

IBM SPSS Forecasting 21



注：この情報とサポートされている製品をご使用になる前に、「注意事項」（p.124）の一般情報をお読みください。

本版は IBM® SPSS® Statistics 21 ,および新版で指示されるまで後続するすべてのリリースおよび変更に対して適用されます。

Adobe 製品のスクリーンショットは Adobe Systems Incorporated の許可を得て転載しています。

Microsoft 製品のスクリーンショットは Microsoft 社の許可を得て転載しています。

Licensed Materials - Property of IBM

© Copyright IBM Corporation 1989, 2012.

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

はじめに

IBM® SPSS® Statistics は、データ分析の包括的システムです。見込みは、このマニュアルで説明されている追加の分析手法を提供するオプションのアドオン モジュールです。見込み アドオン モジュールは SPSS Statistics Core システムと組み合わせて使用し、Core システムに完全に統合されます。

IBM Business Analytics について

IBM Business Analytics ソフトウェアは、意思決定者がビジネス パフォーマンスを向上させるために信頼する完全で、一貫した正確な情報を提供します。ビジネス インテリジェンス、予測分析、財務実績および戦略管理、および 分析アプリケーションの包括的なポートフォリオを利用することによって、現在の実績を明確、迅速に理解し、将来の結果を予測することができます。豊富な業界のソリューション、実績ある実例、専門サービスと組み合わせ、さまざまな規模の組織が、高い生産性を実現、意思決定を自信を持って自動化し、より良い決定をもたらします。

このポートフォリオの一部として、IBM SPSS Predictive Analytics ソフトウェアを使用する組織は、将来のイベントを予測し、その洞察に基づいて積極的に行動し、より優れた業績を実現することができます。全世界の企業、政府、学術分野のお客様が IBM SPSS の技術を活用し、不正行為を減少させ、リスクを軽減させながら、顧客の獲得、保持、成長において、競争優位を高めることができます。IBM SPSS ソフトウェアを日々の業務に取り入れることによって、組織は業務目標を達成し、大きな競争的優位を獲得することができるよう、意思決定を方向付け、自動化することができるようになります。お問い合わせは、<http://www.ibm.com/spss> を参照してください。

テクニカル サポート

テクニカル サポートのサービスをご利用いただけます。IBM Corp. 製品の使用方法や、対応しているハードウェア環境へのインストールに関して問い合わせることもできます。テクニカル サポートの詳細については、IBM Corp. Web サイト (<http://www.ibm.com/support>) を参照してください。連絡の際は、所属団体名、サポート契約などを確認できるよう、あらかじめ手元にご用意ください。

学生向けテクニカル サポート

IBM SPSS ソフトウェア製品の Student 版、アカデミック版、Grad パック版を使用している学生の場合、学生用の特別オンライン ページ、[Solutions for Education \(http://www.ibm.com/spss/rd/students/\)](http://www.ibm.com/spss/rd/students/) ページを参照してください。大学提供の IBM SPSS ソフトウェアのコピーを使用している場合、大学の IBM SPSS 製品コーディネータにお問い合わせください。

カスタマ サービス

配送やアカウントに関するご質問は、お近くの営業所にお問い合わせください。お問い合わせの際には、シリアル番号をご用意ください。

トレーニング セミナー

IBM Corp. では一般公開およびオンサイトで トレーニング セミナーを実施しています。セミナーでは実践的な講習を行います。セミナーは主要都市で定期的に行われます。セミナーに関する詳細については、<http://www.ibm.com/software/analytics/spss/training> を参照してください。

内容

パート I: ユーザー ガイド

1	時系列の概要	1
	時系列データ	1
	データの変換	2
	推定期間と妥当性期間	2
	モデルの構築と予測の実行	3
2	時系列モデラー	4
	エキスパート モデラーのオプションの指定	8
	モデルの選択とイベントの指定	9
	エキスパート モデラーで外れ値を処理する	10
	ユーザー指定による指数平滑法モデル	11
	ユーザー指定の ARIMA モデル	13
	ユーザー指定の ARIMA モデルに対するモデル指定	14
	ユーザー指定の ARIMA モデルにおける伝達関数	16
	ユーザー指定の ARIMA モデルの外れ値	18
	出力	19
	統計表と予測表	20
	作図	22
	出力を最適モデルまたは最不適合モデルに制限する	24
	モデルの予測とモデルの指定を保存する	25
	オプション	26
	TSMODEL コマンドの追加機能	28
3	時系列モデルの適用	29
	出力	32
	統計表と予測表	33
	作図	35
	出力を最適モデルまたは最不適合モデルに制限する	37
	モデルの予測とモデルの指定を保存する	38

オプション	40
TSAPPLY コマンドの追加機能	41
4 [季節性の分解]	42
季節性の分解の保存	43
SEASON コマンドの追加機能	44
5 スペクトルプロット	45
SPECTRA コマンドの追加機能	48
パート II: 例	
6 エキスパート モデラーによるバルク予測	50
データの検証	50
分析の実行	52
モデルの要約図	59
モデルの予測	60
要約	62
7 保存したモデルの適用によるバルク再予測	63
分析の実行	63
モデルの適合度統計量	66
モデルの予測	67
要約	68
8 エキスパート モデラーによる有意な予測変数の特定	69
データのプロット	69

分析の実行	71
系列プロット	77
モデルの説明表	78
モデル統計量の表	78
ARIMA モデルのパラメータ表	79
要約	79
9 保存したモデルの適用による予測変数の検証	80
予測系列の拡張	80
予測期間内の予測値の修正	84
分析の実行	86
10 季節性の分解	91
売上データからの季節性の除去	91
周期性の調査と設定	91
分析の実行	96
出力について	97
要約	99
関連手続き	99
11 スペクトルプロット	100
スペクトルプロットを使用した周期性に関する期待値の確認	100
分析の実行	100
ペリオドグラムとスペクトル密度について	102
要約	104
関連手続き	104

付録

A	適合度統計量	105
B	外れ値の種類	107
C	ACF/PACF プロットについて	108
D	サンプル ファイル	113
E	注意事項	124
	参考文献	127
	索引	129

パート I: ユーザー ガイド

時系列の概要

時系列は、一定期間にわたって定期的に 1 つの変数を測定することによって得られる、一連の観測値です。たとえば、供給量の時系列データでは、観測値が、数か月間にわたる 1 日ごとの供給レベルを表すことがあります。ある製品の市場占有率を示す時系列が、数年間の 1 週間ごとの市場占有率で構成されることもあります。総売上高の時系列が、長年の間の 1 か月ごとの観測値で構成されることもあります。これらの例に共通することは、特定の期間にわたって、定期的な既知の間隔で、ある変数が観測されたことです。したがって、一般的な時系列のデータの形式は、1 つの時系列、または定期的な間隔で測定された結果を表す観測値のリストになります。

テーブル 1-1
1 日ごとの供給の時系列

時刻 (M)	週	日	供給レベル
t ₁	1	MONDAY	160
t ₂	1	TUESDAY	135
t ₃	1	WEDNESDAY	129
t ₄	1	THURSDAY	122
t ₅	1	FRIDAY	108
t ₆	2	MONDAY	150
		...	
t ₆₀	12	FRIDAY	120

時系列を分析する最も重要な目的は、時系列における将来の値を予測することにあります。過去の値を説明する時系列モデルは、今後の値がどの程度増加するのか、またはどの程度減少するのかの予測にも使えます。このような予測を正しく行う能力は、ビジネスまたは科学の分野において明らかに重要です。

時系列データ

Forecasting アドオン モジュールで使用するために時系列データを定義する場合、各系列は個別の変数に対応します。たとえば、データ エディタで時系列を定義するには、[変数ビュー] タブをクリックし、任意の空白行に変数名を入力します。時系列内の各観測はケース（データ エディタ内の行）に対応します。

時系列データが含まれたスプレッドシートを開くと、各系列はそのスプレッドシートの列に配置されます。すでに時系列が行に配置されたスプレッドシートを保持している場合は、そのスプレッドシートを開き、

[データ] メニューの [行と列の入れ換え] を使用して、行と列を入れ換えることができます。

データの変換

Core システムで提供されている多くのデータ変換手続きは、時系列分析に役立ちます。

- [日付の定義] 手続き ([データ] メニュー) では、周期の設定や、履歴期間、妥当性期間、および予測期間の区別を使用する日付変数を生成できます。Forecasting は、[日付の定義] 手続きにより作成された変数を操作できるように設計されています。
- [時系列の作成] 手続き ([変換] メニュー) では、新しい時系列変数を既存の時系列変数の関数として作成できます。作成された時系列変数には、平滑化、平均化、および差分に隣接観測を使用する関数が含まれています。
- [欠損値の置き換え] 手続き ([変換] メニュー) では、システム欠損値とユーザー欠損値が、いくつかある方法のいずれか 1 つに基づいた推定値に置き換えられます。系列の先頭または最終点の欠損データにより特定の問題が発生することはありません。これらの欠損データにより、使用できる系列の長さが短くなるだけです。系列の途中にあるギャップ (埋め込まれた欠損データ) は、より深刻な問題となる可能性があります。

時系列のデータ変換の詳細は、『Core System User's Guide』を参照してください。

推定期間と妥当性期間

多くの場合、時系列を履歴期間 (または推定期間) と妥当性期間に分けることは有効です。推定 (履歴) 期間の観測に基づいてモデルを構築した上で、妥当性期間においてそのモデルがどの程度有効に機能するかについてのテストを行います。すでに結果がわかっている点 (妥当性期間内の点) に対してモデルを適用して予測を行うことにより、モデルによる予測の有効度を知ることができます。

妥当性期間のケースは、モデルを構築するプロセスから除外されているため、通常ホールドアウト ケースと呼ばれています。推定期間は、アクティブなデータセットで現在選択しているケースから構成されます。選択された最後尾のケースより後にある残りのケースはすべて、ホールドアウト ケースとして使用できます。モデルによる予測の有効性を確認した後で、推定期間を再定義してホールドアウト ケースを取り込み、最終モデルを構築することができます。

モデルの構築と予測の実行

予測アドオン モジュールには、モデルの作成と予測の実行の 2 つの作業を行うための手続きが用意されています。

- **時系列モデラー**手続きでは、時系列に対するモデルを作成し、予測を実行することができます。この手続きには、時系列のそれぞれに対して最適なモデルを自動的に特定するエキスパート モデラーを使用します。また、より詳細な制御を必要とする分析経験豊富なユーザーに対しては、ユーザー指定によるモデル構築を行うためのツールが用意されています。
- **時系列モデルの適用**手続きでは、既存の時系列モデル（時系列モデラーで作成）をアクティブなデータセットに適用できます。これにより、モデルを再構築することなく、新しいデータや変更されたデータに対して系列の予測を行うことができます。モデルが変更されたと判断する根拠がある場合は、時系列モデラーを使用してモデルを再構築することができます。

時系列モデラー

[時系列モデラー] 手続きでは、時系列に対する指数平滑法モデル、1 変量 ARIMA (Autoregressive Integrated Moving Average (自己回帰和分移動平均)) モデル、および多変量 ARIMA モデル (または伝達関数モデル) の推定が行われ、予測が生成されます。この手続きでは、1 つ以上の従属変数に対して最適な ARIMA または指数平滑モデルを自動的に特定および推定する Expert Modeler を含むことにより、試行錯誤によって適切なモデルを特定する必要性を排除します。また、カスタム ARIMA または指数平滑モデルを指定することもできます。

例: 製品マネージャとして、100 品目の支店取扱い製品のそれぞれに関して来月の売上高と収益を予測する責任がありますが、時系列のモデリングの経験はほとんどありません。100 品目の支店売上履歴データは、1 枚の Excel スプレッドシートにまとめて保存されています。スプレッドシートを IBM® SPSS® Statistics で開いた後、エキスパート モデラーを使用して今後 1 か月の予測を要求します。エキスパート モデラーは、品目ごとに支店売上の最善のモデルを探し出し、それらのモデルを使用して予測を生成します。エキスパート モデラーでは、複数の入力系列を処理できるため、手続きを一度実行するだけで、全品目の予測が取得できます。予測した結果は、アクティブなデータセットに予測を保存することを選択すれば、元の Excel に簡単にエクスポートできます。

統計量。 適合度: 定常 R 2 乗、R 2 乗 (R^2)、平均 2 乗誤差平方根 (RMSE)、平均絶対誤差 (MAE)、平均絶対パーセント誤差 (MAPE)、最大絶対誤差 (MaxAE)、最大絶対パーセント誤差 (MaxAPE)、正規化されたベイズ情報量基準 (BIC)。残差: 自己相関関数、偏自己相関関数、Ljung-Box Q。ARRIMA モデルの場合: 従属変数に対する ARIMA の次数、独立変数に対する伝達関数の次数、および外れ値の推定値。さらに、指数平滑法モデル用の平滑法パラメータ推定値。

作図。 すべてのモデルの要約図: 定常 R 2 乗、R 2 乗 (R^2)、平均 2 乗誤差平方根 (RMSE)、平均絶対誤差 (MAE)、平均絶対パーセント誤差 (MAPE)、最大絶対誤差 (MaxAE)、最大絶対パーセント誤差 (MaxAPE)、および正規化されたベイズ情報量基準 (BIC) のヒストグラム。残差自己相関および偏自己相関の箱ひげ図。個々の値の結果: 予測値、適合値、観測値、信頼区間の上限と下限、残差自己相関関数と偏自己相関関数。

時系列モデラー データの考慮事項

データ。 従属変数とすべての独立変数は数値型である必要があります。

仮定。 従属変数とすべての独立変数は時系列として扱われます。つまり、各ケースは時点を表し、連続するケースは一定の時間間隔で区切られています。

- **定常性。** ユーザーの指定による ARIMA モデルでは、モデリングする時系列は定常的である必要があります。非定常時系列から定常時系列への変換で最も効率的なのは、差分変換です。差分変換は、[時系列の作成] ダイアログ ボックスから実行できます。
- **予測。** 独立 (予測) 変数を持つモデルを使用して予測を作成するためには、アクティブなデータセットには予測期間内のすべてのケースについてそれらの変数の値が含まれている必要があります。加えて、推定期間内では独立変数に欠損値が含まれていないことが必要です。

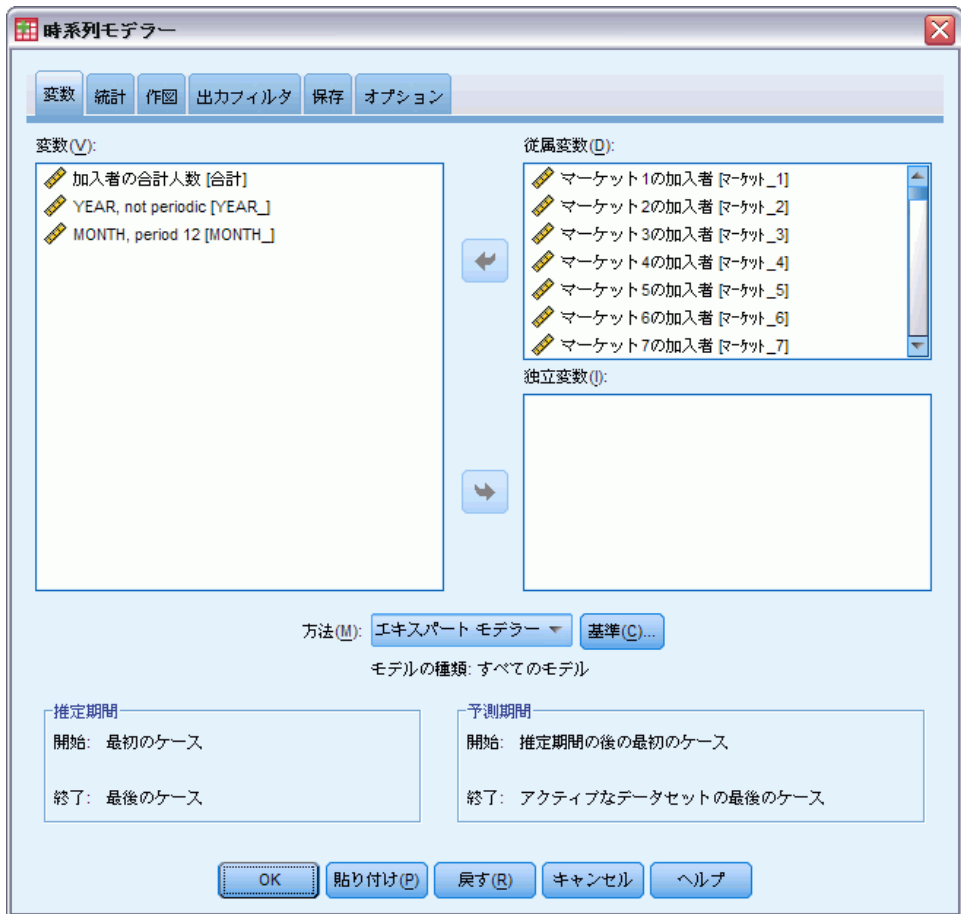
日付の定義

最初のケースに関連する日付、および連続するケース間の時間間隔を指定する場合は、[日付の定義] ダイアログ ボックスを使用することをお勧めします。ただし必須ではありません。この操作は時系列モデラーを使用する前に行われ、その結果、各ケースに関連する日付にラベル付けする変数のグループが生成されます。また、データの想定周期も設定されます。たとえば、連続ケース間の時間間隔が 1 か月なら 12 という周期になります。この周期は、季節モデルの作成に関心がある場合に必要です。季節モデルの作成には関心がなく、出力に日付ラベルも必要ない場合は、[日付の定義] ダイアログ ボックスは使用する必要がありません。その場合、各ケースに関連付けるラベルは単純にケース番号になります。

時系列モデラーを使用するには

- ▶ メニューから次の項目を選択します。
分析 > 見込み > モデルの作成(C)...

図 2-1
時系列モデラーの [変数] タブ



- ▶ [変数] タブでは、モデリングする従属変数を 1 つ以上選択します。
- ▶ [方法] ドロップダウン ボックスから、モデリング方法を選択します。自動モデリングの場合は、デフォルトの方法の「エキスパート モデラー」のままにします。これにより、エキスパート モデラーが起動され、各従属変数に対して最適となるモデルが決定されます。

予測を作成するには、次の手順に従ってください。

- ▶ [オプション] タブをクリックします。
- ▶ 予測期間を指定します。これにより、予測値と観察値を含む図表が作成されます。

オプションとして、次の選択が可能です。

- 1 つ以上の独立変数を選択します。独立変数は、回帰分析の予測変数とよく似た扱われ方をしますが、省略可能です。ARIMA モードに含めることができますが、指数平滑法モデルには含めることができません。モデリング手法として [エキスパート モデラー] を指定し、独立変数を含める場合には、ARIMA モードのみが考慮されます。
- モデリングの詳細を指定するために、[基準] をクリックします。
- 予測、信頼区間、およびノイズ残差を保存します。
- 推定モデルを XML 形式で保存します。モデルを再構築することなく更新された予測を得るために、保存したモデルを新しいデータまたは変更されたデータに適用することができます。この操作は、時系列モデルの適用 手続きを使用して行えます。
- すべての推定モデルから、要約統計量を取得します。
- ユーザーの指定による ARIMA モデルの独立変数に、伝達関数を指定します。
- 外れ値の自動検出を有効にします。
- 特定の時点を、ユーザーの指定による ARIMA モデルの外れ値としてモデリングします。

モデリング方法

利用できるモデリング方法は次のとおりです。

エキスパート モデラー。エキスパート モデラーは、各従属系列に対して最適となるモデルを自動的に見つけます。独立（予測）変数が指定された場合は、従属系列と統計的に有意な関係を持つ独立変数を、ARIMA モデルへ取り込むために選択します。必要に応じて差分を取ることによって、また必要に応じて平方根変換または自然対数変換を使用して、モデル変数は変換されます。デフォルトでは、エキスパート モデラーは指数平滑法モデルと ARIMA モデルの両方を考慮します。ただし、ARIMA モデルのみを検索するか、指数平滑法モデルのみを検索するように、エキスパート モデラーに制限を加えることができます。また、外れ値の自動検出も指定できます。

指数平滑法。ユーザー指定の指数平滑法モデルを指定するためのオプションです。トレンドと季節性の扱い方が異なる各種の指数平滑法モデルから選択することができます。

ARIMA。ユーザー指定の ARIMA モデルを指定するためのオプションです。これを使用すると、差分の次数に加えて、自己回帰の次数と移動平均の次数も明示的に指定する必要があります。独立（予測）変数を含めて、その一部または全部に対する伝達関数を定義することができます。また、外れ値の自動検出を指定したり、外れ値のグループを明示的に指定したりもできます。

推定期間と予測期間

推定期間。 推定期間により、モデルを決定するためのケースのグループが定義されます。デフォルトでは、推定期間はアクティブなデータセットのすべてのケースを含みます。推定期間を設定するには、[ケースの選択] ダイアログ ボックスで、[日付かケース番号の範囲] を選択します。使用可能なデータによっては、この手続きによって使用される推定期間は従属変数ごとに異なる場合があるため、表示される値と異なる場合があります。指定された従属変数に対して、本当の推定期間は、指定された推定期間の先頭または末尾で連続する変数の欠損値を削除した後に残った期間です。

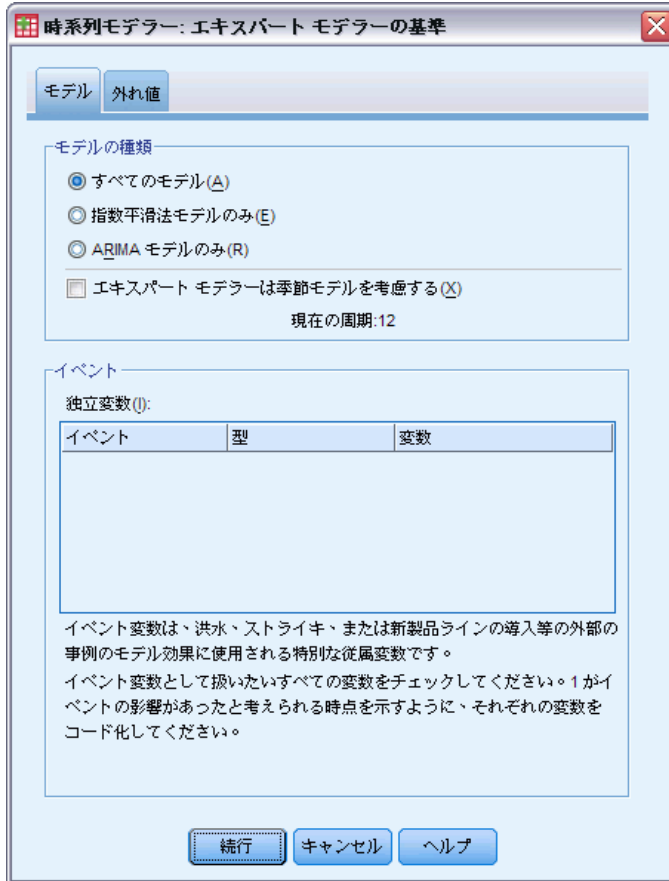
予測期間。 予測期間は、推定期間の後の最初のケースから始まり、デフォルトでは、アクティブなデータセット内の最後のケースまで続きます。予測期間の末尾は、[オプション] タブから設定できます。

エキスパート モデラーのオプションの指定

エキスパート モデラーは、候補モデルのグループに制約を加えたり、外れ値の処理を指定したり、イベント変数を含めたりするためのオプションを提供します。

モデルの選択とイベントの指定

図 2-2
[エキスパート モデラー基準] ダイアログ ボックスの [モデル] タブ



[モデル] タブでは、エキスパート モデラーからの考慮対象とするモデルの種類とイベント変数を指定できます。

モデルの種類。 次のオプションを使用できます。

- **すべてのモデル。** エキスパート モデラーは ARIMA モデルと指数平滑法モデルの両方を考慮します。
- **指数平滑法モデルのみ。** エキスパート モデラーは指数平滑法モデルのみを考慮します。
- **ARIMA モデルのみ。** エキスパート モデラーは ARIMA モデルのみを考慮します。

エキスパート モデラーは季節モデルを考慮する。 このオプションは、アクティブなデータセットに対して周期性が定義されている場合にのみ有効になります。このオプションを選択 (チェック) すると、エキスパート モデラーは

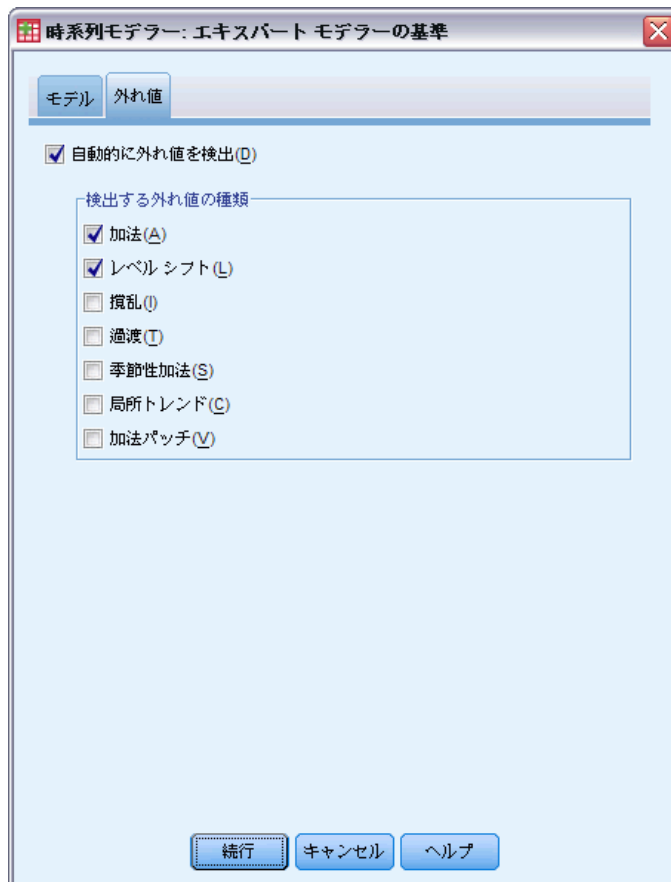
季節モデルと非季節モデルの両方を考慮します。このオプションを選択しないと、エキスパート モデラーは非季節モデルのみを考慮します。

現在の周期。 アクティブ データセットに対して現在定義されている周期（存在する場合）。現在の周期は、整数値として与えられます。たとえば、各ケースが 1 か月を表す年間の周期に対しては 12 となります。周期性が設定されていない場合は、「なし」と表示されます。季節も出るには周期が必要です。

イベント。 イベント変数として扱う独立変数を選択します。イベントの変数の値が 1 のケースは、従属系列がイベントの影響を受けると予想される時間を示します。1 以外のケースは、影響がないことを示します。

エキスパート モデラーで外れ値を処理する

図 2-3
[エキスパート モデラーの基準] ダイアログ ボックスの [外れ値] タブ



[外れ値] タブでは、外れ値の自動検出を選択できるほか、検出対象の外れ値の種類も選択できます。

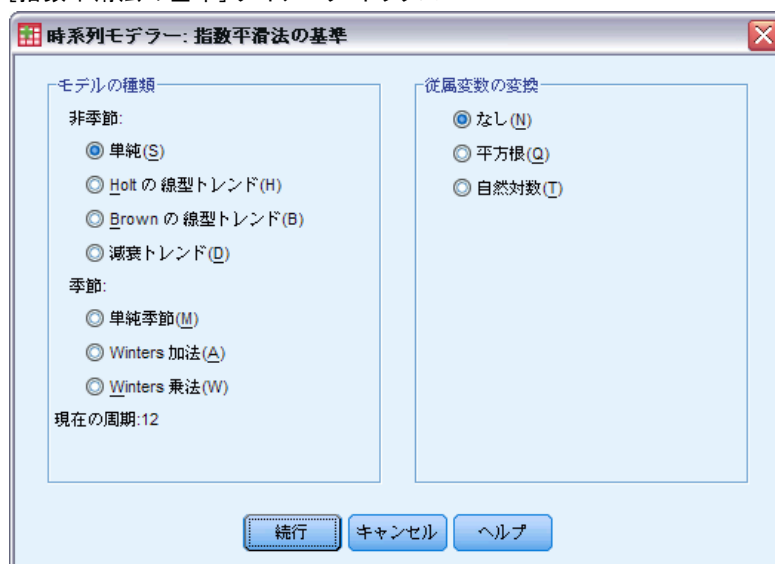
自動的に外れ値を検出。 デフォルトでは、外れ値の自動検出は実行されません。外れ値の自動検出を実行するには、このオプションを選択（チェック）し、以下の外れ値の種類を 1 つ以上選択します。

- 相加的
- レベル シフト (L)
- 攪乱 (I)
- 過渡 (T)
- 季節性加法 (S)
- 局所トレンド (C)
- 加法パッチ (P)

詳細は、B 付録 p.107 外れ値の種類 を参照してください。

ユーザー指定による指数平滑法モデル

図 2-4
[指数平滑法の基準] ダイアログ ボックス



モデルの種類: 指数平滑法モデル (Gardner, 1985) は、季節または非季節のいずれかに分類されます。季節モデルは、周期が作業データセットに対して定義されている場合にのみ使用可能です（以下の「現在の周期」を参照）。

- **単純.** このモデルはトレンドまたは季節性のない系列（時系列）に適しています。唯一の平滑化パラメータは水準です。単純指数平滑法は、0 次数の自己回帰、1 次数の差分、1 次数の移動平均、および定数無しの ARIMA モデルに類似しています。
- **Holt の線型トレンド.** このモデルは線型トレンドで季節性のない系列（時系列）に適しています。その平滑化パラメータは、水準パラメータとトレンドパラメータであり、互いの値には制約されません。Holt のモデルは Brown のモデルより普及していますが、大きな系列の計算には時間がかかります。Holt の減衰指数平滑法は、次数無しの自己回帰、2 次数の差分、および 2 次数の移動平均の ARIMA モデルに類似しています。
- **Brown の線型トレンド.** このモデルは線型トレンドで季節性のない系列（時系列）に適しています。その平滑化パラメータは、水準パラメータとトレンドパラメータであり等しいものと仮定されます。したがって、Brown モデルは、Holt モデルの特別な場合です。Brown の指数平滑法モデルは、次数無しの自己回帰と移動平均の次数 2、および、1 次移動平均の相関係数の半分に等しい 2 次移動平均の相関係数に関しては ARIMA モデルに類似しています。
- **減衰トレンド.** このモデルは減衰する線型トレンドで季節性のない系列（時系列）に適しています。その平滑化パラメータは、水準、トレンド、および減衰トレンドです。減衰指数平滑法は、1 次数の自己回帰、1 次数の差分、および 2 次数の移動平均の ARIMA モデルに類似しています。
- **単純季節.** このモデルは、トレンドがなく、時間に対して一定な季節効果を持つ系列に（時系列）に適しています。その平滑化パラメータは、水準と季節です。単純季節指数平滑化は、自己回帰が 0 次、差分が 1 次、季節差分が 1 次、そして季節区間（月データ： $p = 12$ ）の周期の数が p のとき、移動平均が 1 次、 p 次、および $p + 1$ 次の ARIMA モデルに最も良く似ています。
- **Winters 加法.** このモデルは、線型トレンドを持ち、系列の水準に依存しない季節効果のある系列（時系列）に適しています。その平滑化パラメータは、水準、トレンドおよび季節です。Winters 加法指数平滑化モデルは、自己回帰が 0 次、差分が 1 次、季節差分が 1 次、および季節区間（月データ： $p = 12$ ）の周期の数が p のとき、移動平均が $p + 1$ 次の ARIMA モデルに最も類似しています。
- **Winters 乗法.** このモデルは、線型トレンドを持ち、系列の水準に依存する季節効果のある系列（時系列）に適しています。その平滑化パラメータは、水準、トレンドおよび季節です。Winters の相乗指数平滑法は、いかなる ARIMA モデルにも類似しません。

