IBM SPSS Statistics Base 19

Note: Before using this information and the product it supports, read the general information under Notices第 284 頁.

序

IBM® SPSS® Statistics為分析資料的強大系統。Base 的選用性附加模組能提供其他本手冊所說明的分析技術。Base 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

內容

章節
1

Codebook

「編碼簿」會報告字典資訊（例如，變數名稱、變數標記、數值標記、遺漏值），以及作用中資料集中所有或特定變數和複選題集的摘要統計量。對於名義和次序變數與複選題集，摘要統計量會包括個數與百分比。對於尺度變數，摘要統計量會包括平均數、標準差及四分位數。

注意：編碼簿會忽略分割檔狀態。其包括針對遺漏值多重插補所建立的分割檔群組（提供於「遺漏值」附加選項）。

若要取得編碼簿

- 從功能表選擇：
  分析(A) > 報表 > Codebook
- 按一下「變數」索引標籤。

圖表 1-1
「編碼簿」對話方塊的「變數」索引標籤

▶ 選取一或多個變數和/或複選題集。

您可以：

- 控制要顯示的變數資訊。
- 控制要顯示的統計量（或排除所有摘要統計量）。
- 控制顯示變數與複選題集的順序。
- 變更來源清單中任何變數的測量水準，以變更顯示的摘要統計量。

變更測量水準

您可以暫時變更變數的測量水準。（您無法變更複選題集的測量水準。系統一律會將他們視為名義。）

▶ 在來源清單的變數上按一下滑鼠右鍵。

▶ 選取快顯功能表上的測量水準。

這會暫時變更測量水準。在實際情況中，這只對數值變數有用。名義或次序會限制字串變數的測量水準，但在「編碼簿」程序裡名義和次序是一樣的。

編碼簿輸出索引標籤

「輸出」索引標籤可控制要包含以供各個變數和複選題集使用的變數資訊、顯示變數與複選題集的順序，以及選擇性檔案資訊表格的內容。

圖表 1-2
「編碼簿」對話方塊的「輸出」索引標籤

變數資訊

這會控制為各個變數顯示的字典資訊。

位置。代表檔案順序中變數位置的整數。這無法用於複選題集。

標記。與變數或複選題集相關聯的描述性標記。

類型。基本資料類型。這可以是「數值」、「字串」或「複選題集」。

格式。變數的顯示格式，例如，A4、F8.2 或 DATE11。這無法用於複選題集。

測量水準。可能值為「名義」、「次序」、「尺度」及「未知」。顯示的數值是字典中所儲存的測量水準，而且不會因為在「變數」索引標籤上變更來源變數中的測量水準時，所指定的任何暫時測量水準覆寫而有所影響。這無法用於複選題集。

注意：若尚未明確設定測量水準，則在第一次傳遞資料之前，數字變數的測量水準可能是「未知的」，例如，從外部來源讀取的資料或者新建的變數。

角色。 部分對話方塊支援依據定義的角色預先選取變數進行分析的功能。

數值標記。與特定資料值相關聯的描述性標記。

- 若已在「統計量」索引標籤上選取「個數」或「百分比」，則定義的數值標記會包含於輸出中，即使您並未在此處選取「數值」標記也一樣。
- 對於多重二分集而言，「數值標記」是集合中基本變數的變數標記或個數值的標記（視定義集合的方式而定）。

遺漏值。使用者定義的遺漏值。若已在「統計量」索引標籤上選取「個數」或「百分比」，則定義的數值標記會包含於輸出中，即使您並未在此處選取「遺漏值」也一樣。這無法用於複選題集。

自訂屬性。使用者定義的自訂變數屬性。輸出包括與各個變數相關聯之任何自訂變數屬性的名稱和數值。這無法用於複選題集。

保留的屬性。保留的系統變數屬性。您可以顯示系統屬性，但不應進行改變。系統屬性名稱會以貨幣符號 ($) 開頭。不包括非顯示用的屬性，其名稱是以 “@” 或 “$@” 開頭。輸出包括與各個變數相關聯之任何系統屬性的名稱和數值。這無法用於複選題集。

檔案資訊

選擇性的檔案資料表格可包括下列任何一個檔案屬性：

檔案名稱。IBM® SPSS® Statistics 資料檔的名稱。若資料集未曾以 SPSS Statistics 格式儲存，則不會有任何資料檔名稱(如果「資料編輯程式」視窗的標題列中未顯示任何檔案名稱，則作用中資料集便不會有檔案名稱)。

位置。SPSS Statistics 資料檔的字典（資料夾）位置。若資料集未曾以 SPSS Statistics 格式儲存，就不會有位置。

觀察值個數。作用中資料集中的觀察值個數。此為觀察值總數，包括可能因過濾條件而從摘要統計量中排除的任何觀察值。

標記。這是由 FILE LABEL 指令所定義的檔案標記（如果有的話）。

文件。 資料檔案文件文字。

加權狀態。若已啟用加權，即會顯示加權變數的名稱。

自訂屬性。使用者定義的自訂資料檔案屬性。利用 DATAFILE ATTRIBUTE 指令定義的資料檔案屬性。

保留的屬性。保留的系統資料檔案屬性。您可以顯示系統屬性，但不應進行改變。系統屬性名稱會以貨幣符號 ($) 開頭。不包括非顯示用的屬性，其名稱是以 “@” 或 “$@” 開頭。輸出會同時包括所有系統資料檔案屬性的名稱與數值。

變數顯示順序

以下選項可用於控制顯示變數與複選題集的順序。

字母順序。依變數名稱排序的字母順序。

檔案。變數出現在資料集中的順序（變數在「資料編輯程式」中顯示的順序）。使用遞增順序時，複選題集會在所有選取的變數之後，於最後一個顯示。

測量水準。依測量水準排序。這會建立四個排序組別：名義、次序、尺度及未知。系統會將複選題集視為名義。

注意：若尚未明確設定測量水準，則在第一次傳遞資料之前，數字變數的測量水準可能是「未知的」，例如，從外部來源讀取的資料或者新建的變數。

變數清單。變數與複選題集在「變數」索引標籤上，選取變數清單中顯示的順序。

自訂變數名稱。排序順序選項的清單也包括任何使用者定義之自訂變數屬性的名稱。使用遞增順序時，不含屬性的變數會排序在最頂端，接著是含屬性但未定義屬性之數值的變數，之後則是含有已定義屬性之數值的變數（會以數值的字母順序排序）。

最大類別個數

若輸出包含每個唯一值的數值標記、個數或百分比，當數值的個數超過指定的數值時，您可以不要顯示這項來自表格的資訊。依照預設，若變數的唯一值個數超過 200，即不會顯示此項資訊。

編碼簿的統計量索引標籤

「統計量」索引標籤允許您控制輸出中所包含的摘要統計量，或者完全不顯示摘要統計量。

圖表 1-3
「編碼簿」對話方塊的「統計量」索引標籤

個數和百分比

對於名義和次序變數、複選題集，以及尺度變數標記的數值而言，變數統計量是：

個數。 具有每個變數值（或數值範圍）之觀察值的計數或個數。

百分比。 具有某個數值的觀察值百分比。

集中趨勢和分散情形

對於尺度變數而言，可用的統計量是：

平均數。 一種集中趨勢的量數。算術平均數，總和除以觀察值個數。

標準差。 在平均數四周離散的量數。在常態分配中，68% 的觀察值會落在平均數的一個標準差內，95% 的觀察值會落在兩個標準差內。例如，若平均年齡為 45 歲，標準差是 10 的話，在常態分配中 95% 的觀察值會介於 25 到 65 歲之間。

分位數。 顯示對應至第 25 個、第 50 個和第 75 個百分位數的數值。

注意：您可以在「變數」索引標籤的來源變數清單中，暫時變更與變數相關聯的測量水準（如此就可變更針對該變數所顯示的摘要統計量）。

章節
2

次數分配表(F)

「次數分配表」程序所提供的統計量和圖形畫面，在描述多種變數的類型時是很好用的。「次數分配表」程序是檢視資料的好地方。

在次數報表和長條圖中，您可以依遞增或遞減的順序排列不同值，或者您可以按照次數，來排列類別順序。當變數中有許多明確的數值時，可以不列出次數分配表。此外，您可以使用次數（預設值）或百分位數，在圖表中加上標記。

範例。如果把客戶按照行業來分類，則分配情形為何？從輸出中，您可能會發現，客戶中，有 37.5% 來自政府單位、24.9% 來自公司行號、28.1% 來自學術機構、9.4% 來自醫療院所。此外，如果觀察連續性的數值資料（如銷貨收益），就可以得知，產品平均銷售額為 NT$3.576，而標準差為 NT$1,078。

統計量與圖形。您可以使用次數個數、百分位數、累積百分比、平均數、中位數、眾數、總和、標準差、變異數、範圍、最小值和最大值、平均數的標準誤、偏態和峰度（兩者都具標準誤）、四分位數、使用者指定的百分位數、長條圖、圓餅圖和直方圖。

資料。 使用數值代碼或字串做為類別變數的代碼（名義或次序的水準測量）。

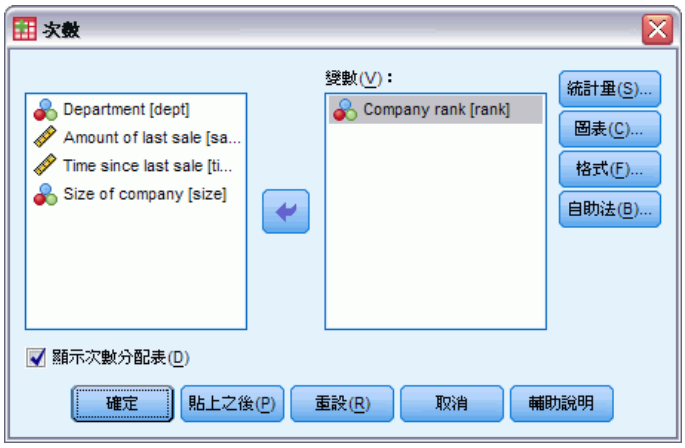
假設。 列表和百分比為來自任何分配的資料提供有用的說明，尤其是含有排序或未排序類別的變數。大部分選擇性的摘要統計量（如平均數和標準差），都是以常態理論為基礎，而且適合用來分析對稱分配的數值變數。至於穩健統計量（如中位數、四分位數、和百分位數等），則適合用來分析數值變數（該變數可能符合，或者不符合常態性的假設）。

若要取得次數分配表

▶ 從功能表選擇：

分析(A) > 敘述統計 > 次數分配表...

圖表 2-1
「次數分配表」主對話方塊



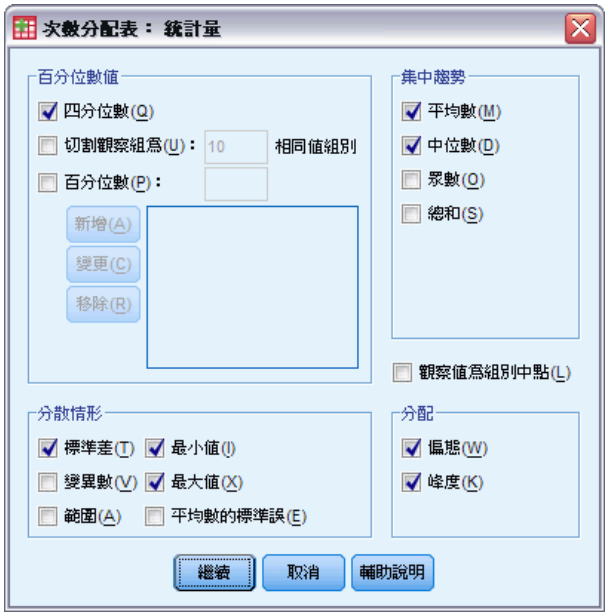
- 選取一個以上的類別或數值變數。

您可以：

- 按一下**統計分析**，選擇數值變數的敘述統計。
- 按一下**圖表**，選擇長條圖、圓餅圖和直方圖。
- 按一下**格式**，選擇結果的顯示順序。

次數分配表的統計量

圖表 2-2
次數分配表統計量對話方塊



百分位數值。此處指的是數值變數的值。 系統使用數值變數，將順序排列的資料分組，因此某個百分比的資料會在這個數值以上，而其他百分比的資料則在它以下。四分位數（第 25 個、第 50 個和第 75 個百分位數）可以將觀察值分成四個組別（大小相等）。如果您想將觀察值均分成其他等分（除了四等分以外），請選取**切割觀察值為 n 組相同組別**。或者，您也可以指定其他的百分位數（例如第 95 個百分位數，就是有 95% 的觀察值在這個值以下）。

集中趨勢。這些統計量可以描述分配位置，包括：平均數、中位數、眾數、和所有值的總和。

- **平均數。** 一種集中趨勢的量數。算術平均數，總和除以觀察值個數。
- **中位數。** 半數觀察值落點上下的值，第 50 個百分位數。如果觀察值個數是偶數的話，中位數是中間兩個觀察值的平均值，此處的兩個中間是指，當觀察值按照遞增或遞減順序排列時，位於最中間的兩個值。中位數為集中趨勢的量數，其不會感應到偏離值（相反地，平均數會受到幾個高低極端數值所影響）。
- **眾數(O)。** 最常出現的值。如果幾個數值的最大出現次數相同，則每一個都是眾數。「次數分配表」程序僅報告這種多個眾數的最小值。
- **總和。** 遍及所有包含遺漏值之觀察值的數值總和或總數。

分散情形。這些統計量可以測量變化數量，或是資料的擴散程度，包括：標準差、變異數、範圍、最小值、最大值和平均數標準誤。

- **標準差(T)。** 在平均數四周離散的量數。在常態分配中，68% 的觀察值會落在平均數的一個標準差內，95% 的觀察值會落在兩個標準差內。例如，若平均年齡為 45 歲，標準差是 10 的話，在常態分配中 95% 的觀察值會介於 25 到 65 歲之間。
- **變異數。** 平均數四周離散的量數，它等於平均數的平方離差總和除以觀察值個數減一。變異數的測量單位是變數本身的平方。
- **範圍。** 數值變數最大和最小值之間的差異，也就是最大值減去最小值。
- **最小值。** 數字變數的最小值。
- **最大值。** 數字變數的最大值。
- **標準誤平均數。** 測量從同一個分配取出來的不同樣本間平均數的變化大小。它可以用來大略地將觀察平均數與假設值相比較（也就是如果標準誤與差異的比值小於 -2 大於 +2 的話，您就可以下結論說兩個值不同）。

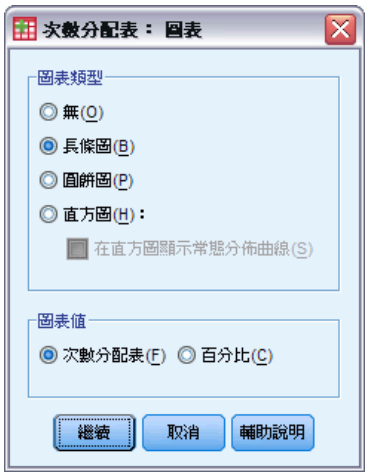
分配。偏態和峰度這兩種統計量，可以用來描述分配類型和對稱性。這些統計量會與其標準誤一起顯示。

- **偏態。** 分配不對稱性的量數。常態分配是對稱的，且偏態值為 0。有顯著正偏態值的分配會有一個長的偏右尾部。有顯著負偏態值的分配會有一個長的偏左尾部。偏態值有如指標，若大於它的兩倍標準誤，則表示背離對稱。
- **峰度。** 中心點四周之觀察值集群的程度量數。對常態分配而言，峰度統計量數值為零。正的峰度值表示相較於常態分配，觀察值較為集中在分配的中央，並且有延伸至分配極端值的較薄尾部，而在此時高峽峰分配的尾部則較常態分配為厚。負的峰度值表示相較於常態分配，觀察值較為分散，並且有延伸至分配極端值的較厚尾部，而在此時低闊峰分配的尾部則較常態分配為薄。

觀察值為組別中點。如果您資料中的值為組別中點的話（例如，所有三十多歲的人，都編碼成 35），那麼選擇這個選項，便可以估計資料在未分組前的中位數和百分位數。

次數分配表圖表

圖表 2-3
次數分配表圖表對話方塊

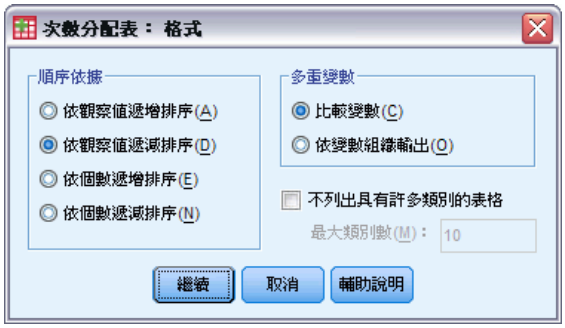


圖表類型。圓餅圖可顯示整體圖形中，每部分的構成比例。圓餅圖的每一個圖塊，都會對應到單一分組變數所定義的組別。而長條圖會把每個不同數值（或類別）的個數，顯示成不同的條狀，讓您可以一眼就看出不同的類別，然後進行比較。直方圖也會顯示長條，但是長條之間的距離是相等的。每個長條的高度，就等於區間裡面的數值變數值個數。直方圖可顯示出分配的類型、中心、和分散情形。直方圖上會重疊顯示常態曲線，以幫助您判斷資料是否為常態分配。

圖表值。如果您使用長條圖，可以利用次數個數或百分比，為尺度軸加上標記。

次數分配表的格式

圖表 2-4
次數分配表格式對話方塊



順序依據。 次數分配表可以根據資料中的實際數值，或是那些數值的個數（意指發生的次數）加以排列，而表格可以遞增或遞減的順序排列。但是如果您要使用直方圖或百分位數的話，「次數分配表」會假設變數為數值變數，並以遞增順序來顯示這些變數值。

多重變數。如果您建立多重變數的統計量表格，就可以在單一表格中，顯示所有的變數（**比較變數**），或者顯示每個變數的不同統計量表格（**依變數組成輸出**）。

不列出超過 n 個類別的表格。這個選項的作用在於：如果類別個數超過指定值，則不會顯示這個表格。

章節
3

描述性統計量(D)

在單一表格中，您可以使用「描述性統計量」程序，來顯示數個變數的單變量摘要統計量，以及計算標準化數值 (z 分數)。其中，變數還可以依其平均數的大小 (遞增或遞減順序)、字母順序，或者根據您選取變數的順序 (預設值)，來加以排列。

當您儲存 z 分數後，程式會將它們加入「資料編輯程式」的資料中，以供圖表、資料清單和分析使用。當變數是以不同的單位來記錄時 (例如，每人的本國生產毛額和實際百分比)，z 分數轉換會將變數置於一般尺度上，方便您以目視的方式加以比較。

範例。 以銷售人員的業績為例。如果資料中的每個觀察值都含有每位銷售人員 (例如，其中一個項目代表 Bob，另一個代表 Kim，而另一個為 Brian) 的當日銷售總額，持續收集數個月之後，您就可以使用「描述性統計量」程序，來計算每位銷售人員的每日平均銷售額，並將結果按照他們的平均銷售額，從最高排到最低。

統計量。包括：樣本大小、平均數、最小值、最大值、標準差、變異數、範圍、總和、平均數的標準誤，以及標準誤的峰度與偏態。

資料。在您將變數製成圖表，以記錄誤差、偏離值和非分配性常態之後，請使用數值變數。對大型表格 (具數千個觀察值) 而言，「描述性統計量」程序非常有效率。

假設。 大部分可以使用的統計量 (包括 z 分數)，都是以一般性理論為基礎，並且適用於對稱性分配的數值變數 (區間或比例水準的測量)。避免未排序非類別或非對稱性分配的變數。z 分數的分配類型，與原始資料的類型相同；因此，計算 z 分數並不能解決問題資料。

若要取得敘述統計

▶ 從功能表選擇：
分析(A) > 敘述統計 > 描述性統計量...

圖表 3-1
描述性統計量對話方塊

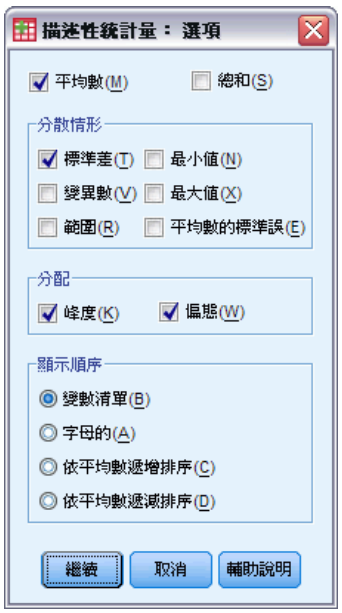
- 選取一個或多個變數。

  您可以：

  - 選取**將標準化數值存成變數**，以將 z 分數存成新的變數。
  - 按一下**選項**，以取得選項的統計量，並顯示順序。

描述性統計量選項

圖表 3-2
描述性統計量選項對話方塊



平均數和總和。這個選項會依照預設值，顯示平均數和算術平均數。

分散情形。這個選項乃是用來測量資料分佈和變化的統計量。 這些統計量包括標準差、變異數、範圍、最小值、最大直，以及平均數的標準誤。

- **標準差(T)。** 在平均數四周離散的量數。在常態分配中，68% 的觀察值會落在平均數的一個標準差內，95% 的觀察值會落在兩個標準差內。例如，若平均年齡為 45 歲，標準差是 10 的話，在常態分配中 95% 的觀察值會介於 25 到 65 歲之間。
- **變異數。** 平均數四周離散的量數，它等於平均數的平方離差總和除以觀察值個數減一。變異數的測量單位是變數本身的平方。
- **範圍。** 數值變數最大和最小值之間的差異，也就是最大值減去最小值。
- **最小值。** 數字變數的最小值。
- **最大值。** 數字變數的最大值。
- **平均數的標準誤(E)。** 測量從同一個分配取出來的不同樣本間平均數的變化大小。它可以用來大略地將觀察平均數與假設值相比較（也就是如果標準誤與差異的比值小於 -2 大於 +2 的話，您就可以下結論說兩個值不同）。

分配。峰度和偏態都是統計量，可用來描述分配的形狀和對稱性。這些統計量會與其標準誤一起顯示。

- **峰度。** 中心點四周之觀察值集群的程度量數。對常態分配而言，峰度統計量數值為零。正的峰度值表示相較於常態分配，觀察值較為集中在分配的中央，並且有延伸至分配極端值的較薄尾部，而在此時高峽峰分配的尾部則較常態分配為厚。負的峰度值表示相較於常態分配，觀察值較為分散，並且有延伸至分配極端值的較厚尾部，而在此時低闊峰分配的尾部則較常態分配為薄。
- **偏態。** 分配不對稱性的量數。常態分配是對稱的，且偏態值為 0。有顯著正偏態值的分配會有一個長的偏右尾部。有顯著負偏態值的分配會有一個長的偏左尾部。偏態值有如指標，若大於它的兩倍標準誤，則表示背離對稱。

顯示次序。依照預設值，變數是依照您所選取的次序排列。除此之外，您也可以依照字母順序、平均數的遞增或遞減順序，來顯示變數。

DESCRIPTIVES 指令的其他功能

指令語法語言也可以讓您：

- 儲存部分變數的標準化分數（z 分數），而不是全部的變數（使用 VARIABLES 次指令）。
- 指定包含有標準化分數的新變數名稱（使用 VARIABLES 次指令）。
- 從任何變數中含有遺漏值的分析觀察值排除（使用 MISSING 次指令）。
- 依任何統計量的數值排序顯示的變數，而不是依據平均數（使用 SORT 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
4

預檢資料

不論是哪一個觀察值或是觀察值組別，都可以使用「預檢資料」程序，來產生摘要統計量和圖形顯示。使用「預檢資料」程序的原因有很多，如資料篩檢、離群值識別、說明、假設檢查、以及描述子母體（觀察值組別）之間差異的特徵等。當您資料篩檢時，可能會顯示一些不尋常的值、極端值、資料間隙、或其他特殊狀況。所以，預檢資料能協助您，判斷用來分析資料的統計技術是否適當。如果技術需要常態分配，預檢資料可能會提示您必須轉換資料。或者，詢問您是否需要無母數檢定。

範例。以老鼠學習走迷宮為例，在四種不同研究主題的實驗下，對時間的分配進行研究。對這四組中的任何一組，您都可以觀察時間的分配是否近似常態分配，以及這四個變異數是否相等。您也可以識別五個最長時間和五個最短時間的觀察值。然後再以盒形圖和莖葉圖的圖形，記錄每一組學習時間分配的摘要。

統計量與圖形。在統計量方面，共有平均數、中位數、5% 修整平均數、標準誤、變異數、標準差、最小值、最大值、範圍、四分位範圍、偏態與峰度及其標準誤、平均數的信賴區間（和指定的信賴水準）、百分位數、Huber M 估計式、Andrews wave 估計式、Hampel’s 再下降 M 估計式、Tukey’s 二權數估計式、五個最大值和五個最小值、檢定常態性的 Lilliefors 顯著水準的 Kolmogorov-Smirnov 統計量，以及 Shapiro-Wilks 統計量。在圖形方面，共有莖葉圖、直方圖、常態機率圖，以及包含 Levene 檢定與轉換的離散對水準之圖形。

資料。「預檢資料」程序適用於數值變數（區間或比例水準測量）。因子變數（用來將資料分成數個觀察值組別）應該有一些不同的數值（類別）。.這些值可以是短的字串或數字。而盒形圖中用來標示離群值的觀察值標記變數，則可以是短字串、長字串（前 15 個位元組）或數字。

假設。您的資料分配不一定是對稱式或常態性的。

若要預檢您的資料

▶ 在功能表上，選擇：
分析(A) > 敘述統計 > 預檢資料...

圖表 4-1
「預檢資料」對話方塊

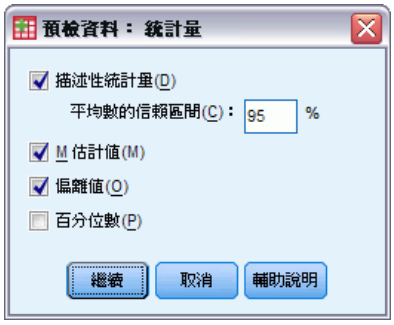
- 選擇一個或多個依變數。

  您可以：

  - 選取一個（或多個）因子變數，該值會定義觀察值組別。
  - 選取識別變數來標記觀察值。
  - 如需穩健的估計式、離群值、百分位數和次數分配表，請按一下「統計量」。
  - 如需直方圖、常態機率圖和檢定，以及包含 Levene 統計量的離散對水準之圖形，請按一下「圖形」。
  - 按一下「選項」，以指定處理遺漏值的方式

預檢資料統計量

圖表 4-2
「預檢資料統計量」對話方塊



描述性統計量。這些集中趨勢和分散測量將依照預設值顯示。集中趨勢的測量會指出分配的位置，包括平均數、中位數，以及 5% 修整平均數。分散測量則顯示數值的相異性，包括標準誤、變異數、標準差、最小值、最大值、範圍，以及四分位範圍。敘述統計也包括分配形狀的測量；偏態和峰度則會顯示它們的標準誤。同時也會顯示平均數的 95% 信賴區間，但是您也可以指定不同的信賴區間。

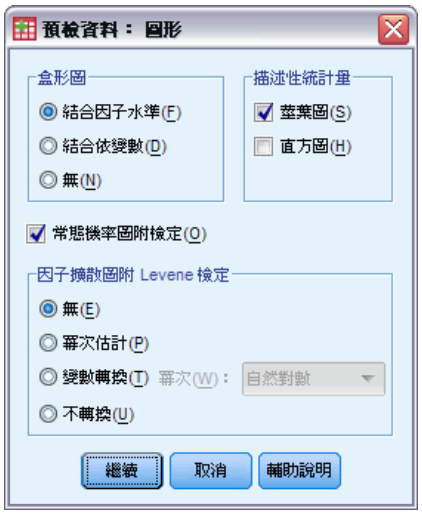
M 估計式。 您可以使用這個數值，取代樣本平均數和中位數，來推算位置。每個估計式會算出不同的加權值，然後再套用到觀察值上。顯示多個估計式，包括 Huber's M 估計式、Andrews' wave 估計式、Hampel's 再下降 M 估計式以及 Tukey's 二權數估計式。

偏離值。 使用觀察值標記，來顯示五個最大值和五個最小值。

百分位數。顯示第 5 個、第 10 個、第 25 個、第 50 個、第 75 個、第 90 個，以及第 95 個百分位數的值。

預檢資料的圖形

圖表 4-3
「預檢資料圖形」對話方塊



盒形圖。當您的依變數不止一個時，就可以使用這些項目，來控制盒形圖的顯示方式。「結合因子水準」會分別顯示每一個依變數。在每一個顯示畫面中，因子變數所定義的每個組別，都會顯示盒形圖。「結合依變數」會將因子變數所定義的每個組別，分別顯示出來。在每一個顯示畫面中，每個依變數的盒形圖會並排顯示出來。當不同的變數代表不同時間所測量的特性時，這種顯示方式就特別有用。

描述性統計量。「描述性統計量」組別，能讓您選擇莖葉圖和直方圖。

常態機率圖附檢定。顯示常態機率圖，以及去除趨勢的常態機率圖。該圖也會顯示包含檢定常態性之 Lilliefors 顯著水準的 Kolmogorov-Smirnov 統計量。如果指定非整數加權，當加權樣本大小落在 3 和 50 之間時，會計算 Shapiro-Wilk 統計量；如果不加權或指定整數加權，則要當加權樣本大小落在 3 和 5,000 之間時，才會計算此統計量。

因子擴散圖附 Levene 檢定。控制離散對水準之圖形的資料轉換。所有離散對水準之圖形，都會顯示迴歸線的坡度，以及變異數均齊性的 Levene's robust 檢定。如果您選取轉換，Levene 檢定就會以轉換的資料為基礎。如果沒有選取任何因子變數，就不會產生離散對水準之圖形。「冪次估計」會產生四分位數範圍的自然對數、對照所有儲存格中位數之自然對數的圖形，以及達成儲存格中相等變異數的冪次轉換估計值。離散對水準之圖形能協助您，判斷穩定組別間變異數（得到更多相等的變異數）所需的轉換冪次。「變數轉換」能讓您選取「冪次」中的其中一個項目（或遵照「冪次」估計所提出的建議），

以產生轉換資料圖形。此外，它也會繪製四分位範圍、以及轉換資料中位數的圖形。而「不轉換」會產生原始資料的圖形。相當於冪次 1 的轉換。

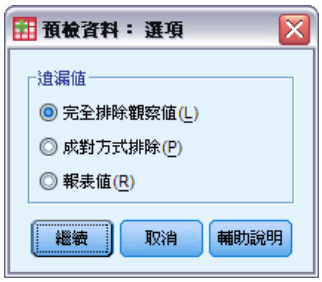
預檢資料的冪次轉換

這些是離散對水準之圖形的冪次轉換。若要轉換資料，您必須先選取要轉換的冪次。您可以任選下列其中一個選項：

- **自然對數。** 自然對數轉換。此為預設值。
- **1/平方根。**計算每個資料值平方根的倒數。
- **倒數。**計算每個資料值的倒數。
- **平方根。**計算每個資料值的平方根。
- **平方。**將每個資料值平方。
- **立方。**將每個資料值立方。

預檢資料的選項

圖表 4-4
「預檢資料選項」對話方塊



遺漏值。控制處理遺漏值的方式。

- **完全排除遺漏值。**如果有任何一個觀察值含有依變數或因子變數的遺漏值，那麼所有的分析都會排除這個觀察值。此為預設值。
- **成對方式排除。**如果組別（儲存格）中變數的觀察值沒有包含遺漏值，那麼該組分析就會將這個觀察值包含在內。對於其他組別中的變數，該觀察值可能就含有遺漏值。
- **報表值。**因子變數的遺漏值，會被視為一個不同的類別。所有的輸出都是因這個另外的類別而產生的。次數分配表就包含遺漏值的類別。雖然因子變數的遺漏值會被包括在內，但是也會將它標示為 “遺漏”。

EXAMINE 指令的其他功能

「預檢資料」程序使用 EXAMINE 指令語法。指令語法語言也可以讓您：

- 要求由因子變數所定義之輸出和圖形群組之外的輸出和圖形總和（使用 TOTAL 次指令）。
- 指定盒形圖群組的一般尺度（使用 SCALE 次指令）。
- 指定因子變數的交互作用（使用 VARIABLES 次指令）。

- 指定預設值以外的百分位數（使用 PERCENTILES 次指令）。
- 根據任五項方法來計算百分位數（使用 PERCENTILES 次指令）。
- 指定任何離散對水準之圖形的冪次轉換（使用 PLOT 次指令）。
- 指定要顯示之極端數值（使用 STATISTICS 次指令）。
- 指定 M 估計式參數，位置的穩健估計式（使用 MESTIMATORS 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
5

交叉表

您可以利用「交叉表」程序，形成二因子和多因子的表格，並為二因子表格提供數種檢定和關聯測量。表格的結構以及是否依類別排列，都是用來決定要使用哪一種檢定或測量的依據。

只有二因子的表格，才能計算「交叉表」的統計量和關聯測量。如果您指定列、行和階層因子（控制變數），則「交叉表」程序會形成一個面板，包含與階層因子（或兩個以上控制變數的數值組合）中每個值相關的統計量和測量。例如，若 gender 是 married（是、否）與 life（生活刺激、規律或無聊）所構成表格的階層因子，則系統會分別計算女性和男性的二因子表格，並以面板的方式依次列印出來。

範例。以公司大小與服務品質為例。小公司的客戶會比大公司的客戶，更容易感受到高品質的銷售服務嗎（例如訓練或諮詢）？從交叉表中，您可以得知，多數的小公司（少於 500 人）提供高品質的服務，而多數的大公司（超過 2,500 人）只提供低品質的服務。

統計量和關聯測量。包括 Pearson 卡方、概似比卡方、列與行變數間之線性關聯檢定、Fisher’s 精確檢定、Yates’ 修正卡方、Pearson’s r 值、 Spearman’s rho、列聯係數、Phi、Cramér’s V 係數、對稱性和非對稱性 lambdas、Goodman 與 Kruskal’s tau 測量、不確定係數、Gamma 分配、Somers’ d 值、Kendall’s tau-b 統計測量、Kendall’s tau-c 統計測量、Eta 係數、Cohen’s Kappa 統計測量、相對風險估計值、odds 比率、McNemar 檢定和 Cochran’s 和 Mantel-Haenszel 統計量。

資料。 若要定義各表格變數的類別，請使用數值或字串（八個位元組以下）變數。例如，您可以將 gender 的資料編碼為 1 和 2，或只用 male 和 female 來表示。

假設。 如同本節所討論的統計量一樣，有一些統計量和測量，會先假設類別已經排序（次序資料）或者為數值（間隔或比例資料）。其他的統計量或測量，則在表格變數類別未排序（名義資料）時才有效。如果是以卡方分配為基礎的統計量（Phi 值、Cramér’s V 係數和列聯係數），資料應該是取自多項式分配的隨機樣本。

注意：次序變數可以是代表類別的數字編碼（例如，1 = 低、2 = 中、3 = 高）或字串變數。但是，系統會假設字串變數的字母順序反映著類別的實際順序。例如，數值為低、中、高的字串變數，其類別排序會被解譯為高、低、中 — 這並不是正確的順序。一般來說，最好使用數值碼來表示次序資料會比較妥當。

若要取得交叉表

▶ 從功能表選擇：
分析(A) > 敘述統計 > 交叉表...

圖表 5-1
交叉表對話方塊

▶ 選取一個或多個列變數，以及一個或多個行變數。

您可以：

- 選取一個或多個控制變數。
- 按一下**統計量**，以取得二因子表格或次表的檢定和關聯測量。
- 按一下**格**，以取得觀察值與期望值、百分比和殘差。
- 按一下**格式**，以控制類別的順序。

交叉表階層

如果您選取一個以上的階層變數，那麼每個階層變數（控制變數）的每個類別，都會產生不同的交叉表列。舉例來說，假設您有一個列變數、一個行變數，以及一個有兩種類別的階層變數，您就會得到階層變數每個類別的二因子表格。若要製作控制變數的另一層，請按一下**下一個**，對於每個第一層變數、每個第二層變數，系統會針對其各種類別組合產生次表格，其餘依此類推。如需統計量和關聯測量，請注意，它們只適用於二因子次表。

交叉表集群長條圖

顯示集群長條圖。集群長條圖能協助您建立觀察值組別資料的摘要。您在「列」之下所指定變數的每個數值，都會有一個集群長條圖。而各集群中定義長條的變數，就是您在「行」下所指定的變數。此變數的每個值都有一組不同顏色或樣式的長條圖。如

果您在「行」或「列」下所指定的變數不止一個，那麼每兩個變數的組合就會產生一個集群長條圖。

以表格階層顯示階層變數的交叉表

以表格階層顯示階層變數。 您可以選擇將階層變數（控制變數）顯示為交叉表表格中的表格階層。這可讓您建立用以顯示列和行變數整體統計量的檢視，並允許下探階層變數的類別。

使用資料檔案 demo.sav () 的範例如下所示，並以下列方式取得：

- 選取 收入類別（千元）(inccat)」做為列變數，「擁有 PDA (ownpda)」做為行變數，並以「教育程度 (ed)」做為階層變數。
- 選取「**以表格階層顯示階層變數**」。
- 在「儲存格顯示」子對話方塊中，選取「**行**」。
- 執行「交叉表」程序，連按兩下交叉表表格，並從「教育程度」下拉式清單中選取「**大學學位**」。

圖表 5-2
表格階層中具有階層變數的交叉表表格

收入類別(千元) * 擁有PDA * 教育程度 交叉表

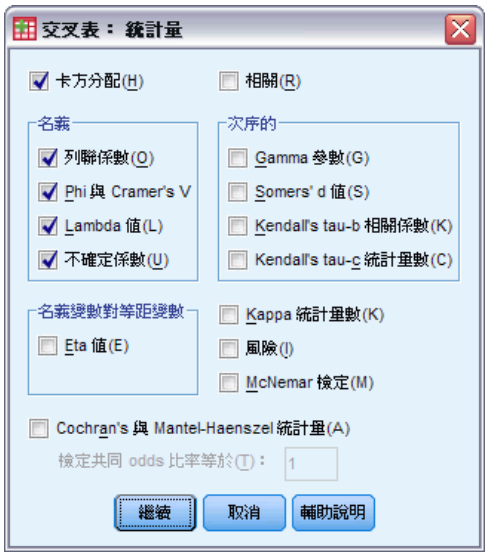
教育程度 大學

			擁有PDA		
		統計	否	是	總和
收入類別(千元)	$25以下	個數	146	50	196
		在 擁有PDA 之內的	15.8%	11.6%	14.5%
	$25 - $49	個數	335	155	490
		在 擁有PDA 之內的	36.3%	35.9%	36.2%
	$50 - $74	個數	187	72	259
		在 擁有PDA 之內的	20.3%	16.7%	19.1%
	$75以上	個數	255	155	410
		在 擁有PDA 之內的	27.6%	35.9%	30.3%
總和		個數	923	432	1355
		在 擁有PDA 之內的	100.0%	100.0%	100.0%

選取的交叉表表格檢視顯示具有大學學位之應答者的統計量。

交叉表統計量

圖表 5-3
交叉表統計量對話方塊



卡方分配。如果表格中有兩個列和兩個行，請選取「**卡方檢定**」，以計算 Pearson 卡方分配、概似比卡方分配、Fisher’s 精確檢定，以及 Yates’ 修正卡方檢定（連續修正）。但對於 2 × 2 的表格，只有當表格不是從大型表格中的遺漏列或行所產生時（該種表格儲存格的期望次數少於 5），才會計算 Fisher’s 精確檢定。會針對其他所有的 2 × 2 的表格計算 Yates’ 修正卡方。至於那些具有任何列和行個數的表格，請選取「**卡方分配**」，以計算 Pearson卡方和概似比卡方分配。此外，當這兩個表格變數都是數值時，**卡方分配**便會產生列與行變數間之線性關聯檢定。

相關。如果表格中的列和行都含有排序數值，**相關**便會產生 Spearman’s 相關係數與 rho 係數（僅數值資料）。其中，Spearman’s rho 係數是等級順序之間的關聯測量。當表格變數（因子）都是數值時，**相關**會產生 Pearson 相關係數 r 值，這是變數間線性相關的測量。

名義。如果是名義資料（它們無法真的排序，例如 Catholic（天主教徒）、Protestant（清教徒）、和 Jewish（猶太教徒）），您可以選取 **列聯係數**、**Phi（係數）**、**Cramér’s V**、**Lambda** 值（對稱性和非對稱性 Lambdas 值與 Goodman and Kruskal’s tau 測量），以及**不確定係數**。

- **列聯係數 (O)。** 以卡方為基礎的關聯性量數。數值範圍介於 0 和 1 之間，若數值為 0 則表示列變數與行變數之間沒有關聯；若數值趨近 1，則表示變數之間具有高度關聯性。可能的最大值視表格中的列與行個數而定。
- **Phi 和 Cramer’s V。** Phi 是以卡方為基礎的關聯性量數；將卡方統計量除以樣本大小，再取結果的平方根。Cramer’s V 是以卡方為基礎的關聯性量數。

- **Lambda 值。** 當使用自變數值來預測依變數值時，此關聯性量數會反映比例誤差縮減。數值 1 表示自變數可完全預測依變數。數值 0 表示自變數無法預測依變數。
- **不確定係數(U)。** 在某個變數值是用來預測其他變數值時，可表示誤差中縮減比例的關聯量數。例如，數值為 0.83 時，表示某個變數的知識會減少預測其他變數值 83 % 的誤差。程式會計算不確定係數的對稱與非對稱這兩種版本。

次序。如果表格中的列和行都含有排序數值，請選取「Gamma」(二因子表為零階，而三因子至十因子表則視情況而定)、「Kendall’s tau-b 統計測量」和「Kendall’s tau-c 統計測量」。如要從列類別中預測行類別，請選取「Somers’ d 值」。

- **Gamma。** 一種兩次序變數的對稱關聯性量數，其範圍介於 -1 和 1 之間。若數值趨近 1 的絕對值，則代表兩變數之間存有極大關係。若數值趨近於 0，則代表兩變數之間關係較疏遠或無任何關係。若為 2 因子表格，則會顯示零階 gamma。若為 3 因子至無因子表格，則會顯示條件 gamma。
- **Somers’ d 值。** 兩個次序變數間的關聯性量數，範圍介於 -1 到 1 之間。得到的數值若接近 1 的絕對值，表示兩變數之間關係極大；若數值接近 0，表示兩變數之間關係極小或沒有關係。Somers’ d 是 gamma 的不對稱延伸，唯一的不同在於它包含不與自變數等值的成對個數。系統也會計算此統計量的對稱版本。
- **Kendall’s tau-b。** 針對將同分值列入估計之次序變數或評等變數的非母數相關性量數。係數符號表示關係的方向，而其絕對值表示長度，若絕對值越大，即表示關係越密切。可能值範圍介於 -1 到 1 之間，但僅可從平方表格取得 -1 或 +1 的數值。
- **Kendall’s tau-c。** 針對忽略同分值之次序變數的非母數關聯性量數。係數符號表示關係的方向，而其絕對值表示長度，若絕對值越大，即表示關係越密切。可能值範圍介於 -1 到 1 之間，但僅可從平方表格取得 -1 或 +1 的數值。

名義變數對區間變數。當一個變數為類別，而另一個為數值時，請選取「Eta 值」。類別變數必須以數值方式來編碼。

- **Eta 值(E)。** 一種範圍從 0 至 1 的關聯性量數，其中 0 代表列變數與行變數之間沒有關聯性，而趨近於 1 的數值則代表具有高度關聯性。Eta 適用於以區間尺度測量的依變數 (例如：收入)，以及擁有一定類別個數的自變數 (例如：性別)。且會計算兩個 Eta 數值：其中一個 Eta 數值會將列變數視為區間變數，而另一個 Eta 數值則會將行變數視為區間變數。

Kappa。 當兩名評估者針對相同物件進行評等時，Cohen’s kappa 會測量其評估間的一致性。數值 1 表示完全一致。數值 0 表示一致性好壞不定。Kappa 僅適用於兩變數使用相同類別數值，且其擁有相同類別號碼的表格。

風險(R)。 對於 2 x 2 的表格，這是顯示因子和發生事件之間的關聯性強度量數。如果統計值的信賴區間包含數值 1，則您無法假設該因子與該事件是否有關聯。您可以使用 odds 比率做為當因子甚少發生時，其相對風險的估計值。

McNemar 檢定(M)。 針對兩相關二分變數的非母數檢定。使用卡方分配檢定回應值的變更。在設計為「事件前後」的實驗類型中，若要偵測因實驗中斷所導致的回應值變更，這種檢定方法非常有用。若為較大的平方表格，則會報告 McNemar-Bowker 對稱檢定。

Cochran’s 與 Mantel-Haenszel 統計量。 在由一個以上階層 (控制) 變數定義共變量樣式的條件下，Cochran’s 與 Mantel-Haenszel 統計量可用來檢定二分因子變數與二分反應變數之間的獨立性。請注意，在逐層計算其他統計量時，會針對所有階層計算一次 Cochran’s 與 Mantel-Haenszel 統計量。

交叉表儲存格顯示

圖表 5-4
交叉表儲存格顯示對話方塊

為了協助您找出有助於顯著性卡方檢定的資料樣式，「交叉表」程序會顯示期望次數和三種殘差（偏差）類型，後者可用來測量觀察和期望次數之間的差異。至於表格中的每個儲存格，則可以包含選取個數、百分比和殘差的任何組合。

個數。 如果列和行變數彼此不相關的話，此值為觀察值的實際觀察個數，以及觀察值的期望個數 。

比較行比例。 這個選項會計算行比例的成對比較，並指出哪對行（針對指定列）有顯著差異。顯著差異會在交叉表表格中以 APA 樣式格式的下標字母表示，並以 0.05 顯著水準計算。

- **調整 p 值 (Bonferroni 方法)。** 行比例的成對比較使用 Bonferroni 修正，可調整觀察顯著水準，並實際進行多重比較。

百分比。 百分比可以整列或整行增加。此外，在表格（一層）中，您也可以使用觀察值總個數的百分比。

殘差。 原始的未標準化殘差，為觀察和期望數值之間的差異。此外，您也可以用標準化和調整後標準化殘差。

- **未標準化 (U)。** 觀察值和期望值之間的差異。期望值是指當兩變數之間沒有關係時，您期望儲存格中會有的觀察值個數。如果列和行變數是獨立的，則正殘差表示儲存格中會出現超過應有的觀察值。

- **標準化(A)。** 殘差除以其標準差的估計值。標準化殘差（也稱為 Pearson 殘差）的平均數為 0，標準差為 1。
- **調整的標準化(A)。** 除以本身標準誤估計值後所得到的儲存格殘差（觀察值減去期望值）。所得的標準化殘差，是以在平均數上下的幾個標準差單位來表示。

非整數權重。 儲存格個數通常是整數，因為它們代表每個儲存格中的觀察值個數。但如果資料檔目前使用有分數值的加權變數來加權（如 1.25），則儲存格個數也會成為分數值。您可以在計算儲存格個數之前或之後，將小數的部分完全捨去或四捨五入，或在顯示表格和計算統計量時使用分數的儲存格個數。

- **捨入儲存格個數。** 系統會使用原有的觀察值加權，但會在計算任何統計量之前將儲存格中的累計加權捨入。
- **截斷儲存格個數。** 系統會使用原有的觀察值加權，但會在計算任何統計量之前將儲存格中的累計加權截斷。
- **捨入的觀察值加權。** 觀察值加權在使用前會先捨入。
- **截斷觀察值加權。** 觀察值加權會在使用後被截斷。
- **無調整(M)。** 使用原有觀察值加權與分數儲存格個數。不過，若要求「精確統計量」(僅適用於「精確檢定」選項)，則會在計算「精確」檢定統計量前，先截斷或捨入儲存格中的累計加權。

交叉表格格式

圖表 5-5
交叉表格格式對話方塊

您可以依照列變數值的遞增或遞減順序來排列列。

章節
6

摘要

「摘要」程序可計算類別中變數的次組別統計量，而此類別中有一個或多個分組變數。所有等級的分組變數都是交叉表列的。您可以選擇顯示統計量的順序。同時會顯示涵蓋所有類別的各種變數統計量。至於每個類別中的資料值，您可以選擇是否顯示出來。但是對於大型資料集，您可以選擇只列出前 n 個觀察值。

範例。 依區域和客戶企業來劃分，其平均產品銷售額為多少？您可能會發現，在西部地區的平均銷售額，比其他地區的稍微高一點，最後導致在西部地區的合作客戶，有最高的平均銷售額。

統計量。 總和、觀察值個數、平均數、中位數、組別中位數、平均數的標準誤、最小值、最大值、範圍、分組變數第一個類別的變數值、分組變數最後一個類別的變數值、標準差、變異數、 峰度、峰度的標準誤、偏態、偏態的標準誤、總和百分比、N 總數百分比、中總和百分比、中 N 的百分比、幾何平均數、調和平均數。

資料。 分組變數是其值為數值或字串的類別變數。類別的數目應該很小、但在合理的範圍內。至於其他的變數，則應該可以被分成數個等級。

假設。 某些選擇性的次組別統計量（像是平均數和標準差），乃是以常態理論為基礎。通常對稱分配的數值變數適合使用這些統計量，至於不一定符合常態性的數值變數，則適合使用如中位數、範圍之類的穩健統計量

若要取得觀察值摘要

▶ 從功能表選擇：
分析(A) > 報表 > 觀察值摘要...

圖表 6-1
「摘要觀察值」對話方塊

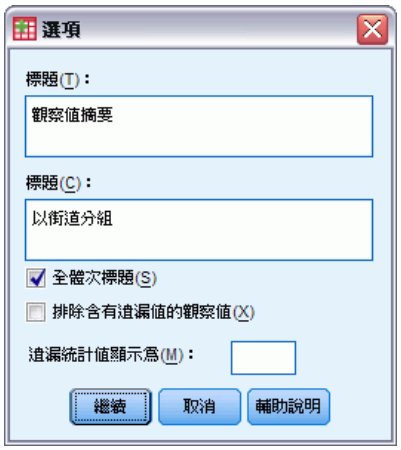
▶ 選取一個或多個變數。

您可以：

- 選取一個（或多個）分組變數，以便把資料分成數個次組別。
- 按一下「**選項**」以改變輸出標題，在輸出底下加入標題，或者排除內含遺漏值的觀察值。
- 按一下「**統計量**」，以取得選擇性的統計量。
- 選取「**顯示觀察值**」，以列出每個次組別中的觀察值。依照預設值，系統只會列出檔案中，前 100 個觀察值。另外，您也可以增減「**限制輸出前 n 個觀察值**」的值，或者取消選取該項目以列出所有的觀察值。

摘要的選項

圖表 6-2
「選項」對話方塊

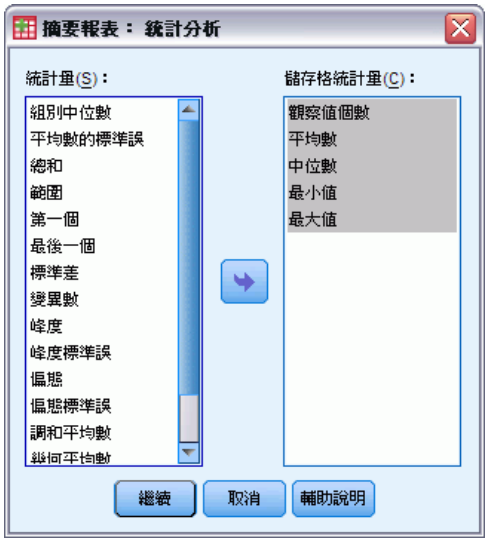


「摘要」的功能，可用來改變輸出的標題，或在輸出表格底下加入標題。如果想要控制表格標題的換行方式的話，只需在文字中，需要插入分行符號的地方，鍵入 \n 即可。

此外，您也可以選擇是否要顯示總計的次標題，或者替任何一種分析所使用的任何一個變數，排除包含遺漏值的觀察值。通常，輸出會用句點或者星號，來表示遺漏值。請輸入字元、片語或者代碼，如此碰到遺漏值時，它們就會顯示出來；不然，將不會對輸出中遺漏的觀察值，進行任何特殊處理。

摘要統計量

圖表 6-3
「摘要報表統計量」對話方塊



您可以從下列次組別統計量中，選擇其中一個或多個，作為各分組變數中每個類別變數的統計量：總和、觀察值個數、平均數、中位數、組別中位數、平均數的標準誤、最小值、最大值、範圍、分組變數第一個類別的變數值、分組變數最後一個類別的變數值、標準差、變異數、 峰度、峰度的標準誤、偏態、偏態的標準誤、總和百分比、N 總數百分比、中總和百分比、中 N 的百分比、幾何平均數、調和平均數。在「格統計量」清單中所顯示的統計量順序，為它們將在輸出中顯示的順序。不論類別為何，每個變數的摘要統計量都會顯示出來。

第一個。 顯示資料檔中出現的第一個資料值。

幾何平均數。 資料值乘積的 n 次方根，其中 n 代表觀察值的個數。

分組中位數。 針對組別中已編碼資料所計算出的中位數。例如針對年齡資料而言，若將每個 30~39 的數值編碼為 35，並將每個 40~49 的數值編碼為 45（依此類推），則分組中位數即為計算自編碼資料的中位數。

調和平均數。 此數值在組別內的樣本大小不相等時，會用來估計平均組別大小。調和平均數等於樣本總數除以樣本大小倒數總和。

峰度。 中心點四周之觀察值集群的程度量數。對常態分配而言，峰度統計量數值為零。正的峰度值表示相較於常態分配，觀察值較為集中在分配的中央，並且有延伸至分配極端值的較薄尾部，而在此時高峽峰分配的尾部則較常態分配為厚。負的峰度值表示相較於常態分配，觀察值較為分散，並且有延伸至分配極端值的較厚尾部，而在此時低闊峰分配的尾部則較常態分配為薄。

最後一個。 顯示資料檔中出現的最後一個資料值。

最大值。 數字變數的最大值。

平均數。 一種集中趨勢的量數。算術平均數，總和除以觀察值個數。

中位數。 半數觀察值落點上下的值，第 50 個百分位數。如果觀察值個數是偶數的話，中位數是中間兩個觀察值的平均值，此處的兩個中間是指，當觀察值按照遞增或遞減順序排列時，位於最中間的兩個值。中位數為集中趨勢的量數，其不會感應到偏離值（相反地，平均數會受到幾個高低極端數值所影響）。

最小值。 數字變數的最小值。

N。 觀察值個數（觀察值或記錄）。

N 總和百分比。 各類別中觀察值總數的百分比。

總和百分比。 各類別中總和的百分比。

範圍。 數值變數最大和最小值之間的差異，也就是最大值減去最小值。

偏態。 分配不對稱性的量數。常態分配是對稱的，且偏態值為 0。有顯著正偏態值的分配會有一個長的偏右尾部。有顯著負偏態值的分配會有一個長的偏左尾部。偏態值有如指標，若大於它的兩倍標準誤，則表示背離對稱。

峰度的標準誤。 峰度對其標準誤的比率可用於檢定常態性（也就是說，如果比率小於 -2 或大於 +2，則您可以否決常態性）。峰度若為大的正值，表示該分配的尾部比常態分配的尾部更長；峰度若為負值，表示尾部較短（變成類似盒形均勻分配的尾部）。

偏態的標準誤。 偏態對其標準誤的比率可用於檢定常態性（也就是說，如果比率小於 -2 或大於 +2，則您可以否決常態性）。偏態若為大的正值，表示有長的偏右尾部：極端負值則表示有長的偏左尾部。

總和。 遍及所有包含遺漏值之觀察值的數值總和或總數。

變異數。 平均數四周離散的量數，它等於平均數的平方離差總和除以觀察值個數減一。變異數的測量單位是變數本身的平方。

章節
7

平均數(M)

如果使用「平均數」程序，就可以在一個（或多個）自變數的類別之中，計算依變數的次組別平均數，及相關的單變量統計量。此外，您還可以使用單因子變異數分析、eta 值和直線性檢定。

範例。 測量三種食用油的平均脂肪吸收量，並執行單因子變異數分析，以觀察這些平均數是否不同。

統計量。 總和、觀察值個數、平均數、中位數、組別中位數、平均數的標準誤、最小值、最大值、範圍、分組變數第一個類別的變數值、分組變數最後一個類別的變數值、標準差、變異數、 峰度、峰度的標準誤、偏態、偏態的標準誤、總和百分比、N 總數百分比、中總和百分比、中 N 的百分比、幾何平均數、調和平均數。選項包括變異數分析、eta 值、eta 平方、直線性 R 及 R^2。

資料。 依變數是數值形式，而自變數是類別形式。其中類別變數的值，也可以是數字或字串。

假設。 某些選擇性的次組別統計量（像是平均數和標準差），乃是以常態理論為基礎。至於穩健統計量（如中位數），則適合用來分析數值變數（該變數可能符合，或者不符合常態性的假設）。變異數分析雖然不會受到偏離常態性的影響，但是每個儲存格中的資料，應該是呈對稱性的。變異數分析也會假設組別之母群，其變異數相同。若要檢定這個假設，請使用「單因子變異數分析」程序中的 Levene 變異數均齊性檢定。

若要取得次組別平均數

▶ 從功能表選擇：

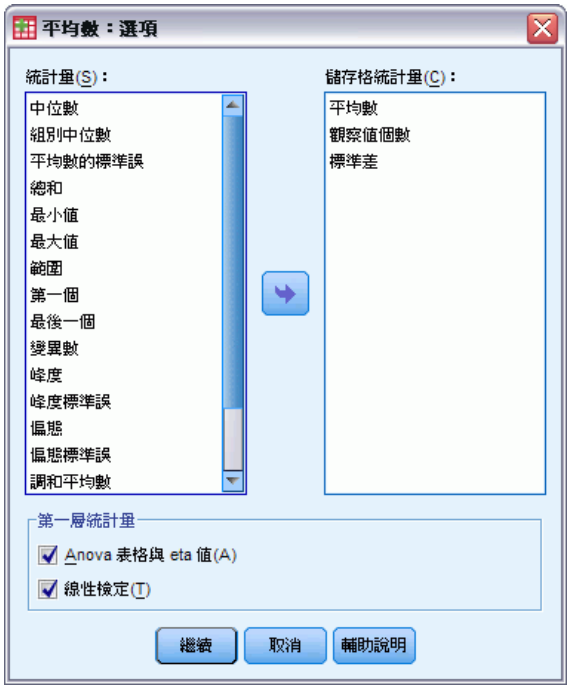
分析(A) > 比較平均數法 > 平均數...

圖表 7-1
「平均數」對話方塊

- 選擇一個或多個依變數。
- 使用下列一種方法來選擇類別自變數：
  - 選取一個或多個自變數。每個自變數顯示的結果不同。
  - 選取一層或多層自變數。每一層會再分割樣本。若您在階層 1 有一個自變數，在階層 2 有一個自變數，則結果會顯示於交叉表格中，與每個自變數的個別表格對比。
- 您也可以按一下「選項」，以取得選擇性的統計量、變異數表格分析、eta 值、eta 平方、R 和 R^2。

平均數選項

圖表 7-2
「平均數選項」對話方塊



您可以從下列次組別統計量中，選擇其中一個或多個，作為各分組變數中每個類別變數的統計量：總和、觀察值個數、平均數、中位數、組別中位數、平均數的標準誤、最小值、最大值、範圍、分組變數第一個類別的變數值、分組變數最後一個類別的變數值、標準差、變異數、 峰度、峰度的標準誤、偏態、偏態的標準誤、總和百分比、N 總數百分比、範圍總和百分比、N範圍百分比、幾何平均數及調和平均數。此外，您也可以改變次組別統計量的出現順序。其中，統計量出現在「格統計」清單中的順序，也就是它們的輸出顯示順序。不論類別為何，每個變數的摘要統計量都會顯示出來。

第一個。 顯示資料檔中出現的第一個資料值。

幾何平均數。 資料值乘積的 n 次方根，其中 n 代表觀察值的個數。

分組中位數。 針對組別中已編碼資料所計算出的中位數。例如針對年齡資料而言，若將每個 30~39 的數值編碼為 35，並將每個 40~49 的數值編碼為 45（依此類推），則分組中位數即為計算自編碼資料的中位數。

調和平均數。 此數值在組別內的樣本大小不相等時，會用來估計平均組別大小。調和平均數等於樣本總數除以樣本大小倒數總和。

峰度。 中心點四周之觀察值集群的程度量數。對常態分配而言，峰度統計量數值為零。正的峰度值表示相較於常態分配，觀察值較為集中在分配的中央，並且有延伸至分配極端值的較薄尾部，而在此時高峽峰分配的尾部則較常態分配為厚。負的峰度值表示相較於常態分配，觀察值較為分散，並且有延伸至分配極端值的較厚尾部，而在此時低闊峰分配的尾部則較常態分配為薄。

最後一個。 顯示資料檔中出現的最後一個資料值。

最大值。 數字變數的最大值。

平均數。 一種集中趨勢的量數。算術平均數，總和除以觀察值個數。

中位數。 半數觀察值落點上下的值，第 50 個百分位數。如果觀察值個數是偶數的話，中位數是中間兩個觀察值的平均值，此處的兩個中間是指，當觀察值按照遞增或遞減順序排列時，位於最中間的兩個值。中位數為集中趨勢的量數，其不會感應到偏離值（相反地，平均數會受到幾個高低極端數值所影響）。

最小值。 數字變數的最小值。

N。 觀察值個數（觀察值或記錄）。

總數 N 百分比。 各類別中觀察值總數的百分比。

總和百分比。 各類別中總和的百分比。

範圍。 數值變數最大和最小值之間的差異，也就是最大值減去最小值。

偏態。 分配不對稱性的量數。常態分配是對稱的，且偏態值為 0。有顯著正偏態值的分配會有一個長的偏右尾部。有顯著負偏態值的分配會有一個長的偏左尾部。偏態值有如指標，若大於它的兩倍標準誤，則表示背離對稱。

峰度的標準誤。 峰度對其標準誤的比率可用於檢定常態性（也就是說，如果比率小於 -2 或大於 +2，則您可以否決常態性）。峰度若為大的正值，表示該分配的尾部比常態分配的尾部更長；峰度若為負值，表示尾部較短（變成類似盒形均勻分配的尾部）。

偏態的標準誤。 偏態對其標準誤的比率可用於檢定常態性（也就是說，如果比率小於 -2 或大於 +2，則您可以否決常態性）。偏態若為大的正值，表示有長的偏右尾部：極端負值則表示有長的偏左尾部。

總和。 遍及所有包含遺漏值之觀察值的數值總和或總數。

變異數。 平均數四周離散的量數，它等於平均數的平方離差總和除以觀察值個數減一。變異數的測量單位是變數本身的平方。

第一層統計量

Anova 表格與 eta 值(A)。 顯示單因子變異數分析表，並計算第一層之每項自變數的 Eta 值與 Eta 平方值（關聯性量數）。

線性檢定(T)。 可計算與線性及非線性成份相關的平方和、自由度及平均平方，以及 F 比例、R 及 R 平方。如果自變數為短字串，就不會計算直線性。

章節
8

OLAP 多維度報表

「OLAP（線上分析處理）多維度報表」程序，可計算類別中連續摘要變數的總和、平均數、和其他單變量統計量，而此類別屬一個或多個類別分組變數。同時也會在表格中，替每個分組變數的所有類別，都產生一個階層。

範例。不同區域的總銷售額和平均銷售額，還有每個區域內，不同生產線的總銷售額和平均銷售額。

統計量。包括了：總和、觀察值個數、平均數、中位數、組別的中位數、平均數的標準誤、最小值、最大值、範圍、分組變數的第一個類別的變數值、分組變數的最後一個類別的變數值、標準差、變異數、峰度、峰度的標準誤、偏態、偏態的標準誤、總觀察值百分比、總和百分比、分組變數中總觀察值百分比、分組變數中總和百分比、幾何平均數及調和平均數。

資料。摘要變數必須是數值（依據間隔或比例尺度所測知的連續變數），而分組變數則是類別式的。其中類別變數的值，也可以是數字或字串。

假設。某些選擇性的次組別統計量（像是平均數和標準差），乃是以常態理論為基礎。至於穩健統計量(如中位數和範圍)，則適合用來分析數值變數(該變數可能符合，或者不符合常態性的假設)。

若要取得 OLAP 多維度報表

- 從功能表選擇：

  分析(A) > 報表 > OLAP 多維度報表...

圖表 8-1
「OLAP 多維度報表」對話方塊

- 選擇一個（或多個）連續摘要變數。

▶ 選擇一個（或多個）類別分組變數。

隨意：

- 選取不同的摘要統計量（按一下「**統計量**」）。選取摘要統計量之前，您必須先選取一個或多個分組變數。
- 計算成對變數和由分組變數所定義之成對組別之間的差異（按一下「**差異**」）。
- 建立自訂的表格標題（按一下「**標題**」）。

OLAP 多維度報表統計量

圖表 8-2
「OLAP 多維度報表統計量」對話方塊

您可以從下表中選擇一個（或多個）摘要變數的次組別統計量，此摘要變數分屬各分組變數的各種類別：總和、觀察值個數、平均數、中位數、組別中位數、平均數的標準誤、最小值、最大值、範圍、分組變數第一個類別的變數值、分組變數最後一個類別的變數值、標準差、變異數、峰度、峰度的標準誤、偏態、偏態的標準誤、觀察值總數的百分比、總和百分比、分組變數中的觀察值總數百分比、分組變數中的總和百分比、幾何平均數、和調和平均數。

此外，您也可以改變次組別統計量的出現順序。其中，統計量出現在「格統計」清單中的順序，也就是它們的輸出顯示順序。不論類別為何，每個變數的摘要統計量都會顯示出來。

第一個。 顯示資料檔中出現的第一個資料值。

幾何平均數。 資料值乘積的 n 次方根，其中 n 代表觀察值的個數。

分組中位數。 針對組別中已編碼資料所計算出的中位數。例如針對年齡資料而言，若將每個 30~39 的數值編碼為 35，並將每個 40~49 的數值編碼為 45（依此類推），則分組中位數即為計算自編碼資料的中位數。

調和平均數。 此數值在組別內的樣本大小不相等時，會用來估計平均組別大小。調和平均數等於樣本總數除以樣本大小倒數總和。

峰度。 中心點四周之觀察值集群的程度量數。對常態分配而言，峰度統計量數值為零。正的峰度值表示相較於常態分配，觀察值較為集中在分配的中央，並且有延伸至分配極端值的較薄尾部，而在此時高峽峰分配的尾部則較常態分配為厚。負的峰度值表示相較於常態分配，觀察值較為分散，並且有延伸至分配極端值的較厚尾部，而在此時低闊峰分配的尾部則較常態分配為薄。

最後一個。 顯示資料檔中出現的最後一個資料值。

最大值。 數字變數的最大值。

平均數。 一種集中趨勢的量數。算術平均數，總和除以觀察值個數。

中位數。 半數觀察值落點上下的值，第 50 個百分位數。如果觀察值個數是偶數的話，中位數是中間兩個觀察值的平均值，此處的兩個中間是指，當觀察值按照遞增或遞減順序排列時，位於最中間的兩個值。中位數為集中趨勢的量數，其不會感應到偏離值（相反地，平均數會受到幾個高低極端數值所影響）。

最小值。 數字變數的最小值。

N。 觀察值個數（觀察值或記錄）。

範圍 N 百分比。 其他分組變數類別內特定分組變數的觀察值個數百分比。如果您只有一個分組變數，則此數值和觀察值總數的百分比相同。

範圍總和百分比。 其他分組變數的類別內特定分組變數的總和百分比。如果您只有一個分組變數，則此數值和總和百分比相同。

N 總和百分比。 各類別中觀察值總數的百分比。

總和百分比。 各類別中總和的百分比。

範圍。 數值變數最大和最小值之間的差異，也就是最大值減去最小值。

偏態。 分配不對稱性的量數。常態分配是對稱的，且偏態值為 0。有顯著正偏態值的分配會有一個長的偏右尾部。有顯著負偏態值的分配會有一個長的偏左尾部。偏態值有如指標，若大於它的兩倍標準誤，則表示背離對稱。

峰度的標準誤。 峰度對其標準誤的比率可用於檢定常態性（也就是說，如果比率小於 -2 或大於 +2，則您可以否決常態性）。峰度若為大的正值，表示該分配的尾部比常態分配的尾部更長；峰度若為負值，表示尾部較短（變成類似盒形均勻分配的尾部）。

偏態的標準誤。 偏態對其標準誤的比率可用於檢定常態性（也就是說，如果比率小於 -2 或大於 +2，則您可以否決常態性）。偏態若為大的正值，表示有長的偏右尾部：極端負值則表示有長的偏左尾部。

總和。 遍及所有包含遺漏值之觀察值的數值總和或總數。

變異數。 平均數四周離散的量數，它等於平均數的平方離差總和除以觀察值個數減一。變異數的測量單位是變數本身的平方。

OLAP 多維度報表差異

圖表 8-3
「OLAP 多維度報表差異」對話方塊

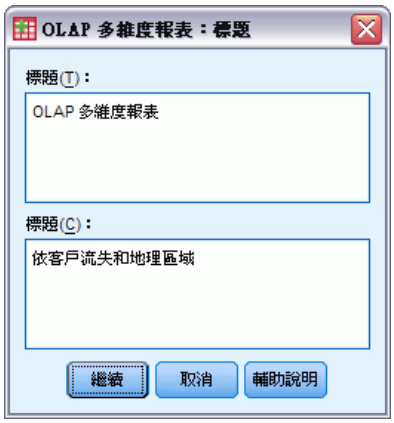
這個對話方塊可讓您計算摘要變數之間或分組變數所定義之組別之間的百分比或算術差異。程式會計算「OLAP 多維度報表統計量」對話方塊中選擇之所有量數的差異。

變數間之差異。計算成對變數之間的差異。將每一對中第一個變數的摘要統計量值減去該對第二個變數（減變數）的摘要統計量值。若計算百分比差異，將使用減變數的摘要變數值做為分母。指定變數之間的差異之前，您必須先在主對話方塊中選取至少兩個摘要變數。

觀察值組別間差異。計算分組變數所定義之成對組別之間的差異。將每一對中第一個類別的摘要統計量值減去該對第二個類別（減組類）的摘要統計量值。百分比差異使用減組類的摘要統計量值做為分母。指定組別之間的差異之前，您必須先在主對話方塊中選取一個或多個分組變數。

OLAP 多維度報表標題

圖表 8-4
「OLAP 多維度報表標題」對話方塊



您可以改變輸出的標題，也可以加入標題，它會出現在輸出表格底下。此外，如果想控制各種標題的換行方式，只需在文字中、任何想要插入分行符號的地方，鍵入 \n 即可。

章節
9

T 檢定

可以使用的 t 檢定類型有三種：

獨立樣本 T 檢定（二樣本 t 檢定）。 比較兩組觀察值中，某個變數的平均數。它也會提供每個群組的敘述統計、變異數相等的 Levene 檢定、相等和不等變異數的 t 值，和平均數差異的 95% 信賴區間。

成對樣本 T 檢定（相依 t 檢定）。 比較單一組別中，兩個變數的平均數。這個檢定也可以用來研究配對組別或者控制觀察值。其輸出包括：檢定變數的敘述統計、它們之間的相關性、配對差異的敘述統計、t 檢定，以及 95% 信賴區間。

單一樣本 T 檢定。 它會把某個變數的平均數，跟已知數值（或假設值）相比較。檢定變數的敘述統計會連同 t 檢定一併顯示。部分預設輸出包括：檢定變數平均數間差異的 95% 信賴區間、假設的檢定值等。

獨立樣本 T 檢定

「獨立樣本 T 檢定」程序，乃是用來比較兩組觀察值的平均數。理論上，這個檢定的受試者應該隨機地指定給兩個小組，這樣一來，反應差異都是來自處理方式（或者未予處理），而不是其他因素造成的。如果您要比較男性和女性的平均收入，事情就不是這樣。因為您不可能把某個人隨機地指定成男性或女性。在這種情況下，您應該確定其他因素中的差異，不會顯著地遮蔽或加強平均數的差異。因為像教育之類的因素也會影響到平均收入，而不只是性別一項而已。

範例。 以高血壓患者為例。我們把高血壓患者隨機地指定成安慰劑組和治療組。安慰劑受試者會收到沒有作用的藥丸，而治療組的受試者則收到預期會降低血壓的新藥。經過兩個月的治療之後，我們使用二樣本 t 檢定，來比較安慰劑組和治療組受試者的平均血壓。每位患者測量一次，而且只屬於一組。

統計量。 對於每個變數來說，統計量包括：樣本大小、平均數、標準差，以及平均數的標準差。對於平均數間的差異而言，統計量包括：平均數、標準誤，和信賴區間（您可以指定信賴水準）。就檢定方面來說，包括：變異性相等的 Levene 檢定，以及平均數相等的綜合變異數和個別變異數 t 檢定。

資料。 相關的數值變數值，都存在資料檔中的單一欄位。此程序會使用分組變數（其內含有兩個數值），把觀察值分成兩組。分組變數可以是數字（如 1 跟 2 或 6.25 跟 12.5），也可以是短字串（如 yes 和 no）。不然還有種方式，就是您可以用 age 之類的數值變數，並指定其分割點，把觀察值分成兩組（分割點 21 會把 age 分成 21 歲以下的人一組，等於或大於 21 歲的人分成另一組）。

假設。 對相等變異數 t 檢定而言，觀察值應該是來自獨立、隨機的樣本，而樣本母體又是呈現常態分配、而且變異數相同。對於不等變異數 t 檢定而言，觀察值應該是來自獨立、隨機的樣本，而樣本母體又是呈現常態分配的。二樣本 t 檢定雖然不

太會受到偏離常態性影響。但在以圖形方式檢查分配情形時，請查看它們是否對稱，以及是否包含離群值。

若要取得獨立樣本 T 檢定

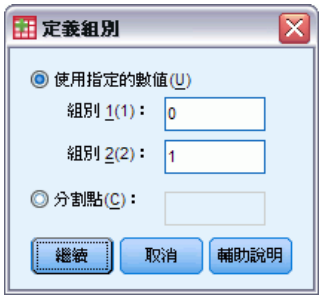
- 從功能表選擇：
  分析(A) > 比較平均數法 > 獨立樣本 T 檢定...

