660	.	0	0	7	8	1	14	2	6	10	
37										520	59.4595	9	14	5	8	1	0	0	15	22	

未顯示少於 1% 觀察值 (10 或更少) 的形式。

- 依遺漏形式來排序變數。
- 未使用該形式 (標記為 X) 中遺漏變數時的完整觀察值個數。
- 唯一形式的均值
- 每個唯一形式的比率分布

表列形式表顯示個別觀察值的多個變數是否容易遺漏資料。也就是說，這可協助您判斷是否有資料連帶遺漏。

有三種連帶遺漏資料的形式出現在 1% 以上的觀察值中。相較於其他成對變數，employ (服務於目前雇主的年數) 和 retire (退休) 這對變數更容易連帶遺漏。這不令人意外，因為 retire 和 YearsWithEmployer 記錄類似的資訊。如果您不知道應答者是否已退休，則可能也不知道應答者受雇於目前雇主的年數。

平均收入 (以千為單位的家庭收入) 似乎會因為遺漏值形式而出現極大的差異。尤其在 marital (婚姻狀況) 遺漏時，觀察值中有 6% (1000 個中有 60 個) 的平均 Income 更高。(tenure (服務期間 (以月為單位)) 遺漏時也會比較高，但是此形式只能解釋其中 1.7% 的觀察值。) 先前提到教育程度較高的應答者很少回應關於婚姻狀況的問題。您可以在 ed (教育程度) 呈現的頻率中看出這個趨勢。我們可以假設，具有較高教育程度的應答者收入較高，且較不可能回答婚姻狀況的問題，這才能夠解釋收入增加的現象。

考慮到遺漏資料的敘述統計和形式，我們或許能夠提出資料不是完全隨機遺漏的結論。透過 Little 的 MCAR 檢定 (與 EM 評估一同印出)，可以證實這個結論的正確性。

## 針對 Little 的 MCAR 檢定重新執行分析

圖表 4-12  
遺漏值分析對話方塊



- ▶ 叫回「遺漏值分析」對話方塊。
- ▶ 按一下「EM 方法」。
- ▶ 按一下「確定」。

圖表 4-13  
EM 平均數表格

MonthsWithService	Age	YearsAtAddress	Income	YearsWithEmployer	PeopleInHousehold
36.12	41.91	11.58	77.3941	11.22	2.29

a. Little's MCAR 檢定: 卡方 = 179.836、DF = 107、Sig. = .000

Little 的 MCAR 檢定結果出現在各個 EM 估計表的註腳中。Little 的 MCAR 檢定有一項虛無假設，也就是假設資料完全是隨機遺漏 (MCAR)。當遺漏值的形式與資料值無關時，資料為 MCAR。由於我們範例中的顯著值低於 0.05，因此我們可以推斷資料不是完全隨機遺漏。這可證實我們從敘述統計和表列形式得到的結論。

在這個時候，由於資料並不是完全隨機遺漏，因此完全刪除清單中含有遺漏值的觀察值或單獨插補遺漏值並不妥當。然而您可以使用「[多重插補](#)」來深入分析此資料集。

# 多重插補

## 使用多重插補完成並分析資料集

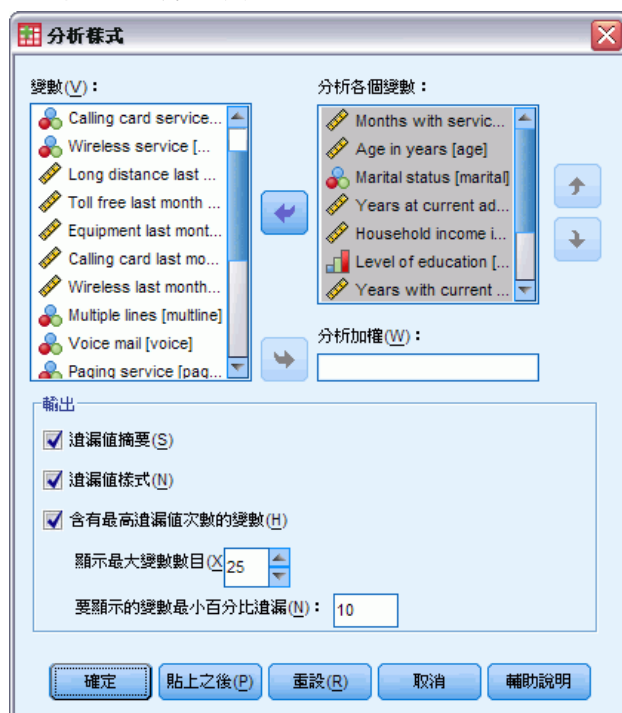
某電信公司想要更了解客戶資料庫中的服務使用形式。電信公司擁有客戶使用的服務之完整資料，但其收集的人口資訊內有幾個遺漏值。而且這些值並不是隨機遺漏值，因此可以在此使用多重插補完成資料集。

取自客戶資料庫的隨機樣本包含在 `telco_missing.sav` 中。如需詳細資訊，請參閱第 76 頁附錄 A 中的[範例檔案](#)。

### 分析遺漏值形式

- ▶ 首先請先檢視遺漏值的形式。從功能表選擇：  
分析 > 多重插補 > 分析形式...

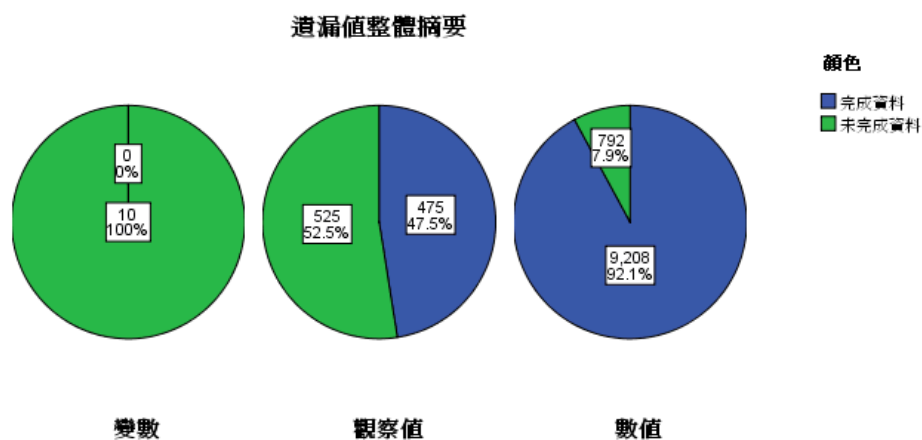
圖表 5-1  
「分析形式」對話方塊



- ▶ 選擇「服務期間 (以月為單位) [tenure]」到「家庭人數 [reside]」作為分析變數。

## 整體摘要

圖表 5-2  
遺漏值的整體摘要



遺漏值的整體摘要會顯示三個圓餅圖，其中包含資料中遺漏值的三種不同層面。

- 變數圖表顯示 10 個分析變數的觀察值中各有至少一個遺漏值。

- 觀察值圖表顯示在 1000 個觀察值內，有 525 個觀察值的變數中各有至少一個遺漏值。
- 值圖表顯示在 10,000 個值內，遺漏了 792 個值（觀察值 × 變數）。

一般而言，每個包含遺漏值的觀察值中，大約每 10 個變數就有 1.5 個遺漏值。這表示刪除全部遺漏值會讓資料集中更多的資訊流失。

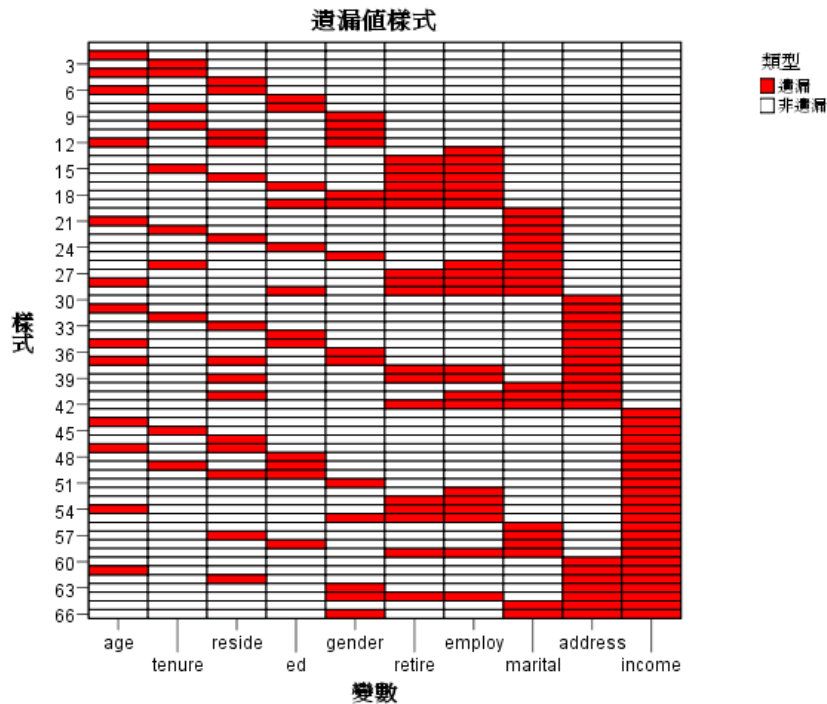
## 變數摘要

圖表 5-3  
變數摘要

	遺漏		有效個數	平均數	標準離差
	個數	百分比			
家庭收入(千元)	179	17.9%	821	71.1462	83.14424
現址年數	150	15.0%	850	11.47	9.965
婚姻狀態	115	11.5%	885		

