

IBM SPSS Forecasting 21



Remarque : Avant d'utiliser ces informations et le produit qu'elles concernent, lisez les informations générales sous Remarques sur p. 115.

Cette version s'applique à IBM® SPSS® Statistics 21 et à toutes les publications et modifications ultérieures jusqu'à mention contraire dans les nouvelles versions.

Les captures d'écran des produits Adobe sont reproduites avec l'autorisation de Adobe Systems Incorporated.

Les captures d'écran des produits Microsoft sont reproduites avec l'autorisation de Microsoft Corporation.

Matériel sous licence - Propriété d'IBM

© Copyright IBM Corporation 1989, 2012.

Droits limités pour les utilisateurs au sein d'administrations américaines : utilisation, copie ou divulgation soumise au GSA ADP Schedule Contract avec IBM Corp.

Préface

IBM® SPSS® Statistics est un système complet d'analyse de données. Le module complémentaire facultatif Prévisions fournit les techniques d'analyse supplémentaires décrites dans ce manuel. Le module complémentaire Prévisions doit être utilisé avec le système central SPSS Statistics auquel il est entièrement intégré.

A propos de IBM Business Analytics

Le logiciel IBM Business Analytics offre des informations complètes, cohérentes et précises permettant aux preneurs de décision d'améliorer leurs performances professionnelles. Un portefeuille complet de solutions de [business intelligence](#), [d'analyses prédictives](#), [de performance financière et de gestion de la stratégie](#), et [d'applications analytiques](#) permet une connaissance claire et immédiate et offre des possibilités d'actions sur les performances actuelles et la capacité de prédire les résultats futurs. En combinant des solutions du secteur, des pratiques prouvées et des services professionnels, les entreprises de toute taille peuvent générer la plus grande productivité, automatiser les décisions en toute confiance et apporter de meilleurs résultats.

Dans le cadre de ce portefeuille, le logiciel IBM SPSS Predictive Analytics aide les entreprises à prédire des événements futurs et à agir de manière proactive en fonction de ces prédictions pour apporter de meilleurs résultats. Des clients dans les domaines commerciaux, gouvernementaux et académiques se servent de la technologie IBM SPSS comme d'un avantage concurrentiel pour attirer ou retenir des clients, tout en réduisant les risques liés à l'incertitude et à la fraude. En intégrant le logiciel IBM SPSS à leurs opérations quotidiennes, les entreprises peuvent effectuer des prévisions, et sont capables de diriger et d'automatiser leurs décisions afin d'atteindre leurs objectifs commerciaux et d'obtenir des avantages concurrentiels mesurables. Pour plus d'informations ou pour contacter un représentant, visitez le site <http://www.ibm.com/spss>.

Support technique

Un support technique est disponible pour les clients du service de maintenance. Les clients peuvent contacter l'assistance technique pour obtenir de l'aide concernant l'utilisation des produits IBM Corp. ou l'installation dans l'un des environnements matériels pris en charge. Pour contacter l'assistance technique, visitez le site IBM Corp. à l'adresse <http://www.ibm.com/support>. Votre nom, celui de votre société, ainsi que votre contrat d'assistance vous seront demandés.

Support technique pour les étudiants

Si vous êtes un étudiant qui utilise la version pour étudiant, personnel de l'éducation ou diplômé d'un produit logiciel IBM SPSS, veuillez consulter les pages [Solutions pour l'éducation](#) (<http://www.ibm.com/spss/rd/students/>) consacrées aux étudiants. Si vous êtes un étudiant utilisant une copie du logiciel IBM SPSS fournie par votre université, veuillez contacter le coordinateur des produits IBM SPSS de votre université.

Service clients

Si vous avez des questions concernant votre livraison ou votre compte, contactez votre bureau local. Veuillez préparer et conserver votre numéro de série à portée de main pour l'identification.

Séminaires de formation

IBM Corp. propose des séminaires de formation, publics et sur site. Tous les séminaires font appel à des ateliers de travaux pratiques. Ces séminaires seront proposés régulièrement dans les grandes villes. Pour plus d'informations sur ces séminaires, accédez au site <http://www.ibm.com/software/analytics/spss/training>.

Contenu

Partie I: Guide de l'utilisateur

1 Introduction aux séries chronologiques 1

| | |
|---|---|
| Données de séries chronologiques | 1 |
| Transformations de données. | 2 |
| Périodes d'estimation et de validation. | 2 |
| Création de modèles et prévisions | 2 |

2 Time Series Modeler 4

| | |
|--|----|
| Spécification des options pour le modélisateur expert | 8 |
| Sélection de modèle et spécification d'événement. | 8 |
| Traitement des valeurs éloignées avec le modélisateur expert. | 10 |
| Modèles de lissage exponentiel personnalisés | 11 |
| Modèles ARIMA personnalisés. | 12 |
| Spécification de modèle pour les modèles ARIMA personnalisés. | 13 |
| Fonctions de transfert dans des modèles ARIMA personnalisés. | 15 |
| Valeurs éloignées dans les modèles ARIMA personnalisés | 17 |
| Résultats | 18 |
| Tableaux de statistiques et de prévision. | 19 |
| Diagrammes | 21 |
| Limitation des résultats aux modèles les plus ou les moins appropriés. | 23 |
| Enregistrement des prévisions et des spécifications de modèle | 24 |
| Options. | 25 |
| Fonctionnalités supplémentaires de la commande TSMODEL | 27 |

3 Appliquer des modèles de séries chronologiques 28

| | |
|--|----|
| Résultats | 31 |
| Tableaux de statistiques et de prévision. | 31 |
| Diagrammes | 33 |
| Limitation des résultats aux modèles les plus ou les moins appropriés. | 35 |
| Enregistrement des prévisions et des spécifications de modèle | 36 |
| Options. | 37 |
| Fonctionnalités supplémentaires de la commande TSAPPLY | 38 |

| | | |
|--------------------------------|--|-----------|
| 4 | <i>Désaisonnalisation</i> | 39 |
| | Enregistrer la désaisonnalisation | 40 |
| | Fonctionnalités supplémentaires de la commande SEASON | 41 |
| 5 | <i>Analyse spectrale</i> | 42 |
| | Fonctionnalités supplémentaires de la commande SPECTRA | 44 |
| Partie II: Exemples | | |
| 6 | <i>Prévision globale avec le modélisateur expert</i> | 47 |
| | Examen de vos données | 47 |
| | Exécution de l'analyse | 49 |
| | Diagrammes récapitulatifs du modèle. | 55 |
| | Prévisions du modèle | 56 |
| | Récapitulatif | 57 |
| 7 | <i>Reprévisions globales en appliquant des modèles enregistrés</i> | 58 |
| | Exécution de l'analyse | 58 |
| | Statistiques d'ajustement du modèle. | 61 |
| | Prévisions du modèle | 62 |
| | Récapitulatif | 62 |
| 8 | <i>Utilisation du modélisateur expert pour déterminer les variables prédites significatives</i> | 63 |
| | Représentation de vos données | 63 |
| | Exécution de l'analyse | 65 |
| | Tracé de séries. | 71 |
| | Tableau de description du modèle. | 71 |
| | Tableau de statistiques du modèle | 72 |

| | |
|--|-----------|
| Tableau des paramètres du modèle ARIMA | 73 |
| Récapitulatif | 73 |
| 9 Test des variables indépendantes en appliquant des modèles enregistrés | 74 |
| Extension des séries de variables indépendantes | 74 |
| Modification des valeurs de prévision dans la période de prévision | 78 |
| Exécution de l'analyse | 80 |
| 10 Désaisonnalisation | 85 |
| Suppression de la saisonnalité dans les données de vente | 85 |
| Détermination et paramétrage de la périodicité | 85 |
| Exécution de l'analyse | 90 |
| Présentation des résultats | 91 |
| Récapitulatif | 92 |
| Procédures apparentées | 92 |
| 11 Analyse spectrale | 93 |
| Utilisation de la procédure Analyse spectrale pour vérifier les prévisions en matière de périodicité | 93 |
| Exécution de l'analyse | 93 |
| Présentation du périodogramme et de la densité spectrale | 95 |
| Récapitulatif | 97 |
| Procédures apparentées | 97 |

Annexes

| | | |
|-----------------|---|-------------------|
| <i>A</i> | <i>Mesures de la qualité d'ajustement</i> | <i>98</i> |
| <i>B</i> | <i>Types de valeurs éloignées</i> | <i>99</i> |
| <i>C</i> | <i>Guide relatif aux diagrammes ACF/PACF</i> | <i>100</i> |
| <i>D</i> | <i>Fichiers d'exemple</i> | <i>104</i> |
| <i>E</i> | <i>Remarques</i> | <i>115</i> |
| | <i>Bibliographie</i> | <i>118</i> |
| | <i>Index</i> | <i>119</i> |

Partie I: ***Guide de l'utilisateur***

Introduction aux séries chronologiques

Une **série chronologique** est un ensemble d'observations obtenues par mesure régulière d'une variable sur une certaine période. Dans une série de données de stock, par exemple, les observations peuvent représenter des niveaux de stock quotidiens pour plusieurs mois. Une série indiquant la part de marché d'un produit peut représenter la part de marché hebdomadaire relevée sur un certain nombre d'années. Une série comprenant des chiffres relatifs aux ventes totales peut représenter une observation par mois pendant plusieurs années. Tous ces exemples indiquent qu'une variable a été observée à intervalles réguliers au cours d'une certaine période. Par conséquent, les données d'une série chronologique standard représentent une séquence unique ou une liste d'observations représentant des mesures effectuées à intervalles réguliers.

Table 1-1
Série chronologique quotidienne de stock

| Heure | Semaine | Jour | Niveau de stock |
|-----------------|---------|----------|-----------------|
| t ₁ | 1 | Lundi | 160 |
| t ₂ | 1 | Mardi | 135 |
| t ₃ | 1 | Mercredi | 129 |
| t ₄ | 1 | Jeudi | 122 |
| t ₅ | 1 | Vendredi | 108 |
| t ₆ | 2 | Lundi | 150 |
| | | ... | |
| t ₆₀ | 12 | Vendredi | 120 |

L'analyse d'une série chronologique est généralement due au besoin de prévoir des valeurs ultérieures. Un modèle de série ayant expliqué les valeurs précédentes peut également indiquer si les valeurs suivantes augmenteront ou diminueront, et à quel degré. La capacité d'effectuer de telles prévisions est évidemment importante pour les secteurs commercial et scientifique.

Données de séries chronologiques

Lorsque vous définissez les données d'une série chronologique à utiliser avec le module complémentaire Prédiction, chaque série correspond à une variable distincte. Par exemple, pour définir une série chronologique dans l'éditeur de données, cliquez sur l'onglet *Affichage des variables*, puis entrez un nom de variable dans une ligne vide. Chaque observation d'une série chronologique correspond à une observation apparaissant sous la forme d'une ligne dans l'éditeur de données.

Si vous ouvrez une feuille de calcul contenant des données de séries chronologiques, chaque série doit être disposée dans une colonne de la feuille de calcul. Si vous avez déjà une feuille de calcul contenant des séries chronologiques disposées en lignes, vous pouvez l'ouvrir et utiliser l'option *Transposer* du menu *Données* pour transformer les lignes en colonnes.

Transformations de données

Un certain nombre de procédures de transformation de données fournies dans le système central sont utiles dans les analyses de séries chronologiques.

- La procédure Définir des dates (menu Données) génère des variables de date utilisées pour établir une périodicité, et faire la distinction entre les périodes historiques, de validation et de prévision. La prévision est conçue pour être utilisée avec les variables créées par la procédure Définir des dates.
- La procédure Créer des séries chronologiques (dans le menu Transformer) permet de créer des variables de séries chronologiques sous forme de fonctions de variables de séries chronologiques existantes. Elle inclut les fonctions qui utilisent les remarques de voisinage pour le lissage, le calcul de la moyenne et la différenciation.
- La procédure Remplacer les valeurs manquantes (menu Transformer) remplace les valeurs manquantes spécifiées et par défaut par des estimations basées sur une méthode. Les données manquantes situées au début ou à la fin d'une série ne posent aucun problème : elles réduisent simplement la longueur utile de la série. Les trous situés au milieu d'une série (données manquantes *intégrées*) peuvent représenter un problème beaucoup plus sérieux.

Pour obtenir des informations détaillées sur la transformation des données des séries chronologiques, reportez-vous au manuel *Guide de l'utilisateur du système central*.

Périodes d'estimation et de validation

Il est souvent utile de diviser une série chronologique en une période *historique* ou *d'estimation* et en une période de *validation*. Développez un modèle sur la base des remarques figurant dans la période d'estimation (historique), puis testez-le pour vérifier son fonctionnement dans la période de validation. En forçant le modèle à effectuer des prévisions pour les points que vous connaissez déjà (les points de la période de validation), vous obtenez une idée de la capacité de prévision du modèle.

Les observations de la période de validation sont habituellement qualifiées d'observations traitées, car elles sont récupérées du processus de création du modèle. La période d'estimation est constituée des observations actuellement sélectionnées dans l'ensemble de données actif. Toute observation suivant la dernière observation sélectionnée constitue une observation exclue. Une fois que la capacité de prévision du modèle vous apporte entière satisfaction, vous pouvez redéfinir la période d'estimation afin d'inclure les observations exclues et de créer ensuite votre modèle final.

Création de modèles et prévisions

Le module complémentaire Prévisions offre deux procédures de création de modèles et de production de prévisions.

- La procédure [Time Series Modeler](#) crée des modèles pour les séries chronologiques et produit des prévisions. Elle inclut un modélisateur expert qui détermine automatiquement le modèle le plus adapté à chaque série chronologique. Pour les analystes expérimentés, souhaitant

bénéficier d'un plus grand degré de contrôle, elle comprend également des outils de création de modèles personnalisés.

- La procédure [Appliquer des modèles de séries chronologiques](#) applique des modèles de séries chronologiques existants, créés à l'aide du modélisateur de séries chronologiques, à l'ensemble de données actif. Ceci vous permet d'obtenir des prévisions pour des séries pour lesquelles de nouvelles données ou des données révisées sont disponibles, sans avoir à recréer les modèles. Tout modèle susceptible d'avoir changé peut ainsi être recréé à l'aide du modélisateur de séries chronologiques.

Time Series Modeler

La procédure de modélisation de série chronologique évalue le lissage exponentiel, le processus autorégressif moyenne mobile intégré (ARIMA - AutoRegressive Integrated Moving Average) univarié et les modèles ARIMA multivariés (ou modèles des fonctions de transfert) pour les séries chronologiques, et produit des prévisions. La procédure inclut un modélisateur expert qui identifie et évalue automatiquement le meilleur ARIMA ou modèle de lissage exponentiel pour une ou plusieurs séries de variables dépendantes, éliminant ainsi le besoin d'identifier un modèle approprié via des tentatives et des erreurs. Vous pouvez également spécifier un modèle ARIMA ou un modèle de lissage exponentiel personnalisé.

Exemple : Vous êtes un chef de produit et êtes responsable des prévisions des ventes de référence et des revenus du mois à venir pour 100 produits distincts et avez peu ou pas du tout d'expérience en modélisation de séries chronologiques. Les données historiques des ventes de référence pour les 100 produits sont stockées dans une feuille de calcul Excel unique. Après avoir ouvert la feuille de calcul dans IBM® SPSS® Statistics, utilisez le modélisateur expert et demandez les prévisions pour un mois à l'avance. Le modélisateur expert trouve le meilleur modèle de vente de référence pour chacun de vos produits et utilise ces modèles pour produire les prévisions. Le modélisateur expert peut traiter plusieurs séries d'entrées, vous n'avez besoin d'exécuter la procédure qu'une seule fois afin d'obtenir des prévisions pour tous vos produits. En choisissant d'enregistrer les prévisions dans l'ensemble de données actif, vous pouvez facilement exporter à nouveau les résultats sur Excel.

Statistiques. Mesures de la qualité de l'ajustement : R -deux stationnaire, R -deux (R^2), erreur quadratique moyenne (RMSE), Erreur Absolue Moyenne (MAE), Erreur de Pourcentage en Moyenne absolue (MAPE), Erreur Absolue Maximum (MaxAE), Pourcentage d'Erreur Absolue Maximum (MaxAPE), critère d'information bayésien normalisé (BIC). Résidus : fonction d'autocorrélation, fonction d'autocorrélation partielle, Q de Ljung-Box. Pour les modèles ARIMA : ARIMA ordonne des variables dépendantes, la fonction de transfert ordonne des variables indépendantes et les variables éloignées estiment. De même, le paramètre de lissage estime les modèles de lissage exponentiel.

Diagrammes : Diagrammes récapitulatifs parmi tous les modèles : histogrammes de R -deux stationnaire, R -deux (R^2), erreur quadratique moyenne (RMSE), Erreur Absolue Moyenne (MAE), Erreur de Pourcentage en Moyenne absolue (MAPE), Erreur Absolue Maximum (MaxAE), Pourcentage d'Erreur Absolue Maximum (MaxAPE), critère d'information bayésien normalisé (BIC) ; diagrammes à surfaces des autocorrélations partielles et résiduelles. Résultats des modèles individuels : valeurs de prévision, valeurs d'ajustement, valeurs observées, limites de confiance supérieure et inférieure, autocorrélations partielles et résiduelles.

Remarques sur les données du modélisateur de séries chronologiques

Données. La variable dépendante et les variables indépendantes doivent être numériques.

Hypothèses : La variable dépendante et toute variable indépendante sont traitées en tant que séries chronologiques, ce qui veut dire que chaque observation représente un point dans le temps, avec des observations successives séparées par un intervalle de temps constant.

- **Stationnarité.** Pour les modèles ARIMA personnalisés, les séries chronologiques à modéliser doivent être stationnaires. La méthode la plus efficace pour transformer une série non stationnaire en une série stationnaire nécessite l'utilisation d'une différence, disponible dans la boîte de dialogue Créer la série chronologique.
- **Prévisions.** Pour produire des prévisions à l'aide de modèles avec des variables indépendantes (explicatives), l'ensemble de données actif doit contenir des valeurs de ces variables pour toutes les observations de la période de prévision. En outre, les variables indépendantes ne doivent pas contenir de valeurs manquantes dans la période d'estimation.

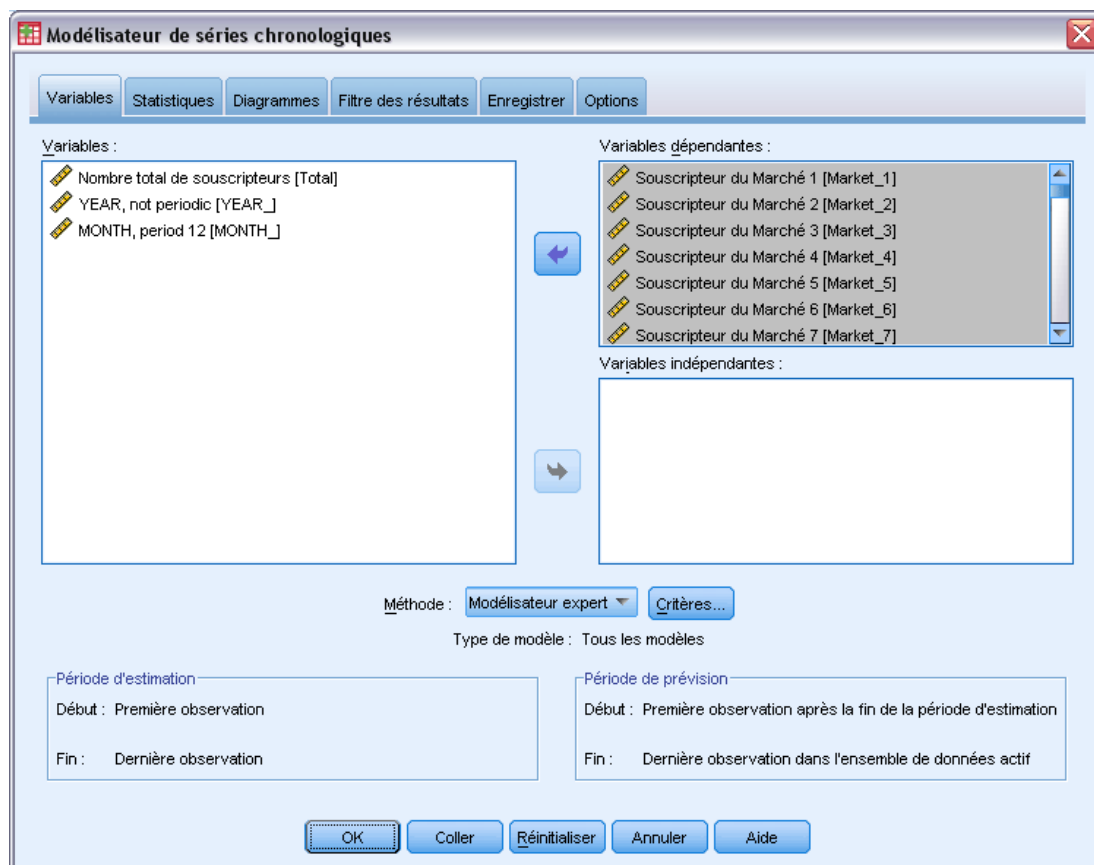
Définition des dates

Bien que ce ne soit pas une obligation, il est recommandé d'utiliser la boîte de dialogue Définir des dates pour indiquer la date correspondant à la première observation et l'intervalle de temps entre les observations suivantes. Ceci est effectué avant d'utiliser le modélisateur de séries chronologiques et résulte en un ensemble de variables qui indiquent la date associée à chaque observation. Ceci définit également une périodicité supposée des données, par exemple une périodicité de 12 si l'intervalle de temps entre les observations suivantes est d'un mois. Cette périodicité est requise si vous voulez créer des modèles saisonniers. Si vous ne voulez pas créer de modèles saisonniers et ne voulez pas d'étiquette de date dans vos résultats, vous pouvez ignorer la boîte de dialogue Définir des dates. L'étiquette associée à chaque observation est alors simplement le numéro de l'observation.

Pour utiliser le modélisateur de séries chronologiques

- ▶ A partir des menus, sélectionnez :
Analyse > Prévisions > Créer des modèles...

Figure 2-1
Modélisateur de séries chronologiques, onglet Variables



- ▶ Dans l'onglet Variables, sélectionnez une ou plusieurs variables dépendantes à modéliser.
- ▶ Depuis la liste déroulante des méthodes, sélectionnez une méthode de modélisation. Pour la modélisation automatique, conservez la méthode du modélisateur expert par défaut. Vous invoquerez ainsi le modélisateur expert pour déterminer le modèle le plus approprié pour chacune des variables dépendantes.

Pour produire des prévisions :

- ▶ Cliquez sur l'onglet Options.
- ▶ Indiquez la période de prévision. Un diagramme sera produit, incluant les prévisions et les valeurs observées.

Sinon, vous pouvez :

- Sélectionnez une ou plusieurs variables indépendantes. Les variables indépendantes sont traitées comme des variables explicatives dans les analyses de régression, mais sont facultatives. Elles peuvent être incluses dans les modèles ARIMA mais pas dans les modèles de lissage exponentiel. Si vous spécifiez le modélisateur expert comme méthode de modélisation et incluez des variables indépendantes, seuls les modèles ARIMA seront pris en considération.

- Cliquez sur Critères pour indiquer les détails de modélisation.
- [Enregistrer les prévisions, les intervalles de confiance et les résidus du bruit.](#)
- [Enregistrez les modèles estimés au format XML.](#) Les modèles enregistrés peuvent être appliqués aux données nouvelles ou revus pour obtenir des prévisions mises à jour sans besoin de reconstruire les modèles. Ceci est effectué via la procédure [Appliquer des modèles de séries chronologiques.](#)
- [Obtenir des statistiques récapitulatives dans tous les modèles estimés.](#)
- [Spécifier les fonctions de transfert pour les variables indépendantes dans les modèles ARIMA personnalisés.](#)
- [Activer la détection automatique de valeurs éloignées.](#)
- [Points dans le temps spécifiques à un modèle comme valeurs éloignées pour les modèles ARIMA personnalisés.](#)

Méthodes de modélisation

Les méthodes de modélisation disponibles sont :

Modélisateur expert Le modélisateur expert trouve automatiquement le modèle le plus approprié pour chaque série dépendante. Si les variables indépendantes (explicatives) sont indiquées, le modélisateur expert sélectionne des variables ayant une relation statistiquement significative avec la série dépendante pour les inclure dans les modèles ARIMA. Les variables de modèle sont transformées le cas échéant à l'aide de transformation par différenciation, par racine carrée ou par log népérien. Par défaut, le modélisateur expert prend en compte les modèles de lissage exponentiel et les modèles ARIMA. Toutefois, vous pouvez limiter le modélisateur expert pour qu'il ne recherche que les modèles ARIMA ou les modèles de lissage exponentiel. Vous pouvez également indiquer la détection automatique des valeurs éloignées.

Lissage exponentiel. Utilisez cette option pour indiquer un modèle de lissage exponentiel personnalisé. Vous pouvez choisir parmi une variété de modèles exponentiels qui diffèrent dans leur manière de traiter la tendance et la saison.

ARIMA. Utilisez cette option pour indiquer un modèle ARIMA personnalisé. Ceci implique de spécifier de manière explicite les ordres autorégressifs et de moyenne mobile ainsi que le degré de différenciation. Vous pouvez inclure des variables indépendantes (explicatives) et définir les fonctions de transfert pour certaines d'entre elles ou pour la totalité. Vous pouvez également spécifier la détection automatique des valeurs éloignées ou spécifier un ensemble explicite de valeurs éloignées.

Périodes d'estimation et de prévision

Période d'estimation. La période d'estimation définit l'ensemble des observations utilisées pour déterminer le modèle. Par défaut la période d'estimation inclut toutes les observations dans l'ensemble de données actif. Pour définir la période d'estimation, sélectionnez Dans un intervalle de temps ou d'observations dans la boîte de dialogue Sélectionner des observations. En fonction des données disponibles, la période d'estimation utilisée par la procédure peut varier selon la variable dépendante et donc être différente de la valeur affichée. Pour une variable dépendante donnée, la vraie période d'estimation est la période restante après l'élimination de toutes les

valeurs manquantes contiguës de la variable qui se produit au début ou à la fin de la période d'estimation spécifiée.

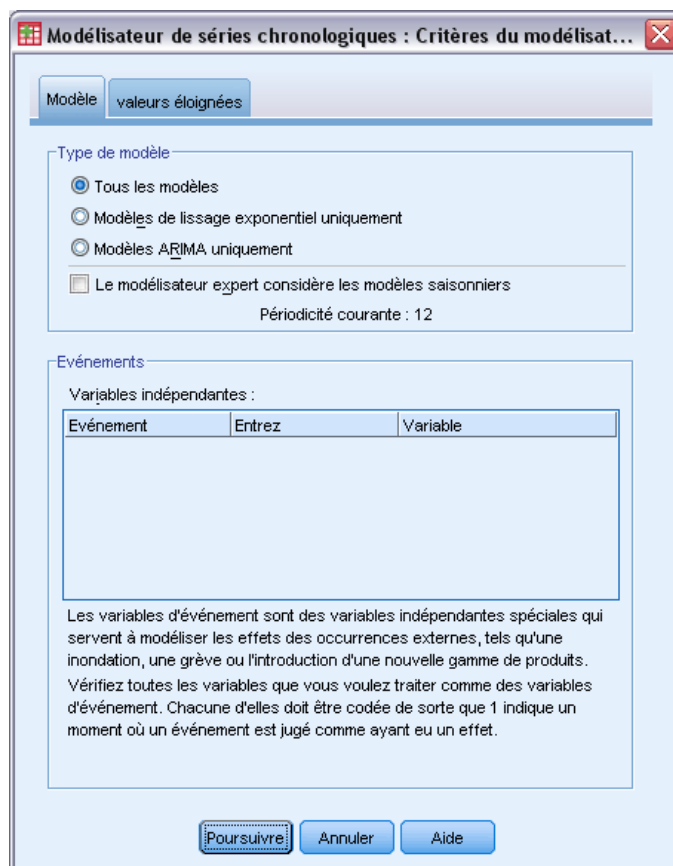
Période de prévision. La période de prévision commence à la première observation après chaque période d'estimation et, par défaut, continue jusqu'à la dernière observation dans l'ensemble de données actif. Vous pouvez définir la fin d'une période de prévision à partir de l'onglet [Options](#).

Spécification des options pour le modélisateur expert

Le modélisateur expert fournit des options pour contraindre l'ensemble de modèles possibles, spécifiant le traitement des valeurs éloignées, y compris les variables d'événement.

Sélection de modèle et spécification d'événement

Figure 2-2
Boîte de dialogue Critères du modélisateur expert, onglet Modèle



L'onglet Modèle vous permet d'indiquer les types de modèles pris en compte par le modélisateur expert et de spécifier les variables d'événement.

Type de modèle. Les options suivantes sont disponibles :

- **Tous les modèles.** Le modélisateur expert considère aussi bien les modèles ARIMA que ceux de lissage exponentiel.
- **Modèles de lissage exponentiel uniquement.** Le modélisateur expert considère uniquement les modèles de lissage exponentiel.
- **Modèles ARIMA uniquement.** Le modélisateur expert considère uniquement les modèles ARIMA.

Le modélisateur expert considère les modèles saisonniers. Cette option est uniquement activée si une périodicité a été définie pour l'ensemble de données actif. Une fois cette option sélectionnée (cochée), le modélisateur expert considère les modèles à la fois saisonniers et non saisonniers. Si cette option n'est pas sélectionnée, le modélisateur expert considère uniquement les modèles non saisonniers.

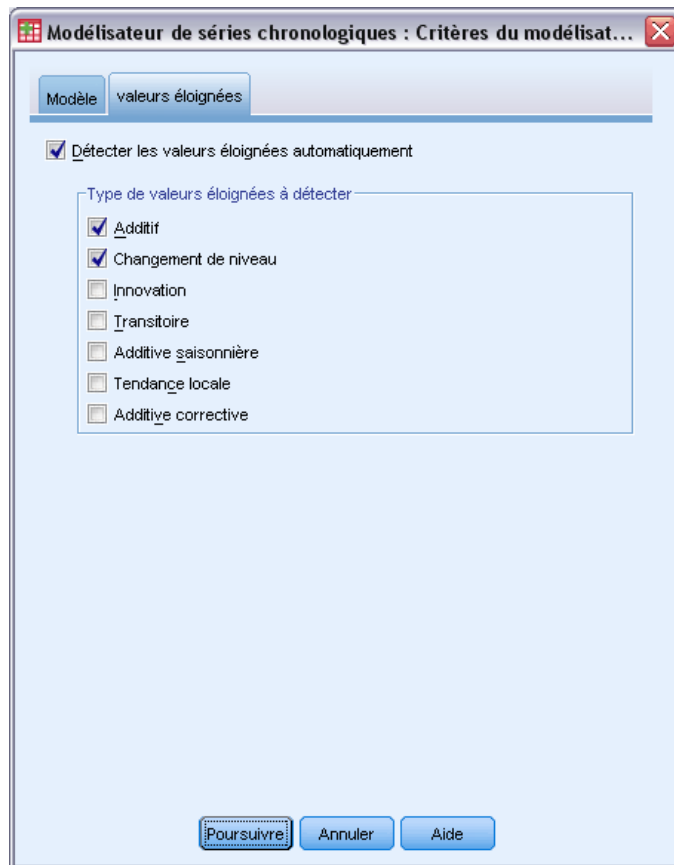
Périodicité courante. Indique la périodicité courante (le cas échéant) définie pour le fichier de travail. La périodicité courante est indiquée par une valeur entière, par exemple 12 pour la périodicité annuelle, chaque observation représentant un mois. La valeur *Aucune* est affichée si aucune périodicité n'a été définie. Les modèles saisonniers nécessitent une périodicité. Vous pouvez définir la périodicité dans la boîte de dialogue Définir des dates.

Événements. Sélectionnez les variables indépendantes qui doivent être traitées comme variables d'événement. Pour les variables d'événement, les observations avec une valeur de 1 indiquent les moments auxquels les séries indépendantes seront affectées par l'événement. Les valeurs autres que 1 n'indiquent aucun effet.

Traitement des valeurs éloignées avec le modélisateur expert

Figure 2-3

Boîte de dialogue Critères du modélisateur expert, onglet Valeurs éloignées



L'onglet Valeurs éloignées vous permet de choisir la détection automatique de valeurs éloignées ainsi que le type de valeurs éloignées à détecter.

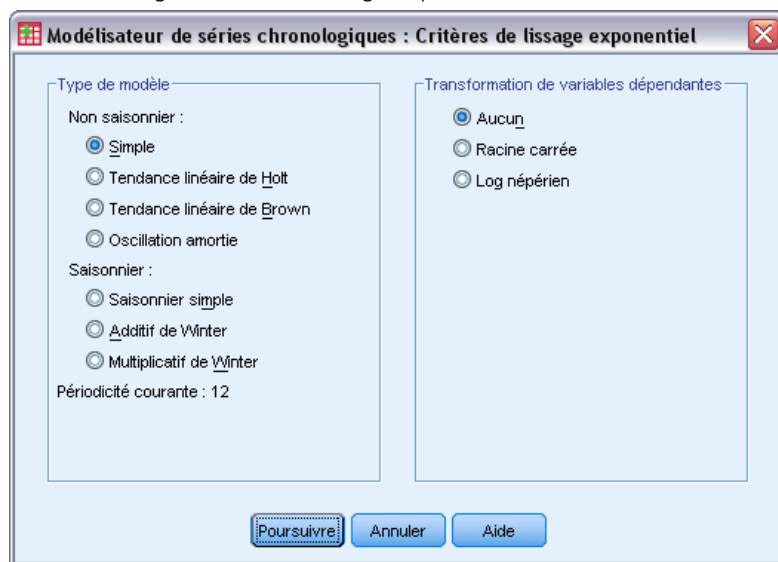
Détecter les valeurs éloignées automatiquement. Par défaut, la détection automatique de valeurs éloignées n'est pas exécutée. Sélectionnez (cochez) cette option pour exécuter la détection automatique de valeurs éloignées et sélectionnez un ou plusieurs des types de valeurs éloignées suivants :

- Additif
- Changement de niveau
- Innovation
- Transitoire
- Additive saisonnière
- Tendance locale
- Additive corrective

Pour plus d'informations, reportez-vous à la section Types de valeurs éloignées dans l'annexe B sur p. 99.

Modèles de lissage exponentiel personnalisés

Figure 2-4
Boîte de dialogue Critère de lissage exponentiel



Type de modèle. Les modèles de lissage exponentiel (Gardner, 1985) sont classés comme saisonniers ou non saisonniers. Les modèles saisonniers sont uniquement disponibles si une périodicité a été définie pour l'ensemble de données actif (voir « Périodicité courante » ci-dessous).

- **Simple.** Ce modèle est approprié pour les séries ne contenant aucune tendance ni saison. Son seul paramètre de lissage est le niveau. Le lissage exponentiel simple est très similaire à un modèle ARIMA avec zéro ordre d'autorégression, un ordre de différenciation, un ordre de moyenne mobile et aucune constante.
- **Tendance linéaire de Brown.** Ce modèle est approprié pour les séries contenant une tendance linéaire et pas de saison. Ses paramètres de lissage sont le niveau et la tendance qui ne sont pas contraints par leurs valeurs réciproques. Le modèle de Holt est plus général que le modèle de Brown, mais prend plus de temps à calculer pour les grandes séries. Le lissage exponentiel de Holt est très similaire à un modèle ARIMA avec zéro ordre d'autorégression, deux ordres de différenciation et deux ordres de moyenne mobile.
- **Tendance linéaire de Holt.** Ce modèle est approprié pour les séries contenant une tendance linéaire et aucun effet saisonnier. Ses paramètres de lissage sont le niveau et la tendance que l'on considère comme égaux. Par conséquent, le modèle de Brown est un cas particulier du modèle de Holt. Le lissage exponentiel de Brown est très similaire à un modèle ARIMA avec zéro ordre d'autorégression, deux ordres de différenciation et deux ordres de moyenne mobile, où le coefficient pour le second ordre de moyenne mobile est égal au carré de la moitié du coefficient du premier ordre.

- **Tendance amortie.** Ce modèle est approprié pour les séries avec une tendance linéaire qui décroît et pas de saison. Ses paramètres de lissage sont le niveau, la tendance et l'oscillation amortie. Le modèle de lissage exponentiel est très similaire à un modèle ARIMA avec un ordre d'autorégression, 1 ordre de différentiation et 2 ordres de moyenne mobile.
- **Saisonnier simple.** Ce modèle est approprié pour les séries sans tendance et avec un effet saisonnier constant dans le temps. Ses paramètres de lissage sont le niveau et la saison. Le lissage exponentiel saisonnier simple est similaire à un modèle ARIMA d'ordre 0 pour le terme d'auto régression, d'ordre 1 pour la différence première et d'ordre 1, p et $p+1$ pour la moyenne mobile, où p est le nombre de périodes dans l'intervalle saisonnier (pour des séries mensuelles, $p = 12$).
- **Additif de Winter.** Ce modèle est approprié pour les séries avec une tendance linéaire et un effet saisonnier qui ne dépend pas du niveau de la série. Ses paramètres de lissage sont le niveau, la tendance et la saison. Le lissage exponentiel additif de Winter est similaire à un modèle ARIMA d'ordre 0 pour le terme d'auto régression, d'ordre 1 pour la différence première, d'ordre 1 pour la différence saisonnière, et $p+1$ pour la moyenne mobile, où p est le nombre de périodes dans l'intervalle saisonnier (pour des séries mensuelles, $p = 12$).
- **Multiplicatif de Winter.** Ce modèle est approprié pour les séries avec une tendance linéaire et un effet saisonnier qui dépend du niveau de la série. Ses paramètres de lissage sont le niveau, la tendance et la saison. Le lissage exponentiel multiplicatif de Winter n'est pas similaire à un modèle ARIMA.

Périodicité courante. Indique la périodicité courante (le cas échéant) définie pour le fichier de travail. La périodicité courante est indiquée par une valeur entière, par exemple 12 pour la périodicité annuelle, chaque observation représentant un mois. La valeur *Aucune* est affichée si aucune périodicité n'a été définie. Les modèles saisonniers nécessitent une périodicité. Vous pouvez définir la périodicité dans la boîte de dialogue Définir des dates.

Transformation de variables dépendantes. Vous pouvez spécifier une transformation exécutée sur chaque variable dépendante avant de la modéliser.

- **Aucune.** Aucune transformation n'est effectuée.
- **Racine carrée :** Transformation racine carrée.
- **Log népérien :** Transformation par log naturel.

Modèles ARIMA personnalisés

Le modélisateur de séries chronologiques vous permet de construire des modèles ARIMA (AutoRegressive Integrated Moving Average - processus autorégressif moyenne mobile intégré) saisonniers et non saisonniers personnalisés (également connus sous le nom de modèles Box-Jenkins (Box, Jenkins, et Reinsel, 1994)) avec ou sans ensemble fixe de variables explicatives. Vous pouvez définir des fonctions de transfert pour une ou toutes les variables explicatives et indiquer la détection automatique de valeurs éloignées ou spécifier un ensemble explicite de valeurs éloignées.

- Toutes les variables indépendantes (explicatives) indiquées dans l'onglet Variables sont explicitement incluses dans le modèle. Ceci se démarque de l'utilisation du modélisateur expert dans lequel les variables indépendantes sont uniquement incluses si elles ont une relation statistiquement significative avec la variable dépendante.

Spécification de modèle pour les modèles ARIMA personnalisés

Figure 2-5
Boîte de dialogue Critères ARIMA, onglet Modèle

Modélisateur de séries chronologiques : Critères ARIMA

Modèle valeurs éloignées

Ordres ARIMA

Structure :

| | Non saisonnier | Saisonnier |
|--------------------|----------------|------------|
| Autorégressif (p) | 0 | 0 |
| Différence (d) | 0 | 0 |
| Moyenne mobile (q) | 0 | 0 |

Périodicité courante : 12

Transformation

Aucun

Racine carrée

Log népérien

Inclure terme constant dans le modèle

Poursuivre Annuler Aide

L'onglet Modèle vous permet de spécifier la structure d'un modèle ARIMA personnalisé.

Ordres ARIMA. Entrez des valeurs pour les différentes composantes ARIMA de votre modèle dans les cellules correspondantes de la grille Structure. Toutes les valeurs doivent être des entiers non négatifs. Pour les composants autorégressifs et de moyenne mobile, la valeur représente l'ordre maximum. Tous les ordres inférieurs positifs seront inclus dans le modèle. Par exemple, si vous indiquez 2, le modèle inclut les ordres 2 et 1. Les cellules de la colonne Saisonnier sont uniquement activées si une périodicité a été définie pour l'ensemble de données actif (voir « Périodicité courante » ci-dessous).

- **Autorégressif (p).** Le nombre d'ordres autorégressifs dans le modèle. Les ordres autorégressifs indiquent quelles valeurs précédentes de la série seront utilisées pour prévoir les valeurs en cours. Par exemple, un ordre autorégressif de 2 indique que la valeur de la série Deux points dans le temps dans le passé sera utilisée pour prévoir la valeur en cours.
- **Différence (d).** Spécifie l'ordre de différenciation appliqué à la série avant d'estimer les modèles. La différenciation est nécessaire lorsque les tendances sont présentes (les séries avec tendances sont en général non stationnaires et la modélisation ARIMA suppose la stationnarité) et est utilisée pour supprimer leurs effets. L'ordre de différenciation correspond au degré de tendance de série, aux comptes de différenciation de premier ordre pour les

tendances linéaires, aux comptes de différenciation de second ordre pour les tendances quadratiques, etc.

