

# IBM SPSS Forecasting 20



注意：使用此資訊和支援的產品之前，請先閱讀 注意事項第 101 頁 包含的一般資訊。

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

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

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

授權內容：IBM 資產

**Copyright IBM Corporation 1989, 2011.**

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

---

# 序

IBM® SPSS® Statistics 為分析資料的強大系統。Forecasting 的選用性附加模組能提供其他本手冊所說明的分析技術。Forecasting 的附加模組必須與 SPSS Statistics Core 系統搭配使用，而且是完全整合到系統中。

## 關於 IBM Business Analytics

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

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

## 技術支援

技術支援可提供客戶維護的服務。客戶可以電洽技術支援以取得 IBM Corp. 產品在使用上的協助，或是支援硬體環境的安裝說明。若要取得技術支援，請參閱 IBM Corp. 網站內容，網址：<http://www.ibm.com/support>。請求協助時，請準備好的您個人、組織和支援合約的相關資訊。

## 針對學生用戶的技術支援

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

## 客戶服務

如果您對於自己的貨品或帳號有任何疑問，請聯絡您的當地辦公室。請備妥您的序號以供識別。

## 訓練研討會

IBM Corp. 同時提供公開與線上訓練研討會。所有的研討會皆以傳達工作群為其特色。研討會將定期在各主要城市舉辦。如需研討會的詳細資訊，請移至 <http://www.ibm.com/software/analytics/spss/training>。

## 其他出版品

SPSS Statistics: Guide to Data Analysis (資料分析指南)、SPSS Statistics: Statistical Procedures Companion (統計程序指南) 以及 SPSS Statistics: Advanced Statistical Procedures Companion (進階統計程序指南) 是由 Marija Norušis 撰寫, 由 Prentice Hall 發行, 為推薦的輔助資料。這些出版品涵蓋 SPSS Statistics Base 模組、進階統計量模組和迴歸模組中的統計程序。不論您是資料分析的新手, 還是已經準備使用高階應用程式, 這些書籍都能幫助您善加利用 IBM® SPSS® Statistics 系列產品中的功能。如需其他資訊 (包括出版品內容和章節樣本), 請參閱作者的網站: <http://www.norusis.com>

---

# 內容

## 部 I：使用手冊

<b>1</b>	<b>時間數列簡介</b>	<b>1</b>
	時間數列資料 . . . . .	1
	資料轉換 . . . . .	1
	估計週期和有效期限 . . . . .	2
	建立模式和產生預測 . . . . .	2
<b>2</b>	<b>時間數列 Modeler</b>	<b>3</b>
	指定 Expert Modeler 的選項 . . . . .	5
	模式選擇和事件規格 . . . . .	6
	以 Expert Modeler 處理偏離值 . . . . .	7
	自訂指數平滑化模式 . . . . .	8
	自訂 ARIMA 模式 . . . . .	9
	自訂 ARIMA 模式的模式規格 . . . . .	10
	自訂 ARIMA 模式中的轉換函數 . . . . .	11
	自訂 ARIMA 模式中的偏離值 . . . . .	13
	輸出 . . . . .	14
	統計量和預測表 . . . . .	14
	圖形 . . . . .	16
	將輸出限制於最適或最不適模式 . . . . .	18
	儲存模式預測與模式規格 . . . . .	19
	選項 . . . . .	20
	TSMODEL 指令的其他功能 . . . . .	21
<b>3</b>	<b>套用時間數列模式</b>	<b>23</b>
	輸出 . . . . .	25
	統計量和預測表 . . . . .	26
	圖形 . . . . .	28
	將輸出限制於最適或最不適模式 . . . . .	29
	儲存模式預測與模式規格 . . . . .	31
	選項 . . . . .	32
	TSAPPLY 指令的其他功能 . . . . .	33

<b>4</b>	<b>週期性分解</b>	<b>34</b>
	週期性分解儲存 . . . . .	35
	SEASON 指令的其他功能 . . . . .	36
<b>5</b>	<b>光譜圖</b>	<b>37</b>
	SPECTRA 指令的其他功能 . . . . .	39
 <b>部 II：範例</b>		
<b>6</b>	<b>以 Expert Modeler 進行大量預測</b>	<b>41</b>
	檢驗您的資料 . . . . .	41
	執行分析 . . . . .	42
	模式摘要圖表 . . . . .	49
	模式預測值 . . . . .	50
	摘要 . . . . .	51
<b>7</b>	<b>藉由套用已儲存的模式進行大量重新預測</b>	<b>52</b>
	執行分析 . . . . .	52
	模式適合度統計量 . . . . .	55
	模式預測值 . . . . .	56
	摘要 . . . . .	56
<b>8</b>	<b>使用 Expert Modeler 判斷顯著預測值</b>	<b>57</b>
	繪製您的資料 . . . . .	57
	執行分析 . . . . .	59
	數列圖 . . . . .	65
	模式說明表 . . . . .	65
	模式統計量表 . . . . .	66

ARIMA 模式參數表 . . . . .	66
摘要 . . . . .	66
<b>9 套用已儲存模式來對預測值進行測試</b>	<b>67</b>
延長預測值數列 . . . . .	67
修正預測週期內的預測值 . . . . .	71
執行分析 . . . . .	73
<b>10 週期性分解</b>	<b>77</b>
移除銷售額資料中的週期性 . . . . .	77
決定與設定週期性 . . . . .	77
執行分析 . . . . .	81
瞭解輸出 . . . . .	81
摘要 . . . . .	83
相關程序 . . . . .	83
<b>11 光譜圖</b>	<b>84</b>
使用光譜圖確認週期預期 . . . . .	84
執行分析 . . . . .	84
認識週期值和光譜密度 . . . . .	85
摘要 . . . . .	87
相關程序 . . . . .	87

## 附錄

A	適合度量數	88
B	偏離值類型	89
C	ACF/PACF 圖形指引	90
D	範例檔案	94
E	注意事項	101
	參考書目	103
	索引	105



# 部 1: 使用手冊



# 時間數列簡介

**時間數列**是藉由在一段時間裏，規律地測量某個變數而取得的一組觀察值。例如，在庫存資料的數列中，觀察值可能是代表幾個月來每天的庫存量。顯示產品市場佔有率的數列可能是以幾年來每週的市場佔有率組成。總銷售數字的數列可能是由多年來每月一個觀察值所組成。這些範例中共通的因素就是：在一段時間中，每隔已知的間隔，規律地觀察某個變數。因此，典型時間數列的資料形式是單一個序列，或是代表定期測量結果的觀察值清單。

表格 1-1  
每日庫存時間數列

時間	週	日	庫存量
t <sub>1</sub>	1	週一	160
t <sub>2</sub>	1	週二	135
t <sub>3</sub>	1	週三	129
t <sub>4</sub>	1	週四	122
t <sub>5</sub>	1	週五	108
t <sub>6</sub>	2	週一	150
		...	
t <sub>60</sub>	12	週五	120

進行時間數列分析的最重要理由之一就是要嘗試預測數列的未來數值。能夠解釋過去值的數列模式也可能預測未來幾個值會增加或減少及其增減的幅度。順利作出預測的能力顯然對任何企業或科學領域都很重要。

## 時間數列資料

當您定義時間數列資料以便與「預測附加模組」一起使用時，每一個數列都會對應至不同的變數。例如，若要在「資料編輯程式」中定義時間數列，按一下「變數檢視」索引標籤，然後在任何空白列中輸入變數名稱。在時間數列中的每一個觀察值會對應至觀察值（「資料編輯程式」中的列）。

如果您開啟的試算表中含有時間數列資料，每一個數列必須排列在試算表的行之中。如果您的試算表已有時間數列排列在列中，您可以直接開啟，然後使用「資料」功能表的「轉置」，將橫列的資料翻轉成直行。

## 資料轉換

在 Core 系統提供許多資料轉換程序，對於時間數列分析十分有用。

- 「」 「定義日期」程序（在「資料」功能表中）會產生日期變數，可用來建立週期，並區分過去期間、有效期限和預測週期。「預測」的設計是為處理「定義日期」程序建立的變數。

- 「建立時間數列」程序（在「轉換」功能表中）會建立新的時間數列變數作為現有時間數列變數的函數。其中包含使用相鄰觀察值以供平滑化、平均化及差異化的函數。
- 「置換遺漏值」程序（在「轉換」功能表中）會根據各種方法算出的估計值來置換系統遺漏值和使用者的遺漏值。在數列開始或結束處的遺漏資料不會引發特定的問題，他們只會縮短數列中有用的長度。數列中間的資料間隙（內嵌的遺漏資料）會是更嚴重的問題。

如需有關時間數列資料轉換的詳細資訊，請參閱《Core 系統使用手冊》。

## 估計週期和有效期限

將時間數列分成估計週期或過去期間，以及有效期限經常是很有用的作法。您可以依據過去期間的觀察值發展出模式，然後測試模式在有效期限內是否仍然運作正常。強制讓模式對您已知的時間點（有效期限的時間點）進行預測，能夠了解模式進行預測時的運作狀況。

有效期限中的觀察值一般都稱為保留觀察值，因為是從模式建立過程中保留下來的觀察值。估計週期是由作用中資料集內目前選擇的觀察值組成。最後選擇的觀察值之後的任何觀察值都可以用來作為保留觀察值。一旦您對模式覺得滿意，認為能夠適當執行預測工作，就可以重新定義估計週期，將保留觀察值包含在內，然後建立最終模式。

## 建立模式和產生預測

「預測附加模組」提供兩個程序，以完成建立模式與產生預測的工作。

- **時間數列 Modeler** 程序會建立時間數列的模式，然後產生預測。它包含 Expert Modeler，可自動決定每個時間數列最適合的模式。對想要提高控制程度的資深分析者來說，也提供自訂建立模式的工具。
- **套用時間數列模式** 程序會將現有時間數列模式（由 Time Series Modeler 建立）套用至作用中資料集。這項作業可以在新的或修正過的資料可用時，讓您取得數列預測，而不需再重新建置模式。如果有理由相信模式已經改變，則可使用 Time Series Modeler 重新建置。

# 時間數列 Modeler

時間數列模式工具程序會為時間數列評估指數平滑化、單變量自身迴歸整合移動平均 (ARIMA)，以及多變量 ARIMA (或稱轉換函數模式) 模式，然後產生預測。此程序包括 Expert Modeler 可以自動識別和評估最合適的 ARIMA、或一個或多個依變數之數列的指數平滑化模式，如此一來，就不需要透過嘗試與錯誤來識別適當的模型。另外，您也可以指定自訂 ARIMA 或指數平滑化模式。

**範例。**您是產品經理，負責就 100 種不同產品中的每一種產品，預測下個月的單位銷售量和收益，但是您對建立時間數列模式幾乎可說沒有經驗。過去所有 100 種產品的單位銷售資料儲存在一份 Excel 試算表中。以 IBM® SPSS® Statistics 開啟試算表之後，您使用「Expert Modeler」，並要求進行未來一個月的預測。「Expert Modeler」為您所負責每一項產品的單位銷售量找出最適合的模式，然後使用這些模式來產生預測。由於「Expert Modeler」能夠處理多重輸入數列，您只需執行程序一次，就能取得所有產品的預測。如果選擇將預測儲存至作用中資料集，您就能輕易地匯出結果存回 Excel 之中。

**統計量。**適合度量數：平穩 R 平方、R 平方 ( $R^2$ )、均方誤差平方根 (RMSE)、平均絕對誤差 (MAE)、平均絕對百分比誤差 (MAPE)、最大絕對誤差 (MaxAE)、最大絕對百分比誤差 (MaxAPE)，以及正規化 Bayesian 訊息準則 (BIC)。殘差：自我相關函數、偏自我相關函數、Ljung-Box Q。若是 ARIMA 模式：依變數的 ARIMA 階數、自變數的轉換函數階數，以及偏離值估計。還有指數平滑化模式的平滑化參數估計。

**圖形。**跨所有模式的摘要圖形：平穩 R 平方、R 平方 ( $R^2$ )、均方誤差平方根 (RMSE)、平均絕對誤差 (MAE)、平均絕對百分比誤差 (MAPE)、最大絕對誤差 (MaxAE)、最大絕對百分比誤差 (MaxAPE)、常態化 Bayesian 資訊條件 (BIC) 的直方圖；殘差自身相關和偏自身相關的盒形圖。各個模式的結果：預測值、適合度值、觀察值、信賴區間上限和下限、殘差自身相關和偏自身相關。

## 時間數列 Modeler 資料考量

**資料。** 依變數和任何自變數應該都是數值。

**假設。** 依變數和任何自變數全部都視為時間數列，也就是每個觀察值都代表一個時間點，而接連相續的觀察值都以常數時間區間分開。

- **穩定性。**對於自訂 ARIMA 模式，要作為模式的時間數列必須為平穩型的。轉換非平穩型數列為平穩型的最有效方法是透過差分轉換 — 在「建立時間數列」對話方塊「中提供使用。
- **預測。**若要利用包含自變數 (預測值) 的模式來產生預測，作用中資料集必須包含預測期間內所有觀察值的變數值。此外，自變數也不應該在估計期間中包含任何遺漏值。

## 定義日期

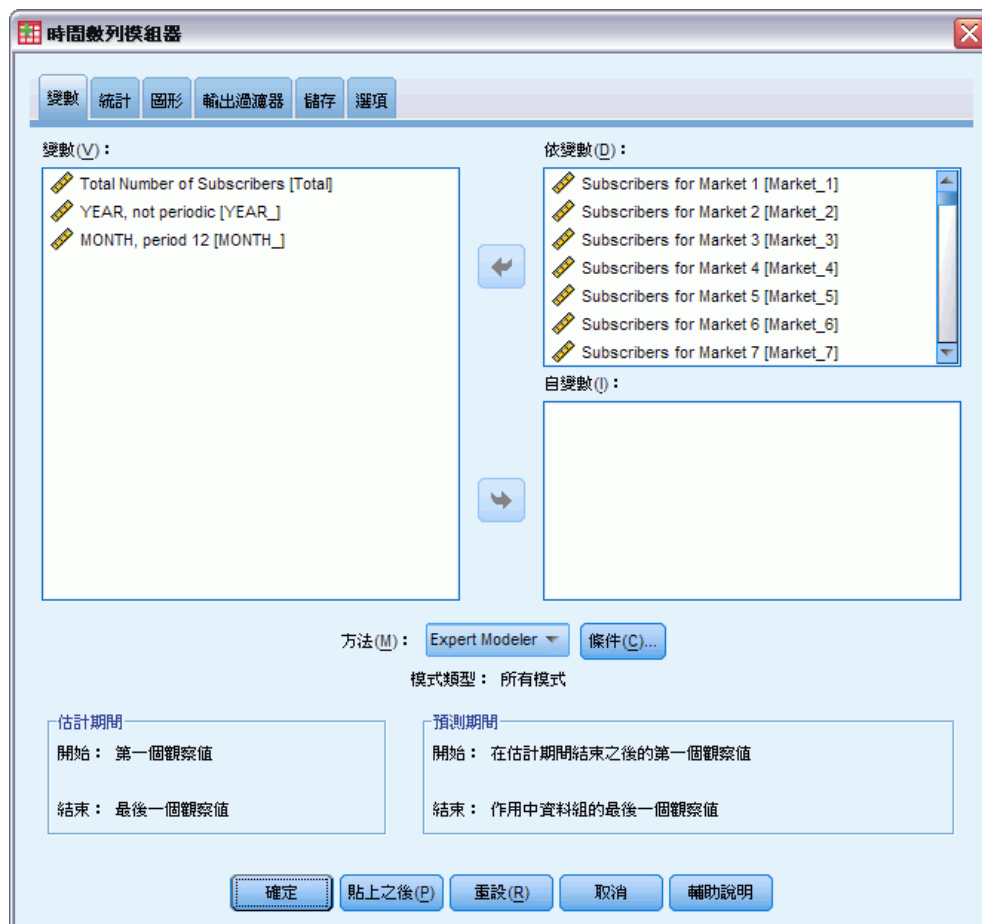
雖然並非必要，但是建議您使用「」 「定義日期」對話方塊，指定與第一個觀察值相關聯之資料以及連續觀察值間的時間區間。這是在使用「時間數列 Modeler」之前完成的作業，結果會得到一組標示與各觀察值相關日期的變數。同時也設定資料的假設週期 —

例如，如果連續觀察值之間的時間區間是一個月，則週期為 12。如果您想要建立季節性模式，就必須有這個週期。如果並不想要季節性模式，輸出上也不需要日期標記，就可以略過「定義日期」對話方塊。此時，與各個觀察值相關聯的標記就是觀察值編號。

### 若要使用時間數列 Modeler

- ▶ 從功能表選擇：  
分析 > Forecasting > 建立模式...

圖表 2-1  
「時間數列 Modeler」，「變數」索引標籤



- ▶ 在「變數」索引標籤上，選擇一個或多個要建模變數。
- ▶ 從「方法」下拉式清單方塊中選擇建模方法。在自動化建模下，保留 **Expert Modeler** 的預設方法。這樣會呼叫「Expert Modeler」，判斷各個依變數的最適合模式。  
若要產生預測：
  - ▶ 按一下「選項」索引標籤。
  - ▶ 指定預測期間。這樣會產生圖表，其中包含預測和觀察值。

您可以：

- 選取一個或多個自變數。自變數在迴歸分析中幾乎被視為預測變數處理，但為選擇性的。可以包含在 ARIMA 模式中，但不能包含在指數平滑化模式中。如果指定 Expert Modeler 作為建模方法，並包含自變數，則只會考慮使用 ARIMA 模式。
- 按一下「條件」，指定模式詳細資料。
- 儲存預測、信賴區間和噪音殘差。
- 將估計模式存成 XML 格式。儲存的模式可以套用至新的或修正過的資料，以取得更新的預測而毋需重新建置模式。這項作業利用[套用時間數列模式](#)程序加以完成。
- 取得所有估計模式的摘要統計量。
- 指定自訂 ARIMA 模式中自變數的轉換函數。
- 啟用自動偵測偏離值。
- 以模式特定時間點作為自訂 ARIMA 模式的偏離值。

### 建模方法

可用的建模方法包括：

**Expert Modeler。**「Expert Modeler」會自動為各個依變數列找尋最適合的模式。如果指定了自（預測）變數，「Expert Modeler」會選擇與依變數列有統計顯著關係的自變數，納入 ARIMA 模式中。若適當，可利用差分及/或平方根或自然對數進行模式變數轉換。依照預設值，「Expert Modeler」會同時考量指數平滑化和 ARIMA 模式。但是您可以將「Expert Modeler」限制為只搜尋 ARIMA 模式，或是只搜尋指數平滑化模式。您也可以指定自動偵測偏離值。

**指數平滑化。**使用此選項，可指定自訂指數平滑化模式。您可以從數種指數平滑化模式中加以選擇，模式的差異在於對趨勢與季節性的處理方法。

**ARIMA 程序**使用此選項，可指定自訂 ARIMA 模式。這項作業涉及指定自我迴歸和移動平均階數，以及差分次數。您可以納入自（預測）變數，並定義任何或全部變數的轉換函數。您也可以指定自動偵測偏離值，或是明確指定一組偏離值。

### 估計和預測期間

**估計期間。**估計期間定義用來決定模式的觀察值集。依照預設值，估計期間包括作用中資料集內所有觀察值。若要設定估計期間，請選擇「」「選擇觀察值」對話方塊中的「以時間或觀察值範圍為準」。依可用的資料而定，程序所使用的估計期間可能會因依變數而異，因此與所顯示的值不同。對一個特定的依變數來說，真正的估計期間是除去任何發生於指定估計期間的開始或結束時的連續變數遺漏值之後所餘下的期間。

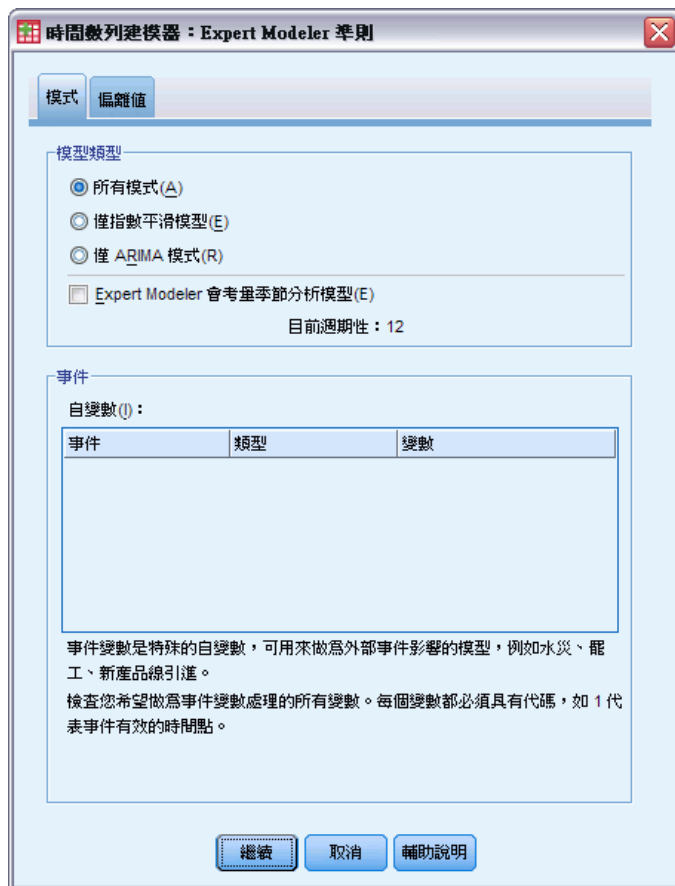
**預測期間。**預測期間是從估計期間之後的第一個觀察值開始，依照預設值會一直到作用中資料集的最後一個觀察值才結束。您可以從「[選項](#)」索引標籤，設定預測期間的結束時間。

## 指定 Expert Modeler 的選項

「Expert Modeler」提供選項，可限制候選模式集、指定如何處理偏離值、並包含事件變數。

## 模式選擇和事件規格

圖表 2-2  
「Expert Modeler 條件」對話方塊，「模式」索引標籤



「模式」索引標籤可以讓您指定由「Expert Modeler」考量的模式類型，並指定事件變數。

**模式類型。**您可以使用的選項如下：

- **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」只會考量指數平滑化模式。
- **僅 ARIMA 模式。**「Expert Modeler」只會考量 ARIMA 模式。

**「Expert Modeler」會考量季節性模式。**此選項只有在已為作用中資料集定義週期後才能啟用。選擇（核取）此選項時，「Expert Modeler」會同時考量季節性和非季節性模式。如果未選取此選項，則「Expert Modeler」只考量非季節性模式。

**目前的週期。**表示定義作用中資料集目前的週期性（若有的話）。目前的週期性是以整數表示——例如，12 表示年週期性，而各觀察值代表一個月。如果尚未設定週期性，則值顯示為無。週期模式一定要有週期性。您可以在「定義日期」對話方塊中設定週期性。

**事件。**選擇任何視為事件變數處理的自變數。若是事件變數，值為 1 的觀察值表示依變數列預期受事件影響的次數。1 以外的值表示沒有影響。



## 以 Expert Modeler 處理偏離值

圖表 2-3

「Expert Modeler 條件」對話方塊，「偏離值」索引標籤



「偏離值」索引標籤可以讓您選擇自動偵測偏離值以及要偵測的偏離類型。

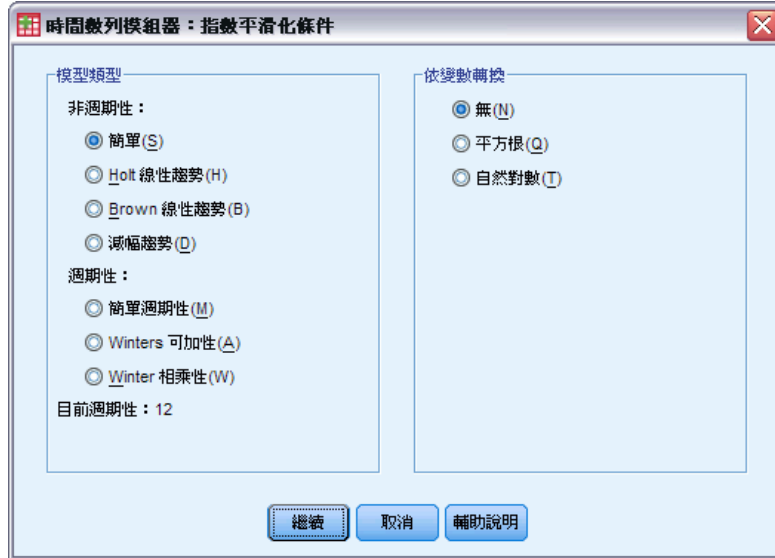
**自動偵測偏離值。**依照預設值，不會執行自動偵測偏離值。選擇（核取）此選項以執行自動偵測偏離值，然後從下面所列選擇一個或多個偏離值類型：

- 可加性
- 水準變遷
- 創新性
- 瞬變
- 週期可加性
- 局部趨勢
- 可加性補充

如需詳細資訊，請參閱第 89 頁附錄 B 中的偏離值類型。

## 自訂指數平滑化模式

圖表 2-4  
「指數平滑化條件」對話方塊



**模式類型。**指數平滑化模式(Gardner, 1985)分成季節性或非季節性兩種。季節性模式只有在已定義作用中資料集的週期後才可使用（請參閱下文中「目前的週期」一節）。

- **簡單。**簡單模式適用於不具有趨勢或週期性的數列。水準是它唯一的平滑化參數。簡單指數平滑化與 ARIMA 模式最為類似之處在於：具有自身迴歸階數 0、差分階數 1、移動平均數階數 1，以及無常數。
- **Holt's 線性趨勢。**此模式適用於具有線性趨勢且沒有週期的數列。其平滑化參數為水準與趨勢，且不受彼此的數值所限制。Holt's 模式較 Brown's 模式更為普遍，但對於大型數列的計算時間會較長。Holt's 指數平滑化與 ARIMA 模式最為類似之處在於：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。
- **Brown's 線性趨勢。**此模式適用於具有線性趨勢且沒有週期的數列。其平滑化參數為水準與趨勢，且皆已假定為相等。因此，Brown's 模式為 Holt's 模式的特殊觀察值。Brown's 指數平滑化與 ARIMA 模式最為相似，其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。
- **減幅趨勢。**此模式適用於具有漸失線性趨勢且沒有週期的數列。其平滑化參數為水準、趨勢和減幅趨勢。減幅指數平滑化與 ARIMA 模式最為類似之處在於：具有自身迴歸階數 1、差分階數 1，以及移動平均數階數 2。
- **簡單週期性。**此模式適用於不包含趨勢、且週期性效應在一段時間保持不變的數列。其平滑化參數為水準和週期。簡單週期性指數平滑化與 ARIMA 模式最為類似之處在於：具有自身迴歸階數 0、差分階數 1、週期差分階數 1，以及移動平均階數 1、 $p$ 、和  $p + 1$ ，其中  $p$  為週期區間內的期間個數（以每月資料來說， $p=12$ ）。