變數摘要會顯示包含至少 10% 遺漏值的變數，並在表格中列出各個變數的遺漏值數目及百分比。此摘要也會顯示尺度變數有效值的平均數及標準離差，以及所有變數的有效值數目。家庭收入（以千為單位）、現址居住年數和婚姻狀況依序包含最多的遺漏值。

## 形式

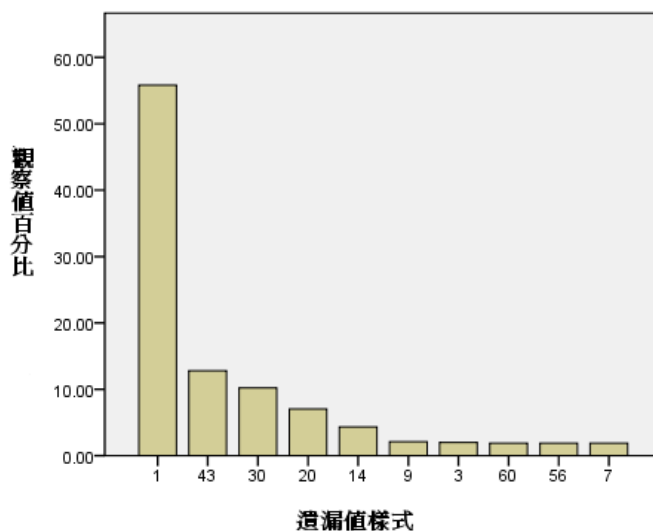
圖表 5-4  
遺漏值形式

形式圖表會顯示分析變數的遺失值形式。各形式會對應到具有相同形式的不完整與完整資料的一組觀察值。例如，形式 1 代表不包含遺漏值的觀察值、形式 33 代表在 reside (家庭人數) 和 address (現址居住年數) 中包含遺漏值的觀察值，而形式 66 代表在 gender (性別)、marital (婚姻狀況)、address 和 income (家庭收入以千為單位) 中包含遺漏值的觀察值。資料集可能會包含兩種變數個數形式。在 10 個分析變數中，形式為  $2^{10}=1024$ ，但在資料集中，1000 個觀察值內只出現了 66 種形式。

圖表順序分析變數和形式可顯示單調性出現的位置。而且變數是以從左到右遞增遺漏值的方式排列順序。然後以最後一個變數來排序形式 (從非遺漏值開始，然後才輪到遺漏值)，接著是從第二個變數到最後一個變數，依此類推從右到左進行。這麼做可以顯示可否在您的資料上使用單調插補方法，以及若不行的話，可得知您的資料與單調形式的近似程度。如果資料為單調資料，則圖表中所有的遺漏和非遺漏儲存格為連續儲存格，意即圖表的右下側不會出現零散分佈的非遺漏儲存格，左上側也不會出現零散分佈的遺漏儲存格。

此資料集不是單調形式資料集，為了讓資料集包含單調性，您必須插補許多的值。

圖表 5-5  
形式次數



當您想要使用形式時，副長條圖會顯示各個形式的觀察值百分比。這會顯示資料集中超過半數的觀察值具有形式 1，而遺漏值形式圖表顯示出這就是不具有遺漏值之觀察值的形式。形式 43 代表在 income 中包含遺漏值的觀察值、形式 30 代表在 address 中包含遺漏值的觀察值，而形式 20 則代表在 marital 中包含遺漏值的觀察值。大部分的觀察值（大約是 5 分之 4 的比例）會由這四種形式表示。形式 14、形式 60 和形式 56 是 10 種最常出現的形式中，唯一能夠代表在一個以上變數中包含遺漏值的觀察值。

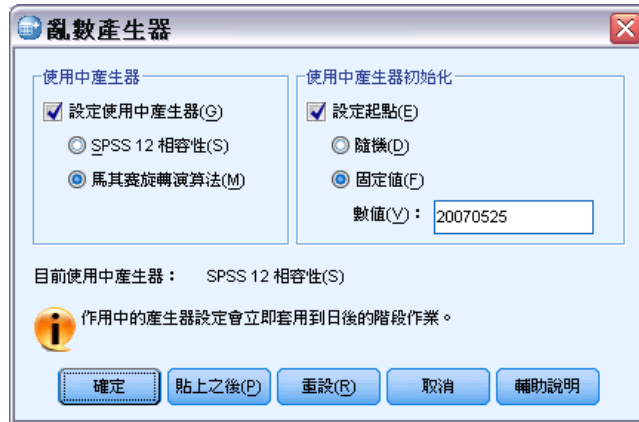
遺漏值分析並未顯示任何可能造成多重插補進行困難的問題，但單調方法可能不太適用。

## 遺漏值自動插補

您現在就可以開始插補值的工作，首先我們將以自動設定進行工作，但在要求插補之前，我們必須先設定亂數種子。設定亂數種子可讓您複製出完全相同的分析。

- ▶ 若要設定亂數種子，從功能表選擇：  
轉換 > 亂數產生器...

圖表 5-6  
「亂數產生器」對話方塊



- ▶ 選擇「設定作用中產生器」。
- ▶ 選擇 Mersenne Twister。
- ▶ 選擇「設定起始點」。
- ▶ 選取「固定值」，再輸入「20070525」作為值。
- ▶ 按一下「確定」。
- ▶ 若要多重插補遺漏的資料值，請在功能表選擇：  
分析 > 多重插補 > 插補遺漏資料值...



圖表 5-7  
「插補遺漏資料值」對話方塊



- ▶ 選擇「服務期間 (以月為單位) [tenure]」到「家庭人數 [reside]」作為插補模式變數。
- ▶ 輸入「telcolimputed」作為資料集，並在此儲存插補資料。
- ▶ 按一下「輸出」索引標籤。

圖表 5-8  
「輸出」索引標籤



- ▶ 選擇「含插補值的變數之敘述統計」。
- ▶ 按一下「確定」。

## 插補規格

圖表 5-9  
插補規格

插補方法	完整條件規格
插補數量	5
尺度變數模型	線性迴歸
模型中包含的交互作用	(無)
遺漏值的最大百分比	100.0%

插補規格表十分實用，可讓您檢閱您要求的項目，以確認規格正確。

## 插補結果

圖表 5-10  
插補結果

插補方法	完整條件規格
完整條件規格方法疊代	10
依變數 已插補	tenure,age,marital,address,income,ed,employ,retire,gender,reside
未插補 (遺漏值太多)	
未插補 (沒有遺漏值)	
插補順序	age,tenure,reside,ed,gender,retire,employ,marital,address,income

插補結果可讓我們一覽插補過程中發生的所有動作。請特別注意下列幾點：

- 規格表中的插補方法是「Automatic」，而自動方法選項實際選擇的方法為「Fully Conditional Specification (完整條件規格)」。
- 已插補所有要求使用的變數。
- 插補序列的順序是以變數顯示在遺漏值形式圖表 x 軸上的順序而定。

## 插補模式

圖表 5-11  
插補模式

	模式		遺漏值	插補值
	類型	效果		
年紀	線性迴歸	ed,gender,retire,marital,tenure, reside,employ,address,income	25	125
服務月數	線性迴歸	ed,gender,retire,marital,age,reside,employ,address,income	32	160
家庭人數	線性迴歸	ed,gender,retire,marital,age,tenure,employ,address,income	34	170
教育程度	Logistic迴歸	gender,retire,marital,age,tenure, reside,employ,address,income	35	175
性別	Logistic迴歸	ed,retire,marital,age,tenure,reside,employ,address,income	42	210
退休	Logistic迴歸	ed,gender,marital,age,tenure, reside,employ,address,income	84	420
現職年數	線性迴歸	ed,gender,retire,marital,age,tenure, reside,address,income	96	480
婚姻狀態	Logistic迴歸	ed,gender,retire,age,tenure,reside,employ,address,income	115	575
現址年數	線性迴歸	ed,gender,retire,marital,age,tenure, reside,employ,income	150	750
家庭收入(千元)	線性迴歸	ed,gender,retire,marital,age,tenure, reside,employ,address	179	895

插補模式表可提供變數插補方式的詳細資訊。請特別注意下列幾點：

- 變數會以插補序列的順序列出。
- 尺度變數會使用線性迴歸設為模型，類別變數則會使用 Logistic 迴歸設為模型。
- 所有模型都會使用其他變數作為主效果。
- 回報各個變數的遺漏值數目，以及變數內插補之值的總數（遺漏個數 × 插補數目）。

## 敘述統計

圖表 5-12  
Tenure 的敘述統計 (服務期間以月為單位)

