

IBM SPSS Advanced Statistics 24

IBM

Important

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 109.

Notice d'édition

La présente édition s'applique à la version 24.0.0 d'IBM SPSS Statistics et à toutes les éditions et modifications ultérieures sauf mention contraire dans les nouvelles éditions.

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.

Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. Les informations qui y sont fournies sont susceptibles d'être modifiées avant que les produits décrits ne deviennent eux-mêmes disponibles. En outre, il peut contenir des informations ou des références concernant certains produits, logiciels ou services non annoncés dans ce pays. Cela ne signifie cependant pas qu'ils y seront annoncés.

Pour plus de détails, pour toute demande d'ordre technique, ou pour obtenir des exemplaires de documents IBM, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial.

Vous pouvez également consulter les serveurs Internet suivants :

- <http://www.fr.ibm.com> (serveur IBM en France)
- <http://www.ibm.com/ca/fr> (serveur IBM au Canada)
- <http://www.ibm.com> (serveur IBM aux Etats-Unis)

*Compagnie IBM France
Direction Qualité
17, avenue de l'Europe
92275 Bois-Colombes Cedex*

© Copyright IBM France 2015. Tous droits réservés.

Table des matières

Avis aux lecteurs canadiens v

Chapitre 1. Introduction aux Statistiques avancées 1

Chapitre 2. Analyses GLM - Tests multivariés 3

Modèle d'analyse GLM - Multivariés 4
Termes construits 4
Somme des carrés 5
GLM - Tests multivariés : Contrastes 6
Types de contraste 6
GLM - Tests multivariés : Tracés de profil 6
Comparaisons multivariées post hoc GLM 7
Enregistrement GLM 8
GLM - Tests multivariés : Options 9
Fonctions supplémentaires de la commande GLM 10

Chapitre 3. GLM - Mesures répétées . . . 11

GLM - Mesures répétées : Définir des facteurs 13
GLM - Mesures répétées : Modèles 14
Termes construits 14
Somme des carrés 14
GLM - Mesures répétées : Contrastes 15
Types de contraste 15
GLM - Mesures répétées : Tracés de profil 16
GLM - Mesures répétées : Comparaisons post hoc 16
GLM - Mesures répétées : Enregistrer 17
Options de GLM - Mesures répétées 18
Fonctions supplémentaires de la commande GLM 19

Chapitre 4. Analyse des composantes de variance 21

Modèle des composantes de variance 22
Termes construits 22
Options des composantes de variance 22
Somme des carrés (Composantes de variance) 23
Composantes de variance : Enregistrer dans un nouveau fichier 24
Fonctions supplémentaires de la commande VARCOMP 24

Chapitre 5. Modèles mixtes linéaires . . . 25

Modèles mixtes linéaires : Sélectionner des variables répétées/des sujets 26
Modèles mixtes linéaires : Effets fixes 27
Termes de construction non imbriquée 27
Construction de termes imbriqués 27
Somme des carrés 28
Modèles mixtes linéaires : Effets aléatoires 28
Estimation des modèles mixtes linéaires 29
Modèles mixtes linéaires : Statistiques 30
Moyennes EM des modèles mixtes linéaires 30
Modèles mixtes linéaires : Enregistrer 31

Fonctions supplémentaires de la commande MIXED 31

Chapitre 6. Modèles linéaires généralisés 33

Modèles linéaires généralisés - Réponse 35
Modèles linéaires généralisés - Catégorie de référence 36
Prédicteurs de modèles linéaires généralisés 36
Modèles linéaires généralisés - Options 37
Modèle des modèles linéaires généralisés 37
Estimation des modèles linéaires généralisés 38
Modèles linéaires généralisés - Valeurs initiales 39
Statistiques des modèles linéaires généralisés 39
Moyennes EM des modèles linéaires généralisés 41
Enregistrement des modèles linéaires généralisés 42
Export de modèles linéaires généralisés 43
Fonctions supplémentaires de la commande GENLIN 43

Chapitre 7. Equations d'estimation généralisées 45

Equations d'estimation généralisées : type de modèle 47
Réponse des équations d'estimation généralisées 49
Catégorie de référence des équations d'estimation généralisées 49
Prédicteurs des équations d'estimation généralisées 50
Options des équations d'estimation généralisées 50
Modèle des équations d'estimation généralisées 50
Estimation des équations d'estimation généralisées 51
Valeurs initiales des équations d'estimation généralisées 52
Statistiques des équations d'estimation généralisées 53
Moyennes EM des équations d'estimation généralisées 54
Enregistrement des équations d'estimation généralisées 55
Export d'équations d'estimation généralisées 56
Fonctions supplémentaires de la commande GENLIN 57

Chapitre 8. Modèles linéaires mixtes généralisés 59

Obtention d'un modèle linéaire mixte généralisé 60
Cible 60
Effets fixes 63
Ajout d'un terme personnalisé 63
Effets aléatoires 64
Bloc d'effets aléatoires 64
Pondération et décalage 65
Options de création générales 66
Estimation 66
Moyennes estimées 67
Enregistrement 68

Vue Modèle	69
Récapitulatif du modèle	69
Structure des données	69
Valeurs prédites en fonction des valeurs observées	69
Classification	70
Effets fixes.	70
Coefficients fixes	70
Covariances des effets aléatoires	71
Paramètres de covariance.	71
Moyennes estimées :Effets significatifs	72
Moyennes estimées : Effets personnalisés	72
Chapitre 9. Analyse log-linéaire	
Sélection du modèle	75
Analyse log-linéaire : Définir plage	76
Modèle d'analyse log-linéaire	76
Termes construits	76
Options de l'analyse log-linéaire de sélection de modèle	76
Fonctions supplémentaires de la commande HILOGLINEAR	77
Chapitre 10. Analyse log-linéaire généralisée	79
Modèle d'analyse log-linéaire généralisée	80
Termes construits	80
Options de l'analyse log-linéaire généralisée	80
Analyse log-linéaire généralisée : Enregistrer	81
Fonctions supplémentaires de la commande GENLOG	81
Chapitre 11. Analyse log-linéaire Logit 83	
Modèle d'analyse log-linéaire Logit	84
Termes construits	84
Options de l'analyse log-linéaire Logit	85
Analyse log-linéaire Logit : Enregistrer	85
Fonctions supplémentaires de la commande GENLOG	86
Chapitre 12. Tables de survie	87
Tables de survie : Définir l'événement pour la variable de statut	88
Tables de survie : Définir plage.	88
Tables de survie : Options	88
Fonctions supplémentaires de la commande SURVIVAL	89

