

IBM SPSS Decision Trees 26

IBM

附註

使用此資訊和支援的產品之前，請先閱讀 第 17 頁的『注意事項』中的資訊。

產品資訊

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

目錄

決策樹狀結構	1	輸出	11
建立決策樹狀結構	1	注意事項	17
選取類別	3	商標	18
驗證	4	索引	19
樹狀結構成長條件	5		
選項	7		
儲存模式資訊	10		

決策樹狀結構

下列「決策樹狀結構」功能包含在 SPSS® Statistics Professional Edition 或「決策樹狀結構」選項中。

建立決策樹狀結構

「決策樹狀結構」程序會建立樹狀結構的分類模式。它會根據自變數（預測值）的值，將個案分組，或因變數（目標）的預測值。這個程序會提供用於解釋與確認分類分析的驗證工具。

這個程序可以用於：

分段。 識別有可能是特殊群組成員的人員。

分層。 將觀察值指定給其中一個類別，例如高風險、中風險與低風險群組。

預測。 建立規則並用這些規則來預測未來的事件，例如某人可能會借貸，或是汽車或房屋的潛在重新銷售值。

資料縮減與變數篩檢。 從一個大型的變數集中選取一個有用的預測值子集，用於建立一個正式的參數模式。

交互作用識別。 識別只與特定子群組有關的關係並在正式的參數模式中指定這些項目。

類別合併與離散化連續的變數。 使用遺失最少資訊的方式將組別預測值類別與連續變數重新編碼。

範例。 假設某家銀行打算根據信用申請人是否有合理的信用風險，來將這些申請人加以分類。根據各種因素，包括過去客戶的已知信用評等，您就可以建立一個模式來預測未來的客戶是否有可能進行借貸。

樹狀結構分析會提供一些引人注意的功能：

- 它可以讓您識別具有高風險或低風險的同質組別。
- 它可以更容易建構有關個別觀察值進行預測的規則。

資料考量

資料。 因變數和自變數可以是：

- **名義。** 當變數值代表實質上並未等級化的種類時（例如，有員工工作的公司部門），則此變數可視為名義。名義變數的範例包括區域、郵遞區號及宗教團體。
- **序數。** 當變數值代表實質上已等級化的種類時（例如，服務滿意度從非常不滿意到非常滿意分級），則此變數可視為序數。序數變數的範例包括代表滿意度或信賴程度的態度分數以及偏好等級分數。
- **尺度。** 如果某一變數可視為尺度（連續），表示它的值代表含有實際意義矩陣的已排序種類，因此適合比較值之間的距離。尺度變數的範例包括以年份表示的年齡及以千元為單位的收入。

頻率加權 如果加權生效的話，則分數加權就會捨入為最接近的整數，所以，加權值少於 0.5 的觀察值就會被指定一個 0 的加權，進而從分析中被排除在外。

假設。 這個程序會假設已經將適當的測量層級指定給所有分析變數，而且某些功能會假設包含在分析中的因變數的所有值都已經定義數值標籤。

- **測量層級。** 測量層級會影響樹狀結構計算作業，因此所有變數都應該指定適當的測量層級。根據預設，數值變數是假設為尺度變數而字串變數則假設為名義變數，它們可能無法精確反映真正測量的層級。變數清單中各變數旁圖示會指明變數類型。

您可以藉由在來源變數清單按一下滑鼠右鍵，從蹦現功能表選取測量層級，暫時變更變數的測量層級。

- **值標籤。** 此程序的對話框介面假設類別（名義、序數）因變數的「所有」非遺漏值均已定義數值標籤，或者「沒有一個」非遺漏值已定義。有些功能至少需要類別因變數的兩個遺漏值具有數值標籤，否則無法使用。如果至少兩個非遺漏值已定義數值標籤，當有任何觀察值具有其他無數值標籤的數值時，該觀察值會從分析中排除。

取得決策樹狀結構

1. 在功能表上，選擇：

分析 > 分類 > 樹狀結構...

2. 選取應變數。
3. 選取一個或多個自變數。
4. 選取一個成長方法。

您可以：

- 變更來源清單中任何變數的測量層級。
- 強制自變數清單中的第一個變數進入模式中當作第一個分割變數。
- 選取定義觀察值影響樹狀結構成長過程之程度的影響變數。觀察值的影響變數數值較低，則影響力較小；反之則影響力較大。影響變數值必須為正數。
- 驗證樹狀結構。
- 自訂樹狀結構成長條件。
- 將終端節點數、預測值以及預測機率另存成變數。
- 以 XML (PMML) 格式儲存模式。

具有未知測量層級的欄位

若在資料集中出現一或多個不明的變數（欄位）測量層次，就會顯示「測量層次」警示。由於測量層級會影響此程序的結果計算，因此所有變數皆必須具有已定義的測量層級。

掃描資料

讀取作用中資料集的資料，並且針對目前具有未知測量層級的任何欄位指派預設的測量層級。若為大型資料集，則讀取時可能需要一些時間。

手動指派

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

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

變更測量層次

1. 在來源清單的變數上按一下滑鼠右鍵。
2. 選取蹦現功能表上的測量層級。

這會暫時變更測量層級，以供在「決策樹狀結構」程序中使用。

成長方法

可用的成長方法包括：

CHAID

卡方自動互動偵測。CHAID 會在每個步驟中，選擇與因變數具有最強交互作用的自 (預測) 變數。如果每個預測變數的種類相對於應變數沒有明顯不同，則會合併這些種類。

耗盡的 CHAID

對 CHAID 的修改，它會檢查每個預測值所有可能的分割。

CRT 分類與迴歸樹狀結構。CRT 會盡量將資料分割成與因變數相關的同質資料區段。所有觀察值皆具有相同應變數值的終端節點，是「純」同質節點。

QUEST

快速、不偏且有效的統計之樹狀結構。此方法不但計算快速，而且能避免如其他方法偏好有許多類別的預測變數。只有在名義應變數才能指定 QUEST。

每一個方法都有其優點與限制，包括：

表 1. 增長方法的特性.

