

IBM SPSS Direct Marketing 20



注意：使用此資訊和支援的產品之前，請先閱讀 注意事項第 98 頁 包含的一般資訊。

若新版本未聲明，則此版本便適用於 IBM® SPSS® Statistics 20 以及之後發行的所有版本和修正。

Adobe 產品的擷取畫面已取得 Adobe Systems Incorporated 之翻印許可。

Microsoft 產品的擷取畫面已取得 Microsoft Corporation 之翻印許可。

授權內容：IBM 資產

Copyright IBM Corporation 1989, 2011.

美國政府使用者有限權利：使用、複製或披露內容皆受 IBM Corp 簽署之 GSA ADP Schedule Contract 限制約束。

序

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

IBM Business Analytics 軟體提供完整、一致且確實的資訊，決策者可信任此資訊，並藉以改善營運績效。包括商業智慧、預測分析、財務績效和策略管理，以及分析應用程式的整合型產品組合，為目前績效提供了清晰、即時且具行動性的前瞻眼界，以及預測未來成果的能力。結合了豐富的業界解決方案、有效實證和專業服務，每種規模的組織都能引爆最高效能，確實自動化執行決策，並且交付更棒的成果。

在這項產品組合中，IBM SPSS Predictive Analytics 軟體有助於組織預測未來事件，並且針對前瞻概念提前行動，創造更棒的營運成果。全球的商業、政府和學術客戶相當倚重 IBM SPSS 技術所帶來的競爭優勢，藉此做為吸引、保有和發展更多客戶，同時降低可能的不實詐欺風險。藉由將 IBM SPSS 軟體併入每天作業，這些組織成為預測型企業 – 足以駕馭決策並使決策自動化處理，以符合營運目標，並且達到可測知的競爭優勢。如需更多資訊，或是聯絡代表人員，請造訪 <http://www.ibm.com/spss>。

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 IBM Corp. 產品在使用上的協助，或是支援硬體環境的安裝說明。若要取得技術支援，請參閱 IBM Corp. 網站內容，網址：<http://www.ibm.com/support>。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。

針對學生用戶的技術支援

如果您是使用任何 IBM SPSS 軟體產品之學生版、學術版或研究套件版本的學生，請參閱適用於學生的特殊線上「[教育解決方案 \(http://www.ibm.com/spss/rd/students/\)](http://www.ibm.com/spss/rd/students/)」頁面。如果您是使用 IBM SPSS 軟體之大學提供副本的學生，請聯絡您大學的 IBM SPSS 產品協調人員。

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。請備妥您的序號以供識別。

訓練研討會

IBM Corp. 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。如需研討會的詳細資訊，請移至 <http://www.ibm.com/software/analytics/spss/training>。

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

1	直效行銷	1
2	RFM 分析	2
	RFM 分數取自交易資料	3
	RFM 分數取自客戶資料	4
	RFM Binning	6
	以交易資料儲存 RFM 分數	8
	以客戶資料儲存 RFM 分數	9
	RFM 輸出	11
3	集群分析	13
	設定	16
4	潛在客戶分析概要	17
	設定	21
	建立類別回應欄位	21
5	郵遞區號回應率	23
	設定	27
	建立類別回應欄位	28
6	傾向機率	30
	設定	34
	建立類別回應欄位	35

7 控制套件檢定 37

部 II：範例

8 交易資料的 RFM 分析 41

交易資料	41
執行分析	41
評估結果	43
合併分數資料和客戶資料	45

9 集群分析 48

執行分析	48
輸出	50
根據集群來選取記錄	57
在集群模式瀏覽器中建立過濾器	57
根據集群欄位值選取記錄	60
摘要	63

10 潛在客戶分析概要 64

資料考量	64
執行分析	64
輸出	66
摘要	68

11 郵遞區號回應率 69

資料考量	69
執行分析	69
輸出	72
摘要	75

12 傾向機率	76
資料考量	76
建立預測模式	76
評估模式	79
套用模式	80
摘要	87
13 控制套件檢定	88
執行分析	88
輸出	90
摘要	90
附錄	
A 範例檔案	91
B 注意事項	98
索引	100

部 1: 使用手冊

直效行銷

「直效行銷」選項提供一組專為改善直效行銷活動結果而設計的工具，可以透過識別人口統計、購買、其他定義不同消費者群組的特性以及鎖定特定群組，藉以最大化回應率。

RFM 分析。 此技術可識別現有客戶中最有可能回應新優惠的客戶。如需詳細資訊，請參閱第 2 頁第 2 章中的 RFM 分析。

集群分析。 這種探索工具旨在顯示資料中的自然分組（或集群）。例如，可以根據不同人口資料和購買特性，識別出不同組別的客戶。如需詳細資訊，請參閱第 13 頁第 3 章中的集群分析。

潛在客戶分析概要。 此方法使用上一個活動或測試活動的結果來建立描述性分析概要。您可以在未來的活動中會使用這些分析概要來鎖定特定的群組。如需詳細資訊，請參閱第 17 頁第 4 章中的潛在客戶分析概要。

郵遞區號回應率。 此方法使用上一個活動的結果來計算郵遞區號回應率。在未來的活動中會使用這些回應率來鎖定特定的郵遞區號。如需詳細資訊，請參閱第 23 頁第 5 章中的郵遞區號回應率。

傾向機率。 此方法使用測試郵件或上一個活動的結果來產生傾向分數。分數會識別哪些聯絡人最有可能回應。如需詳細資訊，請參閱第 30 頁第 6 章中的傾向機率。

控制套件檢定。 此方法會比較行銷活動，瞭解不同包裝或優惠的效果是否有顯著性差異。如需詳細資訊，請參閱第 37 頁第 7 章中的控制套件檢定。

RFM 分析

RFM 分析是一項技術，用來識別最有可能對新優惠方案作出反應的現有客戶。此技術常用在直效行銷上。RFM 分析是根據下列簡單的理論：

- 要識別最有可能對新優惠方案作出反應的現有客戶最重要的因子是 **新近值**。愈在近期曾經購買的客戶會比在過去較久前曾經購買的客戶愈有可能再度購買。
- 第二個最重要的因子是 **頻率值**。在過去曾經購買過較多次數的客戶會比那些較少購買的客戶更可能有反應。
- 第三個最重要的因子是總花費金額，通常稱為 **金融值**。在過去曾花費（所有購買的總和）較多的客戶比花費較少的客戶更可能有反應。

RFM 分析作用方式

- 依據最近一次購買的日期或距最近一次購買的時間間隔指定一新近值分數給客戶。此分數是將新近值簡單的分等級至少量的類別。例如，如果您使用五種類別，對於最近購買日的客戶得到新近值等級為 5，而那些過去購買日最久遠的客戶得到的新近值等級為 1。
- 以類似方式為客戶指定頻率值等級，以較高的值代表較高的購買頻率。例如，在五種類別的等級架構下，購買最頻繁的客戶得到的頻率值等級為 5。
- 最後，為客戶依金融值評定等級，最高的金融值得到最高的等級。繼續這五種類別的範例，花費最多的客戶會得到金融值等級為 5。

結果就是每一個客戶可以有四個分數。新近值、頻率值、金融值和組合的 RFM 分數（它只是將這三項個別分數連接成單一值）。「最好」的客戶（指那些最有可能對優惠方案作出反應者）是那些得到最高組合 RFM 分數者。例如，在五種類別等級下，總共有 125 個可能的組合 RFM 分數，而最高的組合 RFM 分數是 555。

資料考量

- 如果資料列代表交易（每一列代表一個單筆交易，而每一個客戶都可能有多筆交易），請使用交易來得到 RFM。[如需詳細資訊，請參閱第 3 頁 RFM 分數取自交易資料。](#)
- 如果資料列代表具有所有交易之摘要資訊的客戶（欄中含有總花費金額值、總交易次數和最近交易日期），請使用客戶資料來得到 RFM。[如需詳細資訊，請參閱第 4 頁 RFM 分數取自客戶資料。](#)

圖表 2-1
交易與客戶資料

行是交易			
ID	Gender	Date	Amount
1	Male	9/25/2005	21
2	Male	1/15/2006	297
4	Male	2/5/2006	249
4	Male	5/7/2005	172
6	Male	4/16/2005	164
6	Male	4/12/2005	286
7	Female	7/4/2005	400
9	Male		
9	Male		
9	Male		
10	Female		
10	Female		

具有交易小结的客户行				
ID	Gender	Most Recent	Total Amount	Number of Purchases
1	Male	9/25/2005	21	1
2	Male	1/15/2006	297	1
4	Male	2/5/2006	421	2
6	Male	4/16/2005	450	2

RFM 分數取自交易資料

資料考量

資料集必須包含內含下列資訊的變數：

- 可以識別每一個觀察值（客戶）的一個變數或變數的組合。
- 一個具有每一個交易日期的變數。
- 一個具有每一個交易金融值的變數。

圖表 2-2
RFM 交易資料

ID	Date	Amount
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/10/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

以交易資料建立 RFM 分數

- ▶ 從功能表選擇：
直效行銷 > 選擇技術
- ▶ 選取「協助識別我的最佳聯絡人 (RFM 分析)」，並按一下「繼續」。
- ▶ 選取「交易資料」，再按一下「繼續」。

圖表 2-3
交易資料，「變數」索引標籤



- ▶ 請選取包含交易日期的變數。
- ▶ 請選取包含每一個交易金額的變數。
- ▶ 請選取每一個客戶交易數量的摘要方法：總和（所有交易的加總）、平均數、中位數或最大值（最高交易金額）。
- ▶ 請選取可以唯一識別每一個客戶的變數或變數的組合。例如，觀察值可藉唯一的ID 碼或姓和名的組合來識別。

RFM 分數取自客戶資料

資料考量

資料集必須包含內含下列資訊的變數：

- 最近購買日期或是距最近購買日期的時間間隔。這會用來計算新近值分數。
- 購買次數的總和。這會用來計算頻率值分數。
- 所有購買的摘要金融值。這會用來計算金融值分數。一般而言，這是所有購買的總和（總數），但它也可為平均數（平均值）、最大值（最大金額）或是其他摘要量數。

圖表 2-4
RFM 客戶資料

ID	TotalAmount	MostRecent	NumberOfPurchases
1	485.00	09/04/2006	5
2	350.00	11/10/2005	4
3	233.00	06/04/2005	2
4	936.00	08/18/2006	7
5	359.00	07/07/2006	3
6	249.00	07/16/2006	3
7	1089.00	02/15/2006	7
8	423.00	08/21/2006	4
9	689.00	08/31/2006	7
10	325.00	10/13/2005	3

如果您要將 RFM 分數寫入一個新的資料集，則該作用中資料集必須也包含一個可以識別每一個觀察值（客戶）的變數或變數組合。

以客戶資料建立 RFM 分數

- ▶ 從功能表選擇：
直效行銷 > 選擇技術
- ▶ 選取「協助識別我的最佳聯絡人 (RFM 分析)」，並按一下「繼續」。
- ▶ 選取「客戶資料」，再按一下「繼續」。

圖表 2-5
客戶資料，「變數」索引標籤



- ▶ 請選取包含最近交易日期或是一個可代表距最近一次交易的時間間隔數字的變數。

- ▶ 請選取包含每一個客戶之總交易次數的變數。
- ▶ 請選取包含每一個客戶之摘要金額的變數。
- ▶ 如果您要將 RFM 分數寫入一個新的資料集，請選取可以唯一識別每一個客戶的變數或變數的組合。例如，觀察值可藉唯一的ID 碼或姓和名的組合來識別。

RFM Binning

將大數量的數值分組為小數量類別的過程有時也稱為 **Binning**。在 RFM 分析中，Bin 是已分等級的類別。您可使用「Binning」索引標籤來修改用以指定新近值、頻率值和金融值至那些 Bin 的方法。

圖表 2-6
「RFM Binning」索引標籤

Binning 方法

巢狀。 在巢狀 Binning 時會指定給新近值一個簡單等級。在每一個新近值等級之內，再指定一頻率值等級給客戶，而在每一個頻率值等級之內，指定一金融值等級給客戶。這樣將會給組合 RFM 分數提供更為均勻的分配，但它卻有造成頻率值和金融值等級分數難以解讀的缺點。例如，一個頻率值等級為 5 且具有新近值等級為 5 的客戶，其含意可能不同於一個頻率值等級為 5 而新近值等級為 4 的客戶，因為頻率值等級取決於新近值等級。

自變數。 為新近值、頻率值和金融值指定簡單等級。這三個等級是獨立指定的。因此對這三個 RFM 成份每一項的解讀是清楚的；一個頻率值分數為 5 的客戶與另一個頻率值分數為 5 的客戶其含意是相同的，不論他們的新近值分數是多少。對於較小的樣本數，此法會造成組合 RFM 分數分配不夠均勻的缺點。

Bin 的數目

用於建立 RFM 分數之每一個成份的類別 (Bin) 數量。所有可能組合 RFM 分數的總數是這三個值的乘積。例如，5 個新近值 Bin，4 個頻率值 Bin 和 3 個金融值 Bin 將會建立起總共 60 個可能組合 RFM 分數，範圍由 111 至 543。

- 每一個成份的預設值為 5，它會建立起 125 個可能組合 RFM 分數，範圍由 111 至 555。
- 每一個分數成份所容許的最大 Bin 數目是九。

等值結

「同分」就是兩個或多個相等的新近值、頻率值或金融值。最理想的是，您想讓每一個 Bin 中有大約相同數目的客戶，但是大量的同分數值會影響 Bin 的分配。有兩個選項來處理同分：

- **指定等值結至相同的分組 (bin)。** 這個方法一定會將等值指派給相同的分組，而不管這樣會對分組分配造成哪些影響。這樣可以提供一致的 Binning (分組) 方法：如果有兩名客戶擁有相同新近值，則他們一定會指派到相同的新近分數。但一個極端例子來說，您可能擁有 1,000 位客戶，其中有 500 位在相同的一天中完成了其最近一次的購買記錄。在 5-bin 排名中的客戶中，有 50% 會收到新近分數 5，而不是只有預期中 20% 收到該分數。

請注意，對頻率值和金融值而言，巢狀分組方法的「一致性」在某種程度上更為複雜，因為頻率分數會指派到新近分數分組，而金融分數是指派到頻率分數分組中。因此兩個擁有相同頻率值的客戶如果沒有相同的新近分數，則不管那些等值是如何處理，那兩位客戶就不會擁有相同的頻率分數。

- **隨機指派等值。** 這樣便可由事先指派極小的隨機變異數因子到同分，接著再進行排名，確保分組分配平均；而若是要將值指派到已排名的分組中，則不會有任何的等值。這個程序不會影響原來的值。它只能用於釐清等值。當其產生平均分組分配 (每個分組中有幾乎相同數目的客戶)，它便能針對似乎擁有相似或相同新近值、頻率值，及/或金融值的客戶產生完全不同的分數結果 — 尤其是當客戶總數相對的小，及/或等值數目相對的高。

表格 2-1
指派等值到相同的分組與隨機指派等值的比較

ID	最近一次購買 (新近值)	指派等值到相 同的分組	隨機指派等值
1	10/29/2006	5	5
2	10/28/2006	4	4
3	10/28/2006	4	4
4	10/28/2006	4	5
5	10/28/2006	4	3
6	9/21/2006	3	3
7	9/21/2006	3	2
8	8/13/2006	2	2
9	8/13/2006	2	1
10	6/20/2006	1	1

- 在這個範例中，將等值指派到相同分組會造成分組分配不平均：5 (10%)，4 (40%)，3 (20%)，2 (20%)，1 (10%)。
- 隨機指派等值會使得每個分組中擁有 20%，但是為了達到這個結果，日期值為 10/28/2006 的四筆觀察值被指派到三個不同的分組，而日期值為 8/13/2006 的二筆觀察值也會指派到不同的分組。

請注意，等值被指派到不同分組的行為完全是隨機進行（採用條件是最後結果是每個分組中有相同數目的觀察值）。若您使用相同方法計算第二組分數，具有等值之任何特定觀察值的排名可能會變更。例如，觀察值 4 和 5 的新近值排名 5 與 3，在第二次排名可能會互相調換。

以交易資料儲存 RFM 分數

取自交易資料的 RFM 一律會建立新的整合資料集，每一位客戶一行。請使用「儲存」索引標籤來指定您要儲存哪個分數和其他的變數以及您要將其存到何處。

圖表 2-7
交易資料，「儲存」索引標籤



變數

能唯一識別每一個客戶的 ID 變數會自動儲存到此新的資料集內。下列其他的變數可存在新的資料集內：

- **每一個客戶最近一次交易的日期。**
- **交易次數。** 每一個客戶之交易列的總數。
- **金額。** 依據您在「變數」索引標籤上選取的摘要方法而得之每一個客戶的摘要金額。
- **新近值分數。** 依據最近一次交易日期而指定給每一個客戶的分數。較高的分數表示較近的交易日期。

- **頻率值分數。** 依據總交易次數而指定給每一個客戶的分數。較高的分數表示較多的交易次數。
- **金融值分數。** 依據所選金融值摘要量數而指定給每一個客戶的分數。較高的分數表示較高的金融值摘要量數值。
- **RFM 分數。** 這三個個別的分數組合成一個單一值： $(\text{recency} \times 100) + (\text{frequency} \times 10) + \text{monetary}$ 。

依照預設值，所有可用的變數都包含在新的資料集內；所以請取消選取（取消勾選）這些您不想納入的變數。或者，您可以指定您自己的變數名稱。變數名稱必須符合標準變數命名規則。

位置

取自交易資料的 RFM 一律會建立新的整合資料集，每一位客戶一列。您可以在目前階段作業建立一個新的資料集或將 RFM 分數資料儲存在外部資料檔中。資料集名稱必須符合標準變數命名規則。（此限制並不適用於外部資料檔的名稱。）

以客戶資料儲存 RFM 分數

若是客戶資料，您可以將 RFM 分數變數新增至作用中資料集內或建立一個包含所選分數變數的新資料集。請使用「儲存」索引標籤來指定您要儲存哪些分數變數以及您要將其存到何處。

圖表 2-8
客戶資料，「儲存」索引標籤

客戶資料的 RFM 分析

變數 Binning 儲存 輸出

已儲存變數的名稱

自動產生唯一的名稱(A)

自訂名稱(C)

變數(V):

儲存	內容	已儲存變數的名稱
<input checked="" type="checkbox"/>	新近程度分數	Recency_score
<input checked="" type="checkbox"/>	次數分數	Frequency_score
<input checked="" type="checkbox"/>	貨幣分數	Monetary_score
<input type="checkbox"/>	RFM 分數	RFM_score

位置

寫入至作用中資料集(A)

建立新的資料集(C) 名稱(M):

撰寫新的資料檔(W) 瀏覽(B):...