Chapitre 13. Analyse de survie de Kaplan-Meier	91
Kaplan-Meier : Définir un événement pour la variable de statut	92
Kaplan-Meier : Comparer les niveaux du facteur	92
Kaplan-Meier : Enregistrer les nouvelles variables	92
Kaplan-Meier : Options	93
Fonctions supplémentaires de la commande KM	93
Chapitre 14. Analyse de la régression de Cox.	95
Modèle de Cox... Définir des variables qualitatives	96
Tracés de la régression de Cox	96
Modèle de Cox : Enregistrer les nouvelles variables	97
Options de la régression de Cox	97
Modèle de Cox : Définir un événement pour la variable de statut	98
Fonctions supplémentaires de la commande COXREG	98
Chapitre 15. Calcul des prédicteurs chronologiques	99
Calculer un prédicteur chronologique.	99
Régression de Cox munie de fonctions supplémentaires pour les prédicteurs chronologiques	100
Chapitre 16. Méthodes de codification des variables qualitatives	101
Déviation.	101
Simple.	101
Helmert	102
Différence	102
Polynomial	102
Répété.	103
Spécial	103
Indicateur	104
Chapitre 17. Structures de covariance 105	
Remarques	109
Marques	111
Index	113

Avis aux lecteurs canadiens

Le présent document a été traduit en France. Voici les principales différences et particularités dont vous devez tenir compte.

Illustrations

Les illustrations sont fournies à titre d'exemple. Certaines peuvent contenir des données propres à la France.

Terminologie

La terminologie des titres IBM peut différer d'un pays à l'autre. Reportez-vous au tableau ci-dessous, au besoin.

IBM France	IBM Canada
ingénieur commercial	représentant
agence commerciale	succursale
ingénieur technico-commercial	informaticien
inspecteur	technicien du matériel

Claviers

Les lettres sont disposées différemment : le clavier français est de type AZERTY, et le clavier français-canadien de type QWERTY.

OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

- les pages de codes 850 (multilingue) et 863 (français-canadien),
- le code pays 002,
- le code clavier CF.

Nomenclature

Les touches présentées dans le tableau d'équivalence suivant sont libellées différemment selon qu'il s'agit du clavier de la France, du clavier du Canada ou du clavier des États-Unis. Reportez-vous à ce tableau pour faire correspondre les touches françaises figurant dans le présent document aux touches de votre clavier.

France	Canada	Etats-Unis
⌂ (Pos1)	⌂	Home
Fin	Fin	End
⬆ (PgAr)	⬆	PgUp
⬇ (PgAv)	⬇	PgDn
Inser	Inser	Ins
Suppr	Suppr	Del
Echap	Echap	Esc
Attn	Intrp	Break
Impr écran	ImpEc	PrtSc
Verr num	Num	Num Lock
Arrêt défil	Défil	Scroll Lock
🔒 (Verr maj)	FixMaj	Caps Lock
AltGr	AltCar	Alt (à droite)

Brevets

Il est possible qu'IBM détienne des brevets ou qu'elle ait déposé des demandes de brevets portant sur certains sujets abordés dans ce document. Le fait qu'IBM vous fournisse le présent document ne signifie pas qu'elle vous accorde un permis d'utilisation de ces brevets. Vous pouvez envoyer, par écrit, vos demandes de renseignements relatives aux permis d'utilisation au directeur général des relations commerciales d'IBM, 3600 Steeles Avenue East, Markham, Ontario, L3R 9Z7.

Assistance téléphonique

Si vous avez besoin d'assistance ou si vous voulez commander du matériel, des logiciels et des publications IBM, contactez IBM direct au 1 800 465-1234.

Chapitre 1. Introduction aux Statistiques avancées

L'option Statistiques avancées fournit des procédures comportant des options de modélisation plus avancées que celles disponibles dans l'option Statistiques de base.

- La procédure GLM – Tests multivariés étend le modèle linéaire général fourni par GLM - Univarié pour permettre l'utilisation de plusieurs variables dépendantes. Une autre extension, GLM - Mesures répétées, permet des mesures répétées de plusieurs variables dépendantes.
- L'analyse des composantes de variance est un outil spécial de décomposition de la variabilité d'une variable dépendante en composantes fixes et aléatoires.
- La procédure des modèles mixtes linéaires développe le modèle linéaire général pour permettre aux données de présenter des variabilités en corrélation et des variabilités non constantes. Le modèle linéaire mixte offre donc la possibilité de modéliser non seulement les moyennes des données, mais également leurs variances et covariances.
- Les modèles linéaires généralisés (GZLM) assouplissent l'hypothèse de normalité du terme d'erreur et nécessitent seulement que la variable dépendante soit linéairement liée aux prédicteurs au moyen d'une transformation ou d'une fonction de lien. Les équations d'estimation généralisées (GEE) étendent la procédure GZLM pour permettre les mesures répétées.
- L'analyse log-linéaire généralisée vous permet d'ajuster des modèles pour des données d'effectif classées de manière croisée. L'analyse log-linéaire de sélection de modèle peut vous aider à choisir entre plusieurs modèles.
- L'analyse log-linéaire Logit vous permet d'ajuster des modèles log-linéaires pour analyser la relation entre une variable dépendante catégorielle et un ou plusieurs prédicteurs catégoriels.
- L'analyse de survie est disponible via l'option de tables de survie et permet d'examiner la distribution de variables de la durée à l'événement, éventuellement par niveau d'un facteur. L'analyse de survie de Kaplan-Meier permet d'examiner la distribution de variables de la durée à l'événement, éventuellement par niveau d'un facteur, ou de produire des analyses distinctes par niveau d'une variable de stratification. La régression de Cox permet de modéliser la durée jusqu'à un événement spécifié en fonction des valeurs de certaines covariables.