資料	插補	個數	平均數	標準離差	最小值	最大值
原始資料		968	35.56	21.268	1.00	72.00
插補值	1	32	36.06	24.218	-6.72	90.02
	2	32	37.64	22.229	-.19	88.03
	3	32	30.82	27.245	-40.99	104.77
	4	32	39.97	20.585	1.29	80.50
	5	32	37.87	20.669	3.44	94.21
插補後完成資料	1	1000	35.58	21.355	-6.72	90.02
	2	1000	35.63	21.291	-.19	88.03
	3	1000	35.41	21.484	-40.99	104.77
	4	1000	35.70	21.251	1.00	80.50
	5	1000	35.63	21.243	1.00	94.21

敘述統計表會顯示包含插補值的變數摘要。每個變數都會產生一個不同的表格。依據變數為尺度變數或類別尺度而定，會顯示不同的統計類型。

尺度變數的統計包括個數、平均數、標準差、最小值和最大值，為原始資料、每組插補值和完整資料集（合併原始資料和插補值）而顯示。

tenure（服務期間以月為單位）敘述統計表會顯示每組插補值的平均數和標準差，此數字和原始資料幾乎完全相等，但當您在檢視最小值的同時發現已經在 tenure 中插補了負值時，您就立刻知道發生錯誤了。

圖表 5-13  
Marital 的敘述統計 (婚姻狀況)

資料	插補	類別	個數	百分比
原始資料	0		456	51.5
	1		429	48.5
插補值	1	0	51	44.3
		1	64	55.7
	2	0	41	35.7
		1	74	64.3
	3	0	49	42.6
		1	66	57.4
	4	0	43	37.4
		1	72	62.6
	5	0	53	46.1
		1	62	53.9
插補後完成資料	1	0	507	50.7
		1	493	49.3
	2	0	497	49.7
		1	503	50.3
	3	0	505	50.5
		1	495	49.5
	4	0	499	49.9
		1	501	50.1
	5	0	509	50.9
		1	491	49.1

在類別變數中，統計包含了個數，以及原始資料、插補值及完整資料的百分比（依類別分類）。marital（婚姻狀況）會出現有趣的結果，在此結果中，插補值中估計出的已婚觀察值數目大於原始資料中的數目。此結果的原因可能是隨機變異，也可能是因為遺漏機會與此變數值相關。

圖表 5-14  
Income 的敘述統計 (家庭收入以千為單位)

資料	插補	個數	平均數	標準離差	最小值	最大值
原始資料		821	71.1462	83.14424	9.0000	944.0000
插補值	1	179	87.6574	91.13179	-189.1959	373.2412
	2	179	101.6724	94.20599	-122.0010	346.4294
	3	179	100.9445	95.00789	-127.8572	342.5208
	4	179	107.0787	90.23638	-113.0959	369.9674
	5	179	101.1043	90.40865	-167.6978	314.2533
插補後完成資料	1	1000	74.1017	84.81851	-189.1959	944.0000
	2	1000	76.6104	85.98067	-122.0010	944.0000
	3	1000	76.4801	86.10024	-127.8572	944.0000
	4	1000	77.5781	85.52821	-113.0959	944.0000
	5	1000	76.5087	85.22154	-167.6978	944.0000

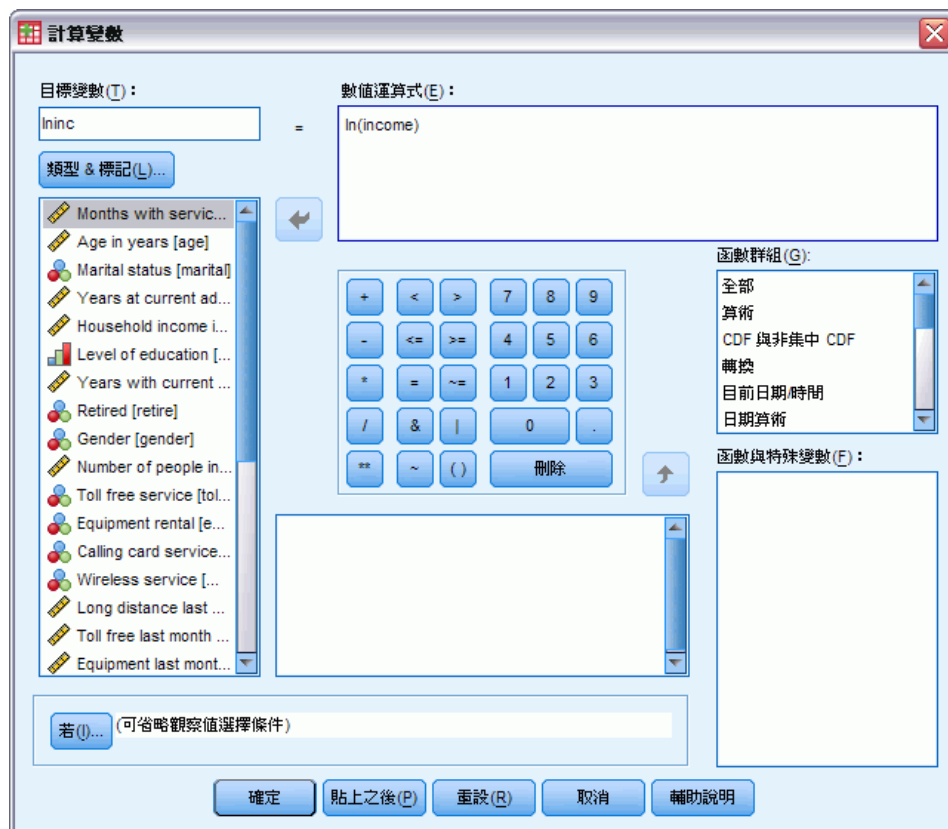
和 tenure 及其他所有尺度變數相同，income（家庭收入以千為單位）會詳細的顯示負插補值——因此，我們必須執行在某些變數中包含限制的自訂模式。但是，income 會顯示其他的潛在問題。所有插補的平均值皆會比原始資料更高，最大值則會比原始資料低很多。收入的分配比例會明顯偏右，這可能就是錯誤的原因。

## 自定插補模式

為了避免插補值落至變數的合理值範圍外，我們必須指定在變數中包含限制的自訂插補模式。除此之外，家庭收入以千為單位會明顯偏右，其他的深入分析很有可能會使用 income 對數，因此直接插補  $\log$ -income 是合理的處理方式。

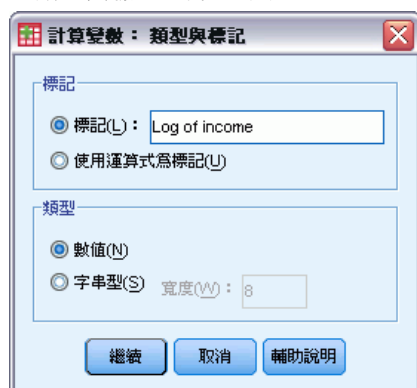
- ▶ 確保原始資料集保持作用中
- ▶ 若要建立  $\log$ -income 變數，請從功能表選擇：  
轉換 > 計算變數...

圖表 5-15  
「計算變數」對話方塊



- ▶ 輸入 lninc 作為目標變數。
- ▶ 輸入 ln(income) 作為數值運算式。
- ▶ 按一下「類型 & 標記」..

圖表 5-16  
「類型與標記」對話方塊



- ▶ 輸入 Log of income 作為標記。
- ▶ 按一下「繼續」。
- ▶ 在「計算變數」對話方塊中，按一下「確定」。

圖表 5-17

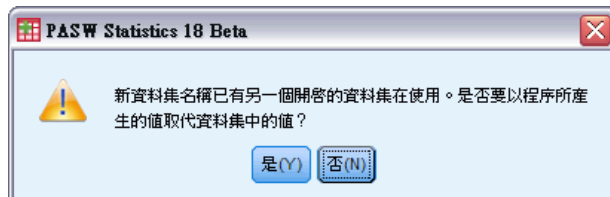
「變數」索引標籤中包含要在插補模式中用來取代「家庭收入以千為單位」的收入記錄。



- ▶ 叫回「插補遺漏資料值」對話方塊，然後按一下「變數」索引標籤。
- ▶ 取消選擇「家庭收入以千為單位 [income]」，然後選擇「收入記錄 [lninc]」作為模式的變數。
- ▶ 按一下「方法」索引標籤。

圖表 5-18

取代現有的資料集時出現警示



- ▶ 在顯示的警示方塊中按一下「是」。



圖表 5-19  
「方法」索引標籤

The screenshot shows the 'Impute Missing Data Values' dialog box with the 'Method' tab selected. The 'Imputation Method' section has three radio buttons: 'Automatic (A)', 'Custom (C)', and 'Single (M)'. 'Custom (C)' is selected, and under it, 'Full Condition Specification (MCMC) (E)' is selected. The 'Maximum Iterations (X)' is set to 10. Below this, there is a checkbox for 'Include Two-Factor Interaction in Categorical Prediction (N)' which is unchecked. The 'Scale Variable Model Type (D)' is set to 'Linear Regression' and the 'Single Error (S)' is set to '1E-012'. At the bottom, there are buttons for 'OK', 'Apply (P)', 'Reset (R)', 'Cancel', and 'Help'.