圖表 9-1
「獨立樣本 T 檢定」對話方塊

- 選取一個或多個數值檢定變數。每個變數都會分別計算 t 檢定。
- 選取單一分組變數，然後再按一下「定義組別」，如此可替需要比較的組別，指定兩個代碼。
- 或者，按一下「選項」，以便控制處理遺漏資料的方式和信賴區間的水準。

獨立樣本 T 檢定的定義組別

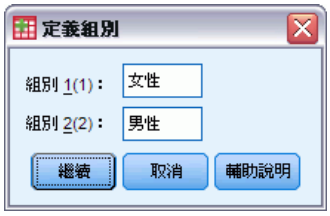
圖表 9-2
數字變數的「定義組別」對話方塊



對於數字分組變數而言，請指定兩個數值或是分割點，以定義 t 檢定的兩個組別：

- **使用指定的數值。**請在組別 1 輸入一個數值，在組別 2 輸入另一個數值。 觀察值如果包含其他數值的話，則都不予以分析。此外，數字不必是整數（例如 6.25 或 12.5 都算有效數字）。
- **分割點。** 您可以輸入一個數字，以便把分組變數值，分成兩個小組。觀察值如果小於分割點的話，就被分成一組，而大於或等於分割點的觀察值，則被分成另外一組。

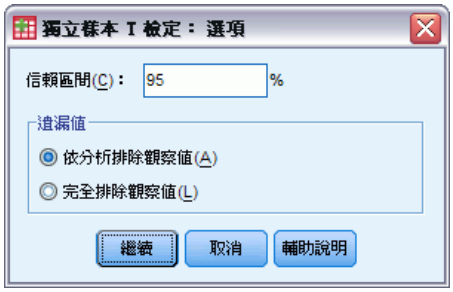
圖表 9-3
字串變數的「定義組別」對話方塊



對於字串分組變數，請替組別 1 輸入一個數值，再替組別 2 輸入另一個數值（如 yes 和 no）。含有其他字串的觀察值會從分析中排除。

獨立樣本 T 檢定的選項

圖表 9-4
「獨立樣本 T 檢定選項」對話方塊



信賴區間。 依照預設值，顯示平均數間差異的 95% 信賴區間。請輸入一個介於 1 到 99 之間的數值，就可以要求不同的信賴水準。

遺漏值。 當您檢定數個變數，而且其中一個（或多個）變數又含有遺漏值時，您可以告訴程序，哪些觀察值需要加進來，或者予以排除。

- **依分析排除觀察值。** 每個 t 檢定所使用的觀察值，對檢定變數而言，其內資料必須是有效的。樣本大小可能會隨著檢定類型而改變。
- **完全排除遺漏值。** 每個 t 檢定所使用的觀察值，對於任何 t 檢定會用到的變數而言，其內資料必須是有效的。樣本大小不會隨著檢定類型而改變。

成對樣本 T 檢定

「成對樣本 T 檢定」程序，乃是用來比較單一組別中兩個變數的平均數。它會計算每個觀察值兩個變數值之間的差異，以及檢定平均是否為 0。

範例。 以高血壓的研究為例。 本次研究中，所有患者都會在研究開始時，先量一次血壓，接受治療之後，再量一次。因此，每位受試者都會有兩個測量值，通常稱為前和後測量值。使用這個檢定的一個其他設計為配對組別或控制觀察值的研究，此研究中資料檔案的每一筆記錄都包含病患，以及與其搭配之控制受試者的反應。以血壓研究為例，患者組和控制組可以根據年齡來配對（75 歲患者搭配 75 歲的控制組成員）。

統計量。 對於每個變數來說，統計量包括：樣本大小、平均數、標準差，以及平均數的標準差。對於平均數間的差異而言，統計量包括：相關、平均數中的平均差異、t 檢定和平均數差異的信賴區間（您可以指定信賴水準）。此外，還有標準差和平均數差異的標準誤。

資料。 對於每個成對檢定而言，請指定兩個數值變數（測量值的間隔水準或測量值的比例水準）。研究如果需要配對組別，或者控制觀察值的話，則每位檢定受試者的反應，還有與其搭配的控制受試者的反應，都必須存在資料檔的同一個觀察值中。

假設。 每一對觀察值應該是在相同的條件下，產生出來的。平均數差異應該是常態分配。但是每個變數的變異數不需要相等。

若要取得成對樣本 T 檢定

▶ 從功能表選擇：

分析(A) > 比較平均數法 > 成對樣本 T 檢定...

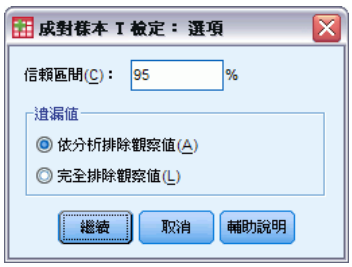
圖表 9-5
「成對樣本 T 檢定」對話方塊

▶ 選取一或多對變數

▶ 或者，按一下「選項」，以便控制處理遺漏資料的方式和信賴區間的水準。

成對樣本 T 檢定的選項

圖表 9-6
「成對樣本 T 檢定選項」對話方塊



信賴區間。 依照預設值，顯示平均數間差異的 95% 信賴區間。請輸入一個介於 1 到 99 之間的數值，就可以要求不同的信賴水準。

遺漏值。 當您檢定數個變數，而且其中一個（或多個）變數又含有遺漏值時，您可以在程序中指定要加入哪些觀察值，或者予以排除：

- **依分析排除觀察值。** 每個 t 檢定所使用的觀察值，對檢定的變數對而言，其內資料必須是有效的。樣本大小可能會隨著檢定類型而改變。
- **完全排除遺漏值。** 每個 t 檢定所使用的觀察值，對於任何 t 檢定會用到的變數而言，其內資料必須是有效的。樣本大小不會隨著檢定類型而改變。

單一樣本 T 檢定

「單一樣本 T 檢定」程序，乃是用來檢定單一變數的平均數，是否跟指定的常數不一樣。

範例。 研究人員可能想檢定，一組學生的平均 IQ 是否為 100。 或者，穀類早餐製造商可能從生產線抽選幾盒，當作樣本，檢定它們的平均重量在 95% 信賴水準下，是否為 1.3 磅。

統計量。 對每個變數來說，統計量包括：樣本大小、平均數、標準差，以及平均數的標準差。此外還有：每個資料值和假設檢定值之間的平均差異，檢定這個差異是否為 0 的 t 檢定，以及這個差異的信賴區間（您可以指定信賴水準）。

資料。 若要依據一個假設的檢定值，來檢定數值變數的值，請選擇一個數值變數，然後再輸入一個假設檢定值。

假設。 這個檢定假設資料是常態性分配；但是，這個檢定不太會受到偏離常態性的影響。

若要取得單一樣本 T 檢定

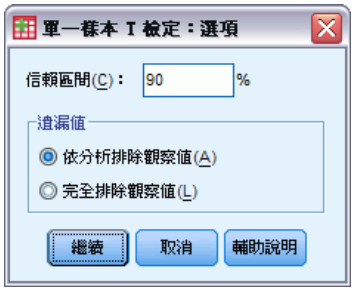
▶ 從功能表選擇：
分析(A) > 比較平均數法 > 單一樣本 T 檢定...

圖表 9-7
「單一樣本 T 檢定」對話方塊

- 選擇一個（或多個）變數，以便根據同一個假設值進行檢定。
- 請輸入一個數字檢定值，以便根據該值，比較每個樣本平均數。
- 或者，按一下「選項」，以便控制處理遺漏資料的方式和信賴區間的水準。

單一樣本 T 檢定的選項

圖表 9-8
「單一樣本 T 檢定」對話方塊



信賴區間。 依照預設值，顯示平均數和假設檢定值間差異的 95% 信賴區間。請輸入一個介於 1 到 99 之間的數值，就可以要求不同的信賴水準。

遺漏值。 當您檢定數個變數，而且其中一個（或多個）變數又含有遺漏值時，您可以告訴程序，哪些觀察值需要加進來，或者予以排除。

- **依分析排除觀察值。** 每個 t 檢定所使用的觀察值，對檢定變數而言，其內資料必須是有效的。樣本大小可能會隨著檢定類型而改變。
- **完全排除遺漏值。** 每個 t 檢定所使用的觀察值，對於任何 t 檢定會用到的變數而言，其內資料必須是有效的。樣本大小不會隨著檢定類型而改變。

T-TEST 指令的其他功能

指令語法語言也可以讓您：

- 在執行單一指令時，同時產生單一樣本和獨立樣本 t 檢定。
- 在成對 t 檢定中，依據清單中的每個變數來檢定變數（使用 PAIRS 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
10

單因子變異數分析

「單因子變異數分析」程序，會根據單一因子變數（自變數），來產生數量依變數的單因子變異數分析。變異數分析是一種假設，用於檢定數個平均數是否相等。這項技術是二個樣本 t 檢定的延伸。

除了確定平均數之間是否存在差異之外，您可能想知道哪個平均數不一樣。比較平均數的檢定類型有兩種：演繹式對比和 post hoc 檢定。對比是在進行實驗之前設定的檢定，而 post hoc 檢定則是在實驗完畢之後進行。您也可以檢定不同類別之間的趨勢。

範例。甜甜圈在調理的過程中，是否會吸收不等量的油脂?我們在實驗中設定三種油脂類型：花生油、玉米油、和豬油。花生油和玉米油是不飽和脂肪，豬油是飽和性脂肪。除了判斷油脂的吸收量，是否與所用的油脂種類有關之外，您還可以設定演繹式對比，來判斷飽和脂肪與不飽和脂肪的油脂吸收量是否不相同。

統計量。對每個組別而言，共有：觀察值個數、平均數、標準差、平均數的標準誤、最小值、最大值、和平均數的 95% 信賴區間。變異數均齊性的 Levene 檢定、各依變數的變異數分析摘要表和平均數相等性穩健檢定、使用者指定的演繹式對比、post hoc 全距檢定、和多重比較：Bonferroni 法、Sidak 檢定、Tukey 最誠實顯著性差異、Hochberg’s GT2 檢定、Gabriel 檢定、Dunnett 檢定、Ryan-Einot-Gabriel-Welsch F 檢定 (R-E-G-W F 值)、Ryan-Einot-Gabriel-Welsch 全距檢定 (R-E-G-W Q 值)、Tamhane’s T2 檢定、Dunnett’s T3 檢定、Games-Howell 檢定、Dunnett’s C 檢定、Duncan’s 多重全距檢定、Student-Newman-Keuls (S-N-K) 檢定、Tukey’s b 檢定、Waller-Duncan 檢定、Scheffé 法和最小顯著差異。

資料。因子變數值應該為整數，而依變數應該是數值變數（測量的區間水準）。

假設。每個組別都是取自常態母群體的獨立隨機樣本。雖然資料應該是對稱的，但變異數分析不受偏離常態性的影響。各組別應該來自具相等變異數的母群體。若要檢定這個設設，請使用 Levene 的變異數均齊性檢定。

若要取得單因子變異數分析

▶ 從功能表選擇：

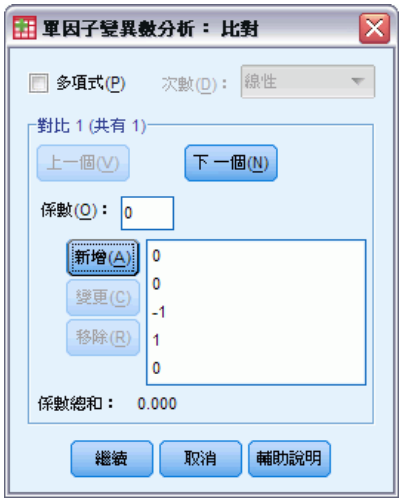
分析(A) > 比較平均數法 > 單因子變異數分析...

圖表 10-1
「單因子變異數分析」對話方塊

- 選擇一個或多個依變數。
- 選擇一個獨立因子變數。

單因子變異數分析的對比

圖表 10-2
「單因子變異數分析對比」對話方塊



您可以把組間平方和區隔成趨勢成份，或指定演繹式對比。

多項式。將組間平方和區隔成趨勢成份。您可以檢定依變數在按照順序排列的因子變數水準中，它們的趨勢。例如，您可以檢定在最高學歷階層中（已經排序過），所獲薪水的線性趨勢（升冪或降冪）。

- **次數。**您可以選擇第一、第二、第三、第四或第五次多項式。

係數。t 統計量所要檢定的使用者定義演繹式對比。請為因子變數的各組別（類別）輸入一個係數，並在每次選入之後按一下**新增**。每個新值會加到係數清單的底部。若要指定其他的對比集合，請按一下**下一個**。使用**下一個**和**前一個**就可以在對比集合間移動。

係數的順序很重要，因為它跟因子變數類別數值的升冪順序互相對應。清單中的第一個係數，會對應至因子變數組別中的最低值，而最後一個係數則跟最高值互相對應。例如，如果因子變數有六個類別的話，那麼係數 -1、0、0、0、0.5 和 0.5，就會將第一組跟第五、第六組互相對照。對大部分的應用程式而言，係數的總和應該為 0，如果不是 0，這個集合還是可以使用，但是會顯示警告訊息。

單因子變異數分析的 Post Hoc 檢定

圖表 10-3
「單因子變異數分析 Post Hoc 多重比較」對話方塊

一旦您判斷平均數之間的確存有差異之後，post hoc 全距檢定和成對多重比較便可以決定到底是哪些平均數不一樣。您可以用全距檢定，來判斷哪些是等值平均數的均勻子集。成對多重比較則檢定每對平均數之間的差異，並且產生一個矩陣。 該矩陣以星號代表顯著不同的組別平均數，其 alpha 水準為 0.05。

假設相同的變異數

Tukey 最誠實顯著性差異、Hochberg’s GT2 檢定、Gabriel 檢定和 Scheffé 檢定既是多重比較檢定，也是全距檢定。其他可以使用的全距檢定還包括：Tukey’s b 檢定、S-N-K (Student-Newman-Keuls 多重比較法)、Duncan、R-E-G-W F 值 (Ryan-Einot-Gabriel-Welsch F 值檢定)、R-E-G-W Q 值 (Ryan-Einot-Gabriel-Welsch 全距檢定) 和 Waller-Duncan 檢定。可以使用的多重比較檢定有 Bonferroni 法、Tukey 最誠實顯著性差異檢定、Sidak 法、Gabriel 檢定、Hochberg 檢定、Dunnett 檢定、Scheffé 法和 LSD (最小顯著差異)。

- **LSD(L)。** 使用 t 檢定執行組別平均數之間的所有成對比較。不會調整多重比較的錯誤率。
- **Bonferroni。** 使用 t 檢定執行組別平均數間的成對比較，並且控制整體錯誤率，其方法是將每項檢定的錯誤率設為除以總檢定個數後得出的實驗錯誤率。因此，觀察顯著水準會經過調整，並實際進行多重比較。

- **Sidak 檢定(I)。** 以 t 統計量為基礎的成對多重比較檢定。Sidak 會調整多重比較的顯著性水準，並提供比 Bonferroni 更緊密的界線。
- **Scheffe。** 為所有可能的成對平均數組合進行同步聯合成對比較。使用 F 取樣分配。可用於檢查組別平均數的所有線性組合，並不是只進行成對比較。
- **R-E-G-W F 值(R)。** 根據 F 檢定的 Ryan-Einot-Gabriel-Welsch 多重減期程序。
- **R-E-G-W Q 值(Q)。** 根據 Studentized 範圍的 Ryan-Einot-Gabriel-Welsch 多重減期程序。
- **S-N-K(S)。** 會使用 Studentized 範圍分配在平均數之間進行所有的成對比較。在樣本大小相同的情況下，它也會使用逐步程序來比較同性質子集中成對的平均數。平均數會從最高到最低來排序，而且會先檢定極端差分。
- **Tukey 法(T)。** 使用 Studentized 範圍統計量在組別之間進行所有的成對比較。會為所有成對比較集的錯誤率設定實驗誤差率。
- **Tukey's b。** 使用 Studentized 範圍分配在組別之間進行成對比較。關鍵值為 Tukey's 誠實顯著差異檢定和 Student-Newman-Keuls 的平均對應值。
- **Duncan(D)。** 運用與 Student-Newman-Keuls 檢定所用順序相同的逐步比較順序，來進行成對比較，不過會針對檢定集合錯誤率（非個別檢定錯誤率）設定保護水準。使用 Studentized 範圍統計量。
- **Hochberg's GT2 檢定(H)。** 採用 Studentized 最大絕對值的多重比較和範圍檢定。其與 Tukey's 最誠實顯著性差異檢定類似。
- **Gabriel 檢定(G)。** 採用 Studentized 最大絕對值的成對比較檢定，一般而言，當儲存格大小不相等時，此檢定較 Hochberg's GT2 更具強大功能。當儲存格大小有相當大的不同時，Gabriel's 檢定可能會變成不拘泥形式的。
- **Waller-Duncan 檢定(W)。** 依據 t 統計量的多重比較檢定，使用的是 Bayesian 方法。
- **Dunnett。** 成對多重比較 t 檢定，可針對單一控制平均數來比較處理組合。最後一個類別是預設的控制類別。當然您也可以改用第一個類別。**雙邊檢定**可檢定因子在任何水準（除了控制類別以外）的平均數跟控制類別的平均數是否相等。**<控制**可檢定因子在任何水準的平均數是否小於控制類別的平均數。**>控制**可檢定因子在任何水準的平均數是否小於控制類別的平均數。

未假設相同的變異數

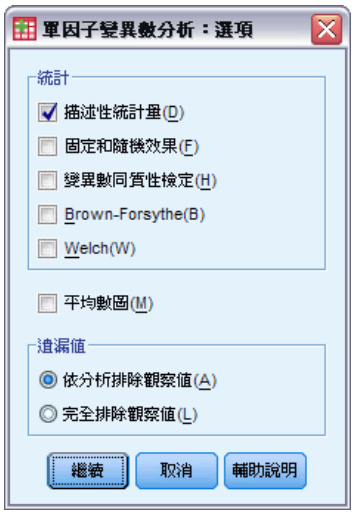
對於多重比較檢定而言（它們不會假設變異數相等），則有：Tamhane's T2 檢定、Dunnett's T3 檢定、Games-Howell 檢定和 Dunnett's C 檢定。

- **Tamhane's T2 檢定(M)。** 以 t 檢定為基礎的 Conservative 成對比較。變異數不相等時適合使用此檢定。
- **Dunnett's T3 檢定(3)。** 以 Studentized 最大絕對值為基礎的成對比較檢定。變異數不相等時適合使用此檢定。
- **Games-Howell 檢定(A)。** 成對比較檢定有時候會顯得不夠嚴謹。變異數不相等時適合使用此檢定。
- **Dunnett's C 檢定(U)。** 以 Studentized 範圍為基礎的成對比較檢定。變異數不相等時適合使用此檢定。

注意：如果在「表格性質」對話方塊（在啟動的樞軸表中，選擇「格式」功能表上的「**表格性質**」）中取消選擇「**隱藏空白的列與欄**」，您可能會發現這樣比較容易說明 post hoc 檢定的輸出。

單因子變異數分析的選項

圖表 10-4
「單因子變異數分析選項」對話方塊



統計量。 選擇一個（或多個）以下項目：

- **描述性統計量。**計算觀察值個數、平均數、標準差、平均數的標準誤、最小值、最大值、和各組別每個依變數的 95% 信賴區間。
- **固定和隨機效應。**顯示固定效應的標準差、標準誤和 95% 信賴區間，以及隨機效應的標準誤、95% 信賴區間和成分間變異數之估計。
- **變異數均齊性檢定。** 計算 Levene 統計量來檢定組別變異數的相等性。這個檢定跟常態性假設無關。
- **Brown-Forsythe。**計算 Brown-Forsythe 統計量以檢定組別平均數的相等性。當變數異相等的假設不成立時，一般慣用這個統計量，而不使用 F 統計量。
- **Welch。**計算 Welch 統計量以檢定組別平均數的相等性。當變數異相等的假設不成立時，一般慣用這個統計量，而不使用 F 統計量。

平均數圖。顯示一個圖表，該表會繪出次組別的平均數（根據因子變數值，來定義各組別的平均數）。

遺漏值。控制處理遺漏值的方式。

- **依分析排除觀察值。**對於該項分析而言，如果依變數或因子變數的某個觀察值含有遺漏值的話，它就不會使用這個觀察值。此外，如果觀察值超過因子變數所指定的範圍，也不會被分析。
- **完全排除遺漏值。**如果因子變數、或主對話方塊中，依變數清單所含的任何依變數，其觀察值含有遺漏值的話，則所有分析都會將這些觀察值排除在外。如果您未指定多重依變數的話，這個選項無效。

ONEWAY 指令的其他功能

指令語法語言也可以讓您：

- 取得固定和隨機效應的統計量。固定效應模式果標準差、平均數的標準誤、和 95% 信賴區間。隨機效應模式的標準誤、95% 信賴區間、和元件變異數間的估計值（使用 STATISTICS=EFFECTS）。
- 指定 alpha 水準的最小顯著差異、Bonferroni 法、Duncan、和 Scheff é多重比較檢定（使用 RANGES 次指令）。
- 寫入平均數、標準差、和次數分配的矩陣，或讀取平均數、次數分配、綜合變異數、和綜合變異數的自由度的矩陣。這些矩陣可以用來取代原始資料，以取得單因子變異數分析（使用 MATRIX 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
11

GLM 單變量分析

「GLM 單變量」程序，可對一個依變數進行迴歸分析與變異數分析，而此依變數可對應一個或多個因子(和/或變數)。因子變數會將母群加以分組。使用這個「一般線性模式」程序，您可以檢定單一依變數的變數群組平均數上，有關其他變數效果的虛無假設。您可以調查因子之間的交互作用、以及個別因子的效果，這其中有部分是隨機的。此外，您還可以研究共變量的效應，以及共變量與因子之間的交互作用。在迴歸分析中，自變數（預測變數）會被指定成共變量。

您也可以檢定平衡或不平衡的模式。所謂平衡模式，就是模式中的每個儲存格所含之觀察值個數相同。除了檢定假設外，「GLM 單變量」還會產生參數的估計值。

您可以用常見的演繹式對比來檢定受試者間因子的假設。此外，當您發現全面 F 檢定結果是顯著的時後，就可以用 post hoc 檢定來評估指定平均數之間的差異。邊際平均數估計值會算出模式內儲存格的預測平均值，而這些平均數的剖面圖（交互作用圖）可以讓您很輕鬆地以目視的方式，看出部分的關係。

您可以將殘差、預測值、Cook’s 距離、影響量數當成新的變數，存入資料檔中，以便驗證假設。

「加權最小平方法之權數」能讓您指定變數，此變數將提供觀察值不同的加權最小平方 (WLS) 分析的加權數，或許能補償不同的測量精確度。

範例。 我們收集數年來，芝加哥地區馬拉松跑者的個人相關資料。每位跑者完成的時間，屬於依變數。其他因子包括天氣（冷、溫和、或熱）、訓練的月數、之前跑過的馬拉松次數以及性別。年齡視為共變量。您可能會發現性別的影響是顯著的，以及性別與天氣的交互作用也非常顯著。

方法。 您可以使用類型 I、類型 II、類型 III 和類型 IV 平方和來評估多種假設。預設值為類型 III。

統計量。 對 Post hoc 範圍檢定和多重比較而言，共有：最小顯著差異、Bonferroni 法、Sidak 檢定、Scheffé 法、Ryan-Einot-Gabriel-Welsch 多重 F 檢定、Ryan-Einot-Gabriel-Welsch 多重範圍、Student-Newman-Keuls 檢定、Tukey 最誠實顯著性差異、Tukey’s b、Duncan、Hochberg’s GT2 檢定、Gabriel 檢定、Waller-Duncan t 檢定、Dunnett（單邊和雙邊）、Tamhane’s T2 檢定、Dunnett’s T3 檢定、Games-Howell 檢定，以及 Dunnett’s C 檢定。敘述統計：對敘述統計而言，則有：觀察平均數、標準差和所有儲存格中全部依變數的個數變異數均齊性的 Levene 檢定。

圖形。 包括離散對水準之圖形、殘差圖、剖面圖（交互作用）。

資料。 依變數是數值變數。因子是類別的。它們可以是數值或最多八個字元的字串值。而共變量是與依變數相關的數值變數。

假設。 資料是採自常態母群體的隨機樣本；在母群體中，所有儲存格變數都是相同的。雖然資料應該是對稱的，但變異數分析不受偏離常態性的影響。若要檢查假設，您可以使用變異數均齊性檢定和離散對水準之圖形。您也可以檢驗殘差和殘差圖。

若要取得 GLM 單變量表格

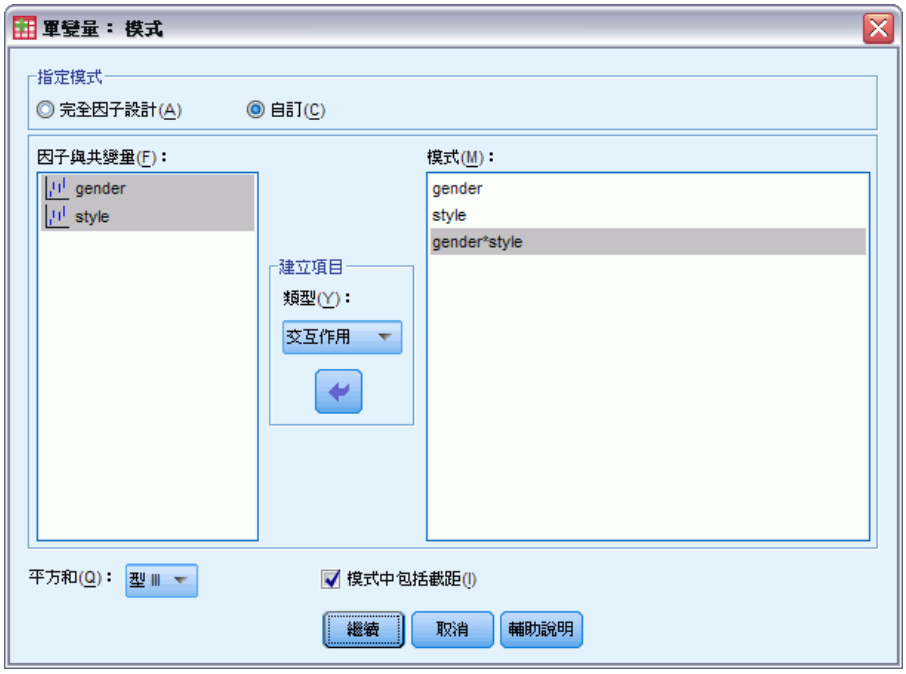
- 從功能表選擇：
  分析(A) > 一般線性模式 > 單變量...

圖表 11-1
「GLM 單變量」對話方塊

- 選取依變數。
- 為「固定因子」、「隨機因子」和「共變量」，選取適合您資料的變數。
- 您也可以選擇性使用「加權最小平方之權數」，來指定加權最小平方法分析的加權變數。如果加權變數為零、負數或遺漏值，那麼分析時，會排除這個觀察值。已用於模式中的變數，不能再當作加權變數使用。

GLM 模式

圖表 11-2
「單變量模式」對話方塊



指定模式。 完全因子模式包括所有因子主效果、所有共變數主效果、以及所有因子對因子交互作用。但卻不包含共變量的交互作用。若要只指定交互作用子集，或指定不同共變量之因子的交互作用，請選擇「自訂」。但是您必須指出所有會被放入模式的項目。

因子與共變量。因子與共變量均會列出。

模式。 模式端視您的資料性質而定。在選擇「自訂」之後，您就可以選擇欲分析之主效果和交互作用。

平方和。 計算平方和之方法。對於不含遺漏值的平衡或不平衡模式而言，類型 III 平方和法是最常用的方法。

模式中通常會包括截距。 模式中通常會包括截距，但是如果假設資料會穿過原點的話，就可以將截距排除在外。

建立效果項

對所選擇的因子和共變量而言：

交互作用。 建立所有選定變數的最高階交互作用項。此為預設值。

主效果。 為每個選擇的變數，建立主效果。

完全二因子。為所選的變數，建立所有可能的二因子交互作用。

完全三因子。為所選的變數，建立所有可能的三因子交互作用。

完全四因子。為所選的變數，建立所有可能的四因子交互作用。

完全五因子。為所選的變數，建立所有可能的五因子交互作用。

平方和

對於此種模式，您可以選擇一種平方和。其中，類型 III 最常使用，也是預設值。

類型 I。 這個方法也稱為平方和方法的階層式分解。模式中的每一項都只能針對它的前一項加以調整。類型 I 平方和常用於下列情形：

- 在平衡的 ANOVA 模式中，任何主效應都應在任何第一階交互作用效應之前指定，而任何第一階交互作用效應都需在任何第二階交互作用效應之前指定，然後依此類推。
- 在多項式迴歸模式中，您必須在指定較高階項之前，先指定較低階項。
- 在純巢狀模式中，第一個指定的效應會套在第二個指定的效應裏，第二個指定的效應會套在第三個裏，依此類推。(這種巢狀形式只能透過語法來指定)。

類型 II。 在已為所有其他「適當」效果調整的模式中，此方法可計算該模式中效果的平方和。適當的效果是指與所有不包含要檢驗效果的效果相對應的效果。類型 II 平方和方法通常用於：

- 平衡的變異數分析模式。
- 任何只有主因子效果的模式。
- 任何迴歸模式。
- 純巢狀設計(這種巢狀形式能透過語法來指定)。

類型 III。 預設值。這個方法是用來計算設計中某個效應的平方和，此乃其他效應（不包含該效應）和與任何包含它的效應（如果有的話）正交調整後的平方和。類型 III 平方和的主要優點在於：只要估計的一般形式保持不變，它們在儲存格次數方面就是不變的。所以一般認為，這個平方和類型對於沒有遺漏儲存格的不平衡模式而言，是相當好用的。在沒有遺漏儲存格的因子設計中，這個方法等於是「Yates 加權平方和」的技術。類型 III 平方和方法通常用於：

- 任何類型 I 和類型 II 中列出的模式。
- 任何沒有空儲存格的平衡（或不平衡）模式。

類型 IV。 此方法是為了遺漏儲存格的狀況設計的。對於設計中的任何效果 F，如果任何其他效果中不包含 F，則類型 IV = 類型 III = 類型 II。當其他效果中包含 F 時，類型 IV 會將在 F 參數之間產生的對比平均的分配在所有較高階效果中。類型 IV 平方和方法通常用於：

- 任何類型 I 和類型 II 中列出的模式。
- 任何有空儲存格的平衡或不平衡模式。

GLM 對比

圖表 11-3
「單變量對比」對話方塊

對比是用以檢定因子水準之間的差異。您可以在模式中指定各因子的對比（若要在各受試者因子之間進行對比，請使用重複量數模式）。對比代表參數的線性組合。

假設檢定是根據虛無假設 LB = 0，此處 L 是對比係數矩陣，而 B 是參數向量。在指定對比時，會建立 L 矩陣。與因子相對應的 L 矩陣的行符合對比。剩餘的行會被調整，不然無法估計 L 矩陣。

每組對比的輸出中都包含一個 F 統計量。同時顯示的對比差異，還有根據 Student’s t 分配的 Bonferroni 類型同時信賴區間。

可用對比

可供您使用的對比包括離差、簡單、差分、Helmert、重複和多項式。對於離差和簡單對比而言，您可以選擇是否讓參考類別變成第一個（或最後一個）類別。

對比類型

離差。 比較每個水準的平均數（除了參考類別）跟所有水準的平均數（總平均）。因子水準以任何一種方式排列都可以。

簡單。 比較每個水準的平均數與指定水準的平均數。這類對比在有控制組時相當有用。您可以選擇第一個或最後一個類別當做參考。

差分。 比較每個水準的平均數（除了第一個）與先前水準的平均數。（這種對比有時候稱為反 Helmert 對比）。

Helmert。 比較每個因子水準的平均數（除了最後一個）與隨後水準的平均數。

重複。 比較每個水準的平均數（除了最後一個）與隨後水準的平均數。

多項式。 比較線性效應、二次效應、三次效應，依此類推。第一自由度包含所有類別的線性效應；第二自由度包含二次效應；依此類推。這些對比常用來估計多項式趨勢。

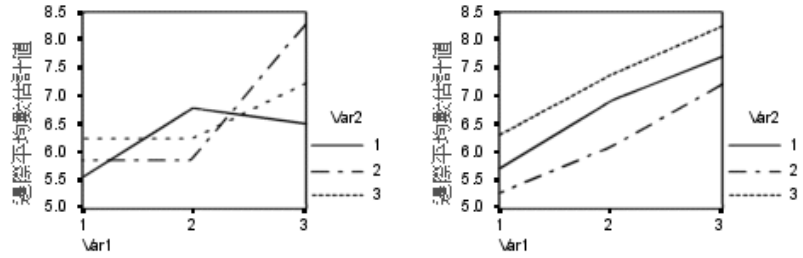
GLM 剖面圖

圖表 11-4
「單變量剖面圖」對話方塊

在比較模式中的邊際平均數時，剖面圖（交互作用圖）非常有用。剖面圖是線形圖，其上的每個點表示某個因子水準上依變數的估計邊際平均數。第二個因子的水準可用於產生個別線條。第三個因子中的每個水準可用於建立個別圖形。所有固定因子和隨機因子（若有的話）可供繪圖使用。對於共變量分析，可針對每個依變數建立剖面圖。在重複測度分析中，受訪者間因子和受訪者內因子均可用於剖面圖中。您必須先安裝「進階統計量」選項，然後才能使用「GLM 多變量」和「GLM 重複量數」。

一個因子的剖面圖會顯示是否要跨水準增加或減少估計邊際平均數。對於兩個以上的因子，平行線表示因子間沒有任何交互作用，這表示您可以僅調查某個因子的水準。非平行線表示交互作用。

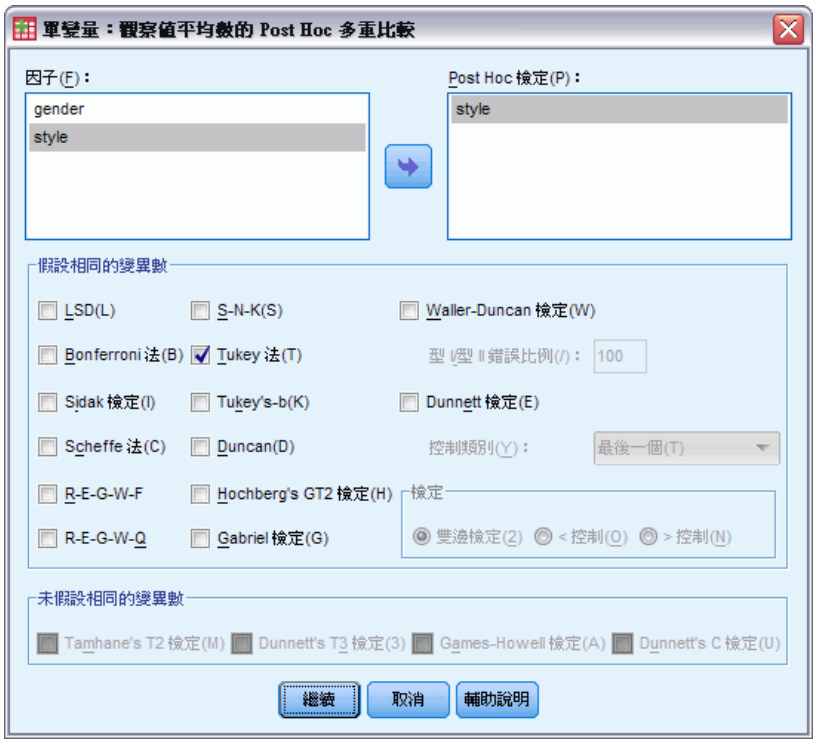
圖表 11-5
非平行圖（左側）與平行圖（右側）



在藉由選取水平軸的因子及（選擇性）個別線與個別圖來指定圖形之後，必須將該圖形新增至「圖形」清單中。

GLM Post Hoc 比較

圖表 11-6
Post Hoc 對話方塊



Post Hoc 多重比較檢定。一旦您判斷平均數之間的確存有差異之後，post hoc 全距檢定和成對多重比較便可以決定到底是哪些平均數不一樣。比較會根據未調整的值來進行。這些檢定僅供固定的受試者間因子使用。在「GLM 重複測度模式」中如果沒有受試者間因子，則無法使用這些檢定，而且會針對跨受試者內因子水準的平均來執行 post hoc 多重比較檢定。對於「GLM 多變量」，post hoc 檢定會分別測試每個依變數 。您必須先安裝「進階統計量」選項，然後才能使用「GLM 多變量」和「GLM 重複量數」。

Bonferroni 與 Tukey’s 最誠實顯著性差異常用於多重比較檢定。**Bonferroni 檢定**（以 Student’s t 統計量為依據）會調整產生多重比較之因子的觀察顯著水準。**Sidak’s t 檢定**也會調整顯著水準，並提供較 Bonferroni 檢定更嚴謹的界限。**Tukey’s 最誠實顯著性差異檢定**會使用 Studentized 範圍統計量，來進行群組間的所有成對比較，並將實驗誤差比設定為所有成對比較集合的誤差比。檢定大量的成對平均數時，Tukey’s 最誠實顯著性差異檢定會較 Bonferroni 檢定的功能更強大。對於少量的配對而言，Bonferroni 的功能較為強大。

Hochberg’s GT2類似於 Tukey’s 最誠實顯著性差異檢定，但是會使用 Studentized 最大模數。一般來說，Tukey’s 的功能較為強大。**Gabriel’s 成對比較檢定**也會使用 Studentized 最大模數，而且在儲存格大小不相等時，其功能通常較 Hochberg’s GT2 的強大。當儲存格大小有相當大的不同時，Gabriel’s 檢定可能會變成不拘泥形式的。

Dunnett’s 成對多重比較檢定會根據單一控制平均數來比較一組處理。最後一個類別是預設的控制類別。當然您也可以改用第一個類別。您也可以選擇雙邊或單邊檢定。若要檢定因子之任意水準（控制類別除外）上的平均數不等於控制類別上的平均數，可使用雙邊檢定。若要檢定因子之任意水準上的平均數小於控制類別上的平均數，可

選取「< **控制**」。同樣的，若要檢定因子之任意水準上的平均數是否大於控制類別的平均數，也可選取「> **控制**」。

Ryan、Einot、Gabriel 及 Welsch (R-E-G-W) 發展出兩個多重細分全距檢定。多重細分程序會先檢定所有平均數是否相等。若所有平均數並非相等，則會針對相等性來檢定平均數的子集。**R-E-G-W F** 是以 F 檢定為依據，而 **R-E-G-W Q** 是以 Studentized 全距為依據。這些檢定會較 Duncan's 多重全距檢定和 Student-Newman-Keuls (亦為多重細分程序) 的功能更為強大，但是不建議您針對儲存格大小不相等的情況使用他們。

當變異數不相等時，可使用 **Tamhane's T2** (以 t 檢定為依據的保存成對比較檢定)、**Dunnett's T3** (以 Studentized 最大模數為依據的成對比較檢定)、**Games-Howell 成對比較檢定** (有時是形式不拘的) 或 **Dunnett's C** (以 Studentized 全距為依據的成對比較檢定)。請注意，如果模式中有多個因子，則這些檢定會無效且不會產生。

Duncan's 多重全距檢定、Student-Newman-Keuls (**S-N-K**) 及 **Tukey's b** 均為全距檢定，可排列組別平均數的等級，並計算全距值。這些檢定的使用頻率不會像先前討論的檢定一樣頻繁。

Waller-Duncan t 檢定會使用 Bayesian 方法。當樣本大小不相等時，這個全距會檢定樣本大小的調和平均數。

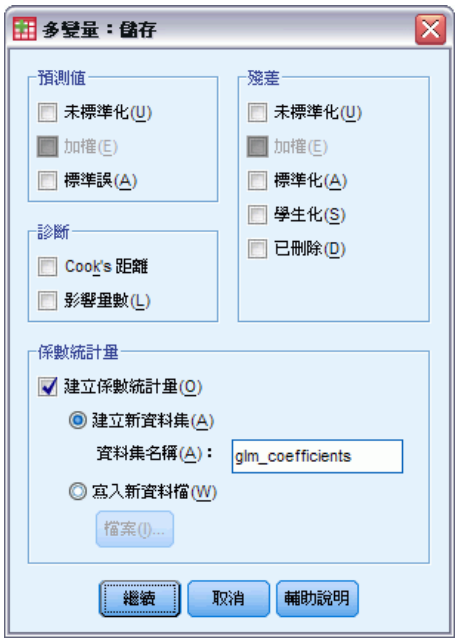
Scheffé 檢定之顯著水準的設計目的是允許所有要檢定之群組平均數的所有可能線性組合，而不只是此功能中可用的成對比較。結果是 Scheffé 檢定會較其他檢定更能保存，這表示需要顯著的平均數間較大差異。

最大的顯著差異 (**LSD**) 成對多重比較檢定相當於所有成對群組間的多重個別 t 檢定。此檢定的缺點是不會嘗試調整多重比較的觀察顯著水準。

顯示的檢定。會針對 LSD、Sidak、Bonferroni、Games-Howell、Tamhane's T2 與 T3、Dunnett's C 及 Dunnett's T3 提供成對比較。會針對 S-N-K、Tukey's b、Duncan、R-E-G-W F、R-E-G-W Q 及 Waller 提供全距檢定的同質性子集。Tukey's 誠實顯著性差異檢定、Hochberg's GT2、Gabriel's 檢定及 Scheffé's 檢定同時為多重比較檢定與全距檢定。

GLM 儲存

圖表 11-7
「儲存」對話方塊



您可以把由模式、殘差和相關量數所預測出來的值存成「資料編輯程式」中的新變數。這種變數可以用來檢驗資料的假設。若要把數值存起來以供別的 IBM® SPSS® Statistics 作業階段使用，就必須儲存目前的資料檔。

預測值。 模式為每個觀察值所預測出來的數值。

- **未標準化。** 模式所預測的依變數數值。
- **已加權。** 未標準化的加權預測值。先前選取了 WLS 變數才可使用。
- **標準誤。** 估計依變數平均值的標準差，它是為與自變數有相同數值的觀察值而進行的估計。

診斷。 此測量可以找出包含自變數異常組合值的觀察值，還有可能對模式有重大影響的觀察值。

- **Cook’s 距離。** 若自迴歸係數計算中排除特定觀察值，則其會測量所有觀察值的殘差變更程度。若 Cook’s D 較大，則表示自迴歸統計量計算中排除某觀察值已足以造成係數變更。
- **影響量數。** 未置中的影響量數。模式適合度中之每個觀察值的相對影響。

殘差。 未標準化的殘差是依變數跟模式預測值相減後，產生出來的實際值。您也可以使用標準化殘差、Studentized 殘差和已刪除殘差。如果已選擇 WLS 變數，則可使用加權的非標準化殘差。

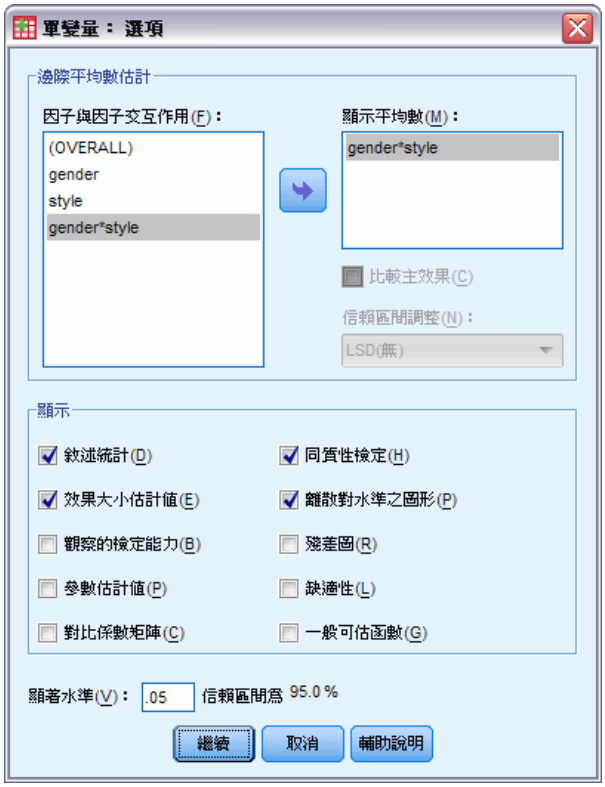
- **未標準化。** 觀察值和模式所預測的數值之間的差異。
- **已加權。** 未標準化加權殘差。先前選取了 WLS 變數才可使用。

- **標準化。** 殘差除以其標準差的估計值。標準化殘差（也稱為 Pearson 殘差）的平均數為 0，標準差為 1。
- **Studentized。** 殘差會根據自變數的平均數到自變數中每個觀察值的數值之距離除以隨其觀察值類型變化之標準差的估計值。
- **刪除。** 從迴歸係數計算中排除之觀察值的殘差。其為依變數值與已調整預測值間的差異。

係數統計量。 將模式中參數估計的變異數共變異數矩陣寫入目前階段作業中的新資料集或外部 SPSS Statistics 資料檔。對於每個依變數而言，資料檔中會有一列參數估計值、一列 t 統計量的顯著性值（跟參數估計值相對應），及一列殘差自由度。對於多變量模式而言，每個依變數都有類似的列。您可以在其他需要讀取矩陣檔的程序中使用這個矩陣檔。

GLM 選項

圖表 11-8
「選項」對話方塊



您從這個對話方塊中取得選用性的統計量。統計量是根據固定效應的模式算出來的。

邊際平均數估計。 選取因子和交互作用，以便估計儲存格中母群的邊際平均數。這些平均數會根據共變量調整（如果有的話）。

- **比較主效應。** 它提供模式中，所有主效應的邊際平均數估計之間，任何未修正的成對比較（這個功能適用於受試者間和受試者內因子）。只有在「顯示平均數」清單選取了主效應之後，才能使用這個選項。
- **信賴區間調整。** 您可以選取最小顯著差異 (LSD)、Bonferroni 法或 Sidak 調整（對信賴區間和顯著性而言）。只有在選取了「**比較主效應**」之後，才能使用這個選項。

顯示。如果選取「**敘述統計**」，就會產生觀察平均數、標準差和所有儲存格中每個依變數的個數。**效果項大小估計值**會提供所有效應項和所有參數估計值的偏 eta-平方值。Eta-平方統計量會說明可歸因於某個因子之總變異性比例。當替代假設是根據觀察值設定時，如果選取「**觀察的檢定能力**」，就可以取得檢定冪次。如果選取「**參數估計值**」，就以產生參數估計值、標準誤、 t 檢定、信賴區間和觀察的各檢定能力。選取「**對比係數矩陣**」可取得 L 矩陣。

均齊性檢定會在受試者間因子的所有水準組合之間(只針對受試者間因子)，產生每個依變數的變異數均齊性 Levene 檢定 。如需檢查關於資料的假設，則「離散對水準之圖形」和殘差圖選項會相當有用。如果沒有因子，這個選項會被停用。選取「**殘差圖**」可產生各依變數的觀察值對預測值對標準化殘差的圖形。這些圖形對於研究變異數相等之類的假設幫助很大。如果選取「**適缺性**」，就可以檢查模式是否已經適當地說明出依變數與自變數之間的關係。**一般可估函數**讓您可以根據一般可估計的函數，建立出自訂的假設檢定。對比係數矩陣中的每一列，都是一般可估計函數的線性組合。

顯著水準。 您也許想調整 post hoc 檢定中所使用的顯著水準，或者用以建立信賴區間的信賴水準。所指定的值也會用來計算觀察的檢定能力。當您指定信賴水準時，信賴區間的相關水準就會顯示在對話方塊中。

UNIANOVA 指令的其他功能

指令語法語言也可以讓您：

- 在設計中指定巢狀效應（使用 DESIGN 次指令）。
- 指定效應檢定，或是效應（或值）的線性組合（使用 TEST 次指令）。
- 指定多重對比（使用 CONTRAST 次指令）。
- 包含使用者自訂的遺漏值（使用 MISSING 次指令）。
- 指定 EPS 條件（使用 CRITERIA 次指令）。
- 建構自訂的 L 矩陣、 M 矩陣或 K 矩陣（使用 LMATRIX、MMATRIX 和 KMATRIX 次指令）。
- 為離差或簡單對比指定其中間參考類別（使用 CONTRAST 次指令）。
- 指定多項式對比的矩陣（使用 CONTRAST 次指令）。
- 指定 post hoc 比較的誤差項（使用 POSTHOC 次指令）。
- 為因子清單中的任何因子或因子間的交互作用計算邊際平均數估計（使用 EMMEANS 次指令）。
- 指定暫存變數的名稱（使用 SAVE 次指令）。
- 建立相關矩陣資料檔（使用 OUTFILE 次指令）。
- 建立矩陣資料檔，其內含有來自受試者間 ANOVA 摘要表之統計量（使用 OUTFILE 次指令）。
- 將設計矩陣存入新的資料檔（使用 OUTFILE 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
12

雙變數相關分析

您可以利用「雙變數相關分析」程序，算出 Pearson 的相關係數，以及 Spearman’s rho 與 Kendall’s tau-b 及其顯著水準。「相關」程序可用來測量變數或等級順序之間的關聯。在計算相關係數之前，請先篩選您要用於偏離值（會造成誤導結論）及線性關係證據的資料。Pearson 的相關係數是一種線性關聯的量數。在 Pearson 的相關係數下，兩個變數可以精確的關聯，但如果關係不是線性的話，那麼就不能用它來測量關聯。

範例。 以籃球得分為例。一個籃球隊獲勝場次與每場的平均得分有關連嗎？從散佈圖中可看出，它們具有線性關聯。我們再從 1994 - 1995 NBA 球季分析資料得知，Pearson 的相關係數 (0.581) 在 0.01 水準時是顯著的。於是您可能猜想，每季所贏得的場次愈多，則對手的得分愈少。這些變數是負相關 (－0.401)，而且此相關在 0.05 水準時是顯著的。

統計量。 對於每個變數來說，統計量包括：觀察值的數量與非遺漏值、平均數、標準差。對於平均數間的差異而言，統計量包括：Pearson 的相關係數、Spearman’s rho、Kendall’s tau-b、叉積離差與共變異數。

資料。 對於 Pearson 的相關係數，請使用對稱性數值變數；而對於 Spearman’s rho 及 Kendall’s tau-b，請使用數值變數或已依類別排序的變數。

假設。 Pearson 的相關係數會假設每對變數都是常態雙變數。

若要取得雙變數相關分析

從功能表選擇：

分析(A) > 相關 > 雙變數...

圖表 12-1
「雙變數相關分析」對話方塊

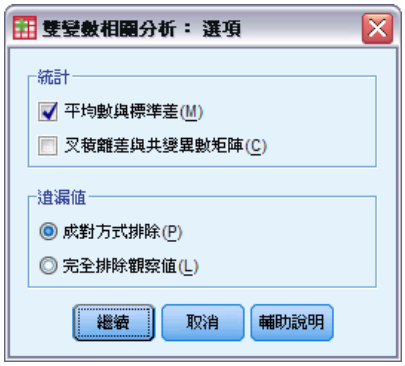
▶ 選取兩個或兩個以上的數值變數。

您也可以使用下列的選項：

- **相關係數。**對於常態分配的數值變數，請選擇 Pearson 相關係數。如果您的資料不是常態分配，或已依類別排序，請選擇 Kendall’s tau-b 或 Spearman，以便測量等級排列之間的關聯。相關係數的值域為 −1（完全負相關）到 +1（完全正相關）。其中，數值 0 表示沒有任何線性關係。當您在解析結果時，請不要因為顯著的相關，而逕下任何跟因果相關的結論。
- **顯著性檢定。**您可以選擇雙尾或單尾機率。如果事先已知道關聯的方向，請選取**單尾**。否則，請選取**雙尾**。
- **相關顯著性訊號。** 當相關係數的顯著性水準為 0.05 時，會以一個星號做為識別，而顯著性水準為 0.01 時，則以兩個星號做為識別。

雙變數相關分析選項

圖表 12-2
「雙變數相關分析選項」對話方塊



統計量。對於 Pearson 相關，您可以選擇下列其中一個選項（或兩者都選）：

- **平均數與標準差。** 每個變數都會顯示這兩個數值。同時也會顯示不具遺漏值的觀察值個數。無論您的遺漏值設定設定為何，系統都會逐一處理變數的遺漏值。
- **叉積離差與共變異數。** 每對變數都會顯示這兩個值。交乘積離差的值，等於正確平均數變數的乘積總和。這是 Pearson 相關係數的分子。而共變量則是兩變數間關係的非標準化量數，其值等於交乘積離差除以 N－1。

遺漏值。 您可以任選一個選項：

- **成對方式排除。** 若觀察值具有相關係數之成對變數中，其中一個或兩個變數的遺漏值，則分析時會排除此觀察值。既然每個係數都是根據在特定成對變數上、具有有效代碼的所有觀察值而來，那麼您就可以在每個計算式中使用最大值資訊。如果觀察值個數不同，就會產生不同的係數集合。
- **完全排除遺漏值。** 如果任何變數的觀察值中，含有遺漏值，它們就會從所有相關係數中排除。

CORRELATIONS 和 NONPAR CORR 指令的其他功能

指令語法語言也可以讓您：

- 編寫可用來代替原始資料的 Pearson 相關矩陣，以取得因子分析等其他分析（使用 MATRIX 次指令）。
- 取得清單上的每個變數，與第二個清單上的每個變數間的關聯（在 VARIABLES 次指令上使用關鍵字 WITH）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
13

偏相關

您可以利用「偏相關」程序，在兩個變數控制一個（或更多個）其他變數的效果時，計算這兩個變數之間，描述其線性關係的偏相關係數。此處相關指的是線性關係的測量。這兩個變數可以完全相關，但只要它們之間的關係不是線性，那麼相關係數就不適合用來測量它們之間的關係。

範例。 健保基金和疾病率之間有無關係？雖然您可能預期此關係會是負相關，不過研究報告卻發現兩者之間有顯著的正相關：隨著健保基金的增加，疾病率也明顯的增加。控制造訪醫療保健機構的比例，但是，實際上是降低觀察的正相關。健保基金與疾病率只會顯示為正相關，因為當基金增加時，就會有更多人使用醫療保健服務，導致醫生與醫院提出更多的疾病醫療報告。

統計量。 對於每個變數來說，統計量包括：觀察值的數量與非遺漏值、平均數、標準差。具自由度和顯著水準的偏相關和零階相關矩陣。

資料。 使用對稱的數值變數。

假設。 「偏相關」程序假設：各對變數都是常態雙變數。

若要取得偏相關

▶ 從功能表選擇：

分析(A) > 相關 > 偏相關...

圖表 13-1
「偏相關」對話方塊

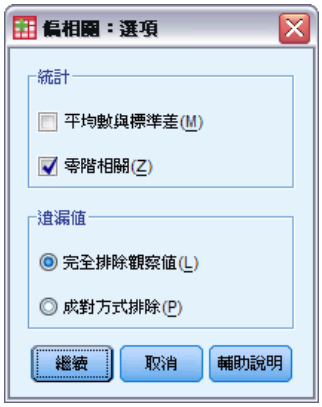
- ▶ 選取兩個（或以上）偏相關所要計算的數字變數。
- ▶ 選取一個（或以上）數值控制變數。

您也可以使用下列的選項：

- **顯著性檢定。**您可以選擇雙尾或單尾機率。如果事先已知道關聯的方向，請選取**單尾**。否則，請選取**雙尾**。
- **顯示實際的顯著水準。** 在預設情況下，各相關係數都會顯示機率和自由度。如果您取消選擇此項目，而且係數顯著性的水準又是 0.05 的話，則以一個星號以便識別，如果係數顯著性的水準為 0.01，則以雙星號識別，此時自由度則不予列出。此設定會同時影響偏相關和零階相關矩陣。

偏相關的選項

圖表 13-2
「偏相關選項」對話方塊



統計量。您可以選擇下列其中一個選項，或兩者都選：

- **平均數與標準差。** 每個變數都會顯示這兩個數值。同時也會顯示不具遺漏值的觀察值個數。
- **零階相關。** 顯示所有變數（包括控制變數）之間的簡單相關矩陣。

遺漏值。 您可以任選下列其中一個選項：

- **完全排除遺漏值。** 所有的計算都會排除遺漏任何變數值（包括控制變數）的觀察值。
- **成對方式排除。** 在計算以偏相關為基礎的零階相關時，如果觀察值的成對變數中，其中有一個或兩個都具有遺漏值的話，就不使用這個觀察值。如果選用以成對方式刪除觀察值，則盡其可能地使用資料。但是觀察值的個數，會隨著係數而改變。如果以成對方式刪除觀察值，則個別偏相關係數的自由度，會以最小觀察值個數為基礎（該值用來計算任何零階相關）。

PARTIAL CORR 指令的其他功能

指令語法語言也可以讓您：

- 讀取零階相關矩陣、或寫入偏相關矩陣（使用 MATRIX 次指令）。
- 取得兩變數清單中的偏相關（使用 VARIABLES 次指令上的關鍵字 WITH）。

- 取得多重分析（使用多個 VARIABLES 次指令）。
- 當您有兩個控制變數（使用 VARIABLES 次指令）時，指定要求的階數（例如，第一、第二階偏相關都包括在內）。
- 隱藏多餘的係數（使用 FORMAT 次指令）。
- 當某些係數無法計算時（使用 STATISTICS 次指令），顯示簡單相關的矩陣。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
14

距離

此程序可用來計算成對變數或成對觀察值之間的相似性或相異性（距離）。然後，我們再將這些相似性或距離測量結果，用於像是因子分析、集群分析、或多元尺度方法之類的其他程序，以便進一步協助我們分析更複雜的資料集。

範例。在本車輛範例中，研究人員想要根據某些特性（如引擎大小、MPG 和馬力），來測量成對車輛之間的相似性。並藉由計算車輛之間的相似性來進一步得知，哪些車輛相類似，而哪些車輛又彼此不同。如果想要進行比較正式的分析，您可能需要考慮是否要使用階層集群分析、或多元尺度方法，以探測其基本結構。

統計量。 在區間資料的相異性（距離）測量方面，其統計量包括：歐基里得直線距離、歐基里得直線距離平方、Chebychev、區塊、Minkowski 或自訂式；如果是個數資料，統計量則有：卡方、或 Phi 的平方距離測量；如果是二項資料，統計量則有：歐基里得直線距離、歐基里得直線距離平方、大小差異、形式差異、變異數、形狀、或 Lance 與 Williams。在區間資料的相似性測量方面，其統計量包括：Pearson 相關或餘弦；如果是二項資料，統計量則有：Russel 與 Rao、簡單配對、Jaccard、骰子、Rogers 與 Tanimoto、Sokal 與 Sneath 1、Sokal 與 Sneath 2、Sokal 與 Sneath 3、Kulczynski 1、Kulczynski 2、Sokal 與 Sneath 4、Hamann、Lambda 值、Anderberg D 值、Yule Y 係數、Yule Q 值、Ochiai、Sokal 與 Sneath 5、Phi 四點相關、分散情形。

若要取得距離矩陣

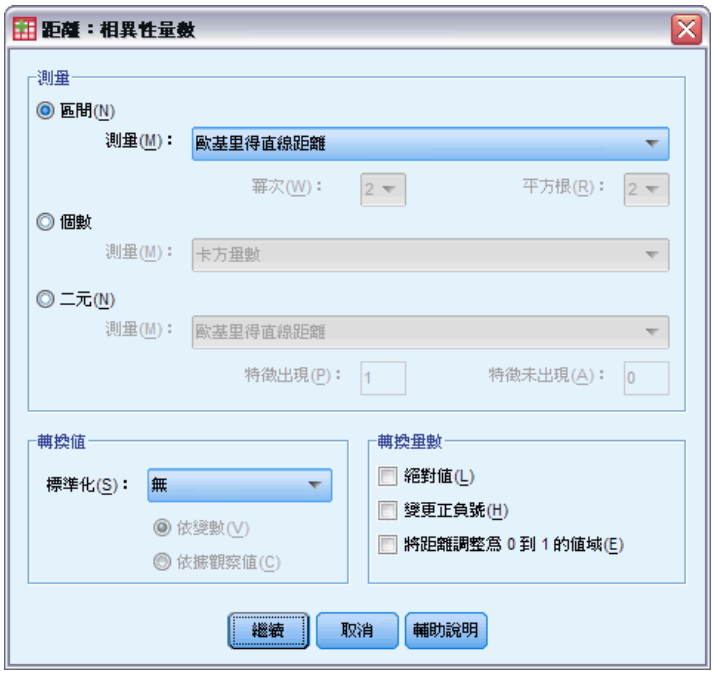
▶ 從功能表選擇：
分析(A) > 相關 > 距離...

圖表 14-1
距離對話方塊

- 若要計算觀察值之間的距離，請至少選取一個數字變數；或者，若要計算變數之間的距離，請至少選取兩個數字變數，。
- 選取「計算距離」組別中的其中一個項目，以計算觀察值或變數之間的近似性。

距離相異性測量

圖表 14-2
距離相異性測量對話方塊



從「測量」組別中，選取對應到您資料類型（間隔、個數、或二元）的項目；然後，從下拉式清單中，選取對應到此資料類型的其中一個量數。如果根據資料類型來區分的話，您可以使用的測量有：

- **間隔資料。**歐基里得直線距離、歐基里得直線距離平方、Chebychev、區塊、Minkowski、或自訂式。
- **個數資料。**卡方量數、Phi 平方測量。
- **二元資料。** 歐基里得直線距離、歐基里得直線距離平方、大小差異、形式差異、變異數、形狀、或 Lance 與 Williams。(請輸入「特徵出現」或「特徵未出現」的值，以指定哪兩個值是有意義的；「距離」將會忽略其他的數值)。

「轉換值」組別，能讓您在計算近似性之前，先將觀察值或變數的資料值標準化。不過，請注意，這些轉換並不適用於二元資料。在這裡，您可以使用的標準化方法，包括：z 分數、範圍 - -1 到 1、範圍 0 到 1、1 的最大級數、1 的平均數、1 的標準差。

「轉換測量」組別能讓您轉換距離測量所產生的數值。但是，請注意，您必須在完成距離測量計算之後，才能套用這些值。在這裡，您可以使用的選項，包括：絕對值、變更正負號、以及將距離重新尺度化為 0 - 1 的範圍。

距離相似性測量

圖表 14-3
距離相似性測量對話方塊

從「測量」組別中，選取對應到您資料類型（間隔或二元）的項目；然後，從下拉式清單中，選取對應到此資料類型的測量。如果根據資料類型來區分的話，您可以使用的測量有：

- **間隔資料。**Pearson 相關或餘弦。
- **二元資料。**Russel 與 Rao、簡單配對、Jaccard、骰子、Rogers 與 Tanimoto、Sokal 與 Sneath 1、Sokal 與 Sneath 2、Sokal 與 Sneath 3、Kulczynski 1、Kulczynski 2、Sokal 與 Sneath 4、Hamann、Lambda 值、Anderberg D 值、Yule Y 係數、Yule Q 值、Ochiai、Sokal 與 Sneath 5、Phi 四點相關、或分散情形。（請輸入「特徵出現」或「特徵未出現」的值，以指定哪兩個值是有意義的；「距離」將會忽略其他的數值）。

「轉換值」組別，能讓您在計算近似性之前，先將觀察值或變數的資料值標準化。不過，請注意，這些轉換並不適用於二元資料。在這裡，您可以使用的標準化方法，包括：z 分數、範圍 - -1 到 1、範圍 0 到 1、1 的最大級數、1 的平均數、1 的標準差。

「轉換測量」組別能讓您轉換距離測量所產生的數值。但是，請注意，您必須在完成距離測量計算之後，才能套用這些值。在這裡，您可以使用的選項，包括：絕對值、變更正負號、以及將距離重新尺度化為 0 - 1 的範圍。

PROXIMITIES 指令的其他功能

「距離」程序使用 PROXIMITIES 指令語法。指令語法語言也可以讓您：

- 指定任何整數作為 Minkowski 距離量數的冪次。
- 指定任何整數作為自訂距離量數的冪次和根。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
15

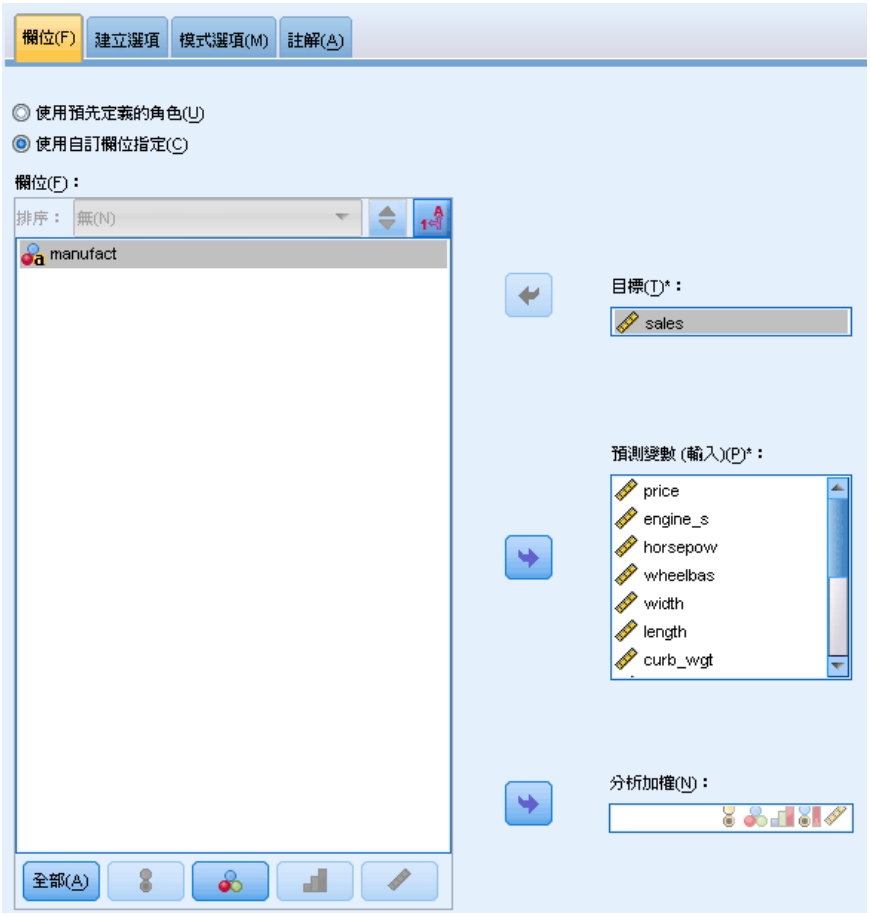
線性模式

線性模式會根據目標與一或多個預測值之間的線性關係預測連續目標。

線性模式相當簡便，且提供了簡易的評分數學公式。與同一個資料集上的其他模式類型（如神經網路或決策樹狀結構）相較，這些模式的內容比較容易瞭解，通常可以快速建立。

範例。 某資源有限的保險公司，打算調查屋主的保險理賠，希望建立估計理賠成本的模式。透過在服務中心部署這個模式，代表就能在和客戶通電話的同時輸入資訊，並根據過去的資料立即取得「預期的」理賠成本。

圖表 15-1
欄位索引標籤



欄位需求。必須具有「目標」以及至少一個的「輸入」。依照預設值，不會使用預先定義角色為「兩者」或「無」的欄位。目標必須是連續的（尺度）。預測值（輸入）無測量水準限制。

若要取得線性模式

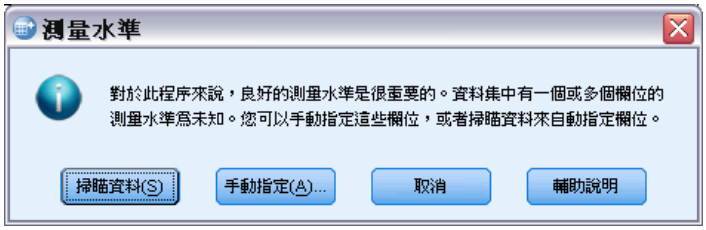
本功能需要 Statistics Base 選項。

從功能表選擇：
分析(A) > 迴歸 > 自動線性模式...