- **Moyenne mobile (q).** Le nombre d'ordres de moyenne mobile dans le modèle. Les ordres de moyenne mobile indiquent comment les écarts de la moyenne de la série pour les valeurs précédentes sont utilisés pour prévoir les valeurs courantes. Par exemple, des ordres de moyenne mobile de 1 et 2 indiquent que les écarts de la valeur de la moyenne de la série pour chacune des deux dernières périodes doivent être considérés lors de la prévision des valeurs actuelles de la série.

Ordres saisonniers. Les composants autorégressifs, de moyenne mobile et de différenciation saisonniers tiennent le même rôle que leurs équivalents non saisonniers. Cependant, pour les ordres saisonniers, les valeurs courantes de la série sont affectées par les valeurs de série précédentes séparées par une ou plusieurs périodes saisonnières. Par exemple, pour des données mensuelles (période saisonnière de 12), un ordre saisonnier de 1 indique que la valeur de série en cours est affectée par les 12 périodes de la valeur de série précédant celle en cours. Un ordre saisonnier de 1, pour des données mensuelles, est alors le même que lorsqu'on spécifie un ordre non saisonnier de 12.

Périodicité courante. Indique la périodicité courante (le cas échéant) définie pour le fichier de travail. La périodicité courante est indiquée par une valeur entière, par exemple 12 pour la périodicité annuelle, chaque observation représentant un mois. La valeur *Aucune* est affichée si aucune périodicité n'a été définie. Les modèles saisonniers nécessitent une périodicité. Vous pouvez définir la périodicité dans la boîte de dialogue Définir des dates.

Transformation de variables dépendantes. Vous pouvez spécifier une transformation exécutée sur chaque variable dépendante avant de la modéliser.

- **Aucune.** Aucune transformation n'est effectuée.
- **Racine carrée :** Transformation racine carrée.
- **Log népérien :** Transformation par log naturel.

Inclure terme constant dans le modèle. L'inclusion d'une constante est normale, à moins que vous ne soyez sûr que la valeur de série de la moyenne générale est 0. L'exclusion de la constante est recommandée lorsque la différenciation s'applique.

Fonctions de transfert dans des modèles ARIMA personnalisés

Figure 2-6
Boîte de dialogue Critères ARIMA, onglet Fonction de transfert

The screenshot shows the 'Modélisateur de séries chronologiques : Critères ARIMA' dialog box with the 'Fonction de transfert' tab selected. The 'Variables indépendantes' section contains a list with one item: 'Nombre total de sousc...'. The 'Ordres de fonction de transfert' section features a table for defining the transfer function structure.

| | Non saisonnier | Saisonnier |
|---------------------|----------------|------------|
| Numérateur | 0 | 0 |
| Dénominateur | 0 | 0 |
| Différencié d'ordre | 0 | 0 |

Below the table, the 'Périodicité courante' is set to 12, and the 'Décal' is set to 0. The 'Transformation' section has three radio buttons: 'Aucun' (selected), 'Racine carrée', and 'Log népérien'. At the bottom, there are buttons for 'Poursuivre', 'Annuler', and 'Aide'.

L'onglet Fonction de transfert (uniquement présent si les variables indépendantes sont spécifiées) vous permet de définir les fonctions de transfert pour une ou toutes les variables indépendantes spécifiées dans l'onglet Variables. Les fonctions de transfert vous permettent de spécifier la manière avec laquelle les valeurs passées des variables indépendantes (explicatives) sont utilisées pour prévoir les valeurs futures de la série dépendante.

Ordres de fonction de transfert. Entrez des valeurs pour les différents composants de la fonction de transfert dans les cellules correspondantes de la grille Structure. Toutes les valeurs doivent être des entiers non négatifs. Pour les composants numérateur et dénominateur, la valeur représente l'ordre maximum. Tous les ordres inférieurs positifs seront inclus dans le modèle. En outre, l'ordre 0 est toujours inclus pour les composants numérateur. Par exemple, si vous indiquez 2 pour le numérateur, le modèle inclut les ordres 2, 1 et 0. Si vous indiquez 3 pour le dénominateur, le modèle inclut les ordres 3, 2 et 1. Les cellules de la colonne Saisonnier ne sont activées que si une périodicité a été définie pour le fichier de travail (voir « Périodicité courante » ci-dessous).

- Numérateur.** L'ordre de numérateur de la fonction de transfert. Spécifie quelles valeurs précédentes de la série indépendante (explicative) sélectionnée sont utilisées pour prévoir les valeurs courantes de la série dépendante. Par exemple, un ordre de numérateur de 1 indique que la valeur d'une série indépendante Un point dans le temps dans le passé, ainsi que la valeur courante de la série indépendante, sont utilisées pour prévoir la valeur courante de chaque série dépendante.

- **Dénominateur.** L'ordre de dénominateur de la fonction de transfert. Spécifie comment les écarts de la moyenne de la série, pour les valeurs précédentes de la série indépendante (explicative) sélectionnée, sont utilisés pour prévoir les valeurs courantes de la série dépendante. Par exemple, un ordre de dénominateur de 1 indique que les écarts de la valeur moyenne d'une série indépendante Un point dans le temps dans le passé doivent être considérés lors de la prévision de la valeur courante de chaque série dépendante.
- **Différence :** Spécifie l'ordre de différenciation appliqué à la série indépendante (explicative) avant d'estimer les modèles. La différenciation est nécessaire lorsque les tendances sont présentes et est utilisée pour supprimer leur effet.

Ordres saisonniers. Les composants numérateur, dénominateur et de différenciation saisonniers tiennent le même rôle que leurs équivalents non saisonniers. Cependant, pour les ordres saisonniers, les valeurs courantes de la série sont affectées par les valeurs de série précédentes séparées par une ou plusieurs périodes saisonnières. Par exemple, pour des données mensuelles (période saisonnière de 12), un ordre saisonnier de 1 indique que la valeur de série en cours est affectée par les 12 périodes de la valeur de série précédant celle en cours. Un ordre saisonnier de 1, pour des données mensuelles, est alors le même que lorsqu'on spécifie un ordre non saisonnier de 12.

Périodicité courante. Indique la périodicité courante (le cas échéant) définie pour le fichier de travail. La périodicité courante est indiquée par une valeur entière, par exemple 12 pour la périodicité annuelle, chaque observation représentant un mois. La valeur *Aucune* est affichée si aucune périodicité n'a été définie. Les modèles saisonniers nécessitent une périodicité. Vous pouvez définir la périodicité dans la boîte de dialogue Définir des dates.

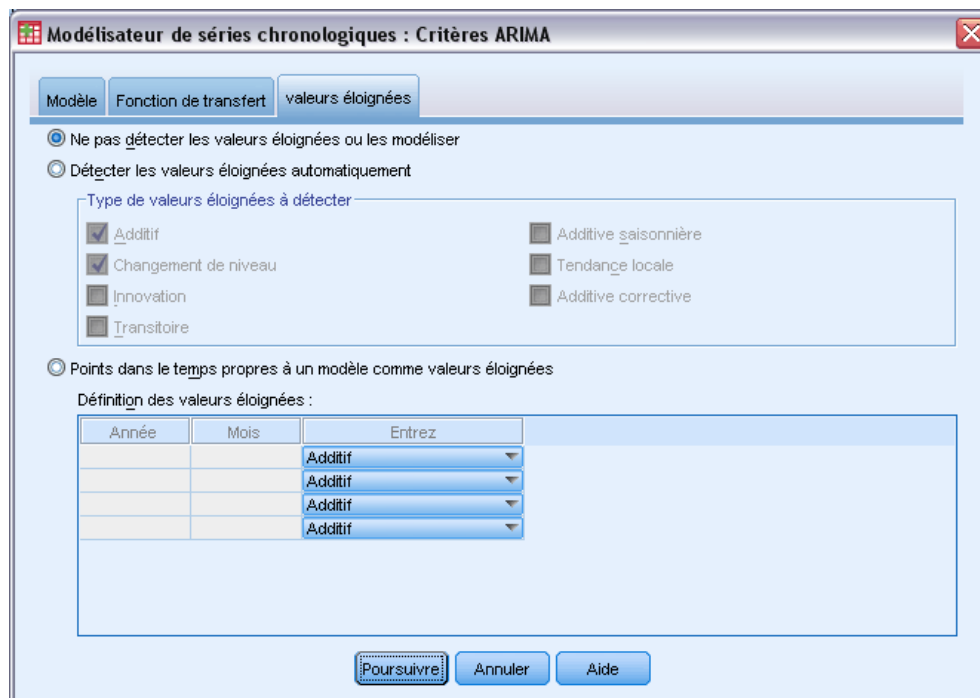
Délai. La définition d'un délai retarde l'influence de la variable indépendante selon le nombre d'intervalles spécifié. Par exemple, si le délai est défini sur 5, la valeur de la variable indépendante au moment t n'affecte pas les prévisions tant que les cinq périodes ne se sont pas écoulées ($t + 5$).

Transformation. La spécification d'une fonction de transfert, pour un ensemble de variables indépendantes, inclut également l'exécution d'une transformation optionnelle sur ces variables.

- **Aucune.** Aucune transformation n'est effectuée.
- **Racine carrée :** Transformation racine carrée.
- **Log népérien :** Transformation par log naturel.

Valeurs éloignées dans les modèles ARIMA personnalisés

Figure 2-7
Boîte de dialogue Critères ARIMA, onglet Valeurs éloignées



L'onglet Valeurs éloignées propose les options suivantes pour le traitement des valeurs éloignées (Pena, Tiao, et Tsay, 2001) : les détecter automatiquement, spécifier des points particuliers en tant que valeurs éloignées ou ne pas les détecter ou les modéliser.

Ne pas détecter ou modéliser les valeurs éloignées. Par défaut, les valeurs éloignées ne sont ni détectées ni modélisées. Sélectionnez cette option pour désactiver la détection ou la modélisation des valeurs éloignées.

Détecter les valeurs éloignées automatiquement. Sélectionnez cette option pour exécuter la détection automatique des valeurs éloignées et sélectionnez un ou plusieurs des types de valeurs éloignées suivants :

- Additif
- Changement de niveau
- Innovation
- Transitoire
- Additive saisonnière
- Tendance locale
- Additive correctrice

Pour plus d'informations, reportez-vous à la section Types de valeurs éloignées dans l'annexe B sur p. 99.

Points dans le temps spécifiques à un modèle comme valeurs éloignées. Sélectionnez cette option pour indiquer des points dans le temps particuliers comme valeurs éloignées. Utilisez une ligne distincte de la grille Définition des valeurs éloignées pour chaque valeur éloignée. Entrez des valeurs dans toutes les cellules d'une ligne donnée.

- **Type.** Le type de valeur éloignée. Les types pris en charge sont : additive (par défaut), changement de niveau, innovation, transitoire, additive saisonnière et tendance locale.

Remarque 1 : Si aucune spécification de date n'a été définie pour l'ensemble de données actif, la grille Définition des valeurs éloignées affiche la colonne unique *Observation*. Pour indiquer une valeur éloignée, entrez le numéro de la ligne (affiché dans l'Editeur de données) de l'observation correspondante.

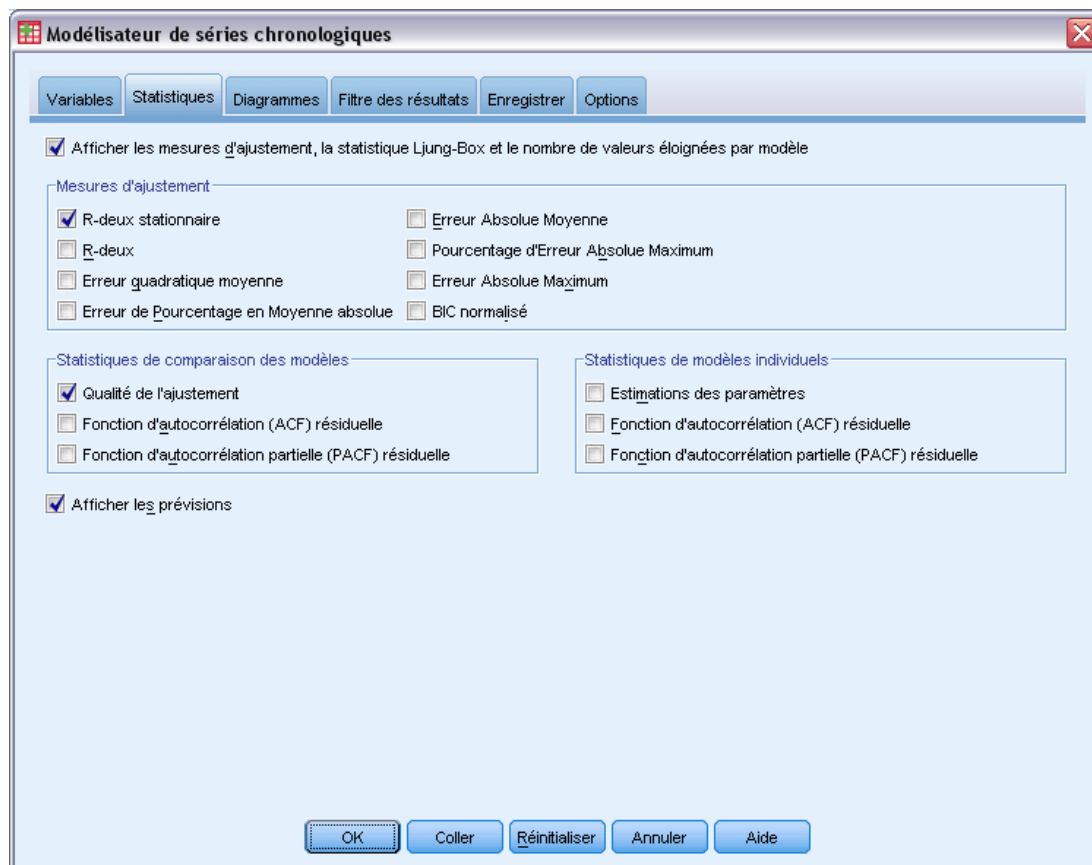
Remarque 2 : La colonne *Cycle* (si elle est présente) dans la grille Définition des valeurs éloignées se rapporte à la valeur de la variable *CYCLE_* de l'ensemble de données actif.

Résultats

Les résultats disponibles incluent des résultats pour les modèles individuels ainsi que des résultats calculés parmi tous les modèles. Les résultats pour les modèles individuels peuvent être limités à l'ensemble des modèles les plus ou les moins appropriés en fonction de critères définis par l'utilisateur.

Tableaux de statistiques et de prévision

Figure 2-8
Modélisateur de séries chronologiques, onglet Statistiques



L'onglet Statistiques fournit des options d'affichage pour les tableaux des résultats de la modélisation.

Afficher les mesures d'ajustement, la statistique Ljung-Box et le nombre de valeurs éloignées par modèle. Sélectionnez (cochez) cette option pour afficher un tableau contenant les mesures d'ajustement, la valeur Ljung-Box et le nombre de valeurs éloignées sélectionnés pour chaque modèle estimé.

Mesures d'ajustement. Vous pouvez sélectionner une ou plusieurs des valeurs suivantes pour les inclure dans le tableau contenant les mesures d'ajustement pour chaque modèle estimé :

- *R*-deux stationnaire
- *R*-deux
- Erreur quadratique moyenne
- Erreur de Pourcentage en Moyenne absolue
- Erreur Absolue Moyenne
- Pourcentage d'Erreur Absolue Maximum

- Erreur Absolue Maximum
- BIC normalisé

Pour plus d'informations, reportez-vous à la section [Mesures de la qualité d'ajustement dans l'annexe A sur p. 98](#).

Statistiques de comparaison des modèles. Ce groupe d'options contrôle l'affichage des tableaux contenant des statistiques calculées parmi tous les modèles estimés. Chaque option génère un tableau distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- **Qualité de l'ajustement.** La table des statistiques et centiles récapitulatifs pour le R -deux stationnaire, le R -deux, l'erreur quadratique moyenne, l'erreur de pourcentage Moyenne absolue, l'erreur Moyenne absolue, l'erreur de pourcentage Maximum absolue, l'erreur Maximum absolue et le critère d'information bayésien normalisé.
- **Fonction d'autocorrélation (ACF) résiduelle.** Tableau des statistiques et centiles récapitulatifs pour les autocorrélations résiduelles dans tous les modèles estimés.
- **Fonction d'autocorrélation partielle (PACF) résiduelle.** Tableau des statistiques et centiles récapitulatifs pour les autocorrélations partielles résiduelles dans tous les modèles estimés.

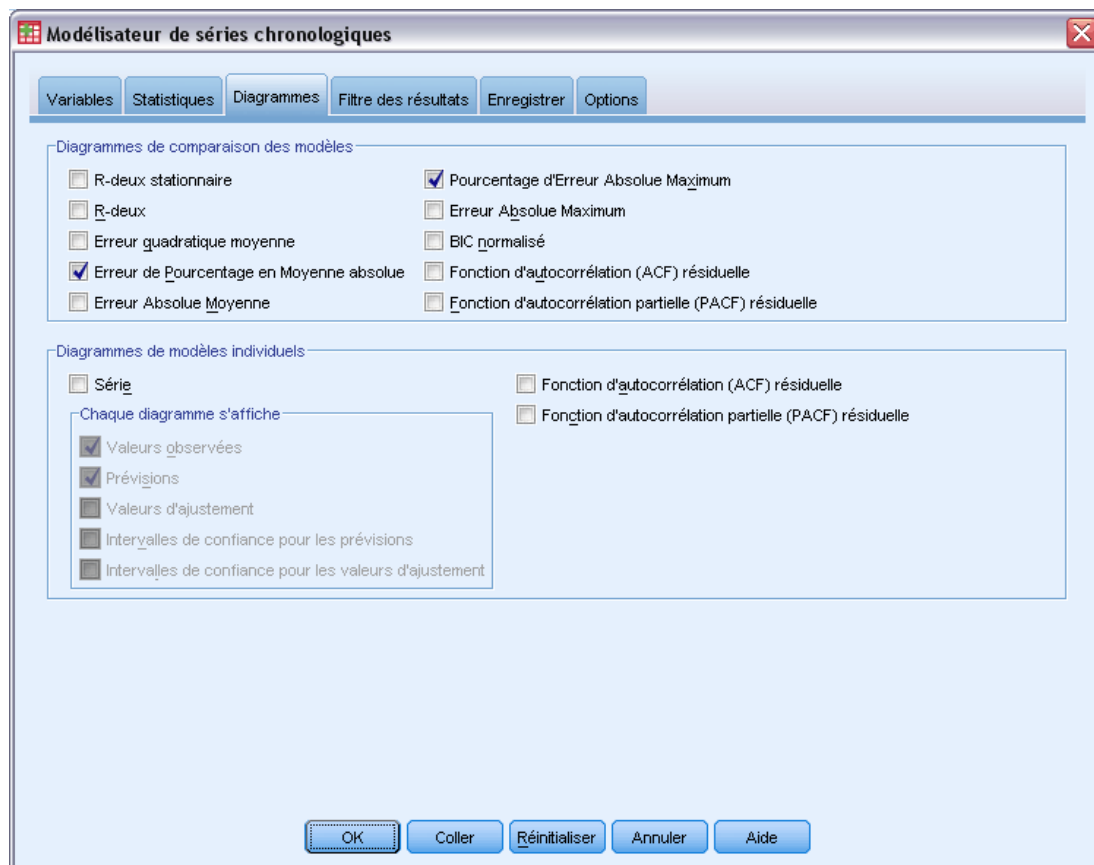
Statistiques de modèles individuels. Ce groupe d'options contrôle l'affichage des tableaux contenant des informations détaillées pour chaque modèle estimé. Chaque option génère un tableau distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- **Estimations des paramètres :** Affiche un tableau d'estimations des paramètres pour chaque modèle estimé. Des tableaux distincts sont affichés pour les modèles de lissage exponentiel et ARIMA. Si des valeurs éloignées existent, leurs estimations de paramètres sont également affichées dans un tableau distinct.
- **Fonction d'autocorrélation (ACF) résiduelle.** Affiche un tableau des autocorrélations résiduelles par décalage pour chaque modèle estimé. Le tableau inclut les intervalles de confiance pour les autocorrélations.
- **Fonction d'autocorrélation partielle (PACF) résiduelle.** Affiche un tableau des autocorrélations partielles résiduelles par décalage pour chaque modèle estimé. Le tableau inclut les intervalles de confiance pour les autocorrélations partielles.

Afficher les prévisions. Affiche un tableau des prévisions du modèle et les intervalles de confiance pour chaque modèle estimé. La période de prévision est définie dans l'onglet Options.

Diagrammes

Figure 2-9
Modélisateur de séries chronologiques, onglet Diagrammes



L'onglet Diagrammes fournit des options d'affichage pour les diagrammes des résultats de la modélisation.

Diagrammes de comparaison des modèles

Ce groupe d'options contrôle l'affichage des diagrammes contenant des statistiques calculées parmi tous les modèles estimés. Chaque option génère un diagramme distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- R-deux stationnaire
- R-deux
- Erreur quadratique moyenne
- Erreur de Pourcentage en Moyenne absolue
- Erreur Absolue Moyenne
- Pourcentage d'Erreur Absolue Maximum
- Erreur Absolue Maximum

- BIC normalisé
- Fonction d'autocorrélation (ACF) résiduelle
- Fonction d'autocorrélation partielle (PACF) résiduelle

Pour plus d'informations, reportez-vous à la section [Mesures de la qualité d'ajustement dans l'annexe A sur p. 98](#).

Diagrammes de modèles individuels

Série : Sélectionnez (cochez) cette option pour obtenir des diagrammes des prévisions pour chaque modèle estimé. Vous pouvez sélectionner un ou plusieurs des éléments suivants pour les insérer dans le diagramme :

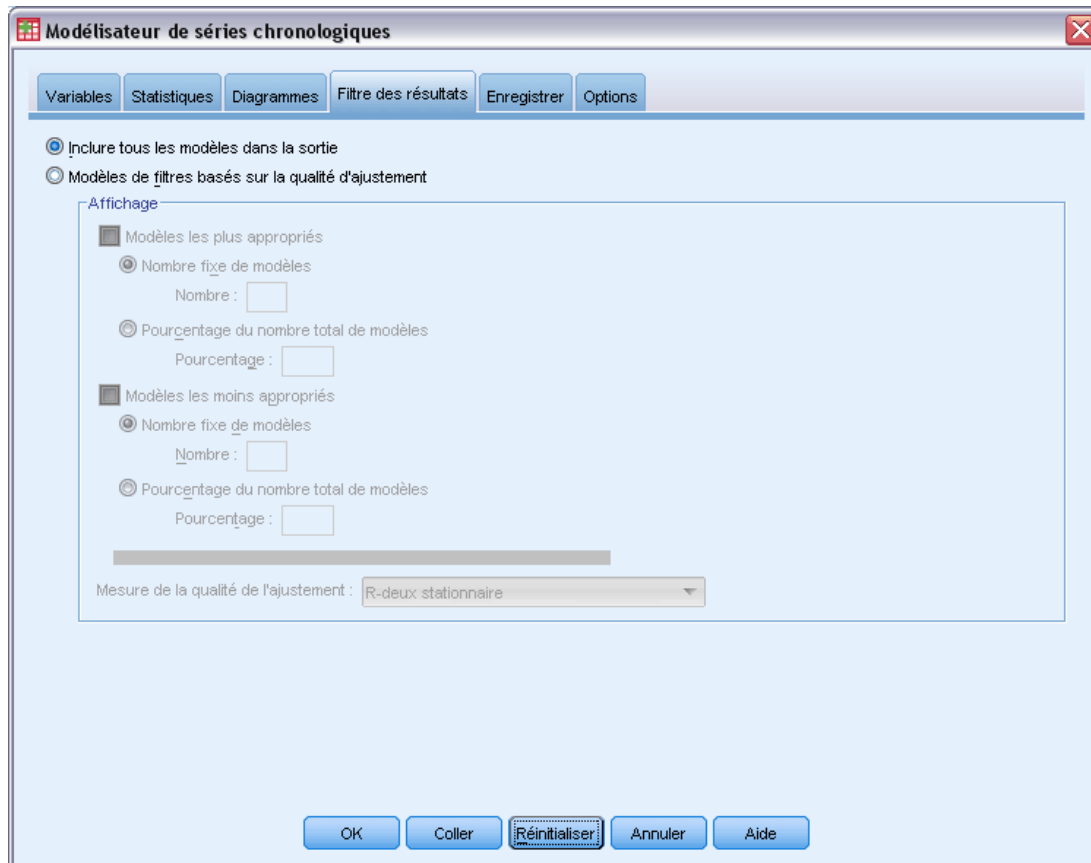
- **Valeurs observées.** Les valeurs observées des séries dépendantes.
- **Prévisions.** Les prévisions du modèle pour la période de prévision.
- **Valeurs d'ajustement.** Les prévisions du modèle pour la période d'estimation.
- **Intervalles de confiance pour les prévisions.** Les intervalles de confiance pour la période de prévision.
- **Intervalles de confiance pour les valeurs d'ajustement.** Les intervalles de confiance pour la période d'estimation.

Fonction d'autocorrélation (ACF) résiduelle. Affiche un diagramme des autocorrélations résiduelles pour chaque modèle estimé.

Fonction d'autocorrélation partielle (PACF) résiduelle. Affiche un diagramme des autocorrélations partielles résiduelles pour chaque modèle estimé.

Limitation des résultats aux modèles les plus ou les moins appropriés

Figure 2-10
Modélisateur de séries chronologiques, onglet Filtre des résultats



L'onglet Filtre des résultats fournit des options pour restreindre à la fois des résultats de tableau ou de diagramme à un sous-ensemble de modèles estimés. Vous pouvez choisir de limiter des résultats aux modèles les plus ou les moins appropriés en fonction de critères d'ajustement que vous fournissez. Par défaut, tous les modèles estimés sont inclus dans les résultats.

Modèles les plus appropriés. Sélectionnez (cochez) cette option pour inclure les modèles les plus appropriés dans les résultats. Sélectionnez une mesure de qualité d'ajustement et indiquez le nombre de modèles à inclure. Sélectionner cette option n'empêche pas de sélectionner également les modèles les moins appropriés. Dans ce cas, les résultats concernent aussi bien les modèles les moins appropriés que les modèles les plus appropriés.

- **Nombre fixe de modèles.** Indique que les résultats sont affichés pour les n modèles les plus appropriés. Si le nombre dépasse le nombre de modèles estimés, tous les modèles sont affichés.
- **Pourcentage du nombre total de modèles.** Indique que les résultats sont affichés pour les modèles possédant des valeurs de qualité d'ajustement comprises dans les n pourcentages supérieurs dans tous les modèles estimés.

Modèles les moins appropriés. Sélectionnez (cochez) cette option pour inclure les modèles les moins appropriés dans les résultats. Sélectionnez une mesure de qualité d'ajustement et indiquez le nombre de modèles à inclure. Sélectionner cette option n'empêche pas de sélectionner également les modèles les plus appropriés. Dans ce cas, les résultats sont constitués des modèles les plus appropriés ainsi que des modèles les moins appropriés.

- **Nombre fixe de modèles.** Indique que les résultats sont affichés pour les n modèles les moins appropriés. Si le nombre dépasse le nombre de modèles estimés, tous les modèles sont affichés.
- **Pourcentage du nombre total de modèles.** Indique que les résultats sont affichés pour les modèles possédant des valeurs de qualité d'ajustement comprises dans les n pourcentages inférieurs dans tous les modèles estimés.

Mesure de la qualité de l'ajustement. Sélectionnez la mesure de la qualité d'ajustement à utiliser pour filtrer les modèles. La valeur par défaut est R -deux stationnaire.

Enregistrement des prévisions et des spécifications de modèle

L'onglet Enregistrer vous permet d'enregistrer des prévisions de modèle en tant que nouvelles variables dans l'ensemble de données actif et d'enregistrer les spécifications de modèle dans un fichier externe au format XML.

Enregistrer les variables. Vous pouvez enregistrer des prévisions de modèle, des intervalles de confiance et des résidus en tant que nouvelles variables dans l'ensemble de données actif. Chaque série dépendante génère son propre ensemble de nouvelles variables et chaque nouvelle variable contient des valeurs pour les périodes de prévision et d'estimation. Les nouvelles observations sont ajoutées si la période de prévision s'étend au-delà de la longueur des séries de variables dépendantes. Choisissez d'enregistrer de nouvelles variables en cochant la case Enregistrer correspondant à chaque variable. Par défaut, aucune nouvelle variable n'est enregistrée.

- **Prévisions :** Les prévisions de modèle.
- **Limites de confiance inférieures.** Limites de confiance inférieures pour les prévisions.
- **Limites de confiance supérieures.** Limites de confiance supérieures pour les prévisions.
- **Résidus du bruit.** Les résidus de modèle. Lors de la transformation des variables dépendantes (par exemple Log népérien), ces derniers sont les résidus des séries transformées.
- **Préfixe de nom de variable** Indiquez les préfixes à utiliser pour des nouveaux noms de variable ou conservez les préfixes par défaut. Les noms de variable sont constitués du préfixe, du nom de la variable dépendante correspondante et d'un identificateur de modèle. Si nécessaire, le nom de variable est développé pour éviter les conflits de noms de variables. Le préfixe doit se conformer aux règles pour les noms de variable valides.

Exporter un fichier de modèle Les spécifications de modèle pour tous les modèles estimés sont exportées vers le fichier spécifié au format XML. Les modèles enregistrés peuvent être utilisés pour obtenir des prévisions mises à jour, en fonction de données plus actuelles, en utilisant la procédure [Appliquer des modèles de séries chronologiques](#).

- **Fichier XML.** Les spécifications de modèle sont enregistrées dans un fichier XML qui peut être utilisé avec les applications IBM SPSS.
- **Fichier PMML.** Les spécifications de modèle sont enregistrées dans un fichier XML compatible-PMML qui peut être utilisé avec les applications compatibles PMML, y compris les applications IBM SPSS.

Options

Figure 2-11
Modélisateur de séries chronologiques, onglet Options

Modélisateur de séries chronologiques

Variables Statistiques Diagrammes Filtre des résultats Enregistrer Options

Période de prévision

Première observation après la fin de la période d'estimation jusqu'à la dernière observation de l'ensemble de données actif
 Première observation après la fin de la période d'estimation jusqu'à une date spécifique

Date :

| Année | Mois |
|-------|------|
| 2004 | 3 |

Valeurs manquantes spécifiées

Traiter comme non valide
 Traiter comme valide

Largeurs des intervalles de confiance (%) : 95

Préfixe pour les identificateurs de modèle dans les résultats : Modèle

Nombre maximum de décalages affichés dans les résultats ACF et PACF : 24

OK Coller Réinitialiser Annuler Aide

L'onglet Options vous permet de définir la période de prévision, spécifier le traitement des valeurs manquantes, définir la largeur de l'intervalle de confiance, spécifier un préfixe personnalisé pour les identificateurs de modèle et définir le nombre de décalages affiché pour les autocorrélations.

Période de prévision. La période de prévision commence toujours par la première observation après la fin de la période d'estimation (l'ensemble d'observations utilisé pour déterminer le modèle) et continue jusqu'à la dernière observation de l'ensemble de données actif ou jusqu'à une date indiquée par l'utilisateur. Par défaut, la fin de la période d'estimation est la dernière observation dans l'ensemble de données actif, mais peut être modifiée via la boîte de dialogue Sélectionner des observations en sélectionnant Dans un intervalle de temps ou d'observations.

- **Première observation après la fin de la période d'estimation jusqu'à la dernière observation de l'ensemble de données actif.** Sélectionnez cette option lorsque la fin de la période d'estimation est antérieure à la dernière observation de l'ensemble de données actif et que vous souhaitez avoir des prévisions jusqu'à la dernière observation. Cette option est généralement utilisée pour produire des prévisions pour une période traitée, permettant une comparaison des prévisions du modèle avec un sous-ensemble des valeurs courantes.
- **Première observation après la fin de la période d'estimation jusqu'à une date spécifique.** Sélectionnez cette option pour indiquer explicitement la fin de la période de prévision. Cette option est généralement utilisée pour produire des prévisions après la fin de la série en cours. Entrez les valeurs dans toutes les cellules de la grille Date.

Si aucune spécification de date n'a été définie pour l'ensemble de données actif, la grille Date affiche la colonne unique *Observation*. Pour indiquer la fin de la période de prévision, entrez le numéro de la ligne (affiché dans l'Editeur de données) de l'observation correspondante.

La colonne *Cycle* (si elle est présente) dans la grille Date se rapporte à la valeur de la variable *CYCLE_* de l'ensemble de données actif.

Valeurs manquantes spécifiées. Ces options contrôlent le traitement des valeurs manquantes spécifiées par l'utilisateur.

- **Traiter comme non valide.** Les valeurs manquantes spécifiées par l'utilisateur sont traitées comme des valeurs manquantes par défaut.
- **Traiter comme valide.** Les valeurs manquantes spécifiées par l'utilisateur sont traitées comme des données valides.

Stratégie de valeur manquante. Les règles suivantes s'appliquent au traitement des valeurs manquantes (inclut les valeurs manquantes par défaut et les valeurs manquantes spécifiées par l'utilisateur traitées comme non valides) lors de la procédure de modélisation :

- Les observations avec des valeurs manquantes de variable dépendante qui se produisent au cours de la période d'estimation sont incluses dans le modèle. Le traitement spécifique de la valeur manquante dépend de la méthode d'estimation.
- Un avertissement est émis si une variable indépendante a des valeurs manquantes pour la période d'estimation. Pour le modélisateur expert, les modèles utilisant la variable indépendante sont estimés sans la variable. Pour ARIMA personnalisé, les modèles utilisant la variable indépendante ne sont pas estimés.
- Si une variable indépendante a des valeurs manquantes pour la période d'estimation, la procédure émet un avertissement et effectue les prévisions dans la limite du possible.

Largeur des intervalles de confiance (%). Les intervalles de confiance sont calculés pour les prévisions et les autocorrélations résiduelles du modèle. Vous pouvez indiquer toute valeur positive inférieure à 100. Par défaut, un intervalle de confiance de 95 % est utilisé.

Préfixe pour les identificateurs de modèle dans les résultats. Chaque variable dépendante indiquée dans l'onglet Variables génère un modèle estimé séparé. Les modèles sont différenciés par des noms uniques constitués d'un préfixe personnalisable et d'un suffixe entier. Vous pouvez entrer un préfixe ou conserver le préfixe par défaut du *modèle*.

Nombre maximum de décalages affichés dans les résultats ACF et PACF. Vous pouvez définir le nombre maximum de décalages affichés dans les tableaux et diagrammes d'autocorrélations et d'autocorrélations partielles.

Fonctionnalités supplémentaires de la commande TSMODEL

Vous pouvez personnaliser la modélisation de votre série chronologique en collant vos sélections dans une fenêtre de syntaxe, puis en modifiant la syntaxe de la commande `TSMODEL` qui en résulte. Le langage de syntaxe de commande vous permet de :

- Indiquer la période saisonnière des données (avec le mot de passe `SEASONLENGTH` de la sous-commande `AUXILIARY`). Cette opération remplace la périodicité courante (le cas échéant) pour l'ensemble de données actif.
- Indiquer les décalages non consécutifs pour ARIMA personnalisé et les composants de la fonction de transfert (avec les sous-commandes `ARIMA` et `TRANSFERFUNCTION`). Par exemple, vous pouvez spécifier un modèle ARIMA personnalisé avec des décalages autorégressifs des ordres 1, 3 et 6 ou une fonction de transfert avec les décalages de numérateur des ordres 2, 5 et 8.
- Fournir plus d'un ensemble de spécifications de modélisation (par exemple, la méthode de modélisation, les ordres ARIMA, les variables indépendantes, etc.) pour une exécution unique de la procédure Modélisateur de séries chronologiques (avec la sous-commande `MODEL`).

Reportez-vous à la *Référence de syntaxe de commande* pour une information complète concernant la syntaxe.

Appliquer des modèles de séries chronologiques

La procédure Appliquer des modèles de séries chronologiques charge des modèles de séries chronologiques existants à partir d'un fichier externe et les applique à l'ensemble de données actif. Vous pouvez utiliser cette procédure pour obtenir des prévisions pour des séries pour lesquelles de nouvelles données ou des données révisées sont disponibles, sans avoir à recréer les modèles. Les modèles sont générés à l'aide de la procédure [Time Series Modeler](#).

Exemple : Vous êtes gestionnaire de stock pour un distributeur mondial et êtes responsable de 5 000 produits. Vous utilisez le modélisateur expert pour créer des modèles qui calculent les prévisions des ventes sur les trois mois à venir pour chaque produit. Votre entrepôt de données est actualisé tous les mois avec les données de vente réelles que vous souhaitez utiliser pour produire des prévisions actualisées tous les mois. La procédure Appliquer des modèles de séries chronologiques vous permet d'accomplir ceci en utilisant des modèles originaux et de réestimer facilement les paramètres des modèles pour représenter les nouvelles données.

Statistiques. Mesures de la qualité de l'ajustement : R -deux stationnaire, R -deux (R^2), erreur quadratique moyenne (RMSE), Erreur Absolue Moyenne (MAE), Erreur de Pourcentage en Moyenne absolue (MAPE), Erreur Absolue Maximum (MaxAE), Pourcentage d'Erreur Absolue Maximum (MaxAPE), critère d'information bayésien normalisé (BIC). Résidus : fonction d'autocorrélation, fonction d'autocorrélation partielle, Q de Ljung-Box.

Diagrammes : Diagrammes récapitulatifs parmi tous les modèles : histogrammes de R -deux stationnaire, R -deux (R^2), erreur quadratique moyenne (RMSE), Erreur Absolue Moyenne (MAE), Erreur de Pourcentage en Moyenne absolue (MAPE), Erreur Absolue Maximum (MaxAE), Pourcentage d'Erreur Absolue Maximum (MaxAPE), critère d'information bayésien normalisé (BIC) ; diagrammes à surfaces des autocorrélations partielles et résiduelles. Résultats des modèles individuels : valeurs de prévision, valeurs d'ajustement, valeurs observées, limites de confiance supérieure et inférieure, autocorrélations partielles et résiduelles.

Remarques sur les données Appliquer des modèles de séries chronologiques

Données. Les variables (indépendantes et dépendantes) auxquelles les modèles seront appliqués doivent être numériques.

Hypothèses : Les modèles sont appliqués à des variables dans l'ensemble de données actif ayant les mêmes noms que les variables indiquées dans le modèle. Toutes les variables de ce type sont traitées en tant que séries chronologiques, ce qui veut dire que chaque observation représente un point dans le temps, avec des observations successives séparées par un intervalle de temps constant.

- **Prévisions.** Pour produire des prévisions à l'aide de modèles avec des variables indépendantes (explicatives), l'ensemble de données actif doit contenir des valeurs de ces variables pour toutes les observations de la période de prévision. Si les paramètres du modèle sont réestimés,

les variables indépendantes ne doivent contenir aucune valeur manquante dans la période d'estimation.

Définition des dates

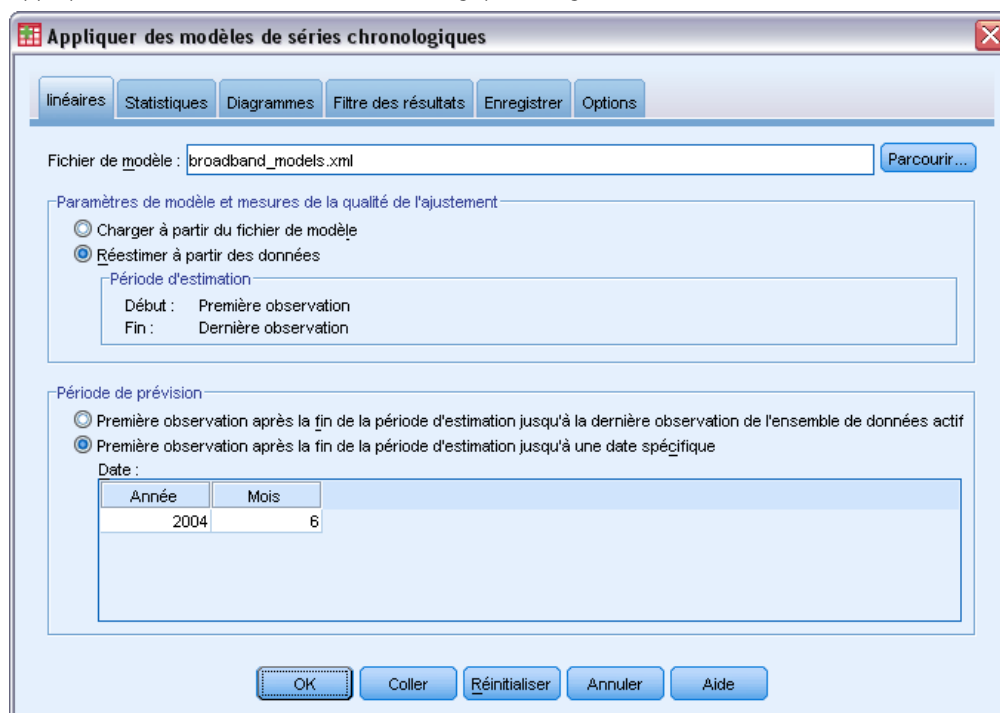
La procédure Appliquer des modèles de séries chronologiques requiert que la périodicité, le cas échéant, de l'ensemble de données actif corresponde à la périodicité des modèles à appliquer. Si vous calculez les prévisions en utilisant simplement le même ensemble de données (éventuellement avec des données nouvelles ou revues) que celui utilisé pour construire ce modèle, alors cette condition est remplie. S'il n'existe aucune périodicité pour l'ensemble de données actif, vous aurez la possibilité d'accéder à la boîte de dialogue Définir des dates pour en créer une. Si, cependant, les modèles ont été créés sans indiquer de périodicité, alors l'ensemble de données actif ne devrait pas en comporter non plus.