- **Winters 可加性。** 此模式可用於具有線性趨勢的數列，以及不依數列水準的週期效應。其平滑化參數可為水準、趨勢和週期。Winters 可加性指數平滑化與 ARIMA 模式最為相似，都具有自身迴歸階數 0、差分階數 1、週期差分階數 1 和移動平均數階數  $p + 1$ ，其中  $p$  為週期區間內的期間個數（以每月資料來說， $p = 12$ ）。
- **Winters 相乘性。** 此模式可用於具有線性趨勢的數列，以及依據數列水準的週期效應。其平滑化參數可為水準、趨勢和週期。Winters 相乘性指數平滑化和 ARIMA 模式不相似。

**目前的週期。** 表示定義作用中資料集目前的週期性（若有的話）。目前的週期性是以整數表示 — 例如，12 表示年週期性，而各觀察值代表一個月。如果尚未設定週期性，則值顯示為無。週期模式一定要有週期性。您可以在「定義日期」對話方塊中設定週期性。

**依變數轉換。** 您可以指定先在各個依變數上執行轉換，然後再進行模式化作業。

- **無。** 沒有執行任何轉換。
- **平方根。** 平方根轉換。
- **自然對數。** 自然對數轉換。

## 自訂 ARIMA 模式

「時間數列 Modeler」可以讓您建立自訂非季節性或季節性 ARIMA（自我迴歸整合移動平均）模式，又稱為 Box-Jenkins（盒形圖，Jenkins, 和 Reinsel, 1994）模式，無論使用或不使用一固定的預測變數集合。您可以為任何或全部預測變數定義轉換函數，並指定自動偵測偏離值，或是明確指定一組偏離值。

- 在「變數」索引標籤上指定的所有自（預測）變數都會明確地包含在模式中。這種作法與使用「Expert Modeler」時相反，後者只在自變數與依變數有統計上顯著關係時才包含自變數。

## 自訂 ARIMA 模式的模式規格

圖表 2-5  
「ARIMA 條件」對話方塊，「模式」索引標籤

時間數列建模器：ARIMA 條件

模式 偏離值

ARIMA 順序

結構(S):

	非週期性	週期性
自身迴歸 (p)	0	0
差分 (d)	0	0
移動均數 (q)	0	0

目前週期性：12

依變數轉換

無 (N)

平方根 (Q)

自然對數 (T)

常數項納入模式中(I)

繼續 取消 輔助說明

「模式」索引標籤可以讓您指定自訂 ARIMA 模式的結構。

**ARIMA 階數。**將模式中各種不同 ARIMA 成份的數值輸入「結構」網格中的相對應儲存格。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。所有正較低階數都會包含在模式中。例如，如果指定 2，模式就會包含階數 2 和 1。「季節性」欄中的儲存格只有在已定義作用中資料集週期後才能啟用（請參閱下文中「目前的週期」一節）。

- **自我迴歸 (p)。**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，自我迴歸階數 2 指定要使用過去兩段時間的數列值來預測目前值。
- **差分 (d)。**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。差分的階數對應於數列趨勢的程度，第一階差分代表線性趨勢，第二階差分代表二次趨勢，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。

**季節性階數。**季節性自我迴歸、移動平均數，和差分成份所扮演角色與其非季節性對等項目相同。但對季節性階數來說，目前數列值是受由一個或多個季節性期間分隔的先前數列值影響。例如，以每月資料（季節性期間為 12）來說，季節性階數 1 代表目前

數列值受 12 個週期之前的數列值影響。如此對每月資料來說，季節性階數 1 就與指定非季節性階數 12 相同。

**目前的週期。** 表示定義作用中資料集目前的週期性（若有的話）。目前的週期性是以整數表示 — 例如，12 表示年週期性，而各觀察值代表一個月。如果尚未設定週期性，則值顯示為無。週期模式一定要有週期性。您可以在「定義日期」對話方塊中設定週期性。

**依變數轉換。** 您可以指定先在各個依變數上執行轉換，然後再進行模式化作業。

- **無。** 沒有執行任何轉換。
- **平方根。** 平方根轉換。
- **自然對數。** 自然對數轉換。

**常數項納入模式中。** 納入常數是標準作業，除非您確定整體平均數列值是 0；套用差分時，則建議排除常數。

## 自訂 ARIMA 模式中的轉換函數

圖表 2-6  
「ARIMA 條件」對話方塊，「轉換函數」索引標籤

時間數列建模器：ARIMA 條件

模式 轉換函數 偏離值

自變數

Number of Catalogs M...

轉換函數順序

結構(S):

	非週期性	週期性
分子	0	0
分母	0	0
差異	0	0

目前週期性：無

延遲(D): 0

依變數轉換

無(N)

平方根(Q)

自然對數(I)

繼續 取消 輔助說明

「轉換函數」索引標籤（只有在指定自變數時才可以會出現）可以讓您定義在「變數」索引標籤上指定的任何或全部自變數之轉換函數。轉換函數可以讓您指定自（預測）變數過去值的使用方式，以預測依變數數列的未來值。

**轉換函數階數。** 將轉換函數中各種不同成份的數值輸入「結構」網格中的相對應儲存格。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。所有正較低階數都會包含在模式中。此外，一定會包含階數 0 供分子成份使用。例如，如

果為分子指定 2，模式就會包含階數 2、1 和 0。如果為分母指定 3，則模式會包含 3、2 和 1。「季節性」欄中的儲存格只有在已定義作用中資料集的週期後才能啟用（請參閱下文中「目前的週期」一節）。

- **分子。**轉換函數的分子階數。指定所選取自（預測）變數列中，哪些先前值用來預測依變數列的目前值。例如，分子階數 1 指定過去一段時間的自變數列值 — 以及自變數列的目前值 — 會用來預測各依變數列的目前值。
- **分母。**轉換函數的分母階數。指定要如何使用所選取自（預測）變數列先前值的數列平均數離差來預測依變數列的目前值。例如，分母階數 1 指定預測各依變數列的目前值時，必須考慮過去一段時間自變數數平均值的離差。
- **差分。**指定在估計模式之前套用至所選擇自（預測）變數列的差分階數。當趨勢存在時，就必須對數列進行差分，以移除趨勢的影響。

**季節性階數。** 季節性分子、分母，和差分成份所扮演角色與其非季節性對等項目相同。但對季節性階數來說，目前數列值是受由一個或多個季節性期間分隔的先前數列值影響。例如，以每月資料（季節性期間為 12）來說，季節性階數 1 代表目前數列值受 12 個週期之前的數列值影響。如此對每月資料來說，季節性階數 1 就與指定非季節性階數 12 相同。

**目前的週期。** 表示定義作用中資料集目前的週期性（若有的話）。目前的週期性是以整數表示 — 例如，12 表示年週期性，而各觀察值代表一個月。如果尚未設定週期性，則值顯示為無。週期模式一定要有週期性。您可以在「定義日期」對話方塊中設定週期性。

**延遲。** 設定延遲會造成自變數的影響依指定區間數延遲。例如，如果延遲設定為 5，在時間  $t$  的自變數值不會影響預測，要一直到經過五個週期 ( $t + 5$ ) 之後才会有影響。

**轉換。** 一個自變數集合的轉換函數規格，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。** 平方根轉換。
- **自然對數。** 自然對數轉換。

## 自訂 ARIMA 模式中的偏離值

圖表 2-7  
「ARIMA 條件」對話方塊，「偏離值」索引標籤



「偏離值」索引標籤提供下列選擇，可供處理偏離值 (Pena, Tiao, 和 Tsay, 2001)：自動偵測偏離值、指定特定時間點作為偏離值，或是不偵測或模式化偏離值。

**不偵測或模式化偏離值。** 依照預設值，偏離值既不加以偵測，也不進行模式化。選擇此選項，可停用任何偏離值偵測或模式化。

**自動偵測偏離值。** 選擇此選項以執行自動偵測偏離值，然後從下面所列選取一個或多個偏離值類型：

- 可加性
- 水準變遷
- 創新性
- 瞬變
- 週期可加性
- 局部趨勢
- 可加性補充

如需詳細資訊，請參閱第 89 頁附錄 B 中的偏離值類型。

**模式特定時間點作為偏離值。** 選擇此選項，可指定特定時間點作為偏離值。每個偏離值在「偏離值定義」網格中分別使用一列。在一列的所有儲存格中輸入數值。

- **類型。** 偏離值類型。支援的類型如下：可加性（預設值）、水準變遷、創新性、瞬變、週期可加性，以及局部趨勢。

注意 1：如果沒有為作用中資料集定義日期規格，「偏離值定義」網格只顯示一欄「觀察」。若要指定偏離值，請輸入相關觀察值的列編號（顯示於「資料編輯程式」中）。

注意 2：「偏離值定義」網格中的「循環」欄（若有）是指向作用中資料集內 CYCLE\_變數的值。

## 輸出

可用的輸出包括各個模式的結果以及計算所有模式的結果。各個模式的結果可以根據使用者指定的條件，限制為最適或最不適模式集。

## 統計量和預測表

圖表 2-8  
「時間數列 Modeler」，「統計量」索引標籤





「統計量」索引標籤提供顯示模式化結果的表格之選項。

**顯示適合度量數、Ljung-Box 統計量，以及各模式的偏離值數目。**選擇（核取）此選項，可顯示包含所選擇適合度量數、Ljung-Box 值，以及各個估計模式偏離值數目的表格。

**適合量數。**您可以選擇下列一個或多個以納入含有各估計模式適合度的表格中：

- 平穩 R 平方
- R 平方
- 根平均平方誤
- 平均絕對百分比誤差
- 平均絕對誤差
- 最大絕對百分比誤差
- 最大絕對誤差
- 正規化 BIC

如需詳細資訊，請參閱第 88 頁附錄 A 中的適合度量數。

**比較模式的統計量。**這組選項控制包含計算所有估計模式統計量表格的顯示。各選項會分別產生一個表格。您可以選取下列一個或多個選項：

- **適合度。**摘要統計量和平穩 R 平方的百分位數、R 平方、根平均平方誤、平均絕對百分比誤差、平均絕對誤差、最大絕對百分比誤差，最大絕對誤差，以及常態化 Bayesian 資訊準則表。
- **殘差自我相關函數 (ACF)。**摘要統計量和所有已估計模式殘差的自身相關百分位數表。
- **殘差偏自我相關函數 (PACF)。**摘要統計量和所有已估計模式殘差的偏自身相關百分位數表。

**各個模式的統計量。**這組選項控制包含各估計模式詳細資料表格的顯示。各選項會分別產生一個表格。您可以選取下列一個或多個選項：

- **參數估計值。**顯示各估計模式的參數估計表。指數平滑化和 ARIMA 模式會分別顯示不同的表格。如果偏離值存在，其參數估計也會另外顯示於一個表格中。
- **殘差自我相關函數 (ACF)。**依落差顯示各個已估計模式的殘差自身相關表。該表中包含自身相關的信賴區間。
- **殘差偏自我相關函數 (PACF)。**依落差顯示各個已估計模式的殘差偏自身相關表。該表中包含偏自身相關的信賴區間。

**顯示預測。**顯示各估計模式的模式預測和信賴區間表。預測期間是在「選項」索引標籤中設定。

## 圖形

圖表 2-9  
「時間數列 Modeler」, 「圖形」索引標籤



「圖形」索引標籤提供顯示模式化結果的圖形之選項。

### 比較模式的圖形

這組選項控制包含計算所有估計模式統計量圖形的顯示。各選項會分別產生一個圖形。您可以選取下列一個或多個選項：

- 平穩 R 平方
- R 平方
- 根平均平方誤
- 平均絕對百分比誤差
- 平均絕對誤差
- 最大絕對百分比誤差
- 最大絕對誤差

- 正規化 BIC
- 殘差自我相關函數 (ACF)
- 殘差偏自我相關函數 (PACF)

如需詳細資訊，請參閱第 88 頁附錄 A 中的適合度量數。

### 各個模式的圖形

**數列。** 選擇（核取）此選項，可取得各估計模式預測值的圖形。您可以選擇下列一個或多個選項，以包含在圖形之中：

- **觀察值。** 依變數列的觀察值。
- **預測。** 預測期間的模式預測值。
- **適合度值。** 估計期間的模式預測值。
- **預測的信賴區間。** 預測期間的信賴區間。
- **適合度值的信賴區間。** 估計期間的信賴區間。

**殘差自我相關函數 (ACF)。** 顯示各個已估計模式的殘差自身相關圖形。

**殘差偏自我相關函數 (PACF)。** 顯示各個已估計模式的殘差偏自身相關圖形。

## 將輸出限制於最適或最不適模式

圖表 2-10  
「時間數列 Modeler」，「輸出過濾器」索引標籤



「輸出過濾器」索引標籤提供選項，可將表格和圖表輸出限制於估計模式子集。您可以選擇根據您所提供的適合度條件，將輸出限制於最適及/或最不適模式。依照預設值，所有估計模式都會包含在輸出之中。

**最適模式。**選擇（核取）此選項，可將最適模式包含在輸出中。選擇適合度量數，並指定要包含的模式數目。選擇此選項並不會阻止同時選取最不適模式。在此情況下，輸出將包含最不適模式及最適模式。

- **固定的模式數目。**指定結果顯示  $n$  個最適模式。如果數目超過估計模式的總數，則顯示所有模式。
- **模式總數的百分比。**指定結果顯示所有估計模式中最高  $n$  百分比之適合度值的模式。

**最不適模式。**選擇（核取）此選項，可將最不適模式包含在輸出中。選擇適合度量數，並指定要包含的模式數目。選擇此選項並不會阻止同時選取最適模式。在此情況下，輸出將包含最適模式及最不適模式。

- **固定的模式數目。**指定結果顯示  $n$  個最不適應模式。如果數目超過估計模式的總數，則顯示所有模式。
  - **模式總數的百分比。**指定結果顯示所有估計模式中最低  $n$  百分比之適合度值的模式。
- 適合度量數。**選擇要用來篩選模式的適合度量數。預設值是平穩  $R$  平方。

## 儲存模式預測與模式規格

圖表 2-11  
「時間數列 Modeler」，「儲存」索引標籤



「儲存」索引標籤可以讓您將模式預測儲存為作用中資料集內的新變數，並將模式規格儲存為 XML 格式的外部檔案。

**儲存變數。**您可以將模式預測、信賴區間及殘差儲存為作用中資料集的新變數。各依變數列會產生其自身的新變數集，而各個新變數都會包含估計和預測期間的數值。如果預測期間延伸至依變數列長度之外，則加入新的觀察值。為每一個新變數選擇相關聯的「儲存」核取方塊，以選擇儲存新變數。依照預設值，不會儲存任何新變數。

- **預測值。**模式預測值。
- **信賴區間下限。**預測值的信賴區間下限。

- **信賴區間上限。** 預測值的信賴區間上限。
- **噪音殘差。** 模式殘差。執行依變數轉換時（例如，自然對數轉換），這些都是轉換數列的殘差。
- **變數名稱字首。** 指定要用來為新變數命名時的字首，或者保留預設字首。變數名稱是由字首、相關依變數的名稱，以及模式識別碼組成。必要時變數名稱會延伸，以避免與變數命名規則相衝突。字首必須與有效的變數名稱規則相符。

**匯出模式檔案。** 所有估計模式的模式規格都匯出至指定的 XML 格式檔案。已儲存的模式可根據更新的資料，使用 [套用時間數列模式](#) 程序，以取得更新的預測。

## 選項

圖表 2-12  
「時間數列 Modeler」，「選項」索引標籤

時間數列模組器

變數 統計 圖形 輸出過濾器 儲存 選項

預測期間

通過現用資料組最後觀察值的估計期間結束之後的第一個觀察值 (F)  
 通過指定日期的估計期間結束之後的第一個觀察值 (C)

日期 (D):

年:	月:
2004	3

使用者遺漏值

視為無效 (I)  
 視為有效 (V)

信賴區間寬度 (%) (W): 95

輸出中的模型識別符號前置符號 (P): 模式

ACF 和 PACF 輸出中顯示的落差最大數目 (L): 24

確定 貼上之後 (P) 重設 (R) 取消 輔助說明

「選項」索引標籤可以讓您設定預測期間、指定如何處理遺漏值、設定信賴區間寬度，自訂模式識別碼字首，以及設定自身相關顯示的落差數目。

**預測期間。** 預測期間一定都是從估計期間結束後的第一個觀察值開始（用來決定模式的觀察值集），然後一直執行到作用中資料集的最後一個觀察值，或是使用者指定的日期為止。依照預設值，估計模式的結束是作用中資料集的最後一個觀察值，但可以從「選擇觀察值」對話方塊中變更，只要選擇「以時間或觀察值範圍為準」即可。

- **估計期間結束後的第一個觀察值一直到作用中資料集的最後一個觀察值。** 若估計期間結束早於作用中資料集的最後一個觀察值，而您要一直預測到最後一個觀察值，請選擇此選項。這個選項一般都是用來為保留週期產生預測，讓模式預測能夠與實際數值子集進行比較。
- **估計期間結束之後第一個觀察值一直到指定日期。** 選擇此選項，明確指定預測期間結束時間。這個選項一般是用來產生實際數列結束之外的預測。輸入「日期」網格中所有儲存格的數值。

如果沒有為作用中資料集定義日期規格，「日期」網格只顯示一欄「觀察」。若要指定預測期間的結束時間，請輸入相關觀察值的列編號（顯示於「資料編輯程式」中）。

「資料」網格中的「循環」欄（若有）是指向作用中資料集內 CYCLE\_ 變數的值。

**使用者遺漏值。** 這些選項可控制使用者遺漏值之處理。

- **視為無效。** 使用者遺漏值會被視為系統遺漏值來處理。
- **視為有效。** 使用者遺漏值被視為有效資料。

**遺漏值原則。** 下列規則在模式化程序中應用於遺漏值之處理（包括系統遺漏值與視為無效的使用者遺漏值）：

- 模式中包含具有發生於估計期間內之依變數遺漏值的觀察值。遺漏值的處理方式依估計方法而定。
- 如果自變數在估計期間內有遺漏值，則發出警告。對於「Expert Modeler」來說，含有自變數的模式是在無變數的情況下進行估計。對於自訂 ARIMA 來說，含有自變數的模式不進行估計。
- 如果任何自變數在預測期間內有遺漏值，則程序發出警告，並盡可能預測到最遠的範圍。

**信賴區間寬度 (%)。** 信賴區間是供模式預測和殘差自身相關計算使用。您可以指定任何小於 100 的正數值。依照預設值，使用的是 95% 信賴區間。

**輸出中模式識別碼的字首。** 在「變數」索引標籤上指定的每個依變數都會分別產生估計模式。模式以獨特的名稱加以區別，名稱由可自訂字首與整數字尾組成。您可以輸入字首，或保留「模式」的預設值。

**顯示於 ACF 與 PACF 輸出中的最大落差數目。** 您可以設定顯示於自身相關與偏自身相關表格和圖形中的最大落差數目。

## TSMODEL 指令的其他功能

如果您將選取部分貼入語法視窗，並且編輯所產生的 TSMODEL 指令語法，就可以自訂時間數列模式化。指令語法語言讓您可以：

- 指定資料的季節性期間（利用 AUXILIARY 次指令上的 SEASONLENGTH 關鍵字）。這樣會覆寫作用中資料集的目前季節性（若有的話）。

- 指定自訂 ARIMA 和轉換函數成份的不連續落差（利用 **ARIMA** 和 **TRANSFERFUNCTION** 次指令）。例如，您可以指定自訂 ARIMA 模式，具有自我迴歸落差 1、3 和 6；或是指定轉換函數，含有分子落差 2、5 和 8。
- 提供多組模式化規格（例如，建模方法、ARIMA 階數、自變數等等），以供單次執行「時間數列 Modeler」程序（利用 **MODEL** 次指令）。

如需完整的語法資訊，請參閱《指令語法參考手冊》。



# 套用時間數列模式

「套用時間數列模式」程序會從外部檔案載入現有時間數列模式，並將其套用至作用中資料集。您可以使用此程序，取得新的或修正過資料的數列預測，而不需再重新建置模式。模式是使用時間數列 [Modeler](#) 程序產生的。

**範例。** 您是一家大零售商店掌管庫存的經理，負責的產品達 5000 種。您使用了 Expert Modeler 建立模式，以預測未來三個月各種產品的銷售量。您的資料倉庫每個月都會以實際銷售資料更新，您要利用資料倉庫來製作每個月的更新預測。「套用時間數列模式」程序可以讓您利用原有的模式完成這項作業，只要重新估計模式參數以涵蓋新資料即可。

**統計量。** 適合度量數：平穩 R 平方、R 平方 ( $R^2$ )、均方誤差平方根 (RMSE)、平均絕對誤差 (MAE)、平均絕對百分比誤差 (MAPE)、最大絕對誤差 (MaxAE)、最大絕對百分比誤差 (MaxAPE)，以及正規化 Bayesian 訊息準則 (BIC)。殘差：自我相關函數、偏自我相關函數、Ljung-Box Q。

**圖形。** 跨所有模式的摘要圖形：平穩 R 平方、R 平方 ( $R^2$ )、均方誤差平方根 (RMSE)、平均絕對誤差 (MAE)、平均絕對百分比誤差 (MAPE)、最大絕對誤差 (MaxAE)、最大絕對百分比誤差 (MaxAPE)、常態化 Bayesian 資訊條件 (BIC) 的直方圖；殘差自身相關和偏自身相關的盒形圖。各個模式的結果：預測值、適合度值、觀察值、信賴區間上限和下限、殘差自身相關和偏自身相關。

## 套用時間數列模式資料考量

**資料。** 要套用模式的變數（依變數和自變數）必須為數值。

**假設。** 模式套用至作用中資料集內（名稱與模式中所指定變數相同）的變數。這些變數全部都視為時間數列，也就是每一個觀察值都代表一個時間點，而接連相續的觀察值都以常數時間間隔分開。

- **預測。** 若要利用包含自變數（預測值）的模式來產生預測，作用中資料集必須包含預測期間內所有觀察值的變數值。如果模式參數已重新估計，則自變數不可在估計週期包含任何遺漏值。

## 定義日期

「套用時間數列模式」程序要求作用中資料集的週期性（若有的話）與所要套用模式的週期性相符。如果就利用與建立模式所用的資料集進行預測（也許用新的或修正過的資料），則已滿足此條件。如果作用中資料集並沒有週期性，您將有機會巡覽至「定義日期」對話方塊，建立一個週期性。但是如果模式是在未指定週期性下建立，則作用中資料集也不可以有週期性。

## 若要套用模式

- ▶ 從功能表選擇：  
分析 > 預測 > 套用模式...