現在の周期。 アクティブ データセットに対して現在定義されている周期（存在する場合）。現在の周期は、整数値として与えられます。たとえば、各ケースが 1 か月を表す年間の周期に対しては 12 となります。周期性が設定されていない場合は、「なし」と表示されます。季節も出るには周期が必要です。

従属変数の変換。 モデル作成する前に、各従属変数に対して実行する変換を指定することができます。

- **なし。**変換は実行されません。
- **平方根。**平方根変換。
- **自然対数。**自然対数変換。

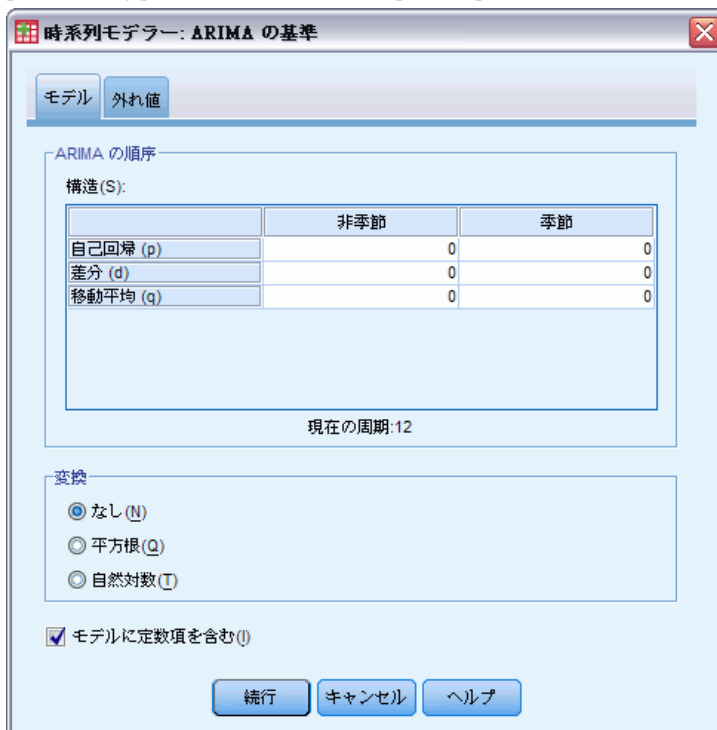
ユーザー指定の ARIMA モデル

時系列モデラーにより、固定された予測変数のグループの有無にかかわらず、ユーザー指定の非季節または季節 ARIMA (Autoregressive Integrated Moving Average) モデル (「Box-Jenkins」 (Box, Jenkins, および Reinsel, 1994) モデルともいいます) を構築することができます。予測変数の一部またはすべての伝達関数を定義し、外れ値の自動検出を指定したり外れ値のグループを明示的に指定したりすることができます。

- [変数] タブで指定されたすべての従属 (予測) 変数は、明示的にモデルに含まれます。これは、従属変数との間に統計的に重要な関係がある場合にのみ独立変数が含まれるエキスパート モデラーの使用と対照的です。

ユーザー指定の ARIMA モデルに対するモデル指定

図 2-5
[ARIMA 基準] ダイアログ ボックスの [モデル] タブ



[モデル] タブでは、ユーザー指定の ARIMA モデルの構造を指定することができます。

ARIMA の順序。さまざまな ARIMA モデルの成分の値を [構造] グリッド内の対応するセルに入力します。すべての値は負でない整数にする必要があります。自己回帰と移動平均の成分については、値は最大次数を表します。すべての正の低い次数はモデルに含まれます。たとえば、2 を指定すると、モデルには次数 2 および 1 が含まれます。[季節] 列のセルは、アクティブなデータセットに対して周期が定義されている場合にだけ有効になります（後述する「現在の周期」を参照）。

- **自己回帰 (p)。**モデル内の自己回帰の次数の数。自己回帰の次数は、系列の使用する過去の値を指定し、現在の値を予測します。たとえば、自己回帰の次数 2 は、現在の値を予測するために系列の値を過去の 2 期間使用するように指定します。
- **差分 (d)。**モデルを推定する前に、系列に適用する差分の次数を指定します。トレンドが存在する場合は差分を取る必要があります（トレンドの存在する系列は通常非定常性であり、ARIMA モデルは定常性を前提としている）、その効果を取り除くために行います。差分の次数は、系

列のトレンドの次数に対応しています（1 次差分は線型トレンドを表し、2 次差分は 2 次トレンドを表す、など）。

- **移動平均 (q)**。モデル内の移動平均の次数の数。移動平均の次数は、過去の値の系列平均の偏差が、現在の値を予測するためにどのように使用されるかを指定します。たとえば、移動平均の次数 1 および 2 は、系列の現在の値を予測する際に最近の 2 期間のそれぞれから取得した系列の平均値の偏差を考慮することを指定します。

季節次数。季節自己回帰、移動平均、および差分成分は、対応する非季節成分と同様の役割を果たします。ただし、季節次数については、現在の系列値は 1 つ以上の季節期間で区切られた過去の系列値に影響されます。たとえば、毎月のデータ（季節期間 12）については、季節次数 1 は、現在の系列値は現在の期間より 12 期間以前の系列値により影響されることを意味しています。毎月のデータについて、季節次数 1 は、非季節次数 12 を指定するのと同じこととなります。

現在の周期。アクティブ データセットに対して現在定義されている周期（存在する場合）。現在の周期は、整数値として与えられます。たとえば、各ケースが 1 か月を表す年間の周期に対しては 12 となります。周期性が設定されていない場合は、「なし」と表示されます。季節も出るには周期が必要です。

従属変数の変換。モデル作成する前に、各従属変数に対して実行する変換を指定することができます。

- **なし**。変換は実行されません。
- **平方根**。平方根変換。
- **自然対数**。自然対数変換。

モデルに定数項を含む。系列値の全体平均が 0 だという確信がない限り、通常は定数項を含めます。差異を適用する場合、定数項を除外することをお勧めします。

ユーザー指定の ARIMA モデルにおける伝達関数

図 2-6
[ARIMA の基準] ダイアログ ボックスの [伝達関数] タブ



[伝達関数] タブ（独立変数が指定された場合にのみ存在する）では、[変数] タブで指定されている独立変数の一部またはすべての伝達関数を定義することができます。伝達関数を使用すると、従属系列の今後の値を予測するために独立（予測）変数の過去の値の使用方法を指定することができます。

伝達関数の順序。 さまざまな伝達関数の成分の値を構造グリッド内の対応するセルに入力します。すべての値は負でない整数にする必要があります。分子と分母の成分については、値は最大次数を表します。すべての正の低い次数はモデルに含まれます。さらに、次数 0 は常に分子成分に含まれます。たとえば、分子に 2 を指定した場合、モデルには順序 2、1、および 0 を含みます。分母に 3 を指定した場合、モデルには順序 3、2、および 1 が含まれます。[季節] 列は、周期がアクティブ データセットに対して定義されている場合にのみ使用可能です（以下の「現在の周期」を参照）。

- **分子。** 伝達関数の分子次数。選択した独立（予測）系列から使用する過去の値を指定し、従属系列の現在の値を予測します。たとえば、分子次数 1 は、過去の 1 期間における独立系列の値（独立系列の現在の値も同様）が各従属系列の現在の値を予測するために使用することを指定します。

- **分母。**伝達関数の分母次数。選択した独立（予測）系列の過去の値に対して、系列の平均の偏差が従属系列の現在の値を予測するためにどのように使用されるのかを指定します。たとえば、分母次数 1 は、各従属系列の現在の値を予測する際に、過去の 1 期間における独立系列の平均値の偏差が考慮されることを指定します。
- **差分。**モデルを推定する前に選択された独立（予測）系列に適用される差分の次数を指定します。トレンドが存在する場合は差分を取る必要があります、トレンドの効果を取り除くために差分を使用します。

季節次数。季節分子、分母、および差分成分は、対応する非季節成分と同様の役割を果たします。ただし、季節次数については、現在の系列値は 1 つ以上の季節期間で区切られた過去の系列値に影響されます。たとえば、毎月のデータ（季節期間 12）については、季節次数 1 は、現在の系列値は現在の期間より 12 期間以前の系列値により影響されることを意味しています。毎月のデータについて、季節次数 1 は、非季節次数 12 を指定するのと同じこととなります。

現在の周期。アクティブ データセットに対して現在定義されている周期（存在する場合）。現在の周期は、整数値として与えられます。たとえば、各ケースが 1 か月を表す年間の周期に対しては 12 となります。周期性が設定されていない場合は、「なし」と表示されます。季節も出るには周期が必要です。

遅延。遅延を設定すると、指定した区間数だけ独立変数の影響が遅延します。たとえば、遅延が 5 に設定された場合、時間 t における独立変数の値は、5 期間が経過するまで予測に影響しません ($t + 5$)。

変換。複数の独立変数に対する伝達関数の指定には、これらの変数に対して実行されるオプションの変換も含まれます。

- **なし。**変換は実行されません。
- **平方根。**平方根変換。
- **自然対数。**自然対数変換。

ユーザー指定の ARIMA モデルの外れ値

図 2-7
[ARIMA の基準] ダイアログ ボックスの [外れ値] タブ



[外れ値] タブは、外れ値 (Pena, Tiao, および Tsay, 2001) の処理に、自動検出、特定のポイントを外れ値として指定、または検出もモデル化もしないを選択します。

外れ値を検出したりモデル化しない。 デフォルトでは、外れ値は検出もモデリングもされません。このオプションを選択すると、外れ値のすべての検出またはモデリングが無効になります。

自動的に外れ値を検出。 外れ値の自動検出を実行するには、このオプションを選択し、以下の外れ値の種類を 1 つ以上選択します。

- 相加的
- レベル シフト (L)
- 攪乱 (I)
- 過渡 (T)
- 季節性加法 (S)
- 局所トレンド (C)
- 加法パッチ (P)

詳細は、B 付録 p.107 外れ値の種類 を参照してください。

特定の時点を外れ値としてモデル化。 特定の時点を外れ値として指定するには、このオプションを選択します。それぞれの外れ値に対して、[外れ値の定義] グリッドの個別の行を使います。所定の行のすべてのセルに値を入力します。

- **タイプ:** 外れ値の種類。サポートされている種類は、相加的（デフォルト）、レベル シフト、攪乱、過渡、季節性加法、および局所トレンドです。

注 1 : アクティブなデータセットに対して日付が指定されていない場合、[外れ値の定義] グリッドには [観測] 列だけが表示されます。外れ値を指定するには、対応するケースの行番号（データ エディタに表示）を入力します。

注 2 : [外れ値の定義] グリッドの [サイクル] 列（存在する場合は、アクティブなデータセットの CYCLE_ 変数の値を指します。

出力

使用可能な出力には、すべてのモデルにわたって計算された結果に加えてそれぞれのモデルの結果が含まれます。それぞれのモデルの結果は、ユーザーが指定した基準に基づいて、最適合モデルまたは最不適合モデルに制限することができます。

統計表と予測表

図 2-8
[時系列モデラー] の [統計量] タブ



[統計量] タブは、クラスタ化の結果を示す表を表示するためのオプションを提供します。

モデルごとの適合度、Ljung-Box 統計量、および外れ値の数を表示。 それぞれの推定モデルについて選択された適合度、Ljung-Box 統計量、および外れ値の数を示す表を表示するには、このオプションを選択（チェック）します。

適合度。 推定されたそれぞれのモデルに対する適合度を示す表に含めるために次のものを 1 つ以上選択することができます。

- 定常 R 2 乗
- R 2 乗
- 平均 2 乗誤差平方根
- 平均絶対パーセント誤差

- 平均絶対誤差
- 最大絶対パーセント誤差
- 最大絶対誤差
- 標準化 BIC

詳細は、A 付録 p.105 適合度統計量 を参照してください。

モデルの比較の統計量。 この一連のオプションは、推定されたすべてのモデルにわたって計算された統計量を示す表の表示を制御します。それぞれのオプションは個別の表を生成します。次のオプションから 1 つ以上を選択できます。

- **適合度。** 定常 R2 乗、R2 乗、平均 2 乗誤差平方根、平均絶対パーセント誤差、平均絶対誤差、最大絶対パーセント誤差、最大絶対誤差、および正規化されたベイズ情報量基準に関する要約統計量およびパーセンタイルの表。
- **残差自己相関関数 (ACF)。** 推定されたすべてのモデルにわたる要約統計量の表と残差の自己相関に対するパーセンタイルです。
- **残差偏自己相関関数 (PACF)。** 推定されたすべてのモデルにわたる要約統計量の表と残差の偏自己相関に対するパーセンタイルです。

個別モデルの統計量。 この一連のオプションは、推定されたそれぞれのモデルの詳細を示す表の表示を制御します。それぞれのオプションは個別の表を生成します。次のオプションから 1 つ以上を選択できます。

- **パラメータ推定値。** 推定されたそれぞれのモデルに対するパラメータ推定値の表を表示します。指数平滑法と ARIMA に対してそれぞれ別の表が表示されます。外れ値が存在する場合、それらに対するパラメータ推定値も別の表に表示されます。
- **残差自己相関関数 (ACF)。** 推定されたそれぞれのモデルについてラグごとの残差自己相関の表を表示します。この表には、自己相関に対する信頼区間が含まれます。
- **残差偏自己相関関数 (PACF)。** 推定されたそれぞれのモデルについてラグごとの偏残差自己相関の表を表示します。この表には、偏自己相関に対する信頼区間が含まれます。

予測値を表示。 推定されたそれぞれのモデルに対するモデル予測値と信頼区間の表を表示します。予測期間は [オプション] タブで設定します。

作図

図 2-9
[時系列モデラー] の [作図] タブ



[作図] タブは、クラスタ化の結果を示す図を表示するためのオプションを提供します。

モデルの比較の作図

この一連のオプションは、推定されたすべてのモデルにわたって計算された統計量を示す図の表示を制御します。それぞれのオプションは個別の図を生成します。次のオプションから 1 つ以上を選択できます。

- 定常 R 2 乗
- R 2 乗
- 平均 2 乗誤差平方根
- 平均絶対パーセント誤差

- 平均絶対誤差
- 最大絶対パーセント誤差
- 最大絶対誤差
- 標準化 BIC
- 残差自己相関関数 (U) (ACF)
- 残差偏自己相関関数 (F) (PACF)

詳細は、A 付録 p.105 適合度統計量 を参照してください。

個別モデルの作図

系列： 推定されたそれぞれのモデルについて予測された値の図を取得するには、このオプションを選択（チェック）します。図に含めるものを次から 1 つ以上選択できます。

- **観測値。** 従属系列の観測値です。
- **予測。** 予測期間に対するモデル予測値です。
- **当てはめ値。** 推定期間に対するモデル予測値です。
- **予測の信頼区間。** 予測期間に対する信頼区間です。
- **当てはめ値の信頼区間。** 推定期間に対する信頼区間です。

残差自己相関関数 (ACF)。 推定されたそれぞれのモデルについて残差自己相関の図を表示します。

残差偏自己相関関数 (PACF)。 推定されたそれぞれのモデルについてラグごとの残差偏自己相関の図を表示します。

出力を最適モデルまたは最不適合モデルに制限する

図 2-10
[時系列モデラー] の [出力フィルタ] タブ



[出力フィルタ] タブは、表出力と図表出力の両方を推定されたモデルのサブセットに制限するためのオプションを提供します。指定した適合基準に基づいて出力を最適モデルおよび最不適合モデルに制限することができます。デフォルトでは、推定されたすべてのモデルが出力に含まれます。

最適モデル。 出力に最適モデルを含めるには、このオプションを選択（チェック）します。適合度を選択し、含めるモデルの数を指定します。このオプションを選択しても、最不適合モデルの選択は排除されません。その場合、出力は、最適モデルに加えて最不適合モデルで構成されます。

- **モデルの固定数。**n 個の最適モデルに対して結果が表示されるように指定します。この数が推定されたモデルの数を超える場合、すべてのモデルが表示されます。
- **モデルの合計数のパーセント。**適合度が全推定モデル中の上位 n パーセントのモデルについて結果が表示されるように指定します。

最不適合モデル。出力に最不適合モデルを含めるには、このオプションを選択（チェック）します。適合度を選択し、含めるモデルの数を指定します。このオプションを選択しても、最適モデルの選択は排除されません。その場合、出力は、最不適合モデルに加えて最適モデルで構成されます。

- **モデルの固定数。**n 個の最不適合モデルに対して結果が表示されるように指定します。この数が推定されたモデルの数を超える場合、すべてのモデルが表示されます。
- **モデルの合計数のパーセント。**適合度が全推定モデル中の下位 n パーセントのモデルについて結果が表示されるように指定します。

適合度。モデルを絞り込むために使用する適合度を選択します。デフォルト設定は、定常 R² 乗です。

モデルの予測とモデルの指定を保存する

[保存] タブでは、モデルの予測を新しい変数としてアクティブなデータセット内に保存したり、モデル指定を外部ファイルに XML 形式で保存したりできます。

変数を保存。モデルの予測、信頼区間、および残差を新しい変数としてアクティブなデータセットに保存することができます。それぞれの従属系列によって新しい変数のグループが発生し、新しいそれぞれの変数には推定期間と予測期間の両方に対する値が含まれます。予測期間が従属変数系列の長さを超えている場合、新しいケースが追加されます。それぞれについて関連する [保存] チェック ボックスを選択することによって新しい変数を保存します。デフォルトでは、新しい変数は保存されません。

- **予測値。**モデルの予測値です。
- **信頼限界の下限。**予測値に対する信頼限界の下限です。
- **信頼限界の上限。**予測値に対する信頼限界の上限です。
- **ノイズ残差。**モデルの残差です。従属変数の変換が実行されたとき（自然対数など）、これらは変換された系列の残差です。
- **変数名の接頭辞。**新しい変数名に対して使用される接頭辞を指定するか、デフォルトの接頭辞のままにします。変数名は、接頭辞、関連付けられた従属変数の名前、およびモデル識別子で構成されます。この変数名は、変数名の競合を避けるために、必要に応じて拡張されます。接頭辞は、有効な変数名に従っている必要があります。

モデル ファイルをエクスポート。 推定されたすべてのモデルに対するモデル指定は、指定されたファイルに XML 形式でエクスポートされます。保存されたモデルは、**時系列モデルの適用** 手続きを使用することによってより多くの現在のデータに基づいて更新された予測を取得するために使用することができます。

- **XML ファイル:** モデルの指定は、IBM SPSS アプリケーションで使用できる XML ファイルに保存されます。
- **PMML ファイル:** モデルの指定は、IBM SPSS アプリケーションを含む PMML 準拠のアプリケーションで使用できる、PMML 準拠 XML ファイルに保存されます。

オプション

図 2-11
[時系列モデラー] の [オプション] タブ

The screenshot shows the '時系列モデラー' (Time Series Modeler) dialog box with the 'オプション' (Options) tab selected. The dialog has a title bar with a close button (X) and a menu icon. Below the title bar are tabs for '変数' (Variables), '統計' (Statistics), '作図' (Plots), '出力フィルタ' (Output Filter), '保存' (Save), and 'オプション' (Options). The 'オプション' tab is active and contains the following settings:

予測期間 (Forecast Period):

- 推定期間の後の最初のケースからアクティブなデータセット内の最後のケースまで (F)
- 推定期間の後の最初のケースから指定された日付まで (C)

日付 (Date):

年	月
2004	3

ユーザー欠損値 (User Missing Values):

- 無効として扱う (I)
- 有効として扱う (V)

信頼区間の幅 (W) (%): 95

出力内のモデル識別子の接頭辞 (P): モデル

AFC および PACF の出力に表示されるラグの最大数 (X): 24

At the bottom of the dialog are buttons for 'OK', '貼り付け (P)', '戻す (R)', 'キャンセル', and 'ヘルプ'.

[オプション] タブでは、予測期間を設定したり、欠損値の処理を指定したり、信頼区間の幅を設定したり、モデル識別子のユーザー指定の接頭辞を指定したり、自己相関に対して表示されるラグの数を設定したりできます。

予測期間。 それぞれのモデルの予測期間は、推定期間（モデルを判定するために使用されるケースのグループ）の終了の後の最初のケースから始まり、アクティブなデータセットまたはユーザーが指定した日付の最後のケースまで続きます。デフォルトでは、推定期間の末尾はアクティブなデータセット内の最後のケースですが、[ケースの選択] ダイアログ ボックスで、[日付かケース番号の範囲] を選択することによって変更できます。

- **推定期間の後の最初のケースからアクティブなデータセット内の最後のケースまで。** 推定期間の末尾がアクティブなデータセットの最後のケースより前で、最後のケースまでの予測が欲しいときは、このオプションを選択してください。このオプションは、通常、ホールドアウト期間の予測を生成して、モデルの予測値と実際の値の部分集合との比較を可能にするために使われます。
- **推定期間の後の最初のケースから指定された日付まで。** 予測期間の末尾を明示的に指定するには、このオプションを選択してください。このオプションは、通常、実際の系列の末尾を超えて予測を生成したい場合に使われます。[日付] グリッドのすべてのセルに値を入力します。

アクティブなデータセットに対して日付が指定されていない場合、[日付] グリッドには[観測] 列だけが表示されます。予測期間の末尾を指定するには、関連するケースの行番号（データ エディタに表示された）を入力します。

[日付] グリッドの [サイクル] 列（存在する場合）は、アクティブなデータセットの CYCLE_ 変数の値を指します。

ユーザー欠損値。 これらのオプションは、ユーザー欠損値の処理を制御します。

- **無効として扱う。** ユーザー欠損値とシステム欠損値を同様に扱います。
- **有効として扱う。** ユーザー欠損値は有効なデータとして取り扱われます。

欠損値のポリシー。 モデリング手続きの間に欠損値（無効として扱われるシステム欠損値とユーザー欠損値を含む）の扱いに対して次の規則が適用されます。

- 推定期間内に発生する従属変数の欠損値を持つケースは、分析に含まれます。欠損値の特定の処理は、推定期間によって異なります。
- 推定期間で独立変数に欠損値がある場合には、警告が発せられます。エキスパート モデラーでは、独立変数を含むモデルの推定はその変数なしで行われます。ユーザー指定の ARIMA では、独立変数を含むモデルの推定は行われません。
- 予測期間内で独立変数に欠損値が含まれる場合、この手続きは警告を発し、可能な限り遠くまで予測します。

信頼区間の幅 (%)。 信頼区間は、モデルの予測と残差自己相関に対して計算されます。100 未満の正の値を指定できます。デフォルトでは、95% 信頼区間が使用されます。

出力内のモデル識別子の接頭辞。 [変数] タブで指定された従属変数に対しては、それぞれ別個の推定モデルが生成されます。それぞれのモデルは、ユーザー指定の接頭辞と整数の接尾辞からなる一意の名前によって識別されます。接尾辞は、独自のものを入力するか、デフォルトの「Model」のままにしておくことができます。

AFC および PACF の出力に表示されるラグの最大数。 自己相関および偏自己相関の表および図に表示されるラグの最大数を指定することができます。

TSMODEL コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、TSMODEL コマンド シンタックスの結果を編集すると、時系列モデリングをカスタマイズできます。コマンド シンタックスを使用すると、次の作業が実行できます。

- データの季節期間を指定する (AUXILIARY サブコマンドの SEASONLENGTH キーワードを使う)。これによって、アクティブなデータセットの現在の周期 (ある場合) が上書きされます。
- ユーザー指定の ARIMA と伝達関数の成分に対して不連続ラグを指定する (ARIMA サブコマンドと TRANSFERFUNCTION サブコマンドを使用)。たとえば、次数 1、3、6 の自己回帰ラグを使用してユーザー指定の ARIMA モデルを指定したり、次数 2、5、8 の分子ラグを使用して伝達関数を指定したりすることができます。
- [時系列モデラー] 手続きの 1 回の実行に対して、モデリングに関する指定項目のグループ (モデリング方法、ARIMA 次数、独立変数など) を複数指定する (MODEL サブコマンドを使用)。

複雑なシンタックス情報については、「コマンド シンタックス リファレンス」を参照してください。

時系列モデルの適用

[時系列モデルの適用] 手続きでは、既存の時系列モデルが外部ファイルからロードされ、アクティブなデータセットに適用されます。この手続きを使用すると、モデルを再構築することなく、新しいデータや変更されたデータに対して系列の予測を行うことができます。モデルは、**時系列モデル** 手続きを使用して生成されます。

例: 大手小売店の在庫の管理者として、5,000 種類の製品を管理するとします。各製品の今後 3 ヶ月の売り上げを予測するモデルを作成するためにエキスパートモデラーを使用しました。データウェアハウスは、実際の売り上げデータによって毎月更新されます。このデータは、月次予測を作成するために使用することができます。時系列モデルの適用手続きを使用すると、元のモデルを使用し、新しいデータの詳細を明らかにするために単にモデルパラメータを再評価することによってこれを行うことができます。

統計量。 適合度：定常 R^2 乗、 R^2 乗 (R^2)、平均 2 乗誤差平方根 (RMSE)、平均絶対誤差 (MAE)、平均絶対パーセント誤差 (MAPE)、最大絶対誤差 (MaxAE)、最大絶対パーセント誤差 (MaxAPE)、正規化されたベイズ情報量基準 (BIC)。残差：自己相関関数、偏自己相関関数、Ljung-Box Q。

作図。 すべてのモデルの要約図：定常 R^2 乗、 R^2 乗 (R^2)、平均 2 乗誤差平方根 (RMSE)、平均絶対誤差 (MAE)、平均絶対パーセント誤差 (MAPE)、最大絶対誤差 (MaxAE)、最大絶対パーセント誤差 (MaxAPE)、および正規化されたベイズ情報量基準 (BIC) のヒストグラム。残差自己相関および偏自己相関の箱ひげ図。個々の値の結果：予測値、適合値、観測値、信頼区間の上限と下限、残差自己相関関数と偏自己相関関数。

時系列モデルの適用のデータの考慮事項

データ。 モデルが適用される変数（従属変数と独立変数）は数値型である必要があります。

仮定。 モデルは、モデル内で指定された変数と同じ名前のアクティブなデータセット内の変数に適用されます。そのような変数は、すべて時系列として扱われます。これは、それぞれのケースが時点を表し、以降のケースは一定の時間間隔で区切られることを意味します。

- **予測。** 独立（予測）変数を持つモデルを使用して予測を作成するためには、アクティブなデータセットには予測期間内のすべてのケースについてそれらの変数の値が含まれている必要があります。モデルのパラメータが再度推定された場合、独立変数には予測期間内で欠損値が含まれてはいけません。

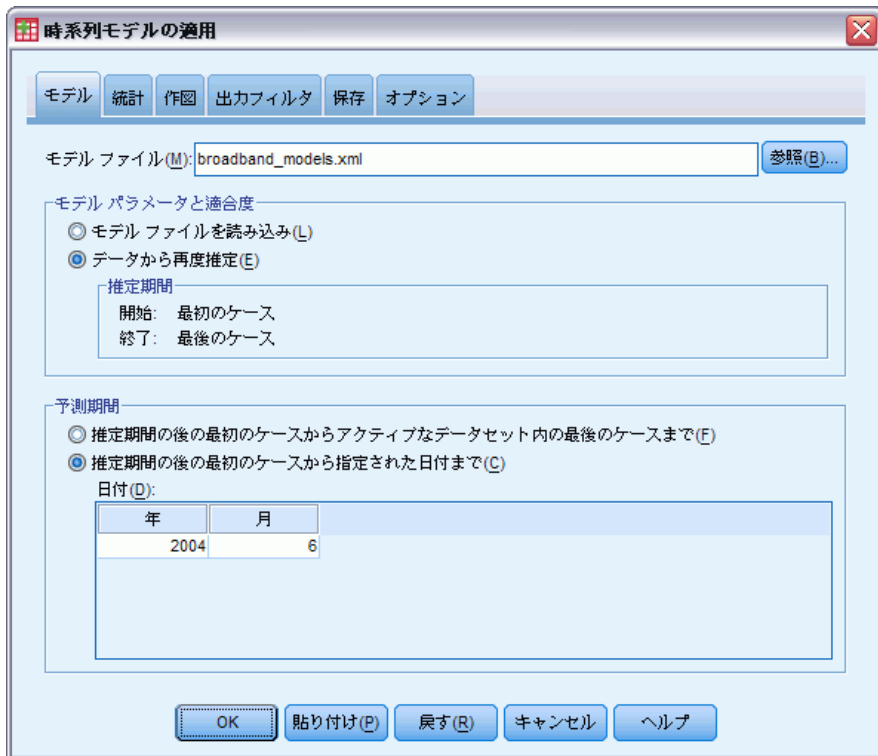
日付の定義

[時系列モデルの適用] 手続きでは、アクティブなデータセットの周期（ある場合）が適用するモデルの周期に一致していることが必要となります。単にモデルをビルドするために使用されたデータセットと同じデータセット（新規または改訂されたデータを含む場合もある）を使用して予測する場合、この条件は満たされます。アクティブなデータセットに対して周期が存在しない場合は、[日付の定義] ダイアログ ボックスを開き、そこで周期を作成できます。ただし、周期を指定しないでモデルを作成した場合、アクティブなデータセットも周期を指定しないで作成する必要があります。

モデルを適用するには

- ▶ メニューから次の項目を選択します。
分析 > 見込み > モデルの適用(A)...