確定 貼上之後(P) 重設(R) 取消 輔助說明

所儲存變數的名稱

- **自動產生唯一的名稱。** 將分數變數新增至作用中資料集時，這會確保新的變數名稱是唯一的。如果您要將多個不同組的 RFM 分數（依據不同的條件）新增至作用中資料集時，這個作法特別有用。
- **自訂名稱。** 這可讓您為分數變數指定您自己的變數名稱。變數名稱必須符合標準變數命名規則。

變數

請選取（勾選）您想儲存的分數變數：

- **新近值分數。** 依據在「變數」索引標籤所選取的交易日期或間隔值變數而指定給每一個客戶的分數。對較近的日期或較低的間隔值指定較高的分數。
- **頻率值分數。** 依據在「變數」索引標籤所選取的「交易次數」變數而指定給每一個客戶的分數。對較高的值指定較高的分數。
- **金融值分數。** 依據在「變數」索引標籤所選取的「金額」變數而指定給每一個客戶的分數。對較高的值指定較高的分數。
- **RFM 分數。** 這三個個別的分數組合成一個單一值： $(\text{新近值} \times 100) + (\text{頻率值} \times 10) + \text{金融值}$ 。

位置

若是客戶資料，有三個選項可讓您決定新的 RFM 分數要存到何處：

- **作用中資料集。** 將選取的分數變數新增至作用中資料集內。
- **新資料集。** 已選取的分數變數和能唯一識別每一個客戶（觀察值）的 ID 變數，將會被寫入至目前階段作業中的某個新資料集內。資料集名稱必須符合標準變數命名規則。此選項只有在您於「變數」索引標籤上選取一或多個「客戶識別碼」變數後才可使用。
- **檔案。** 已選取的分數和能唯一識別每一個客戶（觀察值）的 ID 變數將被儲存到外部資料檔內。此選項只有在您於「變數」索引標籤上選取一或多個「客戶識別碼」變數後才可使用。

RFM 輸出

圖表 2-9
「RFM 輸出」索引標籤



Bin 的資料

bin 資料的圖表與表格是以計算的新近值、頻率值與金融值分數為基礎。

依新近程度和次數所建立之平均貨幣值的熱圖。平均金融值分配的熱圖會顯示由新近值與頻率值分數所定義類別的平均金融值。區域顏色愈深，表示平均金融值愈高。

bin 個數的圖表。bin 個數的圖表顯示所選 bin 方法的 bin 分配。每個長條代表著將指定給每個組合 RFM 分數的觀察值數目。

- 雖然您通常想要相當均勻的分配，即所有（或大部分）的列高度都相等，在使用將同分數值指定至相同之 Bin 的預設 Binning 方法時，可預期應該會有某些程度的差異。
- Bin 分配和/或空白 Bin 中的極端變動可能表示您應嘗試另一種 Binning 方法（較少的 Bin 和/或隨機指定同分），或重新考慮 RFM 分析的適用性。

bin 個數的表格。在 bin 個數圖表中的資訊與每個儲存格中的 bin 個數相同，不同處在於前者以表格形式表示。

沒有 bin 的資料

沒有 bin 的資料其圖表與表格是以用於建立新近值、頻率值與金融值分數的原始變數為基礎。

直方圖。直方圖會顯示用於計算新近值、頻率值與金融值分數的三個變數值其相對分配。這些直方圖會指出有些偏斜的分配，而非非常態或對稱分配，這種狀況並非異常。

每個直方圖的水平軸排序方式一律為左邊的數值低，右邊的數值高。但有了新近值，對圖表的解讀會取決於新近值量數的類型：日期或時間間隔。若為日期，左側的長條代表較早的數值（之前日期的值會小於最近日期的值）。若為時間間隔，左側的長條代表最近的值（時間間隔愈小，交易愈接近）。

成對變數的散佈圖。這些散佈圖顯示用於計算新近值、頻率值與金融值分數的三個數值之間的關係。

因為次數通常代表相當小的離散值範圍，因此次數尺度上常出現明顯的點線性分組。例如，如果總交易次數不超過 15，則只會有 15 個可能的次數值（除非您也將分數交易算在內），但卻會有數百個可能的最近值和上千個金融值。

對新近值軸的解讀取決於新近值量數的類型：日期或時間間隔。若為日期，愈接近原點的點代表愈早的日期。若為時間間隔，愈接近原點的點代表愈近的數值。

集群分析

「集群分析」這種探索工具旨在顯示資料中的自然分組（或集群）。例如，可以根據不同人口資料和購買特性，識別出不同組別的客户。

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

集群分析的資料考量












資料。 此程序可用在連續變數及類別欄位上。每個記錄（列）各代表要集群的物件，欄位（變數）則代表集群所依據的屬性。

記錄順序。 請注意，結果可能會依據記錄順序而定。為最小化順序效應，您可能會考慮以隨機方式排序記錄。您可能想要多次執行分析，並以不同的隨機順序排序記錄，以確認給定解答的穩定性。

測量水準。 正確的測量水準指派是很重要的，因為其會影響結果的計算。

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

每個欄位旁的圖示代表目前的測量水準。

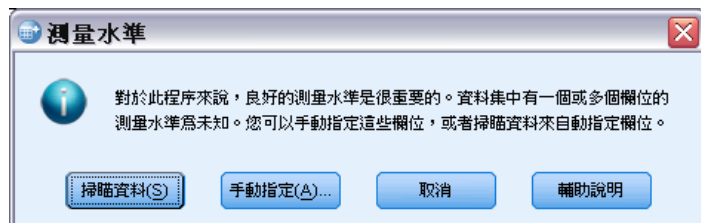
	數字的	字串	日期	時間
尺度（連續）		無		
次序				
名義				

您可以在「資料編輯程式」的「變數檢視」中變更測量水準，或可使用「定義變數性質」對話方塊來建議每個欄位適當的測量水準。

具有未知測量水準的欄位

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

圖表 3-1
測量水準警示



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

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

若要取得集群分析

從功能表選擇：

直效行銷 > 選擇技術

- ▶ 選取「將我的聯絡人區隔為數個集群」。

圖表 3-2
「集群分析欄位」索引標籤



- ▶ 選取要用來建立區隔的類別（名義、次序）和連續（尺度）欄位。
- ▶ 按一下「執行」執行程序。

設定

圖表 3-3
「集群分析設定」索引標籤



「設定」索引標籤可讓您顯示或隱藏描述區隔的圖表或表格、在資料集中儲存新欄位，以識別資料集中每個記錄的區隔（集群），以及指定要在集群解中包含多少區隔。

顯示圖表與表格。 顯示描述區隔的表格與圖表。

區隔會員資格。 儲存新欄位（變數），以識別每個記錄所屬的區隔。

- 欄位名稱必須符合 IBM® SPSS® Statistics 命名規則。
- 片段成員欄位名稱不得與資料集中已存在的欄位名稱重複。如果您要在同一個資料集上執行此程序多次，則必須每次都指定不同的名稱。
- **區隔個數。** 控制如何決定區隔數目。
- **自動決定。** 此程序將會自動決定最佳的區隔數（不超過指定的最大值）。

指定固定數。 此程序將會產生指定的區隔數。

潛在客戶分析概要

此方法使用上一個活動或測試活動的結果來建立描述性分析概要。您可以在未來的活動中會使用這些分析概要來鎖定特定的群組。「回應」欄位會識別誰有回應上一個活動或測試活動。「分析概要」清單包含您可用來建立分析概要的特性。

範例。公司的直效行銷部門根據測試郵件的結果，以人口資料來產生哪一些客戶最可能回應優惠的分析概要。

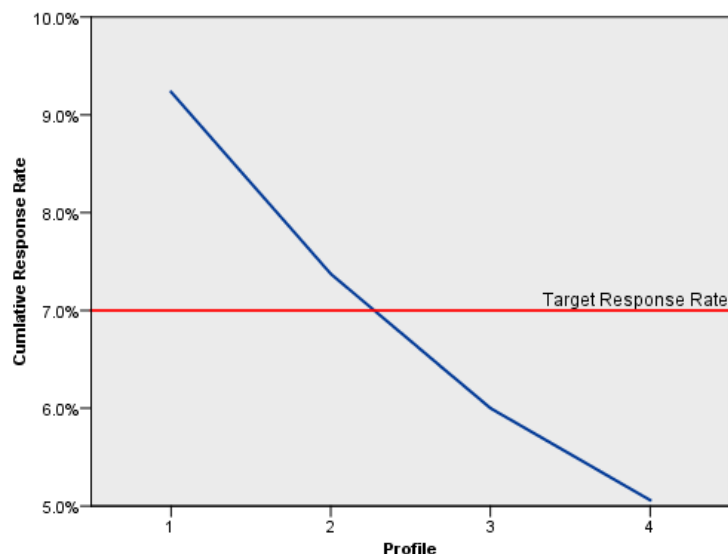
輸出

輸出結果中會包含一張表格，表格中會說明每個分析概要群組，並顯示回應率（正回應的百分比）與累積回應率，此外也會提供累積回應率的圖表。如果您包含了目標最小回應率，則系統會將表格予以彩色編碼，以顯示有哪些分析概要達到最小累積回應率，圖表也會在指定的最小回應率值包含參考線。

圖表 4-1
回應率表格和圖表

數目	回應率			
	說明	群組大小	回應率	累積回應率
1	Region = "West","South","East" Gender = "Female" Married = "No"	379	9.2%	9.2%
2	Region = "West","South","East" Gender = "Female" Married = "Yes"	299	5.0%	7.4%
3	Region = "West","South","East" Gender = "Male"	722	4.7%	6.0%
4	Region = "North"	517	2.5%	5.1%

綠色：達成目標回應率。
紅色：未達成目標回應率。



分析概要資料考量

回應欄位。 回應欄位必須是名義或次序。這個變數可以為字串或數值。如果這個欄位包含的值是表示購買的數目或數量，您必須建立一個新的欄位，其中以單一值代表所有的正回應。如需詳細資訊，請參閱第 21 頁建立類別回應欄位。

正回應值。 正回應值表示正向回應的客戶（例如會購買）。所有其他的非遺漏回應值皆假設為負回應。如果該回應欄位有任何已定義的數值標記，下拉式清單中會顯示這些標記。












建立分析概要方法。 這些欄位可以是名義、次序或連續（尺度）。這些欄位可以為字串或數值。

測量水準。 正確的測量水準指派是很重要的，因為其會影響結果的計算。

- **名義。** 當變數數值代表實質上並未等級化的類別時（例如，有員工工作的公司部門），則此變數可視為名義。名義變數的範例包括地區、郵遞區號以及宗教團體。

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

每個欄位旁的圖示代表目前的測量水準。

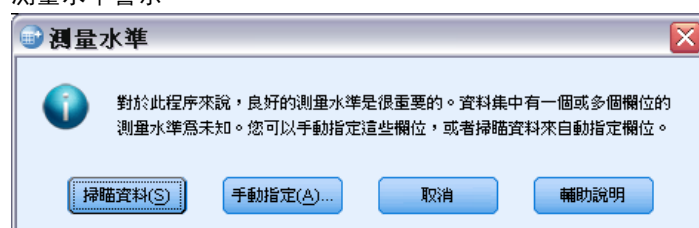
	數字的	字串	日期	時間
尺度（連續）		無		
次序				
名義				

您可以在「資料編輯程式」的「變數檢視」中變更測量水準，或可使用「定義變數性質」對話方塊來建議每個欄位適當的測量水準。

具有未知測量水準的欄位

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

圖表 4-2
測量水準警示



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

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

取得潛在客戶分析概要

從功能表選擇：

直效行銷 > 選擇技術

- ▶ 選取「產生回應優惠的聯絡人分析概要」。

圖表 4-3
「潛在客戶分析概要欄位」索引標籤



- ▶ 選取識別哪一些聯絡人對優惠有回應的欄位。此欄位必須是名義或次序。
- ▶ 輸入指示正回應的值。如果有任何值已定義數值標記，您可以從下拉式清單中選取該數值標記，對應的值也隨即顯示。
- ▶ 選取您想要用來建立分析概要的欄位。
- ▶ 按一下「執行」執行程序。

設定

圖表 4-4
「潛在客戶分析概要設定」索引標籤

潛在客戶分析概要

欄位 設定 (N)

最小側寫群組大小(M): 100

在結果中包含最小回應率門檻資訊(I)

指定目標回應率 (%) (T): 7

執行(R) 貼上之後(P) 重設(R) 取消 輔助說明

「設定」索引標籤可讓您控制最小的分析概要群組大小，並在輸出結果中包含最小回應率門檻。

最小分析概要群組大小。每個分析概要皆代表資料集中一組聯絡人共用的特性（例如，居住在西部 40 歲以下的女性）。依照預設值，最小的分析概要群組大小為 100。群組大小愈小，可顯示愈多的群組，但群組大小愈大，所提供的結果愈可靠。數值必須是正整數。

在結果中包含最小回應率門檻資訊。結果中會包含一張表格，表格中會顯示回應率（正回應的百分比）與累積回應率，此外也會提供累積回應率的圖表。如果您輸入了目標最小回應率，則系統會將表格予以彩色編碼，以顯示有哪些分析概要達到最小累積回應率，圖表也會在指定的最小回應率值包含參考線。數值必須大於 0 且小於 100。

建立類別回應欄位

回應欄位應為類別，其中以值一個代表所有的正回應。任何其他非遺漏值皆假設為負回應。如果回應欄位代表連續（尺度）值，例如購買次數或購買金額，您必須建立一個新欄位將單一正回應值指定給所有非零的回應值。

- 如果將負回應記錄為 0（不是空白，空白會視為遺漏），則可使用下列公式計算：

$$\text{NewName} = \text{OldName} > 0$$

其中 `NewName` 是新欄位的名稱，`OldName` 是原始欄位的名稱。這個邏輯運算式會將 1 的值指定給所有大於 0 的非遺漏值，將 0 指定給所有小於或等於 0 的非遺漏值。

- 如果沒有任何一個值記錄為負回應，則這些值會視為遺漏，公式會有一點複雜：

`NewName=NOT(MISSING(OldName))`

在此邏輯運算式中，會將所有非遺漏的回應值指定為 1 的值，所有遺漏的回應值會指定為 0 的值。

- 如果您無法區分負 (0) 回應值與遺漏值，則無法計算出準確的回應值。如果真正的遺漏值相當的少，則可能對計算的回應率沒有顯著的影響。但是如果有許多遺漏值 — 例如當只記錄了總資料集其一小部分檢定樣本的回應資訊時 — 則計算的回應率將沒有意義，因為這些回應率的顯著性低於真正的回應率。

建立類別回應欄位

- ▶ 從功能表選擇：
轉換(T) > 計算變數
- ▶ 對於「目標變數」，請輸入新的欄位（變數）名稱。
- ▶ 如果將負回應記錄為 0，則將 Numeric Expression 輸入為 `OldName>0`，其中的 `OldName` 是原始欄位名稱。
- ▶ 如果將負回應記錄為遺漏（空白），則將 Numeric Expression 輸入為 `NOT(MISSING(OldName))`，其中 `OldName` 是原始欄位名稱。

郵遞區號回應率

此方法使用上一個活動的結果來計算郵遞區號回應率。在未來的活動中會使用這些回應率來鎖定特定的郵遞區號。「回應」欄位會識別誰有回應上一個活動。「郵遞區號」欄位會識別包含郵遞區號的欄位。

範例。公司的直效行銷部門可以根據先前的郵件結果，來產生郵遞區號回應率。接著，該部門可以依據郵件中所包含的各種條件，如最小可接受回應率和/或最大聯絡人個數，來鎖定特定的郵遞區號。

輸出

此程序的輸出結果中會有一個新的資料集，其中包含了郵遞區號的回應率，以及按照十分位數等級（前 10%、前 20% 等）來摘要結果的表格與圖表。系統會根據使用者指定的最小累積回應率或最大聯絡人個數，將表格加以彩色編碼。

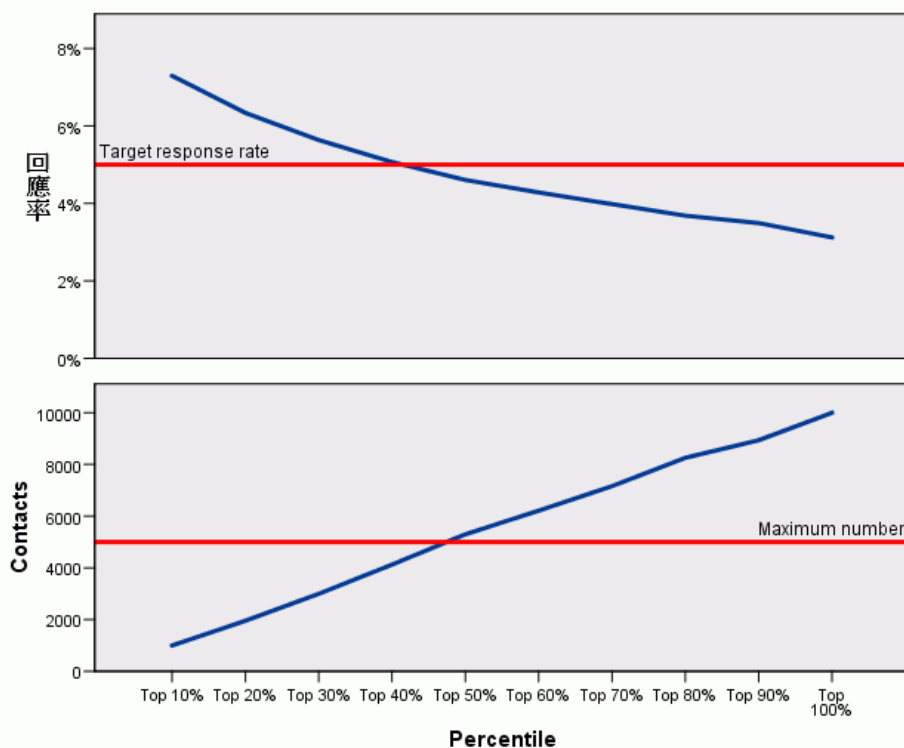
圖表 5-1
含郵遞區號回應率的資料集

	PostalCode	ResponseRate	Responses	Contacts	Index	Rank	
1	932	10%	4	40	3.6	Top 10%	
2	098	9%	6	68	5.5	Top 10%	
3	740	8%	9	116	8.3	Top 10%	
4	100	8%	7	91	6.5	Top 10%	
5	110	8%	5	65	4.6	Top 10%	
6	954	8%	4	53	3.7	Top 10%	
7	108	7%	6	82	5.6	Top 10%	
8	107	7%	5	71	4.6	Top 10%	
9	090	7%	4	58	3.7	Top 10%	
10	966	7%	4	58	3.7	Top 10%	
11	760	7%	8	119	7.5	Top 10%	
12	113	6%	5	80	4.7	Top 10%	

圖表 5-2
摘要表格和圖表

Percentile	回應率	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

綠色：達成目標回應率。
紅色：未達成目標回應率。



新資料集包含下列欄位：

- **郵遞區號**。如果郵遞區號群組只採用完整數值的一部分，則這會是該部分郵遞區號的值。Excel 檔案中這一行的頁首列標記為原始資料集中的郵遞區號欄位名稱。
- **回應率**。各郵遞區號的正回應百分比。
- **反應值**。各郵遞區號的正回應數目。
- **聯絡人**。各郵遞區號中包含回應欄位其非遺漏值的聯絡人總數目。

- **索引。**採用公式 $N \times P \times (1-P)$ 「加權後」的回應，其中 N 是聯絡人個數，P 是以比例表示的回應率。
- **等級。**累積郵遞區號回應率的十分位等級（前 10%、後 20% 等），以遞減順序排序。

郵遞區號回應率資料考量

回應欄位。回應欄位可以為字串或數值。如果這個欄位包含的值是表示購買的數目或貨幣值，您必須建立一個新的欄位，其中以單一值代表所有的正回應。[如需詳細資訊，請參閱第 28 頁建立類別回應欄位。](#)

正回應值。正回應值表示正向回應的客戶（例如會購買）。所有其他的非遺漏回應值皆假設為負回應。如果該回應欄位有任何已定義的數值標記，下拉式清單中會顯示這些標記。

郵遞區號欄位。郵遞區號欄位可以為字串或數值。

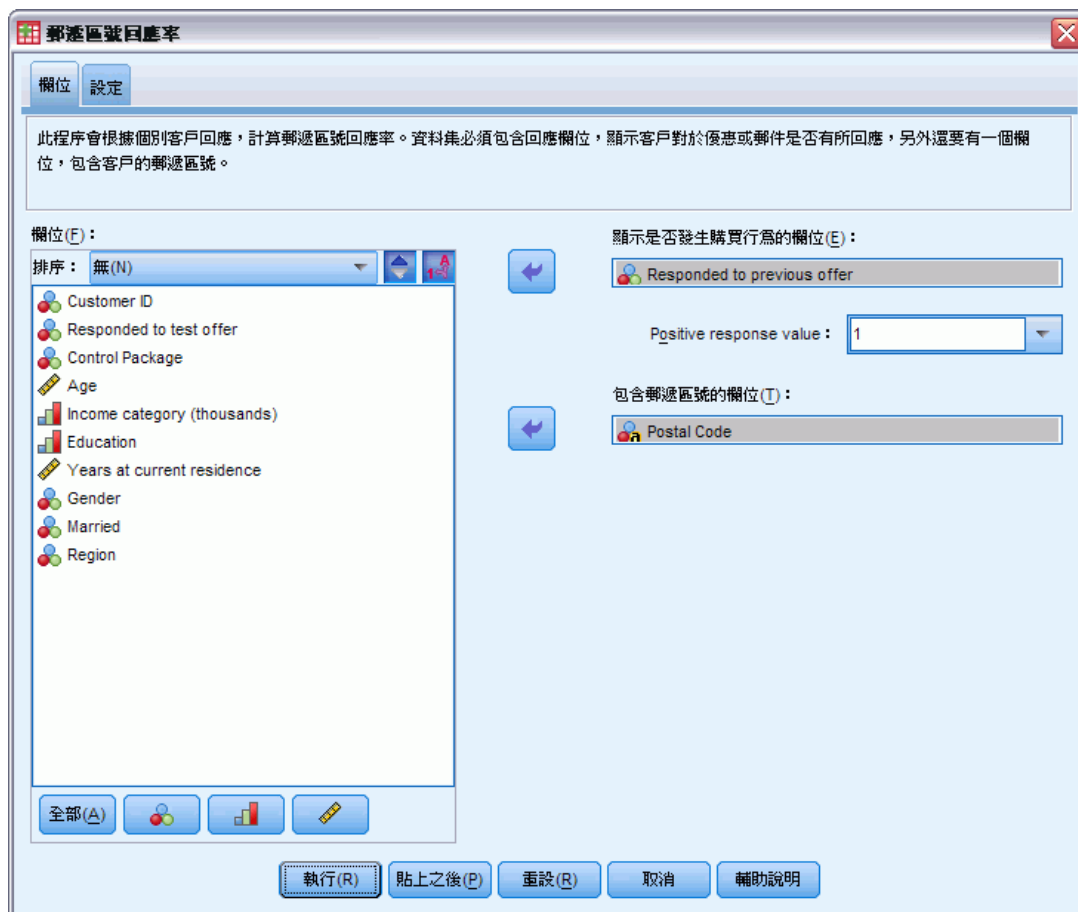
取得郵遞區號回應率

從功能表選擇：

直效行銷 > 選擇技術

- ▶ 選取「識別高分回應郵遞區號」。

圖表 5-3
「郵遞區號回應率欄位」索引標籤



- ▶ 選取識別哪一些聯絡人對優惠有回應的欄位。
- ▶ 輸入指示正回應的值。如果有任何值已定義數值標記，您可以從下拉式清單中選取該數值標記，對應的值也隨即顯示。
- ▶ 選取包含郵遞區號的欄位。
- ▶ 按一下「執行」執行程序。

