IBM SPSS Missing Values 26

附註

使用此資訊和支援的產品之前，請先閱讀 第 17 頁的『注意事項』 中的資訊。

產品資訊

若新版本未聲明，則此版本便適用於 IBM SPSS Statistics 26.0.0 版以及之後發行的所有版本和修正。

目錄

遺漏值. 1
遺漏值簡介 1
遺漏值分析 1
顯示遺漏值形式 3
顯示遺漏值的描述性統計量. 3
估計統計量和插補遺漏值 4
MVA 指令的其他功能 6
多個插補. 7
分析形式. 7
插補遺漏資料值 8
MULTIPLE IMPUTATION 指令的其他功能. . . 10
使用多重插補資料 11
分析多重插補資料 12
多重插補選項. 15

注意事項 17
商標. 18

索引 19

遺漏值

下列遺漏值功能包含在 SPSS® Statistics Premium Edition 或「遺漏值」選項中。

遺漏值簡介

具有遺漏值的觀察值會面臨嚴重問題，因為一般建模程序會簡單地從分析中捨棄這些觀察值。若有遺漏值（大致上小於觀察值總數的 5%），且這些值可以視作隨機遺漏；即是否遺漏值和其他值無關，則一般的清單化刪除方法相對「安全」。「遺漏值」可幫助您判定整批刪除是否足夠，並提供在此方法不足時處理遺漏值的方法。

遺漏值分析和多重插補程序

「遺漏值」提供兩組程序來處理遺漏值：

- 多重插補程序提供對遺漏資料型樣的分析，面相最終結果的遺漏值多重插補。即，產生多個版本的資料集，每一個資料集包含自己的一組插補值。當執行統計分析時，將對所有插補資料集的參數估計進行聯合排存，提供一般比單次插補更為準確的估計。
- 遺漏值分析提供一組略有不同的敘述性工具，用於分析遺漏資料（最特別的是 Little's MCAR 檢定），並包含各種單一插補方法。請注意，一般認為多重插補優於單一插補。

遺漏值作業

您可以透過下列基本步驟開始分析遺漏值：

1. **檢查遺漏。** 使用「遺漏值分析」和「分析型樣」來探索資料中遺漏值的型樣，並判定多重插補是否必要。
2. **插補遺漏值。** 使用「插補遺漏資料值」來乘以插補遺漏值。
3. **分析「完整」資料。** 使用支援多重插補資料的任何程序。如需分析多重插補資料集與支援這些資料的程序清單，請參閱第 12 頁的『分析多重插補資料』。

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劃，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，以及極端數值個數。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，以及相關性矩陣。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定量變數插補遺漏資料。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* 回應值編碼為 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。

次數加權。 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「因素分析」，就讓您用平均值來置換遺漏值。此外，「預測附加程式模組」也提供數種方式，讓您置換時間序列中的遺漏值。

取得遺漏值分析

1. 在功能表上，選擇：

   分析 > **遺漏值分析...**

2. 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**」將遺漏資料的形式列表。如需相關資訊，請參閱主題第 3 頁的『顯示遺漏值形式』。
- 按一下「**描述性統計量**」來顯示遺漏值的描述性統計量。如需相關資訊，請參閱主題第 3 頁的『顯示遺漏值的描述性統計量』。
- 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。如需相關資訊，請參閱主題第 4 頁的『估計統計量和插補遺漏值』。
- 如果您選取 **EM** 或**迴歸**，請按一下**變數**以指定要用於估計的子集。如需相關資訊，請參閱主題第 6 頁的『預測與已預測變數』。
- 選取一個觀察值標籤變數。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

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。

表格化觀察值，依照遺漏值樣式分組

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，依照遺漏值樣式排序

每一個含有遺漏值或極端數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。

所有觀察值，選擇性地依照選擇的變數排序

每一個觀察值都表列，並為每一個變數指示遺漏值和極端數值。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. 第一類使用者遺漏值

B. 第二類使用者遺漏值

C. 第三類使用者遺漏值

變數(B)

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。

排序方式

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**所有觀察值，選擇性地依照選擇的變數排序**。

顯示遺漏值型樣

1. 在「遺漏值分析」主對話框中，選擇您想要顯示其遺漏值型樣的變數。
2. 按一下「**形式**」。
3. 選取您想要顯示的形式表。

顯示遺漏值的描述性統計量

單變量統計資料

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
- 標準差
- 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。

百分比不符

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的組做 T 檢定

您可以使用 Student's *t* 統計量，來比較每個定量變數中，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。