Chapitre 2. Analyses GLM - Tests multivariés

GLM - Tests multivariés fournit un modèle de régression et une analyse de variance pour plusieurs variables dépendantes par un ou plusieurs facteurs ou covariables. Les facteurs divisent la population en groupes. Le modèle linéaire général vous permet de tester les hypothèses nulles pour les effets des facteurs sur la moyenne des divers regroupements d'une distribution jointe de variables dépendantes. Vous pouvez rechercher les interactions entre les facteurs ainsi que les effets des différents facteurs. En outre, les effets et les interactions des covariables avec les facteurs peuvent être inclus. Pour l'analyse de la régression, les variables indépendantes (prédicteur) sont spécifiées comme covariables.

Vous pouvez tester les modèles équilibrés comme déséquilibrés. Un plan est équilibré si chaque cellule de ce modèle contient le même nombre d'observations. Dans un modèle multivarié, les sommes des carrés dues aux effets dans le modèle et les sommes des carrés résiduelles sont représentées sous forme matricielle plutôt que sous forme d'échelle comme dans l'analyse univariée. Ces matrices sont appelées matrices SSCP (sommes des carrés et produits croisés). Si vous avez spécifié plusieurs variables dépendantes, les analyses de variance multivariées à l'aide de la trace de Pillai, du lambda de Wilks, de la trace de Hotelling et de la plus grande racine de Roy avec la statistique F approchée sont fournies, de même que l'analyse de variance univariée pour chaque variable dépendante. L'analyse GLM - Tests multivariés teste non seulement les hypothèses mais elle produit également des estimations.

Vous disposez de contrastes *a priori* communs pour effectuer les tests d'hypothèse. En outre, lorsqu'un test F global se révèle significatif, vous pouvez utiliser les tests post hoc pour évaluer les différences entre les moyennes spécifiques. Les moyennes marginales estimées fournissent des estimations des valeurs moyennes estimées pour les cellules dans le modèle et les tracés de profil (tracés d'interaction) de ces moyennes vous permettent de visualiser plus facilement certaines des relations. Les différents tests de comparaison post hoc sont effectués séparément pour chaque variable dépendante.

Les résidus, les prévisions, la distance de Cook et les valeurs influentes peuvent être enregistrées sous forme de nouvelles variables dans votre fichier de données pour vérifier les hypothèses. Vous disposez également d'une matrice SSCP des résidus (matrice carrée des sommes des carrés et des produits croisés des résidus), d'une matrice de covariance des résidus (matrice SSPC des résidus divisée par les degrés de liberté des résidus) et d'une matrice de corrélation des résidus (forme standardisée de la matrice de covariance des résidus).

La pondération WLS vous permet de spécifier une variable utilisée pour pondérer les observations pour une analyse pondérée (WLS) des moindres carrés, peut-être pour compenser les différents niveaux de précision des mesures.

Exemple : Un fabricant de plastique mesure trois propriétés des films plastiques : la résistance, la brillance et l'opacité. Deux taux d'extrusion et deux différentes quantités d'additif sont testés. Les trois propriétés sont mesurées pour chaque combinaison de taux d'extrusion et de quantité d'additif. Le fabricant découvre que le taux d'extrusion et la quantité d'additif produisent des résultats significatifs mais que l'interaction de ces deux facteurs n'est pas significative.

Méthodes : Les sommes des carrés de type I, II, III et IV peuvent servir à évaluer les différentes hypothèses. Le type III est la valeur par défaut.

Statistiques : Tests de plage post hoc et comparaisons multiples : différence la moins significative, Bonferroni, Sidak, Scheffé, F multiple de Ryan-Einot-Gabriel-Welsch, plage multiple de Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls, test de Tukey, b de Tukey, Duncan, GT2 de Hochberg, Gabriel, test t de Waller Duncan, Dunnett (unilatéral et bilatéral), T2 de Tamhane, T3 de Dunnett, Games-Howell et C de Dunnett. Statistiques descriptive : moyennes observées, écarts types et

effectifs de toutes les variables dépendantes dans toutes les cellules ; test de Levene pour l'homogénéité de variance ; test *M* de Box pour l'homogénéité des matrices de covariance des variables dépendantes et enfin, test de Bartlett pour la sphéricité.

Tracés : Dispersion par niveau, résiduels et profils (interaction).

Remarques sur les données de l'analyse GLM - Tests multivariés

Données : Les variables dépendantes doivent être quantitatives. Les facteurs sont qualitatifs, et peuvent comporter des valeurs numériques ou alphanumériques. Les covariables sont des variables quantitatives liées à la variable dépendante.

Hypothèses : Pour les variables dépendantes, les données forment un échantillon aléatoire de vecteurs dans une population normale multivariée. Dans cette population, les matrices de variance/covariance de toutes les cellules sont les mêmes. L'analyse de variance supporte les écarts à la normalité, bien que les données doivent être symétriques. Pour vérifier les hypothèses, vous pouvez utiliser des tests de l'homogénéité des variances (y compris le *M* de Box) et les tracés de dispersion par niveau. Vous pouvez également étudier les résidus et les tracés résiduels.

Procédures apparentées : Utilisez la procédure Explorer pour étudier les données avant de mener une analyse de variance. Si vous n'avez qu'une seule variable dépendante, utilisez GLM - Univarié. Si vous avez mesuré les mêmes variables dépendantes à plusieurs occasions pour chaque sujet, utilisez GLM - Mesures répétées.