図 3-1
[時系列モデルの適用] の [モデル] タブ



- ▶ モデル ファイルのパス名を入力するか、[参照] をクリックしてモデル ファイルを選択します（モデル ファイルは [時系列モデラー](#) 手続きを使用して作成されます）。

オプションとして、次の選択が可能です。

- アクティブなデータセットのデータを使用してモデルのパラメータを再度推定する。予測は、再度推定されたパラメータを使用して作成されます。
- 予測、信頼区間、およびノイズ残差を保存します。
- 再度推定されたモデルを XML 形式で保存する。

モデル パラメータと適合度

モデル ファイルをロード。 予測は、モデル ファイルから取得したモデルのパラメータを再度推定なしで使用して作成されます。出力に表示され、モデル（最適または最不適）を絞り込むために使用される **適合度**は、モデル ファイルから取得され、それぞれのモデルが作成された（または最後に更新された）ときに使用されたデータを反映します。このオプションを使用すると、アクティブなデータセットの従属変数と独立変数のどちらに対しても、予測で履歴データは考慮されません。履歴データを予測に反映させたい場合は、[データから再度推定]を選択する必要があります。さらに、予測では、予測期間内の従属系列の値は考慮されませんが、予測期間内の独立変数の値は考慮されます。従属系列の現在の値の方が多く、それらを予測に組み込みたい場合は、それらの値が含まれるように推定期間を調節することによって再度推定する必要があります。

データから再度推定。 アクティブなデータセットのデータを使用してモデルのパラメータが再度推定されます。モデルのパラメータの再度推定は、モデルの構造には影響しません。たとえば、ARIMA(1, 0, 1) モデルはそのままになりますが、自己回帰パラメータと移動平均パラメータは再度推定されます。再度推定によって新しい外れ値が検知されることはありません。外れ値は、常にモデル ファイルから取得されます。

- **推定期間。** 推定期間は、モデルのパラメータを再度推定するために使用されるケースのグループを定義します。デフォルトでは、推定期間はアクティブなデータセットのすべてのケースを含みます。推定期間を設定するには、[ケースの選択] ダイアログ ボックスで、[日付かケース番号の範囲]を選択します。使用可能なデータによっては、この手続きによって使用される推定期間はモデルごとに異なる場合があるため、表示される値と異なる場合があります。指定されたモデルに対して、本当の推定期間は、モデルの従属変数から、指定された推定期間の先頭または末尾で連続する欠損値を削除した後に残った期間です。

予測期間

それぞれのモデルの予測期間は、推定期間の終了の後の最初のケースから始まり、アクティブなデータセットまたはユーザーが指定した日付の最後のケースまで続きます。パラメータが再度推定されていない場合（デフォ

ルト)、それぞれのモデルの推定期間は、モデルが作成された（または更新された）ときに使用されたケースのグループです。

- **推定期間の後の最初のケースからアクティブなデータセット内の最後のケースまで。** 推定期間の末尾がアクティブなデータセットの最後のケースより前で、最後のケースまでの予測が欲しいときは、このオプションを選択してください。
- **推定期間の後の最初のケースから指定された日付まで。** 予測期間の末尾を明示的に指定するには、このオプションを選択してください。[日付]グリッドのすべてのセルに値を入力します。

アクティブなデータセットに対して日付が指定されていない場合、[日付]グリッドには[観測]列だけが表示されます。予測期間の末尾を指定するには、関連するケースの行番号（データ エディタに表示された）を入力します。

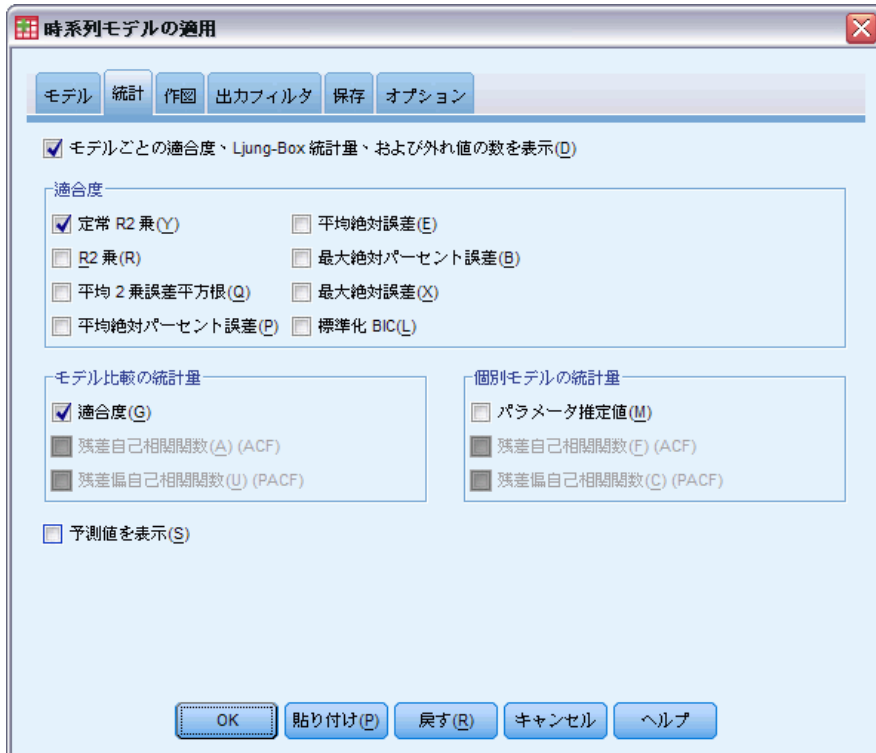
[日付]グリッドの[サイクル]列（存在する場合）は、アクティブなデータセットの CYCLE_ 変数の値を指します。

出力

使用可能な出力には、すべてのモデルにわたる結果に加えてそれぞれのモデルの結果が含まれます。それぞれのモデルの結果は、ユーザーが指定した基準に基づいて、最適合モデルまたは最不適合モデルに制限することができます。

統計表と予測表

図 3-2
[時系列モデルの適用] の [統計量] タブ



[統計量] タブは、モデル適合統計量、モデルのパラメータ、自己相関関数、および予測の表を表示するオプションを提供します。モデルのパラメータが再度推定されない限り（[モデル] タブの [データから再度推定]）、表示される適合測度の値、Ljung-Box 値、およびモデルのパラメータは、モデル ファイルから取得した値で、それぞれのモデルが作成された（または最後に更新された）ときに使用されたデータを反映します。外れ値の情報は、常にモデル ファイルから取得されます。

モデルごとの適合度、Ljung-Box 統計量、および外れ値の数を表示。 それぞれのモデルについて選択された適合度、Ljung-Box 統計量、および外れ値の数を示す表を表示するには、このオプションを選択（チェック）します。

適合度。 それぞれのモデルに対する適合度を示す表に含めるために次のものを 1 つ以上選択することができます。

- 定常 R² 乗
- R² 乗
- 平均 2 乗誤差平方根

- 平均絶対パーセント誤差
- 平均絶対誤差
- 最大絶対パーセント誤差
- 最大絶対誤差
- 標準化 BIC

詳細は、A 付録 p.105 適合度統計量を参照してください。

モデルの比較の統計量。 この一連のオプションは、すべてのモデルにわたる統計量を示す表の表示を制御します。それぞれのオプションは個別の表を生成します。次のオプションから 1 つ以上を選択できます。

- **適合度。** 定常 R2 乗、R2 乗、平均 2 乗誤差平方根、平均絶対パーセント誤差、平均絶対誤差、最大絶対パーセント誤差、最大絶対誤差、および正規化されたベイス情報量基準に関する要約統計量およびパーセンタイルの表。
- **残差自己相関関数 (ACF)。** 推定されたすべてのモデルにわたる要約統計量の表と残差の自己相関に対するパーセンタイルです。この表は、モデルのパラメータが再度推定された（[モデル] タブの [データから再度推定]）場合のみ使用することができます。
- **残差偏自己相関関数 (PACF)。** 推定されたすべてのモデルにわたる要約統計量の表と残差の偏自己相関に対するパーセンタイルです。この表は、モデルのパラメータが再度推定された（[モデル] タブの [データから再度推定]）場合のみ使用することができます。

個別モデルの統計量。 この一連のオプションは、それぞれのモデルの詳細を示す表の表示を制御します。それぞれのオプションは個別の表を生成します。次のオプションから 1 つ以上を選択できます。

- **パラメータ推定値。** それぞれのモデルに対するパラメータ推定値の表を表示します。指数平滑法と ARIMA に対してそれぞれ別の表が表示されます。外れ値が存在する場合、それらに対するパラメータ推定値も別の表に表示されます。
- **残差自己相関関数 (ACF)。** 推定されたそれぞれのモデルについてラグごとの残差自己相関の表を表示します。この表には、自己相関に対する信頼区間が含まれます。この表は、モデルのパラメータが再度推定された（[モデル] タブの [データから再度推定]）場合のみ使用することができます。
- **残差偏自己相関関数 (PACF)。** 推定されたそれぞれのモデルについてラグごとの偏残差自己相関の表を表示します。この表には、偏自己相関に対する信頼区間が含まれます。この表は、モデルのパラメータが再度推定された（[モデル] タブの [データから再度推定]）場合のみ使用することができます。

予測値を表示。 それぞれのモデルに対するモデル予測値と信頼区間の表を表示します。

作図

図 3-3
[時系列モデルの適用] の [作図] タブ



[作図] タブは、モデル適合統計量、モデルのパラメータ、自己相関関数、および系列値（予測を含む）の表を表示するオプションを提供します。

モデルの比較の作図

この一連のオプションは、すべてのモデルにわたる統計量を示す図の表示を制御します。モデルのパラメータが再度推定されない限り（[モデル] タブの [データから再度推定]）、表示される値は、モデル ファイルから取得した値で、それぞれのモデルが作成された（または最後に更新された）ときに使用されたデータを反映します。さらに、自己相関図は、モデルのパラメータが再度推定された場合のみ使用することができます。それぞれのオプションは個別の図を生成します。次のオプションから 1 つ以上を選択できます。

- 定常 R 2 乗
- R 2 乗
- 平均 2 乗誤差平方根
- 平均絶対パーセント誤差

- 平均絶対誤差
- 最大絶対パーセント誤差
- 最大絶対誤差
- 標準化 BIC
- 残差自己相関関数 (U) (ACF)
- 残差偏自己相関関数 (F) (PACF)

詳細は、A 付録 p.105 適合度統計量 を参照してください。

個別モデルの作図

系列：それぞれのモデルについて予測された値の図を取得するには、このオプションを選択（チェック）します。観測値、当てはめ値、当てはめ値に対する信頼区間、および自己相関は、モデルのパラメータが再度推定された（[モデル] タブの [データから再度推定]）場合のみ使用することができます。図に含めるものを次から 1 つ以上選択できます。

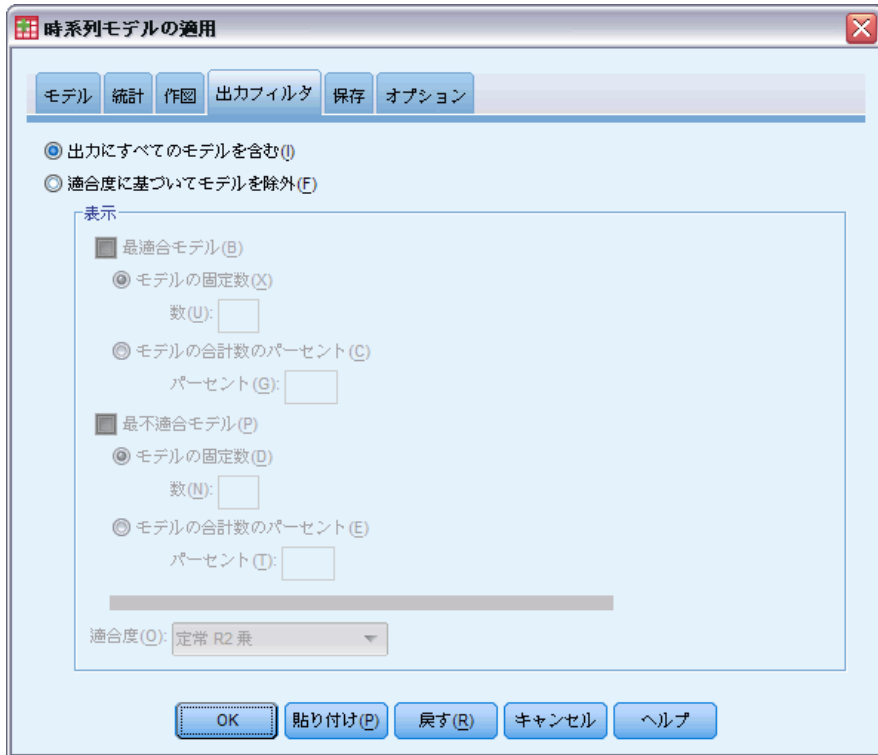
- **観測値。** 従属系列の観測値です。
- **予測。** 予測期間に対するモデル予測値です。
- **当てはめ値。** 推定期間に対するモデル予測値です。
- **予測の信頼区間。** 予測期間に対する信頼区間です。
- **当てはめ値の信頼区間。** 推定期間に対する信頼区間です。

残差自己相関関数 (ACF)。 推定されたそれぞれのモデルについて残差自己相関の図を表示します。

残差偏自己相関関数 (PACF)。 推定されたそれぞれのモデルについてラグごとの残差偏自己相関の図を表示します。

出力を最適モデルまたは最不適合モデルに制限する

図 3-4
[時系列モデルの適用] の [出力フィルタ] タブ



[出力フィルタ] タブは、表出力と図表出力の両方をモデルのサブセットに制限するためのオプションを提供します。指定した適合基準に基づいて出力を最適モデルおよび最不適合モデルに制限することができます。デフォルトでは、すべてのモデルが出力に含まれます。モデルのパラメータが再度推定されない限り（[モデル] タブの [データから再度推定]）、モデルを絞り込むために使用される適合度の値は、モデル ファイルから取得した値で、それぞれのモデルが作成された（または最後に更新された）ときに使用されたデータを反映します。

最適モデル。 出力に最適モデルを含めるには、このオプションを選択（チェック）します。適合度を選択し、含めるモデルの数を指定します。このオプションを選択しても、最不適合モデルの選択は排除されません。その場合、出力は、最適モデルに加えて最不適合モデルで構成されます。

- **モデルの固定数。** n 個の最適モデルに対して結果が表示されるように指定します。この数がモデルの総数を超える場合、すべてのモデルが表示されます。
- **モデルの合計数のパーセント。** 適合度が全モデル中の上位 n パーセントのモデルについて結果が表示されるように指定します。

最不適合モデル。 出力に最不適合モデルを含めるには、このオプションを選択（チェック）します。適合度を選択し、含めるモデルの数を指定します。このオプションを選択しても、最適モデルの選択は排除されません。その場合、出力は、最不適合モデルに加えて最適モデルで構成されます。

- **モデルの固定数。** n 個の最不適合モデルに対して結果が表示されるように指定します。この数がモデルの総数を超える場合、すべてのモデルが表示されます。
- **モデルの合計数のパーセント。** 適合度が全モデル中の下位 n パーセントのモデルについて結果が表示されるように指定します。

適合度。 モデルを絞り込むために使用する適合度を選択します。デフォルト設定は、定常 R^2 乗です。

モデルの予測とモデルの指定を保存する

[保存] タブでは、モデルの予測を新しい変数としてアクティブなデータセット内に保存したり、モデル指定を外部ファイルに XML 形式で保存したりできます。

変数を保存。 モデルの予測、信頼区間、および残差を新しい変数としてアクティブなデータセットに保存することができます。それぞれのモデルは、独自の新しい変数のグループの元となります。予測期間がモデルに関連付けられた従属変数系列の長さを超えている場合、新しいケースが追加されます。モデルのパラメータが再度推定されない限り（[モデル] タブの [データから再度推定]）、予測値と信頼限界は、予測期間に対してのみ作成されます。それぞれについて関連する [保存] チェック ボックスを選択することによって新しい変数を保存します。デフォルトでは、新しい変数は保存されません。

- **予測値。** モデルの予測値です。
- **信頼限界の下限。** 予測値に対する信頼限界の下限です。
- **信頼限界の上限。** 予測値に対する信頼限界の上限です。

- **ノイズ残差。** モデルの残差です。従属変数の変換が実行されたとき（自然対数など）、これらは変換された系列の残差です。この選択は、モデルのパラメータが再度推定された（[モデル] タブの [データから再度推定]）場合のみ使用することができます。
- **変数名の接頭辞。** 新しい変数名に対して使用される接頭辞を指定するか、デフォルトの接頭辞のままにします。変数名は、接頭辞、関連付けられた従属変数の名前、およびモデル識別子で構成されます。この変数名は、変数名の競合を避けるために、必要に応じて拡張されます。接頭辞は、有効な変数名に従っている必要があります。

モデル ファイルをエクスポート。 再度推定されたパラメータと適合統計量を含むモデル指定は、指定されたファイルに XML 形式でエクスポートされます。このオプションは、モデルのパラメータが再度推定された（[モデル] タブの [データから再度推定]）場合のみ使用することができます。

- **XML ファイル:** モデルの指定は、IBM SPSS アプリケーションで使用できる XML ファイルに保存されます。
- **PMML ファイル:** モデルの指定は、IBM SPSS アプリケーションを含む PMML 準拠のアプリケーションで使用できる、PMML 準拠 XML ファイルに保存されます。

オプション

図 3-5
[時系列モデルの適用] の [オプション] タブ



[オプション] タブでは、欠損値の処理を指定したり、信頼区間の幅を設定したり、自己相関に対して表示されるラグの数を設定したりできます。

ユーザー欠損値。 これらのオプションは、ユーザー欠損値の処理を制御します。

- **無効として扱う。** ユーザー欠損値とシステム欠損値を同様に扱います。
- **有効として扱う。** ユーザー欠損値は有効なデータとして取り扱われます。

欠損値のポリシー。 欠損値（無効として扱われるシステム欠損値とユーザー欠損値を含む）の扱いに対して次の規則が適用されます。

- 推定期間内に発生する従属変数の欠損値を持つケースは、分析に含まれます。欠損値の特定の処理は、推定期間によって異なります。
- ARIMA モデルでは、推定期間内で予測変数に欠損値が含まれる場合に、警告が発せられます。予測変数を伴うモデルは、再度推定されません。
- 予測期間内で独立変数に欠損値が含まれる場合、この手続きは警告を発生し、可能な限り遠くまで予測します。

信頼区間の幅 (%)。 信頼区間は、モデルの予測と残差自己相関に対して計算されます。100 未満の正の値を指定できます。デフォルトでは、95% 信頼区間が使用されます。

AFC および PACF の出力に表示されるラグの最大数。 自己相関および偏自己相関の表および図に表示されるラグの最大数を指定することができます。このオプションは、モデルのパラメータが再度推定された（[モデル] タブの [データから再度推定]）場合のみ使用することができます。

TSAPPLY コマンドの追加機能

選択内容をシンタックス ウィンドウに貼り付け、TSAPPLY コマンド シンタックスの結果を編集すると、追加の機能を使用することができます。コマンド シンタックスを使用すると、次の作業が実行できます。

- モデル ファイルから取得したモデルのサブセットだけがアクティブなデータセットに適用されるように指定する (MODEL サブコマンドで DROP キーワードと KEEP キーワードを使用)。
- 2 つ以上のモデル ファイルに保存されているモデルをデータに適用する (MODEL サブコマンドを使用)。たとえば、あるモデル ファイルに売上数量を表す系列のモデルが含まれていて、別のモデル ファイルには収入を表す系列のモデルが含まれている場合があります。

複雑なシンタックス情報については、「コマンド シンタックス リファレンス」を参照してください。

[季節性の分解]

季節性の分解手続きでは、系列が季節成分、結合されたトレンド成分と周期成分、および“誤差”成分に分解されます。この手続きは、移動平均比率法としても知られているセンサス局が開発した Method I の実装です。

例。 ある科学者が、特定の気象観測所で月に 1 回測定されるオゾンレベルの分析に関心を持っています。目的は、データになんらかのトレンドが存在するかどうかを判断することです。実際になんらかのトレンドを見つけるには、まず科学者は季節的影響による測定値の変化について説明する必要があります。季節性の分解手続きを使用して、システムティックな季節変動を取り除くことができます。トレンド分析は、その後季節調整系列で実行されます。

[統計量] 一連の季節因子。

データ。 変数は数値型にする必要があります。

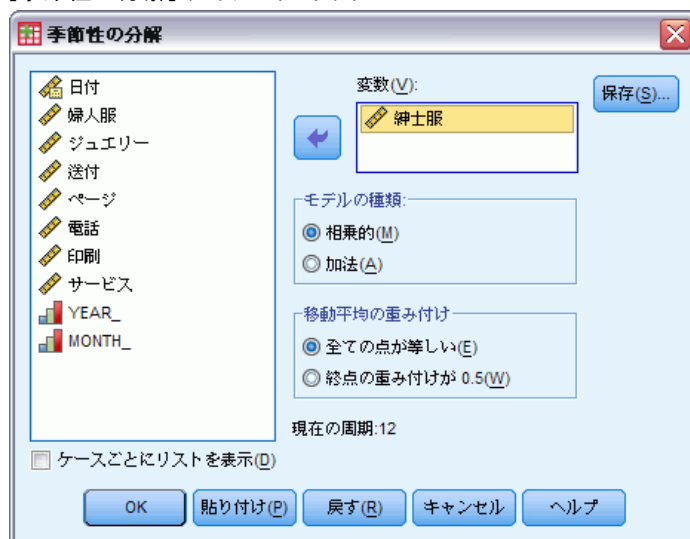
仮定。 変数には埋め込まれた欠損データを含めないようにする必要があります。少なくとも 1 つの周期的な日付成分を定義する必要があります。

季節因子の推定

- ▶ メニューから次の項目を選択します。

[分析] > 予測 > [季節性の分解...]

図 4-1
[季節性の分解] ダイアログ ボックス



- ▶ 使用できるリストから 1 つ以上の変数を選択し、[変数] リストに移動します。このリストに含まれているのは数値型変数のみであることに注意してください。

モデルの種類。 季節性の分解手続きでは、季節因子のモデリング用に、相乗的と相加的という異なる 2 種類のアプローチが提供されます。

- **相乗的。** 季節成分は、季節調整系列がオリジナル系列を生じるために掛け算された要因です。その結果、系列の全体の水準に比例する季節成分となります。季節変動のない観測は 1 の季節成分を持ちます。
- **加法。** 季節調整はオリジナル系列を得るために季節調整系列に加えられます。季節成分によって覆われている可能性のある重要な他の特性を見るために、この調整で系列から季節効果を取り除きます。その結果、系列の全体の水準に依存していない季節成分を推定します。季節変動のない観測は 0 の季節成分を持ちます。

移動平均の重み付け。 [移動平均の重み付け] オプションを使用すると、移動平均を計算する際の系列の処理方法を指定できます。これらのオプションは、系列の周期が偶数の場合にだけ使用できます。周期が奇数の場合は、すべての点が同じ重みになります。

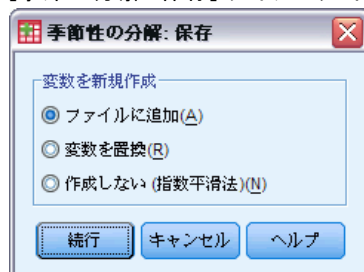
- **全ての点が等しい (季節性の分解)。** 移動平均は周期性に等しいスパンと平等に重み付けして計算されたすべての点で計算されます。この方法は、周期性が奇数の場合、常に使用できます。
- **終点の重み付けが 0.5 (季節性の分解)。** 周期性が偶数の移動平均は周期性に 1 を足した値に等しいスパンと 0.5 で重み付けされたスパンの終点で計算されます。

オプションとして、次の選択が可能です。

- [保存] をクリックし、新しい変数の保存方法を指定します。

季節性の分解の保存

図 4-2
[季節性の分解: 保存] ダイアログ ボックス



変数の作成。 新しい変数を扱う方法を選択できます。

- **ファイルに追加 (季節性の分解).** 季節性の分解で作成された新しい系列は、作業データ ファイルでは通常変数として保存されます。変数名は、3 文字の接頭辞、下線、および数字で構成されます。
- **変数を置換 (季節性の分解).** 季節性の分解で作成した新しい系列は、アクティブなデータセットに一時変数として保存されます。同時に、予測手続きによって作成された一時変数は破棄されます。変数名は、3 文字の接頭辞、ポンド記号 (#) および数字から形成されます。
- **作成しない (指数平滑法).** 新しい時系列は、作業データ ファイルに追加されません。

新しい変数名

季節の分解手続きにより、4 つの新しい変数 (時系列) が作成され、指定した時系列ごとに、次の 3 文字の接頭辞が割り当てられます。

SAF. 季節性の調整因子。この値は、時系列の水準に与える各周期の効果を示します。

SAS 季節調整系列。系列の季節変動を取り除いた後に取得される値です。

STC 平滑化したトレンド サイクル成分。この値は、時系列に存在するトレンドおよび周期動作を示します。

ERR 残差または「誤差」値。季節成分、トレンド成分、およびサイクル成分が時系列から取り除かれた後の残った値です。

SEASON コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 日付の定義手続きにより提供されるオプションのいずれかを選択せずに、SEASON コマンド内で任意の周期を指定する。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

スペクトルプロット

スペクトルプロット手続きは、時系列の周期的な動作を識別するために使用します。ある時点から次の時点までの変数を分析するのではなく、系列の変動を全体的な数値として、さまざまな周波数の周期成分を分析します。平滑化系列では低い周波数に周期成分が多く存在し、ランダムな変動（“ホワイト ノイズ”）では周波数全体に周期成分が分散します。

欠損データが含まれる系列は、この手続きを使用して分析することはできません。

例。新しい家の建築率は、経済状態を表す重要な指標です。一般に、住宅着工件数のデータにより、季節成分がはっきりと示されます。しかし、現在の数値を評価したときに分析者が気付く必要のあるデータには、より長いサイクルが存在するのでしょうか。

統計量。周波数と周期成分ごとのサイン変換とコサイン変換、ペリオドグラムの値、およびスペクトル密度の推定値。2 変数分析の選択時：周波数と周期成分ごとのクロス - ペリオドグラムの実部と虚部、共スペクトル密度、直角位相スペクトル、ゲイン、平方コヒアランス、および位相スペクトル。

作図。1 変数分析と 2 変数分析：ペリオドグラムとスペクトル密度。2 変数分析：平方コヒアランス、直角位相スペクトル、クロス振幅、共スペクトル密度、位相スペクトル、およびゲイン。

データ。変数は数値型にする必要があります。

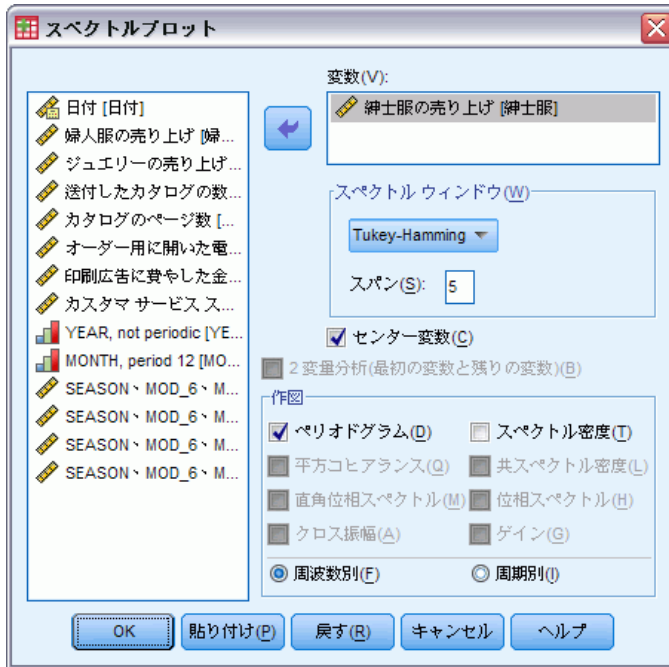
仮定。変数には埋め込まれた欠損データを含めないようにする必要があります。分析する時系列は定常的である必要があります。また、0 以外の平均値はすべてその時系列から差し引かれます。

- **定常的。**ARIMA モデルを適合させるとき、時系列が満たされる必要がある条件です。純粋 MA 系列は定常的であっても、AR 系列と ARMA 系列は定常的でない可能性があります。定常時系列は一定の平均と分散を持ちます。

スペクトル分析を実行するには

- ▶ メニューから次の項目を選択します。
分析 > 時系列 > スペクトル分析...

図 5-1
[スペクトルプロット] ダイアログ ボックス



- ▶ 使用できるリストから 1 つ以上の変数を選択し、[変数] リストに移動します。このリストに含まれているのは数値型変数のみであることに注意してください。
- ▶ [スペクトル ウィンドウ] オプションのいずれかを選択してペリオドグラムの平滑化方法を指定し、スペクトル密度の推定を行います。使用できる平滑化オプションは、[Tukey-Hamming]、[Tukey]、[Parzen]、[Bartlett 法]、[Daniell (Unit)]、および [なし] です。
 - **Tukey-Hamming.** 重みを $W_k = .54D_p(2 \pi f_k) + .23D_p(2 \pi f_k + \pi/p) + .23D_p(2 \pi f_k - \pi/p)$, for $k = 0, \dots, p$ のように計算します。p はスパンを 2 で割った値の整数部、 D_p は次数 p の Dirichlet 核です。
 - **Tukey.** $W_k = 0.5D_p(2 \pi f_k) + 0.25D_p(2 \pi f_k + \pi/p) + 0.25D_p(2 \pi f_k - \pi/p)$, for $k = 0, \dots, p$ のように重みを計算するウィンドウです。p は、スパンを 2 で割った値の整数部、 D_p は次数 p の Dirichlet 核です。
 - **Parzen.** 重み付けは、 $W_k = 1/p(2 + \cos(2 \pi f_k)) (F[p/2](2 \pi f_k))^{**2}$, for $k = 0, \dots, p$ です。p はスパンの半分の整数部で、 $F[p/2]$ は次数 p/2 の Fejer カーネルです。

- **Bartlett.** ウィンドウの上半分のみが $k = 0, \dots, p$ のとき、 $W_k = F_p(2\pi f_k)$ として計算される (p はスパンの整数部分、 F_p は次数 p の Feje のカーネル) スペクトル ウィンドウの形状です。下半分は上半分と対称です。
- **Daniell (Unit).** 重み付けがすべて 1 であるスペクトル ウィンドウ。
- **なし (コレスポンデンス分析).** 平滑化なし。このオプションを選択した場合、スペクトル密度の推定値は、ペリオドグラムと同じになります。

スパン (個別の図表、移動範囲図表). 平滑化が行われる連続した観測値の範囲です。一般に、奇数が使われます。大きいスパンは小さいスパンよりスペクトル密度プロットを平滑にします。

センター変数 (スペクトル). スペクトルを計算する前に平均値をゼロにするため、および系列平均値と関連する大きな項を除去するため、系列を調整します。

2 変数分析 (最初の変数と残りの変数). 複数の変数を選択した場合、このオプションを選択して 2 変数相関のスペクトル分析を要求できます。

- [変数] リストの最初の変数は独立変数として扱われ、残りの変数はすべて従属変数として扱われます。
- 最初の系列より後の各系列は、指定された他の系列とは関係なく、最初の系列を使用して分析されます。各系列の 1 変数分析も実行されます。

作図. ペリオドグラムとスペクトル密度は、1 変数分析と 2 変数分析の両方で使用できます。他のすべての選択肢は、2 変数分析でのみ使用できます。

- **ペリオドグラム.** 周波数または周期に対するスペクトルの大きさを滑らかにしないで作図したもの。低い周波数の変動は、平滑な系列を特徴付けるものです。すべての周波数にわたって一様に広がる変動はホワイトノイズを示します。
- **平方コヒーレンス.** 2 つの系列のゲインの積です。
- **直角位相.** クロス - ペリオドグラムの虚部であり、2 つの時系列の同相周波数成分に関する相関の測定値です。成分は、 $\pi/2$ ラジアンによる同相です。
- **クロス振幅.** 平方共スペクトル密度と平方直角位相スペクトルの和の平方根です。
- **スペクトル密度.** 不規則変動を除去するために、ペリオドグラムを平滑化したものです。
- **共スペクトル密度.** 2 つの時系列の in(内部)-位相周波数成分の相関を測定したものです。クロスペリオドグラムの実部です。
- **位相スペクトル.** 各周波数成分が他の周波数成分のラグ (遅れ)、またはリード (進み) という程度を測定したものです。
- **ゲイン.** 系列の 1 つのスペクトル密度でクロス振幅を割った商です。2 つの系列の各々にはそれぞれゲイン値があります。

周波数別 (スペクトル). すべてのプロットは、周波数 0 (定数または平均値項) から周波数 0.5 (2 つの観測サイクル間の項) の範囲で周波数別に作成されます。

周期別 (スペクトル). すべてのプロットは、2 (2 つの観測サイクル間の項) から観測数に等しい期間 (定数または平均値項) までの範囲で周期別に作成されます。周期は、対数スケールで表示されます。

SPECTRA コマンドの追加機能

コマンド シンタックスを使用すると、次の作業も実行できます。

- 計算されたスペクトル分析変数を、後で使用できるようにアクティブデータセットに保存する。
- スペクトル ウィンドウの重みをユーザー指定する。
- 周波数と周期の両方を基準に作図を行う。
- プロットに示される各値の詳細なリストを出力する。

シンタックスの詳細は、『Command Syntax Reference』を参照してください。

パート II: 例

エキスパート モデラーによるバルク予測

ここでは、全国規模のブロードバンド プロバイダから依頼を受けた分析担当者が、帯域の利用状況を予測するために、ユーザー契約数の予測値を求めるというシナリオを設定します。予測は、全国の加入者基盤を構成する 85 の地方市場のそれぞれについて行う必要があります。過去の月ごとのデータは、broadband_1.sav に収集されています。詳細は、D 付録 サンプル ファイル in IBM SPSS Forecasting 21 を参照してください。

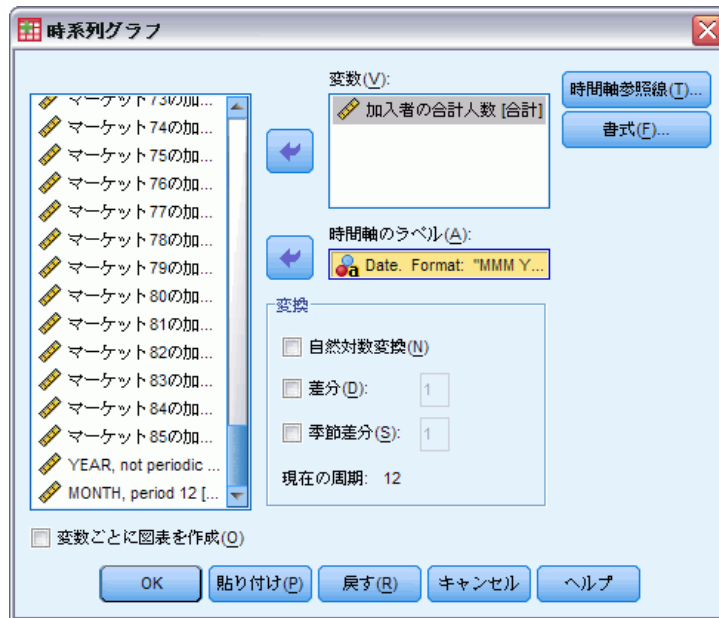
この例では、エキスパート モデラーを使用して、85 ある各地方市場について今後 3 か月間の予測値を求め、生成されたモデルを外部の XML ファイルに保存します。この作業が完了したら、この例の続編にあたる p.63 7 章保存したモデルの適用によるバルク再予測についても、実際に作業してみることをお勧めします。この例では、ここで保存したモデルを、更新されたデータセットに適用することで、モデルを再構築することなく予測期間をさらに 3 か月延長します。

データの検証

モデルを構築する場合は常に、季節変動が見られないかなど、使用するデータの性質をあらかじめ大まかに把握しておくことが重要です。エキスパート モデラーでは、系列ごとに最適な季節モデルまたは非季節モデルが自動的に特定されますが、使用するデータに季節性が見られない場合は、探索の対象を非季節モデルに限定することにより、より短時間のうちに最適モデルを特定できます。すべての市場における総加入者数をプロットすれば、85 ある地方市場のそれぞれについてデータを検証しなくても、おおまかな全体像を把握することができます。

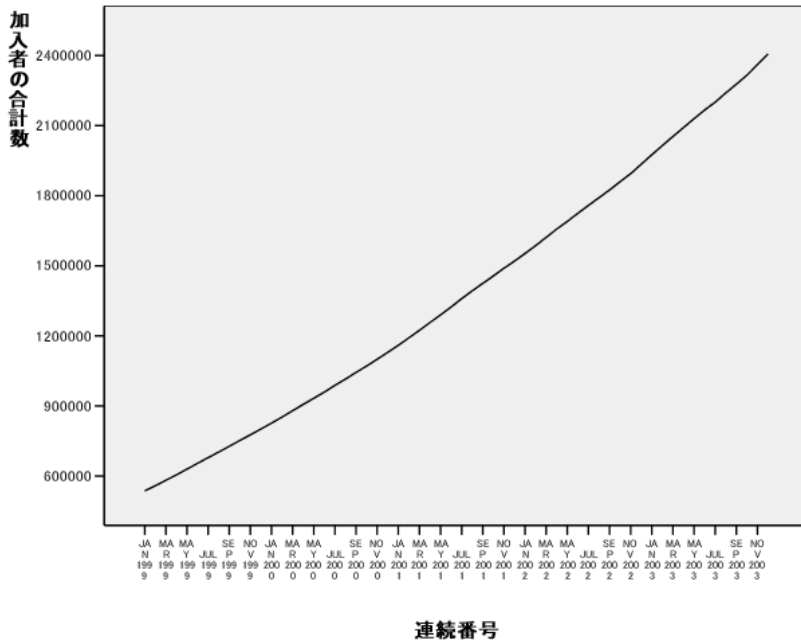
- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > 時系列グラフ...