圖表 3-1  
「套用時間數列模式」，「模式」索引標籤



- ▶ 輸入模式檔的檔案規格，或按一下「瀏覽」，然後選擇模式檔案（模式檔案是以時間數列 Modeler 程序建立）。

您可以：

- 使用作用中資料集的資料，重新估計模式參數。預測是使用重新估計過的參數建立的。
- 儲存預測、信賴區間和噪音殘差。
- 將重新估計過的模式存成 XML 格式。

### 模式參數和適合度量數

**從模式檔案載入。** 預測是使用模式參數從未重新估計參數之模式檔案產生的。顯示於輸出中並用來篩選模式（最適或最不適）的**適合度量數**是取自模式檔案，並反映開發（或最後更新）各模式時所使用的資料。在此選項下，預測不會涵蓋作用中資料集中（不管是依變數或自變數）的過去資料。如果要讓過去的資料影響預測，就必須選擇「從資料重新估計」。此外，預測也不會估計到預測週期的依變數數列值，但是會將預測週期的自變數值估計在內。如果有更多目前依變數數列的值，而且要將這些值納入預測中，就必須重新估計、調整估計週期，以便將這些值包含在內。

**從資料中重新估計。** 模式參數是使用作用中資料集的資料重新估計。模式參數重新估計並不影響模式結構。舉例來說，ARIMA(1, 0, 1) 模式會維持原結構，但自身迴歸和移動平均參數會進行重新估計。重新估計不會造成偵測新偏離值。如果有任何偏離值，一定都是取自模式檔案。

- **估計期間。** 估計週期定義用來重新估計模式參數的觀察值集。依照預設值，估計期間包括作用中資料集內所有觀察值。若要設定估計期間，請選擇「」「選擇觀察值」對話方塊中的「**以時間或觀察值範圍為準**」。依可用的資料而定，程序所使用的估計週期可能會因模式而異，因此與所顯示的值不同。在一個模式中，真正的估計週期是從模式的依變數除去任何發生於指定估計週期，其開始或結束時的連續遺漏值之後所餘下的期間。

### 預測週期

各模式的預測週期一定都是從估計週期結束後第一個觀察值開始，然後一直執行到作用中資料集的最後一個觀察值或是使用者指定的日期為止。如果參數未經過重新估計（這是預設值），則各模式的估計週期就是開發（或最後更新）模式時所用的觀察值集。

- **估計期間結束後的第一個觀察值一直到作用中資料集的最後一個觀察值。** 若估計期間結束早於作用中資料集的最後一個觀察值，而您要一直預測到最後一個觀察值，請選擇此選項。
- **估計期間結束之後第一個觀察值一直到指定日期。** 選擇此選項，明確指定預測期間結束時間。輸入「日期」網格中所有儲存格的數值。

如果沒有為作用中資料集定義日期規格，「日期」網格只顯示一欄「觀察」。若要指定預測期間的結束時間，請輸入相關觀察值的列編號（顯示於「資料編輯程式」中）。

「資料」網格中的「循環」欄（若有）是指向作用中資料集內 CYCLE\_ 變數的值。

## 輸出

可用的輸出包括各個模式的結果以及所有模式的結果。各個模式的結果可以根據使用者指定的條件，限制為最適或最不適應模式集。

## 統計量和預測表

圖表 3-2  
「套用時間數列模式」，「統計量」索引標籤



「統計量」索引標籤提供選項可顯示模式適合度統計表、模式參數、自身相關函數和預測。除非模式參數經過重新估計（「模式」索引標籤上的「從資料重新估計」），所顯示的適合度量數值、Ljung-Box 值，以及模式參數都是取自模式檔，並反映各模式在開發（或最後更新）時所用的資料。偏離值資訊一定都是取自模式檔案。

**顯示適合度量數、Ljung-Box 統計量，以及各模式的偏離值數目。** 選擇（核取）此選項，以顯示包含所選定適合度量數、Ljung-Box 值和各模式偏離值數的表格。

**適合量數。** 您可以選擇下列一個或多個納入包含各模式適合度的表格中：

- 平穩 R 平方
- R 平方
- 根平均平方誤
- 平均絕對百分比誤差
- 平均絕對誤差
- 最大絕對百分比誤差
- 最大絕對誤差
- 正規化 BIC

如需詳細資訊，請參閱第 88 頁附錄 A 中的適合度量數。

**比較模式的統計量。** 這組選項控制包含所有模式統計量表格的顯示。各選項會分別產生一個表格。您可以選取下列一個或多個選項：

- **適合度。** 摘要統計量和平穩 R 平方的百分位數、R 平方、根平均平方誤、平均絕對百分比誤差、平均絕對誤差、最大絕對百分比誤差，最大絕對誤差，以及常態化 Bayesian 資訊準則表。
- **殘差自我相關函數 (ACF)。** 摘要統計量和所有已估計模式殘差的自身相關百分位數表。此表只能用於模式參數已重新估計時（「模式」索引標籤上的「從資料重新估計」）。
- **殘差偏自我相關函數 (PACF)。** 摘要統計量和所有已估計模式殘差的偏自身相關百分位數表。此表只能用於模式參數已重新估計時（「模式」索引標籤上的「從資料重新估計」）。

**各個模式的統計量。** 這組選項控制包含各模式詳細資料表格的顯示。各選項會分別產生一個表格。您可以選取下列一個或多個選項：

- **參數估計值。** 顯示各模式的參數估計表。指數平滑化和 ARIMA 模式會分別顯示不同的表格。如果偏離值存在，其參數估計也會另外顯示於一個表格中。
- **殘差自我相關函數 (ACF)。** 依落差顯示各個已估計模式的殘差自身相關表。該表中包含自身相關的信賴區間。此表只能用於模式參數已重新估計時（「模式」索引標籤上的「從資料重新估計」）。
- **殘差偏自我相關函數 (PACF)。** 依落差顯示各個已估計模式的殘差偏自身相關表。該表中包含偏自身相關的信賴區間。此表只能用於模式參數已重新估計時（「模式」索引標籤上的「從資料重新估計」）。

**顯示預測。** 顯示各模式的模式預測和信賴區間表。

## 圖形

圖表 3-3  
「套用時間數列模式」，「圖形」索引標籤



「圖形」索引標籤提供選項，可顯示模式適合度統計量的圖形、自身相關函數，以及數列值（包括預測）。

### 比較模式的圖形

這組選項控制包含所有模式統計量圖形的顯示。除非模式參數經過重新估計（「模式」索引標籤上的「從資料重新估計」），所顯示的值都是取自模式檔案，並反映各模式在開發（或最後更新）時所用的資料。此外，自身相關圖形只能用於模式參數已重新估計時。各選項會分別產生一個圖形。您可以選取下列一個或多個選項：

- 平穩 R 平方
- R 平方
- 根平均平方誤
- 平均絕對百分比誤差
- 平均絕對誤差
- 最大絕對百分比誤差
- 最大絕對誤差
- 正規化 BIC

- 殘差自我相關函數 (ACF)
- 殘差偏自我相關函數 (PACF)

如需詳細資訊，請參閱第 88 頁附錄 A 中的適合度量數。

### 各個模式的圖形

**數列。** 選取(核取) 此選項，可取得各模式預測值的圖形。觀察值、適合度值、適合度值的信賴區間，以及自身相關只能用於模式參數已重新估計時（「模式」索引標籤上的「從資料重新估計」）。您可以選擇下列一個或多個選項，以包含在圖形之中：

- **觀察值。** 依變數列的觀察值。
- **預測。** 預測期間的模式預測值。
- **適合度值。** 估計期間的模式預測值。
- **預測的信賴區間。** 預測期間的信賴區間。
- **適合度值的信賴區間。** 估計期間的信賴區間。

**殘差自我相關函數 (ACF)。** 顯示各個已估計模式的殘差自身相關圖形。

**殘差偏自我相關函數 (PACF)。** 顯示各個已估計模式的殘差偏自身相關圖形。

## 將輸出限制於最適或最不適模式

圖表 3-4  
「套用時間數列模式」，「輸出過濾器」索引標籤

圖表 3-4 顯示了「套用時間數列模式」對話框中的「輸出過濾器」索引標籤。該對話框包含以下內容：

- 對話框標題為「套用時間數列模式」。
- 標籤頁包括：模式、統計、圖形、輸出過濾器、儲存、選項。
- 主要選項：
  - 在匯出中包含所有模型(I)
  - 根據適合度的過濾器模型(F)
- 顯示區域：
  - 最適模型(B)
    - 固定模型數目(X)
      - 數目(U):
    - 模型總數目百分比(C)
      - 百分比(P):
  - 最不適模型(P)
    - 固定模型數目(X)
      - 數目(U):
    - 模型總數目百分比(C)
      - 百分比(P):
- 適合度量度(O): 平穩形 R 平方
- 底部按鈕：確定、貼上之後(P)、重設(R)、取消、輔助說明

「輸出過濾器」索引標籤提供選項，可將表格和圖表輸出限制於模式子集。您可以選擇根據您所提供的適合度條件，將輸出限制於最適及/或最不適模式。按照預設，所有模式都會包含在輸出之中。除非模式參數經過重新估計（「模式」索引標籤上的「從資料重新估計」），用來篩選模式的適合度量數值都是取自模式檔案，並反映各模式在開發（或最後更新）時所用的資料。

**最適模式。**選擇（核取）此選項，可將最適模式包含在輸出中。選擇適合度量數，並指定要包含的模式數目。選擇此選項並不會阻止同時選取最不適模式。在此情況下，輸出將包含最不適模式及最適模式。

- **固定的模式數目。**指定結果顯示  $n$  個最適模式。如果數目超過模式的總數，則顯示所有模式。
- **模式總數的百分比。**指定結果顯示所有模式中最高  $n$  百分比之適合度值的模式。

**最不適模式。**選擇（核取）此選項，可將最不適模式包含在輸出中。選擇適合度量數，並指定要包含的模式數目。選擇此選項並不會阻止同時選取最適模式。在此情況下，輸出將包含最適模式及最不適模式。

- **固定的模式數目。**指定結果顯示  $n$  個最不適模式。如果數目超過模式的總數，則顯示所有模式。
- **模式總數的百分比。**指定結果顯示所有模式中最低  $n$  百分比之適合度值的模式。

**適合度量數。**選擇要用來篩選模式的適合度量數。預設值是平穩  $R$  平方。



## 儲存模式預測與模式規格

圖表 3-5  
「套用時間數列模式」，「儲存」索引標籤



「儲存」索引標籤可以讓您將模式預測儲存為作用中資料集內的新變數，並將模式規格儲存為 XML 格式的外部檔案。

**儲存變數。**您可以將模式預測、信賴區間及殘差儲存為作用中資料集的新變數。各模式會產生其自身的新變數集。如果預測週期延伸至與模式相關聯的依變數之數列長度之外，則加入新的觀察值。除非模式參數經過重新估計（「模式」索引標籤上的「從資料重新估計」），否則預測值和信賴區間限制只為預測週期特別建立。為每一個新變數選擇相關聯的「儲存」核取方塊，以選擇儲存新變數。依照預設值，不會儲存任何新變數。

- **預測值。** 模式預測值。
- **信賴區間下限。** 預測值的信賴區間下限。
- **信賴區間上限。** 預測值的信賴區間上限。
- **噪音殘差。** 模式殘差。執行依變數轉換時（例如，自然對數轉換），這些都是轉換數列的殘差。此選擇只能用於模式參數已重新估計時（「模式」索引標籤上的「從資料重新估計」）。
- **變數名稱字首。** 指定要用來為新變數命名時的字首，或者保留預設字首。變數名稱是由字首、相關依變數的名稱，以及模式識別碼組成。必要時變數名稱會延伸，以避免與變數命名規則相衝突。字首必須與有效的變數名稱規則相符。

**匯出包含已重新估計參數的模式檔案。** 模式規格（其中包含重新估計的參數和適合度統計量）是匯出至指定的 XML 格式檔案。此選項只能用於模式參數已重新估計時（「模式」索引標籤上的「從資料重新估計」）。

## 選項

圖表 3-6  
「套用時間數列模式」，「選項」索引標籤



「選項」索引標籤可以讓您指定遺漏值的處理、設定信賴區間寬度，並設定供自身相關顯示之落差數目。

**使用者遺漏值。** 這些選項可控制使用者遺漏值之處理。

- **視為無效。** 使用者遺漏值會被視為系統遺漏值來處理。
- **視為有效。** 使用者遺漏值被視為有效資料。

**遺漏值原則。** 下列規則適用於遺漏值之處理（包括系統遺漏值與視為無效的使用者遺漏值）：

- 模式中包含具有發生於估計期間內之依變數遺漏值的觀察值。遺漏值的處理方式依估計方法而定。
- 對於 ARIMA 模式，如果預測值在估計週期內有任何遺漏值，就發出警告。任何涉及預測值的模式都不重新估計。
- 如果任何自變數在預測期間內有遺漏值，則程序發出警告，並盡可能預測到最遠的範圍。

**信賴區間寬度 (%)。**信賴區間是供模式預測和殘差自身相關計算使用。您可以指定任何小於 100 的正數值。依照預設值，使用的是 95% 信賴區間。

**顯示於 ACF 與 PACF 輸出中的最大落差數目。**您可以設定顯示於自身相關與偏自身相關表格和圖形中的最大落差數目。此選項只能用於模式參數已重新估計時（「模式」索引標籤上的「從資料重新估計」）。

## TSAPPLY 指令的其他功能

如果將選取範圍貼入語法視窗，並編輯所得到的 TSAPPLY 指令語法，則可使用其他功能。指令語法語言讓您可以：

- 指定模式檔中唯一的模式子集要套用於作用中資料集（利用 MODEL 次指令上的 DROP 和 KEEP 關鍵字）。
- 從兩個或以上的模式檔套用模式至資料（利用 MODEL 次指令）。例如，一個模式檔可能包含了代表單位銷售量數列的模式，而另一個可能包含代表收益數列的模式。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

# 週期性分解

「週期性分解」程序會將數列分解為週期性成份，結合趨勢與循環的成份，和「誤差」成份。此程序是「普查方法 I」的實作，也稱為移動平均比法。

**範例。** 某科學家希望分析特定氣象台的每月臭氧量。目的是要判斷數據中是否有趨勢。為了瞭解實際趨勢，科學家首先必須根據季節性效應計算讀數變異。「週期性分解」程序可用來移除任何系統化的週期變異性。接著會在依週期性調整的數列上進行趨勢分析。

**統計量。** 週期性因子集合。

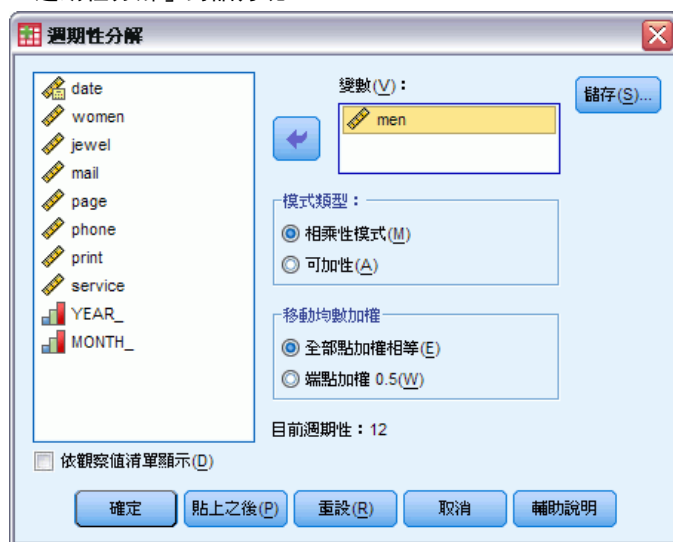
**資料。** 必須為數字變數。

**假設。** 變數不應包含有內嵌遺漏資料。必須至少定義一個週期性日期成份。

## 估計週期性因子

- ▶ 從功能表選擇：  
分析 > Forecasting > 週期性分解...

圖表 4-1  
「週期性分解」對話方塊



- ▶ 在可用的清單中，選擇一個或多個變數，然後將它們移到「變數」清單中。注意，清單只包含數值變數。

**模式類型。** 「週期性分解」程序提供兩種不同方法模式化週期性因子：相乘性或可加性。

- **相乘性。** 週期性成份是一個因子，週期性調整數列乘以此因子便可以得到原始數列。就效果而言，週期性成份對評估和數列整體等級成等比例。無週期變異的觀察值，其週期性成份為 1。
- **可加性。** 會將週期性調整新增至週期性調整的數列中，以取得觀察值。此項調整會試圖從數列中移除週期效應，以便觀察其他遭週期性成份「遮蔽」的相關特性。就效果而言，週期性成份不以數列整體等級為依據。無週期變異的觀察值，其週期性成份為 0。

**移動平均加權。**「移動平均加權」選項允許您指定在計算移動平均時要如何處理數列。這些選項只有在數列週期是均等時才提供使用。如果週期是不固定的，則所有點都有相同的加權。

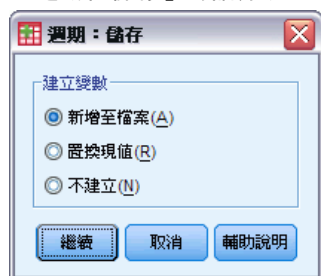
- **所有相等點。** 使用與週期相同的指距，以及所有加權相同的點計算而得的移動平均數。若週期不固定，則會一律採用此方法。
- **使用 .5 加權後的端點。** 使用指距（其等於週期加上 1）和以 0.5 加權後的指距端點，所計算出的週期均等數列移動平均數。

您可以：

- 按一下「儲存」指定如何儲存新變數。

## 週期性分解儲存

圖表 4-2  
「週期性儲存」對話方塊



**建立變數。** 可讓您選擇處理新變數的方式。

- **新增至檔案。** 會將「週期性分解」建立的新數列儲存為作用中資料集的一般變數。變數名稱由三個字母的字首、底線和數字組成。
- **置換現有。** 會將「週期性分解」建立的新數列儲存為作用中資料集的暫存變數。並同時剔除由「預測」程序建立的任何現有暫存變數。變數名稱由三個字母的字首、井字符號 (#) 和數字組成。
- **不要建立。** 不會將新數列新增至作用中資料集。

### 新變數名稱

「週期性分解」程序會為每個指定數列建立四個新變數（數列），名稱分別以下列三個字母為字首：

- SAF。** 週期性調整因子。這些變數會在數列層級上指定各期間的效果。
- SAS。** 週期性調整數列。這些是移除數列的週期變異後獲得的數值。

**STC**。平滑趨勢週期成份。這些數值顯示數列中所表現的趨勢與循環行為。

**ERR**。殘差或「誤差」值。週期性、趨勢與循環成份所遺留的數值已自數列中移除。

## SEASON 指令的其他功能

指令語法語言也可以讓您：

- 以「SEASON」指令指定任何週期性，而不從「定義日期」程序提供的替代方案中選擇。

如需完整的語法資訊，請參閱《指令語法參考手冊》。

# 光譜圖

「光譜圖」程序用於辨識時間數列中的週期性行為。它會以整體來分析數列的變異到不同頻率的週期成份，而不是分析從一時間點到下一個時間點的變異。平滑數列在低頻率有較強的週期性元件；隨機變異（「白噪音」）將成份強度擴展到所有頻率。

含有遺漏資料的數列不能以這個程序分析。

**範例。** 新房屋建造速率是經濟狀態的重要指標。開始建造日期通常為有力的週期性成份。但資料中是否存在較長循環，分析師在評估目前圖表時需要注意？

**統計量。** 各頻率或週期成份的正弦與餘弦轉換、週期值以及光譜密度估計。選擇雙變數分析時：各頻率或週期成份的交叉週期值中實數與虛數部分、共同光譜密度、弦光譜、增益、平方一致性與相光譜。

**圖形。** 對於單變數和雙變數分析：週期值與光譜密度。對於雙變數分析：平方一致性、弦光譜、交叉共變量、共同光譜密度、相光譜與增益。

**資料。** 必須為數字變數。

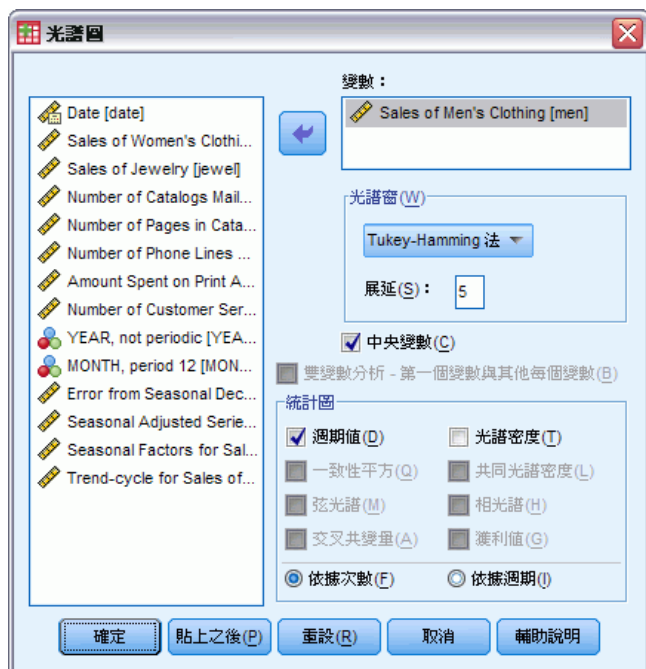
**假設。** 變數不應包含有內嵌遺漏資料。要分析的時間數列應為平穩型，且應從數列中減去所有非零平均數。

- **平穩型。** 時間序列必須符合您配適 ARIMA 模式的條件。純 MA 序列會是恆定的，但是 AR 及 ARMA 序列則不一定。經過一段時間，恆定數列的平均數和變異數都會是常數。

## 取得光譜分析

- ▶ 從功能表選擇：  
分析 > 時間數列 > 光譜分析...

圖表 5-1  
「光譜圖」對話方塊



- ▶ 在可用的清單中，選擇一個或多個變數，然後將它們移到「變數」清單中。注意，清單只包含數值變數。
- ▶ 選擇「光譜視窗」選項之一以選擇如何將週期值平滑化，以得到光譜密度估計。可用的平滑化選項有 Tukey-Hamming、Tukey、Parzen、Bartlett、Daniell (Unit) 和無。
  - **Tukey-Hamming**。加權為  $W_k = .54D_p(2 \pi f_k) + .23D_p(2 \pi f_k + \pi/p) + .23D_p(2 \pi f_k - \pi/p)$ , ( $k = 0, \dots, p$ ), 其中  $p$  為半數指距的整數部分,  $D_p$  為 Dirichlet 階數  $p$  的核心。
  - **Tukey**。加權為  $W_k = 0.5D_p(2 \pi f_k) + 0.25D_p(2 \pi f_k + \pi/p) + 0.25D_p(2 \pi f_k - \pi/p)$ , ( $k = 0, \dots, p$ ), 其中  $p$  為半數指距的整數部分,  $D_p$  為 Dirichlet 階數  $p$  的核心。
  - **Parzen**。加權值為  $W_k = 1/p(2 + \cos(2 \pi f_k)) (F_{[p/2]}(2 \pi f_k))^{*2}$ , ( $k = 0, \dots, p$ ), 其中  $p$  為半數指距的整數部分,  $F_{[p/2]}$  為 Fejer 階數  $p/2$  的核心。
  - **Bartlett**。一種光譜視窗類型, 會以  $W_k = F_p(2\pi f_k)$ , for  $k = 0, \dots, p$  來計算位於視窗上半部的加權, 其中  $p$  為半數指距的整數部分, 而  $F_p$  為  $p$  階數的 Fejer 核心。視窗下半部會與視窗上半部對稱。
  - **Daniell (Unit)**。加權皆等於 1 的光譜視窗類型。
  - **無**。不進行平滑化。如果您選擇了這個選項, 您得到的光譜密度估計就會和原本的週期值一樣。

**指距**。全部納入平滑化的連續數值範圍。指距一般使用奇數整數。較大的指距對光譜密度圖的平滑程度比較小的指距來得大

**中心變數**。在計算光譜並移除可能與數列平均值相關聯之大型項目前, 調整數列以使其平均數為 0。



**雙變數分析 — 第一個變數與其他每個變數。** 若您選擇了二個以上的變數，您可選擇此選項來要求雙變數光譜分析。

- 在「變數」清單中的第一個變數會以自變數來處理，而其他的變數則以依變數來處理。
- 在第一個數列之後的每個數列都會與其他已命名數列無關的以第一個數列做分析。每個數列的單變數分析亦會執行。

**圖形。** 單變數與雙變數分析都可使用週期值與光譜密度。只有雙變數分析可使用其他所有選項。

- **週期值。** 比對頻率或週期的不平滑光譜振幅圖（以對數尺度繪製）。低頻率變異描述平滑數列的特徵。平均擴散在所有頻率的變異表示「白噪音」。
- **平方一致性。** 兩個數列增加的結果。
- **弦光譜。** 交叉週期值的虛數部分，其會測量兩個時間數列的階段外頻率成份相關性。該成份超出階段外  $\pi/2$  的弧度。
- **交叉共變量。** 共同光譜密度與弦光譜平方和的平方根。
- **光譜密度。** 經過平滑化，以移除不規則變化的週期值。
- **共同光譜密度。** 交叉週期值的實數部分，其會測量兩時間數列的同步頻率成份相關性。
- **相光譜。** 一數列的各頻率成分領先或落後另一成分的程度量數。
- **增益。** 將交叉共變量除以其中一個數列之光譜密度所得出的商。兩數列皆擁有專屬的增益數值。

**依次數。** 依據次數來製作所有圖表，範圍從次數 0（常數或平均數項目）到次數 0.5（兩個觀察值週期項目）。

**依區間。** 依據區間來製作所有圖表，範圍從 2（兩個觀察值週期項目）到與觀察值個數相等的區間（常數或平均數項目）。在對數尺度中會顯示區間。

## SPECTRA 指令的其他功能

指令語法語言讓您也可以：

- 儲存計算好的光譜分析變數到現用資料集中，提供日後使用。
- 為光譜視窗指定自訂的加權。
- 以頻率和週期來產生圖。
- 列印圖中每一個顯示值的完整清單。

若需完整的語法資訊，請參閱《指令語法參考手冊》。

# 部 11: 範 例

# 以 Expert Modeler 進行大量預測

全國性寬頻服務提供業者的分析師必須提出用戶訂購量的預測，以預測頻寬的使用量。分析師必須針對組成全國用戶群的 85 個地方市場做出預測。每月的歷史資料收集於 broadband\_1.sav 中。如需詳細資訊，請參閱附錄 D 中的範例檔案中的 IBM SPSS Forecasting 20。

在這個範例中，您將使用 Expert Modeler 來為 85 個本地市場建立未來三個月的預測值，並將產生的模式存入外部 XML 檔案中。完成後，您可能會想要實際操作下一個範例，藉由套用已儲存的模式進行大量重新預測在 第 7 章第 52 頁內，該範例會套用已儲存的模式到已更新的資料集，以便在不需重新建立模式的情況下，再延長三個月的預測值。

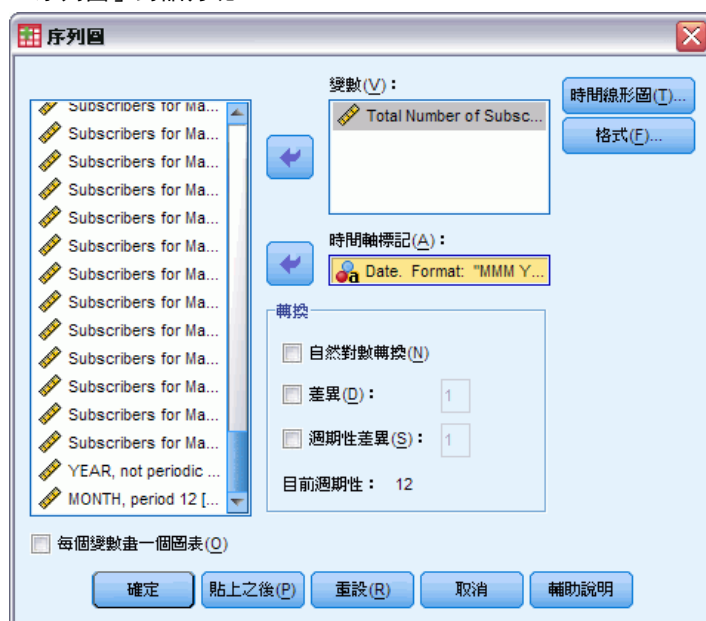
## 檢驗您的資料

建立模式前，最好先了解您資料的性質。資料是否顯示週期性變異？雖然 Expert Modeler 將自動找出每個數列的最佳週期性或非週期性模式，但是當資料內不含週期性時，您可將搜尋限制在非週期性模式，以更快取得結果。我們可繪製每個市場的用戶總數，就可了解大概的情況，而不需檢驗 85 個本地市場中每個市場的資料。

- ▶ 從功能表選擇：

分析(A) > 預測 > 序列圖...