- ▶ 確認至少具有一個目標和輸入。
- ▶ 按一下「**建立選項**」，以指定選用的建立與模式設定。
- ▶ 按一下「**模式選項**」以將分數儲存至作用中資料集，並將模式匯出至外部檔案。
- ▶ 按一下「**執行**」以執行程序，並且建立「模式」物件。

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 15-2
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「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。

目標

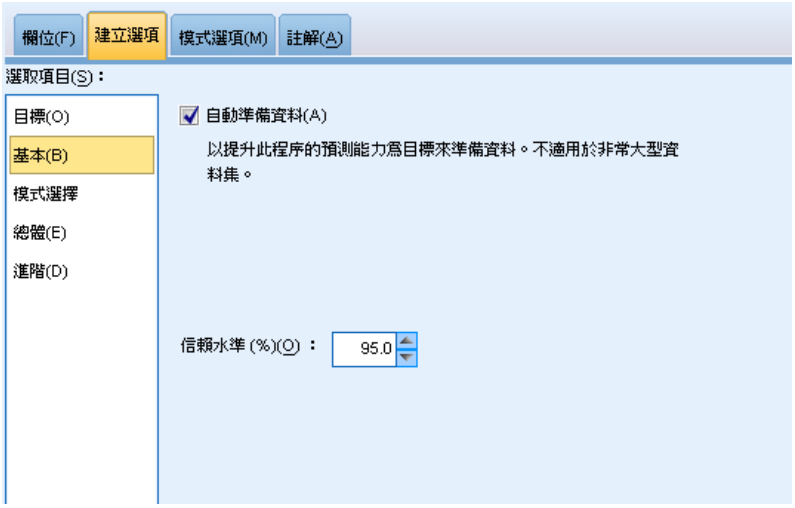
您的主要目標是什麼？

- **建立標準模式。** 此方法會建立單一模式，以預測使用預測值的目標。一般而言，標準模式在解讀上較為容易，且評分速度比 Boosted、Bagged 或大型資料集集合更快。
- **強化模式準確性 (Boosting)。** 此方法會使用 Boosting 來建立集合模式，其會產生一系列的模式，以取得更為準確的預測。集合的建立和評分時間均長於標準模式。

- **強化模式穩定性 (Bagging)。** 此方法會使用 Bagging（自助法整合）來建立集合模式，其會產生多個模式，以取得更為可靠的預測結果。集合的建立和評分時間均長於標準模式。
- **為超大型資料集建立模式（需要 IBM SPSS Statistics Server）。**此方法會將資料集分割為個別的資料區塊，以建立集合模式。若資料集過大而無法建立上述任何模式，或是想要建立增量模式，請選擇此選項。此選項的建立時間較短，但評分時間會長於標準模式。此選項需要 「 SPSS Statistics 伺服器」連線。

基本

圖表 15-3
基本設定



自動準備資料。 此選項可讓程序在內部轉換目標和預測值，以發揮最高的模式預測能力；所有轉換都會和模式一起儲存，並套用至新資料以進行評分。系統會將已轉換欄位的原始版本自模式中排除。依預設，系統會執行下列的自動資料準備作業。

- **日期與時間處理。**每個日期預測值皆會轉換至新的連續預測值，其中內含參考日期 (1970-01-01) 之後的經過時間。每個時間預測值皆會轉換為新的連續預測值，其中包含參考時間 (00:00:00) 之後的經過時間。
- **調整測量水準。**系統會將不同值數目小於 5 的連續預測值重新分配為次序預測值。系統會將不同值數目大於 10 的次序預測值重新分配為連續預測值。
- **偏離值處理。**系統會將超出分割值（與平均值相距 3 個標準差）的連續預測值設為分割值。
- **遺漏值處理。** 系統會以訓練區隔的眾數來取代名義預測值的遺漏值。系統會以訓練區隔的中位數來取代次序預測值的遺漏值。系統會以訓練區隔的平均數來取代連續預測值的遺漏值。
- **受監督合併。**此項目會透過減少要處理的目標相關欄位數目，建立較為精簡的模式。相同的類別是根據輸入和目標之間的關係來識別。系統會合併沒有顯著差異（即 p 值大於 0.1）的類別。若將所有類別合而為一，則會從模式中排除欄位的原始和衍生版本，這是因為它們沒有可作為預測值的數值。

信賴水準。 此信賴水準用於在係數檢視中計算模式係數的區間估計值。指定大於 0 且小於 100 的一個數值。預設值為 95。

模式選擇

圖表 15-4
模式選擇設定

模式選擇方法。 選擇其中一種模式選擇方法（詳細資訊如下）或「無」，以僅將所有可用的預測值輸入為主效果模式項目。依預設，系統會使用「向前逐步」。

向前逐步選擇。 此方法一開始並不會對模式產生任何效果，但會根據逐步條件一步一步新增或移除效果，直至無法再新增或移除任何效果為止。

- **選入/移除的條件。** 此統計量決定是否應在模式中新增或移除效果。資訊準則 (AICC) 是以指定模式訓練集的概似為基礎，且會經過調整以懲罰過於複雜的模式。F 統計量是以模式錯誤改善的統計檢定為基礎。已調整 R 平方是以訓練集的配適度為基礎，且會經過調整以懲罰過於複雜的模式。過適預防準則 (ASE) 是以過適預防集的配適度（平均平方誤，或 ASE）為基礎。過適預防集是原始資料集 30% 左右的隨機次樣本，不用來訓練模式。

  若選擇任何非「F 統計量」的條件，則系統會將每個步驟中對應至最大正向增加條件的效果新增至模式。系統會移除模式中任何對應至減少條件的效果。

若選擇「F 統計量」作為條件，則系統會在每個步驟中，將最小 p 值小於指定門檻（「包含 p 值小於此值的效果」）的效果新增至模式。預設值是 0.05。系統會移除任何 p 值大於指定門檻（「移除 p 值大於此值的效果」）之模式中的效果。預設值是 0.10。

- **自訂最終模式的最大效果數目。** 依預設，系統會將所有可用的效果輸入至模式。或者，若逐步演算法以指定的最大效果數目來結束步驟，則演算法在停止時會保有目前的效果集。
- **自訂步驟的最大數目。** 逐步演算法在經過特定步驟數目後即會停止。依預設，此為可用效果數目的 3 倍。或者，請指定正整數的步驟最大數目。

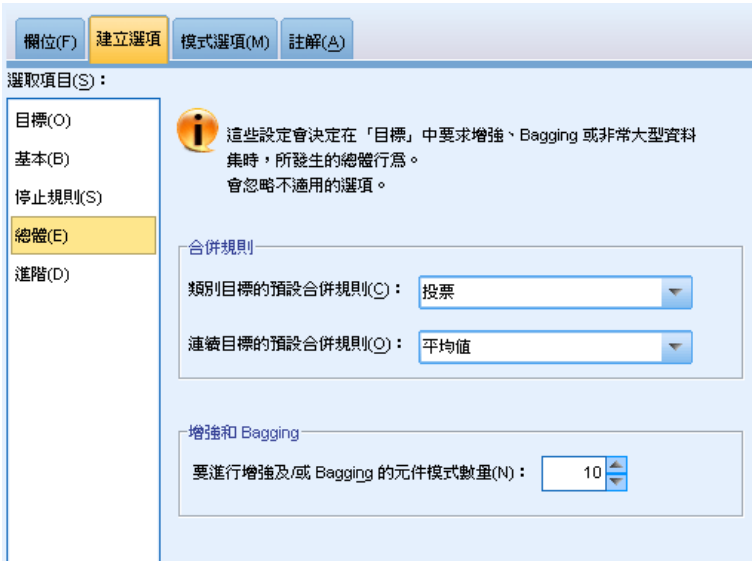
最佳子集選擇。 這會檢查「所有可能」模式或是大於向前逐步的可能模式子集，以根據最佳子集條件來選擇最佳子集。資訊準則 (AICC) 是以指定模式訓練集的概似為基礎，且會經過調整以懲罰過於複雜的模式。已調整 R 平方是以訓練集的配適度為基礎，且會經過調整以懲罰過於複雜的模式。過適預防準則 (ASE) 過適預防準則 (平均平方誤，或 ASE) 為基礎。過適預防集是原始資料集 30% 左右的隨機次樣本，不用來訓練模式。

系統會將具有最大條件值的模式選作最佳模式。

注意：最佳子集選擇較向前逐步選擇更需要大量計算。搭配 boosting、bagging 或極大資料集執行最佳子集時，其建立時間會長於使用向前逐步選擇建立的標準模式。

集合

圖表 15-5
集合設定



系統在「目標」中要求 boosting、bagging 或極大資料集時，這些設定會決定所發生的集合行為。系統會忽略無法套用至所選目標的選項。

Bagging 與極大資料集。 系統執行集合評分時，可使用此規則來合併基底模式的預測值，以運算集合分數值。

- **連續目標的預設合併規則。** 系統會使用基底模式預測值的平均數或中位數，來合併連續目標的集合預測值。

請注意，若目標是用於強化模式準確性，則系統會忽略合併規則選擇。Boosting 會一律使用大部分的加權投票來為類別目標評分，並且使用加權中位數來為連續目標評分。

Boosting 與 Bagging。 當目標用於強化模式準確性或穩定性時，可指定欲建立的基底模式數目；若為 bagging，則此為自助法範例的數目。其應為正整數。

進階

圖表 15-6
進階設定

複製結果。 設定亂數種子以供您複製分析。系統會使用亂數產生器來選擇過適預防集中的記錄。請指定一個整數，或是按一下「**產生**」以建立介於 1 和 2147483647 之間（含）的虛擬亂數整數。預設值是 54752075。

模式選項

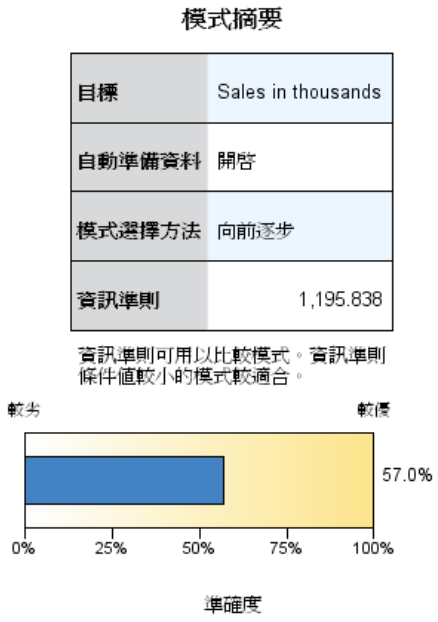
圖表 15-7
模式選項索引標籤

儲存預測值至資料集。預設變數名稱為 PredictedValue。

匯出模式。這會將模式寫入外部的 .zip 檔案。您可以使用這個模式檔案，將模式資訊套用到其他資料檔案中以進行評分工作。 指定唯一且有效的檔案名稱。若檔案規格參照至現有檔案，則系統會覆寫檔案。

模式摘要

圖表 15-8
模式摘要檢視



「模式摘要」檢視是一種快照、模式及其配適的一覽摘要。

表格。此表格指出幾個高階模式設定，包括：

- 的目標名稱，
- 自動資料準備是否如「基本」設定之指定執行，
- 「模式選擇」設定中指定的模式選擇方法和選擇準則。同時也會顯示最終模式的選擇準則值，以越小越佳的格式表示。

圖表。圖表會顯示最終模式的準確性，以越大為越佳的格式表示。此值為 100 ×(最終模式的已調整 R^2)。

自動式資料準備

圖表 15-9
自動式資料準備檢視

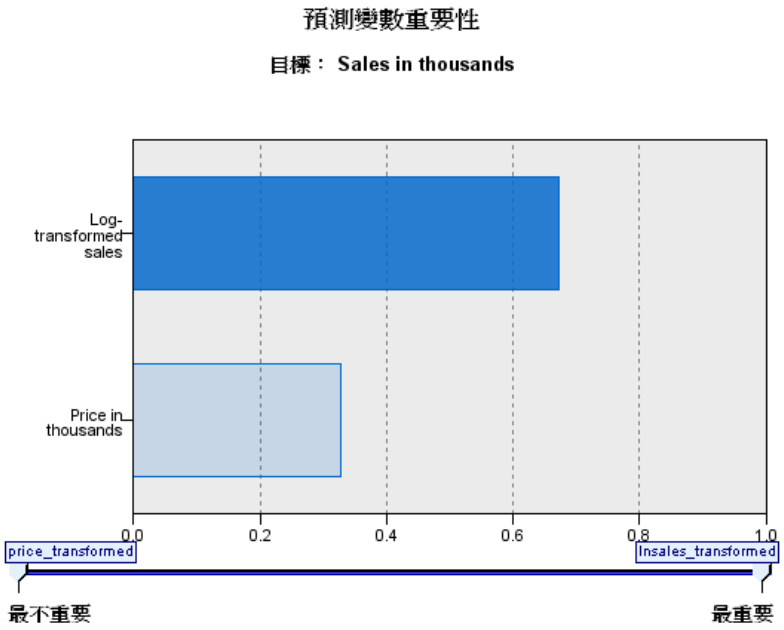
此檢視會顯示已排除欄位，以及在自動式資料準備 (ADP) 步驟中衍生轉換欄位方式的相關資訊。針對每個已轉換或排除的欄位，此表格會列出欄位名稱、分析當中的欄位角色，以及 ADP 步驟所採取的動作。系統會依照欄位名稱的字母順序以遞增方式來排序欄位。可能為每個欄位採取的行動包括：

- **取得期間：月份**計算從包含日期的欄位值到目前系統日期的經過時間 (以月為單位)。
- **取得期間：小時**計算從包含時間的欄位值到目前系統時間的經過時間 (以小時為單位)。
- **將測量水準從連續變更為次序**將唯一值少於 5 個的連續欄位重新分配為次序欄位。
- **將測量水準從次序變更為連續**將唯一值多於 10 個的次序欄位重新分配為連續欄位。
- **修整偏離值**會將超出分割值 (與平均值相距 3 個標準差) 的連續預測值設為分割值。
- **置換遺漏值**以模式置換名義欄位的遺漏值，以中位數置換次序欄位，以平均數置換連續欄位。

- **合併類別以最大化和目標的關聯**根據輸入和目標之間的關係來識別「類似的」預測值類別。系統會合併沒有顯著差異（即 p 值大於 0.05）的類別。
- **排除常數預測值 / 偏離值處理之後 / 合併類別之後**移除具有單一值的預測值，可能在採取其他 ADP 行動之後。

預測值重要性

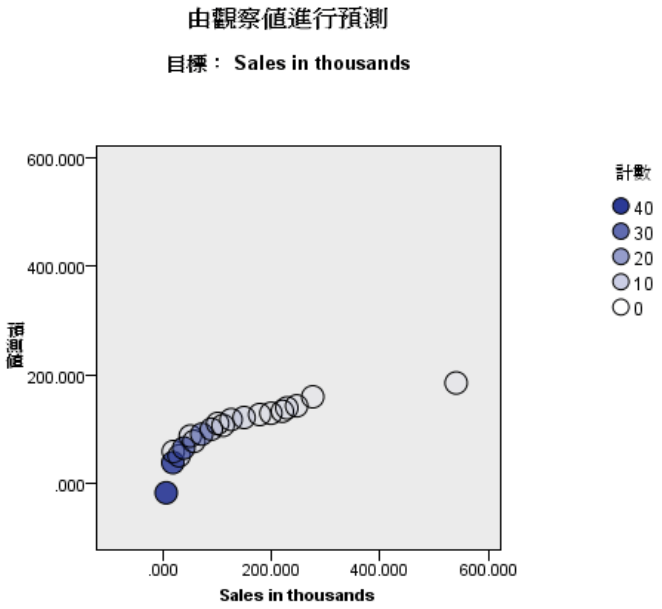
圖表 15-10
預測值重要性檢視



一般而言，您會想要將焦點著重在建模過程中最重要的預測值欄位，並考慮捨棄或忽略最不重要的預測值欄位。預測值重要性圖可協助您指出評估模式時各預測值的相對重要性，以達成此目標。由於其中的值都是相對值，因此顯示中所有預測值的值總和為 1.0。預測值重要性與模式準確性無關。這只涉及進行預測時各預測值的重要性，而不涉及預測是否正確。

依觀察預測

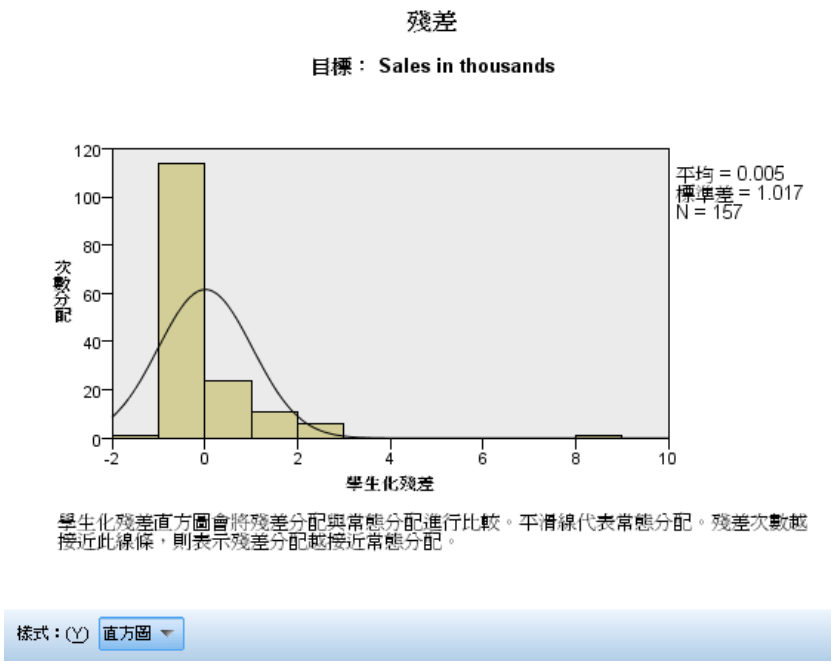
圖表 15-11
依觀察檢視預測



這會根據水平軸上的觀察值，來顯示垂直軸上經過 bin 處理的預測值散佈圖。理想的狀況下，點應排列在 45 度的線上；此檢視可以告訴您模式是否有預測結果特別差的記錄。

殘差

圖表 15-12
殘差檢視、直方圖樣式



這會顯示模式殘差的診斷圖表。

圖表樣式。提供各種不同的顯示樣式，您可以從「樣式」下拉式清單中存取這些樣式。

- **直方圖。**此為經過 bin 處理的 studentized 殘差直方圖，其中含有常態分配的重疊圖。線性模式會假定殘差具有常態分配，因此直方圖應會十分貼近平滑線條。
- **P-P 圖。**此為經過 bin 處理的機率-機率圖，其會將 studentized 殘差與常態分配加以比較。若圖中各點所構成的線條斜率小於一般線條，則殘差所顯示的變異性會高於常態分配；若斜率越大，則殘差所顯示的變異性就會越低於常態分配。若繪製的點構成 S 形曲線，則殘差分配會呈現偏斜。

偏離值

圖表 15-13
偏離值檢視

離群值

目標： Sales in thousands

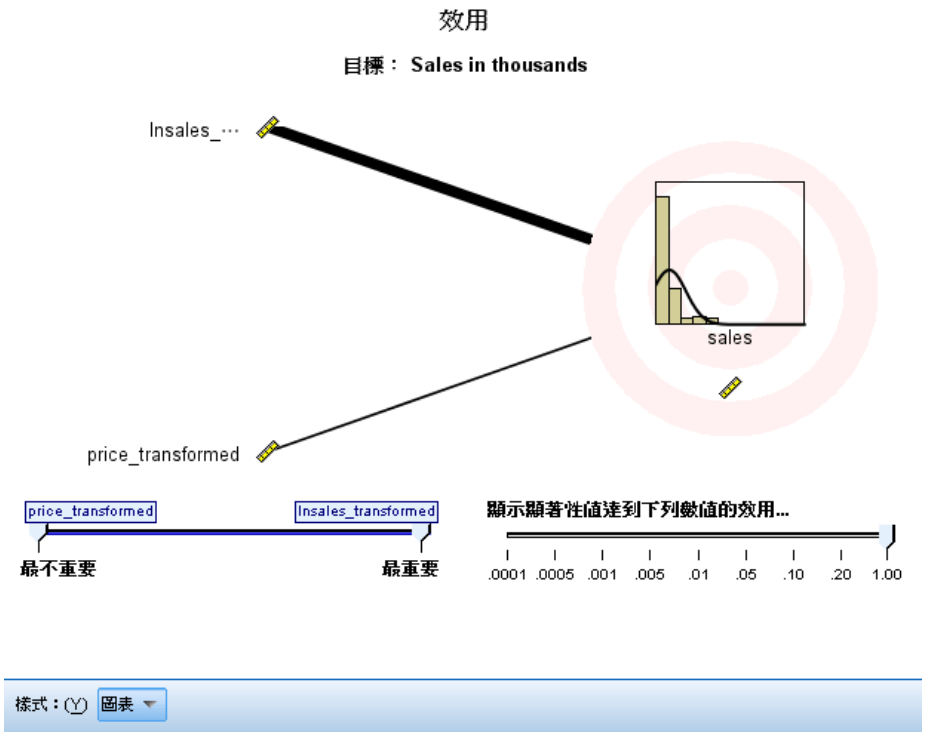
Sales in thousands	庫克距離
540.561	1.369
0.110	0.218
1.112	0.168
276.747	0.116
0.916	0.063
0.954	0.059

此表格會列出對模式產生不當影響的記錄，並會顯示記錄 ID（若已於「欄位」索引標籤上指定）、目標值和 Cook’s 距離。Cook’s 距離是一種測量方式，若自模式係數計算中排除特定記錄，則其會測量所有記錄的殘差變更程度。若 Cook’s 距離較大，則代表排除記錄已足以造成係數變更，因此應將其視為具有影響力。

您應該仔細檢查具有影響力的記錄，以決定是否要在估計模式時給予較少的加權，或是要將偏離值截斷至某些可接受的門檻，或是完整移除具有影響力的記錄。

效果

圖表 15-14
效果檢視、圖樣式



此檢視會顯示模式中每個效果的大小。

樣式。提供各種不同的顯示樣式，您可以從「樣式」下拉式清單中存取這些樣式。

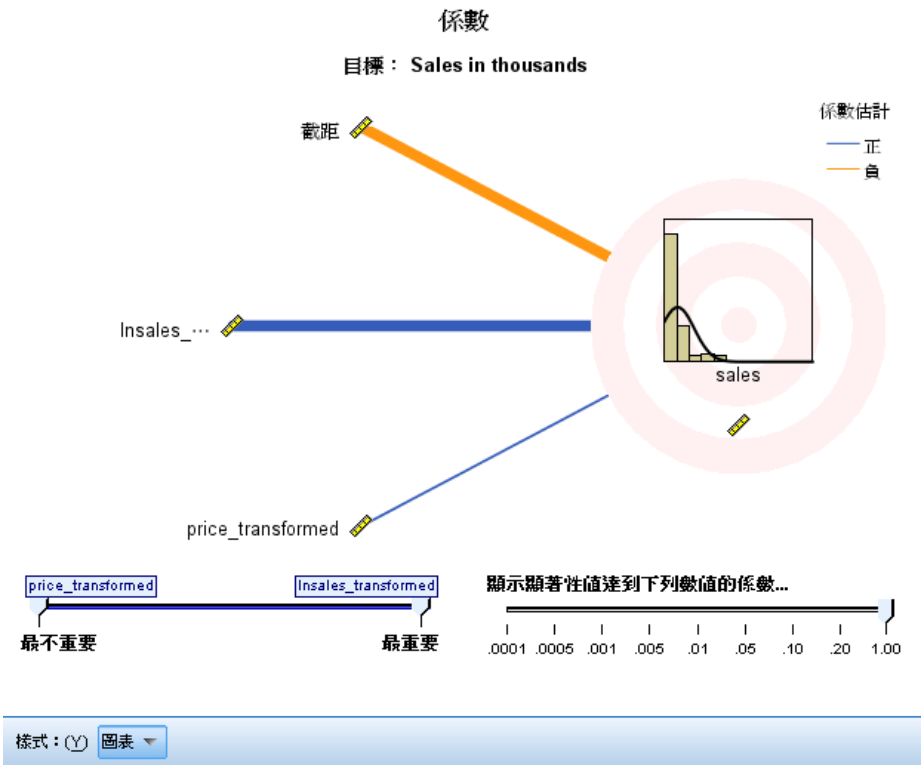
- **圖。**此圖表會依降冪的預測值重要性，來從高至低進行效果排序。系統會根據效果顯著性來加權處理圖中的連接線條，線條寬度越大代表符合越多的顯著效果（較小的 p 值）。游標停留於連接線上時，會以工具提示的方式顯示 p 值和效果的重要性。此為預設值。
- **表格。**此為整體模式與個別模式效果的 ANOVA 表格。系統會依降冪的預測值重要性，來從高至低進行個別效果排序。請注意，依預測表格會收合起來，僅顯示整體模式的結果。若要觀看個別模式效果的結果，請按一下表格中的「已修正模式」儲存格。

預測值重要性。具有「預測值重要性」滑動軸，可控制檢視中所顯示的預測值。這不會變更模式，而僅會讓您更能著重於最重要的預測值。依預設，系統會顯示前 10 個效果。

顯著。「顯著」滑動軸除了根據預測值重要性所顯示的效果之外，還可進一步控制檢視中可顯示的效果。系統會隱藏顯著值大於滑動軸值的效果。這不會變更模式，而僅會讓您更能著重於最重要的效果。預設值為 1.00，因此系統不會根據顯著性來過濾任何效果。

係數

圖表 15-15
係數檢視、圖樣式



此檢視會顯示模式中每個係數的值。請注意，模式當中的各項因子（類別預測值）皆已經過指標編碼，因此一般來說內含因子的**效果**會具有多個相關**係數**；除了對應於冗餘（參考）參數的類別外，每個類別會具有一個係數。

樣式。提供各種不同的顯示樣式，您可以從「樣式」下拉式清單中存取這些樣式。

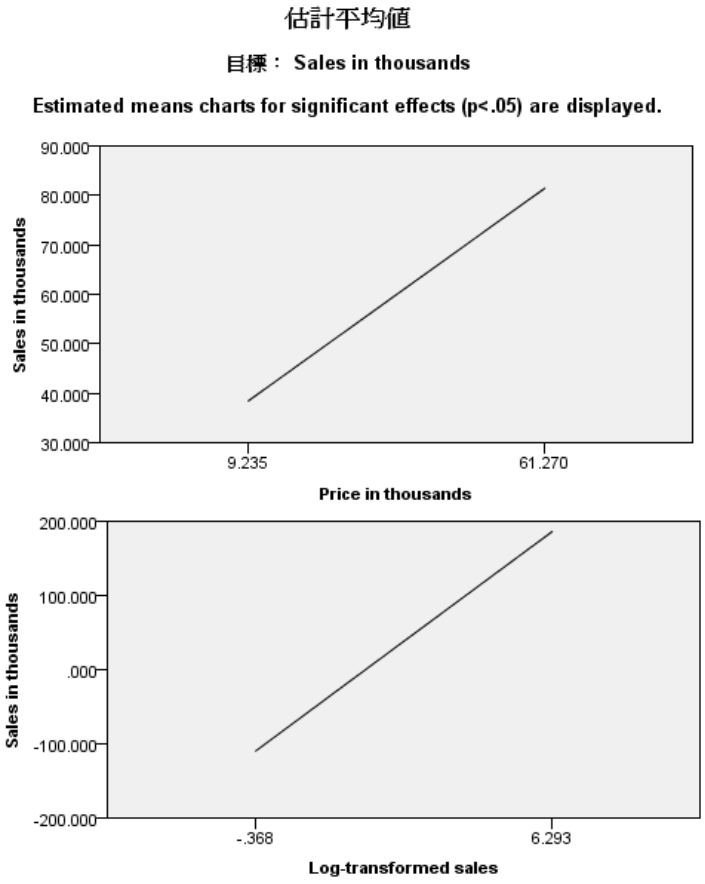
- **圖。**此圖表會先顯示截距，然後再依降冪的預測值重要性，來從高至低進行效果排序。在內含因子的效果當中，系統會依升冪的資料值順序進行係數排序。圖表中的連接線會根據係數符號（請參閱圖表鍵）以彩色顯示和根據係數顯著性加權處理，線條寬度越大代表符合越多的顯著係數（較小的 p 值）。游標停留在連接線上時，會以工具提示的方式顯示係數值、其 p 值，以及與參數相關之效果的重要性。此為預設樣式。
- **表格。** 此表格會顯示個別模式係數的值、顯著性檢定和信賴區間。在截距之後，系統會依降冪的預測值重要性來從高至低進行效果排序。在內含因子的效果當中，系統會依升冪的資料值順序進行係數排序。請注意，依預設表格會收合起來，僅顯示係數、顯著性和每個模式參數的重要性。若要觀看標準誤、t 統計量和信賴區間，請按一下表格中的「係數」儲存格。游標停留在表格中的模式參數名稱上時，會以工具提示的方式顯示參數名稱、與參數相關的效果，以及（在類別預測值方面）與模式參數相關的數值標記。在自動資料準備合併類別預測值的類似類別時，這對於觀看所建立的新類別建立特別有用。

預測值重要性。具有「預測值重要性」滑動軸，可控制檢視中所顯示的預測值。這不會變更模式，而僅會讓您更能著重於最重要的預測值。依預設，系統會顯示前 10 個效果。

顯著。具有「顯著」滑動軸，其除了根據預測值重要性所顯示的係數之外，還可進一步控制檢視中顯示的係數。系統會隱藏顯著值大於滑動軸值的係數。這不會變更模式，而僅會讓您更能著重於最重要的係數。預設值為 1.00，因此系統不會根據顯著性來過濾任何係數。

估計的平均數

圖表 15-16
估計的平均數檢視



此為顯示顯著預測值的圖表。此圖表會針對水平軸上的每個預測值顯示垂直軸上的目標模式估計值，並且保留其他所有的預測值常數。其針對目標中每個預測值係數的效果提供實用的視覺化內容。

注意：若無任何顯著預測值，則不會產生任何的估計平均數。

建立模式摘要

圖表 15-17
建立模式摘要檢視、向前逐步演算法

模式建立摘要

目標： Sales in thousands

		步驟	
		1	2
資訊準則		1,199.485	1,195.838
效用	lnsales_transformed	✓	✓
	price_transformed		✓

模式建立方法是使用「資訊準則」的「向前逐步」。
勾號標示表示效用在此步驟時出現在模型中。

若在「模式選擇」設定中選擇有別於「無」的其他模式選擇演算法，則此項目會提供有關建立模式程序的部分詳細資訊。

向前逐步。若採用向前逐步作為選擇演算法，則表格會顯示逐步演算法中的最後 10 個步驟。系統會針對每個步驟，顯示步驟中模式的選擇條件值和效果。這有助於使您瞭解每個步驟對於模式的貢獻度。您可以在每個行中排序列，以便更加輕鬆地查看指定步驟中的模式效果。

最佳子集。若採用最佳子集作為選擇演算法，則表格會顯示前 10 個模式。系統會針對每個模式，顯示模式中的選擇條件值和效果。這有助於使您瞭解頂端模式的穩定性；若這些模式當中的類似效果差異不大，則您便可安心信賴這些「頂端」模式；若這些模式當中的效果差異過大，則表示某些效果過於近似而應加以合併（或移除）。您可以在每個行中排序列，以便更加輕鬆地查看指定步驟中的模式效果。

章節
16

線性迴歸

您可以利用「線性迴歸」來估計線性方程式的係數，此線性方程式跟一個或多個自變數有關（這些自變數可以正確地預測依變數值）。例如，您可以根據年齡、教育水準、經歷之類的自變數，來預測一位推銷員的年銷售額（依變數）。

範例。 一個藍球隊在一季中贏球的次數，與球隊每場的平均得分有關嗎？從散佈圖中我們可以知道，這兩者是呈線性相關的。獲勝的場次與對手的平均得分也是呈線性相關的。但這些變數為負相關。也就是說，當贏球的次數增加時，對手的平均得分便隨之減少。利用線性迴歸，您可以從這些變數關係中，找出一種模式。一個好的模式，可以用來預測球隊的贏球次數。

統計量。 對於每個變數來說，統計量包括：有效觀察值個數、平均數、和標準差。對每種模式而言：迴歸係數、相關矩陣、部分和偏相關、複相關係數 R、R^2、調整的 R^2、R^2 變量、估計值的標準誤、變異數分析摘要表、預測值、殘差。同時還有每個迴歸係數的 95% 信賴區間、變異數共變異數矩陣、變異數擴張因子、允差、Durbin-Watson 檢定、距離測量 (Mahalanobis、Cook、和影響量數)、DfBeta 值、Df 適合度、預測區間、和依觀察值順序診斷。圖形(T)：散佈圖、殘差散佈圖、直方圖、和常態機率圖。

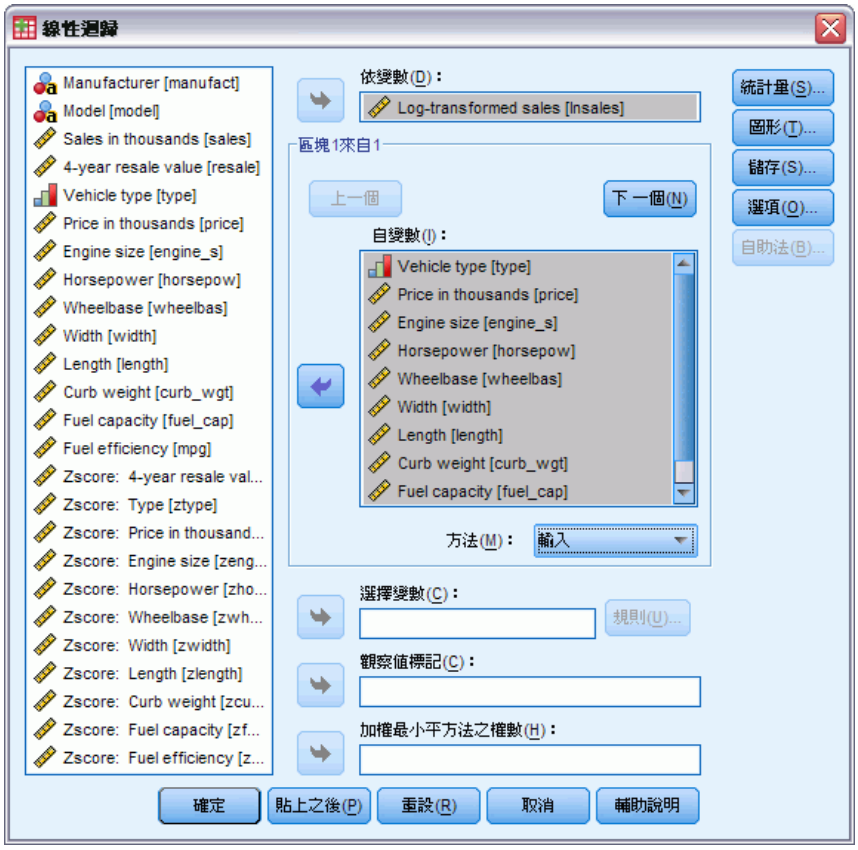
資料。 依變數和自變數應該都是數值變數。類別變數（如宗教、主修範圍、居住地區）都需要重新編碼為二元（虛擬）變數、或其他類型的對比變數。

假設。 對自變數的每個值而言，依變數的分配必須是常態的。對所有自變數數值而言，依變數分配的變異性，應該都是常數。依變數和每個自變數之間的關係，應該是線性的，而且所有觀察值應該互不相關。

線性迴歸分析的執行程序

▶ 從功能表選擇：
分析(A) > 迴歸 > 線性...

圖表 16-1
「線性迴歸」對話方塊



- ▶ 從「線性迴歸」對話方塊中，選取數值的依變數。
- ▶ 選取一個（或以上的）數值自變數。

您可以：

- 在區塊中輸入一組自變數，並為不同的變數子集指定不同的輸入方式。
- 選擇一個選取變數，以便限制只能分析具有此變數之某數值的觀察值子集。
- 選取觀察值識別變數，用以識別圖形上的某些點點。
- 選取一個「加權最小平方法之權數」數值變數以用於加權最小平方法分析。

WLS。 能讓您取得加權最小平方模式。資料點是由其變異數倒數來加權。這表示有大變異數的觀察值與小變異數相關的觀察值相比，前者對於分析的影響比較小。如果加權變數為零、負數或遺漏值，那麼分析時，會排除這個觀察值。

線性迴歸變數的選取法

您可以利用選擇方法來指定如何在分析中輸入自變數。您可以使用不同方法，從相同變數集來建構各種迴歸模式。

- **輸入（迴歸）。** 一種變數選取程序，在其單一步驟中會輸入區塊中的所有變數。

- **逐步。** 在每個步驟中，會輸入不在具有最小 F 機率值之方程式中的自變數（如果機率夠小）。如果其 F 機率值變得夠大，則會移除已在迴歸方程式中的變數。此方法會在沒有變數適合納入或移除之後終止。
- **移除。** 一種變數選取程序，在其單一步驟中會移除區塊中的所有變數。
- **向後消去法。** 一種變數選取程序，其中選取的所有變數在輸入方程式後，即會依序移除。若變數帶有與依變數的最小偏相關，則會首先考量移除。若其符合消去準則，則會予以移除。移除第一個變數之後，接下來會考量移除方程式中帶有最小偏相關的其他剩餘變數。若方程式中沒有任何滿足移除準則的變數，程序即會停止。
- **向前選取法。** 一種逐步變數選取程序，其會循序將變數輸入模式中。考量輸入方程式中的第一個變數是否會與依變數具有最大正相關或負相關。僅在滿足輸入準則時，才會將此變數輸入方程式中。若已輸入第一個變數，則接下來會考量具有最大偏相關且不在方程式中的自變數。若沒有變數符合輸入準則，則會停止此程序。

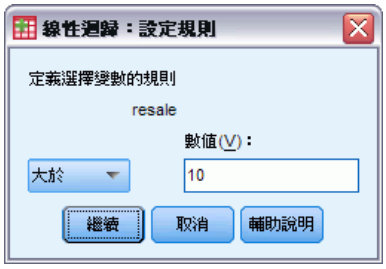
您輸出中的顯著值會符合某個單一模式。因此，如果使用逐步迴歸分析法（逐步迴歸分析法、向前法、或向後法）的話，一般而言，顯著值都是無效的。

無論所指定的輸入方法為何，所有的變數都必須通過允差準則，才能輸入方程式中。預設的容忍度為 0.0001。同時，如果某變數會導致其他已存在方法中之變數的允差度下降，並降到允差度準則以下，則這個變數便不會輸入。

所有選用的自變數，都會被加入單一迴歸模式中。然而，您可以替不同的變數子集，指定多種不同的輸入方式。例如，您可以透過「逐步選取」這種方式，將某個區塊中的變數輸入迴歸模式中，而第二個區塊則使用「向前選擇」方式。若要將第二個變數區塊加入迴歸模式中，請按一下**下一個**。

線性迴歸的設定規則

圖表 16-2
線性迴歸設定規則對話方塊



本次分析會包含所有透過選擇方式所定義出來的觀察值。例如，假設選擇條件為變數，再加上 = 符號和數字 5，如此代表只有選定變數之觀察值為 5 時，該觀察值才會被納入分析中。您也可以用字串值。

線性迴歸圖

圖表 16-3
線性迴歸圖」話方塊

您可以透過圖形，協助證實常態分配、直線性、變異數的相等之類的假設是否正確。圖形在偵測離群值、不尋常的觀察值、和具影響力的觀察值等方面，也非常有用。當您將上述觀察值存成新的變數之後，就可以使用「資料編輯程式」中的預測值、殘差、其他診斷，來建構附有自變數的圖形。您可以使用的圖形如下：

散佈圖。 您可以繪製下列任二個項目：依變數、標準化預測值、標準化殘差、已刪除殘差、調整預測值、Studentized 殘差、或 Studentized 已刪除殘差。請對照標準化預測值，繪製標準化殘差的圖形，藉以檢查變異數的直線性和相等性。

來源變數清單。 列出依變數 (DEPENDNT)，以及以下的預測變數與殘差變數。標準化預測值 (*ZPRED)、標準化殘差 (*ZRESID)、已刪除殘差 (*DRESID)、調整預測值 (*ADJPRED)、Studentized 殘差 (*SRESID)、Studentized 已刪除殘差 (*SDRESID)。

印出全部的殘差散佈圖。 當自變數和依變數分別迴歸至其他自變數上時，顯示每個自變數殘差和依變數殘差的散佈圖。如果要產生殘差散佈圖，該方程式至少需含有兩個自變數。

標準化殘差圖。 您可以取得標準化殘差直方圖，和常態機率圖（該圖比較標準化殘差分配跟常態分配的不同之處）。

若要求任一圖形，則會顯示標準化預測值和標準化殘差 (*ZPRED 和 *ZRESID) 的摘要統計量。

線性迴歸:若要儲存新變數

圖表 16-4
線性迴歸儲存對話方塊

您可以儲存預測值、殘差、和其他統計量以便診斷。每個選項都會新增一或多個新變數到您的作用中資料檔。

預測值。 迴歸模式會預測每個觀察值之未來數值。

- **未標準化 (U)。** 模式所預測的依變數數值。
- **標準化 (A)。** 轉換每個預測值成其標準化形式。也就是說，平均數預測值會從預測值中減去，而差分會除以預測值的標準差。標準化預測值的平均數為 0，標準差為 1。
- **調整後 (J)。** 從迴歸係數計算中排除之觀察值的預測值。
- **平均數與預測值的標準誤 (P)。** 預測值的標準誤。估計依變數平均值的標準差，它是為與自變數有相同數值的觀察值而進行的估計。

距離。 識別有異常自變數組合值之觀察值、或對迴歸模式有重大影響之觀察值時，所使用的量數。

- **Mahalanobis(H)。** 測量自變數之觀察值數值與整體觀察值平均數的變異程度。若 Mahalanobis 距離較大，則會將觀察值識別為在一個以上自變數中具有極端值。
- **Cook's(K)。** 若自迴歸係數計算中排除特定觀察值，則其會測量所有觀察值的殘差變更程度。若 Cook's D 較大，則表示自迴歸統計量計算中排除某觀察值已足以造成係數變更。
- **影響量數(L)。** 測量迴歸適合度中的點影響。置中影響量數的範圍介於 0（適合度中無影響）至 (N-1)/N 之間。

預測區間。 平均數、個別預測區間的上界和下界。

- **平均數。** 預測回應平均數之預測區間的上界與下界（兩變數）。
- **個別(I)。** 單一觀察值依變數的預測區間上界和下界（兩個變數）。
- **信賴區間。** 輸入介於 1 和 99.99 之間的數值，以指定兩個「預測區間」的信賴水準。輸入此數值之前，必須選取「平均數」或「個別值」。一般的信賴區間值為 90、95 和 99。

殘差。 該值算法為依變數的實際數值，再減去迴歸方琵式所預測之數值。

- **未標準化(U)。** 觀察值和模式所預測的數值之間的差異。
- **標準化(A)。** 殘差除以其標準差的估計值。標準化殘差（也稱為 Pearson 殘差）的平均數為 0，標準差為 1。
- **學生化(S)。** 殘差會根據自變數的平均數到自變數中每個觀察值的數值之距離除以隨其觀察值類型變化之標準差的估計值。
- **已刪除(D)。** 從迴歸係數計算中排除之觀察值的殘差。其為依變數值與已調整預測值間的差異。
- **學生化去除殘差(E)。** 觀察值除以其標準誤的已刪除殘差。介於 Studentized 已刪除殘差和其相關 Studentized 殘差之間的差分，表示刪除觀察值對其自身的預測所造成之差異。

影響統計量。 因為排除特定觀察值之故，造成迴歸係數 (DfBeta) 和預測值 (Df 適合度) 的改變。標準化 DfBeta 和 DfFit 值也可以跟共變異數比值一起使用。

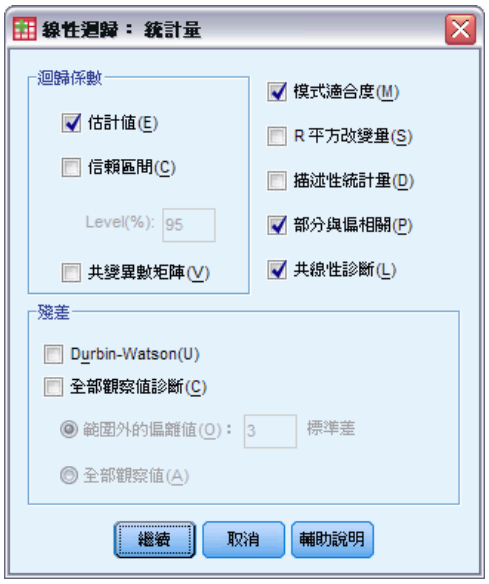
- **DfBeta(s)。** beta 值的差異即為因排除特定觀察值而導致的迴歸係數變更。數值是針對模式中的每個項目計算（包含常數）。
- **標準化 DfBeta。** 以 beta 值標準化的差分。因為排除特定觀察值之故，造成迴歸係數的改變。您可能要檢驗絕對值除以 N 的平方根大於 2 之觀察值，其中 N 為觀察值個數。數值是針對模式中的每個項目計算（包含常數）。
- **DfFit。** 適合度值的差異即為因排除特定觀察值而導致的預測值變更。
- **標準化 DfFit。** 以適合度值標準化的差分。因為排除特定觀察值之故，造成預測值的改變。您可能要檢驗絕對值超過 p/N 平方根兩倍的標準化值，其中 p 為模式中參數的數量，N 為觀察值的個數。
- **共變異數比值(V)。** 帶有自迴歸係數計算中排除之特定觀察值的共變異數矩陣行列式，與包含所有觀察值之共變異數矩陣行列式的比例。若此比例趨近於 1，則觀察值不會明顯變更共變異數矩陣。

係數統計量。 將迴歸係數儲存到資料集或資料檔案中。資料集可供同一階段作業中之後續的作業使用，但是不會儲存為檔案，除非特別在階段作業結束之前儲存。資料集名稱必需符合變數命名規則。

匯出模式資訊至 XML 檔。 將參數估計與（選擇性）其共變異數以 XML（PMML）格式匯出到指定檔案中。您可以使用這個模式檔案，將模式資訊套用到其他資料檔案中以進行評分工作。

線性迴歸的統計量

圖表 16-5
「統計量」對話方塊



您可以使用的統計量如下：

迴歸係數。 **估計值**可顯示迴歸係數 B、B 的標準誤、標準化係數 beta、B 的 t 值，以及 t 的雙尾顯著性水準。**信賴區間**可顯示每個迴歸係數或共變異數矩陣具有指定信賴水準的信賴區間。**共變異數矩陣**可顯示迴歸係數的變異數 - 共變異數矩陣，此迴歸係數中的共變異數在對角線以外、而變異數在對角線上。此外，還會顯示一個相關矩陣。

模式適合度。 列出輸入模式、和從模式中移除的變數，並顯示下列適合度統計量。複相關係數 R、R^2 和調整過的 R^2、估計值的標準誤，以及變數分析表。

R 平方改變量。 R^2 統計量的變更是藉由新增或刪除自變數而產生的。如果與變數相關的 R^2 變更很大，就表示該變數是依變數的良好預測變數。

描述性統計量。 該選項會提供分析中，每個變數的有效觀察值個數、平均數、標準差。此外，還會顯示一個相關矩陣，該矩陣含有單尾顯著水準，以及每個相關的觀察值個數。

偏相關。 在移除兩項變數與其他變數相互關聯性的相關後，餘留在兩變數之間的相關。當模式內其他自變數的線性效應已從依變數和自變數移除後，兩變數之間的相關。

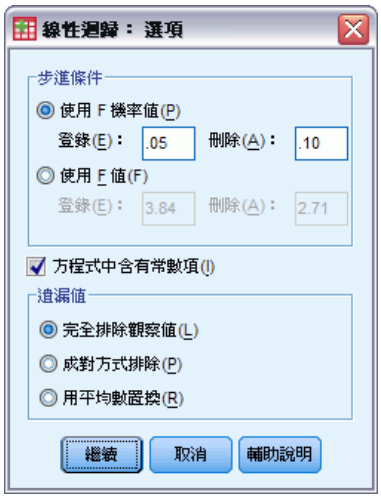
部分相關。 當模式內其他自變數的線性效應已從一個自變數移除後，依變數和自變數之間的相關。它跟方程式加入一個變數，而 R 平方改變有關。有時亦稱作半偏相關。

共線性診斷。 共線性（或多重共線性）是當自變數是其他自變數的線性函數時的不理想狀況。該選項會將（尺度化和未集中的）交叉相乘矩陣、條件索引、變異數分解比例的特徵值，跟變異數擴張因子 (VIF)、個別變數的允差，一起顯示出來。

殘差。顯示殘差和依觀察值診斷之數列相關的 Durbin-Watson 檢定，此處的觀察值必須符合選擇條件（偏離值在 n 標準差以上者）。

線性迴歸的選項

圖表 16-6
線性迴歸選項對話方塊



您可以使用的選項如下：

步進條件。 當指定向前、向後或逐步變數選擇方法時，會套用這些選項。變數可以自模式中輸入或移除，需視 F 值的顯著性（機率）或 F 值本身來決定。

- **使用 F 機率。** 當變數之 F 值顯著水準小於「輸入」值時，系統會將變數輸入模式，而當顯著水準大於「移除」值時，系統會將變數移除。「輸入」必須小於「移除」，而且兩個數值都必須是正數。若要將更多變數輸入模式，請調高「輸入」值。若要從模式中移除更多變數，請調低「移除」值。
- **使用 F 值。** 當變數的 F 值大於「輸入」值時，系統會將該變數輸入模式，而當 F 值小於「移除」值時，系統會將變數移除。「輸入」必須大於「移除」，而且兩個數值都必須是正數。若要將更多變數輸入模式，請調低「輸入」值。若要從模式中移除更多變數，請調高「移除」值。

方程式中含有常數項。 根據預設值，迴歸模式裡面會包含一個常數項。取消選取這個選項，會強制迴歸通過原點，因此我們很少這樣做。您不能將通過原點之迴歸的某些結果，拿來跟含常數項之迴歸結果相比較。例如，您不能用一般的方式來解譯 R^2 。

遺漏值。 您可以任選一個選項：

- **完全排除遺漏值。** 如果觀察值對於每個變數而言，其值都是有效的，則該觀察值才會納入分析中。

- **成對方式排除。** 含有成對相關變數完整資料的觀察值，可用來計算相關係數，而迴歸將根據此係數來進行分析。自由度的依據則是成對的 N的最小值。
- **用平均數置換。** 這個選項會計算每一個觀察值，並且用變數的平均數來取代漏掉的觀察值。

REGRESSION 指令的其他功能

指令語法語言也可以讓您：

- 寫入相關矩陣、或讀取矩陣，以取代原始資料，並取得迴歸分析（利用 MATRIX 次指令）。
- 指定允差水準（利用 CRITERIA 次指令）。
- 取得相同（或不同）依變數的多重模式（利用 METHOD 和 DEPENDENT 次指令）。
- 取得其他統計量（利用 DESCRIPTIVES 和 STATISTICS 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
17

次序迴歸

「次序的迴歸」可讓您做出多分類次序的反應值與一組預測值（可為因子或共變量）的從屬模型。「次序的迴歸」之設計係以 McCullagh (1980, 1998) 方法論為基礎，而且其程序在語法上被稱為 **PLUM**。

標準的線性迴歸分析，會將反應值（依變數）和加權預測（自）變數間差異的平方和最小化。其中，估計係數可反應出預測值的改變，對依變數所造成的影響。因反應值水準之變更在所有依變數範圍內都是相同的，所以依變數係假設為數字的。例如，150 公分高的人及 140 公分高的人之間，身高差距為 10 公分，這與身高 210 公分高及 200 公分高的兩人之間，身高差距的意義是一樣的。但這些關係對次序的變數並非是必要的，因其選擇及依變數類別個數可能是非常多變的。

範例。「次序的迴歸」可用來研究病患對藥物劑量的反應。可能的反應可分為「無」、「輕微」、「溫和」或「嚴重」。但輕微與溫和反應之間的差距是很難量化或無法量化，而且是以感覺為依據的。此外，輕微與溫和反應之間的差距可能會大於或小於溫和與嚴重反應之間的差距。

統計量與圖形。 共有觀察與期望的次數分配及累積次數分配，次數分配及累積次數的 Pearson 殘差，觀察與期望機率，依共變量樣式所劃分的每個反應類別的觀察與期望累積機率，參數估計值的漸進線相關與共變異數矩陣，Pearson’s 卡方與概似比卡方，適合度統計量，疊代過程，平行線假設檢定，參數估計值，標準誤，信賴區間，以及 Cox 與 Snell’s、Nagelkerke’s 及 McFadden’s R^2 統計量。

資料。 依變數係假設為次序的，而且可以是數字或字串。其次序是依遞增順序排序依變數數值而決定的，最低值會定義第一個類別。因子變數係假設為類別的。共變量變數必須為數字變數。請注意，使用一個以上的連續共變量，很容易導致建立很大的儲存格機率表。

假設。 僅能有一個反應值，而且它必須是經過指定的。此外，對所有自變數數值的每個不同樣式而言，依變數是假設為多項式自變數。

相關程序。 名義 logistic 迴歸也會對名義依變數使用類似模式。

若要取得次序的迴歸

▶ 從功能表選擇：
分析(A) > 迴歸 > 次序的...

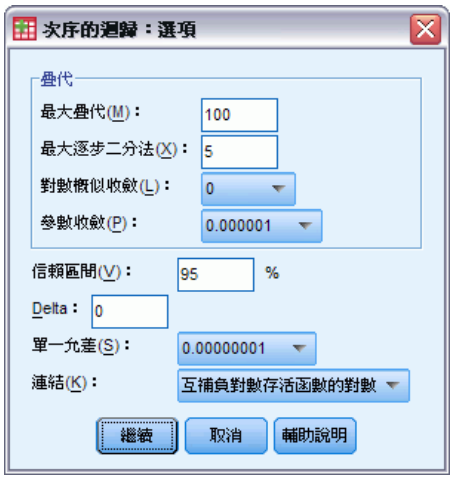
圖表 17-1
「次序的迴歸」對話方塊

- 選擇一個依變數。
- 按一下「確定」。

次序的迴歸的選項

「選項」對話方塊可讓您調整疊代估計演算法所用的參數，選擇您參數估計值的信心水準，以及選擇連結函數。

圖表 17-2
「次序的迴歸選項」對話方塊



疊代。 您可以自訂疊代的演算法。

- **最大疊代。** 指定一個非負的整數。如果指定 0，此程序就會傳回起始估計值。
- **最大的半階次數。** 指定一個正整數。

- **對數概似收斂。** 如果對數概似的絕對或相對變更小於這個數值的話，演算法就會停止。如果指定 0 便不會使用這個條件。
- **參數收斂條件。** 如果參數估計值的絕對或相對變更小於這個數值的話，演算法就會停止。如果指定 0 便不會使用這個條件。

信賴區間。 指定大於等於 1 且小於 100 的一個數值。

Delta 值。 此為加到零格次數的數值，請指定小於 1 的非負數值。

奇異性容忍值。 可用於檢查高度依賴的預測值，請從選項清單中選擇一個數值。

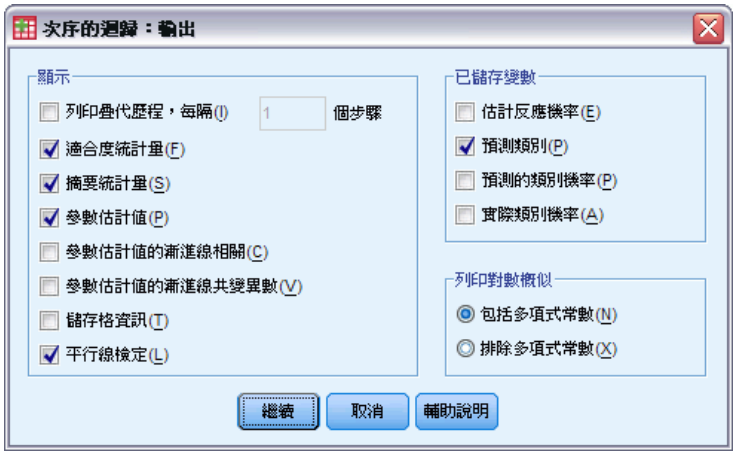
連結函數。連結函數是一個累積機率的轉換，可允許模式估計。有 5 個連結函數可使用，摘要顯示於下表。

函數	形式	一般應用程式
Logit	$\log(\xi / (1-\xi))$	平均分配類別
互補對數存活函數的對數	$\log(-\log(1-\xi))$	類別愈高，愈可能
負對數存活函數的對數	$-\log(-\log(\xi))$	類別愈低，愈可能
Probit	$\Phi^{-1}(\xi)$	潛在變數會常態分配
Cauchit (倒數 Cauchy)	$\tan(\pi(\xi-0.5))$	潛在變數有許多極端值

次序的迴歸輸出

「輸出」對話方塊可讓您產生在「瀏覽器」中顯示的表格，以及將變數儲存至工作檔中。

圖表 17-3
「次序的迴歸輸出」對話方塊



顯示。 可產生下列項目的表格：

- **列印疊代過程。** 對數概似及參數估計值會依指定的列印疊代次數印出，但第一步及最後一步疊代永遠會列出來。
- **適合度統計量。** Pearson 及概似比卡方統計量，會依變數清單中指定的分類而計算。
- **摘要統計量。** Cox 與 Snell’s、Nagelkerke’s 及 McFadden’s R^2 統計量。
- **參數估計值。** 參數估計值、標準誤及信賴區間。

- **參數估計值的漸進線相關。**參數估計值相關矩陣。
- **參數估計值的漸進線共變異數。**參數估計值共變異數矩陣。
- **儲存格資訊。**共有觀察與期望的次數分配及累積次數分配，次數分配及累積次數的 Pearson 殘差，觀察與期望機率，以及依共變量樣式所劃分的每個反應值類別的觀察與期望累積機率。請注意，對有許多共變量樣式（例如有連續共變量的模式）的模式而言，這個選項可能會產生很大而不便使用的表格。
- **平行線檢定。** 位置參數在依變數所有水準下均相同的假設檢定。此檢定僅可使用於位置唯一模式。

已儲存變數。 可將下列變數儲存至工作檔中：

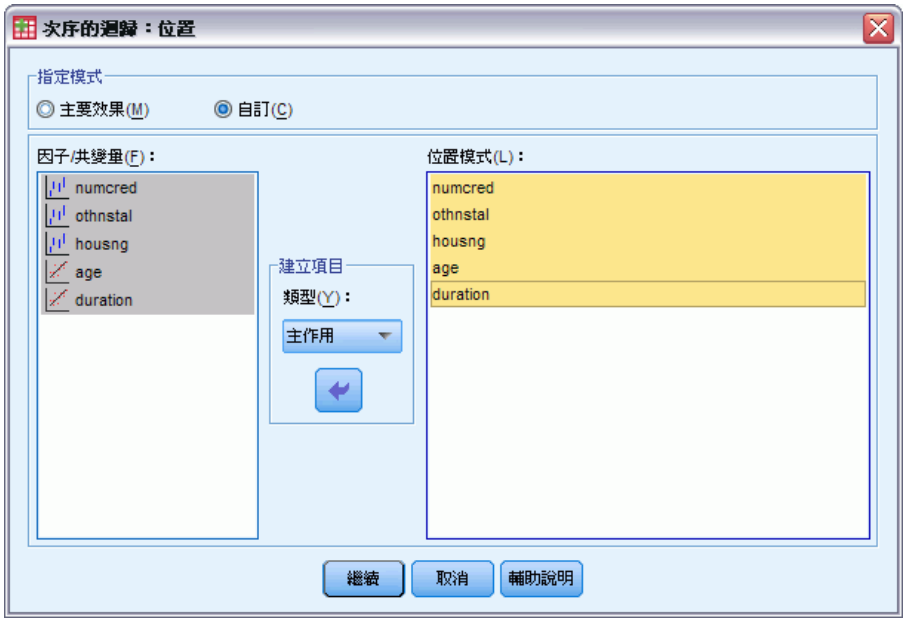
- **估計各類別的反應機率。** 此為將因子/自變數型態組合分類至反應值類別的模式估計機率。機率個數和反應類別個數一樣多。
- **預測應答組類。** 此為具因子/自變數型態組合最大估計機率的反應類別。
- **預測應答機率。** 此為將因子/自變數型態組合分類至預測應答組類的估計機率。此機率也是因子/自變數型態組合估計機率的最大值。
- **實際應答機率。** 此為將因子/自變數型態組合分類至實際類別的估計機率。

列印對數概似。 這會控制對數概似的顯示。**包含多項式常數項**可帶給您完全的概似值。若要在所有不包括此常數的結果中比較您的結果，您可以選擇排除此常數。

次序的迴歸的位置模式

「位置」對話方塊可讓您指定分析的位置模式。

圖表 17-4
「次序的迴歸位置」對話方塊



指定模式。 主效果模式會包括共變量及因子的主效果，但不包括交互作用效應項。您可以建立自訂模式，以指定因子或共變量交互作用子集。

因子/共變量。因子與共變量均會列出。

位置模式。此模式會依您所選的主效果及交互作用項而定。

建立效果項

對所選擇的因子和共變量而言：

交互作用。 建立所有選定變數的最高階交互作用項。此為預設值。

主效果。 為每個選擇的變數，建立主效果。

完全二因子。為所選的變數，建立所有可能的二因子交互作用。

完全三因子。為所選的變數，建立所有可能的三因子交互作用。

完全四因子。為所選的變數，建立所有可能的四因子交互作用。

完全五因子。為所選的變數，建立所有可能的五因子交互作用。

次序的迴歸尺度模式

「尺度」對話方塊可讓您指定分析的尺度模式。

圖表 17-5
「次序的迴歸尺度」對話方塊

因子/共變量。因子與共變量均會列出。

尺度模式。此模式會依您所選的主要及交互作用項而定。

建立效果項

對所選擇的因子和共變量而言：

交互作用。 建立所有選定變數的最高階交互作用項。此為預設值。

主效果。 為每個選擇的變數，建立主效果。

完全二因子。為所選的變數，建立所有可能的二因子交互作用。

完全三因子。為所選的變數，建立所有可能的三因子交互作用。

完全四因子。為所選的變數，建立所有可能的四因子交互作用。

完全五因子。為所選的變數，建立所有可能的五因子交互作用。

PLUM 指令的其他功能

如果您將選擇部分貼入語法視窗，並且編輯所產生的 PLUM 指令語法的話，就可以自訂次序的迴歸。指令語法語言也可以讓您：

- 藉由將虛無假設指定成參數線性組合，來建立自訂的假設檢定

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
18

曲線估計

您可以利用「曲線估計」程序，來產生曲線估計迴歸統計分析，以及 11 種不同曲線估計迴歸模式的相關圖形。每個依變數都會產生一個不同的模式。此外，您也可以將預測值、殘差和預測區間存成新的變數。

範例。 網際網路服務提供者會隨時間追蹤網路上已被病毒感染之電子郵件的流量百分比。散佈圖顯示關係為非線性關係。反之，您可以將某個二次或三次模式套用到這些資料上，並檢查假設的有效程度，以及模式的適合度。

統計量。 對於每個模式：迴歸係數、複相關係數 R、R^2、調過後的 R^2、估計值的標準誤、變異數分析摘要表、預測值、殘差和預測區間。模式方面，則有：線性、對數、倒數、二次、三次、冪次、複合、S 曲線、Logistic、成長以及指數等。

資料。 依變數和自變數應該都是數值變數。如果您從作用中的資料集中選擇「時間」做為自變數（而不選擇變數），那麼「曲線估計」程序就會產一個時間變數，而每個觀察值之間的時間長度都會相同。如果您選擇 時間，那麼依變數就應該是一個時間數列的測量。時間數列分析需要一個資料檔案，其結構中的每個觀察值（列）分別代表一組不同時間的觀察值，而且每個觀察值之間的時間長度都相同。

假設。 將您的資料繪製成圖形，以判斷自變數和依變數的關係（線性、指數等等）。一個良好模式的殘差，應該呈常態的隨機分配。如果您使用線性模式，就應該符合下列的假設：對自變數的每個值而言，依變數的分配必須是常態的。對所有自變數數值而言，依變數分配的變異性，應該都是常數。自變數和依變數之間的關係應該為線性，而所有的觀察值應該各自獨立。

若要取得曲線估計

▶ 從功能表選擇：
分析 > 迴歸方法 > 曲線估計...

