

IBM SPSS Regression 30

IBM

Remarque

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations figurant dans la section [«Remarques»](#), à la page 51.

Notice d'édition

La présente édition s'applique à la version 30.0.0 d'IBM® SPSS Statistics et à toutes les éditions et modifications ultérieures sauf mention contraire dans les nouvelles éditions.

© **Copyright International Business Machines Corporation .**

Table des matières

Chapitre 1. Régression.....	1
Choix d'une procédure pour la régression logistique binaire.....	1
Régression logistique.....	2
Définition de la règle de régression logistique.....	3
Méthodes de sélection des variables de régression logistique.....	3
Régression logistique : définition des variables catégorielles.....	4
Régression logistique : enregistrer les nouvelles variables.....	4
Options de régression logistique.....	5
Fonctions supplémentaires de la commande REGRESSION LOGISTIQUE.....	6
Régression logistique multinomiale.....	6
Régression logistique multinomiale.....	7
Régression logistique multinomiale : Catégorie de référence.....	8
Régression logistique multinomiale : Statistiques.....	8
Régression logistique multinomiale : Critères.....	9
Options de régression logistique multinomiale.....	9
Régression logistique multinomiale : Enregistrer.....	10
Fonctions supplémentaires de la commande NOMREG.....	10
Régression probit.....	10
Analyse par la méthode des probits : définir une plage.....	11
Options des analyses par la méthode des probits.....	12
Fonctions supplémentaires de la commande NLR.....	12
Régression quantile.....	12
Régression quantile : Critères.....	14
Régression quantile : Modèle.....	15
Régression quantile : Affichage.....	16
Régression quantile : Enregistrer.....	17
Régression quantile : Exporter.....	18
Régression non linéaire.....	18
Logique conditionnelle (régression non linéaire).....	19
Paramètres de régression non linéaire.....	20
Modèles courants de régression non linéaire.....	20
Fonction de perte de la régression non linéaire.....	21
Options de contraintes de la régression non linéaire.....	21
Régression non linéaire : enregistrer les nouvelles variables.....	22
Options de régression non linéaire.....	22
Interprétation des résultats de la régression non linéaire.....	22
Fonctions supplémentaires de la commande NLR.....	23
Pondération estimée.....	23
Options de la pondération estimée.....	24
Fonctions supplémentaires de la commande WLS.....	25
Régression par les doubles moindres carrés.....	25
Options de régression par les doubles moindres carrés.....	26
Fonctions supplémentaires de la commande 2SLS.....	26
Méthodes de codification des variables catégorielles.....	26
Déviation.....	26
Simple.....	27
Helmert.....	27
Différence.....	27
Polynomial.....	28
Répété.....	28
Spécial.....	29

Indicateur.....	29
Kernel Ridge Regression.....	30
Paramètres de noyau.....	31
Régression de la crête du noyau: Options.....	31
Modèles de temps d'échec accéléré paramétriques.....	32
Modèles de temps d'échec accéléré paramétriques: Critères.....	33
Modèles de temps d'échec accéléré paramétriques: Modèle.....	34
Modèles de temps d'échec accélérés paramétriques: estimation.....	35
Modèles de temps d'échec accéléré paramétriques: Imprimer.....	36
Modèles de temps d'échec accéléré paramétriques: Prévoir.....	37
Modèles de temps d'échec accéléré paramétriques: Tracé.....	37
Modèles de temps d'échec accéléré paramétriques: Exportation.....	38
Survie AFT Définir des événements pour les variables de statut.....	38
Modèles de temps d'échec accélérés paramétriques: Sélectionner une catégorie.....	38
Modèles de fragilité partagés paramétriques.....	39
Modèles de fragilité partagés paramétriques: Critères.....	40
Modèles de fragilité partagés paramétriques: Modèle.....	41
Modèles de fragilité partagés paramétriques: estimation.....	42
Modèles de fragilité partagés paramétriques: Imprimer.....	43
Modèles de fragilité partagés paramétriques: Prévoir.....	44
Modèles de fragilité partagés paramétriques: Tracé.....	45
Modèles de fragilité partagés paramétriques: Exportation.....	45
Modèles de fragilité partagés paramétriques: Définir des événements.....	45
Modèles de fragilité partagés paramétriques-Exemples.....	46
Modèles de fragilité partagés paramétriques-Étude de cas pour les données récurrentes.....	48
Remarques.....	51
Marques.....	52
Index.....	55

Chapitre 1. Régression

Les fonctions de régression suivantes sont incluses dans SPSS Statistics Standard Edition ou l'option Régression.

Choix d'une procédure pour la régression logistique binaire

Les modèles de régression logistique binaire peuvent être ajustés à l'aide des procédures Régression logistique et Régression logistique multinomiale. Chacune de ces procédures comporte des options qui lui sont propres. Il convient de faire une importante distinction théorique entre les deux procédures : la procédure de régression logistique génère toutes les prévisions, résidus, statistiques d'influence et tests de qualité d'ajustement en utilisant les données au niveau des observations individuelles, quelle que soit la façon dont ces données ont été entrées et que le nombre de motifs de covariable soit inférieur ou non au nombre total d'observations ; alors que la procédure de régression logistique multinomiale agrège les observations au niveau interne pour constituer des sous-populations présentant des motifs de covariable identiques pour les prédicteurs, générant ainsi des prévisions, des résidus et des tests de qualité d'ajustement en fonction de ces sous-populations. Si tous les prédicteurs sont catégoriels ou que des prédicteurs continus prennent en compte un nombre limité de valeurs, de sorte qu'il existe plusieurs observations pour chaque motif de covariable distinct, la méthode de constitution de sous-populations peut générer des tests de qualité d'ajustement et des résidus informatifs valides, ce qui n'est pas le cas de la procédure effectuée au niveau des observations individuelles.

Régression logistique

Offre les fonctions spécifiques suivantes :

- test Hosmer-Lemeshow de la qualité d'ajustement du modèle ;
- analyses pas à pas ;
- contrastes permettant de définir le paramétrage du modèle ;
- césures alternatives pour le classement ;
- tracés de classement ;
- modèle ajusté sur un ensemble d'observations par rapport à un ensemble d'observations restant ;
- enregistrement des prévisions, des résidus et des statistiques d'influence.

Régression logistique multinomiale

Offre les fonctions spécifiques suivantes :

- tests du khi-deux de Pearson et de déviance pour la qualité d'ajustement du modèle ;
- définition de sous-populations pour le regroupement de données afin d'effectuer des tests de qualité d'ajustement ;
- énumération des effectifs, des effectifs prédits et des résidus par sous-population ;
- correction des estimations de variance pour la surdispersion ;
- matrice de covariance des estimations de paramètres ;
- tests des combinaisons linéaires de paramètres ;
- définition explicite des modèles imbriqués ;
- ajustement 1-1 de modèles de régression logistique conditionnels correspondants au moyen de variables différenciées.

Remarques :

- Ces deux procédures permettent d'ajuster un modèle pour des données binaires qui est un modèle linéaire généralisé avec une distribution binomiale et la fonction de lien logit. Si une autre fonction de lien est plus appropriée pour vos données, utilisez la procédure Modèles linéaires généralisés.

- Si vous possédez des mesures répétées de données binaires ou des enregistrements corrélés, vous devez envisager d'utiliser les procédures Modèles linéaires mixtes généralisés ou Equations d'estimation généralisées.

Régression logistique

La régression logistique est utile lorsque vous souhaitez être capable de prévoir la présence ou l'absence d'une caractéristique ou d'un résultat en fonction de certaines valeurs ou d'un groupe de variables de prédicteur. Elle est similaire à la régression linéaire mais elle convient aux modèles dans lesquelles les variables sont dichotomiques. Les coefficients de la régression logistique peuvent servir à estimer des rapports des cotes pour chacune des variables indépendantes d'un modèle. La régression logistique s'applique à une plus large gamme de situations de recherche que l'analyse discriminante.

Exemple : Quelles sont les caractéristiques du mode de vie qui constituent des facteurs de risques coronariens ? Sur un échantillon de patients choisis en fonction de leur statut de fumeur, leur régime alimentaire, leur consommation d'alcool et leur historique cardiaque, vous pouvez construire un modèle à l'aide de quatre variables du mode de vie pour expliquer la présence ou l'absence de déficiences coronariennes sur l'échantillon de patients. Le modèle peut alors servir à dériver les prévisions des rapports des cotes pour chaque facteur afin de vous indiquer, par exemple, que les fumeurs sont plus susceptibles de développer des déficiences coronariennes que les non-fumeurs.

Statistiques : Pour chaque analyse : observations totales, observations sélectionnées, observations valides. Pour chaque variable catégorielle : codage de paramètre. Pour chaque pas : variable(s) introduites ou éliminées, historique des itérations, log de vraisemblance -2 , qualité de l'ajustement, statistique de qualité d'ajustement de Hosmer-Lemeshow, khi-deux du modèle, khi-deux d'amélioration, table de classification, corrélations entre variables, groupes observés et graphique des probabilités prévues, khi-deux résiduel. Pour chaque variable de l'équation : coefficient (B), erreur standard de B , statistique de Wald, rapport des cotes estimé ($\exp(B)$), intervalle de confiance pour $\exp(B)$, log de vraisemblance si un terme a été éliminé du modèle. Pour chaque variable hors de l'équation : statistiques de scores. Pour chaque observation : groupe observé, probabilité prédite, groupe prévu, résidu, résidu standard.

Méthodes : Vous pouvez estimer des modèles à l'aide des entrées en bloc de variables ou de n'importe laquelle des méthodes détaillées pas à pas suivantes : ascendante conditionnelle, ascendante rapport de vraisemblance, ascendante Wald, descendante conditionnelle, descendante rapport de vraisemblance, descendante Wald.

Remarques sur les données de régression logistique

Données : Les variables dépendantes et indépendantes doivent être dichotomiques. Les variables indépendantes peuvent être de niveaux d'intervalles ou des variables catégorielles. Dans ce dernier cas, elles doivent être factices ou codées numériquement (il existe une option dans la procédure pour recoder les variables catégorielles automatiquement).

Hypothèses : La régression logistique ne s'appuie pas sur des hypothèses de distribution au même sens que l'analyse discriminante. Cependant, votre solution peut être plus stable si vos prédicteurs suivent une distribution multivariée gaussienne. De surcroît, comme avec les autres formes de régression, la multicolinéarité parmi les prédicteurs peut entraîner une altération des estimations et l'augmentation des erreurs standard. La procédure est plus efficace lorsque l'appartenance au groupe est une variable purement catégorielle, si l'appartenance au groupe est fondée sur des valeurs d'une variable continue (par exemple "QI élevé" opposé à "QI faible"), vous devez envisager d'utiliser la régression linéaire pour profiter de la richesse des informations offertes par la variable continue elle-même.

Procédures apparentées : Utilisez le nuage de points pour étudier la multicolinéarité de vos données. Si les hypothèses de normalité multivariées et d'égalité des matrices de variance/covariance sont satisfaites, vous devez obtenir une solution plus rapide à l'aide de la procédure d'analyse discriminante. Si toutes vos variables de prédicteur sont catégorielles, vous pouvez également utiliser la procédure log-linéaire. Si votre variable dépendante est continue, utilisez la procédure de régression linéaire. Vous pouvez utiliser la procédure Courbe ROC pour tracer sous forme graphique les probabilités enregistrées avec la procédure Régression logistique.

Obtenir une analyse de la régression logistique

1. À partir des menus, sélectionnez :

Analyser la > régression > logistique binaire...

2. Sélectionnez une Variable dépendante dichotomique. Il peut s'agir d'une variable numérique ou d'une chaîne.
3. Sélectionnez une ou plusieurs covariables. Pour ajouter des termes d'interaction, sélectionnez toutes les variables impliquées dans l'interaction, puis sélectionnez **>a*b>**.

Pour saisir les variables en groupe (**blocs**), sélectionnez les covariables pour un bloc, puis cliquez sur **Suivant** pour spécifier un nouveau bloc. Répétez jusqu'à ce que tous les blocs soient spécifiés.

Vous pouvez éventuellement sélectionner des observations pour analyse. Choisissez une variable de sélection, puis entrez les critères de règle.

Définition de la règle de régression logistique

Les observations définies par la règle de sélection sont incluses dans l'estimation du modèle. Par exemple, si vous avez sélectionné une variable ainsi que l'opérateur **égal à** et que vous avez spécifié la valeur 5, seules les observations pour lesquelles la variable sélectionnée a une valeur égale à 5 sont incluses dans l'estimation du modèle.

Les résultats des statistiques et de classification sont générés pour les observations sélectionnées et celles qui ne le sont pas. Cette procédure met en oeuvre un mécanisme de classification des nouvelles observations à partir des données précédemment existantes, ou de partitionnement de vos données en sous-ensembles d'apprentissage et de test, afin d'effectuer la validation du modèle généré.

Méthodes de sélection des variables de régression logistique

La sélection d'une méthode vous permet de spécifier la manière dont les variables indépendantes sont entrées dans l'analyse. En utilisant différentes méthodes, vous pouvez construire divers modèles de régression à partir du même groupe de variables.

- *Entrer*. Procédure de sélection de variables au cours de laquelle toutes les variables d'un bloc sont introduites en une seule opération.
- *sélection ascendante (conditionnelle)*. Méthode de sélection pas à pas avec test d'entrée fondé sur la signification de la statistique de significativité et avec test de suppression fondé sur la probabilité d'une statistique du rapport de vraisemblance s'appuyant sur des estimations de paramètres conditionnels.
- *sélection ascendante (rapport de vraisemblance)*. Méthode de sélection pas à pas avec test d'entrée fondé sur la signification de la statistique de significativité et avec test de suppression fondé sur la probabilité d'une statistique du rapport de vraisemblance s'appuyant sur des estimations de vraisemblance partielle maximale.
- *sélection ascendante (Wald)*. Méthode de sélection pas à pas avec test d'entrée fondé sur la signification de la statistique de significativité et avec test de suppression fondé sur la probabilité de la statistique de Wald.
- *élimination descendante (conditionnelle)*. Sélection pas à pas descendante. Le test de suppression se base sur la probabilité du rapport de vraisemblance calculé à partir d'estimations de paramètres conditionnels.
- *élimination descendante (rapport de vraisemblance)*. Sélection pas à pas descendante. Le test de suppression se base sur la probabilité de la statistique du rapport de vraisemblance calculé à partir des estimations de vraisemblance partielle maximale.
- *élimination descendante (Wald)*. Sélection pas à pas descendante. Le test de suppression se base sur la probabilité de la statistique de Wald.

Les valeurs de signification dans vos sorties sont basées sur l'adéquation à un modèle unique. Par conséquent, les valeurs de signification ne sont généralement pas valides lorsqu'une méthode détaillée pas à pas est utilisée.

Toutes les variables indépendantes sélectionnées sont ajoutées dans un seul modèle de régression. Cependant, vous pouvez spécifier différentes méthodes d'entrée pour les sous-groupes de variables. Par exemple, vous pouvez entrer un bloc de variables dans le modèle de régression en utilisant la sélection pas à pas, et un second bloc en utilisant la sélection ascendante. Pour ajouter un second bloc de variables au modèle de régression, cliquez sur **Suivant**.

Régression logistique : définition des variables catégorielles

Vous pouvez spécifier les détails de la manière dont la procédure de régression logistique gère les variables catégorielles :

Covariables : Contient la liste de toutes les covariables spécifiées dans la boîte de dialogue principale, soit par elles-mêmes, soit comme partie d'une interaction, à n'importe quelle couche. Si certaines de ces covariables sont des variables de chaîne, vous pouvez utiliser des covariables qualitatives.

Covariables catégorielles : Etablit la liste de toutes les variables identifiées comme étant catégorielles. Chaque variable comprend une notation entre parenthèses indiquant la codification de contraste à utiliser. Les variables de chaîne (identifiées par le symbole < suivi de leurs noms) sont déjà présentes dans la liste des covariables catégorielles. Sélectionnez n'importe quelle autre covariable catégorielle à partir de la liste des covariables catégorielles.

Modifier le contraste : Permet de modifier la méthode de contraste. Les méthodes de contraste disponibles sont :

- **Indicateur** : Les contrastes indiquent la présence ou l'absence d'appartenance à la catégorie. La catégorie de référence est représentée par la matrice de contraste sous la forme d'une ligne de zéros.
- **Simple** : Chaque catégorie de la variable de prédicteur (hormis la catégorie de référence) est comparée à la catégorie de référence.
- **Différence** : Chaque catégorie de la variable de prédicteur (hormis la première catégorie) est comparée avec l'effet moyen des catégories précédentes. (Aussi connu sous le nom de contrastes inversés d'Helmert.)
- **Helmert** : Chaque catégorie de la variable de prédicteur (hormis la dernière catégorie) est comparée avec l'effet moyen des catégories suivantes.
- **Répété** : Chaque catégorie de la variable de prédicteur (hormis la dernière catégorie) est comparée à la catégorie suivante.
- **Polynomial** : Contraste polynomial orthogonal. On part de l'hypothèse que les modalités sont espacées de manière équivalente. Les contrastes polynomiaux sont utilisables pour les variables numériques seulement.
- **Déviaton** : Chaque catégorie de la variable de prédicteur (hormis la catégorie de référence) est comparée à l'effet global.

Si vous sélectionnez **Déviaton**, **Simple** ou **Indicateur**, sélectionnez **Première** ou **Dernière** comme catégorie de référence. Remarquez que vous ne changez pas réellement de méthode avant de cliquer sur **Changer**.

Les covariables de chaîne doivent impérativement être des covariables qualitatives. Pour supprimer une variable de chaîne de la liste des covariables catégorielles, vous devez supprimer tous les termes contenant cette variable de la liste des covariables de la boîte de dialogue principale.

Régression logistique : enregistrer les nouvelles variables

Vous pouvez enregistrer les résultats de la régression logistique sous forme de nouvelles variables dans le jeu de données actif :

Prévisions : Enregistre les valeurs prévues par le modèle. Les options disponibles sont Probabilités et Groupe d'affectation.

- *Probabilités*. Enregistre pour chaque observation la probabilité d'occurrence prévue de l'événement. Dans les sorties, un tableau affiche le nom et le contenu de toutes les nouvelles variables.

L'"événement" est la catégorie de la variable dépendante avec la valeur supérieure ; par exemple, si la variable dépendante prend les valeurs 0 et 1, la probabilité prédite de la catégorie 1 est enregistrée.

- *Appartenance au groupe prévu.* Groupe ayant la probabilité a posteriori la plus élevée, basé sur les scores discriminants. Le groupe prévu par le modèle est celui auquel appartient l'observation.

Influence : Enregistre des valeurs à partir des statistiques qui mesurent l'influence des observations sur les prévisions. Les options disponibles sont Statistique de Cook, Bras de levier et Différence de bêta.

- *Cook.* Régression logistique analogue aux statistiques d'influence de Cook. Mesure permettant de savoir de combien les résidus de toutes les observations seraient modifiés si une observation donnée était exclue du calcul des coefficients de régression.
- *valeur influente.* Mesure de l'influence d'un point sur l'ajustement de la régression.
- *DfBetas.* La différence de bêta correspond au changement du coefficient de régression qui résulte de l'exclusion d'une observation particulière. Une valeur est calculée pour chaque terme du modèle, y compris la constante.

Résidus : Enregistre les résidus. Les options disponibles sont Non standardisés, Logit, Studentisés, Standardisés et Déviance.

- *résiduels non standardisés.* Différence entre une valeur observée et la valeur prévue par le modèle.
- *Résidu Logit.* Résidu de l'observation lorsqu'il est prévu dans l'échelle de logit. Le résidu logit est le résidu divisé par la probabilité prévue fois 1, moins la probabilité prévue.
- *résidu de Student.* Modification de la déviance du modèle lorsqu'une observation est exclue.
- *Résidus standardisés.* Résidu, divisé par une estimation de son écart type. Egalement appelés résiduels de Pearson, les résiduels standardisés ont une moyenne de 0 et un écart type de 1.
- *Deviance.* Résidus fondés sur la déviance du modèle.

Exporter les informations du modèle dans un fichier XML : Les estimations de paramètres et leurs covariances (facultatif) sont exportées vers le fichier spécifié au format XML (PMML). Vous pouvez utiliser ce fichier de modèle pour appliquer les informations du modèle aux autres fichiers de données à des fins d'évaluation. .

Options de régression logistique

Vous pouvez sélectionner les options suivantes pour votre analyse :

Tracés et statistiques : Vous permet de demander statistiques et tracés. Les options disponibles sont Tracés de classement, Qualité d'ajustement d'Hosmer-Lemeshow, Liste des résidus par observation, Corrélations des estimations, Historique des itérations et CI pour $\exp(B)$. Sélectionnez l'une des options dans le groupe Affichage pour consulter les statistiques et les tracés soit A chaque étape, soit uniquement pour le modèle final, A la dernière étape.

- *Statistique de qualité d'ajustement de Hosmer-Lemeshow.* Cette statistique de qualité d'ajustement est plus robuste que la statistique de qualité d'ajustement traditionnellement utilisée pour la régression logistique, particulièrement pour les modèles ayant des covariables continues et les études d'échantillons de petite taille. Elle est basée sur le regroupement des observations en déciles de risque et la comparaison de la probabilité observée avec la probabilité théorique à l'intérieur de chaque décile.

Probabilité dans étape par étape : Vous permet de contrôler les critères d'insertion ou de suppression des variables dans l'équation. Vous pouvez spécifier les critères d'insertion ou de suppression des variables.

- *Probabilité pour la méthode détaillée étape par étape.* Une variable est introduite dans un modèle si la probabilité de sa statistique de score est inférieure à la valeur Entrée et elle est éliminée si la probabilité est supérieure à la valeur Suppression. Pour remplacer les paramètres par défaut, indiquez des valeurs entières positives pour Entrée et Suppression. La valeur Entrée doit être inférieure à Suppression.