功能	CHAID*	CRT	QUEST
以卡方分配為基礎**	X		
代理自 (預測值) 變數		X	X
樹狀結構修正		X	X
多因素節點分割	X		
二元節點分割		X	X
影響變數	X	X	
事前機率		X	X
錯誤分類成本	X	X	X
快速計算	X		X

*包括 Exhaustive CHAID。

**QUEST 也會將卡方量數用於名義自變數。

選取類別

如果是類別 (名義、序數) 因變數，您可以：

- 控制要包含在分析中的類別。
- 識別相關的目標類別。

包括/排除種類

您可以將分析限制為因變數的特定類別。

- 在「排除」清單中之因變數值的個案不會包含在分析中。
- 如果是名義因變數，您也可以包含使用者遺漏的類別 (依預設，使用者遺漏的類別會顯示在「排除」清單中)。

目標類別

已選取的 (已核取的) 類別會被視為分析中主要相關的類別。例如，如果您主要是要識別最有可能借貸的個人，您可以選取信用評等類別「差」來當作目標類別。

- 沒有預設的目標類別。如果沒有選取任何類別，某些分類規則選項與獲利相關選項就無法使用。

- 如果選取多個類別，就會為每一個目標類別產生個別的獲利表與圖表。
- 將一個或多個類別指定為目標類別，對於樹狀模式、風險估計或錯誤分類結果並不會有任何影響。

類別

這個對話框需要應變數的已定義數值標籤。除非類別因變數的至少兩個值已經定義數值標籤，否則無法使用這個對話框。

包括/排除種類以及選取目標種類

1. 在主要「決策樹狀結構」對話框中，選取具有兩個或多個數值標籤的種類（名義、序數）應變數。
2. 按一下「類別」。

驗證

驗證可以讓您評估樹狀結構對大型母群概化的程度。有兩種驗證方法可供使用：交叉驗證和分割樣本驗證。

交叉驗證

交叉驗證會將樣本分成子樣本數或折疊。接著會產生樹狀結構模型，然後從每個子樣本中排除資料。第一個樹狀結構是以第一個樣本折疊中以外的所有觀察值為基礎，第二個樹狀結構是以第二個範例折疊中以外的所有觀察值為基礎，依此類推。對於每一個樹狀結構，都會藉由將樹狀結構套用至在產生時被排除的子樣本來評估錯誤分類風險。

重要：選取修正時，CRT 與 QUEST 方法無法使用交叉驗證。

- 您最多可以指定 25 個樣本折疊。如果值越高，則從每一個樹狀結構模型中排除的觀察值數量也就越少。
- 交叉驗證會產生一個單一、最終的樹狀結構模型。最終樹狀結構的交叉驗證風險評估是計算所有樹狀結構風險的平均值。

分割樣本驗證

透過分割樣本驗證，就可以使用訓練樣本來產生模式，以及在保留樣本上測試模式。

- 您可以指定一個訓練樣本大小（以總樣本大小的百分比表示），以及指定一個會將樣本分割成訓練與測試樣本的變數。
- 如果您使用變數來定義訓練與測試樣本，則具有變數值 1 的觀察值就會被指定給訓練樣本，而所有其他觀察值則會被指定給測試樣本。變數不可以是因變數、加權變數、影響變數或強制自變數。
- 您可以同時顯示訓練樣本與測試樣本，或是僅顯示測試樣本。
- 在小型資料檔（含有少量觀察值的資料檔）上使用分割樣本驗證時，必須小心。小型的訓練樣本大小可能會產生品質不佳的模式，因為這些樣本大小在某些類別中可能沒有足夠的觀察值，所以無法適當地讓樹狀結構成長。

驗證決策樹狀結構

1. 在主要「決策樹狀結構」對話框中，按一下驗證。
2. 選取「交叉驗證」或「分割樣本驗證」。

註：兩種驗證方法都會隨意將觀察值指定給樣本組別。如果想要能夠在後續的分析中重製完全一樣的結果，則在第一次執行分析之前，應該要設定亂數種子（「轉換」功能表、「亂數產生器」），然後再為後續的分析將種子重設為該值。

樹狀結構成長條件

可用的成長條件是根據成長方法、因變數的測量層級或兩者的組合而定。

成長限制

「成長限制」對話框可以讓您限制樹狀結構中的層次數量，以及控制個上層節點與子節點的最小觀察值數量。

樹狀結構深度上限

控制根節點底下成長的最大層級數量。「自動」設定會將樹狀結構限制在 CHAID 與 Exhaustive CHAID 方法之根節點底下的三個層級數量，以及 CRT 與 QUEST 方法的五個層級數量。

最小觀察值個數

控制節點的最小觀察值個數。沒有符合這些條件的節點將不會被分割。

- 如果增加最小值的數字，將會產生具有較少節點的樹狀結構。
- 如果減少最小值的數字，將會產生具有較多節點的樹狀結構。

對於具有少數觀察值的資料檔，上層節點的預設觀察值 100 與子節點的預設觀察值 50 有時候可能會在根節點底下產生不具有任何節點的樹狀結構；在這種情況下，如果減少最小值的數字，可能會產生更多有用的結果。