図 6-1
[時系列グラフ] ダイアログ ボックス



- ▶ 「加入者の合計数」を選択し、[変数] リストに移動します。
- ▶ 「日付」を選択し、[時間軸のラベル] ボックスに移動します。
- ▶ [OK] をクリックします。

図 6-2
全市場におけるブロードバンドの総加入者数



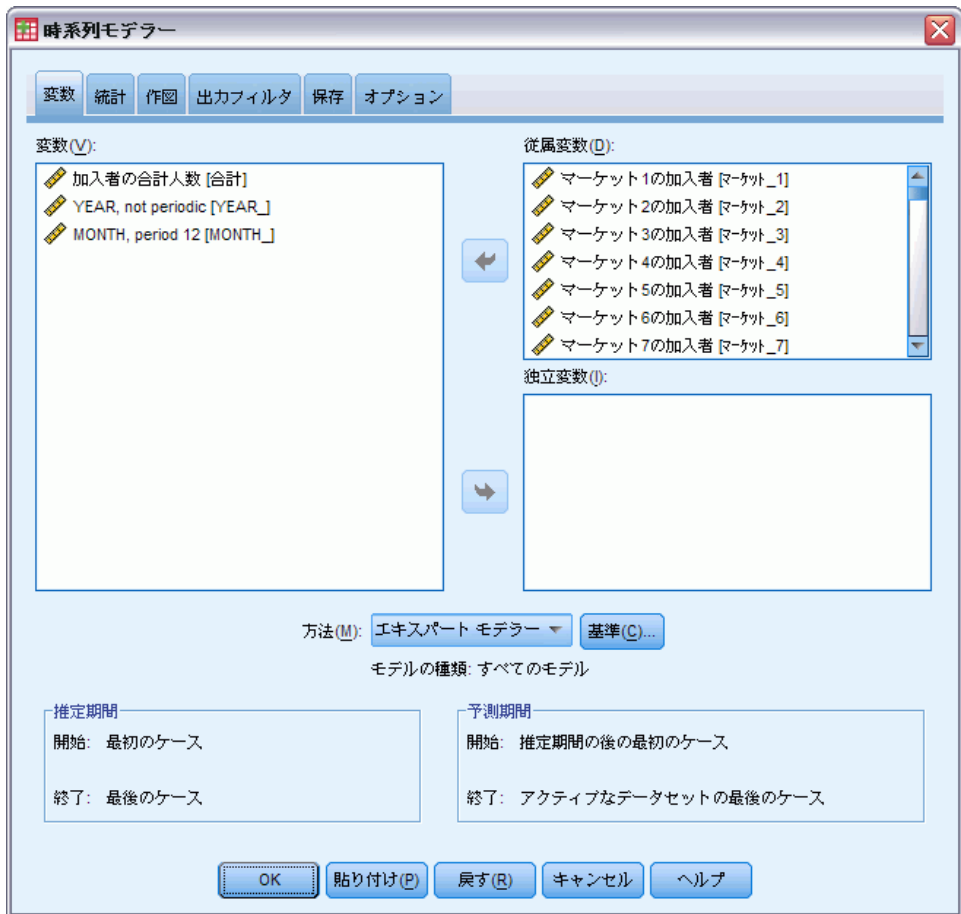
この系列は、非常に滑らかな上昇傾向を示しており、季節変動の存在を示すものではありません。個々の市場について見れば、季節変動を持つ系列も存在する可能性はありますが、データ全般では季節性は顕著な特徴ではないと考えられます。もちろん、季節モデルを除外してしまう前に、個々の系列を調べることは必要です。それにより、季節性が現れる系列を抽出し、それらを別個にモデル化することができます。ここでは、調べてみればわかるように、85 個の系列の中に季節性を示すものは存在しません。

分析の実行

エキスパート モデラーを使用するには

- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > モデルの作成(C)...

図 6-3
[時系列モデラー] ダイアログ ボックス



- ▶ 従属変数として、「マーケット 1 の加入者」から「マーケット 85 の加入者」までの変数を選択します。
- ▶ [方法] ドロップダウン リストで [エキスパート モデラー] が選択されていることを確認します。エキスパート モデラーでは、従属系列のそれぞれに対して最適なモデルが自動的に特定されます。

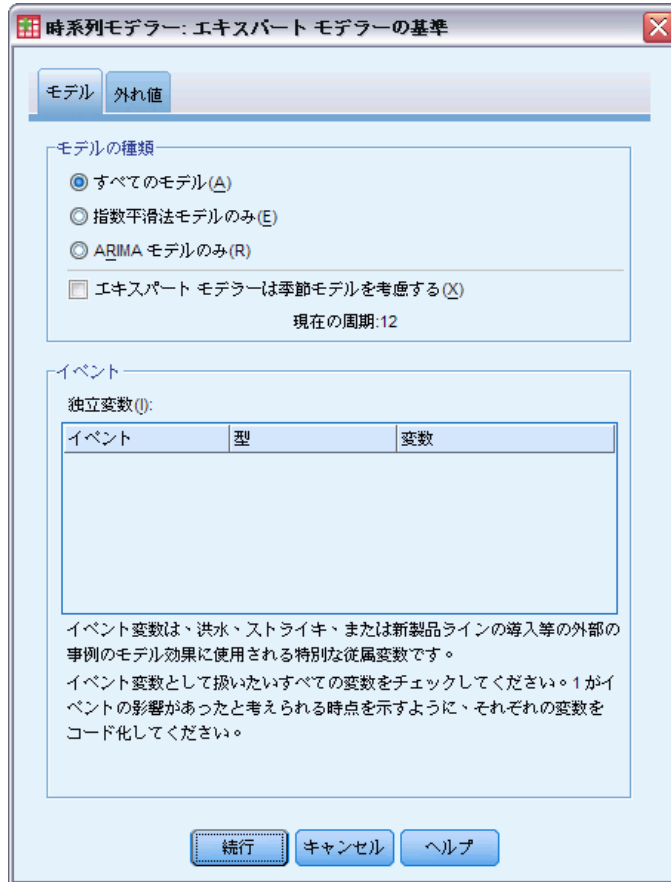
モデルの推定に使用されるケース全体は、**推定期間**と呼ばれます。デフォルトでは、推定期間には、アクティブなデータセットのすべてのケースが含まれます。推定期間を設定するには、[ケースの選択] ダイアログ ボックスで、[日付かケース番号の範囲] を選択します。この例では、デフォルトに従います。

またデフォルトの予測期間は、推定期間の直後から始まり、アクティブなデータセット内の最後のケースまで続きます。最後のケースより後にあるケースについても予測を行う場合は、予測期間を延長する必要があります。

あります。この操作は、この例で後述するように、[オプション] タブから実行できます。

- ▶ [基準] をクリックします。

図 6-4
[エキスパート モデラー基準] ダイアログ ボックスの [モデル] タブ

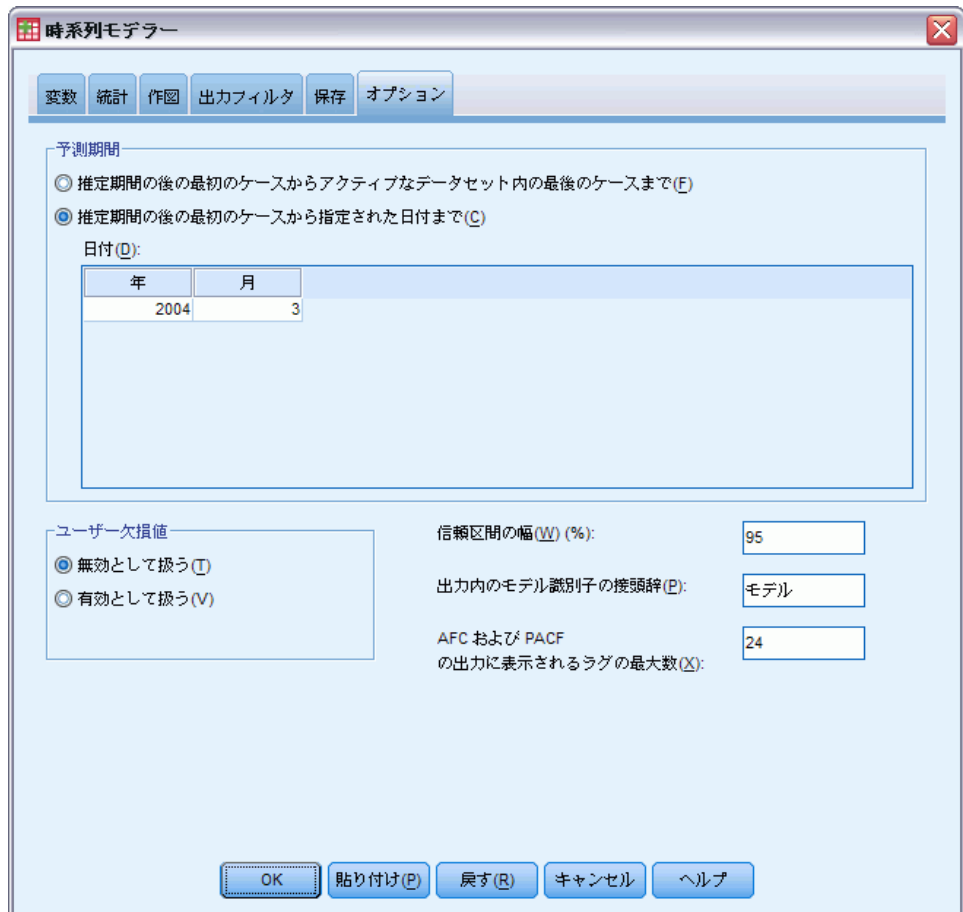


- ▶ [モデルの種類] グループの [エキスパート モデラーは季節モデルを考慮する] の選択を解除します。

データは月次データであり現在の周期は 12 ですが、データには季節性がまったく見られないため、季節モデルを考慮する必要はありません。これにより、エキスパート モデラーが探索するモデルの範囲が狭められるため、処理時間を大幅に短縮できます。

- ▶ [続行] をクリックします。
- ▶ [時系列モデラー] ダイアログ ボックスの [オプション] タブをクリックします。

図 6-5
[時系列モデラー] の [オプション] タブ



- ▶ [予測期間] グループの [推定期間の後の最初のケースから指定された日付まで] を選択します。
- ▶ [日付] グリッドで、年として「2004」、月として「3」を入力します。
データセットには、1999年1月から2003年12月までのデータが含まれているため、上記の設定により、2004年1月から2004年3月までが予測期間となります。
- ▶ [保存] タブをクリックします。

図 6-6
[時系列モデラー] の [保存] タブ



- ▶ [保存] 列の [予測値] を選択し (チェックを入れ)、[変数名の接頭辞] の値にはデフォルトの「予測値」をそのまま使用します。

モデルの予測結果は、新しい変数としてアクティブなデータセットに保存されます。その際、変数名には接頭辞として「予測値」が付加されます。各モデルに関する指定内容は、外部の XML ファイルに保存することもできます。これにより、新しいデータが使用できるようになった時点で、保存したモデルを再利用して、さらに先の予測を行えます。

- ▶ [保存] タブの [参照] ボタンをクリックします。
ファイル保存用の標準のダイアログ ボックスが開きます。
- ▶ XML モデル ファイルを保存するフォルダに移動し、ファイル名を入力して [保存] をクリックします。
- ▶ [統計] タブをクリックします。

図 6-7
[時系列モデラー] の [統計量] タブ



- ▶ [予測値を表示] を選択します。

このオプションを使用すると、各従属変数系列の予測値の表を作成できます。また、予測変数を新しい変数として保存するオプションだけでなく、これらの値を取得するためのオプションも使用できるようになります。

[適合度] ([モデルの比較の統計量] グループ) をデフォルトの選択状態にしておくと、R²乗、平均絶対パーセント誤差、正規化 BIC など、すべてのモデルを対象として計算される適合統計量の表が生成されます。この表には、モデルがどの程度データに適合しているかが簡潔にまとめられています。

- ▶ [作図] タブをクリックします。

図 6-8
[時系列モデラー] の [作図] タブ



- ▶ [個別モデルの作図] グループの [系列] の選択を解除します。

これにより、モデルごとの系列図は生成されなくなります。この例では、予測図を生成することよりも、予測結果を新しい変数として保存することを重視します。

[モデルの比較の作図] グループでは、すべてのモデルについて計算される適合統計量の中から、プロット（ヒストグラム形式）の対象を選択できます（複数可）。

- ▶ [モデルの比較の作図] グループから、[平均絶対パーセント誤差] および [最大絶対パーセント誤差] を選択します。

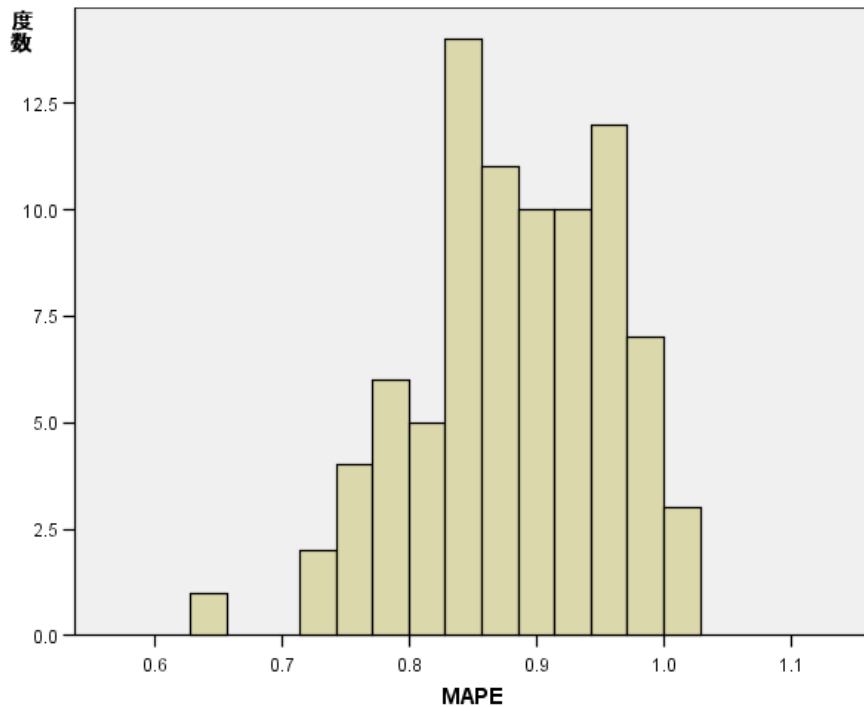
絶対パーセント誤差とは、従属系列がそのモデルの予測水準からどの程度隔たっているのかを表す尺度です。すべてのモデルに関する平均値および最大値を検証することで、予測の不確定性についての目安を得ることがで

きます。従属系列は大小さまざまな市場の加入者数を表しているため、絶対誤差よりもパーセント誤差の要約図を観察するほうが有益です。

- ▶ [時系列モデラー] ダイアログ ボックスで [OK] をクリックします。

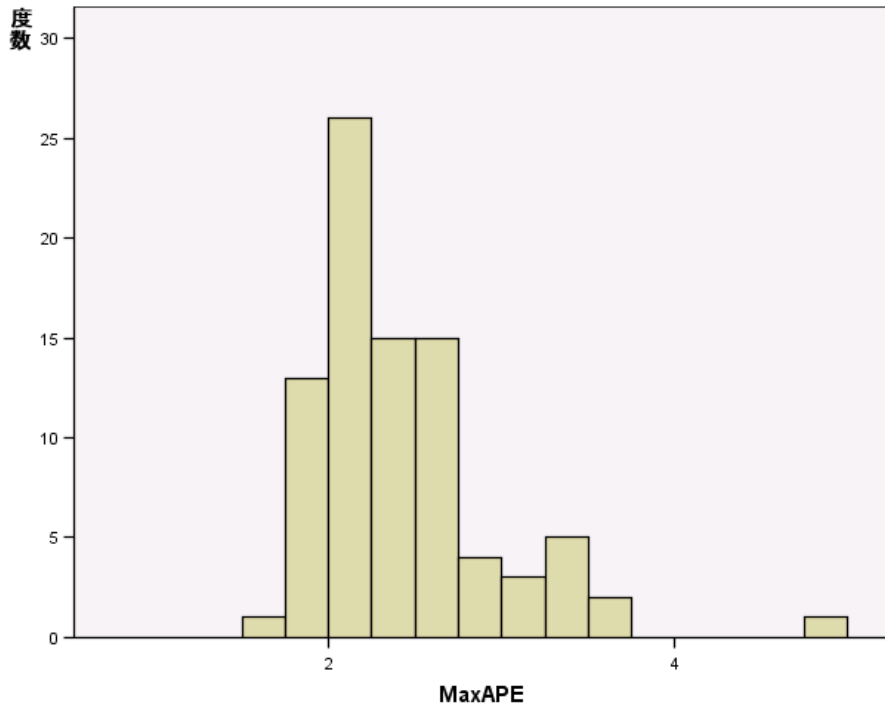
モデルの要約図

図 6-9
平均絶対パーセント誤差のヒストグラム



このヒストグラムには、すべてのモデルについての平均絶対パーセント誤差 (MAPE) が表示されます。ここでは、すべてのモデルについて、不確定性の平均が 1% 前後であることがわかります。

図 6-10
最大絶対パーセント誤差のヒストグラム



このヒストグラムには、すべてのモデルについての最大絶対パーセント誤差 (MaxAPE) が表示されます。これは、予測に対する最悪のケースのシナリオをイメージするのに有効です。ここでは、各モデルについての最大パーセント誤差が 1 ~ 5% の範囲内にあることがわかります。これらの値が表す不確定度は許容できる範囲のものでしょうか。許容できるリスクの程度は問題によって異なるため、この点については各自のビジネスセンスに基づいて判断することになります。

モデルの予測

図 6-11
モデルの予測結果を表す新しい変数

YEAR_	MONTH_	DATE_	予測値_マーケット_1_モデル_1	予測値_マーケット_2_モデル_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

データ エディタに、モデルの予測結果を表す新しい変数が表示されています。ここでは 2 つしか表示されていませんが、新しい変数は 85 個の各従属系列に 1 つずつ対応しており、したがって全部で 85 個あります。変数名は、デフォルトの接頭辞「予測値」、対応する従属変数の名前（たとえば「マーケット_1」）、およびモデル ID（たとえば「モデル_1」）を先頭から順に並べて構成されます。

2004 年 1 月から 2004 年 3 月までの予測結果を含む 3 つの新しいケースがデータセットに追加され、同時にデータ ラベルも自動的に生成されました。新しい変数にはそれぞれ、推定期間（1999 年 1 月から 2003 年 12 月まで）に対するモデル予測値が表示されます。これを見れば、すでにわかっている値に対してモデルがどの程度適合しているかを判断できます。

図 6-12
予測値の表

モデル		1 月1999	2 月1999	3 月1999
マーケット1の加入者-モデル1	予測値	11503	11447	11390
	UCL	11686	11767	11870
	LCL	11321	11126	10910
マーケット2の加入者-モデル2	予測値	54893	55856	56704
	UCL	55632	57195	58575
	LCL	54154	54518	54832
マーケット3の加入者-モデル3	予測値	59656	59305	58954
	UCL	60457	60753	61158
	LCL	58856	57857	56750
マーケット4の加入者-モデル4	予測値	18235	18424	18628
	UCL	18413	18731	19121
	LCL	18058	18116	18136

予測された値の表を作成するよう選択できます。この表には、予測期間内の予測値は表示されますが、モデル予測値を表示する新しい変数とは異なり、推定期間内の予測値は表示されません。結果はモデルごとにまとめられ、モデル名別に表示されます。モデル名は、モデルの予測結果が表示される新しい変数の名前とまったく同じように、対応する従属変数の名前（またはラベル）の後にモデル識別子を付加したものになります。またこの表には、予測値に関する信頼限界の上限（UCL）および信頼限界の下限（LCL）も表示されます（デフォルトは 95%）。

予測値を取得するには、見込みをアクティブ データセットに新しい変数として保存する方法、見込みテーブルを作成する方法の 2 つの方法があります。どちらの方法を使用する場合も、さまざまなオプションを使用して予測値のエクスポート（Excel のスプレッドシートへのエクスポートなど）を行うことができます。

要約

ここでは、エキスパート モデラーを使用して、複数の系列に対する予測値を求める方法について説明しました。また、作成したモデルを外部の XML ファイルに保存しました。次の例では、新しいデータを入手した場合に、[時系列モデルの適用] 手続きを使用して、モデルを再構築することなく、さらに先の予測を行う方法について説明します。

保存したモデルの適用によるバ ルク再予測

これまでに、時系列モデラーを使用して時系列データに関するモデルを作成し、利用可能なデータに基づいて初期予測を行いました。また、より新しいデータを入手した時点で、これらのモデルを再利用してさらに先の予測を行えるよう、モデルを外部ファイルに保存しました。保存したモデルは、これ以降いつでも適用できます。

ここで使用する例は、[p. 50 6 章のエキスパート モデラーによるバルク予測](#)で使用したものを自然な形で拡張したものですが、それとは関係なく単独で使用することもできます。この例では、全国規模のブロードバンド プロバイダから依頼を受けた分析担当者が、85 ある各地方市場ごとに月次のユーザー契約件数を予測するというシナリオが設定されています。ただし、すでにエキスパート モデラーを使用してモデルは作成されており、今後 3 か月の予測結果も出ています。またデータ ウェアハウスには、当初の予測対象であった期間の実際のデータがすでに登録されているため、分析担当者はそれらのデータを使用して、さらに次の 3 か月にまで予測期間を延長したいと考えています。

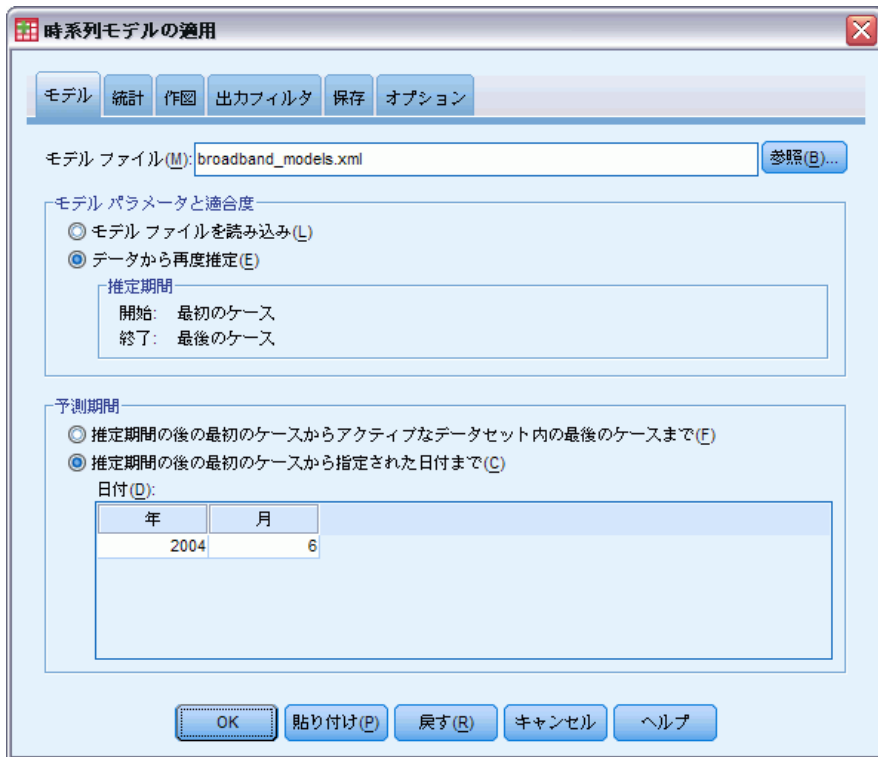
更新された過去の月ごとのデータはすべて、`broadband_2.sav` に収集されています。またモデルは、`broadband_models.xml` に保存されています。[詳細は、D 付録 サンプル ファイル in IBM SPSS Forecasting 21 を参照してください。](#) ただし、前述の例に沿って作業を行い独自のモデル ファイルを保存している場合は、`broadband_models.xml` の代わりにそのファイルを使用することもできます。

分析の実行

モデルを適用するには

- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > モデルの適用(A)...