您也可以：

- 根據郵遞區號的前 n 個字元或位數來產生回應率，而非根據完整的數值
- 自動將結果儲存至 Excel 檔
- 控制輸出顯示選項

設定

圖表 5-4
「郵遞區號回應率設定」索引標籤

郵遞區號分組依據

這會決定要如何群組記錄以計算回應率。依照預設值，系統會使用整個郵遞區號，並將所有具有同一個郵遞區號的記錄群組在一起，以計算出群組回應率。或者，您可以只根據完整郵遞區號的一部分（包含前 n 個位數或字元）來群組記錄。例如，您可以只根據 10 個字元郵遞區號的前 5 個字元，或 5 位數郵遞區號的前 3 位數來群組記錄。輸出資料集會包含各郵遞區號群組的一筆記錄。如果您輸入的是數值，則數值必須為正整數。

郵遞區號數字格式

如果郵遞區號欄位是數值，且您要根據前 n 位數來群組郵遞區號，而非根據整個數值，則您必須指定原值的位數數目。位數數目是郵遞區號中最大可能數目的位數。例如，如果郵遞區號欄位中混合了 5 位數與 9 位數的郵遞區號，則您應指定 9 作為位數數目。

注意：視顯示格式的不同，有些 5 位數的郵遞區號看起來只有 4 位數，但實際上是隱含了前置的零。

輸出

除了顯示包含郵遞區號回應率的新資料集外，您還可以顯示按照十分位等級（前 10%、前 20% 等）來摘要結果的表格與圖表。表格會顯示各十分位中的回應率、累積回應率、記錄數目與累積記錄數目。圖表會顯示各十分位中的累積回應率與累積記錄數目。

最小可接受回應率。如果您輸入的是目標最小回應率或損益平衡公式，則系統會將表格予以彩色編碼，以顯示有哪些十分位達到最小累積回應率，圖表也會在指定的最小回應率值包含參考線。

- **目標回應率。** 回應率以百分比表示（各郵遞區號群組的正回應百分比）。數值必須大於 0 且小於 100。
- **從公式計算收支平衡率。** 根據公式計算最小累積回應率： $(\text{郵寄包裹的成本}/\text{每回應淨收入}) \times 100$ 。兩個數值皆必須為正數。結果應為大於 0 且小於 100 的值。例如，如果郵寄包裹的成本為 0.75 美元，每回應淨收入為 56 美元，則最小回應率為： $(0.75/56) \times 100 = 1.34\%$ 。

最大聯絡人個數。如果您指定最大聯絡人個數，則系統會將表格予以彩色編碼，以顯示哪些十分位沒有超過累積最大聯絡人個數（記錄），圖表也會在該值處包含參考線。

- **聯絡人百分比。** 最大值以百分比表示。例如，您可能想要知道有哪些含最高回應率的十分位包含的聯絡人不到所有聯絡人的 50%。數值必須大於 0 且小於 100。
- **聯絡人數目。** 最大值以聯絡人數目表示。例如，如果您不想要寄出超過 10,000 個包裹，您可以將該值設為 10000。該值必須為正整數（無群組符號）。

如果您指定了最小可接受回應率與最大聯絡人個數，則系統會根據第一個符合的條件將表格予以彩色編碼。

匯出至 Excel

這個程序會自動建立一個新資料集，其中包含郵遞區號回應率。資料集中的每筆記錄（列）代表一個郵遞區號。您可以將相同的資訊自動儲存在 Excel 檔中。此檔案會以 Excel 97-2003 格式儲存。

建立類別回應欄位

回應欄位應為類別，其中以值一個代表所有的正回應。任何其他非遺漏值皆假設為負回應。如果回應欄位代表連續（尺度）值，例如購買次數或購買金額，您必須建立一個新欄位將單一正回應值指定給所有非零的回應值。

- 如果將負回應記錄為 0（不是空白，空白會視為遺漏），則可使用下列公式計算：

$\text{NewName} = \text{OldName} > 0$

其中 NewName 是新欄位的名稱，OldName 是原始欄位的名稱。這個邏輯運算式會將 1 的值指定給所有大於 0 的非遺漏值，將 0 指定給所有小於或等於 0 的非遺漏值。

- 如果沒有任何一個值記錄為負回應，則這些值會視為遺漏，公式會有一點複雜：

`NewName=NOT (MISSING (OldName))`

在此邏輯運算式中，會將所有非遺漏的回應值指定為 1 的值，所有遺漏的回應值會指定為 0 的值。

- 如果您無法區分負 (0) 回應值與遺漏值，則無法計算出準確的回應值。如果真正的遺漏值相當的少，則可能對計算的回應率沒有顯著的影響。但是如果有許多遺漏值 — 例如當只記錄了總資料集其一小部分檢定樣本的回應資訊時 — 則計算的回應率將沒有意義，因為這些回應率的顯著性低於真正的回應率。

建立類別回應欄位

- ▶ 從功能表選擇：
轉換(T) > 計算變數
- ▶ 對於「目標變數」，請輸入新的欄位（變數）名稱。
- ▶ 如果將負回應記錄為 0，則將 Numeric Expression 輸入為 `OldName>0`，其中的 OldName 是原始欄位名稱。
- ▶ 如果將負回應記錄為遺漏（空白），則將 Numeric Expression 輸入為 `NOT (MISSING (OldName))`，其中 OldName 是原始欄位名稱。

傾向機率

「傾向機率」使用測試郵件或上一個活動的結果來產生分數。分數會識別哪些聯絡人最有可能回應。「回應」欄位會識別誰有回覆測試郵件或上一個活動。「傾向」欄位中的特性可用來預測具有類似特性的聯絡人其回應機率。

這項技術採用二元 logistic 迴歸來建立預測模式。建立和套用預測模式的程序具有兩個基本步驟：

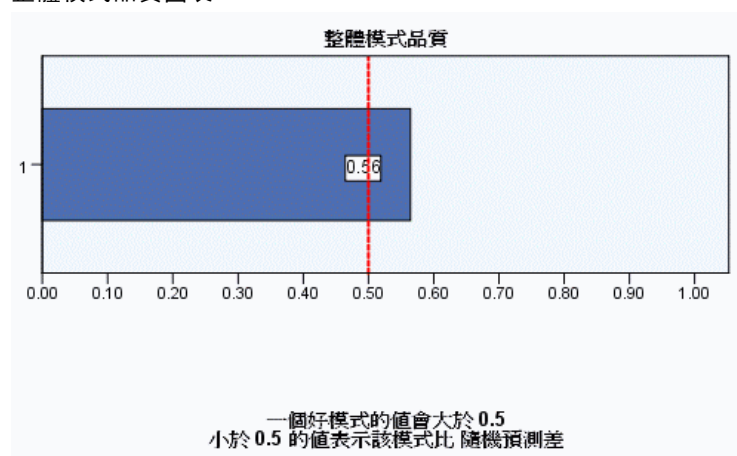
- ▶ 建立模式並儲存模式檔案。您使用已知相關結果（通常稱為**目標**）的資料及來建立模式。例如，如果您要建立一個模式以預測哪些人可能回應直效郵件活動，那麼您必須從已經包含相關資訊（哪些人有回應，哪些人沒有回應）的資料集開始。例如，這可能是對一小群客戶所進行的測試郵件結果，或是過去類似活動的回應資訊。
- ▶ 將該模式套用至不同的資料集（未知其相關結果）以取得預測結果。

範例。 公司的直效行銷部門使用測試郵件的結果，將傾向分數指派給聯絡人資料庫中其他的人，使用各種人口統計特色來識別出最有可能回應與購買的聯絡人。

輸出

此程序會自動在資料集中建立新的欄位，其中會包含檢定資料及 XML 模式檔案（可用於為其他資料集評分）的傾向分數。選用的診斷輸出包括整體模式品質圖表與分類表，這些表可比較預測回應與實際回應。

圖表 6-1
整體模式品質圖表



傾向機率資料考量

回應欄位。 回應欄位可以為字串或數值。如果這個欄位包含的值是表示購買的數目或貨幣值，您必須建立一個新的欄位，其中以單一值代表所有的正回應。[如需詳細資訊，請參閱第 35 頁建立類別回應欄位。](#)












正回應值。 正回應值表示正向回應的客戶（例如會購買）。所有其他的非遺漏回應值皆假設為負回應。如果該回應欄位有任何已定義的數值標記，下拉式清單中會顯示這些標記。

預測傾向方法。 這個欄位用於預測傾向，可以為字串或數值，此欄位為名義、次序或連續（尺度）— 但指派適當的測量水準至所有的預測量欄位，這點十分重要。

測量水準。 正確的測量水準指派是很重要的，因為其會影響結果的計算。

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

每個欄位旁的圖示代表目前的測量水準。

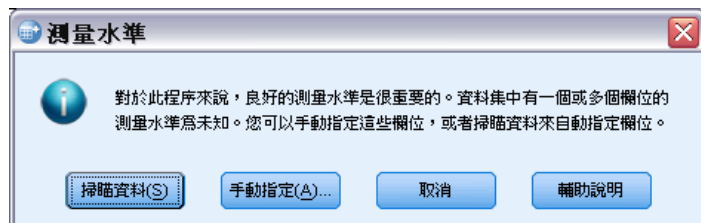
	數字的	字串	日期	時間
尺度（連續）		無		
次序				
名義				

您可以在「資料編輯程式」的「變數檢視」中變更測量水準，或可使用「定義變數性質」對話方塊來建議每個欄位適當的測量水準。

具有未知測量水準的欄位

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

圖表 6-2
測量水準警示



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

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

取得傾向機率分數

從功能表選擇：

直效行銷 > 選擇技術

- ▶ 選取「選取最有可能購買的聯絡人」。

圖表 6-3
「傾向機率欄位」索引標籤



- ▶ 選取識別哪一些聯絡人對優惠有回應的欄位。
- ▶ 輸入指示正回應的值。如果有任何值已定義數值標記，您可以從下拉式清單中選取該數值標記，對應的值也隨即顯示。
- ▶ 選取您想要用來預測傾向的欄位。
如要儲存模式 XML 檔案來為其他資料檔案評分：
- ▶ 選取（勾選）「匯出模式資訊至 XML 檔案」。
- ▶ 輸入目錄路徑和檔案名稱，或是按一下「瀏覽」，以瀏覽至要儲存模式 XML 檔案的所在位置。

- ▶ 按一下「執行」執行程序。
如要使用模式檔案來為其他資料集評分：
- ▶ 開啟您要評分的資料集。
- ▶ 使用「評分精靈」將模式套用至資料集。從功能表選擇：
公用程式(U) > 評分精靈。

設定

圖表 6-4
「傾向機率」的「設定」索引標籤

購買傾向

欄位 設定(N)

模式驗證

您可以驗證用來產生分數的模式。若要驗證模式，您必須將資料分成數個分割。訓練分割用於訓練（也就是建置）模式。測試分割用於驗證模式。如果您要驗證模式，此方法會自動將記錄指派給分割。

驗證模式(V)

訓練樣本分割大小(%) (T):

設定要複製結果的種子(S)

種子(E):

診斷輸出

整體模式品質(Q)

分類表(C)

最小機率(M):

重新編碼的回應欄位名稱與標記

此方法會自動將回應欄位重新編碼為新欄位，其中 1 代表正回應，0 代表負回應。

新欄位名稱(N):

新欄位標記(V):

儲存分數

此方法使用測試郵件或上一個活動的結果來產生分數。系統會自動儲存分數，讓您可以使用這些分數。此索引標籤上的其他控制項可額外控制所儲存的內容。

分數欄位名稱(S):

執行(R) 貼上之後(P) 重設(R) 取消 輔助說明

模式驗證

模式驗證會建立訓練群組與測試群組，以供診斷之用。如果您在「診斷輸出」區段中選取了分類表，則表格會分成訓練（已選取）與測試（未選取）區段，讓您能夠進行比較。請不要選取模式驗證，除非您也選取分類表。分數是根據從訓練樣本產生的模式計算而得，此樣本包含的記錄永遠少於可用記錄的總數。例如，預設的訓練樣本大小為 50%，則僅以一半的可用記錄來建置模式，此模式的可靠性比不上以所有可用記錄建置的模式。

- **訓練樣本分割大小。**指定指派給訓練樣本的記錄百分比。回應欄位中其餘含非遺漏值的記錄會指派至測試樣本。數值必須大於 0 且小於 100。
- **設定要複製結果的種子。**由於記錄是隨機指派給訓練樣本與測試樣本，所以每次您執行程序所得到的結果會不一樣，除非您一直指定相同的起始亂數種子值。

診斷輸出

整體模式品質。顯示整體模式品質的長條圖，以 0 到 1 之間的值表示。一個好的模式其值應大於 0.5。

分類表。顯示一個表格，其中會將預測的正回應與負回應和實際的正回應與負回應加以比較。整體的準確率會指出模式良好運作的程度，但您可能對於正確的預測正回應百分比比較有興趣。

- **最小機率。**將分數值大於指定值的記錄指派給分類表中預測的正回應類別。此程序所產生的分數代表聯絡人正回應的機率（例如會購買）。以一般的規則而言，您應指定接近最小目標回應率的值（以比例表示）。例如，如果您有興趣的回應率至少為 5%，請指定 0.05。數值必須大於 0 且小於 1。

重新編碼的回應欄位名稱與標記

此程序會將回應欄位重新編碼為新欄位，其中 1 代表正回應，0 代表負回應，分析是在重新編碼的欄位上執行。您可以用自己的名稱與標記來覆寫預設的名稱與標記。名稱必須符合 IBM® SPSS® Statistics 命名規則。

儲存分數

包含傾向分數的新欄位會自動儲存至原始資料集。分數代表正回應的機率（以比例表示）。

- 欄位名稱必須符合 SPSS Statistics 命名規則。
- 欄位名稱不得與資料集中已存在的欄位名稱重複。如果您要在同一個資料集上執行此程序多次，則必須每次都指定不同的名稱。

建立類別回應欄位

回應欄位應為類別，其中以值一個代表所有的正回應。任何其他非遺漏值皆假設為負回應。如果回應欄位代表連續（尺度）值，例如購買次數或購買金額，您必須建立一個新欄位將單一正回應值指定給所有非零的回應值。

- 如果將負回應記錄為 0（不是空白，空白會視為遺漏），則可使用下列公式計算：
$$\text{NewName} = \text{OldName} > 0$$

其中 `NewName` 是新欄位的名稱，`OldName` 是原始欄位的名稱。這個邏輯運算式會將 1 的值指定給所有大於 0 的非遺漏值，將 0 指定給所有小於或等於 0 的非遺漏值。

- 如果沒有任何一個值記錄為負回應，則這些值會視為遺漏，公式會有一點複雜：

`NewName=NOT(MISSING(OldName))`

在此邏輯運算式中，會將所有非遺漏的回應值指定為 1 的值，所有遺漏的回應值會指定為 0 的值。

- 如果您無法區分負 (0) 回應值與遺漏值，則無法計算出準確的回應值。如果真正的遺漏值相當的少，則可能對計算的回應率沒有顯著的影響。但是如果有許多遺漏值 — 例如當只記錄了總資料集其一小部分檢定樣本的回應資訊時 — 則計算的回應率將沒有意義，因為這些回應率的顯著性低於真正的回應率。

建立類別回應欄位

- ▶ 從功能表選擇：
轉換(T) > 計算變數
- ▶ 對於「目標變數」，請輸入新的欄位（變數）名稱。
- ▶ 如果將負回應記錄為 0，則將 Numeric Expression 輸入為 `OldName>0`，其中的 `OldName` 是原始欄位名稱。
- ▶ 如果將負回應記錄為遺漏（空白），則將 Numeric Expression 輸入為 `NOT(MISSING(OldName))`，其中 `OldName` 是原始欄位名稱。

控制套件檢定

此方法會比較行銷活動，瞭解不同包裝或優惠的效果是否有顯著性差異。活動效果是根據回應測量而得。「活動」欄位會識別不同的活動，例如「優惠 A」和「優惠 B」。「回應」欄位會識別聯絡人是否已回應活動。若將回應記錄為交易金額，例如「99.99」，則選取「交易金額」。當回應僅指示聯絡人是否正面回應（例如「是」或「否」）時，請選取「回覆」。

範例。 公司的直效行銷決策部門想了解新套件的設計是否將產生較現有套件更正面的回應。因此他們送出一封測試郵件，以判斷新套件是否能產生明顯較高的正回應率。測試郵件包含一個接受現有套件的控制群組，以及一個接受新套件設計的測試群組。然後會將這兩個群組的結果做比較，以了解是否有明顯的差異。

輸出

輸出包含一個表格，其顯示「活動欄位」定義之每個群組的正和負回應的個數與百分比，以及另一個識別哪些群組與彼此之間有明顯差異的表格。

圖表 7-1
「控制套件檢定」輸出

	Control Package			
	Control		Test	
	個數	行百分比	個數	行百分比
效果 (1=Yes 0=No) 0	875	96.2%	945	93.8%
1	35	3.8%	62	6.2%

Control 和 Test 之間存在統計顯著性差異。

控制套件檢定的資料考量與假設

活動欄位。 「活動欄位」應是類別的（名義或次序）。

效果回應欄位。 若您為「效果欄位」選取「交易金額」（此欄位需為數值），則測量水準應為連續（尺度）。

如果您無法區分負（對交易金額而言為 0 值）回應值與遺漏值，則無法計算出準確的回應率。如果真正的遺漏值相當的少，則可能對計算的回應率沒有顯著的影響。但是如果有許多遺漏值 — 例如當只記錄了總資料集其一小部分檢定樣本的回應資訊時 — 則計算的回應率將沒有意義，因為這些回應率的顯著性低於真正的回應率。

假設。 此程序假設聯絡人已隨機指派給每個活動群組。換言之，沒有特定的人口統計、交易記錄或其他特性會影響群組指派，且所有聯絡人指派給任一群組的機率均相等。

若要取得控制套件檢定

從功能表選擇：
直效行銷 > 選擇技術

- ▶ 選取「比較活動的效果」。

圖表 7-2
「控制套件檢定」對話方塊

- ▶ 選取可識別每個聯絡人分屬哪一活動群組（例如優惠 A 與優惠 B 等）的欄位。此欄位必須是名義或次序的。
- ▶ 選取指定回應效果的欄位。

若回應欄位是交易金額，則欄位需為數值。

若回應欄位僅指示聯絡人是否提供正回應（例如「是」或「否」）時，請選取「回覆」，並輸入代表正回應的值。如果有任何值已定義數值標記，您可以從下拉式清單中選取該數值標記，對應的值也隨即顯示。