指定成長限制

1. 在主要「決策樹狀結構」對話框中，按一下**成長限制**。

CHAID 條件

對於 CHAID 與 Exhaustive CHAID 方法，您可以控制：

顯著性層次

您可以控制用於分割節點與合併類別的顯著值。對這兩個條件而言，預設的顯著性層級都是 0.05。

分割節點

該值必須大於 0 並小於 1。較小的值會產生節點較少的樹狀結構。

合併類別

該值必須大於 0 並小於或等於 1。若要避免合併種類，請指定一個 1 的值。如果是尺度自變數，這是表示最終樹狀結構中變數種類的數量就是區間的指定數量（預設值為 10）。如需相關資訊，請參閱主題第 6 頁的『CHAID 分析的尺度區間』。

卡方統計量

如果是序數因變數，則用於決定節點分割與類別合併的卡方，就是使用概似比方法所計算的。如果是序數因變數，您可以選擇方法：

Pearson

這個方法會提供更快速的計算，但是用在小型樣本時則必須小心。此為預設的方法。

概似比

這個方法比 Pearson 方法更為穩健，但是在計算時必須花費較多的時間。對於小型的樣本，這是較佳的方法。

模型估計值

對於名義與序數因變數，您可以指定：

最大疊代數

預設值是 100。如果樹狀結構因為已經到達最大疊代次數而停止成長，您可以增加最大值的數字，或是變更一個或兩個控制樹狀結構成長的條件。

預期資料格次數中的最小變更

值必須大於 0 並且小於 1。預設值為 0.05。較低的值會產生具有較少節點的樹狀結構。

使用 Bonferroni 方法調整顯著性值

對於多重比較而言，用於合併與分割條件的顯著值是透過使用 Bonferroni 方法所調整。此為預設值。

允許重新分割在節點內合併的種類

除非您明確地防止類別進行合併，否則程序會嘗試一起合併自 (預測值) 變數類別以產生會描述模式之最簡單的樹狀結構。這個選項允許程序重新分割已經合併的類別 (如果那樣會提供更佳解答的話)。

指定 CHAID 條件

1. 在主要「決策樹狀結構」對話框中，選取 **CHAID** 或消耗 **CHAID** 作為增長方法。
2. 按一下 **CHAID**。

CHAID 分析的尺度區間： 在 CHAID 分析中，在進行分析之前，尺度自 (預測值) 變數一定會先被分成離散的組別 (例如 0-10、11-20、21-30 等)。您可以控制群組的初始/最大數 (即使程序可能會在初始分割之後合併連續的組別)：

固定數目

所有的尺度自變數一開始都會被分成相同的組別個數。預設值是 10。

自訂 每一個尺度自變數一開始都會被分為該變數指定的組別數。

指定尺度自變數的區間

1. 在主要「決策樹狀結構」對話框中，選取一個或多個尺度自變數。
2. 如果是成長方法，請選取 **CHAID** 或 **Exhaustive CHAID**。
3. 按一下區間。

在 CRT 與 QUEST 分析中，所有的分割都是二元分割，且尺度與序數自變數都是以相同方式處理，所以，您無法為尺度自變數指定區間數。

CRT 條件

CRT 成長方法會嘗試最大化節點內的同質性。如果某個節點無法表示觀察值的同質子集，就表示有雜質。例如，如果某個終端節點中的所有個案都具有因變數的相同值，由於該個案為純同質節點，因此不需要進一步分割。

您可以選取用來測量雜質的方法，以及進行分割節點時所需的最小雜質減少量

雜質測量

如果是尺度因變數，會使用雜質的最小平方差 (LSD) 測量，它會計算節點內變異數，並為任何頻率加權或影響值調整。如果是類別 (名義、序數) 因變數，您可以選取雜質測量：

Gini 會找到與因變數值相關之子節點同質性最大化的分割。Gini 是根據因變數之每個類別成員的平方機率為基礎。當節點中的所有觀察值都落在單一個類別中時，它就會達到最小值 (零)。此為預設的測量。

Twoing

應變數的種類會被分成兩個子類別群組。會找到能夠以最佳方式分開兩個組別的分割。

排序的 Twoing

與 Twoing 類似，差別在於只有相鄰類別才可以加分組。這個測量只能用於序數因變數。

最小改進變更

這是分割節點時所需的最小雜質減少量。預設值是 0.0001。較高的值會產生具有較少節點的樹狀結構。

指定 CRT 準則

1. 如果是成長方法，請選取 **CRT**。
2. 按一下 **CRT**。

QUEST 條件

如果是 QUEST 方法，您可以指定用於分割節點的顯著性層級。除非顯著性層級小於或等於指定的層級，否則您無法使用自變數來分割節點。值必須大於 0 並且小於 1。預設值為 0.05。較小的值會從最終模型中排除更多的自變數。

指定 QUEST 準則

1. 在主要「決策樹狀結構」對話框中，選取名義應變數。
2. 如果是成長方法，請選取 **QUEST**。
3. 按一下 **QUEST**。

修正樹狀結構

您可以使用 CRT 和 QUEST 方法避免透過修剪樹狀結構，過度適配模型：在達到停止準則時，樹狀結構會就會停止成長，然後樹狀結構會根據風險中所指定的最大差異自動修正為最小的子樹狀結構。風險值是以標準誤來表示。預設值為 1，而且必須是正數。若要取得具有最低風險的子樹狀結構，請指定 0。