- ▶ 選擇「自訂」，然後讓插補方式維持為「完整條件規格」。
- ▶ 按一下「限制」索引標籤。

圖表 5-20  
「限制」索引標籤

插補遺漏資料值

變數 方法 限制 輸出

變數摘要的資料掃描

重新掃描資料(S)  限制掃描的觀察值數目(L) 觀察值(C): 5000

變數摘要(V):

模式中的變數	百分比遺漏	最小觀察值	最大觀察值
tenure	3.20	1	72
age	2.50	18	77
marital	11.50	0	1
...	...	...	...

已掃描的觀察值: 1000

定義限制(D):

模式中的變數	角色	最小值	最大值	捨入
gender	插補做為預測			
reside	插補做為預測	1		1
lnc	插補做為預測	0		

排除含有大量遺漏值的變數(E)

最大百分比遺漏(M):

最大觀察值取樣(A): 50

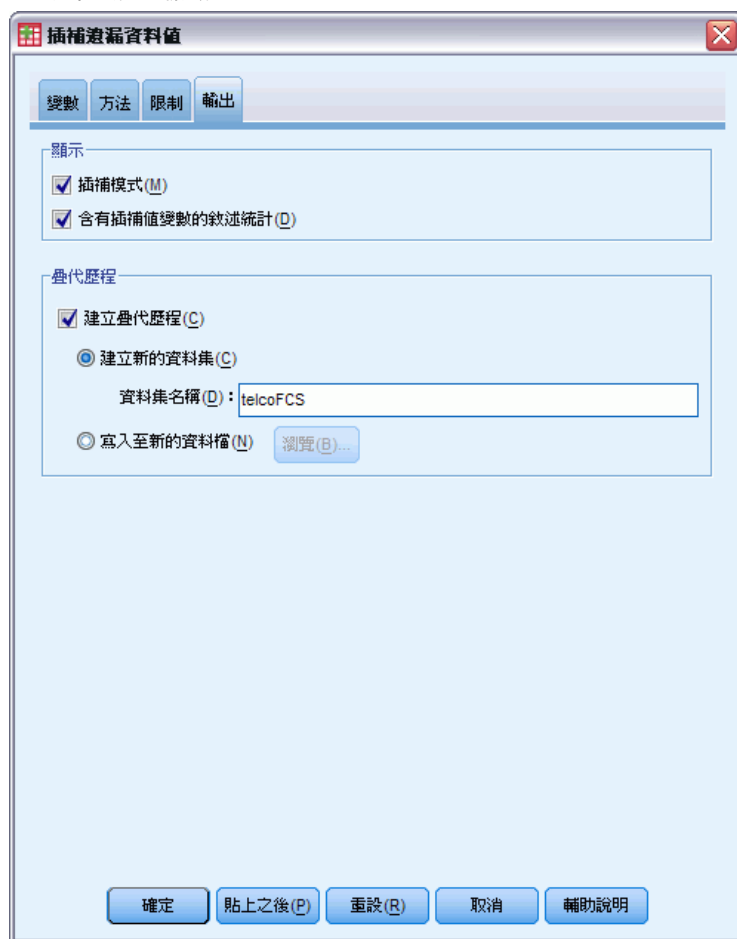
最大參數取樣(X): 2

增加最大參數取樣可明顯增加分析時間。

確定 貼上之後(P) 重設(R) 取消 輔助說明

- ▶ 按一下「掃描資料」。
- ▶ 在「定義限制」網格中，輸入 1 作為「服務期間（以月為單位）[tenure]」的最小值。
- ▶ 輸入 18 作為「age（以年為單位）」的最小值。
- ▶ 輸入 0 作為「address（現址居住年數）」的最小值。
- ▶ 輸入 0 作為「employ（服務於目前雇主的年數）」的最小值。
- ▶ 輸入 1 作為最小值，並以 1 作為「reside（家庭人數）」的四捨五入標準。請注意，當有多個其他尺度變數在整數值中回報時，您可以合理地推論某人已經在居住地現址住了 13.8 年，但其人數不會是 2.2 之類的小數數字。
- ▶ 輸入 0 作為「lnc（收入記錄）」的最小值。
- ▶ 按一下「輸出」索引標籤。

圖表 5-21  
「輸出」索引標籤



- ▶ 選擇「建立疊代歷程」，然後輸入 telcoFCS 作為新資料集的名稱。
- ▶ 按一下「確定」。

## 插補限制

圖表 5-22  
插補限制

	插補中角色		插補值		
	依變數	預測變數	最小值	最大值	捨入
服務月數	是	是	1	(無)	
年紀	是	是	18	(無)	
婚姻狀態	是	是			
現址年數	是	是	0	(無)	
教育程度	是	是			
現職年數	是	是	0	(無)	
退休	是	是			
性別	是	是			
家庭人數	是	是	1	(無)	整數
Log of income	是	是	0	(無)	

自訂插補模式會產生新表格，此新表格可檢閱此插補模式上的限制。出現的所有項目都與您指定的規格相符。

## 敘述統計

圖表 5-23  
Tenure 的敘述統計 (服務期間以月為單位)

資料	插補	個數	平均數	標準離差	最小值	最大值
原始資料		968	35.56	21.268	1.00	72.00
插補值	1	32	36.53	23.000	1.12	89.97
	2	32	41.53	27.370	1.24	126.01
	3	32	39.10	19.183	14.73	93.25
	4	32	40.17	22.113	7.78	89.33
	5	32	41.34	23.936	3.05	87.90
插補後完成資料	1	1000	35.59	21.314	1.00	89.97
	2	1000	35.75	21.499	1.00	126.01
	3	1000	35.67	21.205	1.00	93.25
	4	1000	35.71	21.300	1.00	89.33
	5	1000	35.75	21.370	1.00	87.90

位在包含限制的自訂插補模式下的「tenure (服務期間以月為單位)」敘述統計表格會顯示「tenure」的負插補值問題已解決。

圖表 5-24  
Marital 的敘述統計 (婚姻狀況)

資料	插補	類別	個數	百分比
原始資料		0	456	51.5
		1	429	48.5
插補值	1	0	56	48.7
		1	59	51.3
	2	0	41	35.7
		1	74	64.3
	3	0	52	45.2
		1	63	54.8
	4	0	51	44.3
		1	64	55.7
	5	0	49	42.6
		1	66	57.4
插補後完成資料	1	0	512	51.2
		1	488	48.8
	2	0	497	49.7
		1	503	50.3
	3	0	508	50.8
		1	492	49.2
	4	0	507	50.7
		1	493	49.3
	5	0	505	50.5
		1	495	49.5

「marital (婚姻狀況)」表格內包含一個插補 (3)，此插補的分配明顯與原始資料對齊，但大部分的插補分配顯示被估計為「已婚」的觀察值數目大於原始資料中的數目。原因可能是隨機變異，但我們需要對資料進行更深入的研究，以判斷這些值是否隨機遺漏 (MAR)。所以我們先就此打住。

圖表 5-25  
lninc 的敘述統計 (收入記錄)

資料	插補	個數	平均數	標準離差	最小值	最大值
原始資料		821	3.9291	.75305	2.1972	6.8501
插補值	1	179	4.1101	.90290	.9481	6.1834
	2	179	4.1957	.93415	2.1554	7.0682
	3	179	4.1496	.94692	1.3730	6.7174
	4	179	4.1591	.87912	2.0418	6.4203
	5	179	4.2245	.90974	1.9923	6.6590
插補後完成資料	1	1000	3.9615	.78457	.9481	6.8501
	2	1000	3.9768	.79462	2.1554	7.0682
	3	1000	3.9685	.79523	1.3730	6.8501
	4	1000	3.9702	.78164	2.0418	6.8501
	5	1000	3.9819	.79106	1.9923	6.8501