図 7-1
[時系列モデルの適用] ダイアログ ボックス



- ▶ [参照] をクリックし、broadband_models.xml（または前述の例で保存した独自のモデル ファイル）に移動して選択します。詳細は、D 付録 サンプルファイル in IBM SPSS Forecasting 21 を参照してください。

- ▶ [データから再度推定] を選択します。

時系列の新しい値を予測に組み入れるには、[時系列モデルの適用] 手続きにおいて、モデル パラメータを再推定する必要があります。ただし、モデルの構造に変更はないため、再推定の計算に要する時間は最初にモデルを構成した場合の計算時間に比べ、はるかに短くなります。

再推定に使用するケース群には、新しいデータが含まれている必要があります。デフォルトの推定期間（[最初のケース] から [最後のケース] まで）を使用する場合は、再推定に使用するケース群には必ず新しいデータが含まれます。推定期間をデフォルト以外に設定する必要がある場合は、[ケースの選択] ダイアログ ボックスで [日付かケース番号の範囲] を選択します。

- ▶ [予測期間] グループの [推定期間の後の最初のケースから指定された日付まで] を選択します。
- ▶ [日付] グリッドで、年として「2004」、月として「6」を入力します。

データセットには、1999年1月から2004年3月までのデータが含まれているため、上記の設定により、2004年4月から2004年6月までが予測期間となります。

- ▶ [保存] タブをクリックします。

図 7-2

[時系列モデルの適用] の [保存] タブ



- ▶ [保存] 列の [予測値] を選択 (チェック) し、[変数名の接頭辞] の値にはデフォルトの「予測値」をそのまま使用します。

モデルの予測結果は、新しい変数としてアクティブなデータセットに保存されます。その際、変数名には接頭辞として「予測値」が付加されます。

- ▶ [作図] タブをクリックします。

図 7-3
[時系列モデルの適用] の [作図] タブ



- ▶ [個別モデルの作図] グループの [系列] の選択を解除します。
これにより、モデルごとの系列図は生成されなくなります。この例では、予測図を生成することよりも、予測結果を新しい変数として保存することを重視します。
- ▶ [時系列モデルの適用] ダイアログ ボックスで [OK] をクリックします。

モデルの適合度統計量

図 7-4
モデルの適合度表

適合度の統計量	平均値	SE	最小値	最大値	パーセンタイル							
					5	10	25	50	75	90	95	
適合度の統計量												
定常 R2 乗	.181	.141	-4.34E-015	.609	-9.59E-016	1.22E-016	.070	.193	.252	.359	.453	
R2 乗	.999	.000	.998	1.000	.999	.999	.999	.999	.999	.999	1.000	
RMS E	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 BIC	10.086	1.343	7.749	13.412	8.049	8.285	9.193	9.968	10.717	12.210	12.494	

「モデルの適合度」表には、すべてのモデルを対象として計算された適合度統計量が表示されます。この表には、パラメータを再推定した後のモデルがどの程度データに適合しているかが簡潔にまとめられています。また統計量ごとに、すべてのモデルについて算出した平均値、標準誤差 (SE)、最小値、および最大値が表示されています。さらには、すべてのモデルに対して統計量がどのような分布に従うかについての情報となるパーセンタイル値も表示されます。各パーセンタイルについて、その割合（パーセント）に当たるモデルでは、適合度統計量の値が表示されている値より低くなります。たとえば、全体の 95% に当たるモデルでは、MaxAPE（最大絶対パーセント誤差）の値が 3.676 未満となります。

多くの統計量が報告されるなか、MAPE（平均絶対パーセント誤差）と MaxAPE（最大絶対パーセント誤差）に焦点を当てます。絶対パーセント誤差とは、従属系列がそのモデルの予測水準からどの程度隔たっているかの尺度であり、予測の不確定性を測る指標となります。平均絶対パーセント誤差は、すべてのモデルについて見ると、最小値が 0.669%、最大値が 1.026% となっています。また最大絶対パーセント誤差は、すべてのモデルについて見ると、最小値が 1.742%、最大値が 4.373% となっています。したがって、各モデルの予測に関する不確定性の平均は約 1%、最大値は約 2.5% (MaxAPE の平均値) で、最悪のケースでは約 4% となります。これらの値が不確実度として許容できるものかどうかは、どの程度のリスクを許容するかによって決まります。

モデルの予測

図 7-5
モデルの予測結果を表す新しい変数

YEAR_	MONTH_	DATE_	予測値_マーケット_1_モデル_1	予測値_マーケット_2_モデル_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

データ エディタに、モデルの予測結果を表す新しい変数が表示されています。ここでは 2 つしか表示されていませんが、新しい変数は 85 個の各従属系列に 1 つずつ対応しており、したがって全部で 85 個あります。変数名は、デフォルトの接頭辞「予測値」、対応する従属変数の名前（たとえば「マーケット_1」）、およびモデル ID（たとえば「モデル_1」）を先頭から順に並べて構成されます。

2004 年 4 月から 2004 年 6 月までの予測結果を含む 3 つの新しいケースがデータセットに追加され、同時にデータ ラベルも自動的に生成されました。

要約

より新しいデータを入手した時点で、保存しておいたモデルを再利用し、さらに先の予測を行う方法について説明しました。また、モデルを再構築することなく予測を行いました。もちろん、モデルが変更されたと判断する根拠がある場合は、[時系列モデラー] 手続きを実行してモデルを再構築する必要があります。

エキスパート モデラーによる有意な予測変数の特定

予測モデルの開発に関心を寄せる通信販売会社が、売り上げの変動の説明に使用できると考えられるいくつかの系列とともに、紳士服の毎月の売り上げデータを収集しました。予測変数の候補としては、送付したカタログの数、カタログのページ数、オーダー用に開いた電話回線の数、印刷広告に費やした金額、およびカスタマ サービス スタッフの数があります。これらの中に、予測に有効な予測変数はあるでしょうか。

この例では、エキスパート モデラーを使用して予測変数の候補すべてを調べ、最適なモデルを特定します。エキスパート モデラーでは、従属系列と統計的に有意な関係を持つ予測変数だけが選択されるため、どの予測変数が有効かを知ることができるほか、それらを使用した予測を行うためのモデルを作成できます。この作業が完了したら、この例の続編にあたる [p.80 9 章の保存したモデルの適用による予測変数の検証](#)についても、実際に作業してみることをお勧めします。この例では、ここで構築したモデルを使用して、シナリオに使用する予測変数を変えることで売上にどのような影響があるかについて調べます。

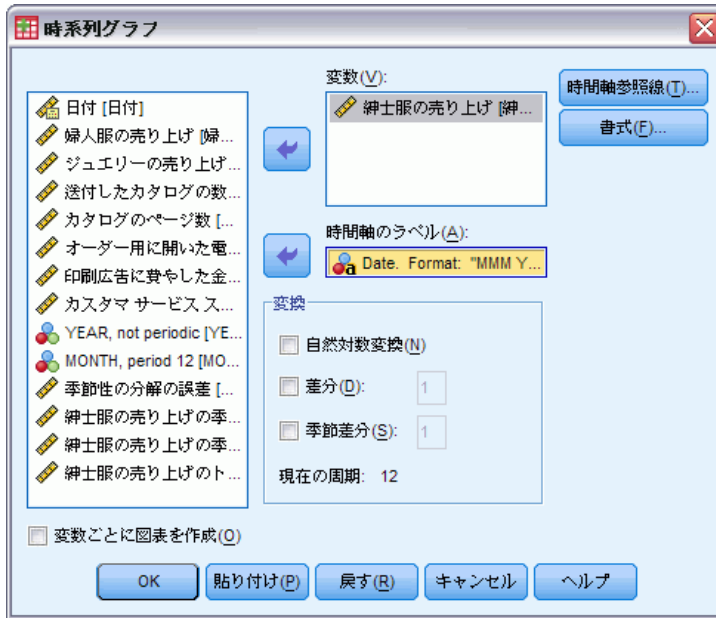
ここでの例で使用するデータは、`catalog_seasfac.sav` に収集されています。詳細は、[D 付録 サンプル ファイル in IBM SPSS Forecasting 21](#) を参照してください。

データのプロット

扱う系列が 1 つだけの場合は特に、次の手順に従ってデータをプロットすると有益です。

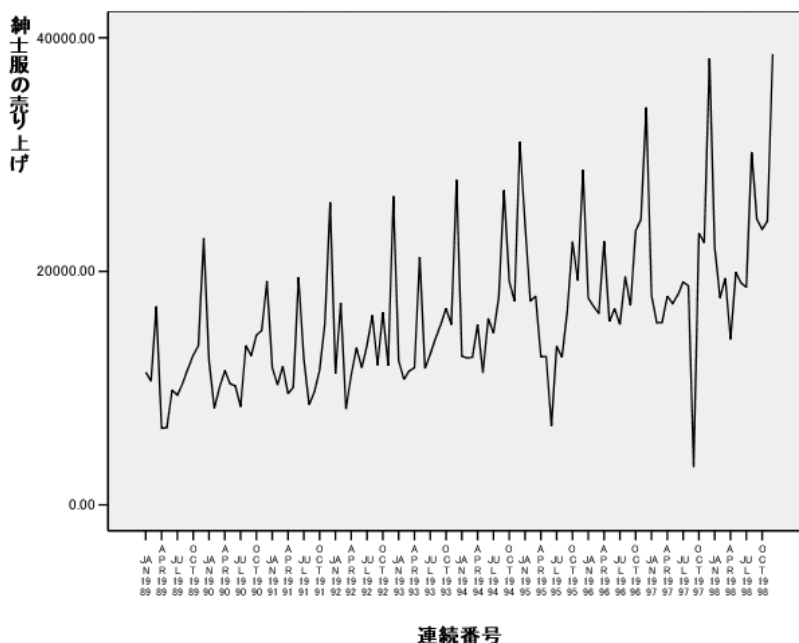
- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > 時系列グラフ...

図 8-1
[時系列グラフ] ダイアログ ボックス



- ▶ 「紳士服の売り上げ」を選択し、[変数] リストに移動します。
- ▶ 「日付」を選択し、[時間軸のラベル] ボックスに移動します。
- ▶ [OK] をクリックします。

図 8-2
紳士服の売り上げ (単位: US ドル)



系列には、多数のピークが表示され、その多くは等間隔に出現しています。また、明らかな上昇傾向が見られます。ピークが等間隔に出現することは、時系列に周期成分が存在することを示唆しています。年末には必ず伸びを見せるなど、売上の季節的な特徴を考えれば、データに年間の季節成分が見出せることは十分に予想できます。

また、季節パターンの中に含まれないピークや、付近のデータ点から著しい隔たりを示すピークもあります。これらの点は、外れ値である可能性もあります。外れ値は、エキスパート モデラーで処理することが可能であり、また処理する必要があります。

分析の実行

エキスパート モデラーを使用するには

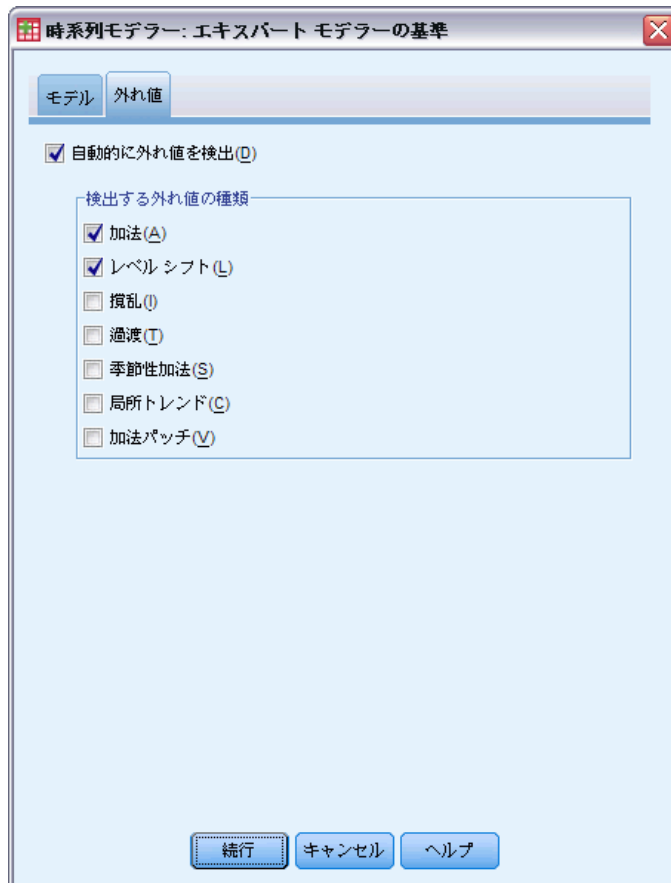
- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > モデルの作成(C)...

図 8-3
[時系列モデラー] ダイアログ ボックス



- ▶ 従属変数として「紳士服の売り上げ」を選択します。
- ▶ 独立変数として、「送付したカタログの数」および「カスタマ サービス スタッフの数」を選択します。
- ▶ [方法] ドロップダウン リストで [エキスパート モデラー] が選択されていることを確認します。エキスパート モデラーでは、従属変数系列に対して最適な季節モデルまたは非季節モデルが自動的に特定されます。
- ▶ [基準] をクリックし、さらに [外れ値] タブをクリックします。

図 8-4
[エキスパート モデラーの基準] ダイアログ ボックスの [外れ値] タブ



- ▶ [自動的に外れ値を検出] を選択し、[検出する外れ値の種類] の各項目の選択状態はデフォルトのままにしておきます。

データを目視で検証する限りでは、外れ値が存在すると思われます。上記のように選択した場合、エキスパート モデラーでは、最も一般的なタイプの外れ値が検索され、検出された外れ値が最終モデルに取り込まれます。外れ値の検出により、エキスパート モデラーに必要な処理時間が大幅に増大することがあります。そのため、多数の系列を同時にモデル化する場合などは特に、この機能は慎重に使用する必要があります。デフォルトでは、外れ値は検出されません。

- ▶ [続行] をクリックします。
- ▶ [時系列モデラー] ダイアログ ボックスの [保存] をクリックします。

図 8-5
[時系列モデラー] の [保存] タブ



モデルを再構築することなく、[時系列モデルの適用] 手続きを使用して予測変数に対するさまざまな値を検証できるようにするために、推定したモデルを外部の XML ファイルに保存します。

- ▶ [保存] タブの [参照] ボタンをクリックします。
ファイル保存用の標準のダイアログ ボックスが開きます。
- ▶ XML モデル ファイルを保存するフォルダに移動し、ファイル名を入力して [保存] をクリックします。
- ▶ [統計] タブをクリックします。

図 8-6
[時系列モデラー] の [統計量] タブ



- ▶ [パラメータ推定値] を選択します。

これにより、有意な予測変数など、エキスパート モデラーで選択されたモデルのすべてのパラメータを表示する表が生成されます。

- ▶ [作図] タブをクリックします。

図 8-7
[時系列モデラー] の [作図] タブ



- ▶ [予測] の選択を解除します。

この例では、有意な予測変数の決定およびモデルの構築についてだけ説明します。予測は行いません。

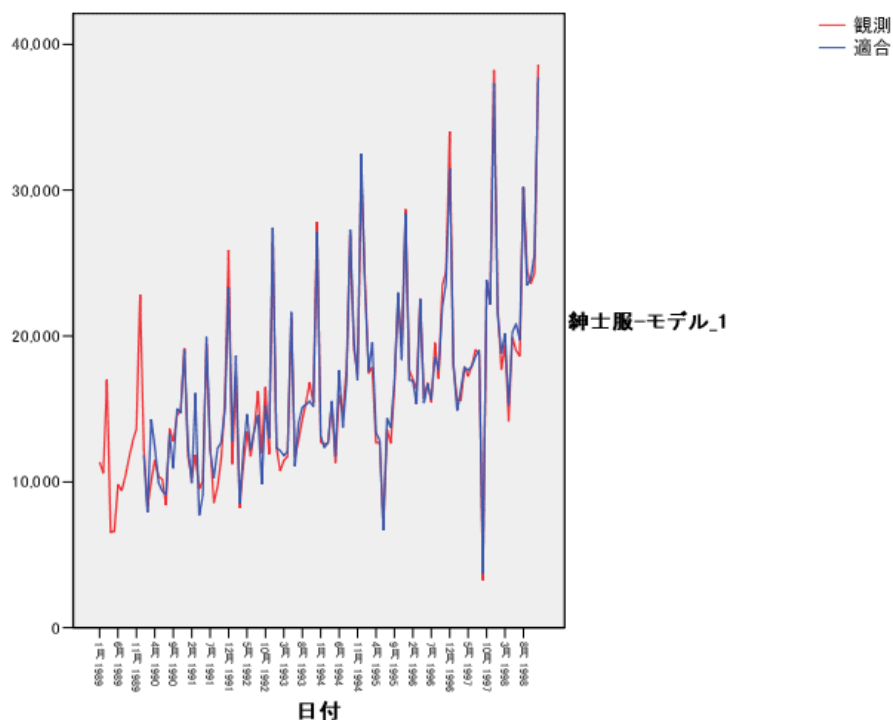
- ▶ [当てはめ値] を選択します。

これにより、モデルの推定に使用される期間内の予測値が表示されます。この期間は**推定期間**と呼ばれ、この例で使用するアクティブなデータセット内のすべてのケースが含まれます。これらの値は、モデルが観測値にどの程度適合しているかを表しており、**当てはめ値**と呼ばれます。作成されるプロットは、観測値と当てはめ値で構成されます。

- ▶ [時系列モデラー] ダイアログ ボックスで [OK] をクリックします。

系列プロット

図 8-8
予測値と観測値



予測値と観測値が非常に一致しています。これは、モデルの予測能力が満足できるものであることを示しています。注目すべきは、このモデルにより季節的なピークが的確に予測されている点です。また、データの上昇傾向が適切に反映されていることにも注目してください。

モデルの説明表

図 8-9
モデルの説明表

モデルの説明			モデルの種類
モデル ID	紳士服の売り上げ	モデル_1	ARIMA(0,0,0)(0,1,0)

モデルの説明表では、推定モデルごとに、モデル識別子とモデルの種類が表示されます。モデル識別子は、対応する従属変数の名前（またはラベル）とシステム割り当て名で構成されます。この例では、従属変数は「紳士服の売り上げ」、システム割り当て名は「モデル_1」です。

時系列モデラーは、指数平滑法モデルと ARIMA モデルをともにサポートしています。指数平滑法モデルの種類は、Holt-Winters の加法など、一般的に使用される名前で一覧表示されます。また ARIMA モデルの種類は、ARIMA (p, d, q) (P, D, Q) という標準的な記法を使用して一覧表示されます。ここで、p は自己回帰の次数、d は差分（または和分）の次数、q は移動平均の次数を表します。また、(P, D, Q) はそれらの季節成分を表します。

エキスパート モデラーでは、紳士服の売り上げを最もよく説明するのは、差分が 1 次の季節 ARIMA モデルであると判定されました。モデルの季節的な特徴が、系列プロットに現れた季節的なピークの説明を与え、1 次の差分が、データに顕著であった上昇傾向を反映しています。

モデル統計量の表

図 8-10
モデル統計量の表

モデル	予測変数の数	モデル適合統計量	Ljung-Box Q(18)			外れ値の数
		定常 R ² 乗	統計量	DF	有意確率	
紳士服の売り上げ-モデル_1	2	.948	7.589	18	.984	9

モデル統計量の表には、推定された各モデルの要約情報と適合度統計量が表示されます。各モデルの結果には、モデルの説明表に表示されるモデル識別子がラベルとして表示されます。まず注目すべきは、モデルには、最初に指定した 5 つの予測変数の候補のうち、2 つの予測変数が含まれていることです。このことから、エキスパート モデラーにより、予測に対して有効だと思われる独立変数が 2 つ特定されたと考えられます。

時系列モデラーでは、さまざまな適合度統計量を出力できますが、その中から定常 R² 乗値だけを選択しました。この統計量は、モデルによって説明される系列の全変動の比率を推定するためのものです。このケースの

ようにトレンドや季節パターンが見られる場合には、通常の R² 乗値よりもこの統計量を使用する方が適しています。定常 R² 乗値が大きいほど（最大値は 1）、適合度が高いことを表しています。値 0.948 は、そのモデルが系列内で観測された変動を非常によく説明していることを示します。

Ljung-Box 統計量（補正 Box-Pierce 統計量ともいう）は、モデルが適切に指定されているかどうかの指標となるものです。有意確率が 0.05 未満であれば、それは、観測された系列の中に、モデルでは説明されない構造が含まれていることを表しています。ここに示されている 0.984 という値は有意ではないため、モデルが適切に指定されていると判断できます。

エキスパート モデラーでは、外れ値と判断された点が 9 個検出されました。これらの各点は、適切にモデル化されているため、系列から除く必要はありません。

ARIMA モデルのパラメータ表

図 8-11
ARIMA モデルのパラメータ表

パラメータ推定値

		推定値	標準誤差	t	近似有意確率
非季節ラグ	AR1	.266	.088	3.013	.003
回帰係数	送付したカタログの数	2.107	.196	10.743	.000
	オーダー用に開いた 電話回線の数	336.823	42.778	7.874	.000
定数		-16821.978	1849.971	-9.093	.000

ARIMA モデルのパラメータ表には、モデル識別子でラベル付けされた推定モデルごとに、モデル内の各パラメータに対する値が表示されます。ここでは、従属変数およびエキスパート モデラーで有意と判断された独立変数を含む、モデル内のすべての変数が一覧表示されています。すでに、モデル統計量の表から、有意な予測変数が 2 つあることがわかっています。モデルのパラメータ表からは、その 2 つの予測変数が「送付したカタログの数」と「オーダー用に開いた電話回線の数」であることがわかります。

要約

ここでは、エキスパート モデラーを使用してモデルを構築した上で有意な予測変数を特定する方法について説明し、作成したモデルを外部ファイルに保存しました。今度は、[時系列モデルの適用] 手続きを使用して、予測系列をその他のシナリオを使って検証し、それらのシナリオによって売り上げ予測がどう影響を受けるのかを観察します。

保存したモデルの適用による予測変数の検証

これまでに、時系列モデラーを使用してデータに関するモデルを作成し、予測に使用する上で有効な予測変数を特定しました。予測変数は制御の対象となる因子を表します。そのため、予測期間内にそれらが取る値を検証し、従属変数の予測にどの程度影響を与えるのかを確認することが必要となります。この作業は、[時系列モデラー] 手続きで作成したモデル ファイルと [時系列モデルの適用] 手続きを使用することで容易に行うことができます。

ここで使用する例は、[p. 69 8 章のエキスパート モデラーによる有意な予測変数の特定](#)で使用した例を自然な形で拡張したのですが、それとは関係なく単独で使うこともできます。シナリオは、ある通信販売会社に関するもので、1989 年 1 月から 1998 年 12 月までの紳士服の月次売上に関するデータがすでに収集してあるほか、将来の売上の予測変数として十分候補となりうるいくつかの系列がすでに取得されています。エキスパート モデラーは、5 つの予測変数のうち、送付したカタログ数と受付用電話回線数の 2 つのみが重要であると判断しました。

さらに、次年度の販売戦略計画を立てる際、財源の用途をカタログの印刷と注文受付用電話回線の維持に限定しました。1999 年の最初の 3 か月の予算は、当初の見通しにプラスして、カタログを 2000 部追加するか、または電話回線を 5 本増設するかのどちらかに使用できるようになっています。この場合、この 3 か月の期間により多くの販売収益を上げるためにはどちらを選択したらよいでしょうか。

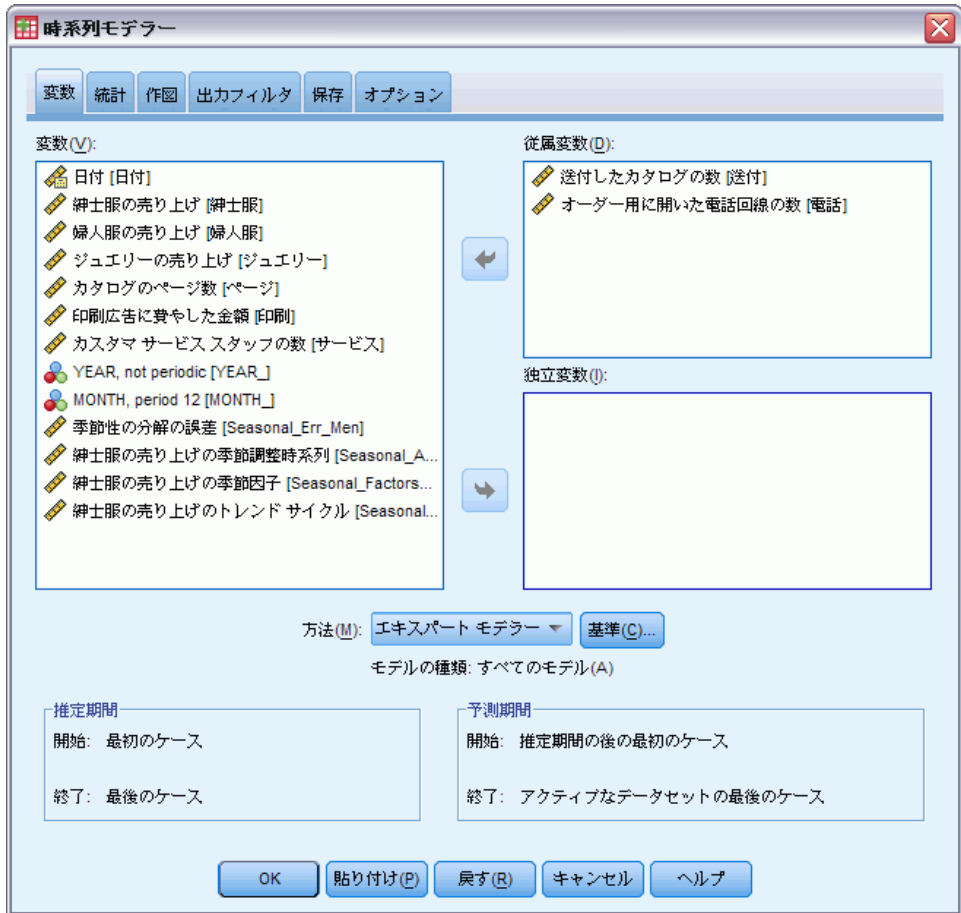
この例で使用するデータは、`catalog_seasfac.sav` に収集されています。また `catalog_model.xml` には、エキスパート モデラーを使用して構築された月次売上のモデルが保存されています。[詳細は、D 付録 サンプルファイル in IBM SPSS Forecasting 21 を参照してください。](#) ただし、前述の例に沿って作業を行い独自のモデル ファイルを保存している場合は、`catalog_model.xml` の代わりにそのファイルを使用することもできます。

予測系列の拡張

予測変数を使用して従属系列の予測値を計算する場合、各予測系列を予測期間にまで拡張する必要があります。予測変数が今後どのような値を取るか正確にわからなくても、推定する必要があります。推定を行った後は、その値を修正しながら、予測変数のさまざまなシナリオを検定できます。初期見通しは、エキスパート モデラーを使用することで、容易に作成できます。

- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > モデルの作成(C)...

図 9-1
[時系列モデラー] ダイアログ ボックス



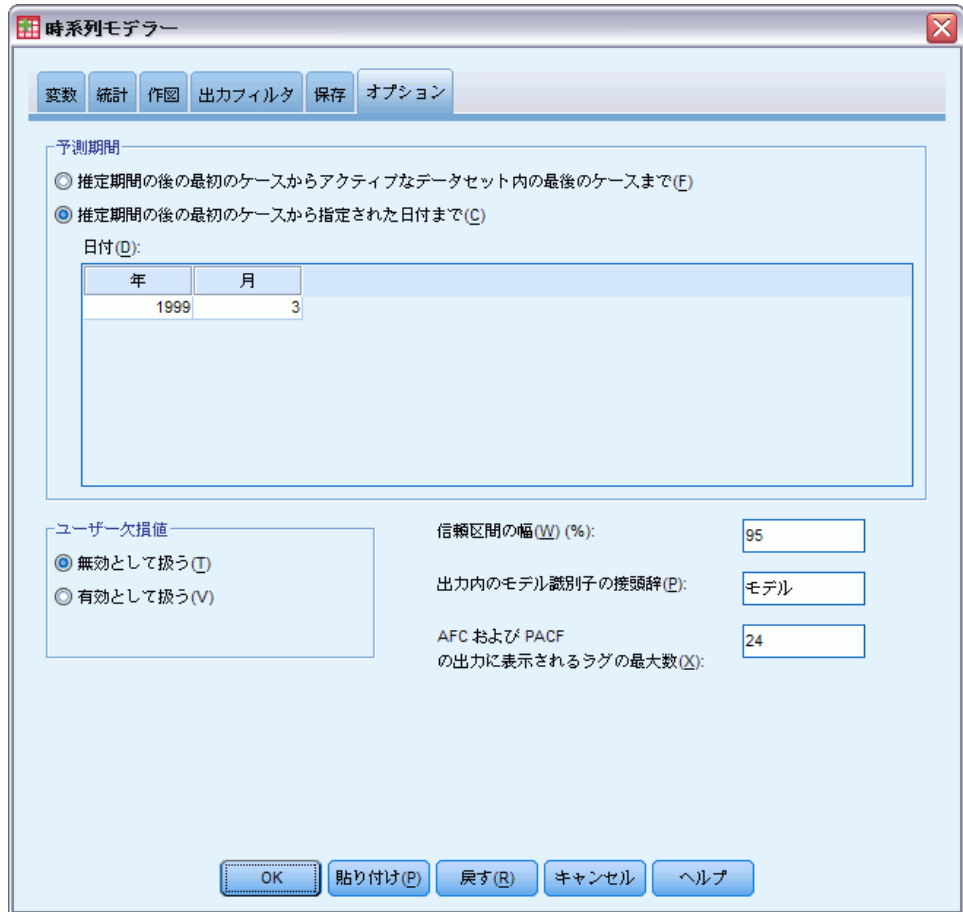
- ▶ 従属変数として、「送付したカタログの数」および「オーダー用に開いた電話回線の数」を選択します。
- ▶ [保存] タブをクリックします。

図 9-2
[時系列モデラー] の [保存] タブ



- ▶ [保存] 列で [予測値] のチェック ボックスを選択し、[変数名の接頭辞] の値にはデフォルトの「予測値」をそのまま使用します。
- ▶ [オプション] タブをクリックします。

図 9-3
[時系列モデラー] の [オプション] タブ



- ▶ [予測期間] グループの [推定期間の後の最初のケースから指定された日付まで] を選択します。
- ▶ [日付] グリッドで、年として「1999」、月として「3」を入力します。
データセットには、1989年1月から1998年12月までのデータが含まれているため、上記の設定により、1999年1月から1999年3月までが予測期間となります。
- ▶ [OK] をクリックします。

図 9-4
予測系列の予測値を表す新しい変数

	予測値_送付_モデル_1	予測値_電話_モデル_2
121	11742	51
122	11853	45
123	11965	45

データ ビュー(D) 変数 ビュー(V)

データ エディタには、新しい変数である、「予測値_送付_モデル_1」と「予測値_電話_モデル_2」が表示されており、各変数には、送付カタログ数および電話回線数のそれぞれに関するモデル予測値が表示されています。予測系列を拡張するのに必要となるのは、1999年1月から1999年3月までの値だけです。これは、ケース121からケース123までの値になります。

- ▶ これら3つのケースの値を「予測値_送付_モデル_1」からコピーし、変数「送付」に追加します。
- ▶ 「予測値_電話_モデル_2」の場合も同様に、最後の3つのケースの値をコピーし、変数「電話」に追加します。

図 9-5
予測期間まで延長された予測系列

	送付	ページ	電話	印刷	サービス	YEAR_	MONTH_	DATE_
121	11742	.	51	.	.	1999	1	JAN 1999
122	11853	.	45	.	.	1999	2	FEB 1999
123	11965	.	45	.	.	1999	3	MAR 1999

データ ビュー(D) 変数 ビュー(V)

これで、予測変数が予測期間にまで拡張されました。

予測期間内の予測値の修正

送付カタログ数の増加および電話回線の増設という2つのシナリオの検定を行うには、それぞれ予測変数「送付」および「電話」の推定値を修正する必要があります。3つのケース（月）に対する予測値だけ修正するので、データ エディタの該当するセルに新しい値を直接入力しても手間はかかりません。ここでは説明を兼ねて、[変数の計算] ダイアログ ボックスを使用します。修正する値の数が多い場合に、この[変数の計算] ダイアログ ボックスは特に役立ちます。

- ▶ メニューから次の項目を選択します。
変換(T) > 変数の計算(C)...

図 9-6
[変数の計算] ダイアログ ボックス



- ▶ 目標変数として、「mail」と入力します。
- ▶ [数式] テキスト ボックスに「mail + 2000」と入力します。
- ▶ [IF] をクリックします。

図 9-7
[変数の計算: IF 条件] ダイアログ ボックス

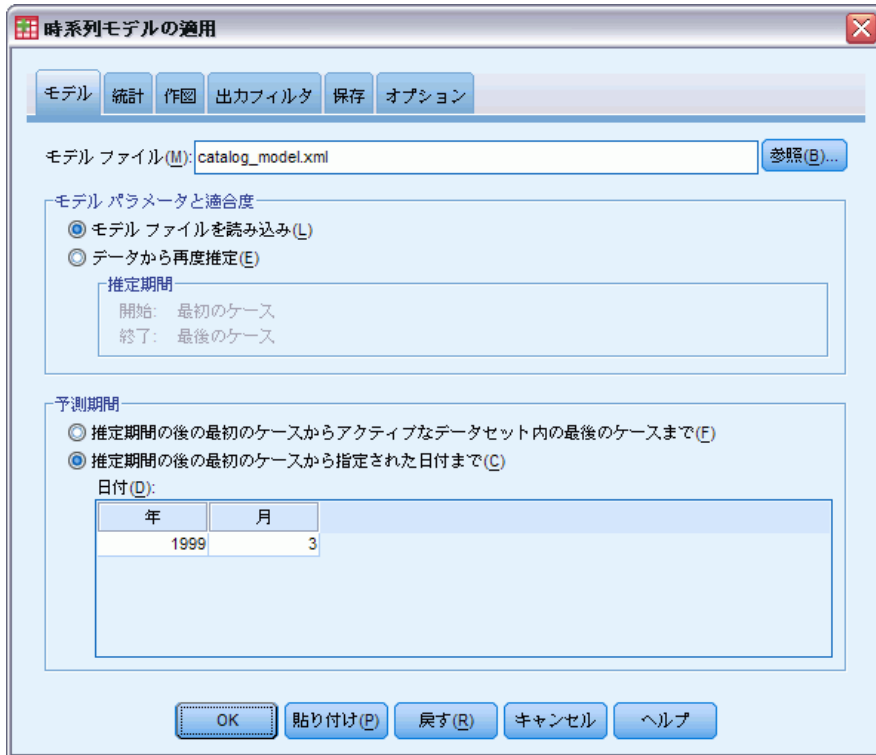


- ▶ [If 条件を満たしたケースを含む] チェック ボックスをオンにします。
- ▶ テキスト ボックスに「\$CASENUM > 120」と入力します。
これにより、変数「送付」を変更できるのは、予測期間内のケースに限られます。
- ▶ [続行] をクリックします。
- ▶ [変数の計算] ダイアログ ボックスで [OK] をクリックし、現在の変数を変更するかどうかを尋ねるメッセージが表示されたら、もう一度 [OK] をクリックします。
これにより、予測期間の 3 つの月のそれぞれについて、「送付」（送付したカタログの数）の値が 2000 ずつ増加します。これで、1 つめのシナリオを検定するためのデータがそろい、分析を実行する準備が整いました。

分析の実行

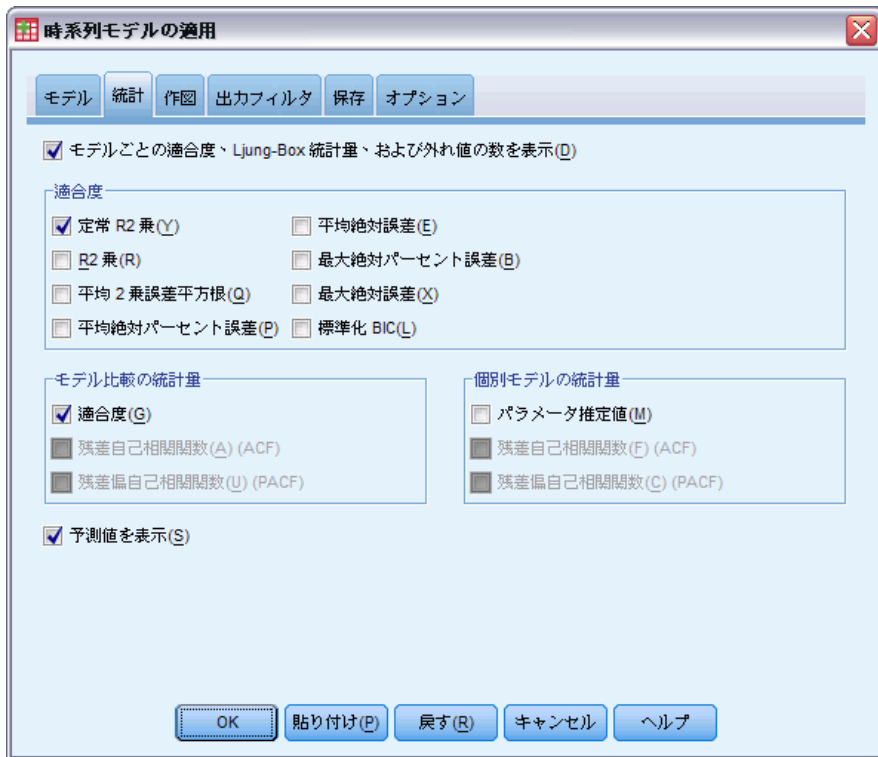
- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > モデルの適用(A)...

