

***IBM SPSS Modeler 18.1.1* 來源、程序和輸出節點**

**IBM**

**附註**

使用本資訊及其所支援的產品之前，請先閱讀第 345 頁的『注意事項』中的資訊。

**產品資訊**

此版本適用於 IBM SPSS Modeler 18.1.1 版以及所有後續版次和修訂，直到新版中另有指示。

# 目錄

前言 . . . . .	vii	匯入 Cognos 報告 . . . . .	34
第 1 章 關於 IBM SPSS Modeler . . . . .	1	Cognos 連線 . . . . .	34
IBM SPSS Modeler 產品 . . . . .	1	Cognos 位置選擇 . . . . .	35
IBM SPSS Modeler . . . . .	1	指定資料或報表的參數 . . . . .	35
IBM SPSS Modeler Server . . . . .	1	IBM Cognos TM1 來源節點 . . . . .	35
IBM SPSS Modeler Administration Console . . . . .	2	匯入 IBM Cognos TM1 資料 . . . . .	36
IBM SPSS Modeler Batch . . . . .	2	TWC 來源節點 . . . . .	36
IBM SPSS Modeler Solution Publisher . . . . .	2	SAS 來源節點 . . . . .	37
IBM SPSS Collaboration and Deployment Services 的 IBM SPSS Modeler Server 配接器 . . . . .	2	為 SAS 來源節點設定選項 . . . . .	37
IBM SPSS Modeler 版本 . . . . .	2	Excel 來源節點 . . . . .	38
說明文件 . . . . .	3	XML 來源節點 . . . . .	39
SPSS Modeler Professional 文件 . . . . .	3	從多個根元素中選取 . . . . .	40
SPSS Modeler Premium 文件 . . . . .	3	從 XML 來源資料中移除不需要的空格 . . . . .	40
應用程式範例 . . . . .	4	使用者輸入節點 . . . . .	40
Demos 資料夾 . . . . .	4	為使用者輸入節點設定選項 . . . . .	41
授權追蹤 . . . . .	4	「模擬產生」節點 . . . . .	45
為「模擬產生」節點設定選項 . . . . .	46		
第 2 章 來源節點 . . . . .	5	複製欄位 . . . . .	50
概述 . . . . .	5	配適詳細資料 . . . . .	50
設定欄位儲存類型和格式 . . . . .	7	指定參數 . . . . .	51
清單儲存以及相關聯的測量層次 . . . . .	9	分佈 . . . . .	53
不受支援的控制字元 . . . . .	10	延伸匯入節點 . . . . .	55
Analytic Server 來源節點 . . . . .	11	延伸匯入節點 - 語法標籤 . . . . .	55
選取資料來源 . . . . .	11	延伸匯入節點 - 主控台輸出標籤 . . . . .	55
修正認證 . . . . .	11	過濾或重新命名欄位 . . . . .	55
受支援的節點 . . . . .	11	資料視圖節點 . . . . .	56
資料庫來源節點 . . . . .	15	為「資料視圖」節點設定選項 . . . . .	56
設定資料庫節點選項 . . . . .	16	「地理空間」來源節點 . . . . .	57
新增資料庫連線 . . . . .	16	設定「地理空間」來源節點的選項 . . . . .	57
潛在資料庫問題 . . . . .	18	一般來源節點標籤 . . . . .	58
為資料庫連線指定預設值 . . . . .	18	在來源節點中設定測量層次 . . . . .	58
選取資料庫表格 . . . . .	21	從來源節點中過濾欄位 . . . . .	59
查詢資料庫 . . . . .	21	第 3 章 記錄作業節點 . . . . .	61
「變數檔案」節點 . . . . .	23	記錄作業概觀 . . . . .	61
設定「變數檔案」節點的選項 . . . . .	23	選取節點(S) . . . . .	63
將地理空間資料匯入到「變數檔案」節點中 . . . . .	25	樣本節點 . . . . .	63
固定檔案節點 . . . . .	26	樣本節點選項 . . . . .	64
為固定檔案節點設定選項 . . . . .	26	叢集與分層設定 . . . . .	65
Statistics 檔案節點 . . . . .	27	Strata 的樣本大小 . . . . .	66
Data Collection 節點 . . . . .	28	平衡節點 . . . . .	67
Data Collection 匯入檔案選項 . . . . .	28	為平衡節點設定選項 . . . . .	67
Data Collection 匯入 meta 資料內容 . . . . .	30	「聚合」節點 . . . . .	68
資料庫連線字串 . . . . .	31	「聚合」節點的設定選項 . . . . .	68
進階特性 . . . . .	31	聚合最佳化設定 . . . . .	70
匯入複選題分析集 . . . . .	31	RFM 「聚合」節點 . . . . .	70
Data Collection 欄匯入說明 . . . . .	31	為 RFM 「聚合」節點設定選項 . . . . .	71
IBM Cognos 來源節點 . . . . .	32	排序節點 . . . . .	72
Cognos 物件圖示 . . . . .	32	排序最佳化設定 . . . . .	72
匯入 Cognos 資料 . . . . .	33	合併節點 . . . . .	72
		結合類型 . . . . .	73

指定合併方法和鍵值	74	衍生節點	131
選取用於局部結合的資料	75	為衍生節點設定基本選項	133
指定合併的條件	75	衍生多個欄位	133
為「合併」指定排名式條件	76	設定衍生公式選項	134
過濾合併節點中的欄位	77	設定衍生旗標選項	136
設定輸入順序和標籤	78	設定衍生名義選項	136
合併最佳化設定	78	設定衍生狀態選項	136
附加節點	79	設定衍生計數選項	137
設定附加選項	79	設定衍生條件選項	137
差異節點	80	使用衍生節點對值進行重新編碼	137
相異的最佳化設定	81	填充器節點	137
相異的複合設定	82	使用填入節點進行儲存類型轉換	138
多媒體串流時間序列節點	83	再分類節點(C)	139
「串流式時間序列」節點 - 欄位選項	84	為再分類節點設定選項	139
「串流式時間序列」節點 - 資料規格選項	84	對多個欄位進行再分類	140
「串流式時間序列」節點 - 建置選項	87	再分類欄位的儲存類別型和測量層次	140
「串流式時間序列」節點 - 模型選項	90	匿名化節點	141
SMOTE 節點	91	匿名化節點的設定選項	141
SMOTE 節點設定	92	對欄位值進行匿名化	142
延伸轉換節點	93	分組節點(B)	142
延伸轉換節點 - 語法標籤	93	為 bin 節點設定選項	143
延伸轉換節點 - 主控台輸出標籤	93	固定寬度 Bin	143
「空間時間限制」節點	94	分位顯示 (相等計數或總和)	144
定義空間時間限制密度	96	等級觀察值	145
串流式 TCM 節點	96	均數/標準差	146
串流式 TCM 節點 - 「時間序列」選項	96	最適 Binning	146
串流式 TCM 節點 - 「觀察值」選項	97	預覽產生的 Bin	147
串流式 TCM 節點 - 「時間間隔」選項	98	RFM 分析節點	147
串流式 TCM 節點 - 「聚集和分佈」選項	98	RFM 分析節點設定	148
串流式 TCM 節點 - 「遺漏值」選項	98	RFM 分析節點分級	149
串流式 TCM 節點 - 「一般資料」選項	99	總體節點	149
串流式 TCM 節點 - 「一般建立」選項	99	總體節點設定	149
串流式 TCM 節點 - 「估計期間」選項	100	分割區節點	150
串流式 TCM 節點 - 「模型」選項	100	分割區節點選項	151
CPLEX 最佳化節點	100	設為旗標節點	152
CPLEX Optimization 節點的設定選項	101	為設為旗標節點設定選項	152
<b>第 4 章 欄位作業節點</b>	<b>103</b>	重新結構化節點	152
欄位作業概觀	103	為重新結構化節點設定選項	153
自動資料準備	105	轉置節點	154
欄位標籤	106	為「轉置」節點設定選項	154
設定標籤	107	歷史節點	155
分析標籤	110	為歷史節點設定選項	155
產生一個衍生節點	116	欄位重新排序節點	156
類型節點	117	設定欄位欄位重新排序選項	156
測量層次	118	時間間隔節點	157
轉換連續資料	121	時間間隔 - 欄位選項	158
何謂實例化?	122	時間間隔 - 建置選項	158
資料值	122	「重新投影」節點	159
定義遺漏值	126	為「重新投影」節點設定選項	159
勾選類型值	126	<b>第 5 章 「圖表」節點</b>	<b>161</b>
設定欄位角色	127	通用圖形節點功能	161
複製類型屬性	127	外觀、套版、畫面和動畫	162
欄位格式設定標籤	128	使用「輸出」標籤	163
過濾或重新命名欄位	129	使用「註解」標籤	164
設定過濾選項	129	三維圖形	164



圖形板節點	166	t-SNE 節點輸出選項	233
圖板基本標籤	166	存取及繪製 t-SNE 資料	233
圖板詳細資訊標籤	169	t-SNE 模型塊	235
可用的內建圖板視覺化類型	170	E-Plot (測試版) 節點	235
建立地圖視覺化	175	E-Plot (測試版) 節點繪圖標籤	235
圖板範例	176	E-Plot (測試版) 節點選項標籤	235
圖版「外觀」標籤	186	E-Plot (測試版) 外觀標籤	236
設定範本、樣式表與地圖的位置	187	使用 E-Plot 圖形	236
管理範本、樣式表與地圖檔案	188	探索圖形	239
轉換及發佈地圖形狀檔案	189	使用級區	240
地圖的關鍵概念	189	使用區域	243
使用地圖轉換公用程式	190	使用標示元素	245
發佈地圖檔案	194	從圖形產生節點	246
統計圖節點	194	編輯視覺化	248
統計圖節點標籤	197	編輯視覺化的一般相關規則	249
統計圖「選項」標籤	198	編輯和設定文字格式	249
統計圖「外觀」標籤	199	變更顏色、樣式、線條粗細和透明度	250
使用統計圖	199	旋轉並變更點元素的形狀和圖面比	251
多重繪圖節點	200	變更圖形元素的大小	251
多重繪圖標籤	200	指定邊緣和填塞大小	251
多重繪圖「外觀」標籤	202	設定數字格式	252
使用多重繪圖	202	變更軸及尺度設定	252
時間繪圖節點	203	編輯類別	254
時間繪圖標籤	203	變更窗格方向	255
時間繪圖「外觀」標籤	204	轉換座標系統	255
使用時間繪圖圖形	204	變更統計量和圖形元素	256
分佈節點	205	變更圖註位置	257
分佈圖形標籤	205	複製視覺化及視覺化資料	257
分佈「外觀」標籤	205	圖表板編輯器鍵盤快速鍵	257
使用分佈節點	206	新增標題和註腳	258
「直方圖」節點	208	使用圖形樣式表	258
直方圖「繪圖形」標籤	208	列印、儲存、複製及匯出圖形	259
直方圖「選項」標籤	208		
直方圖「外觀」標籤	208		
使用直方圖	209		
收集節點	209	<b>第 6 章 輸出節點</b>	<b>263</b>
收集統計圖標籤	210	輸出節點概觀	263
收集選項標籤	210	管理輸出	264
收集「外觀」標籤	210	檢視輸出	265
使用收集圖	211	發行至網路	265
網絡節點	212	在 HTML 瀏覽器中檢視輸出	266
網絡統計圖標籤	213	匯出輸出	266
Web 選項標籤	214	選取單元和欄	267
Web「外觀」標籤	215	表格節點	267
使用網路圖	216	表格節點的「設定」標籤	267
評估節點	220	表格節點格式標籤	267
評估統計圖標籤	223	輸出節點的「輸出」標籤	267
「評估選項」標籤	224	表格瀏覽器	269
評估「外觀」標籤	225	矩陣節點	269
讀取模型評估結果	225	矩陣節點的設定標籤	269
使用評估表	226	矩陣節點的外觀標籤	270
「地圖視覺化」節點	227	矩陣節點輸出瀏覽器	271
地圖視覺化「繪圖形」標籤	227	「分析」節點	271
地圖視覺化「外觀」標籤	231	「分析」節點的「分析」標籤	272
t-SNE 節點	231	分析輸出瀏覽器	273
t-SNE 節點專家選項	231	資料審核節點	274
		資料審核節點的設定標籤	274
		資料審核的品質標籤	275

資料審核輸出瀏覽器	275
轉換節點	280
轉換節點的選項標籤	280
轉換節點的輸出標籤	280
轉換節點的輸出檢視器	280
統計資料節點	282
統計資料節點的設定標籤	282
統計資料輸出瀏覽器	283
「平均數」節點	284
比較獨立群組的平均數	284
在成對欄位之間比較平均數	284
「平均數」節點選項	285
「平均數」節點輸出瀏覽器	285
報告節點	286
報告節點的範本標籤	287
報告節點輸出瀏覽器	288
設定廣域節點	288
設定廣域節點的設定標籤	288
「模擬配適」節點	288
分佈配適	289
「模擬配適」節點的「設定」標籤	290
「模擬評估」節點	291
「模擬評估」節點的「設定」標籤	291
「模擬評估」節點輸出	293
延伸輸出節點	296
延伸輸出節點 - 語法標籤	296
延伸輸出節點 - 主控台輸出標籤	297
延伸輸出節點 - 輸出標籤	297
延伸輸出瀏覽器	298
IBM SPSS Statistics 說明應用程式	298
<b>第 7 章 匯出節點</b>	<b>301</b>
匯出節點概觀	301
資料庫匯出節點	302
資料庫節點的「匯出」標籤	302
資料庫匯出合併選項	303
資料庫匯出綱目選項	304
資料庫匯出指標選項	306
資料庫匯出進階選項	307
大量載入器程式設計	308
純文字檔案匯出節點	314
「純文字檔案匯出」標籤	314
Statistics 匯出節點	315
Statistics 匯出節點 - 「匯出」標籤	315
重新命名或過濾 IBM SPSS Statistics 的欄位	316
Data Collection 匯出節點	316
Analytic Server 匯出節點	317
IBM Cognos 匯出節點	317
Cognos 連線	318
ODBC 連線	318
IBM Cognos TM1 匯出節點	319
連接至 IBM Cognos TM1 多維度報表來匯出資料	320
對映要匯出的 IBM Cognos TM1 資料	320
SAS 匯出節點	321

SAS 匯出節點「匯出」標籤	321
Excel 匯出節點	321
Excel 節點「匯出」標籤	321
延伸匯出節點	322
延伸匯出節點 - 語法標籤	322
延伸匯出節點 - 主控台輸出標籤	323
XML 匯出節點	323
寫入 XML 資料	323
XML 對映記錄選項	324
XML 對映欄位選項	324
XML 對映預覽	324
「一般匯出」節點標籤	324
發佈串流	325

## 第 8 章 IBM SPSS Statistics 個節點 327

IBM SPSS Statistics 節點 - 概觀	327
Statistics 檔案節點	328
Statistics 轉換節點	329
Statistics 轉換節點 - 「語法」標籤	329
容許的語法	329
Statistics 模型節點	331
Statistics 模型節點 - 「模型」標籤	331
Statistics 模型節點 - 模型塊彙總	332
Statistics 輸出節點	332
Statistics 輸出節點 - 「語法」標籤	332
Statistics 輸出節點 - 「輸出」標籤	334
Statistics 匯出節點	334
Statistics 匯出節點 - 「匯出」標籤	335
重新命名或過濾 IBM SPSS Statistics 的欄位	335

## 第 9 章 SuperNode 337

SuperNode 概觀	337
SuperNode 的類型	337
來源 SuperNode	337
處理 SuperNode	337
終端 SuperNode	338
建立 SuperNode	338
巢狀 SuperNode	339
鎖定 SuperNode	339
鎖定和解除鎖定 SuperNode	339
編輯鎖定的 SuperNode	340
編輯 SuperNode	340
修改 SuperNode 類型	340
添加註解和更名 SuperNode	340
SuperNode 參數	341
SuperNode 和快取	342
SuperNode 和 Script 編寫	343
儲存和載入 SuperNode	343

## 注意事項 345

商標	346
產品說明文件的條款	346

## 索引 351

---

## 前言

IBM® SPSS® Modeler 是 IBM Corp. 企業強度的資料採礦工作平台。SPSS Modeler 可以透過深入瞭解資料來協助組織改進客戶和居民關係。組織使用從 SPSS Modeler 取得的見解來保留贏利客戶、識別交叉銷售機會、吸引新客戶、偵測詐騙、減少風險，以及改進政府服務交付。

SPSS Modeler 的視覺化介面會邀請使用者套用其特定的商業專門知識，從而獲得功能更強大的預測模型並縮短解決方案時間。SPSS Modeler 提供了許多建模技術，例如預測、分類、客群劃分，以及關聯偵測演算法。建立模型之後，IBM SPSS Modeler Solution Publisher 就可以從企業層面交付給決策制定者或資料庫。

## 關於 IBM Business Analytics

IBM Business Analytics 軟體提供完整、一致且準確的資訊，決策者可信任此資訊，並藉以改善營運績效。商業智慧、預測分析、財務績效及策略管理，以及分析應用程式的綜合性產品組合會對現行績效提供清晰、即時而可行的洞察，且能夠預測未來結果。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。

作為此產品組合的一部分，IBM SPSS Predictive Analytics 軟體可協助組織預測未來事件，並根據促進較佳業務結果的洞察，主動採取行動。全球的商業、政府及學術客戶都依賴於 IBM SPSS 技術並將其作為吸引、保留及擴充客戶且同時減少詐騙及減輕風險的競爭優勢。透過將 IBM SPSS 軟體引入其每天的作業，組織成為具有預測能力的企業，能夠直接或自動進行決策，以符合業務目標，並達成可測量的競爭優勢。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

技術支援可用於維護客戶。客戶可能會聯絡技術支援，以取得使用 IBM Corp. 產品的協助，或其中一個受支援硬體環境的安裝協助。若要呼叫技術支援，請參閱 IBM Corp. 網站，網址為 <http://www.ibm.com/support>。要求協助時，請準備好識別您自己、您的組織及您的支援合約。



---

## 第 1 章 關於 IBM SPSS Modeler

IBM SPSS Modeler 是一組資料採礦工具，通過這些工具可以採用商業技術快速建立預測性模型，並將其應用於商業活動，從而改進決策過程。IBM SPSS Modeler 參照行業標準 CRISP-DM 模型設計而成，可支援從資料到更優商業成果的整個資料採礦過程。

IBM SPSS Modeler 提供擷取自機器學習人工智慧以及統計資料的各種建模方法。「建模」選用區上提供的方法可讓您根據資料衍生新資訊，以及開發預測模型。每種方法都具有特定的強度且最適合因應特定類型的問題。

SPSS Modeler 可以作為單獨產品購買，也可以作為用戶端與 SPSS Modeler Server 一起使用。同時提供了大量其他選項，下列各節將對這些選項進行概述。有關進一步資訊，請參閱<https://www.ibm.com/analytics/us/en/technology/spss/>。

---

### IBM SPSS Modeler 產品

IBM SPSS Modeler 系列產品及關聯的軟體包括下列各項。

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console (包含在 IBM SPSS Deployment Manager 中)
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- IBM SPSS Collaboration and Deployment Services 的 IBM SPSS Modeler Server 配接器

### IBM SPSS Modeler

SPSS Modeler 是具有完整功能的產品，它安裝並執行於個人電腦上。您可以在本端方式作為單獨產品執行 SPSS Modeler，也可以在分佈方式下將其與 IBM SPSS Modeler Server 一起使用來提高大型資料集的效能。

借助 SPSS Modeler，您可以快速直接地建立準確的預測模型，而不進行程式設計。通過使用唯一可視介面，您可以輕鬆地查看資料採礦過程。借助該產品隨附的進階分析支援，您可以探索資料中先前隱藏的型樣和趨勢。您可以構建結果模型並瞭解影響結果的因素，從而利用業務機會並降低風險。

SPSS Modeler 推出了兩個版本：SPSS Modeler Professional 和 SPSS Modeler Premium。請參閱第 2 頁的『IBM SPSS Modeler 版本』主題，以取得更多資訊。

### IBM SPSS Modeler Server

SPSS Modeler 使用用戶端/伺服器架構將資源集約型作業的要求分發給功能強大的伺服器軟體，因而使大資料集的傳輸速度大大加快。

SPSS Modeler Server 是一個個別授權的產品，在分佈分析方式下，該產品在安裝了一個或多個 IBM SPSS Modeler 的伺服器主機上持續執行。這種運行方式大大提高了 SPSS Modeler Server 對大型資料集的處理速度，因為在伺服器上可以運行耗用記憶體體的作業，並且無需將資料下載到用戶端電腦上。IBM SPSS Modeler Server 還提供對 SQL 最佳化和資料庫內建模功能的支援，從而在效能和自動化方面帶來更多優勢。

## IBM SPSS Modeler Administration Console

Modeler Administration Console 是一個圖表使用者介面，用於管理多個 SPSS Modeler Server 配置選項，這些選項還可以通過選項檔案進行配置。主控台包含在 IBM SPSS Deployment Manager，可以用於監視和配置 SPSS Modeler Server 安裝，並且可供目前 SPSS Modeler Server 客戶免費使用。應用程式僅可以在 Windows 電腦上安裝；但它可以管理在任何受支援平台上安裝的伺服器。

## IBM SPSS Modeler Batch

資料採礦通常是交互過程，因此，還可以從指令行執行 SPSS Modeler 而不需要圖形使用者介面。例如，您可能具有長時間執行或重複作業，並且希望在使用者不進行人為介入的情況下執行這些作業。SPSS Modeler Batch 是該產品的一個特殊版本，可提供對 SPSS Modeler 完整分析性能的支援，而無需存取一般的使用者介面。要使用 SPSS Modeler Batch，需要 SPSS Modeler Server。

## IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher 是一個支持建立 SPSS Modeler 串流的打包版本的工具，該版本的串流可以由外部執行時期引擎執行或內含到外部應用程式中。通過這種方式，您可以發行和部署完整的 SPSS Modeler 串流以用於未安裝 SPSS Modeler 的環境。SPSS Modeler Solution Publisher 作為 IBM SPSS Collaboration and Deployment Services - 評分 服務的組成部分分發，需要個別的授權。通過此授權，您可以接收 SPSS Modeler Solution Publisher Runtime，它使您能夠執行已發佈的串流。

有關 SPSS Modeler Solution Publisher 的進一步資訊，請參閱 IBM SPSS Collaboration and Deployment Services 文件。IBM SPSS Collaboration and Deployment Services Knowledge Center 包含名為 "IBM SPSS Modeler Solution Publisher" 和 "IBM SPSS Analytics Toolkit" 的部分。

## IBM SPSS Collaboration and Deployment Services 的 IBM SPSS Modeler Server 配接器

IBM SPSS Collaboration and Deployment Services 的一些配接器使 SPSS Modeler 和 SPSS Modeler Server 能夠與 IBM SPSS Collaboration and Deployment Services 儲存庫進行交互。通過這種方式，部署到儲存庫的 SPSS Modeler 串流可以由多個使用者共用，或者從瘦用戶端應用程式 IBM SPSS Modeler Advantage 進行存取。請將配接器安裝在管理儲存庫的系統上。

---

## IBM SPSS Modeler 版本

SPSS Modeler 推出了下列版本。

### SPSS Modeler Professional

SPSS Modeler Professional 提供處理大多數類型的結構化資料所需要的所有工具，例如 CRM 系統中追蹤的行為和互動、個人背景資訊、採購行為和銷售資料。

### SPSS Modeler Premium

SPSS Modeler Premium 是一個個別授權的產品，它對 SPSS Modeler Professional 進行了延伸，以便後者能夠處理專門的資料和非結構化文字資料。SPSS Modeler Premium 包含 IBM SPSS Modeler Text Analytics：

**IBM SPSS Modeler Text Analytics** 採用了先進語言技術和自然語言正在處理 (NLP)，以快速正在處理大量非結構化文字資料，擷取和群組織關鍵概念，以及將這些概念分為各式各樣的種類。擷取的概念和種類可以和現有結構化資料中進行已結合（例如人口統計學），並且可用於借助 IBM SPSS Modeler 的一整套資料採礦工具來進行建模，以此實現更好更集中的決策。

## IBM SPSS Modeler Subscription

IBM SPSS Modeler Subscription 會提供與傳統 IBM SPSS Modeler 用戶端完全相同的預測分析功能。W 使用 Subscription 版，您可以定期下載產品更新項目。

---

### 說明文件

文件可以從 SPSS Modeler 中的「說明」功能表獲取。這樣會開啟可在產品外部公開存取的 Knowledge Center。

作為產品下載的一部分，還會在個別的壓縮資料夾中以 PDF 格式提供每個產品的完整文件（包括安裝指示）。也可以從 Web 下載 PDF 文件，地址為：<http://www.ibm.com/support/docview.wss?uid=swg27046871>。

### SPSS Modeler Professional 文件

SPSS Modeler Professional 文件套組（安裝指示除外）如下。

- **IBM SPSS Modeler 使用者手冊**。關於使用 SPSS Modeler 的一般簡介，包括如何建置資料串流、處理遺漏值、建置 CLEM 表示式使用專案與報告，以及包裝串流以部署至 IBM SPSS Collaboration and Deployment Services 或 IBM SPSS Modeler Advantage。
- **IBM SPSS Modeler Source、Process 和 Output 節點**。說明用於以不同格式讀取、處理和輸出資料的所有節點。實際上這表示所有節點而非建模節點。
- **IBM SPSS Modeler Modeling 節點**。說明所有用於建立資料採礦模型的節點。IBM SPSS Modeler 提供擷取自機器學習人工智慧以及統計資料的各種建模方法。
- **IBM SPSS Modeler 應用程式手冊**。本手冊中的範例旨在為具體的建模方法和技術提供具有針對性的簡介。還可以在「說明」功能表中查閱本手冊的線上版本。請參閱主題第 4 頁的『應用程式範例』，瞭解進一步資訊。
- **IBM SPSS Modeler Python Scripting 和自動化**。通過編寫 Python Script 實現系統自動化的相關資訊，其中包含可以用於操作節點和串流的內容的資訊。
- **IBM SPSS Modeler 部署手冊**。有關在 IBM SPSS Deployment Manager 下以正在處理工作的步驟形式執行 IBM SPSS Modeler 串流的資訊。
- **IBM SPSS Modeler CLEF 開發者手冊**。CLEF 提供了將第三方程式（例如，資料處理常式或建模演算法）作為節點整合到 IBM SPSS Modeler 的功能。
- **IBM SPSS Modeler 資料庫內挖掘手冊**。有關如何利用資料庫的功能通過第三方演算法來改進效能並增強分析功能的資訊。
- **IBM SPSS Modeler Server 管理和效能手冊**。提供有關如何配置和管理 IBM SPSS Modeler Server 的資訊。
- **IBM SPSS Deployment Manager 使用手冊**。有關使用 Deployment Manager 應用程式中包含的管理主控台使用者介面來監視和配置 IBM SPSS Modeler Server 的資訊。
- **IBM SPSS Modeler CRISP-DM 手冊**。借助 CRISP-DM 方法進行 SPSS Modeler 資料採礦的分步手冊。
- **IBM SPSS Modeler Batch 使用者手冊**。提供在批次模式下使用 IBM SPSS Modeler 的完整指導，包含批次模式執行和指令行引數的詳細資料。本手冊僅以 PDF 格式提供。

### SPSS Modeler Premium 文件

SPSS Modeler Premium 文件套組（安裝指示除外）如下。

- **SPSS Modeler Text Analytics 使用者手冊**。提供有關將文字分析與 SPSS Modeler 配合使用的資訊，包括文字採集節點、互動式工作台、範本和其他資源。

---

## 應用程式範例

SPSS Modeler 中的資料採礦工具可以說明解決很多業務和組織問題，應用程式範例將提供有關特定建模方法和技術的簡要的針對性說明。此處使用的資料集比某些資料挖掘器管理的大量資料儲存庫小得多，但涉及的概念和方法可擴展到實際應用程式。

要存取範例，請在 SPSS Modeler 中按一下「說明」功能表中的**應用程式範例**。

資料檔案和樣本串流安裝在產品安裝目錄下的 Demos 資料夾中。如需相關資訊，請參閱『Demos 資料夾』。

**資料庫建模範例**。請參閱 *IBM SPSS Modeler 資料庫內挖掘手冊* 中的範例。

**Scripting 範例**。請參閱 *IBM SPSS Modeler Script 編寫和自動化手冊* 中的範例。

---

## Demos 資料夾

與應用程式範例一起使用的資料檔案和樣本串流安裝在產品安裝目錄下的 Demos 資料夾中（例如：C:\Program Files\IBM\SPSS\Modeler\\Demos）。可以從 Windows「開始」功能表上的 IBM SPSS Modeler 程式群組存取此資料夾，也可以通過按一下**檔案 > 開啟串流對話框**中最近的目錄的清單上的 Demos 來進行存取。

---

## 授權追蹤

當您使用 SPSS Modeler 時，系統會定期追蹤並記錄授權使用情況。所記錄的授權度量值為 *AUTHORIZED\_USER* 和 *CONCURRENT\_USER*，並且記錄的度量值類型取決於您針對 SPSS Modeler 具有的授權類型。

產生的日誌檔可由 IBM License Metric Tool 處理，通過該工具可產生授權使用情形報告。

授權日誌檔建立在記錄 SPSS Modeler 用戶端日誌檔的目錄（依預設為 %ALLUSERSPROFILE%/IBM/SPSS/Modeler/<version>/log）中。



---

## 第 2 章 來源節點

---

### 概述

使用來源節點能夠匯入以多種格式儲存的資料，例如純文字檔、IBM SPSS Statistics (.sav)、SAS、Microsoft Excel 和 ODBC 相容關聯式資料庫。也可以使用使用者輸入節點產生綜合資料。

「來源」選用區包含下列節點：



Analytic Server 來源可讓您在「Hadoop 分散式檔案系統 (HDFS)」上執行串流。Analytic Server 資料來源中的資訊可以來源於變化位置，例如文字檔和資料庫。請參閱第 11 頁的『Analytic Server 來源節點』主題，以取得更多資訊。



可使用「資料庫」節點並利用 ODBC（開放式資料庫連線功能）來匯入各種其他套件中的資料，其中包括 Microsoft SQL Server、Db2、Oracle 以及其他。請參閱第 15 頁的『資料庫來源節點』主題，以取得更多資訊。



「變數檔案」節點可讀取自由欄位文字檔（亦即，其記錄包含固定數目的欄位但字元數目不同的檔案）中的資料。對於頁首文字為固定長度且具有特定註釋類型的檔案，此節點也很有用。請參閱第 23 頁的『「變數檔案」節點』主題，以取得更多資訊。



「固定檔案」節點將匯入固定欄位文字檔中的資料，亦即，其欄位未定界但在同一位置啟動且具有固定長度的檔案。機器產生的資料或舊版資料通常以固定欄位格式儲存。請參閱第 26 頁的『固定檔案節點』主題，以取得更多資訊。



Statistics 檔案節點從 IBM SPSS Statistics 使用的 .sav 或 .zsav 檔案格式以及儲存在 IBM SPSS Modeler 中的快取檔（也使用同一格式）讀取資料。



Data Collection 節點從符合 Data Collection 資料模型的市場意見調查軟體所用的各種格式中匯入意見調查資料。必須安裝 Data Collection Developer Library 才可使用此節點。請參閱第 28 頁的『Data Collection 節點』主題，以取得更多資訊。



IBM Cognos 來源節點從 Cognos Analytics 資料庫匯入資料。



IBM Cognos TM1 來源節點從 Cognos TM1 資料庫匯入資料。



SAS 檔案節點可將 SAS 資料匯入到 IBM SPSS Modeler 中。請參閱第 37 頁的『SAS 來源節點』主題，以取得更多資訊。



Excel 節點可從 Microsoft Excel 以 .xlsx 檔案格式匯入資料。不需要 ODBC 資料來源。請參閱第 38 頁的『Excel 來源節點』主題，以取得更多資訊。



XML 來源節點將 XML 格式的資料匯入到串流中。可以匯入某個目錄中的單個檔案或所有檔案。還選擇指定綱目檔，以從中讀取 XML 結構。



「使用者輸入」節點可讓您輕鬆從頭啟動建立綜合資料，或透過變更現有資料來建立。例如，當您要為建模建立一個測試資料集時，此項很有用。請參閱第 40 頁的『使用者輸入節點』主題，以取得更多資訊。



「模擬產生」節點提供了一種產生模擬資料的簡單方法 - 使用使用者指定的統計分佈從頭開始產生資料，或者使用對現有歷程資料執行「模擬配適」節點而獲取的分佈自動產生資料。對於輸型輸入中存在不確定性的情況，此節點在對預測模型的結果進行評估時非常有用。



「資料視圖」節點可用於存取 IBM SPSS Collaboration and Deployment Services 分析資料視圖中定義的資料來源。分析資料視圖定義了一個用於存取資料的標準介面，並將多個實體資料來源與該介面進行關聯。請參閱第 56 頁的『資料視圖節點』主題，以取得更多資訊。



使用「地理空間」來源節點將地圖或空間資料匯入資料採礦階段作業。請參閱第 57 頁的『「地理空間」來源節點』主題，以取得更多資訊。

要開始執行串流，可將來源節點新增到串流畫布中。然後，按兩下該節點以開啟其對話框。通過此對話框中各式各樣的標籤能夠讀取資料；檢視欄位和值；設定各式各樣的選項，包含過濾器、資料類型、欄位角色和遺漏值檢查。

---

## 設定欄位儲存類型和格式

使用固定「檔案節點」、「變數檔案」、「XML 來源」和「使用者輸入」節點的「資料」標籤中的選項，可以在 IBM SPSS Modeler 中匯入或建立欄位時為欄位指定儲存類型。對於「固定檔案」節點、「變數檔案」節點和「使用者輸入」節點，您還可以指定欄位格式化以及其他 meta 資料。

對於從其他來源中讀取的資料，儲存類型自動決定，但可使用轉換函數（例如 `to_integer`）在充填節點或衍生節點中對其進行變更。

**欄位** 使用欄位欄可以檢視並選取現行資料集中的欄位。

**置換** 選中置換欄中的勾選框可以啟動儲存和輸入格式欄中的選項。

### 資料儲存

儲存體說明資料儲存在欄位中的方式。例如，值為 1 和 0 的欄位會儲存整數資料。它與測量層次不同，說明資料使用情形，並不會影響儲存體。例如，您可能想要使用值 1 及 0，將整數欄位的測量層次設為旗標。這通常指出  $1 = True$  及  $0 = False$ 。儲存格式必須在數據來源中確定，而測量層次可以使用「類型」節點在串流中的任意點上進行變更。請參閱第 118 頁的『測量層次』主題，以取得更多資訊。

可用儲存類型有：

- **字串** 用於包含非數值資料（也稱為英數資料）的欄位。字串可以包含任何字元序列，比如 *fred*、*Class 2* 或 *1234*。注意：字串中的數字不能用於計算。
- **整數** 值為整數的欄位。
- **實數** 值為可能包含小數（不限於整數）的數字。顯示格式在「串流內容」對話框中指定，並且可以被「類型」節點（「格式」標籤）中的各個欄位覆蓋。
- **日期** 以標準格式指定的日期值，例如年月日（例如 2007-09-26）。具體格式在「串流內容」對話框中指定。
- **時間** 以持續時間形式測量的時間。例如，某個服務電話持續 1 小時 26 分 38 秒，該時間可以根據「串流內容」對話框中指定的目前時間格式代表為：01:26:38。
- **時間戳記** 同時包含日期與時間元件的值，例如 2007-09-26 09:04:00，具體同樣取決於「串流內容」對話框中的現行日期和時間格式。請注意，需要用雙將時間戳記值括起來，以確保將此值解釋為單一值而非個別的日期與時間值。（同樣適用於在使用者輸入節點中輸入值時的情況。）
- **清單** 在 SPSS Modeler 第 17 版中，隨新測量層次「地理空間」和「收集」一起引入了「清單」儲存欄位，對於單個記錄，此欄位包含多個值。存在所有其他儲存類型的清單版本。

表 1. 清單儲存類型圖示

圖示	儲存體類型
[A]	字串清單
[D]	整數清單
[F]	實數清單
[T]	時間清單
[C]	日期清單
[M]	時間戳記清單
[>]	深度大於零的清單

另外，為了與「收集」測量層次配合使用，提供了下列測量層次的清單版本。

表 2. 清單測量層次圖示

圖示	測量層級
[R]	連續清單
[C]	種類清單
[F]	旗標清單
[N]	名義清單
[O]	序數清單

可以通過三個來源節點（Analytic Server、「地理空間」或「變數檔案」）中的某一個將清單匯入到 SPSS Modeler 中，也可以在串流中使用「衍生」或「填入器」欄位作業節點建立清單。

有關「清單」及其與「收集」和「地理空間」測量層次的互動的進一步資訊，請參閱『清單儲存以及相關聯的測量層次』。

**儲存轉換。**用戶可使用各式各樣的轉換函數來轉換某個欄位的儲存格式，比如「填入器」節點中的 `to_string` 和 `to_integer`。請參閱第 138 頁的『使用填入節點進行儲存類型轉換』主題，以取得更多資訊。請注意，轉換函數（以及需要指定類型輸入（如日期或時間值）的任何其他函數）取決於「串流內容」對話框中指定的現行形式。例如，如果想將值為 *Jan 2003*、*Mar 2003* 等字串欄位轉換為日期儲存格式，請選取 **MON YYYY** 作為串流的預設日期格式。轉換函數也可從「衍生」節點取得，用於在衍生計算期間進行暫時轉換。另外，還可以使用「衍生」節點來執行其他操作，比如使用種類值對字串欄位進行重新編碼。請參閱第 137 頁的『使用衍生節點對值進行重新編碼』主題，以取得更多資訊。

**混合資料中的讀數。**請注意，讀取數值儲存格式（整數、實數、時間、時間戳記或日期）的欄位中的讀數時，任何非數值將被設定為無效或系統遺漏值。這時因為 IBM SPSS Modeler 與某些應用程式不同，它不容許欄位中含有混合儲存類型。為了避免出現混合儲存類型，必須根據需要更改來源節點中或外部應用程式中的儲存類型，從而將任何具有混合資料的欄位以字串的格式讀取入。

## 欄位輸入格式（僅限「固定檔案」節點、「變數檔案」節點和「使用者輸入」節點）

對於除字串和整數以外的所有儲存類型，都可以使用下拉清單為選定的欄位指定格式選項。例如，從不同的語言環境中合併資料時，可能需要為一個欄位指定句號 (.) 作為小數分隔字元，而為另一個欄位指定逗點分隔字元。

在來源節點中指定的輸入選項會置換在串流內容對話框中指定的格式選項；但是這些指定的輸入選項不會在串流的其他位置中保留。根據所掌握的資料知識，這些選項可用來正確地剖析輸入資料。指定的格式可用作將資料讀取到 IBM SPSS Modeler 時剖析資料的指導，但不能確定將資料讀取到 IBM SPSS Modeler 之後應對其格式化的方式。要在串流的其他位置處基於每個欄位指定格式，可使用「類型」節點的「格式」標籤。請參閱第 128 頁的『欄位格式設定標籤』主題，以取得更多資訊。

選項根據儲存類型的變化而變化。例如，對於實數儲存類型，可以選取句號 (.) 或 逗點 (,) 作為小數分隔字元。對於時間戳記欄位，從下拉清單中選取指定將開啟一個個別的對話框。請參閱第 128 頁的『設定欄位格式選項』主題，以取得更多資訊。

對於所有儲存類型，也可以選取串流預設值以便匯入時使用串流預設值。串流設定可在串流內容對話框中指定。

## 其他選項

使用「資料」標籤可指定其他幾個選項：

- 要檢視不再通過目前節點連接的資料（例如，訓練資料）的儲存設定，可選取檢視未使用的欄位設定。可通過按一下清除清除遺產欄位。
- 在此對話框內工作時，可隨時按一下重新整理以重新載入資料來源中的欄位。在更改到來源節點的資料連線時，或在對話框的標籤之間進行操作時，此操作都非常有用。

## 清單儲存以及相關聯的測量層次

為了處理新測量層次「地理空間」和「收集」，在 SPSS Modeler 第 17 版中引入了「清單儲存」欄位，對於單個記錄，此欄位包含多個值。清單以方括弧 ([]) 括住。清單的範例包括：`[1,2,4,16]` 和 `["abc", "def"]`。

可以通過三個來源節點（Analytic Server、「地理空間」或「變數檔案」）中的某一個將清單匯入到 SPSS Modeler 中，可以在串流中使用「衍生」或「填入器」欄位作業節點建立清單，或者在使用「排名式條件」合併方法時由「合併」節點產生該清單。

清單被視為具有深度；例如，將各個項目括在單個方括弧內的簡式清單（格式為 [1,3]）在 IBM SPSS Modeler 中記錄為深度為零。除了深度為零的簡式清單以外，您還可以使用巢套清單，這種清單中的每個值本身也是清單。

巢套清單的深度取決於相關聯的測量層次。對於「無類型」，未設定深度限制；對於「收集」，深度為零；對於「地理空間」，深度必須介於 0 與 2（含首尾值）之間，具體取決於巢套項目數。

對於零深度清單，可以將測量層次設定為「地理空間」或「收集」。這兩個層次都是父測量層次，您可以在「值」對話框中設定測量層次資訊。「收集」的測量層次確定該清單中的元素的測量層次。除「無類型」和「地理空間」以外的所有測量層次都可以用作「收集」的子層次。「地理空間」測量層次有 6 個子層次，分別為 Point、LineString、Polygon、MultiPoint、MultiLineString 和 MultiPolygon；如需相關資訊，請參閱第 120 頁的『地理空間測量層次』。

註：「收集」測量層次只能與深度為 0 的清單配合使用，「地理空間」測量層次只能與最大深度為 2 的清單配合使用，「無類型」測量層次可以與任意深度的清單配合使用。

下列範例使用「地理空間」測量層次「點」和「線串」的結構來顯示深度為零的清單與巢套清單之間的差別：

- 「地理空間」測量層次「點」的欄位深度為 0：

[1,3] 兩個座標

[1,3,-1] 三個座標

- 「地理空間」測量層次「線串」的欄位深度為 1：

[ [1,3], [5,0] ] 兩個座標

[ [1,3,-1], [5,0,8] ] 三個座標

「點」欄位（深度為零）是正常清單，其中的每個值都由兩個或三個座標構成。「線串」欄位（深度為 1）是點清單，其中的每個點都由進一步的清單值數列構成。

有關建立清單的進一步資訊，請參閱第 135 頁的『衍生清單或地理空間欄位』。

---

## 不受支援的控制字元

SPSS Modeler 中的某些過程無法處理包含多種控制字元的資料。如果您的資料使用這些字元，可能會看到錯誤訊息，例如下列範例：

```
Unsupported control characters found in values of field {0}
```

不受支援的字元包括：0x0 到 0x3F（不包括 0x0 和 0x3F）以及 0x7F；但是，Tab 鍵 (0x9(\t))、新行 (0xA(\n)) 以及換行 (0xD(\r)) 不會導致出現問題。

如果看到與不受支援的字元有關的錯誤訊息，請在串流中「來源」節點之後使用「填入器」節點和 CLEM 表示式 `stripctrlchars` 取代這些字元。

---

## Analytic Server 來源節點

Analytic Server 來源可讓您在「Hadoop 分散式檔案系統 (HDFS)」上執行串流。Analytic Server 資料來源中的資訊可來源於各個位置，其中包含：

- HDFS 上的文字檔
- 資料庫
- HCatalog

通常，將在 HDFS 上執行具有 Analytic Server 來源的串流；但是，如果串流包含不支援在 HDFS 上執行的節點，那麼系統會將串流盡可能多的部分「推送回」Analytic Server，然後 SPSS Modeler Server 將試圖處理剩餘的串流。您將需要對超大資料集進行二次抽樣；例如，通過將「樣本」節點放在串流中進行二次抽樣。

如果您想要使用自己的 Analytic Server 連線來代替管理者定義的預設連線，請取消選取使用預設 **Analytic Server** 並選取您的連線。有關設定多個 Analytic Server 連線的詳細資料，請參閱連線到 Analytic Server。

**資料來源。**假定您或者 SPSS Modeler Server 管理者已建立連線，您可以選取包含要使用的資料的資料來源。資料來源包含與該來源相關聯的檔案和 meta 資料。按一下選取以顯示可用資料來源的清單。請參閱『選取資料來源』主題，以取得更多資訊。

如果您需要建立新的資料來源或編輯現有資料來源，請按一下**啟動資料來源編輯器...**。

請注意，使用多個 Analytic Server 連線在控制資料流程方面十分有用。例如，使用 Analytic Server「來源」及「匯出」節點時，您可能想要在串流的不同分支中使用不同的 Analytic Server 連線，因此每一個分支執行時，它會使用自己的 Analytic Server，且不會將任何資料拉取到 IBM SPSS Modeler Server。請注意，如果分支包含多個 Analytic Server 連線，則會從 Analytic Server 拉取到 IBM SPSS Modeler Server。如需相關資訊（包括限制），請參閱 Analytic Server 串流內容。

### 選取資料來源

「資料來源」表格顯示了可用資料來源的清單。選取要使用的來源，然後按一下**確定**。

按一下**顯示擁有者**以顯示資料來源擁有人。

**過濾依據**使您能夠按**關鍵字**對資料來源清單進行過濾，這將根據資料來源名稱和資料來源說明或者擁有人檢查過濾準則。您可以輸入下方說明的字串、數值或萬用字元的組合作為過濾準則。搜尋字串區分觀察值。按一下**重新整理**以更新「資料來源」表格。

\_ 可以使用底線來代表搜尋字串中的任意單個字元。

% 可以使用百分號來代表搜尋字串中零個以上字元的任意序列。

### 修正認證

如果用於存取 Analytic Server 的認證與用於存取 SPSS Modeler Server 的認證不同，那麼在 Analytic Server 上執行串流時，您將需要輸入 Analytic Server 認證。如果您不知道您的認證，請聯絡伺服器管理者。

### 受支援的節點

受支援在 HDFS 上執行多數 SPSS Modeler 節點，但是某些節點的執行過程可能存在一些差異，而且一些節點目前不受支援。本主題詳細介紹了目前支援層次。

#### 一般化

- Analytic Server 將不接受加引號的 Modeler 欄位名稱中通常接受的某些字元。

- 要在 Analytic Server 中執行建模器串流，該串流必須以一個或多個 Analytic Server 「來源」節點開頭，並以單個建模節點或 Analytic Server 「匯出」節點結尾。
- 建議您將連續目標的儲存設定為實數而不是整數。評分模型始終會將實數值寫入連續目標的輸出資料檔案，而評分的輸出資料模型將遵循目標儲存設置。因此，如果連續目標具有整數儲存，那麼寫入值與分數的資料模型將出現不符，並且此不符會在您試圖讀取已評分資料時導致發生錯誤。
- 如果欄位測量是「地理空間」，那麼不支援 @OFFSET 的函數。

## 來源

- 以 Analytic Server 來源節點之外的任何內容開始的串流將在本端執行。

## 記錄作業

受支援所有記錄作業，但串流 TS 和 Space-Time-Boxes 節點除外。以下是關於受支援節點功能的更多注意事項。

### 選取

- 支援「衍生」節點所支援的同一組函數。

### 樣本

- 不支援區塊層次取樣。
- 不支援複合取樣方法。
- 不支援前設定了 "Discard sample" 的前 n 取樣。
- 不支援  $N > 20000$  的前 n 取樣。
- 如果未設定 "Maximum sample size"，則不支援 1-in-n 取樣。
- 如果  $N * \text{"Maximum sample size"} > 20000$ ，則不支援 1-in-n 取樣。
- 不支援隨機 % 區塊層次取樣。
- 隨機 % 目前支援提供種子值。

### 聚集

- 不支援連續鍵。如果您要重複使用對資料進行排序而設定的現有串流，然後在「聚合」節點中使用此設定，請變更此串流以移除「排序」節點。
- 順序統計資料（中位數、第 1 個四分位數和第 3 個四分位數）以近似方式計算，並透過「最佳化」標籤進行支援。

### 排序

- 不支援「最佳化」標籤。

在分散式環境中，存在有限號碼的作業，這些作業將保留「排序」節點所確定的記錄順序。

- 後跟「匯出」節點的「排序」將生成已排序的資料來源。
- 後跟「樣本」節點（包含第一個記錄取樣）的「排序」將傳回前 N 條記錄。

通常，放置「排序」節點的位置應該盡可能靠近需要已排序記錄的作業。

### 合併(M)

- 不支援按順序合併。
- 不支援「最佳化」標籤。
- 合併作業的速度相對較慢。如果 HDFS 中有可用空間，那麼與合併下列每個串流中的資料來源相比，合併資料來源一次並在這些串流中使用合併後的來源的速度更快。

### R 轉換

此節點中的 R 語法應該包含一次記錄作業。



## 欄位作業

受支援所有欄位作業，但「匿名化」、「轉置」、「時間間隔」和「歷史」節點除外。以下是關於受支援節點功能的更多注意事項。

### 自動資料準備

- 不支援對節點進行訓練。受支援將經過訓練的「自動資料準備」節點中的轉換套用於新資料。

### 衍生

- 受支援除序列函數以外的所有衍生函數。
- 衍生新欄位以作為「計數」本質上是一個序列作業，因此不受支援。
- 不能在使用分割欄位作為分割的同一串流中衍生這些欄位；您將需要建立兩個串流，一個用於衍生分割欄位，另一個使用該欄位作為分割。

### 填充器

- 支援「衍生」節點所支援的同一組函數。

**分組** 不支援下列功能。

- 最佳化分級
- 等級
- 分位顯示 - > 分位：值總和
- 分位顯示 - > 同分值：保留在目前分級中並隨機分配
- 分位顯示 - > 自訂 N：超過 100 的值，以及任何 100 % N 不等於 0 的 N 值。

### RFM 分析

- 不支援用於處理同分值的「保留在目前分級中」選項。RFM 近因、頻率和貨幣分數並非始終與 Modeler 根據同一資料計算得出的分數相比對。分數範圍相同，但分數分配（分級數）可能相差 1。

**圖形** 受支援所有「圖表」節點。

**建模** 支援下列「建模」節點：時間序列、TCM、Isotonic-AS、延伸模型、樹狀結構-AS、C&R 樹狀結構、Quest、CHAID、線性、線性 AS、類神經網路、GLE、LSVM、二階 AS、隨機樹狀結構、STP、關聯規則、XGBoost-AS、隨機森林及 K 平均數 AS。下面提供了有關這些節點的功能的進一步說明。

**線性** 在海量資料上建置模型時，您通常想要將目標變更為極大型資料集，或指定分割。

- 不支援繼續訓練現有模型。
- 僅在以下情況時才建議使用標準模型建置目標：已定義分割欄位，以便每個分割中的記錄數不會太大，其中「太大」的定義取決於 Hadoop 叢集中個別節點的冪。相反，您還需要小心確保未過於精細地定義分割，以防記錄太少而無法建置模型。
- 不支援 Boosting 目標。
- 不支援 Bagging 目標。
- 當記錄很少時，不建議使用極大型資料集目標；它通常不會建置模型，或建置降級模型。
- 不支援自動資料準備。嘗試根據具有多個遺漏值的資料建立模型時，這可能會產生問題；通常，將在自動資料準備過程中插補這些遺漏值。暫行解決方法是將樹狀結構模型或類神經網路與進階設定配合使用，以插補所選的遺漏值。
- 不會計算分割模型的準確性統計量。

### 神經網路

在海量資料上建置模型時，您通常想要將目標變更為極大型資料集，或指定分割。

- 不支援對現有標準或 PSM 模型進行持續訓練。
- 僅在以下情況時才建議使用標準模型建置目標：已定義分割欄位，以便每個分割中的記錄數不會太大，其中「太大」的定義取決於 Hadoop 叢集中個別節點的冪。相反，您還需要小心確保未過於精細地定義分割，以防記錄太少而無法建置模型。
- 不支援 Boosting 目標。
- 不支援 Bagging 目標。
- 當記錄很少時，不建議使用極大型資料集目標；它通常不會建置模型，或建置降級模型。
- 如果資料中存在多個遺漏值，請使用進階設定來插補遺漏值。
- 不會計算分割模型的準確性統計量。

### C&R 樹狀結構、CHAID 和 Quest

在海量資料上建置模型時，您通常想要將目標變更為極大型資料集，或指定分割。

- 不支援繼續訓練現有模型。
- 僅在以下情況時才建議使用標準模型建置目標：已定義分割欄位，以便每個分割中的記錄數不會太大，其中「太大」的定義取決於 Hadoop 叢集中個別節點的冪。相反，您還需要小心確保未過於精細地定義分割，以防記錄太少而無法建置模型。
- 不支援 Boosting 目標。
- 不支援 Bagging 目標。
- 當記錄很少時，不建議使用極大型資料集目標；它通常不會建置模型，或建置降級模型。
- 不支援互動式階段作業。
- 不會計算分割模型的準確性統計量。
- 出現分割欄位時，在 Modeler 本端建立的樹狀結構模型與 Analytic Server 所建立的樹狀結構模型略有不同，從而生生成不同的分數。在這兩種情況下，演算法是有效演算法；Analytic Server 所使用的演算法更新。假定樹狀結構演算法傾向於具有多數啟發式規則，兩個元件之間的差異是正常的。

### 模型評分

受支援進行建模的模型也受支援進行評分。另外，對於下列節點，受支援對本端構建的模型塊進行評分：C&RT、Quest、CHAID、線性和類神經網路（無論是標準模型、增強型組裝模型還是對於超大資料集均如此）、迴歸方法、C5.0、Logistic、Genlin、GLMM、Cox、SVM、貝葉斯網絡、二階、KNN、決策清單、區別元件、自學、異常偵測、Apriori、Carma、K-Means、Kohonen、R 和文字採集。

- 不會對原始傾向或調整傾向進行評分。作為暫行解決方法，可以使用「衍生」節點搭配下列表示式，手動計算原始傾向來獲得相同效果：`if 'predicted-value' == 'value-of-interest' then 'prob-of-that-value' else 1-'prob-of-that-value' endif`

**R** 模型塊中的 R 語法應該包含一次記錄作業。

**輸出** 受支援下列節點：矩陣、分析、資料審核、轉換、設定廣域、統計資料、平均數和表格。以下是關於受支援節點功能的更多注意事項。

#### 資料審核

「資料審核」節點無法生成連續欄位的方式。

#### 平均數

「平均數」節點無法生成標準誤差或 95% 信賴區間。

**表格** 通過寫入包含上游作業結果的暫時 Analytic Server 資料來源來受支援「表格」節點。然後，「表格」節點將分頁讀取該資料來源的內容。

**匯出** 串流可以開始於 Analytic Server 來源節點，並以 Analytic Server 匯出節點之外的匯出節點結束，但資料將從 HDFS 移至 SPSS Modeler Server，並最終移至匯出位置。

---

## 資料庫來源節點

可使用「資料庫」來源節點並利用 ODBC（開放式資料庫連線功能）來匯入各種其他套件中的資料，其中包括 Microsoft SQL Server、Db2、Oracle 以及其他。

若要讀取或寫入資料庫，您必須已針對相關資料庫，安裝並配置 ODBC 資料來源，並視需要具有讀取或寫入權。IBM SPSS Data Access Pack 包括一組 ODBC 驅動程式，可用於此目的，且這些驅動程式可從下載網站上取得。如果您在建立或設定 ODBC 資料來源許可權方面存在問題，請聯絡您的資料庫管理者。

## 受支援的 ODBC 驅動程式

如需受支援且已測試可與 IBM SPSS Modeler 搭配使用之資料庫及 ODBC 驅動程式的最新資訊，請參閱公司支援網站上產品相容性矩陣 (<http://www.ibm.com/support>)。

## 在何處安裝驅動程式

註：ODBC 驅動程式必須在可能發生處理的每一部電腦上安裝並配置。

- 如果以本端（獨立式）模式執行 IBM SPSS Modeler，則驅動程式必須安裝在本端電腦上。
- 如果針對遠端 IBM SPSS Modeler Server 以分散式模式執行 IBM SPSS Modeler，則 ODBC 驅動程式需要安裝在已安裝 IBM SPSS Modeler Server 的電腦上。對於 UNIX 系統上的 IBM SPSS Modeler Server，另請參閱本節中稍後的「在 UNIX 系統上配置 ODBC 驅動程式」。
- 如果您需要從 IBM SPSS Modeler 及 IBM SPSS Modeler Server 存取相同的資料來源，則 ODBC 驅動程式必須安裝在這兩部電腦上。
- 如果您透過終端機服務執行 IBM SPSS Modeler，則 ODBC 驅動程式需要安裝在已安裝 IBM SPSS Modeler 的終端機服務伺服器上。

## 存取資料庫中的資料

要存取資料庫中的資料，請完成下列步驟。

- 安裝 ODBC 驅動程式並向您要使用的資料庫配置資料來源。
- 在「資料庫」節點對話框中，使用「表格」方式或「SQL 查詢」方式連接至資料庫。
- 從資料庫中選取表格。
- 通過使用「資料庫」節點對話框中的標籤，您可以更改用法類型和過濾資料欄位。

在相關文件主題中提供了更多有關上述步驟的詳細資料。

註：如果從 SPSS Modeler 中呼叫資料庫儲存程序 (SP)，那麼您可能會看到傳回了名為 RowsAffected 的單個輸出欄位，而不是傳回預期的 SP 輸出。當 ODBC 未傳回足夠的資訊，導致無法確定 SP 的輸出資料模型時，將發生這種情況。SPSS Modeler 對傳回了輸出的 SP 僅提供了有限的支援，因此建議您從 SP 中擷取 SELECT 並使用下列其中一項動作，而不要使用 SP。

- 建立基於 SELECT 的視圖並在「資料庫來源」節點中選擇該視圖
- 直接在「資料庫來源」節點中使用 SELECT。

## 設定資料庫節點選項

可以使用「資料庫來源」節點對話框的「資料」標籤中的選項，來獲取對資料庫的存取並從選定的表格中讀取資料。

**眾數。** 選取**表格**可通過對話框控制項連接至表格。

選取 **SQL 查詢**以使用 SQL 查詢下方選取的資料庫。請參閱第 21 頁的『查詢資料庫』主題，以取得更多資訊。

**資料來源。**對於表格方式和 SQL 查詢方式，都可以在資料來源欄位中輸入名稱或者從下拉清單中選取新增新的資料庫連線。

下列選項用於連接至資料庫和選取表格（使用對話框）：

**表格名。**如果知道要存取的表格的名稱，那麼可在表格名欄位中輸入此名稱。否則，可按一下**選取**按鈕開啟列出了可用的表格的對話框。

**用引號括住表格與直欄名稱。**在資料庫中進行查詢時，指定是否要將表格名和直欄名稱括入引號內（例如，這些名稱包含空格或標點符號）。

- 選中**視需要**選項將僅在表格名和欄位名稱包含非標準字元時參照它們。非標準字元包含非 ASCII 字元、空格字元和除全角句點 (.) 以外的所有非英數字元。
- 如果想給所有表格名和欄位名稱加引號，則選中**始終**。
- 如果從不想給表格名和欄位名稱加引號，則選中**從不**。

**刪除前導及尾端空格。**選中選項以捨棄字串中開頭和結尾的空格。

**附註。**在使用與不使用 SQL 回送的字串之間的對比可能產生存在尾部空格的不同結果。

**從 Oracle 中讀取空字串。**在 Oracle 資料庫中進行值的讀取寫入時，要注意，與 IBM SPSS Modeler 及大多數其他資料庫不同，Oracle 將字符型空值等同於空值對待並儲存。這表示同樣的資料從 Oracle 資料庫中解壓縮和從檔案或其他資料庫中解壓縮其表現可能有所不同，可能會傳回不同的結果。

## 新增資料庫連線

要開啟資料庫，請先選取要連接的資料來源。從「資料」標籤的資料來源下拉清單中選取新增新的資料庫連線。

此時將開啟「資料庫連線」對話框。

註：要獲取開啟此對話框的替代方案，請從主功能表中選擇：**工具 > 資料庫...**

**資料來源。**列出可用的資料來源。如果看不到所需資料庫，請向下捲動。一旦選取了資料來源並已輸入了任何密碼，即可按一下**連接**。按一下**重新整理**以更新此清單。

**使用者和密碼。**如果資料來源受密碼保護的，請輸入使用者名稱和關聯密碼。

**認證。**如果在 IBM SPSS Collaboration and Deployment Services 中已經配置了認證，那麼您可以選取此選項以便在儲存庫中瀏覽到該認證。此認證的使用者名稱和密碼必須與存取資料庫所需的使用者名稱和密碼相符。

**連線。**顯示目前連接的資料庫。

- **預設值。** 選擇性地選擇一個連線作為預設值。執行此操作將導致資料庫來源或匯出節點使用此預先定義連線作為其資料來源，然而，可根據需要對此連線進行編輯。
- **儲存。** 選擇性地選取一個或多個希望在後續階段作業中重新顯示的連線。
- **資料來源。** 目前連線的資料庫的連線字串。
- **預設。** 指示（使用一個 \* 字元）是否為資料庫連線指定了預設值。要指定預設值，請按一下對應於資料庫連線的列中的此直欄，然後從清單中選擇「指定」。請參閱第 18 頁的『為資料庫連線指定預設值』主題，以取得更多資訊。

要刪除連線，可從清單選取一個連線，然後按一下刪除。

完成選擇後，請按一下**確定**。

若要讀取或寫入資料庫，您必須已針對相關資料庫，安裝並配置 ODBC 資料來源，並視需要具有讀取或寫入權。IBM SPSS Data Access Pack 包括一組 ODBC 驅動程式，可用於此目的，且這些驅動程式可從下載網站上取得。如果您在建立或設定 ODBC 資料來源許可權方面存在問題，請聯絡您的資料庫管理者。

## 受支援的 ODBC 驅動程式

如需受支援且已測試可與 IBM SPSS Modeler 搭配使用之資料庫及 ODBC 驅動程式的最新資訊，請參閱公司支援網站上產品相容性矩陣 (<http://www.ibm.com/support>)。

## 在何處安裝驅動程式

註：ODBC 驅動程式必須在可能發生處理的每一部電腦上安裝並配置。

- 如果以本端（獨立式）模式執行 IBM SPSS Modeler，則驅動程式必須安裝在本端電腦上。
- 如果針對遠端 IBM SPSS Modeler Server 以分散式模式執行 IBM SPSS Modeler，則 ODBC 驅動程式需要安裝在已安裝 IBM SPSS Modeler Server 的電腦上。對於 UNIX 系統上的 IBM SPSS Modeler Server，另請參閱本節中稍後的「在 UNIX 系統上配置 ODBC 驅動程式」。
- 如果您需要從 IBM SPSS Modeler 及 IBM SPSS Modeler Server 存取相同的資料來源，則 ODBC 驅動程式必須安裝在這兩部電腦上。
- 如果您透過終端機服務執行 IBM SPSS Modeler，則 ODBC 驅動程式需要安裝在已安裝 IBM SPSS Modeler 的終端機服務伺服器上。

## 在 UNIX 系統上配置 ODBC 驅動程式

依預設，在 UNIX 系統上，不會針對 IBM SPSS Modeler Server 配置 DataDirect Driver Manager。若要配置 UNIX 以載入 DataDirect Driver Manager，請輸入下列指令：

```
cd <modeler_server_install_directory>/bin
rm -f libspssodbc.so
ln -s libspssodbc_datadirect.so libspssodbc.so
```

這樣做會移除預設鏈結，並會建立 DataDirect Driver Manager 的鏈結。

註：對於某些資料庫，需要 UTF16 驅動程式封套，才能使用 SAP HANA 或 IBM Db2 CLI 驅動程式。DashDB 需要 IBM Db2 CLI 驅動程式。若要建立 UTF16 驅動程式封套的鏈結，請改為輸入下列指令：

```
rm -f libspssodbc.so
ln -s libspssodbc_datadirect_utf16.so libspssodbc.so
```

若要配置 SPSS Modeler Server：

1. 透過將下列行新增至 modelersrv.sh，來配置 SPSS Modeler Server 啟動 Script modelersrv.sh，以找到 IBM SPSS Data Access Pack odbc.sh 環境檔案：

```
. /<pathtoSDAPinstall>/odbc.sh
```

其中，<pathtoSDAPinstall> 是 IBM SPSS Data Access Pack 安裝的完整路徑。

2. 重新啟動 SPSS Modeler Server。

此外，僅限 SAP HANA 及 IBM Db2，新增下列參數定義至 odbc.ini 檔案中的 DSN，以避免在連線期間發生緩衝區溢位：

```
DriverUnicodeType=1
```

註：libspssodbc\_datadirect\_utf16.so 封套也與其他 SPSS Modeler Server 受支援的 ODBC 驅動程式相容。

## 潛在資料庫問題

根據您使用的資料庫，可能存在一些您應該知道的潛在問題。

### IBM Db2

當試圖在串流中快取從 Db2 資料庫讀取資料的節點時，可能會看到下列錯誤訊息：

```
A default table space could not be found with a pagesize of at least 4096 that authorization ID TEST is authorized to use
```

若要配置 Db2 以使資料庫內快取能夠在 SPSS Modeler 中正常工作，資料庫管理者應建立「使用者暫時」表格空間，並將該表格空間的存取權授與相關 Db2 帳戶。

我們建議在新表格空間中使用頁大小 32768，因為這會增加對能夠成功快取的欄位數量的限制。

### IBM Db2 for z/OS

- 在啟用信賴度的情況下，使用產生的 SQL 對部分演算法評分可能會在執行時傳回錯誤。該問題特定於 Db2 for z/OS；若要解決此問題，請使用 SPSS Modeler Server Scoring Adapter for Db2 on z/OS。
- 當對 Db2 for z/OS 執行串流時，如果閒置資料庫連線的逾時已啟用並設定得較低，那麼可能會遇到資料庫錯誤。在 Db2 for z/OS 第 8 版中，預設值從無逾時變更為 2 分鐘。該解決方案是為了增加 Db2 系統參數 IDLE THREAD TIMEOUT (IDTHTOIN) 的值，或將該值重設為 0。

### Oracle

如果執行包含「聚合」節點的串流，那麼將 SQL 回送到 Oracle 資料庫時，針對第一個和第三個四分位數傳回的值可能不同於以原生模式傳回的值。

## 為資料庫連線指定預設值

對於某些資料庫，可以指定用於資料庫連線的一些預設值。這些設定將套用至資料庫匯出。

支援此功能的資料庫類型如下所示。

- SQL Server 企業版和開發人員版本。請參閱第 19 頁的『用於 SQL Server 的設定』主題，以取得更多資訊。
- Oracle 企業版或個人版。請參閱第 19 頁的『用於 Oracle 的設定』主題，以取得更多資訊。
- IBM Db2 for z/OS 及 Teradata 都以類似方式連接到資料庫或綱目。請參閱第 19 頁的『用於 IBM Db2 for z/OS、IBM Db2 LUW 及 Teradata 的設定』主題，以取得更多資訊。

如果連接至不支援此功能的資料庫或綱目，則會提示訊息無法為此資料庫連線配置預設。

## 用於 SQL Server 的設定

這些設定針對 SQL Server Enterprise 和 Developer 版本顯示。

**使用壓縮。** 選取之後，建立表格以使用壓縮匯出。

**壓縮目標。** 選擇壓縮層次。

- **列。** 啟用低層次壓縮（例如，相當於 SQL 中的 CREATE TABLE MYTABLE(...) WITH (DATA\_COMPRESSION = ROW);)。
- **頁。** 啟用頁級壓縮（例如 SQL 中的 CREATE TABLE MYTABLE(...) WITH (DATA\_COMPRESSION = PAGE);)。

## 用於 Oracle 的設定

### Oracle 設定 - 基本選項

這些設定針對使用「基本」選項的 Oracle Enterprise 或 Personal 版本顯示。

**使用壓縮。** 選取之後，建立表格以使用壓縮匯出。

**壓縮目標。** 選擇壓縮層次。

- **預設值。** 啟用預設壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS;）。在此情況下，它與基本選項的效果相同。
- **基本。** 啟用基本壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS BASIC;）。

### Oracle 設定 - 進階選項

這些設定針對使用「進階」選項的 Oracle Enterprise 或 Personal 版本顯示。

**使用壓縮。** 選取之後，建立表格以使用壓縮匯出。

**壓縮目標。** 選擇壓縮層次。

- **預設值。** 啟用預設壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS;）。在此情況下，它與基本選項的效果相同。
- **基本。** 啟用基本壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS BASIC;）。
- **OLTP。** 啟用 OLTP 壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS FOR OLTP;）。
- **查詢低/高。**（僅限 Exadata 伺服器）對查詢啟用 Hybrid 欄壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS FOR QUERY LOW; 或 CREATE TABLE MYTABLE(...) COMPRESS FOR QUERY HIGH;）。查詢壓縮非常適合用在資料倉儲環境中；HIGH 提供比 LOW 更高的壓縮比。
- **封存低/高。**（僅限 Exadata 伺服器）對儲存啟用 Hybrid 欄壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS FOR ARCHIVE LOW; 或 CREATE TABLE MYTABLE(...) COMPRESS FOR ARCHIVE HIGH;）。儲存壓縮非常適合用於壓縮那些需要長時間儲存的資料；HIGH 提供比 LOW 更高的壓縮比。

## 用於 IBM Db2 for z/OS、IBM Db2 LUW 及 Teradata 的設定

在為 IBM Db2 for z/OS、IBM Db2 LUW 或 Teradata 指定預設值時，系統將提示您選取下列各項目：

**使用伺服器評分配器資料庫或使用伺服器評分配器綱目。** 如果已選取，那麼將啟用伺服器評分配器資料庫或伺服器評分配器架構選項。

**伺服器評分配器資料庫或伺服器評分配器綱目** 請從下拉清單中選取所需的連線。

另外，對於 Teradata，還可以設定查詢連接詳細資料以提供附加 meta 資料，從而協助各項目，例如工作量管理；對照、識別、解析查詢；以及追蹤資料庫使用量。

**拼字查詢連接。**選擇是為您使用 Teradata 資料庫連線的整個時間段設定一次查詢連接（針對階段作業），還是每一個執行串流時都進行此設定（針對交易）。

**註：**如果對串流設定了查詢連接，那麼將該串流複製到另一台機器時，該連接將丟失。為了避免發生這種情況，您可以套用 Script 來執行串流，並在 Script 中套用關鍵單字 *querybanding* 來套用所需的設定。

## 需要的資料庫權限

為使 SPSS Modeler 資料庫功能正常運行，請向使用的任何使用者 ID 授權對下列項目的存取權：

### Db2 LUW

SYSIBM.SYSDUMMY1  
SYSIBM.SYSFOREIGNKEYS  
SYSIBM.SYSINDEXES  
SYSIBM.SYSKEYCOLUSE  
SYSIBM.SYSKEYS  
SYSIBM.SYSPARMS  
SYSIBM.SYSRELS  
SYSIBM.SYSROUTINES  
SYSIBM.SYSROUTINES\_SRC  
SYSIBM.SYSSYNONYMS  
SYSIBM.SYSTABCONST  
SYSIBM.SYSTABCONSTPKC  
SYSIBM.SYSTABLES  
SYSIBM.SYSTRIGGERS  
SYSIBM.SYSVIEWDEP  
SYSIBM.SYSVIEWS  
SYSCAT.TABLESPACES  
SYSCAT.SCHEMATA

**Db2/z** SYSIBM.SYSDUMMY1  
SYSIBM.SYSFOREIGNKEYS  
SYSIBM.SYSINDEXES  
SYSIBM.SYSKEYCOLUSE  
SYSIBM.SYSKEYS  
SYSIBM.SYSPARMS  
SYSIBM.SYSRELS  
SYSIBM.SYSROUTINES  
SYSIBM.SYSROUTINES\_SRC



SYSIBM.SYSSYNONYMS  
SYSIBM.SYSTABCONST  
SYSIBM.SYSTABCONSTPKC  
SYSIBM.SYSTABLES  
SYSIBM.SYSTRIGGERS  
SYSIBM.SYSVIEWDEP  
SYSIBM.SYSVIEWS  
SYSIBM.SYSDUMMYU  
SYSIBM.SYSPACKSTMT

## Teradata

DBC.Functions  
DBC.USERS

## 選取資料庫表格

已連線到資料來源後，可以選擇從特定的表格或視圖中匯入欄位。在「資料庫」對話框的「資料」標籤上，可以在表格名欄位中輸入表格的名稱，也可以按一下**選取**以開啟「選中表格/視圖」對話框，此對話框將列出可用的表格和視圖。

**顯示表格所有人**。如果資料來源要求在存取表格之前必須指定表格的擁有人，請選中此選項。如果資料來源沒有此要求，則取消選中此選項。

註：SAS 和 Oracle 資料庫通常會要求顯示表格所有人。

**表格與檢視畫面**。選取要匯入的表格或視圖。

**顯示**。列出目前連接到的資料來源中的欄。按一下下列選項之一可以自訂對可用表格的檢視：

- 按一下**使用者表格**檢視由資料庫使用者建立的普通資料庫表格。
- 按一下**系統表格**檢視系統擁有的資料庫表格（例如，可提供有關資料庫的資訊的表格，如指標的詳細資料）。此選項可用於檢視 Excel 資料庫中使用的標籤。（注意，也可以使用個別的 Excel 來源節點。請參閱主題第 38 頁的『Excel 來源節點』，瞭解進一步資訊。）
- 按一下**視圖**以檢視基於查詢（涉及到一個或多個普通表格）的虛擬表格。
- 按一下**同義字**檢視在資料庫中建立的任何現有表格的同義字。

**名稱/擁有人過濾器**。使用這些欄位可以按名稱或擁有人過濾顯示的表格的清單。例如，鍵入 SYS 僅列出具有該擁有人的表格。使用萬用字元搜尋時，可使用底線 (\_) 代表所有的單字元，使用百分比符號 (%) 代表以任何順序排列的零個或多個字元。

**設為預設值**。為目前使用者儲存目前設定作為預設值。當使用者將來開啟新的表格選取器對話框時（僅在資料來源名稱和使用者登入名相同的情況下），可還原使用這些設定。

## 查詢資料庫

已連線到資料來源後，可以選擇使用 SQL 查詢匯入欄位。從主對話框中，選取 **SQL 查詢**作為連線模式。這將在對話框中新增一個查詢編輯器視窗。使用查詢編輯器可建立或載入一個或多個 SQL 查詢，其結果集將被讀取到資料串流中。

如果指定了多個 SQL 查詢，請使用分號 (;) 進行分隔並確保不存在多 SELECT 陳述式。

要取消和關閉查詢編輯器視窗，可選取表格作為連線模式。

可以在 SQL 查詢中包含 SPSS Modeler 串流參數（一種使用者定義變數）。請參閱『在 SQL 查詢中使用串流參數』主題，以取得更多資訊。

**載入查詢。**按一下此選項可開啟檔案瀏覽器，使用該瀏覽器可載入先前儲存的查詢。

**儲存查詢。**按一下此選項可開啟「儲存查詢」對話框，使用該對話框可儲存目前的查詢。

**匯入預設值。**按一下此選項可匯入使用對話框中選取的表格和欄自動構建的範例 SQL SELECT 陳述式。

**清除。**清除工作區的內容。當想要重新開始編輯時，可使用此選項。

**分割文字。**預設選項從不表示查詢將完整地傳送到資料庫。另外，您還可以選擇**根據需要**，這表示 SPSS Modeler 會試圖對查詢進行剖析並確定是否存在應該逐條傳送給資料庫的 SQL 陳述式。

## 在 SQL 查詢中使用串流參數

在撰寫 SQL 查詢來匯入欄位時，可以包含之前定義的 SPSS Modeler 串流參數。受支援所有類型的串流參數。

下表顯示如何在 SQL 查詢中對串流參數的一些範例進行解釋。

表 3. 串流參數的範例.

串流參數名稱（範例）	儲存體	串流參數值	解釋為
PString	字串	ss	'ss'
PInt	整數	5	5
PReal	實數	5.5	5.5
PTime	時間(M)	23:05:01	t{'23:05:01'}
PDate	日期	2011-03-02	d{'2011-03-02'}
PTimeStamp	時間戳記	2011-03-02 23:05:01	ts{'2011-03-02 23:05:01'}
PColumn	不明	IntValue	IntValue

在 SQL 查詢中，可以採用與 CLEM 表示式中相同的方式來指定串流參數，即 '\$P-<parameter\_name>'，其中 <parameter\_name> 是為串流參數定義的名稱。

參照欄位時，儲存類型必須定義為「未知」，並且參數值必須根據需要用引號括起。因此，如果使用表格中所顯示的範例已輸入了 SQL 查詢：

```
select "IntValue" from Table1 where "IntValue" < '$P-PInt';
```

那麼此查詢將評估為：

```
select "IntValue" from Table1 where "IntValue" < 5;
```

如果要使用 PColumn 參數來參照 IntValue 欄位，那麼需要以如下方式指定查詢以取得同一結果：

```
select "IntValue" from Table1 where "'$P-PColumn'" < '$P-PInt';
```

---

## 「變數檔案」節點

可以使用「變數檔案」節點從自由欄位文字檔（其記錄包含的欄位個數不變，但包含的字元數可改變）中讀取資料，該檔案又稱為分隔文字檔。此類型的節點也可用於具有固定長度的頁首文字和特定類型的註解的檔案。每次讀取一條記錄，並將這些記錄傳送到串流中，直到讀取完整個檔案。

### 有關讀取地理空間資料的說明

如果此節點包含地理空間資料，並且此節點是作為純文字檔案的匯出而建立，那麼您必須執行一些額外的步驟以設定地理空間 meta 資料。如需相關資訊，請參閱第 25 頁的『將地理空間資料匯入到「變數檔案」節點中』。

### 有關讀取定界文字資料的說明

- 必須在每行末尾處用換行字元分隔記錄。換行字元不可轉作他用（例如，包含在任何欄位名稱或欄位值內）。最好刪除開頭和結尾處的空格以節省空間（儘管這不是必需步驟）。此節點可以選擇性地去除這些空格。
- 必須使用逗點或其他字元（最好是僅用作定界字元，即該字元不能出現在欄位名稱或欄位值中）分隔欄位。如果做不到這一點，那麼可以將所有文字欄位都使用雙引號括住，前提是所有欄位名稱或文字值均不包含雙。如果欄位名稱或欄位值包含雙，那麼可以改為將文字欄位括在單個內，前提同樣是欄位值中的其他位置未使用單個。如果單個和雙都不能使用，那麼需要對文字值進行修正以移除或取代定界字元或者單個/雙。
- 每一列（包含標頭列）都應包含相同的欄位個數。
- 第一行應包含欄位名稱。如果不是這種情況，請取消選中從檔案中讀取欄位名稱以便對每個欄位指定一個通用名稱，例如 Field1 和 Field2 等等。
- 第二行必須包含資料的第一條記錄。不得存在空白行或備註。
- 數值不能包含千位分隔字元或分組符號，例如，3,000.00 中不能使用逗點。小數指示符（美式英語或英式英語中的期間或句號）只能在適當的情況下使用。
- 日期值和時間值應該採用「串流選項」對話框中可以識別的某種格式，例如 DD/MM/YYYY 或 HH:MM:SS。檔案中的所有日期欄位和時間欄位最好採用同一種格式，並且任何包含日期的欄位內的所有值必須採用同一格式。

### 設定「變數檔案」節點的選項

請在「變數檔案」節點對話框的「檔案」標籤上設定選項。

**檔案** 指定檔案名。可以輸入檔名或按一下省略按鈕 (...) 來選取檔案。您一旦選取了檔案，檔案路徑就會顯示，並且檔案內容將與分隔符號一起顯示在下方面的窗格中。

您可以複製所顯示的來自資料來源的樣本文字，並將其貼上到下列控制項中：EOL 備註字元和使用者指定的分隔符號。使用 Ctrl-C 和 Ctrl-V 進行複製和貼上。

**從檔案中讀取檔案名** 此選項依預設處於已選取狀態，用於將資料檔案中的第一列作為欄的標籤進行處理。如果第一列不是標題，則取消選中此選項，針對資料集中的欄位數為每個欄位自動分配一個一般名稱，例如 *Field1*, *Field2*。

**指定欄位個數**。指定每個記錄中的欄位個數。只要記錄以新行結束，就可以自動偵測欄位個數。也可以手動設定字段數。

**跳過標題字元**。指定要忽略第一個記錄的開頭處的多少個字元。

**EOL 註解字元。**指定字元（例如 # 或 !）表示資料中的註解。無論這些字元之一出現在資料檔案的何處，從該字元起直到下一個新行字元（不包括）之前的所有字元都將被忽略。

**刪除前導及尾端空格。**選中選項可廢棄匯入字串中開頭和結尾的空格。

**附註。**在使用與不使用 SQL 回送的字串之間的對比可能產生存在尾部空格的不同結果。

**無效字元。**選取捨棄以刪除資料來源中的無效字元。選取取代為以將無效字元取代為指定的符號（僅限一個字元）。無效字元為無效字元或指定的編碼方法中不存在的任何字元。

**編碼。**指定使用的文字編碼方法。您可以選擇系統預設、串流預設值或 UTF-8。

- 系統預設值是在 Windows 控制台中指定，如果是分散式模式執行，則是在伺服器電腦上指定。
- 串流預設值在「串流內容」對話框中指定。

**小數符號** 請選取資料來源中使用的小數分隔字元類型。**串流預設值**是從串流內容對話框的「選項」標籤中選取的字元。否則，在此對話框中選取句號 (.) 或逗點 (,) 作為小數分隔字元讀取所有的資料。

**行分隔符號是換行字元** 要將換行字元用作行分隔符號，而非用作欄位定界字元，請選中此選項。例如，如果由於列中的定界字元數為奇數而導致換列，那麼此選項非常有用。請注意，選取此選項表示您將無法選取「定界字元」清單中的換行。

**註：** 如果選中此選項，那麼將去除資料列末尾的所有空白值。

**分隔符號。**通過使用針對此控制項列出的勾選框，可以指定哪些字元（例如逗點 (,)）定義檔案中的欄位範圍。也可以為使用多個分隔符號的記錄指定一個以上的分隔符號，例如「, |」。預設的分隔符號是逗點。

**注意：**如果還將逗點定義為小數分隔字元，那麼此處的預設設定將不起作用。如果逗點既是欄位分隔符號又是小數分隔字元，則可在分隔符號清單中選取其他。然後在輸入欄位中手動指定逗點。

選取允許多個空白分隔符號可將多個相鄰的空白分隔符號字元看作一個分隔符號。例如，如果在一個資料值之後隔四個空格又有一個資料值，則這群組資料將被看作是兩個而不是五個欄位。

**要在其中掃描欄和類型的行數** 請指定要在其中掃描所指定資料類型的行數和欄數。

**自動辨識日期與時間** 要使 IBM SPSS Modeler 能夠自動試圖將資料項目識別為日期或時間，請選中此勾選框。例如，這意味著 07-11-1965 之類的項目將被識別為日期，而 02:35:58 之類的項目將被識別為時間；然而，不明確的項目（例如 07111965 或 023558）由於數字之間沒有定界字元而將顯示為整數。

**註：**為了避免使用來自先前 IBM SPSS Modeler 版本的資料檔案時出現潛在的資料問題，依預設，對於在第 13 版以前的版本中儲存的資訊，未勾選此方框。

**將方括弧視為清單** 如果選中此勾選框，那麼會將括在左右方括弧之間的資料視為單一值，即使該內容包含逗點和雙引號之類的定界字元也是如此。例如，這可能包含兩個或三個維度地理空間資料，在這些資料中，括在方括弧內的座標作為單個清單項目進行處理。如需相關資訊，請參閱第 25 頁的『將地理空間資料匯入到「變數檔案」節點中』

**引號。** 通過使用下拉清單，可以指定匯入時如何處理單個和雙。可以選擇捨棄所有引號，選擇納入為文字將這些引號包含在欄位值內，或選擇成對捨棄比對成對引號然後刪除它們。如果引號不符合，則將收到錯誤訊息。選擇捨棄和配對並捨棄都會將欄位值（不帶引號）按一個字串儲存。

註：使用配對並捨棄時，將保留空格。使用捨棄時，將移除參照內外結尾的空格（例如：" ab c" ， "d ef " ， " gh i "' 將變成 'ab c, d ef, gh i'）。使用納入為文字時，參照將視為常態符號，因此將自然去除開頭和結尾的空格。

在此對話框中操作的任何時刻，都可按一下重新整理以從資料來源重新載入欄位。在更改到來源節點的資料連線時，或在對話框的標籤之間進行操作時，此操作都非常有用。

## 將地理空間資料匯入到「變數檔案」節點中

如果節點包含地理空間資料，作為純文字檔案的匯出而建立，並且在建立該節點的串流中使用，那麼該節點將保留地理空間 meta 資料，並且無需執行更多配置步驟。

但是，如果將該節點匯出並在另一個串流中使用，那麼地理空間清單資料將自動轉換為字串格式；您必須執行一些額外的步驟以還原清單儲存類型及相關地理空間 meta 資料。

有關清單的進一步資訊，請參閱第 9 頁的『清單儲存以及相關聯的測量層次』。

有關可以設定為地理空間 meta 資料的詳細資料的進一步資訊，請參閱第 120 頁的『地理空間測量層次』。

要設定地理空間 meta 資料，請完成下列步驟。

1. 在「變數檔案」節點的「檔案」標籤上，選中將方括弧作為清單進行處理勾選框。選中此勾選框意味著將括在左右方括弧之間的資料視為單個值，即使該內容包含逗點和雙之類的定界字元也是如此。未選中此勾選框意味著將資料作為字串儲存類型進行讀取，欄位中的所有逗點都將作為分隔符號進行處理，這將導致不正確地解釋資料結構。
2. 如果資料包含單引號或雙引號，請相應地在單引號和雙引號欄位中選中配對並廢棄選項。
3. 在「變數檔案」節點的「資料」標籤上，對於地理空間資料欄位，請選中置換勾選框，並將儲存類型由字串變更為清單。
4. 依預設，清單儲存類型設定為實數清單，並且清單欄位的底層值儲存類型設定為實數。要變更底層值儲存類型或深度，請按一下指定... 以顯示「儲存」子對話框。
5. 在「儲存」子對話框中，可以修改下列設定：
  - **儲存** 指定資料欄位的整體儲存類型。依預設，儲存類型設定為「清單」；但是，下拉清單包含所有其他儲存類型（「字串」、「整數」、「實數」、「日期」、「時間」和「時間戳記」）。如果您選取了除「清單」以外的任何儲存類型，那麼值儲存和深度選項不提供。
  - **值儲存** 指定清單中的元素的儲存類型，而不是欄位的整體儲存類型。匯入地理空間欄位時，相關的儲存類型只有「實數」和「整數」；預設設定為「實數」。
  - **深度** 指定清單欄位的深度。所需的深度取決於地理空間欄位的類型並遵循下列準則：
    - 點 - 0
    - 線串 - 1
    - 多邊形 - 1
    - 多點 - 1
    - 多線串 - 2
    - 多多邊形 - 2

註：您必須瞭解要重新轉換為清單的地理空間欄位的類型以及該類欄位的所需深度。如果未正確設定此資訊，那麼將無法使用此欄位。

6. 在「變數檔案」節點的「類型」標籤上，對於地理空間資料欄位，請確保測量 Cell 包含正確的測量層次。要變更測量層次，請在測量 Cell 中按一下指定... 以顯示「值」對話框。

- 在「值」對話框中，將顯示該清單的測量、儲存和深度。選取指定值及標籤選項，並從類型下拉清單中為測量選取正確的類型。根據類型不同，系統可能會提示您輸入更多詳細資料，例如資料是否代表 2 個或 3 個維度以及要使用的座標系統。

---

## 固定檔案節點

可以使用「固定檔案」節點從固定欄位文字檔（其欄位沒有被分隔，但開始位置相同且長度固定）中匯入資料。機器產生的資料或舊版資料通常以固定欄位格式儲存。使用固定檔案節點的「檔案」標籤，可以輕鬆地指定資料中欄的位置和長度。

### 為固定檔案節點設定選項

使用固定檔案節點的「檔案」標籤能夠將資料導入 IBM SPSS Modeler，並指定欄的位置和記錄的長度。使用位於對話框中心的資料預覽窗格，可以按一下以新增箭頭用來指定欄位間的岔斷點。

**檔案。**指定檔案的名稱。您可以輸入檔名或按一下省略符號按鈕 (...) 以選取檔案。一旦選定了一個檔案，即可顯示此檔案的路徑，並且檔案的內容將使用分隔符號分隔顯示在下方的窗格中。

資料預覽窗格可用來指定欄的位置和長度。預覽視窗頂部的尺規有助於測量變數的長度並指定變數間的斷點。通過按一下欄位上方的尺規區域可以指定岔斷點行。通過拖曳可移動岔斷點，而將其拖曳到資料預覽區域之外則可捨棄岔斷點。

- 每個斷點行會自動將一個新欄位新增到下方的欄位表格中。
- 由箭頭表示的開始位置會被自動新增到下方表格中的開始欄中。

**以行導向。**如果要跳過每個記錄末尾的新行字元，請選中此選項。

**跳過頁首行。**指定要忽略第一個記錄的開頭處的行數。這對忽略欄標題非常有用。

**記錄長度。**指定每個記錄中的字元數。

**欄位。**已為此資料檔案定義的所有欄位都在此處列出。有以下兩種定義欄位的方式：

- 使用上述資料預覽窗格交互指定欄位。
- 通過向下方的表格新增空白欄位列手動指定欄位。按一下欄位窗格右側的按鈕新增新欄位。然後在空白欄位中輸入欄位名稱、開始位置和長度。這些選項會自動在資料預覽窗格中新增箭頭，並且可以輕鬆地調整這些箭頭。

要刪除以前定義的欄位，可在清單選取該欄位，然後按一下紅色的刪除按鈕。

**開始。**指定欄位中第一個字元的位置。例如，如果記錄的第二個欄位從第十六個字元開始，則可以輸入 16 作為起點。

**長度。**為每個欄位指定最長值中的字元。該值可為下一個欄位確定截止點。

**刪除前導及尾端空格。**選中此選項以捨棄匯入時字串的開頭和結尾的空格。

**附註。**在使用與不使用 SQL 回送的字串之間的對比可能產生存在尾部空格的不同結果。

**無效字元。**選取捨棄可從資料輸入中移除無效字元。選取取代為以將無效字元取代為指定的符號（僅限一個字元）。無效字元是無效字元 (0) 或所有目前編碼中不存在的字元。

**編碼。**指定使用的文字編碼方法。您可以選擇系統預設、串流預設值或 UTF-8。

- 系統預設值是在 Windows 控制台中指定，如果是分散式模式執行，則是在伺服器電腦上指定。

- 串流預設值在「串流內容」對話框中指定。

**小數符號。**選取在資料來源中使用的小數分隔字元類型。**串流預設值**是從串流內容對話框的「選項」標籤中選取的字元。否則，在此對話框中選取句號 (.) 或逗點 (,) 作為小數分隔字元讀取所有的資料。

**自動識別日期與時間。**要使 IBM SPSS Modeler 能夠自動試圖將資料項目識別為日期或時間，請選中此勾選框。例如，這意味著 07-11-1965 之類的項目將被識別為日期，而 02:35:58 之類的項目將被識別為時間；然而，不明確的項目（例如 07111965 或 023558）由於數字之間沒有定界字元而將顯示為整數。

注意：為避免當使用來自先前 IBM SPSS Modeler 版本的資料檔案時出現潛在的資料問題，依預設，對於 13 之前的版本中儲存的資訊將不勾選此方框。

**類型的掃描行數。**指定對於指定的資料類型要掃描的行數。

在此對話框內工作時，可隨時按一下**重新整理**以重新載入資料來源中的欄位。在更改到來源節點的資料連線時，或在對話框的標籤之間進行操作時，此操作都非常有用。

---

## Statistics 檔案節點

可以使用 Statistics 檔案節點從已儲存的 IBM SPSS Statistics 檔案 (.sav 或 .zsav) 中直接讀取資料。現在可使用該格式取代 IBM SPSS Modeler 早期版本中的快取檔。如果想要匯入已儲存的快取檔，則應使用 IBM SPSS Statistics 檔案節點。

**匯入檔案。**指定檔案的名稱。您可以輸入檔名或按一下省略符號按鈕 (...) 以選取檔案。一旦選定了一個檔案，即可顯示此檔案的路徑。

**檔案已通過密碼進行加密。**如果您知道該檔案受密碼保護的，請選中此框；系統將提示您輸入密碼。如果該檔案受密碼保護的，但您未輸入密碼，那麼在試圖切換至另一標籤、重新整理資料、預覽節點內容或試圖執行包含節點的串流時，將顯示一條警告訊息。

註：受密碼受保護的檔案只能由 IBM SPSS Modeler 第 16 版或更高版本開啟。

**變數名稱。**選取從 IBM SPSS Statistics .sav 或 .zsav 檔案匯入變數名稱和標籤時的處理方法。您選擇要在這裡包括的 meta 資料會存留於 IBM SPSS Modeler 中的整個工作中，可以再次匯出以便在 IBM SPSS Statistics 中使用。

- **讀取姓名與標籤。**選中此選項可將變數名稱和標籤同時讀取入 IBM SPSS Modeler。依預設，此選項已選取，且變數名稱在「類型」節點中顯示。根據串流內容對話框中指定的選項，標籤將顯示在圖表、模型瀏覽器和其他類型的輸出中。依預設，將禁止在輸出中顯示標籤。
- **讀取用作名稱的標籤。**選取從 IBM SPSS Statistics .sav 或 .zsav 檔案中讀取描述性統計量變數標籤（而不是短欄位名稱），並在 IBM SPSS Modeler 中將這些標籤用作變數名稱。

**值。**選取從 IBM SPSS Statistics .sav 或 .zsav 檔案匯入值與標籤時的處理方法。您選擇要在這裡包括的 meta 資料會存留於 IBM SPSS Modeler 中的整個工作中，可以再次匯出以便在 IBM SPSS Statistics 中使用。

- **讀取資料與標籤。**選中此選項可將實際值和值標籤同時讀取入 IBM SPSS Modeler。依預設將已選取此選項，並且這些值本身將顯示在「類型」節點中。根據串流內容對話框中指定的選項，值標籤將顯示在表示式建置器、圖表、模型瀏覽器和其他類型的輸出中。
- **讀取用作資料的標籤。**如果要使用 .sav 或 .zsav 檔案中的值標籤而不是用於代表值的數值或符號代碼，請選中此選項。例如，對於含性別欄位（其值 1 和 2 實際上分別代表陽性和陰性）的資料，選中此選項可將該欄位轉換為字串，並將陽性和陰性作為實際值匯入。

選中此選項前考慮 IBM SPSS Statistics 資料中的遺漏值非常重要。例如，如果數值欄位僅對遺漏值使用標籤 (0 = *No Answer*, -99 = *Unknown*)，則選中上述選項將僅匯入值標籤 *No Answer* 和 *Unknown*，並會將欄位轉換為字串。在這種情況下，應在類型節點中匯入值本身並設定遺漏值。

**使用欄位格式資訊確定儲存類型。**如果取消勾選此方框，那麼將使用整數儲存匯入 *.sav* 檔案中格式化為整數的欄位值（即，在 IBM SPSS Statistics 的「變數視圖」中指定為 *Fn.0* 的欄位）。將除字串外的所有其他欄位值作為實數匯入。

如果已選取了此框（預設情況），那麼無論是否已在 *.sav* 檔案中格式化為整數，除字串以外的所有欄位值都將作為實數匯入。

**多個回應集。**匯入檔案後，IBM SPSS Statistics 檔案中定義的任何複選題分析集都將自動被保留。您可以使用「過濾器」標籤來檢視並編輯任何節點中的回應集。請參閱第 131 頁的『編輯多重回應集』主題，以取得更多資訊。

---

## Data Collection 節點

Data Collection 來源節點根據隨附於 Data Collection 產品的 Survey Reporter Developer Kit 匯入意見調查資料。此格式將案例資料（對意見調查期間所收集的問題的實際回應）與 *meta* 資料（說明如何收集和組織案例資料）進行識別。*meta* 資料包括問題文字、變數名稱和說明、多回應變數定義、字串的變換以及觀察值資料結構的定義等資訊。

註：此節點需要隨 Data Collection 產品分發的 Survey Reporter Developer Kit。除安裝 Developer Kit 外，不需要其他配置。

### 備註(O)

- 可以從一般表狀 VDATA 格式或者從階層式 HDATA 格式中的來源（如果這些來源包含 *meta* 資料來源）讀取意見調查資料。
- 通過使用 *meta* 資料資訊，可以自動實例化類型。
- 將意見調查資料匯入到 SPSS Modeler 時，問題將呈現為欄位，同時每個被意見調查者對應一個記錄。

## Data Collection 匯入檔案選項

使用 Data Collection 節點的「檔案」標籤可為要匯入的 *meta* 資料和觀察值資料指定選項。

### Metadata 設定

註：如需可用提供者檔案類型的完整清單，您需要安裝隨附於 Data Collection 軟體產品的 Survey Reporter Developer Kit。

**meta 資料提供者。**可從 Data Collection Survey Reporter Developer Kit 所受支援的多種格式匯入意見調查資料。包含下列可用的提供者類型：

- **DataCollectionMDD。**從調查表定義檔 (*.mdd*) 中讀取 *meta* 資料。這是標準的 Data Collection 資料模型格式。
- **ADO 資料庫。**從 ADO 檔案中讀取觀察值資料和 *meta* 資料。指定包含 *meta* 資料的 *.adoinfo* 檔案的名稱和位置。此 DSC 的內部名稱是 *mrADODsc*。
- **In2data 資料庫。**讀取 In2data 觀察值資料和 *meta* 資料。此 DSC 的內部名稱是 *mrI2dDsc*。
- **資料收集日誌檔。**從標準 Data Collection 日誌檔中讀取 *meta* 資料。通常，日誌檔的副檔名為 *.tmp*。但是，部分日誌檔可能有其他副檔名。必要的話，您可以重新命名檔案，讓它具有 *.tmp* 副檔名。此 DSC 的內部名稱是 *mrLogDsc*。



- **Quancept 定義檔**。將 meta 資料轉換為 Quancept Script。指定 Quancept *.qdi* 檔案的名稱。此 DSC 的內部名稱是 *mrQdiDrsDsc*。
- **Quanvert 資料庫**。讀取 Quanvert 觀察值資料和 meta 資料。指定 *.qvinfo* 或 *.pkd* 檔案的名稱和位置。此 DSC 的內部名稱是 *mrQvDsc*。
- **資料收集參與資料庫**。讀取專案的「樣本和歷程表」表格並建立與這些表格中的欄相對應的衍生種類變數。此 DSC 的內部名稱是 *mrSampleReportingMDSC*。
- **統計資料檔案**。從 IBM SPSS Statistics *.sav* 檔案中讀取觀察值資料和 meta 資料。將觀察值資料寫入 IBM SPSS Statistics *.sav* 檔案以便在 IBM SPSS Statistics 中分析。將 IBM SPSS Statistics *.sav* 檔案中的 meta 資料寫入 *.mdd* 檔案。此 DSC 的內部名稱是 *mrSavDsc*。
- **Surveycraft 檔案**。讀取 SurveyCraft 觀察值資料和 meta 資料。指定 SurveyCraft *.vq* 檔案的名稱。此 DSC 的內部名稱是 *mrSCDsc*。
- **資料收集 Script 檔**。從 *mrScriptMetadata* 檔案中讀取 meta 資料。通常，這些檔案具有 *.mdd* 或 *.dms* 副檔名。此 DSC 的內部名稱是 *mrScriptMDSC*。
- **Triple-S XML 檔案**。從 XML 格式的 Triple-S 檔案中讀取 meta 資料。此 DSC 的內部名稱是 *mrTripleSDsc*。

**Meta 資料內容**。這是選用選項，選取內容可指定要匯入的意見調查版本及要使用的語言、環境定義和標籤類型。請參閱第 30 頁的『Data Collection 匯入 meta 資料內容』主題，以取得更多資訊。

## 個案資料設定

註：如需可用提供者檔案類型的完整清單，您需要安裝隨附於 Data Collection 軟體產品的 Survey Reporter Developer Kit。

**取得個案資料設定**。僅從 *.mdd* 檔案中讀取 meta 資料時，按一下取得個案資料設定可確定哪些觀察值資料來源與選定的 meta 資料關聯，並確定存取給定的來源所需的特定設定。此選項僅用於 *.mdd* 檔案。

**觀察值資料提供者**。受支援下列提供者類型：

- **ADO 資料庫**。使用 Microsoft ADO 介面讀取觀察值資料。選取 OLE-DB UDL 作為觀察值資料類型，並在觀察值資料 UDL 欄位中指定連線字串。請參閱第 31 頁的『資料庫連線字串』主題，以取得更多資訊。此元件的內部名稱是 *mrADODsc*。
- **分隔的文字檔 (Excel)**。從以逗點分隔的 (.CSV) 檔案中讀取觀察值資料，例如可通過 Excel 輸出的檔案。內部名稱是 *mrCsvDsc*。
- **資料收集資料檔案**。從原生 Data Collection 資料格式檔案中讀取案例資料。內部名稱是 *mrDataFileDsc*。
- **In2data 資料庫**。從 In2data 資料庫 (*.i2d*) 檔案中讀取觀察值資料和 meta 資料。內部名稱是 *mrI2dDsc*。
- **資料收集日誌檔**。從標準 Data Collection 日誌檔中讀取觀察值資料。通常，日誌檔的副檔名為 *.tmp*。但是，部分日誌檔可能有其他副檔名。必要的話，您可以重新命名檔案，讓它具有 *.tmp* 副檔名。內部名稱是 *mrLogDsc*。
- **量子 資料檔案**。從任何量子 格式的 ASCII 檔案 (*.dat*) 中讀取觀察值資料。內部名稱是 *mrPunchDsc*。
- **Quancept 資料檔案**。從 Quancept *.drs*、*.drz* 或 *.dru* 檔案中讀取觀察值資料。內部名稱是 *mrQdiDrsDsc*。
- **Quanvert 資料庫**。從 Quanvert *qvinfo* 或 *.pkd* 檔案中讀取觀察值資料。內部名稱是 *mrQvDsc*。
- **資料收集資料庫 (MS SQL Server)**。從 Microsoft SQL Server 資料庫中讀取觀察值資料。請參閱第 31 頁的『資料庫連線字串』主題，以取得更多資訊。內部名稱是 *mrRdbDsc2*。
- **統計資料檔案**。從 IBM SPSS Statistics *.sav* 檔案中讀取觀察值資料。內部名稱是 *mrSavDsc*。

- **Surveycraft 檔案**。從 Surveycraft *.qdt* 檔案中讀取觀察值資料。*.vq* 檔案和 *.qdt* 檔案必須在同一個目錄下，並且都具有讀取寫入權限。依預設，這兩個由 SurveyCraft 建立的檔案位於不同目錄之下，因此需要將一個檔案移動至另一個檔案所在目錄下以便能夠匯入 SurveyCraft 資料。內部名稱是 *mrScDsc*。
- **Triple-S 資料檔案**。以長度固定的格式或逗點分隔的格式從 Triple-S 資料檔案中讀取觀察值資料。內部名稱是 *mr TripleDsc*。
- **資料收集 XML**。從 Data Collection XML 資料檔案中讀取觀察值資料。通常，此格式可用於將觀察值資料從一個位置傳輸到另一個位置。內部名稱是 *mrXmlDsc*。

**觀察值資料類型**。指定從檔案、資料夾、OLE-DB UDL 或 ODBC DSN 中讀取觀察值資料，並相應地更新對話框選項。有效選項取決於提供者的類型。對於資料庫提供者，可以為 OLE-DB 或 ODBC 連線指定選項。請參閱第 31 頁的『資料庫連線字串』主題，以取得更多資訊。

**觀察值資料專案**。從 Data Collection 資料庫中讀取觀察值資料時，可以輸入專案的名稱。對於所有其他觀察值資料類型，此設定應該留為空白。

## 變數匯入

**匯入系統變數**。指定是否匯入系統變數，其中包含表示存取狀態的變數（進行中、已完成和完成日期等）。您可以選擇 **無**、**所有** 或 **通用**。

**匯入 "Codes" 變數**。控制項匯入代表代碼（用於種類變數的開啟放式「其他」回應）的變數。

**匯入 "SourceFile" 變數**。控制項匯入包含已掃描回應影像檔名的變數。

**多回應變數匯入方式**。多回應變數可作為多旗標欄位（一個多二分法集）匯入，此方法是用於新串流的預設方法。在 12.0 之前的 IBM SPSS Modeler 版本中建立的串流已將多個回應匯入用單個欄位，並以逗點區隔多個值。仍然受支援使用舊方法，以容許現有串流按照之前的方式執行，但建議更新舊串流以使用新方法。請參閱第 31 頁的『匯入複選題分析集』主題，以取得更多資訊。

## Data Collection 匯入 meta 資料內容

匯入 Data Collection 意見調查資料時，您可以在「Meta 資料內容」對話框中指定要匯入的意見調查版本及要使用的語言、環境定義和標籤類型。請注意，一次只能匯入一種語言、環境定義和標籤類型。

**版本**。每個意見調查版本都可看作是用於收集觀察值資料特定集合的 meta 資料的一個 Snapshot。隨著調查表的變更，可建立多個調查版本。可以匯入最新版本、所有版本或特定的版本。

- **所有版本**。如果要使用所有可用版本的組合（超集），請選中此選項。（該父集有時稱作父版本）。版本之間存在衝突時，最新的版本通常優先於較早的版本。例如，如果種類標籤在所有的版本中各不相同，那麼將使用最新的版本中的文字。
- **最新版本**。如果要使用最新版本，請選中此選項。
- **指定版本**。如果要使用特定的意見調查版本，請選中此選項。

選擇所有版本非常有用，例如，當您要為一個以上的版本匯出觀察值資料，且變數和種類定義已發生變更（這意味著在一個版本中收集的觀察值資料在另一個版本中無效）時，即可選中此選項。選取要為其匯出觀察值資料的所有版本意味著，在不發生因版本間差異而導致的有效性錯誤的情況下，通常可同時匯出在不同版本中收集的觀察值資料。但是，因為版本有所變更，某些有效性錯誤仍可能發生。

**語言**。問題和關聯的文字可以多種語言儲存在 meta 資料中。對於意見調查，可使用預設語言，也可指定某種特定的語言。如果某個項目在指定的語言中不提供，那麼將使用預設語言。

**環境定義。** 選取要使用的使用者環境定義。使用者環境定義可控制項顯示哪些文字。例如，選取問題可顯示問題文字，選取分析可顯示適合在分析資料時顯示的較短文字。

**標籤類型。** 列出已定義的標籤類型。預設類型為 標籤 ，它可用於問題使用者環境定義中的問題文字和分析使用者環境定義中的變數說明。針對指導、說明等等，可定義其他標籤類型。

## 資料庫連線字串

使用 Data Collection 節點通過 OLE-DB 或 ODBC 從資料庫中匯入觀察值資料時，選取「檔案」標籤上的編輯可存取連線字串對話框，通過此對話框可以自訂傳送到提供者的連線字串以便對連線進行微調。

## 進階特性

使用 Data Collection 節點從需要明確地登入的資料庫中匯入觀察值資料時，選取進階以提供使用者 ID 和密碼來存取資料來源。

## 匯入複選題分析集

通過對每個可能的變數值使用一個個別的旗標欄位，多回應變數可作為多二分法集從 Data Collection 匯入。例如，如果要求受訪者從清單中選取他們已經參觀過的博物館，那麼該集中就會包含與每個列出的博物館一一對應的個別旗標欄位。

匯入資料後，您可以從包含「過濾器」標籤的任意節點新增或編輯多重回應集。請參閱第 131 頁的『編輯多重回應集』主題，以取得更多資訊。

## 將多個回應匯入單個欄位（適用於在前發行版中建立的串流）

在舊版本的 SPSS Modeler 中，並不是按上方所述匯入多個回應，而是將它們匯入到單一欄位中，並且用逗點分隔值。為受支援現有的串流，該方法仍舊適用，但是建議更新所有這種串流以使用新的方法。

## Data Collection欄匯入說明

Data Collection 資料中的直欄會讀入 SPSS Modeler，如下表所匯總。

表 4. Data Collection 直欄匯入摘要

Data Collection 欄類型	SPSS Modeler 儲存器	測量層級
Boolean 旗標 (yes/no)	字串	旗標 (值為 0 和 1)
種類	字串	名義(M)
日期或時間戳記	時間戳記	連續的
倍精準數 (指定範圍內的浮點值)	實數	連續的
長整數 (指定範圍內的整數值)	整數	連續的
文字 (自由文字說明)	字串	無類型
等級 (指示問題內的網格或迴圈)	不發生在 VDATA 中且不匯入到 SPSS Modeler	
物件 (二進位資料，例如顯示不規則文字的傳真或聲音記錄)	未匯入至 SPSS Modeler	
無 (未知類型)	未匯入至 SPSS Modeler	
Respondent.Serial 欄 (為每個被調查者關聯一個唯一的 ID)	整數	無類型

為避免從 meta 資料中讀取和從實際值中讀取的值標籤之間可能出現的不一致現象，可將所有 meta 資料值轉換為小寫。例如，可將值標籤 *E1720\_years* 轉換為 *e1720\_years*。

## IBM Cognos 來源節點

通過 IBM Cognos 來源節點可將 Cognos 資料庫資料或單一清單報告導入到資料採礦階段作業中。這樣，可將 Cognos 的商業智慧功能與 IBM SPSS Modeler 的預測分析功能融為一體。可以匯入關係資料、維度建模關係 (DMR) 資料和 OLAP 資料。

從 Cognos 伺服器連線中，首先選取從中匯入資料或報告的位置。位置包含 Cognos 模型以及所有與該模型關聯的資料夾、查詢、報告、視圖、快速鍵、URL 和工作定義。Cognos 模型定義商業規則、資料說明、資料關係、業務維度和階層以及其他管理作業。

如果要匯入資料，那麼可以從所選套件中選取要匯入的物件。可以匯入的物件包含查詢主旨（代表資料庫表格）或各個查詢項目（代表表格直欄）。如需相關資訊，請參閱『Cognos 物件圖示』。

如果套件定義了過濾器，那麼可以匯入一個或多個這些過濾器。如果匯入的過濾器與匯入的資料關聯，那麼該過濾器在匯入資料之前套用。要匯入的資料必須是 UTF-8 格式。

如果要匯入報告，請選取包含一個或多個報告的套件或套件內的資料夾。接著，可以選取要匯入的單個報告。只能匯入一個清單報告；不支援多個清單。

如果為資料物件或報告定義了參數，那麼您可以先指定這些參數的值，然後再匯入物件或報告。

註：Cognos 來源節點僅支援 Cognos CQM 包。不支援 DQM 包。

## Cognos 物件圖示

可從 Cognos Analytics 資料庫匯入的各種物件類型以不同的圖示代表，如下表格所述。

表 5. Cognos 物件圖示.







圖示	物件
	套件
	命名空間
	查詢主題
	查詢項目
	測量維度
	測量

表 5. Cognos 物件圖示 (繼續).

圖示	物件
	維度
	層次階層
	層次
	過濾器
	報告
	單獨計算

## 匯入 Cognos 資料

要從 IBM Cognos Analytics 資料庫匯入資料 (支援第 11 版或更高版本)，請確保 IBM Cognos 對話框的「資料」標籤上的方式設定為資料。

**連線。**按一下編輯將顯示一個對話框，可在其中定義從其匯入資料或報告的全新 Cognos 連線的詳細資料。如果已經通過 IBM SPSS Modeler 登入 Cognos 伺服器，也可編輯現行連線的明細。請參閱第 34 頁的『Cognos 連線』主題，以取得更多資訊。

**位置。**建立 Cognos 伺服器連線後，按一下此欄位旁邊的編輯可顯示從中匯入內容的可用套件清單。如需相關資訊，請參閱第 35 頁的『Cognos 位置選擇』。

**內容。**顯示已選取套件的名稱，以及和該套件關聯的命名空間。按兩下命名空間可顯示能夠匯入的物件。各個物件類型通過不同的圖示表示。如需相關資訊，請參閱第 32 頁的『Cognos 物件圖示』。

在選擇要匯入的物件時，應選中物件，並按一下兩個向右箭頭的上部，以便將物件移動到待匯入欄位窗格中。選取查詢主旨將匯入它的所有查詢項目。按兩下查詢主旨將展開該主旨，以便您可以選擇它的一個或多個查詢項目。通過按住 Ctrl 並按一下 (選擇個別項目)、按住 Shift 並按一下 (選擇一區塊項目) 以及按住 Ctrl 和 A (選擇所有項目)，可以執行多種選擇。

在選擇要套用的過濾器時 (如果套件定義了過濾器)，應在「內容」窗格中導覽至過濾器，選中過濾器並按一下兩個向右箭頭的下部，以便將過濾器移動到需套用的過濾器窗格中。通過按住 Ctrl 並按一下 (選擇個別過濾器) 和按住 Shift 並按一下 (選擇一區塊過濾器)，可以執行多種選擇。

**待匯入欄位。**列出您已選定要匯入 IBM SPSS Modeler 以供正在處理的資料庫物件。如果不再需要一個特定物件，請將其選中並按一下向左箭頭以將其移回到內容窗格中。您可以執行多重選項，方式與針對內容的執行方式相同。

**要套用的過濾器。** 列出您已選定要在匯入資料之前套用的過濾器。如果不再需要一個特定過濾器，請將其選中並按一下向左箭頭以將其移回到內容窗格中。您可以執行多重選項，方式與針對內容的執行方式相同。

**參數。** 如果啟用此按鈕，則會為選取的物件定義參數。匯入資料前，您可以使用參數來進行調整（例如：執行參數化計算）。如果參數已定義但未提供任何預設值，按鈕就會顯示三角形的警告標誌。按一下該按鈕以顯示參數，並選擇性地編輯參數。如果已停用按鈕，則表示報告未定義參數。

在匯入前聚集資料。如果要匯入聚集資料而不是原始資料，勾選此方框。

## 匯入 Cognos 報告

要從 IBM Cognos 資料庫匯入預先定義報告，請確保 IBM Cognos 對話框的「資料」標籤上的方式設定為報告。只能匯入一個清單報告；不支援多個清單。

**連線。** 按一下編輯將顯示一個對話框，可在其中定義從其匯入資料或報告的全新 Cognos 連線的詳細資料。如果已經通過 IBM SPSS Modeler 登入 Cognos 伺服器，也可編輯現行連線的明細。請參閱『Cognos 連線』主題，以取得更多資訊。

**位置。** 建立 Cognos 伺服器連線後，按一下此欄位旁邊的編輯可顯示從中匯入內容的可用套件清單。如需相關資訊，請參閱第 35 頁的『Cognos 位置選擇』。

**內容。** 顯示包含報告的選定套件或資料夾的名稱。導覽到相應的報告，選中它並按一下向右箭頭以將其移動到待匯入報告欄位中。

**要匯入的報告。** 指示您已選定要匯入 IBM SPSS Modeler 的報告。如果不再需要該報告，請將其選中並按一下向左箭頭以將其移回到內容窗格中，或者將其他報告移動到該欄位中。

**參數。** 如果此按鈕處於啟用狀態，那麼選定報告定義了參數。可以先使用參數來進行調整，然後再匯入報告（例如，指定報告資料的開始日期和結束日期）。如果參數已定義但未提供任何預設值，按鈕就會顯示三角形的警告標誌。按一下該按鈕以顯示參數，並選擇性地編輯參數。如果已停用按鈕，則表示報告未定義參數。

## Cognos 連線

在「Cognos 連線」對話框中，您可以選取在其中匯入或匯出資料庫物件的 Cognos Analytics 伺服器（支援第 11 版或更高版本）。

**Cognos 伺服器 URL** 輸入要在其中進行匯入或匯出的 Cognos Analytics 伺服器的 URL。這是 Cognos 伺服器上 IBM Cognos 配置的「外部分派器 URI」環境內容的值。如果不確定要使用哪個 URL，請與 Cognos 系統管理者聯絡。

**方式** 如果您希望使用特定的 Cognos 命名空間、使用者名稱和密碼（例如，以管理者身分）登入，請選取設定認證。選取使用匿名連線登入而不使用使用者認證，在這種情況您不需要填寫其他欄位。

另外，如果您具有儲存在 IBM SPSS Collaboration and Deployment Services 儲存庫中的 IBM Cognos 認證，那麼可以使用此認證，而不是輸入使用者名稱和密碼資訊或建立匿名連線。要使用現有認證，請選取儲存的認證，然後輸入認證名稱或瀏覽認證名稱。

Cognos 命名空間由 IBM SPSS Collaboration and Deployment Services 中的某個網域進行建模。

**命名空間 ID** 指定用於登入伺服器的 Cognos 安全鑑別提供者。驗證提供者是用來定義並維持使用者、群組和角色，並控制驗證程序。請注意，這是命名空間 ID，而不是命名空間名稱（ID 並非始終與名稱相同）。

**使用者名稱** 輸入用於登入伺服器的 Cognos 使用者名稱。

密碼 輸入與指定使用者名關聯的密碼。

**儲存為預設值** 按一下此按鈕可以將這些設定儲存為預設值，以避免在您每次開啟節點時重新輸入這些設定。

## Cognos 位置選擇

使用「指定位置」對話框可以選取要從中匯入資料的 Cognos 套件，或要從中匯入報告的套件或資料夾。

**公用資料夾。**如果要匯入資料，這將列出選定伺服器上的可用套件和資料夾。選取希望使用的套件，然後按一下**確定**。對於每個 Cognos 來源節點，只能選取一個套件。

如果要匯入報告，這將列出選定伺服器上可用的包含報告的資料夾和套件。選取套件或報告資料夾並按一下**確定**。雖然報告資料夾可以包含其他報告資料夾和各個報告，但是對於每個 Cognos 來源節點，只能選取一個套件或報告資料夾。

## 指定資料或報表的參數

如果在 Cognos Analytics 中定義了資料物件或報告的參數，則您可以先指定這些參數的值，然後再匯入物件或報告。報表內容的啟動日期與結止日期就是其中一個報表參數的例子。

**名稱。** 參數名稱，與在 Cognos 資料庫中指定的名稱相同。

**類型。** 參數說明。

**值。** 指定給參數的值。若要輸入或編輯值，請在表格中按兩下該值的儲存格。未在此處對值進行驗證，因此，執行時期將偵測到任何無效值。

**自動移除表格中無效的參數。** 此選項預設為已選取，並且會移除任何在資料物件或報表中發現的無效參數。

---

## IBM Cognos TM1 來源節點

IBM Cognos TM1 來源節點可讓您將 Cognos TM1 資料帶入資料採礦階段作業。如此一來，您可以將 Cognos 的企業規劃功能與 IBM SPSS Modeler 的預測分析功能結合起來。您可以匯入壓縮版本的多維度 OLAP 多維度報表資料。

**註：**TM1 使用者需要下列許可權：多維度報表的寫入權、維度的讀取權及維度元素的寫入權。此外，您需要 IBM Cognos TM1 10.2 Fix Pack 3 或更新版本，然後 SPSS Modeler 才能匯入及匯出 Cognos TM1 資料。基於舊版的現有串流仍將發揮作用。

此節點不需要管理者認證。但如果您仍然使用舊式的 17.1 以前的 TM1 節點，則仍需要管理者認證。

您需要在匯入資料之前修改 TM1 中的資料。

**註：**要匯入的資料必須是 UTF-8 格式。

從 IBM Cognos TM1 管理主機連線，先選取要從中匯入資料的 TM1 伺服器；一台伺服器包含一個或多個 TM1 資料方塊。然後選取需要的多維度報表，在多維度報表中，您選取要匯入的直欄和列。

**註：**您必須先在 tm1s.cfg 檔案（這是 TM1 伺服器的根目錄中的 TM1 伺服器配置檔案）中驗證部分設定，然後才能使用 SPSS Modeler 中的 TM1「來源」或「匯出」節點。

- HTTPPortNumber - 設定有效的埠號；一般為 1-65535。請注意，這不是您之後在節點的連線中指定的埠號；它是 TM1 使用的依預設停用的內部埠。如有需要，請聯絡 TM1 管理者來確認此埠的有效設定。

- UseSSL - 如果您將此選項設為 *True*，則會使用 HTTPS 作為傳輸通訊協定。在此情況下，您必須將 TM1 憑證匯入到 SPSS Modeler Server JRE。

## 匯入 IBM Cognos TM1 資料

要從 IBM Cognos TM1 資料庫中匯入資料，請在 IBM Cognos TM1 對話框的「資料」標籤上選取相關的 TM1 管理主機和關聯的伺服器、資料方塊和資料詳細資料。

註：在匯入資料之前，您必須在 TM1 內執行部分前處理，才能確保 IBM SPSS Modeler 可以辨識出資料的格式。這包括使用子集編輯器過濾資料，來檢視正確的匯入大小和形狀。

請注意，從 TM1 中匯入的零 (0) 值將被視為「空」值 (TM1 無法區分空白值與零值)。另請注意，一般維度中的非數值資料 (或 meta 資料) 可以匯入至 IBM SPSS Modeler。但目前不支援匯入非數值測量。

**管理主機。** 輸入要連接的 TM1 伺服器將要安裝到的管理主機 URL。對於所有 TM1 伺服器，管理主機均定義為單一 URL。可以從這個 URL 探索及存取您的環境中已安裝及正在執行的所有 IBM Cognos TM1 伺服器。

**TM1 伺服器。** 與管理主機建立連線後，請選取包含所要匯入的資料的伺服器，然後按一下**登入**。如果您先前不曾連接至此伺服器，則系統會提示您輸入**使用者名稱及密碼**；或者，您可以搜尋另存為儲存的認證的之前輸入的登入詳細資料。

**選取要匯入的 TM1 資料方塊視圖。** 顯示可以從中匯入資料的 TM1 伺服器內資料方塊的名稱。按兩下多維度報表可顯示您可以匯入的視圖資料。

註：

只有包含維度的多維度報表可以匯入至 IBM SPSS Modeler。

如果已在 TM1 資料方塊中定義某個元素的別名 (例如，值 23277 具有別名 Sales)，則將匯入值而不是別名。

若要選擇要匯入的資料，請選取視圖，然後按一下向右箭頭，將其移至**要匯入的視圖**窗格中。如果看不到所需要的視圖，請按兩下多維度報表來展開其視圖清單。可以選取公用或專用視圖。

**列維度。** 列出您選擇匯入的資料中列維度的名稱。捲動整個層次清單並選取您需要的層次。

**欄維度。** 列出您已選擇匯入的資料中欄維度的名稱。捲動整個層次清單並選取您需要的層次。

**環境定義維度。** 僅用於顯示。顯示與所選取直欄和列相關的環境定義維度。

---

## TWC 來源節點

TWC 來源節點將匯入來自 The Weather Company (一家 IBM 企業) 的天氣資料。您可以使用它來取得某個位置的歷程或預測天氣資料。這可以協助您使用最準確且精確的可用天氣資料來制定受天氣驅動的商業解決方案，以做出更好的決策。

通過該節點，可以輸入天氣相關的資料，例如，latitude、longitude、time、day\_ind (指示夜間或白天)、temp、dewpt (露點)、rh (相對濕度)、feels\_like 氣溫、heat\_index、wc (風寒)、wx\_phrase (多雲、局部多雲等)、pressure、clds (雲)、vis (能見度)、wspd (風速)、gust、wdir (風向)、wdir\_cardinal (NW、NNW、N 等)、uv\_index (紫外線指數) 和 uv\_desc (低、高等)。

TWC 來源節點使用下列 API：

- TWC Historical Observations Airport (<http://goo.gl/DplOKj>)，顯示歷程天氣資料



- TWC Hourly Forecast (<http://goo.gl/IJhhvZ>)，顯示預測天氣資料

## 位置

緯度。以格式 [-90.0~90.0] 輸入想要獲取其天氣資料的位置的緯度值。

經度。以格式 [-180.0~180.0] 輸入想要獲取其天氣資料的位置的經度值。

## 細項

**授權金鑰。**需要授權金鑰。輸入您從 The Weather Company 獲取的授權金鑰。如果您沒有金鑰，請聯絡管理者或 IBM 代表。

作為將金鑰發放給所有使用者的替代方法，管理者可能會在 IBM SPSS Modeler Server 上的新 config.cfg 檔案中包含指定金鑰，在此情況下，您可以將該欄位留空白。如果在兩個位置中都進行了指定，該對話框中的金鑰具有優先權。管理者注意：要在伺服器上新增授權金鑰，建立名為 config.cfg 的新檔案，其中的內容是 LicenseKey=<LICENSEKEY>（其中，<LICENSEKEY> 是授權金鑰），該檔案位置為 <ModelerServerInstallation>\ext\bin\pasw.twcdata。

**單位。** 選取要使用的測量單位：**English**、**Metric** 或 **Hybrid**。預設值為 **Metric**。

## 資料類型

**歷程。**如果想要匯入歷程天氣資料，請選取**歷程**，然後指定 YYYYMMDD 格式的開始和結束日期（例如，20120101 表示 2012 年 1 月 1 日）。

**預測。**如果想要匯入預測天氣資料，請選取**預測**，然後指定要預測的小時數。

---

## SAS 來源節點

SPSS Modeler Professional 和 SPSS Modeler Premium 中提供了此功能。

使用 SAS 來源節點可以將 SAS 資料導入資料挖掘階段作業中。可以匯入以下四種類型檔案：

- 適用於 Windows/OS2 的 SAS (.sd2)
- 適用於 UNIX 的 SAS (.ssd)
- SAS Transport File (.tpt)
- SAS 版本 7/8/9 (.sas7bdat)

匯入資料時，將保留所有變數，並且不會變更任何變數類型。將選定全部觀察值。

### 為 SAS 來源節點設定選項

**匯入。**選取要傳輸的 SAS 檔案的類型。選擇適用於 **Windows/OS2** 的 **SAS (.sd2)**、適用於 **UNIX** 的 **SAS (.SSD)**、**SAS Transport File (.tpt)** 或 **SAS 版本 7/8/9 (.sas7bdat)**。

**匯入檔案。**指定檔案的名稱。您可以輸入檔名或按一下省略符號按鈕 (...) 以瀏覽至檔案的位置。

**會員。**從上方選定的 SAS 傳輸檔案中選取要匯入的會員。可以輸入成員名稱或按一下**選取**瀏覽檔案中的所有會員。

從 **SAS 資料檔案讀取使用者格式**。選中此選項以讀取使用者格式。SAS 檔案將資料和資料格式（例如變數標籤）儲存在不同的檔案中。在很多時候，可能希望同時匯入格式。不過，如果擁有的資料集較大，則可能希望取消選中此選項以節省記憶體。

**格式檔案**。如果需要格式化檔案，請啟動此文字框。您可以輸入檔名或按一下省略符號按鈕 (...) 以瀏覽至檔案的位置。

**變數名稱**。選取從 SAS 檔案中匯入時所使用的處理變數名稱和標籤的方法。選擇在此處包含的 meta 資料會保留在 IBM SPSS Modeler 的整個過程中，並且可以再次匯出以在 SAS 中使用。

- **讀取姓名與標籤**。選中此選項將變數名稱和標籤同時讀取入 IBM SPSS Modeler。依預設，此選項已選取，且變數名稱在「類型」節點中顯示。根據串流內容對話框中指定的選項，標籤將顯示在表示式建置器、圖表、模型瀏覽器和其他類型的輸出中。
- **讀取用作名稱的標籤**。選取從 SAS 檔案中讀取說明性的變數標籤而不是短欄位名稱，並將這些標籤作為變數名稱在 IBM SPSS Modeler 中使用。

---

## Excel 來源節點

使用 Excel 來源節點可從 Microsoft Excel 以 .xlsx 檔案格式匯入資料。

**檔案類型**。選取要匯入的 Excel 檔案類型。

**匯入檔案**。指定要匯入的試算表檔案的名稱和位置。

**使用已命名範圍**。選中此選項可以指定在 Excel 工作表中定義的單元的指定範圍。按一下省略按鈕 (...) 從可用範圍清單中進行選擇。如果使用指定範圍，則其他工作表和資料範圍設定將不再可用並最終被停用。

**選擇工作表**。依索引或者依名稱指定要匯入的工作表。

- **依據索引**。指定要匯入的工作表的指標值，開頭的 0 表示第一個工作表，1 表示第二個工作表，依此類推。
- **依名稱**。指定要匯入的工作表的名稱。按一下省略按鈕 (...) 從可用工作表清單中進行選擇。

**工作表中的範圍**。可以第一個非空白列作為開始匯入資料，也可通過單元的明確地範圍匯入資料。

- **範圍從第一個非空白的列啟動**。找到第一個非空白 Cell，並將此 Cell 作為資料範圍的左上角 Cell。
- **單元的明確地範圍**。選中此選項可按列和欄指定明確地範圍。例如，要指定 Excel 範圍 A1:D5，您可以在第一個欄位中輸入 A1，在第二個欄位中輸入 D5，（或，R1C1 和 R5C4）。指定範圍內的所有列都將傳回，包含空白列。

**在空白列中**。如果遇到多個空白列，則可選擇**停止讀取**，或選擇**傳回空白列**以繼續讀取所有資料（包含空白列）直到工作表的末尾。

**第一列包含直欄名稱**。表示指定範圍中的第一列應作為欄位（欄）名使用。如果未選取此選項，那麼將自動產生欄位名稱。

## 欄位儲存和測量層次

從 Excel 中讀數值時，依預設將按連續的測量層次讀取以數值型儲存的欄位，按列名讀取以字串儲存的欄位。可以在「類型」標籤上手動變更測量層次（連續和列名），但儲存類型是自動決定的（雖然必要時可在充填節點或衍生節點中使用轉換函數，例如 to\_integer，來變更此類型）。請參閱第 7 頁的『設定欄位儲存類型和格式』主題，以取得更多資訊。

依預設，將按數值類型讀取以數值和字串值混合儲存的欄位，這意味著在 IBM SPSS Modeler 中所有字串值都將被設定為無效（系統遺漏值）值。這是因為與 Excel 不同，IBM SPSS Modeler 不容許欄位中存在混合的儲存類型。要避免此問題，可以在 Excel 試算表中手動將 Cell 格式設定為文字，這樣將按字串讀取所有的值（包括數字）。

---

## XML 來源節點

SPSS Modeler Professional 和 SPSS Modeler Premium 中提供了此功能。

使用 XML 來源節點可將 XML 格式檔案中的資料匯入到 IBM SPSS Modeler 串流中。XML 是用於資料交換的標準語言，對多數組織而言，它是用於此目的所選取的格式。例如，政府稅務機構可能希望分析線上提交的納稅申報單中的資料，並且這些申報單的資料處於 XML 格式（請參閱 <http://www.w3.org/standards/xml/>）。

通過將 XML 資料匯入 IBM SPSS Modeler 串流，允許您針對資料來源執行多種預測分析功能。系統將 XML 資料剖析為表格式，在此格式中，欄對應於 XML 元素和屬性的不同巢狀層次。XML 項目以 XPath 格式進行顯示（參閱 <http://www.w3.org/TR/xpath20/>）。

**重要：**「XML 來源」節點不會考慮命名空間宣告。因此，（例如）您的 XML 檔案不能在 name 標籤中包含冒號（:）字元。如果包含，在執行期間將會收到與無效字元有關的錯誤。

**讀取單個檔案。**依預設，SPSS Modeler 讀取您在 **XML 資料來源**欄位中指定的單個檔案。

**讀取目錄中的所有 XML 檔案。**如果您要讀取某個特定目錄中的所有 XML 檔案，請選擇此選項。在顯示的目錄欄位中指定位置。選取包含子目錄勾選框，以另外讀取指定目錄的所有子目錄中的 XML 檔案。

**XML 資料來源。**輸入您要匯入的 XML 來源檔案的完整路徑和檔案名稱，或使用「瀏覽」按鈕尋找檔案。

**XML 綱目。**（選用）指定您要從中讀取 XML 結構的 XSD 或 DTD 檔案的完整路徑和檔案名稱，或使用「瀏覽」按鈕尋找此檔案。如果您保留此欄位為空白，那麼將從 XML 來源檔案中讀取結構。XSD 或 DTD 檔案可以具有多個根元素。在這種情況下，將光標移動到另一欄位時，將顯示一個對話框，您可以在此對話框中選擇要使用的根元素。請參閱第 40 頁的『從多個根元素中選取』主題，以取得更多資訊。

**註：**SPSS Modeler 將忽略 XSD 指示符

**XML 結構。**顯示 XML 來源檔案（或架構，如果您在 **XML 綱目**欄位中進行了指定）結構的層次結構樹狀結構。要定義記錄界限，選取某個元素，並按一下右方向箭頭按鈕將此項目複製到記錄欄位。

**顯示屬性。**在 **XML 結構**欄位中顯示或隱藏 XML 元素的屬性。

**記錄（XPath 表示式）。**顯示從 XML 結構欄位複製的元素的 XPath 語法。隨後，此元素將在 XML 結構中強調顯示，並定義記錄界限。每次在來源檔案中遇到此元素，都會建立一個新記錄。如果此欄位為空白，那麼將使用根元素下的第一個子元素作為記錄界限。

**讀取所有資料。**依預設，來源檔案中的所有資料都將讀取到串流中。

**指定要讀取的資料。**如果您要匯入個別元素和/或屬性，請選擇此選項。選擇此選項將啟用「欄位」表格，您可以在該表格中指定要匯入的資料。

**欄位。**如果您選取了**指定要讀取的資料**選項，此表格將列出選取用於匯入的元素與屬性。可以直接在 XPath 欄中輸入元素或屬性的 XPath 語法，也可以在 XML 結構中選取元素或屬性，然後按一下右箭頭按鈕將其複製到表格中。要複製元素的所有子元素和屬性，請在 XML 結構中選中此元素並按一下雙向箭頭按鈕。

- **XPath**。要匯入的項目的 XPath 語法。
- **位置**。要匯入的項目在 XML 結構中的位置。**固定路徑**顯示相對於在 XML 結構中強調顯示的元素（或者如果沒有強調顯示的元素，則為根下面的第一個子元素）的項目路徑。**任何位置**表示在 XML 結構中任何位置上給定名稱的項目。如果您在 XPath 欄中直接輸入位置，則顯示自訂。

## 從多個根元素中選取

格式正確的 XML 檔案僅能具有單一根元素，而 XSD 或 DTD 檔案可包含多個根元素。如果其中一個根元素與 XML 來源檔案中的根元素相符，那麼將使用此根元素，否則，您需要選取一個要使用的根元素。

選擇要顯示的根元素。選取要使用的根元素。預設根元素是 XSD 或 DTD 結構中的第一個根元素。

## 從 XML 來源資料中移除不需要的空格

XML 來源資料中的換行可以通過 [CR][LF] 字元組合實現。在某些情況下，這些換行可以出現在字串中間，例如：

```
<description>An in-depth look at creating applications[CR][LF]with XML.</description>
```

在某些程序（例如，Web 瀏覽器）中開啟此檔案時，這些換行可能不可見。但是，通過 XML 來源節點將資料讀取到串流後，換行將轉換為一系列空格字元。

可以使用「過濾器」節點移除這些不需要的空格來進行更正：

下面是如何實現此操作的範例：

1. 將「填入器」節點附加到 XML 來源節點。
2. 開啟「填入器」節點並使用欄位選擇器選擇帶有不需空格的欄位。
3. 將取代設定為基於條件，並將條件設定為 **true**。
4. 在取代為欄位中，輸入 `replace("","",&FIELD)` 並按一下「確定」。
5. 將「表格」節點附加到「過濾器」節點，然後執行串流。

在「表格」節點的輸出中，此時顯示的文字不包含額外的空格。

---

## 使用者輸入節點

「使用者輸入」節點提供了建立綜合資料的簡便方式 - 從頭啟動建立綜合資料，或通過更改現有的資料建立綜合資料。例如，當您要為建模建立一個測試資料集時，此項很有用。

### 從零啟動建立資料

可在來源選用區中找到使用者輸入節點，並將此節點直接新增到串流畫布中。

1. 按一下節點選用區的來源標籤。
2. 使用拖放或按兩下操作將使用者輸入節點新增到串流畫布中。
3. 按兩下以開啟此節點的對話框並指定欄位和值。

注意：在「來源」選用區中選取的「使用者輸入」節點將是完全空白的，不含欄位也不含任何資料資訊。您可以完全從頭啟動建立綜合資料。

### 從現有的資料來源產生資料

還可以從串流的任何非終端節點產生使用者輸入節點：

1. 確定要在串流的哪個點上取代節點。
2. 用滑鼠右鍵按一下可將其資料提供給「使用者輸入」節點的節點，然後從功能表中選擇產生使用者輸入節點。
3. 此時將出現「使用者輸入」節點，其所有下游過程都將附加到此節點上，從而可取代資料串流的該點上現有的節點。此節點產生時可從 meta 資料中繼承所有的資料結構和欄位類型資訊（如果可用）。

注意：如果資料尚未在串流的所有節點中從頭到尾地執行，那麼這些節點未完全實例化，這表示當使用「使用者輸入」節點取代原來的節點時，儲存類型和資料值可能不可用。

## 為使用者輸入節點設定選項

通過使用「使用者輸入」節點對話框中包含的幾個工具，可為綜合資料輸入值並定義資料結構。對於產生的節點，「資料」標籤上的表格包含來自原始資料來源的欄位名稱。對於從「來源」選用區中新增的節點，該表格是空白的。通過使用表格中的選項可執行下列作業：

- 使用表格右側的「新增新欄位」按鈕新增新的欄位。
- 更名現有欄位。
- 為每個欄位指定資料儲存類型。
- 指定值。
- 在顯示畫面上變更欄位的順序。

### 輸入資料

可使用表格右側的值選取器按鈕從原始資料集中為每個欄位指定值或插入值。有關指定值的詳細資訊，請參閱下方說明的規則。也可以選擇將欄位留為空白 - 留為空白的欄位將填入系統空白值 (\$null\$)。

要指定字串值，僅需要在「值」欄中輸入這些值並以空格進行分隔：

Fred Ethel Martin

含有空格的字串可以用雙引號括起來：

"Bill Smith" "Fred Martin" "Jack Jones"

針對數值欄位，您可以相同的方式（用空格區隔列出）輸入多個值：

10 12 14 16 18 20

或者您可以透過設定限制 (10, 20) 和之間的步長 (2) 指定相同的數列。使用此方法，您會鍵入：

10,20,2

這兩個方法可以透過彼此內嵌而結合使用，例如：

1 5 7 10,20,2 21 23

此項目將產生下列值：

1 5 7 10 12 14 16 18 20 21 23

使用在「串流內容」對話框中選定的目前預設形式已輸入日期值和時間值。

11:04:00 11:05:00 11:06:00

2007-03-14 2007-03-15 2007-03-16

timestamp 值既包含日期元件又包含時間元件，所以必須對其使用雙：

"2007-03-14 11:04:00" "2007-03-14 11:05:00" "2007-03-14 11:06:00"

有關其他詳細資訊，請參閱下方面資料儲存的相關備註。

**產生資料。**通過此選項可指定執行串流時產生記錄的方式。

- **所有組合。**產生包含欄位值的各種可能組合的記錄，此時每個欄位值將出現在幾個記錄中。選中此選項有時可使產生的資料比希望產生的更多，所以通常可能要在此節點後附加一個樣本節點。
- **依照順序。**按指定的資料欄位值的順序產生記錄。每個欄位值僅出現在一個記錄中。記錄的總數與單個欄位值的最大數相等。如果欄位包含的記錄數小於最大記錄數，則插入未定義的 (\$null\$) 值。

顯示範例

例如，下列項目將產生下列兩個表格範例中列出的記錄。

- **年齡。** 30,60,10
- **血壓。** LOW
- **膽固醇。** NORMAL HIGH
- **藥品。** (留空白)

表 6. 產生資料欄位設為「所有組合」。

年齡	BP	膽固醇	藥物
30	低	NORMAL	\$null\$
30	低	高	\$null\$
40	低	NORMAL	\$null\$
40	低	高	\$null\$
50	低	NORMAL	\$null\$
50	低	高	\$null\$
60	低	NORMAL	\$null\$
60	低	高	\$null\$

表 7. 產生資料欄位設為「依照順序」。

年齡	BP	膽固醇	藥物
30	低	NORMAL	\$null\$
40	\$null\$	高	\$null\$
50	\$null\$	\$null\$	\$null\$
60	\$null\$	\$null\$	\$null\$

資料儲存

儲存體說明資料儲存在欄位中的方式。例如，值為 1 和 0 的欄位會儲存整數資料。它與測量層次不同，說明資料使用情形，並不會影響儲存體。例如，您可能想要使用值 1 及 0，將整數欄位的測量層次設為旗標。這通常指出 1 = *True* 及 0 = *False*。儲存格式必須在數據來源中確定，而測量層次可以使用「類型」節點在串流中的任意點上進行變更。請參閱第 118 頁的『測量層次』主題，以取得更多資訊。

可用儲存類型有：

- **字串** 用於包含非數值資料（也稱為英數資料）的欄位。字串可以包含任何字元序列，比如 *fred*、*Class 2* 或 *1234*。注意：字串中的數字不能用於計算。
- **整數** 值為整數的欄位。

- **實數** 值為可能包含小數（不限於整數）的數字。顯示格式在「串流內容」對話框中指定，並且可以被「類型」節點（「格式」標籤）中的各個欄位覆蓋。
- **日期** 以標準格式指定的日期值，例如年月日（例如 2007-09-26）。具體格式在「串流內容」對話框中指定。
- **時間** 以持續時間形式測量的時間。例如，某個服務電話持續 1 小時 26 分 38 秒，該時間可以根據「串流內容」對話框中指定的目前時間格式代表為：01:26:38。
- **時間戳記** 同時包含日期與時間元件的值，例如 2007-09-26 09:04:00，具體同樣取決於「串流內容」對話框中的現行日期和時間格式。請注意，需要用雙引號將時間戳記值括起來，以確保將此值解釋為單一值而非個別日期與時間值。（同樣適用於在使用者輸入節點中輸入值時的情況。）
- **清單** 在 SPSS Modeler 第 17 版中，隨新測量層次「地理空間」和「收集」一起引入了「清單」儲存欄位，對於單個記錄，此欄位包含多個值。存在所有其他儲存類型的清單版本。

表 8. 清單儲存類型圖示




圖示	儲存體類型
[A]	字串清單
[I]	整數清單
[D]	實數清單
[T]	時間清單
[D]	日期清單
[T]	時間戳記清單
[L]	深度大於零的清單

另外，為了與「收集」測量層次配合使用，提供了下列測量層次的清單版本。

表 9. 清單測量層次圖示

圖示	測量層級
[C]	連續清單
[C]	種類清單

表 9. 清單測量層次圖示 (繼續)

圖示	測量層級
	旗標清單
	名義清單
	序數清單

可以通過三個來源節點 (Analytic Server、「地理空間」或「變數檔案」) 中的某一個將清單匯入到 SPSS Modeler 中，也可以在串流中使用「衍生」或「填入器」欄位作業節點建立清單。

有關「清單」及其與「收集」和「地理空間」測量層級的互動的進一步資訊，請參閱第 9 頁的『清單儲存以及相關聯的測量層次』。

**儲存轉換。**用戶可使用各式各樣的轉換函數來轉換某個欄位的儲存格式，比如「填入器」節點中的 `to_string` 和 `to_integer`。請參閱第 138 頁的『使用填入節點進行儲存類型轉換』主題，以取得更多資訊。請注意，轉換函數 (以及需要指定類型輸入 (如日期或時間值) 的任何其他函數) 取決於「串流內容」對話框中指定的現行形式。例如，如果想將值為 *Jan 2003*、*Mar 2003* 等字串欄位轉換為日期儲存格式，請選取 **MON YYYY** 作為串流的預設日期格式。轉換函數也可從「衍生」節點取得，用於在衍生計算期間進行暫時轉換。另外，還可以使用「衍生」節點來執行其他操作，比如使用種類值對字串欄位進行重新編碼。請參閱第 137 頁的『使用衍生節點對值進行重新編碼』主題，以取得更多資訊。

**混合資料中的讀數。**請注意，讀取數值儲存格式 (整數、實數、時間、時間戳記或日期) 的欄位中的讀數時，任何非數值將被設定為無效或系統遺漏值。這時因為 IBM SPSS Modeler 與某些應用程式不同，它不容許欄位中含有混合儲存類型。為了避免出現混合儲存類型，必須根據需要更改來源節點中或外部應用程式中的儲存類型，從而將任何具有混合資料的欄位以字串的格式讀取入。

注意：在產生的「使用者輸入」節點中可能已包含了從來源節點 (如果已實例化) 獲取的儲存類型資訊。未實例化的節點不包含儲存類型資訊或使用類型資訊。

用於指定值的規則

對於符號欄位，應在多個值之間保留空格，例如：

HIGH MEDIUM LOW

針對數值欄位，您可以相同的方式 (用空格區隔列出) 輸入多個值：

10 12 14 16 18 20

或者您可以透過設定限制 (10, 20) 和之間的步長 (2) 指定相同的數列。使用此方法，您會鍵入：

10,20,2

這兩個方法可以透過彼此內嵌而結合使用，例如：

1 5 7 10,20,2 21 23



此項目將產生下列值：

1 5 7 10 12 14 16 18 20 21 23

## 「模擬產生」節點

「模擬產生」節點提供了一種產生模擬資料的簡單方法，即在沒有歷程資料的情況下，使用使用者指定的統計分佈產生資料；或者使用對現有歷程資料執行「模擬配適」節點而獲取的分佈自動產生資料。如果輸型輸入中存在不確定性，那麼在您要對預測模型的結果進行評估時，生成模擬資料非常有用。

在沒有歷程資料的情況下建立資料

在「來源」選用區中找到「模擬產生」節點，並且可以將其直接新增到串流畫布中。

1. 按一下節點選用區的來源標籤。
2. 使用拖放或按兩下操作將「模擬產生」節點新增到串流畫布中。
3. 按兩下以開啟此節點的對話框，並指定欄位、儲存類型、統計分佈和分佈參數。

註：在「來源」選用區中選取的「模擬產生」節點將是完全空白的，不包含任何欄位和任何分佈資訊。這使您能夠在完全沒有歷程資料的情況下建立模擬資料。

使用現有歷程資料創建模擬資料

還可以通過執行「模擬配適」終端節點來建立「模擬產生」節點：

1. 用滑鼠右鍵按一下「模擬配適」節點，然後從功能表中選擇執行。
2. 「模擬產生」節點將顯示在包含指向「模擬配適」節點的更新鏈結的串流畫布中。
3. 產生「模擬產生」節點後，此節點將從「模擬配適」節點中繼承所有欄位、儲存類型和統計分佈資訊。

定義指向「模擬配適」節點的更新鏈結

您可以在「模擬產生」節點與「模擬配適」節點之間建立一個鏈結。如果要使用最佳配適分佈（由到歷程資料的配適確定）的資訊更新一個或多個欄位，那麼此鏈接很有用。

1. 用滑鼠右鍵按一下「模擬產生」節點。
2. 從功能表中，選取定義更新鏈結。游標將變更為鏈結游標。
3. 按一下其他節點。如果此節點是「模擬配適」節點，那麼將建立鏈結。如果此節點不是「模擬配適」節點，那麼不會建立鏈結，並且游標將變更回正常游標。

如果「模擬配適」節點中的欄位不同於「模擬產生」節點中的欄位，那麼將顯示一條訊息，告知您存在差異。

使用「模擬配適」節點更新已鏈結的「模擬產生」節點時，結果取決於是否在這兩個節點中顯示了相同的欄位，以及是否在「模擬產生」節點中解除鎖定了這些欄位。下表格顯示了「模擬配適」節點的更新結果。

表 10. 更新「模擬適合」節點的結果

	「模擬適合」節點 中的字段	
「模擬產生」節點中的欄位	出現	遺漏
出現並已解除鎖定。	已改寫欄位。	已刪除欄位。
遺漏。	已新增欄位。	未進行變更。
出現並已鎖定。	未覆蓋欄位的分配。「配適詳細資料」對話框中的資訊以及相關性已更新。	未改寫欄位。相關性設定為 0。

表 10. 更新「模擬適合」節點的結果 (繼續)

	「模擬適合」節點 中的字段
已選取請不要在重新配適時清除下限/上限值複選框。	已覆蓋欄位，「下限」和「上限值」欄中的值除外。
已選取請不要在重新配適時重新計算相關性勾選框。	如果欄位已解除鎖定，那麼將改寫該欄 未改寫相關性。位。

移除指向「模擬配適」節點的更新鏈結

通過完成下列步驟，可以移除「模擬產生」節點與「模擬配適」節點之間的鏈結：

1. 用滑鼠右鍵按一下「模擬產生」節點。
2. 從功能表中選取**移除更新鏈結**。將移除此鏈結。

### 為「模擬產生」節點設定選項

使用「模擬產生」節點對話框的「資料」標籤上的選項可以執行下列操作：

- 檢視、指定和編輯欄位的統計分佈資訊。
- 檢視、指定和編輯欄位之間的相關性。
- 指定要模擬的疊代次數和觀察值數。

**選取項目。** 使您可以在「模擬產生」節點的以下三個視圖之間進行切換：模擬欄位、相關性和進階選項。

#### 「模擬欄位」視圖

如果使用歷程資料根據「模擬配適」節點產生或更新了「模擬產生」節點，那麼您可以在「模擬欄位」視圖中檢視並編輯每個欄位的統計分佈資訊。下列有關每個欄位的資訊將從「模擬配適」節點複製到「模擬產生」節點的**類型**標籤中：

- 測量層級
- 值
- 遺漏
- 檢查
- 角色

如果您沒有歷程資料，那麼可以通過選取儲存類型，選取分佈類型並鍵入必要的參數來定義欄位並指定這些欄位的分佈。以這種方式產生資料意味著，在對資料進行實例化之前（例如，在**類型**標籤或「類型」節點中對資料進行實例化），有關每個欄位的測量層次的資訊將不可用。

「模擬欄位」視圖包含一些工具，您可以使用這些工具來執行下列作業：

- 新增和移除欄位。
- 在顯示畫面上變更欄位的順序。
- 為每個欄位指定儲存類型。
- 為每個欄位指定統計分佈。
- 為每個欄位的統計分佈指定參數值。

**模擬欄位。** 如果已將「模擬產生」節點從「來源」選用區新增到串流畫布中，那麼此表格將包含一個空白列。編輯此列時，將在表格的底部新增一個新的空白列。如果已根據「模擬配適」節點建立了「模擬產生」節點，那麼此表格將為歷程資料的每個欄位提供一列。通過按一下**新增新欄位**圖示，您可以向表格中新增額外的列。

模擬欄位表包含下列欄：

- **欄位。** 包含欄位的名稱。可以通過在單元中鍵入內容來編輯欄位名稱。
- **儲存體。** 此欄中的儲存格包含儲存類型的下拉清單。可用的儲存類型為**字串、整數、實數、時間、日期和時間戳記**。儲存類型的選取確定了「分佈」欄中可用的分佈。如果已比「模擬配適」節點建立了「模擬產生」節點，那麼將從「模擬配適」節點複製儲存類型。

註：對於儲存類型為日期時間的欄位，必須將分佈參數指定為整數。例如，要指定平均日期 1970 年 1 月 1 日，請使用整數 0。帶符號的整數代表自 1970 年 1 月 1 日午夜以來（或在此之前）經過的秒數。

- **狀態。** 「狀態」欄中的圖示指示每個欄位的擬合狀態。



未對欄位指定分佈或者缺少一個或多個分佈參數。要執行模擬，您必須為此欄位指定分佈並輸入參數的有效值。



欄位設定為最接近的配適分佈。

註：只有在根據「模擬配適」節點建立了「模擬產生」節點的情況下，才會顯示此圖示。



已將最接近的配適分佈取代之為「配適詳細資料」子對話框中的替代分佈。請參閱第 50 頁的『配適詳細資料』主題，以取得更多資訊。



已手動指定或編輯了分佈，該分佈可能包含在多個層次指定的參數。

- **已鎖定。** 通過選中帶有鎖定圖示的欄中的複選框來鎖定模擬欄位，可以阻止已鏈結的「模擬配適」節點自動更新該欄位。如果您手動指定分佈並希望確保在執行已鏈結的「模擬配適」節點時自動分佈配適不會影響該分佈，那麼此列非常有用。
- **分配。** 此欄中的儲存格包含統計分佈的下拉清單。儲存類型的選取決定了指定欄位的此欄中可用的分佈。請參閱第 53 頁的『分佈』主題，以取得更多資訊。

註：您不能為每個欄位指定「固定」分配。如果希望產生的資料中每個欄位都是固定欄位，那麼您可以使用後跟「平衡」節點的「使用者輸入」節點。









- **參數。** 此欄中顯示與每個已配適的分佈關聯的分佈參數。參數的多個值之間以逗點分隔。為參數指定多個值將為模擬產生多個疊代。請參閱第 52 頁的『疊代』主題，以取得更多資訊。如果缺少參數，那麼「狀態」欄中顯示的圖示將反映此情況。要為參數指定值，請在對應於相關欄位的列中按一下此欄，然後從清單中選擇**指定**。這將開啟「指定參數」子對話框。請參閱第 51 頁的『指定參數』主題，以取得更多資訊。如果在「分佈」欄中選擇了「以經驗為依據的」，那麼此欄將處於停用狀態。
- **下限，上限值。** 對於某些分佈，您可以在此欄中指定模擬資料的最小值和/或最大值。小於最小值以及大於最大值的模擬資料將被拒絕，即使這些資料對於指定分佈有效也是如此。要指定最小值和最大值，請在對應於相關欄位的列中按一下此欄，然後從清單中選擇**指定**。這將開啟「指定參數」子對話框。請參閱第 51 頁的『指定參數』主題，以取得更多資訊。如果在「分佈」欄中選擇了「以經驗為依據的」，那麼此欄將處於停用狀態。

**使用最接近擬合。**只有在已使用歷程資料根據「模擬配適」節點自動建立了「模擬產生」節點，並且在模擬欄位表中選取了單個列的情況下，才會啟用此按鈕。此按鈕用於將所選列中欄位的資訊取代為該欄位的最佳配適分佈資訊。如果對所選列中的資訊進行了編輯，那麼按一下此按鈕會將資訊重設為根據「模擬配適」節點確定的最佳配適分佈。

**配適詳細資料。**只有在根據「模擬配適」節點自動建立了「模擬產生」節點的情況下，才會啟用此列。此列用於開啟「擬合詳細資料」子對話框。請參閱第 50 頁的『配適詳細資料』主題，以取得更多資訊。

使用「模擬欄位」視圖右側的圖示可以執行一些有用的作業。下表格說明了這些圖示。

表 11. 「模擬欄位」視圖上的圖示.

