IBM SPSS Complex Samples 19

Note: Before using this information and the product it supports, read the general information under Notices第 258 頁.

序

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

部 I:
使用手冊

章節
1

複合樣本程序簡介

在傳統軟體套件中，統計程序固有的假設便是：資料檔中的觀察值代表所需母群的簡單隨機樣本。對於數量不斷增加的公司而言，這項假設是站不住腳的；而對於一些發現該假設符合成本效益且能以更有結構的方式方便取得樣本的研究人員而言，這項假設也無法成立。

「複合樣本」選項可讓您根據複合式設計選取樣本，並將設計的規格併入資料分析中，因而確保您的結果是有效的。

複合樣本的性質

複合樣本與簡單亂數樣本在許多方面上是不同的。在簡單隨機樣本中，是以相等的機率隨機選出各個樣本單位，而不直接以整個母群體置換 (WOR)。相反的，指定的樣本會有部份或所有的下列特色：

分層。 階層化樣本會在母群體非重疊的次群別 (分層) 中，個別選擇樣本。例如，分層可能是社經群組、工作類別、年齡群組或種族。有了分層，您可以確定所需的次群組有足夠的樣本大小，增加整體估計值的精確性，並針對各個分層使用不同的取樣方法。

集群。 集群取樣會選擇取樣單位 (或集群) 的群組。例如，集群可以是學校、醫院或地理區域，取樣單位可以是學生、病人或市民。集群在多階和區域 (地理) 樣本中很常見的。

多階。在多階取樣中，您可以根據集群選擇第一階樣本。接著，您可以從選擇的集群中抽出次樣本，以建立第二階樣本。如果第二階樣本是以次集群為基礎，接著您便可將第三階加入樣本中。例如，在調查的第一階中，可以抽出城市樣本。接著再從選擇的城市中，抽出家庭為樣本。最後，從選擇的家庭中，抽出個人來訪談。「取樣和分析準備」精靈可讓您在設計中指定三階。

非隨機取樣。 當難以隨機選取時，單位能夠以有系統的 (以固定間距) 或是依照順序來取樣。

不相等的選擇機率。 當取樣的集群中包含不相等的單位數時，您可以使用機率 - 比例 - 大小 (PPS) 取樣方式，讓集群的選擇機率與其包含單位的比例相等。PPS 取樣也可使用更一般的加權架構來選擇單位。

未限制取樣。 未限制取樣選擇取樣後放回 (WR) 的單位。因此，個體單位可多次被選為樣本。

取樣加權。 當抽出複合樣本時，會自動計算出取樣加權，並能夠與目標母群體中的每一個取樣單位所代表的「次數分配」理想的對應。因此，樣本加權的總和應可估計母群體大小。複合樣本分析程序需要取樣加權，才能適當的分析複合樣本。請注意，在「複合樣本」選項中，應該完全使用這些加權，且不應該透過「觀察值加權」程序以其他分析程序來使用這些加權，該程序會將加權當做觀察值的複製。

使用複合樣本程序

根據實際的需要使用「複合樣本」程序。主要的使用者類型為：

- 根據複合設計來計劃和執行的調查，可能於稍後分析樣本。調查員的主要工具是取樣精靈。
- 分析先前根據複合設計取得的資料檔。使用「複合樣本」分析程序前，您可能需要使用分析準備精靈。

不管您是那一類的使用者，您都必須提供設計資訊給「複合樣本」程序。此資訊儲存於**計劃檔**中，以方便重複使用。

計劃檔

計劃檔包含複合樣本規格。計劃檔有兩種：

取樣計劃。 「取樣精靈」中指定的規格定義用於抽出複合樣本的樣本設計。取樣計劃檔包含這些規格。取樣計劃檔也包含預設的分析計劃，該計劃使用適合指定樣本設計的估計方法。

分析計劃。 這些計劃檔包含「複合樣本」分析程序所需資訊，以適當的計算出複本樣本的變數估計值。計劃包含樣本結構、每一階的估計方法和取得變數的參照，例如樣本加權。「分析準備」精靈可讓您建立和編輯分析計劃。

將規格儲存於計劃檔有幾項優點：

- 調查員可以指定多階取樣計劃的第一階，並抽出第一階單位，搜集第二階取樣單位的資訊，再修改包含第二階的取樣計劃。
- 無法存取樣計劃檔的分析者可以指定分析計劃，並參照每個「複合樣本」分析程序的計劃。
- 大規模公用樣本的設計工具可以發行取樣計劃檔，再簡化分析者的指示，避免每個分析者都需指定其個人的分析計劃。

詳細資訊

如需取樣技術的詳細資訊，請參閱下列文字：

Cochran, W. G. 1977. Sampling Techniques (取樣技巧), 第三版 ed. 紐約: John Wiley and Sons.

Kish, L. 1965. Survey Sampling (調查取樣). 紐約: John Wiley and Sons.

Kish, L. 1987. Statistical Design for Research (研究的統計設計). 紐約: John Wiley and Sons.

Murthy, M. N. 1967. Sampling Theory and Methods (取樣理論與方法). Calcutta, India (印度加爾各達): Statistical Publishing Society (統計出版社).

Särndal, C., B. Swensson, 和 J. Wretman. 1992. Model Assisted Survey Sampling (模式輔助調查取樣). 紐約: Springer-Verlag.

章節
2

從複合設計中取樣

圖表 2-1
「取樣精靈 - 歡迎」步驟

「取樣精靈」會帶領您進行建立、修改或執行取樣計劃檔案步驟。在您使用「精靈」之前，您腦海中應具備已定義完全的目標母群、取樣單位清單、及合適的樣本設計。

建立新樣本計劃

- 從功能表選擇：
  分析 > 複合樣本 > 選取樣本...
- 選擇「設計樣本」並選擇計劃檔名以儲存樣本計劃。
- 按一下「下一步」繼續使用「精靈」。

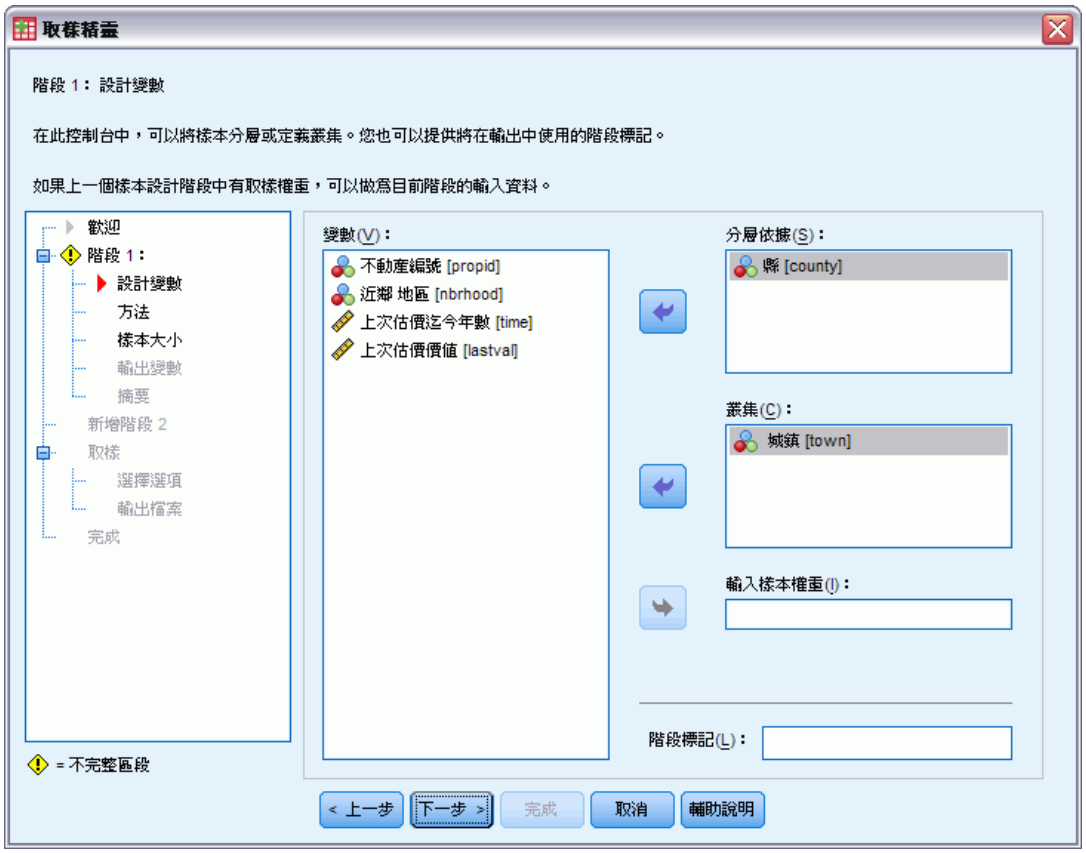
- ▶ 或者在「設計變數」步驟，隨意定義層、集群，並輸入樣本權重。定義完成之後，按一下「**下一步**」。
- ▶ 在「取樣方法」步驟，隨意選擇選取項目的方法。

  若您選擇 **PPS Brewer** 或 **PPS Murthy**，可按一下「**完成**」進行抽樣。否則，請按一下「**下一步**」並接著：
- ▶ 在「樣本大小」步驟，請為樣本指定個數或單位比例。
- ▶ 您現在可按一下「**完成**」抽樣。

  在接下來的步驟中，您可以：
  - 選擇要儲存的輸出變數。
  - 新增設計的第二或第三階段。
  - 設定不同的選擇選項，包括要從哪個階段抽樣、亂數種子，及是否將使用者遺漏值視為設計變數的有效值。
  - 選擇要儲存輸出資料的位置。
  - 將您的選擇貼上為指令語法。

取樣精靈：設計變數

圖表 2-2
取樣精靈，設計變數步驟



此步驟可讓您選擇分層及集群變數，並定義輸入樣本權重。您也可以指定階段的標記。

依...分層。 分層變數的交叉分類會定義不同的子母體或層。而會為每個層取得個別的樣本。若要改善估計值的精確度，興趣特性層中的單位需盡可能同質。

集群。 集群變數定義觀察單位或集群的組別。在直接從高價或不可行的母群中取樣觀察單位時，集群是非常實用的；反之，您可以從母群中取樣集群，並從選定的集群中取樣觀察單位。然而，使用集群會導出取樣單位間的相關性，造成精確度的損失。若要將此效應降到最低，興趣特性集群中的單位應儘可能異質。為了計劃多階段設計，您至少須定義一個集群變數。使用多種不同取樣方法時亦需要集群。 如需詳細資訊，請參閱第 6 頁中的取樣精靈：取樣方法.

輸入樣本權重。 若目前的樣本設計是更大型樣本設計的一部份，您可能也有更大型設計先前階段的樣本權重。您可在目前設計的第一階段指定包含這些權重的數字變數。在目前設計的後續階段中會自動計算樣本權重。

階段標記。 您可指定各階段的選項字串標記。這會用於輸出中，以協助維持階段相關資訊的一致性。

注意：來源變數清單在「精靈」的數個步驟中均具有相同內容。換句話說，特定步驟中刪除的來源清單變數將會從所有步驟中刪除。傳回來源清單的變數會出現在所有步驟的清單中。

供瀏覽取樣精靈的樹狀控制

在「取樣精靈」每個步驟的左側會有所有步驟的概要。您可按一下概要中啟動步驟的名稱以瀏覽「精靈」。只要先前步驟是有效的，步驟即可啟動 — 意即，若已為先前的每個步驟提供各步驟最小所需規格的話。若要了解為何有的步驟無效，請參閱個別步驟的「輔助說明」以取得更多資訊。

取樣精靈：取樣方法

圖表 2-3
取樣精靈，取樣方法步驟

此步驟可讓您指定如何從現用資料集中選擇觀察值。

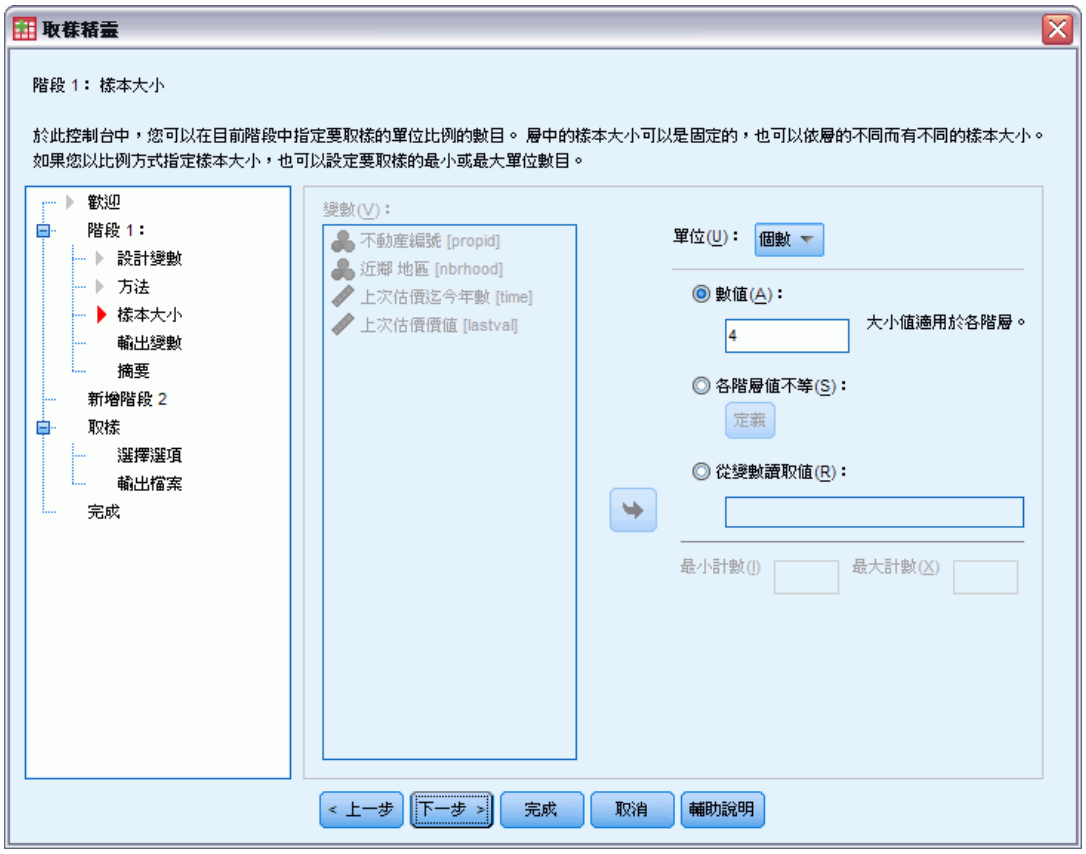
方法。 分組中的控制用於選擇一種選擇方法。有些取樣類型可讓您選擇取樣後放回 (WR) 或取樣後不放回 (WOR)。請參閱類型描述以取得更多資訊。請注意，有些機率與單位大小成比例 (PPS) 類型只有當已定義集群時才可使用，且所有的 PPS 類型只有在設計的第一階段中可用。還有，WR 方法只在設計的最後階段可用。

- **簡單亂數取樣。** 以相等機率選定單位。選擇後可放回或不放回。
- **簡單系統性。** 在整個取樣框（或層，若已指定的話）中以固定的間隔選擇單位，且以選擇後不放回的方式萃取單位。在第一區間中隨機選擇的單位會被選為起始點。
- **簡單循序。** 會以選擇後不放回的方式，以相同的機率選擇單位。
- **PPS。** 這是以機率與單位大小成比例的方式，隨機選擇單位的第一階段方法。所有的單位均可以取樣並放回的方式選擇；唯有集群可以不放回的方式取樣。
- **PPS 系統性。** 這是以機率與單位大小成比例的方式，系統性選擇單位的第一階段方法。可以選擇後不放回的方式選擇。
- **PPS 循序。** 這是以機率與集群大小成比例的方式，以不放回方式循序選擇單位的第一階段方法。
- **PPS Brewer。** 這是以機率與集群大小成比例的方式，以不放回方式從各層選擇兩個集群的第一階段方法。若要使用此方式，需指定集群變數。
- **PPS Murthy。** 這是以機率與集群大小成比例的方式，以不放回方式從各層選擇兩個集群的第一階段方法。若要使用此方式，需指定集群變數。
- **PPS Sampford。** 這是以機率與集群大小成比例的方式，以不放回方式從各層選擇兩個以上集群的第一階段方法。這是 Brewer 方法的延伸。若要使用此方式，需指定集群變數。
- **使用 WR 估計進行分析。** 依照預設，會在與所選取樣方法一致的計劃檔中指定估計方法。即使取樣方法指的是 WOR 估計，但這樣可讓您使用取樣並放回估計。此選項僅在階段 1 可用。

大小測量法 (MOS)。 若已選擇 PPS 方法，您必須指定定義各單位大小的大小測量法。這些大小可明確的在變數中定義，或從資料中計算出。您可隨意設定 MOS 的下界或上界，以置換 MOS 變數中找到的任何數值或資料中計算出的任何數值。這些選項僅在階段 1 可用。

取樣精靈：樣本大小

圖表 2-4
取樣精靈，樣本大小步驟



此步驟可讓您指定單位的個數或比例，以在目前階段中取樣。各層中的樣本大小可以固定或不同。為了指定樣本大小，可用先前階段中選擇的集群定義層。

單位。 您可為樣本指定精確的樣本大小或單位比例。

- **數值。** 所有層會套用單一值。若計量單位選為**個數**，則您應輸入正整數。若選擇**比例**，則您應輸入非負數值。除非取樣後放回，否則比例值亦不應大於 1。
- **層的不相等數值。** 讓您以各層為基礎，透過「定義不相等大小」對話方塊輸入大小值。
- **從變數讀取數值。** 讓您選擇包含層大小值的數值變數。

若選擇**比例**，您就具備設定取樣單位個數上界及下界的選項。

定義不相等大小

圖表 2-5
「定義不相等大小」對話方塊

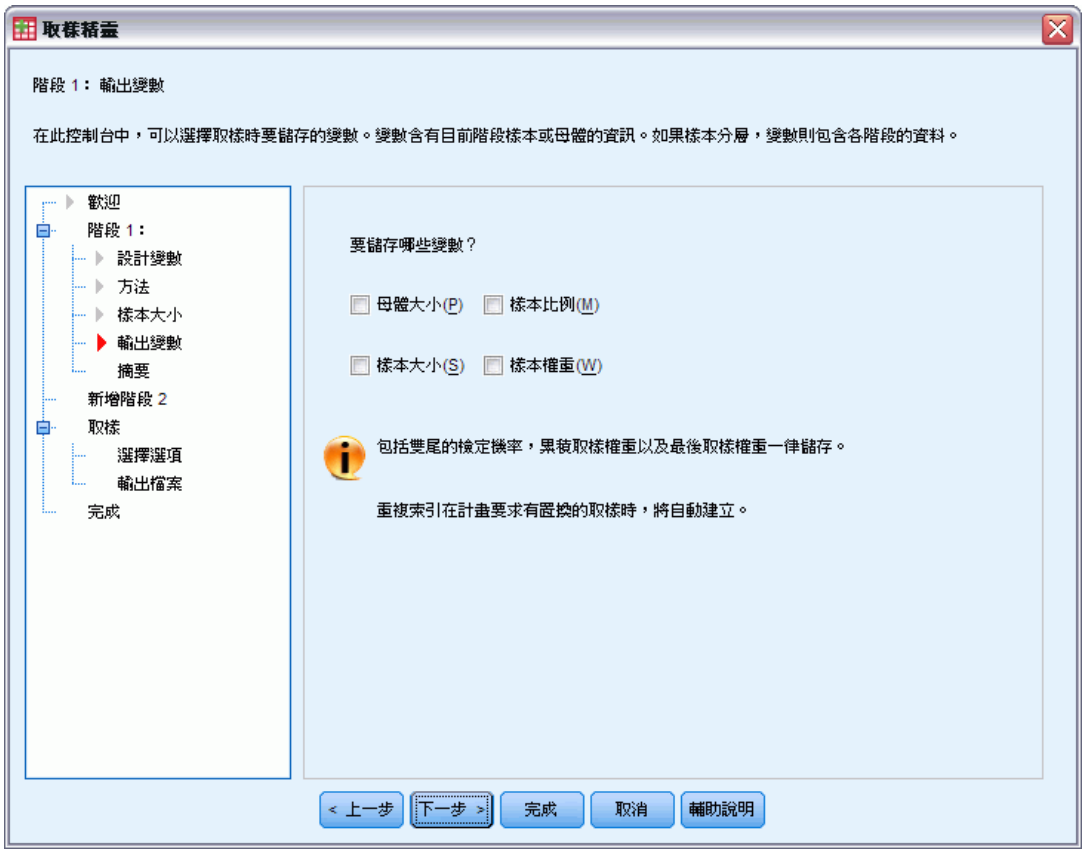
「定義不相等大小」對話方塊可讓您輸入每一分層的大小。

「大小規格」網格。 網格顯示最多五個分層或集群變數的交叉分類—每一列有一個分層/集群組合。合格的網格變數包括所有目前和先前階段中的分層變數，以及先前階段中的集群變數。變數可於網格中重新排序，或移到「排除」清單。在最右邊的行中輸入大小。按一下「**標記**」或「**數值**」，以顯示網格中分層和集群變數的數值標記和資料值。包含未標記數值的儲存格永遠顯示數值。按一下「**重新整理分層**」，以網格中變數的已標記資料值每個組合來重新設定網格。

排除。 若要指定分層/集群組合的子集大小，請將一或多個變數移至「排除」清單。這些變數不用於定義樣本大小。

取樣精靈：輸出變數

圖表 2-6
取樣精靈，輸出變數步驟



此步驟可讓您在取樣後選擇要儲存的變數。

母群大小。 指定階段中母群裡單位的估計數量。所儲存變數的根名稱為 PopulationSize_。

樣本比例。 指定階段中的取樣率。所儲存變數的根名稱為 SamplingRate_。

樣本大小。 指定階段抽出的單位數。所儲存變數的根名稱為 SampleSize_。

樣本權重。 這是包含機率的倒數。所儲存變數的根名稱為 SampleWeight_。

會自動產生某些階段性變數。包括：

包含機率。 指定階段抽出的單位比例。所儲存變數的根名稱為 InclusionProbability_。

累積權重。 包含現階段的先前階段所累積的樣本權重。所儲存變數的根名稱為 SampleWeightCumulative_。

索引。 在指定階段中數次識別所選單位。所儲存變數的根名稱為 Index_。

注意：所儲存變數的根名稱會包含反映階段數的整數字尾 — 例如，PopulationSize_1_ 指的是階段 1 所儲存的母群大小。

取樣精靈：計劃摘要

圖表 2-7
取樣精靈，計劃摘要步驟

這是各階段的最後一個步驟，提供現階段樣本設計規格的摘要。您可從這裡繼續下一階段（若需要的話請建立該階段）或設定抽樣選項。

取樣精靈：抽樣選擇選項

圖表 2-8
「取樣精靈 - 抽樣 - 選擇選項」步驟

此步驟可讓您選擇是否抽樣。您亦可控制其他取樣選項，如亂數種子及遺漏值處理方式。

抽樣。 除了選擇是否抽樣以外，您亦可選擇執行部分的取樣設計。您必須依階段順序取樣 — 也就是說，除非已抽樣階段 1，否則無法抽樣階段 2。編輯或執行計劃時，您無法重新取樣已鎖定的階段。

種子。 這可讓您選擇用來產生亂數的種子數值。

包含使用者自訂的遺漏值。 這決定使用者遺漏值是否有效。若有效，使用者遺漏值會被視為一個不同的類別。

已排序的資料。 若您的樣本框已預先依分層變數的數值排序，則此選項可讓您加速選擇程序。

取樣精靈：抽樣輸出檔案

圖表 2-9
取樣精靈，抽樣輸出檔案步驟

此步驟可讓您選擇要將已取樣的觀察值、加權變數、聯合機率、觀察值選擇規則引導至何處。

範例資料。 這些選項可讓您決定要將樣本輸出寫入至何處。其可新增至現用資料集、寫入至新資料集，或儲存至外部 IBM® SPSS® Statistics 資料檔案。可以在目前階段作業期間使用資料集，但是在後續的階段作業中則無法使用，除非您明確地將其儲存為資料檔案。資料集名稱必須符合變數命名規則。 若指定了外部檔案或新資料集，則會寫入取樣輸出變數和現用資料集中所選觀察值的變數。

聯合機率。 這些選項可讓您決定要將聯合機率寫入至何處。其可儲存為外部 SPSS Statistics 檔案資料。若選擇 PPS WOR、PPS Brewer、PPS Sampford 或 PPS Murthy，且未指定 WR 估計，則會產生聯合機率。

觀察值選擇規則。 若您一次只在一階段中建立樣本，則您可能會希望將觀察值選擇規則儲存為文字檔。這樣對建立後續階段的副框而言非常實用。

取樣精靈：完成

圖表 2-10
取樣精靈 - 完成步驟

這是最後的步驟。您現可儲存計劃檔並抽樣，或將您的選擇貼上至語法視窗。

在現有計劃檔的階段建立變更時，您可將編輯過的計劃儲存為新檔案或覆寫現有檔案。若不對現有階段進行變更而直接新增階段，「精靈」會自動覆寫現有計劃檔。若您希望將計劃儲存為新檔案，請選擇「**將精靈產生的語法貼到語法視窗**」，並在語法指令中變更檔案名稱。

修改現有計劃檔

- 從功能表選擇：
  分析 > 複合樣本 > 選取樣本...
- 選擇「**編輯樣本設計**」並選擇要編輯的計劃檔。
- 按一下「**下一步**」繼續使用「精靈」。

▶ 在「計劃摘要」步驟中檢閱取樣計劃，並按一下「下一步」。

後續步驟大致上與新設計所使用的步驟相似。請參閱「輔助說明」以取得個別步驟的詳細資訊。

▶ 瀏覽至「完成」步驟，並為編輯過的計劃檔指定新名稱，或選擇覆寫現有計劃檔。

您可以：

- 指定已取樣過的階段。
- 將階段從計劃中刪除。

取樣精靈：計劃摘要

圖表 2-11
取樣精靈，計劃摘要步驟

此階段可讓您檢閱取樣計劃，並指出已取樣過的階段。若您正在編輯計劃，亦可從計劃中刪除階段。

先前已取樣過的階段。 若無法使用延伸的取樣框，則您需一次一階段地執行多階段取樣設計。從下拉式清單中選擇哪些階段已取樣過。所有已執行過的階段都會被鎖定；這些階段無法在「抽樣選擇選項」步驟中使用，且在編輯計劃時無法改變。

刪除階段。 您可從多階段設計中刪除階段 2 和 3。

執行現有計劃檔

- 從功能表選擇：
  分析 > 複合樣本 > 選取樣本...
- 選擇「**抽出樣本**」並選擇要執行的計劃檔。
- 按一下「**下一步**」繼續使用「精靈」。
- 在「計劃摘要」步驟中檢閱取樣計劃，並按一下「**下一步**」。
- 執行樣本計劃時會跳過包含階段資訊的個別步驟。您現可隨時前往「完成」步驟。

  或者，您可以指定已取樣過的階段。

CSPLAN 與 CSSELECT 指令的其他功能

指令語法語言讓您也可以：

- 指定輸出變數的自訂名稱。
- 在「瀏覽器」中控制輸出。例如，您可在某樣本已經設計或修改的情況下決定不顯示計劃的階段性摘要、在已執行某樣本設計的情況下決定不顯示依層取樣的觀察值分配摘要，以及要求觀察值程序摘要。
- 在現用資料集中選擇要寫入外部樣本檔或不同資料集的變數子集。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

章節
3

備妥複合樣本以進行分析

圖表 3-1
分析準備精靈，歡迎步驟

「分析準備精靈」會引導您進行一連串步驟，建立或修改與多種「複合樣本」分析程序共同運作的分析計劃。在使用「精靈」之前，您應該先根據複雜設計抽樣。

當您無法存取用以抽樣的取樣計劃檔案時，建立新計劃是最有用的（回想一下，取樣計劃含有預設的分析計劃）。如果您可以存取用以抽樣的取樣計劃檔案時，您可以使用取樣計劃檔案中所含的預設分析計劃，或覆寫預設的分析指定，然後將變更儲存到新檔案中。

建立新分析計劃

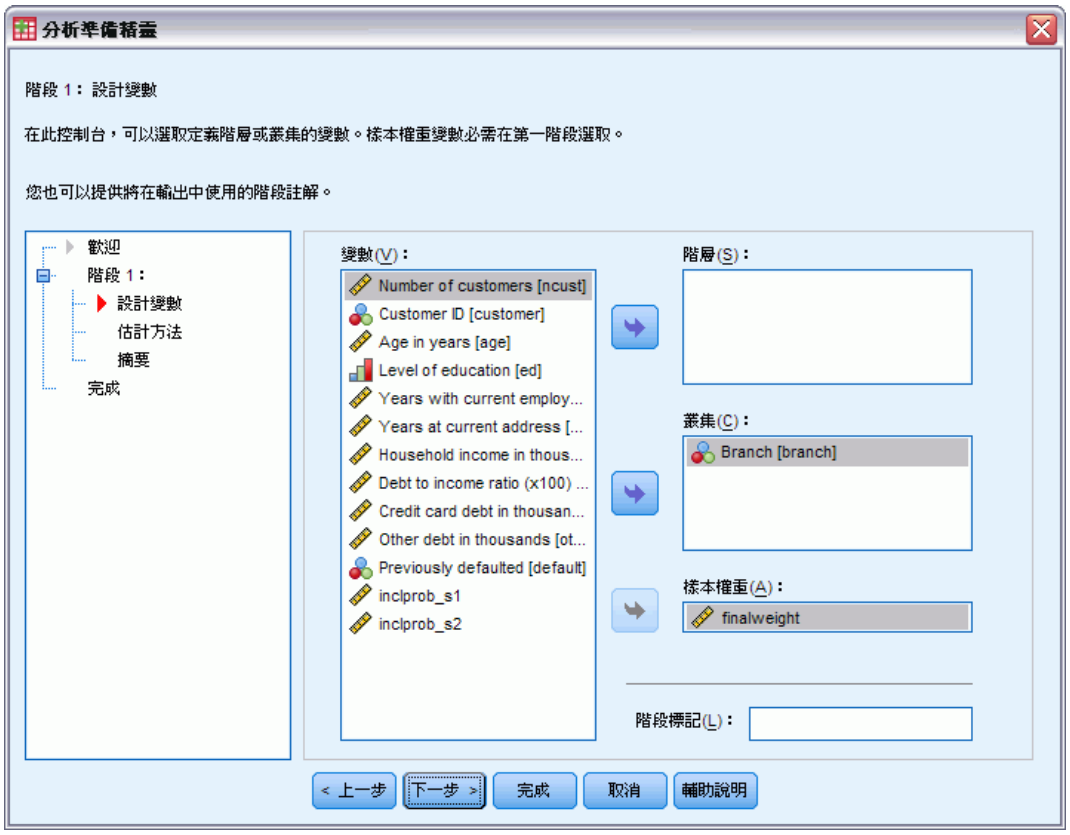
▶ 從功能表選擇：
分析 > 複合樣本 > 準備分析...

- 選擇「**建立計劃檔案**」，並選擇您要儲存分析計劃的計劃檔案名稱。
- 按一下「**下一步**」繼續使用「精靈」。
- 在「設計變數」步驟中指定包含樣本權重的變數，或是選擇定義層或集群。
- 您淚可按一下「**完成**」儲存計劃。

  在接下來的步驟中，您可以：
  - 在「估計方法」步驟中選擇估計標準誤的方法。
  - 在「大小」步驟中，指定取樣的單位數，或每單位包括雙尾的檢定機率。
  - 新增設計的第二或第三階段。
  - 將您的選擇貼上為指令語法。

分析準備精靈：設計變數

圖表 3-2
分析準備精靈，設計變數步驟



此步驟可讓您識別分層和集群變數，並定義樣本權重。您也可以指定階段的標記。

層。 分層變數的交叉分類會定義不同的子母體或層。您的總樣本代表各層獨立樣本的組合。

集群。 集群變數定義觀察單位或集群的組別。在多階段中所抽出的樣本，會選擇以前階段中的集群，然後再從選定集群中二次取樣單位。當分析利用取樣集群並放回的方式取得的資料檔時，您應該加入重複的索引做為集群變數。

樣本權重。 您必須在第一階段中提供樣本權重。在目前設計的後續階段中會自動計算樣本權重。

階段標記。 您可指定各階段的選項字串標記。這會用於輸出中，以協助維持階段相關資訊的一致性。

注意：來源變數清單在「精靈」的數個步驟中均具有相同內容。換句話說，特定步驟中刪除的來源清單變數將會從所有步驟中刪除。傳回來源清單的變數會出淚在所有步驟中。

用於瀏覽分析精靈的樹狀結構控制

「分析精靈」每個步驟的左側會有所有步驟的概要。您可按一下概要中啟動步驟的名稱以瀏覽「精靈」。只要先前步驟是有效的 — 意即，若已為先前的每個步驟提供各步驟最小所需規格的話，即可啟動步驟。若要了解提供的步驟為何為無效資訊時，請參閱個別步驟的「輔助說明」。

分析準備精靈：估計方法

圖表 3-3
分析準備精靈，估計方法步驟

此步驟能讓您指定階段的估計方法。

WR（取樣後放回）。 使用複雜取樣設計估計變異數時，WR 估計不包含有限母群 (FPC) 的取樣修正。使用簡單隨機取樣 (SRS) 估計變異數時，您可以選擇包含或排除 FPC。

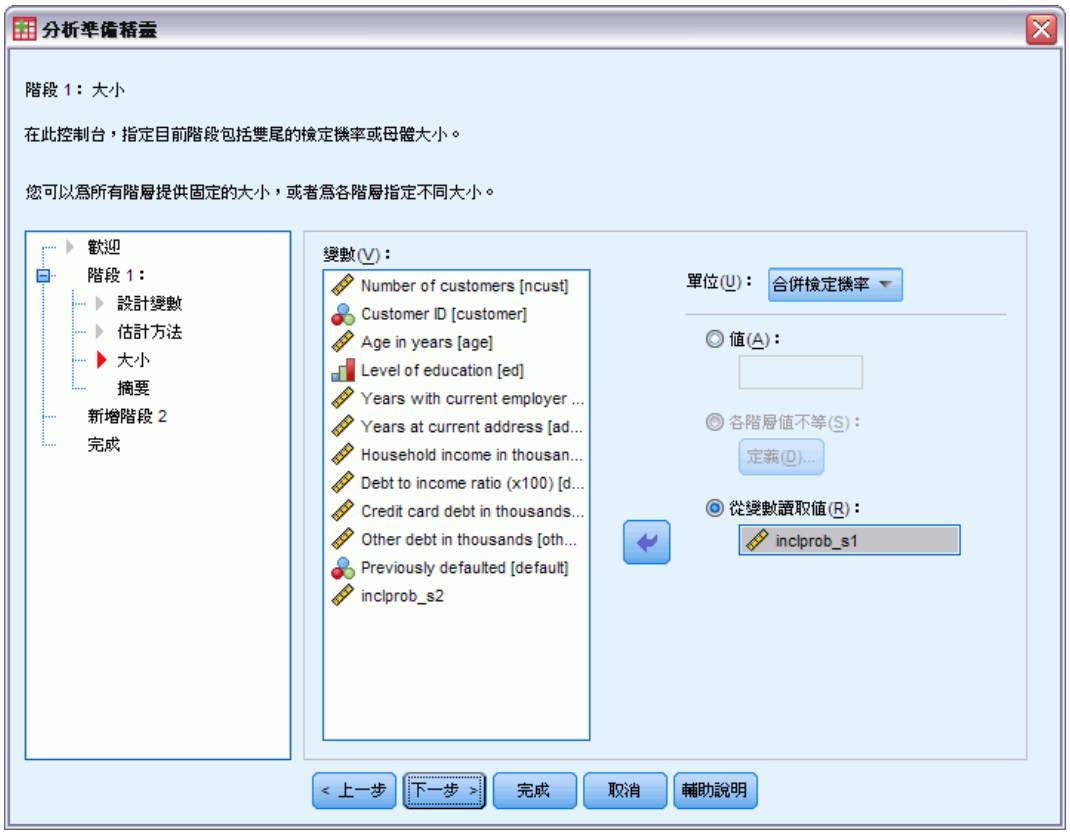
當已尺度化分析加權時，建議選擇不包含 SRS 的 FRC 變異數估計，讓加權總和不會到達母群大小。SRS 變異數估計用於計算如設計效應等統計量。只有設計的最後一個階段才能指定 WR 估計；如果您選擇 WR 估計，「精靈」將不會讓您再新增其他階段。

相等 WOR（相等機率，取樣後不放回）。 相等 WOR 估計包括有限母群修正，並假設取樣單位的機率相等。您可以在設計的任何階段指定相等 WOR。

不相等 WOR（不相等機率，取樣後不放回）。 除了使用有限母群修正外，「不相等 WOR」使用不相等機率來選擇取樣單位（通常為集群）。只有在第一階段中才能使用此估計方法。

分析準備精靈：大小

圖表 3-4
分析準備精靈，大小步驟



此步驟是用來指定目前階段的包括包含機率或母群大小。大小可以是固定的，或隨著層而改變。為了指定大小，可以使用在之前階段中所指定的集群來定義層。請注意，只有在選擇 Equal WOR 做為「估計方法」時，才需要此步驟。

單位。 您可以指定已取樣單位的精確母群大小或機率。

- **數值。** 所有層會套用單一值。若計量單位選為「母群體大小」，則您應該輸入非負數的整數。如果選擇「包含機率」，則您應該輸入介於（含）0 與 1 的值。
- **層的不相等數值。** 讓您以各層為基礎，透過「定義不相等大小」對話方塊輸入大小值。
- **從變數讀取數值。** 讓您選擇包含層大小值的數值變數。

定義不相等大小

圖表 3-5
「定義不相等大小」對話方塊

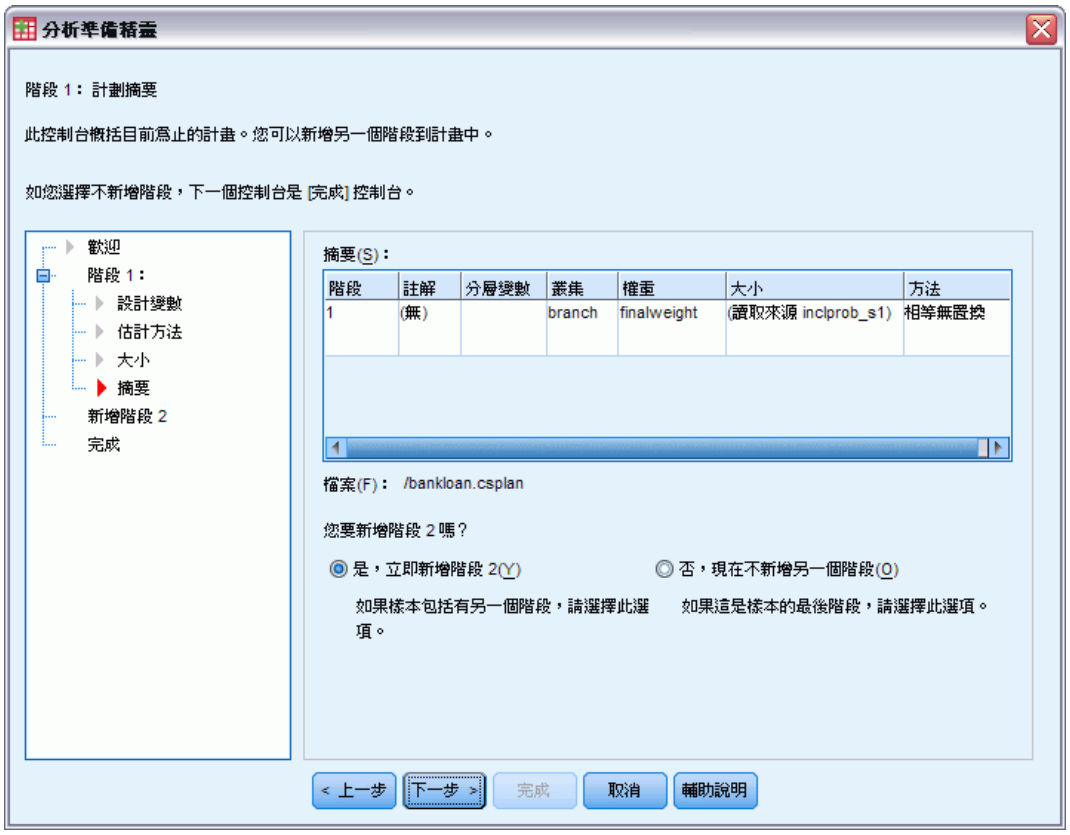
「定義不相等大小」對話方塊可讓您輸入每一分層的大小。

「大小規格」網格。 網格顯示最多五個分層或集群變數的交叉分類—每一列有一個分層/集群組合。合格的網格變數包括所有目前和先前階段中的分層變數，以及先前階段中的集群變數。變數可於網格中重新排序，或移到「排除」清單。在最右邊的行中輸入大小。按一下「標記」或「數值」，以顯示網格中分層和集群變數的數值標記和資料值。包含未標記數值的儲存格永遠顯示數值。按一下「重新整理分層」，以網格中變數的已標記資料值每個組合來重新設定網格。

排除。 若要指定分層/集群組合的子集大小，請將一或多個變數移至「排除」清單。這些變數不用於定義樣本大小。

分析準備精靈：計劃摘要

圖表 3-6
分析準備精靈，計劃摘要步驟



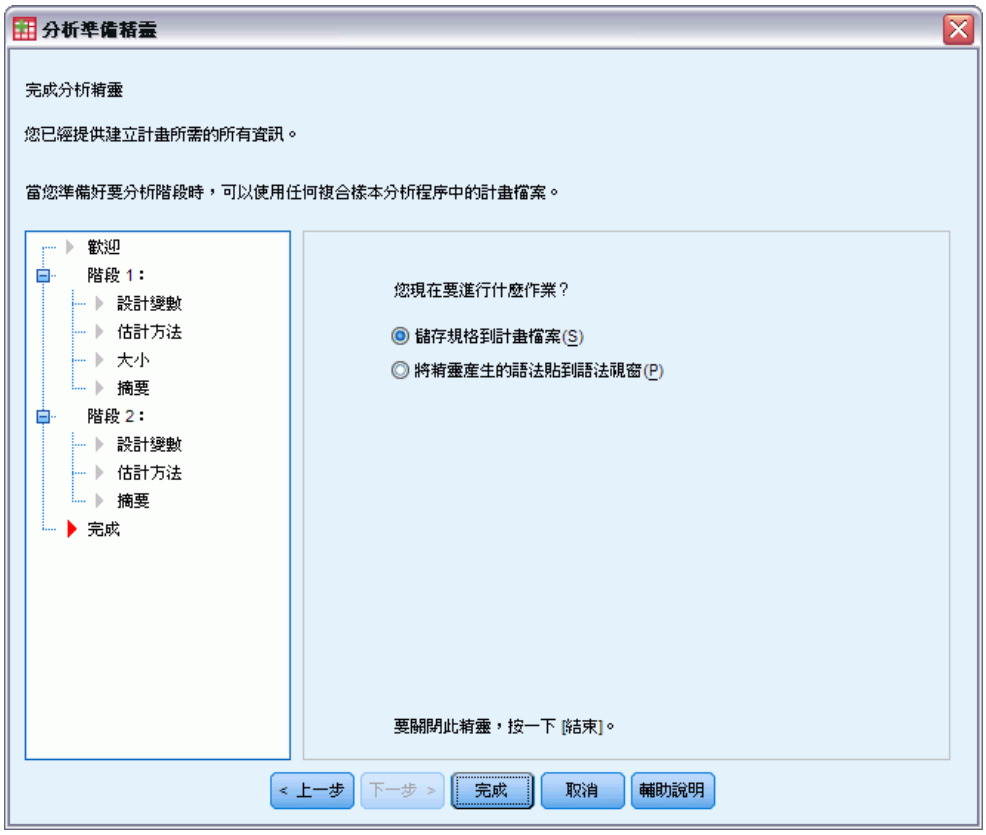
這是各階段的最後一個步驟，提供現階段分析設計規格的摘要。您可以從這裡繼續下一階段（若需要的話請建立該階段）或儲存分析指定。

如果您無法新增其他階段，可能是因為：

- 在「設計變數」步驟中沒有指定集群變數。
- 在「估計方法」步驟中選擇 WR 估計。
- 這是分析的第三個階段，而「精靈」最多支援三個階段。

分析準備精靈：完成

圖表 3-7
分析準備精靈，完成步驟



這是最後的步驟。您現在可以儲存計劃檔案，或將您的選擇貼到語法視窗中。

在現有計劃檔的階段建立變更時，您可將編輯過的計劃儲存為新檔案或覆寫現有檔案。若不對現有階段進行變更而直接新增階段，「精靈」會自動覆寫現有計劃檔。若您希望將計劃儲存為新檔案，請選擇「**將精靈產生的語法貼到語法視窗**」，並在語法指令中變更檔案名稱。

修改現有分析計劃

- 從功能表選擇：
  分析 > **複合樣本** > **準備分析...**
- 選擇「**編輯計劃檔案**」，並選擇您要儲存分析計劃的計劃檔案名稱。
- 按一下「**下一步**」繼續使用「精靈」。

- 在「計劃摘要」步驟中檢閱分析計劃，並按一下「**下一步**」。

  後續步驟大致上與新設計所使用的步驟相似。請參閱「輔助說明」以取得個別步驟的詳細資訊。

- 瀏覽至「完成」步驟，並為編輯過的計劃檔指定新名稱，或選擇覆寫現有計劃檔。

  或者，您可以將階段從計劃中移除。

分析準備精靈：計劃摘要

圖表 3-8
分析準備精靈，計劃摘要步驟

在此步驟中，您可以檢視分析計劃並刪除計劃中的階段。

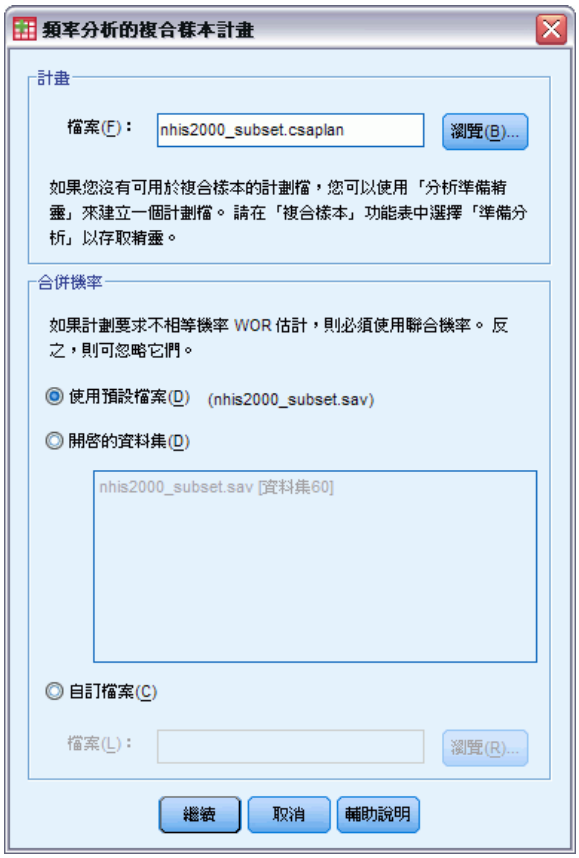
刪除階段。 您可從多階段設計中刪除階段 2 和 3。因為計劃最少必須要有一個階段，所以您可以編輯設計中的階段 1，但不可以刪除。

章節
4

複合樣本計劃

「複合樣本」分析程序需要分析或樣本計劃檔中的分析指定，才能提供有效的結果。

圖表 4-1
複合樣本計劃對話方塊



計劃。 指定分析或樣本計劃檔案的路徑。

聯合機率。 若要使用 PPS WOR 方法來繪製 Unequal WOR 估計的集群，您必須另外指定包含聯合機率的檔案或開放資料集。此檔案或資料集是「取樣精靈」在取樣時建立。

章節
5

複合樣本次數分配表

「複合樣本次數分配表」程序可產生所選變數的次數分配表，並顯示單變量統計。您可隨意依一或多類別變數所定義的次組別要求統計量。

範例。 使用「複合樣本次數分配表」程序，您便可以根據「國民健康訪問調查 (NHIS)」的結果，並對此公共使用資料使用適當的分析計劃，取得美國國民使用維生素情形的單變量表格統計量。

統計量。 該程序會產生儲存格母群體大小的估計值和表格百分比，加上標準差、信賴區間、變異係數、設計效應、設計效應平方根、累積值和每個估計值的未加權個數。此外，亦計算卡方與概似比統計量做為相等儲存格比例的檢定。

資料。 可產生次數分配表的變數應為類別變數。子母體變數可為類別字串或數值。

假設。資料檔中的觀察值代表複合設計中，應根據「複合樣本計劃」對話方塊所選檔案裡規格進行分析的樣本。

取得「複合樣本次數分配表」

- 從功能表選擇：
  分析 > 複合樣本 > 次數分配表...
- 選取計劃檔。或者，選取自訂聯合機率檔案。
- 按一下「繼續」。

圖表 5-1
「次數分配表」對話方塊

- 至少要選取一個次數分配變數。

或者，您可指定變數以定義子母體。分別計算每一個子母體的統計量。

複合樣本次數分配表統計

圖表 5-2
次數分配表統計量對話方塊

儲存格。 該群組允許您要求儲存格母群體大小的估計值，與表格百分比。

統計量。 此群組產生與母群體大小或表格百分比有關的統計量。

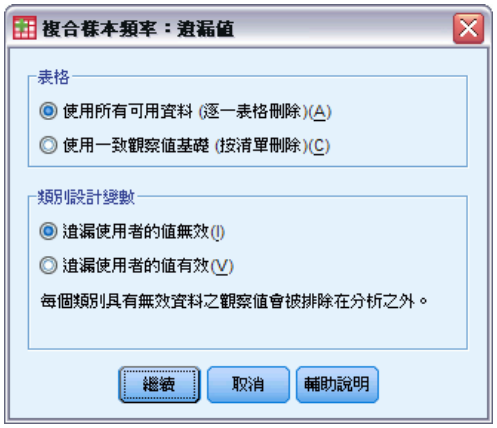
- **標準誤。** 估計值的標準誤。
- **信賴區間。** 使用指定水準時，估計值的信賴區間。

- **變異係數。** 估計值的標準誤與估計值的比例量數。
- **未加權個數。** 用於計算估計值的單位數。
- **設計效應。** 由簡單隨機樣本所取得變異數的估計值變異量比。這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。
- **設計效應的平方根。** 這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。
- **累積值。** 該變數每個值的累積估計值。

相等儲存格比例檢定。 對於變數類別有相同次數分配的虛無假設而言，這會產生卡方和概似值比檢定。對每個變數執行不同的檢定。

複合樣本遺漏值

圖表 5-3
「遺漏值」對話方塊



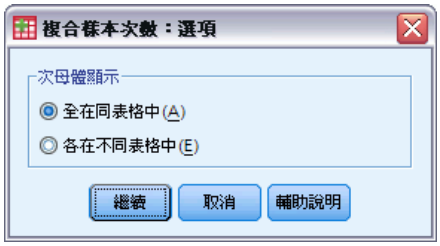
表格。 此組別決定用以分析的觀察值。

- **使用全部的可用資料。** 遺漏值是根據每個表格來決定。因此，次數分配表或交叉表中用來計算統計量的觀察值可能會有所不同。
- **使用一致的觀察值庫。** 遺漏值是根據所有變數判斷的。因此，用來計算統計量的觀察值，就會在各個表格內保持一致。

類別設計變數。 此組別決定使用者遺漏值有效或無效。

複合樣本選項

圖表 5-4
「選項」對話方塊



子母體顯示。 您可以選擇在同一個表格或不同的表格中顯示子母體。

章節
6

複合樣本敘述統計量

「複合樣本敘述統計量」程序顯示數個變數的單變量摘要統計量。您可隨意依一或多類別變數所定義的次組別要求統計量。

範例。使用「複合樣本敘述統計」程序，您便可以根據國民健康訪問調查 (NHIS) 的結果，以及對該公開使用的資料進行適當分析，來得到美國國民活動量的單變量敘述統計。

統計量。 此程序會產生平均數和總和，以及 t-檢定、標準誤、信賴區間、變異係數、未加權的個數、人口數量、設計效應和每個估計值設計效應的平方根。

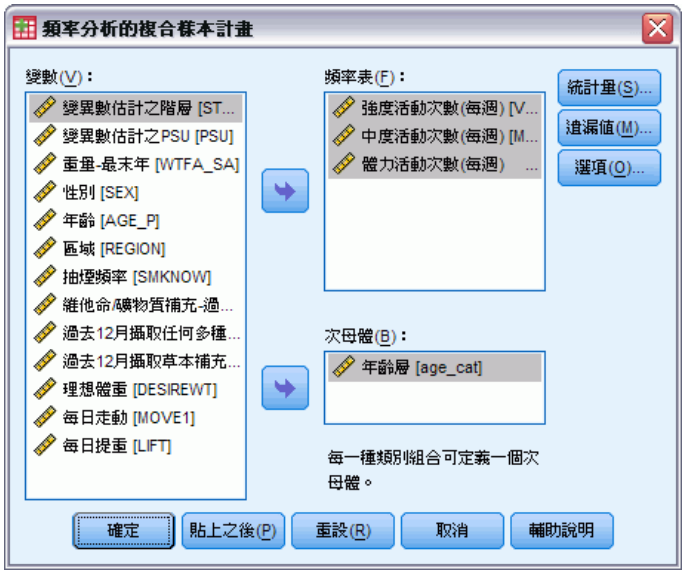
資料。 量數應該是尺度變數。子母體變數可為類別字串或數值。

假設。資料檔中的觀察值代表複合設計中，應根據「複合樣本計劃」對話方塊所選檔案裡規格進行分析的樣本。

取得複合樣本敘述統計量

- 從功能表選擇：
  分析 > 複合樣本 > 描述性統計量...
- 選取計劃檔。或者，選取自訂聯合機率檔案。
- 按一下「繼續」。

圖表 6-1
描述性統計量對話方塊



▶ 至少要選取一個測量變數。

或者，您可指定變數以定義子母體。分別計算每一個子母體的統計量。

複合樣本敘述統計量

圖表 6-2
「敘述統計量」對話方塊

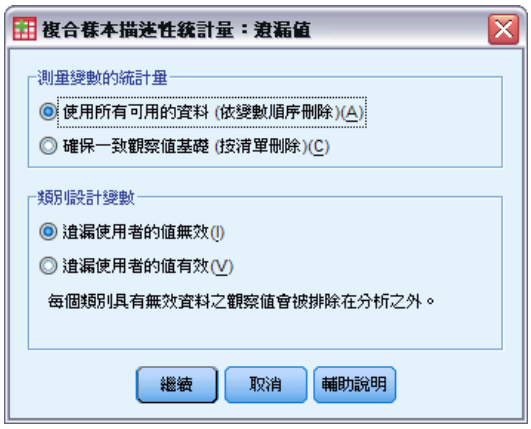
摘要。 此群組可讓您要求平方值的估計值和測量變數總和。此外，您也可以要求指定值的估計值 t 檢定。

統計量。 此群組會產生與平均值或總和相關的統計量。

- **標準誤。** 估計值的標準誤。
- **信賴區間。** 使用指定水準時，估計值的信賴區間。
- **變異係數。** 估計值的標準誤與估計值的比例量數。
- **未加權個數。** 用於計算估計值的單位數。
- **母群大小。** 母群中單位的估計數量。
- **設計效應。** 由簡單隨機樣本所取得變異數的估計值變異量比。這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。
- **設計效應的平方根。** 這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。

複合樣本敘述統計量遺漏值

圖表 6-3
「敘述統計量遺漏值」對話方塊



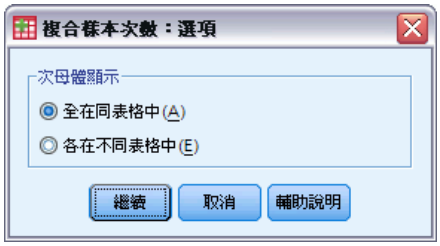
測量變數的統計量。 此組別決定用以分析的觀察值。

- **使用全部的可用資料。** 遺漏值是根據變數逐一決定的，因此用於計算統計量的觀察值會隨測量變數而改變。
- **確保觀察值基準一致。** 遺漏值是依所有變數決定的，因此用於計算統計量的觀察值會一致。

類別設計變數。 此組別決定使用者遺漏值有效或無效。

複合樣本選項

圖表 6-4
「選項」對話方塊



子母體顯示。 您可以選擇在同一個表格或不同的表格中顯示子母體。

章節
7

複合樣本交叉表

「複合樣本交叉表」程序能產生成對選定變數的交叉表列表，並顯示二因子統計量。您可隨意依一或多類別變數所定義的次組別要求統計量。

範例。 利用「複合樣本交叉表」程序，您便可以根據「國民健康訪問調查 (NHIS)」的結果，並對此公共使用資料使用適當的分析計劃，取得美國國民使用維生素與吸煙頻率的交叉分類統計量。

統計量。 此程序可產生儲存格母群大小和列、值行的估計值以及表格百分比，加上標準誤、信賴區間、變異係數、期望值、設計效應、設計效應的平方根、殘差、調整後殘差，和各估計值的未加權個數。使用 2x2 表格來計算 odds 比率、相對風險和風險差異。此外，並計算列和行變數之獨立性檢定的 Pearson 和概似比統計量。

資料。 列和行變數應該是類別變數。子母體變數可為類別字串或數值。

假設。資料檔中的觀察值代表複合設計中，應根據「複合樣本計劃」對話方塊所選檔案裡規格進行分析的樣本。

取得複合樣本交叉表

- 從功能表選擇：
  分析 > 複合樣本 > 交叉表...
- 選取計劃檔。或者，選取自訂聯合機率檔案。
- 按一下「繼續」。

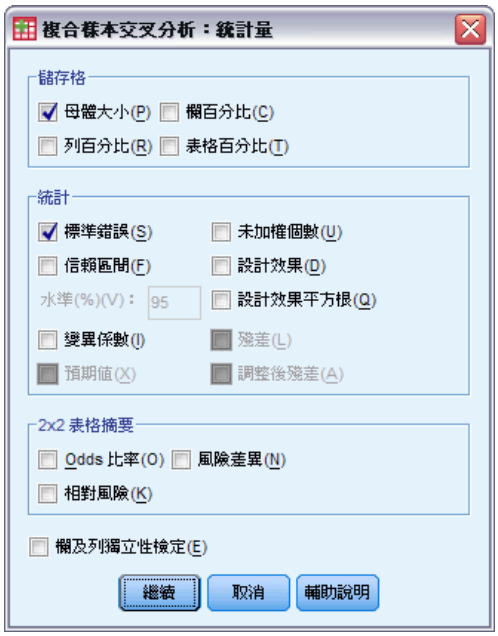
圖表 7-1
交叉表對話方塊

▶ 選擇一個以上的列變數和一個行變數。

或者，您可指定變數以定義子母體。分別計算每一個子母體的統計量。

複合樣本交叉表統計量

圖表 7-2
交叉表統計量對話方塊



儲存格。 此組別能讓您要求儲存格母群大小的估計值，以及列、行與表格的百分比。

統計量。 此組別能產生與母群體大小和列、行及表格百分比相關的統計量。

- **標準誤。** 估計值的標準誤。
- **信賴區間。** 使用指定水準時，估計值的信賴區間。
- **變異係數。** 估計值的標準誤與估計值的比例量數。
- **期望值。** 假設列與行變數不相關的情況下，估計值的期望值。
- **未加權個數。** 用於計算估計值的單位數。
- **設計效應。** 由簡單隨機樣本所取得變異數的估計值變異量比。這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。
- **設計效應的平方根。** 這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。
- **殘差。** 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。
- **調整後殘差。** 儲存格（觀察值減去期望值）除以其標準誤的估計值。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。

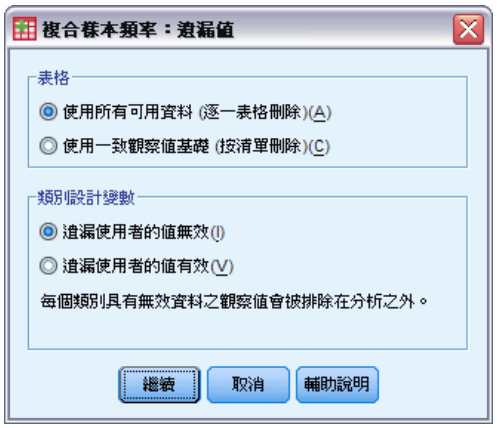
2x2 表格的摘要。 此組別可產生表格的統計量，而此表格中的列與行變數各有兩種類別。各為所顯示之因子與發生事件之間關聯性強度的測量。

- **Odds 比率。** 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。
- **相對風險。** 存在某因子時事件的風險，與該因子不存在時事件風險的比率。
- **風險差異。** 存在某因子時事件的風險，與該因子不存在時事件風險的差異。

列與行的獨立性檢定。 假設列與行變數不相關時，這可產生卡方和概似比檢定。執行各對變數的不同檢定。

複合樣本遺漏值

圖表 7-3
「遺漏值」對話方塊



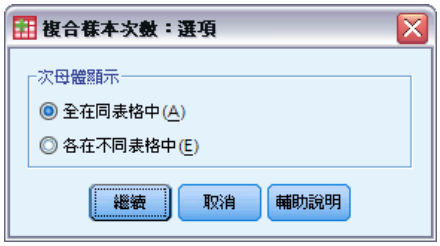
表格。 此組別決定用以分析的觀察值。

- **使用全部的可用資料。** 遺漏值是根據每個表格來決定。因此，次數分配表或交叉表中用來計算統計量的觀察值可能會有所不同。
- **使用一致的觀察值庫。** 遺漏值是根據所有變數判斷的。因此，用來計算統計量的觀察值，就會在各個表格內保持一致。

類別設計變數。 此組別決定使用者遺漏值有效或無效。

複合樣本選項

圖表 7-4
「選項」對話方塊



子母體顯示。 您可以選擇在同一個表格或不同的表格中顯示子母體。

章節
8

複合樣本比例量數

「複合樣本比例量數」程序會顯示變數比例量數的單變量摘要統計量。您可隨意依一或多類別變數所定義的次組別要求統計量。

範例。 使用「複合樣本比例量數」程序，您即可以根據複合設計並具有資料的適當分析計劃所完成的全州調查為基礎，取得目前屬性值對最後評估值的比例量數之敘述統計。

統計量。 此程序會產生比例量數估計值、t 檢定、標準誤、信賴區間、變異係數、未加權個數、母群大小、設計效應及設計效應的平方根。

資料。 分子和分母均應為正值的尺度變數。子母體變數可為類別字串或數值。

假設。資料檔中的觀察值代表複合設計中，應根據「複合樣本計劃」對話方塊所選檔案裡規格進行分析的樣本。

取得複合樣本比例量數

- 從功能表選擇：
  分析 > 複合樣本 > 比例量數...
- 選取計劃檔。或者，選取自訂聯合機率檔案。
- 按一下「繼續」。

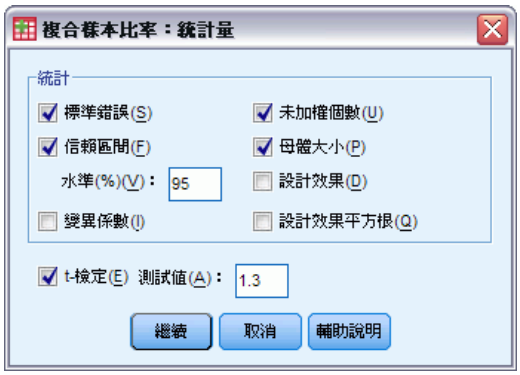
圖表 8-1
比例量數對話方塊

▶ 選擇至少一分子變數及分母變數。

或者，您可指定變數以定義統計量產生的次組別。

複合樣本比例量數統計量

圖表 8-2
比例量數統計量對話方塊



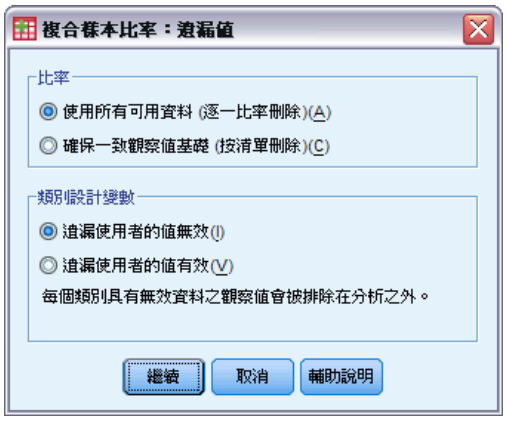
統計量。 此組別會產生與估計值比例量數相關的統計量。

- **標準誤。** 估計值的標準誤。
- **信賴區間。** 使用指定水準時，估計值的信賴區間。
- **變異係數。** 估計值的標準誤與估計值的比例量數。
- **未加權個數。** 用於計算估計值的單位數。
- **母群大小。** 母群中單位的估計數量。
- **設計效應。** 由簡單隨機樣本所取得變異數的估計值變異量比。這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。
- **設計效應的平方根。** 這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。

T 檢定。 您可要求估計值對特定數值的 t 檢定。

複合樣本比例量數遺漏值

圖表 8-3
比例量數遺漏值對話方塊



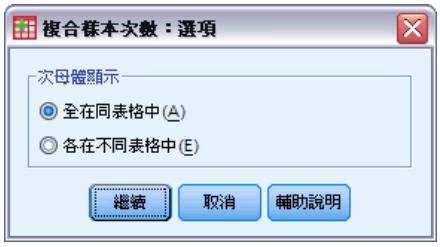
比例量數。 此組別決定用以分析的觀察值。

- **使用全部的可用資料。** 遺漏值是根據逐比例量數基礎判斷的。因此，分子-分母配對中用來計算統計量的觀察值可能會有所不同。
- **確保觀察值基準一致。** 遺漏值是根據所有變數判斷的。因此，用來計算統計量的觀察值會一致。

類別設計變數。 此組別決定使用者遺漏值有效或無效。

複合樣本選項

圖表 8-4
「選項」對話方塊



子母體顯示。 您可以選擇在同一個表格或不同的表格中顯示子母體。

章節
9

複合樣本一般線性模式

「複合樣本一般線性模式 (CSGLM)」程序會為複合取樣方法取得的樣本執行線性迴歸分析，以及變異數與共變異數分析。或者，您也可以要求對子母體進行分析。

範例。 某連鎖雜貨店根據複合設計，調查一組客戶的購買習慣。已知調查結果以及客戶在上個月的花費，該店希望對照客戶性別並結合取樣設計，了解客戶購物的頻率是否與整月支出的量有關。

統計量。 此程序會產生估計量、標準誤、信賴區間、t 檢定、設計效應、模型參數的設計效應平方根，以及參數估計量之間的相關與共變異數。也可取得依變數和自變數的模型適性量數以及描述統計量。此外，您亦可要求模型因子水平與因子間交互作用的邊際平均數估計值。

資料。 依變數是數值變數。因子是類別的。而共變量是與依變數相關的數值變數。子母體變數可為類別字串或數值。

假設。資料檔中的觀察值代表複合設計中，應根據「複合樣本計劃」對話方塊所選檔案裡規格進行分析的樣本。

取得複合樣本一般線性模式

從功能表選擇：

分析(A) > 複合樣本 > 一般線性模式...