重要：選取修正時，CRT 與 QUEST 方法無法使用交叉驗證。

修正樹狀結構

1. 在主要「決策樹狀結構」對話框中，如果是成長方法，請選取 **CRT** 或 **QUEST**。
2. 按一下刪改。

修正與隱藏節點的比較

當您建立已經修正的樹狀結構時，從樹狀結構中所修正的任何節點都無法用在最終樹狀結構中。您可以使用互動方式來隱藏或顯示在最終樹狀結構中所選擇的子節點，但是您無法顯示在樹狀結構建立過程中所修正的節點。

代理

CRT 與 QUEST 可以將代理用於自（預測值）變數。對於該變數值已遺漏的觀察值而言，會使用其他具有與原始變數高度關聯的自變數來進行分類。這些替代的預測值稱為代理。您可以指定要用在模式中的最大代理數。

- 根據預設，最大的代理數是自變數的數量減去1 的數字。換句話說，對於每一個自變數，所有其他的自變數都可以當作代理來使用。
- 如果您不希望模式使用代理，請為代理數指定 0。

指定代理

1. 在主要「決策樹狀結構」對話框中，如果是成長方法，請選取 **CRT** 或 **QUEST**。
2. 按一下代理。

選項

可以使用的選項會因為成長方法、因變數的測量層級，及/或是否有因變數之值的已定義數值標籤等而有所不同。

錯誤分類成本

如果是類別（名義、序數）因變數，則錯誤分類成本可以讓您包含與不正確分類相關的相對懲罰資訊。例如：

- 拒絕信用良好之客戶所花費的成本，可能不同於擴展日後會借貸之客戶的信用所花費的成本。
- 將具有高度心臟疾病風險的人員錯誤分類為具有低度心臟疾病風險人員所付出的成本，可能會高於將具有低度心臟疾病風險的人員錯誤分類為具有高度心臟疾病風險人員所付出的成本還要高。
- 將大量郵件傳送給不太可能回應的人，成本可能比較低，但是如果將大量郵件傳送給可能會回應的人，成本相對上可能會比較高（以損失的收益而言）。

註：除非種類應變數至少有兩個值已經定義數值標籤，否則無法使用這個「錯誤分類成本」對話框。

指定錯誤分類成本

1. 在主要「決策樹狀結構」對話框中，選取具有兩個或多個數值標籤的種類（名義、序數）應變數。
2. 按一下**錯誤分類成本**標籤。
3. 按一下「自訂」。
4. 在「預測種類」網格中輸入一個或多個錯誤分類成本。輸入的值必須是非負數。（正確的分類會顯示在對角線上，而且一定是 0）。

填滿矩陣

在許多情況中，您可能會想要讓成本變成對稱 也就是說，將 A 錯誤分類為 B 所用的成本，與將 B 錯誤分類為 A 所用的成本是一樣的。以下的控制可以讓您更輕易地指定對稱的成本矩陣：

複製下半三角形

將下半三角形矩陣（對角線底下）的值複製到對應的上三角形儲存格中。

複製上半三角形

將矩陣下半三角形（對角線底下）內的值複製到對應的上三角形儲存格中。

使用資料格平均值

對於每一半矩陣的每一個儲存格而言，兩個值（上與下三角形）是相加之後的平均值，而這個平均值會取代原來的兩個值。例如，如果將 A 錯誤分類為 B 所付出的成本為 1，而將 B 錯誤分類為 A 所付出的成本為 3，則這個控制會使用平均值 2 ($(1+3)/2 = 2$) 來取代原來的兩個值。

利潤

如果是類別因變數，您可以指定因變數層級的收益與支出值。

- 利潤是以收益減去支出的方式來計算。
- 利潤值會影響獲利表中利潤與 ROI（投資報酬率）的值，但是不會影響基本樹狀結構的模式結構。
- 收益與支出值都必須是數值，而且也都必須指定給網格中所顯示之因變數的所有類別。

註：這個對話框需要應變數的已定義數值標籤。除非類別因變數的至少兩個值已經定義數值標籤，否則無法使用這個對話框。

指定利潤

1. 在主要「決策樹狀結構」對話框中，選取具有兩個或多個數值標籤的種類（名義、序數）應變數。
2. 按一下**利潤**。
3. 按一下「自訂」。
4. 為網格中所列出的所有因變數類別輸入收益與支出值。

事前機率

如果是具有類別因變數的 CRT 與 QUEST 樹狀結構，您可以指定組別成員的事前機率。**事前機率** 就是在瞭解自變數（預測值）之前，對因變數之每一個類別總體相對次數的估計。使用事前機率可以協助更正由非整體母群之取樣中的資料所造成的任何樹狀結構成長情況。

從訓練樣本取得（事前經驗）

如果資料檔中的因變數值分配是表示母群分配，請使用這個設定。如果您是使用分割樣本驗證，就會使用訓練樣本中的觀察值分配。

註：由於觀察值是隨機指定給分割樣本驗證中的訓練樣本，因此無法事先知道訓練樣本中實際的觀察值分配。如需相關資訊，請參閱主題第 4 頁的『驗證』。

各種類都相等

如果因變數的類別在母群中都是顯示為相等，請使用這個設定。例如，如果一共有四個類別，則每各類別中都會有大約 25% 的觀察值。

自訂 為網格中所列的每一個因變數類別輸入一個非負數值。值可以是比例、百分比、次數個數，或是在所有類別中表示數值分布的其他值。

使用錯誤分類成本調整事前

如果您定義自訂錯誤分類成本，就可以根據這些成本來調整事前機率。如需相關資訊，請參閱主題第 8 頁的『錯誤分類成本』。

註：這個對話框需要應變數的已定義數值標籤。除非類別因變數的至少兩個值已經定義數值標籤，否則無法使用這個對話框。

指定事前機率

1. 在主要「決策樹狀結構」對話框中，選取具有兩個或多個數值標籤的種類（名義、序數）應變數。
2. 如果是**成長方法**，請選取 **CRT** 或 **QUEST**。
3. 按一下**事前機率**。