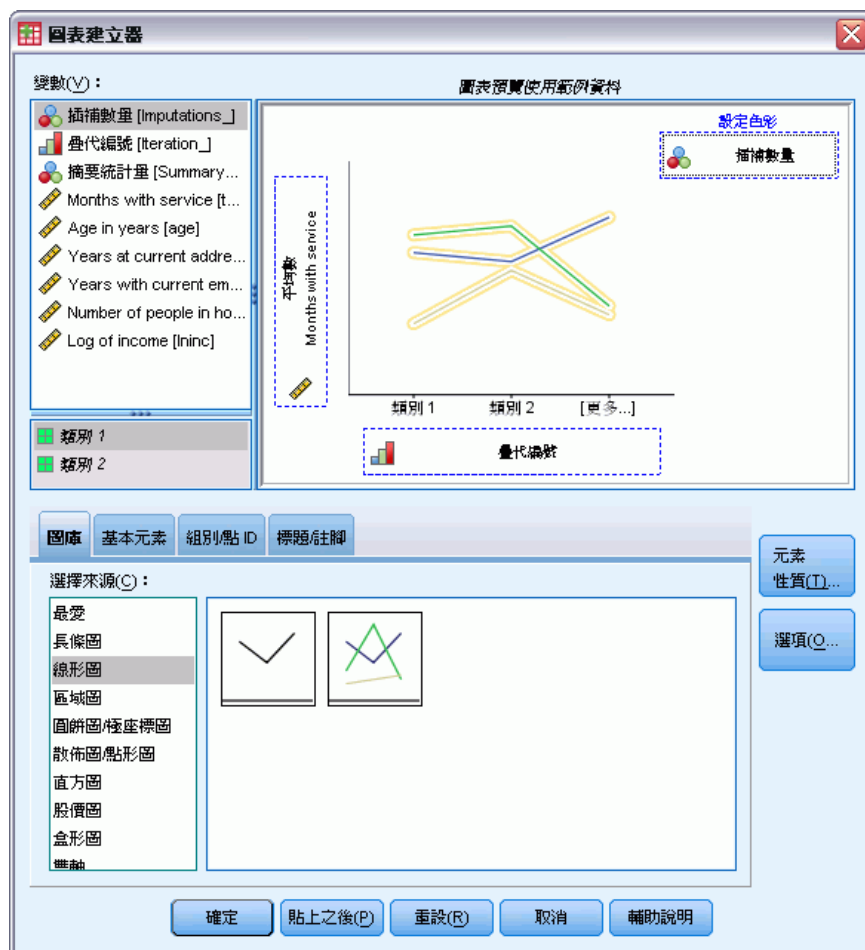
和「tenure」及其他的尺度變數一樣，「lninc (收入記錄)」無法顯示負插補值。另外，插補的平均值較近似於原始資料的平均數，接近程度大於「income」尺度中執行的自動插補一，「lninc」原始資料的平均數大約是  $e^{3.9291}=50.86$ ，而一般插補的平均值大概是  $e^{4.2}=66.69$ 。各個插補的最大值則較為接近原始資料的最大值。

## 檢查 FCS 收斂

使用完全條件規格方法時，建議您使用疊代檢查平均數及標準差的圖形，然後檢查各個尺度依變數（包含插補值）中的插補，以便您存取模式收斂。

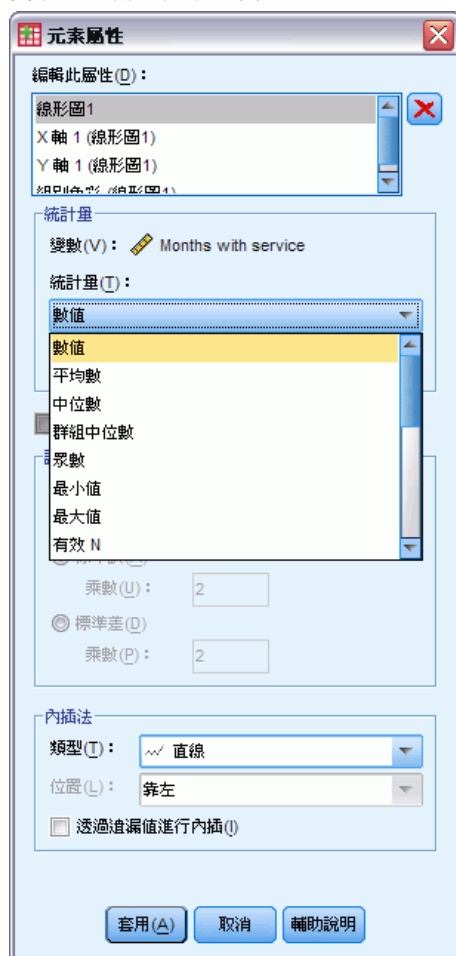
- ▶ 若要建立這種類型的圖表，請啟用 telcoFCS 資料集，然後在功能表中選擇以下選項：  
統計圖 > 圖表建立器...

圖表 5-26  
圖表建立器，多線圖形



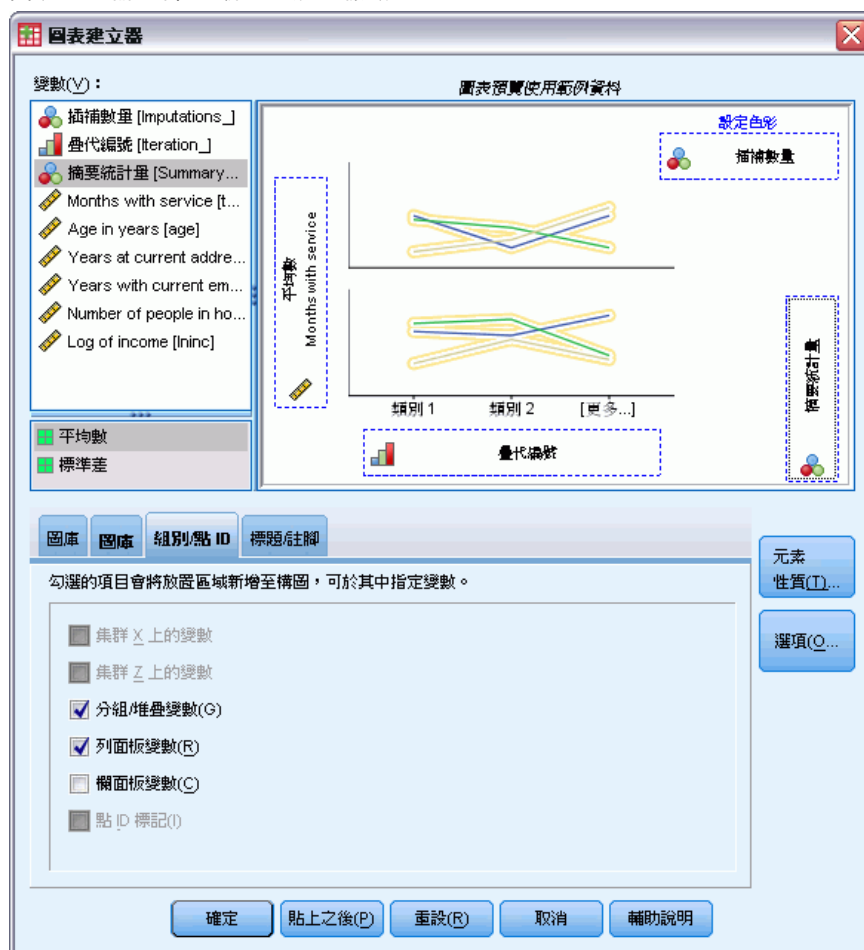
- ▶ 選擇「線形圖」圖庫，然後選擇「多線圖形」。
- ▶ 選擇「服務期間以月為單位 [tenure]」作為圖形在 Y 軸上的變數。
- ▶ 選擇「疊代個數 [Iteration\_]」作為圖形在 X 軸上的變數。
- ▶ 選擇「插補數 [Imputations\_]」作為設定顏色依據的變數。

圖表 5-27  
圖表建立器，元素性質



- ▶ 在「元素性質」中，選擇「Value」作為您要顯示的統計量。
- ▶ 按一下「套用」。
- ▶ 在「圖表建立器」中，按一下「群組/點 ID」索引標籤。

圖表 5-28  
圖表建立器，群組/點 ID 索引標籤

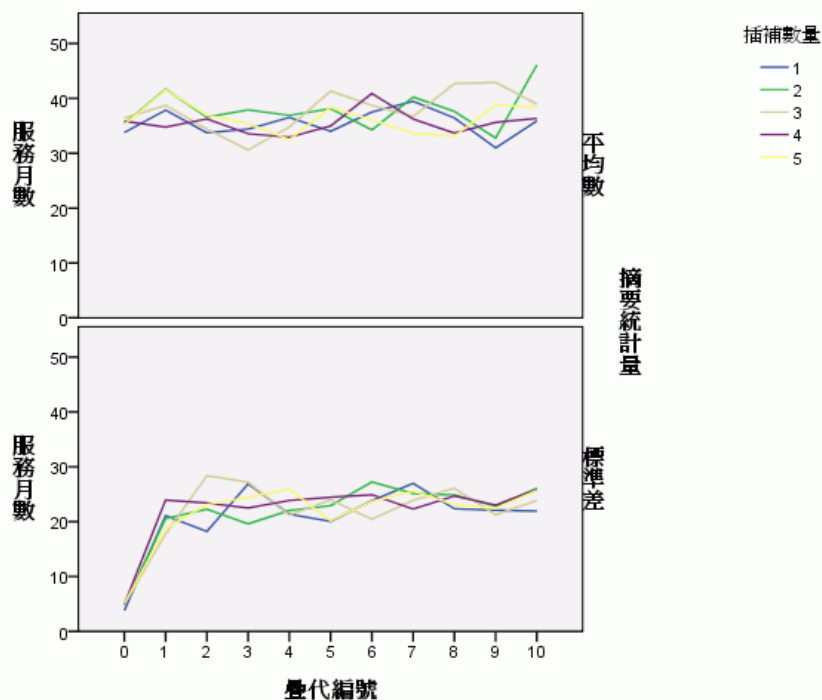


- ▶ 選擇「列面板變數」。
- ▶ 選擇「摘要統計量 [SummaryStatistic\_]」作為面板變數。
- ▶ 按一下「確定」。



## FCS 收斂圖表

圖表 5-29  
FCS 收斂圖表



您已經建立了一組多線圖，圖中顯示五個要求的插補分別在其 FCS 插補方法的疊代中，顯示「服務期間以月為單位 [tenure]」的插補平均數及標準差。此圖形的目的是為了找出線的模式。實際上不應該有任何模式，而這個圖形中的線條也符合「隨機」的要求。您可以為其他尺度變數建立類似的圖形，請注意，這些圖形也可顯示出不明顯的模式。