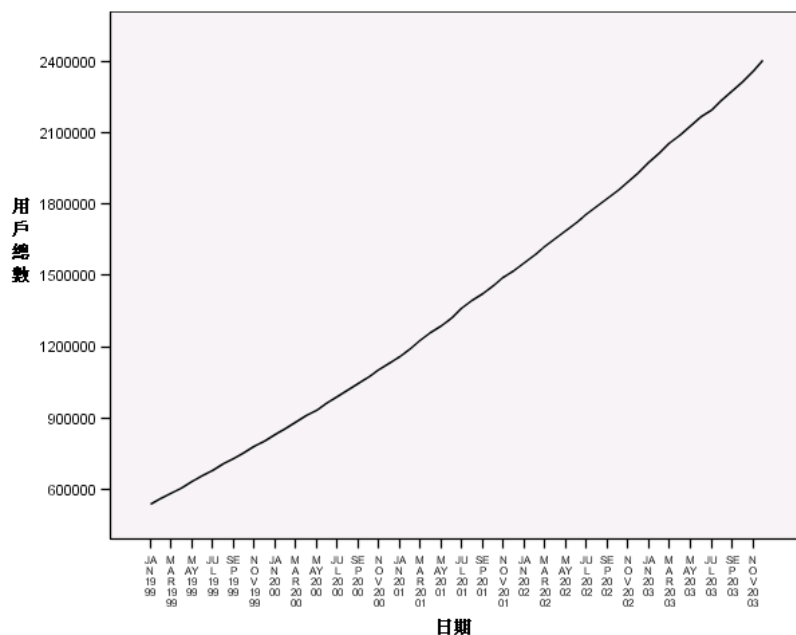
圖表 6-1  
「序列圖」對話方塊



- ▶ 選擇「用戶總數」並將其移到「變數」清單中。

- ▶ 選擇「日期」並將它移到「時間軸標記」方塊中。
- ▶ 按一下「確定」。

圖表 6-2  
所有市場內寬頻用戶的總數



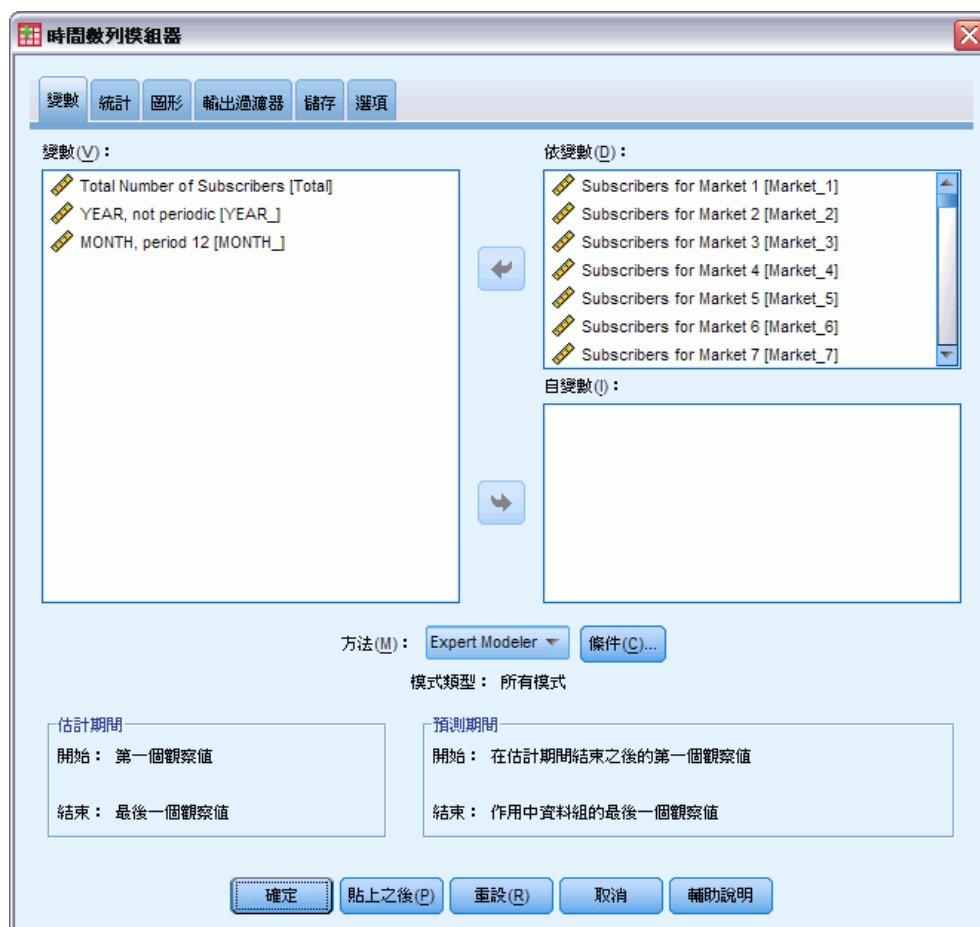
數列展示了非常平滑的向上趨勢，且未顯示任何週期性變異。個別數列可能有週期性，但週期性一般而言似乎不是資料的顯著特色。當然，你也應該在排除週期性模式前，檢閱每個數列。然後您可以分離出具備週期性的數列並分別建立他們的模式。在目前的狀況中，檢閱 85 個數列後發現並沒有任何數列具備週期性。

## 執行分析

若要使用 Expert Modeler:

- ▶ 從功能表選擇：  
分析(A) > 預測 > 建立模式(C)...

圖表 6-3  
「Time Series Modeler」對話方塊



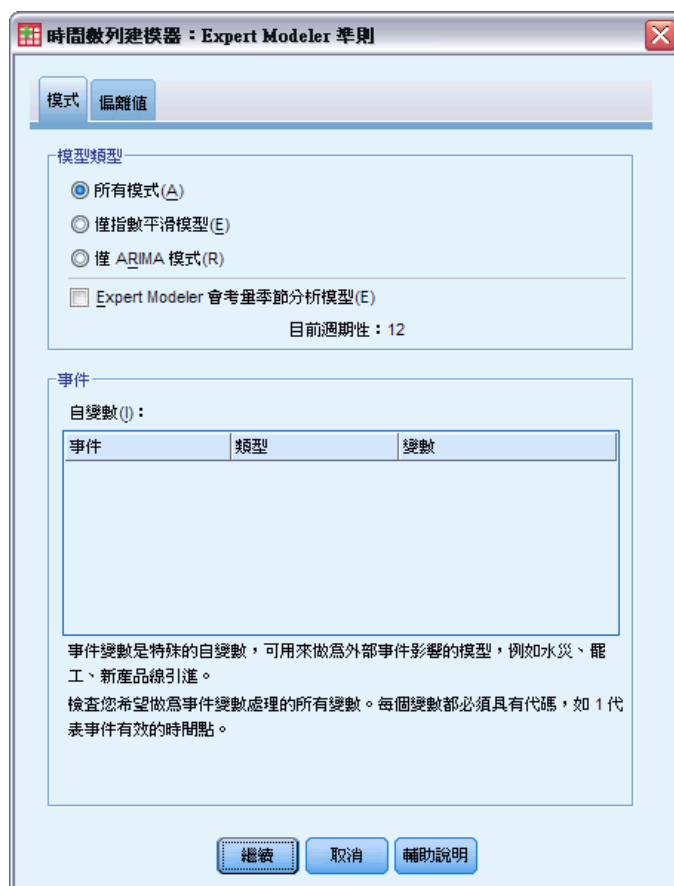
- ▶ 選擇「市場 1 的用戶」到「市場 85 的用戶」作為依變數。
- ▶ 請確認已選擇「方法」下拉式清單上的「Expert Modeler」。Expert Modeler 將自動為每個依變數列找到最適合的模式。

用於估計模式的觀察值集合稱為**估計週期**。根據預設值，這個集合包括作用中資料集內的所有觀察值。您可以選擇「選擇觀察值」對話方塊中的「根據時間或觀察值範圍」來設定估計週期。至於這個範例，我們將沿用預設值。

另外也請注意預設預測週期於估計週期結束後開始，並持續到作用中資料集的最後一個觀察值。如果您正在預測最後一個觀察值之後的數值，您需要延長預測週期。您稍後在本範例中將看到可在「選項」索引標籤中這麼做。

- ▶ 按一下「條件」。

圖表 6-4  
「Expert Modeler 條件」對話方塊，「模式」索引標籤



- ▶ 取消選擇「模式類型」組別中的「Expert Modeler 考量週期性模式」。  
雖然資料是以月為單位且目前的週期是 12，但是我們已經發現資料並不具任何週期性，所以不需要考量週期模式。如此一來會降低由 Expert Modeler 搜尋的模式空間，並可顯著降低計算時間。
- ▶ 按一下「繼續」。
- ▶ 按一下「Time Series Modeler」對話方塊中的「選項」索引標籤。

圖表 6-5  
「時間數列 Modeler」，「選項」索引標籤

時間數列模組器

變數 統計 圖形 輸出過濾器 儲存 選項

預測期間

透過現用資料組最後觀察值的估計期間結束之後的第一個觀察值 (F)  
 透過指定日期的估計期間結束之後的第一個觀察值 (C)

日期 (D):

年:	月:
2004	3

使用者遺漏值

視為無效 (I)  
 視為有效 (V)

信賴區間寬度 (%) (W): 95

輸出中的模型識別符號前置符號 (P): 模式

ACF 和 PACF 輸出中顯示的落差最大數目 (X): 24

確定 貼上之後 (P) 重設 (R) 取消 輔助說明

► 在「預測週期」組別中，選擇「估計週期結束後的第一個觀察值到指定日期」。

► 在「日期」網格中，輸入年份 2004 及月份 3。

資料集包含自 1999 年 1 月至 2003 年 12 月的資料。依據目前設定，預測週期將為 2004 年 1 月至 2004 年 3 月。

► 按一下「儲存」索引標籤。

圖表 6-6  
「時間數列 Modeler」，「儲存」索引標籤



- ▶ 選擇（勾選）「儲存」列中的「預測值」項目，並以預設值「Predicted」作為「變數名稱字首」。  
模式預測值會以「Predicted」作為變數名稱，儲存為作用中資料集的新變數。您也可以將每個模式的規格存到外部 XML 檔案中。如此一來，您將可在獲得新資料時，重複使用模式以延長您的預測。
- ▶ 按一下「儲存」索引標籤上的「瀏覽」按鈕。  
您將進入儲存檔案的標準對話方塊。
- ▶ 請瀏覽到您想要儲存 XML 模式檔的資料夾，輸入檔案名稱，並按一下「儲存」。
- ▶ 按一下「統計量」索引標籤。



圖表 6-7  
「時間數列 Modeler」，「統計量」索引標籤



- ▶ 選擇「顯示預測」。

這個選項會為每個依變數列建立預測數值表格，並提供另一個選項 — 除了將預測值儲存為新變數之外 — 以取得這些數值。

「適合度」的預設選擇（在「比較模式的統計量」組別中）以所有模式內所計算的 — 如 R-平方、平均絕對百分比誤差，及常態化 BIC — 等適合度統計量建立表格。其中提供模式與資料適合度的簡明摘要。

- ▶ 按一下「圖形」索引標籤。

圖表 6-8  
「時間數列 Modeler」，「圖形」索引標籤



- ▶ 取消選擇「個別模式圖形」組別中的「數列」。

這會抑制每個模式產生數列圖形。在這個範例中，我們比較有興趣的是將預測值存為新變數，而不是產生預測值的圖形。

「比較模式的圖形」組別提供好幾個圖形（以直方圖形式），說明所有模式內計算的適合度統計量。

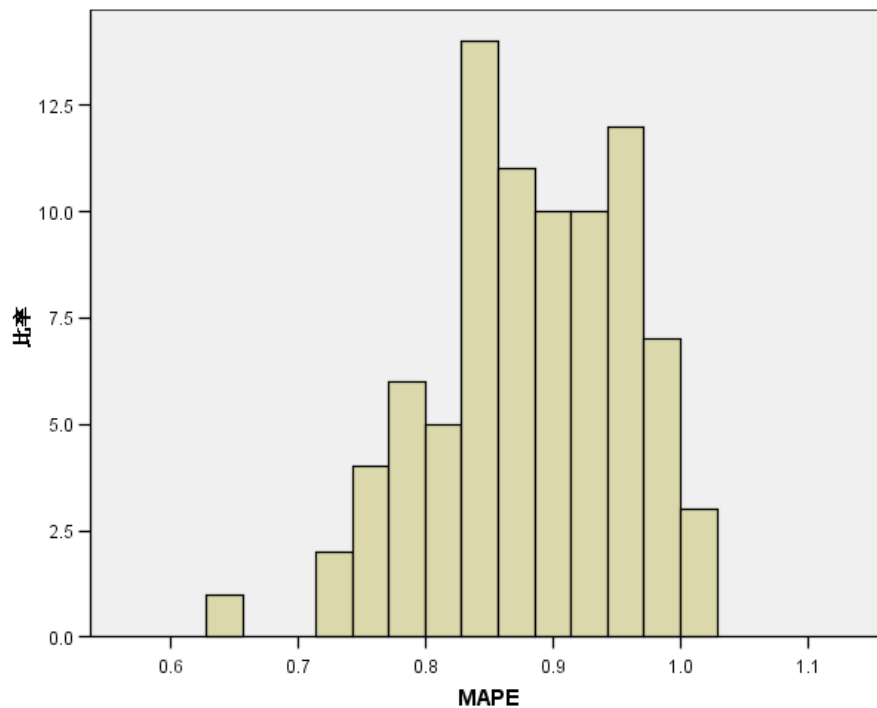
- ▶ 在「比較模式的圖形」組別中選擇「平均絕對百分比誤差」及「最大絕對百分比誤差」。

絕對百分比誤差是依變數列與其模式預測水準差異程度的量數。您可透過檢驗所有模式內的平均數值及最大數值，了解您預測值內的不確定性。而且，由於依變數列代表不同大小市場的用戶數量，因此建議您檢視百分比誤差的摘要圖形，而不是絕對誤差的摘要圖形。

- ▶ 在「Time Series Modeler」對話方塊中，按一下「確定」。

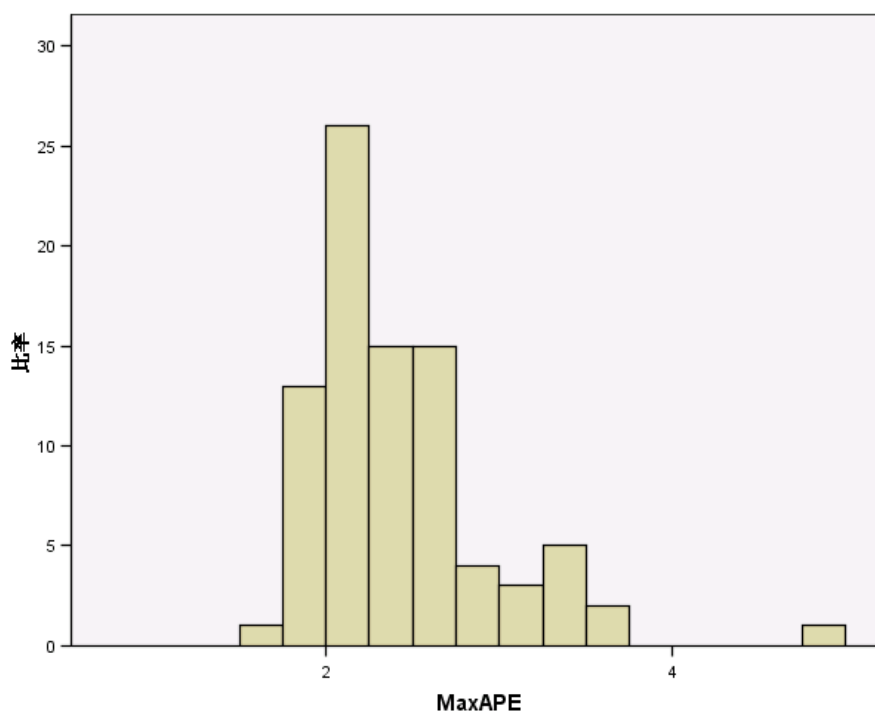
## 模式摘要圖表

圖表 6-9  
平均絕對百分比誤差的直方圖



這個直方圖顯示所有模式內的平均絕對百分比誤差 (MAPE)。並顯示所有模式顯示的平均不確定性大約是 1%。

圖表 6-10  
最大絕對百分比誤差的直方圖



這個直方圖顯示所有模式內的最大絕對百分比誤差 (MaxAPE)，而且對於想像您預測值的最糟情況是很有用的。並顯示每個模式的最大百分比誤差落在 1% 到 5% 的範圍內。這些數值是否代表不確定性的數量令人不可接受？由於可接受的風險在每個問題中都不一樣，這時您的商業智慧便可派上用場。

## 模式預測值

圖表 6-11  
包含模式預測值的新變數

YEAR_	MONTH_	DATE_	Predicted_Market_1_Model_1	Predicted_Market_2_Model_2
2003	10	OCT 2003	11820	51084
2003	11	NOV 2003	11857	51273
2003	12	DEC 2003	11687	53082
2004	1	JAN 2004	11503	54893
2004	2	FEB 2004	11447	55856
2004	3	MAR 2004	11390	56704

「資料編輯器」顯示包含模式預測值的新變數。雖然此處只顯示兩個新變數，但總共有 85 個新變數，每個依變數列都有一個新變數。變數名稱會包括預設字首「Predicted」，後面接上相關的依變數名稱（例如 Market\_1），後面接上模式識別碼（例如 Model\_1）。

三個新觀察值，包括 2004 年 1 月到 2004 年 3 月的預測值，已經與自動產生的資料標記一起新增到資料集中。每個新變數都包含估計週期（1999 年 1 月到 2003 年 12 月）的模式預測值，可讓您看到模式與已知數值的適合程度。

圖表 6-12  
預測表

Model		Jan 2004	Feb 2004	Mar 2004
Subscribers for Market 1-模式_1	Forecast	11503	11447	11390
	UCL	11686	11767	11870
	LCL	11321	11126	10910
Subscribers for Market 2-模式_2	Forecast	54893	55856	56704
	UCL	55632	57195	58575
	LCL	54154	54518	54832
Subscribers for Market 3-模式_3	Forecast	59656	59305	58954
	UCL	60457	60753	61158
	LCL	58856	57857	56750
Subscribers for Market 4-模式_4	Forecast	18235	18424	18628
	UCL	18413	18731	19121
	LCL	18058	18116	18136

您也曾選擇以預測值建立表格。表格包含預測週期內的預測值，但一不同於包含模式預測值的新變數一，表格並不包含估計週期內的預測值。結果會依模式分組並以模式名稱來識別，模式名稱包括相關依變數的名稱（或標記），後面接上模式識別碼——就像是包含模式預測值的新變數名稱。表格也包括預測值（預設值為 95%）的信賴上限（UCL）及信賴下限（LCL）。

您現在已經看過取得預測值的兩種方法：將預測值儲存為作用中資料集內的新變數以及建立預測表。無論使用哪一種方法，您在輸出預測值時都會有幾個可用的選項（例如，輸出成 Excel 試算表）。

## 摘要

您現在已經學會使用 Expert Modeler 來為多重數列建立預測值，且您已經將結果模式存到外部 XML 檔案中。在下個範例中，您將學到如何在獲得新資料時，利用「套用時間數列模式」程序，一不需重新建立模式一，來延長您的預測。

# 藉由套用已儲存的模式進行大量重新預測

您已經使用 Time Series Modeler 來為您的時間數列資料建立模式，並根據可用的資料建立初始預測。您計劃在獲得更多當前的資料時，重複使用這些模式以延長您的預測，所以您已經將模式儲存為外部檔案。您現在已經準備好套用已儲存的模式。

這個範例是上一個範例的自然延伸，以 [Expert Modeler 進行大量預測](#) 在 [第 6 章第 41 頁](#)，但也可以獨立使用。在這個情境中，您是全國性寬頻服務提供業者的分析師，每個月都必須預測 85 個本地市場的用戶訂購量。您已經使用了 Expert Modeler 來建立模式及未來三個月的預測。您的資料倉儲已經以原始預測週期的實際資料重新整理，所以您想要使用那些資料將預測值範圍再延長三個月。

已更新的每月歷史資料收集於 broadband\_2.sav 中，已儲存的模式則收集於 broadband\_models.xml 中。[如需詳細資訊，請參閱附錄 D 中的範例檔案中的 IBM SPSS Forecasting 20](#)。當然，如果您曾實際操作上一個範例並儲存您自己的模式檔案，您可以使用自己的模式檔案，而不需使用 broadband\_models.xml。

## 執行分析

若要套用模式：

- ▶ 從功能表選擇：  
分析(A) > 預測 > 套用模式(A)...

圖表 7-1  
「套用時間數列模式」對話方塊



- ▶ 按一下「瀏覽」，然後瀏覽至 broadband\_models.xml 並加以選取（或在上個範例儲存的檔案中選擇您自己的模式檔案）。如需詳細資訊，請參閱附錄 D 中的範例檔案中的 IBM SPSS Forecasting 20。

- ▶ 選擇「從資料中重新估計」。

若要將您的時間數列新數值合併到預測值中，「套用時間數列模式」程序必須重新估計模式參數。但是模式的結構維持不變，所以重新估計的計算時間比建立模式的原始計算時間快很多。

用於重新估計的觀察值集合需要使用新資料。如果您使用「第一觀察值」到「最後觀察值」的預設估計週期，便可確保這一點。如果您需要設定預設值之外的估計週期，您可以在「選擇觀察值」對話方塊中選擇「根據時間或觀察值範圍」。

- ▶ 在「預測週期」組別中，選擇「估計週期結束後的第一個觀察值到指定日期」。

- ▶ 在「日期」網格中，輸入年份 2004 及月份 6。

資料集包含自 1999 年 1 月至 2004 年 3 月的資料。依據目前設定，預測週期將為 2004 年 4 月至 2004 年 6 月。

- ▶ 按一下「儲存」索引標籤。

圖表 7-2  
「套用時間數列模式」，「儲存」索引標籤

儲存變數

變數(V):

說明	儲存	變數名稱前置字元
預測值	<input checked="" type="checkbox"/>	預測值
信賴區間下限	<input type="checkbox"/>	LCL
信賴區間上限	<input type="checkbox"/>	UCL
雜訊殘差	<input type="checkbox"/>	NResidual

對於所選的每個項目，會由依變數儲存一變數。

匯出模型檔案

檔案(F):  瀏覽(B)...

確定 貼上之後(P) 重設(R) 取消 輔助說明

- ▶ 選擇（勾選）「儲存」列中的「預測值」項目並以預設值「Predicted」作為「變數名稱字首」。

模式預測值將以「Predicted」作為變數名稱，儲存為作用中資料集的新變數。

- ▶ 按一下「圖形」索引標籤。



圖表 7-3  
「套用時間數列模式」，「圖形」索引標籤



- ▶ 取消選擇「個別模式圖形」組別中的「數列」。

這會抑制每個模式產生數列圖形。在這個範例中，我們比較有興趣的是將預測值存為新變數，而不是產生預測值的圖形。

- ▶ 在「套用時間數列模式」對話方塊中，按一下「確定」。

## 模式適合度統計量

圖表 7-4  
「模式適合度」表格

適合度統計量	平均數	SE	最小值	最大值	百分位數						
					5	10	25	50	75	90	95
平穩的 R 平方	.181	.141	-4.34E-015	.609	-9.59E-016	1.22E-016	.070	.193	.252	.359	.453
R 平方	.999	.000	.998	1.000	.999	.999	.999	.999	.999	.999	1.000
RMSE	187.766	149.959	45.048	764.698	52.558	59.848	93.155	138.913	205.742	433.883	493.904
MAPE	.886	.074	.669	1.026	.754	.792	.832	.894	.948	.976	.996
MaxAPE	2.446	.512	1.742	4.373	1.840	1.952	2.103	2.321	2.663	3.313	3.676
MAE	146.373	113.749	35.573	612.989	41.209	48.742	73.553	111.569	162.306	331.207	364.582
MaxAE	499.714	434.112	116.966	2143.993	131.261	144.918	226.960	345.221	574.071	1165.207	1587.793
正規化的 BIC	10.086	1.343	7.749	13.412	8.049	8.285	9.193	9.968	10.717	12.210	12.494

「模式適合度」表格提供所有模式內計算出的適合度統計量。其中提供模式及重新估計的參數與資料適合度的簡明摘要。對每個統計量而言，表格提供所有模式內的平均數、標準誤 (SE)、最小值及最大值。其中也包含百分比值，提供各模式內統計量分佈的資

訊。對每個百分位數而言，該百分比的模式擁有的適合度統計量數值低於敘述的數值。例如，95% 的模式擁有的 MaxPE（最大絕對百分比誤差）低於 3.676。

出現幾個統計量的報告時，我們將著重於以下兩者：MAPE（平均絕對百分比誤差）及 MaxAPE（最大絕對百分比誤差）。絕對百分比誤差是依變數列與其模式預測水準差異程度的量數，且這個量數會指出您預測值中的不確定性。所有模式中的平均絕對百分比誤差介於最低值 0.669% 與最高值 1.026% 之間。所有模式中的最大絕對百分比誤差介於 1.742% 與 4.373% 之間。所以每個模式預測值中的平均不確定性為約 1%，而最大不確定性為約 2.5% (MaxAPE 的平均數值)，最糟為約 4%。這些數值是否代表可接受的不確定性數量將取決於您願意接受的風險層級。

## 模式預測值

圖表 7-5  
包含模式預測值的新變數

YEAR_	MONTH_	DATE_	Predicted_Market_1_Model_1	Predicted_Market_2_Model_2
2004	1	JAN 2004	11513	54947
2004	2	FEB 2004	11806	56810
2004	3	MAR 2004	11950	57344
2004	4	APR 2004	12312	59631
2004	5	MAY 2004	12501	60717
2004	6	JUN 2004	12689	61659

「資料編輯器」顯示包含模式預測值的新變數。雖然此處只顯示兩個新變數，但總共有 85 個新變數，每個依變數列都有一個新變數。變數名稱會包括預設字首「Predicted」，後面接上相關的依變數名稱（例如 Market\_1），後面接上模式識別碼（例如 Model\_1）。

三個新觀察值，包括 2004 年 4 月到 2004 年 6 月的預測值，已經與自動產生的資料標記一起新增到資料集中。

## 摘要

您已經學會如何在取得更多當前的資料後，套用已儲存的模式來延長您之前的預測。而且您已經完成套用，而不需重新建立您的模式。當然，如果有理由認為模式已經變更，則您應該利用 Time Series Modeler 程序重新建立模式。

# 使用 Expert Modeler 判斷顯著預測值

某個郵購公司想要開發預測模式，所以收集了男性服飾的每月銷售額資料，以及數個可以用來解釋銷售量中部分差異的數列。可能的預測值包括郵寄的型錄數、型錄頁數、開放訂購的電話線數、花在印刷廣告的金額，以及客服人員數。還有任何預測值對預測有幫助嗎？

在此範例中，您會使用 Expert Modeler 與所有候選預測值，來找出最佳模式。由於 Expert Modeler 只會選擇在統計上與依變數列有顯著相關的預測值，所以您會知道哪一些預測值是有用的，您也會得到可以使用這些預測值進行預測的模式。完成時，您可能會想接著進行下一個範例，[套用已儲存模式來對預測值進行測試](#) 於第 9 章第 67 頁，；使用本例中建構的模式，來調查在不同預測值情境中的銷售效果。

目前範例的資料收集於 catalog\_seasfac.sav 中。[如需詳細資訊，請參閱附錄 D 中的範例檔案中的 IBM SPSS Forecasting 20。](#)

## 繪製您的資料

將您的資料繪製成圖形是絕對錯不了的，尤其是當您只使用一個數列時：

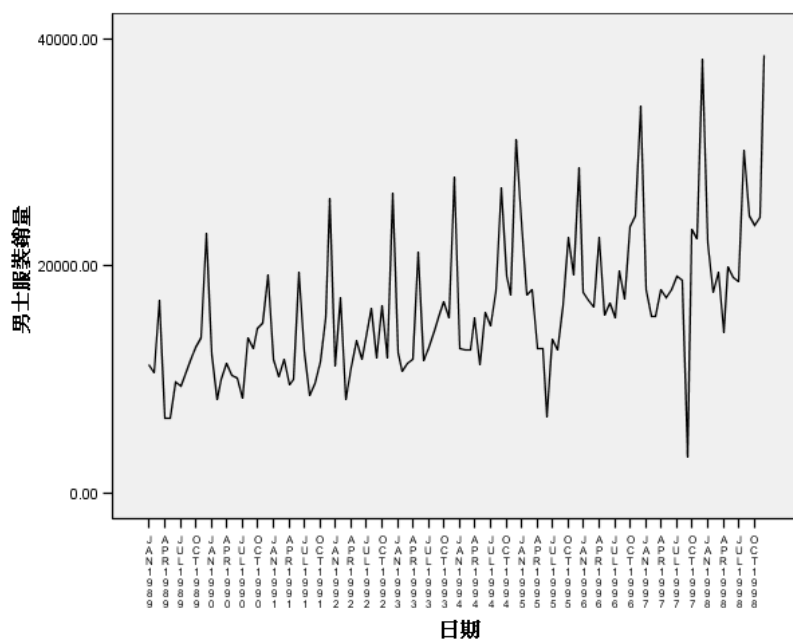
- ▶ 從功能表選擇：  
分析(A) > 預測 > 序列圖...