分數

如果是具有序數因變數的 CHAID 與 Exhaustive CHAID，您可以自訂因變數之每一個類別的分數。分數會定義因變數各類別之間的順序與距離。您可以使用分數來增加或減少序數值之間的相對距離，或是變更值的順序。

各種類都使用序數等級

因變數的最低類別會被指定一個 1 的分數，下一個較高的類別會被指定一個 2 的分數，依此類推。此為預設值。

自訂 為網格中所列的每一個因變數類別輸入一個數值分數。

註：這個對話框需要因變數的已定義數值標籤。除非類別因變數的至少兩個值已經定義數值標籤，否則無法使用這個對話框。

範例

表 2. 自訂評分值

數值標籤	原始值	分數
非技術人員	1	1
技術人員	2	4
事務人員	3	4.5

表 2. 自訂評分值 (繼續)

數值標籤	原始值	分數
Professional	4	7
管理人員	5	6

- 分數會增加非技術人員與技術人員之間的相對距離，而且會減少技術人員與 事務人員之間的相對距離。
- 分數會將管理人員與專業人員的順序反轉。

指定評分

1. 在主要「決策樹狀結構」對話框中，選取具有兩個或多個已定義值標籤的序數應變數。
2. 如果是成長方法，請選取 **CHAID** 或 **Exhaustive CHAID**。
3. 按一下評分。

遺漏值

「遺漏值」對話框會控制名義值、使用者遺漏值與自（預測值）變數值的處理方式。

- 序數與尺度使用者遺漏的自變數值的處理方式會因為成長方法而有所不同。
- 名義應變數的處理方式是在「種類」對話框中所指定。如需相關資訊，請參閱主題 第 3 頁的『選取類別』。
- 如果是序數與尺度因變數，一定會排除具有系統遺漏或使用者遺漏的因變數值的觀察值。

名義自變數的使用者遺漏值

視為遺漏值

使用者遺漏值會被視為系統遺漏值來處理。系統遺漏值的處理方式會因為成長方法而有所不同。

視為無效值

名義自變數的使用者遺漏值會被視為樹狀結構成長與分類中的普通值來處理。

方法相依規則

如果某些（非全部）自變數值是系統遺漏值或使用者遺漏值：

- 如果是 CHAID 與 Exhaustive CHAID，系統遺漏與使用者遺漏的自變數值會以單一、組合的類別包含在分析中。如果是尺度與序數自變數，則演算法會先使用有效的值來產生類別，然後決定是否要將遺失的類別和其最相似的（有效的）類別加以合併，或是將其維持為一個個別的類別。
- 如果是 CRT 與 QUEST，則具有遺失自變數值的觀察值會從樹狀結構成長過程中被排除，但是會使用代理來加以分類（如果方法中含有代理的話）。如果名義使用者遺漏值是被視為遺失值來處理的話，也會使用這個方法來處理這些值。如需相關資訊，請參閱主題第 7 頁的『代理』。

指定名義、自變數使用者遺漏處理

1. 在主要「決策樹狀結構」對話框中，選取至少一個名義自變數。
2. 按一下遺漏值。

儲存模式資訊

您可以將模型的資訊儲存為工作資料檔中的變數，也可以將整個方法以 XML (PMML) 格式儲存至某個外部檔案。

已儲存變數

終端節點數

每個觀察值指定的終端節點。值就是樹狀結構節點數。

預測值

由模式所預測之因變數的類別（群組）或值。

預測機率

與模型的預測相關聯的機率。系統會為因變數的每一個類別儲存一個變數。不適用於尺度因變數。

樣本指派(訓練/測試)

如果是分割樣本驗證，這個變數會指出訓練或測試樣本中是否有使用觀察值。訓練樣本的值為 1，測試樣本的值則為 0。除非您已經選取分割樣本驗證，否則無法使用。如需相關資訊，請參閱主題第 4 頁的『驗證』。

以 XML 格式匯出樹狀模型

您可以將整個樹狀模型儲存為 XML (PMML) 格式。您可以使用這個模型檔案，將模式資訊套用到其他資料檔中以進行評分工作。

訓練樣本

將模式寫入至指定的檔案。如果是分割樣本驗證樹狀結構，這是用於訓練樣本的模式。

檢定樣本

將測試樣本的模式寫入至指定的檔案。除非您已經選取分割樣本驗證，否則無法使用。

輸出

可用的輸出選項依成長方法、因變數的測量層級與其他設定而定。

樹狀結構顯示

您可以控制樹狀結構的初始外觀，或完全隱藏樹狀結構顯示。

樹狀結構

依預設，樹狀結構表是包含在「輸出」標籤中所顯示的輸出中。取消選取這個選項，就可以從輸出中排除樹狀結構圖。

顯示 這些選項會控制「輸出」標籤中樹狀結構圖的初始外觀。所有這些屬性都可以藉由編輯產生的樹狀結構來加以修改。

方向 樹狀結構可以顯示為根節點在頂端，可以從上到下展開、或根節點在左右兩側，可以從左到右或從右到左展開。

節點內容

節點可以顯示表格、圖表，或同時顯示兩者。如果是類別因變數，表格會顯示次數個數與百分比，而圖表則為長條圖。如果是尺度因變數，表格會顯示平均數、標準差、觀察值數與預測的值，而圖表則為直方圖。

尺度(S)