Obtenir des tables GLM - Tests multivariés

1. A partir des menus, sélectionnez :
Analyse > Modèle linéaire général > Multivarié...
2. Sélectionnez au moins deux variables dépendantes.

Vous pouvez éventuellement spécifier Facteur(s) fixé(s), Covariable(s) et Pondération WLS.

Modèle d'analyse GLM - Multivariés

Spécifier le modèle : Un modèle factoriel général contient tous les effets principaux des facteurs, des covariables et toutes les interactions facteur/facteur. Il ne contient pas de d'interactions de covariable. Sélectionnez **Autre** pour indiquer un sous-ensemble d'interactions ou des interactions facteur/covariable. Vous devez indiquer tous les termes à inclure dans le modèle.

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

Modèle : Le modèle dépend de la nature de vos données. Après avoir sélectionné **Autre**, vous pouvez choisir les effets principaux et les interactions qui présentent un intérêt pour votre analyse.

Somme des carrés Méthode de calcul des sommes des carrés. Pour les modèles équilibrés ou non, auxquels aucune cellule ne manque, le type III est la méthode le plus fréquemment utilisée.

Inclure une constante au modèle : La constante est généralement incluse dans le modèle. Si vous partez du principe que les données passent par l'origine, vous pouvez exclure la constante.

Termes construits

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau de toutes les variables sélectionnées. Il s'agit de la valeur par défaut.

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.

Somme des carrés

Pour ce modèle, vous pouvez choisir un type de sommes des carrés. Le type III est le plus courant et c'est la valeur par défaut.

Type I : Cette méthode est également appelée décomposition hiérarchique de la somme des carrés. Chaque terme est ajusté uniquement pour le terme qui le précède dans le modèle. La somme des carrés de type I est généralement utilisée pour :

- Une analyse de la variance équilibrée dans laquelle tout effet principal est spécifié avant les effets d'interaction de premier ordre, et chaque effet de premier ordre spécifié avant ceux de second ordre, et ainsi de suite.
- Un modèle de régression polynomial dans lequel les termes d'ordre inférieur sont spécifiés avant ceux d'ordre supérieur.
- Un modèle par imbrication pur dans lequel le premier effet spécifié est imbriqué dans le second et le second spécifié dans le troisième, etc. (Cette forme d'imbrication peut être spécifiée par la syntaxe uniquement.)

Type II : Cette méthode calcule les sommes des carrés d'un effet dans le modèle ajusté pour tous les autres effets « appropriés ». Un effet approprié est un effet qui correspond à tous les effets qui ne contiennent pas l'effet à étudier. La méthode des sommes des carrés de type II sert généralement pour :

- Une analyse de la variance équilibrée.
- Tout modèle qui contient des effets factoriels principaux uniquement.
- Tout modèle de régression.
- Un plan par imbrication pur. (Cette forme d'imbrication peut être spécifiée par la syntaxe.)

Type III : Valeur par défaut. Cette méthode calcule les sommes des carrés d'un effet dans le plan comme les sommes des carrés, ajustée pour tout autre effet qui ne contient pas le plan et orthogonale pour tous les effets qui le contiennent. Les sommes de carrés de type III présentent l'avantage essentiel qu'elles ne varient pas avec les fréquences de cellule tant que la forme générale d'estimabilité reste constante. Ce type de somme des carrés est donc souvent considéré comme utile pour les modèles déséquilibrés auxquels aucune cellule ne manque. Dans le plan factoriel sans cellule manquante, cette méthode est équivalente à la technique de Yates des carrés moyens pondérés. La méthode des sommes des carrés de type III sert généralement pour :

- Tous les modèles énumérés dans les types I et II.
- Tous les modèles équilibrés ou non qui ne contiennent pas de cellules vides.

Type IV : Cette méthode est conçue pour une situation dans laquelle il manque des cellules. Pour chaque effet F dans le plan, si F n'est inclus dans aucun autre effet, Type IV = Type III = Type II. Si F est inclus dans d'autres effets, le Type IV distribue les contrastes à effectuer parmi les paramètres dans F sur tous les effets de niveau supérieur de façon équitable. La méthode des sommes des carrés de type IV sert généralement pour :

- Tous les modèles énumérés dans les types I et II.
- Tous les modèles équilibrés ou non qui contiennent des cellules vides.

GLM - Tests multivariés : Contrastes

Les contrastes servent à tester si les niveaux d'un effet sont suffisamment différents les uns des autres. Vous pouvez spécifier un contraste pour chaque facteur dans un modèle. Les contrastes représentent des combinaisons linéaires des paramètres.

Le test des hypothèses est fondé sur l'hypothèse nulle $\mathbf{LBM} = \mathbf{0}$, où \mathbf{L} représente la matrice des coefficients des contrastes, \mathbf{M} la matrice d'identité (dont la dimension est égale au nombre de variables dépendantes) et \mathbf{B} le vecteur de paramètre. Si un contraste est spécifié, une matrice \mathbf{L} est créée de telle sorte que les colonnes correspondant au facteur correspondent au contraste. Les colonnes restantes sont ajustées de telle sorte que la matrice \mathbf{L} puisse être estimée.

Outre le test univarié qui utilise les statistiques F et des intervalles de confiance simultanée de type Bonferroni fondé sur la distribution t de Student pour les différences de contraste sur toutes les variables dépendantes, vous disposez des tests multivariés à l'aide de la trace de Pillai, du lambda de Wilks, de la trace de Hotelling et de la plus grande racine de Roy.

Les contrastes fournis sont déviation, simple, différence, Helmert, répétée et modèle polynomial. Pour les contrastes d'écart et simple, vous pouvez choisir si la catégorie de référence est la première ou la dernière.

Types de contraste

Déviation : Compare la moyenne de chaque niveau (hormis une catégorie de référence) à la moyenne de tous les niveaux (grande moyenne). Les niveaux du facteur peuvent être de n'importe quel ordre.