圖表 8-1  
「序列圖」對話方塊



- ▶ 選取「男性服飾銷售額」，並將它移到「變數」清單中。
- ▶ 選擇「日期」並將它移到「時間軸標記」方塊中。
- ▶ 按一下「確定」。

圖表 8-2  
男裝的銷售額 (以美元為單位)



數列顯示出數個高峰，其中有許多以等距出現，並呈現明顯上升的趨勢。等距的高峰意謂著該時間數列有週期性成份。若該銷售額有週期性的本質，而且銷售額通常會在假期間提高，那麼您便找到了資料的年度週期性成份。

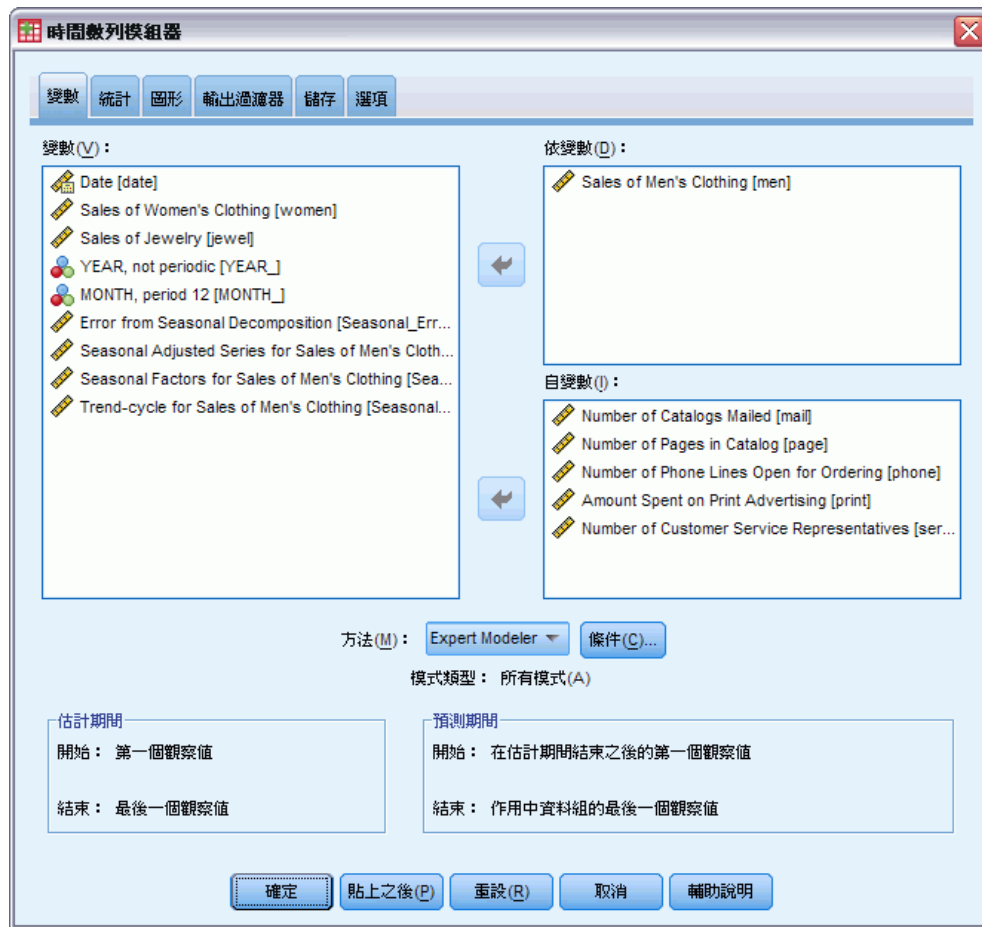
此外也有一些高峰並未呈現週期性樣式，並表示與相鄰的資料點有顯著離異。這些點可能是偏離值，可以且應該由 Expert Modeler 加以說明。

## 執行分析

若要使用 Expert Modeler:

- ▶ 從功能表選擇：  
分析(A) > 預測 > 建立模式(C)...

圖表 8-3  
「Time Series Modeler」對話方塊



- ▶ 選擇「男性服飾銷售額」作為依變數。
- ▶ 選擇「寄出型錄數」到「客服人員數」作為自變數。
- ▶ 請確認已選擇「方法」下拉式清單上的「Expert Modeler」。Expert Modeler 會自動為依變數列找到最適週期性或非週期性模式。
- ▶ 按一下「條件」，再按一下「偏離值」索引標籤。

圖表 8-4  
「Expert Modeler 條件」對話方塊，「偏離值」索引標籤



- ▶ 選擇「自動偵測偏離值」，並為偏離值類型保留預設選擇，以進行偵測。  
我們透過目視方式檢測資料的結果，顯示資料中可能有偏離值。使用目前的選擇，Expert Modeler 會搜尋最常見的偏離值類型，並將任何找到的偏離值併入最終模式中。偏離值偵測會大幅增加 Expert Modeler 所需的計算時間，所以使用此功能時需經過考量，尤其是一次要建構許多數列的模式時。依據預設值，不會偵測偏離值。
- ▶ 按一下「繼續」。
- ▶ 按一下「Time Series Modeler」對話方塊中的「儲存」索引標籤。

圖表 8-5  
「時間數列 Modeler」，「儲存」索引標籤



您要將估計的模式儲存到外部 XML 檔案，讓您可以實驗不同的預測值—使用「套用時間序列模式」程序—而不用重建模式。

- ▶ 按一下「儲存」索引標籤上的「瀏覽」按鈕。  
您將進入儲存檔案的標準對話方塊。
- ▶ 請瀏覽到您想要儲存 XML 模式檔的資料夾，輸入檔案名稱，並按一下「儲存」。
- ▶ 按一下「統計量」索引標籤。



圖表 8-6  
「時間數列 Modeler」，「統計量」索引標籤



- ▶ 選擇「參數估計值」。

此選項會產生一個表格，顯示 Expert Modeler 所選模式的所有參數，其中包括顯著預測值。

- ▶ 按一下「圖形」索引標籤。

圖表 8-7  
「時間數列 Modeler」，「圖形」索引標籤



- ▶ 取消選擇「預測」。

在目前範例中，我們只對判斷顯著預測值和建構模式有興趣。我們不作任何的預測。

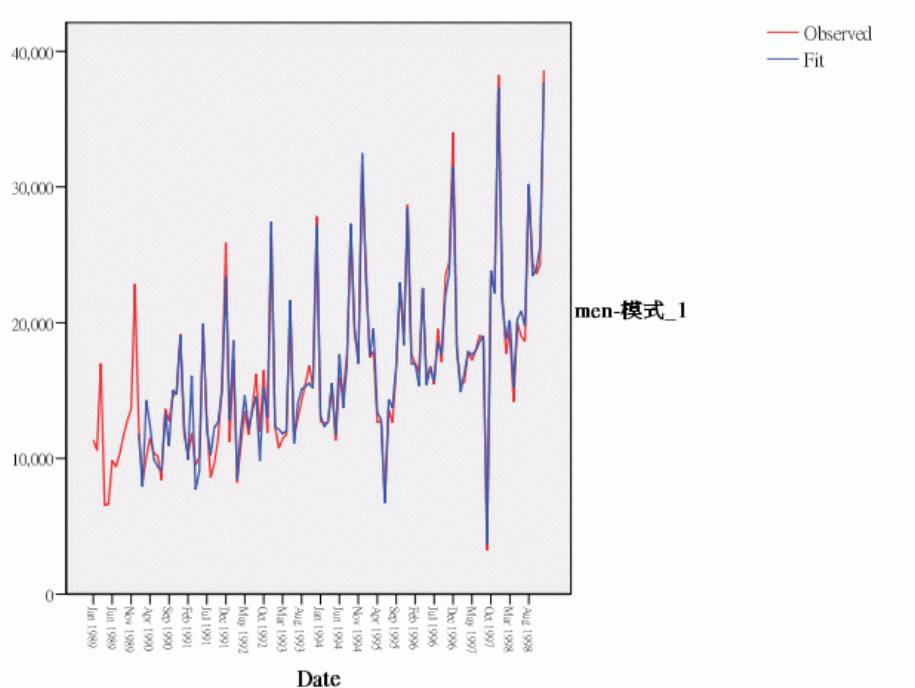
- ▶ 選擇「適合度值」。

此選項顯示某期間內可用於估計模式的預測值。此期間稱為**估計週期**，它包括此例中作用中資料集的所有觀察值。這些數值可顯示模式適合觀察值的程度，所以它們被稱為**適合度值**。所產生的圖形將包括觀察值和適合度值。

- ▶ 在「Time Series Modeler」對話方塊中，按一下「確定」。

## 數列圖

圖表 8-8  
預測和觀察值



預測值與觀察值之間有相當的一致性，表示模式有良好的預測能力。請注意模式可良好預測週期性高峰。此外，它在捕捉資料上升趨勢方面也有不錯的表現。

## 模式說明表

圖表 8-9  
模式說明表

Model ID	Sales of Men's Clothing	模式_1	Model Type
			ARIMA(0,0,0)(0,1,0)

模式說明表中包含代表每個估計模式的項目，也包括模式識別碼和模式類型。模式識別碼由相關依變數的名稱（或標記）和系統指定名稱所組成。在目前範例中，依變數是男性服飾銷售額，系統指定名稱是 Model\_1。

Time Series Modeler 同時支援指數平滑化和 ARIMA 程序模式。指數平滑化模式類型會按照它們常使用的名稱列出，如 Holt 和 Winters 可加性。ARIMA 模式類型是以 ARIMA(p, d, q) (P, D, Q) 標準記號方式列出，其中 p 是自身迴歸的階數，d 是差分（或整合）的階數，q 是移動平均的階數，(P, D, Q) 是其週期性的對應項目。

Expert Modeler 已判定出一階差分的週期性 ARIMA 程式模式最能夠描述男性服飾銷售額。該模式週期性的本質說明了我們在數列圖中看到的週期性高峰，而單一階的差分反映出資料中明顯的上升趨勢。

## 模式統計量表

圖表 8-10  
模式統計量表

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Sales of Men's Clothing-模式_1	2	.948	7.589	18	.984	9

模式統計量表提供每個估計模式的摘要資訊以及適合度統計量。每個模式的結果會以模式說明表所提供的模式識別碼加以標記。首先，請注意在五個您原本指定的候選預測值中，模式包含了其中兩個預測值。所以，這表示 Expert Modeler 已找出兩個可能有助於預測的自變數。

雖然 Time Series Modeler 提供一些不同的適合度統計量，但我們只選擇平穩 R 平方值。此統計量可估計序列中總變異部分，此部分由模式所解釋，而且當有趨勢和週期性樣式時，此統計量比普通 R 平方更好（就此例的狀況）。平穩 R 平方值愈大（最多可達最大值 1），表示越適配。若值為 0.948，表示模式在解釋數列中的觀察差異上表現優異。

Ljung-Box 統計量，也就是所謂的修改 Box-Pierce 統計量，可顯示模式是否正確指定。小於 0.05 的顯著值意謂著觀察數列中有模式無法說明的結構。在此顯示的 0.984 值為不顯著，所以我們可以很有信心模式已正確指定。

Expert Modeler 偵測出有 9 個點應視為偏離值。這些點每一個都適當的設置在模式中，所以您不需要將它們從數列中移除。

## ARIMA 模式參數表

圖表 8-11  
ARIMA 模式參數表

				Estimate	SE	t	Sig.
Sales of Men's Clothing-模式_1	Sales of Men's Clothing	No Transformation	Seasonal Difference	1			
		No Transformation	Numerator Lag 0	1.549	.071	21.943	.000
	Number of Phone Lines Open for Ordering	No Transformation	Seasonal Difference	1			
		No Transformation	Numerator Lag 0	315.262	15.298	20.607	.000
		No Transformation	Seasonal Difference	1			
		No Transformation	Seasonal Difference	1			

ARIMA 模式參數表顯示模式中所有參數的值，每一個由模式識別碼所標記的估計模式在表中皆有一個項目代表。為了我們的目的，該表會列出模式中所有的變數，包括 Expert Modeler 判定為顯著的依變數和任何自變數。我們已從模式統計量表中得知有兩個顯著預測值。模式參數表顯示這兩個預測值是郵寄型錄數和開放訂購電話線數。

## 摘要

您已學會如何使用 Expert Modeler 來建構模式並找出顯著預測值，且您已經將結果模式儲存到外部檔案中。現在您已可以使用「套用時間數列模式」程序，以不同的情境來實驗預測值數列，並觀察情境的不同如何影響銷售預測。

# 套用已儲存模式來對預測值進行測試

您已經使用 Time Series Modeler 來為您的資料建立模式，並找出哪些預測值可能對預測有用。預測值代表您可控制的因子，因此您會想要在預測週期內對其數值進行測試，以了解依變數的預測將如何受到影響。這項工作只要使用 Time Series Modeler 程序所建立的模式檔案，就可利用「套用時間數列模式」程序輕鬆達成。

這個範例是上一個範例的自然延伸，使用 [Expert Modeler 判斷顯著預測值](#) 在 [第 8 章第 57 頁](#)，但這個範例也可以獨立使用。這個情境與一家目錄公司有關，該公司已經收集 1989 年 1 月到 1998 年 12 月男性服飾的每月銷售資料，及其認為可用於未來銷售預測值的幾個數列。Expert Modeler 已經判定五個候選的預測值中，只有兩個是顯著的：郵寄的目錄數量及開放訂購的電話線數量。

您在規劃明年的銷售策略時，可用於列印目錄及開放訂購電話線的資源有限。1999 年前 3 個月中，您所擁有的預算只允許在您的初始規劃之外，多列印 2000 份目錄或多開放 5 條電話線。哪一個選擇會為這三個月創造較多的銷貨收益？

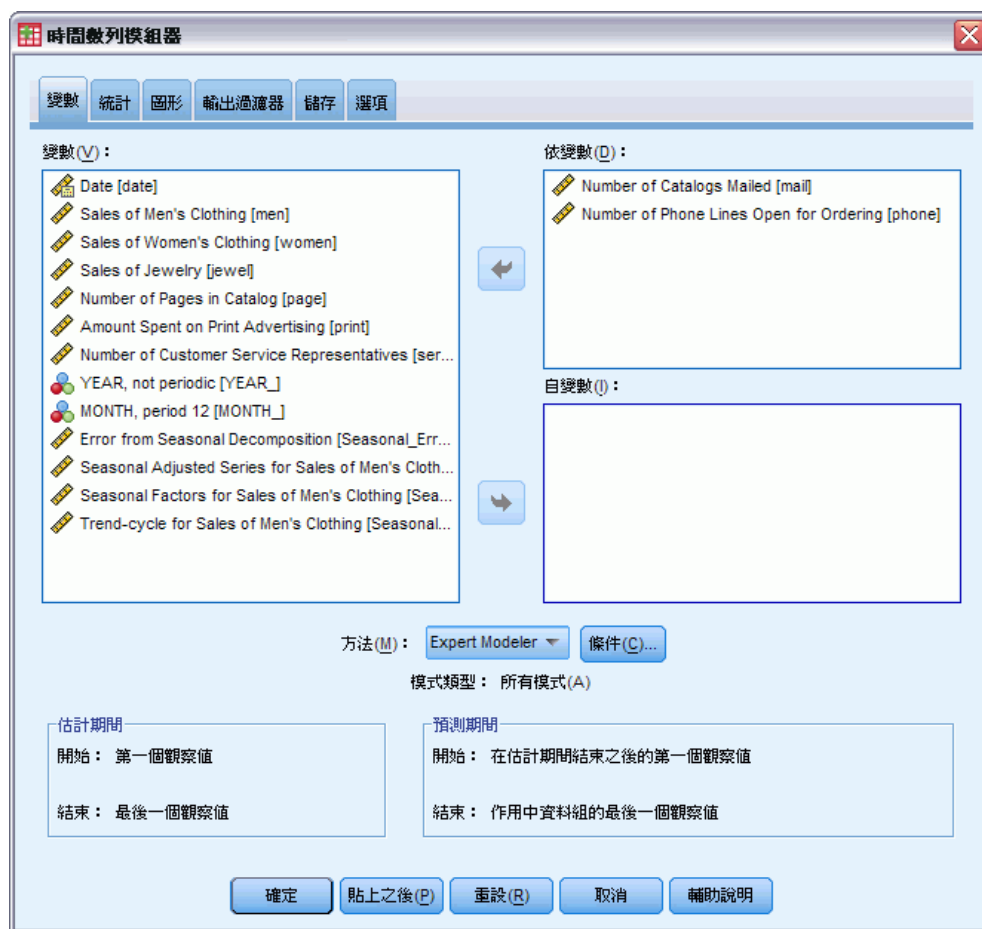
這個範例的資料收集於 catalog\_seasfac.sav 中，且 catalog\_model.xml 包含使用 Expert Modeler 所建立的每月銷售模式。[如需詳細資訊，請參閱附錄 D 中的範例檔案中的 IBM SPSS Forecasting 20](#)。當然，如果您曾實際操作上一個範例並儲存您自己的模式檔案，您可以使用自己的模式檔案，而不需使用 catalog\_model.xml。

## 延長預測值數列

當利用預測值建立依變數數列的預測時，需在預測週期內延長每個預測值數列。除非您很清楚未來的預測值為何，否則您需要估計這些數值。然後您可以修改估計值以測試不同的預測值情境。利用 Expert Modeler 可輕鬆建立初始規劃。

- ▶ 從功能表選擇：
  - 分析 (A) > 預測 > 建立模式 (C)...

圖表 9-1  
「Time Series Modeler」對話方塊



- ▶ 選取「郵寄的目錄數量」及「開放訂購的電話線數量」做為依變數。
- ▶ 按一下「儲存」索引標籤。

圖表 9-2  
「時間數列 Modeler」，「儲存」索引標籤



- ▶ 在「儲存」列中，選擇（檢查）「預測值」項目並以預設值「預測」作為「變數名稱字首」。
- ▶ 按一下「選項」索引標籤。

圖表 9-3  
「時間數列 Modeler」，「選項」索引標籤

時間數列模組器

變數 統計 圖形 輸出過濾器 儲存 選項

預測期間

透過現用資料組最後觀察值的估計期間結束之後的第一個觀察值 (F)

透過指定日期的估計期間結束之後的第一個觀察值 (C)

日期 (D):

年:	月:
1999	3

使用者遺漏值

視為無效 (I)

視為有效 (V)

信賴區間寬度 (%) (W): 95

輸出中的模型識別符號前置符號 (P): 模式

ACF 和 PACF 輸出中顯示的落差最大數目 (X): 24

確定 貼上之後 (P) 重設 (R) 取消 輔助說明

- ▶ 在「預測週期」組別中，選擇「估計週期結束後的第一個觀察值到指定日期」。
- ▶ 在「日期」網格中，輸入年份 1999 及月份 3。  
資料集包含自 1989 年 1 月至 1998 年 12 月的資料，因此依據目前設定，預測週期將為 1999 年 1 月至 1999 年 3 月。
- ▶ 按一下「確定」。



圖表 9-4  
新變數包含預測值數列的預測

	Predicted_mail_Model_1	Predicted_mail_Model_2
121	11742	51
122	11853	45
123	11965	45

「資料編輯程式」顯示新變數 Predicted\_mail\_Model\_1 及 Predicted\_phone\_Model\_2，其中包含郵寄的目錄數量及電話線數量的模式預測值。若要延長預測值數列，只需要 1999 年 1 月到 1999 年 3 月的數值，總計為觀察值 121 至 123。

- ▶ 複製 Predicted\_mail\_Model\_1 內這三個觀察值的數值，再附加到變數 mail。
- ▶ 重複此 Predicted\_mail\_Model\_1 的程序，複製最後三個觀察值，再附加到變數 phone。

圖表 9-5  
預測週期內延長的預測值數列

	mail	page	phone	print	service	YEAR_	MONTH_	DATE_
121	11742	.	51	.	.	1999	1 JAN	1999
122	11853	.	45	.	.	1999	2 FEB	1999
123	11965	.	45	.	.	1999	3 MAR	1999

已延長預測週期內的預測值。

## 修正預測週期內的預測值

為了測試郵寄更多目錄或提供更多電話線這兩個情境，需要分別修正預測值 mail 或 phone 的估計值。由於我們現在只修正三個觀察值（月）的預測值，因此直接在「資料編輯程式」中適當的儲存格內輸入新數值比較容易。為了達到說明的目的，我們將使用「計算變數」對話方塊。您有好幾個數值需要修改時，您可能會發現使用「計算變數」對話方塊更為方便。

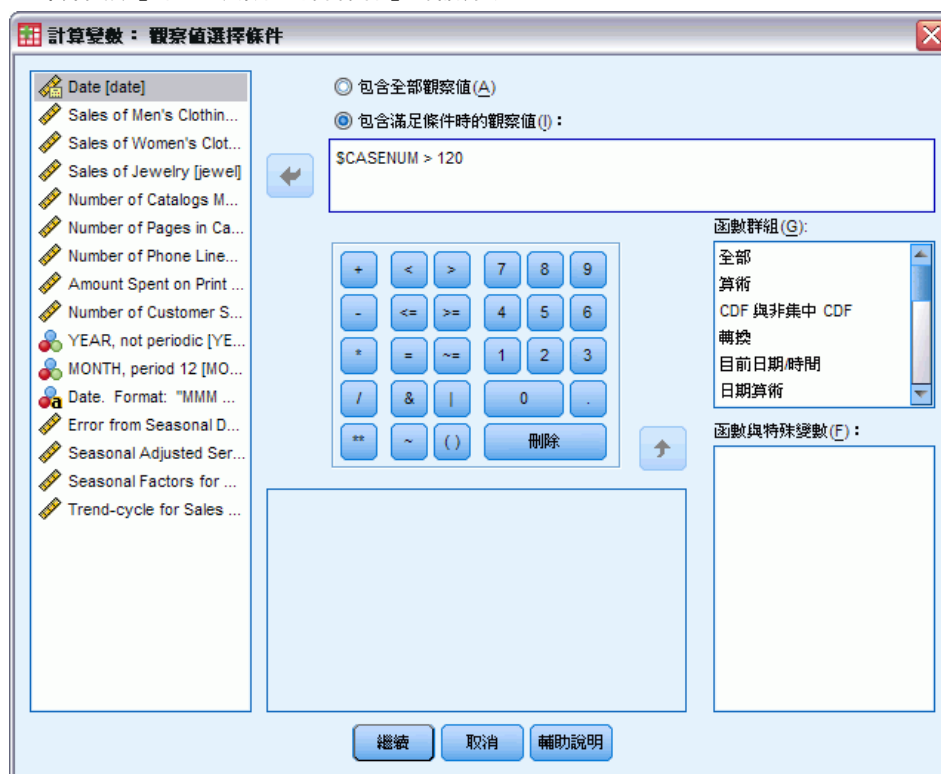
- ▶ 從功能表選擇：  
轉換(T) > 計算變數(C)...

圖表 9-6  
計算變數對話方塊



- ▶ 輸入 mail 做為目標變數。
- ▶ 在「數值運算式」文字方塊中輸入 mail + 2000。
- ▶ 按一下「如果」。

圖表 9-7  
「計算變數」和「觀察值選擇條件」對話方塊



- ▶ 選取「包含滿足條件的觀察值」。
- ▶ 在文字方塊中輸入 `$CASENUM > 120`。  
這將會限制變更預測週期內觀察值的變數 mail。
- ▶ 按一下「繼續」。
- ▶ 在「計算變數」對話方塊中按一下「確定」，及當詢問您是否要變更現有變數時，按一下「確定」。

如此一來會導致在預測週期內的三個月中，每個月 mail 的數值一郵寄的目錄數量一都增加 2000。您現在已經準備好資料可以測試第一個情境，而且您已經準備好執行分析。

## 執行分析

- ▶ 從功能表選擇：  
分析(A) > 預測 > 套用模式(A)...