Césure de classification. Vous permet de définir la césure pour les observations de la classification. Les observations avec des prévisions qui excèdent la limite de classification sont classées positives tandis

que celles dont les prévisions sont inférieures à la limite sont classées négatives. Pour modifier la valeur par défaut, entrez une valeur entre 0.01 et 0.99.

Itérations maximales. Vous permet de modifier le nombre maximal d'itérations du modèle avant interruption.

Inclure la constante dans le modèle. Vous permet d'indiquer si le modèle doit inclure un terme constant. Si cette option est désactivée, la constante est égale à 0.

Fonctions supplémentaires de la commande REGRESSION LOGISTIQUE

Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Identifier la sortie en fonction des observations par les valeurs ou les libellés d'une variable.
- Contrôler l'espacement des rapports d'itération. Plutôt que d'imprimer les estimations après chaque itération, vous pouvez demander les estimations après chaque *énième* itération.
- Modifier les critères d'interruption d'une itération et de contrôle de la redondance.
- Spécifier une liste de variables pour les listes par observations.
- Garder une trace en plaçant les données de chaque groupe de fichiers scindés dans un fichier vierge au cours du traitement.

Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Régression logistique multinomiale

La régression logistique multinomiale est utile dans le cas où vous souhaitez classer des sujets en fonction des valeurs d'un groupe de variables de prédicteur. Ce type de régression est similaire à la régression logistique, mais s'avère plus général puisque la variable dépendante n'est pas limitée à deux catégories.

Exemple : Afin de mieux rentabiliser la commercialisation de leurs films, les studios souhaitent prévoir le type de film que les cinéphiles sont susceptibles d'aller voir. En effectuant une régression logistique multinomiale, le studio peut déterminer l'impact de l'âge, du sexe et de la situation de famille d'une personne sur les types de films qu'elle préfère. Le studio peut alors orienter la campagne promotionnelle d'un film particulier en fonction du groupe de spectateurs susceptibles d'aller le voir.

Statistiques : Historique des itérations, coefficients de paramètre, covariance asymptotique et matrices de corrélation, tests du rapport de vraisemblance pour les effets de modèle et les effets partiels, log de vraisemblance -2 . Qualité d'ajustement du khi-deux de Pearson et de déviance. R^2 de Cox et Snell, de Nagelkerke et de McFadden. Classification : effectifs observés par rapport aux effectifs prédits par catégorie de réponse. Tableaux croisé : effectifs observés et prédits (avec résidus) et proportions par motif de covariables et par catégorie de réponse.

Méthodes : Un modèle logit multinomial est ajusté pour le modèle factoriel complet ou pour un modèle défini par l'utilisateur. L'estimation des paramètres est effectuée au moyen d'un algorithme itératif calculant le maximum de vraisemblance.

Remarques sur les données de régression logistique multinomiale

Données : La variable dépendante doit être catégorielle. Les variables indépendantes peuvent correspondre à des facteurs ou à des covariables. En général, les facteurs doivent être des variables catégorielles et les covariables, des variables continues.

Hypothèses : On suppose que les rapports des cotes de deux catégories quelconques sont indépendants de toutes les autres catégories de réponse. Par exemple, lorsqu'un nouveau produit est introduit sur un marché, ce postulat signifie que les parts de marché de tous les autres produits sont toutes affectées proportionnellement de la même façon. En outre, d'après un motif de covariable, les réponses sont supposées correspondre à des variables multinomiales indépendantes.

Obtention d'une régression logistique multinomiale

1. A partir des menus, sélectionnez :

Analyser > Régression > Régression logistique multinomiale ...

2. Sélectionnez une variable dépendante.

3. Les facteurs sont facultatifs et peuvent être numériques ou catégoriels.

4. Les covariables sont facultatives, mais doivent être numériques si elles sont spécifiées.

Régression logistique multinomiale

Par défaut, la procédure Régression logistique multinomiale crée un modèle contenant des effets principaux de covariable et de facteur, mais vous pouvez spécifier un modèle personnalisé ou choisir un modèle pas à pas dans cette boîte de dialogue.

Spécifier le modèle : Un modèle comportant des effets principaux contient des effets principaux de covariable et de facteur, mais aucun effet d'interaction. Un modèle factoriel complet contient tous les effets principaux et toutes les interactions entre facteurs. Il ne contient pas de d'interactions de covariable. Vous pouvez créer un modèle personnalisé pour définir des sous-groupes d'interactions entre facteurs ou de covariables, ou demander une sélection pas à pas de termes de modèle.

Facteurs & covariables. Les facteurs et les covariables sont répertoriés.

Termes de l'introduction forcée : Les termes ajoutés à la liste d'introduction forcée sont systématiquement inclus dans le modèle.

Termes étape par étape : Les termes ajoutés à la liste pas à pas sont inclus dans le modèle, en fonction de l'une des méthodes détaillées pas à pas suivantes sélectionnées par l'utilisateur :

- **Introduction ascendante :** A la première étape de cette méthode, le modèle ne contient aucun terme pas à pas. A chaque étape, le terme le plus significatif est ajouté au modèle jusqu'à ce qu'aucun terme pas à pas exclu du modèle n'ait de contribution statistiquement significative s'il est inséré dans ce modèle.
- **Elimination descendante :** La première étape de cette méthode consiste à insérer dans le modèle tous les termes de la liste pas à pas. A chaque étape, le terme pas à pas le moins significatif est supprimé du modèle jusqu'à ce que tous les termes pas à pas restants aient une contribution statistiquement significative pour ce modèle.
- **Pas à pas ascendante :** La première étape de cette méthode consiste à sélectionner le modèle par la méthode d'introduction ascendante. A partir de là, l'algorithme alterne entre élimination descendante des termes pas à pas du modèle et introduction ascendante des termes exclus de ce modèle. Ce processus se poursuit jusqu'à ce que plus aucun terme ne réponde aux critères d'ajout ou de suppression.
- **Pas à pas descendante :** La première étape de cette méthode consiste à sélectionner le modèle par la méthode d'élimination descendante. A partir de là, l'algorithme alterne entre introduction ascendante des termes exclus du modèle et élimination descendante des termes pas à pas de ce modèle. Ce processus se poursuit jusqu'à ce que plus aucun terme ne réponde aux critères d'ajout ou de suppression.

Inclure ordonnée à l'origine dans le modèle : Cette option vous permet d'inclure ou d'exclure une constante pour le modèle.

Termes construits

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau de toutes les variables sélectionnées.

Effets principaux : Crée un terme d'effet principal pour chaque variable sélectionnée.

Toutes d'ordre 2 : Crée toutes les interactions d'ordre 2 possibles des variables sélectionnées.

Toutes d'ordre 3 : Crée toutes les interactions d'ordre 3 possibles des variables sélectionnées.

Toutes d'ordre 4 : Crée toutes les interactions d'ordre 4 possibles des variables sélectionnées.

Toutes d'ordre 5 : Crée toutes les interactions d'ordre 5 possibles des variables sélectionnées.

Régression logistique multinomiale : Catégorie de référence

Par défaut, la procédure Régression logistique multinomiale utilise la dernière catégorie comme catégorie de référence. Cette boîte de dialogue vous permet de contrôler la catégorie de référence et le type de tri des catégories.

Catégorie de référence : Spécifiez la première ou la dernière catégorie, ou une catégorie personnalisée.

Ordre des catégories : Dans l'ordre croissant, la valeur minimale définit la première catégorie et la valeur maximale, la dernière catégorie. Dans l'ordre décroissant, la valeur maximale définit la première catégorie et la valeur minimale, la dernière catégorie.

Régression logistique multinomiale : Statistiques

Les statistiques pouvant être définies pour la régression logistique multinomiale sont les suivantes :

Récapitulatif du traitement des observations : Ce tableau contient les informations relatives aux variables catégorielles fournies.

Modèle : Statistiques du modèle global.

- **Pseudo R-carré :** Imprime les statistiques R^2 de Cox et Snell, de Nagelkerke et de McFadden.
- **Récapitulatif des pas :** Ce tableau récapitule les effets ajoutés à chaque étape d'une méthode détaillée pas à pas ou supprimés de cette dernière. Il n'est créé que si un modèle pas à pas est spécifié dans la boîte de dialogue Modèle.
- **Informations sur l'ajustement du modèle :** Ce tableau compare les modèles ajustés et les modèles avec constante seulement ou les modèles nuls.
- **Critères d'information :** Ce tableau imprime le critère d'information d'Akaike (AIC) et le critère d'information bayésien de Schwarz (BIC).
- **Probabilités des cellules :** Imprime un tableau des effectifs observés et des effectifs théoriques (avec résidu), et des proportions par motif de covariable et par catégorie de réponse.
- **Table de classification :** Imprime un tableau comparatif des réponses observées et des réponses prédites.
- **Qualité d'ajustement de statistiques de khi-carré :** Imprime les statistiques khi-carré de Pearson et khi-carré du rapport de vraisemblance. Les statistiques sont calculées pour les motifs de covariable déterminés par tous les facteurs et covariables ou par un sous-ensemble de facteurs et de covariables défini par l'utilisateur.
- **Mesures de monotonie :** Affiche un tableau contenant des informations sur les nombres de paires concordantes, de paires discordantes et de paires liées. Le D de Somers, le Gamma de Goodman et Kruskal, le Tau-a de Kendall et l'Indice de concordance C apparaissent également dans ce tableau.

Paramètres : Statistiques liées aux paramètres du modèle.

- **Estimations :** Imprime les estimations des paramètres du modèle, avec un niveau de confiance défini par l'utilisateur.
- **Test du rapport de vraisemblance :** Imprime les tests du rapport de vraisemblance pour les effets partiels du modèle. Le test du modèle global est imprimé automatiquement.
- **Corrélations asymptotiques :** Imprime la matrice de corrélation des estimations de paramètres.
- **Covariances asymptotiques :** Imprime la matrice de covariance des estimations de paramètres.

Définir les sous-populations : Cette option vous permet de sélectionner un sous-ensemble de facteurs et de covariables afin de définir les motifs de covariable utilisés par les probabilités des cellules et par les tests de qualité d'ajustement.

Régression logistique multinomiale : Critères

Les critères pouvant être définis pour la régression logistique multinomiale sont les suivants :

Itérations : Cette option vous permet d'indiquer le nombre de fois où vous souhaitez répéter l'algorithme, le nombre maximal d'étapes de la procédure de méthode dichotomique, les tolérances de convergence relatives aux modifications du log de vraisemblance et des paramètres, l'effectif des impressions de l'état d'avancement de l'algorithme itératif, ainsi que l'itération à laquelle la procédure doit commencer à rechercher une séparation complète ou quasi-complète des données.

- **Convergence de log de vraisemblance** : La convergence est supposée si la variation absolue de la fonction log de vraisemblance est inférieure à une valeur donnée. Le critère n'est pas utilisé si la valeur est 0. Spécifiez une valeur non négative.
- **Convergence des paramètres** : La convergence est prise en compte si la modification absolue des estimations du paramètre est inférieure à cette valeur. Le critère n'est pas utilisé si la valeur est 0.

Delta : Permet de spécifier une valeur non négative inférieure à 1. Cette valeur est ajoutée à chaque cellule vide du tableau croisé de la catégorie de réponse par le modèle de covariable. Ceci vous permet de stabiliser l'algorithme et d'éviter les estimations biaisées.

Tolérance de singularité : Cette option vous permet de définir la tolérance utilisée lors du contrôle des particularités.

Options de régression logistique multinomiale

Les statistiques pouvant être définies pour la régression logistique multinomiale sont les suivantes :

Echelle de dispersion : Cette option vous permet de définir la valeur d'échelle de dispersion qui sera utilisée pour corriger l'estimation de la matrice de covariance des paramètres. L'option **Déviance** estime la valeur d'échelle au moyen de la statistique fonction de déviance (khi-deux du rapport de vraisemblance). L'option **Pearson** estime la valeur d'échelle à l'aide de la statistique khi-carré de Pearson. Vous pouvez également spécifier votre propre valeur d'échelle. Il doit s'agir d'une valeur numérique positive.

Options étape par étape : Ces options vous permettent de contrôler les critères statistiques lorsque des méthodes détaillées pas à pas servent à créer un modèle. Elles sont ignorées sauf si un modèle pas à pas est spécifié dans la boîte de dialogue Modèle.

- **Probabilité d'entrée** : Il s'agit de la probabilité de la statistique du rapport de vraisemblance pour l'introduction de variables. La facilité avec laquelle une variable est ajoutée au modèle dépend directement de la valeur de la probabilité fournie. Plus cette valeur est élevée, plus la variable a de chances d'être insérée dans le modèle. Ce critère est ignoré sauf si la méthode d'introduction ascendante, ou la méthode détaillée pas à pas ascendante ou descendante est sélectionnée.
- **Test de saisie** : Il s'agit de la méthode permettant de saisir des termes selon des méthodes détaillées pas à pas. Choisissez entre le test du rapport de vraisemblance et le test de score. Ce critère est ignoré sauf si la méthode d'introduction ascendante, ou la méthode détaillée pas à pas ascendante ou descendante est sélectionnée.
- **Probabilité de suppression** : Il s'agit de la probabilité de la statistique du rapport de vraisemblance pour la suppression de variables. La facilité avec laquelle une variable est conservée dans le modèle dépend directement de la valeur de la probabilité fournie. Plus cette valeur est élevée, plus la variable a de chances de rester dans le modèle. Ce critère est ignoré sauf si la méthode d'élimination descendante, ou la méthode détaillée pas à pas ascendante ou descendante est sélectionnée.
- **Test de suppression** : Il s'agit de la méthode permettant d'éliminer des termes selon des méthodes détaillées pas à pas. Choisissez entre le test du rapport de vraisemblance et le test de Wald. Ce critère est ignoré sauf si la méthode d'élimination descendante, ou la méthode détaillée pas à pas ascendante ou descendante est sélectionnée.
- **Effets étape par étape minimum dans le modèle** : Lorsque la méthode détaillée pas à pas descendante ou la méthode d'élimination descendante est utilisée, cette option spécifie le nombre minimal de termes à inclure dans le modèle. La constante n'est pas considérée comme terme de modèle.

- **Effets étape par étape maximum dans le modèle :** Lorsque la méthode détaillée pas à pas ascendante ou la méthode d'introduction ascendante est utilisée, cette option spécifie le nombre maximal de termes à inclure dans le modèle. La constante n'est pas considérée comme terme de modèle.
- **Appliquer une contrainte hiérarchique à l'entrée et à la suppression des termes :** Cette option vous permet d'indiquer si des restrictions doivent s'appliquer à l'ajout de termes de modèle. La hiérarchie exige que, pour tout terme à inclure, l'ensemble des termes de niveau inférieur appartenant à ce terme figure avant tout dans le modèle. Par exemple, si cette exigence de la hiérarchie est appliquée, les facteurs *Situation familiale* et *Sexe* doivent être contenus dans le modèle pour que l'interaction *Situation familiale*Sexe* puisse être ajoutée. Les trois boutons radio déterminent le rôle que jouent les covariables dans l'établissement de la hiérarchie.

Régression logistique multinomiale : Enregistrer

La boîte de dialogue Enregistrer vous permet d'enregistrer des variables dans le fichier de travail et d'exporter les informations du modèle vers un fichier externe.

Variables enregistrées : Les variables pouvant être enregistrées sont les suivantes :

- **Probabilités des réponses estimées :** Il s'agit des probabilités estimées de classement d'un motif de facteur/covariable dans les catégories de réponse. Il y a autant de probabilités estimées que de catégories de variables de réponse ; jusqu'à 25 probabilités seront enregistrées.
- **Catégorie estimée :** Il s'agit de la catégorie de réponse dont le nombre de probabilités théorique est le plus élevé pour un motif de facteur/covariable.
- **Probabilité de catégorie estimée :** Il s'agit du nombre maximum de probabilités de réponses estimées.
- **Probabilité de catégorie actuelle :** Il s'agit de la probabilité estimée sur un motif de classement d'un motif de facteur/covariable dans la catégorie observée.

Exporter les informations du modèle dans un fichier XML : Les estimations de paramètres et leurs covariances (facultatif) sont exportées vers le fichier spécifié au format XML (PMML). Vous pouvez utiliser ce fichier de modèle pour appliquer les informations du modèle aux autres fichiers de données à des fins d'évaluation. .

Fonctions supplémentaires de la commande NOMREG

Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Spécifier la catégorie de référence de la variable dépendante.
- Inclure les observations avec valeurs manquantes de l'utilisateur.
- Personnaliser les tests d'hypothèse en spécifiant des hypothèses nulles comme combinaisons linéaires de paramètres.

Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Régression probit

Cette procédure mesure la relation entre l'intensité d'un stimulus et la proportion des observations montrant une certaine réponse au stimulus. Elle est utile lorsque vous avez une sortie dichotomique qu'on pense être influencée ou causée par des niveaux de certaines variables indépendantes et être particulièrement bien adaptée aux données expérimentales. Elle est de ce fait bien adaptée aux données expérimentales. Cette procédure vous permet d'estimer la force d'un stimulus requise pour induire une certaine proportion de réponses, telle que la dose médiane efficace.

Exemple : Quelle est l'efficacité d'un nouveau pesticide contre les fourmis et quelle concentration doit-on utiliser ? Vous devez mener une expérience dans laquelle vous exposez des échantillons de fourmis à différentes concentrations de pesticide et vous enregistrez le nombre de fourmis tuées et le nombre de fourmis exposées. En appliquant la méthode des probits à ces données, vous pouvez déterminer la force de la relation entre la concentration du produit et l'élimination de fourmis, ainsi que la concentration de pesticide nécessaire si vous souhaitez être sûr de vous débarrasser de, disons, 95 % des fourmis exposées.

Statistiques : Coefficients de régression et erreurs standard, constante et erreur standard, khi-carré de la qualité de l'ajustement de Pearson, fréquences attendues et théoriques et intervalle de confiance pour les niveaux efficaces des variables indépendantes. Tracés : tracés de réponse transformés.

Cette procédure utilise les algorithmes proposés et implémenté dans NPSOL[®] par Gill, Murray, Saunders & Wright pour estimer les paramètres du modèle.

Remarques sur les données de régression probit

Données : Pour chaque valeur de la variable indépendante (ou chaque combinaison de valeurs de plusieurs variables indépendantes), votre variable de réponse doit être l'effectif du nombre d'observations avec celles des valeurs qui montrent la réponse d'intérêt, et la variable observée totale doit être un effectif du nombre total d'observations avec celles des valeurs de la variable indépendante. Le facteur doit être catégoriel, codé sous la forme de nombres entiers.

Hypothèses : Les observations doivent être indépendantes. Si vous avez un grand nombre de valeurs pour les variables indépendantes relatives au nombre d'observations, comme c'est peut-être le cas dans votre étude, le khi-deux et les statistiques de qualité d'ajustement ne sont peut-être pas valables.

Procédures apparentées : Les analyses par la méthode des probits sont étroitement liés à la régression logistique. En fait, si vous sélectionnez la transformation logit, cette procédure calculera essentiellement une régression logistique. En général, les analyses par la méthode des probits s'adaptent à des plans d'expériences, tandis que la régression logistique est plus appropriée pour des études par observation. Les différences au niveau de la sortie reflètent ces différentes emphases. La procédure des analyses par la méthode des probits offre des estimations de valeurs effectives pour divers niveaux de réponse (incluant la dose effective médiane), tandis que la procédure de la régression logistique offre des estimations des rapports des cotes pour les variables indépendantes.

Obtention d'une analyse de régression probit

1. À partir des menus, sélectionnez :

Analyser > Régression > Probit...

2. Sélectionnez une variable de fréquence des réponses. Cette variable indique le nombre d'observations présentant une réponse au stimulus test. Les valeurs de cette variable ne peuvent pas être négatives.
3. Sélectionnez une variable totale observée. Cette variable indique le nombre d'observations auxquelles le stimulus a été appliqué. Les valeurs de cette variable ne peuvent pas être négatives et ne peuvent pas être inférieures à la variable de fréquence de réponse pour chaque observation.

Vous pouvez également sélectionner un facteur. Dans ce cas, utilisez **Définir plage** pour définir la plage pour les groupes.

4. Sélectionnez une ou plusieurs covariables. Cette variable contient le niveau du stimulus appliqué à chaque observation. Si vous désirez transformer la covariable, sélectionnez une transformation dans la liste déroulante **Transformation**. Si vous n'appliquez aucune transformation et qu'il existe un groupe de contrôle, ce groupe de contrôle est alors inclus dans l'analyse.
5. Sélectionnez le modèle **Probit** ou le modèle **Logit**.

Modèle Probit

Applique la transformation Probit (inverse de la fonction de distribution normale standard cumulée) aux proportions de réponses.

modèle logit

Applique la transformation logit (probabilités logarithmiques) aux proportions de réponses.

Analyse par la méthode des probits : définir une plage

Cela vous permet de spécifier les facteurs qui seront analysés. Les niveaux de facteur doivent être codés sous la forme de nombres entiers consécutifs, et tous les niveaux que vous indiquez doivent être analysés.

Options des analyses par la méthode des probits

Vous pouvez spécifier certaines options pour vos analyses par la méthode des probits :

Statistiques : Vous permet de demander les options statistiques suivantes : Effectifs, Impact relatif médian, Test de parallélisme et Intervalles de confiance de référence.

- *Impact relatif médian.* Affiche le rapport des impacts moyens pour chaque paire de niveaux de facteurs. Montre également les intervalles de confiance à 95 % pour chacun des impacts relatifs médians. Les impacts relatifs médians ne sont pas disponibles s'il n'y a pas de facteur ou s'il existe plusieurs covariables.
- *test de parallélisme.* Test de l'hypothèse selon laquelle tous les niveaux de facteur ont une pente commune.
- *intervalles de confiance de référence.* Intervalles de confiance pour que le dosage de l'agent nécessaire produise une certaine probabilité de réponse.