## 分析完整資料

您插補的值顯示令人滿意的結果，接下來，您就可以準備對「完整」資料執行分析。此資料集包含客戶類別 [custcat] 變數，此變數依照服務使用方式來切割客戶數量，並將客戶分為四個群組。如果您能夠使用人口資訊來預測組別成員資訊，並套入模式，您就可以為個別自訂潛在客戶自訂報價。

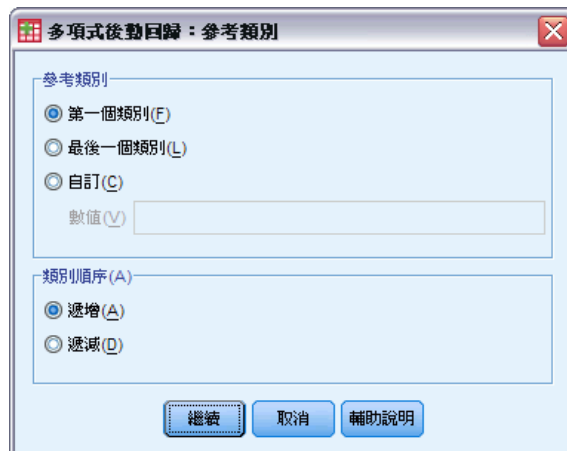
- ▶ 啟用 telcoImputed 資料集。如果想要建立完整資料的多項式 Logistic 迴歸模式，請在功能表中選擇下列選項：  
分析 > 迴歸方法 > 多項式 Logistic...

圖表 5-30  
「多項式 Logistic 迴歸」對話方塊



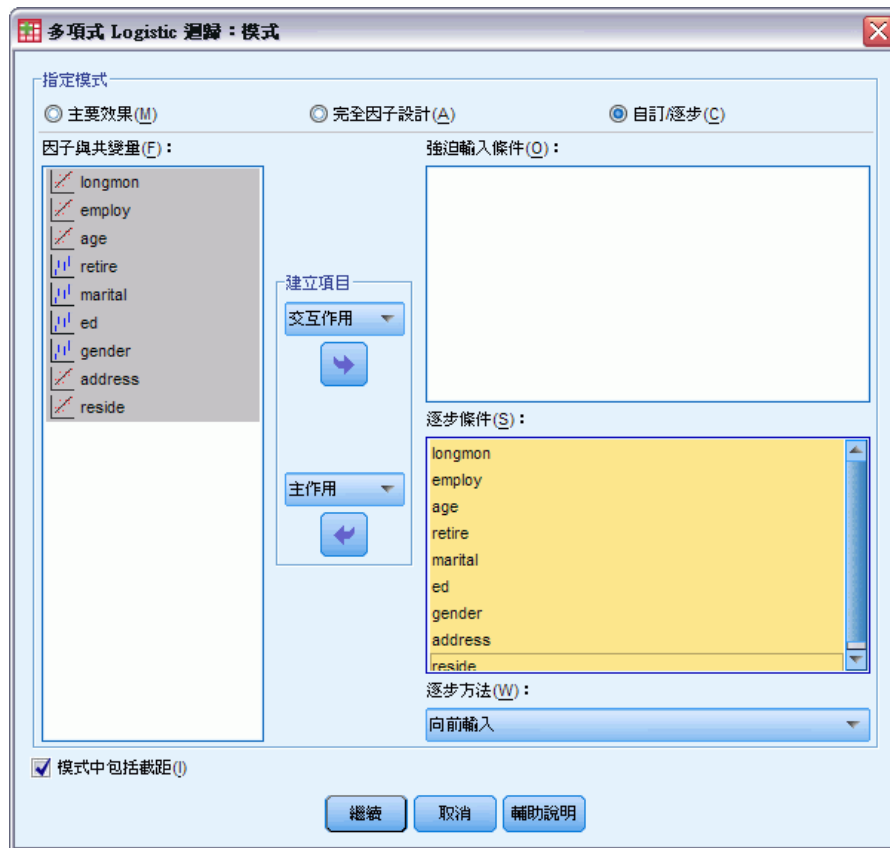
- ▶ 選擇「客戶類別」作為依變數。
- ▶ 選擇「婚姻狀況」、「教育程度」、「退休」和「性別」作為因子。
- ▶ 選擇「年齡（年為單位）」、「現址居住年數」、「服務於目前雇主的年數」、「家庭人數」和「收入記錄」作為共變數。
- ▶ 您要將其他客戶與基本服務訂戶比較，所以選擇「客戶類別」並按一下「參考類別」。

圖表 5-31  
「參考類別」對話方塊



- ▶ 選擇「第一個類別」。
- ▶ 按一下「繼續」。
- ▶ 在「多項式 Logistic 迴歸」對話方塊中，按一下「模式」。

圖表 5-32  
「模式」對話方塊

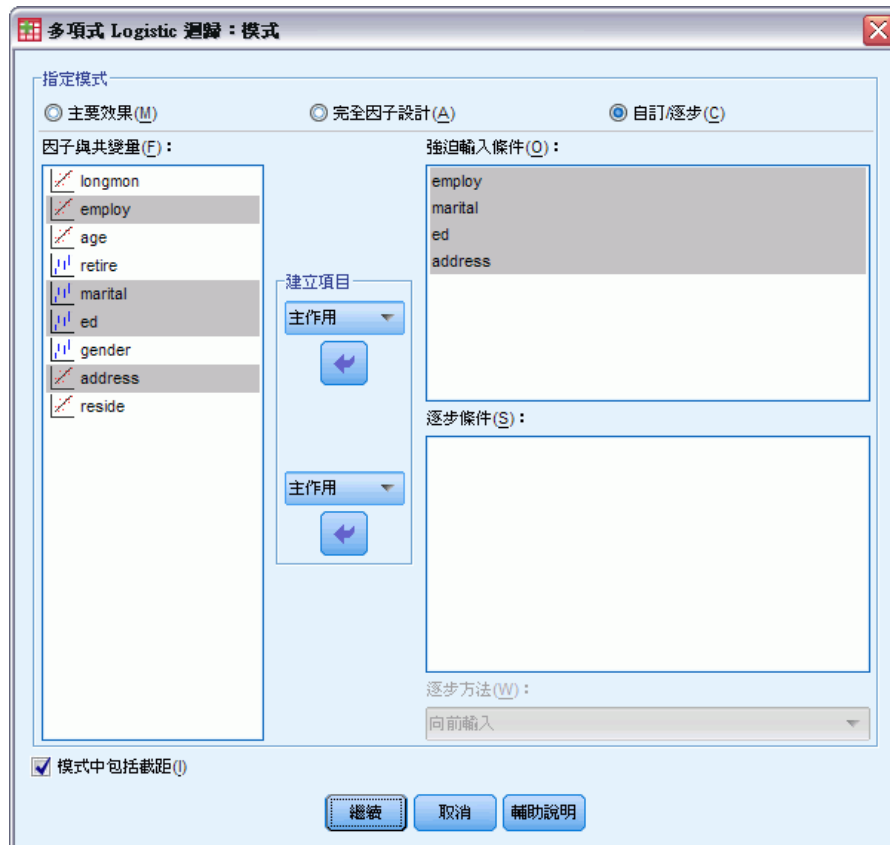


- ▶ 選擇「自訂/逐步」。
- ▶ 從「逐步模式項目 - 建立效果項」下拉清單中，選擇「主效果」。
- ▶ 選擇「lnc」到「reside」作為逐步模式項目。
- ▶ 按一下「繼續」。
- ▶ 在「多項式 Logistic 迴歸」對話方塊中，按一下「確定」。



## 使用預測值子集來執行模式

圖表 5-34  
「模式」對話方塊



- ▶ 叫回「多項式 Logistic 迴歸」對話方塊，並按一下「模式」。
- ▶ 在「逐步模式項目」清單中取消選擇變數。
- ▶ 從「強制選入項目 - 建立效果項」下拉清單中，選擇「主效果」。
- ▶ 選擇「employ」、「marital」、「ed」和「address」作為強制選入項目。
- ▶ 按一下「繼續」。
- ▶ 在「多項式 Logistic 迴歸」對話方塊中，按一下「確定」。