圖表 9-8  
「套用時間數列模式」對話方塊



- ▶ 按一下「瀏覽」，然後瀏覽至 catalog\_model.xml 並加以選取，或選擇您自己的模式檔案（在上一個範例中儲存的檔案）。如需詳細資訊，請參閱附錄 D 中的範例檔案中的 IBM SPSS Forecasting 20。
- ▶ 在「預測週期」組別中，選擇「估計週期結束後的第一個觀察值到指定日期」。
- ▶ 在「日期」網格中，輸入年份 1999 及月份 3。
- ▶ 按一下「統計量」索引標籤。

圖表 9-9  
「套用時間數列模式」，「統計量」索引標籤



- ▶ 選擇「顯示預測」。  
產生依變數的預測值表。
- ▶ 在「套用時間數列模式」對話方塊中，按一下「確定」。

圖表 9-10  
預測表

模式		JAN 1999	FEB 1999	MAR 1999
Sales of Men's Clothing-Model_1	預測	25279.91	22064.72	21580.96
	UCL	27591.62	24376.42	23892.66
	LCL	22968.21	19753.02	19269.25

預測表包含依變數列的預測值，並將預測週期內 mail 及 phone 預測值的數值列入考量。表格也包括預測值的信賴上限 (UCL) 及信賴下限 (LCL)。

您已經為每個月多郵寄 2000 份目錄的情境建立了銷售預測值。您現在想要準備增加電話線數量的情境資料，這代表重新設定變數 mail 至原始值且將變數 phone 增加 5。您可以重設 mail，只要複製預測週期內的 Predicted\_mail\_Model\_1 再將數值貼上並覆蓋預測週期內 mail 的目前數值即可。然後您可以依照我們增加目錄數量的方法，直接在資料編輯程式內或利用「計算變數」對話方塊，在預測週期內每個月增加 5 條電話線。

若要執行分析，請依如下所示重新開啟「套用時間數列模式」對話方塊：

- ▶ 按一下「叫回對話」工具列按鈕。

- ▶ 選擇「套用時間數列模式」。

圖表 9-11  
「套用時間數列模式」對話方塊



- ▶ 在「套用時間數列模式」對話方塊中，按一下「確定」。

圖表 9-12  
兩個情境的預測表

具有更多目錄的預測

模式		JAN 1999	FEB 1999	MAR 1999
Sales of Men's Clothing-Model_1	預測	25279.91	22064.72	21580.96
	UCL	27591.62	24376.42	23892.66
	LCL	22968.21	19753.02	19269.25

具有更多電話線的預測

模式		JAN 1999	FEB 1999	MAR 1999
Sales of Men's Clothing-Model_1	預測	23757.25	20542.06	20058.29
	UCL	26068.95	22853.76	22370.00
	LCL	21445.55	18230.35	17746.59

顯示的兩個情境預測表指出，預計在預測的三個月中，每月增加郵寄目錄數量所創造的銷售額，會比增加開放訂購電話線數量多出約 \$1500 元。根據分析，將資源分配給多郵寄 2000 份目錄似乎會是明智的作法。

# 週期性分解

## 移除銷售額資料中的週期性

某郵購公司想將預測變數集中男裝產品線銷售額的向上趨勢模式化（變數包括如已寄出之目錄數量，以及開放訂購之電話專線的數量）。為這個目的，該公司收集了十年間男裝的每月銷售額。這個資訊收集於 catalog.sav 中。如需詳細資訊，請參閱附錄 D 中的範例檔案中的 IBM SPSS Forecasting 20。

如需執行趨勢分析，您必須移除資料中存在的任何週期性變數。使用「週期性分解」程序，可以讓您輕鬆完成此工作。

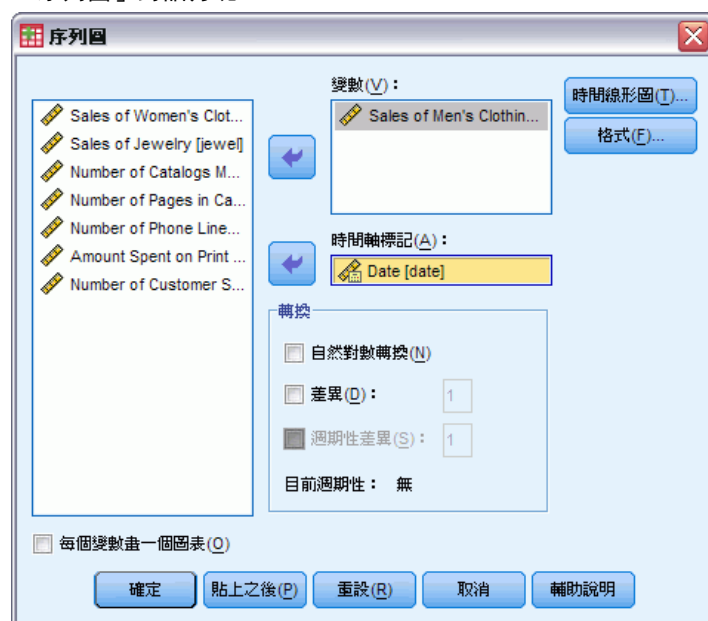
## 決定與設定週期性

「週期性分解」程序需要作用中資料集含有週期性日期成份—例如 12（個月）的年週期、7（天）的星期週期，依此類推。先繪製時間數列是個好主意，因為檢視時間數列圖經常會引起有關基礎週期性的合理猜測。

若要取得一段時間內男裝銷售額的圖表：

- ▶ 從功能表選擇：  
分析(A) > 預測 > 序列圖...

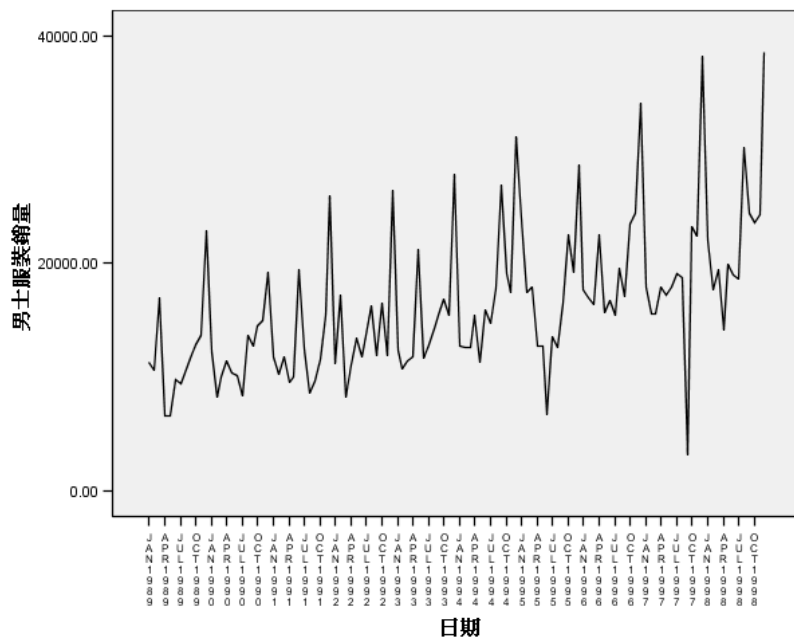
圖表 10-1  
「序列圖」對話方塊



- ▶ 選取「男性服飾銷售額」，並將它移到「變數」清單中。

- ▶ 選取「日期」並將其移到「時間軸標記」清單中。
- ▶ 按一下「確定」。

圖表 10-2  
男裝的銷售額（以美元為單位）



數列顯示有數個高峰，但是這些高峰似乎並非平均分佈。這個輸出表示如果數列含有週期性成份時，同時也會含有非週期的變化 — 這是即時數列的典型範例。小規模的變化除外，顯著高峰顯示不僅以數月作為分隔。以銷售額的週期性本質（在 12 月份的假日季節期間含有典型高點）為前提，時間數列可能具有年度週期性。也請注意，週期性變化會隨著向上的數列趨勢成長，表示週期性變化可能與數列水準成正比，也表示是相乘性模式而不是可加性模式。

檢查時間數列的自身相關和偏自身相關，能提供有關基礎週期性之更詳細的數值結論。

- ▶ 從功能表選擇：  
分析(A) > 預測 > 自身相關(A)...

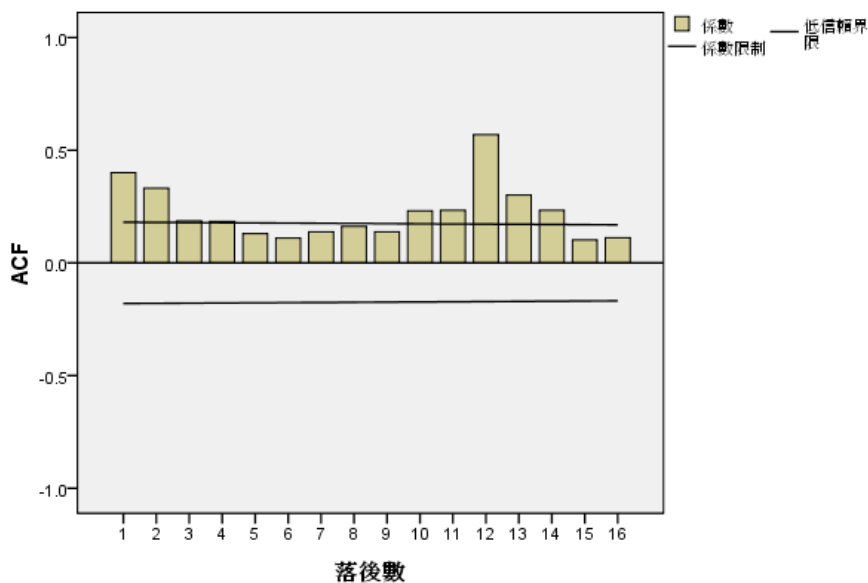


圖表 10-3  
「自身相關」對話方塊



- ▶ 選取「男性服飾銷售額」，並將它移到「變數」清單中。
- ▶ 按一下「確定」。

圖表 10-4  
男人的自身相關圖



自身相關函數會顯示含有較長指數尾之落差 1 的顯著高峰 — 是時間數列的典型。落差 12 的顯著高峰表示資料中含有年度週期性成份。檢查偏自身相關函數能提供更可靠的結論。

圖表 10-5  
男人的偏自身相關圖

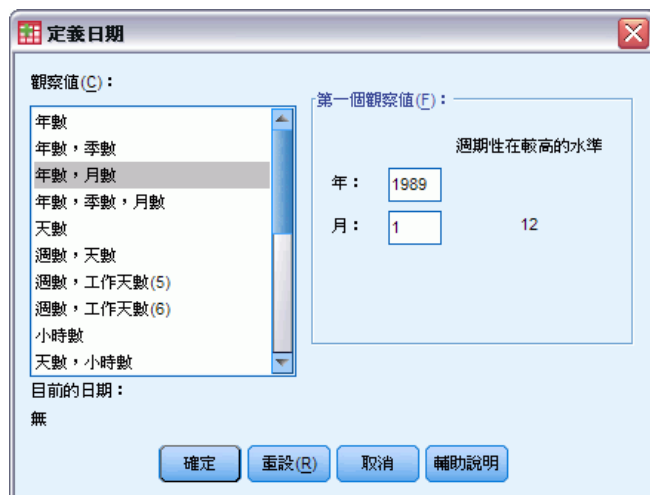


偏自身相關函數之落差 12 的顯著高峰能確定資料中含有年度週期性成份。

如需設定年度週期表：

- ▶ 從功能表選擇：  
資料 > 定義日期 (A)...

圖表 10-6  
定義日期對話方塊



- ▶ 選取「觀察值為」清單中的「年份和月份」。

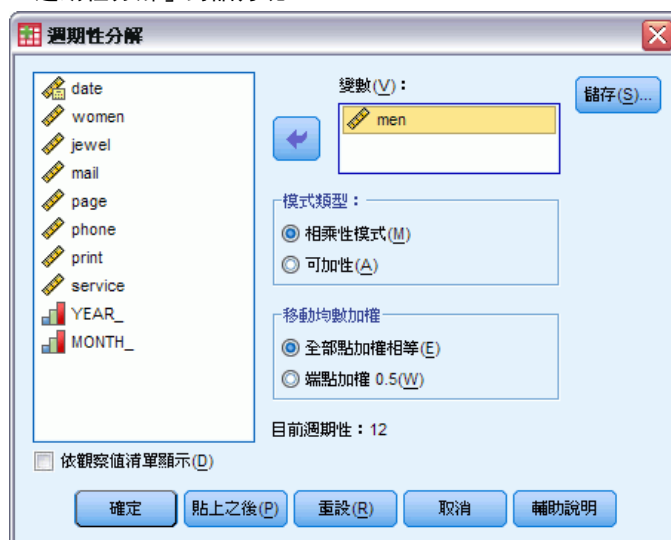
- ▶ 在年份中輸入 1989，並且在月份中輸入 1。
- ▶ 按一下「確定」。  
這會將週期設定為 12，並且會建立設計與「預測」程序一起使用的日期變數集合。

## 執行分析

如需執行「週期性分解」程序：

- ▶ 從功能表選擇：  
分析(A) > 預測 > 週期性分解(S)...

圖表 10-7  
「週期性分解」對話方塊



- ▶ 在來源變數清單中的任一處按一下滑鼠右鍵，並從快顯功能表中選取「Display Variable Names (顯示變數名稱)」。
- ▶ 選取「男人」並將其移到「變數」清單之中。
- ▶ 選取「模式類型」組別中的「相乘性」。
- ▶ 按一下「確定」。

## 瞭解輸出

「週期性分解」程序會為程式所分析的各個原始變數建立四個新變數。依照預設值，會將新變數加入到作用中資料集。新數列含有以下列字首開始的名稱：

**SAF。** 週期調整因子，代表週期性變化。如果是相乘性模式，數值 1 代表沒有週期性變化，如果是可加性模式，則數值 0 代表沒有週期性變化。

**SAS。** 週期調整數列，代表已移除含有週期性變化的原始數列。例如，搭配週期調整數列一起使用時，可以獨立趨勢成份，並且不需任何週期性成份就可以分析趨勢成份。

**STC。** 平滑的趨勢循環成份，是已經平滑化之週期性調整數列的版本，同時顯示趨勢與循環成份。

**ERR。** 特定觀察值之數列殘差成份。

如果是目前的觀察值，週期性調整數列最為適當，因為它表示已經移除含有週期性變化的原始數列。

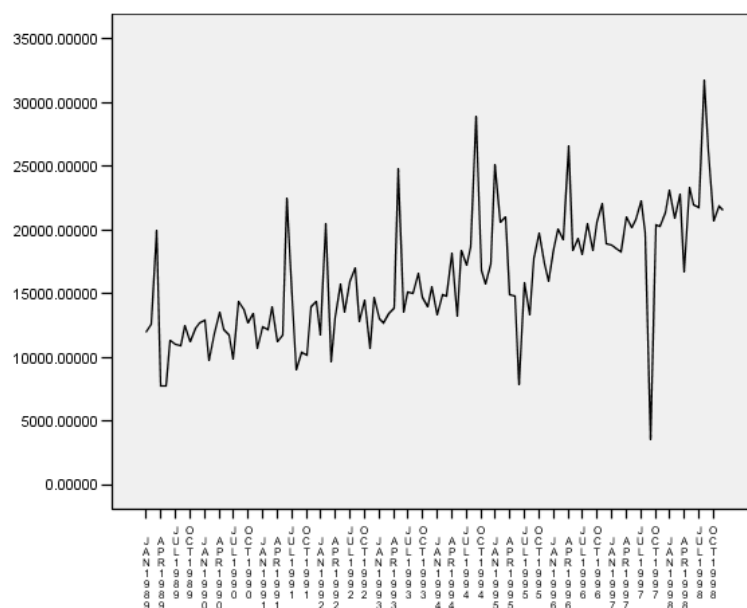
圖表 10-8  
「序列圖」對話方塊



如需繪製週期性調整數列：

- ▶ 開啟「序列圖」對話方塊。
- ▶ 按一下「重設」以清除之前的所有選擇。
- ▶ 在來源變數清單中的任一處按一下滑鼠右鍵，並從快顯功能表中選取「顯示變數名稱」。
- ▶ 選取 SAS\_1，並將它移到「變數」清單中。
- ▶ 按一下「確定」。

圖表 10-9  
週期性調整數列。



週期性調整數列顯示明顯的向上趨勢。有一些明顯的高峰，但是這些高峰的間隔顯示為隨機間隔，沒有顯示出年度樣式。

## 摘要

您已經使用「週期性分解」程序，移除週期性時間數列中的週期性成份，而產生更適合趨勢分析的數列。檢查時間數列之自身相關與偏自身相關對於決定基礎週期性（在此範例中為年度）很實用。

## 相關程序

- 「週期性分解」程序對於從週期性時間數列中移除單一週期性成份很實用。
- 如需執行比部分自身相關函數所提供更深入之時間數列的週期性分析時，請使用「光譜圖」程序。如需詳細資訊，請參閱 [第 11 章](#)。

# 光譜圖

## 使用光譜圖確認週期預期

代表零售商的時間數列通常具有基礎年週期，因為一般的銷售高峰都出現在假期旺季時。建立銷售額預測是指模式化時間數列，也就是找出任何週期性成份。時間數列圖也可能無法發現年週期，因為時間數列包含隨機波動，通常會遮蔽基礎結構。

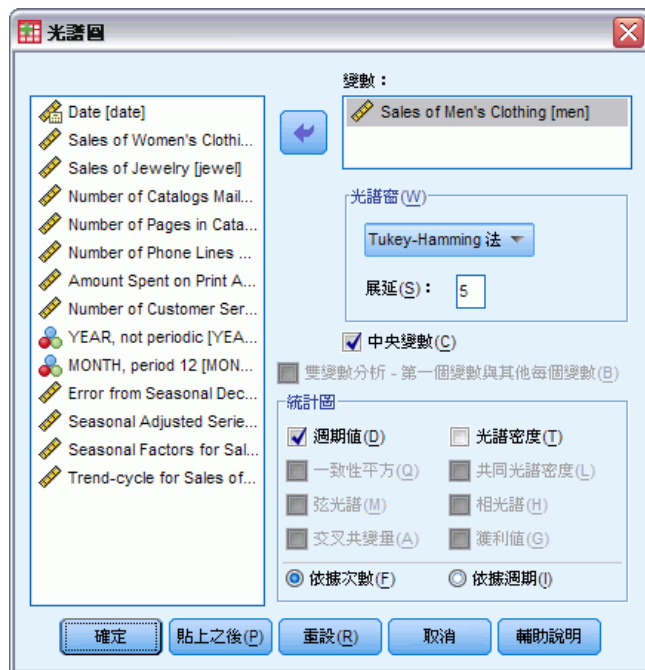
目錄公司每月的銷售資料儲存於 catalog.sav 中。如需詳細資訊，請參閱附錄 D 中的範例檔案中的 IBM SPSS Forecasting 20。在處理銷售額預測之前，您先要確認銷售額資料呈現年週期。時間數列圖顯示出許多波峰，波距不固定，因此若有任何的基礎週期也不明顯。使用「光譜圖」程序找出銷售額資料中的任何週期。

## 執行分析

若要執行「光譜圖」程序：

- ▶ 從功能表選擇：  
分析(A) > 預測 > 光譜分析(T)...

圖表 11-1  
「光譜圖」對話方塊



- ▶ 選取「男性服飾銷售額」，並將它移到「變數」清單中。
- ▶ 選擇「圖形」組別中的「光譜密度」。

► 按一下「確定」。

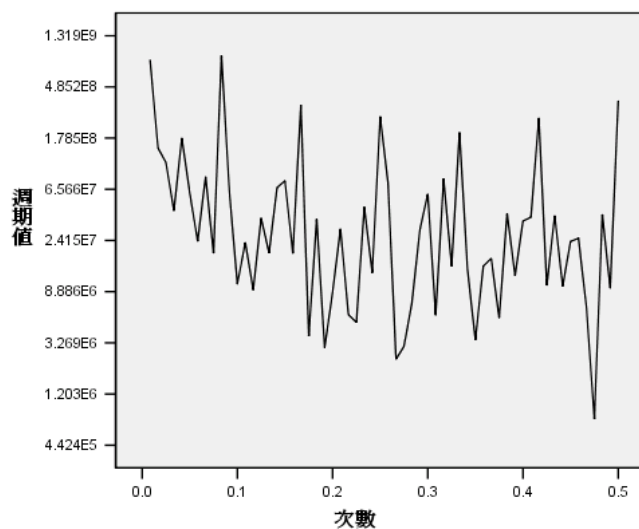
這些選擇會產生下列指令語法：

```
* Spectral Analysis.
TSET PRINT=DEFAULT.
SPECTRA
/VARIABLES=men
/WINDOW=HAMMING(5)
/CENTER
/PLOT=P S BY FREQUENCY.
```

請注意，為了取得輸出中的單變量統計量表，需要變更 TSET 命令以讀取 TSET PRINT=DETAILED。

## 認識週期值和光譜密度

圖表 11-2  
週期值



週期值圖中顯示一系列的高峰從背景噪音中突出，其中最低的頻率高峰出現頻率只不過小於 0.1。您懷疑該資料包括年週期性成份，所以考量年成份對週期值做出的貢獻。時間數列中的每個時間點都代表一個月，所以年週期性會對應到目前資料集中為 12 的週期。由於週期和頻率互為倒數，週期 12 對應的頻率為 1/12 (或 0.083)。所以年成份意謂著在週期值 0.083 處便會出現一個高峰，這似乎與波峰出現的高率小於 0.1 吻合。

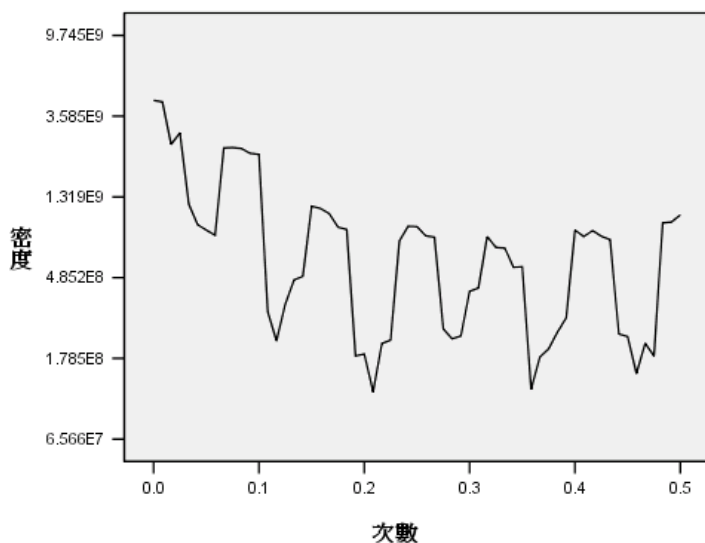
圖表 11-3  
單變量統計量表

數列名稱: men

	次數	正弦轉換	餘弦轉換	週期圖	光譜密度估計值
1	,00000	,000	16242,813	,000	4,361E+09
2	,00833	-3696,643	370,153	828131182	4,278E+09
3	,01667	-1496,277	454,831	146743038	2,526E+09
4	,02500	-1336,400	252,087	110970821	2,921E+09
5	,03333	-662,146	529,734	43143315,6	1,210E+09
6	,04167	-1654,614	501,571	179359225	935924937
7	,05000	-784,814	-636,729	61281367,3	875492029
8	,05833	-335,646	532,062	23744855,0	820375352
9	,06667	-1094,178	-451,489	84064108,9	2,420E+09
10	,07500	264,554	492,876	18774933,5	2,429E+09
11	,08333	-3053,934	2370,483	896742149	2,401E+09
12	,09167	-978,882	-287,035	62435897,6	2,263E+09
13	,10000	-403,128	93,036	10270064,4	2,235E+09

單變量統計量表包含用來繪製週期值圖的資料點。請注意，對於小於 0.1 的頻率，週期值行中的最大值以 0.08333 的頻率出現—剛好與您預期發現到有年週期性成份符合。這項資訊確認找出具有年週期性成份的最低頻率高峰。但其他以較高頻率出現的高峰又如何呢？

圖表 11-4  
光譜密度



光譜密度函數最能夠分析出其餘的高峰，此函數只是平滑化的週期值。平滑化能夠消除週期值圖中的背景噪音，得以更清楚地分離出基礎結構。

光譜密度包括五個明顯的高峰，彼此間距相等。最低頻率高峰僅是於 0.08333 處顯示的平滑化高峰。若要了解前四波高頻率高峰的顯著性，請記住週期值是透過以正弦函數總和模式化時間數列所計算出的。有正弦或餘弦函數形狀（正弦曲線）的週期性成份在週期值圖中會顯示為單一高峰。非正弦曲線的週期性成份會顯示一系列高度不同的



等距高峰，其中以週期性成份的頻率產生一系列的最低頻率波峰。所以在光譜密度中前四波較高的頻率高峰只是指示年週期性成份不是正弦曲線。

您現在已經說明光譜密度圖中的所有可識別的結構，並推斷出資料包含 12 個月為週期的單一週期性成份。

## 摘要

透過使用「光譜圖」程序，您已確定時間數列有年週期性成份，並確認了沒有其他顯著的週期。此外也已看出光譜密度在發現基礎結構上，比週期值更有用，因為光譜密度會將資料中非週期性成份所導致的波動平滑化。

## 相關程序

「光譜圖」程序對於找出時間數列的週期性成份很有用。

- 若要將週期性成份從時間數列中移除—例如，執行趨勢分析—使用「週期性分解」程序。請參閱第 10 章以了解詳細資料。

## 適合度量數

本節將提供用於時間數列模式化中的適合度量數定義。

- **平穩 R 平方。** 將模式中平穩的部份比作簡單平均數模式的測量法。有趨勢或週期性樣式時，此測量法比普通 R 平方更好。平穩 R 平方可以是範圍介於負無限大到 1 之間的負數。負值表示考量的模式比基準線模式更差。正值表示考量的模式比基準線模式更佳。
- **R 平方。** 這是經過模式解釋的數列，當中的變異數總數的比例估計。在平穩數列的情況下，此測量最有用。R 平方可以是範圍介於負無限大到 1 之間的負數。負值表示考量的模式比基準線模式更差。正值表示考量的模式比基準線模式更佳。
- **RMSE。** 根平均平方誤。平均平方誤的平方根。是依變數列與其模式預測水準差異程度的量數，用與依變數列相同的單位來表示。
- **MAPE。** 平均絕對百分比誤差。測量相依數列與其模式預測水準之間的變異程度。其與使用的單元無關，因此可用來比較具有不同單元的數列。
- **MAE。** 平均絕對誤差。測量數列與模式預測水準之間的變異程度。在原始數列單元中會報告 MAE。
- **MaxAPE。** 最大絕對百分比誤差。其為以百分比表示的最大預測誤差。此量數對於想像您預測值的最糟情況非常有用。
- **MaxAE。** 最大絕對誤差。最大預測誤差，其會以作為依變數列的相同單元來表示。其與 MaxAPE 類似，對於想像您預測值的最糟情況非常有用。最大絕對誤差與最大絕對百分比誤差可能發生於不同的數列點—例如當大型數列值之絕對誤差僅略微大於小型數列值之絕對誤差時。在此情況下，最大絕對誤差會發生於較大的數列值，而最大絕對百分比誤差則會發生於較小的數列值。
- **正規化 BIC。** 正規化 Bayesian 資訊準則。它是一種模式整體適合度的一般測量，嘗試說明模式的複雜性。它是依據平均平方誤得到的分數，且包含對模式參數個數與數列長度設下的懲罰。該懲罰會移除參數過多的模式優點，讓統計量較能輕易地比較同一數列的不同模式。