Simple : Compare la moyenne de chaque niveau à celle d'un niveau donné. Ce type de contraste est utile lorsqu'il y a un groupe de contrôle. Vous pouvez prendre la première ou la dernière catégorie en référence.

Différence : Compare la moyenne de chaque niveau (hormis le premier) à la moyenne des niveaux précédents. (Parfois appelé contrastes d'Helmert inversé.)

Helmert : Compare la moyenne de chaque niveau de facteur (hormis le dernier) à la moyenne des niveaux suivants.

Répété : Compare la moyenne de chaque niveau (hormis le premier) à la moyenne du niveau suivant.

Polynomial : Compare l'effet linéaire, l'effet quadratique, l'effet cubique etc. Le premier degré de liberté contient l'effet linéaire sur toutes les catégories, le second degré l'effet quadratique, etc. Ces contrastes servent souvent à estimer les tendances polynomiales.

GLM - Tests multivariés : Tracés de profil

Les tracés de profil (tracés d'interaction) sont utiles pour comparer les moyennes marginales dans votre modèle. Un tracé de profil est un tracé en ligne dont chaque point indique la moyenne marginale estimée d'une variable dépendante (ajustée pour les covariables) à un niveau du facteur. Les niveaux d'un second facteur peuvent servir à dessiner des courbes distinctes. Chaque niveau dans un troisième facteur peut servir à créer un tracé distinct. Tous les facteurs sont disponibles pour les tracés. Les tracés de profil sont créés pour chaque variable dépendante.

Un tracé de profil pour un facteur montre si la moyenne marginale estimée est croissante ou décroissante sur les niveaux. Pour au moins deux facteurs, des courbes parallèles indiquent qu'il n'y a pas d'interaction entre les facteurs, ce qui signifie que vous recherchez les niveaux d'un seul facteur. Les courbes non parallèles indiquent une interaction.

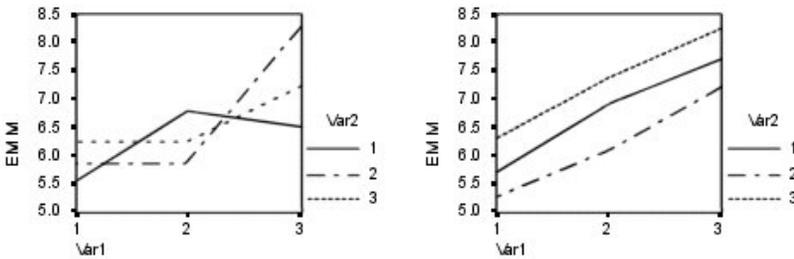


Figure 1. Tracé non parallèle (gauche) et tracé parallèle (droite)

Après avoir sélectionné des facteurs pour l'axe horizontal afin de spécifier un tracé et, éventuellement, des facteurs pour des courbes ou des tracés distincts, vous devez ajouter le tracé à la liste Tracés.

Comparaisons multivariées post hoc GLM

Test de comparaison multiple post hoc : Lorsque vous avez déterminé qu'il existe des différences parmi les moyennes, les tests de plages post hoc et de comparaisons multiples appariées peuvent déterminer les moyennes qui diffèrent. Les comparaisons sont effectuées sur des valeurs non-ajustées. Les tests post hoc sont effectués séparément pour chaque variable dépendante.

Les tests de différence significative de Bonferroni et Tukey servent généralement comme tests de comparaison multiples. Le **test de Bonferroni**, fondé sur la statistique t de Student, ajuste le niveau de signification observé en fonction du nombre de comparaisons multiples qui sont effectuées. Le **test t de Sidak** ajuste également le niveau de signification et fournit des limites plus strictes que le test de Bonferroni. Le **test de Tukey** utilise la statistique de plage de Student pour effectuer des comparaisons appariées entre les groupes et fixe le taux d'erreur empirique au taux d'erreur du regroupement de toutes les comparaisons appariées. Lorsque vous testez un grand nombre de paires de moyennes, le test de Tukey est plus efficace que celui de Bonferroni. Lorsqu'il y a peu de paires, Bonferroni est plus efficace.

Le **GT2 de Hochberg** est similaire au test de Tukey mais il utilise un modulo maximum selon Student. Le test de Tukey est généralement plus efficace. Le **test de comparaison appariée de Gabriel** utilise également le modulo maximum selon Student. Il est plus efficace que le GT2 de Hochberg lorsque les tailles des cellules sont inégales. Le test de Gabriel offre plus de souplesse lorsque les tailles des cellules divergent beaucoup.

Le **test de comparaison multiple appariée de Dunnett** compare un ensemble de traitements à une simple moyenne de contrôle. La dernière catégorie est la catégorie de contrôle par défaut. Vous pouvez également choisir la première catégorie. Vous pouvez également choisir un test unilatéral ou bilatéral. Pour tester que la moyenne à un certain niveau (hormis la catégorie de contrôle) du facteur n'est pas égale à celle de la catégorie de contrôle, utilisez le test double-face. Pour tester si la moyenne est inférieure, à un certain niveau du facteur, à celle de la catégorie de contrôle, sélectionnez **< Contrôle**. Pour tester si la moyenne est supérieure, à un certain niveau du facteur, à celle de la catégorie de contrôle, sélectionnez **> Contrôle**.

Ryan, Einot, Gabriel et Welsch (R-E-G-W) ont développé deux tests de plages multiples descendants. Les procédures multiples descendantes testent d'abord que toutes les moyennes sont égales. Si toutes les moyennes ne sont pas égales, l'égalité est testée sur des sous-ensembles de moyennes. Le **F de R-E-G-W** est fondé sur le test F et le **Q de R-E-G-W** est fondé sur la plage de Student. Ces tests sont plus efficaces que le test de plages multiples de Duncan et Student-Newman-Keuls (procédures multiples descendantes), mais ils sont conseillés lorsque les cellules sont de taille inégale.