圖表 18-1
「曲線估計」對話方塊

- 選擇一個或多個依變數。每個依變數都會產生一個不同的模式。
- 選擇一個自變數（在作用中的資料集中選擇變數或選擇「時間」）。
- 隨意：
  - 選取一個變數，以便在散佈圖中為觀察值加上標記。至於散佈圖中的各點，您可以使用「點選擇」工具來顯示「觀察值標記」變數值。
  - 按一下「儲存」，以便將預測值、殘差、和預測區間存成新的變數。

您也可以使用下列的選項：

- **方程式中含有常數項。**在迴歸方程式中的估計常數項。依照預設值，這個常數應該已經包含在內。
- **印出模式之線性圖。**繪製依變數的數值，以及根據自變數所選取的每個模式。而且每個依變數都會產生一個圖表。
- **顯示 ANOVA 摘要表。** 針對每個選擇的模式，顯示一個變異數分析摘要表。

曲線估計模式

您可以選取一個或多個曲線估計迴歸模式。若要判斷應使用何種模式，請先將您的資料繪製成圖形。如果您的變數呈線性相關，請使用簡單的線性迴歸模式。不過，如果您的變數並非線性相關時，請先嘗試轉換資料。如果轉換後也沒差別時，您可能需要使用較複雜的模式。那麼，請先檢視資料的散佈圖。 如果圖形跟您所知的數學函數很像，請將資料套用到該模式類型上。譬如，我們假設資料類似指數函數，就請使用指數模式。

線性。 方程式為 Y = b0 + (b1 * t) 的模式。此數列值會模式化為時間線性函數。

對數。 方程式為 Y = b0 + (b1 * ln(t)) 的模式。

倒數。 方程式為 Y = b0 + (b1 / t) 的模式。

二次模式。 此模式的方程式為 Y = b0 + (b1 * t) + (b2 * t**2)。二次模式可用於為「起飛」的數列或減幅的數列建立模式。

三次模式。 由 Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3) 方程式定義的模式。

冪次。 此模式的方程式為 Y = b0 * (t**b1) 或 ln(Y) = ln(b0) + (b1 * ln(t))。

複合。 方程式為 Y = b0 * (b1**t) 或 ln(Y) = ln(b0) + (ln(b1) * t) 的模式。

S 曲線。 方程式為 Y = e**(b0 + (b1/t)) 或 ln(Y) = b0 + (b1/t) 的模式。

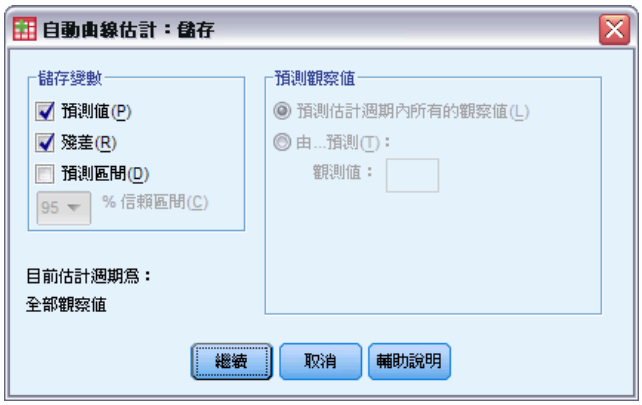
Logistic。 方程式為 Y = 1 / (1/u + (b0 * (b1**t))) 或 ln(1/y-1/u) = ln (b0) + (ln(b1) * t) 的模式，其中 u 為上界值。選取 Logistic 後，指定上界值以便用於迴歸方程式。此數值必須為大於最大依變數值的正數。

成長。 方程式為 Y = e**(b0 + (b1 * t)) 或 ln(Y) = b0 + (b1 * t) 的模式。

指數。 方程式為 Y = b0 * (e**(b1 * t)) 或 ln(Y) = ln(b0) + (b1 * t) 的模式。

曲線估計儲存

圖表 18-2
「曲線估計儲存」對話方塊



儲存變數。對每一種選擇的模式，您都可以儲存預測值、殘差（依變數的觀察值減去模式的預測值）和預測區間（上界與下界）。新的變數名稱和描述性標記，都會顯示在輸出視窗的表格中。

預測觀察值。如果您在作用中的資料集中選擇「時間」而非變數來做為自變數，那麼您就可以在時間數列結束後，指定一個預測期間。您可以任選下列其中一個選項：

- **預測估計週期內所有的觀察值。**以預測期間內的觀察值為基礎，來預測檔案中所有觀察值的數值。在此處的預測期間（顯示在對話方塊底部），是由「資料」功能表上「選擇觀察值」的「範圍」次對話方塊來定義的。如果沒有預測期間定義，則所有觀察值都會用來預測數值。
- **由...預測。**以預測期間內的觀察值為基礎，來預測經過指定日期、指定時間、或觀察個數的數值。這個功能可以用來預測時間數列中、最後一個觀察值以外的數值。目前定義的日期變數能決定哪一個文字方塊可以用來指定結束預測期間。如果沒有定義資料變數，您可以指定結束的觀察（觀察值）號碼。

若要建立資料變數，請使用「資料」功能表上的「定義日期」選項。

章節
19

偏最小平方迴歸

「偏最小平方迴歸」程序能夠估計偏最小平方 (PLS，亦稱為「對潛在結構的投影」) 迴歸模式。PLS 是一種預測技術，能替代普通最小平方法 (OLS) 迴歸、典型相關或結構方程式模式，而且當預測變數有高度相關或當預測變數數目超過觀察值數目時，PLS 特別有用。

PLS 結合主成分分析和多重迴歸的特性。首先它會先萃取出一組能夠盡可能高度解釋依變數和自變數之間共變異數的潛在因子。接著，迴歸會使用自變數的分解逐步預測依變數的值。

可用性。 PLS 是一種延伸指令，執行時需要在您計劃執行 PLS 的系統上安裝 IBM® SPSS® Statistics - Integration Plug-In for Python。「PLS 延伸模組」必須個別安裝，其可下載自 http://www.spss.com/devcentral.

注意：「PLS 延伸模組」會視 Python 軟體而異。SPSS Inc. 並非 Python 軟體的擁有者或授權者。Python 的任何使用者必須同意 Python 網站上的 Python 授權合約條款。SPSS Inc. 未對 Python 程式的品質做出任何聲明。SPSS Inc. 對於您對 Python 軟體之使用概不負責。

表格。 (由潛在因子) 解釋的變異數比例、潛在因子加權值、潛在因子負荷量、投影中自變數重要性 (VIP) 和迴歸參數估計值 (根據依變數) 等，皆會依照預設值產生。

圖表。 投影中變數重要性 (VIP)、因子分數、前三個潛在因子的因子加權值和到模式的距離，全都從「選項」索引標籤中產生。

測量水準。 依變數和自變數 (預測變數) 可以是尺度、名義或次序。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。本程序會以相同方式處理類別 (名義或次序) 變數。

類別變數編碼。 本程序在整個程序期間，會使用其中一個 c 編碼來暫時記錄類別依變數。如果一個變數有多個 c 類別，則變數會儲存為 c 向量，其中第一個類別標示為 (1,0,...,0)，第二個類別標示為 (0,1,0,...,0)，...，最後一個類別標示為 (0,0,...,0,1)。類別依變數會使用虛擬編碼來表示，也就是只略過與參考類別相關的指標。

次數加權。 使用加權值前，會將加權值捨入為最接近的整數。分析中不會使用遺漏加權值或加權值小於 0.5 的觀察值。

遺漏值。 使用者遺漏值和系統遺漏值會視為無效。

調整。 所有模式變數會集中並標準化，包括代表類別變數的指標變數。

若要取得偏最小平方迴歸

從功能表選擇：
分析(A) > 迴歸 > 偏最小平方...

圖表 19-1
「偏最小平方迴歸變數」索引標籤

- ▶ 選取至少一個依變數。
- ▶ 至少要選取一個自變數。

您可以：

- 指定類別（名義或次序）依變數的參考類別。
- 指定變數做為觀察值輸出和所儲存資料集的唯一識別碼。
- 指定要萃取出潛在因子數目的上限。

模式

圖表 19-2
「偏最小平方迴歸模式」索引標籤

指定模式效應。 主效果模式包含所有因子和共變數主效果。選取「自訂」以指定交互作用。但是您必須指出所有會被放入模式的項目。

因子與共變量。因子與共變量均會列出。

模式。 模式端視您的資料性質而定。在選擇「自訂」之後，您就可以選擇欲分析之主效果和交互作用。

建立項目

對所選的因子和共變量而言：

交互作用。 建立所有選定變數的最高階交互作用項。此為預設值。

主效果。 為每個所選的變數，建立主效果。

完全二因子。為所選的變數，建立所有可能的二因子交互作用。

完全三因子。為所選的變數，建立所有可能的三因子交互作用。

完全四因子。為所選的變數，建立所有可能的四因子交互作用。

完全五因子。為所選的變數，建立所有可能的五因子交互作用。

選項

圖表 19-3
「偏最小平方迴歸選項」索引標籤

「選項」索引標籤可讓使用者儲存和繪製個別觀察值、潛在因子和預測變數的模式估計值。

為每種類型的資料指定資料集的名稱。資料集名稱必須獨一無二。如果您指定現有資料集的名稱，則會取代其內容，否則會建立新的資料集。

- **儲存個別觀察值的估計值。** 儲存以下的觀察值模式估計值：預測值、殘差、到潛在因子模式的距離和潛在因子分數。它也會繪製潛在因子分數。
- **儲存潛在因子的估計值。** 儲存潛在因子負荷量和潛在因子加權值。它也會繪製潛在因子加權值。
- **儲存自變數的估計值。** 儲存迴歸參數估計值和投影的變數重要性 (VIP)。它也會按照潛在因子繪製 VIP。

章節
20

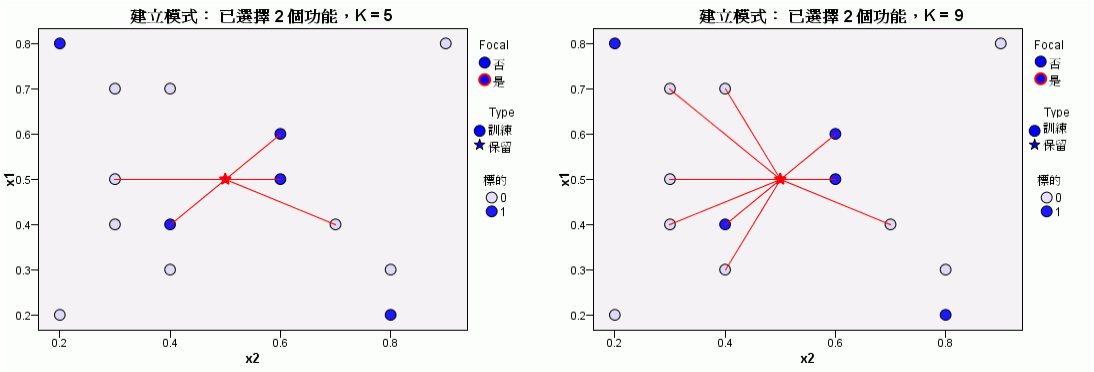
最近鄰法分析

最近鄰法分析是以和其他觀察值的相似性為基礎來分類觀察值的方法。在機器學習中，這是辨認資料形式的方法，完全不需要確切符合任何已儲存的形式或觀察值。相似的觀察值會彼此相鄰，相異的觀察值則會彼此相隔。因此，兩個觀察值相距的距離可用來判斷彼此的相異性。

互相接近的觀察值稱為 “neighbors。”新的觀察值（保留）存在時，會計算模式中各觀察值的距離。計算最相似觀察值的分類 - 最近鄰法 - ，新觀察值會放在包含最近鄰法中個數最多的類別。

您可以指定要檢驗的最近鄰法個數；此數值稱為 k。圖片顯示新觀察值如何使用兩種不同的 k 值來分類。當 k = 5 時，新的觀察值放在類別 1，因為大多數的最近鄰法屬於類別 1。但是，當 k = 9 時，新的觀察值放在類別 0，因為大多數的最近鄰法屬於類別 0。

圖表 20-1
在分類中變更 k 的效果



最近鄰法分析也可以用來計算連續目標的數值。在此狀況下，會使用最近鄰的平均數或中位數目標值來取得新觀察值的預測值。

目標與功能。目標與功能包括：

- **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。

由最近鄰法分析以相同方式處理名義和次序變數。本程序假設已指定給各個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				

類別變數編碼。本程序在整個程序期間，會使用 one-of-c 編碼來暫時記錄類別預測變數和依變數。如果一個變數有多個 c 類別，則變數會儲存為 c 向量，其中第一個類別標示為 (1,0,...,0)，第二個類別標示為 (0,1,0,...,0)，...，最後一個類別標示為 (0,0,...,0,1)。

此編碼架構會增加功能空間的維度。特別是，維度的總數等於尺度預測數目加上所有類型預測的類型數目。因此，此編碼架構會導致訓練變慢。如果您的最近鄰法訓練進行的非常慢，在執行程序前，您可以將類似的類別組合在一起，或捨棄具有極少類別的觀察值，以嘗試減少類別預測變數中的類別個數。

所有的 one-of-c 編碼都是以訓練資料為基礎，即使已定義保留樣本也是如此（請參閱 區隔 ）。因此，如果保留樣本所包含的觀察值之預測值類別不在訓練資料中，則不會評定這些觀察值。如果保留樣本所包含的觀察值之依變數類別不在訓練資料中，便會評定這些觀察值。

調整。尺度功能預設會經過常態化。所有的調整都是以訓練資料為基礎，即使已定義保留樣本也是如此（請參閱 區隔 第 121 頁）。如果您指定變數來定義分割，很重要的是這些功能必須在訓練樣本和保留樣本之間有類似的分配。 例如使用「預檢資料」程序來檢驗跨越分割的分配。

次數加權。 此程序會忽略次數加權。

複製結果。在隨機指派分割和交叉驗證折疊期間，此程序會使用亂數產生器。如果您要精確複製結果，除了使用相同的程序設定外，另外需要設定 Mersenne Twister 的種子（請參閱 區隔 第 121 頁），或使用定義分割和交叉驗證折疊的變數。

取得最近鄰法分析

從功能表選擇：

分析(A) > 分類 > 最近鄰法(N)...

圖表 20-2
最近鄰法分析變數索引標籤

- 指定一個或多個功能，如果有目標，可將功能視為獨立變數或預測變數。

目標（選用）。如果未指定任何目標（依變數或回應），則程序只會找出 k 個最近鄰 - 不會進行任何分類或預測。

將尺度功能常態化。經過常態化的功能具有相同的值範圍，可改善估計演算法的效能。使用調整後常態化 [2*(x–min)/(max–min)]–1。調整後常態化的值介於 –1 和 1 之間。

焦點觀察值識別碼（選用）。這可供您標記特別感興趣的觀察值。例如，研究人員想要判斷一個學區 - 焦點觀察值 - 的測驗分數是否及於類似學區的測驗分數。他使用最近鄰法分析，找出在多個特定功能方面最相似的學區。然後他比較焦點學區的測驗分數與最鄰近學區的測驗分數。

焦點觀察值也可用於臨床研究，以選擇與臨床病例相似的控制觀察值。焦點觀察值會顯示於 k 個最近鄰與距離表、功能空間圖、對等圖和象限地圖中。焦點觀察值的資訊會儲存在「輸出」索引標籤上指定的檔案中。

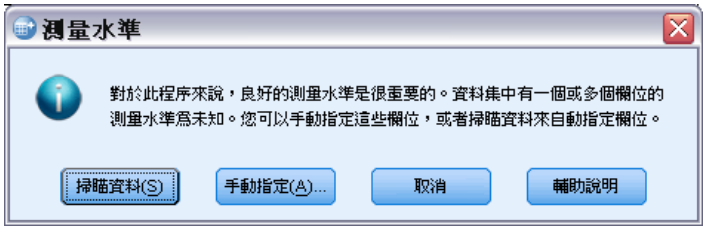
其中會將指定變數為正值的觀察值視為焦點觀察值，而無法指定不含正值的變數。

觀察值標記（選用）。觀察值的標記是以功能空間圖、對等圖和象限地圖中的值予以設定。

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 20-3
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「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。

相鄰

圖表 20-4
最近鄰法分析相鄰索引標籤

最近鄰數目 (k)。指定最近鄰的數目。請注意，使用較大相鄰數目未必可得出較精確的模式。

如果在「變數」標籤指定目標，您也可以指定數值範圍，並允許程序選擇範圍內最適當的相鄰數目。決定相鄰數目的方法需視「功能」索引標籤上是否需要功能選擇而定。

- 如果功能選擇有效，則會針對所要求範圍中 k 的各個值、k 和相伴隨功能集執行功能選擇，其中會選擇最低的錯誤率（如果目標是錯誤，則會選擇最低的平方和錯誤）。
- 如果功能選擇無效，則會使用 V 折疊交叉驗證，以選擇相鄰的 “最佳” 數目。請參閱「區隔」索引標籤，以瞭解折疊指派的控制。

距離計算。這是指定距離矩陣所用的矩陣，可用來測量觀察值的相似性。

- **歐基里得矩陣。**兩個觀察值 x 和 y 之間的距離就是觀察值之間差異平方的所有維度總和平方根。
- **城市區塊矩陣。**兩個觀察值之間的距離就是觀察值數值之間差異的所有維度總和。也稱為 Manhattan 距離。

如果在「變數」索引標籤上指定目標，您也可以在計算距離時按照經過常態化的重要性選擇權重功能。預測的功能重要性，是以已從模式中去除預測的模式錯誤率或平方和錯誤與完全模式的錯誤率或平方和錯誤相比的比例進行計算。經過常態化的重要性，是以重新加權功能重要性值進行計算，因此總和為 1。

尺度目標的預測。如果在「變數」索引標籤上指定尺度目標，則這會指定是否根據最近鄰法的平均數或中位數值計算經過預測的值。

功能

圖表 20-5
最近鄰法分析功能索引標籤

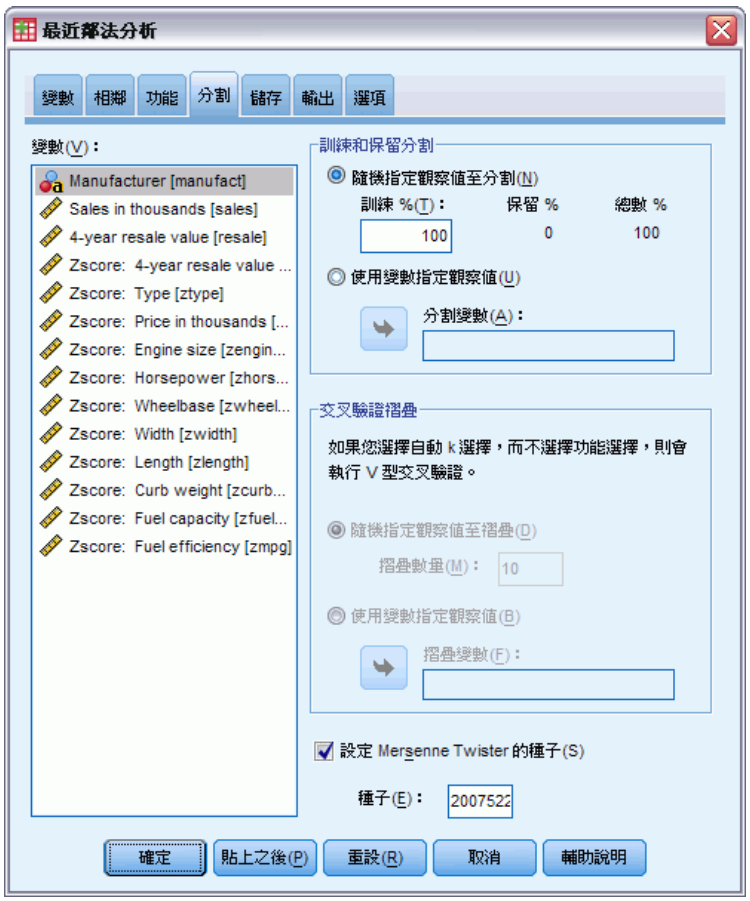
在「變數」標籤上指定目標時，「功能」索引標籤可供您要求和指定功能選擇的選項。依照預設值，會針對功能選擇考量所有功能，但是您可以選擇強制用於模式中的功能子集。

停止準則。在各個步驟中，會將最小錯誤中加總為模式結果的功能（計算成為類別目標的錯誤率與尺度目標的平方和錯誤）視為包含在模式集中。向前選取法會持續進行，直到符合指定的條件為止。

- **指定的功能數目。**在強制用於模式的功能之外，此演算法會加上固定數目的功能。指定一個正整數。降低數目的值會建立較精簡的模式，但是有可能遺漏重要功能。增加數目的值將擷取所有的重要功能，但是有可能加入會實際造成模式錯誤的功能。
- **絕對錯誤比例的最小變更。**當絕對錯誤比例中的變更指示加入更多功能也無法進一步改善模式，則此演算法會停止。指定正數。降低最小變更的值會加入較多功能，但是有可能加入不會在模式中加入數值的功能。提高最小變更的值會排除較多功能，但是有可能遺漏對於模式很重要的功能。最小變更的 “最佳” 值需視您的資料和應用方式而定。請參閱輸出的「功能選擇錯誤記錄」，也協助您評估那些功能最重要。

區隔

圖表 20-6
最近鄰法分析區隔索引標籤



「區隔」索引標籤可供您將資料集區分為訓練集和保留集，並且在必要時，將觀察值指派給交叉驗證折疊

訓練和保留區隔。此組別指定將作用中資料集區隔成訓練和保留樣本的方法。**訓練樣本**由用來訓練最近鄰法模式的資料記錄所組成；在資料集中有某些比例的觀察值必須指定為訓練樣本以取得模式。**保留樣本**是獨立的資料記錄集，用來存取最終的模式；由於保留觀察值並未用來建立模式，保留樣本的錯誤為模式的預測能力提供「誠實」的估計。

- **將觀察值隨機指派給區隔。**指定指派給訓練樣本的觀察值百分比。剩餘的觀察值則指派給保留樣本。
- **使用變數指派觀察值。**指定一數值變數將作用中資料集的每一個觀察值指定給訓練或保留樣本。將變數中含有正值的觀察值指定給訓練樣本，將值為 0 或負值的觀察值指定給保留樣本。含有系統遺漏值的觀察值會從分析中排除。任何分割變數的使用者遺漏值永遠視為有效。

交叉驗證折疊。V 折疊交叉驗證可用來判斷相鄰的「最佳」數目。基於效能考量，這無法搭配功能選擇使用。

交叉驗證會將樣本分成次樣本數或折疊。然後會產生最近鄰法模式，並且從每個次樣本中排除資料。第一個模式是以第一個樣本折疊中以外的所有觀察值為基礎，第二個模式是以第二個範例折疊中以外的所有觀察值為基礎，依此類推。對於各個模式，都會將模式套用至在產生時被排除的次樣本，以評估錯誤。最近鄰法的「最佳」數目是在折疊產生最低錯誤的數目。

- **將觀察值隨機指派給折疊。**指定應該用於交叉驗證的折疊數目。此程序會將觀察值指派給折疊，數目介於 1 到 V（折疊的數目）。
- **使用變數指派觀察值。**指定一數值變數將作用中資料集的每一個觀察值指定給折疊。此變數必須是數值，其中的值必須介於 1 到 V。如果遺漏此範圍的任何值，而且，在任何分割上，如果有分割檔案有作用，則這會造成錯誤。

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

儲存

圖表 20-7
最近鄰法分析儲存索引標籤

已儲存變數的名稱。 自動名稱產生會確保您能保留所有的工作。自訂名稱可讓您捨棄/取代先前執行的結果，而不必先刪除在「資料編輯程式」中儲存的變數。

要儲存的變數

- **預測值或類別。**這會儲存尺度目標的預測值和類別目標的預測類別。
- **預測機率。**這會儲存類別目標的預測機率。對於前 n 個類別，系統會為每個類別儲存個別的變數，其中 n 是在「**儲存類別目標的目標上限**」控制項中指定。
- **訓練/保留區隔變數。**如果將觀察值隨機指派給「區隔」索引標籤上的訓練和保留樣本，則這會儲存已將觀察值指派至的區隔（訓練或保留）值。
- **交叉驗證折疊變數。**如果將觀察值隨機指派給「區隔」索引標籤上的交叉驗證折疊，則這會儲存已將觀察值指派至的折疊值。

輸出

圖表 20-8
最近鄰法分析輸出索引標籤

瀏覽器輸出 (C)

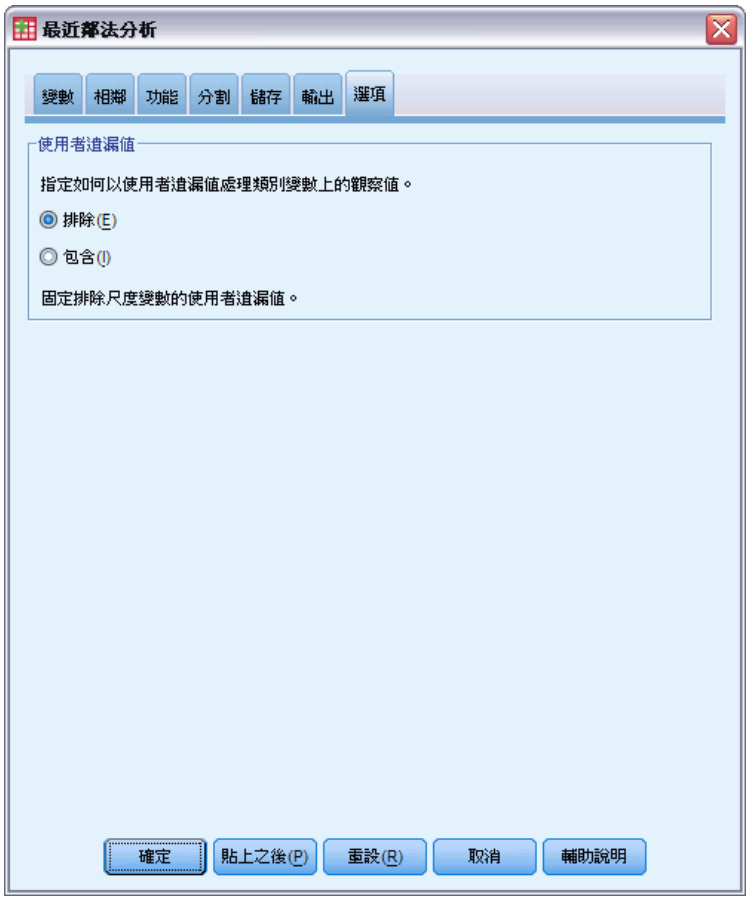
- **觀察值處理摘要。** 顯示觀察值處理摘要表，其中摘要出分析中包含和排除的觀察值數、總數，以及是依訓練和保留樣本包含和排除。
- **圖表和表格。**顯示模式相關輸出，包括表格和圖表。模式檢視中的表格包含焦點觀察值的 k 個最近鄰和距離、類別回應變數的分類，以及錯誤摘要。模式檢視的圖形輸出包含選擇錯誤記錄、功能重要性圖、功能空間圖、對等圖和象限地圖。

檔案

- **將模式匯出為 XML。**您可以使用這個模式檔案，將模式資訊套用到其他資料檔案中以進行評分工作。 如果已定義分割檔，則此選項無法使用。
- **匯出焦點觀察值和 k 個最近鄰之間相距的距離。**對於各個焦點觀察值，會針對各個焦點觀察值的 k 個最近鄰（從訓練樣本）和對應的 k 個最近距離建立個別變數。

選項

圖表 20-9
最近鄰法分析選項索引標籤

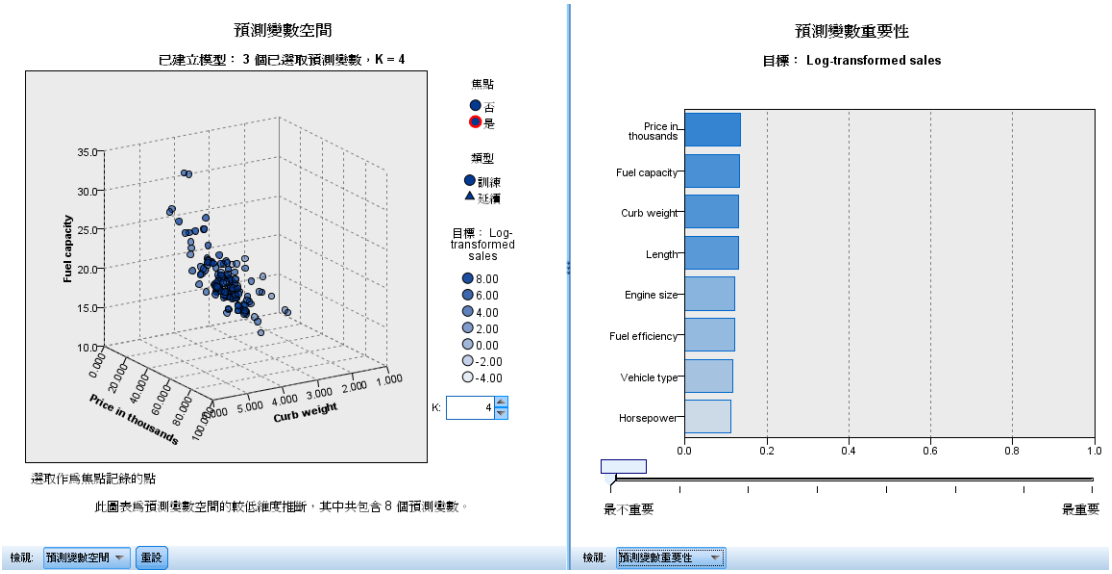


使用者遺漏值。類別變數必須具有要納入分析之觀察值的有效值。這些控制可讓您決定是否要在類別變數中，將使用者遺漏值視為有效值。

系統遺漏值和尺度變數的遺漏值可永遠視為無效。

模式檢視

圖表 20-10
最近鄰法分析模式檢視



在「輸出」索引標籤選擇「**圖表和表格**」時，此程序會在瀏覽器中建立「最近鄰法模式」物件。啟動（連按兩下）此物件，即可取得模式的互動檢視。模式檢視有 2 個面板：

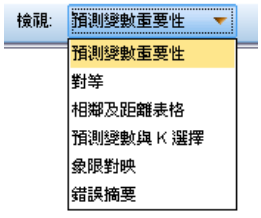
- 第一個面板會顯示模式的概述，稱為主要檢視。
- 第二個面板會顯示兩種檢視類型的其中一種：

  輔助模式檢視會顯示模式的詳細資訊，但是焦點不著重於模式本身。

  連結檢視會顯示使用者探索主要顯示各部分時模式功能的詳細資訊。

依照預設值，第一個面板會顯示功能空間，第二個面板顯示變數重要性圖表。如果沒有變數重要性圖表，也就是未在「功能」索引標籤選取「**按照重要性的權重功能**」時，會顯示「檢視」下拉式清單的第一個可用檢視。

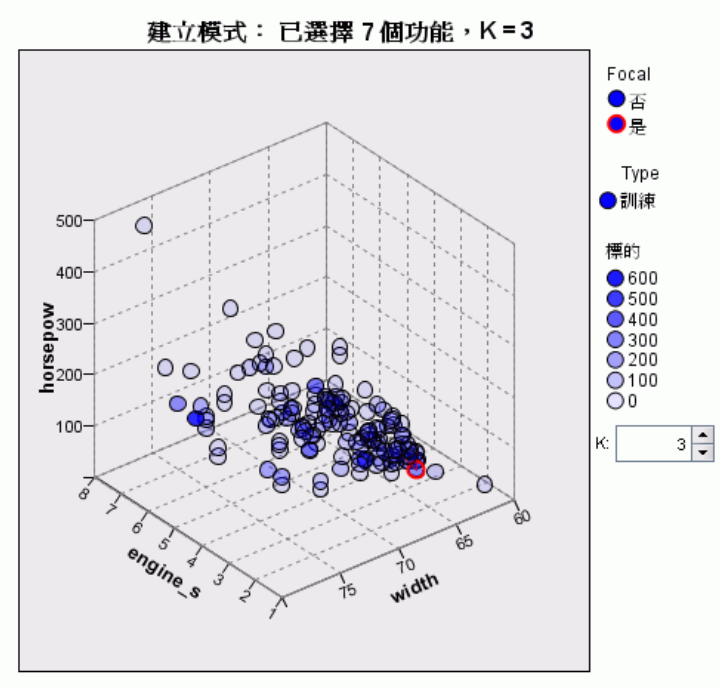
圖表 20-11
最近鄰法分析模式檢視下拉式清單



如果檢視沒有任何可用資訊，則「檢視」下拉式清單中的項目文字會停用。

功能空間

圖表 20-12
功能空間



功能空間圖是功能空間（如果有 3 個以上的功能，則為子空間）的互動式圖表。各個軸表示模式的功能，而圖表中的點位置顯示訓練和保留區隔中觀察值的功能值。

鍵。除了功能值之外，圖中的點也提供其他資訊。

- 形狀表示點所屬的訓練或保留區隔。
- 點的色彩/陰影表示該觀察值的目標值，其中不同色彩值等於類別目標的類別，而形狀表示連續目標的值範圍。訓練區隔的指示值是觀察值；對於保留區隔，這是預測值。如果未指定任何目標，則不會顯示任何鍵。
- 較粗的外框表示觀察值為焦點。顯示的焦點觀察值會連結本身的 k 個最近鄰。

控制和互動性。圖表中多個控制項可供您探索功能空間。

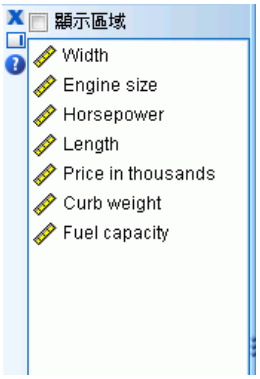
- 您可以選擇哪些功能子集顯示在圖表中，並變更在維度上表示哪些功能。
- “焦點觀察值” 只是「功能空間」圖表中所選取的點。如果您指定焦點觀察值變數，則會選取表示焦點觀察值的點。然而，如果您選取任何點，這個點都可暫時成為焦點觀察值。其中適用點選擇的 “一般” 控制；按一下點會選取該點，並取消選取其他所有點；按住 Ctrl 並按一下點會將該點加入所選取的多個點中。對等圖等連結檢視會根據「功能空間」中選取的觀察值自動更新。
- 您可以變更針對焦點觀察值顯示的最近鄰數目 (k)。
- 游標停留於圖表中的點時，會顯示觀察值標籤值的工具提示（如果未定義觀察值標籤，則顯示觀察值號碼），以及觀察和預測目標值。
- “重設” 按鈕可供您將「功能空間」回復為原有狀態。

新增與移除欄位/變數

您可以將新欄位/變數新增至功能空間，或移除目前顯示的欄位/變數。

「變數」色板

圖表 20-13
「變數」色板



「變數」色板必須顯示，您才能新增與移除變數。若要顯示「變數」色板，「模式瀏覽器」必須為「編輯」模式，在功能空間中必須已選取一個觀察值。

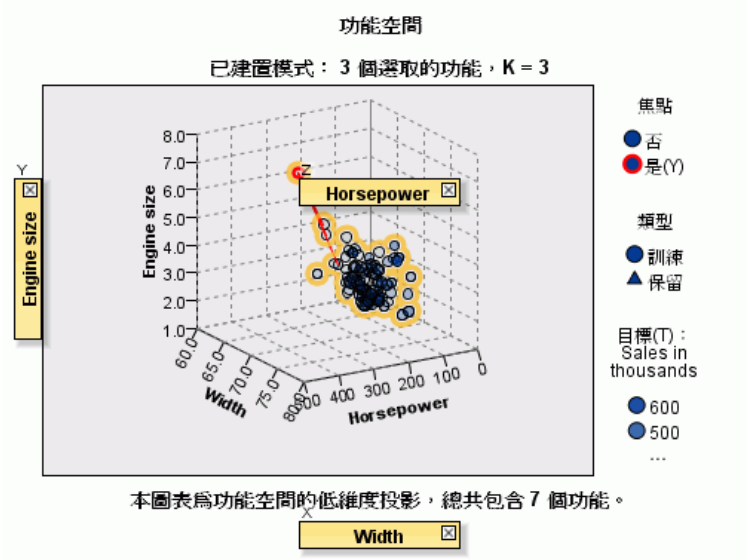
- 若要將「模式瀏覽器」設為「編輯」模式，請從功能表選擇：
  檢視 > 編輯模式
- 在進入「編輯」模式後，按一下功能空間中的任何觀察值。
- 若要顯示「變數」色板，從功能表選擇：
  檢視 > 調色板 > 變數

  「變數」色板會列出功能空間中所有的變數。變數名稱旁的圖示代表變數的測量水準。
- 若要暫時變更變數的測量水準，請在變數色板的變數上按一下滑鼠右鍵，並選擇一個選項。

變數區域

變數會新增至功能空間的「區域」中。若要顯示區域，請開始拖曳「變數」色板中的變數，或選取「**顯示區域**」。

圖表 20-14
變數區域



功能空間有 x、y 與 z 軸的區域。

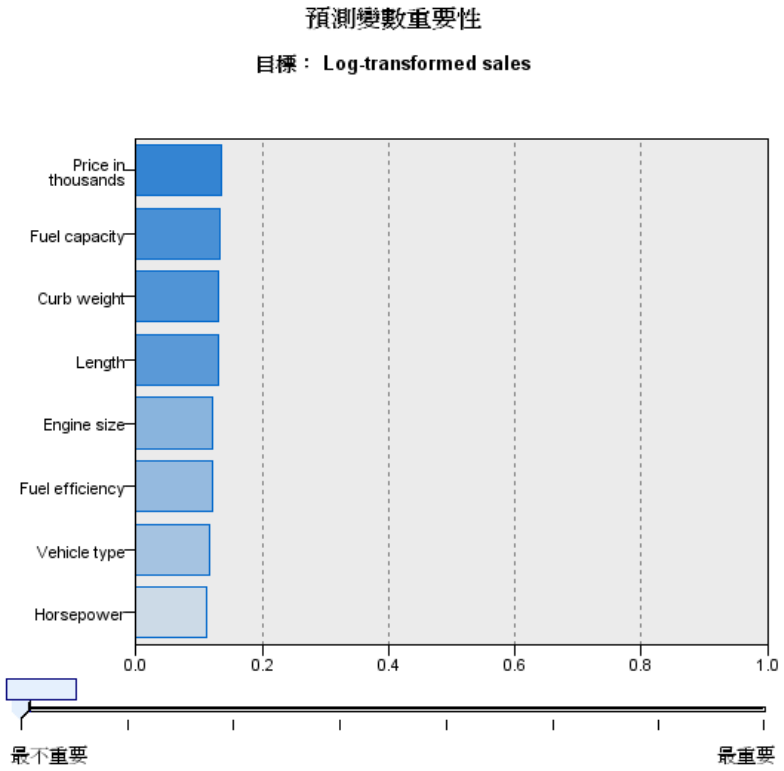
移動變數至區域

以下是將變數移動至區域的一般規則與秘訣：

- 若要將變數移至區域，請按一下「變數」色板中的變數，將變數拖曳至區域中，然後放開變數。如果您選擇「**顯示區域**」，也可以在區域上按一下滑鼠右鍵，並選取您要新增至區域中的變數。
- 如果您將變數從「變數」色板拖曳至一個已有其他變數的區域，則新變數會取代舊變數。
- 如果您將變數從一個區域拖曳另一個已有其他變數的區域，則變數會交換位置。
- 按一下區域中的 X 會將變數從該區域中移除。
- 如果視覺化中有多個圖形元素，則每個圖形元素會有自己的相關聯變數區域。先選取所需的圖形元素。

變數重要性

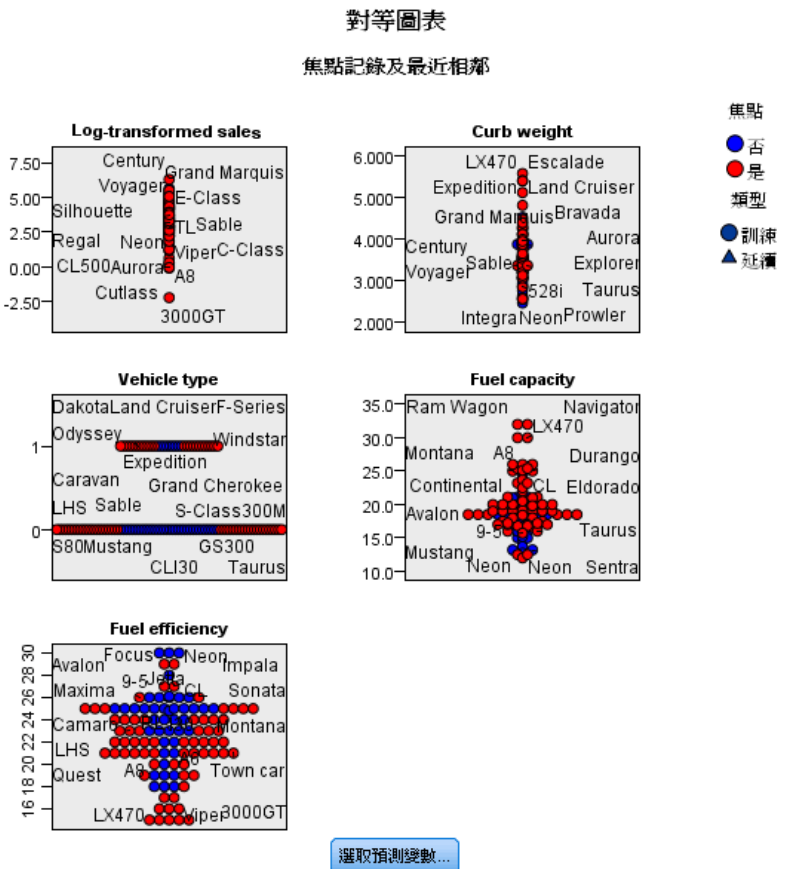
圖表 20-15
變數重要性



一般而言，您會想要將焦點著重在模式中最重要的變數，並考慮捨棄或忽略最不重要的變數。變數重要性圖可協助您指出評估模式時各變數的相對重要性，以達成此目標。由於其中的值都是相對值，因此顯示中所有值的值總合為 1.0。變數重要性與模式準確性無關。這只涉及進行預測時各變數的重要性，而不涉及預測是否正確。

對等

圖表 20-16
對等圖



此圖顯示各功能與目標的焦點觀察值及其 k 個最近鄰。如果在「功能空間」中選擇焦點觀察值，則可使用此圖表。

連結行為。「對等圖」連結「功能空間」的方式有兩種。

- 「功能空間」中選擇的觀察值（焦點）會顯示在「對等圖」中，其中包含觀察值的 k 個最近鄰。
- 「功能空間」中選擇的 k 值會用於「對等圖」中。

最近鄰距離

圖表 20-17
最近鄰距離

k 最近相鄰及距離

針對起始焦點記錄顯示

132 的第 1 到第 100 個資料列

焦點記錄	最近相鄰				最近距離			
	1	2	3	4	1	2	3	4
RL	Catera	Aurora	DeVille	Eldorado	0.021	0.038	0.040	0.044
A6	C70	Diamante	S-Type	Regal	0.034	0.038	0.040	0.055
A8	LS400	S-Class	CL500	SL-Class	0.062	0.065	0.068	0.132

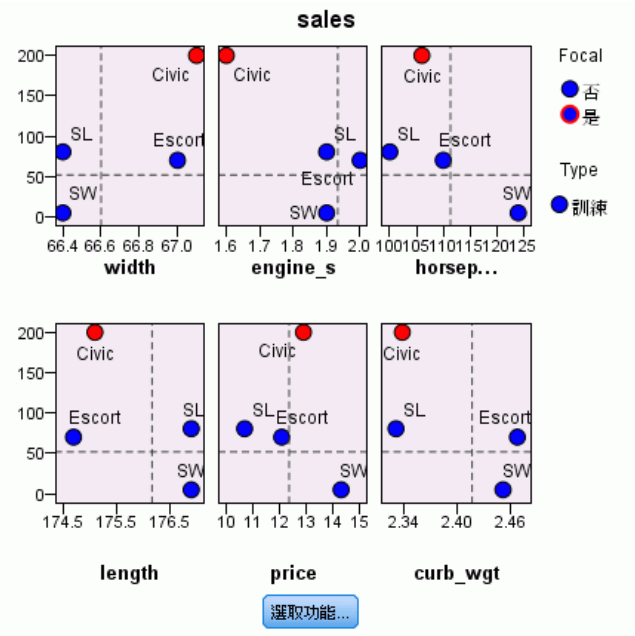
此表只顯示焦點觀察值的 k 個最近鄰和距離。如果在「變數」標籤指定焦點觀察值識別碼，則可使用此表，表中只會顯示此變數識別的焦點觀察值。

各列的：

- **焦點觀察值**欄含有焦點觀察值的觀察值標記變數值；如果未定義觀察值標記，則此欄包含焦點觀察值的觀察值數目。
- 「最近鄰」群組下的第 i 欄含有焦點觀察值第 i 個最近鄰的觀察值標記變數值；如果未定義觀察值標記，則此欄包含焦點觀察值第 i 個最近鄰的觀察值號碼。
- 「最近距離」群組下的第 i 欄包含第 i 個最近鄰與焦點觀察值相距的距離

象限地圖

圖表 20-18
象限地圖

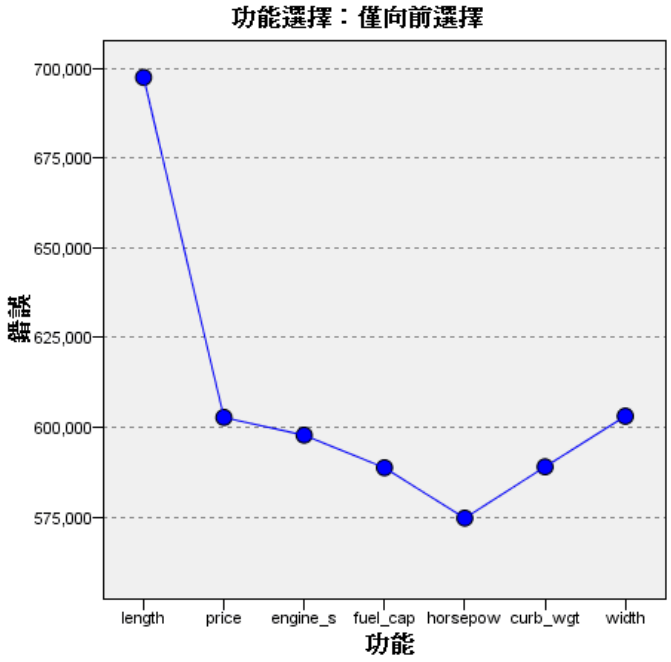


此圖表顯示散佈圖（或點狀圖，視目標的測量水準而定）上的焦點觀察值及其 k 個最近鄰，其中的目標位於 y 軸，尺度功能位於 x 軸，並且按照功能標示。如果有目標，而且在「功能空間」中選取焦點觀察值，則可使用此圖表。

- 對於連續變數會在訓練區隔的變數平均數繪製參考線。

功能選擇錯誤記錄

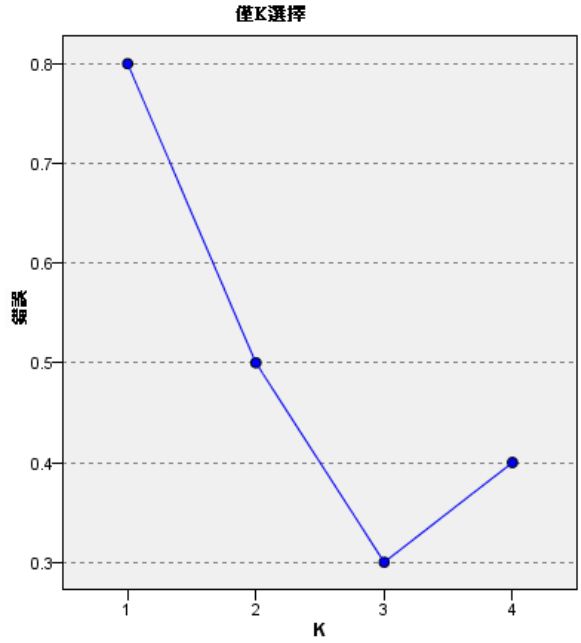
圖表 20-19
功能選擇



圖表上的點會顯示模式 y 軸上的錯誤（錯誤率或平方和錯誤，視目標的測量水準而定），其中的功能列於 x 軸（包含 x 軸左方的所有功能）。如果目標和功能選擇有效，則可使用此圖表。

k 選擇錯誤記錄

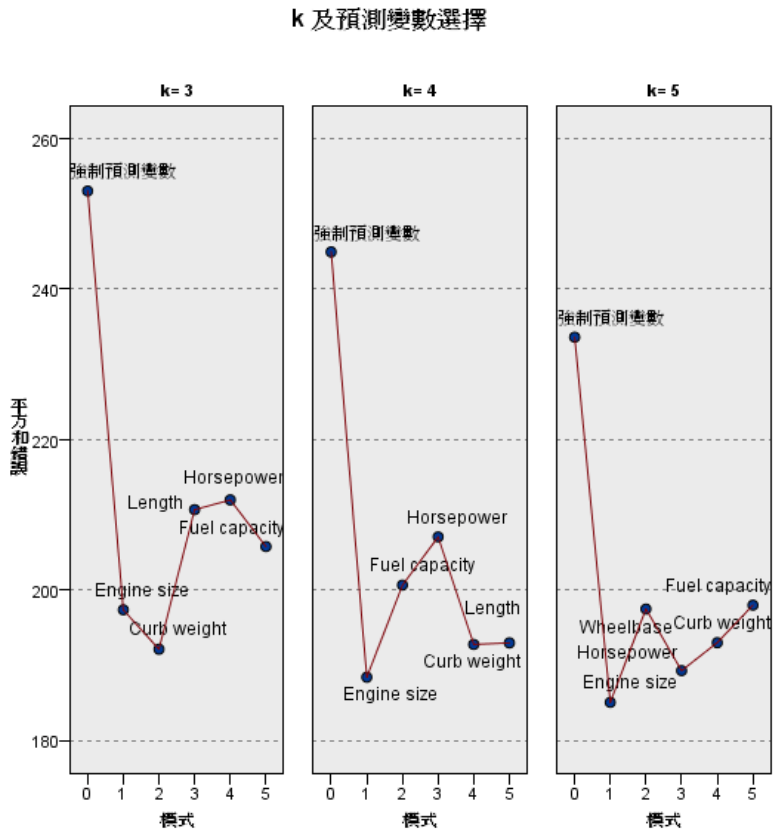
圖表 20-20
k 選擇



圖表上的點會顯示模式 y 軸上的錯誤 (錯誤率或平方和錯誤，視目標的測量水準而定)，其中包含 x 軸上的最近鄰數目 (k)。如果目標和 k 選擇有效，則可使用此圖表。

k 和功能選擇錯誤記錄

圖表 20-21
k 和功能選擇



這些是功能選擇圖（請參閱 功能選擇錯誤記錄 第 134 頁），並且按照 k 標示。如果有目標以及 k 和功能選擇，而且兩者都有效，則可使用此圖表。

分類表

圖表 20-22
分類表

分割		預測		
		0	1	校正百分比
訓練	0	111	1	99.11%
	1	7	33	82.50%
	總體百分比	77.64%	22.37%	94.74%

此表按照區隔顯示目標觀察值與預測值的交叉分類。如果有目標，而且屬於類別，則可使用此表。

- 「保留」區隔的（遺漏）列包含具有目標遺漏值的保留觀察值。這些觀察值會構成保留樣本：整體百分比，而非正確百分比值。

錯誤摘要

圖表 20-23
錯誤摘要

分割	校正百分比
訓練	622043

如果有目標變數，則可使用此表。此表會顯示模式相關的錯誤，以及類別目標的連續目標與錯誤率（100% – 正確的整體百分比）平方和。

章節
21

判別分析

判別分析會建立組別成員的預測模式。此模式包括判別函數，是以預測變數的線性組合為基礎（如果有兩個組別以上，將產生一組判別），而該預測變數，必須能提供組別之間最佳判斷依據。上述函數從觀察值樣本中產生，而且觀察值的組別成員是已知的，這些函數隨後便可以套用到新的觀察值上，新觀察值會有預測變數的測量，但是，組別成員是未知的。

注意：分組變數可以有兩個（或以上）的數值。但是，分組變數的代碼必須為整數，而且必須指定其最小值和最大值。觀察值的數值如果超出這個範圍，就不會分析它。

範例。一般而言，溫帶國家的人，每天消耗的卡路里會比熱帶的人多，而且溫帶地區中，住在都市的人口比例也會比較高。研究人員想將這些資訊，併成一個函數，以便判斷受訪者對這兩個國家的人民，能細分到什麼樣的程度。研究人員認為，人口數量和經濟資訊，應該也相當重要。因此，使用判別分析，能讓您估計線性判別函數的係數，判別函數的運算式，看起來跟多重線性迴歸方程式的右側內容很像。亦即是，它也使用 a、b、c 和 d 係數，函數如下：

D = a * 氣候 + b * 都市 + c * 人口 + d * 每人國民生產毛額

如果這些變數，有助於判別兩種不同的氣候區域，那麼溫帶國家和熱帶國家的 D 值就會不同。如果您使用逐步的變數選取法，可能會發現此函數中，不需要包含四個變數。

統計量。對於每個變數來說，統計量包括：平均數、標準差、單變量 ANOVA。對每一個分析而言：則有：Box’s M、組內相關矩陣、組內共變異數矩陣、各組共變異數矩陣、全體觀察值的共變異數矩陣。對於每種典型判別函數而言，則有：特徵值、變異數百分比、典型相關、Wilks’ Lambda 值、卡方。對每一個步驟：事前機率、Fisher 函數係數、未標準化函數係數、每個典型函數的 Wilks’ Lambda 值。

資料。分組變數的類別必須很明確，而且其類別數也受到限制，它的類別必須以整數編碼。名義式的自變數，必須重新編碼成虛擬或對比變數。

假設。觀察值應該是獨立的。預測變數應該呈多變量常態分配，而各組別間的組內變異數－－共變異數矩陣，應該都是相等的。各組別成員是假設為互相排斥的（也就是說，每個觀察值只屬於一個組別），而且整體無遺漏（也就是說，所有觀察值都是某個組別的成員）。當組別成員為真的類別變數時，此程序將可發揮最高效能；但如果組別成員是以連續變數的值為基礎（例如高 IQ 對低 IQ），則請考慮使用線性迴歸，以便利用連續變數本身所提供的豐富資訊。

若要取得判別分析

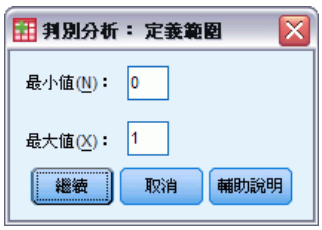
- 從功能表選擇：
  分析(A) > 分類 > 判別...

圖表 21-1
判別分析對話方塊

- 選取整數值的「分組變數」，再按一下**定義範圍**，指定需要處理的類別。
- 選取自變數或預測變數。(如果您的分組變數不是整數值，「轉換」功能表上的「自動重新編碼」將可建立變數的分組變數)。
- 選取輸入自變數的方法。
  - **一起輸入自變數。** 同時輸入所有達到允差條件的自變數。
  - **使用逐步迴歸分析法。**使用逐步迴歸分析法控制變數的輸入和刪除。
- 此外，選取具有選擇變數的觀察值。

判別分析的定義範圍

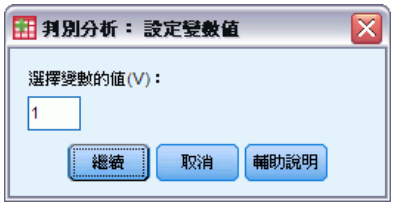
圖表 21-2
判別分析定義範圍對話方塊



請指定分析所用之分組變數的最小值和最大值。超過這個範圍的觀察值不會用於判別分析，但是會依照分析結果分類至其中一個現有組別。最小值和最大值必須是整數。

判別分析的選取觀察值

圖表 21-3
判別分析設定數值對話方塊



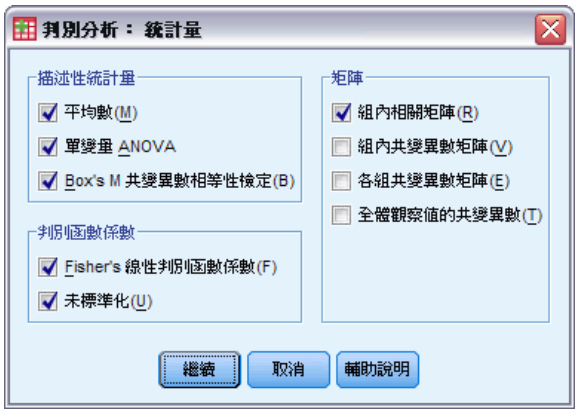
若要選取分析的觀察值：

- 在「判別分析」對話方塊中，選取一個選擇變數。
- 按一下「**數值**」，以便輸入整數作為選擇數值。

觀察值必須具有選擇變數的指定值，才能用來推導判別函數。選取和未選取的觀察值會產生統計量和分類結果。這個程序可讓您依照之前舊資料來分類新的觀察值，或將資料分類為訓練和測試子集，以驗證所產生的模式。

判別分析的統計量

圖表 21-4
判別分析統計量對話方塊



描述性統計量。 可用的選項包括：平均數（包括標準差）、單變量 ANOVA、Box’s M 檢定。

- **平均數。** 顯示總和與組別平均數，以及自變數的標準差。
- **單變量 ANOVA。** 針對每個自變數的組別平均數之相等性執行單因子變異數分析檢定。
- **Box’s M。** 組別共變異數矩陣的相等性檢定。若樣本夠大，則非顯著 p 值代表沒有充足證據顯示矩陣有所不同。此檢定可感應到多變量常態的偏差。

函數係數。可用的選項包括：Fisher 分類係數、未標準化係數。

- **Fisher’s。** 顯示可直接供分類使用的 Fisher’s 分類函數係數。每個組別會取得個別的分類函數係數集，且會針對擁有最高判別分數的組別指定觀察值（分類函數值）。
- **未標準化。** 顯示未標準化的判別函數係數。

矩陣。可用的自變數係數矩陣包括：組內相關矩陣、組內共變異數矩陣、各組共變異數矩陣、總和共變異數矩陣。

- **組內相關矩陣。** 顯示合併的組別相關值矩陣，此矩陣是將所有組別的個別共變異數矩陣平均，再計算相關值而來。
- **組內共變異數具矩陣。** 顯示合併的組別內共變異數矩陣，此矩陣可能與總共變異數矩陣不同。您可將所有組別的個別共變異數矩陣平均來取得矩陣。
- **各組共變異數矩陣。** 顯示各組不同的共變異數矩陣。
- **全體觀察值的共變異數。** 顯示所有觀察值中像是來自單一樣本的共變異數矩陣。

判別分析逐步迴歸分析法

圖表 21-5
判別分析逐步迴歸分析法對話方塊

方法。選取統計量，以輸入或刪除新變數。您可以使用的項目包括：Wilks’ lambda 值、無法解釋的變異數、Mahalanobis 距離、最小 F 比例和 Rao’s V 量數。利用 Rao’s V 量數，您可以指定要輸入的變數 V 最小增量。

- **Wilks’ lambda 值。** 一種進行逐步迴歸分析判別分析的「變數選取方法」，它會根據變數將 Wilks’ lambda 值降低的程度來為輸入等式的項目選擇變數。每一個步驟都要輸入將整體 Wilks’ lambda 值最小化的變數。
- **無法解釋的變異數。** 在每個步驟中輸入會將組別間未說明的變異數總和最小化的變數。
- **Mahalanobis 距離。** 測量自變數之觀察值數值與整體觀察值平均數的變異程度。若 Mahalanobis 距離較大，則會將觀察值識別為在一個以上自變數中具有極端值。
- **最小 F 比例。** 一種逐步分析的變數選取方法，其所根據的作法，是將從組別間之 Mahalanobis 距離計算而來的 F 比例最大化。
- **Rao’s V 量數。** 組別平均數之間差異的量數。也稱為 Lawley-Hotelling 跡。每一個步驟，都要輸入將 Rao’s V 中的增加最大化的變數。選取此選項後，請輸入變數必須輸入分析的最小值。

條件。您可以使用的項目包括：「使用 F 值」和「使用 F 機率值」。請輸入用於輸入和刪除變數的數值。

- **使用 F 值。** 當變數的 F 值大於「輸入」值時，系統會將該變數輸入模式，而當 F 值小於「移除」值時，系統會將變數移除。「輸入」必須大於「移除」，而且兩個數值都必須是正數。若要將更多變數輸入模式，請調低「輸入」值。若要從模式中移除更多變數，請調高「移除」值。
- **使用 F 機率值。** 當變數之 F 值顯著水準小於「輸入」值時，系統會將變數輸入模式，而當顯著水準大於「移除」值時，系統會將變數移除。「輸入」必須小於「移除」，而且兩個數值都必須是正數。若要將更多變數輸入模式，請調高「輸入」值。若要從模式中移除更多變數，請調低「移除」值。

顯示。步驟摘要可顯示每個步驟執行之後，所有變數的統計量；而顯示配對組 F 值則可以顯示每對組別配對的 F 比例矩陣。

判別分析分類

圖表 21-6
判別分析分類對話方塊

事前機率。 此選項會決定分類係數是否調整，以用於組別成員的演繹式知識。

- **所有組別皆同。** 相同的事前機率的假設適用於所有組別；這不會影響係數。
- **從組別大小計算。** 樣本的觀察組別大小會決定組別成員的事前機率。例如，若包括在分析中 50% 的觀察值落入第一個組別，25% 為第二組，25% 為第三組，分類係數便會調整來增加第一組別中與其他二組相關成員的概似。

顯示。可用的顯示選項包括：逐觀察值的結果、摘要表、Leave-one-out 分類方法。

- **逐觀察值的結果 (E)。** 顯示每個觀察值之實際組別、預測組別、事後機率和判別分數的編碼。
- **摘要表 (U)。** 指定至以判別分析為基礎之每個組別的觀察值個數（包括正確與不正確）。有時亦稱作「混淆性矩陣」。
- **Leave-one-out 分類方法 (V)。** 分析中的每個觀察值皆由該觀察值除外之所有觀察值的衍生函數來分類。其亦稱作「U 方法」。

用平均數置換遺漏值。選取這個選項，可將遺漏值替換成自變數的平均數，但是只限於分類階段。

使用共變異數矩陣。您可以選擇使用組內共變異數矩陣、或各組共變異數矩陣，來將觀察值分類。

- **組內。** 合併的組別內共變異數矩陣可用於分類觀察值。
- **各組。** 各組共變異數矩陣在分類時使用。因為分類是以判別函數為依據（而不是以原始變數為依據），因此這個選項並不是永遠相當於二次判別。

圖形。可用的圖形選項包括：組合組別、各組、地域圖。

- **合併組散佈圖(O)。** 建立前兩個判別函數數值的所有組別散佈圖。若僅有一個函數，則會改為顯示直方圖。
- **各組散佈圖(S)。** 建立前兩個判別函數數值的各組散佈圖。在只有一個函數的情況下，則會改為顯示直方圖。
- **地域圖(T)。** 以函數值為基礎，用來將觀察值分類至組別的邊界圖。將觀察值分類至對應之組別的數目。每個組別的平均數是用其邊界中的星號來表示。如果只有一個判別函數，就不會顯示地圖。

判別分析的儲存

圖表 21-7
判別分析對話方塊

您可以在作用資料檔中加入新的變數。可供您使用的選項包括：預測的組群成員（單一變數）、判別分數（各判別函數在解中的變數）、已知判別分數的組別成員機率（各組的一個變數）。

您也可以選擇是否將模式資訊匯出至指定的 XML (PMML) 格式檔案。您可以使用這個模式檔案，將模式資訊套用到其他資料檔案中以進行評分工作。

DISCRIMINANT 指令的其他功能

指令語法語言也可以讓您：

- 執行多個判別分析（僅使用一個指令），以及控制變數輸入的次序（使用 ANALYSIS 次指令）。
- 指定分類的事前機率（使用 PRIORS 次指令）。
- 顯示轉軸後樣式和結構矩陣（使用 ROTATE 次指令）。

- 限制萃取的判別函數數量（使用 FUNCTIONS 次指令）。
- 限制欲分析的選定（或未選定）觀察值類別（使用 SELECT 次指令）。
- 讀取和分析相關矩陣（使用 MATRIX 次指令）。
- 寫入相關矩陣以備日後分析之用（使用 MATRIX 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
22

因子分析

因子分析會試著找出基本變數（或「**因子**」），因為這些變數（或因子）可以說明觀察變數組中，所存在的相關型態。因子分析常用於資料縮減，以找出少量的因子來說明大量的明顯變數中所觀察到的變動。您也可以使用因子分析，產生跟經常性機制相關的假設，或者篩檢變數，以便日後分析（例如，在線性迴歸分析之前，使用因子分析找出共線性）。

因子分析程序的彈性很大，因為它：

- 可以使用七種因子萃取法。
- 可以使用五種旋轉法，包括直接斜交、和非正交旋轉的 Promax。
- 有三種方法來計算因子分數，而且分數可以存成變數以供進一步分析。

範例。受訪者要抱持哪些信念，才會願意接受政治性的問卷調查？檢驗調查項目之間的相關性後發現，各種不同項目次組別之間，出現明顯的重疊現象 — 稅務的相關問題多半彼此關聯、軍事的議題也是彼此相關等。利用因子分析，您可以對基本因子的個數進行研究，並且很多時候，您都可以找出因子所代表的理念。此外，您可以計算每個受訪者的因子分數，然後於後續的分析中，使用此分數。例如，您可能會建立一個 Logistic 迴歸模式，以便根據因子分數預測投票行為。

統計量。 對於每個變數來說，統計量包括：有效觀察值個數、平均數、和標準差。對於每個因子分析來說，共有：變數的相關矩陣，包括顯著水準、行列式、反矩陣；重製的相關矩陣，包括逆映像矩陣；未轉軸之統計量（共同性、特徵值和說明的變異數百分比）；取樣恰當性的 Kaiser-Meyer-Olkin 測量、和 Bartlett 球形檢定；對於未轉軸的解而言，包括：因子負荷、共同性、和特徵值；對於轉軸後的解，則包括：轉軸後樣式矩陣、和轉換矩陣。對於斜交轉軸而言：轉軸後樣式和結構矩陣；因子分數係數矩陣、和因子共變異數矩陣。圖形(T)：對於圖形來說，則有：特徵值的陡坡圖、前兩個（或三個）因子的負荷圖。

資料。在「**區間**」或「**比例**」水準的變數，應該為數值變數。類別資料（如宗教或祖國）不適合做因子分析。如果資料可以用來計算 Pearson 相關係數的話，則可用於因子分析。

假設。資料中每對變數，應該都具雙變數常態分配，而且觀察值應該互不相關。因子分析模式中，會指定由共同因子（由模式所估計的因子），或者由唯一因子（在觀察變數之間不重疊的因子）來決定變數；估計值的假設基礎為：所有唯一因子彼此不相關，跟共同因子也無關。

若要取得因子分析

▶ 從功能表選擇：
分析(A) > 維度縮減(D) > 因子...