# 偏離值類型

本節將提供用於時間數列模式化中的偏離值類型定義。

- **可加性。** 影響單一觀察值的離群值。例如，若出現資料編碼錯誤，則其可能會識別為可加性離群值。
- **水準變遷。** 藉由常數來變遷所有觀察值的偏離值，其以特定數列點為起始處。當變更原則時會導致發生水準變遷。
- **創新性。** 特定數列點之噪音項目的額外偏離值。對恆定數列而言，創新性偏離值會影響數個觀察值。對非恆定數列而言，創新性偏離值會影響所有起始於特定數列點的觀察值。
- **瞬變。** 其影響會將指數減少至 0 的偏離值。
- **週期可加性。** 這是一個離群值，它會影響一個特定的觀察值，以及與這個觀察值相隔一個或多個週期的所有後續觀察值。它對所有這樣的觀察值影響都相同。若在某年年初，每個一月的銷售量都比較高，就可能產生週期可加性離群值。
- **局部趨勢。** 在特定數列點起始局部趨勢的偏離值。
- **可加性補充。** 兩個以上連續的可加性離群值組別。若選取此離群值類型，則除了會偵測個別的可加性離群值外，也會偵測補充的可加性離群值。

## ACF/PACF 圖形指引

此處顯示的圖形是純或理論 ARIMA 處理的圖形。下面是一些識別此處理程序的一般準則：

- 非平穩型數列有 ACF，對六個或以上落差仍保持顯著，而不是迅速衰退至 0。您必須區別這類數列直到平穩，才能識別此處理程序。
- 自身迴歸處理有指數型衰退 ACF，而在 PACF 的最初一個或多個落差中達到波峰。波峰數表示自身迴歸的次序。
- 移動平均數處理在 ACF 的最初一個或多個落差中有波峰，並有指數型衰退的 PACF。波峰數表示移動平均數的次序。
- 混合 (ARMA) 處理一般在 ACF 和 PACF 中都會顯示指數型衰退。

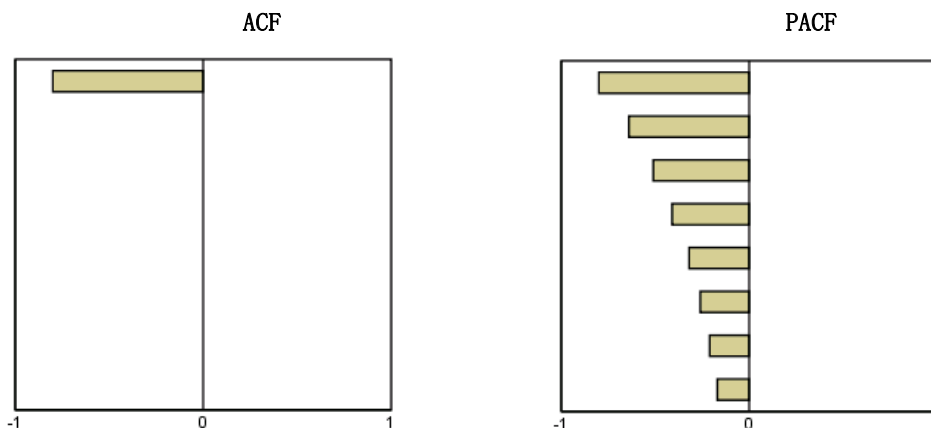
在識別階段時，您不需要擔心 ACF 或 PACF 的訊號，也不必擔心指數型衰退 ACF 或 PACF 接近 0 的速度。這些都是依 AR 和 MA 係數的訊號和實際值而定。在某些情況下，指數型衰退 ACF 會在正負值之間交替互換。

擷取自實際資料的 ACF 和 PACF 圖形絕不會像此處所顯示一般整齊清楚。您必須學習如何在所取得的任何圖形中找出重要資訊。一定要檢查殘差的 ACF 和 PACF，以免識別有誤。請將下列情形謹記在心：

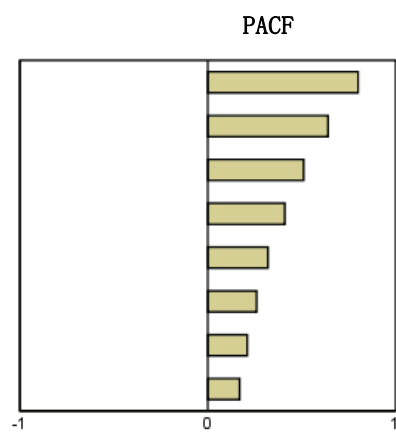
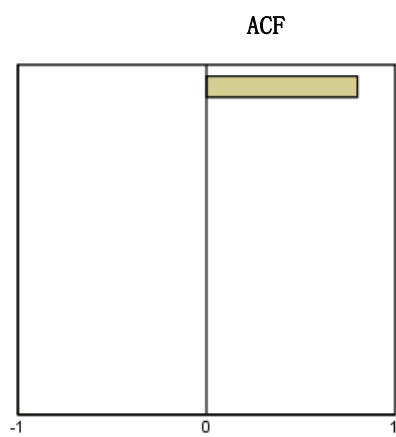
- 週期性處理會在週期性落差（多個週期性期間）顯示這些模式。
- 您有權將非顯著值視為 0，也就是說，您可以忽略在圖形上落於信賴區間內的值。但特別是在持續統計上顯著值模式時，則不必予以忽略。
- 偶發性自身相關將會僅因機率而成為統計上顯著值。如果是單獨出現，您可以忽略統計上顯著自身相關，尤其是高落差，且不是在週期性落差上發生。

參考 ARIMA 分析上的任何文字，取得 ACF 和 PACF 圖形更完整的討論。

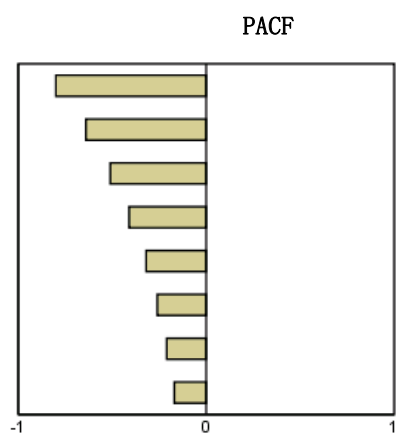
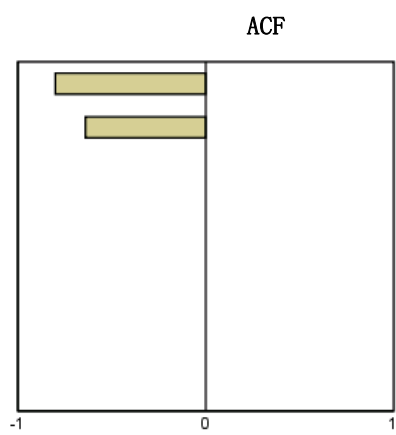
表格 C-1  
ARIMA (0, 0, 1),  $q > 0$



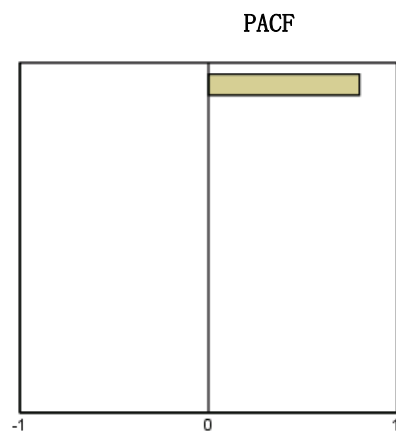
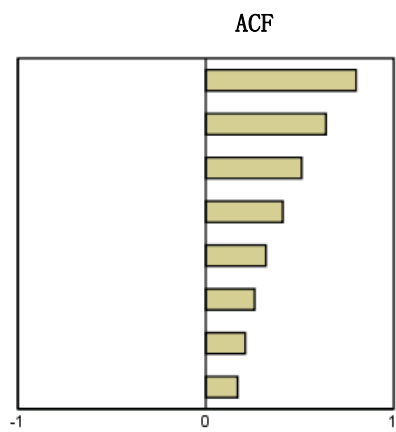
表格 C-2  
ARIMA (0, 0, 1),  $q < 0$



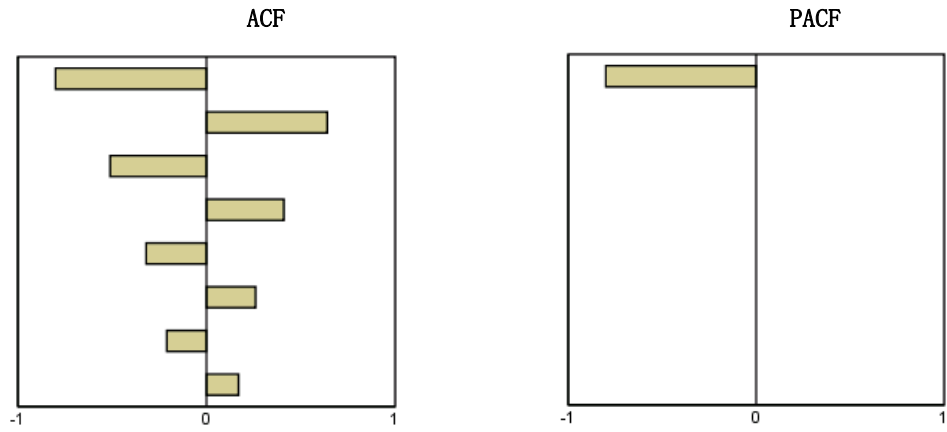
ARIMA (0, 0, 2),  $\theta_1\theta_2 > 0$



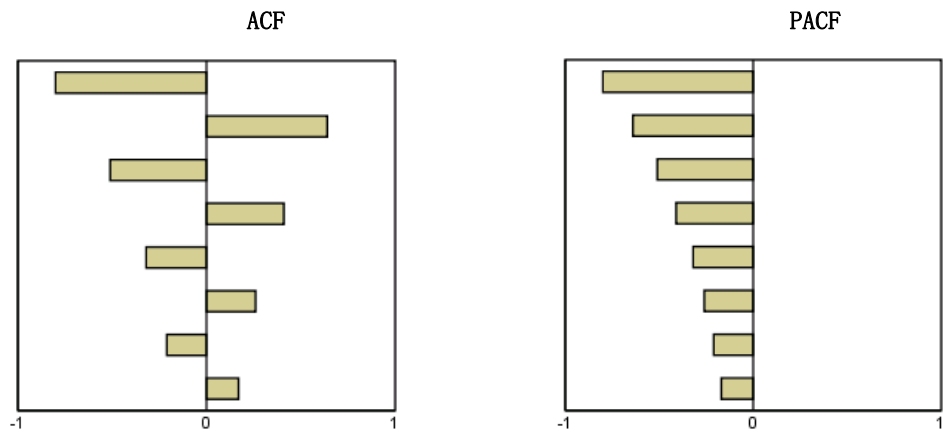
表格 C-3  
ARIMA (1, 0, 0),  $f > 0$



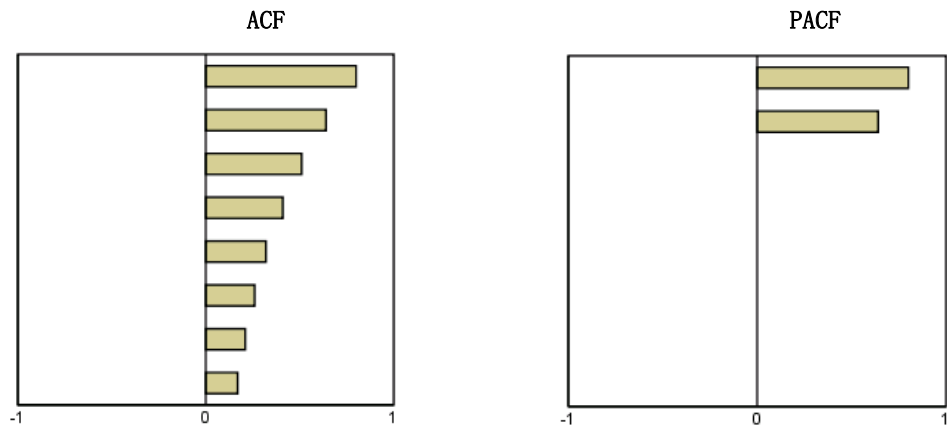
表格 C-4  
 ARIMA(1, 0, 0),  $f < 0$



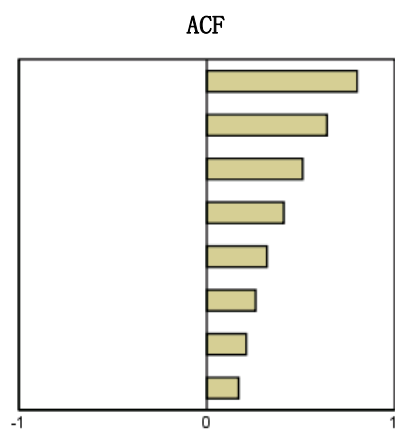
ARIMA(1, 0, 1),  $\phi < 0, \theta > 0$



ARIMA(2, 0, 0),  $\phi_1\phi_2 > 0$



表格 C-5  
ARIMA (0, 1, 0) (整合的數列)



# 範例檔案

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

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

## 說明

以下是使用於本文件中不同範例的範例檔之簡要描述。

- **accidents.sav**。這是有關某保險公司研究年齡和性別風險因子對給定地區汽車意外事件的假設資料檔。每一個觀察值對應至一個年齡類別和性別的交叉分類。
- **adl.sav**。這是有關致力於確定一個建議中風病患治療類型之效益的假設資料檔。醫師隨機指定女性中風病患至兩個組別之一。第一組接受標準的物理治療，而第二組則接受額外的情緒治療。在治療了三個月後，將每一個病患進行日常活動的能力記分為次序變數。
- **advert.sav**。這是有關一家零售商致力於調查廣告費與廣告後銷售情形之間的關係的假設資料檔。為了這個目的，他們收集了過往銷售數字和相關的廣告費用。
- **aflatoxin.sav**。這是有關檢定玉米作物是否有黃麴毒素（一種毒物，其濃度在介於和處於作物產量中都有很大的差異）的假設資料檔。一名穀物加工者收到來自 8 個作物產量各 16 個樣本，並以十億當量（PPB）來測量黃麴毒素的水準。
- **anorectic.sav**。在將厭食/暴食行為症狀學標準化的過程中，研究人員（Van der Ham, Meulman, Van Strien, 和 Van Engeland, 1997）研究了 55 個飲食失調的青少年。每個病患在四年之中被訪問四個回合，所以得到總數為 220 的觀察值。在每次觀察中，為病患在 16 種症狀上逐一評分。目前遺漏了第二次訪察的病患 71，第二次訪察的病患 76，以及第三次訪察的病患 47 的症狀分數，因此只剩下 217 個有效觀察值。
- **bankloan.sav**。這是有關一家銀行致力於減少放款利率預設值的假設資料檔。本檔包含 850 位以前的客戶與現在的準客戶的財務和人口資料。前 700 個觀察值為以前有借貸的客戶。最後 150 個觀察值是銀行需要作信用風險優良與不良分類的準客戶。
- **bankloan\_binning.sav**。這是包含 500 位以前客戶的財務和人口資料的假設資料檔。
- **behavior.sav**。在典型範例（Price 和 Bouffard, 1974）中，52 名學生被要求為 15 種情境與 15 種行為組合評等，等級共分為 10 點，從 0 = 「非常適當」到 9 = 「非常不適當」。平均值超過個別值，值會被視為相異性。
- **behavior\_ini.sav**。本資料檔包含 behavior.sav 之二維解的起始組態。
- **brakes.sav**。這是有關一間生產高性能汽車碟型煞車片工廠中品質管制的假設資料檔。資料檔包含由 8 個生產機器分別取得 16 個碟片的直徑測量。煞車的目標直徑是 322 公釐。



- **breakfast.sav**。在經典研究中(Green 和 Rao, 1972), 21 名 Wharton 學院 MBA 學生及其配偶被要求為 15 項早餐食品按喜愛程度分出等級: 從 1 = 「最喜愛」到 15 = 「最不喜愛」。他們的喜愛程度分六種不同情況記錄, 從「整體喜愛」到「點心, 僅配飲料」。
- **breakfast-overall.sav**。本資料檔只包含第一種情況—「整體喜愛」—所喜愛的早餐項目。
- **broadband\_1.sav**。這是包含全國性寬頻服務地區用戶數目的假設資料檔。本資料檔包含四年期間 85 個地區每月的用戶數目。
- **broadband\_2.sav**。本資料檔與 broadband\_1.sav 相同, 但多了三個月的資料。
- **car\_insurance\_claims.sav**。一個在別處 (McCullagh 和 Nelder, 1989) 出現和分析過, 有關汽車損害理賠的資料集。理賠金額的平均數可建立模式為具有 gamma 分配, 使用反連結函數將依變數的平均數相關至一被保險人年齡、車輛類型、和車齡的線性組合。提出理賠的數量可以用作尺度權重。
- **car\_sales.sav**。本資料檔包含假設性的銷售估計、定價、和不同的品牌與車輛型式的實體規格。定價和實體規格是由 edmunds.com 和製造商處輪流取得。
- **car\_sales\_uprepared.sav**。這是 car\_sales.sav 的修改版本, 其中不包含任何欄位的轉換版本。
- **carpet.sav**。在一個普遍的範例 (Green 和 Wind, 1973) 中, 計劃銷售全新地毯清潔機的公司想要檢驗影響消費者偏好的五個因子—包裝設計、品牌名稱、價格、「優秀家用品」獎章及退費保證。包裝設計有三個因子水準, 每個水準中的清潔刷位置都不相同; 三個品牌名稱 (K2R、Glory、及 Bissell); 三個價格水準; 且最後兩個因子各有兩個水準 (無論無或有)。十名消費者將這些因子所定義的 22 種組合分級。「偏好」變數包含每個組合平均排名的等級。排名數值較小者會對應高偏好程度。這個變數反映每個組合偏好的整體量數。
- **carpet\_prefs.sav**。本資料檔是根據 carpet.sav 所描述的相同範例, 但它包含 10 個消費者每一個人的實際等級。消費者被要求將 22 個產品組合從最喜歡排列到最不喜歡。變數「PREF1」到「PREF22」包含相關組合的識別碼, 如 carpet\_plan.sav 中所定義。
- **catalog.sav**。本資料檔包含郵購公司銷售三項產品的每月假設銷售數字。也包含五個可能預測變數的資料。
- **catalog\_seasfac.sav**。本資料檔與 catalog.sav 相同, 不過多了一組由「週期性分解」程序所計算的週期性因子以及隨附的資料變數。
- **cellular.sav**。這是有關一家手機公司致力於減少顧客不忠的假設資料檔。顧客不忠傾向分數套用於帳戶, 範圍由 0 至 100。帳戶分數 50 或以上有可能正尋求變更供應商。
- **ceramics.sav**。這是有關一家製造商致力於確定一種新的優良合金是否較標準的合金有較大的耐熱性的假設資料檔。每一個觀察值代表對合金之一的不同檢定; 記錄了讓軸承失效的溫度。
- **cereal.sav**。這是有關對 880 人的早餐喜好進行訪談的假設資料檔, 也記下他們的年齡、性別、婚姻狀況、和是否有活躍的生活型態 (根據他們是否一週運動兩次)。每一個觀察值代表一位不同的應答者。
- **clothing\_defects.sav**。這是有關一家服裝工廠品質管制過程的假設資料檔。由該工廠所生產的每一批產品中, 檢查員取出一件服裝的樣本並計算不合格的服裝個數。

- **coffee.sav**。本資料檔是關於六種冰咖啡品牌的感覺印象(Kennedy, Riquier, 和 Sharp, 1996)。對 23 種冰咖啡中每一種的印象屬性, 由群眾來選取依其屬性描述的所有品牌。該六種品牌已標示為 AA、BB、CC、DD、EE、和 FF, 以保持機密。
- **contacts.sav**。這是有關一群公司電腦銷售代表聯絡清單的假設資料檔。每一個聯絡人依他們在公司所服務的部門及其公司的等級而分類。最後一次銷售的金額、到最後一次銷售的時間、和該聯絡人公司的規模也都被列入記錄。
- **creditpromo.sav**。這是有關一家百貨公司致力於評估近期信用卡促銷活動效果的假設資料檔。為達此目標, 隨機選取了 500 位持卡人。有半數收到廣告, 促銷在未來三個月購買將獲得降低利率的優惠。半數收到標準的週期性廣告。
- **customer\_dbase.sav**。這是有關一家公司致力於使用其資料倉庫的資訊來對最有可能回應的客戶提供優惠的假設資料檔。隨機選取客戶庫的子集, 提供優惠, 再將他們的回應記錄下來。
- **customer\_information.sav**。本檔案是包含客戶郵寄資訊的假設資料檔, 例如姓名和地址。
- **customer\_subset.sav**。80 個 customer\_dbase.sav 的觀察值子集。
- **debate.sav**。這是有關一項政治辯論會參與者辯論前和辯論後接受調查之成對反應的假設資料檔。每一個觀察值對應至一位不同的應答者。
- **debate\_aggregate.sav**。這是將 debate.sav 中之反應作整合的假設資料檔。每一個觀察值對應至辯論前和辯論後對偏好之交叉分類的反應。
- **demo.sav**。這是有關提供郵寄每月優惠之購買客戶資料庫的假設資料檔。記錄了客戶是否對該優惠回應, 以及各種的人口資訊。
- **demo\_cs\_1.sav**。這是有關一家公司致力於匯編調查資訊資料庫之第一步的假設資料檔。每一個觀察值對應至一個不同的城市, 也記錄了其地區、省、區、和城市識別。
- **demo\_cs\_2.sav**。這是有關一家公司致力於匯編調查資訊資料庫之第二步的假設資料檔。每一個觀察值對應至在第一步中選取的城市中的一個不同的家庭單位, 也記錄了其地區、省、區、分區、和單位識別。也納入了由該設計的前兩階段所得之取樣資訊。
- **demo\_cs.sav**。這是包含以複合取樣設計所收集之調查資訊的假設資料檔。每一個觀察值對應至一個不同的家庭單位, 也記錄了各種的人口和取樣資訊。
- **dmdata.sav**。這是包含直效行銷公司之人口和購買資訊的假設資料檔。dmdata2.sav 包含收到測試郵件的連絡人子集資訊, 而 dmdata3.sav 則包含剩下未收到測試郵件的連絡人資訊。
- **dietstudy.sav**。本假設資料檔包含對「Stillman 飲食法」(Rickman, Mitchell, Dingman, 和 Dalen, 1974) 研究的結果。每一個觀察值對應至一個不同的受試者, 並記錄下他或她飲食法前、後之體重(磅)和三酸甘油酯水準(毫克/100 毫升)。
- **dvdplayer.sav**。這是有關新 DVD 播放器開發的假設資料檔。市場行銷團隊使用原型收集了焦點組別資料。每一個觀察值對應至不同調查到的使用者, 並記錄下一些有關他們的人口資訊和他們對有關原型問題的回應。
- **german\_credit.sav**。本資料檔取自(Blake 和 Merz, 1998) 艾文(Irvine) 在加州大學機器學習資料庫儲存器的「德國信用」資料集。
- **grocery\_1month.sav**。本假設資料檔是將 grocery\_coupons.sav 資料檔和每週購買的「彙總」, 因此每一個觀察值對應至一個不同的客戶。結果部份每週變更的變數消失了, 而目前所記錄的銷售量是在研究的四週期間銷售量之總和。

- **grocery\_coupons.sav**。這是包含某連鎖雜貨店想要知道他們客戶購買習慣所收集之調查資料的假設資料檔。每一個客戶被追蹤了四週，每一個觀察值對應至一個不同的客戶一週，並記錄有關客戶在何處及如何購物的資訊，包含那一週在雜貨店花了多少錢。
- **guttman.sav**。Bell (Bell, 1961) 以此表說明可能的社會團體。Guttman (Guttman 值, 1968) 過去曾使用此表的一部分，在這部分中有 5 個變數，分別說明 7 個理論社會團體的社會互動、團體歸屬感、成員實際接觸和關係正式性，而這 7 個群組包括：群眾（例如，足球場上的人）、觀眾（例如在戲院中和課堂上的人）、公眾（例如，報紙讀者和電視觀眾）、暴民（和群眾相似，但互動較為激烈）、原級團體（親密性）、次級團體（自願性）和現代社群（因親密的身體接近而導致鬆散的結盟和特殊服務的需求）。
- **health\_funding.sav**。這是包含醫療保健基金（每 100 個人口的金額）、疾病率（每 10,000 個人口的比率）、造訪醫療保健機構的比例（每 10,000 個人口的比率）的假設資料檔。每一個觀察值代表一個不同的城市。
- **hivassay.sav**。這是有關一家製藥實驗室致力於開發一種偵測 HIV 感染快速檢驗的假設資料檔。檢驗結果是八個紅色加深的陰影，陰影愈深表示感染的可能性愈大。進行了一項實驗室的試驗，在 2,000 個血液樣本中，有半數遭到 HIV 的感染，而半數則未感染。
- **hourlywagedata.sav**。這是有關在辦公室和醫院任職的護士依經驗水準不同之鐘點費的假設資料檔。
- **insurance\_claims.sav**。這是有關一家保險公司想要建立模式來標示可疑及可能的詐欺理賠之假設資料檔。每一個觀察值代表個不同的理賠。
- **insure.sav**。這是有關一家保險公司正在研究表示客戶是否必定理賠 10 年壽險合約之風險因子的假設資料檔。在資料檔中的每一個觀察值代表二份合約，其一記錄了理賠而另一則否，二者的年齡和性別相符。
- **judges.sav**。這是有關受過訓練的裁判（加上一位熱心人士）為 300 個體操表演評分的假設資料檔。每一列代表一個不同的表演；裁判們觀看相同的表演。
- **kinship\_dat.sav**。Rosenberg 與 Kim (Rosenberg 和 Kim, 1975) 致力於分析 15 個親屬關係稱呼（姑/姨、兄弟、堂/表兄弟姐妹、女兒、父親、孫女、祖父、祖母、孫子、母親、姪子/外甥、姪女/外甥女、姐妹、兒子、叔/舅父）。他們請四組大學生（兩組女性、兩組男性）根據其相似性來分類整理這些稱謂。他們請其中兩組（一組女性、一組男性）作兩次分類整理，第二次要根據與第一次不同的準則進行分類整理。因此，總共得到六個「來源」。每一個來源對應至一個  $15 \times 15$  的相似性矩陣，其儲存格等於來源中人數減去物件在該來源中分為同組的次數。
- **kinship\_ini.sav**。本資料檔包含 kinship\_dat.sav 之三維解的起始組態。
- **kinship\_var.sav**。本資料檔包含自變數「性別」、「世代」、和可用來解讀 kinship\_dat.sav 解答維度的（分離）「度」。尤其，它們可用來將解答空間限制為這些變數的線性組合。
- **marketvalues.sav**。本資料檔有關於一項在伊立諾州阿爾岡京 (Algonquin, Ill.) 的新屋開發案自 1999 年至 2000 年之房屋銷售情況。這些銷售與公共記錄有關。
- **nhis2000\_subset.sav**。「國民健康訪問調查 (NHIS)」為美國民間人口的一大型民眾調查。其以具全國代表性的家庭為樣本，面對面的完成訪問。而取得各家庭中成員的人口統計學資訊及健康行為、健康狀態方面等觀察報告。本資料檔包含一個 2000 年調查資訊的子集。國家衛生統