圖示	工具提示	說明
	編輯分佈參數	僅當在「模擬欄位」表格中選取單一系列時，才會啟用此項。用於開啟所選列的「指定參數」子對話框。請參閱第 51 頁的『指定參數』主題，以取得更多資訊。
	新增新欄位	僅當在「模擬欄位」表格中選取單一系列時，才會啟用此項。用於向模擬欄位表的底部新增新的空白列。
	建立多個副本	僅當在「模擬欄位」表格中選取單一系列時，才會啟用此項。用於開啟「複製欄位」子對話框。請參閱第 50 頁的『複製欄位』主題，以取得更多資訊。
	刪除所選欄位	用於從「模擬欄位」表格中刪除所選列。
	移至頂端	只有在所選列不在模擬欄位表的頂部時，才會啟用此圖標。用於將所選列移動到模擬欄位表的頂部。此動作會影響欄位在模擬資料中的順序。
	向上移動	只有在所選列不在模擬欄位表的頂部時，才會啟用此圖標。用於在模擬欄位表中將所選列上移一個位置。此動作會影響欄位在模擬資料中的順序。
	向下移動	只有在所選列不在模擬欄位表的底部時，才會啟用此圖標。用於在模擬欄位表中將所選列下移一個位置。此動作會影響欄位在模擬資料中的順序。
	移至底部	只有在所選列不在模擬欄位表的底部時，才會啟用此圖標。用於將所選列移動到模擬欄位表的底部。此動作會影響欄位在模擬資料中的順序。

**重新配適時不清除下限和上限值。**選取此圖標後，通過執行已連線的「模擬配適」節點來更新分佈時，將不會改寫最小值和最大值。

## 「相關性」視圖

我們知道，預測模型的輸入欄位通常具有相關性 - 例如，身高和體重。必須考慮將模擬的欄位之間的相關性，以確保模擬值保留這些相關性。

如果使用歷程資料根據「模擬配適」節點產生或更新了「模擬產生」節點，那麼您可以在「相關性」視圖中檢視並編輯成對欄位之間的已計算相關性。如果您沒有歷程資料，那麼可以根據您對欄位相關係數方式的瞭解來手動指定相關性。

註：在產生任何資料之前，將自動檢查相關性矩陣以確認它是否為半正定矩陣，並且是否能因此進行反轉。如果矩陣的欄線性獨立，那麼可以將該矩陣反轉。如果無法反轉相關性矩陣，那麼它將自動進行調整以實現可反轉。

您可以選擇以矩陣格式或清單格式顯示相關性。

**相關性矩陣。**用於顯示矩陣中成對欄位之間的相關性。欄位名稱字母的列出，從矩陣的左上方向下列出。只能編輯對角線下方的單元；必須已輸入介於 -1.000（含本數）到 1.000（含本數）之間的值。對角線上方的單元將在對焦點離開對角線下方該單元的鏡映單元時進行更新；這兩個單元將顯示同一值。對角線單元始終處於停用狀態，並始終具有相關性 1.000。所有其他單元的預設值為 0.000。值 0.000 指定關聯的欄位對之間不存在相關性。矩陣中僅併入連續欄位和序數欄位。列名、種類和旗標欄位以及分配了「固定」分佈的欄位不會顯示在表格中。

**相關性清單。**用於顯示表格中成對欄位之間的相關性。表格中的每列顯示欄位對之間的相關性。無法新增或刪除列。標題為「欄位 1」和「欄位2」的欄包含無法編輯的欄位名稱。「相關性」欄包含可以編輯的相關性；必須已輸入介於 -1.000（含本數）到 1.000（含本數）之間的值。所有單元的預設值為 0.000。清單中僅併入連續欄位和序數欄位。列名、種類和旗標欄位以及分配了「固定」分佈的欄位不會顯示在清單中。

**重設相關性。**用於開啟「重設相關性」對話框。如果歷程資料可用，那麼您可以選擇以下三個選項中的一個：

- **已配適。**用於將目前相關性取代為那些使用歷程資料計算出的相關性。
- **0。**用於將目前相關性取代為 0。
- **取消。**關閉對話框。相關性保持不變。

如果歷程資料無法使用，但您已變更相關性，那麼可以選擇將目前相關性取代為 0，或者選擇取消。

**顯示為。**選取表格可將相關性顯示為矩陣。選取清單可將相關性顯示為清單。

**重新配適時不重新計算相關性。**如果您要手動指定相關性，並阻止使用「模擬配適」節點和歷程資料自動配適分佈時覆蓋這些相關性，請選取此選項。

**對包含種類配送的輸入使用適合的多方應急表格。**依預設，所有具有種類分佈的欄位將併入在列聯表或多路列聯表中，具體取決於具有種類分佈的欄位的號碼。執行「模擬配適」節點時，將構造列聯表（與相關性相似）。無法檢視列聯表。如果選取了此選項，那麼將使用列聯表中的實際百分比來模擬具有種類分佈的欄位。也就是說，將在新的模擬資料中重建列名欄位之間的關聯。如果取消選擇此選項，那麼將使用列聯表格中的預期百分比來模擬具有種類分佈的欄位。如果修改了某個欄位，那麼將從列聯表中移除該欄位。

## 「進階選項」視圖

**要模擬的觀察值數目。**此選項顯示用於指定要模擬的觀察值數目以及任何疊代的命名方式的選項。

- **觀察值數目上限。**這表示要產生的模擬資料最大觀察值個數和相關目標值。預設值為 100,00，最小值為 1000，最大值為 2,147,483,647。
- **疊代。**此數字由系統自動計算，並且無法進行編輯。每次對分佈參數指定多個值時，將自動建立疊代。

- **總列數。**只有在疊代次數大於 1 的情況下，才會啟用此選項。將使用所顯示的方程式自動計算此數字，並且此數字無法進行編輯。
- **建立疊代欄位。**只有在疊代次數大於 1 的情況下，才會啟用此選項。選取此選項後，將啟用名稱欄位。請參閱第 52 頁的『疊代』主題，以取得更多資訊。
- **名稱。**只有在已選取了建立疊代欄位勾選框，並且疊代次數大於 1 的情況下，才會啟用此選項。可以通過在此文字欄位中鍵入內容來編輯疊代欄位的名稱。請參閱第 52 頁的『疊代』主題，以取得更多資訊。

**隨機種子。** 設定亂數種子以供您複製模擬。

- **複製結果。** 選取此選項後，將啟用產生按鈕和隨機種子欄位。
- **隨機種子。** 僅在選取了複製結果勾選框時啟用。您可以在此欄位中指定要用作隨機種子的整數。預設值為 629111597。
- **產生。** 僅在選取了複製結果勾選框時啟用。用於在隨機種子欄位中建立介於 1（含本數）到 999999999（含本數）之間的偽隨機整數。

## 複製欄位

您可以在「複製欄位」對話框中指定要建立的選取的欄位副本數以及每個副本的命名方式。調查複合作用（例如一些連續時段內的利率或成長率）時，具有欄位的多個副本非常有用。

此對話框的標題列包含所選欄位的名稱。

**要創建的副本數。** 包含要建立的欄位副本數。按一下箭頭可選取要建立的副本數。副本數目上限 1，上限為 512。副本數目最初設定為 10。

**複製字尾字元。** 包含新增到每個副本的欄位名稱末尾的字元。這些字元用於分隔欄位名稱與副本編號。可以通過在此欄位中鍵入內容來編輯字尾字元。可以將此欄位留空白；在這種情況下，欄位名稱與副本編號之間將不存在任何字元。預設字元為底線。

**初始副本編號。** 包含第一個副本的字尾編號。按一下箭頭可選取初始副本編號。下限初始副本編號為 1，上限編號為 1000。預設初始副本編號為 1。

**副本編號步進值。** 包含字尾編號的增量。按一下箭頭可選取增量。下限增量為 1，上限增量為 255。增量最初設定為 1。

**欄位。** 包含副本的欄位名稱預覽，對「複製欄位」對話框的任何欄位進行編輯時，此預覽將進行更新。此文字由系統自動產生，並且無法進行編輯。

**確定。** 按照對話框中指定的內容產生所有副本。這些副本將新增到「模擬產生」節點對話框的模擬欄位表中，位於包含已複製欄位的列的正下方。

**取消。** 關閉對話框。會捨棄已做出的任何變更。

## 配適詳細資料

只有在已通過執行「模擬配適」節點來建立或更新「模擬產生」節點的情況下，「配適詳細資料」對話框才可用。此對話框顯示選取的欄位的自動分佈配適結果。分配以適合度排序，最接近配適分配會列於第一個。您可以在此對話框中執行下列作業：

- 檢查擬合到歷程資料的分佈。
- 選取其中一個已配適的分佈。

**欄位。** 包含所選取欄位的名稱。無法編輯此文字。

**處理方法（測量）。**顯示選取的欄位的測量類型。此類型來自「模擬產生」節點對話框中的模擬欄位表。可以通過按一下箭頭並從下拉清單中選取測量類型來變更此測量類型。提供了以下三個選項：**連續**、**列名和序數**。

**分佈。**「分佈」表格顯示適合於此測量類型的所有分佈。已配適到歷程資料的分佈將按配適度從最佳到最差的順序進行排序。配適度由「模擬配適」節點中選擇的適合度統計資料確定。不配適到歷程資料的分佈字母的列示在表格中已配適的分佈下方。

「分佈」表格包含下列欄：

- **使用。** 所選圓鈕指示目前為欄位選取的分佈。通過在「使用」欄中選取與所需分佈對應的圓鈕，您可以置換最接近的配適分佈。在「使用」欄中選取圓鈕還將顯示選取的欄位的歷程資料直方圖（或長條圖）上疊加的分佈圖形。一次只能選取一個分佈。
- **分配。** 包含分佈的名稱。無法編輯此直欄。
- **配適統計資料。** 包含針對分佈計算的配適統計資料。無法編輯此直欄。此 Cell 的內容取決於欄位的測量類型：
  - **連續。** 包含 Anderson-Darling 測試和 Kolmogorov-Smirnoff 測試的結果。也會顯示和測試相關的  $p$  值。最先顯示的是選擇作為「模擬配適」節點中的配適度準則的適合度統計資料，它用於對分佈進行排序。Anderson-Darling 統計資料顯示為  $A=aval$   $P=pval$ 。Kolmogorov-Smirnoff 統計資料顯示為  $K=kval$   $P=pval$ 。如果無法計算某個統計資料，那麼將顯示一個點來代替數字。
  - **列名和序數。** 包含卡方測試測試的結果。還將顯示與此測試關聯的  $p$  值。統計資料顯示為  $Chi-Sq=val$   $P=pval$ 。如果不配適分佈，那麼將顯示不配適。如果無法以數學方法擬合分佈，那麼將顯示無法擬合。

註：對於經驗分佈，此 Cell 始終為空白。

- **參數。** 包含與每個已配適分佈關聯的分佈參數。這些參數顯示為  $parameter\_name = parameter\_value$ ，參數之間以單空格分隔。對於類別分配，參數名稱為類別且參數值為關聯的機率。如果分佈未擬合到歷程資料，那麼此 Cell 為空白。無法編輯此直欄。

**直方圖縮圖。**顯示所選欄位的歷程資料直方圖上疊加在所選分佈圖形。

**分佈縮圖。**顯示所選分佈的說明和圖示。

**確定。** 用於關閉對話框，並使用所選分佈中的資訊對選取的欄位的模擬欄位表中以下欄的值進行更新：測量、分佈、參數以及下限和上限值。另外，還將更新「狀態」欄中的圖示，以反映所選分佈是否為與資料擬合程度最近的分佈。

**取消。** 關閉對話框。會捨棄已做出的任何變更。

## 指定參數

您可以在「指定參數」對話框中手動為選取的欄位的分配指定參數值。另外，也可以為選取的欄位選擇其他分佈。

「指定參數」對話框可通過以下三種方式開啟：

- 按兩下「模擬產生」節點對話框中模擬欄位表內的欄位名稱。
- 按一下「模擬欄位」表格的「參數」或「下限，上限值」欄，然後從清單中選擇**指定**。
- 在模擬欄位表中選取一行，然後按一下**編輯分佈參數圖示**。

**欄位。** 包含所選取欄位的名稱。無法編輯此文字。

**分配。** 包含所選欄位的分配。此資訊來自模擬欄位表。可以通過按一下箭頭並從下拉清單中選取分佈來變更分佈。可用分佈取決於選取的欄位的儲存類型。

**面數。**只有在從分佈欄位中選取了骰子分佈的情況下，此選項才可用。按一下箭頭可選取要將欄位分割為的面數（種類）。下限面數為 2，上限面數為 20。面數最初設定為 6。

**分佈參數。**分佈參數表格為所選分佈的每個參數都提供了一列。

註：分配使用比率參數，形狀參數  $\alpha = k$  且反向尺度參數  $\beta = 1/\theta$ 。

表格包含兩行：

- **參數。**包含參數的名稱。無法編輯此直欄。
- **值。**包含參數的值。如果已根據「模擬配適」節點建立或更新了「模擬產生」節點，那麼此欄中的儲存格包含已通過將分佈配適到歷程資料來確定的參數值。如果已將「模擬產生」節點新增到「來源」節點選用區內的串流畫布中，那麼此欄中的單元將為空白。可以通過在單元中鍵入內容來編輯值。請參閱主題第 53 頁的『分佈』以獲取有關每個分佈所需要的參數以及可接受的參數值的進一步資訊。

參數的多個值之間必須以逗點進行分隔。為參數指定多個值將定義模擬的多個疊代。只能為一個參數指定多個值。

註：對於儲存類型為日期時間的欄位，必須將分佈參數指定為整數。例如，要指定平均日期 1970 年 1 月 1 日，請使用整數 0。

註：如果選取了骰子分佈，那麼分佈參數表格將略有不同。此表格為每個面（或種類）都提供了一列。此表格包含「值」欄和「機率」欄。「值」欄包含每個種類的標籤。標籤的預設值為 1 - N 之間的整數，其中 N 是面數。可以通過在單元中輸入內容來編輯標籤。可以在單元中已輸入任意值。如果希望使用非數字值，那麼在儲存類型未設定為字串的情況下，必須將資料欄位的儲存類型變更為字串。「機率」欄包含每個種類的機率。無法編輯機率，並且這些機率按照  $1/N$  進行計算。

**預覽。**根據指定的參數顯示分佈的樣本圖形。如果為一個參數指定了兩個或兩個以上的值，那麼將顯示此參數的每個值的樣本圖形。如果歷程資料可用於選取的欄位，那麼分佈圖形將疊加在歷程資料直方圖上。

**選用設定。**使用這些選項可以指定模擬資料的最小值和/或最大值。小於最小值以及大於最大值的模擬資料將被拒絕，即使這些資料對於指定分佈有效也是如此。

- **指定最小值。**選取此選項可啟用拒絕小於最小值的值欄位。該勾選框會在選取了「經驗」分配的情況下停用。
- **拒絕小於最小值的值。**只有在選取了指定最小值的情況下，才會啟用此字段。請輸入模擬資料的最小值。任何小於此值的模擬值將被拒絕。
- **指定最大值。**選取此選項可啟用拒絕大於最大值的值欄位。該勾選框會在選取了「經驗」分配的情況下停用。
- **拒絕大於最大值的值。**只有在選取了指定最大值的情況下，才會啟用此字段。請輸入模擬資料的最大值。任何大於此值的模擬值將被拒絕。

**確定。**用於關閉對話框，並對選取的欄位的「模擬欄位」表格的以下欄的值進行更新：分佈、參數以及「下限，上限值」。另外，還將更新「狀態」欄中的圖示以反映所選分佈。

**取消。**關閉對話框。會捨棄已做出的任何變更。

## 疊代

如果您已經為固定欄位或分佈參數指定了多個值，那麼將針對每個指定值產生一組獨立的模擬觀察值（有效的獨立模擬）。這使您可以調查更改欄位或參數所產生的效果。每組模擬觀察值稱為疊代。在模擬資料中，疊代是重疊的。



如果已選取了「模擬產生」節點對話框的「進階選項」視圖中的建立疊代欄位勾選框，那麼疊代欄位將作為具有數值儲存的列名欄位新增到模擬資料中。通過在「進階選項」視圖中的名稱欄位內鍵入內容，可以對此欄位的名稱進行編輯。此欄位包含一個標籤，指示每個模擬觀察值所屬的疊代。標籤的格式取決於疊代的類型：

- 疊代固定欄位。標籤包含：欄位名稱，後跟等號，然後是該疊代的欄位值，即

$field\_name = field\_value$

- 疊代分佈參數。標籤包含：欄位名稱，後跟冒號，其次是疊代參數的名稱，然後是等號，最後是該疊代的參數值，即

$field\_name;parameter\_name = parameter\_value$

- 疊代種類分佈或範圍分佈的分佈參數。標籤包含：欄位名稱，後跟冒號，然後是 "Iteration"，最後是疊代次數，即

$field\_name: Iteration\ iteration\_number$

## 分佈

通過開啟任何欄位的「指定參數」對話框，並從分佈清單中選取所需分佈，然後在分佈參數表格中輸入分佈參數，您可以手動指定該欄位的機率分配。下列為特定分配的某些備註：

- 類別。種類分佈說明數值號碼固定的輸入欄位，該輸入欄位稱為種類。每個類別都具有相關的機率，使得所有類別的機率總合等於 1。

註：如果您為種類指定的機率的總和不等於 1，那麼將接收到警告。

- 負二項式 - 失敗。觀察到指定的成功個數前，請描述在一系列的試驗中失敗個數的分配。參數 *Threshold* 是指定的成功個數，而參數 *Probability* 是任何指定試驗中的成功機率。
- 負二項式 - 試驗。說明在觀察到指定次數的成功之前，所需失敗次數的分配。參數 *Threshold* 是指定的成功個數，而參數 *Probability* 是任何指定試驗中的成功機率。
- 範圍。此分配包含一組具有指定到每個區間機率的區間，使得所有區間的機率總合等於 1。指定區間中的數值抽取自定義於該區間的均勻分配。區間是以輸入下限、上限和相關機率來進行指定。

例如，您認為原料成本落入單位價格 \$10 - \$15 的範圍內的概率為 40%，而落入單位價格 \$15 - \$20 的範圍內的概率為 60%。您可以使用包含兩個區間 [10 - 15] 和 [15 - 20] 的「範圍」分配建立成本模式，將和第一個區間相關的機率設定為 0.4，而和第二個區間相關的機率為 0.6。區間不必連續，甚至可以重疊。例如，您可能指定了區間 \$10 - \$15 和 \$20 - \$25，或者指定了 \$10 - \$15 和 \$13 - \$16。

- **Weibull**。參數 *Location* 是一個選用的位置參數，用於指定分佈源所在的位置。

下表顯示了可用於自訂分佈配適的分佈以及參數的可接受值。其中某些分佈可用於到特定儲存類型的自訂配適，即使「模擬配適」節點未自動將它們配適到這些儲存類型也是如此。

表 12. 可用於自訂適合的分配

分佈	自訂配適受支援的儲存類型	參數	參數限制	附註
白努利	整數、實數、日期時間	機率	$0 \leq \text{機率} \leq 1$	
Beta	整數、實數、日期時間	形狀 1 形狀 2 下限 上限	$\geq 0$ $\geq 0$ < 上限 > 下限	下限和上限都是選用項目。

表 12. 可用於自訂適合的分配 (繼續)

分佈	自訂配適受支援的儲存類型	參數	參數限制	附註
二項式	整數、實數、日期時間	試驗次數 (n) 機率 下限 上限	$> 0$ ，整數 $0 \leq \text{機率} \leq 1$ $<$ 上限 $>$ 下限	試驗次數必須是整數。下限和上限都是選用項目。
種類	整數、實數、日期時間、字串	種類名稱 (或標籤)	$0 \leq \text{值} \leq 1$	值表示種類的機率。值的總和必須等於 1，否則將產生警告。
骰子	整數，字串	面數	$2 \leq \text{面數} \leq 20$	每個種類 (面) 的機率按照 $1/N$ 進行計算，其中 $N$ 表示面數。無法編輯機率。
以經驗為依據的	整數、實數、日期時間			無法編輯經驗分佈或者選取它作為某種類型。  只有在存在歷程資料的情況下，經驗分佈才可用。
指數	整數、實數、日期時間	尺度 下限 上限	$> 0$ $<$ 上限 $>$ 下限	下限和上限都是選用項目。
固定	整數、實數、日期時間、字串	數值		您不能為每個欄位指定「固定」分配。如果希望產生的資料中每個欄位都是固定欄位，那麼您可以使用後跟「平衡」節點的「使用者輸入」節點。
Gamma	整數、實數、日期時間	形狀 尺度 下限 上限	$\geq 0$ $\geq 0$ $<$ 上限 $>$ 下限	下限和上限都是選用項目。  分配使用比率參數，形狀參數 $\alpha = k$ 且反向尺度參數 $\beta = 1/\theta$ 。
對數常態	整數、實數、日期時間	形狀 1 形狀 2 下限 上限	$\geq 0$ $\geq 0$ $<$ 上限 $>$ 下限	下限和上限都是選用項目。
負二項式 - 失敗	整數、實數、日期時間	臨界值 機率 下限 上限	$\geq 0$ $0 \leq \text{機率} \leq 1$ $<$ 上限 $>$ 下限	下限和上限都是選用項目。
負二項式 - 試用	整數、實數、日期時間	臨界值 機率 下限 上限	$\geq 0$ $0 \leq \text{機率} \leq 1$ $<$ 上限 $>$ 下限	下限和上限都是選用項目。

表 12. 可用於自訂適合的分配 (繼續)

分佈	自訂配適受支援的儲存類型	參數	參數限制	附註
常態	整數、實數、日期時間	平均數 標準差 下限 上限	$\geq 0$ $> 0$ $< \text{上限}$ $> \text{下限}$	下限和上限都是選用項目。
Poisson	整數、實數、日期時間	平均數 下限 上限	$\geq 0$ $< \text{上限} > \text{下限}$	下限和上限都是選用項目。
範圍	整數、實數、日期時間	Begin(X) End(X) Probability(X)	$0 \leq \text{值} \leq 1$	X 是每個分級的指數。機率值的總和必須等於 1。
三角形	整數、實數、日期時間	模式  下限 上限	下限 $\leq$ 值 $\leq$ 上限 $< \text{上限}$ $> \text{下限}$	
均勻分配	整數、實數、日期時間	下限 上限	$< \text{上限}$ $> \text{下限}$	
Weibull	整數、實數、日期時間	速率 尺度 位置 下限 上限	$> 0$ $> 0$ $\geq 0$ $< \text{上限}$ $> \text{下限}$	位置、上限和下限是選用的。

## 延伸匯入節點

使用「延伸匯入」節點，您可以執行 R 或 Python for Spark Script 以匯入資料。

### 延伸匯入節點 - 語法標籤

選取語法的類型 - **R** 或 **Python for Spark**。然後輸入或貼上用於匯入資料的自訂 Script。當您的語法備妥時，可以按一下執行，以執行「延伸匯入」節點。

### 延伸匯入節點 - 主控台輸出標籤

主控台輸出標籤包含「語法」標籤上的 R Script 或 Python for Spark Script 執行時接收的所有輸出（例如，如果使用 R Script，它顯示執行語法標籤上 **R** 語法欄位中的 R Script 時，從 R 主控台接收的輸出）。此輸出可能包括執行 R 或 Python Script 時產生的 R 或 Python 錯誤訊息或警告。輸出可以主要用於 Script 除錯。主控台輸出標籤也包含 **R** 語法或 **Python** 語法欄位中的 Script。

每次執行延伸匯入 Script 時，都會使用從 R 主控台或 Python for Spark 接收的輸出來改寫主控台輸出標籤的內容。無法編輯輸出。

### 過濾或重新命名欄位

您可以在串流中的任何點重新命名或執行欄位。例如，作為醫療研究人員，您可能不考量病患（記錄層次資料）的鉀水平（欄位層次資料）；因此，您可以過濾掉 K（鉀）欄位。可以使用單獨的「過濾器」節點，或者來源或輸出節點上的「過濾器」標籤，完成此作業。功能相同，而無論存取自哪個節點。

- 您可以從來源節點，例如「變數檔」、「固定檔」、Statistics 檔、XML 或「延伸匯入」，在將資料讀取至 IBM SPSS Modeler 時，重新命名或過濾欄位。
- 使用「過濾器」節點，您可以在串流中的任何點重新命名或過濾欄位。
- 從「Statistics 匯出」、「Statistics 轉換」、「Statistics 模型」及「Statistics 輸出」節點，您可以過濾或重新命名欄位，以符合 IBM SPSS Statistics 命名標準。請參閱第 316 頁的『重新命名或過濾 IBM SPSS Statistics 的欄位』主題，以取得更多資訊。
- 您可以使用上述任何節點中的「過濾器」標籤，以定義或編輯多重回應集。請參閱第 131 頁的『編輯多重回應集』主題，以取得更多資訊。
- 最後，您可以使用「過濾器」節點，將欄位從一個來源節點對映至另一個。

---

## 資料視圖節點

使用「資料視圖」節點可以將 IBM SPSS Collaboration and Deployment Services 分析資料視圖中定義的資料包含在串流中。分析資料視圖定義用於存取資料的結構，此結構說明了預測模型和商業規則中使用的實體。該視圖將資料結構與實體資料來源關聯以進行分析。

預測性分析需要在表格中對資料進行組織，表格中的每一列對應於一個對其進行預測的實體。表格中的每一欄各代表一個可測量實體屬性。某些屬性可能通過聚合另一個屬性的值進行衍生。例如，表格中的列可以代表客戶，而欄對應於客戶姓名、性別、郵遞區號以及客戶在過去一年內購物費用超過 \$500 的次數。最後一欄一欄衍生自客戶訂單歷史，該歷史通常儲存在一張或多張相關表格中。

預測分析過程涉及在模型整個生命週期中使用不同資料集。在預測模型的初始開發過程中，可以使用歷史資料，此資料通常已經知道待預測事件的結果。要對模型的有效性和精確度進行評估，您可以根據不同的資料驗證候選模型。驗證此模型之後，將它部署到正式作業使用，以便為批次程序中的多個實體或即時程序中的單個實體產生分數。如果您在決策管理過程中將模型與商業規則進行組合，那麼可以使用模擬資料來驗證該組合的結果。但是，儘管各個模型開發過程階段所使用的資料不同，每個資料集都必須為模型提供同一屬性集。此屬性集保持不變；要分析的資料記錄將發生變更。

分析資料視圖包含下列元件，這些元件能夠滿足預測性分析的特殊需求：

- 資料視圖綱目或資料模型，定義了一個邏輯介面，用於將資料作為組織到相關表格中的一組屬性進行存取。該模型中的屬性可衍生自其他屬性。
- 一個或多個資料存取方案，這些方案提供了具有實體值的資料模型屬性。通過指定哪個資料存取方案對於特定應用程式處於作用中狀態，您可以控制可供資料模型使用的資料。

**重要：**要使用「資料視圖」節點，必須先在您的網站安裝並配置 IBM SPSS Collaboration and Deployment Services 儲存庫。此節點參照的分析資料視圖通常是使用 IBM SPSS Deployment Manager 在儲存庫中建立和儲存的。

## 為「資料視圖」節點設定選項

使用「資料視圖」節點對話框的資料標籤上的選項可以為從 IBM SPSS Collaboration and Deployment Services 儲存庫 中選取的分析資料視圖指定資料設定。

**分析資料視圖。**按一下省略號按鈕 (...) 可選取分析資料視圖。如果您目前未連接至儲存庫伺服器，請在「儲存庫：伺服器」對話框中指定伺服器的 URL，並按一下**確定**，然後在「儲存庫：認證」對話框中指定連線認證。有關登入儲存庫並擷取物件的進一步資訊，請參閱《IBM SPSS Modeler 使用手冊》。

**表格名稱。**請從分析資料視圖中的資料模型中選取一個表格。資料模型中的每一個表格都代表預測分析程序中涉及的概念或實體。表格的欄位對應於這些表格所代表的實體的屬性。例如，如果要對客戶訂單進行分析，那

麼資料模型可以包含一張客戶表格和一張訂單表格。客戶表格可能包含客戶 ID、年齡、性別、婚姻狀況及居住國家等屬性。訂單表格可能包含訂單 ID、訂單中的商品數、總成本以及下訂單的客戶的 ID 等屬性。客戶 ID 屬性可用於將客戶表格中的客戶與他們在訂單表格中的訂單進行關聯。

**資料存取方案。**從分析資料視圖中選取資料存取計劃。資料存取計劃會將分析資料視圖中的資料模型表格與實體資料來源相關聯。分析資料視圖通常包含多個資料存取計劃。變更使用的資料存取計劃時，您會變更串流所用的資料。例如，如果分析資料視圖包含用於訓練模型的資料存取計劃以及用於測試模型的資料存取計劃，則您可以透過變更使用的資料存取計劃來從訓練資料切換至測試資料。

**選用屬性。**如果使用分析資料視圖的應用程式不需要某個特定屬性，那麼您可以將此屬性標示為選用。與必要屬性不同，選用屬性可以包含空值。您可能需要對應用程式進行調整，以包含選用屬性的空值處理。例如，呼叫 IBM Operational Decision Manager 中建立的商業規則時，IBM Analytical Decision Management 將查詢規則服務以確定必要的輸入。對於任何規則服務的必填欄位，如果要進行評分的記錄包含空值，那麼將不會呼叫該規則，並且將使用預設值移入該規則的輸出欄位。如果選用欄位包含空值，那麼將呼叫該規則。該規則可以進行空值檢查以控制處理。

要將屬性指定為選用屬性，請按一下**選用屬性**，然後選取選用屬性。

**在欄位中包含 XML 資料。**選取此選項可建立一個包含每列資料的可執行物件模型 XML 資料的欄位。如果資料將與 IBM Operational Decision Manager 配合使用，那麼此資訊是必要的。請為此新欄位指定名稱。

---

## 「地理空間」來源節點

您可以使用「地理空間」來源節點將地圖或空間資料引入到資料採礦階段作業中。您可以通過兩種方法中的一種來匯入資料：

- 通過形狀檔案 (.shp) 進行導入
- 通過連接至包含地圖檔案的階層式檔案系統所在的 ESRI 伺服器進行導入。

註：您只能連接到公用地圖服務。

空間-時間預測 (STP) 模型可在其預測中包括地圖元素或空間元素。有關這些模型的進一步資訊，請參閱《Modeler 建模節點》手冊 (ModelerModelingNodes.pdf) 的『「時間序列」模型』部分中標題為『空間-時間預測建模節點』的主題。

## 設定「地理空間」來源節點的選項

**資料來源類型** 您可以從形狀檔案 (.shp) 匯入資料，也可以連接至地圖服務。

如果您使用的是**形狀檔案**，請輸入該檔案的檔案名稱和檔案路徑，或者進行瀏覽以選取檔案。該檔案必須位於本端目錄上或從對映驅動器進行存取；您無法使用統一命名慣例 (UNC) 路徑存取該檔案。

註：形狀資料同時需要 .shp 和 .dbf 檔案。這兩個檔案必須同名並且在同一個資料夾中。您選取 .shp 檔案時，將自動匯入 .dbf 檔案。另外，可以存在一個 .prj 檔案，用於指定形狀資料的座標系統。

如果您使用的是**地圖服務**，請輸入該服務的 URL，並按一下**連接**。連接至該服務後，該服務中的層將顯示在對話框底部**可用地圖**窗格中的樹狀結構結構中；請展開該樹狀結構並選取所需的層。

註：您只能連接到公用地圖服務。

## 自動定義地理空間資料

依預設，在有可能時，SPSS Modeler 將使用正確的 meta 資料自動定義來源節點中的所有地理空間資料欄位。meta 資料可能包含地理空間欄位的測量層次（例如「點」或「多邊形」）以及這些欄位所使用的座標系統，其中包含原點（例如緯度 0 經度 0）和度量單位之類的詳細資料。有關測量層次的進一步資訊，請參閱第 120 頁的『地理空間測量層次』。

構成形狀檔案的 .shp 和 .dbf 檔案包含用作鍵的一般 ID 欄位。例如，.shp 檔案可能包含國家，並將國家名稱欄位用作 ID，而 .dbf 檔案可能包含有關這些國家的資訊，並且同樣將國家名稱用作 ID。

註：如果座標系統與預設的 SPSS Modeler 座標系統不同，那麼您可能必須重新投影資料以使用所需的座標系統。如需相關資訊，請參閱第 159 頁的『重新投影』節點。

---

## 一般來源節點標籤

通過按一下對應的標籤可為所有來源節點指定下列選項：

- 「資料」標籤。用於變更預設儲存類型。
- 「過濾」標籤。用於刪除或更名資料欄位。此標籤所提供的功能與「過濾器」節點相同。請參閱第 129 頁的『設定過濾選項』主題，以取得更多資訊。
- 「類型」標籤。用於設定測量層次。此標籤所提供的功能與類型節點相同。
- 「註解」標籤。用於所有節點，此標籤提供一些選項來重新命名節點、提供自訂工具提示，以及儲存長註釋。

## 在來源節點中設定測量層次

欄位內容可在來源節點中指定也可在個別的「類型」節點中指定。兩種節點的功能相似。可用的內容如下：

- 欄位 按兩下任何欄位名稱均可指定 IBM SPSS Modeler 中的資料的值標籤和欄位標籤。例如，從 IBM SPSS Statistics 匯入的欄位 meta 資料可在此處檢視或修改。與之相似，您也可以為欄位及其值建立新的標籤。您在此處指定的標籤將根據您在「串流內容」對話框中的選項顯示在整個 IBM SPSS Modeler 中。
- 測量 這是測量層次，用於說明給定欄位中的資料的特性。如果已經瞭解某個欄位的所有詳細資料，則稱為已完全實例化。如需相關資訊，請參閱第 118 頁的『測量層次』。

註：欄位的測量層次與欄位的儲存類型不同，後者指示資料是作為字串、整數、實數、日期、時間、時間戳記還是清單進行儲存。

- 值 此欄使您能夠指定用於從資料集中讀取資料值的選項，或者使用指定選項在個別的對話框中指定測量層次和值。您還可以選擇傳遞欄位，而不讀取它們的值。如需相關資訊，請參閱第 122 頁的『資料值』。

註：如果對應的欄位項目包含清單，則無法在此直欄中修正資料格。

- 遺漏 用於指定如何處理此欄位的遺漏值。如需相關資訊，請參閱第 126 頁的『定義遺漏值』。

註：如果對應的欄位項目包含清單，則無法在此直欄中修正資料格。

- 檢查 在此欄中，您可以設定選項，以確保欄位值符合指定的值或範圍。如需相關資訊，請參閱第 126 頁的『勾選類型值』。

註：如果對應的欄位項目包含清單，則無法在此直欄中修正資料格。

- 角色 用於向建模節點指示欄位將成為用於機器學習過程的輸入（預測值欄位）還是目標（預測的欄位）。兩者、無以及分割區也是可用角色，最後一個可用角色表明欄位用於將記錄分割區到不同的樣本中，以用於進行訓練、測試和驗證。值 分割 指定將為欄位的每個可能值建立個別的模式。如需相關資訊，請參閱第 127 頁的『設定欄位角色』。

請參閱第 117 頁的『類型節點』主題，以取得更多資訊。

### 何時在來源節點實例化

您可用透過兩種方法瞭解欄位的資料儲存體及值。此實例化可能發生在任一來源節點，可能發生在首次將資料放入 IBM SPSS Modeler 時，或者透過將「類型」節點插入資料串流進行。

下列情況下，來源節點的實例化有用：

- 資料集較小。
- 您計劃使用「表示式建置器」衍生新的欄位（實例化可讓「表示式建置器」提供欄位值）。

通常，如果資料集不是非常大，並且您計劃稍後在串流中新增欄位，於來源節點的實例化是最方便的方法。

### 從來源節點中過濾欄位

使用來源節點對話框上的「過濾」標籤可以根據對資料的初始檢查排除下游作業中的欄位。此功能非常有用，例如，如果資料中存在複製的欄位，或假設您已非常熟悉資料並能夠排除不相關的欄位，那麼可選擇此功能。此外，還可以稍後在串流中新增一個個別的「過濾器」節點。此節點的功能與上述兩種情況下所使用的功能相似。請參閱第 129 頁的『設定過濾選項』主題，以取得更多資訊。





---

## 第 3 章 記錄作業節點

---

### 記錄作業概觀

記錄作業節點用於在記錄層次對資料進行變更。這些作業在資料採礦的資料理解和資料預備階段非常重要，因為通過這些作業可以根據您的特定商業需要裁剪資料。

例如，根據使用「資料審核」節點（「輸出」選用區）處理的資料審核結果，您可能決定合併過去三個月的客戶購買記錄。使用合併節點，可以基於某個索引鍵欄位（如 客戶 ID）的值合併記錄。您還可能會探索無法管理一個包含超過一百萬條網網站擊資訊記錄的資料庫。使用「樣本」節點，可以選取要用於建模的資料子集合。

記錄作業選用區包含下列節點：



「選取」節點可基於特定條件從資料串流中選取或捨棄記錄子集合。例如，可以選取有關特定銷售區域的記錄。



「樣本」節點選取記錄的子集合。支援各種樣本類型，其中包括分層樣本、叢集樣本以及非隨機（結構化）樣本。取樣對於提高效能和選取相關記錄群組或交易群組用於分析會很有用。



「平衡」節點糾正資料集中的不平衡度，因而它遵循指定的條件。平衡指令調整根據指定系數條件為 true 的記錄的比例。



「合計」節點用聚合和合計的輸出記錄替代一系列輸入記錄。



使用「近因、頻率和貨幣 (RFM) 聚合」節點，您可以採用客戶的歷程記錄交易資料，刪除所有無用資料以及將所有他們保留的交易資料組合成一系列，且該列中列出了他們與您上次談業務的時間、所完成的交易量以及這些交易的總貨幣價值。



「排序」節點可根據一個或多個欄位的值將記錄按遞增或遞減排序。



「合併」節點獲取多個輸入記錄並建立包含某些或全部輸入欄位的單個輸出記錄。這對於合併來源不同的資料 非常有用，例如內部客戶資料和已購買人群統計資料。



「附加」節點用於連接各組記錄。也可以用於合併結構類似但資料不同的資料集。



「相異的」節點將移除複製的記錄，方法是將第一個可相異的記錄傳遞到資料串流，或者捨棄第一個記錄而將任何複製記錄傳遞到資料串流。



「多媒體串流時間序列」節點可在一個步驟中建置時間序列模型並進行評分。您可以在本端或分散式環境中使用帶有資料的節點；在分散式環境中，可以利用 IBM SPSS Analytic Server 的能力



Space-Time-Boxes (STB) 是一個 Geohash 形式的空間位置延伸。具體而言，STB 是一個英數字串，其代表分佈均勻的空間與時間區域。



「串流式 TCM」節點在一個步驟中對時間因果模型同時進行建立和分數。



CPLEX 最佳化節點可讓您透過最佳化程式設計語言 (OPL) 模型檔使用複雜的數學 (CPLEX) 式最佳化。IBM Analytical Decision Management 產品中提供了此功能，但您現在還可以在 SPSS Modeler 中使用 CPLEX 節點而不需要 IBM Analytical Decision Management。

有關 CPLEX Optimization 和 OPL 的進一步資訊，請參閱 IBM Analytical Decision Management 文件。

記錄作業選用區中的很多節點都需要使用 CLEM 表示式。如果您熟悉 CLEM，則可以在欄位中鍵入表示式。但是，所有表示式欄位都提供了一個開啟 CLEM 表示式建置器的按鈕，可以協助您自動建立此類表示式。



圖 1. 「表示式建置器」按鈕

---

## 選取節點(S)

您可以使用選取節點，根據某個特定的條件（如 BP（血壓）= "HIGH" 選取或捨棄資料串流中的部分記錄。

**眾數。** 指定將符合條件的記錄併入還是不併入在資料串流中。

- **包含。** 選擇包含符合選擇條件的記錄。
- **捨棄。** 選擇排除符合選擇條件的記錄。

**條件。** 顯示將要用於測試每個記錄的選擇條件，您可以使用 CLEM 表示式進行指定。在視窗中輸入表示式，或者按一下視窗右側的計算機（表示式建置器）按鈕，使用表示式建置器。

如果您選擇根據條件捨棄記錄，例如下列條件：

```
(var1='value1' and var2='value2')
```

依預設，「選取」節點也會捨棄所有選取欄位均為空值的記錄。為了避免這種情況，將下列條件附加到原始條件：

```
and not(@NULL(var1) and @NULL(var2))
```

「選取」節點還用於選擇記錄的比例。通常情況下，對於此作業要使用另外一個節點，樣本節點。但如果您要指定的條件比提供的參數更複合，那麼可以使用「選取」節點建立自己的條件。例如，您可以建立類似下面的條件：

```
BP = "HIGH" and random(10) <= 4
```

此條件將選取近似 40% 顯示高血壓的記錄，並向下游傳送這些記錄進行進一步分析。

---

## 樣本節點

您可以使用「樣本」節點來選取記錄的子集合進行分析，或指定要捨棄的記錄的比例。支援各種樣本類型，其中包括分層樣本、叢集樣本以及非隨機（結構化）樣本。需要使用取樣的原因有以下幾點：

- 通過評估資料子集合上的模型提高效能。通過樣本評估的模型通常與利用全部資料集得到的模型一樣準確，並且如果提高的效能容許您體驗尚未試圖的不同方法，那麼所得的模型還有可能更為準確。
- 選取相關的記錄或交易群組來進行分析，例如選取線上購物車（或市場購物籃）中的所有項目，或特定芳鄰的所有內容。
- 指定單元或觀察值以進行隨機檢驗，從而確保品質、防止欺詐和保證安全。

注意：如果僅希望將資料分割區到訓練樣本和測試樣本以進行驗證，那麼可以改用「分割區」節點。請參閱第 150 頁的『分割區節點』主題，以取得更多資訊。

樣本的類型

**叢集樣本。** 的樣本群組或叢集，而不是單個單位。例如，假設您有一個資料檔案，其中每個學生對應一條記錄。如果按學校叢集並且樣本大小為 50%，那麼將已選取一半的學校並從每所選定的學校中選出所有學生。而

去除未選取學校的學生。一般而言，您可能預期選出大約一半的學生，但由於學校規模不同，百分比也可能不太準確。同樣，您可以按交易 ID 對購物車項目進行叢集，以確保保留所選交易的所有項目。有關按鎮對內容叢集的範例，請參閱 *complexsample\_property.str* 樣本串流。

**分層樣本。**在總體或分階層的沒有重疊的子群組中自功能選擇樣本。例如，您可以確保以同樣的比例對男性和女性進行抽樣，或者可以確保在城市總體中顯示每個區域或社會經濟群體。還可以為每階層指定一個不同的樣本大小（例如，如果您認為一個群組在原始資料中被低估了）。有關按縣對內容分層的範例，請參閱 *complexsample\_property.str* 樣本串流。

**系統性或  $n$  中取 1 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。

**取樣加權。**在繪製複合樣本時會自動計算取樣加權，並且這些加權會與每個取樣單元在原始資料中所代表的「頻率」大致對應。因此，樣本的加權總和應該可以估計原始資料的大小。

### 取樣框

取樣框定義將包含在樣本或研究中的觀測對象的潛在來源。在某些情況下，取樣框可以識別總體中的每個單獨會員並且可以包含樣本中的任何會員 - 例如，對來自某條產品行的產品進行取樣。更普遍的情況是，您將無法存取每一個可能的觀測對象。例如，在選舉之前，您無法確定誰將在選舉中投票。在這種情況下，您可以將選民名冊作為取樣框，即使在下列情況下也是如此：有些登錄人不會投票，而有些人在您停止登錄時還尚未登錄，但可能會投票。您無法對取樣框之外的任何人進行取樣。取樣框是否在本質上與您嘗試評估的總體足夠相似，是必須要為每個現實的觀測對象解決的問題。

## 樣本節點選項

您可以根據需要，選擇簡單或複合方法。

### 簡單取樣選項

通過「簡單」方法，您可以選取記錄的隨機百分比、連續記錄或所有第  $n$  條記錄。

**眾數。**選取對於下面的模式傳遞（包含）還是捨棄（排除）記錄：

- **包含樣本。**包含資料串流中的選定記錄並捨棄所有其他項目記錄。例如，如果您將模式設定為納入樣本並將  **$n$  中取 1** 選項設定為 5，則每隔五個記錄便有一個記錄被包含進來，結果將生產大約為原大小五分之一的資料集。此模式為對資料進行取樣的預設模式，並且是使用複合方法時的唯一模式。
- **捨棄樣本。**排除選定記錄並包含所有其他項目記錄。例如，如果您將模式設定為捨棄樣本並將  **$n$  中取 1** 選項設定為 5，則每隔五條記錄便有一條被捨棄（排除）。此模式僅適用於簡單方法。

**樣本。**從下列選項中選取取樣方法：

- **從第一條記錄開始連續抽取。**選取此選項將使用連續資料取樣。例如，如果最大樣本大小設定為 10000，則前 10000 條記錄會被選取。
- **$n$  中取 1。**選取此選項會按照這樣的方式抽樣資料：每隔  $n$  個記錄進行一次傳遞或捨棄。例如，如果  $n$  設定為 5，則每隔五條記錄便會已選取一條。
- **隨機 %。**選取此選項將隨機抽樣指定百分比的資料。例如，如果百分比設定為 20，那麼根據選取的模式，將 20% 的資料傳送到資料串流或將其捨棄。使用該欄位可指定取樣百分比。您還可以使用設定隨機種子控制項指定一個種子值。

**使用區塊層次取樣（僅限資料庫中）。**在 Oracle 或 IBM Db2 資料庫上執行資料庫內採礦時，只在您選擇隨機百分比取樣時才啟用此選項。在這些情況，區塊層次取樣的效率會更高。

註：每次執行相同的隨機樣本設定時，系統不會傳回確切的列個數。這是因為每個輸入記錄包含在樣本中的可能性為  $N/100$ （其中， $N$  是您在節點中指定的隨機 %），而且可能性是獨立的；因此結果不是確切的  $N\%$ 。

**最大樣本大小。** 指定樣本中所包含的上限記錄數。此選項為多餘選項，因此在選定第一個和包含時會被停用。另外，當與隨機 % 選項結合使用時還請注意，此設定可能會阻止已選取某些記錄。例如，如果資料集中有一千萬條記錄，而您選取了 50% 的記錄且最大樣本大小為三百萬條記錄，那麼將已選取前六百萬條記錄中的 50% 的記錄，剩餘的四百萬條記錄便不會再被選取。為避免這種限制，請選取複合取樣方法，然後對三百萬條記錄進行隨機樣本，無需指定叢集或分層變數。

## 複合取樣選項

通過複合樣本選項，您可以與其他選項一起更好地控制樣本，包含叢集樣本、分層樣本和加權樣本。

**叢集與分層。** 容許您指定叢集和分層並根據需要輸入加權欄位。請參閱『叢集與分層設定』主題，以取得更多資訊。

**樣本類型。**

- **隨機。** 在每一階層內隨機選取叢集或記錄。
- **系統性。** 以固定間隔選取記錄。除了會根據隨機種子變更第一條記錄的位置之外，此選項工作原理與  $n$  中取 1 方法基本相似。 $n$  的值會根據樣本大小和比例自動決定。

**樣本單元。** 可以選取比例或計數作為基本樣本單元。

**樣本大小。** 您可以按以下幾種方式指定樣本大小：

- **固定。** 容許您將樣本總大小指定為計數或比例。
- **自訂。** 容許您為每個子群組或分階層指定樣本大小。此選項只有在「叢集」和「階層」子對話框中指定了層欄位時才可用。
- **變數。** 容許使用者選取一個欄位來為每個子群組或階層定義樣本大小。對於特定階層內的每條記錄，此欄位應該都有相同的值；例如，如果樣本按縣分層，那麼具有 *county = Surrey* 的所有記錄必須具有相同值。該欄位必須為數值型並且它的值必須與所選樣本單元相符。比例的值應該大於 0 少於 1；計數的最小值為 1。

**各層最小樣本。** 指定記錄的最小值（如果已指定了叢集分析欄位，可指定叢集的最小值）。

**各層最大樣本。** 指定記錄或叢集的最大值。如果在沒有指定叢集或分層欄位的情況下選取了此選項，那麼將選取指定大小的隨機或系統性樣本。

**設定隨機種子。** 當取樣或分割根據隨機百分比來記錄時，此選項可讓您將相同的結果複製到其他階段作業中。透過指定亂數產生器所用的起始值，您可以確保每次執行節點時都指派相同的記錄。輸入所需的種子值，或按一下產生按鈕來自動產生一個隨機值。如果未選取此選項，則每次執行節點時，都將產生不同樣本。

附註：搭配使用設定隨機種子選項與讀取自資料庫的記錄時，在取樣之前可能需要一個「排序」節點才能確保每次執行節點的結果相同。這是因為隨機種子取決於記錄的順序，而並不能保證順序在關聯式資料庫中保持相同。請參閱第 72 頁的『排序節點』主題，以取得更多資訊。

## 叢集與分層設定

通過「叢集」和「階層」對話框，您可以在繪製複合樣本時選取叢集、層和加權欄位。

**叢集。** 指定用於叢集記錄的種類欄位。會根據叢集成員資格對記錄進行抽樣，有些叢集包含在內，有些叢集不包含在內。但是如果包含了指定叢集的所有記錄，那麼所有的叢集都將包含在內。例如，當分析購物車中的產品關聯時，您可以按交易 ID 對項目進行叢集從而確保可以維護選定交易中的所有項目。如果改為對記錄進行取樣，則將破壞一起銷售的項目的資訊，您可以對交易進行取樣以確保已儲存選定交易的所有記錄。

**分層方式。** 指定用於分階層記錄的種類欄位，這樣將在總體或階層的沒有重疊的子群組中自功能選擇樣本。例如，如果選中一個按 50% 的比例抽取的按性別分層的樣本，那麼採用兩個按 50% 的比例抽取樣本，一個為男性，另一個為女性。例如，階層可以為社會經濟學群體、工作種類、年齡群組或種族群組，從而可以確保所關注的子群組有足夠的樣本大小。如果在原始資料集中女性人數為男性人數的三倍，那麼此比例將通過從每個群組中取樣而得以儲存。您還可以指定多個層欄位（例如，按各區域內的产品行或各個產品行所在的區域進行取樣）。

注意：如果按包含遺漏值（空白值或系統遺漏值、空字串、空白以及空白值或使用者定義的遺漏值）的欄位進行分階層，那麼無法為階層指定自訂樣本大小。如果您想要在依包含遺漏或空白值的欄位分層時使用自訂樣本大小，則需要將它們填入上游。

**使用輸入加權。** 指定在取樣之前加權記錄的欄位。例如，如果加權欄位值的範圍為 1 到 5，那麼加權為 5 的記錄被選取的機率是其他記錄的 5 倍。該欄位的值將被節點產生的最終輸出加權改寫（請參閱下列章節）。

**新的輸出加權。** 指定在未指定輸入加權欄位的情況下，記錄最終加權的欄位的名稱。（如果已指定輸入加權欄位，則上述的最終加權將取代其值，並且無法建立獨立的輸出加權欄位。）輸出加權值代表原始資料中每一個抽樣記錄所代表的記錄數。通過加權值總和，可以評估樣本的大小。例如，如果按 10% 的比例隨機抽取樣本，則所有記錄的輸出加權將為 10，代表每個抽取的記錄大體上代表原始資料中的 10 條記錄。在分層或加權樣本中，輸出加權值可能會發生變化，具體取決於每階層的樣本比例。

#### 備註(O)

- 如果無法獲得所需取樣的總體的完整清單，但可以得到某些群組或叢集的完整清單，那麼叢集取樣會非常有用。如果隨機樣本生成一系列測試主項，而要聯絡所有的主題又不切實際，那麼可以使用聚類抽樣。例如，選擇拜訪一個縣的所有農民要比選擇全國範圍內所有縣的農民要容易的多。
- 為了在每個階層內對叢集進行獨立抽樣，您可以同時指定叢集欄位和分階層欄位。例如，您可以在每個縣內對按縣分層、按鎮叢集的內容值進行抽樣。這將確保從每個縣所提取的城鎮樣本保持獨立。樣本中將包含某些鎮，而其他鎮將不會包含在其中，但對於所包含的每個鎮，鎮內的所有內容也一定包含在其中。
- 要從每個叢集內選取單元的隨機樣本，您可以將兩個「樣本」節點聯系起來。例如，如上方所述，您可以首先對按縣分層的鎮進行抽樣。然後附加另一個樣本節點並選取鎮作為分層欄位，這樣您可以在每一個鎮中按一定比例對記錄進行抽樣。
- 如果需要使用欄位組合來唯一識別叢集，那麼可以使用「衍生」節點產生新的欄位。例如，如果多個商店在交易時使用的是相同的數字系統，那麼可以衍生結合了商店 ID 和交易 ID 的新欄位。

## Strata 的樣本大小

提取分層樣本時，預設選項是對每個階層中相同比例的記錄和叢集進行抽樣。例如，如果某個群組的號碼超出另一個群組號碼的 3 倍，那麼通常希望在樣本中保留同一比例。但如果不是這種情況，那麼可以為每個階層個別指定樣本大小。

「階層的樣本大小」對話框列出了階層欄位的每個值，您可以置換階層的預設值。如果選取了多個層欄位，那麼將列出每個可能的值組合，這樣您可以指定具體的大小，例如每個城市內每一種族群組的大小，或每個縣內的每個鎮的大小。可以將大小指定為比例或計數，具體取決於「樣本」節點中現有設定。

指定階層的樣本大小

1. 在「樣本」節點，選取**複合**，然後選取一個或多個層欄位。請參閱第 65 頁的『叢集與分層設定』主題，以取得更多資訊。
2. 選取**自訂**，然後選取**指定大小...**。
3. 在「階層的樣本大小」對話框中，按一下左下角的**讀取值**按鈕移入屏幕。如有必要，您可能需要在上游來源節點或「類型」節點中實例化值。請參閱第 122 頁的『何謂實例化？』主題，以取得更多資訊。
4. 按一下任意一列以置換該階層的預設大小。

#### 有關樣本大小的注意事項

例如，如果不同的階層具有不同的變異數，為了使樣本大小與標準差成比例，自訂樣本大小可能會十分有用。（如果階層中的觀察值變化比較大，則需要抽樣更多的觀察值以獲得具有代表性的樣本。）或者階層比較小，而您可能想要使用更大的樣本比例以確保將觀察數目的下限包含在內。

注意：如果按包含遺漏值（空白值或系統遺漏值、空字串、空白以及空白值或使用者定義的遺漏值）的欄位進行分階層，那麼無法為階層指定自訂樣本大小。如果您想要在依包含遺漏或空白值的欄位分層時使用自訂樣本大小，則需要將它們填入上游。

---

## 平衡節點

您可以使用平衡節點修正資料集中的不平衡，以便它們符合指定的測試準則。例如，假設某個資料集只有兩個值 (*low* 或 *high*)，並且 90% 的觀察值為 *low*，而只有 10% 的觀察值為 *high*。很多建模技術處理此類偏倚資料都有困難，因為它們傾向於只學習這些 *low* 的結果，而忽略 *high* 的結果（因為這些結果少的可憐）。如果資料已平衡很好，*low* 和 *high* 結果具有大致相同的數目，那麼模型將更有可能找出分辨這兩個群組的型樣。這種情況下，平衡節點對於建立平衡指令，從而減少帶有 *low* 結果的觀察值數量非常有用。

平衡是通過複製記錄，然後根據指定的條件捨棄記錄完成執行的。將始終傳遞不符合任何條件的記錄。因為此過程的工作模式為複製和/或捨棄記錄，所以在下游作業中將丟失資料的原始順序。在向資料串流新增平衡節點之前，請確保衍生任何與序列相關的值。

附註：平衡節點可從分佈圖和直方圖自動產生。例如，您可以平衡資料以顯示某一種類欄位所有種類的相同比例，如分佈圖形所示。

**範例。** 建立 RFM 串流以識別積極回應以往營銷活動的最新客戶時，銷售公司的市場行銷處可以使用平衡節點來平衡資料中 true 假回應之間的差異。

### 為平衡節點設定選項

**記錄平衡指引。** 列出目前平衡指引。每個指令都包含一個因子和一個條件，該條件告知軟體「在該條件為 true 的情況下以指定的因子增加記錄比例」。如果因子小於 1.0，那麼表示指定記錄的比例要降低。例如，如果您要減少治療藥為藥物 Y 的記錄數，則可以使用因子 0.7 和條件 Drug = "drugY" 建立一個平衡指令。此指令表示對於所有下游作業，治療藥為藥物 Y 的記錄數將減少到 70%。

注意：用於減少的平衡因子可以指定為四位小數。小於 0.0001 的因子設定會產生錯誤，因為這樣的結果無法正確計算。

- **建立條件**，此操作通過按一下該文字欄位右側的按鈕完成。此操作將插入一個用於輸入新條件的空白列。要為條件建立 CLEM 表示式，請按一下表示式建置器按鈕。
- **刪除指引**，此操作通過使用紅色刪除按鈕完成。
- **對指引排序**，此操作通過上下箭頭按鈕完成。

僅平衡訓練資料。如果串流中出現分割區欄位，那麼此選項僅平衡訓練分割區中的資料。尤其是，當產生需要不平衡測試或驗證分割區的調整傾向分數時，此選項非常有用。如果串流中不出現分割區欄位（或已指定多個分割區欄位），那麼將忽略此選項並已平衡所有的資料。

## 「聚合」節點

聚集是一項資料預備作業，經常用於減小資料集的大小。繼續執行聚集之前，您應該花一些時間來清理資料，尤其要關注遺漏值。完成聚合後，或許會丟失可能有用的遺漏值資訊。

您可以使用「聚合」節點將一系列輸入記錄取代為聚集的摘要輸出記錄。例如，您可能有一組輸入銷售記錄，例如下表格中所顯示的記錄。

表 13. 小時記錄輸入範例

年齡	性別	地區	分支	銷售量
23	M	S	8	4
45	M	S	16	4
37	M	S	8	5
30	M	S	5	7
44	M	N	4	9
25	M	N	2	11
29	F	S	16	6
41	F	N	4	8
23	F	N	6	2
45	F	N	4	5
33	F	N	6	10

您可以將 *Sex* 和 *Region* 作為索引鍵欄位對這些記錄進行聚合。然後選擇使用平均數模式聚合年齡，並使用合計模式聚合銷售。在「聚合」節點對話框中選取在欄位中包含記錄計數後，聚合的輸出將顯示在下表格中。

表 14. 聚集記錄範例

年齡（平均數）	性別	地區	銷售量（總和）	記錄計數
35.5	F	N	25	4
29	F	S	6	1
34.5	M	N	20	2
33.75	M	S	20	4

例如，您可從中瞭解到，北部區域四名陰性銷售人員的平均值年齡為 35.5 歲，其銷售量總和為 25 件產品。

註：如果未指定聚合方式，那麼將自動捨棄分支之類的欄位。

## 「聚合」節點的設定選項

在「聚合」節點上，您可以指定下列內容。

- 一個或多個用作聚集種類的索引鍵欄位
- 一個或多個要為其計算聚合值的聚集欄位
- 一種或多種聚集模式（聚集類型），用於每個聚集欄位的輸出



您還可以指定用於新增欄位的預設聚集模式，並使用表示式（類似於公式）對聚集進行分類。

請注意，為了增加效能，啟用平行處理可能會有益於聚集作業。

**索引鍵欄位。**列出可用作聚集種類的欄位。連續（數值）欄位和種類欄位都可用作索引鍵。如果您選擇多個索引鍵欄位，那麼這些值將進行合併，以生成用於聚集記錄的索引鍵值。對於每個唯一的索引鍵欄位，將會產生一條聚合記錄。例如，如果 *Sex* 和 *Region* 是索引鍵欄位，那麼 *M* 和 *F* 與 *N* 和 *S* 區域的每個唯一組合（四個唯一組合）都將具有一條聚合記錄。要新增索引鍵欄位，請使用視窗右側的欄位選擇器按鈕。

對話框的剩餘部分分為兩個主要區域 - **基本聚合和聚集表示式**。

## 基本聚集

**聚集欄位。**列出將聚集其值的欄位以及所選的聚集方式。要向此清單中新增欄位，請使用右側的「欄位選擇器」按鈕。可用的聚集模式如下。

註：某些模式不適用於非數值欄位（例如，**總和**不適用於日期/時間欄位）。不能用於所選聚合欄位的模式將被停用。

- **總和。** 選取此選項可傳回每個索引鍵欄位組合的合計值。總和是遍及所有包含非遺漏值之觀察值的值總計。
- **平均數。** 選取此選項可傳回每個索引鍵欄位組合的平均數。平均數是集中趨勢的量數，是算數平均值（總和除以觀察值數）。
- **最小。** 選取此選項可傳回每個索引鍵欄位組合的最小值。
- **最大。** 選取此選項可傳回每個索引鍵欄位組合的最大值。
- **標準差。** 選取此選項可傳回每個索引鍵欄位組合的標準差。標準差是對圍繞平均數的離差的測量，該值等於變異數測量結果的平方根。
- **中位數。** 選取此選項可傳回每個索引鍵欄位組合的中值。中位數為集中趨勢的量數，其不會感應到偏離值（相反地，平均數會受到幾個高低極端值的影響）。其亦稱作第 50 個百分位數或第 2 個四分位數。
- **計數。** 選取此選項可傳回每個索引鍵欄位組合的非空值計數。
- **變異數。** 選取此選項可傳回每個索引鍵欄位組合的變異值。變異數是對圍繞平均數的離差的測量，該值等於平均離差的平方和除以觀察值數減一。
- **第 1 個四分位數。** 選取此選項可傳回每個索引鍵欄位組合的第 1 個四分位數（第 25 個百分位數）值。
- **第 3 個四分位數。** 選取此選項可傳回每個索引鍵欄位組合的第 3 個四分位數（第 75 百分位數）值。

註：在執行包含「聚合」節點的串流的情況下，將 SQL 推送回 Oracle 資料庫時針對第一個四分位數和第 3 個四分位數傳回的值可能與原生模式下傳回的那些值不同。

**預設方式。**指定要用於新增欄位的預設聚集方式。如果您經常使用同一聚集，請在此處選取一個或多個方式，然後使用右側的「全部套用」按鈕，以便將所選方式套用至上方列出的所有欄位。

**新欄位名稱延伸。** 選取此選項可新增字尾或字首（例如，"1" 或 "new"）以複製聚合的欄位。例如，如果您選取了字尾選項，並將 "1" 指定為延伸，那麼針對欄位年齡的最小值聚集結果將生成欄位名稱 *Age\_Min\_1*。註：聚集延伸（例如，*\_Min* 或 *\_Max*）將自動新增至新欄位，以指示所執行的聚集的類型。選取字尾或字首以指出偏好的延伸樣式。

**包含欄位中的記錄計數。**依預設，選取此選項可在每條輸出記錄中包含一個額外的欄位 *Record\_Count*。此欄位指出聚集多少個輸入記錄來形成每一個聚集記錄。在編輯欄位中鍵入內容，以便為此欄位建立自訂名稱。

註：計算聚合時將排除系統空值，但它們會併入在記錄計數中。另一方面，空白值既併入在聚集中也併入在記錄計數中。要排除空白值，您可以使用「填入器」節點將空白值取代為空值。您還可以使用「選取」節點移除空白值。

## 聚集表示式

表示式類似於根據值、欄位名稱、運算子和函數建立的公式。聚集表示式與函數不同之處在於，函數一次只能操作一條記錄，而聚集表示式能夠對記錄群組、記錄集或記錄收集進行操作。

註：只有在串流包含資料庫連線（通過「資料庫來源」節點），您才能建立聚集表示式。

將新的表示式建立為衍生欄位；要建立表示式，請使用表示式建置器中提供的資料庫聚合函數。

有關表示式建置器的進一步資訊，請參閱《IBM SPSS Modeler 使用者手冊》(ModelerUsersGuide.pdf)。

請注意，由於聚集表示式按索引鍵欄位進行群組，因此索引鍵欄位與您建立的任何聚集表示式之間存在關聯。

有效聚集表示式可評估為聚集結果；以下是幾個有效聚集表示式範例以及用於控管它們的規則：

- 您可以使用標量函數將多個聚集函數組合在一起，以生成單個聚集結果。例如：

```
max(C01) - min(C01)
```

- 聚集函數可對多個標量函數的結果執行操作。例如：

```
sum (C01*C01)
```

## 聚合最佳化設定

在「最佳化」標籤上，您可以指定下列內容。

**鍵是連續的。**如果您知道鍵值相同的所有記錄都在輸入中群組在一起（例如，如果對索引鍵欄位上的輸入進行排序）。這樣做可以改善效能。

**容許使用中位數和四分位數的近似值。**在 Analytic Server 中正在處理資料時，目前不支援順序統計資料（中位數、第 1 個四分位數和第 3 個四分位數）。如果您使用的是 Analytic Server，那麼可以選中此勾選框以改為使用這些統計資料的近似值，而近似值的計算方式如下：對資料進行 Bin 然後根據各個 Bin 之間的分佈為統計資料計算估計值。依預設，未勾選此選項。

**Bin 個數。**僅當選中了容許使用中位數和四分位數的近似值勾選框時，此選項才可用。選取對統計資料進行估計時使用的 Bin 個數；Bin 個數會影響**最大錯誤百分比**。依預設，Bin 個數為 1000，這對應於最大錯誤，即範圍的 0.1%。

---

## RFM 「聚合」節點

通過近因、頻率、貨幣 (RFM) 「聚合」節點，您可以利用客戶的歷程記錄交易資料，去除所有無用的資料，然後一個將他們的所有剩餘交易資料合併到一列並以唯一的客戶 ID 作為關鍵字，從而列出他們最後一欄一次與您交易的時間（近因），交易的次數（頻率）以及這些交易的總值（貨幣）。

繼續執行任一聚集之前，應該花一些時間來清理資料，尤其要關注所有遺漏值。

一旦使用「RFM 整合」節點 ID 和轉換資料之後，您可以使用「RFM 分析」節點執行進一步分析。請參閱第 147 頁的『RFM 分析節點』主題，以取得更多資訊。

請注意，如果已通過「RFM 整合」節點執行資料檔案，那麼資料檔案將不會再具有任何目標值；因此，在利用它作為使用所有建模節點（如 C5.0 或 CHAID）進行進一步預測分析的輸入之前，需要將其與其他客戶的資料進行合併（例如，通過比對用戶 ID）。請參閱第 72 頁的『合併節點』主題，以取得更多資訊。

將 IBM SPSS Modeler 中的 RFM 「聚合」節點和 RFM 分析節點設定為使用獨立的分級；即，它們分別按近因、頻率、貨幣值對資料進行排等級和分級，而無需考慮它們的值或其他兩種標準。

## 為 RFM 「聚合」節點設定選項

「RFM 整合」節點的「設定」標籤包含下列欄位。

**近因計算的相對日期** 指定計算交易近因的日期。該日期可以是您輸入的**固定日期**，也可以是系統設定的**當前日期**。當前日期由系統預設已輸入，並在執行節點時自動更新。

註：在不同的語言環境中，**固定日期**的顯示可能有所不同。例如，如果值 2007-8-10 在串流中儲存為 Fri Aug 10 00:00:00 CST 2007，那麼這是時區 'UTC+8' 中的時間和日期。但是，在時區 'UTC-8' 中顯示為 Thu Aug 9 12:00:00 EDT 2007。

**ID 是連續的** 如果您的資料進行了預先排序，以便所有具有同一 ID 的記錄一起出現在資料串流中，那麼選取此選項可以加快正在處理速度。如果您的資料未經預先排序（或者您不確定），請將此選項保持不選狀態，則該節點將自動對該資料進行排序。

**ID** 選取該欄位以用來識別客戶及其交易。若要顯示您可從中選取的欄位，請使用右側的「欄位選擇器」按鈕。

**日期** 選取將要用來計算近因的日期欄位。若要顯示您可從中選取的欄位，請使用右側的「欄位選擇器」按鈕。

請注意，這需要具有適當格式的儲存日期或時間戳記的欄位以用作輸入。例如，如果某個字串欄位具有 Jan 2007、Mar 2007 等類似值，則可以使用填入器節點或 to\_date() 函數將其轉換為日期欄位。請參閱第 138 頁的『使用填入器節點進行儲存類型轉換』主題，以取得更多資訊。

**值** 選取該欄位以用來計算客戶交易的總貨幣值。若要顯示您可從中選取的欄位，請使用右側的「欄位選擇器」按鈕。注意：該值必須是一個數值。

**新欄位名稱延伸** 選取該欄位可將字首或字尾（如 "12\_month"）附加到新產生的近因、頻率和貨幣欄位。選取字尾或字首以指出偏好的延伸樣式。例如，這在檢查多個時間期間時將可能有用。

**捨棄具有以下方值的記錄** 如有需要，可在計算 RFM 總計時指定一個最小值，凡低於該值的交易詳細資料都不再被使用。該值單元與所選的**值**欄位相關。

**僅包含最新的交易** 如果分析的是大型資料庫，那麼可以指定只使用最近的記錄。無論是在某個特定的日期之後還是在最近的期間內，您都可以選擇使用記錄的資料：

- **該交易日期之後** 指定交易日期以在分析時包含其之後的記錄。
- **以下過去時間段內的交易** 指定從近因計算的**相對日期**開始之前的期間數和期間類型（天、周、月或年），在此日期之最後一個的記錄將被包含在您的分析中。

**儲存第二個最新交易的日期** 如果希望瞭解每個客戶第二個最近交易的日期，請選中此框。此外，您還可以選取**儲存第三個最近交易的日期**複選框。例如，這樣有助於您識別在很長一段時間之前進行多數交易的客戶，但僅限於一個最近交易。

---

## 排序節點

您可以使用「排序」節點，根據一個或多個欄位的值，按照遞增或者遞減對記錄進行排序。例如，「排序」節點經一般於檢視和選取帶有最常見資料值的記錄。通常情況下，您首先要使用「聚合」節點聚集資料，然後使用「排序」節點按照記錄計數的遞減對聚合後的資料進行排序。如果在一個表格中顯示這些結果，您則可以探索這些資料並作出決策，如選取前 10 個最佳客戶的記錄。

「排序」節點的「設定」標籤包含下列欄位。

**依次排序。** 在表格中顯示所有選作排序關鍵的欄位。如果索引鍵欄位為數值欄位，那麼它最適用於排序。

- 使用右側的「自動選擇器」按鈕向此清單**新增欄位**。
- 通過按一下表格中順序欄中的**遞增或遞減箭頭**來選取順序。
- 使用紅色的刪除按鈕來刪除欄位。
- **對指引排序**，此操作通過上下箭頭按鈕完成。

**預設排序順序。** 選取遞增或遞減用作在上方新增新欄位時的預設排序。

註：如果模型串流的下游中存在「相異的」節點，那麼不會套用「排序」節點。有關「相異的」節點的資訊，請參閱第 80 頁的『差異節點』。

## 排序最佳化設定

如果您要對您知道已經按照某些索引鍵欄位排序的資料進行操作，那麼可以指定哪些欄位已經排序，從而使得系統能夠更高效地對剩下的資料進行排序。例如，您要按照 *Age*（遞減）和 *Drug*（遞增）進行排序，但知道這些資料已經按照 *Age*（遞減）進行了排序。

**資料預先經過排序。** 指定資料是否已經按照一個或多個欄位進行排序。

**指定現有排序。** 指定已排序的欄位。使用「選取欄位」對話框將欄位新增至清單。在順序直欄中，指定每個直欄是以遞增還是遞減順序排序。如果您要指定多個欄位，請確保以正確的排序列出這些欄位。使用清單右側的箭頭以正確的順序來排列欄位。如果您在指定正確的現有排序時犯錯，則當您執行串流時將顯示一個錯誤，顯示排序與您指定的順序不一致的記錄編號。

注意：啟用平行處理有益於提高排序速度。

---

## 合併節點

「合併」節點的功能是採用多個輸入記錄，然後建立一個包含全部或其中部分輸入欄位的輸出記錄。當您要合併來源不同的資料（如內部客戶資料和購買的人口統計資料）時，此作業非常有用。可以通過下列方式合併資料。

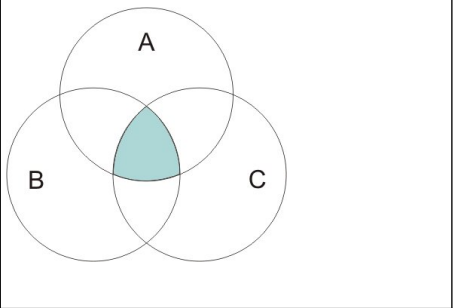
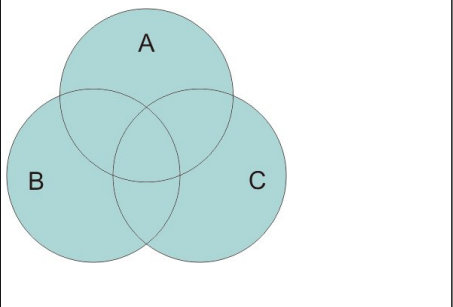
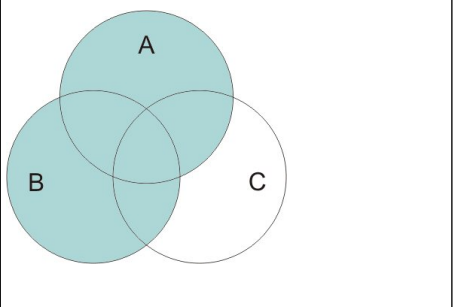
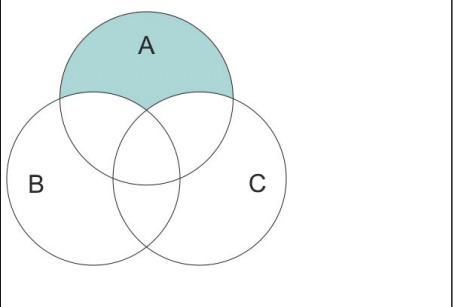
- 按**順序合併**按輸入順序並置來自所有來源的相應記錄，直到窮盡最小的資料來源為止。如果使用此選項，務必確保已使用排序節點完成了對資料的排序。
- 使用**索引鍵欄位**（例如客戶 *ID*）的合併，可指定如何將一個資料來源的記錄與其他資料來源的記錄進行比對。結合的類型有許多，其中包含內部結合、完全外部加入、局部外部結合和反結合。請參閱第 73 頁的『結合類型』主題，以取得更多資訊。
- 按**條件合併**意味著您可以指定執行合併前所需滿足的條件。可以在節點中直接指定條件，也可以使用「表示式建置器」建立條件。
- 按**排名式條件合併**是一個左端外部結合，在此結合中，您指定執行合併前所需滿足的條件，並指定用於實現從低到高排序的排名表示式。這種合併最常用於合併地理空間資料，您可以直接在節點中指定條件，也可以使用表示式建置器來建立條件。

## 結合類型

當資料合併使用一個索引鍵欄位時，最好先花一些時間來考慮要排除和併入哪些記錄。結合的類型有很多種，詳細資訊將在下方討論。

兩種基本的結合類型稱為內部結合和外部加入。這些方法經一般於根據索引鍵欄位（如 客戶 ID）的一般值，合併來自相關資料集的表格。通過內部結合，可以實現清理合併，以及僅包含完整記錄的輸出資料集。外部加入也包含合併資料中的完整紀錄，但它們還容許包含來自一個或多個輸入表格的唯一性資料。

以下內容詳細介紹了容許的結合類型。

	<p><b>內部結合</b> 只包含其中索引鍵欄位的值對於所有輸入表格都共有的記錄。即，不符合的記錄不會併入在輸出資料集中。</p>
	<p><b>完全外部結合</b> 包含輸入表格中的所有記錄，既有相符的記錄也有不相符的記錄。左外部加入和右側外部部結合稱為局部外部結合，將在下方說明。</p>
	<p><b>局部外部結合</b> 包含使用索引鍵欄位符合的所有記錄，以及指定的表格中的不符合記錄。（換句話說，包括部分表格中的所有記錄，以及其他表格中的僅相符記錄。）使用「合併」標籤上的「選取」按鈕，可選取要包括在外部加入中的表格（如此處顯示的 A 和 B）。如果只合併兩個表格，局部結合也稱為左外部加入或右側外部部結合。因為 IBM SPSS Modeler 容許合併兩個以上的表格，所以我們稱此為局部外部結合。</p>
	<p><b>反結合</b> 僅包含第一個輸入表格（此處顯示的表A）的不符合記錄。這種結合類型與內部結合正好相反，在輸出資料集中不包含完整紀錄。</p>

例如，如果您在一個資料集中包含有關農場的資訊，在另一個資料集中包含農場相關的保險索賠資訊，則可以使用合併選項將第一個來源中的記錄與第二個來源相比對。

要確定您農場樣本中的客戶是否已經提出了保險索賠，請使用內部結合選項傳回一個清單，其中顯示兩個樣本中所有 ID 相符的記錄。

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalue
1	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0...
2	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0...
3	id620	name620	north	880.000	74.000	6.000	426988.000	rapeseed	arable_de	118885.0...

圖 2. 內部結合合併的輸出示例

使用完全外部加入選項既會傳回輸入表格中的相符記錄也會傳回不相符的記錄。對於任何不完整的值，都將使用系統遺漏值 (\$null\$)。

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalue
1	id601	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	decomm...	74703.1C
2	id602	name602	north	1780.000	42.000	9.000	734118.000	maize	\$null\$	\$null\$
3	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0
4	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0
5	id606	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	arable_de	122135.0

圖 3. 完全外部加入合併的輸出示例

局部外部結合包含使用索引鍵欄位符合的所有記錄，以及指定的表格中的不符合記錄。該表格顯示了 ID 欄位中所有符合的記錄，以及第一個資料集中符合的記錄。

	id	claimtype	claimvalue	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	\$null\$	\$null\$	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id604	decomm...	281082.0...	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes
3	id605	decomm...	122006.0...	name605	north	1700.000	46.000	8.000	621148.000	wheat
4	id607	\$null\$	\$null\$	name607	southeast	1820.000	29.000	6.000	211605.000	maize
5	id608	\$null\$	\$null\$	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
6	id609	\$null\$	\$null\$	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
7	id615	\$null\$	\$null\$	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
8	id618	\$null\$	\$null\$	name618	southeast	1180.000	98.000	3.000	368646.000	maize

圖 4. 局部外部結合合併的輸出示例

如果使用反結合選項，該表格則只傳回第一個輸入表格的不符合記錄。

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id607	name607	southeast	1820.000	29.000	6.000	211605.000	maize
3	id608	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
4	id609	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
5	id615	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
6	id618	name618	southeast	1180.000	98.000	3.000	368646.000	maize
7	id619	name619	north	840.000	64.000	8.000	457552.000	potatoes

圖 5. 反結合合併的輸出示例

## 指定合併方法和鍵值

「合併」節點的「合併」標籤包含下列欄位。

**合併方法** 請選取用於合併記錄的方法。選取鍵或條件將啟動此對話框的下半部分。

- **順序** 按順序合併記錄，以便將每個輸入中的第  $n$  個記錄合併到一起，從而生第  $n$  個輸出記錄。當任何記錄用完相符輸入記錄後，將不會再生成任何輸出記錄。這意味著，建立的記錄數是最小資料集中的記錄數。
- **鍵** 使用鍵欄位（例如交易 ID）將鍵欄位中的值相同的記錄合併。它相當於資料庫 "equi-join"。如果索引鍵值出現多次，那麼傳回所有可能的組合。例如，如果具有相同鍵欄位值  $A$  的記錄的其他欄位中包含不同的值  $B$ 、 $C$  和  $D$ ，那麼合併後的欄位對於  $A$  與值  $B$ 、 $A$  與值  $C$  以及  $A$  與值  $D$  的每個組合都將生成一個個別的記錄。

注意：在按索引鍵合併的方法中，空值不會被視為相同的值，因此不會結合。

- **條件** 使用此選項可以指定合併條件。如需相關資訊，請參閱『指定合併的條件』。
- **排名式條件** 使用此選項可以指定是否對主資料集與所有輔助資料集中的每一對列進行合併；使用排名表示式可以將任意的多個相符項按從低到高順序排列。如需相關資訊，請參閱第 76 頁的『為「合併」指定排名式條件』。

**可能的鍵** 僅列出那些在所有輸入資料來源中都有完全相符的欄位名稱的欄位。從此清單中選取一個欄位，並使用箭頭按鈕將其新增為用於合併記錄的索引鍵欄位。可以使用多個鍵欄位。您可以使用「過濾器」節點或者來源節點的「過濾器」標籤對不相符的輸入欄位進行更名。

**用於合併的鍵** 根據鍵欄位的值，列出所有用於對來自所有輸入資料來源的記錄進行合併的欄位。若要從清單中移除索引鍵，請選取一個金鑰並使用方向鈕將它傳回至「可能的索引鍵」清單。選取多個索引鍵欄位時，會啟用下列選項。

**組合複製鍵欄位** 在上方選取了多個鍵欄位之後，此選項確保只有一個具有該名稱的輸出欄位。依預設，此選項為啟用狀態，但已從以前版本的 IBM SPSS Modeler 匯入串流的情況下除外。如果停用了此選項，那麼必須使用「合併」節點對話框中的「過濾器」標籤將複製的鍵欄位更名或排除。

**僅包含相符的記錄（內結合）** 選取此項將僅合併完整的記錄。

**包含相符的記錄和不相符的記錄（全外部結合）** 選取此項將執行「全外部結合」。這意味著，如果索引鍵欄位值沒有出現在所有輸入表格中，那麼將仍然保留不完整記錄。未定義的值 (\$null\$) 會新增到索引鍵欄位，並併入在輸出記錄中。

**包含相符記錄和不相符記錄（局部外部結合）** 選取此項表示對子對話框中選取的表格執行「局部外部結合」。按一下**選取**可指定將在合併中為其保留不完整記錄的表格。

**包含第一個資料集中與任何其他記錄都不相符的記錄（反結合）** 選取此項表示執行某種「反結合」，在此類結合中，只有第一個資料集中的不相符記錄將傳送到下游。您可以使用「輸入」標籤上的箭頭指定輸入資料集的順序。這種結合類型在輸出資料集中不包含完整紀錄。如需相關資訊，請參閱第 73 頁的『結合類型』。

### 選取用於局部結合的資料

對於局部外部結合，您必須選取要為其保留不完整記錄的表格。例如，您可能想保留 Customer 表格中的所有記錄，同時僅保留 Mortgage Loan 表格中符合的記錄。

**外部加入欄。** 在外部加入欄中，選取要作為整體包含在內的資料集。對於局部結合，重疊的記錄以及此處已選取的資料集的不完整記錄都將被保留。請參閱第 73 頁的『結合類型』主題，以取得更多資訊。

### 指定合併的條件

通過將合併方法指定為條件，可指定要執行合併必須滿足的一項或多項條件。

可以直接在「條件」欄位中輸入條件，也可按一下此欄位右側的計算機圖標以借助表示式建置器建立條件。

**向重複欄位名稱新增標籤以避免合併衝突** 如果兩個以上要合併的資料集包含相同的欄位名稱，勾選此勾選框可以在欄位欄標題的開頭新增另一個字首標籤。例如，如果存在兩個名為 *Name* 的欄位，那麼合併結果將包含 *1\_Name* 和 *2\_Name*。如果在資料來源中將該標籤更名，那麼將使用新名稱，而不是使用進行了編號的字首標籤。如果您未選取此勾選框，且資料中存在重複的名稱，則會在勾選框右側顯示一個警告。

## 為「合併」指定排名式條件

您可以將「排名式條件」合併視為按條件進行的左端外部結合合併；此合併的左端是主資料集，其中的每筆記錄都是一個事件。例如，對於用來在罪案資料中尋找型樣的模型，主資料集中的每個記錄都是一項罪案及其相關資訊（地點和類型等等）。在此範例中，右端可能包含相關的地理空間資料集。

此合併同時使用合併條件和排名表示式。合併條件可以使用 *within* 或 *close\_to* 之類的地理空間函數。在合併期間，右端資料集中的所有欄位都將新增到左端資料集，但是多項相符將產生清單欄位。例如：

- 左端：罪案資料
- 右端：城鎮資料集和道路資料集
- 合併條件：城鎮內 (*within*) 且鄰近 (*close\_to*) 道路的罪案資料，以及所謂「鄰近」(*close\_to*) 的定義。

在此範例中，如果罪案發生在三條道路（要傳回的相符項數至少設定為 3）的所需鄰近 *close\_to* 距離內，那麼將全部三條道路作為清單項目傳回。

通過將合併方法設定為**排名式條件**，可以指定執行合併前必須滿足的一項或多項條件。

**主要資料集** 請選取用於合併的主要資料集；所有其他資料集中的欄位都將新增到您選取的資料集中。這可視為外部結合合併的左側。

您選取主資料集時，所有其他連接至「合併」節點的輸入資料集都將自動列示在**合併表格**中。

**向重複欄位名稱新增標籤以避免合併衝突** 如果兩個以上要合併的資料集包含相同的欄位名稱，勾選此勾選框可以在欄位欄標題的開頭新增另一個字首標籤。例如，如果存在兩個名為 *Name* 的欄位，那麼合併結果將包含 *1\_Name* 和 *2\_Name*。如果在資料來源中將該標籤更名，那麼將使用新名稱，而不是使用進行了編號的字首標籤。如果您未選取此勾選框，且資料中存在重複的名稱，則會在勾選框右側顯示一個警告。

## 合併

### 資料集

顯示作為輸入連接到「合併」節點的輔助資料集的名稱。預設情況下，存在多個輔助資料集時，這些資料集將按它們連接至「合併」節點的順序列出。

### 合併條件

請輸入用於將表格中各個資料集與主資料集合併的唯一條件。您可以直接在資料庫中輸入條件，或是在「表示式建置器」的協助下透過按一下資料庫右側的計算機符號來建置條件。例如，您可以使用地理空間述詞來建立合併條件，用於將一個資料集中的罪案資料放入另一個資料集的城鎮資料。預設合併條件取決於地理空間測量層次，如以下方清單所示。

- 「點」、「線串」、「多點」和「多線串」- *close\_to* 的預設條件。
- 「多邊形」和「多多邊形」- *within* 的預設條件。

有關這些層次的進一步資訊，請參閱第 120 頁的『地理空間測量子層次』。

如果一個資料集包含多個不同類型的地理空間欄位，那麼使用的預設條件取決於按下列遞減在資料中找到的第一個測量層次。

- 點
- 線串



- 多邊形

註：唯有在次要資料庫中存在地理空間資料欄位時，預設值才可用。

### 排名表示式

請指定一個表示式，用於對資料集的合併進行排名；此表示式將根據排名準則對多個相符項進行排序。您可以直接在資料庫中輸入條件，或是在「表示式建置器」的協助下透過按一下資料庫右側的計算機符號來建置條件。

距離和面積的預設排名表示式在表示式建置器中提供，這兩種表示式都從低到高排名，例如，這表示距離的最上層相符項是最小的值。下面是按距離進行排名的一個範例：主資料集包含罪案及其相關地點，另外每個資料集都包含具有地點的物件；在這種情況下，罪案與物件之間的距離可以用作排名準則。預設排名表示式取決於地理空間測量層次，如以下方清單所示。

- 「點」、「線串」、「多點」和「多線串」- 預設表示式為 *distance*。
- 「多邊形」和「多多邊形」- 預設表示式為 *area*。

註：唯有在次要資料庫中存在地理空間資料欄位時，預設值才可用。

### 相符項數

根據條件和排名表示式，指定傳回的相符項數。預設相符項數取決於輔助資料集中的地理空間測量層次，如以下方清單所示；但是，您可以在 Cell 中按兩下以輸入自己的值，上限為 100。

- 「點」、「線串」、「多點」和「多線串」- 預設值為 3。
- 「多邊形」和「多多邊形」- 預設值為 1。
- 不包含地理空間欄位的資料集 - 預設值為 1。

例如，如果您設定了基於合併條件 *close\_to* 和排名表示式 *distance* 的合併，那麼主資料集中的每個記錄在輔助資料集中的前三個（最鄰近）相符項將作為結果清單欄位中的值傳回。

### 過濾合併節點中的欄位

「合併」節點包含了一種簡便方法，用來過濾或重新命名由於合併多個資料來源導致的重複欄位。按一下對話框中的過濾器標籤可選取過濾選項。

此處顯示的選項幾乎與「過濾器」節點的選項完全相同。但還有其他選項在「過濾」功能表中存在，而此處沒有討論。請參閱第 129 頁的『過濾或重新命名欄位』主題，以取得更多資訊。

**欄位。**顯示目前連接的資料來源中的輸入欄位。

**標籤。**列出與資料來源鏈結相關聯的標籤名稱（或編號）。按一下輸入標籤可更改到此合併節點的作用中鏈結。

**來源節點。**顯示要合併其資料的來源節點。

**已連接的節點。**顯示與「合併」節點連接的節點的節點名稱。複合的資料採礦經常需要若干可能包含同一個來源節點的合併或附加作業。連接的節點名稱提供一種方式來區分這些內容。

**過濾。**顯示輸入欄位和輸出欄位之間的現行連線。作用中連線會顯示一個未切斷的箭頭。帶有紅色 X 的連線表示經過過濾的欄位。

**欄位。**列出合併或附加之後的輸出欄位。複製欄位顯示為紅色。按一下上方的「過濾器」欄位可停用重複的欄位。

**檢視現行欄位。**選取此選項可檢視被選作索引鍵欄位的欄位資訊。

檢視未用的欄位設定。選取此選項可檢視目前未使用的欄位的相關資訊。

## 設定輸入順序和標籤

使用合併節點和附加節點對話框中的「輸入」標籤，可以指定輸入資料來源的順序，還可以對每個來源的標籤名稱進行任意變更。

輸入資料集的標籤與順序。選取此選項將只合併或附加完整的記錄。

- **標籤。**列出每個輸入資料來源的目前標籤名稱。標籤名稱（即標籤）是一種唯一識別用於合併或附加作業的資料鏈結的方法。例如，這就好像來自不同管道的水在一個點處進行合併，然後流程到單個管道中。IBM SPSS Modeler 中的資料也按照相似的方式流動，合併點通常是不同資料來源之間的複合互動。標籤提供了一種用於管理「合併」節點或「附加」節點的輸入（「管道」）的方法，因此，如果儲存或切斷該節點，這些鏈結將被保留並可以輕鬆識別。

將附加資料來源與「合併」節點或「附加」節點相連時，將使用編號自動建立預設標籤，以代表您連接這些節點的順序。此順序與欄位在輸入或輸出資料集中的順序無關。通過在 標籤欄中輸入新名稱，可以變更預設標籤。

- **來源節點。**顯示要合併其資料的來源節點。
- **已連接的節點。**顯示與「合併」節點或「附加」節點連接的節點的節點名稱。複合的資料採礦經常需要若干可能包含同一個來源節點的合併作業。連接的節點名稱提供一種方式來區分這些內容。
- **欄位。**列出每個資料來源中的欄位個數。

檢視目前標籤。選取此選項可檢視正在由「合併」節點或「附加」節點使用的作用中標籤。換而言之，目前標籤識別指向有資料流程過的節點的鏈結。用管道比喻一下，目前標籤就相當於現在有水流程過的管道。

檢視 **UNUSED** 的標籤設定。選取此選項可檢視以前用於連接「合併」節點或「附加」節點、但目前未與資料來源連接的標籤（或鏈結）。這就相當於排水系統中仍然存在的空白管道。您可以選擇將這些「管道」與新來源連接，也可以選擇將其移除。要從節點刪除 **UNUSED** 的標籤，請按一下清除。此操作將馬上清除所有 **UNUSED** 的標籤。

## 合併最佳化設定

系統提供了兩個選項，可協助您在特定的狀況下以更高效的方式合併資料。通過這些選項，您可以在一個輸入資料集明顯大於其他資料集，或者您的資料已經按照將要用於合併的所有或部分索引鍵欄位進行排序的情況下最佳化合併。

註：從此標籤中進行的最佳化僅適用於 IBM SPSS Modeler 原生節點執行（即，「合併」節點不推送回到 SQL）。最佳化設定對 SQL 產生不起作用。

一個輸入資料集相當大。選取此選項可表明其中一個輸入資料集比其他資料集大很多。系統會在記憶體中快取較小的資料集，然後在不快取或不對其進行排序的情況下正在處理較大的資料集來執行合併。您經常會對使用星狀綱目或相似方案設計的資料使用這種類型的結合，這種資料中會存在一個較大的共用資料中心表格（例如交易資料）。如果選取此選項，請按一下選取指定該較大資料集。請注意，您只能選取一個大的資料集。下表彙總了哪些結合可以通過此方法進行最佳化。

表 15. 結合最佳化的摘要。

結合類型	是否可針對大型輸入資料集進行最佳化？
Inner	是
Partial	如果大的資料集中沒有不完整記錄，是。
完整	否
反結合	如果大的資料集是第一個輸入，是。

**所有輸入已按索引鍵欄位排序。** 選取此選項可表明輸入資料已經按照將要用於合併的一個或多個索引鍵欄位進行排序。請確保 所有 輸入資料集均已排序。

**指定現有排序。** 指定已排序的欄位。使用「選取欄位」對話框將欄位新增至清單。您可以僅從將要用於合併的索引鍵欄位（在「合併」標籤中指定）中選取。在順序 直欄中，指定每個直欄是以遞增還是遞減順序排序。如果您要指定多個欄位，請確保以正確的排序列出這些欄位。使用清單右側的箭頭以正確的順序來排列欄位。如果您在指定正確的現有排序時犯錯，則當您執行串流時將顯示一個錯誤，顯示排序與您指定的順序不一致的記錄編號。

根據資料庫使用的排序方法是否區分觀察值，當有一個或多個輸入由資料庫排序時，最佳化可能不會正常工作。例如，如果有兩個輸入分別為區分觀察值和不區分觀察值，那麼排序結果可能有所不同。合併最佳化將導致使用記錄排序後的順序來處理記錄。因此，如果輸入採用不同的排序方法來排序，則合併節點會報告錯誤，並顯示排序不一致處的記錄編號。如果所有輸入均來自相同來源，或使用互容的排序方法來排序，則可以成功合併記錄。

注意：啟用平行正在處理有益於提高合併速度。

---

## 附加節點

您可以使用 Append 節點連接記錄集。「合併」節點將不同來源的記錄結合在一起，而「附加」節點與之不同，它讀取一個來源中的所有記錄並向下游傳遞，直到再也沒有記錄為止。然後會使用與第一個輸入（即主輸入）相同的資料結構（記錄數、欄位個數等）讀取下一個來源中的記錄。當主來源的欄位比另一輸入來源中的欄位多時，對於任何不完整的值都會使用系統無效字串（\$null\$）。

附加節點對於合併具有類似結構但不同資料的資料集非常有用。例如，您可能將不同時段的交易資料儲存在不同檔案中，如三月份一個銷售資料檔案，而四月份另外一個檔案。假設它們具有相同的結構（欄位相同，順序也相同），那麼 Append 節點會將它們結合為一個較大的檔案，然後您可以對該檔案進行分析。

注意：要附加檔案，欄位測量層次必須相似。例如，名義欄位無法附加測量層次為連續的欄位。

## 設定附加選項

**比對欄位方法。** 選取比對要附加的欄位時要使用的方法。

- **位置。** 選取此選項將根據欄位在主資料來源中的位置附加資料集。使用此方法時，您的資料應該進行排序，以確保正確的附加。
- **名稱。** 選取此選項將根據欄位在輸入資料集中的名稱附加資料集。同樣，選取大小寫相符可在比對欄位名稱時啟用大小寫的區分。

**輸出欄位。** 列出連接至附加節點的來源節點。清單上的第一個節點為主輸入來源。您可以通過按一下欄標題，對顯示中的欄位進行排序。此排序並不真正對資料集中的欄位進行重新排序。

**欄位來源。** 選取僅主要資料集可根據主要資料集中的欄位生成輸出欄位。主要資料集是在「輸入」標籤上指定的第一個輸入。選取全部資料集可為全部資料集中的所有欄位生成輸出欄位，而不管在所有輸入資料集中是否存在相符欄位。

**生成新欄位，顯示記錄的來源資料集。** 選取此選項可向輸出檔案新增一個附加欄位，該欄位的值將指示每個記錄的來源資料集。在文字欄位中指定一個名稱。該預設欄位名為 輸入 。

---

## 差異節點

必須移除資料集中的複製記錄後才能開始資料採礦。例如，在某個行銷資料庫中，個別可能以不同的位址或公司資訊多次出現。您可以使用「相異的」節點來尋找或移除資料中的複製記錄，或者根據一群組複製記錄建立單個複合記錄。

要使用「相異」節點，您必須先定義一組索引鍵欄位，用於確定何時將兩個記錄視為重複項。

如果您僅挑選了部分欄位用作鍵欄位，那麼兩個「複製」記錄可能並非確實完全相同，這是因為它們的其餘欄位的值仍可能有所不同。在這種情況下，您還可以定義在每群組複製記錄中套用的排序。此排序使您能夠進行微調，以確定要將哪個記錄視為群組中的第一個記錄。否則，會將所有複製項都視為可交換，並可能已選取任意記錄。不會對記錄的傳入順序加以考慮，因此，使用上游「排序」節點並無說明（請參閱下方文中的『在「相異的」節點中進行記錄排序』）。

**眾數。** 指定是建立複合記錄，還是併入或排除（捨棄）第一個記錄。

- **為每個群組建立複合記錄。** 提供一種對非數值欄位進行聚合的方式。選中此選項將使「複合」標籤可用，您可以在該標籤上指定如何建立複合記錄。如需相關資訊，請參閱第 82 頁的『相異的複合設定』。
- **僅包含每個群組中的第一個記錄。** 選取每群組複製記錄中的第一個記錄並捨棄餘下的記錄。第一筆記錄由下面定義的排序決定，而不是由送入記錄的順序決定。
- **僅捨棄每個群組中的第一個記錄。** 捨棄每群組複製記錄中的第一個記錄並選中餘下的記錄。第一筆記錄由下面定義的排序決定，而不是由送入記錄的順序決定。此選項對於 找出 資料中的複製非常有幫助，因為您隨後即可在串流中檢查這些內容。

**關鍵分組欄位。** 列出用於確定記錄是否相同的一個或多個欄位。您可以：

- 使用右側的欄位選取器按鈕將欄位新增至此清單。
- 使用紅色 X（移除）按鈕從清單中刪除欄位。

**在群組內記錄的排序依據。** 列出用於確定各個記錄在每群組複製項目中的排序方式以及是按遞增還是遞減排序的欄位。您可以：

- 使用右側的欄位選取器按鈕將欄位新增至此清單。
- 使用紅色 X（移除）按鈕從清單中刪除欄位。
- 如果您按多個欄位排序，則使用上下按鈕移動欄位。

如果您已選擇包含或排除每個群組中的第一個記錄，並且將哪個記錄視為第一個記錄對您而言十分重要，那麼必須指定排序。

對於「複合」標籤上的特定選項，如果您已選擇建立複合記錄，那麼也可能希望指定排序。如需相關資訊，請參閱第 82 頁的『相異的複合設定』。

**預設排序順序。** 指定依預設各個記錄是按排序鍵值的遞增還是遞減排列。

## 在「相異的」節點中進行記錄排序

如果一群組複製項中的記錄順序至關重要，那麼您必須使用「相異的」節點中的**在群組內記錄的排序依據**選項來指定順序。請勿依賴於上游「排序」節點。請注意，不會對記錄的傳入順序加以考慮 - 僅考慮此節點中指定的順序。

如果未指定任何排序欄位（或者指定的排序欄位不夠充分），那麼每群組複製項中的記錄均為無序（或者不完全地進行了排序），而結果可能不可預測。

例如，假定存在大量與眾多機器相關的日誌記錄。此日誌包含資料，例如：

表 16. 機器日誌資料

時間戳記	機器	溫度
17:00:22	機器 A	31
13:11:30	機器 B	26
16:49:59	機器 A	30
18:06:30	機器 X	32
16:17:33	機器 A	29
19:59:04	機器 C	35
19:20:55	機器 Y	34
15:36:14	機器 X	28
12:30:41	機器 Y	25
14:45:49	機器 C	27
19:42:00	機器 B	34
20:51:09	機器 Y	36
19:07:23	機器 X	33

要將記錄數減少為每台機器的最新的記錄，請使用 Machine 作為鍵欄位，並使用 Timestamp 作為排序欄位（遞減）。輸入順序不會影響結果，這是因為，排序選擇指定了對於給定機器應傳回多數列中的哪些列，最終資料輸出如下所示。

表 17. 排序的機器日誌資料

時間戳記	機器	溫度
17:00:22	機器 A	31
19:42:00	機器 B	34
19:59:04	機器 C	35
19:07:23	機器 X	33
20:51:09	機器 Y	36

## 相異的最佳化設定

如果您正在處理的資料只有少量記錄，或已排序，您可以最佳化處理資料的方式，以使 IBM SPSS Modeler 更有效地處理資料。

注意：如果您選取了輸入資料集具有少量相異的鍵值，或使用節點的 SQL 產生，那麼可能傳回相異的鍵值內的任何列；要控制相異的鍵值內傳回的列，需要使用「設定」標籤上在群組內記錄的排序依據欄位指定排序。只要您在「設定」標籤上指定了排序，則最佳化選項不會影響按照「相異的」節點輸出的結果。

輸入資料集包含少量相異的鍵值。如果您具有少量記錄和/或少量鍵欄位的唯一性值，請選取此選項。這樣做可以改善效能。

已經按「設定」標籤上的分組欄位和排序欄位對輸入資料集進行排序。只有當您的資料已按「設定」標籤上的在群組內記錄的排序方式下面列出的所有欄位進行排序，且資料的遞增或遞減排序依據相同，才能選取此選項。這樣做可以改善效能。

停用 SQL 產生。選取此選項以取消節點的 SQL 產生。

## 相異的複合設定

如果要處理的資料包含多條記錄（例如針對同一人員的多條記錄），那麼可以通過建立要處理的單個複合（或聚合）記錄來最佳化資料處理方式。

註：只有在您從「設定」標籤中選取了為每個群組建立複合記錄的情況下，此標籤才可用。

### 為「複合」標籤設定選項

**欄位。** 此欄將以自然排序順序顯示資料模型中除索引鍵欄位以外的所有欄位。如果節點未連接，則不顯示任何欄位。要按欄位名稱的按字母順序對列進行排序，請按一下欄標題。可以按住 Shift 進行按一下或者按住 Ctrl 進行按一下以選取多個列。此外，用滑鼠右鍵按一下某個欄位時，將顯示一個功能表，您可以從中選擇所有列，按欄位名稱或值的遞增或遞減對這些列進行排序，按測量或儲存類型選取欄位，或者選擇某個值以自動將同一值填充依據項目新增到每個所選列。

**值填充依據。** 選取要用於欄位的複合記錄的值類型。可用選項取決於欄位類型。

- 對於數值範圍欄位，您可以從下列選項中進行選擇：
  - 群組中的第一筆記錄
  - 群組中的最後一筆記錄
  - 總計
  - 平均數
  - 下限
  - 上限
  - 自訂
- 對於時間或日期欄位，您可以從下列選項中進行選擇：
  - 群組中的第一筆記錄
  - 群組中的最後一筆記錄
  - 最早
  - 最近
  - 自訂
- 對於字串或無類型欄位，您可以從下列選項中進行選擇：
  - 群組中的第一筆記錄
  - 群組中的最後一筆記錄
  - 第一個英數
  - 最後一個英數
  - 自訂

在每種情況下，您都可以使用自訂選項來更好地控制用於填充複合記錄的值。如需相關資訊，請參閱『相異的複合 - 自訂標籤』。

**包含欄位中的記錄計數。** 選取此選項可在每條輸出記錄中包含一個額外的欄位，依預設該欄位名為 Record\_Count。此欄位指出聚集多少個輸入記錄來形成每一個聚集記錄。要為此欄位建立自訂名稱，請在編輯欄位中輸入內容。

### 相異的複合 - 自訂標籤

您可以通過「自訂填入」對話框更好地控制使用哪個值完成新的複合記錄。請注意，如果僅在「複合」標籤上自訂了單個欄位列，那麼必須在對資料進行實例化後使用此選項。

註：只有您在「複合」標籤上的**值填充依據欄**中選取了自訂值的情況下，此對話框才可用。

根據欄位類型，您可以選擇下列其中一個選項。

- **依頻率選取**。根據在資料記錄中的出現頻率選擇值。

註：對於類型為「連續」、「無類型」或「日期/時間」的欄位，此選項無法使用。

– **使用**。選取「頻率最高」或「頻率最低」。

– **鏈結**。如果有兩個或兩個以上的記錄出現頻率相同，請指定如何選取所需記錄。您可以從下列四個選項中進行選擇：使用第一條記錄、使用最後一欄一條記錄、使用頻率最低的記錄或使用頻率最高的記錄。

- **併入值 (T/F)**。選取此選項可將欄位轉換為旗標，該旗標用於確定群組中是否包含具有指定值的記錄。然後，可以從選取的欄位的值清單中選取值。

註：如果您在「合成」標籤上選取多個「欄位」列，則無法使用

- **清單中的第一個符合項**。選取此選項可對要指定給複合記錄的值劃分設定優先順序級。然後，可以從選取的欄位的項目清單中選取某個項目。

註：如果您在「合成」標籤上選取多個「欄位」列，則無法使用

- **連接值**。選取此選項可將群組中的所有值連接成一個字串，從而保留這些值。必須指定要在每個值之間使用的分隔符號。

註：在您選取了一個或多個類型為「連續」、「無類型」或「日期/時間」的欄位列的情況下，這是唯一可用的選項。

- **使用分隔符號**。可以選擇使用空格或逗點作為合併字串中的分隔符號值。或者，您可以在**其他欄位**中輸入自己的分隔符號值字元。

註：僅在選取了**連接值**選項的情況下可用。

---

## 多媒體串流時間序列節點

「串流式時間序列」節點用於在一個步驟中對「時間序列」模型同時進行建立和分數。針對每一個目標欄位會建置個別的時間序列模型，但模型塊不會新增至產生的模型選用區，且無法瀏覽模型資訊。

用於對時間序列資料進行建模的方法要求每個測量之間的時間一致，並且任何遺漏值都由空列指出。如果資料尚未符合此需求，那麼將必須根據需要對值進行轉換。

與「時間序列」節點連線時要注意的其他點：

- 欄位必須為數值。
- 日期欄位不能用作輸入。
- 會忽略分割區。

「串流式時間序列」節點對時間序列的指數平滑化模型、單變異數自身迴歸整合移動均數 (ARIMA) 模型和多變異數 ARIMA (或轉換函數) 模型進行估計，並根據時間序列資料產生預測。另外提供的 Expert Modeler 會嘗試自動為一或多個目標欄位識別與估計最適合的 ARIMA 或指數平滑化模型。

有關時間序列建模的進一步資訊，請參閱《SPSS Modeler 建模節點手冊》的「「時間序列」模型」部分。

通過 IBM SPSS Modeler Solution Publisher，使用 IBM SPSS Collaboration and Deployment Services Scoring Service 受支援在串流式部署環境中使用「串流式時間序列」節點。

## 「串流式時間序列」節點 - 欄位選項

使用預先定義的角色：此選項使用上游類型節點（或上游來源節點的「類型」標籤）的角色設定（目標、預測值等等）。

使用自訂欄位指派。要手動分配目標、預測值和其他角色，請選中此選項。

註：如果已對資料分割區，那麼選取使用預先定義的角色時將考慮分割區，但選取使用自訂欄位分配時則不考慮。

欄位。使用箭頭按鈕，將此清單中的項目手動指派給畫面右側上的各個角色欄位。這些圖示指出每一個角色欄位的有效測量層次。

若要選取清單中的全部欄位，請按一下全部按鈕，或按一下個別測量層次按鈕來選取具有該測量層次的所有欄位。

目標 選取一個或多個欄位作為預測目標。

候選輸入 選取一個或多個欄位作為預測輸入。

事件和人為介入 使用此區域將特定輸入欄位指定為事件欄位或人為介入欄位。此指定可將欄位視為包含可受事件（可預期的重複出現的狀況，例如，促銷）或人為介入（一次性事件，例如，電源中斷或員工罷工）影響的時間序列資料。

## 「串流式時間序列」節點 - 資料規格選項

通過「資料規格」標籤，您可以設定用於將資料包含在模型中的所有選項。只要同時指定日期/時間欄位和時間間隔，便可以按一下執行按鈕來建立包含所有預設選項的模型，但通常您會想要根據自己的用途自訂建立。

該標籤包含數個不同窗格，您可以在其中設定特定於您模型的自訂作業。

## 「串流式時間序列」節點 - 觀察值

使用此窗格中的設定可以指定用於定義觀察的欄位。

### 由日期/時間欄位指定的觀察

您可以指定觀察由日期、時間或時間戳記欄位定義。除了定義觀察的欄位以外，還可以選取說明觀察的適當時間間隔。根據指定的時間間隔，您還可以指定其他設定，例如觀察間隔（增量）或每週的天數。下列考量僅適用於時間間隔：

- 如果觀察的時間（例如處理銷售單的時間）間隔無規律，請使用值無規律。如果選取無規律，您必須從「資料規格」標籤上的時間間隔設定中，指定用於分析的時間間隔。
- 如果觀察代表日期和時間，而時間間隔是小時數、分鐘數或秒數，則使用每天小時數、每天分鐘數或每天秒數。如果觀察代表無日期參照的時間（期間），且時間間隔為小時數、分鐘數或秒數，則使用小時數（非週期）、分鐘數（非週期）或秒數（非週期）。
- 根據選取的時間間隔，程序可以偵測遺漏的觀察。由於此程序假定所有觀察之間的時間間距相等，並假定未遺漏觀察，因此有必要偵測遺漏的觀察。例如，如果時間間隔為「天」，並且日期 2015-10-27 後跟 2015-10-29，那麼表示遺漏 2015-10-28 的觀察。對於任何遺漏觀察，將插補值；使用「資料規格」標籤的遺漏值處理區域可以指定用於處理遺漏值的設定。
- 指定的時間間隔可讓程序偵測相同時間間隔中需要總計在一起的多個觀察，並依據間隔界限（例如月份的第一天）排列觀察，以確保觀察間隔相等。例如，如果時間間隔為「月」，則會將相同月份內的多個日期總計在一起。這種類型的總計稱為分組。依預設，在分組觀察時會加總觀察。您可以從「資料規格」標籤上的總計及分佈設定中指定不同的分組方法，如觀察平均數。



- 對於部分時間間隔，其他設定可定義一般相等時間間隔中的岔斷。例如，如果時間間隔為「日」，但只有工作日有效，則您可以指定一週內有五天，一週從星期一啟動。

### 定義為期間或週期性時期的觀察

觀察可由一或多個代表期間或重複循環期間（多達任意數目的循環水準）的整數欄位定義。借助此結構，您可以說明任何標準時間間隔都無法支持的觀察數列。例如，可使用代表年份的循環欄位以及代表月份的期間欄位來說明只有 10 個月的財政年度，其中一個循環的長度為 10。

指定循環期間的欄位可定義週期性水準的階層，其中最低水準由期間欄位定義。下一個最高水準由水準為 1 的循環欄位指定，接著由水準為 2 的循環欄位指定，依此類推。除最高層次以外，每個層次的欄位值對於次高層次都必須具有週期性。最高水準的值不能為週期性值。例如，對於由 10 個月組成的財年，月在年中具有週期性，而年不具有週期性。

- 特定水準上的循環長度是下一個最低水準的週期。以財政年度為例，只有一個循環水準且循環長度為 10，因為下一個最低水準代表月份，而指定的財政年度有 10 個月。
- 請指定任何並非起始於 1 的週期性欄位的起始值。要偵測遺漏值，有必要進行此設定。例如，如果週期性欄位從 2 啟動，但起始值指定為 1，則程序會假設該欄位每個循環中的第一個期間都存在遺漏值。

### 「串流式時間序列」節點 - 分析時間間隔

用於分析的時間間隔可以與觀察的時間間隔不同。例如，觀察的時間間隔為「天」時，您可以選擇「月」用作進行分析的時間間隔。系統先將每日資料聚合為每月資料，然後再建立模型。您也可以選擇將資料分佈從較長時間間隔改成較短時間間隔。例如，如果觀察是按季度發生的，則您可以將資料從按季度分佈改成按月分佈。

使用此窗格中的設定指定用於分析的時間間隔。聚集或分佈資料的方法是在「資料規格」標籤上的聚集和分佈設定中指定的。

執行分析的時間間隔的可用選項，視觀察的定義方式以及這些觀察的時間間隔而定。特別是，如果觀察由週期性時期定義，那麼僅受支援聚集。在此情況下，分析的時間間隔必須大於或等於觀察的時間間隔。

### 「串流式時間序列」節點 - 聚集和分佈選項

使用此窗格中的設定可以指定用於對觀察的時間間隔的相關輸入資料進行聚合或分佈的設定。

#### 聚集函數

當用於分析的時間間隔長於觀察的時間間隔時，會總計輸入資料。例如，如果觀察時間間隔是「日」，而分析的時間間隔是「月」，則會執行總計。提供了下列聚集函數：平均數、加總、眾數、最小值或最大值。

#### 分佈函數

當用於分析的時間間隔短於觀察的時間間隔時，會分佈輸入資料。例如，如果觀察時間間隔是「季度」，而分析的時間間隔是「月」，則會執行分佈。提供了下列分佈函數：平均數或加總。

#### 分組函數

如果觀察由日期/時間定義且在相同的時間間隔發生多個觀察，則可套用分組。例如，如果觀察的時間間隔是「月」，則相同月份中的多個日期會分組在一起，並與其發生所在的月份相關聯。您可以使用的分組函數如下：mean、sum、mode、min 或 max。如果觀察由日期/時間定義，且觀察的時間間隔指定為「無規律」，則一律會執行分組。

註：雖然分組是一種總計形式，但它是在處理所有遺漏值之前執行，而正式的總計是在處理完所有遺漏值之後執行。當觀察的時間間隔指定為「無規律」時，只能使用分組函數來執行總計。

#### 總計跨日期觀察與前一天的值

指定是否將時間跨天界限的觀察值聚合到前一天的值。例如，對於在 20:00 開始的 8 小時一天的每小

時觀察值，此設定指定是否將介於 00:00 與 04:00 之間的觀察值包含在前一天的聚合結果中。只有在觀察時間間隔是「每天小時數」、「每天分鐘數」或「每天秒數」，且分析的時間間隔為「日」時，此設定才適用。

#### 所指定欄位的自訂設定

您可以逐個欄位指定聚集函數、分佈函數及分組函數。這些設定會置換聚集函數、分佈函數及分組函數的預設值。

### 「串流式時間序列」節點 - 遺漏值選項

使用此窗格中的設定可以指定輸入資料中要取代為插補值的遺漏值數。提供了下列取代方法：

#### 線性插補

使用線性插補取代遺漏值。該選項會用遺漏值出現之前的最後一個有效值，還有遺漏值出現之後的第一個有效值當作內插。如果數列中的第一個或最後一個觀察具有遺漏值，則會使用數列開頭或結尾的兩個最近非遺漏值。

#### 數列平均數

以整個數列的平均數置換遺漏值。

#### 鄰近點平均數

以有效周圍值的平均數置換遺漏值。附近點的指距是指用來計算平均數的遺漏值，其之前及之後的有效值個數。

#### 鄰近點中位數

以有效周圍值的中位數置換遺漏值。附近點的指距是指用來計算中位數的遺漏值，其之前及之後的有效值個數。

#### 線性趨勢

此選項使用數列中的所有非遺漏觀察來擬合簡單線性迴歸模型，該模型隨後用於插補遺漏值。

其他設定：

#### 最低資料品質分數 (%)

針對時間變數以及對應於每個時間序列的輸入資料計算資料品質測量。如果資料品質分數低於此臨界值，那麼將捨棄對應的時間序列。

### 「串流式時間序列」節點 - 估計期間

在「估計期間」窗格中，您可以指定要在模型估計中使用的記錄的範圍。依預設，估計期間啟動於所有數列中最早觀察的時間，在所有數列中最晚觀察的時間結束。

#### 依啟動與結束時間

您可以同時指定估計期間的啟動與結束時間，也可以只指定啟動時間或結束時間。如果您省略估計期間的啟動或結束時間，則會使用預設值。

- 如果觀察值由日期/時間欄位定義，請以用於該日期/時間欄位的格式輸入開始時間值和結束時間值。
- 如果觀察是由循環期間定義，請為每個循環期間欄位指定一個值。每個欄位都會顯示在個別的直欄中。

#### 依最晚或最早時間間隔

將估計期間定義為指定號碼的時間間隔，這些時間間隔從資料中的最早時間間隔開始或以最晚時間間隔結束，並具有選用偏移量。在此環境定義中，時間間隔是指分析時間間隔。例如，假定觀察是按月份，而分析的時間間隔是按季度。指定最晚並為時間間隔數指定值 24 表示最晚 24 個季度。