▶ 選取因子分析的變數。

圖表 22-1
「因子分析」對話方塊

因子分析選擇觀察值

圖表 22-2
「因子分析設定變數值」對話方塊

若要選擇分析的觀察值：

- 選取一個選擇變數。
- 按一下「**數值**」，以便輸入整數作為選擇數值。

觀察值必須擁有「選擇變數」值，才能用來進行因子分析。

因子分析描述性統計量

圖表 22-3
「因子分析描述性統計量」對話方塊

統計量。 「單一變量描述性統計量」包括平均數、標準差、每個變數的有效觀察值個數。「未轉軸之統計量」會顯示初始的共同性、特徵值、已知變異數百分比。

相關矩陣。可使用的選項包括：係數、顯著水準、行列式、KMO 和 Bartlett 的球形檢定、反矩陣、重製，和逆映像。

- **KMO 和 Bartlett 的球形檢定。** 檢定變數間偏相關是否較小的 Kaiser-Meyer-Olkin 取樣恰當性量數。Bartlett’s 球形檢定會檢定相關矩陣是否為識別矩陣，其會指出因子模式是否不當。
- **重製的(R)。** 因子解中的估計相關矩陣。此矩陣也會顯示殘差（估計相關和觀察相關之間的差異）。
- **反映像(A)。** 逆映像相關矩陣包含負數的偏相關係數，而逆映像共變異數矩陣則包含負數的偏共變異數。在較佳的因子模式中，大多數位於對角線外的元素皆較小。在逆映像相關矩陣的對角線上，會顯示變數取樣恰當性的量數。

因子分析萃取

圖表 22-4
「因子分析萃取」對話方塊

方法。允許您指定因子萃取法。可使用的方法包括：主成份、未加權最小平方、概化最小平方、最大概似、主軸因子、alpha 因素萃取法、映像因素萃取法。

- **主成份分析。** 一種萃取因子的方式，用於構成觀察變數的無關線性組合。第一個成份的變異數最大。後續的成份說明變異數的比例逐漸減少，而且互不相關。使用主成份分析可取得未轉軸之統計量。當相關矩陣奇異時，可使用此方法。
- **未加權最小平方法。** 將觀察值和重製相關矩陣（忽略對角線）之間差異的平方和最小化的因子萃取法。
- **概化最小平方法。** 將觀察值和重製相關矩陣之間差異的平方和最小化的因子萃取法。系統會使用相關性本身的反向唯一性來予以加權，因此擁有高唯一性的變數即會獲得較少的加權，而擁有低唯一性的變數則會獲得較多的加權。
- **最大概似法。** 一種因子萃取法，若樣本是來自多變量常態分配，則此方法會產生最可能具有觀察相關矩陣的參數估計值。此外，會藉由變數的反向唯一性來加權相關，且會使用疊代演算法。
- **主軸因子。** 此方法使用在對角線的複相關係數平方值，作為共同性的起始估計值，從原始相關矩陣中萃取因子。這些因子負荷量用於估計新的共同性，以替換在對角線的舊共同性估計值。疊代會一直繼續，直到從上一個疊代到下一個疊代當中，共同性變更可滿足萃取因子的收斂準則為止。
- **Alpha。** 一種因子萃取法，其在進行分析時，會從潛在變數範圍中考量要作為樣本的變數。此方法會將 alpha 信度因子最大化。
- **映像因素萃取法。** 由 Guttman 根據映像理論發展出的一種因子萃取法。變數的常見部分稱作偏映像，且會定義為本身在其他變數中的線性迴歸，而非假設因子函數。

分析。允許您指定相關矩陣、或共同變異數矩陣。

- **相關矩陣。**如果您的分析用不同的尺度法來測量變數，就很有用。
- **共變異數矩陣。**如果要將因子分析套用到多個群組，而且群組中每個變數的變異數各不相同，就很有用。

萃取。因子的特徵值如果超過指定值，則可以將它們全部保留起來，或者您可以保留指定個數。

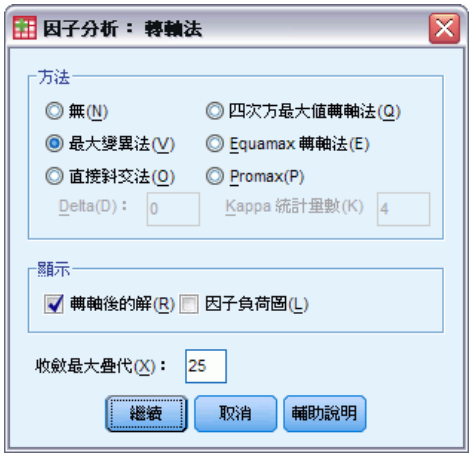
顯示。允許您要求未轉軸前的因子統計量、以及特徵值的陡坡圖。

- **未轉軸前的因子統計量。** 顯示未轉軸前因子負荷量（因子樣式矩陣）、共同性，以及因子統計量的特徵值。
- **陡坡圖(S)。** 與每個因子有關的變異數圖。此圖用於決定應保留多少因子。一般情況下，此圖會顯示大型因子的陡坡間之限定分段，以及（陡坡）其餘部分的漸性尾隨。

收斂最大疊代。可讓您指定在評估解時，演算法可執行的步驟數上限。

因子分析旋轉

圖表 22-5
「因子分析旋轉」對話方塊



方法。讓您選取因子旋轉的方法。可使用的方法包括：最大變異法、相等最大法、四方最大法、直接斜交、或 Promax 旋轉法。

- **最大變異法。** 會將每個因子中具有高負荷量之變數個數最小化的正交旋轉法。這種方法可以簡化因子的解讀。
- **直接斜交旋轉法。** 一種斜交（非正交）旋轉法。當 delta 等於 0（預設值），則大部分會是斜交解。若 delta 越趨近負數，則因子就越不會趨向斜交。若要覆寫 0 的預設 delta，請輸入小於或等於 0.8 的數字。
- **四方最大法。** 此旋轉法可將需要解釋每個變數的因子個數減到最少。這種方法可以簡化對觀察變數的解讀。
- **相等最大法。** 一種結合最大變異法與四方最大法的旋轉法（前者會簡化因子，而後者會簡化變數）。此方法會將因子中高度載入的變數個數，以及變數說明所需因子的個數予以最小化。
- **Promax 旋轉。** 一種斜交轉軸，可讓因子間產生相關。此旋轉法比直接斜交旋轉法計算得更快，所以對大型資料集有幫助。

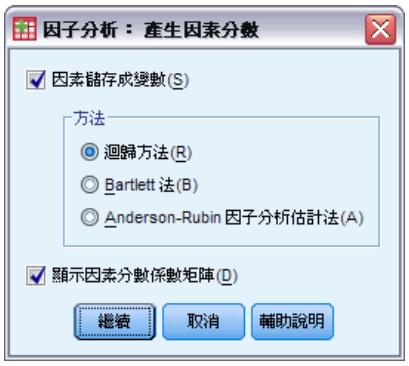
顯示。讓轉軸後的解可以包含：輸出，第一、二（或第三個）因子的負荷圖。

- **轉軸後的解。** 您可以選取旋轉法，以取得轉軸後的解。若是正交旋轉，則會顯示轉軸後的樣式矩陣和因子轉換矩陣。若是斜交旋轉，則會顯示樣式、結構，以及因子相關矩陣。
- **因子負荷量圖。** 前三個因子的三維因子負荷量圖。若為二因子統計量，則會顯示二維圖表。若僅萃取一個因子，則不會顯示圖表。若要求旋轉，則圖表會顯示轉軸後的統計量。

收斂最大疊代。可讓您指定在執行旋轉時，演算法可執行的步驟數上限。

因子分析分數

圖表 22-6
「因子分析因子數」對話方塊



因子儲存成變數。這個選項會替最後解中的每個因子，分別建立一個新變數。

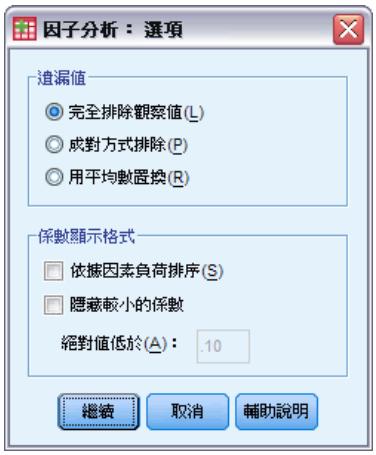
方法。其他計算因子分數的方法為迴歸、Bartlett 和 Anderson-Rubin。

- **迴歸方法。** 估計因子分數係數的方法。產生出來的分數包括 0 平均數以及一個變異數，此變異數等於因子分數估計值和實際因子數值之間的複相關平方值。即使因子為正交，此分數仍可能具有相關性。
- **Bartlett 分數。** 一種因子分數係數的估計方法。產生的分數會具有 0 平均數。若唯一因子的平方和超出變數範圍，則會予以最小化。
- **Anderson-Rubin 方法。** 一種估計因子分數係數的方法；一種經過修改的 Bartlett 方法，其可確保已估計因子的正交性。產生出來的分數包括 0 平均數，以及 1 標準差，且彼此不相關。

顯示因素分數係數矩陣。這個選項會顯示一個數值（變數跟該值相乘，以取得因子分數），同時也會顯示因子分數之間的關聯。

因子分析選項

圖表 22-7
「因子分析選項」對話方塊



遺漏值。 讓您指定處理遺漏值的方式。可使用的選擇包括：「**完全**」排除遺漏值、「**成對**」方式排除、或以平均數置換。

係數顯示格式。讓您控制輸出矩陣的外觀。您可以按照大小，將係數排序，如果係數的絕對值小於指定值，則不列出來。

FACTOR 指令的其他功能

指令語法語言也可以讓您：

- 在萃取和旋轉期間，指定疊代的收斂準則。
- 指定個別的旋轉因子圖形。
- 指定所要儲存之因子分數的數量。
- 指定主軸因子分析法的對角線值。
- 將相關矩陣或因子負荷矩陣寫入磁碟供之後分析使用。
- 讀取和分析相關矩陣或因子負荷矩陣。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
23

選擇集群程序

您可以使用「TwoStep 分析」、「階層集群分析」或「K 平均數集群分析」程序來執行集群分析。每個程序會使用不同的演算法來建立集群，並且各程序有其他兩個沒有的選項。

TwoStep 集群分析。大部分的應用程式適用於「TwoStep 集群分析」程序這個方法。此程序提供下列獨特的功能：

- 除了選擇集群模式的測量方法以外，還可自動選取最佳的集群數。
- 以類別和連續變數為基礎來同時建立集群模式的能力。
- 可將集群模式儲存成外部 XML 檔，之後再使用更新的資料來讀取該檔案並更新集群模式。

此外，「TwoStep 集群分析」程序還可以分析大型的資料檔。

階層集群分析。「階層集群分析」程序限用於小型的資料檔（要集群的個體只有數百個），但具有下列獨特的功能：

- 能夠集群觀察值或變數。
- 能夠計算某個範圍的可能解，並儲存每個解的集群組員。
- 提供數種方式以執行集群構成、變數轉換，並測量集群間的相異性。

只要所有變數的類型相同，「階層集群分析」程序便可以分析區間（連續）、個數或二元變數。

K 平均數集群分析。「K 平均數集群分析」限用於連續資料，並且需要事先指定集群數，但具有下列獨特的功能：

- 能夠儲存每個個體與集群中心點間的距離。
- 能夠讀取各集群初始的中心，並將最後的集群中心存成外部 IBM® SPSS® Statistics 檔。

此外，「K 平均數集群分析」程序還可以分析大型的資料檔。

章節
24

TwoStep 集群分析

「TwoStep 集群分析」程序是設計用來顯示資料集中自然分組（或集群）的探索工具（原本不會加以顯示）。此程序所採用的演算法有數種理想的功能，這使它們與傳統的集群技術因而有所區分。

- **處理類別和連續變數。** 藉由假設變數為自變數，結合的多項式-常態分配就可以放置在類別和連續變數上。
- **自動選擇集群數目。** 藉由比較不同集群解之間的模型-選項準則的值，此程序可自動決定最適集群數目。
- **擴展性** 藉由建構可摘要記錄的集群功能（CF）樹狀結構，TwoStep 演算法可讓您分析大型資料檔。

範例。 零售商和消費者產品公司通常會將集群技術，套用到描述它們客戶之消費習慣、性別、年齡、收入水準等的資料。這些公司對每個消費者組別量身訂做行銷和產品跚

圖表 24-1
「TwoStep 集群分析」對話方塊

距離量數。 此選項可決定兩個集群間計算的相似琵度。

- **對數概似。** 概似量數會對變數進行機率分配。連續變數假設為常態分配，而類別變數則假設為多項式分配。所有變數皆假設為自變數。
- **歐基里得。** 歐基里得量數是兩個集群間的「直線」距離。它只能在所有變數皆為連續時使用。

集群數目。 此選項可讓您指定集群數目決定的方式。

- **自動決定。** 此程序將使用指定於「集群準則」組別中的準則，自動決定「最適」集群數目。您也可以選擇性輸入一個正整數，來指定此程序應考慮的集群數目最大值。
- **指定固定數。** 可讓您固定解中的集群數目。輸入正整數。

連續變數的個數。 此組別可提供「選項」對話方塊中制定的連續變數標準化規格的摘要。

集群準則。 此選項可決定自動集群演算法決定集群數目的方式。您可以指定「Bayesian 資訊準則」(Bayesian Information Criterion，BIC)，或指定「Akaike 資訊準則」(Akaike Information Criterion，AIC)。

資料。 此程序可用在連續變數及類別變數上。觀察值代表要集群的物件，變數則代表集群所依據的屬性。

觀察值順序。 請注意，集群樹狀結構和最終解可能會依觀察值的順序而定。若要將順序效應降到最低，請以隨機方式排列觀察值。您可以取得一些觀察值之不同的解答，這些觀察值會以不同的隨機順序排序，以確認給定解答的穩定性。在因為檔案過大而顯得困難的情況下，可以用執行多個使用不同隨機順序排序的觀察值範例來取代。

假設。 概似距離量數假設集群模式中的變數，皆為自變數。再者，每個連續變數都假設具有常態 (Gaussian) 分配，每個類別變數都假設具有多項式分配。經驗內部檢定指出此程序很少受到獨立性假設及分配假設偏差的影響，但是您應該要嘗試注意這些假設符合的程序。

使用「雙變數相關分析」程序來檢定兩個連續變數的獨立性。Use the Crosstabs procedure to test the independence of two categorical variables. 使用「平均數」程序來檢定連續變數和類別變數間的獨立性。 使用「預檢資料」程序來檢定連續變數的常態性。使用「卡方檢定」程序來檢視類別變數是否有指定的多項式分配。

若要取得 TwoStep 集群分析

▶ 從功能表選擇：
分析(A) > 分類 > TwoStep 集群...

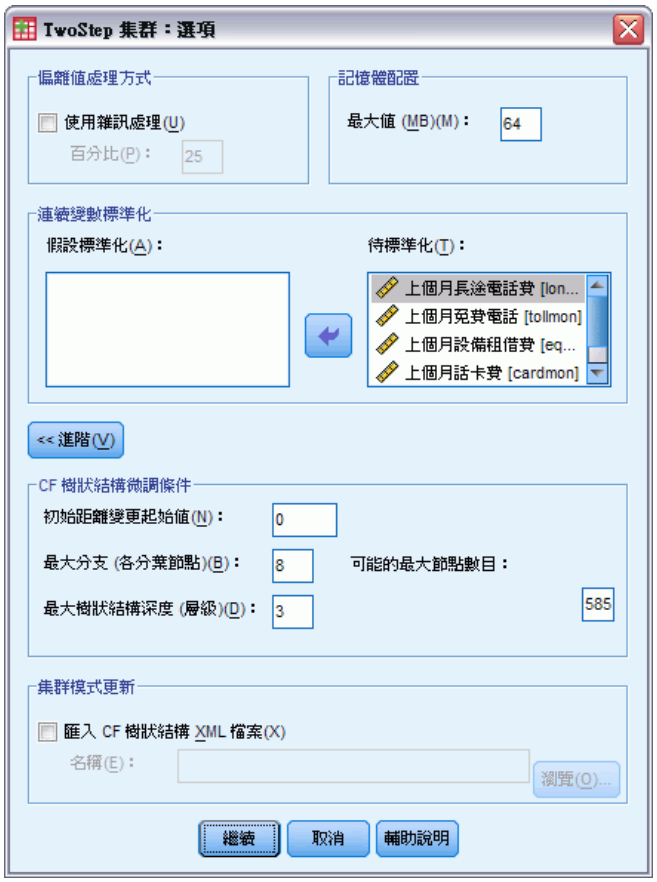
▶ 選取一或多個類別或連續變數。

您可以：

- 調整建構集群所依據的準則。
- 選取噪音處理、記憶體配置、變數標準化和集群模式輸入的設定。
- 要求模式瀏覽器輸出。
- 將模式結果儲存到工作檔或外部的 XML 檔。

「TwoStep 集群分析選項」

圖表 24-2
「TwoStep 集群選項」對話方塊



偏離值處理。 此組別可讓您處理偏離值，特別是在集群過程中集群功能 (CF) 樹狀結構填滿時。CF 樹狀結構的分葉節點如果無法再接受任何觀察值，且沒有可分割的分葉節點時，就是填滿的狀態。

- 如果您選取噪音處理且 CF 樹狀結構為填滿狀態，分葉的中觀察值在放置到「噪音」分葉後會重新成長。如果分葉包含少於最大分葉大小指定的觀察值百分比，此種分葉就屬於稀疏。在樹狀結構重新成長之後，偏離值將會放到 CF 樹狀結構中（如果可能的話）。如果沒有的話，偏離值會被捨棄。
- 如果您沒有選取噪音處理而 CF 樹狀結構為填滿狀況，其在將變更門檻值變大時會重新成長。在最終集群之後，無法指定到集群的值就會標記為偏離值。偏離值集群的識別碼為 -1，且不算在集群數目的個數中。

記憶體配置。 此組別可讓您指定記憶體的最大容量（以千位元組 MB 為單位），這也是集群演算法應該使用的。如果程序超過此最大值，它將會使用磁碟來儲存不適合存在記憶體的資訊。請指定一個大於或等於 4 的數值。

- 有關您可在系統中指定的最大值，請諮詢您的系統管理員。
- 如果此值過低，演算法可能會找不到正確或需要的集群數目。

變數標準化。 集群演算法可以處理標準化連續變數。任何未標準化的連續變數都應保留為在「待標準化」清單中的變數。為了節省時間和便於計算，您可以選取任何已標準化為「假設已標準化」清單中之變數的連續變數。

進階選項

CF 樹狀結構調整準則。 以下集群演算法設定專門套用在集群功能 (CF) 樹狀結構，並且應小心地加以變更：

- **起始距離變更門檻。** 這是 CF 樹狀結構成長所用的起始門檻。如果將一個觀察值插入 CF 樹狀結構的分葉，其所產生的緊密程度小於門檻標準，因此分葉不會分割。而如果緊密的程度超過門檻，分葉就會分割。
- **最大分支 (每個分葉節點)。** 分葉節點所能擁有的子節點最大數目。
- **最大樹狀結構深度。** CF 樹狀結構所能擁有的階層最大數目。
- **可能節點的最大數目。**這可指出程序有可能產生的 CF 樹狀結構節點最大數目，所依據的函數是 $(b^{d+1} - 1) / (b - 1)$，此處的 b 是最大分支，而 d 則是最大樹狀結構深度。請注意，過大的 CF 樹狀結構可能會消耗系統資源，這樣反而會影響程序的效能。每個節點至少需要 16 個位元組。

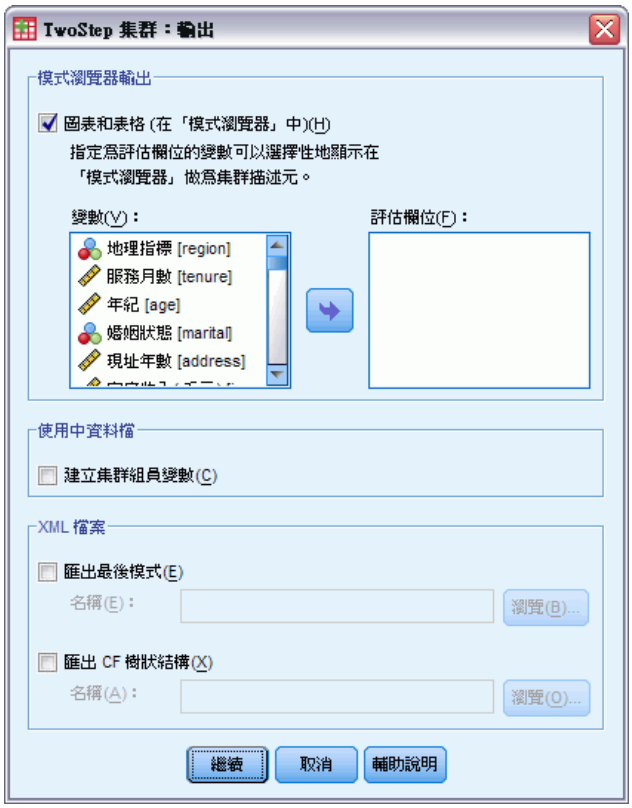
集群模式升級。 此組別可讓您匯入並更新事前分析所產生的集群模式。輸入檔包含 XML 格式的 CF 樹狀結構。此模式將會使用作用中檔案中的資料來進行更新。您必須在主對話方塊中，以在事前分析中所指定的次序來選取變數名稱。除非您特別寫入新模式資訊到同樣的檔名中，否則 XML 檔維持不變。

如果您指定集群模式更新，會使用適合產生為原始模式所指定的 CF 樹狀結構之選項。更特別的是，還會使用已儲存模式的距離量數、噪音處理、記憶體配置或 CF 樹狀結構調整準則設定，而對話方塊中這些選項的任何設定均予以忽略。

注意：執行集群模式更新時，程序會假設未使用作用中資料集中的任何選定觀察值來建立原始集群模式。程序也會假設模式更新中使用的觀察值，和用來建立原始模式的觀察值是來自相同的母群；也就是說，會假設連續變數的平均數及變異數和類別變數的水準在觀察值集之間是都一致的。如果您的「新」觀察值集和「舊」觀察值集來自異質的母群，您應該對此組合的觀察值集執行「TwoStep 集群分析」程序，才能得到最好的結果。

TwoStep 集群分析輸出

圖表 24-3
「TwoStep 集群輸出」對話方塊



模式瀏覽器輸出。 此組別提供顯示集群結果的選項。

- **圖表和表格。**顯示模式相關輸出，包括表格和圖表。模式檢視中的表格包含模式摘要與集群對功能網格。模式檢視中的圖形輸出包含集群品質圖表、集群大小、變數重要性、集群比較網格以及儲存格資訊。
- **評估欄位。** 這會計算未用於集群建立之變數的集群資料。在「顯示」子對話方塊中選取評估欄位和輸入功能，便可在模式瀏覽器中同時顯示這兩項資訊。會忽略有遺漏值的欄位。

使用中資料檔。此組別可讓您將變數儲存到作用中資料集。

- **建立各集群組員變數。** 此變數包含各觀察值的集群識別碼。此變數的名稱是 tsc_n，此處的 n 是一個正整數，可指出在某個指定作業階段中，由此程序完成的作用中資料集儲存作業之次序。

XML 檔案。 最終集群模式和 CF 樹狀結構是兩種可以 XML 格式匯出的輸出檔。

- **匯出最終模式。**最終集群模式會匯出到指定的 XML (PMML) 格式檔案中。您可以使用這個模式檔案，將模式資訊套用到其他資料檔案中以進行評分工作。
- **匯出 CF 樹狀結構。** 此選項可讓您儲存集群樹狀結構的目前狀態，並在稍後使用更新的資料加以更新。

集群瀏覽器

集群模式通常用於根據檢驗的變數尋找類似記錄的群組（或集群），其中相同群組中成員之間的相似性高，不同群組中成員之間的相似性低。尋找的結果可用於找出不明顯的關聯性。例如，透過集群分析客戶喜好、收入水準與購買習慣，可找出較可能回應特定市場行銷活動的客戶類型。

有兩種方法可解讀集群顯示的結果：

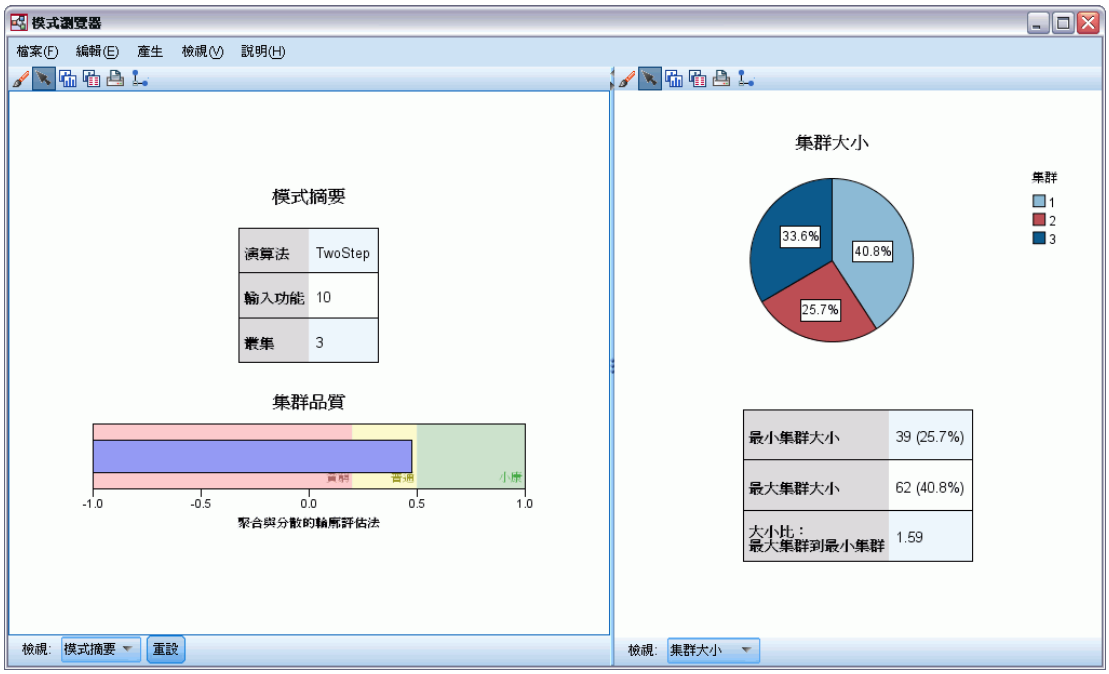
- 檢驗集群，以判定該集群獨一無二的特性。一個集群是否包含所有高收入借款人？此集群包含的記錄是否比其他集群多？
- 檢驗各集群的欄位，以判定數值在集群之間如何分佈。一個人的教育程度是否決定了在集群中的成員資格？高信用評分是否區分不同集群中的成員資格？

您可以在「集群瀏覽器」中使用主檢視與各種連結的檢視來深入瞭解，以協助您回答這些問題。

若要檢視集群模式的相關資訊，請在「瀏覽器」中啟動（連按兩下）「模式瀏覽器」物件。

集群瀏覽器

圖表 24-4
「集群瀏覽器」的預設顯示



「集群瀏覽器」由兩個面板組成，主檢視位於左邊，連結或輔助檢視位於右邊。主檢視有兩種：

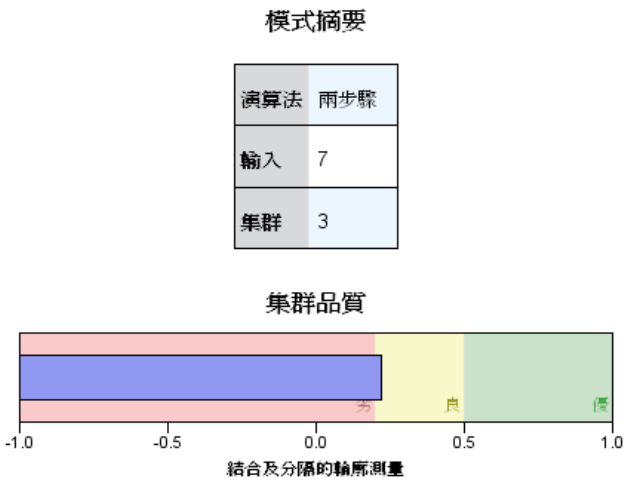
- 模式摘要（預設值）。
- 集群。

連結/輔助檢視有四種：

- 預測值重要性。
- 集群大小（預設值）。
- 儲存格分配。
- 集群比較。

模式摘要檢視

圖表 24-5
主面板中的「模式摘要」檢視



「模式摘要」檢視會顯示集群模式的快照，也就是摘要，包括集群結合與分組的 Silhouette 量數，此量數會以不同顏色來表示差、可或佳的結果。此快照可讓您快速檢查品質是否為差，如果為差，您可以決定回到模式節點修正集群模式設定，以產生更好的結果。

差、可與佳的結果是依據 Kaufman 與 Rousseeuw (1990) 關於集群結構解讀的著作而決定。在「模式摘要」檢視中，結果「佳」則表示到達 Kaufman 與 Rousseeuw 的資料為集群結構合理或強力證據等級，結果「可」則表示到達薄弱證據等級，結果「差」則表示到達非顯著證明的等級。

所有記錄的 Silhouette 量數平均為 (B–A) / max(A,B)，其中 A 為記錄距離其集群中心的距離，B 為距離最近的非其所屬集群中心的距離。Silhouette 係數 1 是指所有的觀察值直接位於其集群中心。–1 的值表示所有觀察值位於另一個集群的集群中心。平均而言，平均數值為 0 表示觀察值距離其集群中心與另一個最近集群中心的距離是相等的。

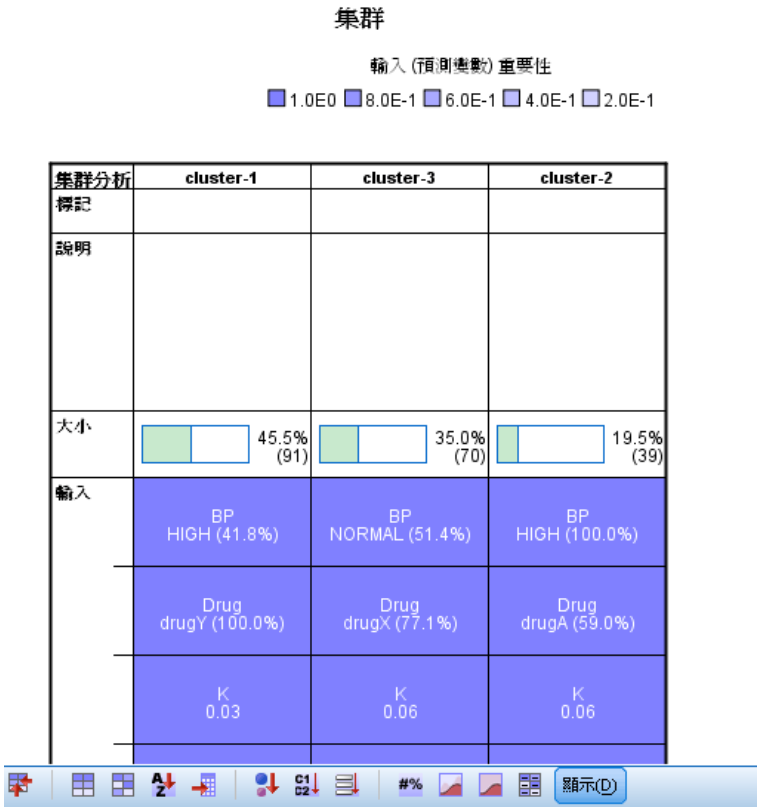
摘要會提供一張表格，其中包含下列資訊：

- **演算法。** 所使用的集群演算法，例如 “TwoStep”。

- **輸入功能。** 欄位數目，也就是所謂的**輸入**或**預測值**。
- **集群。** 解中的集群數目。

集群檢視

圖表 24-6
主面板中的「集群中心」檢視



「集群」檢視包含集群對特徵網格中，網格包含集群名稱、大小和每個集群的分析概要。

網格中的行包含下列資訊：

- **集群。** 演算法建立的集群數目。
- **標記。** 任何套用到每個集群上的標記（預設為空白）。在儲存格內部連按兩下以輸入可說明集群內容的標記，例如 Luxury 汽車買主。
- **說明。** 任何集群內容的說明（預設為空白）。在儲存格內連按兩下以輸入集群的說明，例如「55 歲以上，專業人員，收入超過 100,000 美元」。
- **大小。** 每個集群大小，以佔整體集群樣本的百分比表示。網格內每個大小儲存格會顯示一個垂直列，會顯示在集群內的大小百分比、以數值格示表示的大小百分比，以及集群觀察值個數。
- **功能。** 個別的輸入值或預測值，依照預設會按照整體重要性排序。如果有任何行的大小相同，則會以集群數目的遞增排序順序顯示。

整體特徵重要性會以儲存格背景顏色來表示，最重要的特徵顏色最深，最不重要的特徵則沒有顏色。上表的網格指出每個特徵儲存格色彩的重要性。

將滑鼠停留在儲存格上方，會顯示特徵的完整名稱/標記與儲存格的重要性值。是否顯示進一步的資訊，取決於檢視與特徵的類型。在「集群中心」檢視中，這包含儲存格統計量與儲存格值：“平均數：4.32”。若是類別特徵，則儲存格會顯示最多（典型）的類別及其百分比。

在「集群」檢視內，您可以選取各種方式來顯示集群資訊：

- 轉置集群與特徵。
- 排序特徵。
- 排序集群。
- 選取儲存格內容。

轉置集群與特徵。

依照預設值，集群會顯示為行，特徵會顯示為列。若要將此顯示方式顛倒，請按一下「**排序特徵依據**」按鈕左側的「**轉置集群與特徵**」按鈕。例如，當您有許多集群要顯示時，可能需要這麼做，如此不需要太多水平捲動就能看到資料。

圖表 24-7
主面板中已轉置的集群

聚类	标签	说明	大小	
cluster-1			45.0% (91)	BP HIGH (41.8%)
cluster-3			35.0% (70)	BP NORMAL (51.4%)
cluster-2			19.0% (39)	BP HIGH (100.0%)

排序特徵

「**排序特徵依據**」按鈕可讓您選取顯示特徵儲存格的方式：

- **整體重要性。** 這是預設的排序順序。特徵是按照整體重要性遞減排序，各集群之間的排序順序皆相同。如果有任何特徵的重要性值同分，則會以特徵名稱的遞增排序順序列出同分的特徵。
- **集群內重要性。** 系統會根據特徵對每個集群的重要性來排序特徵。如果有任何特徵的重要性值同分，則會以特徵名稱的遞增排序順序列出同分的特徵。選擇此選項後，排序順序通常會隨著集群改變。
- **名稱。** 特徵是依照名稱的字母順序排序。
- **資料順序。** 特徵是按照其資料集順序排序。

排序集群

依照預設，集群會依照大小的遞減順序排序。「**排序集群依據**」按鈕可讓您按照名稱字母順序排序集群，如果您已為集群建立唯一的標記，則改為以標記字母順序排序。

標記相同的特徵會依照集群名稱排序。如果集群是按照標記排序的，當您編輯了集群的標記後，排序順序會自動更新。

儲存格內容

「**儲存格**」按鈕可讓您變更特徵的儲存格內容與評估欄位的顯示方式。

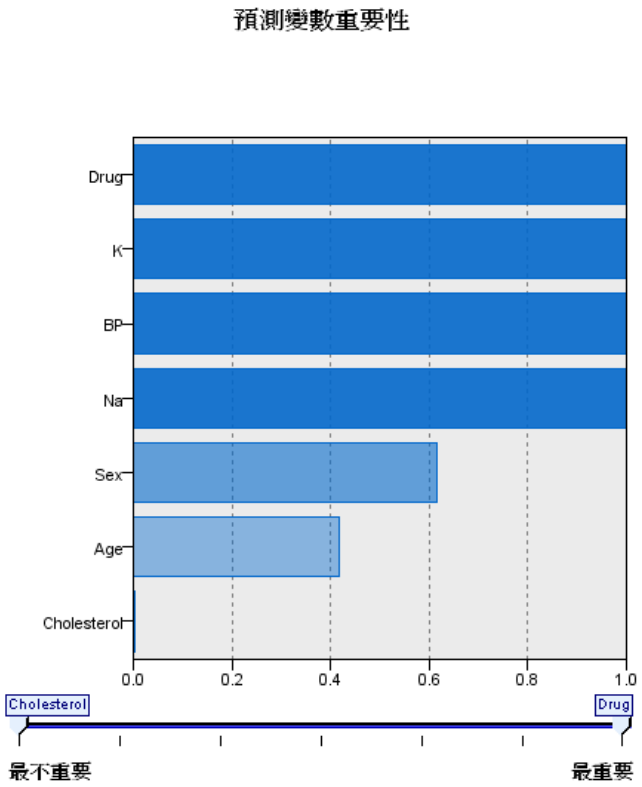
- **集群中心。** 依照預設值，儲存格會顯示特徵名稱/標記，以及每個集群/特徵組合的集中趨勢。系統會針對連續欄位與眾數（最常出現的類別）顯示平均數，以及類別欄位的欄表百分比。
- **絕對分配。** 顯示每個集群內的特徵名稱/標記，以及特徵的絕對分配。若是類別特徵，顯示畫面會出現與類別重疊的長條圖，這些類別按照資料值的遞增順序排序。若是連續特徵，顯示畫面會顯示平滑密度圖，此圖會為每個集群使用相同的端點與間隔。

  著上實心紅色的顯示畫面會顯示集群分配，較淡色的顯示畫面則表示整體資料。
- **相對分配。** 在儲存格中顯示特徵名稱/標記與相對分配。一般而言，顯示畫面與顯示絕對分配的分畫面類似，不同處為顯示的是相對分配。

  著上實心紅色的顯示畫面會顯示集群分配，較淡色的顯示畫面則表示整體資料。
- **基本檢視。** 當集群很多時，如果沒有捲動畫面，很難看見所有的詳細資訊。若要減少捲動的次數，請選取此檢視將顯示畫面變更為表格的精簡版。

集群預測值重要性檢視

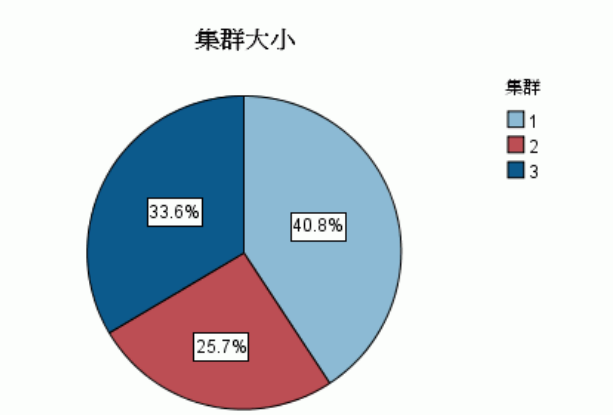
圖表 24-8
連結面板中的「集群預測值重要性」檢視



「預測值重要性」檢視會顯示每個欄位在估計模式時的相對重要性。

集群大小檢視

圖表 24-9
連結面板中的「集群大小」檢視



最小集群大小	39 (25.7%)
最大集群大小	62 (40.8%)
大小比： 最大集群到最小集群	1.59

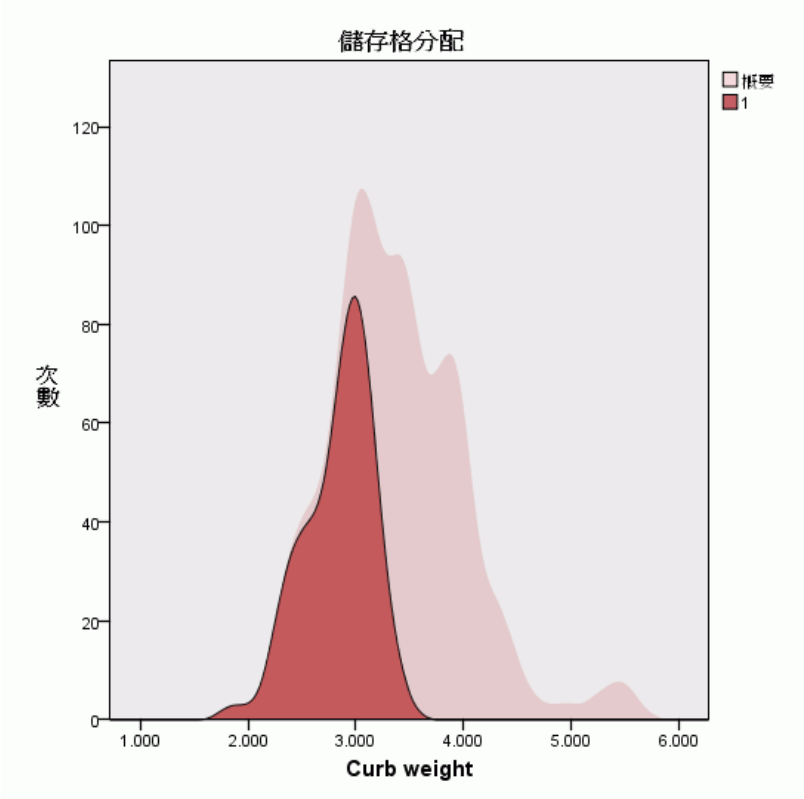
「集群大小」檢視會顯示圓餅圖，其中包含每個集群。每個圖塊上會顯示每個集群的百分比大小，將滑鼠停在每個圖塊上方，圖塊中會顯示個數。

在圖表下方的表格會列出下列大小資訊：

- 最小集群大小（個數與佔整體的百分比）。
- 最大集群大小（個數與佔整體的百分比）。
- 最大集群對最小集群的大小比例。

儲存格分配檢視

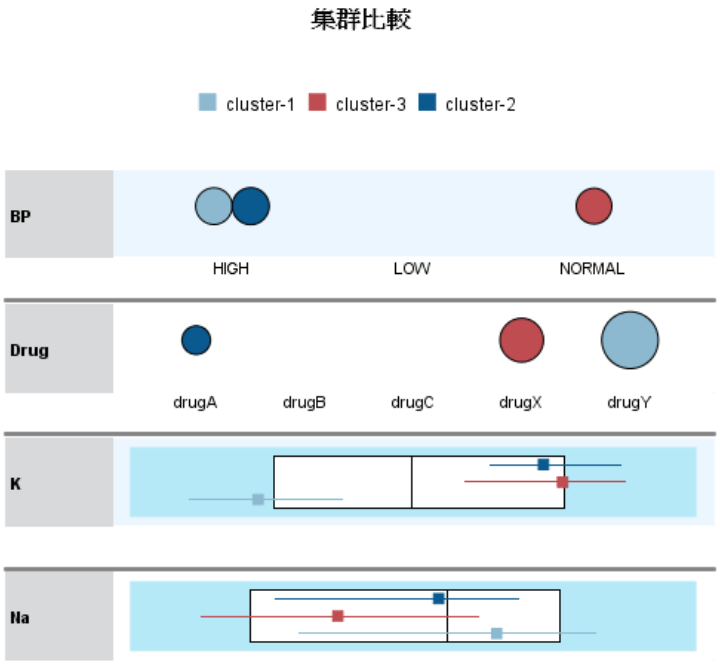
圖表 24-10
連結面板中的「儲存格分配」檢視



「儲存格分配」檢視會為您在「集群」主面板的表格中選取的任何特徵儲存格，以展開的方式顯示更詳細的資料分配圖。

集群比較檢視

圖表 24-11
連結面板中的「集群比較」檢視



「集群比較」檢視包含網格模式配置，此配置會在列中顯示特徵，在行中顯示選取的集群。此檢視可幫助您更瞭解構成集群的因子，它也可以讓您看見集群之間的差異，這些差異不只是集群與整體資料的比較，也會有集群與集群間的比較。

若要選取要顯示的集群，請按一下「集群」主面板上集群行的頂端。使用 Ctrl-按一下或 Shift-按一下來選取或取消選取多個要進行比較的集群。

注意：您最多可以選擇顯示五個集群。

系統會以選取集群的順序來顯示集群，欄位順序則是由「**排序特徵依據**」選項所決定。選取「**集群內重要性**」時，一律依據整體重要性排序欄位。

背景圖會顯示每個特徵的整體分配：

- 類別特徵會顯示為點形圖，點的大小代表每個集群最多/最常用的類別（按照特徵）。
- 連續特徵會顯示為盒形圖，圖中會顯示整體中位數與四分數範圍。

所選取集群的盒形圖會在這些背景檢視上重疊：

- 若是連續特徵，方形點標記與水平線會表示每個集群的中位數與四分數範圍。
- 每個集群都以不同的色彩表示，並在檢視頂端顯示。

瀏覽集群瀏覽器

「集群瀏覽器」是互動式的顯示畫面。您可以：

- 選取欄位或集群可檢視更多詳細資訊。
- 比較集群以選取感興趣的項目。
- 改變顯示畫面。
- 轉置軸。

使用工具列

您可以使用工具列選項來控制在左面板與右面板中顯示的資訊。您可以使用工具列控制項來變更顯示畫面的方向（上到下、左到右或右到左）。此外，您也可以將瀏覽器重設為預設設定，並開啟對話方塊來指定主面板中「集群」檢視的內容。

圖表 24-12
「集群瀏覽器」中可控制資料的工具列

「排序特徵依據」、「排序集群依據」、「儲存格」與「顯示」選項只有在您選取了主面板中的「集群」檢視後才能使用。

	請參閱轉置集群與特徵。第 161 頁
	請參閱排序特徵依據。第 161 頁
	請參閱排序集群依據。第 162 頁
	請參閱儲存格。第 162 頁

控制集群檢視顯示畫面

若要控制主面板「集群」檢視中顯示的內容，請按一下「顯示」按鈕；「顯示」對話方塊隨即開啟。

圖表 24-13
集群瀏覽器 - 「顯示」選項

功能。 預設為選取。若要隱藏所有輸入特徵，請取消選取此核取方塊。

評估欄位。 選擇要顯示的評估欄位（建立集群模式時未使用，但已傳送到模式瀏覽器以評估集群的欄位）；預設不會顯示任何欄位。注意：如果沒有任何的評估欄位，則無法使用此核取方塊。

集群說明。 預設為選取。若要隱藏所有集群說明儲存格，請取消選取此核取方塊。

集群大小。 預設為選取。若要隱藏所有集群大小儲存格，請取消選取此核取方塊。

最大類別個數。 指定在類別特徵的圖表中顯示的類別數目上限，預設值為 20。

過濾記錄

圖表 24-14
集群瀏覽器 - 過濾觀察值

如果您想要進一步瞭解特定集群或集群群組中的觀察值，您可以依據選擇的集群來選取要進一步分析的記錄子集。

- 選取「集群瀏覽器」其「集群」檢視中的集群。若要選取多個集群，可以按住 Ctrl 鍵來選取。
- 從功能表選擇：
  產生 > 過濾記錄...

- 輸入過濾變數名稱。所選集群中的記錄會收到此欄位的數值 1。所有其他記錄會得到數值 0，而且這些記錄會在後續的分析中排除，直到您變更過濾狀態為止。
- 按一下「**確定**」。

章節
25

階層集群分析法

這個程序會根據您所選取的特性，試圖找出具有相對同質性的觀察值（或變數）組別。它所使用的演算法，會從個別集群中的每一個觀察值（或變數）開始，然後再與集群組合，直到只剩下一個為止。您可以分析原始資料，或從各種不同的標準化轉換中選擇。「相似性」程序會產生距離或相似性測量。每個階段都會顯示統計量，以協助您選出最適用的數值。

範例。要如何找出哪些電視節目可吸引群組中類似的觀眾？您可以使用階層集群分析，並根據觀眾特性將電視節目（觀察值）分成同質的組別，結果可以用來區隔市場。或者您可以將城市（觀察值）分成同質群組，以便從中選出可以比較的城市，並檢驗不同的市場策略。

統計量。包括：群數凝聚過程、距離（或相似性）矩陣、單一解（或解的範圍）的各集群組員。圖形(T)：樹狀圖和冰柱圖。

資料。變數可以是數量、二元或個數資料。變數的尺度是重點所在，因為尺度的差異可能會影響您的集群解。如果您的變數尺度差異極大（例如，一個變數以元為單位來測量，另一個以年為單位），那麼您應該考慮把它們標準化（您可以使用「階層集群分析」程序來自動執行）。

觀察值順序。 如果輸入資料中有同分距離或相似性，或者合併期間在更新的集群中發生，則產生的集群解會依據檔案中觀察值的次序而定。您可以取得一些觀察值之不同的解答，這些觀察值會以不同的隨機順序排序，以確認給定解答的穩定性。

假設。所用的距離或相似性測量，應該要適合所分析的資料（如需有關選擇距離和相似性測量的資訊，請參閱「相似性」程序）。此外，您應該將所有相關的變數納入分析。若是遺漏了舉足輕重的變數，分析的結果可能偏頗。再者，因為階層集群分析是一種勘察方法，所以在與獨立的樣本確認之前，都應該將結果看成是暫時性的結果。

若要取得階層集群分析

▶ 從功能表選擇：
分析(A) > 分類 > 階層集群分析法...

圖表 25-1
「階層集群分析」對話方塊

- 如果您將觀察值分成集群的話，請至少選擇一個數字變數。如果您將變數分成集群的話，請至少選擇三個數字變數。

  （選用性）您可以選出一個識別變數，以標記觀察值。

階層集群分析法

圖表 25-2
「階層集群分析法」對話方塊

集群方法。可用的選項包括：群間連結、組內變數連結、最近鄰法、最遠鄰法、重心集群化、中位數集群化，以及 Ward’s 法。

量數。可讓您指定集群方法中，所要使用的距離或相似性測量。選擇資料類型和適當的距離或相似性測量：

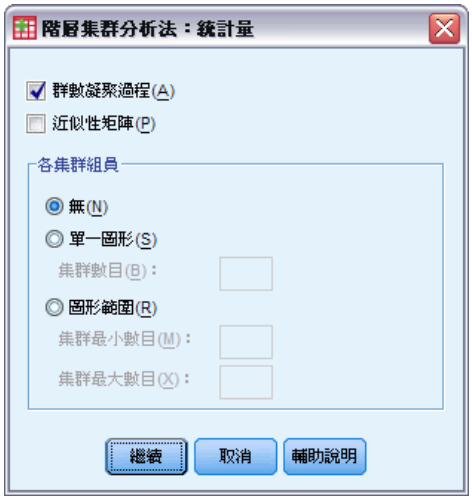
- **間隔。**可用的選項包括：歐基里得直線距離、歐基里得直線距離平方、餘弦、Pearson 相關、Chebychev 距離、區塊距離、Minkowski 距離和自訂的距離。
- **個數。**可用的選項包括：卡方量數和 Phi 平方量數。
- **二進位。**可用的選項包括：歐基里得直線距離、歐基里得直線距離平方、大小差異、型態差異、變異數、離散、類型、簡單配對、Phi 四點相關、Lambda 值、Anderberg’s D、骰子、Hamann、Jaccard、Kulczynski 1、Kulczynski 2、Lance 與 Williams、Ochiai、Rogers 與 Tanimoto、Russel 與 Rao、Sokal 與 Sneath 1、Sokal 與 Sneath 2、Sokal 與 Sneath 3、Sokal 與 Sneath 4、Sokal 與 Sneath 5 等相似性測量，以及 Yule’s Y 與 Yule’s Q 。

轉換值。讓您在計算相似性之前，先將觀察值或數值的資料值標準化（不適用於二元資料）。在這裡，您可以使用的標準化方法，包括：z 分數、範圍– -1 到 1、範圍 0 到 1、1 的最大級數、1 的平均數、1 的標準差。

轉換測量。可讓您轉換由距離測量所產生的值。但是，請注意，您必須在完成距離測量計算之後，才能套用這些值。可用的選項包括：絕對值、變更正負號，和將距離尺度化為 0- 到 1 的範圍。

階層集群分析統計量

圖表 25-3
「階層集群分析統計量」對話方塊



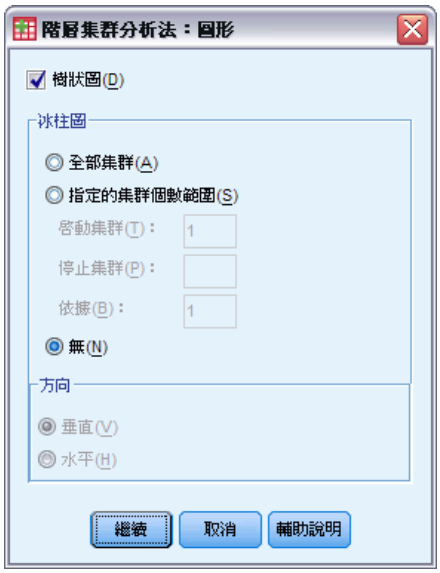
群數凝聚過程。這個選項會顯示每個階段所結合的觀察值或集群、要結合的觀察值或集群之間的距離，以及觀察值（或變數）與集群結合時的最後集群水準。

相似矩陣。提供項目之間的距離或相似性。

各集群組員。在集群組合時，顯示在一個或多個階段中，指定至各觀察值的集群。可用的選項包括單一解或解的範圍。

階層集群分析圖

圖表 25-4
「階層集群分析圖形」對話方塊

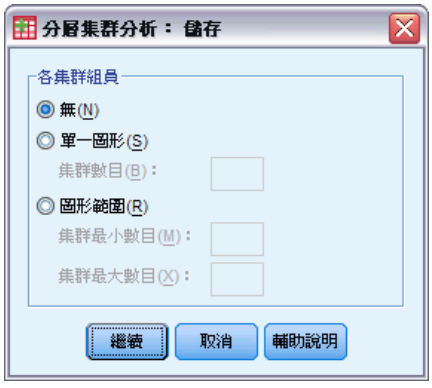


樹狀圖。顯示一個**樹狀圖**。樹狀圖可用來評估所形成之集群的內聚性，並提供適當集群數的相關資訊以供保存。

冰柱圖。顯示一個**冰柱圖**，包括所有集群或指定範圍的集群。冰柱圖會顯示有關每次疊代分析時，觀察值如何組合成集群的資訊。您可以用定位來選擇垂直或水平的圖形。

階層集群分析儲存新變數

圖表 25-5
「階層集群分析的儲存」對話方塊



各集群組員。可讓您儲存單一解或解的範圍之各集群組員。這樣便可以在後續分析中，用儲存的變數來探索組間的其他差異。

CLUSTER 指令語法其他功能

「階級集群」程序會使用 CLUSTER 指令語法。指令語法語言也可以讓您：

- 使用一些單一分析中的集群方法。
- 讀取與分析相似矩陣。
- 將相似矩陣寫入磁碟，以供稍後分析。
- 在自訂（電源）距離測量中指定電源和根目錄的任何數值。
- 指定儲存變數的名稱。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
26

K 平均數集群分析

您可以利用這個程序，讓演算法（該法可以處理大量觀察值）根據所選用的特性，找出性質相似的觀察值群組。但是，演算法會要求您指定集群數目。如果您知道這個資訊，便可以指定初始集群中心。觀察值分類方式有兩種，一種就是反覆更新各集群中心，另一種只會分類，您可以任選其中一種。您可以儲存集群組員、距離資訊、和最後集群的中心。除此之外，還可以指定變數，並用其值來標示依觀察值順序的輸出。您也可以要求分析變異數 F 統計量。雖然這些統計量都是隨機的（此程序會試著建立不同的組別），但是由統計量的相對大小，就可以看出每個變數在分組上的用處。

範例。 會吸引不同群組中類似觀眾的部分相同電視節目為何？利用 k 平均數集群分析，您可以根據檢視者的特性，將電視節目（觀察值）分成 k 同質群組。這個程序可以用來區隔市場。或者您可以將城市（觀察值）分成同質群組，以便從中選出可以比較的城市，並檢驗不同的市場策略。

統計量。 完成解答：各集群初始的中心、ANOVA 摘要表。各觀察值：集群資訊，距集群中心的距離。

資料。 變數在區間或比例水準時應該是量化的。如果您的變數是二元資料或個數的話，請使用「階層集群分析」程序。

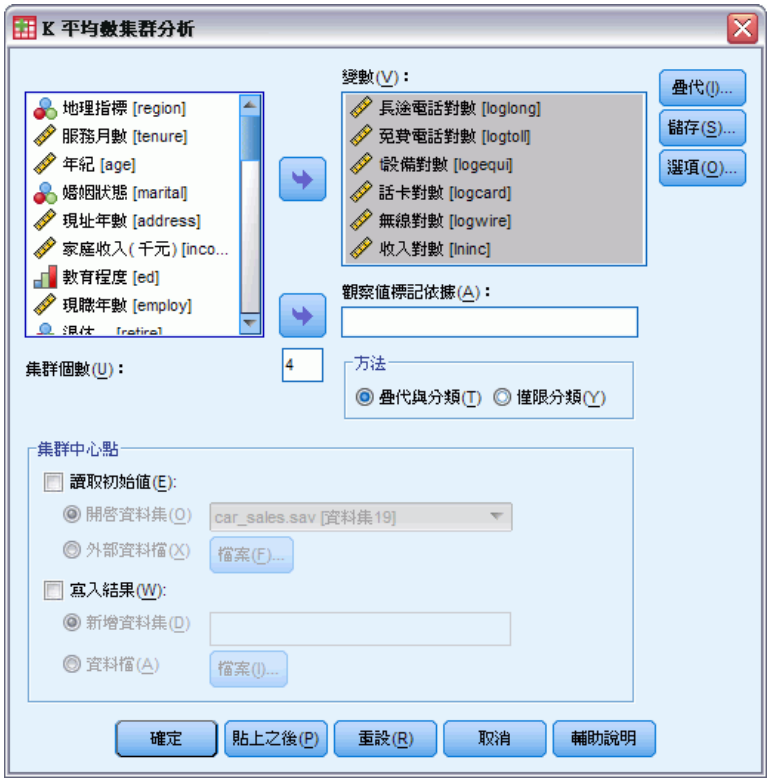
觀察值與集群初始中心順序 選擇集群初始中心的預設演算會改變觀察值的順序。「疊代」對話方塊中的「**使用可動平均數**」選項，會讓產生的解答必須依觀察值的順序而定，無論選取集群初始中心的方式為何。如果使用這些方法，您可以取得一些觀察值之不同的解答，這些觀察值會以不同的隨機順序排序，以確認給定解答的穩定性。指定各集群初始的中心而不使用「**使用可動平均數**」選項，可避免關於觀察值順序的問題。但是，如果從觀察值到集群中心為同分距離時，集群初始中心的順序可能會影響解答。若要取得給定解答的穩定性，您可以將分析的結果與初始中心值的不同排列互相比較。

假設。 使用簡單歐基里得直線距離來計算距離。如果您想要使用其他距離或類似的量數，請使用「階層集群分析法」程序。變數的尺度，應該是考慮重點之一。如果您用不同的尺度來測量變數的話（例如，某個變數以貨幣來表示，而另一個變數則是以年代來表示），則結果可能會令人誤解。在這類情況下，您應該考慮再執行 k 平均數集群分析前，將變數標準化（可於「敘述統計」程序中執行這項工作）。本程序假設您已選取適當的集群個數，而且已經納入所有相關的變數。如果所選用的集群個數不適當，或者省掉重要的變數的話，則您的結果會令人誤解。

若要取得 K 平均數集群分析

▶ 從功能表選擇：
分析(A) > 分類 > K 平均數集群...

圖表 26-1
K 平均數集群分析對話方塊



- 選取集群分析要使用的變數。
- 指定集群的個數。(集群個數至少必須等於 2，而且必須小於（或等於）資料檔中，觀察值的個數)。
- 選取 **疊代與分類**或**只限於分類**。
- 您可隨意地選擇識別變數來標記觀察值。

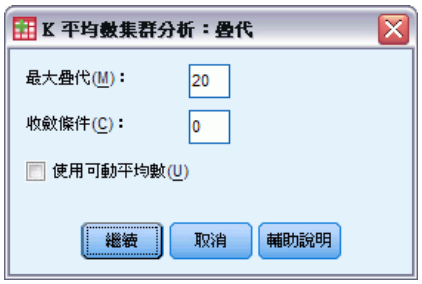
K 平均數集群分析效率

使用 k 平均數集群分析指令非常有效率，主要是因為它不會計算所有成對觀察值之間的距離（有很多集群演算法會)，包括階層式集群指令所使用的演算法。

為了產生最佳效能，請採用單一觀察值樣本，並使用「**疊代與分類**」方法，來決定每個集群的中心。選擇「**將最終值寫成**」。然後將整個資料檔案還原，並選擇「**只限於分類**」做為方法，以及選擇「**從這裡讀取初始值**」，使用從範例估計的中心，將整個檔案分類。您可以從檔案或資料集寫入或讀取。資料集可供同一階段作業中之後續的作業使用，但是不會儲存為檔案，除非特別在階段作業結束之前儲存。資料集名稱必須符合變數命名規則。

K 平均數集群分析疊代

圖表 26-2
K 平均數集群分析疊代對話方塊



注意：只有當您從「K 平均數集群分析」對話方塊中選擇「**疊代與分類**」方法時，才能使用這些選項。

最大疊代。限制 k 平均數演算法中的疊代次數。即使未達到收斂準則，疊代也會在達到這個次數之後停止。此數字必須介於 1 到 999 之間。

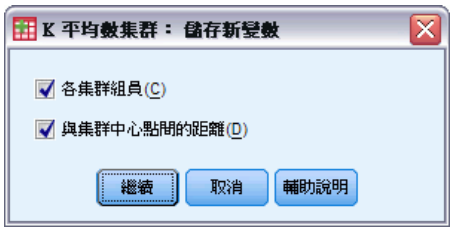
若要重製 5.0 版之前 Quick Cluster 指令所使用的演算法，請將**最大疊代**設為 1。

收斂準則。 決定疊代停止的時間。它代表集群初始中心之間最小的距離比例，所以它必須大於 0 但小於或等於 1。例如，如果此準則等於 0.02，則當完整疊代未依照指定距離來移動任何集群中心時（該距離必須比集群初始中心之間的最小距離，高出兩個百分比），疊代就會終止。

使用可動平均數。 您可要求在指定每個觀察值之後更新集群的中心。如果沒有選用本選項，則您必須先指定所有觀察值，才能計算新的集群中心。

K 平均數集群分析的儲存

圖表 26-3
K 平均數集群分析儲存新變數對話方塊



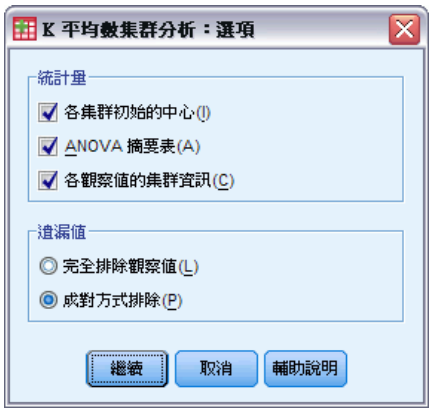
您可以儲存解答的相關資訊，並在後續分析中，當成新變數使用。

各集群組員。 建立新變數，用來代表每個觀察值的最後一位集群組員。此新變數值應該介於 1 跟集群個數之間。

與集群中心點間的距離。 可建立可表示各觀察值間歐基里得直線距離及其分類中心的新變數。

K 平均數集群分析的選項

圖表 26-4
K 平均數集群分析選項對話方塊



統計量。 您可以選取下列統計量：各觀察值的初始集群中心、ANOVA 摘要表、和集群資訊。

- **各集群初始的中心(I)。** 每個集群之變數平均數的第一個估計值。根據預設值，經妥善安排間隔的觀察值個數會等於自資料中選取的集群個數。在進行第一次分類時會使用初始集群中心，之後便會將其更新。
- **ANOVA 表。** 顯示包含每項集群變數之單變量 F 檢定的變異數分析表。F 檢定僅具描述性質，且不應解譯產生機率。若針對單一集群指定所有觀察值，則不會顯示 ANOVA 表。
- **各觀察值的集群資訊(C)。** 顯示每個觀察值的最終集群指定，以及觀察值與用來分類觀察值之集群中心間的歐基里得直線距離。此外亦會顯示最終集群中心間的歐基里得直線距離。

遺漏值。可用的選項為「**完全排除遺漏值**」或「**成對方式排除**」。

- **完全排除遺漏值。** 不要分析任何集群變數中含遺漏值的觀察值。
- **成對方式排除。** 根據從不含遺漏值的所有變數中計算出的距離，以將觀察值指定到集群。

QUICK CLUSTER 指令的其他功能

「K 平均數集群」程序會使用 QUICK CLUSTER 指令語法。指令語法語言也可以讓您：

- 讓第一個 k 觀察值，當作初始化的集群中心，如此就可避免傳遞資料（一般用於估計該值）。
- 直接指定各集群初始的中心，並當成指令語法的一部份。
- 指定儲存變數的名稱。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
27

無母數檢定

無母數檢定可做出關於資料基本分配的最小假設。這些方塊中提供的檢定可以根據資料組織的方式分成三個廣義的類別：

- 單一樣本檢定會分析一個欄位。
- 相關樣本檢定會比較相同觀察值組合的二或多個欄位。
- 獨立樣本檢定會分析另一個欄位的類別組成的欄位。

單一樣本無母數檢定

單一樣本無母數檢定可以使用一或多個無母數檢定識別單一欄位中的差異。無母數檢定不會假設您的資料以常態分配。

圖表 27-1
「單一樣本無母數檢定目標」索引標籤

您的目標是什麼？ 目標除可讓您快速指定其他的常用檢定設定。

- **自動比較觀察資料和假設資料。** 此目標可將二項式檢定套用至僅含有兩種類別的類別欄位、將卡方檢定套用至其他所有類別欄位，以及將 Kolmogorov-Smirnov 檢定套用至連續欄位。
- **檢定隨機順序。** 此目標使用連檢定來檢定觀察的資料值順序的隨機性。
- **自訂分析。** 當您想在「設定」索引標籤中手動修正檢定設定時，請選取此選項。請注意，若您之後對「設定」索引標籤中的選項進行變更，但該變更與目前選擇的目標不符時，會自動選取此設定。

取得單一樣本無母數檢定

從功能表選擇：
分析(A) > 無母數檢定 > 單一樣本...

▶ 按一下「執行」。

您可以：

- 在「目標」索引標籤上指定目標。
- 在「欄位」索引標籤上指定欄位指派。
- 在「設定」索引標籤上指定匯出設定。

欄位索引標籤

圖表 27-2
「單一樣本無母數檢定欄位」索引標籤

「欄位」索引標籤指定應檢定哪些欄位。

使用預先定義的角色。 此選項使用現有的欄位資訊。角色預先定義為「輸入」、「目標」或「兩者」的所有欄位，將會做為檢定欄位。至少需要一個檢定欄位。

使用自訂欄位指派。此選項可讓您覆寫欄位角色。選取此選項後，請指定下列欄位：

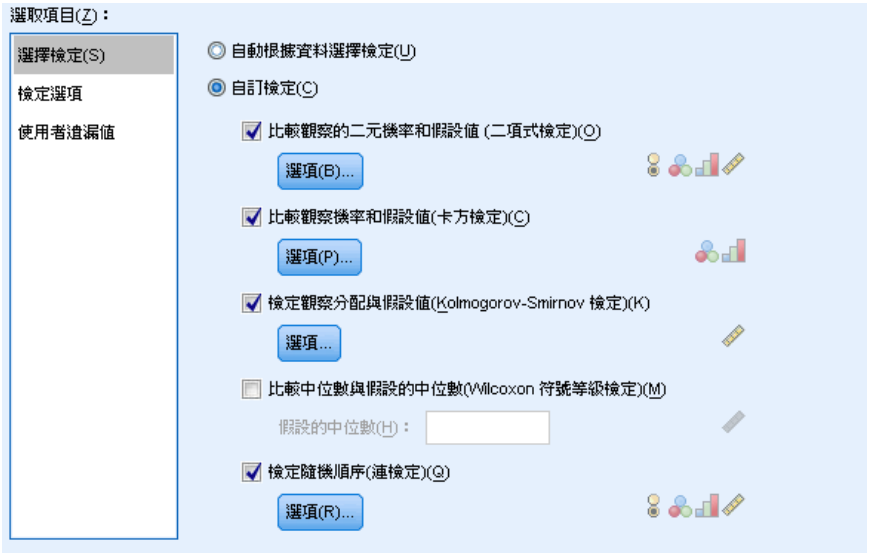
- **檢定欄位。**選取一或多個欄位。

設定索引標籤

「設定」索引標籤含有多種設定群組，可讓您修改以微調演算法處理資料的方式。若您對預設設定所做的任何變更與目前選取的目標不符，「目標」索引標籤會自動更新為選取「自訂分析」選項。

選擇檢定

圖表 27-3
單一樣本無母數檢定的「選擇檢定」設定



這些設定會指定在「欄位」索引標籤中指定之欄位所執行的檢定。

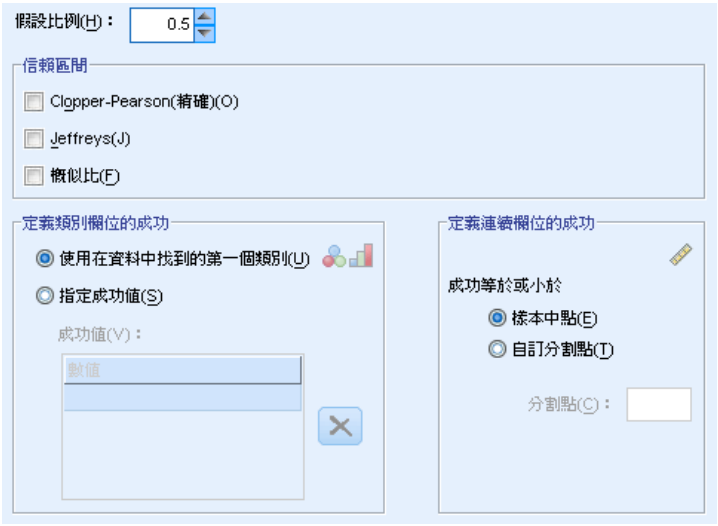
自動依據資料選擇檢定。此設定可將二項式檢定套用至僅含有兩種有效（非遺漏）類別的類別欄位、將卡方檢定套用至其他所有類別欄位，以及將 Kolmogorov-Smirnov 檢定套用至連續欄位。

自訂檢定。此設定可以讓您選擇要執行的特定檢定。

- **比較觀察的二元機率和假設值（二項式檢定）。**二項式檢定可以套用至所有欄位。這會產生單一樣本檢定，以檢定觀察的旗標欄位（一種只含有兩種類別的類別欄位）分配是否與指定之二項式分配的預期分配相同。此外，您也可以要求信賴區間。請參閱二項式檢定選項，了解檢定設定的詳細資訊。
- **比較觀察機率和假設值(卡方檢定)。**卡方檢定會套用至名義和次序欄位。這會產生單一樣本檢定，以根據觀察和預期的欄位類別次數間的差異來計算卡方統計量。請參閱卡方檢定選項，了解檢定設定的詳細資訊。
- **檢定觀察分配與假設值(Kolmogorov-Smimov 檢定)。**Kolmogorov-Smirnov 檢定會套用至連續欄位。這會產生單一樣本檢定，以檢定欄位的樣本累積分配函數是否與均勻、名義、Poisson 或指數分配同質。請參閱Kolmogorov-Smirnov 選項，了解檢定設定的詳細資訊。
- **比較中位數與假設的中位數(Wilcoxon 符號等級檢定)。** Wilcoxon 符號等級檢定會套用至連續欄位。這會產生欄位中位數值的單一樣本檢定。請指定一個數字做為假設的中位數。
- **檢定隨機順序(連檢定)。**連檢定會套用至所有欄位。這會產生單一樣本檢定，以檢定二元欄位的值順序是否為隨機。請參閱連檢定選項，了解檢定設定的詳細資訊。

二項式檢定選項

圖表 27-4
單一樣本無母數檢定的「二項式檢定」選項



二項式檢定適用於旗標欄位（只含有兩種類別的類別欄位），但使用定義「成功」的規則便可套用至所有欄位。

假設比例。 這會將已定義記錄的預期比例指定為「成功」，也就是 p。指定大於 0 且小於 1 的一個數值。預設值為 0.5。

信賴區間。 可使用下列計算二元資料信賴區間的方法：

- **Clopper-Pearson(精確)。**一種以累積分配函數為基礎的精確區間。
- **Jeffreys。**一種以先前使用 Jeffreys 計算的 p 事後分配為基礎的 Bayesian 區間。
- **概似比。** 一種以 p 的概似函數為基礎的區間。

定義類別欄位的成功。這會指定類別欄位是如何定義「成功」(檢定的資料值對照假設比例)。

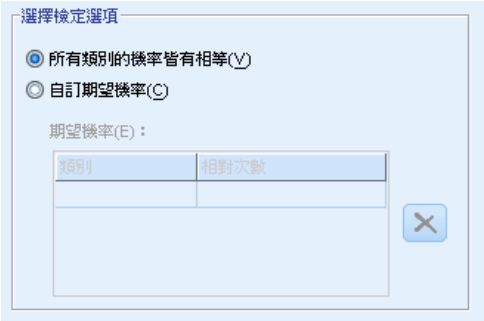
- 「使用在資料中找到的第一個類別」會使用在樣本中找到的第一個值執行二項式檢定，以定義「成功」。此選項僅適用於僅含有兩個值的名義或次序欄位；將不會檢定「欄位」索引標籤中指定之其他所有（已使用此選項的）類別欄位。此為預設值。
- 「指定成功值」會使用指定的值清單來執行二項式檢定，以定義「成功」。請指定字串或數值清單。清單中的值不需存在樣本中。

定義連續欄位的成功。這會指定連續欄位是如何定義「成功」(檢定的資料值對照檢定值)。「成功」會定義為等於或小於分割點的值。

- 「樣本中點」會在最小值和最大值的平均值設定分割點。
- 「自訂分割點」可以讓您指定分割點的值。

卡方檢定選項

圖表 27-5
單一樣本無母數檢定的「卡方檢定」選項



所有類別都有相同的機率。這會在樣本的所有類別之間產生相等次數。此為預設值。

自訂期望機率。這可讓您為指定的類別清單指定不相等次數。請指定字串或數值清單。清單中的值不需存在樣本中。在「**類別**」行中，指定類別值。在「**相對次數**」行中，為每個類別指定大於 0 的值。自訂次數會視為比率處理，因此，例如指定次數 1、2 和 3 等同於指定次數 10、20 和 30，且兩者都會指定 1/6 的記錄預期會落入第一個類別、1/3 的記錄落入第二個，1/2 的記錄落入第三個。指定自訂的預期機率時，自訂類別值必須包含資料中所有的欄位值；否則不會為該欄位執行檢定。

Kolmogorov-Smirnov 選項

圖表 27-6
單一樣本無母數檢定的 Kolmogorov-Smirnov 選項

此對話方塊指定應檢定哪些分配，以及假設分配的參數。

常態。 「使用樣本資料」會使用觀察的平均值和標準差，「自訂」可讓您指定值。

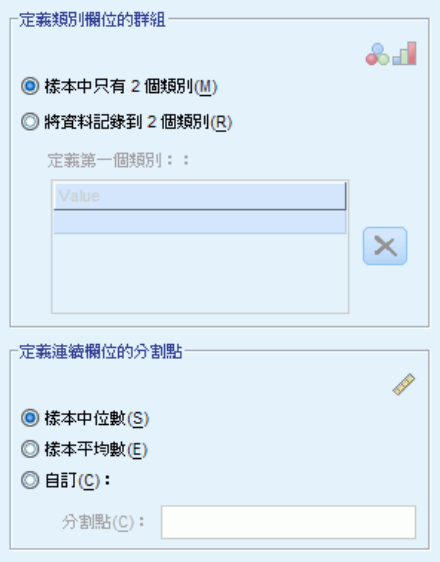
均勻。 「使用樣本資料」會使用觀察的最小值和最大值，「自訂」可讓您指定值。

指數模式。 「樣本平均數」會使用觀察的平均數，「自訂」可讓您指定值。

Poisson。 「樣本平均數」會使用觀察的平均數，「自訂」可讓您指定值。

連檢定選項

圖表 27-7
單一樣本無母數檢定的「連檢定」選項



連檢定適用於旗標欄位（只含有兩種類別的類別欄位），但使用定義群組的規則便可套用至所有欄位。

定義類別欄位的群組

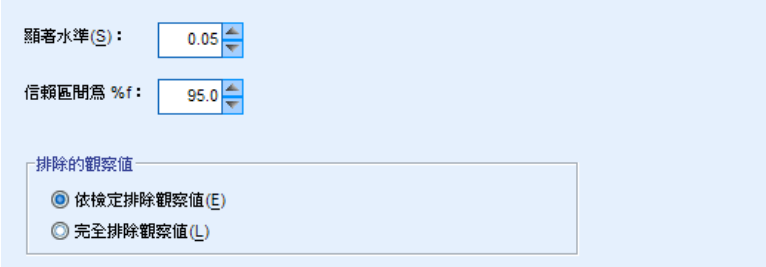
- 「樣本中只有 2 個類別」會使用在樣本中找到的值執行連檢定，以定義群組。此選項僅適用於僅含有兩個值的名義或次序欄位；將不會檢定「欄位」索引標籤中指定之其他所有（已使用此選項的）類別欄位。
- 「將資料記錄到 2 個類別」會使用指定的值清單執行連檢定，以定義其中一個群組。樣本中其他所有的值會定義其他群組。清單中的值不需全部存在於樣本中，但每個群組中至少需有一個記錄。

定義連續欄位的分割點。這會指定連續欄位的群組如何定義。第一個群組會定義為等於或小於分割點的值。

- 「樣本中位數」會在樣本中位數設定分割點。
- 「樣本平均數」會在樣本平均數設定分割點。
- 「自訂」可以讓您指定分割點的值。

檢定選項

圖表 27-8
單一樣本無母數檢定的「檢定選項」設定



顯著水準。這會指定所有檢定的顯著水準 (alpha)。請指定 0 到 1. 0.05 之間的數值做為預設值。

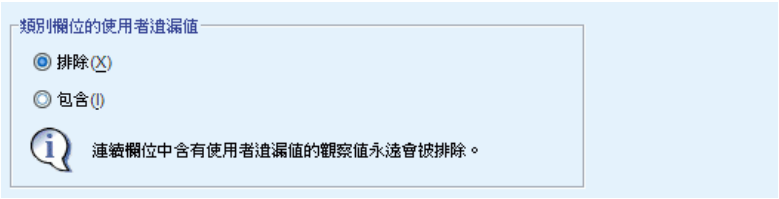
信賴區間 (%)。這會指定所有產生之信賴區間的信賴水準。請指定 0 到 100. 95 之間的數值做為預設值。

排除的觀察值。這會指定如何決定檢定的觀察值基礎。

- 「完全排除遺漏值」表示分析會排除任何在「欄位」索引標籤上所指定任何欄位內含有遺漏值的記錄。
- 「依檢定排除遺漏值」表示該檢定會略過用於特定檢定之欄位內含有遺漏值的記錄。在分析中指定多項檢定時，會個別評估每項檢定。

使用者遺漏值

圖表 27-9
單一樣本無母數檢定的「使用者遺漏值」設定



類別欄位的使用者遺漏值。類別欄位必須具有要納入分析之記錄的有效值。這些控制可讓您決定是否要在類別欄位中，將使用者遺漏值視為有效值。系統遺漏值和連續欄位的遺漏值會永遠視為無效。

獨立樣本無母數檢定

獨立樣本無母數檢定可以使用一或多個無母數檢定，以識別兩個或多個群組間的差異。無母數檢定不會假設您的資料以常態分配。

圖表 27-10
「獨立樣本無母數檢定目標」索引標籤

您的目標是什麼？ 目標除可讓您快速指定其他的常用檢定設定。

- **自動比較群組間的分配。** 此目標會將 Mann-Whitney U 檢定套用至含有 2 個群組的資料，或將 Kruskal-Wallis 單因子變異數分析套用至含有 k 個群組的資料。
- **比較群組間的中位數。** 此目標使用中位數檢定，以比較在群組間觀察的中位數。
- **自訂分析。** 當您想在「設定」索引標籤中手動修正檢定設定時，請選取此選項。請注意，若您之後對「設定」索引標籤中的選項進行變更，但該變更與目前選擇的目標不符時，會自動選取此設定。

取得獨立樣本無母數檢定

從功能表選擇：
分析(A) > 無母數檢定 > 獨立樣本...