- 選取計畫檔案。您可以選擇性選取自訂的聯合機率檔案。
- 按一下「**繼續**」。

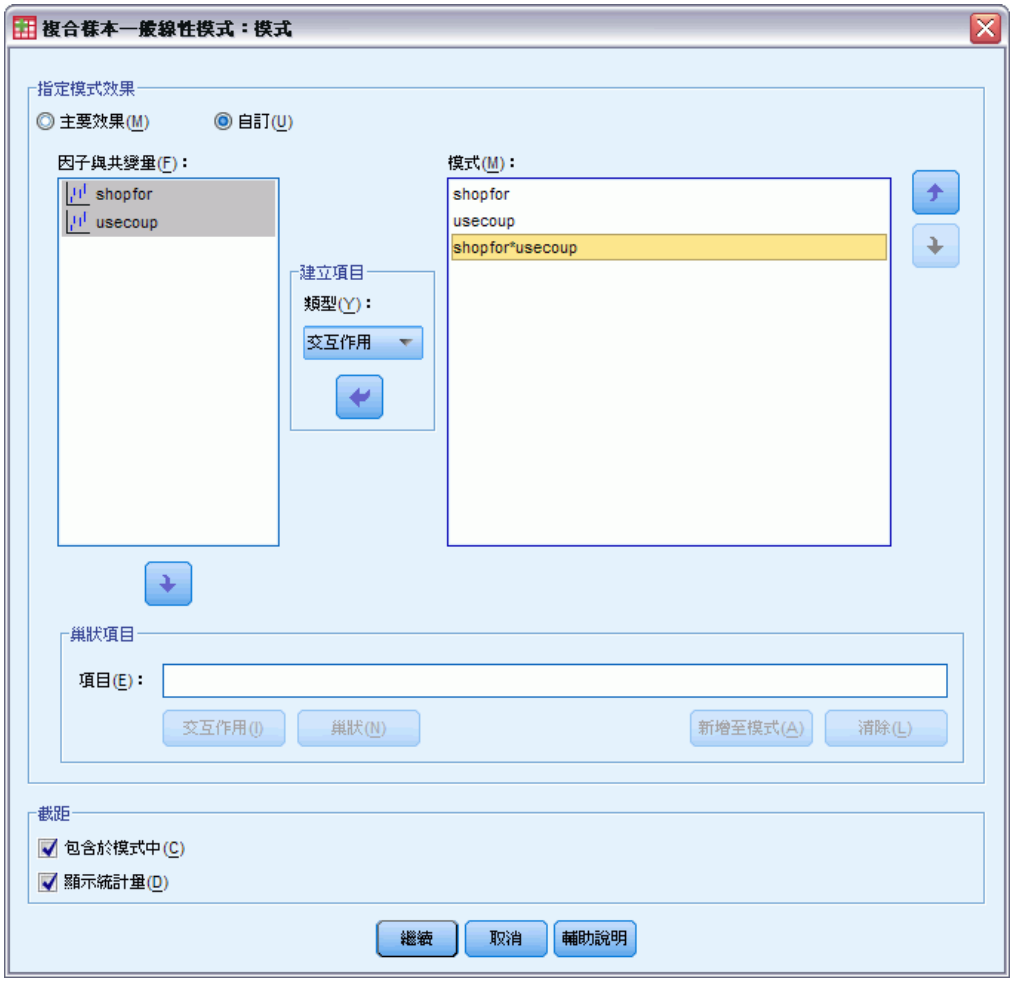
圖表 9-1
一般線性模式對話方塊

▶ 選取依變數。

您可以：

- 根據資料選取適當的因子或共變量變數。
- 指定變數以定義子母體。只會對子母體變數的所選類別執行分析。

圖表 9-2
「模式」對話方塊



指定模式效應。 依照預設值，程序會使用主對話方塊中指定的因子和共變量，建立主效果模式。否則您可建立包含交互作用效應與巢狀項次的自訂模式。

非巢狀項次

對所選擇的因子和共變量而言：

交互作用。 建立所有選取變數的最高階交互作用項。

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。

巢狀項次

您可以在這個程序中，為您的模式建立巢狀的項次。通常巢狀項次在建立因子或共變量效應項的模式時非常有用，但因子或共變量的值不可以與其他因子水準交互作用。例如，連鎖雜貨店可能會追蹤他們客戶在數個商店位置的消費習慣。因為每個客戶通常只在其中一個地點消費，因此您可以說客戶效果項是**巢狀**於商店位置效果項內。

此外，您可以包含交互作用效能，例如多項式項目與同一個共變量有關，或將多層的巢狀新增至巢狀項次。

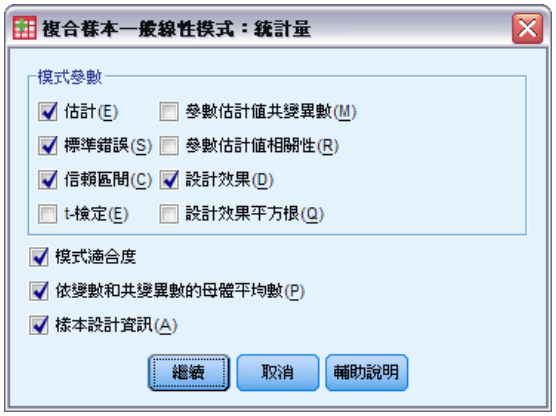
限制。 巢狀項次有下列限制：

- 交互作用內的所有因子都必須是唯一的。因此，如果 A 是因子，那麼指定 A*A 是無效的。
- 巢狀效應項中的所有因子都必須是唯一的。因此，如果 A 是因子，那麼指定 A(A) 是無效的。
- 共變量內不可巢狀效果項。因此，如果 A 是因子，而 X 是共變量，那麼指定 A(X) 是無效的。

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。即使您將截距納入模式，還是可以選擇隱藏其相關的統計量。

複合樣本一般線性模式統計量

圖表 9-3
一般線性模式統計量對話方塊



模式參數。 此群組可讓您控制與模式參數相關的統計量顯示。

- **估計值。** 顯示係數估計值。
- **標準誤。** 顯示各係數估計值的標準誤。
- **信賴區間。** 顯示各係數估計值的信賴區間。區間的信賴水準設定於「選項」對話方塊。
- **T 檢定。** 顯示各係數估計值的 t 檢定。各檢定的虛無假設其係數值為 0。
- **參數估計值的共變異數。** 顯示模式係數的共變異數矩陣估計值。
- **參數估計值的相關性。** 顯示模式係數的相關矩陣估計值。

- **設計效應。** 由簡單隨機樣本所取得變異數的估計值變異量比。這是所指定複合設計的效應量數值，其中遠離 1 的數值表示較大的效應。
- **設計效應的平方根。** 這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。

模式適合度。 顯示 R^2 與均方誤差根統計值。

依變數與共變異數的母群平均數。 顯示依變數、共變異數和因子的摘要資訊。

樣本設計資訊。 顯示樣本相關摘要資訊，包括未加權個數和母群大小。

複合樣本假設檢定

圖表 9-4
「假設檢定」對話方塊

檢定統計量。 這個組別可以讓您選擇用來檢定假設的統計量類型。您可以在 F、調整後 F、卡方和調整後卡方之間選擇。

取樣自由度。 這個組別讓您可以控制用於為所有檢定統計量計算 p 值的取樣設計自由度。如果根據取樣設計時，數值為主要取樣單位的數目與第一階段取樣中的層數目之間的差異。此外，您可以透過指定正整數來設定自訂自由度。

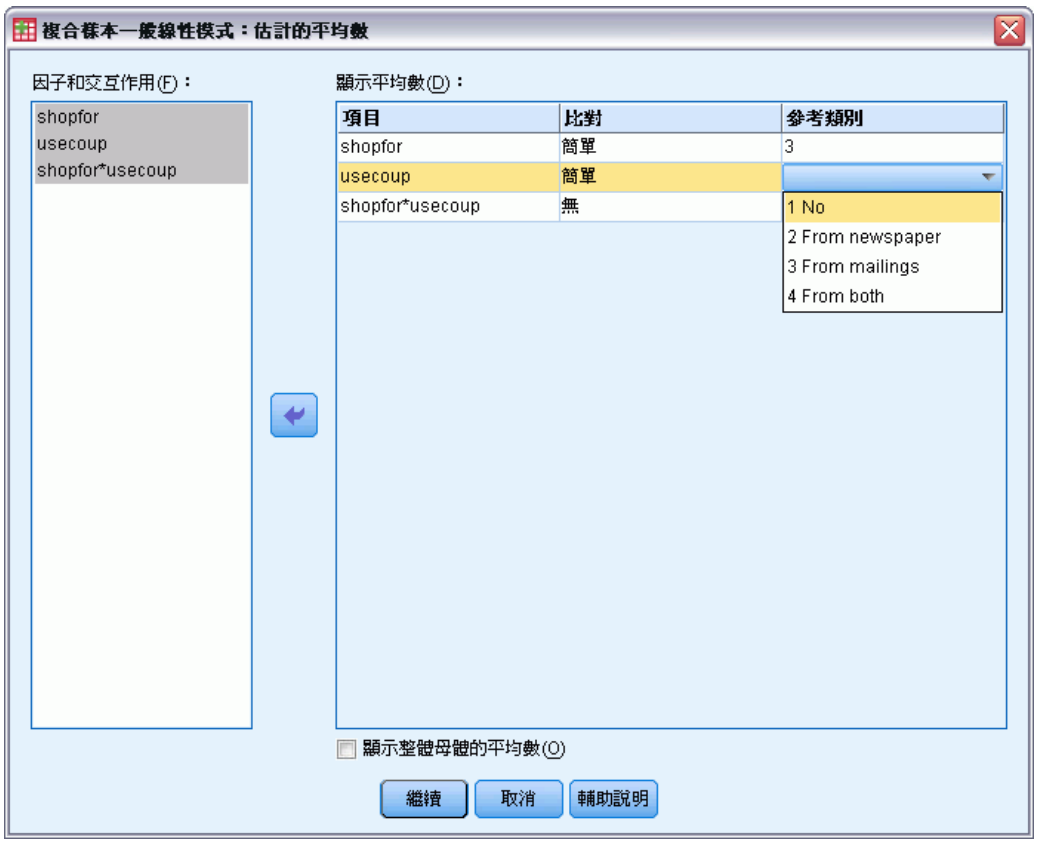
多重比較的調整。 在使用多重對比執行假設檢定時，可從顯著水準針對包含的對比調整整體顯著水準。此組別可讓您選擇調整方法。

- **最小顯著差異。** 這個方法無法控制以下假設之整體可能性，此假設為某些線性的對比不同於虛無假設值，。
- **循序 Sidak。** 這是循序逐步拒絕的 Sidak 程序；就拒絕個別假設而言，此程序做法相當不保守，但整體顯著水準仍維持相同。

- **循序 Bonferroni。** 這是循序逐步拒絕的 Bonferroni 程序；就拒絕個別假設而言，此程序做法相當不保守，但整體顯著水準仍維持相同。
- **Sidak。** 此方法的界限比 Bonferroni 法的界線更緊密
- **Bonferroni。** 此方法會調整觀察到的顯著水準，並實際檢定多重對比。

複合樣本一般線性模式估計平均數

圖表 9-5
一般線性模式估計平均數對話方塊



「估計平均數」對話方塊可讓您顯示「模式」子對話方塊中指定的因子間交互作用程度之模式-邊際平均數估計值。您亦可要求顯示總母群平均數。

項目。 估計平均數是由所選因子間交互作用所計算出來的。

對比 對比會決定如何設定假設檢定，以比較估計平均數。

- **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。
- **離差。** 比較每個水準的平均數（除了參考類別）與所有水準的平均數（總平均）。因子水準以任何一種方式排列都可以。
- **差異。** 比較每個水準的平均數（除了第一個）與前一個水準的平均數。這種對比有時叫作反「Helmert 對比」。

- **Helmert。** 比較每個因子水準的平均數（除了最後一個）與後續水準的平均數。
- **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。

參考類別。 簡單與離差對比需要參考類別，或用來比較的因子水準。

複合樣本一般線性模式儲存

圖表 9-6
一般線性模式儲存對話方塊

儲存變數。 此群組可讓您將模式預測值與殘差儲存為工作檔案中的新變數。

將模式匯出為 SPSS Statistics 資料。以 IBM® SPSS® Statistics 格式寫入資料集，其中包含參數相關性、或共變異數矩陣，以及參數估計值、標準誤、顯著值和自由度。矩陣檔案的變數順序如下。

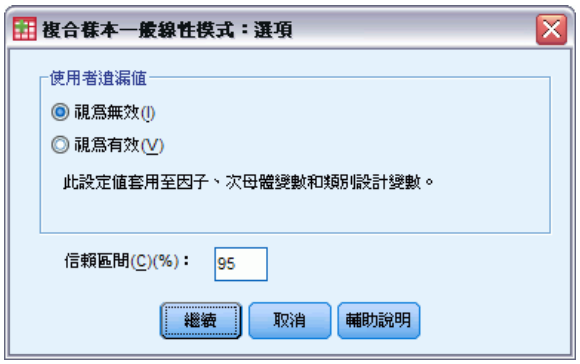
- **rowtype_。**取得值（和數值標記）、COV（共變量）、CORR（相關性）、EST（參數估計值）、SE（標準誤）、SIG（顯著水準）和 DF（取樣設計自由度）。每個模式參數都有列類型為 COV（或 CORR）的個別觀察值，加上每個其他列類型的個別觀察值。
- **varname_。**取得值 P1、P2、...，與所有模式參數的排序清單相對應，對於列類型 COV 或 CORR，其數值標記與參數估計表中顯示的參數字串相對應。其他列類型的儲存格為空白。
- **P1、P2、...** 這些變數與所有模式參數的次序清單相對應，其中的變數標記對應到在參數估計表中顯示的參數字串，並根據列類型取得值。對於冗餘參數，所有共變量會設為零；相關性會設為系統遺漏值；所有參數估計值會設為零；所有標準誤、顯著水準和殘差自由度會設為系統遺漏值。

注意：此檔案無法立即用於在其他讀取矩陣檔案的程序中做進一步的分析，除非這些程序接受所有在此匯出的列類型。

將「模式」匯出為 XML。 將參數估計值與參數共變異數（若選取的話）以 XML (PMML) 格式儲存為矩陣。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。

複合樣本一般線性模式選項

圖表 9-7
一般線性模式選項對話方塊



使用者遺漏值。 所有設計變數與依變數和任何共變異數都必須具備有效資料。這些變數具備無效資料的觀察值都將從分析中刪除。這些控制項可讓您決定是否要將使用者遺漏值視為層、集群、子母體和因子變數間的有效值。

信賴區間。 這是係數估計值與邊際平均數估計值的信賴區間水準。請指定大於或等於 50 且小於 100 的一個數值。

CSGLM 指令的其他功能

指令語法語言也可以讓您：

- 指定效應自訂檢定，或是效應（或值）的線性組合（使用 CUSTOM 次指令）。
- 計算邊際平均數估計值時，將共變異數固定在其平均數以外的數值（使用 EMMEANS 次指令）。
- 指定多項式對比的矩陣（使用 EMMEANS 次指令）。
- 指定檢查單一性時的允差值（使用 CRITERIA 次指令）。
- 建立使用者指定的已存變數名稱（使用 SAVE 次指令）。
- 產生一般估計量函數表（使用 PRINT 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

複合樣本 Logistic 迴歸

「複合樣本 Logistic 迴歸」程序會在所抽取樣本的二分或多項式依變數上執行 Logistic 迴歸分析，此抽取的樣本是透過複雜取樣方法抽取的。或者，您也可以要求對子母體進行分析。

範例。 某放款員根據複合設計，收集了多家不同分行過去的貸款客戶資料。該放款員希望結合樣本設計，看出顧客預設機率是否與年齡、工作資歷、信貸額度相關。

統計量。 此程序會產生估計量、指數化估計量、標準誤、信賴區間、t 檢定、設計效應、模型參數的設計效應平方根，以及參數估計量之間的相關與共變異數。亦可取得依變數和自變數的假 R^2 統計量、分類表和描述統計量。

資料。 依變數是類別的。因子是類別的。而共變量是與依變數相關的數值變數。子母體變數可為類別字串或數值。

假設。資料檔中的觀察值代表複合設計中，應根據「複合樣本計劃」對話方塊所選檔案裡規格進行分析的樣本。

取得複合樣本 Logistic 迴歸

從功能表選擇：

分析(A) > 複合樣本 > Logistic 迴歸...

- ▶ 選取計畫檔案。您可以選擇性選取自訂的聯合機率檔案。
- ▶ 按一下「繼續」。

圖表 10-1
Logistic 迴歸對話方塊

- 選取依變數。

  您可以：

  - 根據資料選取適當的因子或共變量變數。
  - 指定變數以定義子母體。只會對子母體變數的所選類別執行分析。

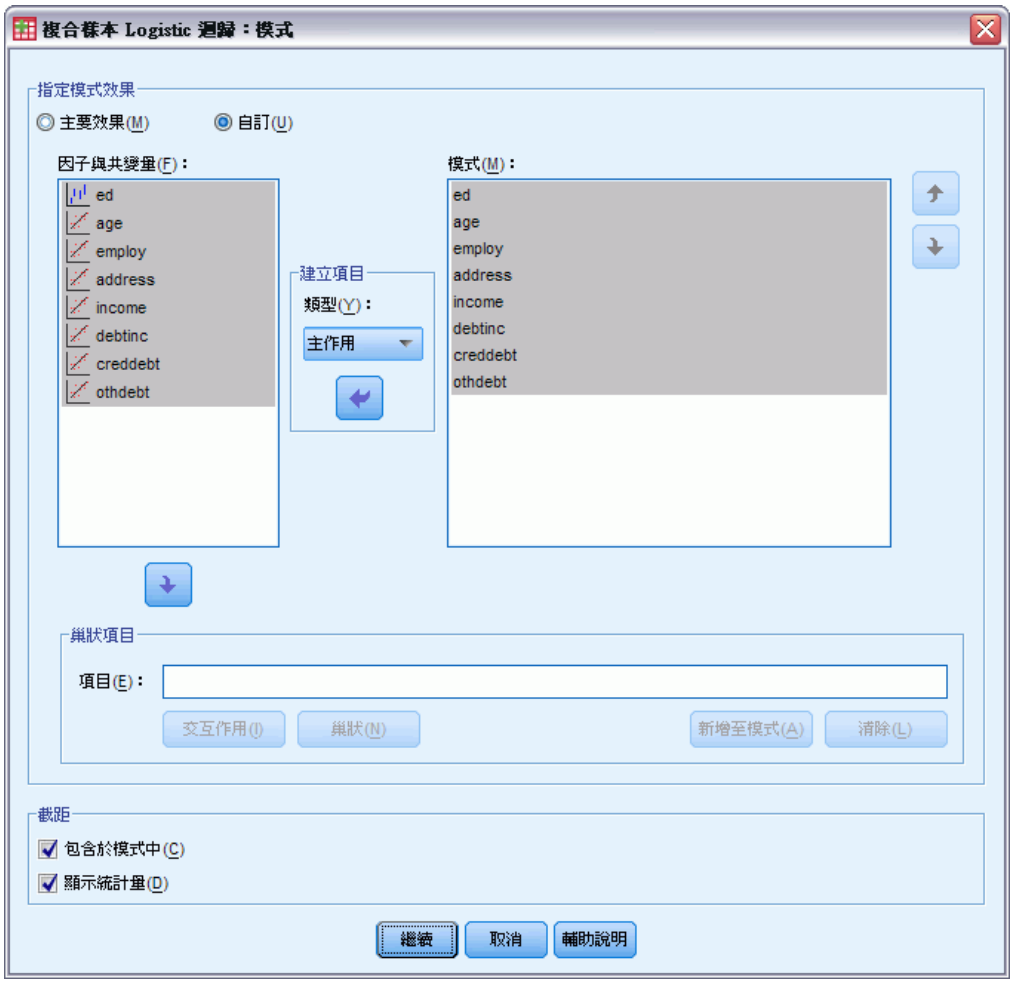
複合樣本 Logistic 迴歸參考類別

圖表 10-2
Logistic 迴歸參考類別對話方塊

根據預設，「複合樣本 Logistic 迴歸」程序會將數值最高的類別當作參考類別。此對話方塊可讓您指定最高值、最低值或自訂類別，作為參考類別。

複合樣本 Logistic 迴歸模式

圖表 10-3
Logistic 迴歸模式對話方塊



指定模式效應。 依照預設值，程序會使用主對話方塊中指定的因子和共變量，建立主效果模式。否則您可建立包含交互作用效應與巢狀項次的自訂模式。

非巢狀項次

對所選擇的因子和共變量而言：

交互作用。 建立所有選取變數的最高階交互作用項。

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。

巢狀項次

您可以在這個程序中，為您的模式建立巢狀的項次。通常巢狀項次在建立因子或共變量效應項的模式時非常有用，但因子或共變量的值不可以與其他因子水準交互作用。例如，連鎖雜貨店可能會追蹤他們客戶在數個商店位置的消費習慣。因為每個客戶通常只在其中一個地點消費，因此您可以說客戶效果項是**巢狀**於商店位置效果項內。

此外，您可以包含交互作用效能，例如多項式項目與同一個共變量有關，或將多層的巢狀新增至巢狀項次。

限制。 巢狀項次有下列限制：

- 交互作用內的所有因子都必須是唯一的。因此，如果 A 是因子，那麼指定 A*A 是無效的。
- 巢狀效應項中的所有因子都必須是唯一的。因此，如果 A 是因子，那麼指定 A(A) 是無效的。
- 共變量內不可巢狀效果項。因此，如果 A 是因子，而 X 是共變量，那麼指定 A(X) 是無效的。

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。即使您將截距納入模式，還是可以選擇隱藏其相關的統計量。

複合樣本 Logistic 迴歸統計量

圖表 10-4
Logistic 迴歸統計量對話方塊

模式適合度。 控制測量總體模式效能的統計量顯示。

- **虛擬迴歸係數 (Pseudo R-square)。** 線性迴歸中的 R^2 統計量在 logistic 迴歸模式中不具有確切的對等值。反而有多量數嘗試模仿 R^2 統計量的特性。
- **分類表。** 根據依變數上的預測模式類別，顯示表格形式的觀察類型交叉分類。

參數。 此群組可讓您控制與模式參數相關的統計量顯示。

- **估計值。** 顯示係數估計值。
- **指數化估計量。** 顯示自然對數的係數估計值次方的基準。其中估計量具有統計檢定、指數化估計量或 exp(B) 的良好特性，較易解譯。
- **標準誤。** 顯示各係數估計值的標準誤。
- **信賴區間。** 顯示各係數估計值的信賴區間。區間的信賴水準設定於「選項」對話方塊。
- **T 檢定。** 顯示各係數估計值的 t 檢定。各檢定的虛無假設其係數值為 0。
- **參數估計值的共變異數。** 顯示模式係數的共變異數矩陣估計值。
- **參數估計值的相關性。** 顯示模式係數的相關矩陣估計值。
- **設計效應。** 由簡單隨機樣本所取得變異數的估計值變異量比。這是所指定複合設計的效應量數值，其中遠離 1 的數值表示較大的效應。
- **設計效應的平方根。** 這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。

模式變數的摘要統計量。 顯示依變數、共變異數和因子的摘要資訊。

樣本設計資訊。 顯示樣本相關摘要資訊，包括未加權個數和母群大小。

複合樣本假設檢定

圖表 10-5
「假設檢定」對話方塊

檢定統計量。 這個組別可以讓您選擇用來檢定假設的統計量類型。您可以在 F、調整後 F、卡方和調整後卡方之間選擇。

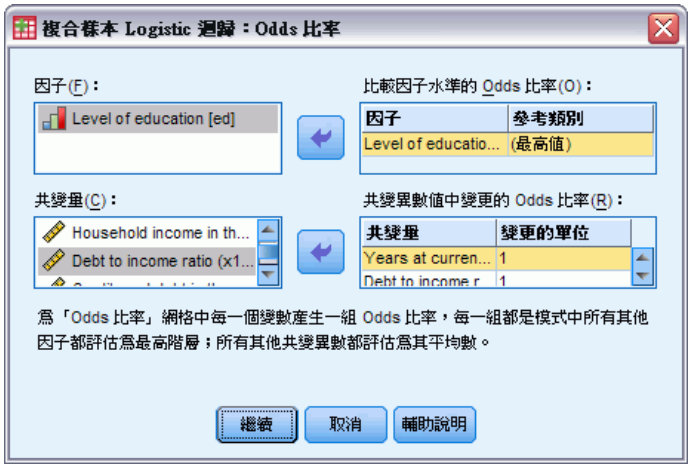
取樣自由度。 這個組別讓您可以控制用於為所有檢定統計量計算 p 值的取樣設計自由度。如果根據取樣設計時，數值為主要取樣單位的數目與第一階段取樣中的層數目之間的差異。此外，您可以透過指定正整數來設定自訂自由度。

多重比較的調整。 在使用多重對比執行假設檢定時，可從顯著水準針對包含的對比調整整體顯著水準。此組別可讓您選擇調整方法。

- **最小顯著差異。** 這個方法無法控制以下假設之整體可能性，此假設為某些線性的對比不同於虛無假設值，。
- **循序 Sidak。** 這是循序逐步拒絕的 Sidak 程序；就拒絕個別假設而言，此程序做法相當不保守，但整體顯著水準仍維持相同。
- **循序 Bonferroni。** 這是循序逐步拒絕的 Bonferroni 程序；就拒絕個別假設而言，此程序做法相當不保守，但整體顯著水準仍維持相同。
- **Sidak。** 此方法的界限比 Bonferroni 法的界線更緊密
- **Bonferroni。** 此方法會調整觀察到的顯著水準，並實際檢定多重對比。

複合樣本 Logistic 迴歸 Odds 比率

圖表 10-6
Logistic 迴歸 Odds 比率對話方塊



「Odds 比率」對話方塊可讓您顯示特定因子與共變異數的模式估計 odds 比率。除了參考類別以外，各依變數類別會各自計算不同組的 odds 比率。

因子。 對於所選的每個因子，顯示在各因子類別中 odds 與特定參考類別中 odds 的比率。

共變量。 對於所選的每個共變量，顯示在平均共變量值加上特定變更單位中 odds 與平均數中 odds 的比率。

計算因子或共變量的 odds 比率時，程序會將其他所有因子固定在其最高值，以及將其他所有共變量固定在其平均值。若因子或共變量與模式中其他預測值交互作用，則 odds 比率不僅根據指定變數的變化而異，也會因交互作用的變數值而異。若指定的共變量在模式中自行交互作用（例如，age*age），則 odds 比率會根據共變量與共變量值的變化而異。

複合樣本 Logistic 迴歸儲存

圖表 10-7
Logistic 迴歸儲存對話方塊

儲存變數。 此群組可讓您將模式預測值類別與預測機率儲存為作用中資料集中的新變數。

將模式匯出為 SPSS Statistics 資料。以 IBM® SPSS® Statistics 格式寫入資料集，其中包含參數相關性、或共變異數矩陣，以及參數估計值、標準誤、顯著值和自由度。矩陣檔案的變數順序如下。

- **rowtype_。**取得值（和數值標記）、COV（共變量）、CORR（相關性）、EST（參數估計值）、SE（標準誤）、SIG（顯著水準）和 DF（取樣設計自由度）。每個模式參數都有列類型為 COV（或 CORR）的個別觀察值，加上每個其他列類型的個別觀察值。
- **varname_。**取得值 P1、P2、...，與所有模式參數的排序清單相對應，對於列類型 COV 或 CORR，其數值標記與參數估計表中顯示的參數字串相對應。其他列類型的儲存格為空白。
- **P1、P2、...** 這些變數與所有模式參數的次序清單相對應，其中的變數標記對應到在參數估計表中顯示的參數字串，並根據列類型取得值。對於冗餘參數，所有共變量會設為零；相關性會設為系統遺漏值；所有參數估計值會設為零；所有標準誤、顯著水準和殘差自由度會設為系統遺漏值。

注意：此檔案無法立即用於在其他讀取矩陣檔案的程序中做進一步的分析，除非這些程序接受所有在此匯出的列類型。

將「模式」匯出為 XML。 將參數估計值與參數共變異數（若選取的話）以 XML (PMML) 格式儲存為矩陣。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。

複合樣本 Logistic 迴歸選項

圖表 10-8
Logistic 迴歸選項對話方塊

估計。 此群組可讓您控制模式估計中使用的變數準則。

- **最大疊代。** 將執行運算的疊代最大值。指定一個非負的整數。
- **Step-Halving 最大值。** 每次疊代時，步驟大小會因乘以因子 0.5 而減少，直到對數概似增加或到達最大半階次數。指定一個正整數。
- **根據參數估計值的變更限制疊代。** 選取的話，演算法會在參數估計值之絕對或相對變更小於所指定數值的疊代之後停止，該數值不得為負數。
- **根據對數概似的變更限制疊代。** 選取的話，演算法會在對數概似函數之絕對或相對變更小於所指定數值的疊代之後停止，該數值不得為負數。
- **檢查資料點的完整分組。** 選取的話，演算法會執行檢定，以確定參數估計值具有唯一值。當程序能夠產生可正確分組各觀察值的模式，就會開始分組。
- **顯示疊代歷程。** 從第 0 個疊代（初始估計值）開始，每 n 個疊代顯示參數估計值與統計量。如果您選擇列印疊代歷程，則無論 n 的值為何，都一定會列印最後一次疊代。

使用者遺漏值。 所有設計變數與依變數和任何共變異數都必須具備有效資料。這些變數具備無效資料的觀察值都將從分析中刪除。這些控制項可讓您決定是否要將使用者遺漏值視為層、集群、子母體和因子變數間的有效值。

信賴區間。 這是係數估計值、指數化係數估計值與 odds 比率的信賴區間水準。請指定大於或等於 50 且小於 100 的一個數值。

CSLOGISTIC 指令的其他功能

指令語法語言也可以讓您：

- 指定效應自訂檢定，或是效應（或值）的線性組合（使用 CUSTOM 次指令）。
- 計算因子和共變量的 odds 比率時，固定其他模式變數的值（使用 ODDSRATIOS 次指令）。
- 指定檢查單一性時的允差值（使用 CRITERIA 次指令）。
- 建立使用者指定的已存變數名稱（使用 SAVE 次指令）。
- 產生一般估計量函數表（使用 PRINT 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

複合樣本次序迴歸

「複合樣本次序迴歸」程序會在所抽取樣本的二分或次序依變數上執行迴歸分析，此抽取的樣本是透過複合取樣方法抽取的。或者，您也可以要求對子母體進行分析。

範例。 議會代表們在考慮將一項法案交付立法之前，會想知道是否有民眾支持該法案，及該項支持與投票人口如何相關。民調機構根據複雜取樣設計，設計出一份問卷並開始進行訪查。您可以使用「複合樣本次序迴歸」，根據投票人口為法案的支援等級套入模式。

資料。 依變數是次序的。因子是類別的。而共變量是與依變數相關的數值變數。子母體變數可為類別字串或數值。

假設。資料檔中的觀察值代表複合設計中，應根據「複合樣本計劃」對話方塊所選檔案裡規格進行分析的樣本。

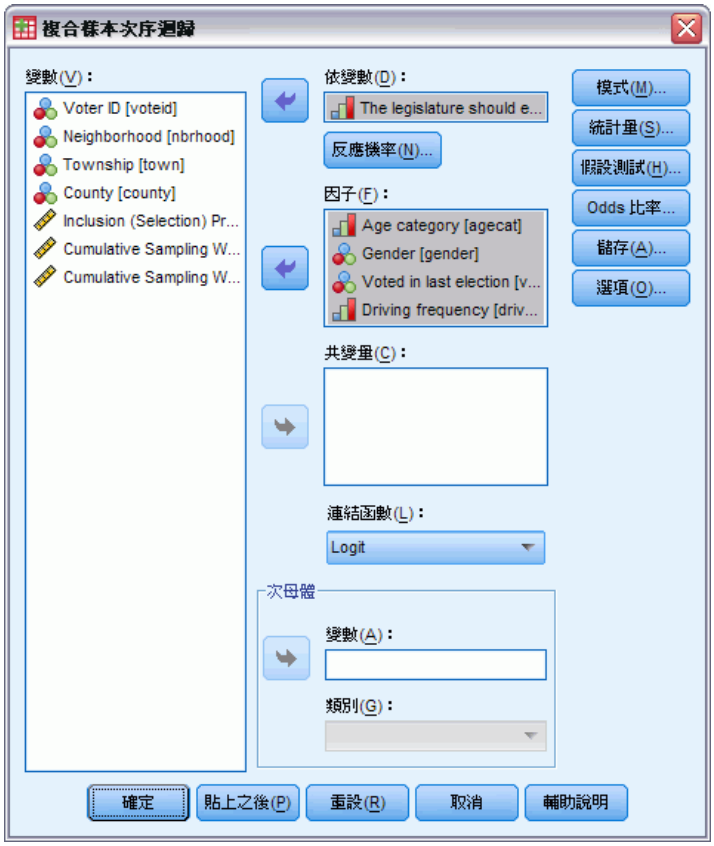
取得複合樣本次序迴歸

從功能表選擇：

分析(A) > 複合樣本 > 次序迴歸...

- 選取計畫檔案。您可以選擇性選取自訂的聯合機率檔案。
- 按一下「繼續」。

圖表 11-1
「次序的迴歸」對話方塊



▶ 選取依變數。

您可以：

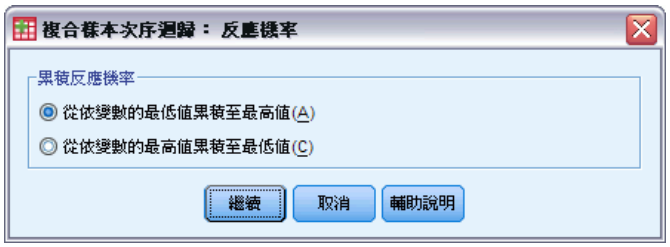
- 根據資料選取適當的因子或共變量變數。
- 指定變數以定義子母體。只會對子母體變數的所選類別執行分析，雖然仍會根據整個資料集對變異數進行適當預估。
- 選取連結函數。

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

複合樣本次序迴歸：反應機率

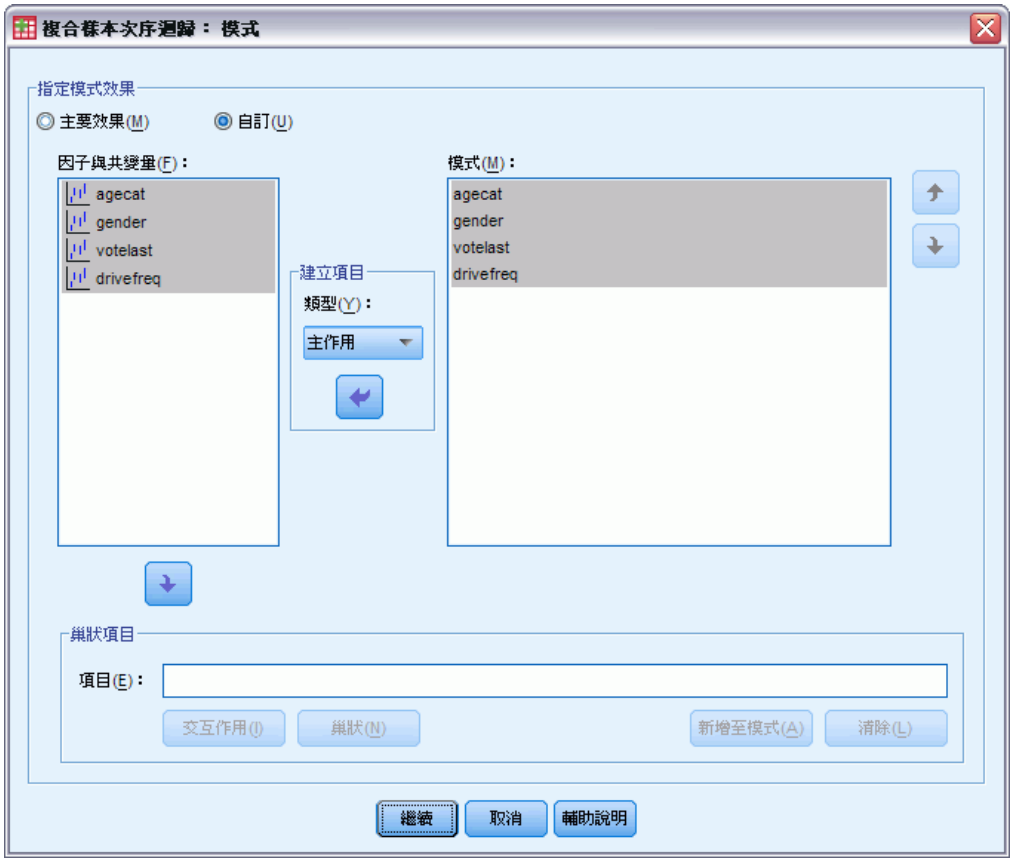
圖表 11-2
「次序迴歸反應機率」對話方塊



「反應機率」對話方塊能讓您指定反應的累積機率（即歸入並包含依變數特定類別的機率）是否要隨著依變數值的增減而增加。

複合樣本次序迴歸模式

圖表 11-3
「次序迴歸模式」對話方塊



指定模式效應。 依照預設值，程序會使用主對話方塊中指定的因子和共變量，建立主效果模式。否則您可建立包含交互作用效應與巢狀項次的自訂模式。

非巢狀項次

對所選擇的因子和共變量而言：

交互作用。 建立所有選取變數的最高階交互作用項。

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。

巢狀項次

您可以在這個程序中，為您的模式建立巢狀的項次。通常巢狀項次在建立因子或共變量效應項的模式時非常有用，但因子或共變量的值不可以與其他因子水準交互作用。例如，連鎖雜貨店可能會追蹤他們客戶在數個商店位置的消費習慣。因為每個客戶通常只在其中一個地點消費，因此您可以說客戶效果項是**巢狀**於商店位置效果項內。

此外，您可以包含交互作用效能，例如多項式項目與同一個共變量有關，或將多層的巢狀新增至巢狀項次。

限制。 巢狀項次有下列限制：

- 交互作用內的所有因子都必須是唯一的。因此，如果 A 是因子，那麼指定 A*A 是無效的。
- 巢狀效應項中的所有因子都必須是唯一的。因此，如果 A 是因子，那麼指定 A(A) 是無效的。
- 共變量內不可巢狀效果項。因此，如果 A 是因子，而 X 是共變量，那麼指定 A(X) 是無效的。

複合樣本次序迴歸統計量

圖表 11-4
「次序迴歸統計量」對話方塊

模式適合度。 控制測量總體模式效能的統計量顯示。

- **虛擬迴歸係數 (Pseudo R-square)。** 線性迴歸中的 R^2 統計量在次序迴歸模式中不具有確切的對等值。反而有多量數嘗試模仿 R^2 統計量的特性。
- **分類表。** 根據依變數上的預測模式類別，顯示表格形式的觀察類型交叉分類。

參數。 此群組可讓您控制與模式參數相關的統計量顯示。

- **估計值。** 顯示係數估計值。
- **指數化估計量。** 顯示自然對數的係數估計值次方的基準。其中估計量具有統計檢定、指數化估計量或 exp(B) 的良好特性，較易解譯。
- **標準誤。** 顯示各係數估計值的標準誤。
- **信賴區間。** 顯示各係數估計值的信賴區間。區間的信賴水準設定於「選項」對話方塊。
- **T 檢定。** 顯示各係數估計值的 t 檢定。各檢定的虛無假設其係數值為 0。
- **參數估計值的共變異數。** 顯示模式係數的共變異數矩陣估計值。
- **參數估計值的相關性。** 顯示模式係數的相關矩陣估計值。
- **設計效應。** 由簡單隨機樣本所取得變異數的估計值變異量比。這是所指定複合設計的效應量數值，其中遠離 1 的數值表示較大的效應。
- **設計效應的平方根。** 這是所指定複合設計的效應測量值，以類似標準誤的單位表示，其中遠離 1 的數值表示較大的效應。

平行線。 這個群組允許您以非平行線要求與模式相關聯的統計量，其中不同的迴歸線會套用至每個反應類別（除了最後一個以外）。

- **Wald 檢定。** 產生所有累積反應的迴歸參數都相等的虛無假設檢定。針對具有非平行線的模式加以估計，並套用相等參數的 Wald 檢定。
- **參數估計值。** 顯示具有非平行線之模式的係數和標準誤估計值。
- **參數估計值的共變異數。** 針對具有非平行線之模式的係數，顯示共變異數矩陣估計值。

模式變數的摘要統計量。 顯示依變數、共變異數和因子的摘要資訊。

樣本設計資訊。 顯示樣本相關摘要資訊，包括未加權個數和母群大小。

複合樣本假設檢定

圖表 11-5
「假設檢定」對話方塊

檢定統計量。 這個組別可以讓您選擇用來檢定假設的統計量類型。您可以在 F、調整後 F、卡方和調整後卡方之間選擇。

取樣自由度。 這個組別讓您可以控制用於為所有檢定統計量計算 p 值的取樣設計自由度。如果根據取樣設計時，數值為主要取樣單位的數目與第一階段取樣中的層數目之間的差異。此外，您可以透過指定正整數來設定自訂自由度。

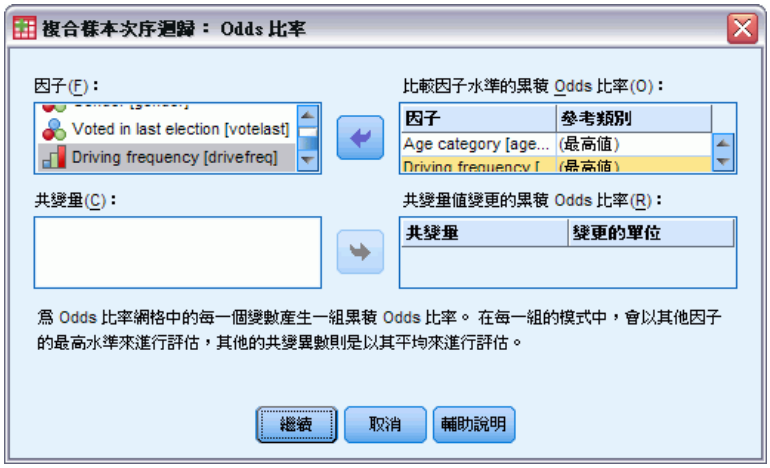
多重比較的調整。 在使用多重對比執行假設檢定時，可從顯著水準針對包含的對比調整整體顯著水準。此組別可讓您選擇調整方法。

- **最小顯著差異。** 這個方法無法控制以下假設之整體可能性，此假設為某些線性的對比不同於虛無假設值，。
- **循序 Sidak。** 這是循序逐步拒絕的 Sidak 程序；就拒絕個別假設而言，此程序做法相當不保守，但整體顯著水準仍維持相同。

- **循序 Bonferroni。** 這是循序逐步拒絕的 Bonferroni 程序；就拒絕個別假設而言，此程序做法相當不保守，但整體顯著水準仍維持相同。
- **Sidak。** 此方法的界限比 Bonferroni 法的界線更緊密
- **Bonferroni。** 此方法會調整觀察到的顯著水準，並實際檢定多重對比。

複合樣本次序迴歸 Odds 比率

圖表 11-6
次序迴歸 Odds 比率對話方塊



「Odds 比率」對話方塊可讓您顯示特定因子與共變異數的模式估計累積 odds 比率。這項功能僅適用於使用 Logit 連結函數的模式。會針對所有依變數類別（除了最後一個）計算單一累積 odds 比率；比例 odds 模式會假設它們都相等。

因子。 對於所選的每個因子，顯示在各因子類別中 odds 與特定參考類別中累積 odds 的比率。

共變量。 對於所選的每個共變量，顯示在平均共變量值加上特定變更單位中 odds 與平均數中累積 odds 的比率。

計算因子或共變量的 odds 比率時，程序會將其他所有因子固定在其最高值，以及將其他所有共變量固定在其平均值。若因子或共變量與模式中其他預測值交互作用，則 odds 比率不僅根據指定變數的變化而異，也會因交互作用的變數值而異。若指定的共變量在模式中自行交互作用（例如，age*age），則 odds 比率會根據共變量與共變量值的變化而異。

複合樣本次序迴歸儲存

圖表 11-7
「次序迴歸儲存」對話方塊

儲存變數。 此群組可讓您將模式預測值類別、預測類別機率、觀察類別機率、累積機率和預測機率儲存為作用中資料集中的新變數。

將模式匯出為 SPSS Statistics 資料。 以 IBM® SPSS® Statistics 格式寫入資料集，其中包含參數相關性、或共變異數矩陣，以及參數估計值、標準誤、顯著值和自由度。矩陣檔案的變數順序如下。

- **rowtype_。** 取得值（和數值標記）、COV（共變量）、CORR（相關性）、EST（參數估計值）、SE（標準誤）、SIG（顯著水準）和 DF（取樣設計自由度）。每個模式參數都有列類型為 COV（或 CORR）的個別觀察值，加上每個其他列類型的個別觀察值。
- **varname_。** 取得值 P1、P2、...，與所有模式參數的排序清單相對應，對於列類型 COV 或 CORR，其數值標記與參數估計表中顯示的參數字串相對應。其他列類型的儲存格為空白。
- **P1、P2、...** 這些變數與所有模式參數的次序清單相對應，其中的變數標記對應到在參數估計表中顯示的參數字串，並根據列類型取得值。對於冗餘參數，所有共變量會設為零；相關性會設為系統遺漏值；所有參數估計值會設為零；所有標準誤、顯著水準和殘差自由度會設為系統遺漏值。

注意：此檔案無法立即用於在其他讀取矩陣檔案的程序中做進一步的分析，除非這些程序接受所有在此匯出的列類型。

將模式匯出為 XML。 將參數估計值與參數共變異數（若選取的話）以 XML (PMML) 格式儲存為矩陣。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。

複合樣本次序迴歸選項

圖表 11-8
「次序的迴歸選項」對話方塊

估計方法。 您可以選取參數估計方法；在 Newton-Raphson、Fisher 分數或混合方法（會在切換至 Newton-Raphson 方法前，先執行 Fisher 分數疊代）之間選擇。如果在混合方法的 Fisher 分數階段期間，尚未到達 Fisher 疊代的最大數量就已達到收斂，則演算法會繼續進行 Newton-Raphson 方法。

估計。 此群組可讓您控制模式估計中使用的變數準則。

- **最大疊代。** 將執行運算的疊代最大值。指定一個非負的整數。
- **Step-Halving 最大值。** 每次疊代時，步驟大小會因乘以因子 0.5 而減少，直到對數概似增加或到達最大半階次數。指定一個正整數。
- **根據參數估計值的變更限制疊代。** 選取的話，演算法會在參數估計值之絕對或相對變更小於所指定數值的疊代之後停止，該數值不得為負數。
- **根據對數概似的變更限制疊代。** 選取的話，演算法會在對數概似函數之絕對或相對變更小於所指定數值的疊代之後停止，該數值不得為負數。

- **檢查資料點的完整分組。** 選取的話，演算法會執行檢定，以確定參數估計值具有唯一值。當程序能夠產生可正確分組各觀察值的模式，就會開始分組。
- **顯示疊代歷程。** 從第 0 個疊代（初始估計值）開始，每 n 個疊代顯示參數估計值與統計量。如果您選擇列印疊代歷程，則無論 n 的值為何，都一定會列印最後一次疊代。

使用者遺漏值。 尺度設計變數與依變數和任何共變異數都必須具備有效資料。這些變數具備無效資料的觀察值都將從分析中刪除。這些控制項可讓您決定是否要將使用者遺漏值視為層、集群、子母體和因子變數間的有效值。

信賴區間。 這是係數估計值、指數化係數估計值與 odds 比率的信賴區間水準。請指定大於或等於 50 且小於 100 的一個數值。

CSORDINAL 指令的其他功能

指令語法語言也可以讓您：

- 指定效應自訂檢定，或是效應（或值）的線性組合（使用 CUSTOM 次指令）。
- 計算因子和共變量的累積 odds 比率時，將其他模式變數的值固定在其平均值以外的值（使用 ODDSRATIOS 次指令）。
- 在要求 odds 比率時，將未標記的值當成因子的自訂參照類別使用（使用 ODDSRATIOS 次指令）。
- 指定檢查單一性時的允差值（使用 CRITERIA 次指令）。
- 產生一般估計量函數表（使用 PRINT 次指令）。
- 儲存 25 個以上的機率變數（使用 SAVE 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

複合樣本 Cox 迴歸

「複合樣本 Cox 迴歸」程序會針對透過複雜取樣方法抽取的樣本執行存活分析。或者，您也可以要求對子母體進行分析。

範例。某個政府法令執行機構想要瞭解其轄區內的再犯率。測量再犯率的其中一個量數，就是違法者第二次被逮補的間隔時間。該機構想要使用「Cox 迴歸」將再次被補的時間建立出一個模式，但卻擔心年齡類別之間的成比例風險假設無效。

醫學研究人員正在調查缺血性中風的病患在結束康復計畫後的存活時間。每個受試者可能會有多個觀察值，因為病患的病史會隨著重大的未死事件發生次數，以及這些事件記錄的次數而改變。樣本也是左側被截斷的，因為觀察的存活時間會被康復的時間長度而「誇大」，因為雖然風險是在缺血性中風時開始，但樣本只會有康復計畫後存活的病患。

存活時間。此程序會將 Cox 迴歸套用到存活時間的分析，存活時間就是事件發生前的時間長短。根據間隔開始時間的不同，而有兩種指定存活時間的方法：

- **Time=0。**一般而言，您會有每個受訪者間隔開始的完整資訊，並只有包含結束時間的變數 (或使用「日期 & 時間」變數的結束時間建立單一變數；如下所示)。
- **依受試者而定。**此方法適用於有**左截斷**時，亦稱作**延遲進入**；例如，如果您分析的是中風的病患在結束康復計畫後的存活時間，您可能會考量到他們的風險是在中風時開始。但是，如果您的樣本只包含康復計畫後存活的病患，則您樣本的左側會被截斷，因為觀察的存活時間會被康復長度而「誇大」。您可以透過將病患結束康復的時間指定為進入研究的時間，來說明此現象。

日期 & 時間變數。「日期 & 時間」變數無法用於直接定義間隔的開始與結束時間；如果您有「日期 & 時間」變數，您應使用這些變數來建立包含存活時間的變數。如果左側沒有截斷，只要根據進入研究的日期與觀察日期之間的差異，來建立包含結束時間的變數。如果左側被截斷，則可根據研究開始日期與進入日期之間的差異，建立包含開始時間的變數，並根據研究開始日期與觀察日期之間的差異，建立包含結束時間的變數。

事件狀態。您需要有一個變數，此變數會記錄受訪者在該間隔內是否經歷過所需的事件。發生事件的受訪者會右側受限。

受試者識別碼。您可以透過將單一受訪者的觀察值分成多個觀察值，輕鬆的納入成段常數、依時預測值。例如，您正在分析中風後病患的存活時間，代表醫療病史的變數應與預測值一樣的有用。經過一段時間後，他們開始碰到改變其醫療病史的重要醫療事件。下表顯示如果建構這類資料集的結構：病患 ID 為受訪者識別碼，結束時間定

義觀察的間隔，狀態記錄主要的醫療事件，心臟病的過去病史與出血的過去病史為成段常數、依時預測值。

病患 ID	結束時間	狀態	心臟病的過去病史	出血的過去病史
1	5	心臟病	否	否
1	7	出血	是(Y)	否
1	8	死亡	是(Y)	是(Y)
2	24	死亡	否	否
3	8	心臟病	否	否
3	15	死亡	是(Y)	否

假設。資料檔中的觀察值代表複合設計中，應根據「複合樣本計劃」對話方塊所選檔案裡規格進行分析的樣本。

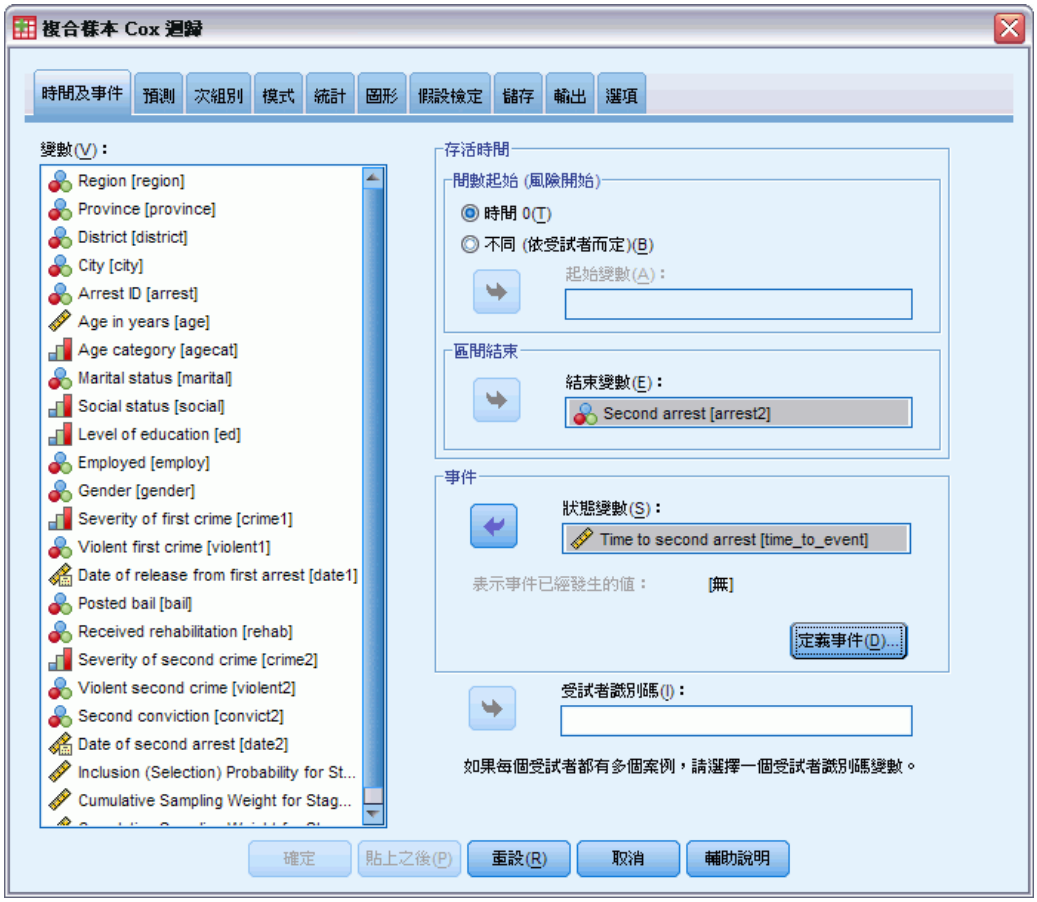
基本上，Cox 迴歸模式假設比例風險—亦即，從一個觀察值到另一個觀察值的風險比例不會隨著時間而改變。如果這個假設不成立，您必須將依時預測值新增至模式。

Kaplan-Meier 分析。如果您沒有選取任何預測值（或沒有將任何選取的預測值輸入到模式中），並在「選項」索引標籤上選擇乘積界限法來計算基準存活曲線，此程序會執行存活分析的 Kaplan-Meier 類型。