Les intervalles de confiance de référence et l'impact relatif médian ne sont pas disponibles si vous avez sélectionné plus d'une covariable. L'impact relatif médian et le test de parallélisme sont disponibles uniquement lorsque vous avez sélectionné un facteur.

Taux de réponse naturel : Vous permet d'indiquer un taux de réponse naturel même en l'absence de stimulus. Les options disponibles sont Aucun, Calculer à partir des données ou Valeur.

- *calculer à partir des données.* Estime le taux de réponse naturel à partir des données de l'échantillon. Vos données doivent contenir un cas représentant le niveau de contrôle, pour lequel la valeur des covariables est 0. Probit estime le taux de réponse naturel en utilisant la proportion de réponses pour le niveau de contrôle comme valeur initiale.
- *Valeur.* Définit le taux de réponse naturel dans le modèle (sélectionnez cet élément lorsque vous connaissez le taux de réponse naturel à l'avance). Tapez la proportion de réponse naturelle (cette proportion doit être inférieure à 1). Si, par exemple, une réponse existe dans 10 % des cas lorsque le stimulus est de 0, tapez 0,10.

Critères. Vous permet de contrôler les paramètres de l'algorithme itératif d'estimations. Vous pouvez passer outre les valeurs par défaut pour le maximum des itérations, la stabilité des coefficients et la précision à l'optimum.

Fonctions supplémentaires de la commande NLR

Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Demander une analyse parmi les deux analyses, Probit et Logit.
- Contrôler le traitement des valeurs manquantes.
- Transformer les covariables par des bases différentes de la base 10 ou des logarithmes naturels.

Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Régression quantile

La régression est une méthode statistique couramment utilisée en modélisation quantitative. La régression linéaire multiple est une approche de base standard où les chercheurs utilisent les valeurs de plusieurs variables pour expliquer ou prédire les valeurs d'une échelle de résultats. Cependant, dans de nombreux cas, nous sommes davantage intéressés par la médiane, ou un quantile arbitraire de l'échelle de résultat.

La régression quantile modélise la relation entre un ensemble de variables de prédicteur (indépendantes) et des percentiles spécifiques (ou "quantiles") d'une variable cible (dépendante), le plus souvent la médiane. Cette méthode présente deux principaux avantages par rapport à la régression aux moindres carrés :

- Elle n'émet aucune hypothèse sur la distribution de la variable cible.
- Elle tend à résister à l'influence des observations aberrantes.

Elle est couramment utilisée pour les recherches dans des secteurs tels que l'écologie, la santé et l'économie financière.

Exemple

Quelle est la relation entre le revenu total du ménage et la proportion du revenu consacrée à l'alimentation ? La loi d'Engel est une observation en économie selon laquelle plus le revenu augmente, plus la part du revenu consacrée à l'alimentation diminue, même si les dépenses absolues en nourriture augmentent. Appliquée à ces données, la régression quantile permet de déterminer la dépense alimentaire qui concerne 90 % des familles (pour 100 familles disposant d'un revenu donné) lorsque la dépense alimentaire moyenne n'est pas pertinente.

Statistiques

Régression quantile, approche simplex, algorithme d'optimisation non linéaire de points intérieurs de Frisch et Newton, Barrodale et Roberts, Bofinger, Hall Sheather, bande passante, niveau de signification, manipulation de matrices, critère de convergence, pondérations de régression, terme de la constante, cible prévue, résidus de prévision, mise en tableau, tracés de prévisions, estimations de paramètres, matrice de covariance, matrice de corrélation, valeurs observées, intervalle de confiance.

Cette procédure utilise les algorithmes proposés par Roger Koenker et George Basset (1978). Regression quantiles, *Econometrica*, 46, 33–50.

Remarques sur les données de régression quantile

Données

Une seule variable dépendante numérique est requise. La variable cible doit être une variable continue. Les prédicteurs peuvent être des variables continues ou des variables dummy pour des prédicteurs catégoriels. Pour exécuter une analyse, vous devez disposer du terme de constante ou d'au moins un prédicteur.

Hypothèses

La régression quantile n'émettez pas d'hypothèse sur la distribution de la variable cible et résiste à l'influence des observations aberrantes.

Procédures apparentées

L'analyse quantile est liée à la régression par la méthode des moindres carrés ordinaires (OLS).

Obtention d'une analyse de régression quantile

1. À partir des menus, sélectionnez :

Analyser > Régression > Quantile...

La boîte de dialogue permet d'indiquer les variables (cible, facteur, covariable et pondération) à utiliser pour l'analyse de régression quantile. Elle permet également de conserver de la mémoire pour les analyses complexes ou les jeux de données volumineux.

2. Sélectionnez une variable cible numérique. Une seule variable cible est requise pour exécuter une analyse. Seules les variables numériques sont autorisées.
3. Vous pouvez également sélectionner une ou plusieurs variables de facteur. Les variables d'échelle ne sont pas autorisées.
4. Vous pouvez également sélectionner une ou plusieurs variables de covariable. Les variables de chaîne ne sont pas autorisées.

Remarque : Lorsque les listes **Facteur(s)** et **Covariable(s)** sont vides et que **Inclure une constante dans le modèle** est sélectionné dans la boîte de dialogue Modèle, le message suivant s'affiche :

```
No effects have been specified. Therefore, an intercept only model will be fit.
Do you want to fit an intercept-only model?
```

5. Vous pouvez également sélectionner une variable de pondération de régression. Les variables de chaîne ne sont pas autorisées.
6. Eventuellement, sélectionnez **Conserver de la mémoire pour les analyses complexes et les jeux de données importants**. Ce paramètre indique si les données doivent être placées dans un fichier

externe pendant le traitement. Cela peut vous permettre d'économiser des ressources mémoire lorsque vous effectuez des analyses complexes ou avec des jeux de données volumineux.

Régression quantile : Critères

La boîte de dialogue Critères propose les options suivantes :

Quantile

Propose des options pour indiquer le ou les quantiles.

Spécifier des quantiles uniques

Si cette option est sélectionnée, vous devez entrer au moins une valeur pour exécuter l'analyse. Les valeurs multiples sont admises et chaque valeur doit être comprise dans l'intervalle $[0, 1]$. Vous pouvez spécifier plusieurs valeurs en les séparant par un ou plusieurs espaces. Utilisez les boutons **Ajouter**, **Changer** et **Retirer** pour utiliser les valeurs de la liste de valeurs de quantile.

Toutes les valeurs doivent être uniques (les valeurs en double sont interdites). La valeur par défaut est 0, 5.

Spécifier les quantiles de grille

Si cette option est sélectionnée, une grille de quantiles peut être spécifiée entre une valeur **Début** (value1) et une valeur **Fin** (value2) avec l'incrément **Par** (value3). Si cette option est spécifiée, un seul ensemble $[value1 \text{ TO } value2 \text{ BY } value3]$ valide est autorisé. Il doit satisfaire la condition $0 \leq value1 \leq value2 \leq 1$. Si $value1 = value2$, cela revient à spécifier une valeur value1 unique, quelle que soit la valeur value3.

Méthode d'estimation

Propose des options pour indiquer la méthode d'estimation du modèle.

Choisie automatiquement par le programme

Permet à la procédure de sélectionner automatiquement la méthode d'estimation appropriée. Il s'agit du paramètre par défaut.

Algorithme Simplex

Appelle l'algorithme Simplex développé par Barrodale et Roberts.

Optimisation non linéaire de points intérieurs de Frisch et Newton

Appelle l'algorithme d'optimisation non linéaire de points intérieurs de Frisch et Newton.

Post-estimation

Propose des options pour la post-estimation de la variance-covariance des estimations de paramètre et des intervalles de confiance pour les valeurs cible prévues.

Supposer que les observations sont IID

Sélectionné, ce paramètre permet de supposer que les termes d'erreur sont indépendants et identiquement distribués (IID). Lorsque ce paramètre n'est pas sélectionné, le temps de calcul peut augmenter considérablement pour les modèles volumineux. Ce paramètre est sélectionné par défaut.

Type de bande passante

Détermine la méthode de bande passante utilisée pour estimer la matrice variance-covariance des estimations de paramètre (**Bofinger** ou **Hall-Sheather**). **Bofinger** est le paramètre par défaut.

Méthode numérique

Propose les options suivantes :

Tolérance de singularité

Indique la valeur de tolérance pour les manipulations de matrice dans la méthode des points intérieurs. Il doit s'agir d'une seule valeur double précision comprise dans l'intervalle $(0, 10^{-3})$. La valeur par défaut est 10^{-12} .

Convergence

Indique le critère de convergence pour la méthode numérique. Il doit s'agir d'une seule valeur double précision comprise dans l'intervalle $(0, 10^{-3})$. La valeur par défaut est 10^{-6} .

Itérations maximales

Spécifie le nombre maximal d'itérations. Il doit s'agir d'un nombre entier positif. La valeur par défaut est 2000.

Valeurs manquantes

Permet d'indiquer comment traiter les valeurs manquantes.

Exclure les valeurs manquantes de l'utilisateur et les valeurs système manquantes

Lorsque cette option est sélectionnée, les valeurs manquantes de l'utilisateur et système sont exclues.

Les valeurs manquantes de l'utilisateur sont traitées comme des données valides.

Lorsque cette option est sélectionnée, les valeurs manquantes de l'utilisateur sont traitées comme des données non valides.

Intervalle de confiance (%)

Indique le niveau d'importance. Si vous indiquez ce paramètre, il doit s'agir d'une valeur double précision comprise entre 0 et 100. La valeur par défaut est 95.

Régression quantile : Modèle

La boîte de dialogue Modèle propose des options pour définir les effets et les pondérations utilisés dans le modèle. Omis ou auto-défini, le modèle contient le terme d'interception ainsi que tous les effets principaux, avec les covariables de la liste des covariables et les facteurs de la liste des facteurs.

Spécifier les effets du modèle

Le modèle par défaut est à constante seulement, c'est pourquoi vous devez spécifier explicitement les autres effets du modèle. Vous pouvez aussi construire des termes imbriqués ou non imbriqués. Lorsque **Termes construits** est sélectionné, l'effet et les options d'interaction ci-après sont disponibles pour les termes non imbriqués.

Effets principaux

Crée un terme d'effet principal pour chaque variable sélectionnée.

Interaction

Crée le terme d'interaction du plus haut niveau pour toutes les variables sélectionnées.

Factoriel

Crée toutes les interactions et les effets principaux possibles pour les variables sélectionnées.

Tous d'ordre 2

Crée toutes les interactions d'ordre 2 possibles des variables sélectionnées.

Tous d'ordre 3

Crée toutes les interactions d'ordre 3 possibles des variables sélectionnées.

Tout ordre 4

Crée toutes les interactions d'ordre 4 possibles des variables sélectionnées.

Tout ordre 5

Crée toutes les interactions d'ordre 5 possibles des variables sélectionnées.

Lorsque **construction de termes imbriqués** est sélectionné, vous pouvez construire des termes imbriqués. Les termes imbriqués sont utiles pour modéliser l'effet d'un facteur ou d'une covariable dont les valeurs n'interagissent pas avec les niveaux d'un autre facteur. Par exemple, une chaîne d'épicerie peut suivre les habitudes d'achat de ses clients dans divers magasins. Puisque chaque client ne fréquente qu'un seul de ces magasins, l'effet *Client* peut être considéré comme étant **imbriqué dans** l'effet *Emplacement des magasins*.

En outre, vous pouvez inclure des effets d'interaction, tels que des termes polynomiaux impliquant la même covariable, ou ajouter plusieurs niveaux d'imbrication au terme imbriqué.

Imbrication de notes :

- Afin d'inclure un effet relatif à une interaction entre des facteurs, utilisez le mot clé BY ou un astérisque (*) afin de joindre les facteurs impliqués dans l'interaction.
- Les facteurs d'un effet d'interaction doivent être distincts.

- Pour imbriquer un terme dans un autre, incluez l'effet entre une paire de parenthèses.
- En présence de plusieurs paires de parenthèses, placez ou imbriquez chaque paire dans une autre paire de parenthèses.
- L'imbrication multiple est autorisée.
- Les interactions entre effets imbriqués ne sont pas prises en charge.

Limitations : Les termes imbriqués comportent les restrictions suivantes :

- Tous les facteurs d'une interaction doivent être uniques. Ainsi, si A est un facteur, la spécification $A*A$ n'est pas valide.
- Tous les facteurs d'un effet imbriqué doivent être uniques. Ainsi, si A est un facteur, la spécification $A(A)$ n'est pas valide.
- Aucun effet ne peut être imbriqué dans un effet de covariable. Ainsi, si A est un facteur et X une covariable, la spécification $A(X)$ n'est pas valide.

Inclure l'interception dans le modèle

Lorsque cette option est sélectionnée, les termes de constante sont inclus dans le modèle. Lorsqu'elle n'est pas sélectionnée, vous devez entrer au moins un prédicteur pour exécuter l'analyse. Ce paramètre est activé par défaut.

Régression quantile : Affichage

La boîte de dialogue Affichage propose des paramètres de contrôle de tracé et de sortie.

Imprimer

Les options de sortie suivantes sont disponibles :

Estimations de paramètres

Affiche les estimations des paramètres, ainsi que les statistiques de test et les intervalles de confiance correspondants. Vous pouvez éventuellement afficher les estimations de paramètres exponentielles, en plus des estimations de paramètres brutes.

Matrice de covariance des estimations des paramètres

Affiche la matrice de covariance des paramètres estimés.

Matrice de corrélation des estimations des paramètres

Affiche la matrice de corrélation des paramètres estimés.

Tracé et tableau

Les options de traçage suivantes sont disponibles :

Tracer le graphique des estimations de paramètres pour

Vous pouvez tracer les estimations de paramètre pour un nombre précis d'effets principaux, ou pour tous les effets du modèle. Le paramètre identifiant les **xx effets principaux** permet de contrôler le nombre de catégories (ou de combinaisons de catégories) qui, dans un effet mixte, sont tracées dans l'interaction d'une covariable et d'un facteur (ou de deux facteurs). La valeur doit être un entier positif unique (50 est le paramètre par défaut).

Remarques :

- Les tracés de prévisions sont créés pour tous les effets lorsque la valeur entière indiquée est supérieure au nombre de catégories ou de combinaisons.
- Ce paramétrage n'est valide que lorsque plusieurs valeurs sont indiquées pour le paramètre **Valeurs quantiles** dans la boîte de dialogue Critères. Aucun tracé n'est créé lorsqu'un seul quantile est spécifié.

Afficher les prévisions par tracé observé

Permet de contrôler la création du tracé des valeurs observées par rapport aux prévisions.

Activée, cette option crée un tracé unique contenant les points (les quantiles étant représentés en pointillés de différentes couleurs). Il est désactivé par défaut.

Prédire les effets dans le modèle

Lorsque ce paramètre est activé, les options suivantes sont disponibles :

Présenter sous forme de graphique ou de tableau les x effets principaux

Spécifiez le nombre d'effets supérieurs à partir desquels le diagramme de prédiction ou le tableau de prédiction sera créé. 3 est la valeur par défaut.

Remarque : Les tracés (ou les tables) de prévisions sont créés pour tous les effets lorsque la valeur entière indiquée est supérieure au nombre d'effets valides dans le modèle.

Tracer les effets spécifiés par l'utilisateur ou les présenter sous forme de tableau

Voici les règles relatives aux effets valides :

- Effet à une seule covariable (y compris une puissance élevée de la covariable proprement dite) : créez un tracé unique contenant les lignes prévues par quantiles.
- Effet à un seul facteur : mettez en tableau les prévisions pour les catégories du facteur par quantiles.
- Effet avec interaction de deux facteurs : pour chaque quantile, mettez en tableau les prévisions pour les catégories des deux facteurs.
- Effet avec interaction d'une covariable et d'un ou de deux facteurs : pour chaque quantile, créez un tracé contenant les lignes pour chaque catégorie ou combinaison de catégories dans l'effet d'interaction.
- Le nombre maximal des combinaisons à tracer est contrôlé par la valeur indiquée dans **Présenter sous forme de graphique au maximum xx catégories ou combinaisons de catégories dans un effet mixte**.

Les effets déplacés de **Effets de modèle** vers la liste **Lignes de prévision** sont utilisés pour le traçage. Les tracés ne sont pas créés lorsque les effets spécifiés sont constants (supprimés de la génération de modèle).

Les effets déplacés depuis **Effets de modèle** vers la liste **Tables de prévision** sont utilisés pour la mise en tableau. Les tables ne sont pas créées lorsque les effets spécifiés sont constants (supprimés de la génération de modèle).

Présenter sous forme de graphique au maximum xx catégories ou combinaisons de catégories dans un effet mixte

Contrôle le nombre maximal de combinaisons de catégories à tracer. La valeur par défaut est 10.

Régression quantile : Enregistrer

La boîte de dialogue Enregistrer propose des options pour évaluer le modèle.

Valeur prévue de la réponse

Lorsque cette option est sélectionnée, la valeur cible prévue est évaluée.

Résiduel

Si cette option est sélectionnée, les résidus de prévision sont évalués.

Limite inférieure de l'intervalle de prévision

Lorsque cette option est sélectionnée, les limites inférieures des intervalles de prévision sont évaluées.

Limite supérieure de l'intervalle de prévision

Lorsque cette option est sélectionnée, les limites supérieures des intervalles de prévision sont évaluées.

Remarque : Un nom de variable peut être indiqué pour chaque option d'enregistrement. Si un nom de racine est spécifié, il doit s'agir d'un nom de variable valide. Un nom de racine, suivi d'un trait de soulignement "_" et d'un suffixe de quantile significatif sont utilisés lorsque plusieurs valeurs sont indiquées pour les **valeurs de quantile** dans la boîte de dialogue Critères.

Régression quantile : Exporter

La boîte de dialogue Exporter propose des options pour définir les statistiques à exporter et leur mode d'exportation (dans un fichier de données externe ou dans des jeux de données). Elle permet également de contrôler le mode de gestion des données au cours du traitement (traitées en mode normal ou conservées dans un fichier temporaire).

Matrice de covariance des estimations de paramètres

Sélectionnée, cette option active les options d'écriture de la matrice de covariance des estimations de paramètre dans un fichier de données externe ou dans un jeu de données déjà déclaré.

Matrice de corrélation des estimations de paramètres

Sélectionnée, cette option active les options d'écriture de la matrice de corrélation des estimations de paramètre dans un fichier de données externe ou dans un jeu de données déjà déclaré.

La matrice de covariance/corrélation sera enregistrée dans un seul jeu de données ou fichier externe en présence de plusieurs quantiles de régression

En présence de plusieurs quantiles, cette option permet d'activer l'enregistrement des matrices de covariance/corrélation dans un ou plusieurs fichiers de données externes ou jeux de données. Lorsqu'elle n'est pas activée, les matrices sont enregistrés dans un seul fichier de données externe ou jeu de données. Lorsque cette option est activée, les matrices sont enregistrées dans plusieurs fichiers de données externes ou jeux de données. Ce paramétrage n'est valide qu'en présence de plusieurs **valeurs de quantile** dans la boîte de dialogue Critères.

Remarque : Cette option n'est valide que lorsque **Matrice de covariance des estimations de paramètres** ou **Matrice de corrélation des estimations de paramètres** est sélectionnée.

Exporter les informations du modèle dans un fichier XML

Lorsqu'elle est sélectionnée, cette option propose des options pour exporter les informations du modèle dans un fichier XML (ayant un nom et un emplacement spécifiques).

Exporter au format XML

Lorsque l'option **Exporter les informations du modèle dans un fichier XML** est sélectionnée, vous pouvez exporter les estimations de paramètre et les matrices de covariance, ou les estimations de paramètre uniquement. La valeur par défaut est **Estimations des paramètres et matrice de covariance**.

Conventions relatives aux noms de fichier

- Lorsqu'une seule **valeur de quantile** est indiquée dans la boîte de dialogue Critères, `savefile` et `dataset` sont utilisés pour nommer le fichier de données externe ou le jeu de données.
- Lorsque plusieurs **valeurs de quantile** sont indiquées dans la boîte de dialogue Critères, chaque quantile est enregistré dans un fichier de données externe ou un jeu de données.
- Un caractère de soulignement "_", suivi d'un suffixe significatif de quantile, est automatiquement ajouté au nom du fichier de données ou du jeu de données. Par exemple, lorsque la **valeur quantile** 0,25, 0,50 ou 0,75 est indiquée, le suffixe `_25`, `_50` ou `_75` est ajouté au nom du fichier de données (avant l'extension `.sav`).
- Des chiffres supplémentaires peuvent être indiqués pour chaque quantile de suffixe (si nécessaire).
- Le zéro non significatif et le séparateur décimal des **valeurs de quantile** ne sont pas utilisés dans le suffixe.
- Lorsque la notation scientifique est indiquée pour les **valeurs de quantile**, elle est convertie en valeur décimale lorsqu'elle s'affiche dans le suffixe.

Régression non linéaire

La régression non linéaire est une méthode permettant de déterminer un modèle non linéaire de relation entre la variable dépendante et un groupe de variables indépendantes. A l'inverse de la régression linéaire classique, qui se limite aux modèles linéaires de prévision, la régression non linéaire peut élaborer des modèles avec des relations arbitraires entre variables dépendantes et indépendantes. Elle

emploie pour cela des algorithmes itératifs d'estimation. Notez que cette procédure n'est pas nécessaire pour les modèles polynomiaux simples de la forme $Y = A + BX^2$. En définissant $W = X^2$, nous obtenons un modèle linéaire simple, $Y = A + BW$, qui peut être estimé à l'aide de méthodes traditionnelles telles que la procédure de régression linéaire.