## 合併參數估計值

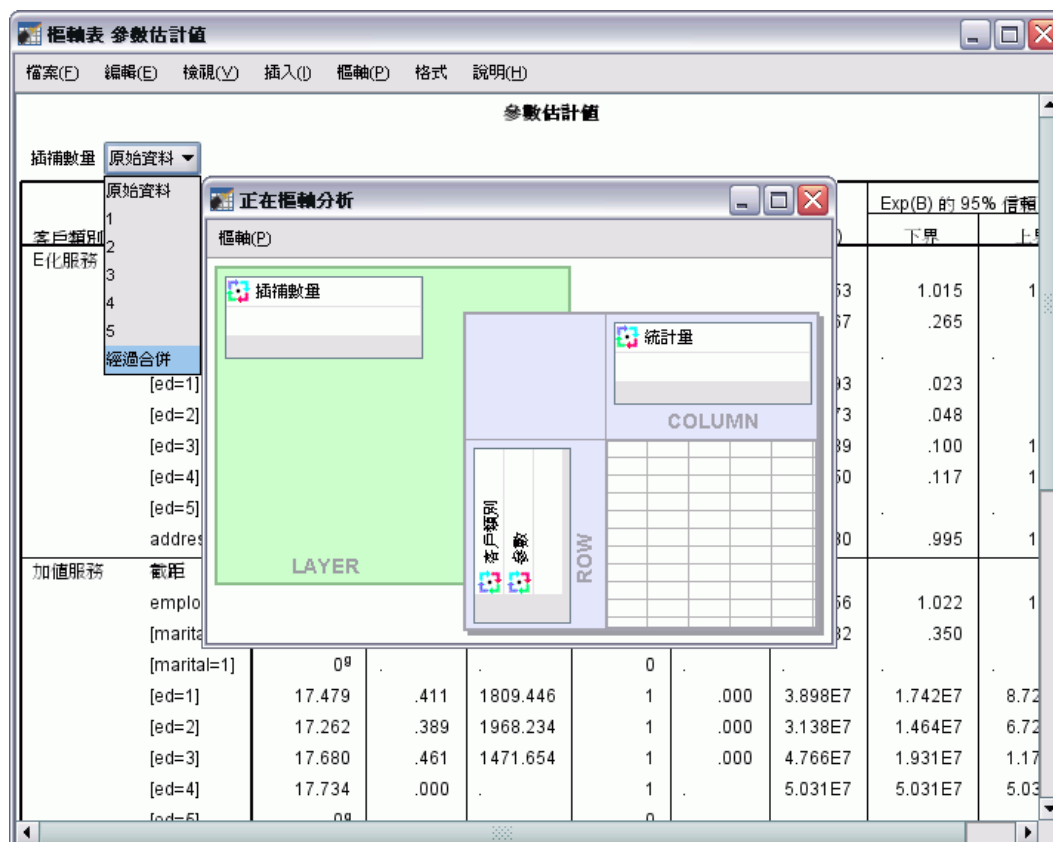
這個表格很大，但樞軸可以提供我們幾個不同的輸出有效檢視方式。

圖表 5-35  
合併參數估計值

		平方和	Wald	df	顯著性	Exp(B)	Exp(B) 的 95% 下界
原始資料	E化服務	.640	1.630	1	.202		
		.018	7.819	1	.005	1.053	1.015
		.289	6.914	1	.009	.467	.265
		.708	11.263	1	.001	.093	.023
		.656	7.161	1	.007	.173	.048
		.692	1.857	1	.173	.389	.100
		.687	1.350	1	.245	.450	.117
				0			
		.018	2.784	1	.095	1.030	.995
加值服務		.381	2171.237	1	.000		
		.016	11.037	1	.001	1.056	1.022
		.259	4.368	1	.037	.582	.350
				0			
		.411	1809.446	1	.000	3.898E7	1.742E7
		.389	1968.234	1	.000	3.138E7	1.464E7
		.461	1471.654	1	.000	4.766E7	1.931E7
		.000		1		5.031E7	5.031E7
				0			
	address	.022	.016	1	.167	1.023	.991
全方位服務	截距	1.528	.601	1	.011		

- ▶ 啟用（連按兩下）表格，然後在「內容」功能表中選擇「樞軸分析匣」。

圖表 5-36  
合併參數估計值



- ▶ 將「插補數」從「列」移動至「階層」。
- ▶ 在「插補數」下拉清單中選擇「合併」。

圖表 5-37  
合併參數估計值

客戶類別 <sup>a,b,c,d,e,f</sup>	B 之估計值	標準誤差	顯著性	Exp(B)	Exp(B) 的 95% 信賴區間		分數遺漏資訊	相對增量變異數	相對效能
					下界	上界			
E化服務	截距	.452	.487	.354			.118	.126	.977
	employ	.036	.014	.011	1.036	1.008	1.065	.102	.980
	[marital=0]	-.677	.210	.001	.508	.337	.768	.015	.997
	[marital=1]	0 <sup>g</sup>							
	[ed=1]	-1.894	.519	.000	.151	.054	.416	.049	.990
	[ed=2]	-1.098	.492	.026	.334	.127	.877	.083	.984
	[ed=3]	-.531	.502	.290	.588	.220	1.573	.078	.985
	[ed=4]	-.388	.504	.442	.678	.252	1.828	.099	.981
	[ed=5]	0 <sup>g</sup>							
address	.028	.014	.049	1.029	1.000	1.058	.190	.216	.963
加值服務	截距	-1.099	.707	.121			.144	.157	.972
	employ	.049	.012	.000	1.051	1.026	1.076	.028	.994
	[marital=0]	-.406	.200	.043	.666	.450	.987	.036	.993
	[marital=1]	0 <sup>g</sup>							
	[ed=1]	.654	.723	.367	1.923	.462	8.001	.156	.172
	[ed=2]	.706	.717	.326	2.027	.493	8.332	.155	.171
	[ed=3]	.762	.734	.300	2.143	.504	9.112	.159	.176
	[ed=4]	.672	.729	.358	1.958	.466	8.230	.136	.148
	[ed=5]	0 <sup>g</sup>							
address	.014	.012	.242	1.014	.990	1.039	.036	.037	.993
全方位服務	截距	1.097	.451	.015			.093	.098	.982
	employ	.041	.015	.005	1.041	1.012	1.072	.100	.106
	[marital=0]	-.664	.217	.002	.515	.336	.788	.057	.989
	[marital=1]	0 <sup>g</sup>							
	[ed=1]	-3.354	.564	.000	.035	.012	.106	.034	.034
	[ed=2]	-1.731	.471	.000	.177	.070	.447	.097	.103
	[ed=3]	-1.184	.479	.014	.306	.120	.784	.087	.091
	[ed=4]	-.437	.457	.339	.646	.263	1.583	.059	.061
	[ed=5]	0 <sup>g</sup>							
address	.008	.015	.565	1.008	.980	1.038	.142	.155	.972

這個檢視可表示出合併結果的所有統計量。您可以在不包含遺漏值的資料集中使用此表格的方式來使用並解讀這些係數。

參數估計值的表格會將各個預測值的效果做成摘要。係數對其標準誤的比率，平方後等於 Wald 統計量。如果 Wald 統計的顯著性水準較小（小於 0.05），則參數與 0 不同。

- 含有顯著負係數的參數會降低與參考類別相關之反應類別的概似度。
- 含正係數的參數會增加與反應類別的概似度。
- 如果有截距項，則與每個因子最後一個類別相關的參數為多餘。

這三個欄位是表格中的其他欄位，其中包含有關合併輸出的詳細資訊。**遺漏資訊的分數**是遺漏資訊對「完整」資訊的比例估計值，其基礎為**變異數相對增量**（因為無回應），換句話說，它也是在迴歸係數的插補變異數內，位於插補和平均之間的（修改後）比例。**相對效能**是對某一估計值及某一使用不限數目的插補計算出的理論估計值做出的比較。相對效能是由遺漏資訊的分數及用來取得合併結果的插補所決定，當遺漏資訊的分數值較大時，您就需要使用較大的插補數，以取得近似 1 的相對效能和接近理想估計值的合併估計值。



圖表 5-38  
合併參數估計值

圖表 5-38 顯示了 SPSS 軟體的「樞軸表 參數估計值」對話框。左側顯示了統計量列表，其中「B 之估計值」被選中。右側顯示了「正在樞軸分析」對話框，其中「統計量」被移動到「階層」(LAYER) 區域，而「插補數」被移動到「列」(COLUMN) 區域。

統計量	B 之估計值	標準誤差	Wald	df	顯著性	Exp(B)	Exp(B) 的 95% 信賴區間 下界	Exp(B) 的 95% 信賴區間 上界
客戶類別	[ed=2]	-1.754						
	[ed=3]	-.943						
	[ed=4]	-.798						
	[ed=5]	0 <sup>a</sup>						
	address	.030						
加值服務	截距	-17.768						
	employ	.054						
	[marital=0]	-.542						
	[marital=1]	0 <sup>a</sup>						
	[ed=1]	17.479	.380	.723	.925	.844	.397	.654
	[ed=2]	17.262	.453	.733	.997	.896	.453	.706
	[ed=3]	17.680	.486	.807	1.092	.908	.517	.762
[ed=4]	17.734	.486	.742	.927	.841	.362	.672	
[ed=5]	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>	0 <sup>a</sup>	