取得複合樣本 Cox 迴歸

- 從功能表選擇：
  分析(A) > 複合樣本 > Cox 迴歸...
- 選取計畫檔案。您可以選擇性選取自訂的聯合機率檔案。
- 按一下「繼續」。

圖表 12-1
「Cox 迴歸」對話方塊，「時間及事件」索引標籤

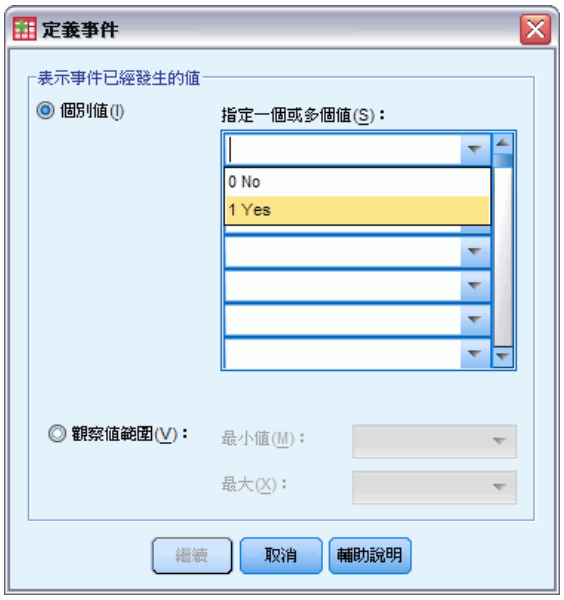


- 選取進入與離開研究的時間，來指定存活時間。
- 選取事件狀態變數。
- 按一下「定義事件」，然後至少定義一個事件值。

  您可選擇性地選取受訪者識別碼。

定義事件

圖表 12-2
「定義事件」對話方塊

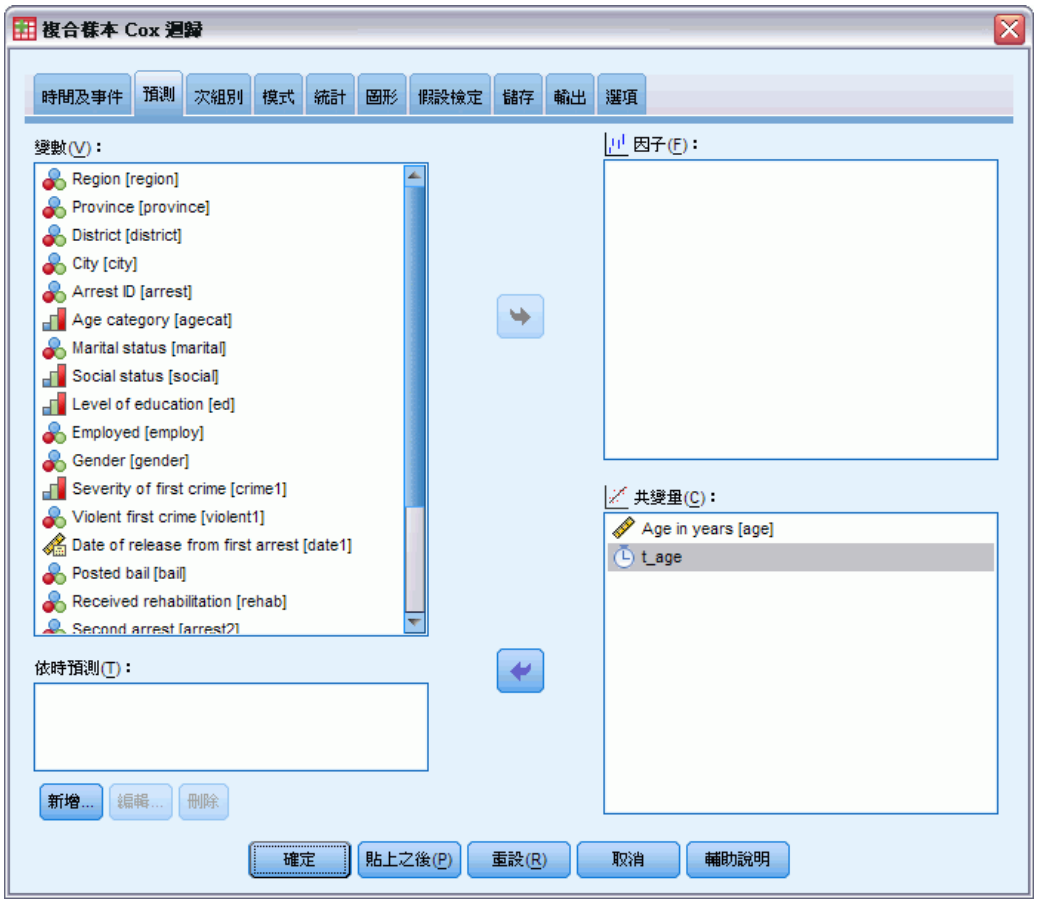


指定會指示終端事件已發生的值。

- **個別值。**指定一或多個值，方法是將這些值輸入網格中，或從含已定義值標記的值清單中選取這些值。
- **觀察值範圍。**指定一個範圍的值，方法是將這些值輸入網格中，或從含已定義值標記的清單中選取這些值。

預測值

圖表 12-3
「Cox 迴歸」對話方塊，「預測值」索引標籤



「預測值」索引標籤可讓您指定用來建立模式效應的因子與共變量。

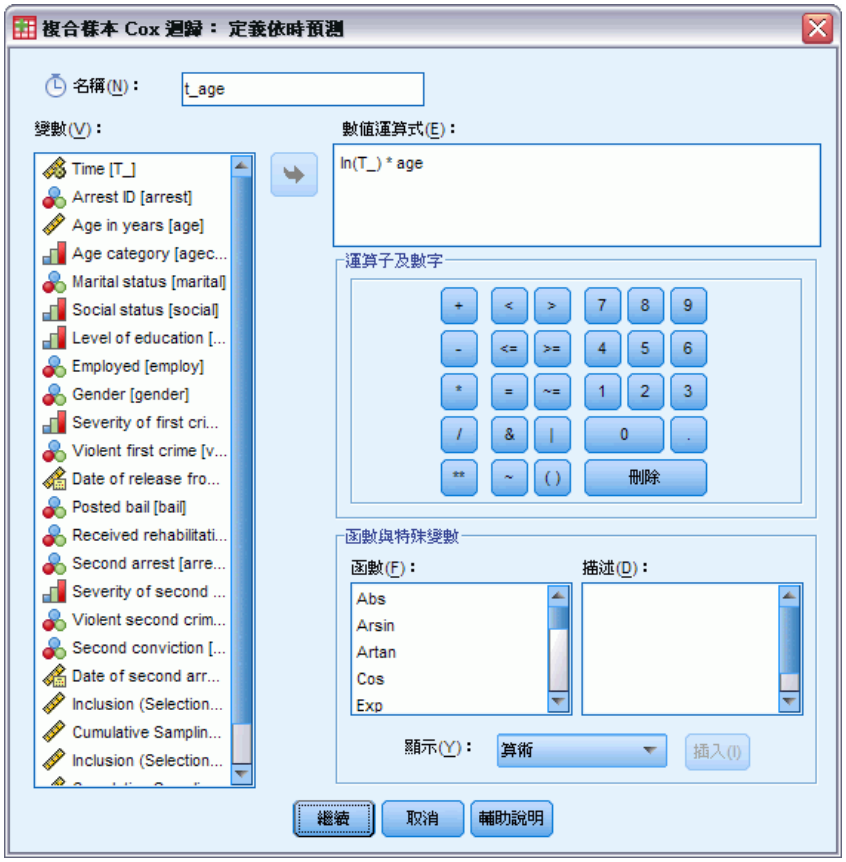
因子。因子為類別預測值，可為數值或字串。

共變量。共變量為尺度預測值，必須為數值。

依時預測。在某些狀況下，比例風險假設是不成立的。也就是說，風險比會隨著時間而改變，所以在不同時間點上，預測值中的某個值（或多個值）也會有所不同。在這種狀況下，您必須指定依時預測值。如需詳細資訊，請參閱第 72 頁中的定義依時預測值。可選取依時預測值作為因子或共變量。

定義依時預測值

圖表 12-4
Cox 迴歸「定義依時預測值」對話方塊



「定義依時預測值」對話方塊可讓您建立依存於內建時間變數 T_ 的預測值。您可以使用這個變數來定義依時共變量，常用的方式有下列兩種：

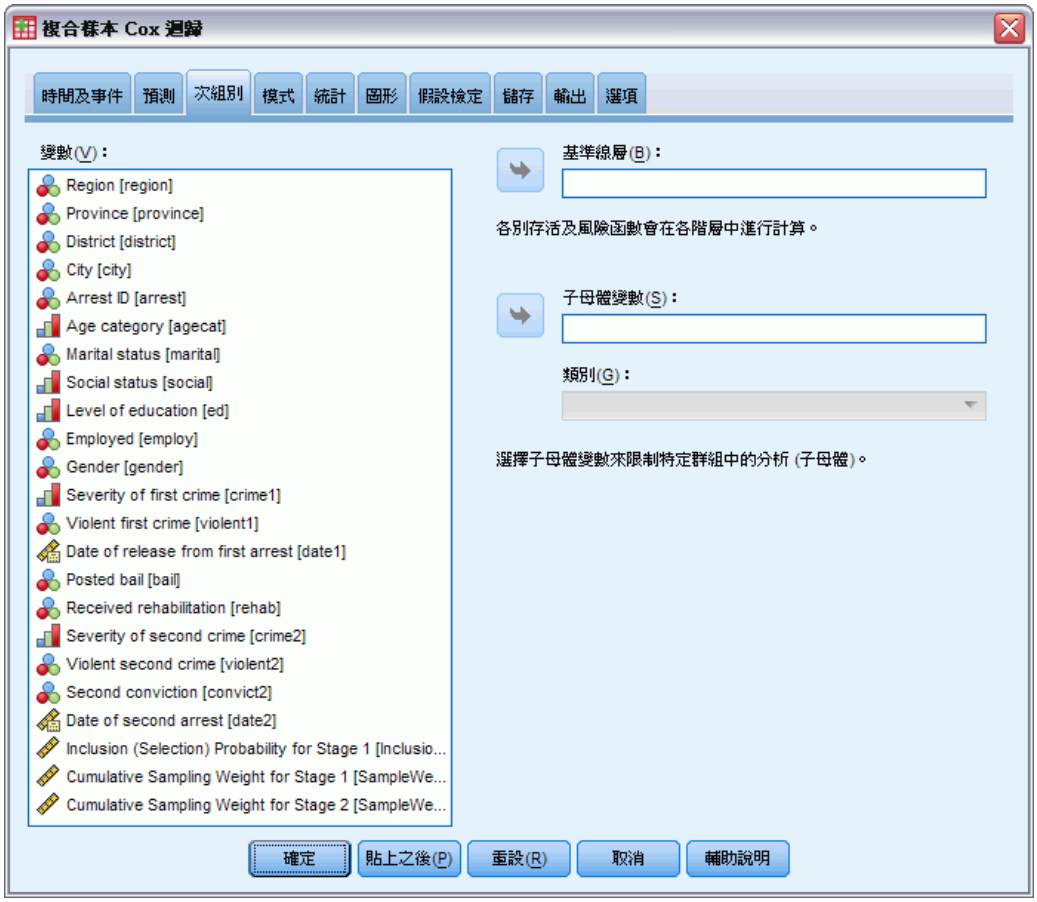
- 如果您想要估計允許不成比例風險的延伸式 Cox 迴歸模式，您可以將依時預測值定義為時間變數 T_ 的函數和問題共變量。例如時間變數和預測值的乘積，就是常見的簡單範例，但是您也可以指定較複雜的函數。
- 有些變數在不同的時間區內，可能會有不同的值，但系統上不見得一定跟時間有關。在這種情況下，您必須先定義**片段依時預測值**，而這種預測值可以使用邏輯運算式來加以定義。如果邏輯運算式的結果為真，其值為 1；如果是偽，其值是 0。然後再使用一連串的邏輯運算式，從一組測量值中建立依時預測值。例如，若您在進行研究的四個星期中，每星期測量一次血壓（識別為 BP1 到 BP4），則您可以將依時預測值定義為 (T_ < 1) * BP1 + (T_ >= 1 & T_ < 2) * BP2 + (T_ >= 2 & T_ < 3) * BP3 + (T_ >= 3 & T_ < 4) * BP4。請注意，對於任何的已知觀察值，括弧中的數值會剛好只有一項等於 1，剩下的都會等於 0。換句話說，我們可以將這個函數以文字解釋成「如果時間在一星期以內，使用 BP1；如果介於一星期和兩星期之間，使用 BP2，依此類推。

  注意：如果您的分段依時預測值是片段內的常數，如同上述的血壓範例，則您可以很容易透過將受訪者分割成數個觀察值，來指定成段常數、依時預測值。如需詳細資訊，請參閱複合樣本 Cox 迴歸第 67 頁中對「受試者識別碼」的討論。

在「定義依時預測值」對話方塊中，你可以使用函數建立控制項來建立依時共變量的運算式，或者將它直接輸入「數值運算式」文字區域中。注意字串常數必須以在引號或省略符號中，而數值常數必須以美制格式鍵入，並且以點做為小數分隔。結果變數名稱是您指定的名稱，且在「預測值」索引標籤中會包含結果變數作為因子或共變量。

次組別

圖表 12-5
「Cox 迴歸」對話方塊，「次組別」索引標籤

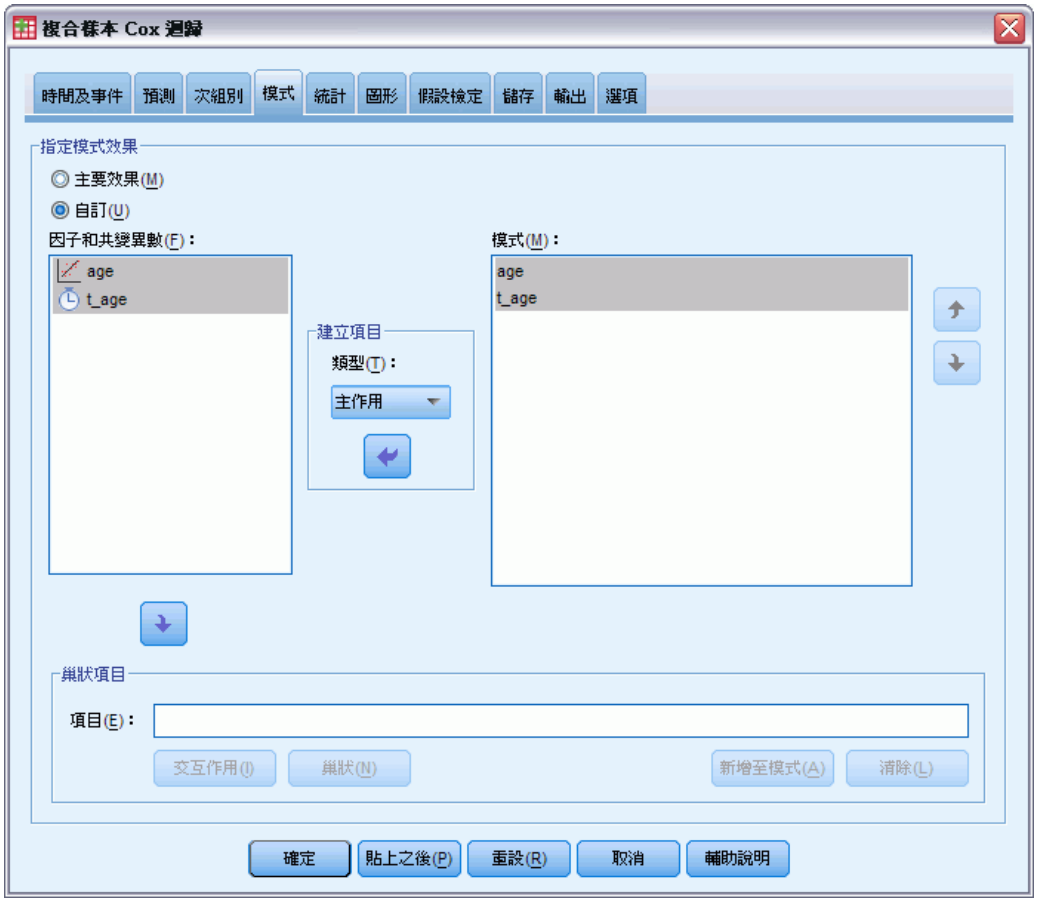


基準線層。針對此變數之每項數值所計算的個別基準線風險與存活函數，並會估計整層的單一模式係數集合。

子母體變數。指定變數以定義子母體。只會對子母體變數的所選類別執行分析。

模式

圖表 12-6
「Cox 迴歸」對話方塊，「模式」索引標籤



指定模式效應。 依照預設值，程序會使用主對話方塊中指定的因子和共變量，建立主效果模式。否則您可建立包含交互作用效應與巢狀項次的自訂模式。

非巢狀項次

對所選擇的因子和共變量而言：

交互作用。 建立所有選取變數的最高階交互作用項。

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。

巢狀項次

您可以在這個程序中，為您的模式建立巢狀的項次。通常巢狀項次在建立因子或共變量效應項的模式時非常有用，但因子或共變量的值不可以與其他因子水準交互作用。例如，連鎖雜貨店可能會追蹤他們客戶在數個商店位置的消費習慣。因為每個客戶通常只在其中一個地點消費，因此您可以說客戶效果項是**巢狀**於商店位置效果項內。

此外，您可以包含交互作用項（例如與相同的共變量有關的多項式項目）或新增多層巢狀結構到巢狀項次中。

限制。 巢狀項次有下列限制：

- 交互作用內的所有因子都必須是唯一的。因此，如果 A 是因子，那麼指定 A*A 是無效的。
- 巢狀效應項中的所有因子都必須是唯一的。因此，如果 A 是因子，那麼指定 A(A) 是無效的。
- 共變量內不可巢狀效果項。因此，如果 A 是因子，而 X 是共變量，那麼指定 A(X) 是無效的。

統計量

圖表 12-7
「Cox 迴歸」對話方塊，「統計量」索引標籤

樣本設計資訊。 顯示樣本相關摘要資訊，包括未加權個數和母群大小。

事件及設限摘要。顯示有關受限觀察值數量與百分比的摘要資訊。

事件時間的風險集。顯示每個基準線層中每個事件時間的事件數目與有風險數目。

參數。 此群組可讓您控制與模式參數相關的統計量顯示。

- **估計值。** 顯示係數估計值。
- **指數化估計量。** 顯示自然對數的係數估計值次方的基準。其中估計量具有統計檢定、指數化估計量或 exp(B) 的良好特性，較易解譯。
- **標準誤。** 顯示各係數估計值的標準誤。
- **信賴區間。** 顯示各係數估計值的信賴區間。區間的信賴水準設定於「選項」對話方塊。
- **t-檢定。** 顯示各係數估計值的 t 檢定。各檢定的虛無假設其係數值為 0。
- **參數估計值的共變異數。** 顯示模式係數的共變異數矩陣估計值。
- **參數估計值的相關性。** 顯示模式係數的相關矩陣估計值。
- **設計效應。** 由簡單隨機樣本所取得變異數的估計值變異量比。這是所指定複合設計的效應量數值，其中遠離 1 的數值表示較大的效應。
- **設計效應的平方根。** 這是所指定複合設計的效應測量值，其中遠離 1 的數值表示較大的效應。

模式假設。此組別可讓您產生成比例風險假設的檢定。該檢定會將適合的模式與替代模式比較，其中包括依時預測值 x* 每個預測值 x 的 _TF，其中 _TF 是特定的時間函數。

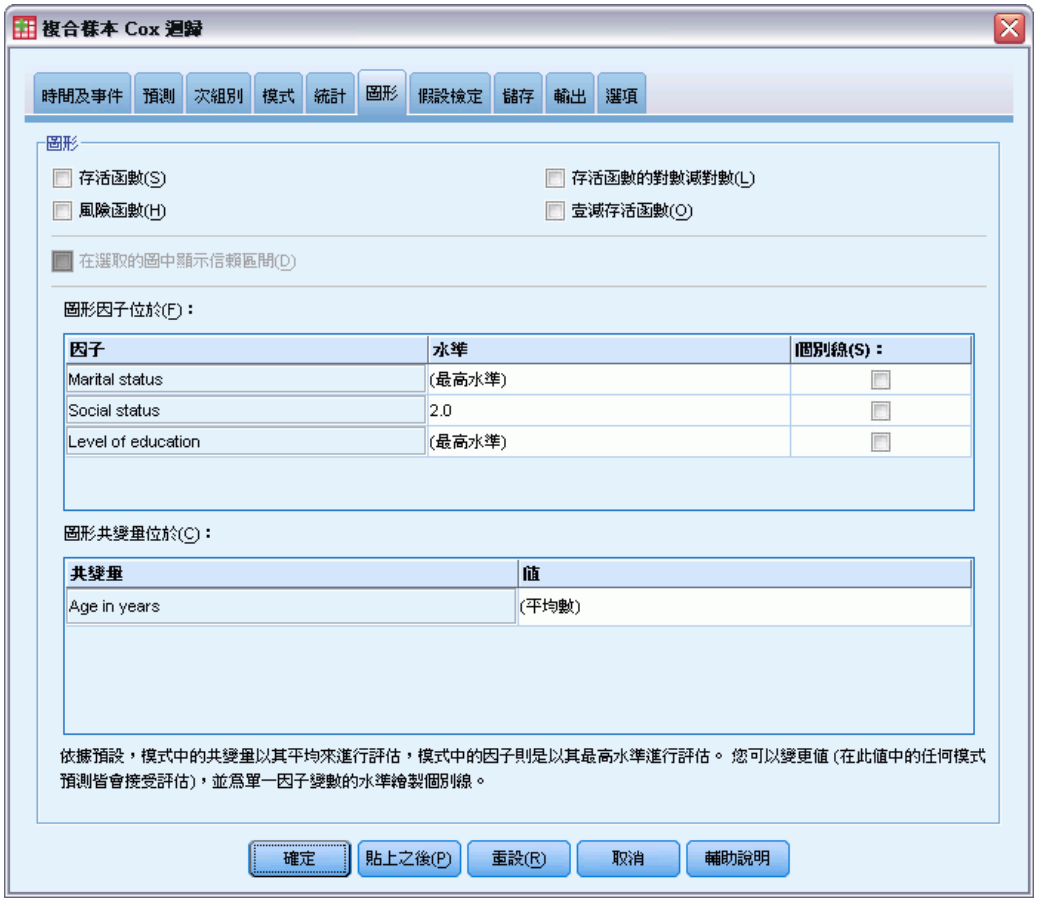
- **時間函數。**為替代模式指定 _TF 的形式。對於**識別**函數，_TF=T_。對於**對數**函數，_TF=log(T_)。對於 **Kaplan-Meier**，_TF=1−S_{KM}(T_)，其中 S_{KM}(.) 是存活函數的 Kaplan-Meier 估計值。對於**等級**，_TF 是所觀察結束時間之間的 T_ 等級排列。
- **替代模式的參數估計。**顯示替代模式中每個參數的估計值、標準誤與信賴區間。
- **替代模式的共變異數矩陣。**顯示替代模式中參數之間的估計共變異數矩陣。

基準線存活及累積風險函數。顯示基準線存活函數與基準線累積風險函數，以及其標準誤。

注意：如果模式中包含在「預測值」索引標籤上定義的依時預測值，則無法使用此選項。

圖形

圖表 12-8
「Cox 迴歸」對話方塊，「圖形」索引標籤



「圖形」索引標籤可讓您要求風險函數、存活函數、存活函數的對數減對數，與壹減存活函數的圖形。您也可以選擇繪製信賴區間與特定的函數；信賴區間是在「選項」索引標籤上設定。

預測值樣式。您可以在「匯出」索引標籤上指定供所要求圖形使用的預測值樣式，以及匯出的存活檔案。請注意，如果模式中包含在「預測值」索引標籤上定義的依時預測值，則無法使用這些選項。

- **圖形因子位於。**依照預設值，每個因子是以其最高水準評估的。視需要輸入或選取不同的水準。或者，您可以選擇為單一因子的每個水準繪製個別線，方法是選取該因子的核取方塊。
- **圖形共變量位於。**每個共變量是以其平均數評估的。視需要輸入或選取不同的值。

假設檢定

圖表 12-9
「Cox 迴歸」對話方塊，「假設檢定」索引標籤

檢定統計量。 這個組別可以讓您選擇用來檢定假設的統計量類型。您可以在 F、調整後 F、卡方和調整後卡方之間選擇。

取樣自由度。 這個組別讓您可以控制用於為所有檢定統計量計算 p 值的取樣設計自由度。如果根據取樣設計時，數值為主要取樣單位的數目與第一階段取樣中的層數目之間的差異。此外，您可以透過指定正整數來設定自訂自由度。

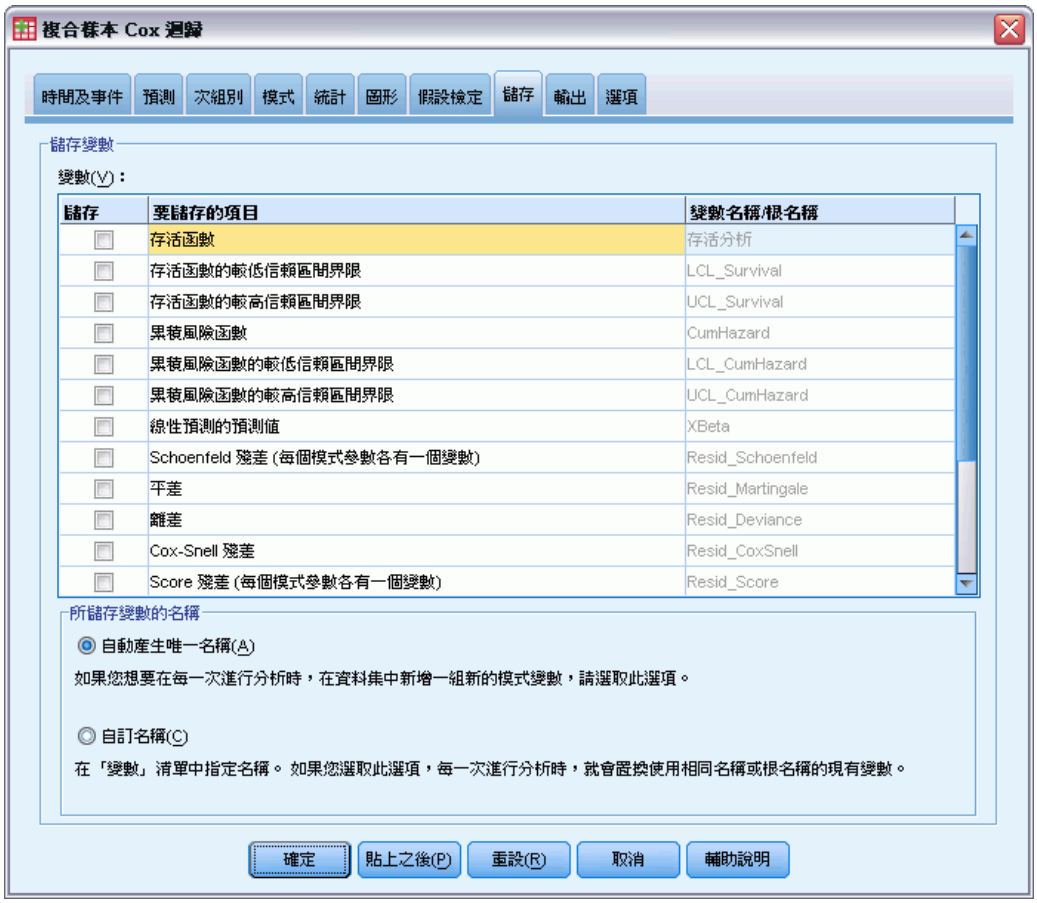
多重比較的調整。使用多重對比執行假設檢定時，則可從所包含對比的顯著水準調整整體的顯著水準。此組別可讓您選擇調整方法。

- **最小顯著差異。** 這個方法無法控制以下假設之整體可能性，此假設為某些線性的對比不同於虛無假設值，。
- **循序 Sidak。** 這是循序逐步拒絕的 Sidak 程序；就拒絕個別假設而言，此程序做法相當不保守，但整體顯著水準仍維持相同。
- **循序 Bonferroni。** 這是循序逐步拒絕的 Bonferroni 程序；就拒絕個別假設而言，此程序做法相當不保守，但整體顯著水準仍維持相同。

- **Sidak。** 此方法的界限比 Bonferroni 法的界線更緊密
- **Bonferroni。** 此方法會調整觀察到的顯著水準，並實際檢定多重對比。

儲存

圖表 12-10
「Cox 迴歸」對話方塊，「儲存」索引標籤



儲存變數。此組別可讓您將模式相關的變數儲存至作用中資料集，以進一步用於診斷與報告結果。請注意，當模式中包含依時預測值時，沒有任何一個變數可以使用。

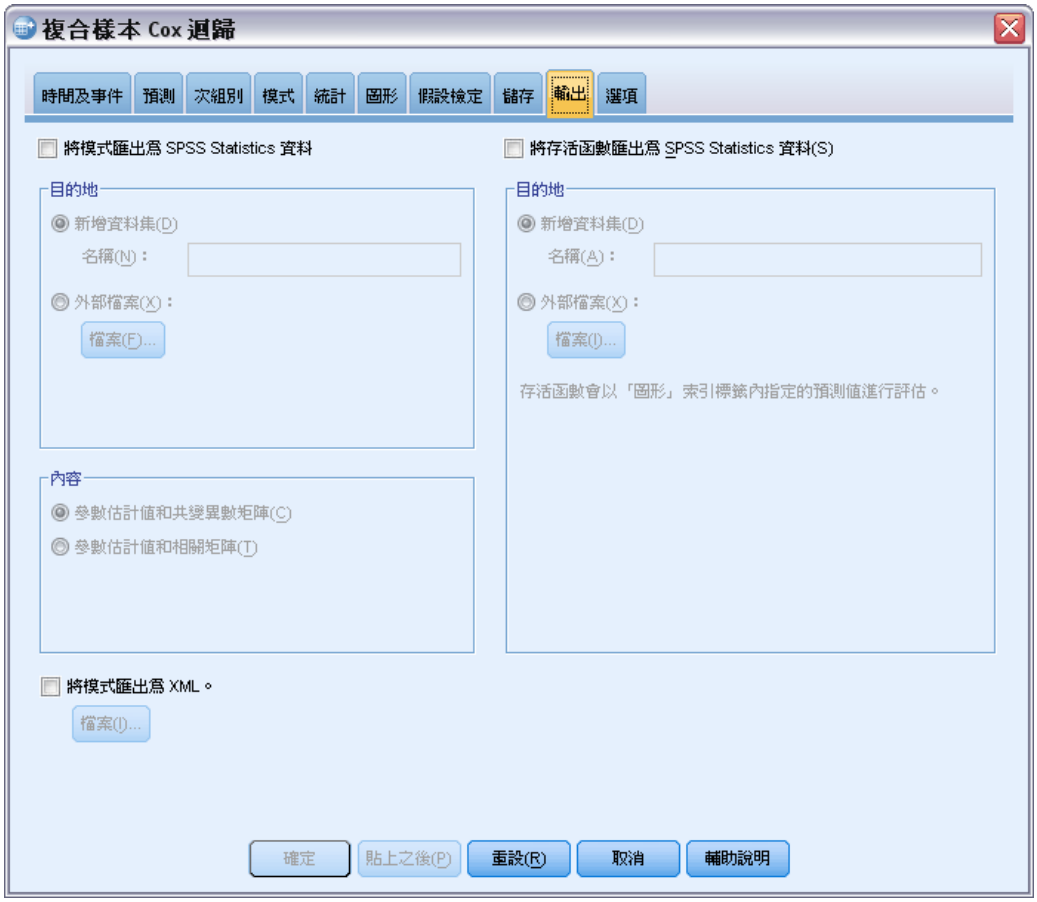
- **存活函數。**在每個觀察值的觀察時間與預測值儲存存活機率（存活函數值）。
- **存活函數的信賴區間下界。**在每個觀察值的觀察時間與預測值儲存存活函數的信賴區間下界。
- **存活函數的信賴區間上界。**在每個觀察值的觀察時間與預測值儲存存活函數的信賴區間上界。
- **累積風險函數。**在每個觀察值的觀察時間與預測值儲存累積風險，或−ln(存活)。
- **累積風險函數的信賴區間下界。**在每個觀察值的觀察時間與預測值儲存累積風險函數的信賴區間下界。

- **累積風險函數的信賴區間上界。**在每個觀察值的觀察時間與預測值儲存累積風險函數的信賴區間上界。
- **線性預測的預測值。**儲存參考值已修正的預測值乘以迴歸係數的線性組合。線性預測值是風險函數對基準線風險的比率。在比例風險模式之下，此值在一段時間裡都是常數。
- **Schoenfeld 殘差。**對於模式中每個非受限觀察值與每個非多餘參數，Schoenfeld 殘差是與模式參數相關的預測值觀察值，與觀察事件時間中所設定有風險觀察值其預測值期望值之間的差異。Schoenfeld 殘差可用於協助評估成比例風險假設，例如對於預測值 x，如果成比例風險成立的話，依時預測值 x*ln(T_) 對時間的 Schoenfeld 殘差圖形，應在 0 顯示水平線。在模式中，會為每個非多餘參數儲存個別的變數。僅針對非受限觀察值計算 Schoenfeld 殘差。
- **平差。**對於每個觀察值，平差是指在觀察期間，觀察到的設限（如果受限，則為 0，如果非受限，則為 1）與事件期望值之間的差異。
- **離差。**離差是指「調整後」的平差，看起來會與 0 更為對稱。離差對預測值的圖形應沒有顯示樣式。
- **Cox-Snell 殘差。**對於每個觀察值，Cox-Snell 殘差是觀察期間事件的期望值，或是觀察到的受限減去平差。
- **Score 殘差。**對於模式中的每個觀察值與每個非多餘參數，Score 殘差是對虛擬概似值其第一個微分的觀察值貢獻。在模式中，會為每個非多餘參數儲存個別的變數。
- **DFBeta 殘差。**對於模式中的每個觀察值與每個非多餘參數，DFBeta 殘差近似於當觀察值從模式中移除時參數估計值的變更。DFBeta 殘差相當大的觀察值會在分析上運用不當的影響力。在模式中，會為每個非多餘參數儲存個別的變數。
- **整合殘差。**當多個觀察值代表單一受訪者時，該受訪者的整合殘差就是所有屬於同一個受訪者的觀察值其相對應觀察值殘差的加總。對於 Schoenfeld 殘差，整合的 Schoenfeld 殘差與非整合的 Schoenfeld 殘差相同，因為只會對非受限的觀察值定義 Schoenfeld 殘差。只有已在「時間及事件」索引標籤上指定受訪者識別碼時，這些殘差才能使用。

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。

匯出

圖表 12-11
「Cox 迴歸」對話方塊，「匯出」索引標籤



將模式匯出為 SPSS Statistics 資料。以 IBM® SPSS® Statistics 格式寫入資料集，其中包含參數相關性、或共變異數矩陣，以及參數估計值、標準誤、顯著值和自由度。矩陣檔案的變數順序如下。

- **rowtype_。**取得值（和數值標記）、COV（共變量）、CORR（相關性）、EST（參數估計值）、SE（標準誤）、SIG（顯著水準）和 DF（取樣設計自由度）。每個模式參數都有列類型為 COV（或 CORR）的個別觀察值，加上每個其他列類型的個別觀察值。
- **varname_。**取得值 P1、P2、...，與所有模式參數的排序清單相對應，對於列類型 COV 或 CORR，其數值標記與參數估計表中顯示的參數字串相對應。其他列類型的儲存格為空白。
- **P1、P2、...** 這些變數與所有模式參數的次序清單相對應，其中的變數標記對應到在參數估計表中顯示的參數字串，並根據列類型取得值。對於冗餘參數，所有共變量會設為零；相關性會設為系統遺漏值；所有參數估計值會設為零；所有標準誤、顯著水準和殘差自由度會設為系統遺漏值。

注意：此檔案無法立即用於在其他讀取矩陣檔案的程序中做進一步的分析，除非這些程序接受所有在此匯出的列類型。

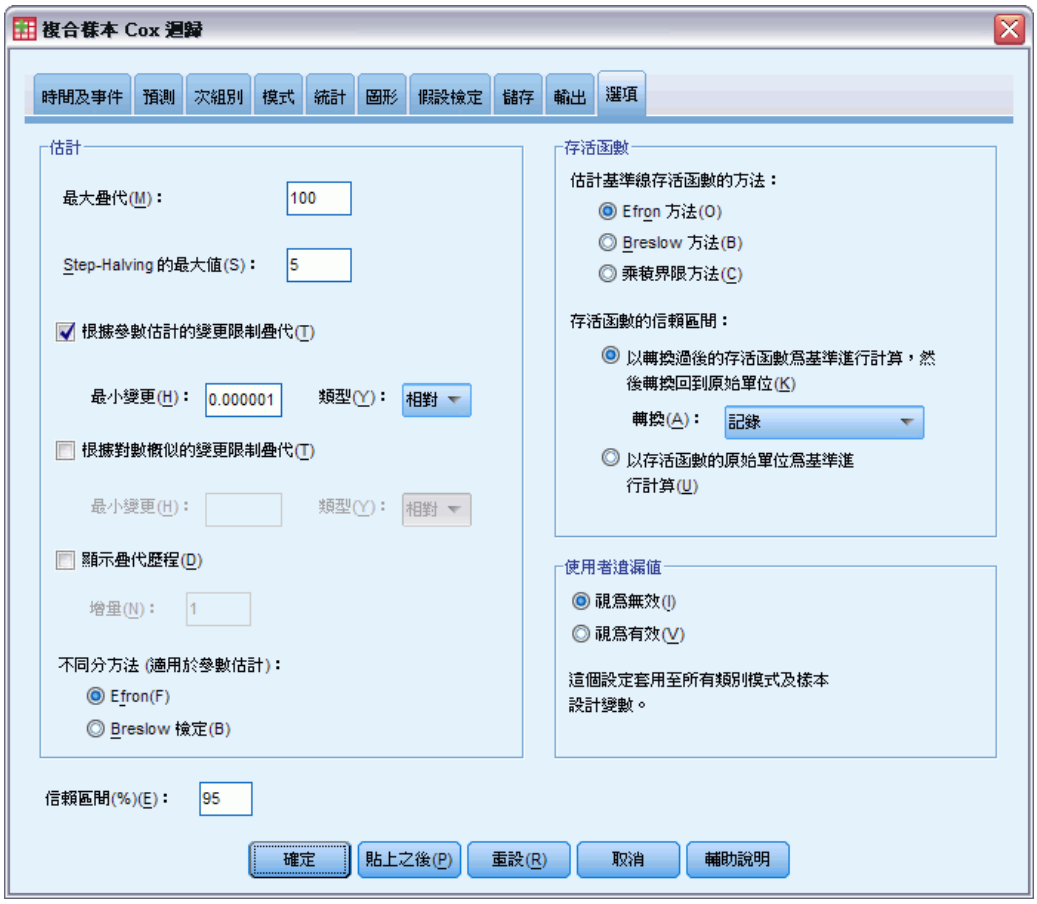
匯出存活函數為 SPSS Statistics 資料。以包含存活函數的 SPSS Statistics 格式寫入資料集；存活函數的標準誤；存活函數信賴區間的上界與下界；每個失敗或事件時間的累積風險函數，以在「圖形」索引標籤上指定的基準線與預測值樣式加以評估。矩陣檔案的變數順序如下。

- **基準線層變數。**為層變數的每個值產生不同的存活表格。
- **存活時間變數。**事件時間；為每個唯一的事件時間建立不同的觀察值。
- **Sur_0，LCL_Sur_0，UCL_Sur_0。**基準線存活函數與其信賴區間的上界和下界。
- **Sur_R，LCL_Sur_R，UCL_Sur_R。** 以「參考」樣式（請參閱在輸出中的樣式值表）與其信賴區間的上界和下界評估的存活函數。
- **Sur_#.#，LCL_Sur_#.#，UCL_Sur_#.#，…**以在「圖形」索引標籤指定的每個預測值樣式與其信賴區間的上界和下界評估的存活函數。請參閱在與編號為 #.# 樣式符合的輸出中的樣式值表格。
- **Haz_0，LCL_Haz_0，UCL_Haz_0。**基準線累積風險函數與其信賴區間的上界和下界。
- **Haz_R，LCL_Haz_R，UCL_Haz_R。**以「參考」樣式（請參閱在輸出中的樣式值表）與其信賴區間的上界和下界評估的累積風險函數。
- **Haz_#.#，LCL_Haz_#.#，UCL_Haz_#.#，…**以在「圖形」索引標籤指定的每個預測值樣式與其信賴區間的上界和下界評估的累積風險函數。請參閱在與編號為 #.# 樣式符合的輸出中的樣式值表格。

將模式匯出為 XML。以 XML (PMML) 格式儲存預測存活函數所需的所有資訊，包括參數估計值與基準線存活函數。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。

選項

圖表 12-12
「Cox 迴歸」對話方塊，「選項」索引標籤



估計。這些控制項可指定估計迴歸係數的條件。

- **最大疊代。** 將執行運算的疊代最大值。指定一個非負的整數。
- **Step-Halving 最大值。** 每次疊代時，步驟大小會因乘以因子 0.5 而減少，直到對數概似增加或到達最大半階次數。指定一個正整數。
- **根據參數估計值的變更限制疊代。**選取的話，演算法會在參數估計值之絕對或相對變更小於所指定數值的疊代之後停止，該數值必須為正數。
- **根據對數概似的變更限制疊代。**選取的話，演算法會在對數概似函數之絕對或相對變更小於所指定數值的疊代之後停止，該數值必須為正數。
- **顯示疊代歷程。**顯示參數估計值與虛擬對數概似值的疊代歷程，並列印在參數估計值與虛擬對數概似值中變更的最後估計。疊代歷程表格會從第 0 個疊代（初始疊代）開始，列印出每 n 個疊代，其中 n 為增量值。若要求疊代歷程，則永遠會顯示最後一個疊代，無論 n 的值為何。
- **不同分方法（適用於參數估計）。**當有同分的觀察失敗次數時，會使用其中一個方法讓它們變成不同分。Efron 方法的計算代價較高。

存活函數。這些控制項可指定涉及存活函數計算的條件。

- **估計基準線存活函數的方法。**Breslow（或 Nelson-Aalan 或經驗）方法會以非下降逐步函數，並以觀察到的失敗次數來逐步估計基準線累積風險，然後使用關係存活=exp(−累積風險）計算基準線存活。Efron 方法的計算代價較高，在不同分時會降低為 Breslow 方法。**乘積界限**方法會估計非上升正確連續函數的基準線存活；當模式中沒有預測值時，此方法會降低為 Kaplan-Meier 估計。
- **存活函數的信賴區間。**信賴區間的計算方式有三種：以原始值、透過對數轉換，或負對數轉換。只有負對數轉換才會保證信賴區間的界限會介於 0 到 1 之間，但對數轉換一般看起來會執行「最適」。

使用者遺漏值。所有變數必須具有要納入分析之觀察值的有效值。這些控制項可讓您決定是否要將使用者遺漏值視為類別模式（包括因子、事件、層與子母體變數）與取樣設計變數間的有效值。

信賴區間 (%)。這是供係數估計值、指數化係數估計值、存活函數估計值與累積風險函數估計值使用的信賴區間水準。指定大於等於 1 且小於 100 的一個數值。

CSCOXREG 指令的其他功能

指令語言讓您也可以：

- 執行自訂的假設檢定（使用 CUSTOM 次指令與 /PRINT LMATRIX）。
- 允差規格（使用 /CRITERIA SINGULAR）。
- 一般估計量函數表（使用 /PRINT GEF）。
- 多重預測值樣式（使用多個 PATTERN 次指令）。
- 指定根名稱時已存變數數目上限（使用 SAVE 次指令）。對話方塊允許 25 個變數的 CSCOXREG 預設值。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

部 II:
範例

章節
13

複合樣本取樣精靈

「取樣精靈」會帶領您進行建立、修改或執行取樣計劃檔案步驟。在您使用「精靈」之前，您腦海中應具備已定義完全的目標母群、取樣單位清單、及合適的樣本設計。

從「完整取樣框」獲得樣本

州立機構會確定各郡財產稅均合理公平後才徵收。稅額根據財產的估計值而得，故機構會希望調查各郡的財產樣本，以確定各郡的紀錄均保持最新。然而，取得目前估價的資源有限，故可用的資源通常會被明智的應用是很重要的。機構決定使用複雜取樣方法選取財產樣本。

屬性清單收集於 property_assess_cs.sav 中。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 使用「複合樣本取樣精靈」選取樣本。

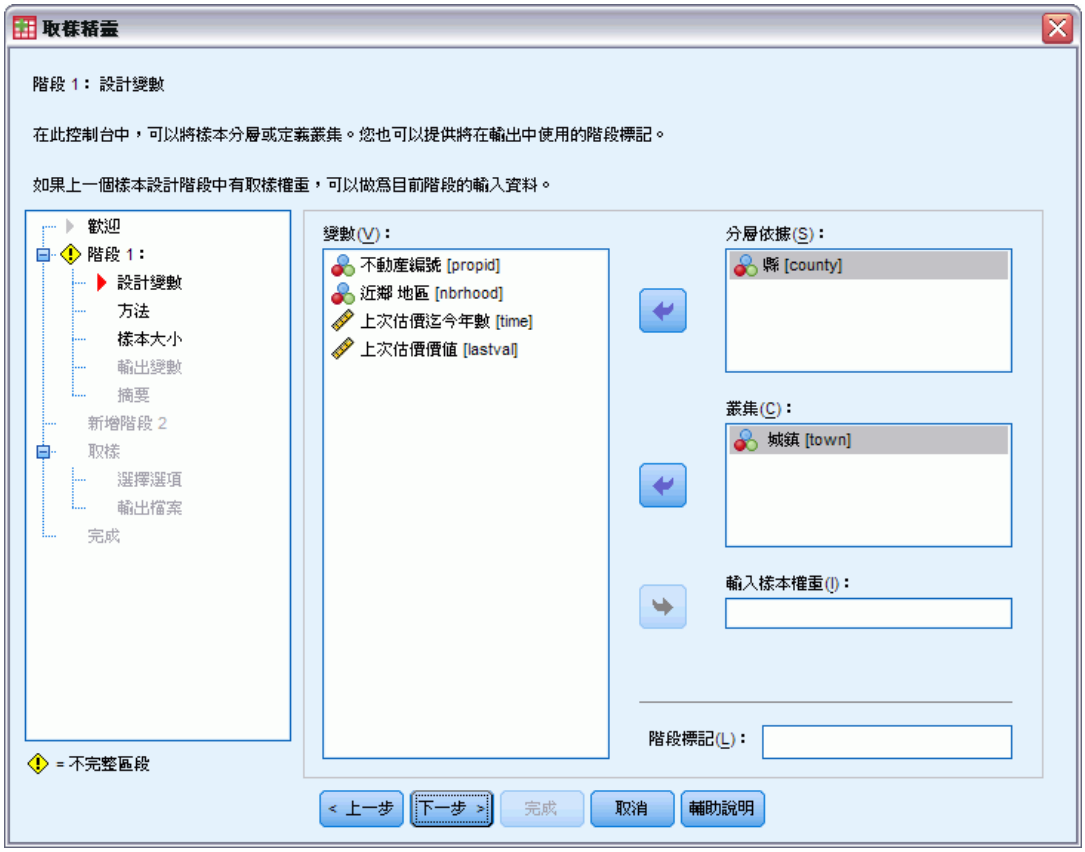
使用精靈

- 若要執行「複合樣本取樣精靈」，請從功能表選擇：
  分析(A) > 複合樣本 > 選擇樣本...

圖表 13-1
「取樣精靈 - 歡迎」步驟

- 選取「設計樣本」，瀏覽至您要儲存檔案的位置，然後輸入 property_assess.csplan 做為計劃檔的名稱。
- 按一下「下一步」。

圖表 13-2
取樣精靈 - 設計變數步驟（階段 1）



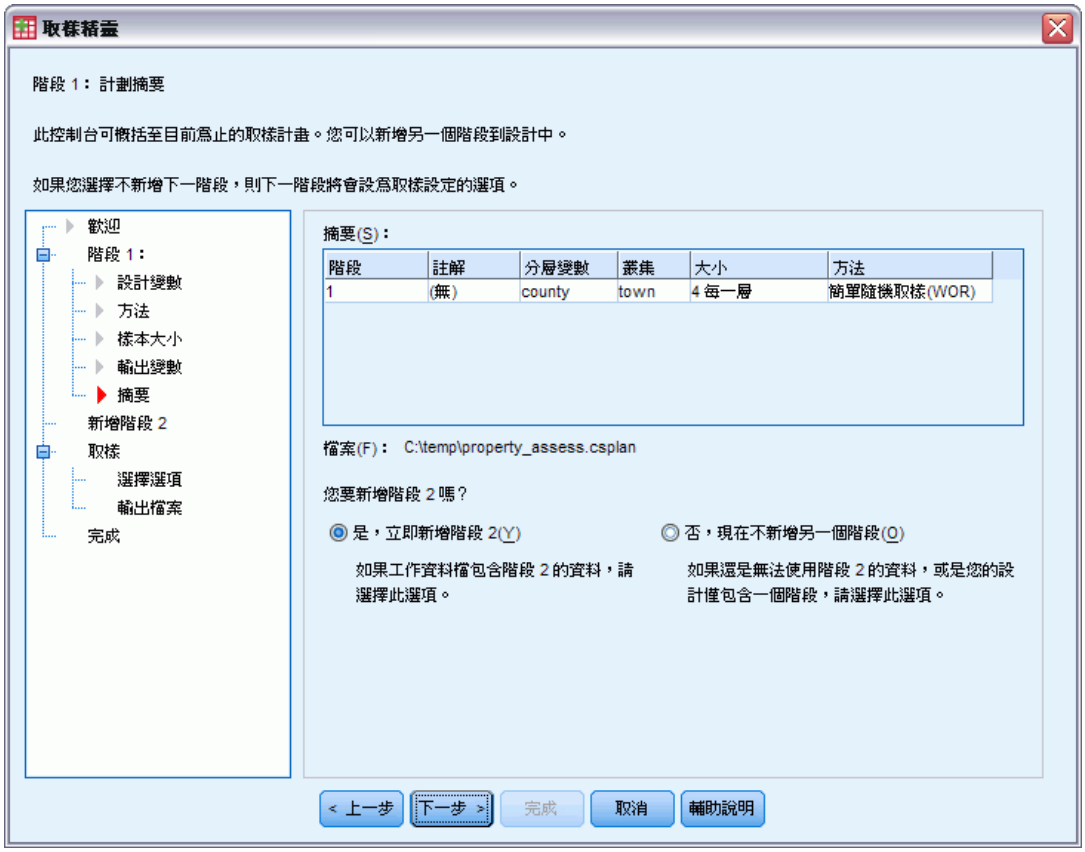
- 選取郡做為分層變數。
- 選取鎮做為集群變數。
- 按一下「**下一步**」，並在「取樣方法」步驟按一下「**下一步**」。

  設計結構表示從各郡抽出獨立樣本。在此階段中，會使用預設方法「簡單隨機取樣」抽出鎮，做為主要取樣單位。

圖表 13-3
取樣精靈 - 取樣大小步驟（階段 1）

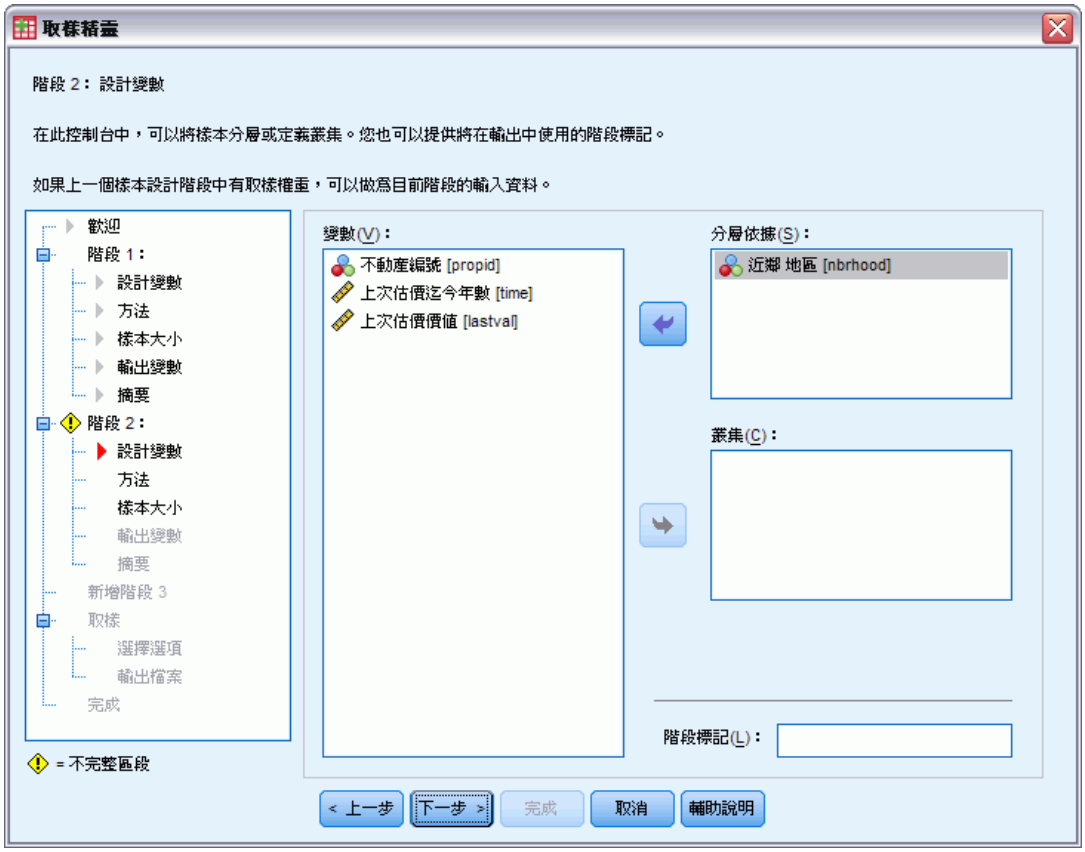
- 從「單位」下拉式清單中選取**個數**。
- 輸入 **4** 做為要在此階段選取的單位個數數值。
- 按一下「**下一步**」，並在「輸出變數」步驟按一下「**下一步**」。

圖表 13-4
取樣精靈 - 計劃摘要步驟（階段 1）



- ▶ 選取「是，立即新增階段 2」。
- ▶ 按一下「下一步」。

圖表 13-5
取樣精靈 - 設計變數步驟（階段 2）



- 選取鄰近做為分層變數。
- 按一下「**下一步**」，並在「取樣方法」步驟按一下「**下一步**」。

  此設計結構表示階段 1 中抽出的鎮之鄰近各範圍所抽出的獨立樣本。在此階段中，會使用「簡單隨機取樣」抽出財產做為主要取樣單位。

圖表 13-6
取樣精靈 - 取樣大小步驟（階段 2）

- 從「單位」下拉式清單中選取**財產**。
- 輸入 **0.2** 做為各層中樣本單元的比例數值。
- 按一下「**下一步**」，並在「輸出變數」步驟按一下「**下一步**」。

圖表 13-7
取樣精靈 - 計劃摘要步驟（階段 2）

► 檢查取樣設計，再按一下「下一步」。

圖表 13-8
「取樣精靈 - 抽樣 - 選擇選項」步驟

- 在要使用的亂數種子類型選取「**自訂值**」，並輸入 **241972** 做為數值。

  使用自訂值可讓您確實複製此範例的結果。

- 按一下「**下一步**」，並在「抽樣輸出檔案」步驟按一下「**下一步**」。

圖表 13-9
取樣精靈 - 完成步驟

▶ 按一下「完成」。

這些選擇會產生取樣計劃檔案 property_assess.csplan，並根據此計畫抽出樣本。

計畫摘要

圖表 13-10
計劃摘要

摘要

			階段 1	階段 2
設計變數	分層	1	省	分區
	集群	1	區	
樣本資訊	選擇方法		簡單隨機，取樣後不放回	簡單隨機，取樣後不放回
	取樣的單位數		4	
	已建立或修改的變數	階段性包含 (選擇) 機率	階段 1 的包含 (選擇) 機率	包含機率_2_
		階段性累積樣本權重	階段 1 的累積樣本權重	樣本權重累積_2_
	取樣的單位比例			,2
分析資訊	估計式假設		相等機率，取樣後不放回	相等機率，取樣後不放回
	包含機率		自階段 1 的變數包含 (選擇) 機率取得	自變數包含機率_2_取得

計劃檔案：C:\property_assess.csplan
加權變數：最後取樣權重

摘要表會檢閱您的取樣計劃，且在確認該計劃是否能夠表達您目的時非常實用。

取樣摘要

圖表 13-11
階段摘要

	取樣的單位數		取樣的單位比例	
County	要求的	實際	要求的	實際
Eastern	4	4	44,4%	44,4%
Central	4	4	57,1%	57,1%
Western	4	4	25,0%	25,0%
Northern	4	4	44,4%	44,4%
Southern	4	4	50,0%	50,0%

此摘要表會檢閱取樣的第一階段，且在檢查取樣是否按照計劃進行時非常實用。依照您的要求，各郡均取樣四個鎮。

圖表 13-12
階段摘要

縣	城鎮	鄰近地區	Number of Units Sampled		Proportion of Units Sampled	
			Requested	Actual	Requested	Actual
南部	88	613	14	14	20.0%	20.6%
		612	15	15	20.0%	19.5%
		611	11	11	20.0%	20.4%
		610	16	16	20.0%	20.3%
	82	571	15	15	20.0%	20.3%
		570	14	14	20.0%	19.7%
		569	15	15	20.0%	19.5%
		568	8	8	20.0%	20.5%
	84	585	6	6	20.0%	20.7%
		582	11	11	20.0%	19.3%
		584	11	11	20.0%	19.3%
		583	14	14	20.0%	19.7%
		587	11	11	20.0%	20.4%
		586	8	8	20.0%	21.1%
	81	564	6	6	20.0%	21.4%
		563	5	5	20.0%	20.8%
		562	12	12	20.0%	19.4%
		561	15	15	20.0%	19.5%
		565	13	13	20.0%	19.7%
東部	6	36	13	13	20.0%	20.3%
		38	13	13	20.0%	20.6%
		37	14	14	20.0%	20.6%
	7	44	11	11	20.0%	19.6%

此摘要表（最上面部分顯示於此）可檢閱取樣的第二階段。在檢查取樣是否依照計劃進行時亦十分有用。依照要求，約取樣出第一階段所取樣的各鎮各鄰近範圍之 20% 的財產。

樣本結果

圖表 13-13
具有樣本結果的資料編輯程式

	propid	nbrhood	town	county	time	lastval	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
273	577.0	8	2	1	4	181.70	.	.	.	.	.
274	578.0	8	2	1	5	189.60	.	.	.	.	.
275	579.0	8	2	1	4	200.10	.	.	.	.	.
276	580.0	8	2	1	5	211.50	.	.	.	.	.
277	581.0	8	2	1	4	181.50	.	.	.	.	.
278	641.0	9	2	1	7	192.40	.	.	.	.	.
279	642.0	9	2	1	6	236.70	.44	2.25	.21	10.93	10.93
280	643.0	9	2	1	6	150.40	.44	2.25	.21	10.93	10.93
281	644.0	9	2	1	8	204.80	.	.	.	.	.
282	645.0	9	2	1	6	225.40	.	.	.	.	.
283	646.0	9	2	1	7	180.80	.44	2.25	.21	10.93	10.93
284	647.0	9	2	1	5	176.90	.	.	.	.	.

資料檢視 變數檢視

您可在「資料編輯程式」中檢視取樣結果。 會有五個新變數儲存至工作檔中，表示各階段的包含機率和累積樣本權重，加上最後的取樣權重。

- 將會選出具備這些變數值的觀察值進行取樣。
- 而不會選取具備變數系統遺漏值的觀察值。

機構現可使用其資源，在樣本中收集所選財產的目前估價。只要取得這些估價，您就可使用 property_assess.csplan，以「複合樣本」分析程序處理這些樣本。

從「部分取樣框」獲得樣本

有家公司喜歡匯集並販賣高品質調查資訊的資料庫。他們必須需有效率地取得具代表性的調查樣本，故使用複雜取樣方法。完整取樣設計需要下列結構：

階段	分層變數	叢集
1	地區	省
2	區	城市
3	分區	

在第三階段中，家庭為主要取樣單位，而會調查所選家庭。然而，因為較易取得的資料只到城市層級，故該公司計畫先執行設計的前兩階段，再從取樣的城市中收集分區和家庭個數等資訊。城市層級的資訊收集於 demo_cs_1.sav 中。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 請注意，在本檔案中有一個變數「分區」，內容全部為 1。這是「真」變數的預留位置，其值將在該設計前兩階段執行後才會取得，此讓您可以現在就指定完整的三階取樣設計。請使用「複合樣本取樣精靈」來指定完整的複合取樣設計，然後在前兩階段取樣。

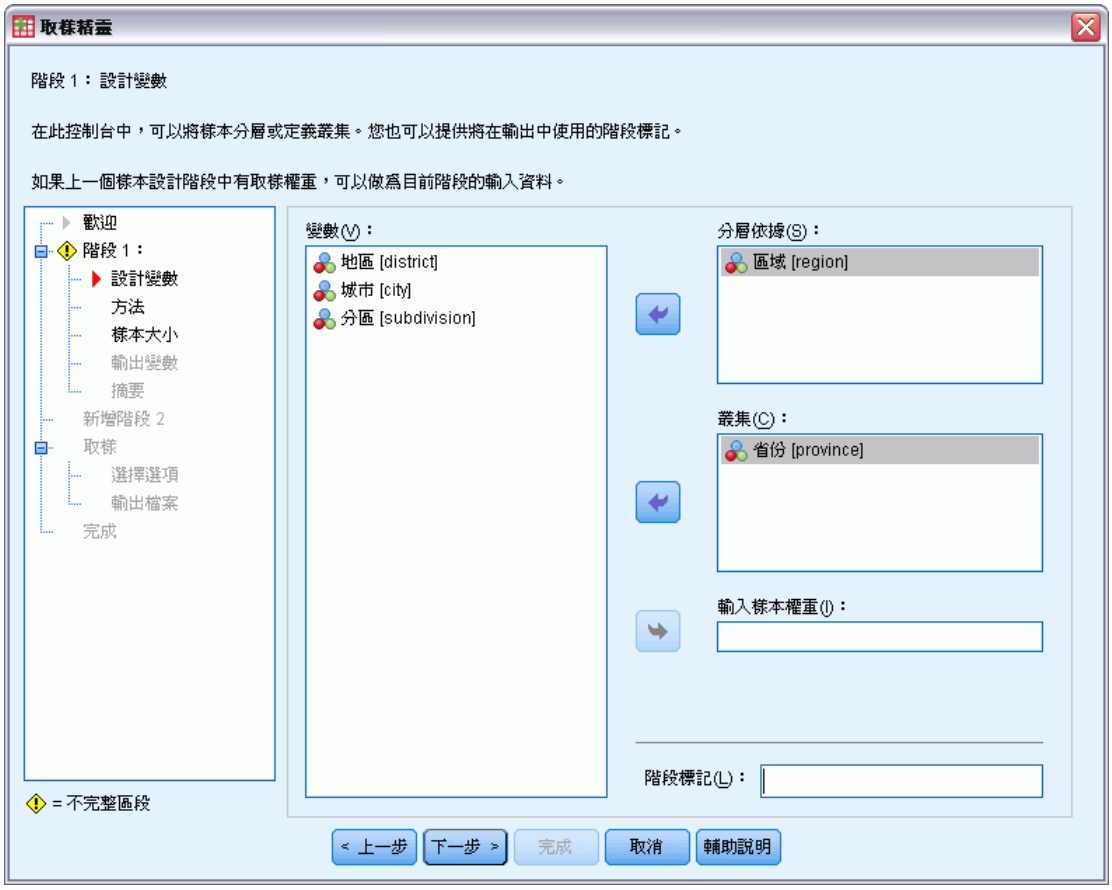
從第一個部分框使用精靈取樣

▶ 若要執行「複合樣本取樣精靈」，請從功能表選擇：
分析(A) > 複合樣本 > 選擇樣本...