図 9-8
[時系列モデルの適用] ダイアログ ボックス



- ▶ [参照] をクリックし、catalog_model.xml または（前述の例で保存した）独自のモデル ファイルに移動して選択します。詳細は、D 付録 サンプル ファイル in IBM SPSS Forecasting 21 を参照してください。
- ▶ [予測期間] グループの [推定期間の後の最初のケースから指定された日付まで] を選択します。
- ▶ [日付] グリッドで、年として「1999」、月として「3」を入力します。
- ▶ [統計] タブをクリックします。

図 9-9
[時系列モデルの適用] の [統計量] タブ



- ▶ [予測値を表示] を選択します。
これにより、従属変数に対する予測値の表が作成されます。
- ▶ [時系列モデルの適用] ダイアログ ボックスで [OK] をクリックします。

図 9-10
予測値の表

モデル		1 月1999	2 月1999	3 月1999
紳士服の売り上げ-モデル1	予測値	25279.91	22064.72	21580.96
	UCL	27591.62	24376.42	23892.66
	LCL	22968.21	19753.02	19269.25

予測値の表には、予測期間内における 2 つの予測変数「送付」および「電話」の値も考慮に入れた、従属系列に対する予測値が表示されます。またこの表には、予測に対する信頼限界の上限 (UCL) および信頼限界の下限 (LCL) も表示されます。

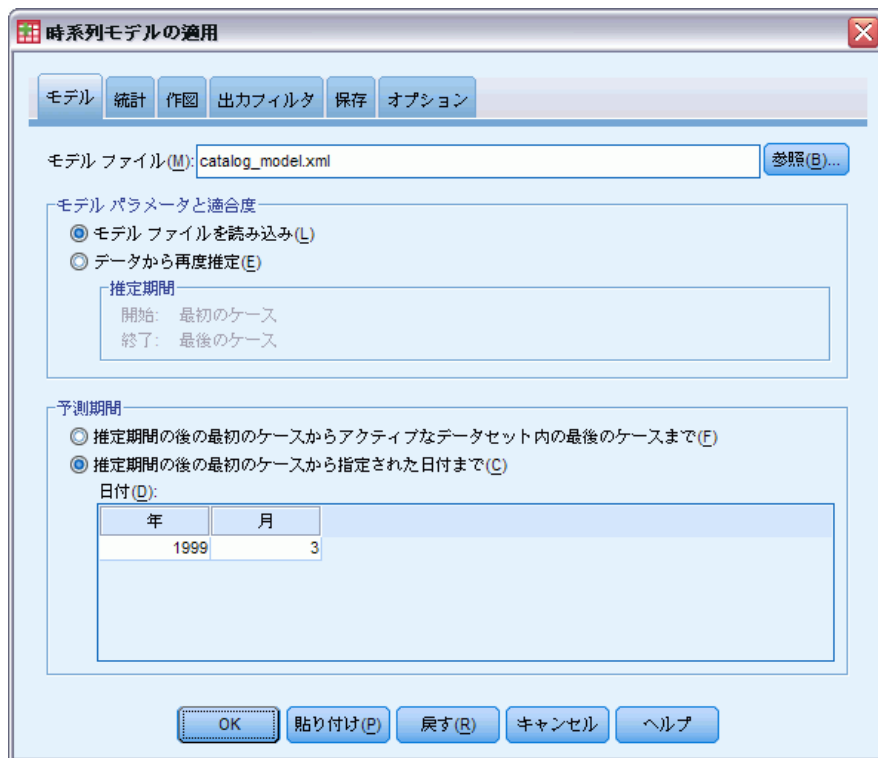
送付するカタログ数を毎月 2000 ずつ増やすというシナリオにおける売上の予測値が得られました。次は、電話回線数を増やすというシナリオについてのデータの準備をします。そのために、変数「送付」を元の値にリセットし、変数「電話」の値を 5 ずつ増やします。「送付」の値をリ

セットするには、予測期間内の「予測値_送付_モデル_1」の値をコピーし、予測期間内の「送付」の現在の値の上に貼り付けます。電話回線の数を（予測期間内のそれぞれの月について 5 ずつ）増やす場合も、カタログ数の場合と同様、データ エディタに直接値を入力するか、または [変数の計算] ダイアログ ボックスを使用します。

分析を行うには、再度 [時系列モデルの適用] ダイアログ ボックスを開いて、次の操作を行います。

- ▶ [ダイアログのリコール] ツールバー ボタンをクリックします。
- ▶ [時系列モデルの適用] を選択します。

図 9-11
[時系列モデルの適用] ダイアログ ボックス



- ▶ [時系列モデルの適用] ダイアログ ボックスで [OK] をクリックします。

図 9-12
2 つのシナリオに対する予測値の表

カタログ予測

モデル		1月1999	2月1999	3月1999
紳士服の売り 上げ-モデル1	予測値	25279.91	22064.72	21580.96
	UCL	27591.62	24376.42	23892.66
	LCL	22968.21	19753.02	19269.25

電話線予測

モデル		1月1999	2月1999	3月1999
紳士服の売り 上げ-モデル1	予測値	23757.25	20542.06	20058.29
	UCL	26068.95	22853.76	22370.00
	LCL	21445.55	18230.35	17746.59

2 つのシナリオに対する予測値の表を表示すると、予測対象となった 3 つの月のそれぞれについて、送付するカタログ数を増やしたほうが、オーダー用の電話回線を増設するよりも、利益が約 1500 ドル多くなると期待されます。この分析を基にすれば、送付するカタログの数を 2000 部増やすことに財源を割り当てた方が賢明のようです。

季節性の分解

売上データからの季節性の除去

ある通信販売会社は、いくつかの予測変数（送付したカタログの数やオーダー用にかいた電話回線の数など）に基づいて、紳士服商品の売り上げの増加傾向をモデリングすることに関心を持っています。このため、この会社では過去 10 年分の紳士服の月次売り上げデータを収集しました。この情報は、catalog.sav に収集されています。詳細は、D 付録 サンプルファイル in IBM SPSS Forecasting 21 を参照してください。

トレンド分析を実行するには、データ内にあるすべての季節変動を取り除く必要があります。この処理は、季節性の分解手続きを使用することで容易に行えます。

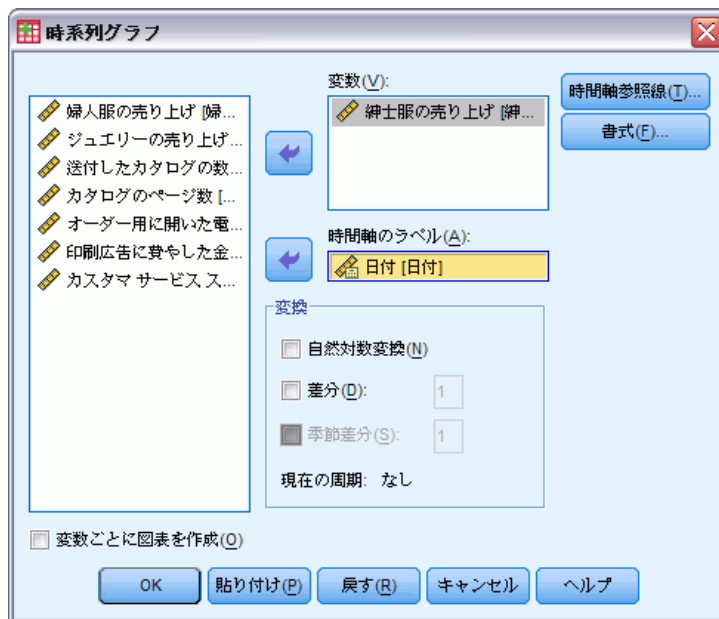
周期性の調査と設定

季節性の分解手続きを実行するには、アクティブなデータセットの中に周期的な日付成分が存在している必要があります。たとえば、年間の周期性であれば 12（月数）、週間の周期性であれば 7（日数）などです。時系列プロットを表示することで、基本的な周期性を的確に推測できることが多いため、まずは時系列のプロットから始めることをお勧めします。

紳士服の売り上げを時系列でプロットするには

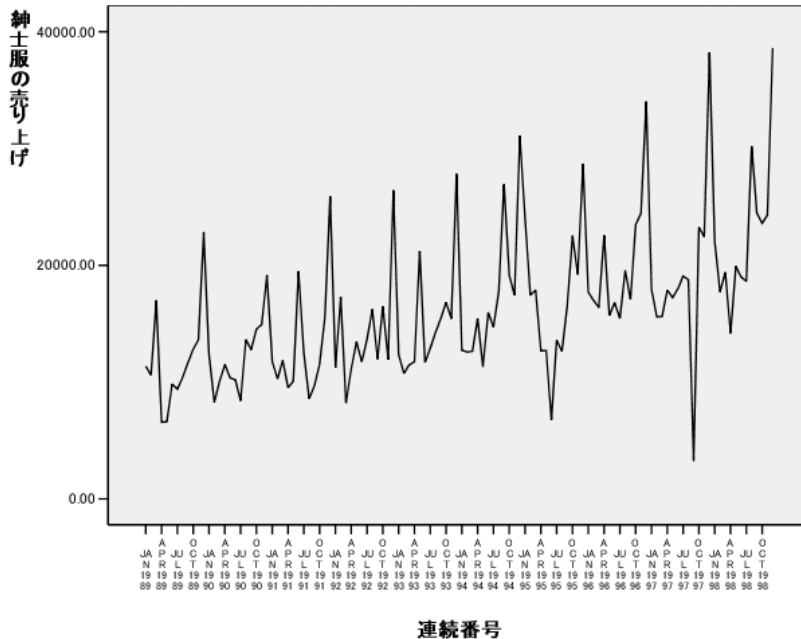
- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > 時系列グラフ...

図 10-1
[時系列グラフ] ダイアログ ボックス



- ▶ 「紳士服の売り上げ」を選択し、[変数] リストに移動します。
- ▶ 「日付」を選択し、[時間軸のラベル] リストに移動します。
- ▶ [OK] をクリックします。

図 10-2
紳士服の売り上げ (単位: US ドル)

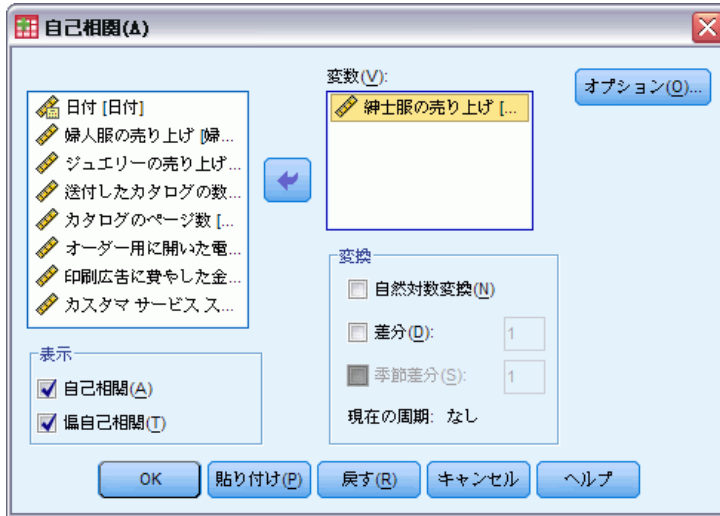


この時系列から、売上げのピーク回数はわかりますが、その間隔は等しくありません。この図は、系列に周期成分があれば、そこには非周期的な変動も伴うことを示唆しています。これは、実際の系列でも一般的に見られる現象です。小さな変動を除けば、有意なピーク間には数か月以上の開きがあります。年末には必ず伸びを見せるなど、売上げの季節的な特徴を考慮すれば、この時系列には年間の周期性があると考えられます。また、系列の上昇傾向が進むにつれて季節変動が大きくなる様子も見られます。このことは、季節変動が時系列のレベルに比例しており、ここでは相加的モデルよりも相乗的モデルの方が適していることを示唆しています。

時系列の自己相関と偏自己相関を調べると、基本的な周期性について量的な結論を得られます。

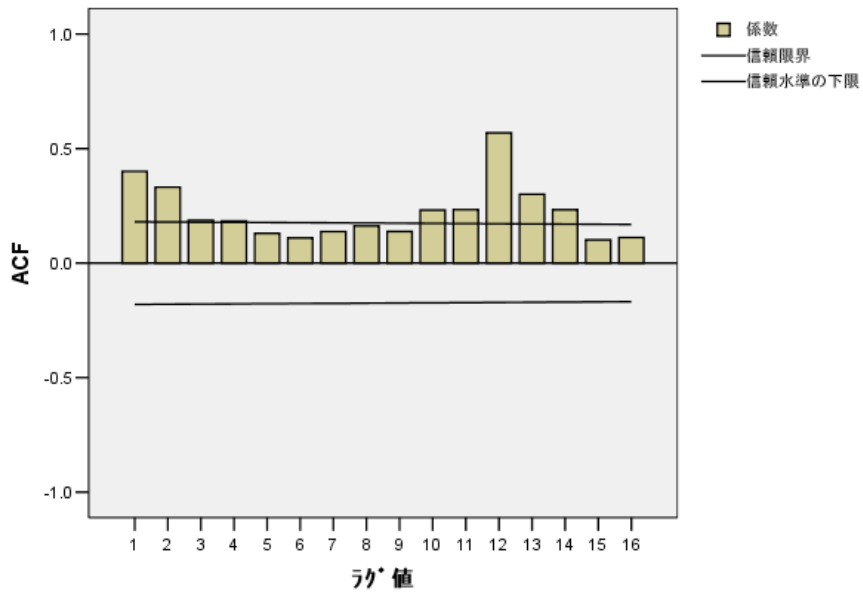
- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > 自己相関(O)...

図 10-3
[自己相関] ダイアログ ボックス



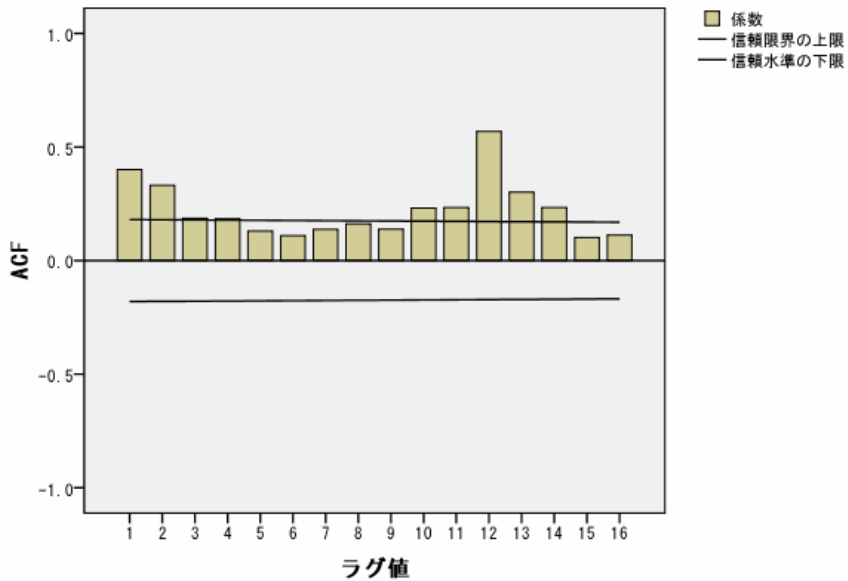
- ▶ 「紳士服の売り上げ」を選択し、[変数] リストに移動します。
- ▶ [OK] をクリックします。

図 10-4
紳士服の自己相関プロット



自己相関関数では、ラグが 1 で、指数の裾が長い場合に有意なピークが表示されます。これは、時系列でよく見られるパターンです。ラグ 12 の場合の有意なピークは、データ内に年間の季節成分が存在することを示します。偏自己相関関数を調べると、より信頼性の高い結論を得られます。

図 10-5
紳士服の偏自己相関プロット

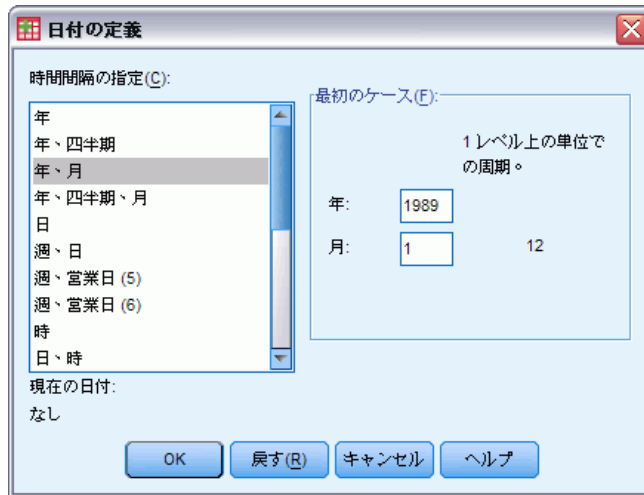


偏自己相関関数では、ラグ 12 の有意なピークは、データ内に年間の季節成分が確実に存在することを示します。

年間の周期性を設定するには、次の手順を実行します。

- ▶ メニューから次の項目を選択します。
データ > 日付の定義(E)...

図 10-6
[日付の定義] ダイアログ ボックス



- ▶ [時間間隔の指定] リストから [年、月] を選択します。
- ▶ 年として「1989」、月として「1」を入力します。
- ▶ [OK] をクリックします。

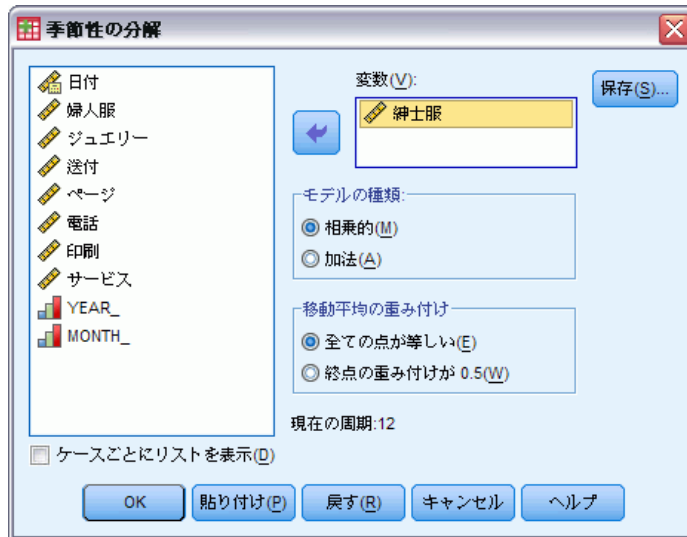
この操作により、周期が 12 に設定され、予測の手続きで使用するための日付変数のセットが作成されます。

分析の実行

季節性の分解手続きを実行するには、次の手順を実行します。

- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > 季節性の分解(S)...

図 10-7
[季節性の分解] ダイアログ ボックス



- ▶ ソース変数リストの任意の場所を右クリックし、コンテキストメニューから [変数名を表示] を選択します。
- ▶ 「紳士服」を選択し、[変数] リストに移動します。
- ▶ [モデルの種類] グループから [相乗的] を選択します。
- ▶ [移動平均の重み付け] グループから [全ての点が等しい] を選択します。
- ▶ [OK] をクリックします。

出力について

季節性の分解手続きでは、手続きで分析した元の変数ごとに 4 つの新しい変数が作成されます。デフォルトでは、これらの新しい変数はアクティブなデータセットに追加されます。これらの新しい変数は、次の接頭辞で始まります。

SAF。 季節変動を表す季節調整因子。相乗的モデルの場合、値 1 は季節変動がないことを表しますが、相加的モデルの場合、値 0 が季節変動がないことを表します。

SAS。 季節変動を取り除いた元の系列を表す季節調整時系列。たとえば、季節調整時系列を使用して作業すると、トレンド成分を切り離して、すべての季節成分から独立した分析を行えます。

STC。 平滑化したトレンド サイクル成分（トレンド成分と周期成分の両方を示す季節調整時系列を平滑化したもの）。

ERR。 特定の観測値に対する系列の残差成分。

このケースでは、季節変動を取り除いた元の系列を表す季節調整時系列が最適な変数です。

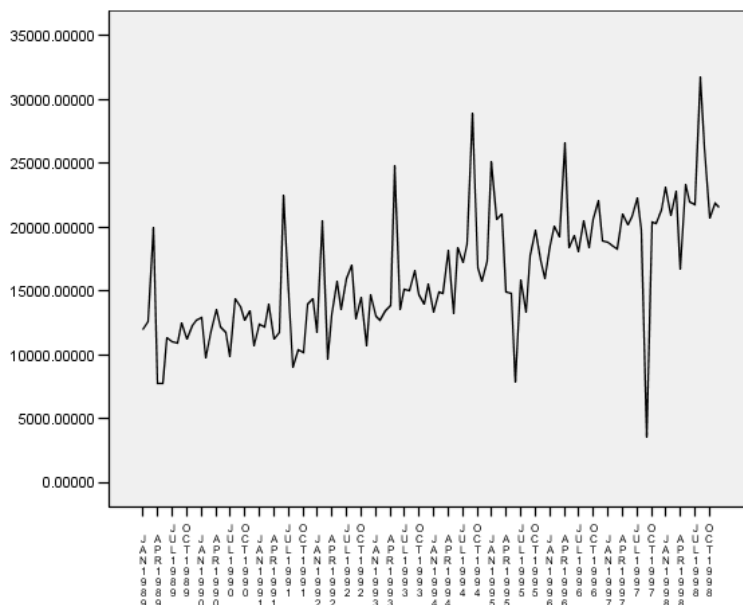
図 10-8
[時系列グラフ] ダイアログ ボックス



季節調整時系列をプロットするには、次の手順を実行します。

- ▶ [時系列グラフ] ダイアログ ボックスを開きます。
- ▶ [戻す] をクリックし、これまでの選択をすべてクリアします。
- ▶ ソース変数リストの任意の場所を右クリックし、コンテキスト メニューから [変数名を表示] を選択します。
- ▶ 「SAS_1」を選択し、[変数] リストに移動します。
- ▶ [OK] をクリックします。

図 10-9
季節調整時系列



季節調整時系列では、明確な上昇傾向が示されます。ピークが多数存在することは明らかですが、それらの間隔に規則性は見られないため、明確な年間パターンは示されません。

要約

季節性の分解手続きを使用して、周期的な時系列の季節成分を取り除き、トレンド分析に適した系列を作成しました。時系列の自己相関と偏自己相関を調べることは、基本的な周期性（ここでは年間の周期性）を判断するのに役立ちました。

関連手続き

季節性の分解手続きは、周期的な時系列から単一の季節成分を取り除く場合に役立ちます。

- 偏相関関数によって提供された時系列の周期のより深い分析を実行するには、スペクトラルプロットの手順を実行します。詳細は、「11章」を参照してください。

スペクトルプロット

スペクトルプロットを使用した周期性に関する期待値の確認

通常、小売店の売り上げを表す時系列には、売り上げのピークが休暇シーズンに見られる傾向があるので、基本的な年間周期性があるといえます。売り上げの投影を作成することは、時系列のモデルを構築することになります。つまり、すべての周期成分を識別することになります。時系列には、その基本構造をあいまいにするランダムな変動が含まれるため、時系列のプロットを表示しても、年間の周期性を見つけ出すことができるとは限りません。

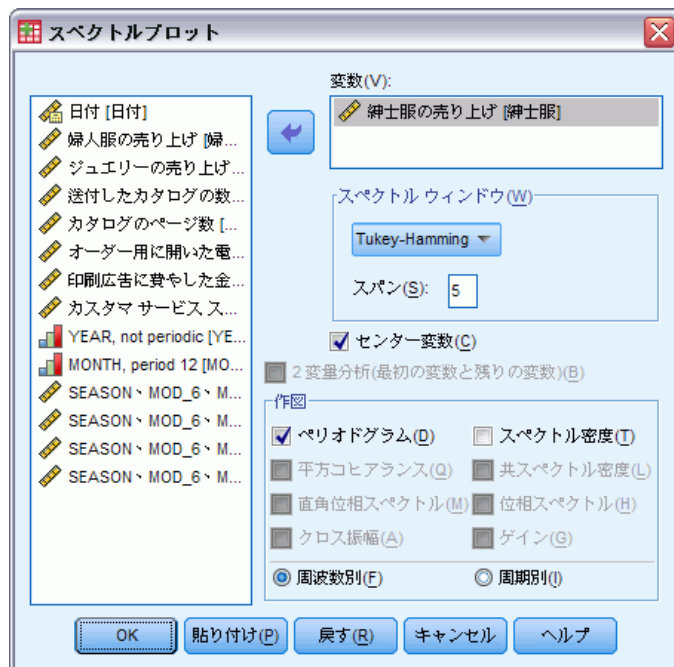
ある通信販売会社の月次売上データは、`catalog.sav` に保存されています。詳細は、[D 付録 サンプル ファイル in IBM SPSS Forecasting 21](#) を参照してください。売り上げの見通しを作成する前に、売り上げデータに年間周期性が現れていることを確認する必要があります。時系列のプロットには、多数のピークが不規則な間隔で現れるため、基本的な周期性は明らかではありません。スペクトルプロット手続きを使用して、売り上げデータの周期性を識別します。

分析の実行

スペクトルプロット手続きを実行するには、次の手順を実行します。

- ▶ メニューから次の項目を選択します。
分析(A) > 見込み > スペクトル分析(T)...