您可以選擇性地排除指定的時間間隔數。例如，指定最晚 24 個時間間隔並指定 1 代表要排除的數目，表示估計期間由最後一個間隔之前的 24 個間隔組成。

## 「串流式時間序列」節點 - 建置選項

通過「數據規範」標籤，您可以設定用於建立模型的所有選項。您只需按一下執行按鈕即可使用所有預設選項來建置模型，但通常您想要自訂建置以用於您的專屬用途。

此標籤包含兩種不同的窗格，您可以在這些窗格中設定特定於模型的自訂作業內容。

## 「串流式時間序列」節點 - 一般建置選項

此窗格中的可用選項取決於您從方法清單中選擇下列三項設定中的哪一項：

- **Expert Modeler**。選擇此選項以使用 Expert Modeler，此組件將自動為每個相依數列尋找配適度最高的模型。
- **指數平滑化**。使用此選項，可指定自訂指數平滑化模式。
- **ARIMA 程序** 使用此選項，可指定自訂 ARIMA 模式。

## Expert Modeler

在模型類型下，選取您要建立的模型的類型：

- **所有模式**。Expert Modeler 會同時考量 ARIMA 和指數平滑化模式。
- **僅指數平滑化模式**。Expert Modeler 僅考慮指數平滑化模型。
- **僅 ARIMA 模式**。Expert Modeler 僅考慮 ARIMA 模型。

**Expert Modeler 會考量週期性模型**。只有在為作用中資料集定義了週期性時才啟用此選項。選取此選項時，Expert Modeler 會同時考量週期性和非週期性模型。如果未選取此選項，那麼 Expert Modeler 將僅考慮非週期性模型。

**Expert Modeler 考慮複雜指數平滑化模型**。選取了此選項時，Expert Modeler 將搜尋所有 13 個指數平滑化模型（其中 7 個存在於原始時間序列節點中，而剩下 6 個是 18.1 版中新增的節點）。如果未選取此選項，那麼 Expert Modeler 將搜尋原始的 7 個指數平滑化模型。

在離群值下，從下列選項中進行選取

**自動偵測偏離值**。依預設，不會執行自動偵測偏離值。選取此選項以執行自動偵測離群值，然後選取所期望的離群值類型：

輸入欄位必須具有旗標、列名或序數測量層次，並且必須是數值（例如，對於旗標欄位，必須為 1/0，而非 True/False），才能包含在此清單中。

對於在欄位標籤上 ID 為事件欄位或人為介入欄位的輸入，Expert Modeler 僅考慮簡單迴歸方法而不是任意變換函數。

## 指數平滑化

**模式類型**。指數平滑化模型分類為週期性模型或非週期性模型。<sup>1</sup>只有使用「資料規格」標籤上的「時間間隔」窗格定義的週期性為週期性時，才可以使用週期性模型。週期為：循環期間、年、季度、月、每週天數、每日小時數、每日分鐘數及每日秒數。提供了下列模型類型：

---

1. Gardner, E. S. 1985. Exponential smoothing: The state of the art. *Journal of Forecasting*, 4, 1-28.

- **簡單。** 此模型適用於不具有趨勢或週期性的數列。層級是它唯一的相關平滑化參數。簡式指數平滑化與 ARIMA 最為相似，都具有零個自身迴歸階數、一個差分階數、一個移動平均數階數，且沒有常數。
- **Holt 線性趨勢。** 此模型適用於具有線性趨勢且沒有週期性的序列。其相關平滑化參數為層級與趨勢，且在此模型中不受彼此的數值所限制。Holt's 模式較 Brown's 模式更為普遍，但對於大型數列的預估值計算時間會較長。Holt 指數平滑化與 ARIMA 最為相似，都具有零個自身迴歸階數、兩個差分階數，以及兩個移動平均數階數。
- **減幅趨勢。** 此模型適用於具有線性趨勢且具有漸失線性趨勢但沒有週期性的數列。其相關平滑化參數為層級、趨勢和減幅趨勢。減幅指數平滑化與 ARIMA 最為相似，都具有一個自身迴歸階數、一個差分階數，以及兩個移動平均數階數。
- **相乘性趨勢。** 該模型適合於具有一種隨數列長度而變的趨勢且沒有週期性的數列。其相關的平滑參數是層次和趨勢。相乘性趨勢指數平滑化與任何 ARIMA 模型都不相似。
- **Brown 線性趨勢。** 此模型適用於具有線性趨勢且沒有週期性的序列。其相關平滑化參數為層級與趨勢，且在此模型中皆已假定為相等。因此，Brown's 模式為 Holt's 模式的特殊觀察值。Brown 指數平滑化與 ARIMA 最為相似，都具有零個自身迴歸階數、兩個差分階數以及兩個移動平均數階數，且移動平均數的第二階係數與第一階的二分之一係數平方相等。
- **簡單週期性。** 此模型適用於沒有趨勢的數列，以及在一段時間保持不變的週期效果。其相關平滑化參數為層級和週期。週期可加性指數平滑化與 ARIMA 最為相似，都具有零個自身迴歸階數、一個差分階數、一個週期差分階數及一個、 $p$  個及  $p + 1$  個移動平均數階數，其中  $p$  為週期區間內的期間個數。對於每月資料， $p=12$ 。
- **Winters 可加性。** 此模型適用於具有線性趨勢的數列，以及在一段時間保持不變的週期效果。其相關平滑化參數可為層級、趨勢和週期。Winters 可加性指數平滑化與 ARIMA 最為相似，都具有零個自身迴歸階數、一個差分階數、一個週期差分階數及  $p + 1$  個移動平均數階數，其中  $p$  為週期區間內的期間個數。對於每月資料， $p=12$ 。
- **具有可加性週期性的減幅趨勢。** 此模型適合於具有逐漸消退的線性趨勢且季節作用不隨時間變化的數列。它的相對平滑化參數為層級、趨勢、減幅趨勢及週期。阻尼趨勢和可加性週期性指數平滑化與任何 ARIMA 模型都不相似。
- **具有可加性週期性的相乘性趨勢。** 該模型適合於具有隨數列長度而變的趨勢且季節作用不隨時間變化的數列。其相關平滑化參數可為層級、趨勢和週期。相乘性趨勢和可加性週期性指數平滑化與任何 ARIMA 模型都不相似。
- **相乘性週期性。** 此模型適合於不具有趨勢且季節作用隨數列長度而變的數列。其相關平滑化參數為層級和週期。相乘性週期性指數平滑化與任何 ARIMA 模型都不相似。
- **Winters 相乘性。** 此模型適用於具有線性趨勢的數列，以及隨著數列長度而變更的週期效果。其相關平滑化參數可為層級、趨勢和週期。Winters 的可乘指數平滑法與任何 ARIMA 模型都不相似。
- **具有相乘性週期性的減幅趨勢。** 此模型適合於具有逐漸消退的線性趨勢且季節作用隨數列的大小而變化的數列。它的相對平滑化參數為層級、趨勢、減幅趨勢及週期。減幅趨勢和相乘性週期性指數平滑化與任何 ARIMA 模型都不相似。
- **具有相乘性週期性的相乘性趨勢。** 此模型適合於具有隨數列的長度發生變化的趨勢和季節作用的數列。其相關平滑化參數可為層級、趨勢和週期。相乘性趨勢和相乘性週期性指數平滑化與任何 ARIMA 模型都不相似。

**目標轉換。** 您可以指定先在各個因變數上執行轉換，然後再進行模式化作業。

- 無。 沒有執行任何轉換。
- 平方根。 執行平方根轉換。
- 自然對數。 執行自然對數轉換。

## ARIMA

指定自訂 ARIMA 模型的結構。

**ARIMA 階數。** 在網格的相應單元中，輸入模型的各個 ARIMA 成分的值。所有數值都必須為非負的整數。對自我迴歸和移動平均成份來說，該值表示最大階數。所有正較低階數都包括在模型中。例如，如果指定 2，那麼模型將包含階數 2 和 1。只有在為作用中資料集定義了週期性的情況下，才會啟用「週期性」欄中的單元。

- **自我迴歸。** 模式中自我迴歸階數的個數。自我迴歸階數指定要從數列中取用哪個先前值來預測目前值。例如，自身迴歸階數 2 指定數列中過去兩個時間段的值用於預測目前的值。
- **差分。** 指定在估計模式之前套用至數列的差分階數。當趨勢存在時（包含趨勢的數列一般都是非平穩性數列，但 ARIMA 模式中假定數列為穩定性），就必須對數列進行差分，以移除趨勢的影響。差異分析階數對應於數列趨勢的程度；第一階數差異分析表示線性趨勢，第二階數差異分析表示二次趨勢，依此類推。
- **移動平均數 (q)。** 模式中移動平均階數的個數。移動平均階數指定如何使用先前數值的數列平均數離差來預測目前值。例如，移動平均階數 1 和 2 指定在預測數列的目前值時，要考慮最後兩段時間中各個數列平均值的離差。

**週期性。** 週期性自我迴歸、移動平均數，和差分成份所扮演角色與其非週期性對等項目相同。但對週期性階數來說，目前序列值是受由一或多個週期性期間分隔的先前數列值影響。例如，對於月度資料（週期性期間為 12），週期性階數 1 表示目前數列值將受到目前期間之前 12 個期間的數列值的影響。如此對每月資料來說，週期性階數 1 就與指定非週期性階數 12 相同。

**自動偵測偏離值。** 選中此選項可以對離群值執行自動偵測，並選取可用的一個或多個離群值類型。

**要偵測的偏離值的類型。** 選取要偵測的離群值類型。支援的類型為：

- 可加性（預設值）
- 層級變遷（預設值）
- 創新
- 短暫
- 週期性相加(S)
- 局部趨勢
- 可加性補充(V)

**轉換函數階數數和轉換。** 要指定轉換並為 ARIMA 模型中的任何或所有輸入欄位定義轉換函數，請按一下設定；這將顯示另一個對話框，您可以在其中輸入變換和轉換詳細資料。

**常數項納入模式中。** 納入常數是標準作業，除非您確定總平均序列值是 0；套用差分時，則建議排除常數。

**轉換函數：** 使用「轉換函數順序和轉換」對話框可以指定轉換以及為 ARIMA 模型中的任何或所有輸入欄位定義轉換函數。

**目標轉換。** 在此窗格中，您可以指定對每個目標變數進行建模之前要對其執行的轉換。

- 無。沒有執行任何轉換。
- 平方根。執行平方根轉換。
- 自然對數。執行自然對數轉換。

**候選輸入轉換函數及轉換。** 通過使用轉換函數，您可指定以何種方式使用輸入欄位的過去值來預測目標序列的未來值。左端窗格的清單中顯示了所有的輸入欄位。此窗格中的其餘資訊特定於您選取的輸入欄位。

**轉換函數階數。** 在結構網格的相應單元中，輸入轉換函數的各個成分的值。所有數值都必須為非負的整數。對分子和分母成份來說，該值表示最大階數。所有正較低階數都包括在模型中。此外，一定會包含階數 0 供分子成份使用。例如，如果將分子指定為 2，那麼模型將包含階數數 2、1 和 0。如果將分子指定為 3，那麼模型將包含階數數 3、2 和 1。只有在為作用中資料集定義了週期性的情況下，才會啟用季節欄中的單元。

**分子。** 轉換函數的分子階數指定所選取不相依系列（預測值）中，哪些先前值用來預測不相依系列的現行值。例如，分子階數數 1 指定使用過去某個時段的相依系列的值（以及相依系列的目前的值）來預測各不相依系列的目前的值。

**分母。** 轉換函數的分母階數指定要如何使用所選取不相依系列（預測值）先前值的數列平均數離差來預測相依系列的現行值。例如，分母階數數 1 指定在預測每個相依系列的目前的值時，需要考慮與過去一個時間期間的不相依系列的平均數離差。

**差異。** 指定在估計模型之前套用至所選擇不相依系列（預測值）的差分階數。當趨勢存在時，就必須對數列進行差分，以移除趨勢的影響。

**週期性。** 週期性分子、分母，和差分成份所扮演角色與其非週期性對等項目相同。但對週期性階數來說，目前序列值是受由一或多個週期性期間分隔的先前數列值影響。例如，以每月資料（週期性期間為 12）來說，週期性階數 1 代表目前序列值受 12 個週期之前的序列值影響。如此對每月資料來說，週期性階數 1 就與指定非週期性階數 12 相同。

**延遲。** 設定延遲會造成輸入欄位的影響依指定區間數延遲。例如，如果延遲設定為 5，在時間  $t$  的輸入欄位值不會影響預測，要一直到經過五個週期 ( $t + 5$ ) 之後才会有影響。

**轉換。** 一組自變數的轉換函數規格，同時也包括要在這些變數上執行的選擇性轉換。

- 無。沒有執行任何轉換。
- 平方根。執行平方根轉換。
- 自然對數。執行自然對數轉換。

## 「串流式時間序列」節點 - 模型選項

**信賴限制寬度 (%)。** 信賴區間是供模型預測和殘差自我相關計算使用。您可以指定小於 100 的任何正數值。依預設，將使用 95% 信賴區間。

用於將記錄延伸至未來的選項可設定要在估計期間結束以外預測的時間間隔數目。在這種情況下，時間間隔是您在「資料規格」標籤上指定的分析的時間間隔。要求預測時，會針對任何不是目標的輸入序列自動建置自動迴歸模型。然後，在預測期間使用這些模型產生那些輸入數列的值。此設定沒有上限制。

## 要在用於預測的未來值

- **計算輸入的未來值** 如果選取此選項，那麼會自動計算預測值、雜訊預測、差異估計和未來時間值的預測值。要求預測時，會針對任何不是目標的輸入序列自動建置自動迴歸模型。然後，在預測期間使用這些模型產生那些輸入數列的值。
- **選取要將其值新增到資料中的欄位。** 對於要預測的每條記錄（holdout 除外），如果您使用預測值欄位（角色設定為輸入），那麼可以為每個預測值的預測期間指定估計值。您可以手動指定值，也可以從清單中選擇。
  - **欄位。** 按一下欄位選取器按鈕，並選擇可以用作預測值的任何欄位。請注意，在這裡選取的欄位在建模時可以使用，也可能無法使用；若要實際將欄位用作預測值，則必須在下游建模節點中選取它。此對話框會為您直接提供方便的地方來指定未來值，因此可由多個下游建模節點共用而無需每一個節點中單獨指定它們。另請注意，可用欄位；的清單可能會受到「建置選項」標籤中選項的約束。

請注意，如果為串流中不再可用（因為已將其刪除或由於「建置選項」標籤中的選項更新）的欄位指定了未來值，那麼此欄位將顯示為紅色。

- **值。** 針對每一個欄位，您可以從函數清單中選擇，或按一下**指定**以手動輸入值或從預定值清單中選擇。如果預測值欄位與您可以控制或可以提前知道的項目相關，則應手動輸入值。例如，如果您要根據房間預約數預測旅館下個月的收入，則可以指定該期間實際擁有的預約數。相反地，如果預測值欄位與您無法控制的某個項目相關，例如股價，則可以使用最近的值或最近點的平均數等函數。

可用的函數視欄位的測量層級而定。

表 18. 可用於測量層次的函數

測量層級	函數
連續或名義欄位	空白 最近點的平均數 最新值 指定
旗標欄位	空白 最新值 TrueFalse指定

**最近點的均數**根據最後一欄三個資料點的均數計算未來值。

**最近值**將未來值設定為最近資料點的值。

**true/假**將旗標欄位的未來值設定為指定的 true 值或 false。

**指定**開啟一個對話框，用於手動指定未來值或從預先定義清單中選擇未來值。

## 可用於評分

您可以在此設定模型塊的對話框中顯示的評分選項的預設值。

- **計算置信限制的上限和下限。** 如果選取了此選項，那麼對於每個目標欄位，將為信賴區間上限和下限建立新欄位（帶有預設字首 \$TSLCI- 和 \$TSUCI-）。
- **計算雜訊殘差。** 如果已選取了此選項，那麼對於每個目標欄位，此選項將為模型殘差建立新欄位（帶有預設字首 \$TSResidual-），並同時建立這些值的總計。

## 模型設定

**要在輸出中顯示的上限模式個數。** 指定您要包含在輸出中的上限模式個數。請注意，如果建立的模型個數超過了此臨界值，那麼模型不會顯示在輸出中，但它們仍可用於評分。預設值是 10。顯示大量模型可能會導致效能不佳或不穩定。

## SMOTE 節點

Synthetic Minority Over-sampling Technique (SMOTE) 節點提供一個超額取樣演算法來處理不平衡的資料集。它提供一個進階方法來平衡資料。SMOTE 程序節點在 Python 中實作並需要 imbalanced-learn© Python 程式庫。如需 imbalanced-learn 程式庫的詳細資料，請參閱 <http://contrib.scikit-learn.org/imbalanced-learn/about.html><sup>1</sup>。

節點選用區上的 Python 標籤包含 SMOTE 節點和其他 Python 節點。

<sup>1</sup>Lemaître, Nogueira, Aridas. "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning." *Journal of Machine Learning Research*, vol. 18, no. 17, 2017, pp. 1-5. (<http://jmlr.org/papers/v18/16-365.html>)

## SMOTE 節點設定

在 SMOTE 節點的設定標籤上定義下列設定。

### 目標設定

**目標欄位。** 選取目標欄位。支援所有旗標、名義、序數及離散測量類型。如果在「分割區」區段中選取了使用分割的資料選項，則將僅對訓練資料進行超額取樣。

### 超額取樣比例

選取自動以自動選取超額取樣比例，或選取設定比例（少數相較多數）以設定自訂比例值。比例是少數類別中的樣本數除以多數類別中的樣本數。值必須大於 0 且小於或等於 1。

### 隨機種子

設定隨機種子。選取此選項並按一下產生可以產生由亂數字產生器使用的種子。

### 方法

**演算法類型。** 選取您要使用的 SMOTE 演算法類型。

### 樣本規則

**K 鄰接項。** 指定要用於建構合成樣本的最近鄰接項的數目

**M 鄰接項。** 指定要用於判定少數樣本是否處於危險狀態的最近鄰接項的數目。僅當選取了 **Borderline1** 或 **Borderline1** SMOTE 演算法類型時，才會使用此選項。

### 分割區

使用分割的資料。如果您只想對訓練資料進行超額取樣，請選取此選項。

SMOTE 節點需要 imbalanced-learn© Python 程式庫。下表顯示 SPSS Modeler SMOTE 節點對話框中的設定與 Python 演算法之間的關係。

表 19. 對映至 Python 程式庫參數的節點內容

SPSS Modeler 設定	Script 名稱 (內容名稱)	Python API 參數名稱
超額取樣比例 (數字輸入控制項)	sample_ratio_value	ratio
隨機種子	random_seed	random_state
K 鄰接項	k_neighbours	k
M 鄰接項	m_neighbours	m
演算法類型	algorithm_kind	kind



---

## 延伸轉換節點

使用「延伸轉換」節點，您可以從 IBM SPSS Modeler 串流中取得資料，並使用 R Scripting 或 Python for Spark Scripting 將轉換套用至資料。資料修改後即會傳回至串流以便進一步處理、進行模型建置及模型評分。

「延伸轉換」節點讓使用者能夠使用以 R 或 Python for Spark 撰寫的演算法來轉換資料，並可讓使用者開發適合特定問題的資料轉換方法。

若要搭配使用此節點與 R，您必須安裝 IBM SPSS Modeler - Essentials for R。請參閱 *IBM SPSS Modeler - Essentials for R: Installation Instructions*，以取得安裝指示與相容性資訊。您還必須已在電腦上安裝相容版本的 R。

### 延伸轉換節點 - 語法標籤

選取語法的類型 - **R** 或 **Python for Spark**。如需相關資訊，請參閱下列章節。當您的語法備妥時，可以按一下執行，以執行「延伸轉換」節點。

#### R 語法

**R 語法**。您可以在此欄位中輸入或貼上用於分析資料的自訂 R Scripting 語法。

**轉換旗標欄位**。指定如何處理旗標欄位。有兩個選項：**字串至因數**，**整數與實數至倍精準數**，以及**邏輯值 (True、False)**。如果您選取**邏輯值 (True、False)**，則會遺失旗標欄位的原始值。例如，如果欄位具有值「男性」與「女性」，則這些值會變更為 True 與 False。

**將遺漏值轉換為 R「無法使用」值 (NA)**。當選取時，會將所有遺漏值轉換為 R NA 值。R 使用值 NA 來識別遺漏值。您使用的部分 R 函數可能具有一個引數，可用來控制資料包含 NA 時的函數行為。例如，函數可讓您選擇自動排除包含 NA 的記錄。如果未選取此選項，則任何遺漏值都會依現狀傳遞至 R，並可能在執行 R Script 時造成錯誤。

**將日期/時間欄位轉換為具有時區特殊控制項的 R 類別**。選取此選項時，具有日期或日期時間格式的變數會轉換為 R 日期/時間物件。您必須選取下列其中一個選項：

- **R POSIXct**。具有日期或日期時間格式的變數轉換為 R POSIXct 物件。
- **R POSIXlt (清單)**。具有日期或日期時間格式的變數轉換為 R POSIXlt 物件。

註：POSIX 格式是進階選項。僅在您的 R Script 指定以需要這些格式的方式處理日期時間欄位時，才使用這些選項。POSIX 格式不適用於具有時間格式的變數。

#### Python 語法

**Python 語法**。您可以在此欄位中輸入或貼上供資料分析使用的自訂 Python Scripting 語法。如需 Python for Spark 的相關資訊，請參閱 Python for Spark 及 Python for Spark Scripting。

### 延伸轉換節點 - 主控台輸出標籤

**主控台輸出標籤**包含「語法」標籤上的 R Script 或 Python for Spark Script 執行時接收的所有輸出（例如，如果使用 R Script，它顯示執行語法標籤上 **R** 語法欄位中的 R Script 時，從 R 主控台接收的輸出）。此輸出可能包括執行 R 或 Python Script 時產生的 R 或 Python 錯誤訊息或警告。輸出可以主要用於 Script 除錯。**主控台輸出標籤**也包含 **R 語法**或 **Python 語法**欄位中的 Script。

每次執行延伸轉換 Script 時，都會使用從 R 主控台或 Python for Spark 接收的輸出來改寫**主控台輸出標籤**的內容。無法編輯輸出。

---

## 「空間時間限制」節點

Space-Time-Boxes (STB) 是一個 Geohash 形式的空間位置延伸。具體而言，STB 是一個英數字串，其代表分佈均勻的空間與時間區域。

例如，STB **dr5ru7l2013-01-01 00:00:00l2013-01-01 00:15:00** 由下列三個部分組成：

- geohash **dr5ru7**
- 啟動時間戳記 **2013-01-01 00:00:00**
- 結束時間戳記 **2013-01-01 00:15:00**

例如，您可以使用空間和時間資訊來提高兩個實體是同一實體的信賴度，因為這兩個實體幾乎在同一時間位於同一位置。另外，通過顯示兩個實體由於其空間和時間接近而相關，您可以提高關係確定的精確度。

您可以根據需要選擇各個記錄或逗留方式。這兩種方式需要相同的基本詳細資料，如下所示：

**緯度欄位。**請選取用於識別WGS84 座標系統中的緯度的欄位。

**經度欄位。**請選取用於識別WGS84 座標系統中的經度的欄位。

**時間戳記欄位。**選取用於識別時間或日期的欄位。

## 「各個記錄」選項

使用此方式可向記錄新增一個附加欄位，用於識別該記錄在指定時間的位置。

**衍生。**選取一個或多個空間和時間密度，將根據所選內容衍生新欄位。如需相關資訊，請參閱第 96 頁的『定義空間時間限制密度』。

**欄位名稱延伸。**鍵入您要新增至新欄位名稱的延伸。您可以選擇將此延伸新增為字尾，也可以將其新增為字首。

## 「逗留」選項

可以將逗留視作實體在其中持續或重複出現的位置和/或時間。例如，可以使用它來識別進行定期運輸的車輛，並確定與標準值的任何離差。

逗留偵測器監視實體的移動，並旗標觀察到實體在某個區域內「逗留」的情況。逗留偵測器將旗標的每次逗留自動分配給一個或多個 STB，並使用記憶體內的實體和事件追蹤功能以最優效率偵測逗留。

**STB 密度。**選取空間和時間密度，將根據所選內容衍生新欄位。例如，值 **STB\_GH4\_10MINS** 對應於大小約為 20 公里乘 20 公里且時間視窗為 10 分鐘的四字元 geohash 限制空間。如需相關資訊，請參閱第 96 頁的『定義空間時間限制密度』。

**實體 ID 欄位。**選取要用作逗留 ID 的實體。此識別欄位用於識別事件。

**下限事件數目。**事件是資料中的列。選取將實體視為逗留實體時，該實體的某一事件發生的最少次數。逗留還必須滿足下列**最短停留時間**欄位所指定的條件。

**最短停留時間。**指定實體必須在同一位置停留的最短持續時間。這有助於排除操作，例如可以將等待紅綠燈的車輛從逗留車輛中排除的操作。逗留還必須滿足以上**下限事件數目**欄位所指定的條件。

下列是更多有關逗留所必須滿足的條件的詳細資訊：

假定  $e_1, \dots, e_n$  指示期間  $(t_1, t_n)$  內從給定實體 ID 接收到的所有按時間排序的事件。在以下情況下，這些事件符合逗留條件：

- $n \geq$  下限事件數目
- $t_n - t_1 \geq$  下限停留時間
- 所有事件  $e_1, \dots, e_n$  都在同一個 STB 中發生

**容許逗留跨越 STB 範圍。** 選取此選項後，逗留的定義將不再那麼嚴格並且可以包含一些實體，例如在多個空間時間限制中逗留的實體。例如，如果 STB 定義為整小時，那麼選中此選項會將逗留一小時的實體識別為有效，即使這一小時由午夜前的 30 分鐘和午夜後的 30 分鐘組成也是如此。如果未選取此選項，那麼 100% 的逗留時間都必須在單個空間時間限制之內。

**滿足時間限制的事件的下限比例 (%)。** 只有在選取了容許逗留跨越 STB 範圍的情況下，此選項才可用。使用此選項可控制一個 STB 中報告的逗留事實上可能與另一逗留重疊的程度。選取必須在單一 STB 中發生的事件的下限比例可以識別逗留。如果設定為 25%，並且事件比例為 26%，那麼符合逗留條件。

例如，假定將逗留偵測器配置為至少需要兩個事件（下限事件數目為 = 2），並且連續徘徊時間是在 4 位元組 geohash 限制空間和 10 分鐘限制時間 (STB\_NAME = STB\_GH4\_10MINS) 內至少徘徊 2 分鐘。偵測到逗留時，假定下午 4:57 與 5:07 之間的 10 分鐘時間展開範圍內，在下午 4:58、5:01 和 5:03 發生三個合格事件時，即可認為實體在同一個 4 位元組 geohash 限制空間內停留。合格時間百分比值指定允許用於逗留的 STB，如下所示：

- **100%**。在下午 5:00 - 5:10 的限定時間內報告逗留，而在下午 4:50 - 5:00 的限定時間內不報告逗留（下午 5:01 和 5:03 發生的事件符合合格逗留所需的所有條件，並且其中 100% 的事件在 5:00 - 5:10 限定時間內發生）。
- **50%**。兩個限定時間內的逗留都將進行報告（下午 5:01 和 5:03 發生的事件符合合格逗留所需的所有條件，在這些事件中，至少 50% 發生在 4:50 - 5:00 限定時間內，並且至少 50% 發生在 5:00 - 5:10 限定時間內）。
- **0%**。兩個限定時間內的逗留都將進行報告。

如果指定了 0%，那麼逗留報告將包含那些代表合格持續時間所觸及的每個限定時間的 STB。合格持續時間必須小於或等於 STB 中的限定時間的相應持續時間。換而言之，如果配置的合格持續時間為 20 分鐘，那麼不得將 STB 配置為 10 分鐘。

一旦滿足合格條件就會報告逗留，並且對於每個 STB 不會多次進行報告。假定有三個事件符合逗留條件，並且合格持續時間內發生的全部 10 個事件在同一個 STB 內。在這種情況下，將在發生第三個合格事件時報告逗留。另外 7 個事件都不會觸發、觸發程式逗留報告。

註：

- 逗留偵測器的記憶體內事件資料在處理程序之間不共用。因此，特定實體與特定逗留偵測器節點之間存在親緣性。即，一個實體的傳入移動資料必須始終一致地傳送到追蹤該實體的逗留偵測器節點，在整個執行過程中，這一律是同一個節點。
- 逗留偵測器的記憶體內事件資料並非持久儲存。每當逗留偵測器結束和重新啟動時，所有工作中的逗留都將丟失。這意味著停止並重新啟動處理程序可能會導致系統錯過報告實際逗留。潛在的補救涉及重播部分歷程移動資料（例如，後退 48 小時並重播適用於任何已重新啟動的節點的移動記錄）。
- 必須以時間順序向逗留偵測器輸送資料。

## 定義空間時間限制密度

通過指定要併入在每個空間時間限制 (STB) 中的實體區域和經歷時間，可以選擇空間時間限制的大小（密度）。

**地理密度。** 選取要併入在每個 STB 中的區域的大小。

**時間間隔。** 選取要併入在每個 STB 中的小時數。

**欄位名稱。** 以 STB 作為字首，將根據前兩個欄位中的選擇自動完成此名稱。

---

## 串流式 TCM 節點

「串流式 TCM」節點可用於在一個步驟中對時間因果模型進行建立和分數。

有關時間原因建模的進一步資訊，請參閱《SPSS Modeler 建模節點》手冊的「「時間序列」模型」部分中的「時間因果模型」主題。

### 串流式 TCM 節點 - 「時間序列」選項

在「欄位」標籤上，使用時間序列設定來指定要併入模型系統中的數列。

為適用於您資料的資料結構選取選項。若為多維度資料，請按一下**選取維度**以指定維度欄位。指定的維度欄位順序可定義維度欄位在所有後續對話框及輸出中的出現順序。使用「選取維度」子對話框上的向上和向下箭頭按鈕，可對維度欄位進行重新排序。

若為直欄型資料，術語數列與術語欄位的意義相同。若為多維度資料，包含時間序列的欄位是指度量值欄位。多維度資料的時間序列由度量值欄位和每個維度欄位的值來定義。下列考量同時適用於直欄型資料和多維度資料。

- 指定為候選輸入或目標及輸入的數列可考慮併入每一個目標的模型中。每個目標的模型一律包括目標本身的階階值。
- 指定為強制輸入的數列一律併入每個目標的模型中。
- 至少一個數列必須指定為目標或目標和輸入。
- 如果選取了**使用預先定義角色**，則角色為「輸入」的欄位會設為候選輸入。預先定義的角色未對映為強制輸入。

### 多維度資料

若為多維度資料，可以在網格中指定度量值欄位與相關聯的角色，其中，網格中的每一列指定單一度量值和角色。依預設，模型系統包括網格中每一列的維度欄位所有組合的數列。例如，如果存在 *region* 和 *brand* 的維度，則依預設，將度量值 *sales* 指定為目標表示每個 *region* 和 *brand* 的組合都有個別的銷售目標數列。

針對網格中的每一列，您可以按一下維度的省略符號按鈕，來自訂任何維度欄位的值集。執行此動作會開啟「選取維度值」子對話框。您還可以新增、刪除或複製網格列。

數列計數直欄會顯示目前為相關聯度量值指定的維度值集數。顯示的值可以大於數列的實際數目（每個集一個數列）。如果指定的部分維度值組合未對應於相關聯度量值包含的數列，則會發生這種情況。

### 串流式 TCM 節點 - 選取維度值

若為多維度資料，您可以指定將套用至具有特定角色的特定度量值欄位的維度值，以自訂分析。例如，如果 *sales* 為度量值欄位，而 *channel* 是值為「零售」和「網路」的維度，則您可以指定「網路」銷售是輸入，而「零售」銷售是目標。您也可以指定套用至分析中所使用之所有度量值欄位的維度子集。例如，如果 *region* 是指示地理空間區域的維度欄位，則您可以將分析限制為特定區域。

## 所有值

指定包括現行維度欄位的所有值。此選項是預設值。

## 選取要包括或排除的值

使用此選項來指定現行維度欄位的值集。如果針對眾數選取了**包括**，則只會包括選取的值清單中指定的值。如果針對眾數選取了**排除**，則會包括除選取的值清單中指定的值以外的所有值。

您可以過濾可從中選擇的值集。符合過濾條件的值會出現在**未選取**的值清單的**相符**標籤中，而不符合過濾條件的值則會出現在**不符合**標籤中。無論任何過濾條件，**所有**標籤都會列出所有未選取的值。

- 指定過濾器時，您可以使用星號 (\*) 來表示萬用字元。
- 若要清除現行過濾器，請在「過濾顯示的值」對話框上，為搜尋詞彙指定空白值。

## 串流式 TCM 節點 - 「觀察值」選項

在「欄位」標籤上，使用**觀察**設定來指定用於定義觀察的欄位。

### 由日期/時間定義的觀察

您可以指定觀察由日期、時間或時間戳記欄位定義。除了定義觀察的欄位以外，還可以選取說明觀察的適當時間間隔。根據指定的時間間隔，您還可以指定其他設定，例如觀察間隔（增量）或每週的天數。下列考量僅適用於時間間隔：

- 如果觀察的時間（例如處理銷售單的時間）間隔無規律，請使用值**無規律**。如果選取**無規律**，您必須從「資料規格」標籤上的**時間間隔**設定中，指定用於分析的時間間隔。
- 如果觀察代表日期和時間，而時間間隔是小時數、分鐘數或秒數，則使用**每天小時數**、**每天分鐘數**或**每天秒數**。如果觀察代表無日期參照的時間（期間），且時間間隔為小時數、分鐘數或秒數，則使用**小時數（非週期）**、**分鐘數（非週期）**或**秒數（非週期）**。
- 根據選取的時間間隔，程序可以偵測遺漏的觀察。必須偵測遺漏觀察，因為該程序假設所有觀察的時間間隔相等，且沒有任何遺漏觀察。例如，如果時間間隔為「日」，而日期 2014-10-27 後接 2014-10-29，則遺漏 2014-10-28 的觀察。會針對任何遺漏觀察插補值。可從「資料規格」標籤指定用來處理遺漏值的設定。
- 指定的時間間隔可讓程序偵測相同時間間隔中需要總計在一起的多個觀察，並依據間隔界限（例如月份的第一天）排列觀察，以確保觀察間隔相等。例如，如果時間間隔為「月」，則會將相同月份內的多個日期總計在一起。這種類型的總計稱為**分組**。依預設，在分組觀察時會加總觀察。您可以從「資料規格」標籤上的**總計及分佈**設定中指定不同的分組方法，如**觀察平均數**。
- 對於部分時間間隔，其他設定可定義一般相等時間間隔中的**岔斷**。例如，如果時間間隔為「日」，但只有工作日有效，則您可以指定一週內有五天，一週從星期一啟動。

### 由期間或循環期間定義的觀察

觀察可由一或多個代表期間或重複循環期間（多達任意數目的循環水準）的整數欄位定義。您可以使用此結構來說明不適合其中一個標準時間間隔的觀察數列。例如，可使用代表年份的循環欄位以及代表月份的期間欄位來說明只有 10 個月的財政年度，其中一個循環的長度為 10。

指定循環期間的欄位可定義週期性水準的階層，其中最低水準由**期間**欄位定義。下一個最高水準由水準為 1 的循環欄位指定，接著由水準為 2 的循環欄位指定，依此類推。每個水準（除最高水準以外）的欄位值相對於下一個最高水準必須是週期性的。最高水準的值不能為週期性值。例如，在財政年度為 10 個月的情況下，月份在年份中是週期性的，但年份不是週期性的。

- 特定水準上的循環長度是下一個最低水準的週期。以財政年度為例，只有一個循環水準且循環長度為 10，因為下一個最低水準代表月份，而指定的財政年度有 10 個月。
- 針對不從 1 啟動的任何週期性欄位指定起始值。偵測遺漏值時需要此設定。例如，如果週期性欄位從 2 啟動，但起始值指定為 1，則程序會假設該欄位每個循環中的第一個期間都存在遺漏值。

## 串流式 TCM 節點 - 「時間間隔」選項

用於分析的時間間隔可能與觀察的時間間隔不同。例如，如果觀察的時間間隔為「日」，您可以為分析的時間間隔選擇「月」。然後，在模型建置之前，將資料總計從每日改為每月。您也可以選擇將資料分佈從較長時間間隔改成較短時間間隔。例如，如果觀察是按季度發生的，則您可以將資料從按季度分佈改成按月分佈。

執行分析的時間間隔的可用選項，視觀察的定義方式以及這些觀察的時間間隔而定。特別地，如果觀察是由循環期間定義，則只支援總計。在此情況下，分析的時間間隔必須大於或等於觀察的時間間隔。

從「資料規格」標籤上的時間間隔設定中，指定用於分析的時間間隔。您可以從「資料規格」標籤上的總計及分佈設定指定總計或分佈資料所採用的方式。

## 串流式 TCM 節點 - 「聚集和分佈」選項

### 聚集函數

當用於分析的時間間隔長於觀察的時間間隔時，會總計輸入資料。例如，如果觀察時間間隔是「日」，而分析的時間間隔是「月」，則會執行總計。提供了下列聚集函數：平均數、加總、眾數、最小值或最大值。

### 分佈函數

當用於分析的時間間隔短於觀察的時間間隔時，會分佈輸入資料。例如，如果觀察時間間隔是「季度」，而分析的時間間隔是「月」，則會執行分佈。提供了下列分佈函數：平均數或加總。

### 分組函數

如果觀察由日期/時間定義且在相同的時間間隔發生多個觀察，則可套用分組。例如，如果觀察的時間間隔是「月」，則相同月份中的多個日期會分組在一起，並與其發生所在的月份相關聯。您可以使用的分組函數如下：mean、sum、mode、min 或 max。如果觀察由日期/時間定義，且觀察的時間間隔指定為「無規律」，則一律會執行分組。

註：雖然分組是一種總計形式，但它是在處理所有遺漏值之前執行，而正式的總計是在處理完所有遺漏值之後執行。當觀察的時間間隔指定為「無規律」時，只能使用分組函數來執行總計。

### 總計跨日期觀察與前一天的值

指定是否將在跨日期界限的時間內的觀察與前一天的值總計在一起。例如，針對一天 8 小時制（在 20:00 啟動）的每小時觀察，這個設定會指定在 00:00 和 04:00 之間的觀察是否併入前一天的總計結果。只有在觀察時間間隔是「每天小時數」、「每天分鐘數」或「每天秒數」，且分析的時間間隔為「日」時，此設定才適用。

### 所指定欄位的自訂設定

您可以逐個欄位指定聚集函數、分佈函數及分組函數。這些設定會置換聚集函數、分佈函數及分組函數的預設值。

## 串流式 TCM 節點 - 「遺漏值」選項

輸入資料中的遺漏值會取代為插補值。提供了下列取代方法：

### 線性插補

使用線性插補來取代遺漏值。該選項會用遺漏值出現之前的最後一個有效值，還有遺漏值出現之後的第一個有效值當作內插。如果數列中的第一個或最後一個觀察具有遺漏值，則會使用數列開頭或結尾的兩個最近非遺漏值。

### 數列平均數

以整個數列的平均數置換遺漏值。

### 鄰近點平均數

以有效周圍值的平均數置換遺漏值。附近點的指距是指用來計算平均數的遺漏值，其之前及之後的有效值個數。

### 鄰近點中位數

以有效周圍值的中位數置換遺漏值。附近點的指距是指用來計算中位數的遺漏值，其之前及之後的有效值個數。

### 線性趨勢

此選項會使用數列中的所有非遺漏觀察來擬合簡式線性迴歸模型，然後使用該模型來插補遺漏值。

其他設定：

### 遺漏值百分比上限 (%)

指定任何數列可接受的遺漏值百分比上限。其遺漏值多於所指定上限的數列將會從分析中排除。

## 串流式 TCM 節點 - 「一般資料」選項

### 每個維度欄位的最大相異值數目

此設定適用於多維度資料，並指定任何一個維度欄位可接受的最大相異值數目。依預設，此限制設為 10000，但它可以增加到任意大數目。

## 串流式 TCM 節點 - 「一般建立」選項

### 信賴區間寬度 (%)

此設定可控制預測及模型參數兩者的信賴區間。您可以指定小於 100 的任何正數值。依預設，會使用 95% 信賴區間。

### 每個目標的最大輸入數

此設定指定每個目標模型可接受的最大輸入數目。您可以指定 1 - 20 範圍內的整數。每個目標的模型一律包括本身的落階值，因此，將此值設為 1 表示唯一的輸入是目標本身。

### 模型容差

此設定可控制用來判定每個目標最佳輸入集的疊代處理程序。您可以指定大於零的任何值。預設值是 0.001。模型容差是用於預測值選取的停止準則。它會影響最終模型中包含的預測值數目。但如果某個目標能很好地預測自身，則最終模型中可能不會包含其他預測值。可能需要進行一些嘗試和錯誤（例如，如果您將此值設定為較高值，則您可以嘗試將其設為較小的值，以查看是否可以包含其他預測值）。

### 離群值臨界值 (%)

如果觀察為離群值的機率（從模型計算得出）超出此臨界值，則會將觀察標示為離群值。您可以指定 50 - 100 範圍內的值。

### 每個輸入的落階數

此設定指定每個目標模型中每個輸入的落階項目數。依預設，會透過用於分析的時間間隔自動判定落階項目數。例如，如果時間間隔為月數（增量為一個月），則落階數為 12。您可以選擇性地明確指定落階數。指定的值必須是 1-20 範圍內的整數。

### 使用現有模型繼續估計

如果您產生了時間原因模型，選取此選項可重複使用指定給該模型的準則設定，而不必建置新的模型。使用這種方式，您可以根據與之前相同的模型設定，但使用更新的資料，來重新估計並產生新預測，從而節省時間。

## 串流式 TCM 節點 - 「估計期間」選項

依預設，估計期間啟動於所有數列中最早觀察的時間，在所有數列中最晚觀察的時間結束。

### 依啟動與結束時間

您可以同時指定估計期間的啟動與結束時間，也可以只指定啟動時間或結束時間。如果您省略估計期間的啟動或結束時間，則會使用預設值。

- 如果觀察是由日期/時間欄位來定義，請以日期/時間欄位使用的相同格式輸入啟動時間與結束時間的值。
- 如果觀察是由循環期間定義，請為每個循環期間欄位指定一個值。每個欄位都會顯示在個別的直欄中。

### 依最晚或最早時間間隔

將估計期間定義為資料中從最早時間間隔開始或在最晚時間間隔的指定時間間隔數（可包含選用偏移）結束。在此環境定義中，時間間隔是指分析的時間間隔。例如，假定觀察是按月份，而分析的時間間隔是按季度。指定最晚並為時間間隔數指定值 24 表示最晚 24 個季度。

您可以選擇性地排除指定的時間間隔數。例如，指定最晚 24 個時間間隔並指定 1 代表要排除的數目，表示估計期間由最後一個間隔之前的 24 個間隔組成。

## 串流式 TCM 節點 - 「模型」選項

### 模型名稱

您可以對模型指定自訂名稱，也可以接受自動產生的名稱，即 *TCM*。

**預測** 用於將記錄延伸至未來的選項可設定要在估計期間結束以外預測的時間間隔數目。在這種情況下，時間間隔是分析的時間間隔，它在「資料規格」標籤上指定。要求預測時，會針對也不是目標的任何輸入數列自動建置自迴歸模型。然後，在預測期間使用這些模型產生那些輸入數列的值。此設定沒有上限限制。

---

## CPLEX 最佳化節點

CPLEX 最佳化節點可讓您透過最佳化程式設計語言 (OPL) 模型檔使用複雜的數學 (CPLEX) 式最佳化。IBM Analytical Decision Management 產品中提供了此功能，但您現在還可以在 SPSS Modeler 中使用 CPLEX 節點而不需要 IBM Analytical Decision Management。

如需 CPLEX 最佳化和 OPL 的相關資訊，請參閱 ILOG CPLEX Optimization Studio 文件。

CPLEX Optimization 節點支援多個資料來源，或多維度送入資料。可以將數個節點連接至 CPLEX Optimization 節點，每一個先前節點都可用來提供資料給 OPL 模型計算 – 設定為具有個別欄位對映的個別元組集。

輸出由 CPLEX Optimization 節點產生的資料時，可以將資料來源中的原始資料一起作為單一索引輸出，或作為結果的多維度索引輸出。

**註：**在 **IBM SPSS Modeler Server** 上執行包含 CPLEX Optimization 節點的串流時，依預設會使用內嵌的 Community 版 CPLEX 程式庫。它具有以下限制：1000 個變數及 1000 個限制項。如果您安裝完整版本的 IBM ILOG CPLEX 且想要改為使用完整版本的 CPLEX 引擎（其沒有此類限制），請針對您的平台完成下列步驟。

- 在 Windows 上，編輯 options.cfg 並新增 OPL 程式庫路徑。例如：

```
cpLEX_opl_lib_path="<CPLEX_path>\opl\bin\<Platform_dir>"
```

其中 <CPLEX\_path> 是 CPLEX 安裝目錄（例如 C:\Program Files\IBM\ILOG\CPLEX\_Studio127），<Platform\_dir> 是平台專用目錄（例如 x64\_win64）。



- 在 Linux 上，編輯 modelersrv.sh 並新增 OPL 程式庫路徑。例如：

```
CPLEX_OPL_LIB_PATH=<CPLEX_path>/opl/bin/<Platform_dir>
```

其中 <CPLEX\_path> 是 CPLEX 安裝目錄（例如 /root/Libs\_127\_FullEdition/Linux\_x86\_64），<Platform\_dir> 是平台專用目錄（例如 x86-64\_linux）。

註：在 **SPSS Modeler** 中執行包含 CPLEX 最佳化的串流時，依預設，會使用內嵌的 Community 版本 CPLEX 程式庫。它具有以下限制：1000 個變數及 1000 個限制項。如果您安裝完整版本的 IBM ILOG CPLEX 且想要改為使用完整版本的 CPLEX 引擎（其沒有此類限制），請針對您的平台完成下列步驟。

- 在 Windows 上，新增 OPL 程式庫路徑作為 modelerrun.exe 的指令行引數。例如：

```
-o cplex_opl_lib_path="<CPLEX_path>\opl\bin\<Platform_dir>"
```

其中 <CPLEX\_path> 是 CPLEX 安裝目錄（例如 C:\Program Files\IBM\ILOG\CPLEX\_Studio127），<Platform\_dir> 是平台專用目錄（例如 x64\_win64）。

- 在 Linux 上，編輯 modelerrun 並新增 OPL 程式庫路徑。例如：

```
CPLEX_OPL_LIB_PATH=<CPLEX_path>/opl/bin/<Platform_dir>
```

其中 <CPLEX\_path> 是 CPLEX 安裝目錄（例如 /root/Libs\_127\_FullEdition/Linux\_x86\_64），<Platform\_dir> 是平台專用目錄（例如 x86-64\_linux）。

## CPLEX Optimization 節點的設定選項

CPLEX Optimization 節點的標籤「選項」包含下列欄位。

**OPL 模型檔。** 選取最佳化程式設計語言 (OPL) 模型檔案。

**OPL 文件。** 選取 OPL 模型後，內容將顯示在此處。

### 輸入資料

在「輸入資料」標籤上，**資料來源**下拉清單會列出已連接至現行 CPLEX Optimization 節點的所有資料來源（先前節點）。從下拉清單中選取資料來源會重新整理下面的**輸入對映**區段。按一下**套用所有欄位**以自動產生所選資料來源的所有欄位對映。將會自動移入**輸入對映**表格。

在 OPL 模型中輸入對應於送入資料的元組集名稱。然後，必要的話，可以驗證所有元組欄位是否已根據其在元組定義中的順序對映至資料輸入欄位。

設定資料來源的輸入對映之後，您可以從下拉清單中選取不同資料來源並重複此程序。將自動儲存之前的資料來源對映。完成時，按一下**套用**或**確定**。

### 其他資料

在「其他資料」標籤上，如果您需要指定任何其他資料進行最佳化，請使用 **OPL 資料**部分。

### 輸出

當輸出為決策變數時，它必須採用先前資料來源（送入的資料）作為索引，並且這些索引必須已在「輸入資料」標籤上的**輸入對映**區段中預先定義。目前不支援其他類型的決策變數。決策變數可以具有單一索引或多個索引。SPSS Modeler 將一起輸出 CPLEX 結果與所有或部分原始送入資料，這與其他 SPSS Modeler 節點一致。所參照的對應索引必須已在下面說明的**輸出元組**欄位中指定。

在「輸出」標籤上，選擇輸出模式（原始輸出或決策變數）並指定其他選項（如果適用）。「原始輸出」選項將直接輸出目標函數，而與名稱無關。

**OPL 中的目標函數值變數名稱。**如果您已選取決策變數輸出模式，則將啟用此欄位。輸入 OPL 模型中目標函數值變數的名稱。

**輸出的目標函數值欄位名稱。**輸入要用在輸出中的欄位名稱。預設值是 `_OBJECTIVE`。

**輸出元組。**輸入送入的資料中預先定義元組的名稱。它充當決策變數的索引，並預期與變數輸出一起輸出。「輸出元組」應該與 OPL 中的決策變數定義一致。如果有多個索引，則元組名稱必須以逗點 (,) 結合。

**變數輸出。**新增一個或多個要包含在輸出中的變數。

---

## 第 4 章 欄位作業節點

---

### 欄位作業概觀

經過起始資料研究之後，您可能需要在準備分析的過程中選取、清除或構造資料。「欄位作業」選用區包含多數適用於這種轉換和準備的節點。

例如，使用「衍生」節點，可以建立目前資料中並未呈現的屬性。或者，使用「分級」節點可以自動針對目標分析進行欄位值的重新編碼。您可能會發現自己使用「類型」節點的頻率很高，該節點可用於為資料集中每個欄位分配測量層次、值和建模角色。其作業對於處理遺漏值和下游建模十分實用。

「欄位作業」選用區包含下列節點：



自動資料預備 (ADP) 節點可分析您的資料並識別修正，篩選出存在問題或可能無用的欄位，並在適當的情況下衍生新的屬性，通過智能篩選和取樣技術改進效能。您可以完全自動化地套用節點，容許節點選擇並套用修正，或者也可在修正前預覽變更，按照需要接受、拒絕或修正。



「類型」節點指定欄位 meta 資料和內容。例如，您可以指定每個欄位的測量層次（連續、列名、序數或旗標）、設定用於處理遺漏值和系統空值的選項、設定用於建模的欄位的角色、指定欄位和值標籤，以及為欄位指定值。



「過濾器」節點用於過濾（捨棄）欄位、更名欄位並將欄位從一個來源節點對映到另一來源節點。



「衍生」節點將修改資料值或根據一個或多個現有欄位建立新欄位。它可建立的欄位類型包括公式、旗標、列名、狀態、計數和條件。



「總體」節點可結合使用兩個或兩個以上模型塊，這樣所獲得的預測會比通過任意一個模型獲得的預測更為準確。



填入節點會取代欄位值並變更儲存。您可以選擇基於 CLEM 條件（例如 @BLANK(@FIELD)）的取代數值。或者，您可以選擇將所有空白或空值取代為特定值。「填入器」節點經常結合「類型」節點使用以取代遺漏值。



「匿名化」節點用於轉換欄位名稱和欄位值的下游代表方式，從而掩飾了原始資料。如果要容許其他使用者使用敏感資料（如客戶名稱或其他詳細資料）建立模型，這種節點將十分有用。



「再分類」節點可將一組種類值轉換為另一組值。重新分類對收合種類或重新分組分析資料非常有用。



「分級」節點根據一個或多個現有連續（數值型範圍）欄位的值自動建立新的列名（集合）欄位。例如，用戶可將連續收入欄位轉換為一個包含各群組收入的新的種類欄位，作為其與平均數之間的離差。一旦您為新欄位建立了 Bin，就可以根據切點來產生「衍生」節點。



「回頭率、頻率及消費金額 (RFM) 分析」節點可讓您在數量上判定哪些客戶有可能是最佳客戶，方法是檢查距離客戶上次向您購買的時間（回頭率）、客戶多長時間購買一次（頻率），以及客戶在所有交易中的花費金額（消費金額）。



分割區節點可產生分割區欄位，該欄位可將資料分割為個別的子集以便在模型建置的訓練、正在測試和驗證階段使用。



「設為旗標」節點根據為一個或多個列名欄位定義的種類值獲取多個旗標欄位。



「重組」節點可將一個列名欄位或旗標欄位轉換為一群組欄位（該欄位群組由已成為另一欄位的值移入）。例如，給定一個名為 支付類型的欄位，其值為貸方、現金和借方，則將建立三個新欄位（貸方、現金、借方），每個欄位可能包含實際支付的值。



轉置節點交換列與欄中的資料，以便記錄變成欄位，欄位變成記錄。



使用「時間間隔」節點可以指定時間間隔並衍生用於估計或預測的新時間欄位。支援從秒到年的全系列時間區間。



「歷史」節點將建立新欄位，其中包含之前記錄中的欄位資料。「歷程」節點通常用於序列資料，例如時間序列資料。使用「歷史」節點前，您可能希望使用「排序」節點對此資料進行排序。



「欄位重新排序」節點定義了用於顯示下游欄位的自然順序。此順序會影響欄位在各個位置（例如表格、清單與「欄位」選擇器）中的顯示方式。製作大型資料集時，此作業有助於使所需欄位更為直觀。



在 SPSS Modeler 中，表示式建置器空間函數、「空間-時間預測」(STP) 節點和「地圖視覺化」節點之類的項目使用投影座標系統。使用「重新投影」節點可以變更所匯入的任何使用了地理座標系統的資料的座標系統。

其中某些節點可以通過「資料審核」節點所建立的審核報告直接產生。請參閱第 279 頁的『產生其他用於資料預備的節點』主題，以取得更多資訊。

## 自動資料準備

準備資料以供分析是任何專案中最重要的一步驟之一，也是傳統上最耗時的步驟之一。「自動資料準備」(ADP) 可為您處理工作、分析您的資料並識別修正、篩選出有問題或可能無用的欄位、在適當時衍生新屬性，以及透過智慧型篩選技術增進效能。您可以全自動方式使用演算法，以允許其選擇並套用修正，或以互動方式使用演算法，以在進行變更前先行預覽，然後視需要接受或拒絕變更。

使用 ADP 可讓您快速、輕鬆地準備資料以建置模式，不需事先了解統計相關概念。模式將可更快地建置並進行資料評分，此外，使用 ADP 可提高自動建模程序，例如：模型更新與冠軍/挑戰者 (champion / challenger) 的穩健性。

註：ADP 準備要進行分析的欄位時，會建立包含調整或轉換的新欄位，而非取代舊欄位現有的值和性質。舊欄位不會用於進一步分析；其角色會設定為「無」。

範例。某資源有限的保險公司，打算調查屋主的保險理賠，希望建立標示可疑潛在詐欺理賠的模式。建置模型之前，他們將使用自動資料準備來準備建模用的資料。由於他們希望在套用轉換前檢閱提議的轉換，因此會在互動式模式使用自動資料準備。

某汽車業集團會追蹤各種個人汽車的銷售額。為了能夠識別表現超前與表現不佳的模式，他們希望建立汽車銷售額與汽車特性之間的關係。他們會使用自動的資料準備來準備分析用的資料，以及使用準備「之前」與「之後」的資料建置模型，以瞭解結果有何差異。

您的目標是什麼？自動資料準備會建議資料準備步驟，這些步驟將影響其他演算法建置模型的速度，並提升這些模型的預測能力。其中包含轉換、建立和選取功能。亦可轉換目標。您可以指定資料準備步驟遵循的模式建立優先順序。

- **權衡速度與準確度。**此選項準備資料時，會兼顧模式建立演算法處理資料的速度，以及預測的準確度。
- **最佳化速度。**此選項準備資料時，會優先考慮模式建立演算法處理資料的速度。當您正在處理非常大型的資料集或想快速找到答案時，請選取此選項。
- **最佳化準確度。**此選項準備資料時，會優先考慮模式建立演算法所產生預測的準確度。
- **自訂分析。**當您想在「設定」標籤中手動變更演算法時，請選取此選項。請注意，若您之後對「設定」標籤中的選項進行變更，但該變更與任一項目標不符時，會自動選取此設定。

## 訓練節點

ADP 節點會執行為程序節點，且運作方式和「類型」節點類似；訓練 ADP 節點相當於實例化「類型」節點。當分析執行後，只要上游的資料模型未變更，指定的轉換便會套用到資料而不會再進一步分析。就像「類型」和「過濾器」節點一樣，ADP 節點中斷連線時會記住資料模型與轉換，因此重新連線時就不需要再重新訓練；這可讓您在典型資料的子集上訓練 ADP 節點，然後再複製或部署該 ADP 節點以頻繁地用於作用中資料。

## 使用工具列

工具列可讓您執行和更新資料分析的顯示，並產生可搭配原始資料使用的節點。

- **產生** 從這個功能表，您可以產生「過濾」或「衍生」節點。請注意，當「分析」標籤中有顯示分析時，才會提供此功能表。

「過濾器」節點會移除轉換後的輸入欄位。若您將 ADP 節點設定為在資料集中保留原始輸入欄位，這樣將會還原原始的輸入組，以供您依據輸入來解讀分數欄位。例如，若您想產生分數欄位對照各種輸入的圖形，此功能會很有用。

「衍生」節點可以還原原始資料集和目標單位。當 ADP 節點包含重新調整範圍目標 (亦即在「準備輸入與目標」窗格中選取 Box-Cox 重新調整時) 的分析時，您才可以產生「衍生」節點。若目標不是範圍或未選取 Box-Cox 重新調整，您就無法產生「衍生」節點。請參閱第 116 頁的『產生一個衍生節點』主題，以取得更多資訊。

- **檢視** 包含的選項可控制「分析」標籤中顯示的項目。這包含主畫面與鏈結的視圖的圖形編輯控制項和顯示選項。
- **預覽** 顯示將套用到輸入資料的轉換範例。
- **分析** 資料使用目前的設定建立分析，並在「分析」標籤中顯示結果。
- **清除** 分析刪除現有的分析 (目前有分析存在時才可使用)。

## 節點狀態

IBM SPSS Modeler 構圖區中的 ADP 節點狀態是以箭頭或圖示上的刻度來表示，顯示是否執行過分析。

如需使用「自動資料準備」節點執行之計算的相關資訊，請參閱 *IBM SPSS Modeler Algorithms Guide* 中的 *Automated Data Preparation algorithms* 小節。該手冊以 PDF 格式提供，位於安裝磁碟的 \Documentation 中，作為產品下載的一部分，或位於 Web 上。

## 欄位標籤

您必須先指定要用做目標與輸入的欄位，然後才能建置模型。在某些例外狀況下，所有建模節點會使用來自上游「類型」節點的欄位資訊。若您使用「類型」節點來選取輸入和目標欄位，則無須變更此標籤的任何內容。

**使用類型節點設定。**此選項會告知節點使用來自上游「類型」節點的欄位資訊。此為預設值。

**使用自訂設定。**此選項會告知節點使用此處指定的欄位資訊，而不使用任何上游「類型」節點的指定欄位資訊。選取此選項後，請視需要於下方指定欄位。

**目標。**針對需要使用一或多個目標欄位的模式，請選取一個或多個目標欄位。這與在「類型」節點中將欄位角色設為目標類似。

**輸入。**選取一個或多個輸入欄位。這與在「類型」節點中將欄位角色設為輸入類似。

## 設定標籤

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

## 欄位設定

**使用頻率欄位。**此選項可讓您選取某個欄位作為一個頻率加權。如果您的訓練資料中的每個記錄代表多個單位（例如，如果您正使用聚集資料），請使用此選項。欄位值應該為每筆記錄代表的單位數。

**使用加權欄位。**此選項可讓您選取某個欄位作為一個觀察值加權。觀察值加權是用來說明輸出欄位不同等級間的變異數差異。

**何處理建模中排除的欄位。**指定如何處理排除的欄位；您可以選擇將它們從資料中過濾掉或僅把它們的角色設定為無。

注意：此動作還將套用到已轉換的目標。例如，如果目標的新衍生版本用作目標欄位，原始目標將被過濾或設定為無。

**如果傳入欄位與現有分析不相符。**指定在您執行經過培訓的 ADP 節點時，如果接收資料集中遺漏一個或多個所需輸入欄位會怎樣。

- 停止執行並保留現有分析。這將停止執行過程，保留目前分析資訊並顯示錯誤。
- 清除現有分析並分析新資料。這將清除現有分析、分析接收資料並對該資料套用建議的變換。

## 準備日期與時間

許多模式建立演算法均無法直接處理日期與時間詳細資料；這些設定可讓您衍生新的期間資料，以作為您現有資料中日期和時間的模式輸入。包含日期與時間的欄位必須預先定義日期或時間儲存類型。原始的日期與時間欄位在自動資料準備之後將不建議作為模式輸入。

**準備建模的日期與時間。**取消選取此選項會停用全部其他「準備日期與時間」控制項，同時維持選擇。

**計算至參考日期需經過的時間。**這會產生自各包含日期變數的參考日期至今的年/月/天數。

- 參考日期。指定輸入資料的日期資訊中，作為計算持續期間起始日的日期。選取今天日期表示執行 ADP 時，永遠會使用目前的系統日期。若要使用特定日期，請選取「固定日期」並輸入必要的日期。第一次建立節點時，就會在「固定日期」欄位中自動輸入目前日期。
- 日期持續期間的單位。指定 ADP 應自動決定日期持續期間的單位，或從「年數」、「月」或「天數」的「固定單位」中選取。

**計算至參考時間需經過的時間。**這會產生自各包含時間變數的參考時間至今的小時/分鐘/秒數。

- **參考時間。**指定輸入資料的時間資訊中，作為計算持續期間起始時間的時間。選取「**目前時間**」表示執行 ADP 時，永遠會使用目前的系統時間。若要使用特定時間，請選取「**固定時間**」並輸入必要的詳細資料。第一次建立節點時，就會在「**固定時間**」欄位中自動輸入目前時間。
- **時間持續期間的單位。**指定 ADP 應自動決定時間持續期間的單位，或從「**時數**」、「**分鐘數**」或「**秒數**」的「**固定單位**」中選取。

**擷取循環時間元素。**使用這些設定將單一日期或時間欄位分割為一或多個欄位。例如，若您選取這三個日期的勾選框，輸入日期欄位 "1954-05-23" 會分割為三個欄位：1954、5 和 23，且會分別使用「**欄位名稱**」畫面中定義的字尾，並且會忽略原始日期。

- **從日期擷取。**對於任何日期輸入，指定您要擷取年、月、日或任何組合。
- **從時間擷取。**對於任何時間輸入，指定您要擷取小時、分鐘、秒或任何組合。

## 排除欄位

品質不佳的資料會影響預測的準確度；因此，您可以指定可接受的輸入等級品質功能。所有常數欄位或含有 100% 遺漏值的欄位都會自動被排除。

**排除低品質的輸入欄位。**取消選取此選項會停用全部其他「**排除欄位**」控制項，同時維持選擇。

**排除具有太多遺漏值的欄位。**超過指定遺漏值百分比的欄位會被移除，不執行進一步分析。即使指定大於或等於 0 (等於取消選取此選項)，而且小於或等於 100 的數值，所有含有遺漏值的欄位還是會遭自動排除。預設值是 50。

**排除具有太多唯一類別的名義欄位。**超過指定類別數目的名義欄位會被移除，不執行進一步分析。指定一個正整數。預設值為 100。這對自動從建模移除包含記錄唯一資訊 (例如 ID、位址或名稱) 的欄位很實用。

**排除單一類別中具有太多數值的類別欄位。**含有超過指定記錄百分比之類別的序數和名義欄位會被移除，不執行進一步分析。即使指定大於或等於 0 (等於取消選取此選項)，而且小於或等於 100 的數值，常數欄位還是會遭自動排除。預設值是 95。

## 準備輸入和目標

由於沒有資料處於適合正在處理的完美狀態，您可能希望在執行分析之前調整一些設定。例如，這可能包含移除離群值，指定如何處理遺漏值或調整類型。

附註：如果您變更此畫面上的值，則會自動更新目標標籤以選取自訂分析選項。

**對輸入和目標欄位進行建模準備。**開啟或關閉畫面上的所有欄位。

**調整類型並提高資料品質。**對於輸入和目標，您可以分別指定幾個資料變換；這是因為您可能不希望變更目標值。例如，以美元為單位的收入預測變量比以對數 (美元) 測量的預測變量更有意義。此外，如果目標有遺漏值，將沒有預測增益來填充遺漏值，而在輸入中填充遺漏值可啟用一些演算法來處理可能會丟失的資訊。

這些變換的其他設定 (如離群值截斷值) 對於目標和輸入都通用。

您可以為輸入或目標、或兩者選取下列設定：

- **調整數值欄位的類型。**選取此選項以確定序數測量層次的數值型欄位是否可以轉換為連續，反之亦然。您可以指定下限和上限臨界值以控制轉換。
- **重新排序名義欄位。**選取此選項以按從小到大的種類順序排序名義 (集合) 欄位。
- **取代連續欄位中的離群值。**指定是否取代離群值；將其與以下方的取代離群值的方法結合使用。
- **連續欄位：用平均數取代遺漏值。**選取本選項以取代連續 (範圍) 特徵的遺漏值。



- **名義欄位：**用模式取代遺漏值。選取本選項以取代名義（集合）特徵的遺漏值。
- **序數欄位：**用中位數取代遺漏值。選取本選項以取代依序（依序集合）特徵的遺漏值。

**序數欄位數值的最大數量。**指定重新定義序數（排定次序的集）欄位為連續（範圍）的臨界值。預設值為 10；因此，如果一個序數欄位有超過 10 個種類，它將被重新定義為連續（範圍）。

**連續欄位數值的最小數量。**指定重新定義尺度或連續（範圍）欄位為依序（依序集合）的臨界值。預設值是 5；因此，如果連續欄位有少於 5 個值，它將被重新定義為依序（依序集合）。

**離群值分割值。**指定離群值截斷準則（採用標準差測量），預設值是 3。

**替換離群值的方法。**選取是否通過修整（強制）截斷值、將其刪除或設定為遺漏值來取代離群值。在任何離群值被設定為遺漏值後，將按照以上方所選的遺漏值處理設定進行處理。

**將所有連續輸入欄位置於一般尺度上。**要正規化連續輸入欄位，選中本勾選框並選擇正規化方法。預設為 **z** 分數變換，其中您可以指定最終平均數（預設值是 0），以及最終標準差（預設值是 1）。或者也可以選擇使用下限/上限變換並指定最小值和最大值，預設值分別為 0 和 100。

當您在「構建&選取特徵」窗格上選取執行特徵構建時，本欄位特別有用。

**重新調整具有 Box-Cox 變換的連續目標。**要正規化連續（尺度或範圍）目標欄位，請選中此勾選框。Box-Cox 變換的最終平均數預設值為 0，同時最終標準差預設值為 1。

注意：如果您選擇正規化目標，那麼將會轉換目標的維度。這時，您可能需要產生「衍生」節點以套用逆模型變換，將變換後的單位轉回可識別的格式，以供進一步正在處理。請參閱第 116 頁的『產生一個衍生節點』主題，以取得更多資訊。

## 構建和功能選擇

為提高資料預測能力，您可以根據現有欄位轉換輸入欄位或構建新欄位。

附註：如果您變更此畫面上的值，則會自動更新目標標籤以選取自訂分析選項。

**轉換、構造和選取輸入欄位以提高預測能力。**開啟或關閉畫面上的所有欄位。

**合併稀疏類別，以最大化與目標之間的關聯。**選中此選項可以減少與目標關聯的需製作的變數數目，得到更簡約的模型。必要的話，請變更預設機率值 0.05。

注意，如果所有種類合併為一個種類，欄位的原始和衍生版本將被排除，因為它們沒有作為預測值的值。

**若無目標，則根據個數合併稀疏類別。**如果您處理的是沒有目標的資料，那麼可以選擇合併依序（依序集合）和/或名義（集合）特徵的鬆散種類。指定識別要合併種類的資料中觀察值或記錄的下限百分比，預設值為 10。

使用下列規則合併種類：

- 合併不能在二元欄位上執行。
- 如果在合併過程中只有兩個種類，合併將停止。
- 如果沒有原始種類，或者合併期間所建立種類的觀察值百分比均不少於指定下限觀察值百分比，合併將停止。

**在保留預測能力的同時分級連續欄位。**如果您擁有的資料包含種類目標，那麼可以採用強關聯對連續輸入分級，以改進正在處理效能。如果需要，變更 0.05 的預設齊次子集機率值。

若 binning 作業會導致特定欄位有一個 bin，則會排除次序和經過 bin 處理之版本的欄位，因為它們沒有作為預測值的值。

注意：ADP 中的分級與 IBM SPSS Modeler 其他部分的最佳分級不同。最佳離散化使用熵資訊將連續變數轉換為種類變數；這需要在記憶體中對全部資料進行排序和儲存。ADP 使用齊次子集來離散化連續變數，這意味著 ADP 離散化不需要在記憶體中對全部資料進行排序和儲存。通過使用齊次子集方法離散化連續變數，離散化後個種類數總是小於或等於目標種類數。

**執行功能選擇。** 選取此選項將移除相關係數低的特徵。必要的話，請變更預設機率值 0.05。

該選項僅適用於目標為連續的連續輸入特徵，以及種類輸入特徵。

**執行功能建構。** 選中此選項，以從包含多個現有特徵的組合中衍生新特徵，現有特徵隨後將從建模過程中捨棄。

此選項只適用於目標是連續或沒有目標的連續輸入功能。

## 欄位名稱

為輕鬆識別新功能和轉換功能，ADP 會建立並套用基本新名稱、字首及字尾。您可以修正這些名稱，以更符合您的需求與您的資料。若您要指定其他標記，必須在下游的「類型」節點中進行。

**已轉換與已建構的欄位。** 指定要套用至轉換後的目標和輸入欄位的副檔名。

請注意，在 ADP 節點中，依據您選擇處理未使用欄位的方式而定，將字串欄位設為空值可能會導致錯誤。若「設定」標籤的「欄位設定」窗格中，「如何處理從模式建立排除的欄位」設為「過濾出未使用的欄位」，則輸入和目標的副檔名可以設為空值。原始的欄位會被過濾出來，而轉換後的欄位將以它們的名稱儲存，如此一來，轉換後的欄位名稱就會和原始欄位相同。

不過，若您選擇將未使用欄位的方向設為「無」，因為您將嘗試建立重複的欄位名稱，因此目標和輸入欄位若含有空白或虛無的副檔名將會導致錯誤。

此外，請指定要套用至透過「選取」和「建構」設定建構之任何功能的字首名稱。如此便會透過將數值字尾附加到此字首根名稱的方式來建立新名稱。數字的格式會根據衍生多少新功能而定，例如：

- 1-9 個建構的功能將命名為：功能 1 到功能 9。
- 10-99 個建構的功能將命名為：功能 01 到功能 99。
- 100-999 個建構的功能將命名為：功能 001 到 999，依此類推。

這可確保無論有多少個功能，建構的功能將依據合理的順序排序。

**從日期與時間計算的持續時間。** 指定副檔名以套用至從日期與時間計算的持續時間。

**從日期與時間擷取的循環元素。** 指定副檔名以套用至從日期與時間擷取的循環元素。

## 分析標籤

1. 當 ADP 設定 (包括對「目標」、「欄位」及「設定」標籤的任何變更) 滿足您的需求時，請按一下「分析資料」；演算法會將設定套用至資料輸入，並在「分析」標籤中顯示結果。

「分析」標籤包含表格和圖形輸出，這些輸出摘要說明資料的處理，並顯示關於可如何修改或改善資料以進行評分的建議。您之後可以檢視及接受或拒絕這些建議。

「分析」標籤由兩個畫面組成，主要視圖位於左邊，鏈結或輔助視圖位於右邊。主要視圖有三種：

- 欄位處理摘要 (預設值)。請參閱『欄位處理摘要』主題，以取得更多資訊。
- 欄位。請參閱『欄位』主題，以取得更多資訊。
- 動作摘要。請參閱第 112 頁的『動作摘要』主題，以取得更多資訊。

鏈結/輔助視圖有四種：

- 預測能力 (預設值)。請參閱第 113 頁的『預測能力』主題，以取得更多資訊。
- 欄位表格。請參閱第 113 頁的『欄位表格』主題，以取得更多資訊。
- 欄位詳細資料。請參閱第 113 頁的『欄位詳細資料』主題，以取得更多資訊。
- 動作詳細資料。請參閱第 114 頁的『動作詳細資料』主題，以取得更多資訊。

檢視之間的鏈結

在主要視圖中，表格內加底線的文字會控制鏈結視圖中的顯示。按一下文字可讓您取得特定欄位、欄位集或處理步驟的詳細資料。您最後選取的鏈結會以較暗的顏色顯示，這可協助您識別兩個檢視窗格內容之間的關係。

重設檢視

若要重新顯示原始的「分析」建議並捨棄您對「分析」視圖所做的任何變更，請按一下主視圖畫面下方的「重設」。

## 欄位處理摘要

「欄位處理摘要」表格提供投射的整體處理影響Snapshot，包括功能狀態的變化和建構的功能數目。

請注意，實際上不會建立任何模式，因此資料準備之前和之後都沒有整體預測能力的變更測量值或圖形；相反地，您可以顯示個別的建議預測值的預測能力圖形。

表格會顯示下列資訊：

- 目標欄位數目。
- 原始 (輸入) 預測值的數目。
- 建議用於分析和模式建立的預測值。這包括建議的欄位總數；建議的原始、未轉換、欄位數目；建議的已轉換欄位數目 (不包括任何欄位的中間版本、從日期/時間預測值衍生的欄位，以及建構的預測值)；從日期/時間欄位衍生的建議欄位數目；以及建議的已建構預測值數目。
- 輸入預測值的數目不建議以任何格式使用，無論是以其原始格式 (衍生的欄位) 或以建構預測值的輸入格式。

在加底線的任一「欄位」資訊按一下，即可在鏈結的檢視中顯示更多詳細資料。「目標」、「輸入功能」和「未使用的輸入功能」會顯示於「欄位表格」鏈結檢視中。請參閱第 113 頁的『欄位表格』主題，以取得更多資訊。建議用於分析的功能會顯示在「預測能力」鏈結視圖中。請參閱第 113 頁的『預測能力』主題，以取得更多資訊。

## 欄位

「欄位」主要視圖顯示處理的欄位，以及 ADP 是否建議將它們用於下游模型中。您可以改寫任何欄位的建議；例如，排除建構的功能或包含 ADP 建議排除的功能。若欄位已經過轉換，您可以決定要接受建議的轉換或使用原始版本。

「欄位」視圖包含兩個表格，一個代表目標，一個代表已處理或建立的預測值。

目標表格

當資料中有定義目標時，才會顯示「目標」表格。

表格包含兩行：

- **名稱。** 這是目標欄位的名稱或標記；原始名稱永遠會顯示，即使欄位已經過轉換也一樣。
- **測量層次。** 這會顯示代表測量層級的圖示；將滑鼠移到圖示上方即可顯示描述資料的標記 (連續、次序、名義等)。

若目標經過轉換，則**測量層級**欄會反映最終的轉換版本。附註：您無法關閉目標的轉換功能。

### 預測值表格

永遠都會顯示**預測值**表格。表格的每一列代表一個欄位。依預設，列是以預測能力的遞減順序排序。

對於一般的功能，原始名稱永遠會作為名義稱。原始和衍生版本的日期/時間欄位會顯示於表格中 (以個別的列顯示)；表格也會包含建構的預測值。

請注意，表格中顯示的已轉換版本欄位永遠代表最終的版本。

依預設，只有建議的欄位會顯示在「預測值」表格。若要顯示其餘的欄位，請選取表格上方的「**在表格中包含非建議的欄位**」方框；接著就會在表格下方顯示這些欄位。

表格包含下列行：

- **要使用的版本。** 這會顯示下拉清單，此下拉清單控制欄位是否用於下游，以及是否使用建議的轉換。依預設，下拉清單會反映建議。

對於已轉換的一般預測值，下拉清單有三個選項：**轉換**、**原始**，以及**不使用**。

對於未轉換的一般預測值，選項為：**原始**和**不使用**。

對於衍生的日期/時間欄位和建構的預測值，選項為：**轉換**和**不使用**。

對於原始日期欄位，下拉清單是停用的，並且設為「**不使用**」。

附註：對於含有原始和轉換版本的預測值，變更**原始**和**轉換**版本會自動更新那些功能的**測量層級**和**預測能力**設定。

- **名稱。** 每個欄位名稱都是一個鏈結。在名稱上按一下可以在鏈結的檢視中顯示欄位的相關資訊。請參閱第 113 頁的『欄位詳細資料』主題，以取得更多資訊。
- **測量層次。** 這會顯示代表資料類型的圖示；將滑鼠移到圖示上方即可顯示描述資料的標記 (連續、次序、名義等)。
- **預測能力。** 只有 ADP 建議的欄位會顯示預測能力。若未定義任何目標，則不會顯示此行。預測能力範圍介於 0 到 1，較大的數值代表「較佳」的預測值。一般來說，預測能力對於在 ADP 分析內比較預測值非常實用，但不應在分析中比較預測能力值。

### 動作摘要

系統會針對自動資料準備所採取之各動作，轉換和/或過濾出輸入預測值；動作後留下來的欄位會用於下一個動作。之後，系統便會建議將留到最後一個步驟的欄位用於模式建立，並且過濾出轉換和建構預設值的輸入。

「動作摘要」是個簡單的表格，會列出 ADP 所採取的處理動作。按一下其中任何加底線的**動作**，便會在鏈結的檢視中顯示更多關於執行動作的詳細資料。請參閱第 114 頁的『動作詳細資料』主題，以取得更多資訊。

附註：只有原始和最終轉換版本的每個欄位會顯示，不會顯示分析期間使用的任何中間版本。

## 預測能力

在第一次執行分析，或是選取「欄位處理摘要」主要視圖的建議用於分析的預測值時，則會依預設顯示。此圖表顯示建議預設值的預測能力。欄位會依照預測能力排序，具有最高值的欄位會顯示於上方。

對於轉換版本的一般預測值，欄位名稱反映您在「設定」標籤的「欄位名稱」窗格選擇的字尾；例如：*\_transformed*。

測量層級圖示會顯示在個別的欄位名稱之後。

依據目標是連續或類別而定，系統會從線性迴歸或 naïve Bayes 模式中計算每個建議預測值的預測能力。

## 欄位表格

當您在「欄位處理摘要」主視圖中按一下「目標」、「預測值」或「未使用的預測值」時，就會顯示「欄位表格」視圖，其會顯示一個列出相關功能的簡單表格。

表格包含兩行：

- **名稱。** 預測值名稱。

對於目標會使用欄位原始名稱或標記，即使目標已經過轉換也一樣。

對於轉換版本的一般預測值，名稱反映您在「設定」標籤的「欄位名稱」畫面選擇的字尾；例如：*\_transformed*。

對於從日期與時間中衍生的欄位，會使用最終轉換版本的名稱；例如：*bdate\_years*。

對於建構的預測值，會使用建構預測值的名稱；例如：*Predictor1*。

- **測量層次。** 這會顯示代表資料類型的圖示。

對於目標，測量層級永遠反映轉換的版本（若目標已經過轉換）；例如，從次序（排序集合）變更為連續（範圍、尺度），反之亦然。

## 欄位詳細資料

當您在「欄位」主視圖中按一下任何「名稱」時，就會顯示「欄位詳細資料」。「欄位詳細資料」視圖包含所選欄位的分配、遺漏值或預測能力圖表（如果適用）。此外，也會顯示欄位的處理記錄和轉換欄位的名稱（如果適用）。

對於每個圖表集合，會以並排的方式顯示兩個版本，以比較套用和未套用轉換的欄位；若轉換版本的欄位不存在，則只會顯示原始版本的圖表。對於衍生的日期或時間欄位及建構的預測值，則只會顯示新預測值的圖表。

附註：若某個欄位因為有太多類別而被排除，便只會顯示處理記錄。

### 分配圖表

連續欄位分配會顯示為直方圖並重疊常態曲線，而垂直參考線代表平均值；類別欄位則顯示為長條圖。

直方圖標示為顯示標準差及偏斜度，然而，如果值的數目少於 2，或者原始欄位的變異數少於 10-20，則不會顯示偏斜度。

將滑鼠移到圖表上方，即可顯示直方圖的平均數，或是長條圖中類別記錄總數的個數及百分比。

### 遺漏值圖表

圓餅圖會比較套用轉換和未套用轉換的遺漏值百分比；圖表標記會顯示百分比。

若 ADP 執行了遺漏值處理，則轉換過後的圓餅圖也會包含置換值以作為標記，也就是說，會使用此值取代遺漏值。

將滑鼠移到圖表上方，會顯示遺漏值個數與記錄總數百分比。

#### 預測能力圖表

對於建議的欄位，長條圖會顯示轉換前後的預測能力。若目標已經過轉換，則計算的預測能力會和轉換後的目標有關。

附註：若未定義目標或是在主要視圖畫面中按一下目標，則不會顯示預測能力圖表。

將滑鼠移到圖表上方，會顯示預測能力值。

#### 處理記錄表格

表格會顯示轉換版本的欄位如何衍生。ADP 執行的動作會以它們執行的順序列出；不過，某些步驟的特定欄位可能會執行多個動作。

附註：未經過轉換的欄位不會顯示此表格。

表格中的資訊分為二或三行：

- **動作**。動作的名稱。例如「連續預測值」。請參閱『動作詳細資料』主題，以取得更多資訊。
- **詳細資料**。所執行處理的清單。例如，轉換為標準單位。
- **函數**。只有建構的預測值會顯示函數。函數會顯示輸入欄位的線性組合，例如  $.06*age + 1.21*height$ 。

#### 動作詳細資料

當您在「動作摘要」主要視圖中選取任何加底線的動作時，就會顯示「動作詳細資料」，「動作詳細資料」鏈結的視圖會顯示每個執行之處理步驟的動作特定資訊和一般資訊；系統會先顯示動作專屬的詳細資料。

針對各動作，會在鏈結檢視上方使用說明作為標題。動作專屬的詳細資料會顯示於標題下方，並且可能包含下列詳細資料：衍生預測值的數目、欄位重新分配、目標轉換、合併或重新排序的類別以及建構或排除之預測值。

當每個動作處理完後，處理過程中使用的預測值數目可能會變更，例如將預測值排除或合併時。

附註：若關閉某個動作或未指定任何目標，則在「動作摘要」主要視圖中按一下該動作時，便會在動作詳細資料處顯示錯誤訊息。

有 9 個可能的動作，但不一定每個分析都會用到。

#### 文字欄位表格

表格會顯示下列項目的數目：

- 刪除的尾隨空白值。
- 從分析中排除的預測值。

#### 日期與時間預測值表格

表格會顯示下列項目的數目：

- 從日期和時間衍生的持續期間預測值。
- 日期和時間元素。
- 衍生的日期和時間預測值總計。

若已計算任何日期持續期間，則參考日期或時間會顯示為註腳。

#### 預測值篩選表格

此表格會顯示下列從處理排除的預測值數目：

- 常數。
- 具有太多遺漏值的預測值。
- 單一類別中具有太多觀察值的預測值。
- 具有太多類別的名義欄位 (集合)。
- 篩選出的預測值總數。

#### 檢查測量層級表格

此表格會顯示重新分配的欄位數目，內容分為：

- 序數欄位 (排序集合) 重新分配為連續欄位。
- 連續欄位重新分配為序數欄位。
- 總數重新分配。

如果沒有連續或次序輸入欄位 (目標或預測值)，這就會顯示為註腳。

#### 偏離值表格

此表格會顯示已處理的偏離值個數。

- 根據您在「設定」標籤的「準備輸入與目標」窗格中的設定而定，可能是已發現並刪除其偏離值的連續欄位個數，或是已發現其偏離值並設為遺漏的連續欄位個數。
- 在偏離值處理之後，因為連續欄位的個數會是常數，因此將被排除。

有一個註腳會顯示偏離值分割值；如果沒有連續的輸入欄位 (目標或預測值)，則會顯示另一個註腳。

#### 遺漏值表格

此表格會顯示已置換遺漏值的欄位數目，內容分為：

- 目標。如果沒有指定目標則不會顯示此列。
- 預測值。這會進一步分為名義 (集合)、次序 (排序集合) 及連續的數目。
- 置換的遺漏值總個數。

#### 目標表格

這個表格會顯示目標是否已轉換，顯示為：

- 博克斯-考克斯 (Box-Cox) 轉換為常態。這又進一步分為顯示指定條件 (平均數和標準差) 的直欄和 Lambda ( $\lambda$ ) 值。
- 目標類別會重新排序以提升穩定性。

#### 類別預測值表格

此表格會顯示下列類別預測值的數目：

- 其類別經過重新排序 (最低至最高) 以提升穩定性。
- 其類別經過合併以最大化和目標之關聯的功能。
- 其類別經過合併以處理稀疏類別的功能。
- 因為和目標的關聯性低而排除的功能。
- 因為合併後是常數而排除的功能。

若沒有類別預測值，則會顯示註腳。

#### 連續預測值表格

有兩個表格。第一個顯示下列其中一項轉換的數目：

- 預測值轉換為標準單位。此外，這也會顯示轉換的預測值數目、指定的平均數以及標準差。
- 對應到一般範圍的預測值。此外，這也會顯示使用最小/最大值轉換來轉換的預測值數，以及指定的最小值與最大值。
- 經過 bin 處理的預測值與經過 bin 處理的預測值數。

第二個表格會顯示預測值空間建構詳細資料，並顯示為下列預測值的數目：

- 建構的功能。
- 因為和目標的關聯性低而排除的功能。
- 因為 bin 處理後是常數而排除的功能。
- 因為建構後是常數而排除的功能。

若沒有連續預測值為輸入，則會顯示註腳。

#### 產生一個衍生節點

當您產生衍生節點時，它會將目標逆模型變換套用到分數欄位。依預設，節點輸入由自動建模節點（如「自動分類器」或「自動數值」）或「總體」節點生成的分數欄位名稱。如果尺度（範圍）目標已轉換，則分數欄位以轉換後的單位進行顯示，例如，以  $\log(\$)$  代替  $\$$ 。為解讀和使用結果，您必須將預測值轉換回原始尺度。

注意：只有當 ADP 節點包含對範圍目標重新尺度（即，在「準備輸入和目標」窗格上選取了 Box-Cox 重定比）的分析時，才能產生「衍生」節點。若目標不是範圍或未選取 Box-Cox 重新調整，您就無法產生「衍生」節點。

在「多個」模式下建立衍生節點，並在表示式中使用 @FIELD，以便可以在需要時新增轉換後的目標。例如，使用下列詳細資料：

- 目標欄位名稱：response
- 轉換後的目標欄位名稱：response\_transformed
- 分數欄位名稱：\$XR-response\_transformed

「衍生」節點將建立一個新欄位：\$XR-response\_transformed\_inverse。

注意：如果未使用自動建模節點或總體節點，那麼您將需要編輯「衍生」節點以便為模型轉換正確的分數欄位。

#### 正規化連續目標



依預設，如果您在「準備輸入和目標」窗格上選中了使用 **Box-Cox** 變換重新尺度連續目標勾選框，那麼這將變換目標，並且您所建立的新欄位將成為模型建置的目標。例如，如果原始目標為 *response*，則新目標將為 *response\_transformed*。ADP 節點的模型下游將自動選取該新目標。

但這可能會引發問題，具體取決於原始目標。例如，如果目標為 *Age*，則新目標的值將不是 *Years*，而是 *Years* 的轉換版本。這意味著，由於它們的單位不可識別，因此您無法看到分數和解釋。這時，您可以套用逆模型變換，將變換後的單位轉回到它們原來的含義。若要這麼做，請執行下列動作：

1. 按一下分析資料以執行 ADP 分析，然後從產生功能表中選取衍生節點。
2. 在模型畫布上，將衍生節點放置在模型塊後面。

衍生節點會將分數欄位還原為原始維度，這樣預測值的單位將為 *Years*。

依預設，「衍生」節點會轉換由自動建模節點或總體模型產生的分數欄位。如果建立個別的模型，那麼需要編輯「衍生」節點，以便從實際分數欄位中衍生。如果要對您的模型進行評估，則應當將轉換後的目標新增到衍生節點的衍生自欄位中。這也會將相同的逆模型變換套用到目標，並且任何下游評估或分析節點都能正確地套用已轉換資料，只要您將其切換到使用欄位名，而不是 *meta* 資料。

如果還想還原原始名稱，那麼可以使用「過濾器」節點移除原始目標欄位（如果仍然存在），然後重新命名目標和分數欄位。

---

## 類型節點

欄位內容可在來源節點中指定也可在個別的「類型」節點中指定。兩種節點的功能相似。可用的內容如下：

- **欄位** 按兩下任何欄位名稱均可指定 IBM SPSS Modeler 中的資料的值標籤和欄位標籤。例如，從 IBM SPSS Statistics 匯入的欄位 *meta* 資料可在此處檢視或修改。與之相似，您也可以為欄位及其值建立新的標籤。您在此處指定的標籤將根據您在「串流內容」對話框中的選項顯示在整個 IBM SPSS Modeler 中。
- **測量** 這是測量層次，用於說明給定欄位中的資料的特性。如果已經瞭解某個欄位的所有詳細資料，則稱為 **已完全實例化**。如需相關資訊，請參閱第 118 頁的『測量層次』。

註：欄位的測量層次與欄位的儲存類型不同，後者指示資料是作為字串、整數、實數、日期、時間、時間戳記還是清單進行儲存。

- **值** 此欄使您能夠指定用於從資料集中讀取資料值的選項，或者使用指定選項在個別的對話框中指定測量層次和值。您還可以選擇傳遞欄位，而不讀取它們的值。如需相關資訊，請參閱第 122 頁的『資料值』。

註：如果對應的欄位項目包含清單，則無法在此直欄中修正資料格。

- **遺漏** 用於指定如何處理此欄位的遺漏值。如需相關資訊，請參閱第 126 頁的『定義遺漏值』。

註：如果對應的欄位項目包含清單，則無法在此直欄中修正資料格。

- **檢查** 在此欄中，您可以設定選項，以確保欄位值符合指定的值或範圍。如需相關資訊，請參閱第 126 頁的『勾選類型值』。

註：如果對應的欄位項目包含清單，則無法在此直欄中修正資料格。

- **角色** 用於向建模節點指示欄位將成為用於機器學習過程的輸入（預測值欄位）還是目標（預測的欄位）。兩者、無以及分割區也是可用角色，最後一個可用角色表明欄位用於將記錄分割區到不同的樣本中，以用於進行訓練、測試和驗證。值 **分割** 指定將為欄位的每個可能值建立個別的模型。如需相關資訊，請參閱第 127 頁的『設定欄位角色』。

可以使用「類型」節點視窗指定數個其他選項：

- 使用工具功能表按鈕，您可以在實例化「類型」節點（透過規格、讀取值或執行串流）後，選擇**忽略唯一的欄位**。忽略唯一的欄位將自動忽略僅具有一個值的欄位。
- 使用工具功能表按鈕，您可以選擇在實例化「類型」節點之後**忽略大型集**。忽略大型集將自動忽略具有大量成員的集。
- 使用工具功能表按鈕，您可以選擇在實例化「類型」節點之後將**連續整數轉換為序數**。請參閱第 121 頁的『轉換連續資料』主題，以取得更多資訊。
- 使用工具功能表按鈕，您可以產生「過濾器」節點，以捨棄所選取的欄位。
- 使用太陽鏡切換按鈕，您可以將所有欄位的預設值設為「讀取」或「傳遞」。依預設，來源節點中的「類型」標籤傳遞欄位，而「類型」節點依預設會自行讀取值。
- 使用清除值按鈕，您可以清除在此節點中對欄位值進行的變更（非繼承值），並從上游作業重新讀取值。重設可能已對特定欄位上游進行的變更時，此選項有用。
- 使用清除所有值按鈕，您可以將所有欄位讀取的值重設為節點。此選項將所有欄位的值直欄有效地設為讀取。重設所有欄位的值，並從上游作業重新讀取值及類型時，此選項有用。
- 使用快速功能表，您可選擇從一個欄位至另一個複製屬性。請參閱第 127 頁的『複製類型屬性』主題，以取得更多資訊。
- 使用檢視未用的欄位設定選項時，您可以檢視資料中不再呈現或已連接至此「類型」節點之欄位的類型設定。為已變更的資料集重複使用「類型」節點時，這有用。

## 測量層次

測量層次（早期稱為「資料類型」或「用量類型」）說明 IBM SPSS Modeler 中資料欄位的使用情形。可以在來源或「類型」節點的「類型」標籤上指定測量層次。例如，您可能想要使用值 1 及 0，將整數欄位的測量層次設為旗標。這通常指出 1 = *True* 及 0 = *False*。

**儲存體與測量。**請注意，欄位的測量層次與其儲存體類型不同，這指出資料是儲存為字串、整數、實數、日期、時間還是時間戳記。由於可以使用「類型」節點，在串流中的任意點修改資料類型，必須在將資料讀取至 IBM SPSS Modeler 時，在來源判定儲存體（雖然可以後續使用轉換功能予以變更）。請參閱第 7 頁的『設定欄位儲存類型和格式』主題，以取得更多資訊。

部分建模節點透過其「欄位」標籤上的圖示，指出其輸入及目標欄位允許的測量層次類型。

### 測量層次圖示

表 20. 測量層次圖示










圖示	測量層級
	預設值
	連續的
	種類
	旗標

表 20. 測量層次圖示 (繼續)

圖示	測量層級
	名義(M)
	序數(O)
	無類型
	集合
	地理空間

可以使用下列測量層次：

- **預設值** 儲存類型和值未知（例如，由於尚未讀取）的資料將顯示為 <預設值>。
- **連續** 用於說明數值，例如範圍 0-100 或 0.75-1.25。連續值可以是整數、實數或日期/時間。
- **種類** 在不同值的準確號碼未知時用於字串值。這是一種**非實例化**資料類型，表示有關資料儲存類型和使用的所有可用資訊均未知。讀取資料後，測量層次將為旗標、列名、或無類型，具體取決於「串流內容」對話框中指定的列名欄位會員最大數目。
- **旗標** 用於具有兩個不同值的資料，這兩個值用於指示特性存在與否（例如 true 與 false、Yes 與 No 或者 0 與 1）。所用值可能有所不同，但其中總會有個值代表 "true" 值，另一個代表 "false" 值。資料可代表為文字、整數、實數、日期、時間或時間戳記。
- **列名** 用於說明具有多個不同值的資料，其中的每個值都被視為集合的一個會員，例如 small/medium/large。列名資料可具有任何儲存數值型、字串或日期/時間。請注意，將測量層次設定為列名不會自動將值變更為字串儲存。
- **序數** 用於說明具有多個順序固定的不同值的資料。例如，工資種類或滿意度分等級可以歸類為序數資料。順序由資料元素的自然排序定義。例如，1, 3, 5 是某個整數集合的預設排序，而 HIGH, LOW, NORMAL（按英文字母遞增）是某個字串集合的順序。使用序數測量層次可以將一組種類資料定義為序數資料，以進行視覺化處理、模型建置以及匯出到將序數資料識別為不同類型的其他應用程式（如 IBM SPSS Statistics）。您可以在任何能夠使用列名欄位的位置使用序數欄位。此外，可以將任何儲存類型（實數、整數、字串、日期、時間等等）的欄位定義為序數。
- **無類型** 用於不的上述任何類型的資料、具有單一值的欄位或者集合會員數超過定義的數目上限的列名資料。當測量層次為包含多數會員（如帳號）的集合時，這種類型也將十分有用。When you select 欄位的**無類型**，角色將自動設定為無，並且記錄 ID將作為唯一的替代項。集合的上限預設容量為 250 個唯一值。可在「串流內容」對話框（通過「工具」功能表存取）的「選項」標籤中調整或停用該數字。
- **收集** 用於識別清單中記錄的非地理空間資料。收集實際上是深度為零的清單欄位，該清單中的元素具有另外某種測量層次。

有關清單的更多資訊，請參閱《SPSS Modeler 的「來源」節點、「過程」節點和「輸出」節點》手冊的『「來源」節點』部分中的『清單儲存以及相關聯的測量層次』主題。

- **地理空間** 與「清單」儲存類型配合使用以識別地理空間資料。清單可以是清單深度介於 0 與 2（含首尾值）之間的「整數清單」或「實數清單」欄位。

有關進一步資訊，請參閱《SPSS Modeler 的「來源」節點、「過程」節點和「輸出」節點》手冊的『「類型」節點』部分中的『地理空間測量子層次』主題。

您可以手動指定測量層次，或者您可以容許軟體讀取資料，並根據它讀取的值判定測量層次。

或者，當您有數個應該被視為種類資料的連續資料欄位時，可以選擇一個選項予以轉換。請參閱第 121 頁的『轉換連續資料』主題，以取得更多資訊。

## 若要使用自動鍵入

1. 在來源節點的「類型」標籤或「類型」節點中，將想要的欄位的值直欄設為 **<Read>**。這將讓 meta 資料可用於所有節點下游。您可以使用對話框上的太陽鏡按鈕，快速地將所有欄位都設為 **<Read>** 或 **<Pass>**。
2. 按一下讀取值，以立即從資料來源讀取值。

## 若要手動設定欄位的測量層次

1. 選取表格中的某個欄位。
2. 從測量直欄的下拉清單中，選取該欄位的測量層次。
3. 或者，您可以在使用下拉清單選取測量層次之前，使用 Ctrl-A 或 Ctrl-按一下，以選取多個欄位。

## 地理空間測量子層次

與「清單」儲存類型配合使用的「地理空間」測量層次具有 6 個子層次，這些子層次用於識別不同類型的地理空間資料。

- **點** - 指示特定位置；例如城市中心。
- **多邊形** - 點數列，用於識別區域及其位置（例如城鎮）的單個界限。
- **線串** - 也稱為「線條」或簡稱「線條」，「線串」是點的數列，用於識別線條路徑。例如，線串可能是固定項目，例如道路、河流或鐵路；或者是移動的物件的軌跡，例如飛機的飛行路徑或輪船的航行。
- **多點** - 在資料中的每一列都包含每個區域的多個點時使用。例如，如果每一列都代表城市街道，那麼可以使用每條街道的多個點來識別每個街燈。
- **多多邊形** - 在資料中的每一列都包含多個多邊形時使用。例如，如果每一列都代表國家的輪廓，那麼可以將美國記錄為多個多邊形以識別不同的地區域圖，例如本土、阿拉斯加和夏威夷。
- **多線串** - 在資料中的每一列都包含多條線條時使用。由於線條無法分支，因此您可以使用「多線串」來識別一群組線條。例如，每個國家的通航水道或鐵路網路之類的資料。

這些測量子層次與「清單」儲存類型配合使用。如需相關資訊，請參閱第 9 頁的『清單儲存以及相關聯的測量層次』。

## 限制

在使用地理空間資料時，您必須瞭解一些限制。

- 座標系統可能會影響資料格式。例如，投影座標系使用座標值  $x$  和  $y$  並在需要時使用  $z$ ，而地理座標系使用座標值經度和緯度並在需要時使用高度值或深度值。

如需座標系統的相關資訊，請參閱《SPSS Modeler 使用手冊》之「使用串流」小節中的「設定串流的地理空間選項」主題。

- 線串不得與自身交叉。
- 多邊形不會自行閉合；對於每個多邊形，您必須確保將第一個點與最後一個點定義為相同。

- 多多邊形中的資料方向至關重要；順時針方向表示實心形狀，而逆時針方向表示空心形狀。例如，如果您記錄國家中存在湖泊的地區域圖，那麼可以按順時針方向記錄大陸地區域圖的邊界，並按逆時針方向記錄每個湖泊的形狀。
- 多邊形不得與自身相交。這種相交的一個範例是，您嘗試將多邊形界限繪製成圖形 8 所示的連續線條形狀。
- 多多邊形不得相互重疊。
- 若為「地理空間」欄位，相關的儲存體類型只有實數和整數（預設值為實數）。

## 地理空間測量子層次圖示

表 21. 地理空間測量子層次圖示

圖示	測量層級
	點
	多邊形
	線串
	多點
	多個多邊形
	MultiLineString

## 轉換連續資料

將種類資料視為連續可能對模型的品質產生嚴重效果，特別是當它為目標欄位時；例如，產生回歸模型，而非二進位模型。若要防止此情況，您可以將整數範圍轉換為種類類型，例如序數或旗標。

1. 從「作業」及「產生」功能表按鈕（具有工具符號），選取將連續整合轉換為序數。即會顯示轉換值對話框。
2. 指定將自動轉換的範圍大小；這適用於不超出您輸入的大小的任何範圍。
3. 按一下「確定」。受影響的範圍會轉換為旗標或序數，並顯示在「類型」節點的「類型」標籤上。

### 轉換的結果

- 具有整數儲存體的連續會變更為序數，系統會擴充下限值與上限值，以包括從下限值到上限值的所有整數值。例如，如果範圍是 1、5，則值集為 1、2、3、4、5。
- 如果連續欄位變更為旗標，則旗標欄位的下限值及上限值會變成 false 及 true 值。

## 何謂實例化？

實例化是讀取或指定資訊的處理程序，例如資料欄位的儲存體類型及值。為了最佳化系統資源，實例化是使用者引導的處理程序，您透過在來源節點的「類型」標籤上指定選項，或者透過「類型」節點執行資料，從而告知軟體讀取值。

- 具有不明類型的資料也被稱為**未實例化**。儲存體類型及值不明的資料顯示在「類型」標籤的測量直欄中，作為 **<Default>**。
- 當您具有欄位儲存體的部分相關資訊時，例如字串或數值，資料被稱為**部分實例化**。**種類或連續**是部分實例化的測量層次。例如，**種類**指定欄位是符號，但是您不知道它是名義、序數還是旗標。
- 當有關某個類型的所有詳細資料都已知時（包括值），會在此直欄中顯示**完全實例化**測量層次，即名義、序數、旗標或連續。附註：連續類型用於部分實例化及完全實例化的資料欄位。連續資料可以是整數或實數。

在執行具有「類型」節點的資料串流期間，根據起始資料值，未實例化的類型會立即變成部分實例化。所有資料都通過節點之後，所有資料都會變成完整實例化，除非值設為 **<Pass>**。如果岔斷執行，則資料將保留部分實例化。已實例化「類型」標籤之後，在串流的該點，欄位的值為靜態。這表示任何上游變更都不會影響特定欄位的值，即使您重新執行串流。若要根據新的資料或新增的操作變更或更新值，您必須在「類型」標籤自身中編輯它們，或者將欄位的值設為 **<Read>** 或 **<Read +>**。

### 何時實例化

通常，如果資料集不是非常大，並且您計劃稍後在串流中新增欄位，於來源節點的實例化是最方便的方法。然而，下列情況下，在單獨的「類型」節點中實例化有用：

- 資料集較大，且串流過濾「類型」節點之前的子集。
- 已在串流中過濾資料。
- 已在串流中合併或附加資料。
- 處理期間衍生了新的資料欄位。

## 資料值

您可以使用「類型」標籤的**值**直欄，從資料自動讀取值，或者您可以在單獨的對話框中指定測量層次及值。

「值」下拉清單中提供的選項會提供自動鍵入指示，如下表中所示。

表 22. 自動鍵入的指示

選項	函數
<Read>	執行節點時讀取資料。
<Read+>	讀取節點，並附加至現行資料（如果存在的話）。
<Pass>	不讀取任何資料。
<Current>	保留現行資料值。
指定...	為您開啟的單獨的對話框，以指定值及測量層次選項。

執行「類型」節點或按一下**讀取值**，會根據您的選擇，從資料來源自動鍵入及讀取值。也可以使用「指定」選項，或者在欄位直欄的資料格中按兩下，手動指定這些值。

對「類型」節點中的欄位進行變更之後，您可以使用對話框工具列上的下列按鈕，重設值資訊：

- 使用**清除值**按鈕，您可以清除在此節點中對欄位值進行的變更（非繼承值），並從上游作業重新讀取值。重設可能已對特定欄位上游進行的變更時，此選項有用。

- 使用**清除所有值**按鈕，您可以將**所有**欄位讀取的值重設為節點。此選項將所有欄位的值直欄有效地設為讀取。重設所有欄位的值，並從上游作業重新讀取值及測量層次時，此選項有用。

## 將值直欄中的文字變成灰色

在「類型」節點或「來源」節點內，如果值直欄中的資料以黑色文字顯示，它指出已讀取該欄位的值，並儲存在該節點內。如果此欄位中未呈現任何黑色文字，則尚未讀取該欄位的值，並進一步在上游判定。

有時，您可能看到資料為灰色文字。當 SPSS Modeler 可以識別或推斷欄位的有效值，而不實際讀取及儲存資料時，可能發生此情況。如果您使用以下其中一個節點，則很可能發生此情況：

- 使用者輸入節點。由於在節點內定義資料，因此一律已知欄位的值範圍，即使值尚未儲存在節點中。
- 統計資料檔案來源節點。如果呈現資料類型的 meta 資料，則會啟用 SPSS Modeler 以推斷值的可能範圍，而不讀取或儲存資料。

在任一節點中，值以灰色文字顯示，直到您按一下**讀取值**。

註：如果您未實例化串流中的資料，並且資料值以灰色顯示，則不適用勾選您在**勾選直欄**中設定的類型值。

## 使用值對話框

在「類型」標籤中按一下**值或遺漏欄顯示預定值**的下拉清單。在此清單中選取**指定...** 選項將開啟一個個別的對話框，在該對話框中，可以為選取的欄位設定用於讀取、指定、標註和處理值的選項。

很多控制項是所有資料類型通用的。下面介紹這些通用控制項。

**測量** 顯示目前選取的測量層次。您可以變更設定以反應希望使用資料的方式。例如，如果名為 `day_of_week` 的欄位包含代表各天的數字，您可能希望將此變更為列名資料，以建立用於分別檢查每個種類的分佈節點。

**儲存** 顯示儲存類型（如果已知）。儲存類型不受您選擇的測量層次的影響。要改變儲存類型，可以使用「固定檔案和可變檔案」來源節點中的「資料」標籤或使用「過濾器」節點中的轉換函數。

**模型欄位** 對於對模型塊進行評分時產生的欄位，還可以檢視模型欄位的詳細資料。這些詳細資訊包含目標欄位的名稱以及建模時此欄位的角色（預測值、機率和傾向等等）。

**值** 選取用於確定選取的欄位的值的方法。您在此做出的選擇將置換之前在類型節點對話框的值欄中進行的任何選擇。用於讀數值的選項包含：

- **從資料中讀取** 選取此項表示在執行節點時讀取值。此選項與 **<讀取>** 相同。
- **經過** 選取此項表示不讀取目前欄位的資料。此選項與 **<傳送>** 相同。
- **指定值及標籤** 這裡的選項用於指定選取的欄位的值與標籤。將此選項與值檢查功能配合使用，可以根據您對目前欄位的瞭解指定值。此選項可針對不同欄位類型啟動該類型所特有的控制項。後續主題將分別介紹用於值與標籤的選項。

註：對於測量層次為「無類型」或「<預設>」的欄位，無法指定值或標籤。

- **使用資料延伸值** 選取此項可以對現行資料附加此處輸入的值。例如，如果 `field_1` 的範圍為 (0,10)，您輸入的值範圍為 (8,16)，則會通過新增 16 來延伸範圍，而不刪除原始最小值。新的範圍將是 (0,16)。選擇此選項會自動將自動歸類選項設定為 **<讀取+>**。

**最大清單長度** 僅適用於測量層次為「地理空間」或「收集」的資料。透過指定清單可以包含的元素數目來設定清單的長度上限。

**最大字串長度** 僅適用於無類型資料；在產生 SQL 以建立表格時使用該欄位。輸入資料中最大字串的值；這樣會在表格中產生一個足夠容納該字串的欄。如果字串長度值不可用，將使用可能不適用於該資料的預設字串大小（例如，如果值太小，向表格中寫入資料時可能會發生錯誤；如果值太大，可能會對效能產生不利影響。）

**檢查值** 請選取強制轉換值以使其符合指定的連續、旗標或列名值的方法。此選項與類型節點對話框中的檢查欄對應，在此進行的設定將置換該對話框中的設定。通過將值檢查功能與**指定值及標籤**選項配合使用，可以使資料中的值與期望值的值一致。例如，如果指定值為 1、0，然後使用**捨棄**選項，則可以捨棄所有值不是 1 或 0 的記錄。

**定義空白值** 選取此項可以啟動下列控制項，這些控制項可用於宣告資料中的遺漏值或空白值。

- **遺漏值** 使用此表格可以將特定的值（例如 99 或 0）定義為空白值。該值應適用於欄位的儲存類型。
- **範圍** 用於指定遺漏值的範圍，例如，年齡 1-17 或大於 65。如果將某個界限值保留為空白，那麼範圍將不受限；例如，如果僅指定下限為 100 而未指定上限，那麼會將所有大於或等於 100 的值定義為遺漏。界限值包含在內；例如，下限為 5 且上限為 10 的範圍定義將包含 5 和 10。可以為任何儲存類型定義遺漏值範圍，這些類型包含日期/時間和字串（在這種情況下，將採用英文字母排序來確定某個值是否在範圍內）。
- **空值/空格** 您還可以將系統空值（在資料中顯示為 \$null\$）和空白（不帶可見字元的字串值）指定為空白值。

註：為了執行分析，「類型」節點還會將空字串視為空白，儘管它們在內部以不同方式進行儲存，並且在某些情況下以不同方式進行處理。

註：要將空白值編碼為未定義值 \$null\$，請使用「填入器」節點。

**說明** 使用此文字框可以指定欄位標籤。這些標籤將根據您在「串流內容」對話框中選擇的選項出現在各種位置，例如出現在圖形、表格、輸出和模型瀏覽器中。

## 指定連續資料的適用於下列項目的值與標籤

連續測量層次僅用於數值型欄位。連續資料的儲存類型有以下三種：

- 實數
- 整數
- 日期/時間

所有連續欄位都將通過同一個對話框進行編輯，顯示的儲存類型僅供參照。

### 指定值

下列控制項是連續欄位所獨有的，用於指定值的範圍：

**低值。** 指定值範圍的下限。

**高值。** 指定值範圍的上限。

### 指定標籤

您可以為範圍欄位的任何值指定標籤。按一下**標籤**按鈕，可開啟個別對話框來指定值標籤。

**值與標籤子對話框：** 在範圍欄位的「值」對話框中按一下**標籤**將開啟一個新的對話框，您可在其中指定該範圍內任意值的標籤。



可以使用此表格中的 值和 標籤欄定義值和標籤成對。目前已定義的成對將在此顯示。通過按一下空白 Cell 並輸入值及其對應標籤，可以新增新的標籤成對。注意：向此表格新增值/值-標籤成對不會向欄位新增任何新值。該操作只是建立欄位值的 meta 資料而已。

您在「類型」節點中指定的標將按按照您在「串流內容」對話框中選擇的選項顯示在多個位置（顯示為工具提示、輸出標籤等）。

## 指定列名和序數資料的適用於下列項目的值與標籤

列名（集合）和依序（依序集合）測量層次表明資料值將分別用作集合的一個會員。集合的儲存類型可以是字串、整數、實數或日期/時間。

下列控制項是列名和序數欄位獨有的，用於指定值及標籤：

**值。** 使用表格中的值欄，可以根據您對目前欄位的瞭解指定值。使用此表格，可以輸入欄位的預期值，並使用「檢查值」下拉清單檢查資料集與這些值的一致性。使用箭頭及刪除按鈕，您可以修改現有值以及重新排序或刪除值。

**標籤。** 使用標籤欄可以為集合中的每個值指定標籤。這些標籤在圖形、表格、輸出及模型瀏覽器等各種位置中出現，視您在串流內容對話框中所做的選擇而定。

## 指定旗標的值

旗標欄位用於顯示具有兩個不同值的資料。旗標的儲存類型可以是字串、整數、實數或日期/時間。

**True。** 指定條件成立時欄位的旗標值。

**False。** 指定條件不成立時欄位的旗標值。

**標籤。** 為旗標欄位中的每個值指定標籤。這些標籤在圖形、表格、輸出及模型瀏覽器等各種位置中出現，視您在串流內容對話框中所做的選擇而定。

## 指定收集資料的值

收集欄位用於顯示清單中的非地理空間資料。

您可以為集合測量層次設定的唯一項目是清單測量。依預設，此測量設定為「無類型」，但您可以選取另一個值，以設定清單中的元素的測量層次。您可以選取下列其中一個選項：

- 無類型
- 連續的
- 名義(M)
- 序數(O)
- 旗標

## 指定地理空間資料的值

地理空間欄位用於顯示清單中的地理空間資料。

對於地理空間測量層次，您可以設定下列選項，以設定該清單中的元素的測量層次：

**類型** 選取地理空間欄位的測量子層次。可用的子層次由清單欄位的深度確定；預設值為：「點」（深度為零）、「線串」（深度為 1）和「多邊形」（深度為 1）。

如需子層次的相關資訊，請參閱第 120 頁的『地理空間測量子層次』。

如需清單深度的相關資訊，請參閱第 9 頁的『清單儲存以及相關聯的測量層次』。

**座標系統** 只有當您將測量層次從地理空間層次變更成非地理空間層次時，此選項才可用。若要將座標系統套用到地理空間資料，請選取此勾選框。依預設，會顯示在工具 > 串流內容 > 選項 > 地理空間窗格中設定的座標系統。要使用另一個座標系統，請按一下變更按鈕以顯示「選擇座標系統」對話框，並選擇所需的座標系統。

如需座標系統的相關資訊，請參閱《SPSS Modeler 使用手冊》之「使用串流」小節中的「設定串流的地理空間選項」主題。

## 定義遺漏值

「類型」標籤的遺漏欄指示是否已為欄位定義遺漏值處理。可能的設定為：

**On (\*)**。指示已為此欄位定義遺漏值處理。可使用下方遊填入節點，或使用「指定」選項（見下方文）通過明確規格來進行此操作。

**Off**。欄位沒有定義遺漏值處理。

**指定**。選擇此選項將顯示對話框，您可在其中宣告將明確值視為此欄位的遺漏值。

## 勾選類型值

針對每一個欄位開啟「勾選」選項，可檢查該欄位中的所有值，以判定它們是否符合您在「指定值」對話框中指定的現行類型設定或值。清除資料集及減小單一作業內的資料集大小時，這有用。

「類型節點」對話框中勾選直欄的設定會判定探索到類型限制外部的值時會發生什麼。若要變更欄位的「勾選」設定，請使用勾選直欄中該欄位的下拉清單。若要設定所有欄位的「勾選」設定，請在欄位直欄中按一下，然後按下 Ctrl-A。然後將下拉清單用於勾選直欄中的任何欄位。

下列「勾選」設定可用：

**無**。無需勾選即通過值。這是預設值。

**設為空值**。將限制外部的值變更為系統空值 (\$null\$)。

**強制轉型**。將針對在指定範圍外部的值，勾選測量層次為完整實例化的欄位。將使用下列規則，將未指定的值轉換為該測量層次的合法值：

- 對於旗標，true 或 false 值之外的任何其他值都會轉換為 false 值。
- 對於集（名義或序數），任何不明值都會轉換為集的第一個成員的值。
- 大於範圍上限的數字會取代為上限。
- 小於範圍下限的數字會取代為下限。
- 範圍中的空值會取得該範圍的中間點值。

**捨棄**。找到不正確的值時，會捨棄整個記錄。

**警告**。已讀取所有資料後，會在串流內容對話框中計數及報告不合法項目的數目。

**中斷**。發現的第一個不正確的值會終止串流的執行。系統會在串流內容對話框中報告錯誤。

## 設定欄位角色

欄位的角色指定它如何在模型建置中使用，例如，欄位是輸入還是目標（正在預測的項目）。

註：「分割區」、「頻率」及「記錄 ID」角色每一個都僅能套用至單一欄位。

下列角色可用：

**輸入。**該欄位將用作機器學習的輸入（一個預測工具欄位）。

**目標。**該欄位將用作機器學習的輸出或目標（模型將嘗試預測的其中一個欄位）。

**兩者。**Apriori 節點會將該欄位用作輸入與輸出。所有其他建模節點將忽略該欄位。

**無。**機器學習將忽略該欄位。在**角色直欄**中，測量層次已設為**無類型**的欄位會自動設為**無**。

**分割區。**指出欄位用於將資料分割為單獨的範例，以進行訓練、測試及（選用）驗證。該欄位必須是已實例化的集類型，並具有兩個或三個可能的值（如「欄位值」對話框中所定義）。第一個值代表訓練範例，第二個值代表測試範例，第三個（如果呈現的話）代表驗證範例。系統會忽略任何其他值，且無法使用旗標欄位。請注意，要在分析中使用分割區，必須在適當模型建置或分析節點中的「模型選項」標籤上啟用分割。啟用分割時，分析會排除分割區欄位為空值的記錄。如果已在串流中定義多個分割欄位，則必須在每一個適用建模節點的「欄位」標籤上指定單一分割欄位。如果您的資料中不存在適用的欄位，您可以使用「分割區」節點或「裝置」節點建立一個。請參閱第 150 頁的『分割區節點』主題，以取得更多資訊。

**分割。**（僅限名義、序數及旗標欄位）指定要為欄位的每一個可能值建置一個模型。

**次數。**（僅限數值欄位）設定此角色可讓欄位值用作記錄的頻率加權係數。只有「C&R 樹狀結構」、CHAID、QUEST 及「線性」模型支援此特性；所有其他節點都會忽略此角色。透過支援該特性的那些建模節點「欄位」標籤上的**使用頻率加權**選項，啟用頻率加權。

**記錄 ID。**該欄位將用作唯一的記錄 ID。大部分節點會忽略此特性；然而，「線性」模型支援該特性，且 IBM Netezza 資料庫內採礦節點需要該特性。

## 複製類型屬性

可以輕鬆地將某種類型的屬性（如值、檢查選項和遺漏值）從一個欄位複製到另一個欄位：

1. 用滑鼠右鍵按一下要複製其屬性的欄位。
2. 在快速功能表中選擇**複製**。
3. 用滑鼠右鍵按一下要變更其屬性的欄位。
4. 在快速功能表中選擇**選擇性貼上**。注意：可以採用按住 Ctrl 並按一下的方法或使用快速功能表中的**選取欄位**選項選取多個欄位。

此時將開啟一個新的對話框，您可在其中選取要貼上的特定屬性。如果要粘貼至多個欄位，在此選取的選項將套用至所有目標欄位。

**貼上以下屬性。**在下方面的清單中進行選取，將屬性從一個欄位貼上至另一個欄位。

- **類型。**選取此選項可貼上測量層次。
- **值。**選取此選項可貼上欄位值。
- **遺漏值。**選取此選項可貼上遺漏值設定。
- **檢查。**選取此選項可貼上值檢查選項。
- **角色。**選取此選項可貼上欄位的角色。

## 欄位格式設定標籤

「表格」節點和「類型」節點的「格式」標籤將顯示目前欄位或未用欄位的清單，以及每個欄位的格式設置選項。下面是欄位格式設置表格中每個欄的說明：

**欄位。** 此列顯示選取的欄位的名稱。

**格式。** 通過按兩下此欄中的 Cell，可以使用開啟的對話框指定各個欄位的格式設置。請參閱『設定欄位格式選項』主題，以取得更多資訊。在此指定的格式設置將置換整體串流內容中指定的格式設置。

注意：Statistics 匯出節點和 Statistics 輸出節點匯出的 .sav 檔案的 meta 資料中包含根據欄位選擇格式設置。如果指定了 IBM SPSS Statistics .sav 檔案格式所不支援的根據欄位格式，那麼節點將採用 IBM SPSS Statistics 預設格式。

**排列整齊。** 使用此欄可指定表格欄中的值的排列整齊方式。預設值為自動，它會向左調整符號值以及向右調整數值。您可通過選取左、右或中心來置換預設設置。

**欄寬。** 依預設，根據欄位值自動計算欄寬。要置換自動寬度計算，請按一下表格資料格，然後使用下拉清單選取新的寬度。要輸入此處未列出的自訂寬度，請通過按兩下「欄位」或「格式」欄中的表格資料格開啟「欄位格式」子對話框。或者，也可以用滑鼠右鍵按一下某個 Cell，然後選擇 **設定格式**。

**檢視目前欄位。** 依預設，對話框將顯示目前處於作用中狀態的欄位的清單。要檢視 UNUSED 欄位的清單，請選取檢視未使用的欄位設定。

**快速功能表。** 此標籤的快速功能表提供了多種選擇和設定更新選項。在欄中用滑鼠右鍵按一下可顯示此功能表。

- **全部選中。** 選中所有欄位。
- **全部全部不選。** 清除選擇內容。
- **選取欄位。** 依據類型或儲存特性選取欄位。選項包括選取種類、選取連續（數值型）、選取無類型、選取字串、選取數值型或選取日期/時間。請參閱第 118 頁的『測量層次』主題，以取得更多資訊。
- **設定格式。** 開啟用於指定每個欄位的日期、時間和小數選項的子對話框。
- **設定排列整齊方式。** 設定選取的欄位的對齊方式。選項包括自動、中心、左或右。
- **設定直欄寬度。** 設定選取的欄位的欄位寬度。指定自動將從資料中讀取寬度。您也可以將欄位寬度設為 5、10、20、30、50、100 或 200。

## 設定欄位格式選項

欄位格式設置在一個子對話框中進行指定，該對話框在「類型」節點和「表格」節點的「格式」標籤中提供。如果在開啟此對話框之前已選取多個欄位，那麼已選取的第一個欄位的設定將用於所有已選取欄位。在此進行指定後按一下**確定**會將這些設定套用於「格式」標籤中選定的所有欄位。

下列選項針對每個欄位提供。其中很多設定也可以在「串流內容」對話框中指定。任何在欄位層次進行的設定都將置換串流指定的預設設定。

**日期格式。** 選取日期儲存欄位要使用的日期格式或 CLEM 日期函數將字串解析為日期時使用的日期格式。

**時間格式。** 選取時間儲存欄位要使用的時間格式或 CLEM 時間函數將字串解析為時間時使用的時間格式。

**數字顯示格式。** 您可以選擇標準 (#####.###)、科學 (#.###E+##) 或貨幣顯示格式 (\$###.##)。

**小數點符號。** 選取逗點 (,) 或句號 (.) 作為小數分隔字元。

**分組符號。** 針對為顯示格式編號，選取用於對值進行群組的符號（例如，3,000.00 中的逗點）。選項包括無、句點、逗點、空格以及語言環境所定義（在此情況下會使用現行語言環境的預設值）。

**小數位數（標準、科學、貨幣、匯出）。** 針對數字顯示格式，指定要在顯示實數時使用的小數位數。此選項將針對每個顯示格式個別指定。請注意，**匯出小數位**設定僅適用於純文字檔案匯出，將置換串流內容。純文字檔案匯出的串流預設值是為串流內容中的**標準小數位數**設定指定的任何值。「XML 匯出」節點匯出的小數位數始終為 6。

**排列整齊。** 指定欄中的值應採用的排列整齊方式。預設值為**自動**，它會向左調整符號值以及向右調整數值。您可通過選取「左」、「右」或「中心」來置換預設設置。

**直欄寬度。** 依預設，根據欄位值自動計算欄寬。您可使用清單框右邊的箭頭指定以五為間隔的自訂寬度。

---

## 過濾或重新命名欄位

您可以在串流中的任何點重新命名或執行欄位。例如，作為醫療研究人員，您可能不考量病患（記錄層次資料）的鉀水平（欄位層次資料）；因此，您可以過濾掉 K（鉀）欄位。可以使用單獨的「過濾器」節點，或者來源或輸出節點上的「過濾器」標籤，完成此作業。功能相同，而無論存取自哪個節點。

- 您可以從來源節點，例如「變數檔」、「固定檔」、Statistics 檔、XML 或「延伸匯入」，在將資料讀取至 IBM SPSS Modeler 時，重新命名或過濾欄位。
- 使用「過濾器」節點，您可以在串流中的任何點重新命名或過濾欄位。
- 從「Statistics 匯出」、「Statistics 轉換」、「Statistics 模型」及「Statistics 輸出」節點，您可以過濾或重新命名欄位，以符合 IBM SPSS Statistics 命名標準。請參閱第 316 頁的『重新命名或過濾 IBM SPSS Statistics 的欄位』主題，以取得更多資訊。
- 您可以使用上述任何節點中的「過濾器」標籤，以定義或編輯多重回應集。請參閱第 131 頁的『編輯多重回應集』主題，以取得更多資訊。
- 最後，您可以使用「過濾器」節點，將欄位從一個來源節點對映至另一個。

## 設定過濾選項

「過濾」標籤中使用的表格可顯示每個欄位進入和離開節點時的名稱。可以使用此表格中的選項對複製的或下游作業不需要的欄位進行更名或過濾。

- **欄位。** 顯示目前連接的資料來源中的輸入欄位。
- **過濾。** 顯示所有輸入欄位的過濾狀態。過濾後的欄位在此欄中帶有紅色 X，指示該欄位不會向下游傳送。按一下選定欄位的過濾欄可開啟和關閉過濾功能。此外，也可以採用按住 Shift 並按一下的選擇方法同時選擇多個欄位的選項。
- **欄位。** 在欄位離開「過濾器」節點時顯示這些欄位。複製名稱顯示為紅色。可以通過按一下此欄並輸入新名稱來編輯欄位名稱。也可以通過按一下過濾欄刪除欄位，以取消複製欄位。

通過按一下欄標題，可以對表格中所有直欄進行排序。

**檢視現行欄位。** 選取此選項可檢視當前連接到「過濾器」節點的資料集的欄位。此選項預設處於已選取狀態，是最一般的過濾節點使用方法。

**檢視未用的欄位設定。** 選取此選項可檢視曾連接至「過濾器」節點但已切斷連接的資料集的欄位。將過濾器節點從一個串流複製到另一個串流時，或儲存並重新載入過濾器節點時，此選項將十分有用。

## 過濾按鈕功能表

按一下對話框左上角落的「過濾」按鈕可存取含有多個快捷鍵和其他選項的功能表。

可以選擇執行下列操作：

- 刪除所有欄位。
- 納入全部欄位。
- 切換全部欄位。
- 刪除副本。請注意，選取此選項將移除具有該複製名稱的所有字段，包含第一次出現該名稱的字段。
- 更名欄位和複選題分析集以滿足其他應用程式的要求。請參閱第 316 頁的『重新命名或過濾 IBM SPSS Statistics 的欄位』主題，以取得更多資訊。
- 截斷欄位名稱。
- 匿名化欄位和多重回應集名稱。
- 使用輸入欄位名稱。
- 編輯多重回應集。請參閱第 131 頁的『編輯多重回應集』主題，以取得更多資訊。
- 設定預設過濾狀態。

您還可以使用對話框頂部的箭頭切換按鈕指定要依預設包含還是捨棄欄位。對於要在下游納入的欄位很少的大型資料集，這種方法十分有用。例如，可以僅選取要保留的欄位，並指定所有其他項目欄位都應捨棄（而不是逐一選取所有要捨棄的欄位）。

### 截斷欄位名稱

通過過濾按鈕功能表（「過濾」標籤的左上角落），您可以選擇截斷欄位名稱。

**最大長度。**指定欄位名稱長度的限制字元數。

**數字小數位數。**如果截斷後的欄位名稱不再唯一，那麼將對其進行進一步的截斷，並通過向名稱新增數字位來進行區分。您可指定使用的數字小數位數。使用箭頭按鈕可調整該位數。

例如，下表格說明如何使用預設設定（最大長度=8，數字小數位數=2）對某個醫學資料集中的欄位名稱進行截斷。

表 23. 欄位名稱截斷

欄位名稱	截斷後的欄位名稱
Patient Input 1	Patien01
Patient Input 2	Patien02
Heart Rate	HeartRat
BP	BP

### 匿名化欄位名稱

通過按一下左上角落的過濾按鈕功能表並選擇對欄位名稱進行匿名化，您可以對包含「過濾」標籤的任何節點中的欄位名稱進行匿名化。已匿名化的欄位名稱由一個字串字首及一個基於數值的唯一值組成。

對以下各項的名稱進行匿名化。選擇僅限已選取的資料欄位將僅對「過濾」標籤中的選取的欄位的名稱進行匿名化。預設設置為所有欄位，這表示將對所有欄位名稱進行匿名化。

欄位名稱字首。已匿名化的欄位名稱的預設字首為 **anon\_**；如果要使用其他字首，請選擇自訂並鍵入您自己的字首。

對複選題分析集進行匿名化。按照對欄位進行匿名化的方法對複選題分析集名稱進行匿名化。請參閱『編輯多重回應集』主題，以取得更多資訊。

要還原原始欄位名稱，請在過濾按鈕功能表中選擇使用輸入欄位名稱。

## 編輯多重回應集

您可以透過按一下左上角中的「過濾器」按鈕功能表，並選擇編輯多重回應集，從包括「過濾器」標籤的任何節點中新增或編輯多重回應集。

多重回應集用來記錄資料，針對每一個案例可以具有多個值，例如，當詢問意見調查調查受訪者他們參觀了哪些博物館或者閱讀了哪些雜誌時。多重回應集可以透過使用 Data Collection 來源節點或「統計資料檔案」節點，匯入至 IBM SPSS Modeler，並且可以透過使用「過濾器」節點，在 IBM SPSS Modeler 中進行定義。

按一下新建，以建立新的多重回應集，或者按一下編輯以修改現有集。

名稱及標籤。指定集的名稱及說明。

類型。可以使用以下兩種方法之一來處理多重回應問題：

- **多重二分法集**。針對每一個可能的回應建立單獨的旗標欄位；因此，如果有 10 本雜誌，則有 10 個旗標欄位，其中每一個都可以具有值，例如使用 0 或 1 表示 *true* 或 *false*。計數值容許您指定將哪個值計為 *true*。當您想要容許受訪者選擇所有適用選項時，此方法有用。
- **多重種類集**。為每一位受訪者建立一個名義欄位，可納入給定受訪者的回答數上限。每一個名義欄位具有代表可能的回答的值，例如 1 表示時間，2 表示新聞週，3 表示 PC 週。當您想要限制回答的數目時，例如要求受訪者選擇他們最經常閱讀的三本雜誌時，此方法最有用。

集中的欄位。使用右側的圖示，可新增或移除欄位。

備註(O)

- 多重回應集中包括的所有欄位都必須具有相同的儲存體。
- 集與它們包括的欄位不同。例如，刪除集不會導致刪除它包括的欄位，只會刪除那些欄位之間的鏈結。集在刪除點的上游仍可見，但在下游則不可見。
- 如果使用「過濾器」節點（直接在標籤上，或者透過在「過濾器」功能表上選擇「重新命名 IBM SPSS Statistics」、截斷或匿名選項）重新命名欄位，也將更新多重回應集中使用之這些欄位的所有參照。然而，不會從多重回應集中移除「過濾器」節點捨棄之多重回應集中的任何欄位。此類欄位雖然在串流中不再可見，但仍由多重參照集參照；例如，匯出時可以考量此項。

---

## 衍生節點

IBM SPSS Modeler 中最強大的功能之一是可以修改資料值並從現有資料中衍生新欄位。在漫長的資料採礦專案中，執行若干衍生動作是很常見的，如從 Web 日誌資料的字串中擷取客戶 ID，或根據交易和人口統計資料建立客戶生命期限值。所有這些轉換均可使用各式各樣的欄位作業節點完成。

若干節點可提供衍生新欄位的功能：



「衍生」節點將修改資料值或根據一個或多個現有欄位建立新欄位。它可建立的欄位類型包括公式、旗標、列名、狀態、計數和條件。



「再分類」節點可將一組種類值轉換為另一組值。重新分類對收合種類或重新分組分析資料非常有用。



「分級」節點根據一個或多個現有連續（數值型範圍）欄位的值自動建立新的列名（集合）欄位。例如，用戶可將連續收入欄位轉換為一個包含各群組收入的新的種類欄位，作為其與平均數之間的離差。一旦您為新欄位建立了 Bin，就可以根據切點來產生「衍生」節點。



「設為旗標」節點根據為一個或多個列名欄位定義的種類值獲取多個旗標欄位。



「重組」節點可將一個列名欄位或旗標欄位轉換為一群組欄位（該欄位群組由已成為另一欄位的值移入）。例如，給定一個名為 支付類型的欄位，其值為貸方、現金和借方，則將建立三個新欄位（貸方、現金、借方），每個欄位可能包含實際支付的值。



「歷史」節點將建立新欄位，其中包含之前記錄中的欄位資料。「歷程」節點通常用於序列資料，例如時間序列資料。使用「歷史」節點前，您可能希望使用「排序」節點對此資料進行排序。

## 使用衍生節點

使用衍生節點，可以根據一個或多個現有欄位建立六種類型的新欄位：

- 公式。新欄位是任意 CLEM 表示式的結果。
- 旗標。新欄位是代表指定條件的旗標。
- 名義。新欄位是名義的，表示其會員是一組指定值。
- 狀態。新欄位是兩種狀態之一。通過指定條件觸發、觸發程式這兩種狀態之間的切換。
- 計數。新欄位以滿足某個條件的次數為基礎。
- 條件。新欄位是兩個表示式的其中一個的值，具體取決於條件的值。

其中每個節點在 Derive 節點對話框中都包含一組特殊選項。後續的主題會討論這些選項。

請注意，使用下列各項可能會變更列順序：



- 通過 SQL 回送在資料庫中執行
- 通過遠端 IBM SPSS Analytic Server 執行
- 使用在內含式 IBM SPSS Analytic Server 中執行的函數
- 衍生清單（例如，請參閱第 135 頁的『衍生清單或地理空間欄位』）
- 呼叫空間函數所述的任何函數

## 為衍生節點設定基本選項

在衍生節點的對話框頂部，有用於選取所需衍生節點類型的若干選項。

**眾數。** 根據是否要衍生多個欄位，選取單個或多個。選取多個時，對話框將變為顯示用於多個衍生欄位的選項。

**衍生欄位。** 對於簡單的 Derive 節點，請指定要衍生並新增到每條記錄的欄位的名稱。預設名稱為 DeriveN，其中 N 是您在現行階段作業期間迄今為止已建立的「衍生」節點數。

**衍生為。** 從下拉清單中選取 Derive 節點的類型，例如「公式」或「列名」。對於每個類型，都會根據您在該類型特定的對話框中指定的條件建立新欄位。

從下拉清單中選取某個選項會將一組新的控制項新增到主對話框，具體取決於每種 Derive 節點類型的內容。

**欄位類型。** 為新衍生的節點選取測量層次，如「連續」、「種類」或「旗標」。此選項對於衍生節點的所有形式是通用的。

注意：衍生新欄位通常需要使用特殊的函數或數學表示式。為了協助您建立這些表示式，所有類型的衍生節點的對話框中都提供了表示式建置器，且此建置器提供了規則檢查和 CLEM 表示式的完整清單。

## 衍生多個欄位

在衍生節點內將模式設定為多個可以依據同一節點內的同一條件衍生多個欄位。如果要對資料集中的多個欄位進行相同轉換，使用此功能可節省時間。例如，如果要建立一個迴歸模型，用於根據起始工資和原有經驗預測目前工資，那麼對所有三個非對稱變數套用對數轉換可能很有幫助。您可以對所有欄位同時套用同一函數，而不是針對每種變換新增一個新的「衍生」節點。只需選取要從中衍生新欄位的所有欄位，然後使用 @FIELD 函數在欄位括弧內鍵入衍生表示式。

注意：對於同時衍生多個欄位，@FIELD 函數是一種重要的工具。它使您可以在無需指定確切欄位名稱的情況下參照目前欄位的內容。例如，用於對多個欄位套用對數變換的 CLEM 表示式為  $\log(@FIELD)$ 。