計中心。2000 年「國民健康訪問調查 (NHIS)」。公用資料檔案和文件。  
[ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NHIS/2000/](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/)。2003 年曾存取。

- **ozone.sav**。本資料包含對六個氣象變數所作的 330 個觀察值，以自其餘的變數中預測臭氧濃度。先前研究人員中，(Breiman 和 Friedman 檢定(F), 1985)、(Hastie 和 Tibshirani, 1990) 在這些會阻礙標準迴歸方式的變數中發現非線性。
- **pain\_medication.sav**。本假設資料檔包含治療慢性關節炎疼痛之消炎藥物臨床試驗的結果。特別關注於藥物發生作用的時間以及它是如何與現用藥物作比較。
- **patient\_los.sav**。本假設資料檔包含對因可能為心肌梗塞 (MI, 或「心臟病」) 入院病患的治療記錄。每一個觀察值對應至一個不同的病患並記錄許多與其留院期間有關的變數。
- **patlos\_sample.sav**。本假設資料檔包含病患在為心肌梗塞 (MI, 或「心臟病」) 治療期間接受血栓溶解治療的治療記錄樣本。每一個觀察值對應至一個不同的病患並記錄許多與其留院期間有關的變數。
- **poll\_cs.sav**。這是有關民意測驗專家致力於確定交付立法之前公眾對法案支持水準的假設資料檔。觀察值對應至登記選民。每一個觀察值記錄下選民的郡、鎮、和他居住的鄰近範圍。
- **poll\_cs\_sample.sav**。本假設資料檔包含列於 poll\_cs.sav 中的選民樣本。樣本是根據在 poll\_csplan 計劃檔中指定的設計來取得，而本資料檔記錄了包含機率和樣本權重。不過，請注意，由於取樣計劃採用到機率 - 比例 - 大小 (PPS) 方法，也用到一個包含聯合選擇機率的檔案 (poll\_jointprob.sav)。其他與選民人口及其對提議法案之意見有關的變數都在取樣後收集並加入資料檔中。
- **property\_assess.sav**。這是有關郡財產估價人員致力於對限定資源保持財產價值評估維持最新的假設資料檔。觀察值對應至郡內過去一年銷售的財產。資料檔中的每一個觀察值記錄了財產所在的鎮、上次訪查該財產的估價人員、自那次評估後經過的時間、當時定的估價、和該財產銷售價值。
- **property\_assess\_cs.sav**。這是有關州財產估價人員致力於對限定資源保持財產價值評估維持最新的假設資料檔。觀察值對應至州中的財產。資料檔中的每一個觀察值記錄了郡、鎮、和財產所在的鄰近範圍、自最後一次評估後經過的時間、和當時定的估價。
- **property\_assess\_cs\_sample.sav**。本假設資料檔包含列於 property\_assess\_cs.sav 中的財產樣本。樣本是根據在 property\_assess\_csplan 計劃檔中指定的設計來取得，而本資料檔記錄了包含機率和樣本權重。另外的變數「目前價值」是在取樣後收集並加入資料檔中。
- **recidivism.sav**。這是有關政府法令執行機構致力於瞭解其轄區內之再犯率的假設資料檔。每一個觀察值對應至一個先前的違法者並記錄其人口資訊、第一次犯罪的一些細節、然後是直到第二次被捕的時間 (如果它發生在第一次被捕的兩年之內)。
- **recidivism\_cs\_sample.sav**。這是有關政府法令執行機構致力於瞭解其轄區內之再犯率的假設資料檔。每一個觀察值對應到一個先前的違法者，在 2003 年六月第一次被捕後釋放，並記錄其人口資訊、第一次犯罪的一些細節、和第二次被捕日期 (如果它發生在 2006 年六月之前)。違法者是根據在 recidivism\_cs\_csplan 中所指定的取樣計劃之樣本部門來選取；由於取樣計劃採用到機率 - 比例 - 大小 (PPS) 方法，也用到一個包含聯合選擇機率的檔案 (recidivism\_cs\_jointprob.sav)。

- **rfm\_transactions.sav**。本檔案是包含購買交易資料的假設資料檔，包括購買日期、購買項目及每一項交易的金額。
- **salesperformance.sav**。這是有關評估兩個新售貨員訓練課程的假設資料檔。六十個員工，分成三個組別，全部接受標準訓練。此外，組別二得到技術訓練；組別三則是實務輔導簡介。每一個員工在訓練課程結束時接受測驗並記錄他們的分數。在資料檔中每一個觀察值代表一個不同的訓員，並記錄他們所分派的組別和他們在測驗中得到的分數。
- **satisf.sav**。這是有關一家零售公司在 4 個商店位置所作之滿意度調查的假設資料檔。總共有 582 位客戶接受調查，每一個觀察值代表一位客戶的反應。
- **screws.sav**。這個資料檔包含螺絲釘、螺栓、螺帽和圖釘之特色的資訊 (Hartigan, 1975)。
- **shampoo\_ph.sav**。這是有關一家美髮產品工廠品質管制過程的假設資料檔。在固定的時間間隔，記錄下六個不同輸出批次的測量和它們的 pH 值。目標範圍是 4.5 - 5.5。
- **ships.sav**。一個在別處 (McCullagh et al., 1989) 出現和分析過，有關商船因風浪所造成損壞的資料集。事件次數可建立模式為以 Poisson 率發生，給定船型、建造期間、和服務期間。以因子交叉分類所形成的表格的每一個儲存格服務月數的整合，提供了暴露於風險之值。
- **site.sav**。這是有關一家公司致力於為事業擴展選擇新地點的假設資料檔。他們僱請兩位顧問分別評估該地點，除了一份廣泛的報告之外，他們還要將每個地點摘要為前景「佳」、「可」、或「差」。
- **smokers.sav**。本資料檔是由「1998 年全國家庭毒品濫用調查」中摘錄，且是美國家庭的機率樣本。(<http://dx.doi.org/10.3886/ICPSR02934>) 因此，在分析本資料檔的第一步應該是將資料加權以反映母群體傾向。
- **stocks.sav** 本假設資料檔包含一年的股票價格和數量。
- **stroke\_clean.sav**。本假設資料檔包含一個醫療資料庫，其在以「資料準備」選項中的程序清理之後的狀態。
- **stroke\_invalid.sav**。本假設資料檔包含一個醫療資料庫的起始狀態並包含幾個資料輸入錯誤。
- **stroke\_survival**。本假設資料檔是有關缺血性中風的病患，其在結束康復計畫後存活時間方面，面臨許多挑戰。中風後，記載了心肌梗塞、缺血性中風、或出血性中風的發生，以及事件記錄的時間。由於它只包含在康復計劃所管制的中風存活的病患，此樣本的左側被截斷。
- **stroke\_valid.sav**。本假設資料檔包含一個醫療資料庫，在其值以「驗證資料」程序檢查之後的狀態。它仍包含可能的異常觀察值。
- **survey\_sample.sav**。本資料檔包含調查資料，包括人口資料和各種態度測量。雖然已修改一些資料數值，且為人口資料之目的新增了一些額外的虛構變數，但是資料仍是以「1998 NORC 基本社會調查」的變數子集為基礎。
- **telco.sav**。這是有關一家電信公司致力於在客戶庫中減少顧客不忠的假設資料檔。每一個觀察值對應至一位不同的客戶並記錄不同的人口資料和服務使用方式資訊。
- **telco\_extra.sav**。本資料檔類似於 telco.sav 資料檔，但「任期」的對數轉換客戶花費變數已予刪除，並更換為標準的對數轉換客戶花費變數。
- **telco\_missing.sav**。本資料檔是 telco.sav 資料檔的子集，不過某些人口資料值已更換為遺漏值。

- **testmarket.sav**。本假設資料檔有關於一家速食連鎖店計劃在菜單中加入新的項目。有三個可能的活動來促銷此新產品，所以該新項目在幾個隨機選取市場中的地點作介紹。在每一個地點使用不同的促銷，並記錄該新項目前四週的每週銷售量。每一個觀察值對應至一個不同的地點-週。
- **testmarket\_1month.sav**。本假設資料檔是將 testmarket.sav 資料檔和每週購買的「彙總」，因此每一個觀察值對應至一個不同的客戶。結果部份每週變更的變數消失了，而目前所記錄的銷售量是在研究的四週期間銷售量之總和。
- **tree\_car.sav**。這是包含人口資料和車輛購買價格資料的假設資料檔。
- **tree\_credit.sav**。這是包含人口資料和銀行放款歷史資料的假設資料檔。
- **tree\_missing\_data.sav** 這是包含有大量遺漏值的人口資料和銀行放款歷史資料的假設資料檔。
- **tree\_score\_car.sav**。這是包含人口資料和車輛購買價格資料的假設資料檔。
- **tree\_textdata.sav**。一個只有兩個變數的簡單資料檔，主要目的在顯示變數預設狀態（在指定量測水準和數值標記之前）。
- **tv-survey.sav**。這是有關一家電視製片廠考量是否要延長一個成功節目的播送所作之調查的假設資料檔。有 906 位應答者被問到在不同的狀況下他們是否願意觀看這個節目。每一列代表一個不同的應答者；每一行為一個不同的狀況。
- **ulcer\_recurrence.sav**。本檔案包含一項用來比較兩種防止潰瘍復發治療法功效之研究的部分資訊。它是很好的區間受限資料範例，且已在別處 (Collett, 2003) 出現和分析過。
- **ulcer\_recurrence\_recoded.sav**。本檔案是將 ulcer\_recurrence.sav 的資訊重新組織，以讓您為此研究的每一個區間事件機率而非只是研究目的事件機率建立模式。它已在別處 (Collett et al., 2003) 出現和分析過。
- **verd1985.sav**。本資料檔有關於一項調查 (Verdegaal, 1985)。在調查中記錄了來自 15 個受訪者對 8 個變數的回應。所需的變數被分成三組。集 1 包括 age 和 marital，集 2 包括 pet 和 news，集 3 包括 music 和 live。Pet 調整為多重名義量數，age 調整為次序量數，其他的變數調整為單一名義量數。
- **virus.sav**。這是有關一家網際網路服務提供者致力於在其網路上判斷病毒之影響的假設資料檔。他們在其網路上追蹤從發現病毒直到控制威脅的這段時間，被病毒感染之電子郵件的流量（約略）百分比。
- **wheeze\_steubenville.sav**。這是空氣污染對兒童健康之影響 (Ware, Dockery, Spiro III, Speizer, 和 Ferris Jr., 1984) 縱向研究的子集。本資料包含來自俄亥俄州 Steubenville, 年齡 7、8、9 和 10 歲兒童的氣喘聲狀態之重複二元測量，以及其母親在本研究的第一年是否抽煙的固定記錄。
- **workprog.sav**。這是有關一項政府職業計劃，設法將弱勢民眾安置到較好之工作的假設資料檔。一個樣本的可能計劃參與者被追蹤，他們之中某些被選取加入本計劃，而其他的則否。每一個觀察值代表一位不同的計劃參與者。
- **worldsales.sav** 本假設資料檔包含依洲和產品分類之銷貨收益。

# 注意事項

本資訊適用於全球提供之產品與服務。

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

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

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U. S. A.

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

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

**下列條款不適用於英國，或其他任何當地法律規定與下列條款不一致的國家：**

INTERNATIONAL BUSINESS MACHINES 聲明係以「現狀」提供，沒有任何保固；不作任何明示或默示的保證，包括但不限於不侵權、適銷性或適合某一特定用途之保證。某些州不允許特定交易中明示或默示的保固聲明，因此，此聲明或許對您不適用。

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

此資訊中對於非 IBM 網站之任何參考僅為查閱方便而設，而且在任何情況中均不得作為那些網站之背書。該「網站」的內容並非此 IBM 產品的部分內容，使用該「網站」需自行承擔風險。

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

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

IBM Software Group, 收件人：Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA。

在適當條款與條件之下，包括某些情況下（支付費用），或可使用此類資訊。

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

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

此資訊包含用於日常企業運作的資料和報表範例。為了儘可能提供完整說明，範例中包含了人名、公司名稱、品牌名稱和產品名稱。所有的名稱皆為虛構，使用的名稱或地址和實際的企業如有雷同純屬巧合。

如果您正在螢幕上檢視此資訊，則圖片和彩色說明可能不會顯示。

## 商標

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

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

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

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

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

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

UNIX 為 The Open Group 於美國和其他國家的註冊商標。

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

其他產品和服務名稱可能為 IBM 或其他公司的商標。

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

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





---

# 參考書目

- Bell, E. H. 1961. Social foundations of human behavior: (人類行為之社會基礎) Introduction to the study of sociology (社會學研究簡介). 紐約: Harper & Row.
- Blake, C. L., 和 C. J. Merz. 1998. "UCI Repository of machine learning databases (機器學習資料庫 UCI 儲存器)." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Breiman, L., 和 J. H. Friedman 檢定(F). 1985. Estimating optimal transformations for multiple regression and correlation (估計多重迴歸與相關之最適轉換). Journal of the American Statistical Association (美國統計協會彙報), 80, .
- Collett, D. 2003. Modelling survival data in medical research (模式化醫學研究中的存活資料), 2 ed. Boca Raton: Chapman & Hall/CRC.
- Gardner, E. S. 1985. Exponential smoothing: (指數平滑化)The state of the art (最新技術). Journal of Forecasting (預測雜誌), 4, .
- Green, P. E., 和 V. Rao. 1972. Applied multidimensional scaling (應用多元尺度方法). Hinsdale, Ill.: Dryden Press (Dryden 出版社).
- Green, P. E., 和 Y. Wind. 1973. Multiattribute decisions in marketing: (市場行銷之多重屬性決策)A measurement approach (測量方法). Hinsdale, Ill.: Dryden Press (Dryden 出版社).
- Guttman 值, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points (尋找組態點最小座標空間之一般非計量技巧). Psychometrika (心理學計量報導), 33, .
- Hartigan, J. A. 1975. Clustering algorithms (集群演算法). 紐約: John Wiley and Sons.
- Hastie, T., 和 R. Tibshirani. 1990. Generalized additive models (概化附加模式). 倫敦: Chapman and Hall.
- Kennedy, R., C. Riquier, 和 B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research (市場研究類別資料之對應分析實際應用). Journal of Targeting, Measurement, and Analysis for Marketing (市場行銷之目標訂定、測量與分析雜誌), 5, .
- McCullagh, P., 和 J. A. Nelder. 1989. 概化線性模式, 第二版 ed. 倫敦: Chapman & Hall.
- Pena, D., G. C. Tiao, 和 R. S. Tsay, eds. 2001. A course in time series analysis (時間數列分析課程). 紐約: John Wiley and Sons.
- Price, R. H., 和 D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior (行為適切性與情境限制作為社會行為維度). Journal of Personality and Social Psychology (人格與社會心理學雜誌), 30, .
- Rickman, R., N. Mitchell, J. Dingman, 和 J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet (實行 Stillman 飲食法期間血膽固醇改變情形). Journal of the American Medical Association (美國醫學協會彙報), 228, .
- Rosenberg, S., 和 M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research (多變量研究中作為資料收集程序之排序方法). Multivariate Behavioral Research (多變量行為研究), 10, .

## 參考書目

- Van der Ham, T., J. J. Meulman, D. C. Van Strien, 和 H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: (依經驗法將青少年飲食異常次組別化) A longitudinal perspective (縱向觀點). *British Journal of Psychiatry* (英國心理學雜誌), 170, .
- Verdegaal, R. 1985. Meer sets analyse voor kwalitatieve gegevens (in Dutch) (更多性質資料的集合分析 (荷蘭文)). Leiden (萊頓): Department of Data Theory, University of Leiden (萊頓大學資料理論系).
- Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, 和 B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities (六個城市中的兒童二手煙、氣體煮食與呼吸道健康). *American Review of Respiratory Diseases* (美國呼吸道疾病評論), 129, .
- 盒形圖, G. E. P., G. M. Jenkins, 和 G. C. Reinsel. 1994. *Time series analysis: (時間數列分析) Forecasting and control (預測和控制)*, 第三版 ed. Englewood Cliffs, N.J. (紐澤西州鷹木崖): Prentice Hall.

- ACF
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
  - 純 ARIMA 處理的圖形, 90
- ARIMA 模式, 5
  - 偏離值, 13
  - 季節性階數, 10
  - 差分階數, 10
  - 常數, 10
  - 移動平均階數, 10
  - 自我迴歸階數, 10
  - 轉換函數, 11
- ARIMA 模式參數表
  - 在「時間數列 Modeler」中, 66
- Box-Ljung 統計量
  - 在「套用時間數列模式」中, 26
  - 在「時間數列 Modeler」中, 14, 66
- Brown's 指數平滑化模式, 8
- Expert Modeler, 5, 41
  - 偏離值, 7, 61
  - 限制模式空間, 6, 44
- Holt's 指數平滑化模式, 8
- MAE, 88
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
- MAPE, 88
  - 在「套用時間數列模式」中, 26, 28, 56
  - 在「時間數列 Modeler」中, 14, 16, 48
- MaxAE, 88
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
- MaxAPE, 88
  - 在「套用時間數列模式」中, 26, 28, 56
  - 在「時間數列 Modeler」中, 14, 16, 48
- PACF
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
  - 純 ARIMA 處理的圖形, 90
- R<sup>2</sup> 值, 88
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
- RMSE, 88
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
- Winters' 指數平滑化模式
  - 可加性, 8
  - 相乘性, 8
- XML
  - 以 XML 儲存時間數列模式, 19
  - 儲存重新估計模式為 XML, 31
- XML(X)
  - 以 XML 儲存時間數列模式, 46, 62
- 事件, 6
  - 在「時間數列 Modeler」中, 6
- 估計期間, 2
  - 在「套用時間數列模式」中, 24
  - 在「時間數列 Modeler」中, 5, 43
- 保留觀察值, 2
- 信賴區間
  - 在「套用時間數列模式」中, 28, 32
  - 在「時間數列 Modeler」中, 16, 20
- 偏自我相關函數
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
  - 純 ARIMA 處理的圖形, 90
- 偏離值
  - ARIMA 模式, 13
  - Expert Modeler, 7, 61
  - 定義, 89
- 儲存
  - XML 格式的模式規格, 19
  - 新變數名稱, 19, 31
  - 模式預測, 19, 31
  - 重新估計模式為 XML, 31
- 光譜圖, 37, 39
  - 假設, 37
  - 光譜密度, 85
  - 光譜視窗, 37
  - 相關程序, 87
  - 置中轉換, 38
  - 週期值, 85
  - 雙變數光譜分析, 38
- 創新性偏離值, 89
  - 在「時間數列 Modeler」中, 7, 13
- 可加性偏離值, 89
  - 在「時間數列 Modeler」中, 7, 13
- 可加性補充偏離值, 89
  - 在「時間數列 Modeler」中, 7, 13
- 商標, 102
- 套用時間數列模式, 23, 52, 67
  - Box-Ljung 統計量, 26
  - 估計期間, 24
  - 信賴區間, 28, 32
  - 儲存重新估計模式為 XML, 31
  - 儲存預測值, 31, 54
  - 所有模式內的統計量, 26, 28, 55
  - 新變數名稱, 31, 56
  - 最適及最不適模式, 29
  - 模式參數, 26
  - 模式適合度表格, 55
  - 殘差偏自我相關函數, 26, 28
  - 殘差自我相關函數, 26, 28
  - 適合度值, 28

## 索引

- 適合度統計量, 26, 28, 55
- 遺漏值, 32
- 重新估計模式參數, 24, 53
- 預測, 26, 28, 75
- 預測期間, 24, 53, 74
- 預測表, 75
  
- 季節可加性偏離值, 89
  - 在「時間數列 Modeler」中, 7, 13
- 季節性差分轉換
  - ARIMA 模式, 10
- 季節性階數
  - ARIMA 模式, 10
  
- 對數轉換
  - 在「時間數列 Modeler」中, 8, 10 - 11
  
- 局部趨勢偏離值, 89
  - 在「時間數列 Modeler」中, 7, 13
  
- 差分轉換
  - ARIMA 模式, 10
  
- 平均絕對百分比誤差, 88
  - 在「套用時間數列模式」中, 26, 28, 56
  - 在「時間數列 Modeler」中, 14, 16, 48
- 平均絕對誤差, 88
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
- 平方根轉換
  - 在「時間數列 Modeler」中, 8, 10 - 11
- 平穩  $R^2$  值, 88
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16, 66
  
- 指數平滑化模式, 5, 8
  
- 整合
  - ARIMA 模式, 10
  
- 時間數列 Modeler, 3
  - ARIMA 模式參數表, 66
  - ARIMA 程序, 5, 10
  - Box-Ljung 統計量, 14
  - Expert Modeler, 5, 41, 57
  - 事件, 6
  - 以 XML 儲存模式規格, 19, 46, 62
  - 估計期間, 5, 43
  - 信賴區間, 16, 20
  - 偏離值, 7, 13, 61
  - 儲存預測值, 19, 46
  - 所有模式內的統計量, 14, 16, 47, 49
  - 指數平滑化, 5, 8
  - 數列轉換, 8, 10 - 11
  - 新變數名稱, 19, 50
  - 最適及最不適模式, 18
  - 模式參數, 14, 63
  - 模式名稱, 20
  - 模式統計量表, 66
  - 模式說明表, 65
  - 殘差偏自我相關函數, 14, 16
  - 殘差自我相關函數, 14, 16
  - 轉換函數, 11
  - 週期, 6, 8, 10 - 11
  - 適合度值, 16, 64
  - 適合度統計量, 14, 16, 47, 66
  - 遺漏值, 20
  - 預測, 14, 16, 47
  - 預測期間, 5, 20, 43, 45
  - 預測表, 51
  
- 最大絕對百分比誤差, 88
  - 在「套用時間數列模式」中, 26, 28, 56
  - 在「時間數列 Modeler」中, 14, 16, 48
- 最大絕對誤差, 88
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
  
- 有效期限, 2
  
- 根平均平方誤, 88
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
- 模式
  - ARIMA 程序, 5, 10
  - Expert Modeler, 5
  - 指數平滑化, 5, 8
- 模式參數
  - 在「套用時間數列模式」中, 26
  - 在「時間數列 Modeler」中, 14, 63
- 模式名稱
  - 在「時間數列 Modeler」中, 20
- 模式統計量表
  - 在「時間數列 Modeler」中, 66
- 模式說明表
  - 在「時間數列 Modeler」中, 65
- 模式適合度表格
  - 在「套用時間數列模式」中, 55
  
- 正規化 BIC (Bayesian 資訊準則), 88
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
- 歷史資料
  - 在「套用時間數列模式」中, 28
  - 在「時間數列 Modeler」中, 16
  
- 殘差
  - 在「套用時間數列模式」中, 26, 28

- 在「時間數列 Modeler」中, 14, 16
- 水準變遷偏離值, 89
  - 在「時間數列 Modeler」中, 7, 13
- 法律注意事項, 101
- 減幅指數平滑化模式, 8
- 瞬變偏離值, 89
  - 在「時間數列 Modeler」中, 7, 13
- 移動平均
  - ARIMA 模式, 10
- 範例檔案
  - 位置, 94
- 簡單季節性指數平滑化模式, 8
- 簡單指數平滑化模式, 8
- 自我相關函數
  - 在「套用時間數列模式」中, 26, 28
  - 在「時間數列 Modeler」中, 14, 16
  - 純 ARIMA 處理的圖形, 90
- 自我迴歸
  - ARIMA 模式, 10
- 自然對數轉換
  - 在「時間數列 Modeler」中, 8, 10 - 11
- 調和分析, 37
- 變數名稱
  - 在「套用時間數列模式」中, 31
  - 在「時間數列 Modeler」中, 19
- 轉換函數, 11
  - 分子階數, 11
  - 分母階數, 11
  - 季節性階數, 11
  - 差分階數, 11
  - 延遲, 11
- 週期
  - 在「時間數列 Modeler」中, 6, 8, 10 - 11
- 週期性分解, 34 - 36
  - 假設, 34
  - 儲存新變數, 35
  - 建立變數, 35
  - 新變數, 81
  - 模式, 34
  - 相關程序, 83
  - 計算移動平均, 34
  - 週期性日期成份, 77
- 過去期間, 2
- 適合度
  - 在「套用時間數列模式」中, 26, 28, 55
  - 在「時間數列 Modeler」中, 14, 16, 47
  - 定義, 88
- 適合度值
  - 在「套用時間數列模式」中, 28
  - 在「時間數列 Modeler」中, 16, 64
- 遺漏值
  - 在「套用時間數列模式」中, 32
  - 在「時間數列 Modeler」中, 20
- 重新估計模式參數
  - 在「套用時間數列模式」中, 24, 53
- 預測
  - 在「套用時間數列模式」中, 26, 28, 75
  - 在「時間數列 Modeler」中, 14, 16, 47
- 預測期間
  - 在「套用時間數列模式」中, 24, 53, 74
  - 在「時間數列 Modeler」中, 5, 20, 43, 45
- 預測表
  - 在「套用時間數列模式」中, 75
  - 在「時間數列 Modeler」中, 51