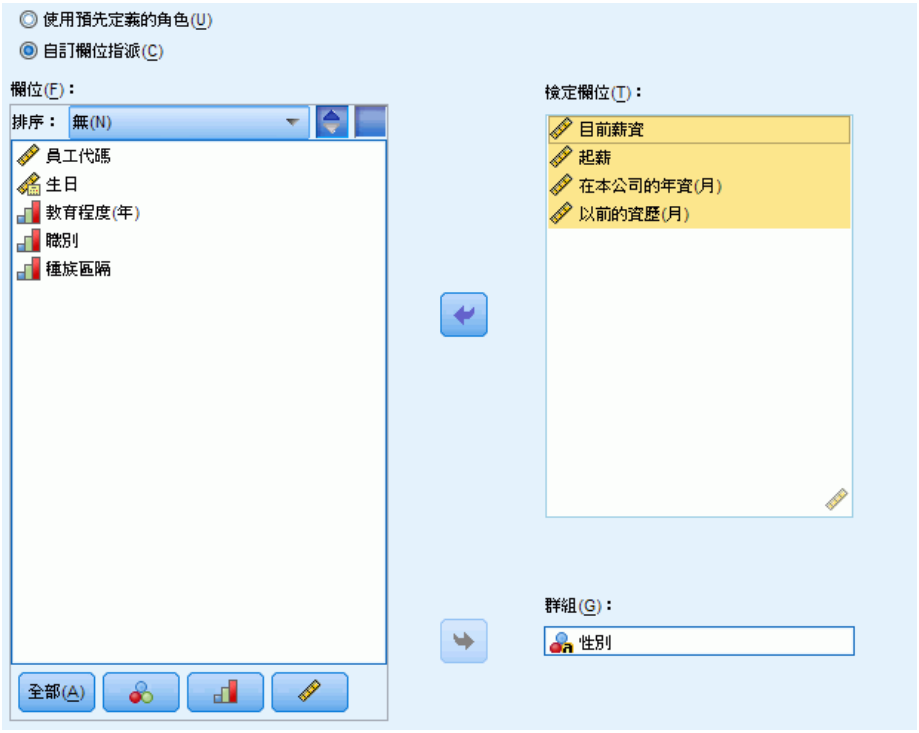
▶ 按一下「執行」。

您可以：

- 在「目標」索引標籤上指定目標。
- 在「欄位」索引標籤上指定欄位指派。
- 在「設定」索引標籤上指定匯出設定。

欄位索引標籤

圖表 27-11
「獨立樣本無母數檢定欄位」索引標籤



「欄位」索引標籤指定應檢定哪些欄位，以及用來定義群組的欄位。

使用預先定義的角色。 此選項使用現有的欄位資訊。角色預先定義為「目標」或「兩者」的所有連續欄位，將會做為檢定欄位。如果有角色預先定義為「輸入」的單一個類別欄位，則此欄位將會作為群組欄位。如果沒有的話，預設不會使用任何群組欄位，您必須使用自訂欄位指派。至少需要一個檢定欄位和群組欄位。

使用自訂欄位指派。此選項可讓您覆寫欄位角色。選取此選項後，請指定下列欄位：

- **檢定欄位。**選取一或多個連續欄位。
- **群組。**請選取類別欄位。

設定索引標籤

「設定」索引標籤含有多種設定群組，可讓您修改以微調演算法處理資料的方式。若您對預設設定所做的任何變更與目前選取的目標不符，「目標」索引標籤會自動更新為選取「**自訂分析**」選項。

選擇檢定

圖表 27-12
獨立樣本無母數檢定的「選擇檢定」設定

這些設定會指定在「欄位」索引標籤中指定之欄位所執行的檢定。

自動依據資料選擇檢定。此設定會將 Mann-Whitney U 檢定套用至含有 2 個群組的資料，或將 Kruskal-Wallis 單因子變異數分析套用至含有 k 個群組的資料。

自訂檢定。此設定可以讓您選擇要執行的特定檢定。

- **比較群組間的分配。**這些會產生樣本是否來自相同母群的獨立樣本檢定。

  「**Mann-Whitney U (2 個樣本)**」使用每個觀察值的等級，以檢定群組是否是由相同的母群中抽取的。群組欄位中以遞增順序排列的第一個值定義第一個群組，第二個則定義第二個群組。若群組欄位有兩個以上的值，則不會產生此檢定。

  「**Kolmogorov-Smirnov (2 個樣本)**」可察覺兩種分配之間的中位數、分散情形及偏態等的任何差異。若群組欄位有兩個以上的值，則不會產生此檢定。

  「**檢定隨機順序 (2 個樣本的 Wald-Wolfowitz)**」會使用群組成員做為準則以產生連檢定。若群組欄位有兩個以上的值，則不會產生此檢定。

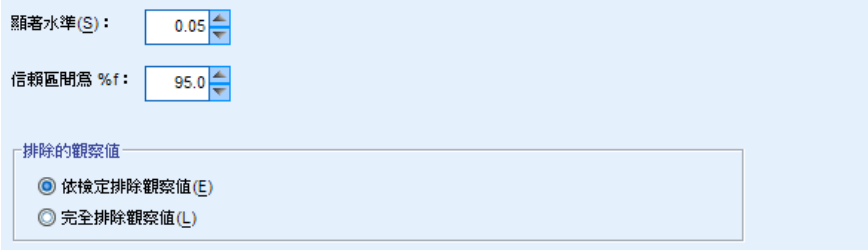
  「**Kruskal-Wallis 單因子變異數分析 (k 個樣本)**」是 Mann-Whitney U 檢定的延伸，以及單因子變異數分析的無母數類比。您可以選擇性要求 k 個樣本的多重比較，無論是「**所有成對**」的多重比較，或是「**逐步細分程序**」的比較。

「**排序選項的檢定 (k 個樣本的 Jonckheere-Terpstra)**」是當 k 個樣本擁有自然排序時比 Kruskal-Wallis 更有力的替代選項。例如，k 母群可能代表所增加的溫度 (k 度)。而我們可以先假設不同溫度，但回應分配是一樣的，以便跟另一個假設：溫度增加、回應等級就隨之增加，來互相對照，並加以檢定。這兩個假設都已經排序過，因此，Jonckheere-Terpstra 是最適用的檢定。請指定其他假設的順序；**最小至最大**會規定一個其他假設，即第一個群組的位置參數不等於第二個，第二個不等於第三個，依此類推；**最大至最小**會規定一個其他假設，即最後一個群組的位置參數不等於倒數第二個，而倒數第二個不等於倒數第三個，依此類推。您可以選擇性要求 k 個樣本的多重比較，無論是「**所有成對**」的多重比較，或是「**逐步細分程序**」的比較。

- **比較群組間的範圍。**這會產生樣本是否擁有相同範圍的獨立樣本檢定。「**Moses 極端反應 (2 個樣本)**」會檢定控制群組與比較群組。群組欄位中以遞增順序排列的第一個值定義控制群組，第二個則定義比較群組。若群組欄位有兩個以上的值，則不會產生此檢定。
- **比較群組間的中位數。**這會產生樣本是否擁有相同中位數的獨立樣本檢定。「**中位數檢定 (k個樣本)**」可以使用合併樣本中位數 (在資料集的所有記錄間計算而得) 或自訂值做為假設的中位數。您可以選擇性要求 k 個樣本的多重比較，無論是「**所有成對**」的多重比較，或是「**逐步細分程序**」的比較。
- **估計群組間的信賴區間。**「**Hodges-Lehman 估計 (2 個樣本)**」會針對兩個群組的中位數差異，產生獨立樣本估計與信賴區間。若群組欄位有兩個以上的值，則不會產生此檢定。

檢定選項

圖表 27-13
獨立樣本無母數檢定的「檢定選項」設定



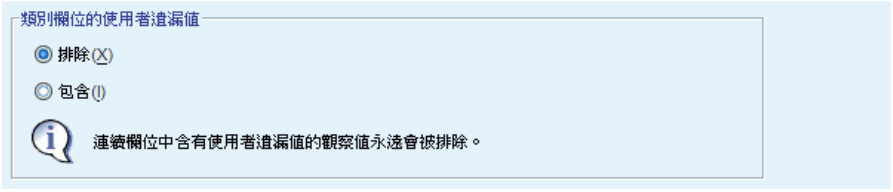
顯著水準。這會指定所有檢定的顯著水準 (alpha)。請指定 0 到 1. 0.05 之間的數值做為預設值。

信賴區間 (%)。這會指定所有產生之信賴區間的信賴水準。請指定 0 到 100. 95 之間的數值做為預設值。

排除的觀察值。這會指定如何決定檢定的觀察值基礎。「**完全排除遺漏值**」表示分析會排除任何次指令中所指定任何欄位內含有遺漏值的記錄。「**依檢定排除遺漏值**」表示該檢定會略過用於特定檢定之欄位內含有遺漏值的記錄。在分析中指定多項檢定時，會個別評估每項檢定。

使用者遺漏值

圖表 27-14
獨立樣本無母數檢定的「使用者遺漏值」設定



類別欄位的使用者遺漏值。類別欄位必須具有要納入分析之記錄的有效值。這些控制可讓您決定是否要在類別欄位中，將使用者遺漏值視為有效值。系統遺漏值和連續欄位的遺漏值會永遠視為無效。

相關樣本無母數檢定

使用一或多個無母數檢定識別兩個以上之相關欄位間的差異。無母數檢定不會假設您的資料以常態分配。

資料考量。每筆記錄所對應的指定受訪者，皆有兩個以上的相關量數儲存在資料集的不同欄位中。例如，如果定期測量每個受訪者的體重，並將受訪者的體重儲存在如「飲食法實施前體重」、「飲食法實施中體重」與「飲食法實施後體重」等欄位，則可以使用相關樣本無母數檢定來分析有關飲食計劃效果的研究。這些欄位是「相關的」。

圖表 27-15
「相關樣本無母數檢定目標」索引標籤

您的目標是什麼？目標除可讓您快速指定其他的常用檢定設定。

- **自動比較觀察資料和假設資料。**當指定 2 個欄位時，此目標會將 McNemar 檢定套用至類別資料，當指定 2 個以上的欄位時，會將 Cochran’s Q 套用至類別資料，當指定 2 個欄位時，會將 Wilcoxon 配對組別符號等級檢定套用至連續資料，當指定 2 個以上的欄位時，會將 Friedman 二因子變異數分析套用至連續資料。
- **自訂分析。** 當您想在「設定」索引標籤中手動修正檢定設定時，請選取此選項。請注意，若您之後對「設定」索引標籤中的選項進行變更，但該變更與目前選擇的目標不符時，會自動選取此設定。

當指定不同測量水準的欄位時，系統會先以測量水準區隔這些欄位，再將適當的檢定套用到每個組別。例如，如果您選擇「**自動比較觀察資料和假設資料**」作為您的目標，並指定 3 個連續欄位與 2 個名義欄位，則 Friedman’s 檢定會套用到連續欄位，McNemar’s 檢定會用到名義欄位。

取得相關樣本無母數檢定

從功能表選擇：
分析 > 無母數檢定 > 相關樣本...

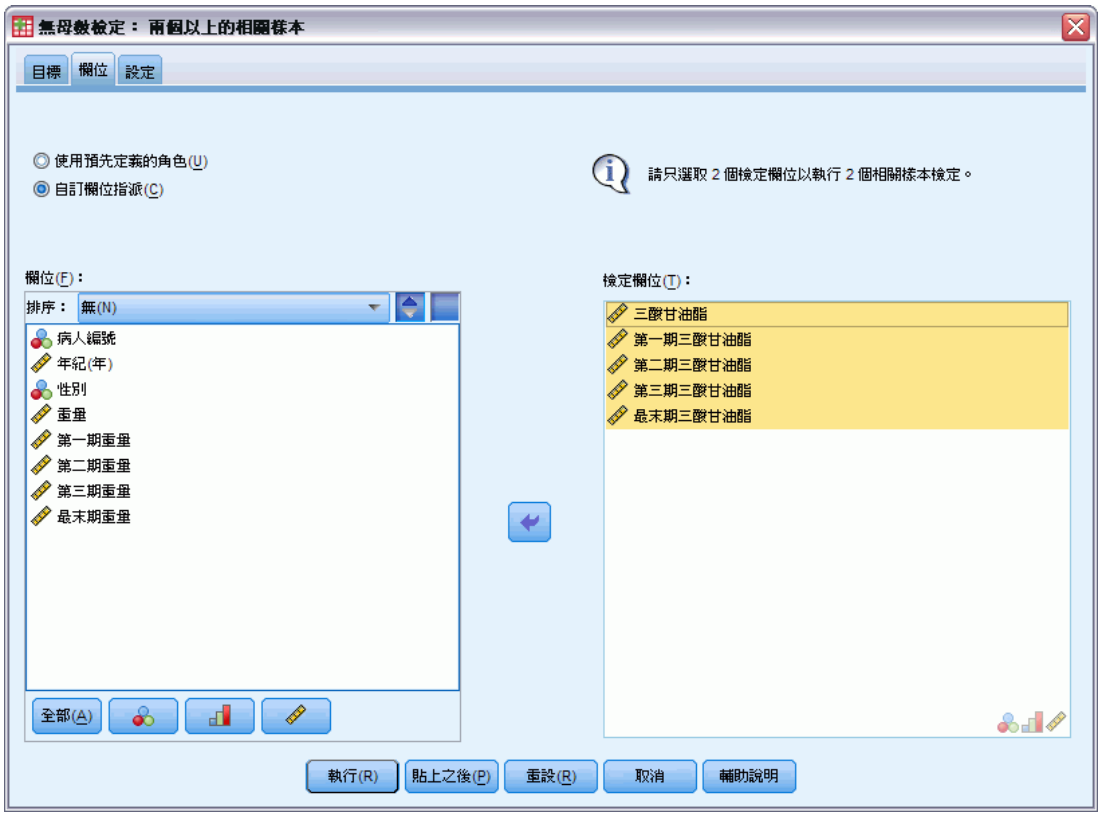
▶ 按一下「**執行**」。

您可以：

- 在「目標」索引標籤上指定目標。
- 在「欄位」索引標籤上指定欄位指派。
- 在「設定」索引標籤上指定匯出設定。

欄位索引標籤

圖表 27-16
「相關樣本無母數檢定欄位」索引標籤



「欄位」索引標籤指定應檢定哪些欄位。

使用預先定義的角色。 此選項使用現有的欄位資訊。角色預先定義為「目標」或「兩者」的所有欄位，將會做為檢定欄位。至少需要兩個檢定欄位。

使用自訂欄位指派。此選項可讓您覆寫欄位角色。選取此選項後，請指定下列欄位：

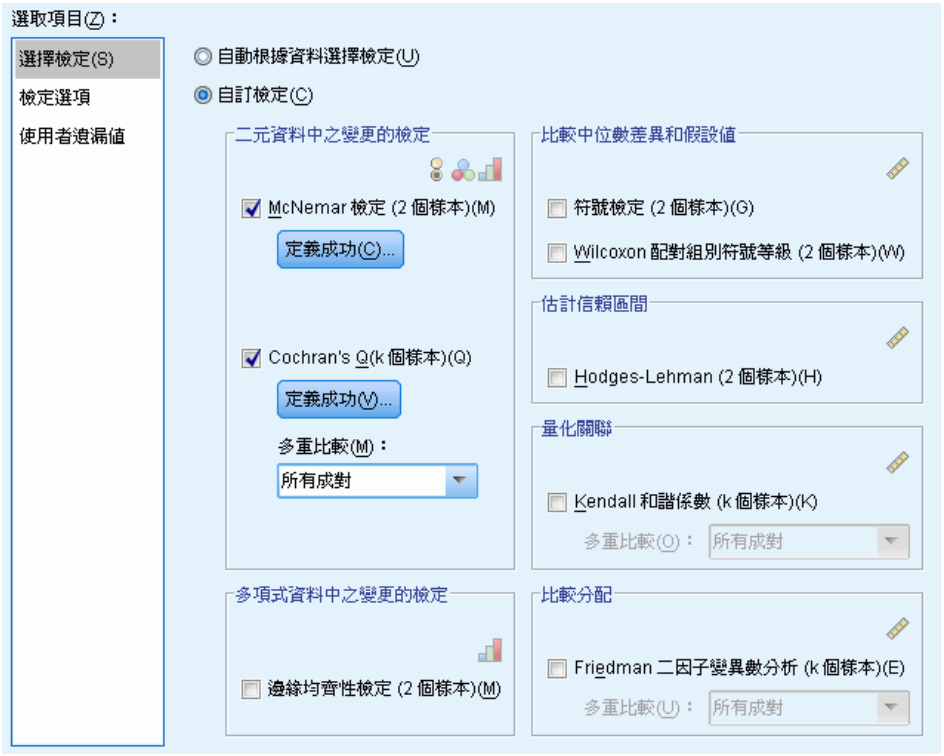
- **檢定欄位。**選取兩個以上的欄位。每個欄位各對應至不同的相關樣本。

設定索引標籤

「設定」索引標籤含有多種設定群組，可讓您修改以微調程序處理資料的方式。若您對預設設定所做的任何變更與其他目標不符，「目標」索引標籤會自動更新為選取「自訂分析」選項。

選擇檢定

圖表 27-17
相關樣本無母數檢定的「選擇檢定」設定



這些設定會指定在「欄位」索引標籤中指定之欄位所執行的檢定。

自動依據資料選擇檢定。 當指定 2 個欄位時，此設定會將 McNemar 檢定套用至類別資料，當指定 2 個以上的欄位時，會將 Cochran’s Q 套用至類別資料，當指定 2 個欄位時，會將 Wilcoxon 配對組別符號等級檢定套用至連續資料，當指定 2 個以上的欄位時，會將 Friedman 二因子變異數分析套用至連續資料。

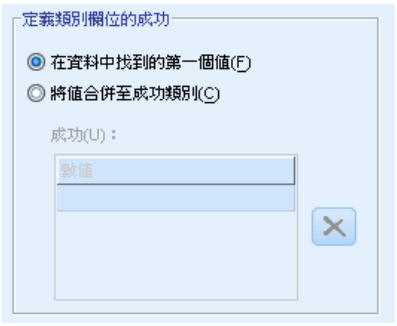
自訂檢定。此設定可以讓您選擇要執行的特定檢定。

- **二元資料中之變更的檢定。** 「McNemar 檢定 (2 個樣本)」可套用至類別欄位。這會產生兩個旗標欄位（僅含有兩個值的類別欄位）間的值組合是否具有相同可能性的相關樣本檢定。若「欄位」索引標籤中指定兩個以上的欄位，則不會執行此檢定。請參閱McNemar 檢定：定義成功，了解檢定設定的詳細資訊。「Cochran’s Q (k 個樣本)」可套用至類別欄位。這會產生 k 個旗標欄位（僅含有兩個值的類別欄位）間的值組合是否具有相同可能性的相關樣本檢定。您可以選擇性要求 k 個樣本的多重比較，無論是「所有成對」的多重比較，或是「逐步細分程序」的比較。請參閱Cochran’s Q：定義成功，了解檢定設定的詳細資訊。
- **多項式資料中之變更的檢定。** 「邊緣均齊性檢定 (2 個樣本)」會產生兩個成對次序欄位之間的值組合是否具有相同可能性的相關樣本檢定。邊緣均齊性檢定通常用於重複性測量。這個檢定是將 McNemar 檢定從二項反應擴充到多項式反應。若「欄位」索引標籤中指定兩個以上的欄位，則不會執行此檢定。

- **比較中位數差異和假設值。**這些檢定會各自產生兩個連續欄位間的中位數差異是否不同於 0 的相關樣本檢定。若「欄位」索引標籤中指定兩個以上的欄位，則不會執行這些檢定。
- **估計信賴區間。**這會產生兩個成對連續欄位間中位數差異的相關樣本估計和信賴區間。若「欄位」索引標籤中指定兩個以上的欄位，則不會執行此檢定。
- **Quanlify 關聯。** 「Kendall 和諧係數 (k 個樣本)」會產生裁判或定等級者間的一致性測量，其中，每個記錄各是多種項目 (欄位) 的一個審查評等。您可以選擇性要求 k 個樣本的多重比較，無論是「所有成對」的多重比較，或是「逐步細分程序」的比較。
- **比較分配。** 「二因子變異數分析 (k 個樣本)」會產生 k 個相關樣本是否是從相同母群抽取的相關樣本檢定。您可以選擇性要求 k 個樣本的多重比較，無論是「所有成對」的多重比較，或是「逐步細分程序」的比較。

McNemar 檢定：定義成功

圖表 27-18
相關樣本無母數檢定的「McNemar 檢定」設定：定義成功設定



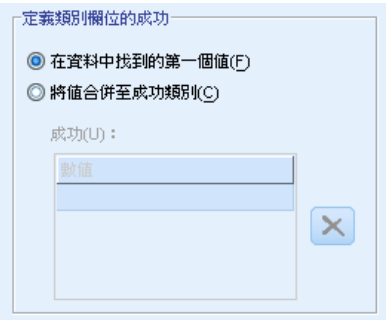
McNemar 檢定適用於旗標欄位 (只含有兩種類別的類別欄位)，但使用定義「成功」的規則便可套用至所有類別欄位。

定義類別欄位的成功。這會指定如何定義類別欄位的「成功」。

- 「使用在資料中找到的第一個類別」會使用在樣本中找到的第一個值執行檢定，以定義「成功」。此選項僅適用於僅含有兩個值的名義或次序欄位；將不會檢定「欄位」索引標籤中指定之其他所有 (已使用此選項的) 類別欄位。此為預設值。
- 「指定成功值」會使用指定的值清單來執行檢定，以定義「成功」。請指定字串或數值清單。清單中的值不需存在樣本中。

Cochran’s Q：定義成功

圖表 27-19
相關樣本無母數檢定的「Cochran’s Q」：定義成功



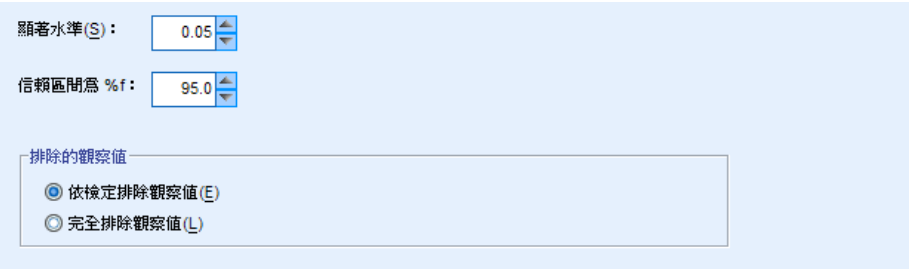
Cochran’s Q 檢定適用於旗標欄位（只含有兩種類別的類別欄位），但使用定義「成功」的規則便可套用至所有類別欄位。

定義類別欄位的成功。這會指定如何定義類別欄位的「成功」。

- 「**使用在資料中找到的第一個類別**」會使用在樣本中找到的第一個值執行檢定，以定義「成功」。此選項僅適用於僅含有兩個值的名義或次序欄位；將不會檢定「欄位」索引標籤中指定之其他所有（已使用此選項的）類別欄位。此為預設值。
- 「**指定成功值**」會使用指定的值清單來執行檢定，以定義「成功」。請指定字串或數值清單。清單中的值不需存在樣本中。

檢定選項

圖表 27-20
相關樣本無母數檢定的「檢定選項」設定



顯著水準。這會指定所有檢定的顯著水準 (alpha)。請指定 0 到 1. 0.05 之間的數值做為預設值。

信賴區間 (%)。這會指定所有產生之信賴區間的信賴水準。請指定 0 到 100. 95 之間的數值做為預設值。

排除的觀察值。這會指定如何決定檢定的觀察值基礎。

- 「**完全排除遺漏值**」表示分析會排除任何次指令中所指定任何欄位內含有遺漏值的記錄。
- 「**依檢定排除遺漏值**」表示該檢定會略過用於特定檢定之欄位內含有遺漏值的記錄。在分析中指定多項檢定時，會個別評估每項檢定。

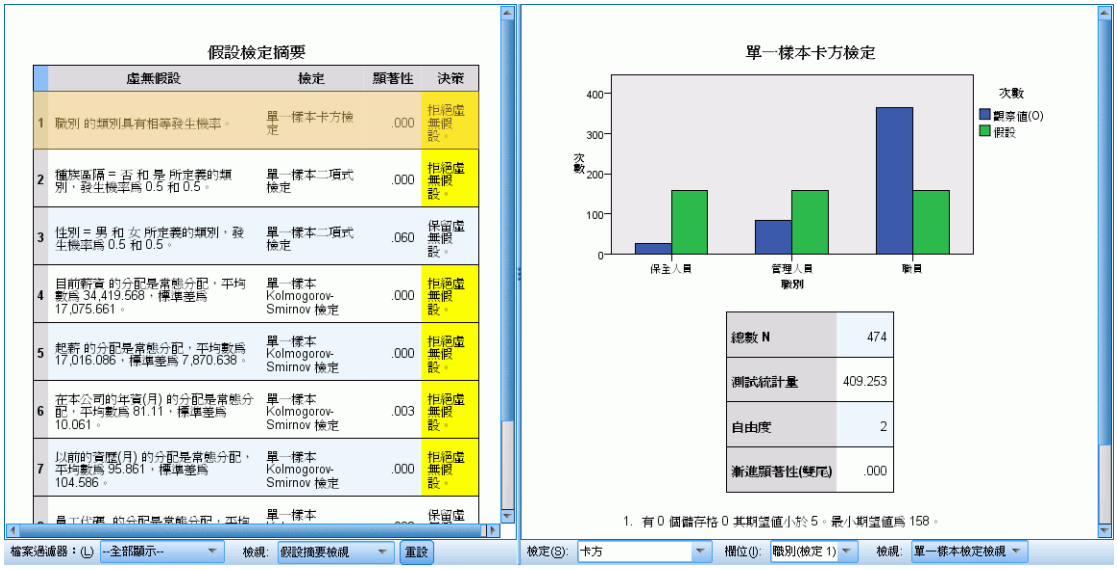
使用者遺漏值

圖表 27-21
相關樣本無母數檢定的「使用者遺漏值」設定

類別欄位的使用者遺漏值。類別欄位必須具有要納入分析之記錄的有效值。這些控制可讓您決定是否要在類別欄位中，將使用者遺漏值視為有效值。系統遺漏值和連續欄位的遺漏值會永遠視為無效。

模式檢視

圖表 27-22
無母數檢定模式檢視



此程序會在「瀏覽器」中建立「模式瀏覽器」物件。啟動（連按兩下）此物件，即可取得模式的互動檢視。模式檢視有 2 個面板視窗，主檢視位於左邊，連結或輔助檢視位於右邊。

主檢視有兩種：

- 假設摘要。此為預設檢視。
- 信賴區間摘要。

連結/輔助檢視有七種：

- 單一樣本檢定。這是要求單一樣本檢定時的預設檢視。
- 相關樣本檢定。這是要求相關樣本檢定和無單一樣本檢定時的預設檢視。

- 獨立樣本檢定。這是要求無相關樣本檢定或單一樣本檢定時的預設檢視。
- 類別欄位資訊。
- 連續欄位資訊。
- 成對比較。
- 同質子集。

假設摘要

圖表 27-23
假設摘要

假設檢定摘要

	虛無假設	檢定	顯著性	決策
1	職別 的類別具有相等發生機率。	單一樣本卡方檢定	.000	拒絕虛無假設。
2	種族區隔 = 否 和 是 所定義的類別，發生機率爲 0.5 和 0.5。	單一樣本二項式檢定	.000	拒絕虛無假設。
3	性別 = 男 和 女 所定義的類別，發生機率爲 0.5 和 0.5。	單一樣本二項式檢定	.060	保留虛無假設。
4	目前薪資 的分配是常態分配，平均數爲 34,419.568，標準差爲 17,075.661。	單一樣本 Kolmogorov-Smirnov 檢定	.000	拒絕虛無假設。
5	起薪 的分配是常態分配，平均數爲 17,016.086，標準差爲 7,870.638。	單一樣本 Kolmogorov-Smirnov 檢定	.000	拒絕虛無假設。
6	在本公司的年資(月) 的分配是常態分配，平均數爲 81.11，標準差爲 10.061。	單一樣本 Kolmogorov-Smirnov 檢定	.003	拒絕虛無假設。
7	以前的資歷(月) 的分配是常態分配，平均數爲 95.861，標準差爲 104.586。	單一樣本 Kolmogorov-Smirnov 檢定	.000	拒絕虛無假設。
8	員工代碼 的分配是常態分配，平均數爲 237.5，標準差爲 136.976。	單一樣本 Kolmogorov-Smirnov 檢定	.082	保留虛無假設。

已顯示精確顯著性。顯著水準爲 .05。

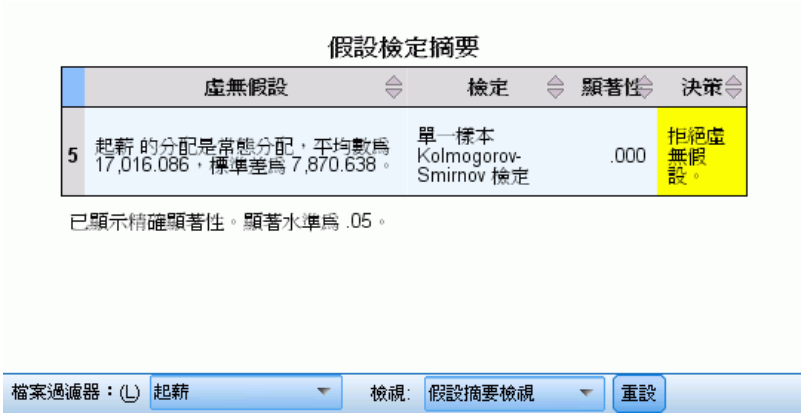
檔案過濾器：(L) --全部顯示-- 檢視: 假設摘要檢視 重設

「模式摘要」檢視是一種快照、無母數檢定的一覽摘要。它強調虛無假設和決定，並將重心放在顯著的 p 值。

- 每一列對應不同的檢定。在列上按一下可以在連結的檢視中顯示檢定的其他相關資訊。
- 在任何行標題上按一下便可按該行的值排序列。

- 「重設」按鈕可讓您將「模式瀏覽器」回復為原有狀態。
- 「欄位過濾器」下拉式清單可讓您只顯示與所選欄位相關的檢定。例如，當在「欄位過濾器」下拉式清單中選取「起薪」時，「假設摘要」中只會顯示兩種檢定。

圖表 27-24
假設摘要；依「起薪」過濾


假設檢定摘要

	虛無假設	檢定	顯著性	決策
5	起薪 的分配是常態分配，平均數為 17,016.086，標準差為 7,870.638。	單一樣本 Kolmogorov-Smirnov 檢定	.000	拒絕虛無假設。

已顯示精確顯著性。顯著水準為 .05。

檔案過濾器：(L) 起薪　檢視: 假設摘要檢視　重設

信賴區間摘要

圖表 27-25
信賴區間摘要


信賴區間摘要

信賴區間類型	參數	估計(E)	漸進 95% 信賴區間 下界	漸進 95% 信賴區間 上界
單一樣本二項式成功率 (Clopper-Pearson)	機率(性別=男)。	.544	.498	.590
單一樣本二項式成功率 (Jeffreys)	機率(性別=男)。	.544	.499	.589
單一樣本二項式成功率 (Profile Likelihood)	機率(性別=男)。	.544	.499	.589
單一樣本二項式成功率 (Clopper-Pearson)	機率(種族區隔=否)。	.781	.741	.817

檢視: 信賴區間摘要檢視　重設

「信賴區間摘要」會顯示無母數檢定所產生的任何信賴區間。

- 每一列對應不同的信賴區間。
- 在任何行標題上按一下便可按該行的值排序列。

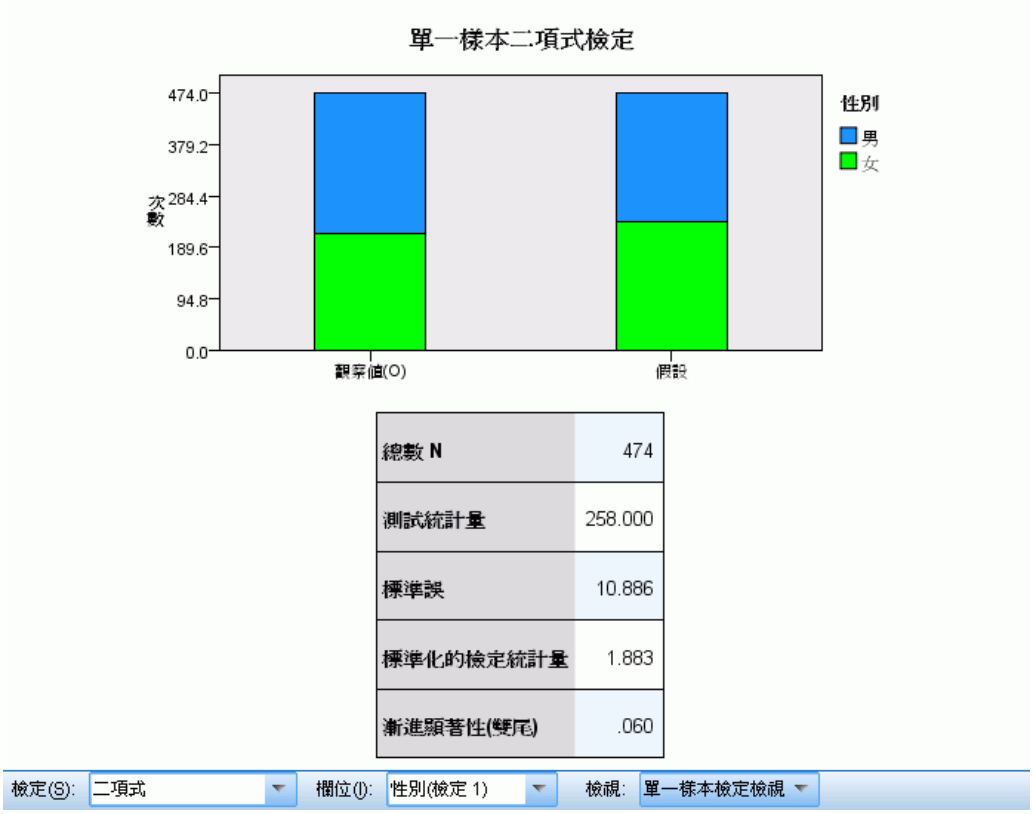
單一樣本檢定

「單一樣本檢定」檢視會顯示與任何所要求的單一樣本無母數檢定相關的詳細資料。顯示的資訊會視選取的檢定而有所不同。

- 「**檢定**」下拉式清單可讓您選取指定類型的單一樣本檢定。
- 您可以使用「**欄位**」下拉式清單，選取使用「**檢定**」下拉式清單中選取檢定進行檢定的欄位。

二項式檢定

圖表 27-26
單一樣本檢定檢視，二項式檢定

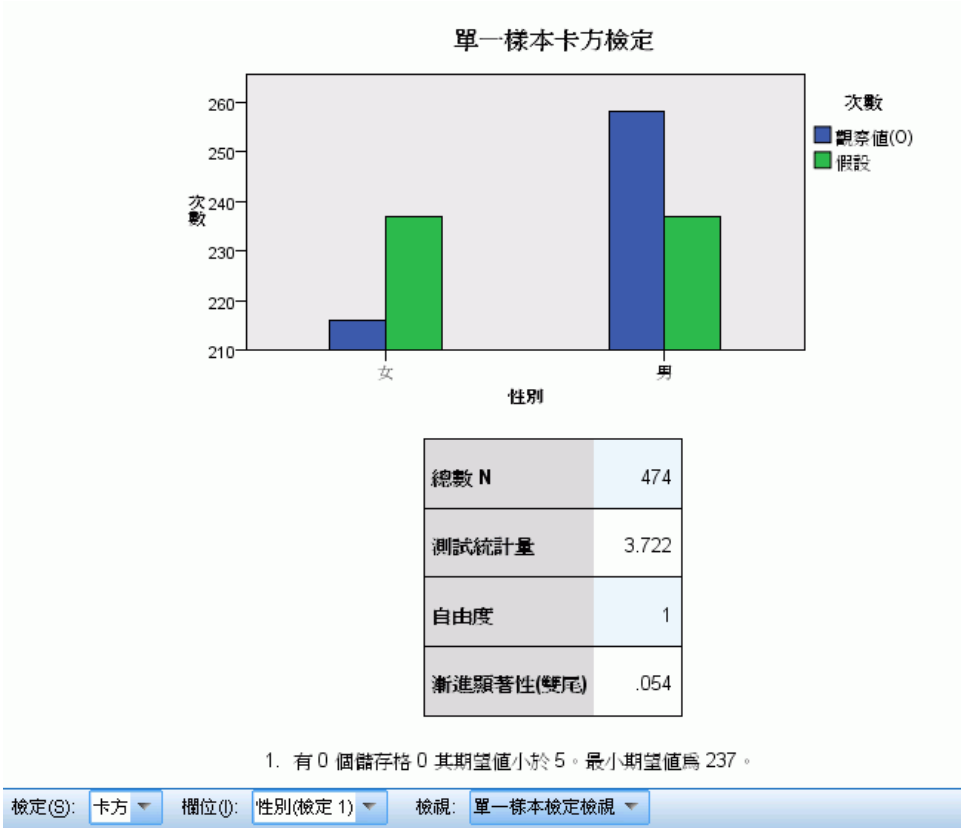


「二項式檢定」會顯示堆疊長條圖和檢定表格。

- 堆疊長條圖會顯示檢定欄位之「成功」和「失敗」類別的觀察與假設次數，圖中「失敗」會堆疊在「成功」上方。游標停留於長條上時，會以工具提示的方式顯示類別百分比。長條中的可見差異表示檢定欄位可能沒有假設的二項式分配。
- 表格會顯示檢定的詳細資料。

「卡方檢定」

圖表 27-27
單一樣本檢定檢視，卡方檢定



總數 N	474
測試統計量	3.722
自由度	1
漸進顯著性(雙尾)	.054

1. 有 0 個儲存格 0 其期望值小於 5。最小期望值為 237。

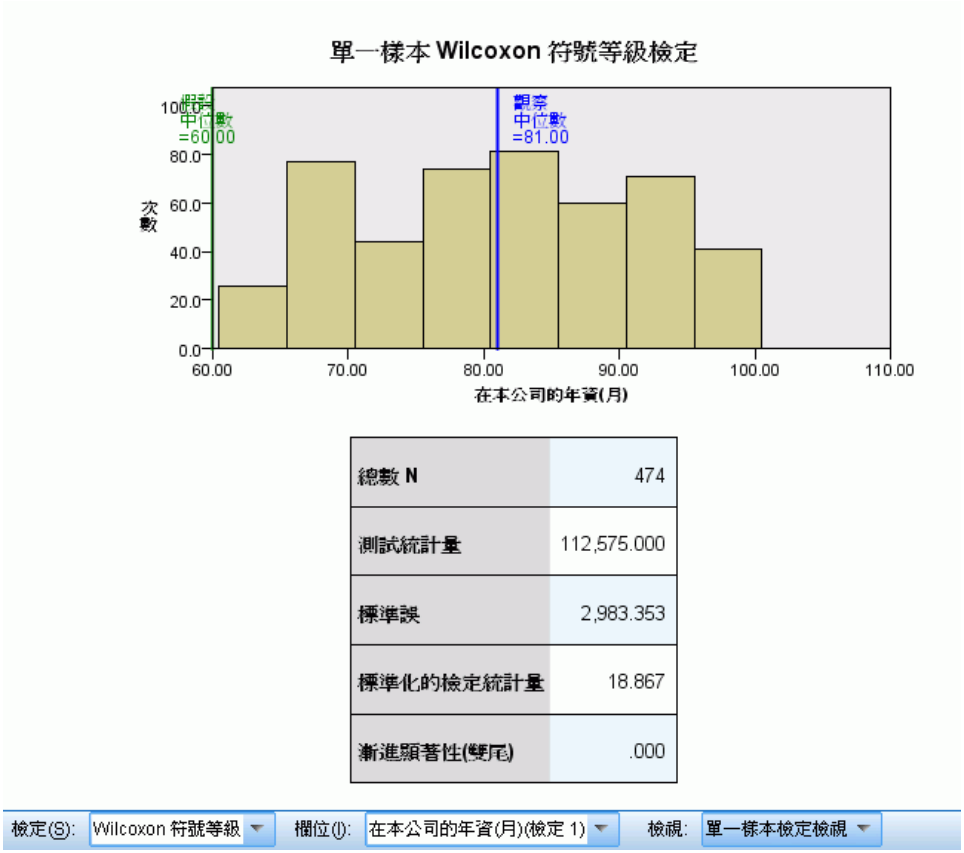
檢定(S): 卡方 | 欄位(I): 性別(檢定 1) | 檢視: 單一樣本檢定檢視

「卡方檢定」檢視會顯示集群長條圖和檢定表格。

- 集群長條圖會顯示檢定欄位各個類別的觀察與假設次數。游標停留於長條上時，會以工具提示的方式顯示觀察和假設次數及其差異（殘差）。觀察與假設長條中的可見差異表示檢定欄位可能沒有假設的分配。
- 表格會顯示檢定的詳細資料。

Wilcoxon 符號等級

圖表 27-28
單一樣本檢定檢視，Wilcoxon 符號等級檢定

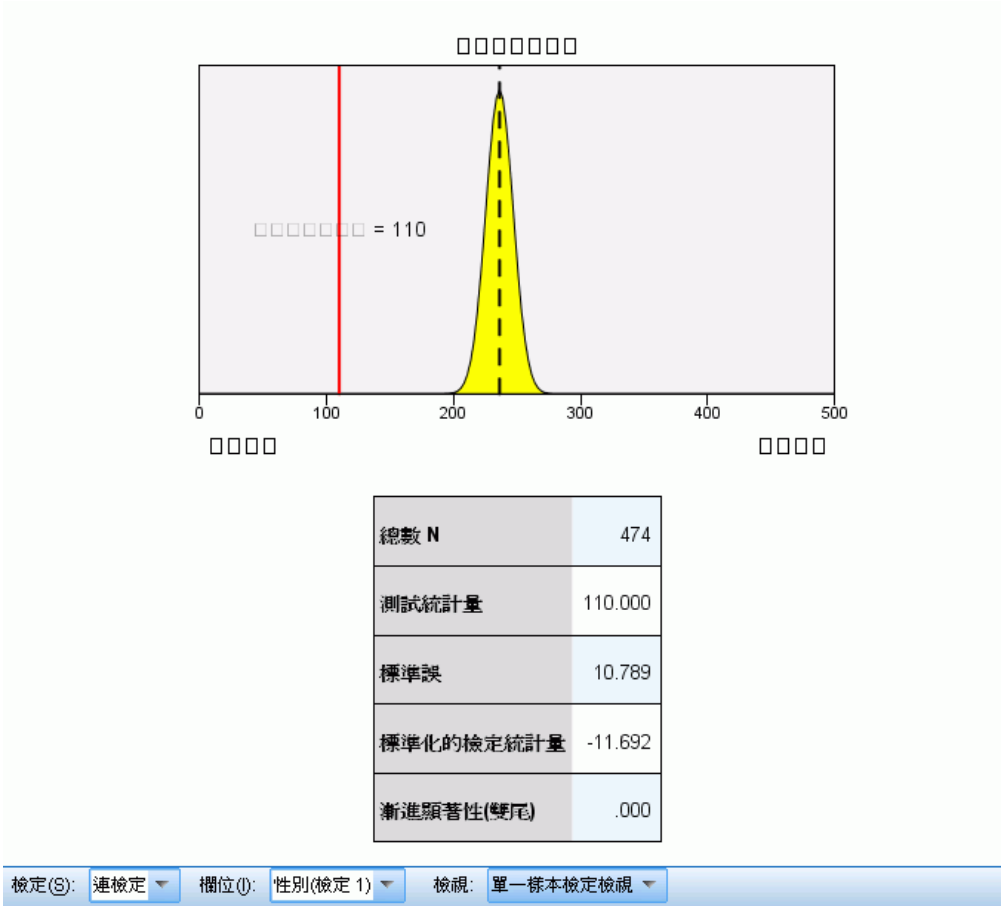


「Wilcoxon 符號等級檢定」檢視會顯示直方圖和檢定表格。

- 直方圖包括顯示觀察和假設中位數的垂直線。
- 表格會顯示檢定的詳細資料。

連檢定

圖表 27-29
單一樣本檢定檢視，連檢定



總數 N	474
測試統計量	110.000
標準誤	10.789
標準化的檢定統計量	-11.692
漸進顯著性(雙尾)	.000

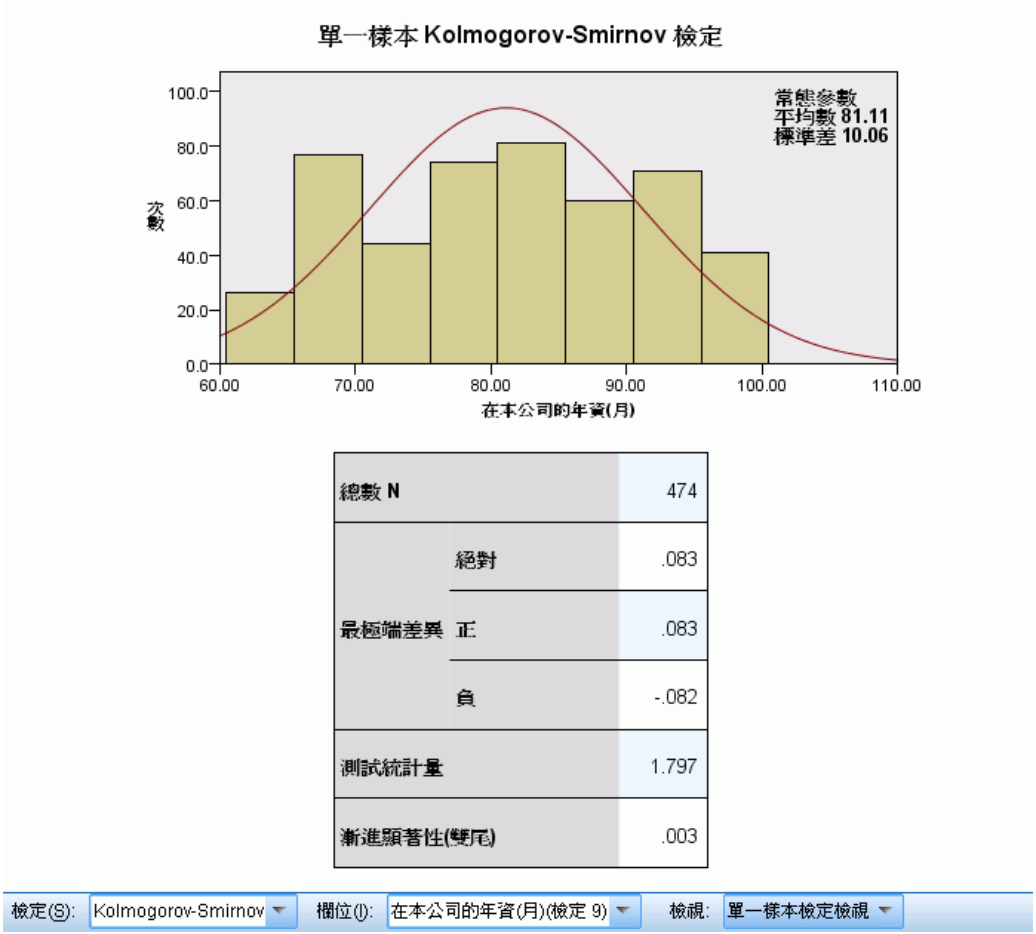
檢定(S): 連檢定 欄位(I): 性別(檢定 1) 檢視: 單一樣本檢定檢視

「連檢定」檢視會顯示圖表和檢定表格。

- 圖表會顯示觀察到的連個數的常態分配，而連個數會以垂直線作標記。請注意，當執行精確檢定時，檢定並不是以常態分配為基礎。
- 表格會顯示檢定的詳細資料。

Kolmogorov-Smirnov 檢定

圖表 27-30
單一樣本檢定檢視，Kolmogorov-Smirnov 檢定



「Kolmogorov-Smirnov 檢定」檢視會顯示直方圖和檢定表格。

- 直方圖包括假設的均勻、常態、Poisson 或指數分配之機率密度函數的重疊圖。請注意，檢定是以累積分配為基礎，而表格中所報告的「最極端差異」應以累積分配的觀點來解釋。
- 表格會顯示檢定的詳細資料。

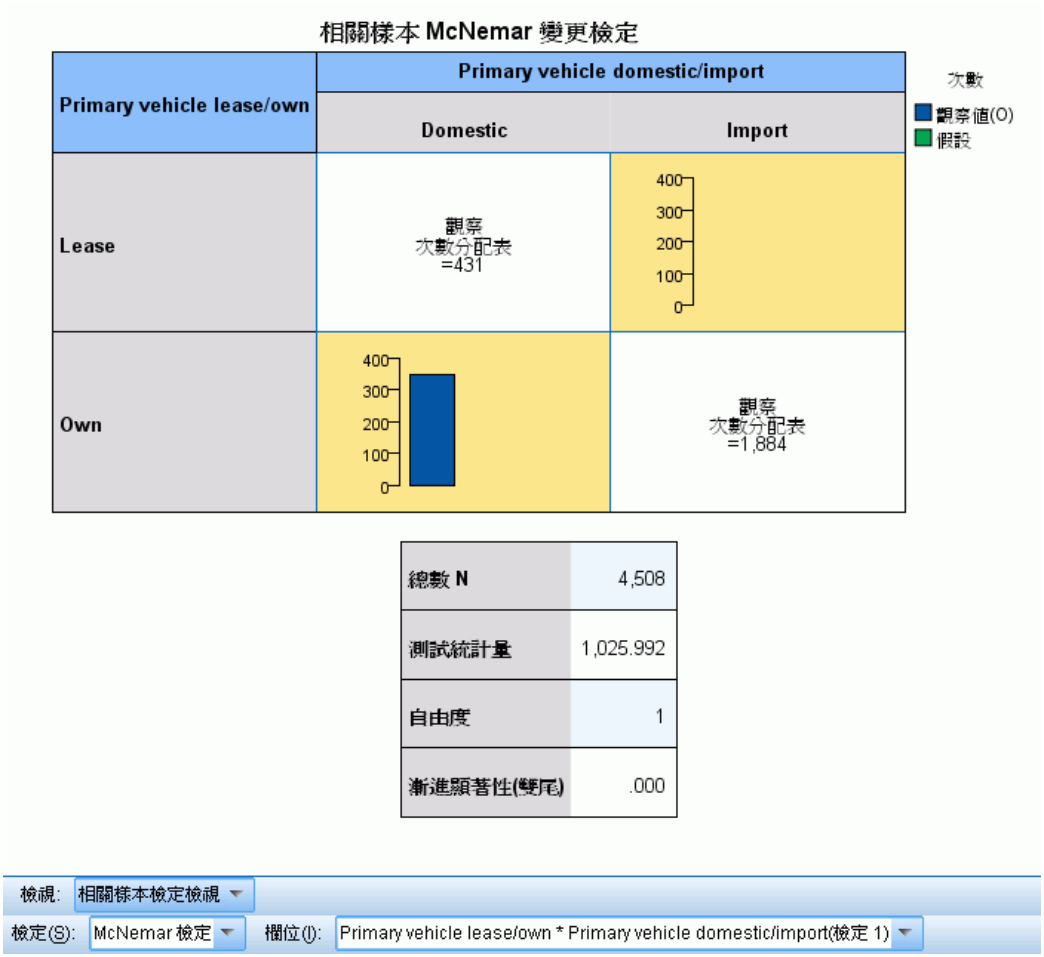
相關樣本檢定

「單一樣本檢定」檢視會顯示與任何所要求的單一樣本無母數檢定相關的詳細資料。顯示的資訊會視選取的檢定而有所不同。

- 「**檢定**」下拉式清單可讓您選取指定類型的單一樣本檢定。
- 您可以使用「**欄位**」下拉式清單，選取使用「**檢定**」下拉式清單中選取檢定進行檢定的欄位。

McNemar 檢定

圖表 27-31
相關樣本檢定檢視，McNemar 檢定

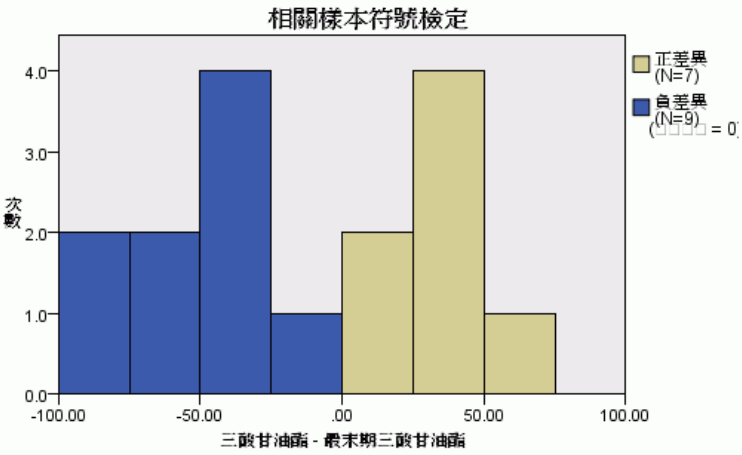


「McNemar 檢定」檢視會顯示集群長條圖和檢定表格。

- 集群長條圖會顯示檢定欄位所定義之 2×2 表格中對角線外儲存格的觀察和假設次數。
- 表格會顯示檢定的詳細資料。

符號檢定

圖表 27-32
相關樣本檢定檢視，符號檢定



總數 N	16
測試統計量	7.000
標準誤	2.000
標準化的檢定統計量	-.250
漸進顯著性(雙尾)	.803
精確顯著性(雙尾)	.804

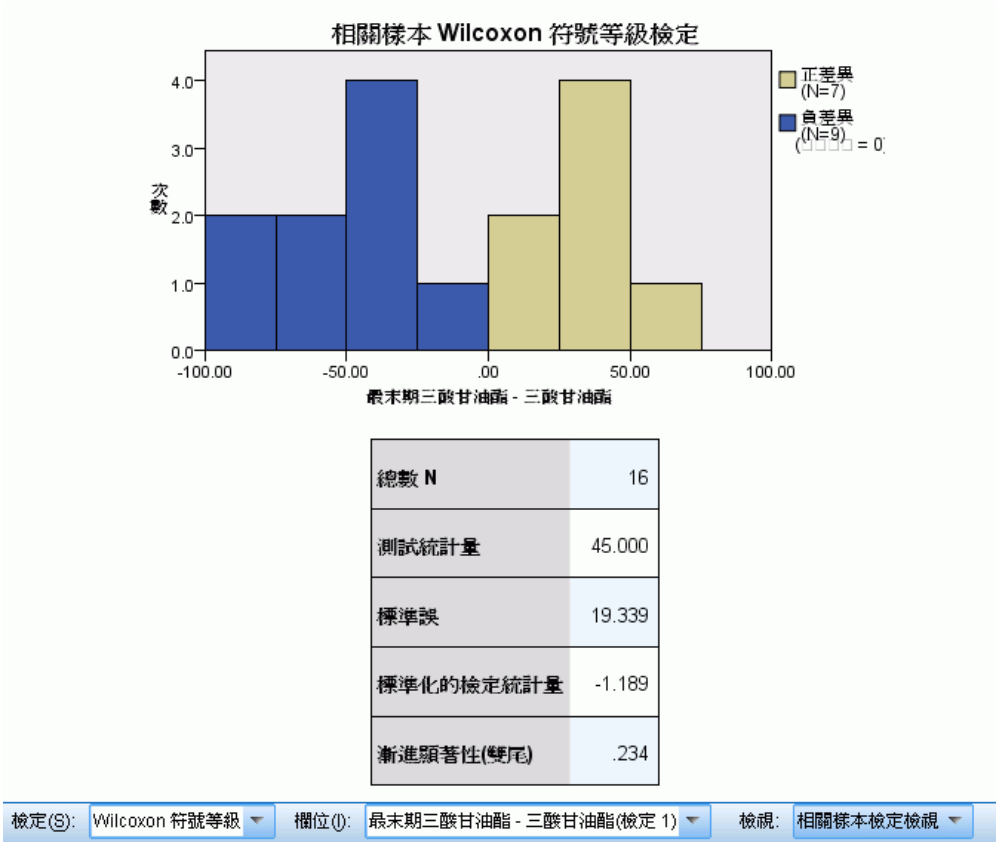
檢定(S): 符號檢定 | 欄位(I): 三酸甘油酯 - 最末期三酸甘油酯(檢定 1) | 檢視: 相關樣本檢定檢視

「符號檢定」檢視會顯示堆疊直方圖和檢定表格。

- 堆疊直方圖會顯示欄位之間的差異，並使用差異的符號當作堆疊欄位。
- 表格會顯示檢定的詳細資料。

Wilcoxon 符號等級檢定

圖表 27-33
相關樣本檢定檢視，Wilcoxon 符號等級檢定



總數 N	16
測試統計量	45.000
標準誤	19.339
標準化的檢定統計量	-1.189
漸進顯著性(雙尾)	.234

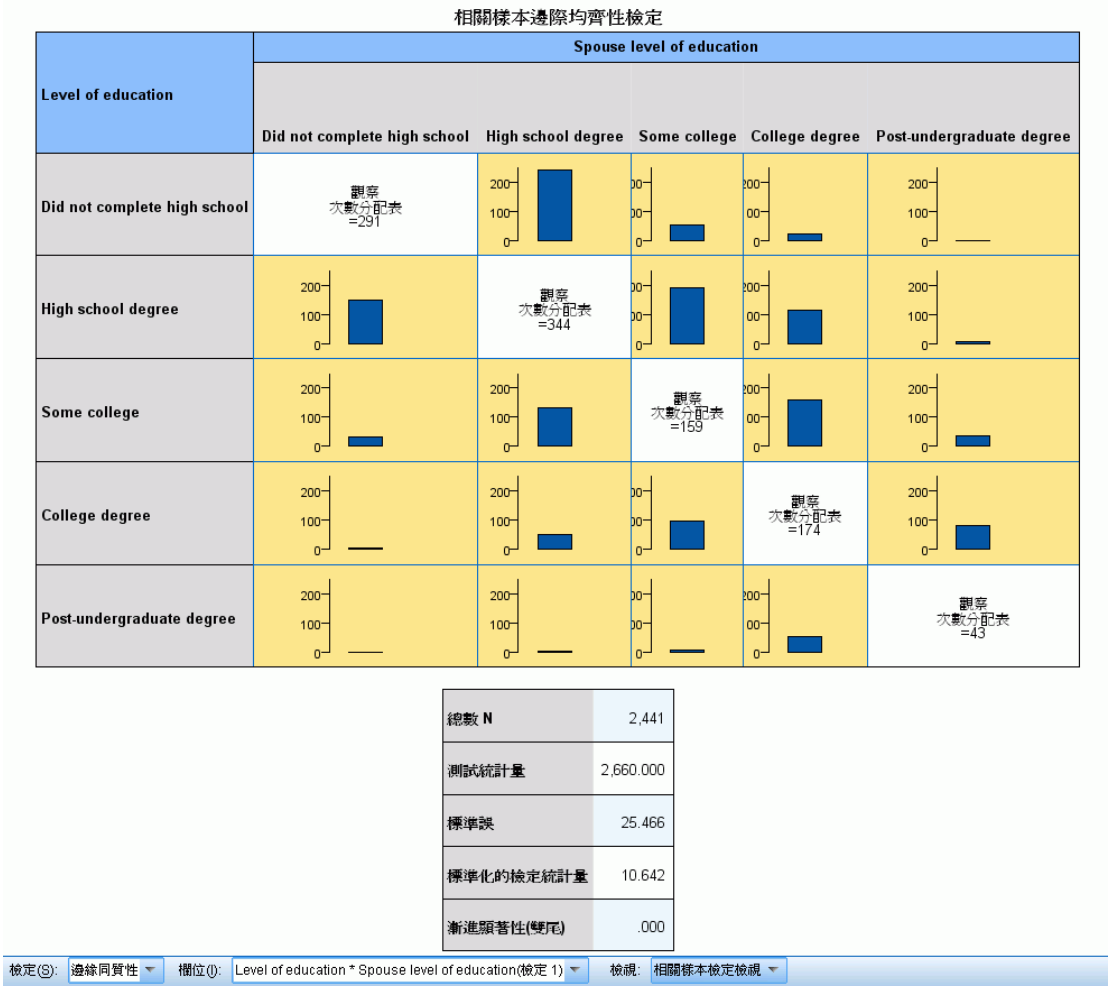
檢定(S): Wilcoxon 符號等級 | 欄位(I): 最末期三酸甘油酯 - 三酸甘油酯(檢定 1) | 檢視: 相關樣本檢定檢視

「Wilcoxon 符號等級檢定」檢視會顯示堆疊直方圖和檢定表格。

- 堆疊直方圖會顯示欄位之間的差異，並使用差異的符號當作堆疊欄位。
- 表格會顯示檢定的詳細資料。

邊際均齊性檢定

圖表 27-34
相關樣本檢定檢視，邊際均齊性檢定

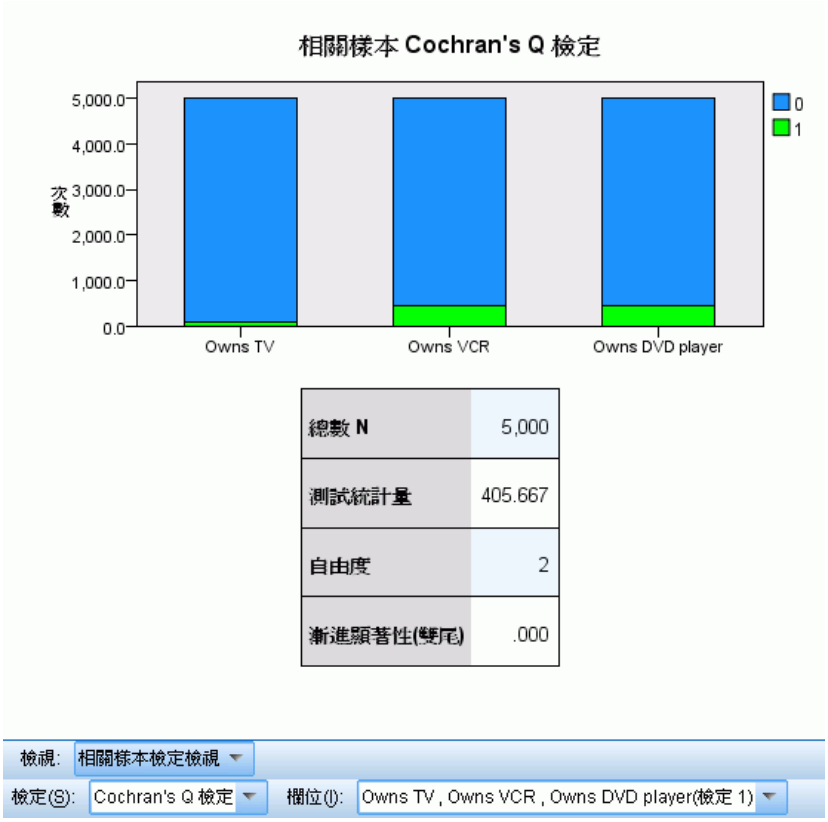


「邊際均齊性檢定」檢視會顯示集群長條圖和檢定表格。

- 集群長條圖會顯示檢定欄位所定義之表格中對角線外儲存格的觀察次數。
- 表格會顯示檢定的詳細資料。

Cochran’s Q 檢定

圖表 27-35
相關樣本檢定檢視，Cochran’s Q 檢定



總數 N	5,000
測試統計量	405.667
自由度	2
漸進顯著性(雙尾)	.000

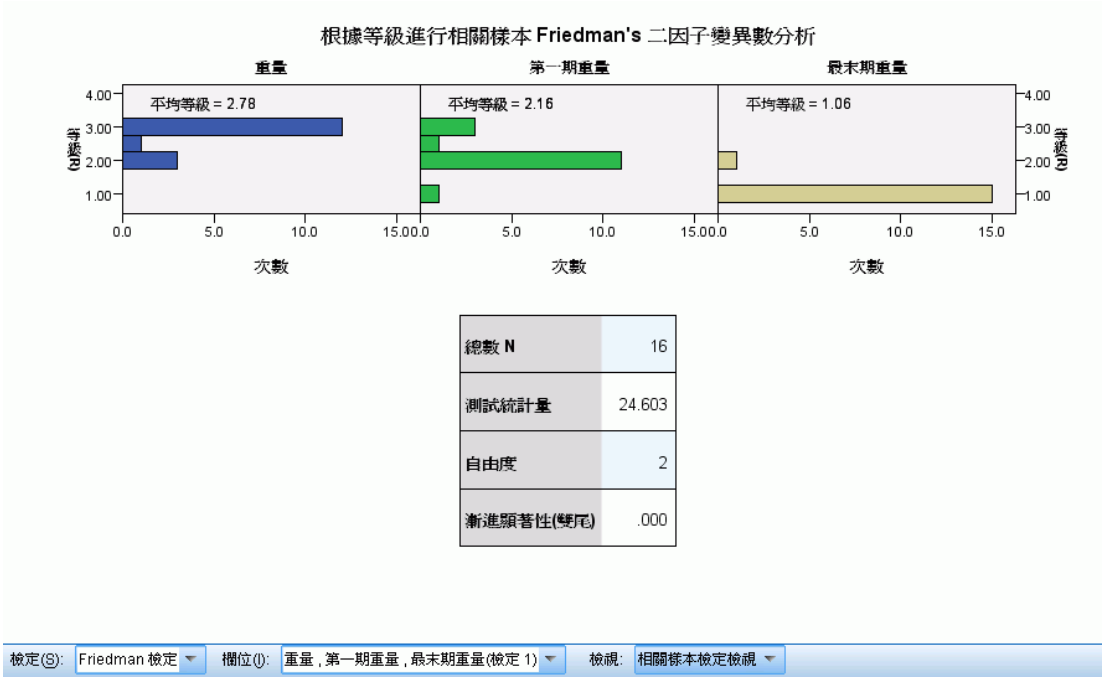
檢視: 相關樣本檢定檢視
檢定(S): Cochran's Q 檢定
欄位(I): Owns TV , Owns VCR , Owns DVD player(檢定 1)