當您選取多個模式時，會將下列選項新增到對話框中：

**衍生自。** 使用「欄位選擇器」選擇要從其中衍生新欄位的欄位。對於每個選定欄位，將產生一個輸出欄位。注意：選取的欄位不需要具有同一儲存類型；但是，如果條件並非對於所有欄位均成立，那麼衍生作業將會失敗。

**欄位名稱延伸。** 鍵入您要新增至新欄位名稱的延伸。例如，對於包含 *Current Salary* 的記錄的新欄位，可以為欄位名稱新增延伸部分 *log\_*，從而產生 *log\_Current Salary*。使用圓鈕選擇要將副檔名新增為欄位名稱的字首（開頭）還是字尾（末尾）。預設名稱為 DeriveN，其中 N 是您在現行階段作業期間迄今為止已建立的「衍生」節點數。

在單個模式的「衍生」節點中，此時需要建立用於衍生新欄位的表示式。根據所選衍生作業的類型，可有多種建立條件的選項。後續的主題會討論這些選項。要建立表示式，可以簡單地鍵入公式欄位，也可以通過按一下計算機按鈕使用表示式建置器。請記住，在參照多個欄位的操作時可使用 @FIELD 函數。

## 選取多個欄位

對於對多個輸入欄位執行作業的所有節點（如衍生節點（多個模式）、「聚合」節點、排序節點、多重繪圖節點和時間統計圖節點），您可以使用「選取欄位」對話框輕鬆選取多個欄位。

**依次排序。** 可以通過選取下列其中一個選項對可用的欄位進行排序以供檢視：

- **自然。** 當欄位沿資料串流向下傳遞至現行節點時，檢視欄位的順序。
- **名稱。** 採用按字母順序對欄位進行排序以便於檢視。
- **類型。** 檢視欄位時按其測量層次排序。此選項在選取採用特定測量層級的欄位時有用。

從清單中一次選取一個欄位或使用 Shift-按一下和 Ctrl-按一下方法來選取多個欄位。您還可以使用清單下面的按鈕來基於測量層次選取欄位群組，或者選取或取消選取表格中的所有欄位。

## 設定衍生公式選項

「衍生公式」節點根據 CLEM 表示式的結果為資料集中的每條記錄建立一個新欄位。此表示式不能是條件表示式。要衍生基於條件式表示式的值，請使用旗標或條件類型的「衍生」節點。

**公式** 使用 CLEM 語言指定用於為新欄位衍生值的公式。

**註：**由於 SPSS Modeler 無法確定要用於所衍生的清單欄位的子測量層次，因此對於「收集」和「地理空間」測量層次，您可以按一下**指定...**以開啟「值」對話框，並設定所需的子測量層次。如需相關資訊，請參閱『設定衍生清單值』。

若為「地理空間」欄位，相關的儲存體類型只有**實數**和**整數**（預設值為**實數**）。

## 設定衍生清單值

您從「衍生」節點的「公式」欄位類型下拉清單中選取**指定...**時，將顯示「值」對話框。在此對話框中，可以設定用於「公式」欄位類型測量層次「收集」或「地理空間」的子測量層次值。

**測量** 請選取**收集**或**地理空間**。如果您選取了任何其他測量層次，那麼對話框將顯示一條訊息，指出沒有可編輯的值。

## 集合

您可以為集合測量層次設定的唯一項目是**清單測量**。依預設，此測量設定為「無類型」，但您可以選取另一個值，以設定清單中的元素的測量層次。您可以選取下列其中一個選項：

- 無類型
- 種類
- 連續的
- 名義(M)
- 序數(O)
- 旗標

## 地理空間

對於「地理空間」測量層次，您可以選取下列選項，以設定該清單中的元素的測量層次：

**類型** 選取地理空間欄位的測量子層次。可用的子層次由清單欄位的深度確定；預設值如下所示：

- 點（深度為零）

- 線串 (深度為 1)
- 多邊形 (深度為 1)
- 多點 (深度為 1)
- 多線串 (深度為 2)
- 多多邊形 (深度為 2)

有關子層次的進一步資訊，請參閱《SPSS Modeler 的「來源」節點、「過程」節點和「輸出」節點》手冊的『「類型」節點』部分中的「地理空間測量子層次」主題。

有關清單深度的進一步資訊，請參閱《SPSS Modeler 的「來源」節點、「過程」節點和「輸出」節點》手冊的『「來源」節點』部分中的「清單儲存以及相關聯的測量層次」主題。

**座標系統** 只有當您將測量層次從地理空間層次變成非地理空間層次時，此選項才可用。若要將座標系統套用至地理空間資料，請選取此勾選框。依預設，會顯示在**工具 > 串流內容 > 選項 > 地理空間**窗格中設定的座標系統。要使用另一個座標系統，請按一下**變更**按鈕以顯示「選擇座標系統」對話框，並選擇與資料相符的座標系統。

如需座標系統的相關資訊，請參閱《SPSS Modeler 使用手冊》之「使用串流」小節中的「設定串流的地理空間選項」主題。

## 衍生清單或地理空間欄位

在某些情況下，應該記錄為清單項目的資料以錯誤的屬性匯入到 SPSS Modeler 中。例如，作為個別的地理空間欄位（例如 x 座標和 y 座標或者經度和緯度）而導入，或者作為 .csv 檔案中的各個列而導入。在這種狀況下，必須將各個欄位組合成單個清單欄位；完成此任務的一種方法是使用「衍生」節點。

**註：**組合地理空間資料時，您必須知道哪個是 x（或經度）欄位，哪個是 y（或緯度）欄位。您必須對該資料進行組合，以使生成的清單欄位的元素順序為 [x, y] 或 [經度, 緯度]，這是地理空間座標的標準格式。

下列步驟顯示衍生清單欄位的簡單範例。

1. 在串流中，將「衍生」節點連接到「來源」節點。
2. 在「衍生」節點的「設定」標籤上，從衍生為清單中選取**公式**。
3. 在欄位類型中，選取**收集**（對於非地理空間清單）或**地理空間**。依預設，SPSS Modeler 使用「最佳猜測」方法來設定正確的清單詳細資訊；您可以選取**指定...**以開啟「值」對話框。此對話框可以用於收集以輸入有關清單中資料的更多資訊，對於地理空間，可以用於設定資料類型並指定資料的座標系統。

**註：**對於地理空間，您指定的座標系統必須與資料的座標系統完全相符。如果不是完全匹配，地理空間功能可能會產生不正確的結果。

4. 在**公式**窗格中，輸入用於將資料組合成正確清單格式的公式。並且，還可以按一下**計算機**按鈕以開啟表示式建置器。

用於衍生清單的公式的一個簡單範例是 [x, y]，其中 x 和 y 是資料來源中的個別欄位。建立的新衍生欄位是一個清單，其中每個記錄的值都是該記錄的並置 x 和 y 值。

**註：**以此方式已結合成清單的欄位必須具有同一種儲存類型。

有關清單和清單深度的進一步資訊，請參閱第 9 頁的『清單儲存以及相關聯的測量層次』。

## 設定衍生旗標選項

衍生旗標節點用於指示特定條件，如高血壓或客戶帳號閒置。對於每條記錄都會建立一個旗標欄位，當條件為 true 時，會在欄位中新增代表 true 的旗標值。

**True 值。** 指定針對滿足以下方指定條件的記錄要在旗標欄位中包含的值。預設值為 T。

**False 值。** 指定要併入不符合下方指定條件的記錄的旗標欄位的值。預設值是 F。

以下情況時為 **true**。指定某個 CLEM 條件，用於評估每條記錄的某些值，並為記錄賦予 true 值或 false（定義如上方）。請注意，對於非假數值，會將 true 值賦予記錄。

注意：要傳回空字串，您應該鍵入一對參照，並且中間不包含任何內容，如 ""。例如，空字串通常可用作 false，以使 true 值在表格中更為明顯。類似地，如果希望某個字串值在其他情況下被視為數字，應使用參照

範例

在 IBM SPSS Modeler 12.0 之前的版本中，可用逗點來分隔值將多個回應匯入一個欄位中。例如：

```
museum_of_design,institute_of_textiles_and_fashionmuseum_of_designarcheological_museum$null$  
national_art_gallery,national_museum_of_science,other
```

要準備分析該資料，可以使用 `hassubstring` 函數為每個回應產生個別的旗標欄位，所使用的表示式如下：

```
hassubstring(museums,"museum_of_design")
```

## 設定衍生名義選項

Derive Nominal 節點用於執行一組 CLEM 條件，以確定每條記錄滿足的條件。當每條記錄符合某個條件時，會將一個值（指示符合哪組條件）新增到新的衍生欄位。

**預設值。** 指定不滿足任何條件時要用於新欄位的值。

**將欄位設定為。** 指定符合某個特定條件時要在新欄位中輸入的值。清單中的每個值都有一個關聯條件，該條件由用戶在相鄰欄中指定。

**若此條件為 true。** 為集合欄位中要列出的每個會員指定條件。使用表示式建置器在可用的函數和欄位中進行選取。可以使用箭頭和刪除按鈕對條件進行重新排序或刪除。

條件的工作原理是對資料集中特定欄位的值進行測試。測試每個條件時，都會為新欄位分配上述指定值，以指示符合哪個條件（如果有）。如果不滿足任何條件，則會使用預設值。

## 設定衍生狀態選項

Derive State 節點與 Derive Flag 節點相當類似。旗標節點根據現行記錄對單個條件的滿足情況設定值，而衍生狀態節點可以根據欄位對兩個獨立的條件的滿足方式變更該欄位的值。這意味著符合每個條件時，該值都會發生變更（開啟或關閉）。

**初始狀態。** 選取初始時要為新欄位的每條記錄指定 **On** 還是 **Off**。請注意，此值可能在符合每個條件時發生變更。

**"On" 值。** 指定符合 On 條件時新欄位的值。

以下情況時切換 **"On"**。指定會在條件為 true 時將狀態變更為「開啟」的 CLEM 條件。按一下計算機按鈕以開啟「表示式建置器」。

**"Off"** 值。指定符合 Off 條件時新欄位的值。

以下情況時切換 **"Off"**。指定會在條件為假時將狀態變更為「關」的 CLEM 條件。按一下計算機按鈕以開啟「表示式建置器」。

注意：要指定空字串，您應該鍵入一對參照，並且中間不包含任何內容，如 ""。類似地，如果希望某個字串值在其他情況下被視為數字，那麼應使用參照。

## 設定衍生計數選項

Derive Count 節點用於對資料集中數值欄位的值套用一系列條件。滿足每個條件時，衍生的計數欄位的值會根據集合增量的大小而相應增加。這種類型的 Derive 節點對於時間序列資料十分有用。

**初始值**。設定執行新欄位時使用的值。初始值必須是數值常數。使用箭頭按鈕可以增加或減少該值。

**遞增時機**。指定某個 CLEM 條件，滿足該條件時將根據「增量」中指定的數字變更衍生值。按一下計算機按鈕以開啟「表示式建置器」。

**增加的值**。設定用於增加計數的值。可以使用數值常數或 CLEM 表示式的結果。

**重設時機**。指定某個條件，滿足該條件時會將衍生值重設為初始值。按一下計算機按鈕以開啟「表示式建置器」。

## 設定衍生條件選項

Derive Conditional 節點使用一系列 If-Then-Else 陳述式衍生新欄位的值。

**If**。指定一個 CLEM 條件，在執行時會逐一為每條記錄評估該條件。如果條件為 true（對於數字的情況為非假），那麼會為新欄位賦予下方通過 Then 表示式指定的值。按一下計算機按鈕以開啟「表示式建置器」。

**Then**。指定上述 If 陳述式為 true（或非假）時新欄位的值或 CLEM 表示式。按一下計算機按鈕以開啟「表示式建置器」。

**Else**。指定上述 If 陳述式為假時新欄位的值或 CLEM 表示式。按一下計算機按鈕以開啟「表示式建置器」。

## 使用衍生節點對值進行重新編碼

Derive 節點還可用於對值進行重新編碼，例如，通過將具有種類值的字串欄位轉換為數值型名義（集合）欄位。

1. 在「衍生為」中選取適當的欄位類型（「名義」、「旗標」等）。
2. 指定對值進行重新編碼的條件。例如，可以設定為：如果 Drug='drugA' 則值為 1，如果 Drug='drugB' 則值為 2，依此類推。

---

## 填充器節點

填入節點用於取代欄位值和變更儲存類型。您可以選擇基於指定的 CLEM 條件（如 @BLANK(FIELD)）取代數值。或者，您可以選擇將所有空白或空值取代為特定值。「填入器」節點通常與「類型」節點結合使用，用於取代遺漏值。例如，可以通過指定表示式（如 @GLOBAL\_平均值）為空白值填充欄位的平均數。此表示式將為所有空白值填充通過設定廣域節點計算的平均數。

**填入欄位**。使用欄位選擇器（文字欄位右邊的按鈕）從資料集中選擇要檢查並取代其值的欄位。預設行為是根據「條件」和「取代」將值取代為下方指定的表示式。另外，也可以使用下方的「取代」選項選取替代的取代方法。

註：選取要取代為使用者定義值的多個欄位時，欄位類型相似（均為數值或均為符號）是很重要的。

**取代。**選取此選項可使用下列其中一種方法來取代選取的欄位的值：

- **基於條件。**此選項將啟動「條件」欄位和表示式建置器，使用它們可建立用作將值取代為指定值的條件的表示式。
- **一律。**取代選取的欄位的所有值。例如，可以通過此選項使用下列 CLEM 表示式將 *income* 的儲存類型轉換為字串：`(to_string(income))`。
- **空白值。**取代所選欄位中的所有使用者指定的空白值。採用標準條件 `@BLANK(@FIELD)` 選取空白值。注意：您可以使用來源節點的「類型」標籤或使用「類型」節點定義空白值。
- **空值。**取代選取的欄位中的所有系統空值。採用標準條件 `@NULL(@FIELD)` 選取空值。
- **空白值與空值。**取代選取的欄位中的空白值和系統空白值。當您不確定是否已將無效定義為遺漏值時，此選項將十分有用。

**條件。**此選項在已選取**根據條件**選項後可用。使用此文字框指定用於評估選取的欄位的 CLEM 表示式。按一下計算機按鈕以開啟「表示式建置器」。

**取代為。**指定某個 CLEM 表示式來為選取的欄位賦予新值。此外，也可以通過在文字框中鍵入 `undef` 將值取代為空值。按一下計算機按鈕以開啟「表示式建置器」。

註：如果所選欄位為字串，那麼應將其取代為字串值。使用預設值 0 或其他數值作為字串欄位的取代值將產生錯誤。

請注意，使用下列各項可能會變更列順序：

- 通過 SQL 回送在資料庫中執行
- 通過遠端 IBM SPSS Analytic Server 執行
- 使用在內含式 IBM SPSS Analytic Server 中執行的函數
- 衍生清單（例如，請參閱第 135 頁的『衍生清單或地理空間欄位』）
- 呼叫空間函數所述的任何函數

## 使用填入節點進行儲存類型轉換

使用「填入器」節點的「取代」條件，可以輕鬆轉換單個或多個欄位的欄位儲存類型。例如，借助下列 CLEM 表示式，使用轉換函數 `to_integer` 可以將 *income* 從字串轉換為整數：`to_integer(income)`。

可以使用表示式建置器檢視可用的轉換函數並自動建立 CLEM 表示式。在「函數」下拉清單中，選取轉換可檢視儲存類型轉換函數的清單。可用的轉換函數如下：

- `to_integer(ITEM)`
- `to_real(ITEM)`
- `to_number(ITEM)`
- `to_string(ITEM)`
- `to_time(ITEM)`
- `to_timestamp(ITEM)`
- `to_date(ITEM)`
- `to_datetime(ITEM)`

**轉換日期與時間值。**請注意，轉換函數（以及需要指定類型輸入（如日期或時間值）的任何其他函數）取決於「串流選項」對話框中指定的現行形式。例如，如果要將值為 *Jan 2003*、*Mar 2003* 等的字串欄位轉換為日期儲存類型，請選取 **MON YYYY** 作為串流的預設日期格式。

轉換函數也可從「衍生」節點取得，用於在衍生計算期間進行暫時轉換。另外，還可以使用「衍生」節點來執行其他操作，例如使用種類值對字串欄位進行重新編碼。請參閱第 137 頁的『使用衍生節點對值進行重新編碼』主題，以取得更多資訊。

---

## 再分類節點(C)

「再分類」節點可實現一組種類值到另一組種類值的變換。重新分類對收合種類或重新分組分析資料非常有用。例如，可以將 *Product* 的值再分類為三群組，如 *Kitchenware*、*Bath and Linens* 和 *Appliances*。通常情況下，此作業按分組值直接通過「分佈」節點執行，並且生成一個「再分類」節點。請參閱第 206 頁的『使用分佈節點』主題，以取得更多資訊。

可以對一個或多個符號欄位執行重新分類。您也可以選擇為現有欄位替換新值或產生新欄位。

### 何時使用再分類節點

在使用再分類節點之前，請考慮是否有更適用於當前作業的其他欄位作業節點：

- 要採用自動方法將數值型範圍轉換為集合（如等級或百分位數），應使用「分級」節點。請參閱第 142 頁的『分組節點(B)』主題，以取得更多資訊。
- 要將數值型範圍手動分類為集合，應使用「衍生」節點。例如，如果要將工資值壓縮至特定的工資範圍種類，應使用「衍生」節點手動定義每個種類。
- 要根據種類欄位（如 *Mortgage\_type*）的值建立一個或多個旗標欄位，應使用設為旗標節點。
- 要將種類欄位轉換為數值型範圍，可以使用「衍生」節點。例如，可以將 *No* 和 *Yes* 值分別轉換為 0 和 1。請參閱第 137 頁的『使用衍生節點對值進行重新編碼』主題，以取得更多資訊。

## 為再分類節點設定選項

再分類節點的使用分為以下三個步驟：

1. 首先，選取要對多個欄位還是單個欄位進行再分類。
2. 下面，選擇是在現有欄位內重新編碼還是建立新欄位。
3. 然後，根據需要使用再分類節點對話框中的動態選項對映集合。

**眾數。** 選取單個可對一個欄位進行再分類。選取多個將啟動若干選項，它們可實現同時變換多個欄位。

**再分類成。** 選取新欄位將保留原始列名欄位，並衍生包含重新分類別的值的新欄位。選取現有欄位將使用新的分類改寫原始欄位中的值。此選項實質上是一種「填充」作業。

指定模式和取代選項後，必須選取變換欄位並使用對話框下半部分的動態選項指定新的分類值。這些選項會依據前面所選模式的不同而變化。

**再分類欄位。** 使用右邊的欄位選擇器按鈕選擇一個（「單個」模式）或多個（「多個」模式）種類欄位。

**新欄位名稱。** 為包含重新編碼的值的新列名欄位指定名稱。如果前面選取了新欄位，此選項僅在「多個」模式下可用。如果選取了現有欄位，則會保留原始欄位名稱。採用「多個」模式時，此選項將被其他控制項取代，以指定向每個新欄位新增的副檔名。請參閱第 140 頁的『對多個欄位進行再分類』主題，以取得更多資訊。

**再分類值。**使用此表格，可以實現從舊集合值到此處指定的集合值的明確對映。

- **原始值。** 此直欄列出所選欄位的現有值。
- **新建值。** 使用此直欄來輸入新的種類值或從下拉清單選取值。當您使用「分佈」圖表中的值自動產生「重新分類」節點時，這些值會併入下拉清單中。這可讓您快速將現有值對映至已知的值集。例如，保健組織有時候會根據網路或語言環境以不同方式對診斷進行分組。在合併或收購之後，所有相關方都需要以一致的方式來重新分類新的甚至是現有的資料。無需手動輸入長清單中的每個目標值，您可以將主要值清單讀取至 IBM SPSS Modeler，執行診斷 欄位的「分佈」圖表，並直接從圖表為這個欄位產生「重新分類（值）」節點。此程序可讓所有目標「診斷」值都能在「新建值」下拉清單中可用。

4. 按一下**取得**讀取前面選取的一個或多個欄位的原始值。
5. 按一下**複製**針對尚未對映的欄位將原始值貼上至 新值欄。未對映的原始值將新增到下拉清單中。
6. 按一下**清除**將擦除 新值欄中的所有指定值。注意：此操作不會將值從下拉清單中擦除。
7. 按一下**自動**可自動產生代表每個原始值的連續整數。只能產生整數值，不能產生實數值（如 1.5、2.5 等）。

例如，可以自動產生代表供應項目名的連續供應項目 ID，或代表大學課程的課程編號。此功能對應於 IBM SPSS Statistics 中集合的自動重新編碼變換。

**用於未指定的值。**此選項用於在新欄位中填充未指定的值。可以選擇保留原始值（選擇**原始值**），也可以指定預設值。

## 對多個欄位進行再分類

要一次對映多個欄位的種類值，請將模式設定為**多個**。這時「再分類」對話框中將顯示新的設定，下方介紹這些設定。

**再分類欄位。**使用右邊的「欄位選擇器」按鈕選擇要轉換的欄位。使用欄位選擇器，可以同時選中所有欄位或的相似類型的欄位，如列名或旗標。

**欄位名稱延伸。**同時對多個欄位進行重新編碼時，指定新增到所有新欄位的副檔名比指定各個欄位名稱更為高效。指定延伸部分（如 `_recode`），然後選取要將其附加到原始欄位名稱的後面還是前面。

## 再分類欄位的儲存類別型和測量層次

「再分類」節點總是通過重新編碼作業建立列名欄位。在某些情況下，這種方法可能會在使用**現有欄位**重新分類模式時變更欄位的測量層次。

新欄位的儲存類型（資料的儲存方式，而不是使用方式）基於「設定」標籤的下列選項進行計算：

- 如果將未指定的值設定為使用預設值，儲存類型將按以下方式決定：檢查新值和預設值，並確定適當的儲存類型。例如，如果所有值均可剖析為整數，欄位將獲得整數儲存類型。
- 如果將未指定的值設定為使用原始值，儲存類型將以原始欄位的儲存類型為基礎。如果所有值均可剖析為原始欄位的儲存類型，那麼將保留該儲存類型；否則，儲存類型將通過尋找同時包含舊值和新值的最適合的儲存類型來確定。例如，使用再分類  $4 \Rightarrow 0, 5 \Rightarrow 0$  對整數集合  $\{ 1, 2, 3, 4, 5 \}$  進行再分類將產生新的整數集合  $\{ 1, 2, 3, 0 \}$ ，而使用  $4 \Rightarrow \text{"Over 3"}, 5 \Rightarrow \text{"Over 3"}$  則會產生字串集合  $\{ "1", "2", "3", \text{"Over 3"} \}$ 。

注意：如果原始類型是非實例化類型，那麼新類型也將是非實例化類型。



---

## 匿名化節點

通過使用匿名化節點，您可以在處理要包含在節點的下游模型中的資料時對欄位名稱和/或欄位值進行掩飾。這樣，可以隨意分發所產生的模型（例如，分發至技術支援部門），而未授權使用者無法檢視機密資料（例如，員工記錄或病患的醫療記錄）。

您可能需要對其他節點進行變更，具體取決於匿名化節點在串流中的位置。例如，如果在「選取」節點上游中插入一個匿名化節點，那麼該「選取」節點中的選擇準則作用於現已匿名化的值時，這些準則需要進行變更。

用於匿名化的方法取決於多種因素。對於欄位名稱以及除「連續」測量層次外的所有欄位值，資料將取代之為以下形式的字串：

*prefix\_Sn*

其中 *prefix\_* 是使用者指定的字串或預設字串 *anon\_*，*n* 是從 0 開始並在遇到每個唯一值時遞增（例如，*anon\_S0*、*anon\_S1* 等）的整數值。

類型為「連續」的欄位值必須進行轉換，因為數值型範圍處理的是整數或實數值，而不是字串。因此，只能通過將範圍轉換為不同範圍對字段值進行匿名化，從而掩飾原始資料。範圍內的值 *x* 的轉換按下列方法執行：

$$A*(x + B)$$

其中：

*A* 是比例係數，必須大於 0。

*B* 是要為值增加的轉換偏移量。

範例

對於年齡欄位，如果比例係數 *A* 設定為 7 而轉換偏移量 *B* 設定為 3，那麼年齡的值將轉換為：

$$7*(AGE + 3)$$

## 匿名化節點的設定選項

您可在此處選擇要在較下游位置對其值進行掩飾的欄位。

請注意，必須先在匿名化節點的上游對資料欄位進行實例化，然後才能執行匿名化作業。您可以通過在類型節點或在來源節點的「類型」標籤上按一下讀取值按鈕對資料進行實例化。

**欄位。** 列出現行資料集中的欄位。如果有任何欄位名稱已匿名化，那麼會在此處顯示匿名化名稱。

**測量。** 欄位的測量層次。

**匿名值。** 選擇一個或多個欄位，並按一下此欄，然後選擇是以使用預設值字首 **anon\_** 對欄位值進行匿名化；選擇**指定**以顯示一個對話框，您可在其中輸入自己的字首，或者對於類型為連續的欄位值，也可以指定欄位值的變換將採用隨機值還是使用者指定值。請注意，不能在同一作業中對連續和非連續欄位類型進行指定，必須分別針對每種欄位類型執行此作業。

**檢視現行欄位。** 選取此選項可檢視當前連接至匿名化節點的資料集的欄位。依預設會選取這個選項。

**檢視未用的欄位設定。** 選取此選項可檢視曾連接至該節點但已切斷連接的資料集的欄位。將節點從一個串流複製到另一個串流時，或儲存並重新載入節點時，此選項十分有用。

## 指定如何對欄位值進行匿名化

通過使用「取代數值」對話框，您可以選擇對匿名化的欄位值使用預設值字首還是自訂字首。在此對話框中按一下**確定**可針對選取的欄位將「設定」標籤中的「已匿名化值」設定變更為是。

**欄位值字首。**匿名化的欄位值的預設字首為 **anon\_**；如果要使用其他字首，請選擇**自訂**並輸入您自己的字首。

「轉換值」對話框僅針對類型為「連續」的欄位顯示，可用於指定欄位值的轉換使用隨機值還是使用者指定值。

**隨機。**選擇此選項將對轉換使用隨機值。**設定隨機種子**依預設處於已選取狀態；請在**種子**中指定一個值，或使用預設值。

**固定。**選擇此選項可為轉換指定您自己的值。

- **乘數。**欄位值在轉換中的乘數。最小值為 1；最大值通常為 10，但有時需要減小該值以避免溢位。
- **轉換幅度。**在轉換中將為欄位值增加的數字。最小值為 0；最大值通常為 1000，但有時需要減小該值以避免溢位。

## 對欄位值進行匿名化

已在「設定」標籤上選取進行匿名化的欄位，其值將在以下情況下進行匿名化：

- 執行包含匿名化節點的串流時
- 對值進行預覽時

要對值進行預覽，請按一下「已匿名化值」標籤中的**已匿名化值**按鈕。然後，從下拉清單中選取一個欄位名稱。

如果測量層次為「連續」，那麼將顯示以下內容：

- 原始範圍的最小值和最大值
- 用於轉換值的方程式

如果測量層次不是「連續」，那麼畫面將顯示該欄位的原始值和匿名化的數值。

如果屏幕顯示黃色背景，那麼這表示選取的欄位的設定自上次已匿名化值以來已發生變更，或者表示已對匿名化節點的上游資料進行變更，導致匿名化的數值發生錯誤。此時將顯示目前值集合；再次按一下**已匿名化值**按鈕可根據目前設定產生一組新值。

**匿名值。**為選取的欄位建立匿名化的數值並在表格中顯示這些值。如果針對類型為「連續」的欄位使用隨機播種，那麼每次按一下此按鈕都會建立一組不同的值。

**清除值。**從表格中清除原始值和匿名化的數值。

---

## 分組節點(B)

使用 Binning 節點，可以根據一個或多個現有連續（數值型範圍）欄位的值自動建立新的名義欄位。例如，可以將連續收入欄位轉換為包含若干等寬收入群組的新的種類欄位，或轉換為與平均數之間的離差。或者，也可以選取一個「監督者」種類欄位，以保持兩個欄位之間原始關聯的強度。

分級的實用性源於以下幾個原因：

- **演算法要求。**某些特定演算法（如素樸貝葉斯、邏輯迴歸）要求種類輸入。

- **效能。** 如果減少輸入欄位的不同值數量，演算法（如多項式 Logistic）的性能可能會提高。例如，對每個分級使用中位數或平均數，而不使用原始值。
- **資料隱私。** 敏感類個人資訊（如工資）可採用範圍的報告形式，而不使用實際工資數字，以保護個人隱私。

提供了一些 Binning 方法。一旦您為新欄位建立了 Bin，就可以根據切點來產生「衍生」節點。

何時使用 bin 節點

在使用分組節點之前，請考慮是否有更適用於當前作業的其他技術：

- 要為種類手動指定分割點（如特定的預先定義工資範圍），請使用 Derive 節點。請參閱第 131 頁的『衍生節點』主題，以取得更多資訊。
- 要為現有集合建立新的種類，請使用再分類節點。請參閱第 139 頁的『再分類節點(C)』主題，以取得更多資訊。

遺漏值處理

分組節點處理遺漏值的方法如下：

- **使用者指定的空白值。** 變換過程中將併入指定為空白值的遺漏值。例如，若使用類型 節點指定 -99 表示空白值，那麼會在分級過程中併入此值。要在分級過程中忽略空白值，應使用填充節點將空白值取代為系統空值。
- **系統遺漏值 (\$null\$)。** 在分級變換過程中，空值將被忽略，並在變換之後保持空值。

「設定」標籤提供了有關適用技術的選項。「視圖」標籤將顯示針對先前通過節點的資料建立的割點。

## 為 bin 節點設定選項

使用分組節點，可以採用下列技術自動產生 Bin（種類）：

- 固定寬度 Binning
- 分位顯示（相等計數或總和）
- 平均數和標準差
- 等級
- 相對於種類「監督者」欄位最佳化

對話框的下部分會根據所選的 Binning 方法動態變化。

**Bin 欄位。** 此處將顯示待變換的連續（數值型範圍）欄位。使用分組節點，可以同時對多個欄位進行分級。使用右側的按鈕可新增或刪除欄位。

**分組方法。** 選取用於確定新欄位 Bin（種類）的分割點的方法。後續主題說明每個觀察值中提供的選項。

**分組臨界值。** 指定如何計算 Bin 臨界值。

- **始終重新計算。** 執行節點時，將一律重新計算分割點和 bin 分配。
- **從「Bin 值」標籤讀取（如果可用）。** 僅在必要時（例如，新增新資料後）計算分割點和分級配置。

下列主題將介紹可用分級方法的選項。

## 固定寬度 Bin

選擇 **固定寬度** 作為 Binning 方法時，對話框中會顯示一組新的選項。

**副檔名。** 指定用於所產生欄位的副檔名。 *\_BIN* 是預設副檔名。您還可以指定是否將副檔名新增至欄位名稱的開頭（字首）或結尾（字尾）。例如，可以產生名為 *income\_BIN* 的新欄位。

**Bin 寬度。** 指定用於計算分級「寬度」的值（整數或實數）。例如，可以使用預設值 10 對欄位 *Age* 進行分級。由於 *Age* 的範圍為 18–65，因此產生的 Bin 如下表格所示。

表 24. 經歷時間範圍為 18–65 的 Bin

Bin 1	Bin 2	Bin 3	分級 4	分級 5	分級 6
>=13 至 <23	>=23 至 <33	>=33 至 <43	>=43 至 <53	>=53 至 <63	>=63 至 <73

分級間隔起點的計算方法為：掃描到的最低值減去 Bin 寬度的一半（指定值）。例如，在上方顯示的 Bin 中，使用 13 作為間隔的起點，依據的計算方法如下： $18 [\text{最低資料值}] - 5 [0.5 \times (\text{Bin 寬度 } 10)] = 13$ 。

**Bin 數。** 使用此選項可指定用於確定新欄位的固定寬度 Bin（種類）數的整數。

在串流中執行分組節點後，即可通過按一下分組節點對話框中的 **預覽** 標籤來檢視產生的 Bin 臨界值。請參閱第 147 頁的『預覽產生的 Bin』主題，以取得更多資訊。

### 分位顯示（相等計數或總和）

分位顯示 Binning 方法用於建立名義欄位，這些欄位可用於將掃描到的記錄分割為百分位數（或四分位數、十分位數等）群組，使每個群組包含相同數量的記錄，或使每個群組中值的總和相等。記錄根據指定的分級欄位值按升冪排列，因此所選分級變數的值最低的記錄將獲得等級 1，下一組記錄等級為 2，依此類推。每個分級的臨界值將根據所用的資料和分位方法自動產生。

**分位延伸名。** 指定用於使用標準 *p* 分位顯示產生的欄位的副檔名。預設副檔名是 *\_TILE* 加上 *N*，其中 *N* 是圖磚號。您還可以指定是否將副檔名新增至欄位名稱的開頭（字首）或結尾（字尾）。例如，可以產生名為 *income\_BIN4* 的新欄位。

**自訂分位延伸名。** 指定用於自訂分位顯示範圍的副檔名。預設值是 *\_TILEN*。請注意，此處的 *N* 將不會被自訂數字取代。

可用的 *p* 分位顯示如下：

- **四分位數。** 產生 4 個 Bin，每個包含 25% 的觀察值。
- **五分位數。** 產生 5 個 Bin，每個包含 20% 的觀察值。
- **十分位數。** 產生 10 個 Bin，每個包含 10% 的觀察值。
- **二十分位數。** 產生 20 個 Bin，每個包含 5% 的觀察值。
- **百分位數。** 產生 100 個 Bin，每個包含 1% 的觀察值。
- **自訂 N。** 選取此選項可指定 Bin 個數。例如，值為 3 將產生 3 個劃分種類（2 個割點），每個包含 33.3% 的觀察值。

請注意，如果資料中的離散值少於指定的分位顯示，那麼不會使用任何分位顯示。在這種情況下，新的分佈很可能反映資料的原始分佈。

**分位方法。** 指定用於為 Bin 分配記錄的方法。

- **記錄計數。** 盡量為每個分級分配相等號碼的記錄。
- **值總和。** 為 Bin 分配記錄時，盡量使每個 Bin 中值的總和相等。例如，以銷售業績為目標時，此方法可用於根據每條記錄的值為十分位數群組分配預期業績，最高分級獲得價值最高的預期業績。例如，某制藥公司可根據所開啟處方的數量將醫師分入十分位數群組。儘管每個十分位數包含的底方數大致相同，但提

供這些底方的個人數並不會相同，所開底方最多的個人集中在十分位數 10 中。請注意，此方法會假定所有值均大於零，如果實際情況不是這樣則可能產生意外結果。

**鏈結。** 當分割點兩端的值相同時，將產生結條件。例如，如果是分配十分位數，且超過 10% 的記錄的分級欄位具有相同值，那麼除非對臨界值進行向上或向下的強制轉換，否則無法將這些記錄全部分配至同一分級。可以將同分值上移至下一個 Bin，也可以保留在目前 Bin 中，但必須將其解析，使具有相同值的所有記錄位於同一 Bin 內，即使這樣會導致某些 Bin 的記錄數超過預期值也是如此。後續 Bin 的臨界值可能也會因此發生調整，導致對相同數字集合進行不同的值分配，具體取決於用於解析同分值的方法。

- **新增至下一個。** 選取此項以將鏈結值向上移至下一個 Bin。
- **保留在現行中。** 將鏈結值保留在現行（下方）Bin 中。此方法可能會導致建立的 Bin 總數較少。
- **隨機分配。** 選取此選項可將同分值隨機配置至一個分級。這將試圖使每個分級中的記錄數數量相等。

範例：按記錄計數分位

下表說明按記錄計數進行分位時如何將簡單欄位值分為四分位數。請注意，結果將隨選取的同分值選項而變化。

表 25. 依記錄計數範例排列。

值	新增至下一個	保留在目前分級中
10	1	1
13	2	1
15	3	2
15	3	2
20	4	3

每個分級的項目數的計算方法如下：

值的總數 / 分位顯示

在上方的簡單範例中，每個分級的所需項目數為 1.25（5 個值 / 4 個四分位數）。值 13（值編號為 2）跨越 1.25 的所需計數臨界值，因此將根據所選的同分值選項進行不同處理。在 **新增到下一個** 模式下，會將其新增到分級 2 中。在 **保留在目前分級中** 模式下，會將其留在分級 1 中，從而將分級 4 的值範圍推到現有資料值範圍外。結果是，僅建立三個 Bin，每個 Bin 的臨界值將進行相應調整，如下表格中所示。

表 26. Binning 範例結果。

分級	下限	上限
1	>=10	<15
2	>=15	<20
3	>=20	<=20

注意：啟用平行處理可提高按分位顯示分級的速度。

## 等級觀察值

選擇 **排等級** 作為 Binning 方法時，對話框中會顯示一組新的選項。

分等級可建立包含數字欄位的分等級值、分數分等級值和百分位數值的新欄位，具體取決於下方指定的選項。

**排等級順序。** 選取遞增（將最低值標示為 1）或遞減（將最高值標示為 1）。

等級。 選取此選項將按上方指定的遞增或遞減對觀察值進行排等級。新欄位中的值的範圍將是 1–N，其中 N 是原始欄位中離散值的號碼。結值將獲得其排等級值的平均值。

分數等級值。選取此選項將對觀察值進行排等級，其中新欄位的值等於排等級值除以非遺漏觀察值的加權和。分數排名範圍在 0–1 之間。

分數排等級百分比值。每個等級皆除以具有有效值的記錄數目，然後乘以 100。分數排名百分比範圍在 1–100 之間。

副檔名。對於所有排等級選項，還可以建立自訂副檔名，並指定將其新增到欄位名稱的開頭（字首）還是末尾（字尾）。例如，可以產生名為 *income\_P\_RANK* 的新欄位。

## 均數/標準差

選擇均數/標準差作為 Binning 方法時，對話框中會顯示一組新的選項。

此方法可根據指定欄位分佈的均數和標準差的值產生具有劃分種類的一個或多個新欄位。選取下方要使用的離差數。

副檔名。指定用於所產生欄位的副檔名。\_SDBIN 是預設副檔名。您還可以指定是否將副檔名新增至欄位名稱的開頭（字首）或結尾（字尾）。例如，可以產生名為 *income\_SDBIN* 的新欄位。

- +/- 1 標準差。選取此選項將產生三個 Bin。
- +/- 2 標準差。選取此選項將產生五個 Bin。
- +/- 3 標準差。選取此選項將產生七個 Bin。

例如，選取 +/-1 標準差將產生三個 Bin，計算方法如下表格所示。

表 27. 標準偏差 Bin 範例。

Bin 1	Bin 2	Bin 3
$x < (\text{Mean} - \text{Std. Dev})$	$(\text{Mean} - \text{Std. Dev}) \leq x \leq (\text{Mean} + \text{Std. Dev})$	$x > (\text{Mean} + \text{Std. Dev})$

在常態分佈中，68% 的觀察值落入與均數相距不到一個標準差的範圍內，95% 落入兩個標準差的範圍內，99% 落入三個標準差的範圍內。但請注意，根據標準差建立帶狀種類可能會使某些 Bin 在實際的資料範圍外，甚至在可能的資料值範圍（例如，負值工資範圍）外定義。

## 最適 Binning

如果要 Bin 的欄位與另一個種類欄位強關聯，則可選取種類欄位作為「監督者」欄位以便以類似於保留兩個欄位間的原始關聯強度的方式建立 Bin。

例如，假定已採用叢集分析根據家庭貸款的拖欠率對狀態進行群組，那麼最高拖欠率將位於第一個叢集中。在這種情況下，可以選擇 過期百分比 和 取消贖取權百分比 作為 Bin 欄位和模型產生的作為管理員欄位的叢集會員資格欄位。

名稱延伸 指定要用於所產生欄位的副檔名，以及是將其新增到欄位名稱開頭（字首）還是末尾（字尾）。例如，可以產生名為 *pastdue\_OPTIMAL* 的新欄位以及名為 *inforeclosure\_OPTIMAL* 的另一個欄位。

管理員欄位 這是用於構造 Bin 的種類欄位。

對於大型資料集，預先對欄位進行 Binning 以提高效能指示是否應使用前置處理來簡化最佳化 Binning。該方法會採用簡單的非監督式 Binning 方法將尺度值群組為大量 Bin，以平均數代表每個 Bin 中的值，並在繼續

監督式 Bin 之前對觀察值加權進行相應調整。在實際應用中，此方法會犧牲一定的精準度以換取速度，建議用於大型資料集。使用此選項時，也可以指定任意變數前置處理後的最大 bin 個數。

將觀察值計數相對較小的 Bin 與較大的相鄰 Bin 進行合併。如果啟用，則指示當該分級大小（觀察值的個數）與相鄰分級大小的比值小於指定的臨界值時，將合併分級；請注意臨界值越大合併的分級越多。

## 分割點設定

使用「分割點設定」對話框，可以指定最佳化分級演算法的進階選項。這些選項將指示演算法如何使用目標欄位計算 Bin。

**Bin 終點。**可以指定應包括（下限  $\leq x$ ）還是排除（下限  $< x$ ）端點下限或上限。

**第一個 Bin 和最後一個 Bin。**對於第一個和最後一個 Bin，可以指定該 Bin 應無限制（向正無窮或負無窮延伸）還是按最低或最高資料點進行限制。

## 預覽產生的 Bin

使用 Binning 節點中的「Bin 值」標籤，可以檢視產生的 Bin 的臨界值。套用「產生」功能表還可以產生 Derive 節點，該節點可用於將一個資料集中的臨界值套用於另一個資料集。

**Binned 欄位。** 使用下拉清單選取要檢視的欄位。為明確起見，顯示的欄位名稱採用原始欄位名稱。

**分位顯示。**使用下拉清單選取用於檢視的分位顯示，如 10 或 100。僅當 Bin 採用分位顯示方法（相等計數或總和）產生時，此選項才可用。

**分組臨界值。** 此處顯示每個產生的 Bin 的臨界值，以及每個 Bin 內的記錄數。僅對於最佳化 Binning 方法，每個 Bin 中的記錄數顯示為總數的百分比。請注意，採用排定分級 Binning 方法時，臨界值不適用。

**讀取值。** 從資料集中讀取 Bin 值。請注意，當新資料通過串流時，臨界值也將被改寫。

## 產生一個衍生節點

可以根據目前臨界值使用「產生」功能表建立 Derive 節點。將已建立的一組資料的 Bin 臨界值套用於另一組資料時，此操作十分有用。此外，如果這些分割點已知，那麼使用大型資料集時衍生作業比 Binning 作業更為高效（即更迅速）。

---

## RFM 分析節點

「回頭率、頻率及消費金額 (RFM) 分析」節點可讓您在數量上判定哪些客戶有可能是最佳客戶，方法是檢查距離客戶上次向您購買的時間（回頭率）、客戶多長時間購買一次（頻率），以及客戶在所有交易中的花費金額（消費金額）。

RFM 分析的原理是曾經購買過產品或服務的客戶更有可能會再次進行購買。分類的客戶資料會分為多個 Bin，其中 Bin 準則可根據您的需要進行調整。在每個 Bin 中，會分配給客戶一個分數；然後將這些分數已結合在一起，從而得到 RFM 的總分數。此分數代表為每個 RFM 參數建立各 Bin 中的客戶會員資格。這種已分級的資料可以充分滿足您的需求，例如識別購買頻率最高的高價值客戶；另外，它還可以在串流中進行傳送以便進一步建模和分析。

不過需要注意的是，儘管分析 RFM 分數並對這些分數進行排等級的功能非常有用，但使用時還必須注意一些特定的因素。可以對排名最高的目標客戶進行一些促銷活動；但過度誘導這些客戶可能會適得其反，導致他們在重複的交易過程中出現反感或不進行購買。另外，我們還需要牢記：不應忽視分數低的客戶，因為他們經過培養可能會成為更好的客戶。相反，根據市場反饋，僅具有高分值的客戶不一定能帶來好的預期銷售業績。例

如，近因中分級為 5 的客戶（即最近購買過產品或服務的客戶）對有些銷售人員（如銷售汽車或電視等昂貴且使用期較長的產品的人員）來說可能並不是真正的最佳目標客戶。

注意：根據資料儲存的方式，可能需要在「RFM 分析」節點之前先使用「RFM 整合」節點將資料轉換為可用格式。例如，輸入資料必須是客戶格式，一列一個客戶；如果客戶的資料是交易處理資料，請使用「RFM 整合」節點在上游衍生近因、頻率和貨幣欄位。請參閱第 70 頁的『RFM 「聚合」節點』主題，以取得更多資訊。

將 IBM SPSS Modeler 中的 RFM 「聚合」節點和 RFM 分析節點設定為使用獨立的分級；即，它們分別按近因、頻率、貨幣值對資料進行排等級和分級，而無需考慮它們的值或其他兩種標準。

## RFM 分析節點設定

**近因。** 使用欄位選擇器（文字框右側的按鈕）選擇近因欄位。這可以是日期、時間戳記或簡單數字。請注意，如果日期或時間戳記代表的是最近交易的日期，則將最高值視為最近；如果指定一個數字，它會代表自最近交易以來過去的時間，並將最低值視為最近。

注意：如果串流中「RFM 整合」節點的位置在「RFM 分析」節點之前，那麼應將「RFM 整合」節點產生的近因、頻率和貨幣欄位選作「RFM 分析」節點的輸入。

**次數。** 使用欄位選擇器選擇要使用的頻率欄位。

**貨幣。** 使用欄位選擇器選擇要使用的貨幣欄位。

**分組數目。** 為三種輸出類型分別選取要建立的 Bin 個數。預設值為 5。

注意：Bin 的最小值為 2，最大值為 9。

**加權。** 依預設，計算分值時會將近因資料的重要性視為最高，其次是頻率，最後是貨幣。如果需要，可以修正影響上述一個或多個字段的加權，來變更重要性級別。

RFM 分數的計算方法如下： $(\text{近因分值} \times \text{近因加權}) + (\text{頻率分值} \times \text{次數加權}) + (\text{貨幣分值} \times \text{貨幣加權})$ 。

**鏈結。** 指定如何分級相同的分數。選項是：

- **新增至下一個。** 選取此項以將鏈結值向上移至下一個 Bin。
- **保留在現行中。** 將鏈結值保留在現行（下方）Bin 中。此方法可能會導致建立的 Bin 總數較少。（這是預設值。）

**分組臨界值。** 指定在執行節點時是始終重新計算 RFM 分值和分級配置，還是僅在需要時進行計算（如在新增了新資料時）。如果選取**如果可用，從「Bin 值」標籤讀取**，則可以在「Bin 值」標籤上編輯不同 Bin 的上、下割點。

執行時，RFM Bin 節點可 Bin 原始近因、頻率和貨幣欄位，並將下列新欄位新增到資料集：

- 最近購買日分數。近因排等級（分級值）
- 頻率分數。頻率排等級（分級值）
- 消費金額分數。貨幣排等級（分級值）
- RFM 分數。近因、頻率和貨幣分數的加權和。

**將離群值新增到最終 Bin。** 如果選中此勾選框，那麼可將位於較低分級下方的記錄新增到較低的分級中，同時將最高分級上方的記錄新增到最高分級中；否則，會將空值分配給這些記錄。只有選取**如果可用，從「Bin 值」標籤讀取**時，才可使用此複選框。



## RFM 分析節點分級

通過「Bin 值」標籤，您可以檢視並在某些情況下修正產生的 Bin 的臨界值。

注意：如果在「設定」標籤中選中了**如果可用**，從「Bin 值」標籤讀取，那麼只能在此標籤上修正值。

**Binned 欄位**。使用下拉清單選取要 Bin 的欄位。可用的值是「設定」標籤上選定的值。

**Bin 值表格**。此處顯示每個產生的 Bin 的臨界值。如果在「設定」標籤上選中了**如果可用**，從「Bin 值」標籤讀取，則可以通過按兩下相應 Cell 修正每個分級的上、下割點。

**讀取值**。從資料集中讀取已 Bin 值並填寫 Bin 值表格。請注意，如果在「設定」標籤上選中了**永遠要重新計算**，則當新資料通過串流時，Bin 臨界值也將被改寫。

---

## 總體節點

「總體」節點可結合使用兩個或兩個以上模型塊，這樣所獲得的預測會比通過任意一個模型獲得的預測更為準確。通過結合多個模型的預測，可以避免單個模型的局限性，從而使整體精確度更高。一般情況下，以這種方式已結合的模型所得的結果不但可以與使用單個模型所得的最佳結果相媲美，而且結果通常會更理想。

這種節點結合在「自動分類器」、「自動數值」和「自動叢集」自動建模節點中自動產生。

使用「總體」節點後，可以通過「分析」節點或「評估」節點將這些綜合結果的精確度與每個輸入模型進行比較。要執行此操作，請確保未選取總體節點中「設定」標籤上的**過濾出總體模型產生的欄位**選項。

### 輸出欄位

每個總體節點都可以產生含有綜合分數的欄位。該欄位名稱基於指定的目標欄位，其字首根據具體的欄位測量層次（旗標、列名（集合）或連續（範圍））可以為 \$XF\_、\$XS\_ 或 \$XR\_。例如，若目標欄位是一個名為 *response* 的旗標欄位，則輸出欄位將為 \$XF\_response。

**信賴度或偏好欄位**。對於旗標和列名欄位，附加的信賴度或偏好欄位將根據總體方法進行建立，詳情如下表格所示。

表 28. 全體方法欄位建立.

總體方法	欄位名稱(F)
投票 信賴度加權投票 原始傾向加權投票調整傾向加權投票最高獲勝信賴度	\$XFC_<field>
原始傾向平均值	\$XFRP_<field>
平均值調整的原始傾向	\$XFAP_<field>

## 總體節點設定

**整合的目標欄位**。選取單一欄位以用作兩個或兩個以上游模型的目標欄位。上游模型可以使用旗標、列名或連續目標字段，但其中至少要有兩個模型必須共用同一目標字段以便獲得綜合分數。

**過濾掉組合模型所產生的欄位**。從輸出中移除由個別模型（為組合節點提供資訊來源）產生的所有其他欄位。如果您僅對所有輸入模型中的結合評分感興趣，請選取這個勾選框。確保在下列情況下取消選取此選項：例如，如果您要使用「分析」節點或「評估」節點來比較結合評分的準確性與每一個個別輸入模型的評分準確性。

可用設定取決於選作目標欄位的欄位的測量層次。

### 連續目標

對於連續目標，將會對其分數求平均值。這是求綜合分數唯一可用的方法。

求分數或估計值的平均值時，「總體」節點使用標準誤計算得出測量值或估計值與 true 值之間的差分，並顯示這些估計值的接近程度。預設對新模型產生標準誤差計算；但是，您可以為現有模型取消選取勾選框，例如當要對其重新產生時。

### 種類目標

種類目標可受支援包含 **投票** 在內的許多方法，其工作原理是計算每種可能的預測值的選取次數並選取總數最高的值。例如，若五個模型中有三個都預測為 **是**，其他兩個預測為 **否**，則是以 3 比 2 的投票結果取勝。另外也可以根據每個預測的信賴度或傾向值進行 **加權** 投票。然後對加權求和，並再次選取總數最高的值。最終預測的信賴度是指取勝值的加權總和除以總體模型中包含的模型數量。

所有種類欄位。旗標和列名欄位均受支援下列方法：

- 投票
- 信賴度加權投票
- 最高獲勝信賴度

僅旗標欄位。對於僅限旗標欄位的情況，還支持以下多種基於傾向的方法：

- 採用原始傾向加權的原始投票
- 採用調整後傾向加權的投票
- 原始傾向平均值
- 平均調整後傾向

投票同數。根據投票方法，可以指定解析投票同數的方法。

- **隨機選擇**。隨機選擇其中一個同分值。
- **最高信賴度**。選擇使用最高信賴度進行預測的同分值。請注意，該信賴度值無需與所有預測項值的最高信賴度值相同。
- **原始或已調整的傾向（僅限旗標欄位）**。使用最大絕對傾向預測的同分值，其中絕對傾向的計算方法如下：

$\text{abs}(0.5 - \text{propensity}) * 2$

或者，對於已調整的傾向，絕對傾向計算方法如下：

$\text{abs}(0.5 - \text{adjusted propensity}) * 2$

---

## 分割區節點

「分割區」節點用於產生分割區欄位，將資料分割為個別的子集或樣本，以供模型建置的訓練、正在測試和驗證階段使用。通過用某個樣本產生模型並用另一個樣本對模型進行測試，可以預判此模型對類似於現行資料的大型資料集的擬合優劣。

分割區節點會產生列名欄位，其角色設定為**分割區**。此外，如果資料中已經存在適當的欄位，可以使用「類型」節點將其指定為分割區。在這種情況下，不需要個別的「分割區」節點。可以將任何具有兩個或三個值的實例化列名欄位用作分割區，但不能使用旗標欄位。請參閱第 127 頁的『設定欄位角色』主題，以取得更多資訊。

可以在一個串流中定義多個分割區欄位，但如果這麼做，那麼必須在每個用於分區的建模節點的「欄位」標籤中選取一個分割區欄位。（如果僅存在一個分割區，則每當啟用分隔時，都會自動使用該分割區。）

**啟用分區。**要在分析中使用分區，必須在適當的模型建立或分析節點的「模型選項」標籤中啟用分區。取消選取此選項可以在不刪除欄位的條件下取消分區功能。

要基於其他準則（如數據範圍或位置）建立分割區欄位，還可以使用「衍生」節點。請參閱第 131 頁的『衍生節點』主題，以取得更多資訊。

**範例。**建立 RFM 串流以識別積極回應以往營銷活動的最新客戶時，銷售公司的市場行銷處可以使用「分割區」節點將資料分割到訓練分割區和測試分割區。

## 分割區節點選項

**分割區欄位。**指定由該節點建立的欄位的名稱。

**分割區。**可以將資料分割為兩個樣本（訓練與測試）或三個樣本（訓練、測試和驗證）。

- **訓練與測試。**將資料分割為兩個樣本，使您能夠用一個樣本訓練模型並用另一個樣本測試模型。
- **訓練、測試和驗證。**將資料分割為三個樣本，使您能夠用一個樣本訓練模型，用第二個樣本測試並精練模型，然後用第三個樣本驗證得到的結果。這種方式會相應減小每個分割區的大小，但在使用超大型資料集時最為適用。

**分割區大小。**指定每個分割區的相對大小。如果分割區大小之和少於 100%，那麼未包含在分割區中的記錄將被捨棄。例如，如果使用者擁有一千萬條記錄，並已指定 5% 的訓練分割區大小和 10% 的正在測試分割區大小，那麼在執行該節點之後，大約會有五十萬條訓練記錄和一百萬條正在測試記錄，其餘記錄則被捨棄。

**值。**指定用於代表資料中每個分割區樣本的值。

- **使用系統定義的值 ("1"、"2" 和 "3")。**使用整數代表每個分割區；例如，位於訓練樣本中的所有記錄的分割區欄位值均為 1。這樣可確保資料能夠在不同語言環境之間移動，而且如果分割區欄位在其他位置進行重新實例化（例如從資料集讀取回資料），將保留排序（因此 1 仍將代表訓練分割區）。但是，這種值需要一定的解釋。
- **向系統定義的值附加標籤。**將整數與標籤組合；例如，訓練分割區記錄的值為 1\_Training。這樣，查看資料的人可能識別出具體的值，並且資料可以保留排序。但是，這種值僅適用於給定的語言環境。
- **把標籤當作值使用。**使用不帶整數的標籤；例如，**Training**。這使您能夠通過編輯標籤來指定值。但是，這也使資料特定於語言環境，而分割區欄的重新實例化會使值具有自然排序，而不對應其「語義」順序。

**種子。**僅在選取可重複的分割區指派的情況下可用。當取樣或分割根據隨機百分比來記錄時，此選項可讓您將相同的結果複製到其他階段作業中。透過指定亂數產生器所用的起始值，您可以確保每次執行節點時都指派相同的記錄。輸入所需的種子值，或按一下產生按鈕來自動產生一個隨機值。如果未選取此選項，則每次執行節點時，都將產生不同樣本。

**註：**對從資料庫中讀取的記錄使用種子選項時，可能需要在取樣前使用「排序」節點以確保每次執行節點時都獲得相同的結果。這是因為隨機種子取決於記錄的順序，而並不能保證順序在關聯式資料庫中保持相同。請參閱第 72 頁的『排序節點』主題，以取得更多資訊。

**使用唯一欄位來分配分割區。**僅在選取可重複的分割區指派的情況下可用。（僅適合第 1 層級資料庫）勾選此方框以使用 SQL 回送分配記錄到分割區。從下拉清單中，選擇具有唯一值的欄位（例如 ID 欄位）以確保以隨機且可重複的方式分配記錄。

「資料庫來源」節點的說明中對資料庫分層級進行了說明。請參閱第 15 頁的『資料庫來源節點』主題，以取得更多資訊。

## 產生選取節點

使用「分割區」節點中的「產生」功能表，可以自動為每個分割區產生一個「選取」節點。例如，可以選取訓練分割區中的所有記錄，以便僅使用此分割區獲得進一步的求值或分析。

---

## 設為旗標節點

「設為旗標」節點用於根據為一個或多個集合欄位定義的種類值衍生旗標欄位。例如，您的資料集可能包含一個列名欄位 *BP*（血壓），其值為 *High*、*Normal* 和 *Low*。為簡化資料操作，可以建立一個代表高血壓的旗標欄位，用於指示病患是否患有高血壓。

### 為設為旗標節點設定選項

**集合欄位。**列出測量層次為列名（集合）的所有資料欄位。從清單中選取一個字段，以顯示集合中的值。您可以在這些值中進行選擇，以建立旗標欄位。請注意，必須先使用上游來源節點或「類型」節點對資料進行完全實例化，然後才能請參閱可用的列名欄位（及其值）。請參閱第 117 頁的『類型節點』主題，以取得更多資訊。

**欄位名稱延伸。**選取此選項將啟用用於指定副檔名的控制項，該副檔名將作為字尾或字首新增到新的旗標欄位。依預設，會透過將原始欄位名稱與欄位值合併到某個標籤來自動建立新欄位名稱，例如 *Fieldname\_fieldvalue*。

**可用集合值。**在上方選取的集中的值會在這裡顯示。選取要為其產生旗標的一個或多個值。例如，如果名為 *blood\_pressure* 的欄位中的值為 *High*、*Medium* 和 *Low*，則可以選取 *High* 並將其新增到右側的清單中。此操作會為具有表示高血壓的值的記錄建立一個帶標籤欄位。

**建立旗標欄位。**此處列出新建立的旗標欄位。可以指定使用欄位名稱延伸控制項命名新欄位的選項。

**True 值。**指定設定旗標時節點所用的 *true* 值。預設情況下，此值為 **T**。

**False 值。**指定設定旗標時節點所用的 *false* 值。預設情況下，此值為 **F**。

**整合鍵值。**選取此選項將根據下方指定的索引鍵欄位對記錄進行群組。已選取**按鍵值聚合欄位**時，只要任何記錄被設為 *true*，便會「開啟」群組中的所有旗標欄位。使用欄位選擇器可指定將用於聚合記錄的索引鍵欄位。

---

## 重新結構化節點

「重組」節點可用於根據列名欄位或旗標欄位的值產生多個欄位。新產生的欄位可包含來自另一個欄位或數值型旗標（0 和 1）的值。此節點的功能與「設為旗標」節點類似，但更加靈活。使用這種節點，可以使用另一個欄位的值建立任意類型的欄位（包括數值型旗標）。隨後，您可以對其他下游節點執行聚集或其他操作。

（設為旗標節點允許您在一個步驟中聚合欄位，因此如果要建立旗標欄位，使用設為旗標節點更為方便。）

例如，下列資料集包含一個列名欄位 *Account*，該欄位的值為 *Savings* 和 *Draft*。每個帳戶均記錄了期初餘額和目前餘額，而且有些客戶在每種類型中均有多个帳戶。假設您希望瞭解每個客戶是否擁有特定的帳戶類型，如果有，每種帳戶類型中有多少資金。可以使用重新結構化節點為每個 *Account* 值產生一個欄位，並選取 *Current\_Balance* 作為值。這樣會用給定記錄的目前餘額移入每個新欄位。

表 29. 在重組之前取樣資料。

CustID	Account	Open_Bal	Current_Bal
12701	Draft	1000	1005.32
12702	Savings	100	144.51
12703	Savings	300	321.20

表 29. 在重組之前取樣資料 (繼續).

CustID	Account	Open_Bal	Current_Bal
12703	Savings	150	204.51
12703	Draft	1200	586.32

表 30. 在重組之後取樣資料.

CustID	Account	Open_Bal	Current_Bal	Account_Draft_Current_Bal	Account_Savings_Current_Bal
12701	Draft	1000	1005.32	1005.32	\$null\$
12702	Savings	100	144.51	\$null\$	144.51
12703	Savings	300	321.20	\$null\$	321.20
12703	Savings	150	204.51	\$null\$	204.51
12703	Draft	1200	586.32	586.32	\$null\$

將重新結構化節點與「聚合」節點一起使用

在多數情況下，可能需要將「重組」節點與「聚合」節點配對使用。在上一個範例中，一個客戶 (ID 為 12703) 有三個帳戶。可以使用「聚合」節點計算每種帳戶類型的總餘額。索引鍵欄位為 *CustID*，且聚集欄位是重新架構欄位 *Account\_Draft\_Current\_Bal* 和 *Account\_Savings\_Current\_Bal*。下表顯示了結果。

表 31. 在重組和聚集之後取樣資料.

CustID	Record_Count	Account_Draft_Current_Bal_Sum	Account_Savings_Current_Bal_Sum
12701	1	1005.32	\$null\$
12702	1	\$null\$	144.51
12703	3	586.32	525.71

## 為重新結構化節點設定選項

**可用的欄位。** 列出測量層次為列名 (集合) 或旗標的所有資料欄位。從清單中選擇一個欄位，以顯示集合 (或旗標) 中的值；隨後在這些值中進行選擇，以建立重組欄位。請注意，必須先使用上游來源節點或「類型」節點對資料進行完全實例化，然後才能請參閱可用的欄位 (及其值)。請參閱第 117 頁的『類型節點』主題，以取得更多資訊。

**可用的值。** 在上方選取的集中的值會在這裡顯示。選取要產生重新架構欄位的一個或多個值。例如，如果名為 *Blood Pressure* 的欄位中的值為 *High*、*Medium* 和 *Low*，您可以選取 *High* 並將其新增到右側的清單中。此操作會為值為 *High* 的記錄建立一個具有指定值的欄位 (請參閱下方文)。

**建立重組的欄位。** 此處列出新建立的重組欄位。依預設，會透過將原始欄位名稱與欄位值合併到某個標籤來自動建立新欄位名稱，例如 *Fieldname\_fieldvalue*。

**包含欄位名稱。** 取消選取此選項將禁止將原始欄位名稱用作新的欄位名稱的字首。

**使用來自其他欄位的值。** 指定一個或多個欄位，其值將用於移入重組的欄位。使用欄位選擇器選擇一個或多個欄位。對於選擇的每個欄位，都會建立一個新欄位。值欄位名稱將附加到重組的欄位名稱之後；例如，*BP\_High\_Age* 或 *BP\_Low\_Age*。每個新欄位都會繼承原始值欄位的類型。

**建立數值旗標。** 選取此選項可使用數值旗標 (0 表示假，1 表示 true) 移入新欄位，而不是使用另一個欄位的值。

---

## 轉置節點

依預設，欄為欄位，而列為記錄或觀察。如有必要，可使用轉置節點交換列與欄中的資料，使欄位變為記錄、記錄變為欄位。例如，如果有時間序列資料，其中每個數列均為一列而不是一欄，則可以在分析之前轉置資料。

### 為「轉置」節點設定選項

在轉置方法下拉清單中，選取想要「轉置」節點執行的方法：欄位和記錄、記錄到欄位或欄位至記錄。下列部分中說明了所有這三種方法的設定。

限制：僅在 Windows 64 位元、Linux 64 位元和 Mac 上受支援記錄到欄位和欄位至記錄方法。

### 欄位和記錄

可以根據指定的字首自動產生新的欄位名稱，也可以從資料中的現有欄位讀取新的欄位名稱。

**使用字首。**此選項將根據指定的字首 (Field1 和 Field2 等) 自動產生新的欄位名稱。您可以根據需要自訂字首。如果使用此選項，那麼必須指定要建立的欄位個數，而與原始資料中的列個數無關。例如，如果新欄位個數設為 100，前 100 列以外的所有資料都將被捨棄。如果原始資料中的列數少於 100，那麼有些欄位將為無效。(可以根據需要增加欄位數，但此設定的目的是避免將一百萬條記錄轉置為一百萬個欄位，因為這會產生無法管理的結果。)

例如，假定資料中包含按列顯示的數列，並且每個月有一個個別的欄位 (欄)。您可以轉置此數據，使每個數列包含在一個個別的欄位中，每一列表示一個月。

**從欄位讀取。**從現有欄位讀取欄位名稱。使用此選項，新欄位數將由資料決定，最多可達到指定的最大值。選定欄位的每個值都變為輸出資料中的一個新欄位。選定的欄位可具有任何儲存類型 (整數、字串、日期等)，但為避免出現重複的欄位名稱，選定欄位的每個值都必須唯一 (換言之，值的號碼應與列數相符)。如果遇到重複的欄位名稱，將顯示警告消息。

- **讀取值。** 如果選定的欄位尚未實例化，那麼選取此選項將移入新的欄位名稱的清單。如果欄位已實例化，那麼不必執行此步驟。
- **要讀取的值的最大數量。** 從資料中讀取欄位名稱時，需要指定上限以避免建立過多的欄位。(如上方所述，將一百萬條記錄轉置為一百萬個欄位會產生無法管理的結果。)

例如，如果資料中的第一欄指定了每個數列的名稱，您可以將這些值用作調換資料中的欄位名稱。

**轉置。**預設情況下，僅對連續 (數值型範圍) 欄位 (儲存類型為整數或實數) 進行已轉置。另外，您也可以選擇數字欄位的子集合，或改為轉置字串欄位。但是，所有已轉置的欄位都必須具有相同的儲存類型，可以是數值或字串，但不能同時是這兩者，因為混合輸入欄位會在每個輸出欄中產生混合的值，而這會違規一個欄位中的所有值必須具有相同儲存類型的原則。其他儲存類型 (日期、時間、時間戳記) 不能進行已轉置。

- **所有數值。** 轉置所有數字欄位 (儲存類型為整數或實數)。輸出中的列個數必須與原始資料中的數字欄位數相符。
- **所有字串。** 轉置所有字串欄位。
- **自訂。** 容許您選取數字欄位的子集合。輸出中的列數必須與所選的欄位個數相符。該選項僅適用於數值欄位。

**橫列 ID 名稱。** 指定由節點建立的列 ID 欄位的名稱。此欄位的值由原始資料中的欄位名稱確定。

**提示：**將時間序列資料從列轉置為欄時，如果原始資料中有一列 (如日期、月或年) 帶有每個測量期間的標籤，請確保將這些標籤作為欄位名稱讀取入 IBM SPSS Modeler (如前面的範例所示，將原始資料中的月或日

期分別顯示為欄位名稱），而不是在第一列資料中包含標籤。這樣將避免在每一欄中混合標籤與值（這會將數字強制讀取為字串，因為一欄中不能混合儲存類型）。

## 記錄至欄位

**欄位。** 「欄位」清單包含進入「轉置」節點的所有欄位。

**索引。** 使用「索引」區段可選取您想要用作索引欄位的欄位。

**欄位。** 使用「欄位」部分來選取要用作欄位的欄位。

**值。** 使用「值」區段可選取您想要用作值欄位的欄位。

**聚合函數。** 如果某指標對應多個記錄，必須將這些記錄聚合為一條記錄。使用聚合函數下拉清單來指定如何使用下列某個函數來聚合記錄。請注意，聚集影響所有欄位。

- **平均數。** 傳回每個索引鍵欄位組合的平均數。平均數是集中趨勢的量數，是算數平均值（總和除以觀察值數）。
- **總和。** 傳回每個索引鍵欄位組合的合計值。總和是遍及所有包含非遺漏值之觀察值的值總計。
- **最小。** 傳回每個索引鍵欄位組合的最小值。
- **最大。** 傳回每個索引鍵欄位組合的最大值。
- **中位數。** 傳回每個索引鍵欄位組合的中值。中位數為集中趨勢的量數，其不會感應到偏離值（相反地，平均數會受到幾個高低極端值的影響）。其亦稱作第 50 個百分位數或第 2 個四分位數。
- **計數。** 傳回每個索引鍵欄位組合的非空值計數。

## 欄位至記錄

**欄位。** 「欄位」清單包含進入「轉置」節點的所有欄位。

**索引。** 使用「索引」區段可選取您想要用作索引欄位的欄位。

**值。** 使用「值」區段可選取您想要用作值欄位的欄位。如果不選取任何數值欄位，那麼所有未分配的數字欄位將用作值。但如果數字欄位不可用，那麼將使用所有未分配的字串欄位。

---

## 歷史節點

「歷程」節點通常用於序列資料，例如時間序列資料。這種節點用於建立包含先前記錄中欄位的資料的新欄位。使用「歷史」節點時，可能需要使用按特定欄位預先排序的資料。可以使用排序節點執行此操作。

### 為歷史節點設定選項

**選取的欄位。** 使用「欄位選擇器」（文字框右邊的按鈕）選擇需要使用其歷史的欄位。每個選取的欄位將用於建立資料集中所有記錄的新欄位。

**偏移。** 指定要從中擷取歷程欄位值的現行記錄之前的最新的記錄。例如，如果「偏移量」設為 3，則當每條記錄通過此節點時，之前第三條記錄的欄位值將併入在現行記錄中。使用「展開範圍」設定可指定從中擷取記錄的記錄後退展開範圍。使用箭頭可調整偏移值。

**展開範圍。** 指定要從中擷取值的以前記錄的數目。例如，如果「偏移量」設為 3 且「展開範圍」設為 5，那麼通過該節點的每條記錄將針對「所選取的欄位」清單中指定的每個欄位新增五個欄位。這表示，當節點正在處理記錄 10 時，將從記錄 7 至記錄 3 新增欄位。使用箭頭可調整展開範圍值。

**歷史不提供時。** 選取下列其中一個選項以用於處理沒有歷史值的記錄。這通常是指資料集頂端的前幾條記錄，它們沒有可用作歷史的先前記錄。

- **捨棄記錄。** 選取此選項將捨棄對於所選欄位沒有可用歷史值的記錄。
- **未定義歷史記錄。** 選取此選項將保留沒有可用歷史值的記錄。歷史欄位將填入未定義值，顯示為 \$null\$。
- **填入值。** 指定要用於沒有可用歷史值的記錄的值或字串。預設的取代值為系統空值 *undef*。使用字串 \$null\$ 顯示空值。

為實現正確執行，在選取取代值時請記住下列規則：

- 選取的欄位應的同一儲存類型。
- 如果所有所選取的欄位的儲存類型均為數值，取代值必須剖析為整數。
- 如果所有所選取的欄位的儲存類型均為實數，取代值必須剖析為實數。
- 如果所有所選取的欄位的儲存類型均為符號，取代值必須剖析為字串。
- 如果所有所選取的欄位的儲存類型均為日期/時間，取代值必須剖析為日期/時間欄位。

如果上述任一條件不成立，則會在執行歷史節點時收到錯誤警告。

---

## 欄位重新排序節點

使用「欄位重新排序」節點，可以定義用於顯示下游欄位的自然順序。此順序會影響欄位在各個位置（例如表格、清單與「欄位」選擇器）中的顯示方式。例如，使用大型資料集時，此作業有助於使所需欄位更直觀。

### 設定欄位欄位重新排序選項

對欄位進行重新排序的方法有兩種：自訂排序和自動排序。

#### 自訂排序

選取自訂順序可啟用一個包含欄位名稱和類型的表格，您可在其中檢視所有欄位並使用箭頭按鈕建立自訂順序。

要對欄位進行重新排序，請執行下列操作：

1. 選取表格中的某個欄位。採用按住 **Ctrl** 並按一下的方法可選取多個欄位。
2. 使用簡單的箭頭按鈕可將欄位上移或下移一列。
3. 使用列箭頭按鈕可將欄位移至清單底部或頂部。
4. 通過將分隔線（標為 [其他欄位]）上移或下移，指定此處不包含的欄位的順序。

有關 [其他欄位] 的詳細資訊

**其他欄位。** [其他欄位] 分割線的用途是將表格分為兩半。

- 顯示在分隔線以上方的欄位（以它們出現在表格中的順序）排序後，將在此節點下游欄位的所有自然順序的頂端顯示。
- 顯示在分隔線以下方的欄位（以它們出現在表格中的順序）排序後，將在此節點下方遊欄位的所有自然順序的底端顯示。



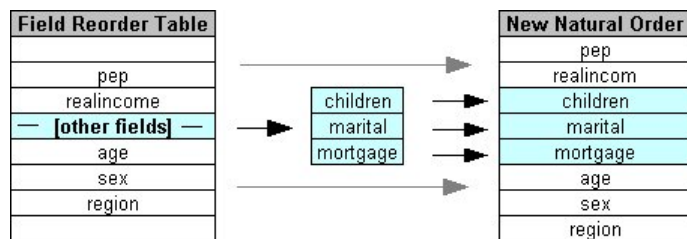


圖 6. 說明「其他欄位」在新欄位順序中的結合方式的圖

- 未出現在欄位重新排序表格中的所有其他欄位將在這些「頂端」和「底端」欄位之間顯示，按分隔線的位置標示。

其他自訂排序選項包含：

- 通過按一下每個欄標題（**類型、名稱和儲存類型**）上方的箭頭可按遞增或遞減對欄位進行排序。按欄排序時，未在此指定的欄位（按 [其他欄位] 列標識）將排在其自然順序的最後一欄。
- 按一下清除 **UNUSED** 的可將所有 UNUSED 的欄位從欄位重新排序節點中刪除。UNUSED 的欄位在表格中以紅色字型顯示。這表示該欄位已在上游作業中被刪除。
- 指定任意新欄位（顯示有閃電圖示，表示新的或未指定的欄位）的排序方式。按一下**確定**或**套用**時，該圖示將消失。

注意：如果應用自訂序列後在上游新增了欄位，那麼會將新欄位附加到自訂清單的底部。

#### 自動排序

選取**自動排序**以指定排序參數。對話框選項將動態變化，以提供用於自動排序的選項。

**排序依據**。選取三種方式之一，對讀取入「欄位重新排序」節點的欄位進行排序。箭頭按鈕將指示順序為遞增還是遞減。選取一種以進行變更。

- 名稱
- 類型
- 儲存體

套用自動排序後在欄位重新排序節點上游新增的欄位將根據所選排序類型被自動置於適當位置。

## 時間間隔節點

SPSS Modeler 17.1 版和更早版本中的原始「時間間隔」節點與 Analytic Server (AS) 不相容，並在 SPSS Modeler 版本 18.0 中已淘汰。

取代「時間間隔」節點包含對原始「時間間隔」節點的一系列變更。該新節點可以與 Analytic Server 一起使用，也可以由 SPSS Modeler 單獨使用。

「時間間隔」節點用於指定時間間隔並衍生用於估算或預測的新時間欄位。支援從秒到年的全系列時間區間。

使用此節點可以衍生新的時間欄位；新欄位的儲存類型與您選擇的輸入時間欄位相同。此節點將產生下列各項目：

- 在「欄位」標籤上指定為**時間欄位**的欄位以及選擇的字首/字尾。依預設，字首為 \$TI\_。
- 在「欄位」標籤上指定為**維度欄位**的欄位。
- 在「欄位」標籤上指定為**要聚合的欄位**的欄位。

另外，還可以產生若干其他欄位，具體取決於選取的時間間隔或期間（例如測量值所處的分鐘或秒）。

## 時間間隔 - 欄位選項

使用「時間間隔」節點中的「欄位」標籤選取從中衍生新時間間隔的資料。

**欄位** 顯示此節點的所有輸入欄位及其測量類型圖示。所有時間欄位都具有「連續」測量類型。請選取要用作輸入的欄位。

**時間欄位** 顯示要從中衍生新時間間隔的輸入欄位；只容許使用單個連續欄位。「時間間隔」節點將此欄位作用於轉換時間間隔的聚集鍵。新欄位的儲存類型與選擇的輸入時間欄位相同。如果您選取了整數欄位，那麼該欄位將被視為時間指標。

**維度欄位** 您可以選擇性地在此處新增欄位，以便建立單個基於欄位值的時間序列。舉一個簡單的例子，對於地理空間資料，可以使用點欄位作為維度。在此範例中，來自「時間間隔」節點的資料輸出排序成點欄位中每個點值的時間序列。

當您使用平面化多維度資料（類似於 TM1 節點產生的資料）或者為了支援類似地理空間之類的更加複合的資料類型時，維度是理想選擇。從本質上來說，您可以考慮使用維度欄位作為 SQL 查詢中的 **Group By** 子句的等效項，或者類似於「聚合」節點中的索引鍵欄位；但是，維度欄位在本質上更加複雜，因為它能夠處理比只有傳統列和直欄資料更加複雜的資料結構。

**要聚合的欄位** 選取在更改時間欄位的时间段過程中要聚合的欄位。只有此處選取的欄位才可用於「建立」標籤上的指定欄位的自訂設定表格。在離開此節點的資料中，將過濾掉所有未在此處併入的欄位。這意味著，將從該資料中過濾掉欄位清單中的所有餘下欄位。

## 時間間隔 - 建置選項

使用「建立」標籤可以指定用於變更時間間隔的選項，以及指定如何根據資料的測量類型對其中的欄位進行聚合。

對資料進行聚合時，現有的所有日期、時間或時間戳記欄位都將被產生的欄位取代，並從輸出中刪除。其他欄位將根據此標籤中指定的選項進行聚合。

**時間間隔** 請選取用於建立數列的時間間隔和週期性長度。

**預設值** 請選取要套用至不同類型的資料的預設聚集。預設聚集將根據測量層次進行套用；例如，連續欄位按合計進行聚合，而名義欄位則使用眾數。您可以為三種不同的測量層次設定預設彙總：

- **連續** 可用於連續欄位的函數包含合計、平均數、下限、上限值、中位數、第 1 個四分位數和第 3 個四分位數。
- **名義** 選項包含眾數、下限和上限值。
- **旗標** 選項為如果任意一項為 **true** 則為 **true** 和如果任意一項為 **false** 則為 **false**。

**指定欄位的自訂設定** 您可以指定各個欄位的預設聚集設定的例外狀況。請使用右側的圖示在表格中新增或移除欄位，或者按一下相應欄中的 Cell 以變更用於該欄位的聚集函數。無類型欄位已從清單中排除，無法新增至表格。

**新欄位名稱延伸** 請指定要對此節點所產生的所有欄位套用的字首或字尾。

---

## 「重新投影」節點

對於地理空間資料或地圖資料，最一般的兩種指示座標的方法是投影座標系統和地理座標系統。在 IBM SPSS Modeler 中，表示式建置器空間函數、「空間-時間預測」(STP) 節點和「地圖視覺化」節點之類的项目使用投影座標系統，因此必須對匯入的所有使用地理座標系統進行記錄的資料執行重新投影。在有可能的情況下，將在使用（而不是匯入）地理空間欄位（任何具有地理空間測量層次的欄位）時自動對其進行重新投影。存在任何無法自動進行重新投影的欄位時，您可以使用「重新投影」節點來變更其座標系統。以此方式進行重新投影意味著您可以解決由於使用了不正確的座標系統而出錯的狀況。

下列清單顯示了可能必須進行重新投影以變更座標系統的範例狀況：

- 附加 如果您嘗試附加兩個將不同座標系統用於某個地理空間欄位的資料集，那麼 SPSS Modeler 將顯示下列錯誤訊息：<Field1> 和 <Field2> 的座標系統不相容。請將其中某個或全部欄位重新投影到同一個座標系統。  
  
<Field1> 和 <Field2> 是引起錯誤的地理空間欄位的名稱。
- If/else 表示式 在使用的表示式中，如果該表示式的兩個部分都包含帶有地理空間欄位或傳回類型的 if/else 陳述式，但是這些欄位或傳回類型採用不同的座標系統，那麼 SPSS Modeler 將顯示下列錯誤訊息：條件式表示式包含不相容的座標系統：<arg1> 和 <arg2>。  
  
<arg1> 和 <arg2> 是傳回採用不同座標系統的地理空間類型的 then 或 else 引數。
- 構造一系列地理空間欄位 要建立由許多地理空間欄位組成的清單欄位，向清單表示式提供的所有地理空間欄位引數必須採用同一個座標系統。否則，將顯示下列錯誤訊息：<Field1> 和 <Field2> 的座標系統不相容。請將其中某個或全部欄位重新投影到同一個座標系統。

如需座標系統的相關資訊，請參閱《SPSS Modeler 使用手冊》之「使用串流」小節中的「設定串流的地理空間選項」主題。

## 為「重新投影」節點設定選項

### 欄位

#### 地理空間欄位

依預設，此清單為空白。您可以將地理空間欄位從要重新投影的欄位清單移入此清單，以確保不對這些欄位進行重新投影。

#### 要重新投影的欄位

依預設，此清單包含所有作為此節點的輸入的地理空間欄位。此清單中的所有欄位都將重新投影到座標系統區域中設定的座標系統。

### 座標系統

#### 串流預設值

選取此選項表示使用預設座標系統。

**指定** 如果選擇此選項，那麼可以使用變更按鈕來顯示「選擇座標系統」對話框，並選擇要用於進行重新投影的座標系統。

如需座標系統的相關資訊，請參閱《SPSS Modeler 使用手冊》之「使用串流」小節中的「設定串流的地理空間選項」主題。



---

## 第 5 章 「圖表」節點

---

### 通用圖形節點功能

資料採礦過程的多個階段都會使用圖形和圖表探索導入到 IBM SPSS Modeler 中的資料。例如，可將「圖形」或「分佈」節點連接到資料來源，以瞭解資料類型和資料分佈。然後，可以執行記錄和欄位作業，為下游建模作業準備資料。圖形的另一個常見用途是檢查新衍生的欄位的分佈和它們之間的關係。

「圖形」選用區含有下列節點：



「圖版」節點可在一個節點中提供多數不同類型的圖形。使用此節點，可以選擇要探索的資料欄位，然後從適用於選定資料的欄位中選擇一個圖表。該節點會自動濾出不適用於欄位選項的任何圖形類型。



統計圖節點可顯示數值型欄位間的關係。可通過使用點（散點）或行建立散佈圖。



「分佈」節點顯示了標誌（種類）值（例如抵押類型或性別）的出現次數。通常，可以使用「分佈」節點來顯示資料中的不平衡度，然後在建立模型前使用「平衡」節點來糾正此類不平衡度。



「直方圖」節點顯示了數值型欄位的值的出現次數。它經常用來在資料操作和模型建置之前探索資料。與「分佈」節點相似，「直方圖」節點經常用來揭示資料中的不平衡度。



「收集」節點顯示一個數值欄位的值相對於另一個數值欄位的值的分配。（它建立類似於直方圖的圖形。）圖示說明值不斷變化的變數或欄位時，它是有益的。使用 3-D 製圖，您還可以包括符號軸，用來依種類顯示分佈項目。



使用多重繪圖節點可建立在一個 X 欄位上顯示多個 Y 欄位的統計圖。Y 欄位被繪製為彩色的行；每條行相當於「樣式」設定為行且「X 模式」設定為排序的統計圖節點。要研究幾個變數隨時間的變化情況時，多重繪圖非常有用。



Web 節點說明兩個或多個符號（種類）欄位值之間關係的強度。該圖使用不同粗細的行來表示關係強度。例如，您可以使用 Web 節點來研究電子商務網站上一系列商品的購買之間的關係。



「時間繪圖」節點顯示一個或多個時間序列資料集。通常情況下，您首先要使用時間間隔節點建立一個 *TimeLabel* 欄位，該欄位用於為 *x* 軸設置標籤。



「評估」節點有助於評估和比較預測模型。評估表示了模型對特定結果的預測優劣。它根據預測值和預測信賴度來對記錄進行排序。它將記錄分成若干個相同大小的群組（分位數），然後從高到底為每個分位數劃分業務準則值。多個模型在圖形中顯示為個別的行。



「地圖視覺化」節點可以接受多個輸入連線並在地圖上將地理空間資料顯示為一系列層。每一層都是單一地理空間欄位；例如，基本層可能是某個國家或地區的地圖，而在該層之上，可能有一層代表道路，一側代表河流，以及一層代表城鎮。



E-Plot（測試版）節點顯示數值欄位之間的關係。它與「繪圖」節點類似，但選項不同，且其輸出使用特定於此節點的新圖形繪製介面。使用測試版層次節點來與新圖形繪製功能搭配使用。



t-Distributed Stochastic Neighbor Embedding (t-SNE) 是一套用於視覺化高維度資料的工具。它會將資料點的親緣性轉換為可能性。此 t-SNE 節點在 SPSS Modeler 中使用 Python 進行實現並且需要 scikit-learn© Python 程式庫。

將圖表節點新增到串流後，可按兩下節點以開啟用於指定選項的對話框。絕大多數圖形都含有一些獨特的選項，這些選項會顯示在一個或多個標籤上。除此以外，還有若干通用於所有圖形的標籤選項。下列章節包含有關這些通用選項的進一步資訊。

配置圖表節點的選項後，可通過對話框執行該節點或將它作為串流的組成部分來執行。可在產生的圖表視窗中根據選擇或資料區域產生「衍生」（集合和旗標）和「選取」節點，有效地將資料劃分為多個「子集」。例如，您可使用這個強大的功能來識別以及排除離群值。

## 外觀、套版、畫面和動畫

### 套版和外觀

審美標準 (與重疊) 會將範圍新增到視覺化。審美標準 (群組、叢集或堆疊) 的效果根據視覺化類型、欄位 (變數) 類型, 以及圖形元素類型與統計而異。例如, 依色彩的類別欄位將散佈圖上的點分組, 或在堆疊長條圖中建立堆疊。色彩的連續數值範圍也可當作散佈圖中各點範圍值的指標。

您應該實驗各種審美標準與重疊, 找出符合您需求者。下列說明或許可協助您找到正確的選項。

附註: 並非所有視覺化類型都可以使用所有的審美標準及重疊。

- **顏色。** 當顏色已由某個類別欄位定義時, 其會根據個別的類別分割視覺化, 每種類別各一種顏色。當顏色是連續數值範圍時, 其會根據範圍欄位的值來變更顏色。若圖形元素 (例如長條圖或方框) 代表一個以上的記錄/觀察值, 並將範圍欄位用於顏色, 則顏色會依據範圍欄位的平均數改變。
- **形狀。** 形狀是由類別欄位所定義, 其可將視覺化分割成不同形狀的元素, 每種類別各一種形狀。
- **透明度。** 當透明度已由某個類別欄位定義時, 其會根據個別的類別分割視覺化, 每種類別各一種透明度層級。當透明度是連續數值範圍時, 其會根據範圍欄位的值來變更透明度。若圖形元素 (例如長條圖或方框) 代表一個以上的記錄/觀察值, 並將範圍欄位用於透明度, 則顏色會依據範圍欄位的平均數改變。在最大值時, 圖形元素會完全透明。在最小值時, 它們會完全不透明。
- **資料標籤。** 資料標籤是由任一類型的欄位所定義, 其值則用來建立附加至圖形元素的標記。
- **大小。** 當大小已由某個類別欄位定義時, 其會根據個別的類別分割視覺化, 每種類別各一種大小。當大小是連續數值範圍時, 其會根據範圍欄位的值來變更大小。若圖形元素 (例如長條圖或方框) 代表一個以上的記錄/觀察值, 並將範圍欄位用於大小, 則大小會依據範圍欄位的平均數改變。

## 畫面化和動畫

**畫面化。** 畫面化 (亦稱為分面) 會建立圖形表格。畫面化欄位中的每個類別會產生一個圖形, 但所有窗格會同步顯示。畫面化可用於檢查視覺化是否適用於畫面化欄位的狀況。例如, 您可依性別畫面化一個直方圖, 以判斷男性和女性之間的次數分配是否相等。換言之, 您可查看薪水是否因性別而有差異。選取種類欄位進行畫面化。

**動畫。** 動畫和畫面化的相似處, 在於多個圖形是從動畫欄位的值而建立, 但這些圖形不會一併顯示。您使用「探索」模式中的控制項為輸出加上動畫效果, 並透過連續的個別圖形翻轉。此外, 和畫面化不同的是, 動畫不需要類別欄位。您可指定連續的欄位, 其值會自動分割為範圍。您可使用探索模式中的動畫控制項來變更範圍的大小。並非所有視覺化都提供動畫效果。

## 使用「輸出」標籤

對於所有圖形類型, 都可以為產生的圖形的檔名和顯示指定下列選項。

註: 分佈節點圖具有其他設定。

**輸出名稱。** 指定執行節點時產生的圖表名稱。**自動**會根據產生輸出的節點來選擇名稱。您可以選擇性地選取自訂以指定不同的名稱。

**輸出至畫面。** 選取此選項將在新視窗中產生並顯示圖表。

**輸出至檔案。** 選取此選項可將輸出另存為檔案。

- **輸出圖形。** 選取此選項將以圖表格式產生輸出。僅適用於分佈節點。
- **輸出表格。** 選取此選項將以表格式產生輸出。僅適用於分佈節點。
- **檔名。** 指定用於所產生圖表或表格的檔名。使用省略號按鈕 (...) 指定具體檔案和位置。
- **檔案類型。** 在下拉清單中指定檔案類型。對於所有圖表節點, 除了使用**輸出表格**選項的「分佈」節點外, 可用的圖表檔案類型如下。

- 點陣圖 (.bmp)
- PNG (.png)
- 輸出物件 (.cou)
- JPEG (.jpg)
- HTML (.html)
- ViZml 文件 (.xml)，可在其他 IBM SPSS Statistics 應用程式中使用。

對於「分佈」節點中的輸出表格選項，可用的檔案類型如下。

- 以 tab 空格區隔的資料 (.tab)
- 以逗點區隔的資料 (.csv)
- HTML (.html)
- 輸出物件 (.cou)

**編頁碼輸出。**如果將輸出儲存為 HTML，那麼將啟用此選項以使您可以控制每個 HTML 頁的大小。（僅適用於分佈節點。）

**每頁行數。**如果選取了編頁碼輸出，那麼將啟用此選項以使您能夠確定每個 HTML 頁的長度。該預設設定為 400 列。（僅適用於分佈節點。）

### 使用「註解」標籤

用於所有節點，此標籤提供一些選項來重新命名節點、提供自訂工具提示，以及儲存長註釋。

### 三維圖形

IBM SPSS Modeler 中的統計圖和收集圖形能夠在第三坐標軸上顯示資訊。這樣就可以更加靈活地對資料進行視覺化，以選取子集或衍生用於建模的新欄位。

建立三維圖表後，即可按一下圖表並拖曳滑鼠旋轉圖表，以便從任何角度檢視圖表。



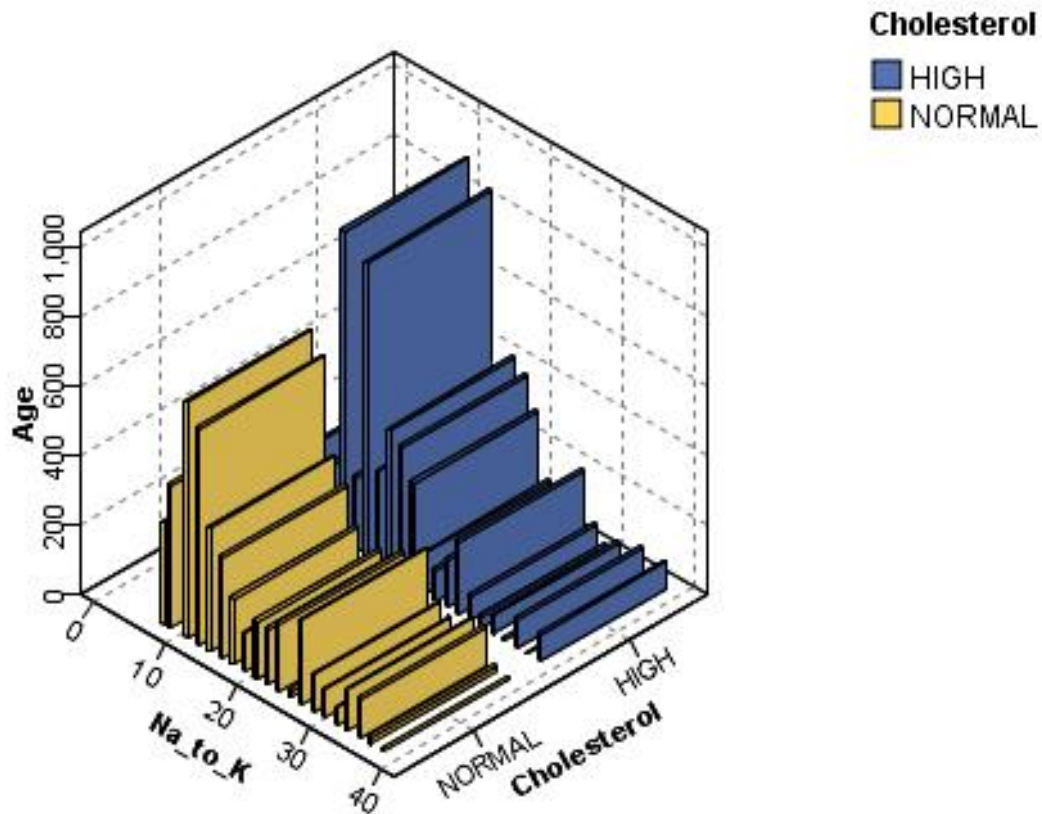


圖 7. 帶有  $x$ 、 $y$  和  $z$  軸的收集圖

在 IBM SPSS Modeler 中建立三維圖形有兩種方法：在第三坐標軸上繪製資訊（true 實三維圖形）和以三維效果顯示圖形。統計圖和收集圖形都可以使用這兩種方法。

在第三坐標軸上繪製資訊

1. 按一下圖表節點對話框中的統計圖標籤。
2. 按一下三維按鈕以啟用  $z$  軸選項。
3. 使用「欄位選擇器」按鈕以選擇  $z$  軸的欄位。某些情況下，這裡只容許使用符號欄位。「欄位選擇器」將顯示適合的欄位。

向圖表新增三維效果

1. 建立圖表後，即可按一下輸出視窗中的圖表標籤。
2. 按一下三維度按鈕以將視圖切換為三維度圖表。

---

## 圖形板節點

通過「圖版」節點，您可以從單個節點中的多數不同圖形輸出（長條圖、圓餅圖、直方圖、散佈圖和熱力圖等）中進行選擇。在第一個標籤中，首先選取想要探索的資料欄位，然後節點將提供一個適用於您的資料的圖形類型選項。該節點會自動濾出不適用於欄位選項的任何圖形類型。在「詳細」標籤上，可以定義詳細或較進階的圖表選項。

註：為了編輯節點或選取圖形類型，必須將「圖版」節點連接到包含資料的串流。

有兩個按鈕讓您控制哪些視覺化範本（以及樣式表與地圖）可用：

**管理。**管理您電腦中的視覺化範本、樣式表和地圖。您可以匯入、匯出、重新命名或刪除您本機電腦中的視覺化範本、樣式表與地圖。請參閱第 188 頁的『管理範本、樣式表與地圖檔案』主題，以取得更多資訊。

**位置。**變更儲存視覺化範本、樣式表與地圖的位置。目前位置列於按鈕右邊。請參閱第 187 頁的『設定範本、樣式表與地圖的位置』主題，以取得更多資訊。

## 圖板基本標籤

若您不確定哪一種視覺化類型最適合呈現您的資料，可使用「基本」標籤。選取資料時，您會看到適用於此資料的視覺化類型子集。如需範例，請參閱第 176 頁的『圖板範例』。

1. 從清單中選取一或多個欄位 (變數)。按住 Ctrl 並按一下可選取多個欄位。

請注意，欄位的測量層級決定了可用的視覺化類型。您可在清單中的欄位上按右鍵，並選擇選項以變更測量層級。如需可用測量層次類型的相關資訊，請參閱第 167 頁的『欄位 (變數) 類型』。

2. 選取一種視覺化類型。如需可用類型的說明，請參閱第 170 頁的『可用的內建圖板視覺化類型』。
3. 對於某些視覺化，您可選擇摘要統計量。統計量的不同子集之可用性，會根據統計量是以個數為基準或是從連續欄位計算得來而有所不同。可用的統計量也會根據範本本身而異。您可以在下一個步驟中找到可用統計量的完整清單。
4. 若您希望定義更多選項，如選用審美標準與窗格欄位，請按一下「細節」。請參閱第 169 頁的『圖板詳細資訊標籤』主題，以取得更多資訊。

從連續欄位計算的摘要統計量

- 平均數( $M$ )。集中趨勢的測量。算術平均數，總和除以觀察值數。
- 中位數。半數觀察值落點上下的值，第 50 個百分位數。如果觀察值個數是偶數的話，中位數是中間兩個觀察值的平均值，此處的兩個中間是指，當觀察值按照遞增或遞減順序排列時，位於最中間的兩個值。中位數為集中趨勢的量數，其不會感應到偏離值（相反地，平均數會受到幾個高低極端值所影響）。
- 眾數( $O$ )。最常出現的值。如果幾個值的最大出現次數相同，則每一個都是眾數。
- 最小值。數值變數的最小值。
- 最大值。數值變數的最大值。
- 範圍。最小與最大值之間的差異。
- 中間範圍。範圍的中間值，亦即與最小值之間的差異值等於與最大值之間的差異值。
- 總和( $U$ )。遍及所有包含遺漏值之觀察值的值總和或總數。
- 累積總和。值的累積總和。每一個圖形元素都顯示出一個子群組加上所有之前群組總和的和。
- 百分比總和。以總和欄位為基礎，與所有群組的總和相比較的每個子群組中的百分比。
- 累積百分比總和。根據與所有群組總和相較的總和欄位所得的每個子群組中的累積百分比。每一個圖形元素都顯示出一個子群組加上所有之前群組百分比的百分比。

- 變異數. 測量平均數四周的離散情形，它等於平均數的平方離差總和除以觀察值數目減一。變異數的測量單位是變數本身的平方。
- 標準差. 測量平均數四周的離散情形。在常態分配中，68% 的觀察值會落在平均數的一個標準差內，95% 的觀察值會落在兩個標準差內。例如，如果平均年齡為 45 歲，標準差是 10 的話，在常態分配中 95% 的觀察值會介於 25 到 65 歲之間。
- 標準誤. 不同樣本間測試統計量值的變化大小量數。這是一項統計量取樣分配的標準差。例如，平均數的標準誤就是樣本平均數的標準差。
- 峰度. 離群值範圍的一種測量。對常態分配而言，峰度統計量數值為零。峰度為正數指出資料比常態分佈展示更多的極端離群值。峰度為負數指出資料比常態分佈展示較少的極端離群值。
- 偏斜度. 分配不對稱性的量數。常態分佈是一種對稱性分佈，其偏斜度值為 0。具有顯著性正偏斜度的分佈右側尾部較長。有顯著負偏斜度值的分配會有一個長的偏左尾部。偏斜度值有如指標，若大於它的兩倍標準誤，則表示背離對稱。

以下區域統計量可能導致每個子群組產生一個以上的圖形元素。當使用區間、區域或邊緣圖形元素時，區域統計量會導致產生一個顯示範圍的圖形元素。其他所有圖形元素會導致產生兩個單獨的元素，其中一個顯示範圍的啟動，另一個顯示範圍的結束。

- 區域：範圍。最小與最大值之間的數值範圍。
- 區域：平均數的 95% 信賴區間。有 95% 機會包含人口平均數的數值範圍。
- 區域：個別觀察值的 95% 信賴區間。在個別觀察值的情況下，有 95% 機會包含預測值的數值範圍。
- 區域：高於/低於平均數的 1 個標準差。介於高於和低於 平均數 1 個標準差之間的數值範圍。
- 區域：高於/低於平均數的 1 個標準錯誤。介於高於和低於 平均數 1 個標準誤之間的數值範圍。

以個數為基準的摘要統計量

- 個數。列/觀察值的個數。
- 累積個數。列數/觀察值的累積個數。每一個圖形元素都顯示出一個子群組加上所有之前群組個數的個數。
- 個數百分比。每一個子群組中，與總列數/觀察值相比較的列/觀察值的百分比。
- 累積個數百分比。每一個子群組中與總列數/觀察值相比較列/觀察值的累積百分比。每一個圖形元素都顯示出一個子群組加上所有之前群組百分比的百分比。

## 欄位 (變數) 類型

欄位清單中位於欄位旁的圖示，表示欄位類型與資料類型。圖示同時也可識別多重回應集。

表 32. 測量層次圖示.














測量層級	數值型(N)	字串	日期	時間(M)
連續的		無		
排序集合				
集合				

表 33. 多重回應集圖示。

多重回應集類型	圖示
多重回應集，多重類別	
多重回應集，多重二分集	

## 測量層級

欄位的測量層級在您建立視覺化時很重要。以下為測量層級的說明。您可在欄位清單中的欄位上按一下右鍵，並選擇選項，暫時變更測量層級。在大部分的情況中，您只需要考量兩個最廣泛的欄位分類（也就是類別和連續）：

**類別。**資料具有幾個不同值或類別的變數（例如，性別或宗教）。類別欄位可以是字串（文數字）或數值欄位，使用數字編碼來代表類別（例如，0 = 男而 1 = 女）。也稱為定性資料。集合、排序集合與旗標都是類別欄位。

- **集合。**一種欄位/變數，其值可表示不包含等級化種類（例如，某個員工在公司內任職的部門）的種類。名義變數的範例包括地區、郵遞區號以及宗教團體。也稱為名義變數。
- **排定次序的集。**一種欄位/變數，其值可表示包含等級化種類（例如，服務滿意度，可從非常不滿意至非常滿意的程度進行排序）的種類。排序集合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。也稱為序數變數。
- **旗標。**包含兩個不同值（如 Yes 及 No 或 1 及 2）的欄位/變數。一般也稱為二分/二進位變數。

**連續。**資料以區間或比例尺度測量，其資料值同時表示數值順序及數值之間的距離。例如，薪資 \$72,195 比 \$52,398 高，而兩數值之間的距離為 \$19,797。也稱為數值、尺度或數值範圍資料。

類別欄位會定義視覺化中的類別，一般的方式是取得各個圖形元素或將圖形元素分組。連續欄位通常會在類別欄位的類別中作成摘要。例如，在性別類別的收入預設視覺化中，會分別顯示男性平均收入以及女性平均收入。也可以繪製連續欄位的原始值，就如同散佈圖。例如，散佈圖可以顯示各個觀察值的目前薪資與剛啟動工作時的新資。可以使用類別欄位將觀察值以性別分組。

## 資料類型

測量層級並不是判斷欄位類型唯一的特質。欄位亦可儲存為指定資料類型。可能的資料類型有字串（字母等非數值資料）、數值（實數）和日期。與測量層級不同的是，欄位的資料類型無法暫時變更。您必須變更資料在原始資料集中的儲存方式。

## 多重回應集

有些資料檔支援特殊的「欄位」，稱為「**多重回應集**」。多重回應集不是一般所熟知的「欄位」。多重回應集會使用多重欄位，來記錄應答者對問題的回應（應答者的回答可能不止一個）。多重回應集是視為類別欄位來處理，類別欄位能做的事，多重回應集大半也都能做。

多重回應集可以是多重二分集或多類別集合。

**多重二分集。**多重二分集一般是以多重二分欄位組成：只有兩個可能值的欄位：是/否、在場/不在場、選取/未選取性質。雖然欄位可能不完全是二分欄位，但集合中所有欄位都會以相同方式編碼。

例如，調查會對「您從下列何處獲取新聞？」這個問題提供五個可能的回答。應答者可以在每個選取旁邊的方框打勾，來表示多重選取。五個回答變成資料檔中的五個欄位，編碼 0 代表否 (不選取) 以及 1 代表是 (選取)。

**多類別集合。**多類別集合是由多重欄位組成，全部都是用相同方式編碼，經常有許多可能的回應類別。例如，調查項目陳述：「列出最多三個能代表你們種族傳統的國家名稱。」回答可能會有數百種，但是因為編碼的關係，清單限制在 40 個最常見的國家中，其餘一切都歸的「其他」類別。在資料檔中，三個選取變成三個欄位，每個都有 41 個類別 (40 個編碼國家和一個「其他」類別)。

## 圖板詳細資訊標籤

若您知道您要建立的視覺化的何種類型，或您希望在視覺化中新增選用的審美標準、窗格和/或動畫，可使用「詳細資訊」標籤。如需範例，請參閱第 176 頁的『圖板範例』。

1. 若您在「基本」標籤中選擇一種視覺化類型，則會顯示該類型。否則請從下拉清單選擇一種類型。如需視覺化類型的相關資訊，請參閱第 170 頁的『可用的內建圖板視覺化類型』。
2. 視覺化縮圖影像的正右方，有控制項可以指定視覺化類型所需的欄位 (變數)。您必須指定這些所有欄位。
3. 您可針對某些視覺化選擇摘要統計量。某些情況下 (例如有長條圖時)，您可使用這些摘要選項之一當作透明度審美標準。如需摘要統計量的說明，請參閱第 166 頁的『圖板基本標籤』。
4. 您可以選擇一個或多個選用審美標準：這可讓您納入視覺化中的其他欄位，以增加範圍。例如，您可使用欄位，變化散佈圖中點的大小。如需選用審美標準的相關資訊，請參閱第 162 頁的『外觀、套版、畫面和動畫』。請注意，無法透過 Scripting 來支援透通性審美標準。
5. 如果您建立地圖視覺化，「地圖檔案」群組會顯示將使用的一或多個地圖檔案。如果有預設地圖檔案，會顯示此檔案。若要變更地圖檔案，請按一下「選取地圖檔案」來顯示「選取地圖」對話框。您也可在此對話框中指定預設地圖檔案。請參閱『為地圖視覺化選取地圖檔案』主題，以取得更多資訊。
6. 您可選擇一或多個畫面化或動畫選項。如需畫面化與動畫選項的更多資訊，請參閱第 162 頁的『外觀、套版、畫面和動畫』。

## 為地圖視覺化選取地圖檔案

如果您選取地圖視覺化範本，您需要有一個地圖檔案能夠針對繪製地圖定義地理資訊。如果有預設地圖檔案，這將會用於地圖視覺化。若要選擇不同的地圖檔案，請按一下「詳細資訊」標籤上的「選取地圖檔案」來顯示「選取地圖」對話框。

「選取地圖」對話框可以讓您選擇主要地圖檔案及參考地圖檔案。地圖檔案可針對繪製地圖定義地圖資訊。您的應用程式會安裝一組標準地圖檔案。如果您有要使用的其他 ESRI 形狀檔案，您首先需要將形狀檔案轉換為 SMZ 檔案。請參閱第 189 頁的『轉換及發佈地圖形狀檔案』主題，以取得更多資訊。轉換地圖之後，按一下「範本選擇器」對話框上的管理...，來將地圖匯入到「管理」系統中，使其能夠在「選取地圖」對話框中出現。

以下是指定地圖檔案時需要考慮的幾點：

- 所有地圖範本需要至少一個地圖檔案。
- 地圖檔案一般會將地圖鍵屬性鏈結至資料鍵。
- 如果範本不需要鏈結至資料鍵的地圖鍵，它需要參考地圖檔案與指定座標 (例如經度與緯度) 以供在參考地圖上繪製元素的欄位。
- 重疊圖範本需要兩個地圖：主要地圖檔案及參照地圖檔案。會先繪製參考地圖，以使其位於主要地圖檔案之後。

如需屬性與特徵等地圖專有名詞的相關資訊，請參閱第 189 頁的『地圖的關鍵概念』。

**地圖檔案。** 您可以選取位於「管理」系統中的任何地圖檔案。這包括預先安裝的地圖檔案與您匯入的地圖檔案。如需管理地圖檔案的相關資訊，請參閱第 188 頁的『管理範本、樣式表與地圖檔案』。

**地圖鍵。** 指定您要作為將地圖檔案鏈結至資料鍵的鍵使用的屬性。

**將此地圖檔案與設定儲存為預設值。** 如果您要將所選地圖檔案作為預設值使用，請選取此勾選框。如果您已指定預設地圖檔案，您不需要在每次建立地圖視覺化時指定地圖檔案。

**資料鍵。** 此控制項可列出與在「範本選擇器」的「詳細資訊」標籤上顯示的相同值。如果您因選擇的特定地圖檔案而需要變更鍵，則會在此提供以方便您使用。

**在視覺化中顯示所有地圖功能。** 勾選此選項時，會在視覺化中呈現地圖中的所有特徵，即使沒有相符的資料鍵值也是如此。如果您只想要看見有資料的功能，請清除此選項。顯示於「不相符的地圖鍵」清單中由地圖鍵識別的功能將不會以視覺化呈現。

**比較地圖與資料值。** 地圖鍵與資料鍵會彼此鏈結，以建立地圖視覺化。地圖鍵與資料鍵應該由相同的網域取得(例如，國家或是區域)。按一下「比較」來測試資料鍵與地圖鍵是否相符。顯示的圖示可通知您比較的狀態。以下說明這些圖示。如果已執行比較，且資料鍵值不含相符地圖鍵值，資料鍵值會顯示在「不相符的資料鍵」清單中。在「不相符的地圖鍵」清單中，您也可以看到哪些地圖鍵的值沒有相符的資料鍵值。如果取消勾選在視覺化中顯示所有地圖功能，將不會呈現由這些地圖鍵值識別的功能。

表 34. 比較圖示。

圖示	說明
	未執行比較。這是在您按一下「比較」之前的預設狀態。繼續時應注意，因為您並不知道資料鍵與地圖鍵值是否相符。
	已執行比較，且資料鍵與地圖鍵值完全相符。針對資料鍵的每一個值，都有由地圖鍵所識別的相符特徵。
	已執行比較，且部分資料鍵與地圖鍵值不相符。針對部分資料鍵值，並沒有由地圖鍵所識別的相符特徵。繼續時應注意。如果您繼續，地圖視覺化將不包括所有資料值。
	已執行比較，且沒有資料鍵與地圖鍵值相符。您應選擇不同的資料鍵或地圖鍵，因為如果繼續，將不會呈現地圖。

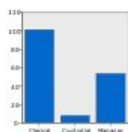
## 可用的內建圖板視覺化類型

您可建立多種不同的視覺化類型。下列所有內建類型，都可在「基本」和「詳細資訊」標籤中使用。範本(特別是地圖範本)的某些說明，可識別在「詳細資訊」標籤上使用「特殊文字」指定的欄位(變數)。



表 35. 可用圖形類型。

**圖表圖示**



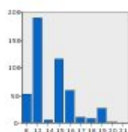
**說明**

**條形圖**

計算連續數值欄位的摘要統計量，並以長條圖顯示類別欄位的各類別結果。

需要：種類欄位及連續欄位。

**圖表圖示**

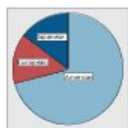


**說明**

**個數長條**

以長條圖顯示類別欄位中各類別的列/觀察值比例。您亦可使用「分配」圖形節點製作此圖形。節點提供一些其他選項。請參閱第 205 頁的『分佈節點』主題，以取得更多資訊。

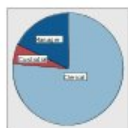
需要：單一種類欄位。



**圓餅圖**

計算連續數值欄位的總和，並以圓餅圖切片顯示類別欄位的各類別總和散佈。

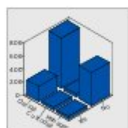
需要：種類欄位及連續欄位。



**個數圓餅**

以圓餅圖切片顯示類別欄位中各類別的列/觀察值比例。

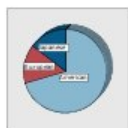
需要：單一種類欄位。



**立體長條圖**

計算連續數值欄位的摘要統計量，並顯示兩類別欄位之間的交叉類別結果。

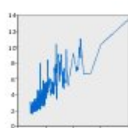
需要：一對種類欄位與一個連續欄位。



**3-D 圓餅**

與圓餅圖相同，只是增加了 3-D 效果。

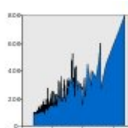
需要：種類欄位及連續欄位。



**線形圖**

針對另一個欄位的每個值計算欄位的摘要統計量，並繪製連接值的線。您亦可使用「圖形」圖形節點製作線形圖。節點提供一些其他選項。請參閱第 194 頁的『統計圖節點』主題，以取得更多資訊。

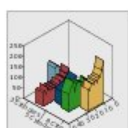
需要：一對任何類型的欄位。



**區域圖**

針對另一個欄位的每個值計算欄位的摘要統計量，並繪製連接值的區域。直線與區域之間的差異極小，可將區域看做下方有彩色空間的直線。但若您使用色彩審美標準，這樣會導致直線遭到分割，而區域產生堆疊。

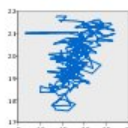
需要：一對任何類型的欄位。



**3-D 區域**

顯示對照其他連續數值範圍繪製出來的欄位，並以類別欄位進行分割。每個類別都會繪製區域元素。

需要：一個種類欄位及一對任何類型的欄位。



**路徑**

顯示對照其他欄位值繪製出來的欄位值，並以直線依照出現在原始資料集的順序連接這些值。路徑與直線最主要的差異在於順序。

需要：一對任何類型的欄位。

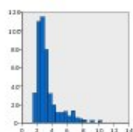
表 35. 可用圖形類型 (繼續).

圖表圖示	說明	圖表圖示	說明
	<p><b>綵帶</b></p> <p>針對另一個欄位的每個值計算欄位的摘要統計量，並繪製連接值的帶狀。帶狀的本質是直線，加了 3-D 效果。並不是真正的 3-D 圖形。</p> <p>需要：一對任何類型的欄位。</p>		<p><b>表面</b></p> <p>顯示對照其他欄位繪製出來的三個欄位值，並以表面連接這些值。</p> <p>需要：任何類型的三個欄位。</p>
	<p><b>散佈圖</b></p> <p>顯示對照另一個數值所繪製的某欄位數值。此圖形可強調出欄位之間的關係 (若有的話)。您亦可使用「圖形」圖形節點製作散佈圖。節點提供一些其他選項。請參閱第 194 頁的『統計圖節點』主題，以取得更多資訊。</p> <p>需要：一對任何類型的欄位。</p>		<p><b>泡泡圖</b></p> <p>與基本散佈圖相似，顯示對照另一個數值所繪製的某欄位數值。差別在於第三個欄位的數值會用來變化個別點的大小。</p> <p>需要：任何類型的三個欄位。</p>
	<p><b>Bin 散佈圖</b></p> <p>與基本散佈圖相似，顯示對照另一個數值所繪製的某欄位數值。差別在於相似數值會 bin 在一起，並使用色彩或大小審美標準判定各 bin 的觀察值個數。</p> <p>需要：一對連續欄位。</p>		<p><b>六角形 Bin 散佈圖</b></p> <p>請參閱「Bin 散佈圖」的說明。差別在於基本 bin 的形狀像是六角形，而不是圓形。這樣會讓六角形 Bin 散佈圖看起來與 Bin 散佈圖很相似。但由於基本 bin 的形狀，這兩種圖形之中的 bin 數值數也有所不同。</p> <p>需要：一對連續欄位。</p>
	<p><b>3-D 散佈圖</b></p> <p>顯示互相對照所繪製的三個欄位值。此圖形可強調出欄位之間的關係 (若有的話)。您亦可使用「圖形」圖形節點製作 3-D 散佈圖。節點提供一些其他選項。請參閱第 194 頁的『統計圖節點』主題，以取得更多資訊。</p> <p>需要：任何類型的三個欄位。</p>		<p><b>散佈圖矩陣 (SPLOM)</b></p> <p>針對每一個欄位顯示對照另一個數值所繪製的某欄位數值。SPLOM 就像散佈圖的表格。SPLOM 亦包含各欄位的直方圖。</p> <p>需要：兩個或多個連續欄位。</p>



表 35. 可用圖形類型 (繼續).

圖表圖示



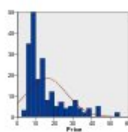
說明

**直方圖**

顯示欄位的次數分配。直方圖可助於判斷散佈類型，並瞭解散佈是否偏斜。您亦可使用「直方圖」圖形節點製作此圖形。節點提供一些其他選項。請參閱第 208 頁的『直方圖「繪圖形」標籤』主題，以取得更多資訊。

需要：任何類型的單一欄位。

圖表圖示

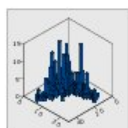


說明

**具有常態分佈的直方圖**

顯示包括常態分佈重疊曲線的連續欄位次數分配。

需要：單一連續欄位。



**3-D 直方圖**

顯示一對欄位的次數分配。

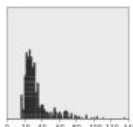
需要：一對連續欄位。



**3-D 密度**

顯示一對欄位的次數分配。這與 3-D 直方圖相似，唯一的差異是使用表面而不是長條來顯示散佈。

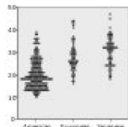
需要：一對連續欄位。



**點形圖**

顯示個別觀察值/列，並將它們堆疊在  $x$  軸的不同資料點上。此圖形顯示資料分佈方面與直方圖相似，但它是顯示各觀察值/列，而不是特定 bin (數值範圍) 的聚集個數。

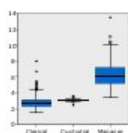
需要：任何類型的單一欄位。



**2-D 點形圖**

顯示個別觀察值/列，並根據類別欄位的各類別，將它們堆疊在  $y$  軸的不同資料點上。

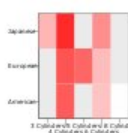
需要：種類欄位及連續欄位。



**盒形圖**

計算類別欄位各類別的連續欄位之五種統計量 (最小值、第一個四分位數、中位數、第三個四分位數和最大值)。以盒形圖/架構元素顯示結果。盒形圖可協助您瞭解類別間連續資料的散佈情形。

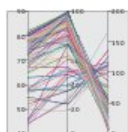
需要：種類欄位及連續欄位。



**熱圖**

計算兩類別欄位之間交叉的連續欄位平均數。

需要：一對種類欄位與一個連續欄位。



**平行模式測試**

建立各欄位的平行軸，並繪製一條直線，通過資料中各列/觀察值的欄位值。

需要：兩個或多個連續欄位。



**個數面量圖**

計算類型欄位 (資料鍵) 的每個類別的個數，並繪製使用色彩飽和度來表示地圖特徵中對應於類別之個數的地圖。

需要：種類欄位。其鍵符合「資料鍵」類別的地圖檔案。

表 35. 可用圖形類型 (繼續).


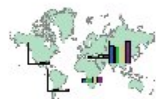
圖表圖示	說明	圖表圖示	說明
	<p><b>平均數/中位數/總和的面量圖</b> 針對類別欄位 (資料鍵) 的每一個類別計算連續欄位 (色彩) 的平均數、中位數或總和，並繪製使用色彩飽和度來表示地圖特徵中對應於類別之計算統計量的地圖。</p> <p>需要：種類欄位及連續欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>		<p><b>值面量圖</b> 針對對應於由其他類別欄位 (資料鍵) 所定義之值的地圖特徵，繪製使用色彩來表示類別欄位 (色彩) 的值的地圖。如果每個特徵都有「色彩」欄位的多個類別值，則會使用典型值。</p> <p>需要：一對種類欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>
	<p><b>個數面量圖上的座標</b> 這與「個數面量圖」類似，但有其他兩個連續欄位 (「經度」與「緯度」) 可識別在個數面量圖上繪製點的座標。</p> <p>需要：一個種類欄位及一對連續欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>		<p><b>平均數/中位數/總和的面量圖上的座標</b> 這與「平均數/中位數/總和的面量圖」類似，但其他兩個連續欄位 (「經度」與「緯度」) 可識別在面量圖上繪製點的座標。</p> <p>需要：一個種類欄位及三個連續欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>
	<p><b>值面量圖上的座標</b> 這與「值面量圖」類似，但其他兩個連續欄位 (「經度」與「緯度」) 可識別在面量圖上繪製點的座標。</p> <p>需要：一對種類欄位及一對連續欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>		<p><b>地圖上的個數長條</b> 針對每一個地圖特徵 (資料鍵) 計算類別欄位 (類別) 的每一個類別中的列/觀察值比例，並在每一個地圖特徵的中心繪製地圖與長條圖。</p> <p>需要：一對種類欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>
	<p><b>地圖上的長條</b> 計算連續欄位 (值) 的摘要統計量，並將每一個地圖特徵 (資料鍵) 的類別欄位 (類別) 的每一個類別之結果顯示為置於每一個地圖特徵中心的長條圖。</p> <p>需要：一對種類欄位與一個連續欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>		<p><b>地圖上的個數圓餅</b> 針對每一個地圖特徵 (資料鍵) 顯示類別欄位 (類別) 的每一個類別中的列/觀察值比例，並在每一個地圖特徵的中心繪製地圖與圓餅圖的切片比例。</p> <p>需要：一對種類欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>

表 35. 可用圖形類型 (繼續).