種類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。

顯示描述性統計量

1. 在「遺漏值分析」主對話框中，選擇您想要顯示遺漏值描述性統計量的變數。
2. 按一下「**描述性統計量**」。
3. 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。

成批法

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 層級時是顯著的。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

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。

分佈 EM 根據指定分配之概似值進行推論。依預設，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有更長尾部的分配。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(M)

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。

儲存已完成的資料

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。

指定 EM 選項

1. 在「遺漏值分析」主對話框中，選擇您想要使用 EM 方法來估計其遺漏值的變數。
2. 在**估計**群組中選取 **EM**。
3. 若要指定已預測變數和預測值變數，請按一下**「變數」**。如需相關資訊，請參閱主題第 6 頁的『預測與已預測變數』。
4. 按一下**「EM 方法」**。
5. 選取您要的 EM 選項。

迴歸方法估計選項

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性矩陣。

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，隨機選擇誤差項再加入迴歸估計值。

常態變量 *(Normal Variates)*

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中，隨機取出誤差項。

學生 *t* 變量 *(Student's t Variates)*

誤差項是以指定的自由度從 t 分配隨機抽出，並按均方根誤差 (RMSE) 縮放。

預測值數目上限

替估計程序中所需使用之預測變數（自變數）個數，設定它的最大值。

儲存已完成的資料

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。

指定迴歸選項

1. 在「遺漏值分析」主對話框中，選擇您想要採用的迴歸方法來估計遺漏值的變數。
2. 從「估計」組別中，選取「**迴歸方法**」。
3. 若要指定已預測變數和預測值變數，請按一下「**變數**」。如需相關資訊，請參閱主題『預測與已預測變數』。
4. 按一下「**迴歸方法**」。
5. 選取您要的迴歸選項。

預測與已預測變數

依預設，所有定量變數都是用來做 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。

指定已預測變數和預測值變數

1. 在「遺漏值分析」主對話框中，選擇您想要採用的迴歸方法來估計遺漏值的變數。
2. 在「估計」組別中，選取「**EM 方法**」或「**迴歸方法**」。
3. 按一下「**變數**」。
4. 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，來指定容差和收斂。
- 使用 REGRESSION 次指令，來指定容差和 *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。

多重插補資料考量

分析變數。 分析變數可以是：

- 名義. 當變數值代表實質上並未等級化的種類時（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區域、郵遞區號及宗教團體。
- 序數. 當變數值代表實質上已等級化的種類時（例如，服務滿意度從非常不滿意到非常滿意分級），則此變數可視為序數。序數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. 如果某一變數可視為尺度（連續），表示它的值代表含有實際意義矩陣的已排序種類，因此適合比較值之間的距離。尺度變數的範例包括以年份表示的年齡及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層級，但您可以在來源變數清單的變數上按一下滑鼠右鍵，並選取蹦現功能表上的測量層級，暫時變更變數的測量層級。若要暫時變更變數的測量層級，

變數清單中各變數旁的圖示可識別測量層級和資料類型：

次數加權。 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型中。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 第 12 頁的『分析多重插補資料』。

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型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。

有兩個程序專用於多重插補。

- 分析型樣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，請參閱第 12 頁的『分析多重插補資料』。

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。

在功能表上，選擇：

分析 > **多重插補** > **分析型樣...**

1. 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。

遺漏值樣式

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型樣的近似程度。程序會將分析變數排序為顯露或趨近單調型樣。若重新排序之後沒有出現單調型樣，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型樣。

含有最高遺漏值次數的變數

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描述性統計量（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，請參閱第 12 頁的『分析多重插補資料』。

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。

在功能表上，選擇：

分析 > **多重插補** > **插被遺漏資料值...**

1. 在插補模型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。
2. 指定要計算的插補數目。此值預設為 5。
3. 指定要寫入插補資料的資料集或 IBM SPSS Statistics-格式資料檔。

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* 為具備插補值的觀察值）。

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。

具有未知測量層級的欄位

若在資料集中出現一或多個不明的變數（欄位）測量層次，就會顯示「測量層次」警示。由於測量層級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層級。

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層級的任何欄位指派預設的測量層級。若為大型資料集，則讀取時可能需要一些時間。