圖表 13-14
「取樣精靈 - 歡迎」步驟

- 選取「設計樣本」，瀏覽至您要儲存檔案的位置，然後輸入 demo.csplan 做為計劃檔的名稱。
- 按一下「下一步」。

圖表 13-15
取樣精靈 - 設計變數步驟（階段 1）



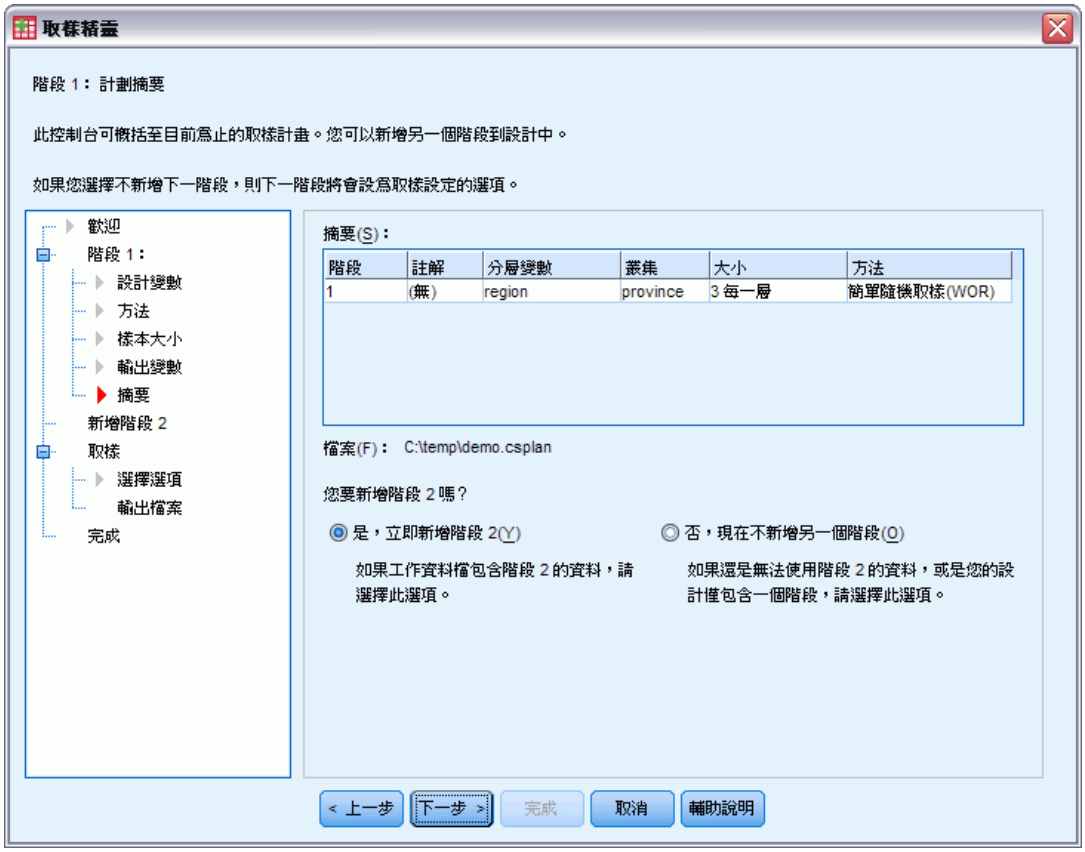
- 選取範圍做為分層變數。
- 選取省做為集群變數。
- 按一下「**下一步**」，並在「取樣方法」步驟按一下「**下一步**」。

  設計結構表示從各範圍抽出獨立樣本。在此階段中，會使用預設方法「簡單隨機取樣」抽出省，做為主要取樣單位。

圖表 13-16
取樣精靈 - 取樣大小步驟（階段 1）

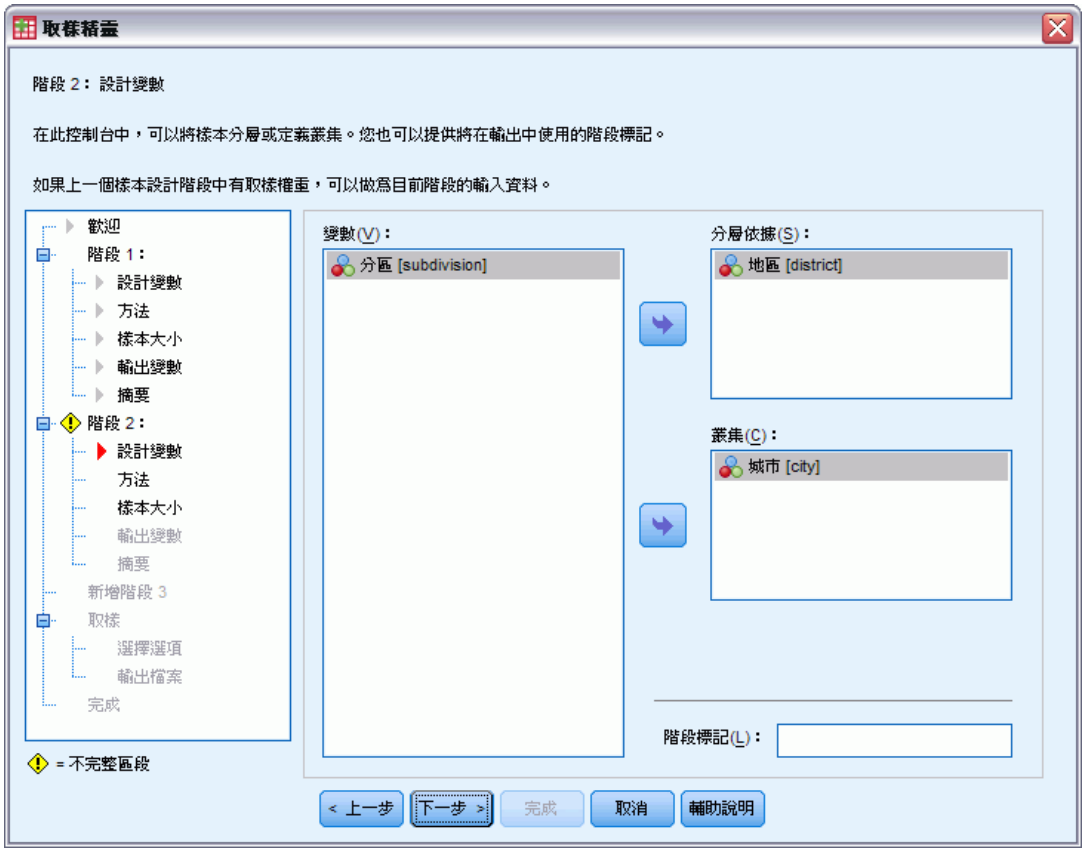
- 從「單位」下拉式清單中選取**個數**。
- 輸入 3 做為要在此階段選取的單位個數數值。
- 按一下「**下一步**」，並在「輸出變數」步驟按一下「**下一步**」。

圖表 13-17
取樣精靈 - 計劃摘要步驟（階段 1）



- ▶ 選取「是，立即新增階段 2」。
- ▶ 按一下「下一步」。

圖表 13-18
取樣精靈 - 設計變數步驟（階段 2）



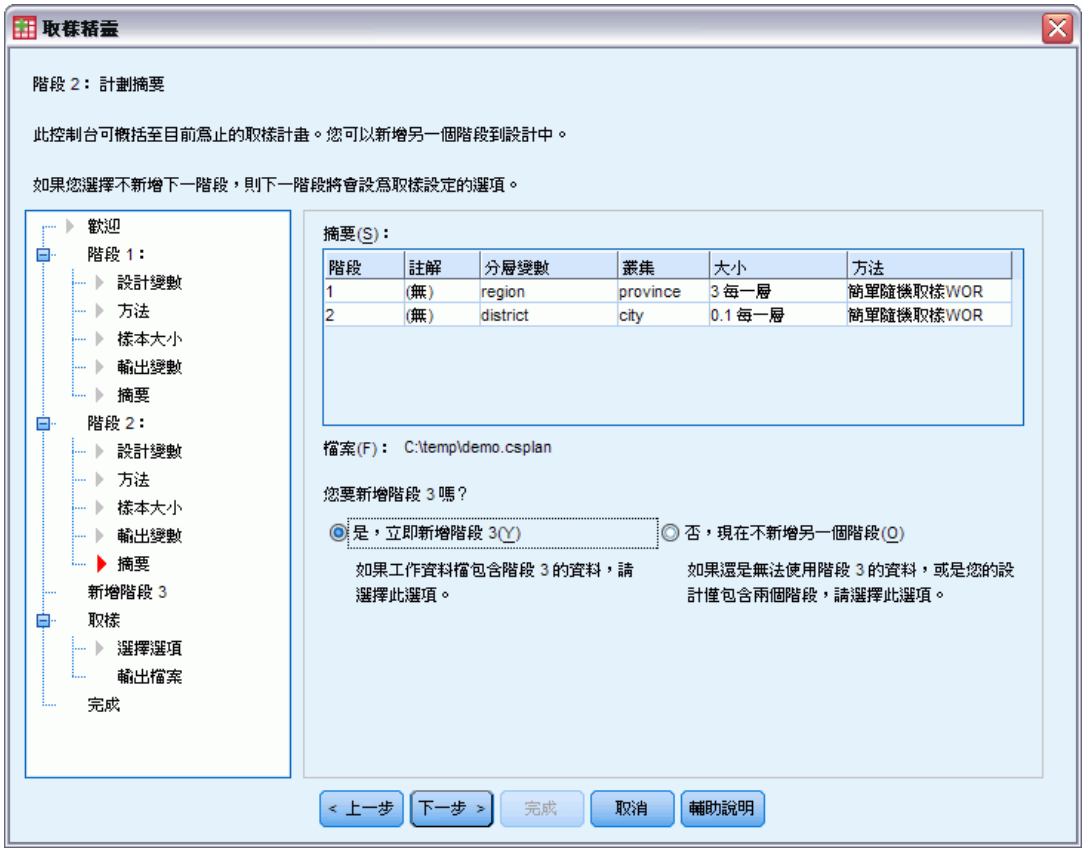
- 選取區做為分層變數。
- 選取市做為集群變數。
- 按一下「**下一步**」，並在「取樣方法」步驟按一下「**下一步**」。

  設計結構表示從各區抽出獨立樣本。在此階段中，會使用預設方法「簡單隨機取樣」抽出市，做為主要取樣單位。

圖表 13-19
取樣精靈 - 取樣大小步驟（階段 2）

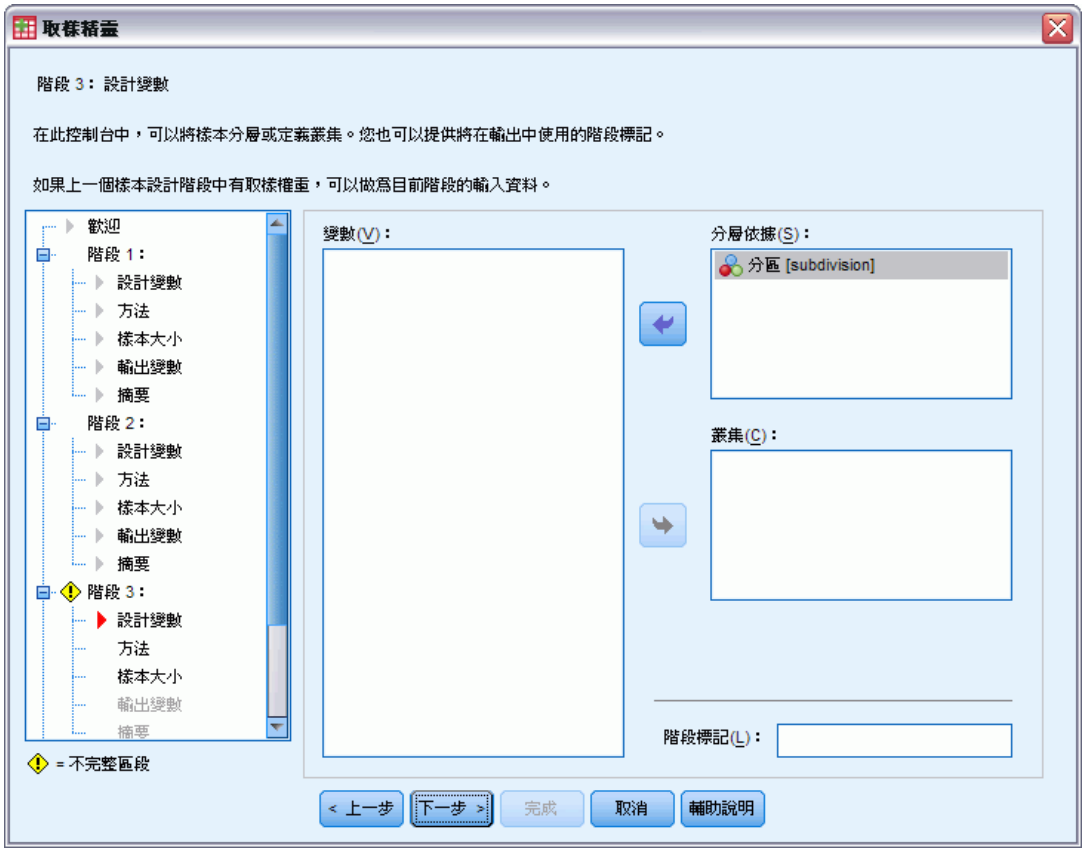
- 從「單位」下拉式清單中選取**財產**。
- 輸入 0.1 做為各層中樣本單元的比例數值。
- 按一下「**下一步**」，並在「輸出變數」步驟按一下「**下一步**」。

圖表 13-20
取樣精靈 - 計劃摘要步驟（階段 2）



- ▶ 選取「是，立即新增階段 3」。
- ▶ 按一下「下一步」。

圖表 13-21
取樣精靈 - 設計變數步驟（階段 3）



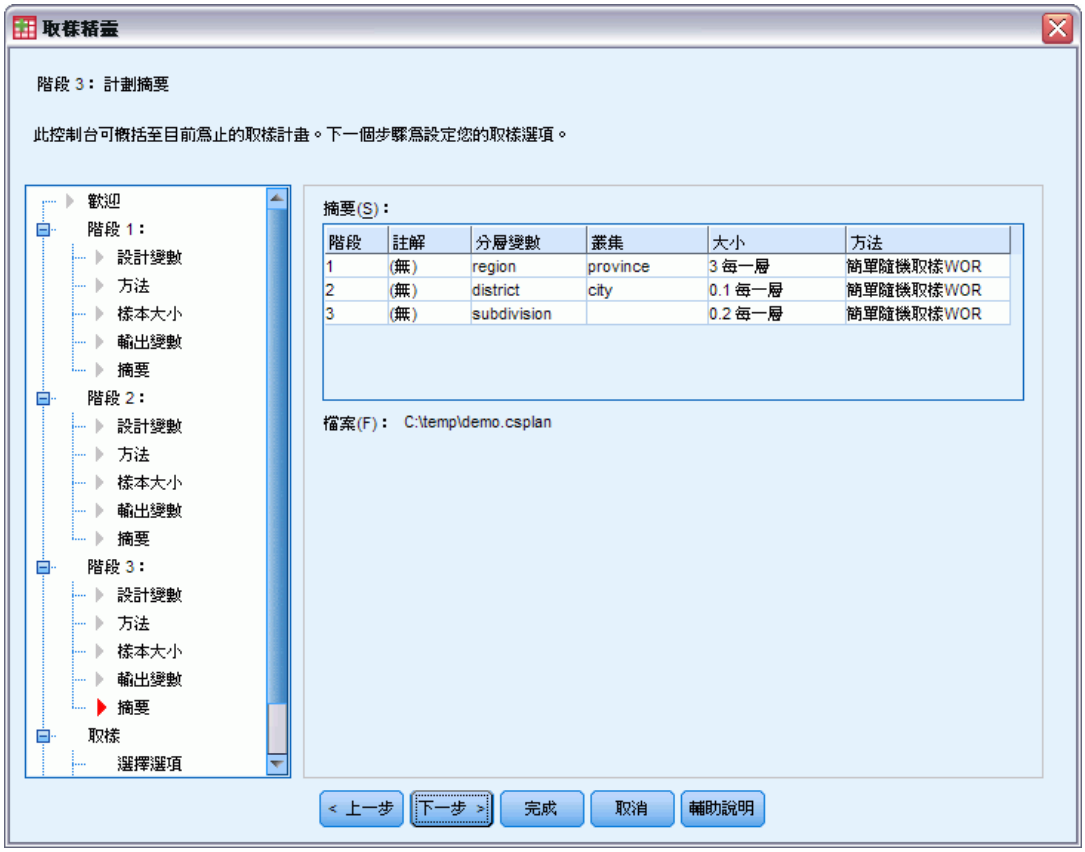
- 選取分區做為分層變數。
- 按一下「**下一步**」，並在「取樣方法」步驟按一下「**下一步**」。

  設計結構表示從各土地分區抽出獨立樣本。在此階段中，會使用預設方法「簡單隨機取樣」抽出家庭單位，做為主要取樣單位。

圖表 13-22
取樣精靈 - 取樣大小步驟（階段 3）

- 從「單位」下拉式清單中選取**財產**。
- 輸入 **0.2** 做為要在此階段選擇的單位比例數值。
- 按一下「**下一步**」，並在「輸出變數」步驟按一下「**下一步**」。

圖表 13-23
取樣精靈 - 計劃摘要步驟（階段 3）



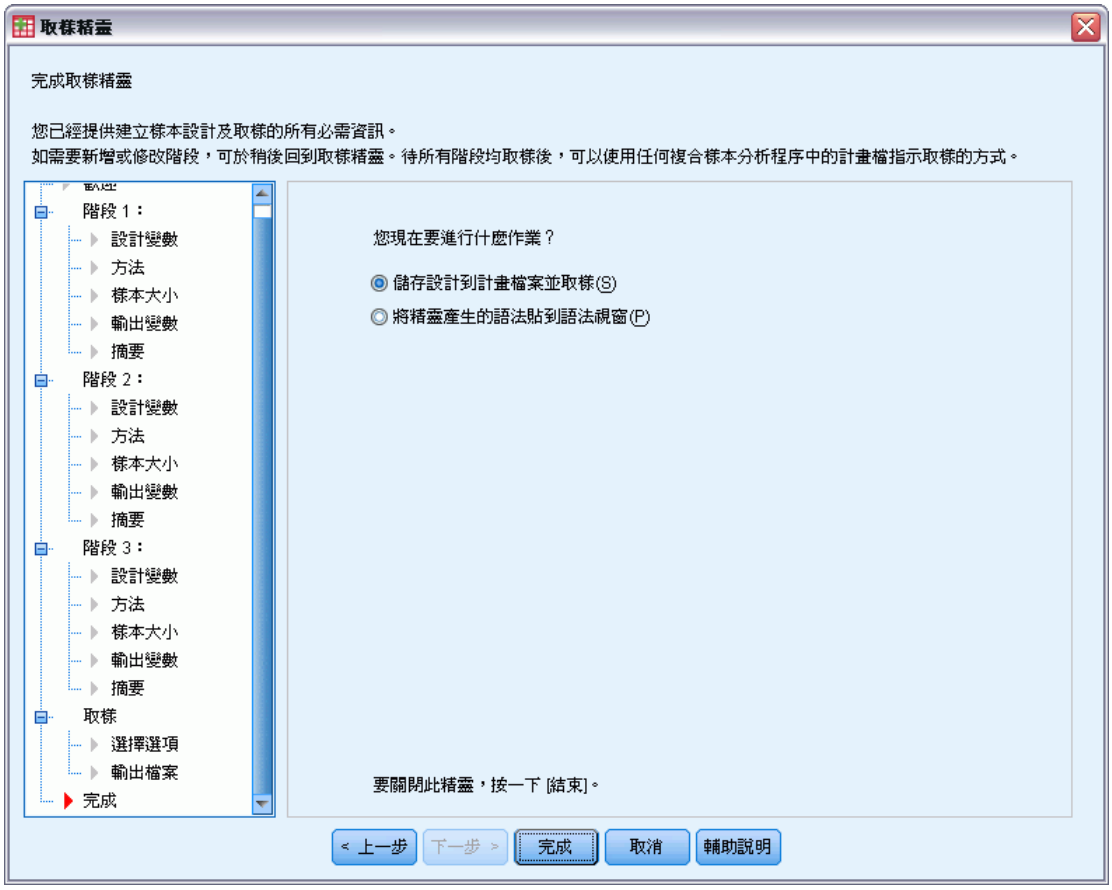
► 檢查取樣設計，再按一下「下一步」。

圖表 13-24
「取樣精靈 - 抽樣 - 選擇選項」步驟

- 選取 「1、2 」作為現在要取樣的階段。
- 在要使用的亂數種子類型選取「**自訂值**」，並輸入 241972 做為數值。

  使用自訂值可讓您確實複製此範例的結果。
- 按一下「**下一步**」，並在「抽樣輸出檔案」步驟按一下「**下一步**」。

圖表 13-25
取樣精靈 - 完成步驟



▶ 按一下「完成」。

這些選擇會產生取樣計劃檔案 demo.csplan，並根據此計畫的前兩階段抽出樣本。

樣本結果

圖表 13-26
具有樣本結果的資料編輯程式

	region	province	district	city	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
295	1	2	10	295	.	.	.	.	.
296	1	2	10	296	.	.	.	.	.
297	1	2	10	297	.	.	.	.	.
298	1	2	10	298	.20	5.00	.10	50.00	50.00
299	1	2	10	299	.	.	.	.	.
300	1	2	10	300	.20	5.00	.10	50.00	50.00
301	1	2	11	301	.	.	.	.	.
302	1	2	11	302	.	.	.	.	.
303	1	2	11	303	.	.	.	.	.
304	1	2	11	304	.	.	.	.	.
305	1	2	11	305	.	.	.	.	.
306	1	2	11	306	.	.	.	.	.
307	1	2	11	307	.20	5.00	.10	50.00	50.00
308	1	2	11	308					

資料檢視 變數檢視

您可在「資料編輯程式」中檢視取樣結果。 會有五個新變數儲存至工作檔中，表示各階段的包含機率和累積樣本權重，加上前兩階段的「最後」取樣權重。

- 將會選出具備這些變數值的市進行取樣。
- 而不會選取具備變數系統遺漏值的市。

對於選取的各城市，此公司已取得分區和家庭單位資訊，並放置在 demo_cs_2.sav 中。使用此檔案及「取樣精靈」取樣此設計的第三階段。

從第二個部分框使用精靈取樣

▶ 若要執行「複合樣本取樣精靈」，請從功能表選擇：
分析(A) > 複合樣本 > 選擇樣本...

圖表 13-27
「取樣精靈 - 歡迎」步驟

- 選取「抽出樣本」，瀏覽至您儲存計劃檔的位置，然後選取您建立的 demo.csplan 計劃檔。
- 按一下「下一步」。

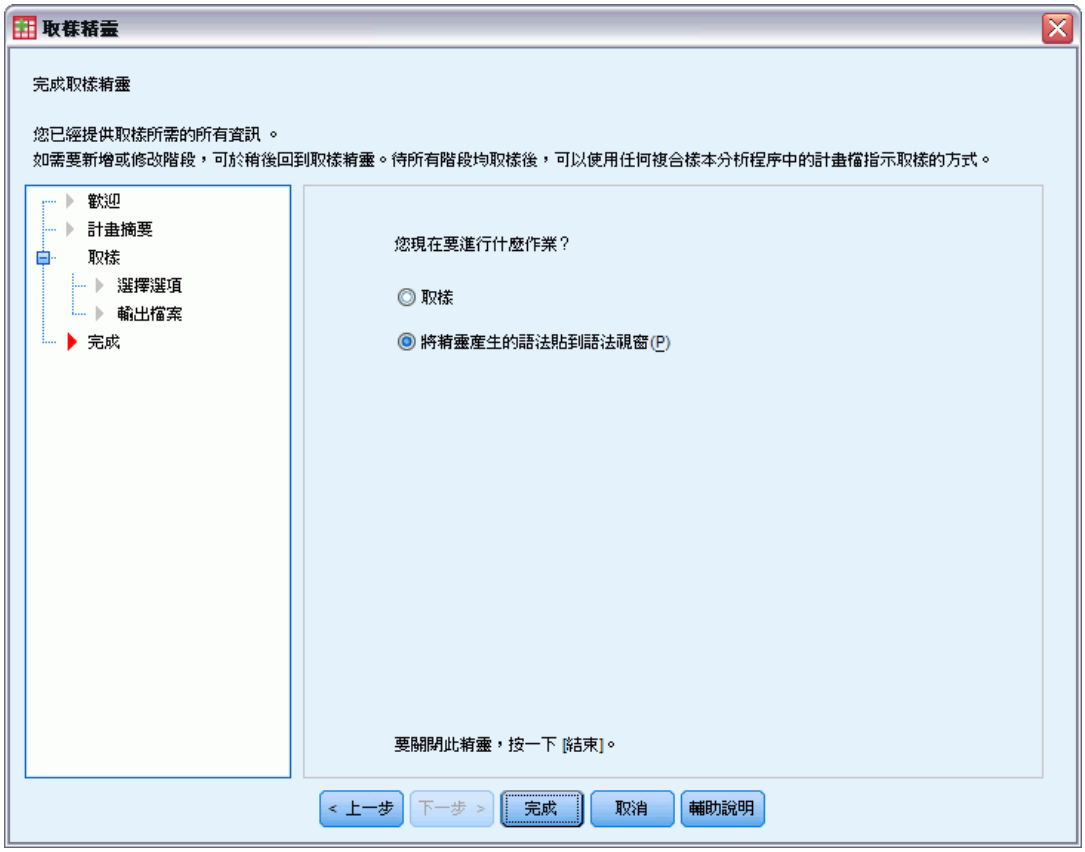
圖表 13-28
取樣精靈 - 計劃摘要步驟（階段 3）

- ▶ 選取「1、2」作為已取樣的階段。
- ▶ 按一下「下一步」。

圖表 13-29
「取樣精靈 - 抽樣 - 選擇選項」步驟

- 在要使用的亂數種子類型選取「**自訂值**」，並輸入 **4231946** 做為數值。
- 按一下「**下一步**」，並在「抽樣輸出檔案」步驟按一下「**下一步**」。

圖表 13-30
取樣精靈 - 完成步驟



- 選取「將精靈所產生的語法貼入語法視窗」。
- 按一下「完成」。

產生下列語法：

```
* 取樣精靈。
CSSELECT
 /PLAN FILE='demo.csplan'
 /CRITERIA STAGES = 3 SEED = 4231946
 /CLASSMISSING EXCLUDE
 /DATA RENAMEVARS
 /PRINT SELECTION.
```

在此狀況下列印取樣摘要會產生繁瑣的表格，在「輸出瀏覽器」中造成問題。若要關閉取樣摘要的顯示，請將「PRINT」次指令中的「SELECTION 」以「CPS」取代。接著在語法視窗中執行該語法。

這些選擇將依據 demo.csplan 取樣計劃的第三階段抽樣。

樣本結果

圖表 13-31
具有樣本結果的資料編輯程式

	city	subdivision	unit	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	InclusionProbability_3_	SampleWeightCumulative_3_	SampleWeight_Final_
14	190	946	94514	0.20	5.00	0.10	50.00	.	.	.
15	190	946	94515	0.20	5.00	0.10	50.00	.	.	.
16	190	946	94516	0.20	5.00	0.10	50.00	0.20	244.44	244.44
17	190	946	94517	0.20	5.00	0.10	50.00	.	.	.
18	190	946	94518	0.20	5.00	0.10	50.00	.	.	.
19	190	946	94519	0.20	5.00	0.10	50.00	.	.	.
20	190	946	94520	0.20	5.00	0.10	50.00	.	.	.
21	190	946	94521	0.20	5.00	0.10	50.00	.	.	.
22	190	946	94522	0.20	5.00	0.10	50.00	.	.	.
23	190	946	94523	0.20	5.00	0.10	50.00	.	.	.
24	190	946	94524	0.20	5.00	0.10	50.00	0.20	244.44	244.44
25	190	946	94525	0.20	5.00	0.10	50.00	.	.	.
26	190	946	94526	0.20	5.00	0.10	50.00	.	.	.
27	190	946	94527	0.20	5.00	0.10	50.00	.	.	.
28	190	946	94528	0.20	5.00	0.10	50.00	.	.	.
29	190	946	94529	0.20	5.00	0.10	50.00	0.20	244.44	244.44
30	190	946	94530	0.20	5.00	0.10	50.00	.	.	.

資料檢視 變數檢視

您可在「資料編輯程式」中檢視取樣結果。會有三個新變數儲存至工作檔中，表示第三階段的包含機率和累積樣本權重，加上最後的取樣權重。在取樣前兩步驟時，這些新的權重會考慮到所計算出的權重。

- 將會選出具備這些變數值的單位進行取樣。
- 而不會選取具備變數系統遺漏值的單位。

此公司現可使用其資源，以取得樣本中所選家庭單位的調查資料。只要收集到這些調查資料，您就可使用 demo.csplan，以「複合樣本」分析程序處理這些樣本。

以「機率與單位大小成比例」(PPS) 的方式取樣

議會代表們在考慮將一項法案交付立法之前，會想知道是否有民眾支持該法案，及該項支持與投票人口如何相關。民調機構根據複雜取樣設計，設計出一份問卷並開始進行訪查。

登記選民清單收集於 poll_cs.sav 中。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 使用「複合樣本取樣精靈」選取樣本供進一步分析。

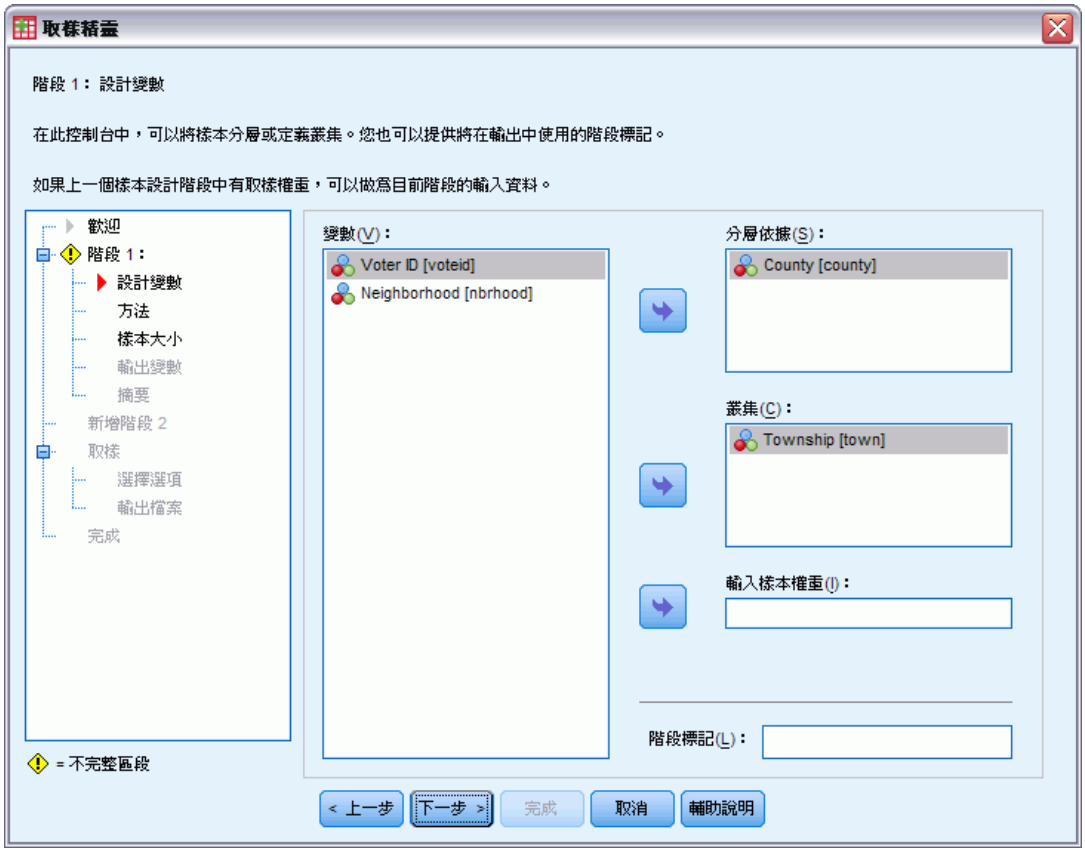
使用精靈

▶ 若要執行「複合樣本取樣精靈」，請從功能表選擇：
分析(A) > 複合樣本 > 選擇樣本...

圖表 13-32
「取樣精靈 - 歡迎」步驟

- 選取「設計樣本」，瀏覽至您要儲存檔案的位置，然後輸入 poll.csplan 做為計劃檔的名稱。
- 按一下「下一步」。

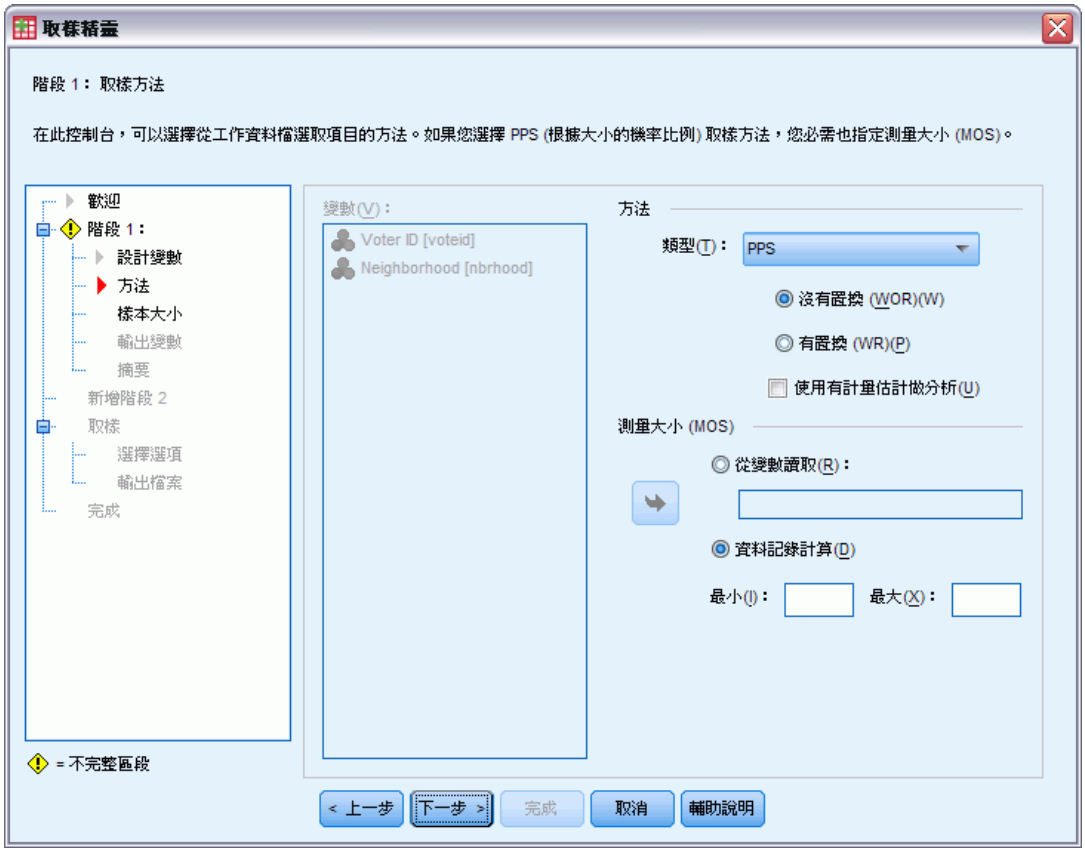
圖表 13-33
取樣精靈 - 設計變數步驟（階段 1）



- 選取郡做為分層變數。
- 選取鎮做為集群變數。
- 按一下「下一步」。

設計結構表示從各郡抽出獨立樣本。在這個階段，抽出鎮來做為主要取樣單位。

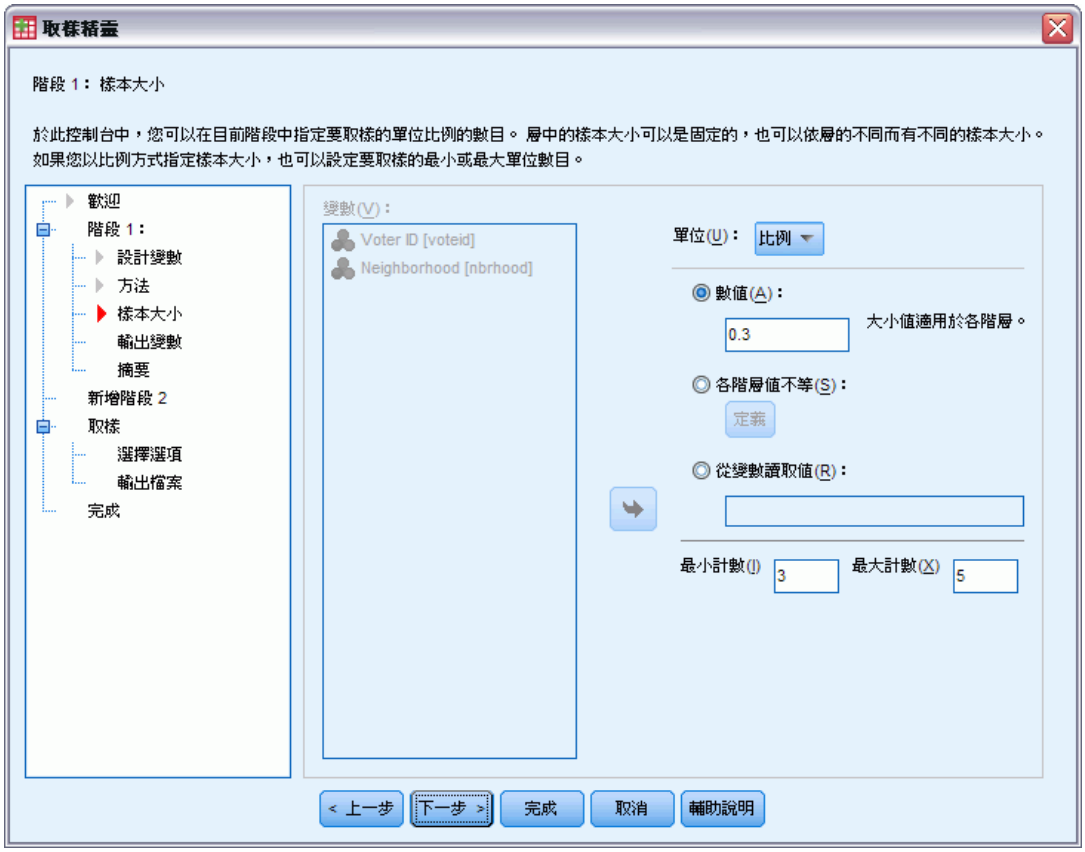
圖表 13-34
取樣精靈 - 取樣方法步驟（階段 1）



- 選取「PPS」作為取樣方法。
- 選取「資料記錄計數」作為大小測量法。
- 按一下「下一步」。

在每一個郡中，以不放回取樣、機率與每一個鎮的數目成比例的方式抽取出鎮。使用 PPS 方法會產生這些鎮的聯合取樣機率；您可在「輸出檔案」步驟中指定這些值的存放處。

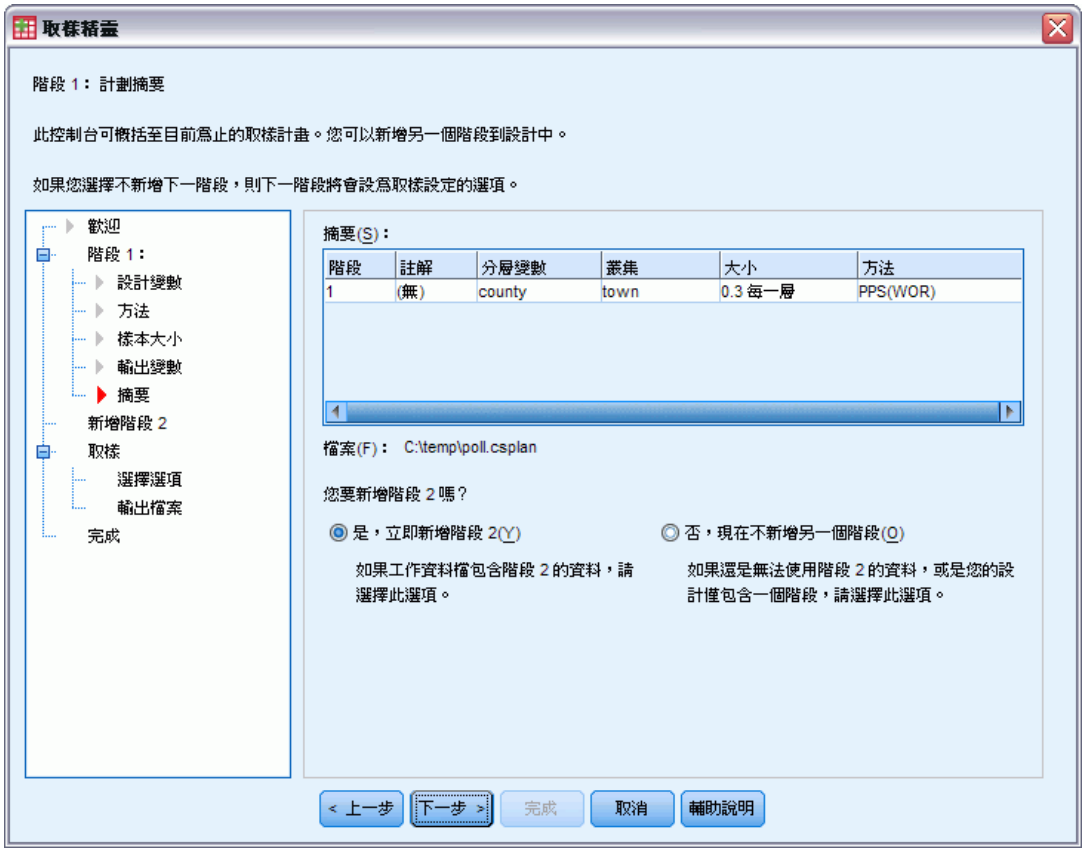
圖表 13-35
取樣精靈 - 取樣大小步驟（階段 1）



- 從「單位」下拉式清單中選取**財產**。
- 輸入 0.3 作為要在此階段選取的鎮在每一個郡中的比例數值。

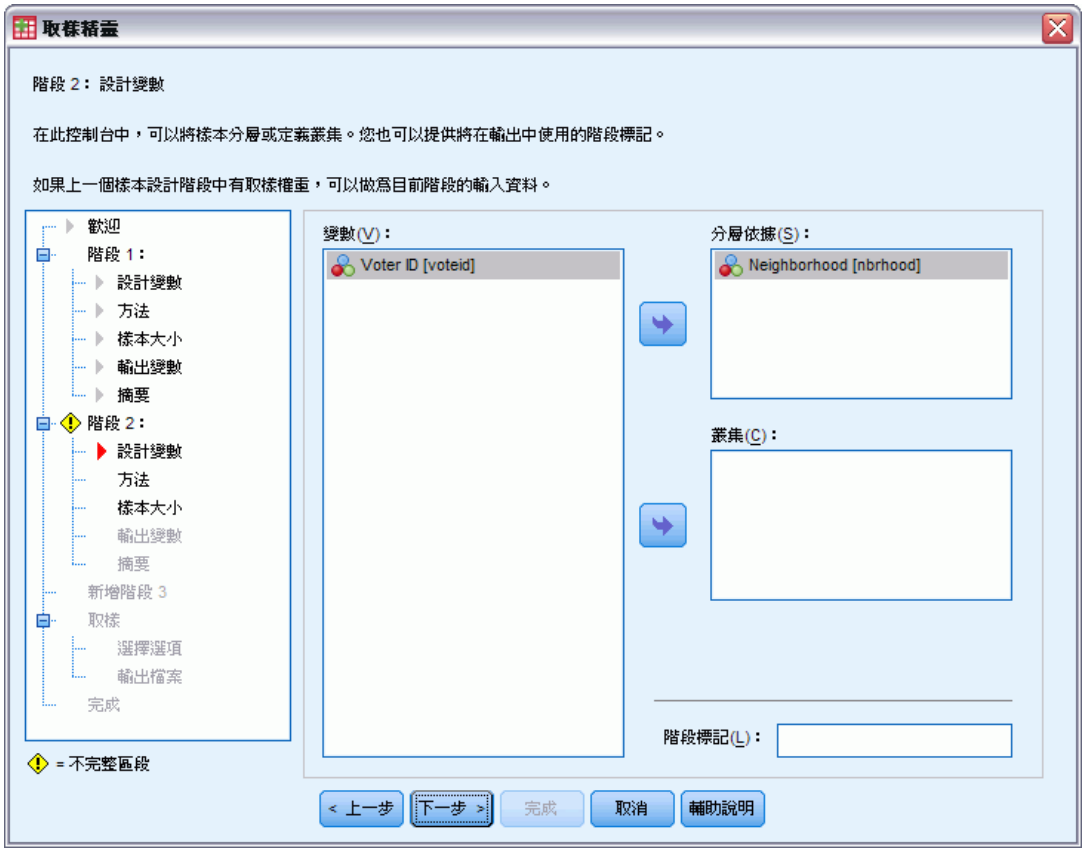
  西部郡的議員們指出，在他們的郡中，鎮的數目比其它的郡少。為了確保有足夠的代表性，他們想要在每一個郡中，建立至少 3 個鎮的樣本。
- 鍵入「3」作為要選取鎮的最小數目，和「5」作為最大數目。
- 按一下「**下一步**」，並在「輸出變數」步驟按一下「**下一步**」。

圖表 13-36
取樣精靈 - 計劃摘要步驟（階段 1）



- ▶ 選取「是，立即新增階段 2」。
- ▶ 按一下「下一步」。

圖表 13-37
取樣精靈 - 設計變數步驟（階段 2）



- 選取鄰近做為分層變數。
- 按一下「**下一步**」，並在「取樣方法」步驟按一下「**下一步**」。

  此設計結構表示階段 1 中抽出的鎮之鄰近各範圍所抽出的獨立樣本。在此階段中，會使用「簡單隨機取樣」抽出選民做為主要取樣單位。

圖表 13-38
取樣精靈 - 取樣大小步驟（階段 2）

- 從「單位」下拉式清單中選取**財產**。
- 輸入 **0.2** 做為各層中樣本單元的比例數值。
- 按一下「**下一步**」，並在「輸出變數」步驟按一下「**下一步**」。

圖表 13-39
取樣精靈 - 計劃摘要步驟（階段 2）

► 檢查取樣設計，再按一下「下一步」。

圖表 13-40
「取樣精靈 - 抽樣 - 選擇選項」步驟

- 在要使用的亂數種子類型選取「**自訂值**」，並輸入 **592004** 做為數值。

  使用自訂值可讓您確實複製此範例的結果。

- 按一下「**下一步**」。

圖表 13-41
「取樣精靈 - 抽樣 - 選擇選項」步驟

- 選擇將此樣本存到一個新的資料集，再輸入 poll_cs_sample 做為資料集的名稱。
- 瀏覽至您要儲存聯合機率的位置，然後輸入 poll_jointprob.sav 做為聯合機率檔案的名稱。
- 按一下「下一步」。

圖表 13-42
取樣精靈 - 完成步驟

▶ 按一下「**完成**」。

這些選擇產生取樣計劃檔 poll.csplan，並依據該計劃抽取樣本將樣本結果存入新的資料集 poll_cs_sample，再將聯合機率檔存入外部檔案 poll_jointprob.sav。

計畫摘要

圖表 13-43
計劃摘要

			第 1 階段	第 2 階段
設計變數	階層化	1	County	Neighborhood
	叢集	1	Township	
樣本資訊	階層化	選擇方法	Sample PPS_WOR	Sample SIMPLE_WOR
		測量大小	取自資料	
		取樣的單位比例	.3	.2
		取樣的最小單位數	3	
		取樣的最大單位數	5	
	建立或修改過的變數	階段性包括雙尾的檢定 (選擇) 機率	InclusionProbability_1_	InclusionProbability_2_
		階段性的累積取樣權重	SampleWeight Cumulative_1_	SampleWeight Cumulative_2_
分析資訊	階層化	估計量假設	沒有置換的不相等機率取樣 (使用包括雙尾的合併檢定機率)	沒有置換的相等機率取樣
		包括雙尾的檢定機率	取自變數 InclusionProbability_1_	取自變數 InclusionProbability_2_
設計變數	階層化			

規劃檔案\： C:\Samples\en\poll.csplan
加權變數\： SampleWeight_Final_

摘要表會檢閱您的取樣計劃，且在確認該計劃是否能夠表達您目的時非常實用。

取樣摘要

圖表 13-44
階段摘要

	取樣的單位數		取樣的單位比例	
County	要求的	眞實	要求的	眞實
Eastern	4	4	30.0%	30.8%
Central	4	4	30.0%	30.8%
Western	3	3	30.0%	50.0%
Northern	5	5	30.0%	33.3%
Southern	3	3	30.0%	50.0%

規劃檔案\： C:\Samples\en\poll.csplan

此摘要表會檢閱取樣的第一階段，且在檢查取樣是否按照計劃進行時非常實用。請試著回想，除了在西部和南部的郡以外，您要求以郡抽取 30% 的鎮樣本；實際取樣的比例接近 30%。這是因為這些郡各自只有 6 個鎮，而您又指定每個郡至少要選取三個鎮。

圖表 13-45
階段摘要

County	Township	Neighborhood	取樣的單位數		取樣的單位比例	
			要求的	眞實	要求的	眞實
Eastern	9	1	49	49	20.0%	19.9%
		2	143	143	20.0%	20.0%
		3	113	113	20.0%	20.0%
		4	77	77	20.0%	20.0%
		5	139	139	20.0%	20.0%
		6	120	120	20.0%	20.0%
	10	1	149	149	20.0%	20.1%
		2	117	117	20.0%	20.0%
		3	116	116	20.0%	20.0%
		4	69	69	20.0%	19.9%
	11	1	65	65	20.0%	19.9%
		2	72	72	20.0%	19.9%
		3	109	109	20.0%	20.0%
		4	140	140	20.0%	20.0%
		5	42	42	20.0%	19.8%
		6	142	142	20.0%	20.0%
	12	1	145	145	20.0%	20.1%
		2	69	69	20.0%	20.1%
		3	98	98	20.0%	20.1%
		4	134	134	20.0%	20.0%
		5	114	114	20.0%	20.0%
		6	137	137	20.0%	19.9%
Central	2	1	119	119	20.0%	20.1%
		2	153	153	20.0%	19.9%
		3	101	101	20.0%	20.0%
		4	52	52	20.0%	19.8%
		5	144	144	20.0%	20.0%

規劃檔案\： C:\Samples\en\poll.csplan

此摘要表（最上面部分顯示於此）可檢閱取樣的第二階段。在檢查取樣是否依照計劃進行時亦十分有用。依照要求，約取樣出第一階段所取樣的各鎮各鄰近範圍之 20% 的選民。

樣本結果

圖表 13-46
具有樣本結果的資料編輯程式

	voteid	nbrhood	town	county	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
376	368	4	9	1	.44	2.26	.20	11.28	11.28
377	369	4	9	1	.44	2.26	.20	11.28	11.28
378	374	4	9	1	.44	2.26	.20	11.28	11.28
379	376	4	9	1	.44	2.26	.20	11.28	11.28
380	379	4	9	1	.44	2.26	.20	11.28	11.28
381	380	4	9	1	.44	2.26	.20	11.28	11.28
382	382	4	9	1	.44	2.26	.20	11.28	11.28
383	13	5	9	1	.44	2.26	.20	11.26	11.26
384	18	5	9	1	.44	2.26	.20	11.26	11.26
385	23	5	9	1	.44	2.26	.20	11.26	11.26
386	38	5	9	1	.44	2.26	.20	11.26	11.26
387	39	5	9	1	.44	2.26	.20	11.26	11.26
388	40	5	9	1	.44	2.26	.20	11.26	11.26
389	41	5	9	1	.44	2.26	.20	11.26	11.26
390	43	5	9	1	.44	2.26	.20	11.26	11.26

資料檢視 變數檢視

您可在新建立的資料集中檢視取樣結果。 會有五個新變數儲存至工作檔中，表示各階段的包含機率和累積樣本權重，加上最後的取樣權重。未被選入樣本的選民，將被排除在資料集外。

在相同鄰近範圍的選民其最終取樣權重相同，因為他們是在鄰近範圍依據簡單隨機抽樣方法所選取的。不過，他們在同一鎮跨鄰近範圍是不相同的，因為取樣比例在所有的鄰近範圍內並不是剛好 20%。

圖表 13-47
具有樣本結果的資料編輯程式

	voteid	nbrhood	town	county	InclusionProbability_1_	SampleWeightCumulative_1_	InclusionProbability_2_	SampleWeightCumulative_2_	SampleWeight_Final_
635	577	6	9	1	.44	2.26	.20	11.30	11.30
636	578	6	9	1	.44	2.26	.20	11.30	11.30
637	582	6	9	1	.44	2.26	.20	11.30	11.30
638	590	6	9	1	.44	2.26	.20	11.30	11.30
639	594	6	9	1	.44	2.26	.20	11.30	11.30
640	597	6	9	1	.44	2.26	.20	11.30	11.30
641	600	6	9	1	.44	2.26	.20	11.30	11.30
642	4	1	10	1	.31	3.21	.20	16.00	16.00
643	5	1	10	1	.31	3.21	.20	16.00	16.00
644	9	1	10	1	.31	3.21	.20	16.00	16.00
645	10	1	10	1	.31	3.21	.20	16.00	16.00
646	12	1	10	1	.31	3.21	.20	16.00	16.00
647	16	1	10	1	.31	3.21	.20	16.00	16.00
648	17	1	10	1	.31	3.21	.20	16.00	16.00
649	19	1	10	1	.31	3.21	.20	16.00	16.00

資料檢視 變數檢視

與第二階段中的選民不同的是，在相同的郡中，鎮的第一階段取樣權重並不相同，因為它們是採機率與單位大小成比例的方式選取的。

圖表 13-48
聯合機率檔案

	county	town	Unit_No_	Joint_Prob_1_	Joint_Prob_2_	Joint_Prob_3_	Joint_Prob_4_	Joint_Prob_5_
1	1	10	1	.31	.10	.11	.12	.
2	1	11	2	.10	.39	.15	.16	.
3	1	9	3	.11	.15	.44	.21	.
4	1	12	4	.12	.16	.21	.48	.
5	2	12	1	.22	.04	.07	.08	.
6	2	6	2	.04	.23	.07	.08	.
7	2	7	3	.07	.07	.41	.19	.
8	2	2	4	.08	.08	.19	.45	.
9	3	5	1	.58	.31	.32	.	.
10	3	3	2	.31	.61	.36	.	.
11	3	4	3	.32	.36	.63	.	.
12	4	14	1	.26	.06	.06	.07	.09
13	4	8	2	.06	.29	.07	.08	.10
14	4	4	3	.06	.07	.29	.08	.10
15	4	2	4	.07	.08	.08	.33	.12
16	4	13	5	.09	.10	.10	.12	.43
17	5	3	1	.74	.25	.27	.	.
18	5	6	2	.25	.41	.13	.	.
19	5	4	3	.27	.13	.43		

資料檢視 變數檢視

檔案 poll_jointprob.sav 包含郡內選取鄉鎮的第一階段聯合機率。郡是第一階段分層變數，鎮是集群變數。這些變數的組合可專門識別所有第一階段 PSU。Unit_No_ 可標記每一層內的 PSU，並用於比對 Joint_Prob_1_、Joint_Prob_2_、Joint_Prob_3_、Joint_Prob_4_ 和 Joint_Prob_5_。前兩層均有 4 個 4 PSU；因此這些層的聯合包含機率矩陣為 4×4，而這些列的 Joint_Prob_5_ 欄會保留空白。同樣地，層 3 和 5 有 3×3 聯合包含機率矩陣，而層 4 有一個 5×5 聯合包含機率矩陣。

藉由細讀聯合包含機率矩陣，可發現聯合機率檔案的需求。當取樣方法不是 PPS WOR 方法時，PSU 的選擇與其他 PSU 的選擇沒有關係，且其聯合包含機率只是其包含機率的乘積。反之，郡 1 的鎮 9 和鎮 10 的聯合包含機率約為 0.11（請參閱 Joint_Prob_3_ 的第一個觀察值或Joint_Prob_1_ 的第三個觀察值），或小於其個別包含機率的乘積（Joint_Prob_1_ 的第一個觀察值和 Joint_Prob_3_ 的第三個觀察值為 0.31×0.44=0.1364）。

民意測驗專家現在將對選取的樣本進行訪問。一旦結果出來，您可運用「複合樣本」分析程序來處理樣本，使用取樣計劃 poll.csplan 提供取樣規格，及 poll_jointprob.sav 提供所需的聯合包含機率。

相關程序

「複合樣本取樣精靈」程序在建立取樣計劃檔及抽樣時十分有用。

- 如需在您不具備取樣計劃檔存取權的情況下備妥要分析的樣本，請使用分析準備精靈。

章節
14

複合樣本分析準備精靈

「分析準備精靈」會引導您進行一連串步驟，建立或修改與多種「複合樣本」分析程序共同運作的分析計劃。這在您沒有用於抽樣的取樣計劃檔案存取權時最為實用。

使用「複合樣本分析準備精靈」以備妥 NHIS 公用資料

「國民健康訪問調查 (NHIS)」為美國民間人口的一大型民眾調查。其以具全國代表性的家庭為樣本，面對面的完成訪問。而取得各家庭中成員的人口統計學資訊及健康行為、健康狀態方面等觀察報告。

在 nhis2000_subset.sav 檔中收集了 2000 年調查的子集。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.使用「複合樣本分析準備精靈」建立此資料檔的分析計劃，以使用「複合樣本」分析程序處理。

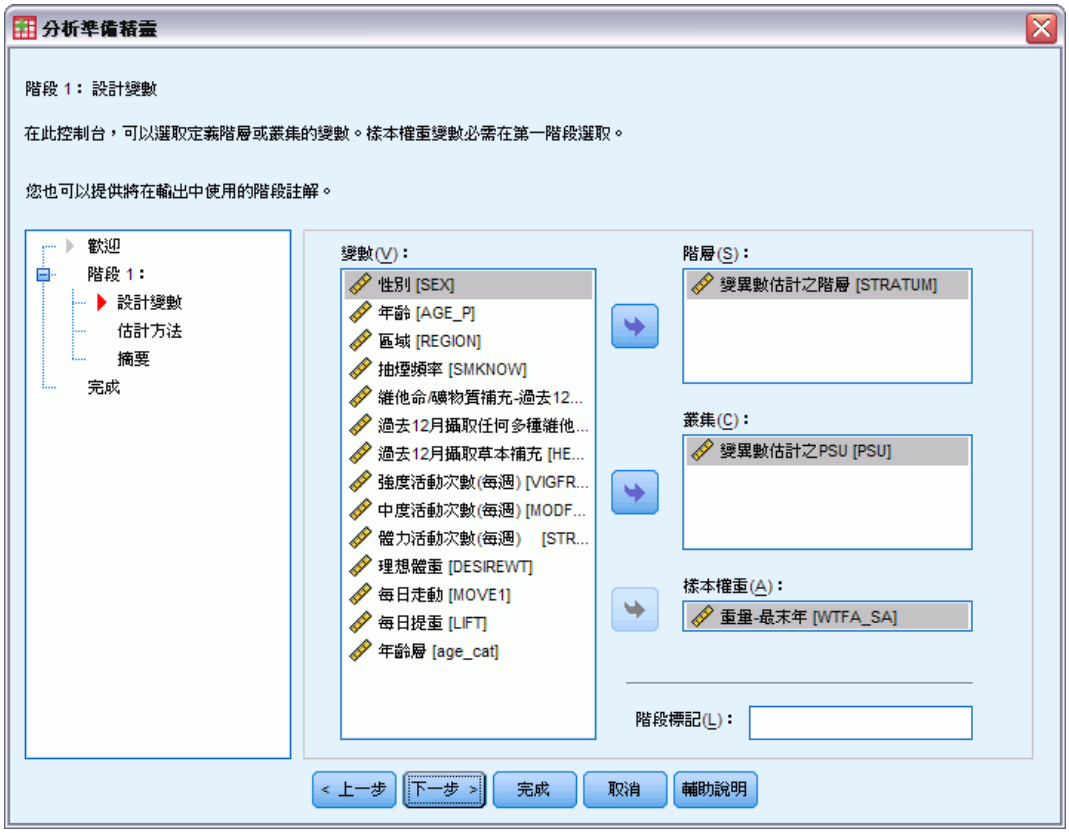
使用精靈

- 若要準備用於「複合樣本分析準備精靈」的樣本，請從功能表中選擇：
  分析(A) > 複合樣本 > 準備分析...