Exemple : Peut-on estimer l'évolution de la population par rapport au temps ? Un nuage de points montre qu'il semble y avoir une forte relation entre la population et le temps mais cette relation n'est pas linéaire. Il faut donc employer des méthodes d'estimation particulières de la procédure de régression non linéaire. En déterminant une équation appropriée, tel qu'un modèle logistique d'évolution de la population, nous pouvons obtenir une bonne approximation du modèle, ce qui nous permet de prévoir la population à des dates pour lesquelles elle n'a pas encore été mesurée.

Statistiques : Pour chaque itération : estimations des paramètres et somme résiduelle des carrés. Pour chaque modèle : somme des carrés pour la régression, résidu, total correct ou incorrect, estimations de paramètres, erreurs standard asymptotiques et matrice de corrélation asymptotique des estimations de paramètres.

Remarque : La régression non linéaire restreinte fait appel aux algorithmes proposés et mis en oeuvre dans NPSOL[®] par Gill, Murray, Saunders et Wright pour estimer les paramètres du modèle.

Remarques sur les données de régression non linéaire

Données : Les variables dépendantes et indépendantes doivent être quantitatives. Les variables catégorielles, comme la religion, la qualification, la zone de résidence, doivent être enregistrées sous forme de variables binaires (factices) ou sous de tout autre type de variables de contraste.

Hypothèses : Les résultats ne sont valides que si vous avez indiqué une fonction qui décrit correctement la relation entre les variables dépendantes et indépendantes. De surcroît, le choix des bonnes valeurs de départ est très important. Même si vous avez spécifié la forme fonctionnelle correcte du modèle, si vous utilisez de mauvaises valeurs de départ, votre modèle risque de ne pas réussir à converger et vous n'obtiendrez qu'un modèle optimal locale et non pas globale.

Procédures apparentées : De nombreux modèles qui n'apparaissent pas linéaires à première vue peuvent être transformés en modèles linéaires et analysés à l'aide une procédure de régression linéaire. Si vous n'êtes pas sûr du modèle à employer, la procédure d'estimation de courbe peut vous permettre d'identifier les relations fonctionnelles utiles dans vos données.

Obtenir une analyse de la régression non linéaire

1. À partir des menus, sélectionnez :

Analyser > Régression > Non linéaire ...

2. Sélectionnez une variable dépendante (numérique) dans la liste des variables de votre jeu de données actif.
3. Pour construire une expression de modèle, entrez celle-ci dans le champ **Expression de modèle** ou collez des composantes (variables, paramètres, fonctions) dans le champ.
4. Pour identifier les paramètres de votre modèle, cliquez sur **Paramètres**.

Un modèle segmenté (qui prend différentes formes dans les différentes parties du domaine) peut être spécifié à l'aide d'une logique conditionnelle au sein d'une même instruction de modèle.

Logique conditionnelle (régression non linéaire)

Vous pouvez spécifier un modèle segmenté à l'aide d'une logique conditionnelle. Pour employer une logique conditionnelle dans l'expression d'un modèle ou une fonction de perte, vous formez la somme d'une série de termes pour chaque condition. Chaque terme contient une expression logique (entre parenthèses) multipliée par l'expression qui doit résulter lorsque l'expression logique est vraie.

Par exemple, considérons un modèle segmenté qui est égal à 0 pour $X \leq 0$, X pour $0 < X < 1$, and 1 for $X \geq 1$. L'expression à cet égard est la suivante :

$$(X \leq 0) * 0 + (X > 0 \ \& \ X < 1) * X + (X \geq 1) * 1.$$

Les expressions logiques entre parenthèses ont toutes pour résultat 1 (vrai) ou 0 (faux). Donc :

Si $X \leq 0$, l'expression se réduit à $1 * 0 + 0 * X + 0 * 1 = 0$.

Si $0 < X < 1$, elle se réduit à $0 * 0 + 1 * X + 0 * 1 = X$.

Si $X \geq 1$, il est réduit à $0 * 0 + 0 * X + 1 * 1 = 1$.

Des exemples plus complexes peuvent se construire facilement en substituant les différentes expressions logiques et les expressions de sortie. Rappelez-vous que les doubles inégalités, telles que $0 < X < 1$, must be written as compound expressions, such as $(X > 0 \ \& \ X < 1)$.

Les variables de chaîne peuvent être utilisées dans les expressions logiques :

$(ville='Paris') * pouvach + (ville='Maubeuge') * 0.59 * pouvach$

Cela produit l'expression (la valeur de la variable *pouvach*) pour les Parisiens et une autre (59 % de cette valeur) pour les habitants de Maubeuge. Les constantes alphanumériques doivent être présentées entre guillemets ou apostrophes, comme dans cet exemple.

Paramètres de régression non linéaire

Les paramètres constituent les parties de votre modèle que la procédure de régression non linéaire estime. Ces paramètres peuvent être des constantes additives, des coefficients multiplicateurs, des exposants ou des valeurs utilisés dans les fonctions d'évaluation. Tous les paramètres que vous avez définis apparaissent (avec leurs valeurs initiales) dans la liste Paramètres de la boîte de dialogue.

Nom : Vous devez attribuer un nom à chaque paramètre. Ce nom doit être un nom de variable valide et doit être le nom utilisé dans l'expression de modèle de la boîte de dialogue principale.

Valeur initiale : Vous permet de spécifier une valeur initiale pour le paramètre, de préférence aussi proche que possible de la solution finale escomptée. De mauvaises valeurs initiales peuvent entraîner des problèmes de convergence (impossibilité de converger ou convergence locale plutôt que globale).

Utiliser l'analyse précédente pour spécifier les valeurs initiales : Si vous avez déjà exécuté une régression non linéaire à partir de cette boîte de dialogue, vous pouvez sélectionner cette option pour obtenir les valeurs initiales des paramètres à partir des valeurs de la précédente exécution. Cela vous permet de continuer la recherche lorsque l'algorithme converge lentement. (Les valeurs initiales de départ apparaissent toujours dans la liste Paramètres de la boîte de dialogue principale.)

Remarque : Cette sélection persiste dans la boîte de dialogue pour le reste de la session. Si vous changez de modèle, assurez-vous de le désélectionner.

Modèles courants de régression non linéaire

Le tableau suivant présente un exemple de syntaxe de plusieurs modèles de régression non linéaire. Un modèle choisi au hasard a peu de chance de s'adapter à vos données. Les valeurs de départ appropriées pour les paramètres sont indispensables et certains modèles requièrent des contraintes afin de converger.

Tableau 1. Exemple de syntaxe

Nom	Expression
Régression asymptotique	$b1 + b2 * \exp(b3 * x)$
Régression asymptotique	$b1 - (b2 * (b3 ** x))$
Densité	$(b1 + b2 * x) ** (-1 / b3)$
Gauss	$b1 * (1 - b3 * \exp(-b2 * x ** 2))$
Gompertz	$b1 * \exp(-b2 * \exp(-b3 * x))$
Johnson-Schumacher	$b1 * \exp(-b2 / (x + b3))$

Tableau 1. Exemple de syntaxe (suite)

Nom	Expression
Log modifié	$(b1 + b3 * x) ** b2$
Log logistique	$b1 - \ln(1 + b2 * \exp(-b3 * x))$
Loi des réponses décroissantes de Metcherlich	$b1 + b2 * \exp(-b3 * x)$
Michaelis Menten	$b1 * x / (x + b2)$
Morgan-Mercer-Florin	$(b1 * b2 + b3 * x ** b4) / (b2 + x ** b4)$
Peal-Reed	$b1 / (1 + b2 * \exp(-(b3 * x + b4 * x ** 2 + b5 * x ** 3)))$
Rapport cubique	$(b1 + b2 * x + b3 * x ** 2 + b4 * x ** 3) / (b5 * x ** 3)$
Rapport quadratique	$(b1 + b2 * x + b3 * x ** 2) / (b4 * x ** 2)$
Richards	$b1 / ((1 + b3 * \exp(-b2 * x)) ** (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1 ** (1 - b4) - b2 * \exp(-b3 * x)) ** (1 / (1 - b4))$
Weibull	$b1 - b2 * \exp(-b3 * x ** b4)$
Densité de rendement	$(b1 + b2 * x + b3 * x ** 2) ** (-1)$

Fonction de perte de la régression non linéaire

La **fonction de perte** dans la régression non linéaire est la fonction minimisée par l'algorithme. Sélectionnez soit **Somme des carrés des résidus** pour minimiser la somme des carrés résiduels, soit **Fonction de perte spécifiée par l'utilisateur** pour minimiser une fonction différente.

Si vous sélectionnez **Fonction de perte spécifiée par l'utilisateur**, vous devez définir la fonction de perte dont la somme (sur toutes les observations) doit être minimisée par le choix des valeurs du paramètre.

- La plupart des fonctions de perte impliquent la variable spéciale *RESID_*, qui représente le résidu. (La fonction de perte par défaut Somme des carrés résiduels doit être saisie explicitement sous la forme *RESID_**2*.) Si vous avez besoin d'employer la valeur prévisionnelle dans votre fonction de perte, cette valeur est égale à la variable dépendante moins le résidu.
- Il est possible de spécifier une fonction de perte conditionnelle à l'aide de la logique conditionnelle.

Vous pouvez soit taper une expression dans la zone de la fonction de perte personnalisée (spécifiée par l'utilisateur), soit coller les composants de cette expression dans la zone. Les constantes alphanumériques doivent être saisies entre guillemets ou apostrophes, tandis que les constantes numériques doivent être en format Américain avec un point en tant que délimiteur décimal.

Options de contraintes de la régression non linéaire

Une **contrainte** est une restriction émise sur les valeurs permises d'un paramètre au cours du processus itératif de recherche d'une solution. Les expressions linéaires sont évaluées avant chaque étape. Vous pouvez donc utiliser les contraintes linéaires pour éviter les étapes qui risquent d'entraîner des dépassements positifs. Les expressions non linéaires sont évaluées après chaque étape.

Chaque équation ou inégalité requièrent les éléments suivants :

- Une expression impliquant au moins un paramètre dans le modèle. Saisissez l'expression ou employez le clavier qui vous permet de coller des nombres, des opérateurs ou des parenthèses dans une expression. Vous pouvez soit taper les paramètres requis avec le reste de l'expression ou les coller depuis la liste Paramètres sur la gauche. Vous ne pouvez pas utiliser de variables courantes dans une contrainte.
- Un des trois opérateurs logiques $< =$, $=$, ou $> =$.

- Une constante numérique à laquelle l'expression est comparée à l'aide de l'opérateur logique. Tapez la constante. Les constantes numériques doivent être saisies en format Américain avec un point en tant que délimiteur décimal.

Régression non linéaire : enregistrer les nouvelles variables

Vous pouvez enregistrer un certain nombre de nouvelles variables dans votre fichier de données actif. Les options disponibles sont Prévisions, Résidus, Calculées, et Valeurs de la fonction de perte. Ces variables peuvent servir dans les analyses suivantes pour tester l'adéquation du modèle ou pour identifier les observations problématiques.

- *résiduels*. Les résiduels sont enregistrés avec le nom de variable resid.
- *Valeurs prédites*. Enregistre les valeurs prévues avec les noms de variable pred_.
- *dérivées*. Une dérivée est enregistrée pour chaque paramètre de modèle. Les noms de dérivée sont créés en ajoutant le préfixe 'd.' au six premiers caractères des noms de paramètre.
- *valeurs de fonction de perte*. Cette option est disponible si vous spécifiez votre propre fonction de perte. Le nom de variable loss_ est affecté aux valeurs de la fonction de perte.

Options de régression non linéaire

Ces options permettent de contrôler les différents aspects de votre analyse de régression non linéaire :

estimations de bootstrap. Méthode d'estimation de l'erreur standard d'une statistique par échantillonnage répété du jeu de données d'origine. Pour cela, un échantillonnage (avec remise) est réalisé afin d'obtenir de nombreux échantillons de la même taille que le jeu de données d'origine. Une estimation de l'équation non linéaire est réalisée pour chacun de ces échantillons. L'erreur standard de chaque estimation de paramètres est alors calculée comme l'écart type estimé par le bootstrap. Les valeurs des paramètres des données d'origine servent de valeurs initiales à chaque échantillon du bootstrap. Cela nécessite un algorithme de programmation quadratique séquentielle.

Méthode d'estimation : Permet de sélectionner la méthode d'estimation, si c'est possible. (Certain choix dans cette boîte de dialogue comme dans d'autres impliquent l'utilisation d'un algorithme de programmation quadratique séquentielle). Les alternatives disponibles sont la programmation quadratique séquentielle et l'algorithme de Levenberg-Marquardt.

- *programmation quadratique séquentielle*. Cette méthode est disponible pour les modèles avec ou sans contraintes. La programmation quadratique séquentielle est utilisée automatiquement si vous spécifiez un modèle avec contraintes, une fonction de perte définie par l'utilisateur ou un bootstrap. Vous pouvez saisir de nouvelles valeurs pour le Maximum d'itérations et la stabilité des coefficients. Vous pouvez également modifier la sélection dans les listes déroulantes de précision à l'optimum, de Précision de la fonction et de critère de convergence.
- *Levenberg-Marquardt*. Il s'agit de l'algorithme par défaut pour les modèles non contraints. La méthode Levenberg-Marquardt n'est pas utilisable si vous sélectionnez un modèle avec contraintes, une fonction de perte définie par l'utilisateur ou un bootstrap. Vous pouvez saisir de nouvelles valeurs pour le Maximum des itérations et vous pouvez également modifier la sélection dans les listes Convergence de la somme des carrés et Convergence des paramètres.

Interprétation des résultats de la régression non linéaire

Les problèmes de régression non linéaire présentent souvent des difficultés de calcul :

- Le choix des valeurs initiales pour les paramètres influence la convergence. Essayez de choisir des valeurs raisonnables et, si possible, proches de la solution finale escomptée.
- Certains algorithmes se révèlent parfois meilleurs que d'autres pour résoudre un problème particulier. Dans la boîte de dialogue Options, sélectionnez l'autre algorithme, s'il est disponible. (Si vous indiquez une fonction de perte ou certains types de contrainte, vous ne pouvez pas employer l'algorithme de Levenberg-Marquardt.)

- Lorsque l'itération ne s'interrompt que lorsque le nombre maximal d'itérations est atteint, le modèle final n'est probablement pas une solution satisfaisante. Sélectionnez **Utiliser l'analyse précédente pour spécifier les valeurs initiales** dans la boîte de dialogue Paramètres pour poursuivre l'itération ou, encore mieux, choisissez des valeurs initiales différentes.
- Les modèles qui requièrent une mise en exposant de ou par des valeurs importantes peuvent engendrer des dépassements positifs ou négatifs (nombres trop grands ou trop petits pour être représentés sur l'ordinateur). En général, pour éviter cela, vous devez fixer des valeurs initiales appropriées ou fixer des contraintes sur les paramètres.

Fonctions supplémentaires de la commande NLR

Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Nommer un fichier à partir duquel les valeurs initiales pour les estimations sont lues.
- Spécifier plusieurs instructions de modèle et fonctions de perte. Cela facilite la spécification d'un modèle segmenté.
- Employer vos propres dérivées plutôt que celles calculées par le programme.
- Spécifier le nombre d'échantillons de départ à générer.
- Indiquer les critères d'itération supplémentaires, notamment la définition d'une valeur critique pour le contrôle de la dérivée et la définition d'un critère de convergence pour la corrélation entre les résidus et les dérivées.

Les critères supplémentaires de la commande CNLR (régression non linéaire restreinte) vous permettent d'effectuer les opérations suivantes :

- Indiquer le nombre maximal d'itérations mineures permises dans une itération majeure.
- Fixer une valeur critique pour le contrôle de dérivée (calculée).
- Définir une limite d'étape.
- Indiquer une tolérance pour établir si les valeurs initiales se situent dans les limites déterminées.

Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Pondération estimée

Les modèles de régression linéaire standard partent du principe que la variance est constante au sein de la population étudiée. En cas contraire (par exemple, lorsque les observations élevées sur un certain attribut montrent plus de variabilité que les observations faibles sur cet attribut), la régression linéaire par la méthode des moindres carrés ordinaires ne fournit plus des estimations optimales. Si les différences de variabilité peuvent être prévues à partir d'une autre variable, la procédure de pondération estimée peut calculer les coefficients d'un modèle de régression linéaire par la méthode des moindres carrés pondérés, de sorte que les observations les plus précises (c'est-à-dire celles offrant le moins de variabilité) ont plus de pondération dans la détermination des coefficients de régression. La procédure de pondération estimée teste une fourchette de transformations de la pondération et indique celle qui correspond le mieux aux données.

Exemple : Quels sont les effets de l'inflation et du chômage sur les fluctuations des cours de la bourse ? Les actions à forte valeur montrant plus de variabilité que celles de faible valeur, les moindres carrés ordinaires ne fournissent pas d'estimations optimales. La pondération estimée vous permet de prendre en compte les effets du prix de l'action sur la variabilité des fluctuations de prix dans le calcul du modèle linéaire.

Statistiques : Valeurs de log de vraisemblance de la variable source pondérée testée, R multiple, R carré, R carré ajusté, tableau d'ANOVA pour le modèle WLS, estimations standardisées et non standardisées et log de vraisemblance pour le modèle WLS.

Remarques sur les données de pondération estimée

Données : Les variables dépendantes et indépendantes doivent être quantitatives. Les variables catégorielles, comme la religion, la qualification, la zone de résidence, doivent être enregistrées sous forme de variables binaires (factices) ou sous de tout autre type de variables de contraste. La variable de pondération doit être quantitative et doit être associée à la variabilité de la variable dépendante.

Hypothèses : Pour chaque valeur de la variable indépendante, la distribution de la variable dépendante doit être normale. La relation entre la variable dépendante et chaque variable indépendante doit être linéaire et toutes les observations doivent être indépendantes. La variance de la variable dépendante peut varier selon les niveaux de la ou des variables indépendantes mais les différences doivent être prévisibles en fonction de la variable de pondération.

Procédures apparentées : La procédure d'exploration peut être utilisée pour analyser vos données. L'exploration vous propose des tests de normalité et d'homogénéité de la variance, ainsi que des illustrations graphiques. Si votre variable dépendante semble avoir la même variance sur tous les niveaux des variables indépendantes, utilisez la procédure de régression linéaire. Si vos données apparaissent ne pas satisfaire une hypothèse (telle que la normalité), essayez de les modifier. Si vos données ne sont pas liées linéairement et qu'une modification ne change rien, utilisez un autre modèle dans la procédure d'estimation de courbe. Si votre variable dépendante est dichotomique, telle que Utilisable ou Défectueux, utilisez la procédure de régression logistique. Si votre variable dépendante est censurée (par exemple, la durée de survie après opération), utilisez les procédures Tables de survie, Kaplan-Meier ou Régression de Cox, disponibles dans Tableaux personnalisés et statistiques avancées. Si vos données ne sont pas indépendantes (par exemple, si vous observez le même individu sous différentes conditions), utilisez la procédure de mesures répétées, dans Tableaux personnalisés et statistiques avancées.

Obtenir une analyse de pondération estimée

1. À partir des menus, sélectionnez :

Analyser > Régression > Estimation du poids ...

2. Sélectionnez une variable dépendante.

3. Sélectionner une ou plusieurs variables indépendantes.

4. Sélectionnez la variable qui est la source de l'hétéroscédasticité comme variable de pondération.

variable de pondération

Les données sont pondérées par la réciproque de cette variable élevée à une puissance. L'équation de régression est calculée pour chacune des valeurs d'une plage spécifiée d'exposants et indique l'exposant qui maximise la fonction log de vraisemblance.

Plage d'alimentation

S'utilise conjointement avec la variable de pondération pour calculer les pondérations. Plusieurs équations de régression seront acceptables, une par valeur de la plage d'exposants. Les valeurs indiquées dans la case de test de variation d'exposant et dans la zone de texte doivent être comprises entre 6,5 et 7,5 (limites incluses). Les valeurs d'exposant varient de la plus faible à la plus élevée, l'incrément étant déterminé par la valeur spécifiée. Le nombre total de valeurs dans la variation de l'exposant est limité à 150.

Options de la pondération estimée

Vous pouvez sélectionner les options de votre analyse de pondération estimée :

Enregistrer la meilleure pondération en tant que nouvelle variable : Ajoute la variable de pondération au fichier actif. Cette variable s'appelle *WGT_n*, *n* étant le nombre attribué à la variable pour qu'elle ait un nom univoque.

Afficher ANOVA et les estimations : Vous permet de contrôler le mode d'affichage des statistiques dans la sortie. Vous pouvez choisir entre Pour le meilleur exposant et Pour chaque valeur de l'exposant.

Fonctions supplémentaires de la commande WLS

Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Fournir une valeur unique à l'exposant.
- Spécifier la liste des valeurs d'exposant ou combiner une plage de valeurs avec une liste de valeurs pour cet exposant.

Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Régression par les doubles moindres carrés

Les modèles de régression linéaire standard partent du principe que les erreurs au niveau de la variable dépendante ne sont pas corrélées à la ou les variables indépendantes. En cas contraire (par exemple, lorsque les relations entre les variables sont bidirectionnelles), la régression linéaire par la méthode des moindres carrés ne constitue plus un modèle de prévision optimal. La régression par les doubles moindres carrés emploie des variables instrumentales non corrélées aux termes d'erreurs pour calculer les valeurs prévisionnelles du prédicteur problématique (première étape) puis utilise ces valeurs calculées pour évaluer le modèle de régression linéaire de la variable dépendante (seconde étape). Les valeurs calculées étant fondées sur des variables non corrélées aux erreurs, les résultats du modèle double sont optimaux.