Pour appliquer des modèles

- A partir des menus, sélectionnez :
Analyse > Prévisions > Appliquer des modèles...

Figure 3-1

Appliquer des modèles de séries chronologiques, onglet Modèles



- Entrez les spécifications de fichier pour un modèle de fichier ou cliquez sur Parcourir et sélectionnez un fichier de modèle (les fichiers de modèle sont créés avec la procédure [Time Series Modeler](#)).

Sinon, vous pouvez :

- Réestimer les paramètres de modèle en utilisant les données de l'ensemble de données actif. Les prévisions sont créées en utilisant les paramètres réestimés.
- [Enregistrer les prévisions, les intervalles de confiance et les résidus du bruit.](#)
- [Enregistrer les modèles réestimés au format XML.](#)

Paramètres de modèle et mesures de la qualité de l'ajustement

Charger à partir du fichier de modèle. Les prévisions sont produites à l'aide des paramètres de modèle du fichier de modèle, sans réestimation de ces paramètres. [Mesures de la qualité de l'ajustement](#) affichées dans les résultats et utilisées pour filtrer les modèles (les plus appropriés ou les moins appropriés) sont extraites du fichier de modèle et reflètent les données utilisées pour le développement de chaque modèle (ou pour leur dernière actualisation). Grâce à cette option, les prévisions ne prennent pas en compte les données historiques, pour les variables dépendantes comme pour les variables indépendantes, dans l'ensemble de données actif. Vous devez choisir Réestimer à partir des données pour que les données historiques aient un impact sur les prévisions. En outre, les prévisions ne prennent pas en compte les valeurs des séries dépendantes dans la période de prévision, cependant elles prennent en compte celles des variables indépendantes. Si vous disposez de valeurs plus actuelles pour les séries dépendantes et que vous souhaitez les inclure dans les prévisions, vous devez réestimer la période d'estimation en l'ajustant pour y inclure ces valeurs.

Réestimer à partir des données. Les paramètres de modèle sont réestimés en utilisant les données de l'ensemble de données actif. La réestimation des paramètres de modèle n'a aucun effet sur la structure de modèle. Par exemple, un modèle ARIMA(1,0,1) reste tel quel, mais les paramètres autorégressifs et de moyenne mobile seront réestimés. La réestimation ne résulte pas dans la détection de nouvelles valeurs éloignées. Les valeurs éloignées, s'il y en a, sont toujours extraites du fichier de modèle.

- **Période d'estimation.** La période d'estimation définit l'ensemble des observations utilisées pour réestimer les paramètres du modèle. Par défaut la période d'estimation inclut toutes les observations dans l'ensemble de données actif. Pour définir la période d'estimation, sélectionnez Dans un intervalle de temps ou d'observations dans la boîte de dialogue Sélectionner des observations. En fonction des données disponibles, la période d'estimation utilisée par la procédure peut varier selon le modèle et donc être différente de la valeur affichée. Pour un modèle donné, la vraie période d'estimation est la période restante après l'élimination de toutes les valeurs manquantes contiguës, à partir de la variable dépendante du modèle, qui se produit au début ou à la fin de la période d'estimation spécifiée.

Période de prévision

La période de prévision pour chaque modèle commence toujours par la première observation après la fin de la période d'estimation et va jusqu'à la dernière observation de l'ensemble de données actif ou jusqu'à une date indiquée par l'utilisateur. Si les paramètres ne sont pas réestimés (ce qui se produit par défaut), alors la période d'estimation pour chaque modèle est l'ensemble des observations utilisées lors du développement du modèle (ou lors de sa dernière actualisation).

- **Première observation après la fin de la période d'estimation jusqu'à la dernière observation de l'ensemble de données actif.** Sélectionnez cette option lorsque la fin de la période d'estimation est antérieure à la dernière observation de l'ensemble de données actif et que vous souhaitez avoir des prévisions jusqu'à la dernière observation.
- **Première observation après la fin de la période d'estimation jusqu'à une date spécifique.** Sélectionnez cette option pour indiquer explicitement la fin de la période de prévision. Entrez les valeurs dans toutes les cellules de la grille Date.

Si aucune spécification de date n'a été définie pour l'ensemble de données actif, la grille Date affiche la colonne unique *Observation*. Pour indiquer la fin de la période de prévision, entrez le numéro de la ligne (affiché dans l'Editeur de données) de l'observation correspondante.

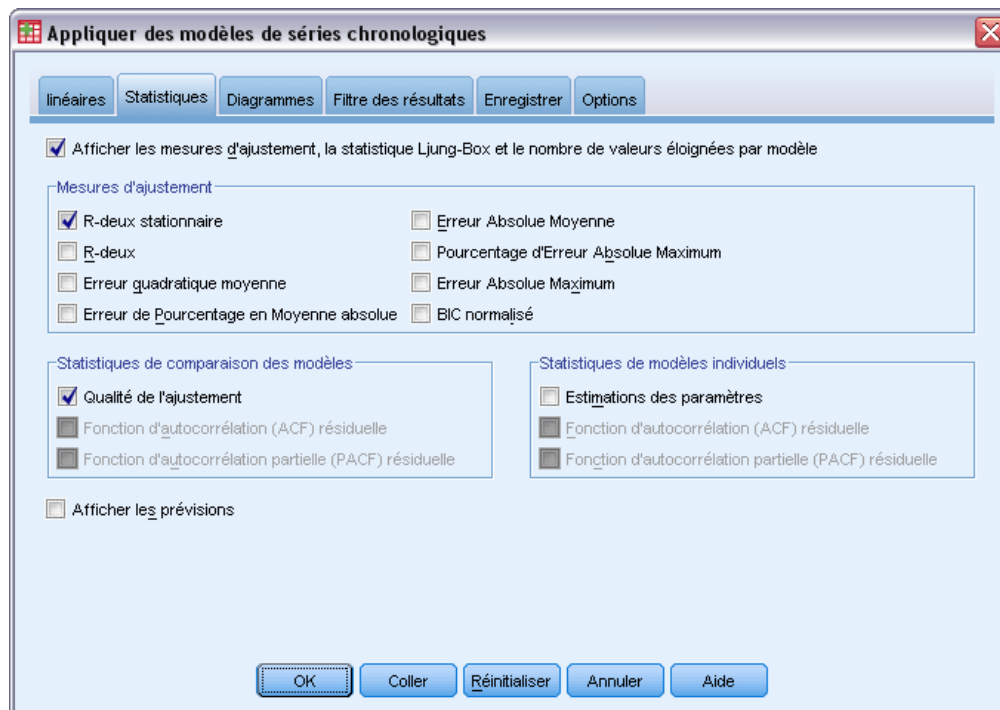
La colonne *Cycle* (si elle est présente) dans la grille Date se rapporte à la valeur de la variable *CYCLE_* de l'ensemble de données actif.

Résultats

Les résultats disponibles incluent des résultats pour les modèles individuels ainsi que des résultats parmi tous les modèles. Les résultats pour les modèles individuels peuvent être limités à l'ensemble des modèles les plus ou les moins appropriés en fonction de critères définis par l'utilisateur.

Tableaux de statistiques et de prévision

Figure 3-2
Appliquer des modèles de séries chronologiques, onglet Statistiques



L'onglet Statistiques fournit des options pour afficher des tableaux de statistiques d'ajustement du modèle, de paramètres de modèle, de fonctions d'autocorrélation et de prévisions. A moins que les paramètres de modèle ne soient réestimés (Réestimer à partir des données dans l'onglet Modèles), les valeurs affichées des mesures d'ajustement, les valeurs Ljung-Box et des paramètres de modèle sont celles du fichier de modèle et reflètent les données utilisées lors du développement de chaque modèle (ou lors de sa dernière actualisation). Les informations de valeurs éloignées sont toujours prises dans le fichier de modèle.

Afficher les mesures d'ajustement, la statistique Ljung-Box et le nombre de valeurs éloignées par modèle. Sélectionnez (cochez) cette option pour afficher un tableau contenant les mesures d'ajustement, la valeur Ljung-Box et le nombre de valeurs éloignées sélectionnés pour chaque modèle.

Mesures d'ajustement. Vous pouvez sélectionner une ou plusieurs des valeurs suivantes pour les inclure dans le tableau contenant les mesures d'ajustement pour chaque modèle :

- *R*-deux stationnaire
- *R*-deux
- Erreur quadratique moyenne
- Erreur de Pourcentage en Moyenne absolue
- Erreur Absolue Moyenne
- Pourcentage d'Erreur Absolue Maximum
- Erreur Absolue Maximum
- BIC normalisé

Pour plus d'informations, reportez-vous à la section Mesures de la qualité d'ajustement dans l'annexe A sur p. 98.

Statistiques de comparaison des modèles. Ce groupe d'options contrôle l'affichage des tableaux contenant des statistiques servant à tous les modèles. Chaque option génère un tableau distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- **Qualité de l'ajustement.** La table des statistiques et centiles récapitulatifs pour le *R*-deux stationnaire, le *R*-deux, l'erreur quadratique moyenne, l'erreur de pourcentage Moyenne absolue, l'erreur Moyenne absolue, l'erreur de pourcentage Maximum absolue, l'erreur Maximum absolue et le critère d'information bayésien normalisé.
- **Fonction d'autocorrélation (ACF) résiduelle.** Tableau des statistiques et centiles récapitulatifs pour les autocorrélations résiduelles dans tous les modèles estimés. Ce tableau est uniquement disponible si les paramètres de modèle sont réestimés (Réestimer à partir des données dans l'onglet Modèles).
- **Fonction d'autocorrélation partielle (PACF) résiduelle.** Tableau des statistiques et centiles récapitulatifs pour les autocorrélations partielles résiduelles dans tous les modèles estimés. Ce tableau est uniquement disponible si les paramètres de modèle sont réestimés (Réestimer à partir des données dans l'onglet Modèles).

Statistiques de modèles individuels. Ce groupe d'options contrôle l'affichage des tableaux contenant des informations détaillées pour chaque modèle. Chaque option génère un tableau distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

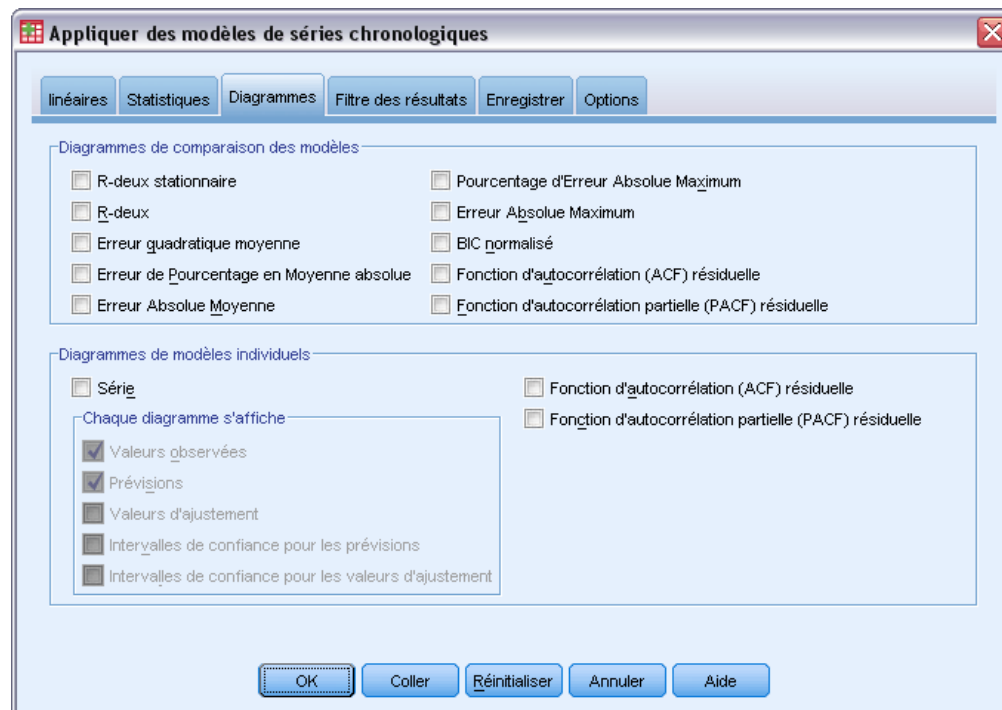
- **Estimations des paramètres :** Affiche un tableau d'estimation des paramètres pour chaque modèle. Des tableaux distincts sont affichés pour les modèles de lissage exponentiel et ARIMA. Si des valeurs éloignées existent, leurs estimations de paramètres sont également affichées dans un tableau distinct.
- **Fonction d'autocorrélation (ACF) résiduelle.** Affiche un tableau des autocorrélations résiduelles par décalage pour chaque modèle estimé. Le tableau inclut les intervalles de confiance pour les autocorrélations. Ce tableau est uniquement disponible si les paramètres de modèle sont réestimés (Réestimer à partir des données dans l'onglet Modèles).
- **Fonction d'autocorrélation partielle (PACF) résiduelle.** Affiche un tableau des autocorrélations partielles résiduelles par décalage pour chaque modèle estimé. Le tableau inclut les intervalles de confiance pour les autocorrélations partielles. Ce tableau est uniquement disponible si les paramètres de modèle sont réestimés (Réestimer à partir des données dans l'onglet Modèles).

Afficher les prévisions. Affiche un tableau des prévisions du modèle et les intervalles de confiance pour chaque modèle.

Diagrammes

Figure 3-3

Appliquer des modèles de séries chronologiques, onglet Diagrammes



L'onglet Diagrammes fournit des options pour l'affichage de diagrammes de statistiques d'ajustement de modèle, de fonctions d'autocorrélation et de valeurs de séries, y compris les prévisions.

Diagrammes de comparaison des modèles

Ce groupe d'options contrôle l'affichage des diagrammes contenant des statistiques servant à tous les modèles. A moins que les paramètres de modèle ne soient réestimés (Réestimer à partir des données dans l'onglet Modèles), les valeurs affichées sont celles du fichier de modèle et reflètent les données utilisées lors du développement de chaque modèle (ou lors de sa dernière actualisation). En outre, les diagrammes d'autocorrélation sont uniquement disponibles si les paramètres de modèle sont réestimés. Chaque option génère un diagramme distinct. Vous pouvez sélectionner une ou plusieurs des options suivantes :

- *R*-deux stationnaire
- *R*-deux
- Erreur quadratique moyenne
- Erreur de Pourcentage en Moyenne absolue
- Erreur Absolue Moyenne
- Pourcentage d'Erreur Absolue Maximum
- Erreur Absolue Maximum
- BIC normalisé
- Fonction d'autocorrélation (ACF) résiduelle
- Fonction d'autocorrélation partielle (PACF) résiduelle

Pour plus d'informations, reportez-vous à la section Mesures de la qualité d'ajustement dans l'annexe A sur p. 98.

Diagrammes de modèles individuels

Série : Sélectionnez (cochez) cette option pour obtenir des diagrammes des prévisions pour chaque modèle. Les valeurs observées, les valeurs d'ajustement, les intervalles de confiances des valeurs d'ajustement et les autocorrélations sont uniquement disponibles si les paramètres du modèle sont réestimés (Réestimer à partir des données dans l'onglet Modèles). Vous pouvez sélectionner un ou plusieurs des éléments suivants pour les insérer dans le diagramme :

- **Valeurs observées.** Les valeurs observées des séries dépendantes.
- **Prévisions.** Les prévisions du modèle pour la période de prévision.
- **Valeurs d'ajustement.** Les prévisions du modèle pour la période d'estimation.
- **Intervalles de confiance pour les prévisions.** Les intervalles de confiance pour la période de prévision.
- **Intervalles de confiance pour les valeurs d'ajustement.** Les intervalles de confiance pour la période d'estimation.

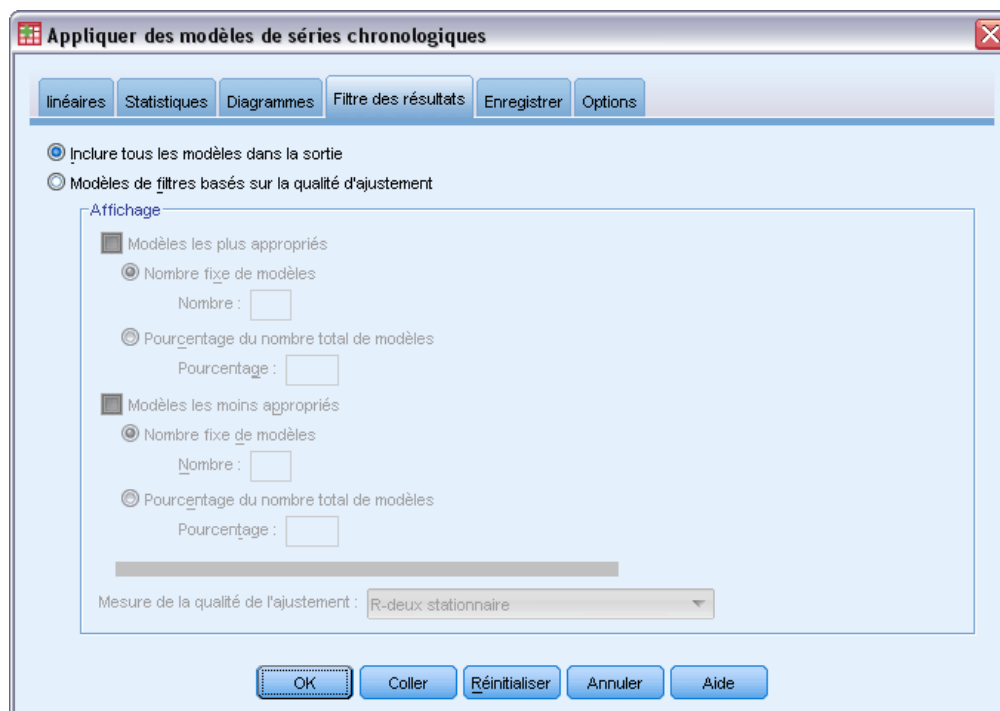
Fonction d'autocorrélation (ACF) résiduelle. Affiche un diagramme des autocorrélations résiduelles pour chaque modèle estimé.

Fonction d'autocorrélation partielle (PACF) résiduelle. Affiche un diagramme des autocorrélations partielles résiduelles pour chaque modèle estimé.

Limitation des résultats aux modèles les plus ou les moins appropriés

Figure 3-4

Appliquer des modèles de séries chronologiques, onglet Filtre des résultats



L'onglet Filtre des résultats fournit des options pour restreindre à la fois des résultats de tableau ou de diagramme à un sous-ensemble de modèles. Vous pouvez choisir de limiter des résultats aux modèles les plus ou les moins appropriés en fonction de critères d'ajustement que vous fournissez. Par défaut, tous les modèles sont inclus dans les résultats. A moins que les paramètres de modèle ne soient réestimés (Réestimer à partir des données dans l'onglet Modèles), les valeurs des mesures d'ajustement utilisées pour filtrer les modèles sont celles du fichier de modèle et reflètent les données utilisées lors du développement de chaque modèle (ou lors de sa dernière actualisation).

Modèles les plus appropriés. Sélectionnez (cochez) cette option pour inclure les modèles les plus appropriés dans les résultats. Sélectionnez une mesure de qualité d'ajustement et indiquez le nombre de modèles à inclure. Sélectionner cette option n'empêche pas de sélectionner également les modèles les moins appropriés. Dans ce cas, les résultats concernent aussi bien les modèles les moins appropriés que les modèles les plus appropriés.

- **Nombre fixe de modèles.** Indique que les résultats sont affichés pour les n modèles les plus appropriés. Si le nombre dépasse le nombre total de modèles, tous les modèles sont affichés.
- **Pourcentage du nombre total de modèles.** Indique que les résultats sont affichés pour les modèles avec des valeurs de qualité d'ajustement dans les n pourcentages supérieurs dans tous les modèles.

Modèles les moins appropriés. Sélectionnez (cochez) cette option pour inclure les modèles les moins appropriés dans les résultats. Sélectionnez une mesure de qualité d'ajustement et indiquez le nombre de modèles à inclure. Sélectionner cette option n'empêche pas de sélectionner également les modèles les plus appropriés. Dans ce cas, les résultats sont constitués des modèles les plus appropriés ainsi que des modèles les moins appropriés.

- **Nombre fixe de modèles.** Indique que les résultats sont affichés pour les n modèles les moins appropriés. Si le nombre dépasse le nombre total de modèles, tous les modèles sont affichés.
- **Pourcentage du nombre total de modèles.** Indique que les résultats sont affichés pour les modèles avec des valeurs de qualité d'ajustement dans les n pourcentages inférieurs dans tous les modèles.

Mesure de la qualité de l'ajustement. Sélectionnez la mesure de la qualité d'ajustement à utiliser pour filtrer les modèles. La valeur par défaut est R -deux stationnaire.

Enregistrement des prévisions et des spécifications de modèle

L'onglet Enregistrer vous permet d'enregistrer des prévisions de modèle en tant que nouvelles variables dans l'ensemble de données actif et d'enregistrer les spécifications de modèle dans un fichier externe au format XML.

Enregistrer les variables. Vous pouvez enregistrer des prévisions de modèle, des intervalles de confiance et des résidus en tant que nouvelles variables dans l'ensemble de données actif. Chaque modèle génère son propre ensemble de nouvelles variables. Les nouvelles observations sont ajoutées si la période de prévision s'étend au-delà de la longueur des séries de variables dépendantes associées au modèle. A moins que les paramètres de modèle soient réestimés (Réestimer à partir des données dans l'onglet Modèles), les prévisions et les limites de confiance sont uniquement créées pour la période de prévision. Choisissez d'enregistrer de nouvelles variables en cochant la case Enregistrer correspondant à chaque variable. Par défaut, aucune nouvelle variable n'est enregistrée.

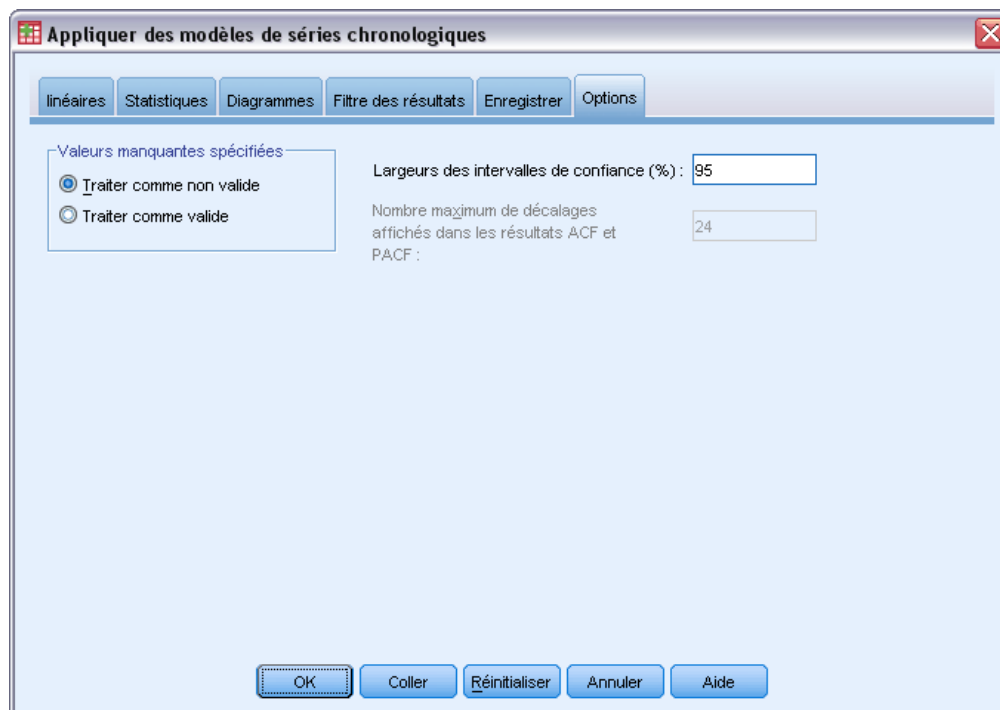
- **Prévisions :** Les prévisions de modèle.
- **Limites de confiance inférieures.** Limites de confiance inférieures pour les prévisions.
- **Limites de confiance supérieures.** Limites de confiance supérieures pour les prévisions.
- **Résidus du bruit.** Les résidus de modèle. Lors de la transformation des variables dépendantes (par exemple Log népérien), ces derniers sont les résidus des séries transformées. Ce choix est uniquement disponible si les paramètres de modèle sont réestimés (Réestimer à partir des données dans l'onglet Modèles).
- **Préfixe de nom de variable** Indiquez les préfixes à utiliser pour des nouveaux noms de variables ou conservez les préfixes par défaut. Les noms de variable sont constitués du préfixe, du nom de la variable dépendante correspondante et d'un identificateur de modèle. Si nécessaire, le nom de variable est développé pour éviter les conflits de noms de variables. Le préfixe doit se conformer aux règles pour les noms de variable valides.

Exporter un fichier de modèle Les spécifications de modèle, contenant des paramètres réestimés et des statistiques d'ajustement sont exportés vers le fichier indiqué au format XML. Cette option est uniquement disponible si les paramètres de modèle sont réestimés (Réestimer à partir des données dans l'onglet Modèles).

- **Fichier XML.** Les spécifications de modèle sont enregistrées dans un fichier XML qui peut être utilisé avec les applications IBM SPSS.
- **Fichier PMML.** Les spécifications de modèle sont enregistrées dans un fichier XML compatible-PMML qui peut être utilisé avec les applications compatibles PMML, y compris les applications IBM SPSS.

Options

Figure 3-5
Appliquer des modèles de séries chronologiques, onglet Options



L'onglet Options vous permet de spécifier le traitement des valeurs manquantes, définir la largeur de l'intervalle de confiance et définir le nombre de décalage affiché pour les autocorrélations.

Valeurs manquantes spécifiées. Ces options contrôlent le traitement des valeurs manquantes spécifiées par l'utilisateur.

- **Traiter comme non valide.** Les valeurs manquantes spécifiées par l'utilisateur sont traitées comme des valeurs manquantes par défaut.
- **Traiter comme valide.** Les valeurs manquantes spécifiées par l'utilisateur sont traitées comme des données valides.

Stratégie de valeur manquante. Les règles suivantes s'appliquent au traitement des valeurs manquantes (inclut les valeurs manquantes par défaut et les valeurs manquantes spécifiées par l'utilisateur traitées comme non valides) :

- Les observations avec des valeurs manquantes de variable dépendante qui se produisent au cours de la période d'estimation sont incluses dans le modèle. Le traitement spécifique de la valeur manquante dépend de la méthode d'estimation.
- Pour les modèles ARIMA, un avertissement est émis si la variable explicative a une valeur manquante pour la période d'estimation. Aucun modèle utilisant cette variable explicative ne sera réestimé.
- Si une variable indépendante a des valeurs manquantes pour la période d'estimation, la procédure émet un avertissement et effectue les prévisions dans la limite du possible.

Largeur des intervalles de confiance (%). Les intervalles de confiance sont calculés pour les prévisions et les autocorrélations résiduelles du modèle. Vous pouvez indiquer toute valeur positive inférieure à 100. Par défaut, un intervalle de confiance de 95 % est utilisé.

Nombre maximum de décalages affichés dans les résultats ACF et PACF. Vous pouvez définir le nombre maximum de décalages affichés dans les tableaux et diagrammes d'autocorrélations et d'autocorrélations partielles. Cette option est uniquement disponible si les paramètres de modèle sont réestimés (Réestimer à partir des données dans l'onglet Modèles).

Fonctionnalités supplémentaires de la commande TSAPPLY

Les fonctionnalités supplémentaires sont disponibles si vous collez vos sélections dans une fenêtre de syntaxe et modifiez la syntaxe de la commande `TSAPPLY`. Le langage de syntaxe de commande vous permet de :

- Spécifier que seul un sous-ensemble des modèles dans un fichier de modèle doit s'appliquer à l'ensemble de données actif (avec les mots-clés `DROP` et `KEEP` dans la sous-commande `MODEL`).
- Appliquer des modèles de deux ou plusieurs fichiers de modèles à vos données (avec la sous-commande `MODEL`). Par exemple, un fichier de modèle peut contenir des modèles pour des séries représentant des ventes de référence et un autre des modèles pour des séries représentant les revenus.

Reportez-vous à la *Référence de syntaxe de commande* pour une information complète concernant la syntaxe.

Désaisonnalisation

La procédure Désaisonnalisation décompose une série en une composante saisonnière, une combinaison d'une composante de tendance et de cycle, et une composante d'erreur. Cette procédure représente l'application de la méthode rapport/moyenne mobile (Census Method I).

Exemple : Un scientifique souhaite analyser les mesures mensuelles du niveau d'ozone dans une station météorologique donnée. Son objectif est de savoir si les données affichent une tendance. Afin de découvrir une tendance effective, il doit d'abord représenter la variation des mesures due aux effets saisonniers. Ce scientifique peut utiliser la procédure Désaisonnalisation pour supprimer toute variation saisonnière systématique. La tendance peut ensuite être analysée sur une série désaisonnalisée.

Statistiques. Ensemble de composantes saisonnières.

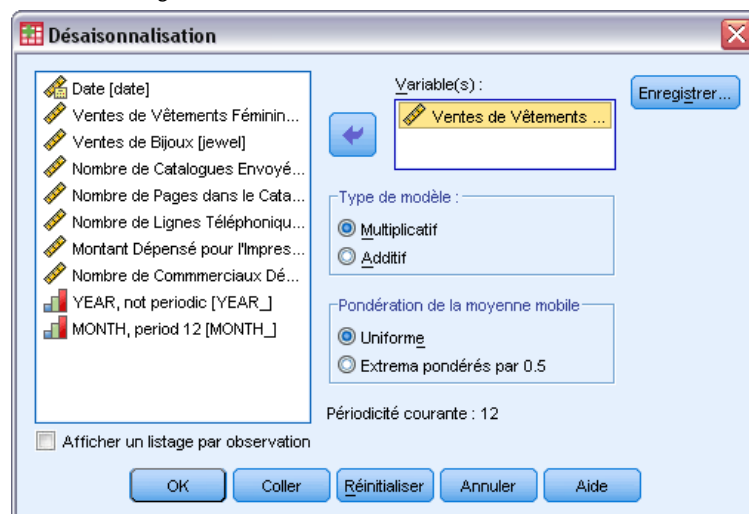
Données : Les variables doivent être numériques.

Hypothèses : Les variables ne doivent contenir aucune donnée manquante intégrée. Au moins une composante de date périodique doit être définie.

Estimer des composantes saisonnières

- A partir des menus, sélectionnez :
Analyse > Prévisions > Désaisonnalisation...

Figure 4-1
Boîte de dialogue Désaisonnalisation



- Sélectionnez des variables dans la liste disponible, puis déplacez-les vers la liste des variables. Cette liste contient uniquement des variables numériques.

Type de modèle. La procédure Désaisonnalisation propose deux approches différentes pour modéliser les composantes saisonnières : l'approche multiplicative ou l'approche additive.

- **Multiplicatif.** Facteur par lequel les séries désaisonnalisées sont multipliées pour fournir les séries d'origine. Ainsi, les composantes saisonnières qui sont proportionnelles au niveau global des séries. Les observations sans variation saisonnière ont une composante saisonnière de 1.
- **Additif.** Les ajustements saisonniers sont ajoutés aux séries désaisonnalisées pour obtenir les valeurs observées. Cet ajustement permet de filtrer les effets saisonniers d'une série dans le but de mettre en évidence d'autres caractéristiques pouvant être masquées par la composante saisonnière. Les composantes saisonnières qui ne dépendent pas du niveau global de la série. Les observations sans variation saisonnière ont une composante saisonnière de 0.

Pondération de la moyenne mobile. Les options du groupe Pondération de la moyenne mobile vous permettent de spécifier le mode de traitement de la série lors du calcul des moyennes mobiles. Ces options sont uniquement disponibles si la périodicité de la série est régulière. Si la périodicité est impaire, tous les points sont pondérés de façon égale.

- **Egal pour tous les points.** Les moyennes mobiles sont calculées avec un intervalle égal à la périodicité et avec tous les points pondérés de façon égale. Cette méthode est toujours utilisée si la périodicité est impaire.
- **Extrema pondérés par 0,5.** Les moyennes mobiles des séries dont la périodicité est paire sont calculées sur un intervalle égal à la périodicité plus un, les extrema de l'intervalle étant pondérés par 0,5.

Sinon, vous pouvez :

- Cliquez sur Enregistrer pour indiquer le mode d'enregistrement des nouvelles variables.

Enregistrer la désaisonnalisation

Figure 4-2
Boîte de dialogue Désaisonnalisation - Enregistrer



Créer des variables. Vous permet de choisir le mode de traitement des nouvelles variables.

- **Ajouter au fichier.** Les nouvelles séries créées par la procédure Désaisonnalisation sont enregistrées comme variables régulières dans votre ensemble de données actif. Le nom des variables est formé d'un préfixe de trois lettres, d'un trait de soulignement et d'un nombre.
- **Remplacer les anciennes.** Les nouvelles séries créées par la procédure Désaisonnalisation sont enregistrées comme variables temporaires dans votre ensemble de données actif. Simultanément, les variables temporaires existantes créées par les procédures Prévission

sont ignorées. Le nom des variables est formé d'un préfixe à trois lettres, d'un signe dièse (#) et d'un nombre.

- **Ne pas créer.** Les nouvelles séries ne sont pas ajoutées au fichier de travail.

Noms des nouvelles variables

La procédure Désaisonnalisation crée quatre nouvelles variables (séries) avec les préfixes à trois lettres suivants, pour chaque série spécifiée :

SAF. *Composantes de désaisonnalisation.* Ces valeurs indiquent l'effet de chaque période sur chaque niveau de la série.

SAS. *Séries désaisonnalisées.* Il s'agit des valeurs obtenues après la suppression de la variation saisonnière d'une série.

STC. *Composantes tendance-cycle lissées.* Ces valeurs indiquent la tendance et le comportement cyclique présents dans la série.

ERR. *Valeurs de résidu ou d'« erreur ».* Les valeurs restantes après que les composantes saisonnières, de tendance et de cycle ont été supprimées de la série.

Fonctionnalités supplémentaires de la commande SEASON

Le langage de syntaxe de commande vous permet aussi de :

- Indiquez une périodicité dans la commande `SEASON` au lieu de sélectionner l'une des solutions offertes par la procédure Définir des dates.

Pour obtenir des renseignements complets sur la syntaxe, reportez-vous au manuel *Command Syntax Reference*.

Analyse spectrale

La procédure Analyse spectrale permet d'identifier le comportement périodique des séries chronologiques. Au lieu d'analyser la variation d'un point à un autre, elle permet d'analyser la variation de la série dans son ensemble en tant que composantes périodiques des fréquences différentes. Les séries de lissage disposent de composantes périodiques plus fortes à basse fréquence ; la variation aléatoire (« white noise ») répartit la force de ces composantes sur toutes les autres fréquences.

Les séries incluant des données manquantes ne peuvent pas être analysées avec cette procédure.

Exemple : La vitesse à laquelle les nouvelles maisons sont construites représente un baromètre important de l'état de l'économie. Les données relatives aux mises en chantier affichent généralement une composante saisonnière forte. Mais ces données contiennent-elles des cycles plus longs que les analystes doivent connaître lors de l'évaluation des chiffres en cours ?

Statistiques : Transformations sinus et cosinus, valeur du périodogramme et estimation de densité spectrale pour chaque composante de fréquence ou de période. Lorsque l'analyse bivariée est sélectionnée : parties réelles et imaginaires du périodogramme croisé, densité cospectrale, spectre en quadrature, gain, cohérence carrée et spectre de phase pour chaque composante de fréquence ou de période.

Diagrammes : Pour les analyses univariées et bivariées : périodogramme et densité spectrale. Pour les analyses bivariées : cohérence carrée, spectre en quadrature, amplitude croisée, densité cospectrale, spectre de phase et gain.

Données : Les variables doivent être numériques.

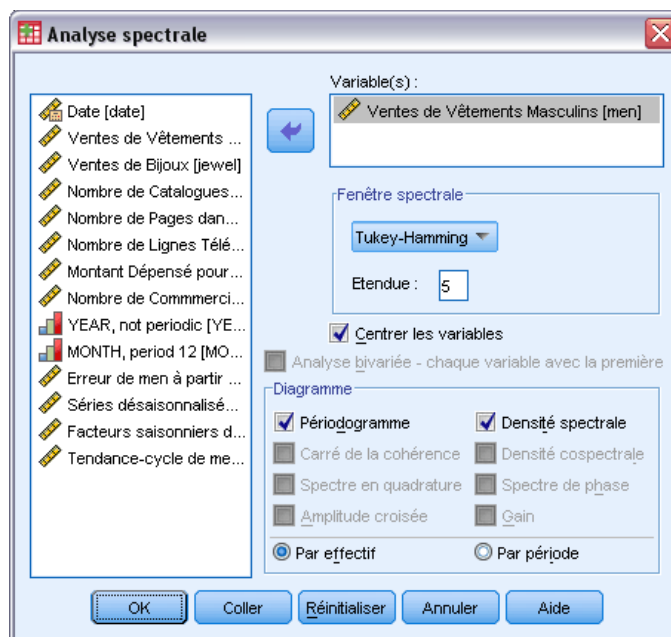
Hypothèses : Les variables ne doivent contenir aucune donnée manquante intégrée. La série chronologique à analyser doit être stationnaire et toute moyenne différente de zéro doit en être soustraite.

- **Stationnaire.** Condition qui doit être satisfaite par les séries chronologiques auxquelles vous ajustez un modèle ARIMA. Les séries MA pures sont stationnaires, mais les séries AR et ARMA ne le sont pas toujours. Une série stationnaire a une moyenne et une variance constantes dans le temps.

Obtention d'une analyse spectrale

- ▶ A partir des menus, sélectionnez :
Analyse > Séries chronologiques > Analyse spectrale...

Figure 5-1
Boîte de dialogue Analyse spectrale



- ▶ Sélectionnez des variables dans la liste disponible, puis déplacez-les vers la liste des variables. Cette liste contient uniquement des variables numériques.
- ▶ Sélectionnez l'une des options Fenêtre spectrale pour choisir le mode de lissage du périodogramme, en vue d'obtenir une estimation de densité spectrale. Les options de lissage disponibles sont les suivantes : Tukey-Hamming, Tukey, Parzen, Bartlett, Daniell (unité) et Aucun.
 - **Tukey-Hamming.** Les pondérations sont $W_k = 0,54D_p(2 \pi f_k) + 0,23D_p(2 \pi f_k + \pi/p) + 0,23D_p(2 \pi f_k - \pi/p)$, pour $k = 0, \dots, p$, où p est la partie entière de la moitié de l'intervalle et D_p est le noyau de Dirichlet d'ordre p .
 - **Tukey.** Les pondérations sont égales à $W_k = 0,5D_p(2 \pi f_k) + 0,25D_p(2 \pi f_k + \pi/p) + 0,25D_p(2 \pi f_k - \pi/p)$, pour $k = 0, \dots, p$, où p est la partie entière de la moitié de l'intervalle et D_p est le noyau de Dirichlet d'ordre p .
 - **Parzen.** Les pondérations sont $W_k = 1/p(2 + \cos(2 \pi f_k)) (F_p/2(2 \pi f_k))^{**2}$ pour $k = 0, \dots, p$, p étant la partie entière du nombre de largeurs divisé par 2 et $F[p/2]$ le noyau de Fejer d'ordre $p/2$.
 - **Bartlett.** Forme d'une fenêtre spectrale pour laquelle les pondérations de la moitié supérieure de la fenêtre sont calculées de la façon suivante : $W_k = F_p(2 * \pi * f_k)$ pour $k = 0$ jusqu'à p , p étant la partie entière de la moitié de l'intervalle et F_p le noyau de Fejer d'ordre p . La moitié inférieure est symétrique à la moitié supérieure.
 - **Daniell (unité).** Forme d'une fenêtre spectrale pour laquelle les pondérations sont égales à 1.
 - **Aucun.** Pas de lissage. Si cette option est sélectionnée, la densité spectrale estimée est identique à celle évaluée par le périodogramme.

Etendue : Plage de valeurs consécutives sur laquelle un lissage est effectué. Généralement, un entier impair est utilisé. Les grands intervalles appliquent à la densité spectrale du diagramme un lissage plus important que les intervalles réduits.

Centrer les variables. Ajuste la série de manière à obtenir une moyenne égale à zéro avant de calculer le spectre et d'éliminer le terme élevé qui peut être associé avec la moyenne de la série.

Analyse bivariée—chaque variable avec la première. Si vous avez sélectionné au moins deux variables, vous pouvez sélectionner cette option pour demander des analyses spectrales bivariées.

- La première variable de la liste Variable(s) est traitée comme une variable explicative et toutes les autres variables comme des variables dépendantes.
- Chaque série suivant la première est analysée avec la première série indépendamment de l'autre série nommée. Des analyses univariées de chaque série sont également lancées.