依預設，大型的樹狀結構都會自動調整，嘗試讓樹狀結構能夠符合頁面的大小。您可以指定高達 200% 的自訂尺度百分比。

自變數統計量

如果是 CHAID 與 Exhaustive CHAID，統計量包括 F 值（用於尺度因變數）或卡方值（用於類別因變數，以及顯著值和自由度。如果是 CRT，則會顯示改善值。如果是 QUEST，則會為尺度與序數自變數顯示 F 、顯著值與自由度；如果是名義自變數，則會顯示卡方值、顯著值與自由度。

節點定義

節點定義會顯示每個節點分割使用之自變數的值。

表格格式的樹狀結構

樹狀結構中每個節點的摘要資訊，包括上層節點數、自變數統計量、節點的自變數值、尺度因變數的平均數與標準差，或是類別因變數的個數與百分比。

控制起始樹狀結構顯示

1. 在主要「決策樹狀結構」對話框中，按一下**樹狀結構**。

Statistics

可用的統計量表格根據因變數的測量層級、成長方法與其他設定而定。

模型

摘要 摘要包括使用的方法、模式中所包括的變數，以及模式中所指定但未包括的變數。

風險 風險估計與其標準誤。樹狀結構預測準確性的測量。

- 如果是類別因變數，風險估計就是在事前機率和錯誤分類成本調整之後，不正確分類之觀察值的比例。
- 如果是尺度因變數，風險估計是在節點變異數的範圍中。

分類表

如果是類別（名義、序數）因變數，這個表格就會顯示為每個因變數類別正確分類與不正確分類之的觀察值數。不適用於尺度因變數。

成本、事前機率、分數和利潤值

如果是類別因變數，這個表格會顯示分析中所使用的成本、事前機率、分數與利潤值。不適用於尺度因變數。

自變數(V)

對模型的重要性

如果是 CRT 成長方法，則會根據其對模型的重要性來將每個自（預測值）變數分等。不適用於 QUEST 或 CHAID 方法。

依分割列出代理變數

對於 CRT 與 QUEST 成長方法，如果模式包括代理，則會列出樹狀結構中每個分割的代理。不適用於 CHAID 方法。如需相關資訊，請參閱主題 第 7 頁的『代理』。

節點效能

摘要 如果是尺度因變數，這個表格會包括節點數、觀察值數，以及因變數的平均值。如果是具有已定義之利潤的類別因變數，則表格會包括節點數、平均利潤以及 ROI（投資報酬率）值。不適用於沒有已定義之利潤的類別因變數。如需相關資訊，請參閱主題第 8 頁的『利潤』。

依目標類別

如果是具有已定義之目標分類的類別因變數，則表格會包括百分比獲利、回應百分比以及根據節點或百分位數組別所區分的指數百分比（提升）。每個目標類別都會產生個別的表格。不適用於沒有已定義之目標類別的尺度因變數或類別因變數。如需相關資訊，請參閱主題第 3 頁的『選取類別』。

列數 節點效能表可以根據終端節點、百分位數或兩者來顯示結果。如果您選取同時使用兩者，每個目標類別就會產生兩個表格。百分位數表會根據排序順序，為每個百分位數顯示累積值。

排序 值根據應變數的測量層次改變，且增益摘要與增益表格不同。

百分位數增量

如果是百分位數表，您可以選取百分位數增量：1、2、5、10、20 或 25。

顯示累積統計量。

如果是終端節點表，會在每個表格顯示更多的欄位，來顯示累積結果。

選取統計量輸出

1. 在主要「決策樹狀結構」對話框中，按一下**統計量**。

圖表

可用的圖表是根據因變數的測量層級、成長方法與其他設定而定。

自變數對模型的重要性

根據自變數（測量值）之模式重要性長條圖。只適合與 CRT 成長方法搭配使用。

節點效能

獲利 「獲利」是指每個節點之目標類別的總觀察值的百分比，計算方式為： $(\text{節點目標 } n / \text{目標總計 } n) \times 100$ 。獲利圖表是獲利累積百分位的折線圖，計算方式為： $(\text{累積百分位目標 } n / \text{目標總計 } n) \times 100$ 。每個目標類別都會產生分隔線圖表。只適用於有定義之目標類別的類別因變數。如需相關資訊，請參閱主題第 3 頁的『選取類別』。

獲利圖表會繪製與您在百分位數表之獲利的「獲利百分比」行中所見相同的值，這個百分位數表也會報告累積值。

指數 指標為目標種類之節點回應值百分比與整個樣本之整體目標種類回應值百分比相較之下，所得出的比例。指數圖表就是累積百分位數指數值的線性圖。僅適用於類別因變數。累積百分位指數的計算方式為： $(\text{累積百分位回應百分比} / \text{回應百分比總計}) \times 100$ 。每個目標類別都會產生獨立圖表，且必須定義目標類別。

指數圖表會繪製與您在百分位數表之獲利的「指數」行中所見相同的值。

回應 指定的目標種類中，節點內的觀察值百分比。回應圖表就是累積百分位數回應的折線圖，計算方式為： $(\text{累積百分位目標 } n / \text{累積百分位總計 } n) \times 100$ 。只適用於有定義之目標類別的類別因變數。

回應圖表會繪製與您在百分位數表之獲利的「回應」行中所見相同的值。

平均值

因變數的累積百分位數平均值線性圖。僅適用於尺度因變數。

平均利潤

累積平均利潤的線性圖。僅適用於具有已定義之利潤的類別因變數。如需相關資訊，請參閱主題第 8 頁的『利潤』。

平均利潤圖表會繪製與您在百分位數表之獲利摘要的「利潤」行中所見相同的值。

投資報酬率 (ROI)