Exemple : La demande pour un article est-elle liée au prix et au revenu du consommateur ? La difficulté ici réside dans le fait que le prix et la demande agissent mutuellement l'un sur l'autre. En effet, le prix influence la demande mais la demande influence également le prix. Un modèle de régression par les doubles moindres carrés peut utiliser le revenu du consommateur comme un représentant du prix qui n'est pas corrélé avec les erreurs de mesure de la demande. Ce représentant joue le rôle du prix lui-même dans le modèle originalement spécifié, celui-ci est alors évalué.

Statistiques : Pour chaque modèle : coefficients de régression standardisés et non standardisés, R multiple, R^2 , R^2 ajusté, erreur standard de l'estimation, tableau d'analyse de variance, prévisions et résidus. Également, intervalles de confiance de 95 % pour chaque coefficient de régression, matrices de corrélation et de covariance des estimations des paramètres.

Remarques sur les données de régression par les doubles moindres carrés

Données : Les variables dépendantes et indépendantes doivent être quantitatives. Les variables catégorielles, comme la religion, la qualification, la zone de résidence, doivent être enregistrées sous forme de variables binaires (factices) ou sous de tout autre type de variables de contraste. Les variables explicatives *endogènes* doivent également être quantitatives (pas catégorielles).

Hypothèses : Pour chaque valeur de la variable indépendante, la distribution de la variable dépendante doit être normale. La variance de la distribution de la variable dépendante doit être constante pour toutes les valeurs de la variable indépendante. La relation entre la variable dépendante et chaque variable indépendante doit être linéaire.

Procédures apparentées : Si vous estimez qu'aucune de vos variables de prédicteur n'est corrélée avec les erreurs de la variable dépendante, vous pouvez employer une procédure de régression linéaire. Si vos données ne semblent pas répondre aux hypothèses formulées (telles que la normalité et la constance de la variance), essayez de les modifier. Si vos données ne sont pas liées linéairement et qu'une modification ne change rien, utilisez un autre modèle dans la procédure d'estimation de courbe. Si votre variable dépendante est dichotomique, telle que Vendue ou Non vendue, utilisez la procédure de régression logistique. Si vos données ne sont pas indépendantes (par exemple, si vous observez le même individu sous différentes conditions), utilisez la procédure de mesures répétées.

Obtenir une analyse de la régression par les doubles moindres carrés

1. A partir des menus, sélectionnez :

Analyser > Régression > des moindres carrés à 2 étapes...

2. Sélectionnez une variable dépendante.
3. Sélectionnez une ou plusieurs Variables explicatives (prédicteur).
4. Sélectionnez une ou plusieurs Variables instrumentales.
 - *instrumentale*. Il s'agit des variables utilisées pour calculer les prévisions des variables endogènes dans la première étape de l'analyse des moindres carrés à deux étapes. Les mêmes variables peuvent apparaître à la fois dans les zones de liste Variables explicatives et Variables instrumentales. Le nombre de variables instrumentales doit être au moins aussi élevé que celui des variables explicatives. Si toutes les variables explicatives et instrumentales répertoriées sont identiques, les résultats sont les mêmes que ceux obtenus par la procédure de régression linéaire.

Les variables explicatives non spécifiées comme instrumentales sont considérées comme étant endogènes. En principe, toutes les variables exogènes de la liste Explicatif sont également spécifiées en tant que variables instrumentales.

Options de régression par les doubles moindres carrés

Vous pouvez sélectionner les options suivantes pour votre analyse :

Enregistrer les nouvelles variables : Permet d'ajouter de nouvelles variables au fichier actif. Les options disponibles sont Prévisions et Résidus.

Afficher la covariance des paramètres : Permet d'imprimer la matrice de covariance des estimations des paramètres.

Fonctions supplémentaires de la commande 2SLS

Le langage de syntaxe de commande vous permet également d'estimer plusieurs équations en même temps. Reportez-vous au manuel *Command Syntax Reference* pour plus d'informations sur la syntaxe.

Méthodes de codification des variables catégorielles

Dans de nombreuses procédures, vous pouvez demander le remplacement automatique d'une variable indépendante catégorielle par un ensemble de variables de contraste, qui seront ensuite introduites dans une équation, ou en seront supprimées, en tant que bloc. Vous pouvez indiquer comment le groupe de variables de contraste doit être codé, généralement à l'aide de la sous-commande CONTRAST. Cette annexe explique et illustre le fonctionnement des différents types de contraste que vous pouvez appeler via la sous-commande CONTRAST.

Déviation

Déviations par rapport à la moyenne générale : Dans les matrices, ces contrastes ont la forme suivante :

```

mean ( 1/k  1/k  ...  1/k  1/k)
df(1) (1-1/k -1/k  ... -1/k -1/k)
df(2) (-1/k  1-1/k  ... -1/k -1/k)
.
.
df(k-1) (-1/k  -1/k  ...  1-1/k -1/k)

```

où k est le nombre de catégories de la variable indépendante, la dernière catégorie étant omise par défaut. Par exemple, les contrastes de déviation d'une variable indépendante comportant trois catégories sont les suivants :

```

( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3  2/3 -1/3)

```

Pour omettre une catégorie autre que la dernière, indiquez son numéro entre parenthèses après le mot-clé DEVIATION. Par exemple, la sous-commande suivante permet d'obtenir les déviations de la première et de la troisième catégorie, et d'omettre la deuxième :

```
/CONTRAST(FACTOR)=DEVIATION(2)
```

Supposons que le facteur (*FACTOR*) comporte trois catégories. La matrice de contraste obtenue est la suivante :

```
( 1/3  1/3  1/3)
( 2/3 -1/3 -1/3)
(-1/3 -1/3  2/3)
```

Simple

Contrastes simples : Compare chaque niveau d'un facteur au dernier. La forme de la matrice générale est la suivante :

```
mean (1/k  1/k  ...  1/k  1/k)
df(1) ( 1  0  ...  0  -1)
df(2) ( 0  1  ...  0  -1)
      .
      .
df(k-1) ( 0  0  ...  1  -1)
```

où k est le nombre de catégories de la variable indépendante. Par exemple, les contrastes simples d'une variable indépendante comportant quatre catégories sont les suivants :

```
(1/4  1/4  1/4  1/4)
( 1  0  0  -1)
( 0  1  0  -1)
( 0  0  1  -1)
```

Pour utiliser comme catégorie de référence une autre catégorie que la dernière, indiquez entre parenthèses, après le mot-clé SIMPLE, le numéro de séquence de la catégorie de référence ; il ne s'agit pas nécessairement de la valeur associée à la catégorie. Par exemple, la sous-commande CONTRAST suivante permet d'obtenir une matrice de contraste qui omet la deuxième catégorie :

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

Supposons que le facteur (*FACTOR*) comporte quatre catégories. La matrice de contraste obtenue est la suivante :

```
(1/4  1/4  1/4  1/4)
( 1  -1  0  0)
( 0  -1  1  0)
( 0  -1  0  1)
```

Helmert

Contrastes de Helmert : Compare les catégories d'une variable indépendante avec la moyenne des catégories suivantes. La forme de la matrice générale est la suivante :

```
mean (1/k  1/k  ...  1/k  1/k  1/k)
df(1) ( 1  -1/(k-1)  ...  -1/(k-1)  -1/(k-1)  -1/(k-1))
df(2) ( 0  1  ...  -1/(k-2)  -1/(k-2)  -1/(k-2))
      .
      .
df(k-2) ( 0  0  ...  1  -1/2  -1/2)
df(k-1) ( 0  0  ...  0  1  -1)
```

où k est le nombre de catégories de la variable indépendante. Par exemple, une variable indépendante comportant quatre catégories présente une matrice de contraste de Helmert ayant la forme suivante :

```
(1/4  1/4  1/4  1/4)
( 1  -1/3  -1/3  -1/3)
( 0  1  -1/2  -1/2)
( 0  0  1  -1)
```

Différence

Contrastes de différence ou contrastes inversés de Helmert : Compare les catégories d'une variable indépendante avec la moyenne des catégories précédentes de la variable. La forme de la matrice générale est la suivante :

```
mean ( 1/k  1/k  1/k  ...  1/k)
df(1) (  -1  1  0  ...  0)
```

```
df(2) ( -1/2 -1/2 1 ... 0)
      .
      .
df(k-1) (-1/(k-1) -1/(k-1) -1/(k-1) ... 1)
```

où k est le nombre de catégories de la variable indépendante. Par exemple, les contrastes de différence d'une variable indépendante comportant quatre catégories sont les suivants :

```
( 1/4 1/4 1/4 1/4)
(-1 1 0 0)
(-1/2 -1/2 1 0)
(-1/3 -1/3 -1/3 1)
```

Polynomial

Contraste polynomial orthogonal : Le premier degré de liberté contient l'effet linéaire sur toutes les catégories, le second degré l'effet quadratique, le troisième degré l'effet cubique, et ainsi de suite pour les effets d'ordre supérieur.

Vous pouvez définir l'espacement entre les niveaux du traitement mesuré par la variable qualitative donnée. Vous pouvez indiquer l'espacement égal (espacement par défaut en cas d'omission de la mesure), sous la forme d'une suite d'entiers allant de 1 à k , où k est le nombre de catégories. Si la variable *médicament* comporte trois catégories, la sous-commande

```
/CONTRAST(DRUG)=POLYNOMIAL
```

est identique à

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

Toutefois, l'espacement égal n'est pas systématiquement nécessaire. Par exemple, supposons que la variable *médicament* représente différents dosages d'un médicament administré à trois groupes. Si le dosage administré au deuxième groupe est le double de celui administré au premier groupe, et si celui administré au troisième groupe est le triple de celui administré au premier groupe, les catégories de traitement sont espacées de manière égale et, dans cette situation, une mesure appropriée se compose d'une suite d'entiers :

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

Toutefois, si le dosage administré au deuxième groupe est le quadruple de celui administré au premier groupe, et si celui administré au troisième groupe est le septuple de celui administré au premier groupe, une mesure appropriée se présente sous la forme suivante :

```
/CONTRAST(DRUG)=POLYNOMIAL(1,4,7)
```

Dans les deux cas, une fois le contraste défini, le premier degré de liberté de la variable *médicament* contient l'effet linéaire des niveaux de dosage, tandis que le deuxième degré contient l'effet quadratique.

Les contrastes polynomiaux sont particulièrement utiles pour réaliser des tests de tendances et analyser la nature des surfaces de réponses. Vous pouvez également utiliser les contrastes polynomiaux pour effectuer un ajustement de courbe non linéaire, comme une régression curviligne.

Répété

Compare les niveaux adjacents d'une variable indépendante : La forme de la matrice générale est la suivante :

```
mean (1/k 1/k 1/k ... 1/k 1/k)
df(1) ( 1 -1 0 ... 0 0)
df(2) ( 0 1 -1 ... 0 0)
      .
      .
df(k-1) ( 0 0 0 ... 1 -1)
```


où k est le nombre de catégories de la variable indépendante. Par exemple, les contrastes répétés d'une variable indépendante comportant quatre catégories sont les suivants :

```
(1/4  1/4  1/4  1/4)
(  1   -1   0   0)
(  0    1  -1   0)
(  0    0   1  -1)
```

Ces contrastes sont utiles dans l'analyse des profils et lorsque des statistiques de différence sont nécessaires.

Spécial

Contraste défini par l'utilisateur : Permet la saisie de contrastes spéciaux sous la forme de matrices carrées comportant autant de lignes et de colonnes que le nombre de catégories de la variable indépendante spécifiée. Pour MANOVA et LOGLINEAR, la première ligne saisie est toujours l'effet de moyenne ou de constante, et représente le groupe de pondérations indiquant comment déterminer, par rapport à la variable spécifiée, la moyenne des autres variables indépendantes (le cas échéant). Généralement, ce contraste est un vecteur.

Les autres lignes de la matrice contiennent les contrastes spéciaux indiquant les comparaisons entre les catégories de la variable. Généralement, les contrastes orthogonaux sont les plus utiles. Ils ne sont pas redondants et sont statistiquement indépendants. Les contrastes sont orthogonaux si :

- Pour chaque ligne, la somme des coefficients de contraste est égale à 0.
- La somme des produits des coefficients correspondant à toutes les paires de lignes disjointes est aussi égale à 0.

Par exemple, supposons que la variable traitement comporte quatre niveaux et que vous souhaitez comparer les différents niveaux de traitement. Un contraste spécial approprié peut avoir la forme suivante :

```
(1  1  1  1) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0  2 -1 -1) compare 2nd with 3rd and 4th
(0  0  1 -1) compare 3rd with 4th
```

que vous définissez à l'aide de la sous-commande CONTRAST suivante pour MANOVA, LOGISTIC REGRESSION et COXREG :

```
/CONTRAST(TREATMNT)=SPECIAL( 1  1  1  1
                             3 -1 -1 -1
                             0  2 -1 -1
                             0  0  1 -1 )
```

Pour LOGLINEAR, vous devez indiquer :

```
/CONTRAST(TREATMNT)=BASIS SPECIAL( 1  1  1  1
                                    3 -1 -1 -1
                                    0  2 -1 -1
                                    0  0  1 -1 )
```

Chaque ligne, à l'exception des moyennes, correspond à 0. Les produits de chaque paire de rangées disjointes sont également égaux à 0 :

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

Il n'est pas nécessaire que les contrastes spéciaux soient orthogonaux. Toutefois, ils ne doivent pas constituer des combinaisons linéaires les uns avec les autres. Si tel est le cas, la procédure signale la dépendance linéaire et interrompt le traitement. Les contrastes polynomiaux, de différence et de Helmert sont tous des contrastes orthogonaux.

Indicateur

Codification des variables indicateur : Egalement appelé codification factice, ce type de codification n'est pas disponible dans LOGLINEAR ni MANOVA. Le numéro des nouvelles variables codées est $k-1$. Les

observations de la catégorie de référence sont codées 0 pour toutes les variables $k-1$. Une observation dans la $n^{\text{ième}}$ catégorie est codée 0 pour toutes les variables indicateur, sauf la $n^{\text{ième}}$, codée 1.

Kernel Ridge Regression

Kernel Ridge Regression est une procédure d'extension qui utilise la classe `sklearn.kernel_ridge.KernelRidge` Python pour estimer les modèles de régression de la crête du noyau. Les modèles de régression de la crête du noyau sont des modèles de régression non paramétriques qui sont capables de modéliser les relations linéaires et non linéaires entre les variables prédictives et les résultats. Les résultats peuvent être très sensibles aux choix des hyperparamètres du modèle. La régression de la crête du noyau facilite le choix des valeurs de l'hyperparamètre par le biais de la validation croisée k-fold sur des grilles spécifiées de valeurs à l'aide de la classe `sklearn.model_selection.GridSearchCV`.

Exemple

Statistiques

Additive_CHI2, CHI2, Cosinus, Laplacien, Linéaire, Polynomial, RBF, Sigmoid, Alpha, Gamma, Coef0, Degré, validation croisée, observé contre prédit, résidus contre prédit, coefficients de poids doubles, coefficients de poids de l'espace noyau.

Remarques sur les données

Données

- Vous pouvez spécifier l'une ou l'autre des huit fonctions différentes du noyau.
- La fonction noyau sélectionnée détermine quels hyperparamètres sont actifs.
- Les hyperparamètres comprennent des alpha pour la régularisation des crêtes qui sont communs à tous les noyaux, plus jusqu'à trois autres hyperparamètres pour chaque fonction noyau spécifique.
- Lorsque plusieurs sous-commandes de noyau sont spécifiées, ou plus d'une valeur pour n'importe quel paramètre est spécifiée, une recherche de grille avec validation croisée pour évaluer les modèles est effectuée, et le modèle de meilleur ajustement basé sur les données de sortie retenues est sélectionné.
- L'extension accepte les variables partagées de la procédure de partage de fichiers et les pondérations à l'aide de la procédure de pondération.
- Lorsque les poids sont inclus, ils sont utilisés pour créer des valeurs ajustées dans toutes les analyses. En raison des limitations de la méthode de score dans la classe `sklearn.model_selection.GridSearchCV`, les évaluations par validation croisée qui sont utilisées pour la sélection du modèle ne sont pas pondérées.

Hypothèses

Obtention d'une régression de la crête du noyau

1. À partir des menus, sélectionnez :

Analyser > Régression > Arête du noyau ...

2. Sélectionnez une **variable** dépendante.
3. Sélectionnez une ou plusieurs variables **Indépendante(s)**.
4. Le paramètre **Modèle unique** par défaut est utilisé lorsqu'une seule valeur pour chaque paramètre de fonction de noyau est spécifiée. Lorsque le paramètre **Modèle unique** est sélectionné, vous ne pouvez pas spécifier d'autres fonctions **Noyau (s)** et les pondérations sont appliquées intégralement tout au long de l'analyse, de l'évaluation et de la notation des résultats. Vous pouvez également utiliser les commandes de flèche vers le haut et vers le bas pour réorganiser les fonctions du noyau.

Vous pouvez également sélectionner **Sélection de modèle** dans la liste **Mode**.

Lorsque **Sélection de modèle** est sélectionné dans la liste **Mode**, vous pouvez ajouter plusieurs fonctions de noyau à la liste **Noyau (s)**.

- a. Cliquez sur le contrôle d'ajout (+) pour inclure des fonctions de noyau supplémentaires.
- b. Cliquez sur la cellule vide dans la colonne **Noyau** pour sélectionner une fonction de noyau.
- c. Cliquez deux fois sur une cellule de ligne de fonction du noyau pour spécifier les valeurs des paramètres de fonction du noyau pour la colonne correspondante (**Alpha, Gamma, Coef0, Degré**). Pour plus d'informations, voir «Paramètres de noyau», à la page 31. Les paramètres d'optimisation de la fonction de noyau par défaut sont répertoriés ci-dessous.

Additive_CHI2

ALPHA=1 GAMMA= 1

CHI2

ALPHA=1 GAMMA= 1

cosinus

ALPHA=1

Laplacien

ALPHA= 1 GAMMA=1/p

Linéaire

Fonction de noyau par défaut. ALPHA=1

Polynomial

ALPHA= 1 GAMMA=1/p COEFO = 1 DEGREE= 3

fonction de base radiale

ALPHA= 1 GAMMA=1/p

Sigmoïdale

ALPHA= 1 GAMMA=1/p COEFO = 1

Remarque : Lorsque plus d'une valeur pour un paramètre de fonction de noyau est spécifiée, une recherche de grille avec validation croisée pour évaluer les modèles est effectuée et le modèle de meilleur ajustement basé sur les données conservées est sélectionné.

5. Si vous le souhaitez, cliquez sur **Options** pour indiquer le nombre de plis de validation croisée, les options d'affichage, les paramètres de tracé et les éléments à sauvegarder. Pour plus d'informations, voir «Régression de la crête du noyau: Options», à la page 31.
6. Cliquez sur **OK**.

Paramètres de noyau

La boîte de dialogue **Paramètres du noyau** fournit des options permettant de spécifier des valeurs de paramètre de fonction de noyau uniques et de spécifier que la sélection de modèle est effectuée à l'aide d'une recherche de grille sur les combinaisons de noyaux et de valeurs de paramètres de grille spécifiées.

Spécifier les paramètres uniques

Activez le paramètre pour spécifier des valeurs pour le paramètre de fonction de noyau sélectionné.

- Entrez une valeur et cliquez sur **Ajouter** pour inclure la valeur dans le paramètre de fonction du noyau.
- Sélectionnez une valeur de paramètre et cliquez sur **Changement** pour mettre à jour la valeur.
- Sélectionnez une valeur de paramètre et cliquez sur **Supprimer** pour supprimer la valeur.

Spécifier les paramètres de grille

Activez le paramètre pour spécifier que la sélection du modèle est effectuée à l'aide d'une recherche de grille sur les combinaisons de noyaux et de valeurs de paramètre de grille spécifiées.

Régression de la crête du noyau: Options

La boîte de dialogue **Tracements** fournit des options permettant de spécifier le nombre de plis de validation croisée, les options d'affichage, les paramètres de tracé et les éléments à sauvegarder.

Nombre de plis de validation croisée

Nombre de divisions ou de plis dans la validation croisée avec la recherche de grille pour la sélection de modèle. Entrez une valeur entière supérieure à 1. Valeur par défaut : 5. Le paramètre est disponible uniquement lorsque **Sélection de modèle** est choisi en tant que **Mode** dans la boîte de dialogue **Régression de la crête du noyau** principale.

Affichage

Fournit des options permettant de spécifier la sortie à afficher lorsque la validation croisée est en vigueur.

Meilleur

Le paramètre par défaut affiche uniquement les résultats de base pour le meilleur modèle choisi.

Comparer

Affiche les résultats de base pour tous les modèles évalués.

Comparer des modèles et des plis

Affiche les résultats complets pour chaque division ou chaque pli pour chaque modèle évalué.

Tracé

Fournit des options permettant de spécifier des tracés de valeurs observées ou résiduelles par rapport aux valeurs prédites.

Observé par rapport à prévu

Affiche un nuage de points des valeurs observées par rapport aux valeurs prédites pour le modèle spécifié ou le meilleur modèle.

Résidus par rapport à prévu

Affiche un nuage de points des valeurs résiduelles par rapport aux valeurs prédites pour le modèle spécifié ou le meilleur modèle.

Sauvegarder

La table fournit des options permettant de spécifier des variables à enregistrer dans le jeu de données actif.

Valeurs prédites

Enregistre les valeurs prédites du modèle spécifié ou du meilleur modèle dans le jeu de données actif. Un nom de variable facultatif peut être inclus.

Résiduels

Enregistre les résidus des prévisions spécifiées ou des meilleures prédictions du modèle dans le jeu de données actif. Un nom de variable facultatif peut être inclus.

Coefficients doubles

Enregistre les coefficients de pondération de l'espace du noyau ou du noyau à partir du modèle spécifié vers le jeu de données actif. Un nom de variable facultatif peut être inclus. Le paramètre n'est pas disponible lorsque **Sélection de modèle** est choisi en tant que **Mode** dans la boîte de dialogue **Régression de la crête du noyau** principale.