Diagramme : Le périodogramme et la densité spectrale sont disponibles pour les analyses univariées et bivariées. Toutes les autres options sont disponibles uniquement pour les analyses bivariées.

- **Périodogramme.** Diagramme non lissé d'amplitude spectrale (sur une échelle logarithmique) par rapport à l'effectif ou à la période. De faibles variations caractérisent une série lissée. Des variations étendues sur tous les effectifs indiquent un « bruit blanc ».
- **Cohérence carrée.** Le produit des gains des deux séries.
- **Quadrature spectrum.** Partie imaginaire du périodogramme-croisé, qui est une mesure de la corrélation des composantes de fréquence décalés de deux séries chronologiques. Les composantes sont décalées de $\pi/2$ radians.
- **Amplitude croisée.** La racine carrée de la somme du carré de la densité cospectrale et du carré du spectre en quadrature.
- **Densité spectrale.** Périodogramme qui a été lissé pour en éliminer les variations irrégulières.
- **Densité cospectrale.** La partie réelle d'un périodogramme croisé qui correspond à la corrélation entre les composantes des effectifs en phase de deux séries.
- **Spectre de phase.** Une mesure de l'étendue pour laquelle la composante d'effectif d'une série succède ou précède une autre.
- **Gain.** Quotient de la division de l'amplitude croisée par la densité spectrale pour l'une des séries. Chacune des deux séries dispose de sa propre valeur de gains.

Par effectif. Tous les diagrammes sont créés par effectif, de l'effectif 0 (constante ou moyenne) à l'effectif 0,5 (terme correspondant à un cycle de deux observations).

Par période. Tous les diagrammes sont produits par période, allant de 2 (cycle de deux observations) à une période égale au nombre d'observations (constante ou moyenne). La période est affichée sur une échelle logarithmique.

Fonctionnalités supplémentaires de la commande SPECTRA

Le langage de syntaxe de commande vous permet aussi de :

- Enregistrer les variables d'analyse spectrale calculées dans l'ensemble de données actif en vue d'une prochaine utilisation.
- Spécifier les pondérations personnalisées pour la fenêtre spectrale.

- Produire les diagrammes par fréquence et par période.
- Imprimer un listing complet de chaque valeur affichée dans le diagramme.

Pour obtenir des renseignements complets sur la syntaxe, reportez-vous au manuel *Command Syntax Reference*.

Partie II: Exemples

Prévision globale avec le modélisateur expert

Un analyste d'un fournisseur national de connexions haut débit doit réaliser des prévisions des abonnements utilisateurs afin de prédire le développement de l'utilisation du haut débit. Des prévisions sont nécessaires pour chacun des 85 marchés locaux qui constituent la base nationale des abonnements. Les données historiques mensuelles sont collectées dans le fichier *broadband_1.sav*. [Pour plus d'informations, reportez-vous à la section Fichiers d'exemple dans l'annexe D dans IBM SPSS Forecasting 21.](#)

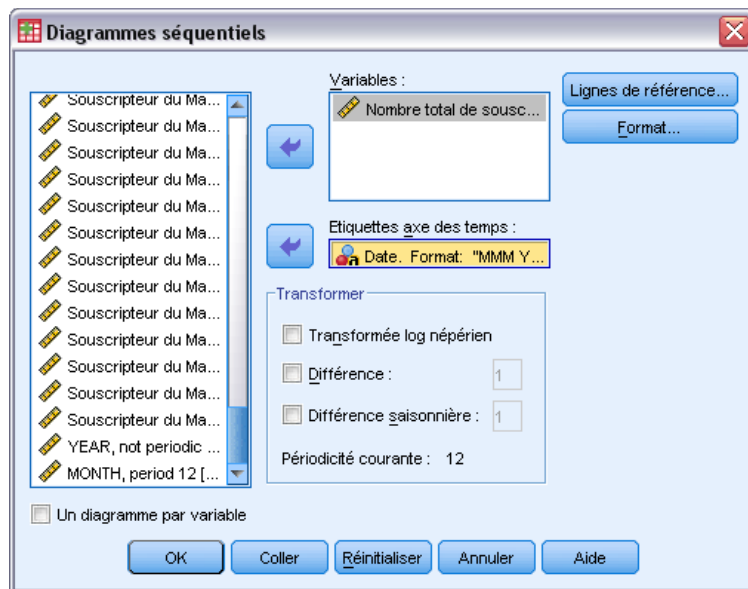
Dans cet exemple, vous utiliserez le modélisateur expert pour produire des prévisions pour les trois prochains mois pour chacun des 85 marchés locaux et vous enregistrerez les modèles générés dans un fichier XML externe. Une fois ceci terminé, continuez avec l'exemple suivant, [Reprévisions globales en appliquant des modèles enregistrés dans le chapitre 7 sur p. 58](#), qui applique les modèles enregistrés à un ensemble de données mis à jour afin d'étendre les prévisions à trois mois supplémentaires sans avoir à reconstruire les modèles.

Examen de vos données

Il s'avère toujours utile d'examiner la nature de vos données avant de construire un modèle. Ces données présentent-elles des variations saisonnières ? Bien que le modélisateur expert recherche automatiquement le meilleur modèle saisonnier ou non-saisonnier pour chaque série, vous obtiendrez des résultats plus rapides en limitant la recherche aux modèles non-saisonniers si vos données ne présentent pas cette composante. Sans avoir à examiner les données pour chacun des 85 marchés locaux, vous obtiendrez une image approximative en représentant le nombre total d'abonnés sur tous les marchés.

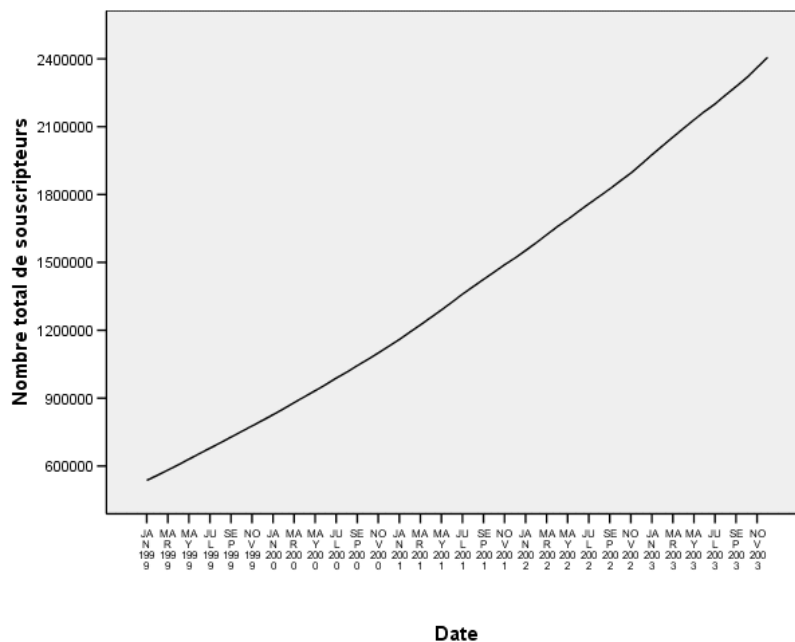
- ▶ A partir des menus, sélectionnez :
Analyse > Prévisions > Diagrammes séquentiels...

Figure 6-1
Boîte de dialogue Diagrammes séquentiels



- ▶ Sélectionnez *Nombre total d'abonnés* et déplacez-le dans la liste Variables.
- ▶ Sélectionnez la variable *Date* et placez-la dans la case Etiquettes axe des temps.
- ▶ Cliquez sur OK.

Figure 6-2
Nombre total d'abonnés au haut débit sur tous les marchés



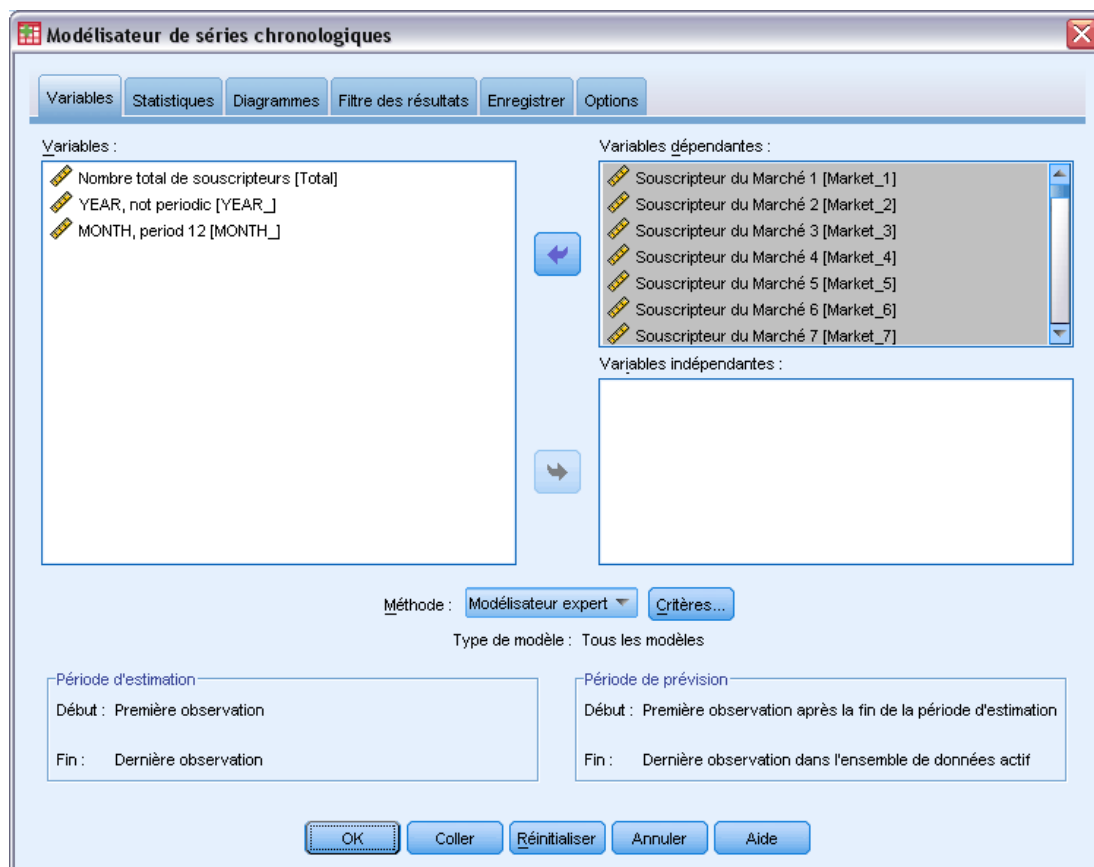
La série montre une tendance à la hausse très lisse avec aucun pic de variation saisonnière. Il se peut qu'il y ait des séries individuelles présentant une composante saisonnière mais cette composante n'est pas une particularité dominante dans les données en général. Vous devez bien sûr inspecter chaque série avant d'exclure les modèles saisonniers. Vous pouvez alors séparer les séries présentant une composante saisonnière et les modéliser à part. Dans le cas présent, l'inspection des 85 séries a prouvé qu'aucune d'entre elles ne présentait de composante saisonnière.

Exécution de l'analyse

Pour utiliser le modélisateur expert :

- ▶ A partir des menus, sélectionnez :
Analyse > Prévisions > Créer des modèles...

Figure 6-3
Boîte de dialogue Modélisateur de séries chronologiques



- ▶ Sélectionnez *Abonnés du marché 1* jusqu'à *Abonnés du marché 85* comme variables dépendantes.

- Vérifiez que le Modélisateur expert est sélectionné dans la liste déroulante Méthode. Le modélisateur expert trouve automatiquement le modèle le plus approprié pour chaque série de variables dépendantes.

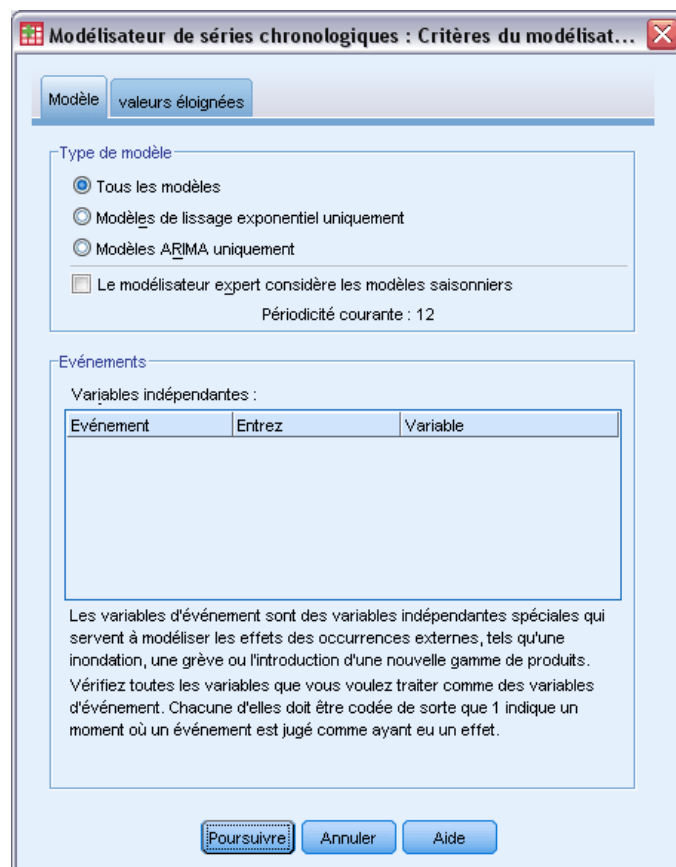
L'ensemble des observations utilisées pour estimer le modèle est appelé **période d'estimation**. Par défaut, elle inclut toutes les observations de l'ensemble de données actif. Vous pouvez définir la période d'estimation en sélectionnant Dans un intervalle de temps ou d'observations dans la boîte de dialogue Sélectionner des observations. Pour cet exemple, nous garderons la période d'estimation par défaut.

Notez également que la période de prévision par défaut débute après la période d'estimation et s'étend jusqu'à la dernière observation de l'ensemble de données actif. Si vous faites des prévisions au-delà de la dernière observation, vous devrez étendre la période de prévision. Cela se fait dans l'onglet Options comme nous le verrons plus tard dans cet exemple.

- Cliquez sur Critères.

Figure 6-4

Boîte de dialogue Critères du modélisateur expert, onglet Modèle



- Désélectionnez Le modélisateur expert considère les modèles saisonniers dans le groupe Type de modèle.

Bien que les données soient mensuelles et la périodicité actuelle soit de 12, nous avons déterminé que les données ne présentaient pas de composante saisonnière. Il n'est donc pas nécessaire de prendre en considération les modèles saisonniers. Cela réduit le nombre de modèles recherchés par le modélisateur expert et permet de réduire de manière significative les temps de calcul.

- ▶ Cliquez sur Poursuivre.
- ▶ Cliquez sur l'onglet Options dans la boîte de dialogue Modélisateur des séries chronologiques.

Figure 6-5
Modélisateur de séries chronologiques, onglet Options

Modélisateur de séries chronologiques

Variables Statistiques Diagrammes Filtre des résultats Enregistrer Options

Période de prévision

Première observation après la fin de la période d'estimation jusqu'à la dernière observation de l'ensemble de données actif

Première observation après la fin de la période d'estimation jusqu'à une date spécifique

Date :

| Année | Mois |
|-------|------|
| 2004 | 3 |

Valeurs manquantes spécifiées

Traiter comme non valide

Traiter comme valide

Largeurs des intervalles de confiance (%) : 95

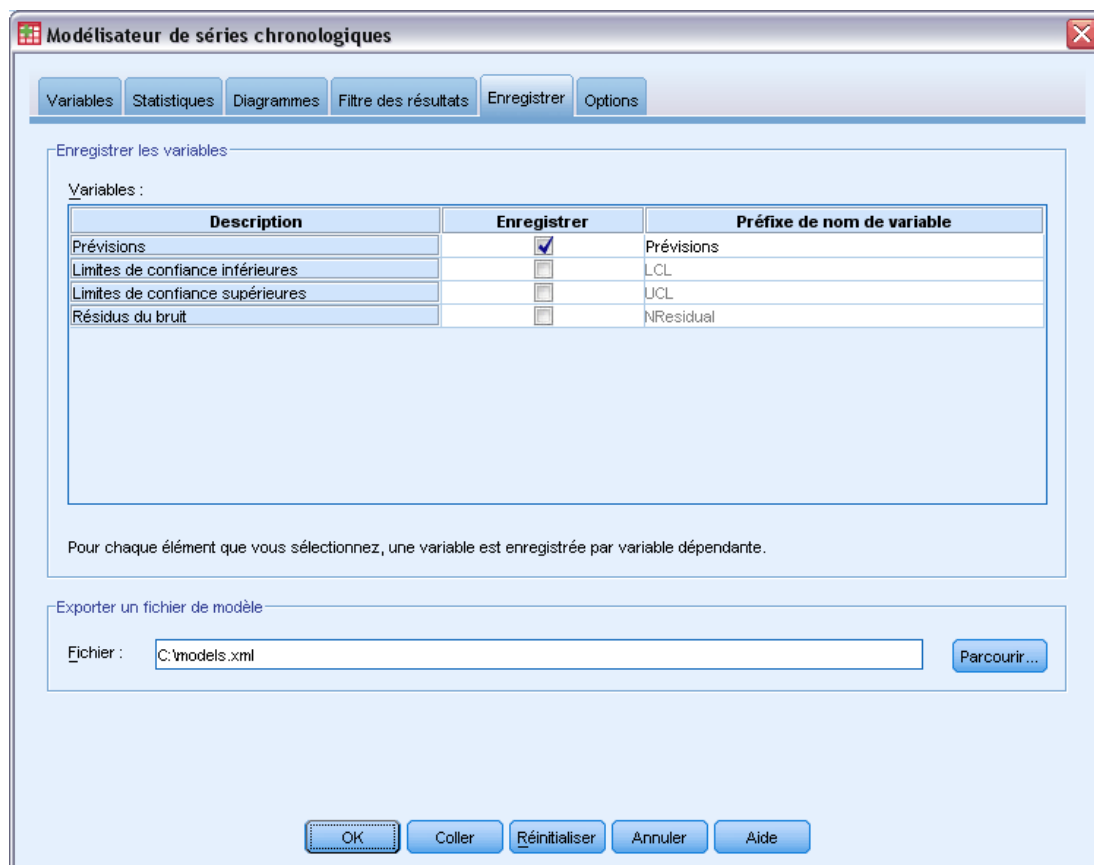
Préfixe pour les identificateurs de modèle dans les résultats : Modèle

Nombre maximum de décalages affichés dans les résultats ACF et PACF : 24

OK Coller Réinitialiser Annuler Aide

- ▶ Sélectionnez Première observation après la fin de la période d'estimation jusqu'à une date spécifique dans le groupe Période de prévision.
 - ▶ Dans la grille Date, saisissez 2004 pour l'année et 3 pour le mois.
- L'ensemble de données contient des données de janvier 1999 à décembre 2003. Selon les paramètres actuels, la période de prévision sera de janvier 2004 à mars 2004.
- ▶ Cliquez sur l'onglet Enregistrer.

Figure 6-6
Modélisateur de séries chronologiques, onglet Enregistrer

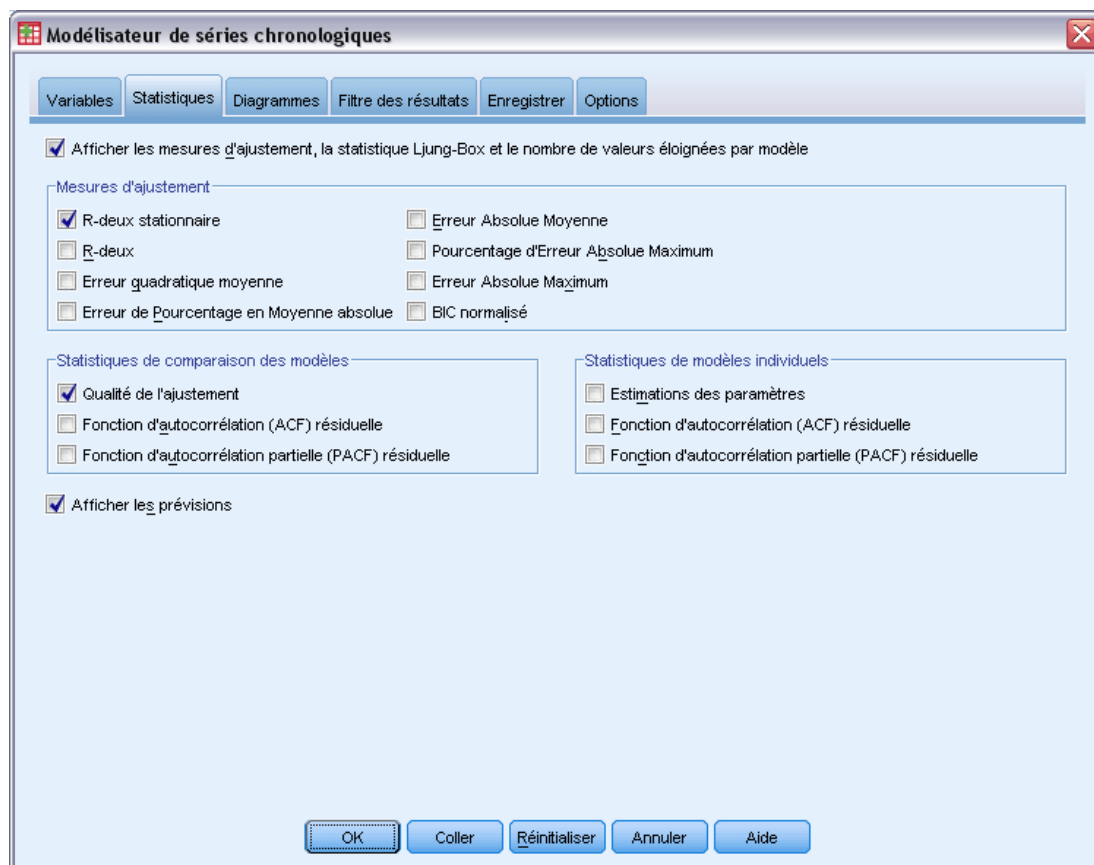


- Sélectionnez (cochez) l'entrée pour les Prévisions dans la colonne *Enregistrer* et laissez la valeur par défaut *Estimé* comme Préfixe de nom de variable.

Les prévisions du modèle sont enregistrées comme de nouvelles variables dans l'ensemble de données actif et utilisent le préfixe *Estimé* pour les noms de variables. Vous pouvez également enregistrer les spécifications pour chacun des modèles dans un fichier XML externe. Cela vous permettra de réutiliser les modèles pour étendre les prévisions lorsque de nouvelles données seront disponibles.

- Cliquez sur le bouton Parcourir dans l'onglet Enregistrer.
Cela fera apparaître une boîte de dialogue standard pour enregistrer un fichier.
- Recherchez le dossier où vous souhaitez enregistrer le fichier XML de modèle, saisissez un nom de fichier puis cliquez sur Enregistrer.
- Cliquez sur l'onglet Statistiques.

Figure 6-7
Modélisateur de séries chronologiques, onglet Statistiques



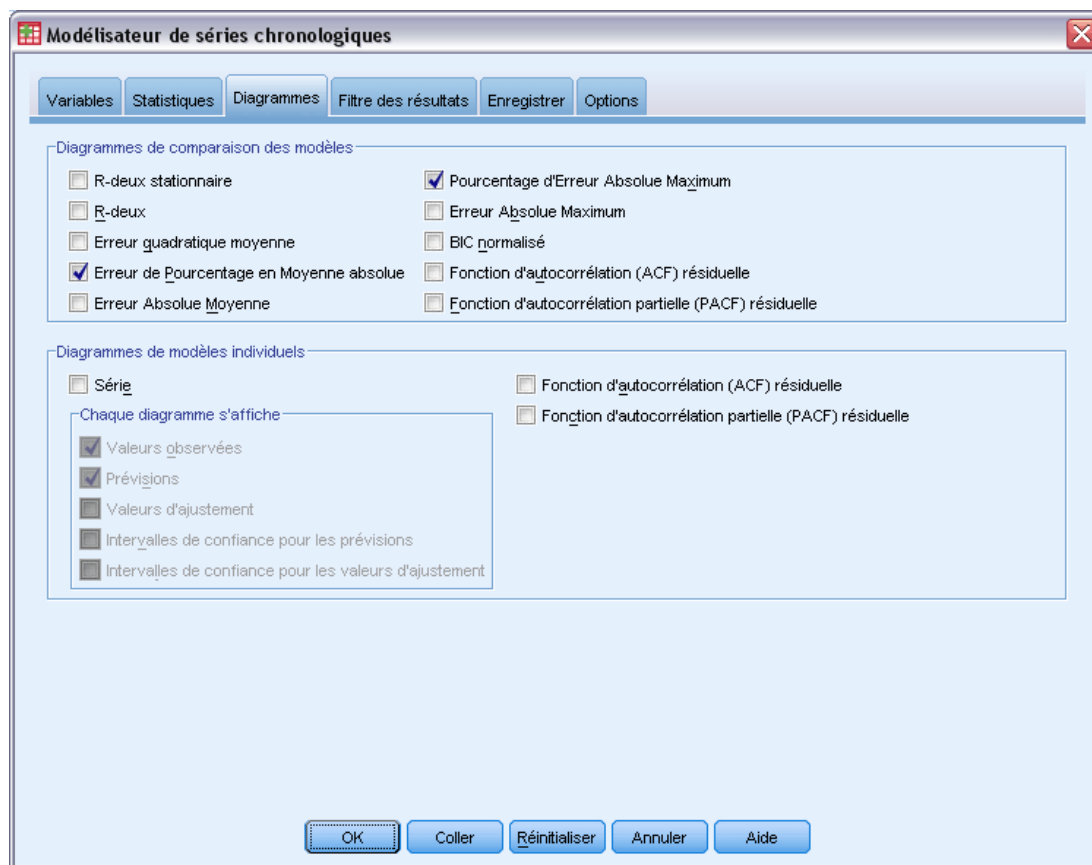
- Sélectionnez Afficher les prévisions.

Cette option produit un tableau des valeurs de prévision pour chaque série de variables dépendantes et fournit une autre option, différente de l'enregistrement des prévisions en tant que nouvelles variables, pour l'obtention de ces valeurs.

La sélection par défaut Qualité de l'ajustement (dans le groupe Statistiques pour la comparaison des modèles) produit un tableau avec les statistiques d'ajustement—comme le *R*-deux, l'Erreur de Pourcentage en Moyenne absolue et le BIC normalisé—calculées sur tous les modèles. Ce tableau offre un résumé concis de l'adéquation des modèles aux données.

- Cliquez sur l'onglet Diagrammes.

Figure 6-8
Modélisateur de séries chronologiques, onglet Diagrammes



- Désélectionnez **Séries** dans le groupe **Diagrammes de modèles individuels**.

Ceci supprime la création de diagrammes de séries pour chaque modèle. Dans cet exemple, nous sommes plus intéressés par l'enregistrement des prévisions en tant que nouvelles variables que par la création de diagrammes pour les prévisions.

Le groupe **Diagrammes** pour la comparaison des modèles fournit plusieurs diagrammes (sous la forme d'histogrammes) de statistiques d'ajustement calculées sur tous les modèles.

- Sélectionnez **Erreur de Pourcentage en Moyenne absolue** et **Pourcentage d'Erreur Absolue Maximum** dans le groupe **Diagrammes** pour la comparaison des modèles.

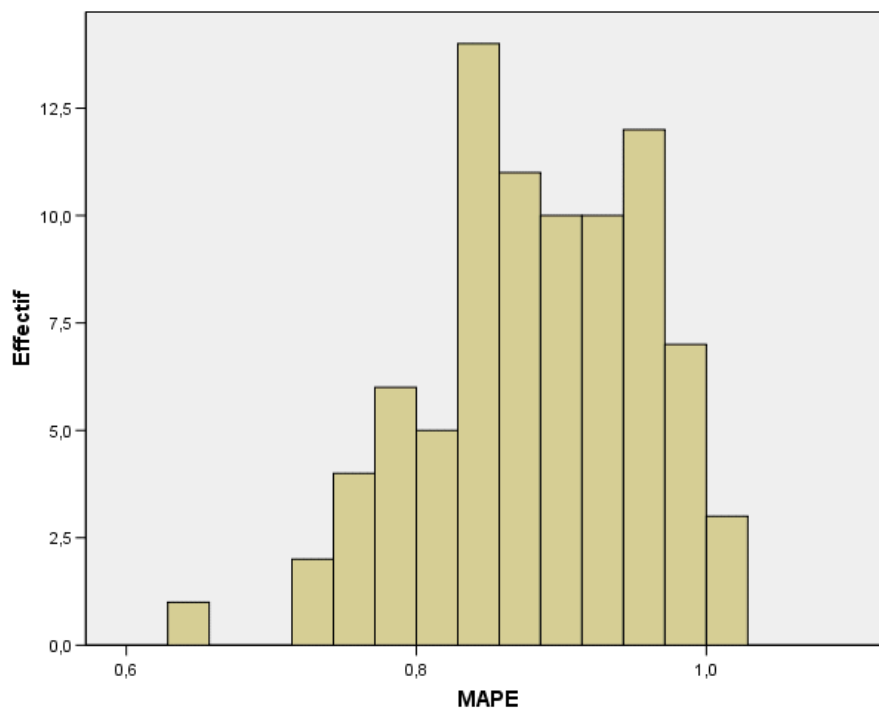
Le pourcentage d'erreur absolue mesure la variation de la série dépendante par rapport au niveau prévu par le modèle. En examinant la moyenne et le maximum parmi tous les modèles, vous obtiendrez une indication de l'incertitude de vos prévisions. Le fait de regarder les diagrammes récapitulatifs d'erreurs de pourcentage plutôt que les erreurs absolues est conseillé car les séries dépendantes représentent le nombre d'abonnés pour des marchés de taille différente.

- Cliquez sur **OK** dans la boîte de dialogue **Modélisateur de séries chronologiques**.

Diagrammes récapitulatifs du modèle

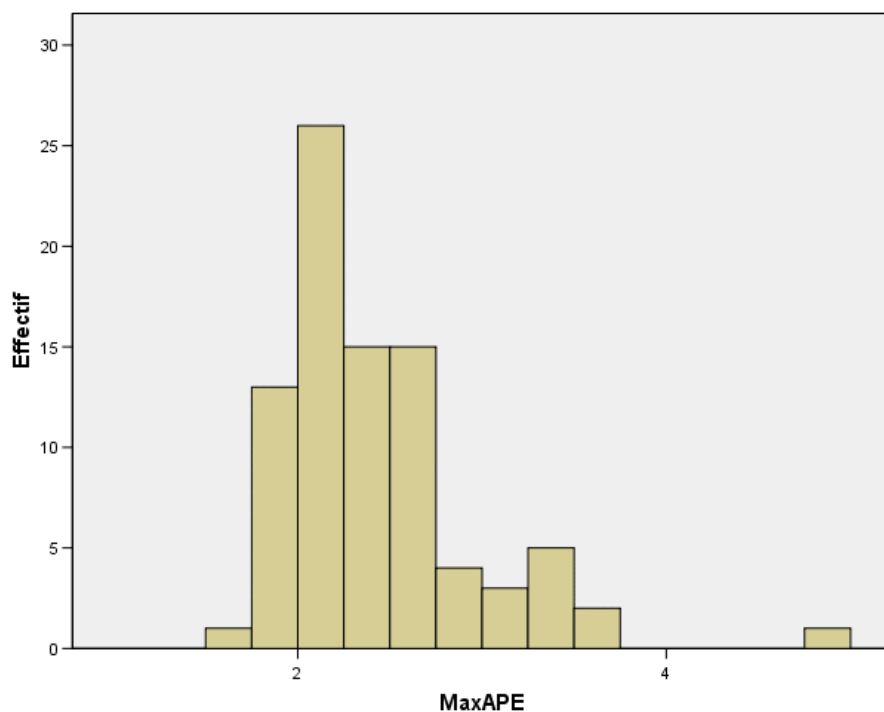
Figure 6-9

Histogramme de l'Erreur de Pourcentage en Moyenne absolue



Cet histogramme affiche l'Erreur de Pourcentage en Moyenne absolue (MAPE) de tous les modèles. Il indique que tous les modèles affiche une incertitude moyenne d'environ 1 %.

Figure 6-10
Histogramme de Pourcentage d'Erreur Absolue Maximum



Cet histogramme affiche le Pourcentage d'Erreur Absolue Maximum (MaxAPE) parmi tous les modèles. Il est utile pour envisager le pire scénario pour vos prévisions. Il indique que le pourcentage d'erreur le plus important pour chaque modèle est de l'ordre de 1 à 5 %. Ces valeurs représentent-elles un niveau d'incertitude acceptable ? C'est une situation où votre sens commercial doit entrer en œuvre car le risque acceptable est une notion qui varie de problème à problème.

Prévisions du modèle

Figure 6-11
Nouvelles variables contenant les prévisions du modèle

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

L'Editeur de données indique les nouvelles variables contenant les prévisions du modèle. Bien que seules deux soient affichées, il y a 85 nouvelles variables, une pour chacune des séries dépendantes. Le nom des variables comprend le préfixe *Estimé*, suivi du nom de la variable

dépendante associée (par exemple *Marché_1*), suivi de l'identificateur de modèle (par exemple *Modèle_1*).

Trois nouvelles observations, contenant les prévisions de janvier 2004 à mars 2004 ont été ajoutées à l'ensemble de données, ainsi que des étiquettes de date générées automatiquement. Chacune des nouvelles variables contient les prévisions du modèle pour la période d'estimation (janvier 1999 à décembre 2003), vous permettant de voir l'adéquation du modèle aux valeurs connues.

Figure 6-12
Tableau des prévisions

| Modèle | | JAN 2004 | FEV 2004 | MAR 2004 |
|--------------------------------------|-----------|----------|----------|----------|
| Souscripteur du Marché 1-Modèle_1 | Prévision | 11503 | 11447 | 11390 |
| | UCL | 11686 | 11767 | 11870 |
| | LCL | 11321 | 11126 | 10910 |
| Souscripteur du Marché 2-Modèle_2 | Prévision | 54893 | 55856 | 56704 |
| | UCL | 55632 | 57195 | 58575 |
| | LCL | 54154 | 54518 | 54832 |
| Souscripteur du Marché 3-Modèle_3 | Prévision | 59656 | 59305 | 58954 |
| | UCL | 60457 | 60753 | 61158 |
| | LCL | 58856 | 57857 | 56750 |
| Souscripteur du Marché 4-Modèle_4 | Prévision | 18235 | 18424 | 18628 |
| | UCL | 18413 | 18731 | 19121 |
| | LCL | 18058 | 18116 | 18136 |

Vous avez aussi choisi de créer un tableau avec les valeurs de prévision. Le tableau comprend les prévisions pour la période de prévision mais—contrairement aux nouvelles variables contenant les prévisions du modèle—n'inclut pas les prévisions pour la période d'estimation. Les résultats sont organisés par modèles et identifiés par le nom du modèle qui comprend le nom (étiquette) de la variable dépendante associée suivi par l'identificateur de modèle, comme le nom des nouvelles variables contenant les prévisions du modèle. Le tableau inclut aussi les limites de confiance supérieures (UCL) et les limites de confiance inférieures (LCL) pour les valeurs des prévisions (95 % par défaut).

Vous avez désormais vu deux approches pour obtenir des valeurs de prévision : en enregistrant les prévisions en tant que nouvelles variables dans l'ensemble de données actif et en créant un tableau des prévisions. Avec chaque approche, vous disposerez d'un nombre d'options disponibles pour exporter vos prévisions (par exemple, dans une feuille de calcul Excel).

Récapitulatif

Vous avez appris à utiliser le modélisateur expert pour produire des prévisions pour des séries multiples et vous avez enregistré les modèles en résultant dans un fichier XML externe. Dans le prochain exemple, vous apprendrez à étendre vos prévisions lorsque de nouvelles données seront disponibles—sans avoir à reconstruire vos modèles—en utilisant la procédure Appliquer des modèles de séries chronologiques.

Reprévisions globales en appliquant des modèles enregistrés

Vous avez utilisé le modélisateur de séries chronologiques pour créer des modèles pour vos données de séries chronologiques et produire les prévisions initiales basées sur les données disponibles. Vous planifiez de réutiliser ces modèles pour étendre vos prévisions lorsque des données actualisées seront à votre disposition, donc vous les avez enregistrés dans un fichier externe. Vous pouvez désormais réappliquer les modèles enregistrés.

Cet exemple est le prolongement naturel de l'exemple précédent, [Prévision globale avec le modélisateur expert dans le chapitre 6 sur p. 47](#) mais il peut être également utilisé indépendamment. Dans ce scénario, vous êtes l'analyste d'un fournisseur Internet haut-débit national, chargé de produire des prévisions mensuelles concernant les abonnements utilisateurs pour chacun des 85 marchés locaux. Vous avez déjà utilisé le modélisateur expert pour créer des modèles et effectuer des prévisions à trois mois. Votre banque de données a été actualisée avec des données vérifiées pour la période de prévision d'origine, donc vous souhaitez utiliser ces données pour étendre votre horizon de prévision de trois mois supplémentaires.

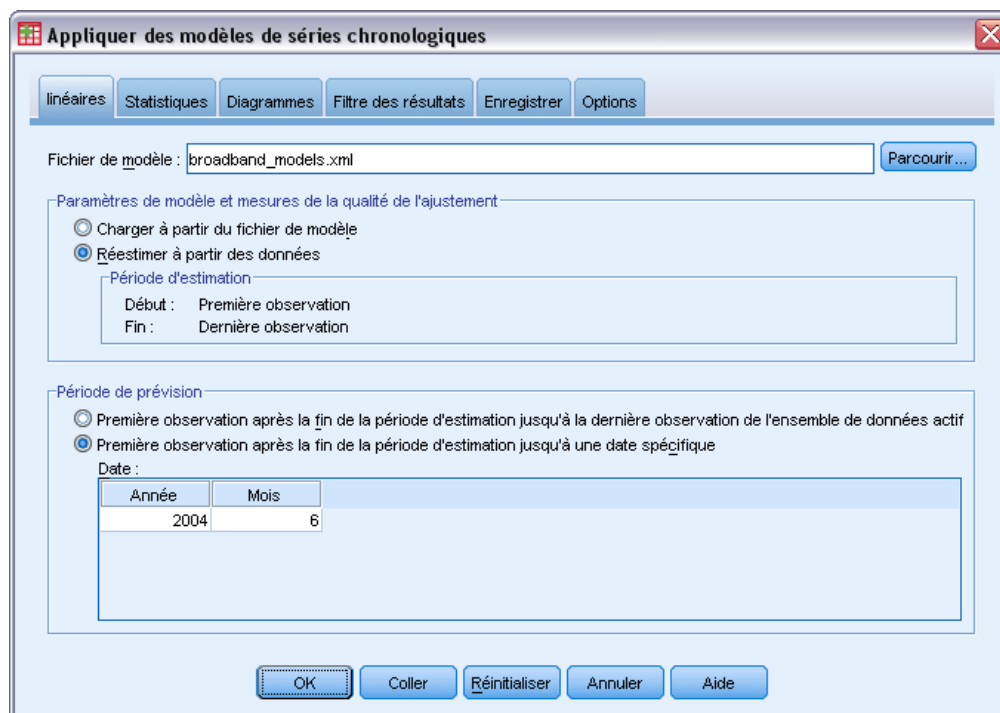
Les données historiques mensuelles mises à jour sont rassemblées dans le fichier *broadband_2.sav* et les modèles enregistrés dans le fichier *broadband_models.xml*. [Pour plus d'informations, reportez-vous à la section Fichiers d'exemple dans l'annexe D dans IBM SPSS Forecasting 21](#). Bien sûr, si vous avez travaillé sur l'exemple précédent et que vous avez enregistré votre propre fichier modèle, vous pouvez utiliser ce dernier au lieu du fichier *broadband_models.xml*.

Exécution de l'analyse

Pour appliquer des modèles :

- ▶ A partir des menus, sélectionnez :
Analyse > Prévisions > Appliquer des modèles...

Figure 7-1
Boîte de dialogue Appliquer des modèles de séries chronologiques



- Cliquez sur Parcourir, puis recherchez et sélectionnez le fichier *broadband_models.xml* (ou choisissez votre propre fichier de modèle enregistré dans l'exemple précédent). [Pour plus d'informations, reportez-vous à la section Fichiers d'exemple dans l'annexe D dans IBM SPSS Forecasting 21.](#)

- Sélectionnez Réestimer à partir des données.

Pour incorporer les nouvelles valeurs de vos séries chronologiques dans vos prévisions, la procédure Appliquer des modèles des séries chronologiques devra réestimer les paramètres du modèle. La structure des modèles reste cependant la même, ce qui implique un calcul de réestimation beaucoup plus rapide que le calcul initial pour la création des modèles.

L'ensemble des observations utilisées pour la réestimation doit inclure les nouvelles données. Cette condition est respectée si vous utilisez la période d'estimation par défaut s'étendant de la première observation à la dernière observation. Si vous devez définir une période d'estimation différente de la période par défaut, sélectionnez Dans un intervalle de temps ou d'observations dans la boîte de dialogue Sélectionner des observations.

- Sélectionnez Première observation après la fin de la période d'estimation jusqu'à une date spécifique dans le groupe Période de prévision.
- Dans la grille Date, saisissez 2004 pour l'année et 6 pour le mois.