図 11-1
[スペクトルプロット] ダイアログ ボックス



- ▶ 「紳士服の売り上げ」を選択し、[変数] リストに移動します。
- ▶ [作図] グループから [スペクトル密度] を選択します。
- ▶ [OK] をクリックします。

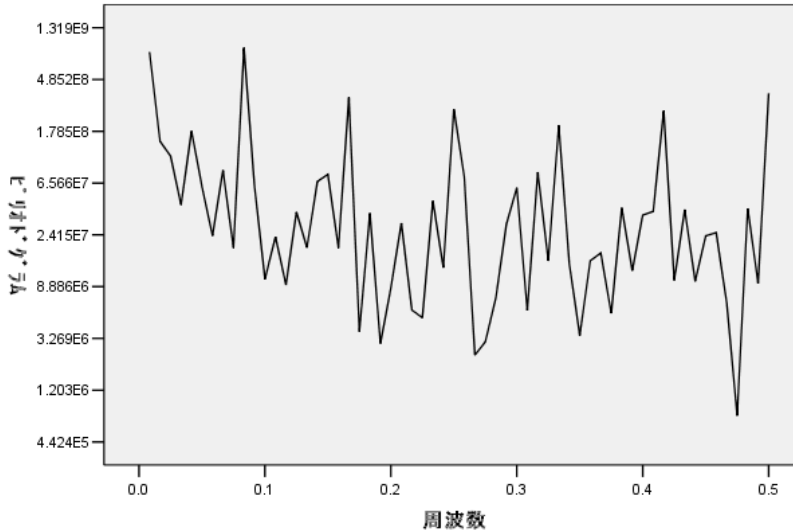
以上の選択により、次のコマンド シンタックスが生成されます。

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

出力に 1 変量統計統計テーブルを追加するには、TSET コマンドを変更して、TSET PRINT=DETAILED を読み込む必要があります。

ペリオドグラムとスペクトル密度について

図 11-2
ペリオドグラム(D)



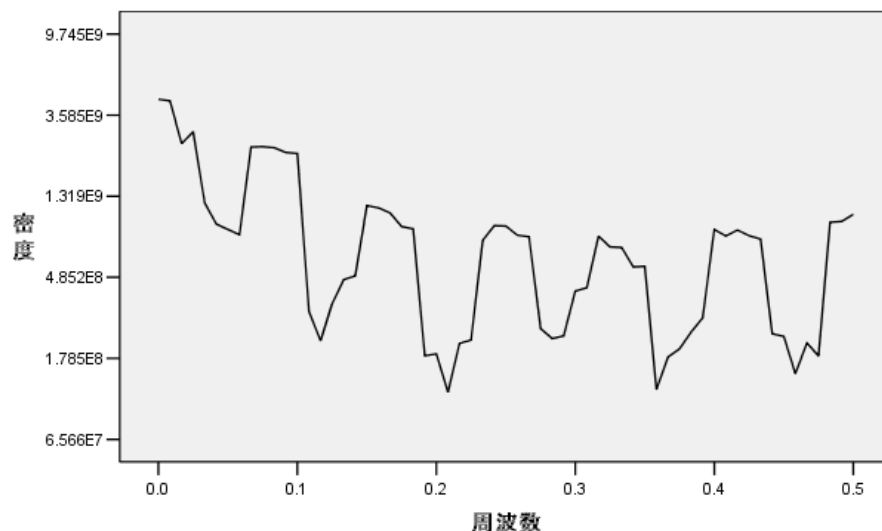
ペリオドグラムのプロットは、バックグラウンド ノイズから突出している最低周波数のピークがちょうど 0.1 以下のときに存在する一連のピークを示します。データに年間の周期成分が含まれている可能性がある場合、年間の周期成分がペリオドグラムに与える影響を考慮する必要があります。時系列の各データ点は月を表しているので、このデータセットでは年間の周期性は 12 の周期に対応します。周波数と周期は互いに逆数の関係にあるため、周期 12 は周波数 $1/12$ (つまり 0.083) に対応します。したがって、ペリオドグラムでは年間の周期成分は周波数が 0.083 の場合にピークが存在することを示します。これは、周波数がちょうど 0.1 以下の場合にピークが存在することと一致しています。

図 11-3
1 変量統計テーブル

Series Name: 紳士服の売り上げ					
	度数	正弦変換 (サイン変換)	コサイン変換	ペリオドグラム	スペクトル密度 推定値
1	.00000	.000	16242.813	.000	4360770849
2	.00833	-3696.643	370.153	828131181.656	4278121402
3	.01667	-1496.277	454.831	146743037.650	2526314733
4	.02500	-1336.400	252.087	110970820.672	2920767853
5	.03333	-662.146	529.734	43143315.613	1210312726
6	.04167	-1654.614	501.571	179359225.256	935924937.4
7	.05000	-784.814	-636.729	61281367.291	875492028.5
8	.05833	-335.646	532.062	23744854.980	820375351.9
9	.06667	-1094.178	-451.489	84064108.939	2420402781
10	.07500	264.554	492.876	18774933.531	2428984191
11	.08333	-3053.934	2370.483	896742148.972	2400843900
12	.09167	-978.882	-287.035	62435897.565	2263220346
13	.10000	-403.128	93.036	10270064.412	2235491523

1 変量統計テーブルには、ペリオドグラムをプロットする場合に使用するデータ点が含まれています。0.1 未満の周波数について見ると、[ペリオドグラム] 列が最大値を取るのは周波数が 0.08333 の場合です。これは、年間周期成分があるという仮定のもとで予測した結果と完全に一致します。これにより、年間の周期成分がある場合に周波数が最低となるピークを確認できます。今度は、他の周波数が高い場合のピークについて見てみましょう。

図 11-4
スペクトル密度(T)



その他のピークの分析には、スペクトル密度関数を使用するのが最適な方法です。これは、単にペリオドグラムを平滑化したものです。平滑化により、ペリオドグラムからバックグラウンド ノイズを取り除き、基本構造をより明確に分離させることができます。

スペクトル密度は、等間隔で現れる 5 つの異なるピークで構成されています。最低周波数のピークは、単に周波数が 0.08333 の場合ピークを平滑化したものを示します。他の 4 つの周波数のピークの有意確率を判断するには、ペリオドグラムがコサイン関数とサイン関数の合計として時系列をモデリングして計算されていることを覚えておいてください。サイン関数またはコサイン関数（正弦曲線）の形をした周期成分は、ペリオドグラムでは単一のピークとして現れます。正弦曲線を描かない周期成分は、等間隔で高さの異なる一連のピークとして現れます。この際、これらの一連のピークのうち最低周波数のピークは周期成分と同じ周波数で発生します。したがって、スペクトル密度では、最低周波数以外の 4 つの周波数のピークは、年間の周期成分が正弦曲線を描かないことを示しています。

ここでは、スペクトル密度プロットで目に見える構造を説明し、このデータには 12 か月周期の単一の周期成分が含まれていると結論付けられます。

要約

[スペクトルプロット] 手続きを使用して、1 つの時系列に 1 つの年間の周期成分があることを確認し、他の有意な周期性が存在しないことを確認しました。またスペクトル密度は、データ内にある非周期成分による変動を平滑化するため、基本構造を見出す上ではペリオドグラムよりも有効であることがわかりました。

関連手続き

スペクトルプロット手続きは、時系列の周期成分を識別するのに役立ちます。

- ートレンド分析を実行するなど、時系列から周期成分を削除するには、季節性の分解の手順を実行します。詳細は、[10 章](#)を参照してください。

適合度統計量

この項では、時系列モデリングで使用される適合度統計量を定義します。

- **定常 R-2 乗**. モデルの定常部分と単純平均モデルを比較する測定。トレンドか季節のパターンがある場合は、この寸法は通常の R2 乗に推奨されます。定常的 R2 乗は、負の無限大から 1 までの範囲で負になる場合があります。負の値は、検討中のモデルがベースラインモデルより悪い場合を意味します。正の値は、検討中のモデルがベースラインモデルより良い場合を意味します。
- **R-2 乗**. モデルで説明される系列のすべての変動の割合（寄与率）の推定値。系列が静止している場合、この測定法はとても役に立ちます。R2 乗は、負の無限大から 1 までの範囲で負になる場合があります。負の値は、検討中のモデルがベースラインモデルより悪い場合を意味します。正の値は、検討中のモデルがベースラインモデルより良い場合を意味します。
- **平均 2 乗誤差平方根 (RMSE)**. 平均 2 乗誤差平方根。平均平方誤差の平方根。従属系列がそのモデルの予測水準からどれくらい変るかを測定。従属系列と同じ単位で表現されます。
- **平均絶対パーセント誤差 (MAPE)**. 平均絶対パーセント誤差。従属系列がそのモデルの予測水準からどれくらい変るかを測定。使用されている単位に依存しないので、異なる単位の系列との比較に使用することができます。
- **平均絶対誤差 (MAE)**. 平均絶対誤差。系列はそのモデルの予測水準からどれくらい変るかを測定します。MAE は、もとの系列単位で報告されます。
- **最大絶対パーセント誤差 (MaxAPE)**. 最大絶対パーセント誤差。最大予測誤差であり、パーセントとして表現することができます。この測定値は、予測に対する最悪のケースのシナリオをイメージするのに有効です。
- **最大絶対誤差 (MaxAE)**. 最大絶対誤差。最大予測誤差であり、従属系列と同じ単位で表現されます。MaxAPE と同様に、予測に対して最悪のケースのシナリオをイメージするのに有効です。最大絶対値誤差と最大絶対値パーセント誤差は、他の系列点で生じる場合があります。たとえば、大きな系列値の絶対誤差が小さな系列値の絶対誤差よりわずかに大きい場合などです。そのような場合、最大絶対値誤差はより大きな系列値で発生し、最大絶対値パーセント誤差はより小さな系列値で発生します。
- **標準化ベイズ情報基準 (BIC)**. 正規化されたベイズの情報量基準。モデルの複雑さを説明しようとするのは一般的にはモデルに適する方法です。それは誤差平均平方和に基づいたスコアであり、モデルおよび系列の長さのパラメーターの数に対するペナルティーが含まれます。ペ

ナルティは、より多くのパラメーターを持つモデルの利点を無くしますが、同じ系列のさまざまなモデル全体を統計的に簡単に比較することができます。

外れ値の種類

この項では、時系列モデリングで使用される外れ値の種類を定義します。

- **加法.** 単一観測値に影響する外れ値。たとえば、データコーディングエラーは加法的外れ値として識別される場合があります。
- **レベルシフト.** 定数によりすべての観測値をシフトさせる外れ値。特定の系列点から開始します。水準の変化は方針の変更によってもたらされます。
- **攪乱.** 特定の系列ポイントでノイズ項へ追加の役割をする外れ値。定常時系列では、革新の外れ値がいくつかの観測値に影響します。非定常時系列では、特別の系列ポイントから始まるすべての観測値に影響する場合があります。
- **一時的.** インパクトが 0 に指数的に減衰する外れ値。
- **季節加法.** 特定の観測値と 1 つ以上の季節期間によって、それから分離されたすべてのその後の観測値に影響する外れ値。そのような観察はすべて等しく影響されます。ある年から開始された販売が毎年 1 月に高い場合は、加法的季節成分が発生する場合があります。
- **局所トレンド.** 特定の系列点で局所トレンドを開始する外れ値。
- **加法パッチ.** 2 つ以上の連続する加法的外れ値のグループ。この外れ値型を選択すると、それらのパッチだけでなく個々の加法外れ値パッチ（変量のある値が密集するある特定の区間）の発見にもつながります。

ACF/PACF プロットについて

ここに示すプロットは、純粋な ARIMA 手続きまたは理論的 ARIMA 手続きのものです。手続きを識別するための一般的なガイドラインが記載されています。

- 非定常時系列には、すぐに 0 に減少するのではなく、より多くのラグの有意度を残す ACF があります。プロセスを特定する前に、定常時系列となるまでこの時系列を区分する必要があります。
- 自己回帰プロセスには、PACF の最初のいくつかのラグの垂線、および指数関数的に減少する ACF があります。垂線の数は、自己回帰の次数を示しています。
- 移動平均プロセスには、ACF の最初のいくつかのラグの垂線、および指数関数的に減少する PACF があります。垂線の数は、移動平均の次数を示しています。
- 混合 (ARMA) プロセスは、通常、ACF と PACF の両方で指数関数的な減少を示します。

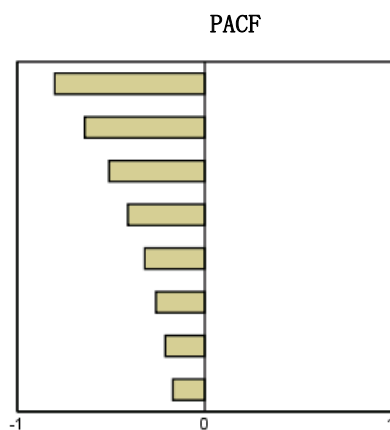
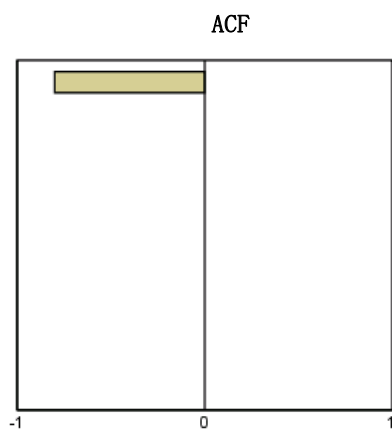
特定の段階では、ACF または PACF の記号について、または ACF または PACF の減少が指数関数的に 0 に達する速度について気にする必要はありません。これは、AR または MA 係数の記号および実際の値によって異なります。場合によっては、指数関数的に減少する ACF の値が正と負を交互に繰り返すことがあります。

実際の値から生成される ACF、および PACF のプロットは、ここに示すほど簡潔ではありません。与えられたプロットの中から必要なものを選択する必要があります。識別が誤っている場合は、必ず残差の ACF、および PACF を調べてください。以下の点に注意してください。

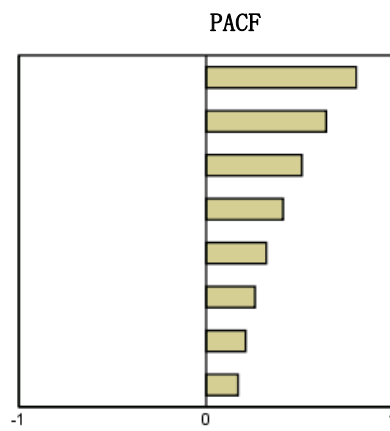
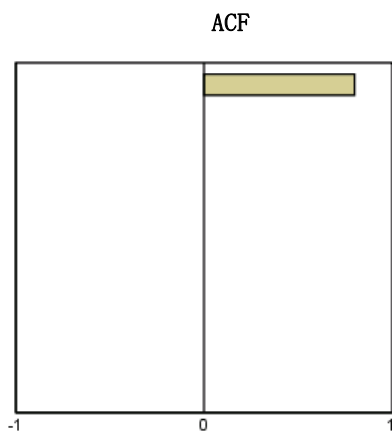
- 季節性のプロセスは、季節ラグでこのようなパターンを示す (季節期間の倍数)。
- 有意でない値を 0 として扱います。つまり、プロットの信頼区間内にある値は無視できます。しかし、値が統計上有意なパターンを維持している場合などは、無視する必要はありません。
- 偶発的な自己相関は、偶然以外に統計的な意味を持たない。統計的に有意な自己相関が (できれば高いラグで) 孤立している場合、および季節ラグが発生していない場合は無視できます。

ACF、および PACF のプロットの詳細は、ARIMA 分析のテキストを参照してください。

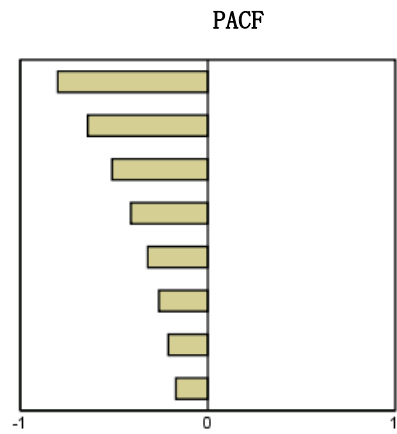
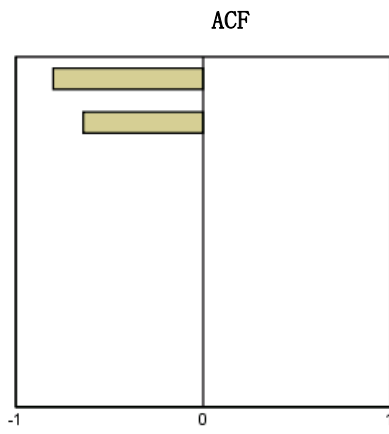
テーブル C-1
ARIMA(0,0,1), $q > 0$



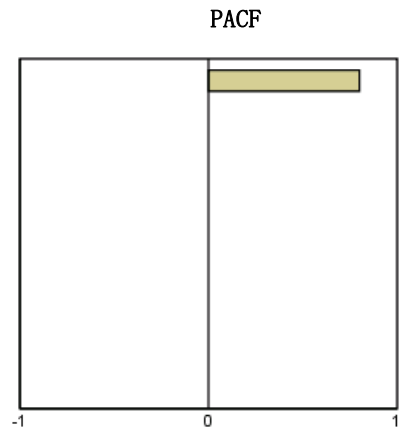
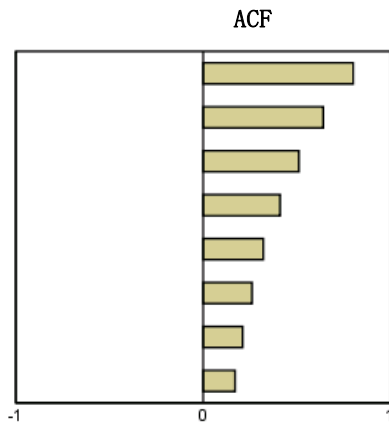
テーブル C-2
ARIMA(0,0,1), $q < 0$



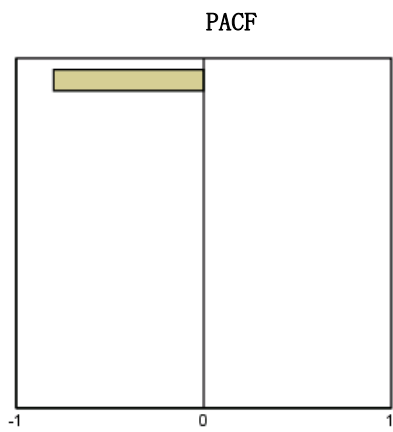
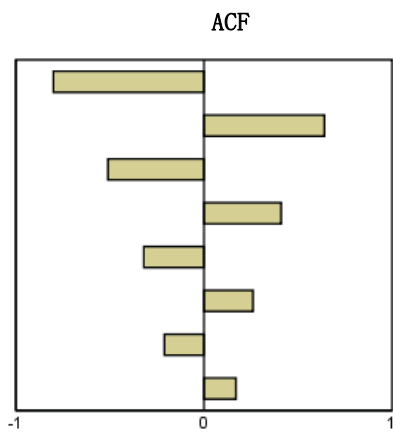
ARIMA(0, 0, 2), $\theta_1\theta_2 > 0$



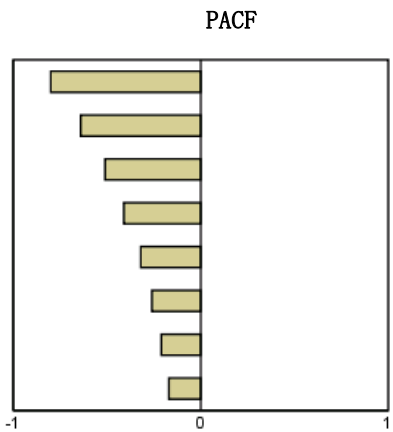
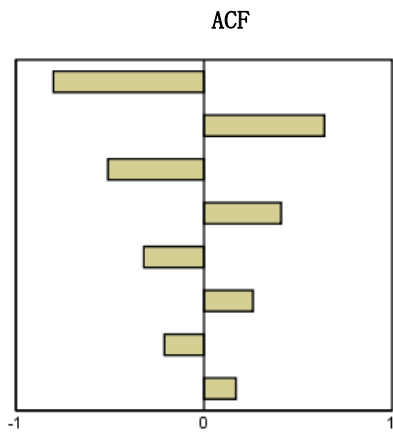
テーブル C-3
ARIMA(1,0,0), $\phi > 0$



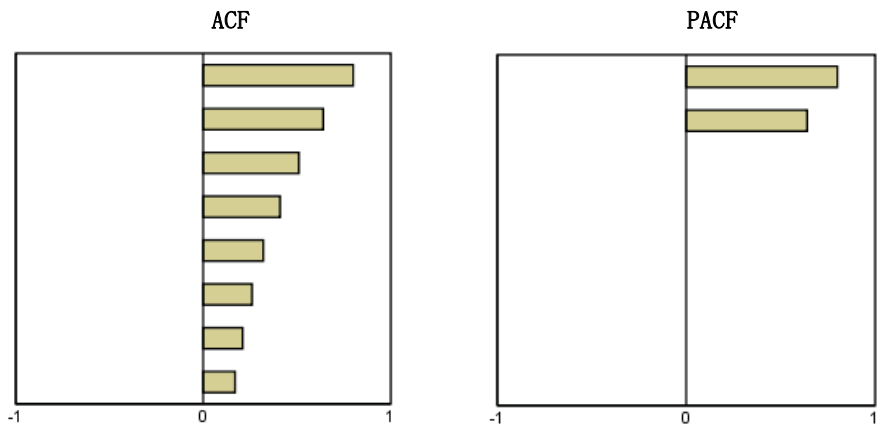
テーブル C-4
ARIMA(1,0,0), $\phi < 0$



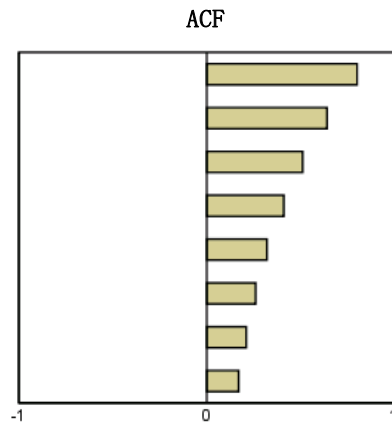
ARIMA(1,0,1), $\phi < 0, \theta > 0$



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



テーブル C-5
ARIMA(0,1,0) (和分系列)



サンプル ファイル

製品とともにインストールされるサンプル ファイルは、インストールディレクトリの Samples サブディレクトリにあります。[サンプル] サブディレクトリ内に次の各言語の別のフォルダがあります。英語、フランス語、ドイツ語、イタリア語、日本語、韓国語、ポーランド語、ロシア語、簡体字中国語、スペイン語、そして繁体中国語です。

すべてのサンプル ファイルが、すべての言語で使用できるわけではありません。サンプル ファイルがある言語で使用できない場合、その言語のフォルダには、サンプル ファイルの英語バージョンが含まれています。

説明

以下は、このドキュメントのさまざまな例で使用されているサンプル ファイルの簡単な説明です。

- **accidents.sav**。与えられた地域での自動車事故の危険因子を年齢および性別ごとに調べている保険会社に関する架空のデータ ファイルです。各ケースが、年齢カテゴリと性別のクロス分類に対応します。
- **adl.sav**。脳卒中患者に提案される治療の効果を特定するための取り組みに関する架空のデータ ファイルです。医師団は、女性の脳卒中患者たちを、2 つのグループのいずれかにランダムに割り当てました。一方のグループは標準的な理学療法を受け、もう一方のグループは感情面の治療も追加で受けました。治療の 3 か月後に、各患者が日常生活の一般的な行動をどの程度とることができるかを、順序変数として得点付けしました。
- **advert.sav**。広告費とその売上成果の関係を調べるための小売業者の取り組みに関する架空のデータ ファイルです。この小売業者は、そのために、過去の売上と、それに関する広告費のデータを収集しました。
- **aflatoxin.sav**。収穫物によって濃度が大きく異なる毒物であるアフラトキシンを、トウモロコシの収穫物に関して検定することに関する架空のデータ ファイルです。ある穀物加工業者は、8 つそれぞれの収穫物から 16 のサンプルを受け取って、10 億分の 1 単位でアフラトキシン レベルを測定しました。
- **anorectic.sav**。拒食行動または過食行動の標準的な症状の特定を目指して、調査員 (Van der Ham, Meulman, Van Strien, および Van Engeland, 1997) が、摂食障害を持つ大人 55 人の調査を行いました。各患者が 4 年間で 4 回診察を受けたので、観測値は合計で 220 になりました。観

測値ごとに、16 種類の症状に関して患者の得点が記録されました。患者 71 (2 回目)、患者 76 (2 回目)、患者 47 (3 回目) の症状の得点が見つからなかったため、残っている 217 回分の観測値が有効です。

- **bankloan.sav.** 債務不履行率を低減させるための銀行の取り組みに関する架空のデータ ファイルです。このファイルには、過去の顧客および見込み客 850 人に関する財務情報と人口統計情報が含まれています。最初の 700 ケースは、以前に貸付を行った顧客です。残りの 150 ケースは見込み顧客で、これらの顧客に関して銀行は信用リスクの良し悪しを分類する必要があります。
- **bankloan_binning.sav.** 過去の顧客 5,000 人に関する財務情報と人口統計情報を含む架空のデータ ファイルです。
- **behavior.sav.** 52 人の学生に 15 の状況と 15 の行動の組み合わせについて、0 = 「非常に適切」から 9 = 「非常に不適切」までの 10 段階でランク付けするよう依頼した研究があります (Price および Bouffard, 1974)。個人間の平均を取ったため、値は非類似度としてみなされます。
- **behavior_ini.sav.** このデータ ファイルには、behavior.sav の 2 次元の解の初期配置が含まれています。
- **brakes.sav.** 高性能自動車のディスク ブレーキを生産している工場での品質管理に関する架空のデータ ファイルです。このデータ ファイルには、8 台の機械で生産した 16 個のディスクの直径測定値が含まれています。ブレーキの目標の直径は 322 ミリメートルです。
- **breakfast.sav.** 21 人の Wharton School MBA の学生およびその配偶者に、15 種類の朝食を好みの順に (1 = 「最も好き」から 15 = 「最も嫌い」まで) ランク付けするよう依頼した研究があります (Green および Rao, 1972)。調査対象者の嗜好は、「すべて」から「スナックとドリンクのみ」まで、6 つの異なるシナリオに基づいて記録されました。
- **breakfast-overall.sav.** このデータ ファイルには、最初のシナリオ (「すべて」) のみの朝食の好みが含まれています。
- **broadband_1.sav.** 全国規模のブロードバンド サービスの地域ごとの契約者数を含む架空のデータ ファイルです。このデータ ファイルには、85 地域の月々の契約者数が 4 年間分含まれています。
- **broadband_2.sav.** このデータ ファイルは broadband_1.sav と同じですが、データが 3 か月分追加されています。
- **car_insurance_claims.sav.** 他の場所 (McCullagh および Nelder, 1989) で表示および分析される、自動車の損害請求に関するデータセットです。逆リンク関数を使用して従属変数の平均値を保険契約者の年齢、車種、製造年の線型結合と関連付けることにより、平均請求数はガンマ分布としてモデリングできます。申請された請求の数は、尺度重み付けとして使用できます。
- **car_sales.sav.** このデータ ファイルには、自動車のさまざまな車種やモデルの架空の売上推定値、定価、仕様が含まれています。定価と仕様はそれぞれ、edmunds.com と製造元のサイトから入手しました。

- **car_sales_upprepared.sav.** 変換したバージョンのフィールドを含まない car_sales.sav の修正したバージョンです。
- **carpet.sav.** 一般的な例 (Green および Wind, 1973) としては、新しいカーペット専用洗剤を市販することに関心のある企業が消費者の嗜好に関する 5 種類の因子 (パッケージのデザイン、ブランド名、価格、サービスシール、料金の払い戻し) の影響について調べたい場合があります。パッケージのデザインには、3 つの因子レベルがあります。それぞれ塗布用ブラシの位置が異なります。また、3 つのブランド名 (K2R、Glory、および Bissell)、3 つの価格水準があり、最後の 2 つの因子のそれぞれに対しては 2 つのレベル (「なし」または「あり」) があります。10 人の消費者が、これらの因子により定義された 22 個のプロファイルに順位を付けます。変数「嗜好」には、各プロファイルの平均順位の序列が含まれています。順位が低いほど、嗜好度は高くなります。この変数には、各プロファイルの嗜好測定値がすべて反映されます。
- **carpet_prefs.sav.** このデータ ファイルは carpet.sav と同じ例に基づいていますが、10 人の消費者それぞれから収集した実際のランキングが含まれています。消費者は、22 種類の製品プロファイルを、一番好きなものから一番嫌いなものまで順位付けすることを依頼されています。変数 PREF1 から PREF22 には、carpet_plan.sav で定義されている、関連するプロファイルの ID が含まれています。
- **catalog.sav.** このデータ ファイルには、あるカタログ会社が販売した 3 つの製品の、架空の月間売上高が含まれています。5 つの予測変数のデータも含まれています。
- **catalog_seasfac.sav.** このデータ ファイルは catalog.sav と同じですが、季節性の分解手続きとそれに付随する日付変数から計算した一連の季節因子が追加されています。
- **cellular.sav.** 解約率を削減するための携帯電話会社の取り組みに関する架空のデータ ファイルです。解約の傾向スコアは、0 ~ 100 の範囲でアカウントに適用されます。スコアリングが 50 以上のアカウントはプロバイダの変更を考えている場合があります。
- **ceramics.sav.** 新しい上質の合金に標準的な合金より高い耐熱性があるかどうかを特定するための、ある製造業者の取り組みに関する架空のデータ ファイルです。各ケースが 1 つの合金の別々のテストを表し、軸受けの耐熱温度が記録されます。
- **cereal.sav.** 880 人を対象に、朝食の好みについて、年齢、性別、婚姻状況、ライフスタイルが活動的かどうか (週 2 回以上運動するか) を含めて調査した、架空のデータ ファイルです。各ケースが別々の回答者を表します。
- **clothing_defects.sav.** ある衣料品工場での品質管理工程に関する架空のデータ ファイルです。工場で生産される各ロットから、調査員が衣料品のサンプルを取り出し、不良品の数を数えます。

- **coffee.sav.** このデータ ファイルは、6 つのアイスコーヒー ブランド (Kennedy, Riquier, および Sharp, 1996) について受けた印象に関連しています。回答者は、アイス コーヒーに対する 23 の各印象属性に対して、その属性が言い表していると思われるすべてのブランドを選択しました。機密保持のため、6 つのブランドを AA、BB、CC、DD、EE、および FF で表しています。
- **contacts.sav.** 企業のコンピュータ営業グループの担当者リストに関する架空のデータ ファイルです。各担当者は、所属する会社の部門および会社のランクによって分類されています。また、最新の販売金額、最後の販売以降の経過時間、担当者の会社の規模も記録されています。
- **creditpromo.sav.** 最近のクレジット カード プロモーションの有効性を評価するための、あるデパートの取り組みに関する架空のデータ ファイルです。このために、500 人のカード所有者がランダムに選択されました。そのうち半分には、今後 3 か月間の買い物に関して利率を下げることをプロモーションする広告を送付しました。残り半分には、通常どおりの定期的な広告を送付しました。
- **customer_dbase.sav.** 自社のデータ ウェアハウスにある情報を使用して、反応がありそうな顧客に対して特典を提供するための、ある会社の取り組みに関する架空のデータ ファイルです。顧客ベースのサブセットをランダムに選択して特典を提供し、顧客の反応が記録されています。
- **customer_information.sav.** 名前や住所など、顧客の連絡先情報を含む架空のデータ ファイルです。
- **customer_subset.sav.** customer_dbase.sav の 80 件のケースのサブセット。
- **debate.sav.** 政治討論の出席者に対して行った調査の、討論の前後それぞれの回答に関する架空のデータ ファイルです。各ケースが別々の回答者に対応します。
- **debate_aggregate.sav.** debate.sav 内の回答を集計する、架空のデータ ファイルです。各ケースが、討論前後の好みのクロス分類に対応しています。
- **demo.sav.** 月々の特典を送付することを目的とした、購入顧客のデータベースに関する架空のデータ ファイルです。顧客が特典に反応したかどうか、さまざまな人口統計情報と共に記録されています。
- **demo_cs_1.sav.** 調査情報のデータベースをコンパイルするための、ある会社の取り組みの最初のステップに関する架空のデータ ファイルです。各ケースが別々の都市に対応し、地域、地方、地区、および都市の ID が記録されています。
- **demo_cs_2.sav.** 調査情報のデータベースをコンパイルするための、ある会社の取り組みの第 2 のステップに関する架空のデータ ファイルです。各ケースが、最初のステップで選択した都市の別々の世帯単位に対応し、地域、地方、地区、都市、区画、および単位の ID が記録されます。計画の最初の 2 つの段階からの抽出情報も含まれています。

- **demo_cs.sav**。コンプレックス サンプル計画を使用して収集された調査情報を含む架空のデータ ファイルです。各ケースが別々の世帯単位に対応し、さまざまな人口統計情報および抽出情報が記録されています。
- **dmdata.sav**。これは、ダイレクト マーケティング企業の人口統計情報および購入情報を含む架空のデータです。dmdata2.sav には、テストメールを受け取った連絡先のサブセットの情報を含み、dmdata3.sav には、テストメールを受け取らなかった残りの連絡先に関する情報を含みます。
- **dietstudy.sav**。この架空のデータ ファイルには、“Stillman diet” (Rickman, Mitchell, Dingman, および Dalen, 1974) の研究結果が含まれています。各ケースが別々の被験者に対応し、被験者のダイエット前後の体重 (ポンド単位) と、トリグルセリド レベル (mg/100 ml 単位) が記録されています。
- **dvdplayer.sav**。新しい DVD プレーヤーの開発に関する架空のデータ ファイルです。プロトタイプを使用して、マーケティング チームはフォーカス グループ データを収集しました。各ケースが別々の調査対象ユーザーに対応し、ユーザーの人口統計情報と、プロトタイプに関する質問への回答が記録されています。
- **german_credit.sav**。このデータ ファイルは、カリフォルニア大学アーバイン校の Repository of Machine Learning Databases (Blake および Merz, 1998) にある “German credit” データセットから取ったものです。
- **grocery_1month.sav**。この架空のデータ ファイルは、grocery_coupons.sav データ ファイルの週ごとの購入を「ロールアップ」して、各ケースが別々の顧客に対応するようにしたものです。その結果、週ごとに変わっていた変数の一部が表示されなくなり、買物の総額が、調査を行った 4 週間の買物額の合計になっています。
- **grocery_coupons.sav**。顧客の購買習慣に関心を持っている食料雑貨店チェーンが収集した調査データを含む架空のデータ ファイルです。各顧客を 4 週間に渡って追跡し、各ケースが別々の顧客の週に対応しています。その週に食料品に費やした金額も含め、顧客がいつどこで買物をするかに関する情報が記録されています。
- **guttman.sav**。Bell (Bell, 1961) は、予想される社会グループを示す表を作成しました。Guttman (Guttman, 1968) は、この表の一部を使用しました。この表では、社会相互作用、グループへの帰属感、メンバとの物理的な近接性、関係の形式化などを表す 5 個の変数が、理論上の 7 つの社会グループと交差しています。このグループには、観衆 (例、フットボールの試合の観戦者)、視聴者 (例、映画館または授業の参加者)、公衆 (例、新聞やテレビの視聴者)、暴徒 (観衆に似ているが、より強い相互作用がある)、第一次集団 (親密な関係)、第二次集団 (自発的な集団)、および近代コミュニティ (物理的により密接した近接性と特化されたサービスの必要性によるゆるい同盟関係) があります。