累積的 ROI（投資報酬）的線性圖。ROI 是以利潤對支出的比率來計算。僅適用於具有已定義之利潤的類別因變數。

ROI 圖表會繪製與您在百分位數表之獲利摘要的「ROI」行中所見相同的值。

百分位數增量

對於所有的百分位數圖表，這個設定會控制圖表上所顯示的百分位數增量：1、2、5、10、20 或 25。

選取圖表輸出

1. 在主要「決策樹狀結構」對話框中，按一下圖表。

選項與評分規則

「規則」對話框會提供以指令語法、SQL 或範例（純英文）文字等形式來產生選項或分類/預測規則的功能。您可以在「輸出」標籤中顯示這些規則及/或將這些規則儲存至某個外部檔案。

產生分類規則

選取此項以啟用選項與評分規則設定。

語法 控制在「輸出」標籤中顯示之輸出以及儲存為外部檔案兩者的選擇規則。

SPSS Statistics

指令語法語言。規則是以定義用於選取觀察值子集之過濾條件的一組指令來表示，或以用於為觀察值評分的 COMPUTE 陳述式來表示。

SQL 標準的 SQL 規則是用來從資料庫中選取或擷取記錄，或是將值指定給這些記錄。產生的 SQL 規則不包含任何表格名稱或其他資料來源資訊。

簡單文字

純英文虛擬程式碼。規則表示為一組邏輯 "if...then" 陳述式，這一組陳述式可以描述模型的分類或每一個節點的預測。這種形式的規則可以用來定義變數和數值標籤或變數名稱和資料值。

類型(T)

若是 IBM® SPSS Statistics 和 SQL 規則，可控制所產生規則的類型：選擇或評分規則。

指派值給觀察值

此規則可用來指定模型的預測給符合節點成員資格條件的觀察值。另外會為符合節點成員資格條件的各節點產生不同的規則。

選取觀察值

此規則可用來選取符合節點成員資格條件的觀察值。有關 IBM SPSS Statistics 或 SQL 規則，會產生單一規則，以選取符合選擇條件的所有觀察值。

在 SPSS Statistics 和 SQL 規則中納入代理

您可以在 CRT 和 QUEST 中，包含規則中模型的代理預測值。包含代理的規則可能會相當複雜。一般來說，如果只要推導有關樹狀結構的概念資訊，請排除代理。如果有些觀察值有不完整的自變數（預測值）資料，而您要模擬樹狀結構的規則，請包含代理。如需相關資訊，請參閱主題 第 7 頁的『代理』。

節點 控制產生規則的範圍。個別的規則會針對範圍中所包括的每個節點而產生。

所有終端節點

為每個終端節點產生規則。

最佳終端節點

根據指數值，為前 n 個終端節點產生規則。如果數量超過樹狀結構中終端節點的數量，就會為所有的終端節點產生規則

最佳終端節點會往上移至指定的觀察值百分比。

根據指數值，為前 n 百分比觀察值的終端節點產生規則

其指數值符合或超過分割值的終端節點。

為具有指數值大於或等於指定值的所有終端節點產生規則。大於 100 的指數值是表示該節點中目標類別內的觀察值百分比已經超過根節點中的百分比

所有節點

為所有節點產生規則。

附註：

- 以指數值為根據的節點選項功能僅適用於有已定義之目標類別的類別應變數。如果您已經指定多個目標類別，就會為每個目標類別產生一組個別的規則。
- 如果是用於選擇觀察值的 IBM SPSS Statistics 與 SQL 規則（不是用於指定值的規則），則**所有節點與所有終端節點**將會有效率地產生可以選擇分析中所使用之所有觀察值的規則。

匯出規則至檔案

將規則儲存在某個外部文字檔案中。

您也可以根據最終樹狀結構模型中已經選擇的節點，以互動方式產生並儲存選項或評分規則。

註：如果您將指令語法格式的規則套用到另一個資料檔，則該資料檔必須包含與最終模型中之自變數相同名稱、使用相同之單位測量，並且有使用者定義遺漏值（如果有的話）的變數。

指定選項或評分規則

1. 在主要「決策樹狀結構」對話框中，按一下規則。

注意事項

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

IBM 可能並未在其他國家提供在本文件中討論到的產品、服務或功能。有關目前在 貴地區可供使用的產品與服務相關資訊，請洽您當地的 IBM 服務代表。對於 IBM 產品、程式或服務的任何參考，目的並不是要陳述或暗示只能使用 IBM 產品、程式或服務。任何功能相等且未侵犯 IBM 智慧財產權的產品、程式或服務皆可使用。但是，評估及確認任何非 IBM 產品、程式或服務的操作之責任應由使用者承擔。

IBM 可能有一些擁有專利或專利申請中的項目包含本文件所描述的內容。本文件的提供並不表示授與您對於這些專利的權利。您可以將書面的授權查詢寄至：

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US*

對於與雙位元組 (DBCS) 資訊相關的授權查詢，請與貴國的 IBM 智慧財產部門聯絡，或將查詢郵寄至：

*Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan*

International Business Machines Corporation 只依「現況」提供本出版品，不提供任何明示或默示之保證，其中包括且不限於不侵權、可商用性或特定目的之適用性的隱含保證。有些地區不允許特定交易中明示或默示的保固聲明，因此，此聲明或許對您不適用。

此資訊內容可能包含技術失準或排版印刷錯誤。此處資訊會定期變更，這些變更將會納入新版的聲明中。IBM 可能會隨時改善和 / 或變更此聲明中所述的產品和 / 或程式，恕不另行通知。