L'ensemble de données contient des données de janvier 1999 à mars 2004. Selon les paramètres actuels, la période de prévision sera d'avril 2004 à juin 2004.

- Cliquez sur l'onglet Enregistrer.

Figure 7-2
Appliquer des modèles de séries chronologiques, onglet Enregistrer

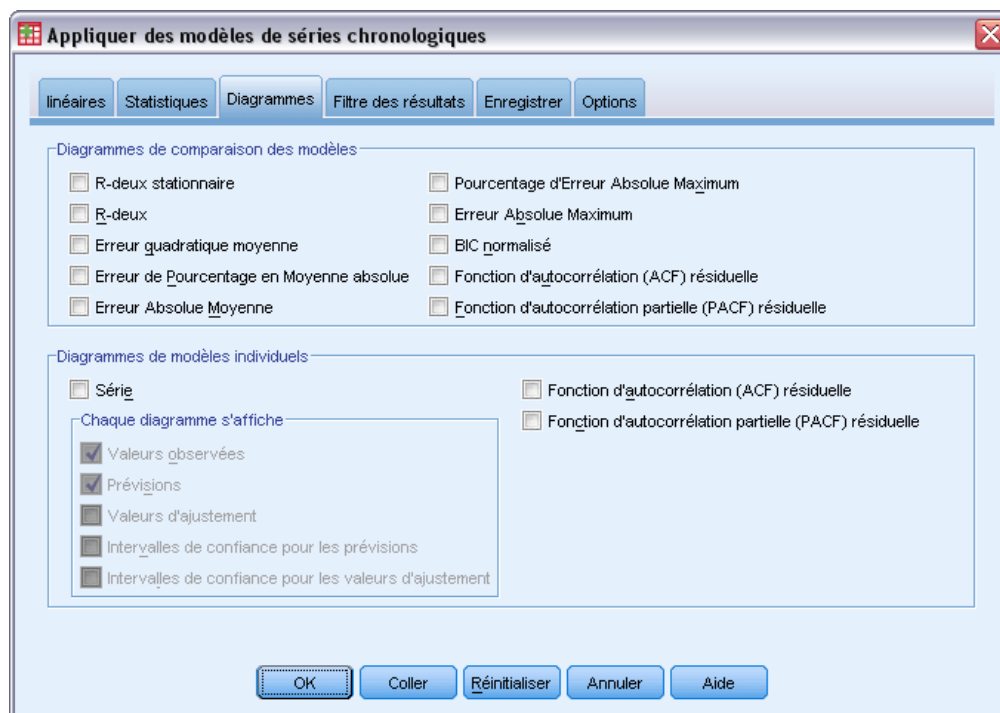


- Sélectionnez (cochez) l'entrée pour les Prévisions dans la colonne *Enregistrer* et laissez la valeur par défaut *Prévision* comme Préfixe de nom de variable.

Les prévisions du modèle sont enregistrées comme de nouvelles variables dans l'ensemble de données actif et utilisent le préfixe *Prévision* pour les noms de variables.

- Cliquez sur l'onglet Diagrammes.

Figure 7-3
Appliquer des modèles de séries chronologiques, onglet Diagrammes



- ▶ Désélectionnez **Séries** dans le groupe **Diagrammes de modèles individuels**.
Ceci supprime la création de diagrammes de séries pour chaque modèle. Dans cet exemple, nous sommes plus intéressés par l'enregistrement des prévisions en tant que nouvelles variables que par la création de diagrammes pour les prévisions.
- ▶ Cliquez sur **OK** dans la boîte de dialogue **Appliquer des modèles de séries chronologiques**.

Statistiques d'ajustement du modèle

Figure 7-4
Tableau d'ajustement du modèle

| Statistique de qualité de l'ajustement | Moyenne | SE | Minimum | Maximum | Centile | | | | | | |
|--|---------|---------|-----------|----------|-----------|---------|---------|---------|---------|----------|----------|
| | | | | | 5 | 10 | 25 | 50 | 75 | 90 | 95 |
| R-deux stationnaire | ,183 | ,144 | -2,7E-015 | ,629 | -6,0E-016 | ,000 | ,068 | ,188 | ,247 | ,376 | ,478 |
| R-deux | ,999 | ,000 | ,998 | 1,000 | ,998 | ,999 | ,999 | ,999 | ,999 | 1,000 | 1,000 |
| RMSE | 177,951 | 138,821 | 42,088 | 737,540 | 51,471 | 58,625 | 90,316 | 135,016 | 195,507 | 402,111 | 480,796 |
| MAPE | ,883 | ,075 | ,647 | 1,007 | ,748 | ,775 | ,831 | ,885 | ,946 | ,980 | ,991 |
| MaxAPE | 2,426 | ,516 | 1,708 | 4,765 | 1,798 | 1,937 | 2,097 | 2,307 | 2,604 | 3,260 | 3,676 |
| MAE | 139,634 | 106,313 | 34,033 | 589,708 | 40,608 | 46,836 | 71,445 | 107,377 | 150,785 | 316,144 | 343,171 |
| MaxAE | 456,170 | 378,664 | 108,378 | 1813,295 | 117,841 | 137,366 | 206,780 | 340,308 | 517,760 | 1070,229 | 1356,113 |
| BIC standardisé | 10,003 | 1,316 | 7,618 | 13,345 | 7,994 | 8,256 | 9,111 | 9,902 | 10,619 | 12,062 | 12,445 |

Le tableau d'ajustement du modèle fournit des statistiques d'ajustement calculées à partir de tous les modèles. Il offre un résumé concis de l'adéquation des modèles comprenant les paramètres réestimés aux données. Pour chaque statistique, le tableau fournit la moyenne, l'erreur standard (SE), les valeurs minimum et maximum sur l'ensemble des modèles. Il contient également des

valeurs de centiles fournissant des informations sur la distribution des statistiques sur l'ensemble des modèles. Une valeur de centile indique que le pourcentage de modèles correspondant comporte une valeur de statistique de l'ajustement inférieure à la valeur annoncée. Par exemple, 95 % des modèles ont un pourcentage d'erreur absolue maximum (MaxAPE) inférieur à 3,676.

Un certain nombre de statistiques sont rapportées, mais nous nous concentrerons sur deux d'entre elles : MAPE (Erreur de Pourcentage en Moyenne absolue) et MaxAPE (Pourcentage d'Erreur Absolue Maximum). Le pourcentage d'erreur absolue est une mesure de la variation d'une série dépendante par rapport au niveau prévu par son modèle et fournit une indication de l'incertitude de vos prévisions. L'Erreur de Pourcentage en Moyenne absolue varie entre le minimum de 0,669 % et le maximum de 1,026 % sur l'ensemble des modèles. Le Pourcentage d'Erreur Absolue Maximum varie entre 1,742 % et 4,373 % sur l'ensemble des modèles. Ainsi, l'incertitude moyenne des prévisions de chaque modèle est d'environ 1 % et l'incertitude maximum d'environ 2,5 % (la valeur moyenne de MaxAPE), avec le plus mauvais scénario aux environs de 4 %. Le degré de risque que vous êtes disposé à accepter détermine l'acceptabilité de ces degrés d'incertitude.

Prévisions du modèle

Figure 7-5

Nouvelles variables contenant les prévisions du modèle

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

L'Editeur de données indiquent les nouvelles variables contenant les prévisions du modèle. Bien que seules deux soient affichées, il y a 85 nouvelles variables, une pour chacune des séries dépendantes. Le nom des variables comprend le préfixe *Estimé*, suivi du nom de la variable dépendante associée (par exemple *Marché_1*), suivi de l'identificateur de modèle (par exemple *Modèle_1*).

Trois nouvelles observations, contenant les prévisions d'avril 2004 à juin 2004 ont été ajoutées au fichier de données, ainsi que des étiquettes de date générées automatiquement.

Récapitulatif

Vous avez appris à appliquer des modèles enregistrés afin d'étendre vos prévisions précédentes en cas de mise à disposition de données actualisées et ceci, sans avoir à recréer des modèles. Bien sûr, si le modèle est susceptible d'avoir changé, il est conseillé de le recréer à l'aide de la procédure Modélisateur de séries chronologiques.

Utilisation du modélisateur expert pour déterminer les variables prédites significatives

Une société de vente par correspondance, intéressée par le développement d'un modèle de prévision, a collecté des données sur les ventes mensuelles de vêtements pour hommes, ainsi que plusieurs séries susceptibles d'expliquer certaines variations des ventes. Les variables prédites peuvent par exemple concerner le nombre de catalogues envoyés, le nombre de pages du catalogue, le nombre de lignes téléphoniques ouvertes pour le passage des commandes, le budget consacré à la publicité imprimée et le nombre de représentants du service clients. L'une des variables prédites est-elle utile pour la prévision ?

Dans cet exemple, utilisez le modélisateur expert avec toutes les variables prédites candidates pour trouver le meilleur modèle. Puisque que le modélisateur expert sélectionne uniquement les variables prédites qui ont une relation statistiquement importante avec les séries dépendantes, les variables prédites les plus utiles ainsi qu'un modèle pour réaliser des prévisions avec ces variables vous seront proposés. Une fois ceci terminé, continuez avec l'exemple suivant, [Test des variables indépendantes en appliquant des modèles enregistrés dans le chapitre 9 sur p. 74](#), qui étudie l'effet sur les ventes de différents scénarios de variables prédites utilisant le modèle construit dans cet exemple.

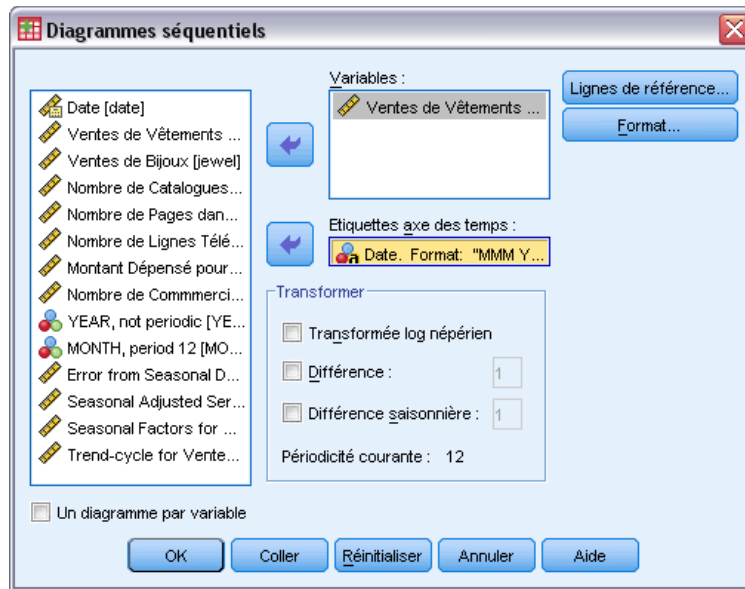
Les données pour cet exemple sont collectées dans *catalog_seasfac.sav*. [Pour plus d'informations, reportez-vous à la section Fichiers d'exemple dans l'annexe D dans IBM SPSS Forecasting 21.](#)

Représentation de vos données

Il s'avère toujours utile de représenter vos données sous forme graphique, particulièrement si vous travaillez avec une seule série :

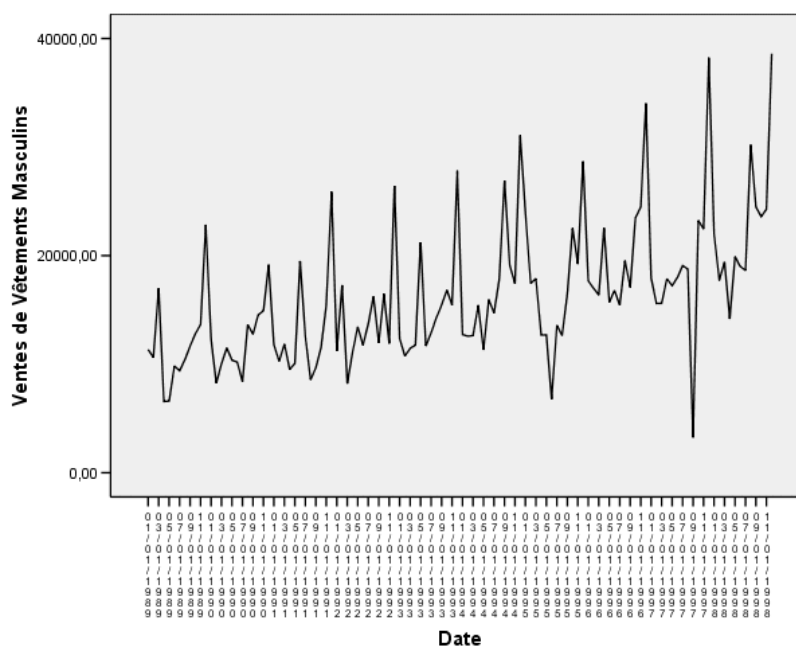
- ▶ A partir des menus, sélectionnez :
Analyse > Prévisions > Diagrammes séquentiels...

Figure 8-1
Boîte de dialogue Diagrammes séquentiels



- ▶ Sélectionnez la variable *Ventes de vêtements hommes* et placez-la dans la liste Variables.
- ▶ Sélectionnez la variable *Date* et placez-la dans la case Etiquettes axe des temps.
- ▶ Cliquez sur OK.

Figure 8-2
Ventes de vêtements pour hommes (en dollars U.S.)



Cette série présente de nombreux pics, qui apparaissent espacés de la même manière, ainsi qu'une nette tendance vers le haut. Cet espacement équitable suggère la présence d'une composante périodique dans la série chronologique. Étant donné la nature saisonnière des ventes, avec des pics survenant généralement pendant les vacances, il ne serait pas surprenant de constater que les données comportent une composante saisonnière.

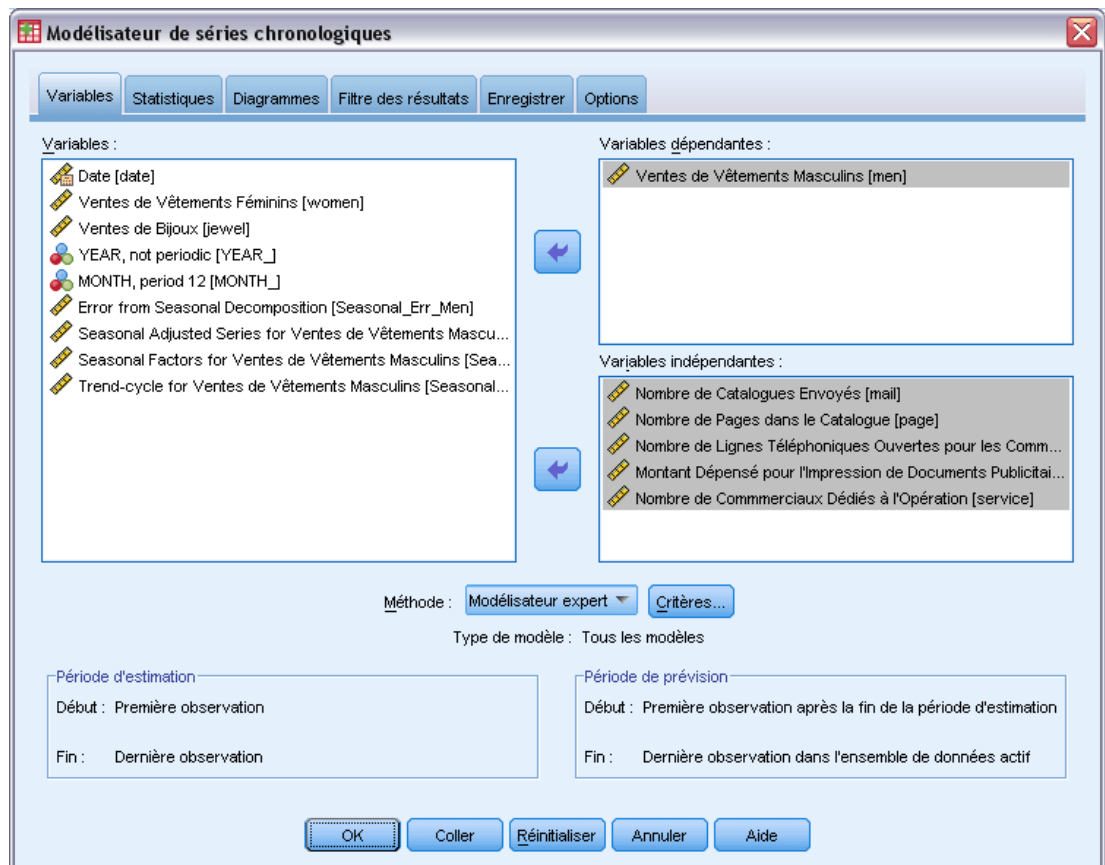
Certains pics ne semblent pas faire partie de la composante saisonnière et représentent des écarts significatifs des points de données avoisinants. Ces points peuvent être des valeurs éloignées, qui peuvent et doivent être traitées avec le modélisateur expert.

Exécution de l'analyse

Pour utiliser le modélisateur expert :

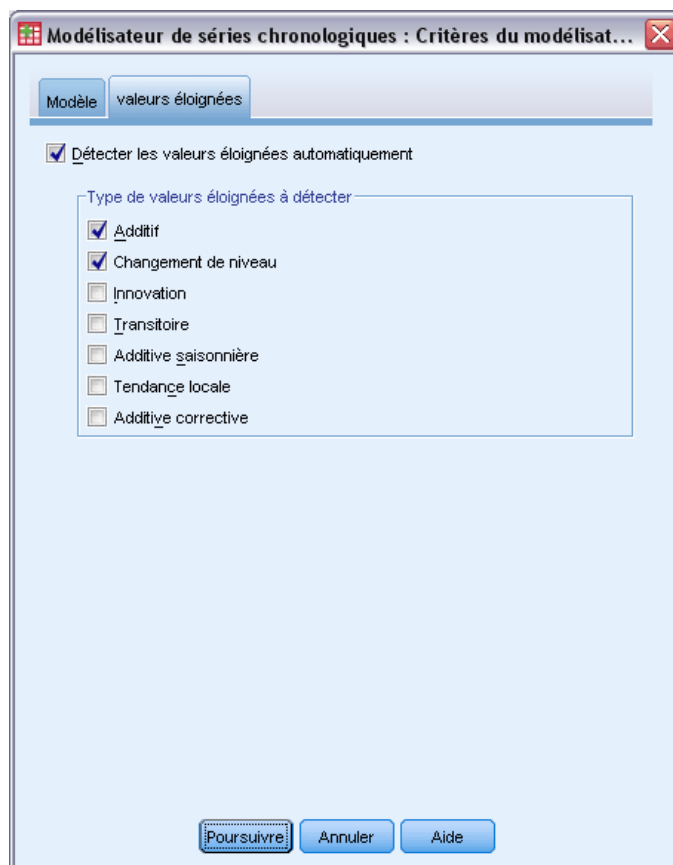
- A partir des menus, sélectionnez :
Analyse > Prévisions > Créer des modèles...

Figure 8-3
Boîte de dialogue Modélisateur de séries chronologiques



- ▶ Sélectionnez *Ventes de vêtements pour homme* pour la variable dépendante.
- ▶ Sélectionnez *Nombre de catalogues expédiés* jusqu'à *Nombre de représentants du service clients* comme variables indépendantes.
- ▶ Vérifiez que le Modélisateur expert est sélectionné dans la liste déroulante Méthode. Le modélisateur expert trouvera automatiquement le modèle saisonnier ou non saisonnier le mieux adapté pour la série de variables dépendantes.
- ▶ Cliquez sur Critères puis sur l'onglet Valeurs éloignées.

Figure 8-4
Boîte de dialogue Critères du modélisateur expert, onglet Valeurs éloignées

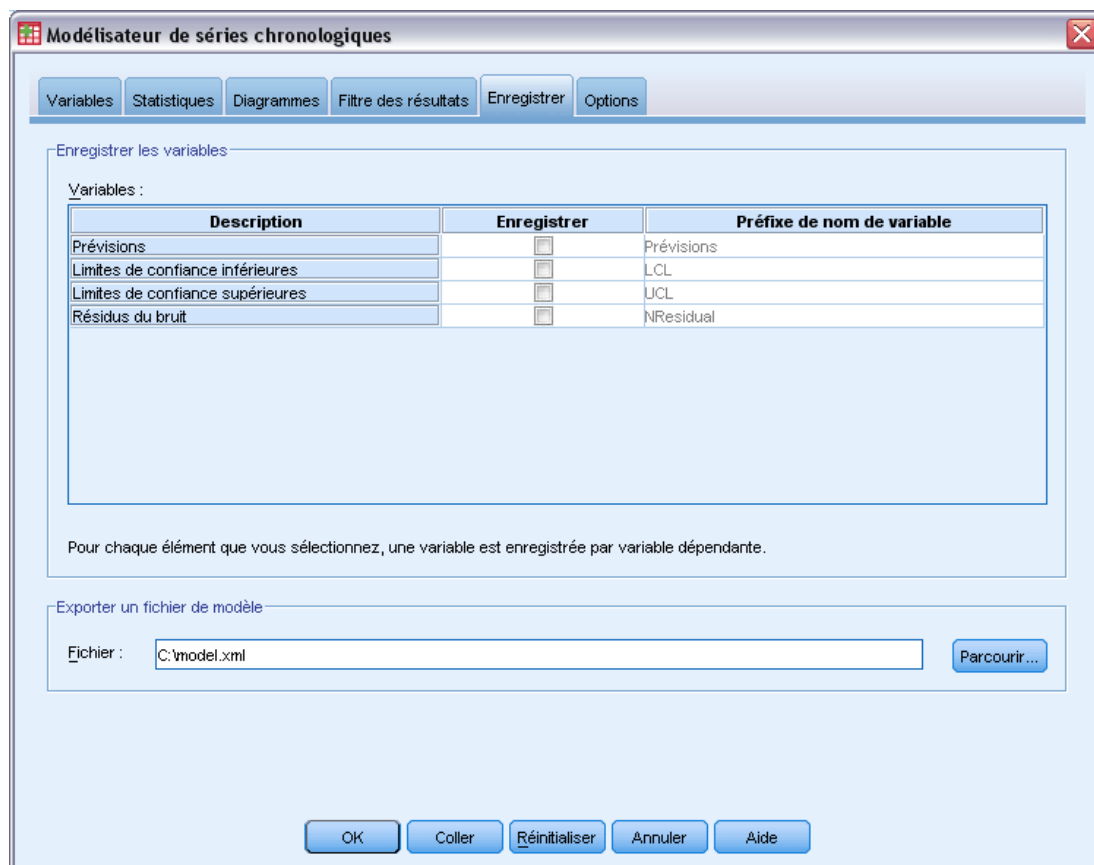


- Sélectionnez Détecter automatiquement les valeurs éloignées et gardez les sélections par défaut pour les types de valeurs éloignées à détecter.

Notre inspection visuelle des données suggèrent qu'il y a des valeurs éloignées. En se basant sur les choix actuels, le modélisateur expert recherchera les types de valeurs éloignées les plus communes et les incorporera dans le modèle final. La détection des valeurs éloignées peut augmenter de manière significative le temps de calcul nécessaire du modélisateur expert, ainsi cette fonction doit être utilisée avec parcimonie, particulièrement lors de la modélisation de plusieurs séries en même temps. Par défaut, les valeurs éloignées ne sont pas détectées.

- Cliquez sur Poursuivre.
- Cliquez sur l'onglet Enregistrer dans la boîte de dialogue Modélisateur des séries chronologiques.

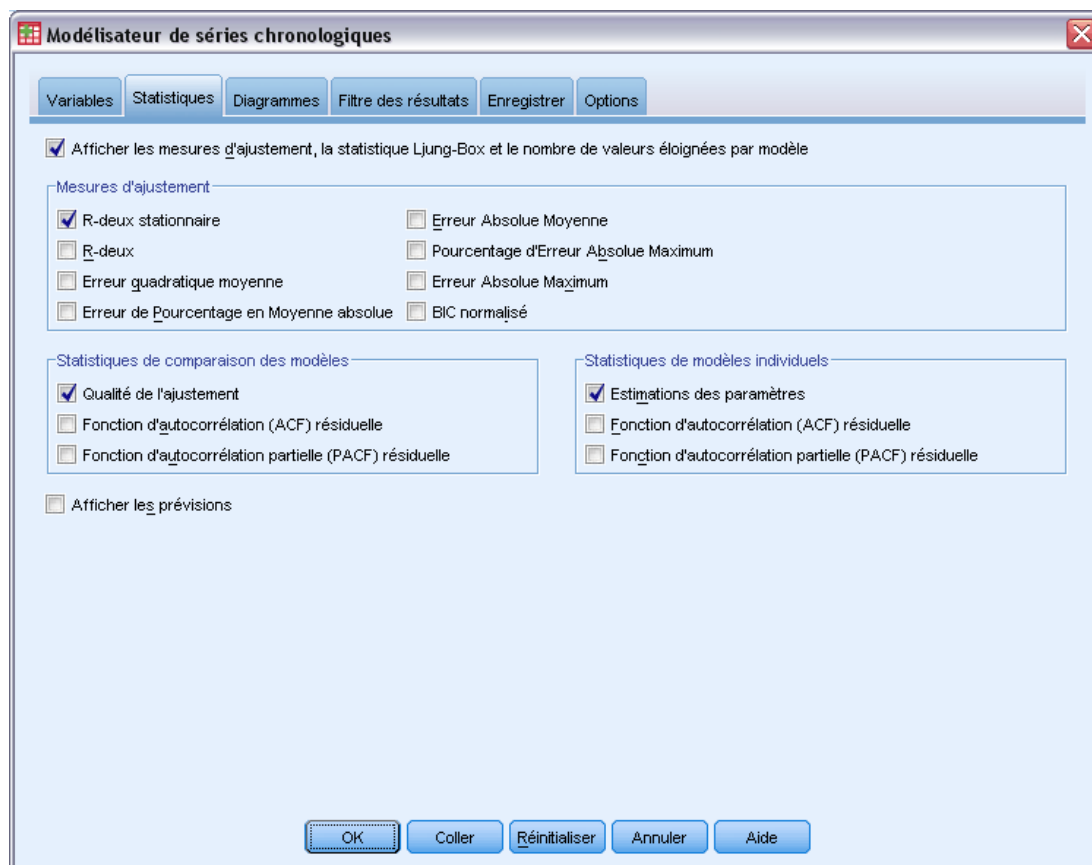
Figure 8-5
Modélisateur de séries chronologiques, onglet Enregistrer



Vous allez enregistrer le modèle estimé dans un fichier XML externe afin de pouvoir le tester avec différentes valeurs de variables prédites, en utilisant la procédure Appliquer des modèles de séries chronologiques, sans avoir à reconstruire le modèle.

- ▶ Cliquez sur le bouton Parcourir dans l'onglet Enregistrer.
Cela fera apparaître une boîte de dialogue standard pour enregistrer un fichier.
- ▶ Recherchez le dossier où vous souhaitez enregistrer le fichier XML de modèle, saisissez un nom de fichier puis cliquez sur Enregistrer.
- ▶ Cliquez sur l'onglet Statistiques.

Figure 8-6
Modélisateur de séries chronologiques, onglet Statistiques

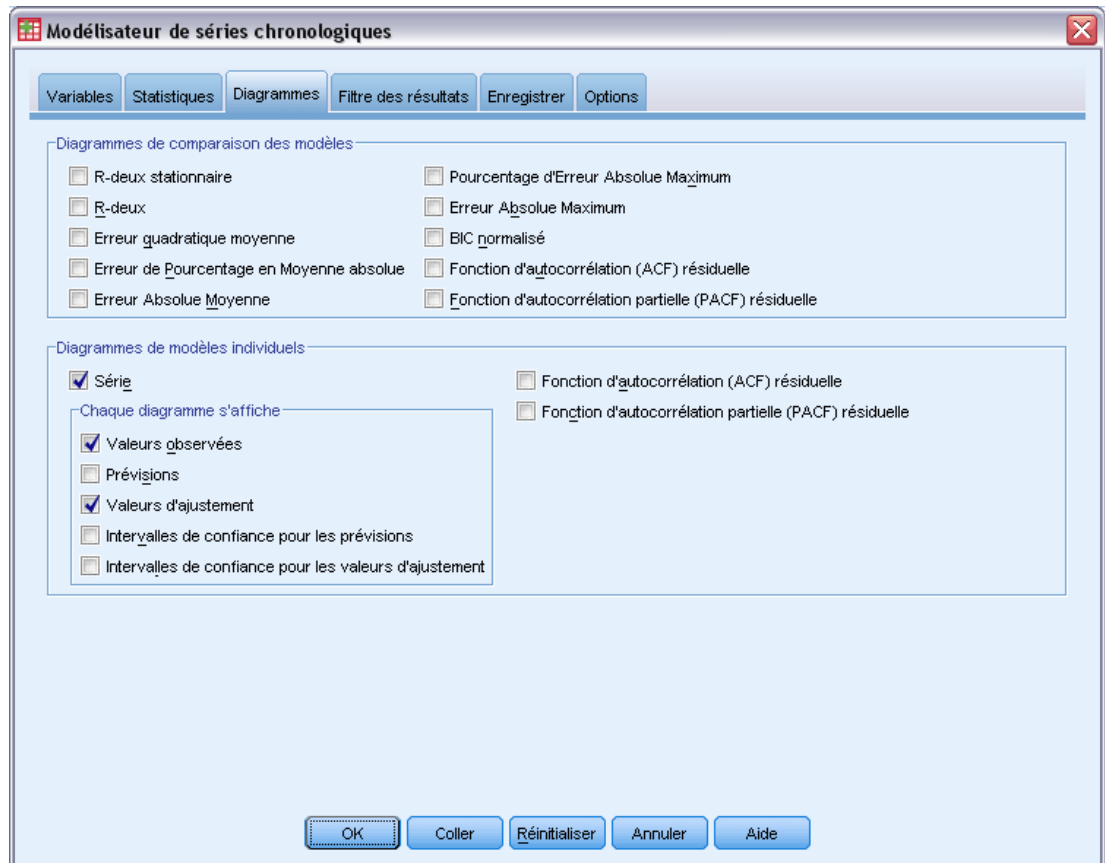


- Sélectionnez Estimations des paramètres.

Cette option produit un tableau affichant tous les paramètres, y compris les variables prédites significatives, pour le modèle choisi par le modélisateur expert.

- Cliquez sur l'onglet Diagrammes.

Figure 8-7
Modélisateur de séries chronologiques, onglet Diagrammes



- Désélectionnez l'option Prévisions.

Dans cet exemple, seules la détermination des variables prédites significatives et la construction du modèle nous intéressent. Nous n'effectuons aucune prévision.

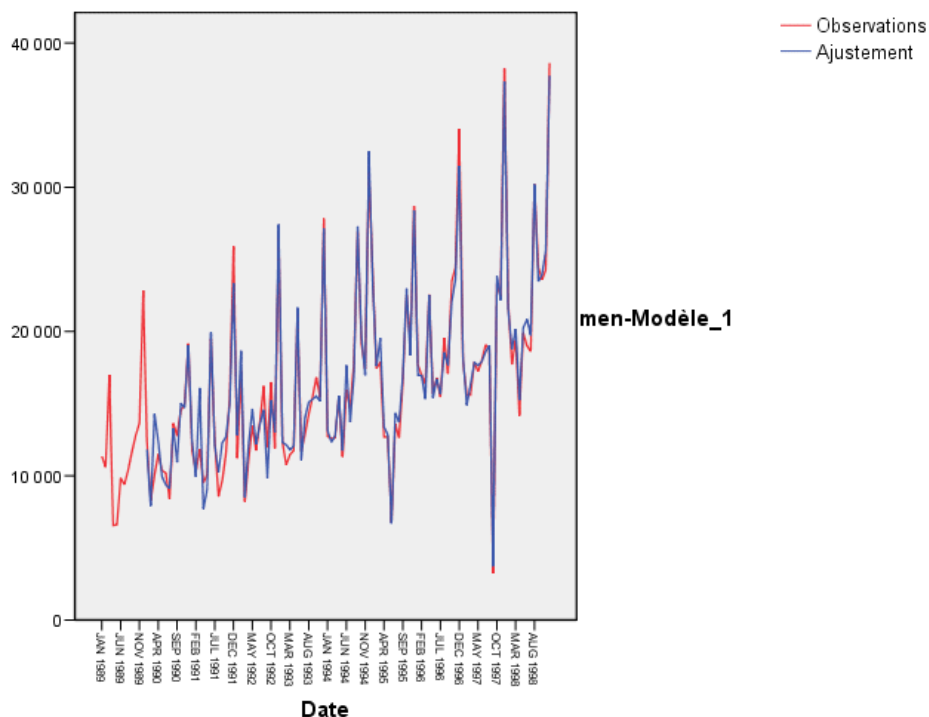
- Sélectionnez Valeurs d'ajustement.

Cette option affiche les prévisions dans la période utilisée pour estimer le modèle. Cette période est appelée **période d'estimation** et elle inclut toutes les observations de l'ensemble de données actif pour cet exemple. Ces valeurs offrent une indication sur l'adéquation du modèle aux valeurs observées, elles sont donc appelées **valeurs d'ajustement**. Le diagramme en résultant représente à la fois les valeurs observées et les valeurs d'ajustement.

- Cliquez sur OK dans la boîte de dialogue Modélisateur de séries chronologiques.

Tracé de séries

Figure 8-8
Prévisions et valeurs observées



Les prévisions sont en adéquation avec les valeurs observées, ce qui montre que le modèle offre une capacité de prévision satisfaisante. Notez la justesse avec laquelle le modèle prédit les pics saisonniers. Il réalise également une bonne capture de la tendance à la hausse des données.

Tableau de description du modèle

Figure 8-9
Tableau de description du modèle

| ID de modèle | Ventes de Vêtements Masculins | Modèle_1 | Type de modèle |
|--------------|-------------------------------|----------|---------------------|
| | | | ARIMA(0,0,0)(0,1,0) |

Le tableau de description du modèle contient une entrée pour chaque modèle estimé et comporte à la fois un identificateur de modèle et le type du modèle. L'identificateur de modèle se compose du nom (étiquette) de la variable dépendante associée et d'un nom assigné par le système. Dans cet exemple, la variable dépendante est *Ventes des vêtements homme* et le nom assigné par le système est *Modèle_1*.

Le Modélisateur de séries chronologiques prend en charge à la fois les modèles de lissage exponentiel et ARIMA. Les types de modèles de lissage exponentiel sont listés sous leurs noms communs comme Holt ou Additif de Winter. Les types de modèles ARIMA sont listés en utilisant la notation standard $ARIMA(p,d,q)(P,D,Q)$, où p est l'ordre d'autorégression, d l'ordre de

différenciation (ou intégration) et q l'ordre de la moyenne mobile, alors que (P,D,Q) sont leurs équivalents saisonniers.

Le modélisateur expert a déterminé que les ventes de vêtements homme sont les mieux décrites par un modèle saisonnier ARIMA avec un ordre de différenciation. La nature saisonnière des comptes du modèle pour les pics saisonniers qui apparaissent dans le diagramme de la série et l'ordre unique de différenciation reflètent la tendance à la hausse qui était évidente dans les données.

Tableau de statistiques du modèle

Figure 8-10
Tableau de statistiques du modèle

| Modèle | Nombre de variables indépendantes | Statistiques de qualité d'ajustement du modèle | Ljung-Box Q(18) | | | Nombre de valeurs éloignées |
|--|-----------------------------------|--|-----------------|-----|------|-----------------------------|
| | | R-deux stationnaire | Statistiques | DDL | Sig. | |
| Ventes de Vêtements Masculins-Modèle_1 | 2 | ,948 | 7,589 | 18 | ,984 | 9 |

Le tableau de statistiques du modèle fournit des informations récapitulatives et des statistiques de qualité de l'ajustement pour chaque modèle estimé. Les résultats pour chaque modèle sont étiquetés avec l'identificateur de modèle fourni dans le tableau de description du modèle. Tout d'abord, notez que le modèle contient deux variables prédites sur les cinq variables prédites candidates que vous avez spécifiées. Il se trouve donc que le modélisateur expert a identifié deux variables indépendantes qui pourraient s'avérer utiles pour la prévision.

Bien que le Modélisateur de séries chronologiques offre de nombreuses statistiques de qualité de l'ajustement, nous optons uniquement pour la valeur R -deux stationnaire. Cette statistique offre une estimation de la proportion de la variation totale de la série qui est expliquée par le modèle. Elle est préférable à la valeur R -deux ordinaire lorsqu'il y a une tendance ou une composante saisonnière comme c'est le cas ici. De plus grandes valeurs R -deux stationnaire (jusqu'à la valeur maximale de 1) indiquent un meilleur ajustement. Une valeur de 0,948 signifie que le modèle explique très bien la variation observée dans la série.

La statistique de Ljung-Box, appelée aussi statistique de Box-Pierce corrigée, fournit une indication sur la bonne spécification du modèle. Une valeur de signification inférieure à 0,05 implique qu'il existe une structure dans la série observée qui n'est pas prise en compte par le modèle. La valeur de 0,984 affichée ici n'est pas significative, donc nous sommes assurés que le modèle est correctement spécifié.

Le modélisateur expert a détecté neuf points considérés comme des valeurs éloignées. Chacun de ces points a été modélisé de façon appropriée, il n'est donc pas nécessaire de les supprimer de la série.

Tableau des paramètres du modèle ARIMA

Figure 8-11

Tableau des paramètres du modèle ARIMA

| | | | | Estimation | SE | t | Sig. |
|--|---|-----------------------|------------------------|------------|--------|--------|------|
| Ventes de Vêtements Masculins-Modèle_1 | Ventes de Vêtements Masculins | Aucune transformation | Différence saisonnière | 1 | | | |
| | Nombre de Catalogues Envoyés | Aucune transformation | Numérateur Lag 0 | 1,549 | ,071 | 21,943 | ,000 |
| | | | Différence saisonnière | 1 | | | |
| | Nombre de Lignes Téléphoniques Ouvertes | Aucune transformation | Numérateur Lag 0 | 315,262 | 15,298 | 20,607 | ,000 |
| | | | Différence saisonnière | 1 | | | |

Le tableau des paramètres du modèle ARIMA affiche les valeurs pour tous les paramètres du modèle, avec une entrée pour chaque modèle estimé étiqueté par l'identificateur de modèle. En ce qui nous concerne, il liste toutes les variables du modèle, y compris la variable dépendante et les variables indépendantes que le modélisateur expert a déterminé comme significatives. Nous savons déjà grâce au tableau des statistiques du modèle qu'il y a deux variables prédites significatives. Le tableau des paramètres du modèle nous montre que ce sont les variables *Nombre de catalogues expédiés* et *Nombre de lignes téléphoniques ouvertes pour les commandes*.

Récapitulatif

Vous avez appris à utiliser le modélisateur expert pour construire un modèle et identifier des variables prédites significatives et vous avez enregistré le modèle en question dans un fichier externe. Vous pouvez maintenant utiliser la procédure Appliquer des modèles de séries chronologiques pour le tester avec des scénarios alternatifs pour des séries de variables prédites et ainsi voir comment ces alternatives affectent les prévisions des ventes.

Test des variables indépendantes en appliquant des modèles enregistrés

Vous avez utilisé le Modélisateur de séries chronologiques pour créer un modèle pour vos données et pour identifier les variables indépendantes utiles pour les prévisions. Les variables indépendantes représentent des facteurs que vous contrôlez et dont vous souhaitez tester les valeurs dans la période de prévision pour voir comment les prévisions affectent les variables dépendantes. Cette tâche est facilement réalisable avec la procédure Appliquer des modèles de séries chronologiques, en utilisant le fichier du modèle créé avec la procédure Modélisateur de séries chronologiques.

Cet exemple est le prolongement naturel de l'exemple précédent, [Utilisation du modélisateur expert pour déterminer les variables prédites significatives dans le chapitre 8 sur p. 63](#), mais il peut être également utilisé indépendamment. Le scénario implique une entreprise de catalogue qui a recueilli des données sur les ventes mensuelles de vêtements homme entre janvier 1989 et décembre 1998, ainsi que de nombreuses séries que l'on pense être potentiellement utiles comme variables indépendantes pour les ventes futures. Le modélisateur expert a déterminé que seules deux des cinq variables indépendantes candidates étaient significatives : le nombre de catalogues expédiés et le nombre de lignes téléphoniques ouvertes pour les commandes.

Lors de la planification de la stratégie des ventes pour l'année suivante, vous avez des ressources limitées pour imprimer les catalogues et garder les lignes téléphoniques ouvertes pour les commandes. Votre budget pour les trois premiers mois de 1999 vous permet soit 2000 catalogues additionnels ou 5 lignes téléphoniques supplémentaires selon vos projections initiales. Quel choix vous permettra de générer plus de revenu pour les ventes sur cette période de trois mois ?