手動指派

列出測量層次不明的所有欄位。您可以將測量層次指派給那些欄位。您也可以在「資料編輯器」的「變數清單」窗格中指派測量層次。

由於測量層次是此程序的重要項目，因此在所有欄位皆擁有已定義的測量層次之前，您無法執行此程序。

方法

「方法」對話框會指定遺漏值的插補方式，包括使用的模型類型。類別預測變數為經過編碼的指標（虛擬）。

插補方法

自動 掃描資料，若資料顯示遺漏值的單調型樣，則使用單調方法；否則就會使用完整條件規格。

自訂 若您確定要使用的方法，則選取此項。

完全條件規格 (MCMC)

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（單一因變數）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(M)

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」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型樣時使用。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，那麼就表示單調型樣存在。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（單一因變數）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。

包含種類預測中的雙向交互作用

自動選擇插補方法時，各變數的插補模型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雙向交互作用。

尺度變數的模型類型

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，其中會將迴歸模型計算出的插補值與最接近的觀察值進行比對。

羅吉斯迴歸一向用來當作類別變數的單變量模型。不論模型類型為何，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（或不可翻轉的）矩陣具有線性相關直欄，這可能導致預估演算法發生嚴重問題。即使接近奇異的矩陣也可能導致較差的結果，因此程序會將行列式小於容錯的矩陣視為奇異。指定一個正值。

限制

「限制」對話框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。

排除含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型的足夠數值，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型中的預測值。

輸出

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）、已插補的因變數、排除不予插補的因變數、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。

插補模型

這裡會顯示因變數與預測值的插補模型，並包含單變量類型、模型效應，以及插補值數量。

含有插補值變數的描述性統計量

這裡會顯示插補值的因變數之描述性統計量。對於尺度變數，描述性統計量包含原始輸入資料（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，描述性統計量包含原始輸入資料（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應變數，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。

MULTIPLE IMPUTATION 指令的其他功能

指令語法語言也可以讓您：

- 指定會顯示描述性統計量的變數子集（IMPUTATIONSUMMARIES 次指令）。
- 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型參數最大數目（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。插補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_* 需作為分組變數，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。

在功能表上，選擇：

資料 > **分割檔案...**

1. 選取**比較**組別。
2. 選擇插補數 *[Imputation_]* 作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。

區分插補值與觀察值

您可根據儲存格背景顏色、字型與粗體字（用於插補值），區分插補值與觀察值。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時，會關閉標示功能。

若要開啟標示，請從「資料編輯器」功能表選擇：

檢視 > **標示插補資料**

或可按一下「資料編輯器」中「資料視圖」其編輯列右緣的插補標示按鈕，開啟標示。

在插補之間移動

1. 在功能表上，選擇：

   編輯 > **直接跳到插補...**

2. 從下拉清單中選取插補（或原始資料）。

或者，可以從「資料編輯器」其「資料視圖」中編輯列的下拉清單選取插補。

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，也會保存直欄位置，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「資料編輯器」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「描述性統計量」子功能表中，「次數分配表」、「描述性統計量」、「探索」、「交叉表」都支援合併，而「比率」、「P-P 圖」和「Q-Q 圖」則否。

表列式輸出和模型 PMML 都可以合併。要求合併輸出並不需要新程序；而是「選項」對話框上會有個新標籤，讓您全面控制多重插補輸出。

- **表列式輸出的合併。** 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。支援下列功能：

- 「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。

敘述性統計量。支援下列功能：

- 「敘述性統計量」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 N。

交叉表。支援下列功能：

- 「交叉表列」表格在單純合併中支援個數。

平均數。支援下列功能：

- 「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。支援下列功能：

- 「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。

獨立樣本 T 檢定。支援下列功能：

- 「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。支援下列功能：

- 「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。

單向變異數分析。支援下列功能：

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

UNIANOVA。支援下列功能：

- 單純合併中的描述性統計量。
- 單純合併中的受試者間因素。
- 單變量合併中的參數估計值。
- 單變量合併中的邊際平均數估計值。
- 單變量合併中的邊際平均數估計值比較。

GLM 單變數。支援下列功能：

- 「參數估計值」表格在單變量合併中支援係數 B。

線性混合模型。支援下列功能：

- 「敘述性統計量」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。
- 估計邊緣平均數：「估計值」表格在單變量合併中支援平均數。
- 估計邊緣平均數：「成對比較」表格在單變量合併中支援平均數差異。

「概化線性模型」與「廣義估計方程式」。 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。
- 估計邊緣平均數：「估計係數」表格在單純合併中支援平均數。
- 估計邊緣平均數：「估計值」表格在單變量合併中支援平均數。
- 估計邊緣平均數：「成對比較」表格在單變量合併中支援平均數差異。