- ▶ 請重新啟用 (連按兩下) 表格，然後在「內容」功能表中選擇「樞軸分析匣」。
- ▶ 將「插補數」從「階層」移動至「列」。
- ▶ 將「統計量」從「列」移動至「階層」。
- ▶ 從「統計量」下拉清單中選擇「B」。

圖表 5-39  
合併的參數估計值、位於列中的插補數以及而位於階層中的統計量。

客戶類別 a,b,c,d,e,f		插補數量						經過合併
		原始資料	1	2	3	4	5	
E化服務	截距	.817	.244	.578	.387	.441	.610	.452
	employ	.051	.038	.039	.038	.034	.029	.036
	[marital=0]	-.760	-.653	-.685	-.681	-.709	-.654	-.677
	[marital=1]	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>
	[ed=1]	-2.376	-1.838	-1.992	-1.751	-1.893	-1.993	-1.894
	[ed=2]	-1.754	-1.004	-1.200	-.968	-1.056	-1.261	-1.098
	[ed=3]	-.943	-.436	-.637	-.393	-.506	-.684	-.531
	[ed=4]	-.798	-.197	-.463	-.299	-.424	-.559	-.388
	[ed=5]	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>
address	.030	.035	.023	.022	.030	.031	.028	
加值服務	截距	-17.768	-.901	-1.152	-1.407	-1.215	-.823	-1.099
	employ	.054	.052	.050	.048	.047	.050	.049
	[marital=0]	-.542	-.404	-.391	-.367	-.460	-.407	-.406
	[marital=1]	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>
	[ed=1]	17.479	.380	.723	.925	.844	.397	.654
	[ed=2]	17.262	.453	.733	.997	.896	.453	.706
	[ed=3]	17.680	.486	.807	1.092	.908	.517	.762
	[ed=4]	17.734	.486	.742	.927	.841	.362	.672
	[ed=5]	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>
address	.022	.016	.013	.017	.013	.012	.014	
全方位服務	截距	1.528	.947	1.224	1.013	1.083	1.219	1.097
	employ	.038	.042	.047	.037	.039	.038	.041
	[marital=0]	-.459	-.687	-.718	-.663	-.664	-.590	-.664
	[marital=1]	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>
	[ed=1]	-3.664	-3.364	-3.380	-3.280	-3.491	-3.253	-3.354
	[ed=2]	-2.305	-1.633	-1.896	-1.611	-1.667	-1.848	-1.731
	[ed=3]	-1.295	-1.119	-1.309	-1.023	-1.152	-1.315	-1.184
	[ed=4]	-.710	-.291	-.500	-.389	-.462	-.545	-.437
	[ed=5]	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>	0 <sup>9</sup>
address	.005	.015	.003	.010	.009	.005	.008	

此表格檢視可讓您於所有插補中比較值，並透過從插補至插補或對照原始資料的方式，迅速地檢視迴歸係數估計值中的變異。您更可以將統計量從階層切換至標準。您也可以錯誤部分中，了解多重插補如何減低係數估計值對刪除全部遺漏值（原始資料）中的變化性。

圖表 5-40  
警告

下列變數: retire, gender, tenure, age, income, reside, lninc 只被用來定義次母群體，而非用來建構模式。

對於分割檔案 插補數量 = 原始資料，Hessian 矩陣中發生非預期的異常狀況。這表示若非某些預測變數應該被排除在外，便是某些類別應該被合併。

NOMREG 繼續不理會上述警告。隨後顯示的結果是以最後一個疊代為準。不確定模型配適的有效性。

然而，這個範例中的原始資料集產生錯誤，這就解釋了「加值服務」截距中出現極大的參數估計值，以及表格的原始資料列中出現非重複「ed（教育程度）」階層的原因。

## 摘要

藉由使用多重插補程序，您可以分析遺漏值的模式，並發現使用簡式刪除全部遺漏值時可能造成遺失許多資訊的後果。在一開始就執行自動多重插補，您會發現您需要使用限制以在合理範圍內保留插補值。使用限制來執行工作可產生正確的值，另外也沒有確切證據證實 FCS 方法無法收斂。使用包含多重插補值的「完整」資料集可讓您將多項式 Logistic 迴歸套用至資料以及取得合併迴歸估計值，並了解若是在原始資料中使用刪除全部遺漏值，就無法產生最終的模式適合度。

# 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。
- **aflatoxin20.sav**。這個資料檔包含由 aflatoxin.sav 取得，來自 4 和 8 作物產量的 16 個樣本，每一個樣本的黃麴毒素測量。
- **anorectic.sav**。在將厭食/暴食行為症狀學標準化的過程中，研究人員研究了 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。
- **autoaccidents.sav**。這是有關一位保險分析師致力於為每個駕駛的汽車意外事件次數建立模式，同時考量駕駛的年齡和性別的假設資料檔。每一個觀察值代表一位不同的駕駛，記錄了駕駛的性別、年齡、和近五年內的汽車意外事故次數。
- **band.sav**。本資料檔包含某樂團音樂 CD 假設性的每週銷售數字。也包含三個可能預測變數的資料。
- **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，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。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。
- **creditpromo.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。
- **customers\_model.sav**。本檔案包含一市場行銷活動所鎖定之個人的假設資料。這些資料包含人口資訊、購買歷史摘要、和每一個人是否對該活動有回應。每一個觀察值代表一位不同的個人。
- **customers\_new.sav**。本檔案包含一市場行銷活動潛在候選人之個人的假設資料。這些資料包含每一位個人的人口資訊和購買歷史摘要。每一個觀察值代表一位不同的個人。
- **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」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 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-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 以此表說明可能的社會團體。Guttman 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 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，、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。
- **polishing.sav**。這是取自「資料和故事圖書館」的「Nambeware 打磨時間」資料檔。它是有關一家金屬餐具製造商 (Namb Mills, 聖塔非, 新墨西哥州) 致力於規畫其生產排程。每一個觀察值代表生產線上一個不同的產品。每一個產品都記錄下直徑、打磨時間、價格、和產品類別。
- **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。
- **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。
- **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。它已在別處出現和分析過。
- **verd1985.sav**。本資料檔有關於一項調查。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 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。

---

# Notices

Licensed Materials - Property of SPSS Inc., an IBM Company. © Copyright SPSS Inc. 1989, 2010.

Patent No. 7,023,453

**The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:** SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

**COPYRIGHT LICENSE:**

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

**Trademarks**

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

This product uses WinWrap Basic, Copyright 1993–2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

Other product and service names might be trademarks of IBM, SPSS, or other companies.

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.



- EM(E)
  - 平均數, 8
- FCS 收斂圖表
  - 多重插補中, 65
- legal notices, 84
- Little' s MCAR 檢定, 7
  - 平均數, 2, 42
- MCAR 檢定
  - 平均數, 2, 42
- Student' s t 檢定
  - 平均數, 35
- t 檢定
  - 平均數, 6, 35
- trademarks, 85
  
- 分析樣式, 13
  
- 合併估計值
  - 多重插補中, 70
- 合併結果
  - 多重插補中, 65
- 單調插補
  - 多重插補中, 17
- 單變量統計量
  - 平均數, 34
  
- 在遺漏值分析中
  - 平均數, 2, 4, 6, 8-9, 36
  
- 多個插補, 22, 25
- 多重插補, 12, 43
  - FCS 收斂圖表, 65
  - 分析形式, 13
  - 合併估計值, 70
  - 合併結果, 65
  - 插補結果, 51
  - 插補規格, 50
  - 插補遺漏資料值, 15
  - 敘述統計, 52, 60
  - 模式, 51
  - 變數摘要, 45
  - 選項, 29
  - 遺漏值形式, 46
  - 遺漏值的整體摘要, 44
  - 限制, 60
  
- 完整條件規格
  - 多重插補中, 17
  
- 平均數
  - 平均數, 6, 8-9
  
- 排序觀察值
  - 平均數, 4
- 插補遺漏資料值, 15
  - 插補方法, 17
  - 輸出, 21
  - 限制, 19
  
- 標準差
  - 平均數, 6
  
- 次數表
  - 平均數, 6
  
- 殘差
  - 平均數, 9
  
- 疊代歷程
  - 多重插補中, 21
  
- 相關
  - 平均數, 8-9
  
- 範例檔案
  - 位置, 76
  
- 資料不完整
  - 請參閱「遺漏值分析」, 2
  
- 迴歸
  - 平均數, 9
- 選項
  - 多重插補, 29
- 遺漏值
  - 單變量統計量, 6, 34
- 遺漏值分析, 2, 32
  - EM(E), 8
  - expectation-maximization, 10
  - MCAR 檢定, 7
  - 估計統計量, 7
  - 形式, 4, 39
  - 指令的其他功能, 11
  - 插補遺漏值, 7

---

索引

敘述統計, 6, 32  
方法, 7  
迴歸, 9  
遺漏值形式, 41