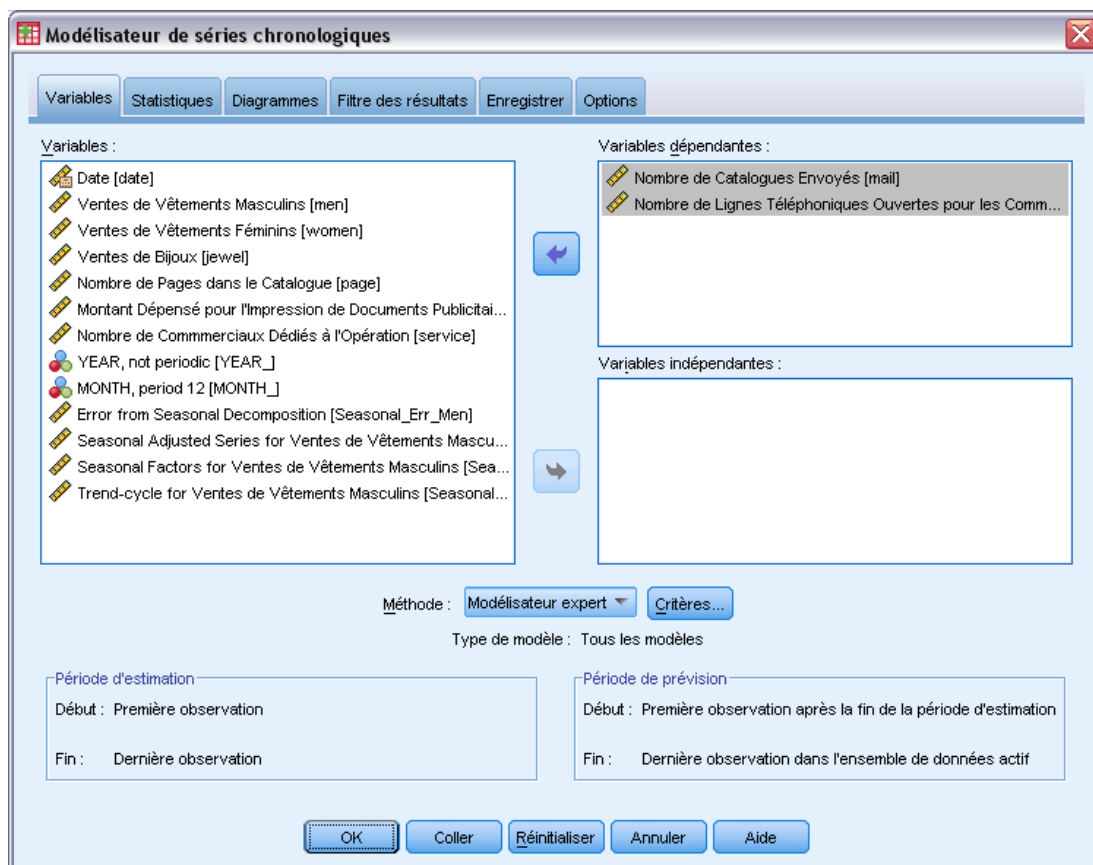
Les données de cet exemple sont rassemblées dans le fichier *catalog_seasfac.sav* et le fichier *catalog_model.xml* contient le modèle des ventes mensuelles construit avec le modélisateur expert. [Pour plus d'informations, reportez-vous à la section Fichiers d'exemple dans l'annexe D dans IBM SPSS Forecasting 21](#). Bien sûr, si vous avez travaillé sur l'exemple précédent et que vous avez enregistré votre propre fichier modèle, vous pouvez utiliser ce dernier au lieu du fichier *catalog_model.xml*.

Extension des séries de variables indépendantes

Lorsque vous créez des prévisions pour les séries de variables dépendantes avec des variables indépendantes, chaque série de variables indépendantes doit être étendue à la période de prévision. A moins de connaître précisément les futures valeurs des variables indépendantes, vous devez les estimer. Vous pouvez ensuite modifier les estimations afin de tester différents scénarios de prévision. Les projections initiales sont facilement créées avec le modélisateur expert.

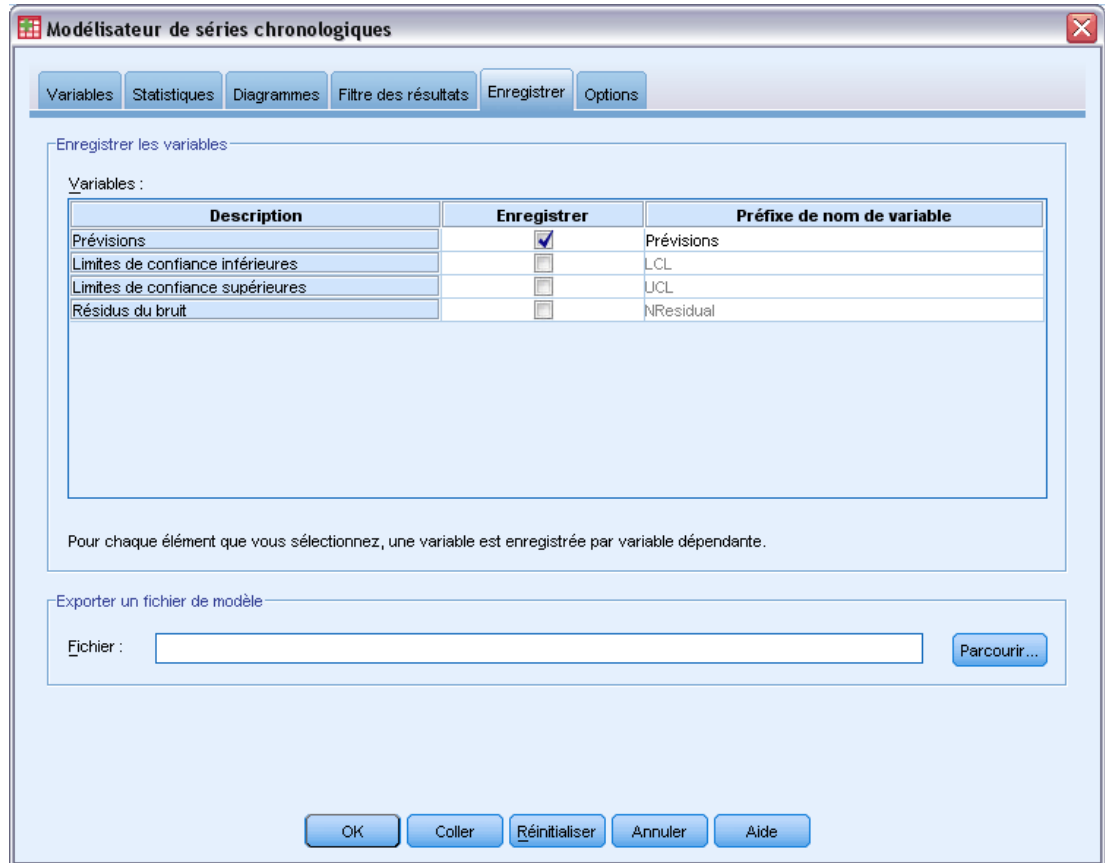
- ▶ A partir des menus, sélectionnez :
Analyse > Prévisions > Créer des modèles...

Figure 9-1
Boîte de dialogue Modélisateur de séries chronologiques



- ▶ Sélectionnez *Nombre de catalogues expédiés* et *Nombre de lignes téléphoniques ouvertes pour les commandes* comme variables dépendantes.
- ▶ Cliquez sur l'onglet *Enregistrer*.

Figure 9-2
Modélisateur de séries chronologiques, onglet Enregistrer



- ▶ Sélectionnez (cochez) l'entrée pour les Prévisions dans la colonne *Enregistrer* et laissez la valeur par défaut *Estimé* comme Préfixe de nom de variable.
- ▶ Cliquez sur l'onglet Options.

Figure 9-3
Modélisateur de séries chronologiques, onglet Options

- Sélectionnez Première observation après la fin de la période d'estimation jusqu'à une date spécifique dans le groupe Période de prévision.

- Dans la grille Date, saisissez 1999 pour l'année et 3 pour le mois.

L'ensemble de données contient des données de janvier 1999 à décembre 1998. Selon les paramètres actuels, la période de prévision sera de janvier 1999 à mars 1999.

- Cliquez sur OK.

Figure 9-4
Nouvelles variables contenant les prévisions pour les séries de variables indépendantes

| | Prévu_mail_Modèle_1 | Prévu_phone_Modèle_2 | |
|-----|---------------------|----------------------|---|
| 121 | 11742 | 51 | ▲ |
| 122 | 11853 | 45 | ■ |
| 123 | 11965 | 45 | ▼ |

L'Editeur de données affiche les nouvelles variables *Estimé_courrier_Modèle_1* et *Estimé_téléphone_Modèle_2* contenant les valeurs de prévision du modèle pour le nombre de catalogues expédiés et le nombre de lignes téléphoniques. Pour étendre nos séries de variables

indépendantes, nous n'avons besoin que des valeurs de janvier 1999 à mars 1999, c'est-à-dire les observations 121 à 123.

- ▶ Copiez les valeurs de ces trois observations à partir de *Estimé_courrier_Modèle_1* and ajoutez-les à la variable *courrier*.
- ▶ Répétez ce procédé pour *Estimé_téléphone_Modèle_2* en copiant les trois dernières observations et en les ajoutant à la variable *téléphone*.

Figure 9-5

Les séries de variables indépendantes s'étendent maintenant à la période de prévision.

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

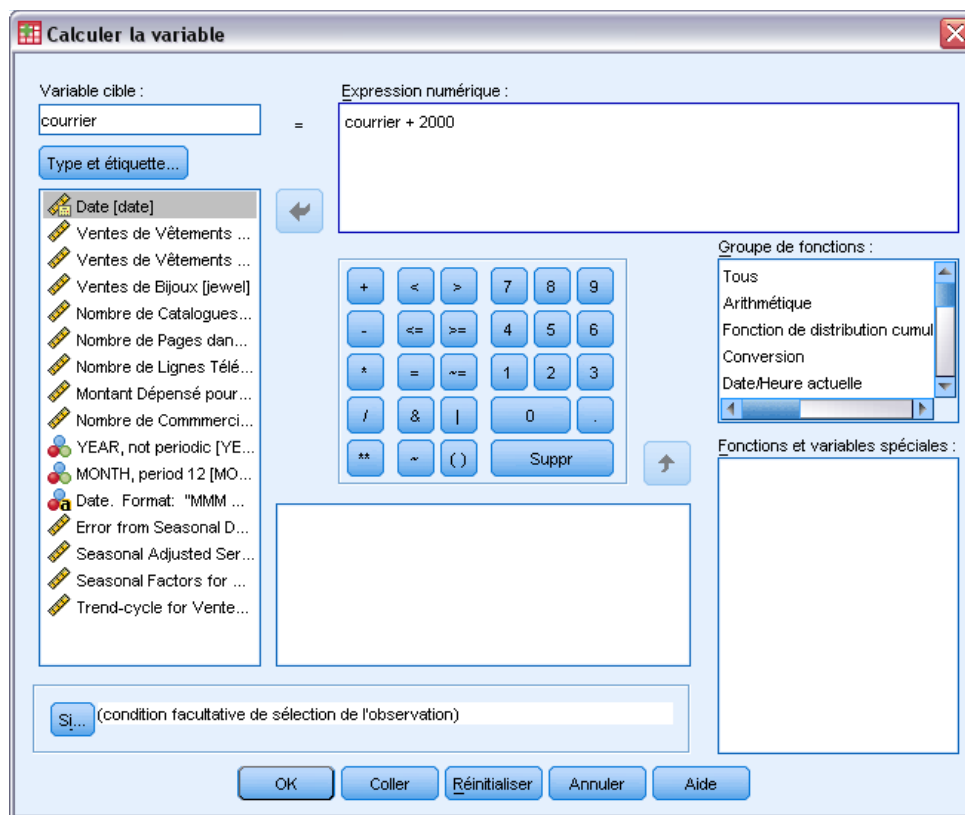
Les variables indépendantes s'étendent maintenant à la période de prévision.

Modification des valeurs de prévision dans la période de prévision

Le test des deux scénarios (expédier plus de catalogues ou ouvrir plus de lignes téléphoniques) nécessite la modification des estimations pour les deux variables indépendantes *courrier* et *téléphone*. Puisque nous modifions uniquement les valeurs prévisionnelles pour trois observations (mois), il serait facile d'entrer directement les nouvelles valeurs dans les cellules appropriées de l'Editeur de données. Mais pour des raisons pédagogiques, nous allons utiliser la boîte de dialogue Calculer la variable. Si vous devez modifier un nombre conséquent de valeurs, cet outil vous semblera plus pratique.

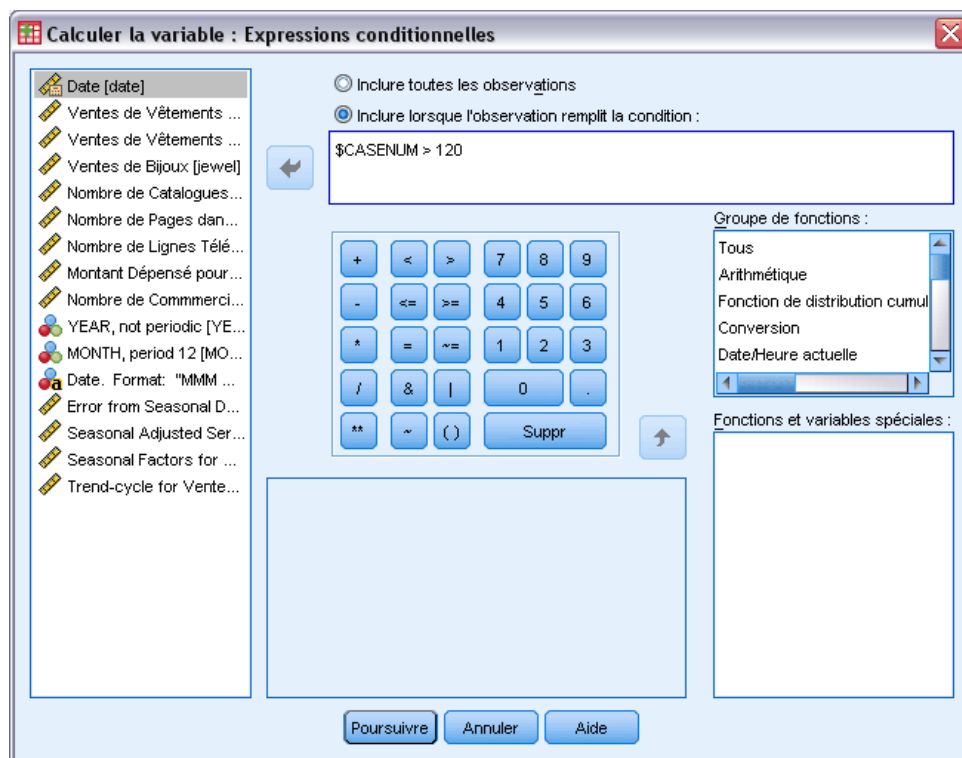
- ▶ A partir des menus, sélectionnez :
Transformer > Calculer la variable...

Figure 9-6
Boîte de dialogue Calculer la variable



- ▶ Entrez courrier pour la variable cible.
- ▶ Dans la zone de texte Expression numérique, saisissez courrier + 2000.
- ▶ Cliquez sur Si.

Figure 9-7
Boîte de dialogue Expression conditionnelle Calculer la variable



- Sélectionnez Inclure si l'observation remplit la condition :

- Dans la zone de texte, saisissez `$CASENUM > 120`.

Cela limitera les modifications à la variable *courrier* pour les observations de la période de prévision.

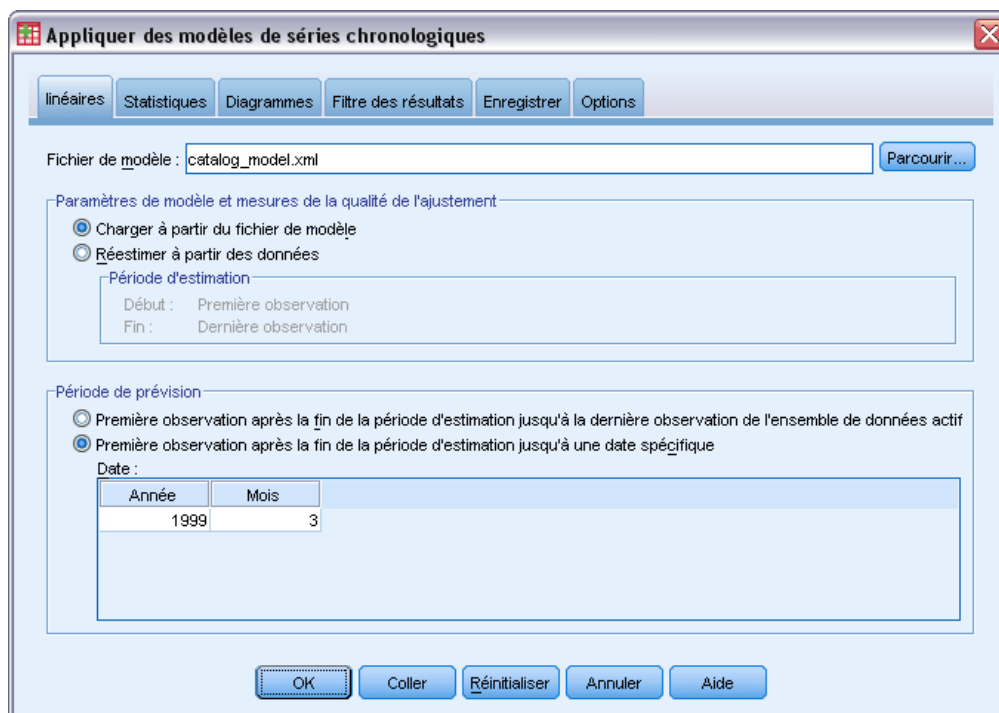
- Cliquez sur Poursuivre.
- Cliquez sur OK dans la boîte de dialogue Calculer la variable puis cliquez sur OK à l'invite pour modifier la variable existante.

Cela augmente les valeurs de la variable *courrier* (le nombre de catalogues expédiés) de 2 000 pour chacun des trois mois de la période d'estimation. Vous avez maintenant préparé les données pour tester le premier scénario et vous êtes prêt à exécuter l'analyse.

Exécution de l'analyse

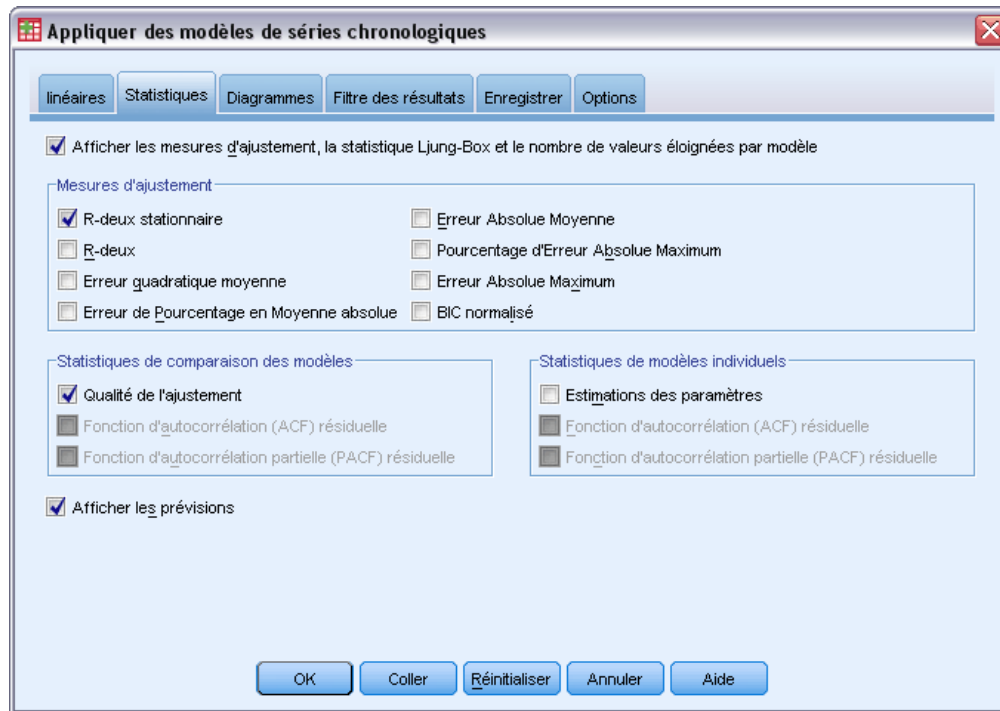
- A partir des menus, sélectionnez :
Analyse > Previsions > Appliquer des modèles...

Figure 9-8
Boîte de dialogue Appliquer des modèles de séries chronologiques



- ▶ Cliquez sur Parcourir, puis recherchez et sélectionnez le fichier *catalog_model.xml*, ou choisissez votre propre fichier de modèle (enregistré dans l'exemple précédent). [Pour plus d'informations, reportez-vous à la section Fichiers d'exemple dans l'annexe D dans IBM SPSS Forecasting 21.](#)
- ▶ Sélectionnez Première observation après la fin de la période d'estimation jusqu'à une date spécifique dans le groupe Période de prévision.
- ▶ Dans la grille Date, saisissez 1999 pour l'année et 3 pour le mois.
- ▶ Cliquez sur l'onglet Statistiques.

Figure 9-9
Appliquer des modèles de séries chronologiques, onglet Statistiques



- Sélectionnez Afficher les prévisions.

Un tableau des valeurs de prévision pour la variable dépendante s'affiche.

- Cliquez sur OK dans la boîte de dialogue Appliquer des modèles de séries chronologiques.

Figure 9-10
Tableau des prévisions

| Modèle | | JAN 1999 | FEV 1999 | MAR 1999 |
|---------------------|-----------|----------|----------|----------|
| Ventes de Vêtements | Prévision | 25213,59 | 21966,89 | 21514,68 |
| Masculins-Model_1 | UCL | 27525,30 | 24278,59 | 23826,38 |
| | LCL | 22901,89 | 19655,18 | 19202,97 |

Ce tableau de prévisions contient les prévisions pour la série de variables dépendantes en prenant en compte les valeurs des deux variables indépendantes *courrier* et *téléphone* pour la période de prévision. Le tableau inclut aussi la limite de confiance supérieure (UCL) et la limite de confiance inférieure (LCL) pour les prévisions.

Vous avez établi les prévisions de ventes pour le scénario consistant à envoyer 2 000 catalogues de plus par mois. Il vous faut maintenant préparer les données pour le scénario consistant à augmenter le nombre de lignes téléphoniques en redéfinissant la variable *courrier* sur sa valeur originale et en augmentant la variable *téléphone* de 5. Vous pouvez rétablir la variable *courrier* en copiant les valeurs de la variable *Estimé_courrier_Modèle_1* dans la période de prévision et en les collant à la place des valeurs actuelles de *courrier* dans la période de prévision. Vous pouvez également augmenter le nombre de lignes téléphoniques (de 5 pour chaque mois de la période de

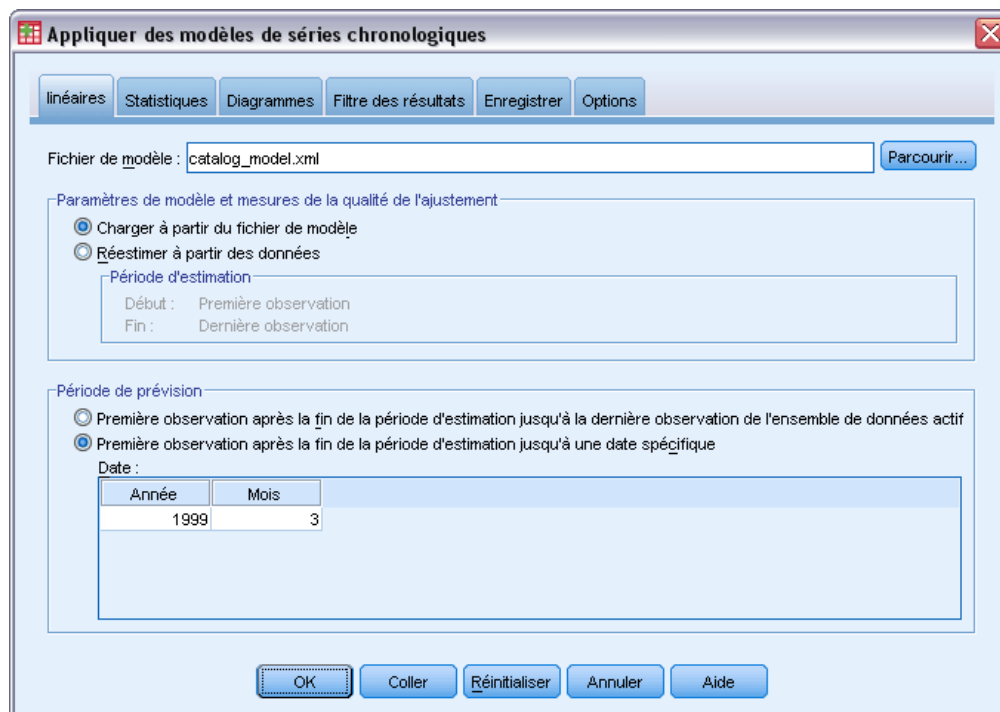
prévision) soit directement dans l'éditeur de données, soit en utilisant la boîte de dialogue Calculer la variable comme pour le nombre de catalogues.

Pour exécuter l'analyse, réouvrez la boîte de dialogue Appliquer des modèles des séries chronologiques comme suit :

- ▶ Cliquez sur l' icône Rappeler boîte de dialogue dans la barre d'outils.
- ▶ Choisissez Appliquer des modèles de séries chronologiques.

Figure 9-11

Boîte de dialogue Appliquer des modèles de séries chronologiques



- ▶ Cliquez sur OK dans la boîte de dialogue Appliquer des modèles de séries chronologiques.

Figure 9-12
Tableaux de prévisions pour les deux scénarios

| Prévision avec plus catalogues | | | | |
|---------------------------------------|-----------|----------|----------|----------|
| Modèle | | JAN 1999 | FEV 1999 | MAR 1999 |
| Ventes de Vêtements | Prévision | 25279,91 | 22064,72 | 21580,96 |
| Masculins-Model_1 | UCL | 27591,62 | 24376,42 | 23892,66 |
| | LCL | 22968,21 | 19753,02 | 19269,25 |

| Prévision avec plus lignes téléphoniques | | | | |
|---|-----------|----------|----------|----------|
| Modèle | | JAN 1999 | FEV 1999 | MAR 1999 |
| Ventes de Vêtements | Prévision | 23757,25 | 20542,06 | 20058,29 |
| Masculins-Model_1 | UCL | 26068,95 | 22853,76 | 22370,00 |
| | LCL | 21445,55 | 18230,35 | 17746,59 |

En affichant les tableaux de prévisions pour les deux scénarios, on s'aperçoit que pour chacun des trois mois de la période de prévision, l'augmentation du nombre de catalogues expédiés générera environ 1 500 US\$ de ventes en plus que l'augmentation du nombre de lignes téléphoniques ouvertes pour les commandes. En se basant sur cette analyse, il semble plus sage d'allouer des ressources pour l'envoi de 2000 catalogues supplémentaires.

Désaisonnalisation

Suppression de la saisonnalité dans les données de vente

Une société de vente par correspondance est intéressée par la modélisation de la tendance à la hausse des ventes de sa ligne de vêtements pour homme en fonction d'un ensemble de variables prédites, comme le nombre de catalogues envoyés et le nombre de lignes de téléphone ouvertes pour l'enregistrement des commandes. Pour cela, cette société a rassemblé les ventes mensuelles de vêtements pour hommes sur une période de 10 ans. Ces informations sont regroupées dans le fichier *catalog.sav*. [Pour plus d'informations, reportez-vous à la section Fichiers d'exemple dans l'annexe D dans IBM SPSS Forecasting 21.](#)

Pour réaliser une analyse de tendance, vous devez supprimer toute variation saisonnière présente dans les données. Cette tâche est facilement réalisable grâce à la procédure Désaisonnalisation.

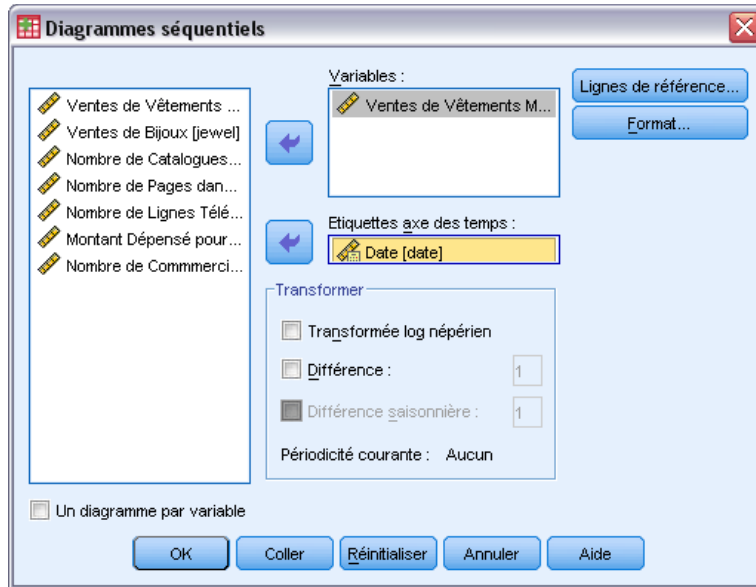
Détermination et paramétrage de la périodicité

La procédure Désaisonnalisation requiert la présence d'une composante de date périodique dans l'ensemble de données actif. Par exemple, une périodicité annuelle de 12 (mois), une périodicité hebdomadaire de 7 (jours), etc. Il est conseillé de d'abord tracer la série chronologique, car l'apparence du tracé d'une série chronologique donne généralement une bonne idée de la périodicité sous-jacente.

Pour représenter les ventes de vêtements pour hommes dans le temps :

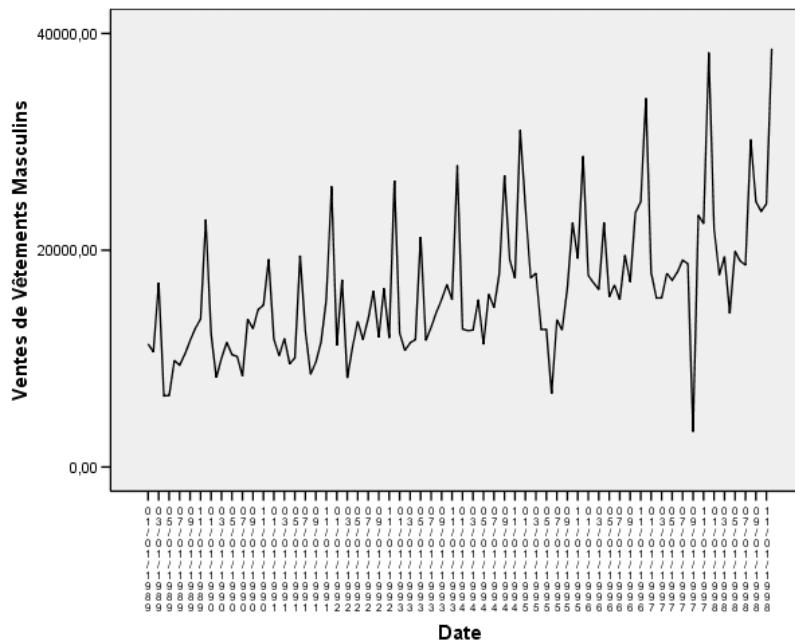
- ▶ A partir des menus, sélectionnez :
Analyse > Prévisions > Diagrammes séquentiels...

Figure 10-1
Boîte de dialogue Diagrammes séquentiels



- ▶ Sélectionnez la variable *Ventes de vêtements hommes* et placez-la dans la liste Variables.
- ▶ Sélectionnez la variable *Date* et placez-la dans la liste Etiquettes axe des temps.
- ▶ Cliquez sur OK.

Figure 10-2
Ventes de vêtements pour hommes (en dollars U.S.)

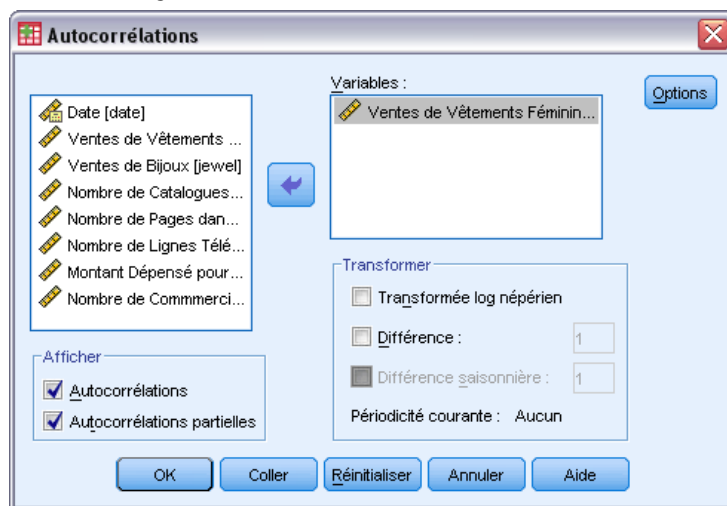


La série comporte un certain nombre de projections, mais elles ne sont pas régulièrement espacées. Ce résultat suggère que, si la série comporte une composante périodique, elle comporte également des fluctuations non périodiques, ce qui est classique pour les séries en temps réel. A l'exception des fluctuations à petite échelle, les projections significatives sont séparées de plus de quelques mois. Etant donné le caractère saisonnier des ventes, avec des hauts pendant la période des congés de fin d'année, il est probable que la série chronologique ait une périodicité annuelle. Notez également que les variations saisonnières croissent en même temps que la tendance à la hausse de la série, ce qui laisse supposer que les variations saisonnières sont proportionnelles au niveau de la série, ce qui implique un modèle multiplicatif plutôt qu'additif.

L'examen des autocorrélations et des autocorrélations partielles d'une série chronologique fournit une conclusion plus quantitative à propos de la périodicité sous-jacente.

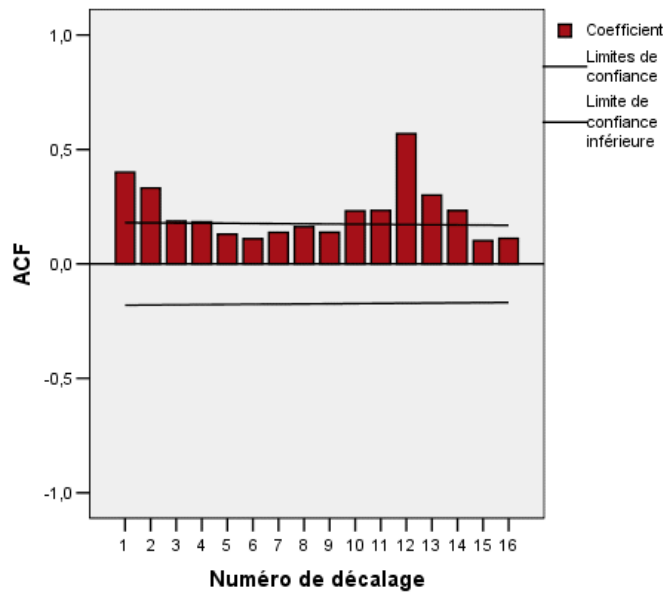
- A partir des menus, sélectionnez :
Analyse > Prévisions > Autocorrélations...

Figure 10-3
Boîte de dialogue Autocorrélations



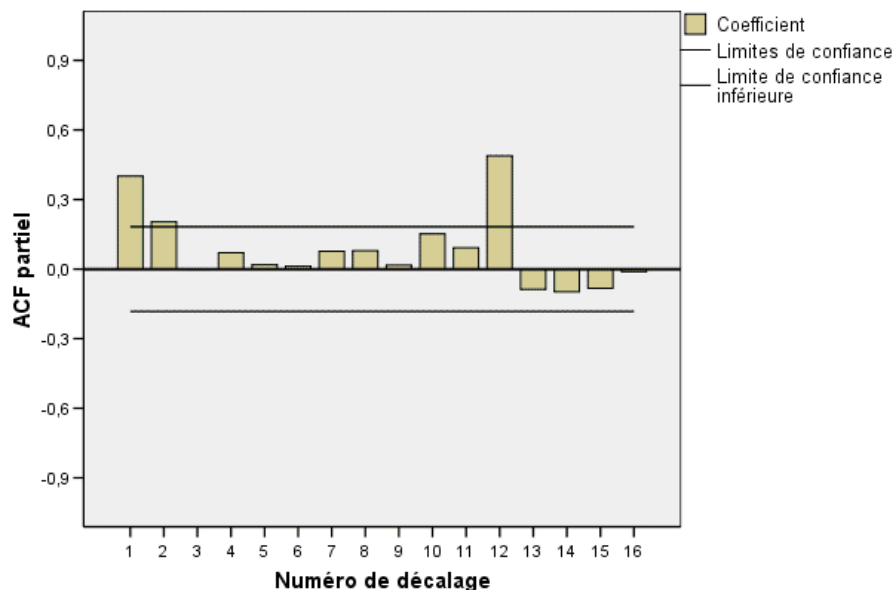
- Sélectionnez la variable *Ventes de vêtements hommes* et placez-la dans la liste Variables.
- Cliquez sur OK.

Figure 10-4
Autocorrélogramme des vêtements pour hommes



La fonction d'autocorrélation présente une projection significative au niveau du retard 1 avec une longue extrémité exponentielle—ce qui est classique pour les séries chronologiques. La projection significative au niveau du retard 12 suggère la présence d'une composante saisonnière annuelle dans les données. L'examen de la fonction d'autocorrélation partielle offre une conclusion plus définitive.

Figure 10-5
Autocorrélogramme partiel des vêtements pour homme

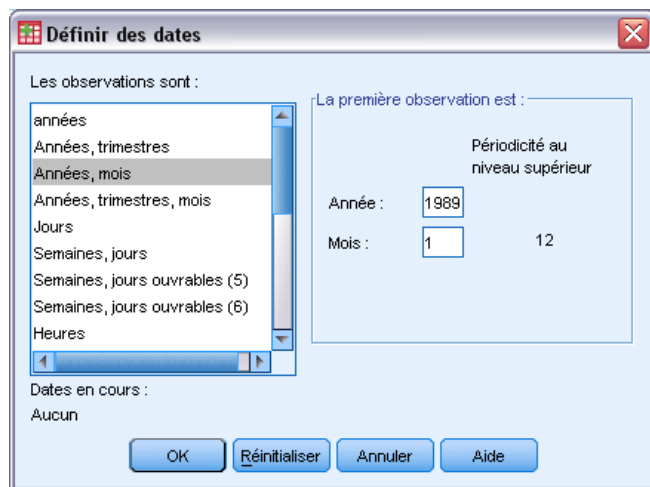


La projection significative au niveau du retard 12 dans la fonction d'autocorrélation partielle confirme la présence d'une composante saisonnière annuelle dans les données.

Pour définir une périodicité annuelle :

- A partir des menus, sélectionnez :
Données > Définir des dates

Figure 10-6
Boîte de dialogue Définir des dates



- Sélectionnez Années, mois dans la liste Les observations sont.
- Entrez 1989 pour l'année et 1 pour le mois.

- ▶ Cliquez sur OK.

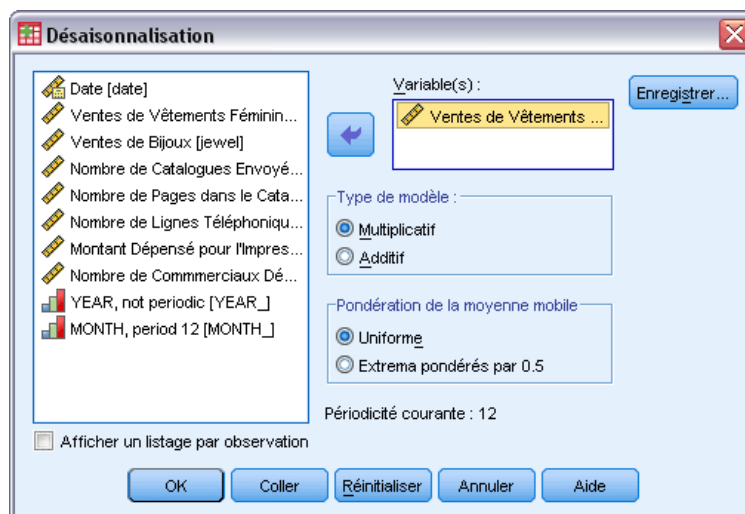
Cela définit la périodicité sur 12 et crée un ensemble de variables de date conçues pour être utilisées avec les procédures Prévisions.

Exécution de l'analyse

Pour exécuter la procédure Désaisonnalisation :

- ▶ A partir des menus, sélectionnez :
Analyse > Prévisions > Désaisonnalisation...

Figure 10-7
Boîte de dialogue Désaisonnalisation



- ▶ Cliquez avec le bouton droit de la souris n'importe où dans la liste des variables source et, à partir du menu contextuel, sélectionnez Afficher les noms des variables.
- ▶ Sélectionnez la variable *hommes* et placez-la dans la liste Variables.
- ▶ Sélectionnez Multiplicatif dans le groupe Type de modèle.
- ▶ Cliquez sur OK.

Présentation des résultats

La procédure Désaisonnalisation crée quatre nouvelles variables pour chacune des variables d'origine analysées par la procédure. Par défaut, les nouvelles variables sont ajoutées au fichier de travail. Les nouvelles séries portent des noms commençant par les préfixes suivants :

SAF. (Pour Seasonal Adjustment Factors) Composantes de désaisonnalisation, représentant la variation saisonnière. Pour le modèle multiplicatif, la valeur 1 représente l'absence de variation saisonnière ; pour le modèle additif, il s'agit de la valeur 0.