圖表圖示	說明	圖表圖示	說明
	<p><b>地圖上的圓餅</b></p> <p>針對每一個地圖特徵 (資料鍵) 計算類別欄位 (類別) 的每一個類別中的連續欄位 (值) 的總和, 並在每一個地圖特徵的中心繪製地圖與圓餅圖的切片總和。</p> <p>需要: 一對種類欄位與一個連續欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>		<p><b>地圖上的線條圖表</b></p> <p>針對每一個地圖特徵 (資料鍵) 計算另一個欄位 (X) 的每一個值的一個連續欄位 (Y) 的摘要統計量, 並在每一個地圖特徵的中心繪製連接值的地圖與線條圖表。</p> <p>需要: 一個種類欄位及一對任何類型的欄位。其鍵符合「資料鍵」類別的地圖檔案。</p>
	<p><b>參考地圖上的座標</b></p> <p>使用識別點座標的連續欄位 (「經度」與「緯度」) 繪製地圖與點。</p> <p>需要: 一對範圍欄位。一個地圖檔案。</p>		<p><b>參考地圖上的箭頭</b></p> <p>使用識別每一個箭頭的起點 (「啟動經度」與「啟動緯度」) 及終點 (「結束經度」與「結束緯度」) 的連續欄位繪製地圖與箭頭。資料結果中的每一個記錄/觀察值都會在地圖中產生箭頭。</p> <p>需要: 四個連續欄位。一個地圖檔案。</p>
	<p><b>點重疊圖</b></p> <p>繪製參考地圖並用由類別欄位 (色彩) 上色的點特徵來在參考圖上重疊另一個點圖。</p> <p>需要: 一對種類欄位。其鍵符合「資料鍵」類別的點圖檔案。一個參考地圖檔案。</p>		<p><b>多邊形重疊圖</b></p> <p>繪製參考圖並用由類別欄位 (色彩) 上色的多邊形特徵來在參考圖上重疊另一個多邊形圖。</p> <p>需要: 一對種類欄位。其鍵符合「資料鍵」類別的多邊形圖檔案。一個參考地圖檔案。</p>
	<p><b>線條重疊圖</b></p> <p>繪製參考圖並用由類別欄位 (色彩) 上色的線條特徵來在參考圖上重疊另一個線條圖。</p> <p>需要: 一對種類欄位。其鍵符合「資料鍵」類別的線條圖檔案。一個參考地圖檔案。</p>		

## 建立地圖視覺化

對於許多視覺化而言, 您只需要做出兩個選擇: 感興趣的欄位 (變數) 及視覺化這些欄位的範本。不需要做出其他選擇或動作。地圖視覺化需要至少一個額外的步驟: 選取為地圖視覺化定義地理資訊的地圖檔案。

建立簡單地圖的基本步驟如下:

1. 在「基本」標籤上選取感興趣的欄位。如需不同地圖視覺化所需欄位類型及數目的相關資訊, 請參閱第 170 頁的『可用的內建圖板視覺化類型』。

2. 選取地圖範本。
3. 按一下「詳細資訊」標籤。
4. 檢查「資料鍵」與其他所需下拉清單是否已設定為正確的欄位。
5. 在「地圖檔案」群組中，按一下「選取地圖檔案」。
6. 使用「選取地圖」對話框來選擇地圖檔案與地圖鍵。地圖鍵的值必須與「資料鍵」所指定的欄位值相符。您可以使用「比較」按鈕來比較這些值。如果您選取重疊地圖範本，您也將需要選擇參考地圖。不會將參考地圖嵌入至資料。它會作為主地圖的背景使用。如需「選取地圖」對話框的相關資訊，請參閱第 169 頁的『為地圖視覺化選取地圖檔案』。
7. 按一下「確定」關閉「選取地圖」對話框。
8. 在「圖表板樣本選擇器」中，按一下「執行」建立地圖視覺化。

## 圖板範例

本節包含許多不同的範例，以示範可用選項。這些範例亦提供解讀視覺化結果的相關資訊。

這些範例使用名為 *graphboard.str* 的資料流，其參照到的資料檔名稱分別為 *employee\_data.sav*、*customer\_subset.sav* 和 *worldsales.sav*。這些檔案可從任何 IBM SPSS Modeler Client 安裝的 *Demos* 資料夾取得。您可從 Windows「啟動」功能表的 IBM SPSS Modeler 程式集存取。*graphboard.str* 檔案位於 *streams* 資料夾。

建議您以我們呈現的順序閱讀範例。後續範例建置於先前範例的基礎之上。

### 範例：具有摘要統計量的長條圖

我們會建立一個長條圖，該圖則建立一集合/類別變數的各類別連續數值欄位/變數摘要。明確來說，我們要建立的長條圖，會顯示男性與女性薪資的平均數。

此範例與後續許多範例都使用 *Employee data*，這是一個假設的資料庫，包含公司員工的相關資訊。

1. 新增連接到 *employee\_data.sav* 的 Statistics 檔案來源節點。
2. 新增圖板節點並開啟進行編輯。
3. 在「基本」標籤上，選取性別及現行薪資。(按住 Ctrl 並按一下可選取多個欄位/變數。)
4. 選取長條圖。
5. 從「摘要」下拉清單中，選取功能表。
6. 按一下「執行」。
7. 在結果顯示中，按一下「顯示欄位與數值標籤」工具列按鈕 (工具列中央兩群組間的第二個)。

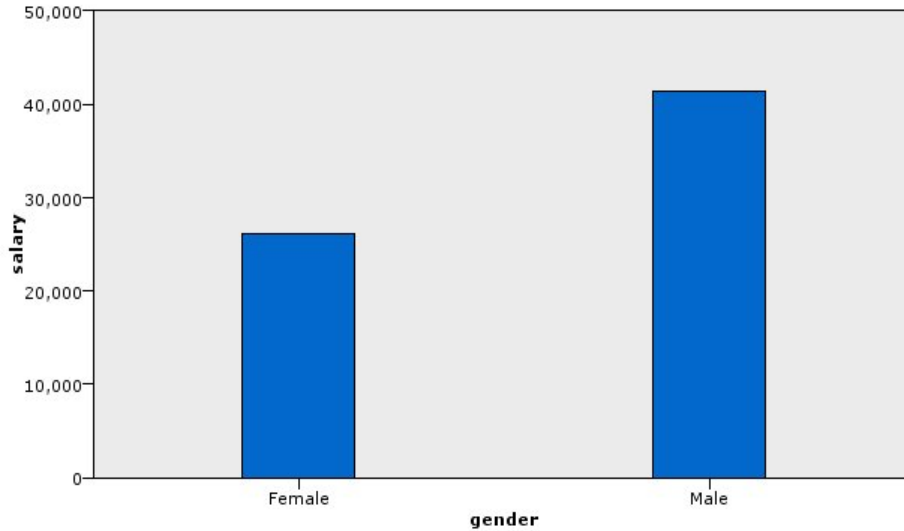


圖 8. 具有摘要統計量的長條圖

我們可以觀察下列資訊：

- 根據長條的高度，可以清楚看到男性的平均薪資大於女性的平均薪資。

### 範例：含彙總統計量的堆疊長條圖

現在我們將建立堆疊長條圖，以瞭解陽性與陰性的平均工資差異是否取決於工作類型。或許平均而言，女性在某些工作類型上的所得會超過男性。

註：本範例使用員工資料。

1. 新增圖板節點並開啟進行編輯。
2. 在「基本」標籤上，選取職業種類和目前工資。(按住 Ctrl 並按一下可選取多個欄位/變數。)
3. 選取長條圖。
4. 從「摘要」清單中，選取功能表。
5. 按一下「詳細資訊」標籤。請注意，您在上一個標籤中所作的選擇會反映在此。
6. 在「選用審美標準」組別中，從「顏色」下拉清單中選擇性別。
7. 按一下「執行」。

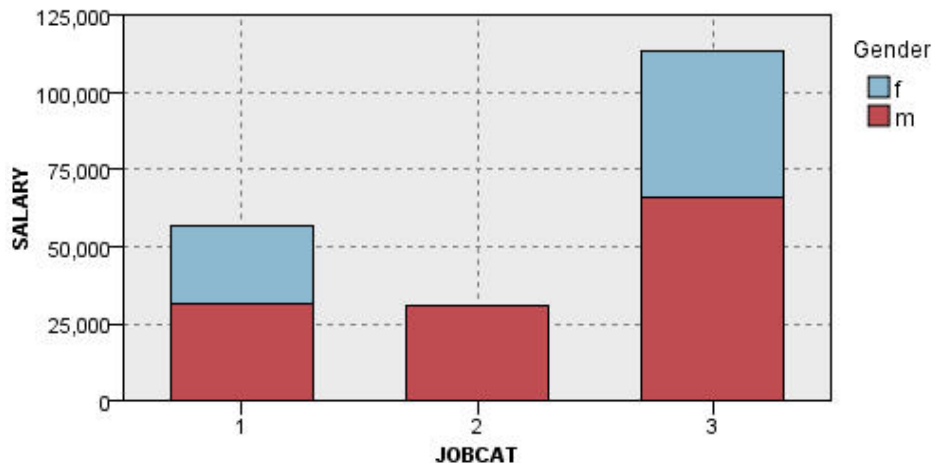


圖 9. 堆疊長條圖

我們可以觀察下列資訊：

- 各工作類型的平均薪資差異就沒有像長條圖中比較所有男性與所有女性平均薪資的差異那麼明顯了。或許各組別中男性與女性的數量有所差別。您可建立個數長條圖來驗證此說法。
- 不考慮工作類型的話，男性平均薪資一律大於女性平均薪資。

### 範例：畫面化直方圖

我們可以建立依性別執行畫面化的直方圖，以比較男性與女性薪資的次數分配。次數分配會顯示特定薪資範圍內的個案數/列數量。畫面化直方圖可協助我們進一步分析性別之間的薪資差異。

附註：本範例使用 *Employee data*。

1. 新增圖板節點並開啟進行編輯。
2. 在「基本」標籤上，選取現行薪資。
3. 選取直方圖。
4. 按一下「詳細資訊」標籤。
5. 在「畫面化與動畫」組別中，從「窗格跨越」下拉清單中選擇性別。
6. 按一下「執行」。

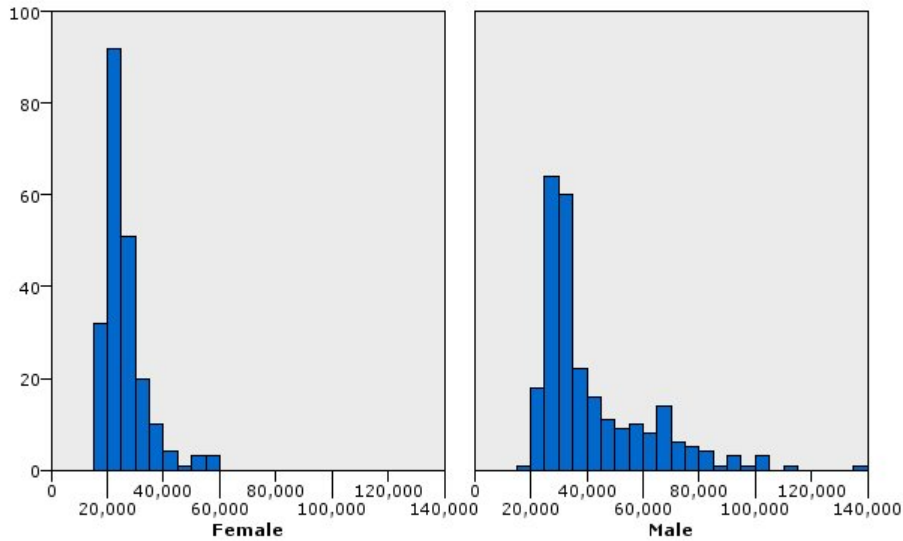


圖 10. 畫面化直方圖

我們可以觀察下列資訊：

- 兩種次數分配都不是常態分配。也就是說，直方圖不會類似鐘型曲線，常態分配才會出現鐘型曲線。
- 每個圖中，較高的長條都位於左邊。因此，無論男女，低所得者的人數都比高所得者多。
- 男性與女性之間的薪資次數分配並不相等。請注意直方圖的形狀。高所得男性人數多過於高所得女性人數。

### 範例：畫面化點形圖

點形圖和直方圖一樣，會顯示連續數值範圍的分配。但與直方圖不同的是，直方圖會顯示已 bin 資料範圍的個數，點形圖則顯示資料中的每一列/觀察值。因此，點形圖可提供比直方圖更細微的比較。事實上，分析次數分配時，最好是從點狀圖著手。

附註：本範例使用 *Employee data*。

1. 新增圖板節點並開啟進行編輯。
2. 在「基本」標籤上，選取現行薪資。
3. 選取「點形圖」。
4. 按一下「詳細資訊」標籤。
5. 在「畫面化與動畫」組別中，從「窗格跨越」下拉清單中選擇性別。
6. 按一下「執行」。
7. 將結果輸出視窗放到最大，以更清晰的檢視圖形。

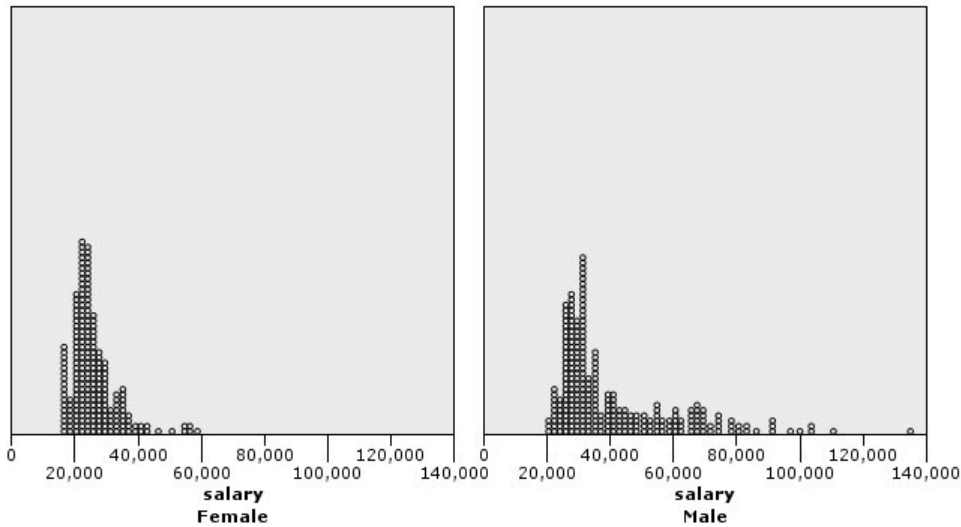


圖 11. 畫面化點形圖

與直方圖比較（請參閱第 178 頁的『範例：畫面化直方圖』），我們可以觀察下列資訊：

- 直方圖中位於 20,000 的女性高峰遠少於點形圖中。有許多觀察值/列集中在該值，但這些值中大部分比較接近 25,000。直方圖並未表現出這麼細微的程度。
- 雖然男性直方圖指出，男性平均薪資在 40,000 之後逐漸減少，但點形圖卻顯示，在該值之後一直到 80,000，分配都相當平均。在這個範圍內的任一個薪資值，每個薪資值的人數都有三個或以上。

### 範例：箱型圖

盒形圖是另一種實用的視覺化方式，可用來檢視資料分配的方式。盒形圖包含許多統計測量，讓我們在建立視覺化之後可以一一探討。

附註：本範例使用 *Employee data*。

1. 新增圖板節點並開啟進行編輯。
2. 在「基本」標籤上，選取性別 及現行薪資。（按住 Ctrl 並按一下可選取多個欄位/變數。）
3. 選取「盒形圖」。
4. 按一下「執行」。



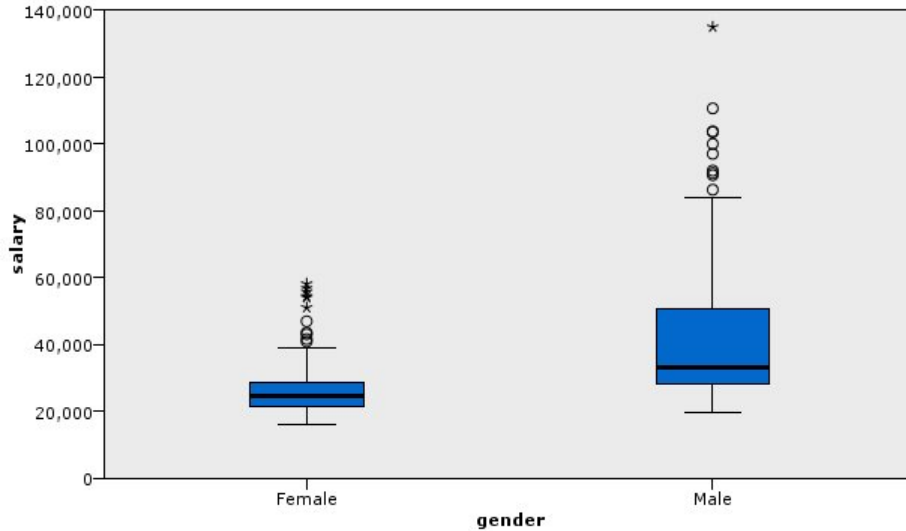


圖 12. 盒形圖

我們來看看盒形圖的不同之處：

- 盒狀中央的深色線是薪資的中位數。觀察值/列半數的值大於中位數，而半數的值小於中位數。和平均數相似的是，中位數是集中趨勢的量數。和平均數不同的是，中位數較不會受到有極端數值的觀察值/列影響。在此範例中，中位數小於平均數（與第 176 頁的『範例：具有摘要統計量的長條圖』相比）。平均數與中位數的差異表示有一些觀察值/列具有極端數值，因此拉高平均值。也就是說，有少數員工的薪資很高。
- 盒狀的底端表示第 25 個百分位數。百分之二十五的觀察值/列的值低於第 25 個百分位數。盒狀的頂端表示第 75 個百分位數。百分之七十五的觀察值/列的值高於第 75 個百分位數。這就表示，盒狀當中的就是 50% 的觀察值/列。女性的盒狀遠比男性的短。這是女性薪資變動比男性少的一個線索。盒狀的上下通常稱為樞紐。
- 從盒狀延伸的 T 長條稱為內籬或鬚。這些長條會延伸至盒狀高度的 1.5 倍，或若該範圍內都沒有觀察值/列的值，則會是值的最大值或最小值。若資料呈常態分配，內籬中應該有約 95% 的資料。在此範例中，內籬延伸情況女性少於男性，這是女性薪資變化較少的另一個指標。
- 這些點是偏離值。即被定義為未落在內籬中的值。偏離值就是極端數值。星號或星形圖案表示極端偏離值。這些觀察值/列的值超過盒狀高度三倍以上。男性和女性都有許多偏離值。請記住，本範例的平均數大於中位數。平均數較大就是由於這些偏離值所影響的。

### 範例：圓餅圖

我們現在要使用不同的資料庫探討另一些視覺化類型。*customer\_subset* 這個資料庫是假設資料檔，包含客戶相關資訊。

我們會先建立一個圓餅圖，查看各地理區域客戶的比例。

1. 新增連接到 *customer\_subset.sav* 的 Statistics 檔案來源節點。
2. 新增圖板節點並開啟進行編輯。
3. 在「基本」標籤上，選取地理指標。
4. 選取「個數圓餅」。
5. 按一下「執行」。

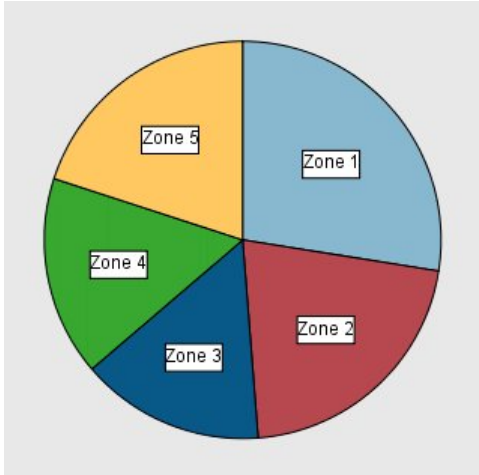


圖 13. 圓餅圖

我們可以觀察下列資訊：

- 區域 1 的客戶人數多於其他每個區域。
- 客戶平均分散於其他區域中。

### 範例：熱圖

現在我們要建立類別熱圖，以檢查不同地理區域和年齡群組客戶的平均收入。

附註：本範例使用 *customer\_subset*。

1. 新增圖板節點並開啟進行編輯。
2. 在「基本」標籤中，依序選取地理指標、年齡種類 及家庭收入（千元）。（按住 Ctrl 並按一下可選取多個欄位/變數。）
3. 選取「熱圖」。
4. 按一下「執行」。
5. 在結果輸出視窗中，按一下「顯示欄位與數值標籤」工具列按鈕（工具列中央兩群組間位於右邊者）。

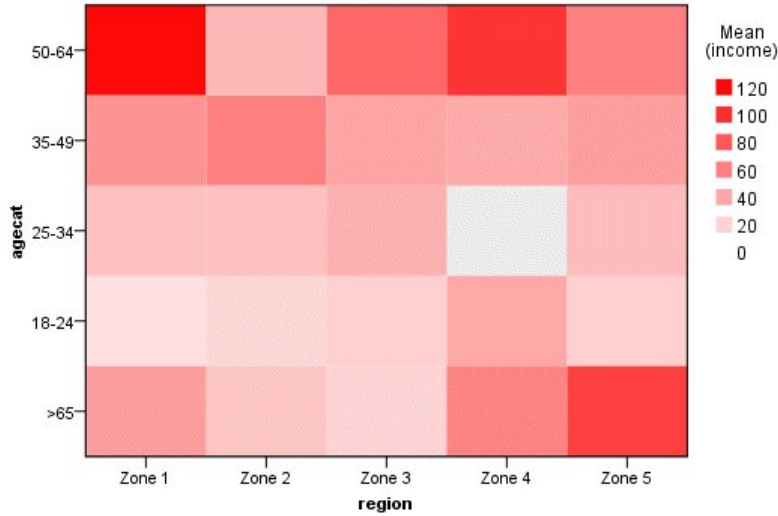


圖 14. 類別熱圖

我們可以觀察下列資訊：

- 熱圖像是一個以顏色呈現各儲存格數值的表格。明亮的深紅色代表最高值，灰色則代表低值。各儲存格的數值都是各類別配對的連續欄位/變數平均數。
- 除了區域 2 和區域 5 之外，年齡介於 50 到 64 之間的客戶群組所擁有的平均家庭收入，高於其他群組的客戶。
- 區域 4 中沒有年齡介於 25 到 34 的客戶。

### 範例：散佈圖矩陣 (SPLOM)

我們要建立一個具有許多不同變數的散佈圖矩陣，判斷資料集中的各個變數之間有沒有關聯性。

附註：本範例使用 *customer\_subset*。

1. 新增圖板節點並開啟進行編輯。
2. 在「基本」標籤中，選取年齡（歲）、家庭收入（千元）及信用卡負債（千元）。（按住 Ctrl 並按一下可選取多個欄位/變數。）
3. 選取 **SPLOM**。
4. 按一下「執行」。
5. 將輸出視窗放到最大，以更清晰的檢視矩陣。

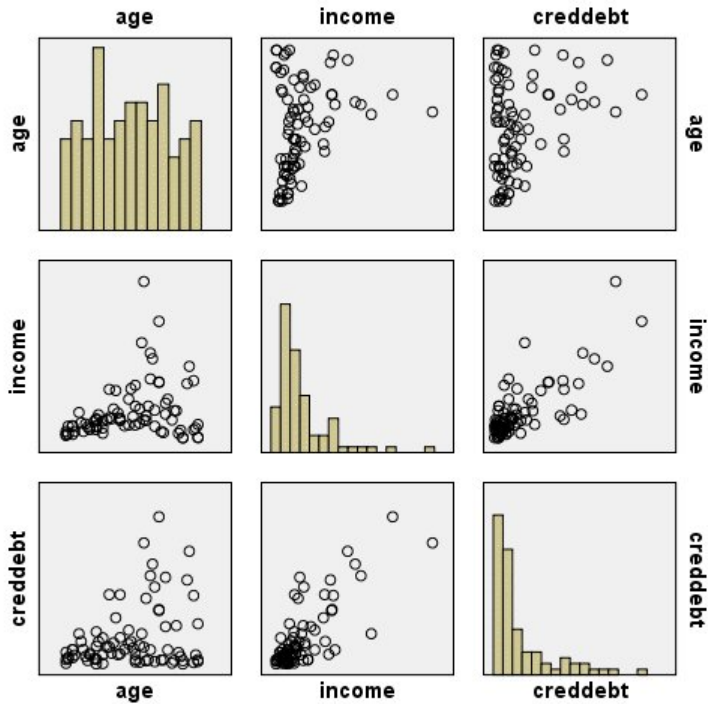


圖 15. 散佈圖矩陣 (SPLOM)

我們可以觀察下列資訊：

- 對角線上的直方圖顯示 SPLOM 中各變數的分配。*age* 的直方圖出現於左上儲存格，而 *income* 出現於中央儲存格，*creddebt* 則出現於右下儲存格。這些變數都不是常態分配。也就是說，沒有一個直方圖呈現鐘型曲線。另外，請注意 *income* 和 *creddebt* 的直方圖呈現正偏斜。
- *age* 和其他變數之間似乎沒有任何關係。
- *income* 和 *creddebt* 之間有線性關係。也就是說，*creddebt* 增加的時候，*income* 也會增加。您可以建立這些變數與其他相關變數的個別散佈圖，進一步探索其中關聯。

### 範例：總和的面量圖（色彩對映）

我們現在會建立地圖視覺化。之後，在後續範例中，我們會建立這個視覺化的變異。資料集是 *worldsales*，此為包含依洲和產品分類之銷貨收益的假設資料檔。

1. 新增圖板節點並開啟進行編輯。
2. 在「基本」標籤上，選取「洲」和「收益」。(按住 Ctrl 並按一下可選取多個欄位/變數。)
3. 選取「總和的面量圖」。
4. 按一下「詳細資訊」標籤。
5. 在「選用審美標準」組別中，從「資料標籤」下拉清單中選擇「洲」。
6. 在「地圖檔案」群組中，按一下「選取地圖檔案」。
7. 在「選取地圖」對話框中，檢查「地圖」是否設定為「洲」，且「地圖鍵」設定為「洲」。
8. 在「比較地圖」與「資料值」群組中，按一下「比較」以確定地圖鍵符合資料鍵。在本範例中，所有資料鍵值都符合地圖鍵和功能。我們也發現沒有適用於「大洋洲」的資料。
9. 在「選取地圖」對話框中，按一下「確定」。
10. 按一下「執行」。



圖 16. 總和的面量圖

從該地圖視覺化，我們可以輕易發現北美洲收益最高，而南美洲及非洲收益最低。每個洲都有加註標記，因為我們已針對資料標籤審美標準使用「洲」。

### 範例：地圖上的長條圖

本範例顯示收益如何依每個洲中的產品分列細目。

附註：本範例使用 *worldsales*。

1. 新增圖板節點並開啟進行編輯。
2. 在「基本」標籤上，選取「洲」、「產品」和「收益」。(按住 Ctrl 並按一下可選取多個欄位/變數。)
3. 選取「地圖上的長條」。
4. 按一下「詳細資訊」標籤。

使用特定類型一個以上的欄位，重要的是檢查每個欄位是否指定給正確插槽。

5. 從「類別」下拉清單中，選擇「產品」。
6. 從「值」下拉清單中，選擇「收益」。
7. 從「資料鍵」下拉清單中，選擇「洲」。
8. 從「摘要」下拉清單中，選擇「總和」。
9. 在「地圖檔案」群組中，按一下「選取地圖檔案」。
10. 在「選取地圖」對話框中，檢查「地圖」是否設定為「洲」，且「地圖鍵」設定為「洲」。
11. 在「比較地圖」與「資料值」群組中，按一下「比較」以確定地圖鍵符合資料鍵。在本範例中，所有資料鍵值都符合地圖鍵和功能。我們也發現沒有適用於「大洋洲」的資料。
12. 在「選取地圖」對話框中，按一下「確定」。

13. 按一下「執行」。
14. 將結果輸出視窗放到最大，以更清晰的檢視顯示。

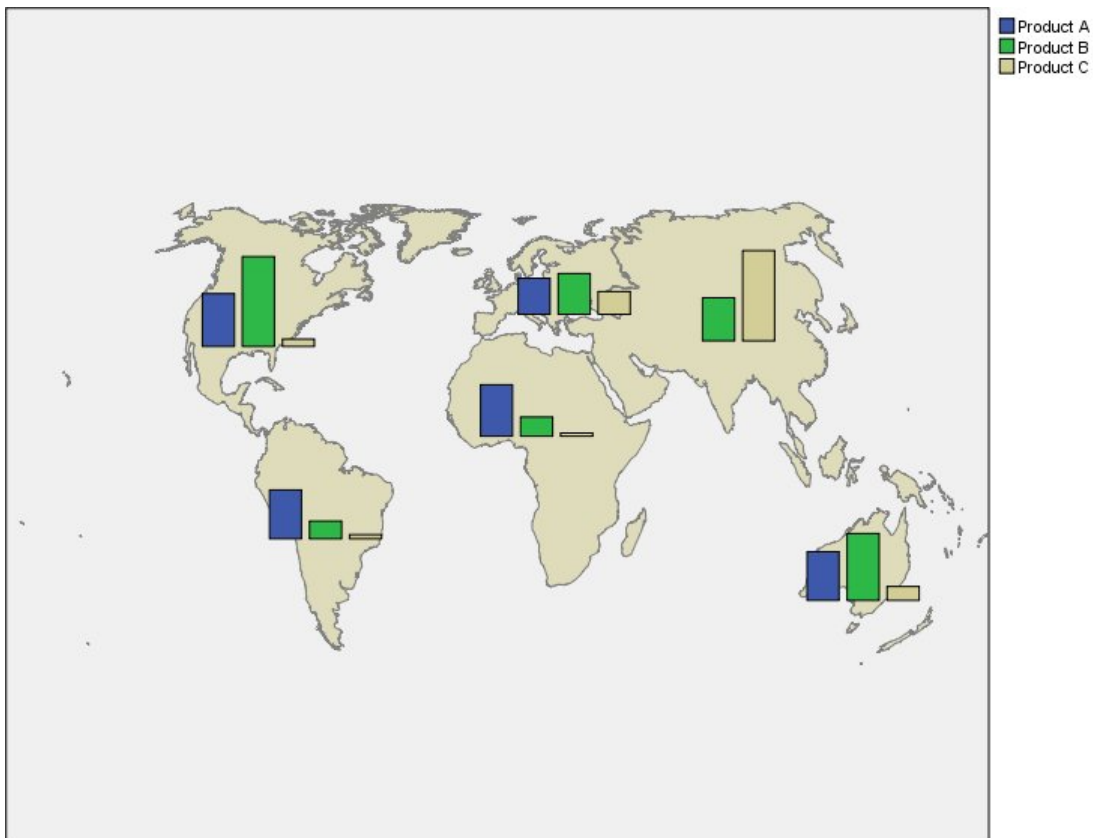


圖 17. 地圖上的長條圖

我們可以觀察下列資訊：

- 在產品中的總收益分配與南美洲及非洲的分配非常類似。
- 產品 C 會在所有地方產生最低收益 (亞洲除外)。
- 產品 A 在亞洲沒有收益或收益非常微小。

## 圖版「外觀」標籤

您可以在建立圖形之前指定外觀選項。

### 一般外觀選項

**標題。** 輸入要用作圖表標題的文字。

**次標題。** 輸入要用作圖表次標題的文字。

**標題。** 輸入要用作圖表標題的文字。

**取樣。** 為大的資料集指定一種方法。可以指定上限資料集大小，或使用預設記錄數。如果選取**抽樣**選項，那麼處理大資料集時的效能將會提高。或者，您可以選擇透過選取**使用所有資料**來繪製所有資料點，但您應該注意到這樣可能會極大地降低軟體效能。

## 樣式表外觀選項

有兩個按鈕讓您控制哪些視覺化範本（以及樣式表與地圖）可用：

**管理。**管理您電腦中的視覺化範本、樣式表和地圖。您可以匯入、匯出、重新命名或刪除您本機電腦中的視覺化範本、樣式表與地圖。請參閱第 188 頁的『管理範本、樣式表與地圖檔案』主題，以取得更多資訊。

**位置。**變更儲存視覺化範本、樣式表與地圖的位置。目前位置列於按鈕右邊。請參閱『設定範本、樣式表與地圖的位置』主題，以取得更多資訊。

下列範例顯示外觀選項在圖表上的位置。（注意：並非所有圖形都使用所有這些選項。）

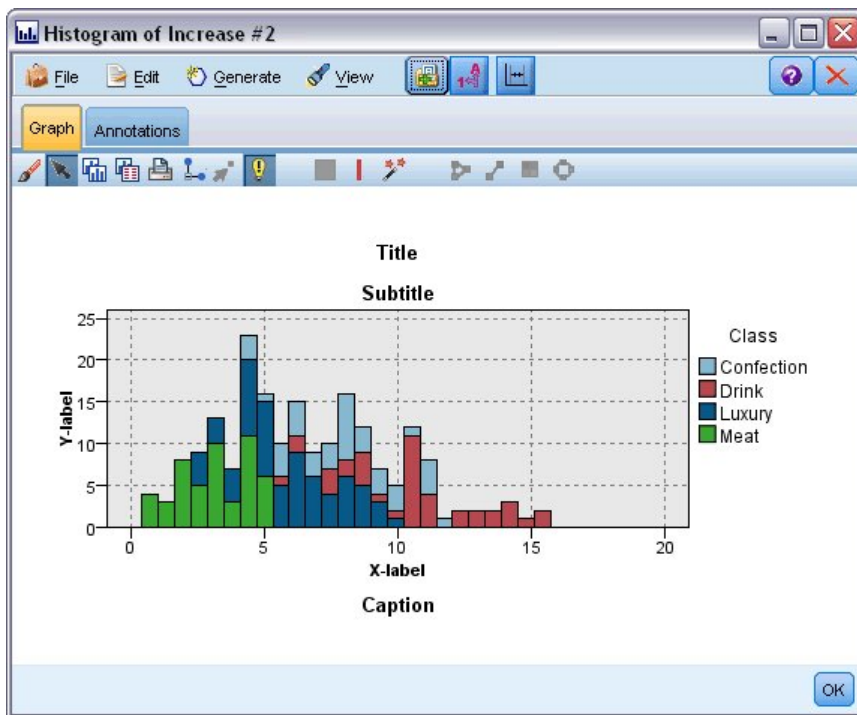


圖 18. 各個圖表外觀選項的位置

## 設定範本、樣式表與地圖的位置

視覺化範本、視覺化樣式表與地圖檔案儲存於特定本機資料夾或 IBM SPSS Collaboration and Deployment Services 儲存庫。選取範本、樣式表與地圖時，只會顯示位於此位置的內建項目。將所有範本、樣式表及地圖檔案保留在同一個位置，可讓 IBM SPSS 應用程式存取更容易。如需將其他範本、樣式表與地圖檔案新增到此位置的更多資訊，請參閱第 188 頁的『管理範本、樣式表與地圖檔案』。

如何設定範本、樣式表與地圖檔案的位置

1. 在範本或樣式表對話框中，按一下「位置...」，顯示範本、樣式表與地圖對話框。
2. 選取範本、樣式表與地圖檔案預設位置的選項：

**本機電腦。**範本、樣式表與地圖檔案位於本機電腦的特定資料夾中。在 Windows XP 上，這個資料夾是 *C:\Documents and Settings\<user>\Application Data\SPSSInc\Graphboard*。此資料夾無法變更。



**IBM SPSS Collaboration and Deployment Services 儲存庫** 範本、樣式表及地圖檔案位於 IBM SPSS Collaboration and Deployment Services 儲存庫 中的使用者指定資料夾。若要識別此指定資料夾，請按一下「資料夾」。如需相關資訊，請參閱『使用IBM SPSS Collaboration and Deployment Services 儲存庫作為範本、樣式表與地圖檔案位置』。

3. 按一下**確定**。

## 使用IBM SPSS Collaboration and Deployment Services 儲存庫作為範本、樣式表與地圖檔案位置

視覺化範本與樣式表可儲存於IBM SPSS Collaboration and Deployment Services 儲存庫。此位置是IBM SPSS Collaboration and Deployment Services 儲存庫中的指定資料夾。若此位置被設為預設位置，則此處所有的範本、樣式表與地圖檔案都可供選用。

如何將 IBM SPSS Collaboration and Deployment Services 儲存庫 中的資料夾設定為範本、樣式表與地圖檔案的位置

1. 在「位置」，按一下「位置...」。
2. 選取 IBM SPSS Collaboration and Deployment Services 儲存庫。
3. 按一下「資料夾」。

附註：如果您尚未連接至 IBM SPSS Collaboration and Deployment Services 儲存庫，將會提示您提供連線資訊。

4. 在「選取資料夾」對話框中，選取要儲存範本、樣式表與地圖檔案的資料夾。
5. 或者，您可以從擷取標籤選取標籤。只有具備此標記的範本、樣式表與地圖檔案會顯示出來。
6. 若您要尋找包含特定範本或樣式表的資料夾，可以在「搜尋」標籤中搜尋範本、樣式表或地圖檔案。「選取資料夾」對話框會自動選取所找範本、樣式表或地圖檔案所在的資料夾。
7. 按一下「選取資料夾」。

## 管理範本、樣式表與地圖檔案

您可以使用「管理範本、樣式表與地圖」對話框，管理您電腦本機位置中的範本、樣式表及地圖檔案。此對話框可讓您匯入、匯出、更名與刪除您電腦本機位置中的視覺化範本、樣式表與地圖檔案。

按一下您選取範本、樣式表或地圖的對話框之一中的「管理...」。

### 管理範本、樣式表與地圖對話框

「範本」標籤會列出所有本機範本。「樣式表」標籤會列出所有本機樣式表，並以樣本資料顯示視覺化範例。您可選取樣式表之一，將其樣式套用到視覺化範例。如需相關資訊，請參閱主題第 258 頁的『套用樣式表』。「地圖」標籤會列出所有本機地圖檔案。此標籤也會顯示地圖鍵，包括樣本值、備註 (如果建立地圖時提供了備註)，以及地圖的預覽。

下列按鈕會作用於目前作用中的標籤。

**匯入**。從檔案系統匯入視覺化範本、樣式表或地圖檔案。匯入範本、樣式表或地圖檔案，就能在 IBM SPSS 應用程式中使用它們。如果其他使用者將範本、樣式表或地圖檔案傳送給您，您必須先將檔案匯入，才能在應用程式中使用。

**匯出**。將視覺化範本、樣式表或地圖檔案匯出至檔案系統。若您希望將範本、樣式表或地圖檔案傳送給其他使用者，可執行匯出。

**更名**。重新命名選取的視覺化範本、樣式表或地圖檔案。您不能將名稱變更為目前已使用的名稱之一。



**匯出地圖鍵。**將地圖鍵以逗點分隔值 (CSV) 檔案匯出。此按鈕僅能在「地圖」標籤上啟用。

**刪除。**刪除選取的視覺化範本、樣式表或地圖檔案。您可以在 Windows 與 Linux 上按住 Ctrl 並按一下，來選取多個範本、樣式表或地圖檔案。刪除動作無法復原，請小心使用。

---

## 轉換及發佈地圖形狀檔案

「圖表板樣本選擇器」可讓您從視覺化範本與 SMZ 檔案的組合建立地圖視覺化。SMZ 檔案與 ESRI 形狀檔案 (SHP 檔案格式) 類似，其中包含了繪製地圖的地理資訊 (例如國界)，但會對地圖視覺化進行最佳化。「圖表板樣本選擇器」會預先安裝一些經過選取的 SMZ 檔案。如果您有一些現有的 ESRI 形狀檔案想要用於地圖視覺化，您首先需要使用「地圖轉換公用程式」將形狀檔案轉換為 SMZ 檔案。「地圖轉換公用程式」支援包含單一圖層的點、多線或多邊形 (形狀類型 1、3 與 5) ESRI 形狀檔案。

除了轉換 ESRI 形狀檔案以外，「地圖轉換公用程式」可讓您修改地圖的細節程度、變更特徵標記、合併特徵及移動特徵，另外還可進行其他許多選擇性的變更。您也可以使用「地圖轉換公用程式」來修改現有 SMZ 檔案 (包括預先安裝的檔案)。

編輯預先安裝的 SMZ 檔案

1. 自「管理」系統匯出 SMZ 檔案。請參閱第 188 頁的『管理範本、樣式表與地圖檔案』主題，以取得更多資訊。
2. 使用「地圖轉換公用程式」開啟和編輯匯出的 SMZ 檔案。建議您以不同的名稱儲存檔案。請參閱第 190 頁的『使用地圖轉換公用程式』主題，以取得更多資訊。
3. 將修改後的 SMZ 檔案匯入至「管理」系統。請參閱第 188 頁的『管理範本、樣式表與地圖檔案』主題，以取得更多資訊。

地圖檔案的其他資源

SHP 檔案格式中的地理空間資料，適用於支援您的對應需求，可從許多私有和公用來源取得。如果您要尋找免費資料，請造訪當地政府網站。此產品中的許多範本是以從 GeoCommons ( ) 及美國人口普查局 (<http://www.census.gov>) 公開取得的資料為依據。

**重要注意事項：**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 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。有關非 IBM 產品的性能問題應直接洽詢該產品供應商。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。那些網站中的資料不是此 IBM 程式的資料，除非隨附於此 IBM 程式之「通知」檔案有所指定，且使用資料網站之風險由您自行承擔。

## 地圖的關鍵概念

瞭解與形狀檔案相關的某些關鍵概念將可協助您有效使用「地圖轉換公用程式」。

「形狀檔案」提供繪製地圖的地理資訊。「地圖轉換公用程式」支援三種類型的形狀檔案：

- **點。** 形狀檔案可識別點的位置，例如城市。
- **折線。** 形狀檔案可識別路徑及其位置，例如河流。
- **多邊形。** 形狀檔案可識別邊界區域及其位置，例如國家。

您最常使用的是多邊形形狀檔案。面量圖地圖由多邊形形狀檔案建立。面量圖地圖使用顏色來代表個別多邊形 (區域) 內的值。點與多線形狀檔案一般會重疊在多邊形形狀檔案上。例如重疊在美國州多邊形形狀檔案上的美國城市的點形狀檔案。

形狀檔案由「特徵」組成。特徵是個別的地理圖實體。例如，特徵可能是國家、州、城市等。形狀檔案也包含有關特徵的資料。這些資料儲存在「屬性」中。屬性與資料檔中的欄位或變數類似。至少有一個屬性是特徵的「地圖鍵」。地圖鍵可能是標記，例如國家或州名稱。您會將地圖鍵鏈結至資料檔中的變數/欄位以建立地圖視覺化。

請注意，您將只能保留 SMZ 檔案中的鍵屬性或屬性。「地圖轉換公用程式」不支援儲存其他屬性。這表示如果您要在不同的程度上整合，您將需要建立多個 SMZ 檔案。例如，如果您要聚集美國的州與區域，您將需要分離 SMZ 檔案：一個 SMZ 檔案用鍵來識別州，一個 SMZ 檔案用鍵來識別區域。

## 使用地圖轉換公用程式

如何啟動「地圖轉換公用程式」

在功能表上，選擇：

工具 > 地圖轉換公用程式

「地圖轉換公用程式」中總共有四個主要的畫面 (步驟)。其中一個步驟還牽涉到一些子步驟，可以更詳細地控制地圖檔案的編輯。

### 步驟 1 - 選擇目的地與來源檔案

您首先必須針對轉換的地圖檔案選取來源地圖檔案與目的地。對於形狀檔案，您將需要 *.shp* 與 *.dbf* 檔案。

針對轉換選取 *.shp* (ESRI) 或 *.smz* 檔案。瀏覽到電腦上的現有地圖檔案。這是您將轉換為並儲存為 SMZ 檔案的檔案。形狀檔案的 *.dbf* 檔案必須儲存在基本檔案名稱與 *.shp* 檔案相符的相同位置。必須要有 *.dbf* 檔案，因為其中包含 *.shp* 檔案的屬性資訊。

為轉換的地圖檔案設定目的地及檔案名稱。為將從原始地圖來源建立的 SMZ 檔案輸入路徑與檔案名稱。

- 匯入至「範本選擇器」。除了在檔案系統上儲存檔案以外，您可以選擇性地將地圖新增至「範本選擇器」的「管理」清單。如果您選擇此選項，地圖將會針對安裝在電腦上的 IBM SPSS 產品自動出現在「範本選擇器」中。若您現在不匯入至「範本選擇器」，稍後您必須手動匯入。如需將地圖匯入至「範本選擇器管理」系統的相關資訊，請參閱第 188 頁的『管理範本、樣式表與地圖檔案』。

### 步驟 2 - 選擇地圖鍵

現在您將選擇要在 SMZ 檔案中包含哪些地圖鍵。然後您可以變更將對地圖的呈現產生影響的某些選項。「地圖轉換公用程式」中的後續步驟包括地圖的預覽。您所選擇的呈現選項將用於產生地圖預覽。

**選擇主地圖鍵。** 選取主要索引鍵用來識別及標記地圖中的特徵的屬性。例如，世界地圖的主要索引鍵可能是識別國家/地區名稱的屬性。主要索引鍵也會將您的資料鏈結至地圖特徵，因此請確保您所選擇的屬性值 (標籤) 將與資料中的值相符。當您選擇屬性時，會顯示範例標記。如果您需要變更這些標記，您將可在稍後步驟中執行。

**選擇要包括的其他鍵。** 除了地圖主索引鍵以外，請檢查您要包括在所產生的 SMZ 檔案中的其他任何鍵屬性。例如，某些屬性可能包含翻譯的標記。如果您期望用其他語言對資料進行編碼，您可能想要保留這些屬性。請注意，您只能選擇將相同特徵表示為主要索引鍵的其他索引鍵。例如，如果主要索引鍵為美國各州的全名，則您只能選取表示美國各州的替代索引鍵，例如：州的縮寫。

**自動平滑處理地圖。** 含有多邊形的形狀檔案一般包含統計量地圖視覺化的太多資料點，以及太多詳細資訊。過量的詳細資訊可能會導致分心，以及對效能造成負面影響。您可以降低詳細的程度，並概化進行平滑處理的地

圖。地圖的顆粒看起來會比較粗，因此呈現的速度會比較快。當對地圖自動進行平滑處理時，最大角度為 15 度，所保持的百分比為 99。如需這些設定的相關資訊，請參閱『平滑處理地圖』。請注意，在稍後的另一個步驟中，您還有機會套用其他平滑處理。

**移除相同功能中彼此接觸的多邊形之間的邊界。** 某些特徵可能包括子特徵，其邊界位於感興趣的主特徵的內部。例如，由各大洲構成的世界地圖可能包含國家的內部邊界，而這些邊界也是每個洲的邊界。如果您選擇此選項，內部邊界將不會顯示在地圖中。在由各大洲構成的世界地圖這個範例中，選擇此選項會移除國家邊界，同時保留洲的邊界。

### 步驟 3 - 編輯地圖

當您為地圖指定基本選項之後，便可編輯更多特定選項。這些修改是選擇性的。「地圖轉換公用程式」的此步驟可引導您完成相關工作，並顯示地圖的預覽，以便能夠確認您的變更。視形狀檔案類型 (點、多線或多邊形) 與座標系統而定，可能無法使用某些工作。

每個工作在「地圖轉換公用程式」的左側都有以下常用控制項。

**在地圖上顯示標籤。** 依預設，特徵標記不會顯示在預覽中。您可以選擇顯示標籤。雖然標籤可以幫助您識別特徵，但它們可能會對在預覽地圖上進行直接選取產生干擾。當您有需要時，例如當您正在編輯特徵標記時，請開啟此選項。

**為地圖預覽上色。** 依預設，預覽地圖會顯示有實色的區域。所有特徵的顏色都一樣。您可以選擇將顏色分類指派給個別的地圖特徵。此選項有助於分辨地圖中的不同特徵。當您合併特徵以及要看見新特徵如何在預覽中呈現時，此選項特別有幫助。

每個工作在「地圖轉換公用程式」的右側都有以下常用控制項。

**復原。** 按一下復原以恢復至上一個狀態。您最多可以復原 100 次變更。

**平滑處理地圖：** 含有多邊形的形狀檔案一般包含統計量地圖視覺化的太多資料點，以及太多詳細資訊。過量的詳細資訊可能會導致分心，以及對效能造成負面影響。您可以降低詳細的程度，並概化進行平滑處理的地圖。地圖的顆粒看起來會比較粗，因此呈現的速度會比較快。點與多線地圖無法使用此選項。

**最大角度。** 最大角度必須是介於 1 與 20 之間的值，並可指定平滑處理幾乎為線性的點集的容差。較大的值允許線性平滑處理有更大的容差，並將導致丟棄更多的點，產生更為概化的地圖。若要套用線性平滑處理，「地圖轉換公用程式」會檢查地圖中由每一組三個點構成的內角。如果 180 減角度小於指定值，「地圖轉換公用程式」將會丟棄中間點。換句話說，「地圖轉換公用程式」會檢查由三點構成的線是否幾乎為直線。如果是，「地圖轉換公用程式」會將這條線當成端點之間的直線處理，並丟棄中間點。

**保持的百分比。** 保持的百分比必須是介於 90 與 100 之間的值，其可決定對地圖進行平滑處理時要保留的土地面積量。此選項只對有多邊形的特徵產生影響，就像特徵包括島的情況一樣。如果特徵的總面積減多邊形大於指定的原始區域百分比，「地圖轉換公用程式」將會丟棄地圖中的多邊形。「地圖轉換公用程式」將永遠不會移除特徵的所有多邊形。也就是說，特徵永遠都至少有一個多邊形，不管套用的平滑處理量是多少。

當您選擇要保留的最大角度與百分比之後，請按一下「套用」。預覽會隨著平滑處理的變更而更新。如果您需要再次平滑處理地圖，請重複到取得平滑處理的所需程度為止。請注意，平滑處理有限制。如果您重複平滑處理，總會到達一個極限，無法再將其他平滑處理套用至地圖。

**編輯特徵標記：** 您可以視需要編輯特徵標記 (或許用以符合您的預期資料)，也可以在地圖中重新放置標記。即使您認為不需要變更標記，您也應該在從地圖中建立視覺化之前檢閱標記。由於依預設不會在預覽中顯示標記，因此您也可能會想要選取「在地圖上顯示標記」來顯示標記。

**鍵。** 選取包含您想要檢閱及/或編輯的特徵標記的鍵。

**功能。** 此清單顯示所選鍵中所包含的特徵標記。若要編輯標記，請在清單中連按兩下該標記。如果標記顯示在地圖上，您也可以直接在地圖預覽中連按兩下特徵標記。如果您要比較標記與實際資料檔，請按一下「比較」。

**X/Y。** 這些文字框列出了地圖上所選功能標籤的現行中心點。這些單元會顯示在地圖的座標中。這些可能是本機的笛卡兒座標 (例如州平面座標系統) 或地理座標 (其中 **X** 是經度，**Y** 是緯度)。請輸入標記新位置的座標。如果顯示標記，您也可以按一下並拖曳地圖上的標記來移動它。文字框會隨著新位置而更新。

**比較。** 如果您想要使資料檔中所包含的資料值與特殊鍵的特徵標記相符，請按一下「比較」來顯示「比較外部資料來源」對話框。在此對話框中，您將能夠開啟資料檔，並直接將其值與地圖鍵的特徵標記中的值進行比較。

**「比較外部資料來源」對話框：** 「比較外部資料來源」對話框可以讓您開啟以分隔符號分隔的值檔案 (副檔名為 *.txt*)、以逗點分隔的值檔案 (副檔名為 *.csv*) 或格式為 IBM SPSS Statistics 的資料檔 (副檔名為 *.sav*)。開啟檔案時，您可以選取資料檔中的欄位來與特定地圖鍵中的特徵標記進行比較。然後您可以更正地圖檔案中的任何差異。

**資料檔中的欄位。** 選擇您要與特徵標記比較值的欄位。如果 *.txt* 或 *.csv* 檔案中的第一列包含每一個欄位的描述性標記，請勾選「使用第一列作為欄標籤」。否則，每一個欄位都將由其在資料檔中的位置識別 (例如「行 1」、「行 2」等)。

**要比較的鍵。** 選擇您要與資料檔欄位值比較特徵標記的地圖鍵。

**比較。** 當您準備好，要比較值時，請按一下。

**比較結果。** 依預設，「比較結果」表格僅列出資料檔中的不相符欄位值。應用程式通常會嘗試透過檢查插入或遺失的空格，來尋找相關特徵標記。按一下「地圖標記」行中的下拉清單，來使地圖檔案中的特徵標記與顯示的欄位值相符。如果地圖檔案中沒有對應特徵標記，請選擇「保持不相符」。如果您想要看見所有欄位值，即使已與功能標籤相符，也請清除僅顯示不相符的觀察值。若您想要置換一個或多個相符項目，可以執行此操作。

您只能使用每個特徵一次，以使其與欄位值相符。如果您要使多個特徵與單一欄位值相符，您可以合併特徵，然後使新的合併特徵與欄位值相符。如需合併功能的相關資訊，請參閱『合併特徵』。

**合併特徵：** 合併特徵對於在地圖中建立較大區域而言很有幫助。例如，如果您要轉換州地圖，可以將州 (本範例中的特徵) 合併為較大的北、南、東與西區域。

**鍵。** 選取包含特徵標記的地圖鍵將可協助您識別您要合併的特徵。

**功能。** 按一下您要合併的第一個特徵。按住 **Ctrl** 並按一下您要合併的其他特徵。請注意，也將會在地圖預覽中選取功能。除了從清單中選取以外，您也可以在地圖預覽中直接按一下特徵，然後按住 **Ctrl** 並按一下 特徵。

選取您要合併的特徵之後，請按一下「合併」以顯示「命名合併特徵」對話框，您可以在此將標記套用至新特徵。合併特徵之後，您可能會想要勾選「為地圖預覽上色」，以確保獲得您預期的結果。

合併特徵之後，您也可以移動新特徵的標記。您可以在「編輯特徵標記」工作中執行此操作。請參閱第 191 頁的『編輯特徵標記』主題，以取得更多資訊。

**「命名合併特徵」對話框：** 「命名合併特徵」對話框可以讓您將標記指派給新合併特徵。

「標記」表格顯示地圖檔案中每一個鍵的資訊，並可讓您為每一個鍵指派標記。

**新標籤。** 為合併特徵輸入新標記來指派給特定地圖鍵。

鍵。 您為其指派新標記的地圖鍵。

舊標籤。 將合併至新特徵的特徵的標記。

移除彼此接觸的多邊形之間的邊界。 勾選此選項可移除合併在一起的特徵的邊界。例如，如果您將州合併至地理區域，此選項會移除個別州周圍的邊界。

移動特徵： 您可以移動地圖中的特徵。當您要將特徵放在一起時，例如大陸與散佈於外海的島嶼，此功能就很有用。

鍵。 選取包含特徵標記的地圖鍵將可協助您識別要移動的特徵。

功能。 按一下您要移動的特徵。請注意，將會在地圖預覽中選取特徵。您也可以直接在地圖預覽中按一下特徵。

**X/Y**。 這些文字框列出了地圖上功能的現行中心點。這些單元會顯示在地圖的座標中。這些可能是本機的笛卡兒座標 (例如州平面座標系統) 或地理座標 (其中 **X** 是經度，**Y** 是緯度)。請輸入特徵新位置的座標。您也可以按一下並拖曳地圖上的特徵來移動它。文字框會隨著新位置而更新。

刪除特徵： 您可以刪除地圖中不想要的特徵。當您要刪除在地圖視覺化中不會感興趣的特徵，來移除某些擁擠的情形時，此功能很有幫助。

鍵。 選取包含特徵標記的地圖鍵將可協助您識別要刪除的特徵。

功能。 按一下您要刪除的特徵。如果您要同時刪除多個特徵，請按住 **Ctrl** 並按一下其他特徵。請注意，也會在地圖預覽中選取功能。除了從清單中選取以外，您也可以在地圖預覽中直接按一下特徵，然後按住 **Ctrl** 並按一下 特徵。

刪除個別元素： 除了刪除完整特徵以外，您也可以刪除構成特徵的某些個別元素，例如湖泊與小島。點地圖無法使用此選項。

元素。 按一下您要刪除的元素。如果您要同時刪除多個元素，請按住 **Ctrl** 並按一下其他元素。請注意，也會在地圖預覽中選取元素。除了從清單中選取以外，您也可以在地圖預覽中直接按一下元素，然後按住 **Ctrl** 並按一下 元素。由於元素名稱清單不是描述性的 (會為功能內的每一個元素指派一個數字)，您應檢查地圖預覽中的選取項目，以確保您已選取所需的元素。

### 設定投影：

地圖投影指定了以二維形式表現三維地球的方式。所有投影都會導致失真。但是，根據您所檢視的是世界地圖還是本地圖，某些投影會更加適合。同時，某些投影還會保留原始特徵的形狀。保留形狀的投影叫做正形投影。只有具有地理座標 (經度與緯度) 的地圖才能使用此選項。

與「地圖轉換公用程式」中的其他選項不同，投影可在建立地圖視覺化之後變更。

投射。 選取地圖投影。如果您正在建立世界或半球地圖，請使用「本地」、「麥卡托」或「溫科爾」投影法。對於較小的地區，請使用「本地」、「蘭伯特正形」或「橫麥卡托」投影法。所有投影都針對基準使用 WGS83 橢面。

- 「本地」投影法始終用於本地座標系統 (例如州平面座標系統) 建立時。這些座標系統由笛卡兒座標定義，而非由地理座標 (經度與緯度) 定義。在「本地」投影法中，水平與垂直線會在笛卡兒座標系統中平均分佈。「本地」投影法並非正形。
- 「麥卡托」投影法是世界地圖的正形投影法。水平與垂直線是直線，且始終彼此成直角。請注意，當「麥卡托」投影法接近北極與南極時，會向無窮遠延伸，因此如果您的地圖包含北極或南極，便無法使用此投影法。當地圖接近這些限制時，失真會達到最大。

- 「溫科爾」投影法是世界地圖的非正形投影法。雖然它不是正形的方法，但它也在形狀與大小之間提供了不錯的平衡。除了「赤道」與「本初子午線」以外，所有線都是曲線。如果您的世界地圖中包括北極或南極，這便是一種不錯的投影選擇。
- 如其名所示，「蘭伯特正形」投影法是一種正形投影法，可用於東西長於南北的大陸或較小地塊之地圖製作。
- 「橫麥卡托」是另一種適用於大陸或較小地塊之地圖製作的正形投影法。請針對南北長於東西的地塊使用此投影法。

## 步驟 4 - 完成

到目前為止，您可以新增備註來描述地圖檔案，也可以從地圖鍵中建立樣本資料檔。

**地圖鍵。** 如果地圖檔案中有多個鍵，請選取您要在預覽中顯示特徵標記的地圖鍵。如果您從地圖中建立資料檔，將會針對資料值使用這些標記。

**備註。** 輸入描述地圖或提供可能與您的使用者相關的其他資訊 (例如原始形狀檔案的來源) 的備註。備註將會顯示在「圖表板樣本選擇器」的「管理」系統中。

**從功能標籤中建立資料集。** 如果您要從顯示的特徵標記中建立資料檔，請勾選此選項。當您按一下瀏覽...時，您就能指定位置及檔案名稱。如果您新增 *.txt* 副檔名，檔案將會另存為以分隔符號分隔的值檔案。如果您新增 *.csv* 副檔名，檔案將會另存為以逗點分隔的值檔案。如果您新增 *.sav* 副檔名，檔案將會以 IBM SPSS Statistics 格式儲存。如果沒有指定任何副檔名，SAV 則為預設值。

## 發佈地圖檔案

在「地圖轉換公用程式」的第一步驟中，您選擇了儲存轉換的 SMZ 檔案的位置。您可能也已經選擇將地圖新增至「圖表板樣本選擇器」的「管理」系統。如果您選擇儲存至「管理」系統，您也可以您在相同電腦上執行的任何 IBM SPSS 產品中看到該地圖。

若要將地圖發佈給其他使用者，您需要將 SMZ 傳送給這些使用者。然後這些使用者可以使用「管理」系統來匯入地圖。您可以直接傳送您在步驟 1 中指定位置的檔案。如果您要傳送的檔案位於「管理」系統中，您首先需要將其匯出：

1. 在「範本選擇器」中，按一下管理...
2. 按一下「地圖」標籤。
3. 選取您想發佈的地圖。
4. 按一下「匯出...」並選擇您要儲存檔案的位置。

現在您可以將實際地圖檔案傳送給其他使用者。使用者將需要反向執行此程序，並將地圖匯入至「管理」系統。

---

## 統計圖節點

「統計圖」節點可顯示各數字欄位之間的關係。可使用點建立圖形 (即散佈圖)，或者也可使用行段。可通過在對話框中指定一種 X 模式來建立三種類型的行統計圖。

X 模式 = 排序

如果將 X 模式設定為排序，則資料將按照繪製在 x 軸上的欄位的值進行排序。這將在圖表上畫出一條貫穿左右的行。如果將列名欄位用作套版，則將在圖表上生成多條不同色調的、貫穿左右的行。

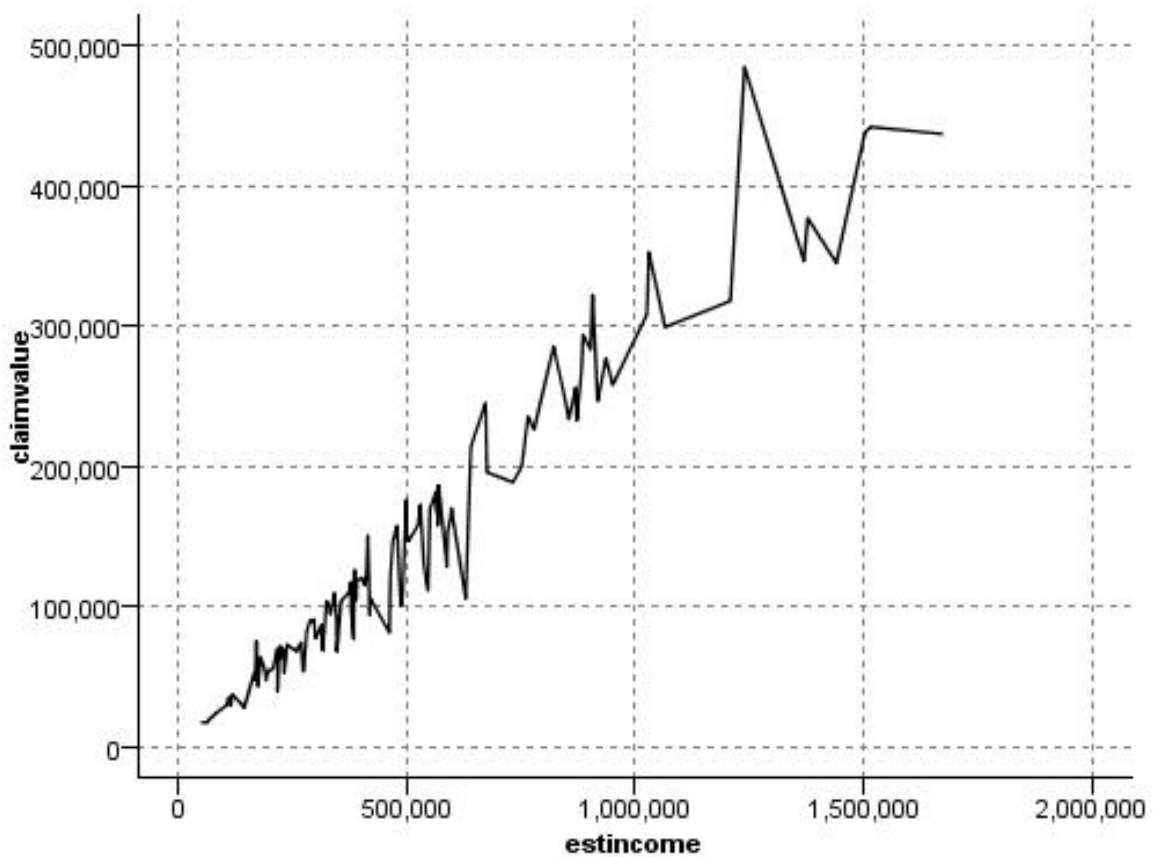


圖 19. X 模式設定為「排序」的行統計圖

X 模式 = 套版

如果將 X 模式設定為 **套版**，則將在同一圖表上建立多行統計圖。不會對套版統計圖的資料進行排序；只要  $x$  上的值不斷增大，資料將繪製在一條行上。如果值減小，那麼將開始一條新行。例如，如果  $x$  從 0 增大到 100，則  $y$  值將繪製在一條行上。 $x$  降低為 100 以下方時，則在第一條行之外，將繪製一條新行。完成的統計圖可能有多個統計圖，它們可用於對比多個  $y$  值數列。此類型的統計圖對具有期間時間成分的資料很有用，比如連續 24 小時期間的用電需求。

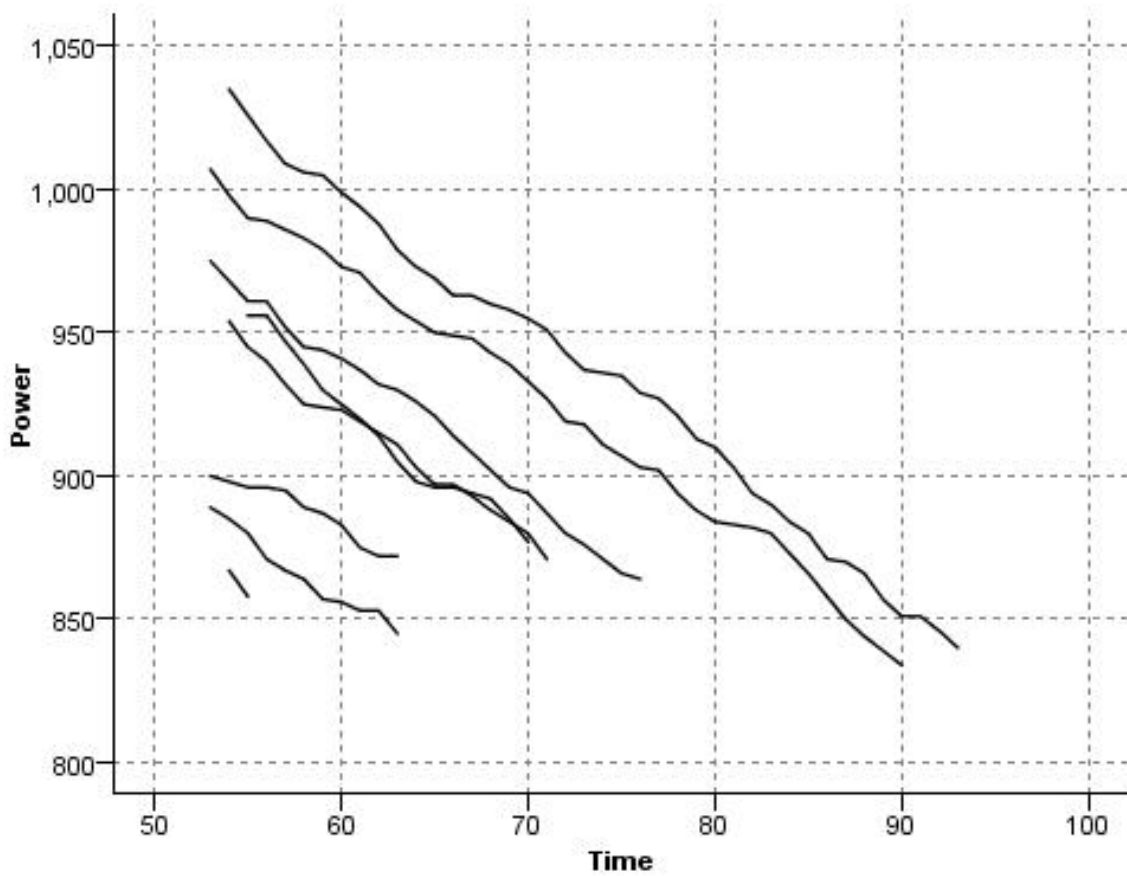


圖 20. X 模式設定為「套版」的行統計圖

X 模式 = 如所讀取

將 X 模式設定為 **如所讀取**，統計圖的  $x$  和  $y$  值將與從資料來源讀取的值一樣。對於具有時間序列成分的資料，此選項將有助於您研究與資料順序關聯的趨勢和樣式。可能需要在創建此類型的統計圖之前對資料進行排序。它也可以用於對比設定為**排序**和**如所讀取**的 X 模式，以便確定樣式對排序的依賴程度。



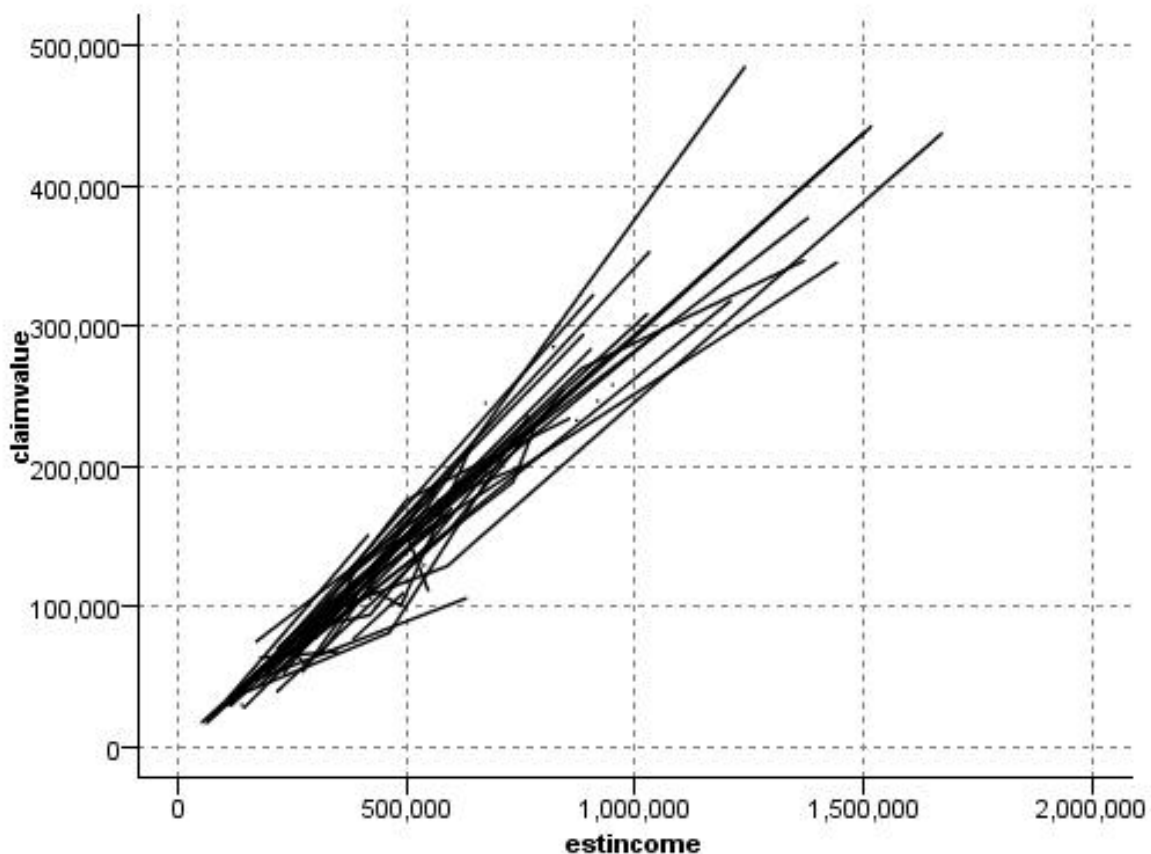


圖 21. 行統計圖起先顯示為「排序」，隨後以「如所讀取」的 X 模式再次執行

也可使用「圖版」節點生成散佈圖和行散佈圖。但是，在此節點中，有多個選項供您選擇。請參閱第 170 頁的『可用的內建圖板視覺化類型』主題，以取得更多資訊。

### 統計圖節點標籤

統計圖對照 X 欄位的值，顯示 Y 的值。通常而言，這兩個欄位分別對應於一個應變數和一個自變數。

**X 欄位。** 從清單中，選取要在水平  $x$  軸上顯示的欄位。

**Y 欄位。** 從清單中選取顯示在垂直  $y$  軸上的欄位。

**Z 欄位。** 按一下 3D 圖表按鈕後，可以從清單中選擇要顯示在  $z$  軸上的欄位。

**重疊。** 有數種方式來說明資料值的種類。例如，可以使用 *maincrop* 作為顏色套版來指定補貼申請人種植的主要作物的 *estincome* 和 *claimvalue* 值。請參閱第 162 頁的『外觀、套版、畫面和動畫』主題，以取得更多資訊。

**套版類型。** 指定是否顯示套版函數或光滑線。光滑線和套版函數總是作為  $y$  的函數來計算。

- 無。不顯示任何套版。
- 光滑線。顯示平滑的擬合行，該擬合行用區域加權迭代穩健最小二乘迴歸方法 (LOESS) 來計算。此方法可有效計算一系列的迴歸方法，每個迴歸方法均集中關注統計圖內的一個小區域。此操作將生成一連串「區

域」回歸行，然後可將這些行結合起來形成一條平滑曲線。

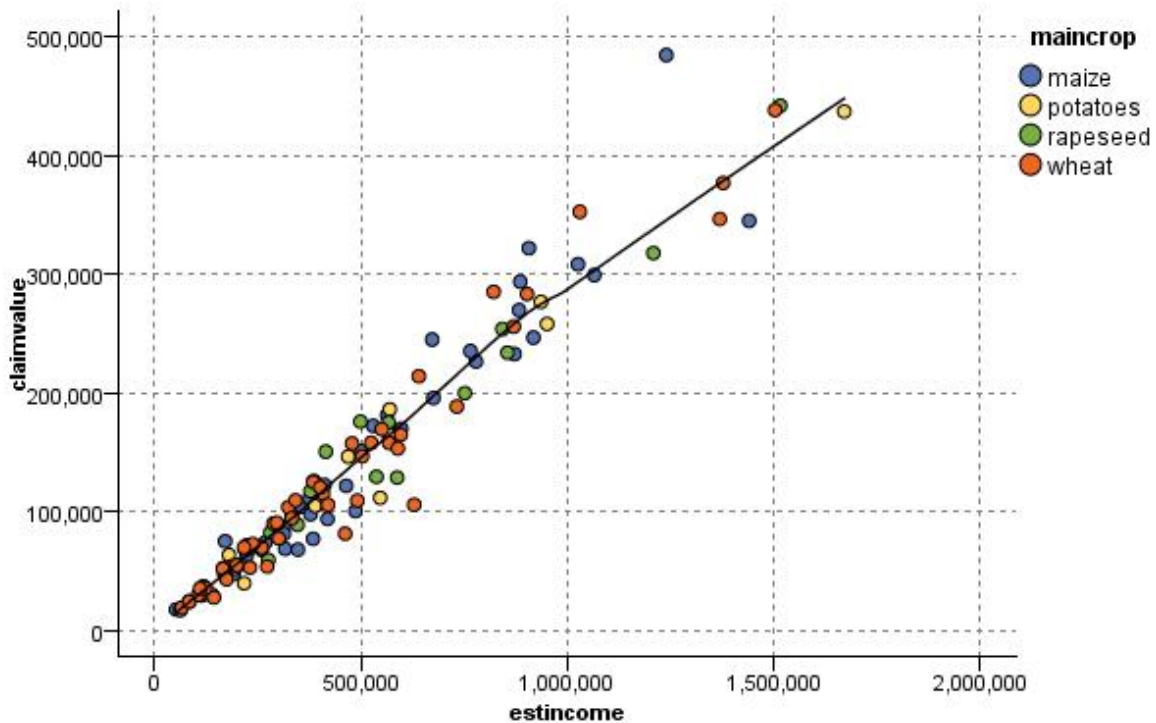


圖 22. 具有 LOESS 光滑套版的圖形

- **函數。**選取以指定已知函數來比較實際值。例如，若要比較實際值與預測值，您可以將函數  $y = x$  繪製為重疊。在文字框中指定函數 **y =**。預設函數為  $y = x$ ，但您可以依據  $x$  指定任何種類的函數，例如，二次函數或任意表示式。

附註：重疊函數不適用於畫面或動畫圖形。

當您為一個統計圖設定了選項後，可以通過按一下對話框中的執行來執行繪製。但是，您也許希望使用「選項」標籤指定更多內容，如進行「分隔」、設置「X 模式」和「樣式」。

### 統計圖「選項」標籤

**樣式。**選取點或行作為繪圖樣式。選取行將啟動 **X 模式** 控制項。選取點將使用加號 (+) 作為預設的點形狀。建立圖表後，可以變更點的形狀並改變其大小。

**X 模式。**對於行統計圖，必須選擇「X 模式」以定義行統計圖的樣式。選取**排序**、**套版**或**如所讀取**。如選擇**套版**或**如所讀取**，應該指定一個上限資料集大小，以用於對前  $n$  條記錄進行抽樣。否則，將使用預設值 2000 條記錄。

**自動 X 範圍。**選取此選項將使用此軸上整個範圍內的資料值。取消選取此項，則將使用由您指定的 **下限** 和 **上限值** 限定的值的精確子集合。您可以直接鍵入值或使用箭頭。依預設，將選取自動範圍以快速建立圖表。

**自動化 Y 範圍。**選取此選項將使用此軸上整個範圍內的資料值。取消選取此項，則將使用由您指定的 **下限** 和 **上限值** 限定的值的精確子集合。您可以直接鍵入值或使用箭頭。依預設，將選取自動範圍以快速建立圖表。

**自動化 Z 範圍。**僅用於在「統計圖」標籤上指定三維圖表的情況。選取此選項將使用此軸上整個範圍內的資料值。取消選取此項，則將使用由您指定的 **下限** 和 **上限值** 限定的值的精確子集合。您可以直接鍵入值或使用箭頭。依預設，將選取自動範圍以快速建立圖表。

**抖動。**又稱為**顫動**，在資料集中有多數重複值的情況下，「抖動」對於統計圖很有用。為了更清楚地看清值的分配，您可利用「抖動」使點隨機分佈在實際值周圍。

使用先前版本 *IBM SPSS Modeler* 的使用者請注意：在 *IBM SPSS Modeler* 的本發行版中，統計圖中使用的抖動值所採用的度量值方式與以前不同。在先前版本中，該值是一個實際數字，但現在它是相對於框大小的比例。這就意味著，舊版本串流中的顫動值在本版本中可能過大。在本版本中，任何非零顫動值都將被轉換為值 0.2。

**要繪製的上限記錄數。**指定用來繪製大量資料集的方法。可以指定上限資料集大小，或使用預設值 2,000 條記錄。當您選取 **Bin** 或**樣本**選項時，大型資料集的效能會加強。或者，您可以選擇透過選取**使用所有資料**來繪製所有資料點，但您應該注意到這樣可能會極大地降低軟體效能。

註：如果「X 模式」設定為**套版**或**如所讀取**，那麼上述選項將處於停用狀態且僅使用前  $n$  條記錄。

- **Bin。** 選取此項以在資料集包含的記錄數大於指定的記錄數時啟用 Binning。bin 使圖表在實際繪製前劃分成較小的網格，並計算每個網格單元中出現的點數。在最終圖表中，將在每個 Cell 的 bin 質心處繪製一個點（該點即代表 bin 中所有點位置的平均值）。所繪製符號的大小表示此區域內的點數（除非您將大小用作套版）。通過使用質心及大小代表點數，使 bin 後的統計圖成為代表大資料集的最佳方式，因為這樣可以避免在密集區域過量繪製（點聚集過多無法區分顏色），也減少了符號誤導的問題（點的密度型樣出現偏差）。當某些符號（特別是加號 [+]) 部分重疊時，其所產生的密集區域並不是原始資料的真實反映，這一現象稱為符號誤導。
- **樣本。** 選取此選項將根據文字框中已輸入的記錄數對資料進行隨機抽樣。預設值是 2,000。

## 統計圖「外觀」標籤

您可以在建立圖形之前指定外觀選項。

**標題。** 輸入要用作圖表標題的文字。

**次標題。** 輸入要用作圖表次標題的文字。

**標題。** 輸入要用作圖表標題的文字。

**X 標籤。** 接受自動產生的  $x$  軸（水平）標籤，或者選取自訂來指定標籤。

**Y 標籤。** 接受自動產生的  $y$  軸（垂直）標籤，或者選取自訂來指定標籤。

**Z 標籤。** 僅可用於 3D 圖形，接受自動產生的  $z$  軸標籤，或者選取自訂來指定自訂標籤。

**顯示格線。** 此選項依預設處於已選取狀態，它顯示統計圖或圖表背後的網格行，以便更輕鬆地確定區域和帶狀分割值點。網格線一律用白色顯示，但圖表背景為白色時除外；此情況下，網格線用灰色顯示。

## 使用統計圖

統計圖和多重繪圖實質都是  $X$  相對於  $Y$  的繪圖。例如，如果要調查農業補助申請中的潛在欺詐行為，那麼您可能希望繪製申請中聲明的收入相對於由類神經網路算法估計的收入的繪圖。可以使用套版，如農作物類型，來證實索賠（值或數量）與作物類型之間是否存在一定關係。

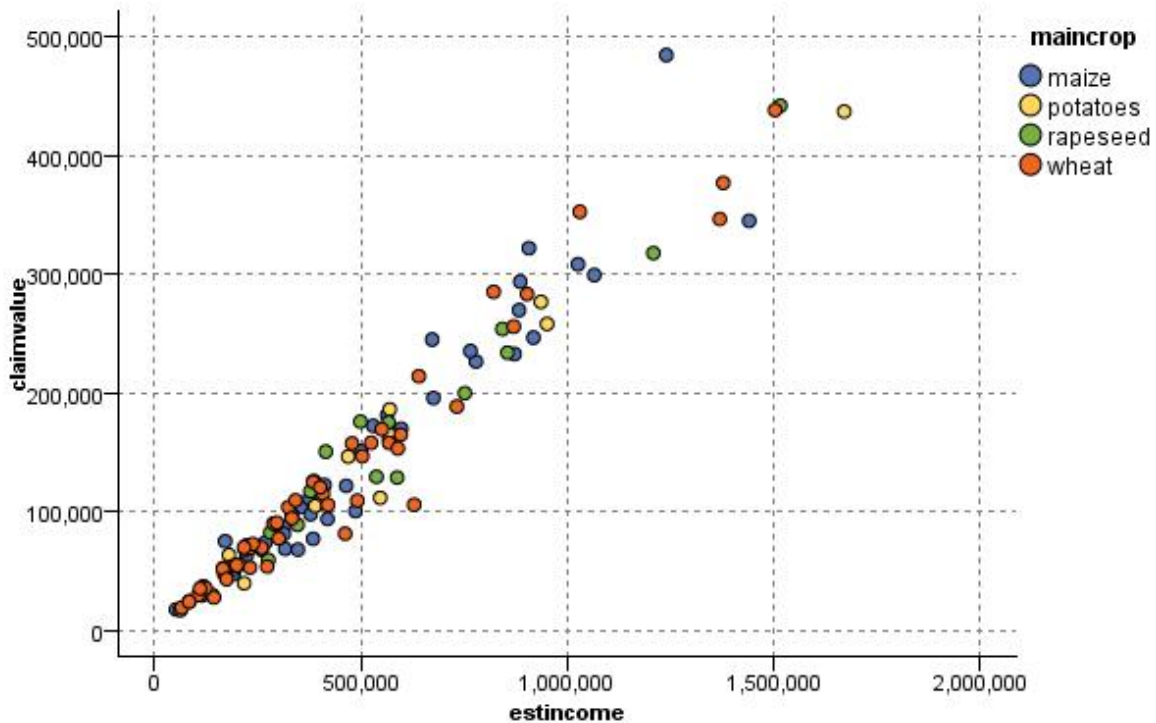


圖 23. 將主要作物類型作為套版來繪製估計收入與索賠值之間關係的繪圖

由於統計圖、多重繪圖和評估表是  $Y$  相對於  $X$  的二維度圖形顯示，通過定義區域，標示元素或繪製條形區可以很容易地與它們進行交互。也可以為這些區域、條形圖或元素所代表的資料產生節點。請參閱第 239 頁的『探索圖形』主題，以取得更多資訊。

## 多重繪圖節點

多點是一種特殊的圖類型，會在單一  $X$  欄位上顯示多個  $Y$  欄位。  $Y$  欄位將被繪製為彩色線條，且每條線都相當於一個「樣式」設定為行而「 $X$  模式」設定為排序的統計圖節點。當您使用時間序列資料並要研究幾個變數在一段時間中的變化時，多重繪圖十分有用。

## 多重繪圖標籤

**X 欄位。** 從清單中，選取要在水平  $x$  軸上顯示的欄位。

**Y 欄位。** 從從清單中選取要顯示的一個或多個超出  $X$  欄位值範圍的欄位。使用「欄位選擇器」按鈕選擇多個欄位。按一下「刪除」按鈕從清單中刪除欄位。

**重疊。** 有數種方式來說明資料值的種類。例如，您可以使用動畫套版顯示資料中所有值的多重統計圖。這對包含 10 個種類以上的集非常有用。如果是在超過 15 個種類的集合中使用，您可能會注意到效能下降了。請參閱第 162 頁的『外觀、套版、畫面和動畫』主題，以取得更多資訊。

**正規化。** 選取以將所有  $Y$  值調整至範圍 0–1，以便在圖形上顯示。正規化可協助您探索可能由於每一個數列的值範圍中的差異而遮蔽的線條之間的關係，建議在相同圖形上繪製多個線條時或者在並列畫面中比較圖形時使用。（當所有資料值都處於類似範圍時，沒有必要使用正規化。）

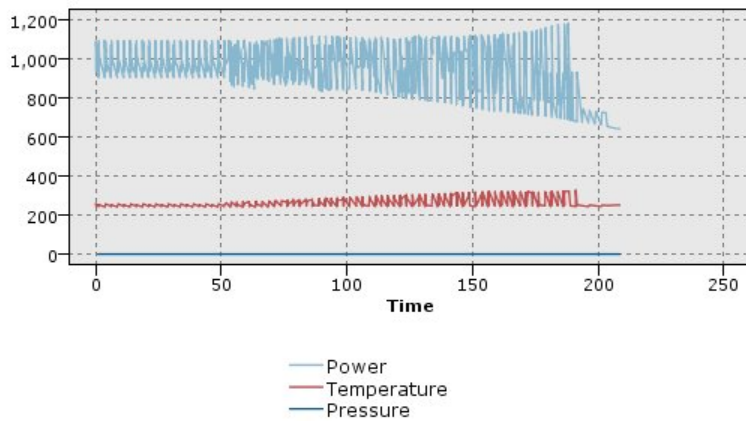


圖 24. 顯示發電廠在一段時間中的變化的標準多重繪圖（請注意，如果未選擇「標準」，則壓力的統計圖不可見）

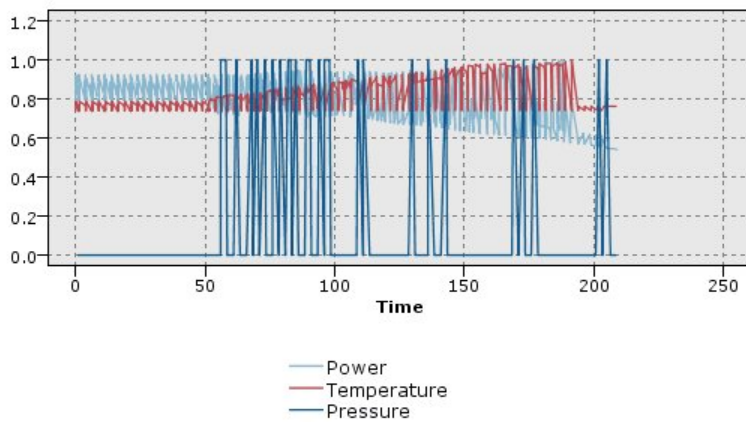


圖 25. 顯示壓力統計圖的正規化多重繪圖

**套版函數。** 選取以指定已知函數來比較實際值。例如，若要比較實際值與預測值，您可以將函數  $y = x$  繪製為重疊。在文字框中指定函數  $y =$ 。預設函數為  $y = x$ ，但您可以依據  $x$  指定任何種類的函數，例如，二次函數或任意表示式。

附註：重疊函數不適用於畫面或動畫圖形。

**當記錄數大於。** 指定用來繪製大量資料集的方法。可以指定資料集大小上限，或使用預設的 2000 點。當您選取 **Bin** 或**樣本** 選項時，大型資料集的效能會加強。或者，您可以選擇透過選取**使用所有資料**來繪製所有資料點，但您應該注意到這樣可能會極大地降低軟體效能。

注意：如果「X 模式」設定為**套版**或**如所讀取**，那麼上述選項將處於停用狀態且僅使用前  $n$  個記錄。

- **Bin。** 選取此項以在資料集包含的記錄數大於指定的記錄數時啟用 Binning。「分級」使圖表在實際繪製前被分散在較小的網格中，並計算在每個單元中將出現的連線數。在最終圖表中，每個網格中的分級矩心處將使用一個連線（該連線即代表分級中所有連線點位置的平均值數）。
- **樣本。** 選取此選項將隨機抽取指定記錄數的資料樣本。

## 多重繪圖「外觀」標籤

您可以在建立圖形之前指定外觀選項。

**標題。** 輸入要用作圖表標題的文字。

**次標題。** 輸入要用作圖表次標題的文字。

**標題。** 輸入要用作圖表標題的文字。

**X 標籤。** 接受自動產生的  $x$  軸（水平）標籤，或者選取自訂來指定標籤。

**Y 標籤。** 接受自動產生的  $y$  軸（垂直）標籤，或者選取自訂來指定標籤。

**顯示格線。** 此選項依預設處於已選取狀態，它顯示統計圖或圖表背後的網格行，以便更輕鬆地確定區域和帶狀分割值點。網格線一律用白色顯示，但圖表背景為白色時除外；此情況下，網格線用灰色顯示。

## 使用多重繪圖

統計圖和多重繪圖實質都是  $X$  相對於  $Y$  的繪圖。例如，如果要調查農業補助申請中的潛在欺詐行為，那麼您可能希望繪製申請中聲明的收入相對於由類神經網路算法估計的收入的繪圖。可以使用套版，如農作物類型，來證實索賠（值或數量）與作物類型之間是否存在一定關係。

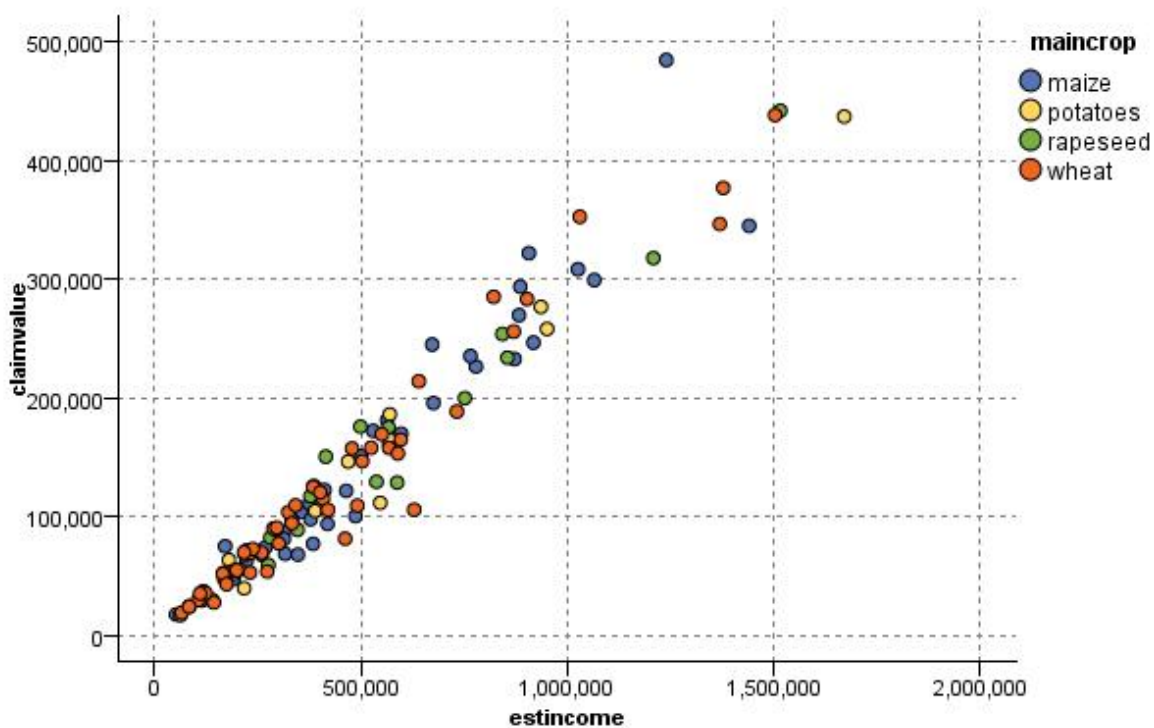


圖 26. 將主要作物類型作為套版來繪製估計收入與索賠值之間關係的繪圖

由於統計圖、多重繪圖和評估表是  $Y$  相對於  $X$  的二維度圖形顯示，通過定義區域，標示元素或繪製條形區可以很容易地與它們進行交互。也可以為這些區域、條形圖或元素所代表的資料產生節點。請參閱第 239 頁的『探索圖形』主題，以取得更多資訊。



## 時間繪圖節點

您可以使用「時間繪圖」節點檢視一個或多個繪製的在一段時間內的時間序列。這些由您繪製的數列包含數值，並且被假定將在一個時間範圍內（其中的期間一致）發生。

在 SPSS Modeler 17.1 版和更低版本中，通常在使用時間繪圖節點前，應使用時間間隔節點，建立 *TimeLabel* 欄位。依預設，在圖形中，該欄位為 *x* 軸標籤。

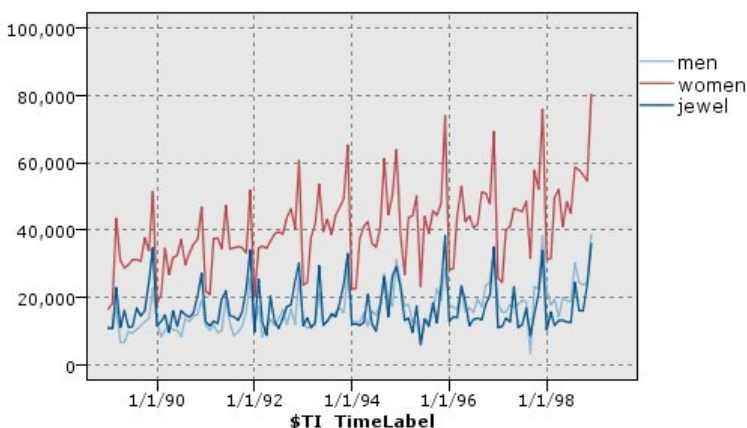


圖 27. 繪製一段時間內男人和女人的衣服及珠寶的銷售

## 建立人為介入和事件

可以在時間繪圖中，使用快速功能表產生衍生（旗標或列名）節點，並由此建立事件和人為介入欄位。例如，您可以將鐵路工人罷工這個情況建立為一個事件欄位。如果事件發生，那麼衍生狀態為 true，如未發生則為假。對於人為介入欄位，以價格增長為例，您可以使用衍生計數識別增長日期，0 代表舊價格，1 代表新價格。請參閱第 131 頁的『衍生節點』主題，以取得更多資訊。

## 時間繪圖標籤

**統計圖。** 提供一個繪製時間序列資料的選取。

- **選定的數列。** 繪製選定時間序列的值。如果選取了此選項，則在繪製信賴區間時，將取消選取正規化。
- **選定的「時間序列」模型。** 與「時間序列」模型結合使用，此選項為一個或多個選定的時間序列繪製所有的相關欄位（實際和預測值以及信賴區間）。此選項取消對話框中的其他選項。如果繪製信賴度區間，則此選項為首選項。

**數列。** 選取您要繪製的一個或多個帶有時間序列資料的欄位。資料必須是數值。

**X 軸標。** 選擇預設標籤或者單一欄位用作統計圖中 *x* 軸的標籤。如果選擇「預設值」，那麼系統將使用「時間間隔」節點中建立的 *TimeLabel* 欄位作為上游（針對 SPSS Modeler 17.1 版和更低版本中建立的串流）。如果沒有上游「時間間隔」節點，那麼將選擇有序整數列。

**在個別畫面中顯示系列。** 指定是否將每個數列顯示在個別的畫面中。另外，如果您未選擇畫面，那麼所有時間序列都將繪製在同一圖表中，而且光滑線也不可用。如果將所有時間序列繪製在同一個圖表中，每個數列都將有不同的顏色代表。

**正規化。** 選取以將所有 Y 值調整至範圍 0–1，以便在圖形上顯示。正規化可協助您探索可能由於每一個數列的值範圍中的差異而遮蔽的線條之間的關係，建議在相同圖形上繪製多個線條時或者在並列畫面中比較圖形時使用。（當所有資料值都處於類似範圍時，沒有必要使用正規化。）

**顯示。** 選取一個或多個要在統計圖中顯示的元素。您可以選擇「行」、「點」和 (LOESS)「光滑行」。只有在您選擇將多個數列顯示在不同畫面中時，「光滑線」才可用。預設情況下將選取「行」元素。請確保在您執行圖表節點前，至少選取一種繪製元素；不然，系統將傳回錯誤，告知您未選取可繪製的內容。

**限制記錄。** 如果要限制繪製的記錄數，請選取此選項。指定記錄條數，從資料檔案開頭讀取，這些將在要繪製的上限記錄數選項中繪製。依預設，此數字設定為 2,000。如果您要繪製資料檔案中倒數  $n$  條記錄，您可以事先使用排序節點將記錄按時間遞減排列。

## 時間繪圖「外觀」標籤

您可以在建立圖形之前指定外觀選項。

**標題。** 輸入要用作圖表標題的文字。

**次標題。** 輸入要用作圖表次標題的文字。

**標題。** 輸入要用作圖表標題的文字。

**X 標籤。** 接受自動產生的  $x$  軸（水平）標籤，或者選取自訂來指定標籤。

**Y 標籤。** 接受自動產生的  $y$  軸（垂直）標籤，或者選取自訂來指定標籤。

**顯示格線。** 此選項依預設處於已選取狀態，它顯示統計圖或圖表背後的網格行，以便更輕鬆地確定區域和帶狀分割值點。網格線一律用白色顯示，但圖表背景為白色時除外；此情況下，網格線用灰色顯示。

**佈置。** 僅對時間繪圖，可指定時間值散點是沿橫軸分佈還是沿縱軸分佈。

## 使用時間繪圖圖形

在建立了「時間繪圖圖形」節點後，有幾個選項可用來調整圖表顯示並生成節點以備進一步分析。請參閱第 239 頁的『探索圖形』主題，以取得更多資訊。

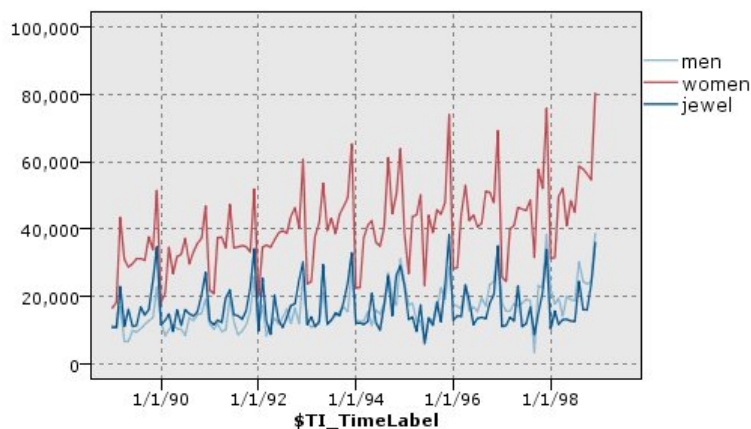


圖 28. 繪製一段時間內男人和女人的衣服及珠寶的銷售



在建立了時間繪圖、定義了級區並查看了結果之後，您可以使用「產生」功能表上的選項及快速功能表建立「選取」或「衍生」節點。請參閱第 246 頁的『從圖形產生節點』主題，以取得更多資訊。

---

## 分佈節點

分佈圖或表格顯示資料集中符號（非數值）值的出現情況，比如抵押類型或性別。「分佈」節點的一個一般用法是：在創建模型前，顯示資料中可使用「平衡」節點糾正的不平衡度。您可以使用分佈圖或表格視窗中的「產生」功能表自動產生平衡節點。

也可以使用「圖形板」節點生成計數圖形的長條。但是，在此節點中，有多個選項供您選擇。請參閱第 170 頁的『可用的內建圖板視覺化類型』主題，以取得更多資訊。

注意：要顯示數值的出現情況，應該使用「直方圖」節點。

## 分佈圖形標籤

**圖形。** 選取分佈類型。選取**所選取的欄位**將顯示所選取的欄位的分配。選取**全部旗標欄位 (true 值)** 顯示資料集中，值為 true 的旗標欄位的分佈。

**欄位。** 選取列名或旗標欄位以為其顯示值的分佈。在欄位清單中，只會出現類型未被明確設定為數值的欄位。

**套版。** 選取用作顏色套版的列名或旗標欄位，從而顯示該欄位值在指定欄位的每個值中的分配情況。例如，您可以使用市場行銷活動回應 (*pep*) 作為兒童數量的套版 (*children*)，以說明依家庭規模的回應性。請參閱第 162 頁的『外觀、套版、畫面和動畫』主題，以取得更多資訊。

**依照顏色正規化。** 選取此選項將按比例縮放長條圖，因此，所有長條圖都將填滿圖表的整個幅寬。套版值相當於每個條形的一部分，使用它可輕鬆比較不同種類。

**排序。** 選取在分佈圖中顯示值的方法。選取**英文字母**使用字母順序，或選取**依照計數**以出現次數的遞減順序列出值。

**比例尺。** 選取此選項將按比例縮放值分佈，因此，計數最大的值將填滿整個圖形的幅寬。所有其他的長條圖都將根據此值進行調整。取消選取此選項，則長條圖將根據每個值的全部計數進行縮放。

## 分佈「外觀」標籤

您可以在建立圖形之前指定外觀選項。

**標題。** 輸入要用作圖表標題的文字。

**次標題。** 輸入要用作圖表次標題的文字。

**標題。** 輸入要用作圖表標題的文字。

**X 標籤。** 接受自動產生的  $x$  軸（水平）標籤，或者選取自訂來指定標籤。

**Y 標籤。** 接受自動產生的  $y$  軸（垂直）標籤，或者選取自訂來指定標籤。

**顯示格線。** 此選項依預設處於已選取狀態，它顯示統計圖或圖表背後的網格行，以便更輕鬆地確定區域和帶狀分割值點。網格線一律用白色顯示，但圖表背景為白色時除外；此情況下，網格線用灰色顯示。

## 使用分佈節點

分佈節點用於顯示資料集中符號值的分配情況。通常在操作節點前使用，以探索資料並更正任何不平衡。例如，如果無子女的受訪者的實例遠多於其他類型的受訪者，您可能想要減少此種實例，以使後續資料採礦作業能夠產生更加有用的規則。分佈節點將協助您檢查此類不平衡並作出相關決策。

分佈節點與眾不同之處在於它能以圖表和表格兩種方式分析資料。

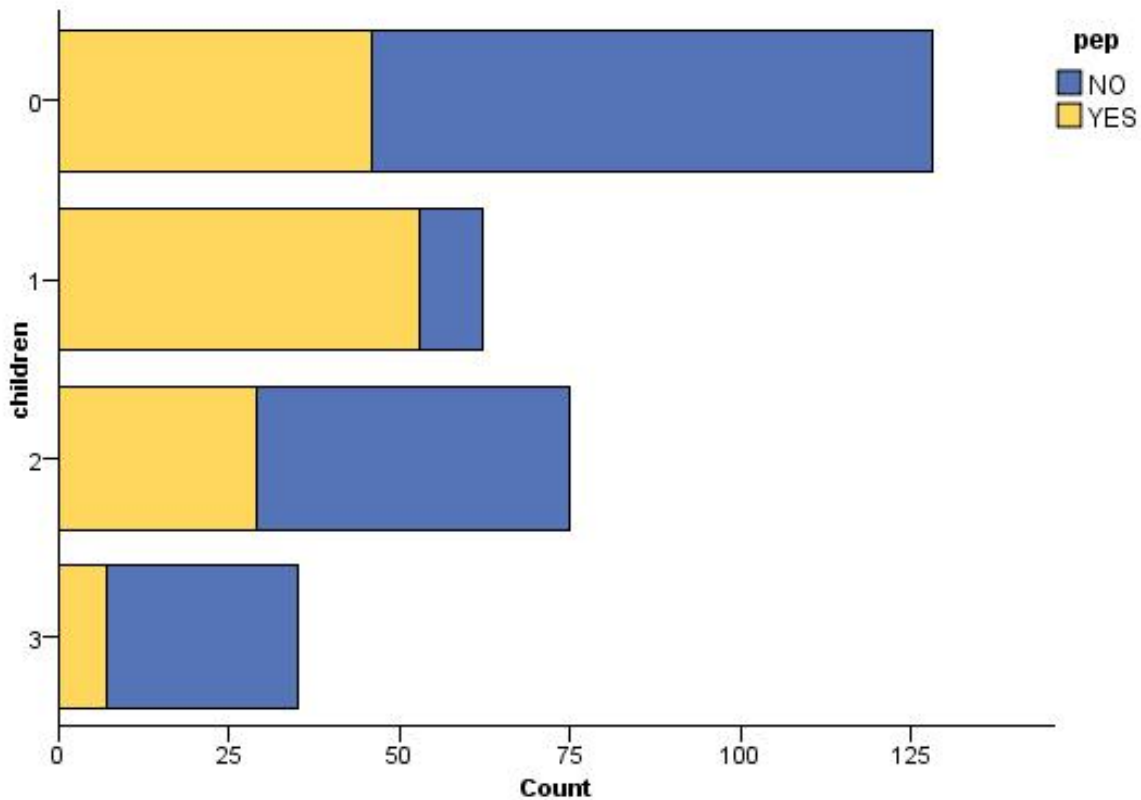


圖 29. 分佈圖顯示對市場活動有回應的人（有孩子或沒孩子）數目

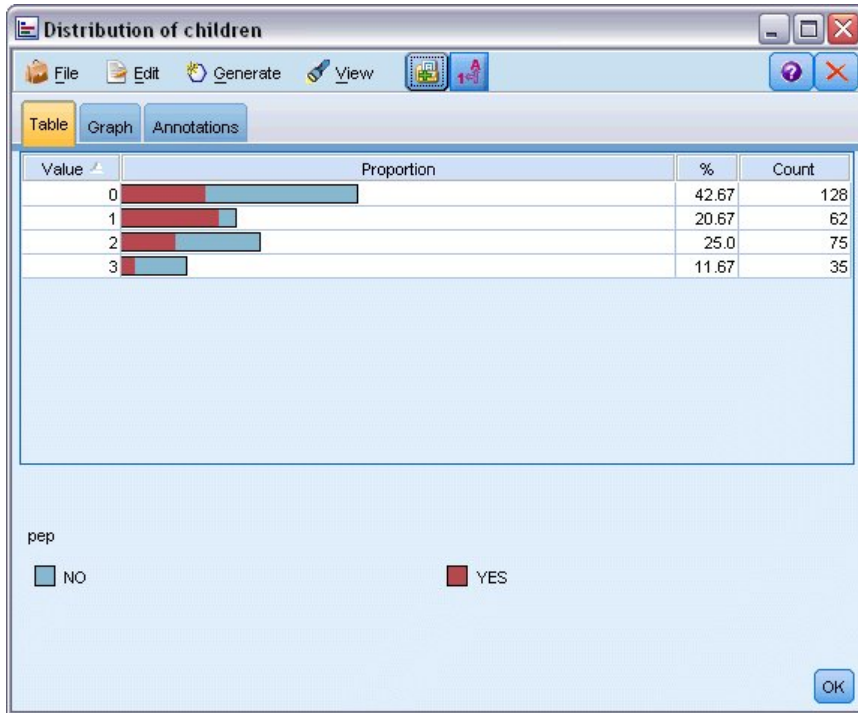


圖 30. 分佈表格顯示對市場活動有回應的人（有孩子或沒孩子）數目

在建立分佈表格和分佈圖並檢查結果後，您可以使用功能表中的選項對值進行群組、複製值並產生一些用於資料準備的節點。此外，您也可以複製或匯出圖形和表格資訊以在其他應用程式（如 MS Word 或 MS PowerPoint）中使用。請參閱第 259 頁的『列印、儲存、複製及匯出圖形』主題，以取得更多資訊。

#### 選取和複製條形圖表格中的值

1. 按一下並按住滑鼠按鈕，同時在列間拖曳滑鼠以選取值的集合。也可以使用「編輯」功能表全選值。
2. 在「編輯」功能表中，選擇複製表格或複製表格（包括欄位名稱）。
3. 貼上到剪貼簿或所需應用程式。

註：不會直接複製長條圖。而是複製表格值。也就是說，在由複製得到的表格中不會顯示重疊值。

#### 從條形圖表格中將值群組

1. 通過按住 Ctrl 鍵並按一下的方法選取要進行分組的值。
2. 在「編輯」功能表中選擇群組。

注意：在對值群組或取消值群組時，將自動重新繪製「圖表」標籤上的圖表以顯示變更。

還可以進行下列操作：

- 在條形圖清單中選取群組名稱並從「編輯」功能表中選取消群組，取消值的群組。
- 在條形圖清單中選取群組名稱並從「編輯」功能表中選取編輯群組，對群組進行編輯。此時，將開啟一個對話框，通過它可將值移入或移出該群組。

#### 「產生」功能表選項

可使用「產生」功能表上的選項選取資料子集合、衍生旗標欄位、重新對值進行群組、重新對值進行分類或平衡圖表或表格中的資料。這些作業將產生一個資料預備節點，並將其放置在串流畫布中。要使用產生的節點，請將其連接到現有串流中。請參閱第 246 頁的『從圖形產生節點』主題，以取得更多資訊。

---

## 「直方圖」節點

「直方圖」節點顯示數字欄位值的出現情況。在進行操作和模型建立之前，通常使用直方圖來探索資料。與「分佈」節點相似，「直方圖」節點常常用來揭示資料中的不平衡情況。也可以使用「圖形板」節點產生直方圖，此節點提供了更多選項可供您選擇。請參閱第 170 頁的『可用的內建圖板視覺化類型』主題，以取得更多資訊。

註：要顯示符號欄位值的出現情況，應該使用「分佈」節點。

### 直方圖「繪圖形」標籤

**欄位。** 選取要為其顯示值分佈的數值欄位。僅列出未明確地定義為符號（種類）的欄位。

**重疊。** 選取一個符號欄位以顯示指定欄位的值個種類。選取套版欄位會將直方圖轉換為堆疊圖，圖中以各種顏色代表套版欄位的不同種類。如果使用「直方圖」節點，那麼將存在三種類型的套版：顏色、畫面和動畫。請參閱第 162 頁的『外觀、套版、畫面和動畫』主題，以取得更多資訊。

### 直方圖「選項」標籤

**自動 X 範圍。** 選取此選項將使用此軸上整個範圍內的資料值。取消選取此項，則將使用由您指定的 下限 和 上限值 限定的值的精確子集合。您可以直接鍵入值或使用箭頭。依預設，將選取自動範圍以快速建立圖表。

**Bin。** 請選取按數字或依寬度。

- 選取**依數目**可顯示固定數量的長條圖，這些長條圖的寬度取決於指定 Bin 的範圍和數量。指示**Bin 數量**選項的圖表中要使用的 Bin 數量。使用箭頭調整該數字。
- 選取**按寬度**可建立一個長條圖寬度固定的圖表。Bin 數量取決於指定的寬度和值的範圍。指示**Bin 寬度**選項中長條圖的寬度。

**依照顏色正規化。** 選取此選項可將所有長條圖調整為同一高度，以便重疊的值在每個長條圖中顯示為佔所有觀察值的百分比。

**顯示常態曲線。** 選取此選項可向圖表新增常態曲線來顯示資料的平均數和變異數。

**每種顏色的獨立級區。** 選取此選項會在圖表中將每個套版的值顯示為一個獨立的級區。

### 直方圖「外觀」標籤

您可以在建立圖形之前指定外觀選項。

**標題。** 輸入要用作圖表標題的文字。

**次標題。** 輸入要用作圖表次標題的文字。

**標題。** 輸入要用作圖表標題的文字。

**X 標籤。** 接受自動產生的  $x$  軸（水平）標籤，或者選取自訂來指定標籤。

**Y 標籤。** 接受自動產生的  $y$  軸（垂直）標籤，或者選取自訂來指定標籤。

**顯示格線。** 此選項依預設處於已選取狀態，它顯示統計圖或圖表背後的網格行，以便更輕鬆地確定區域和帶狀分割值點。網格線一律用白色顯示，但圖表背景為白色時除外；此情況下，網格線用灰色顯示。

## 使用直方圖

直方圖顯示值在數值欄位中的分佈，而數值欄位中的值沿  $x$  軸排列。直方圖的運作方式類似於集合圖形。集合顯示針對一個數值欄位的值相對於另一個數值欄位值的分佈，而不是單一欄位值的出現。

建立圖形之後，您可以檢查結果並定義條帶以沿著  $x$  軸分割值，或定義區域。您也可以在圖形內標示元素。請參閱第 239 頁的『探索圖形』主題，以取得更多資訊。

可使用「產生」功能表上的選項，利用圖形中的資料（或者更具體的條帶、區域或所標示元素內的資料），來建立「平衡」、「選取」或「衍生」節點這種類型的圖形通常用於操作節點之前，用來探索資料，以及透過從圖形產生「平衡」節點以用於串流來更正任何不平衡的情況。您也可以產生「衍生旗標」節點來新增顯示每筆記錄所隸屬條帶的欄位，或是產生「選取」節點來選取特定的值集合或範圍內的所有記錄。此類作業可協助您集中在特定的資料子集以進一步探索。請參閱第 246 頁的『從圖形產生節點』主題，以取得更多資訊。

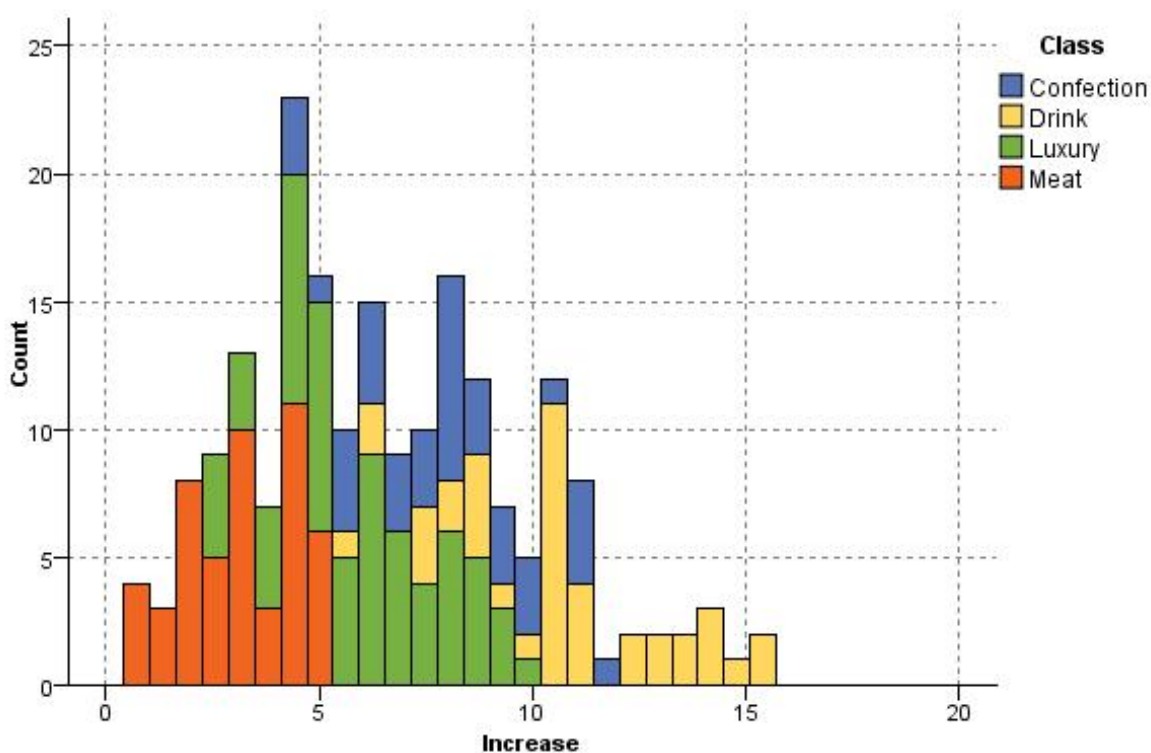


圖 31. 按種類顯示因促銷帶來的購買增長情況分佈的直方圖

## 收集節點

收集與直方圖基本相同，只有一個區別，即收集顯示某個數值欄位的值分佈（相對於另一個數值欄位的值），而不是顯示單一欄位的值出現項目。要通過圖示說明值不斷變化的變數或欄位時，可使用收集。使用 3-D 製圖，您還可以包括符號軸，用來依種類顯示分佈項目。二維度收集顯示為使用套版的堆疊長條圖。請參閱第 162 頁的『外觀、套版、畫面和動畫』主題，以取得更多資訊。

## 收集統計圖標籤

**收集。** 選取一個欄位，其值將收集並顯示在超出中指定的欄位值範圍內。只有欄位類型未被定義為符號的欄位會被列出。

**超出。** 選取一個欄位，其值將用於顯示收集中指定的欄位。

**依據。** 建立 3-D 圖表時，該選項將被啟用，您可以通過它選取用於按種類顯示收集欄位的名義或旗標欄位。

**作業。** 選取收集圖中每個條形代表的內容。選項包含合計、平均數、上限值、下限和標準差。

**重疊。** 選取一個符號欄位以顯示選定欄位的值種類。選取一個套版欄位，將收集進行轉換並為各個種類建立不同顏色的多個長條。此節點可使用三種類型的套版：顏色、畫面和動畫。請參閱第 162 頁的『外觀、套版、畫面和動畫』主題，以取得更多資訊。

## 收集選項標籤

**自動 X 範圍。** 選取此選項將使用此軸上整個範圍內的資料值。取消選取此項，則將使用由您指定的 下限 和 上限值 限定的值的精確子集合。您可以直接鍵入值或使用箭頭。依預設，將選取自動範圍以快速建立圖表。

**Bin。** 請選取按數字或依寬度。

- 選取**依數目**可顯示固定數量的長條圖，這些長條圖的寬度取決於指定 Bin 的範圍和數量。指示**Bin 數量**選項的圖表中要使用的 Bin 數量。使用箭頭調整該數字。
- 選取**按寬度**可建立一個長條圖寬度固定的圖表。Bin 數量取決於指定的寬度和值的範圍。指示**Bin 寬度**選項中長條圖的寬度。

## 收集「外觀」標籤

您可以在建立圖形之前指定外觀選項。

**標題。** 輸入要用作圖表標題的文字。

**次標題。** 輸入要用作圖表次標題的文字。

**標題。** 輸入要用作圖表標題的文字。

**Over 標籤。** 接受自動產生的標籤，或選取自訂以指定標籤。

**Collect 標籤。** 接受自動產生的標籤，或選取自訂以指定標籤。

**By 標籤。** 接受自動產生的標籤，或選取自訂以指定標籤。

**顯示格線。** 此選項依預設處於已選取狀態，它顯示統計圖或圖表背後的網格行，以便更輕鬆地確定區域和帶狀分割值點。網格線一律用白色顯示，但圖表背景為白色時除外；此情況下，網格線用灰色顯示。

下列範例顯示三維圖表上各個外觀選項的位置。

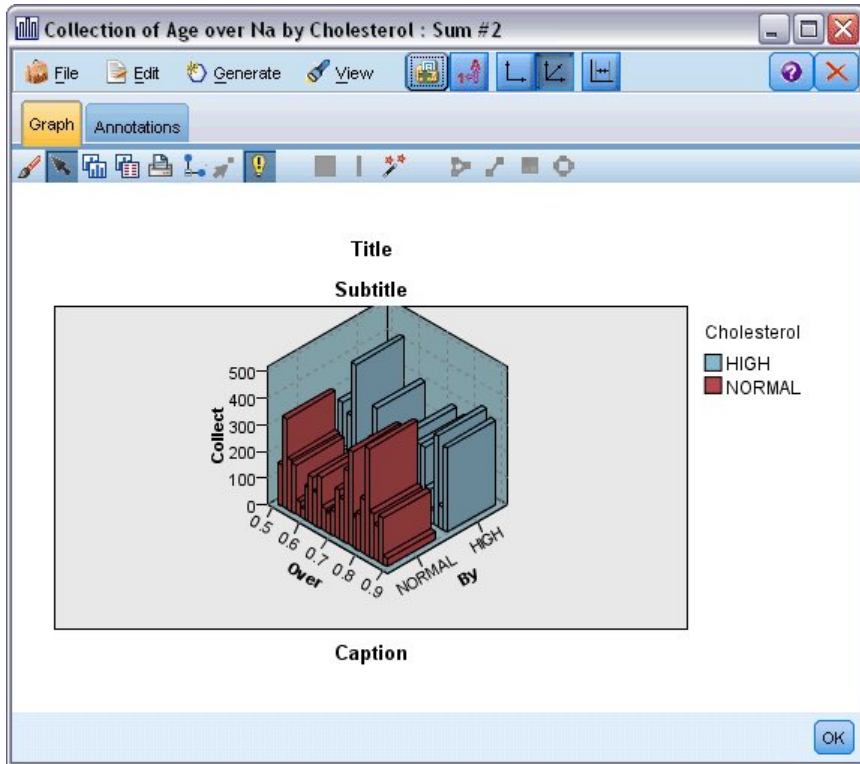


圖 32. 三維收集圖表上各個圖表外觀選項的位置

## 使用收集圖

集合顯示針對一個數值欄位的值相對於另一個數值欄位值的分佈，而不是單一欄位值的出現。直方圖的運作方式類似於集合圖形。直方圖顯示值在數值欄位中的分佈，而數值欄位中的值沿  $x$  軸排列。

建立圖形之後，您可以檢查結果並定義條帶以沿著  $x$  軸分割值，或定義區域。您也可以圖形內標示元素。請參閱第 239 頁的『探索圖形』主題，以取得更多資訊。

可使用「產生」功能表上的選項，利用圖形中的資料（或者更具體的條帶、區域或所標示元素內的資料），來建立「平衡」、「選取」或「衍生」節點這種類型的圖形通常用於操作節點之前，用來探索資料，以及透過從圖形產生「平衡」節點以用於串流來更正任何不平衡的情況。您也可以產生「衍生旗標」節點來新增顯示每筆記錄所隸屬條帶的欄位，或是產生「選取」節點來選取特定的值集合或範圍內的所有記錄。此類作業可協助您集中在特定的資料子集以進一步探索。請參閱第 246 頁的『從圖形產生節點』主題，以取得更多資訊。

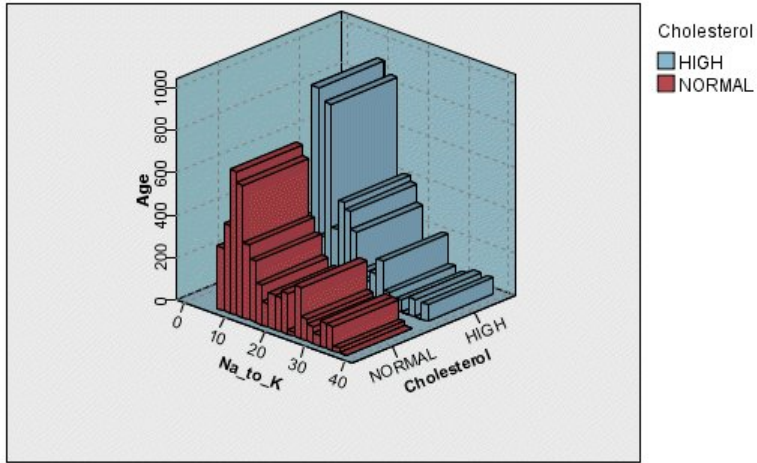


圖 33. 三維收集圖顯示隨著年齡，高和正常膽固醇層次的 *Na\_to\_K* 合計

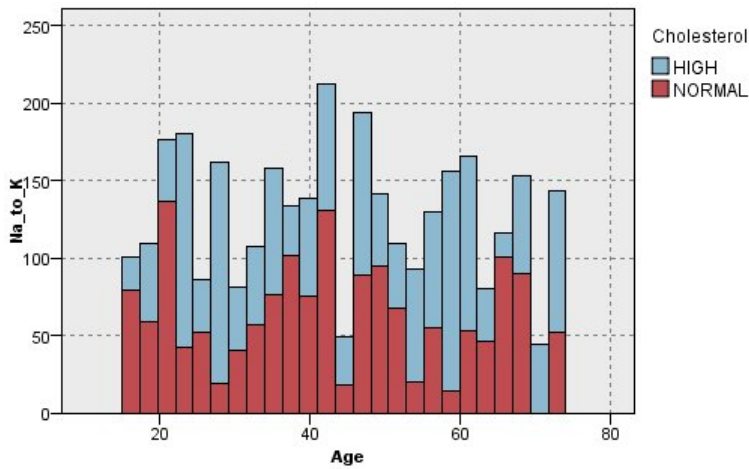


圖 34. 顯示沒有 *z* 軸的收集圖，但將膽固醇作為顏色套版

## 網絡節點

Web 節點顯示兩個或兩個以上符號欄位的值之間關係的緊密程度。其圖表使用不同類型的線條顯示鏈接，說明鏈接強度。例如，您可以使用網絡節點研究通過電子貿易網站（或在傳統零售店）所購買的不同商品之間的關係。



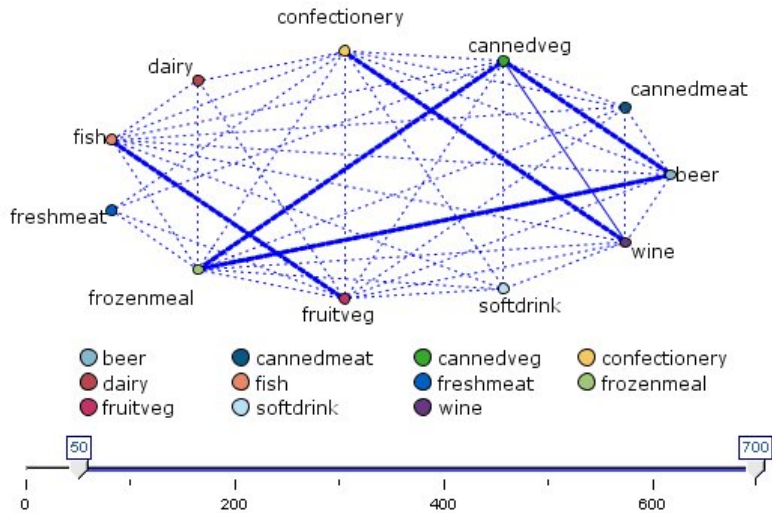


圖 35. Web 圖形顯示購買雜貨物品之間的關係

### 導向 Web

導向 Web 節點與 Web 節點相同，二者都可以顯示符號欄位間關係的緊密程度。但是，導向 Web 圖形只顯示從一個或多個從欄位到一個結束欄位之間的連線。這些鏈接都是有方向性的，即它們都是單向的。

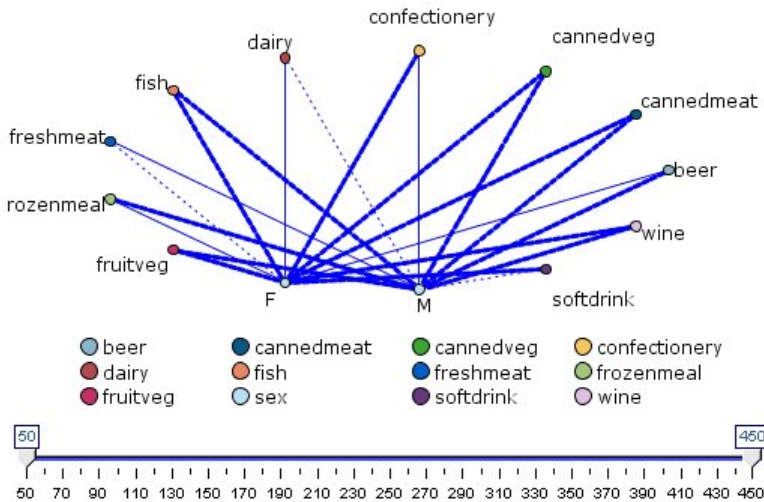


圖 36. 定向 Web 圖形顯示購買雜貨物品與性別之間的關係

與 Web 節點相同，其圖表使用不同類型的線條顯示連線來說明連線強度。例如，您可以使用導向 Web 節點研究性別與對購買某正商品的傾向性之間的關係。

### 網絡統計圖標籤

**Web**。選取此選項將建立一個 Web 圖形，顯示所有指定欄位間的關係緊密程度。

**定向 Web**。選取此選項將建立一個具有方向性的 Web 圖形，顯示多個欄位與某個欄位的值（如性別或宗教信仰）之間的關係緊密程度。如果選取此選項，則將啟動「結束欄位」並且其下方面的欄位控制項也被更名為「從欄位」，以加法以明確區分。

**結束欄位（僅用於定向 Web）。** 選取一個用於導向 Web 的旗標或列名欄位。僅會列出尚未明確設為數值的欄位。

**欄位/從欄位。** 選取欄位來建立 Web 圖形。僅會列出尚未明確設為數值的欄位。使用「欄位選擇器」按鈕選擇多個欄位或根據類型選取欄位。

注意：對於定向 Web 來說，這一控制項用來選取「從欄位」。

**僅顯示 true 值旗標。** 選取此選項將僅顯示旗標欄位中值為 true 的旗標。此選項簡化了網絡圖的顯示，常用於正面資料值的出現意義異常重要的情況。

**線的值為。** 從下拉清單中選取一個臨界值類型。

- **絕對值**選項將根據帶有成對值的記錄數設定臨界值。
- **整體百分比**選項將鏈結所代表的觀察值數的絕對值顯示為相對於 Web 圖形全部對值的出現次數的比例。
- **較小欄位/值的百分比**和**較大欄位/值的百分比**說明要使用哪個欄位/值來估計百分比。例如，欄位 *Drug* 中有 100 條記錄值為 *drugY*，但欄位 *BP* 中只有 10 條記錄值為 *LOW*。有七條記錄同時具有值 *drugY* 和 *LOW*，因此，根據您用來參考的欄位不同（較小：*BP* 或較大：*Drug*），百分比分別為 70% 或 7%。

注意：對於定向 Web 圖來說，上述第三和第四個選項不可用。作為替代，您可以選取「目標」欄位/值的百分比和「源」欄位/值的百分比。

**強鏈結較重。** 此選項依預設處於已選取狀態，即檢視欄位間鏈結的標準方式。

**弱鏈結較重。** 選取此選項將使顯示線條粗細所代表的意義與標準方式相反。經常使用此選項來偵測詐騙或檢查偏離值。

## Web 選項標籤

網絡節點的「選項」標籤包含一些用於自訂輸出圖形的其他選項。

**鏈結數目。** 下列選項用於控制在輸出圖形中顯示的鏈結數。有些選項，如 **以上方弱鏈結** 和 **以上方粗鏈結**，也可在輸出視窗中使用。同時，您也可以在最終圖表中滑動控制項調整顯示的鏈結數目量。

- **可顯示的最大鏈結數目。** 指定一個數字，說明要在輸出圖形中顯示的上限鏈結數目。使用箭頭來調整值。
- **僅顯示以上方鏈結。** 指定一個數字，說明要在 Web 中顯示的連線必須達到的最小值。使用箭頭來調整值。
- **顯示全部鏈結。** 無論上限或最小值是多少，都顯示所有鏈結。如果欄位數量過多，選取此選項將延長正在處理時間。

**若記錄過少則廢棄。** 選取此選項將忽略受少量記錄受支援的連線。在**記錄/行數**下限中輸入一個數字，設定此選項的臨界值。

**若記錄過多則廢棄。** 選取此選項將忽略受到較強受支援的連線。在**記錄/行數**上限中輸入一個數字。

**以下方弱鏈結。** 指定一個數字來指示較弱連線（點虛線）及一般連線（一般線條）的臨界值。將低於此值的所有連線都視為較弱。

**以上方強鏈結。** 指定區分強連線（粗行）與一般連線（普通線條）的臨界值。將高於此值的所有連線都視為較強。

**鏈結大小。** 指定控制鏈結大小的選項：

- **鏈結大小會持續變動。** 選取此選項將顯示鏈結大小範圍，從而反映由實際資料值產生的連線強度變化。

- **鏈結大小顯示強/正常/弱種類。** 選取此選項將顯示三種強度的連線 - 強、正常及弱。這些種類的截斷點可以在上方以及最終圖形中指定。

**Web 顯示。** 選取 Web 顯示的類型：

- **圓圈佈置。** 選取此選項將使用標準 Web 顯示。
- **網路佈置。** 選取此選項將使用一種演算法，將最強的鏈結分在一起。這樣做的目的是以空間差別及加粗線條突出強鏈結。
- **方向佈置。** 選擇此選項將建立一個導向 Web 顯示。此圖形使用「統計圖」標籤至欄位中的選擇作為方向的集中點。
- **格線佈置。** 選取此選項將建立一個以大小相同的網格形式顯示的網路圖。

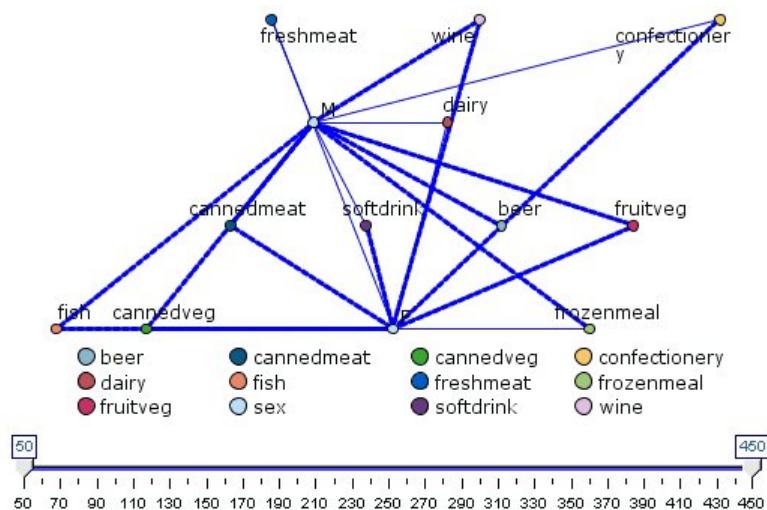


圖 37. 顯示從 *frozenmeal* 及 *cannedveg* 到其他雜貨的強鏈接的網路圖

註：當過濾所顯示的鏈結（使用 Web 圖表中的調節器或 Web 節點「選項」標籤上方的僅顯示上述鏈結控制項）時，最後可能會出現這樣一種狀況，即保持顯示的所有鏈結都是一種值（換言之，全都是弱鏈結、中級鏈結或強鏈結，這取決於 Web 節點「選項」標籤上方的以下方弱鏈結和以上方強鏈結控制項的定義）。如果出現這種情況，所有這些鏈結在 Web 圖形輸出中都顯示為中等寬度的線條。

## Web 「外觀」標籤

您可以在建立圖形之前指定外觀選項。

**標題。** 輸入要用作圖表標題的文字。

**次標題。** 輸入要用作圖表次標題的文字。

**標題。** 輸入要用作圖表標題的文字。

**顯示圖註。** 可以指定是否顯示圖註。對於具有大量欄位的統計圖，隱藏圖註將改善統計圖的外觀。

**把標籤當作節點使用。** 可以將標籤文字包含在每個節點中，而不是顯示鄰近的標籤。對於欄位數量較少的統計圖，此選項可以提高圖表可讀取性。

### Relationship between gender and grocery purchases

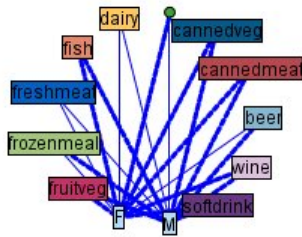


圖 38. 將標籤顯示為節點的 Web 圖形

### 使用網路圖

Web 節點用於顯示兩個或更多符號欄位的值之間關係的緊密程度。在圖表中顯示的鏈接以不同類型的線條表示，依次說明鏈接的強度不同。例如，您可以使用 Web 節點來研究膽固醇層次、血壓及可有效治療病人疾患的藥品之間的關係。

- 強鏈接以粗線條顯示。用以說明兩個值之間關係緊密，應該進一步研究。
- 普通鏈接用普通粗細的線條顯示。
- 弱鏈接以虛行顯示。
- 如果在兩個值之間沒有顯示線條，則說明兩個值從不同時出現在同一記錄中或這種同時出現的情況只發生在有限的記錄中，即記錄數量低於在「網絡節點」對話框中指定的臨界值。

在建立了 Web 節點後，有幾個選項可用來調整圖表顯示並產生節點以備進一步分析。

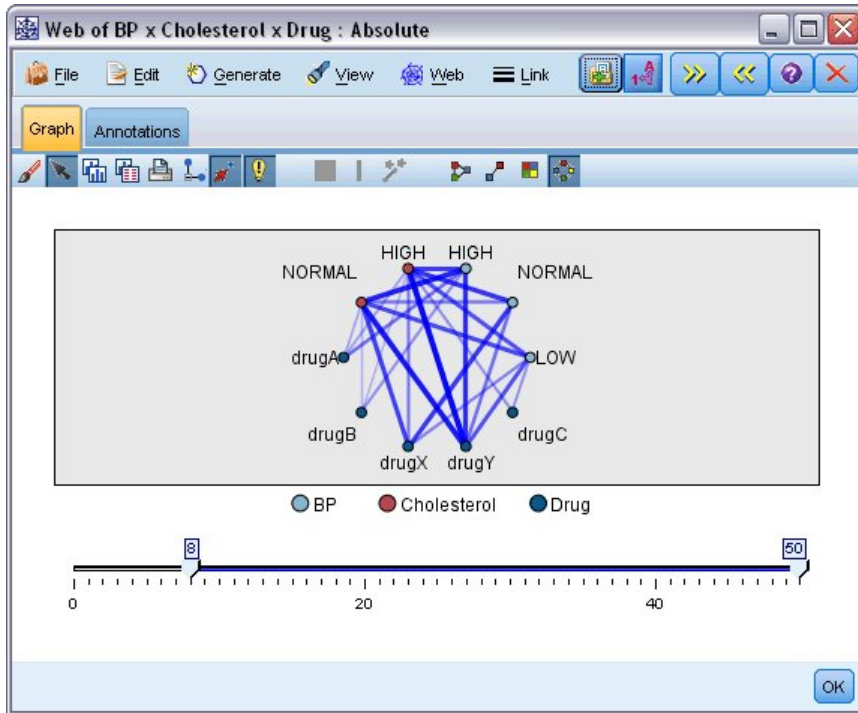


圖 39. 網路圖說明一組緊密程度高的關係，如正常血壓與 *DrugX* 及高膽固醇與 *DrugY*。

對於網絡節點和導向 Web 節點，您都可以：

- 變更 Web 顯示的佈置。
- 隱藏一些點以簡化顯示。
- 變更控制線條樣式的臨界值。
- 強調顯示值之間的線條，說明「選定」關係。
- 產生一個或多個「選定」記錄的「選取」節點或與 Web 中的一種或多種關係關聯的「衍生旗標」節點。

### 調整點

- 在某點上按一下滑鼠並將其拖曳到新的位置來移動點。將重新繪製 Web 以反映這個新的位置。
- 在 Web 中某點上用滑鼠右鍵按一下並從快速功能表中選擇隱藏或隱藏並重新規劃來隱藏點。隱藏僅隱藏所選點及與其相關聯的所有線條。隱藏並重新規劃將根據您所作的變更來重新繪製 Web。所有手動移動都是未完成的。
- 通過選擇圖表視窗中「網絡」功能表上的全部顯現或全部顯現並重新計劃來顯示所有隱藏的點。選取全部顯現並重新計劃將重新繪製網絡，即，將之前隱藏的點和它們的鏈接都包含進來。

### 選取或「強調顯示」行

所選行以紅色強調顯示。

1. 要選取一條行，請左鍵按一下該行。
2. 要選取多條行，進行下列操作之一：
  - 使用游標在您希望選取行的點周圍繪製圓形。
  - 按下 Ctrl 鍵並左鍵按一下您希望選取的單個行。

您可以按一下圖表背景，或從圖表視窗中的 Web 功能表選擇清除選擇取消選擇所有選定行。

## 使用其他佈置檢視 Web

在「網路」功能表中，選擇**圓圈佈置**、**網狀佈置**、**方向佈置**或**格線佈置**變更圖表佈置。

## 開啟或關閉鏈結調節器

在「檢視」功能表中選擇**鏈結調節器**。

## 選取或旗標單個關係的記錄

1. 對於感興趣關係，可在代表該關係的行上用滑鼠右鍵按一下。
2. 在快速功能表中，選擇**為鏈結產生選取節點**或**為鏈結產生推導節點**。

帶有相應選項和指定條件的選取節點或衍生節點將自動新增到串流畫布中：

- 「選取」節點選取指定關係中的所有記錄。
- 「衍生」節點產生一個旗標，為整個資料集中的記錄一一標明所選關係是否存在，即旗標值為 true。旗標欄位的名稱以兩個值的名稱中間加上底線表示，例如，LOW\_drugC 或 drugC\_LOW。

## 選取或旗標一群組關係的記錄

1. 選取 Web 顯示中代表相關關係的線條。
2. 在圖表視窗的「產生」功能表中選擇**選取節點 ("與")**，**衍生節點 ("或")**，或**衍生節點 ("與")**和**衍生節點 ("或")**。
  - 「或」節點將條件進行析取。也就是說，只要在記錄中存在所選關係中的任何一個，即可產生此節點。
  - 「與」節點將條件進行鏈結。也就是說，只有當記錄滿足所有所選關係時，才可產生此節點。如果所選關係中存在任何互斥的關係，則產生錯誤。

在選擇完成之後，帶有相應選項和指定條件的選取節點或衍生節點將自動新增至串流畫布中。

註：當過濾所顯示的鏈結（使用 Web 圖表中的調節器或 Web 節點「選項」標籤上方的**僅顯示上述鏈結控制項**）時，最後可能會出現這樣一種狀況，即保持顯示的所有鏈結都是一種值（換言之，全都是弱鏈結、中級鏈結或強鏈結，這取決於 Web 節點「選項」標籤上方的**以下方弱鏈結和以上方強鏈結控制項**的定義）。如果出現這種情況，所有這些鏈結在 Web 圖形輸出中都顯示為中等寬度的線條。

## 調整網絡臨界值

在您建立了 Web 圖形後，您可以通過工具列調節器調整控制線條樣式的臨界值，以改變下限可見線條。您還可以按一下工具列上的黃色的雙箭頭擴展 Web 圖形視窗，以檢視其他臨界值選項。然後按一下**控制項**標籤檢視其他選項。



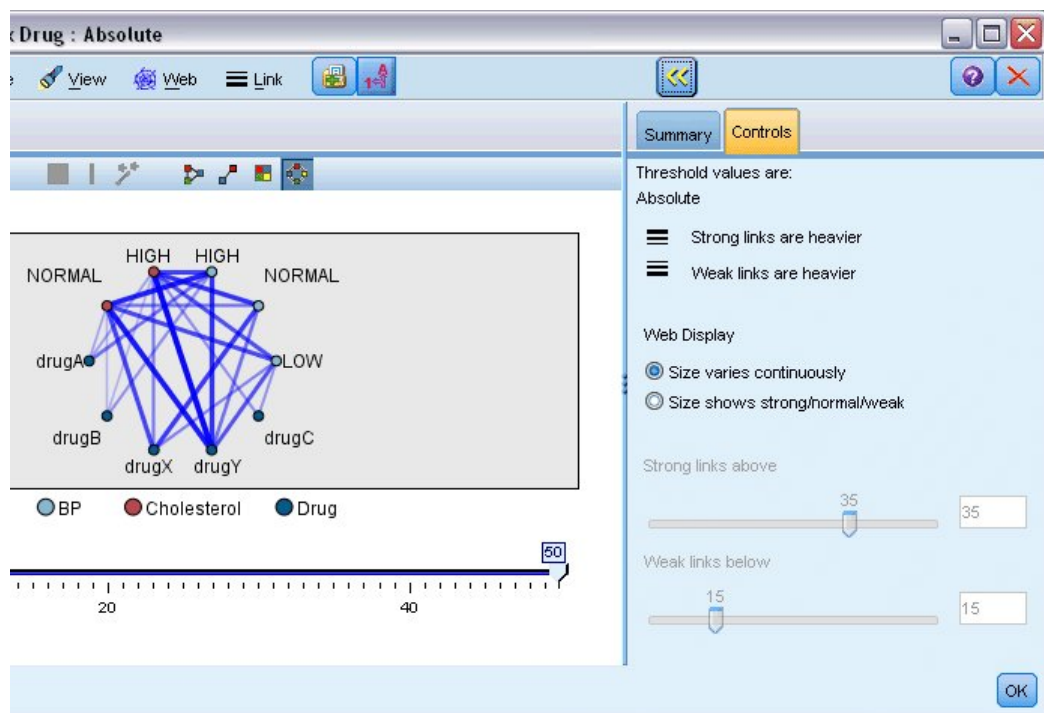


圖 40. 擴展視窗中的選項主要針對顯示和臨界值

臨界值為。顯示在 Web 節點對話框中建立節點時所選的臨界值類型。

粗鏈結較重。此選項依預設處於已選取狀態，即檢視欄位間鏈結的標準方式。

弱鏈結較重。選取此選項將使顯示線條粗細所代表的意義與標準方式相反。經常使用此選項來偵測詐騙或檢查偏離值。

**Web 顯示。**在輸出圖形中指定控制鏈結大小的選項：

- 大小會持續變動。選取此選項將顯示鏈結大小範圍，從而反映由實際資料值產生的連線強度變化。
- 大小顯示強/正常/弱。選取此選項將顯示三種強度的連線 - 強、正常及弱。這些種類的截斷點可以在上方以及最終圖形中指定。

**以上方強鏈結。**指定區分強連線（粗行）與一般連線（普通線條）的臨界值。將高於此值的所有連線都視為較強。使用調節器來調整值或在欄位中輸入數字。

**以下方弱鏈結。**指定一個數字來指示較弱連線（點虛線）及一般連線（一般線條）的臨界值。將低於此值的所有連線都視為較弱。使用調節器來調整值或在欄位中輸入數字。

在您調整了 Web 的臨界值之後，您可以通過 Web 圖形工具列上的 Web 功能表重新計劃或重新繪製基於新臨界值的 Web 顯示。在您確定了能使圖表的意義最為明顯的設定之後，您可以按一下圖表視窗「網絡」功能表中的**更新母節點**來更新網絡節點（也叫作母節點）中的原有設定。

## 建立 Web 彙總

您可以按一下工具列上的黃色雙箭頭按鈕擴展 Web 圖形視窗，以建立 Web 彙總文件，其中將列出粗、中等及弱鏈結。然後按一下彙總標籤，檢視每類鏈結的表格。可以通過每個表格的切換按鈕展開和摺疊表格。

要列印彙總，從網路圖視窗的功能表中選擇：

## 評估節點

「評估」節點為您提供了一個評估並比較預測模型，以選擇最適合模型的便捷方法。評估圖表顯示模型如何執行對特定結果的預測。評估圖表的工作原理是：根據預測值及預測的信賴度排序記錄、將記錄分割為大小相等的群組（分位數）並按由高到低順序為每個分位數繪製業務準則值。多個模型在圖形中顯示為個別的行。

通過將具體值或值的範圍定義為匹配，處理結果。通常，匹配表示相關的某類別（如向顧客銷售）或某事件（如某項醫療診斷）成功執行。您可以在對話框的「選項」標籤上定義匹配準則，或使用以下描述的預設匹配準則：

- 旗標輸出欄位是正向的，即匹配表現為 *true* 值。
- 對於列名輸出欄位，集合中的第一個值確定是否匹配。
- 對於連續輸出欄位，大於欄位範圍中間點的值即為匹配。

有六種類型的評估表，每一種類型針對不同的評估準則。

### 增益圖表

增益的定義是相對於全部匹配，發生於每個分位數中的匹配的百分比。其計算方法為（分位數中的匹配數量 / 全部匹配數量）× 100%。

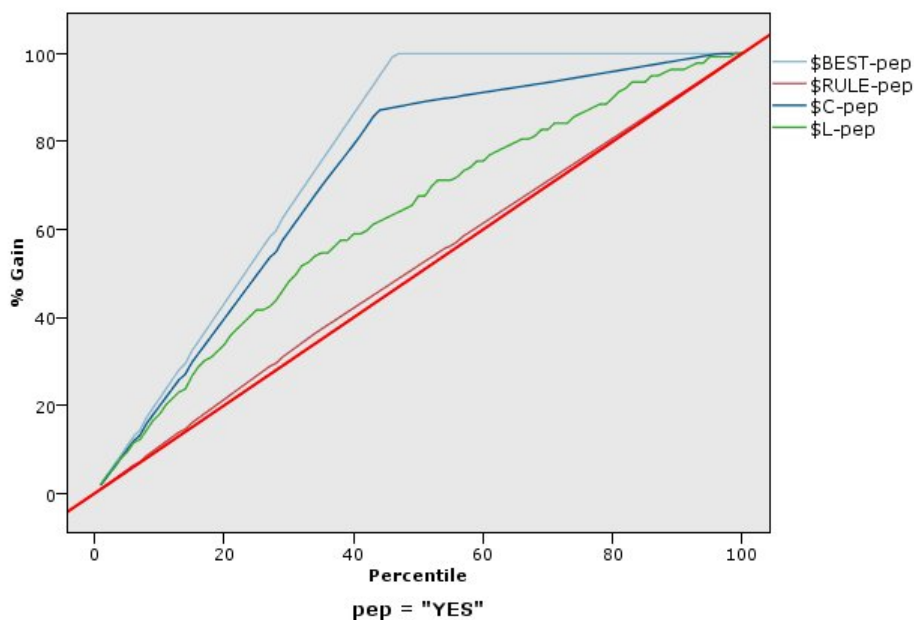


圖 41. 顯示帶有基準線、最佳行及商業規則的增益圖（累加）

### 提升圖

提升將每個分位數中匹配記錄的百分比與在全部訓練資料中匹配的百分比進行比較。其計算方式為（在分位數中的匹配 / 在分位數中的記錄） / （全部匹配 / 全部記錄）。