本資訊中任何對非 IBM 網站的敘述僅供參考，IBM 對該網站並不提供任何保證。該「網站」的內容並非此 IBM 產品的部分內容，使用該「網站」需自行承擔風險。

IBM 可能會以任何其認為適當的方式使用或散佈您提供的任何資訊，無需對您負責。

意欲針對達成以下目的而擁有本程式相關資訊之程式被授權人：(i) 在獨立建立的程式與其他程式 (包括本程式) 之間交換資訊及 (ii) 共用已交換的資訊，應聯絡：

*IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US*

在適當條款與條件之下，包括某些情況下 (支付費用)，或可使用此類資訊。

在本文件中描述的授權程式及其適用之所有授權材料皆由 IBM 在與我方簽訂之 IBM 客戶合約、IBM 國際程式授權合約或任何相等效力合約中提供。

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

本文件所提及之非 IBM 產品資訊，取自產品的供應商，或其發佈的聲明或其他公開管道。IBM 並未測試過這些產品，也無法確認這些非 IBM 產品的執行效能、相容性或任何對產品的其他主張是否完全無誤。有關非 IBM 產品的功能問題應直接洽詢該產品供應商。

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

此資訊包含用於日常企業運作的資料和報表範例。為了儘可能提供完整說明，範例中包含了人名、公司名稱、品牌名稱和產品名稱。這些名稱全為虛構，如與實際人員或企業之名稱有所雷同，純屬巧合。

著作權授權：

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

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

© IBM 2019. 本程式之若干部分係衍生自 IBM 公司的範例程式。

© Copyright IBM Corp. 1989 - 20019. All rights reserved.

商標

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

Adobe、Adobe 標誌、PostScript 以及 PostScript 標誌為 Adobe Systems Incorporated 於美國和 / 或其他國家的註冊商標或商標。

Intel、Intel 標誌、Intel Inside、Intel Inside 標誌、Intel Centrino、Intel Centrino 標誌、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 為 Intel Corporation 或其分公司於美國和其他國家的商標或註冊商標。

Linux 為 Linus Torvalds 於美國和 / 或其他國家的註冊商標。

Microsoft、Windows、Windows NT 和 Windows 標誌為 Microsoft Corporation 於美國和 / 或其他國家的商標。

UNIX 為 The Open Group 於美國和其他國家的註冊商標。

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

索引

索引順序以中文字，英文字，及特殊符號之次序排列。

〔四劃〕

- 分割樣本驗證
 - 樹狀結構 4
- 分數
 - 樹狀結構 9

〔五劃〕

- 加權觀察值
 - 決策樹狀結構中的分數加權 1
 - 用於分割節點的顯著性層級 7

〔六劃〕

- 交叉驗證
 - 樹狀結構 4
- 成本
 - 錯誤分類 8

〔七劃〕

- 利潤
 - 事前機率 9
 - 樹狀結構 8, 12
- 決策樹 1
 - 強制第一個變數進入模式中 1
 - 測量層級 1
 - CHAID 方法 1
 - CRT 方法 1
 - Exhaustive CHAID 方法 1
 - QUEST 方法 1, 7

〔八劃〕

- 法則
 - 建立決策樹狀結構的選項與評分語法 14

〔九劃〕

- 指令語法
 - 建立決策樹狀結構的選項與評分語法 14
- 指數值
 - 樹狀結構 12

- 風險估計
 - 樹狀結構 12

〔十劃〕

- 修正決策樹狀結構
 - 與隱藏節點 7

〔十二劃〕

- 測量層級
 - 決策樹 1

〔十三劃〕

- 亂數種子
 - 決策樹狀結構驗證 4
- 節點數
 - 從決策樹狀結構儲存為變數 10
- 預測的值
 - 從決策樹狀結構儲存為變數 10
- 預測的機率
 - 從決策樹狀結構儲存為變數 10

〔十四劃〕

- 語法
 - 建立決策樹狀結構的選項與評分語法 14

〔十六劃〕

- 樹狀結構 1
 - 分割樣本驗證 4
 - 分數 9
 - 尺度自變數的區間 6
 - 交叉驗證 4
 - 利潤 8
 - 事前機率 9
 - 表格中的樹狀結構目錄 11
 - 指數值 12
 - 限制層級數量 5
 - 風險估計 12
 - 修正 7
 - 控制節點大小 5
 - 控制樹狀結構顯示 11
 - 產生規則 14
 - 終端節點統計量 12
 - 預測值重要性 12
 - 圖表 13

- 樹狀結構 (繼續)
 - 樹狀結構方向 11
 - 遺漏值 10
 - 錯誤分類成本 8
 - 錯誤分類表格 12
 - 儲存模型變數 10
 - 顯示與隱藏分支統計量 11
 - CHAID 成長條件 5
 - CRT 方法 6
- 遺漏值
 - 樹狀結構 10
- 錯誤分類
 - 成本 8
 - 樹狀結構 12

〔十七劃〕

- 隱藏節點
 - 與修正 7

〔十八劃〕

- 雜質
 - CRT 樹狀結構 6

〔二十三劃〕

- 驗證
 - 樹狀結構 4

C

- CHAID 1
 - 分割與合併條件 5
 - 尺度自變數的區間 6
 - 重新分割合併的類別 5
 - 最大疊代 5
 - Bonferroni 法調整 5
- CRT 1
 - 修正 7
 - 雜質測量 6

G

- Gini 6

O

- Ordered Twoing 6

Q

QUEST 1, 7
修正 7

S

SQL
建立選項與評分的 SQL 語法 14

T

Twoing 6



Printed in Taiwan