新的欄位便會自動建立，其中 1 代表正回應，0 代表負回應，並且會針對新欄位執行分析。您可以用自己的名稱與標記來覆寫預設的名稱與標記。名稱必須符合 IBM® SPSS® Statistics 命名規則。

- ▶ 按一下「執行」執行程序。

部 11: 範 例

交易資料的 RFM 分析

在交易資料檔中，每一列代表一個不同的交易而非不同的客戶，而每一個客戶都可能有多筆交易列。此範例使用資料檔 `rfm_transactions.sav`。如需詳細資訊，請參閱第 91 頁附錄 A 中的範例檔案。

交易資料

資料集必須包含內含下列資訊的變數：

- 可以識別每一個觀察值（客戶）的一個變數或變數的組合。
- 一個具有每一個交易日期的變數。
- 一個具有每一個交易金融值的變數。

圖表 8-1
RFM 交易資料

ID	Date	Amount
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/10/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

執行分析

- ▶ 若要計算 RFM 分數，請在功能表中選擇：
直效行銷 > 選擇技術
- ▶ 選取「協助識別我的最佳聯絡人 (RFM 分析)」，並按一下「繼續」。
- ▶ 按一下「交易資料」，再按一下「繼續」。

圖表 8-2
從交易資料取得 RFM, 「變數」索引標籤



- ▶ 按一下**重設**，清除之前的任何設定。
- ▶ 若欲取得交易日期，請選取「**購買日期 [日期]**」。
- ▶ 若欲取得交易金額，請選取「**交易金額 [金額]**」。
- ▶ 若欲取得摘要方法，請選取「**總數**」。
- ▶ 若欲取得客戶識別碼，請選取「**客戶 ID[ID]**」。
- ▶ 然後按一下「**輸出**」索引標籤。

圖表 8-3
從交易資料取得 RFM, 「輸出」索引標籤



- ▶ 選取 (勾選) 「bin 計數資料圖」。
- ▶ 按一下「確定」執行程序。

評估結果

當您從交易資料計算 RFM 分數時，將會建立一個包含新的 RFM 分數的新資料集。

圖表 8-4
從交易資料集取得 RFM

ID	Date_most_recent	Transaction_count	Amount	Recency_score	Frequency_score	Monetary_score	RFM_score
1	05/17/2006	10	1313.00	2	3	5	235
2	09/21/2005	11	1230.00	1	5	4	154
3	08/11/2006	13	1194.00	3	5	2	352
4	05/24/2006	9	794.00	2	3	2	232
5	03/13/2005	3	278.00	1	1	2	112
6	07/28/2006	9	922.00	3	2	4	324
7	06/20/2006	11	961.00	2	4	2	242

依照預設值，資料集包含下列資訊給每位客戶：

- 客戶 ID 變數
- 最近交易的日期
- 購買次數的總和。

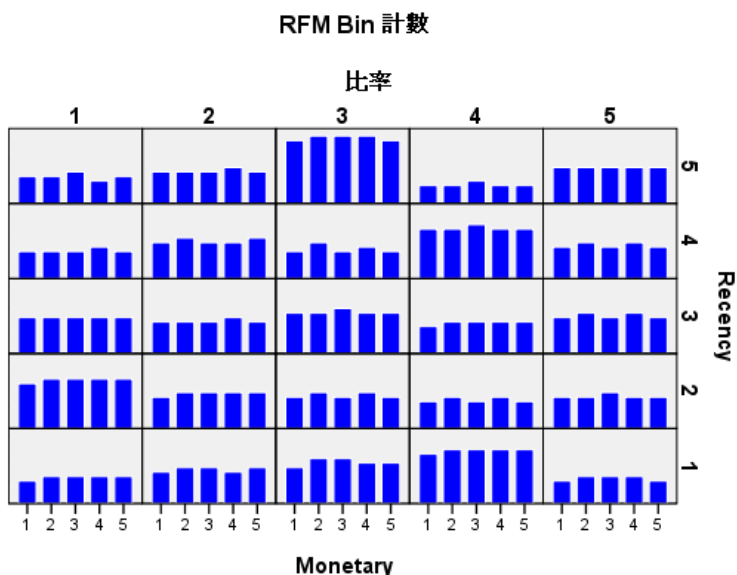
- 交易金額摘要（預設為總和1）
- 新近值、頻率值、金融值和組合 RFM 分數

新的資料集只包含一列（一筆記錄）對應到每位客戶。原始交易資料與「客戶識別碼」變數的數值整合。識別碼變數一定會包含在新資料集裡，否則您將無法用 RFM 分數和客戶進行比對。

每位客戶的組合 RFM 分數就是三筆個別分數的連鎖數值，計算方式為： $(\text{recency} \times 100) + (\text{frequency} \times 10) + \text{monetary}$ 。

bin 計數資料圖的圖表會顯示在「瀏覽器」視窗中，且顯示每個 RFM 分類裡的客戶數量。

圖表 8-5
bin 個數的圖表 (C)



使用預設的五種分數分類方法來計算三個 RFM 的構成成份，會得出 125 個可能的 RFM 分數分類。每一個圖上的長條代表在每個 RFM 分類中的客戶數量。

最理想的狀況是，您希望在每個 RFM 分數分類裡都有相對平均分配的客戶數量。實際上，通常會有某些程度的變異，如同您在此範例裡看到的。如果存在許多空白的分類，您可能要考慮改變 binning 的方法。

有幾個方法可以處理不平均分配的 RFM 分數，包括：

- 用巢狀取代自變的 binning。
- 減少可能分數分類的數量 (bins)。
- 當大量的同分數值出現時，隨機將同分數的觀察值設定到別的分類。

如需詳細資訊，請參閱第 6 頁第 2 章中的 RFM Binning。

合併分數資料和客戶資料

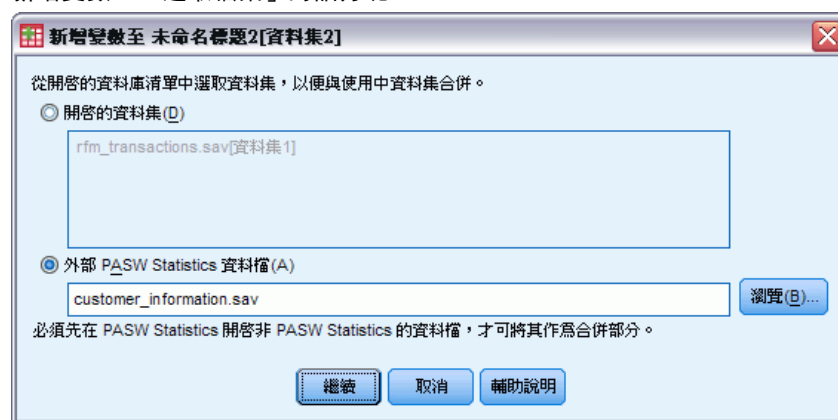
現在您已經擁有一個包含 RFM 分數的資料集，您需要將這些分數與客戶進行比對。您可以回頭把分數與交易資料檔案合併，但一般情況下您會希望將分數資料與一個像 RFM 資料集一樣每列（每筆記錄）都對應到每位客戶還包含客戶姓名、地址的檔案合併。

圖表 8-6
「變數檢視」中的RFM 資料集

名稱	類型	寬度	小數	標記	值
ID	數字的	5	0	Customer ID	無
Date_most_recent	日期	10	0	Date of most re...	無
Transaction_count	數字的	7	0	Number of tran...	無
Amount	數字的	8	2	Amount	無
Recency_score	數字的	3	0	Recency score	無
Frequency_score	數字的	3	0	Frequency score	無
Monetary_score	數字的	3	0	Monetary score	無
RFM_score	數字的	3	0	RFM score	無

- ▶ 將包含 RFM 分數的資料集設為作用中資料集。（在有資料集的「資料編輯程式」視窗中任一處按一下。）
- ▶ 從功能表選擇：
資料 > 合併檔案 > 新增變數

圖表 8-7
新增變數、「選取檔案」對話方塊



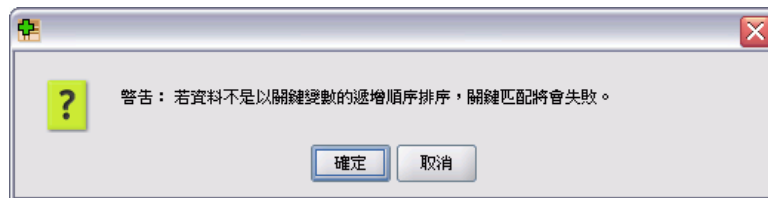
- ▶ 選取「一個外部資料檔」。
- ▶ 使用「瀏覽」按鈕以瀏覽至Samples 資料夾，並選取 customer_information.sav. 如需詳細資訊，請參閱第 91 頁附錄 A 中的範例檔案。
- ▶ 然後按一下繼續。

圖表 8-8
新增變數、「選取檔案」對話方塊



- ▶ 選取（勾選）「依已排序檔案中的關鍵變數匹配觀察值」。
- ▶ 選取「兩個檔案都提供觀察值」。
- ▶ 選取「ID 關鍵變數表」。
- ▶ 按一下「確定」。

圖表 8-9
新增「變數提示訊息」



請注意訊息，提示您兩個檔案必須依照關鍵變數遞增排序。在此範例中，兩個檔案已經依照關鍵變數遞增排序，即依照我們計算 RFM 分數時選取的識別碼變數排序。當您從交易資料計算 RFM 分數時，最新的資料集即會自動依照識別碼變數遞增排序。如果您改變了分數資料集的排序方式、或者您欲和分數資料集合併的檔案並非依照指定方式排序，您首先必須將兩個檔案依照客戶識別碼變數排序。

- ▶ 按一下「確定」以合併兩個資料集。

該包含 RFM 分數的資料集現在也包含了每位客戶的姓名、地址及其他資訊。

圖表 8-10
完成合併的資料集

名稱	類型	寬度	小數	標記	值
ID	數字的	5	0	Customer ID	無
Date_most_recent	日期	10	0	Date of most re...	無
Transaction_count	數字的	7	0	Number of tran...	無
Amount	數字的	8	2	Amount	無
Recency_score	數字的	3	0	Recency score	無
Frequency_score	數字的	3	0	Frequency score	無
Monetary_score	數字的	3	0	Monetary score	無
RFM_score	數字的	3	0	RFM score	無
Name	字串	4	0		無
Address	字串	7	0		無
City	字串	4	0		無
State_Province	字串	14	0		無
PostalCode	字串	11	0		無
Country	字串	7	0		無
Gender	數字的	1	0		{0, Female}...
AgeCategory	數字的	1	0	Age Category	{1, <25}...

集群分析

「集群分析」這種探索工具旨在顯示資料中的自然分組（或集群）。例如，可以根據不同人口資料和購買特性，識別出不同組別的客户。

例如，公司的直效行銷部門想要識別其客户資料庫中的人口統計分組，以協助決定市場行銷策略及開發新產品。

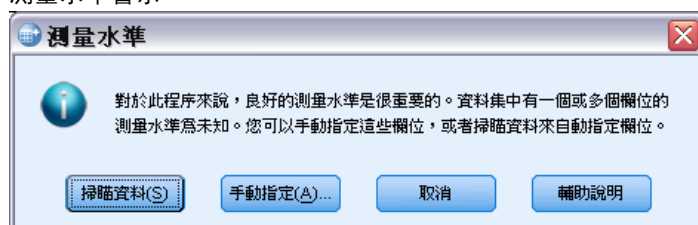
此資訊收集於 dndata.sav 中。如需詳細資訊，請參閱第 91 頁附錄 A 中的範例檔案。

執行分析

- ▶ 若要執行「集群分析」，請從功能表中選擇：
直效行銷 > 選擇技術
- ▶ 選取「將我的聯絡人區隔為數個集群」然後按一下「繼續」。

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

圖表 9-1
測量水準警示



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

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

在此範例檔案中，沒有具有未知測量水準的欄位，且所有的欄位都有正確的測量水準；因此不會顯示測量水準的警示。

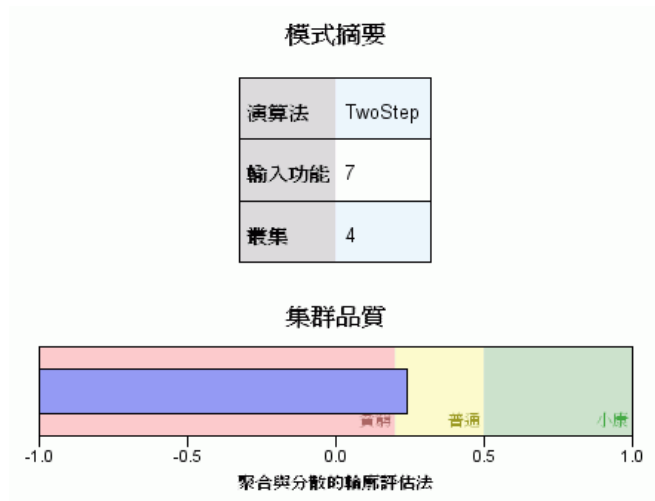
圖表 9-2
「集群分析」，「欄位」索引標籤



- ▶ 選取下列欄位以建立片段：「年齡」、「收入類別」、「教育水準」、「現址居住年數」、「性別」、「已婚」和「子女」。
- ▶ 按一下「執行」執行程序。

輸出

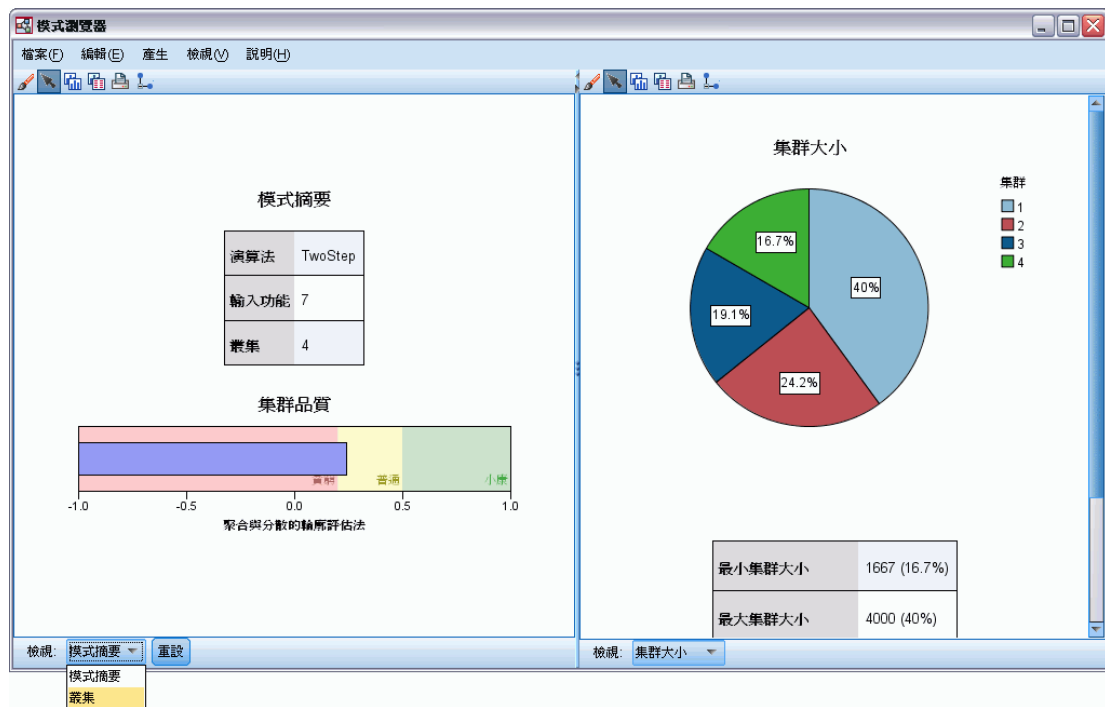
圖表 9-3
 集群模式摘要



結果會在「集群模式瀏覽器」中顯示。

- 模式摘要顯示根據您選取的七種輸入功能（欄位），找到了四個集群。
 - 集群品質圖表顯示整體模式品質位在「可」範圍的中間。
- ▶ 連按兩下「集群模式瀏覽器」來啟動「模式瀏覽器」。

圖表 9-4
 已啟動的「集群模式瀏覽器」



- 從「集群瀏覽器」視窗底部的「檢視」下拉式清單，選取「集群」。

圖表 9-5
「集群」檢視

集群	1	2	3	4
標記				
說明				
大小	40.0% (4000)	24.2% (2424)	19.1% (1909)	16.7% (1667)
功能	Age 50.30	Age 44.07	Age 39.05	Age 33.09
	Children 1.58	Children 1.29	Children 0.39	Children 0.12
	Gender Male (57.0%)	Gender Female (100.0%)	Gender Male (100.0%)	Gender Female (50.9%)
	Income category (thousands) 75+ (56.1%)	Income category (thousands) 50-74 (47.2%)	Income category (thousands) 75+ (34.8%)	Income category (thousands) <25 (100.0%)
	Married Yes (100.0%)	Married No (78.5%)	Married No (100.0%)	Married No (78.5%)
	Education Post-graduate (20.5%)	Education Post-graduate (20.5%)	Education College (21.1%)	Education Post-graduate (20.6%)
	Years at current residence 9.47	Years at current residence 9.51	Years at current residence 9.47	Years at current residence 9.42

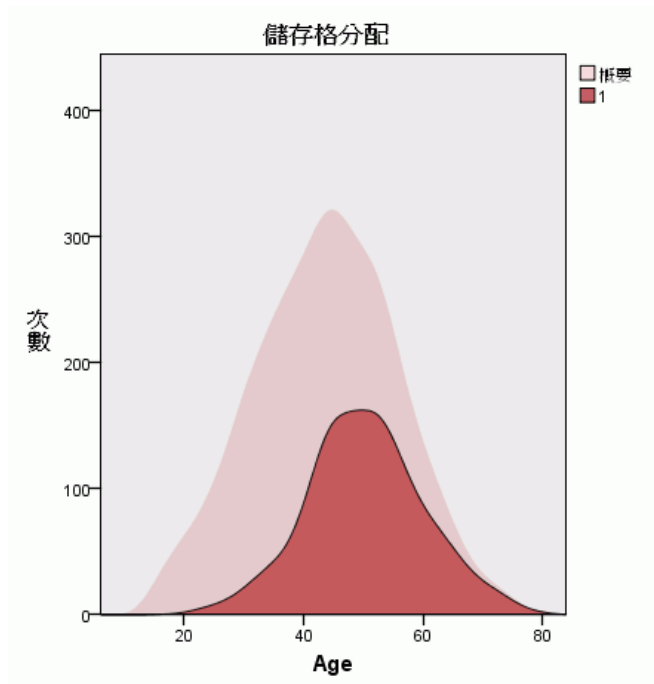
「集群」檢視會顯示每個集群屬性上的資訊。

- 對於連續（尺度）欄位，會顯示平均值。
- 對於類別（名義、次序）欄位，會顯示眾數。眾數是指記錄個數最多的類別。在本範例中，每個記錄都是一位客戶。
- 依照預設值，欄位是以對模式的整體重要性依序顯示。在本範例中，「年齡」擁有最高的整體重要性。您也可以依照集群內重要性或字母順序來排序欄位。