圖表 14-1
分析準備精靈，歡迎步驟

- 瀏覽至您要儲存計劃檔的位置，然後輸入 nhis2000_subset.csaplan 做為分析計劃檔的名稱。
- 按一下「下一步」。

圖表 14-2
分析準備精靈，設計變數步驟（階段 1）



使用複雜多階段取樣獲得資料。然而，在我們給使用者資料中，原始的 NHIS 設計變數已轉換為設計及加權變數的簡化集合，其結果近似於原始設計結構的結果。

- 選取「變異數估計層」做為層變數。
- 選取「變異數估計 PSU」 做為集群變數。
- 選取「權重 - 最後年度」做為樣本加權變數。
- 按一下「**完成**」。

摘要

圖表 14-3
摘要

摘要

			階段 1
設計變數	分層	1	變異數估計層
	集群	1	變異數估計 PSU
分析資訊	估計式假設		取樣後放回[a]

計劃檔案：c:\nhis2000_subset.csaplan
加權變數：加權 - 最後年度
a. 階段 1 中需要的估計式假設取樣後放回。將不會使用任何後續階段於估計。

摘要表會檢閱您的分析計劃。計劃由一具備分層變數之設計的階段及一集群變數所組成。計劃使用取樣並放回 (WR) 估計，並儲存至 c:\nhis2000_subset.csaplan 中。您現可使用此計劃檔，以「複合樣本」分析程序處理 nhis2000_subset.sav。

當資料檔案中無取樣權重時準備分析

某放款員根據複雜設計收集了客戶記錄；然而，檔案中未包含取樣權重。本資訊存放在 bankloan_cs_noweights.sav 中。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 該放款員希望從她所知道的取樣設計開始，使用「複合樣本分析準備精靈」建立此資料檔的分析計劃，以使用「複合樣本」分析程序處理。

該放款員知道記錄是由兩階段選出的，第一階段從 100 家銀行分行以相同機率與選取後不放回的方式選出 15 家。第二階段以相同機率與選取後不放回的方式，從各銀行中選出一百位客戶，資料檔案中並有各銀行客戶的人數資訊。第一步驟是要建立分析計劃，計算階段性包含機率，以及最終取樣權重。

計算包含機率以及取樣權重

- 若要計算第一階段的包含機率，請從功能表中選擇：
  轉換(T) > 計算變數(C)...

圖表 14-4
計算變數對話方塊

第一階段以選取後不放回的方式從 100 家銀行分行中選出 15 家，因此已知銀行被選取的機率為 15/100 = 0.15。

- 輸入 inclprob_s1 做為目標變數。
- 輸入 0.15 作為數值運算式。
- 按一下「確定」。

圖表 14-5
計算變數對話方塊

第二階段從各分行中選出 100 位客戶；因此階段 2 已知銀行中已知客戶被選取的包含機率為 100/銀行客戶數。

- 叫回「計算變數」對話方塊。
- 輸入 inclprob_s2 做為目標變數。
- 輸入 100/ncust 做為數值運算式。
- 按一下「確定」。

圖表 14-6
計算變數對話方塊

現在您具有各階段的包含機率，就可輕鬆算出最終取樣權重。

- 叫回「計算變數」對話方塊。
- 輸入 finalweight 做為目標變數。
- 輸入 1/(inclprob_s1 * inclprob_s2) 做為數值運算式。
- 按一下「確定」。

  現在您可建立分析計劃。

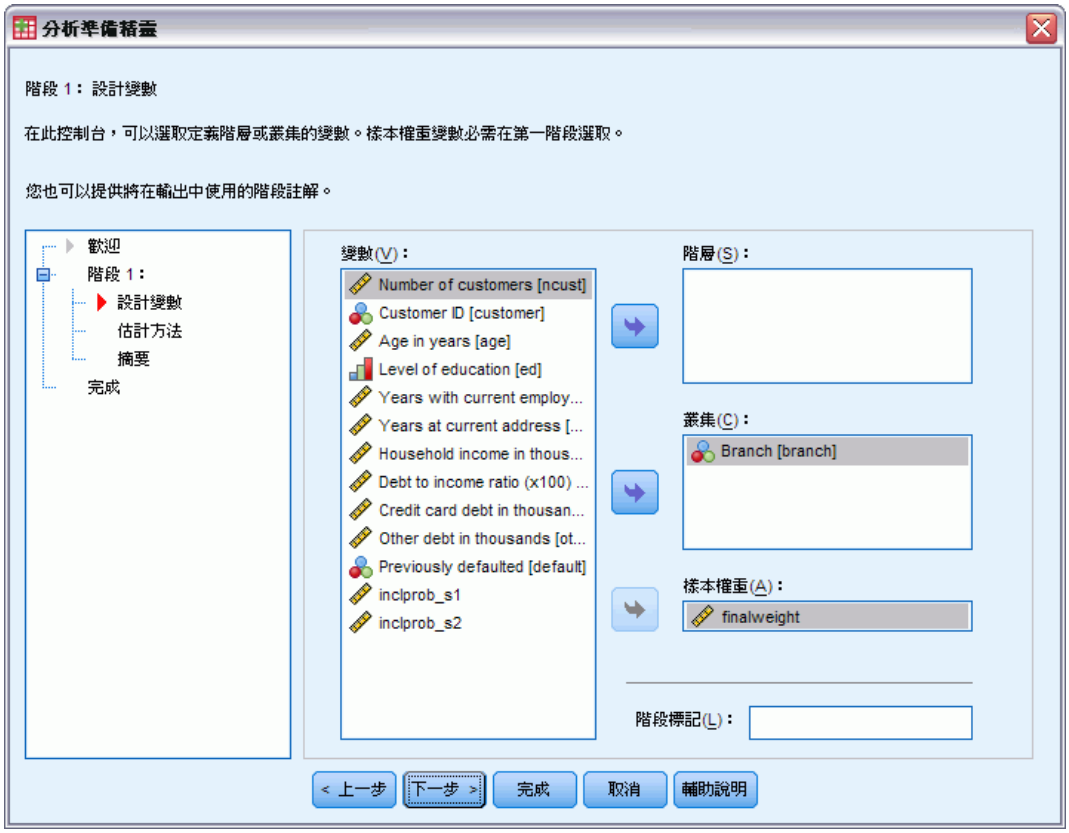
使用精靈

- 若要準備用於「複合樣本分析準備精靈」的樣本，請從功能表中選擇：
  分析(A) > 複合樣本 > 準備分析...

圖表 14-7
分析準備精靈，歡迎步驟

- 瀏覽至您要儲存計劃檔的位置，然後輸入 bankloan.csaplan 做為分析計劃檔的名稱。
- 按一下「下一步」。

圖表 14-8
分析準備精靈，設計變數步驟（階段 1）

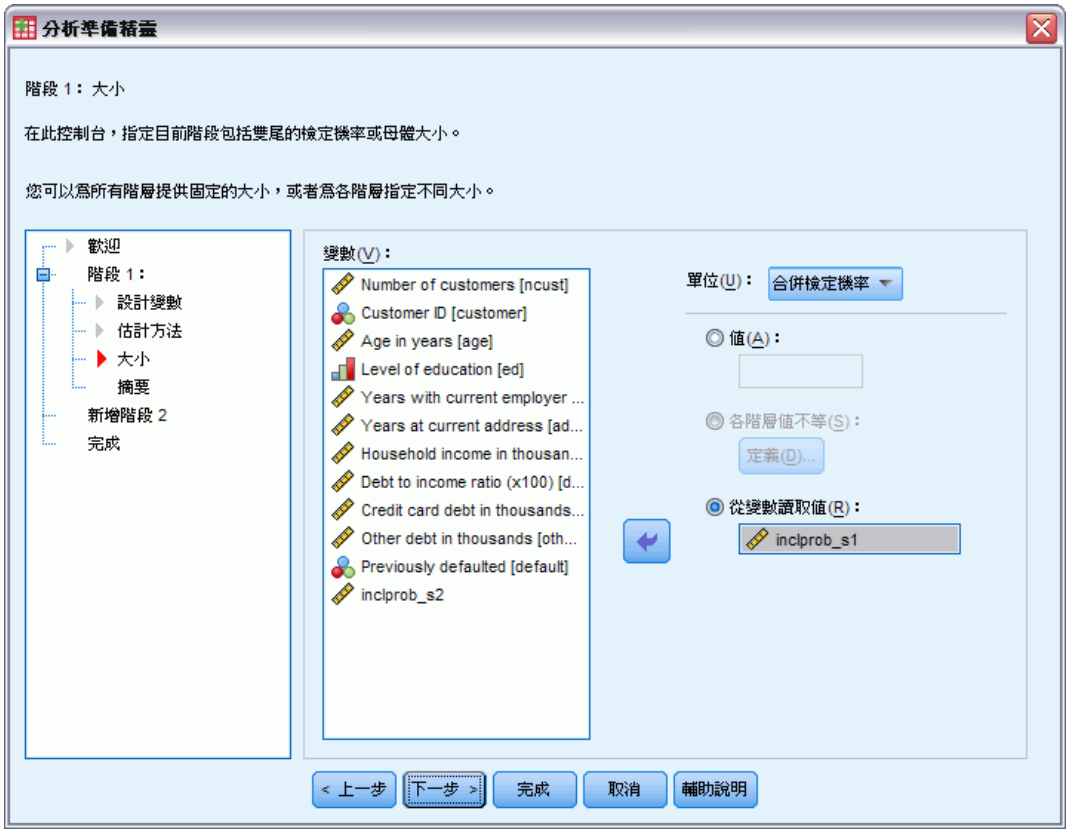


- 選取「分行」做為集群變數。
- 選取「finalweight」做為樣本權重變數。
- 按一下「**下一步**」。

圖表 14-9
分析準備精靈，估計方法步驟（階段 1）

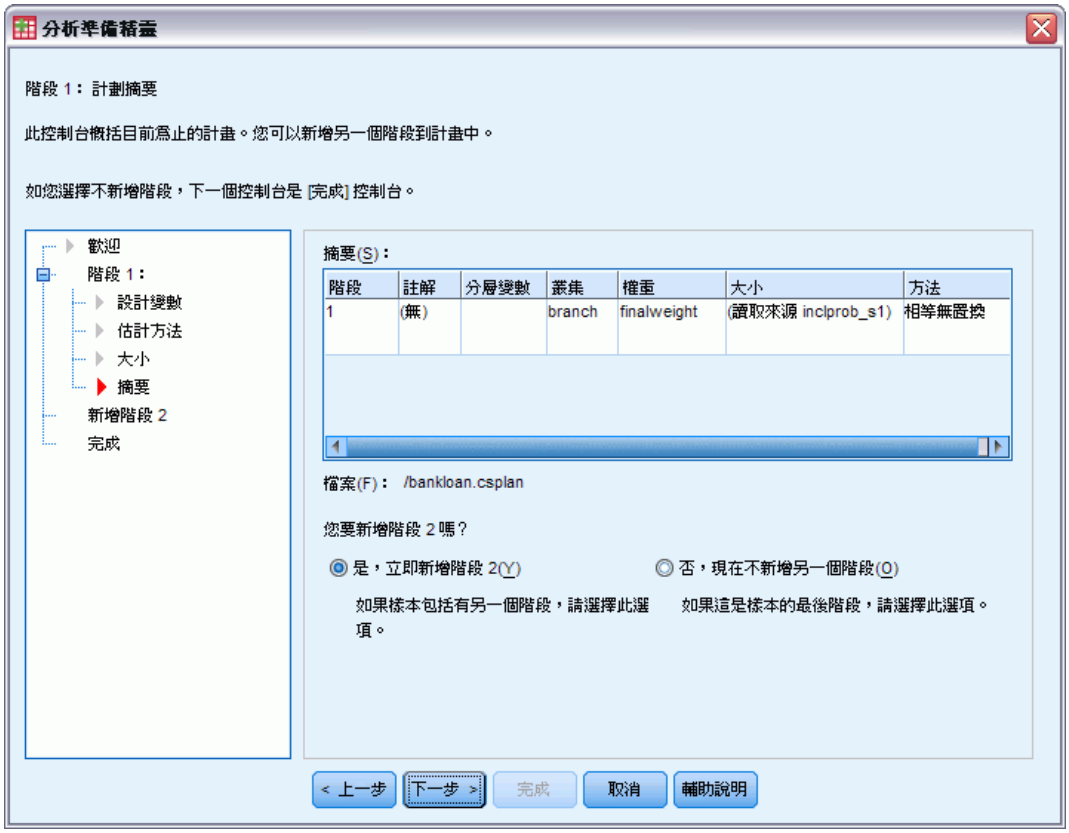
- 選取「相等 WOR」 做為評估方法的第一步驟。
- 按一下「下一步」。

圖表 14-10
分析準備精靈，大小步驟（階段 1）



- 選取「從變數讀取數值」並選取「inclprob_s1」做為含有第一階段包含機率的變數。
- 按一下「下一步」。

圖表 14-11
分析準備精靈，計畫摘要步驟（階段 1）

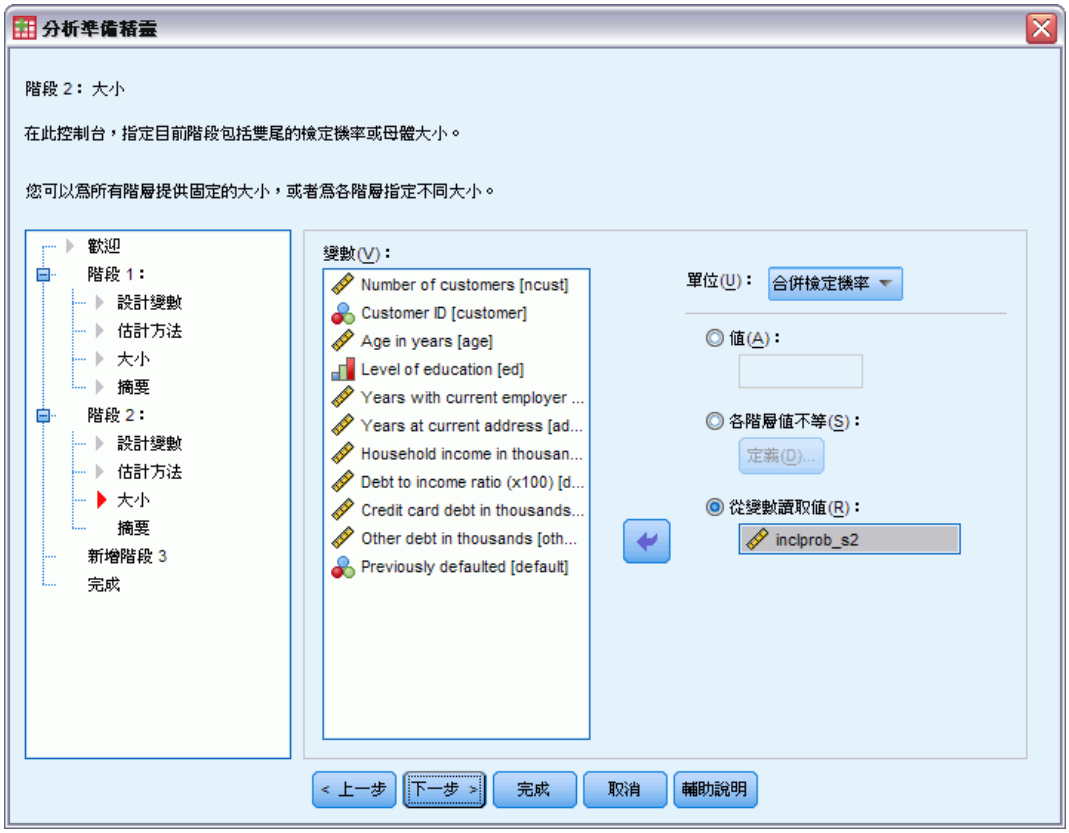


- 選取「是，立即新增階段 2」。
- 按一下「下一步」，並在「設計變數」步驟按一下「下一步」。

圖表 14-12
分析準備精靈，估計方法步驟（階段 2）

- 選取「相等 WOR」做為評估方法的第二步驟。
- 按一下「下一步」。

圖表 14-13
分析準備精靈，大小步驟（階段 2）



- 選取「從變數讀取數值」，並選取「inclprob_s2」做為含有第二階段包含機率的變數。
- 按一下「完成」。

摘要

圖表 14-14
摘要表(U)

			第 1 階段	第 2 階段
設計變數	集群	1	分店	
分析資訊	估計量假設		沒有置換的相等機率取樣	沒有置換的相等機率取樣
	包括雙尾的檢定機率		取自變數 inclprob_s1	取自變數 inclprob_s2

規劃檔案：c:\bankloan.csaplan
加權變數：finalweight

摘要表會檢閱您的分析計劃。計劃包含一集群變數設計的兩階段。使用相等機率、取樣後不放回 (WOR) 來估計，並將計劃儲存至 c:\bankloan.csaplan。現在您可使用計劃檔案，以「複合樣本」分析程序來處理 bankloan_noweights.sav（使用您計算出的包含機率和取樣權重）。

相關程序

「複合樣本分析準備精靈」程序在您不具有取樣計劃檔案存取權而須備妥要分析的樣本時，是相當實用的工具。

- 若要建立取樣計劃檔並抽樣，請使用「取樣精靈」。

章節
15

複合樣本次數分配表

「複合樣本次數分配表」程序可產生所選變數的次數分配表，並顯示單變量統計。您可隨意依一或多類別變數所定義的次組別要求統計量。

使用「複合樣本次數分配表」分析營養補充品使用情形。

一研究員希望使用「國民健康訪問調查 (NHIS)」與已預先建立的分析計劃來研究美國國民的營養補充品使用情形。如需詳細資訊，請參閱第 132 頁第 14 章中的使用「複合樣本分析準備精靈」以備妥 NHIS 公用資料.

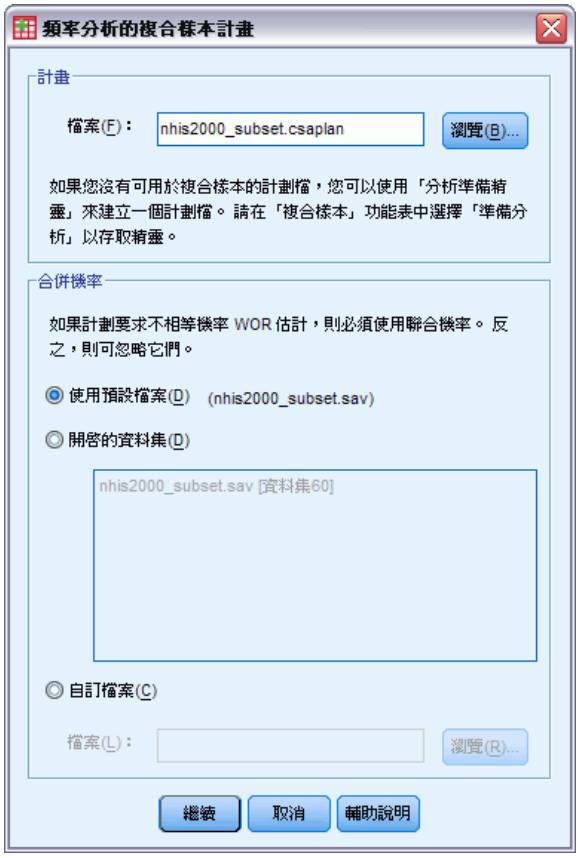
在 nhis2000_subset.sav 檔中收集了 2000 年調查的子集。分析計劃儲存在 nhis2000_subset.csaplan 中。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.

使用「複合樣本次數」產生營養補充品使用情形的統計量。

執行分析

- 若要執行「複合樣本次數」分析，請從功能表選擇：
  分析 > 複合樣本 > 次數分配表...

圖表 15-1
複合樣本計劃對話方塊



- 瀏覽並選取 nhis2000_subset.csaplan。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.
- 按一下「繼續」。

圖表 15-2
「次數分配表」對話方塊

- 選取過去 12 個月補充的維他命/礦物質做為次數變數。
- 選取年齡類別做為子母體變數。
- 按一下「統計量」。

圖表 15-3
次數分配表統計量對話方塊

- 在「儲存格」組別中選取**表格百分比**。
- 在「統計量」組別中選取**信賴區間**。
- 按一下「繼續」。
- 按一下「次數分配表」對話方塊中的**確定**。

次數表

圖表 15-4
變數/情況次數表

		估計	標準誤	%95 信賴區間	
				上界	下界
母群大小	是	1.0E+08	1185126.7	1.0E+08	1.1E+08
	否	9.1E+07	1094401.9	8.9E+07	9.3E+07
	總和	1.9E+08	1789098.7	1.9E+08	2.0E+08
總和百分比	是	53.1%	.4%	52.4%	53.8%
	否	46.9%	.4%	46.2%	47.6%
	總和	100.0%	.0%	100.0%	100.0%

會為每個所選儲存格測量計算所選統計量。第一行包含服用與不服用維他命/礦物質補充品的母群個數及百分比估計。 信賴區間不重疊，因此您可概要地推斷出，服用維他命/礦物質補充品的美國人比不服用者多。

子母體次數表

圖表 15-5
子母體的次數表

年紀類別			估計	標準誤	%95 信賴區間	
					下界	上界
18-24	母群大小	是	1.0E+07	350602.35	9328682	1.1E+07
		否	1.5E+07	499182.39	1.4E+07	1.6E+07
		總和	2.5E+07	680732.81	2.4E+07	2.7E+07
	總和百分比	是	39.3%	1.0%	37.4%	41.2%
		否	60.7%	1.0%	58.8%	62.6%
		總和	100.0%	.0%	100.0%	100.0%
25-44	母群大小	是	3.9E+07	660855.72	3.8E+07	4.0E+07
		否	4.0E+07	645934.19	3.8E+07	4.1E+07
		總和	7.9E+07	961114.33	7.7E+07	8.1E+07
	總和百分比	是	49.8%	.6%	48.7%	50.9%
		否	50.2%	.6%	49.1%	51.3%
		總和	100.0%	.0%	100.0%	100.0%
45-64	母群大小	是	3.4E+07	598603.73	3.3E+07	3.5E+07
		否	2.4E+07	497723.83	2.3E+07	2.5E+07
		總和	5.8E+07	814680.41	5.7E+07	6.0E+07
	總和百分比	是	58.7%	.6%	57.5%	60.0%
		否	41.3%	.6%	40.0%	42.5%
		總和	100.0%	.0%	100.0%	100.0%
65+	母群大小	是	1.9E+07	439459.79	1.9E+07	2.0E+07
		否	1.2E+07	314238.08	1.1E+07	1.2E+07
		總和	3.1E+07	587623.44	3.0E+07	3.2E+07
	總和百分比	是	62.2%	.7%	60.7%	63.6%
		否	37.8%	.7%	36.4%	39.3%
		總和	100.0%	.0%	100.0%	100.0%

由次母群組計算統計量時，會從年齡類別值計算各所選儲存格測量的各所選統計量。第一行包含服用與不服用維他命/礦物質補充品的各類別母群個數及百分比估計。 表格百分比的信賴區間完全不重疊，因此您可概要地推斷出，維生素/礦物質補充品的使用量也隨著年齡而增加。

摘要

使用「複合樣本次數」程序，您可獲得美國國民營養補充品使用情形的統計量。

- 總體而言，服用維生素/礦物質補充品的美國人比不服用者多。
- 若以年齡區分的話，有很大比例的美國人隨著年齡的增加開始服用維生素/礦物質補充品。

相關程序

「複合樣本次數」程序對於取得由複雜取樣設計所得的觀察值之類別變數的單變量敘述統計而言，是相當實用的工具。

- 「複合樣本取樣精靈」用以指定複雜取樣設計規格並取得樣本。「取樣精靈」建立的取樣計劃檔包含拖欠分析計劃，並可在分析以此計劃取得的樣本時，於「計劃」對話方塊中指定。
- 「複合樣本分析準備精靈」可用來為現有複合樣本設定分析指定。您可在分析此計劃相關的樣本時，在「計劃」對話方塊中指定由「取樣精靈」建立的分析計劃檔。
- 「複合樣本交叉表」程序可提供類別變數的交叉表列敘述統計。
- 「複合樣本敘述統計量」程序提供尺度變數的單變量敘述統計。

章節
16

複合樣本敘述統計量

「複合樣本敘述統計量」程序顯示數個變數的單變量摘要統計量。您可隨意依一或多類別變數所定義的次組別要求統計量。

使用複合樣本描述性統計量分析活動水準

一研究員希望使用「國民健康訪問調查 (NHIS)」與已預先建立了分析計劃來研究美國國民的活動量。如需詳細資訊，請參閱第 132 頁第 14 章中的使用「複合樣本分析準備精靈」以備妥 NHIS 公用資料.

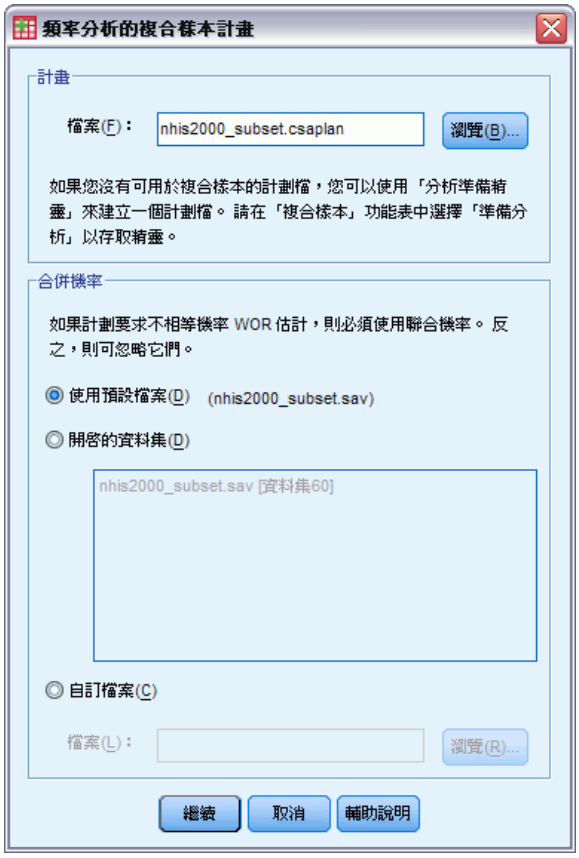
在 nhis2000_subset.sav 檔中收集了 2000 年調查的子集。分析計劃儲存在 nhis2000_subset.csaplan 中。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.

使用「複合樣本敘述統計量」產生活動水準的單變量敘述統計。

執行分析

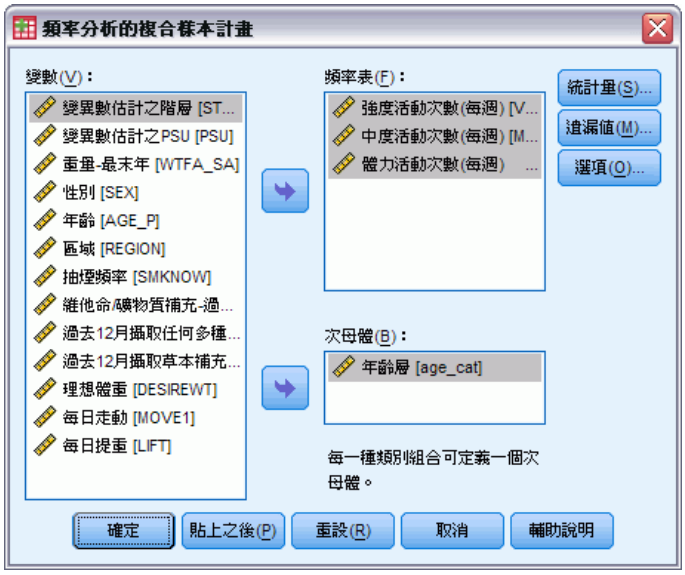
- 若要執行「複合樣本敘述統計量」分析，請從功能表選擇：
  分析 > 複合樣本 > 敘述統計量...

圖表 16-1
複合樣本計劃對話方塊



- 瀏覽並選取 nhis2000_subset.csaplan。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.
- 按一下「繼續」。

圖表 16-2
描述性統計量對話方塊



- 選取劇烈活動次數（每週）至肌力活動次數（每週）做為測量變數。
- 選取年齡類別做為子母體變數。
- 按一下「統計量」。

圖表 16-3
「敘述統計量」對話方塊

- 在「統計量」組別中選取信賴區間。
- 按一下「繼續」。
- 在「複合樣本敘述統計量」對話方塊中按一下確定。

單變量統計量

圖表 16-4
單變量統計量

		估計	標準誤	95% 信賴區間	
				下界	上界
平均	劇烈活動（每週）	3.73	.033	3.66	3.79
	溫和活動（每週）	4.90	.041	4.82	4.98
	肌力活動（每週）	3.52	.042	3.43	3.60

會為每個測量變數計算所選統計量。第一行包含每人每週從事特定類型活動的平均個數估計。 平均數的信賴區間不重疊。因此您可概要地推斷出，美國人從事肌力活動的次數少於激烈活動，且他們從事劇烈活動的次數少於溫和活動。

子母體的單變量統計量

圖表 16-5
子母體的單變量統計量

年紀類別			估計	標準誤	95% 信賴區間	
					下界	上界
18-24	平均	劇烈活動（每週）	3.92	.087	3.75	4.09
		溫和活動（每週）	5.18	.137	4.91	5.45
		肌力活動（每週）	3.45	.085	3.28	3.62
25-24	平均	劇烈活動（每週）	3.55	.048	3.46	3.65
		溫和活動（每週）	4.73	.056	4.62	4.84
		肌力活動（每週）	3.28	.052	3.18	3.38
45-64	平均	劇烈活動（每週）	3.79	.063	3.66	3.91
		溫和活動（每週）	4.88	.070	4.74	5.02
		肌力活動（每週）	3.65	.092	3.47	3.84
65+	平均	劇烈活動（每週）	4.18	.111	3.96	4.39
		溫和活動（每週）	5.22	.084	5.06	5.39
		肌力活動（每週）	4.66	.155	4.36	4.97

會為年齡類別值的各測量變數計算各所選統計量。第一行包含各類別人士每週從事特定類型活動平均個數的估計。 平均數的信賴區間可讓您推斷出一些有趣的結論。

- 在劇烈及溫和活動方面，25 - 44 歲者活動次數比 18 - 24 歲及 45 - 64 歲者少，且 45 - 64 歲者的活動量比 65 歲以上者少。
- 在肌力活動方面，25 - 44 歲者活動次數比 45 - 64 歲及 18 - 24 歲者少，且 45 - 64 歲者的活動量比 65 歲以上者少。

摘要

使用「複合樣本敘述統計量」程序，您可獲得美國國民活動量的統計量。

- 整體而言，美國人花在各類活動上的時間各異。
- 若以年齡區分的話，粗略顯示出美國大學生在學時較少運動，但隨著年齡的增長會較為勤奮。

相關程序

「複合樣本敘述統計量」程序對於取得由複雜取樣設計所得的觀察值之尺度測量的單變量敘述統計而言，是相當實用的工具。

- 「複合樣本取樣精靈」用以指定複雜取樣設計規格並取得樣本。「取樣精靈」建立的取樣計劃檔包含拖欠分析計劃，並可在分析以此計劃取得的樣本時，於「計劃」對話方塊中指定。
- 「複合樣本分析準備精靈」可用來為現有複合樣本設定分析指定。您可在分析此計劃相關的樣本時，在「計劃」對話方塊中指定由「取樣精靈」建立的分析計劃檔。
- 「複合樣本比例量數」程序提供尺度測量比例量數的敍述統計。
- 「複合樣本次數分配表」程序可為類別變數提供單變量敘述統計。

章節
17

複合樣本交叉表

「複合樣本交叉表」程序能產生成對選定變數的交叉表列表，並顯示二因子統計量。您可隨意依一或多類別變數所定義的次組別要求統計量。

使用複合樣本交叉表測量事件的相對風險

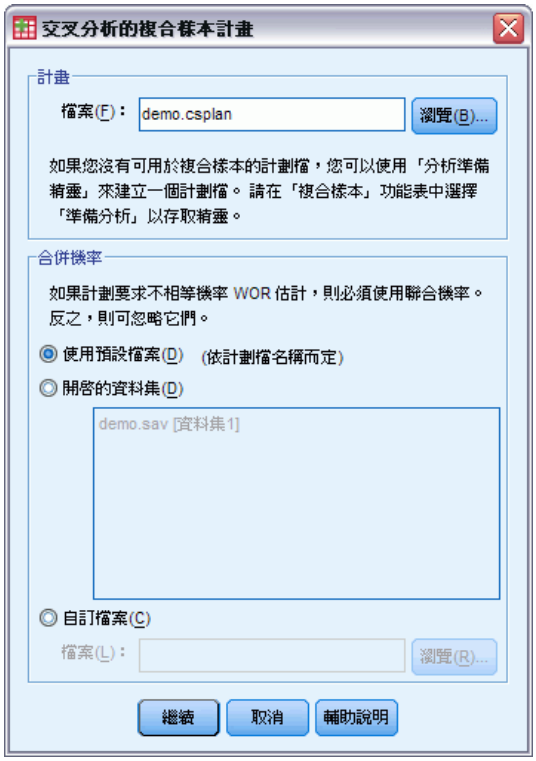
以販售訂閱雜誌為主的公司傳統上每月會寄出宣傳郵件給買來資料庫上的名單。一般回應率很低，因此您必須找出新方法，更精確的挑出準客戶。若假設閱讀報紙的人較可能訂閱雜誌，則我們可建議集中發送宣傳郵件給訂閱報紙的人。

使用「複合樣本交叉表」程序，藉由建立訂報者與回應的二乘二表格以測試此理論，並計算 「訂報者會回應宣傳郵件」的相對風險。本資訊收集在 demo_cs.sav 中，並須以取樣計劃檔 demo.csplan 來分析。 如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.

執行分析

▶ 若要執行「複合樣本交叉表」分析，請從功能表選擇：
分析 > 複合樣本 > 交叉表...

圖表 17-1
複合樣本計劃對話方塊

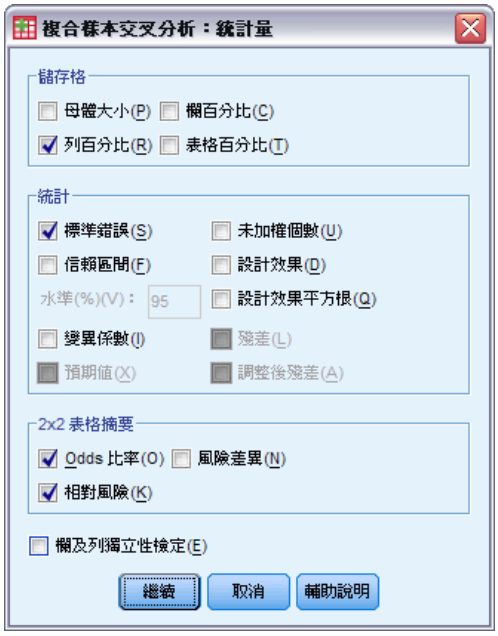


- ▶ 瀏覽並選取 demo.csplan。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.
- ▶ 按一下「繼續」。

圖表 17-2
交叉表對話方塊

- 選取訂報者做為列變數。
- 選取回應做為直欄變數。
- 看著因收入類別分類的結果好像也蠻有趣的，因此我們選取收入類別（千元）做為子母體變數。
- 按一下「統計量」。

圖表 17-3
交叉表統計量對話方塊



- 取消選取**母群大小**並在「儲存格」組別中選取**列百分比**。
- 在 2 乘 2 表組別的「摘要」中選取 **Odds 比率**及**相對風險**。
- 按一下「**繼續**」。
- 在「複合樣本交叉表」對話方塊中按一下**確定**。

這些選擇會產生訂報者對於回應的交叉表列表及風險估計。亦會產生由收入類別（千元）所分割結果的個別表格。

交叉表列

圖表 17-4
訂報者回應的交叉表列

訂報者			回應		
			是	否	總和
是	訂報者的 %	估計	18.6%	81.4%	100.0%
		標準誤	1.2%	1.2%	.0%
否	訂報者的 %	估計	10.6%	89.4%	100.0%
		標準誤	.7%	.7%	.0%
總和	訂報者的 %	估計	13.5%	86.5%	100.0%
		標準誤	.7%	.7%	.0%

交叉表概要顯示出，整體來講，回應宣傳郵件的人並不多。然而，訂報者中回應的比例大出許多。

風險估計

圖表 17-5
訂報者回應的風險估計

			估計
訂報者 *回應	Odds 比率		1.912
	相對風險	回應 = 是	1.743
		回應 = 否	.911

僅計算含所有已觀察儲存格的 2-by-2 表統計量。

相對風險是事件機率的比例量數。回應宣傳郵件的相對風險為，訂報者回應機率對非訂報者回應機率的比例量數。如此，相對風險的估計約為 17.2%/10.3% = 1.673。 同樣的，無回應的相對風險為，訂報者不回應機率對非訂報者不回應的比例量數。此相對風險的估計為 0.923。有了這些結果，您可估計訂報者回應宣傳郵件的可能性為非訂報者的 1.673 倍，或非訂報者不回應的可能性之 0.923 倍。

Odds 比率為事件機會比例量數。事件機會是事件發生的機率對事件不發生的機率之比例量數。因此，訂報者回應宣傳郵件的機會估計為 17.2%/82.8% = 0.208。 同樣的，非訂報者回應的機會估計為 10.3%/89.7% = 0.115。因此，odds 比率的估計值是 0.208/0.115 = 1.812 (請注意，在過程中會有一些捨入誤差)。Odds 比率亦為回應相對風險對無回應相對風險的比例量數，或 1.673/0.923 = 1.812。

Odds 比率對相對風險

因為 odds 比率是比例量數的比例量數，故很難解讀。相對風險較易解讀，因此單獨看 odds 比率比較沒有幫助。然而，有些常見的情況中，相對風險並不良好，而 odds 比率恰可用來估算所需事件的相對風險。符合下列條件時，應使用 odds 比率估算所需事件的相對風險。

- 所需事件的機率很小 (<0.1)。此條件可保證 odds 比率可妥善估算出相對風險。在此範例中，所需事件為對宣傳郵件的回應。
- 研究的設計為觀察值控制。此條件示意相對風險的常用估計很可能並不良好。觀察值控制研究是回溯性的，最常用於所需事件不可能或準實驗的設計不切實際或不道德時。

此範例中兩個條件均不符合，因應答者的整體比例為 12.8%，且研究的設計非觀察值控制，故比起 odds 比率的數值，將相對風險報告為 1.673 會較安全。

由字母群體進行風險估計

圖表 17-6
訂報者對回應的風險估計，收入類別的控制

收入類別				估計
$25 以下	訂報者 * 回應	Odds 比率		2.608
		相對風險	回應 = 是	2.149
			回應 = 否	.824
$25 - $49	訂報者 * 回應	Odds 比率		1.923
		相對風險	回應 = 是	1.737
			回應 = 否	.903
$50 - $74	訂報者 * 回應	Odds 比率		1.538
		相對風險	回應 = 是	1.493
			回應 = 否	.971
$75+	訂報者 * 回應	Odds 比率		1.200
		相對風險	回應 = 是	1.182
			回應 = 否	.985

僅計算含所有已觀察儲存格的 2-by-2 表統計量。

相對風險估計是由各收入類別分別計算出。請注意，隨著收入的增加，訂報者的正反應相對風險呈現逐漸降低的趨勢，指出您可以進一步鎖定郵寄對象。

摘要

使用「複合樣本交叉表」的風險估計，您會發現您可藉由以訂報者為郵寄宣傳目標以增加回應率。接著，您會發現一些證據指出，收入類別中的所有風險估計並非常數，故您可以藉由鎖定收入較低的訂報者，進一步增加回應率。

相關程序

「複合樣本交表」程序對於取得由複雜取樣設計所得的觀察直之類別變數交叉表列的敘述統計而言，是相當實用的工具。

- 「複合樣本取樣精靈」用以指定複雜取樣設計規格並取得樣本。「取樣精靈」建立的取樣計劃檔包含拖欠分析計劃，並可在分析以此計劃取得的樣本時，於「計劃」對話方塊中指定。
- 「複合樣本分析準備精靈」可用來為現有複合樣本設定分析指定。您可在分析此計劃相關的樣本時，在「計劃」對話方塊中指定由「取樣精靈」建立的分析計劃檔。
- 「複合樣本次數分配表」程序可為類別變數提供單變量敘述統計。

章節
18

複合樣本比例量數

「複合樣本比例量數」程序會顯示變數比例量數的單變量摘要統計量。您可隨意依一或多類別變數所定義的次組別要求統計量。

使用複合樣本比例量數協助財產價值評估

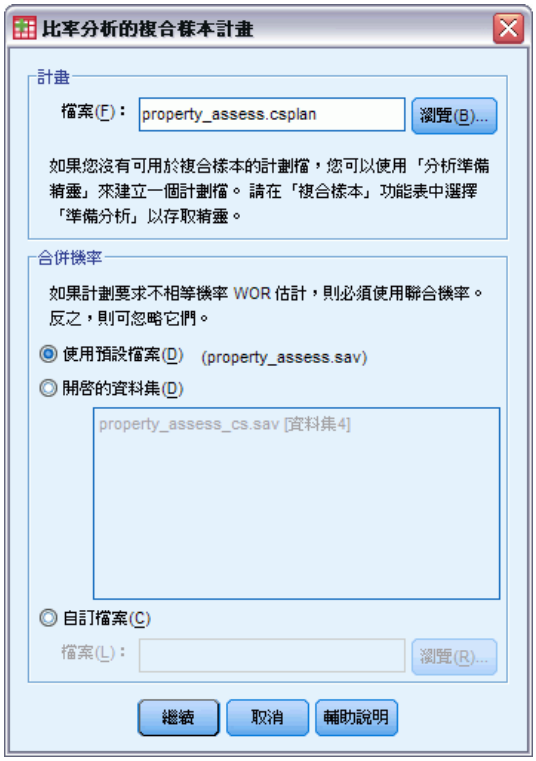
州立機構會確定各郡財產稅均合理公平後才徵收。稅額根據財產的估計值而得，故機構會希望在各郡間追蹤財產價值，以確定各郡的紀錄均保持最新。由於取得目前估價的資源有限，故機構會選擇利用複雜取樣方法以選取財產。

所選取屬性的樣本和它們目前的估價資訊均收錄在 property_assess_cs_sample.sav 中。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 使用「複合樣本比例量數」評估五郡中上次估價後的財產價值變更。

執行分析

▶ 若要執行「複合樣本比例量數」分析，請從功能表選擇：
分析 > 複合樣本 > 比例量數...

圖表 18-1
複合樣本計劃對話方塊

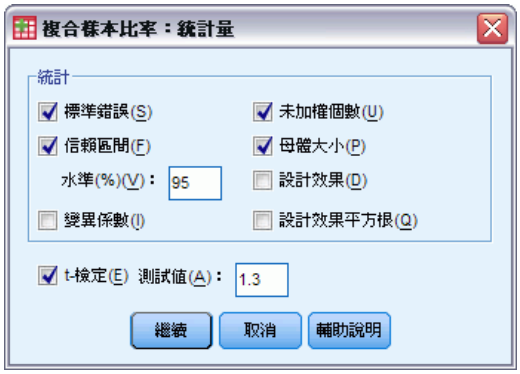


- 瀏覽並選取 property_assess.csplan。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.
- 按一下「繼續」。

圖表 18-2
比例量數對話方塊

- 選取目前價值做為分子變數。
- 選取上次估價價值做為分母變數。
- 選取郡做為子母體變數。
- 按一下「統計量」。

圖表 18-3
比例量數統計量對話方塊



- 在「統計量」組別中選取信賴區間、未加權個數和母群大小。
- 選取 t 檢定並輸入 1.3 做為檢定值。
- 按一下「繼續」。
- 在「複合樣本比例量數」對話方塊中按一下確定。

比例量數

圖表 18-4
比例量數表

郡	分子	分母	估計值比例	標準誤	95% 信賴區間	
					下界	上界
東部	目前數值	上次估價價值	1.381	.068	1.236	1.5
中部	目前數值	上次估價價值	1.364	.064	1.227	1.5
西部	目前數值	上次估價價值	1.524	.053	1.410	1.6
北部	目前數值	上次估價價值	1.277	.032	1.208	1.3
南部	目前數值	上次估價價值	1.195	.029	1.134	1.2

表格的預設顯示非常寬，故您必須進行樞軸分析以方便檢視。

對比例量數表進行樞軸分析

- ▶ 連按兩下以啟動表格。
- ▶ 從「瀏覽器」功能表選擇：
  樞軸 > 樞軸分析匣
- ▶ 依序將分子及分母從列中拖曳至層。
- ▶ 將郡從列中拖曳至行。
- ▶ 將統計量從行中拖曳至列。
- ▶ 關閉樞軸分析匣視窗。

已進行樞軸分析的比例量數表

圖表 18-5
已進行樞軸分析的比例量數表

分子：目前價
分母：上次估價價值

		國家				
		東部	中部	西部	北部	南部
估計值的比例量數		1.381	1.364	1.524	1.277	1.195
標準誤		.068	.064	.053	.032	.029
95% 信賴區間	下界	1.236	1.227	1.410	1.208	1.134
	上界	1.525	1.502	1.638	1.346	1.256
假設檢定	檢定	1.3	1.3	1.3	1.3	1.3
		1.191	.997	4.201	-.702	-3.646
		15	15	15	15	15
		.252	.334	.001	.493	.002
母群大小		1883.250	1557.500	4044.000	2306.250	2204.000
未加權個數		168	179	202	205	220

比例量數表現已進行樞軸分析，故較容易比較各郡的統計量。

- 比例量數估計範圍從南部郡的 1.195 到西部郡的 1.524。
- 標準誤中亦有一些變異性，其範圍從南部郡的 0.029 到東部郡的 0.068。
- 有些信賴區間並不重疊，故您可推論西部郡的比例量數比北部和南部郡的比例量數高。
- 最後，若要進行更客觀的測量，請注意西部及南部郡的 t 檢定顯著值均小於 0.05。因此，您可推論西部郡的比例量數大於 1.3，而南部郡的比例量數小於 1.3。

摘要

使用「複合樣本比例量數」程序，您會取得目前價值至上次估價價值比例量數的不同統計量。結果表示郡至郡的財產稅評估可能有些不公正，也就是：

- 西部郡的比例量數高，表示他們關於財產價值增值部分的紀錄未像其他郡般更新。本郡的財產稅可能太低。
- 南部郡的比例量數低，表示他們關於財產價值增值部分的紀錄比其他郡還新。本郡的財產稅可能太高。
- 南部郡的比例量數低於西部郡的比例量數，但仍在 1.3 的客觀目標內。

用於追蹤南部郡財產價值的資源將會重新指派給西部郡，以使這兩個郡的比例量數與其他郡一致，並達到目標 1.3。

相關程序

「複合樣本比例量數」程序對於透過複雜取樣設計所得的觀察值，其在取得尺度測量比例量數的單變量敘述統計方面，是相當實用的工具。

- 「複合樣本取樣精靈」用以指定複雜取樣設計規格並取得樣本。「取樣精靈」建立的取樣計劃檔包含預設分析計劃，並可在分析以此計劃取得的樣本時，於「計劃」對話方塊中指定。
- 「複合樣本分析準備精靈」可用來為現有複合樣本設定分析指定。您可在分析此計劃相關的樣本時，在「計劃」對話方塊中指定由「取樣精靈」建立的分析計劃檔。
- 「複合樣本敘述統計量」程序提供尺度變數的敘述統計。

章節
19

複合樣本一般線性模式

「複合樣本一般線性模式 (CSGLM)」程序會為複合取樣方法取得的樣本執行線性迴歸分析，以及變異數與共變異數分析。或者，您也可以要求對子母體進行分析。

使用複合樣本一般線性模式以配合二因子 ANOVA（變異數分析）

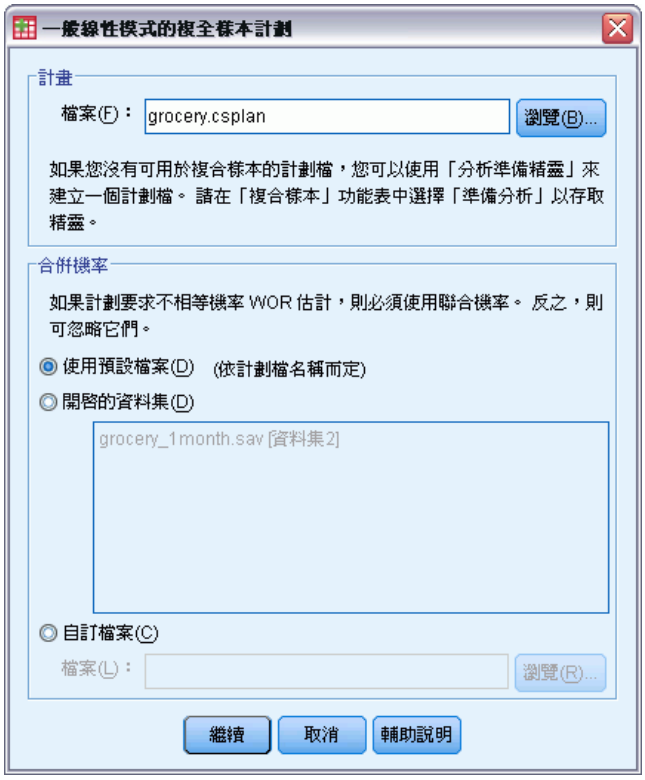
某連鎖雜貨店根據複合設計，調查一組客戶的購買習慣。已知調查結果以及客戶在上個月的花費，該店希望對照客戶性別並結合取樣設計，了解客戶購物的頻率是否與整月支出的量有關。

這項資訊收集於 grocery_1month_sample.sav 中。 如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 使用「複合樣本一般線性模式」程序，對花費的數量執行二因子變異數分析。

執行分析

- 若要執行「複合樣本一般線性模式」分析，請從功能表選擇：
  分析 > 複合樣本 > 一般線性模式...

圖表 19-1
複合樣本計劃對話方塊

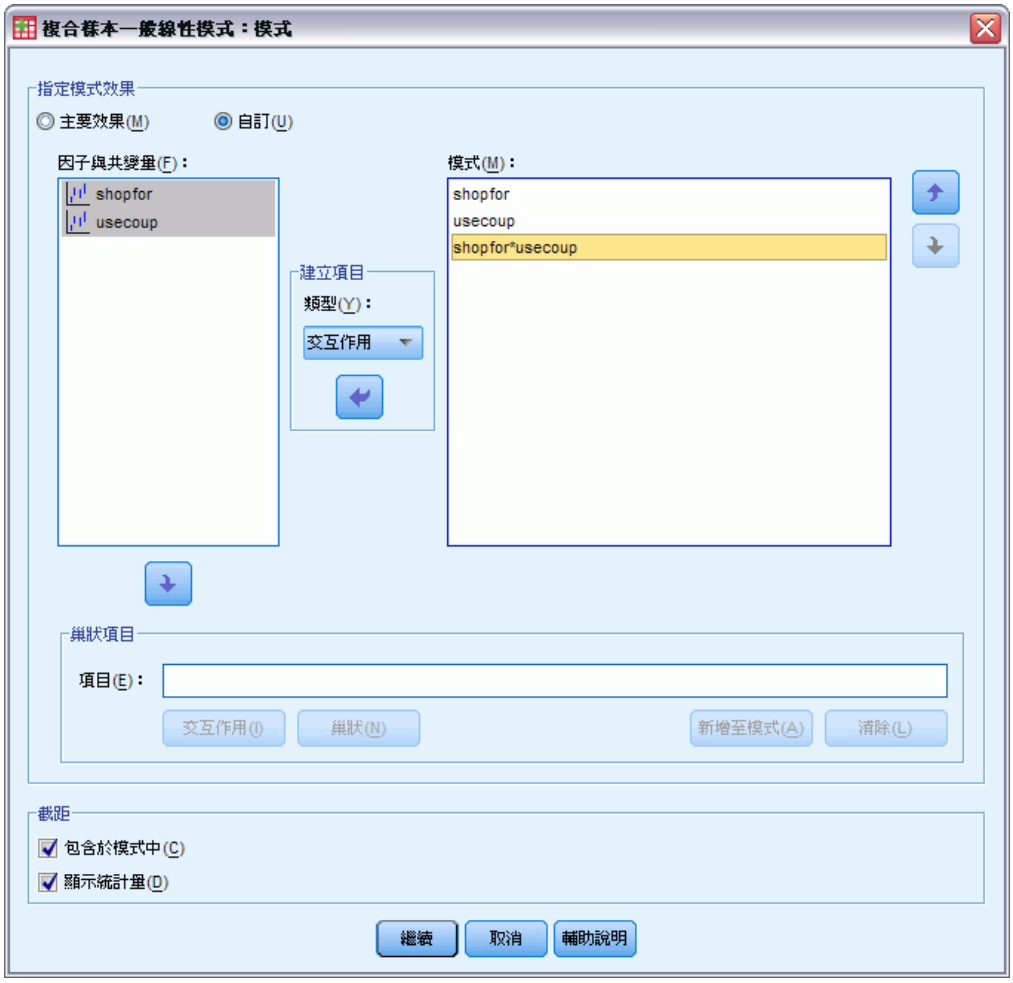


- 瀏覽並選取 grocery.csplan。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.
- 按一下「繼續」。

圖表 19-2
一般線性模式對話方塊

- 選取「花費數量」做為依變數。
- 選取「誰買什麼」和「使用折價券」為因子。
- 按一下「**模式**」。

圖表 19-3
「模式」對話方塊



- 選擇建立自訂模式。
- 選取「主效果」做為建立的項目類型，並選取「shopfor」和「usecoup」做為模式項目。
- 選取「交互作用」做為建立的項目類型，並加入「shopfor*usecoup」交互作用做為模式項目。
- 按一下「繼續」。
- 在「一般線性模式」對話方塊中按一下「統計量」。

ê><
Ô8 3® û õ ã3E?ÜG£ á@E •

E ü Ë õ ã — L Ì3 ù ÈFL ª Ë ?Ü Ì Ã Ë í$j@x Ì Ã Ë µBÈ Kg Ì ` Ë @ ?Ü • Ì Ä

E Ý Ô ß Ë4P4`Ì Ä

E ü Ë Ô8 3® û õ ã Ì á@E • Ý Ô ß Ë ?Ü G L Ì Ä

ê><
Ô8 3® û õ ã ?Ü G L á@E •

E FL ªNC/ V K R S I R U Ã X V H F R X S ` V K R S I R U X V H F R X S x f 0*ü,X G L Ä

- ▶ 選取「簡單對比」和「3 位個人和家庭」，做為 shopfor 的參考類別。請注意，選取之後，類別會在對話方塊中顯示「3」。
- ▶ 選取「簡單對比」和「1 No」，做為 usecoup 的參考類別。
- ▶ 按一下「繼續」。
- ▶ 在「一般線性模式」對話方塊中按一下「確定」。

模式摘要

圖表 19-6
R 平方統計量

R 平方	.601

a. 模式: Amount spent = (截距) + shopfor + usecoup + shopfor * usecoup

R 平方是判斷係數，為模式適合度之強度測量。能顯示在花費數量中約有 60%的變異數可以透過模式來解釋，並能提供您良好的解釋能力。您可能仍想將其他預測值新增到模式中，進而改善適合度。

模式效應的檢定

圖表 19-7
所有受試者間效應的檢定

來源	df1	df2	Wald F	Sig.
(更正的模型)	11.000	3.000	127.231	.001
(截距)	1.000	13.000	6321.597	.000
shopfor	2.000	12.000	643.593	.000
usecoup	3.000	11.000	87.453	.000
shopfor * usecoup	6.000	8.000	10.688	.002

a. 模式: Amount spent = (截距) + shopfor + usecoup + shopfor * usecoup

模式中的每一個項目，加上整個模式，都要檢定其影響值是否為 0。顯著值小於 0.05 的項目會含有某些明顯效應。如此一來，所有模式項目都有助於模式。

參數估計值

圖表 19-8
參數估計值

參數	估計	標準 錯誤	95% 信賴區間		設計效果
			較低	較高	
(截距)	518.249	11.731	492.905	543.592	1.387
[shopfor=1]	-174.757	10.762	-198.006	-151.508	.950
[shopfor=2]	-129.443	11.455	-154.191	-104.696	.925
[shopfor=3]	.000[a]	.	.	.	.
[usecoup=1]	-140.838	10.180	-162.830	-118.846	.649
[usecoup=2]	-63.026	13.195	-91.531	-34.520	.940
[usecoup=3]	-31.375	9.726	-52.387	-10.363	.564
[usecoup=4]	.000[a]	.	.	.	.
[shopfor=1] * [usecoup=1]	41.693	11.170	17.562	65.824	.606
[shopfor=1] * [usecoup=2]	44.505	18.068	5.471	83.539	1.413
[shopfor=1] * [usecoup=3]	9.204	11.057	-14.684	33.092	.594
[shopfor=1] * [usecoup=4]	.000[a]	.	.	.	.
[shopfor=2] * [usecoup=1]					

參數估計值會在「花費數量」上顯示各個預測值的效應。數值為 518.249 的截距項目表示連鎖雜貨店可以預期使用報紙或鎖定郵寄折價券之購物者家庭的平均花費為 518.25 美元。您會發現截距與這些因子水準是互相關連的，因為這些因子水準的參數是多餘的。

- shopfor 係數暗示，在使用報紙折價券與郵寄折價券的顧客之中，沒有家庭的人會傾向比有配偶的人花費少，而有配偶的人又比家裡有其他成員的人花費少。因為模式效應的檢定顯示此項目有助於模式，因此這些差異不是偶然的。
- usecoup 係數暗示，家裡有其他成員之顧客的花費會隨著檢少折價券的使用而減少。估計值中有適量的不確定，但是信賴區間並不包括 0。
- 交互作用係數暗示，不使用折價券或是只從報紙剪下折價券、家裡沒有其他成員的顧客會傾向於比您預期的花費更多。如果有任何一部份的交互作用參數是多餘的，則交互作用參數就是多餘的。
- 設計效應數值中的離差從 1 開始，表示這些參數估計值所計算出的某些標準誤，比當您假設這些觀察是從簡單隨機樣本中所取得的標準誤還大，而其他的標準誤則較小。將取樣設計資訊併入分析中是非常重要的，否則，舉例來說，您可能會推論 usecoup=3 係數為與 0 沒有差別！

參數估計值對於將各個模式項目的效應加以量化很有幫助，但是估計的邊際平均數表格可以讓您以更輕鬆的方式解讀模式結果。

邊際平均數估計

圖表 19-9
透過「誰買什麼」的水準估計邊際平均數

Who shopping for	平均數	標準錯誤	95% 信賴區間	
			較低	較高
Self	308.5326	3.94286	300.0145	317.0506
Self and spouse	370.3361	4.87908	359.7955	380.8767
Self and family	459.4392	7.19769	443.8895	474.9888

這個表格會顯示模式估計的邊際平均數，以及在「誰買什麼」的因子水準中之「花費數量」的標準誤。這個表格對於探索因子水準之間的不同很實用。 在這個例子中，為他/她自己購物的顧客預期會花費約 308.53 美元，而有配偶的顧客則預期會花費約 370.34 美元，家裡還有其他成員的顧客則會花費約 459.44 美元。如果要瞭解是否表示真正的差異，或是由於機會不同所造成，請查閱檢定結果。

圖表 19-10
性別之估計邊際平均數的個別檢定結果

Who shopping for 簡單對照[a]	相對估計值	假設值	差異 (估計 - 假設的)	標準錯誤	df1	df2	Wald F	Sig.
水準 Self 相對於水準 Self and family	-150.907	.000	-150.907	4.903	1.000	13.000	947.409	.000
水準 Self and spouse 相對於水準 Self and family	-89.103	.000	-89.103	5.903	1.000	13.000	227.842	.000

a. 參考類別 = Self and family

個別檢定表格會顯示兩個簡單的花費對比。

- 對比估計為「誰買什麼」所列出之花費水準的差異。
- 數值為 0.00 的假設值，表示花費是沒有差異的。
- Wald F 統計量 (含有顯示的自由度)，會用來檢定對比估計和假設值之間的差異，是否由於機會改變所造成。
- 因為顯著數值小於 0.05，結論就是花費是有所差異。

對比估計值的數值與參數估計值的數值是不同的。這是因為一個交互作用項目中含有「誰買什麼」效應。如此一來，shopfor=1 的參數估計為在「使用折價券」變數上「自兩者」水準中之「自己」水準和「自己與家庭」水準之間的簡單對比。會針對各個「使用折價券」水準，算出表格中對比估計的平均值。

圖表 19-11
性別之估計邊際平均數的整體檢定結果

df1	df2	Wald F	顯著性
2.000	12.000	643.593	.000

整體檢定表格會報告個別檢定表格中所有對比之檢定結果。其小於 0.05 的顯著值能確定「誰買什麼」之水準中的花費有差異。

圖表 19-12
透過購物型式水準估計邊際平均數

Use coupons	平均數	標準 錯誤	95% 信賴區間	
			較低	較高
No	319.6455	6.51429	305.5722	333.7188
From newspaper	386.7469	4.32295	377.4077	396.0861
From mailings	394.5028	5.54218	382.5297	406.4760
From both	416.8486	6.51260	402.7790	430.9182

這個表格會顯示模式估計的邊際平均數，以及在「使用折價券」的因子水準中「花費數量」的標準誤。這個表格對於探索因子水準之間的不同很實用。 本例中，不使用折價券的顧客預期會花費約 319.65 美元，而那些使用折價券的顧客則預期花費較多。

圖表 19-13
購物型式之估計邊際平均數的個別檢定結果

Use coupons 簡單對照[a]	相對估 計值	假設值	差異 (估計 - 假設的)	標準 錯誤	df1	df2	Wald F	Sig.
水準 From newspaper 相對於水準 No	67.101	.000	67.101	6.537	1.000	13.000	105.352	.000
水準 From mailings 相對於水準 No	74.857	.000	74.857	5.875	1.000	13.000	162.328	.000
水準 From both 相對於水準 No	97.203	.000	97.203	5.603	1.000	13.000	300.921	.000

a. 參考類別= No

個別檢定表格會顯示三個對比，以比較使用折價券與不使用折價券之顧客的花費。

因為檢定的顯著值小於 0.05，結論就是使用折價券的顧客傾向於比不使用折價券的顧客花費還多。

圖表 19-14
購物型式之估計邊際平均數的整體檢定結果

df1	df2	Wald F	顯著性
3.000	11.000	87.453	.000

整體檢定表格報告在個別檢定表格中所有對比之檢定結果。其小於 0.05 的顯著值能確定「使用折價券」之水準中的花費有差異。請注意，「使用折價券」和「誰買什麼」的整體檢定會等於模式效應的檢定，因為假設對比值等於 0。

圖表 19-15
透過購物型式使用性別水準來估計邊際平均數

Who shopping for	Use coupons	平均數	標準錯誤	95% 信賴區間	
				較低	較高
Self	No	244.3471	6.00949	231.3644	257.3298
	From newspaper	324.9708	5.94134	312.1353	337.8063
	From mailings	321.3207	4.11028	312.4410	330.2005
	From both	343.4916	6.57845	329.2797	357.7034
Self and spouse	No	337.1783	7.12181	321.7925	352.5640
	From newspaper	380.0468	7.91038	362.9574	397.1361
	From mailings	375.3141	6.22468	361.8665	388.7617
	From both	388.8054	7.12101	373.4214	404.1894
Self and family	No	377.4111	11.58215	352.3894	402.4328
	From newspaper	455.2232	6.14420	441.9494	468.4969
	From mailings	486.8736	10.76529	463.6166	510.1306
	From both	518.2488	11.73120	492.9050	543.5925

這個表格會顯示模式估計的邊際平均數、標準誤，以及在「誰買什麼」和「使用折價券」的因子組合中「花費數量」的信賴區間。這個表格對於探索這兩個因子（發現於模式效應檢定中）之間的交互作用效應很實用。

摘要

此例中，估計邊際平均數顯示「誰買什麼」和「使用折價券」的各個水準中，顧客間的花費不同。模式效應的檢定可以確定這個結果，同時也確定「誰買什麼*使用折價券」交互作用效應的事實。模式摘要表格顯示現存模式大約能解釋資料中超過一半的變異數，此模式很可能可以藉由新增更多預測值而加以改善。

相關程序

在依據複合取樣架構抽取觀察值時，「複合樣本一般線性模式」程序是將尺度變數模式化很好的工具。

- 「複合樣本取樣精靈」用以指定複雜取樣設計規格並取得樣本。「取樣精靈」建立的取樣計劃檔包含拖欠分析計劃，並可在分析以此計劃取得的樣本時，於「計劃」對話方塊中指定。
- 「複合樣本分析準備精靈」可用來為現有複合樣本指定分析指定。您可在分析此計劃相關的樣本時，在「計劃」對話方塊中指定由「取樣精靈」建立的分析計劃檔。
- 「複合樣本羅吉斯迴歸」程序可以讓您將類別反應值模式化。
- 「複合樣本次序迴歸」程序可以讓您將次序反應值模式化。

章節
20

複合樣本 Logistic 迴歸

「複合樣本 Logistic 迴歸」程序會在所抽取樣本的二分或多項式依變數上執行 Logistic 迴歸分析，此抽取的樣本是透過複雜取樣方法抽取的。或者，您也可以要求對子母體進行分析。

使用複合樣本 Logistic 迴歸評估信用風險

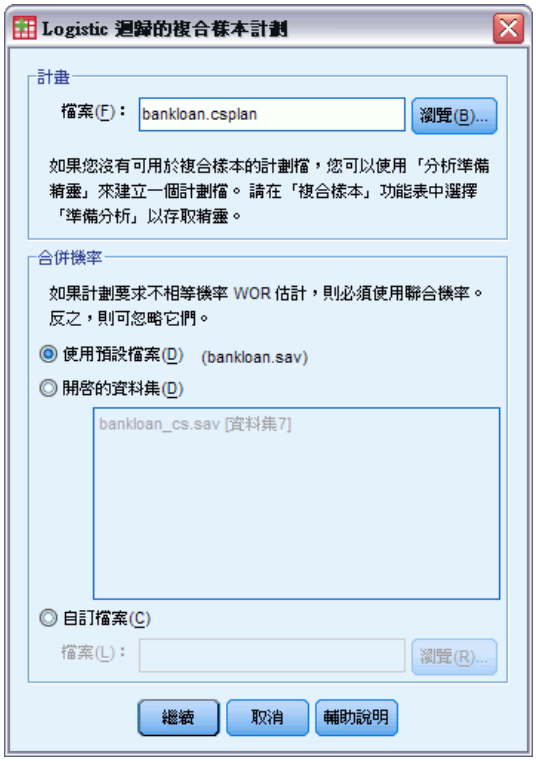
如果您是銀行放貸人員，要能辨識具有哪些特質的人可能會拖欠貸款，並使用這些特質來識別好和壞的信用風險。

假設銀行放貸人員，已經依據複雜設計，收集了一些不同分行客戶的指定貸款記錄。本資訊存放在bankloan_cs.sav中。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 該放款員希望結合樣本設計，看出顧客拖欠機率是否與年齡、工作資歷、信貸額度相關。

執行分析

▶ 若要建立 Logistic 迴歸模式，請從功能表選擇：
分析 > 複合樣本 > Logistic 迴歸...