Lorsque les variances sont inégales, utilisez le **T2 de Tamhane** (test de comparaisons appariées conservatif fondé sur un test t), le **T3 de Dunnett** (test de comparaison appariée fondé sur le modulo maximal de Student), le **test de comparaison appariée de Games-Howell** (parfois flexible) ou le **C de Dunnett** (test de comparaison appariée fondé sur la plage de Student).

Le **test de plages multiples de Duncan**, Student-Newman-Keuls (**S-N-K**) et le **b de Tukey** sont des tests de plage qui classifient les moyennes de groupe et calculent une valeur de plage. Ces tests ne sont pas utilisés aussi souvent que les tests évoqués précédemment.

Le **test t de Waller-Duncan** utilise une approche de Bayes. Ce test de plage utilise la moyenne harmonique de la taille de l'échantillon lorsque les échantillons sont de tailles différentes.

Le niveau de signification du **test de Scheffé** est conçu pour permettre toutes les combinaisons linéaires possibles des moyennes de groupe à tester, pas seulement appariée, disponibles dans cette fonction. Il en résulte que le test de Scheffé est souvent plus strict que les autres, ce qui signifie qu'une plus grande différence de moyenne est nécessaire pour être significative.

Le test de comparaison multiple appariée de différence la moins significative (**LSD**) est équivalent aux divers tests *t* individuels entre toutes les paires des groupes. L'inconvénient de ce test est qu'il n'essaie pas d'ajuster le niveau d'importance observée pour les comparaisons multiples.

Tests affichés : Les comparaisons appariées sont proposées pour LSD, Sidak, Bonferroni, Games et Howell, T2 et T3 de Tamhane, C et T3 de Dunnett. Des sous-ensembles homogènes pour les tests de plage sont proposés pour S-N-K, *b* de Tukey, Duncan, *F* et *Q* de R-E-G-W et Waller. Le test de Tukey, le GT2 de Hochberg, le test de Gabriel et le test de Scheffé sont à la fois des tests de comparaison multiple et des tests de plage.

Enregistrement GLM

Vous pouvez enregistrer les prévisions par le modèle, les résidus et les mesures associées sous forme de nouvelles variables dans l'éditeur de données. La plupart de ces variables peuvent servir à étudier les hypothèses relatives aux données. Pour enregistrer les valeurs afin de les utiliser dans une autre session IBM® SPSS Statistics, vous devez enregistrer le fichier de données en cours.

Prévisions : Valeurs que le modèle estime pour chaque observation.

- *Non standardisés*. Valeur prévue par le modèle pour la variable dépendante.
- *Pondérés*. Valeurs estimées non standardisées pondérées. Disponibles uniquement lorsqu'une variable WLS a été préalablement sélectionnée.
- *Erreur standard*. Estimation de l'écart type de la valeur moyenne de la variable expliquée pour les unités statistiques qui ont les mêmes valeurs pour les valeurs explicatives.

Diagnostics : Mesures permettant d'identifier les observations avec des combinaisons inhabituelles de valeurs pour les variables indépendantes et les observations qui peuvent avoir un impact important sur le modèle.

- *Distance 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. Si la distance de Cook est élevée, l'exclusion d'une observation changerait substantiellement la valeur des coefficients.
- *Valeurs influentes*. Valeurs influentes non centrées. Mesure de l'influence d'un point sur l'ajustement de la régression.

Résidus : Un résidu non standardisé correspond à la valeur réelle de la variable dépendante moins la valeur estimée par le modèle. Les résidus standardisés, de Student et supprimés sont également disponibles. Si vous avez choisi une variable de pondération, les résidus standardisés pondérés sont disponibles.

- *Non standardisés*. Différence entre la valeur observée et la valeur prévue par le modèle.
- *Pondérés*. Résidus estimés non standardisés pondérés. Disponibles uniquement lorsqu'une variable WLS a été préalablement sélectionnée.

- *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.
- *De Student*. Résidu, divisé par une estimation de son écart type, qui varie d'une observation à l'autre, selon la distance entre les valeurs et la moyenne des variables indépendantes pour chaque observation.
- *Supprimées*. Résidu d'une observation lorsque celle-ci est exclue du calcul des coefficients de régression. Il s'agit de la différence entre la valeur de la variable dépendante et la prévision ajustée.

Statistiques à coefficients : Ecrit une matrice variance-covariance des estimations des paramètres du modèle dans un nouveau jeu de données de la session en cours ou dans un fichier de données externe au format IBM SPSS Statistics. D'autre part, pour chaque variable dépendante, il y aura une ligne d'estimations, une ligne de valeurs de signification pour les statistiques t correspondant aux estimations et une ligne de degrés de liberté résiduels. Pour un modèle multivarié, il y a les mêmes lignes pour chaque variable dépendante. Vous pouvez utiliser ces fichiers de matrice dans les autres procédures qui lisent des fichiers de matrice.

GLM - Tests multivariés : Options

Des statistiques facultatives sont disponibles à partir de cette boîte de dialogue. Ces statistiques sont calculées à l'aide de modèle à effets fixes.

Moyenne marginale estimée : Sélectionnez les facteurs et les interactions pour lesquels vous souhaitez obtenir des estimations de la moyenne marginale de la population dans les cellules. Ces moyennes sont ajustées pour les covariables, si elles existent. Les interactions sont disponibles si vous avez spécifié un modèle personnalisé.

- **Comparer les effets principaux** : Propose des comparaisons appariées non corrigées des moyennes marginales estimées pour tout effet principal dans le modèle, à la fois pour les facteurs intersujets et intrasujets. Ceci n'est valable que si les effets principaux sont sélectionnés dans la liste Afficher les moyennes.
- **Ajustement intervalle de confiance** : Sélectionnez l'ajustement aux intervalles de confiance et à la significativité des intervalles en adoptant l'une des méthodes suivantes : la différence de moindre signification (LSD), l'ajustement Bonferroni ou l'ajustement de Sidak. Cet élément est disponible uniquement si **Comparer les effets principaux** est sélectionné.