如果您選取（按一下）「集群」檢視中的任一個儲存格，您就可以看到該集群的圖表，其中摘要顯示該欄位的值。

- ▶ 例如，選取集群 1 的「年齡」儲存格。

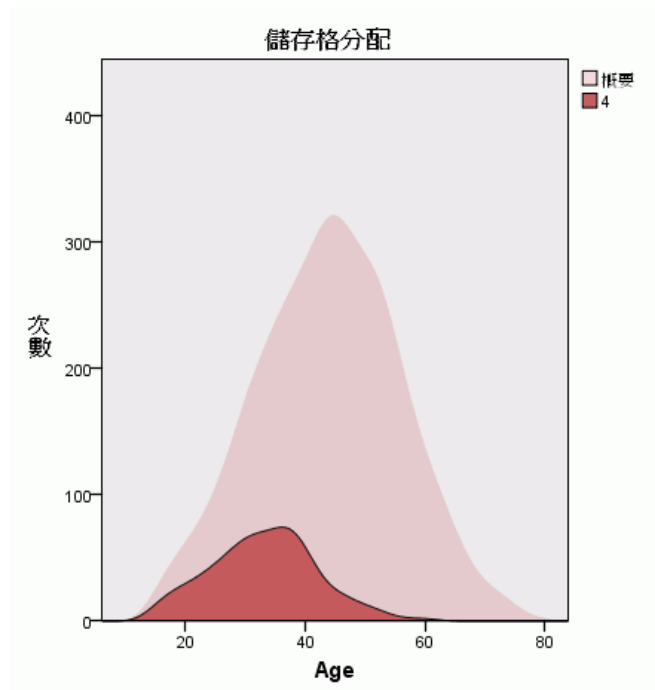
圖表 9-6
集群 1 的「年齡」直方圖



對於連續欄位，會顯示直方圖。直方圖會顯示在該集群中值的分配，以及欄位的整體值分配。直方圖顯示集群 1 中的客戶較為年長。

- ▶ 在「集群」檢視中選取集群 4 的「年齡」儲存格。

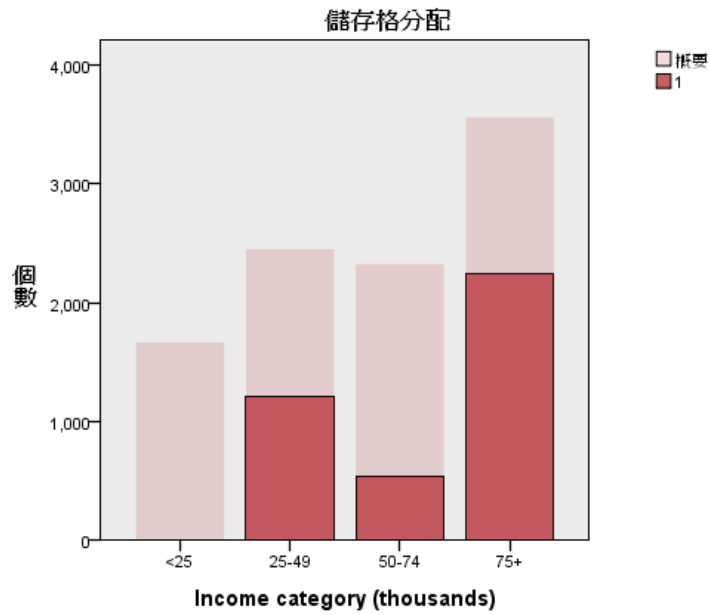
圖表 9-7
集群 4 的「年齡」直方圖



與集群 1 相反，集群 4 中的客戶比整體平均要年輕。

- ▶ 在「集群」檢視中選取集群 1 的「收入類別」儲存格。

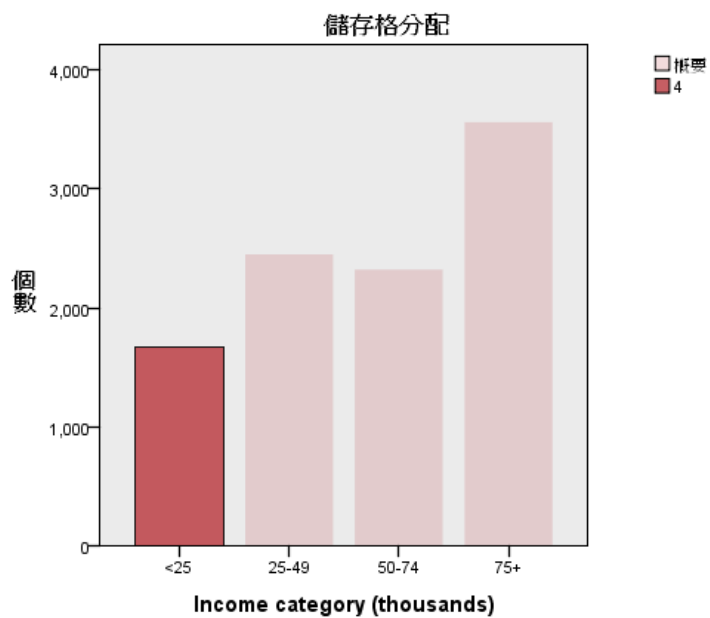
圖表 9-8
集群 1 的「收入類別」長條圖



對於類別欄位，會顯示長條圖。此集群的收入類別直方圖中，最明顯的特色就是完全不含最低收入類別的客戶。

- ▶ 在「集群」檢視中選取集群 4 的「收入類別」儲存格。

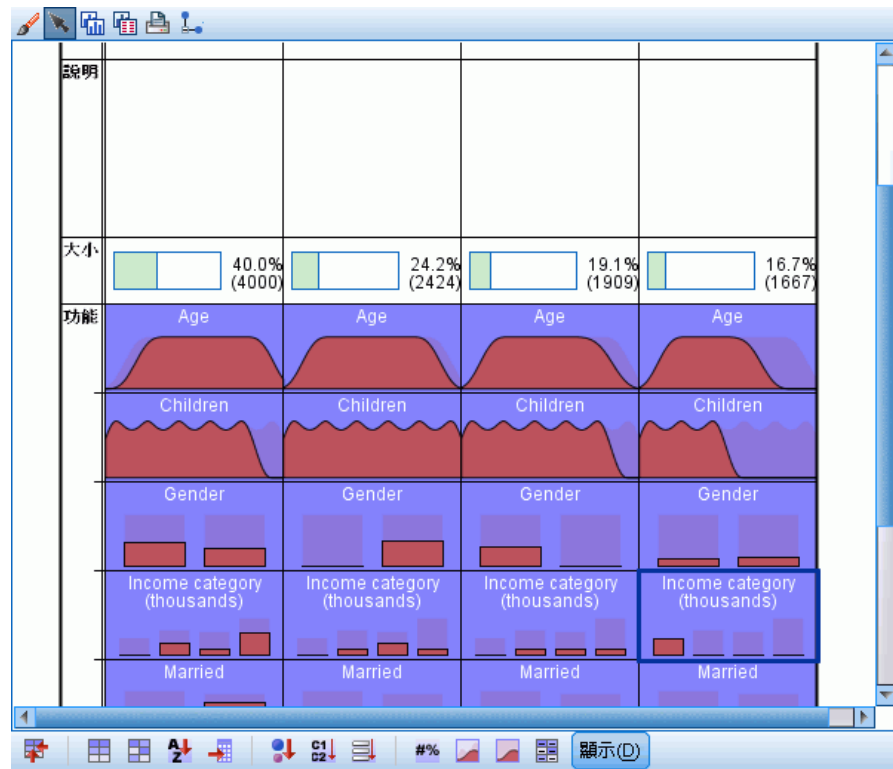
圖表 9-9
集群 4 的「收入類別」長條圖



與集群 1 相反，集群 4 中的所有客戶都是在最低的收入類別中。

您也可以使用「模式瀏覽器」視窗底部的工具列來變更「集群」檢視，以在儲存格中顯示圖表，這樣可以更快速地比較集群之間值的分配。

圖表 9-10
在「集群」中顯示的圖表



從「集群」檢視以及每個儲存格圖表中提供的額外資訊，您便可以得知集群間某些明顯的差異：

- 集群 1 中的客戶年齡較長、已婚且育有子女的人數較多，且收入也較高。
- 集群 2 中的客戶大多為年齡較長的單親母親，具有中等收入。
- 集群 3 中的客戶大多為年輕單身的男性，並且無子女。
- 集群 4 中的客戶大多為年輕單身的女性，並且無子女，收入也較低。

「集群」檢視中的「說明」儲存格是文字欄位，您可以編輯並加入每個集群的說明。

圖表 9-11
含集群說明的「集群」檢視

集群	1	2	3	4
標記				
說明	Older, married, have children, higher income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
大小	40.0% (4000)	24.2% (2424)	19.1% (1909)	16.7% (1667)
功能	Age 50.30	Age 44.07	Age 39.05	Age 33.09
	Children 1.58	Children 1.29	Children 0.39	Children 0.12
	Gender Male (57.0%)	Gender Female (100.0%)	Gender Male (100.0%)	Gender Female (50.9%)
	Income category (thousands) 75+ (56.1%)	Income category (thousands) 50-74 (47.2%)	Income category (thousands) 75+ (34.8%)	Income category (thousands) <25 (100.0%)
	Married Yes (100.0%)	Married No (78.5%)	Married No (100.0%)	Married No (78.5%)
	Education Post-graduate (20.5%)	Education Post-graduate (20.5%)	Education College (21.1%)	Education Post-graduate (20.6%)
	Years at current residence 9.47	Years at current residence 9.51	Years at current residence 9.47	Years at current residence 9.42

根據集群來選取記錄

您可以用兩種方式，根據集群組員來選取記錄：

- 在「集群模式瀏覽器」中互動建立過濾條件。
- 使用程序產生的集群欄位數值來指定過濾或選擇條件。

在集群模式瀏覽器中建立過濾器

在「集群模式瀏覽器」中建立根據特定集群選取記錄的過濾條件：

- ▶ 啟動（連按兩下）「集群模式瀏覽器」。

- ▶ 從「集群瀏覽器」視窗底部的「檢視」下拉式清單，選取「集群」。
- ▶ 在「集群檢視」上方，按一下您要的集群號碼。如果您要選取多個集群，按住 Ctrl 再按一下要選取的每個集群號碼。

圖表 9-12
「集群」檢視中選取的集群

集群 標記	1	2	3	4
說明	Older, married, have children, higher income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
大小	40.0% (4000)	24.2% (2424)	19.1% (1909)	16.7% (1667)
功能	Age 50.30	Age 44.07	Age 39.05	Age 33.09
	Children 1.58	Children 1.29	Children 0.39	Children 0.12
	Gender Male (57.0%)	Gender Female (100.0%)	Gender Male (100.0%)	Gender Female (50.9%)
	Income category (thousands) 75+ (56.1%)	Income category (thousands) 50-74 (47.2%)	Income category (thousands) 75+ (34.8%)	Income category (thousands) <25 (100.0%)
	Married Yes (100.0%)	Married No (78.5%)	Married No (100.0%)	Married No (78.5%)
	Education Post-graduate (20.5%)	Education Post-graduate (20.5%)	Education College (21.1%)	Education Post-graduate (20.6%)
	Years at current residence 9.47	Years at current residence 9.51	Years at current residence 9.47	Years at current residence 9.42

- ▶ 從「集群模式瀏覽器」功能表選擇：
產生 > 過濾記錄

圖表 9-13
「過濾記錄」對話方塊



- ▶ 輸入過濾欄位名稱，並按一下「確定」。名稱必須符合 IBM® SPSS® Statistics 命名規則。

圖表 9-14
「資料編輯程式」中的過濾記錄

	ID	Married	Children	Region	ClusterGroup1	clusters_1_2
14	03623	No	0	West	3	.00
15	01353	No	0	West	3	.00
16	07055	No	0	West	3	.00
17	04455	No	0	West	2	1.00
18	07210	No	1	West	2	1.00
19	08054	No	0	West	4	.00
20	06937	No	0	West	4	.00
21	06512	No	0	West	4	.00
22	08315	No	0	West	4	.00
23	09676	No	3	West	2	1.00
24	09636	No	0	West	4	.00
25	08579	No	1	West	2	1.00
26	01480	No	1	West	2	1.00

這會在資料集中建立新欄位，並且會根據該欄位的值來過濾資料集中的紀錄。

- 過濾欄位值為 1 的記錄會包含在後續的分析、圖表和報告中。
- 過濾欄位值為 0 的記錄則會被排除。
- 排除的紀錄不會由資料集刪除，而會加以保留，並加上過濾器狀態指標；這會在「資料編輯程式」中以對角斜線穿過記錄編號的方式表示。

根據集群欄位值選取記錄

依照預設值，「集群分析」會建立可識別每個記錄的集群群組的新欄位。此欄位的預設名稱是「ClusterGroupn」，其中 n 是形成為一欄位名稱的整數。

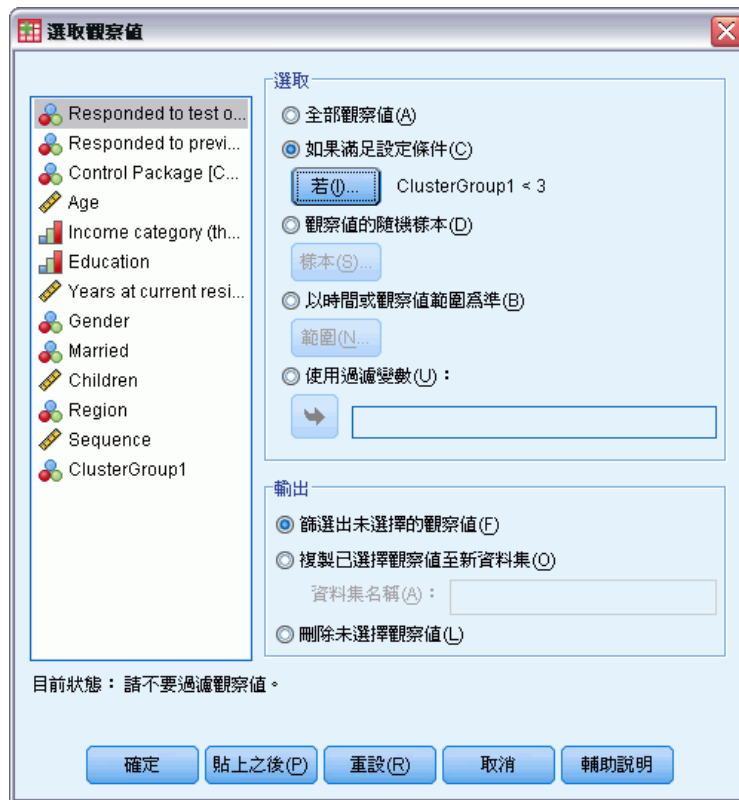
圖表 9-15
新增至資料集的集群欄位

	ID	Gender	Married	Children	Region	ClusterGroup1
1	01359	Female	No	0	West	4
2	06262	Female	No	1	West	2
3	08031	Male	No	0	West	3
4	01971	Male	No	0	West	4
5	09689	Male	No	0	West	3
6	06108	Male	No	1	West	3
7	09853	Male	No	0	West	3
8	06802	Male	No	0	West	4
9	07597	Male	No	0	West	3
10	03692	Male	No	1	West	3
11	00071	Male	No	0	West	4
12	00769	Male	No	0	West	3

使用集群欄位的值來選取特定集群中的紀錄：

- ▶ 從功能表選擇：
資料 > 選擇觀察值

圖表 9-16
「選擇觀察值」對話方塊



- ▶ 在「選擇觀察值」對話方塊中，選取「如果滿足設定條件」，然後按一下「若」。

圖表 9-17
選擇觀察值：「若」對話方塊



- ▶ 輸入選擇條件。

例如，`ClusterGroup1 < 3` 會選取所有集群 1 和 2 中的記錄，並且會排除集群 3 及之後的紀錄。

- ▶ 按一下「繼續」。

在「選擇觀察值」對話方塊中，有數個選項可用於選取及未選取的紀錄：

過濾掉未選取的觀察值。 這會建立指定過濾條件的新欄位。排除的紀錄不會由資料集刪除，而會加以保留，並加上過濾器狀態指標；這會在「資料編輯程式」中以對角斜線穿過記錄編號的方式表示。這等同於在「集群模式瀏覽器」中互動選取集群。

複製選取的觀察值到新的資料集。 這會在目前的作業階段中建立新的資料集，其中僅包含符合過濾條件的記錄。原始資料集不會受到影響。

刪除未選取的觀察值。 自資料集刪除未選取的記錄。如果要復原已刪除的記錄，必須離開檔案且不儲存任何變更，再重新開啟檔案。如果您將變更儲存於資料檔，觀察值便永久刪除。

「選取觀察值」對話方塊也具有使用現有變數作為過濾變數（欄位）的選項。如果您在「集群模式瀏覽器」中互動建立過濾條件，並且將產生的過濾欄位儲存至資料集，您便可以在後續的作業階段中使用該欄位來過濾記錄。

摘要

「集群分析」是能夠顯示資料中的自然分組（或集群）的有用探索工具。您可以使用來自集群的資訊來決定市場行銷策略及開發新產品。您可以根據集群組員來選取記錄，供進一步分析使用，或用於鎖定的市場行銷活動。

潛在客戶分析概要

「潛在客戶分析概要」使用上一個活動或測試活動的結果來建立描述性分析概要。您可以在未來的活動中會使用這些分析概要來鎖定特定的群組。例如，公司的直效行銷部門根據測試郵件的結果，以人口統計資料來產生哪一類型的人最可能回應特定類型優惠的分析概要。接著直效行銷部門會根據結果來判定如有類似的優惠，應使用哪一類的郵寄名單。

例如，公司的直效行銷部門會將測試郵件寄給其總客戶資料庫中約 20% 的客戶。此測試郵件的結果會記錄在資料檔案中，檔案中也包含每個客戶的人口統計特性，例如年齡、性別、婚姻狀況和地理區域。這些結果以簡單的「是/否」方式記錄，指出測試郵件的客戶中哪一些有回應（購買），哪一些沒有回應。

此資訊收集於 dmdata.sav 中。如需詳細資訊，請參閱第 91 頁附錄 A 中的範例檔案。

資料考量

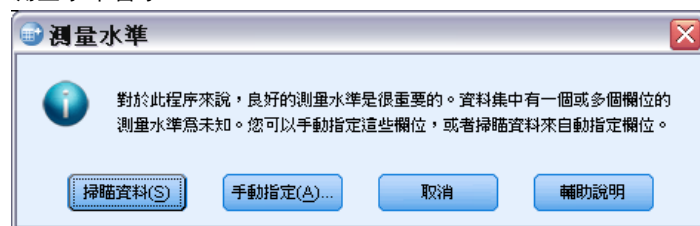
回應欄位應為類別，其中以值一個代表所有的正回應。任何其他非遺漏值皆假設為負回應。如果回應欄位代表連續（尺度）值，例如購買次數或購買金額，您必須建立一個新欄位將單一正回應值指定給所有非零的回應值。如需詳細資訊，請參閱第 21 頁第 4 章中的建立類別回應欄位。

執行分析

- ▶ 若要執行「潛在客戶分析概要」分析，請從功能表中選擇：
直效行銷 > 選擇技術
- ▶ 選取「產生回應優惠的聯絡人分析概要」，再按一下「繼續」。

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

圖表 10-1
測量水準警示



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

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

在此範例檔案中，沒有具有未知測量水準的欄位，且所有的欄位都有正確的測量水準；因此不會顯示測量水準的警示。

圖表 10-2
「潛在客戶分析概要」，「欄位」索引標籤



- ▶ 對於「回應」欄位，選取「已回應測試優惠」。
- ▶ 對於「正」回應值，請選取下拉式清單中的「是」。文字欄位中顯示的是數值 1，這是因為「是」實際上是與所記錄的數值 1 相關聯的數值標記（如果正回應值未定義數值標記，您可以直接在文字欄位中輸入該值）。
- ▶ 對於「建立分析概要方法」，選取「年齡」、「收入類別」、「教育水準」、「現址居住年數」、「性別」、「已婚」、「地區」和「子女」。
- ▶ 按一下「設定」索引標籤。

圖表 10-3
「潛在客戶分析概要」，「設定」索引標籤

潛在客戶分析概要

欄位 設定 (N)

最小側寫群組大小(M): 100

在結果中包含最小回應率門檻資訊(I)

指定目標回應率 (%) (T): 7

執行(R) 貼上之後(P) 重設(R) 取消 輔助說明

- ▶ 選取（勾選）「結果中包含最小回應率門檻資訊」。
- ▶ 輸入 7 作為目標回應率。
- ▶ 然後按一下「執行」執执行程序。

輸出

圖表 10-4
回應率表格

數目	縱斷面			
	說明	群組大小	回應率	累積回應率
1	Region = "West","South","East" Gender = "Female" Married = "No"	379	9.2%	9.2%
2	Region = "West","South","East" Gender = "Female" Married = "Yes"	299	5.0%	7.4%
3	Region = "West","South","East" Gender = "Male"	722	4.7%	6.0%
4	Region = "North"	517	2.5%	5.1%