「Cochran’s Q 檢定」檢視會顯示堆疊長條圖和檢定表格。

- 堆疊長條圖會顯示檢定欄位之「成功」和「失敗」類別的觀察次數，圖中「失敗」會堆疊在「成功」上方。游標停留於長條上時，會以工具提示的方式顯示類別百分比。
- 表格會顯示檢定的詳細資料。

根據等級進行 Friedman 二因子變異數分析

圖表 27-36
相關樣本檢定檢視，根據等級進行 Friedman 二因子變異數分析

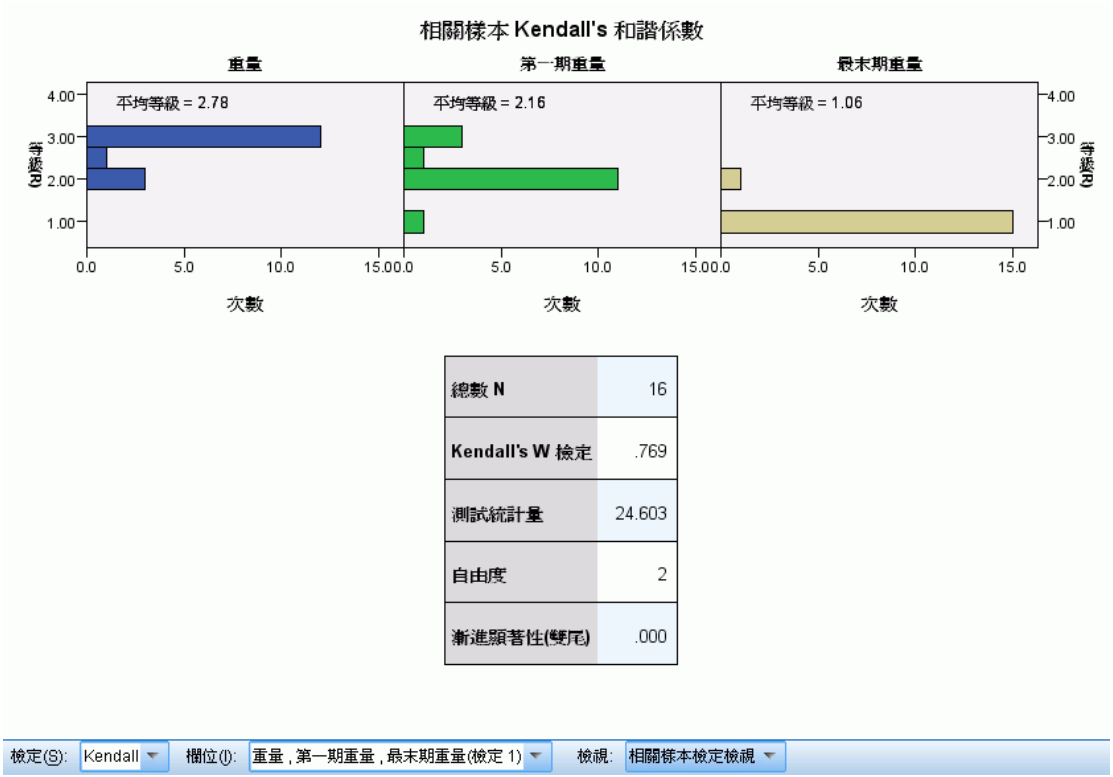


「根據等級進行 Friedman 二因子變異數分析」檢視會顯示面板化的直方圖和檢定表格。

- 直方圖會顯示等級的觀察分配，並依據檢定欄位來面板化。
- 表格會顯示檢定的詳細資料。

Kendall 和諧係數

圖表 27-37
相關樣本檢定檢視，Kendall 和諧係數



「Kendall 和諧係數」檢視會顯示面板化的直方圖和檢定表格。

- 直方圖會顯示等級的觀察分配，並依據檢定欄位來面板化。
- 表格會顯示檢定的詳細資料。

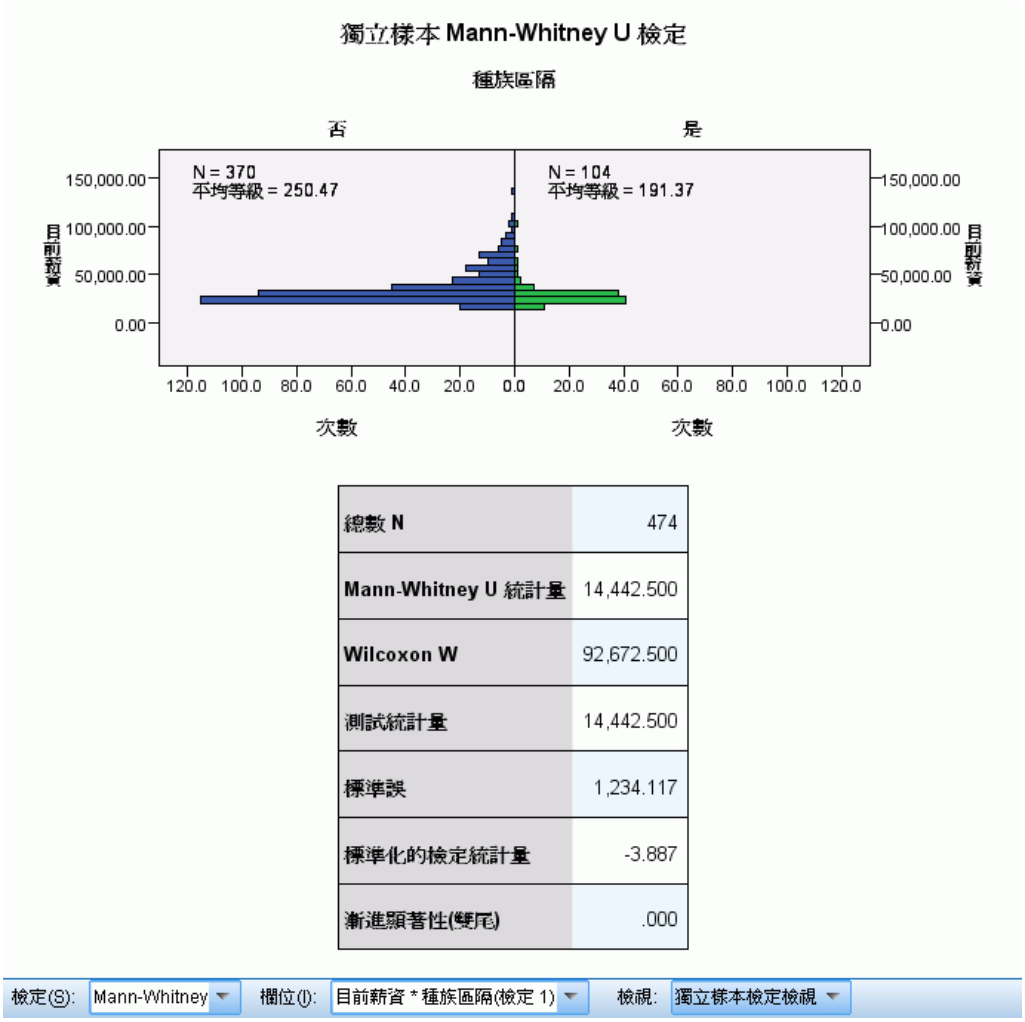
獨立樣本檢定

「獨立樣本檢定」檢視會顯示與任何所要求的獨立樣本無母數檢定相關的詳細資料。顯示的資訊會視選取的檢定而有所不同。

- 「**檢定**」下拉式清單可讓您選取指定類型的獨立樣本檢定。
- 您可以使用「**欄位**」下拉式清單，選取使用「**檢定**」下拉式清單中選取檢定來進行檢定的檢定與分組欄位組合。

Mann-Whitney 檢定

圖表 27-38
獨立樣本檢定檢視，Mann-Whitney 檢定



總數 N	474
Mann-Whitney U 統計量	14,442.500
Wilcoxon W	92,672.500
測試統計量	14,442.500
標準誤	1,234.117
標準化的檢定統計量	-3.887
漸進顯著性(雙尾)	.000

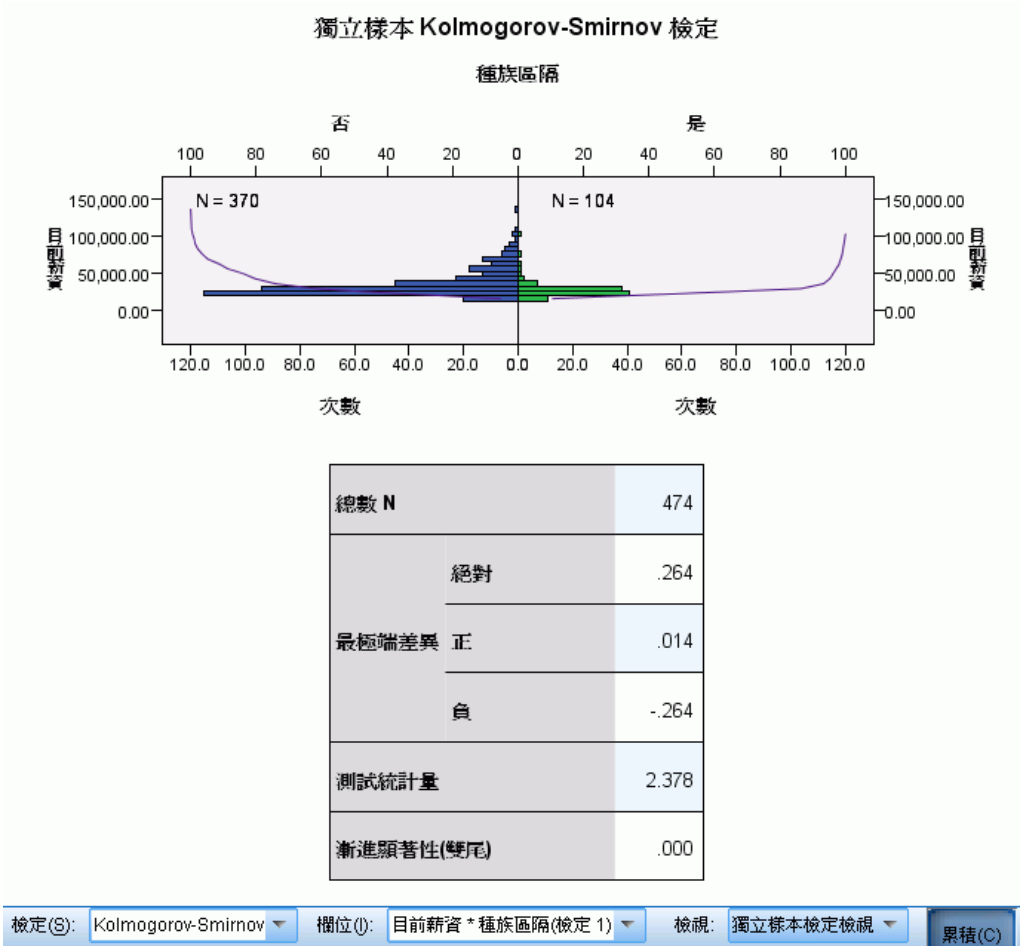
檢定(S): Mann-Whitney 欄位(I): 目前薪資 * 種族區隔(檢定 1) 檢視: 獨立樣本檢定檢視

「Mann-Whitney 檢定」檢視會顯示人口金字塔圖和檢定表格。

- 人口金字塔圖會依據分組欄位的類別來顯示連續的直方圖，並加註每個組別中的記錄數目和組別的平均等級。
- 表格會顯示檢定的詳細資料。

Kolmogorov-Smirnov 檢定

圖表 27-39
獨立樣本檢定檢視，Kolmogorov-Smirnov 檢定



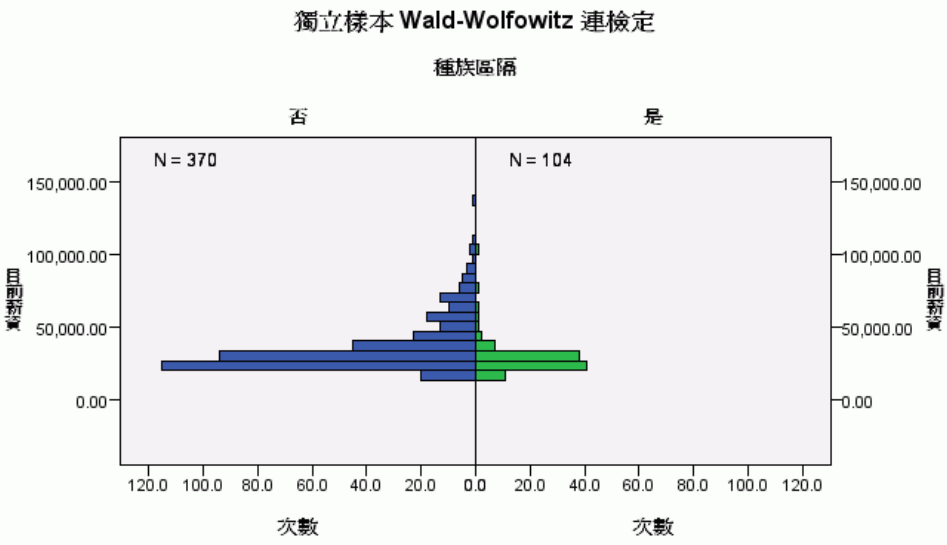
總數 N		474
最極端差異	絕對	.264
	正	.014
	負	-.264
測試統計量		2.378
漸進顯著性(雙尾)		.000

「Kolmogorov-Smirnov 檢定」檢視會顯示人口金字塔圖和檢定表格。

- 人口金字塔圖會依據分組欄位的類別來顯示背對背的直方圖，並加註每個組別中的記錄數目。按一下「**累積**」按鈕便可顯示或隱藏觀察累積分配線。
- 表格會顯示檢定的詳細資料。

Wald-Wolfowitz 連檢定

圖表 27-40
獨立樣本檢定檢視，Wald-Wolfowitz 連檢定



總數 N		474
可能最小值	測試統計量[1]	97.000
	標準誤	7.442
	標準化的檢定統計量	-8.917
	漸進顯著性(雙尾)	.000
可能最大值	測試統計量[1]	199.000
	標準誤	7.442
	標準化的檢定統計量	4.788
	漸進顯著性(雙尾)	1.000

[1]The test statistic is the number of runs.

1. There are 55 inter-group ties involving 228 records.

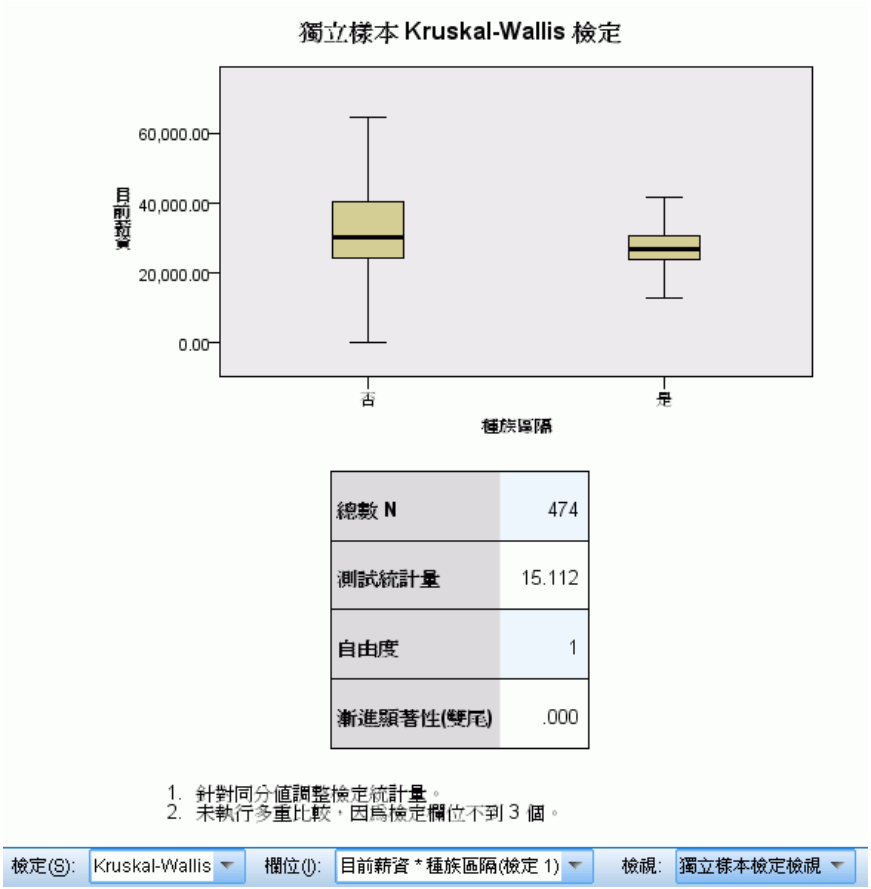
檢定(S): Wald-Wolfowitz 欄位(I): 目前薪資 * 種族區隔(檢定 1) 檢視: 獨立樣本檢定檢視

「Wald-Wolfowitz 連檢定」檢視會顯示堆疊長條圖和檢定表格。

- 人口金字塔圖會依據分組欄位的類別來顯示背對背的直方圖，並加註每個組別中的記錄數目。
- 表格會顯示檢定的詳細資料。

Kruskal-Wallis 檢定

圖表 27-41
獨立樣本檢定檢視，Kruskal-Wallis 檢定

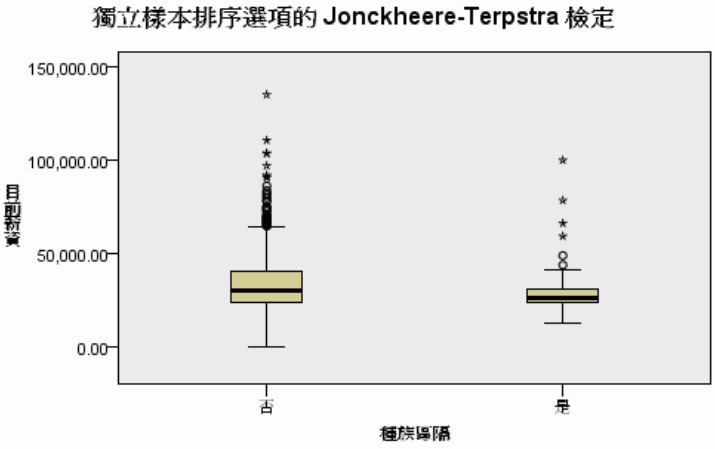


「Kruskal-Wallis 檢定」檢視會顯示盒形圖和檢定表格。

- 針對分組欄位的每個類別顯示不同的盒形圖。游標停留於方塊上時，會以工具提示的方式顯示平均等級。
- 表格會顯示檢定的詳細資料。

Jonckheere-Terpstra 檢定

圖表 27-42
獨立樣本檢定檢視，Jonckheere-Terpstra 檢定



總數 N	474
測試統計量	14,442.500
標準誤	1,234.117
標準化的檢定統計量	-3.887
漸進顯著性(雙尾)	.000

1. 未執行多重比較，因為檢定欄位不到 3 個。

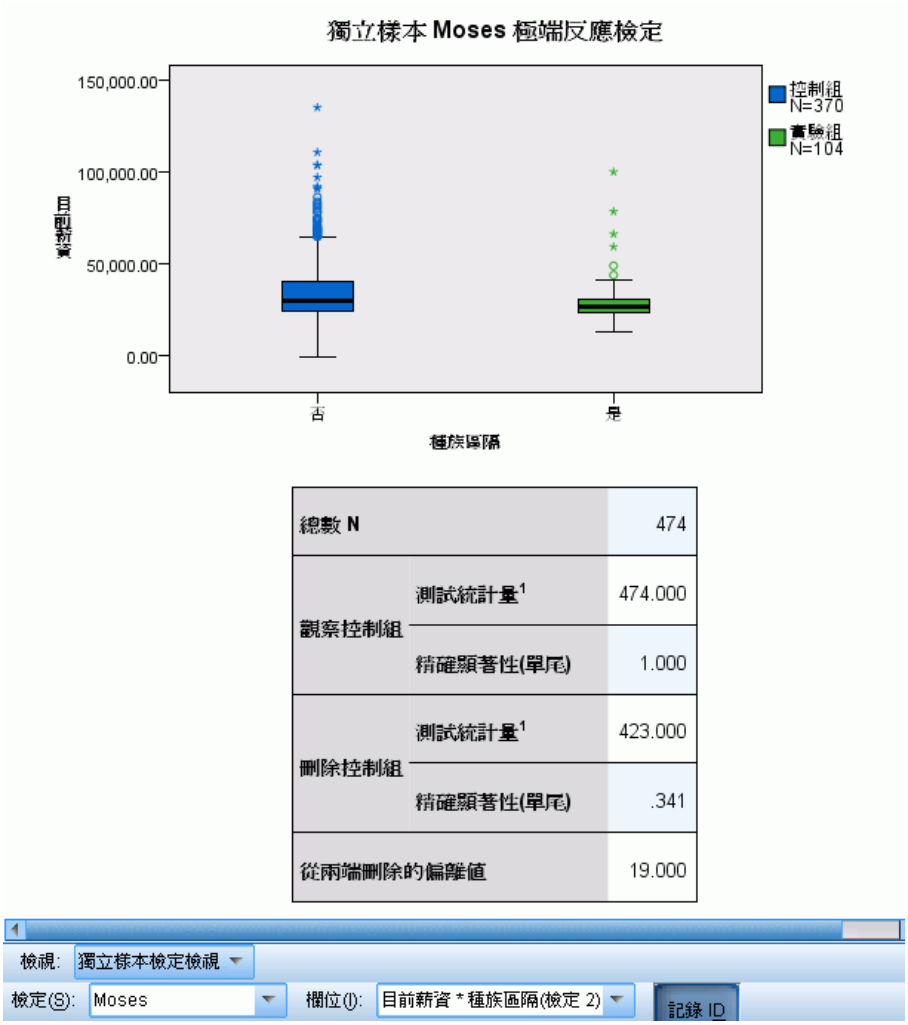
檢定(S): Jonkheere-Terpstra　欄位(I): 目前薪資 * 種族區隔(檢定 1)　檢視: 獨立樣本檢定檢視

「Jonckheere-Terpstra 檢定」檢視會顯示盒形圖和檢定表格。

- 針對分組欄位的每個類別顯示不同的盒形圖。
- 表格會顯示檢定的詳細資料。

Moses 極端反應檢定

圖表 27-43
獨立樣本檢定檢視，Moses 極端反應檢定



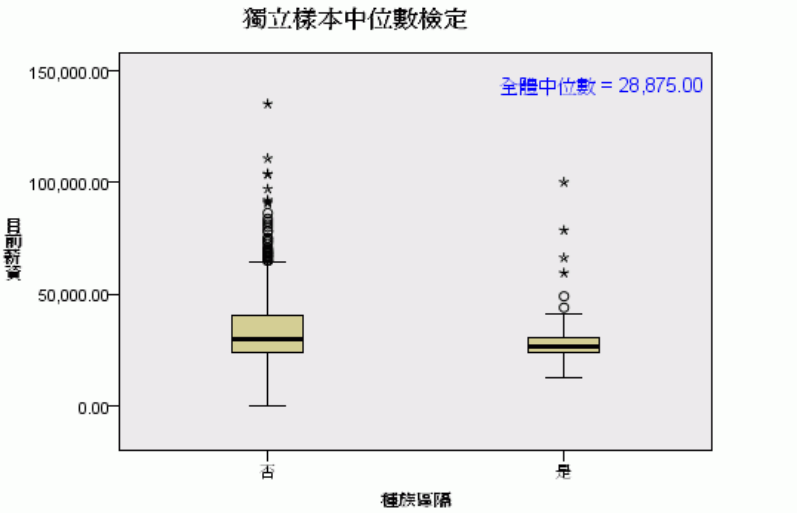
總數 N		474
觀察控制組	測試統計量[1]	474.000
	精確顯著性(單尾)	1.000
刪除控制組	測試統計量[1]	423.000
	精確顯著性(單尾)	.341
從兩端刪除的偏離值		19.000

「Moses 極端反應檢定」檢視會顯示盒形圖和檢定表格。

- 針對分組欄位的每個類別顯示不同的盒形圖。按一下「記錄 ID」按鈕可顯示或隱藏點標記。
- 表格會顯示檢定的詳細資料。

中位數檢定

圖表 27-44
獨立樣本檢定檢視，中位數檢定



總數 N		474
中位數		28,875.000
測試統計量		14.240
自由度		1
漸進顯著性(雙尾)		.000
Yates 連續修正	卡方	13.414
	自由度	1
	漸進顯著性(雙尾)	.000

1. 未執行多重比較，因為檢定欄位不到 3 個。

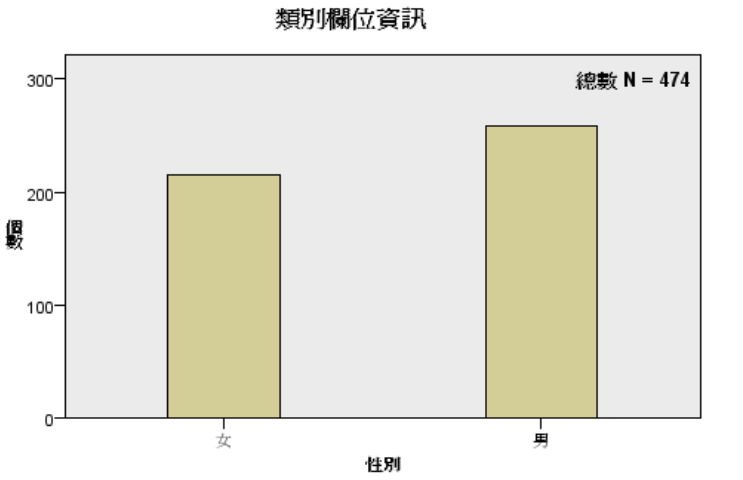
檢定(S): 中位數 | 欄位(I): 目前薪資 * 種族區隔(檢定 3) | 檢視: 獨立樣本檢定檢視

「中位數檢定」檢視會顯示盒形圖和檢定表格。

- 針對分組欄位的每個類別顯示不同的盒形圖。
- 表格會顯示檢定的詳細資料。

類別欄位資訊

圖表 27-45
類別欄位資訊

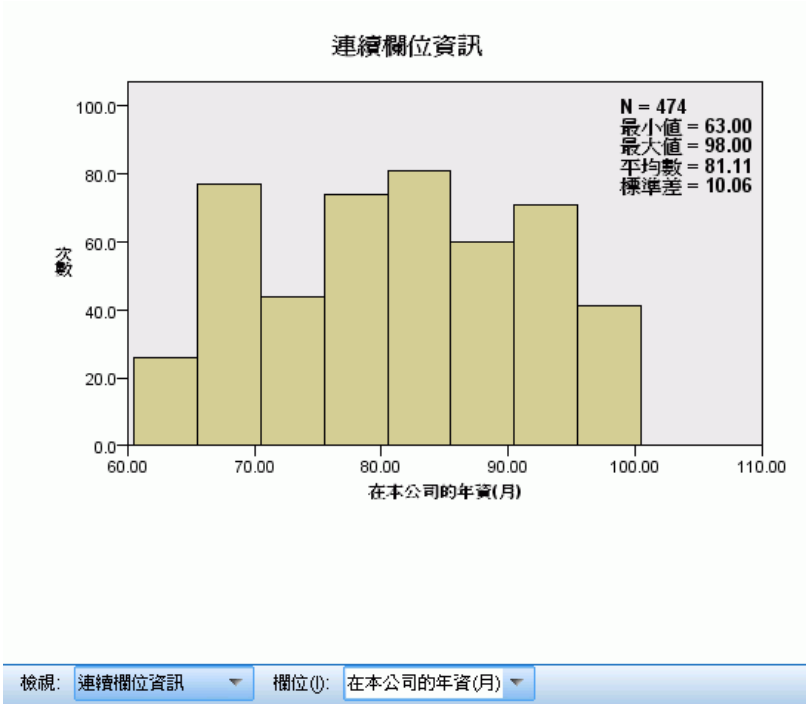


「類別欄位資訊」檢視會顯示「**欄位**」下拉式清單中選取類別欄位的盒形圖。可用欄位清單限定為「假設摘要」檢視中目前選取檢定所使用的類別欄位。

- 游標停留於長條上時，會以工具提示的方式顯示類別百分比。

連續欄位資訊

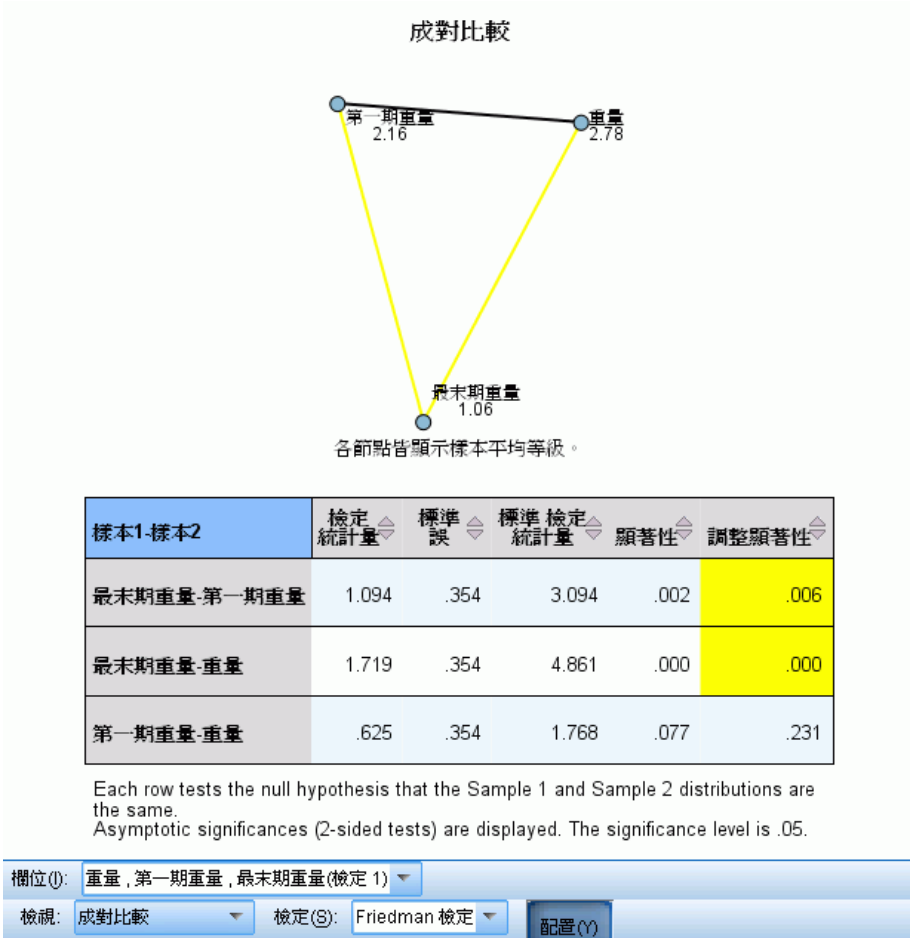
圖表 27-46
連續欄位資訊



「連續欄位資訊」檢視會顯示「**欄位**」下拉式清單中選取連續欄位的直方圖。可用欄位清單限定為「假設摘要」檢視中目前選取檢定所使用的連續欄位。

成對比較

圖表 27-47
成對比較



樣本1-樣本2	檢定統計量	標準誤	標準 檢定統計量	顯著性	調整顯著性
最末期重量-第一期重量	1.094	.354	3.094	.002	.006
最末期重量-重量	1.719	.354	4.861	.000	.000
第一期重量-重量	.625	.354	1.768	.077	.231

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

「成對比較」檢視會顯示距離網路圖，以及當要求成對多重比較時，由 k 樣本無母數檢定所產生的比較表。

- 距離網路圖是以圖形表示的比較表，其中網路中節點之間的距離對應至樣本之間的差異。黃線對應至統計上的顯著差異；黑線對應至非顯著差異。游標停留於網路中的線上時，會以工具提示的方式顯示線所連接節點之間差異的調整顯著性。
- 比較表會顯示所有成對比較的數值結果。每一列對應不同的成對比較。在行標題上按一下便可按該行的值排序列。

同質子集

圖表 27-48
同質子集

同質子集

		子集		
		1	2	3
樣本[1]	最末期重量	1.063		
	第一期重量		2.156	
	重量			2.781
測試統計量		.[2]	.[2]	.[2]
顯著性(2 尾)		.	.	.
調整顯著性(2 尾)		.	.	.

同質子集以漸進顯著性爲基礎。顯著水準爲 .05。

[1]各儲存格皆顯示樣本平均等級。

[2]Unable to compute because the subset contains only one sample.

欄位(I): 重量 , 第一期重量 , 最末期重量(檢定 1)
檢視: 同質子集 檢定(S): Kendall

「同質子集」檢視會顯示當要求逐步減期多重比較時，由 k 樣本無母數檢定所產生的比較表。

- 「樣本」組別中的每一列對應至不同的相關樣本（資料中以不同的欄位表示）。統計上沒有顯著差異的樣本分成顏色相同的子集；每個已識別的子集位於不同行。如果所有樣本統計上皆有顯著差異，每個樣本則有不同的子集。如果沒有樣本有統計上的顯著差異，則只有單一子集。
- 然後針對每個含有一個以上樣本的子集，計算其檢定統計量、顯著值和調整顯著值。

NPTESTS 指令的其他功能

指令語法語言也可以讓您：

- 在單一程序的執行中指定單一樣本檢定、獨立樣本檢定以及相關樣本檢定。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

原有對話方塊

有許多「原有」對話方塊也會執行無母數檢定。這些對話方塊支援「精確檢定」選項所提供的功能。

卡方檢定。依照類別製作變數的表格，並以觀察和期望次數間的差異為基礎，計算卡方統計量。

二項式檢定。將二分變數中各類別的觀察次數，與二項式分配的期望次數作一比較。

連檢定。檢定一變數的兩數值發生順序是否為隨機。

單一樣本 Kolmogorov-Smirnov 檢定。變數的觀察累積分配函數，與指定理論分配作一比較。後者可能為常態分配、均勻分配、指數分配或 Poisson 機率分配。

兩個獨立樣本檢定。比較某個變數上的兩組觀察值。在此您可以使用的檢定，包括 Mann-Whitney U 統計量檢定、兩個樣本 Kolmogorov-Smirnov 檢定、Moses 極端反應檢定，以及 Wald-Wolfowitz 連檢定。

兩個相關樣本檢定。比較兩個變數的分配。在此您可以使用的檢定，包括 Wilcoxon 符號等級檢定、符號檢定，以及 McNemar 檢定。

K 個獨立樣本檢定。比較某個變數上的兩組或兩組以上的觀察值。在此您可以使用的檢定，包括 Kruskal-Wallis 檢定、「中位數」檢定，以及 Jonckheere-Terpstra 檢定。

K 個相關樣本檢定。比較兩個或兩個以上變數的分配。在此您可以使用的檢定，包括 Friedman 檢定、Kendall’s W 檢定，以及 Cochran’s Q 檢定。

以上所有檢定都可以使用四分位數和平均數、標準差、最小值、最大值，以及非遺漏觀察值的個數。

「卡方檢定」

「卡方檢定」程序，會依照類別將變數製成表格，並計算卡方統計量。這種適合度檢定，會比較各類別中的觀察和期望次數，以檢定是否所有類別都包含相同的數值比例，或各類別都包含使用者指定的數值比例。

範例。以一袋軟糖為例。 我們可以使用卡方檢定來判斷，這袋軟糖中所包含的各色糖果比例是否相同 (這些顏色包括藍色、棕色、綠色、橙色、紅色和黃色)。或者，您也可以用來檢定，以便查看袋中軟糖的顏色比例是否為 5% 藍色、30% 棕色、10% 綠色、20% 橙色、15% 紅色和 15% 黃色。

統計量。包括平均數、標準差、最小值、最大值和四分位數。非遺漏和遺漏觀察值的個數與百分比、各類別所觀察和期望的觀察值個數、殘差和卡方統計量。

資料。請使用已排列順序或未排列順序的數值類別變數 (次序或名義測量水準)。若要將字串變數轉換成數字變數，請使用「轉換」功能表上的「自動重新編碼」程序。

假設。 無母數檢定不需要有關底線分配類型的假設。因為資料都已先假設為隨機樣本。各類別的期望次數應該至少為1，而且不應有超過 20% 的類別有少於 5 的期望次數。

若要取得卡方檢定

▶ 從功能表選擇：
分析(A) > 無母數檢定 > 歷史對話記錄 > 卡方...

圖表 27-49
「卡方檢定」對話方塊

▶ 選取一個或多個檢定變數。每個變數會產生不同的檢定。

▶ 或者，按一下「選項」選取敘述統計、四分位數，還可以決定 SPSS 該如何處理遺漏資料。

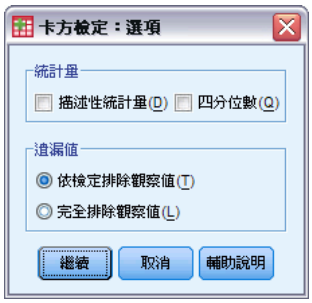
卡方檢定的期望範圍和期望值

期望範圍。依照預設值，變數的每個不同值會被定義為類別。若要在指定的範圍內建立類別，請按一下**使用指定的範圍**，並輸入下限或上限的整數值。各整數值會在所包含的範圍內建立類別，而其數值在範圍之外的觀察值則予以排除。例如，如果您指定的下限值為 1 而上限值為 4，那麼卡方檢定將只會使用介於 1 到 4 之間的整數值。

期望值。依照預設值，所有類別的期望值都相等。此外，類別也可以有使用者指定的期望比例。選擇**數值**，並為檢定變數的每個類別輸入大於 0 的值，然後按一下「**新增**」。每次您新增數值時，該值便會顯示在數值清單的底部。數值的順序非常重要，因為它會對應到檢定變數之類別數值的遞增順序。其中，清單中的第一個數值，會對應到檢定變數的最低組別數值，而最後一個數值，則對應至最高的數值。數值清單中的各項要素會被加總，然後各數值再除以這個總和，以算出在對應類別中的期望觀察值比例。例如，3、4、5、4 的數值清單，指定期望比例為 3/16、4/16、5/16 和 4/16。

卡方檢定的選項

圖表 27-50
「卡方檢定選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（卡方檢定）

指令語法語言也可以讓您：

- 為不同的變數，指定不同的最小值和最大值、或期望次數（利用 CHISQUARE 次指令）。
- 使用不同的期望次數或不同的範圍，來檢定相同的變數（利用 EXPECTED 次指令令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

二項式檢定

您可以利用「二項式檢定」程序，將二分變數兩個類別的觀察次數，跟具有特定機率參數的二項式分配之下的期望次數，互相做一比較。依照預設值，這兩組的機率參數都是 0.5。若要變更機率，您可以為第一組輸入一個檢定比例。至於第二組的機率，則是 1 減去第一組的指定機率。

範例。以擲銅板為例。當您擲銅板的時候，出現頭的機率是 1/2。我們現在用這個前提當作假設，如果您擲 40 下銅板，並記錄每一次的結果（頭或數字）。根據二項式檢定，您或許會發現 40 次的 3/4（30次）出現頭，也就是觀察顯著水準很小（0.0027）。這些結果指出，頭的出現機率不太可能是 1/2，所以這個硬幣可能有問題。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。

資料。 受檢定的變數必須是二分的數字變數。若要將字串變數轉換成數字變數，請使用「轉換」功能表上的「自動重新編碼」程序。**二分變數**是一個只能有兩種數值的變數：yes 或 no、true 或 false、0 或 1 等。在資料集中發現的第一個值會定義第一個群組，

其他值則定義第二個群組。如果不是二分的變數，您必須指定分割點。此分割點會把觀察值分成兩組，第一組的值大於或等於分割點，第二組的值小於分割點。

假設。 無母數檢定不需要有關底線分配類型的假設。因為資料都已先假設為隨機樣本。

若要取得二項式檢定

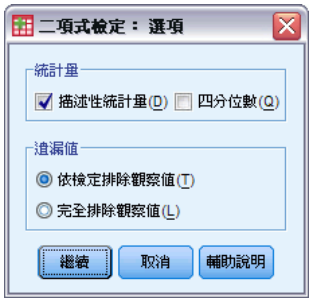
- 從功能表選擇：
  分析(A) > 無母數檢定 > 歷史對話記錄 > 二項式...

圖表 27-51
「二項式檢定」對話方塊

- 選取一個（或以上的）數字檢定變數。
- 或者，按一下「選項」選取敘述統計、四分位數，還可以決定 SPSS 該如何處理遺漏資料。

二項式檢定選項

圖表 27-52
「二項式檢定選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果受檢定變數的觀察值含有遺漏值的話，那麼所有的分析都會排除這個觀察值。

NPAR TESTS 指令的其他功能（二項式檢定）

指令語法語言也可以讓您：

- 當變數有兩個以上的類別時，選出特定的組別（並排除其他組別）（使用 BINOMIAL 次指令）。
- 為不同的變數，指定不同的分割點或機率（使用 BINOMIAL 次指令）。
- 依照不同的分割點或機率，來檢定同一個變數（使用 EXPECTED 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

連檢定

您可以利用「連檢定」程序，來檢定變數中，兩數值的發生順序，是否為隨機。連檢定的值是一連續的觀察值。如果樣本所含的連數太多或者太少，就代表該樣本不是隨機的。

範例。 假設有 20 個人要投票表決，是否購買某項產品。如果這 20 個人的性別都是相同的話，那麼我們就會質疑這個樣本假設的隨機性。此時就可以用連檢定來判斷，此樣本是否是隨機選取的。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。

資料。 必須為數字變數。若要將字串變數轉換成數字變數，請使用「轉換」功能表上的「自動重新編碼」程序。

假設。 無母數檢定不需要有關底線分配類型的假設。請使用連續機率分配的樣本。

若要取得連檢定

▶ 從功能表選擇：
分析(A) > 無母數檢定 > 歷史對話記錄 > 連檢定...

圖表 27-53
新增自訂分割點

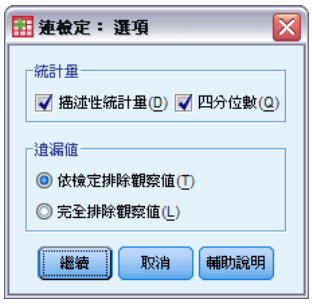
- 選取一個（或以上的）數字檢定變數。
- 或者，按一下「選項」選取敘述統計、四分位數，還可以決定 SPSS 該如何處理遺漏資料。

連檢定的分割點

分割點。 指定分割點，將選用變數一分為二。您可以使用觀察平均數、中位數、或眾數，也可以指定一個數值，當作分割點。如果觀察值小於分割點，則被歸在一組，如果大於分割點，則被歸在另外一組。每個分割點上都會執行一個檢定。

連檢定的選項

圖表 27-54
「連檢定選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（連檢定）

指令語法語言也可以讓您：

- 指定不同變數的分割點（利用 RUNS 次指令）。
- 根據數個自訂分割點（其值不同），檢定相同的變數（利用 RUNS 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

單一樣本 Kolmogorov-Smirnov 檢定

「單一樣本 Kolmogorov-Smirnov 檢定」程序，可使用指定的理論分配，比較變數的觀察累積分配函數。這個指定的分配可能是常態的、均勻的、Poisson 或指數分配。而 Kolmogorov-Smirnov Z 檢定，就是觀察和理論累積分配函數之間的最大差異（絕對值）。這個適合度檢定，可測試觀察值的指定分配是否合理。

範例。在許多參數檢定中，都需要使用常態分配的變數。而單一樣本 Kolmogorov-Smirnov 檢定，則可以用來測試變數的分配是否為常態，（例如 income）。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。

資料。使用數值變數（測量的區間或比例水準）。

假設。在 Kolmogorov-Smirnov 檢定中，系統會假設您已事先指定檢定分配的參數。此程序將會從樣本來估計參數。樣本平均數和樣本標準差，就是常態分配的參數。 而樣本最小值和最大值，則用來定義均勻分配的範圍。Poisson 分配的參數、與指數分配的參數，都是樣本平均數。檢定的冪次在偵測偏離假設分配時可能會嚴重降低。檢定含估計參數的常態分配時，請考慮調整過的 K-S Lilliefors 檢定（可自「預檢資料」程序中取得）。

若要取得單一樣本 Kolmogorov-Smirnov 檢定

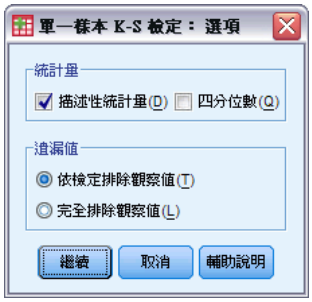
▶ 從功能表選擇：
分析(A) > 無母數檢定 > 歷史對話記錄 > 單一樣本 K-S 檢定...

圖表 27-55
「單一樣本 K-S 檢定」對話方塊

- 選取一個（或以上的）數字檢定變數。每個變數會產生不同的檢定。
- 或者，按一下「選項」選取敘述統計、四分位數，還可以決定 SPSS 該如何處理遺漏資料。

單一樣本 Kolmogorov-Smirnov 檢定選項

圖表 27-56
「單一樣本 K-S 檢定選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（單一樣本 Kolmogorov-Smirnov 檢定）

指令語言也允許您指定檢定分配的參數（使用 K-S 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

兩個獨立樣本檢定

您可以利用「兩個獨立樣本檢定」程序，來比較某個變數的兩組觀察值。

範例。 以新開發的齒列矯正器為例，其目的是要使患者戴起來更舒服、外表看起來更美觀、並且縮短整個療程。為了研究新、舊齒列矯正器的配戴時間是否相同。首先，我們隨機取樣 10 位兒童來配戴舊的齒列矯正器，另外取樣 10 位兒童配戴新的齒列矯正器。從 Mann-Whitney U 統計量檢定中，您可能會發現，平均來說，新齒列矯正器的配戴時間，會比舊齒列矯正器的配戴時間來得短。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。就檢定方面來說，包括：Mann-Whitney U 統計量、Moses 極端反應、Kolmogorov-Smirnov Z 檢定、Wald-Wolfowitz 連檢定等。

資料。 請使用可以排列順序的數字變數。

假設。 請使用獨立的隨機樣本。Mann-Whitney U 檢定是測試兩種分配的相等性。為使用此檢定來測試兩種分配之間位置的差異性，必須假設分配有相同的形狀。

若要取得兩個獨立樣本檢定

- 從功能表選擇：
  分析(A) > 無母數檢定 > 歷史對話記錄 > 二個獨立樣本...

圖表 27-57
「兩個獨立樣本檢定」對話方塊

- 選取一個（或以上的）數字變數。
- 選擇分組變數，並按一下「定義組別」，以便將檔案分成兩個組別或樣本。

兩個獨立樣本檢定的類型

檢定類型。 若要檢定兩個獨立樣本（組別）是否來自相同的母群體，您可以使用下列四種檢定方法。

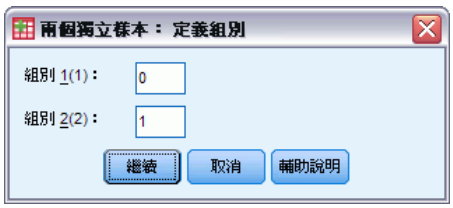
Mann-Whitney U 統計量檢定，是所有兩個獨立樣本檢定中，最常用的方法。它相當於兩個組別的 Wilcoxon 等級和檢定，以及 Kruskal-Wallis 檢定。Mann-Whitney 統計量，會檢定兩個取樣的母群體，是否在相同的位置上。然後，再利用同分觀察值中所指定的均數等級，將這兩個組別的觀察結果進行組合並分級。同分觀察值的個數，應該比觀察的總個數少。如果在位置上的母群體相同，則等級應該在兩樣本間隨機混合。統計量檢定會計算組別 1 分數高於組別 2 分數的次數，以及組別 2 分數高於組別 1 分數的次數，而 Mann-Whitney U 統計量則是這兩個次數中較小的那一個。也會顯示 Wilcoxon 等級總和檢定 W 統計量。W 是平均數等級較小之群組的等級總和，若群組具有相同的平均數等級，則其為「兩個獨立樣本定義組別」對話方塊內最後命名之群組的等級總和。

Kolmogorov-Smirnov Z 檢定和 **Wald-Wolfowitz 連檢定**是屬於比較一般性的檢定，會偵測分散位置和類型差異。Kolmogorov-Smirnov 檢定，乃是以兩樣本在觀察累積分配函數之間的最大值絕對差異為基礎。當差異非常大時，這兩個分配就會被視為不一樣。至於 Wald-Wolfowitz 連檢定，它會組合兩組別中的觀察值，並將其分級。如果這兩個樣本來自相同的母群體，那麼這兩個組別就應該隨機散佈於等級中。

Moses 極端反應檢定，假設實驗變數將會在某一方面影響某些受試者，而在另一個相反方面影響其他受試者。此檢定會在另一個控制組的比較之下，檢定極端反應值。這個檢定方法的焦點，乃是集中在控制組等級差異的範圍，以及在將實驗組與控制組合併下，測量實驗組中極端值數目，如何影響等級差異範圍的方法。此檢定方法中的控制組，是由「兩個獨立樣本定義組別」對話方塊中的 group 1 值所定義的。它會對這兩個組別進行觀察，然後再將觀察結果組合並分級。至於控制組的等級差異範圍，會被計算為控制組數加 1 中、最大和最小值等級間的差異。 因為機會偏離值會輕易地扭曲差異等級範圍，所以會自動地從控制觀察值的兩端各刪除 5%。

兩個獨立樣本檢定的定義組別

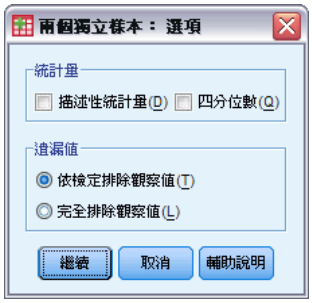
圖表 27-58
「兩個獨立樣本定義組別」對話方塊



若要將檔案分成兩個組別或樣本，請為 Group 1 和 Group 2 分別輸入一個整數值，則其他數值的觀察值便會從此分析中排除。

兩個獨立樣本檢定的選項

圖表 27-59
「兩個獨立樣本選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（兩個獨立樣本檢定）

指令語法語言也可以讓您指定 Moses 檢定中要刪除的觀察值個數（使用 MOSES 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

兩個相關樣本檢定

您可以利用「兩個相關樣本檢定」程序，來比較兩個變數的分配情況。

範例。 以房屋出售為例。一般來說，在出售房子時，屋主會賣到他們所要求的價格嗎？我們將 10 間房屋的相關資料使用 Wilcoxon 符號等級檢定之後，可能會得到以下的結果，有七個家庭收到低於要求的價格，一個家庭收到高於要求的價錢，另外兩個家庭則收到所要求的價錢。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。就檢定方面來說，包括：Wilcoxon 符號等級、符號，和 McNemar。若安裝「精確檢定」選項（只能在 Windows 作業系統中使用），也可以使用邊際均齊性檢定。

資料。 請使用可以排列順序的數字變數。

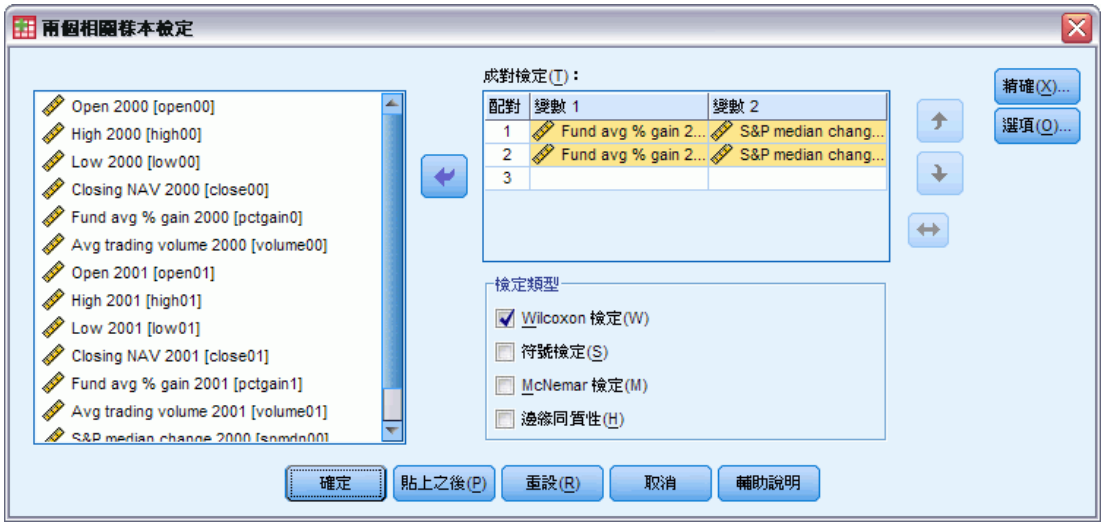
假設。 雖然沒有對這兩個變數假設任何特殊的分配，但是仍須將這對差異的母群體分配假設為對稱。

若要取得兩個相關樣本檢定

▶ 從功能表選擇：

分析(A) > 無母數檢定 > 歷史對話記錄 > 二個相關樣本...

圖表 27-60
「兩個相關樣本檢定」對話方塊



▶ 選取一或多對變數。

兩個相關樣本檢定的類型

在本節中所述的檢定，是用來比較兩個相關變數的分配情形。若要決定使用何種檢定較為合適，應視資料的類型而定。

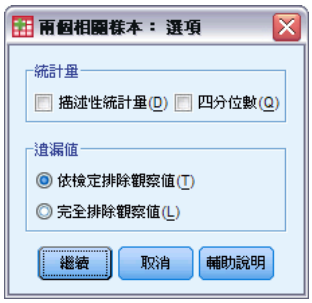
如果您的資料是連續的，請使用符號檢定或 Wilcoxon 符號等級檢定。**符號檢定**會為所有的觀察值，計算這兩個變數間的差異，並將這些差異分類成正、負或同分。如果這兩個變數的分配類似，則正負差異的個數將不會有顯著的不同。**Wilcoxon 符號等級檢定**會考慮有關成對變數間、差異符號和差異級數兩者的相關資訊。因為 Wilcoxon 符號等級檢定合併更多有關資料的資訊，因此比符號檢定的功能更強大。

如果您的資料是二元進位，請使用 **McNemar 檢定**。這種檢定方法，通常是用於重複性測量，此時每位受試者的反應都會出現兩次（在指定事件發生前後各一次）。McNemar 檢定會判斷初始的反應率（在事件之前），是否與最後的反應率（在事件之後）相同。在設計為事件前後的實驗類型中，若要偵測因實驗中斷所導致的反應變更，這種檢定方法非常有用。

如果您的資料是類別形式的話，請使用**邊際均齊性檢定**。這個檢定是將 McNemar 檢定從二項反應擴充到多項式反應。此外，它也會使用卡方分配檢定反應值的變更，而且在設計為事件前後的實驗類型中，若要偵測因實驗中斷所導致的反應值變更，這種檢定方法非常好用。不過，只有在您安裝「Exact Tests」之後，才能使用邊緣均齊性檢定。

兩個相關樣本檢定選項

圖表 27-61
「兩個相關樣本選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（兩個相關樣本）

指令語法語言也可以讓您將變數列成清單，檢定其中每一個變數。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

多個獨立樣本的檢定

您可以透過「多個獨立樣本的檢定」程序，針對某個變數，比較兩個（或以上）的觀察值組別。

範例。 檢定三種品牌的 100 瓦燈泡，其平均使用壽命都不一樣嗎？從 Kruskal-Wallis 單因子變異數分析中，您可以得知，這三種品牌的平均使用壽命的確不一樣。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。就檢定方面來說，包括：Kruskal-Wallis H 檢定、中位數檢定。

資料。 請使用可以排列順序的數字變數。

假設。 請使用獨立的隨機樣本。但是 Kruskal-Wallis H 檢定會要求所檢定的樣本，其類型必須相似。

若要取得多個獨立樣本的檢定

▶ 從功能表選擇：
分析(A) > 無母數檢定 > 歷史對話記錄 > K 個獨立樣本...

圖表 27-62
定義中位數檢定

- 選取一個（或以上的）數字變數。
- 選取分組變數，並按一下「**定義範圍**」，指定分組變數的最小和最大整數值。

多個獨立樣本檢定的類型

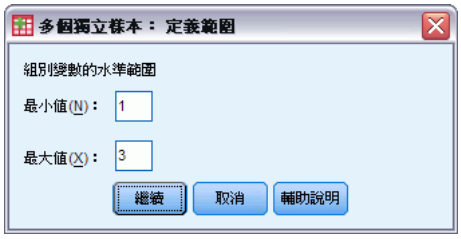
您可以使用三種檢定，來判斷多個獨立樣本是否是取自於相同母群。Kruskal-Wallis H 檢定、中位數檢定，還有 Jonckheere-Terpstra 檢定，都可以用來檢定多個獨立樣本是否是取自相同母群。

Kruskal-Wallis H 檢定：是 Mann-Whitney U 統計量檢定的延伸，為單因子變異數分析的無母數類比，還可以用來偵測分配位置的差異。**中位數檢定**：此檢定比較一般化（但功能卻不夠強大），但是您還是可以用它來偵測分配在位置和類型上面的差異。Kruskal-Wallis H 檢定和中位數檢定：均假設 k 母群（為樣本來源）沒有事前順序。

如果出現 k 母群的自然事前順序（遞增或遞減）時，**Jonckheere-Terpstra test** 檢定功能便會變得比較強大。例如，k 母群可能代表所增加的溫度（k 度）。而我們可以先假設不同溫度，但回應分配是一樣的，以便跟另一個假設：溫度增加、回應等級就隨之增加，來互相對照，並加以檢定。這兩個假設都已經排序過，因此，Jonckheere-Terpstra 是最適用的檢定。請注意，Jonckheere-Terpstra 檢定必須在您安裝「Exact Tests 附加模組」之後才能使用。

多個獨立樣本檢定的定義範圍

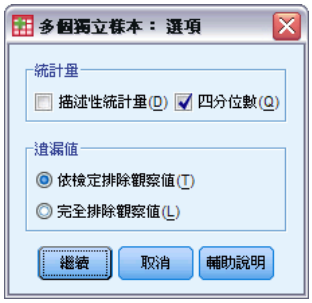
圖表 27-63
「多個獨立樣本定義範圍」對話方塊



若要定義範圍，請輸入**最小值**和**最大值**的整數值，以便跟分組變數最低和最高類別互相對應。如果觀察值落在界限之外，就會被排除。例如，如果您指定最小值為 1，而最大值為 3，那麼只會使用介於 1 跟 3 之間的整數。最小值必須小於最大值，而且您必須指定這兩個數值。

多個獨立樣本檢定的選項

圖表 27-64
「多個獨立樣本選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（K 個獨立樣本）

指令語法語言也可以讓您為中位數檢定指定觀察中位數之外的數值（使用 MEDIAN 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

多個相關樣本的檢定

您可以利用「多個相關樣本的檢定」程序，來比較兩個（或以上）變數的分配情形。

範例。 公共關係會使醫生、律師、警察、和教師的聲望程度有所不同嗎？我們要求 10 個人，排列這四種職業的聲望名次。由 Friedman 的檢定可以看出，公共關係跟這四種職業的聲望程度有關係。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。就檢定方面來說，包括：Friedman 檢定、Kendall’s W 檢定和 Cochran’s Q 檢定。

資料。 請使用可以排列順序的數字變數。

假設。 無母數檢定不需要有關底線分配類型的假設。請使用連續機率分配的樣本。

若要取得多個相關樣本的檢定

▶ 從功能表選擇：

分析(A) > 無母數檢定 > 歷史對話記錄 > K 個相關樣本...

圖表 27-65
選取 Cochran 做為檢定類型

▶ 選取兩個（或以上的）數字檢定變數。

多個相關樣本檢定的類型

有三種可用的檢定，以便您比較多個相關變數的分配。

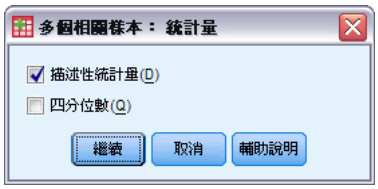
Friedman 檢定是一種無母數檢定，該種檢定跟單一樣本重複量數設計、或每一格具一個觀察值之二因子變異數分析是相等的。Friedman 檢定一個虛無假設：k 個相關變數都是來自相同母群。對於每個觀察值而言，k 變數的等級是從 1 到 k。檢定統計量以這些等級為基礎。

Kendall’s W 檢定是將 Friedman 統計量常態化。Kendall’s W 檢定則可解釋成是一種 “和諧係數”，它測量評估者之間的一致性。每一個觀察值都是一位裁判（或定等級者），而每一個變數都是被審查的項目(或個人)。針對每個變數，計算等級總和。Kendall’s W 檢定範圍介於 0（不同意）和 1（完全同意）之間。

Cochran’s Q 檢定與 Friedman 檢定相同；此外，如果答案都是二元化，也可以使用它。這個檢定是 McNemar 檢定 k 個樣本情況的延伸。Cochran’s Q 檢定的假設如下：多個相關二分變數都有相同平均數。您應該對同一位受訪者（或條件相符的受訪者）測量變數。

多個相關樣本檢定的統計量

圖表 27-66
「多個相關樣本統計量」對話方塊



您可以選擇統計量。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

NPAR TESTS 指令的其他功能（K 個相關樣本）

如需完整的語法資訊，請參閱《指令語法參考手冊》。

二項式檢定

您可以利用「二項式檢定」程序，將二分變數兩個類別的觀察次數，跟具有特定機率參數的二項式分配之下的期望次數，互相做一比較。依照預設值，這兩組的機率參數都是 0.5。若要變更機率，您可以為第一組輸入一個檢定比例。至於第二組的機率，則是 1 減去第一組的指定機率。

範例。以擲銅板為例。當您擲銅板的時候，出現頭的機率是 1/2。我們現在用這個前提當作假設，如果您擲 40 下銅板，並記錄每一次的結果（頭或數字）。根據二項式檢定，您或許會發現 40 次的 3/4（30次）出現頭，也就是觀察顯著水準很小（0.0027）。這些結果指出，頭的出現機率不太可能是 1/2，所以這個硬幣可能有問題。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。

資料。 受檢定的變數必須是二分的數字變數。若要將字串變數轉換成數字變數，請使用「轉換」功能表上的「自動重新編碼」程序。**二分變數**是一個只能有兩種數值的變數：yes 或 no、true 或 false、0 或 1 等。在資料集中發現的第一個值會定義第一個群組，其他值則定義第二個群組。如果不是二分的變數，您必須指定分割點。此分割點會把觀察值分成兩組，第一組的值大於或等於分割點，第二組的值小於分割點。

假設。 無母數檢定不需要有關底線分配類型的假設。因為資料都已先假設為隨機樣本。

若要取得二項式檢定

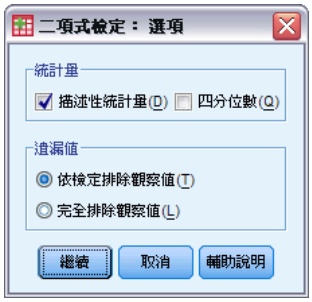
▶ 從功能表選擇：
分析(A) > 無母數檢定 > 歷史對話記錄 > 二項式...

圖表 27-67
「二項式檢定」對話方塊

- 選取一個（或以上的）數字檢定變數。
- 或者，按一下「選項」選取敘述統計、四分位數，還可以決定 SPSS 該如何處理遺漏資料。

二項式檢定選項

圖表 27-68
「二項式檢定選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果受檢定變數的觀察值含有遺漏值的話，那麼所有的分析都會排除這個觀察值。

NPAR TESTS 指令的其他功能（二項式檢定）

指令語法語言也可以讓您：

- 當變數有兩個以上的類別時，選出特定的組別（並排除其他組別）（使用 BINOMIAL 次指令）。
- 為不同的變數，指定不同的分割點或機率（使用 BINOMIAL 次指令）。
- 依照不同的分割點或機率，來檢定同一個變數（使用 EXPECTED 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

連檢定

您可以利用「連檢定」程序，來檢定變數中，兩數值的發生順序，是否為隨機。連檢定的值是一連續的觀察值。如果樣本所含的連數太多或者太少，就代表該樣本不是隨機的。

範例。 假設有 20 個人要投票表決，是否購買某項產品。如果這 20 個人的性別都是相同的話，那麼我們就會質疑這個樣本假設的隨機性。此時就可以用連檢定來判斷，此樣本是否是隨機選取的。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。

資料。 必須為數字變數。若要將字串變數轉換成數字變數，請使用「轉換」功能表上的「自動重新編碼」程序。

假設。 無母數檢定不需要有關底線分配類型的假設。請使用連續機率分配的樣本。

若要取得連檢定

▶ 從功能表選擇：

分析(A) > 無母數檢定 > 歷史對話記錄 > 連檢定...

圖表 27-69
新增自訂分割點

▶ 選取一個（或以上的）數字檢定變數。

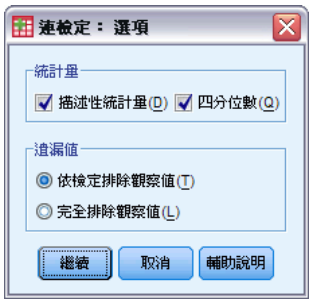
▶ 或者，按一下「選項」選取敘述統計、四分位數，還可以決定 SPSS 該如何處理遺漏資料。

連檢定的分割點

分割點。 指定分割點，將選用變數一分為二。您可以使用觀察平均數、中位數、或眾數，也可以指定一個數值，當作分割點。如果觀察值小於分割點，則被歸在一組，如果大於分割點，則被歸在另外一組。每個分割點上都會執行一個檢定。

連檢定的選項

圖表 27-70
「連檢定選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（連檢定）

指令語法語言也可以讓您：

- 指定不同變數的分割點（利用 RUNS 次指令）。
- 根據數個自訂分割點（其值不同），檢定相同的變數（利用 RUNS 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

單一樣本 Kolmogorov-Smirnov 檢定

「單一樣本 Kolmogorov-Smirnov 檢定」程序，可使用指定的理論分配，比較變數的觀察累積分配函數。這個指定的分配可能是常態的、均勻的、Poisson 或指數分配。而 Kolmogorov-Smirnov Z 檢定，就是觀察和理論累積分配函數之間的最大差異（絕對值）。這個適合度檢定，可測試觀察值的指定分配是否合理。

範例。在許多參數檢定中，都需要使用常態分配的變數。而單一樣本 Kolmogorov-Smirnov 檢定，則可以用來測試變數的分配是否為常態，(例如 income)。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。

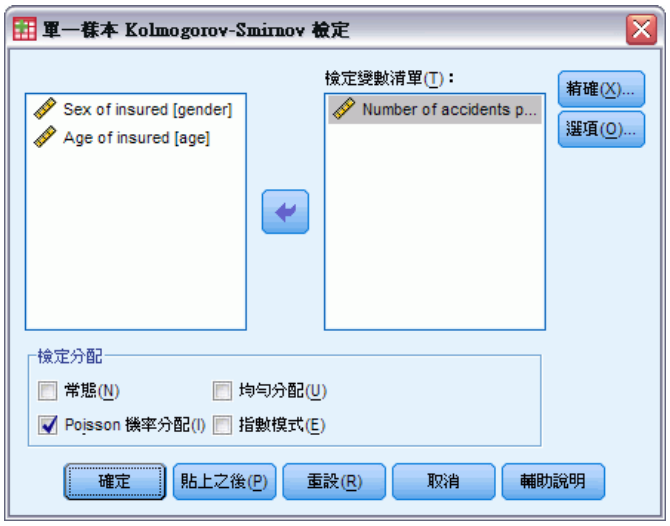
資料。使用數值變數（測量的區間或比例水準)。

假設。在 Kolmogorov-Smirnov 檢定中，系統會假設您已事先指定檢定分配的參數。此程序將會從樣本來估計參數。樣本平均數和樣本標準差，就是常態分配的參數。 而樣本最小值和最大值，則用來定義均勻分配的範圍。Poisson 分配的參數、與指數分配的參數，都是樣本平均數。檢定的冪次在偵測偏離假設分配時可能會嚴重降低。檢定含估計參數的常態分配時，請考慮調整過的 K-S Lilliefors 檢定（可自「預檢資料」程序中取得)。

若要取得單一樣本 Kolmogorov-Smirnov 檢定

▶ 從功能表選擇：
分析(A) > 無母數檢定 > 歷史對話記錄 > 單一樣本 K-S 檢定...