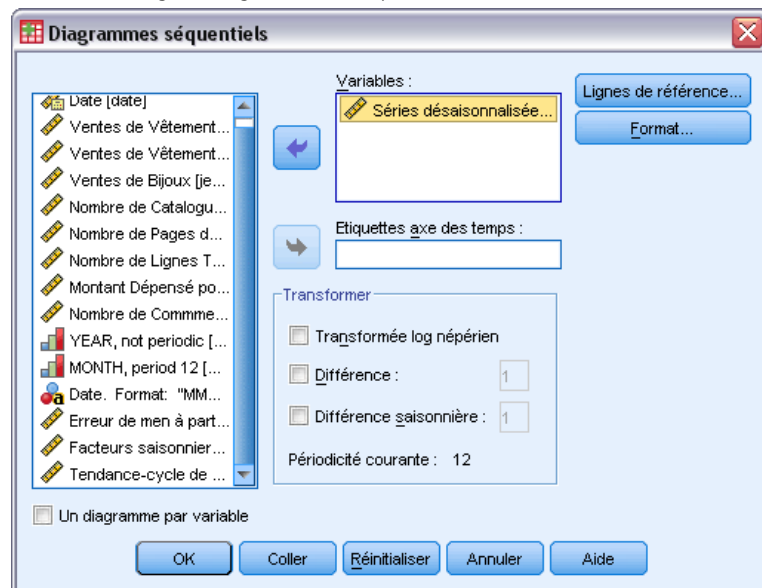
SAS. (Pour Seasonally Adjusted Series) Séries désaisonnalisées, représentant les séries d'origine sans les variations saisonnières. Une série désaisonnée permet par exemple d'isoler et d'analyser une composante de tendance, indépendamment des composantes saisonnières.

STC. (Pour Smoothed Trend-cycle Component) Composante tendance-cycle lissée, une version lissée de la série désaisonnée, représentant la tendance ainsi que les composantes cycliques.

ERR. Composante résiduelle des séries pour une observation particulière.

Pour l'observation qui nous concerne, c'est la série désaisonnée qui est la plus adaptée car elle représente la série d'origine sans les variations saisonnières.

Figure 10-8
Boîte de dialogue Diagrammes séquentiels

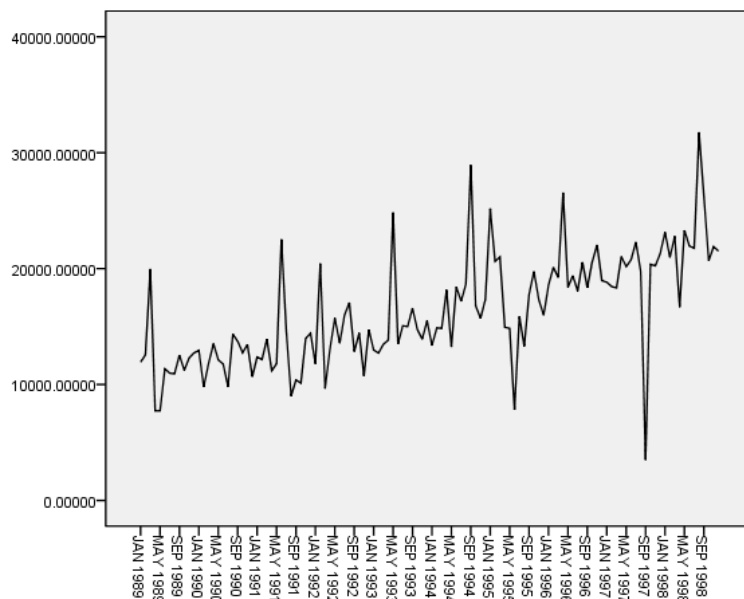


Pour tracer les séries désaisonnées :

- ▶ Ouvrez la boîte de dialogue Diagrammes séquentiels.
- ▶ Cliquez sur Restaurer pour effacer toute sélection antérieure.
- ▶ Cliquez avec le bouton droit de la souris n'importe où dans la liste des variables source et, à partir du menu contextuel, sélectionnez Afficher les noms des variables.
- ▶ Sélectionnez la variable `SAS_I` et placez-la dans la liste Variables.

► Cliquez sur OK.

Figure 10-9
Séries désaisonnalisées



La série désaisonnalisée montre une tendance à la hausse prononcée. De nombreuses projections sont significatives, mais elles se produisent à des intervalles non réguliers, ce qui ne permet pas de dégager un motif annuel.

Récapitulatif

Grâce à la procédure Désaisonnalisation, vous avez supprimé la composante saisonnière d'une série chronologique périodique afin d'obtenir une série plus adaptée à l'analyse de tendances. L'examen des autocorrélations et des autocorrélations partielles de la série chronologique a déterminé avec efficacité la périodicité sous-jacente de cette observation, qui s'est avérée annuelle.

Procédures apparentées

La procédure Désaisonnalisation est utile pour supprimer une composante saisonnière unique d'une série chronologique périodique.

- Pour réaliser une analyse de la périodicité d'une série chronologique plus approfondie que celle fournie par la fonction de corrélation partielle, utilisez la procédure Analyse spectrale. Pour plus d'informations, consultez [le chapitre 11](#).

Analyse spectrale

Utilisation de la procédure Analyse spectrale pour vérifier les prévisions en matière de périodicité

Les séries chronologiques qui représentent des ventes au détail ont généralement une périodicité annuelle sous-jacente en raison des fortes ventes pendant les vacances. Générer des prévisions de vente revient à créer un modèle pour les séries chronologiques, c'est-à-dire à identifier les composantes périodiques. Un diagramme de séries chronologiques ne peut pas toujours permettre de découvrir la périodicité annuelle car les séries chronologiques contiennent des fluctuations aléatoires qui masquent souvent la structure sous-jacente.

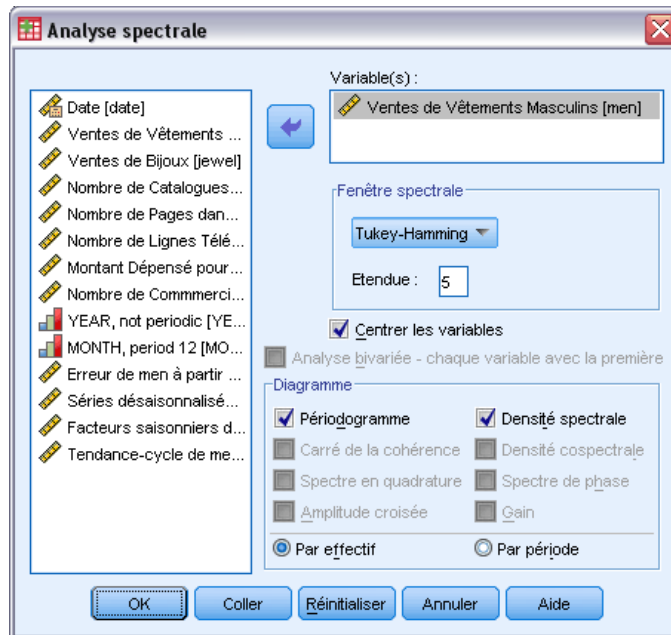
Les données relatives aux ventes mensuelles d'une société de vente par correspondance sont stockées dans le fichier *catalog.sav*. [Pour plus d'informations, reportez-vous à la section Fichiers d'exemple dans l'annexe D dans IBM SPSS Forecasting 21](#). Avant de passer à la prévision des ventes, vous souhaitez vous assurer que les données relatives aux ventes montrent une périodicité annuelle. Un diagramme des séries chronologiques indique de nombreux pics et des intervalles irréguliers, de telle sorte qu'aucune périodicité sous-jacente n'est évidente. Utilisez la procédure Analyse spectrale pour identifier une périodicité dans les données relatives aux ventes.

Exécution de l'analyse

Pour exécuter la procédure Analyse spectrale :

- ▶ A partir des menus, sélectionnez :
Analyse > Prévisions > Analyse spectrale...

Figure 11-1
Boîte de dialogue Analyse spectrale



- ▶ Sélectionnez la variable *Ventes de vêtements hommes* et placez-la dans la liste Variables.
- ▶ Sélectionnez Densité spectrale dans le groupe Diagramme.
- ▶ Cliquez sur OK.

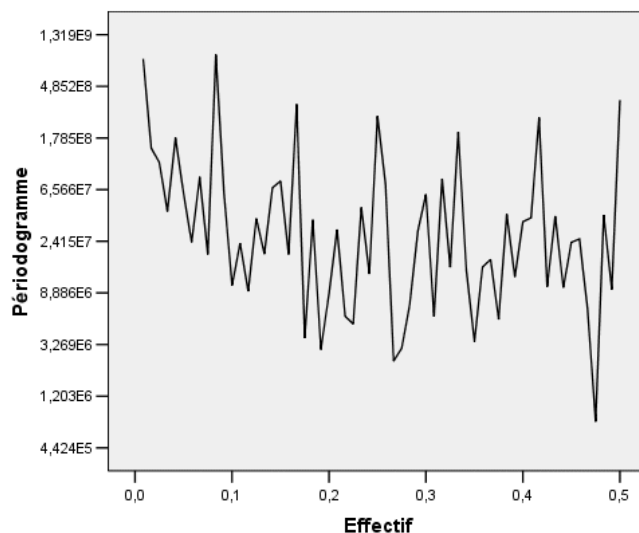
Ces sélections génèrent la syntaxe de commande suivante :

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

Veuillez noter que pour obtenir le tableau des statistiques univariées dans les résultats, la commande TSET doit être modifiée pour lire TSET PRINT=DETAILED.

Présentation du périodogramme et de la densité spectrale

Figure 11-2
Périodogramme



Le diagramme du périodogramme indique une séquence de pics qui ressortent du bruit de fond, le pic de fréquence le moins élevé se situant à une fréquence légèrement inférieure à 0,1. Vous pensez que les données contiennent une composante périodique annuelle. Par conséquent, considérez la contribution d'une composante annuelle au périodogramme. Chaque point de données de la série chronologique représentant un mois, une périodicité annuelle correspond à une période de 12 dans l'ensemble de données en cours. Etant donné que la période et la fréquence sont réciproques, une période de 12 correspond à une fréquence de $1/12$ (ou 0,083). Une composante annuelle suppose donc un pic dans le périodogramme à une fréquence de 0,083, ce qui semble conforme à la présence du pic légèrement en dessous d'une fréquence de 0,1.

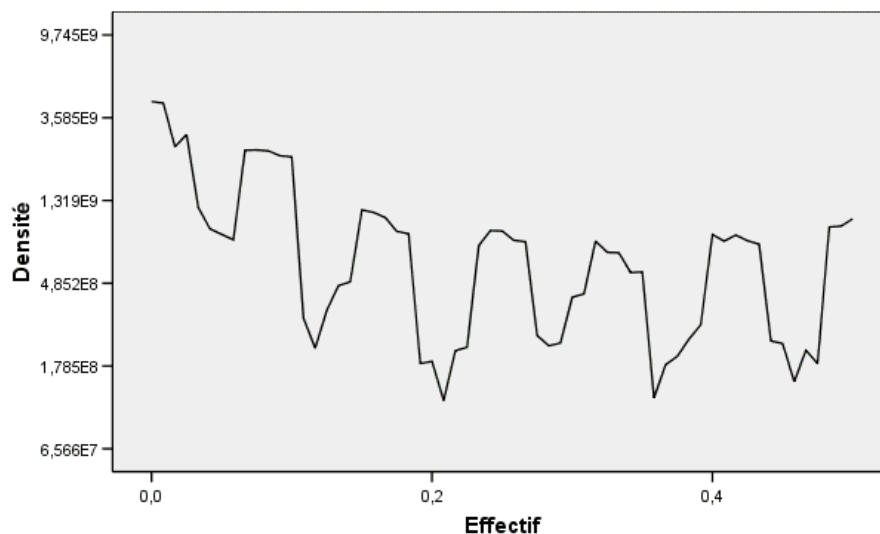
Figure 11-3
Tableau des statistiques univariées

Nom de série : hommes

| | Fréquence | Transformation sinus | Transformation cosinus | Périodogramme | Estimation de la densité spectrale |
|----|-----------|----------------------|------------------------|---------------|------------------------------------|
| 1 | .00000 | .000 | 16242.813 | .000 | 4.361E+09 |
| 2 | .00833 | -3696.643 | 370.153 | 828131182 | 4.278E+09 |
| 3 | .01667 | -1496.277 | 454.831 | 146743038 | 2.526E+09 |
| 4 | .02500 | -1336.400 | 252.087 | 110970821 | 2.921E+09 |
| 5 | .03333 | -662.146 | 529.734 | 43143315.6 | 1.210E+09 |
| 6 | .04167 | -1654.614 | 501.571 | 179359225 | 935924937 |
| 7 | .05000 | -784.814 | -636.729 | 61281367.3 | 875492029 |
| 8 | .05833 | -335.646 | 532.062 | 23744855.0 | 820375352 |
| 9 | .06667 | -1094.178 | -451.489 | 84064108.9 | 2.420E+09 |
| 10 | .07500 | 264.554 | 492.876 | 18774933.5 | 2.429E+09 |
| 11 | .08333 | -3053.934 | 2370.483 | 896742149 | 2.401E+09 |
| 12 | .09167 | -978.882 | -287.035 | 62435897.6 | 2.263E+09 |
| 13 | .10000 | -403.128 | 93.036 | 10270064.4 | 2.235E+09 |

Le tableau des statistiques univariées contient les points de données utilisés pour tracer le périodogramme. Pour les fréquences inférieures à 0,1, la valeur la plus élevée de la colonne *Périodogramme* correspond à une fréquence de 0,08333 (précisément la valeur que vous attendez pour savoir s'il existe une composante périodique annuelle). Ceci confirme l'identification du pic de fréquence le moins élevé avec une composante périodique annuelle. Mais qu'en est-il des autres pics à des fréquences plus élevées ?

Figure 11-4
Densité spectrale



Les autres pics sont mieux analysés avec la fonction de densité spectrale, qui est une version lissée du périodogramme. Le lissage permet d'éliminer le bruit de fond d'un périodogramme et d'isoler plus nettement la structure sous-jacente.

La densité spectrale se compose de cinq pics distincts qui apparaissent à intervalles réguliers. Le pic de fréquence le moins élevé représente simplement la version lissée du pic à une fréquence de 0,08333. Pour comprendre la signification des quatre pics de fréquence les plus élevés, rappelez-vous que le périodogramme est calculé en modélisant la série chronologique en tant que somme des fonctions de cosinus et de sinus. Les composantes périodiques ayant la forme d'une fonction de sinus ou de cosinus (composantes sinusoïdales) apparaissent dans le périodogramme sous la forme de pics uniques. Les composantes périodiques qui ne sont pas sinusoïdales apparaissent sous la forme d'une série de pics (espacés de manière égale) de hauteurs différentes, le pic de fréquence le moins élevé de la série correspondant à la fréquence de la composante périodique. Les quatre pics de fréquence les plus élevés de la densité spectrale nous indiquent donc simplement que la composante périodique annuelle n'est pas sinusoïdale.

Vous venez de représenter l'ensemble de la structure visible dans le diagramme de densité spectrale et vous en concluez que les données contiennent une seule composante périodique (période de 12 mois).

Récapitulatif

A l'aide de la procédure Analyse spectrale, vous avez confirmé l'existence d'une composante périodique annuelle pour une série chronologique et vérifié l'absence de toute autre périodicité significative. La densité spectrale s'est avérée plus utile que le périodogramme pour découvrir la structure sous-jacente, car elle lisse les fluctuations générées par la composante non périodique des données.

Procédures apparentées

La procédure Analyse spectrale permet d'identifier les composantes périodiques d'une série chronologique.

- Pour supprimer une composante périodique d'une série chronologique —par exemple, pour effectuer une analyse de tendances—, utilisez la procédure Désaisonnalisation. Consultez [le chapitre 10](#) pour plus de détails.

Mesures de la qualité d'ajustement

Cette section fournit les définitions des mesures de la qualité d'ajustement utilisées pour la modélisation de séries chronologiques.

- **R-deux stationnaire.** Une mesure qui compare la partie stationnaire du modèle à un simple modèle de moyenne. Cette mesure est préférable à un R-deux ordinaire lorsqu'il y a une tendance ou un motif saisonnier. Le R-deux stationnaire peut être négatif avec une plage d'infinité négative de 1. Les valeurs négatives signifient que la sous-considération du modèle est pire que le modèle de la ligne de base. Les valeurs positives signifient que la sous-considération du modèle est meilleure que le modèle de la ligne de base.
- **R•deux.** Une estimation de la proportion de la variabilité totale de la série expliquée par le modèle. Cette mesure est très utile lorsque la série est stationnaire. Le R-deux peut être négatif avec une plage d'infinité négative de 1. Les valeurs négatives signifient que la sous-considération du modèle est pire que le modèle de la ligne de base. Les valeurs positives signifient que la sous-considération du modèle est meilleure que le modèle de la ligne de base.
- **RMSE.** Erreur quadratique moyenne. La racine carrée de l'erreur quadratique moyenne. Une mesure de la variation de la série dépendante par rapport au niveau de prédiction, exprimée dans les mêmes unités que la série dépendante.
- **MAPE.** Erreur de pourcentage Moyenne absolue. Une mesure de la variation de la série dépendante par rapport au niveau prévu par le modèle. Elle est indépendante des mesures utilisées et peut donc être utilisée pour comparer les séries comportant des unités différentes.
- **Erreur absolue moyenne.** Erreur Moyenne absolue. Mesure la variation de la série par rapport au niveau prévu par le modèle. La MAE est reportée dans les unités de séries d'origine.
- **MaxAE.** Erreur de pourcentage absolu maximale. La plus grande erreur prévue, exprimée en pourcentage. Cette mesure est utile pour imaginer le pire scénario lors de vos prévisions.
- **MaxAE.** Erreur absolue maximale. La plus grande erreur prévue, exprimée dans les mêmes unités que la série dépendante. Comme MaxAPE, elle est utile pour imaginer le pire scénario lors de vos prévisions. L'erreur Maximum absolue et l'erreur de pourcentage Maximum absolue peuvent se produire à différents moments d'une série, par exemple lorsque l'erreur absolue pour la valeur d'une grande série est légèrement supérieure à l'erreur absolue pour la valeur d'une petite série. Dans ce cas, l'erreur maximum absolue se produira à la valeur de la grande série et l'erreur de pourcentage maximum absolue se produira à la valeur de la petite série.
- **Critère d'information bayésien normalisé (BIC).** Critère d'information bayésien normalisé. Une mesure générale de l'ajustement global d'un modèle qui essaye de prendre en compte la complexité du modèle. C'est un résultat basé sur l'erreur quadratique moyenne et qui inclut une pénalité pour le nombre de paramètres du modèle et la longueur de la série. La pénalité supprime l'avantage des modèles disposant de plus de paramètres, rendant les statistiques plus faciles à comparer parmi les différents modèles d'une même série.

Types de valeurs éloignées

Cette section fournit les définitions des types de valeurs éloignées utilisées pour la modélisation de séries chronologiques.

- **Additif.** Une valeur éloignée qui affecte une observation unique. Par exemple, une erreur de codage des données peut être identifiée comme une valeur éloignée additive.
- **Changement de niveau.** Une valeur éloignée qui remplace toutes les observations par une constante en commençant à un point particulier de la série. Un changement de politique peut entraîner un changement de niveau.
- **Innovationnel.** Une valeur éloignée qui agit en tant qu'addition au terme de bruit à un point particulier de la série. Pour les séries stationnaires, une valeur éloignée d'innovation affecte plusieurs observations. Pour les séries non stationnaires, elle peut affecter chaque observation à partir d'un point particulier de la série.
- **Transitoire.** Un point aberrant dont l'impact décroît de façon exponentielle pour atteindre 0.
- **Additif saisonnier.** Une valeur éloignée qui affecte une observation en particulier et toutes les observations suivantes séparées par une ou plusieurs périodes saisonnières. Toutes ces observations sont affectées de la même façon. Une valeur éloignée additive saisonnière peut se produire si, au début d'une année donnée, les ventes sont supérieures chaque mois de janvier.
- **Tendance locale.** Point aberrant qui apparaît en début de tendance locale à un point spécifique de la série.
- **Correctif additif.** Groupe d'au moins deux valeurs éloignées additives consécutives. Sélectionner le type de valeur éloignée permet de détecter des valeurs éloignées additives individuelles ainsi que les correctifs.



Guide relatif aux diagrammes ACF/PACF

Les diagrammes présentés ici se rapportent à des processus ARIMA réels ou théoriques. Voici quelques règles générales permettant d'identifier le processus :

- Une série non stationnaire a un ACF qui reste significatif pour au moins une demi-douzaine de retards, plutôt que de décliner rapidement vers 0. Vous devez différencier une telle série jusqu'à ce qu'elle devienne stationnaire avant de pouvoir identifier le processus.
- Les processus autorégressifs disposent d'une fonction d'autocorrélation (ACF) qui diminue de manière exponentielle et de projections dans les premiers retards de la fonction d'autocorrélation partielle (PACF). Le nombre de projections indique l'ordre de l'autorégression.
- Les processus moyenne mobile disposent de projections dans les premiers retards de la fonction d'autocorrélation et d'une fonction d'autocorrélation partielle qui diminue de manière exponentielle. Le nombre de projections indique l'ordre de la moyenne mobile.
- Les processus mixtes (ARMA) indiquent généralement des baisses exponentielles dans la fonction d'autocorrélation et dans la fonction d'autocorrélation partielle.

Lors de l'étape d'identification, vous pouvez ignorer le signe de l'ACF ou de la PACF, ou la vitesse avec laquelle une ACF ou PACF qui baisse de manière exponentielle se rapproche de zéro. Ceci dépend du signe et de la valeur réelle des coefficients AR et MA. Dans certains exemples, une ACF qui baisse de manière exponentielle adopte tour à tour des valeurs positives et négatives.

Les diagrammes ACF et PACF issus de données réelles ne sont jamais aussi nets que ceux montrés ici. Vous devez apprendre à identifier les informations essentielles de tout diagramme. Vérifiez toujours l'ACF et la PACF des résidus, au cas où votre identification serait incorrecte. Tenez compte des points suivants :

- Les processus saisonniers affichent ces motifs au niveau des retards saisonniers (multiples de la période saisonnière).
- Vous êtes autorisé à traiter les valeurs non significatives en tant que 0. C'est-à-dire que vous pouvez ignorer les valeurs se situant dans les intervalles de confiance sur les diagrammes. Cependant, vous ne devez pas les ignorer si elles poursuivent le motif des valeurs significatives au niveau statistique.
- Une autocorrélation intermittente sera significative au niveau statistique uniquement par hasard. Vous pouvez ignorer une autocorrélation statistiquement significative si elle est isolée, de préférence au niveau d'un retard élevé, et si elle ne se produit pas au niveau d'un retard saisonnier.

Pour plus d'informations sur les diagrammes ACF et PACF, consultez les textes relatifs à l'analyse ARIMA.

Table C-1
 $ARIMA(0,0,1), q>0$

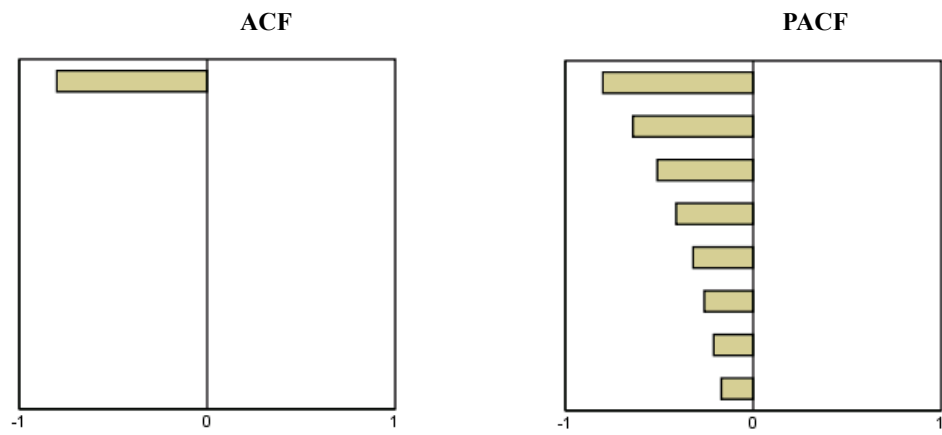
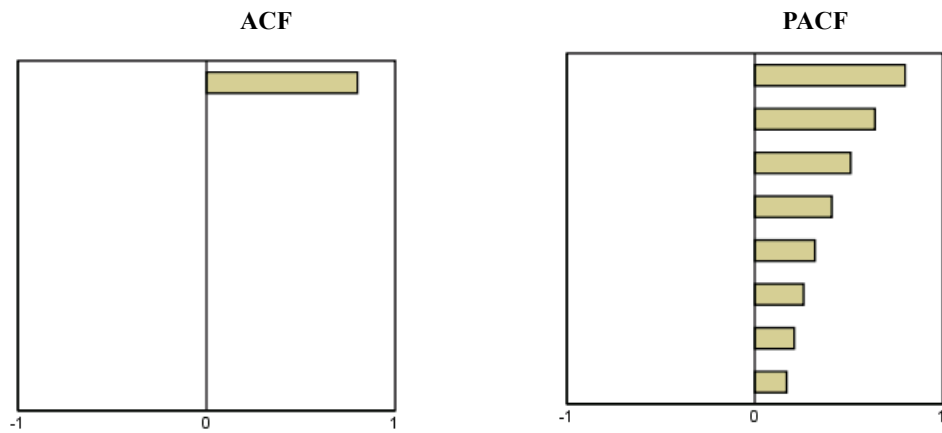


Table C-2
 $ARIMA(0,0,1), q<0$



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

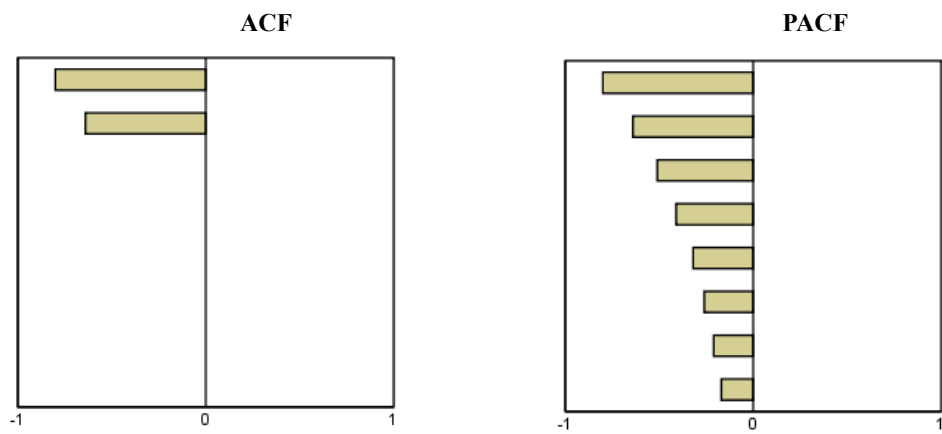


Table C-3
 $ARIMA(1,0,0), f > 0$

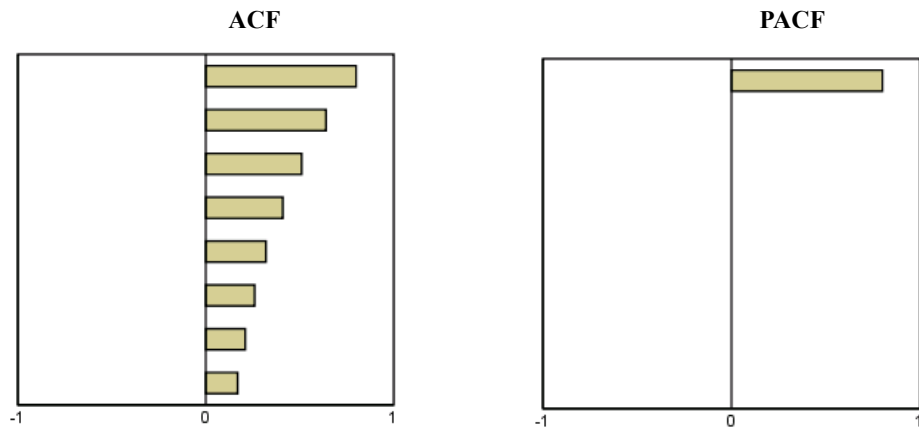
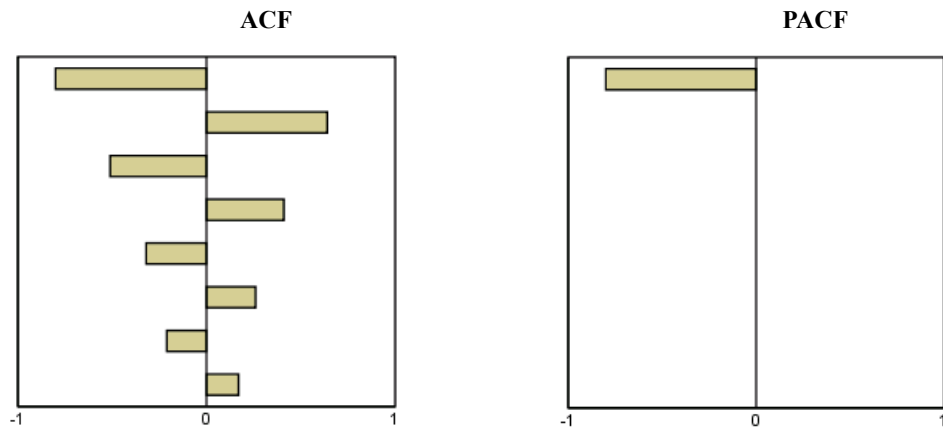
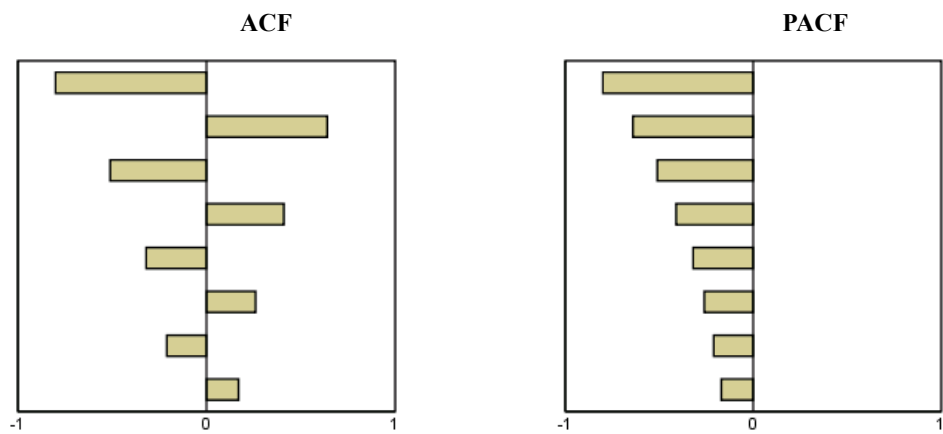


Table C-4
 $ARIMA(1,0,0), f < 0$



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



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

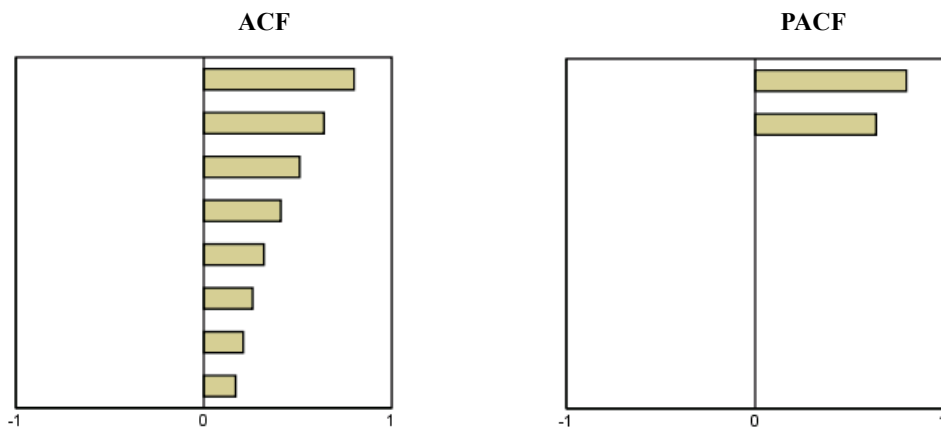
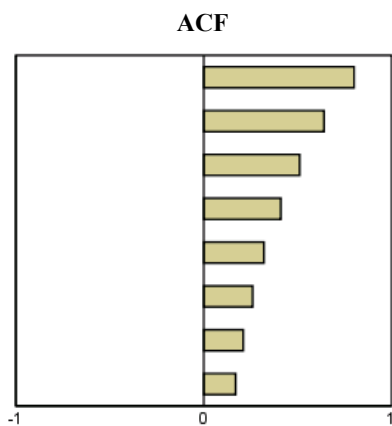


Table C-5
ARIMA(0,1,0) (séries intégrées)



Fichiers d'exemple

Les fichiers d'exemple installés avec le produit figurent dans le sous-répertoire *Echantillons* du répertoire d'installation. Il existe un dossier distinct au sein du sous-répertoire *Echantillons* pour chacune des langues suivantes : Anglais, Français, Allemand, Italien, Japonais, Coréen, Polonais, Russe, Chinois simplifié, Espagnol et Chinois traditionnel.

Seuls quelques fichiers d'exemples sont disponibles dans toutes les langues. Si un fichier d'exemple n'est pas disponible dans une langue, le dossier de langue contient la version anglaise du fichier d'exemple.

Descriptions

Voici de brèves descriptions des fichiers d'exemple utilisés dans divers exemples à travers la documentation.

- **accidents.sav.** Ce fichier de données d'hypothèse concerne une société d'assurance qui étudie les facteurs de risque liés à l'âge et au sexe dans les accidents de la route survenant dans une région donnée. Chaque observation correspond à une classification croisée de la catégorie d'âge et du sexe.
- **adl.sav.** Ce fichier de données d'hypothèse concerne les mesures entreprises pour identifier les avantages d'un type de thérapie proposé aux patients qui ont subi une attaque cardiaque. Les médecins ont assigné de manière aléatoire les patients du sexe féminin ayant subi une attaque cardiaque à un groupe parmi deux groupes possibles. Le premier groupe a fait l'objet de la thérapie standard tandis que le second a bénéficié en plus d'une thérapie émotionnelle. Trois mois après les traitements, les capacités de chaque patient à effectuer les tâches ordinaires de la vie quotidienne ont été notées en tant que variables ordinales.
- **advert.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un détaillant pour examiner la relation existant entre l'argent dépensé dans la publicité et les ventes résultantes. Pour ce faire, il collecte les chiffres des ventes passées et les coûts associés à la publicité.
- **aflatoxin.sav.** Ce fichier de données d'hypothèse concerne le test de l'aflatoxine dans des récoltes de maïs. La concentration de ce poison varie largement d'une récolte à l'autre et au sein de chaque récolte. Un processeur de grain a reçu 16 échantillons issus de 8 récoltes de maïs et a mesuré les niveaux d'aflatoxine en parties par milliard (PPB).
- **anorectic.sav.** En cherchant à développer une symptomatologie standardisée du comportement anorexique/boulimique, des chercheurs (Van der Ham, Meulman, Van Strien, et Van Engeland, 1997) ont examiné 55 adolescents souffrant de troubles alimentaires. Chaque patient a été observé quatre fois sur une période de quatre années, soit un total de 220 observations. A chaque observation, les patients ont été notés pour chacun des 16 symptômes. En raison de l'absence de scores de symptôme pour le patient 71/visite 2, le patient 76/visite 2 et le patient 47/visite 3, le nombre d'observations valides est de 217.

- **bankloan.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une banque pour réduire le taux de défaut de paiement. Il contient des informations financières et démographiques sur 850 clients existants et éventuels. Les premières 700 observations concernent des clients auxquels des prêts ont été octroyés. Les 150 dernières observations correspondent aux clients éventuels que la banque doit classer comme bons ou mauvais risques de crédit.
- **bankloan_binning.sav.** Ce fichier de données d'hypothèse concerne des informations financières et démographiques sur 5 000 clients existants.
- **behavior.sav.** Dans un exemple classique (Price et Bouffard, 1974), on a demandé à 52 étudiants de noter les combinaisons établies à partir de 15 situations et de 15 comportements sur une échelle de 0 à 9, où 0 = « extrêmement approprié » et 9 = « extrêmement inapproprié ». En effectuant la moyenne des résultats de l'ensemble des individus, on constate une certaine différence entre les valeurs.
- **behavior_ini.sav.** Ce fichier de données contient la configuration initiale d'une solution bidimensionnelle pour *behavior.sav*.
- **brakes.sav.** Ce fichier de données d'hypothèse concerne le contrôle qualité effectué dans une usine qui fabrique des freins à disque pour des voitures haut de gamme. Le fichier de données contient les mesures de diamètre de 16 disques de 8 machines de production. Le diamètre cible des freins est de 322 millimètres.
- **breakfast.sav.** Au cours d'une étude classique (Green et Rao, 1972), on a demandé à 21 étudiants en MBA (Master of Business Administration) de l'école de Wharton et à leurs conjoints de classer 15 aliments du petit-déjeuner selon leurs préférences, de 1= « aliment préféré » à 15= « aliment le moins apprécié ». Leurs préférences ont été enregistrées dans six scénarios différents, allant de « Préférence générale » à « En-cas avec boisson uniquement ».
- **breakfast-overall.sav.** Ce fichier de données contient les préférences de petit-déjeuner du premier scénario uniquement, « Préférence générale ».
- **broadband_1.sav.** Ce fichier de données d'hypothèse concerne le nombre d'abonnés, par région, à un service haut débit. Le fichier de données contient le nombre d'abonnés mensuels de 85 régions sur une période de quatre ans.
- **broadband_2.sav.** Ce fichier de données est identique au fichier *broadband_1.sav* mais contient les données relatives à trois mois supplémentaires.
- **car_insurance_claims.sav.** Il s'agit d'un ensemble de données présenté et analysé ailleurs (McCullagh et Nelder, 1989) qui concerne des actions en indemnisation pour des voitures. Le montant d'action en indemnisation moyen peut être modélisé comme présentant une distribution gamma, à l'aide d'une fonction de lien inverse pour associer la moyenne de la variable dépendante à une combinaison linéaire de l'âge de l'assuré, du type de véhicule et de l'âge du véhicule. Le nombre d'actions entreprises peut être utilisé comme pondération de positionnement.
- **car_sales.sav.** Ce fichier de données contient des estimations de ventes hypothétiques, des barèmes de prix et des spécifications physiques concernant divers modèles et marques de véhicule. Les barèmes de prix et les spécifications physiques proviennent tour à tour de *edmunds.com* et des sites des constructeurs.
- **car_sales_uprepared.sav.** Il s'agit d'une version modifiée de *car_sales.sav* qui n'inclut aucune version transformée des champs.

- **carpet.sav.** Dans un exemple courant (Green et Wind, 1973), une société intéressée par la commercialisation d'un nouveau nettoyeur de tapis souhaite examiner l'influence de cinq critères sur la préférence du consommateur : la conception du conditionnement, la marque, le prix, une étiquette *Economique* et une garantie satisfait ou remboursé. Il existe trois niveaux de critère pour la conception du conditionnement, suivant l'emplacement de l'applicateur, trois marques (*K2R*, *Glory* et *Bissell*), trois niveaux de prix et deux niveaux (non ou oui) pour chacun des deux derniers critères. Dix consommateurs classent 22 profils définis par ces critères. La variable *Préférence* indique le classement des rangs moyens de chaque profil. Un rang faible correspond à une préférence élevée. Cette variable reflète une mesure globale de préférence pour chaque profil.
- **carpet_prefs.sav.** Ce fichier de données repose sur le même exemple que celui décrit pour *carpet.sav*, mais contient les classements réels issus de chacun des 10 clients. On a demandé aux consommateurs de classer les 22 profils de produits, du préféré au moins intéressant. Les variables *PREF1* à *PREF22* contiennent les identificateurs des profils associés, tels qu'ils sont définis dans *carpet_plan.sav*.
- **catalog.sav.** Ce fichier de données contient des chiffres de ventes mensuelles hypothétiques relatifs à trois produits vendus par une entreprise de vente par correspondance. Les données relatives à cinq variables explicatives possibles sont également incluses.
- **catalog_seasfac.sav.** Ce fichier de données est identique à *catalog.sav* mais contient en plus un ensemble de facteurs saisonniers calculés à partir de la procédure de désaisonnalisation, ainsi que les variables de date correspondantes.
- **cellular.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un opérateur téléphonique pour réduire les taux de désabonnement. Des scores de propension au désabonnement sont attribués aux comptes, de 0 à 100. Les comptes ayant une note égale ou supérieure à 50 sont susceptibles de changer de fournisseur.
- **ceramics.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un fabricant pour déterminer si un nouvel alliage haute qualité résiste mieux à la chaleur qu'un alliage standard. Chaque observation représente un test séparé de l'un des deux alliages ; le degré de chaleur auquel l'alliage ne résiste pas est enregistré.
- **cereal.sav.** Ce fichier de données d'hypothèse concerne un sondage de 880 personnes interrogées sur leurs préférences de petit-déjeuner et sur leur âge, leur sexe, leur situation familiale et leur mode de vie (actif ou non actif, selon qu'elles pratiquent une activité physique au moins deux fois par semaine). Chaque observation correspond à un répondant distinct.
- **clothing_defects.sav.** Ce fichier de données d'hypothèse concerne le processus de contrôle qualité observé dans une usine de textile. Dans chaque lot produit à l'usine, les inspecteurs prélèvent un échantillon de vêtements et comptent le nombre de vêtements qui ne sont pas acceptables.
- **coffee.sav.** Ce fichier de données concerne l'image perçue de six marques de café frappé (Kennedy, Riquier, et Sharp, 1996). Pour chacun des 23 attributs d'image de café frappé, les personnes sollicitées ont sélectionné toutes les marques décrites par l'attribut. Les six marques sont appelées AA, BB, CC, DD, EE et FF à des fins de confidentialité.
- **contacts.sav.** Ce fichier de données d'hypothèse concerne les listes de contacts d'un groupe de représentants en informatique d'entreprise. Chaque contact est classé selon le service de l'entreprise où il travaille et le classement de son entreprise. Sont également enregistrés le

montant de la dernière vente effectuée, le temps passé depuis la dernière vente et la taille de l'entreprise du contact.