圖表 20-1
複合樣本計劃對話方塊



- 瀏覽並選取 bankloan.csaplan。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.
- 按一下「繼續」。

圖表 20-2
Logistic 迴歸對話方塊

- 選取「先前已拖欠」為依變數。
- 選取「教育程度」為因子。
- 選取「年齡（以年為單位）」到「其他負債（以千為單位）」做為共變量。
- 選取「先前已拖欠」然後按一下「參考類別」。

圖表 20-3
Logistic 迴歸參考類別對話方塊

- 選取「最低值」為參考類別。

  會將「無拖欠」設定為參考類別，如此一來，輸出報告中的 odds 比率將會有這個性質，即 odds 比率提高和拖欠可能性提高相對應。

- 按一下「繼續」。

- 按一下在「Logistic 迴歸」對話方塊中的「統計量」。

圖表 20-4
Logistic 迴歸統計量對話方塊

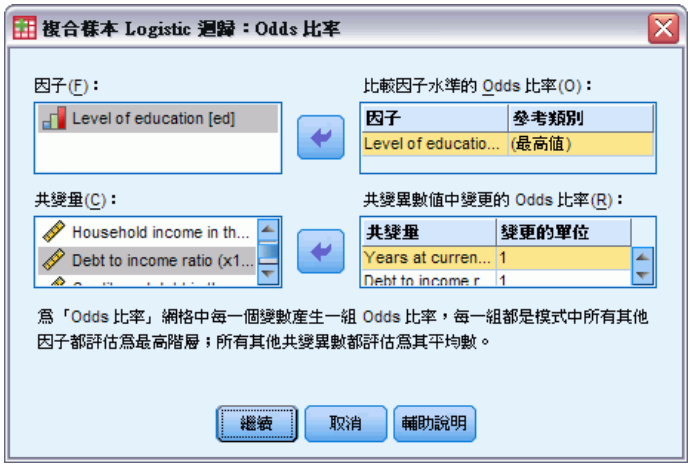
- 選取「模式適合度」組別中的「分類表」

- 在「參數」組別中，選取「估計」、「指數化估計」、「標準誤」、「信賴區間」和「設計效應」。

- 按一下「繼續」。

- 在「Logistic 迴歸」對話方塊中的「Odds 比率」。

圖表 20-5
Logistic 迴歸 Odds 比率對話方塊



- 選取以建立因子 ed 和共變量 employ 與 debtinc 的 odds 比率。
- 按一下「繼續」。
- 在「Logistic 迴歸」對話方塊中按一下「確定」。

虛擬 R 平方

圖表 20-6
虛擬 R 平方統計量

Cox 及 Snell	.330
Nagelkerke	.451
McFadden	.304

依變數: Previously defaulted (參考類別 = No)
模式: (截距), ed, age, employ, address, income, debtinc, creddebt, othdebt

在線性迴歸模式中，判別係數 R^2 摘要了在依變數中的變異數與預測值（自變數）相關的部分，以較大的 R^2 值表示本模式解釋了較多的變異數，最大到 1。對於具類別依變數的迴歸模式，是無法計算出單一個在線性迴歸模式中具有全部 R^2 特性統計量的 R^2，所以改為計算這些近似值。下列方法用來估計判別係數。

- Cox 和 Snell 的 R^2 (Cox 和 Snell, 1989) 是根據本模式的對數概似與基準線模式的對數概似相比較。不過，對於類別結果，它的理論上最大值小於 1，即使在「完美」模式也是如此。
- Nagelkerke 的 R^2 (Nagelkerke, 1991) 是 Cox & Snell R-平方的調整後版本，它調整了統計量的尺度以涵蓋由 0 到 1 的整個範圍。
- McFadden 的 R^2 (McFadden, 1974) 為另一版本，根據只有截距的模式和完全估計模式的對數概似核心。

什麼可構成「良好」的 R^2 值，會隨應用的領域不同而異。雖然這些統計量本身即有暗示性，但在以相同資料來比較競爭模式時最有用。根據本測量，具有最大 R^2 的統計量「最好」。

分類

圖表 20-7
分類表

觀察的	預測的		
	No	Yes	百分比修正
No	188289.667	31871.267	85.5%
Yes	49970.600	77675.133	60.9%
整體百分比	68.5%	31.5%	76.5%

依變數: Previously defaulted (參考類別 = No)
模式: (截距), ed, age, employ, address, income, debtinc, creddebt, othdebt

分類表顯示使用 Logistic 迴歸模式的實際結果。 如果各個觀察值之模式預測的 logit 大於 0，則其預測反應為是。觀察值是由 finalweight 來進行加權，因此類別表會報告母群的期待模式效能。

- 對角線上的儲存格為正確預測。
- 不在對角線上的儲存格為不正確預測。

根據用來建立模型的觀察值，您可以使用這個模式，將母群中 85.5% 的非拖欠者正確地分類。同樣的，您可以將 60.9% 的拖欠者正確地分類。 總體而言，您可以將 76.5% 的觀察值正確地分類，但是因為這個表格是利用建立模式的觀察值來建構，這些估計值可能過於樂觀。

模式效應的檢定

圖表 20-8
所有受試者間效應的檢定

來源	df1	df2	Wald F	Sig.
(更正的模型)	11.000	4.000	14.669	.010
(截距)	1.000	14.000	5.777	.031
ed	4.000	11.000	1.683	.224
age	1.000	14.000	5.352	.036
employ	1.000	14.000	88.244	.000
address	1.000	14.000	1.123	.307
income	1.000	14.000	.007	.932
debtinc	1.000	14.000	27.632	.000
creddebt	1.000	14.000	33.402	.000
othdebt	1.000	14.000	.709	.414

依變數: Previously defaulted (參考類別 = No)
模式: (截距), ed, age, employ, address, income, debtinc, creddebt, othdebt

模式中的每一個項目，加上整個模式，都要檢定其影響是否為 0。顯著值小於 0.05 的項目會含有某些明顯 效應。如此一來，age、employ、debtinc，和 creddebt 都有助於模式，而其他主效果則不會。在資料的進一步分析中，您可能會從模式考量移除 ed、address、income，和 othdebt。

參數估計值

圖表 20-9
參數估計值

Previously defaulted	參數	B	標準錯誤	95% 信賴區間 較低	95% 信賴區間 較高	設計效果	Exp (B)	Exp(B)的 95% 信賴區間 較低	Exp(B)的 95% 信賴區間 較高
Yes	(截距)	-1.140	.399	-1.995	-.284	.665	.320	.136	.753
	[ed=1]	.720	.340	-.010	1.449	.862	2.054	.990	4.259
	[ed=2]	.684	.371	-.112	1.481	1.247	1.983	.894	4.397
	[ed=3]	.518	.307	-.140	1.177	.813	1.679	.869	3.244
	[ed=4]	.789	.302	.142	1.437	.817	2.202	1.152	4.208
	[ed=5]	.000[a]	.	.	.	.	1.000	.	.
	age	-.023	.010	-.043	-.002	.418	.978	.958	.998
	employ	-.225	.024	-.277	-.174	1.200	.798	.758	.840
	address	-.028	.026	-.085	.029	.651	.972	.919	1.029
	income	.000	.003	-.007	.006	1.410	1.000	.993	1.006
	debtinc	.095	.018	.056	.134	1.222	1.100	1.058	1.143
	creddebt	.493	.085	.310	.676	1.373	1.637	1.363	1.966
	othdebt	.026	.031	-.041	.094	1.219	1.027	.960	1.098

依變數: Previously defaulted (參考類別 = No)
模式: (截距), ed, age, employ, address, income, debtinc, creddebt, othdebt

a. 設定為零，因為這個參數是冗餘的。

參數估計值的表格會將各個預測值的效果做成摘要。請注意，參數值會影響對照於「不曾拖欠」類別之「已經拖欠」類別的機率。如此一來，含有正係數的參數會增加拖欠的機率，而含有負值的係數則會減少拖欠的機率。

Logistic 迴歸係數的含意不如線性迴歸係數的含意明確。B 可以方便您進行檢定模式效應，而 Exp(B) 則能輕易地解釋。 針對不是交互作用項之一部分的預測值，Exp(B) 表示可在預測值中增加一單位的利息事件之機會的比率變更。例如，employ 的 Exp(B) 等於 0.798 時，表示其他條件相同時，與目前雇主共事兩年而拖欠的機會，為與目前雇主共事一年而拖欠之機會的 0.798 倍。

設計效應表示這些參數估計值所計算出的某些標準誤，比當您假設這些觀察是從簡單隨機樣本中所取得的標準誤還大，而其他的標準誤則較小。將取樣設計資訊併入分析中是非常重要的，否則，舉例來說，您可能會推論年齡係數為與 0 沒有差別!

Odds 比率

圖表 20-10
教育程度的 Odds 比率

		Previously defaulted	Odds 比率	95% 信賴區間 較低	95% 信賴區間 較高
Level of education	Did not complete high school vs. Post-undergraduate degree	Yes	2.054	.990	4.259
	High school degree vs.	Yes	1.983	.894	4.397
	Some college vs.	Yes	1.679	.869	3.244
	College degree vs.	Yes	2.202	1.152	4.208

依變數: Previously defaulted (參考類別 = No)
模式: (截距), ed, age, employ, address, income, debtinc, creddebt, othdebt

a. 計算中所使用的因子與共變量固定為以下值: Level of education=Post-undergraduate degree; Age in years=34.19; Years with current employer=6.99; Years at current address=6.32; Household income in thousands=60.1581; Debt to income ratio (x100)=9.9341; Credit card debt in thousands=1.9764; Other debt in thousands=3.9164

這個表格會顯示在教育程度的因子水準上之先前已拖欠的 odds 比率。所報告的數值為未完成中學到學院學位之拖欠的 odds 比率，與學士後學位之拖欠的 odds 比率之比較。如此一來，表格列上的 odds 比率為 2.054 時，表示未完成中學的人之拖欠機會，為具有學士後學位的人之拖欠機會的 2.054 倍。

圖表 20-11
與目前雇主共事的時間（以年為單位）之 Odds 比率

變更單位			Odds比率	95% 信賴區間	
		Previously defaulted		較低	較高
Years with current employer	1.000	Yes	.798	.758	.840

依變數: Previously defaulted (參考類別 = No)
模式: (截距), ed, age, employ, address, income, debtinc, creddebt, othdebt
a. 計算中所使用的因子與共變量固定為以下值: Level of education=Post-undergraduate degree; Age in years=34.19; Years with current employer=6.99; Years at current address=6.32; Household income in thousands=60.1581; Debt to income ratio (x100)=9.9341; Credit card debt in thousands=1.9764; Other debt in thousands=3.9164

這個表格會顯示在共變量與目前雇主共事的時間（以年為單位）中單位變更的先前已拖欠之 odds 比率。報告的數值為任現職 7.99 年的人之拖欠 odds 比率，與任現職 6.99年（平均數）的人之拖欠 odds 比率之比較。

圖表 20-12
收入與負債比率之 odds 比率

變更單位			Odds比率	95% 信賴區間	
		Previously defaulted		較低	較高
Debt to income ratio (x100)	1.000	Yes	1.100	1.058	1.143

依變數: Previously defaulted (參考類別 = No)
模式: (截距), ed, age, employ, address, income, debtinc, creddebt, othdebt
a. 計算中所使用的因子與共變量固定為以下值: Level of education=Post-undergraduate degree; Age in years=34.19; Years with current employer=6.99; Years at current address=6.32; Household income in thousands=60.1581; Debt to income ratio (x100)=9.9341; Credit card debt in thousands=1.9764; Other debt in thousands=3.9164

這個表格會顯示在共變量負債與收入比率中之單位變更的先前已拖欠之 odds 比率。報告的數值為負債/收入比率為 10.9341 之人的拖欠 odds 比率，與負債/收入比率為 9.9341（平均數）之人的拖欠 odds 比率之比較。

請注意，因為這些預測值中沒有一個是交互作用項的一部份，這些表格中所報告之 odds 比率的數值會等於指數化的參數估計值之數值。當預測值為交互作用項的一部份時，這些表格中所報告之其 odds 比率也將以組成交互作用之其他預測值的數值為依據。

摘要

您已經使用「複合樣本 Logistic 迴歸程序」，建構預測指定客戶貸款拖欠之機率模式。

放款人員的重要問題是「類型一」與「類型二」之錯誤成本。也就是將拖欠者歸類為非拖欠者（類型一）的成本是多少？將非拖欠者歸類為拖欠者（類型二）的成本是多少？如果以呆帳作為主要考量，則您要降低「類型一」錯誤並且將您的**敏感度**最大化。如果以客戶數量的成長為優先考量，則您要降低「類型二」錯誤，並將您的**明確性**最大化。通常以上兩項都是主要考量，因此您必須選擇並決定一個分類的規則，此規則能讓敏感度與明確性有最佳的組合。

相關程序

在依據複合取樣架構抽取觀察值時，「複合樣本 Logistic 迴歸」程序是將類別變數模式化很好的工具。

- 「複合樣本取樣精靈」用以指定複雜取樣設計規格並取得樣本。「取樣精靈」建立的取樣計劃檔包含拖欠分析計劃，並可在分析以此計劃取得的樣本時，於「計劃」對話方塊中指定。
- 「複合樣本分析準備精靈」可用來為現有複合樣本指定分析指定。您可在分析此計劃相關的樣本時，在「計劃」對話方塊中指定由「取樣精靈」建立的分析計劃檔。
- 「複合樣本一般線性模式」程序可以讓您將尺度反應值模式化。
- 「複合樣本次序迴歸」程序可以讓您將次序反應值模式化。

章節
21

複合樣本次序迴歸

「複合樣本次序迴歸」程序可為以複雜取樣方法所抽取的次序依變數樣本建立預測模式。或者，您也可以要求對子母體進行分析。

使用「複合樣本次序迴歸」來分析調查結果

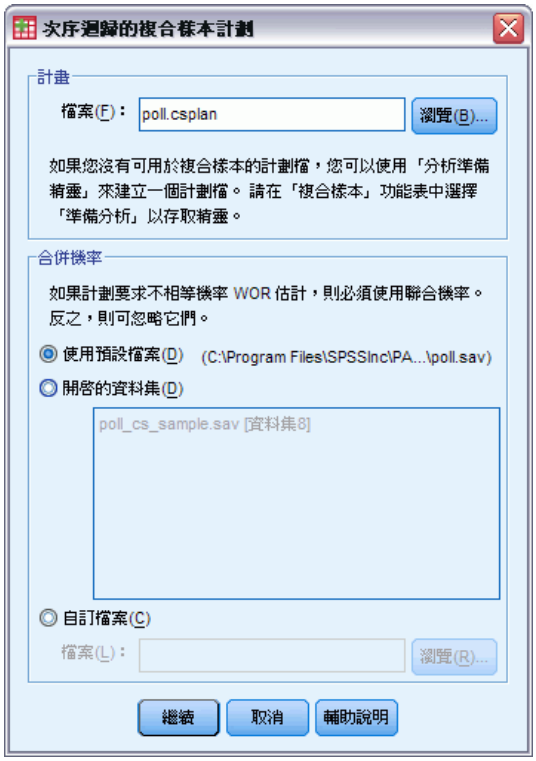
議會代表們在考慮將一項法案交付立法之前，會想知道民眾對該法案的支持度，以及民眾支持度與投票人口的關聯性。民意測驗專家依據複合樣本設計來規劃和進行訪問。

調查結果收集於 poll_cs_sample.sav 中。民意測驗專家的取樣計劃存放在 poll.csplan 中；由於使用到機率 - 比例 - 大小 (PPS) 取樣方式，所以也有一個檔案存放聯合選擇機率 (poll_jointprob.sav)。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 使用「複合樣本次序迴歸」來為投票人口對該法案的支持程度適配出一個模式。

執行分析

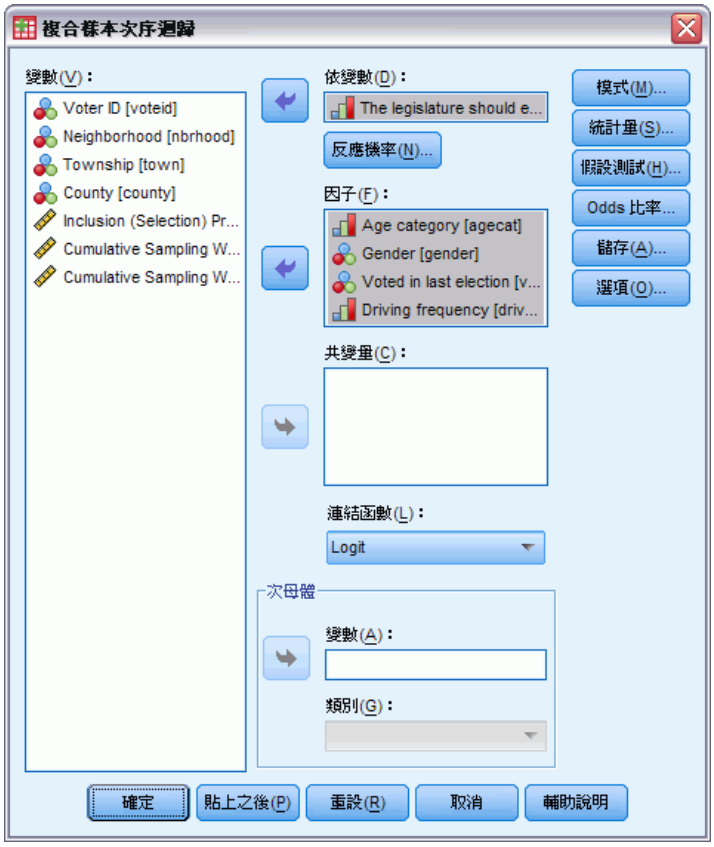
▶ 若要執行「複合樣本次序迴歸」分析，請從功能表選擇：
分析 > 複合樣本 > 次序迴歸...

圖表 21-1
複合樣本計劃對話方塊



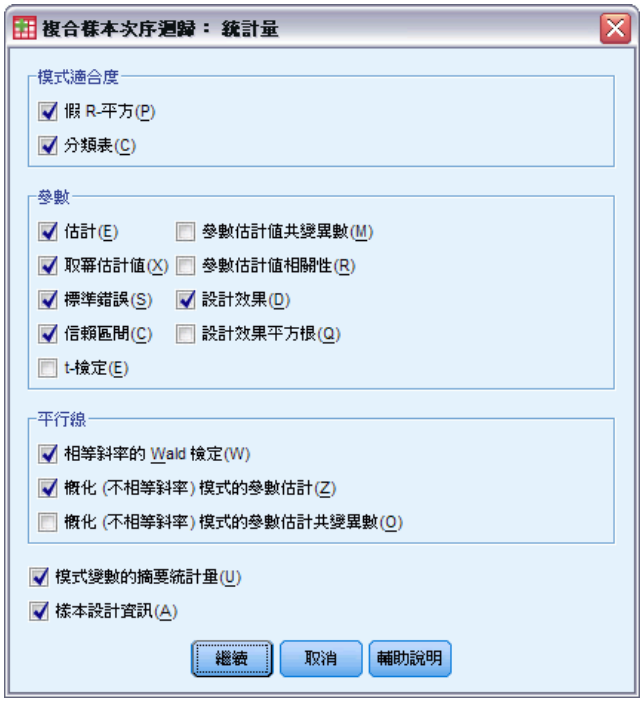
- 瀏覽並選取 poll.csplan 作為計劃檔。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案.
- 選取 poll_jointprob.sav 作為聯合機率檔。
- 按一下「繼續」。

圖表 21-2
「次序迴歸」對話方塊



- 選取「立法機關應制定瓦斯稅」作為依變數。
- 選取「年齡類別」到「駕駛次數」作為因子。
- 按一下「統計量」。

圖表 21-3
「次序迴歸統計量」對話方塊



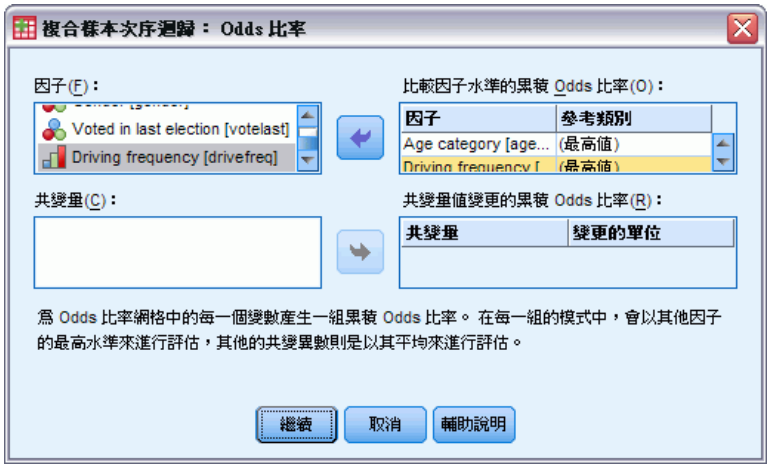
- 選取「模式適合度」組別中的「分類表」。
- 在「參數」組別中，選取「估計」、「指數化估計」、「標準誤」、「信賴區間」和「設計效應」。
- 選取「Wald 斜率相等檢定」和「概化（不相等斜率）模式的參數估計」。
- 按一下「繼續」。
- 在「複合樣本次序迴歸」對話方塊中按一下「假設檢定」。

圖表 21-4
「假設檢定」對話方塊

即使是適量的預測變數和反應類別， 對於平行線的 Wald F 檢定統計量可能還是難以估計的。

- 在「檢定統計量組別」中選取「調整後 F 」。
- 選取「循序 Sidak」作為多重比較的調整方法。
- 按一下「繼續」。
- 在「複合樣本次序迴歸」對話方塊中按一下「Odds 比率」。

圖表 21-5
次序迴歸 Odds 比率對話方塊



- 選擇要產生「年齡類別」和「駕駛次數」的累積 Odds 比率。
- 選取「10-14, 999 哩/年」(一個較最大值更為「一般性」的年哩程數)，作為「駕駛次數」的參考類別。

- ▶ 按一下「繼續」。
- ▶ 在「複合樣本次序迴歸」對話方塊中按一下**確定**。

虛擬 R 平方

圖表 21-6
虛擬 R 平方

Cox和 Snell	.179
Nagelkerke	.191
McFadden	.071

依變數：The legislature should enact a gas taxThe legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, gender, votelast, drivefreq
連結函數\：Logit

在線性迴歸模式中，判別係數 R^2摘要了在依變數中的變異數與預測值（自變數）相關的部分，以較大的 R^2 值表示本模式解釋了較多的變異數，最大到 1。對於具類別依變數的迴歸模式，是無法計算出單一個在線性迴歸模式中具有全部R^2 特性統計量的 R^2，所以改為計算這些近似值。下列方法用來估計判別係數。

- Cox 和 Snell 的 R^2 (Cox 和 Snell, 1989) 是根據本模式的對數概似與基準線模式的對數概似相比較。不過，對於類別結果，它的理論上最大值小於 1，即使在「完美」模式也是如此。
- Nagelkerke 的 R^2 (Nagelkerke, 1991) 是 Cox & Snell R-平方的調整後版本，它調整了統計量的尺度以涵蓋由 0 到 1 的整個範圍。
- McFadden 的 R^2 (McFadden, 1974) 為另一版本，根據只有截距的模式和完全估計模式的對數概似核心。

什麼可構成「良好」的 R^2 值，會隨應用的領域不同而異。雖然這些統計量本身即有暗示性，但在以相同資料來比較競爭模式時最有用。根據本測量，具有最大 R^2 的統計量「最好」。

模式效應的檢定

圖表 21-7
模式效應的檢定

來源	df1	df2	調整過的 Wald F	顯著性	循序 Sidak 顯著性
agecat	2.283	31.966	6.215	.004	.003
gender	1.000	14.000	.046	.834	.834
votelast	1.000	14.000	.076	.787	.787
drivefreq	3.785	52.987	228.015	.000	.000

依變數：The legislature should enact a gas taxThe legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, gender, votelast, drivefreq
連結函數\：Logit

模式中的每一個項目都要檢定其影響是否為 0。顯著值小於 0.05 的項目會含有某些明顯效應。因此，「agecat」和「drivefreq」有助於模式，而其他主效果則不會。在進一步的分析資料時，您可以考量自模式中移除「gender」和「votelast」。

參數估計值

參數估計值的表格會將各個預測值的效果做成摘要。本模式中的係數因連結函數的特性而難以解讀，而共變量係數的正負號和因子水準係數的相對數值卻可提供本模式中預測值效應的重要洞察。

- 對於共變量，正（負）係數表示預測值和結果間的正向（反向）關係。一個增加數值具正係數的共變量對應至存在於一「較高的」累積結果類別中的機率增加。
- 對於因子，一較大係數的因子水準表示存在於一「較高的」累積結果類別中的機率較大。因子水準係數的正負號是根據相對於參考類別的因子水準效應而定。

圖表 21-8
參數估計值

參數		B 之估計值	標準誤差	95% 信賴區間 下界	95% 信賴區間 上界	設計效果	Exp(B)	Exp(B) 的 95% 信賴區間 下界	Exp(B) 的 95% 信賴區間 上界
臨界	[opinion_gastax=1]	-3.343	.104	-3.566	-3.120	1.132	.035	.028	.044
	[opinion_gastax=2]	-1.910	.098	-2.120	-1.700	1.058	.148	.120	.183
	[opinion_gastax=3]	-.674	.090	-.866	-.482	.915	.510	.421	.618
迴歸	[agecat=1]	-.324	.079	-.494	-.154	1.793	.723	.610	.858
	[agecat=2]	-.138	.054	-.255	-.022	1.158	.871	.775	.978
	[agecat=3]	-.095	.076	-.257	.068	2.206	.909	.773	1.070
	[agecat=4]	.000[a]	.	.	.	.	1.000	.	.
	[gender=0]	-.008	.035	-.084	.068	.949	.992	.920	1.071
	[gender=1]	.000[a]	.	.	.	.	1.000	.	.
	[votelast=0]	-.011	.039	-.095	.073	1.103	.989	.909	1.076
	[votelast=1]	.000[a]	.	.	.	.	1.000	.	.
	[drivefreq=1]	-3.751	.153	-4.079	-3.423	1.117	.023	.017	.033
	[drivefreq=2]	-3.003	.116	-3.251	-2.755	1.226	.050	.039	.064
	[drivefreq=3]	-2.295	.114	-2.540	-2.050	1.585	.101	.079	.129
	[drivefreq=4]	-1.570	.092	-1.769	-1.372	1.078	.208	.171	.254
	[drivefreq=5]	-.812	.089	-1.003	-.621	.941	.444	.367	.537
	[drivefreq=6]	.000[a]	.	.	.	.	1.000	.	.

依變數：The legislature should enact a gas taxThe legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, gender, votelast, drivefreq
連結函數\：Logit

a. 設定為零，因為這個參數是冗餘的。

您可以根據參數估計值來作下列解讀：

- 較低年齡類別者會比最高年齡類別者更支持該法案。
- 駕駛次數較少者會比駕駛次數較多者更支持該法案。
- 除了統計量不夠顯著者以外，「gneder」和「votelast」變數的係數顯得比其他的係數小。

設計效應表示這些參數估計值所計算出的某些標準誤，比您從簡單隨機樣本中所取得的標準誤還大，而其他的標準誤則較小。將取樣設計資訊併入分析中是非常重要的，否則，舉例來說，您可能會推論「年齡類別」，「[agecat=3]」的第三水準係數與 0 有顯著不同！

分類

圖表 21-9
類別變數資訊

		加權個數	加權百分比
The legislature should enact a gas tax[a]	Strongly agree	25132.955	21.3%
	Agree	32261.425	27.3%
	Disagree	29477.417	24.9%
	Strongly disagree	31314.203	26.5%
Age category	18-30	20509.504	17.4%
	31-45	35380.506	29.9%
	46-60	34865.792	29.5%
	>60	27430.198	23.2%
Gender	Male	61424.547	52.0%
	Female	56761.453	48.0%
Voted in last election	No	70607.216	59.7%
	Yes	47578.784	40.3%
Driving frequency	Do not own car	3437.137	2.9%
	<10,000 miles/year	10816.349	9.2%
	10-14,999 miles/year	32539.364	27.5%
	15-19,999 miles/year	39179.814	33.2%
	20-29,999 miles/year	25617.804	21.7%
	>=30,000 miles/year	6595.532	5.6%
The legislature should enact a gas tax	母體大小	118186.000	100.0%

a. 以遞增順序排序依變數值。

如果有觀察的資料，「零階」模式（也就是沒有預測值的模式）會將所有的客戶分類到典型組別同意。因此，零階模式會有 27.3% 是正確的。

圖表 21-10
分類表

觀察次數	預測次數				
	Strongly agree	Agree	Disagree	Strongly disagree	百分比修正
Strongly agree	7067.567	12130.814	3875.825	2058.750	28.1%
Agree	4271.234	14464.286	7320.767	6205.137	44.8%
Disagree	2024.816	11703.368	7108.487	8640.746	24.1%
Strongly disagree	889.869	8169.109	6946.522	15308.703	48.9%
整體百分比	12.1%	39.3%	21.4%	27.3%	37.2%

依變數：The legislature should enact a gas taxThe legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, gender, votelast, drivefreq
連結函數\：Logit

分類表顯示使用該模式的實際結果。對於每個觀察值，具有最高模式預測機率的反應類別，即為預測反應。觀察值是以「最終取樣權重」來進行加權，因此類別表會報告母群體的期望模式效能。

- 對角線上的儲存格為正確預測。
- 不在對角線上的儲存格為不正確預測。

此模式將觀察值以超過 9.9%，或是 37.2% 的正確性分類。特別是，此模式在為「同意」或「非常不同意」者作分類時明顯較佳，而對於「不同意」者則略差。

Odds 比率

累積 Odds 定義為依變數的值小於或等於一指定反應類別的機率，與它的值大於該反應類別的機率間之比率。**累積 Odds 比率**為不同預測變數值的累積 odds 之比率，並與指數參數估計值密切相關。有趣的是，累積 odds 比率本身並不依反應類別而定。

圖表 21-11
年齡類別的累積 odds 比率

		累積 Odds 比率	95% 信賴區間		設計效果	平方根設計效果
			下界	上界		
Age category	18-30 vs. >60	1.383	1.166	1.639	1.793	1.339
	31-45 vs. >60	1.148	1.022	1.290	1.158	1.076
	46-60 vs. >60	1.100	.935	1.294	2.206	1.485

依變數：The legislature should enact a gas taxThe legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, gender, votelast, drivefreq
連結函數\：Logit

a. 計算中所使用的因子與共變量固定爲以下值: Age category=>60; Gender=Female; Voted in last election=Yes; Driving frequency=>=30,000 miles/year

本表格顯示「年齡類別」因子水準的累積 odds 比率。所報告的數值即為18 - 30 到 46 - 60 的累積 odds 與>60 的累積 odds 相比之比率。因此，在表內第一行 1.383 的 odds 比率表示對於 18 - 30 歲者的累積 odds 是超過 60 歲者的累積 odds 的 1.383 倍。請注意，由於「年齡類別」並非交互作用項，該 odds 比率只是指數參數估計值的比率。例如，18 - 30與>60 的累積 odds 比率為 1.00/0.723 = 1.383。

圖表 21-12
駕駛次數的 odds 比率

		累積 Odds 比率	95% 信賴區間		設計效果	平方根設計效果
			下界	上界		
Driving frequency	Do not own car vs. 10-14,999 miles/year	4.288	2.878	6.390	2.345	1.531
	<10,000 miles/year vs. 10-14,999 miles/year	2.030	1.656	2.488	1.838	1.356
	15-19,999 miles/year vs. 10-14,999 miles/year	.484	.430	.546	1.450	1.204
	20-29,999 miles/year vs. 10-14,999 miles/year	.227	.193	.267	2.095	1.448
	>=30,000 miles/year vs. 10-14,999 miles/year	.101	.079	.129	1.585	1.259

依變數：The legislature should enact a gas taxThe legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, gender, votelast, drivefreq
連結函數\：Logit

a. 計算中所使用的因子與共變量固定為以下值: Age category=>60; Gender=Female; Voted in last election=Yes; Driving frequency=>=30,000 miles/year

本表格顯示「 駕駛次數」因子水準的累積 odds 比率。使用10 – 14,999 哩/年作為參考類別。由於「駕駛次數」並非交互作用項，該 odds 比率只是指數參數估計值的比率。例如，20 – 29,999 哩/年與 >10 – 14,999 哩/年 的累積 odds 比率為 0.101/0.444 = 0.227。

概化累積模式

圖表 21-13
平行線檢定

df1	df2	調整過的 Wald F	顯著性	循序 Sidak 顯著性
8.769	122.767	1.894	.061	.392

依變數：The legislature should enact a gas taxThe legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, gender, votelast, drivefreq
連結函數\：Logit

此平行線檢定可供您評估「所有反應類別的參數皆相同」的假設是否合理。本檢定將所有類別用同一組係數的估計模式與每一個類別有個別組係數的概化模式作比較。

Wald F 檢定為一可提供漸進式正確 p 值平行線假設對比矩陣的總括檢定；對於小型至中型樣本，調整後 Wald F 統計量很好用。顯著值接近 0.05 表示此概化模式可能有助於改善模式適合度；但循序 Sidak 調整後檢定所記錄的顯著值高達 0.392，因此，整體而言並沒有清楚的證據來推翻平行線假設。「循序 Sidak」檢定由個別的對比 Wald 檢定開始，以提供一整體的 p 值，而這些結果應與總括 Wald 檢定的結果作比較。在此範例中，它們有如此大差異的事實的確有點驚人，但可能是由於在檢定中存在許多對比，相對上卻又只有很小的設計自由度。

圖表 21-14
概化累積模式的參數估計（部分顯示）

The legislature should ena…	參數	B 之估計值	標準誤差	95% 信賴區間	
				下界	上界
Strongly agree	(臨界)	-3.681	.221	-4.155	-3.207
	[agecat=1]	-.320	.096	-.525	-.115
	[agecat=2]	-.075	.071	-.227	.077
	[agecat=3]	-.022	.073	-.180	.135
	[agecat=4]	.000[a]	.		
	[gender=0]	-.082	.054	-.197	.033
	[gender=1]	.000[a]	.		
	[votelast=0]	.008	.052	-.104	.120
	[votelast=1]	.000[a]	.		
	[drivefreq=1]	-4.096	.267	-4.669	-3.523
	[drivefreq=2]	-3.367	.237	-3.876	-2.857
	[drivefreq=3]	-2.678	.224	-3.158	-2.199
	[drivefreq=4]	-1.928	.213	-2.384	-1.471
	[drivefreq=5]	-1.015	.252	-1.555	-.476
	[drivefreq=6]	.000[a]	.		
Agree	(臨界)	-1.963	.153	-2.291	-1.635
	[agecat=1]	-.385	.095	-.587	-.182
	[agecat=2]	-.130	.069	-.279	.018
	[agecat=3]	-.139	.101	-.356	.077
	[agecat=4]	.000[a]	.		
	[gender=0]	-.004	.040	-.090	.082
	[gender=1]	.000[a]	.		
	[votelast=0]	.009	.059	-.117	.135
	[votelast=1]	.000[a]	.		
	[drivefreq=1]	-3.867	.318	-4.549	-3.185
	[drivefreq=2]	-3.005	.175	-3.380	-2.630
	[drivefreq=3]	-2.290	.187	-2.691	-1.888
	[drivefreq=4]	-1.633	.166	-1.988	-1.278
	[drivefreq=5]	-.909	.137	-1.204	-.615
	[drivefreq=6]	.000[a]	.		

此外，概化模式係數的估計值看起來與平行線假設而得的估計值差別不大。

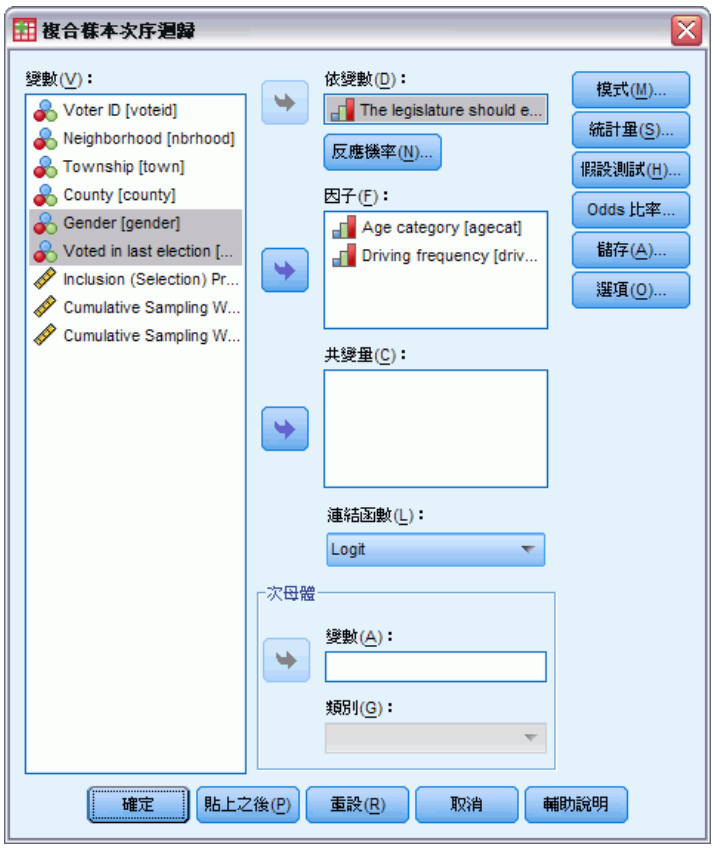
剔除非顯著預測值

模式效應的檢定顯示「性別」和「上次選舉有去投票」的模式係數在統計上未與 0 有顯著不同。

- 若要產生一減少模式，請叫回「複合樣本次序迴歸」對話方塊。

▶ 在「計劃」對話方塊中按一下「繼續」。

圖表 21-15
「次序迴歸」對話方塊



▶ 取消選取「性別」和「上次選舉有去投票」作為因子。

▶ 按一下「選項」。

圖表 21-16
「次序迴歸選項」對話方塊

- 選取「顯示疊代歷程」。

  在診斷估計演算法所遭遇的問題時，疊代歷程非常有用。

- 按一下「繼續」。
- 在「複合樣本次序迴歸」對話方塊中按一下確定。

警告

圖表 21-17
為減少模式作出的警告

在步驟減半方法的步驟上限之後，已無法再增加對數相似度。

儘管出現以上警告，CSORDINAL 程序還是繼續進行。隨後顯示的結果是以最後一個疊代爲準。不確定模型配適的有效性。

下列訊息適用於概化累積模式。

在步驟減半方法的步驟上限之後，已無法再增加對數相似度。

警告會指出減少模式要在參數估計達到收斂條件前結束估計，因為目前參數估計值的對數概似不會增加任何變更或「步驟」。

圖表 21-18
為減少模式作出的警告

疊代編號[a]	N 分享步驟	虛擬 -2 對數概似	臨界			迴歸							
			[opinion_gastax=1]	[opinion_gastax=2]	[opinion_gastax=3]	[agecat=1]	[agecat=2]	[agecat=3]	[drivefreq=1]	[drivefreq=2]	[drivefreq=3]	[drivefreq=4]	[drivefreq=5]
0	0	326640.341	-1.309	-.058	1.020	.000	.000	.000	.000	.000	.000	.000	.000
1	0	303567.549	-3.242	-1.881	-.704	-.323	-.137	-.094	-3.841	-2.970	-2.248	-1.563	-.835
2	0	303336.336	-3.327	-1.897	-.664	-.325	-.139	-.095	-3.740	-2.998	-2.291	-1.568	-.811
3	0	303335.933	-3.333	-1.900	-.664	-.326	-.139	-.096	-3.750	-3.003	-2.295	-1.570	-.812
4	0	303335.933	-3.333	-1.900	-.664	-.326	-.139	-.096	-3.750	-3.003	-2.295	-1.570	-.812
5[b]	5	303335.933	-3.333	-1.900	-.664	-.326	-.139	-.096	-3.750	-3.003	-2.295	-1.570	-.812

不顯示重複的參數。在所有疊代中，他們的值都為零。
依變數\：The legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, drivefreq
連結函數\： Logit

a. 使用 Newton-Raphson 方法估計參數。

b. 在步驟減半方法的步驟上限之後，已無法再增加對數相似度。

查看疊代歷程，參數估計值對最後幾次疊代的變更相當細微，您不必太在意該警告訊息。

比較模式

圖表 21-19
「虛擬 R 平方」用於減少模式

Cox 和 Snell	.179
Nagelkerke	.191
McFadden	.071

依變數：The legislature should enact a gas taxThe legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, gender, votelast, drivefreq
連結函數\： Logit

用於減少模式的「R^2」數值與原始模式所用者相同。這對減少模式是有利的證據。

圖表 21-20
分類表用於減少模式

	預測次數				
觀察次數	Strongly agree	Agree	Disagree	Strongly disagree	百分比修正
Strongly agree	7067.567	12823.258	3183.380	2058.750	28.1%
Agree	4271.234	15684.090	6100.963	6205.137	48.6%
Disagree	2024.816	13157.809	5654.047	8640.746	19.2%
Strongly disagree	889.869	9226.578	5889.053	15308.703	48.9%
整體百分比	12.1%	43.1%	17.6%	27.3%	37.0%

依變數：The legislature should enact a gas taxThe legislature should enact a gas tax (遞增)
模式\：(臨界), agecat, gender, votelast, drivefreq
連結函數\： Logit

分類表會使事情稍微複雜化。減少模式 37.0% 的整體分類率與原始模式差不多，這對減少模式是有利的證據。不過，減少模式將 3.8% 選民的預測反應由「不同意」移轉為「同意」，超過觀察到反應「不同意」或「非常不同意」者的半數。這是很重要的差異，使得在選擇減少模式前需要審慎考量。

摘要

使用「複合樣本次序迴歸程序」，您已建立了投票人口對提議法案支持程度的競爭模式。平行線檢定顯示概化累積模式是不必要的。模式效應的檢定指示「性別」和「上次選舉有去投票」可自本模式中剔除，減少模式在虛擬R^2上表現良好，而整體的分類率與原始模式差不多。不過，減少模式在跨越「同意」/「不同意」分割處時錯誤分類出較多的選民，所以議員們目前會寧願保留原始模組。

相關程序

在依據複合取樣架構抽取觀察值時，「複合樣本次序迴歸」程序是將次序變數模式化很好的工具。

- 「複合樣本取樣精靈」用以指定複雜取樣設計規格並取得樣本。「取樣精靈」建立的取樣計劃檔包含預設分析計劃，並可在分析以此計劃取得的樣本時，於「計劃」對話方塊中指定。
- 「複合樣本分析準備精靈」可用來為現有複合樣本指定分析指定。您可在分析此計劃相關的樣本時，在「計劃」對話方塊中指定由「取樣精靈」建立的分析計劃檔。
- 「複合樣本一般線性模式」程序可以讓您將尺度反應值模式化。
- 「複合樣本 Logistic 迴歸」程序可以讓您將類別反應值模式化。

章節
22

複合樣本 Cox 迴歸

「複合樣本 Cox 迴歸」程序對以複合取樣方法所抽取的樣本執行存活分析。

在「複合樣本 Cox 迴歸」中使用「依時預測值」

政府法令執行機構會關心其轄區內的再犯率。其中一項測量再犯的方法，就是依據違法者第二次被補以前的時間。機構想要將再度被補時間在以複合取樣方法抽取的樣本上，採用「Cox 迴歸」來建立模式，但又耽心成比例風險假設在所有年齡類別間會無效。

在取樣的部門中選出在 2003 年六月間所釋放的第一次被補的犯人，以及檢閱他們在 2006 年六月底前的案件歷史。該樣本收集於recidivism_cs_sample.sav中。用到的取樣計劃存放在 recidivism_cs.csplan 中；由於使用到機率 - 比例 - 大小 (PPS) 取樣方式，所以也有一個檔案存放聯合選擇機率 (recidivism_cs_jointprob.sav)。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 使用「複合樣本 Cox 迴歸」來評估成比例風險假設的有效性，並在適當的時機配適出有依時預測值的模式。

準備資料

資料集包含第一次被補的釋放日期和第二次被補的日期；由於 Cox 迴歸是分析存活時間，故您需要計算這些日期之間的時間量。

不過，「第二次被補日期 [date2]」包含的觀察值有 10/03/1582 值 (日期變數的遺漏值)。有些人沒有第二次犯案，我們當然要將他們納入為本模式中正確受限的觀察值。追蹤期間的結尾是 2006 年 六月 30 日，所以我們要將 10/03/1582 重新編碼為 06/30/2006。

- 若要將這些數值重新編碼，請從功能表選擇：
  轉換 > 計算變數...

圖表 22-1
計算變數對話方塊

- 輸入 date2 作為目標變數。
- 輸入 DATE.DMY(30,6,2006) 作為表示式。
- 按一下「如果」。

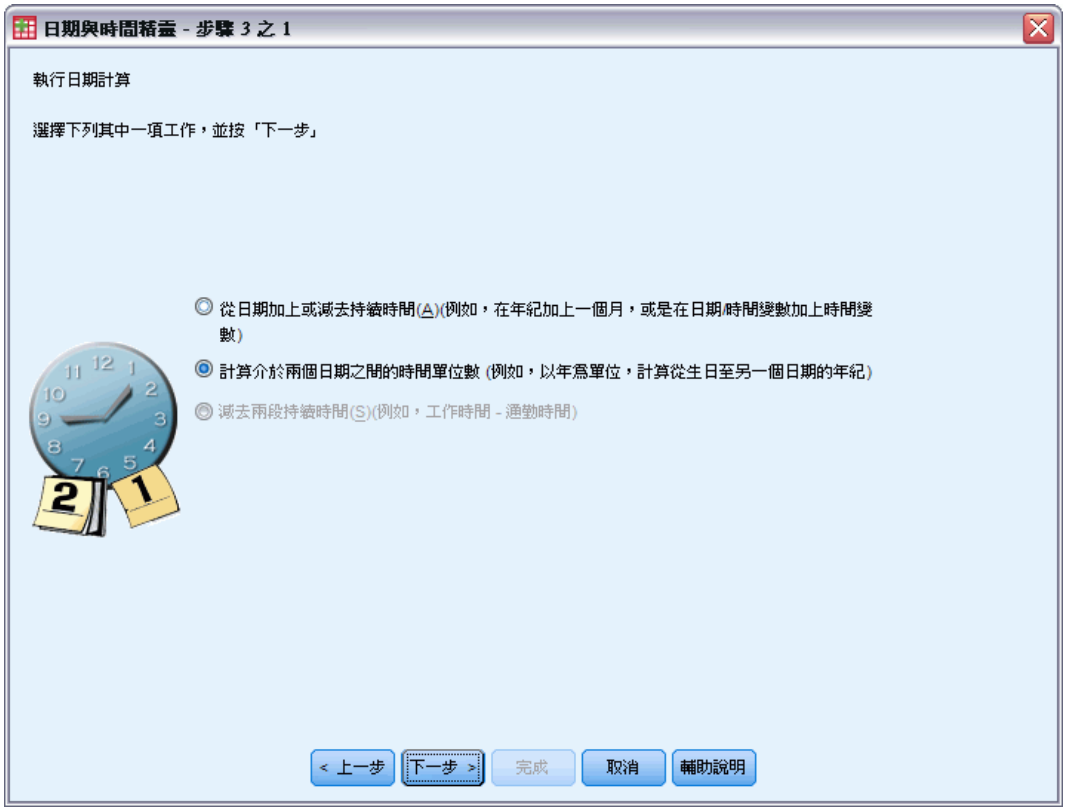
圖表 22-2
「計算變數」和「觀察值選擇條件」對話方塊

- 選取**包含滿足條件的觀察值**。
- 輸入 MISSING(date2) 作為表示式。
- 按一下「繼續」。
- 在「計算變數」對話方塊中，按一下「確定」。
- 接著，要計算第一次和第二次被補的時間間隔，請由功能表中選擇：
  轉換 > 日期和時間精靈...

圖表 22-3
「日期和時間精靈」的「歡迎」步驟

- 選取「使用日期和時間計算」。
- 按一下「下一步」。

圖表 22-4
「日期和時間精靈」的「作日期計算」步驟

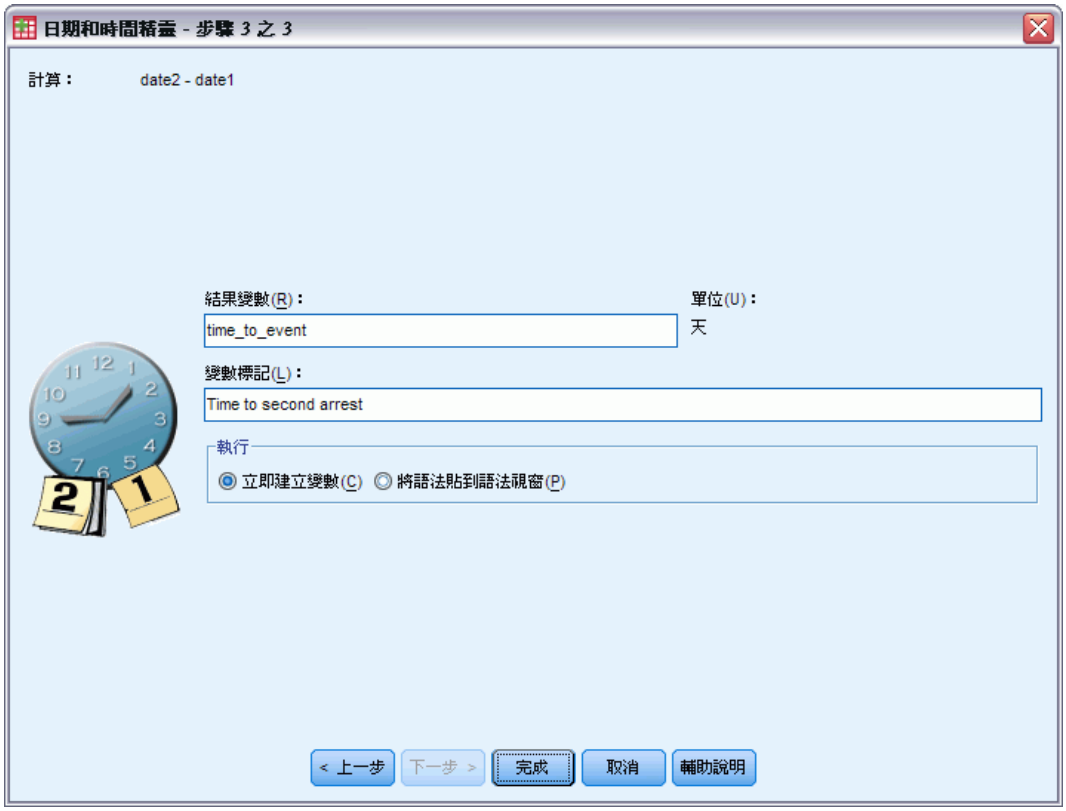


- 選取「計算兩個日期之間的時間單位」。
- 按一下「下一步」。

圖表 22-5
「日期和時間精靈」的「計算兩個日期之間的時間單位」步驟

- 選取「第二次被補日期 [date2]」作為第一個日期。
- 選取「第一次被補的釋放日期 [date1]」作為要從第一個日期減去的日期。
- 選取 **日數** 作為單位。
- 按一下「**下一步**」。

圖表 22-6
「日期和時間精靈」的「計算」步驟

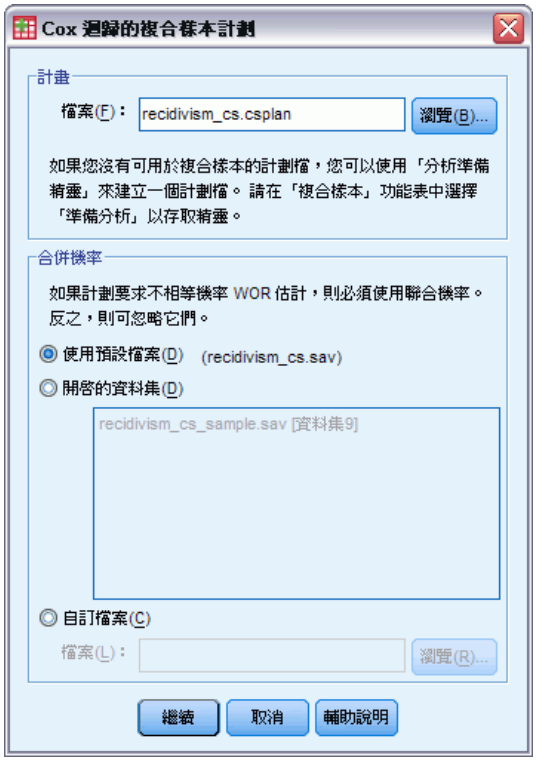


- 輸入 time_to_event 作為變數名稱，代表這兩個日期間隔的時間。
- 輸入「到第二次被補的時間」作為變數標記。
- 按一下「**完成**」。

執行分析

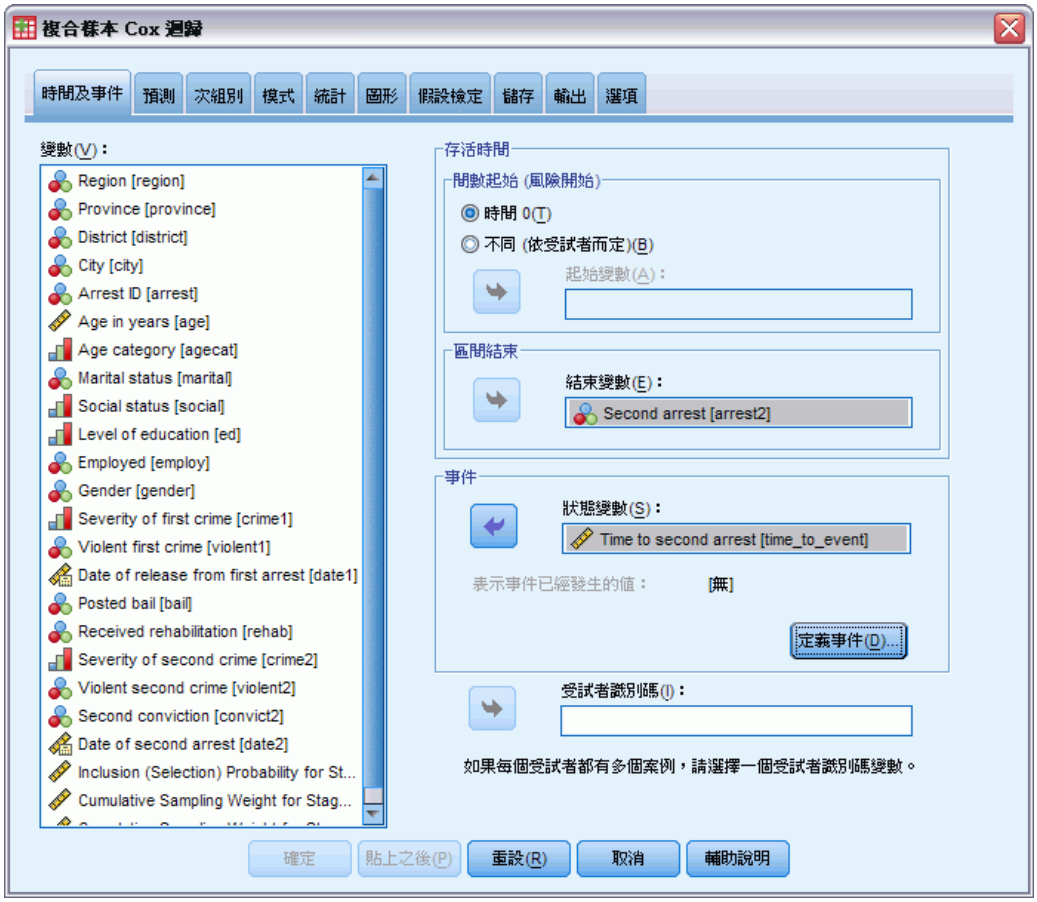
- 若要執行「複合樣本 Cox 迴歸」分析，請從功能表選擇：
  分析 > 複合樣本 > Cox 迴歸...