圖表 27-71
「單一樣本 K-S 檢定」對話方塊

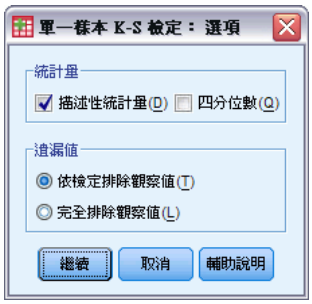


▶ 選取一個（或以上的）數字檢定變數。每個變數會產生不同的檢定。

▶ 或者，按一下「選項」選取敘述統計、四分位數，還可以決定 SPSS 該如何處理遺漏資料。

單一樣本 Kolmogorov-Smirnov 檢定選項

圖表 27-72
「單一樣本 K-S 檢定選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（單一樣本 Kolmogorov-Smirnov 檢定）

指令語言也允許您指定檢定分配的參數（使用 K-S 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

兩個獨立樣本檢定

您可以利用「兩個獨立樣本檢定」程序，來比較某個變數的兩組觀察值。

範例。 以新開發的齒列矯正器為例，其目的是要使患者戴起來更舒服、外表看起來更美觀、並且縮短整個療程。為了研究新、舊齒列矯正器的配戴時間是否相同。首先，我們隨機取樣 10 位兒童來配戴舊的齒列矯正器，另外取樣 10 位兒童配戴新的齒列矯正器。從 Mann-Whitney U 統計量檢定中，您可能會發現，平均來說，新齒列矯正器的配戴時間，會比舊齒列矯正器的配戴時間來得短。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。就檢定方面來說，包括：Mann-Whitney U 統計量、Moses 極端反應、Kolmogorov-Smirnov Z 檢定、Wald-Wolfowitz 連檢定等。

資料。 請使用可以排列順序的數字變數。

假設。 請使用獨立的隨機樣本。Mann-Whitney U 檢定是測試兩種分配的相等性。為使用此檢定來測試兩種分配之間位置的差異性，必須假設分配有相同的形狀。

若要取得兩個獨立樣本檢定

▶ 從功能表選擇：
分析(A) > 無母數檢定 > 歷史對話記錄 > 二個獨立樣本...

圖表 27-73
「兩個獨立樣本檢定」對話方塊

▶ 選取一個（或以上的）數字變數。

▶ 選擇分組變數，並按一下「定義組別」，以便將檔案分成兩個組別或樣本。

兩個獨立樣本檢定的類型

檢定類型。 若要檢定兩個獨立樣本（組別）是否來自相同的母群體，您可以使用下列四種檢定方法。

Mann-Whitney U 統計量檢定，是所有兩個獨立樣本檢定中，最常用的方法。它相當於兩個組別的 Wilcoxon 等級和檢定，以及 Kruskal-Wallis 檢定。Mann-Whitney 統計量，會檢定兩個取樣的母群體，是否在相同的位置上。然後，再利用同分觀察值中所指定的均數等級，將這兩個組別的觀察結果進行組合並分級。同分觀察值的個數，應該比觀察的總個數少。如果在位置上的母群體相同，則等級應該在兩樣本間隨機混合。統計量檢定會計算組別 1 分數高於組別 2 分數的次數，以及組別 2 分數高於組別 1 分數的次數，而 Mann-Whitney U 統計量則是這兩個次數中較小的那一個。也會顯示 Wilcoxon 等級總和檢定 W 統計量。W 是平均數等級較小之群組的等級總和，若群組具有相同的平均數等級，則其為「兩個獨立樣本定義組別」對話方塊內最後命名之群組的等級總和。

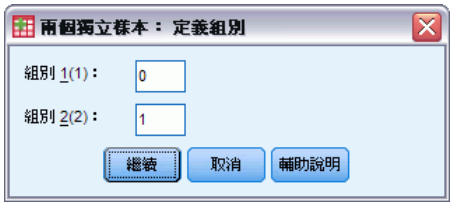
Kolmogorov-Smirnov Z 檢定和 **Wald-Wolfowitz 連檢定**是屬於比較一般性的檢定，會偵測分散位置和類型差異。Kolmogorov-Smirnov 檢定，乃是以兩樣本在觀察累積分配函數之間的最大值絕對差異為基礎。當差異非常大時，這兩個分配就會被視為不一樣。至於 Wald-Wolfowitz 連檢定，它會組合兩組別中的觀察值，並將其分級。如果這兩個樣本來自相同的母群體，那麼這兩個組別就應該隨機散佈於等級中。

Moses 極端反應檢定，假設實驗變數將會在某一方面影響某些受試者，而在另一個相反方面影響其他受試者。此檢定會在另一個控制組的比較之下，檢定極端反應值。這個檢定方法的焦點，乃是集中在控制組等級差異的範圍，以及在將實驗組與控制組合併

下，測量實驗組中極端值數目，如何影響等級差異範圍的方法。此檢定方法中的控制組，是由「兩個獨立樣本定義組別」對話方塊中的 group 1 值所定義的。它會對這兩個組別進行觀察，然後再將觀察結果組合並分級。至於控制組的等級差異範圍，會被計算為控制組數加 1 中、最大和最小值等級間的差異。 因為機會偏離值會輕易地扭曲差異等級範圍，所以會自動地從控制觀察值的兩端各刪除 5%。

兩個獨立樣本檢定的定義組別

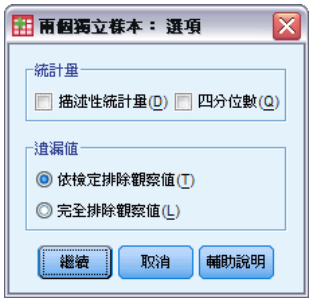
圖表 27-74
「兩個獨立樣本定義組別」對話方塊



若要將檔案分成兩個組別或樣本，請為 Group 1 和 Group 2 分別輸入一個整數值，則其他數值的觀察值便會從此分析中排除。

兩個獨立樣本檢定的選項

圖表 27-75
「兩個獨立樣本選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（兩個獨立樣本檢定）

指令語法語言也可以讓您指定 Moses 檢定中要刪除的觀察值個數（使用 MOSES 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

兩個相關樣本檢定

您可以利用「兩個相關樣本檢定」程序，來比較兩個變數的分配情況。

範例。 以房屋出售為例。一般來說，在出售房子時，屋主會賣到他們所要求的價格嗎？我們將 10 間房屋的相關資料使用 Wilcoxon 符號等級檢定之後，可能會得到以下的結果，有七個家庭收到低於要求的價格，一個家庭收到高於要求的價錢，另外兩個家庭則收到所要求的價錢。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。就檢定方面來說，包括：Wilcoxon 符號等級、符號，和 McNemar。若安裝「精確檢定」選項（只能在 Windows 作業系統中使用），也可以使用邊際均齊性檢定。

資料。 請使用可以排列順序的數字變數。

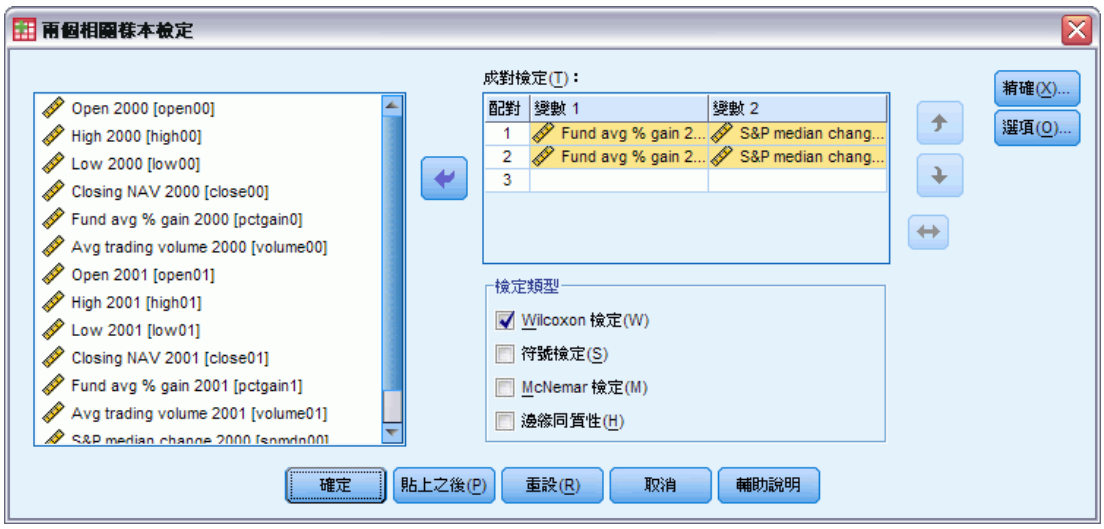
假設。 雖然沒有對這兩個變數假設任何特殊的分配，但是仍須將這對差異的母群體分配假設為對稱。

若要取得兩個相關樣本檢定

▶ 從功能表選擇：

分析(A) > 無母數檢定 > 歷史對話記錄 > 二個相關樣本...

圖表 27-76
「兩個相關樣本檢定」對話方塊



▶ 選取一或多對變數。

兩個相關樣本檢定的類型

在本節中所述的檢定，是用來比較兩個相關變數的分配情形。若要決定使用何種檢定較為合適，應視資料的類型而定。

如果您的資料是連續的，請使用符號檢定或 Wilcoxon 符號等級檢定。**符號檢定**會為所有的觀察值，計算這兩個變數間的差異，並將這些差異分類成正、負或同分。如果這兩個變數的分配類似，則正負差異的個數將不會有顯著的不同。**Wilcoxon 符號等級檢定**

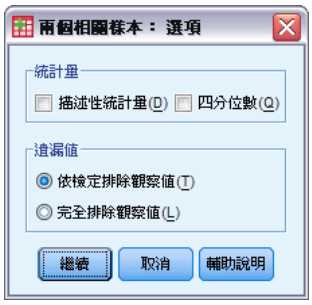
會考慮有關成對變數間、差異符號和差異級數兩者的相關資訊。因為 Wilcoxon 符號等級檢定合併更多有關資料的資訊，因此比符號檢定的功能更強大。

如果您的資料是二元進位，請使用 **McNemar 檢定**。這種檢定方法，通常是用於重複性測量，此時每位受試者的反應都會出現兩次（在指定事件發生前後各一次）。McNemar 檢定會判斷初始的反應率（在事件之前），是否與最後的反應率（在事件之後）相同。在設計為事件前後的實驗類型中，若要偵測因實驗中斷所導致的反應變更，這種檢定方法非常有用。

如果您的資料是類別形式的話，請使用**邊際均齊性檢定**。這個檢定是將 McNemar 檢定從二項反應擴充到多項式反應。此外，它也會使用卡方分配檢定反應值的變更，而且在設計為事件前後的實驗類型中，若要偵測因實驗中斷所導致的反應值變更，這種檢定方法非常好用。不過，只有在您安裝「Exact Tests」之後，才能使用邊緣均齊性檢定。

兩個相關樣本檢定選項

圖表 27-77
「兩個相關樣本選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（兩個相關樣本）

指令語法語言也可以讓您將變數列成清單，檢定其中每一個變數。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

多個獨立樣本的檢定

您可以透過「多個獨立樣本的檢定」程序，針對某個變數，比較兩個（或以上）的觀察值組別。

範例。 檢定三種品牌的 100 瓦燈泡，其平均使用壽命都不一樣嗎？從 Kruskal-Wallis 單因子變異數分析中，您可以得知，這三種品牌的平均使用壽命的確不一樣。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。就檢定方面來說，包括：Kruskal-Wallis H 檢定、中位數檢定。

資料。 請使用可以排列順序的數字變數。

假設。 請使用獨立的隨機樣本。但是 Kruskal-Wallis H 檢定會要求所檢定的樣本，其類型必須相似。

若要取得多個獨立樣本的檢定

- 從功能表選擇：

  分析(A) > 無母數檢定 > 歷史對話記錄 > K 個獨立樣本...

圖表 27-78
定義中位數檢定

- 選取一個（或以上的）數字變數。
- 選取分組變數，並按一下「定義範圍」，指定分組變數的最小和最大整數值。

多個獨立樣本檢定的類型

您可以使用三種檢定，來判斷多個獨立樣本是否是取自於相同母群。Kruskal-Wallis H 檢定、中位數檢定，還有 Jonckheere-Terpstra 檢定，都可以用來檢定多個獨立樣本是否是取自相同母群。

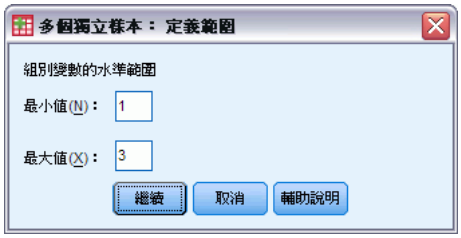
Kruskal-Wallis H 檢定：是 Mann-Whitney U 統計量檢定的延伸，為單因子變異數分析的無母數類比，還可以用來偵測分配位置的差異。**中位數檢定**：此檢定比較一般化（但功能卻不夠強大），但是您還是可以用它來偵測分配在位置和類型上面的差異。Kruskal-Wallis H 檢定和中位數檢定：均假設 k 母群（為樣本來源）沒有事前順序。

如果出現 k 母群的自然事前順序（遞增或遞減）時，**Jonckheere-Terpstra test** 檢定功能便會變得比較強大。例如，k 母群可能代表所增加的溫度（k 度）。而我們可以先假設不同溫度，但回應分配是一樣的，以便跟另一個假設：溫度增加、回應等級就隨之增加，來互相對照，並加以檢定。這兩個假設都已經排序過，因此，Jonckheere-Terpstra

是最適用的檢定。請注意，Jonckheere-Terpstra 檢定必須在您安裝「Exact Tests 附加模組」之後才能使用。

多個獨立樣本檢定的定義範圍

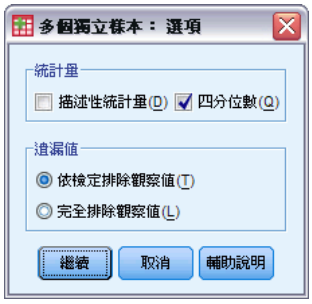
圖表 27-79
「多個獨立樣本定義範圍」對話方塊



若要定義範圍，請輸入**最小值**和**最大值**的整數值，以便跟分組變數最低和最高類別互相對應。如果觀察值落在界限之外，就會被排除。例如，如果您指定最小值為 1，而最大值為 3，那麼只會使用介於 1 跟 3 之間的整數。最小值必須小於最大值，而且您必須指定這兩個數值。

多個獨立樣本檢定的選項

圖表 27-80
「多個獨立樣本選項」對話方塊



統計量。您可以任選摘要統計量中的某一個（或兩者都選）。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

遺漏值。控制處理遺漏值的方式。

- **依檢定排除遺漏值。** 當您指定數種檢定之後，各種檢定會分別對遺漏值進行評估。
- **完全排除遺漏值。** 如果任何變數之觀察值擁有遺漏值的話，則所有分析都會將這個觀察值排除在外。

NPAR TESTS 指令的其他功能（K 個獨立樣本）

指令語法語言也可以讓您為中位數檢定指定觀察中位數之外的數值（使用 MEDIAN 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

多個相關樣本的檢定

您可以利用「多個相關樣本的檢定」程序，來比較兩個（或以上）變數的分配情形。

範例。 公共關係會使醫生、律師、警察、和教師的聲望程度有所不同嗎？我們要求 10 個人，排列這四種職業的聲望名次。由 Friedman 的檢定可以看出，公共關係跟這四種職業的聲望程度有關係。

統計量。 平均數、標準差、最小值、最大值、非遺漏觀察值的個數和四分位數。就檢定方面來說，包括：Friedman 檢定、Kendall’s W 檢定和 Cochran’s Q 檢定。

資料。 請使用可以排列順序的數字變數。

假設。 無母數檢定不需要有關底線分配類型的假設。請使用連續機率分配的樣本。

若要取得多個相關樣本的檢定

▶ 從功能表選擇：
分析(A) > 無母數檢定 > 歷史對話記錄 > K 個相關樣本...

圖表 27-81
選取 Cochran 做為檢定類型

▶ 選取兩個（或以上的）數字檢定變數。

多個相關樣本檢定的類型

有三種可用的檢定，以便您比較多個相關變數的分配。

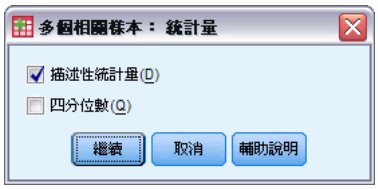
Friedman 檢定是一種無母數檢定，該種檢定跟單一樣本重複量數設計、或每一格具一個觀察值之二因子變異數分析是相等的。Friedman 檢定一個虛無假設：k 個相關變數都是來自相同母群。對於每個觀察值而言，k 變數的等級是從 1 到 k。檢定統計量以這些等級為基礎。

Kendall’s W 檢定是將 Friedman 統計量常態化。Kendall’s W 檢定則可解釋成是一種 “和諧係數”，它測量評估者之間的一致性。每一個觀察值都是一位裁判（或定等級者），而每一個變數都是被審查的項目(或個人)。針對每個變數，計算等級總和。Kendall’s W 檢定範圍介於 0（不同意）和 1（完全同意）之間。

Cochran’s Q 檢定與 Friedman 檢定相同；此外，如果答案都是二元化，也可以使用它。這個檢定是 McNemar 檢定 k 個樣本情況的延伸。Cochran’s Q 檢定的假設如下：多個相關二分變數都有相同平均數。您應該對同一位受訪者（或條件相符的受訪者）測量變數。

多個相關樣本檢定的統計量

圖表 27-82
「多個相關樣本統計量」對話方塊



您可以選擇統計量。

- **描述性統計量。** 顯示平均數、標準差、最小值、最大值、和非遺漏觀察值的個數。
- **四分位數。** 顯示跟第 25 個、第 50 個和第 75 個百分位數相對應的數值。

NPAR TESTS 指令的其他功能（K 個相關樣本）

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
28

複選題分析

當您分析多重二分集和多類別集合時，可以使用兩種程序。「複選題次數分配表」程序會顯示次數分配表。「複選題交叉表」程序則會顯示二維和三維的交叉表。在使用任一個程序之前，您必須先定義複選題集。

範例。下列範例將會說明市場研究調查中，複選題項目的使用方式。但是下面的資料都是虛構的，請勿當作真實範例使用。航空公司為了對競爭對手（別家航空公司）進行評估，可能會調查搭乘特定航線的乘客。在這個範例中，中華航空公司想瞭解它的乘客，在搭乘別家公司的北高航線方面的情形；以及班次和服務品質在乘客選擇航空公司時，所占有的重要性。空服員在乘客登機時，會發給每人一份簡單的問卷。第一個問題是：「請圈選出過去六個月內，您在這條航線上，所有搭乘過（至少一次）的航空公司：中華、遠東、長榮、復興、其他」。因為乘客可以圈選不止一個答案，所以這是複選題。但是因為變數的每個觀察值內，只能擁有一個數值，所以這個問題並無法直接編碼。您必須使用幾個變數，來跟每個問題的答案相對應。您可以使用的方法有兩種：第一種是為每種選擇定義一個變數（例如 China、FarEast、Eva、TransAsia 和 Other）。如果乘客圈選遠東的話，fareast 變數就指定代碼 1，否則指定 0。 這是**多重二分法**的變數對應法。另一種對應答案的方式是**多重類別法**，這個方法是估計問題最多可能有幾個答案，並設定相同的變數數目，再以不同的代碼來指定搭乘的航空公司。當您詳細地看過問卷樣本以後，可能會發現：在過去半年內，沒有任何乘客曾在這條航線上，搭乘過三家以上的航空公司。此外，您發現因為開放天空的緣故，屬於「其他」類別的航空公司名稱已多達 10 家。使用複選題選項法，您將定義三個變數，每個編碼成 1 = china、2 = fareast、3 = eva、4 = transasia、5 = mandarin，依此類推。如果某個乘客圈選中華和長榮的話，第一個變數使用代碼 1，第二個使用代碼 3，第三個就是某個遺漏值代碼。另一個乘客可能圈選了中華，還輸入華信。因此第一個變數使用代碼 1，第二個使用代碼 5，第三個是遺漏值代碼。在另一方面，如果您使用多重二分法的話，就會有 14 個不一樣的變數。雖然這兩種對應的方法對於這項研究來說都很適合，但是到底該用哪一種，則應視答案的分配情形而定。

定義複選題集

「定義複選題集」程序會將基本變數分成多重二分集和多類別集合，讓您可以取得次數分配表和交叉表。您最多可以定義 20 個複選題集。每個集合必須有一個唯一的名稱。若要移除集合，請從複選題集的清單上，將它以反白的方式標記出來，再按一下「**刪除**」。若要變更集合，請在清單上將它反白，再按一下「**變更**」。

您可以把基本變數編成二分變數或類別變數。若要使用二分變數，請選取「**二分法**」以建立多重二分集。如果在計數值中輸入整數值，則計數值至少會出現一次，而且計數值中的每個變數，都會變成多重二分集中的類別。選取「**類別**」，則會建立多類別集合，其值範圍跟成份變數相同。請在多重類別變數集合類別範圍的最小值和最大值

中，輸入整數值。程序會合計範圍內（含括所有成份變數）所有不同的整數值。空的類別將不會列在表裏。

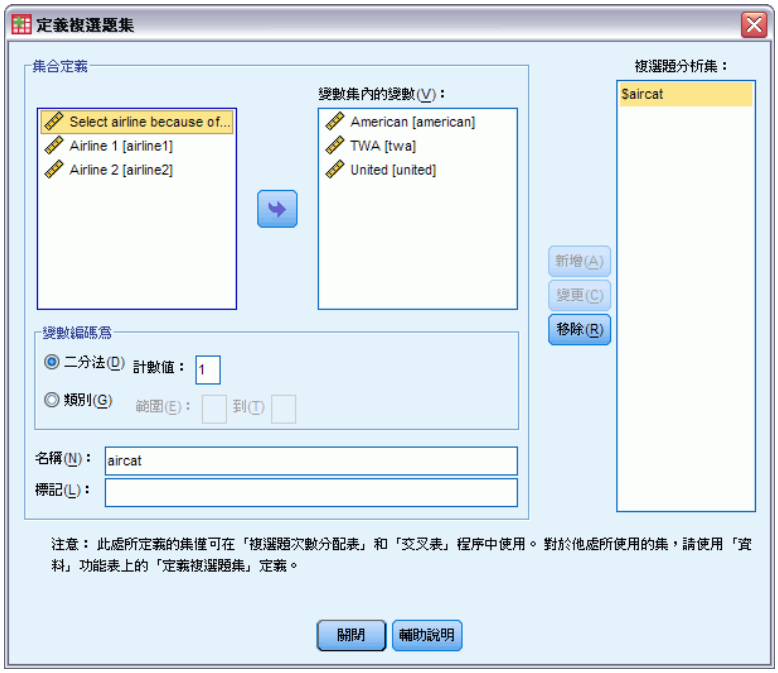
每個複選題集都必須指定專屬的名稱，最多可有七個字元。程序會在您指定的名稱前，加上一個金額符號 ($)。您無法使用下列保留字：casenum、sysmis、jdate、date、time、length 和 width。複選題集名稱之所以存在，只是為了供複選題程序使用。您無法在其他程序中，參考到複選題集的名稱。您可以選擇性地輸入複選題集的描述性變數標記。標記最多可以擁有 40 個字元。

若要定義複選題集

- 從功能表選擇：

  分析(A) > 複選題分析 > 定義變數集...

圖表 28-1
「定義複選題集」對話方塊



- 選取兩個以上的變數。
- 如果您的變數被編成二分變數的話，請指出您想計算的數值。如果變數被編成類別變數的話，請定義類別的範圍。
- 輸入每個複選題集的唯一名稱。
- 按一下「新增」，即可將複選題集加進定義集合的清單中。

複選題次數分配表

「複選題次數分配表」程序可以產生複選題集的次數分配表。您必須先定義一個或多個複選題集（請參閱〈複選題定義集〉）。

針對多重二分集，使用組別中、基本變數定義的變數標記，當作輸出中的類別名稱。如果您沒有定義變數標記的話，變數名稱會被當成標記使用。在多類別集合中，組別中第一個變數的數值標記會被當作類別標記。如果第一個變數沒有某些類別，但組別中的其他變數有這些類別的話，請為遺漏的類別定義數值標記。

遺漏值。 會替每個表格排除含有遺漏值的觀察值。您也可以任選下列其中一個選項（或者兩樣都選）：

- **排除二分法內列出之觀察值。** 如果變數的觀察值包含遺漏值，則會將這些觀察值排除在多重二分集的表格外面。如果複選題集已經定義成二分變數集合的話，就可以使用這個選項。根據預設值，如果觀察值的成份變數都不包含計數值的話，這個觀察值就會被當成是多重二分集的遺漏值。對於含有遺漏值的觀察值（只是部分變數的觀察值含有遺漏值）而言，它們會被納入群組表格中（如果至少有一個變數含有計數值的話）。
- **排除類別內列出之觀察值。** 如果變數的觀察值包含遺漏值，SPSS 會將這些觀察值排除在多類別集合的表格外面。如果複選題集已經定義成類別變數集合的話，就可以使用這個選項。根據預設值，如果觀察值的任何一個成分值都未落入定義範圍內的話，這個觀察值就會被當成多類別集合的遺漏值。

範例。 根據問卷的問題所建立的每個變數都是基本變數。若要分析複選題集項目，您必須把變數結合成兩種複選題集之一：多重二分集或多類別集合。例如，如果一家航空公司詢問受訪者最近六個月內、曾經搭乘三家航空公司（中華、遠東、長榮）中的那幾家。這時如果使用二分變數，並且定義**多重二分集**的話，則集合中的三個變數都會變成組別變數的類別，而且受訪者搭乘三家航空公司的次數和百分比會顯示在一個次數分配表裏。如果您發現，受訪者所提到的航空公司都不超過兩家的話，就可以建立兩個變數，而且每個變數有三個代碼，每個代碼代表一家航空公司。如果您定義**多重類別集**的話，那麼會根據基本變數中，相同代碼的總和，將值列示出來。結果會發現，總和值跟每個基本變數的值加總後相同。例如 30人回答中華航空，是第一家航空公司有五人提到中華，再加上第二家航空公司有 25人回答中華航空。 三家航空公司的次數和百分比會顯示在一個次數分配表裏。

統計量。 顯示次數分配表，其內含有個數、依變數百分比、觀察值百分比、有效觀察值個數和遺漏觀察值個數。

資料。 使用複選題集。

假設。 個數和百分比會提供任何分配資料的說明資訊。

相關程序。「複選題定義集」程序讓您可以定義複選題集。

若要取得複選題次數分配表

▶ 從功能表選擇：
分析 > 複選題 > 次數分配表...

圖表 28-2
「複選題次數分配表」對話方塊

▶ 選取一個或多個複選題集。

複選題交叉表

「複選題交叉表」程序會製作定義的複選題集或基本變數（或兩者組合）的交叉表列。您也可以取得儲存格百分比（以觀察值或回答為基礎）、修改處理遺漏值的方式，或者取得成對的交叉表列。但是在進行上述事項之前，都必須先定義一個（或多個）複選題集（請參閱〈若要定義複選題集〉）。

針對多重二分集，使用組別中、基本變數定義的變數標記，當作輸出中的類別名稱。如果您沒有定義變數標記的話，變數名稱會被當成標記使用。在多類別集合中，組別中第一個變數的數值標記會被當作類別標記。如果第一個變數沒有某些類別，但組別中的其他變數有這些類別的話，請為遺漏的類別定義數值標記。程序會將行類別標記分成三行顯示，每行最多顯示八個字元。若要避免字被切斷，您可以反轉列和行項目，或者重新定義標記。

範例。 本程序可以製作多重二分集合(或多類別集合）跟其他變數的交叉表列。航空公司詢問受訪者下列問題：「請將最近六個月內，您至少搭乘過一次的航空公司（中華、遠東、長榮）全部圈選出來。還有當您選取班機時，哪項因素最重要？班次或是服務？請只選取一項。在您將資料當成二分變數或多重類別輸入之後，(並把它們併成一組)，就可以製作航空公司選項跟服務（或班次）相關問題的交叉表列。

統計量。 交叉表列（其內含有儲存格、列、行和總個數）以及儲存格、列、行和總百分比。您可以根據觀察值或依變數來決定儲存格百分比。

資料。 使用複選題集或數值類別變數。

假設。 個數和百分比會提供任何分配資料的說明資訊。

相關程序。「複選題定義集」程序讓您可以定義複選題集。

若要取得複選題交叉表

- 從功能表選擇：
  分析 > 複選題 > 交叉表...

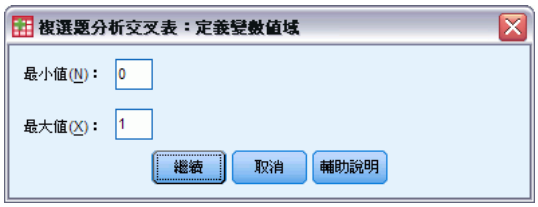
圖表 28-3
「複選題交叉表」對話方塊

- 替交叉表列的每個維度選取數字變數（一個或多個）或複選題集。
- 定義所有基本變數的範圍。

  (選擇性）您可以決定是否取得控制變數（或複選題集的所有類別）的二因子交叉表列。請從「層」清單上選取一個（或多個）項目。

複選題交叉表定義變數範圍

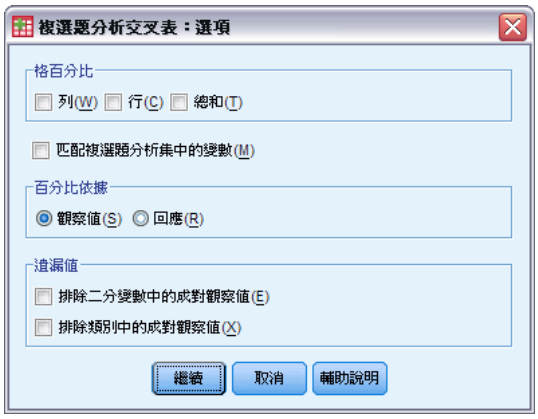
圖表 28-4
「複選題交叉表定義變數範圍」對話方塊



您必須定義交叉表列中，任何基本變數的數值範圍。請輸入該交叉表列的最小、最大整數類別值。不會分析範圍以外的類別。它同時也假設範圍裡面的數值都是整數（非整數會被截斷）。

複選題交叉表選項

圖表 28-5
「複選題交叉表選項」對話方塊



格百分比。永遠都會顯示儲存格個數。但是您可以選擇是否顯示列百分比、行百分比和二因子表格（總計）百分比。

百分比依據。 您可以根據觀察值（或受訪者）來決定格百分比。如果您橫跨多重類別集合以選出匹配變數的話，就無法使用這個選項。您也可以根據依變數決定格百分比。對於多重二分集而言，回答個數會等於不同觀察值之間的計數值個數。對於多類別集合而言，依變數個數代表定義範圍中的數值個數。

遺漏值。您可以選擇下列其中一個選項，或兩者都選：

- **排除二分法內列出之觀察值。** 如果變數的觀察值包含遺漏值，則會將這些觀察值排除在多重二分集的表格外面。如果複選題集已經定義成二分變數集合的話，就可以使用這個選項。根據預設值，如果觀察值的成份變數都不包含計數值的話，這個觀察值就會被當成是多重二分集的遺漏值。對於含有遺漏值的觀察值（只是部分變數的觀察值含有遺漏值）而言，它們會被納入群組表格中（如果至少有一個變數含有計數值的話）。
- **排除類別內列出之觀察值。** 如果變數的觀察值包含遺漏值，SPSS 會將這些觀察值排除在多類別集合的表格外面。如果複選題集已經定義成類別變數集合的話，就可以使用這個選項。根據預設值，如果觀察值的任何一個成分值都未落入定義範圍內的話，這個觀察值就會被當成多類別集合的遺漏值。

根據預設值，當製作兩個多類別集合的交叉表列時，程序會將第一組中的每個變數，跟第二組中的每個變數交叉分析，並將儲存格所含個數加起來；因此，表格中的某些依變數會出現一次以上。您可以選擇的選項如下：

匹配複選題選項集間的變數。 會把第一組中的第一個變數，跟第二組中的第一個變數配對，然後依此類推。如果您選取這個選項的話，程序會根據回答算出格百分比（而不是根據受訪者）。多重二分集或基本變數都無法使用配對選項。

MULT RESPONSE 指令的其他功能

指令語法語言讓您也可以：

- 取得交叉表列，最多可含五個維度（使用 BY 副命令）。
- 改變輸出格式選項，其中一個選項可以不顯示數值標記（使用 FORMAT 副命令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
29

報告結果

觀察值清單和敘述統計，是研究和顯示資料的基本工具。您可以使用「資料編輯程式」或「摘要」程序來取得觀察值清單，使用「次數」程序來取得次數個數和敘述統計，以及使用「平均數」程序來取得子母體統計量。這些格式都是經過特殊設計，可以很清楚地呈現資訊。如果您想要以不同的格式來顯示這些資訊的話，則可以透過控制項：「報表中列摘要」和「報表中行摘要」，來控制呈現資料的方式。

報表中列摘要

「報表中列摘要」會產生一份報表，該表會將不同摘要統計量放到列裡面。也會有可使用的觀察值清單（但是不見得會附上摘要統計量）。

範例。 某家公司擁有連鎖零售商店，並記錄了員工資訊，包括薪資、在職期間，以及該名員工所屬之商店或部門。您可以產生一份報表，它根據商店和部門（分段變數），將員工資訊（清單）分類，並附上每家商店、每個部門以及每家商店中各部門的摘要統計量（例如平均薪資）。

資料行。 它會列出需要製作觀察值清單，或是摘要統計量的報表變數，並控制資料欄的顯示格式。

分割行。 它會列出選擇性分段變數，該變數會將報表分組，並且控制摘要統計量和分段欄的顯示格式。對於多個分段變數來說，此清單中，上個分段變數的類別內，每個分段變數的各個類別都會自成一組。分段變數應該是非連續類別變數，可將觀察值分成數個有意義的類別。每個分段變數的獨立值，會顯示並排序在所有資料行左邊的個別行中。

報表。 控制整張報表的特性，包括全部摘要統計量、是否顯示遺漏值、編頁數及標題等。

顯示觀察值。 顯示每個觀察值之資料欄變數的實際值（或數值標記）。如此會產生一份清單報表，該表可能比摘要報表更長。

預覽。 只顯示報表的第一頁。當您要預覽報表的格式，但不想處理整張報表時，此選項非常有用。

資料已經排序。 對於含有分段變數的報表而言，您必須在產生報表之前，先將資料檔按照分段變數值排序。如果資料檔已經按照分段變數值排序，則藉由選取這個選項，可以節省不少時間。當您預覽過報表之後，這個選項就特別有用。

若要取得摘要報表:列中的摘要

- ▶ 從功能表選擇：
  分析(A) > 報表 > 列的報表摘要...
- ▶ 選取「資料行」中一個（或多個）變數。所選取的每個變數，都會在報表中產生一個行。

- 如果報表是按照組別排序，並加以顯示的話，請選取「分割行」中的一個（或多個）變數。
- 對於由分段變數定義次組別的摘要統計報表，請在「分割行變數」清單選取分段變數並按一下「分割行」群組中的「**摘要**」，以指定摘要量數。
- 對於包含整體摘要統計量的報表，請按一下「**摘要**」以指定摘要量數。

圖表 29-1
「報表中列摘要」對話方塊

報表資料行/分段格式

「格式」對話方塊可控制行標題、行寬、文字對齊方式，還有顯示資料值或觀察值註解的方式。「資料行格式」可控制報表頁面右邊「資料行」的格式。「分段格式」則控制左邊「分割行」的格式。

圖表 29-2
「報表資料行格式」對話方塊

行標題。 控制選用變數的行標題。該行內的長標題會自動換行。如果想手動地插入分行符號，請在標題需要換行的地方，按下 Enter 鍵。

「行」內「數值位置」。 這個選項會控制選用變數，其行內資料值或數值標記的對齊方式。是否對齊數值或註解，並不會影響行標題的對齊方式。您可以根據所指定之字元數，將行內容縮排，或者讓內容向中對齊。

行內容。 這個選項會控制選用變數，其資料值或定義數值標記的顯示方式。任何未定義數值標記的值一律會顯示資料值。(行摘要報表中的資料行無法使用這個選項)。

報表摘要行所屬/最終摘要行

此處的兩個「摘要行」的對話方塊，控制分段組別和整份報表之摘要統計量的顯示方式。「摘要行」則控制每個類別的次組別統計量，而類別則是根據分段變數來定義的。「最終摘要行」控制報表結尾處所顯示的所有統計量。

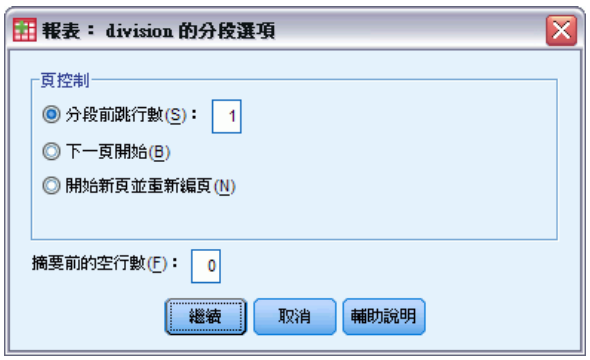
圖表 29-3
「報表摘要行」對話方塊

您可以使用的摘要統計量包括：總和、平均數、最小值、最大值、觀察值個數、指定值之上或之下的觀察值百分比、指定數值範圍內的觀察值百分比、標準差、峰度、變異數和偏態等。

報表分段選項

「分段選項」控制分段類別資訊的間距和頁數。

圖表 29-4
「報表分段選項」對話方塊



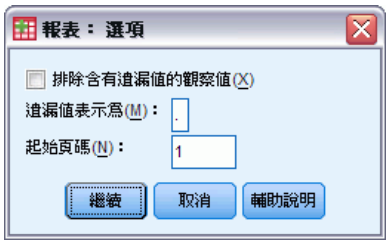
頁控制。 控制選用分段變數的類別間距和頁數。您可以指定分段類別之間的空白行，或者在新頁面上，開始每個分段類別。

摘要前空白行。 控制分段類別標記，還有資料和摘要統計量之間的空白行。如果報表包含分段類別的個別觀察值清單，以及摘要統計量時，這個選項就特別有用；在這些報表中，您可以在觀察值清單和摘要統計量之間，插入空格。

報表選項

「報表選項」控制處理、顯示遺漏值和頁碼的方式。

圖表 29-5
「報表選項」對話方塊



排除含有遺漏值的觀察值。 如果任何報表變數之觀察值含有遺漏值的話，就（自報表）加以刪除。

遺漏值以…顯示 您可以在此指定符號，用來代表資料檔中的遺漏值。這個符號只有一個字元也沒關係，還是可以用來代表**系統遺漏值**和**使用者遺漏**值。

頁數從…。 讓您指定報表第一頁的頁碼。

報表配置

「報表配置」控制每個報表頁面的寬度和長度、報表在頁面上的位置，以及如何插入空行和標記。

圖表 29-6
「報表配置」對話方塊

頁外觀。 控制頁面邊緣，該邊緣是以行（頂端和底部）和字元（左和右）來代表，這個選項還控制邊緣中報表的對齊方式。

頁標題與註腳。 控制報表內文各頁標題與註腳的行數。

分割行。 控制分割行的顯示方式。如果指定多重分段變數，則這些變數可以被放在不同行，或第一個行裡面。如果將所有分段變數都放在第一個行裡面的話，所產生的報表會比較窄。

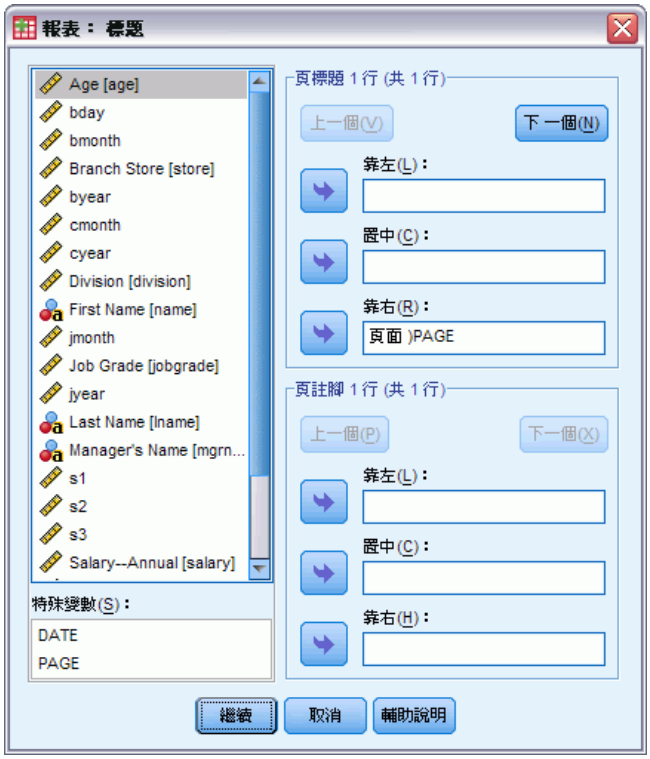
行標題。 控制行標題的顯示方式，其中包括標題是否加底線、標題和報表內文之間的空格數，以及行標題是否垂直對齊。

資料行列與分段標記。 控制資料行資訊（資料值和/或摘要統計量）以及每個分段類別開始處之分段標記相對位置。資料行資訊的第一列，可以跟分段類別標記同行，或者放在自分段類別標記算起，某指定行數上。(行摘要報表無法使用這個選項)。

報表標題

「報表標題」控制報表標題、頁底的內容與位置。頁面標題和頁底最多可以指定 10 行，而且每一行都有向左調整、置中和向右調整等元件。

圖表 29-7
「報表標題」對話方塊



如果您在標題和頁底中插入變數，則目前的觀察值註解、或該變數值，都會出現在標題或頁底中。標題也會顯示觀察值註解，該註解跟頁面開始處變數值相對應。頁底所顯示的觀察值註解，則是跟頁面結束處變數值相對應。若沒有數值標記，會顯示實際值。

特殊變數。 您可以利用特殊變數 DATE 和 DATE，將目前的日期或頁碼，插入報表頁首或頁底任何一行內。如果您的資料檔包含 DATE 或 PAGE 變數，就無法在報表標題或頁底中，使用這些變數。

報表中行摘要

「報表中行摘要」會產生不同摘要統計量，而且這些統計量會顯示在摘要報表的各個行中。

範例。某家公司擁有連鎖零售商店，並記錄了員工資訊，包括薪資、在職期間，以及該名員工所屬之商店或部門。您可以產生提供各部門的摘要薪資統計量（例如，平均數、最小值、最大值）的報表。

資料行。 列出需要製作摘要統計量的報表變數，並控制每個變數的顯示格式和摘要統計量。

分割行。它會列出選擇性分段變數，該變數會將報表分組，並控制分割行的顯示格式。對於多個分段變數來說，此清單中，上個分段變數的類別內，每個分段變數的各個類別都會自成一組。分段變數應該是非連續類別變數，可將觀察值分成數個有意義的類別。

報表。 控制整張報表的特性，包括遺漏值的顯示方式、編頁數和標題等。

預覽。 只顯示報表的第一頁。當您要預覽報表的格式，但不想處理整張報表時，此選項非常有用。

資料已經排序。 對於含有分段變數的報表而言，您必須在產生報表之前，先將資料檔按照分段變數值排序。如果資料檔已經按照分段變數值排序，則藉由選取這個選項，可以節省不少時間。當您預覽過報表之後，這個選項就特別有用。

若要取得摘要報表：行中的摘要

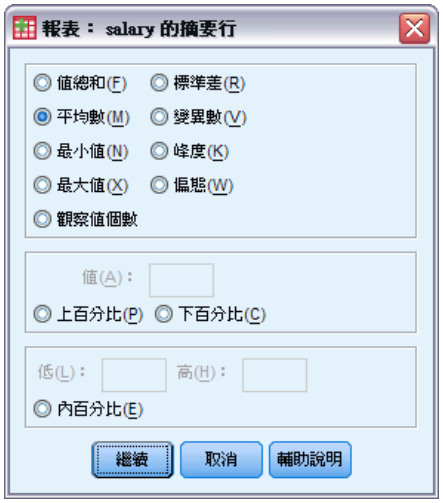
- 從功能表選擇：
  分析(A) > 報表 > 欄的報表摘要...
- 選取「資料行」中一個（或多個）變數。所選取的每個變數，都會在報表中產生一個行。
- 若要變更變數的摘要量數，請選取「資料行變數」清單中的變數，然後按一下「**摘要**」。
- 若要取得變數的摘要量數（一個以上），請從來源清單中選出變數，並重複將此變數移到「資料行變數」清單中，以便逐一分配給每個摘要量數。
- 若要顯示現存行總和、平均數、比例量數、或其他函數的話，請按一下「**插入總和**」。這會將稱為 total 的變數，放到「資料行」清單中。
- 如果報表是按照組別排序，並加以顯示的話，請選取「分割行」中的一個（或多個）變數。

圖表 29-8
「報表中行摘要」對話方塊

資料行摘要函數

您可以利用「摘要行」選項，來控制選用資料行變數所顯示之摘要統計量。

圖表 29-9
「報表摘要行」對話方塊



您可以使用的摘要統計量包括：總和、平均數、最小值、最大值、觀察值個數、指定值之上或之下的觀察值百分比、指定數值範圍內的觀察值百分比、標準差、峰度、變異數和偏態等。

行總和的資料行摘要

「摘要行」選項可以控制兩個以上資料行的摘要統計量總和。

您可以使用的摘要統計量總和包括了：行總和、行平均數、最小值、最大值、兩行中數值的差、一行內的數值除以另一行內數值所得的商，以及行數值的乘積。

圖表 29-10
「報表摘要行」對話方塊

行總和。total 行代表「摘要行」清單中行的總和。

行平均數。total 行代表「摘要行」清單中行的平均數。

行最小值。total 行代表「摘要行」清單中行的最小值。

行最大值。total 行代表「摘要行」清單中行的最大值。

第一行 - 第二行。total 行代表「摘要行」清單中的行差。「摘要行」清單必須確實包含兩個行。

第一行 / 第二行。total 行代表「摘要行」清單中行的商。「摘要行」清單必須確實包含兩個行。

% 第一行 / 第二行。total 行代表「摘要行」清單中，第二行在第一行中，所佔之百分比。「摘要行」清單必須確實包含兩個行。

行結果。total 行代表「摘要行」清單中行的積。

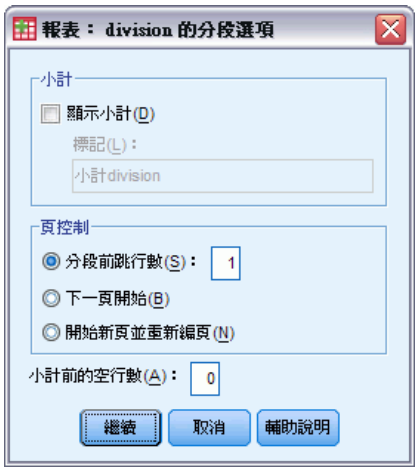
報表行格式

「報表中行摘要」的資料選項和分割行格式化選項，都跟「報表中列摘要」中，所描述的選項一樣。

行分段選項中的報表摘要

「分段選項」控制分段類別小計之顯示方式、間距和頁數。

圖表 29-11
「報表分段選項」對話方塊



小計。 控制分段類別小計的顯示方式。

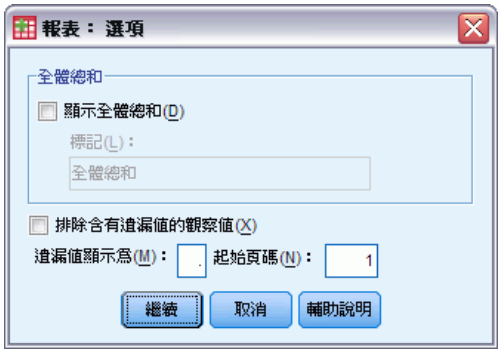
頁控制。 控制選用分段變數的類別間距和頁數。您可以指定分段類別之間的空白行，或者在新頁面上，開始每個分段類別。

小計前的空白行。 控制分段類別資料和小計之間的空白行。

行選項中的報表摘要

「選項」控制全體總和的顯示方式、遺漏值的顯示方式，以及行摘要報表中的頁數。

圖表 29-12
「報表選項」對話方塊



全體總和。 這個選項會顯示（在該行的底部）和標記每個行的全體總和 。

遺漏值。 您可以將遺漏值排除在報表之外，或者選取單一字元，並用它來代表報表中的遺漏值。

行中摘要的報表配置

「報表中行摘要」的報表配置選項，跟「報表中列摘要」中所描述的選項是一樣的。

REPORT 指令的其他功能

指令語法語言也可以讓您：

- 在單一摘要行的行中，顯示不同的摘要函數。
- 將摘要行插入變數的資料行中，而不是資料行變數中，也不是每種摘要函數組合（合成函數）的資料行中。
- 使用「中位數」、「眾數」、「次數」、「百分比」，當作摘要函數。
- 更精確地控制摘要統計量的顯示格式。
- 在報表的每個點上，插入空白行。
- 在清單報表中的每 n 個觀察值之後，插入空白行。

因為 REPORT 語法很複雜，所以當您使用此語法建立新報表（讓報表看起來像是透過對話方塊所產生的），複製或貼上對應語法，還有改進該語法，以產生所需之正確報表時，就會發覺它很有用。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
30

信度分析

信度分析可讓您研究測量尺度的性質，以及組成他們的項目。信度分析程序會計算一些常用的尺度信度量數，同時也會提供有關尺度中個別項目之間關係的資訊。組內相關係數可用來計算定等級者之間的信度估計值。

範例。 我的問卷是以有用的方式來測量消費者滿意度嗎？使用信度分析，您就可以確定問卷中項目彼此有關的程度，而從整個看來，您可以得到信度的概要指標或尺度的內部一致性，而且您還可以識別應排除於該尺度之外的問題項目。

統計量。 計有每個變數與尺度的描述性統計量、所有項目的摘要統計量、項目內之相關值與共變異數、信度估計值、ANOVA 摘要表、組內相關係數、Hotelling's T^2 及 Tukey 的可加性測試。

模式。 下列信度模式是可選用的：

- **Alpha (Cronbach) 值。** 這是內部一致性的模式，以平均項目之相關值為根據。
- **折半信度。** 本模式會將尺度分跚
- **Guttman 值。** 此模式會計算真實信度的 Guttman 值下限。
- **平行模式檢定。** 此模式在所有複製過程，會假設所有項目均有相等的變異數及相等的誤差變異數。
- **嚴密平行模式。** 此模式會做出「平行」模式之假設，並假設所有項目均有相等的平均數。

資料。 資料可以是二分的、次序的或區間的，但資料必須是用數值表示的。

假設。 觀察值應該是獨立的，且誤差應該與觀察項目無關。每一對項目應該具有雙變數常態分配。尺度應該具有可加性，以便每個項目都與總分數線性相關。

相關程序。 如果您想要探究您的尺度項目的維度 (查看是否需以一個以上的概念來說明項目分數的樣式)，請使用因子分析或多元尺度方法。若要識別變數的同質組別，您可以在集群變數上使用階層集群分析法。

取得信度分析

▶ 從功能表選擇：
分析(A) > 尺度 > 信度分析...

圖表 30-1
「信度分析」對話方塊

- 選取當作可加性尺度可能元件的二或多個變數。
- 從「模式」下拉式清單中選擇一種模式。

信度分析統計量

圖表 30-2
「信度分析：統計量」對話方塊

您可以選取不同的統計量以描述您的尺度及項目。依預設值描述的統計量包括下列觀察值個數、項目個數及信度估計值：

- **Alpha 值模式。** alpha 係數；對二分資料而言，這就相當於 Kuder-Richardson 20 (KR20) 係數。
- **折半信度模式。** 共有形式間相關、Guttman 折半信度、Spearman-Brown 信度 (相等及不相等長度)、及每個折半的 alpha 值。
- **Guttman 模式。** 信度係數從 lambda 1 到 lambda 6。
- **平行與嚴密平行模式。** 共有模式適合度檢定、誤差變異數、共同變異數及真實變異數的估計值、估計的共同項目之間相關、估計的信度及不偏的信度估計值。

描述性統計量。 會產生所有觀察值尺度或項目的敘述統計。

- **項目。** 會產生所有觀察值項目的敘述統計。
- **尺度。** 產生尺度的描述性統計量。
- **刪除項目後之尺度摘要** 顯示每個項目與由其他項目所構成的尺度相比較之摘要統計量。這些統計量包括尺度平均數和變異數 (如果是從尺度、項目和由其他項目所構成之尺度間的相關中刪除項目)，以及 Cronbach’s alpha 值 (如果項目是由尺度中刪除)。

摘要。 可提供尺度中所有項目的項目分配敘述統計。

- **平均數。** 項目平均數的摘要統計量。系統會顯示最小、最大和平均項目平均數、項目平均數的範圍與變異數，以及最大與最小項目平均數的比例。
- **變異數(V)。** 項目變異數的摘要統計量。系統會顯示最小、最大及平均項目變異數，項目變異數的範圍與變異，以及最大到最小項目變異數的比例。
- **共變異數(E)。** 項目之間共變異數的摘要統計量。系統會顯示最小、最大及平均項目之間共變異數、項目之間共變異數的範圍與變異，以及最大與最小項目之間共變異數的比例。
- **相關(R)。** 項目之間相關的摘要統計量。系統會顯示最小、最大及平均項目之間相關、項目之間相關的範圍與變異，以及最大與最小項目之間相關的比例。

各分量表內項目之間。 會產生項目之間相關或共變異數矩豹

變異數分析表。 程序檢定相等的平均數。

- **F 檢定(F)。** 顯示重複的變異數分析表量數。
- **Friedman 卡方(Q)。** 顯示 Friedman’s 卡方和 Kendall’s 和諧係數。此選項適用於採等級形式的資料。在 ANOVA 表中，卡方檢定會取代一般的 F 檢定。
- **Cochran 卡方(H)。** 顯示 Cochran’s Q。此選項適用於二分資料。在 ANOVA 表中，Q 統計量會取代一般的 F 統計量。

Hotelling’s T 平方。 會產生虛無假設 (尺度上所有項目均有相同平均數) 的多變量檢定。

Tukey 的可加性測試。會產生假設 (項目間沒有相乘性交互作用) 的檢定。

組內相關係數 會產生觀察值間數值一致性或協定的量數。

- **模式。** 可選取計算類組間相關係數的模式。可用的模式有「二因子混合」、「二因子隨機」及「單因子隨機」。當人的效應項為隨機而項目效應項為固定時，選取「二因子混合」；當人的效應項和項目效應項均為隨機時，選取「二因子隨機」，或當人的效應項為隨機時，選取「單因子隨機」。
- **類型。** 可選取索引類型。可用的類型有「一致性」及「絕對協定」。

- **信賴區間。** 可指定信賴區間的水準。預設值是 95%。
- **檢定值。** 可指定假設檢定的係數假設值。此數值將用來與觀察值做比較，預設值為 0。

RELIABILITY 指令的其他功能

指令語法語言也可以讓您：

- 讀取及分析相關矩陣。
- 寫入相關矩陣以備日後分析之用。
- 指定折半方法除了均等各半之外的分割方式。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
31

多元尺度方法

多元尺度方法試圖尋找物件或觀察值間距離量數集的結構，可藉著將觀察值指定到概念上的空間（通常為二或三個維度）中特定位置來達成此工作，而空間中的點間距須儘可能符合規定的相異性。在許多例子中，這項概念上空間的維度可解釋並用來進一步瞭解您的資料。

如果您有客觀測量的變數，就可將多元尺度方法當作資料縮減技術使用（如果有必要，多元尺度方法程序將會為您計算多變量資料的距離），另多元尺度方法也可以應用在物件或概念間的主觀等級上。此外，多元尺度方法程序可處理來自多個來源的相異性資料，例如來自多位評估者或問卷應答者的資料。

範例。人們如何察覺不同汽車間的關係？如果您有指出不同汽車廠牌及型號之間相似性等級的應答者資料，就可使用多元尺度方法來確定可敘述消費者接受度的維度。例如，您可能會發現汽車價格及大小可定義一個二維空間，以說明應答者所回報的相似性。

統計量。 對於每個模式：資料矩陣、最佳調整的資料矩陣、S 應力 (Young‘s)、應力 (Kruskal’s)、RSQ、刺激座標、平均應力及各項刺激的 RSQ (RMDS models)。適用個別差異 (INDSCAL) 模式：每個受試者的加權值及離奇性索引。適用複製的多元尺度方法模式中各矩陣：每項刺激的應力及 RSQ。圖形(T)：刺激座標 (二或三個維度)、不等性對距離的散佈圖。

資料。如果您的資料是相異性資料，那所有相異性都應為數值，並以相同的計量單位測量。如果您的資料是多變量資料，那麼變數可以是數值、二元或個數資料。變數尺度是一項重要問題——因為尺度的差異可能會影響您的解答。如果您的變數尺度差異極大（例如，一個變數以元為單位來測量，另一個以年為單位），那就應考慮將他們標準化（可使用多元尺度方法程序來自動執行）。

假設。多元尺度方法程序比較沒有分配上的假設，但一定要在「多維度尺度選項」對話方塊中選擇適當的測量水準 (次序的、區間或比例量數)，以確保結果係經過正確計算。

相關程序。如果您的目的是資料縮減，特別是您的變數為數值的話，可考慮使用因子分析的替代方法。如果您想要確認相似觀察值的組別，請考慮以階層或 k 平均數集群分析來補充多元尺度方法的分析。

取得多元尺度方法的分析

▶ 從功能表選擇：
分析(A) > 尺度 > 「多元尺度方法...」

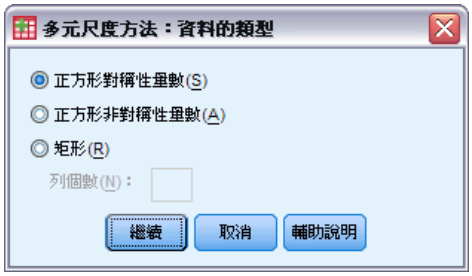
圖表 31-1
「多元尺度方法」對話方塊

- 選取至少四個數值變數進行分析。
- 在「距離」群組中，選擇「**資料為歐基里得直線距離**」或「**將資料轉換成歐基里得直線距離**」其中之一。
- 如果您選取「**將資料轉換成歐基里得直線距離**」，您也可以為個別的矩陣選取分組變數。分組變數可以是數字或字串。

  您也可以：
  - 當資料為歐基里得直線距離時，指定距離矩陣的形狀。
  - 從資料建立歐基里得直線距離時指定距離量數。

多元尺度方法資料的類型

圖表 31-2
「多元尺度方法類型」對話方塊



如果您的作用中資料集代表一組物件之內的距離或代表兩組物件之間的距離，請指定資料矩陣的類型，以便得到正確的結果。

注意：如果「模式」對話方塊指定了列的條件性，您就無法選取「**正方形對稱性量數**」。

多元尺度方法建立測量法

圖表 31-3
「多元尺度方法：從資料建立測量法」對話方塊

多元尺度方法使用相異性資料來建立尺度法解答。如果您的資料是多變量資料（已測量變數的數值），您就必須建立相異性資料，以便計算多元尺度方法的解答。您也可以指定從資料建立相異性量數的詳細內容。

量數。可讓您指定分析所需的相異性量數，從對應您的資料類型的「測量」組別中，選擇一個選項，然後從對應那項測量類型的下拉式清單中選取一種量數。可用的選項有：

- **區間**。共有歐基里得直線距離、歐基里得直線距離平方、Chebychev、區塊、Minkowski 或自訂式。
- **個數**。共有卡方量數或 Phi-square 量數。
- **二進位**。共有歐基里得直線距離、歐基里得直線距離平方、大小差異、形式差異、變異數或 Lance 與 Williams。

距離矩陣的計算法。可讓您選擇分析的單位，選項共有「變數之間」或「觀察值之間」。

轉換值。在某些例子中，例如，當變數是以差異甚大之尺度來測量時，您可能會想在計算近似性（無法應用在二進位資料）之前先將數值標準化。從「標準化」下拉清單上，選取一個標準化方法。如果不需要任何標準化，請選擇「無」）。

多元尺度方法的模式

圖表 31-4
「多元尺度方法：模式」對話方塊

多元尺度方法模式的估計正確與否，取決於資料的樣子及模式本身。

測量水準。可讓您指定資料的水準，選項共有「次序的」、「等距量數」或「比例量數」。如果您的資料是次序的，選擇「**將同分觀察值變為不同分**」會要求將他們視為連續的變數，以便能最佳化地分解同分（不同觀察值的相同數值）。

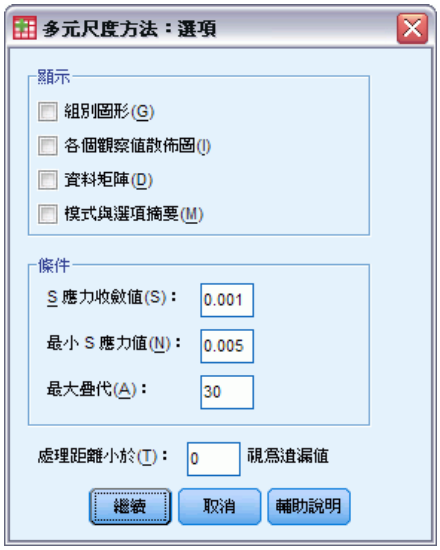
條件性。可讓您指定哪些比較是有意義的，選項共有「矩陣」、「列」或「無條件性限制」。

維度。可讓您指定尺度法解答的維度範圍，對範圍內的每項數字，計算出一項解答。指定 1 到 6 之間的整數，但只有在您將「**歐基里得直線距離**」選定為尺度模式時，才能指定最小值 1。對單一解答，請將最小值及最大值均指定為相同之數字。

尺度模式。可讓您指定尺度運作所依循之假設，可用之選項共有「歐基里得直線距離」或「個別差異歐基里得直線距離」(也稱為 INDSCAL)。對個別差異歐基里得直線距離模式而言，如果適合您的資料，您就可選取「**允許負的加權值**」。

多元尺度方法的選項

圖表 31-5
「多元尺度方法：選項」對話方塊



在此對話方塊中，您可以指定多元尺度方法分析的選項。

顯示。可讓您選取不同類型之輸出，可用之選項共有「組別圖形」、「各個觀察值散佈圖」、「資料矩陣」及「模式與選項摘要」。

條件。可讓您決定何時應停止疊代。若要變更預設值，請輸入 **S 應力收斂值**、**最小 S 應用值**及**最大疊代** 的數值。

歐基里得直線距離小於 n 視為遺漏值。 小於這項數值的距離會從分析中排除。

多維度尺度法指令的其他功能

指令語法語言也可以讓您：

- 使用三項額外的模式類型，即大家所熟知的 ASCAL、AINDS、及 GEMSCAL 多元尺度方法。
- 完成等距與比例資料的多項式轉換。
- 分析次序的資料相似性（非距離）。
- 分析名義資料。
- 將座標及權數矩陣儲存至檔案中，並可將他們讀取回來用於分析。
- 限制多維度展開。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
32

比例量數統計量

「比例量數統計量」程序可提供摘要統計量的完整清單，以描述兩個尺度變數間的比例量數。

您可以藉由分組變數的值，將輸出以遞增或遞減的順序排序。比例量數統計量報表可以不輸出，而將結果儲存在外部檔案。

範例。 在五個縣市中，每個縣市住家房屋之預估價格和銷售價格之間的比例量數是否可以均勻分配？從輸出中你可以發現，比例量數的分配在每個縣市間有相當大的不同。

統計量。 中位數、平均數、加權平均數、信賴區間、 離散係數 (COD)、以中位數為中心的變數係數、以平均數為中心的變數係數、價格關聯微分 (PRD)、標準差、平均絕對離差 (AAD)、範圍、最小值和最大值，以及針對使用者指定範圍或中位數比例量數內百分比的集中度指標。

資料。 使用數值代碼或字串做為分組變數的代碼 (名義或次序的水準測量)。

假設。 定義比例量數的分子和分母之變數，應該是使用正值的尺度變數。

若要取得比例量數統計量

▶ 在功能表上，選擇：

分析(A) > 敘述統計 > 比率...

圖表 32-1
比例量數統計量對話方塊

▶ 選取分子變數。

▶ 選取分母變數。

隨意：

- 選擇分組變數並指定結果中群組的排序方式。
- 選取是否在「瀏覽器」顯示結果。
- 選取是否要將結果儲存到外部檔以供日後使用，並指定儲存結果的檔名。

比例量數統計量

圖表 32-2
比例量數統計量對話方塊

集中趨勢。集中趨勢的測量是說明比例量數分散情形的統計量。

- **中位數。** 不論比例量數的數目少於這個值，或比例量數的數目大於這個值，該值都一樣。
- **平均數。** 比例量數加總並除以比例量數總數目的結果。
- **加權平均數。**把分子的平均數除以分母的平均數之結果。加權平均數也是使用分母加權後的比例量數平均數。
- **信賴區間。** 會顯示平均數、中位數及加權平均數的信賴區間（如果您要求的話）。指定一個大於或等於 0 且小於 100 的數值作為信賴水準。

分散情形 。這是在觀察值中測量變數量或離散量的統計量。

- **AAD。** 平均絕對離差是將中位數相關比例量數的絕對離差加總，並除以總比例量數數目的結果。
- **COD。** 離散係數是將絕對離差表示成中位數百分比的結果。
- **PRD。** 價格關聯微分（另稱為迴歸的指標），是平均數除以加權平均數的結果。

- **以中位數為中心的變異係數。** 以中位數為中心之變異係數，是將中位數離差之平均數平方根，表示成中位數百分比的結果。
- **以平均數為中心的變異係數。**以平均數為中心之變異係數，是將標準差表示成平均數百分比的結果。
- **標準差。** 標準差是將平均數相關比例量數的離差平方加總，除以比例量數總數目減 1，並使用正平方根的結果。
- **範圍。**範圍是將最大值比例量數減去最小值比例量數的結果。
- **最小值。**最小值是最小比例量數。
- **最大值。**最大值是最大比例量數。

集中度指標。集中係數可測量落在區間內的比例量數百分比。可透過使用以下兩種分法來計算：

- **比例量數介於。** 這裡必須明確定義區間的最低值和最高值。請輸入低比例和高比例的值，然後按一下「新增」以取得區間。
- **比例量數包含於。** 這裡不須明確定義區間，只需指定中位數的百分比。請輸入一個介於 0 到 100 之間的數值，然後按一下「新增」。區間的低值等於 (1 - 0.01 × 值) × 中位數，而高值則等於 (1 + 0.01 × 值) × 中位數。

章節
33

ROC 曲線

這個程序是評估分類綱要效能的有效方式，在此綱要中，其中一個變數會有兩種類別，端視所分類的受試者而定。

範例。 銀行依利息將客戶恰分成兩種，會借貸者與不會借貸者，因此便發展出這種特殊方法，以執行這類判斷。ROC 曲線可用來評估這些方法執行的效果。

統計量。 在 ROC 曲線底下的區域，包含 ROC 信賴區間和座標點。圖形(T)：ROC 曲線。

方法。 在 ROC 曲線下的評估區域，可以使用雙負 (binegative) 指數模式，做非參數式 (nonparametrically) 或參數式 (parametrically) 的計算。

資料。 檢定變數為數值變數。檢定變數通常是由判別分析的機率、logistic 迴歸、或任意尺度上的分數所組成，此尺度上會表示定等級者會落在某一類別或其他類別的「信賴強度」。狀態變數可以是任何類型的變數，但要指出受試者所屬的真實類別。狀態變數的值會指出哪一個類別應視為正向。

假設。 假設定等級者尺度上數值的增加，代表相信受試者屬於某類別的程度加深，則尺度上數值的減少，則代表相信受試者屬於其他類別的程度增加。使用者必須選擇正向。同時假設已知各受試者屬於真的類別。

若要取得 ROC 曲線

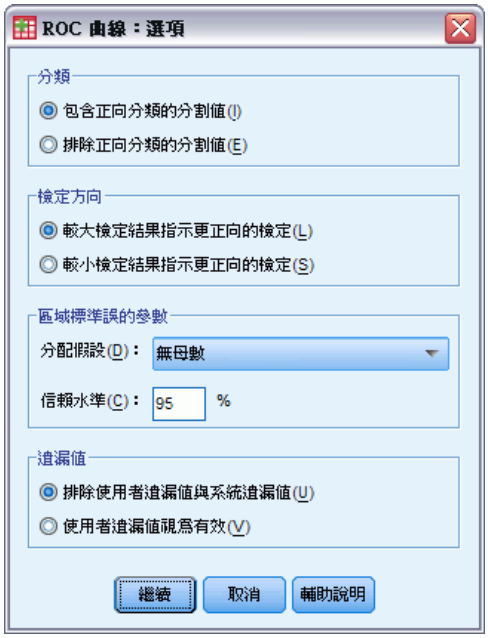
▶ 從功能表選擇：
分析(A) > ROC 曲線(V)...

圖表 33-1
「ROC 曲線」對話方塊

- 選取一個或多個檢定機率變數。
- 選取一個狀態變數。
- 識別狀態變數的正值。

ROC 曲線選項

圖表 33-2
「ROC 曲線選項」對話方塊



您可以指定下列的 ROC 分析選項：

分類。 讓您指定在進行正向分類時，是否要加入或排除分割值。這個設定目前對輸出沒有影響。

檢定方向。 讓您指定與正向類別相關之尺度的方向。

區域標準誤差的參數。 讓您指定估計曲線下區域之標準誤的方法。可用的方法有非參數式 (nonparametric) 和雙負指數。同時您還可以設定信賴區間的等級。有效範圍為 50.1% 到 99.9%。

遺漏值。 讓您指定處理遺漏值的方式。

附錄
A

Notices

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 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.shmtl.

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.

索引