綠色：達成目標回應率。
紅色：未達成目標回應率。

回應率表顯示由此程序所識別每個分析概要群組的資訊。

- 分析概要會以遞減順序或回應率顯示。
- 回應率是正回應（購買）的客戶百分比。
- 累積回應率為目前分析概要群組回應率與所有先前分析概要群組回應率的合併。由於設定檔是以回應率遞減順序顯示的，這表示累積回應率是目前分析概要群組其回應率，加上所有回應率較高分析概要群組其回應率的合併回應率。
- 分析概要說明僅包括對模式有顯著貢獻的欄位特性。在本範例中，模式中包含地區、性別與婚姻狀況。剩餘的欄位 — 年齡、教育與居住在目前地址的年數 — 並不包含在此模式中，因為這些欄位對於模式沒有顯著貢獻。
- 表格的綠色區域代表累積回應率等於或大於指定目標回應率（本範例為 7%）的分析概要集合。
- 表格的紅色區域代表累積回應率小於指定目標回應率的分析概要集合。
- 表格最後一列的累積回應率是測試郵件中包含的所有客戶整體或平均回應率，因為這一系列是所有分析概要群組的回應率。

表格顯示的結果表示如果您的目標女性位在西方、南方與東方，您得到的回應率會略高於目標回應率。

但請注意，這些地區未婚女性（9.2%）與已婚女性（5.0%）的回應率之間有很大的差異。雖然這兩組的累積回應率大於目標回應率，但事實上單單已婚婦女組的回應率就大於目標回應率，這表示您應該找出其他可改進此模式的特性。

智慧型輸出

圖表 10-5
智慧型輸出

回應率表格會顯示由程序所指定之每個側寫群組的資料。側寫說明包含僅對模式提供重大貢獻的欄位特性。不包含無重大貢獻的欄位。

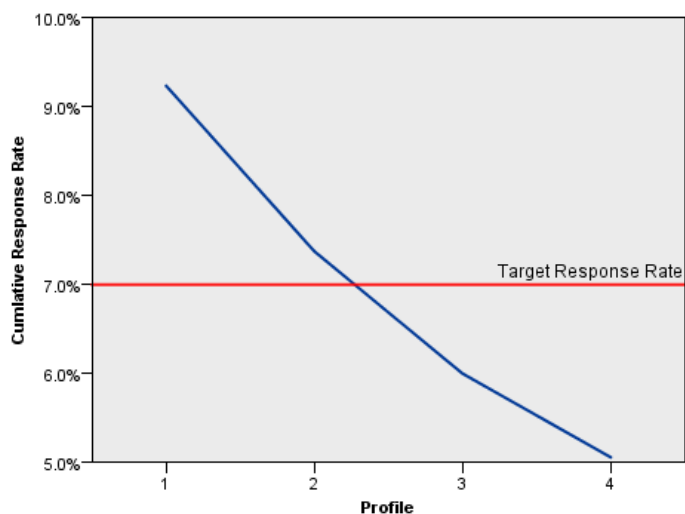
側寫會以回應率的依遞減排序顯示。回應率是具有正向回應（進行購買）的客戶百分比。

累積回應率為目前和先前所有側寫群組的組合回應率。由於側寫會以回應率的依遞減排序顯示，表示累積回應率為目前側寫群組加上所有具有較高回應率的側寫群組的組合回應率。

指定的目標回應率為 7.00%。綠色列的累積回應率高於 7.00%，而紅色列的累積回應率低於 7.00%。雖然有些在綠色區域的側寫群組的個別回應率可能低於 7.00%，但是該點的累積回應率高於 7.00%。

表格隨附「智慧型輸出」可提供有關如何解讀表格的「般資訊，以及有關表格中結果的特定資訊。

圖表 10-6
累積回應率圖表



累積回應率圖表基本上以視覺方式表示表格中顯示的累積回應率。由於系統以回應率遞減順序報告分析概要，因此每個後續分析概要的累積回應率曲線一律向下。和表格一樣，圖表顯示累積回應率落在目標回應率之下，介於分析概要群組 2 與分析概要群組 3 之間。

摘要

系統針對此特定的測試郵件，識別出四個分析概要群組，結果表示與是否回應優惠有關的顯著人口統計特性只有性別、地區與婚姻狀況。最高回應率群組包含居住在南方、東方與西方的未婚女性。在這之後，回應率快速下降，雖然相同地區的已婚女性所產生的累積回應率仍比目標回應率大。

郵遞區號回應率

此方法使用上一個活動的結果來計算郵遞區號回應率。在未來的活動中會使用這些回應率來鎖定特定的郵遞區號。

例如，公司的直效行銷部門可以根據先前的郵件結果，來產生郵遞區號回應率。接著，該部門可以依據郵件中所包含的各種條件，如最小可接受回應率和/或最大聯絡人個數，來鎖定特定的郵遞區號。

此資訊收集於 dmdata.sav 中。如需詳細資訊，請參閱第 91 頁附錄 A 中的範例檔案。

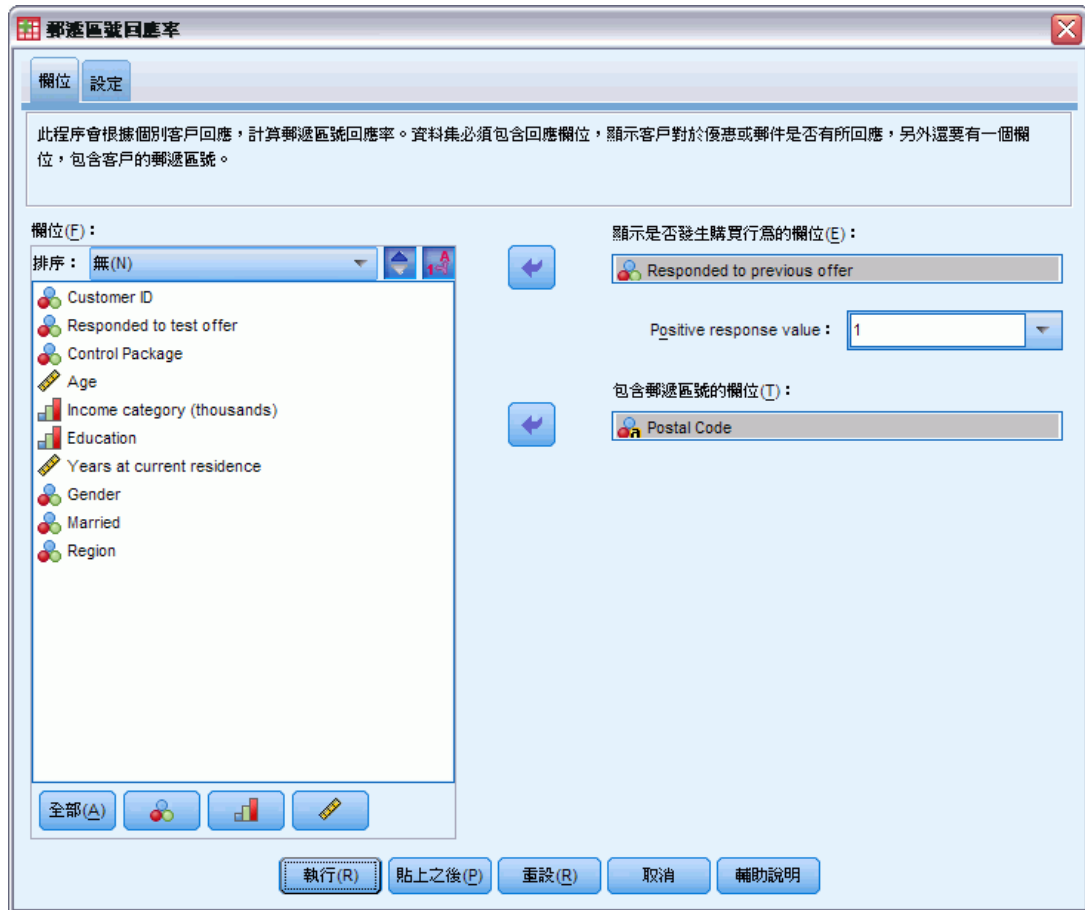
資料考量

回應欄位應為類別，其中以值一個代表所有的正回應。任何其他非遺漏值皆假設為負回應。如果回應欄位代表連續（尺度）值，例如購買次數或購買金額，您必須建立一個新欄位將單一正回應值指定給所有非零的回應值。如需詳細資訊，請參閱第 28 頁第 5 章中的建立類別回應欄位。

執行分析

- ▶ 若要計算郵遞區號回應率，請從功能表選擇：
直效行銷 > 選擇技術
- ▶ 選取「識別高分回應郵遞區號」，再按一下「繼續」。

圖表 11-1
「郵遞區號回應率」，「欄位」索引標籤



- ▶ 對於「回應」欄位，選取「已回應過去優惠」。
- ▶ 對於「正」回應值，請選取下拉式清單中的「是」。文字欄位中顯示的是數值 1，這是因為「是」實際上是與所記錄的數值 1 相關聯的數值標記（如果正回應值未定義數值標記，您可以直接在文字欄位中輸入該值）。
- ▶ 對於「郵遞區號」欄位，選取「郵遞區號」。
- ▶ 按一下「設定」索引標籤。

圖表 11-2
「郵遞區號回應率」，「設定」索引標籤

The screenshot shows the '郵遞區號回應率' (Zip Code Response Rate) application window with the '設定 (N)' (Settings) tab selected. The window contains the following settings:

- 郵遞區號分組依據 (Grouping Criteria):**
 - 完整的值 (C)
 - 前 3 位數或字元 (5)
 - 前 5 位數或字元 (5)
 - 前 N 位數或字元 (E)
 - V:
- 郵遞區號數字格式 (Digit Format):**
 - 郵遞區號原始記錄方式
 - 3 位數 (D)
 - 5 位數 (I)
 - 9 位數 (9)
 - 其他 (O)
 - 位數 (G):
- 輸出 (Output):**
 - 回應率與能力分析 (S)
 - 最小可接受回應率 (Minimum Acceptable Response Rate):
 - 無最小值 (U)
 - 目標回應率 (%) (E)
 - 從公式計算收支平衡率 (K)
 - 郵寄包裹成本 (I):
 - 每回應淨收入 (P):
 - 最大聯絡人個數 (Maximum Number of Contacts):
 - 所有聯絡人 (A)
 - 聯絡人百分比 (T)
 - 聯絡人數目 (M)
- 匯出至 Excel (Export to Excel):**
 - 將郵遞區號回應率儲存至 Excel (A)
 - 檔案名稱 (L):
 - 瀏覽 (B)...

At the bottom of the window, there are five buttons: 執行 (R) (Execute), 貼上之後 (P) (Paste After), 重設 (R) (Reset), 取消 (Cancel), and 輔助說明 (Help).

- ▶ 在「郵遞區號分組依據」群組中，選取「前 3 位數或字元」。這會計算所有郵遞區號開頭前 3 位數或字元相同的聯絡人其合併回應率。例如，美國郵遞區號的前 3 位數所代表的共用地理區域範圍大於整個 5 位數郵遞區號所代表的範圍。
- ▶ 在「輸出」群組中，選取（勾選）「回應率與能力分析」。
- ▶ 選取「目標回應率」，並輸入數值 5。
- ▶ 選取「聯絡人數目」，並輸入數值 5000。
- ▶ 然後按一下「執行」執行程序。

輸出

圖表 11-3
含郵遞區號回應率的新資料集

	PostalCode	ResponseRate	Responses	Contacts	Index	Rank
1	932	10%	4	40	3.6	Top 10%
2	098	9%	6	68	5.5	Top 10%
3	740	8%	9	116	8.3	Top 10%
4	100	8%	7	91	6.5	Top 10%
5	110	8%	5	65	4.6	Top 10%
6	954	8%	4	53	3.7	Top 10%
7	108	7%	6	82	5.6	Top 10%
8	107	7%	5	71	4.6	Top 10%
9	090	7%	4	58	3.7	Top 10%
10	966	7%	4	58	3.7	Top 10%
11	760	7%	8	119	7.5	Top 10%
12	113	6%	5	80	4.7	Top 10%

系統會自動建立新的資料集。此資料集包含每個郵遞區號的單一記錄（列）。在此範例中，每列包含所有開頭前 3 位數或字元相同的郵遞區號的摘要資訊。

除了包含郵遞區號的欄位外，新資料集也包含下列欄位：

- **回應率。**各郵遞區號的正回應百分比。系統會自動以回應率的遞減順序排序記錄，讓回應率最高的郵遞區號會出現在資料集頂端。
- **反應值。**各郵遞區號的正回應數目。
- **聯絡人。**各郵遞區號中包含回應欄位其非遺漏值的聯絡人總數目。
- **索引。**採用公式 $N \times P \times (1-P)$ 「加權後」的回應，其中 N 是聯絡人個數， P 是以比例表示的回應率。對於兩個回應率相同的郵遞區號，此公式會將較高的索引值指定給聯絡人數目較多的郵遞區號。
- **等級。**累積郵遞區號回應率的十分位等級（前 10%、後 20% 等），以遞減順序排序。

由於我們在「郵遞區號回應率」對話方塊的「設定」索引標籤上選取了「回應率與能力分析」，因此在「瀏覽器」中會顯示摘要回應率表格與圖表。

圖表 11-4
回應率表格

Percentile	回應率	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

綠色：達成目標回應率。
紅色：未達成目標回應率。

此表格會以遞減順序，按照十分位等級（前 10%，前 20% 等）來摘要結果。

- 累積回應率是目前列與所有先前列中正回應的合併百分比。由於結果是以回應率的遞減順序顯示的，因此此結果是將目前十分位數的回應率與所有回應率與較高的十分位數其回應率合併在一起。
- 系統會根據您為回應率輸入的值，以及聯絡人上限，將表格予以彩色編碼。累積回應率等於或大於 5% 和 5,000 的列，或累積聯絡人較少的列，會以綠色編碼。系統會根據先達到哪一個門檻值來進行彩色編碼。在此範例中，相同十分位數的兩個門檻值皆有達到。

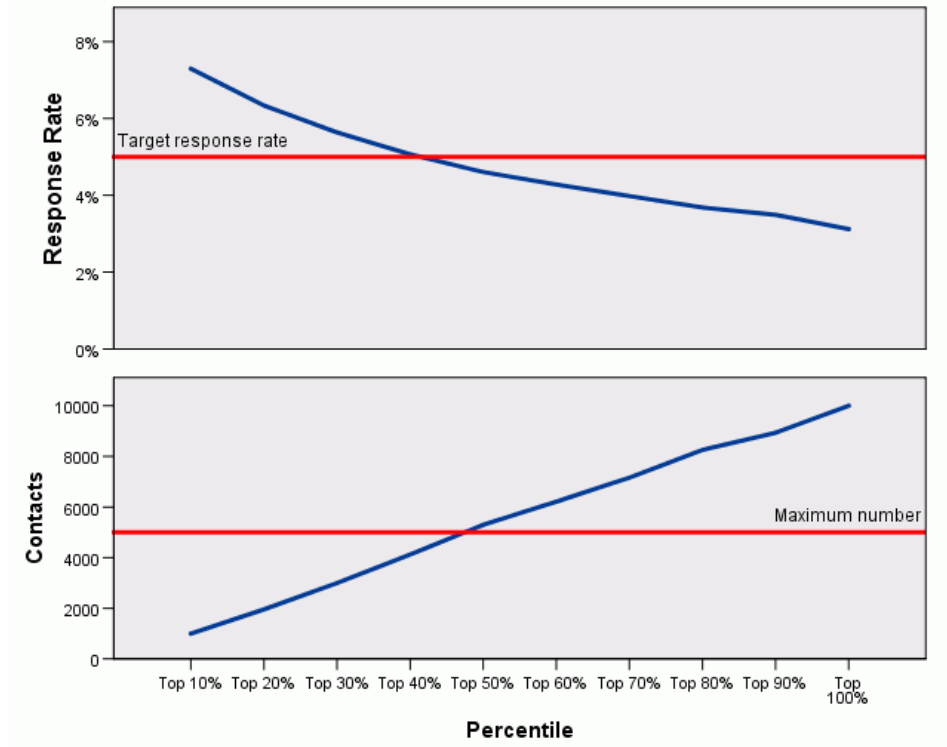
圖表 11-5
回應率表格的智慧型輸出

回應率表會依十分位數遞減順序概括結果（前 10%、前 20% 等）。累積回應率為目前和先前所有列的正回應合併百分比。由於結果會依回應率遞減順序顯示，因此會合併目前的十分位數與所有較高回應率的十分位數。由於十分位數會包含在新資料集中，因此您可以輕易地識別出符合某些累積回應率的郵遞區號。識別十分某些累積回應率的郵遞區號。識別十分位數等級的新資料集欄位稱為 Rank，其中 1= 前 10%、2= 前 20%，依此類推。

指定最小回應率為 5.00%。指定最大聯絡人數目為 5000。表格的色彩編碼會以第一個達到的門檻值為根據。已達到相同類別中的兩個門檻值。綠色列的累積回應率高於或等於指定最小回應率，而聯絡人的累積數目低於或等於聯絡人指定最大數目。紅色列的累積回應率低於指定最小回應率，而聯絡人的累積數目高於聯絡人指定最大數目。

表格隨附提供如何讀取表格的一般說明文字。如果您已指定最小回應率或最大聯絡人個數，表格也將包含一個區段說明結果與指定門檻值之間的相關性。

圖表 11-6
累積回應率圖表



表示每個十分位數的累積回應率與累積聯絡人數目的圖表，會以視覺方式表示在回應率表格中顯示的相同資訊。聯絡人最小累積回應率與最大累積回應率的門檻值落在第 40 與 50 個百分位數之間。

- 由於圖表類型以回應率十分位等級的遞減順序顯示累積回應率，所以每個後續十分位的累積回應率曲線一律向下。
- 由於聯絡人數目的曲線代表累積的聯絡人數目，因此它一律向上。

表格與圖表中的資訊會告訴您如果要到達至少 5% 的回應率，但包含的人數不能過活動中 5,000 個聯絡人，您應注重在前四個十分位數的郵遞區號。由於十分位等級包含在新資料集中，所以您可以很容易識別符合前 40% 需求的郵遞區號。

圖表 11-7
新增資料集 (D)

	PostalCode	ResponseRate	Responses	Contacts	Index	Rank
49	965	3.57%	2.00	56	1.93	Top 40%
50	618	3.54%	4.00	113	3.86	Top 40%
51	603	3.53%	3.00	85	2.89	Top 40%
52	757	3.48%	4.00	115	3.86	Top 40%
53	948	3.39%	2.00	59	1.93	Top 40%
54	103	3.33%	3.00	90	2.90	Top 40%
55	608	3.33%	3.00	90	2.90	Top 40%
56	612	3.28%	4.00	122	3.87	Top 50%
57	762	3.23%	1.00	31	.97	Top 50%
58	033	3.23%	2.00	62	1.04	Top 50%

注意：等級會記錄為 1 到 10 的整數。欄位已定義數值標記，其中 1= 前 10%，2 = 前 20% 等。您會在「資料檢視」或「資料編輯程式」中看到實際的等級值或數值標記（視您在「檢視」中的設定為何）。

摘要

「郵遞區號回應率」使用上一個活動的結果來計算郵遞區號回應率。在未來的活動中會使用這些回應率來鎖定特定的郵遞區號。這個程序會建立一個新資料集，其中包含每個郵遞區號的回應率。根據新資料集中回應率表格與圖表的資訊，以及十分位等級資訊，您可以識別一組符合聯絡人特定最小累積回應率和/或最大聯絡人數目的郵遞區號。

傾向機率

「傾向機率」使用測試郵件或上一個活動的結果來產生傾向分數。此分數會根據各種所選的特性來識別哪些聯絡人最有可能回應。

這項技術採用二元 logistic 迴歸來建立預測模式。建立和套用預測模式的程序具有兩個基本步驟：

- ▶ 建立模式並儲存模式檔案。您使用已知相關結果（通常稱為**目標**）的資料及來建立模式。例如，如果您要建立一個模式以預測哪些人可能回應直效郵件活動，那麼您必須從已經包含相關資訊（哪些人有回應，哪些人沒有回應）的資料集開始。例如，這可能是對一小群客戶所進行的測試郵件結果，或是過去類似活動的回應資訊。
- ▶ 將該模式套用至不同的資料集（未知其相關結果）以取得預測結果。

本範例使用兩個資料檔：利用 `dmdata2.sav` 來建立模式，然後將該模式套用到 `dmdata3.sav`。如需詳細資訊，請參閱第 91 頁附錄 A 中的範例檔案。

資料考量

回應欄位（相關目標結果）應為類別，其中以值一個代表所有的正回應。任何其他非遺漏值皆假設為負回應。如果回應欄位代表連續（尺度）值，例如購買次數或購買金額，您必須建立一個新欄位將單一正回應值指定給所有非零的回應值。如需詳細資訊，請參閱第 35 頁第 6 章中的建立類別回應欄位。