Afficher : Sélectionnez **Statistiques descriptives** pour produire des moyennes, des écarts-types et des effectifs pour toutes les variables dépendantes de toutes les cellules. L'option **Estimation d'effet de taille** fournit une valeur partielle de Eta carré pour chaque effet et chaque estimation. La statistique d'Eta carré décrit la proportion de la variabilité totale imputable au facteur. Sélectionnez **Puissance observée** pour obtenir la puissance du test lorsque l'autre hypothèse est définie sur la base de la valeur observée. Sélectionnez **Estimation des paramètres** pour produire des estimations de paramètres, des erreurs standard, des tests t , des intervalles de confiance et la puissance observée de chaque test. Vous pouvez afficher les **matrices SSCP** d'hypothèse et d'erreur, la **matrice SSCP des résidus** et le test de Bartlett de sphéricité de la matrice de covariance des résidus.

L'option des **tests d'homogénéité** produit le test de Levene d'homogénéité de la variance pour chaque variable dépendante sur toutes les combinaisons de niveaux des facteurs inter-sujets, uniquement pour les facteurs inter-sujets. Les tests d'homogénéité incluent également le test M de Box d'homogénéité des matrices de covariance des variables dépendantes sur toutes les combinaisons de niveaux des facteurs inter-sujets. Les options des tracés de dispersion par niveau et résiduels sont utiles pour vérifier les hypothèses sur les données. Ceci n'est pas valable s'il n'y a pas de facteurs. Sélectionnez **Tracés résiduels** pour produire un tracé résiduel observé/estimé/standardisé pour chaque variable dépendante. Ces tracés sont utiles pour vérifier l'hypothèse de variance égale. Sélectionnez **Manque d'ajustement** pour vérifier si la relation entre la variable dépendante et les variables indépendantes peut être convenablement décrite par le modèle. **Fonction générale estimée** vous permet de construire des tests d'hypothèses personnalisés basés sur la fonction générale estimée. Les lignes de n'importe quelle matrice des coefficients de contraste sont des combinaisons linéaires de la fonction générale estimée.

Niveau de signification : Vous souhaitez peut-être ajuster le niveau de signification utilisé dans les tests post hoc et le niveau de confiance utilisé pour construire des intervalles de confiance. La valeur spécifiée est également utilisée pour calculer l'intensité observée pour le test. Lorsque vous spécifiez un niveau de signification, le niveau associé des intervalles de confiance est affiché dans la boîte de dialogue.

Fonctions supplémentaires de la commande GLM

Ces fonctions s'appliquent aux analyses univariée, multivariée ou de mesures répétées. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Spécifier les effets imbriqués dans un plan (à l'aide de la sous-commande DESIGN).
- Spécifier les tests d'effets par rapport à une combinaison linéaire d'effets ou une valeur (à l'aide de la sous-commande TEST).
- Spécifier de multiples contrastes (à l'aide de la sous-commande CONTRAST).
- Inclure les valeurs manquantes de l'utilisateur (à l'aide de la sous-commande MISSING).
- Spécifier les critères EPS (à l'aide de la sous-commande CRITERIA).
- Construire une matrice **L** personnalisée, une matrice **M** ou une matrice **K** (à l'aide des sous-commandes LMATRIX, MMATRIX ou KMATRIX).
- Pour les contrastes simples ou d'écart, spécifier une catégorie de référence intermédiaire (à l'aide de la sous-commande CONTRAST).
- Spécifier les mesures pour les contrastes polynomiaux (à l'aide de la sous-commande CONTRAST).
- Spécifier des termes d'erreur pour les comparaisons post hoc (à l'aide de la sous-commande POSTHOC).
- Calculer les moyennes marginales estimées pour chaque facteur ou interaction entre facteurs parmi les facteurs de la liste (à l'aide de la sous-commande EMMEANS).
- Attribuer des noms aux variables temporaires (à l'aide de la sous-commande SAVE).
- Construire un fichier de matrice de corrélation (à l'aide de la sous-commande OUTFILE).
- Construire un fichier de type matrice de données qui contient les statistiques provenant de la table ANOVA inter-sujets (à l'aide de la sous-commande OUTFILE).
- Enregistrer la matrice du plan dans un nouveau fichier de données (à l'aide de la sous-commande OUTFILE).

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

Chapitre 3. GLM - Mesures répétées

La procédure GLM - Mesures répétées fournit une analyse de variance lorsque la même mesure est effectuée plusieurs fois avec le même sujet ou observation. Si des facteurs inter-sujets sont spécifiés, ils divisent la population en groupes. Le modèle linéaire général vous permet de tester les hypothèses nulles relatives aux effets des facteurs inter-sujets et intra-sujets. Vous pouvez rechercher les interactions entre les facteurs ainsi que les effets des différents facteurs. En outre, on peut inclure les effets de covariables constantes et ceux des interactions des covariables avec les facteurs.

Dans un plan de mesures répétées et multivarié, les variables dépendantes représentent les mesures de plusieurs variables pour les différents niveaux des facteurs intra-sujets. Par exemple, vous pouvez avoir mesuré à la fois le pouls et la respiration à trois dates différentes sur chaque sujet.

GLM - Mesures répétées fournit à la fois des analyses univariées et multivariées pour les données des mesures répétées. Vous pouvez tester les modèles équilibrés comme déséquilibrés. Un plan est équilibré si chaque cellule de ce modèle contient le même nombre d'observations. Dans un modèle multivarié, les sommes des carrés dues aux effets dans le modèle et les sommes des carrés résiduelles sont représentées sous forme matricielle plutôt que sous forme d'échelle comme dans l'analyse univariée. Ces matrices sont appelées matrices SSCP (sommes des carrés et produits croisés). GLM - Mesures répétées teste non seulement les hypothèses mais elle produit également des estimations.