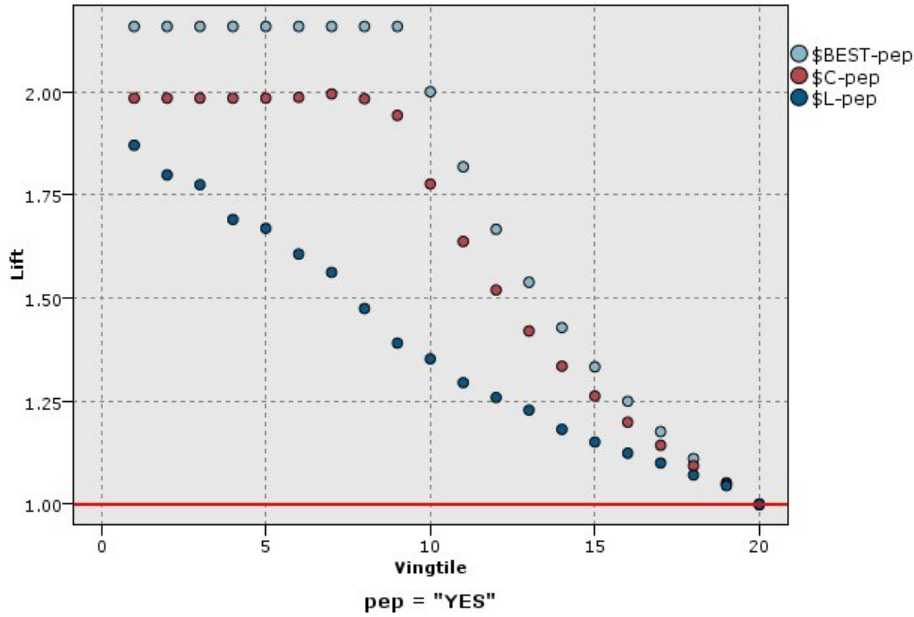


圖 42. 使用點和最佳行的提升圖（累加）

#### 回應圖表

回應即分位數中，匹配記錄的比例。回應計算方式為（命中分位數 / 記錄分位數）× 100%。

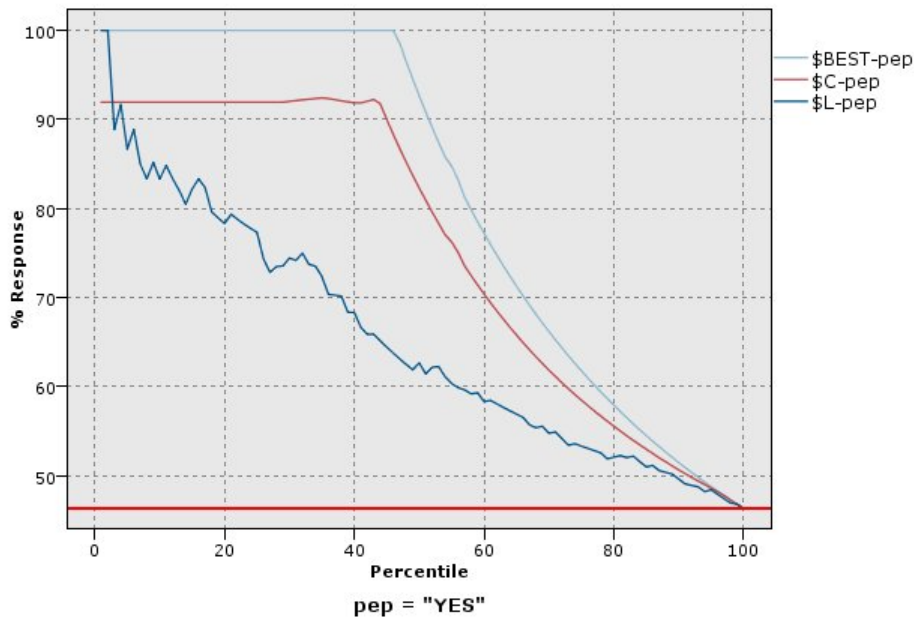


圖 43. 具有最佳行的回應圖表（累加）

#### 利潤圖

利潤等於每個記錄的 營收 減去該記錄的 成本。分位數的利潤僅僅是分位數中所有記錄的利潤總和。這裡假定營收僅套用至命中數，但成本可套用至所有的記錄。利潤及成本都可以是固定的，也可以由資料中的欄位決

定。其計算方法為（分位數中所有記錄營收的總和 - 分位數中所有記錄成本的總和）。

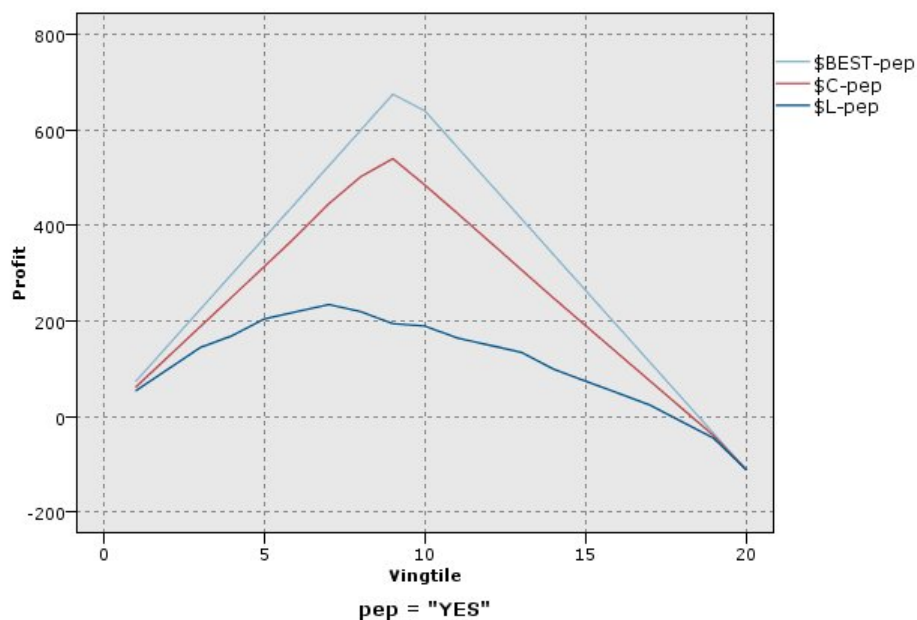


圖 44. 具有最佳行的利潤圖（累加）

#### 投資報酬率圖表

投資返回 (ROI) 也需要確定營收和成本，從這一點上來說，它與利潤相同。ROI 將分位數的成本和利潤進行比較。ROI 計算方式為（利潤分位數 / 成本分位數）× 100%。

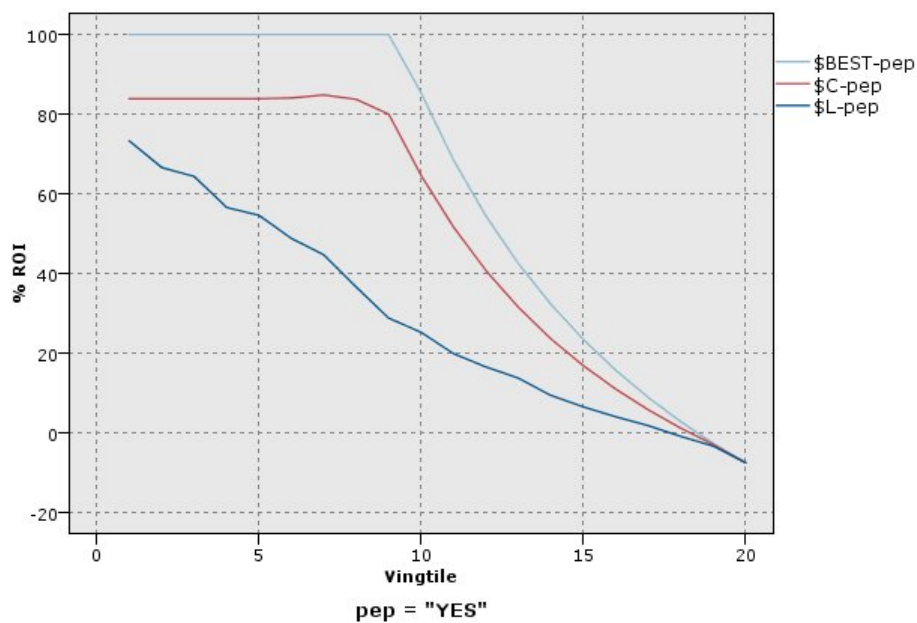


圖 45. 具有最佳行的投資報酬率圖表（累加）

#### ROC 圖表

只能將 ROC (受試者工作特性) 與二元分類器配合使用。ROC 可用於根據效能顯示、組織和選取分類器。ROC 圖表繪製分類器的 true 陽性率 (或靈敏度) 與假陽性率的比率。ROC 圖表描述了收益 (true 陽性) 與成本 (假陽性) 之間的相對平衡。true 陽性是一個匹配實例, 並且分類為命中數。因此, true 陽性率按照 true 陽性數 / 實際匹配的實例數進行計算。假陽性是一個未匹配實例, 並且分類為命中數。因此, 假陽性率按照假陽性數 / 實際未匹配的實例數進行計算。

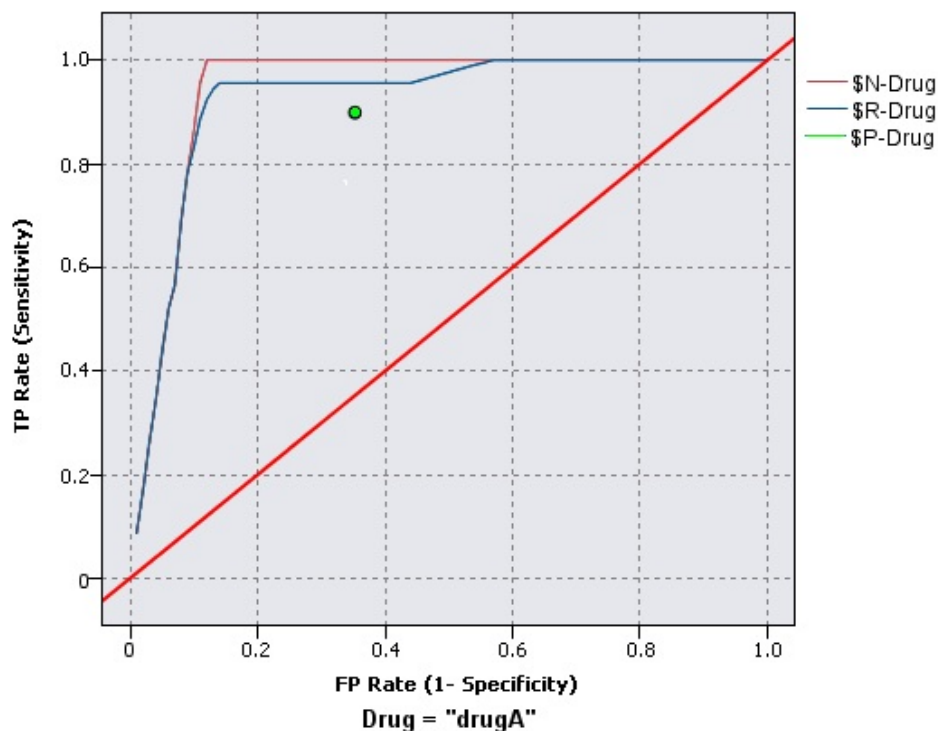


圖 46. 具有最佳行的 ROC 圖表

評估表也可以累加, 因此每個點等於相應分位數的值加號上所有更高分位數的值。累加圖表通常能夠更好的表現模型效能, 而非累加圖則更有利於指出模型中可能存在問題的地方。

### 評估統計圖標籤

**圖表類別。** 選取下列其中一種類型: 提昇圖、回應圖、提升圖、利潤圖、ROI (投資增益率) 或 ROC (受試者工作特性)。

**累加圖形。** 選取此選項將建立累加圖表。累加圖表中繪製的值代表每個分位數與所有更高分位數的和。(累加圖形不適用於 ROC 圖表。)

**含基準線。** 選取此選項將在統計圖中包含基準線, 表示匹配值的完全隨機分配 (此時信賴度並不相關)。(含基準線不適用於利潤圖及投資報酬率圖表。)

**含最佳線。** 選取此選項將在統計圖中包含最佳行, 代表最佳信賴度 (即命中數 = 觀察值的 100%)。(含最佳線不適用於 ROC 圖表。)

**將利潤準則用於所有圖表類別。** 選取此選項可在計算評估測量時使用利潤準則 (成本、營收和加權) 而不是正常匹配計數。對於具有特定數值目標的模型 (例如用於預測從客戶那裡取得的供應營收的模型), 目標欄位的值提供了一種比命中數數更好的模型效能測量方式。選取此選項將對提昇圖、回應圖和提升圖啟用成本、營收

和加權欄位。要對這三種圖表類別使用利潤條件，建議將營收設定為目標欄位，將成本設定為 0.0，從而使利潤等於營收，並且建議指定一個使所有記錄都計為命中數的使用者定義 "true" 匹配條件。（將利潤準則用於所有圖表類別不適用於 ROC 圖表。）

**尋找預測/預測值欄位使用。** 選取模型輸出欄位 **meta** 資料以使用其 meta 資料搜尋圖表中的預測欄位，或者選取欄位名稱格式依名稱進行搜尋。

**圖形分數欄位。** 選擇此勾選框可啟用分數欄位選擇器。然後選取一個或多個範圍或連續型分數欄位；即，不是嚴格預測模型但可用於根據匹配傾向程度對記錄排等級的欄位。「評估」節點可以將一個或多個分數欄位的任意組合與一個或多個預測模型進行對比。一個一般的範例是將幾個 RFM 欄位和最佳預測模型進行對比。

**目標。** 使用欄位選擇器選擇目標欄位。選擇任意實例化旗標或具有兩個或多個值的列名欄位。

注意：此目標欄位只適用於分數欄位（預測模型會定義自己的目標欄位），並且如果在「選項」標籤上設定了自訂匹配準則，那麼將省略目標欄位。

**按分割區分割。** 如果要用分割區欄位將記錄分割為訓練、測試及驗證樣本，請選取此選項，以便為每個分割區顯示一個個別的評估表。請參閱第 150 頁的『分割區節點』主題，以取得更多資訊。

注意：按分割區分割時，將不對分割區欄位中具有空值的記錄進行評估。如果使用「分割區」節點，這永遠都不是問題，因為「分割區」節點不會產生空值。

**圖形。** 從下拉清單中選取要在圖表中繪製的分位數的大小。選項包含四分位數、五分位數、十分位數、二十分位數、百分位數和千分位數。（統計圖不適用於 ROC 圖表。）

**樣式。** 選取行或點。

對於除 ROC 圖表以外的所有圖表類別，提供了使您可以指定成本、營收和加權的更多控制項。

- **成本。** 指定與每個記錄關聯的成本。您可以選取**固定或變數成本**。對於固定成本，請指定成本值。對於變動成本，請按一下「欄位選擇器」按鈕，將某個欄位選擇為成本欄位。（成本不適用於 ROC 圖表。）
- **營收。** 指定與代表命中數的每個記錄關聯的營收。您可以選取**固定或變數成本**。對於固定收益，請指定營收值。對於可變營收，請按一下「欄位選擇器」按鈕，將某個欄位選擇為營收欄位。（營收不適用於 ROC 圖表。）
- **加權。** 如果資料中的記錄代表多個單元，那麼可以使用頻率加權來調整結果。使用**固定或可變加權**，指定與每個記錄關聯的加權。對於固定加權，請指定加權值（每個記錄的單元數）。對於變動加權，請按一下「欄位選擇器」按鈕，將某個欄位選擇為加權欄位。（加權不適用於 ROC 圖表。）

## 「評估選項」標籤

評估表的「選項」標籤提供了定義在圖表中顯示的匹配、評分準則及商業規則的靈活性。您可以設定這些選項，以匯出模型評估結果。

**使用者定義的命中數。** 選取此選項可指定一個用於指示命中數的自訂條件。此選項更適合於定義相關結果，而不是從目標欄位類型和值的順序中推測結果。

- **條件。** 如果選取了上方的**使用者定義的命中數**，那麼您必須指定一個 CLEM 表示式作為匹配條件。例如，@TARGET = "YES" 即是一個有效條件：凡是目標欄位的值為 Yes 的記錄，都將在評估中算作匹配。指定的條件將用於所有目標欄位。若要建立條件，請在欄位中輸入或使用「表示式建置器」來產生條件表示式。如果資料已實例化，則您可以直接從「表示式建置器」插入值。

**使用者定義的分數。** 選取此選項後，用戶可以指定一個在將觀察值分配到分位數之前，對觀察值評分的條件。預設分數將根據預測值及信賴度計算。使用「表示式」欄位建立一個自訂評分表示式。

- **表示式**。指定用於評分的 CLEM 表示式。例如，如果取值範圍在 0 -1 之間的數值輸出是按如下順序排列的：較小的值好於較大的值，那麼您可以定義一個命中數  $@TARGET < 0.5$ ，且相應的分數為  $1 - @PREDICTED$ 。分數表示式必須返回一個數值。若要建立條件，請在欄位中輸入或使用「表示式建置器」來產生條件表示式。

**包含商業規則**。選取此選項後，您可指定滿足相關準則的規則條件。例如，您可能想要顯示用於所有滿足 `mortgage = "Y" and income >= 33000` 這一條件的觀察值的規則。商業規則將在圖表中繪製出來，在鍵字段中，將被標記為 *Rule*。（對於 ROC 圖表，包含商業規則不受支援。）

- **條件**。指定用於定義輸出圖表中商業規則的 CLEM 表示式。請直接在欄位中鍵入或使用表示式建置器產生條件表示式。如果資料已實例化，則您可以直接從「表示式建置器」插入值。

**匯出結果至檔案**。選取此選項後，用戶可將模型評估結果匯出到分隔的文字檔中。您可以讀取此檔案，對計算結果執行特定的分析。為匯出設定如下選項：

- **檔名**。輸入輸出檔案的檔名。使用省略號按鈕 (...)，瀏覽所需的資料夾。
- **定界字元**。輸入一個字元（如逗點或空格）作為欄位定界字元。

**包括欄位名稱**。選取此選項可使欄位名稱在輸出檔案的第一行顯示出來。

**每筆記錄後換新行**。選取此選項後，每條記錄將另起一個新行。

## 評估「外觀」標籤

您可以在建立圖形之前指定外觀選項。

**標題**。輸入要用作圖表標題的文字。

**次標題**。輸入要用作圖表次標題的文字。

**文字**。接受自動產生的文字標籤，或選取自訂指定標籤。

**X 標籤**。接受自動產生的 *x* 軸（水平）標籤，或者選取自訂來指定標籤。

**Y 標籤**。接受自動產生的 *y* 軸（垂直）標籤，或者選取自訂來指定標籤。

**顯示格線**。此選項依預設處於已選取狀態，它顯示統計圖或圖表背後的網格行，以便更輕鬆地確定區域和帶狀分割值點。網格線一律用白色顯示，但圖表背景為白色時除外；此情況下，網格線用灰色顯示。

## 讀取模型評估結果

評估表的解讀方法在某種程度上取決於圖表類型，但是，有些特點是所有評估表共有的。對於累加圖表而言，行位置越高（特別是當圖表左端行位置高時）表明模型越優秀。在很多情況下，在比較多個模型時，行會發生交叉。因此，一個模型的行可能會在某處較高；但在圖表另一處，另一個模型的行較高。如果出現這種情況，您需要考慮要哪個部分的樣本（由此確定 *x* 軸上點的位置），以確定選擇哪個模型。

大多數非累加圖表都極其相似。優秀模型的非累加圖應該是左端較高，右端較低。（如果非累加圖形呈鋸齒狀，您可以減少分位數的數量，重新繪製並執行圖表，由此獲得較為平滑的圖表。）線在圖表左端偏低而在右端偏高，可能意味著模型預測結果較低劣的區域。一條在整個圖表中平直的線條則說明此模型基本不能提供任何資訊。

**增益圖**。累加增益圖的線從左到右的走勢一律是從 0% 到 100%。對於良好的模型，增益圖表向 100% 突增，然後趨於平穩。無法提供有用資訊的模型將呈對角線狀，即從左下角到右上角（選取了含基準線後將顯示類似圖表）。

**增益圖。**累加增益圖的線從左到右的走勢通常為：起始於大於 1.0 的值，並漸漸下降，直到接近 1.0。圖表的右側邊緣代表整個資料集，因此累加分位數的匹配與資料中的匹配的比例為 1.0。對於優秀模型的提升圖，其線從圖表左端大於 1.0 的值開始，且在向右移動的過程中，始終保持在較高的水平；然後，在圖表右端，向 1.0 的方向迅速下降。如果模型不能提供任何資訊，那麼其行在整個圖表中將始終圍繞在 1.0 左右。（如果選取了**包含基準線**，一條值為 1.0 的水平行將顯示在圖表中供您參照。）

**回應圖表。**累加回應圖表通常與增益圖極其類似，只在尺度標準方面有所區別。通常，回應圖從接近 100% 的地方開始，並逐漸下降，最終將在延伸至圖表右側邊緣時達到整體回應率（全部匹配/全部記錄）。對於優秀模型的回應圖，其行從圖表左端接近或等於 100% 的值開始，且在向右移動的過程中，始終保持在較高的水平；然後，在圖表右端，向整體回應率的方向迅速下降。如果模型不能提供任何資訊，那麼其行在整個圖表中將始終圍繞在整體回應率左右。（如果選取了**含基準線**，一條值相當於整體回應率的水平行將顯示在圖表中供您參照。）

**利潤圖。**累加利潤圖線從左到右的走勢代表隨著所選樣本數量的增加，利潤總和的增長。利潤圖通常從 0 附近開始，並在向右延伸的過程中，穩步增長直至在圖表中部到達尖峰或保持較高的值；隨後，在向右側邊緣延伸的過程中，逐漸下降。優秀模型的利潤圖將在圖表中部某處顯示定義良好的尖峰。而無法提供任何資訊的模型，其行相對而言比較平直，也可能由於成本/營收結構的不同增加、降低或保持不變。

**投資報酬率圖表。**累加投資返回 (ROI) 圖通常與回應圖表及提升圖類似，只有在尺度標準方面有所差別。投資報酬率圖表通常從大於 0% 的值開始，並逐漸下降，直到達到整個資料集的整體 ROI（可能為負）。對於優秀模型的投資回報圖，其行從圖表左端大於 0% 的值開始，且在向右移動的過程中，始終保持在較高的水平；然後，在圖表右端，向整體 ROI 的方向迅速下降。如果模型不能提供任何資訊，則其行在整個圖形中將始終圍繞在整體 ROI 左右。

**ROC 圖表**ROC 曲線的形狀通常為累加收益圖表。該曲線從 (0,0) 座標開始，到 (1,1) 座標結束，方向為從左到右。圖表曲線朝 (0,1) 座標急劇上升隨後趨於平穩，這表示分類器較好。將實例隨機分類為命中數或未命中數的模型將呈對角線狀，即從左下角到右上角（如果選取了**含基準線**，那麼此對角線將顯示在圖表中）。如果未提供模型的信賴度欄位，那麼模型將繪製為單個點。具有最優分類臨界值的分類器位於最接近 (0,1) 座標的位置或者圖表的左上角落。此位置代表正確分類為命中數的實例數較多，並且錯誤分類為命中數的實例數較少。對角線上方的點代表較好的分類結果。對角線下方的點代表較低劣的分類結果，這些結果比對實例進行隨機分類的結果較劣。

## 使用評估表

用滑鼠檢查評估表的方法與在直方圖或收集圖中相同。x 軸代表指定分位數（如二十分位數或十分位數）的模型分數。

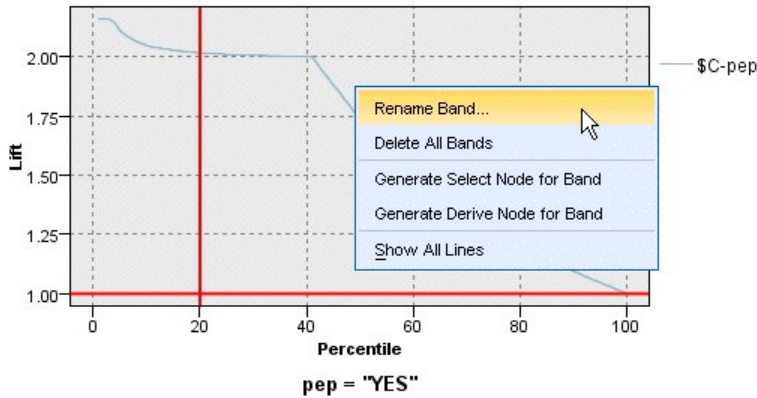


圖 47. 處理評估表

您可以像處理直方圖那樣，使用分割標誌顯示將軸自動分為等寬級區的選項，以將評估圖表的  $x$  軸分為級區。請參閱第 239 頁的『探索圖形』主題，以取得更多資訊。您可以選取「編輯」功能表的圖表級區狀區域來手動編輯級區的範圍。

在建立了評估表、定義了級區並查看了結果之後，您可以使用「產生」功能表上的選項及快速功能表根據圖表中的選擇自動建立節點。請參閱第 246 頁的『從圖形產生節點』主題，以取得更多資訊。

從評估表中產生節點時，將提示您選取圖表中所有可用模型中的一個。

選取一個模型並按一下確定在串流畫布中產生新節點。

## 「地圖視覺化」節點

「地圖視覺化」節點可以接受多個輸入連線並在地圖上將地理空間資料顯示為一系列層。每一層都是單一地理空間欄位；例如，基本層可能是某個國家或地區的地圖，而在該層之上，可能有一層代表道路，一側代表河流，以及一層代表城鎮。

雖然大部分地理空間資料集通常包含單個地理空間欄位，但是單個輸入中存在多個地理空間欄位時，您可以選擇要顯示的欄位。同一個輸入連線中的兩個欄位不能同時顯示；但是，您可以複製並貼上傳入連線並顯示這兩個連線中的不同欄位。

## 地圖視覺化「繪圖形」標籤

### 階層

該表格顯示有關地圖節點的輸入的資訊。各個層的順序指示了執行該節點時各個層在地圖預覽和可視輸出中的顯示順序。表格中的頂列是「頂層」，底列是「底層」；換言之，每一層在地圖上都顯示在表格中正好位於其下方的層前面。

註：如果表格中的層包含三維度地理空間欄位，那麼將僅繪製  $x$  和  $y$  軸。將忽略  $z$  軸。

**名稱** 對於每一層，將自動建立名稱，其格式如下：tag[source node:connected node]。依預設，標籤顯示為數字，1 代表所連接的第一個輸入，2 代表第二個輸入，依此類推。如有需要，請按編輯層按鈕，以便在「變更地圖層選項」對話框中變更標籤。例如，可以將標籤變更為「道路」或「城市」以反映資料輸入。

**類型** 顯示已選作層的地理空間欄位的測量類型圖示。如果輸入資料包含多個具有地理空間測量類型的欄位，那麼預設選擇將使用下列排序：



1. 點
2. Linestring
3. 多邊形
4. 多點
5. MultiLinestring
6. 多個多邊形

註：如果存在兩個具有相同測量類型的欄位，那麼依預設將已選取第一個欄位（依名稱的按字母順序排列）。

## 符號

註：僅對於點和多點欄位才會填寫此欄。

顯示用於點或多點欄位的符號。如有需要，請按編輯層按鈕，以便在「變更地圖層選項」對話框中變更符號。

**色彩** 顯示已選取用於在地圖上代表層的顏色。如有需要，請按編輯層按鈕，以便在「變更地圖層選項」對話框中變更顏色。根據測量類型的不同，此顏色將套用至不同的項目。

- 對於點或多點，此顏色將套用至層的符號。
- 對於線串和多邊形，此顏色將套用至整個形狀。多邊形始終具有黑色輪廓；此欄中顯示的顏色是用於填充該形狀的顏色。

## 預覽

此窗格顯示層表格中目前選擇的輸入的預覽。此預覽將考慮層順序、符號、顏色以及任何其他與層相關聯的顯示設定，如有可能，每當這些設定發生變更時，預覽都會更新顯示。如果您在串流中的其他位置變更了詳細資料，例如變更了要用作層的地理空間欄位，或者修正了詳細資料（例如相關聯的聚集函數），那麼可能需要按一下重新整理資料按鈕以更新預覽。

在執行串流之前，請使用預覽來設定顯示設定。為了避免因使用大型資料集而造成時間延遲，預覽功能將對每個層進行抽樣，並根據前 100 個記錄來建立顯示內容。

## 變更地圖層

您可以使用「變更地圖層選項」對話框來修正「地圖視覺化」節點統計圖標籤上顯示的任何層的各种詳細資料。

## 輸入詳細資料

**標籤** 依預設，標籤是一個數字；您可以將此數字取代為更有意義的標籤，以說明在地圖上識別層。例如，此標籤可以是資料輸入的名稱，如「城市」。

## 層欄位

如果輸入資料中存在多個地理空間欄位，請使用此選項來選取要在地圖上顯示為層的欄位。

依預設，可供選擇的層按下列順序排列。

- 點
- Linestring
- 多邊形
- 多點
- MultiLinestring



- 多個多邊形

## 顯示設定

### 使用六邊形 bin

註：此選項只影響點欄位和多點欄位。

六邊形（六角形）bin 根據臨近點的 x 和 y 座標將這些點組合成單個點以顯示在地圖上。這一個點顯示為六角形，但實際上呈現為多邊形。

由於六角形呈現為多邊形，因此開啟了六邊形 bin 的任何點欄位都將被視為多邊形。這意味著，如果在地圖節點對話框中選擇了**排序方式類型**，那麼套用了六邊形 bin 的任何點層都將呈現在多邊形層上方，但在線串層和點層下方。

如果將六邊形 bin 用於多點欄位，那麼會先通過對多點值進行 bin 來計算中心點，將該欄位轉換為點欄位。中心點用於計算六邊形 Bin。

## 聚集

註：僅當選中了**使用六邊形 bin**複選框，並選中了**套版**時，此欄才可用。

如果您為使用了六邊形 bin 的點層選中了**套版**欄位，那麼該欄位中所有的值都必須針對該六邊形內的所有點進行聚集。請為任何要套用至地圖的任何套版欄位指定聚集函數。可用的聚集函數取決於測量類型。

- 對於「實數」或「整數」儲存，適用於「連續」測量類型的聚集函數包括：
  - 總和
  - 平均數
  - 下限
  - 上限
  - 中位數
  - 第一個四分位數值
  - 第三個四分位數值
- 對於「時間」、「日期」或「時間戳記」儲存，適用於「連續」測量類型的聚集函數包括：
  - 平均數
  - 下限
  - 上限
- 適用於「列名」或「種類」測量類型的聚集函數包括：
  - 模式
  - 下限
  - 上限
- 適用於「旗標」測量類型的聚集函數包括：
  - True（如果任何項為 true）
  - False（如果任何項為 false）

## 色彩

套用此選項可以選擇標準顏色（此顏色將套用至地理空間欄位的所有功能）或套版欄位（此欄位比資料中另一欄位的值對各個特徵進行著色）。

如果您選擇了標準，那麼可以從「使用者選項」對話框「顯示」標籤圖表種類顏色順序窗格顯示的選用區中選擇顏色。

如果您選取重疊，則可以從包含已選取作為層欄位之地理空間欄位的資料來源中選擇任何欄位。

- 對於列名或種類套版欄位，可以從中進行選擇的調色盤與針對標準顏色選項顯示的調色盤相同。
- 對於連續和序數套版欄位，將顯示另一個下拉清單，您可以從中選取顏色。您選取顏色時，將比該連續或序數欄位中的值改變該顏色的飽和度，從而套用套版。最大的值將使用從下拉清單中選擇的顏色，較小的值由相應較低的飽和度顯示。

## 符號

註：僅對點測量類型和多點測量類型啟用。

使用此選項可以選擇是使用標準符號（此符號將應用於地理空間欄位的所有記錄），還是使用套版符號（這將比資料中另一個欄位的值來變更各個點的符號圖示）。

如果您選擇了標準，那麼可以從下拉清單中選擇其中一個預設符號，用於在地圖上代表點資料。

如果您選擇了套版，那麼可以從選作層欄位的地理空間欄位所在的資料來源中選擇任意列名欄位、序數欄位或種類欄位。對於套版欄位中的每個值，都將在地圖上顯示不同的符號。

例如，您的資料可能包含代表商店位置的點欄位，並且套版可能是商店類型欄位。在此範例中，所有飲食店在地圖上可能由十字符號 ID，而所有電器店由方形符號 ID。

## 大小

註：僅對點、多點、線串和多線串測量類型啟用。

使用此選項可以選擇是使用標準大小（此大小將應用於地理空間欄位的所有記錄），還是使用套版大小（這將比資料中另一個欄位的值來變更符號圖示大小或行寬）。

如果您選擇了標準，那麼可以選擇像素寬度值。可用的選項包括 1、2、3、4、5、10、20 或 30。

如果您選取重疊，則可以從包含已選取作為層欄位之地理空間欄位的資料來源中選擇任何欄位。行寬或點的大小根據所選欄位的值不同而有所不同。

## 透明

套用此選項可以選擇是套用標準透明（此透明將套用至地理空間欄位的所有記錄），還是套用套版透明（這將比資料中另一個欄位的值來變更符號、線條或多邊形的透明度）。

如果您選擇了標準，那麼可以從一組由 0%（不透明）開頭並以 10% 遞增到 100%（透明）的透明層次中進行選擇。

如果您選取重疊，則可以從包含已選取作為層欄位之地理空間欄位的資料來源中選擇任何欄位。在地圖上，對於套版欄位中的每個值，將顯示不同層次的透明。此透明將套用至從顏色下拉清單中為點、線條或多邊形選擇的顏色。

## 資料標籤

註：如果選中了使用六邊形 bin 複選框，那麼此選項無法使用。

使用此選項可以選取要在地圖上用作資料標籤的欄位。例如，如果套用至多邊形層，那麼資料標籤可能是名稱欄位，其中包含每個多邊形的名稱。如果您選取了名稱欄位，那麼這些名稱將顯示在地圖上。

## 地圖視覺化「外觀」標籤

您可以在建立圖形之前指定外觀選項。

標題。輸入要用作圖表標題的文字。

次標題。輸入要用作圖表次標題的文字。

標題。輸入要用作圖表標題的文字。

---

## t-SNE 節點

t-Distributed Stochastic Neighbor Embedding (t-SNE)<sup>©</sup> 是一套用於視覺化高維度資料的工具。它會將資料點的親緣性轉換為可能性。原始空間中的親緣性由 Gaussian 聯合機率代表，而內嵌空間中的親緣性則是由「學生 t 分布 (Student's t-distributions)」代表。這可讓 t-SNE 對於本端結構，並且擁有幾個超越現有技術的優點：<sup>1</sup>

- 在單一地圖上以許多比例顯示結構
- 顯示位於多個、不同、各種、叢集的資料
- 降低人群一起指向中心的傾向

此 t-SNE 節點在 SPSS Modeler 中使用 Python 進行實現並且需要 scikit-learn<sup>©</sup> Python 程式庫。有關 t-SNE 和 scikit-learn 程式庫的詳細資料，請參閱：

- <https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html#sklearn.manifold.TSNE>
- <https://scikit-learn.org/stable/modules/manifold.html#t-sne>

節點選用區上的 Python 標籤包含此節點和其他 Python 節點。「圖表」標籤上也可以使用 t-SNE 節點。

<sup>1</sup> 參考資料：

van der Maaten, L.J.P.; Hinton, G. "Visualizing High-Dimensional Data using t-SNE." Journal of Machine Learning Research. 9:2579-2605, 2008.

van der Maaten, L.J.P. "t-Distributed Stochastic Neighbor Embedding."

van der Maaten, L.J.P. "Accelerating t-SNE using Tree-Based Algorithms." Journal of Machine Learning Research. 15(Oct):3221-3245, 2014.

## t-SNE 節點專家選項

選擇簡式模式或專家模式，視您要對 t-SNE 節點設定的選項而定。

視覺化類型。選取 **2D** 或 **3D**，指定要以二維還是三維來繪製圖表。

方法。選取 **Barnes Hut** 或 **Exact**。依預設，梯度計算演算法會使用 Barnes-Hut 近似值，其執行速度遠比 Exact 方法快。Barnes-Hut 近似值可讓 t-SNE 技術能套用到大型的實際資料集。Exact 演算法在避免最近鄰接項錯誤方面的運作較好。

起始設定。選取隨機或 **PCA** 作為內嵌的起始設定。

目標欄位。選取要顯示為輸出圖表上的彩色圖的目標欄位。如果未在這裡指定目標欄位，圖表將會使用一個顏色。

## 最佳化

**複雜度。** 複雜度與其他各種學習演算法中所用的最近鄰接項數目相關。較大的資料集通常需要較大的複雜度。請考量選取介於 **5** 與 **50** 之間的值。預設值是 **30**，範圍是 **2 - 9999999**。

**提早誇大。** 這項設定控制原始空間中的自然叢集在內嵌空間中將會有多緊，以及它們之間將會有多少空間。預設值是 **12**，範圍是 **2 - 9999999**。

**學習率。** 如果學習速率過高，則資料可能會類似於「球狀」，其中任何點與其最近鄰接項的距離均大約相等。如果學習速率過低，則大部分點可能壓縮在一個密集雲端中而幾乎沒有偏離值。如果成本函數停留在一個不當本端下限中，則增加學習速率可能有用。預設值是 **200**，範圍是 **0 - 9999999**。

**最大反覆運算數。** 用於最佳化的反覆運算數目上限。預設值是 **1000**，範圍是 **250 - 9999999**。

**角距大小。** 從一個點所測量的遠距節點的角距大小。輸入介於 **0** 和 **1** 之間的值。預設值是 **0.5**。

## 隨機種子

**設定隨機種子。** 選取此選項並按一下產生可以產生由亂數字產生器使用的種子。

## 最佳化停止條件

**沒有進度的最大疊代數。** 停止最佳化之前要執行的沒有進度的最大反覆運算數，在提前放大的情況下執行 250 次起始反覆運算之後使用。請注意，每 50 次反覆運算才檢查一次進度，因此這個值四捨五入為 50 的下一個倍數。預設值是 **300**，範圍是 **0 - 9999999**。

**最小梯度規範。** 如果梯度規範低於此最小臨界值，最佳化將會停止。預設值是 **1.0E-7**。

**度量值。** 計算特徵陣列中實例之間的距離時要使用的度量。如果度量是字串，則它必須是其度量參數的 `scipy.spatial.distance.pdist` 所容許的其中一個選項，或是 `pairwise.PAIRWISE_DISTANCE_FUNCTIONS` 中列出的度量。選取其中一個可用的度量類型。預設值是 **euclidean**。

**當記錄數大於。** 指定用來繪製大量資料集的方法。可以指定資料集大小上限，或使用預設的 2000 點。當您選取 **Bin** 或**樣本**選項時，大型資料集的效能會加強。或者，您可以選擇透過選取**使用所有資料**來繪製所有資料點，但您應該注意到這樣可能會極大地降低軟體效能。

- **Bin。** 選取此項以在資料集包含的記錄數大於指定的記錄數時啟用 Binning。「分級」使圖表在實際繪製前被分散在較小的網格中，並計算在每個單元中將出現的連線數。在最終圖表中，每個網格中的分級矩心處將使用一個連線（該連線即代表分級中所有連線點位置的平均值數）。
- **樣本。** 選取此選項將隨機抽取指定記錄數的資料樣本。

下表顯示了 SPSS Modeler t-SNE 節點對話框的「專家」標籤上的設定與 Python t-SNE 程式庫參數之間的關係。

表 36. 對映至 Python 程式庫參數的節點內容

SPSS Modeler 設定	Script 名稱 (內容名稱)	Python t-SNE 參數
模式	mode_type	
視覺化類型	n_components	n_components
方法	method	method
內嵌起始設定	init	init
目標	target_field	target_field

表 36. 對映至 Python 程式庫參數的節點內容 (繼續)

SPSS Modeler 設定	Script 名稱 (內容名稱)	Python t-SNE 參數
複雜度	perplexity	perplexity
提早誇大	early_exaggeration	early_exaggeration
學習率	learning_rate	learning_rate
最大疊代數	n_iter	n_iter
角距大小	angle	angle
設定隨機種子	enable_random_seed	
隨機種子	random_seed	random_state
沒有進度的最大疊代數	n_iter_without_progress	n_iter_without_progress
最小梯度規範	min_grad_norm	min_grad_norm
使用多重複雜度來執行 t-SNE	isGridSearch	

## t-SNE 節點輸出選項

在輸出標籤上指定 t-SNE 節點輸出的選項。

**輸出名稱。**指定節點執行時產生的輸出名稱。如果您選取**自動**，會自動設定輸出的名稱。

**輸出至畫面。**選取此選項可在新視窗中產生並顯示輸出。輸出也會新增至「輸出」管理程式。

**輸出至檔案。**選取此選項可將輸出儲存至檔案。這麼做會啟用**檔名**和**檔案類型**欄位。如果您想要使用其他欄位來建立圖形以用於比較目的 – 或想要使用其輸出作為分類或迴歸模型中的預測值，則 t-SNE 節點需要存取此輸出檔。t-SNE 模型會建立一個包含 x、y (及 z) 座標欄位的結果檔案，可使用「固定檔案」來源節點輕鬆地存取此檔案。如需相關資訊，請參閱 @@@@。

下表格顯示了 SPSS Modeler t-SNE 節點對話框的「輸出」標籤上的設定與 Python t-SNE 程式庫參數之間的關係。

表 37. 對映至 Python 程式庫參數的節點內容

SPSS Modeler 設定	Script 名稱 (內容名稱)	Python t-SNE 參數
輸出名稱	output_Rename	output_Rename
輸出模式	output_to	output_to
檔名	full_filename	full_filename
檔案類型	output_file_type	output_file_type
目標	target_field	target_field

## 存取及繪製 t-SNE 資料

如果使用**輸出至檔案**選項將 t-SNE 輸出儲存至檔案，則您隨後可以使用其他欄位來建立圖形以用於比較目的 – 或使用輸出作為分類或迴歸模型中的預測值。t-SNE 模型會建立一個包含 x、y (及 z) 座標欄位的結果檔案，可使用「固定檔案」來源節點輕鬆地存取此檔案。本節提供範例資訊。

- 在 t-SNE 節點對話框中，開啟**輸出**標籤。
- 選取**輸出至檔案**並鍵入檔名。使用預設 HTML 檔案類型。執行模型時，將會在輸出位置中產生三個輸出檔：
  - 文字檔 (result\_xxxxx.txt)

- HTML 檔 (您指定的檔名)
- PNG 檔 (tsne\_chart\_yyyyyy.png)

文字檔將包含您需要的資料，但出於技術原因，此檔案可能採用標準或科學格式。如果採用科學格式 (1.1111111e+01)，則您需要建立新串流來辨識該格式：

### 當文字檔採用科學數字格式時存取 t-SNE 圖形資料

1. 建立新串流 (檔案 > 新串流)。
2. 跳至工具 > 串流內容 > 選項，選取數字格式，並選取科學 (#.###E+##) 作為數字顯示格式。
3. 將「固定檔案」來源節點新增至畫布，並在「檔案」標籤上使用下列設定：
  - 跳過標頭行：1
  - 記錄長度：54
  - tSNE\_x 開始：3，長度：16
  - tSNE\_y 開始：20，長度：16
  - tSNE\_z 開始：36，長度：16
4. 在「類型」標籤上，數字應該辨識為「實數」。按一下「讀取值」，您應看到類似於下列內容的欄位值：

表 38. 欄位值範例

欄位	測量	值
tSNE_x	連續的	[-7.07176703,7.14338837]
tSNE_y	連續的	[-9.2188112,8.89647667]
tSNE_x	連續的	[-9.95892882,9.95742482]

5. 將「選取」節點新增至串流，以便您可以在檔案中刪除底端的下列兩行讀取為空值的文字：

```
*****
Perform t-SNE (total time 9.5s)
```

在「選取」節點的「設定」標籤上，針對眾數選取捨棄並使用條件 @NULL(tSNE\_x) 來刪除列。

6. 將「類型」節點及「純文字檔」匯出節點新增至串流以建立變數。複製「檔案」來源節點並將其貼回原始串流。

### 當文字檔採用標準數字格式時存取 t-SNE 圖形資料

1. 建立新串流 (檔案 > 新串流)。
2. 將「固定檔案」來源節點新增至畫布。以下是存取 t-SNE 資料所需的全部三個節點。



圖 48. Stream for accessing t-SNE plot data in standard numeric format

3. 在「固定檔案」來源節點的「檔案」標籤上使用下列設定：
  - 跳過標頭行：1
  - 記錄長度：29

- tSNE\_x 開始：3，長度：12
  - tSNE\_y 開始：16，長度：12
4. 在「過濾器」標籤上，您可以將 field1 及 field2 重新命名為 tsneX 及 tsneY。
  5. 新增「合併」節點以使用順序合併方法將其連接至串流。
  6. 現在，您可以使用「繪圖」節點來繪製 tsneX 與 tsneY，並使用調查下的欄位對其進行著色。

## t-SNE 模型塊

t-SNE 模型塊包含 t-SNE 模型所擷取的所有資訊。可用的標籤如下。

### 圖表

圖表標籤顯示 t-SNE 節點的圖表輸出。pyplot 散佈圖顯示低維度結果。如果您未選取 t-SNE 節點的專家標籤上的使用多重複雜度來執行 t-SNE 選項，則只會包含一個圖表而不是六個具有不同複雜度的圖表。

### 文字輸出

文字輸出標籤顯示 t-SNE 演算法的結果。如果您在 t-SNE 節點的專家標籤上選擇 **2D** 視覺化類型，則這裡的結果為二維點值。如果您選擇 **3D**，則結果為三維點值。

---

## E-Plot (測試版) 節點

E-Plot (測試版) 節點顯示數值欄位之間的關係。E-Plot (測試版) 節點與繪圖節點類似，但選項不同，且它使用新的圖形繪製功能。將此節點與 SPSS Modeler 中的新圖形繪製功能搭配使用。

E-Plot (測試版) 節點提供散佈圖、折線圖及長條圖來說明數值欄位之間的關係。此節點中的新圖形繪製介面比較直觀且具現代感、非常容易自訂，且資料圖表是互動式圖表。如需相關資訊，請參閱第 236 頁的『使用 E-Plot 圖形』。

### E-Plot (測試版) 節點繪圖標籤

統計圖對照 X 欄位的值，顯示 Y 的值。通常而言，這兩個欄位分別對應於一個應變數和一個自變數。

**X 欄位。** 從清單中，選取要在水平 x 軸上顯示的欄位。

**Y 欄位。** 從清單中選取顯示在垂直 y 軸上的欄位。

**重疊。** 有不同方式來說明資料值的種類。例如，可以使用 *maincrop* 欄位作為顏色套版來指出補貼申請人種植的主要作物的 *estincome* 和 *claimvalue* 值。選取在輸出中用於顏色對映、大小對映及形狀對映的欄位。另外請選取要在互動式輸出中包括的任何其他相關欄位。請參閱第 162 頁的『外觀、套版、畫面和動畫』主題，以取得更多資訊。

當您為 E-Plot 設定了選項後，可以透過按一下對話框中的執行來直接執行繪圖。但是，您也許希望使用「選項」標籤指定更多內容。

### E-Plot (測試版) 節點選項標籤

**要繪製的上限記錄數。** 指定用來繪製大量資料集的方法。可以指定上限資料集大小，或使用預設值 2,000 條記錄。如果選取取樣選項，那麼處理大資料集時的效能將會提高。「取樣」選項會對資料進行隨機取樣，直至達到文字欄位中輸入的記錄數目。或者，您可以選擇透過選取使用所有資料來繪製所有資料點，但您應該注意到這樣可能會極大地降低軟體效能。

## E-Plot (測試版) 外觀標籤

在建立圖形之前，可以根據需要指定標題和子標題。也可以在建立圖形之後指定或變更這些選項。

**標題。** 輸入要用作圖表標題的文字。

**次標題。** 輸入要用作圖表次標題的文字。

## 使用 E-Plot 圖形

E-Plot (測試版) 節點提供散佈圖、折線圖及長條圖來說明數值欄位之間的關係。此測試版節點中引進的新圖形繪製介面包含很多新增及改良功能。

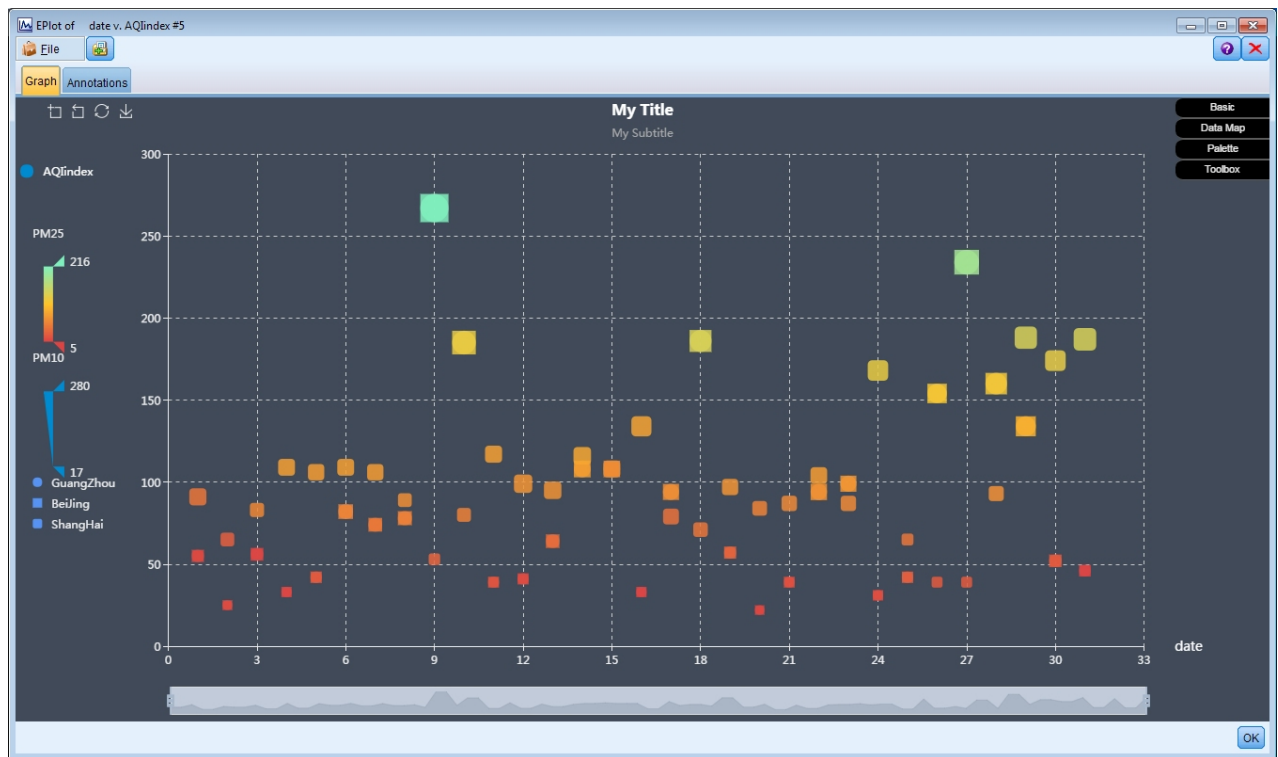


圖 49. E-Plot (Beta) scatterplot graph

「圖形」標籤的左上角提供了一個工具列，用來放大圖表的特定部分、取消縮放、回到起始完整視圖，或儲存圖表以在外部使用：



圖 50. Toolbar

在視窗底端，您可以使用調節器來放大圖表的特定部分。左右移動小矩形控制項以進行縮放。若要使用此調節器，您必須先在「工具箱」選項區域中將其開啟。





圖 51. Zoom slider

視窗左側提供了一些控制項來變更所顯示值的範圍。若要使用這些控制項，您必須先在「資料對映」選項區域中指定選項。在以下範例中，欄位 PM25 選取為用於色彩對映，欄位 PM10 選取為用於大小對映，而欄位 City 選取為用於形狀對映。您可以將滑鼠指標移至垂直顏色條欄上以強調顯示圖形的對應區域，或上下滑動三角形。



圖 52. Range controls

在視窗右側上，提供了一組可展開的選項，您可以用來與資料互動並即時變更圖表外觀：



圖 53. Expandable options

## 基本選項

	<p>選取暗或亮佈景主題，指定標題和子標題，選取圖表類型（散佈圖、折線圖或長條圖），並選取 Y 軸上顯示的序列。如果您選取折線圖，則將僅顯示 Y 軸上的欄位，且色彩對映及大小對映的「資料對映」選項中將只能使用 Y 軸上的欄位。如果您選取長條圖，則「資料對映」選項中將只能使用色彩對映選項。對於序列，您在 E-Plot 節點的「繪圖」標籤上選取的所有相關欄位都將可在這裡使用。</p>
--	---

圖 54. Basic options

## 資料對映選項

	<p>為色彩對映選取連續欄位或類別欄位。如果選取連續欄位，則將顯示從綠色到紅色的顏色。值越低，其顏色就越接近紅色 – 值越高，其顏色就越接近綠色。如果選取類別欄位，則將會根據定義的調色盤顯示欄位顏色。 大小對映僅支援連續欄位。圖表上的值越低，其繪製大小就越小。 形狀對映僅支援類別欄位。對映上顯示的形狀由一個類別欄位定義，該欄位會將視覺化分割成不同形狀的元素，每種類別各一種形狀。</p>
--	--

圖 55. Data map options

## 選用區選項

	<p>如果要自訂用於標題及序列的顏色，請使用「選用區」。從下拉清單中選取標題或序列，按一下編輯預先定義的顏色，然後按一下更多以選取顏色。您也可以使用 RGB 或十六進位欄位來指定確切顏色。</p>
--	--

圖 56. Palette options

## 工具箱選項


	使用「工具箱」選項來開啟或關閉縮放調節器，設定格線內容，以及開啟或關閉滑鼠追蹤。滑鼠追蹤會在您將滑鼠指標移至圖表上時顯示您的確切座標位置。
---	---

圖 57. Toolbox options

## 探索圖形

使用編輯模式可以編輯圖表的佈置和外觀，使用探索模式可以分析視角探索圖表代表的資料和值。探索的主要目的是分析資料，然後使用級區、區域和標示來識別值，以產生「選取」節點、「衍生」節點或「平衡」節點。要選擇此模式，請從功能表中選擇視圖 > 探索模式（或按一下工具列圖示）。

有些圖形可以使用所有探索工具，而有些圖形僅接受一個探索工具。探索模式包含：

- 定義和編輯級區，這些級區用於沿著尺度  $x$  軸分隔值。請參閱第 240 頁的『使用級區』主題，以取得更多資訊。
- 定義和編輯區域，這些區域用於識別矩形區域中的一群組值。請參閱第 243 頁的『使用區域』主題，以取得更多資訊。
- 對元素進行標示和取消標示，以手動選取可用於產生「選取」節點或「衍生」節點的值。請參閱第 245 頁的『使用標示元素』主題，以取得更多資訊。
- 使用由級區、區域、標示元素和 Web 鏈結 ID 的值產生可在串流中使用的節點。請參閱第 246 頁的『從圖形產生節點』主題，以取得更多資訊。

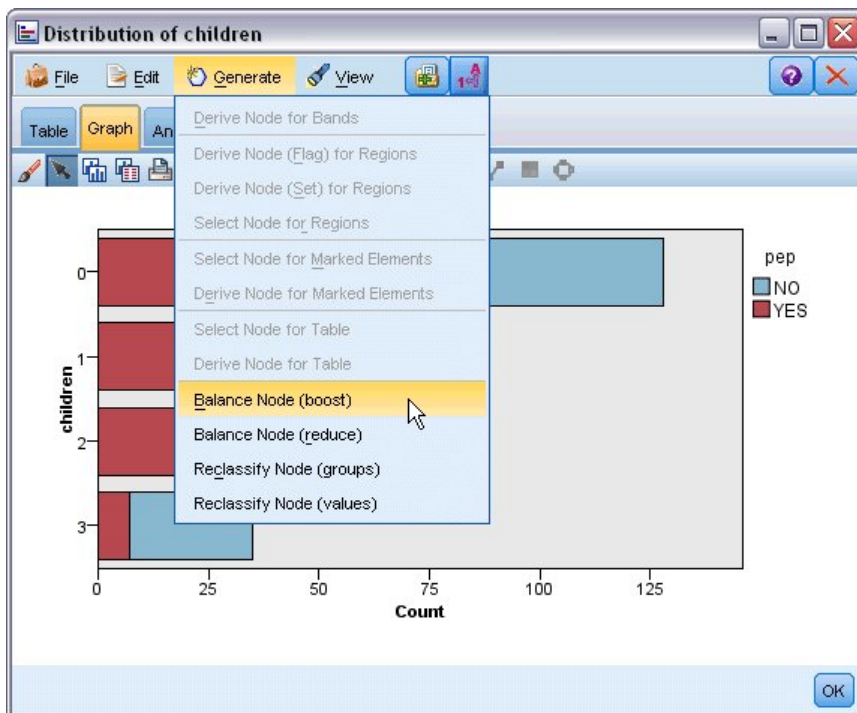


圖 58. 顯示產生功能表的圖形

## 使用級區

對於在  $x$  軸上存在尺度欄位的任一圖表，通過繪製垂直級區行可分割  $x$  軸上的值範圍。如果圖表中包含多個畫面，那麼在一個畫面上繪製的級區行也會顯示在其他畫面上。

不是所有圖形都能夠接受級區。其中一些可以包含級區的圖形包含：直方圖、長條圖和分佈圖形、散點圖（行、散點、時間等）、收集圖形和評估表。在使用畫面的圖形中，級區會顯示在所有畫面中。而在某些情況下，SPLOM 中會顯示為水平級區行，這是因為繪製欄位/變數級區所在的軸已翻轉。

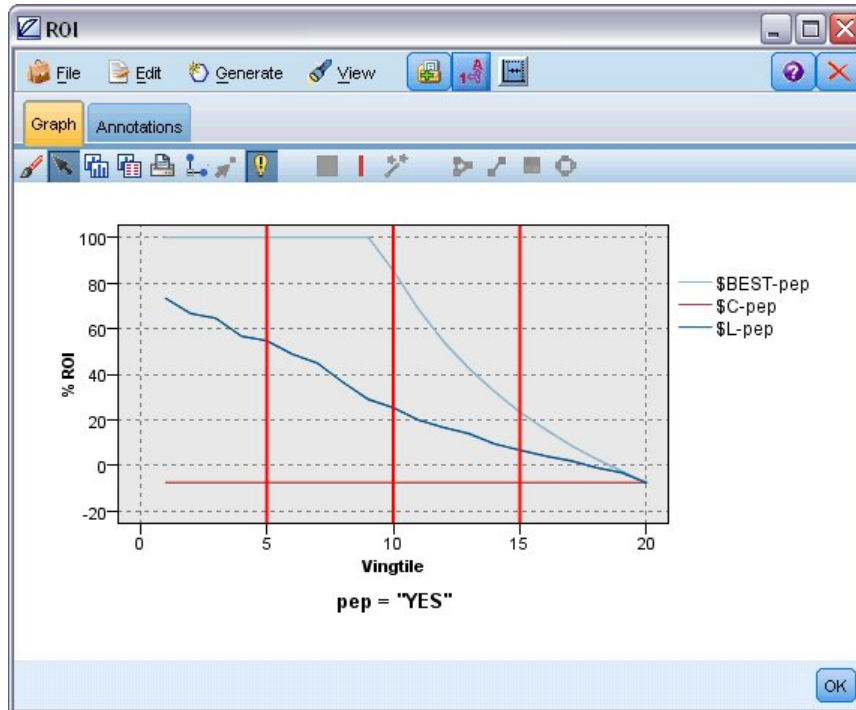


圖 59. 包含三個級區的圖表

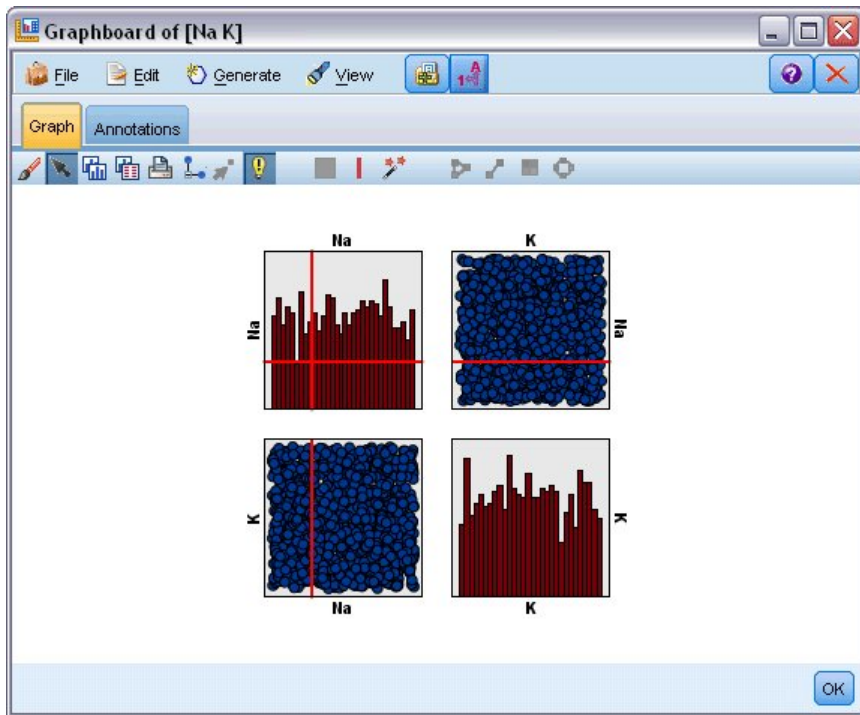


圖 60. 包含級區的 SPLOM

### 定義級區

在不包含級區的圖表中，新增級區行可將圖表分割為兩個級區。級區行值代表從左到右查看圖表時第二個級區的起點，也稱為下限。同樣，在包含兩個級區的圖表中，新增級區行可將其中一個級區分割為兩部分，最終形成三個級區。依預設，級區命名為 *bandN*，其中 *N* 相當於沿 *x* 軸從左到右排列的級區的序號。

定義級區後，通過拖放級區可在 *x* 軸上重新定位。通過在級區內按一下右鍵可以請參閱更多的快速鍵，這些快速鍵用於執行重命名、刪除或產生該特定級區節點之類的作業。

### 要定義級區：

1. 驗證您是否處於瀏覽模式。從功能表中選擇檢視 > 瀏覽模式。
2. 在「探索模式」工具列中，按一下「繪製條帶」按鈕。



圖 61. 「繪製級區」工具列按鈕

3. 在接受級區的圖表中，按一下要定義級區行所在的 *x* 軸值點。

註：或者，按一下將圖表分割為級區工具列圖示，然後輸入所需的相等級區數量，然後按一下分割。



圖 62. 通過用於分割為條帶區域的選項擴展工具列的分割圖示



圖 63. 啟用了級區的「建立相等級區」工具列

編輯、重命名和刪除級區狀區域

可以在「編輯圖表級區」對話框中或通過圖表自身的快速功能表編輯現有級區的內容。

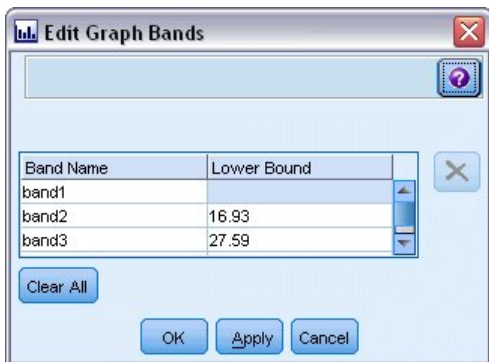


圖 64. 「編輯圖表級區狀區域」對話框

#### 要編輯級區：

1. 驗證您是否處於瀏覽模式。從功能表中選擇檢視 > 瀏覽模式。
2. 在「探索模式」工具列中，按一下「繪製條帶」按鈕。
3. 從功能表中選擇編輯 > 圖表級區狀區域。此時將開啟「編輯圖表級區狀區域」對話框。
4. 如果圖形中有多個欄位（例如 SPLOM 圖形），可以在下拉清單中選取所需欄位。
5. 通過輸入名稱和下限新增新的級區。按 Enter 鍵啟動新列。
6. 通過調整下限值編輯級區的界限。
7. 通過輸入新的級區名稱更名級區。

8. 通過選取表格中的行並按一下「刪除」按鈕，可刪除級區。
9. 按一下**確定**以套用您的變更並關閉對話框。

註：或者，通過用滑鼠右鍵按一下級區所在的行，並從快速功能表中選擇所需選項，可直接刪除和更名圖表中的級區。

## 使用區域

在包含兩個尺度（或範圍）軸的任意圖表中，通過繪製區域可在繪製好的矩形區域（稱為區域）中對值進行群組。區域為圖表中用 X 和 Y 的最小值和最大值說明的一塊區域。如果圖表中包含多個畫面，那麼在一個畫面上繪製的區域也會顯示在其他畫面上。

不是所有圖形都能夠接受區域。其中一些能夠接受區域的圖形包含：散點圖（行、散點、氣泡、時間等）、SPLOM 圖形和收集圖形。這些區域是在 X、Y 空間中繪製的，因此無法在一維、三維或動畫統計圖中定義。在使用畫面的圖形中，區域會顯示在所有畫面中。在散佈圖矩陣圖形 (SPLOM) 中，對應的區域會顯示在對應的上方散佈圖中，而不是顯示在對角線散佈圖中，因為後者只顯示一個尺度欄位。

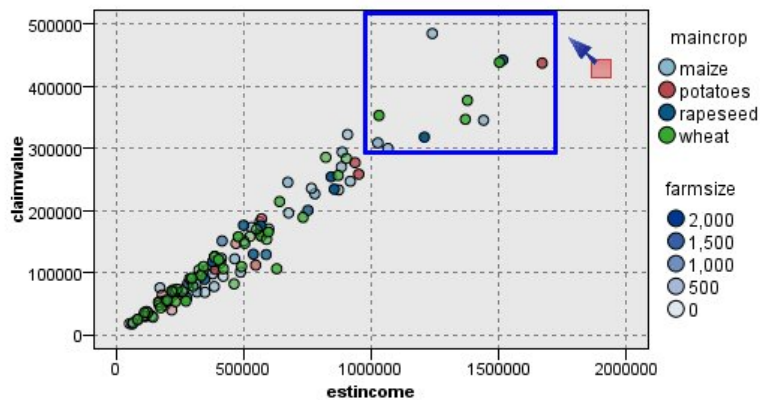


圖 65. 定義包含高索賠值的區域

### 定義區域

無論在哪个位置上定義區域，都是建立值的分組。依預設，每個新區域都稱為 *Region<N>*，其中 *N* 表示已經建立的區域數量。

定義區域後，通過用滑鼠右鍵按一下區域行可存取一些基本的快速鍵。但是，通過在區域內（而不是在行上）按一下右鍵可以看到多數其他的快速鍵，這些快速鍵用於執行重命名、刪除或產生該特定區域的「選取」和「衍生」節點之類的作業。

您可以根據記錄是併入在某個特定區域還是若干區域中的某個區域來選取記錄的子集合。還可以通過產生「衍生」節點，根據記錄是否併入在區域中來旗標記錄，從而併入記錄的區域資訊。請參閱第 246 頁的『從圖形產生節點』主題，以取得更多資訊。



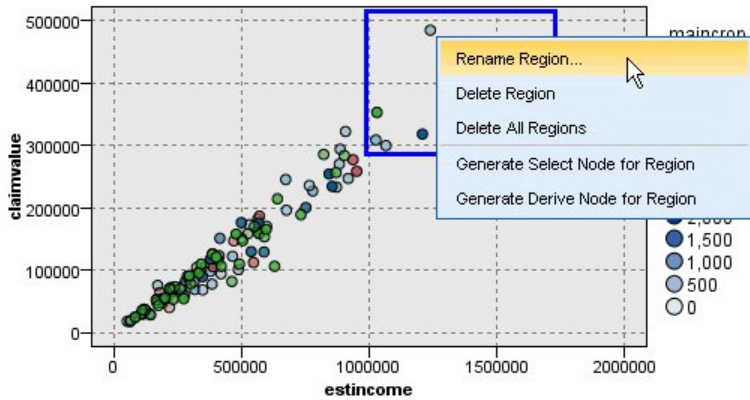


圖 66. 探索包含高索賠值的區域

**要定義區域：**

1. 驗證您是否處於瀏覽模式。從功能表中選擇檢視 > 瀏覽模式。
2. 在「探索模式」工具列中，按一下「繪製區域」按鈕。



圖 67. 「繪製區域」工具列按鈕

3. 在能夠接受區域的圖表中，按一下並拖拽滑鼠可繪製矩形區域。

**編輯、更名和刪除區域**

可以在「編輯圖表區域」對話框中或通過圖表自身的快速功能表編輯現有區域的內容。

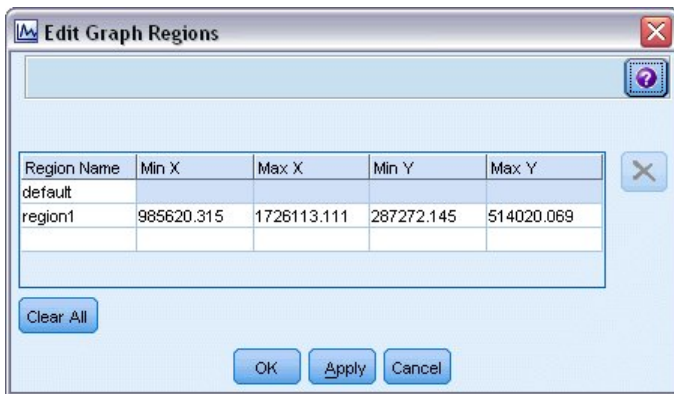


圖 68. 指定定義區域的內容

**要編輯區域：**

1. 驗證您是否處於瀏覽模式。從功能表中選擇檢視 > 瀏覽模式。
2. 在「探索模式」工具列中，按一下「繪製區域」按鈕。
3. 從功能表中選擇編輯 > 圖表區域。此時將開啟「編輯圖表區域」對話框。
4. 如果圖形中包含多個欄位（例如 SPLOM 圖形），必須在欄位 A 欄和欄位 B 欄中定義該區域的欄位。



5. 通過輸入名稱，選取欄位名稱（如果適用）並定義各個欄位的上限和下限，可以在新行上新增新區域。按 Enter 鍵啟動新列。
6. 通過調整 A 和 B 的下限和上限值編輯現有區域範圍。
7. 通過變更表格中區域的名稱，可以對區域更名。
8. 通過選取表格中的行並按一下「刪除」按鈕，可刪除區域。
9. 按一下**確定**以套用您的變更並關閉對話框。

註：或者，通過用滑鼠右鍵按一下區域所在的行，並從快速功能表中選擇所需選項，可直接刪除和更名圖表中的區域。

## 使用標示元素

您可以對任意圖表中的元素（例如長條圖、分區和點）進行標示。除時間繪圖、多統計圖和評估圖之外，不能在其他圖形中標示行、區域和表面，因為在這些圖形中行指代欄位。標示元素時，實質上是強調顯示該元素代表的所有資料。在同一觀察值出現在多個位置的任意圖表（例如 SPLOM）中，標示等同於塗抹。您可以對圖形中、甚至級區和區域中的元素進行標示。標示元素並恢復為編輯模式時，標示仍然可見。

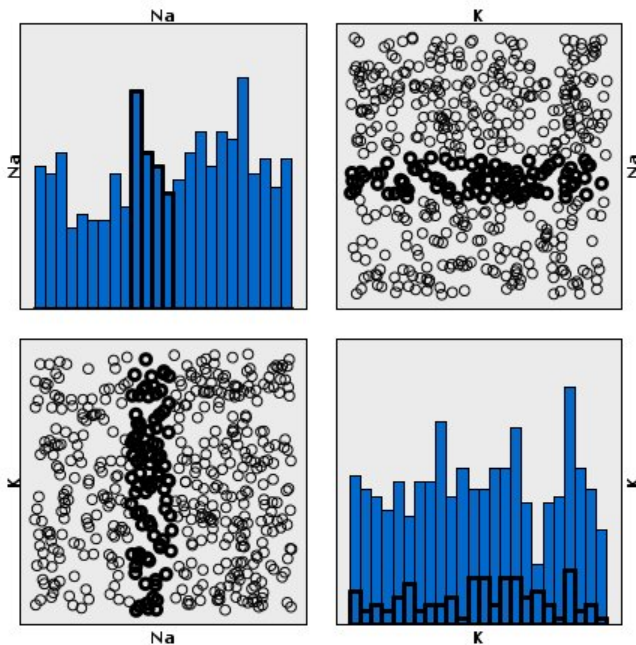


圖 69. 對 SPLOM 中的元素進行標示

通過按一下圖表中的元素，可以對元素進行標示和取消標示。首次按一下元素進行標示時，該元素周圍會出現顏色較深的邊框，表示已標示。如果再次按一下該元素，該邊框會消失，表示該元素不再處於已標示狀態。要對多個元素進行標示，可以按住 Ctrl 鍵並按一下元素，也可以使用「魔棒」，用滑鼠將想要標示的所有元素環繞起來。請記住，如果在沒有按住 Ctrl 鍵的情況下按一下其他區域或元素，會清除之前標示的所有元素。

可以從圖中標示的元素產生「選取」節點和「衍生」節點。請參閱第 246 頁的『從圖形產生節點』主題，以取得更多資訊。

### 要標示元素：

1. 驗證您是否處於瀏覽模式。從功能表中選擇檢視 > 瀏覽模式。
2. 在探索模式工具列上，按一下「標示元素」按鈕。

3. 按一下所需元素，或按一下並拖曳滑鼠，在包含多個元素的區域周圍繪製線條。

## 從圖形產生節點

IBM SPSS Modeler 圖形提供的最強大功能之一是可從圖形中產生節點或從圖形中所選內容產生節點。例如，在時間繪圖圖形中，可以根據選擇的資料或資料區域產生「衍生」節點和「選取」節點，有效地將資料劃分為多個「子集」。例如，您可使用這個強大的功能來識別以及排除離群值。

在繪製級區時，也可以產生「衍生」節點。在包含兩個尺度軸的圖形中，可以從圖形中繪製的區域產生「衍生」節點或「選取」節點。在包含標示元素的圖形中，可以從這些元素中產生「衍生」節點和「選取」節點，有些情況下可以產生「過濾器」節點。可以為顯示計數分佈的任意圖表啟用「平衡」節點產生。

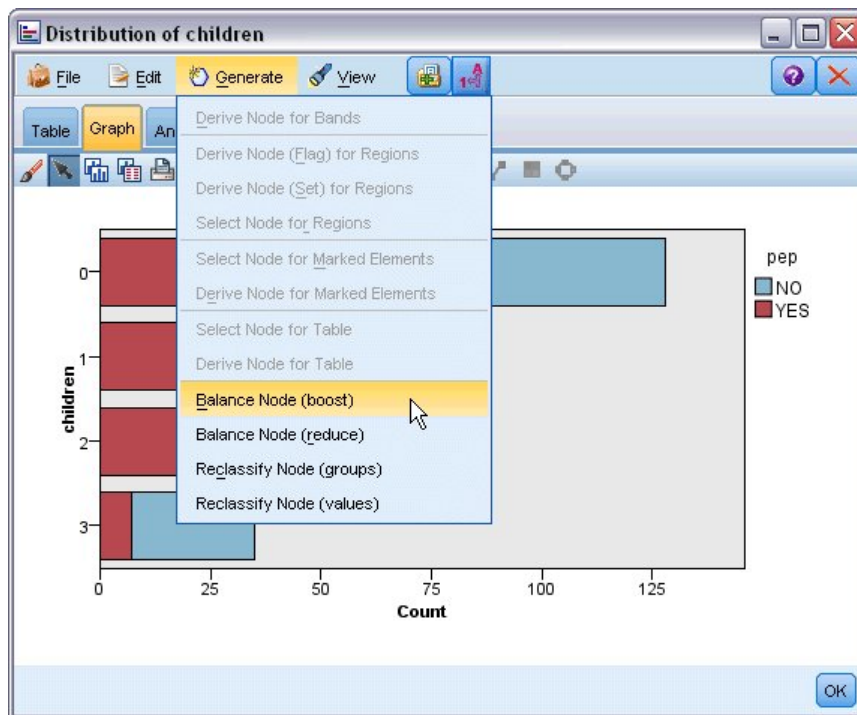


圖 70. 顯示產生功能表的圖形

產生節點時，該節點直接放置在串流畫布中，方便您將其連接到現有的串流。下列節點可從圖形中產生：選取、衍生、平衡、過濾和再分類。

### 「選取」節點

通過產生「選取」節點，可以針對下游正在處理測試記錄是否併入在某一區域中，以及是否排除了外該區域的所有記錄，或者測試與上述相反的情況。

- 針對條帶。可以產生用於包含或排除該級區中的記錄的「選取」節點。僅適用於級區的「選取」節點只能通過快速功能表存取，因為您需要選取要在「選取」節點中使用的級區。
- 針對區域。可以產生用於包含或排除該區域中的記錄的「選取」節點。
- 針對標示元素。您可以產生「選取」節點以擷取與標示元素或 Web 圖形鏈結相對應的記錄。

### 「衍生」節點

「衍生」節點可從區域、級區和標示元素中產生。所有圖形都可產生「衍生」節點。在評估表中，會出現用於選取模型的對話框。在 Web 圖形中，可以生成「衍生」節點（「與」）和「衍生」節點（「或」）。

- **針對條帶。**可以產生「衍生」節點，它通過將「編輯級區」對話框中所列的級區名稱用作種類名稱，為軸上標示的每個間隔產生種類。
- **針對區域。**可以產生「衍生」節點（**衍生為旗標**），用於建立一個名為 *in\_region* 的旗標欄位，其中設定為 *T* 的旗標表示任意區域中部的記錄，設定為 *F* 的旗標表示所有區域外部的記錄。您也可以產生「衍生」節點（**衍生為集**），用於為每個記錄都有對應的新欄位區域的每個區域產生具有值的集，每個記錄都以記錄所在區域的名稱作為自己的值。不在任何區域內的記錄將與預設區域同名。值名稱成為「編輯區域」對話框中所列的區域名稱。
- **針對標示元素。**可以產生「衍生」節點，用於計算對於所有旗標元素均為 *True* 且對於所有其他記錄均為 *False* 的旗標。

#### 「平衡」節點

通過產生「平衡」節點，可以更正資料中的不平衡情況，例如降低一般值的頻率（使用**平衡節點(減少)**功能表選項）或增加不一般值的出現次數（使用**平衡節點(增加)**功能表選項）。可以為顯示計數分佈的任意圖表啟用「平衡」節點產生，這些圖表包括直方圖、點圖形、收集圖形、個數的長條圖形圖形、個數的圓餅圖和多統計圖。

#### 「過濾器」節點

通過產生「過濾器」節點，可以根據圖表中標示的線條或節點對欄位進行更名或過濾。在評估表中，最佳擬合行不會產生過「濾節」點。

#### 「再分類」節點

通過產生「再分類」節點，可以對值進行重新編碼。此選項適用於分佈圖。可以為**群組**產生「再分類」節點，以便根據所顯示欄位的特定值是否併入在某個群組（在**表格標籤**上，按住 **Ctrl** 鍵並同時按一下來選取群組）中，對這些值進行重新編碼。也可以為**值**產生「再分類」節點，以將資料重新編碼到包含多個值的現有集，例如將資料再分類到標準值集，以合併多個公司的財務資料進行分析。

註：如果已經預先定義了值，那麼可以將它們作為純文字檔案讀取到 IBM SPSS Modeler 中，並使用分佈圖顯示所有值。然後直接從圖表中為該欄位產生「再分類（值）」節點。這樣，所有目標值都會出現在「再分類」節點的新值欄（下拉清單）中。

設定「再分類」節點的選項時，表格將啟用從舊集值到指定的新值的明確對映：

- **原始值。**此直欄列出所選欄位的現有值。
- **新建值。**使用此直欄來輸入新的種類值或從下拉清單選取值。當您使用「分佈」圖表中的值自動產生「重新分類」節點時，這些值會併入下拉清單中。這可讓您快速將現有值對映至已知的值集。例如，保健組織有時候會根據網路或語言環境以不同方式對診斷進行分組。在合併或收購之後，所有相關方都需要以一致的方式來重新分類新的甚至是現有的資料。無需手動輸入長清單中的每個目標值，您可以將主要值清單讀取至 IBM SPSS Modeler，執行診斷欄位的「分佈」圖表，並直接從圖表為這個欄位產生「重新分類（值）」節點。此程序可讓所有目標「診斷」值都能在「新建值」下拉清單中可用。

有關「再分類」節點的進一步資訊，請參閱第 139 頁的『為再分類節點設定選項』。

#### 從圖形產生節點

可以使用圖表輸出視窗中的「產生」功能表產生節點。產生的節點將被放置在串流畫布中。要使用此節點，請將其連接到現有串流中。

要從圖表產生節點：

1. 驗證您是否處於瀏覽模式。從功能表中選擇**檢視 > 瀏覽模式**。
2. 在探索模式工具列上，按一下「區域」按鈕。
3. 定義產生節點所需的級區、區域或任何標示元素。
4. 從「產生」功能表中選擇要產生的節點類型。僅啟用允許的類型。

註：或者，也可以通過用滑鼠右鍵按一下並從快速功能表中選擇所需產生選項來直接從圖表中產生節點。

---

## 編輯視覺化

使用「探索」模式可讓您以分析方式探索視覺化所表示的資料和值，「編輯」模式則可讓您變更視覺化的佈置及外觀。例如，您可以變更字型和顏色來搭配您組織樣式使用手冊的外觀。若要選取這個模式，請在功能表中選擇「**檢視 > 編輯模式**」（或按一下工具列圖示）。

在「編輯」模式中，有數個影響視覺化不同佈置方式的工具列。如果其中有您用不到的工具，您可以隱藏這些工具列以增加對話框中的空間，如此一來才能顯示圖表。若要選取或取消選取工具列，請在「檢視」功能表中按一下相關的工具列名稱。

附註：若要在視覺化中新增更多明細，您可以套用標題、註腳及軸標籤。請參閱第 258 頁的『新增標題和註腳』主題，以取得更多資訊。

在「編輯模式」中，您有數種選擇可用來編輯視覺化。您可以：

- 編輯文字以及其格式。
- 變更填滿顏色、透明度、邊框樣式和圖形元素。
- 變更框線及線條的顏色和粗細。
- 旋轉並變更點元素的形狀和圖面比。
- 變更圖形元素大小 (例如長條圖和點)。
- 使用邊緣和填塞調整項目之間的空間。
- 指定數字格式。
- 變更軸及縮放比例設定。
- 在類別軸中分類、排除及收合類別。
- 設定窗格方向。
- 將轉換套用至座標系統。
- 變更統計、圖形元素類型和衝突修飾元。
- 變更圖註位置。
- 套用視覺化樣式表。

下列標題將說明如何執行這些工作。建議您詳讀有關編輯圖形工作的一般法則。

如何切換至編輯模式

在功能表上，選擇：

**檢視 > 編輯模式**

## 編輯視覺化的一般相關規則

### 編輯模式(E)

在編輯模式中完成所有編輯。若要啟用「編輯」模式，從功能表選擇：

#### 檢視 > 編輯模式

### 選擇(S)

依選擇不同，可用的編輯選項也不同。依選擇的項目而定，可啟用不同的工具列和性質色板選項。只有啟用中項目才會套用至目前的選擇。例如，如果已經選取了軸，您就可以使用性質色板上的「尺度」、「主刻度」、「次刻度」標籤。

以下列出幾個提示，讓您為視覺化選擇項目時更得心應手：

- 按一下項目以表示選取。
- 按一下以選取一個圖形元素 (如散佈圖中的點或長條圖中的長條)。初步選擇圖形後，再按一下以將您選項的項目縮小為圖形元素群組或單一圖形元素。
- 按一下 ESC 以取消選取所有選項。

### 調色板

在視覺化中選取項目時會同時更新多個色板，以反映您所進行的選擇。色板中包含可讓您在選擇項目中編輯的控制項。色板可能是工具列，或是包含多種控制項和標籤的窗格。色板不一定會顯示，因此請確認已顯示色板以進行編輯。核取「檢視」功能表以使用目前已顯示的色板。

您可以按一下並拖曳工作列色板或其他色板左方的空白區域，以重新放置色板。您可透過視覺回饋得知您能停駐色板的位置。如果是非工具列色板，您也可以按一下「關閉」按鈕以隱藏色板和解除停駐按鈕，以在獨立視窗中顯示色板。按一下「說明」按鈕以顯示特定色板的說明。

### 自動設定

部分設定提供「-auto-」選項。這表示已套用了自動值。依特定視覺化和資料值而定，使用不同自動設定。您可以輸入數值以改寫自動設定。如果您想要還原自動設定，請刪除目前的數值，然後按一下 Enter。設定會再次顯示為「-auto-」。

### 移除/隱藏項目

您可以在視覺化中移除/隱藏不同項目。例如，您可以隱藏圖註或軸標記。若要刪除項目，請選取此項目並按下「刪除」。如果項目不允許進行刪除，則不會出現任何變更。如果您不小心刪除了項目，按下 Ctrl+Z 以復原此刪除項目。

### 狀態

某些工具列可反映目前選定內容的狀態，其他工具列則不會反映。內容選用區始終反映此狀態。如果工具列「並未」顯示狀態，則描述此工具列的標題會說明其狀態。

## 編輯和設定文字格式

您可以在此編輯文字，並變更整個文字區塊的格式。請注意，您無法編輯直接鏈結至資料值的文字。例如您無法編輯刻度標記，因為標記的內容來自基礎資料。但是，您可以在視覺化中設定任何文字的格式。

### 在位置中編輯文字

1. 連按兩下文字區塊。這個動作會選取所有文字。此時會停用所有的工作列，編輯文字時，您無法變更視覺化的任何部分。
2. 輸入內容以取代現有文字。您也可以再按一下文字以顯示滑鼠游標。將滑鼠游標置於您需要的位置，並輸入其他文字。

如何調整文字格式

1. 選取包含文字的框。請勿在文字上連按兩下。
2. 使用字型工具列設定文字格式。如果尚未啟用工具列，請確定您只選取了包含文字的「框」。如果已選取文字，工具列就會停用。

您可以變更字型：

- 色彩
- 字型系列 (例如 Arial 或 Verdana)
- 大小 (除非您選擇使用如 pc 等不同單位，否則預設單位為 pt)
- 粗細
- 對文字框進行對齊

格式會套用至框內的所有文字。您無法變更任何文字區塊內單一字母或文字的格式。

## 變更顏色、樣式、線條粗細和透明度

視覺化中的許多項目都已填滿並具有框線。最顯而易見的例子是長條圖內的長條。長條內的顏色是填滿的色彩，並且以實心黑色邊框包圍。

視覺化中部分已填滿色彩的項目可能沒有這麼顯目。如果使用的填滿色彩是透明色，您可能就看不出來。例如軸標記中的文字。文字看起來像是浮在平面上，但事實上它出現在使用透明填滿色彩效果的框中。您可以選取軸標記以顯示框。

視覺化中的任何框都可能有填滿色彩效果和邊框樣式，視覺化的邊框也不例外。除此之外，您也可以調整任何填滿色彩效果的不透明度/透明度。

如何變更顏色、樣式、線條粗細和透明度

1. 選取您想調整格式的項目。例如，選取長條圖中的長條，或是包含文字的框。如果視覺化由類別變數或欄位分割，您也可以選取對應至單一類別的群組。這麼一來，您就可以變更此群組的預設審美標準。例如，您可以變更堆疊長條圖中其中一個堆疊群組的顏色。
2. 若要變更填滿顏色、邊框顏色或填滿樣式，請使用顏色工具列。

附註：此工具列無法反映現行選擇項目的狀態。

若要變更顏色或填滿方式，您可以按一下按鈕來選取已顯示的選項，或是按一下下拉式箭頭以選擇另一個選項。在顏色部分，您會看見一個白底、對角線為紅色的顏色選項。這就是透明色。您可以使用這個選項，在直方圖中隱藏長條邊框。

- 第一個按鈕控制填滿顏色效果。如果顏色與連續或序數欄位相關，這個按鈕可變更資料中與最高數值相關顏色的填滿顏色。您可使用性質色板中的「顏色」標籤來變更與最低數值和遺漏資料相關的顏色。元素顏色將隨著基本資料的增加，由「低」顏色增量變更至「高」顏色。
- 第二個按鈕控制邊框顏色。
- 第三個按鈕控制填滿樣式。填滿樣式會使用邊框的顏色。因此，只有在使用不透明的邊框顏色時，才看得見填滿樣式。

- 第四個控制項是滑塊和文字框，可控制填滿顏色和型樣的不透明度。百分比低時，表示透明度較低，且不透明度較高。100% 表示完全不透明 (無透明度)。
3. 若要變更邊框或線條的粗細，請使用線條工具列。

附註：此工具列無法反映現行選擇項目的狀態。

和其他工具列一樣，您可以按一下按鈕來選取已顯示的選項，或是按一下下拉式箭頭以選擇另一個選項。

## 旋轉並變更點元素的形狀和圖面比

您可以旋轉點元素、指派不同預先定義的形狀，或變更圖面比 (圖的長寬比)。

### 如何修改點元素

1. 選取點元素。您無法旋轉或變更單一點元素的形狀和圖面比。
2. 請使用符號工具列為點進行修改。
  - 第一個按鈕可讓您變更點的形狀。按一下下拉式箭頭並選取預先定義的形狀。
  - 第二個按鈕可讓您將點旋轉至特定周圍方位。按一下下拉式箭頭，然後將 needle 拖曳至您所需的位置。
  - 第三個按鈕可讓您變更圖面比。按一下下拉式箭頭，然後在顯示的長方形上按一下並進行拖曳。長方形的形狀表示圖面比。

## 變更圖形元素的大小

您可以變更視覺化中圖形元素的大小。例如長條、線條和點等等。如果圖形元素的大小是由變數或欄位所設定，則指定尺寸為「最小」大小。

### 如何變更圖形元素的大小

1. 選取您要改變大小的圖形元素。
2. 在符號工具列上使用滑塊或輸入特定大小，選擇可用的選項。除非您另行設定不同單位，否則預設單位是像素 (請參閱下列完整可用單位縮寫清單)。您也可以指定百分比 (如 30%)，這表示圖形元素可使用可用空間中特定百分比大小的空間。可用空間依圖形元素類型和指定視覺化而定有所不同。

表 39. 有效的單位縮寫

縮寫	單位(N)
公分	公分
吋	英吋
mm	公釐
pc	1/6 英吋
pt	點
px	像素

## 指定邊緣和填塞大小

如果框的邊緣或內部空間太大或太小，您可以變更其邊緣和填塞設定。「邊緣」是框和周圍其他項目間的空間大小。「填塞」則是邊框和框內「內容」之間的空間大小。

### 如何指定邊緣和填塞大小

1. 選取您想要指定邊緣和填塞大小的框。它可以是文字框、圖註邊框或是顯示圖形元素的資料框 (例如長條和點)。

2. 使用「性質色板」上的「邊緣」標籤來指定設定。如果您沒有指定其他單位 (如公分或英吋)，則所有的尺寸都以像素為單位。

## 設定數字格式

您可以在連續軸或顯示數字的資料值標籤上，指定刻度標記中的數字格式。例如，您可以指定以千為單位顯示刻度標記中的數字。

如何指定數字格式

1. 選取其中含有數字的連續軸刻度標記或資料值標籤。
2. 按一下「性質色板」中的「格式」標籤。
3. 選取您想要使用的數字格式選項：

**字首。**顯示於數字最前方的字元。例如，如果數字是以美元為單位的薪資，則輸入貨幣符號 (\$)。

**字尾。**顯示於數字最後方的字元。例如，如果數字是百分比，則輸入百分比符號 (%)。

**最小整數位數。**在包含小數點的數字中，整數部分應顯示的最小數字位數。如果實際的值不包含最小所需位數，則會用零來填塞數值的整數部分。

**最大整數位數。**在包含小數點的數字中，整數部分應顯示的最大數字位數。如果實際的數值超出位數上限，則將會以星號取代值的整數部分。

**最小小數位數。**小數點後的數字或以科學符號表示的數字中，應顯示的最小數字位數。如果實際的值不包含最小所需位數，則會用零來填塞值的小數部分。

**最大小數位數。**小數點後的數字或以科學符號表示的數字中，應顯示的最大數字位數。如果實際的值超出位數上限，小數會四捨五入為適當的位數。

**科學符號。**以科學符號顯示數字的時機。科學符號在數量很大或很小時非常有用。**-auto-** 可讓應用程式決定使用科學符號的適當時機。

**尺度設定。**尺度因素，原始數值會與此數字相除。當數字大時使用尺度因素，但是標記最好不要過長，以免無法納入數字。如果變更刻度標記的數字格式，一定要編輯軸標題，指示如何解譯數字。例如，假設您的尺度軸顯示的是薪水，而標記是 30,000、50,000 和 70,000，您可能會輸入尺度因素 1000，以顯示 30、50 和 70。此時就要編輯尺度軸標題，以千為單位將文字包含在內。

**以括號標示負數。**在負數值兩側顯示括號的時機。

**分組。**在位數分組之間顯示字元的時機。電腦目前的國別或語言環境會決定使用何種字元進行位數分組。

## 變更軸及尺度設定

用來修改軸和尺度的數個選項。

如何變更軸及尺度設定

1. 選取軸的任何部分 (例如軸標記或刻度標記)。
2. 使用「性質色板」上的「尺度」、「主刻度」和「副刻度」標籤來變更軸及尺度設定。

「尺度」標籤

附註：對於預先聚集資料的圖形 (例如直方圖)，「尺度」標籤不會出現。



**類型。** 指定為線性尺度或轉換尺度。尺度轉換可協助您了解資料或針對統計推論進行所需假設。在散佈圖上，如果自變數與因變數或欄位之間的關係是非線性的，那麼您就可以使用轉換尺度。尺度轉換也可用於使偏斜直方圖更為對稱，使它看起來更像常態分配。請注意，您只轉換顯示資料的尺度，而不是轉換實際的資料。

- **線性。** 指定線性、不轉換的尺度。
- **對數。** 指定以 10 為基底對數轉換而成的尺度。若要納入零與負值，轉換會使用修改過的對數函數版本。這個「安全對數」函數定義為  $\text{sign}(x) * \log(1 + \text{abs}(x))$ 。因此， $\text{safeLog}(-99)$  會等於：

$$\text{sign}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$$

- **次方。** 使用數值為 0.5 的指數指定幕次轉換的尺度。若要納入負值，轉換會使用修改過的幕次函數版本。這個「安全幕次」函數的定義為  $\text{sign}(x) * \text{pow}(\text{abs}(x), 0.5)$ 。因此， $\text{safePower}(-100)$  會等於：

$$\text{sign}(-100) * \text{pow}(\text{abs}(-100), 0.5) = -1 * \text{pow}(100, 0.5) = -1 * 10 = -10$$

**最小/最大/最適低值/最適高值。** 指定尺度範圍。選取「最適低值」和「最適高值」可讓應用程式依資料選取適當的尺度。最小和最大值是最適當的，因為這類的值通常是大於或小於最大及最小資料值的完整值。例如，如果資料範圍為 4 到 92，尺度的最適低值和最適高值應為 0 和 100，而不是實際的最大和最小值。請避免設定太小的範圍，範圍太小就無法顯示重要的項目。另外，如果您選取了「包含零」選項，您就無法明確設定最小和最大值。

**低邊緣/高邊緣。** 以軸的較低和/或較高點建立邊緣。邊緣相對於所選取的軸會以垂直方式顯示。在沒有指定其他單位 (如公分或英吋) 的情況下，預設使用的單位為像素。例如，如果您將垂直軸的「高邊緣」設為「5」，則 5 個像素的水平邊緣會和資料框的頂端平行。

**反轉。** 指定是否反轉尺度。

**包含零。** 指示尺度應包含 0。此選項一般用於長條圖以確保長條圖從 0 處開始，而不是從接近最小長條圖高度的值開始。「最小」和「最大」會在您選取這個選項後停用，因為您無法自訂尺度範圍的最大和最小值。

**主刻度/副刻度標籤**

「刻度」或「刻度標記」是顯示在軸上的線。這些線條表示特定區間或類別的值。「主刻度」是包含標記的刻度標記。這些標記比其他刻度標記來得長。「副刻度」是顯示在主刻度標記之間的刻度標記。某些選項只能在特定刻度類型中使用，但大多數的選項都可用在主刻度和副刻度中。

**顯示刻度。** 指定是否在圖形上顯示主刻度或副刻度。

**顯示網格線。** 指定主刻度或副刻度上是否顯示網格線。「網格線」是圖形中穿越軸與軸之間的線條。

**位置。** 指定相對於軸的刻度標記之位置。

**長度。** 指定刻度標記的長度。在沒有指定其他單位 (如公分或英吋) 的情況下，預設使用的單位為像素。

**基點。** 只適用於主要刻度。指定第一個主刻度出現位置上的數值。

**Delta。** 只適用於主要刻度。指定主刻度之間的不同處。也就是說，主刻度會在出現於第「n」個值的位置，這個位置的「n」則是 Delta 值。

**分割區域。** 只適用於次要刻度。指定副刻度的分割區域在主刻度間的數目。副刻度的數目會比分割區域的數目少 1。例如，假設主要刻度介於 0 與 100 之間。如果您輸入 2 作為次要刻度分割區域的數字，則在 50 的位置會出現 1 個次要刻度、分割 0-100 的範圍，並建立 2 個分割區域。

## 編輯類別

您可以以數種方式，在類別軸上編輯類別：

- 變更顯示類別的排序順序。
- 排除特定類別。
- 新增未顯示在資料集中的類別。
- 將小型類別收合/合併為一個類別。

如何變更類別的排序順序

1. 選取類別軸。「類別」色板會在軸上顯示類別。

附註：如果未顯示選用區，則請先確定您已經啟用選用區。在 IBM SPSS Modeler 的「檢視」功能表中，選擇**種類**。

2. 在「類別」色板中，在下拉清單中選取排序選項。

**自訂**。依類別在色板上出現的順序進行排序。使用箭頭按鈕將類別上下移動，您也可以將其移動至清單的頂端或底端。

**資料**。依類別在資料集上發生的順序進行排序。

**名稱**。依類別在色板中顯示的名稱，以字母順序來排序類別。視您選取的工具列按鈕是顯示值或標記，其顯示名稱可能是值或是標記。

**值**。依色板中顯示的值（加括弧顯示）而定，以基礎資料值為基準來排序類別。只有包含 meta 資料的資料來源（如 IBM SPSS Statistics 資料檔）才支援此選項。

**統計量**。依計算統計量而定，為所有類別排序類別。統計量範例包括：個數、百分比和平均數。只有在圖形中使用統計量時，您才能使用此選項。

如何新增類別

依照預設，您只能使用顯示在資料集中的類別。您可以依需求將類別新增至視覺化。

1. 選取類別軸。「類別」色板會在軸上顯示類別。

附註：如果未顯示選用區，則請先確定您已經啟用選用區。在 IBM SPSS Modeler 的「檢視」功能表中，選擇**種類**。

2. 在「類別」色板中，按一下「新增類別」按鈕：



圖 71. 「新增類別」按鈕

3. 在「新增新類別」對話框中，輸入類別的名稱。
4. 按一下**確定**。

如何排除特定類別

1. 選取類別軸。「類別」色板會在軸上顯示類別。

附註：如果未顯示選用區，則請先確定您已經啟用選用區。在 IBM SPSS Modeler 的「檢視」功能表中，選擇種類。

2. 在「類別」色板中，在「包含」清單中選取類別名稱，然後按一下 X 按鈕。若要移回類別，請在「排除」清單中選取該類別名稱，然後按一下清單右邊的箭頭。

### 如何收合/合併小型類別

您可以合併不需分別顯示的小型類別。例如，如果您有一個包含許多類別的圓餅圖，您可以考慮收合百分比小於 10 的類別。收合僅適用於加總的統計資料。例如，您無法一起加入平均數，因為平均數不是具有可加性的項目。因此，無法使用平均數合併/收合類別。

1. 選取類別軸。「類別」色板會在軸上顯示類別。

附註：如果未顯示選用區，則請先確定您已經啟用選用區。在 IBM SPSS Modeler 的「檢視」功能表中，選擇種類。

2. 在「類別」色板中，選取「收合」，並指定百分比。總百分比小於指定數字的所有類別就會合併為一個類別。百分比是根據圖表中所顯示的統計量計算。收合功能只可用於以個數為基礎的統計量，或是總和統計量。

## 變更窗格方向

如果您正在視覺化中使用窗格，您可以變更窗格的方向。

### 如何變更窗格方向

1. 選取視覺化的任何部分。
2. 按一下「性質色板」中的「窗格」標籤。
3. 在「佈置」中選擇選項：

**表格。** 如果您以表格方式配置窗格，面板中的值會各自擁有各別的列或欄位。

**轉置。** 以表格方式配置窗格，但交換原本的列和欄。這個選項和轉置圖形的功能不同。請注意，在選取這個選項後， $x$  軸和  $y$  軸並不會出現改變。

**清單。** 以清單方式配置窗格，每個儲存格都代表一個合併值。如此並不會將直欄和列指派給個別數值。這個選項允許窗格依需求進行換行。

## 轉換座標系統

許多視覺化都顯示在平面的直角坐標系統中。您可以依需求轉換座標系統。例如，您可以將極轉換套用至座標系統、新增斜交陰影效果並轉置軸。您也可以移除任何已經套用至現有視覺化的轉換。例如，您在極座標系統中繪製了一張圖餅圖。您可以復原此極線轉換，並在直角坐標系統中以單一堆疊長條的方式顯示圖餅圖。

### 如何轉換座標系統

1. 選取您想要轉換的座標系統。您可以以選取圖形的邊框的方式來選擇座標系統。
2. 按一下「性質色板」中的「座標」標籤。
3. 選取您要套用到座標系統上的轉換。您也可以取消選取此轉換以進行取消。

**轉置。** 變更軸方向的動作稱為「轉置」。這個動作和在 2D 視覺化中交換垂直及平行軸的動作類似。

**極。** 極轉換可以從圖形中點啟動，以特定角度及距離繪製圖形元素。圓餅圖是 1-D 視覺化，其中包含極線轉換，可使用特定角度繪製個別長條。雷達圖是 2-D 視覺化，其中包含極線轉換，可以從圖形中點啟動，以特定角度及距離繪製圖形元素。3-D 視覺化也可以包含維度深度。

**斜交。**斜交轉換可在圖形元素中新增 3-D 效果。這種轉換可在圖形元素中加入深度，但新增的深度只有裝飾效果。和特定資料值無關。

**相同比例。**套用相同比例可讓您指定，在資料值中，所有尺度中的相同距離都代表相同差異。例如，在雙方尺度上標示 2 公分的距離，就表示差異為 1000。

**轉換前插頁 %。**如果轉換完成後軸被截斷，您可以在套用轉換前，在圖形中加入插頁。插頁會將維度縮小至某個百分比，然後才在座標系統中套用轉換。您可以依此順序，控制下半部的  $x$  維度、上半部的  $x$  維度、下半部的  $y$  維度及上半部的  $y$  維度。

**轉換後插頁 %。**如果您想要變更圖形的長寬比，您可以在套用轉換後，在圖形中新增插入項。座標系統會先套用轉換，接著插頁會將維度縮小至某個百分比。就算不打算在圖形中套用任何轉換，您仍然可以套用插頁。您可以依此順序，控制下半部的  $x$  維度、上半部的  $x$  維度、下半部的  $y$  維度及上半部的  $y$  維度。

## 變更統計量和圖形元素

您可以將圖形元素轉換為另一種類型，變用來繪製圖形元素的統計量，或指定衝突修飾元以決定圖形元素重疊時的結果。

### 如何轉換圖形元素

1. 選取您想要轉換的圖形元素。
2. 按一下「性質色板」中的「元素」標籤。
3. 在「類型」清單中選取新圖形元素。

表 40. 圖形元素類型

圖形元素類型	說明
點	識別特定資料點的標記。點元素用於散佈圖和其他相關視覺化中。
區間(N)	繪製於特定資料值，並填滿原點和另一資料值之間的空間的矩形。區間元素用於長條圖和直方圖中。
行	連接資料值的線。
路徑	連接資料值的線 (依照資料值顯示於資料集中的順序)。
區域圖(A)	連接資料元素的線，線和原點之間的區域會填滿。
多邊形	括住資料區域的多邊形。多邊形元素可用於 Bin 散佈圖或地圖。
架構	由方框組成的元素，以鬚帽和標記指示偏離值。架構元素用於盒形圖。

### 如何變更統計量

1. 選取圖型元素以變更其統計量。
2. 按一下「性質色板」中的「元素」標籤。

### 如何指定衝突修飾元

衝突修飾元可決定圖形元素重疊時的結果。

1. 請選取您要用衝突修飾元指定的圖形元素。
2. 按一下「性質色板」中的「元素」標籤。
3. 在「修飾器」下拉清單中，選取一個衝突修飾元。**-auto-** 可讓應用程式為圖形元素及統計量決定合適的衝突修飾元。

**重疊。**當圖形元素具備相同的值時，在圖形元素上繪製其他圖形元素並一個個堆疊上去。

**堆疊。**通常，在圖形元素包含相同的資料值時會串聯，此時可使用堆疊。

**迴避。**將包含相同值的圖形元素以相鄰方式顯示，而不是串聯。圖形元素會以對稱方式排列。也就是說，圖形元素會移動至中央位置的兩側。迴避和叢集十分相似。

**堆放。**將包含相同值的圖形元素以相鄰方式顯示，而不是串聯。圖形元素會以不對稱方式排列。也就是說，圖形元素會一個接著一個堆疊而上，最底端的圖形元素則是尺度的指定值。

**波動 (常態)。**以常態分配方式，在同一個資料值中隨機重新放置圖形元素。

**波動 (均勻)。**以均勻分配方式，在同一個資料值中隨機重新放置圖形元素。

## 變更圖註位置

如果圖形中包含圖註，則圖註一般會顯示於圖形右邊。您可以視需要改變位置。

如何變更圖註位置

1. 選取圖註。
2. 按一下「性質色板」中的「圖註」標籤。
3. 選取位置。

## 複製視覺化及視覺化資料

「一般」色板中包含可複製視覺化及其資料的按鈕。



圖 72. 「複製視覺化」按鈕

**複製視覺化。**這個動作會將視覺化以影像方式複製到剪貼簿。有多種影像格式可使用。當您將影像貼上至其他應用程式時，您可以選取「選擇性貼上」選項以選擇一個可用的影像格式然後貼上。



圖 73. 「複製視覺化資料」按鈕

**複製視覺化資料。**這個動作可複製用來繪製視覺化的基礎資料。資料會以純文字或 HTML 格式文字的方式複製到剪貼簿。當您將資料貼上至其他應用程式時，您可以選擇「選擇性貼上」選項以選擇一個可用的格式然後貼上。

## 圖表板編輯器鍵盤快速鍵

表 41. 鍵盤捷徑

捷徑按鍵	函數
Alt+ /	在瀏覽與編輯模式之間切換
刪除	刪除視覺化項目
Ctrl+Z	復原

表 41. 鍵盤捷徑 (繼續)

捷徑按鍵	函數
Ctrl+Y	重做
F2	顯示概要以在圖形中選取項目。

## 新增標題和註腳

對於所有圖形類型，您都可以新增唯一的標題、註腳或軸標籤，以協助識別圖形中顯示的內容。

### 新增標題至圖形

1. 從功能表中選擇**編輯 > 新增圖形標題**。圖形上方會顯示含有 **<TITLE>** 的文字框。
2. 驗證您處於「編輯」模式。從功能表中選擇**視圖 > 編輯模式**。
3. 按兩下 **<TITLE>** 文字。
4. 鍵入所需的標題並按 Return。

### 新增註腳至圖形

1. 從功能表中選擇**編輯 > 新增圖形註腳**。圖形下方會顯示含有 **<FOOTNOTE>** 的文字框。
2. 驗證您處於「編輯」模式。從功能表中選擇**視圖 > 編輯模式**。
3. 按兩下 **<FOOTNOTE>** 文字。
4. 鍵入所需的標題並按 Return。

## 使用圖形樣式表

基本圖形顯示資訊（例如顏色、字型、符號及線條粗細）均由樣式表控制。IBM SPSS Modeler 隨附了一個預設樣式表；但是您可以根據需要對其進行變更。例如，您可以為要在圖形中使用的簡報設定公司色系。請參閱第 248 頁的『編輯視覺化』主題，以取得更多資訊。

在圖形節點中，您可以使用「編輯」模式對圖形外觀進行樣式變更。然後，您可以使用**編輯 > 樣式**功能表將這些變更儲存為樣式表，以套用至隨後從現行圖形節點產生的所有圖形，或者儲存為新的預設樣式表，用於使用 IBM SPSS Modeler 產生的所有圖形。

「編輯」功能表上的樣式選項提供五個樣式表選項：

- **切換樣式表**。此選項顯示已儲存的<sup>不同</sup>樣式表清單，您可以從中選取以變更圖形外觀。請參閱『套用樣式表』主題，以取得更多資訊。
- **將樣式儲存至節點**。此選項會儲存對選定圖形的樣式所進行的修改，以便將這些修改套用至所有將來從現行串流中相同圖形節點建立的圖形。
- **將樣式儲存為預設值**。此選項會儲存對選定圖形的樣式所進行的修改，以便將這些修改套用至所有將來從任何串流中任何圖形節點建立的圖形。選取此選項之後，您可以使用**套用預設樣式**將任何其他現有圖形變更為使用相同樣式。
- **套用預設樣式**。此選項可將選定圖形的樣式變更為目前儲存為預設樣式的樣式。
- **套用原始樣式**。此選項可將圖形的樣式變更回提供作為原始預設值的樣式。

## 套用樣式表

您可以套用視覺化樣式表，以指定視覺化的文體內容。例如，樣式表可以定義字型、破折號及色彩等其他選項。在一定程度上，樣式表提供您手動執行編輯的捷徑。但是請注意，樣式表僅限於樣式變更。其他變更，如圖註位置或尺度範圍不會儲存在樣式表中。

如何套用樣式表

1. 在功能表上，選擇：

編輯 > 樣式 > 切換樣式表

2. 使用「切換樣式表」對話框選取樣式表。

3. 按一下「套用」將樣式表套用至視覺化，而不關閉對話框。按一下「確定」套用樣式表，並關閉對話框。

「切換/選取樣式表」對話框

對話框上方的表格列出目前可用的所有視覺化樣式表。有些樣式表是預先安裝的，其他則是在 IBM SPSS Visualization Designer (不同的產品) 中所建立的。

對話框下方會顯示樣本資料的視覺化範例。請選取樣式表之一，將其樣式套用到視覺化範例。這些範例可協助您判斷樣式表對您實際上視覺化的影響。

對話框同時也提供下列選項。

**現有樣式。**依照預設，樣式表可改寫視覺化中的所有樣式。您可以變更此行為。

- **改寫所有樣式。**套用樣式表時，改寫視覺化中的所有樣式，包括目前編輯階段作業在視覺化中所修改過的樣式。
- **保留修改過的樣式。**套用樣式表時，僅改寫非目前編輯作業階段在視覺化中所修改的樣式。則目前編輯作業階段所修改的樣式會被保留起來。

**管理。**管理您電腦中的視覺化範本、樣式表和地圖。您可以匯入、匯出、重新命名或刪除您本機電腦中的視覺化範本、樣式表與地圖。請參閱第 188 頁的『管理範本、樣式表與地圖檔案』主題，以取得更多資訊。

**位置。**變更儲存視覺化範本、樣式表與地圖的位置。目前位置列於按鈕右邊。請參閱第 187 頁的『設定範本、樣式表與地圖的位置』主題，以取得更多資訊。

## 列印、儲存、複製及匯出圖形

每個圖形都具有多個選項，可讓您儲存或列印圖形，或者將圖形匯出為另一種格式。這些選項大多都可從「檔案」功能表取得。此外，從「編輯」功能表中，您可以選擇複製圖形、其中資料或 Microsoft Office 繪圖物件，以在其他應用程式中使用。

### 正在列印

若要列印圖形，請使用列印功能表項目或按鈕。在列印之前，您可以使用頁面設定及預覽列印來設定列印選項及預覽輸出。

### 儲存圖形

若要將圖形儲存到 IBM SPSS Modeler 輸出檔 (\*.cou)，請從功能表選擇檔案 > 儲存或檔案 > 另存新檔。

或

若要將圖形儲存在儲存庫中，請從功能表選擇檔案 > 儲存輸出。

### 複製圖形

若要複製圖形以用於其他應用程式（例如 MS Word 或 MS PowerPoint），請從功能表選擇編輯 > 複製圖形。

## 複製資料

若要複製資料以用於其他應用程式（例如 MS Excel 或 MS Word），請從功能表選擇編輯 > 複製資料。依預設，該資料會被格式化為 HTML。在其他應用程式中使用選擇性貼上，可在進行貼上作業時檢視其他格式選項。

## 複製 Microsoft Office 圖形物件

您可以將圖形複製為 Microsoft Office 圖形物件，以在 Microsoft Office 應用程式中使用，例如 Excel 或 PowerPoint。若要複製圖形，請從功能表中選擇編輯 > 複製 Microsoft Office 圖形物件。內容將複製到您的剪貼簿，依預設將為二進位格式。在 Microsoft Office 應用程式中使用選擇性貼上，可在進行貼上作業時指定其他格式選項。

請注意，部分內容可能不支援此功能，在這種情況下會停用複製 Microsoft Office 圖形物件功能表選項。此外亦請注意，在貼入 Office 應用程式之後，圖形外觀可能會有所不同，但圖形資料相同。

有六種類型的圖形輸出可複製並貼入 Excel：簡單長條圖、堆疊長條圖、簡單盒形圖、叢集盒形圖、簡單散佈圖及分組散佈圖。如果對這其中任何圖形類型使用「畫面」及「動畫」選項，則 SPSS Modeler 中將顯示複製 Microsoft Office 圖形物件選項。而對於其他設定（例如「選用審美標準」或「重疊」），該選項可局部支援。如需詳細資料，請參閱下表：

表 42. 複製 Microsoft 圖形物件支援

圖形輸出範本	模型器圖形節點	模型器圖形類型	基本設定	選用審美標準	重疊	Microsoft 圖形物件複製支援	備註(O)
簡單長條圖 (B)	圖表板	條形圖(B)	是	否	N/A	是	
		個數長條圖	是	否	N/A	是	
	分佈	條形圖(B)	是	N/A	否	是	
堆疊長條圖 (C)	圖表板	條形圖(B)	是	是	N/A	是，但有限制	僅對「選用審美標準」中的類別變數為「是」。
		個數長條圖	是	是	N/A	是，但有限制	僅對「選用審美標準」中的類別變數為「是」。
	分佈	條形圖(B)	是	N/A	是	是	
盒形圖	圖表板	盒形圖	是	否	N/A	是，但有限制	僅在 Windows 上為「是」。
		盒形圖	是	是	N/A	否	
叢集盒形圖	圖表板	叢集盒形圖	是	否	N/A	是，但有限制	僅在 Windows 上為「是」。
		叢集盒形圖	是	是	N/A	否	



表 42. 複製 Microsoft 圖形物件支援 (繼續)

圖形輸出範本	模型器圖形節點	模型器圖形類型	基本設定	選用審美標準	重疊	Microsoft 圖形物件複製支援	備註(O)
簡單散佈圖	圖表板	泡泡圖	是	否	N/A	是，但有限制	僅對 X 及 Y 欄位中的連續變數以及「大小」中的類別變數為「是」。
		散佈圖	是	否	N/A	是，但有限制	僅對 X 及 Y 欄位中的連續變數為「是」。
	統計圖	點	是	N/A	否	是，但有限制	僅對 X 及 Y 欄位中的連續變數為「是」。
分組散佈圖	圖表板	泡泡圖	是	是	N/A	否	
		散佈圖	是	是	N/A	是，但有限制	僅對 X 及 Y 欄位中的連續變數以及「選用審美標準」中的類別變數為「是」。
	統計圖	點	是	N/A	是	是，但有限制	僅對 X 及 Y 欄位中的連續變數以及「覆蓋」中的類別變數為「是」。

## 匯出圖形

匯出圖形選項可讓您以下列其中一種格式匯出圖形：點陣圖 (.bmp)、JPEG (.jpg)、PNG (.png)、HTML (.html)、PDF (.pdf) 或 ViZml 文件 (.xml)，以便在其他應用程式中使用。

註：選取 PDF 選項時，圖形將匯出為高解析度 PDF 檔案，且檔案大小已裁剪為圖形大小。

若要匯出圖形，請從功能表選擇**檔案 > 匯出圖形**，然後選擇格式。

## 匯出表格

匯出表格選項可讓您以下列其中一種格式匯出表格：Tab 鍵定界 (.tab)、逗點定界 (.csv) 或 HTML (.html)。

若要匯出表格，請從功能表選擇**檔案 > 匯出表格**，然後選擇格式。



---

## 第 6 章 輸出節點

---

### 輸出節點概觀

輸出節點提供了用於獲取資料和模型的相關資訊的方法。還提供了以各種格式匯出資料以與其他軟體工具相互作用的機制。

下列輸出節點可用：



「表格」節點以表格式顯示資料，這些資料還可以寫入到檔案中。每當您需要檢查資料值或將其匯出為可輕鬆讀取的形式時，此節點非常有用。



「矩陣」節點將建立一個欄位關係表格。此節點最一般於顯示兩個符號欄位間的關係，但也可用於顯示旗標欄位或數字欄位間的關係。



「分析」節點評估預測模型產生準確預測的能力。「分析」節點執行一個或多個模型塊的預測值和實際值之間的各種比較。「分析」節點也可以對比各個預測模型。



「資料審核」節點將首先全面檢查資料，這些資料包含每個欄位的彙總統計資料、直方圖和分佈以及有關離群值、遺漏值和極端值的資訊。結果顯示在易於讀取的矩陣中，該矩陣可以排序並且可以用於產生完整大小的圖表和資料預備節點。



通過的轉換節點可首先選取和以可視方式預覽的轉換結果，然後再將其套用於選取的欄位。



「統計資料」節點可提供有關數值型欄位的基本彙總資訊。它可計算單個欄位以及欄位間的相關性的彙總統計資料。



「平均值」節點會比較獨立群組之間的平均值，或比較相關欄位的配對之間的平均值，以測試是否存在顯著性差異。例如，您可以比較開展促銷 前後的平均營收，或者將來自未接受促銷客戶的營收與接受促銷客戶的營收進行比較。



報告節點可建立格式化報告，其中包含固定文字、資料及得自資料的其他表示式。通過使用文字範本定義固定文字和資料輸出建構，可以指定報告的格式。通過使用範本中的 HTML 標籤和在「輸出」標籤上設定選項，可以提供自訂文字格式。通過使用範本中的 CLEM 表示式，可以包含資料值和其他條件輸出。



設定廣域節點掃描資料並計算可在 CLEM 表示式中使用的彙總值。例如，可以用該節點為一個名為年齡的欄位計算統計資料並通過插入函數 @GLOBAL\_平均值(age) 在 CLEM 表示式中使用年齡的總平均數。



「模擬配適」節點檢查每個欄位中資料的統計分佈，並產生（或更新）「模擬產生」節點，同時將最佳配適分佈分配給每個欄位。然後，可以使用「模擬產生」節點來產生模擬資料。



「模擬評估」節點對指定的預測目標欄位進行評估，並顯示有關該目標欄位的分佈和相關性資訊。

---

## 管理輸出

輸出管理器可顯示在 IBM SPSS Modeler 階段作業期間產生的圖表、圖形和表格。您始終可以通過在管理器中按兩下輸出來將它重新開啟，而不必重新執行對應的串流或節點。

### 檢視輸出管理器

開啟「檢視」功能表，然後選擇**管理程式**。按一下**輸出**標籤。

從輸出管理器中，可以：

- 顯示現有輸出物件，如直方圖、評估表和表格。
- 更名輸出物件。
- 將輸出物件儲存到磁碟或 IBM SPSS Collaboration and Deployment Services 儲存庫（如果可用）。
- 將輸出檔案新增到目前專案中。
- 從現行階段作業中刪除未儲存的輸出物件。
- 開啟已儲存的輸出物件或從 IBM SPSS Collaboration and Deployment Services 儲存庫 進行擷取（如果可用）。

要存取這些選項，請在「輸出」標籤上的任意位置按一下右鍵。

## 檢視輸出

畫面上的輸出顯示在輸出瀏覽器視窗中。輸出瀏覽器視窗具有其自己的一組功能表，可讓您列印或儲存輸出，或者以另一種格式匯出輸出。請注意，具體選項會因輸出類型的不同而不同。

列印、儲存和匯出資料。下面提供了詳細資訊：

- 要列印輸出，請使用列印功能表選項或按鈕。在列印之前，您可以使用頁面設定及預覽列印來設定列印選項及預覽輸出。
- 要將輸出儲存在 IBM SPSS Modeler 輸出檔案 (.cou)，請從「檔案」功能表中選擇儲存或另存為。
- 要以另一種格式（如文字或 HTML）儲存輸出，請從「檔案」功能表中選擇匯出。請參閱第 266 頁的『匯出輸出』主題，以取得更多資訊。

請注意，僅當輸出包含可以通過該方式正確匯出的資料時，才能選取這些格式。例如，決策樹的內容可以匯出為文字，但 K-Means 模型的內容則不能匯出為文字。

- 要將輸出儲存在共用儲存庫中使得其他使用者可使用 IBM SPSS Collaboration and Deployment Services 部署入口網站 檢視，從「檔案」功能表中選擇 發行到網路。請注意，該選項需要 IBM SPSS Collaboration and Deployment Services 的個別授權。

選取單元和欄。「編輯」功能表包含用於選取、取消選取和複製單元和欄且適合於目前輸出類型的各種選項。如需相關資訊，請參閱第 267 頁的『選取單元和欄』。

產生新節點。使用「產生」功能表，可以根據輸出瀏覽器的內容產生新節點。這些選項因輸出類型以及目前在輸出中選取的項目的不同而有所差異。有關特定輸出類型的節點產生選項的詳細資料，請參見該輸出的文件。

## 發行至網路

「發行到網路」功能可以將特定類型的串流輸出發行到形成 IBM SPSS Collaboration and Deployment Services 基準的中央共用 IBM SPSS Collaboration and Deployment Services 儲存庫。如果您使用此選項，要檢視該輸出的其他使用者可使用網際網路存取和 IBM SPSS Collaboration and Deployment Services 帳戶進行檢視，而無需安裝 IBM SPSS Modeler。

下表格列出了支援「發行到網路」功能的 IBM SPSS Modeler 節點。這些節點的輸出以輸出物件格式 (.cou) 儲存在 IBM SPSS Collaboration and Deployment Services 儲存庫中，並可直接在 IBM SPSS Collaboration and Deployment Services 部署入口網站中檢視。

其他類型的輸出只能在使用者機器上安裝了相關應用程式（例如，對於串流物件，需要 IBM SPSS Modeler）時才能進行檢視。

表 43. 支援「發佈至網路」的節點.

節點類型	節點
圖形	all
輸出	表格
	矩陣
	資料審核
	轉換
	平均數
	分析
	Statistics
報告 (HTML)	

表 43. 支援「發佈至網路」的節點 (繼續).