建立預測模式

- ▶ 開啟資料檔 `dmdata2.sav`。

這個檔案包含收到測試郵件人員的各種人口特性，同時也包含他們是否回應郵件的資訊。這項資訊記錄在已回應欄位（變數）。數值 1 表示連絡人已回應郵件，數值 0 表示連絡人沒有回應。

圖表 12-1
「資料編輯程式」中的資料檔內容

ID	Responded	Previous	ControlPackage	PostalCode	Age	Income	Education	Reside	Gender
03179	0	0	0	93640	38	3	4	11	1
03647	1	0	1	93760	27	2	5	14	1
01741	0	0	1	93850	52	1	5	12	1
05388	0	0	0	93900	66	3	4	10	1
01942	0	0	0	93900	41	4	5	11	1
06254	0	0	1	94120	48	4	4	12	1
02164	0	0	1	94130	46	1	5	9	1
02865	1	0	1	94150	37	4	3	10	1
03330	0	0	1	94270	43	3	2	13	1

- ▶ 從功能表選擇：
直效行銷 > 選擇技術
- ▶ 選取「選取最有可能購買的聯絡人」，然後按一下「繼續」。

圖表 12-2
「傾向機率」，「欄位」索引標籤



- ▶ 對於「回應」欄位，選取「已回應測試優惠」。
- ▶ 對於「正」回應值，請選取下拉式清單中的「是」。文字欄位中顯示的是數值 1，這是因為「是」實際上是與所記錄的數值 1 相關聯的數值標記（如果正回應值未定義數值標記，您可以直接在文字欄位中輸入該值）。
- ▶ 對於「預測傾向方法」，選取「年齡」、「收入類別」、「教育水準」、「現址居住年數」、「性別」、「已婚」、「地區」和「子女」。
- ▶ 選取（勾選）「匯出模式資訊至 XML 檔案」。
- ▶ 按一下「瀏覽」以瀏覽至要儲存檔案的位置，然後輸入檔案名稱。
- ▶ 在「傾向機率」對話方塊中，按一下「設定」索引標籤。

圖表 12-3
「傾向機率」的「設定」索引標籤

購買傾向

欄位 **設定 (N)**

模式驗證

您可以驗證用來產生分數的模式。若要驗證模式，您必須將資料分成數個分割。訓練分割用於訓練（也就是建置）模式。測試分割用於驗證模式。如果您要驗證模式，此方法會自動將記錄指派給分割。

驗證模式(V)

訓練樣本分割大小(%) (T):

設定要複製結果的種子(S)

種子(E):

診斷輸出

整體模式品質(Q)

分類表(C)

最小機率(M):

重新編碼的回應欄位名稱與標記

此方法會自動將回應欄位重新編碼為新欄位，其中 1 代表正回應，0 代表負回應。

新欄位名稱(N):

新欄位標記(W):

儲存分數

此方法使用測試郵件或上一個活動的結果來產生分數。系統會自動儲存分數，讓您可以使用這些分數。此索引標籤上的其他控制項可額外控制所儲存的內容。

分數欄位名稱(S):

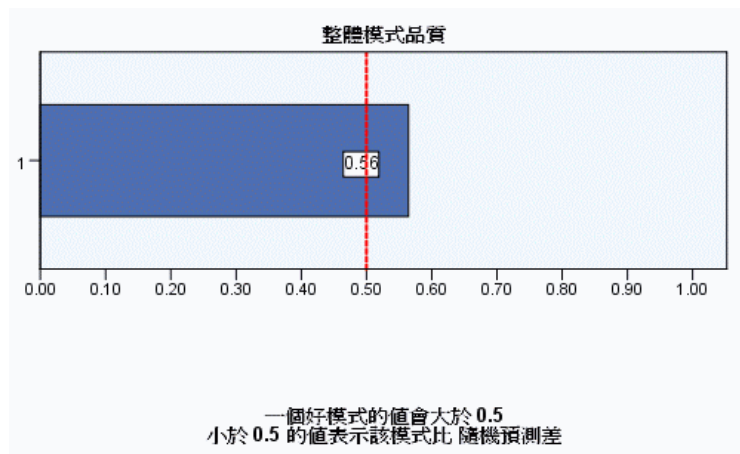
- ▶ 在「模式驗證群組」中，選取（勾選）「驗證模式」和「設定要複製結果的種子」。
- ▶ 使用預設為 50% 的訓練樣本分割大小，以及預設為 2000000 的種子值。
- ▶ 在「診斷輸出」群組中，選取（勾選）「整體模式品質」和「分類表」。
- ▶ 對於「最小機率」，輸入 0.05。以一般的規則而言，您應指定接近最小目標回應率的值（以比例表示）。0.05 的值代表 5% 的回應率。
- ▶ 按一下「**執行**」以執行程序並產生模式。

評估模式

「傾向機率」會產生整體模式品質圖表，而分類表則可用來評估模式。

整體模式品質圖表可快速以視覺方式呈現模式品質。以一般的規則而言，整體模式品質應在 0.5 以上。

圖表 12-4
整體模式品質圖表



如要確認模式是否足以進行評分，應同時檢驗分類表。

圖表 12-5
分類表

		分類表					
		預測次數					
		訓練樣本			測試樣本		
		已記錄回應 (1=Yes, 0=No)		百分比修正	已記錄回應 (1=Yes, 0=No)		百分比修正
否	是	否	是				
觀察次數							
已記錄回應 (1=Yes, 0=No)	否	651	249	72.33	653	267	70.98
	是	19	20	51.28	36	22	37.93
概要百分比		2.84	7.43	71.46	5.22	7.61	69.02

分類表會比較目標欄位的預測值與目標欄位的實際值。整體的準確率會指出模式良好運作的程度，但如果目標是建立建立模式，以識別可能會產生正回應率（等於或大於指定的最小正回應率）的連絡人群組，那麼您可能會對正確的預測正回應百分比更有興趣。

在這個範例中，分類表分為**訓練樣本**與**測試樣本**。訓練樣本會用來建立模式。接著將此模式套用至測試樣本，以瞭解模式良好運作的程度。

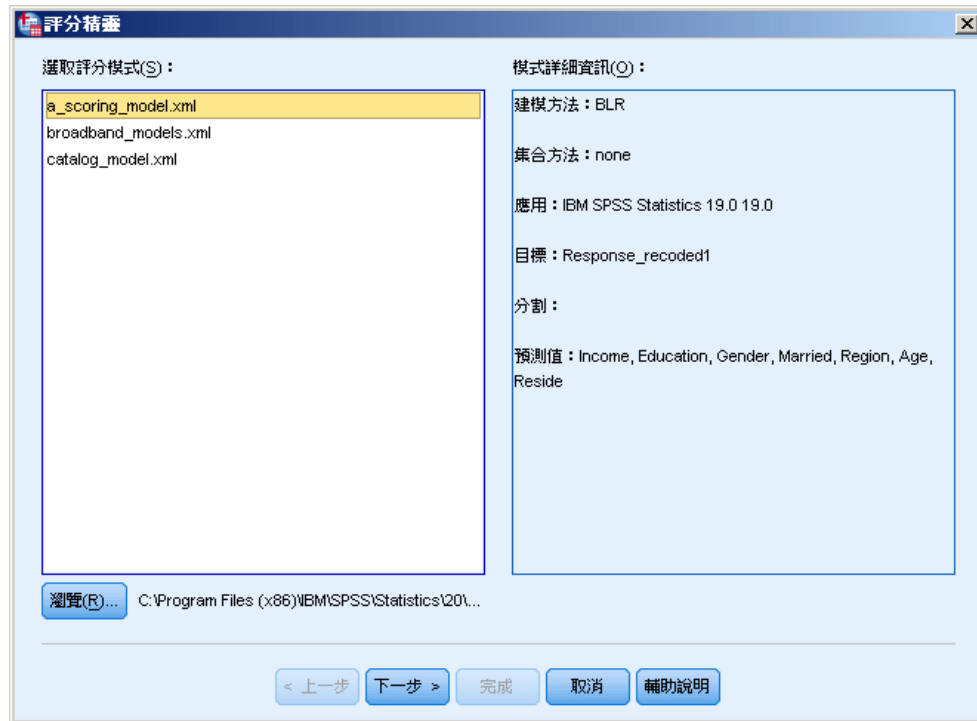
指定的最小回應率是 0.05 或 5%。分類表顯示，在訓練樣本和測試樣本中，正回應的正確分類率分別為 7.43% 和 7.61%。由於測試樣本回應率大於 5%，因此這個模式應該可以識別可能會產生大於 5% 回應率的連絡人群組。

套用模式

- ▶ 開啟資料檔 `dmdata3.sav`。這個資料檔包含不在測試郵件範圍內之所有連絡人的人口和其他資訊。如需詳細資訊，請參閱第 91 頁附錄 A 中的範例檔案。

- ▶ 開啟「評分精靈」。若要開啟「評分精靈」，請從功能表選擇：
公用程式(U) > 評分精靈

圖表 12-6
「評分精靈」，選取「評分模式」



- ▶ 按一下「瀏覽」，以瀏覽至您儲存模式 XML 檔案的位置，然後按一下「瀏覽」對話方塊中的「選取」。

「評分精靈」中會顯示所有副檔名為 .xml 或 .zip 的檔案。如果所選檔案被視為有效的模式檔案，則會顯示模式的說明。

- ▶ 選取您所建立的模式 XML 檔案，然後按「下一步」。

圖表 12-7
評分精靈，使模式欄位相符



為針對作用中的資料集進行評分，資料集必須包含與模式中所有預測值對應的欄位（變數）。如果模式也包含分割欄位，那麼資料集也必須包含與模式之所有分割欄對應的欄位。

- 根據預設，作用中資料集的欄位只要名稱和類型與模式之欄位相同，就會自動相符。
- 使用下拉式清單，比對資料集欄位與模式欄位。模式和資料集中每個欄位的資料類型必須相同，才能與欄位進行比對。
- 模式中所有的預測值（以及分割欄位，如果有的話）都必須與作用中資料集的欄位相符，才能繼續使用精靈，或是為作用中的資料集評分。

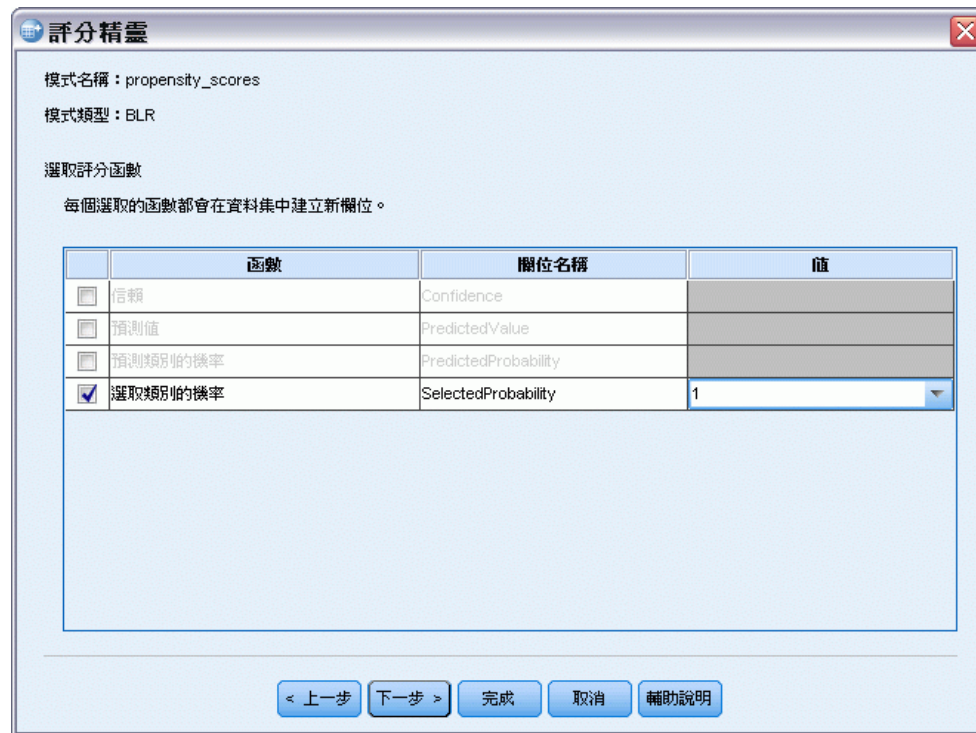
作用中的資料集不包含名為「收入」的欄位。因此最初在「資料集欄位」行中，與模式欄位「收入」對應的儲存格是空的。您必須在作用中的資料集中，選取與該模式欄位相等的欄位。

- ▶ 在「收入」模式欄位列的空白儲存格中，於「資料集欄位」行的下拉清單中選取「IncomeCategory」。

注意：除了欄位名稱和類型外，您應確定所評分之資料集中的實際資料值，都以和用來建立模式之資料集中的資料值相同的方式記錄下來。例如，如果模式是以區分為四個類別的收入欄位所建立，而作用中資料集中 IncomeCategory 的收入共區分為六個類別或四個不同的類別，則這些欄位不是真正彼此相符，所產生的分數也不可靠。

按一下「下一步」繼續到「評分精靈」的下個步驟。

圖表 12-8
評分精靈：選取評分函數



評分函數是可供所選模式使用的「分數」類型。可用的評分函數視模式而定。對於本範例中所使用的二元 logistic 模式，可用的函數為預測值、預測值機率、所選值的機率和信賴。

在這個範例中，我們感興趣的是對郵件正回應的預測機率；因此我們想要的是所選值的機率。

- ▶ 選取（核取）「選取類別的機率」。
- ▶ 在「值」行中，從下拉式清單中選取 1。目標可能值的清單是依據用來建立模式之資料檔中的目標值而定義於模式中。
- ▶ 取消選取（取消勾選）其他所有評分函數。
- ▶ 或者，您可以為將包含作用中資料集中分數值的新欄位指定更具描述性的名稱。例如，Probability_of_responding。
- ▶ 按一下「完成」，將模式套用至作用中的資料集。

內含正回應機率的新欄位會附加到資料集最後。

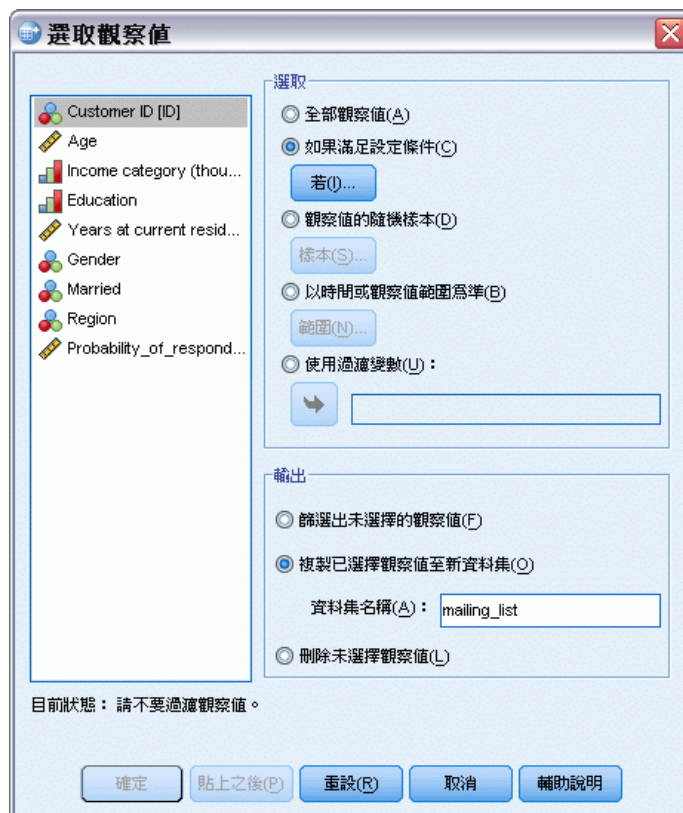
圖表 12-9
具有新的機率欄位的資料集

Reside	Gender	Married	Region	Probability_of_responding
7	1	0	4	.04
9	0	0	4	.03
12	0	0	4	.03
8	0	0	4	.04
13	0	0	4	.07
10	0	0	4	.04
12	0	0	4	.03
15	0	0	4	.05
10	0	0	4	.05
14	0	0	4	.02
5	0	0	4	.12

接著您就可以使用該欄位，來選取可能產生等於或高於特定水準正回應率的連絡人子集。例如，您可以建立新的資料集，其中包含可能產生至少 5% 的正回應率的個案子集。

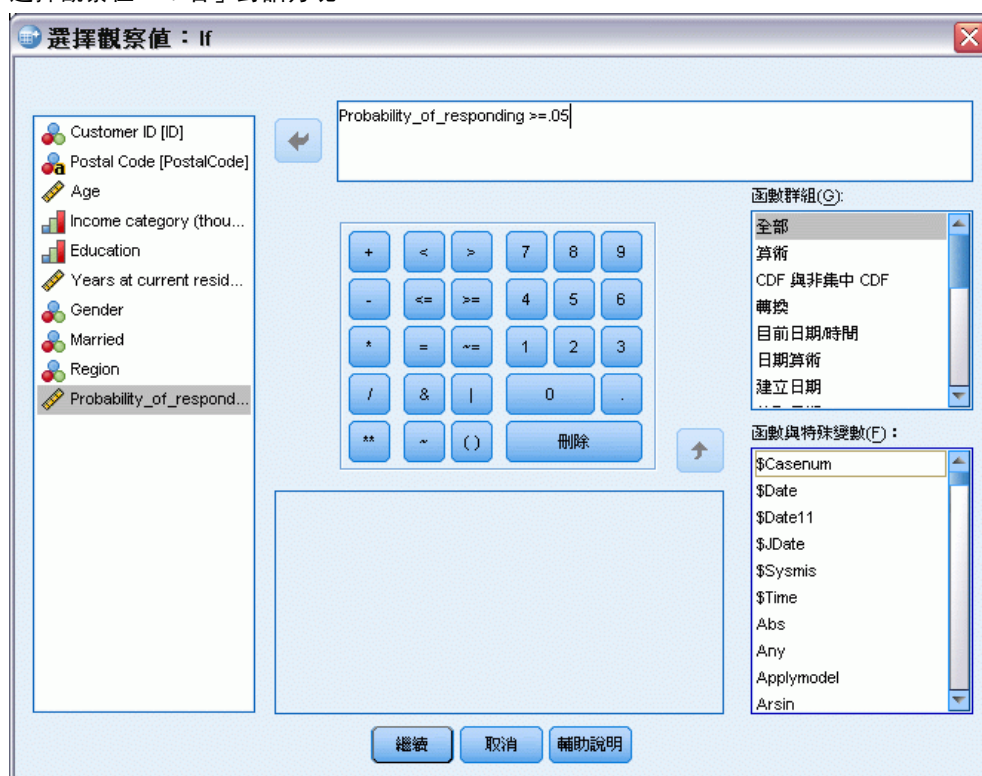
- ▶ 從功能表選擇：
資料 > 選擇觀察值

圖表 12-10
「選擇觀察值」對話方塊



- ▶ 在「選擇觀察值」對話方塊中，選取「如果滿足設定條件」，然後按一下「若」。

圖表 12-11
選擇觀察值：「若」對話方塊



- ▶ 在「選擇觀察值」中：如果對話方塊輸入下列運算式：

Probability_of_responding >=.05

注意： 如果您為包含機率值的欄位使用不同的名稱，請輸入該名稱以取代 Probability_of_responding。預設名稱是 SelectedProbability。

- ▶ 按一下「繼續」。
- ▶ 在「選擇觀察值」對話方塊中，選取「將所選觀察值複製到新的資料集」，並輸入新資料集的名稱。資料集名稱必須符合欄位（變數）命名規則。
- ▶ 按一下「確定」以建立包含所選連絡人的資料集。

新的資料集只包含正回應預測機率至少為 5% 的連絡人。

圖表 12-12
含所選聯絡人的新資料集

Reside	Gender	Married	Region	Probability_of_responding
13	0	0	4	.07
15	0	0	4	.05
10	0	0	4	.05
5	0	0	4	.12
7	0	0	4	.08
10	0	0	4	.10
15	1	0	4	.05
11	1	0	4	.08
9	1	0	4	.08
9	1	0	4	.05