圖表 22-7
「Cox 迴歸」的「複合樣本計劃」對話方塊



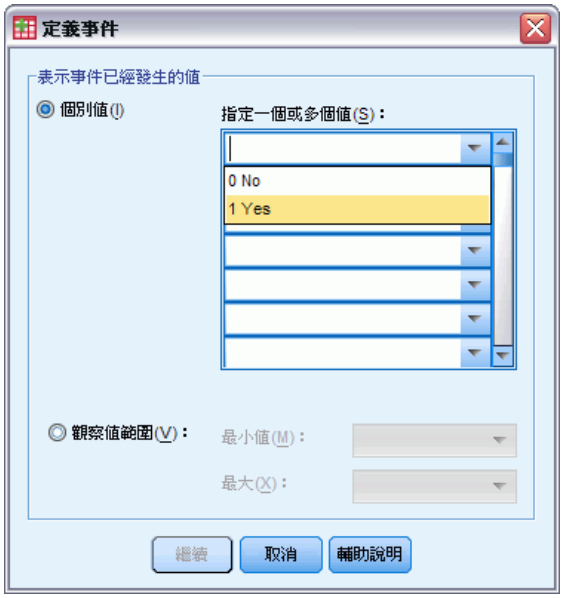
- 瀏覽到樣本檔目錄並選取 recidivism_cs.csplan 作為計劃檔。
- 在「聯合機率」組別中選取「**自訂檔案**」，瀏覽到樣本檔目錄並選取 recidivism_cs_jointprob.sav。
- 按一下「繼續」。

圖表 22-8
「Cox 迴歸」對話方塊的「時間和事件」索引標籤



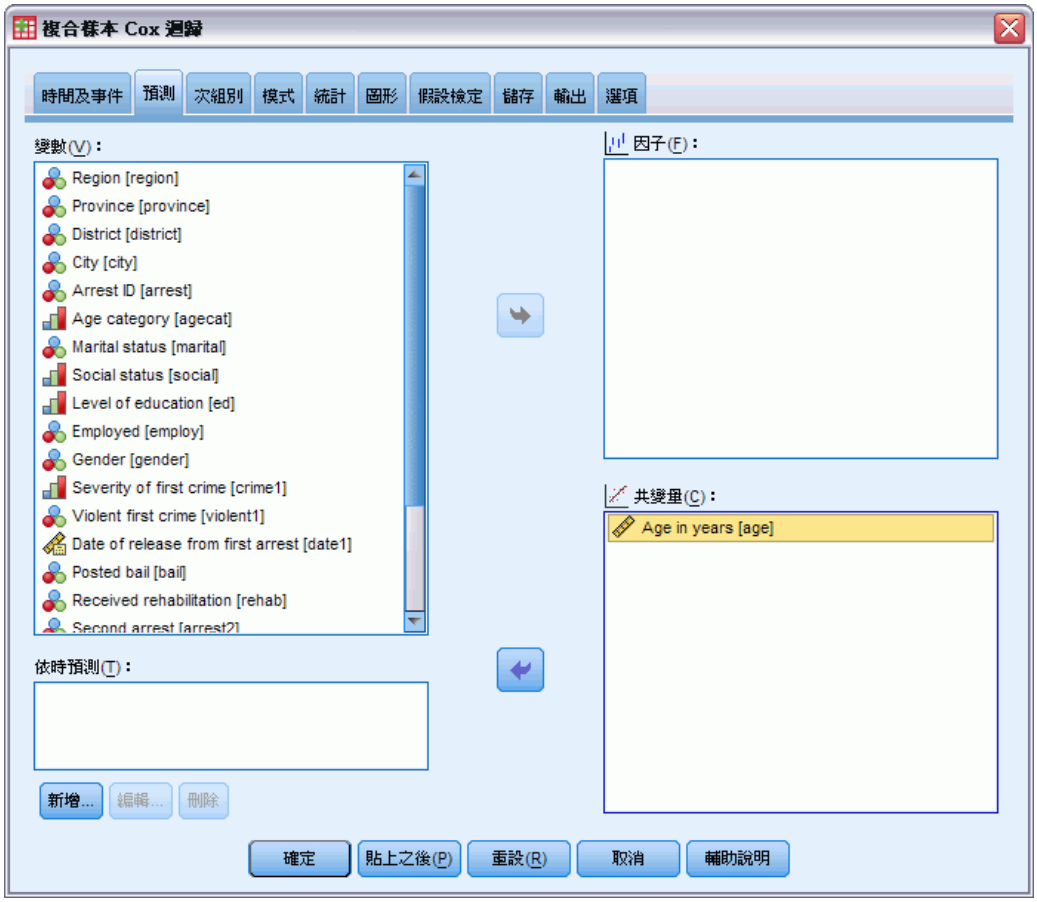
- 選取「到第二次被補的時間 [time_to_event]」作為定義區間終點的變數。
- 選取「第二次被補 [arrest2]」作為定義事件是否發生的變數。
- 按一下「**定義事件**」。

圖表 22-9
「定義事件」對話方塊



- 選取「1 是」作為表示事件（再度被補）發生的值。
- 按一下「繼續」。
- 按一下「預測值」索引標籤。

圖表 22-10
「Cox 迴歸」對話方塊的「預測值」索引標籤



- 選取「年齡 [age]」作為共變量。
- 按一下「統計量」索引標籤。

圖表 22-11
「Cox 迴歸」對話方塊的「統計量」索引標籤

- 選取「成比例風險假設檢定」，再選取「對數」作為模式假設組別的時間函數。
- 選取「其他模式的參數估計量」。
- 按一下「確定」。

樣本設計資訊

圖表 22-12
樣本設計資訊

			N
未加權計數	有效	主題	5687
		觀察值	5687
	有效觀察值		0
	總觀察值		5687
主題大小			307583.898
階段 1	Strata		4
	Units		20
樣本設計自由度			16

本表包含與模式估計有關的樣本設計資訊。

- 每一個受試者有一個觀察值，所有 5,687 個觀察值都在分析中使用到。
- 該樣本代表不到整個估計母群體的 2%。
- 本設計要求 4 個分層，每分層有 5 個單位，在本設計的第一階段共有 20 個單位。取樣設計自由度估計為 20−4=16。

模式效應的檢定

圖表 22-13
模式效應的檢定

源	df1	df2	Wald F	顯著性
age	1.000	16	504.787	1.580E-13

存活時間變數: Time to second arrest
事件狀態變數: Second arrest = 1.0
模式 : age

在成比例風險模式中，預測值「年齡」的顯著值小於 0.05，看起來有助於本模式。

成比例風險的檢定

圖表 22-14
成比例風險的整體檢定

df1	df2	Wald F	顯著性
1.000	16.000	29.924	5.136E-5

存活時間變數: Time to second arrest
事件狀態變數: Second arrest = 1.0
模式 : age, age*_TF

圖表 22-15
其他模式的參數估計量

			90% 置信間隔	
參數	B	標準差	下界	上界
age	-.002	.014	-.025	0.02
t_age	-.012	.002	-.016	-.009

存活時間變數: Time to second arrest
事件狀態變數: Second arrest = 1.0
模式: age, age*_TF

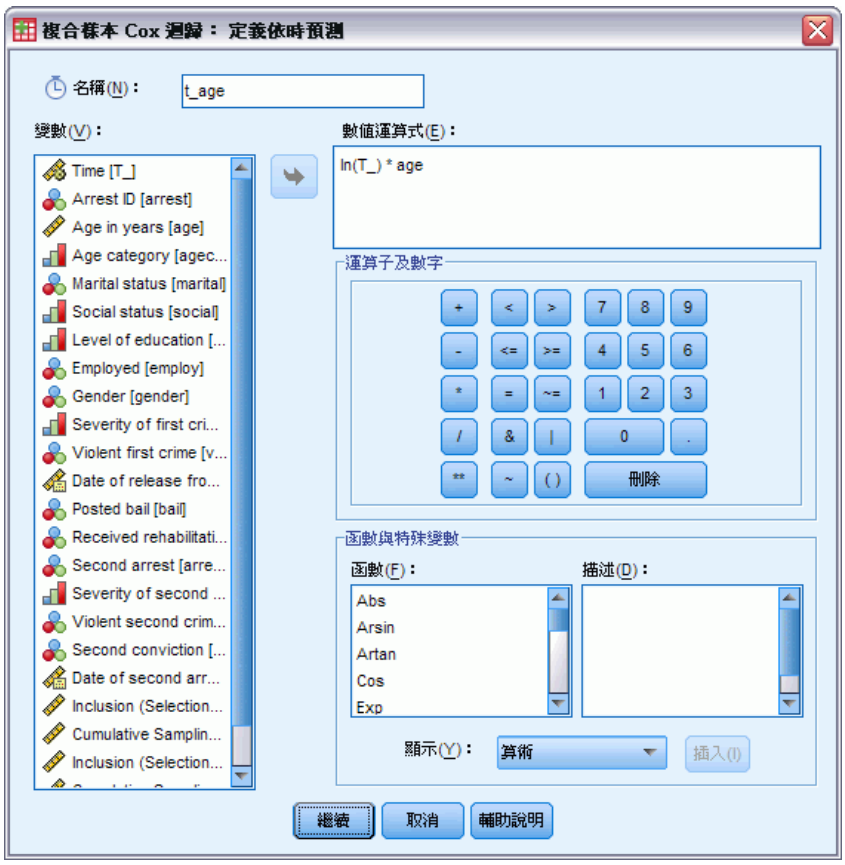
成比例風險整體檢定的顯著值小於 0.05，表示違反了成比例風險假設。其他模式使用了對數時間函數，所以很容易複製此依時預測值。

加入一個依時預測值

▶ 叫回「複合樣本 Cox 迴歸」對話方塊並按一下「**預測值**」索引標籤。

▶ 按一下「新增」。

圖表 22-16
「Cox 迴歸定義依時預測值」對話方塊

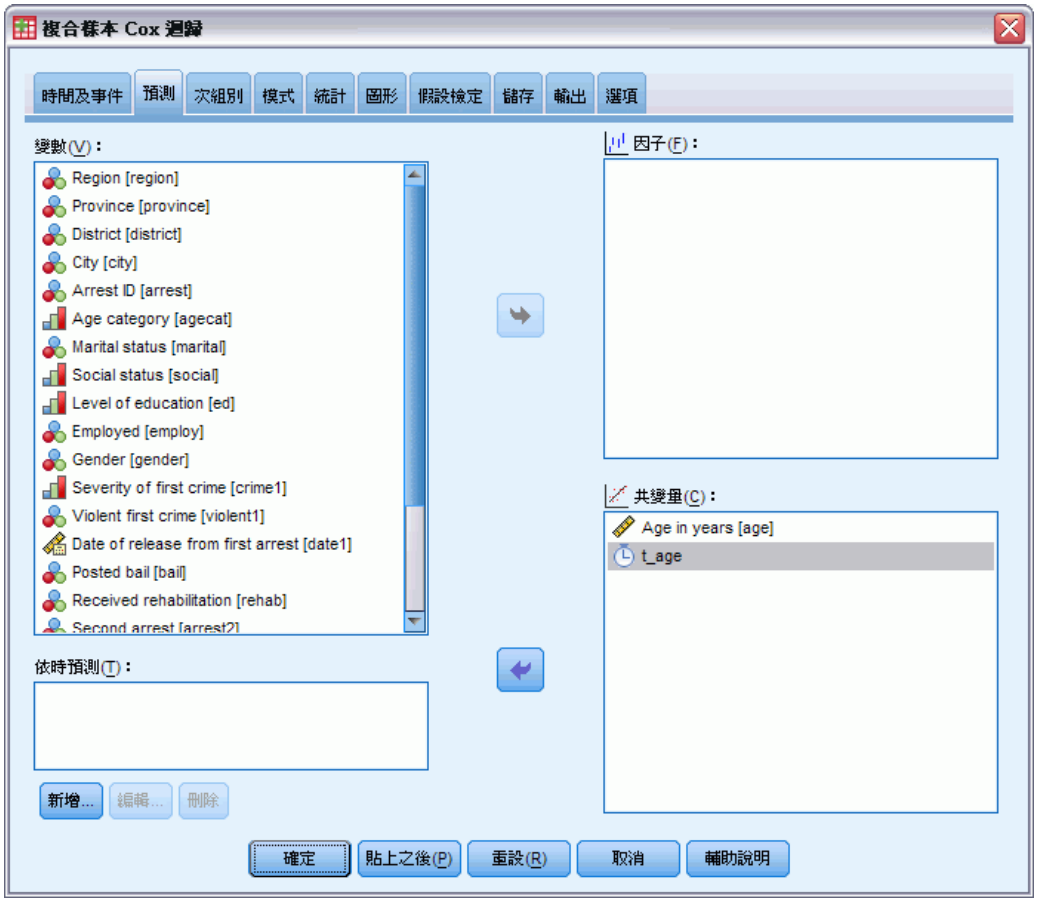


▶ 輸入「t_age」作為您想要定義的依時預測值的名稱。

▶ 輸入 ln(T_)*age 作為數值運算式。

▶ 按一下「繼續」。

圖表 22-17
「Cox 迴歸」對話方塊的「預測值」索引標籤



- 選取「t_age」作為共變量。
- 按一下「統計量」索引標籤。

圖表 22-18
「Cox 迴歸」對話方塊的「預測值」索引標籤

- 在「參數」組別中，選取「估計」、「標準誤」、「信賴區間」和「設計效應」。
- 在「模式假設」組別中，取消選取「成比例風險檢定」和「其他模式的參數估計量」
- 按一下「確定」。

模式效應的檢定

圖表 22-19
模式效應的檢定

源	df1	df2	Wald F	顯著性
age	1.000	16.000	.015	0.91
t_age	1.000	16.000	29.924	5.136E-5

存活時間變數: Time to second arrest
事件狀態變數: Second arrest = 1.0
模式: age, t_age

在加入依時預測值後，「年齡」的顯著值是 0.91，表示其對該模式的貢獻被「t_age」所取代。

參數估計值

圖表 22-20
參數估計值

參數	B	標準差	95% 置信間隔		設計效應
			下界	上界	
age	-.002	0.01	-.030	.027	.702
t_age	-.012	.002	-.017	-.008	.666

存活時間變數: Time to second arrest
事件狀態變數: Second arrest = 1
模式: age, t_age

查看參數估計和標準誤，可看到您已由成比例風險的檢定複製了其他模式。藉著明確指定模式，您可要求其他的參數統計量和圖形。此處，我們要求設計效應；如果您假設資料集為一簡單隨機樣本的話，「t_age」值小於 1 表示「t_age」的標準誤小於您可能得到者。在這種情況下，「t_age」的效應在統計上仍然顯著，但信任區間將會更寬。

在「複合樣本 Cox 迴歸」中每個受試者的多個觀察值

研究人員對於因缺血性中風的病患，其在結束康復計畫後存活時間方面的研究，面臨許多挑戰。

每個受試者有多個觀察值 代表病患醫療史的變數應和預測值一樣很有用。經過一段時間，病患可能經歷改變他們醫療史的主要醫療事件。在這個資料集中，記載了心肌梗塞、缺血性中風、或出血性中風的發生，以及事件記錄的時間。您可以在程序中建立可計算的依時共變量，以將此資訊納入該模式，但若使用每個受試者有多個觀察值將更加方便。請注意，變數原本是編碼的，以便將病患的病歷是記錄在所有的變數中，因此您需要重新架構此資料集。

左側截斷。 風險從發生缺血性中風時開始。但樣本只包含在康復計劃中存活的病患，因此會以樣本左側截斷的方式，表示所觀察的存活時間被康復期所「誇大」。您可以藉著將他們結束康復計劃的時間指定為開始研究的時間來解決此問題。

無取樣計劃 資料集不是經過複合取樣計劃所收集，而是被視為一簡單隨機樣本。您需要建立一個分析計劃來使用「複合樣本 Cox 迴歸」。

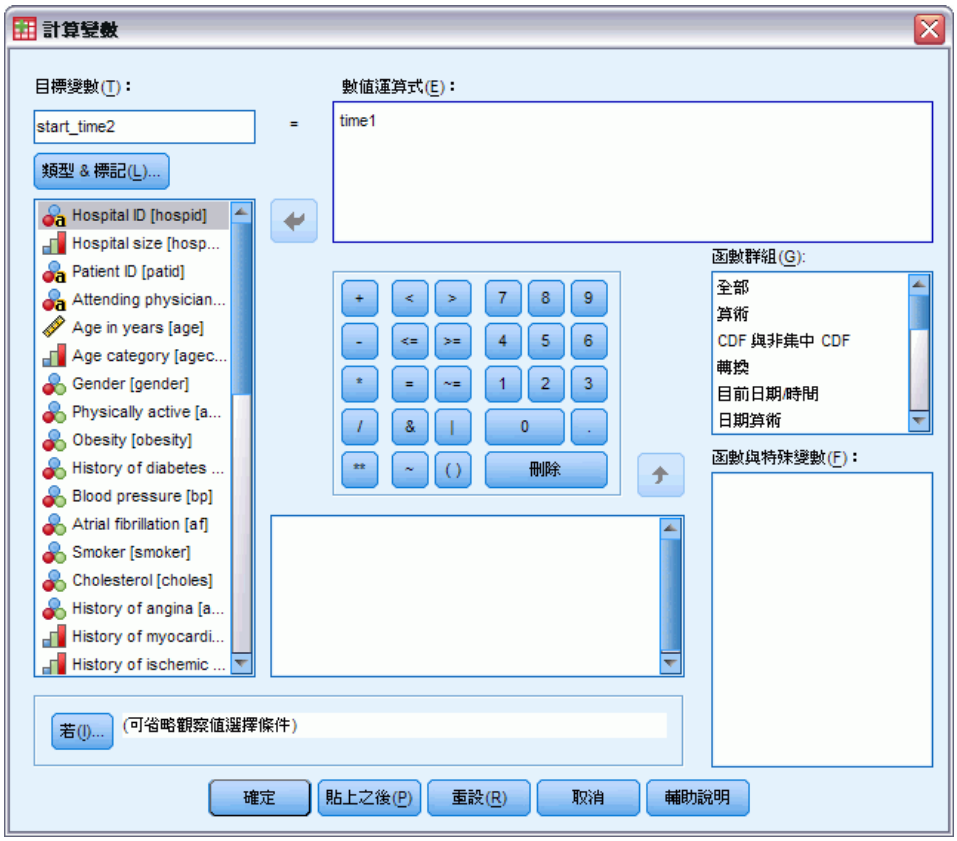
資料集收集在 stroke_survival.sav 中。如需詳細資訊，請參閱第 250 頁附錄 A 中的範例檔案. 運用「重新架構資料精靈」來準備分析資料，接著用「分析準備精靈」來建立一個簡單隨機取樣計劃，最後再用「複合樣本 Cox 迴歸」來建立存活時間的模式。

準備進行分析所用的資料

在重新架構資料前，您需要建立兩個輔助變數來協助重新架構。

▶ 若要計算新的變數，請從功能表選擇：
轉換 > 計算變數...

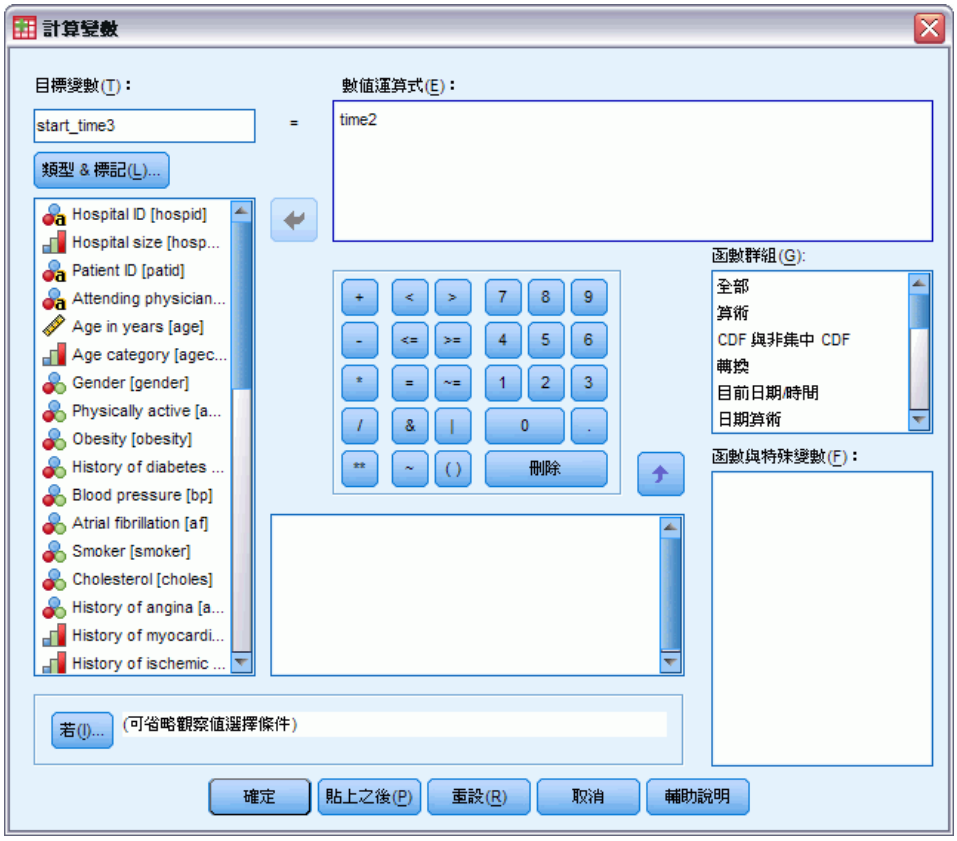
圖表 22-21
計算變數對話方塊



- 輸入「start_time2」作為目標變數。
- 輸入 time1 作為數值運算式。
- 按一下「確定」。

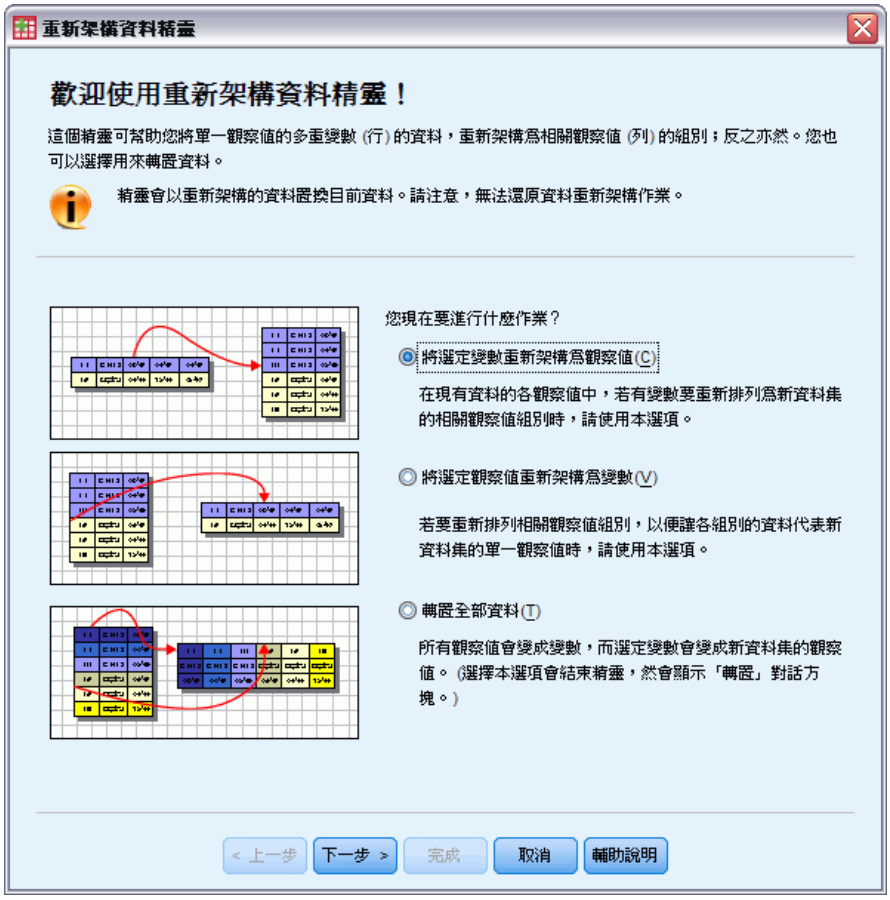
- 叫回「計算變數」對話方塊。

圖表 22-22
計算變數對話方塊



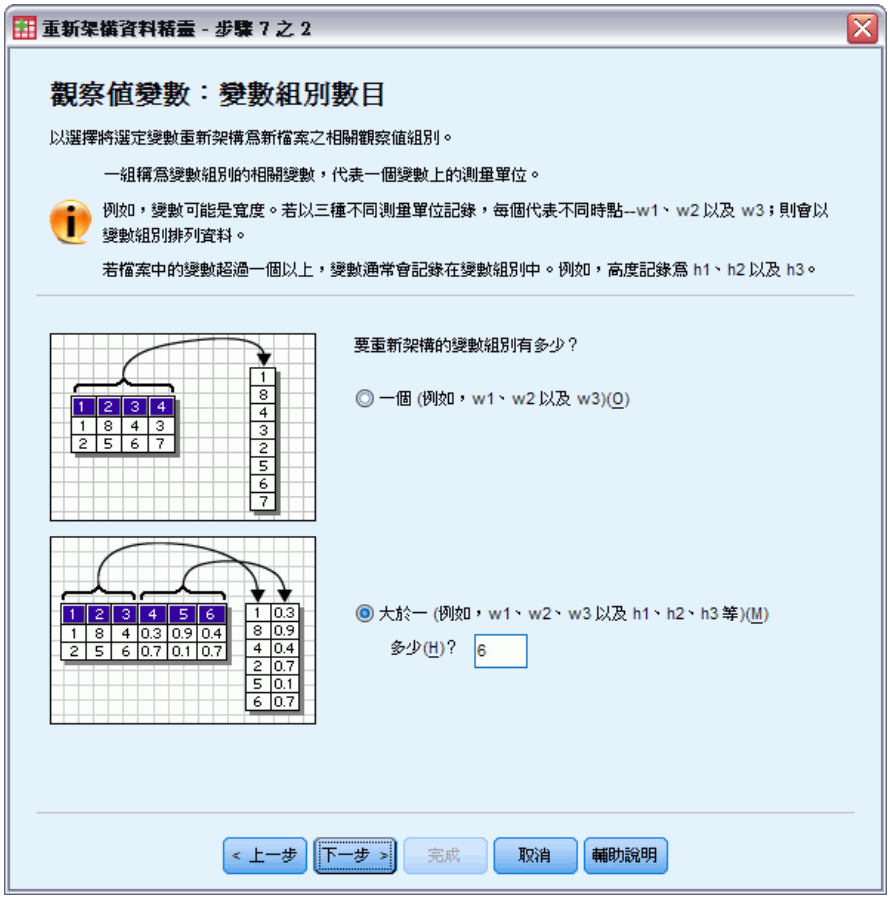
- 輸入「start_time3」作為目標變數。
- 輸入 time2 作為數值運算式。
- 按一下「**確定**」。
- 若要將資料由變數重新架構為觀察值，請從功能表選擇：
  資料 > **重新架構...**

圖表 22-23
「重新架構資料精靈」的「歡迎」步驟



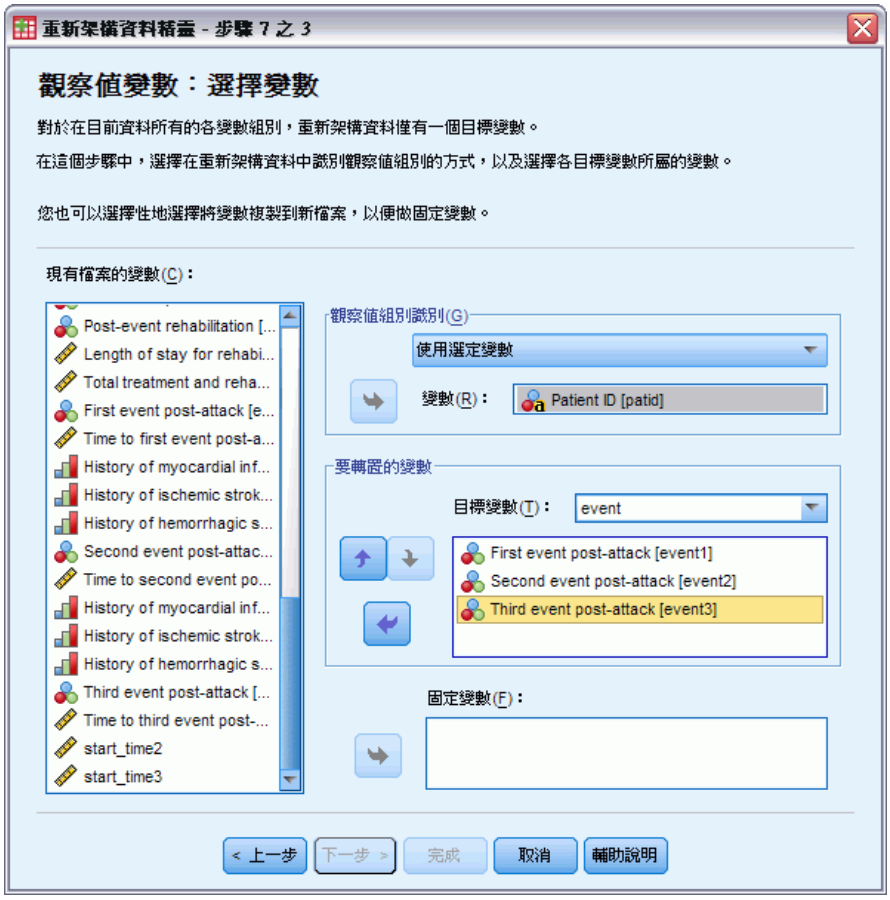
- ▶ 確定已選取「將選取的變數重新架構成為觀察值」。
- ▶ 按一下「下一步」。

圖表 22-24
「重新架構資料精靈」的「變數至觀察值變數組別的個數」步驟



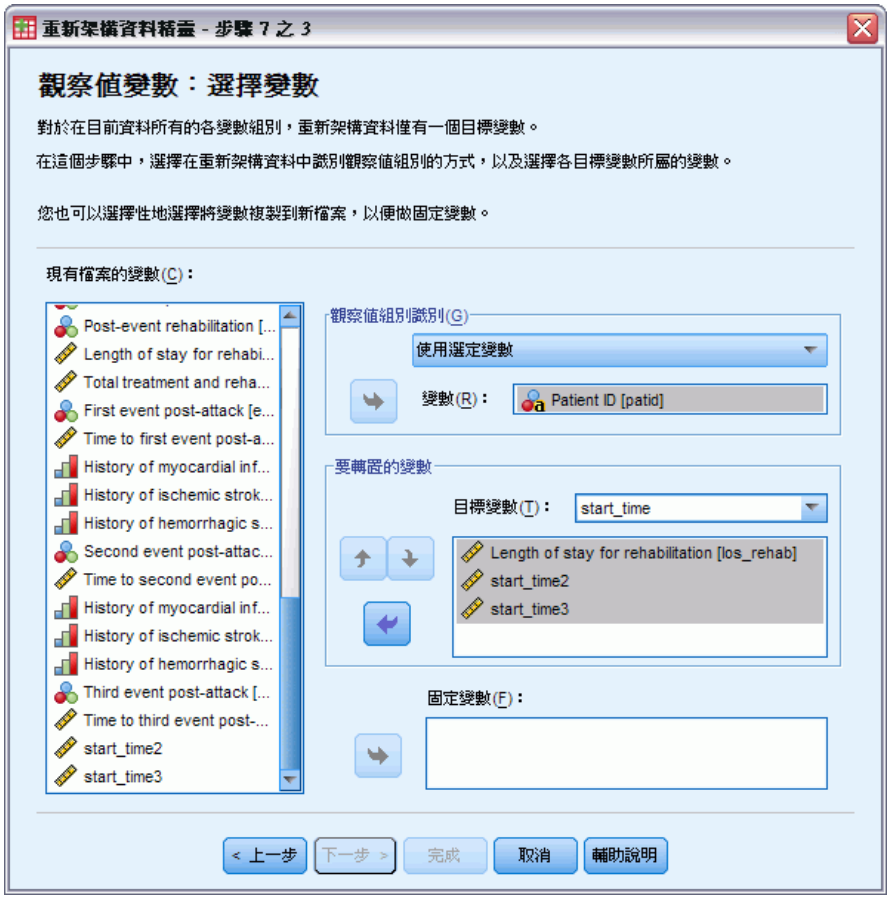
- 選取「一個以上」要重新架構的變數組別。
- 輸入「6」作為組別個數。
- 按一下「下一步」。

圖表 22-25
「重新架構資料精靈」的「變數至觀察值選取變數」步驟



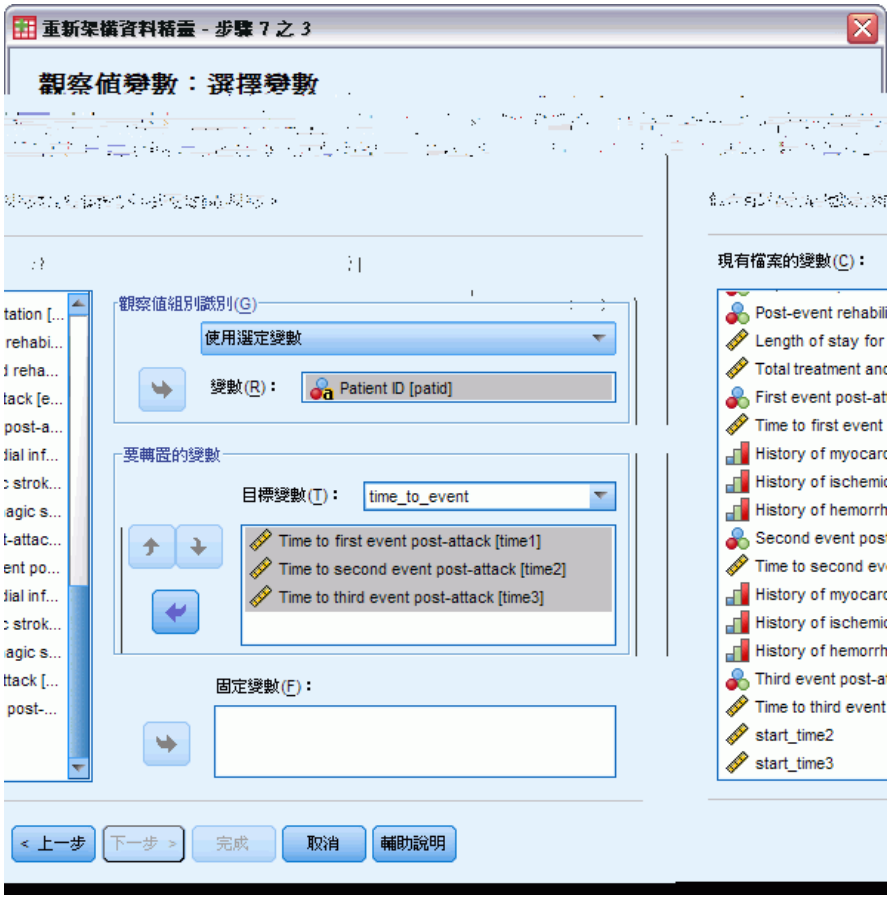
- 在「觀察值組別識別」組別中，選取「**使用選取的變數**」，再選取「病患 ID [patid]」作為受試者識別碼。
- 輸入「**event**」作為第一個目標變數。
- 選取「第一次事件發作後 [event1]」、「第二次事件發作後 [event2]」和「第三次事件發作後 [event3]」作為要轉置的變數。
- 從目標變數清單中選取「trans2」。

圖表 22-26
「重新架構資料精靈」的「變數至觀察值選取變數」步驟



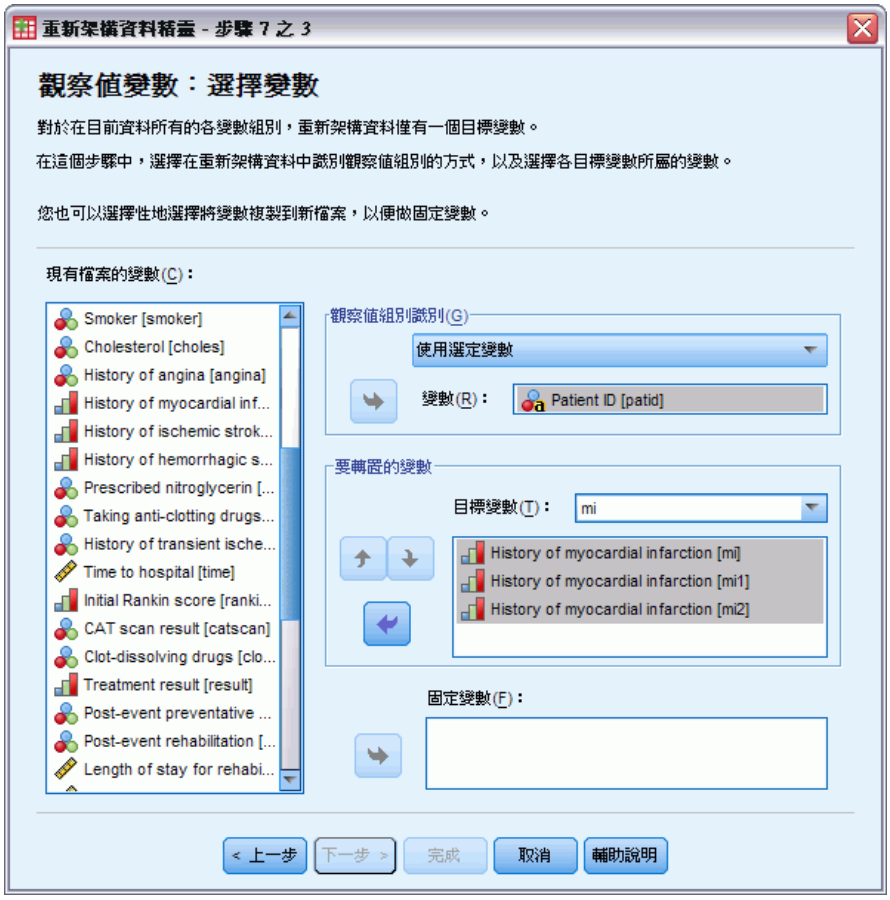
- 輸入「start_time」作為目標變數。
- 選取「康復期的時間長度 [los_rehab]」、「start_time2」和「start_time3」作為要轉置的變數。「至第一次事件發作後時間 [time1]」和「至第二次事件發作後時間 [time2]」將用來建立結束時間，而每一個變數只能出現在一個要轉置的變數清單中，因此「start_time2」和「start_time3」都是必要的。
- 從目標變數清單中選取「trans3」。

圖表 22-27
「重新架構資料精靈」的「變數至觀察值選取變數」步驟



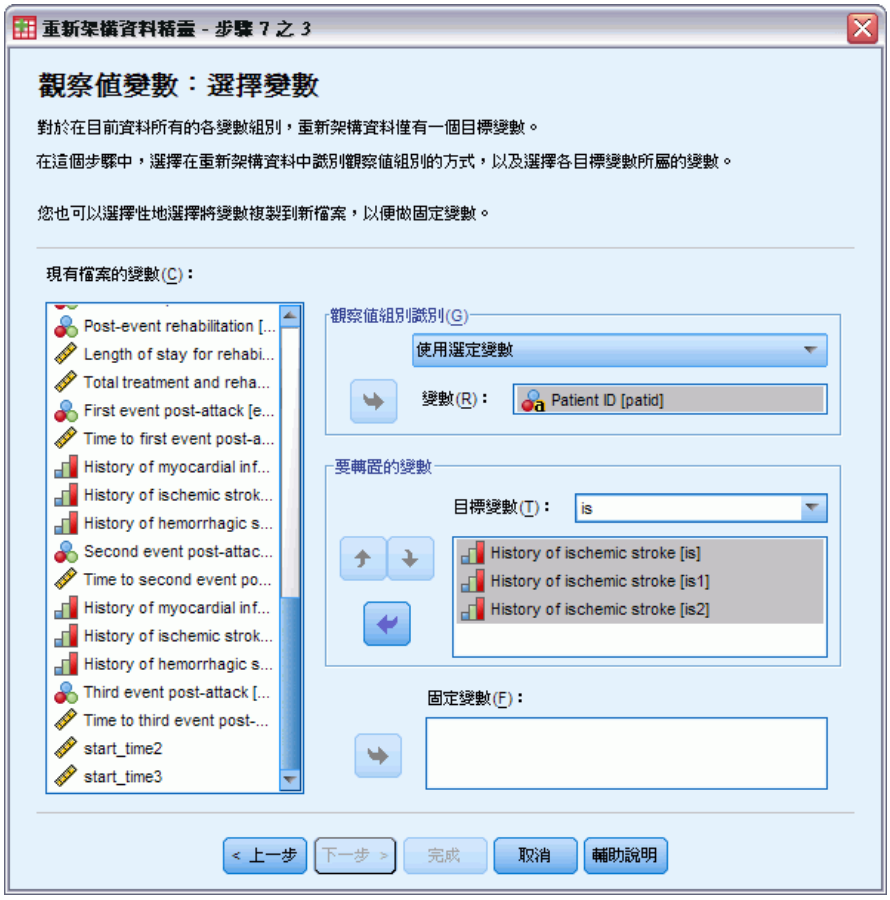
- 輸入「time_to_event」作為目標變數。
- 選取「至第一次事件發作後時間 [time1]」、「至第二次事件發作後時間 [time2]」和「至第三次事件發作後時間 [time 3]」作為要轉置的變數。
- 從目標變數清單中選取「trans4」。

圖表 22-28
「重新架構資料精靈」的「變數至觀察值選取變數」步驟



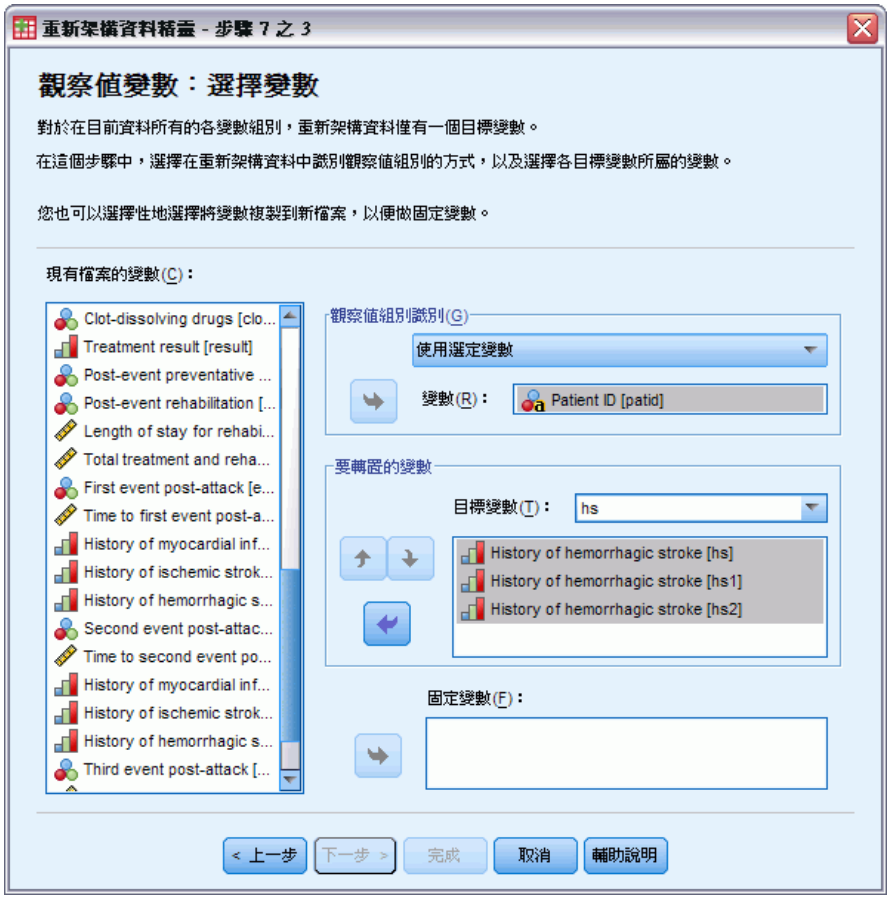
- 輸入「mi」作為目標變數。
- 選取「心肌梗塞病史 [mi]」、「心肌梗塞病史 [mi1]」和「心肌梗塞病史 [mi2]」作為要轉置的變數。
- 從目標變數清單中選取「trans5」。

圖表 22-29
「重新架構資料精靈」的「變數至觀察值選取變數」步驟



- 輸入「is」作為目標變數。
- 選取「缺血性中風病史 [is]」、「缺血性中風病史 [is1]」和「缺血性中風病史 [is2]」作為要轉置的變數。
- 從目標變數清單中選取「trans6」。

圖表 22-30
「重新架構資料精靈」的「變數至觀察值選取變數」步驟

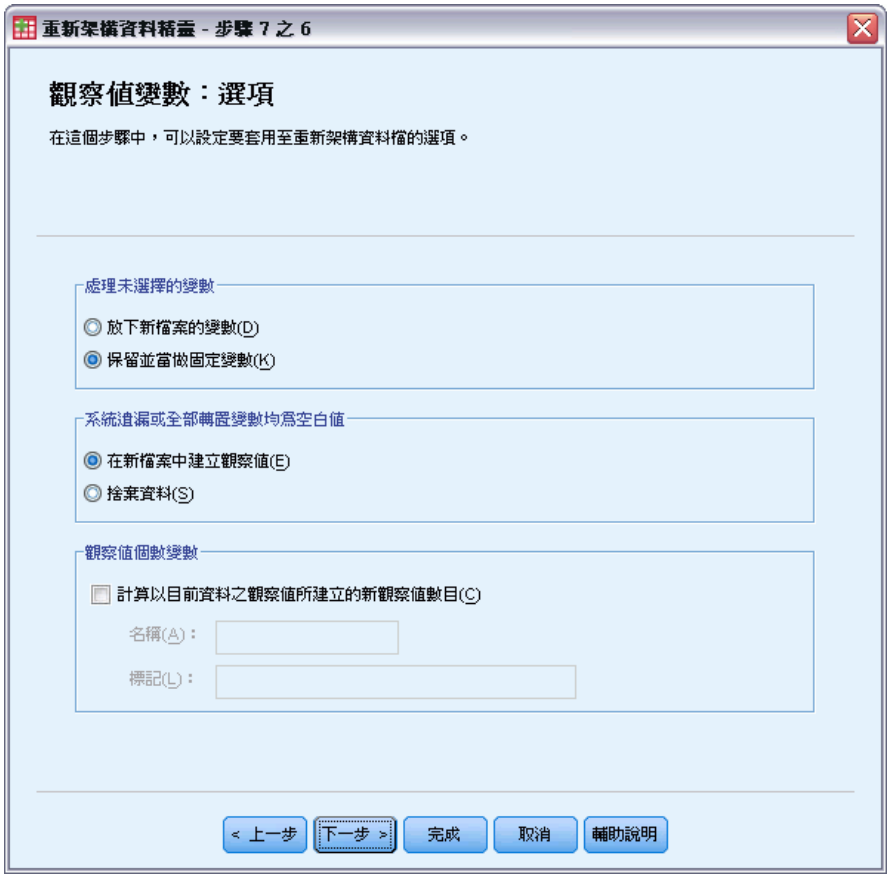


- 輸入「hs」作為目標變數。
- 選取「出血性中風病史 [hs]」、「出血性中風病史 [hs1]」和「出血性中風病史 [hs2]」作為要轉置的變數。
- 按一下「下一步」，並在「建立索引變數」步驟按一下「下一步」。

圖表 22-31
「重新架構資料精靈」的「變數至觀察值建立一個索引變數」步驟

- 輸入「event_index」作為索引變數名稱，並輸入「事件索引」作為變數標記。
- 按一下「下一步」。

圖表 22-32
「重新架構資料精靈」的「變數至觀察值建立一個索引變數」步驟



- 確定已選取「保持並視為固定變數」。
- 按一下「完成」。

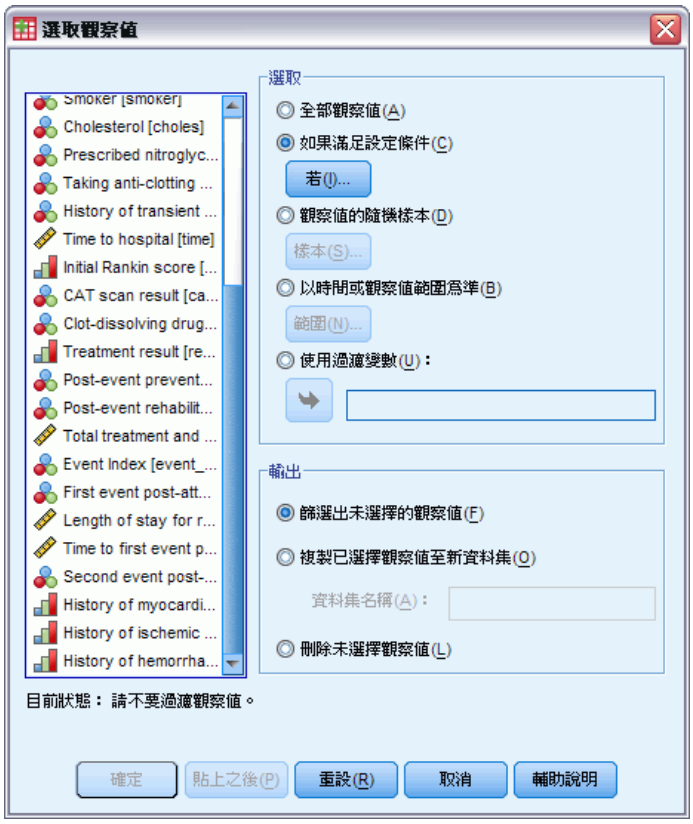
圖表 22-33
重新架構資料

event_index	event	start_time	time_to_event	mi	is	hs
1	0	3	1500	0	1	0
2	-4	1500	-4	-4	-4	-4
3	-4	.	-4	-4	-4	-4
1	1	33	1311	0	1	0
2	4	1311	1325	1	1	0
3	-3	1325	-3	-3	-3	-3
1	4	12	1098	1	1	0
2	-3	1098	-3	-3	-3	-3
3	-3	.	-3	-3	-3	-3
1	4	4	1356	0	1	0
2	-3	1356	-3	-3	-3	-3
3	-3	.	-3	-3	-3	-3

每個病患的重新架構資料包含三個觀察值；不過許多病患經歷不到三個事件，所以有許多觀察值的「事件」是負（遺漏）值。您可以簡單地從資料集中過濾這些觀察值。

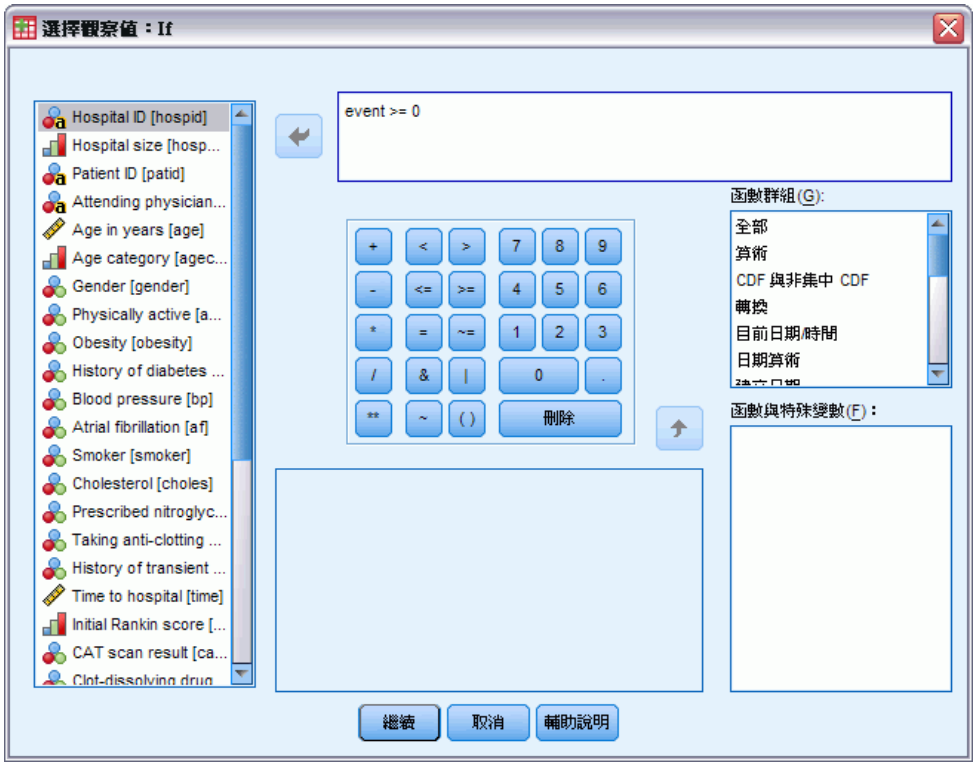
- 若要過濾這些觀察值，請從功能表選擇：
  資料 > 選取觀察值...

圖表 22-34
選取「觀察值」對話方塊



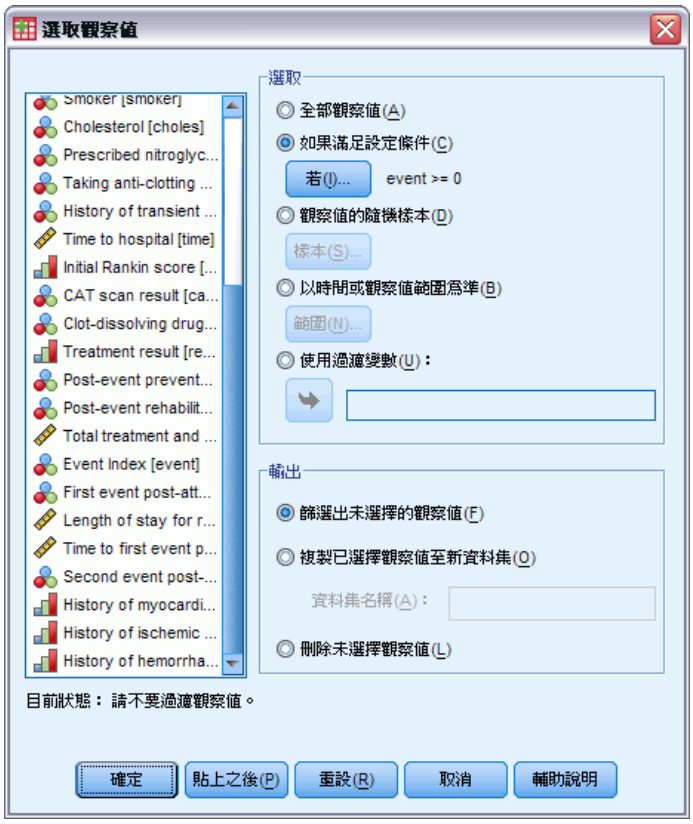
- ▶ 選取「如果滿足設定條件」。
- ▶ 按一下「如果」。

圖表 22-35
選取「觀察值選擇」對話方塊



- 輸入 event >= 0 作為條件運算式。
- 按一下「繼續」。

圖表 22-36
選取「觀察值」對話方塊



- ▶ 選取「**刪除未選擇的觀察值**」。
- ▶ 按一下「**確定**」。

建立「簡單隨機取樣分析計劃」

現在您已準備好建立簡單取樣分析計劃。

- ▶ 首先，您需要建立一個取樣權重變數。從功能表選擇：
  轉換 > **計算變數...**

圖表 22-37
「Cox 迴歸」主對話方塊

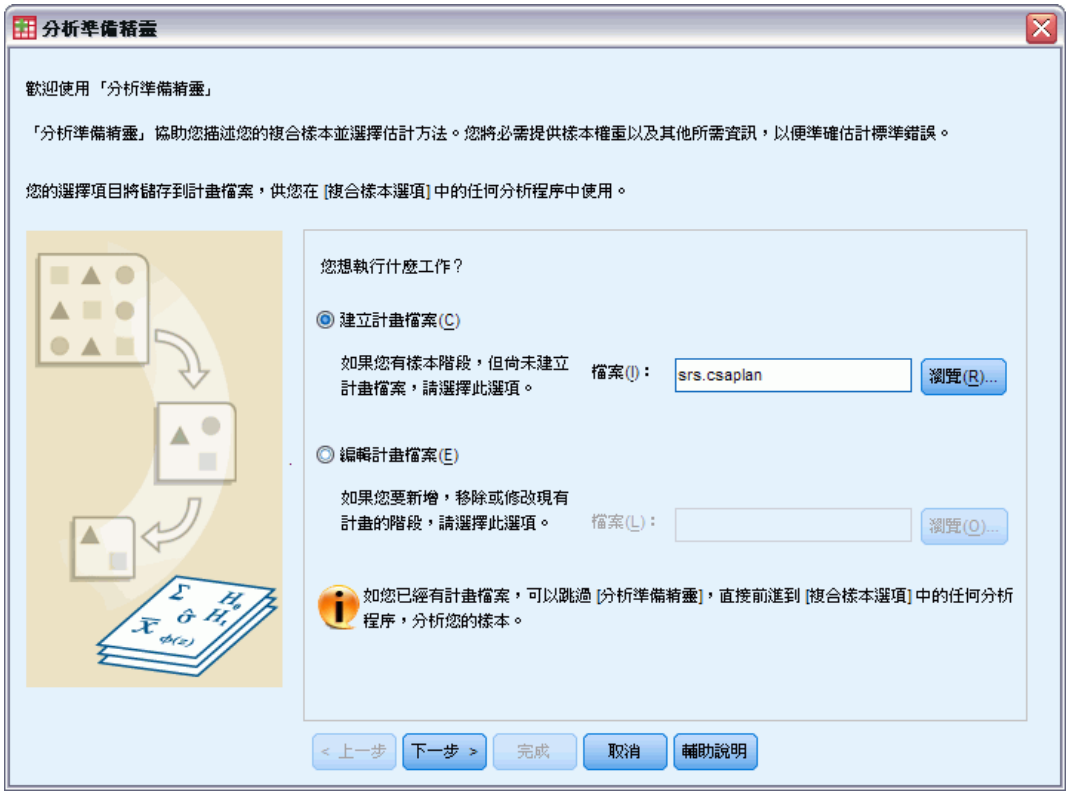
- 輸入「sampleweight」作為目標變數。
- 輸入 1 作為數值運算式。
- 按一下「確定」。

  現在您可建立分析計劃。

  注意：若您想跳過下列指示並繼續分析資料，您可以使用在範例檔目錄中的現有計劃檔 srs.csaplan。

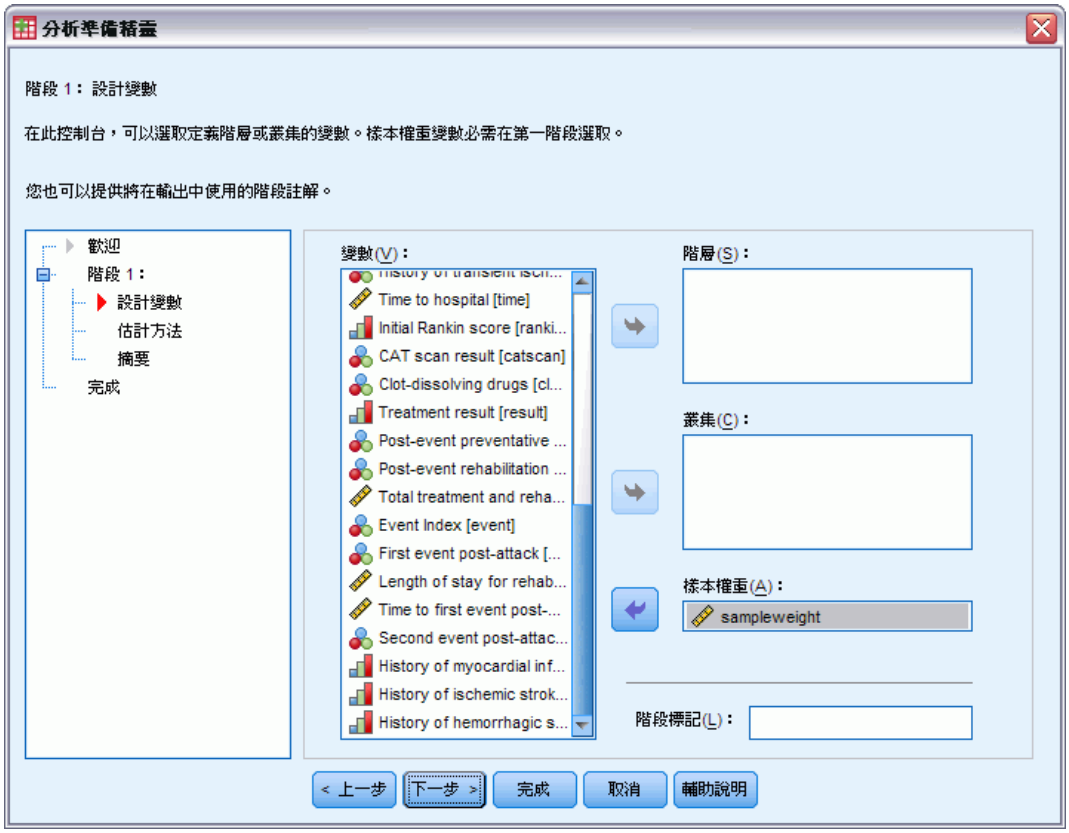
- 若要建立分析計劃，請從功能表選擇：
  分析 > 複合樣本 > 準備分析...