Modèles de temps d'échec accéléré paramétriques

Une analyse de modèle AFT (Parametric Accelerated Failure Time) appelle la procédure des modèles de survie paramétriques avec des données de durée de vie non récurrentes. Les modèles de survie paramétriques supposent que le temps de survie suit une distribution connue, et cette analyse correspond aux modèles de temps d'échec accéléré avec leurs effets de modèle proportionnels par rapport au temps de survie.

Obtention d'une analyse des modèles de temps d'échec accéléré paramétrique

1. A partir des menus, sélectionnez :

Analyse > Survie > Modèles AFT (Parametric Accelerated Failure Time)

2. Sélectionnez une variable source.

Durée

Survie

Variable numérique unique indiquant la durée de la durée de survie.

Début/Fin

Variables numériques indiquant l' **heure de début** et l' **heure de fin**.

Statut

Chaîne facultative unique ou variable numérique qui détermine l'un des paramètres de statut suivants:

Défaillance/Événement

Mappe un enregistrement à une catégorie d'échec ou d'événement. La valeur par défaut d'une variable de statut de chaîne est F.

Censure à droite

Mappe un enregistrement à une catégorie de censure de droite. La valeur par défaut d'une variable de statut de chaîne est R.

Censure à gauche

Mappe un enregistrement à une catégorie de censure de gauche. La valeur par défaut d'une variable de statut de chaîne est L.

Censure par intervalle

Mappe un enregistrement à une catégorie de censure d'intervalle. Pour **Démarrer / Arrêter** uniquement. La valeur par défaut d'une variable de statut de chaîne est I.

Traitement valeurs non mappées

Contrôle la catégorie à laquelle mapper les enregistrements non mappés. Pour supprimer les enregistrements dont le mappage a échoué, sélectionnez **Exclure les enregistrements de l'analyse**.

Pour **Survie**, le statut par défaut de toutes les observations est **Echec / Événement**. Pour **Start / End**, le statut par défaut est **Interval Censoring**. Cliquez sur le bouton **Définir un événement** pour définir un événement pour la variable de statut.

Covariable(s)

Une ou plusieurs variables numériques facultatives à traiter comme covariables. Notez qu'une variable ne peut pas être spécifiée à la fois par **Covariable (s)** et par **Facteur (s) fixe (s)**.

Facteur(s) fixe(s)

Une ou plusieurs variables facultatives à traiter en tant que facteurs. Une variable ne peut pas être spécifiée à la fois par **Facteur (s) fixe (s)** et par **Covariable (s)**.

Troncature à gauche

Variable numérique facultative unique pour la troncature à gauche pour **Survie** uniquement.

Modèles de temps d'échec accéléré paramétriques: Critères

Critères

Panneau facultatif permettant de spécifier les critères généraux.

Intervalle de confiance

Pourcentage facultatif permettant de spécifier le niveau des intervalles de confiance des paramètres de régression. Il doit s'agir d'une valeur numérique unique comprise entre 0 et 100. La valeur par défaut est 95.

Valeurs manquantes

Option permettant de contrôler le traitement des valeurs manquantes de l'utilisateur:

Exclure les valeurs manquantes de l'utilisateur et les valeurs système manquantes

Traite les valeurs manquantes de l'utilisateur comme des valeurs valides. Il s'agit de la valeur par défaut.

Les valeurs manquantes de l'utilisateur sont traitées comme des données valides

Ignore les désignations de valeurs manquantes de l'utilisateur et les traite comme des valeurs valides.

Traitement du statut

Pour **Démarrer / Arrêter** uniquement. Option permettant de contrôler le traitement des enregistrements dont les zones de statut sont incorrectes:

Supprimer l'enregistrement en conflit

Supprime les enregistrements en conflit. Il s'agit du paramètre par défaut.

Obtenir les informations temporelles en fonction du statut

Obtient les informations d'heure en fonction du statut.

Dériver le statut en fonction des informations temporelles

Modifie le statut en fonction des informations d'heure.

Modèles de temps d'échec accéléré paramétriques: Modèle

Modèle

Panneau facultatif permettant de spécifier les options et les paramètres du modèle.

Répartition du temps de survie

Option permettant de spécifier la distribution de la durée de survie.

Weibull

Indique la distribution de Weibull. Il s'agit du paramètre par défaut.

Exponentiel

Indique la distribution exponentielle.

Loi log-normale

Indique la distribution log-normale.

Log logistique

Indique la distribution log-logistique.

Paramètres de covariable

Spécifiez les variables de covariable.

Paramètres de facteur

Spécifiez les variables de facteur.

Valeur initiale de la constante

Option permettant de spécifier la valeur initiale du terme de la constante. S'il est spécifié, il doit s'agir d'une valeur numérique unique et ne peut pas être égal à 0.

Valeur initiale du paramètre d'échelle

Option permettant de contrôler la définition du paramètre d'échelle.

Erreur standard de la régression correspondante des MCO

Utilise l'erreur standard de la régression des moindres carrés ordinaires correspondante comme valeur initiale.

Inverse l'erreur standard de la régression des MCO correspondante

Utilise la réciproque de l'erreur standard.

Valeur fournie par l'utilisateur

Si une valeur numérique unique est spécifiée, elle est utilisée comme valeur initiale. S'il est spécifié, il doit être supérieur à 0.

Modèles de temps d'échec accélérés paramétriques: estimation

Estimation

Panneau facultatif permettant de spécifier les paramètres permettant de contrôler l'estimation des modèles de temps d'échec accéléré et le processus de sélection de fonction facultatif.

Méthode de direction alternée ou multipinces (ADMM)

Rapide

Applique la méthode des multiplicateurs à sens alternatif rapide (ADMM). Il s'agit de la valeur par défaut.

Traditionnel

Applique l'algorithme ADMM traditionnel.

Appliquer la régularisation L-1

Exécute le processus pour contrôler la sélection des fonctions. La zone **Paramètre de pénalité** indique le paramètre de pénalité qui contrôle le processus de régularisation. Il doit s'agir d'une valeur unique supérieure à 0. Le paramètre par défaut est 0.001.

Critères de convergence des modèles

Convergence des paramètres

Spécifie les critères de convergence pour le paramètre. Il doit s'agir d'une valeur numérique unique appartenant à $[0, 1)$. Valeur par défaut : 0.000001. Pour **Type**, vous pouvez sélectionner **ABSOLUTE** pour appliquer la convergence absolue à l'optimisation interne ou **RELATIVE** pour appliquer la convergence relative à l'optimisation interne. La zone facultative **Valeur** indique un mot clé.

Convergence des fonctions d'objectif

Spécifie les critères de convergence pour la fonction objectif. Il doit s'agir d'une valeur numérique unique appartenant à $[0, 1)$. La valeur par défaut est 0, ce qui n'applique pas les critères de convergence. Pour **Type**, vous pouvez sélectionner **ABSOLUTE** pour appliquer la convergence absolue à l'optimisation interne ou **RELATIVE** pour appliquer la convergence relative à l'optimisation interne. La zone facultative **Valeur** indique un mot clé.

Convergence de Hess

Spécifie les critères de convergence de la matrice de Hess. Il doit s'agir d'une valeur numérique unique appartenant à $[0, 1)$. La valeur par défaut est 0, ce qui n'applique pas les critères de convergence. Pour **Type**, vous pouvez sélectionner **ABSOLUTE** pour appliquer la convergence absolue à l'optimisation interne ou **RELATIVE** pour appliquer la convergence relative à l'optimisation interne. La zone facultative **Valeur** indique un mot clé.

Critères de convergence des résidus

Option permettant de contrôler le processus d'optimisation.

Résidu primaire et double

Applique à la fois le critère de convergence des résidus primal et dual. Il s'agit du paramètre par défaut.

Résidu primaire uniquement

Applique le critère de convergence des résidus primaux.

Résidu double uniquement

Applique le critère de convergence des résidus duaux.

Méthode

Paramètre facultatif permettant de spécifier la méthode d'estimation.

Automatique

Sélectionne automatiquement la méthode en fonction de l'exemple de jeu de données. Il s'agit de la valeur par défaut. Le champ **Nombre de prédicteurs de seuil** indique le seuil du nombre de prédicteurs et doit être un entier unique supérieur à 1. La valeur par défaut est de 1000.

Newton-Raphson

Applique la méthode de Newton-Raphson.

L-BFGS

Applique l'algorithme BFGS à mémoire limitée. La zone **Mettre à jour** indique le nombre de mises à jour passées gérées par l'algorithme BFGS à mémoire limitée et doit être un entier unique supérieur ou égal à 1. La valeur par défaut est 5.

Itération

Nombre maximum d'itérations

Spécifie le nombre maximal d'itérations. Il doit s'agir d'un entier unique appartenant à [1, 100]. La valeur par défaut est 20.

Découpage maximal en deux des étapes

Indique le nombre maximal de step-halving. Il doit s'agir d'un entier unique appartenant à [1, 20]. La valeur par défaut est 5.

Nombre maximal de recherches de ligne

Indique le nombre maximal de recherches de ligne. Il doit s'agir d'un entier unique appartenant à [1, 100]. La valeur par défaut est 20.

Convergence absolue pour le processus d'itération

Indique la convergence absolue pour le processus d'itération externe. Il doit s'agir d'une valeur numérique unique appartenant à (0, 1). Le paramètre par défaut est 0.0001.

Convergence relative pour le processus d'itération

Indique la convergence relative pour le processus d'itération externe. Il doit s'agir d'une valeur numérique unique appartenant à (0, 1). Le paramètre par défaut est 0.01.

Modèles de temps d'échec accéléré paramétriques: Imprimer

Imprimer

Panneau facultatif permettant de contrôler les sorties de la table.

Détails du codage de facteur

Si cette option est sélectionnée, elle affiche et imprime les détails de codage des facteurs. Le processus est ignoré si aucun facteur n'est appliqué.

Valeurs initiales affectées aux paramètres de régression

Si cette option est sélectionnée, elle affiche les valeurs initiales utilisées dans le processus d'estimation.

Historique des itérations de modèle

Si cette option est sélectionnée, elle affiche l'historique des itérations de l'analyse de survie. Dans la zone **Nombre d'étapes**, indiquez le nombre d'étapes compris entre 1 et 99999999. Le paramètre par défaut est 1.

Résultats de sélection contenant

Contrôle l'affichage des détails de la sélection de fonction.

Variables sélectionnées et non sélectionnées

Affiche les variables sélectionnées et non sélectionnées dans le tableau.

Uniquement les variables sélectionnées

Affiche uniquement les variables sélectionnées.

Uniquement les variables non sélectionnées

Affiche uniquement les variables non sélectionnées. La zone **Nombre maximal de variables à afficher** indique le nombre maximal de variables imprimées dans le tableau. La valeur par défaut est 30.

Modèles de temps d'échec accéléré paramétriques: Prévoir

Prévision

Panneau facultatif permettant d'évaluer et d'enregistrer les statistiques prévues dans le jeu de données actif.

Valeurs de temps pour le scoring

Valeurs de temps définies par variable(s) dépendante(s)

Attribue un score à **Prévisions** en fonction de la variable de temps spécifiée pour le modèle de survie paramétrique.

Intervalles ordinaires

Attribue des scores aux **prévisions** en fonction des valeurs de temps futures. La zone **Intervalle de temps** indique l'intervalle de temps et doit être une valeur numérique unique supérieure à 0. La zone **Nombre de périodes** indique le nombre de périodes et doit être un entier numérique unique compris entre 2 et 100.

Durée

Calcule les scores des **prévisions** en fonction de la durée afin de définir les valeurs temporelles futures. Il doit s'agir d'une variable numérique unique.

Prévisions

Survie

Calcule et enregistre les statistiques de survie prévues dans le jeu de données actif. Le nom de variable personnalisée par défaut (ou nom racine) est PredSurvival.

Risque

Attribue un score et enregistre les risques prévus dans le jeu de données actif. Le nom de variable personnalisée par défaut (ou nom racine) est PredHazard.

Risque cumulé

Attribue un score et enregistre les risques cumulés prévus dans le jeu de données actif. Le nom de variable personnalisée par défaut (ou nom racine) est PredCumHazard.

Survie conditionnelle

Attribue un score et enregistre les statistiques de survie conditionnelle prévues dans le jeu de données actif. Le nom de variable personnalisée par défaut (ou nom racine) est PredConditionalSurvival. Le processus sera ignoré si PASTTIME n'est pas spécifié. Une valeur **Temps de survie passé** est requise et spécifie les valeurs de temps passé pour le scoring. Il doit s'agir d'une variable numérique unique.

Modèles de temps d'échec accéléré paramétriques: Tracé

Tracé

Tracé résiduel de Cox-Snell

Sélectionnez **Afficher le tracé** pour créer un tracé de résidus de Cox-Snell. Dans la zone **Nombre de points de césure de regroupement en catégories**, indiquez un nombre compris entre 1 et 10000. La valeur par défaut est 100.

Tracés de fonction

Option permettant de contrôler les tracés de fonction.

Type

Survie

Crée le tracé des fonctions de survie.

Risque

Crée le tracé pour les fonctions de risque.

Densité

Crée un tracé pour les fonctions de densité.

Nombre de points à afficher

Indique le nombre de points de fonction compris entre 1 et 200. La valeur par défaut est 100.

Valeurs de covariable pour le tracé

Facultatif permettant de spécifier les valeurs fournies par l'utilisateur et de les affecter aux prédicteurs. Par défaut, les tracés désignés sont créés à la **moyenne** de chaque covariable en vigueur et à la fréquence de catégorie de chaque facteur en vigueur. Si spécifié, les tracés désignés seront créés en fonction du paramètre du modèle. En présence de variables en double, celle spécifiée en premier est reconnue et le reste est ignoré. Une variable valide doit être contenue dans un effet de modèle. Pour une covariable, la valeur fournie par l'utilisateur doit être numérique. L'omission d'une variable indique que la fréquence de la catégorie et la **moyenne** sont utilisées par défaut pour le facteur et la covariable, respectivement. Si une valeur non valide est affectée à une variable, le modèle demandé ne sera pas tracé.

Valeurs de facteur pour le tracé

Facultatif permettant de spécifier les valeurs fournies par l'utilisateur et de les affecter aux prédicteurs. En présence de variables en double, celle spécifiée en premier est reconnue et le reste est ignoré. Une variable valide doit être contenue dans un effet de modèle. L'omission d'une variable indique que la fréquence de la catégorie et la moyenne sont utilisées par défaut pour le facteur et la covariable, respectivement. Si une valeur non valide est affectée à une variable, le modèle demandé ne sera pas tracé.

Séparer les lignes pour

Option permettant de spécifier une variable catégorielle à l'aide de laquelle les tracés de courbes seront tracés.

Nombre maximal de lignes dans un graphique

Indique le nombre maximal de lignes dans un graphique si l'option **Séparer les lignes pour** est spécifiée. Le paramètre par défaut est 10.

Modèles de temps d'échec accéléré paramétriques: Exportation

Exporter

Sélectionnez **Exporter les informations du modèle dans un fichier XML** pour écrire les informations du modèle et des paramètres dans un fichier PMML à des fins d'évaluation. Vous devez indiquer le répertoire et le nom du fichier PMML à sauvegarder.

Survie AFT Définir des événements pour les variables de statut

Les occurrences de la ou des valeurs sélectionnées pour la variable de statut indiquent que l'événement final s'est produit pour ces cas. Toutes les autres observations sont considérées comme censurées. Entrez soit une valeur unique, soit une plage de valeur qui identifie l'événement étudié.

Modèles de temps d'échec accélérés paramétriques: Sélectionner une catégorie

Le paramètre Sélectionner une catégorie fournit des options permettant de choisir une valeur qui indique la catégorie à modéliser comme référence pour la comparaison.

Sélection de la catégorie

Cliquez sur 'Dernière catégorie' pour ouvrir la boîte de dialogue 'Sélectionner une catégorie'.

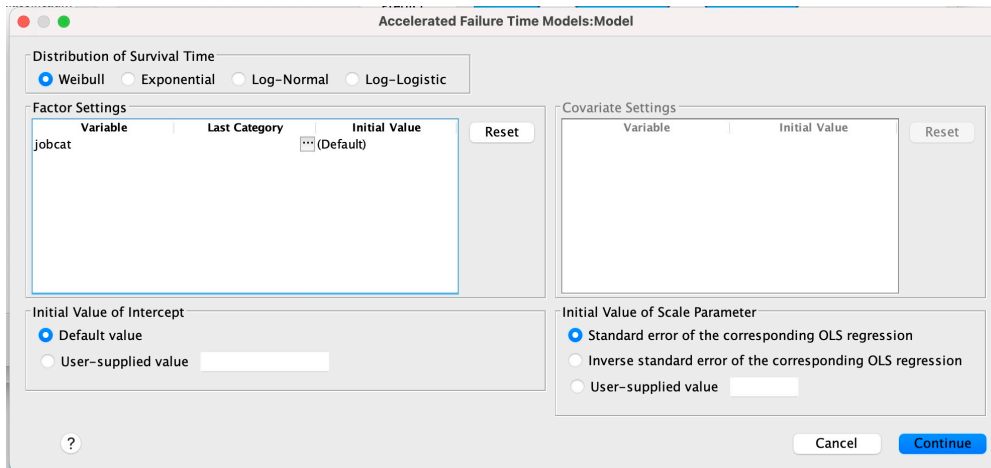


Figure 1. Modèles de durée de vie accélérée-Boîte de dialogue-Catégorie

Pour désigner une catégorie comme version de référence, sélectionnez une valeur dans la boîte de dialogue 'Sélectionner une catégorie'.

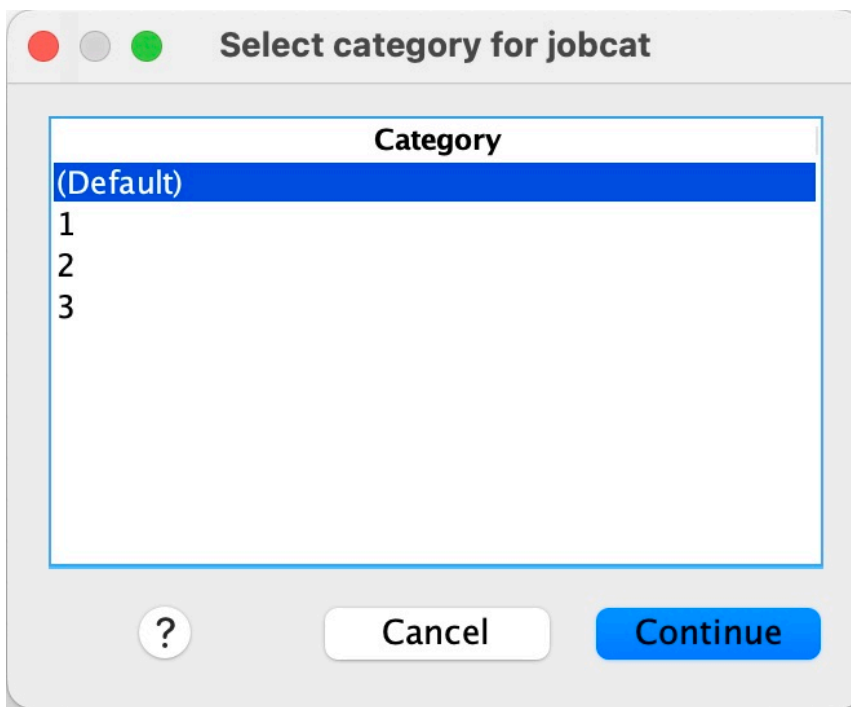


Figure 2. Modèles de durée de vie accélérée-Boîte de dialogue-Sélectionner une catégorie

Cliquez sur Continuer.

Modèles de fragilité partagés paramétriques

Une analyse de survie des modèles de fragilité partagés paramétriques démarre la procédure des modèles de survie paramétriques avec une entrée de données de durée de vie récurrente. Les modèles de survie paramétriques supposent que le temps de survie suit une distribution connue, et cette analyse incorpore un terme de fragilité dans un modèle de survie paramétrique. Il est traité comme un composant aléatoire pour tenir compte d'un effet non observé en raison de la variabilité de niveau individuel ou de groupe.

Obtention d'une analyse Parametric Shared Frailty Models

1. Dans les menus, sélectionnez :

Analyser > Survie > Modèles de fragilité partagés paramétriques

2. Sélectionnez une variable source.

Durée

Survie

La durée de survie est représentée par une variable pour indiquer l'heure de fin. L'heure de début est définie sur 0.

Début / Fin

Variables numériques indiquant l' **heure de début** et l' **heure de fin**.

Objet

Requis pour exécuter la procédure. Indique une variable unique pour l'ID de sujet.

Intervalle

Indique une variable unique et numérique pour le numéro d'intervalle utilisé pour identifier les différents enregistrements récurrents qui partagent le même ID de sujet.

Statut

Chaîne facultative unique ou variable numérique qui détermine l'un des paramètres de statut suivants:

Défaillance/Événement

Mappe un enregistrement à une catégorie d'échec / événement. La valeur par défaut d'une variable de statut de chaîne est F.

Censure à droite

Mappe un enregistrement à une catégorie de censure de droite. La valeur par défaut d'une variable de statut de chaîne est R.

Traitement valeurs non mappées

Contrôle la catégorie à laquelle mapper les enregistrements non mappés. Pour supprimer les enregistrements dont le mappage a échoué, sélectionnez **Exclure les enregistrements de l'analyse**.

Cliquez sur le bouton **Définir un événement** pour définir un événement pour la variable de statut.

Covariable(s)

Une ou plusieurs variables numériques facultatives à traiter comme covariables. Notez qu'une variable ne peut pas être spécifiée à la fois par **Covariable (s)** et par **Facteur (s) fixe (s)**.