摘要

「傾向機率」使用測試郵件或上一個活動的結果來產生傾向分數。此分數會根據各種所選的特性來識別哪些聯絡人最有可能回應。這項技術可建立預測模式，然後再套用至資料集以取得傾向分數。

控制套件檢定

此方法會比較行銷活動，瞭解不同包裝或優惠的效果是否有顯著性差異。活動效果是根據回應測量而得。

例如，公司的直效行銷決策部門想了解新套件的設計是否將產生較現有套件更正面的回應。因此他們送出一封測試郵件，以判斷新套件是否能產生明顯較高的正回應率。測試郵件包含一個接受現有套件的**控制群組**，以及一個接受新套件設計的**測試群組**。然後會將這兩個群組的結果做比較，以了解是否有明顯的差異。

此資訊收集於 `dmdata.sav` 中。如需詳細資訊，請參閱第 91 頁附錄 A 中的範例檔案。

執行分析

- ▶ 若要取得控制套件檢定，請從功能表選擇：
直效行銷 > 選擇技術
- ▶ 選取「比較活動的效果（控制套件檢定）」並按一下「繼續」。

圖表 13-1
「控制套件檢定」，「欄位」索引標籤

此方法會比較行銷活動，瞭解不同包裝或優惠是否有顯著性差異。「活動欄位」會識別不同活動。「效果欄位」會指示客戶對活動是否有回應。

欄位(F):
 排序: 無(N)

- Customer ID
- Responded to previous offer
- Postal Code
- Age
- Income category (thousands)
- Education
- Years at current residence
- Gender
- Married
- Region

活動(C):
Control Package

效果(E):
Responded to test offer

效果測量

交易金額(U)

正回應(P)

數值: []

Name and Label for Recoded Effectiveness Response Field

i 若效果是正回應，此技術會自動建立「是/否」效果欄位，以進行分析。

新欄位名稱(N): Effectiveness_new1

新欄位標記(L): 效果 (1=是 0=否)

執行(R) 貼上之後(P) 重設(R) 取消 輔助說明

- ▶ 對於「活動欄位」，選取「控制套件」。
- ▶ 對於「效果回應」欄位，選取「已回應測試優惠」。
- ▶ 選取「回覆」。
- ▶ 對於「正」回應值，請選取下拉式清單中的「是」。文字欄位中顯示的是數值 1，這是因為「是」實際上是與所記錄的數值 1 相關聯的數值標記（如果正回應值未定義數值標記，您可以直接在文字欄位中輸入該值）。

新的欄位便會自動建立，其中 1 代表正回應，0 代表負回應，並且會針對新欄位執行分析。您可以用自己的名稱與標記來覆寫預設的名稱與標記。對於此範例，我們會使用已經提供的欄位名稱。

- ▶ 按一下「執行」執执行程序。

輸出

圖表 13-2
「控制套件檢定」輸出

	Control Package			
	Control		Test	
	個數	行百分比	個數	行百分比
效果 (1=Yes 0=No) 0	875	96.2%	945	93.8%
1	35	3.8%	62	6.2%

Control 和 Test 之間存在統計顯著性差異。

程序輸出包含一個表格，其顯示「活動欄位」定義之每個群組的正和負回應的個數與百分比，以及另一個顯示群組回應率是否有明顯差異的表格。

- 「效果」是回應欄位重新編碼後的版本，其中 1 代表正回應，0 代表負回應。
- 控制套件的正確回應率是 3.8%，而檢定套件的正確回應率是 6.2%。

表格下方的簡單文字說明表示組別之間的差異明顯地不同，這代表較高的檢定套件回應率可能不是隨機的結果。此文字表格將包含分析中每個可能的成對組別之比較。因為此範例中只有兩個組別，因此只有 1 個比較。如果有超過五個以上的組別，文字說明表格將由「比較行比例」表格所取代。

摘要

「控制套件檢定」會比較行銷活動，瞭解不同包裝或優惠的效果是否有顯著性差異。在本範例中，檢定套件具有 6.2% 的正確回應率，明顯比控制套件的正確回應率 3.8% 要高。這表示您應該使用新的套件設計，而非舊的。但是可能有其他您需要考量的因素，例如任何與新套件設計有關的額外成本。

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。
- **anorectic.sav**。在將厭食/暴食行為症狀學標準化的過程中，研究人員研究了 55 個飲食失調的青少年。每個病患在四年之中被訪問四個回合，所以得到總數為 220 的觀察值。在每次觀察中，為病患在 16 種症狀上逐一評分。目前遺漏了第二次訪察的病患 71，第二次訪察的病患 76，以及第三次訪察的病患 47 的症狀分數，因此只剩下 217 個有效觀察值。
- **bankloan.sav**。這是有關一家銀行致力於減少放款利率預設值的假設資料檔。本檔包含 850 位以前的客戶與現在的準客戶的財務和人口資料。前 700 個觀察值為以前有借貸的客戶。最後 150 個觀察值是銀行需要作信用風險優良與不良分類的準客戶。
- **bankloan_binning.sav**。這是包含 500 位以前客戶的財務和人口資料的假設資料檔。
- **behavior.sav**。在典型範例中，52 名學生被要求為 15 種情境與 15 種行為組合評等，等級共分為 10 點，從 0 = 「非常適當」到 9 = 「非常不適當」。平均值超過個別值，值會被視為相異性。
- **behavior_ini.sav**。本資料檔包含 behavior.sav 之二維解的起始組態。
- **brakes.sav**。這是有關一間生產高性能汽車碟型煞車片工廠中品質管制的假設資料檔。資料檔包含由 8 個生產機器分別取得 16 個碟片的直徑測量。煞車的目標直徑是 322 公釐。

- **breakfast.sav.** 在經典研究中，21 名 Wharton 學院 MBA 學生及其配偶被要求為 15 項早餐食品按喜愛程度分出等級：從 1 = 「最喜愛」到 15 = 「最不喜愛」。他們的喜愛程度分六種不同情況記錄，從「整體喜愛」到「點心，僅配飲料」。
- **breakfast-overall.sav.** 本資料檔只包含第一種情況—「整體喜愛」—所喜愛的早餐項目。
- **broadband_1.sav.** 這是包含全國性寬頻服務地區用戶數目的假設資料檔。本資料檔包含四年期間 85 個地區每月的用戶數目。
- **broadband_2.sav.** 本資料檔與 broadband_1.sav 相同，但多了三個月的資料。
- **car_insurance_claims.sav.** 一個在別處 出現和分析過，有關汽車損害理賠的資料集。理賠金額的平均數可建立模式為具有 gamma 分配，使用反連結函數將依變數的平均數相關至一被保險人年齡、車輛類型、和車齡的線性組合。提出理賠的數量可以用作尺度權重。
- **car_sales.sav.** 本資料檔包含假設性的銷售估計、定價、和不同的品牌與車輛型式的實體規格。定價和實體規格是由 edmunds.com 和製造商處輪流取得。
- **car_sales_uprepared.sav.** 這是 car_sales.sav 的修改版本，其中不包含任何欄位的轉換版本。
- **carpet.sav.** 在一個普遍的範例 中，計劃銷售全新地毯清潔機的公司想要檢驗影響消費者偏好的五個因子—包裝設計、品牌名稱、價格、「優秀家用品」獎章及退費保證。包裝設計有三個因子水準，每個水準中的清潔刷位置都不相同；三個品牌名稱 (K2R、Glory、及 Bissell)；三個價格水準；且最後兩個因子各有兩個水準（無論無或有）。十名消費者將這些因子所定義的 22 種組合分級。「偏好」變數包含每個組合平均排名的等級。排名數值較小者會對應高偏好程度。這個變數反映每個組合偏好的整體量數。
- **carpet_prefs.sav.** 本資料檔是根據 carpet.sav 所描述的相同範例，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。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。
- **creditemo.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。
- **debate.sav**。這是有關一項政治辯論會參與者辯論前和辯論後接受調查之成對反應的假設資料檔。每一個觀察值對應至一位不同的應答者。
- **debate_aggregate.sav**。這是將 debate.sav 中之反應作整合的假設資料檔。每一個觀察值對應至辯論前和辯論後對偏好之交叉分類的反應。
- **demo.sav**。這是有關提供郵寄每月優惠之購買客戶資料庫的假設資料檔。記錄了客戶是否對該優惠回應，以及各種的人口資訊。
- **demo_cs_1.sav**。這是有關一家公司致力於匯編調查資訊資料庫之第一步的假設資料檔。每一個觀察值對應至一個不同的城市，也記錄了其地區、省、區、和城市識別。
- **demo_cs_2.sav**。這是有關一家公司致力於匯編調查資訊資料庫之第二步的假設資料檔。每一個觀察值對應至在第一步中選取的城市中的一個不同的家庭單位，也記錄了其地區、省、區、分區、和單位識別。也納入了由該設計的前兩階段所得之取樣資訊。
- **demo_cs.sav**。這是包含以複合取樣設計所收集之調查資訊的假設資料檔。每一個觀察值對應至一個不同的家庭單位，也記錄了各種的人口和取樣資訊。
- **dmdata.sav**。這是包含直效行銷公司之人口和購買資訊的假設資料檔。dmdata2.sav 包含收到測試郵件的連絡人子集資訊，而 dmdata3.sav 則包含剩下未收到測試郵件的連絡人資訊。
- **dietstudy.sav**。本假設資料檔包含對「Stillman 飲食法」研究的結果。每一個觀察值對應至一個不同的受試者，並記錄下他或她飲食法前、後之體重（磅）和三酸甘油酯水準（毫克/100 毫升）。
- **dvdplayer.sav**。這是有關新 DVD 播放器開發的假設資料檔。市場行銷團隊使用原型收集了焦點組別資料。每一個觀察值對應至不同調查到的使用者，並記錄下一些有關他們的人口資訊和他們對有關原型問題的回應。
- **german_credit.sav**。本資料檔取自 艾文 (Irvine) 在加州大學機器學習資料庫儲存器的「德國信用」資料集。
- **grocery_1month.sav**。本假設資料檔是將 grocery_coupons.sav 資料檔和每週購買的「彙總」，因此每一個觀察值對應至一個不同的客戶。結果部份每週變更的變數消失了，而目前所記錄的銷售量是在研究的四週期間銷售量之總和。

- **grocery_coupons.sav**。這是包含某連鎖雜貨店想要知道他們客戶購買習慣所收集之調查資料的假設資料檔。每一個客戶被追蹤了四週，每一個觀察值對應至一個不同的客戶一週，並記錄有關客戶在何處及如何購物的資訊，包含那一週在雜貨店花了多少錢。
- **guttman.sav**。Bell 以此表說明可能的社會團體。Guttman 過去曾使用此表的一部分，在這部分中有 5 個變數，分別說明 7 個理論社會團體的社會互動、團體歸屬感、成員實際接觸和關係正式性，而這 7 個群組包括：群眾（例如，足球場上的人）、觀眾（例如在戲院中和課堂上的人）、公眾（例如，報紙讀者和電視觀眾）、暴民（和群眾相似，但互動較為激烈）、原級團體（親密性）、次級團體（自願性）和現代社群（因親密的身體接近而導致鬆散的結盟和特殊服務的需求）。
- **health_funding.sav**。這是包含醫療保健基金（每 100 個人口的金額）、疾病率（每 10,000 個人口的比率）、造訪醫療保健機構的比例（每 10,000 個人口的比率）的假設資料檔。每一個觀察值代表一個不同的城市。
- **hivassay.sav**。這是有關一家製藥實驗室致力於開發一種偵測 HIV 感染快速檢驗的假設資料檔。檢驗結果是八個紅色加深的陰影，陰影愈深表示感染的可能性愈大。進行了一項實驗室的試驗，在 2,000 個血液樣本中，有半數遭到 HIV 的感染，而半數則未感染。
- **hourlywagedata.sav**。這是有關在辦公室和醫院任職的護士依經驗水準不同之鐘點費的假設資料檔。
- **insurance_claims.sav**。這是有關一家保險公司想要建立模式來標示可疑及可能的詐欺理賠之假設資料檔。每一個觀察值代表個不同的理賠。
- **insure.sav**。這是有關一家保險公司正在研究表示客戶是否必定理賠 10 年壽險合約之風險因子的假設資料檔。在資料檔中的每一個觀察值代表二份合約，其一記錄了理賠而另一則否，二者的年齡和性別相符。
- **judges.sav**。這是有關受過訓練的裁判（加上一位熱心人士）為 300 個體操表演評分的假設資料檔。每一列代表一個不同的表演；裁判們觀看相同的表演。
- **kinship_dat.sav**。Rosenberg 與 Kim 致力於分析 15 個親屬關係稱呼（姑/姨、兄弟、堂/表兄弟姐妹、女兒、父親、孫女、祖父、祖母、孫子、母親、姪子/外甥、姪女/外甥女、姐妹、兒子、叔/舅父）。他們請四組大學生（兩組女性、兩組男性）根據其相似性來分類整理這些稱謂。他們請其中兩組（一組女性、一組男性）作兩次分類整理，第二次要根據與第一次不同的準則進行分類整理。因此，總共得到六個「來源」。每一個來源對應至一個 15×15 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，、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。
- **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_csplan 中所指定的取樣計劃之樣本部門來選取；由於取樣計劃採用到機率 - 比例 - 大小 (PPS) 方法，也用到一個包含聯合選擇機率的檔案 (recidivism_cs_jointprob.sav)。
- **rfm_transactions.sav**。本檔案是包含購買交易資料的假設資料檔，包括購買日期、購買項目及每一項交易的金額。

- **salesperformance.sav**。這是有關評估兩個新售貨員訓練課程的假設資料檔。六十個員工，分成三個組別，全部接受標準訓練。此外，組別二得到技術訓練；組別三則是實務輔導簡介。每一個員工在訓練課程結束時接受測驗並記錄他們的分數。在資料檔中每一個觀察值代表一個不同的訓員，並記錄他們所分派的組別和他們在測驗中得到的分數。
- **satisf.sav**。這是有關一家零售公司在 4 個商店位置所作之滿意度調查的假設資料檔。總共有 582 位客戶接受調查，每一個觀察值代表一位客戶的反應。
- **screws.sav**。這個資料檔包含螺絲釘、螺栓、螺帽和圖釘之特色的資訊。
- **shampoo_ph.sav**。這是有關一家美髮產品工廠品質管制過程的假設資料檔。在固定的時間間隔，記錄下六個不同輸出批次的測量和它們的 pH 值。目標範圍是 4.5 - 5.5。
- **ships.sav**。一個在別處出現和分析過，有關商船因風浪所造成損壞的資料集。事件次數可建立模式為以 Poisson 率發生，給定船型、建造期間、和服務期間。以因子交叉分類所形成的表格的每一個儲存格服務月數的整合，提供了暴露於風險之值。
- **site.sav**。這是有關一家公司致力於為事業擴展選擇新地點的假設資料檔。他們僱請兩位顧問分別評估該地點，除了一份廣泛的報告之外，他們還要將每個地點摘要為前景「佳」、「可」、或「差」。
- **smokers.sav**。本資料檔是由「1998 年全國家庭毒品濫用調查」中摘錄，且是美國家庭的機率樣本。<http://dx.doi.org/10.3886/ICPSR02934> 因此，在分析本資料檔的第一步應該是將資料加權以反映母群體傾向。
- **stocks.sav** 本假設資料檔包含一年的股票價格和數量。
- **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，並記錄該新項目目前四週的每週銷售量。每一個觀察值對應至一個不同的地點-週。

- **testmarket_1month.sav**。本假設資料檔是將 testmarket.sav 資料檔和每週購買的「彙總」，因此每一個觀察值對應至一個不同的客戶。結果部份每週變更的變數消失了，而目前所記錄的銷售量是在研究的四週期間銷售量之總和。
- **tree_car.sav**。這是包含人口資料和車輛購買價格資料的假設資料檔。
- **tree_credit.sav**。這是包含人口資料和銀行放款歷史資料的假設資料檔。
- **tree_missing_data.sav** 這是包含有大量遺漏值的人口資料和銀行放款歷史資料的假設資料檔。
- **tree_score_car.sav**。這是包含人口資料和車輛購買價格資料的假設資料檔。
- **tree_textdata.sav**。一個只有兩個變數的簡單資料檔，主要目的在顯示變數預設狀態（在指定量測水準和數值標記之前）。
- **tv-survey.sav**。這是有關一家電視製片廠考量是否要延長一個成功節目的播送所作之調查的假設資料檔。有 906 位應答者被問到在不同的狀況下他們是否願意觀看這個節目。每一列代表一個不同的應答者；每一行為一個不同的狀況。
- **ulcer_recurrence.sav**。本檔案包含一項用來比較兩種防止潰瘍復發治療法功效之研究的部分資訊。它是很好的區間受限資料範例，且已在別處 出現和分析過。
- **ulcer_recurrence_recoded.sav**。本檔案是將 ulcer_recurrence.sav 的資訊重新組織，以讓您為此研究的每一個區間事件機率而非只是研究目的事件機率建立模式。它已在別處 出現和分析過。
- **verd1985.sav**。本資料檔有關於一項調查。在調查中記錄了來自 15 個受訪者對 8 個變數的回應。所需的變數被分成三組。集 1 包括 age 和 marital，集 2 包括 pet 和 news，集 3 包括 music 和 live。Pet 調整為多重名義量數，age 調整為次序量數，其他的變數調整為單一名義量數。
- **virus.sav**。這是有關一家網際網路服務提供者致力於在其網路上判斷病毒之影響的假設資料檔。他們在其網路上追蹤從發現病毒直到控制威脅的這段時間，被病毒感染之電子郵件的流量（約略）百分比。
- **wheeze_steubenville.sav**。這是空氣污染對兒童健康之影響 縱向研究的子集。本資料包含來自俄亥俄州 Steubenville，年齡 7、8、9 和 10 歲兒童的氣喘聲狀態之重複二元測量，以及其母親在本研究的第一年是否抽煙的固定記錄。
- **workprog.sav**。這是有關一項政府職業計劃，設法將弱勢民眾安置到較好之工作的假設資料檔。一個樣本的可能計劃參與者被追蹤，他們之中某些被選取加入本計劃，而其他的則否。每一個觀察值代表一位不同的計劃參與者。
- **worldsales.sav** 本假設資料檔包含依洲和產品分類之銷貨收益。

注意事項

本資訊適用於全球提供之產品與服務。

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, Armonk, NY 10504-1785, U. S. A.

對於與雙位元組字元集 (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., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

下列條款不適用於英國，或其他任何當地法律規定與下列條款不一致的國家：

INTERNATIONAL BUSINESS MACHINES 聲明係以「現狀」提供，沒有任何保固；不作任何明示或默示的保證，包括但不限於不侵權、適銷性或適合某一特定用途之保證。某些州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。

此資訊中對於非 IBM 網站之任何參考僅為查閱方便而設，而且在任何情況中均不得作為那些網站之背書。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 Software Group, 收件人：Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA。

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等效力合約中提供。

有關非 IBM 產品的資訊來自於那些產品的供應商、其公佈內容或其他可公開取得的來源。IBM 尚未測試那些產品，且無法確認效能準確度、相容性或任何其他與非 IBM 產品相關的理賠。對於非 IBM 產品之功能有任何問題，應由那些產品之供應商加以解決。

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。所有的名稱皆為虛構，使用的名稱或地址和實際的企業如有雷同純屬巧合。

如果您正在螢幕上檢視此資訊，則圖片和彩色說明可能不會顯示。

商標

IBM、IBM 標誌、ibm.com 和 SPSS 為 IBM Corporation 之註冊商標，已經於世界各地許多法律管轄區域註冊。IBM 註冊商標的清單目前可於「網站」上取得，網址為：<http://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。

Java 和所有 Java 商標和標誌皆為 Sun Microsystems, Inc. 於美國和 / 或其他國家的商標。

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。

本產品使用 WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>。

其他產品和服務名稱可能為 IBM 或其他公司的商標。

Adobe 產品的擷取畫面已取得 Adobe Systems Incorporated 之翻印許可。

Microsoft 產品的擷取畫面已取得 Microsoft Corporation 之翻印許可。



索引

logistic 迴歸 , 30
Logistic 迴歸, 76
RFM, 2, 8-9, 11, 41
 binning, 6
 交易資料, 3, 41
 客戶資料, 4

傾向機率, 30, 76

商標, 99

控制套件檢定, 37, 88

法律注意事項, 98
潛在客戶分析概要, 17, 64

範例檔案
 位置, 91

郵遞區號回應率, 23, 69

集群, 13
集群分析 , 13, 48