- **health_funding.sav**。医療用資金（人口 100 人あたりの金額）、罹患率（人口 10,000 人あたりの人数）、医療サービス機関への訪問率（人口 10,000 人あたりの人数）のデータを含む、架空のデータ ファイルです。各ケースが別々の都市を表します。
- **hivassay.sav**。HIV 感染を発見する迅速な分析方法を開発するための、ある製薬研究所の取り組みに関する架空のデータ ファイルです。分析の結果は、8 段階の濃さの赤で表現され、色が濃いほど感染の可能性が高くなります。研究所では 2,000 件の血液サンプルに関して試験を行い、その半数が HIV に感染しており、半分は感染していませんでした。
- **hourlywagedata.sav**。管理職から現場担当まで、またさまざまな経験レベルの看護師の時給に関する架空のデータ ファイルです。
- **insurance_claims.sav**。不正請求の恐れがある、疑いを区別するためにモデルを作成する必要がある保険会社の仮説データ ファイルです。各ケースがそれぞれの請求を表します。
- **insure.sav**。10 年満期の生命保険契約に対し、顧客が請求を行うかどうかを示す危険因子を調査している保険会社に関する架空のデータ ファイルです。データ ファイルの各ケースは、年齢と性別が一致する、請求を行った契約と行わなかった契約のペアを表します。
- **judges.sav**。訓練を受けた審判（および 1 人のファン）が 300 件の体操の演技に対して付けた得点に関する架空のデータ ファイルです。各行が別々の演技を表し、審判たちは同じ演技を見ました。
- **kinship_dat.sav**。Rosenberg と Kim (Rosenberg および Kim, 1975) は、15 種類の親族関係用語（祖父、祖母、父、母、叔父、叔母、兄弟、姉妹、いとこ、息子、娘、甥、姪、孫息子、孫娘）の分析を行いました。Rosenberg と Kim は、大学生の 4 つのグループ（女性 2 組、男性 2 組）に、類似性に基づいて上記の用語を並べ替えるよう依頼しました。2 つのグループ（女性 1 組、男性 1 組）には、1 回目と違う条件に基づいて、2 回目の並べ替えをするように頼みました。このようにして、合計で 6 つの「ソース」が取得できました。各ソースは、15×15 の近接行列に対応します。この近接行列のセルの数は、ソースの人数から、ソース内でオブジェクトを分割した回数を引いたものです。
- **kinship_ini.sav**。このデータ ファイルには、kinship_dat.sav の 3 次元の解の初期配置が含まれています。
- **kinship_var.sav**。このデータ ファイルには、kinship_dat.sav の解の次元の解釈に使用できる独立変数である性別、世代、および(ation), and 親等が含まれています。特に、解の空間をこれらの変数の線型結合に制限するために使用できます。
- **marketvalues.sav**。1999 ~ 2000 年の間の、イリノイ州アルゴンキンの新興住宅地での住宅売上に関するデータ ファイルです。これらの売上は、公開レコードの問題となります。

- **nhis2000_subset.sav**。National Health Interview Survey (NHIS) は、米国民を対象とした人口ベースの大規模な調査です。全国の代表的な世帯サンプルについて対面式で調査が行われます。各世帯のメンバーに関して、人口統計情報、健康に関する行動および状態の観測値が得られます。このデータ ファイルには、2000 年の調査から得られた情報のサブセットが含まれています。National Center for Health Statistics。National Health Interview Survey, 2000。一般使用データおよびドキュメント。ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/。2003 年にアクセス。
- **ozone.sav**。データには、残りの変数からオゾン濃度を予測するための、6 個の気象変数に対する 330 個の観測値が含まれています。それまでの研究者 (Breiman および Friedman, 1985)、(Hastie および Tibshirani, 1990) が、他の研究者と共に、これらの変数間に非線型性を確認しています。この場合、標準的な回帰アプローチは使用できません。
- **pain_medication.sav**。この架空のデータ ファイルには、慢性関節炎を治療する抗炎症薬の臨床試験の結果が含まれています。特に興味深いことは、薬の効果が出るまでの時間と、既存の薬剤との比較です。
- **patient_los.sav**。この架空のデータ ファイルには、心筋梗塞 (MI、または「心臓発作」) の疑いで入院した患者の治療記録が含まれています。各ケースが別々の患者に対応し、入院に関連する多くの変数が記録されています。
- **patlos_sample.sav**。この架空のデータ ファイルには、心筋梗塞 (MI、または「心臓発作」) の治療中に血栓溶解剤を投薬された患者のサンプルの治療記録が含まれています。各ケースが別々の患者に対応し、入院に関連する多くの変数が記録されています。
- **poll_cs.sav**。市民の法案支持率を議会開会前に特定するための、世論調査員の取り組みに関する架空のデータ ファイルです。各ケースは登録有権者に対応しています。ケースごとに、有権者が居住している郡、町、区域が記録されています。
- **poll_cs_sample.sav**。この架空のデータ ファイルには、poll_cs.sav の有権者のサンプルが含まれています。サンプルは、poll_csplan 計画ファイルで指定されている計画に従って抽出され、このデータ ファイルには包含確率およびサンプル重み付けが記録されています。ただし、抽出計画では確率比例 (PPS) 法を使用するため、結合選択確率を含むファイル (poll_jointprob.sav) もあります。サンプル抽出後、有権者の人口統計および法案に関する意見に対応する追加の変数が収集され、データ ファイルに追加されました。
- **property_assess.sav**。限られたリソースで資産価値評価を最新に保つための、郡の評価担当者の取り組みに関する架空のデータ ファイルです。各ケースは、前年に郡内で売却された資産に対応します。データ ファイル内の各ケースでは、資産が存在する町、最後に訪問した評価

担当者、その評価からの経過時間、当時行われた評価、および資産の売却価値が記録されています。

- **property_assess_cs.sav**。限られたリソースで資産価値評価を最新に保つための、州の評価担当者の取り組みに関する架空のデータ ファイルです。各ケースは州内の資産に対応します。データ ファイル内の各ケースでは、資産が存在する郡、町、および区域、最後の評価からの経過時間、および当時行われた評価が記録されています。
- **property_assess_cs_sample.sav**。この架空のデータ ファイルには、property_assess_cs.sav の資産のサンプルが含まれています。サンプルは、property_assess_csplan 計画ファイルで指定されている計画に従って抽出され、このデータ ファイルには包含確率およびサンプル重み付けが記録されています。サンプル抽出後、現在の価値変数が収集され、データ ファイルに追加されました。
- **recidivism.sav**。管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ ファイルです。各ケースは元犯罪者に対応し、人口統計情報、最初の犯罪の詳細、初犯から 2 年以内の場合には 2 回目の逮捕までの期間が記録されています。
- **recidivism_cs_sample.sav**。管轄地域での累犯率を把握するための、政府の法執行機関の取り組みに関する架空のデータ ファイルです。各ケースは 2003 年の 7 月に最初の逮捕から釈放された元犯罪者に対応し、人口統計情報、最初の犯罪の詳細、2006 年 7 月までの 2 回目の逮捕のデータが記録されています。犯罪者は recidivism_cs.plan で指定された抽出計画に従って抽出された部門から選択されます。調査では確率比例 (PPS) 法を採用したため、結合選択確率を保持したファイル (recidivism_cs_jointprob.sav) も用意されています。
- **rfm_transactions.sav**。購入日、購入品目、各取引のマネタリー量など、購買取引データを含む架空のデータ ファイルです。
- **salesperformance.sav**。2 つの新しい販売トレーニング コースの評価に関する架空のデータ ファイルです。60 人の従業員が 3 つのグループに分けられ、全員が標準のトレーニングを受けます。さらに、グループ 2 は技術トレーニングを、グループ 3 は実践的なチュートリアルを受けます。トレーニング コースの最後に各従業員がテストを受け、得点が記録されました。データ ファイルの各ケースは別々の訓練生を表し、割り当てられたグループと、テストの得点が記録されています。
- **satisf.sav**。ある小売業者が 4 箇所の店舗で行った満足度調査に関する架空のデータ ファイルです。合計で 582 人の顧客を調査し、各ケースは 1 人の顧客からの回答を表します。
- **screws.sav**。このデータ ファイルには、ねじ、ボルト、ナット、鋸 (びょう) (Hartigan, 1975) の特性に関する情報が含まれています。
- **shampoo_ph.sav**。あるヘアケア製品工場での品質管理に関する架空のデータ ファイルです。定期的に、6 つの異なる製品が測定され、pH が記録されます。目標範囲は 4.5 ~ 5.5 です。

- **ships.sav.** 他の場所 (McCullagh など, 1989) で表示および分析される、波による貨物船への損害に関するデータセットです。件数は、船舶の種類、建造期間、およびサービス期間によって、ポワゾン率で発生するものとしてモデリングできます。因子のクロス分類によって形成されたテーブルの各セルのサービス月数の集計によって、危険にさらされる確率の値が得られます。
- **site.sav.** 業務拡大に向けて新たな用地を選択するための、ある会社の取り組みに関する架空のデータ ファイルです。2 人のコンサルタントを雇って、用地を別々に評価させました。広範囲のレポートに加えて、各用地を「良い」、「普通」、「悪い」のいずれかで集計しました。
- **smokers.sav.** このデータ ファイルは、1998 年の National Household Survey of Drug Abuse から抜粋したものであり、アメリカの世帯の確率サンプルです。(<http://dx.doi.org/10.3886/ICPSR02934>) したがって、このデータ ファイルを分析する場合は、まず人口の傾向を反映させてデータを重み付けする必要があります。
- **stocks.sav** このデータ ファイルには、1 年あたりの在庫価格、量が含まれています。
- **stroke_clean.sav.** この架空のデータ ファイルには、[データの準備] オプションの手続きを使用して整理した後の、医療データベースの状態が含まれています。
- **stroke_invalid.sav.** この架空のデータ ファイルには、医療データベースの初期状態が含まれており、データ入力にいくつかエラーがあります。
- **stroke_survival.** この架空のデータ ファイルは、虚血性脳卒中で数回の困難に直面した後リハビリ プログラムを終えた患者の生存時間に関するものです。脳卒中後、心筋梗塞の発生、虚血性脳卒中、または出血性脳卒中が注意され、イベントの時間が記録されます。脳卒中後に実施されたリハビリ プログラムの最後まで生存した患者のみが含まれるため、サンプルは左側が切り捨てられます。
- **stroke_valid.sav.** この架空のデータ ファイルには、[データの検証] 手続きを使用して確認した後の、医療データベースの状態が含まれています。異常である可能性のあるケースが含まれています。
- **survey_sample.sav.** このデータ ファイルには、人口統計データおよびさまざまな態度指標などの調査データが含まれています。これは「1998 NORC General Social Survey」の変数のサブセットに基づいていますが、いくつかのデータ値が変更され、追加の架空変数がデモの目的で追加されています。
- **telco.sav.** 顧客ベースにおける解約率を削減するための電気通信会社の取り組みに関する架空のデータ ファイルです。各ケースが別々の顧客に対応し、人口統計やサービス利用状況などのさまざまな情報が記録されています。

- **telco_extra.sav.** このデータ ファイルは telco.sav データ ファイルに似ていますが、「期間」および対数変換された顧客支出の属性が削除され、標準化された対数変換顧客支出の変数に置き換えられています。
- **telco_missing.sav.** このデータ ファイルは telco.sav データ ファイルのサブセットですが、一部の人口統計データ値が欠損値に置き換えられています。
- **testmarket.sav.** この架空のデータ ファイルは、新しいメニューを追加しようというファースト フード チェーンの計画に関連しています。新製品をプロモーションするためのキャンペーンには 3 つの候補があるため、新メニューはいくつかのランダムに選択した市場にある場所で紹介されます。場所ごとに別々のプロモーションを使用し、最初の 4 週間の新メニューの週間売上高が記録されます。各ケースが場所と週に対応します。
- **testmarket_1month.sav.** この架空のデータ ファイルは、testmarket.sav データ ファイルの週ごとの売上を「ロールアップ」して、各ケースが別々の場所に対応するようにしたものです。その結果、週ごとに変わっていた変数の一部が表示されなくなり、売上高が、調査を行った 4 週間の売上高の合計になっています。
- **tree_car.sav.** これは、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree_credit.sav.** これは、人口統計および銀行ローン履歴のデータを含む架空のデータ ファイルです。
- **tree_missing_data.sav.** これは、人口統計および銀行ローン履歴のデータと、多数の欠損値を含む架空のデータ ファイルです。
- **tree_score_car.sav.** これは、人口統計および自動車購入価格のデータを含む架空のデータ ファイルです。
- **tree_textdata.sav.** 尺度および値ラベルを割り当てる前の、変数のデフォルトの状態を示すことを主な目的とする、変数を 2 つだけ含む単純なデータ ファイルです。
- **tv-survey.sav.** テレビ スタジオで実施された、ヒットした番組の放送期間を延長するかどうかを検討する調査に関する架空のデータ ファイルです。906 人の回答者に、さまざまな条件下でこの番組を視聴するかどうかを質問しました。各行は別々の回答者を表し、各列は別々の条件を表します。
- **ulcer_recurrence.sav.** このファイルには、潰瘍の再発を防ぐための 2 つの治療の有効性を比較するように計画された調査の情報の一部が含まれています。これは区間調査の良い例であり、他の場所 (Collett, 2003) で表示および分析されています。
- **ulcer_recurrence_recoded.sav.** このファイルでは、ulcer_recurrence.sav の情報が、単に調査終了時のイベント確率ではなく調査の区間ごとのイベント確率をモデリングできるように再編成されています。これは他の場所 (Collett など, 2003) で表示および分析されています。

- **verd1985.sav.** このデータ ファイルは調査 (Verdegaal, 1985) に関連しています。8 つの変数に対する 15 人の被験者の回答を記録しました。対象となる変数が 3 つのグループに分類されます。グループ 1 には「年齢」と「婚姻」、グループ 2 には「ペット」と「新聞」、グループ 3 には「音楽」と「居住地域」がそれぞれ含まれます。「ペット」は多重名義として尺度化され、「年齢」は順序として尺度化されます。また、その他のすべての変数は単一名義として尺度化されます。
- **virus.sav.** 自社のネットワーク上のウィルスの影響を特定するための、インターネット サービス プロバイダ (ISP) の取り組みに関する架空のデータ ファイルです。この ISP は、ネットワーク上の感染した E メール トラフィックの (およその) パーセンテージを、発見の瞬間から脅威が阻止されるまで追跡しました。
- **wheeze_steubenville.sav.** これは、子供 (Ware, Dockery, Spiro III, Speizer, および Ferris Jr., 1984) に対する大気汚染の健康上の影響の長期調査から得られたサブセットです。このデータには、オハイオ州 スビューベンビルの 7 歳、8 歳、9 歳、10 歳の子供を対象に行った、喘鳴の状態の反復 2 値測定と、調査の初年に母親が喫煙していたかどうかの固定記録が含まれています。
- **workprog.sav.** 体の不自由な人をより良い仕事に就かせようとする政府の事業プログラムに関する架空のデータ ファイルです。プログラムの参加者候補のサンプルが追跡されました。その中には、ランダムに選ばれてプログラムに登録された人と、そうでない人がいました。各ケースが別々のプログラム参加者を表します。
- **worldsales.sav** このデータ ファイルには、大陸および製品ごとの販売収益が含まれています。

注意事項

この情報は、世界各国で提供される製品およびサービス向けに作成されています。

IBMはこのドキュメントで説明する製品、サービス、機能は他の国では提供していない場合があります。現在お住まいの地域で利用可能な製品、サービス、および、情報については、お近くの IBM の担当者にお問い合わせください。IBM 製品、プログラム、またはサービスに対する参照は、IBM 製品、プログラム、またはサービスのみが使用することができることを説明したり意味するものではありません。IBM の知的所有権を侵害しない機能的に同等の製品、プログラム、またはサービスを代わりに使用することができます。ただし、IBM 以外の製品、プログラム、またはサービスの動作を評価および確認するのはユーザーの責任によるものです。

IBMは、本ドキュメントに記載されている内容に関し、特許または特許出願中の可能性があります。本ドキュメントの提供によって、これらの特許に関するいかなる権利も使用者に付与するものではありません。ライセンスのお問い合わせは、書面にて、下記住所に送ることができます。

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

2 バイト文字セット (DBCS) 情報についてのライセンスに関するお問い合わせは、お住まいの国の IBM Intellectual Property Department に連絡するか、書面にて下記宛先にお送りください。

神奈川県大和市下鶴間1623番14号 日本アイ・ビー・エム株式会社 法務・知的財産 知的財産権ライセンス渉外

以下の条項は、イギリスまたはこのような条項が法律に反する他の国では適用されません。 International Business Machines は、明示的または黙示的に関わらず、第三者の権利の侵害しない、商品性または特定の目的に対する適合性の暗黙の保証を含むがこれに限定されない、いかなる保証なく、本出版物を「そのまま」提供します一部の州では、特定の取引の明示的または暗示的な保証の免責を許可していないため、この文が適用されない場合があります。

この情報には、技術的に不適切な記述や誤植を含む場合があります。情報については変更が定期的に行われます。これらの変更は本書の新版に追加されます。IBM は、本書に記載されている製品およびプログラムについて、事前の告知なくいつでも改善および変更を行う場合があります。

IBM 以外の Web サイトに対するこの情報内のすべての参照は、便宜上提供されているものであり、決してそれらの Web サイトを推奨するものではありません。これらの Web サイトの資料はこの IBM 製品の資料に含まれるものではなく、これらの Web サイトの使用はお客様の責任によるものとします。

IBM はお客様に対する一切の義務を負うことなく、自ら適切と考える方法で、情報を使用または配布することができるものとします。

本プログラムのライセンス取得者が (i) 別途作成されたプログラムと他のプログラム（本プログラムを含む）との間の情報交換および (ii) 交換された情報の相互利用を目的とした本プログラムに関する情報の所有を希望する場合、下記住所にお問い合わせください。

IBM Software Group, Attention:Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

上記のような情報は、該当する条項および条件に従い、有料で利用できるものとします。

本ドキュメントに記載されている許可されたプログラムおよびそのプログラムに使用できるすべてのライセンス認証された資料は、IBM Customer Agreement、IBM International Program License Agreement、および当社とかわした同等の契約の条件に基づき、IBM によって提供されます。

IBM 以外の製品に関する情報は、それらの製品の供給業者、公開済みの発表、または公開で使用できるソースから取得しています。IBM は、それらの製品のテストは行っておらず、IBM 以外の製品に関連する性能、互換性、またはその他の要求については確認できません。IBM 以外の製品の性能に関する質問は、それらの製品の供給業者に通知する必要があります。

この情報には、日常の業務処理で用いられるデータや報告書の例が含まれています。できる限り詳細に説明するため、例には、個人、企業、ブランド、製品などの名前が使用されています。これらの名称はすべて架空のものであり、実際の企業で使用される名称および住所とは一切関係ありません。

この情報をソフトコピーでご覧になっている場合は、写真やカラーのイラストが表示されない場合があります。

商標

IBM、IBM ロゴ、および [ibm.com](http://www.ibm.com)、SPSS は、世界の多くの国で登録された IBM Corporation の商標です。IBM の商標の現在のリストは、<http://www.ibm.com/legal/copytrade.shtml> を参照してください。

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel、Intel のロゴ、Intel Inside、Intel Inside のロゴ、Intel Centrino、Intel Centrino のロゴ、Celeron、Intel Xeon、Intel SpeedStep、Itanium、および Pentium は、米国およびその他の国の Intel Corporation または関連会社の商標または登録商標です。

Java およびすべての Java ベースの商標およびロゴは、米国およびその他の国の Sun Microsystems, Inc. の商標です。

Linux は、米国およびその他の国における Linus Torvalds の登録商標です。

Microsoft、Windows、Windows NT、および Windows のロゴは、米国およびその他の国における Microsoft 社の商標です。

UNIX は、米国およびその他の国における The Open Group の登録商標です。

この製品は、WinWrap Basic (Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>) を使用します。

その他の製品名およびサービス名等は、IBM または他の会社の商標です。

Adobe 製品のスクリーンショットは Adobe Systems Incorporated の許可を得て転載しています。

Microsoft 製品のスクリーンショットは Microsoft 社の許可を得て転載しています。



参考文献

- Bell, E. H. 1961. Social foundations of human behavior: Introduction to the study of sociology. New York: Harper & Row.
- Blake, C. L., および C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Box, G. E. P., G. M. Jenkins, および G. C. Reinsel. 1994. Time series analysis: Forecasting and control, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.
- Breiman, L., および J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association, 80, 580-598.
- Collett, D. 2003. Modelling survival data in medical research, 2 ed. Boca Raton: Chapman & Hall/CRC.
- Gardner, E. S. 1985. Exponential smoothing: The state of the art. Journal of Forecasting, 4, 1-28.
- Green, P. E., および V. Rao. 1972. Applied multidimensional scaling. Hinsdale, Ill.: Dryden Press.
- Green, P. E., および Y. Wind. 1973. Multiattribute decisions in marketing: A measurement approach. Hinsdale, Ill.: Dryden Press.
- Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. Psychometrika, 33, 469-506.
- Hartigan, J. A. 1975. Clustering algorithms. New York: John Wiley and Sons.
- Hastie, T., および R. Tibshirani. 1990. Generalized additive models. London: Chapman and Hall.
- Kennedy, R., C. Riquier, および B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. Journal of Targeting, Measurement, and Analysis for Marketing, 5, 56-70.
- McCullagh, P., および J. A. Nelder. 1989. Generalized Linear Models, 2nd ed. London: Chapman & Hall.
- Pena, D., G. C. Tiao, および R. S. Tsay, 編集者. 2001. A course in time series analysis. New York: John Wiley and Sons.
- Price, R. H., および D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. Journal of Personality and Social Psychology, 30, 579-586.

参考文献

Rickman, R., N. Mitchell, J. Dingman, および J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, 54-58.

Rosenberg, S., および M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

Van der Ham, T., J. J. Meulman, D. C. Van Strien, および H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

Verdegaal, R. 1985. Meer sets analyse voor kwalitatieve gegevens (in Dutch). Leiden: Department of Data Theory, University of Leiden.

Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, およ
び B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and
respiratory health of children living in six cities. *American Review
of Respiratory Diseases*, 129, 366-374.

索引

- 偏自己相関関数
 - 純粹 ARIMA 手続きのプロット, 108
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- 平均絶対誤差
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- 最大絶対誤差
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- 自己相関関数
 - 純粹 ARIMA 手続きのプロット, 108
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- 自然対数変換
 - 時系列モデラー, 11, 14, 16
- 妥当性期間, 2
- 平方根変換
 - 時系列モデラー, 11, 14, 16
- 予測期間
 - 時系列モデラー, 8, 26, 53, 55
 - 時系列モデルの適用, 31, 64, 87
- 伝達関数, 16
 - 分子次数, 16
 - 分母次数, 16
 - 季節次数, 16
 - 差分次数, 16
 - 遅延, 16
- 信頼区間
 - 時系列モデラー, 22, 26
 - 時系列モデルの適用, 35, 40
- 対数変換
 - 時系列モデラー, 11, 14, 16
- 履歴期間, 2
- 推定期間, 2
 - 時系列モデラー, 8, 53
 - 時系列モデルの適用, 31
- 周期性
 - 時系列モデラー, 9, 11, 14, 16
- 変数名
 - 時系列モデラー, 25
 - 時系列モデルの適用, 38
- 欠損値
 - 時系列モデラー, 26
 - 時系列モデルの適用, 40
- 適合度
 - 時系列モデラー, 20, 22, 57
 - 時系列モデルの適用, 33, 35, 66
- 予測
 - 時系列モデラー, 20, 22, 57
 - 時系列モデルの適用, 33, 35, 88
- 保存
 - 再度推定されたモデルを XML 形式で, 38
 - 新しい変数名, 25, 38
 - モデルの予測, 25, 38
 - モデル指定を XML 形式で, 25
- 商標, 125
- 残差
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- 最大絶対誤差 (MaxAE)
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- 平均 2 乗誤差平方根
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- 正規化 BIC (ベイズ情報基準)
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- 定常 R^2
 - 時系列モデラー, 20, 22, 78
 - 時系列モデルの適用, 33, 35
- 平均絶対誤差 (MAE)
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- 平均 2 乗誤差平方根 (RMSE)
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- ACF
 - 純粹 ARIMA 手続きのプロット, 108
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- ARIMA モデル, 7
 - 伝達関数, 16
 - outliers, 18
- ARIMA モデルのパラメータ表
 - 時系列モデラー, 79
- ARIMA モデル
 - 移動平均の次数, 14
 - 季節次数, 14
 - 差分の次数, 14
 - 自己回帰の次数, 14
 - 定数, 14
- Box-Ljung 統計量
 - 時系列モデラー, 20, 79
 - 時系列モデルの適用, 33
- Brown の指数平滑法モデル, 11
- events, 10
 - 時系列モデラー, 9
- Holt の指数平滑法モデル, 11
- outliers
 - ARIMA モデル, 18

索引

- エキスパート モデラー, 10, 73
- PACF
 - 純粹 ARIMA 手続きのプロット, 108
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- R^2 , 105
 - 時系列モデラー, 20, 22
 - 時系列モデルの適用, 33, 35
- Winters の指数平滑法モデル
 - 相乗的, 11
 - 加法, 11
- XML(X)
 - 再度推定されたモデルを XML 形式で保存する, 38
 - 時系列モデルを XML 形式で保存する, 25, 56, 74

- 移動平均
 - ARIMA モデル, 14
- エキスパート モデラー, 7, 50
 - outliers, 10, 73
 - モデル スペースの制限, 9, 54
- 攪乱的な外れ値, 107
- 過渡的な外れ値, 107
- 加法の外れ値, 107

- 局所トレンドの外れ値, 107
- 季節差分変換
 - ARIMA モデル, 14
- 季節次数
 - ARIMA モデル, 14
- 季節性相加的な外れ値, 107
- [季節性の分解], 42-44
 - 移動平均の計算, 42
 - 仮定, 42
 - 新変数の保存, 43
 - 変数の作成, 43
 - モデル, 42
- 最大絶対誤差, 105
- 最大絶対誤差 (MaxAE), 105
- 最大絶対パーセント誤差, 105
- 最大絶対パーセント誤差 (MaxAPE), 105
- 差分変換
 - ARIMA モデル, 14
- サンプル ファイル
 - 位置, 113

- 自己相関誤差の回帰
 - ARIMA モデル, 14
- スペクトルプロット, 45, 48
 - 2 変量相関のスペクトル分析, 47
 - 仮定, 45
 - 関連手続き, 104
 - スペクトル密度, 102
 - スペクトル ウィンドウ, 45
 - ペリオドグラム, 102
 - 変換の中央位置合わせ, 47
- 正規化 BIC (ベイズ情報基準), 105

- 相加的パッチの外れ値, 107
- 調和分析, 45

- 定常 R^2 , 105
- 適合度
 - 定義, 105
- 当てはめ値
 - 時系列モデラー, 22, 76
 - 時系列モデルの適用, 35
- 履歴データ
 - 時系列モデラー, 22
 - 時系列モデルの適用, 35

- 局所トレンドの外れ値
 - 時系列モデラー, 10, 18

- 季節性相加的な外れ値
 - 時系列モデラー, 10, 18
- 攪乱的な外れ値
 - 時系列モデラー, 10, 18
- 過渡的な外れ値
 - 時系列モデラー, 10, 18

- 法律に関する注意事項, 124

- 季節性の分解
 - 関連手続き, 99
 - 新しい変数, 97
 - 周期的な日付成分, 91
- 予測値の表
 - 時系列モデラー, 61
 - 時系列モデルの適用, 88
- 加法の外れ値
 - 時系列モデラー, 10, 18

- 外れ値
 - 定義, 107

- 相加的パッチの外れ値
 - 時系列モデラー, 10, 18
- 平均絶対パーセント誤差
 - 時系列モデラー, 20, 22, 58
 - 時系列モデルの適用, 33, 35, 67
- 最大絶対パーセント誤差
 - 時系列モデラー, 20, 22, 58
 - 時系列モデルの適用, 33, 35, 67
- 平均絶対パーセント誤差 (MAPE)
 - 時系列モデラー, 20, 22, 58
 - 時系列モデルの適用, 33, 35, 67
- 最大絶対パーセント誤差 (MaxAPE)
 - 時系列モデラー, 20, 22, 58
 - 時系列モデルの適用, 33, 35, 67

- 平均 2 乗誤差平方根, 105
- 平均2乗誤差平方根 (RMSE), 105
- 平均絶対誤差, 105
- 平均絶対誤差(MAE), 105
- 平均絶対パーセント誤差, 105
- 平均絶対パーセント誤差 (MAPE), 105

- ホールドアウト ケース, 2

- 時系列モデラー, 4
 - 残差偏自己相関関数, 20, 22
 - 残差自己相関関数, 20, 22
 - 適合度統計量, 20, 22, 57, 78
 - 指数平滑法, 7, 11
 - 予測期間, 8, 26, 53, 55
 - 伝達関数, 16
 - 信頼区間, 22, 26
 - 推定期間, 8, 53
 - 系列変換, 11, 14, 16
 - 周期性, 9, 11, 14, 16
 - 欠損値, 26
 - 予測, 20, 22, 57
 - ARIMA 分析, 7, 14
 - ARIMA モデルのパラメータ表, 79
 - Box-Ljung 統計量, 20
 - events, 9
 - outliers, 10, 18, 73
 - エキスパート モデラー, 7, 50, 69
 - 新しい変数名, 25, 60
 - すべてのモデルにわたる統計量, 20, 22, 57, 59
 - 当てはめ値, 22, 76
 - 予測値の表, 61
 - モデル名, 26
 - 最適モデルと最不適合モデル, 24
 - モデル統計量の表, 78
 - モデルの説明表, 78
 - モデルのパラメータ, 20, 75
 - モデル指定を XML 形式で保存する, 25, 56, 74
 - 予測を保存する, 25, 56
- モデル
 - 指数平滑法, 7, 11
 - ARIMA 分析, 7
 - ARIMA (X11 ARIMA), 14
 - エキスパート モデラー, 7
 - 単純季節指数平滑法モデル, 11
 - 単純指数平滑法モデル, 11
 - 減衰指数平滑法モデル, 11
 - 指数平滑法モデル, 7, 11
 - モデル名
 - 時系列モデラー, 26
 - モデルの適合度表
 - 時系列モデルの適用, 66
 - モデル統計量の表
 - 時系列モデラー, 78
 - モデルの説明表
 - 時系列モデラー, 78
 - 時系列モデルの適用, 29, 63, 80
 - 残差偏自己相関関数, 33, 35
 - 残差自己相関関数, 33, 35
 - 適合度統計量, 33, 35, 66
 - 予測期間, 31, 64, 87
 - 信頼区間, 35, 40
 - 推定期間, 31
 - 欠損値, 40
 - 予測, 33, 35, 88
 - Box-Ljung 統計量, 33
 - 再度推定されたモデルを XML 形式で保存する, 38
 - 新しい変数名, 38, 67
 - すべてのモデルにわたる統計量, 33, 35, 66
 - 当てはめ値, 35
 - 予測値の表, 88
 - 最適モデルと最不適合モデル, 37
 - モデルの適合度表, 66
 - モデルのパラメータ, 33
 - モデルのパラメータを再度推定, 31, 64
 - 予測を保存する, 38, 65
- モデルのパラメータ
 - 時系列モデラー, 20, 75
 - 時系列モデルの適用, 33
- モデルのパラメータを再度推定
 - 時系列モデルの適用, 31, 64

- レベル シフトの外れ値
 - 時系列モデラー, 10, 18
- レベル シフトの外れ値, 107

- 和分
 - ARIMA モデル, 14