節點類型	節點
IBM SPSS Statistics	統計資料輸出

## 發行輸出到網路

要發行輸出到網路：

1. 在 IBM SPSS Modeler 串流中，執行表格中列出的某個節點。這會在新視窗中建立輸出物件（例如，表格、矩陣或報告物件）。
2. 從輸出物件視窗中，選擇：

檔案 > 發行至網路

注意：如果想匯出簡單 HTML 檔案用於標準 Web 瀏覽器，請從「檔案」功能表選擇匯出並選擇 **HTML**。

3. 連接至 IBM SPSS Collaboration and Deployment Services 儲存庫。

在連接成功後，將顯示「儲存庫：儲存」對話框，其中提供了多種儲存選項。

4. 在您選擇所需的儲存選項後，按一下儲存。

## 在網絡上檢視發行的輸出

使用此功能需要設定有 IBM SPSS Collaboration and Deployment Services 帳戶。如果您為要檢視的物件類型安裝有相關應用程式（例如 IBM SPSS Modeler 或 IBM SPSS Statistics），則輸出將顯示在應用程式本身而不是瀏覽器中。

要在網絡上檢視發行的輸出：

1. 將您的瀏覽器指向 `http://<repos_host>:<repos_port>/peb`

其中 `repos_host` 和 `repos_port` 為 IBM SPSS Collaboration and Deployment Services 主機的主機名稱和埠號。

2. 輸入您的 IBM SPSS Collaboration and Deployment Services 帳戶的詳細登入資訊。
3. 按一下內容儲存庫。
4. 導覽到或搜尋您要檢視的物件。
5. 按一下物件名稱。對於某些物件類型，例如圖形，在瀏覽器中呈現物件可能需要一些時間。

## 在 HTML 瀏覽器中檢視輸出

從線性、Logistic 和主成分/因子模型塊的「進階」標籤中，可以用個別的瀏覽器（如 Internet Explorer）檢視所顯示的資訊。該資訊是 HTML 形式的輸出，因而您可以儲存它並在別處（如公司內部網路或網際網路網站）重複使用。

要在瀏覽器中顯示資訊，請按一下模型塊「進階」標籤對話框左上角的模型圖示下方的啟動按鈕。

## 匯出輸出

在輸出瀏覽器視窗中，可以選擇將輸出匯出為另一格式（如文字或 HTML）。匯出格式因輸出類型的不同而不同，但一般類似於在用於產生輸出的節點中選取儲存到檔案時可以使用的檔案類型選項。

註：僅當輸出包含可以通過該方式正確匯出的資料時，才能選取這些格式。例如，決策樹的內容可以匯出為文字，但 K-Means 模型的內容則不能匯出為文字。

## 匯出輸出

1. 在輸出瀏覽器中，開啟「檔案」功能表並選擇**匯出**。然後，選擇要建立的檔案類型：

- **以 Tab 空格區隔(\*.tab)**。此選項會產生包含資料值的格式化文字檔。此樣式通常用於產生可匯入至其他應用程式的資訊的純文字表示法。此選項可用於「表格」、「矩陣」及「平均數」節點。
- **以逗點區隔(\*.dat)**。此選項會產生包含資料值的逗點分隔文字檔。透過此樣式通常可快速產生可匯入至試算表或其他資料分析應用程式的資料檔。此選項可用於「表格」、「矩陣」及「平均數」節點。
- **已調換的 Tab 空格區隔 (\*.tab)**。此選項與「以 Tab 空格區隔」選項相同，但是資料進行了已轉置，以便列代表欄位，欄代表記錄。
- **轉置的逗號分隔 (\*.dat)**。此選項與「逗號分隔」選項相同，但是資料進行了已轉置，以便列代表欄位，欄代表記錄。
- **HTML (\*.html)**。此選項將 HTML 格式的輸出寫入至檔案。

## 選取單元和欄

許多節點（包含「表格」節點、「矩陣」節點和「平均數」節點）產生表格輸出。可通過相似方式檢視並操作這些輸出表格，包含選擇單元、將表格的全部或部分複製到剪貼簿、根據目前選擇產生新節點，以及儲存和列印表格。

**選取單元**。請按一下 Cell 以將其選中。要選取矩形單元區域，請按一下所需區域的一個角落，拖曳滑鼠至該區域的另一個角落，然後鬆開啟滑鼠按鈕。要選取一整欄，請按一下欄標題。要選取多欄，請按住 Shift 鍵或 Ctrl 鍵並按一下直欄標題。

進行新選擇時，舊選擇會被清除。通過在按住 Ctrl 鍵的同時進行選擇，可以將新選擇新增到任何現有選擇中，而不是清除舊選擇。可以使用此方法選取多個非連續的表格區域。「編輯」功能表還包含**全選**和**清除選擇**選項。

**對欄進行重新排序**。使用「表格」節點和「平均數」節點的輸出瀏覽器，可以移動表格中的欄，方法是：按一下欄標題，並將它拖至所需位置。一次只能移動一欄。

---

## 表格節點

表格節點可以建立能夠列出資料中的值的表格。該表中包含了串流中的所有欄位和所有值，從而可以方便檢查資料值或以易於讀取的格式進行匯出。此外，您還可以強調顯示符合特定條件的記錄。

**註：**除非您在使用小資料集，否則建議您選取資料的子集合傳送到「表格」節點。當記錄數超過顯示結構可以包含的大小（例如：1 億列）時，「表格」節點無法準確顯示。

### 表格節點的「設定」標籤

強調顯示符合以下條件的記錄。通過輸入適用於要強調顯示的記錄的 CLEM 表示式，可以強調顯示表格中的記錄。只有在已選取輸出至螢幕時，此選項才會啟用。

### 表格節點格式標籤

「格式」標籤包含用來逐個欄位地指定格式的選項。此標籤與「類型」節點共用。請參閱第 128 頁的『欄位格式設定標籤』主題，以取得更多資訊。

### 輸出節點的「輸出」標籤

對於產生表格樣式輸出的節點，使用「輸出」標籤可指定結果的格式和位置。

**輸出名稱。**指定執行節點時產生的輸出名稱。**自動**會根據產生輸出的節點來選擇名稱。您可以選擇性地選取自訂以指定不同的名稱。

**輸出至螢幕**（預設選項）。建立輸出物件以在線上檢視。執行輸出節點時，輸出物件將在管理程式視窗的「輸出」標籤上出現。

**輸出至檔案。**執行節點時將輸出儲存到檔案。如果選擇此選項，請輸入檔名（或導覽到某目錄，並使用檔案選擇器按鈕指定檔名）並選擇檔案類型。請注意，有些檔案類型可能不適用於某些特定類型的輸出。

註：

來自輸出節點的輸出資料根據下列規則進行編碼：

- 在執行輸出節點時，串流編碼值（在「串流選項」標籤上設定）將設定為輸出。
- 在產生輸出之後，即使串流編碼變更，其編碼也不會變更。
- 匯出輸出節點輸出時，將使用目前定義的編碼來匯出輸出檔案。在建立輸出後，即使變更串流編碼，也不會影響產生的輸出。

請注意這些規則的下列例外狀況：

- 所有 HTML 匯出均採用 UTF-8 格式編碼。
- 來自「延伸」輸出節點的輸出由自訂使用者 Script 產生。因此，編碼由 Script 控制。

下列選項可用於將輸出儲存到檔案：

- **資料（以 Tab 空格區隔）(\*.tab)**。此選項會產生包含資料值的格式化文字檔。此樣式通常用於產生可匯入至其他應用程式的資訊的純文字表示法。此選項可用於「表格」、「矩陣」及「平均數」節點。
- **資料（逗號分隔）(\*.dat)**。此選項會產生包含資料值的逗點分隔文字檔。透過此樣式通常可快速產生可匯入至試算表或其他資料分析應用程式的資料檔。此選項可用於「表格」、「矩陣」及「平均數」節點。
- **HTML (\*.html)**。此選項將 HTML 格式的輸出寫入至檔案。對於來自「表格」節點、「矩陣」節點或「平均數」節點的表格輸出，一組 HTML 檔案包含一個內容畫面，該窗格在 HTML 表格中列出了欄位名稱和資料。如果表格中的列個數超過每頁列個數規格，則可將表格分割為多個 HTML 檔案。在這種情況下，該內容畫面包含指向所有表格頁的鏈結，並提供導覽表格的方法。對於非表格輸出，會建立一個包含節點結果的 HTML 檔案。

註：如果 HTML 輸出只包含第一頁的格式，請選取**頁碼輸出**並調整**每頁行數**規格，以在單一頁上包含所有輸出。或者，如果節點（如報告節點）的輸出範本包含自訂的 HTML 標籤，則一定要指定**自訂**作為格式類型。

- **文字檔 (\*.txt)**。此選項產生包含輸出的文字檔。此樣式經常用於產生可匯入到其他應用程式（如文字處理器或簡報軟體）中的輸出。此選項對於某些節點不適用。
- **輸出物件 (\*.cou)**。對於以這種格式儲存的輸出物件，可在 IBM SPSS Modeler 中開啟並檢視、新增至專案，以及使用 IBM SPSS Collaboration and Deployment Services 儲存庫 發行和追蹤。

**輸出檢視。**對於「平均數」節點，可以指定依預設是顯示簡單輸出還是進階輸出。請注意，也可以在瀏覽產生的輸出時在這些視圖間切換。請參閱第 285 頁的『「平均數」節點輸出瀏覽器』主題，以取得更多資訊。

**格式。**對於「報告」節點，可以選擇是自動設置輸出格式，還是使用範本中併入的 HTML 設置輸出格式。選取自訂可容許使用範本中的 HTML 格式。

**標題。**對於報告節點，可以指定將顯示在報告輸出頂部的選用標題文字。



**反白插入文字。**對於報告節點，選取此選項可在報告範本中強調顯示由 CLEM 表示式產生的文字。請參閱主題第 287 頁的『報告節點的範本標籤』，瞭解進一步資訊。建議不要在使用自訂格式時有此選項。

**每頁行數。**對於報告節點，指定在輸出報告的自動格式設置期間要在每頁上包含的行數。

**調換資料。**此選項在匯出前轉置資料，以便列代表欄位，欄代表記錄。

註：對於大型表格，上述選項的效用可能會較差，尤其是在使用遠端伺服器時。在這種情況下，使用檔案輸出節點可提供更好的效能。請參閱第 314 頁的『純文字檔案匯出節點』主題，以取得更多資訊。

## 表格瀏覽器

表格瀏覽器顯示表格資料，並且可以用於執行標準作業，包括選取和複製單元、重新為欄排序，以及儲存和列印表格。請參閱第 267 頁的『選取單元和欄』主題，以取得更多資訊。這些作業與預覽節點中資料的作業相同。

**匯出表格資料。**要從表格瀏覽器匯出資料，請選擇：

檔案 > 匯出

請參閱第 266 頁的『匯出輸出』主題，以取得更多資訊。

資料以系統預設值編碼格式匯出，編碼格式可在 Windows 控制台中指定，如果以分散式模式執行，則在伺服器電腦上指定。

**搜尋表格。**主工具列上的搜尋按鈕（使用雙筒望遠鏡圖示）可啟動搜尋工具列，從而使您可以在表格中搜尋特定值。您可以在表格中向前或向後搜尋，可以指定區分觀察值的搜尋（**Aa** 按鈕），並且可以使用中斷搜尋按鈕中斷正在進行的搜尋。

**產生新節點。**「產生」功能表包含節點產生作業。

- 「選取」節點（「記錄」）。產生選取節點，該節點對表格中任何已選取的 Cell 的記錄進行選取。
- 選取（「和」）。產生選取節點，該節點選取包含了表格中所有已選取值的記錄。
- 選取（「或」）。產生選取節點，該節點選取包含了表格中任何已選取值的記錄。
- 衍生（「記錄」）。產生「衍生」節點以建立一個新的旗標欄位。旗標欄位包含 *T*（表示表格中任何已選取的 Cell 的記錄）和 *F*（表示其餘記錄）。
- 衍生（「和」）。產生「衍生」節點以建立一個新的旗標欄位。旗標欄位包含 *T*（表示包含了表格中所有已選取值的記錄）和 *F*（表示其餘記錄）。
- 衍生（「或」）。產生「衍生」節點以建立一個新的旗標欄位。旗標欄位包含 *T*（表示包含了表格中任何已選取值的記錄）和 *F*（表示其餘記錄）。

---

## 矩陣節點

使用「矩陣」節點可以建立顯示欄位間關係的表格。它最常用於顯示兩個種類欄位（旗標、列名或序數）之間的關係，但也可用於顯示連續（數值型範圍）欄位之間的關係。

### 矩陣節點的設定標籤

「設定」標籤用於為矩陣結構指定選項。

**欄位。**從下列選項中選取欄位選擇類型：

- **選定字段**。使用此選項可以為矩陣的列和欄分別選取一個種類欄位。矩陣的列與欄由所選種類欄位的值清單定義。矩陣儲存格包含以下所選的彙總統計量。
- **全部旗標 (true 值)**。此選項要求一個滿足如下條件的矩陣：資料中的每個旗標欄位都包含一列和一欄。矩陣儲存格的每個旗標組合都包含倍精準確度正數的計數。換言之，對於與所購面包對應的列和與所購干酪對應的欄，該列和欄交叉處的 Cell 包含所購面包和所購干酪均為 true 值的記錄的個數。
- **所有數值**。此選項要求一個滿足如下條件的矩陣：每個數值欄位都包含一個列和一個欄。矩陣的單元代表相應欄位對的交叉product的總和。換言之，對於矩陣中的每個 Cell，會將該格內每條記錄的列欄位值和欄欄位值相乘，然後對格內所有記錄的product值求和。

**包含遺漏值**。在列和欄輸出中包含使用者遺漏值 (空白值) 和系統遺漏值 (\$null\$)。例如，如果已將值 N/A 定義為所選欄位的使用者遺漏，則表格中將併入標籤為 N/A 的個別欄 (假設此值實際出現在資料中)，就像任何其他種類一樣。如果取消選取此選項，則無論 N/A 欄的出現頻率高低，都會將它排除。

注意：用於包含遺漏值的選項僅在所選取的欄位為交叉列表形式時適用。空白值對映至 \$null\$，並且在模式為已選取且內容設定為函數時從函數欄位的合計中排除，在模式設定為所有數值時從所有數字欄位的合計中排除。

**儲存格內容**。如果已經選取上方的所選欄位，則可以指定要在矩陣的單元中使用的統計資料。選取基於計數的統計量，或選取套版欄位以根據列和欄欄位的值彙總數值欄位的值。

- **交叉列表**。Cell 值是用於統計多少記錄具有對應值組合的計數和/或百分比。您可以使用「外觀」標籤上的選項指定所需的交叉列表彙總。廣域卡方值也會和顯著性一起顯示。請參閱第 271 頁的『矩陣節點輸出瀏覽器』主題，以取得更多資訊。
- **函數**。如果選取彙總函數，那麼 Cell 值是所選套版欄位值的函數 (對於具有適當的列值和欄值的情況)。例如，如果列欄位為區域，欄欄位為產品，套版欄位為營收，則位於東北列和小器具欄中的 Cell 將包含東北區域出售小器具所獲得的營收的總和 (或平均值、最小值或最大值)。預設彙總函數是 **Mean**。您可以選取其他函數來彙總該函數欄位。選項包含 **Mean**、**Sum**、**SDev** (標準差)、**Max** (最大值) 和 **Min** (最小值)。

## 矩陣節點的外觀標籤

使用「外觀」標籤，可以控制矩陣的排序和強調顯示選項，以及為交叉列表矩陣提供的統計資料。

**列與欄**。控制矩陣中列和欄標題的排序。預設值是**未排序**。選取遞增或遞減可按指定的方向為列和直欄標題排序。

**重疊**。用於強調顯示矩陣中的極端值。值根據儲存格計數 (對於交叉列表矩陣) 或計算出的值 (對於函數矩陣) 強調顯示。

- **反白標示頂端**。您可以要求強調顯示矩陣中大小排在前幾位的值 (以紅色顯示)。指定要強調顯示的數目。
- **反白標示底端**。您也可以要求強調顯示矩陣中大小排在後幾位的值 (以綠色顯示)。指定要強調顯示的數目。

注意：對於這兩個強調顯示選項，同分值的存在可能導致強調顯示的值比已要求強調顯示的值多。例如，如果您有一個矩陣，而該矩陣在單元中包括六個零，當您要求**強調顯示後 5 個值**時，則將強調顯示全部六個零。

**交叉列表儲存格內容**。對於交叉列表，可以指定包含在交叉列表的矩陣中的彙總統計資料。當在「設定」標籤上選取了**所有數值或函數**選項時，這些選項無法使用。

- **個數**。儲存格包含其列值具有對應欄位的記錄數。這只是預設 Cell 內容。
- **期望值**。Cell 中記錄數的期望 (假設列和欄之間沒有關係)。預期值基於下列公式：

$$p(\text{row value}) * p(\text{column value}) * \text{total number of records}$$

- 殘差。觀察值和預期值之間的差異。
- 列的百分比。其列值具有對應欄值的所有記錄的百分比。列百分比的和為 100。
- 欄的百分比。其欄值具有對應列值的所有記錄的百分比。欄百分比的和為 100。
- 合計百分比。具有列值和欄值組合的所有記錄的百分比。整個矩陣的百分比和為 100。
- 包含列和欄的合計。將列和欄新增到列和欄合計的矩陣。
- 套用設定。（僅限輸出瀏覽器）使您可以變更「矩陣」節點輸出的外觀，且不必關閉並重新開啟輸出瀏覽器。在輸出瀏覽器的此標籤上進行變更，按一下此按鈕，然後選取「矩陣」標籤以請參閱變更的效果。

## 矩陣節點輸出瀏覽器

矩陣瀏覽器顯示交叉列表資料，並可用於針對矩陣執行作業，包含選擇單元、將矩陣全部或部分複製到剪貼簿、根據矩陣選擇生成新節點，以及儲存和列印矩陣。矩陣瀏覽器還可用於顯示某些特定模型的輸出，如 Oracle 的貝式邏輯分類演算法模型。

「檔案」和「編輯」功能表提供用於列印、儲存和匯出輸出以及用於選取和複製資料的常用選項。請參閱第 265 頁的『檢視輸出』主題，以取得更多資訊。

**卡方分配。** 對於含兩個種類欄位的交叉表，還會在該表格的下方顯示廣域 Pearson 卡方。此測試指出兩個欄位之間不存在關係的機率（根據的是不存在關係時觀測計數和預期計數之間的差分）。例如，如果客戶滿意率和商店位置之間沒有關係，那麼您將預期所有商店的客戶滿意率相似。但是，如果某些特定商店報告的客戶滿意率總是比其他商店高，那麼您可能會懷疑這並非巧合。差分越大，則單純由隨機取樣誤導致此結果的機率就越小。

- 卡方測試指出兩個欄位之間不存在關係的機率，如果兩個欄位之間不存在關係，那麼觀測頻率和預期頻率之間的任何差分完全歸咎於隨機效應。如果此機率非常小（通常少於 5%），那麼說明兩個欄位之間存在顯著的聯系。
- 如果只有一欄或一行（單因素卡方測試），那麼自由度等於單元數減一。對於雙因素卡方，自由度等於列個數和欄個數均減一之後二者的product。
- 如果任何預期的 Cell 頻率少於五，那麼解釋卡方測試統計資料時請務必小心。
- 卡方測試只適用於含兩個欄位的交叉列表。（當在「設定」標籤上選取全部旗標或所有數值時，此測試不會顯示。）

**「產生」功能表** 「產生」功能表包含節點產生作業。這些作業只適用於交叉列表矩陣，並且必須至少在矩陣中選取了一個 Cell。

- **選取節點。** 產生「選取」節點，該節點選取與矩陣中的任何選定儲存格相符的記錄。
- **「衍生」節點（旗標）。** 產生「衍生」節點以建立一個新的旗標欄位。旗標欄位包含  $T$ （表示與矩陣中的任何選定 Cell 相符的記錄）和  $F$ （表示其餘記錄）。
- **「衍生」節點（集合）。** 產生用於新建列名欄位的「衍生」節點。對於矩陣中的每個連續的選定單元集，列名欄位都包含一個種類。

---

## 「分析」節點

通過使用「分析」節點，您可以對模型產生準確預測的能力進行評估。「分析」節點將對一個或多個模型塊的預測值和實際值（您的目標欄位）進行各種比較。「分析」節點也可用於將一些預測模型和其他預測模型進行比較。

執行「分析」節點時，對於已執行的串流中的每個模型塊，會自動將分析結果摘要新增至「摘要」標籤上的「分析」部分。詳細分析結果顯示在管理器視窗的「輸出」標籤中，或者直接寫入檔案。

註：由於「分析」節點將預測值與實際值進行比較，因此它們只適用於受監督模型（需要目標欄位的模型）。對於非受監督模型（例如，叢集演算法），沒有實際結果可用作比較基準。

## 「分析」節點的「分析」標籤

通過使用「分析」標籤，您可以指定分析詳細資料。

**符合矩陣（適用於符號目標或種類目標）。**顯示種類目標（旗標、名義或序數）的各個產生（預測值）欄位與其目標欄位之間的相符型樣。將顯示一個表格，其中包含實際值所定義的列和預測值所定義的欄，以及每個 Cell 中符合該型樣的記錄數。這用於確定預測中的系統錯誤。如果產生了多個與同一輸出欄位相關的欄位，但這些欄位由不同模型產生，那麼將為這些欄位相同和不相同的情況進行計數並顯示總計值。對於它們相同的情況，將顯示另一組正確/錯誤統計資料。

**效能評估。**使用種類輸出顯示模型的效能評估統計資料。此統計資料（針對輸出欄位的每個種類報告）是以位元為單位對模型（用於預測的該種類的記錄）的平均值資訊內容的測量。考慮到分類問題的難度，因此，罕見種類的準確性預測會比常見種類的準確性預測獲得更高的效能評估指數。對於某個種類，如果模型效果比隨機猜測差，那麼該種類的效能評估指數將為 0。

**評估度量值（僅二元分類器 AUC 和 Gini）。**對於二元分類器，此選項將報告 AUC（曲線下區域）和 Gini 係數評估度量值。將對每個二元模型共同計算這兩個評估度量值。將在分析輸出瀏覽器中以表格報告這些度量值。

AUC 評估度量值按照 ROC（受試者工作特性）曲線下方的面積進行計算，它是分類器預期效能的標量代表。AUC 始終介於 0 到 1 之間，數字越大代表分類器越好。座標 (0,0) 與 (1,1) 之間的對角線 ROC 曲線代表隨機分類器，並且其 AUC 為 0.5。因此，實際分類器的 AUC 不會少於 0.5。

有時，Gini 係數評估測量用作 AUC 的替代評估測量，並且這兩個測量密切相關。Gini 係數計算為 ROC 曲線與對角線之間的面積的兩倍，或者按照  $Gini = 2AUC - 1$  進行計算。Gini 係數始終介於 0 到 1 之間，數字越大代表分類器越好。對於 ROC 曲線在對角線下方的不可能事件，Gini 係數為負。

**信賴度數字（如果可用）。**對於產生信賴度欄位的模型，此選項將報告關於信賴度值及其與預測的關係的統計資料。此選項有兩項設定：

- **臨界值。**報告精確度達到指定百分比的信賴等級。
- **提高精確度。**報告精確度提高指定系數的信賴等級。例如，如果總精確度為 90%，而此選項設定為 2.0，那麼所報告的值將是精確度為 95% 時所需的信賴度。

**尋找預測/預測值欄位使用。**確定預測欄位與原始目標欄位符合的方式。

- **模式輸出欄位 meta 資料。**基於模型欄位資訊使預測欄位與目標相符，即便在更名預測欄位的情況下也可以進行相符。通過使用「類型」節點，還可以從「值」對話框存取任何預測欄位的模型欄位資訊。請參閱第 123 頁的『使用值對話框』主題，以取得更多資訊。
- **欄位名稱格式。**根據命名慣例相符欄位。例如，C5.0 模型塊為目標 *response* 產生的預測值必須位於欄位  $\$C-response$  中。

**以分割區區隔。**如果使用分割區欄位將記錄分割為訓練樣本、測試樣本和驗證樣本，那麼選取此選項可個別為每個分割區顯示結果。請參閱第 150 頁的『分割區節點』主題，以取得更多資訊。

註：以分割區區隔時，將從分析中排除分割區欄位中具有空值的記錄。如果使用「分割區」節點，這永遠都不是問題，因為「分割區」節點不會產生空值。

**使用者定義的分析。**您可以指定要在評估模型時使用的分析計算。使用 CLEM 表示式可指定應為每條記錄計算的內容以及如何將記錄層次的分數合併到總分數中。使用函數 **@TARGET** 和 **@PREDICTED** 可分別參照目標（實際輸出）值和預測值。

- **If**。指定需要視某一條件使用不同計算時的條件式表示式。
- **Then**。指定 If 條件為 true 時的計算。
- **Else**。指定 If 條件為 false 時的計算。
- **使用**。選取用於從個別分數計算總分數的統計量。

**依欄位分解分析。**顯示可用於分解分析的種類欄位。除整體分析外，將為每個分解欄位的每個種類報告個別的分析。

## 分析輸出瀏覽器

分析輸出瀏覽器顯示了「分析」節點的執行結果。「檔案」功能表中提供了常用的儲存、匯出和列印選項。請參閱第 265 頁的『檢視輸出』主題，以取得更多資訊。

首次瀏覽分析輸出時，結果會展開。若要在檢視結果之後將其隱藏，請使用項目左側的展開工具控制項來收合要隱藏的特定結果，或按一下**全部收合**按鈕來收合所有結果。若要在收合結果之後再次查看，請使用項目左側的展開工具控制項來顯示結果，或按一下**全部展開**按鈕來顯示所有結果。

**輸出欄位的結果。**對於具有所產生的模型建立的相應預測欄位的每個輸出欄位，分析輸出都包含一個相應部分。

**比較。**在輸出欄位部分中，這是與該輸出欄位關聯的每個預測欄位的子部分。對於種類輸出欄位，此部分的最上層位置包含一張表格，其中顯示正確預測和不正確預測的號碼和百分比以及串流中的記錄總數。對於數值輸出欄位，此部分顯示下列資訊：

- **最小錯誤。**顯示最小錯誤（觀察值和預測值之間的差分）。
- **最大錯誤。**顯示最大錯誤。
- **均數誤。**顯示所有記錄之間的誤差平均數（平均值）。這指示模型中是否有系統**偏移**（過高估計的趨勢強於過低估計的趨勢，或相反）。
- **平均絕對錯誤。**顯示所有記錄的誤絕對值的平均值。指出誤的平均值長度（不考慮方向）。
- **標準差。**顯示誤的標準差。
- **線性相關。**顯示預測值和實際值之間的線性相關。此統計資料介於 -1.0 和 1.0 之間。值接近於 +1.0 表示強正相關，因此，高預測值與高實際值相關，而低預測值與低實際值相關。值接近於 -1.0 表示強負相關，因此，高預測值與低實際值相關，而低預測值與高實際值相關。值接近於 0.0 表示弱相關，因此，預測值幾乎獨立於實際值。註：此處的空白項目表示由於實際或預測值為常數，因此在該案例中無法計算線性相關。
- **出現次數。**顯示分析中使用的記錄數。

**符合矩陣。**對於種類輸出欄位，如果在分析選項中已要求了符合矩陣，那麼此處會顯示一個包含該矩陣的子部分。列代表實際觀察值，而欄代表預測值。表格中的 Cell 表示每個預測值和實際值組合的記錄數。

**效能評估。**對於種類輸出欄位，如果在分析選項中已要求了效能評估統計資料，那麼此處會顯示效能評估結果。每個輸出種類均與其效能評估統計資料一起列出。

**信賴度值報告。**對於種類輸出欄位，如果在分析選項中已要求了信賴度值，那麼此處會顯示這些值。對於模型信賴度值，將報告下列統計資料：

- **範圍。**顯示串流資料中記錄的信賴度值的範圍（最小值和最大值）。

- **正確分類的平均置信度。**顯示正確分類的記錄的平均值信賴度。
- **未正確分類的平均置信度。**顯示未正確分類的記錄的平均值信賴度。
- **總是正確的置信度下限。**顯示預測總是正確的信任臨界值的下限，並顯示符合此準則的案例所佔百分比。
- **永遠錯誤的置信度上限。**顯示預測始終錯誤的信任臨界值的上限，並顯示符合此準則的案例所佔百分比。
- **準確度 X% 以上方。**顯示準確度為 X% 時的信賴等級。X 是分析選項中為**臨界值**指定的近似值。對於某些模型和資料集，無法選擇提供精確臨界值（在選項中指定的）的信賴度值（通常是具有接近臨界值的相同信賴度值的類似觀察值的叢集所致）。所報告的臨界值是最接近於指定精確度準則的值，該值可以通過單個信賴度值臨界值獲取。
- **比整體準確性好 X 倍。**顯示比整個資料集的精確度好 X 倍時相應的信賴度值。X 是在分析選項中為**改進精確度**指定的值。

**間的合約。**如果串流中併入兩個或兩個以上對同一輸出欄位進行預測的產生的模型，那麼您還將看到與模型所產生的預測之間的**合約**相關的統計資料。這包含預測同意的記錄數和百分比（對於種類輸出欄位）或誤彙總統計資料的號碼和百分比（對於連續輸出欄位）。對於種類欄位，它包含分析與模型同意（產生相同預測值）的記錄子集合的實際值相比的預測值。

**評估度量值。**對於二元分類器，如果您已要求分析選項中的評估度量值，那麼 AUC 和 Gini 係數評估度量值將顯示在此區段的表格中。對於每個二元分類器模型，表格都包含與之對應的一列。評估度量值表格針對每個輸出欄位而不是每個模型顯示。

---

## 資料審核節點

通過資料審核節點，可對您放置到 IBM SPSS Modeler 中並且以易於讀取的矩陣（可對該矩陣進行排序並使用它產生正常大小的圖形和各式各樣的資料預備節點）形式顯示的資料有個初步的全面瞭解。

- 「審核」標籤顯示具有彙總統計資料、直方圖和分佈圖的報告，它們有助於獲得對資料的初步瞭解。該報告在欄位名稱之前還顯示儲存圖示。
- 審核報告中的「品質」標籤顯示有關離群值、極端值和遺漏值的資訊，並提供用於處理這些值的工具。

使用資料審核節點

資料審核節點可直接附加到來源節點，或附加到已實例化的類型 節點的下游。您也可以根據結果產生多個資料預備節點。例如，可以產生過濾器節點來排除具有過多遺漏值的欄位，這在建模中很有用，並產生 SuperNode 為任何或所有保留欄位歸因遺漏值。這就是審核的真正作用所在，使您不僅可以評量資料的現行狀態，還可以根據評量執行動作。

**篩選資料或對資料取樣。**因為初始審核在正在處理「大資料」時特別有效，所以可以在初始探索期間使用樣本節點選取部分記錄，以此縮短正在處理時間。在分析的探索階段，也可以將 資料審核節點與 Feature Selection 節點和 Anomaly Detection 節點等節點組合使用。

## 資料審核節點的設定標籤

使用「設定」標籤，可指定用於審核的基本參數。

**預設值。**您可以只是將節點附加到串流中，然後按一下**執行**以根據預設值產生所有欄位的審核報告，如下所示：

- 如果沒有類型 節點設定，那麼報告中將併入所有欄位。
- 如果有「類型」設定（無論它們是否已實例化），則顯示中併入所有輸入、目標和兩者欄位。如果有一個目標欄位，請使用它作為「套版」欄位。如果指定了多個目標欄位，則不指定預設套版。

**使用自訂欄位。**選取此選項可手動選取欄位。使用右側的欄位選擇器按鈕可個別選取欄位或按類型選取欄位。

**套版欄位。**套版欄位用於繪製審核報告中顯示的縮圖圖形。如果是連續（數值型範圍）欄位，則還計算雙變數統計資料（共變數和相關係數）。如果根據「類型」節點設定呈現了單個目標欄位，則使用它作為預設套版欄位，如上方所述。或者，您可以選取使用自訂欄位以指定套版。

**顯示。**用於指定輸出中是否顯示圖形以及選擇預設顯示的統計資料。

- **圖形。**顯示每個選定欄位的圖表；根據資料的情況顯示為分佈（條形）圖形、直方圖或散佈圖。圖形在初始報告中顯示為縮圖，但也可產生標準大小的圖形和圖形節點。請參閱『資料審核輸出瀏覽器』主題，以取得更多資訊。
- **基本/進階統計量。**指定預設顯示在輸出中的統計資料的層次。當此設定確定初始顯示時，無論此設定是什麼，所有統計資料均顯示在輸出中。請參閱第 276 頁的『顯示統計資料』主題，以取得更多資訊。

**中位數和眾數。**計算報告中所有欄位的中位數和眾數。請注意，對於大型資料集，由於這些統計資料的計算時間比其他統計資料長，因此它們可能會延長正在處理時間。在僅含中位數的情況下，報告值有時基於含 2000 條記錄的樣本（而不是完整資料集）。為了防止超過記憶體限制，此取樣是逐個欄位地進行的。當取樣生效時，輸出中的結果將標記為（樣本中位數而不只是中位數）。對於除中位數之外的所有統計資料，始終使用完整資料集進行計算。

**空白欄位或無類型欄位。**當無類型欄位與實例化資料配合使用時，這些欄位不會併入在審核報告中。要包含無類型欄位（包含空白欄位），請在位於上游的所有類型節點中選取**清除所有值**。這可確保不實例化資料，從而導致在報告中併入所有欄位。例如，如果要獲取所有欄位的完整清單或產生將排除空白欄位的過濾器節點，那麼這非常有用。請參閱第 279 頁的『過濾含遺漏資料的欄位』主題，以取得更多資訊。

## 資料審核的品質標籤

資料審核節點中的「品質」標籤提供用於處理遺漏值、離群集和極端值的選項。

### 遺漏值

- **含有效值的記錄的計數。**選取此選項可為每個評估欄位顯示含有效值的記錄數。請注意，空值（未定義）、空白值、空白和空字串總是被視為無效值。
- **含無效值的記錄的分類計數。**選取此選項可為每個欄位顯示含每類無效值的記錄數。

### 離群值與極端值

離群值與極端值的偵測方法。受支援兩種方法：

**與平均數的標準差。**根據與平均數的標準差的個數偵測離群值與極端值。例如，如果您具有一個包含平均數為 100 和標準差 10 的欄位，那麼可以指定 3.0 來指出應將任何低於 70 或高於 130 的值視為離群值。

**內四分位距。**根據內四分位距（即中間兩個四分位數的間距，介於 25% 百分位數和 75% 百分位數之間）偵測離群值與極端值。例如，根據預設設定 1.5，離群值的臨界值下限將為  $Q1 - 1.5 * IQR$ ，臨界值上限將為  $Q3 + 1.5 * IQR$ 。請注意，使用此選項可能會降低大型資料集的效能。

## 資料審核輸出瀏覽器

資料審核瀏覽器是用於獲取資料概觀的強大工具。「審核」標籤顯示所有欄位的縮圖、儲存圖示以及統計資料，而「品質」標籤顯示有關離群值、極端值和遺漏值的資訊。根據初始圖形和彙總統計資料，您可以決定重新為數值欄位編碼、衍生新欄位，或重新為名義欄位的值分類。或者，您可能需要使用更加高級的視覺化功能來進一步進行探索。您可以透過使用「產生」功能表建立可用於轉換或視覺化資料的任意數量節點，直接從審核報告瀏覽器執行此動作。

- 通過按一下欄標題為欄排序，或使用拖放來重新為欄排序。並且受支援大多數標準輸出作業。請參閱第 265 頁的『檢視輸出』主題，以取得更多資訊。

- 通過按兩下「測量」欄或「唯一」欄中的欄位檢視欄位的值和範圍。
- 使用工具列或「編輯」功能表顯示或隱藏值標籤，或選擇要顯示的統計資料。請參閱『顯示統計資料』主題，以取得更多資訊。
- 檢驗欄位名左側的儲存圖示。儲存體說明資料儲存在欄位中的方式。例如，值為 1 和 0 的欄位會儲存整數資料。它與測量層次不同，說明資料使用情形，並不會影響儲存體。請參閱第 7 頁的『設定欄位儲存類型和格式』主題，以取得更多資訊。

## 檢視和產生圖形

如果未選取套版，那麼「審核」標籤顯示長條圖（對於名義欄位或旗標欄位）或直方圖（連續欄位）。

對於名義或旗標欄位套版，將以套版的值為圖形著色。

對於連續欄位套版，會產生二維度散佈圖，而不是一維度長條和直方圖。在這種情況下， $x$  軸對映到套版欄位，從而使您可以在沿著表格向下讀取時，看到的所有  $x$  軸上相同尺度。

- 對於旗標或名義欄位，將滑鼠指針停留在條形圖上可在工具提示中顯示基礎值或標籤。
- 對於旗標或名義欄位，使用工具列可將縮圖的方向從水平切換為垂直。
- 要從任何縮圖產生標準大小的圖表，請按兩下縮圖，或選擇縮圖，然後從「產生」功能表中選擇圖形輸出。注意：如果縮圖基於抽樣資料，那麼在原始資料串流仍然開啟時，產生的圖表將包含全部觀察值。

如果建立輸出的「資料審核」節點已連線到串流，則只能產生圖表。

- 要產生相符的圖表節點，請在「審核」標籤上選擇一個或多個欄位，然後從「產生」功能表中選擇圖表節點。產生的節點會被新增到串流畫布中，並且可在每次執行串流時用於重新建立圖表。
- 如果套版集具有 100 個以上的值，則會發出警告，並且不會併入套版。

## 顯示統計資料

使用「顯示統計資料」對話框，可以選擇顯示在「審核」標籤上的統計資料。初始設定是在 資料審核節點中指定的。請參閱第 274 頁的『資料審核節點的設定標籤』主題，以取得更多資訊。

最小值. 數值變數的最小值。

最大值. 數值變數的最大值。

總和( $U$ ). 遍及所有包含遺漏值之觀察值的值總和或總數。

範圍. 值變數最大與最小值之間的差異，也就是最大值減去最小值。

平均數( $M$ ). 集中趨勢的測量。算術平均數，總和除以觀察值數。

平均數的標準誤 (*Standard Error of Mean*). 測量從同一個分配取出來的不同樣本間平均數的變化大小。它可以用來大略地將觀察平均數與假設值相比較（也就是如果標準誤與差異的比值小於 -2 大於 +2 的話，您就可以下結論說兩個值不同）。

標準差. 在平均數四周離散的量數，等於變異數的平方根。標準差所用的測量單位與原始觀察值相同。

變異數. 測量平均數四周的離散情形，它等於平均數的平方離差總和除以觀察值數目減一。變異數的測量單位是變數本身的平方。

偏斜度. 分配不對稱性的量數。常態分佈是一種對稱性分佈，其偏斜度值為 0。具有顯著性正偏斜度的分佈右側尾部較長。有顯著負偏斜度值的分配會有一個長的偏左尾部。偏斜度值有如指標，若大於它的兩倍標準誤，則表示背離對稱。



偏斜度的標準誤 (*Standard Error of Skewness*). 偏斜度對其標準誤的比例可用於測試常態性 (也就是說, 如果比例小於 -2 或大於 +2, 則您可以拒絕常態性)。偏斜度若為大的正值, 表示有長的偏右尾端; 極端負值則表示有長的偏左尾端。

峰度. 離群值範圍的一種測量。對常態分配而言, 峰度統計量數值為零。峰度為正數指出資料比常態分佈展示更多的極端離群值。峰度為負數指出資料比常態分佈展示較少的極端離群值。

峰度的標準誤 (*Standard Error of Kurtosis*). 峰度對其標準誤的比例可用於測試常態性 (也就是說, 如果比例小於 -2 或大於 +2, 則您可以拒絕常態性)。峰度若為大的正值, 表示該分配的尾端比常態分配的尾端更長; 峰度若為負值, 表示尾端較短 (變成類似箱型均勻分配的尾端)。

唯一 (*Unique*). 針對任何類型的所有其他效果, 調整每個效果來同時評估所有效果。

有效. 沒有系統遺漏值, 也沒有定義為使用者遺漏值的有效觀察值。請注意, 空值 (未定義)、空白值、空白和空字串總是被視為無效值。

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

眾數 (*O*). 最常出現的值。如果幾個值的最大出現次數相同, 則每一個都是眾數。

請注意, 為了提高效率, 依預設不會顯示中位數和眾數, 但是您可以在 資料審核節點的「設定」標籤上將其已選取。請參閱第 274 頁的『資料審核節點的設定標籤』主題, 以取得更多資訊。

套版的統計資料

如果連續 (數值型範圍) 套版欄位正在使用, 則下列統計資料也可用:

共變異數 (*Covariance*). 兩個變數之間關聯的非標準化量數, 它等於交叉product離差除以  $N-1$ 。

## 資料審核瀏覽器的品質標籤

「資料審核」瀏覽器中的「品質」標籤顯示資料品質分析的結果, 並且可用於指定離群值、極端值和遺漏值的處理。

填補遺漏值: 審核報告列出每個欄位完整紀錄的百分比以及有效值、空白和空值的號碼。您可以根據情況選擇填補特定欄位的遺漏值, 然後產生 SuperNode 以套用這些轉換。

1. 在填補遺漏值欄中, 指定要填補的值的類型 (如果有)。您可以選擇填補空白值和/或空白值, 或指定用於選擇待填補值的自訂條件或表示式。

IBM SPSS Modeler 可識別的遺漏值類型有以下幾種:

- **空值或系統遺漏值。**這兩種類型是資料庫或來源檔案中留空白、並且尚未在來源節點或類型節點中專門定義為「遺漏」的非字串值。系統遺漏值顯示為 **\$null\$**。請注意, 空字串在 IBM SPSS Modeler 中不被視為空白值, 但它們可能會被某些資料庫視為空白值。
- **空字串和空白。**空字串值和空白 (帶有不可見字元的字串) 不被視為空值。對於大多數用途, 空字串都視為相當於空白。例如, 如果您選取在來源節點或類型節點中將空白視為空白值的選項, 則此設定也套用至空字串。
- **空白值或使用者定義的遺漏值。**這些是在來源節點或類型節點中被明確定義為遺漏的值 (例如 unknown、99 或 -1)。您還可以將空白和空白視為空白值, 這樣將使得它們被旗標為進行特殊處理並排除在大多數計算之外。例如, 您可以使用 @BLANK 函數將這些值以及其他類型的遺漏值處理為空白值。

2. 在**方法**欄中，指定要使用的方法。

下列方法可用於輸入遺漏值：

**固定**。替換為固定值（可以欄位平均數、範圍中間值，或者您指定的常數）。

**隨機**。替換為基於正常分佈或均勻分佈產生的隨機值。

**表示式**。用於指定自訂表示式。例如，您可以使用設定廣域節點建立的廣域變數取代數值。

**演算法**。基於 C&RT 演算法替換為模型預測的值。對於使用此方法輸入的每個欄位，都會有一個個別的 C&RT 模型，還有一個填入節點會使用該模型預測的值取代空值與無效。然後使用過濾器節點刪除該模型產生的預測欄位。

3. 要產生遺漏值 SuperNode，請從功能表中選擇：

#### 產生 > 遺漏值 SuperNode

這將顯示「遺漏值 SuperNode」對話框。

4. 選取**所有欄位**或**僅所選取的欄位**，並根據需要指定樣本大小。（指定的樣本是百分比，預設情況，將對所有記錄取 10% 的樣本。）
5. 按一下**確定**將產生的 SuperNode 新增到串流畫布中。
6. 將 SuperNode 連接到串流以套用轉換。

在 SuperNode 中，將根據情況使用由模型塊、填入器和過濾器節點形成的組合。要瞭解 SuperNode 如何工作，可以編輯 SuperNode 並按一下**放大**，並且可以在 SuperNode 中新增、編輯或刪除特定節點以對行為進行微調。

**處理離群值與極端值：**對於每個欄位，將根據在「資料審核」節點中指定的偵測選項顯示列有離群值與極端值個數的審核報告。請參閱第 275 頁的『資料審核的品質標籤』主題，以取得更多資訊。您可以根據情況選擇控制、放棄特定欄位的這些值或使其取消，然後產生 SuperNode 以套用這些轉換。

1. 在**動作**欄中，根據需要指定對特定欄位的離群值與極端值的處理。

下列動作可用於處理離群值與極端值：

- **強制轉型**。將離群值與極端值取代為不會被視為極端值的最接近值。例如，如果將離群值定義為高於或低於三個標準差的任何值，則會將所有離群值取代為此範圍中的最高值或最低值。
- **捨棄**。捨棄含指定欄位的離群值或極端值的記錄。
- **設為空值**。將離群值與極端值取代為空值或系統遺漏值。
- **強制離群值/捨棄極端值**。只捨棄極端值。
- **強制離群值/使極端值取消**。僅使極值取消。

2. 要產生 SuperNode，請從功能表中選擇：

#### 產生 > 離群值和極值 SuperNode

這將顯示「離群值 SuperNode」對話框。

3. 選取**所有欄位**或**僅所選取的欄位**，然後按一下**確定**以將產生的 SuperNode 新增到串流畫布中。
4. 將 SuperNode 連接到串流以套用轉換。

（選用）您可以編輯 SuperNode 並進行放大以進行瀏覽或進行變更。在 SuperNode 內，根據情況使用一系列「選取」和/或「填入器」節點捨棄、強制或使值取消。

過濾含遺漏資料的欄位： 從資料審核瀏覽器中，可以根據品質分析的結果通過使用「品質」對話框中的「產生過濾器」建立新的「過濾器」節點。

眾數。 為指定的欄位選取所需的作業（包含或排除）。

- 選取的欄位。「過濾器」節點將包含/排除在「品質」標籤上選取的欄位。例如，您可以根據完成 %欄為表格排序，再通過按住 Shift 鍵並按一下來選取完成率最低的欄位，然後產生排除這些欄位的過濾器節點。
- 品質百分比大於指定值的欄位。「過濾器」節點將包含/排除完整紀錄的百分比高於指定臨界值的欄位。預設臨界值為 50%。

過濾空白欄位或無類型欄位

請注意，在將資料值實例化之後，審核結果或 IBM SPSS Modeler 中的大多數其他輸出會將無類型欄位或空白欄位排除。這些欄位在建模時會被忽略，但它們可能會使資料過多或混亂。如果這樣，可以使用資料審核瀏覽器產生「過濾器」節點，以通過該「過濾器」節點從串流中刪除這些欄位。

1. 要確保將所有欄位（併入空白欄位或無類型欄位）併入在審核中，可在上游的來源節點或「類型」節點中按一下清除所有值，或者將所有欄位的值設定為 <Pass>。
2. 在資料審核瀏覽器中，請根據完成 %欄進行排序，選取含零個有效值（或某個其他臨界值）的欄位，並使用「產生」功能表產生可新增到串流中的過濾器節點。

選取含遺漏資料的記錄： 從資料審核瀏覽器中，可以根據品質分析的結果建立新選取節點。

1. 在「資料審核」瀏覽器中，選擇「品質」標籤。
2. 從功能表中選擇：

產生 > 遺漏值選取節點

這將顯示「產生選取節點」對話框。

請選取。指定當記錄有效或無效時是否應儲存這些記錄。

在以下位置查找無效值。指定在何處檢查無效值。

- 所有欄位。選取欄位將檢查所有欄位中是否有無效值。
- 在表格中選取的欄位。選取節點將只檢查目前在「品質」輸出表格中選取的欄位。
- 品質百分比大於指定值的欄位。選取節點將檢查完整紀錄的百分比大於指定臨界值的欄位。預設臨界值為 50%。

在以下位置發現無效值時將記錄視為無效。指定將記錄確定為無效的條件。

- 任何上述欄位。如果上述任何 指定欄位包含某記錄的無效值，則選取節點會將該記錄視為無效。
- 所有上述欄位。如果上述所有 指定欄位都包含某記錄的無效值，則選取節點會將該記錄視為無效。

## 產生其他用於資料預備的節點

在資料預備中使用的各式各樣的節點可直接從資料審核瀏覽器中產生，包含 Reclassify、Binning 和 Derive 節點。例如：

- 您可以根據 *claimvalue* 和 *farmincome* 的值衍生新欄位，方法為：在審核報告中選取這兩者，並從「產生」功能表中選取衍生。這會將新節點新增到串流畫布中。
- 同樣，您可以根據審核結果確定：將 *farmincome* 重新編碼為基於百分位數的 bin 會提供更加集中的分析。要產生分組節點，請在顯示中選擇欄位列，然後從「產生」功能表中選擇分級。

一旦產生節點並將其新增到串流畫布中，必須將它附加到串流中並開啟該節點以指定選取的欄位的選項。

---

## 轉換節點

將輸入欄位正常化是使用傳統評分技術（如迴歸方法、logistic 迴歸方法和判別分析）之前的一個重要步驟。這些技術採用的資料服從常態分佈的假設，對於多數原始資料檔可能不適用。處理現實世界資料的一種方法是：對原始資料元素作轉換，使其更接近常態分佈。此外，可以輕鬆地在常態化欄位之間進行比較例如，收入和年齡在原始資料檔案中有著完全不同的尺度，但常態化後，可以輕鬆地解釋每個尺度的相對影響。

「轉換」節點提供輸出檢視器，使用該輸出檢視器，可以快速而直觀地評量要使用的最佳轉換。您可以快速請參閱變數是否是正常分佈，並在需要時選擇所需的轉換並進行套用。可以選擇多個欄位並針對每個欄位執行一次轉換。

為欄位選取首選的轉換後，可以產生執行這些的轉換的「衍生」節點或「過濾器」節點，並將這些節點附加到串流中。「衍生」節點建立新欄位，而「過濾器」節點轉換現有欄位。請參閱第 282 頁的『產生圖形』主題，以取得更多資訊。

### 轉換節點的欄位標籤

在「欄位」標籤上，可指定要套用資料的哪些欄位並檢視可能的轉換並套用它們。僅能轉換數字欄位。按一下欄位選取器按鈕，並從顯示的清單中選取一個或多個數字欄位。

### 轉換節點的選項標籤

使用「選項」標籤，可以指定要包含的轉換類型。您可以選擇包含所有可用轉換，或個別選擇各個轉換。

在後一種情況下，也可以輸入一個數字以偏移逆模型轉換和對數轉換的資料。如果資料中有很小比例的零，那麼可能會導致平均數和標準差結果有離差，此時作偏移處理將會非常有用。

例如，假設您有一個名為餘額的欄位，該欄位中包含有一些零，並且您要針對該欄位使用逆模型轉換。為避免不期望的偏移，請選取 **Inverse (1/x)** 並在使用資料偏移欄位中輸入 1。（請注意，此偏移與 IBM SPSS Modeler 中的 @OFFSET 序列函數執行的偏移無關。）

**全部公式。** 標識出所有應計算的可用轉換並將其顯示在輸出中。

**選取公式。** 用於選取要計算並顯示在輸出中的不同轉換。

- **取取倒數(1/x)**。指出逆模型轉換應顯示在輸出中。
- **取對數 (log n)**。指出  $\log_n$  轉換應顯示在輸出中。
- **取對數 (log 10)**。指出  $\log_{10}$  轉換應顯示在輸出中。
- **指數**。指出指數轉換 ( $e^x$ ) 應顯示在輸出中。
- **平方根**。指出平方根轉換應顯示在輸出中。

### 轉換節點的輸出標籤

使用「輸出」標籤，可以指定輸出格式和輸出位置。您還可以選擇將結果顯示在畫面上，或將它們傳送到其中一種標準檔案中。請參閱第 267 頁的『輸出節點的「輸出」標籤』主題，以取得更多資訊。

### 轉換節點的輸出檢視器

使用輸出檢視器，可以檢視「轉換」節點的執行結果。該檢視器是一種功能強大的工具，它在轉換的縮圖視圖中顯示每個欄位的多個轉換，從而使您可以快速地比較欄位。您可以使用其「檔案」功能表上的選項來儲存、匯出或列印輸出。請參閱第 265 頁的『檢視輸出』主題，以取得更多資訊。

對於所選轉換以外的每個轉換，會以下面的格式在其下顯示圖註：

平均數（標準差）

## 為轉換產生節點

輸出檢視器為資料準備提供了有用的起點。例如，您可能想要將欄位年齡正常化，以便可以使用採用常態分佈的評分技術（如邏輯迴歸方法或判別分析）。依據初始圖形和彙總統計資料，您可能會決定根據特定分佈（例如，對數分佈）轉換年齡欄位。選取首選分佈後，可以產生使用標準化轉換的衍生節點以用於評分。

可以從輸出檢視器中產生下列欄位作業節點：

- 衍生
- 填充器

「衍生」節點建立含所需轉換的新欄位，而「填入器」節點轉換現有欄位。節點以受的形式放置在畫布上。

如果為不同的欄位選取同一的轉換，那麼「衍生」節點或「填入器」節點為套用該的轉換的所有欄位包含該的轉換類型的公式。例如，假設您選取了下表格中顯示的欄位和轉換來產生「衍生」節點。

表 44. 衍生節點產生的範例。

欄位	變換
AGE	目前分配
INCOME	對數
OPEN_BAL	反向
BALANCE	反向

受中包含下列節點：

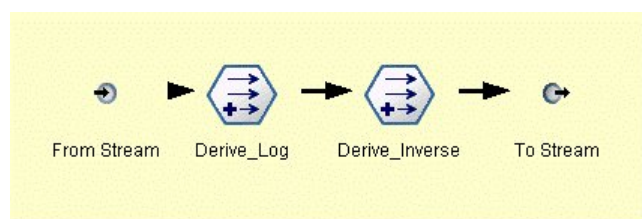


圖 74. 畫布上的受

在此範例中，Derive\_Log 節點具有收入欄位的對數公式，而 Derive\_Inverse 節點具有 OPEN\_BAL 和餘額欄位的逆模型公式。

### 產生節點

1. 對於輸出檢視器中的每個欄位，請選取所需的轉換。
2. 從「產生」功能表中，根據需要選擇衍生節點或填入器節點。

如果這樣做，會相應地顯示「產生衍生節點」或「產生填入節點」對話框。

根據需要選擇非標準化轉換或標準化轉換（z 分）。第二個選項將 z 分套用到轉換；z 分將值代表為與標準差中變數平均數為的差值的函數。例如，如果將對數轉換套用到年齡欄位並選擇標準化轉換，則產生的節點的最終方程式為：

$$(\log(\text{AGE}) - \text{Mean}) / \text{SD}$$

一旦節點產生並顯示在串流畫布上：

1. 將它附加到串流中。
2. 對於受，可以選擇按兩下該節點以檢視它的內容。
3. (選用) 按兩下衍生節點或填入節點以修改選取的欄位的選項。

**產生圖形：** 您可以在輸出檢視器中根據縮略直方圖產生標準大小的直方圖輸出。

**產生圖表**

1. 在輸出檢視器中按兩下縮圖。

或

在輸出檢視器中選取縮圖。

2. 從「產生」功能表中，選擇**圖形輸出**。

執行此操作將顯示該直方圖，並在其上疊放一條常態分佈曲線。這樣，您便可以比較每個可用轉換與常態分佈的相符程度。

注意：僅當建立輸出的「轉換」節點連接至串流時，才能夠產生圖表。

**其他作業：** 從輸出檢視器中，還可以：

- 按「欄位」欄為輸出網格排序。
- 將輸出匯出到 HTML 檔案中。請參閱第 266 頁的『匯出輸出』主題，以取得更多資訊。

---

## 統計資料節點

「統計資料」節點提供與數值欄位相關的基本彙總資訊。您可以取得各個欄位的彙總統計資料以及欄位之間的相關係數。

### 統計資料節點的設定標籤

**檢查。** 選取您需要其個別彙總統計資料的欄位。可以選取多個欄位。

**統計量。** 選取要報告的統計資料。可用選項包含計數、平均數、和、下限、上限值、極差、變異數、標準差、平均數的標準誤、中位數和模式。

**相關。** 選取您希望相關的欄位。可以選取多個欄位。當選取相關係數欄位時，將在輸出中列出每個「檢查」欄位和相關係數欄位之間的相關係數。

**相互關連設定。** 您可以指定用於在輸出中顯示相關係數強度的選項。

### 相互關連設定

IBM SPSS Modeler 可以使用描述性統計量標籤描述相關係數的特徵以說明強調顯示重要關係。**相關係數**測量兩個連續（數值型範圍）欄位之間的關係強度。它的值介於 -1.0 和 1.0 之間。值接近於 +1.0 表示強正相關，因此在兩個欄位之間，大值與大值相關，小值與小值相關。值接近於 -1.0 表示強負相關，因此在兩個欄位之間，大值與小值相關，小值與大值相關。值接近於 0.0 表示弱相關，因此，兩個欄位的值或多或少地彼此獨立。

通過使用「相互關連設定」對話框，您可以控制相關係數標籤的顯示，變更定義種類的臨界值，以及變更改用於每個範圍的標籤。因為刻畫相關係數的方式很大程度上依賴於問題網域，所以您可能需要依據具體狀況來自訂範圍和標籤。

在輸出中顯示相關強度標籤。依預設會選取這個選項。取消選取此選項將在輸出中省略敘述統計標籤。

相關強度。有兩個選項用於定義和標記相關係數強度：

- **按重要性 (1-p) 定義相關強度。**根據重要性標記相關係數，重要性等於 1 減顯著性（即，1 減去平均數的差分完全歸結於機遇變異的機率）。此值越接近於 1，兩個欄位不相依的機率越大，換句話說，它們之間存在某種關係。一般情況下，建議比重要性而不是絕對值標記相關係數，因為重要性考慮了資料的可變性，例如係數 0.6 可能在某個資料集中非常顯著，而在另一個資料集中根本不顯著。依預設，將介於 0.0 和 0.9 之間的重要性值標記為弱，將介於 0.9 和 0.95 之間的重要性值標記為中，將介於 0.95 和 1.0 之間的重要性值標記為強。
- **按絕對值定義相關強度。**如上方所述，根據 Pearson 相關係數（介於 -1 和 1 之間）的絕對值標記相關係數。此測量的絕對值越接近於 1，相關係數就越強。依預設，將介於 0.0 和 0.3333 之間的相關係數（採用絕對值的形式）標記為弱，將介於 0.3333 和 0.6666 之間的相關係數標記為中，將介於 0.6666 和 1.0 之間的相關係數標記為強。但是請注意，要將任何給定值的顯著性從一個資料集擴展到另一個資料集都是非常困難的；因此，在大多數情況下，建議根據機率而不是絕對值定義相關係數。

## 統計資料輸出瀏覽器

統計量節點輸出瀏覽器顯示統計量分析的結果，並且可用於執行作業，包含選取欄位、根據選擇產生新節點，以及儲存和列印結果。「檔案」功能表中提供一般的儲存、匯出及列印選項，「編輯」功能表中提供一般的編輯選項。請參閱第 265 頁的『檢視輸出』主題，以取得更多資訊。

首次瀏覽統計資料輸出時，會展開結果。若要在檢視結果之後將其隱藏，請使用項目左側的展開工具控制項來收合要隱藏的特定結果，或按一下**全部收合**按鈕來收合所有結果。若要在收合結果之後再次查看，請使用項目左側的展開工具控制項來顯示結果，或按一下**全部展開**按鈕來顯示所有結果。

輸出包含每個檢查欄位的一部分，還包含所已要求的統計資料的表格。

- **計數。**具有欄位的有效值的記錄數。
- **平均數。**所有記錄的欄位的平均值。
- **總和。**所有記錄的欄位的值總和。
- **最小。**欄位的最小值。
- **最大。**欄位的最大值。
- **範圍。**最小與最大值之間的差異。
- **變異數。**一種對欄位值可變性的測量。計算方法是：求出每個值和整體平均值之間的差分並平方之，然後對所有的平方值求和，再除以記錄數。
- **標準差。**欄位值的可變性的另一個測量，其值為變異數的平方根。
- **平均數的標準誤差。**一種對欄位平均數估計值的不確定性（如果該平均數套用到新資料）的測量。
- **中位數。**欄位的「中間」值；即，根據欄位值將資料的上半部分與下半部分拆分開啟的值。
- **眾數。**資料中最常見的單一值。

**相關性。**如果指定了相關係數欄位，那麼輸出還包含列出「檢查」欄位和每個相關係數欄位之間的 Pearson 相關係數的部分，以及相關係數值的選用敘述統計標籤。請參閱第 282 頁的『相互關連設定』主題，以取得更多資訊。

「產生」功能表 「產生」功能表包含節點產生作業。

- **過濾。**產生「過濾器」節點以過濾掉與其他欄位無關或相關弱的欄位。

## 根據統計資料產生過濾器節點

從統計資料輸出瀏覽器產生的「過濾器」節點將根據它們與其他欄位的相關性過濾欄位。它的工作方式為：按絕對值的順序為相關係數排序，獲取一些最大的相關係數（根據「統計資料」對話框中的準則集），並建立過濾器（該過濾器傳遞在這些大相關係數中顯示的所有欄位）。

**眾數。** 確定如何選取相關係數。**包含**導致保留在指定的相關係數中顯示的欄位。**排除**導致過濾掉欄位。

**包含/排除**顯示在以下位置的欄位。定義用於選取相關係數的準則。

- **相關係數總數。** 選取指定數量的相關係數並包含/排除在這些相關係數中出現的所有欄位。
- **最大相關係數值的百分比。** 選取指定百分比 ( $n\%$ ) 的相關係數並包含/排除顯示在這些相關係數中出現的所有欄位。
- **大於該值的相關係數。** 選取絕對值大於指定臨界值的相關係數。

---

## 「平均數」節點

「平均數」節點會比較獨立群組之間的平均值，或比較相關欄位的配對之間的平均值，以測試是否存在顯著性差異。例如，您可以將開展促銷前後的營收平均數進行比較，或將從參加促銷的客戶那獲得的營收與從未參加促銷的客戶那獲得的營收進行比較。

您可以根據您的資料以兩種不同的方式比較平均值：

- **在欄位中的群組之間。** 若要比較獨立群組，請選取一個測試欄位和一個分組欄位。例如，您可在開展促銷時排除「拒不參加」客戶樣本，並將「拒不參加」群組的營收平均數與所有其他群組的營收平均數進行比較。在這種情況下，您要指定一個測試欄位（該欄位指出每名客戶的營收），以及一個旗標或列名欄位（該欄位指出他們是否獲得了優惠）。這些樣本是獨立的，原因是：將每條記錄分配給一個群組或另一個群組，並且無法將一個群組的特定會員鏈結到另一個群組的特定會員。也可以指定含兩個以上值的列名欄位以比較多個群組的平均數。當執行該節點時，它會針對選取的欄位進行單因素 ANOVA 檢定。如果只有兩個欄位群組，則單因素 ANOVA 結果與獨立樣本的  $t$  測試本質上一樣。請參閱『比較獨立群組的平均數』主題，以取得更多資訊。
- **成對欄位之間。** 比較兩個相關欄位的平均數時，群組必須以某種方式配對，結果才有意義。例如，可將同一群組客戶在開展促銷前後的營收平均數進行比較，或在夫妻成對之間比較某服務的使用率，以請參閱它們是否不同。每條記錄包含兩個個別但相關的測量，可以對它們進行有意義的比較。當執行該節點時，它針對所選的每個欄位對進行成對樣本  $t$  測試。請參閱『在成對欄位之間比較平均數』主題，以取得更多資訊。

## 比較獨立群組的平均數

在「平均數」節點中選取在欄位中的群組之間以比較兩個或更多個獨立群組的平均數。

**分組欄位。** 選取一個數值旗標或列名欄位，該欄位含兩個或兩個以上的不同值，並且將記錄分為要比較的群組，如獲得優惠的群組和未獲得優惠的群組。無論測試欄位數是多少，都只能選取一個分組欄位。

**測試欄位。** 選取一個或多個包含要測試的測量的數字欄位。對於您選取的每個欄位，將進行個別測試。例如，您可以測試給定促銷對使用情況、營收和買賣的影響。

## 在成對欄位之間比較平均數

在「平均數」節點中選取成對欄位之間以在個別欄位之間比較平均值。這些欄位必須以某種方式相關，結果才有意義，如促銷前後的營收。也可以選取多個欄位成對。

**欄位一。** 選取包含要比較的第一個測量的數值欄位。在前後研究中，該欄位將為「之前」欄位。



欄位二。選取要比較的第二個欄位。

新增。將所選對新增到「測試」欄位對的清單中。

根據需要重複進行欄位選擇以將多個成對新增到該清單中。

相關性設定。使您可以指定用於標註相關性強度的選項。請參閱第 282 頁的『相互關連設定』主題，以取得更多資訊。

### 「平均數」節點選項

使用「選項」標籤可以設定用於將結果標註為重要、邊際或不重要的臨界值  $p$ 。您也可以編輯每個分等級的標籤。重要性可依據百分比尺度進行測量，並且可將重要性定義推廣為 1 減去獲取完全由機遇變異造成的相同或更為極端的結果（如兩個欄位的平均數差分）的機率。例如， $p$  值大於 0.95 表示結果完全歸結於機遇變異的機率少於 5%。

重要性標籤。您可以編輯用於標記輸出中的每個欄位對或群組的標籤。預設標籤為重要、邊際和不重要。

截斷值。指定每個等級的臨界值。通常情況下，大於 0.95 的  $p$  值將歸為重要等級，而小於 0.9 則歸為不重要等級，但可以根據需要調整這些臨界值。

注意：許多節點中都提供了重要性測量。具體計算依賴於節點以及所使用的目標和輸入欄位的類型，但仍舊可以對值進行比較（因為所有值都是根據百分比尺度測量的）。

### 「平均數」節點輸出瀏覽器

平均數輸出瀏覽器以交叉表顯示資料，並且可用於執行標準作業，如一次一列地選取和複製表格，按任何欄進行排序，以及儲存和列印表格。請參閱第 265 頁的『檢視輸出』主題，以取得更多資訊。

表格中的具體資訊依賴於比較的類型（欄位中的群組或個別的欄位）。

依次排序。使您可以根據特定欄對輸出排序。按一下向上或向下箭頭可變更排序的方向。或者，可以按一下任何欄標題以便根據該欄進行排序。（要變更欄中的排序方向，請再次按一下。）

檢視。您可以選擇簡單或進階以控制顯示中的詳細資訊層次。進階視圖包含簡式視圖中的所有資訊，但是還額外提供了詳細資料。

### 比較欄位中的群組的平均數輸出

比較欄位中的群組時，分組欄位的名稱顯示在輸出表格的上方，並且個別為每個群組報告平均數和相關統計資料。該表格為每個測試欄位包含一個個別的列。

下列各欄將會顯示：

- 欄位。列出所選測試欄位的名稱。
- 按群組分類的平均數。顯示分組欄位的每個種類的平均數。例如，可以將獲得特殊優惠的客戶（新促銷）與未獲得特殊優惠的客戶（標準）進行比較。在進階視圖中，還會顯示標準偏差、標準誤以及計數。
- 重要性。顯示重要性值與標籤。請參閱『「平均數」節點選項』主題，以取得更多資訊。

進階輸出

在進階視圖中，還會顯示下列各欄。

- **F 測試**。此測試基於群組之間的變異數與每個群組內的變異數的比例。如果所有群組的平均數相同，則您將預計  $F$  比例接近於 1（因為兩者均是同一總體變異數的估計值）。此比例越大，群組之間的差異就越大，並且存在顯著差別的機率越大。
- **差異**。顯示自由度。

## 比較欄位成對的平均數輸出

比較個別欄位時，輸出表格為每個選取的欄位對包含一列。

- **欄位一/二**。顯示每個對中第一個欄位和第二個欄位的名稱。在進階視圖中，還會顯示標準偏差、標準誤以及計數。
- **平均數一/二**。分別顯示每個欄位的平均數。
- **相關性**。測量兩個連續（數值型範圍）欄位之間的關係強度。值接近於 +1.0 表示強正相關，值接近於 -1.0 表示強負相關。請參閱第 282 頁的『相互關連設定』主題，以取得更多資訊。
- **平均值差異**。顯示兩個欄位平均數之間的差分。
- **重要性**。顯示重要性值與標籤。請參閱第 285 頁的『「平均數」節點選項』主題，以取得更多資訊。

### 進階輸出

進階輸出增加了下列各欄：

**95% 信賴區間**。範圍的下限和上限，對總體中所有具有該樣本量的區間而言，實際平均數落入其中的可能性為 95%。

**T 測試**。通過將平均值差異除以它的標準誤差，可以獲取  $t$  統計資料。此統計資料的絕對值越大，平均數不相同的機率越大。

**差異**。顯示統計資料的自由度。

---

## 報告節點

使用「報告」節點，可以建立包含固定文字以及資料和從該資料衍生的其他表示式的格式報告。通過使用文字範本定義固定文字和資料輸出建構，可以指定報告的格式。通過使用範本中的 HTML 標籤和在「輸出」標籤上設定選項，可以提供自訂文字格式。在使用範本中的 CLEM 表示式的報告中包含有資料值和其他條件輸出。

### 報告節點的替代選項

報告節點最常用於列出串流的記錄或案例輸出，如符合某個特定條件的所有記錄。就此而言，可將「報告」節點視為「表格」節點的結構性較差的替代選項。

- 如果您希望報告列出欄位資訊或在串流而不是資料本身中定義的任何其他內容（如在「類型」節點中指定的欄位定義），那麼可以改用 Script。
- 要產生包含多個輸出物件（如一個或多個串流產生的模型、表格和圖形的集合）並且可以成為採用多種格式（包含文字、HTML 和 Microsoft Word/Office）的輸出的報告，可以使用 IBM SPSS Modeler 專案。
- 要在未使用 Script 的情況下生成欄位名稱清單，可以使用前面帶有「樣本」節點（捨棄所有記錄）的「表格」節點。這會生成一個不含列的表格，該表格可在匯出時已轉置以在單一欄中生成欄位名稱清單。（要這樣做，請在表格節點中的「輸出」標籤上選取調換資料。）

## 報告節點的範本標籤

**建立範本。**要定義報告的內容，請在「報告」節點的「範本」標籤上建立範本。該範本包含數行文字，每一行都指定與報告內容相關的某些資訊，並且用一些特殊標籤行指出內容行的範圍。在每個內容行中，會在將該行發送到報告之前對括在方括弧 ([]) 內的 CLEM 表示式求值。範本中某個行的可能範圍有三個：

**固定。**未標示的行被視為固定行。在對固定行包含的所有表示式求值後，只將這些行向報告複製一次。例如，行

這是我的報告，列印在 [TODAY] 上

將一個行複製到報告中，包含文字和現行日期。

**廣域 (疊代 ALL)。**對於輸入資料的每條記錄，會將包含在特殊標籤 #ALL 與 # 之間的行向報告複製一次。CLEM 表示式 (括在方括弧中) 根據每個輸出行的現行記錄進行求值。例如，行

```
#ALL
For record [INDEX], the value of AGE is [AGE]#
```

將為每個記錄包含一行，指出記錄編號和年齡。

產生所有記錄的清單：

```
#ALL
[Age] [Sex] [Cholesterol] [BP]
#
```

**條件 (疊代 WHERE)。**對於滿足指定條件的每條記錄，會將包含在特殊標籤 #WHERE <condition> 和 # 之間的行向報告複製一次。該條件是指 CLEM 表示式。(在 WHERE 條件中，方括弧是選用的。) 例如，行

```
#WHERE [SEX = 'M'] Male at record no. [INDEX] has age [AGE].#
```

會為每個性別值為 M 的記錄向檔案寫入一行。完整的報告將包含通過將範本套用到輸入資料定義的固定行、廣域行和條件行。

您可以使用各種類型的輸出節點都具備的「輸出」標籤指定用於顯示或儲存結果的選項。請參閱第 267 頁的『輸出節點的「輸出」標籤』主題，以取得更多資訊。

以 HTML 或 XML 格式輸出資料

您可以直接在範本中包含 HTML 或 XML 標籤以使用這兩種格式中的任意一種撰寫報告。例如，下列範本生成 HTML 表格。

This report is written in HTML. Only records where Age is above 60 are included.

```
<HTML>
  <TABLE border="2">
    <TR>
      <TD>Age</TD>
      <TD>BP</TD>
      <TD>Cholesterol</TD>
      <TD>Drug</TD>
    </TR>

    #WHERE Age > 60    <TR>
      <TD>[Age]</TD>
      <TD>[BP]</TD>
      <TD>[Cholesterol]</TD>
      <TD>[Drug]</TD>
```

```
</TR>
#
</TABLE>
</HTML>
```

## 報告節點輸出瀏覽器

報告瀏覽器向您顯示所產生的報告的內容。「檔案」功能表中提供一般的儲存、匯出及列印選項，「編輯」功能表中提供一般的編輯選項。請參閱第 265 頁的『檢視輸出』主題，以取得更多資訊。

---

## 設定廣域節點

設定廣域節點掃描資料並計算可在 CLEM 表示式中使用的彙總值。例如，可以使用「設定廣域」節點為名為 *age* 的欄位計算統計資料，然後通過插入 @GLOBAL\_平均值(*age*) 函數在 CLEM 表示式中使用 *age* 的總平均數。

### 設定廣域節點的設定標籤

所要建立的廣域。選取希望其廣域量可用的欄位。可以選取多個欄位。對於每個欄位，請通過確保在欄位名稱旁邊的欄中已選取所需的統計資料來指定要計算的統計資料。

- **平均值**。所有記錄的欄位的平均值。
- **SUM**。所有記錄的欄位的值總和。
- **MIN**。欄位的最小值。
- **MAX**。欄位的最大值。
- **SDEV**。標準差，它是欄位值的可變性的測量，其值為變異數的平方根。

**預設操作**。在此處選取的選項將在向上方的廣域清單新增新欄位時使用。要變更預設統計資料集，請根據情況選取或取消選取統計資料。也可以套用**套用按鈕**將預設作業套用到清單中的所有欄位。

**註**：某些作業不適用於非數值欄位（例如，「合計」不適用於日期/時間欄位）。無法用於選定欄位的作業將處於停用狀態。

**執行前清除所有廣域量**。選取此選項可在計算新值前刪除所有廣域量。如果不選取此選項，則新計算出的值會取代較舊的值，但未重新計算的廣域量仍保持可用。

**執行後顯示已建立廣域量的預覽**。如果選取此選項，那麼「串流內容」對話框的「廣域量」標籤將在執行後顯示，以顯示計算出的廣域量。

---

## 「模擬配適」節點

「模擬配適」節點將一組候選統計分佈配適到資料中的每個欄位。每個分佈到欄位的擬合將通過擬合度準則進行評估。執行「模擬配適」節點時，將建立一個「模擬產生」節點（或更新現有節點）。將為每個欄位分配其最佳配適分佈。然後，可以使用「模擬產生」節點為每個欄位產生模擬資料。

雖然「模擬配適」節點是一個終端節點，但它不會向已產生的模型選用區新增模型，也不會向輸出標籤新增輸出或圖表或者匯出資料。

**註**：如果歷程資料較為稀疏（即，遺漏值非常多），那麼配適元件可能難以找到足夠多的有效值將分佈配適到資料。對於資料較為稀疏的情況，您應該先移除不需要的稀疏欄位或插補遺漏值，然後再進行配適。通過使用「資料審核」節點的**品質**標籤上的選項，您可以檢視完整紀錄的號碼、識別稀疏欄位並選取插補方法。如果用於分佈配適的記錄數不足，那麼可以使用「平衡」節點來增加記錄數。

使用「模擬配適」節點可自動建立「模擬產生」節點

首次執行「模擬配適」節點時，將使用指向「模擬配適」節點的更新鏈結建立一個「模擬產生」節點。再次執行「模擬配適」節點時，只有在已移除更新鏈結的情況下才會建立新的「模擬產生」節點。另外，「模擬配適」節點還可用於更新已連線的「模擬產生」節點。結果取決於是否在這兩個節點中出現相同的欄位，以及是否在「模擬產生」節點中解除鎖定了這些欄位。請參閱第 45 頁的『「模擬產生」節點』主題，以取得更多資訊。

「模擬配適」節點只能具有一個指向「模擬產生」節點的更新鏈結。要定義指向「模擬產生」節點的更新鏈結，請完成下列步驟：

1. 用滑鼠右鍵按一下「模擬配適」節點。
2. 從功能表中，選取**定義更新鏈結**。
3. 按一下要定義的更新鏈結所指向的「模擬產生」節點。

要移除「模擬配適」節點與「模擬產生」節點之間的更新鏈結，請用滑鼠右鍵按一下該更新鏈結，然後選取**移除鏈結**。

## 分佈配適

統計分佈是某個變數可以使用的值的理論出現頻率。在「模擬配適」節點中，會將一組理論統計分佈與每個資料欄位進行比較。主題第 53 頁的『分佈』中說明了可用於配適的分佈。對理論分佈的參數進行了調整，以便根據擬合度的測量實現與資料的最佳擬合；Anderson-Darling 準則或 Kolmogorov-Smirnov 準則。通過「模擬配適」節點實現的分佈配適的結果顯示配適了哪些分佈、每個分佈的最佳參數估計以及每個分佈與資料的配適度。分佈配適期間，還可以計算具有數值儲存類型的欄位之間的相關性，以及具有種類分佈的欄位之間的偶然性。分佈配適的結果將用於建立「模擬產生」節點。

將任何分佈與資料進行擬合之前，會在前 1000 條記錄中查找遺漏值。如果遺漏值過多，那麼無法進行分佈配適。在這種情況下，您必須確定下列某個選項是否適用：

- 使用上游節點移除包含遺漏值的記錄。
- 使用上游節點針對遺漏值對值進行插補。

分佈配適未排除使用者遺漏值。如果您的資料包含使用者遺漏值，並且您希望從分佈配適中排除這些值，那麼應該將這些值設定為系統遺漏值。

擬合分佈時，將不會考慮欄位的角色。例如，角色為目標的欄位的處理方式與角色為輸入、無、兩者、分割區、分割、頻率和ID的欄位相同。

分佈配適期間，將根據欄位的儲存類型和測量層次以不同方式對這些欄位進行處理。下表說明了分佈配適期間的欄位處理。

表 45. 根據欄位儲存類型及測量層級的分配適合

儲存體類型	測量層級					
	連續的	種類	旗標	名義(M)	序數(O)	無類型
字串	不可能		對種類分佈、 骰子分佈和固 定分佈進行擬 合			
整數						
實數						

表 45. 根據欄位儲存類型及測量層級的分配適合 (繼續)

儲存體類型	測量層級			
時間(M)	對所有分佈進行擬合。將計算相關性和偶然性。	對種類分佈進行擬合。不計算相關性。	對二項分佈、負二項分佈和 Poisson 分佈進行擬合，並計算相關性。	將忽略欄位，並且不會將欄位傳送到「模擬產生」節點。
日期				
時間戳記				
不明	根據資料確定適當的儲存類型。			

對於測量層次為序數的欄位，其處理方式類似於連續欄位，並且它們包含在「模擬產生」節點中的相關性表格內。如果您要將二項分佈、負二項分佈或 Poisson 分佈以外的分佈擬合到序數欄位，那麼必須將欄位的測量層次變更為連續。如果您先前為序數欄位的每個值定義了標籤，並且隨後將測量層次變更為連續，那麼這些標籤將丟失。

在分佈配適到具有多個值的欄位時，將以相同方式處理具有單一值的欄位。具有儲存類型時間、日期或時間戳記的欄位將作為數值進行處理。

#### 將分佈配適到分割欄位

如果您的資料包含分割欄位，並且您希望對每個分割個別執行分佈配適，那麼必須使用上游「重組」節點來轉換資料。使用「重組」節點可以為分割欄位的每個值產生一個新欄位。隨後，可以將此重組資料用於「模擬配適」節點中的分佈配適。

#### 「模擬配適」節點的「設定」標籤

**來源節點名稱。** 通過選取自動，您可以自動產生已產生（或已更新）的「模擬產生」節點的名稱。如果指定了自訂名稱，那麼自動產生的名稱將為「模擬配適」節點中指定的名稱（或者，如果「模擬配適」節點中未指定任何自訂名稱，那麼將為 Sim Gen）。選取自訂可在相鄰的文字欄位中指定自訂名稱。除非對此文字欄位進行編輯，否則預設自訂名為 Sim Gen。

**配適選項** 通過這些選項，您可以指定分佈配適到欄位的方式以及評估分佈配適的方式。

- **要抽樣的觀察值個數。** 此選項指定將分佈配適到資料集中的欄位時要使用的觀察值數。選取**全部觀察值**可將分佈擬合到資料中的所有記錄。資料集非常大時，您可能需要考慮限制用於分佈配適的觀察值數。選取**限制為前 N 個觀察值**可僅使用前 N 個觀察值。按一下箭頭可指定要使用的觀察值數。或者，可以使用上游節點對用於分佈配適的記錄進行隨機抽樣。
- **擬合度準則（僅限連續欄位）。** 對於連續欄位，將分佈配適到欄位時選取配適度 Anderson-Darling 測試或 Kolmogorov-Smirnoff 測試來對分佈進行排等級。依預設將選取 Anderson-Darling 測試，尤其是在您希望確保尾部區域中實現可能的最佳擬合時，建議使用此測試。對於每個候選分佈，計算了這兩種統計資料，但僅使用所選統計資料對分佈進行排序並確定最佳配適分佈。
- **Bin（僅限經驗分佈）。** 對於連續欄位，經驗分佈是歷程資料的累積分佈函數。它是每個值或值範圍的機率，並且直接衍生自資料。通過按一下箭頭，可以指定用於計算連續欄位的經驗分佈的 Bin 個數目。預設值是 100，且最大值是 1000。

- **加權欄位（選用）。**如果資料集包含加權欄位，請按一下欄位選取器圖示並從清單中選取加權欄位。然後，將從分佈配適過程中排除加權欄位。清單顯示了資料集中測量層次為連續的所有欄位。只能選取一個加權欄位。

---

## 「模擬評估」節點

「模擬評估」節點是一個終端節點，用於評估指定欄位、提供該欄位的分配以及生成分佈圖和相關係數圖。此節點主要用於評估連續欄位。因此，它將對評估圖進行補充，該圖由「評估」節點產生並且可用於評估獨立欄位。另一差異在於，「模擬評估」節點跨多個疊代對單一預測進行評估，而「評估」節點對多個預測進行評估，其中每個預測包含單個疊代。在為「模擬產生」節點中的分佈參數指定了多個值的情況下，將產生疊代。請參閱第 52 頁的『疊代』主題，以取得更多資訊。

「模擬評估」節點設計為與從「模擬配適」和「模擬產生」節點中獲取的資料配合使用。但是，此節點可以與任何其他節點配合使用。可以在「模擬產生」節點和「模擬評估」節點之間放置任意號碼的正在處理步驟。

**重要：**「模擬評估」節點至少需要 1000 條具有目標欄位的有效值的記錄。

### 「模擬評估」節點的「設定」標籤

在「模擬評估」節點的「設定」標籤上，您可以指定資料集中每個欄位的角色，並自訂模擬所產生的輸出。

**選取項目。**使您能夠在「模擬評估」節點的以下三個視圖之間進行切換：欄位、密度函數和輸出。

#### 「欄位」視圖

**目標欄位。**這是必要欄位。按一下箭頭可從下拉清單中選取資料集的目標欄位。選取的欄位可以具有連續、次序或名義測量層次，但不能具有日期或未指定的測量層次。

**疊代欄位（選用）。**如果您的資料具有指示資料中每條記錄所屬疊代的疊代欄位，那麼必須在此處選取該欄位。這表示將分別對每次疊代進行評估。只能選取具有連續、次序或名義測量層次的欄位。

**已按疊代對輸入資料進行排序。**僅在疊代欄位（選用）中指定疊代欄位的情況下啟用。只有在您確定輸入資料已按疊代欄位（選用）中指定的疊代欄位進行排序的情況下，才能選取此選項。

**要繪製的上限疊代次數。**僅在疊代欄位（選用）中指定疊代欄位的情況下啟用。按一下箭頭可指定要繪製的疊代的號碼。指定此號碼可避免嘗試在單個圖表上繪製過多的疊代，疊代過多會導致難以解釋繪製的圖形。上限疊代次數可以設定的最低層次為 2；最高層次為 50。要繪製的上限疊代次數最初設定為 10。

**相關性龍捲風的輸入欄位。**相關性龍捲風圖是一個長條圖，它顯示了指定目標與每個指定輸入之間的相關係數。按一下欄位選取器圖示可從可用模擬的輸入的清單中選取要併入在龍捲風圖中的輸入欄位。只能選取具有連續和序數測量層次的輸入欄位。清單中未提供列名、無類型和日期輸入欄位，因此無法選取這些欄位。

#### 「密度函數」視圖

通過此視圖中的選項，您可以針對連續目標自訂機率密度函數和累積分佈函數的輸出，以及針對種類目標自訂預測值的長條圖。

**密度函數。**密度函數是探索模擬結果集合的主要方法。

- **機率密度函數 (PDF)。**選取此選項可針對目標欄位生成機率密度函數。機率密度函數用於顯示目標值的分配。您可以使用機率密度函數來確定目標位於特定區域中的機率。對於種類目標（具有列名或序數測量層次的目標），將產生一個長條圖，用於顯示落入每個目標種類內的觀察值的百分比。

- **累積分佈函數 (CDF)**。選取此選項可針對目標欄位生成累積分佈函數。累積分配函數會顯示目標值小於或等於指定值的機率。僅限連續目標使用。

**參考線 (連續)**。只有在選取了**機率密度函數 (PDF)** 和/或**累加分佈函數 (CDF)** 的情況下，才會啟用這些選項。通過這些選項，您可以向機率密度函數和累積分佈函數新增多條固定的垂直參考線。

- **平均數**。選取此選項可在目標欄位的平均數處新增參照線。
- **中位數**。選取此選項可在目標欄位的中值處新增參照線。
- **標準差**。選取此選項可在目標欄位均數加號上或減去指定標準差數目的位置新增參照行。選取此選項會啟用貼近的數字欄位。按一下箭頭可指定標準差的值。標準差的最小值為 1，最大值為 10。標準差數一啟動設為 3。
- **百分位數**。選取此選項可在目標欄位分佈的兩個百分位值處新增參考線。選取此選項將啟用相鄰的**後幾個值**和**前幾個值**文字欄位。例如，在**前幾個值**文字欄位中輸入值 90 將在目標的第 90 百分位數處新增參照線，90% 的觀察將落入該值以下方。同樣，**後幾個值**文字欄位中的值 10 代表目標的第 10 百分位數，10% 的觀察將落入該值以下方。
- **自訂參考線**。選取此選項可沿水平軸在指定值處新增參照行。選取此選項會啟用貼近的**值表格**。每一個在值表格中輸入一個有效數字時，都會向表格的底部附加一個新的空白列。有效數字是在目標欄位的值範圍內的數字

註：如果在單個圖表上顯示了多個密度函數或分佈函數（來自多個疊代），那麼將分別對每個函數套用參照行（而不是自訂行）。

**種類目標 (僅限 PDF)**。只有在選取了**機率密度函數 (PDF)** 的情況下，才會啟用這些選項。

- **要報告的類別值**。對於具有種類目標欄位的模型，此模型的結果是目標值落入每個種類的一組預測機率（每個種類對應一個機率）。具有最高機率的種類將作為預測種類，並用於產生機率密度函數的長條圖。選取**預測種類**可產生長條圖。選取**預測機率**可針對目標欄位的每個種類產生預測機率的分配直方圖。您也可以選取兩者來產生這兩種類型的圖表。
- **敏感度分析分組**。包含敏感度分析疊代的模擬將針對分析所定義的每個疊代產生一個獨立的目標欄位（或模型中的預測目標欄位）。對於各式各樣的分佈參數的每個值，存在一個疊代。如果出現疊代，那麼種類目標欄位的預測種類長條圖將顯示為一個包含所有疊代結果的複式長條圖。請選取**將種類群組到一起**或**將疊代分組在一起**。

「輸出」視圖

**目標分佈的百分位數值**。通過這些選項，您可以選擇建立目標分佈的百分位值表格，並指定要顯示的百分位數。

**建立百分位數值表**。對於連續目標欄位，選取此選項可取得目標分佈的指定百分位數表格。請選擇下列其中一個選項來指定百分位數：

- **四分位數**。四分位數是指目標欄位分佈的第 25 百分位數、第 50 百分位數和第 75 百分位數。觀察被分為大小相等的四個群組。
- **區間**。如果需要號碼相等的群組（除 4 以外），請選取**區間**。選取此選項會啟用貼近的數字欄位。按一下箭頭可指定區間數目。區間的最小值為 2，最大值為 100。區間的值最初設定為 10。
- **自訂百分位數**。選取自訂百分位數可指定各個百分位數，例如第 99 百分位數。選取此選項會啟用貼近的值表格。每一個在值表格中輸入一個有效數字（介於 1 到 100 之間）時，都會向表格的底部附加一個新的空白列。



## 「模擬評估」節點輸出

執行「模擬評估」節點時，輸出將新增到輸出管理器中。「模擬評估」輸出瀏覽器顯示了執行「模擬評估」節點的結果。檔案功能表中提供了常用的儲存、匯出和列印選項，而編輯功能表中提供了常用的編輯選項。請參閱第 265 頁的『檢視輸出』主題，以取得更多資訊。只有在選取了其中一個圖表的情況下，才會啟用視圖功能表。對於分佈表格或資訊輸出，未啟用此菜單。您可以從視圖功能表中選取編輯模式來變更圖表的佈置和外觀，或者選取探索方式來探索圖表所代表的資料和值。靜態方式會將圖表參考線固定在其現行位置，因此無法移動這些參考線。只有在靜態方式下才能複製、列印或匯出帶有參考線的圖表。要選取此方式，請在視圖功能表中按一下靜態方式。

「模擬評估」輸出瀏覽器視窗包含兩個畫面。導覽窗格位於視窗左側，用於顯示執行「模擬評估」節點時產生的圖表的縮圖代表。選取縮圖後，圖表輸出將顯示在視窗右側的窗格中。

## 導覽畫面

輸出瀏覽器的導覽畫面包含根據模擬產生的圖表的縮圖。導覽畫面上顯示的縮圖取決於目標欄位的測量層次，以及在「模擬評估」節點對話框中選取的選項。下表格中提供了對這些縮圖說明。

表 46. 導覽畫面縮圖

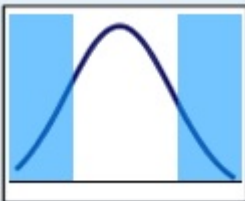
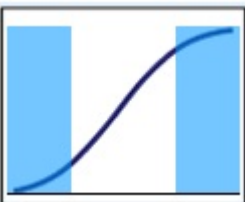
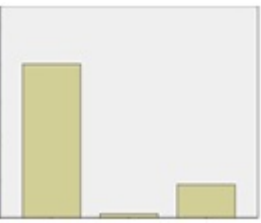
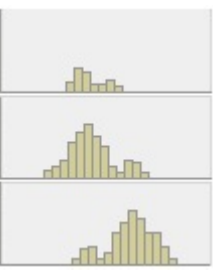
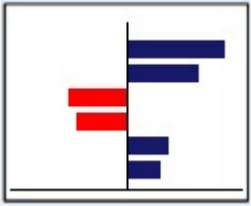
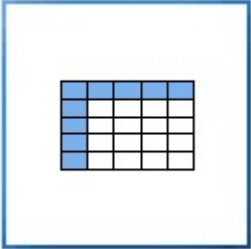

縮圖	說明	備註(O)
	機率密度函數	只有在目標欄位的測量層次為連續，並且在「模擬評估」節點對話框的「密度函數」視圖中選取了機率密度函數 (PDF) 的情況下，才會顯示此縮圖。 如果目標欄位的測量層次是類別，則不會顯示此縮圖。
	累積分佈函數	只有在目標欄位的測量層次為連續，並且在「模擬評估」節點對話框的「密度函數」視圖中選取了累加分佈函數 (CDF) 的情況下，才會顯示此縮圖。 如果目標欄位的測量層次是類別，則不會顯示此縮圖。
	預測的種類值	只有在目標欄位的測量層次為種類，在「模擬評估」節點對話框的「密度函數」視圖中選取了機率密度函數 (PDF)，以及在要報告的種類值區域中選取了預測種類或兩者的情況下，才會顯示此縮圖。 如果目標欄位的測量層級是連續的，則不會顯示此縮圖。
	預測種類機率	只有在目標欄位的測量層次為種類，在「模擬評估」節點對話框的「密度函數」視圖中選取了機率密度函數 (PDF)，以及在要報告的種類值區域中選取了預測機率或兩者的情況下，才會顯示此縮圖。 如果目標欄位的測量層級是連續的，則不會顯示此縮圖。

表 46. 導覽畫面縮圖 (繼續)

縮圖	說明	備註(O)
	龍捲風圖	只有在「模擬評估」節點對話框的「欄位」視圖中的相關性龍捲風的輸入欄位欄位中選取了一個或多個輸入欄位的情況下，才會顯示此縮圖。
	分佈表格	只有在目標欄位的測量層次為連續，並且在「模擬評估」節點對話框的「輸出」視圖中選取了建立百分位值表格的情況下，才會顯示此縮圖。對於此圖表，視圖功能表處於停用狀態。  如果目標欄位的測量層次是類別，則不會顯示此縮圖。
	資訊	始終顯示此縮圖。對於此輸出，視圖功能表處於停用狀態。

## 圖表輸出

可用的輸出圖表的類型取決於目標欄位的測量層次、是否使用了疊代欄位以及在「模擬評估」節點對話框中選取的選項。根據模擬產生的很多圖表都具有互動式功能，您可以使用此功能自訂顯示。按一下圖表選項可以使用互動式功能。所有模擬圖表都是以圖表板視覺化呈現。

**連續目標的機率密度函數圖。** 此圖表同時顯示機率和頻率，機率尺度位於左側的垂直軸上，而頻率尺度位於右側的垂直軸上。此圖表具有兩條滑動垂直參考線，這些參考線將該圖表劃分為一些個別的區域。此圖表下方的表格顯示了每個區域中的分佈所佔的百分比。如果同一圖表上顯示了多個密度函數（由於疊代），那麼此表格將為與每個密度函數關聯的機率提供個別的一列，並提供額外的一欄，用於顯示疊代名稱以及與每個密度函數關聯的顏色。疊代將根據疊代標籤字母的列示在表格中。如果沒有可用的疊代標籤，則會使用疊代值。此表格無法進行編輯。

每條參考線都有一個調節器（倒三角形），您可以使用該調節器輕鬆移動此參考線。每個調節器都有一個標籤，用於指示其現行位置。依預設，調節器位於分佈的第 5 百分位數和第 95 百分位數處。如果存在多個疊代，那麼調節器將位於表格中列出的第一個疊代的第 5 百分位數和第 95 百分位數處。您不能使這兩條行穿過對方。

通過按一下圖表選項，可以使用一些附加功能。特別是，您可以明確地設定調節器位置、新增固定的參考線，以及將圖表視圖由連續曲線變更為直方圖。請參閱第 295 頁的『圖表選項』主題，以取得更多資訊。用滑鼠右鍵按一下圖表可複製或匯出該圖表。

**連續目標的累積分配函數圖。** 此圖表具有兩個可移動的垂直參考線和關聯表格，這些參考線和表格與針對機率密度函數圖說明的參考線和表格相同。存在多個疊代時，調節器控制項和表格的行為與機率密度函數相同。用於識別哪個密度函數的每個疊代的同一顏色也用於分佈函數。

此圖表還提供了對「圖表選項」對話框的存取，通過該對話框您可以明確地設定調節器位置、新增固定的參考線以及指定累積分佈函數是顯示為遞增函數（預設值）還是遞減函數。請參閱『圖表選項』主題，以取得更多資訊。用滑鼠右鍵按一下圖表來複製、匯出或編輯圖表。選取編輯會在浮動的圖表板編輯器視窗中開啟圖表。

**種類目標的預測的種類值圖。**對於種類目標欄位，長條圖顯示了預測值。預測值顯示為預測落入每個種類的目標欄位的百分比。對於具有敏感度分析疊代的種類目標欄位，預測目標種類的結果將顯示為包含所有疊代的結果的複式長條圖。此圖表按種類或按疊代進行叢集，具體取決於在「模擬評估」節點對話框的「密度函數」視圖中的**敏感度分析分組**區域內選取了哪個選項。用滑鼠右鍵按一下圖表來複製、匯出或編輯圖表。選取編輯會在浮動的圖表板編輯器視窗中開啟圖表。

**種類目標的預測種類機率圖。**對於種類目標欄位，直方圖將顯示目標的每個種類的預測機率分佈。對於具有敏感度分析疊代的種類目標欄位，將按種類或按疊代顯示直方圖，具體取決於在「模擬評估」節點對話框的「密度函數」視圖中的**敏感度分析分組**區域內選取了哪個選項。如果直方圖按種類進行群組，那麼您可以通過包含疊代標籤的下拉清單來選擇要顯示的疊代。另外，也可以通過用滑鼠右鍵按一下圖表並從疊代子功能表中選取疊代來選取要顯示的疊代。如果直方圖按疊代進行群組，那麼您可以通過包含種類名稱的下拉清單來選擇要顯示的種類。另外，也可以通過用滑鼠右鍵按一下圖表並從**種類子功能表**中選取種類來選取要顯示的種類。

此圖表僅可用於部分模型，並且必須在模型塊上已選取產生所有群組機率的選項。例如，在 Logistic 模型塊上，您必須選中**附加所有機率**。下列模型塊支援此選項：

- Logistic、SVM、貝葉斯、類神經網路和 KNN
- 用於羅吉斯迴歸、決策樹狀結構及 naïve 貝氏的 Db2/ISW 資料庫內採礦模型

依預設，未在這些模型塊上已選取產生所有群組機率的選項。

**龍捲風圖。**龍捲風圖是一個長條圖，它顯示了目標欄位對每個指定輸入的敏感度。敏感度通過目標與每個輸入之間的相關性進行測量。圖表標題包含目標欄位的名稱。圖表上的每個條形都代表目標欄位與輸入欄位之間的相關性。該圖表中包含的模擬的輸入是在「模擬評估」節點對話框的「欄位」視圖中的**相關性龍捲風**的輸入欄位欄位中選取的輸入。每個條形都通過相關性值進行標註。這些長條圖按相關性絕對值從大到小的順序排序。如果存在疊代，那麼將針對每個疊代產生一個個別的圖表。每個圖表都具有一個次標題，其中包含疊代的名稱。

**分佈表格。**此表格包含目標欄位的值，指定百分比的觀察將落入該值以下方。此表格為「模擬評估」節點對話框的「輸出」視圖中指定的每個百分位數值提供了一列。百分位值可以是四分位數、空間相等的其他百分位數或分別指定的百分位數。分佈表格為每個疊代提供了一欄。

**資訊。**此部分提供在評估中使用的欄位和記錄的整體摘要。它還顯示了輸入欄位和記錄計數，可以針對每個疊代細分這些欄位和計數。

## 圖表選項

您可以在「圖表選項」對話框中自訂根據模擬產生的機率密度函數和累積分佈函數的活動圖表顯示。

**檢視。**視圖下拉清單僅適用於機率密度函數圖。您可以使用此列表將圖表視圖由連續曲線切換為直方圖。在同一圖表上顯示了多個密度函數（來自多個疊代）的情況下，此功能將處於停用狀態。如果存在多個密度函數，那麼只能以連續曲線形式檢視這些密度函數。

**順序。**順序下拉清單僅適用於累積分佈函數圖。該清單會指定累積分配函數是否顯示為遞增函數（預設）或遞減函數。如果顯示為遞減函數，那麼水平軸上指定點處的函數值是目標欄位該點右側的機率。

**滑塊位置。**上限文字欄位包含右滑動參照行的現行位置。下限文字欄位包含左滑動參照線的目前位置。通過在上限和下限文字欄位中輸入值，您可以明確地設定調節器的位置。下限文字欄位中的值必須嚴格少於上限文字欄位中的值。可以選取**負無限**來移除左參照線，從而將位置有效地設定為負無限。此動作將取消下限文字欄

位。可以選取無限來移除右參照線，從而將其位置有效地設定為無限。此動作將取消上限文字欄位。您不能同時移除這兩條參考線；選取負無限將取消無限勾選框，反之亦然。

**參考線。** 您可以將各種固定垂直參照行新增到機率密度函數和累積分配函數。

- **平均數。** 可以在目標欄位的均數處新增參照線。
- **中位數。** 可以在目標欄位的中位數處新增參照線。
- **標準差。** 可以在目標欄位均數加號上或減去指定標準差數目的位置新增參考線。您可以在相鄰的文字欄位中輸入要使用的標準差值。標準差的最小值為 1，最大值為 10。標準差數一啟動設為 3。
- **百分位數。** 通過在後幾個值和前幾個值文字欄位中輸入值，您可以在目標欄位分佈的一個或兩個百分位值處新增參考線。例如，前幾個值文字欄位中的值 95 代表第 95 百分位數，95% 的觀察將落入該值以下方。同樣，後幾個值文字欄位中的值 5 代表第 5 百分位數，5% 的觀察將落入該值以下方。對於後幾個值文字欄位，下限百分位數值為 0，上限百分位數值為 49。對於前幾個值文字欄位，下限百分位數值為 50，上限百分位數值為 100。
- **自訂位置。** 您可以在水平軸上的指定值新增參考線。通過從網格中刪除該項目，您可以移除自訂參照行。

在您按一下**確定**後，調節器、調節器上方的標籤、參考線以及圖表下方的表格將進行更新，以反映「圖表選項」對話框中選取的選項。按一下**取消**可以在不進行任何變更的情況下關閉此對話框。可以通過取消選取「圖表選項」對話框中的關聯選項並按一下**確定**來移除參考線。

**註：**如果在單個圖表上顯示了多個密度函數或分佈函數（由於結果來自敏感度分析疊代），那麼將分別對每個函數應用參考線（而不是自訂行）。將只顯示第一個疊代的參考線。參照線標籤包含疊代標籤。疊代標籤衍生自上游，通常衍生自「模擬產生」節點。如果沒有可用的疊代標籤，則會使用疊代值。對於具有多個疊代的累積分佈函數，均數、中位數、標準差和百分位數選項處於停用狀態。

---

## 延伸輸出節點

如果在「延伸輸出」節點對話框的輸出標籤上選取**輸出至畫面**，則畫面上輸出會顯示在輸出瀏覽器視窗中。輸出也會新增至「輸出」管理程式。輸出瀏覽器視窗具有其自己的一組功能表，可讓您列印或儲存輸出，或者以另一種格式匯出輸出。編輯功能表僅包含複製選項。「延伸輸出」節點的輸出瀏覽器具有兩個標籤；顯示文字輸出的**文字輸出**標籤，以及顯示圖形與圖表的**圖形輸出**標籤。

如果在「延伸輸出」節點對話框的輸出標籤上選取**輸出至檔案**，則在順利執行「延伸輸出」節點之後不會顯示輸出瀏覽器視窗。

### 延伸輸出節點 - 語法標籤

選取語法的類型 - **R** 或 **Python for Spark**。如需相關資訊，請參閱下列章節。當您的語法備妥時，可以按一下**執行**，以執行「延伸輸出」節點。輸出物件會新增至「輸出」管理程式，或者選擇性地新增至輸出標籤上檔名欄位中指定的檔案。

### R 語法

**R 語法。** 您可以在此欄位中輸入或貼上用於分析資料的自訂 R Scripting 語法。

**轉換旗標欄位。** 指定如何處理旗標欄位。有兩個選項：**字串至因數**，**整數與實數至倍精準數**，以及**邏輯值 (True、False)**。如果您選取**邏輯值 (True、False)**，則會遺失旗標欄位的原始值。例如，如果欄位具有值「男性」與「女性」，則這些值會變更為 True 與 False。

將遺漏值轉換為 R「無法使用」值 (NA)。當選取時，會將所有遺漏值轉換為 R NA 值。R 使用值 NA 來識別遺漏值。您使用的部分 R 函數可能具有一個引數，可用來控制資料包含 NA 時的函數行為。例如，函數可讓您選擇自動排除包含 NA 的記錄。如果未選取此選項，則任何遺漏值都會依現狀傳遞至 R，並可能在執行 R Script 時造成錯誤。

將日期/時間欄位轉換為具有時區特殊控制項的 R 類別。選取此選項時，具有日期或日期時間格式的變數會轉換為 R 日期/時間物件。您必須選取下列其中一個選項：

- **R POSIXct**。具有日期或日期時間格式的變數轉換為 R POSIXct 物件。
- **R POSIXlt** (清單)。具有日期或日期時間格式的變數轉換為 R POSIXlt 物件。

註：POSIX 格式是進階選項。僅在您的 R Script 指定以需要這些格式的方式處理日期時間欄位時，才使用這些選項。POSIX 格式不適用於具有時間格式的變數。

## Python 語法

**Python 語法**。您可以在此欄位中輸入或貼上供資料分析使用的自訂 Python Scripting 語法。如需 Python for Spark 的相關資訊，請參閱 Python for Spark 及 Python for Spark Scripting。

### 延伸輸出節點 - 主控台輸出標籤

主控台輸出標籤包含「語法」標籤上的 R Script 或 Python for Spark Script 執行時接收的所有輸出（例如，如果使用 R Script，它顯示執行語法標籤上 R 語法欄位中的 R Script 時，從 R 主控台接收的輸出）。此輸出可能包括執行 R 或 Python Script 時產生的 R 或 Python 錯誤訊息或警告。輸出可以主要用於 Script 除錯。主控台輸出標籤也包含 R 語法或 Python 語法欄位中的 Script。

每次執行延伸輸出 Script 時，都會使用從 R 主控台或 Python for Spark 接收的輸出來改寫主控台輸出標籤的內容。無法編輯輸出。

### 延伸輸出節點 - 輸出標籤

**輸出名稱**。指定執行節點時產生的輸出名稱。當選取自動時，根據 Script 類型，輸出的名稱會自動設為 "R Output" 或 "Python Output"。您可以選擇性地選取自訂以指定不同的名稱。

**輸出至畫面**。選取此選項可在新視窗中產生並顯示輸出。輸出也會新增至「輸出」管理程式。

**輸出至檔案**。選取此選項可將輸出儲存至檔案。執行此動作會啟用輸出圖形與輸出檔圓鈕。

**輸出圖形**。僅當選取了輸出至檔案時才啟用。選取此選項可將執行「延伸輸出」節點產生的所有圖形儲存至檔案。在檔名欄位中指定用於所產生輸出的檔名。按一下省略符號按鈕 (...) 來選擇特定檔案與位置。在檔案類型下拉清單中指定檔案類型。下列檔案類型可用：

- 輸出物件 (.cou)
- HTML (.html)

**輸出文字**。僅當選取了輸出至檔案時才啟用。選取此選項可將執行「延伸輸出」節點產生的所有文字輸出儲存至檔案。在檔名欄位中指定用於所產生輸出的檔名。按一下省略符號按鈕 (...) 來指定特定檔案與位置。在檔案類型下拉清單中指定檔案類型。下列檔案類型可用：

- HTML (.html)
- 輸出物件 (.cou)
- 文字文件 (.txt)

## 延伸輸出瀏覽器

如果在「延伸輸出」節點對話框的輸出標籤上選取輸出至畫面，則畫面上輸出會顯示在輸出瀏覽器視窗中。輸出也會新增至「輸出」管理程式。輸出瀏覽器視窗具有其自己的一組功能表，可讓您列印或儲存輸出，或者以另一種格式匯出輸出。編輯功能表僅包含複製選項。「延伸輸出」節點的輸出瀏覽器具有兩個標籤：

- 文字輸出標籤會顯示文字輸出
- 圖形輸出標籤會顯示圖形及圖表

如果在「延伸輸出」節點對話框的輸出標籤上選取輸出至檔案（而非輸出至畫面），則在順利執行「延伸輸出」節點之後不會顯示輸出瀏覽器視窗。

### 延伸輸出瀏覽器 - 文字輸出標籤

文字輸出標籤會顯示當執行「延伸輸出」節點之語法標籤上的 R Script 或 Python for Spark Script 時，產生的任何文字輸出。

註：透過執行延伸輸出 Script 產生的 R 或 Python for Spark 錯誤訊息或警告都一律顯示在「延伸輸出」節點的主控台輸出標籤上。

### 延伸輸出瀏覽器 - 圖形輸出標籤

圖形輸出標籤會顯示當執行「延伸輸出」節點之語法標籤上的 R Script 或 Python for Spark Script 時，產生的任何圖形或圖表。例如，如果您的 R Script 包含針對 R plot 函數的呼叫，則產生的圖形會顯示在此標籤上。

---

## IBM SPSS Statistics 說明應用程式

如果在您的電腦上安裝並軟體使用權了 IBM SPSS Statistics 的相容版本，則可以使用 Statistics 「轉換」、Statistics 「模型」、Statistics 「輸出」或 Statistics 「匯出」節點將 IBM SPSS Modeler 配置為使用 IBM SPSS Statistics 功能來處理資料。

有關 IBM SPSS Modeler 目前版本的產品相容性的資訊，請存取公司支援網站 (<http://www.ibm.com/support>)。

要配置 IBM SPSS Modeler 以便與 IBM SPSS Statistics 和其他應用程式一起使用，選擇：

### 工具 > 選項 > Helper 應用程式

**IBM SPSS Statistics 交互。**輸入直接在「Statistics 匯出」節點生成的資料檔案上啟動 IBM SPSS Statistics 時要使用的指令的完整路徑和名稱（例如，*C:\Program Files\IBM\SPSS\Statistics\<nn>\stats.exe*）。請參閱第 315 頁的『Statistics 匯出節點』主題，以取得更多資訊。

**連線。**如果 IBM SPSS Statistics Server 位於 IBM SPSS Modeler Server 所在的主機，那麼可以在兩個應用程式之間啟用一個連線，以便在分析過程中將資料保留在伺服器上以增加效率。選取**伺服器**以啟用下方列埠選項。預設設定為 **本端**。

埠。為 IBM SPSS Statistics Server 指定伺服器埠。

**IBM SPSS Statistics 位置公用程式。**若要啟用 IBM SPSS Modeler 使用 Statistics 「轉換」、Statistics 「模型」及 Statistics 「輸出」節點，您必須在執行串流的電腦上，已安裝並授權 IBM SPSS Statistics 的副本。

- 如果以本端（獨立式）方式執行 IBM SPSS Modeler，那麼 IBM SPSS Statistics 的軟體使用權副本必須位於本端電腦上。按一下該按鈕以指定想要用於軟體使用權的本端 IBM SPSS Statistics 安裝的位置。

- 此外，如果是針對遠端 IBM SPSS Modeler Server 以分散式方式執行，那麼還需要在 IBM SPSS Modeler Server 主機上執行中一個公用程式來建立 *statistics.ini* 檔案，此檔案向 IBM SPSS Statistics 指出 IBM SPSS Modeler Server 的安裝路徑。若要這樣做，請從命令提示字元，切換至 IBM SPSS Modeler Server *bin* 目錄，對於 Windows，執行：

```
statisticsutility -location=<IBM SPSS Statistics_installation_path>/
```

或者，對於 UNIX，執行：

```
./statisticsutility -location=<IBM SPSS Statistics_installation_path>/bin
```

如果在本端機器上沒有 IBM SPSS Statistics 的已授權副本，則您仍可以針對 IBM SPSS Statistics 伺服器執行 Statistics 「檔案」節點，但嘗試執行其他 IBM SPSS Statistics 節點會顯示錯誤訊息。

#### 備註(O)

如果在執行 IBM SPSS Statistics 程序節點時遇到困難，可考慮下列提示：

- 如果 IBM SPSS Modeler 中使用的欄位名稱超過八個字元（對於 IBM SPSS Statistics 12.0 之前的版本），或 64 個字元（對於 IBM SPSS Statistics 12.0 及之後的版本），或者欄位名稱中包含無效字元，則在將這些欄位名稱讀取入到 IBM SPSS Statistics 之前有必要將其更名或截斷。請參閱第 316 頁的『重新命名或過濾 IBM SPSS Statistics 的欄位』主題，以取得更多資訊。
- 如果在 IBM SPSS Modeler 之後安裝了 IBM SPSS Statistics，那麼可能需要指定 IBM SPSS Statistics 位置（如上方所述）。





---

## 第 7 章 匯出節點

---

### 匯出節點概觀

「匯出」節點提供一種匯出各種格式的資料以與其他軟體工具相互作用的機制。

可用的匯出節點有：



「資料庫匯出」節點將資料寫入與 ODBC 相容的關係資料來源。要將資料寫入 ODBC 資料來源，資料來源必須存在且您必須對其具有寫入權。



純文字檔案匯出節點將資料輸出到已分隔的文字檔案。這對匯出可由其他分析或試算表軟體讀取的資料非常有用。



Statistics 匯出節點以 IBM SPSS Statistics *.sav* 或 *.zsav* 格式輸出資料。IBM SPSS Statistics Base 和其他產品可以讀取 *.sav* 或 *.zsav* 檔案。這種格式也用於 IBM SPSS Modeler 中的某些快取檔。



Data Collection 匯出節點以 Data Collection 市場調查軟體使用的格式輸出資料。必須安裝 Data Collection 資料程式庫才可使用此節點。



IBM Cognos 匯出節點以 Cognos 資料庫可以讀取的格式匯出資料。



IBM Cognos TM1 匯出節點以 Cognos TM1 資料庫可以讀取的格式匯出資料。



「SAS 匯出」節點可以 SAS 格式輸出資料，以便讀取入 SAS 或與 SAS 相容的套裝軟體中。有三種可用的 SAS 檔案格式：SAS for Windows/OS2、SAS for UNIX 或 SAS Version 7/8。



Excel 匯出節點以 Microsoft Excel .xlsx 檔案格式輸出資料。您還選用擇在執行完此節點後自動啟動 Excel 並開啟匯出的檔案。



XML 匯出節點將資料以 XML 格式輸出到檔案。您可以選擇性地建立 XML 來源節點以將匯出的資料讀回串流。

---

## 資料庫匯出節點

可使用「資料庫」節點將資料寫入與 ODBC 相容的關聯式資料來源，請參閱「資料庫」來源節點相關說明。請參閱第 15 頁的『資料庫來源節點』主題，以取得更多資訊。

請使用下列常用步驟將資料寫入資料庫：

1. 安裝 ODBC 驅動程式並向您要使用的資料庫配置資料來源。
2. 在「資料庫」節點的「匯出」標籤中，指定要寫入的資料來源和表格。可建立新表格或將資料插入現有表格。
3. 根據需要指定其他選項。

在下面的幾個主題中將對這些步驟進行更詳細地說明。

### 資料庫節點的「匯出」標籤

註：您可以匯出到的某些資料庫可能不支援表格中長度超過 30 個字元的直欄名稱。如果顯示提醒您的表格有不正確的直欄名稱的錯誤訊息，請將名稱的大小減小至少於 30 個字元。

**資料來源。** 顯示所選資料來源。輸入名稱或者從下拉清單中選取名稱。如果清單中未顯示所需的資料庫，則選取新增新的資料庫連線並從「資料庫連線」對話框選定資料庫。請參閱第 16 頁的『新增資料庫連線』主題，以取得更多資訊。

**表格名稱。** 輸入接收資料的表格名稱。如果選取插入到表格選項，則可以通過按一下選取按鈕在資料庫中選取一個現有表格。

**建立表格。** 選中此項可建立一個新的資料庫表格或改寫現有的資料庫表格。

**插入到表格。** 選中此項可將資料作為新列插入現有的資料庫表格中。

**合併表格。**（如可用）選取此選項以使用相應來源資料欄位中的值更新所選資料庫直欄。選取此選項會啟用合併按鈕，在其顯示的對話框中您可以將來源資料欄位對映到資料庫直欄。

**刪除現有表格。** 選中此項可在建立新表格時刪除所有名稱相同的現有表格。

**刪除現有列。** 選取此選項可在插入表格時先將現有列從表格中刪除然後匯出。

註：如果選取上述任意兩個選項，那麼在執行節點時將收到**改寫警告**訊息。要想不顯示此警告，請取消選取「使用者選項」對話框的「通知」標籤上的**當節點改寫資料表時發出警告**選項。

**預設字串大小。** 上游「類型」節點中標示為無類型的欄位將作為字串欄位寫入資料庫。請指定無類型欄位要使用的字串大小。

按一下**綱目**可開啟一個對話框，您可在其中設定各種匯出選項（對於支援此功能的資料庫）、設定所需欄位的 SQL 資料類型，並指定創建資料庫編製索引的主要關鍵字。請參閱第 304 頁的『資料庫匯出綱目選項』主題，以取得更多資訊。

按一下**編製索引**可指定創建匯出表格編製索引的選項，以提高資料庫的運行效能。請參閱第 306 頁的『資料庫匯出指標選項』主題，以取得更多資訊。

按一下**進階**可指定批量載入和資料庫提交選項。請參閱第 307 頁的『資料庫匯出進階選項』主題，以取得更多資訊。

**用引號括住表格與直欄名稱。** 選取將 CREATE TABLE 陳述式傳送到資料庫時使用的選項。必須為包含空格和非標準字元的表格和欄添加引號。

- **根據需要。** 選取此選項，以容許 IBM SPSS Modeler 自動根據各個情況確定是否需要添加引號。
- **一律。** 選中此項將始終為表格名和直欄名稱添加參照。
- **從不。** 選中此項將取消參照。

**為此資料產生匯入節點。** 選中此項可在將資料匯出到指定資料來源和表格時產生此資料的資料庫來源節點。執行此操作後，此節點即被新增到串流畫布。

## 資料庫匯出合併選項

此對話框使您能夠將欄位從來源資料對映到目標資料庫表格中的欄。其中來源資料欄位被對映至資料庫直欄，在執行串流時欄值被取代為來源資料值。未對映的來源欄位在資料庫中保持不變。

**對映欄位。** 您可在這裡指定來源資料欄位與資料庫直欄之間的對映。如果來源資料欄位與資料庫中的欄具有相同的名稱，那麼自動將其對映。

- **對映。** 將按鈕左側欄位清單中已選取的來源資料欄位對映到右側清單中已選取的資料庫直欄。您可一次對映多個欄位，但兩個清單中已選取的項目數必須相同。
- **取消對映。** 刪除一個或多個已選取的資料庫欄的對映。在對話框右側的表格中選取欄位或資料庫欄後，將啟動此按鈕。
- **新增。** 將按鈕左側欄位清單中已選取的一個或多個來源資料欄位新增到右側的清單中以待對映。在左側清單中選取欄位，並且右側清單中不存在具有該名稱的欄位時，將啟動此按鈕。按一下此按鈕可以將選定欄位對映到具有同一名稱的新資料庫欄。資料庫直欄名後面將顯示單詞 **<NEW>** 以表示這是新欄位。

**合併列。** 使用鍵欄位（如交易 ID）合併在鍵欄位中具有相同值的記錄。它相當於資料庫 "equi-join"。索引鍵值必須為這些主要索引鍵的值；也就是說，它們必須是唯一的，且不得包含空值。

- **可能的鍵值。** 列出所有輸入資料來源中的所有欄位。從此清單選取一個或多個欄位，並使用箭頭按鈕將其新增為用於合併記錄的鍵欄位。具有相對應對映資料庫直欄的任何對映欄位都可用作索引鍵，但新增為新資料庫直欄的欄位（在名稱後顯示有 **<NEW>**）除外。
- **用於合併的鍵。** 基於索引鍵欄位值，列出所有用於從所有輸入資料來源中合併記錄的欄位。若要從清單中移除索引鍵，請選取一個金鑰並使用方向鈕將它傳回至「可能的索引鍵」清單。選取多個索引鍵欄位時，會啟用下列選項。

- 只包含資料庫中存在的記錄。執行局部結合；如果記錄同時在資料庫和串流中，那麼將更新對映欄位。
- 新增記錄至資料庫。執行外部加入；串流中的所有記錄將被合併（如果資料庫中存在相同記錄）或新增（如果資料庫中尚不存在該記錄）。

要對映來源資料欄位到新資料庫直欄

1. 按一下左側清單（位於對映欄位之下）中的來源欄位名稱。
2. 按一下新增按鈕完成對映。

要對映來源資料欄位到現有資料庫直欄

1. 按一下左側清單（位於對映欄位之下）中的來源欄位名稱。
2. 按一下右側資料庫直欄下的直欄名稱。
3. 按一下對映按鈕完成對映。

要刪除對映

1. 在右側清單中的「欄位」下按一下您要移除對映的欄位名稱。
2. 按一下解除對映按鈕。

在任何清單中取消選取欄位

按下 CTRL 鍵並按一下欄位名稱。

## 資料庫匯出綱目選項

在資料庫匯出綱目對話框中，可以設定資料庫匯出選項（適用於支援這些選項的資料庫），設定欄位的 SQL 資料類型，指定主要鍵欄位，並自訂匯出後產生的 CREATE TABLE 陳述式。

該對話框由幾個部分組成：

- 上面的部分（如顯示）包含匯出到資料庫的選項（適用於支援這些選項的資料庫）。如果沒有連接至此類資料庫，則不會顯示此部分。
- 中間的文字欄位顯示用於產生 CREATE TABLE 指令的範本，該欄位的預設格式為：

```
CREATE TABLE <table-name> <(table columns)>
```

- 下方的表格用於指定每個欄位的 SQL 資料類型，並指明如下方所述的哪些欄位為主要鍵值。該對話框將根據表格中的設定自動產生 <table-name > 和 <(table columns) > 參數的值。

## 設定資料庫匯出選項

如果顯示此部分，那麼可以指定多種匯出到資料庫的設定。支援此功能的資料庫類型如下所示。

- SQL Server 企業版和開發人員版本。請參閱第 305 頁的『用於 SQL Server 的選項』主題，以取得更多資訊。
- Oracle 企業版或個人版。請參閱第 305 頁的『用於 Oracle 的選項』主題，以取得更多資訊。

## 自訂 CREATE TABLE 陳述式

使用此對話框的文字欄位部分，可將其他特定於資料庫的選項新增到 CREATE TABLE 陳述式。

1. 請選中自訂 **CREATE TABLE** 指令勾選框，以啟動文字視窗。
2. 將任意特定於資料庫的選項新增到陳述式中。請務必保留文字參數 <table-name > 和 ( <table-columns > )，因為 IBM SPSS Modeler 會用它們來替代實際表格名稱和直欄定義。

## 設定 SQL 資料類型

依預設，IBM SPSS Modeler 允許資料庫伺服器自動指定 SQL 資料類型。要置換欄位的自動類型，請尋找與該欄位相對應的列並從綱目表格的 類型 欄的下拉清單中選取所需的類型。可使用 Shift-按一下選取多個列。

對於包含長度、精確度或比例引數的類型（BINARY、VARBINARY、CHAR、VARCHAR、NUMERIC 和 NUMBER，應由您指定一個長度，而不可讓資料庫伺服器自動指定長度。例如，如果為長度指定一個理想值，如 VARCHAR(25)，則必會如您所願改寫 IBM SPSS Modeler 中的儲存類型。要置換自動指定的值，請從「類型」下拉清單中選取指定，並將類型定義取代為所需的 SQL 類型定義陳述式。

最簡單的方法是先選取最接近所需類型定義的類型，然後選取指定編輯該定義。例如，如果要將 SQL 資料類型設定為 VARCHAR(25)，則可以先從「類型」下拉清單中將類型設定為 **VARCHAR(length)**，然後選取指定並將文字長度取代為值 25。

### 主要鍵值

如果匯出表格中的每一列必須對應一欄唯一值或多欄值組合，則可以通過為每個欄位選中適用的 主要關鍵字 勾選框來指定。大多數資料庫都不容許對表格進行會導致某個主要關鍵字的限制項無效的修改，並將自動建立有助於加強此限制的主要關鍵字指標。（您可以選擇性地在「索引」對話框中建立其他欄位的索引）。請參閱第 306 頁的『資料庫匯出指標選項』主題，以取得更多資訊。）

### 用於 SQL Server 的選項

使用壓縮。選取之後，建立表格以使用壓縮匯出。

壓縮目標。選擇壓縮層次。

- 列。啟用低層次壓縮（例如，相當於 SQL 中的 CREATE TABLE MYTABLE(...) WITH (DATA\_COMPRESSION = ROW);）。
- 頁。啟用頁級壓縮（例如 SQL 中的 CREATE TABLE MYTABLE(...) WITH (DATA\_COMPRESSION = PAGE);）。

### 用於 Oracle 的選項

#### Oracle 設定 - 基本選項

使用壓縮。選取之後，建立表格以使用壓縮匯出。

壓縮目標。選擇壓縮層次。

- 預設值。啟用預設壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS;）。在此情況下，它與基本選項的效果相同。
- 基本。啟用基本壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS BASIC;）。

#### Oracle 設定 - 進階選項

使用壓縮。選取之後，建立表格以使用壓縮匯出。

壓縮目標。選擇壓縮層次。

- 預設值。啟用預設壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS;）。在此情況下，它與基本選項的效果相同。
- 基本。啟用基本壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS BASIC;）。
- **OLTP**。啟用 OLTP 壓縮（例如，SQL 中的 CREATE TABLE MYTABLE(...) COMPRESS FOR OLTP;）。

- **查詢低/高。**（僅限 Exadata 伺服器）對查詢啟用 Hybrid 欄壓縮（例如，SQL 中的 CREATE TABLE MYTABLE (...) COMPRESS FOR QUERY LOW; 或 CREATE TABLE MYTABLE (...) COMPRESS FOR QUERY HIGH;）。查詢壓縮非常適合用在資料倉儲環境中；HIGH 提供比 LOW 更高的壓縮比。
- **封存低/高。**（僅限 Exadata 伺服器）對儲存啟用 Hybrid 欄壓縮（例如，SQL 中的 CREATE TABLE MYTABLE (...) COMPRESS FOR ARCHIVE LOW; 或 CREATE TABLE MYTABLE (...) COMPRESS FOR ARCHIVE HIGH;）。儲存壓縮非常適合用於壓縮那些需要長時間儲存的資料；HIGH 提供比 LOW 更高的壓縮比。

## 資料庫匯出指標選項

利用「指標」對話框，可建立從 IBM SPSS Modeler 匯出的資料庫表格的指標。可指定要包含的欄位集合併根據需要自訂 CREATE INDEX 指令。

該對話框由兩部分組成：

- 上方的文字欄位顯示可用於產生一個或多個 CREATE INDEX 指令的範本，該欄位的預設格式為：

```
CREATE INDEX <index-name> ON <table-name>
```

- 對話框下方的表格用於為要建立的每個索引新增規格。可指定每個指標的名稱以及要包含的欄位或欄。對話框將依次自動產生 <index-name >和 <table-name >參數的值。

例如，為欄位 *empid* 和 *deptid* 上的單一索引產生的 SQL 類似以下項目：

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID)
```

可以新增多列建立多指標。每一列將個別產生一個 CREATE INDEX 指令。

### 自訂 CREATE INDEX 指令

可以為所有指標或僅為特定指標自訂 CREATE INDEX 指令。這可讓您靈活地適應特定資料庫需求或選項，並根據需要將自訂作業套用至所有索引或只套用至特定索引。

- 選取對話框上方的**自訂 CREATE INDEX 指令**選項，可修改所有後續新增的指標使用的範本。請注意，系統不會自動對已新增到表格中的指標套用變更。
- 選取表格中的一列或多列，然後按一下對話框上方的**更新選定的指標**，即可將目前自訂套用到所有選定的列。
- 選取每一列的**自訂**勾選框可僅修改相應指標的指令範本。

請注意，對話框將根據表格中的設定自動產生 <index-name > 和 <table-name >參數的值，但無法直接編輯這些參數值。

**BITMAP 關鍵字。**如果使用的是 Oracle 資料庫，那麼可以自訂範本來建立點陣圖指標，而不是標準指標，方法如下：

```
CREATE BITMAP INDEX <index-name> ON <table-name>
```

點陣圖編製索引對於創建包含少量不同值的欄的編製索引很有用。產生的 SQL 可能如下所示：

```
CREATE BITMAP INDEX MYTABLE_IDX1 ON MYTABLE(COLOR)
```

**UNIQUE 關鍵字。**大多數資料庫都支援在 CREATE INDEX 指令中使用 UNIQUE 關鍵字。此關鍵字將強制執行一個唯一性限制，類似於基礎表格上的主要索引鍵限制。

```
CREATE UNIQUE INDEX <index-name> ON <table-name>
```

請注意，不必對實際指定為主要鍵值的欄位執行此約束。由於大多數資料庫會自動為 CREATE TABLE 指令中指定的任何主要索引鍵欄位建立指數，因此不必明確建立這些欄位的指標。請參閱第 304 頁的『資料庫匯出綱目選項』主題，以取得更多資訊。

**FILLFACTOR 事實索引鍵。**可以對某些索引的實體參數進行細部調整。例如，利用 SQL Server，使用者可以根據日後變更表格所產生的維護成本來選用指數大小（首次建立後）。

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID) WITH FILLFACTOR=20
```

備註

- 如果已存在指定名稱的指標，那麼指標建立將失敗。任何失敗最初都將被當作警告處理，以便建立後續指標，然後在所有指標都已試圖後在訊息日誌中重新報告為錯誤。
- 為獲得最佳效能，應在將資料載入到表格中後建立指標。指標必須至少包含一欄。
- 在執行節點前，可以在訊息日誌中預覽產生的 SQL。
- 對於寫入資料庫的暫時表格（即啟用節點快取時），指定主要鍵值和指標的選項無法使用。但是，系統可以根據資料在下游節點中的使用方式在暫時表格中建立適當的指標。例如，如果後續由 DEPT 直欄結合快取資料，則有必要在此直欄上建立快取表格的索引。

指標和查詢最佳化

在某些資料庫管理系統中，對資料庫表格進行建立、載入和編製索引操作之後，還需要執行進一步的操作，最佳化程序才能利用指標來提高對新表格的查詢速度。例如，在 Oracle 中，基於成本的查詢最佳化程序要求先對表格進行分析，然後才能在查詢最佳化系統中使用表格的指標。Oracle 的內部 ODBC 內容檔（使用者不可見）包含導致發生此情形的選項，顯示如下：

```
# Defines SQL to be executed after a table and any associated indexes  
# have been created and populated  
table_analysis_sql, 'ANALYZE TABLE <table-name> COMPUTE STATISTICS'
```

在 Oracle 中建立表格（無論是否定義主要鍵值或指標）時，都會執行此步驟。如有必要，可以類似方法自訂其他資料庫的 ODBC 內容檔 - 請聯絡支援獲得協助。

## 資料庫匯出進階選項

按一下資料庫匯出節點對話框的「進階」按鈕時，將顯示一個新的對話框，可在其中指定將結果匯出到資料庫的具體方法。

**使用批次確定。**選中此選項可關閉到資料庫的逐列提交。

**批次大小。**指定在向記憶體確定之前傳送給資料庫的記錄數。此數字越低時，資料完整性越高，但會降低資料傳輸速度。可優化調整該數字，以實現最佳的資料庫效能。

**使用大量載入。**指定一種將資料從 IBM SPSS Modeler 直接批量載入到資料庫的方法。可能需要進行多次嘗試以為特定案例選取合適的批量載入選項。

- **透過 ODBC。**選取此項可使用 ODBC API 執行多列插入，這比正常匯出到資料庫的速度更快。請從以下方選項選擇逐列鏈結或逐直欄式鏈結。
- **透過外部載入器。**選中此項可使用特定於資料庫的自訂批量載入程式。選中此項將啟動下方各個選項。對於 AIX，如果您遇到 UTF-8 資料編碼問題，則您可能需要將 locale, en\_US.UTF-8 新增至 options.cfg。

**進階 ODBC 選項。**這些選項只有在已選取透過 ODBC 時才可用。注意：只有部分 ODBC 驅動程式受支援此功能。

- **逐列綁定。** 選取逐列式鏈結可使用 SQLBulkOperations 呼叫將資料載入到資料庫。相較於使用根據逐記錄插入資料的參數式插入來說，列式鏈結通常可改進速度。
- **逐欄綁定。** 選中此項可使用逐直欄式鏈結將資料載入到資料庫。直欄式連結可通過將每個資料庫直欄（位於參數化 INSERT 陳述式中）連結至由 N 個值組成的陣列，來改善效能。每次執行 INSERT 陳述式時都會在資料庫中插入 N 列。此方法可以大大增加處理速度。

**外部載入器選項。** 指定透過外部載入器時，將顯示各式各樣的選項，這些選項可用於將資料集匯出到檔案，然後指定並執行自訂載入程式以將資料從該檔案載入到資料庫。IBM SPSS Modeler 可透過介面連接到多數一般資料庫系統的外部載入程式。軟體隨附了數個 Script，與技術說明文件一起位於 scripts 子目錄下。請注意，要使用此功能，必須將 Python 2.7 安裝在 IBM SPSS Modeler 或 IBM SPSS Modeler Server 所在的機器上，並且必須在 options.cfg 檔案中設定 python\_exe\_path 參數。請參閱『大量載入器程式設計』主題，以取得更多資訊。

- **使用分隔符號。** 指定已匯出檔案中應使用的分隔符號。選取定位點將以定位點定界，選取空格則以空格定界。選取其他可指定其他符號，比如半角逗點 (,)。
- **指定資料檔案。** 選中此項可輸入批量載入時寫入資料檔案所用的路徑。依預設，將在伺服器的暫時目錄中建立一個暫存檔。
- **指定載入器程式。** 選中此項可指定一個批量載入程式。預設情況下，該軟體會搜尋 IBM SPSS Modeler 安裝架構的 scripts 子目錄，查找給定資料庫要執行的 Python Script。軟體隨附了數個 Script，與技術說明文件一起位於 scripts 子目錄下。
- **產生日誌。** 選中此項可在指定目錄下產生日誌檔。日誌檔包含的錯誤資訊在批量載入作業失敗時可派上用場。
- **檢查表格大小。** 選中此項可執行表格檢查，以確保表格大小會隨從 IBM SPSS Modeler 匯出列個數的增多而相應地增大。
- **額外載入器選項。** 指定載入程式的其他引數。對於含有空格的引數，請使用雙引號。

雙引號包含在選用引數中，並使用反斜線進行跳出。例如，指定為 -comment "This is a \"comment\"" 的選項包含 -comment 旗標，以及呈現為 This is a "comment" 的註解本身。

透過使用另一個反斜線進行跳出，可包含單個反斜線。例如，指定為 -specialdir "C:\\Test Scripts\\" 的選項既包含旗標 -specialdir，還包含顯示為 C:\Test Scripts\ 的目錄。

## 大量載入器程式設計

資料庫匯出節點在「進階選項」對話框中提供了用於批量載入的選項。批量載入程式可以將資料從文字檔載入到資料庫。

**選項使用大量載入 - 透過外部載入器可通過配置 IBM SPSS Modeler 來完成以下三項操作：**

- 建立任何需要的資料庫表格。
- 將資料匯出到文字檔。
- 呼叫批量載入程式，以將資料從此檔案載入到資料庫表格。

一般而言，批量載入程式自身不是資料庫載入公用程式（例如，Oracle 的 sqldr 公用程式），而是一個小型 Script 或程式，該 Script 或程式可以構造正確的引數，建立任何特定於資料庫的輔助檔案（比如控制檔），然後激活資料庫載入公用程式。下列各節中的資訊將協助您編輯現有批量載入程序。

另外，您也可以撰寫自己的程式以用於批量載入。請參閱第 312 頁的『開發批量載入程式』主題，以取得更多資訊。請注意，這不在標準技術支援合約的涵蓋範圍之下，您應該聯絡 IBM 客戶服務代表以獲取協助。

批量載入 Script



IBM SPSS Modeler 附帶了大量的批量載入程式，它們適合不同的資料庫，並使用 Python Script 來實作。當執行包含選取了透過外部載入器選項的資料庫匯出節點的串流時，IBM SPSS Modeler 將透過 ODBC 來建立資料庫表格（如需要），並將資料匯出至執行 IBM SPSS Modeler Server 的主機上的暫時檔案，然後呼叫批量載入 Script。此 Script 進而執行 DBMS 供應商提供的公用程式，以便將資料從暫存檔上傳至資料庫中。

注意：IBM SPSS Modeler 安裝並不包含 Python 執行時期直譯器，因此需要另外安裝 Python。請參閱第 307 頁的『資料庫匯出進階選項』主題，以取得更多資訊。

在 IBM SPSS Modeler 安裝目錄的 \scripts 資料夾下，提供了用於下表格中列出的資料庫的 Script。

表 47. 提供了批次載入器 Script.

資料庫	Script 名稱	進一步資訊
IBM Db2	db2_loader.py	請參閱『向 IBM Db2 資料庫批量載入資料』主題，以取得更多資訊。
IBM Netezza	netezza_loader.py	請參閱第 310 頁的『向 IBM Netezza 資料庫批量載入資料』主題，以取得更多資訊。
Oracle	oracle_loader.py	請參閱第 310 頁的『向 Oracle 資料庫批量載入資料』主題，以取得更多資訊。
SQL Server	mssql_loader.py	請參閱第 311 頁的『向 SQL Server 資料庫批量載入資料』主題，以取得更多資訊。
Teradata	teradata_loader.py	請參閱第 311 頁的『向 Teradata 資料庫批量載入資料』主題，以取得更多資訊。

## 向 IBM Db2 資料庫批量載入資料

下列幾點可協助您使用「資料庫匯出進階選項」對話框中的「外部載入器」選項來配置從 IBM SPSS Modeler 到 IBM Db2 資料庫的批量載入。

### 確保安裝了 Db2 指令行處理器 (CLP) 公用程式

Script db2\_loader.py 呼叫 Db2 LOAD 指令。確保在要執行 db2\_loader.py 的伺服器（通常為執行 IBM SPSS Modeler Server 的主機）上安裝了指令行處理器（對於 UNIX，應為 db2；對於 Windows，應為 db2cmd）。

### 檢查本端資料庫別名是否與實際資料庫名稱相同

Db2 本端資料庫別名是 Db2 用戶端軟體用來參照本端或遠端 Db2 實例中資料庫的名稱。如果本端資料庫別名與遠端資料庫的名稱不同，請提供其他載入程序選項：

```
-alias <local_database_alias>
```

例如，在主機 GALAXY 上的遠端資料庫名為 STARS，但在執行 IBM SPSS Modeler Server 的主機上 DB2 本端資料庫別名為 STARS\_GALAXY。使用其他載入程序選項

```
-alias STARS_GALAXY
```

### 非 ASCII 字元資料編碼

如果要批量載入非 ASCII 格式的資料，應確保在 db2\_loader.py 配置部分中的代碼頁變數在系統中正確設定。

### 空白字串

空白字串會作為空值匯出至資料庫。

## 向 IBM Netezza 資料庫批量載入資料

下列幾點說明如何通過使用「資料庫匯出進階資料選項」對話框中的「外部載入程序」選項來配置從 IBM SPSS Modeler 到 IBM Netezza 資料庫的批量載入。

確保安裝了 Netezza `nzload` 公用程式

Script `netezza_loader.py` 呼叫 Netezza 公用程式 `nzload`。確保在要執行 `netezza_loader.py` 的伺服器上安裝並正確配置 `nzload`。

匯出非 ASCII 資料

如果匯出中包含非 ASCII 格式的資料，那麼可能需要向「資料庫匯出進階選項」對話框的額外載入器選項欄位新增 `-encoding UTF8`。這將確保正確上傳非 ASCII 資料。

日期、時間及時間戳記格式資料

在串流內容中，將日期格式設為 **DD-MM-YYYY**，並將時間格式設為 **HH:MM:SS**。

空白字串

空白字串會作為空值匯出至資料庫。

在現有表格中插入資料時，串流與目標表格中的欄順序不同

如果串流中的欄順序與目標表格不同，那麼資料值將被插入錯誤的欄中。使用欄位重新排序節點可確保串流中的欄順序與目標表格中的欄順序相符。請參閱第 156 頁的『欄位重新排序節點』主題，以取得更多資訊。

追蹤 `nzload` 進度

在本端方式下執行 IBM SPSS Modeler 時，向「數據庫匯出進階資料選項」對話框的其他載入程式選項欄位新增 `-sts`，即可在 `nzload` 公用程式開啟的指令視窗中查看每 10000 列的狀態訊息。

## 向 Oracle 資料庫批量載入資料

下列幾點可協助您使用「資料庫匯出進階選項」對話框中的「外部載入器」選項來配置從 IBM SPSS Modeler 到 Oracle 資料庫的批量載入。

確保安裝了 Oracle `sqlldr` 公用程式

Script `oracle_loader.py` 呼叫 Oracle 公用程式 `sqlldr`。請注意，`sqlldr` 未自動包含在 Oracle 用戶端中。確保在要執行 `oracle_loader.py` 的伺服器上安裝 `sqlldr`。

指定資料庫 SID 或服務名稱

如果要將資料匯出至非本端 Oracle 伺服器，或者本端 Oracle 伺服器有多個資料庫，則需要在「資料庫匯出進階資料選項」對話框的額外載入器選項欄位中指定下列選項，以傳入 SID 或服務名稱：

```
-database <SID>
```

編輯 `oracle_loader.py` 的配置部分

在 UNIX（或 Windows）系統上，編輯 `oracle_loader.py` Script 起點處的配置部分。在這裡，可根據情況指定 `ORACLE_SID`、`NLS_LANG`、`TNS_ADMIN` 和 `ORACLE_HOME` 環境變數值，以及 `sqlldr` 公用程式的完整路徑。

日期、時間及時間戳記格式資料

在串流內容中，通常應將日期格式設為 **YYYY-MM-DD**，並將時間格式設為 **HH:MM:SS**。

如果要使用其他日期與時間格式，請參閱 Oracle 文件並編輯 *oracle\_loader.py* Script 檔案。

非 ASCII 字元資料編碼

如果要批量載入非 ASCII 格式的資料，那麼應該確保在系統中正確地設定了環境變數 `NLS_LANG`。這將由 Oracle 載入公用程式 *sqlldr* 來讀取。例如，對於 Windows，Shift-JIS 的 `NLS_LANG` 的正確值是 `Japanese_Japan.JA16SJIS`。有關 `NLS_LANG` 的更多詳細資料，請查閱 Oracle 文件。

空白字串

空白字串會作為空值匯出至資料庫。

## 向 SQL Server 資料庫批量載入資料

下列幾點說明如何通過使用「資料庫匯出進階資料選項」對話框中的「外部載入程序」選項來配置從 IBM SPSS Modeler 到 SQL Server 資料庫的批量載入。

### 確保安裝了 SQL Server bcp.exe 公用程式

Script *mssql\_loader.py* 呼叫 SQL Server 公用程式 *bcp.exe*。確保在要執行 *mssql\_loader.py* 的伺服器上安裝 *bcp.exe*。

### 使用空格作為定界字元不起作用

請避免在「數據庫匯出進階資料選項」對話框中選擇空格作為定界字元。

### 建議的「檢查表格大小」選項

建議在「數據庫匯出進階選項」對話框中啟用檢查表格大小選項。系統並不是總能偵測到批量載入過程中的失敗，啟用此選項將執行額外的檢查，以確保載入正確號碼的列。

空白字串

空白字串會作為空值匯出至資料庫。

### 指定標準 SQL Server 命名實例

在某些情況下，SPSS Modeler 可能會由於主機名稱未限定而無法存取 SQL Server 並顯示下列錯誤：執行外部批量載入程序時發生錯誤。日誌檔可能提供了更多詳細資料。

要更正此錯誤，請向其他載入程序選項欄位新增下列字串（包含雙）：

```
"-S mhreboot.spss.com\SQLEXPRESS"
```

## 向 Teradata 資料庫批量載入資料

下列幾點可協助您使用「資料庫匯出進階選項」對話框中的「外部載入器」選項來配置從 IBM SPSS Modeler 到 Teradata 資料庫的批量載入。

確保安裝了 Teradata fastload 公用程式

Script *teradata\_loader.py* 呼叫 Teradata 公用程式 *fastload*。確保在要執行 *teradata\_loader.py* 的伺服器上安裝並正確配置 *fastload*。

資料只能批量載入到空白表格中

對於批量載入，只能使用空白表格作為目標。如果目標表格在批量載入之前包含任何資料，那麼作業將失敗。

日期、時間及時間戳記格式資料

在串流內容中，將日期格式設為 **YYYY-MM-DD**，並將時間格式設為 **HH:MM:SS**。

空白字串

空白字串會作為空值匯出至資料庫。

Teradata 程序 ID (tdpid)

依預設，*fastload* 會將資料匯出至 tdpid=dbc 的 Teradata 系統中。通常，在 HOSTS 檔案中會有一個項目，用於將 dbccop1 與 Teradata 伺服器的 IP 位址進行關聯。要使用其他伺服器，請在「資料庫匯出進階選項」對話框的**額外載入器選項**欄位中指定下列選項以傳遞此伺服器的 tdpid：

```
-tdpid <id>
```

表格名和直欄名稱中的空格

如果表格名或直欄名稱包含空格，那麼批量載入作業將失敗。如有可能，請重新命名表格或欄以移除空格。

## 開發批量載入程式

該主題說明如何開發可在 IBM SPSS Modeler 中執行的批量載入程式，以便將文字檔資料載入到資料庫中。請注意，這不在標準技術支援合約的涵蓋範圍之下，您應該聯絡 IBM 客戶服務代表以獲取協助。

使用 Python 建立批量載入程式

依預設，IBM SPSS Modeler 會根據資料庫類型搜尋預設的批量載入程式。請參閱第 309 頁的表 47。

可使用 Script *test\_loader.py* 來協助開發批量載入程式。請參閱第 314 頁的『正在測試批量載入程式』主題，以取得更多資訊。

傳送給批量載入程式的物件

IBM SPSS Modeler 寫入兩個將傳送給批量載入程式的檔案。

- **資料檔案。**該檔案包含要以文字格式載入的資料。
- **綱目檔。**該檔案為 XML 檔案，說明欄名與類型，並提供資料檔案的格式資訊（例如，用作欄位間定界字元的字元）。

此外，IBM SPSS Modeler 還將傳送其他資訊（例如，表格名、使用者名稱和密碼等）作為呼叫批量載入程式時的引數。

注意：為了向 IBM SPSS Modeler 表明載入操作完成，批量載入程式將刪除綱目檔。

傳送給批量載入程式的引數

下表格中列出了傳遞給程式的引數。

表 48. 傳遞給批次載入器的引數.

引數	說明
schemafile	綱目檔的路徑。
data file	資料檔案的路徑。
servername	DBMS 伺服器的名稱；可以為空白。

表 48. 傳遞給批次載入器的引數 (繼續).

引數	說明
databasename	DBMS 伺服器內的資料庫名稱；可以為空白。
username	用於登入資料庫的使用者名稱。
password	用於登入資料庫的密碼。
tablename	要載入的表格的名稱。
ownername	表格所有人的名稱 (也稱為綱目名稱)。
logfilename	日誌檔的名稱 (如果為空白, 那麼不會建立日誌檔)。
rowcount	資料集中的列個數。

傳遞這些標準引數之後, 將向批量載入程式傳遞在「資料庫匯出進階選項」的額外載入器選項欄位中指定的任何選項。

### 資料檔案格式

資料將以文字格式寫入資料檔案, 每個欄位之間以在「資料庫匯出進階資料選項」上指定的定界字元進行分隔。下面是 Tab 定界資料檔案的可能顯示方式的範例。

```
48 F HIGH NORMAL 0.692623 0.055369 drugA
15 M NORMAL HIGH 0.678247 0.040851 drugY
37 M HIGH NORMAL 0.538192 0.069780 drugA
35 F HIGH HIGH 0.635680 0.068481 drugA
```

採用 IBM SPSS Modeler Server (或 IBM SPSS Modeler, 如未連線到 IBM SPSS Modeler Server) 所用的本端編碼來寫入該檔案。某些格式通過 IBM SPSS Modeler 串流設定來控制。

### 綱目檔格式

綱目檔是用於說明資料檔案的 XML 檔案。下面是前述資料檔案附帶的一個範例。

```
<?xml version="1.0" encoding="UTF-8"?>
<DBSCHEMA version="1.0">
  <table delimiter="\t" commit_every="10000" date_format="YYYY-MM-DD" time_format="HH:MM:SS"
  append_existing="false" delete_datafile="false">
    <column name="Age" encoded_name="416765" type="integer"/>
    <column name="Sex" encoded_name="536578" type="char" size="1"/>
    <column name="BP" encoded_name="4250" type="char" size="6"/>
    <column name="Cholesterol" encoded_name="43686F6C65737465726F6C" type="char" size="6"/>
    <column name="Na" encoded_name="4E61" type="real"/>
    <column name="K" encoded_name="4B" type="real"/>
    <column name="Drug" encoded_name="44727567" type="char" size="5"/>
  </table>
</DBSCHEMA>
```

下面的兩個表格列出了綱目檔的 <table> 和 <column> 元素的屬性。

表 49. <table> 元素的屬性.