Facteur(s) fixe(s)

Une ou plusieurs variables facultatives à traiter en tant que facteurs. Une variable ne peut pas être spécifiée à la fois par **Facteur (s) fixe (s)** et **Covariable (s)**.

Modèles de fragilité partagés paramétriques: Critères

Critères

Panneau facultatif permettant de spécifier les critères généraux.

Intervalle de confiance

Pourcentage facultatif permettant de spécifier le niveau des intervalles de confiance des paramètres de régression. Il doit s'agir d'une valeur numérique unique comprise entre 0 et 100. La valeur par défaut est 95.

Niveau d'importance

Option permettant de spécifier le niveau de signification du test du rapport de vraisemblance pour le composant de fragilité. Il doit s'agir d'une valeur numérique unique comprise entre 0 et 1. La valeur par défaut est 0.05.

Valeurs manquantes

Option permettant de contrôler le traitement des valeurs manquantes de l'utilisateur:

Exclure les valeurs manquantes de l'utilisateur et les valeurs système manquantes

Traite les valeurs manquantes de l'utilisateur comme des valeurs valides. Il s'agit de la valeur par défaut.

Les valeurs manquantes de l'utilisateur sont traitées comme des données valides

Ignore les désignations de valeurs manquantes de l'utilisateur et les traite comme des valeurs valides.

Traitement de l'intervalle

Option permettant de contrôler le traitement des enregistrements dont l'intervalle est en conflit avec l'heure de début et l'heure de fin. Elle prend effet s'il existe deux variables de temps avec une variable d'intervalle spécifiée dans la boîte de dialogue principale.

Supprimer les enregistrements en conflit

Supprime l'intégralité des enregistrements de série de sujet si la valeur d'intervalle est en conflit avec l'heure de début et de fin. Il s'agit du paramètre par défaut.

Reconnaître les valeurs d'intervalle en fonction des heures de début et de fin

Reconnaît la valeur d'intervalle à partir de l'heure de début et de fin.

Modèles de fragilité partagés paramétriques: Modèle

Modèle

Panneau facultatif permettant de spécifier les options et les paramètres du modèle.

Répartition du temps de survie

Option permettant de spécifier la distribution de la durée de survie.

Weibull

Indique la distribution de Weibull. Il s'agit du paramètre par défaut.

Exponentiel

Indique la distribution exponentielle.

Loi log-normale

Indique la distribution Log-normal.

Log logistique

Indique la distribution log-logistique.

Paramètres de covariable

Spécifiez les variables de covariable.

Paramètres de facteur

Spécifiez les variables de facteur.

Valeur initiale de la constante

Option permettant de spécifier la valeur initiale du terme de la constante. S'il est spécifié, il doit s'agir d'une valeur numérique unique et ne peut pas être égal à 0.

Valeur initiale du paramètre d'échelle

Option permettant de contrôler la définition du paramètre d'échelle.

Erreur standard de la régression correspondante des MCO

Utilise l'erreur standard de la régression des moindres carrés ordinaires correspondante comme valeur initiale.

Inverse l'erreur standard de la régression des MCO correspondante

Utilise la réciproque de l'erreur standard.

Valeur fournie par l'utilisateur

Si une valeur numérique unique est spécifiée, elle est utilisée comme valeur initiale. S'il est spécifié, il doit être supérieur à 0.

Composant fragilité

Paramètre facultatif permettant de spécifier la **distribution** du composant de fragilité.

Gamma

Indique la distribution Gamma. Il s'agit du paramètre par défaut.

Inverse-Gaussien

Indique la distribution gaussienne inverse.

Valeur initiale de la variance

Indique la valeur initiale de la variance de la composante de fragilité. Il doit s'agir d'une valeur numérique unique supérieure à 0. La valeur par défaut est 1.0 pour la distribution Gamma et 0.1 pour la distribution gaussienne inverse.

Modèles de fragilité partagés paramétriques: estimation

Estimation

Panneau facultatif permettant de spécifier les paramètres permettant de contrôler l'estimation des modèles de fragilité partagés et le processus de sélection des fonctions facultatives.

Méthode de direction alternée ou multipinces (ADMM)

Rapide

Applique la méthode des multiplicateurs à sens alternatif rapide (ADMM). Il s'agit de la valeur par défaut.

Traditionnel

Applique l'algorithme ADMM traditionnel.

Appliquer la régularisation L-1

Exécute le processus pour contrôler la sélection des fonctions. La zone **Paramètre de pénalité** indique le paramètre de pénalité qui contrôle le processus de régularisation. Il doit s'agir d'une valeur unique supérieure à 0. Le paramètre par défaut est 0.001.

Critères de convergence des modèles

Convergence des paramètres

Spécifie les critères de convergence pour le paramètre. Il doit s'agir d'une valeur numérique unique appartenant à $[0, 1)$. Valeur par défaut : 0.000001. Pour **Type**, vous pouvez sélectionner **ABSOLUTE** pour appliquer la convergence absolue à l'optimisation interne ou **RELATIVE** pour appliquer la convergence relative à l'optimisation interne. La **Valeur** facultative spécifie un seuil numérique pour le type de convergence.

Convergence des fonctions d'objectif

Spécifie les critères de convergence pour la fonction objectif. Il doit s'agir d'une valeur numérique unique appartenant à $[0, 1)$. La valeur par défaut est 0, ce qui n'applique pas les critères de convergence. Pour **Type**, vous pouvez sélectionner **ABSOLUTE** pour appliquer la convergence absolue à l'optimisation interne ou **RELATIVE** pour appliquer la convergence relative à l'optimisation interne. La **Valeur** facultative spécifie un seuil numérique pour le type de convergence.

Convergence de Hess

Spécifie les critères de convergence de la matrice de Hess. Il doit s'agir d'une valeur numérique unique appartenant à $[0, 1)$. La valeur par défaut est 0, ce qui n'applique pas les critères de convergence. Pour **Type**, vous pouvez sélectionner **ABSOLUTE** pour appliquer la convergence absolue à l'optimisation interne ou **RELATIVE** pour appliquer la convergence relative à l'optimisation interne. La **Valeur** facultative spécifie un seuil numérique pour le type de convergence.

Critères de convergence des résidus

Option permettant de contrôler le processus d'optimisation.

Résidu primaire et double

Applique à la fois le critère de convergence des résidus primal et dual. Ce paramètre est défini par défaut.

Résidu primaire uniquement

Applique le critère de convergence des résidus primaux.

Résidu double uniquement

Applique le critère de convergence des résidus duaux.

Méthode

Paramètre facultatif permettant de spécifier la méthode d'estimation.

Automatique

Sélectionne automatiquement la méthode en fonction de l'exemple de jeu de données. Cette méthode est sélectionnée par défaut. Le champ **Nombre de prédicteurs de seuil** indique le seuil du nombre de prédicteurs et doit être un entier unique supérieur à 1. La valeur par défaut est de 1000.

Newton-Raphson

Applique la méthode de Newton-Raphson.

L-BFGS

Applique l'algorithme BFGS à mémoire limitée. La zone **Mettre à jour** indique le nombre de mises à jour passées qui sont gérées par l'algorithme BFGS à mémoire limitée et qui doivent être un entier unique supérieur ou égal à 1. La valeur par défaut est 5.

Itération**Nombre maximum d'itérations**

Spécifie le nombre maximal d'itérations. Il doit s'agir d'un entier unique appartenant à [1, 300]. La valeur par défaut est 20.

Découpage maximal en deux des étapes

Indique le nombre maximal de step-halving. Il doit s'agir d'un entier unique appartenant à [1, 200]. La valeur par défaut est 5.

Nombre maximal de recherches de ligne

Indique le nombre maximal de recherches de ligne. Il doit s'agir d'un entier unique appartenant à [1, 300]. La valeur par défaut est 20.

Convergence absolue pour le processus d'itération

Indique la convergence absolue pour le processus d'itération externe. Il doit s'agir d'une valeur numérique unique appartenant à (0, 1). Le paramètre par défaut est 0.0001.

Convergence relative pour le processus d'itération

Indique la convergence relative pour le processus d'itération externe. Il doit s'agir d'une valeur numérique unique appartenant à (0, 1). Le paramètre par défaut est 0.01.

Modèles de fragilité partagés paramétriques: Imprimer

Imprimer

Panneau facultatif qui contrôle les sorties de la table.

Détails du codage de facteur

Si cette option est sélectionnée, affiche et imprime les détails de codage des facteurs. Le processus est ignoré si aucun facteur n'est appliqué.

Valeurs initiales affectées aux paramètres de régression

Si cette option est sélectionnée, elle affiche les valeurs initiales utilisées dans le processus d'estimation.

Historique des itérations de modèle

Si cette option est sélectionnée, elle affiche l'historique des itérations de l'analyse de survie. Dans la zone **Nombre d'étapes**, indiquez le nombre d'étapes comprises entre 1 et 99999999. Le paramètre par défaut est 1.

Modèles de fragilité partagés paramétriques: Prévoir

Prévision

Panneau facultatif permettant d'évaluer et d'enregistrer les statistiques prévues dans le jeu de données actif.

Valeurs de temps pour le scoring

Valeurs de temps définies par variable(s) dépendante(s)

Attribue un score à **Prévisions** en fonction de la variable de temps spécifiée pour le modèle de survie paramétrique.

Intervalle ordinaire

Attribue des scores aux **prévisions** en fonction des valeurs de temps futures. La zone **Intervalle de temps** indique l'intervalle de temps et doit être une valeur numérique unique supérieure à 0. La zone **Nombre de périodes** indique le nombre de périodes et doit être un entier numérique unique compris entre 2 et 100.

Durée

Calcule les scores des **prévisions** en fonction de la durée afin de définir les valeurs temporelles futures. Il doit s'agir d'une variable numérique unique.

Prévisions

Survie

Calcule et enregistre les statistiques de survie prévues dans le jeu de données actif. Le nom de variable personnalisée par défaut (ou nom racine) est `PredSurvival`.

Risque

Évalue et enregistre les risques prévus dans le jeu de données actif. Le nom de variable personnalisée par défaut (ou nom racine) est `PredHazard`.

Risque cumulé

Attribue un score et enregistre les risques cumulés prévus dans le jeu de données actif. Le nom de variable personnalisée par défaut (ou nom racine) est `PredCumHazard`.

Survie conditionnelle

Attribue un score et enregistre les statistiques de survie conditionnelle prédites dans le jeu de données actif. Le nom de variable personnalisée par défaut (ou nom racine) est `PredConditionalSurvival`. Le processus est ignoré si `PASTTIME` n'est pas indiqué. Une valeur **Temps de survie passé** est requise et spécifie les valeurs de temps passé pour le scoring. Il doit s'agir d'une variable numérique unique.

Survie inconditionnelle

Calcule et enregistre les statistiques de survie sans condition prédites dans le jeu de données actif. Le mot clé est supprimé par défaut. S'il est spécifié, il peut être suivi d'un nom de variable facultatif fourni par l'utilisateur (ou d'un nom racine) spécifié entre parenthèses. Le nom par défaut est `PredUnCondSurvival`.

Risque inconditionnel

Calcule et enregistre les statistiques de risque inconditionnelles prévues dans le jeu de données actif. Le mot clé est supprimé par défaut. S'il est spécifié, il peut être suivi d'un nom de variable facultatif fourni par l'utilisateur (ou d'un nom racine) spécifié entre parenthèses. Le nom par défaut est `PredUncondHazard`.

Risque de cumul inconditionnel

Calcule et enregistre les statistiques de risque cumulées inconditionnelles prévues dans le jeu de données actif. Le mot clé est supprimé par défaut. S'il est spécifié, il peut être suivi d'un nom de variable facultatif fourni par l'utilisateur (ou d'un nom racine spécifié entre parenthèses). Le nom par défaut est `PredUncondCumHazard`.

Modèles de fragilité partagés paramétriques: Tracé

Tracé

Tracés de fonction

Option permettant de contrôler les tracés de fonction.

Type

Survie

Crée le tracé pour les fonctions de survie inconditionnelles.

Risque

Crée le tracé des fonctions de risque inconditionnelles.

Densité

Crée un tracé pour les fonctions de densité.

Nombre de points à afficher

Indique le nombre de points de fonction compris entre 1 et 200. La valeur par défaut est 100.

Valeurs de covariable pour le tracé

Option permettant de spécifier les valeurs fournies par l'utilisateur et de les affecter aux prédicteurs. Par défaut, les tracés désignés seront créés à la moyenne de chaque covariable en vigueur. Si spécifié, les tracés désignés seront créés en fonction du paramètre du modèle. En présence de variables en double, celle spécifiée en premier est reconnue et le reste est ignoré. Une variable valide doit être contenue dans un effet de modèle. Pour une covariable, la valeur fournie par l'utilisateur doit être numérique. L'omission d'une variable en vigueur indique que la moyenne est utilisée par défaut pour la covariable. Si une valeur non valide est affectée à une variable, le modèle demandé ne sera pas tracé.

Valeurs de facteur pour le tracé

Option permettant de spécifier les valeurs fournies par l'utilisateur et de les affecter aux prédicteurs. Par défaut, les tracés désignés seront créés à la fréquence de catégorie de chaque facteur en vigueur. Si spécifié, les tracés désignés seront créés en fonction du paramètre du modèle. En présence de variables en double, celle spécifiée en premier est reconnue et le reste est ignoré. Une variable valide doit être contenue dans un effet de modèle. L'omission d'une variable en vigueur indique que la fréquence de la catégorie serait utilisée par défaut pour le facteur. Si une valeur non valide est affectée à une variable, le modèle demandé ne sera pas tracé.

Séparer les lignes pour

Option permettant de spécifier une variable catégorielle à l'aide de laquelle les tracés de ligne seront tracés.

Nombre maximal de lignes dans un graphique

Indique le nombre maximal de lignes dans un graphique si l'option **Séparer les lignes pour** est spécifiée. Le paramètre par défaut est 10.

Modèles de fragilité partagés paramétriques: Exportation

Exporter

Sélectionnez **Exporter les informations du modèle dans un fichier XML** pour écrire les informations du modèle et des paramètres dans un fichier PMML à des fins d'évaluation. Vous devez indiquer le répertoire et le nom du fichier PMML à sauvegarder.

Modèles de fragilité partagés paramétriques: Définir des événements

Option permettant de définir le statut. Si la variable de statut est omise, l'échec ou l'événement devient le statut par défaut pour tous les cas.

1. Dans le menu, choisissez:

Analyser > Survie > Modèles de fragilité partagés paramétriques ...

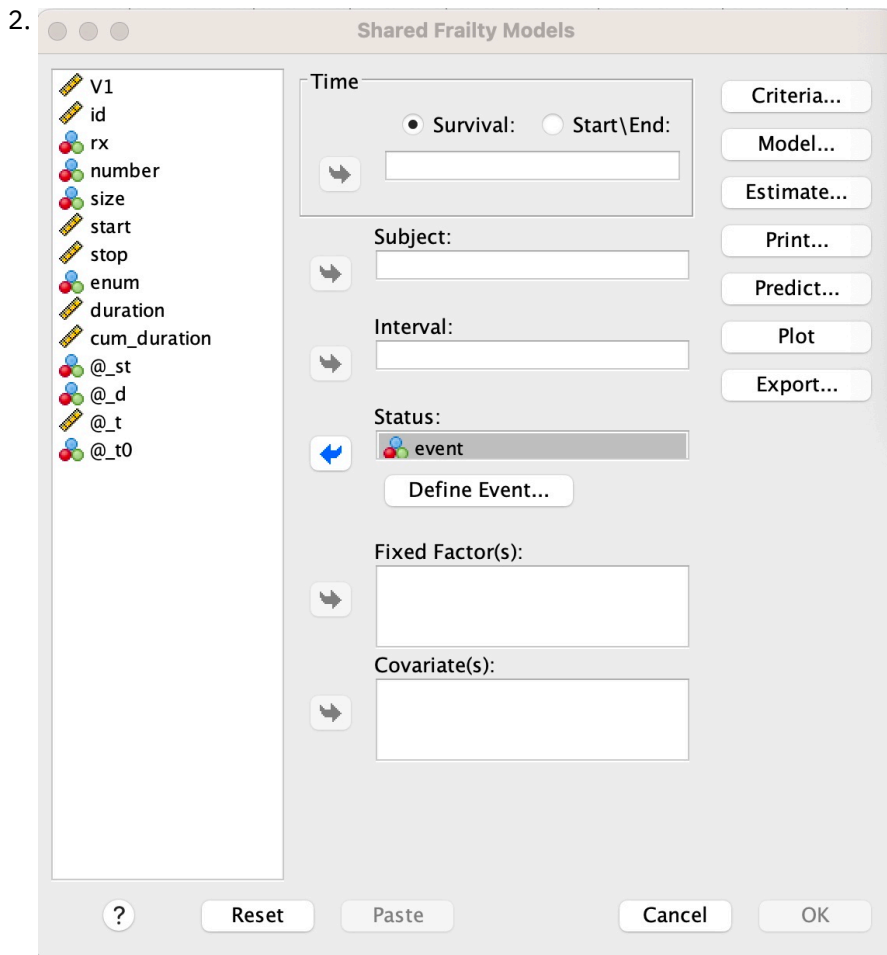


Figure 3. Modèles de fragilité partagés-boîte de dialogue-Statut

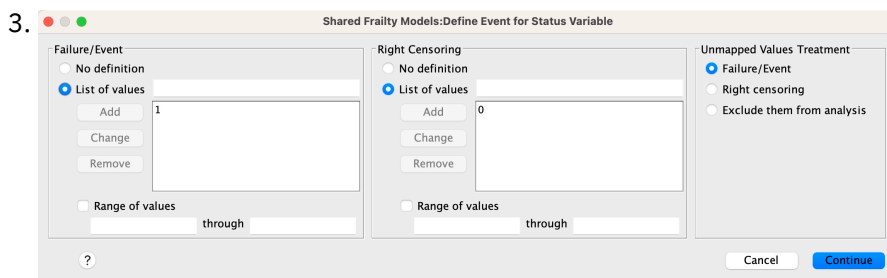


Figure 4. Modèles de fragilité partagés-boîte de dialogue-Statut-Définir un événement

Modèles de fragilité partagés paramétriques-Exemples

Exemple 1

```
SURVREG RECURRENT y WITH x1 BY x2  
/MODEL SUBJECT = id FRAILITY=GAMMA DISTRIBUTION=WEIBULL.
```

Un modèle de survie de fragilité partagée paramétrique est ajusté de y sur une covariable x1 et un facteur x2.

La durée de survie est représentée par une seule variable y.

Les sujets sont identifiés par l'ID variable.

La durée de survie est supposée suivre une distribution de Weibull.

La variance de la fragilité est supposée suivre une distribution Gamma.

Tous les enregistrements valides sont utilisés dans l'analyse de survie.

Exemple 2

```
SURVREG RECURRENT y WITH x1 BY x2
```

```
/MODEL SUBJECT = id FRAILITY=INV_GAUSSIAN DISTRIBUTION=LOG_NORMAL INTERVAL=z.
```

Un modèle de survie de fragilité partagée paramétrique est ajusté de y sur une covariable x1 et un facteur x2.

La durée de survie est représentée par deux variables y1 et y2 indiquant les heures de début et de fin.

Les sujets sont identifiés par l'ID variable.

La durée de survie est supposée suivre une distribution logarithmique normale.

La variance de la fragilité est supposée suivre une distribution gaussienne inverse.

Les intervalles de temps sont définis par la variable z. Pour chaque sujet, la procédure utilise uniquement les enregistrements non conflictuels et exclut de l'analyse tous les enregistrements après le premier statut d'échec.

Exemple 3

```
SURVREG RECURRENT y1 y2 WITH x1 BY x2(1)
```

```
/MODEL SUBJECT = id FRAILITY=INV_GAUSSIAN DISTRIBUTION=LOG_LOGISTIC
```

```
/STATUS VARIABLE=event FAILURE=1 RIGHT=0.
```

Un modèle de survie de fragilité partagée paramétrique est ajusté de y sur une covariable x1 et un facteur x2. Survie

L'heure est représentée par deux variables y1 et y2 indiquant respectivement l'heure de début et l'heure de fin. Pour le facteur x2, la catégorie "1" est désignée comme ligne de base à modéliser.

Les sujets sont identifiés par l'ID variable.

La durée de survie est supposée suivre une distribution log-logistique.

La variance de la fragilité est supposée suivre une distribution gaussienne inverse.

L'événement variable est spécifié pour définir le statut avec 1 et 0 indiquant l'échec et la censure à droite, respectivement.

Exemple 4

```
SURVREG RECURRENT y WITH x1 BY x2
```

```
/MODEL SUBJECT = id
```

```
/STATUS VARIABLE=event FAILURE=1 RIGHT=0
```

```
/PREDICT UNCONDSURVIVAL UNCONDHAZARD UNCONDHAZARD UNCONDCUMHAZARD
```

```
/FUNCTIONPLOT SURVIVAL HAZARD DENSITY PLOTBY (x2).
```

Un modèle de survie de fragilité partagée paramétrique est ajusté de y sur une covariable x1 et un facteur x2. La durée de survie est représentée par une seule variable y.

Les sujets sont identifiés par l'ID variable.

La survie inconditionnelle ou basée sur la population, le risque et le risque cumulé sont évalués et enregistrés dans le jeu de données actif.

Les courbes de survie et de risque inconditionnelles ou basées sur la population sont tracées et séparées par les catégories dans x2.

Exemple 5

```
SURVREG RECURRENT y WITH x1 BY x2
```

```
/MODEL SUBJECT = id FRAILITY=GAMMA DISTRIBUTION=WEIBULL
```

```
/STATUS VARIABLE=event FAILURE=1 RIGHT=0
```

```
/ESTIMATION HCONVERGE=1e-12(RELATIVE) PCONVERGE=0 FCONVERGE=0SELECTFEATURES=TRUE  
PENALTY=0.01.
```

Un modèle de survie de fragilité partagée paramétrique est ajusté de y sur une covariable x1 et un facteur x2. La durée de survie est représentée par une seule variable y.