雙變量相關分析。支援下列功能：

- 「敘述性統計量」表格在單純合併中支援平均數與個數。
- 「相關性」表格在單變量合併中支援相關與 N。請注意，在合併之前，使用 Fisher's *z* 轉換來轉換相關，然後在合併之後進行反向轉換。

偏相關分析。支援下列功能：

- 「敘述性統計量」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。

線性迴歸。 此程序支援合併的 PMML。

- 「敘述性統計量」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。

二元羅吉斯迴歸。 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。

多項式羅吉斯迴歸。 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。

序數迴歸。支援下列功能：

- 「參數估計值」表格在單變量合併中支援係數 B。

區別分析。 此程序支援合併的模式 XML。

- 「組別統計量」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。
- 「典型區別函數係數」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。支援下列功能：

- 「敘述性統計量」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。支援下列功能：

- 「敘述性統計量」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。

執行檢定。支援下列功能：

- 「敘述性統計量」表格在單純合併中支援平均數與個數。

單一樣本 Kolmogorov-Smirnov 檢定。支援下列功能：

- 「敘述性統計量」表格在單純合併中支援平均數與個數。

兩個獨立樣本檢定。支援下列功能：

- 「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。支援下列功能：

- 「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計數。

兩個相關樣本檢定。支援下列功能：

- 「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。支援下列功能：

- 「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

「多重插補」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，所產生的檢視器輸出類型。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」標籤。

注意事項

本資訊係針對 IBM 在美國所提供之產品與服務所開發。IBM 可能會以其他語言提供本資料。但是，您可能需要具有該語言的產品或產品版本，才能存取該產品。

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, MD-NC119
Armonk, NY 10504-1785
US

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.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

International Business Machines Corporation 只依「現況」提供本出版品，不提供任何明示或默示之保證，其中包括且不限於不侵權、可商用性或特定目的之適用性的隱含保證。有些地區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 Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US

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。

本文件中引用的效能資料及用戶範例僅供敘述之目的。特定配置及作業條件下的實際效能結果可能不同。

本文件所提及之非 IBM 產品資訊，取自產品的供應商，或其發佈的聲明或其他公開管道。IBM 並未測試過這些產品，也無法確認這些非 IBM 產品的執行效能、相容性或任何對產品的其他主張是否完全無誤。有關非 IBM 產品的功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。這些名稱全為虛構，如與實際人員或企業之名稱有所雷同，純屬巧合。

著作權授權：

本資訊含有原始語言之範例應用程式，用以說明各作業平台中之程式設計技術。貴客戶可以為了研發、使用、銷售或散布符合範例應用程式所適用的作業平台之應用程式介面的應用程式，以任何形式複製、修改及散布這些範例程式，不必向 IBM 付費。這些範例並未在所有情況下完整測試。故 IBM 不保證或默示保證這些樣本程式之可靠性、服務性或功能。這些程式範例以「現狀」提供，且無任何保證。IBM 對因使用這些程式範例而產生的任何損害概不負責。

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：

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。

〔四劃〕

分析樣式 7

〔五劃〕

平均數
　平均數 3, 5

〔六劃〕

在遺漏值分析中
　平均數 3, 5
多重插補 7
　分析形式 7
　插補遺漏資料值 8
多個插補 11, 12
成對刪除
　平均數 1
次數表
　平均數 3

〔七劃〕

完全刪除
　平均數 1
完整條件規格
　多重插補中 9

〔九劃〕

相關
　平均數 5

〔十劃〕

迴歸
　平均數 5

〔十一劃〕

排序觀察值
　平均數 3

〔十二劃〕

單調插補
　多重插補中 9
插補遺漏資料值 8
　限制 10
　插補方法 9
　輸出 10
殘差
　平均數 5

〔十三劃〕

極端數值數量
　平均數 3
資料不完整
　請參閱「遺漏值分析」 1

〔十五劃〕

標準差
　平均數 3

〔十六劃〕

遺漏值
　單變量統計量 3
遺漏值分析 1
　方法 4
　估計統計量 4
　形式 3
　指令的其他功能 6
　迴歸 5
　描述性統計量 3
　插補遺漏值 4
　EM 5
　expectation-maximization 6
　MCAR 檢定 4

〔二十二劃〕

疊代歷程
　多重插補中 10

E

EM
　平均數 5

L

Little's MCAR 檢定 4
　平均數 1

M

MCAR 檢定
　平均數 1

T

t 檢定
　平均數 3

Printed in Taiwan