屬性	說明
delimiter	欄位定界字元 (TAB 將代表為 \t)。
commit_every	批次大小間隔 (與在「數據庫匯出進階資料選項」對話框上相同)。
date_format	用於代表日期的格式。
time_format	用於代表時間的格式。

表 49. <table> 元素的屬性 (繼續).

屬性	說明
append_existing	true (如果要載入的表格已包含資料) ; 否則, 值為 false。
delete_datafile	true (如果批量載入程式在完成載入時應刪除資料檔案)。

表 50. <column> 元素的屬性.

屬性	說明
name	直欄名稱。
encoded_name	直欄名稱已轉換為與資料檔案相同的編碼, 並輸出為一系列兩位十六進位數字。
type	欄的資料類型, 值為下列其中一個: integer、real、char、time、date 和 datetime。
size	對於 char 資料類型, 這是欄的上限寬度 (字元)。

## 正在測試批量載入程式

可使用位於 IBM SPSS Modeler 安裝目錄的 \scripts 資料夾下的測試 Script *test\_loader.py* 對批量載入進行測試。在嘗試開發、偵錯用於 IBM SPSS Modeler 的批量載入程式或 Script 以及對其進行疑難排解時, 此操作非常有用。

要使用測試 Script, 請繼續以下操作。

1. 執行 *test\_loader.py* Script, 將綱目與資料檔案複製到檔案 *schema.xml* 與 *data.txt*, 並建立 Windows 批次檔 (*test.bat*)。
2. 編輯 *test.bat* 檔案以選取要測試的批量載入程式或 Script。
3. 從指令 shell 執行 *test.bat* 以測試選定的批量載入程式或 Script。

注意: 實際上, 執行 *test.bat* 不會將資料載入資料庫。

## 純文字檔案匯出節點

純文字檔案匯出節點可用於將資料寫入已定界文字檔。特別適用於匯出其他分析或試算表軟體可以讀取的資料。

如果資料包含地理空間資訊, 那麼您可以將其匯出為純文字檔案; 如果產生了「變數檔案」來源節點以便在同一個串流中使用, 那麼所有的儲存、測量和地理空間 meta 資料都將保留在新的來源節點中。但是, 如果您匯出該資料, 然後在另一個串流中將其匯入, 那麼您必須執行一些額外的步驟以便在新的來源節點中設定地理空間 meta 資料。如需相關資訊, 請參閱主題第 23 頁的『「變數檔案」節點』。

註: 無法以過往的快取格式寫入檔案, 這是因為 IBM SPSS Modeler 已不再使用該快取檔案格式。IBM SPSS Modeler 快取檔現以 IBM SPSS Statistics .sav 格式儲存, 您可以通過 Statistics 匯出節點寫入檔案。如需相關資訊, 請參閱主題第 315 頁的『Statistics 匯出節點』。

### 「純文字檔案匯出」標籤

**匯出檔案。** 指定檔案名。輸入檔名, 或按一下檔案選擇器按鈕以瀏覽至該檔案的位置。

**寫入方式。** 如果選取了改寫, 那麼將改寫指定檔案中的任何現有資料。如果選取附加, 則輸出將被新增到現有檔案的末尾, 同時保留該檔案包含的所有資料。

- **包括欄位名稱。** 如果已選取此選項, 那麼欄位名稱將寫入到輸出檔案的第一行。此選項只適用於改寫寫入模式。

每筆記錄後換新行。如果已選取此選項，那麼每條記錄都將對應寫入到輸出檔案的新的一行中。

欄位分隔字元。指定要在產生的文字檔中的欄位值之間插入的字元。選項有逗點、定位點、空格和其他。如果選取其他，則在文字框中輸入所需的分隔符號。

符號參照。指定用於為符號欄位值添加引號的方式。選項有無（不為值添加引號）、單個 (')、雙 (") 和其他。如果選取其他，則在文字框中輸入所需的引號。

編碼。指定使用的文字編碼方法。您可以選擇系統預設、串流預設值或 UTF-8。

- 系統預設值是在 Windows 控制台中指定，如果是分散式模式執行，則是在伺服器電腦上指定。
- 串流預設值在「串流內容」對話框中指定。

小數點符號。指定小數在資料中的代表方法。

- 串流預設值。將使用目前串流預設設定定義的小數分隔字元。通常為電腦語言環境中定義的小數分隔字元。
- 句點(.)。將使用期間字元作為小數分隔字元。
- 逗點(,)。將使用逗點字元作為小數分隔字元。

為此資料產生匯入節點。選中此選項可自動產生將讀取已匯出資料檔案的「變數檔案」來源節點。請參閱第 23 頁的『「變數檔案」節點』主題，以取得更多資訊。

---

## Statistics 匯出節點

Statistics 匯出節點可以以 IBM SPSS Statistics *.sav* 格式匯出資料。IBM SPSS Statistics Base 和其他模組能夠讀取 IBM SPSS Statistics *.sav* 檔案。這也是 IBM SPSS Modeler 快取檔所使用的格式。

有時候，將 IBM SPSS Modeler 欄位名稱對映到 IBM SPSS Statistics 變數名稱時會出錯，這是因為 IBM SPSS Statistics 變數名稱被限制為 64 個字元且不能包含某些字元，比如空格、美元符號 (\$) 和 劃線 (-) 等。有兩種方法可以規避上述限制：

- 可通過按一下「過濾」標籤，根據 IBM SPSS Statistics 變數名稱要求更名欄位。請參閱第 316 頁的『重新命名或過濾 IBM SPSS Statistics 的欄位』主題，以取得更多資訊。
- 選擇同時從 IBM SPSS Modeler 匯出欄位名稱和標籤。

注意：IBM SPSS Modeler 以 Unicode UTF-8 格式寫入 *.sav* 檔案。IBM SPSS Statistics 16.0 版以後的版本只支援 Unicode UTF-8 格式的檔案。為了防止資料可能損壞，使用 Unicode 編碼儲存的 *.sav* 檔案不得用於 16.0 之前的 IBM SPSS Statistics 版本。有關詳細資訊，請參閱 IBM SPSS Statistics 說明。

多個回應集。匯出檔案時，將自動保留定義在串流中的所有多重回應集。您可以使用「過濾器」標籤來檢視並編輯任何節點中的回應集。請參閱第 131 頁的『編輯多重回應集』主題，以取得更多資訊。

## Statistics 匯出節點 - 「匯出」標籤

匯出檔案 指定檔案的名稱。輸入檔名，或按一下檔案選擇器按鈕以瀏覽至該檔案的位置。

檔案類型 請選取是以正常的 *.sav* 格式還是壓縮的 *.zsav* 格式儲存檔案。

使用密碼對加密檔案 要使用密碼保護檔案，請選中此框；系統將提示您在另一對話框中輸入並確認密碼。

註：受密碼受保護的檔案只能由 SPSS Modeler 16 版或更高版本，或者由 SPSS Statistics 21 版或更高版本開啟。

**匯出欄位名稱為變數稱** 指定將變數名稱和標籤從 SPSS Modeler 匯出至 SPSS Statistics *.sav* 或 *.zsav* 檔案時的處理方法。

- **名稱** 選取此選項可同時匯出 SPSS Modeler 欄位名稱和欄位標籤。名稱將以 SPSS Statistics 變數名方式匯出，而標籤則以 SPSS Statistics 變數標籤方式匯出。
- **作為變數標籤的名稱** 選中此選項可將 SPSS Modeler 欄位名稱用作 SPSS Statistics 中的變數標籤。SPSS Modeler 容許在欄位名稱中包含不適用於 SPSS Statistics 變數名稱的字元。為了防止建立無效的 SPSS Statistics 名稱，請選取將名稱用作變數標籤或使用「過濾」標籤調整欄位名。

**啟動應用程式** 如果電腦上安裝了 SPSS Statistics，則可以選取此選項對已儲存資料檔案直接激活應用程式。Helper 應用程式對話框中必須指定用於啟動應用程式的選項。請參閱第 298 頁的『IBM SPSS Statistics 說明應用程式』主題，以取得更多資訊。要在不開啟外部程式的情況下直接建立 SPSS Statistics *.sav* 或 *.zsav* 檔案，請取消選取此選項。

**註：**當以伺服器（分散式）方式同時執行 SPSS Modeler 和 SPSS Statistics 時，寫入出資料並啟動 SPSS Statistics 階段作業不會自動開啟 SPSS Statistics 用戶端以顯示讀取入作用中的資料集的資料集。暫行解決方法是 SPSS Statistics 用戶端啟動時立刻在其中開啟資料檔案。

**為該資料產生匯入節點** 選中此選項可自動產生將讀取已匯出資料檔案的「Statistics 檔案」來源節點。請參閱第 27 頁的『Statistics 檔案節點』主題，以取得更多資訊。

## 重新命名或過濾 IBM SPSS Statistics 的欄位

在將資料從 IBM SPSS Modeler 匯出或部署至外部應用程式（例如 IBM SPSS Statistics）時，可能需要重新命名或調整欄位名稱。「統計資料轉換」、「統計資料輸出」及「統計資料匯出」對話框包含一個「過濾器」標籤，以協助此處理程序。

「過濾器」標籤功能的基本說明在其他部分中討論。請參閱第 129 頁的『設定過濾選項』主題，以取得更多資訊。本主題提供將資料讀取至 IBM SPSS Statistics 的提示。

若要調整欄位名稱以符合 IBM SPSS Statistics 命名慣例：

1. 在「過濾器」標籤上，按一下「過濾器選項功能表」工具列按鈕（工具列上的第一個）。
2. 選取「重新命名 IBM SPSS Statistics」。
3. 在「重新命名 IBM SPSS Statistics」對話框上，您可以選擇將檔名中的無效字元取代為井字號 (#) 字元或底線 (\_)。

**重新命名多重回應集。**如果您想要調整多重回應集的名稱，這些回應集可以利用「統計資料檔案」來源節點匯入至 IBM SPSS Modeler，請選取此選項。它們用來記錄資料，針對每一個案例可以具有多個值，例如在意見調查回應中。

---

## Data Collection 匯出節點

Data Collection 匯出節點基於 Data Collection 資料模型，以 Data Collection 市場調查軟體中使用的格式儲存資料。此格式可以將觀察值資料（對意見調查期間所收集的問題的實際回應）與 meta 資料（說明如何收集和組織觀察值資料）進行識別。meta 資料包括以下資訊，例如問題文字、變數名稱和說明、多重回應集、不同文字的轉換以及觀察值資料結構的定義。請參閱第 28 頁的『Data Collection 節點』主題，以取得更多資訊。

**meta 資料檔。**指定調查表定義檔的名稱 (*.mdd*)，將在該檔案中儲存已匯出的 meta 資料。預設的調查表將基於欄位類型資訊建立。例如，列名（集合）欄位可以代表為單個問題，其中，將欄位說明用作問題文字，並為每個已定義的值採用個別勾選框。



**合併 meta 資料。**指定 meta 資料是將改寫現有版本還是與現有 meta 資料合併。如果選取了合併選項，那麼將在每次執行串流時建立新版本。由此可以在調查表發生變更時追蹤它的各種版本。每個版本都可看作是用於收集觀察值資料特定集合的 meta 資料的一個 Snapshot。

**啟用系統變數。**指定系統變數是否包含在匯出的 *.mdd* 檔案中。這些變數中包含 *Respondent.Serial*、*Respondent.Origin* 和 *DataCollection.StartTime* 等。

**個案資料設定。**指定從其中匯出觀察值資料的 IBM SPSS Statistics 資料 (*.sav*) 檔案。請注意，所有對變數和值名稱的限制均適用於此處，例如您可能需要切換至「過濾」標籤，並套用「過濾選項」功能表上的「為 IBM SPSS Statistics 更名」選項，以更正欄位名稱中的無效字元。

**為此資料產生匯入節點。**選取此選項可自動產生將讀取已匯出資料檔案的 Data Collection 來源節點。

**多個回應集。**匯出檔案時，將自動保留定義在串流中的所有多重回應集。您可以使用「過濾器」標籤來檢視並編輯任何節點中的回應集。請參閱第 131 頁的『編輯多重回應集』主題，以取得更多資訊。

---

## Analytic Server 匯出節點

Analytic Server 的「匯出」節點使您能夠將來自分析的資料寫入現有 Analytic Server 資料來源。例如，此資料源可以是 Hadoop 分散式檔案系統 (HDFS) 或資料庫中的文字檔。

通常，具有 Analytic Server 匯出節點的串流還會以 Analytic Server 來源節點開頭，並被提交到 Analytic Server，然後在 HDFS 上執行。另外，具有「原生」資料來源的串流可以 Analytic Server 匯出節點結束，以便上傳相對較小的資料集（不超過 100,000 條記錄）與 Analytic Server 配合使用。

如果您想要使用自己的 Analytic Server 連線來代替管理者定義的預設連線，請取消選取**使用預設 Analytic Server** 並選取您的連線。有關設定多個 Analytic Server 連線的詳細資料，請參閱連線到 Analytic Server。

**資料來源。**選取包含要使用的資料的資料來源。資料來源包含與該來源相關聯的檔案和 meta 資料。按一下**選取**以顯示可用資料來源的清單。請參閱第 11 頁的『選取資料來源』主題，以取得更多資訊。

如果您需要建立新的資料來源或編輯現有資料來源，請按一下**啟動資料來源編輯器...**

**眾數。**選取附加表示新增到現有資料來源，選取改寫表示取代資料來源的內容。

**針對此資料產生一個匯入節點。**選取此選項可在資料匯出至指定資料來源時產生此資料的來源節點。這會將此節點新增到串流畫布中。

請注意，使用多個 Analytic Server 連線在控制資料流程方面十分有用。例如，使用 Analytic Server 「來源」及「匯出」節點時，您可能想要在串流的不同分支中使用不同的 Analytic Server 連線，因此每一個分支執行時，它會使用自己的 Analytic Server，且不會將任何資料拉取到 IBM SPSS Modeler Server。請注意，如果分支包含多個 Analytic Server 連線，則會從 Analytic Server 拉取到 IBM SPSS Modeler Server。如需相關資訊（包括限制），請參閱 Analytic Server 串流內容。

---

## IBM Cognos 匯出節點

IBM Cognos 匯出節點允許您採用 UTF-8 格式將資料從 IBM SPSS Modeler 串流匯出到 Cognos Analytics。這樣，Cognos 可利用來自 IBM SPSS Modeler 的轉換或分數資料。例如，可使用 Cognos Report Studio 建立一個基於匯出資料的報告，包含預測和信賴度值。然後可在 Cognos 伺服器上儲存報告並分配給 Cognos 使用者。

註：只能匯出關聯式資料，不能匯出 OLAP 資料。

要將資料匯出到 Cognos，需要指定下列各項：

- Cognos 連線 - 到 Cognos Analytics 伺服器的連線（支援第 11 版或更高版本）
- ODBC 連線 - 到 Cognos 伺服器使用的 Cognos 資料伺服器的連線

在 Cognos 連線中指定要使用的 Cognos 資料來源。該資料來源必須使用與 ODBC 資料來源相同的登入資訊。

將實際串流資料匯出到資料伺服器，同時將套件 meta 資料匯出到 Cognos 伺服器。

至於任何其他匯出節點，您還可以使用節點對話框的「發佈」標籤，利用 IBM SPSS Modeler Solution Publisher 來發佈串流以進行部署。

註：Cognos 來源節點僅支援 Cognos CQM 包。不支援 DQM 包。

## Cognos 連線

您可以在此處指定與要用於匯出的 Cognos Analytics 伺服器的連線（支援第 11 版或更高版本）。該程序涉及將 meta 資料匯出到 Cognos 伺服器上的新套件，同時將串流資料匯出到 Cognos 資料伺服器。

**連線。** 按一下編輯按鈕將顯示一個對話框，您可在此對話框中定義要將資料匯出至其中的 Cognos 伺服器的 URL 及其他詳細資料。如果已經通過 IBM SPSS Modeler 登入 Cognos 伺服器，也可編輯現行連線的明細。如需相關資訊，請參閱第 34 頁的『Cognos 連線』。

**資料來源。** 這是將資料匯出到的 Cognos 資料來源（通常為資料庫）的名稱。下拉清單顯示您可從現行連線存取的所有 Cognos 資料來源。按一下**重新整理**按鈕以更新此清單。

**資料夾。** Cognos 伺服器上要建立匯出套件的資料夾路徑和名稱。

**套件名稱。** 要包含匯出 meta 資料的指定資料夾中的套件名稱。這必須是一個帶有單獨查詢主旨的新套件；不能匯出到現有套件。

**眾數。** 指定您希望如何執行匯出：

- **立即發行套件。**（預設）按一下**執行**即開始執行匯出作業。
- **匯出動作 Script。** 建立一個可稍後執行的 XML Script（例如使用 Framework Manager）以執行匯出。在檔案欄位中鍵入 Script 的路徑和檔名，或使用編輯按鈕指定 Script 檔的名稱和位置。

**為此資料產生匯入節點。** 選取此選項以針對匯出至指定資料來源與表格的資料產生來源節點。按一下**執行**時，此節點會新增至串流畫布。

## ODBC 連線

可在此處指定將要匯出串流資料至的 Cognos 資料伺服器（即資料庫）的連線。

註：必須確保此處指定的資料來源與在 **Cognos 連線**畫面上指定的資料來源相同。您還必須確保 Cognos 連線資料來源使用與 ODBC 資料來源相同的登入資訊。

**資料來源。** 顯示所選資料來源。輸入名稱或者從下拉清單中選取名稱。如果清單中未顯示所需的資料庫，則選取新增新的資料庫連線並從「資料庫連線」對話框選定資料庫。請參閱第 16 頁的『新增資料庫連線』主題，以取得更多資訊。

**表格名稱。** 輸入接收資料的表格名稱。如果選取**插入到表格**選項，則可以通過按一下**選取**按鈕在資料庫中選取一個現有表格。

**建立表格。** 選中此項可建立一個新的資料庫表格或改寫現有的資料庫表格。

**插入到表格。**選中此項可將資料作為新列插入現有的資料庫表格中。

**合併表格。**（如可用）選取此選項以使用相應來源資料欄位中的值更新所選資料庫直欄。選取此選項會啟用合併按鈕，在其顯示的對話框中您可以將來源資料欄位對映到資料庫直欄。

**刪除現有表格。**選中此項可在建立新表格時刪除所有名稱相同的現有表格。

**刪除現有列。**選取此選項可在插入表格時先將現有列從表格中刪除然後匯出。

**註：**如果選取上述任意兩個選項，那麼在執行節點時將收到**改寫警告**訊息。要想不顯示此警告，請取消選取「使用者選項」對話框的「通知」標籤上的**當節點改寫資料表時發出警告**選項。

**預設字串大小。**上游「類型」節點中標示為無類型的欄位將作為字串欄位寫入資料庫。請指定無類型欄位要使用的字串大小。

按一下綱目可開啟一個對話框，您可在其中設定各種匯出選項（對於支援此功能的資料庫）、設定所需欄位的 SQL 資料類型，並指定創建資料庫編製索引的主要關鍵字。請參閱第 304 頁的『資料庫匯出綱目選項』主題，以取得更多資訊。

按一下編製索引可指定創建匯出表格編製索引的選項，以提高資料庫的運行效能。請參閱第 306 頁的『資料庫匯出指標選項』主題，以取得更多資訊。

按一下進階可指定批量載入和資料庫提交選項。請參閱第 307 頁的『資料庫匯出進階選項』主題，以取得更多資訊。

**用引號括住表格與直欄名稱。**選取將 CREATE TABLE 陳述式傳送到資料庫時使用的選項。必須為包含空格和非標準字元的表格和欄添加引號。

- **根據需要。**選取此選項，以容許 IBM SPSS Modeler 自動根據各個情況確定是否需要添加引號。
- **一律。**選中此項將始終為表格名和直欄名稱添加參照。
- **從不。**選中此項將取消參照。

**為此資料產生匯入節點。**選取此選項以針對匯出至指定資料來源與表格的資料產生來源節點。按一下執行時，此節點會新增至串流畫布。

---

## IBM Cognos TM1 匯出節點

通過 IBM Cognos 匯出節點，您可以將資料從 SPSS Modeler 串流匯出至 Cognos TM1。這樣，Cognos Analytics 可利用來自 SPSS Modeler 的轉換或分數資料。

**註：**您僅能匯出不是環境定義維度資料的測量；或者，您可以將新元素新增至多維度報表。

要將資料匯出到 Cognos Analytics（支援第 11 版或更高版本），需要指定下列各項：

- Cognos TM1 伺服器連線。
- 將向其匯出資料的多維度報表。
- 從 SPSS 資料名稱對映到對等的 TM1 維度及測量。

**註：**TM1 使用者需要下列許可權：多維度報表的寫入權、維度的讀取權及維度元素的寫入權。此外，您需要 IBM Cognos TM1 10.2 Fix Pack 3 或更新版本，然後 SPSS Modeler 才能匯入及匯出 Cognos TM1 資料。基於舊版的現有串流仍將發揮作用。

此節點不需要管理者認證。但如果您仍然使用舊式的 17.1 以前的 TM1 節點，則仍需要管理者認證。

至於任何其他匯出節點，您還可以使用節點對話框的「發佈」標籤，利用 IBM SPSS Modeler Solution Publisher 來發佈串流以進行部署。

註：您必須先在 `tmls.cfg` 檔案（這是 TM1 伺服器的根目錄中的 TM1 伺服器配置檔案）中驗證部分設定，然後才能使用 SPSS Modeler 中的 TM1「來源」或「匯出」節點。

- HTTPPortNumber - 設定有效的埠號；一般為 1-65535。請注意，這不是您之後在節點的連線中指定的埠號；它是 TM1 使用的依預設停用的內部埠。如有需要，請聯絡 TM1 管理者來確認此埠的有效設定。
- UseSSL - 如果您將此選項設為 *True*，則會使用 HTTPS 作為傳輸通訊協定。在此情況下，您必須將 TM1 憑證匯入到 SPSS Modeler Server JRE。

## 連接至 IBM Cognos TM1 多維度報表來匯出資料

將資料匯出至 IBM Cognos TM1 資料庫的第一步是在 IBM Cognos TM1 對話框的連線標籤中選取相關的 TM1 管理主機以及關聯的伺服器和資料方塊。

註：將資料匯出至 TM1 時僅會捨棄實際「空」值。零 (0) 值會作為有效值匯出。另請注意，只有儲存類型為字串的欄位可以在「對映」標籤上對映到維度。在匯出至 TM1 之前，您必須使用 IBM SPSS Modeler 用戶端將非字串資料類型轉換為字串。

**管理主機** 輸入管理主機的 URL，您要連接的 TM1 伺服器安裝在此主機中。對於所有 TM1 伺服器，管理主機均定義為單一 URL。可以從這個 URL 探索及存取您的環境中已安裝及正在執行的所有 IBM Cognos TM1 伺服器。

**TM1 伺服器** 與管理主機建立連線後，請選取包含所要匯入的資料的伺服器，然後按一下**登入**。如果您先前不曾連接至此伺服器，則系統會提示您輸入**使用者名稱及密碼**；或者，您可以搜尋另存為儲存的認證的之前輸入的登入詳細資料。

**選取要匯出的 TM1 多維度報表** 顯示 TM1 伺服器內可以向其匯出資料的多維度報表的名稱。

若要選擇要匯出的資料，請選取多維度報表，然後按一下向右箭頭，將多維度報表移至**匯出至多維度報表**欄位中。選取資料方塊後，請使用「對映」標籤將 TM1 維度和測量對映到相關 SPSS 欄位或固定值（選取作業）。

## 對映要匯出的 IBM Cognos TM1 資料

選取 TM1 管理主機以及關聯的 TM1 伺服器和資料方塊之後，請使用「IBM Cognos TM1 匯出」對話框的「對映」標籤將 TM1 維度和測量對映到 SPSS 欄位或者將 TM1 維度設定為固定值。

註：只有儲存類型為字串的欄位可以對映到維度。在匯出至 TM1 之前，您必須使用 IBM SPSS Modeler 用戶端將非字串資料類型轉換為字串。

**欄位** 列出 SPSS 資料檔案中可供匯出的資料欄位名稱。

**TM1 維度** 顯示「連線」標籤中選取的 TM1 資料方塊及其一般維度、測量維度及所選測量維度的元素。選取要對映至 SPSS 資料欄位的 TM1 維度（即，測量）的名稱。

您可以使用「對映」標籤上的下列選項。

**選取測量維度** 從所選資料方塊的維度清單中，選取某個維度作為測量維度。

選取測量維度之外的某個維度，並按一下**選取**時，將顯示一個對話框，其中包含所選維度的葉節點元素。您只能選取葉節點元素。所選元素帶有標籤 **S**。

**對映** 將所選 SPSS 資料欄位對映到所選 TM1 維度或測量（一般維度或者測量維度中的特定測量或元素）。對映的欄位使用 **M** 標示。

**取消對映** 取消所選 SPSS 資料欄位與所選 TM1 維度或測量之間的對映。請注意，您一次只能取消一個對映。取消對映的 SPSS 資料欄位將移回到左側欄中。

**建立新的項目** 在 TM1 測量維度中建立新的項目測量。這將顯示一個對話框，您可以在其中輸入新的 **TM1 測量名稱**。此選項僅適用於測量維度，而不適用於一般維度。

有關 TM1 的進一步資訊，請參閱 IBM Cognos TM1 文件 ([http://www-01.ibm.com/support/knowledgecenter/SS9RXT\\_10.2.2/com.ibm.swg.ba.cognos.ctm1.doc/welcome.html](http://www-01.ibm.com/support/knowledgecenter/SS9RXT_10.2.2/com.ibm.swg.ba.cognos.ctm1.doc/welcome.html))。

---

## SAS 匯出節點

SPSS Modeler Professional 和 SPSS Modeler Premium 中提供了此功能。

使用 SAS 匯出節點可以 SAS 格式寫入資料，以將資料讀取入 SAS 套裝軟體或 SAS 相容套裝軟體。可以使用三種 SAS 檔案格式進行匯出：SAS for Windows/OS2、SAS for UNIX 或 SAS。

### SAS 匯出節點「匯出」標籤

**匯出檔案。** 指定檔案的名稱。輸入檔名，或按一下檔案選擇器按鈕以瀏覽至該檔案的位置。

**匯出。** 指定匯出檔案格式。選項是 **SAS for Windows/OS2**、**SAS for UNIX** 或 **SAS Version 7/8/9**。

**匯出欄位名稱為變數稱。** 選取從 IBM SPSS Modeler 匯出欄位名稱為變數稱和標籤以供 SAS 使用的選項。

- **名稱。** 選取此選項可同時匯出 IBM SPSS Modeler 欄位名稱和欄位標籤。名稱將以 SAS 變數名稱方式匯出，而標籤則以 SAS 變數標籤方式匯出。
- **將名稱用作變數標籤。** 選中此選項可將 IBM SPSS Modeler 欄位名稱用作 SAS 中的變數標籤。IBM SPSS Modeler 容許在欄位名稱中包含不適用於 SAS 變數名稱的字元。為了防止創建無效的 SAS 名稱，請選取名稱和變數標籤。

**為此資料產生匯入節點。** 選中此選項可自動產生將讀取已匯出資料檔案的 SAS 來源節點。請參閱第 37 頁的『SAS 來源節點』主題，以取得更多資訊。

註：容許的字串長度上限是 255 個位元組。如果字串超過 255 個位元組，則會在匯出時將其截斷。

---

## Excel 匯出節點

Excel 匯出節點以 Microsoft Excel .xlsx 格式輸出資料。也可以選擇在執行節點時自動啟動 Excel 並開啟匯出的檔案。

### Excel 節點「匯出」標籤

**檔案名稱。** 輸入檔名，或按一下檔案選擇器按鈕以瀏覽至該檔案的位置。預設檔名為 *excelxp.xlsx*。

**檔案類型。** 受支援 Excel .xlsx 檔案格式。

**建立新檔案。** 建立新的項目 Excel 檔案。

**插入現有檔案。** 內容從從 **Cell** 開始欄位指定的 Cell 開始取代。試算表中的其他單元保留其原始內容。

**包括欄位名稱。** 指定欄位名稱是否應該包含在工作表的第一列中。

從 **Cell** 開始。用於首個匯出記錄（若已勾選了**包含欄位名稱**，則為首個欄位名稱）的 **Cell** 位置。資料填入右側並從此初始 **Cell** 向下填充。

**選擇工作表。** 指定您要匯出資料的目標工作表。您可以依索引或依名稱識別工作表：

- **依據索引。** 如果您要建立新檔案，指定從 0 到 9 的數字，以識別您要匯出的目標工作表，開頭的 0 表示第一個工作表，1 表示第二個工作表，依此類推。如果在此位置已存在工作表，您可以使用 10 或更大的值。
- **依名稱。** 如果您要建立新檔案，指定用於工作表的名稱。如果您插入到現有檔案，若工作表存在則資料插入此工作表，否則將建立具有此名稱的新工作表。

**啟動 Excel。** 指定執行節點時是否在匯出檔案中自動啟動 Excel。請注意，以分散式模式執行 IBM SPSS Modeler Server 時，輸出結果將儲存至伺服器檔案系統中，並在用戶端上啟動 Excel，其中顯示已匯出檔案的副本。

**為此資料產生匯入節點。** 選取此選項可自動產生將讀取已匯出資料檔案的 Excel 來源節點。請參閱第 38 頁的『Excel 來源節點』主題，以取得更多資訊。

---

## 延伸匯出節點

使用「延伸匯出」節點，您可以執行 R 或 Python for Spark Script 以匯出資料。

### 延伸匯出節點 - 語法標籤

選取語法的類型 - **R** 或 **Python for Spark**。如需相關資訊，請參閱下列章節。當您的語法備妥時，可以按一下執行，以執行「延伸匯出」節點。

#### R 語法

**R 語法。** 您可以在此欄位中輸入或貼上用於分析資料的自訂 R Scripting 語法。

**轉換旗標欄位。** 指定如何處理旗標欄位。有兩個選項：**字串至因數**，**整數與實數至倍精準數**，以及**邏輯值 (True、False)**。如果您選取**邏輯值 (True、False)**，則會遺失旗標欄位的原始值。例如，如果欄位具有值「男性」與「女性」，則這些值會變更為 True 與 False。

**將遺漏值轉換為 R「無法使用」值 (NA)。** 當選取時，會將所有遺漏值轉換為 R NA 值。R 使用值 NA 來識別遺漏值。您使用的部分 R 函數可能具有一個引數，可用來控制資料包含 NA 時的函數行為。例如，函數可讓您選擇自動排除包含 NA 的記錄。如果未選取此選項，則任何遺漏值都會依現狀傳遞至 R，並可能在執行 R Script 時造成錯誤。

**將日期/時間欄位轉換為具有時區特殊控制項的 R 類別。** 選取此選項時，具有日期或日期時間格式的變數會轉換為 R 日期/時間物件。您必須選取下列其中一個選項：

- **R POSIXct。** 具有日期或日期時間格式的變數轉換為 R POSIXct 物件。
- **R POSIXlt (清單)。** 具有日期或日期時間格式的變數轉換為 R POSIXlt 物件。

註：POSIX 格式是進階選項。僅在您的 R Script 指定以需要這些格式的方式處理日期時間欄位時，才使用這些選項。POSIX 格式不適用於具有時間格式的變數。

#### Python 語法

**Python 語法。** 您可以在此欄位中輸入或貼上供資料分析使用的自訂 Python Scripting 語法。如需 Python for Spark 的相關資訊，請參閱 Python for Spark 及 Python for Spark Scripting。

## 延伸匯出節點 - 主控台輸出標籤

主控台輸出標籤包含「語法」標籤上的 R Script 或 Python for Spark Script 執行時接收的所有輸出（例如，如果使用 R Script，它顯示執行語法標籤上 R 語法欄位中的 R Script 時，從 R 主控台接收的輸出）。此輸出可能包括執行 R 或 Python Script 時產生的 R 或 Python 錯誤訊息或警告。輸出可以主要用於 Script 除錯。主控台輸出標籤也包含 R 語法或 Python 語法欄位中的 Script。

每次執行延伸匯出 Script 時，都會使用從 R 主控台或 Python for Spark 接收的輸出來改寫主控台輸出標籤的內容。無法編輯輸出。

---

## XML 匯出節點

XML 匯出節點允許您以使用 UTF-8 編碼的 XML 格式輸出資料。您可以選擇性地建立 XML 來源節點以將匯出的資料讀回串流。

**XML 匯出檔案。**您要匯出資料的目標 XML 檔案的完整路徑和檔案名稱。

**使用 XML 綱目。**如果您要使用綱目或 DTD 來控制匯出資料的結構，請選取此勾選框。這將啟動下方面所說明的對映按鈕。

如果您不使用綱目或 DTD，則對匯出資料使用下列預設結構：

```
<records>
  <record>
    <fieldname1>value</fieldname1>
    <fieldname2>value</fieldname2>
    :
    <fieldnameN>value</fieldnameN>
  </record>
  <record>
    :
    :
  </record>
  :
  :
</records>
```

欄位名稱中的空格用底線取代；例如，"My Field" 將成為 <My\_Field>。

**對映。**如果您選擇使用 XML 綱目，該按鈕會開啟一個對話框，從中可以指定使用 XML 結構的哪個部分開始每個新記錄。請參閱第 324 頁的『XML 對映記錄選項』主題，以取得更多資訊。

**對映欄位。**表示已對映的欄位個數。

**為此資料產生匯入節點。**選中此選項可自動產生一個 XML 來源節點，該節點會將已匯出資料檔案讀取回到串流中。請參閱第 39 頁的『XML 來源節點』主題，以取得更多資訊。

## 寫入 XML 資料

當指定 XML 元素時，欄位值會放入元素標籤內：

```
<element>value</element>
```

當對映屬性時，欄位值會作為屬性值放置：

```
<element attribute= "value" >
```

如果欄位對映至 <records> 元素上方的元素，則欄位僅寫入一次，並作為所有記錄的常數。該元素的值將來自第一個記錄。

如果要寫入空值，那麼可通過指定空白內容來完成。對於元素，此為：

```
<element></element>
```

對於屬性，則為：

```
<element attribute="" >
```

## XML 對映記錄選項

「記錄」標籤允許您指定使用 XML 結構的哪個部分來開始每個新記錄。要正確對映到綱目，您需要指定記錄分隔符號。

**XML 結構。** 顯示前面畫面中指定的 XML 綱目的結構的層級樹狀結構。

**記錄 (XPath 表示式)。** 要設定記錄分隔符號，請選取 XML 結構中的元素，然後按一下右箭頭按鈕。每次在來源資料中遇到此元素時，都將在輸出檔案中建立新的記錄。

注意：如果您選取了 XML 結構中的根元素，那麼只能寫入單個記錄，所有其他記錄都會跳過。

## XML 對映欄位選項

當使用綱目檔時，「欄位」標籤可用於將資料集中的欄位對映到 XML 結構中的元素或屬性。

只要元素或屬性名稱是唯一的，就會自動對映與元素或屬性名相符的欄位名稱。因此，如果同時存在名為 field1 的元素和屬性，則不會自動對映。如果在結構中只有一個名為 field1 的項目，則自動對映在串流中具有此名稱的欄位。

**欄位。** 模型中的欄位清單。選取一個或多個欄位作為對映的來源部分。您可以使用清單底部的按鈕選取所有欄位，或具有特定測量層次的所有欄位。

**XML 結構。** 選取 XML 結構中的元素作為對映目標。要建立對映，請按一下「對映」。然後將顯示對映。已通過此方式對映的欄位個數顯示在清單下方。

要刪除對映，選取 XML 結構清單中的項目，然後按一下解除對映。

**顯示屬性。** 顯示或隱藏 XML 結構中的 XML 元素的屬性（如果有）。

## XML 對映預覽

在「預覽」標籤上，按一下更新以檢視將寫入的 XML 的預覽。

如果對映不正確，傳回到「記錄」或「欄位」標籤以糾正錯誤，然後再次按一下更新以請參閱結果。

---

## 「一般匯出」節點標籤

通過按一下對應的標籤可以為所有匯出節點指定下列選項：

- 「發行」標籤。用於發行串流的結果。
- 「註解」標籤。用於所有節點，此標籤提供一些選項來重新命名節點、提供自訂工具提示，以及儲存長註釋。



## 發佈串流

使用任何標準匯出節點，直接從 IBM SPSS Modeler 完成發佈串流：「資料庫」、「純文字檔」、「Statistics 匯出」、「延伸匯出」、「資料收集匯出」、「SAS 匯出」、Excel 及「XML 匯出」節點。匯出節點的類型會判定每一次使用 IBM SPSS Modeler Solution Publisher Runtime 或外部應用程式執行發佈的串流時，結果的寫入格式。例如，如果您想要在每一次執行發佈的串流時，將結果寫入資料庫，請使用「資料庫」匯出節點。

## 若要發佈串流

1. 以正常方式開啟或建置一個串流，並在結尾附加匯出節點。
2. 在匯出節點的「發佈」標籤上，指定已發佈檔案的 root 名稱（即，將附加 .pim、.par 及 .xml 副檔名的檔名）。
3. 按一下發佈以發佈串流，或者選取發佈串流，以在每一次執行節點時自動發佈串流。

**發佈的名稱。**指定所發佈映像檔及參數檔的 root 名稱。

- **映像檔 (\*.pim)** 提供「執行時期」執行所發佈串流需要的所有資訊，與匯出時完全一樣。如果您有信心不需要變更串流的任何設定（例如輸入資料來源或輸出資料檔），則可以僅部署映像檔。
- **參數檔 (\*.par)** 包含資料來源、輸出檔及執行選項的可配置資訊。如果您想要能夠控制串流的輸入或輸出，而無需重新發佈串流，您將需要參數檔，以及映像檔。
- **meta 資料檔 (\*.xml)** 說明映像檔的輸入及輸出，以及其資料模型。它設計為由應用程式使用，其中內含執行時期程式庫，並需要瞭解輸入及輸出資料的結構。

註：僅當您選取發佈 **meta** 資料選項時，才產生此檔案。

**發佈參數。**如果需要，您可以在 \*.par 檔中包括串流參數。您可以在透過編輯 \*.par 檔或透過執行時期 API 執行映像檔時，變更這些串流參數值。

此選項會啟用參數按鈕。當您按一下按鈕時，會顯示「發佈的參數」對話框。

在發佈直欄中選取相關選項，選擇要在所發佈映像檔中包括的參數。

**串流上執行。**指定執行節點時是否自動發佈串流。

- **匯出資料。**以標準方式執行匯出節點，而不發佈串流。（基本上，節點利用與 IBM SPSS Modeler Solution Publisher 無法使用時相同的方式，在 IBM SPSS Modeler 中執行。）如果您選取此選項，則不會發佈該串流，除非您在匯出節點對話框中按一下**發佈**來明確這樣做。或者，您可以使用工具列上的「發佈」工具，或者透過使用 Script，發佈現行串流。
- **發佈串流。**使用 IBM SPSS Modeler Solution Publisher 發佈串流以進行部署。如果您想要在每次執行時自動發佈串流，請選取此選項。

註：

- 如果您計劃使用新的或已更新的資料來執行所發佈的串流，請務必注意，輸入檔中的欄位順序必須與所發佈串流中指定之來源節點輸入檔中的欄位順序相同。
- 發佈至外部應用程式時，請考量過濾無關的欄位或重新命名欄位，以符合輸入需求。可以在匯出節點之前使用「過濾器」節點，從而達成兩者。



---

## 第 8 章 IBM SPSS Statistics 個節點

---

### IBM SPSS Statistics 節點 - 概觀

作為 IBM SPSS Modeler 及其資料採礦功能的補充，IBM SPSS Statistics 允許您進一步執行統計分析和資料管理。

安裝 IBM SPSS Statistics 的相容、受軟體使用權副本後，您可以從 IBM SPSS Modeler 與它連接，並執行 IBM SPSS Modeler 不支援的複合、多步資料操作與分析。對於進階使用者，還提供了一個使用指令語法對分析進行進一步修改的選項。請參閱「版本注意事項」以瞭解關於版本相容性的資訊。

如果可用，IBM SPSS Statistics 節點顯示在節點選用區的專門部分中。

註：建議您在使用 IBM SPSS Statistics 的「轉換」、「模型」和「輸出」節點之前，在「類型」節點中實例化資料。使用 AUTORECODE 語法指令時，這還是一項要求。

IBM SPSS Statistics 選用區包含下列節點：



Statistics 檔案節點從 IBM SPSS Statistics 使用的 *.sav* 或 *.zsav* 檔案格式以及儲存在 IBM SPSS Modeler 中的快取檔（也使用同一格式）讀取資料。



Statistics 轉換節點針對 IBM SPSS Modeler 中的資料來源執行所選的 IBM SPSS Statistics 語法指令。此節點需要 IBM SPSS Statistics 的授權副本。



Statistics 模型節點使您能夠通過執行生成 PMML 的 IBM SPSS Statistics 程序分析和處理資料。此節點需要 IBM SPSS Statistics 的授權副本。



Statistics 輸出節點可呼叫 IBM SPSS Statistics 程序以分析您的 IBM SPSS Modeler 資料。可以存取許多不同的 IBM SPSS Statistics 分析程序。此節點需要 IBM SPSS Statistics 的授權副本。



Statistics 匯出節點以 IBM SPSS Statistics *.sav* 或 *.zsav* 格式輸出資料。IBM SPSS Statistics Base 和其他產品可以讀取 *.sav* 或 *.zsav* 檔案。這種格式也用於 IBM SPSS Modeler 中的某些快取檔。

註：如果您的 SPSS Statistics 副本僅授權給單一使用者使用，而您執行的串流帶有兩個或兩個以上分支，並且每個分支均包含 SPSS Statistics 節點，那麼您可能會得到軟體使用權授權錯誤。當某個分支的 SPSS Statistics 階段作業尚未結束，而另一個分支試圖起動時，就會出現此錯誤。如可行，應重新設計串流，以確保帶有 SPSS Statistics 節點的多個分支不會同時執行。

---

## Statistics 檔案節點

可以使用 Statistics 檔案節點從已儲存的 IBM SPSS Statistics 檔案 (.sav 或 .zsav) 中直接讀取資料。現在可使用該格式取代 IBM SPSS Modeler 早期版本中的快取檔。如果想要匯入已儲存的快取檔，則應使用 IBM SPSS Statistics 檔案節點。

**匯入檔案。** 指定檔案的名稱。您可以輸入檔名或按一下省略符號按鈕 (...) 以選取檔案。一旦選定了一個檔案，即可顯示此檔案的路徑。

**檔案已通過密碼進行加密。** 如果您知道該檔案受密碼保護的，請選中此框；系統將提示您輸入密碼。如果該檔案受密碼保護的，但您未輸入密碼，那麼在試圖切換至另一標籤、重新整理資料、預覽節點內容或試圖執行包含節點的串流時，將顯示一條警告訊息。

註：受密碼受保護的檔案只能由 IBM SPSS Modeler 第 16 版或更高版本開啟。

**變數名稱。** 選取從 IBM SPSS Statistics .sav 或 .zsav 檔案匯入變數名稱和標籤時的處理方法。您選擇要在這裡包括的 meta 資料會存留於 IBM SPSS Modeler 中的整個工作中，可以再次匯出以便在 IBM SPSS Statistics 中使用。

- **讀取姓名與標籤。** 選中此選項可將變數名稱和標籤同時讀取入 IBM SPSS Modeler。依預設，此選項已選取，且變數名稱在「類型」節點中顯示。根據串流內容對話框中指定的選項，標籤將顯示在圖表、模型瀏覽器和其他類型的輸出中。依預設，將禁止在輸出中顯示標籤。
- **讀取用作名稱的標籤。** 選取從 IBM SPSS Statistics .sav 或 .zsav 檔案中讀取描述性統計量變數標籤（而不是短欄位名稱），並在 IBM SPSS Modeler 中將這些標籤用作變數名稱。

**值。** 選取從 IBM SPSS Statistics .sav 或 .zsav 檔案匯入值與標籤時的處理方法。您選擇要在這裡包括的 meta 資料會存留於 IBM SPSS Modeler 中的整個工作中，可以再次匯出以便在 IBM SPSS Statistics 中使用。

- **讀取資料與標籤。** 選中此選項可將實際值和值標籤同時讀取入 IBM SPSS Modeler。依預設將已選取此選項，並且這些值本身將顯示在「類型」節點中。根據串流內容對話框中指定的選項，值標籤將顯示在表示式建置器、圖表、模型瀏覽器和其他類型的輸出中。
- **讀取用作資料的標籤。** 如果要使用 .sav 或 .zsav 檔案中的值標籤而不是用於代表值的數值或符號代碼，請選中此選項。例如，對於含性別欄位（其值 1 和 2 實際上分別代表陽性和陰性）的資料，選中此選項可將該欄位轉換為字串，並將陽性和陰性作為實際值匯入。

選中此選項前考慮 IBM SPSS Statistics 資料中的遺漏值非常重要。例如，如果數值欄位僅對遺漏值使用標籤 (0 = *No Answer*, -99 = *Unknown*)，則選中上述選項將僅匯入值標籤 *No Answer* 和 *Unknown*，並會將欄位轉換為字串。在這種情況下，應在類型節點中匯入值本身並設定遺漏值。

**使用欄位格式資訊確定儲存類型。** 如果取消勾選此方框，那麼將使用整數儲存匯入 .sav 檔案中格式化為整數的欄位值（即，在 IBM SPSS Statistics 的「變數視圖」中指定為 *Fn.0* 的欄位）。將除字串外的所有其他欄位值作為實數匯入。

如果已選取了此框（預設情況），那麼無論是否已在 .sav 檔案中格式化為整數，除字串以外的所有欄位值都將作為實數匯入。

多個回應集。匯入檔案後，IBM SPSS Statistics 檔案中定義的任何複選題分析集都將自動被保留。您可以使用「過濾器」標籤來檢視並編輯任何節點中的回應集。請參閱第 131 頁的『編輯多重回應集』主題，以取得更多資訊。

---

## Statistics 轉換節點

使用 Statistics 變換節點，可以使用 IBM SPSS Statistics 指令語法完成資料變換。這樣便有可能完成 IBM SPSS Modeler 不受支援的若干轉換，並實現複合多步轉換的自動化，包含通過單一節點建立多個欄位。這種節點與 Statistics 輸出節點類似，只是資料將傳回 IBM SPSS Modeler 進行進一步的分析，而輸出節點中的資料將作為已要求的輸出物件（如圖形或表格）傳回。

要使用此節點，必須在電腦上安裝 IBM SPSS Statistics 的相容版本並軟體使用權使用。請參閱第 298 頁的『IBM SPSS Statistics 說明應用程式』主題，以取得更多資訊。請參閱版本注意事項以獲得有關相容性的資訊。

必要的話，您可以使用「過濾器」標籤來過濾或重新命名欄位，以使其符合 IBM SPSS Statistics 命名標準。請參閱第 316 頁的『重新命名或過濾 IBM SPSS Statistics 的欄位』主題，以取得更多資訊。

語法參照。有關特定的 IBM SPSS Statistics 程序的詳細資料，請參閱《IBM SPSS Statistics 指令語法參照》手冊，該文檔包含在您的 IBM SPSS Statistics 軟體副本中。要從「語法」標籤上檢視該手冊，請選擇語法編輯程式選項，並按一下「啟動 IBM SPSS Statistics 語法說明」按鈕。

注意：此節點並不支援所有 IBM SPSS Statistics 語法。請參閱『容許的語法』主題，以取得更多資訊。

### Statistics 轉換節點 - 「語法」標籤

IBM SPSS Statistics 對話框選項

如果不熟悉某個程序的 IBM SPSS Statistics 語法，那麼在 IBM SPSS Modeler 中建立語法的最簡單方式是選擇 **IBM SPSS Statistics 對話框** 選項，選擇該程序的對話框，完成對話框並按一下「確定」。這樣可以將語法放入當前在 IBM SPSS Modeler 中使用的 IBM SPSS Statistics 節點的「語法」標籤中。然後，可以執行串流以獲得程序輸出。

IBM SPSS Statistics 語法編輯程式選項

檢查。在對話框上面已輸入語法指令後，使用此按鈕可驗證您的已輸入。所有不正確的語法將在對話框下面標示。

為確保檢查過程不會過長，當您驗證語法時，會對資料的代表性樣本進行檢查（而不是檢查整個資料集），以確保輸入有效。

### 容許的語法

如果有大量繼承自 IBM SPSS Statistics 的語法或熟悉 IBM SPSS Statistics 的資料預備功能，您可以使用 Statistics 變換節點執行很多現有變換。提示：使用該節點可按可預測的方式轉換資料 - 例如，通過執行迴圈指令或通過對資料進行更改、新增、排序、過濾或選取。

可使用的指令範例如下：

- 根據二項式分佈計算亂數：

```
COMPUTE newvar = RV.BINOM(10000,0.1)
```

- 將某個變數重新編碼為新變數：

```
RECODE Age (Lowest thru 30=1) (30 thru 50=2) (50 thru Highest=3) INTO AgeRecoded
```

- 取代遺漏值：

RMV Age\_1=SMEAN(Age)

下方表列出了 Statistics 轉換節點所受支援的 IBM SPSS Statistics 語法。

#### 指令名稱

ADD VALUE LABELS  
APPLY DICTIONARY  
AUTORECODE  
BREAK  
CD  
CLEAR MODEL PROGRAMS  
CLEAR TIME PROGRAM  
CLEAR TRANSFORMATIONS  
COMPUTE  
COUNT  
CREATE  
DATE  
DEFINE-!ENDDFIN  
DELETE VARIABLES  
DO IF  
DO REPEAT  
ELSE  
ELSE IF  
END CASE  
END FILE  
END IF  
END INPUT PROGRAM  
END LOOP  
END REPEAT  
EXECUTE  
FILE HANDLE  
FILE LABEL  
FILE TYPE-END FILE TYPE  
FILTER  
FORMATS  
IF  
INCLUDE  
INPUT PROGRAM-END INPUT PROGRAM  
INSERT  
LEAVE  
LOOP-END LOOP  
MATRIX-END MATRIX

## 指令名稱

MISSING VALUES

N OF CASES

NUMERIC

PERMISSIONS

PRESERVE

RANK

RECODE

RENAME VARIABLES

RESTORE

RMV

SAMPLE

SELECT IF

SET

SORT CASES

字串

  SUBTITLE

TEMPORARY

TITLE

UPDATE

V2C

VALIDATEDATA

VALUE LABELS

VARIABLE ATTRIBUTE

VARSTOCASES

VECTOR

---

## Statistics 模型節點

Statistics 模型節點使您能夠通過執行生成 PMML 的 IBM SPSS Statistics 程序分析和處理資料。然後，建立的模型塊可按常規方式在 IBM SPSS Modeler 串流中進行評分等操作。

要使用此節點，必須在電腦上安裝 IBM SPSS Statistics 的相容版本並軟體使用權使用。請參閱第 298 頁的『IBM SPSS Statistics 說明應用程式』主題，以取得更多資訊。請參閱版本注意事項以獲得有關相容性的資訊。

可用的 IBM SPSS Statistics 分析程序取決於您的授權類型。

### Statistics 模型節點 - 「模型」標籤

**模型名稱** 您可以根據目標或 ID 欄位（如果未指定此類欄位，則根據模型類型）自動產生模型名稱，或者指定自訂名稱。

**選取一個對話框...框。** 按一下以顯示您可以選取並執行的可用 IBM SPSS Statistics 程序清單。此清單僅顯示那些生成 PMML 且您已獲得軟體使用權的程序，並且不包含使用者編寫的程序。

1. 按一下所需程序；顯示相關 IBM SPSS Statistics 對話框。

2. 在 IBM SPSS Statistics 對話框中輸入程序詳細資料。
3. 按一下**確定**傳回到 Statistics 模型節點；在「模型」標籤中顯示 IBM SPSS Statistics 語法。
4. 在任何時候，要傳回到 IBM SPSS Statistics 對話框，例如，要修改您的查詢，請按一下程序選擇按鈕右側的 IBM SPSS Statistics 對話框顯示按鈕。

## Statistics 模型節點 - 模型塊彙總

在執行 Statistics 模型節點時，它執行相關的 IBM SPSS Statistics 程序並建立您可在 IBM SPSS Modeler 串流中進行評分的模型塊。

模型塊的「摘要」標籤顯示有關欄位、建構設定值和模型估計過程的資訊。結果呈現在樹狀結構視圖中，可以按一下特定項目來展開或收合該視圖。

檢視模式按鈕以 IBM SPSS Statistics 輸出檢視器的修改形式顯示結果。有關該檢視器的進一步資訊，請參閱 IBM SPSS Statistics 文件。

「檔案」功能表中提供了常用的匯出和列印選項。請參閱第 265 頁的『檢視輸出』主題，以取得更多資訊。

---

## Statistics 輸出節點

Statistics 輸出節點可呼叫 IBM SPSS Statistics 程序以分析您的 IBM SPSS Modeler 資料。您可以在瀏覽器視窗中檢視結果，或以 IBM SPSS Statistics 輸出檔案格式儲存結果。從 IBM SPSS Modeler 中，可以存取許多不同的 IBM SPSS Statistics 分析程序。

要使用此節點，必須在電腦上安裝 IBM SPSS Statistics 的相容版本並軟體使用權使用。請參閱第 298 頁的『IBM SPSS Statistics 說明應用程式』主題，以取得更多資訊。請參閱版本注意事項以獲得有關相容性的資訊。

必要的話，您可以使用「過濾器」標籤來過濾或重新命名欄位，以使其符合 IBM SPSS Statistics 命名標準。請參閱第 316 頁的『重新命名或過濾 IBM SPSS Statistics 的欄位』主題，以取得更多資訊。

**語法參照。** 有關特定的 IBM SPSS Statistics 程序的詳細資料，請參閱《IBM SPSS Statistics 指令語法參照》手冊，該文檔包含在您的 IBM SPSS Statistics 軟體副本中。要從「語法」標籤上檢視該手冊，請選擇**語法編輯程式**選項，並按一下「啟動 IBM SPSS Statistics 語法說明」按鈕。

## Statistics 輸出節點 - 「語法」標籤

使用此標籤為要用於分析您的資料的 SPSS Statistics 程序建立語法。語法由兩個部分組成：**陳述式**和**關聯的選項**。陳述式指定要執行的分析或作業和要使用的欄位。選項指定所有其他內容，包含要顯示的統計資料、要儲存的衍生欄位，等等。

## SPSS Statistics 對話框選項

如果不熟悉某個程序的 IBM SPSS Statistics 語法，那麼在 IBM SPSS Modeler 中建立語法的最簡單方式是選擇 **IBM SPSS Statistics 對話框**選項，選擇該程序的對話框，完成對話框並按一下「確定」。這樣可以將語法放入當前在 IBM SPSS Modeler 中使用的 IBM SPSS Statistics 節點的「語法」標籤中。然後，可以執行串流以獲得程序輸出。

可以選擇性地產生一個「Statistics 檔案」來源節點，用於匯入產生的資料。此節點非常有用，例如，除顯示輸出之外，某個程序將欄位（例如分數）寫入作用中資料集。

註：

- 當以非英文語言產生輸出時，建議在語法中指定該語言。



- 「統計資料輸出」節點中不支援「輸出樣式」選項。

## 建立語法

1. 按一下**選取一個對話框...框**按鈕。
2. 選擇其中一個選項：
  - **分析** 列出 SPSS Statistics 分析功能表的內容；選擇您要使用的程序。
  - **其他** 如果顯示，則列出在 SPSS Statistics 的自訂對話框建置器中建立的對話框，以及未出現在「分析」功能表上且您具有許可的任何其他 SPSS Statistics 對話框。如果沒有適用的對話框，那麼將不顯示此選項。

註：不會顯示「自動資料準備」對話框。

如果某個 SPSS Statistics 自訂對話框建立有新欄位，則這些欄位不能在 SPSS Modeler 中使用，因為 Statistics 輸出節點為終端節點。

選擇性地勾選**為產生的資料產生匯入節點方框**，以建立可用來將產生的資料匯入另一個串流的「統計資料檔案」來源節點。節點置於螢幕畫布上，資料包含在**檔案欄位**（預設位置是 SPSS Modeler 安裝目錄）中指定的 .sav 檔案中。

## 語法編輯程式選項

要儲存針對頻繁使用的程序建立的語法，請執行下列操作：

1. 按一下**檔案選項**按鈕（工具列上的第一個按鈕）。
2. 從功能表中選擇**儲存或另存新檔**。
3. 將檔案儲存為 .sps 檔案。

要使用先前建立的語法檔案，並同時取代語法編輯器的目前內容（如果有），請執行下列操作：

1. 按一下「**檔案選項**」按鈕（工具列上第一個）。
2. 從功能表中選擇**開啟**。
3. 選取 .sps 檔案以將其內容貼上到「輸出」節點「語法」標籤中。

要插入先前儲存的語法而不取代目前內容，請執行下列操作：

1. 按一下「**檔案選項**」按鈕（工具列上第一個）。
2. 從功能表中選擇**插入**。
3. 選取 .sps 檔案以將其內容貼上到「輸出」節點的游標指定位置。

選擇性地勾選**為產生的資料產生匯入節點方框**，以建立可用來將產生的資料匯入另一個串流的「統計資料檔案」來源節點。節點置於螢幕畫布上，資料包含在**檔案欄位**（預設位置是 SPSS Modeler 安裝目錄）中指定的 .sav 檔案中。

按一下執行時，結果會顯示在 SPSS Statistics 輸出檢視器中。有關檢視器的進一步資訊，請參閱 SPSS Statistics 文件。

註：不支援下列項目（及其對應SPSS Statistics 對話框選項）的語法。它們不會影響輸出。

- OUTPUT ACTIVATE
- OUTPUT CLOSE
- OUTPUT DISPLAY
- OUTPUT EXPORT

- OUTPUT MODIFY
- OUTPUT NAME
- OUTPUT NEW
- OUTPUT OPEN
- OUTPUT SAVE

## Statistics 輸出節點 - 「輸出」標籤

使用「輸出」標籤，可以指定輸出的格式和位置。您可以選擇將結果顯示在畫面上，也可以將其傳送給其中一種可用字段類型。

**輸出名稱。**指定執行節點時產生的輸出名稱。自動會根據產生輸出的節點來選擇名稱。您可以選擇性地選取自訂以指定不同的名稱。

**輸出至螢幕**（預設選項）。建立輸出物件以在線上檢視。執行輸出節點時，輸出物件將在管理程式視窗的「輸出」標籤上出現。

**輸出至檔案。**執行節點時將輸出儲存到檔案。如果選擇此選項，請在檔名欄位中輸入檔名（或導覽到某目錄，並使用檔案選擇器按鈕指定檔名）並選擇檔案類型。

**檔案類型。**選擇您要傳送輸出的目標檔案類型。

- **HTML 文件(\*.html)**。以 HTML 格式寫入輸出。
- **IBM SPSS Statistics 檢視器檔案 (\*.spv)**。以可由 IBM SPSS Statistics 輸出檢視器讀取的格式寫入輸出。
- **IBM SPSS Statistics Web 報告檔案 (\*.spw)**。以 IBM SPSS Statistics Web 報告格式寫入輸出，此類文件可以發行到 IBM SPSS Collaboration and Deployment Services 儲存庫，隨後在 Web 瀏覽器中檢視。請參閱第 265 頁的『發行至網路』主題，以取得更多資訊。

註：如果選取輸出至螢幕，那麼 IBM SPSS Statistics OMS 指令 VIEWER=NO 將不起作用；另外，Scripting API（基本和 *Python SpssClient* 模組）將在 IBM SPSS Modeler 中無法使用。

---

## Statistics 匯出節點

Statistics 匯出節點可以以 IBM SPSS Statistics *.sav* 格式匯出資料。IBM SPSS Statistics Base 和其他模組能夠讀取 IBM SPSS Statistics *.sav* 檔案。這也是 IBM SPSS Modeler 快取檔所使用的格式。

有時候，將 IBM SPSS Modeler 欄位名稱對映到 IBM SPSS Statistics 變數名稱時會出錯，這是因為 IBM SPSS Statistics 變數名稱被限制為 64 個字元且不能包含某些字元，比如空格、美元符號 (\$) 和 劃線 (-) 等。有兩種方法可以規避上述限制：

- 可通過按一下「過濾」標籤，根據 IBM SPSS Statistics 變數名稱要求更名欄位。請參閱第 316 頁的『重新命名或過濾 IBM SPSS Statistics 的欄位』主題，以取得更多資訊。
- 選擇同時從 IBM SPSS Modeler 匯出欄位名稱和標籤。

注意：IBM SPSS Modeler 以 Unicode UTF-8 格式寫入 *.sav* 檔案。IBM SPSS Statistics 16.0 版以後的版本只支援 Unicode UTF-8 格式的檔案。為了防止資料可能損壞，使用 Unicode 編碼儲存的 *.sav* 檔案不得用於 16.0 之前的 IBM SPSS Statistics 版本。有關詳細資訊，請參閱 IBM SPSS Statistics 說明。

**多個回應集。**匯出檔案時，將自動保留定義在串流中的所有多重回應集。您可以使用「過濾器」標籤來檢視並編輯任何節點中的回應集。請參閱第 131 頁的『編輯多重回應集』主題，以取得更多資訊。

## Statistics 匯出節點 - 「匯出」標籤

**匯出檔案** 指定檔案的名稱。輸入檔名，或按一下檔案選擇器按鈕以瀏覽至該檔案的位置。

**檔案類型** 請選取是以正常的 *.sav* 格式還是壓縮的 *.zsav* 格式儲存檔案。

**使用密碼對加密檔案** 要使用密碼保護檔案，請選中此框；系統將提示您在另一對話框中輸入並確認密碼。

註：受密碼受保護的檔案只能由 SPSS Modeler 16 版或更高版本，或者由 SPSS Statistics 21 版或更高版本開啟。

**匯出欄位名稱為變數稱** 指定將變數名稱和標籤從 SPSS Modeler 匯出至 SPSS Statistics *.sav* 或 *.zsav* 檔案時的處理方法。

- **名稱** 選取此選項可同時匯出 SPSS Modeler 欄位名稱和欄位標籤。名稱將以 SPSS Statistics 變數名方式匯出，而標籤則以 SPSS Statistics 變數標籤方式匯出。
- **作為變數標籤的名稱** 選中此選項可將 SPSS Modeler 欄位名稱用作 SPSS Statistics 中的變數標籤。SPSS Modeler 容許在欄位名稱中包含不適用於 SPSS Statistics 變數名稱的字元。為了防止建立無效的 SPSS Statistics 名稱，請選取將名稱用作變數標籤或使用「過濾」標籤調整欄位名。

**啟動應用程式** 如果電腦上安裝了 SPSS Statistics，則可以選取此選項對已儲存資料檔案直接激活應用程式。Helper 應用程式對話框中必須指定用於啟動應用程式的選項。請參閱第 298 頁的『IBM SPSS Statistics 說明應用程式』主題，以取得更多資訊。要在不開啟外部程式的情況下直接建立 SPSS Statistics *.sav* 或 *.zsav* 檔案，請取消選取此選項。

註：當以伺服器（分散式）方式同時執行 SPSS Modeler 和 SPSS Statistics 時，寫入出資料並啟動 SPSS Statistics 階段作業不會自動開啟 SPSS Statistics 用戶端以顯示讀取作用中的資料集的資料集。暫行解決方法是 SPSS Statistics 用戶端啟動時立刻在其中開啟資料檔案。

**為該資料產生匯入節點** 選中此選項可自動產生將讀取已匯出資料檔案的「Statistics 檔案」來源節點。請參閱第 27 頁的『Statistics 檔案節點』主題，以取得更多資訊。

## 重新命名或過濾 IBM SPSS Statistics 的欄位

在將資料從 IBM SPSS Modeler 匯出或部署至外部應用程式（例如 IBM SPSS Statistics）時，可能需要重新命名或調整欄位名稱。「統計資料轉換」、「統計資料輸出」及「統計資料匯出」對話框包含一個「過濾器」標籤，以協助此處理程序。

「過濾器」標籤功能的基本說明在其他部分中討論。請參閱第 129 頁的『設定過濾選項』主題，以取得更多資訊。本主題提供將資料讀取至 IBM SPSS Statistics 的提示。

若要調整欄位名稱以符合 IBM SPSS Statistics 命名慣例：

1. 在「過濾器」標籤上，按一下「過濾器選項功能表」工具列按鈕（工具列上的第一個）。
2. 選取「重新命名 IBM SPSS Statistics」。
3. 在「重新命名 IBM SPSS Statistics」對話框上，您可以選擇將檔名中的無效字元取代為井字號 (#) 字元或底線 (\_)。

**重新命名多重回應集**。如果您想要調整多重回應集的名稱，這些回應集可以利用「統計資料檔案」來源節點匯入至 IBM SPSS Modeler，請選取此選項。它們用來記錄資料，針對每一個案例可以具有多個值，例如在意見調查回應中。



---

## 第 9 章 SuperNode

---

### SuperNode 概觀

IBM SPSS Modeler 可視化程式設計介面容易學習的其中一個原因是，每個節點都有明確定義的功能。但是，要進行複合正在處理，可能需要一個較長的節點序列。最後，這可能會使串流畫布混亂不堪，並導致難以遵循串流圖。要避免長而複合的串流混亂不堪，有兩種方法：

- 您可以將正在處理序列分割為幾個相互融匯的串流。例如，第一個串流建立第二個串流作為輸入使用的資料檔案。第二個流建立第三個流作為輸入使用的檔案，依此類推。您可以通過將多個串流儲存在專案中來對它們進行管理。專案提供多個串流及其輸出的組織。但是，專案檔案只包含對它所包含的物件的參照，並且仍舊有多個串流檔案需要您進行管理。
- 處理複合的串流過程時，可以建立有著更為清晰的處理程序的 **SuperNode** 作為替代方案。

SuperNode 通過封裝資料串流的群組成部分將多個節點群組成一個節點。這會為資料挖掘程序提供許多便利：

- 串流更加整潔並且更易管理。
- 節點可以合併為業務特定的 SuperNode。
- 可將 SuperNode 匯出到程式庫中以在多個資料採礦專案中重複使用。

---

### SuperNode 的類型

SuperNode 在資料串流中由星狀圖示代表。該圖示中具有陰影，用於代表 SuperNode 的類型以及串流必須串流入或串流出的方向。

存在三種類型的 SuperNode：

- 來源 SuperNode
- 處理 SuperNode
- 終端 SuperNode

#### 來源 SuperNode

來源 SuperNode 包含一個類似於普通來源節點的資料來源，並且可在可使用普通來源節點的任意位置使用。來源 SuperNode 的左端具有陰影，表示它在左端「已關閉」，並且資料從 SuperNode 流動到下游。

來源 SuperNode 只在右側具有一個連線點，表示資料離開 SuperNode 並串流向串流中。

#### 處理 SuperNode

處理 SuperNode 只包含處理節點，並且沒有陰影，這表示資料可流入這種類型的 SuperNode 和從這種類型的 SuperNode 流出。

處理 SuperNode 在左側和右側都有連線點，表示資料進入 SuperNode 並離開以串流回到串流中。雖然 SuperNode 可以包含附加串流段，甚至額外的串流，但這兩個連線點都必須通過一條連線開始串流點和結束串流點的路徑流動。

注意：處理 SuperNode 有時也稱為操作 SuperNode。

## 終端 SuperNode

終端 SuperNode 包含一個或多個終端節點（圖形、表格等）並且使用方式可與終端節點相同。終端 SuperNode 的右端具有陰影，表示它在右端「已關閉」並且資料只能流入入 終端 SuperNode 中。

終端 SuperNode 只在左端具有一個連線點，表示資料從串流進入 SuperNode。並在 SuperNode 內終止。

終端 SuperNode 也可以包含 Script，Script 用於指定 SuperNode 內所有終端節點的執行順序。請參閱第 343 頁的『SuperNode 和 Script 編寫』主題，以取得更多資訊。

---

## 建立 SuperNode

建立 SuperNode 時，通過將多個節點封裝為一個節點可「縮短」資料串流。一旦在畫布上建立或載入了串流，便有多種方法來建立 SuperNode。

### 多重選擇

建立 SuperNode 的最簡單方法是選取要封裝的所有節點：

1. 使用滑鼠選取串流畫布上的多個節點。也可以通過在按住 Shift 鍵的同時按一下來選取串流或串流的一部分。

註：您選取的節點必須來自連續或分出的串流。在某種程度上，您無法選取不相鄰或不連接的節點。

2. 然後，使用下面三種方法之一來封裝所選節點：
  - 按一下工具列上的 SuperNode 圖示（形狀像星狀）。
  - 用滑鼠右鍵按一下 SuperNode。然後從快速功能表中選擇：

**建立 SuperNode > 來自選項**

- 從 SuperNode 功能表中，選擇：

**建立 SuperNode > 來自選項**

所有這三個選項均將節點封裝到一個 SuperNode 中，該 SuperNode 具有陰影，用於根據它的內容反映它的類型：來源、過程或終端機。

### 單一選擇

也可以通過以下方式建立 SuperNode：選取一個節點並使用功能表選項確定 SuperNode 的開始和結束，或封裝所選節點的全部下游內容。

1. 按一下確定封裝開始的節點。
2. 從 SuperNode 功能表中，選擇：

**建立 SuperNode > 由此**

也可以通過選取串流部分的開始和結束來封裝節點，以更加交互的方式建立 SuperNode：

1. 按一下要包含在 SuperNode 中的第一個或最後一個節點。
2. 從 SuperNode 功能表中，選擇：

**建立 SuperNode > 選取...**

3. 或者，可以通過用滑鼠右鍵按一下所需的節點來使用快速功能表選項。

- 游標會變為 SuperNode 圖示，表示必須選取串流中的其他點。逆串流或順串流移動到 SuperNode 片段的「另一端」並按一下某節點。此動作將使用 SuperNode 星狀圖示取代所有節點。

註：您選取的節點必須來自連續或分出的串流。在某種程度上，您無法選取不相鄰或不連接的節點。

## 巢狀 SuperNode

SuperNode 可以巢套在其他 SuperNode 中。用於每類 SuperNode（來源、過程和終端機）的規則也套用至巢套 SuperNode。例如，具有巢狀的處理 SuperNode 必須具有通過所有巢狀 SuperNode 的連續資料流入，才能保持為處理 SuperNode。如果其中一個巢套 SuperNode 是終端節點，則資料將不再通過階層流入。

終端 SuperNode 和來源 SuperNode 可以包含其他類型的巢套 SuperNode。但用於建立 SuperNode 的基本規則同樣適用。

---

## 鎖定 SuperNode

一旦您已建立 SuperNode。您可以使用密碼將其鎖定，以防修正。例如，如果您建立串流或部分串流，您可以進行此操作，因為在您組織中其他人使用的固定值範本，這些人設定 IBM SPSS Modeler 查詢的經驗較少。

當鎖定 SuperNode 時，使用者仍然可以在「參數」標籤上為任何已經定義的參數輸入值，且不輸入密碼也可執行鎖定的 SuperNode。

註：不能使用 Script 執行鎖定和解除鎖定。

## 鎖定和解除鎖定 SuperNode

註：丟失的密碼不能回復。

您可以從三個標籤中的任何標籤鎖定或解除鎖定 SuperNode。

- 按一下**鎖定節點**。
- 輸入並確認密碼。
- 按一下**確定**。

受密碼受保護的 SuperNode 在串流畫布上通過 SuperNode 圖示左上角的小掛鎖符號 ID。

## 解除鎖定 SuperNode

- 要永久地刪除密碼保護，請按一下**解除鎖定節點**。將提示您輸入密碼。
- 輸入密碼並按一下**確定**。SuperNode 不再受密碼受保護的，掛鎖符號不再顯示在串流中圖示的旁邊。

對於在包含鎖定 SuperNode 的 SPSS Modeler 第 16 版和 17.0 版中儲存的串流，在其他環境中開啟串流（例如，在 IBM SPSS Collaboration and Deployment Services 中），或者在 Mac 上，在 SPSS Modeler 安裝的 JRE 不同時，首先必須使用上次儲存所在的舊環境上的 17.1 版或更高版本進行開啟、解除鎖定和重新儲存。

某些時候，在解除鎖定第 18 版之前的串流中的 SuperNode 時，將顯示不正確的密碼錯誤。要解決此問題，請在具有相同系統本端設定的相同平台上使用與上次開啟相同的 IBM SPSS Modeler 版本（或更新版本）重新開啟並解除鎖定節點。然後，在第 18 版或更高版本中進行開啟。解鎖節點，然後重新儲存串流。

## 編輯鎖定的 SuperNode

如果您試圖定義參數或放大以顯示鎖定的 SuperNode。將提示您輸入密碼。

輸入密碼並按一下**確定**。

您現在能夠根據需要隨時編輯參數定義並放大和縮小，直到您關閉 SuperNode 所在的串流。

注意這並不會刪除密碼保護，只容許您存取以處理 SuperNode。請參閱第 339 頁的『鎖定和解除鎖定 SuperNode』主題，以取得更多資訊。

---

## 編輯 SuperNode

一旦您已建立了 SuperNode。您可以通過放大更加仔細地檢查；如果 SuperNode 已鎖定，那麼系統將提示您輸入密碼。請參閱『編輯鎖定的 SuperNode』主題，以取得更多資訊。

要檢視 SuperNode 的內容，可以使用 IBM SPSS Modeler 工具列中的放大圖示或下列方法：

1. 用滑鼠右鍵按一下 SuperNode。
2. 從快速功能表中，選擇**放大**。

所選 SuperNode 的內容將在略有不同的 IBM SPSS Modeler 環境中顯示，其中的連接器顯示通過串流或串流段進行的資料流程。在串流畫布的此層次上，可以執行多項作業：

- 修改 SuperNode 類型：來源、過程或終端機。
- 建立參數或編輯參數的值。參數在 Script 和 CLEM 表示式中使用。
- 指定 SuperNode 及其子節點的快捷選項。
- 建立或修改 SuperNode Script（僅限終端 SuperNode）。

## 修改 SuperNode 類型

在一些情況下，改變 SuperNode 的類型會很有用。只有在放大 SuperNode 時，此選項才可用，並且它僅在該層次適用於 SuperNode。下表格對三種類型的 SuperNode 進行了說明。

表 51. SuperNode 的類型.

SuperNode 的類型	說明
來源 SuperNode	一個傳出的連線
處理 SuperNode	兩個連線：一個傳入，一個傳出
終端 SuperNode	一個傳入的連線

## 要變更 SuperNode 的類型

1. 確保放大 SuperNode。
2. 從 SuperNode 功能表中，選擇 **SuperNode 類型**，然後選擇類型。

## 添加註解和更名 SuperNode

您可以在 SuperNode 顯示在串流中時更名 SuperNode。以及撰寫在專案或報告中使用的註解。要存取這些內容，請執行以下操作：

- 用滑鼠右鍵按一下 SuperNode（縮小），然後選擇**更名並註解**。
- 或者，從 SuperNode 功能表中，選擇**更名並註解**。此選項在縮小模式和放大模式下均可用。



在這兩種情況下，會開啟一個對話框，其中「註解」標籤處於選定狀態。使用此處的選項可自訂顯示在串流畫布上的名稱，並提供與 SuperNode 作業相關的文件。

使用帶有 SuperNode 的備註

從帶備註的節點或塊建立 SuperNode 時，如果您想在 SuperNode 中出現備註，那麼必須在建立 SuperNode 的選擇中包含備註。如果您在選擇中省略了備註，那麼備註將在建立 SuperNode 時在串流上保留。

當您展開已併入備註的 SuperNode 時，備註恢復到建立 SuperNode 之前的位置。

當您展開已併入備註物件的 SuperNode。但備註不包含在 SuperNode 中時，物件恢復到之前的位置，但不會再次添加備註。

## SuperNode 參數

在 IBM SPSS Modeler 中，您能夠設定使用者定義變數，如 Minvalue，當在 Script 編寫或 CLEM 表示式中使用這些變數時可以指定其值。這些變數被稱為參數。您可以為串流、階段作業和 SuperNode 設定參數。當在該 SuperNode 或任何巢套節點中建立 CLEM 表示式時，SuperNode 的所有參數集均可用。巢套 SuperNode 的參數集不適用於它們的父級 SuperNode。

為 SuperNode 建立和設定參數分為兩個步驟：

1. 為 SuperNode 定義參數。
2. 然後，為 SuperNode 的每個參數指定值。

然後，可在任何封裝節點的 CLEM 表示式中使用這些參數。

## 定義 SuperNode 參數

在縮小模式和放大模式下均可定義 SuperNode 的參數。所定義的參數適用於所有封裝節點。要定義 SuperNode 的參數，首先需要存取 SuperNode 對話框的「參數」標籤。使用下列方法之一可開啟該對話框：

- 按兩下串流中的 SuperNode。
- 從 SuperNode 功能表中，選擇**設定參數**。
- 或者，當進行放大以向 SuperNode 推進時，請從快速功能表中選擇**設定參數**。

一旦開啟了對話框，「參數」標籤便會顯示，其中包含以前定義的所有參數。

定義新的參數

按一下**定義參數**按鈕以開啟對話框。

**名稱。** 參數名稱在這裡列出。可以通過在本欄位中輸入名稱來建立新的參數。例如，要為下限溫度建立參數，可以鍵入minvalue。請勿包含表示 CLEM 表示式中的參數的 \$P- 字首。該名稱也用於在 CLEM 表示式建置器中顯示。

**長名稱。** 列出每個所建立參數的描述性統計量名稱。

**儲存體。** 在清單中選取儲存類型。儲存類型表明資料值在參數中如何儲存。例如，當所使用的值包含希望保留的先導 0 時（例如 008），應選取**字串**作為儲存類型。否則，先導 0 將從值中剝離。有效的儲存類型為字串、整數、實數、時間、日期及時間戳記。對於日期參數，注意其值必須用下一段落所示的 ISO 標準符號指定。

**值。** 列出每個參數的目的值。根據需要調整參數。注意對於日期參數，其值必須用 ISO 標準符號（即 YYYY-MM-DD）指定。系統不接受以其他格式指定的日期。

**類型 (選用)**。如果計劃將該串流部署到外部應用程式，則從清單中選取測量層次。否則，建議保留 類型欄的值。如果您想為參數指定值限制（如數值型範圍的上限或下限），請從清單中選取**指定**。

注意只能通過使用者介面為參數設定長名稱、儲存類型和類型選項。不能用 Script 設定這些選項。

按一下位於右側的箭頭可使已選取的參數在可用參數清單中上下移動。使用刪除按鈕（標示為 X）可刪除已選取的參數。

## 設定 SuperNode 參數的值

一旦為 SuperNode 定義了參數，便可以使用 CLEM 表示式或 Script 中的參數指定值。

指定 SuperNode 的參數

1. 按兩下 SuperNode 圖示以開啟 SuperNode 對話框。
2. 或者，從 SuperNode 功能表中，選擇**設定參數**。
3. 按一下**參數標籤**。注意：此對話框中的欄位是通過按一下此標籤上的**定義參數**按鈕定義的欄位。
4. 在文字框中為您建立的每個參數輸入值。例如，可以將 *minvalue* 值設定為相關的特定臨界值。然後，可以在許多作業中使用此參數，如果選取高於或低於此臨界值的記錄以進行進一步的研究。

## 使用 SuperNode 參數存取節點內容

SuperNode 參數也可用於定義封裝節點的節點內容（也稱為**位置參數**）。例如，假設您要指定某 SuperNode 使用可用的隨機抽樣資料對封裝在其中的「類神經網路」節點進行適當時間長度的訓練。使用參數，可指定時間長度和抽樣百分比的值。

假設此範例 SuperNode 包含名為抽樣的「樣本」節點和名為訓練的「類神經網路」節點。您可以使用節點對話框將樣本節點的**抽樣**設定指定為**隨機 %**，將類神經網路節點的**停止於條件**設定指定為**時間**。一旦指定了這些選項，便可使用參數存取節點內容並為 SuperNode 指定特定值。在 SuperNode 對話框中，按一下**定義參數**並建立如下表格中所示的參數。

表 52. 要建立的參數

參數	數值	長名稱
Train.time	5	要訓練的時間（分鐘）
Sample.random	10	隨機抽樣百分比

注意：參數名稱（如 *Sample.random*）使用正確的語法參照節點內容，其中 *Sample* 代表節點的名稱，*random* 是節點內容。

定義了這些參數後，可以輕鬆地修改「樣本」節點和「類神經網路」節點內容的值，而不必重新開啟每個對話框。相反，只需從 SuperNode 功能表中選取**設定參數**便可以存取 SuperNode 對話框的「參數」標籤，在該標籤上，可為**隨機 %**和**時間**指定新值。這有助於在模型建立的多次疊代期間研究資料。

## SuperNode 和快取

從 SuperNode 中，可以快取除終端節點以外的所有節點。通過用滑鼠右鍵按一下節點並從「快取」快速功能表中選擇其中一個選項，以控制快取。此功能表選項對 SuperNode 外部以及封裝在 SuperNode 內的節點均適用。

以下是一些用於 SuperNode 快取的指導準則：

- 如果任何封裝在 SuperNode 中的節點啟用了快取，則 SuperNode 也會啟用快取。
- 取消 SuperNode 快取將取消 所有 被封裝節點的快取。

- 啟用 SuperNode 快取會實際啟用最後一個可快取子節點的快取。換而言之，如果最後一個子節點是「選取」節點，那麼將對該「選取」節點啟用快取。如果最後一個子節點是終端節點（不容許快取），則將啟用比鄰的支援快取的上游節點。
- 一旦為 SuperNode 的子節點設定了快取，則該快取節點中的所有上游活動（如新增或編輯節點）都將清除快取。

## SuperNode 和 Script 編寫

您可以使用 SPSS Modeler Script 撰寫語言來撰寫操作並執行終端 SuperNode 內容的簡單程式。例如，您可能想要指定複合串流的執行順序。例如，如果 SuperNode 包含需要在「統計圖」節點之前執行的「設定廣域」節點，那麼可以建立首先執行「設定廣域」節點的 Script。然後由此節點計算的值（例如，平均值或標準差）可以在執行「圖形」節點時使用。

SuperNode 對話框的 Script 標籤只適用於終端 SuperNode。

要為終端 SuperNode 開啟 Scripting 對話框，請執行以下操作：

- 用滑鼠右鍵按一下 SuperNode 畫布，然後選擇 **SuperNode Script**。
- 或者，在放大模式和縮小模式下，均可以從 SuperNode 功能表中選擇 **SuperNode Script**。

註：僅當在對話框中選定執行此 **Script**時，才會對串流和 SuperNode 執行 SuperNode Script。

《Scripting 和自動化手冊》中討論了特定 Scripting 選項及其在 SPSS Modeler 內的使用，該文檔在產品下載過程中以 PDF 檔案形式提供。

---

## 儲存和載入 SuperNode

SuperNode 的其中一個優點為：可將它們儲存下來並在其他串流中重複使用。儲存和載入 SuperNode 時，請注意，它們使用 .slb 副檔名。

### 儲存 SuperNode

1. 放大 SuperNode。
2. 從 SuperNode 功能表中，選擇儲存 **SuperNode**。
3. 在對話框中指定檔名和目錄。
4. 選取是否要將已儲存的 SuperNode 新增到目前專案中。
5. 按一下「儲存」。

### 載入 SuperNode

1. 從 IBM SPSS Modeler 視窗中的「插入」功能表中，選擇 **SuperNode**。
2. 從現行目錄中選取 SuperNode 檔案 (.slb) 或瀏覽到其他 SuperNode 檔案。
3. 按一下載入。

註：所匯入的 SuperNode 具有其所有參數的預設值。要變更參數，請按兩下串流畫布上的 SuperNode。



---

## 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。不過，您可能需要擁有該語言的產品副本或產品版本，才能對它進行存取。

在其他國家，IBM 不見得有提供本文件所提及之各項產品、服務或功能。請洽詢當地的 IBM 業務代表，以取得當地目前提供的產品和服務之相關資訊。本文件在提及 IBM 的產品、程式或服務時，不表示或暗示只能使用 IBM 的產品、程式或服務。只要未侵犯 IBM 之智慧財產權，任何功能相當之產品、程式或服務皆可取代 IBM 之產品、程式或服務。不過，任何非 IBM 之產品、程式或服務，使用者必須自行負責作業之評估和驗證責任。

本文件所說明之主題內容，IBM 可能擁有其專利或專利申請案。提供本文件不代表提供這些專利的授權。您可以書面提出授權查詢，來函請寄到：

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

如果是有關雙位元組 (DBCS) 資訊的授權查詢，請洽詢所在國的 IBM 智慧財產部門，或書面提出授權查詢，來函請寄到：

*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。有些地區在特定交易上，不允許排除明示或暗示的保證，因此，這項聲明不一定適合您。

本資訊中可能會有技術上或排版印刷上的訛誤。因此，IBM 會定期修訂；並將修訂後的內容納入新版中。IBM 隨時會改進及/或變更本出版品所提及的產品及/或程式，不另行通知。

本書對於非 IBM 網站的援引只是為了方便而提供，並不對這些網站作任何認可。這些網站上的內容並非本 IBM 產品內容的一部分，用戶使用該網站時應自行承擔風險。

IBM 得以各種 IBM 認為適當的方式使用或散布貴客戶提供的任何資訊，而無需對貴客戶負責。

如果本程式之獲授權人為了 (i) 在個別建立的程式和其他程式（包括本程式）之間交換資訊，以及 (ii) 相互使用所交換的資訊，因而需要相關的資訊，請洽詢：

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

上述資料之取得有其特殊要件，在某些情況下必須付費方得使用。

IBM 基於 IBM 客戶合約、IBM 國際程式授權合約或雙方之任何同等合約的條款，提供本文件所提及的授權程式與其所有適用的授權資料。

呈現的引用執行資料和客戶範例僅供說明之用。實際效能結果可能會依據特定配置和作業條件而有所不同。

本文件所提及之非 IBM 產品資訊，係取自產品供應商，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 產品的性能問題應直接洽詢該產品供應商。

有關 IBM 未來動向的陳述，僅代表 IBM 的目標而已，並可能於未事先聲明的情況下有所變動或撤回。

此資訊包含日常企業運作所使用的資料和報告的範例。為了盡可能完整地說明，範例中包括了個人、公司行號、品牌以及產品等的名稱。所有這些名稱全為虛構，如有任何類似實際人員或實際企業之處，純屬巧合。

---

## 商標

IBM、IBM 標誌及 [ibm.com](http://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](http://www.ibm.com/legal/copytrade.shtml)。

Adobe、Adobe 標誌、PostScript 及 PostScript 標誌是 Adobe Systems Incorporated 在美國及（或）其他國家或地區的註冊商標或商標。

Intel、Intel 標誌、Intel Inside、Intel Inside 標誌、Intel Centrino、Intel Centrino 標誌、Celeron、Intel Xeon、Intel SpeedStep、Itanium 及 Pentium 是 Intel Corporation 或其子公司在美國及其他國家或地區的商標或註冊商標。

Linux 是 Linus Torvalds 在美國及（或）其他國家或地區的註冊商標。

Microsoft、Windows、Windows NT 及 Windows 標誌是 Microsoft Corporation 在美國及/或其他國家或地區的商標。

UNIX 是 The Open Group 在美國及其他國家或地區的註冊商標。

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。

---

## 產品說明文件的條款

這些出版品的使用授權受到下列條款的限制。

### 適用性

除了 IBM 網站的所有使用條款以外，還適用這些條款。

### 個人使用

貴客戶可以為了非商務性的私人用途而複製這些出版品，但必須保留全部的所有權聲明。如果沒有 IBM 的明文同意，貴客戶不能散布、顯示或衍生這些出版品或其中的任何部分。

### 商業用途

貴客戶可以在企業內複製、散佈和展示這些出版品，但必須保留所有專利注意事項。如果沒有 IBM 的明文同意，貴客戶不能在您的企業外衍生這些出版品，或複製、散佈或顯示這些出版品或其中的任何部分。

## 權利

除了在此明確授與的許可權之外，並未授與（明確或隱含）出版品或其包含的任何資訊、資料、軟體或其他智慧財產的任何其他許可權、軟體授權或權利。

IBM 保留在判定出版品的使用將損害其利益或判定未適當遵守上述指示時，撤銷此處所授予之許可權的權利。

除非完全符合所有適當的法律和規章，其中包括所有美國輸出法律和規章，否則，貴客戶不能下載、輸出或再輸出本項資訊。

IBM 對於該等出版品之內容不作任何保證。這些出版品只依「現狀」提供，不含任何明示或暗示的保證，其中包括（但不限於）可售性、未涉侵權或符合特定效用的暗示保證。





---

## 詞彙

---

### C

共變異數 (Covariance)：兩個變數之間關聯的非標準化量數，它等於交叉product離差除以  $N-1$ 。

---

### K

峰度：離群值範圍的一種測量。對常態分配而言，峰度統計量數值為零。峰度為正數指出資料比常態分佈展示更多的極端離群值。峰度為負數指出資料比常態分佈展示較少的極端離群值。

---

### M

最大值：數值變數的最大值。

平均數(M)：集中趨勢的測量。算術平均數，總和除以觀察值數。

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

最小值：數值變數的最小值。

眾數(O)：最常出現的值。如果幾個值的最大出現次數相同，則每一個都是眾數。

---

### R

範圍：值變數最大與最小值之間的差異，也就是最大值減去最小值。

---

### S

偏斜度：分配不對稱性的量數。常態分佈是一種對稱性分佈，其偏斜度值為 0。具有顯著性正偏斜度的分佈右側尾部較長。有顯著負偏斜度值的分配會有一個長的偏左尾部。偏斜度值有如指標，若大於它的兩倍標準誤，則表示背離對稱。

標準差：在平均數四周離散的量數，等於變異數的平方根。標準差所用的測量單位與原始觀察值相同。

標準差：測量平均數四周的離散情形。在常態分配中，68% 的觀察值會落在平均數的一個標準差內，95% 的觀察值會落在兩個標準差內。例如，如果平均年齡為 45 歲，標準差是 10 的話，在常態分配中 95% 的觀察值會介於 25 到 65 歲之間。

標準誤：不同樣本間測試統計量值的變化大小量數。這是一項統計量取樣分配的標準差。例如，平均數的標準誤就是樣本平均數的標準差。

峰度的標準誤 (Standard Error of Kurtosis)：峰度對其標準誤的比例可用於測試常態性（也就是說，如果比例小於 -2 或大於 +2，則您可以拒絕常態性）。峰度若為大的正值，表示該分配的尾端比常態分配的尾端更長；峰度若為負值，表示尾端較短（變成類似箱型均勻分配的尾端）。

平均數的標準誤 (Standard Error of Mean)：測量從同一個分配取出來的不同樣本間平均數的變化大小。它可以用來大略地將觀察平均數與假設值相比較（也就是如果標準誤與差異的比值小於 -2 大於 +2 的話，您就可以下結論說兩個值不同）。

偏斜度的標準誤 (*Standard Error of Skewness*)：偏斜度對其標準誤的比例可用於測試常態性（也就是說，如果比例小於 -2 或大於 +2，則您可以拒絕常態性）。偏斜度若為大的正值，表示有長的偏右尾端；極端負值則表示有長的偏左尾端。

總和(U)：遍及所有包含遺漏值之觀察值的值總和或總數。

---

## U

唯一 (*Unique*)：針對任何類型的所有其他效果，調整每個效果來同時評估所有效果。

---

## V

有效：沒有系統遺漏值，也沒有定義為使用者遺漏值的有效觀察值。

變異數：測量平均數四周的離散情形，它等於平均數的平方離差總和除以觀察值數目減一。變異數的測量單位是變數本身的平方。

## 索引

索引順序以中文字，英文字，及特殊符號之次序排列。

### 〔二劃〕

二十分位數 Bin 144  
人為介入  
    建立 203  
十六進位控制字元 10  
十分位數 Bin 144

### 〔三劃〕

三維圖形 164  
下限  
    統計量輸出 283  
    設定廣域節點 288  
上限  
    統計量輸出 283  
    設定廣域節點 288  
大小圖形套版 162  
大型資料庫 61  
    執行資料審核 274  
大量載入 307, 308  
小數位數  
    顯示格式 128  
小數點符號 23  
    純文字檔匯出節點 314  
    數字顯示格式 128  
工作表  
    從 Excel 匯入 38  
已分層樣本 63, 64, 65, 66

### 〔四劃〕

不已平衡的資料 67  
不完整記錄 74  
不受支援的控制字元 10  
中位數  
    統計量輸出 283  
中位數近似值 70  
五分位數 Bin 144  
內容  
    節點 342  
內部合併 73  
六角形 Bin 散佈圖 170  
分位顯示  
    分組節點 144  
分佈節點 205  
    外觀標籤 205

分佈節點 (繼續)  
    使用表格 206  
    使用圖形 206  
    繪圖標籤 205  
分析節點 271  
    「分析」標籤 272  
    輸出標籤 267  
分析資料視圖 56  
分析瀏覽器  
    解讀 273  
分配 208  
分組符號  
    數字顯示格式 128  
分組節點  
    均數/標準差 Bin 146  
    固定寬度 Bin 143  
    相等和 144  
    相等計數 144  
    排等級 145  
    設定選項 143  
    最佳化 146  
    概述 142  
    預覽 Bin 147  
分割區節點 150, 151  
分割區欄位 127, 150, 151  
分割資料 150, 151  
    分析節點 272  
    評估圖表 223  
分割點  
    分組節點 142  
分隔的文字資料 23  
分數等級 145  
勾選類型 126  
升冪 72  
反結合 73  
天氣公司來源 36  
尺度因子 67  
引號  
    匯入文字檔 23  
文件 3  
文字  
    分隔 23  
    資料 23, 26  
文字輸出標籤 298  
文字檔 23  
    匯出 321  
日期  
    設定格式 128  
日期識別 23, 26  
日期/時間 118

### 〔五劃〕

主要索引鍵欄位  
    資料庫匯出節點 304  
主要資料集 79  
以逗點區隔檔案  
    匯出 266, 321  
    儲存 267  
代碼變數  
    資料收集來源節點 28  
加權值  
    評估圖表 223  
加權樣本 65  
卡方  
    矩陣節點 271  
四分位數 Bin 144  
四分位數近似值 70  
外部合併 73  
市場調查資料  
    匯入 28, 31  
    資料收集來源節點 28, 31  
平方根轉換  
    時間序列模型器 89  
平行座標圖 170  
平行處理  
    合併 78  
    排序 72  
    聚集節點 68  
平均值  
    分組節點 146  
    統計量輸出 283  
    設定廣域節點 288  
平均數  
    比較 284, 285  
平均數函數  
    時間序列聚集 158  
平均數的標準誤差  
    統計量輸出 283  
平均數節點 284  
    成對欄位 284  
    重要性 285  
    獨立群組 284  
    輸出標籤 267  
    輸出瀏覽器 285  
平衡因子 67  
平衡節點  
    從圖形產生 246  
    設定選項 67  
    概述 67  
未定義值 74

正規化值

- 圖形節點 200, 203
- 用於聚集的中位數值 68, 70
- 用於聚集的四分位數值 68, 70
- 用量類型 7, 118
- 皮爾森 (Pearson) 卡方矩陣節點 271

## (六劃)

交叉列表

- 矩陣節點 269, 270
- 「全體」節點
  - 綜合分數 149
  - 輸出欄位 149
- 列印輸出 265
- 同分值
  - 分組節點 144
- 名義資料 125
- 合成記錄 82
  - 自訂設定 82
- 合併節點 73
  - 設定選項 74, 75, 76
  - 最佳化設定 78
  - 概述 72
  - 過濾欄位 77
  - 標記欄位 78
- 合併資料 79
  - 來自多個檔案 72
- 合併選項，資料庫匯出 303
- 合計值 68
- 回應圖表 220, 225
- 地理空間
  - 設定匯入選項 57
- 地理空間來源節點
  - 地圖服務 57
  - .dbf 檔案 57
  - .shp 檔案 57
- 地理空間測量層次 9, 118, 120, 125, 134
- 地理空間資料 23
  - 合併 76
  - 在變數檔案中 25
  - 衍生 135
  - 限制 120
  - 排名式條件合併 76
  - 匯入 23
  - 匯出 314
  - 變數檔案中的清單 25
- 地理空間資料中的限制 120
- 地理空間資料重新投影 159
- 地理空間資料預備
  - 重新投影節點 159
- 地理空間圖中的層 227
- 地理空間類型 125
- 地理座標系統 159

地圖

- 平滑處理 190, 191
- 合併特徵 192
- 色彩 170
- 刪除個別元素 193
- 刪除特徵 193
- 投影 193
  - 具有長條圖 170
  - 具有圓餅圖 170
  - 具有箭頭 170
  - 具有線條圖表 170
  - 具有點 170
- 削薄處理 190, 191
- 重疊 170
- 特徵標記 191
- 移動特徵 193
- 發佈 194
  - 轉換 ESRI 形狀檔案 189
- 地圖中的地理空間資料 227
- 地圖形狀檔案
  - 使用圖表板樣本選擇器 189
  - 概念 189
  - 編輯預先安裝的 SMZ 地圖 189
  - 類型 189
- 地圖服務
  - 地理空間來源節點 57
- 地圖的層選項 228
- 地圖視覺化
  - 建立 175
  - 範例 184
- 地圖視覺化節點 227
  - 繪圖標籤 227
  - 變更層選項 228
- 地圖資料重新投影 159
- 地圖檔案
  - 在圖表板樣本選擇器中選取 169
  - 位置 187
  - 刪除 188
  - 更名 188
  - 匯入 188
  - 匯出 188
- 地圖轉換公用程式 189, 190
- 多重二分法集 131
- 多重回應集
  - 在視覺化中 167
- 多重二分法集 131
- 多類別集合 131
- 刪除 131
- 定義 131
- 資料收集來源節點 28, 31
- IBM SPSS Statistics 來源節點 27, 328
- 多重衍生 133
- 多個多邊形 120
- 多個輸入 72

多個欄位

- 選擇 134
- 多媒體串流 TCM 節點 96, 97, 98, 99, 100
- 多媒體串流時間序列節點
  - 概述 83
- 多媒體串流時間序列模型
  - 一般建置選項 87
  - 估計期間 86
  - 建置選項 87
  - 指數平滑化 87
  - 時間間隔選項 85
  - 資料規格選項 84
  - 聚集及分配選項 85
  - 模型選項 90
  - 遺漏值選項 86
  - 欄位選項 84
  - 觀察選項 84
  - ARIMA 87
- 多圖節點 200
  - 外觀標籤 202
  - 使用圖形 202
  - 繪圖標籤 200
- 多點 120
- 多邊形 120
- 多類別集合 131
- 如果有任何 true 函數則為 True
  - 時間序列聚集 158
- 成本
  - 評估圖表 223
- 收集類型 125
- 次數分配表
  - 分組節點 144
- 百分位數 Bin 144
- 自由度
  - 平均數節點 285, 286
  - 矩陣節點 271
- 自由欄位文字資料 23
- 自動日期識別 23, 26
- 自動式資料準備
  - 功能選擇 109
  - 目標準備 108
  - 常態化連續目標 108
  - 準備目標 108
  - 準備輸入 108
  - 構建 109
  - 輸入準備 108
- 自動重新編碼 139
- 自動資料準備
  - 目標 105
  - 名稱欄位 110
  - 衍生節點產生 116
  - 重設檢視 110
  - 動作詳細資料 114
  - 動作摘要 112
  - 常態化連續目標 116

- 自動資料準備 (繼續)
  - 排除 UNUSED 欄位 107
  - 排除未用的欄位 107
  - 排除欄位 108
  - 準備日期與時間 107
  - 預測能力 113
  - 模型視圖 110
  - 檢視之間的鏈結 110
  - 欄位 106
  - 欄位分析 111
  - 欄位表格 113
  - 欄位處理摘要 111
  - 欄位設定 107
  - 欄位詳細資料 113
- 「自動資料準備」節點 105
- 自動鍵入 118, 122
- 自然順序
  - 更改 156
- 自然對數轉換
  - 時間序列模型器 89
- 色板
  - 移動 249
  - 隱藏 249
  - 顯示 249
- 色彩圖 170
  - 範例 184

## 〔七劃〕

- 串流參數 21, 22
- 利潤圖表 220, 225
- 刪除
  - 地圖檔案 188
  - 視覺化樣式表 188
  - 視覺化範本 188
  - 輸出物件 264
- 均數 / 標準差
  - 用於 Bin 欄位 146
- 局部結合 73, 75
- 序數資料 125
- 形狀圖形套版 162
- 形狀檔案 189
- 快取
  - SuperNode 342
- 快取檔節點 27, 328
- 抖動 198, 235
- 投影座標系統 159
- 改寫資料庫表格 302
- 更名
  - 地圖檔案 188
  - 視覺化樣式表 188
  - 視覺化範本 188
  - 匯出的欄位 316, 335
- 系統遺漏值
  - 在 Matrix 表格中 269

- 系統變數
  - 資料收集來源節點 28
- 角色
  - 為欄位指定 127

## 〔八劃〕

- 事件
  - 建立 203
- 使用者輸入節點
  - 設定選項 41
  - 概述 40
- 使用者遺漏值
  - 在 Matrix 表格中 269
- 來源節點
  - 天氣公司來源 36
  - 地理空間來源節點 57
  - 使用者輸入節點 40, 41
  - 固定檔案節點 26
  - 統計量檔案節點 27, 328
  - 概述 5
  - 資料庫來源節點 15
  - 實例化類型 59
  - 模擬產生節點 45, 46
  - 變數檔案節點 23
  - Analytic Server 來源 11
  - Excel 來源節點 38
  - IBM Cognos TM1 來源節點 35
  - IBM Cognos 來源節點 32, 34, 35
  - SAS 來源節點 37
  - TWC 來源 36
  - XML 來源節點 39
- 來源檔案變數
  - 資料收集來源節點 28
- 取代欄位值 137
- 取樣訊框 63
- 取樣資料 66
- 命中數
  - 評估圖表選項 224
- 固定檔案節點
  - 自動日期識別 26
  - 設定選項 26
  - 概述 26
- 固定欄位文字資料 26
- 定界字元 23, 307
- 延伸
  - 衍生的欄位 133
- 延伸匯入節點 55
  - 主控台輸出標籤 55
- 延伸匯出節點 322
  - 主控台輸出標籤 323
- 延伸輸出節點 296
  - 主控台輸出標籤 297
  - 語法標籤 296
  - 輸出標籤 297
- 延伸輸出瀏覽器 298

- 延伸轉換節點 93
  - 主控台輸出標籤 93
- 波動 248, 256
- 泡泡圖 170
- 直方圖 170
  - 範例 178
  - 3-D 170
- 直方圖節點 208
  - 外觀標籤 208
  - 使用圖形 209
  - 繪圖標籤 208
- 直欄順序
  - 表格瀏覽器 267, 269
- 空白
  - 在 Matrix 表格中 269
- 空白列
  - Excel 檔案 38
- 空白處理 123
  - 分組節點 143
  - 填充值 137
- 空值 123
  - 在 Matrix 表格中 269
- 空間時間限制中的密度定義 96
- 股價圖 170
- 表示式建置器 61
- 表面圖形 170
- 表格
  - 結合 73
  - 儲存為文字 267
  - 儲存輸出 267
- 表格節點 267
  - 設定標籤 267
  - 輸出設定 267
  - 輸出標籤 267
- 表格輸出
  - 重排直欄 267
  - 選取資料格 267
- 表格瀏覽器
  - 重排直欄 267, 269
  - 「產生」功能表 269
  - 搜尋 269
  - 選取資料格 267, 269
- 長條圖 170
  - 在地圖上 170
  - 個數 170
  - 範例 176, 177
  - 3-D 170
- 附加節點
  - 設定選項 79
  - 概述 79
  - 標記欄位 78
  - 欄位比對 79
- 信賴區間
  - 平均數節點 285, 286
- 品質報告
  - 資料審核瀏覽器 277

- 品質瀏覽器
  - 產生過濾器節點 279
  - 產生選取節點 279

## (九劃)

- 封裝節點 338
- 建立
  - 新建欄位 131, 133
- 建模角色
  - 為欄位指定 127
- 持續時間計算
  - 自動資料準備 107
- 指派資料類型 103
- 查詢
  - 資料庫來源節點 15, 16
- 查詢級區
  - Teradata 19
- 查詢編輯器
  - 資料庫來源節點 21, 22
- 流程地圖 170
- 相等計數
  - 分組節點 144
- 相關 282
  - 平均數節點 286
  - 統計量輸出 283
  - 描述性統計量標籤 282
  - 絕對值 282
  - 顯著性 282
  - probability 282
- 科學顯示格式 128
- 衍生公式的地理空間值 134
- 衍生公式的收集值 134
- 衍生公式的值 134
- 衍生節點
  - 公式值 134
  - 名義 136
  - 地理空間值 134
  - 多重衍生 133
  - 收集值 134
  - 狀態 136
  - 衍生地理空間欄位 135
  - 衍生清單欄位 135
  - 計數 137
  - 從 Web 圖形鏈結中產生 216
  - 從自動資料預備產生 116
  - 從圖形產生 246
  - 條件 137
  - 設定選項 133
  - 通過 Bin 產生 142
  - 通過分組節點產生 147
  - 概述 131
  - 旗標 136
  - 數值重新編碼 137
  - 轉換欄位儲存類型 137
  - formula 134

- 計算持續時間
  - 自動資料準備 107
- 計數
  - 分組節點 144
  - 統計量輸出 283
  - 重命名輸出物件 264
  - 重建資料 152
  - 重要性
    - 比較平均數 285
    - 平均數節點 285, 286
  - 重組節點 152, 153
    - 與「聚合」節點 152
  - 重新分類節點 139, 140
    - 從分佈產生 206
    - 概述 139, 142
  - 重新投影節點 159
  - 重新編碼 139, 142
  - 重疊圖 170
  - 面量圖
    - 範例 184
  - 面量圖地圖 170
- (十劃)
- 值
  - 指定 123
  - 欄位和值標籤 123
  - 讀取 122
- 修改資料值 131
- 差異節點
  - 合成設定 82
  - 排序記錄 80
  - 最佳化設定 81
  - 概述 80
- 座標系統
  - 轉換 255
- 座標圖 170
- 效能
  - 分組節點 147
  - 合併 78
  - 取樣資料 63
  - 「衍生」節點 147
  - 排序 72
  - 聚集節點 68
- 效能評估統計資料 272
- 時間序列 155
- 時間序列資料
  - 聚集 157
- 時間序列模型
  - 轉換 89
  - 轉換函數順序 89
  - ARIMA 89
- 時間原因模型 98
  - 多媒體串流 TCM 節點 96
- 時間格式 128
- 時間間隔節點 157, 158

- 時間間隔節點 (繼續)
  - 概述 157
- 時間圖節點 203
  - 外觀標籤 204
  - 使用圖形 204
  - 繪圖標籤 203
- 時間戳記 118
- 格式
  - 資料 7
- 格式化檔案 37
- 特徵
  - 在地圖中 189
- 矩陣表中的
  - 在 Matrix 表格中 269
- 矩陣節點 269
  - 外觀標籤 270
  - 交叉列表 270
  - 列百分比 270
  - 直欄百分比 270
  - 為列與欄排序 270
  - 強調顯示 270
  - 設定標籤 269
  - 輸出標籤 267
  - 輸出瀏覽器 271
- 矩陣輸出
  - 儲存為文字 267
- 矩陣瀏覽器
  - 「產生」功能表 271
- 索引資料庫表格 306
- 索引鍵方法 72
- 索引鍵欄位 68, 152
- 純文字檔案 23
- 純文字檔匯出節點 314
  - 匯出標籤 314
- 記錄
  - 合併 72
  - 計數 68
  - 標記 127
  - 轉置 154
  - length 26
- 記錄作業節點 61
- 記錄的平均數 68
- 訓練樣本
  - 分割資料 150, 151
  - 平衡 67
- 迴避 248, 256
- 高-低-關閉圖 170

## (十一劃)

- 偏倚資料 67
- 匿名化欄位名稱 130
- 匿名節點
  - 建立匿名化的數值 142
  - 設定選項 141
  - 概述 141

- 區域加權最小二乘迴歸方法 (LOESS)
  - 繪圖節點 197
  - E-Plot 節點 235
- 區域圖 170
  - 3-D 170
- 區間
  - 時間序列資料 157
- 參照
  - 用於資料庫匯出 302
- 參數
  - 在 IBM Cognos 中 35
  - 節點內容 342
  - SuperNode 341, 342
  - SuperNode 設定 341
- 商業規則
  - 評估圖表選項 224
- 唯一性記錄 80
- 堆疊 248, 256
- 堆疊長條圖
  - 範例 177
- 基準線
  - 評估圖表選項 223
- 執行
  - 指定順序 343
- 執行順序
  - 指定 343
- 密度
  - 3-D 170
- 將值分組 206
- 將集轉換為旗標 152
- 常態化連續目標 108, 116
- 帶狀圖 170
- 強制轉型值 126
- 從圖形產生節點 246
  - 「平衡」節點 246
  - 「再分類」節點 246
  - 「衍生」節點 246
  - 「過濾器」節點 246
  - 「選取」節點 246
- 控制字元 10
- 探索資料
  - 資料審核節點 274
- 探索圖形 239
  - 區域 243
  - 圖表級區狀區域 240
  - 標示元素 245
  - 魔棒 245
- 排名式條件
  - 為某個合併指定 76
- 排序
  - 差異節點 80
  - 記錄 72
  - 預先排序的欄位 72, 81
  - 欄位 156
- 排序節點
  - 最佳化設定 72

- 排序節點 (繼續)
  - 概述 72
- 排序資料 72, 156
- 排除未用的欄位
  - 自動資料準備 107
- 排等級觀察值 145
- 捨棄
  - 欄位 129
- 旋轉三維圖形 164
- 條件
  - 已排名的 76
  - 指定數列 137
  - 為某個合併指定 75
- 清除值 58
- 清單 9, 118
  - 地理空間測量層次 120
  - 地理空間資料類型 125
  - 衍生 135
  - 深度 9
  - 最大長度 123
  - 集合資料類型 125
- 清單的深度 9
- 清單儲存格式 9
- 清單儲存類型 25
- 產生旗標 152, 153
- 盒形圖 170
  - 範例 180
- 眾數函數
  - 時間序列聚集 158
- 第一個四分位數
  - 時間序列聚集 158
- 第三個四分位數
  - 時間序列聚集 158
- 符合矩陣
  - 分析節點 272
- 統計量
  - 在視覺化中編輯 256
  - 資料審核節點 274
- 統計量匯出節點 315, 334
  - 「匯出」標籤 315, 335
- 統計量節點 282
  - 相關 282
  - 相關係數標籤 282
  - 統計量 282
  - 設定標籤 282
  - 輸出標籤 267
- 統計量輸出節點 332
  - 語法標籤 332
- 統計量檔案節點 27, 328
- 統計量轉換節點 329
  - 容許的語法 329
  - 設定選項 329
  - 語法標籤 329
- 統計資料
  - 矩陣節點 269

- 統計資料瀏覽器
  - 「產生」功能表 283
  - 產生過濾器節點 284
  - 解讀 283
- 處理遺漏值 103
- 設定為旗標節點 152
- 設定廣域節點 288
  - 設定標籤 288
- 設定隨機種子
  - 取樣記錄 151
- 設定類型 118
- 貨幣顯示格式 128
- 連接記錄 79
- 連續資料 121, 124
- 連續資料取樣 64
- 連續鍵 70
- 逐列式鏈結 307
- 逐直欄式鏈結 307

## (十二劃)

- 最大值函數
  - 時間序列聚集 158
- 最小值函數
  - 時間序列聚集 158
- 最佳化分級 146
- 最佳行
  - 評估圖表選項 223
- 單因素 ANOVA
  - 平均數節點 284
- 報告
  - 儲存輸出 267
- 報告節點 286
  - 「範本」標籤 287
  - 輸出標籤 267
- 報告瀏覽器 288
- 循環時間元素
  - 自動資料準備 107
- 提交大小 307
- 散佈圖 170, 194, 200, 235
  - 六角形 binned 170
  - 3-D 170
  - binned 170
- 散佈圖矩陣
  - 範例 183, 185
- 散佈圖矩陣 (SPLOM) 170
- 期望值
  - 矩陣節點 270
- 殘差
  - 矩陣節點 270
- 減少資料 63
- 測量層級
  - 已定義 118
  - 地理空間 9, 120, 125, 134
  - 地理空間資料中的限制 120
  - 在視覺化中 167

- 測量層級 (繼續)
  - 在視覺化中變更 166
  - 集合 9, 125, 134
- 測試樣本
  - 分割資料 150, 151
- 無偏倚資料 67
- 畫面化 162
- 畫面圖形重疊 162
- 發佈至網路 265
- 發佈串流
  - IBM SPSS Modeler 解決方案發佈者 325
- 結合 72, 73, 74
  - 局部外部 75
- 結合資料集 79
- 虛擬編碼 152
- 視覺化
  - 尺度 252
  - 文字 249
  - 透明度 250
  - 窗格 254, 255
  - 軸 252
  - 填塞 251
  - 圖形和圖表 161
  - 圖註位置 257
  - 種類 254
  - 數字格式 252
  - 編輯 248
  - 編輯模式 248
  - 線條粗細 250
  - 複製 257
  - 點的形狀 251
  - 點的旋轉 251
  - 點的圖面比 251
  - 轉換座標系統 255
  - 顏色和樣式 250
  - 邊緣 251
  - transpose 254, 255
- 視覺化樣式表
  - 位置 187
  - 刪除 188
  - 更名 188
  - 套用 258
  - 匯入 188
  - 匯出 188
- 視覺化範本
  - 位置 187
  - 刪除 188
  - 更名 188
  - 匯入 188
  - 匯出 188
- 註解
  - 用於 SuperNode 340
- 註解字元
  - 在變數檔案中 23

- 評分
  - 評估圖表選項 224
- 評估節點 220
  - 分數表示式 224
  - 匹配條件 224
  - 外觀標籤 225, 231
  - 使用圖形 226
  - 商業規則 224
  - 選項標籤 224
  - 繪圖標籤 223
  - 讀取結果 225
- 評估模型 271
- 週期
  - 時間序列資料 157
  - 時間序列模型器 89
- 量子 資料
  - 匯入 28
- 開啟
  - 輸出物件 264
- 集
  - 轉換 139, 140
  - 轉換為旗標 152
- 集合測量層次 125, 134
- 集合節點 209
  - 外觀標籤 210
  - 使用圖形 211
  - 選項標籤 210
- 順序合併 72

## (十三劃)

- 傳輸檔案
  - SAS 來源節點 37
- 傾向評分
  - 平衡資料 67
- 匯入
  - 地圖檔案 188
  - 視覺化樣式表 188
  - 視覺化範本 188
  - IBM Cognos BI 中的報告 34
  - IBM Cognos TM1 中的資料 36
  - IBM Cognos 中的資料 33
  - SuperNode 343
- 匯出
  - 地圖檔案 188
  - 視覺化樣式表 188
  - 視覺化範本 188
  - 輸出 266
  - IBM Cognos TM1 中的資料 320
  - SuperNode 343
- 匯出小數位 128
- 匯出節點 301
  - Analytic Server 匯出 317
- 匯出資料
  - 文字 321
  - 地理空間 314

- 匯出資料 (繼續)
  - 至 Excel 321
  - 至資料庫 302
  - 到 IBM SPSS Statistics 315, 334
  - 純文字檔案格式 314
  - DAT 檔案 321
  - IBM Cognos TM1 匯出節點 319
  - IBM Cognos 匯出節點 34, 317, 318
  - SAS 格式 321
  - XML 格式 323
- 圓餅圖 170
  - 在地圖上 170
  - 使用個數 170
  - 範例 181
  - 3-D 170
- 填充器節點
  - 概述 137
- 意見調查資料
  - 匯入 28, 31
  - 資料收集來源節點 28
- 搜尋
  - 表格瀏覽器 269
- 新增
  - 記錄 68
- 會員 (SAS 匯入)
  - 設定 37
- 極座標 255
- 節點內容 342
- 解除鎖定 SuperNode 339
- 資料
  - 不受支援的控制字元 10
  - 匿名 141
  - 探索 274
  - 理解 61
  - 準備 61
  - 聚集 68
  - 審核 274
  - 儲存類型 123
  - 儲存體 137, 138
- 資料存取方案 56
- 資料收集來源節點 28, 31
  - 日誌檔 28
  - 多重回應集 31
  - 資料庫連線設定 31
  - 語言 30
  - 標籤類型 30
  - meta 資料檔 28
- 資料收集匯出節點 316
- 資料收集意見調查資料
  - 匯入 28
- 資料來源
  - 資料庫連線 16
- 資料品質
  - 資料審核瀏覽器 277
- 資料庫
  - 大量載入 307, 308



- 資料庫來源節點 15
  - 查詢編輯器 21, 22
  - 潛在問題 18
  - 選取表格和視圖 21
  - SQL 查詢 16
- 資料庫連線
  - 定義 16
  - 預設值 18
- 資料庫匯出節點 302
  - 合併選項 303
  - 表格名稱 302
  - 匯出標籤 302
  - 資料來源 302
  - 對映來源資料欄位到資料庫直欄 303
  - 綱目 304
  - 編製索引表格 306
- 資料視圖節點 56
  - 設定選項 56
- 資料審核節點 274
  - 設定標籤 274
  - 輸出標籤 267
- 資料審核瀏覽器
  - 生成圖形 279
  - 產生節點 279
  - 編輯功能表 275
  - 檔案功能表 275
- 資料類型 26, 103, 118
  - 實例化 122
- 路徑圖形 170
- 過濾器節點
  - 多重回應集 131
  - 設定選項 129
  - 概述 129
- 過濾欄位 77, 129
  - 適用於 IBM SPSS Statistics 316, 335
- 預設值, 資料庫連線 18

## (十四劃)

- 圖示, IBM Cognos 32
- 圖形
  - 列印 259
  - 地圖視覺化 227
  - 多重繪圖 200
  - 收集 209
  - 刪除區域 243
  - 來自圖版 166
  - 直方圖 208
  - 時間序列 203
  - 級區 240
  - 區域 243
  - 從資料審核產生 279
  - 探索 239
  - 旋轉三維影像 164
  - 條形圖 205
  - 產生節點 246

- 圖形 (繼續)
  - 註腳 258
  - 「註解」標籤 164
  - 評估圖表 220
  - 軸註解 258
  - 匯出 259
  - 預設色系 258
  - 圖形 194
  - 圖形元素的大小 251
  - 網絡 212
  - 樣式表 258
  - 標題 258
  - 複製 259
  - 「輸出」標籤 163
  - 儲存 259
  - 儲存佈置變更 258
  - 儲存編輯的佈置 258
  - 儲存輸出 267
  - 3-D 164
  - E-Plot 235
  - 圖形中的級區 240
  - 圖形中的動畫 162
  - 圖形中的區域 243
  - 圖形中的魔棒 245
  - 圖形元素
    - 衝突修飾元 256
    - 轉換 256
    - 變更 256
  - 圖形板節點 166
    - 外觀標籤 186
  - 圖形的套版 162
  - 圖形的透明度 162
  - 圖形節點 161
    - 分佈 205
    - 地圖視覺化 227
    - 多點 200
    - 直方圖 208
    - 套版 162
    - 時間圖 203
    - 動畫 162
    - 統計圖 194
    - 畫面 162
    - 評估 220
    - 集合 209
    - 圖表板 166
    - 網路 212
    - E-Plot 235
  - 圖形輸出標籤 298
  - 圖形類型
    - 圖板 170
  - 圖板
    - 圖形類型 170
  - 圖表
    - 儲存輸出 267
  - 圖表輸出 294
  - 圖表選項 295

- 圖註
  - 位置 257
- 實例化 118, 122
  - 來源節點 59
- 實數範圍 124
- 對映
  - 用來匯出至 IBM Cognos TM1 的資料 320
- 對映欄位 303
- 對數轉換
  - 時間序列模型器 89
- 對模型中使用的資料進行掩飾 141
- 截斷欄位名稱 129, 130
- 摘要統計量
  - 資料審核節點 274
- 摘要資料 68
- 旗標類型 118, 125
- 滯後資料 155
- 監督式 binning 146
- 種子值
  - 取樣與記錄 151
- 管理程式
  - 「輸出」標籤 264
- 綜合資料
  - 使用者輸入節點 40
- 網路圖的方向佈置 214
- 網路圖的佈置 214
- 綱目
  - 資料庫匯出節點 304
- 聚合記錄 152
- 聚集的平均數 68
- 聚集的金鑰值 68
- 聚集的計數值 68
- 聚集的最大值 68
- 聚集的最小值 68
- 聚集的標準差 68
- 聚集的變異數值 68
- 聚集時間序列資料 158
- 聚集節點
  - 中位數近似值 70
  - 四分位數近似值 70
  - 平行處理 68
  - 效能 68
  - 設定選項 68
  - 最佳化設定 70
  - 概述 68
- 語言
  - 資料收集來源節點 30
- 語法標籤
  - 統計量輸出節點 332
- 說明應用程式 298
- 遞減 72

## (十五劃)

- 增益圖表 220, 225
- 增譯圖表 220, 225
- 審核
  - 起始資料審核 274
  - 資料審核節點 274
- 廣域值 288
- 數字顯示格式 128
- 數值標籤
  - 統計量檔案節點 27, 328
- 樣本節點
  - 已分層樣本 63, 64, 65, 66
  - 加權樣本 65
  - 取樣訊框 63
  - 隨機樣本 63, 64
  - 叢集樣本 63, 64, 65
  - nonrandom 樣本 63, 64
  - Strata 的樣本大小 66
  - systematic 樣本 63, 64
- 樣式表
  - 刪除 188
  - 更名 188
  - 匯入 188
  - 匯出 188
- 標示元素 243, 245
- 標記 125
  - 指定 123, 124, 125
  - 匯入 27, 37, 328
  - 匯出 315, 321, 335
- 標準差
  - 分組節點 146
  - 統計量輸出 283
  - 設定廣域節點 288
- 標籤 72, 78
- 標籤類型
  - 資料收集來源節點 30
- 標籤欄位
  - 為輸出中的記錄加上標籤 127
- 模式
  - 統計量輸出 283
- 模型
  - 匿名化資料 141
- 模型視圖
  - 於自動資料準備中 110
- 模型評估 220
- 模型選項
  - Statistics 模型節點 331
  - 「模擬產生」節點
    - 設定選項 46
- 模擬產生節點
  - 概述 45
  - 「模擬評估」節點
    - 設定標籤 291
- 模擬評估節點 291, 293, 294, 295
  - 輸出設定 291

- 模擬資料
  - 模擬產生節點 45
- 模擬適合節點 288
  - 分配配適 289
  - 設定標籤 290
  - 輸出設定 290
- 潛在問題
  - 資料庫來源節點 18
- 熱圖 170
  - 範例 182
- 範本
  - 刪除 188
  - 更名 188
  - 報告節點 287
  - 匯入 188
  - 匯出 188
- 範例
  - 概述 4
  - 應用程式手冊 3
- 範圍 118
  - 統計量輸出 283
  - 遺漏值 123
- 編輯視覺化 248
  - 尺度 252
  - 文字 249
  - 加入 3-D 效果 255
  - 合併類別 254
  - 收合種類 254
  - 自動設定 249
  - 法則 249
  - 排序類別 254
  - 排除類別 254
  - 透明度 250
  - 窗格 255
  - 軸 252
  - 填塞 251
  - 圖註位置 257
  - 種類 254
  - 數字格式 252
  - 線條粗細 250
  - 選擇 249
  - 點的形狀 251
  - 點的旋轉 251
  - 點的圖面比 251
  - 轉換座標系統 255
  - 顏色和樣式 250
  - 邊緣 251
  - transpose 255
- 編輯圖形
  - 圖形元素的大小 251
- 線串 120
- 線形圖 170
  - 在地圖上 170
- 線條圖 194, 200, 235
- 衝突修飾元 256

- 複製
  - 記錄 80
  - 欄位 72, 129
- 複製視覺化 257
- 複製類型屬性 127
- 調整傾向評分
  - 平衡資料 67

## (十六劃)

- 導覽 293
- 整數範圍 124
- 歷程節點 155
  - 概述 155
- 輸入資料的順序 78
- 輸出 293, 294
  - 生成新節點 265
  - 列印 265
  - 匯出 266
  - 儲存 265
  - HTML 266
- 輸出元素 294
- 輸出格式 267
- 輸出節點 263, 267, 269, 271, 274, 282, 286, 288, 289, 290, 291, 293, 294, 295, 332
  - 發佈至網路 265
  - 輸出標籤 267
- 輸出管理器 264
- 輸出檔
  - 儲存 267
  - 「選取」節點
    - 概述 63
- 選取列 (觀察值) 63
- 選取值 240, 243, 245
- 選取節點
  - 從 Web 圖形鏈結中產生 216
  - 從圖形產生 246
- 選項
  - IBM SPSS Statistics 298
- 遺漏值 103, 123, 126
  - 在 Matrix 表格中 269
  - 在「聚合」節點中 68
- 隨機種子值
  - 取樣記錄 151

## (十七劃)

- 儲存
  - 輸出 265
  - 輸出物件 264, 267
- 儲存格式 7
- 儲存體 123
  - 轉換 137, 138

- 儲存體類型
  - 清單 25
- 應用程式範例 3
- 檢視
  - HTML 輸出在瀏覽器中 266
- 營收
  - 評估圖表 223
- 縮放 340
- 總和
  - 統計量輸出 283
  - 設定廣域節點 288
- 總和函數
  - 時間序列聚集 158
- 臨界值
  - 檢視 Bin 臨界值 147
- 點 120
- 點形圖 170
  - 範例 179
  - 2-D 170
- 點陣圖指標
  - 資料庫表格 306
- 點圖 194, 200, 235

## 〔十八劃〕

- 叢集 248, 256
- 叢集樣本 63, 64, 65
- 轉換
  - 重新分類 139, 142
  - 重新編碼 139, 142
- 轉換函數 89
  - 分子階數 89
  - 分母階數 89
  - 延遲 89
  - 差分階數 89
  - 週期性階數 89
- 轉換測量層次 121
- 轉換節點 280
- 轉置節點 154
  - 字串欄位 154
  - 數值欄位 154
  - 欄位名稱 154
- 轉置資料 154
- 鎖定 SuperNode 339
- 顏色圖形套版 162

## 〔十九劃〕

- 繪圖節點 194
  - 外觀標籤 199
  - 使用圖形 199
  - 選項標籤 198
  - 繪圖標籤 197
- 繪製關聯 212

- 鏈結
  - Web 節點 214
- 關聯繪製 212
- 類別資料 121
- 類型 7
- 類型節點
  - 名義資料 125
  - 地理空間資料類型 125
  - 序數資料 125
  - 空白處理 123
  - 清除值 58
  - 設定建模角色 127
  - 設定選項 118, 120, 121
  - 連續資料 124
  - 集合資料類型 125
  - 概述 117
  - 旗標欄位類型 125
  - 複製類型 127
- 類型屬性 127

## 〔二十一劃〕

- 屬性
  - 在地圖中 189
- 欄位
  - 衍生多個欄位 133
  - 重新排序 156
  - 匿名化資料 141
  - 選取多個 134
  - 轉置 154
  - 欄位和值標籤 123
- 欄位名稱 130
  - 匿名 130
  - 資料匯出 302, 314, 315, 321, 335
- 欄位作業節點 103
  - 時間間隔節點 157
  - 從資料審核產生 279
- 欄位的方向 127
- 欄位衍生公式 134
- 欄位重新排序節點 156
  - 自訂排序 156
  - 自動排序 156
  - 設定選項 156
- 欄位儲存
  - 轉換 137
- 欄位類型
  - 在視覺化中 167
- 欄位屬性 127

## 〔二十三劃〕

- 變異數
  - 統計量輸出 283
- 變異數分析
  - 平均數節點 284

- 變數名稱
  - 資料匯出 302, 314, 315, 321, 335
- 變數標籤
  - 統計量匯出節點 315, 334
  - 統計量檔案節點 27, 328
- 變數檔案中的清單 25
- 變數檔案節點 23
  - 地理空間 meta 資料 25
  - 自動日期識別 23
  - 設定選項 23
  - 匯入地理空間資料 25
- 變數類型
  - 在視覺化中 167
- 顯示格式
  - 小數位數 128
  - 分組符號 128
  - 科學 128
  - 貨幣 128
  - 數字 128
- 顯著性
  - 相關強度 282
- 驗證樣本
  - 分割資料 150, 151

## 〔二十五劃〕

- 觀察值資料
  - 資料收集來源節點 28

## 〔數字〕

- 2-D 點形圖 170
- 3-D 直方圖 170
- 3-D 長條圖 170
- 3-D 區域圖
  - 說明 170
- 3-D 密度 170
- 3-D 散佈圖 170
- 3-D 圓餅圖 170

## A

- ADO 資料庫
  - 匯入 28
- Analytic Server 來源 11
- Analytic Server 匯出 317
- ARIMA 模型
  - 轉換函數 89

## B

- Bin 散佈圖 170
  - 六角形 bin 170

## C

- Cell 範圍
  - Excel 檔案 38
- CLEM 表示式 61
- Cognos, 請參閱 IBM Cognos 34
- CPLEX 最佳化節點
  - 設定選項 101
  - 概述 100
- CREATE INDEX 指令 306
- CRISP-DM
  - 資料理解 5
- CRISP-DM 程序模型
  - 資料準備 103
- CSV 資料
  - 匯入 28

## D

- DAT 檔案
  - 匯出 266, 321
  - 儲存 267

## E

- employee\_data.sav 資料檔案 329
- EOL 字元 23
- ESRI 伺服器 57
- ESRI 檔案 189
- Excel
  - 從 IBM SPSS Modeler 啟動 321
  - Excel 來源節點 38
  - Excel 匯入節點
    - 通過輸出產生 321
  - Excel 匯出節點 321
  - Excel 檔案
    - 匯出 321
- E-Plot 節點 235
  - 外觀標籤 236
  - 使用圖形 236
  - 選項標籤 235
  - 繪圖標籤 235

## F

- F 統計量
  - 平均數節點 285
- false 125
- FILLFACTOR 事實索引鍵
  - 索引資料庫表格 306

## H

- hassubstring 函數 136

- HDATA 格式
  - 資料收集來源節點 28
- HTML
  - 儲存輸出 267
- HTML 輸出
  - 在瀏覽器中檢視 266
  - 報告節點 287

## I

- IBM Cognos BI 來源節點
  - 匯入報告 34
- IBM Cognos TM1 來源節點 35
  - 匯入資料 36
- IBM Cognos TM1 匯出節點 319
  - 匯出資料 320
  - 對映匯出資料 320
- IBM Cognos 來源節點 32, 34, 35
  - 匯入資料 33
  - 圖像 32
- IBM Cognos 匯出節點 34, 317, 318
- IBM SPSS Modeler 1
  - 文件 3
- IBM SPSS Modeler Server 1
- IBM SPSS Modeler 解決方案發佈者 325
- IBM SPSS Statistics
  - 有效的欄位名稱 316, 335
  - 從 IBM SPSS Modeler 啟動 298, 315, 332, 335
  - 授權位置 298
- IBM SPSS Statistics 節點 327
- IBM SPSS Statistics 資料檔
  - 匯入意見調查資料 28
- IBM SPSS Statistics 模型 331
  - 進階塊詳細資料 332
  - 模型片段 332
  - 模型選項 331
  - 關於 331
- IBM SPSS Statistics 輸出節點
  - 「輸出」標籤 334
- IBM SPSS 協同作業及部署服務儲存庫
  - 用為視覺化範本、樣式表與地圖的位置 188
- if-then-else 陳述式 137
- In2data 資料庫
  - 匯入 28

## L

- LOESS 光滑線
  - 繪圖節點 197
- E-Plot 節點 235
- lowess 光滑線 請參閱 LOESS 光滑線
  - 繪圖節點 197
- E-Plot 節點 235

## M

- MDD 文件
  - 匯入 28
- meta 資料 123
  - 資料收集來源節點 28
- Microsoft Excel 來源節點 38
- MultiLineString 120

## N

- n 中取 1 取樣 64
- nonrandom 樣本 63, 64

## O

- ODBC
  - 大量載入方式 307, 308
  - 資料庫來源節點 15
  - IBM Cognos 匯出節點的連線 318
- ODBC 匯出節點。請參閱資料庫匯出節點 302
- Oracle 15

## P

- p 值
  - 重要性 285
- Pearson 相關性
  - 平均數節點 286
  - 統計量輸出 283
- python 節點 91, 92, 231, 233, 235
- Python(Y)
  - 大量載入 Script 307, 308

## Q

- Quancept 資料
  - 匯入 28
- Quanvert 資料庫
  - 匯入 28

## R

- recency
  - 設定相關日期 71
- RFM 分析節點
  - 設定 148
  - 概述 147
  - Bin 值 149
- RFM 聚集節點
  - 設定選項 71
  - 概述 70
- ROI
  - 圖表 220, 225

## S

- SAS
  - 設定匯入選項 37
- SAS 來源節點
  - 傳輸檔案 37
  - .sd2 (SAS) 檔 37
  - .ssd (SAS) 檔 37
  - .tpt (SAS) 檔 37
- SAS 匯出節點 321
- Scripting
  - SuperNode 343
- smoother
  - 繪圖節點 197
  - E-Plot 節點 235
- SMOTE 節點 91, 92
- SMZ 檔案
  - 刪除 188
  - 更名 188
  - 建立 189
  - 匯入 188
  - 匯出 188
  - 概述 189
  - 預先安裝的 189
  - 編輯預先安裝的 SMZ 檔案 189
- Space-Time-Boxes 節點
  - 定義密度 96
  - 概述 94
- SPLOM 170
  - 範例 183, 185
- SQL 查詢
  - 資料庫來源節點 15, 16, 21, 22
- SuperNode 337
  - 以下的類型 337
  - 使用備註, 帶有 340
  - 來源 SuperNode 337
  - 放大 340
  - 建立 338
  - 建立快取 342
  - 密碼保護 339, 340
  - 巢狀 339
  - 終端 SuperNode 338
  - 處理 SuperNode 337
  - 設定參數 341
  - 解除鎖定 339
  - 載入 343
  - 編輯 340
  - 儲存 343
  - 鎖定 339
  - Scripting 343
- SuperNode 參數 341, 342
- Surveycraft 資料
  - 匯入 28
- systematic 樣本 63, 64

## T

- t 測試
  - 平均數節點 284, 286
  - 成對樣本 284
  - 獨立樣本 284
- Teradata
  - 查詢級區 19
- The Weather Company 36
- Triple-S 資料
  - 匯入 28
- true 值 125
- TWC 來源 36
- t-SNE 節點 231, 233
- t-SNE 模型塊 235

## U

- UNIQUE 關鍵字
  - 索引資料庫表格 306

## V

- VDATA 格式
  - 資料收集來源節點 28

## W

- Web 節點 212
  - 外觀標籤 215
  - 使用圖形 216
  - 定義鏈結 214
  - 網絡彙總 219
  - 調節器 216
  - 調整臨界值 218
  - 調整點 216
  - 選項標籤 214
  - 繪圖標籤 213
  - 鏈結調節器 216
  - 變更佈置 216

## X

- XLSX 檔案
  - 匯出 321
- XML 來源節點 39
- XML 匯出節點 323
- XML 輸出
  - 報告節點 287
- XPath 語法 39

## [特殊字元]

- auto- 設定 249
- .dbf 檔案 57

- .par 檔 325
- .pim 檔 325
- .sav 檔案 27, 328
- .sd2 (SAS) 檔 37
- .shp 檔案 57
- .sib 檔案 343
- .ssd (SAS) 檔 37
- .tpt (SAS) 檔 37
- .zsav 檔案 27, 328







Printed in Taiwan