- **creditpromo.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un grand magasin pour évaluer l'efficacité d'une promotion récente de carte de crédit. A cette fin, 500 détenteurs de carte ont été sélectionnés au hasard. La moitié a reçu une publicité faisant la promotion d'un taux d'intérêt réduit sur les achats effectués dans les trois mois à venir. L'autre moitié a reçu une publicité saisonnière standard.
- **customer_dbase.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une société pour utiliser les informations figurant dans sa banque de données et proposer des offres spéciales aux clients susceptibles d'être intéressés. Un sous-groupe de la base de clients a été sélectionné au hasard et a reçu des offres spéciales. Les réponses des clients ont été enregistrées.
- **customer_information.sav.** Un fichier de données d'hypothèse qui contient les informations postales du client, telles que le nom et l'adresse.
- **customer_subset.sav.** Un sous-ensemble de 80 observations de *customer_dbase.sav*.
- **debate.sav.** Ce fichier de données d'hypothèse concerne des réponses appariées à une enquête donnée aux participants à un débat politique avant et après le débat. Chaque observation représente un répondant distinct.
- **debate_aggregate.sav.** Il s'agit d'un fichier de données d'hypothèse qui rassemble les réponses dans le fichier *debate.sav*. Chaque observation correspond à une classification croisée de préférence avant et après le débat.
- **demo.sav.** Ce fichier de données d'hypothèse concerne une base de données clients achetée en vue de diffuser des offres mensuelles. Les données indiquent si le client a répondu ou non à l'offre et contiennent diverses informations démographiques.
- **demo_cs_1.sav.** Ce fichier de données d'hypothèse concerne la première mesure entreprise par une société pour compiler une base de données contenant des informations d'enquête. Chaque observation correspond à une ville différente. La région, la province, le quartier et la ville sont enregistrés.
- **demo_cs_2.sav.** Ce fichier de données d'hypothèse concerne la seconde mesure entreprise par une société pour compiler une base de données contenant des informations d'enquête. Chaque observation correspond à un ménage différent issu des villes sélectionnées à la première étape. La région, la province, le quartier, la ville, la sous-division et l'identification sont enregistrés. Les informations d'échantillonnage des deux premières étapes de la conception sont également incluses.
- **demo_cs.sav.** Ce fichier de données d'hypothèse concerne des informations d'enquête collectées via une méthode complexe d'échantillonnage. Chaque observation correspond à un ménage différent et diverses informations géographiques et d'échantillonnage sont enregistrées.
- **dmdata.sav.** Ceci est un fichier de données d'hypothèse qui contient des informations démographiques et des informations concernant les achats pour une entreprise de marketing direct. *dmdata2.sav* contient les informations pour un sous-ensemble de contacts qui ont reçu un envoi d'essai, et *dmdata3.sav* contient des informations sur les contacts restants qui n'ont pas reçu l'envoi d'essai.

- **dietstudy.sav.** Ce fichier de données d'hypothèse contient les résultats d'une étude portant sur le régime de Stillman (Rickman, Mitchell, Dingman, et Dalen, 1974). Chaque observation correspond à un sujet distinct et enregistre son poids en livres avant et après le régime, ainsi que ses niveaux de triglycérides en mg/100 ml.
- **dvdplayer.sav.** Ce fichier de données d'hypothèse concerne le développement d'un nouveau lecteur DVD. À l'aide d'un prototype, l'équipe de marketing a collecté des données de groupes spécifiques. Chaque observation correspond à un utilisateur interrogé et enregistre des informations démographiques sur cet utilisateur, ainsi que ses réponses aux questions portant sur le prototype.
- **german_credit.sav.** Ce fichier de données provient de l'ensemble de données « German credit » figurant dans le référentiel Machine Learning Databases (Blake et Merz, 1998) de l'université de Californie, Irvine.
- **grocery_1month.sav.** Ce fichier de données d'hypothèse est le fichier de données *grocery_coupons.sav* dans lequel les achats hebdomadaires sont organisés par client distinct. Certaines variables qui changeaient toutes les semaines disparaissent. En outre, le montant dépensé enregistré est à présent la somme des montants dépensés au cours des quatre semaines de l'enquête.
- **grocery_coupons.sav.** Il s'agit d'un fichier de données d'hypothèse qui contient des données d'enquête collectées par une chaîne de magasins d'alimentation qui cherchent à déterminer les habitudes de consommation de ses clients. Chaque client est suivi pendant quatre semaines et chaque observation correspond à une semaine distincte. Les informations enregistrées concernent les endroits où le client effectue ses achats, la manière dont il les effectue, ainsi que les sommes dépensées en provisions au cours de cette semaine.
- **guttman.sav.** Bell (Bell, 1961) a présenté un tableau pour illustrer les groupes sociaux possibles. Guttman (Guttman, 1968) a utilisé une partie de ce tableau, dans lequel cinq variables décrivant des éléments tels que l'interaction sociale, le sentiment d'appartenance à un groupe, la proximité physique des membres et la formalité de la relation, ont été croisées avec sept groupes sociaux théoriques, dont les foules (par exemple, le public d'un match de football), l'audience (par exemple, au cinéma ou dans une salle de classe), le public (par exemple, les journaux ou la télévision), les bandes (proche d'une foule, mais qui serait caractérisée par une interaction beaucoup plus intense), les groupes primaires (intimes), les groupes secondaires (volontaires) et la communauté moderne (groupement lâche issu d'une forte proximité physique et d'un besoin de services spécialisés).
- **health_funding.sav.** Ce fichier de données d'hypothèse concerne des données sur le financement des soins de santé (montant par groupe de 100 individus), les taux de maladie (taux par groupe de 10 000 individus) et les visites chez les prestataires de soins de santé (taux par groupe de 10 000 individus). Chaque observation représente une ville différente.
- **hivassay.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un laboratoire pharmaceutique pour développer une analyse rapide de détection d'infection HIV. L'analyse a pour résultat huit nuances de rouge, les nuances les plus marquées indiquant une plus forte probabilité d'infection. Un test en laboratoire a été effectué sur 2 000 échantillons de sang, la moitié de ces échantillons étant infectée par le virus HIV et l'autre moitié étant saine.
- **hourlywagedata.sav.** Ce fichier de données d'hypothèse concerne les salaires horaires d'infirmières occupant des postes administratifs et dans les services de soins, et affichant divers niveaux d'expérience.

- **insurance_claims.sav.** Il s'agit d'un fichier de données hypothétiques qui concerne une compagnie d'assurance souhaitant développer un modèle pour signaler des réclamations suspectes, potentiellement frauduleuses. Chaque observation correspond à une réclamation distincte.
- **insure.sav.** Ce fichier de données d'hypothèse concerne une compagnie d'assurance qui étudie les facteurs de risque indiquant si un client sera amené à déclarer un incident au cours d'un contrat d'assurance vie d'une durée de 10 ans. Chaque observation figurant dans le fichier de données représente deux contrats, l'un ayant enregistré une réclamation et l'autre non, appariés par âge et sexe.
- **judges.sav.** Ce fichier de données d'hypothèse concerne les scores attribués par des juges expérimentés (plus un juge enthousiaste) à 300 performances de gymnastique. Chaque ligne représente une performance distincte ; les juges ont examiné les mêmes performances.
- **kinship_dat.sav.** Rosenberg et Kim (Rosenberg et Kim, 1975) se sont lancés dans l'analyse de 15 termes de parenté (cousin/cousine, fille, fils, frère, grand-mère, grand-père, mère, neveu, nièce, oncle, père, petite-fille, petit-fils, sœur, tante). Ils ont demandé à quatre groupes d'étudiants (deux groupes de femmes et deux groupes d'hommes) de trier ces termes en fonction des similarités. Deux groupes (un groupe de femmes et un groupe d'hommes) ont été invités à effectuer deux tris, en basant le second sur un autre critère que le premier. Ainsi, un total de six "sources" a été obtenu. Chaque source correspond à une matrice de proximité 15×15 , dont le nombre de cellules est égal au nombre de personnes dans une source moins le nombre de fois où les objets ont été partitionnés dans cette source.
- **kinship_ini.sav.** Ce fichier de données contient une configuration initiale d'une solution tridimensionnelle pour *kinship_dat.sav*.
- **kinship_var.sav.** Ce fichier de données contient les variables indépendantes *sexe*, *génér(ation)* et *degré* (de séparation) permettant d'interpréter les dimensions d'une solution pour *kinship_dat.sav*. Elles permettent en particulier de réduire l'espace de la solution à une combinaison linéaire de ces variables.
- **marketvalues.sav.** Ce fichier de données concerne les ventes de maisons dans un nouvel ensemble à Algonquin (Illinois) au cours des années 1999–2000. Ces ventes relèvent des archives publiques.
- **nhis2000_subset.sav.** Le NHIS (National Health Interview Survey) est une enquête de grande envergure concernant la population des États-Unis. Des entretiens ont lieu avec un échantillon de ménages représentatifs de la population américaine. Des informations démographiques et des observations sur l'état de santé et le comportement sanitaire sont recueillies auprès des membres de chaque ménage. Ce fichier de données contient un sous-groupe d'informations issues de l'enquête de 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Fichier de données et documentation d'usage public. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/. Accès en 2003.
- **ozone.sav.** Les données incluent 330 observations portant sur six variables météorologiques pour prévoir la concentration d'ozone à partir des variables restantes. Des chercheurs précédents (Breiman et Friedman, 1985), (Hastie et Tibshirani, 1990), ont décelé parmi ces variables des non-linéarités qui pénalisent les approches standard de la régression.

- **pain_medication.sav.** Ce fichier de données d'hypothèse contient les résultats d'un essai clinique d'un remède anti-inflammatoire traitant les douleurs de l'arthrite chronique. On cherche notamment à déterminer le temps nécessaire au médicament pour agir et les résultats qu'il permet d'obtenir par rapport à un médicament existant.
- **patient_los.sav.** Ce fichier de données d'hypothèse contient les dossiers médicaux de patients admis à l'hôpital pour suspicion d'infarctus du myocarde suspecté (ou « attaque cardiaque »). Chaque observation correspond à un patient distinct et enregistre de nombreuses variables liées à son séjour à l'hôpital.
- **patlos_sample.sav.** Ce fichier de données d'hypothèse contient les dossiers médicaux d'un échantillon de patients sous traitement thrombolytique après un infarctus du myocarde. Chaque observation correspond à un patient distinct et enregistre de nombreuses variables liées à son séjour à l'hôpital.
- **poll_cs.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un enquêteur pour déterminer le niveau de soutien du public pour un projet de loi avant législature. Les observations correspondent à des électeurs enregistrés. Chaque observation enregistre le comté, la ville et le quartier où habite l'électeur.
- **poll_cs_sample.sav.** Ce fichier de données d'hypothèse contient un échantillon des électeurs répertoriés dans le fichier *poll_cs.sav*. L'échantillon a été prélevé selon le plan spécifié dans le fichier de plan *poll_csplan* et ce fichier de données enregistre les probabilités d'inclusion et les pondérations d'échantillon. Toutefois, ce plan faisant appel à une méthode d'échantillonnage de probabilité proportionnelle à la taille (PPS – Probability-Proportional-to-Size), il existe également un fichier contenant les probabilités de sélection conjointes (*poll_jointprob.sav*). Les variables supplémentaires correspondant à la répartition démographique des électeurs et à leur opinion sur le projet de loi proposé ont été collectées et ajoutées au fichier de données une fois l'échantillon prélevé.
- **property_assess.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un contrôleur au niveau du comté pour maintenir les évaluations de valeur de propriété à jour sur des ressources limitées. Les observations correspondent à des propriétés vendues dans le comté au cours de l'année précédente. Chaque observation du fichier de données enregistre la ville où se trouve la propriété, l'évaluateur ayant visité la propriété pour la dernière fois, le temps écoulé depuis cette évaluation, l'évaluation effectuée à ce moment-là et la valeur de vente de la propriété.
- **property_assess_cs.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un contrôleur du gouvernement pour maintenir les évaluations de valeur de propriété à jour sur des ressources limitées. Les observations correspondent à des propriétés de l'état. Chaque observation du fichier de données enregistre le comté, la ville et le quartier où se trouve la propriété, le temps écoulé depuis la dernière évaluation et l'évaluation alors effectuée.
- **property_assess_cs_sample.sav.** Ce fichier de données d'hypothèse contient un échantillon des propriétés répertoriées dans le fichier *property_assess_cs.sav*. L'échantillon a été prélevé selon le plan spécifié dans le fichier de plan *property_assess_csplan* et ce fichier de données enregistre les probabilités d'inclusion et les pondérations d'échantillon. La variable supplémentaire *Valeur courante* a été collectée et ajoutée au fichier de données une fois l'échantillon prélevé.

- **recidivism.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une agence administrative d'application de la loi pour interpréter les taux de récidive dans la juridiction. Chaque observation correspond à un récidiviste et enregistre les informations démographiques qui lui sont propres, certains détails sur le premier délit commis, ainsi que le temps écoulé jusqu'à la seconde arrestation si elle s'est produite dans les deux années suivant la première.
- **recidivism_cs_sample.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une agence administrative d'application de la loi pour interpréter les taux de récidive dans la juridiction. Chaque observation correspond à un récidiviste libéré suite à la première arrestation en juin 2003 et enregistre les informations démographiques qui lui sont propres, certains détails sur le premier délit commis et les données relatives à la seconde arrestation, si elle a eu lieu avant fin juin 2006. Les récidivistes ont été choisis dans plusieurs départements échantillonnés conformément au plan d'échantillonnage spécifié dans *recidivism_cs.csplan*. Ce plan faisant appel à une méthode d'échantillonnage de probabilité proportionnelle à la taille (PPS - Probability proportional to size), il existe également un fichier contenant les probabilités de sélection conjointes (*recidivism_cs_jointprob.sav*).
- **rfm_transactions.sav.** Un fichier de données d'hypothèse qui contient les données de transaction d'achat, y compris la date d'achat, le/les élément(s) acheté(s) et le montant monétaire pour chaque transaction.
- **salesperformance.sav.** Ce fichier de données d'hypothèse concerne l'évaluation de deux nouveaux cours de formation en vente. Soixante employés, divisés en trois groupes, reçoivent chacun une formation standard. En outre, le groupe 2 suit une formation technique et le groupe 3 un didacticiel pratique. À l'issue du cours de formation, chaque employé est testé et sa note enregistrée. Chaque observation du fichier de données représente un stagiaire distinct et enregistre le groupe auquel il a été assigné et la note qu'il a obtenue au test.
- **satisf.sav.** Il s'agit d'un fichier de données d'hypothèse portant sur une enquête de satisfaction effectuée par une société de vente au détail au niveau de quatre magasins. Un total de 582 clients ont été interrogés et chaque observation représente la réponse d'un seul client.
- **screws.sav.** Ce fichier de données contient des informations sur les descriptives des vis, des boulons, des écrous et des clous. (Hartigan, 1975).
- **shampoo_ph.sav.** Ce fichier de données d'hypothèse concerne le processus de contrôle qualité observé dans une usine de produits capillaires. À intervalles réguliers, six lots de sortie distincts sont mesurés et leur pH enregistré. La plage cible est 4,5–5,5.
- **ships.sav.** Il s'agit d'un ensemble de données présenté et analysé ailleurs (McCullagh et al., 1989) et concernant les dommages causés à des cargos par les vagues. Les effectifs d'incidents peuvent être modélisés comme des incidents se produisant selon un taux de Poisson en fonction du type de navire, de la période de construction et de la période de service. Les mois de service totalisés pour chaque cellule du tableau formé par la classification croisée des facteurs fournissent les valeurs d'exposition au risque.
- **site.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une société pour choisir de nouveaux sites pour le développement de ses activités. L'entreprise a fait appel à deux consultants pour évaluer séparément les sites. Ces consultants, en plus de fournir un rapport approfondi, ont classé chaque site comme constituant une éventualité « bonne », « moyenne » ou « faible ».

- **smokers.sav.** Ce fichier de données est extrait de l'étude National Household Survey of Drug Abuse de 1998 et constitue un échantillon de probabilité des ménages américains. (<http://dx.doi.org/10.3886/ICPSR02934>) Ainsi, la première étape dans l'analyse de ce fichier doit consister à pondérer les données pour refléter les tendances de population.
- **stocks.sav** Ce fichier de données hypothétiques contient le cours et le volume des actions pour un an.
- **stroke_clean.sav.** Ce fichier de données d'hypothèse concerne l'état d'une base de données médicales une fois celle-ci purgée via des procédures de l'option Validation de données.
- **stroke_invalid.sav.** Ce fichier de données d'hypothèse concerne l'état initial d'une base de données médicales et comporte plusieurs erreurs de saisie de données.
- **stroke_survival.** Ce fichier de données d'hypothèse concerne les temps de survie de patients qui quittent un programme de rééducation à la suite d'un accident ischémique et rencontrent un certain nombre de problèmes. Après l'attaque, l'occurrence d'infarctus du myocarde, d'accidents ischémiques ou hémorragiques est signalée, et le moment de l'événement enregistré. L'échantillon est tronqué à gauche car il n'inclut que les patients ayant survécu durant le programme de rééducation mis en place suite à une attaque.
- **stroke_valid.sav.** Ce fichier de données d'hypothèse concerne l'état d'une base de données médicales une fois les valeurs vérifiées via la procédure Validation de données. Elle contient encore des observations anormales potentielles.
- **survey_sample.sav.** Ce fichier de données concerne des informations d'enquête dont des données démographiques et des mesures comportementales. Il est basé sur un sous-ensemble de variables de la 1998 NORC General Social Survey, bien que certaines valeurs de données aient été modifiées et que des variables supplémentaires fictives aient été ajoutées à titre de démonstration.
- **telco.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une société de télécommunications pour réduire les taux de désabonnement de sa base de clients. Chaque observation correspond à un client distinct et enregistre diverses informations démographiques et d'utilisation de service.
- **telco_extra.sav.** Ce fichier de données est semblable au fichier de données *telco.sav* mais les variables de permanence et de dépenses des consommateurs transformées log ont été supprimées et remplacées par des variables de dépenses des consommateurs transformées log standardisées.
- **telco_missing.sav.** Ce fichier de données est un sous-ensemble du fichier de données *telco.sav* mais certaines des valeurs de données démographiques ont été remplacées par des valeurs manquantes.
- **testmarket.sav.** Ce fichier de données d'hypothèse concerne une chaîne de fast foods et ses plans marketing visant à ajouter un nouveau plat à son menu. Trois campagnes étant possibles pour promouvoir le nouveau produit, le nouveau plat est introduit sur des sites sur plusieurs marchés sélectionnés au hasard. Une promotion différente est effectuée sur chaque site et les ventes hebdomadaires du nouveau plat sont enregistrées pour les quatre premières semaines. Chaque observation correspond à un site-semaine distinct.
- **testmarket_1month.sav.** Ce fichier de données d'hypothèse est le fichier de données *testmarket.sav* dans lequel les ventes hebdomadaires sont organisées par site distinct. Certaines variables qui changeaient toutes les semaines disparaissent. En outre, les ventes

enregistrées sont à présent la somme des ventes réalisées au cours des quatre semaines de l'enquête.

- **tree_car.sav.** Ce fichier de données d'hypothèse concerne des données démographiques et de prix d'achat de véhicule.
- **tree_credit.sav.** Ce fichier de données d'hypothèse concerne des données démographiques et d'historique de prêt bancaire.
- **tree_missing_data.sav** Ce fichier de données d'hypothèse concerne des données démographiques et d'historique de prêt bancaire avec un grand nombre de valeurs manquantes.
- **tree_score_car.sav.** Ce fichier de données d'hypothèse concerne des données démographiques et de prix d'achat de véhicule.
- **tree_textdata.sav.** Ce fichier de données simples ne comporte que deux variables et vise essentiellement à indiquer l'état par défaut des variables avant affectation du niveau de mesure et des étiquettes de valeurs.
- **tv-survey.sav.** Ce fichier de données d'hypothèse concerne une enquête menée par un studio de télévision qui envisage de prolonger la diffusion d'un programme ou de l'arrêter. On a demandé à 906 personnes si elles regarderaient le programme dans diverses situations. Chaque ligne représente un répondant distinct et chaque colonne une situation distincte.
- **ulcer_recurrence.sav.** Ce fichier contient des informations partielles d'une enquête visant à comparer l'efficacité de deux thérapies de prévention de la récurrence des ulcères. Il fournit un bon exemple de données censurées par intervalle et a été présenté et analysé ailleurs (Collett, 2003).
- **ulcer_recurrence_recoded.sav.** Ce fichier réorganise les informations figurant dans le fichier *ulcer_recurrence.sav* pour que vous puissiez modéliser la probabilité d'événement pour chaque intervalle de l'enquête plutôt que la probabilité d'événement de fin d'enquête. Il a été présenté et analysé ailleurs (Collett et al., 2003).
- **verd1985.sav.** Ce fichier de données concerne une enquête (Verdegaal, 1985). Les réponses de 15 sujets à 8 variables ont été enregistrées. Les variables présentant un intérêt sont divisées en trois ensembles. Le groupe 1 comprend l'âge et la *situation familiale*, le groupe 2 les *animaux domestiques* et la *presse*, et le groupe 3 la *musique* et l'*habitat*. A la variable *animal domestique* est appliqué un codage nominal multiple et à *âge*, un codage ordinal ; toutes les autres variables ont un codage nominal simple.
- **virus.sav.** Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un fournisseur de services Internet pour déterminer les effets d'un virus sur ses réseaux. Il a suivi le pourcentage (approximatif) de trafic de messages électroniques infectés par un virus sur ses réseaux sur la durée, de la découverte à la circonscription de la menace.
- **wheeze_steubenville.sav.** Il s'agit d'un sous-ensemble d'une enquête longitudinale des effets de la pollution de l'air sur la santé des enfants (Ware, Dockery, Spiro III, Speizer, et Ferris Jr., 1984). Les données contiennent des mesures binaires répétées de l'état asthmatique d'enfants de la ville de Steubenville (Ohio), âgés de 7, 8, 9 et 10 ans, et indiquent si la mère fumait au cours de la première année de l'enquête.
- **workprog.sav.** Ce fichier de données d'hypothèse concerne un programme de l'administration visant à proposer de meilleurs postes aux personnes défavorisées. Un échantillon de participants potentiels au programme a ensuite été prélevé. Certains de ces participants ont

été sélectionnés au hasard pour participer au programme. Chaque observation représente un participant au programme distinct.

- **worldsales.sav** Ce fichier de données hypothétiques contient les revenus des ventes par continent et par produit.



Remarques

Ces informations ont été développées pour les produits et services offerts dans le monde.

Il est possible qu'IBM n'offre pas dans les autres pays les produits, services et fonctionnalités décrits dans ce document. Contactez votre représentant local IBM pour obtenir des informations sur les produits et services actuellement disponibles dans votre région. Toute référence à un produit, programme ou service IBM n'implique pas que les seuls les produits, programmes ou services IBM peuvent être utilisés. Tout produit, programme ou service de fonctionnalité équivalente qui ne viole pas la propriété intellectuelle IBM peut être utilisé à la place. Cependant l'utilisateur doit évaluer et vérifier l'utilisation d'un produit, programme ou service non IBM.

IBM peut posséder des brevets ou des applications de brevet en attente qui couvrent les sujets décrits dans ce document. L'octroi de ce document n'équivaut aucunement à celui d'une licence pour ces brevets. Vous pouvez envoyer par écrit des questions concernant la licence à :

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, États-Unis

Pour obtenir des informations de licence concernant la configuration de caractères codés sur deux octets (DBCS), veuillez contacter dans votre pays le département chargé de la propriété intellectuelle chez IBM ou envoyez vos commentaires par écrit à :

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

Le paragraphe suivant ne s'applique pas au Royaume-Uni ni à aucun pays dans lequel ces dispositions sont contraires au droit local : INTERNATIONAL BUSINESS MACHINES FOURNIT CETTE PUBLICATION « EN L'ETAT » SANS GARANTIE D'AUCUNE SORTE, IMPLICITE OU EXPLICITE, Y COMPRIS, MAIS SANS ETRE LIMITE AUX GARANTIES IMPLICITES DE NON VIOLATION, DE QUALITE MARCHANDE OU D'ADAPTATION POUR UN USAGE PARTICULIER. Certains états n'autorisent pas l'exclusion de garanties explicites ou implicites lors de certaines transactions, par conséquent, il est possible que cet énoncé ne vous concerne pas.

Ces informations peuvent contenir des erreurs techniques ou des erreurs typographiques. Ces informations sont modifiées de temps en temps ; ces modifications seront intégrées aux nouvelles versions de la publication. IBM peut apporter des améliorations et/ou modifications des produits et/ou des programmes décrits dans cette publications à tout moment sans avertissement préalable.

Toute référence dans ces informations à des sites Web autres qu'IBM est fournie dans un but pratique uniquement et ne sert en aucun cas de recommandation pour ces sites Web. Le matériel contenu sur ces sites Web ne fait pas partie du matériel de ce produit IBM et l'utilisation de ces sites Web se fait à vos propres risques.

IBM peut utiliser ou distribuer les informations que vous lui fournissez, de la façon dont il le souhaite, sans encourir aucune obligation envers vous.

Les personnes disposant d'une licence pour ce programme et qui souhaitent obtenir des informations sur celui-ci pour activer : (i) l'échange d'informations entre des programmes créés de manière indépendante et d'autres programmes (notamment celui-ci) et (ii) l'utilisation mutuelle des informations qui ont été échangées, doivent contacter :

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, États-Unis.

Ces informations peuvent être disponibles, soumises à des conditions générales, et dans certains cas payantes.

Le programme sous licence décrit dans ce document et toute la documentation sous licence disponible pour ce programme sont fournis par IBM en conformité avec les conditions de l'accord du client IBM, avec l'accord de licence du programme international IBM et avec tout accord équivalent entre nous.

les informations concernant les produits autres qu'IBM ont été obtenues auprès des fabricants de ces produits, leurs annonces publiques ou d'autres sources publiques disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances, leur compatibilité ou toute autre fonctionnalité associée à des produits autres qu'IBM. Les questions sur les capacités de produits autres qu'IBM doivent être adressées aux fabricants de ces produits.

Ces informations contiennent des exemples de données et de rapports utilisés au cours d'opérations quotidiennes standard. Pour les illustrer le mieux possible, ces exemples contiennent des noms d'individus, d'entreprises, de marques et de produits. Tous ces noms sont fictifs et toute ressemblance avec des noms et des adresses utilisés par une entreprise réelle ne serait que pure coïncidence.

Si vous consultez la version papier de ces informations, il est possible que certaines photographies et illustrations en couleurs n'apparaissent pas.

Marques commerciales

IBM, le logo IBM, ibm.com et SPSS sont des marques commerciales d'IBM Corporation, déposées dans de nombreuses juridictions du monde entier. Une liste à jour des marques IBM est disponible sur Internet à l'adresse <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques déposées ou des marques commerciales de Adobe Systems Incorporated aux États-Unis et/ou dans d'autres pays.

Intel, le logo Intel, Intel Inside, le logo Intel Inside, Intel Centrino, le logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium, et Pentium sont des marques commerciales ou des marques déposées de Intel Corporation ou de ses filiales aux États-Unis et dans d'autres pays.

Java et toutes les marques et logos Java sont des marques commerciales de Sun Microsystems, Inc. aux États-Unis et/ou dans d'autres pays.

Linux est une marque déposée de Linus Torvalds aux États-Unis et/ou dans d'autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques commerciales de Microsoft Corporation aux États-Unis et/ou dans d'autres pays.

UNIX est une marque déposée de The Open Group aux États-Unis et dans d'autres pays.

Ce produit utilise WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com/>.

Les autres noms de produits et de services peuvent être des marques d'IBM ou d'autres sociétés.

Les captures d'écran des produits Adobe sont reproduites avec l'autorisation de Adobe Systems Incorporated.

Les captures d'écran des produits Microsoft sont reproduites avec l'autorisation de Microsoft Corporation.



Bibliographie

- Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.
- Blake, C. L., et C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mlearn/MLRepository.html>.
- Box, G. E. P., G. M. Jenkins, et G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*, 3rd éd. Englewood Cliffs, N.J.: Prentice Hall.
- Breiman, L., et J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, .
- Collett, D. 2003. *Modelling survival data in medical research*, 2 éd. Boca Raton: Chapman & Hall/CRC.
- Gardner, E. S. 1985. Exponential smoothing: The state of the art. *Journal of Forecasting*, 4, .
- Green, P. E., et V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.
- Green, P. E., et 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, .
- Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.
- Hastie, T., et R. Tibshirani. 1990. *Generalized additive models*. Londres: Chapman and Hall.
- Kennedy, R., C. Riquier, et B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, .
- McCullagh, P., et J. A. Nelder. 1989. *Generalized Linear Models*, 2nd éd. Londres: Chapman & Hall.
- Pena, D., G. C. Tiao, et R. S. Tsay, édés. 2001. *A course in time series analysis*. New York: John Wiley and Sons.
- Price, R. H., et D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, .
- Rickman, R., N. Mitchell, J. Dingman, et J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, .
- Rosenberg, S., et M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, .
- Van der Ham, T., J. J. Meulman, D. C. Van Strien, et H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, .
- Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (en néerlandais)*. Leiden: Department of Data Theory, University of Leiden.
- Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, et B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, .

Index

- ACF
 - dans Appliquer des modèles de séries chronologiques, 31, 33
 - dans le modélisateur de séries chronologiques, 19, 21
 - diagrammes des processus ARIMA réels, 100
- Analyse harmonique, 42
- Analyse spectrale, 42, 44
 - Analyse spectrale bivariée, 44
 - Centrage de transformation, 44
 - densité spectrale, 95
 - fenêtres spectrales, 42
 - Hypothèses, 42
 - périodogramme, 95
 - Procédures apparentées, 97
- Appliquer des modèles de séries chronologiques, 28, 58, 74
 - enregistrement de modèles réestimés en XML, 36
 - enregistrement de prévisions, 36, 60
 - Fonction d'autocorrélation partielle résiduelle, 31, 33
 - Fonction d'autocorrélation résiduelle, 31, 33
 - Intervalles de confiance, 33, 37
 - modèles les plus ou les moins appropriés, 35
 - nouveaux noms de variable, 36, 62
 - Paramètres de modèle, 31
 - période de prévision, 30, 59, 81
 - période d'estimation, 30
 - Prévisions, 31, 33, 82
 - réestimer les paramètres de modèle, 30, 59
 - Statistique de Ljung-Box, 31
 - Statistiques dans tous les modèles, 31, 33, 61
 - statistiques de qualité de l'ajustement, 31, 33, 61
 - tableau d'ajustement du modèle, 61
 - tableau des prévisions, 82
 - valeurs d'ajustement, 33
 - Valeurs manquantes, 37
- autorégression
 - modèles ARIMA, 13
- BIC (critère d'information bayésien) normalisé, 98
 - dans Appliquer des modèles de séries chronologiques, 31, 33
 - dans le modélisateur de séries chronologiques, 19, 21
- Désaisonnalisation, 39–41
 - Calcul de moyennes mobiles, 39
 - composante de date périodique, 85
 - Création de variables, 40
 - Enregistrement de nouvelles variables, 40
 - Hypothèses, 39
 - Modèles, 39
 - nouvelles variables, 91
 - Procédures apparentées, 92
- différence
 - modèles ARIMA, 13
- différence saisonnière
 - modèles ARIMA, 13
- données historiques
 - dans Appliquer des modèles de séries chronologiques, 33
 - dans le modélisateur de séries chronologiques, 21
- enregistrer
 - modèles réestimés en XML, 36
 - nouveaux noms de variable, 24, 36
 - prévisions de modèle, 24, 36
 - spécifications de modèles au format XML, 24
- Erreur Absolue Maximum, 98
 - dans Appliquer des modèles de séries chronologiques, 31, 33
 - dans le modélisateur de séries chronologiques, 19, 21
- Erreur Absolue Moyenne, 98
 - dans Appliquer des modèles de séries chronologiques, 31, 33
 - dans le modélisateur de séries chronologiques, 19, 21
- Erreur de Pourcentage en Moyenne absolue, 98
 - dans Appliquer des modèles de séries chronologiques, 31, 33, 62
 - dans le modélisateur de séries chronologiques, 19, 21, 54
- erreur quadratique moyenne, 98
 - dans Appliquer des modèles de séries chronologiques, 31, 33
 - dans le modélisateur de séries chronologiques, 19, 21
- événements, 9
 - dans le modélisateur de séries chronologiques, 8
- fichiers d'exemple
 - emplacement, 104
- fonction d'autocorrélation partielle
 - dans Appliquer des modèles de séries chronologiques, 31, 33
 - dans le modélisateur de séries chronologiques, 19, 21
 - diagrammes des processus ARIMA réels, 100
- fonctions d'autocorrélation
 - dans Appliquer des modèles de séries chronologiques, 31, 33
 - dans le modélisateur de séries chronologiques, 19, 21
 - diagrammes des processus ARIMA réels, 100
- fonctions de transfert, 15
 - délai, 15
 - ordres de dénominateur, 15
 - ordres de différence, 15
 - ordres de numérateur, 15
 - ordres saisonniers, 15
- intégration
 - modèles ARIMA, 13
- Intervalles de confiance
 - dans Appliquer des modèles de séries chronologiques, 33, 37
 - dans le modélisateur de séries chronologiques, 21, 25

- MAE, 98
 dans Appliquer des modèles de séries chronologiques, 31, 33
 dans le modélisateur de séries chronologiques, 19, 21
- MAPE, 98
 dans Appliquer des modèles de séries chronologiques, 31, 33, 62
 dans le modélisateur de séries chronologiques, 19, 21, 54
- marques commerciales, 116
- MaxAE, 98
 dans Appliquer des modèles de séries chronologiques, 31, 33
 dans le modélisateur de séries chronologiques, 19, 21
- MaxAPE, 98
 dans Appliquer des modèles de séries chronologiques, 31, 33, 62
 dans le modélisateur de séries chronologiques, 19, 21, 54
- mentions légales, 115
- modèle de lissage exponentiel amorti, 11
- modèle de lissage exponentiel de Brown, 11
- modèle de lissage exponentiel de Holt, 11
- modèle de lissage exponentiel de Winters
 additif, 11
 multiplicatif, 11
- modèle de lissage exponentiel saisonnier simple, 11
- modèle de lissage exponentiel simple, 11
- Modèles
 ARIMA, 7, 13
 lissage exponentiel, 7, 11
 Modélisateur expert, 7
- modèles ARIMA, 7
 constant, 13
 fonctions de transfert, 15
 ordres autorégressifs, 13
 ordres de différenciation, 13
 ordres de moyenne mobile, 13
 ordres saisonniers, 13
 valeurs éloignées, 17
- modèles de lissage exponentiel, 7, 11
- Modélisateur expert, 7, 47
 limitation de l'espace de modèle, 8, 50
 valeurs éloignées, 10, 67
- moyenne mobile
 modèles ARIMA, 13
- noms de modèle
 dans le modélisateur de séries chronologiques, 25
- noms de variable
 dans Appliquer des modèles de séries chronologiques, 36
 dans le modélisateur de séries chronologiques, 24
- observations traitées, 2
- ordres saisonniers
 modèles ARIMA, 13
- PACF
 dans Appliquer des modèles de séries chronologiques, 31, 33
 dans le modélisateur de séries chronologiques, 19, 21
 diagrammes des processus ARIMA réels, 100
- Paramètres de modèle
 dans Appliquer des modèles de séries chronologiques, 31
 dans le modélisateur de séries chronologiques, 19, 69
- période de prévision
 dans Appliquer des modèles de séries chronologiques, 30, 59, 81
 dans le modélisateur de séries chronologiques, 7, 25, 50–51
- période de validation, 2
- période d'estimation, 2
 dans Appliquer des modèles de séries chronologiques, 30
 dans le modélisateur de séries chronologiques, 7, 50
- période historique, 2
- périodicité
 dans le modélisateur de séries chronologiques, 8, 11, 13, 15
- Pourcentage d'Erreur Absolue Maximum, 98
 dans Appliquer des modèles de séries chronologiques, 31, 33, 62
 dans le modélisateur de séries chronologiques, 19, 21, 54
- Prévisions
 dans Appliquer des modèles de séries chronologiques, 31, 33, 82
 dans le modélisateur de séries chronologiques, 19, 21, 53
- Qualité de l'ajustement
 dans Appliquer des modèles de séries chronologiques, 31, 33, 61
 dans le modélisateur de séries chronologiques, 19, 21, 53
 définitions, 98
- R^2 , 98
 dans Appliquer des modèles de séries chronologiques, 31, 33
 dans le modélisateur de séries chronologiques, 19, 21
- R^2 stationnaire, 98
 dans Appliquer des modèles de séries chronologiques, 31, 33
 dans le modélisateur de séries chronologiques, 19, 21, 72
- réestimer les paramètres de modèle
 dans Appliquer des modèles de séries chronologiques, 30, 59
- Résidus
 dans Appliquer des modèles de séries chronologiques, 31, 33
 dans le modélisateur de séries chronologiques, 19, 21
- RMSE, 98
 dans Appliquer des modèles de séries chronologiques, 31, 33
 dans le modélisateur de séries chronologiques, 19, 21

- Statistique de Ljung-Box
 - dans Appliquer des modèles de séries chronologiques, 31
 - dans le modélisateur de séries chronologiques, 19, 72
- tableau d'ajustement du modèle
 - dans Appliquer des modèles de séries chronologiques, 61
- tableau de description du modèle
 - dans le modélisateur de séries chronologiques, 71
- tableau de statistiques du modèle
 - dans le modélisateur de séries chronologiques, 72
- Tableau des paramètres du modèle ARIMA
 - dans le modélisateur de séries chronologiques, 73
- tableau des prévisions
 - dans Appliquer des modèles de séries chronologiques, 82
 - dans le modélisateur de séries chronologiques, 57
- Time Series Modeler, 4
 - ARIMA, 7, 13
 - enregistrement de prévisions, 24, 52
 - enregistrement des spécifications de modèles au format XML, 24, 52, 68
 - événements, 8
 - Fonction d'autocorrélation partielle résiduelle, 19, 21
 - Fonction d'autocorrélation résiduelle, 19, 21
 - fonctions de transfert, 15
 - Intervalles de confiance, 21, 25
 - lissage exponentiel, 7, 11
 - modèles les plus ou les moins appropriés, 23
 - Modélisateur expert, 7, 47, 63
 - noms de modèle, 25
 - nouveaux noms de variable, 24, 56
 - Paramètres de modèle, 19, 69
 - période de prévision, 7, 25, 50–51
 - période d'estimation, 7, 50
 - périodicité, 8, 11, 13, 15
 - Prévisions, 19, 21, 53
 - Statistique de Ljung-Box, 19
 - Statistiques dans tous les modèles, 19, 21, 53, 55
 - statistiques de qualité de l'ajustement, 19, 21, 53, 72
 - tableau de description du modèle, 71
 - tableau de statistiques du modèle, 72
 - Tableau des paramètres du modèle ARIMA, 73
 - tableau des prévisions, 57
 - transformation de séries, 11, 13, 15
 - valeurs d'ajustement, 21, 70
 - valeurs éloignées, 10, 17, 67
 - Valeurs manquantes, 25
- Transformation log :
 - dans le modélisateur de séries chronologiques, 11, 13, 15
- transformation par log népérien
 - dans le modélisateur de séries chronologiques, 11, 13, 15
- transformation racine carrée
 - dans le modélisateur de séries chronologiques, 11, 13, 15
- valeur éloignée additive, 99
 - dans le modélisateur de séries chronologiques, 10, 17
- valeur éloignée additive saisonnière, 99
 - dans le modélisateur de séries chronologiques, 10, 17
- valeur éloignée corrective additive, 99
 - dans le modélisateur de séries chronologiques, 10, 17
- valeur éloignée de changement de niveau, 99
 - dans le modélisateur de séries chronologiques, 10, 17
- valeur éloignée de tendance locale, 99
 - dans le modélisateur de séries chronologiques, 10, 17
- valeur éloignée d'innovation, 99
 - dans le modélisateur de séries chronologiques, 10, 17
- valeur éloignée transitoire, 99
 - dans le modélisateur de séries chronologiques, 10, 17
- valeurs d'ajustement
 - dans Appliquer des modèles de séries chronologiques, 33
 - dans le modélisateur de séries chronologiques, 21, 70
- valeurs éloignées
 - modèles ARIMA, 17
 - Modélisateur expert, 10, 67
- Valeurs éloignées
 - définitions, 99
- Valeurs manquantes
 - dans Appliquer des modèles de séries chronologiques, 37
 - dans le modélisateur de séries chronologiques, 25
- XML
 - enregistrement de modèles réestimés en XML, 36
 - enregistrement des modèles de séries chronologiques au format XML, 24, 52, 68