Les sujets sont identifiés par l'ID variable.

La durée de survie est supposée suivre une distribution de Weibull.

La variance de la fragilité est supposée suivre une distribution Gamma.

Les critères de convergence sont basés sur la matrice de Hess. Il utilise 1e-12 comme convergence relative.

Le modèle inclut un terme de pénalité pour contrôler le processus de régularisation. Le paramètre de pénalité est défini sur 0.01.

Exemple 6

```
SURVREG RECURRENT y WITH x1 BY x2
```

```
/MODEL SUBJECT = id
```

```
/STATUS VARIABLE=infect FAILURE=1 RIGHT=0
```

```
/ESTIMATION MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20.
```

Un modèle de survie de fragilité partagée paramétrique est ajusté de y sur une covariable x1 et un facteur x2. La durée de survie est représentée par une seule variable y.

Les sujets sont identifiés par l'ID variable.

La procédure spécifie le nombre maximal de la recherche de ligne à 100, le nombre maximal d'itérations à 50 et le nombre maximal de step-halving à 20.

Modèles de fragilité partagés paramétriques-Étude de cas pour les données récurrentes

Modèles de fragilité partagés paramétriques-Étude de cas pour les données récurrentes

Nom du cas d'utilisation-Effet secondaire du traitement.

Acteurs-Chercheur et praticien en santé publique.

Préconditions-Ensemble de données nettoyé disponible en fonction de la durée de survie, du statut de l'effet secondaire et des prédicteurs à ajuster.

Description-Patrick, un chercheur en santé publique, enquête sur un échantillon de données qui comprend 20 participants. Ces participants sont recrutés dans une étude sur un effet secondaire léger qui est potentiellement causé par un nouveau traitement. Le concepteur du traitement affirme qu'il n'y aurait pas de différences entre les hommes et les femmes en ce qui concerne l'effet secondaire.

Patrick voudrait évaluer une telle hypothèse. Les variables incluses dans l'échantillon de données sont répertoriées comme suit:

- patID: numéro d'ID permettant d'identifier un participant unique.
- endTime: Durée de survie (en jours) de l'effet secondaire, après un traitement, qui est mesurée depuis le début d'un traitement jusqu'à un effet secondaire rapporté ou censuré dans les 60 jours.
- sideEffect: Statut de l'effet secondaire, statut = 0 si censuré et statut = 1 si l'effet secondaire est signalé.
- âge: âge du participant à la période de recherche.
- femelle: femelle = 0 si mâle et femelle = 1 si femelle.

Plusieurs traitements peuvent s'appliquer, ce qui se traduit par des enregistrements multiples des temps de récurrence mesurés pour un participant donné. L'heure de début est toujours 0 pour chaque enregistrement, qui est omis dans l'échantillon de données. Patrick s'intéresse à la visualisation des fonctions de survie et de danger pour établir une comparaison entre un homme et une femme en contrôlant leur âge et leur fragilité. Il est conscient que les traitements qui sont administrés au même participant sont plus corrélés. En supposant que la durée de survie suit une distribution de Weibull, Patrick décide de créer un modèle de survie à la fragilité partagée paramétrique dans SPSS Statistics pour tenir compte de la dépendance du traitement pour le même participant.

Syntaxe-

```
DATA LIST FREE
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .
BEGIN DATA .
1 45 0 38.00 0
2 26 1 20.00 1
3 58 0 53.00 0
4 31 1 37.00 1
4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILITY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .
```

Synopsis:

La syntaxe spécifiée par Patrick désigne endTime comme une variable de temps dépendante unique. La procédure suppose automatiquement que l'heure de début est 0 pour chaque enregistrement. Les variables âge et femme sont modélisées comme covariable et facteur, respectivement. Les temps de

survie de récurrence sont supposés suivre une distribution de Weibull. Le terme de fragilité non observé est supposé suivre une distribution Gamma, et sa composante de variance est modélisée. En ce qui concerne les sorties, le tableau récapitulatif du modèle fournit les informations sur la procédure et le modèle. Le tableau récapitulatif de traitement des cas fournit une liste complète du statut d'échec ou de censure, ainsi que les cas exclus de l'analyse.

Dans l'échantillon de données de Patrick, tous les enregistrements sont valides et inclus dans l'analyse. En comparant le log de vraisemblance avec celui du modèle correspondant sans le composant de fragilité, le modèle de fragilité partagée ne parvient pas à atteindre un niveau significatif (valeur $p = 0.168$). Patrick se demande s'il est nécessaire d'inclure un terme de fragilité partagée dans le modèle. Le facteur d'accélération estimé d'un participant masculin est 1.017, qui est obtenu en calculant l'exposant du coefficient de régression estimé 0.017 de [femelle = 0.0]. L'intervalle de confiance à 95% qui lui est associé est (.688, 1.504). Ces résultats suggèrent qu'un individu masculin a presque le même facteur d'accélération qu'un individu féminin avec le même âge et la même fragilité. Au niveau de la population, Patrick trace les courbes de survie et de risque inconditionnelles séparément pour les hommes et les femmes qui sont évalués à la moyenne d'âge de l'échantillon (37.45 ans).

Patrick confirme que, pour toute valeur fixe du temps de survie, un mâle et une femelle en moyenne sont censés avoir la même probabilité de survie. Fait intéressant, malgré une forme unimodale qui est montrée dans le tableau des risques inconditionnels, Patrick découvre que dans une période de 60 jours, le risque de la population augmente en fait. Ce comportement peut impliquer l'existence de l'effet de fragilité. Pour étudier plus en détail les effets secondaires causés par les traitements, Patrick peut continuer avec un modèle sans la composante de fragilité et comparer le comportement des mâles et des femelles. En outre, il peut envisager d'effectuer un suivi auprès des participants pendant une période de plus de 60 jours afin de recueillir davantage de données.

Remarques

Le présent document a été développé pour des produits et des services proposés aux Etats-Unis. et peut être mis à disposition par IBM dans d'autres langues. Vous pouvez toutefois devoir détenir une copie du produit ou une version du produit dans cette langue pour pouvoir y accéder.

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Contactez votre interlocuteur IBM pour plus d'informations sur les produits et services disponibles dans votre région. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service IBM puisse être utilisé. Tout autre élément fonctionnellement équivalent peut être utilisé, s'il n'enfreint aucun droit d'IBM. Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier lui-même les installations et applications réalisées avec des produits, logiciels ou services non expressément référencés par IBM.

IBM peut détenir des brevets ou des demandes de brevet couvrant les produits décrits dans le présent document. La remise de ce document ne vous accorde aucun droit de licence sur ces brevets ou demandes de brevet. Si vous désirez recevoir des informations concernant l'acquisition de licences, veuillez en faire la demande par écrit à l'adresse suivante :

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119Armonk, NY 10504-1785U.S.A.

Les informations sur les licences concernant les produits IBM utilisant un jeu de caractères double octet peuvent être obtenues par écrit à l'adresse suivante :

Intellectual Property Licensing

Legal and Intellectual Property Law

IBM Japan Ltd.

19-21, Nihonbashi-Hakozakicho, Chuo-kuTokyo 103-8510, Japon

LE PRESENT DOCUMENT EST LIVRE "EN L'ETAT" SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Le présent document peut contenir des inexactitudes ou des coquilles. Il est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, apporter des améliorations et des modifications aux produits et aux logiciels décrits dans ce document.

Les références à des sites Web non IBM sont fournies à titre d'information uniquement et n'impliquent en aucun cas une adhésion aux données qu'ils contiennent. Les éléments figurant sur ces sites Web ne font pas partie des éléments du présent produit IBM et l'utilisation de ces sites relève de votre seule responsabilité.

IBM pourra utiliser ou diffuser, de toute manière qu'elle jugera appropriée et sans aucune obligation de sa part, tout ou partie des informations qui lui seront fournies.

Les licenciés souhaitant obtenir des informations permettant : (i) l'échange des données entre des logiciels créés de façon indépendante et d'autres logiciels (dont celui-ci) et (ii) l'utilisation mutuelle des données ainsi échangées, doivent adresser leur demande à :

IBM Director of Licensing

IBM Corporation

North Castle Drive, MD-NC119Armonk, NY 10504-1785U.S.A.

Ces informations peuvent être soumises à des conditions particulières, prévoyant notamment le paiement d'une redevance.

Le logiciel sous licence décrit dans ce document et tous les éléments sous licence disponibles pour ce dernier sont fournis par IBM d'après les termes du contrat client IBM, des conditions internationales d'utilisation de logiciels IBM ou de tout contrat équivalent entre nous.

Les données de performances et les exemples de clients ne sont présentés qu'à des fins d'illustration. Les performances réelles peuvent varier en fonction des configurations et des conditions d'exploitation.

Les informations concernant des produits non IBM ont été obtenues auprès des fournisseurs de ces produits, par l'intermédiaire d'annonces publiques ou via d'autres sources disponibles. IBM n'a pas testé ces produits et ne peut confirmer l'exactitude de leurs performances ni leur compatibilité. Elle ne peut recevoir aucune réclamation concernant des produits non IBM. Toute question concernant les performances de produits non IBM doit être adressée aux fournisseurs de ces produits.

Toute instruction relative aux intentions d'IBM pour ses opérations à venir est susceptible d'être modifiée ou annulée sans préavis, et doit être considérée uniquement comme un objectif.

Le présent document peut contenir des exemples de données et de rapports utilisés couramment dans l'environnement professionnel. Ces exemples mentionnent des noms fictifs de personnes, de sociétés, de marques ou de produits à des fins d'illustration ou d'explication uniquement. Toute ressemblance avec des noms de personnes et de sociétés serait purement fortuite.

LICENCE DE COPYRIGHT :

Le présent logiciel contient des exemples de programmes d'application en langage source destinés à illustrer les techniques de programmation sur différentes plateformes d'exploitation. Vous avez le droit de copier, de modifier et de distribuer ces exemples de programmes sous quelque forme que ce soit et sans paiement d'aucune redevance à IBM, à des fins de développement, d'utilisation, de vente ou de distribution de programmes d'application conformes aux interfaces de programmation des plateformes pour lesquelles ils ont été écrits ou aux interfaces de programmation IBM. Ces exemples n'ont pas été testés en détails, ni dans toutes les conditions. C'est pourquoi IBM ne peut pas garantir ou assurer la fiabilité, la serviceabilité ou le fonctionnement de ces programmes. Les modèles de programmes sont fournis "en l'état", sans garantie d'aucune sorte. IBM ne sera en aucun cas responsable des dommages liés à l'utilisation de ces programmes.

Toute copie totale ou partielle de ces programmes exemples et des oeuvres qui en sont dérivées doit comprendre une notice de copyright, libellée comme suit :

© Copyright IBM Corp. 2021. Des parties de ce code sont proviennent d'IBM Corp. Programmes exemples.

© Copyright IBM Corp. 1989 - 2021. All rights reserved.

Marques

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines Corp., dans de nombreux pays. Les autres noms de services et de produits peuvent être des marques d'IBM ou d'autres sociétés. La liste actualisée de toutes les marques d'IBM est disponible sur la page Web "Copyright and trademark information" à www.ibm.com/legal/copytrade.shtml.

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques d'Adobe Systems Incorporated aux Etats-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 d'Intel Corporation ou de ses filiales aux Etats-Unis et dans certains autres pays.

Linux est une marque de Linus Torvalds aux Etats-Unis et/ou dans certains autres pays.

Microsoft, Windows, Windows NT et le logo Windows sont des marques de Microsoft Corporation aux Etats-Unis et/ou dans certains autres pays.

UNIX est une marque enregistrée de The Open Group aux Etats-Unis et/ou dans certains autres pays.

Java ainsi que tous les logos et toutes les marques incluant Java sont des marques d'Oracle et/ou de ses sociétés affiliées.

Index

A

- AFT de survie
 - Boîte de dialogue de survie-Variables de catégorie [38](#)
- Analyse de régression logistique [2](#)
- analyse de survie
 - dans la régression d'arête de noyau [30](#)
- Analyse par la méthode des probits
 - critères [12](#)
 - Définition d'une plage [11](#)
 - fonctions supplémentaires de la commande [12](#)
 - Impact relatif médian [12](#)
 - Intervalle de confiance de référence [12](#)
 - itérations [12](#)
 - statistiques [12](#)
 - Taux de réponse naturel [12](#)
 - Test de parallélisme [12](#)
- Arête de noyau
 - alpha [30](#)
 - choix du modèle [30](#)
 - coef0 [30](#)
 - degré [30](#)
 - gamma [30](#)
 - Modèle unique [30](#)

C

- Cellules contenant 0 observation
 - Dans la régression logistique multinomiale [9](#)
- classification
 - Dans la régression logistique multinomiale [6](#)
- Constante
 - dans la régression linéaire [5](#)
 - Inclusion ou exclusion [7](#)
- Contraintes sur les paramètres
 - Dans la régression non linéaire [21](#)
- contrastes
 - Dans la régression logistique [4](#)
- Covariables
 - Dans la régression logistique [4](#)
- covariables catégorielles [4](#)
- covariables de chaîne
 - Dans la régression logistique [4](#)
- Critère de convergence
 - Dans la régression logistique multinomiale [9](#)

D

- découpage en deux des étapes
 - Dans la régression logistique multinomiale [9](#)
- delta
 - Comme correction pour les cellules contenant 0 observation [9](#)
- Différence de bêta
 - Dans la régression logistique [4](#)
- Distance de Cook

- Distance de Cook (*suite*)
 - Dans la régression logistique [4](#)

E

- élimination descendante
 - Dans la régression logistique [3](#)
- estimations des paramètres
 - Dans la régression logistique multinomiale [8](#)

F

- Fonction de déviance
 - Pour l'estimation de la valeur d'échelle de dispersion [9](#)

H

- historique des itérations
 - Dans la régression logistique multinomiale [9](#)

I

- Impact relatif médian
 - Dans les analyses par la méthode des probits [12](#)
- intervalles de confiance
 - Dans la régression logistique multinomiale [8](#)
- Intervalle de confiance de référence
 - Dans les analyses par la méthode des probits [12](#)
- itérations
 - Dans la régression logistique [5](#)
 - Dans la régression logistique multinomiale [9](#)
 - Dans les analyses par la méthode des probits [12](#)

K

- Kernel Ridge Regression
 - affichage [31](#)
 - enregistrer [31](#)
 - nuages [31](#)
 - paramètres [31](#)
 - paramètres de grille [31](#)
 - Plis de validation croisée [31](#)
- khi-carré de Pearson
 - Pour l'estimation de la valeur d'échelle de dispersion [9](#)
 - qualité d'ajustement [8](#)

L

- Log de vraisemblance
 - Dans la pondération estimée [23](#)
 - Dans la régression logistique multinomiale [8](#)
- Loi des réponses décroissantes de Metcherlich
 - Dans la régression non linéaire [20](#)

M

- matrice de corrélation
 - Dans la régression logistique multinomiale [8](#)
- matrice de covariance
 - Dans la régression logistique multinomiale [8](#)
- modalité de référence
 - Dans la régression logistique multinomiale [8](#)
- Modèle de densité
 - Dans la régression non linéaire [20](#)
- Modèle de densité de rendement
 - Dans la régression non linéaire [20](#)
- Modèle de Gompertz
 - Dans la régression non linéaire [20](#)
- Modèle de Johnson-Schumacher
 - Dans la régression non linéaire [20](#)
- Modèle de log modifié
 - Dans la régression non linéaire [20](#)
- Modèle de Michaelis-Menten
 - Dans la régression non linéaire [20](#)
- Modèle de Morgan-Mercer-Florin
 - Dans la régression non linéaire [20](#)
- Modèle de Peal-Reed
 - Dans la régression non linéaire [20](#)
- Modèle de Richards
 - Dans la régression non linéaire [20](#)
- Modèle de Verhulst
 - Dans la régression non linéaire [20](#)
- Modèle de Von Bertalanffy
 - Dans la régression non linéaire [20](#)
- Modèle de Weibull
 - Dans la régression non linéaire [20](#)
- Modèle des rapports cubiques
 - Dans la régression non linéaire [20](#)
- Modèle des rapports quadratiques
 - Dans la régression non linéaire [20](#)
- Modèle gaussien
 - Dans la régression non linéaire [20](#)
- Modèles avec effets principaux
 - Dans la régression logistique multinomiale [7](#)
- Modèles de fragilité paramétriques
 - variables de statut de survie [45](#)
- Modèles de fragilité partagés paramétriques
 - analyse [39](#)
 - critères [40](#)
 - estimation [42](#)
 - exporter [45](#)
 - imprimer [43](#)
 - modèle [41](#)
 - prévoir [44](#)
 - tracer [45](#)
- Modèles de temps d'échec accéléré paramétriques
 - analyse [32](#)
 - critères [33](#)
 - estimation [35](#)
 - exporter [38](#)
 - imprimer [36](#)
 - modèle [34](#)
 - prévoir [37](#)
 - tracer [37](#)
- modèles factoriels complets
 - Dans la régression logistique multinomiale [7](#)
- Modèles non linéaires
 - Dans la régression non linéaire [20](#)

- modèles personnalisés
 - Dans la régression logistique multinomiale [7](#)

P

- Pondération estimée
 - Afficher ANOVA et les estimations [24](#)
 - Enregistrer meilleure pondération dans nouvelle variable [24](#)
 - exemple [23](#)
 - fonctions supplémentaires de la commande [25](#)
 - historique des itérations [24](#)
 - Log de vraisemblance [23](#)
 - statistiques [23](#)

Q

- qualité d'ajustement
 - Dans la régression logistique multinomiale [8](#)

R

- R2 de Cox et Snell
 - Dans la régression logistique multinomiale [8](#)
- R2 de McFadden
 - Dans la régression logistique multinomiale [8](#)
- R2 de Nagelkerke
 - Dans la régression logistique multinomiale [8](#)
- Rapport de vraisemblance
 - Pour l'estimation de la valeur d'échelle de dispersion [9](#)
 - qualité d'ajustement [8](#)
- Régression à deux niveaux par la méthode des moindres carrés
 - Covariance des paramètres [26](#)
 - enregistrement des nouvelles variables [26](#)
- Régression asymptotique
 - Dans la régression non linéaire [20](#)
- Régression linéaire
 - fonctions supplémentaires de la commande [23](#)
 - Interprétation des résultats [22](#)
 - Pondération estimée [23](#)
 - Régression par les doubles moindres carrés [25](#)
- Régression logistique
 - Binaire [1](#)
 - coefficients [2](#)
 - Constante [5](#)
 - contrastes [4](#)
 - covariables catégorielles [4](#)
 - covariables de chaîne [4](#)
 - Définition de la règle de sélection [3](#)
 - Définition de règle [3](#)
 - enregistrement des nouvelles variables [4](#)
 - exemple [2](#)
 - fonctions supplémentaires de la commande [6](#)
 - itérations [5](#)
 - limite du classement [5](#)
 - Mesures d'influence [4](#)
 - Méthodes de sélection des variables [3](#)
 - Options d'affichage [5](#)
 - Probabilité pour méthode pas à pas [5](#)
 - résiduels [4](#)
 - Statistique de qualité d'ajustement de Hosmer-Lemeshow [5](#)

Régression logistique (*suite*)
statistiques [2](#)
tracés et statistiques [5](#)
valeur prédite [4](#)
Régression logistique binaire [1, 2](#)
Régression logistique multinomiale
critères [9](#)
enregistrer [10](#)
exportation des informations du modèle [10](#)
fonctions supplémentaires de la commande [10](#)
modalité de référence [8](#)
Modèles [7](#)
statistiques [8](#)
Régression logistique SPSS [2](#)
Régression non linéaire
Algorithme de Levenberg-Marquardt [22](#)
Contraintes sur les paramètres [21](#)
Dérivées [22](#)
Enregistrement de nouvelles variables [22](#)
Erreur standard estimée par le bootstrap [22](#)
exemple [18](#)
Fonction de perte [21](#)
logique conditionnelle [19](#)
Méthode d'estimation [22](#)
Modèle segmenté [19](#)
Modèles non linéaires communs [20](#)
paramètres [20](#)
Programmation quadratique séquentielle [22](#)
Résidus [22](#)
Statistiques [18](#)
Valeurs initiales [20](#)
valeurs prédites [22](#)
Régression par les doubles moindres carrés
exemple [25](#)
fonctions supplémentaires de la commande [26](#)
statistiques [25](#)
Variables instrumentales [25](#)
Régression probit
exemple [10](#)
statistiques [10](#)
Régression quantile
affichage [16](#)
critères [14](#)
enregistrer [17](#)
exemple [12](#)
exporter [18](#)
modèle [15](#)
statistiques [12](#)
Régression restreinte
Dans la régression non linéaire [21](#)
rupture du contrat de travail
Dans la régression logistique multinomiale [9](#)

S

sélection ascendante
Dans la régression logistique [3](#)
sélection progressive
Dans la régression logistique [3](#)
Dans la régression logistique multinomiale [7](#)
singularité
Dans la régression logistique multinomiale [9](#)
Statistique de qualité d'ajustement de Hosmer-Lemeshow
Dans la régression logistique [5](#)

T

Table de classification
Dans la régression logistique multinomiale [8](#)
Tableaux de probabilités des cellules
Dans la régression logistique multinomiale [8](#)
Tables de survie
variables de statut de survie [38](#)
Test de parallélisme
Dans les analyses par la méthode des probits [12](#)

V

Valeur d'échelle de dispersion
Dans la régression logistique multinomiale [9](#)
valeur influente
Dans la régression logistique [4](#)