圖表 22-38
分析準備精靈，歡迎步驟



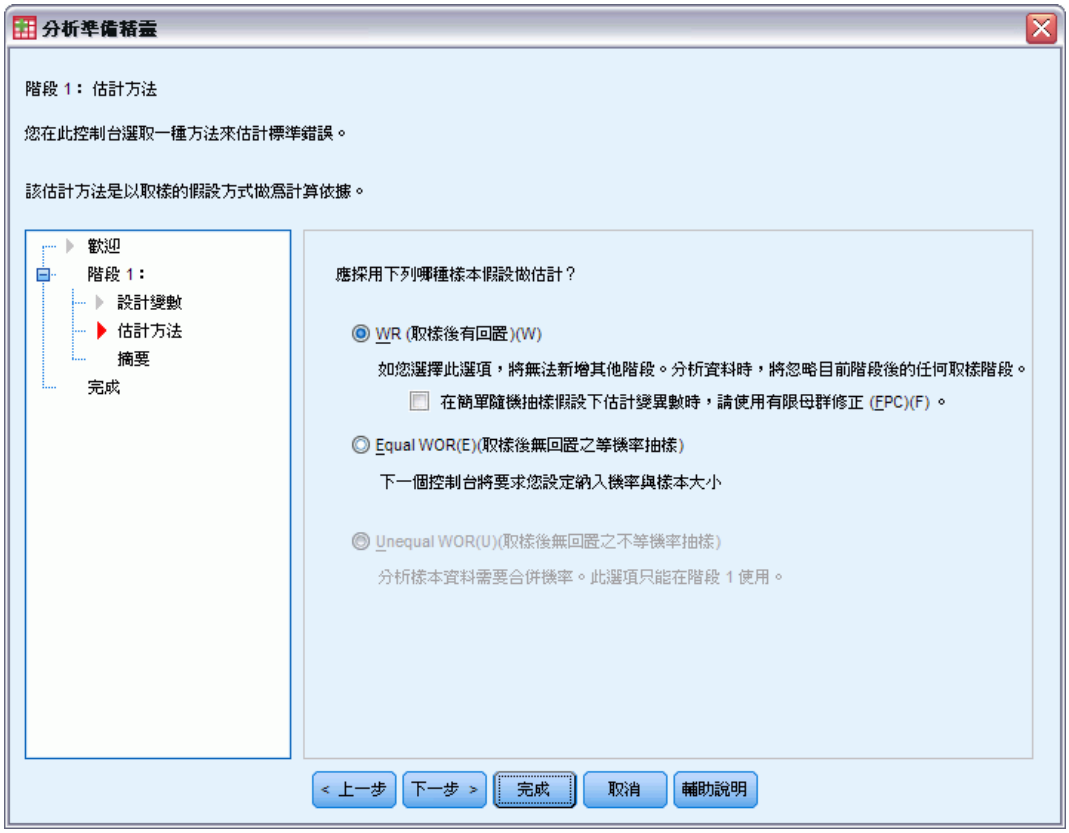
- 選取「建立計劃檔案」，並輸入 srs.csaplan 作為計劃檔的名稱。或者，瀏覽到您要儲存的位置。
- 按一下「下一步」。

圖表 22-39
「分析準備精靈」的「設計變數」



- 選取「sampleweight」作為樣本權重變數。
- 按一下「下一步」。

圖表 22-40
「分析準備精靈」的「估計方法」



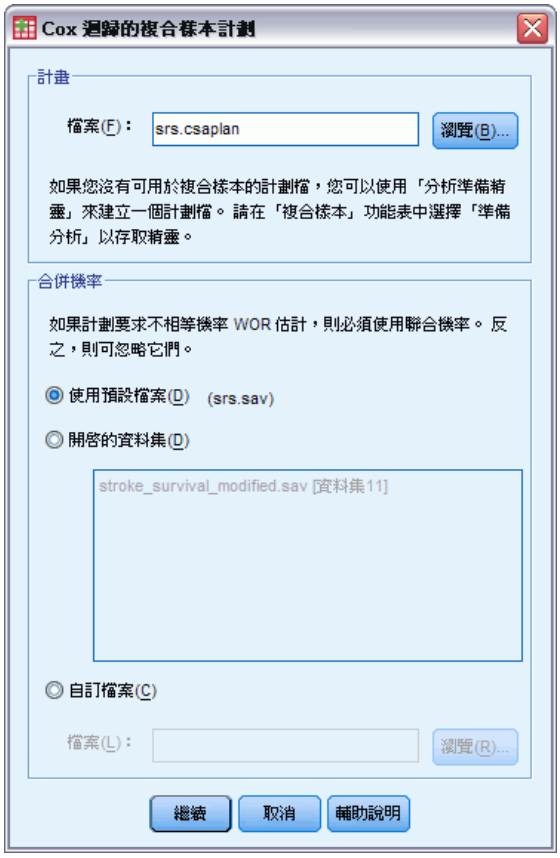
- 取消選取「使用有限母群修正」。
- 按一下「完成」。

現在您已準備好執行分析。

執行分析

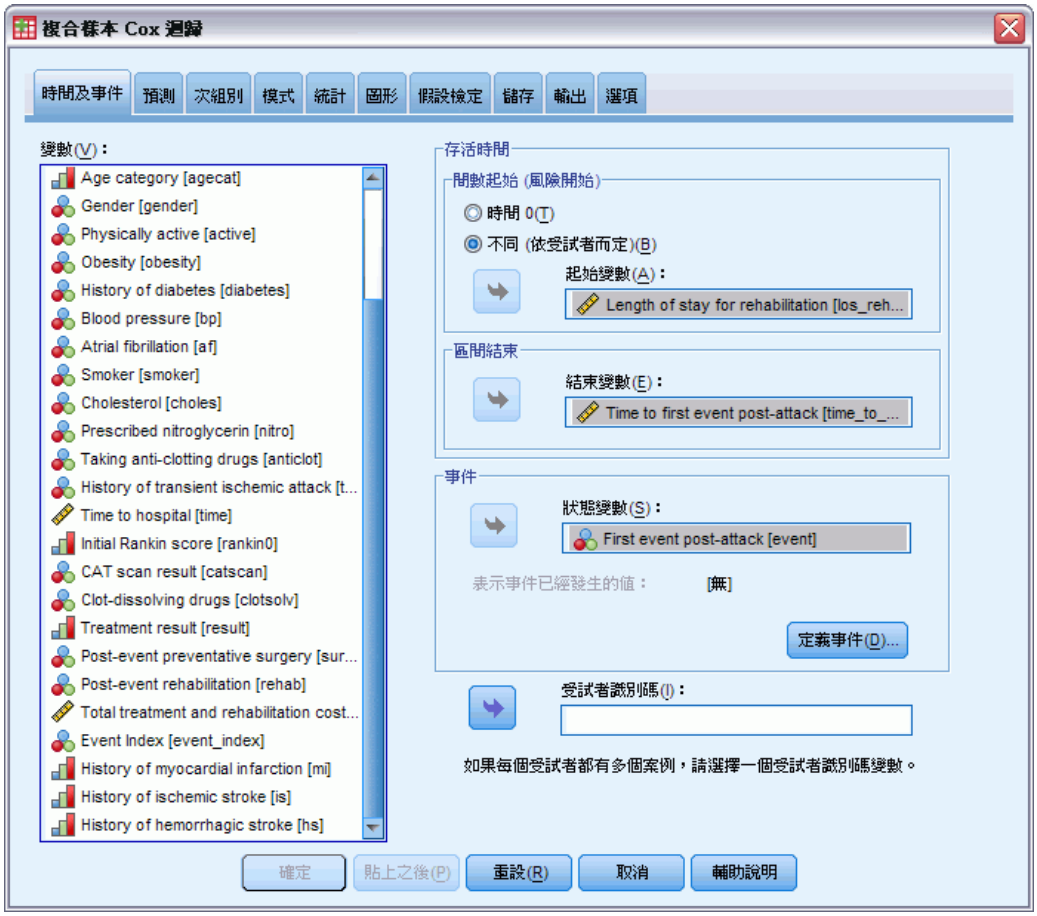
- 若要執行「複合樣本 Cox 迴歸」分析，請從功能表選擇：
  分析 > 複合樣本 > Cox 迴歸...

圖表 22-41
「Cox 迴歸計劃」對話方塊



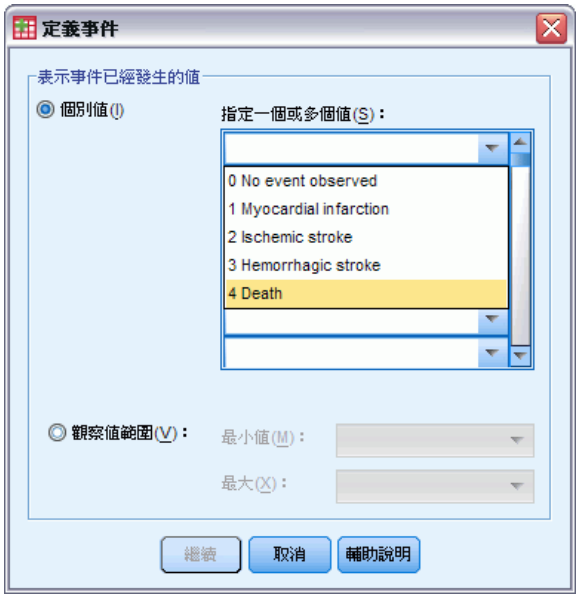
- 瀏覽至您儲存簡單隨機取樣分析計劃的位置，或樣本檔目錄，然後選取 srs.csaplan。
- 按一下「繼續」。

圖表 22-42
「Cox 迴歸」對話方塊的「時間和事件」索引標籤



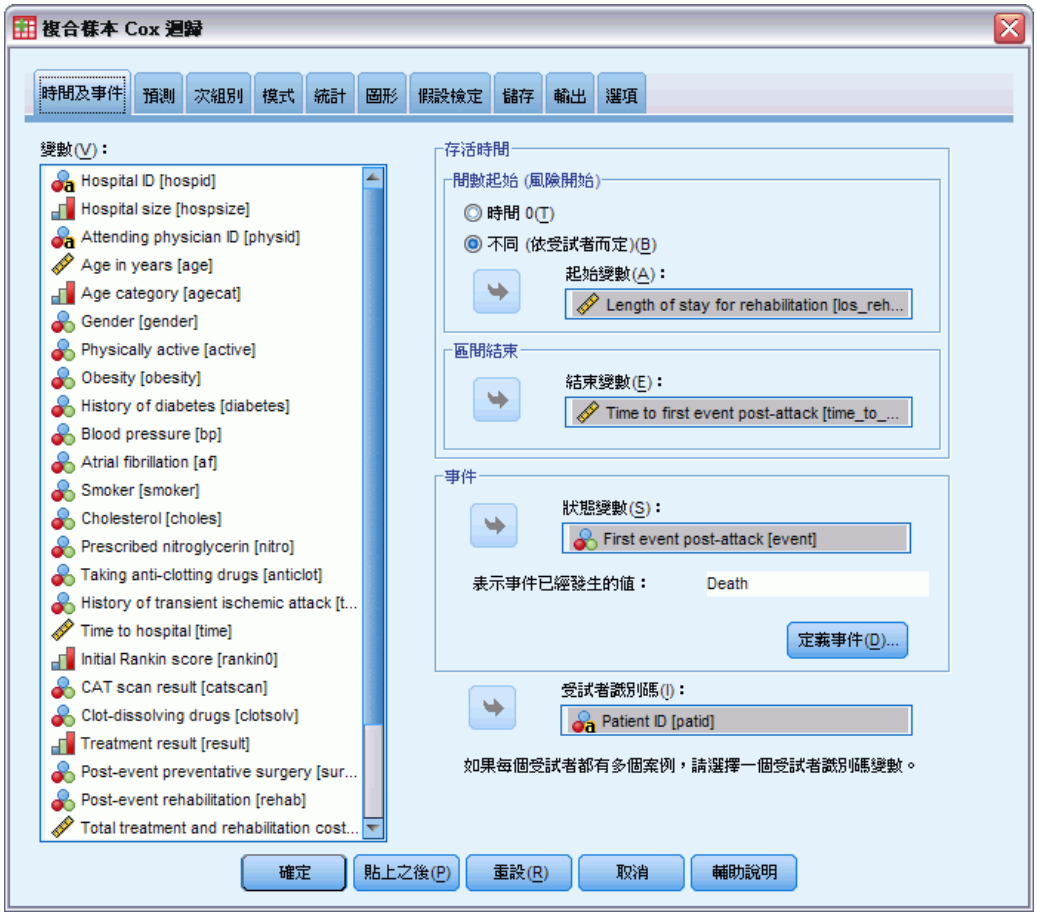
- 選取「**因受試者而異**」，再選取「康復期的時間長度 [los_rehab]」作為開始變數。請注意，重新架構的變數已自第一個用來建構它的變數取得變數標記，雖然該標記未必適於此建構的變數。
- 選取「至第一次事件發作後時間 [time_to_event]」作為結束變數。
- 選取「第一次事件發作後 [event]」作為結束變數。
- 按一下「**定義事件**」。

圖表 22-43
「定義事件」對話方塊



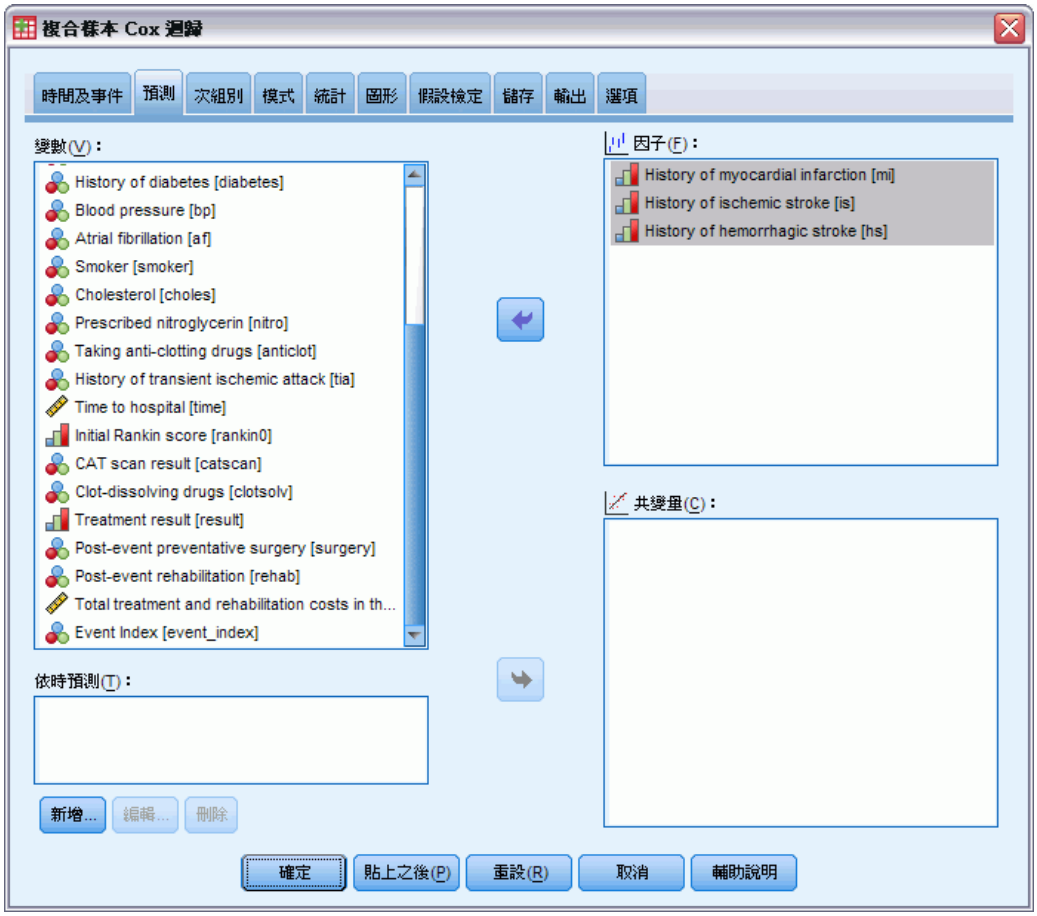
- 選取「4 死亡」作為表示已發生終端事件的值。
- 按一下「繼續」。

圖表 22-44
「Cox 迴歸」對話方塊的「時間和事件」索引標籤



- 選取「病患 ID [patid]」作為受試者識別碼。
- 按一下「**預測值**」索引標籤。

圖表 22-45
「Cox 迴歸」對話方塊的「預測值」索引標籤

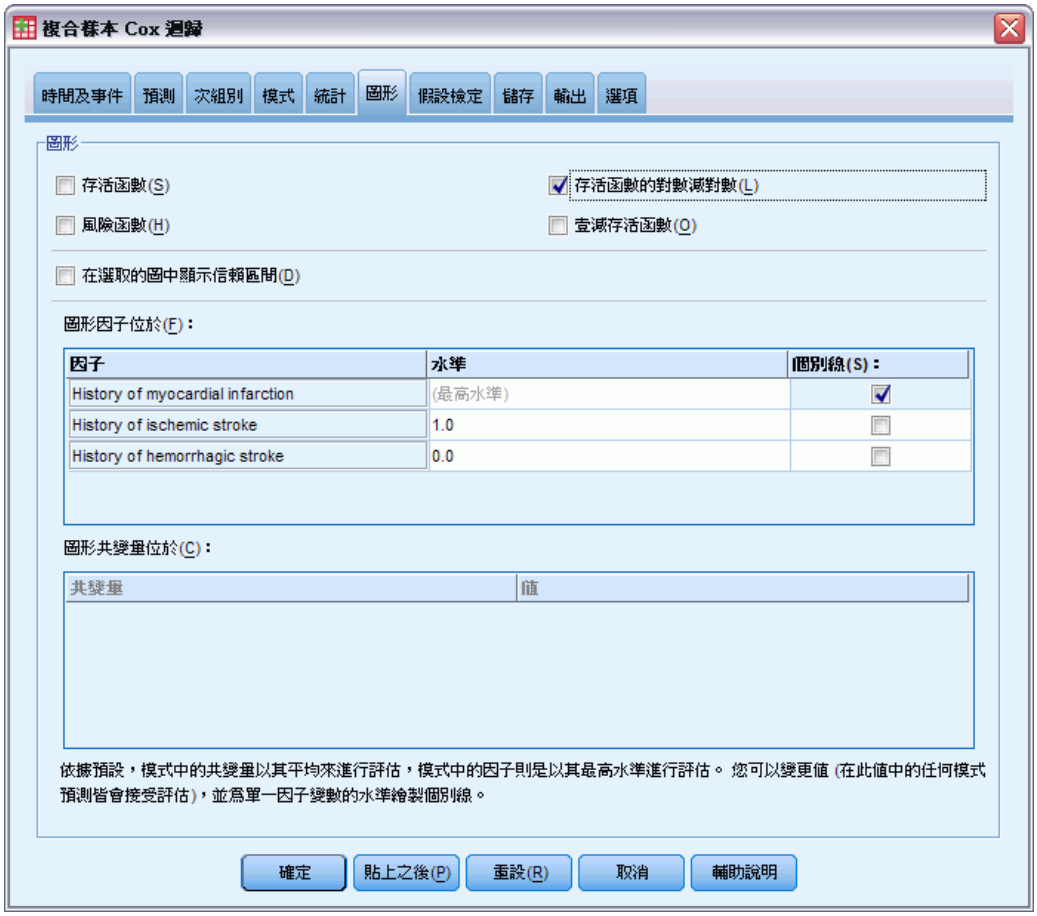


- 選取「心?梗塞病史 [mi]」至「出血性中風病史 [hs]」作為因子。
- 按一下「統計量」索引標籤。

圖表 22-46
「Cox 迴歸」對話方塊的「統計量」索引標籤

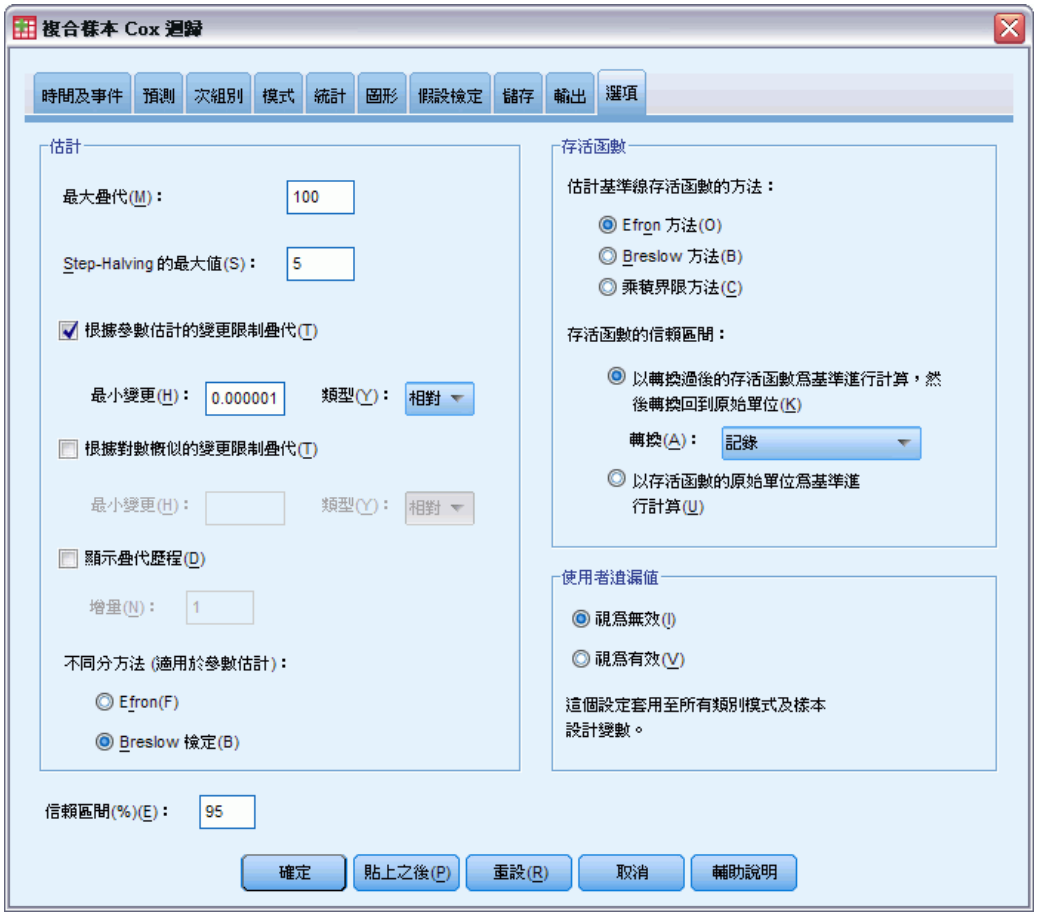
- 在「參數」組別中，選取「估計」、「指數化估計」、「標準誤」和「信賴區間」。
- 按一下「圖形」索引標籤。

圖表 22-47
「Cox 迴歸」對話方塊的「統計量」索引標籤



- 選取「負對數存活函數的對數」。
- 核取「心肌梗塞病史」的「個別線」。
- 選取「1.0」作為「缺血性中風病史」的測量水準。
- 選取「0.0」作為「出血性中風病史」的測量水準。
- 按一下「選項」索引標籤。

圖表 22-48
「Cox 迴歸」對話方塊的「選項」索引標籤



- 選取「Breslow」作為估計組別的公平處理方法。
- 按一下「確定」。

樣本設計資訊

圖表 22-49
樣本設計資訊

			N
未加權計數	有效	主題	2421
		觀察值	3310
	有效觀察值		0
	總觀察值		3310
主題大小			2421.000
階段 1	Strata		1
	Units		2421
樣本設計自由度			2420

本表包含與模式估計有關的樣本設計資訊。

- 部分受試者有多個觀察值，且所有 3,310 個觀察值都在分析中使用到。
- 本設計有一個分層和 2,421 個單位（每一個受試者有一層）。取樣設計自由度估計為 2421−1=2420。

模式效應的檢定

圖表 22-50
模式效應的檢定

源	df1	df2	Wald F	顯著性
mi	3.000	2418.000	452.873	.000
is	2.000	2419.000	1064.936	.000
hs	2.000	2419.000	739.197	.000

存活時間變數: Length of stay for rehabilitation, Time to first event post-attack
事件狀態變數: First event post-attack = 4
主題 ID 變數: Patient ID
模式: mi, is, hs

每一個效應的顯著值都接近 0，暗示其有助於本模式。

參數估計值

圖表 22-51
參數估計值

參數	B	標準差	95% 置信間隔		Exp(B)	Exp(B) 的 95.0% CI	
			下界	上界		下界	上界
[mi=0]	-6.381	.283	-6.935	-5.827	.002	.001	.003
[mi=1]	-5.589	.284	-6.147	-5.032	.004	.002	.007
[mi=2]	-2.119	.344	-2.794	-1.445	.120	.061	.236
[mi=3]	.000[a]	.	.	.	1.000	.	.
[is=1]	-6.421	.202	-6.817	-6.024	.002	.001	.002
[is=2]	-2.803	.222	-3.239	-2.366	.061	.039	.094
[is=3]	.000[a]	.	.	.	1.000	.	.
[hs=0]	-6.148	.355	-6.844	-5.453	0	.001	.004
[hs=1]	-2.232	.373	-2.963	-1.502	.107	.052	.223
[hs=2]	.000[a]	.	.	.	1.000	.	.

存活時間變數: Length of stay for rehabilitation, Time to first event post-attack
事件狀態變數: First event post-attack = 4
主題 ID 變數: Patient ID
模式: mi, is, hs
a. 由于參數冗余設置爲 0
b. 分割方法: Breslow

該程序使用每一個因子的最後的類別作為參考類別；其他類別的效應是相對於此參考類別。請注意，雖然估計對於統計檢定很有用，但指數化估計量－Exp(B)，更容易被解譯為與參考類別相關的風險的預測變更。

- 「[mi=0]」的 Exp(B) 值表示未曾發生心肌梗塞 (mi) 的病患，其死亡風險是曾三度發生心肌梗塞者的 0.002 倍。

- 「[mi=1]」和「[mi=0]」的信賴區間重疊，表示曾有一次心肌梗塞的病患與未曾發生心肌梗塞病患，其風險在統計上難以分辨。
- 「[mi=0]」和「[mi=1]」的信賴區間與「[mi=2]」的區間未重疊，且它們都未包含 0。因此，可以看出曾發生一次或未曾發生心肌梗塞的病患，與曾發生兩次心肌梗塞的病患，其風險是可以分辨的，接下來，它們與曾發生三次心肌梗塞的病患的風險也是可以分辨的。

「is」和「hs」的測量水準也保有相似的關係，增加先前事件的次數也會增加死亡的風險。

形式值

圖表 22-52
形式值

		存活時間間隔		History of myocardial infarction	History of ischemic stroke	History of hemorrhagic stroke
		開始	結束			
引用樣式	1	.000	a	Three	Three	Two
樣式 1.1	1	.000	a	None	One	None
樣式 1.2	1	.000	a	One	One	None
樣式 1.3	1	.000	a	Two	One	None
樣式 1.4	1	.000	a	Three	One	None

未指定的預測在引用樣式處分配值。
每個存活時間間隔定義為 開始 < 存活時間 <= 結束
模式 : mi, is, hs.

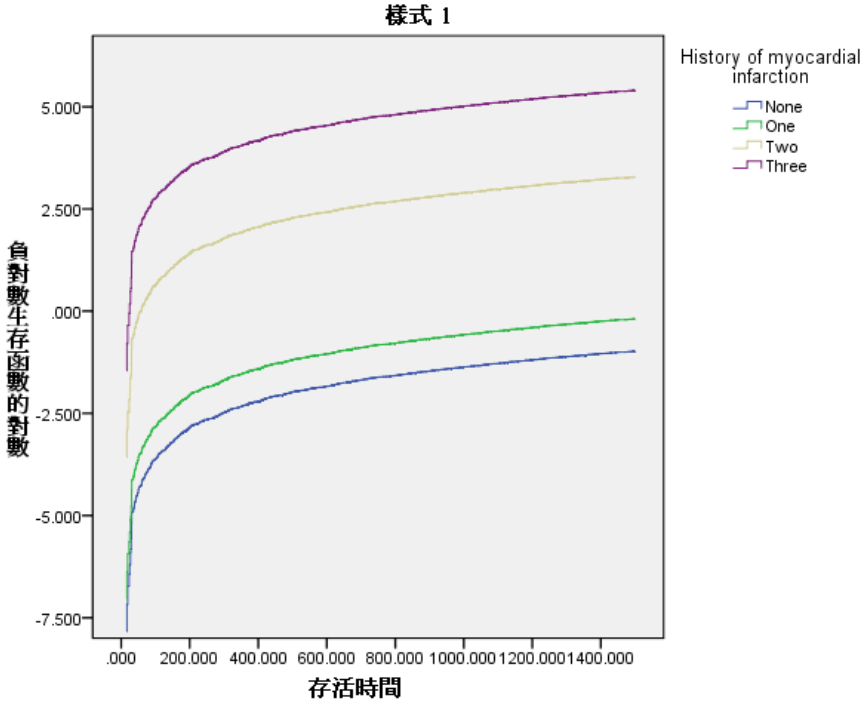
a. 未綁定

形式值表列出定義每一個預測值形式的值。除了模式中的預測值以外，也顯示存活區間的開始和結束時間。對經由對話執行的分析，其開始和結束時間永遠分別是 0 和不受限；您可以透過語法指定成段的常數預測路徑。

- 每一個因子的參考形式設定為參考類別，以及每一個共變量的平均數（在這個模式中沒有共變量）。對此資料集，不會發生所顯示參考模式的因子組合，因此我們將忽略參考形式的負對數存活函數的對數圖形。
- 形式 1.1 至 1.4 只有在「心肌梗塞病史」的值上有差異。每一個「心肌梗塞病史」的值會建立一個個別形式（及在要求圖形中的一條個別線），而其他的變數保持常數。

負對數存活函數的對數圖形

圖表 22-53
負對數存活函數的對數圖形



此圖形顯示負對數存活函數的對數，ln(−ln(suvival))，對存活時間。此特殊圖形針對 心肌梗塞病史、固定於一的缺血性中風病史和固定於無的出血性中風病史的每一個類別顯示一條個別曲線，並且是心肌梗塞病史效應在存活函數上很有用的圖像。如同在參數估計表所見，它顯示出曾發生一次或未曾發生心肌梗塞病患的存活，與曾發生兩次心肌梗塞病患的存活是可以分辨的，接下來，它們與曾發生三次心肌梗塞的病患的存活是可以分辨的。

摘要

您已為中風後存活適配了 Cox 迴歸，以估計變更中風後病患病史的效應。這只是開始，因為研究人員毫無疑問地將在此模式中包含其他潛在的預測值。此外，在進一步分析這個資料集時，您可能考慮對模式結構作更顯著的變更。例如，目前的模式假設病患病史改變事件的效應可以用基準線風險的倍數加以量化。相反地，假設基準線風險的類型會因非死亡事件發生而改變可能是合理的。要達成此目的，您可以將分析根據「事件索引」來分層。

附錄
A

範例檔案

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

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

說明

以下是使用於本文件中不同範例的範例檔之簡要描述。

- **accidents.sav。** 這是有關某保險公司研究年齡和性別風險因子對給定地區汽車意外事件的假設資料檔。每一個觀察值對應至一個年齡類別和性別的交叉分類。
- **adl.sav。** 這是有關致力於確定一個建議中風病患治療類型之效益的假設資料檔。醫師隨機指定女性中風病患至兩個組別之一。第一組接受標準的物理治療，而第二組則接受額外的情緒治療。在治療了三個月後，將每一個病患進行日常活動的能力記分為次序變數。
- **advert.sav。** 這是有關一家零售商致力於調查廣告費與廣告後銷售情形之間的關係的假設資料檔。為了這個目的，他們收集了過往銷售數字和相關的廣告費用。
- **aflatoxin.sav。** 這是有關檢定玉米作物是否有黃麴毒素（一種毒物，其濃度在介於和處於作物產量中都有很大的差異）的假設資料檔。一名穀物加工者收到來自 8 個作物產量各 16 個樣本，並以十億當量 (PPB) 來測量黃麴毒素的水準。
- **aflatoxin20.sav。** 這個資料檔包含由 aflatoxin.sav 取得，來自 4 和 8 作物產量的 16 個樣本，每一個樣本的黃麴毒素測量。
- **anorectic.sav。** 在將厭食/暴食行為症狀學標準化的過程中，研究人員 (Van der Ham, Meulman, Van Strien, 和 Van Engeland, 1997) 研究了 55 個飲食失調的青少年。每個病患在四年之中被訪問四個回合，所以得到總數為 220 的觀察值。在每次觀察中，為病患在 16 種症狀上逐一評分。目前遺漏了第二次訪察的病患 71，第二次訪察的病患 76，以及第三次訪察的病患 47 的症狀分數，因此只剩下 217 個有效觀察值。
- **autoaccidents.sav。** 這是有關一位保險分析師致力於為每個駕駛的汽車意外事件次數建立模式，同時考量駕駛的年齡和性別的假設資料檔。每一個觀察值代表一位不同的駕駛，記錄了駕駛的性別、年齡、和近五年內的汽車意外事故次數。
- **band.sav。** 本資料檔包含某樂團音樂 CD 假設性的每週銷售數字。也包含三個可能預測變數的資料。
- **bankloan.sav。** 這是有關一家銀行致力於減少放款利率預設值的假設資料檔。本檔包含 850 位以前的客戶與現在的準客戶的財務和人口資料。前 700 個觀察值為以前有借貸的客戶。最後 150 個觀察值是銀行需要作信用風險優良與不良分類的準客戶。

- **bankloan_binning.sav。** 這是包含 500 位以前客戶的財務和人口資料的假設資料檔。
- **behavior.sav。**在典型範例 (Price 和 Bouffard, 1974) 中，52 名學生被要求為 15 種情境與 15 種行為組合評等，等級共分為 10 點，從 0 =「非常適當」到 9 =「 非常不適當」。平均值超過個別值，值會被視為相異性。
- **behavior_ini.sav。** 本資料檔包含 behavior.sav 之二維解的起始組態。
- **brakes.sav。** 這是有關一間生產高性能汽車碟型煞車片工廠中品質管制的假設資料檔。資料檔包含由 8 個生產機器分別取得 16 個碟片的直徑測量。煞車的目標直徑是 322 公釐。
- **breakfast.sav。**在經典研究中(Green 和 Rao, 1972)，21 名 Wharton 學院 MBA 學生及其配偶被要求為 15 項早餐食品按喜愛程度分出等級：從 1 = 「最喜愛」到 15 =「最不喜愛」。他們的喜愛程度分六種不同情況記錄，從「整體喜愛」到「點心，僅配飲料」。
- **breakfast-overall.sav。** 本資料檔只包含第一種情況－「整體喜愛」－所喜愛的早餐項目。
- **broadband_1.sav。** 這是包含全國性寬頻服務地區用戶數目的假設資料檔。本資料檔包含四年期間 85 個地區每月的用戶數目。
- **broadband_2.sav。** 本資料檔與 broadband_1.sav 相同，但多了三個月的資料。
- **car_insurance_claims.sav。** 一個在別處 (McCullagh 和 Nelder, 1989) 出現和分析過，有關汽車損害理賠的資料集。理賠金額的平均數可建立模式為具有 gamma 分配，使用反連結函數將依變數的平均數相關至一被保險人年齡、車輛類型、和車齡的線性組合。提出理賠的數量可以用作尺度權重。
- **car_sales.sav。** 本資料檔包含假設性的銷售估計、定價、和不同的品牌與車輛型式的實體規格。定價和實體規格是由 edmunds.com 和製造商處輪流取得。
- **car_sales_uprepared.sav。** 這是 car_sales.sav 的修改版本，其中不包含任何欄位的轉換版本。
- **carpet.sav。**在一個普遍的範例 (Green 和 Wind, 1973) 中，計劃銷售全新地毯清潔機的公司想要檢驗影響消費者偏好的五個因子—包裝設計、品牌名稱、價格、「優秀家用品」獎章及退費保證。包裝設計有三個因子水準，每個水準中的清潔刷位置都不相同；三個品牌名稱 (K2R、Glory、及 Bissell)；三個價格水準；且最後兩個因子各有兩個水準 (無論無或有)。十名消費者將這些因子所定義的 22 種組合分級。「偏好」變數包含每個組合平均排名的等級。排名數值較小者會對應高偏好程度。這個變數反映每個組合偏好的整體量數。
- **carpet_prefs.sav。** 本資料檔是根據 carpet.sav 所描述的相同範例，但它包含 10 個消費者每一個人的實際等級。消費者被要求將 22 個產品組合從最喜歡排列到最不喜歡。變數「PREF1」到「PREF22」包含相關組合的識別碼，如 carpet_plan.sav 中所定義。
- **catalog.sav。** 本資料檔包含郵購公司銷售三項產品的每月假設銷售數字。也包含五個可能預測變數的資料。
- **catalog_seasfac.sav。** 本資料檔與 catalog.sav 相同，不過多了一組由「週期性分解」程序所計算的週期性因子以及隨附的資料變數。
- **cellular.sav。** 這是有關一家手機公司致力於減少顧客不忠的假設資料檔。顧客不忠傾向分數套用於帳戶，範圍由 0 至 100。帳戶分數 50 或以上有可能正尋求變更供應商。

- **ceramics.sav。** 這是有關一家製造商致力於確定一種新的優良合金是否較標準的合金有較大的耐熱性的假設資料檔。每一個觀察值代表對合金之一的不同檢定；記錄了讓軸承失效的溫度。
- **cereal.sav。** 這是有關對 880 人的早餐喜好進行訪談的假設資料檔，也記下他們的年齡、性別、婚姻狀況、和是否有活躍的生活型態 (根據他們是否一週運動兩次)。每一個觀察值代表一位不同的應答者。
- **clothing_defects.sav。** 這是有關一家服裝工廠品質管制過程的假設資料檔。由該工廠所生產的每一批產品中，檢查員取出一件服裝的樣本並計算不合格的服裝個數。
- **coffee.sav。** 本資料檔是關於六種冰咖啡品牌的感覺印象(Kennedy, Riquier, 和 Sharp, 1996)。對 23 種冰咖啡中每一種的印象屬性，由群眾來選取依其屬性描述的所有品牌。該六種品牌已標示為 AA、BB、CC、DD、EE、和 FF，以保持機密。
- **contacts.sav。** 這是有關一群公司電腦銷售代表聯絡清單的假設資料檔。每一個聯絡人依他們在公司所服務的部門及其公司的等級而分類。最後一次銷售的金額、到最後一次銷售的時間、和該聯絡人公司的規模也都被列入記錄。
- **creditpromo.sav。** 這是有關一家百貨公司致力於評估近期信用卡促銷活動效果的假設資料檔。為達此目標，隨機選取了 500 位持卡人。有半數收到廣告，促銷在未來三個月購買將獲得降低利率的優惠。半數收到標準的週期性廣告。
- **customer_dbase.sav。** 這是有關一家公司致力於使用其資料倉庫的資訊來對最有可能回應的客戶提供優惠的假設資料檔。隨機選取客戶庫的子集，提供優惠，再將他們的回應記錄下來。
- **customer_information.sav。** 本檔案是包含客戶郵寄資訊的假設資料檔，例如姓名和地址。
- **customer_subset.sav。** 80 個 customer_dbase.sav 的觀察值子集。
- **customers_model.sav。** 本檔案包含一市場行銷活動所鎖定之個人的假設資料。這些資料包含人口資訊、購買歷史摘要、和每一個人是否對該活動有回應。每一個觀察值代表一位不同的個人。
- **customers_new.sav。** 本檔案包含一市場行銷活動潛在候選人之個人的假設資料。這些資料包含每一位個人的人口資訊和購買歷史摘要。每一個觀察值代表一位不同的個人。
- **debate.sav。** 這是有關一項政治辯論會參與者辯論前和辯論後接受調查之成對反應的假設資料檔。每一個觀察值對應至一位不同的應答者。
- **debate_aggregate.sav。** 這是將 debate.sav 中之反應作整合的假設資料檔。每一個觀察值對應至辯論前和辯論後對偏好之交叉分類的反應。
- **demo.sav。** 這是有關提供郵寄每月優惠之購買客戶資料庫的假設資料檔。記錄了客戶是否對該優惠回應，以及各種的人口資訊。
- **demo_cs_1.sav。** 這是有關一家公司致力於匯編調查資訊資料庫之第一步的假設資料檔。每一個觀察值對應至一個不同的城市，也記錄了其地區、省、區、和城市識別。
- **demo_cs_2.sav。** 這是有關一家公司致力於匯編調查資訊資料庫之第二步的假設資料檔。每一個觀察值對應至在第一步中選取的城市中的一個不同的家庭單位，也記錄了其地區、省、區、分區、和單位識別。也納入了由該設計的前兩階段所得之取樣資訊。
- **demo_cs.sav。** 這是包含以複合取樣設計所收集之調查資訊的假設資料檔。每一個觀察值對應至一個不同的家庭單位，也記錄了各種的人口和取樣資訊。

- **dmdata.sav。** 這是包含直效行銷公司之人口和購買資訊的假設資料檔。dmdata2.sav 包含收到測試郵件的連絡人子集資訊，而 dmdata3.sav 則包含剩下未收到測試郵件的連絡人資訊。
- **dietstudy.sav。** 本假設資料檔包含對「Stillman 飲食法」(Rickman, Mitchell, Dingman, 和 Dalen, 1974) 研究的結果。每一個觀察值對應至一個不同的受試者，並記錄下他或她飲食法前、後之體重 (磅) 和三酸甘油酯水準 (毫克/100 毫升)。
- **dvdplayer.sav。** 這是有關新 DVD 播放器開發的假設資料檔。市場行銷團隊使用原型收集了焦點組別資料。每一個觀察值對應至不同調查到的使用者，並記錄下一些有關他們的人口資訊和他們對有關原型問題的回應。
- **german_credit.sav。** 本資料檔取自 (Blake 和 Merz, 1998) 艾文 (Irvine) 在加州大學機器學習資料庫儲存器的「德國信用」資料集。
- **grocery_1month.sav。** 本假設資料檔是將 grocery_coupons.sav 資料檔和每週購買的「彙總」，因此每一個觀察值對應至一個不同的客戶。結果部份每週變更的變數消失了，而目前所記錄的銷售量是在研究的四週期間銷售量之總和。
- **grocery_coupons.sav。** 這是包含某連鎖雜貨店想要知道他們客戶購買習慣所收集之調查資料的假設資料檔。每一個客戶被追蹤了四週，每一個觀察值對應至一個不同的客戶-週，並記錄有關客戶在何處及如何購物的資訊，包含那一週在雜貨店花了多少錢。
- **guttman.sav。**Bell(Bell, 1961) 以此表說明可能的社會團體。Guttman (Guttman 值, 1968) 過去曾使用此表的一部分，在這部分中有 5 個變數，分別說明 7 個理論社會團體的社會互動、團體歸屬感、成員實際接觸和關係正式性，而這 7 個群組包括：群眾 (例如，足球場上的人)、觀眾 (例如在戲院中和課堂上的人)、公眾 (例如，報紙讀者和電視觀眾)、暴民 (和群眾相似，但互動較為激烈)、原級團體 (親密性)、次級團體 (自願性) 和現代社群 (因親密的身體接近而導致鬆散的結盟和特殊服務的需求)。
- **health_funding.sav。** 這是包含醫療保健基金 (每 100 個人口的金額)、疾病率 (每 10,000 個人口的比率)、造訪醫療保健機構的比例 (每 10,000 個人口的比率) 的假設資料檔。每一個觀察值代表一個不同的城市。
- **hivassay.sav。** 這是有關一家製藥實驗室致力於開發一種偵測 HIV 感染快速檢驗的假設資料檔。檢驗結果是八個紅色加深的陰影，陰影愈深表示感染的可能性愈大。進行了一項實驗室的試驗，在 2,000 個血液樣本中，有半數遭到 HIV 的感染，而半數則未感染。
- **hourlywagedata.sav。** 這是有關在辦公室和醫院任職的護士依經驗水準不同之鐘點費的假設資料檔。
- **insurance_claims.sav.** 這是有關一家保險公司想要建立模式來標示可疑及可能的詐欺理賠之假設資料檔。每一個觀察值代表個不同的理賠。
- **insure.sav。** 這是有關一家保險公司正在研究表示客戶是否必定理賠 10 年壽險合約之風險因子的假設資料檔。在資料檔中的每一個觀察值代表二份合約，其一記錄了理賠而另一則否，二者的年齡和性別相符。
- **judges.sav。** 這是有關受過訓練的裁判 (加上一位熱心人士) 為 300 個體操表演評分的假設資料檔。每一列代表一個不同的表演；裁判們觀看相同的表演。
- **kinship_dat.sav。**Rosenberg 與 Kim (Rosenberg 和 Kim, 1975) 致力於分析 15 個親屬關係稱呼 (姑/姨、兄弟、堂/表兄弟姐妹、女兒、父親、孫女、祖父、祖母、孫子、母親、姪子/外甥、姪女/外甥女、姐妹、兒子、叔/舅父)。他們請四組大學生 (兩組女性、兩組男性) 根據其相似性來分類整理這些稱謂。他們請其中兩組 (一組

女性、一組男性）作兩次分類整理，第二次要根據與第一次不同的準則進行分類整理。因此，總共得到六個「來源」。每一個來源對應至一個 15 × 15 的相似性矩陣，其儲存格等於來源中人數減去物件在該來源中分為同組的次數。

- **kinship_ini.sav。** 本資料檔包含 kinship_dat.sav 之三維解的起始組態。
- **kinship_var.sav。** 本資料檔包含自變數「性別」、「世代」、和可用來解讀 kinship_dat.sav 解答維度的（分離）「度」。尤其，它們可用來將解答空間限制為這些變數的線性組合。
- **marketvalues.sav。** 本資料檔有關於一項在伊立諾州阿爾岡京 (Algonquin, Ill.) 的新屋開發案自 1999 年至 2000 年之房屋銷售情況。這些銷售與公共記錄有關。
- **nhis2000_subset.sav。** 「國民健康訪問調查 (NHIS)」為美國民間人口的一大型民眾調查。其以具全國代表性的家庭為樣本，面對面的完成訪問。而取得各家庭中成員的人口統計學資訊及健康行為、健康狀態方面等觀察報告。本資料檔包含一個 2000 年調查資訊的子集。國家衛生統計中心。2000 年「國民健康訪問調查 (NHIS)」。公用資料檔案和文件。ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/。2003 年曾存取。
- **ozone.sav。** 本資料包含對六個氣象變數所作的 330 個觀察值，以自其餘的變數中預測臭氧濃度。先前研究人員中，(Breiman 和 Friedman 檢定(F), 1985)、(Hastie 和 Tibshirani, 1990) 在這些會阻礙標準迴歸方式的變數中發現非線性。
- **pain_medication.sav。** 本假設資料檔包含治療慢性關節炎疼痛之消炎藥物臨床試驗的結果。特別關注於藥物發生作用的時間以及它是如何與現用藥物作比較。
- **patient_los.sav。** 本假設資料檔包含對因可能為心肌梗塞 (MI，或「心臟病」) 入院病患的治療記錄。每一個觀察值對應至一個不同的病患並記錄許多與其留院期間有關的變數。
- **patlos_sample.sav。** 本假設資料檔包含病患在為心肌梗塞 (MI，或「心臟病」) 治療期間接受血栓溶解治療的治療記錄樣本。每一個觀察值對應至一個不同的病患並記錄許多與其留院期間有關的變數。
- **polishing.sav。** 這是取自「資料和故事圖書館」的「Nambeware 打磨時間」資料檔。它是有關一家金屬餐具製造商 (Nambe Mills, 聖塔非, 新墨西哥州) 致力於規劃其生產排程。每一個觀察值代表生產線上一個不同的產品。每一個產品都記錄下直徑、打磨時間、價格、和產品類別。
- **poll_cs.sav。** 這是有關民意測驗專家致力於確定交付立法之前公眾對法案支持水準的假設資料檔。觀察值對應至登記選民。每一個觀察值記錄下選民的郡、鎮、和他居住的鄰近範圍。
- **poll_cs_sample.sav。** 本假設資料檔包含列於 poll_cs.sav 中的選民樣本。樣本是根據在 poll.csplan 計劃檔中指定的設計來取得，而本資料檔記錄了包含機率和樣本權重。不過，請注意，由於取樣計劃採用到機率 - 比例 - 大小 (PPS) 方法，也用到一個包含聯合選擇機率的檔案 (poll_jointprob.sav)。其他與選民人口及其對提議法案之意見有關的變數都在取樣後收集並加入資料檔中。
- **property_assess.sav。** 這是有關郡財產估價人員致力於對限定資源保持財產價值評估維持最新的假設資料檔。觀察值對應至郡內過去一年銷售的財產。資料檔中的每一個觀察值記錄了財產所在的鎮、上次訪查該財產的估價人員、自那次評估後經過的時間、當時定的估價、和該財產銷售價值。

- **property_assess_cs.sav。** 這是有關州財產估價人員致力於對限定資源保持財產價值評估維持最新的假設資料檔。觀察值對應至州中的財產。資料檔中的每一個觀察值記錄了郡、鎮、和財產所在的鄰近範圍、自最後一次評估後經過的時間、和當時定的估價。
- **property_assess_cs_sample.sav。** 本假設資料檔包含列於 property_assess_cs.sav 中的財產樣本。樣本是根據在 property_assess.csplan 計劃檔中指定的設計來取得，而本資料檔記錄了包含機率和樣本權重。另外的變數「目前價值」是在取樣後收集並加入資料檔中。
- **recidivism.sav。** 這是有關政府法令執行機構致力於瞭解其轄區內之再犯率的假設資料檔。每一個觀察值對應至一個先前的違法者並記錄其人口資訊、第一次犯罪的一些細節、然後是直到第二次被捕的時間（如果它發生在第一次被捕的兩年之內）。
- **recidivism_cs_sample.sav。** 這是有關政府法令執行機構致力於瞭解其轄區內之再犯率的假設資料檔。每一個觀察值對應到一個先前的違法者，在 2003 年六月第一次被捕後釋放，並記錄其人口資訊、第一次犯罪的一些細節、和第二次被捕日期（如果它發生在 2006 年六月之前）。違法者是根據在 recidivism_cs.csplan 中所指定的取樣計劃之樣本部門來選取；由於取樣計劃採用到機率 - 比例 - 大小 (PPS) 方法，也用到一個包含聯合選擇機率的檔案 (recidivism_cs_jointprob.sav)。
- **rfm_transactions.sav。** 本檔案是包含購買交易資料的假設資料檔，包括購買日期、購買項目及每一項交易的金額。
- **salesperformance.sav。** 這是有關評估兩個新售貨員訓練課程的假設資料檔。六十個員工，分成三個組別，全部接受標準訓練。此外，組別二得到技術訓練；組別三則是實務輔導簡介。每一個員工在訓練課程結束時接受測驗並記錄他們的分數。在資料檔中每一個觀察值代表一個不同的訓員，並記錄他們所分派的組別和他們在測驗中得到的分數。
- **satisf.sav。** 這是有關一家零售公司在 4 個商店位置所作之滿意度調查的假設資料檔。總共有 582 位客戶接受調查，每一個觀察值代表一位客戶的反應。
- **screws.sav。** 這個資料檔包含螺絲釘、螺栓、螺帽和圖釘之特色的資訊(Hartigan, 1975)。
- **shampoo_ph.sav。** 這是有關一家美髮產品工廠品質管制過程的假設資料檔。在固定的時間間隔，記錄下六個不同輸出批次的測量和它們的 pH 值。目標範圍是 4.5 - 5.5。
- **ships.sav。** 一個在別處 (McCullagh et al。, 1989) 出現和分析過，有關商船因風浪所造成損壞的資料集。事件次數可建立模式為以 Poisson 率發生，給定船型、建造期間、和服務期間。以因子交叉分類所形成的表格的每一個儲存格服務月數的整合，提供了暴露於風險之值。
- **site.sav。** 這是有關一家公司致力於為事業擴展選擇新地點的假設資料檔。您們僱請兩位顧問分別評估該地點，除了一份廣泛的報告之外，他們還要將每個地點摘要為前景「佳」、「可」、或「差」。
- **smokers.sav。** 本資料檔是由「1998 年全國家庭毒品濫用調查」中摘錄，且是美國家庭的機率樣本。(http://dx.doi.org/10.3886/ICPSR02934) 因此，在分析本資料檔的第一步應該是將資料加權以反映母群體傾向。
- **stroke_clean.sav。** 本假設資料檔包含一個醫療資料庫，其在以「資料準備」選項中的程序清理之後的狀態。
- **stroke_invalid.sav。** 本假設資料檔包含一個醫療資料庫的起始狀態並包含幾個資料輸入錯誤。

- **stroke_survival。** 本假設資料檔是有關缺血性中風的病患，其在結束康復計畫後存活時間方面，面臨許多挑戰。中風後，記載了心肌梗塞、缺血性中風、或出血性中風的發生，以及事件記錄的時間。由於它只包含在康復計劃所管制的中風存活的病患，此樣本的左側被截斷。
- **stroke_valid.sav。** 本假設資料檔包含一個醫療資料庫，在其值以「驗證資料」程序檢查之後的狀態。它仍包含可能的異常觀察值。
- **survey_sample.sav。** 本資料檔包含調查資料，包括人口資料和各種態度測量。雖然已修改一些資料數值，且為人口資料之目的新增了一些額外的虛構變數，但是資料仍是以「1998 NORC 基本社會調查」的變數子集為基礎。
- **telco.sav。** 這是有關一家電信公司致力於在客戶庫中減少顧客不忠的假設資料檔。每一個觀察值對應至一位不同的客戶並記錄不同的人口資料和服務使用方式資訊。
- **telco_extra.sav。** 本資料檔類似於 telco.sav 資料檔，但「任期」的對數轉換客戶花費變數已予刪除，並更換為標準的對數轉換客戶花費變數。
- **telco_missing.sav。** 本資料檔是 telco.sav 資料檔的子集，不過某些人口資料值已更換為遺漏值。
- **testmarket.sav。** 本假設資料檔有關於一家速食連鎖店計劃在菜單中加入新的項目。有三個可能的活動來促銷此新產品，所以該新項目在幾個隨機選取市場中的地點作介紹。在每一個地點使用不同的促銷，並記錄該新項目前四週的每週銷售量。每一個觀察值對應至一個不同的地點-週。
- **testmarket_1month.sav。** 本假設資料檔是將 testmarket.sav 資料檔和每週購買的「彙總」，因此每一個觀察值對應至一個不同的客戶。結果部份每週變更的變數消失了，而目前所記錄的銷售量是在研究的四週期間銷售量之總和。
- **tree_car.sav。** 這是包含人口資料和車輛購買價格資料的假設資料檔。
- **tree_credit.sav。** 這是包含人口資料和銀行放款歷史資料的假設資料檔。
- **tree_missing_data.sav** 這是包含有大量遺漏值的人口資料和銀行放款歷史資料的假設資料檔。
- **tree_score_car.sav。** 這是包含人口資料和車輛購買價格資料的假設資料檔。
- **tree_textdata.sav。** 一個只有兩個變數的簡單資料檔，主要目的在顯示變數預設狀態（在指定量測水準和數值標記之前）。
- **tv-survey.sav。** 這是有關一家電視製片廠考量是否要延長一個成功節目的播送所作之調查的假設資料檔。有 906 位應答者被問到在不同的狀況下他們是否願意觀看這個節目。每一列代表一個不同的應答者；每一行為一個不同的狀況。
- **ulcer_recurrence.sav。** 本檔案包含一項用來比較兩種防止潰瘍復發治療法功效之研究的部分資訊。它是很好的區間受限資料範例，且已在別處 (Collett, 2003) 出現和分析過。
- **ulcer_recurrence_recoded.sav。** 本檔案是將 ulcer_recurrence.sav 的資訊重新組織，以讓您為此研究的每一個區間事件機率而非只是研究目的事件機率建立模式。它已在別處 (Collett et al。, 2003) 出現和分析過。
- **verd1985.sav。** 本資料檔有關於一項調查 (Verdegaal, 1985)。在調查中記錄了來自 15 個受訪者對 8 個變數的回應。所需的變數被分成三組。集 1 包括 age 和 marital，集 2 包括 pet 和 news，集 3 包括 music 和 live。Pet 調整為多重名義量數，age 調整為次序量數，其他的變數調整為單一名義量數。

- **virus.sav。** 這是有關一家網際網路服務提供者致力於在其網路上判斷病毒之影響的假設資料檔。他們在其網路上追蹤從發現病毒直到控制威脅的這段時間，被病毒感染之電子郵件的流量（約略）百分比。
- **wheeze_steubenville.sav。** 這是空氣污染對兒童健康之影響 (Ware, Dockery, Spiro III, Speizer, 和 Ferris Jr., 1984) 縱向研究的子集。本資料包含來自俄亥俄州 Steubenville，年齡 7、8、9 和 10 歲兒童的氣喘聲狀態之重複二元測量，以及其母親在本研究的第一年是否抽煙的固定記錄。
- **workprog.sav。** 這是有關一項政府職業計劃，設法將弱勢民眾安置到較好之工作的假設資料檔。一個樣本的可能計劃參與者被追蹤，他們之中某些被選取加入本計劃，而其他的則否。每一個觀察值代表一位不同的計劃參與者。

附錄
B

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.

參考書目

Bell, E. H. 1961. Social foundations of human behavior: (人類行為之社會基礎)Introduction to the study of sociology (社會學研究簡介). 紐約: Harper & Row.

Blake, C. L., 和 C. J. Merz. 1998. "UCI Repository of machine learning databases (機器學習資料庫 UCI 儲存器)." Available at http://www.ics.uci.edu/~mlearn/MLRepository.html.

Breiman, L., 和 J. H. Friedman 檢定(F). 1985. Estimating optimal transformations for multiple regression and correlation (估計多重迴歸與相關之最適轉換). Journal of the American Statistical Association (美國統計協會彙報), 80, .

Cochran, W. G. 1977. Sampling Techniques (取樣技巧), 第三版 ed. 紐約: John Wiley and Sons.

Collett, D. 2003. Modelling survival data in medical research (模式化醫學研究中的存活資料), 2 ed. Boca Raton: Chapman & Hall/CRC.

Cox, D. R., 和 E. J. Snell. 1989. The Analysis of Binary Data (二元資料分析), 第二版 ed. 倫敦: Chapman and Hall.

Green, P. E., 和 V. Rao. 1972. Applied multidimensional scaling (應用多元尺度方法). Hinsdale, Ill.: Dryden Press (Dryden 出版社).

Green, P. E., 和 Y. Wind. 1973. Multiattribute decisions in marketing: (市場行銷之多重屬性決策)A measurement approach (測量方法). Hinsdale, Ill.: Dryden Press (Dryden 出版社).

Guttman 值, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points (尋找組態點最小座標空間之一般非計量技巧). Psychometrika (心理學計量報導), 33, .

Hartigan, J. A. 1975. Clustering algorithms (集群演算法). 紐約: John Wiley and Sons.

Hastie, T., 和 R. Tibshirani. 1990. Generalized additive models (概化附加模式). 倫敦: Chapman and Hall.

Kennedy, R., C. Riquier, 和 B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research (市場研究類別資料之對應分析實際應用). Journal of Targeting, Measurement, and Analysis for Marketing (市場行銷之目標訂定、測量與分析雜誌), 5, .

Kish, L. 1965. Survey Sampling (調查取樣). 紐約: John Wiley and Sons.

Kish, L. 1987. Statistical Design for Research (研究的統計設計). 紐約: John Wiley and Sons.

McCullagh, P., 和 J. A. Nelder. 1989. 概化線性模式, 第二版 ed. 倫敦: Chapman & Hall.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior (性質選擇行為之條件式 logit 分析). 於: Frontiers in Economics (經濟新疆界), P. Zarembka, ed. 紐約: Academic Press (學術出版社).

Murthy, M. N. 1967. Sampling Theory and Methods (取樣理論與方法). Calcutta, India (印度加爾各達): Statistical Publishing Society (統計出版社).

Nagelkerke, N. J. D. 1991. A note on the general definition of the coefficient of determination (判定係數之一般定義小記). Biometrika (生物統計學雜誌), 78:3, .

Price, R. H., 和 D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior (行為適切性與情境限制作為社會行為維度). Journal of Personality and Social Psychology (人格與社會心理學雜誌), 30, .

Rickman, R., N. Mitchell, J. Dingman, 和 J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet (實行 Stillman 飲食法期間血膽固醇改變情形). Journal of the American Medical Association (美國醫學協會彙報), 228, .

Rosenberg, S., 和 M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research (多變量研究中作為資料收集程序之排序方法). Multivariate Behavioral Research (多變量行為研究), 10, .

Särndal, C., B. Swensson, 和 J. Wretman. 1992. Model Assisted Survey Sampling (模式輔助調查取樣). 紐約: Springer-Verlag.

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

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

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

索引