Vous disposez des contrastes *a priori* les plus courants pour tester les hypothèses relatives aux facteurs inter-sujets. En outre, lorsqu'un test *F* global se révèle significatif, vous pouvez utiliser les tests post hoc pour évaluer les différences entre les moyennes spécifiques. Les moyennes marginales estimées fournissent des estimations des valeurs moyennes estimées pour les cellules dans le modèle et les tracés de profil (tracés d'interaction) de ces moyennes vous permettent de visualiser plus facilement certaines des relations.

Les résidus, les prévisions, la distance de Cook et les valeurs influentes peuvent être enregistrées sous forme de nouvelles variables dans votre fichier de données pour vérifier les hypothèses. Vous disposez également d'une matrice SSCP des résidus (matrice carrée des sommes des carrés et des produits croisés des résidus), d'une matrice de covariance des résidus (matrice SSPC des résidus divisée par les degrés de liberté des résidus) et d'une matrice de corrélation des résidus (forme standardisée de la matrice de covariance des résidus).

La pondération WLS vous permet de spécifier une variable utilisée pour pondérer les observations pour une analyse pondérée (WLS) des moindres carrés, peut-être pour compenser les différents niveaux de précision des mesures.

Exemple : Douze étudiants sont affectés à des groupes soit de faible anxiété, soit de forte anxiété en fonction de leurs scores à des tests de classement de l'anxiété. Ce classement est appelé facteur inter-sujets, car elle répartit les individus en groupes. On propose aux étudiants quatre essais sur une tâche d'apprentissage et on enregistre le nombre d'erreurs à chaque essai. Ces erreurs sont enregistrées dans des variables distinctes et un facteur intra-sujets (essai) est défini avec quatre niveaux pour les quatre essais. On trouve que l'effet de l'essai est significatif alors que l'interaction essai/anxiété ne l'est pas.

Méthodes : Les sommes des carrés de type I, II, III et IV peuvent servir à évaluer les différentes hypothèses. Le type III est la valeur par défaut.

Statistiques : Tests de plage post hoc et comparaisons multiples (pour les facteurs inter-sujets) : différence la moins significative, Bonferroni, Sidak, Scheffé, *F* multiple de Ryan-Einot-Gabriel-Welsch, plage multiple de Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls, test de Tukey, *b* de Tukey, Duncan,

GT2 de Hochberg, Gabriel, test t de Waller Duncan, Dunnett (unilatéral et bilatéral, T2 de Tamhane, T3 de Dunnett, Games-Howell et C de Dunnett. Statistiques descriptives : moyennes observées, écarts types et effectifs pour toutes les variables dépendantes de toutes les cellules, test de Levene pour l'homogénéité de la variance, test M de Box et test de Mauchly pour la sphéricité.

Tracés : Dispersion par niveau, résiduels et profils (interaction).

Remarques sur les données de GLM - Mesures répétées

Données : Les variables dépendantes doivent être quantitatives. Les facteurs inter-sujets divisent l'échantillon en sous-groupes discrets, tels que homme et femme. Ces facteurs sont qualitatifs, et peuvent comporter des valeurs numériques ou alphanumériques. Les facteurs intrasujets sont définis dans la boîte de dialogue GLM Mesures répétées : Définir des facteurs. Les covariables sont des variables quantitatives liées à la variable dépendante. Dans une analyse des mesures répétées, elles doivent rester constantes à chaque niveau d'une variable intra-sujets.

Le fichier de données doit contenir un groupe de variables pour chaque groupe de mesures sur les sujets. L'ensemble dispose d'une variable pour chaque répétition de la mesure dans le groupe. Un facteur intra-sujet est défini pour le groupe, le nombre de niveaux étant égal au nombre de répétitions. Par exemple, on peut mesurer la pondération sur différentes journées. Si des mesures de la même propriété sont effectuées sur 5 jours, le facteur intra-sujets peut être effectué par *jour* et disposer de cinq niveaux.

S'il y a plusieurs facteurs intra-sujets, le nombre de mesures sur chaque sujet est égal au produit du nombre de niveaux de chaque facteur. Par exemple, si on effectue des mesures trois fois par jour pendant 4 jours, le nombre total de mesures est de 12 pour chaque sujet. Les facteurs intrasujets sont alors effectués par *jour* (4) et *heure* (3).

Hypothèses : Une analyse des mesures répétées peut être conduite de deux façons, univariée et multivariée.

L'approche univariée (également connu comme l'approche par tracé de dispersion ou approche mixte) considère les variables dépendantes comme des réponses aux niveaux des facteurs intra-sujets. Les mesures sur un sujet doivent être un échantillon d'une distribution normale multivariée, et les matrices variance/covariance sont les mêmes sur toutes les cellules formées par les effets inter-sujets. Certaines hypothèses sont faites sur la matrice variance/covariance des variables dépendantes. La validité de la statistique F utilisée dans l'approche univariée peut être garantie si la matrice variance/covariance est sphérique dans la forme (Huynh et Mandeville, 1979).

Pour tester cette hypothèse, utilisez le test de sphéricité de Mauchly qui réalise un test de sphéricité sur la matrice variance/covariance d'une variable dépendante une fois orthogonalisée. Le test de Mauchly est automatiquement affiché pour une analyse des mesures répétées. Ce test est très efficace sur les petits échantillons. Pour les plus grands, le test peut être significatif même si l'impact d'une violation de la sphéricité sur les résultats est faible. Si le test est très significatif, l'hypothèse de sphéricité peut être prise en compte. Cependant, si le test est peu significatif et que l'hypothèse de sphéricité ne semble pas respectée, vous pouvez effectuer un ajustement des degrés de liberté du numérateur et du dénominateur afin de valider la statistique F univariée. Vous disposez de trois estimations de cet ajustement, appelé **epsilon** dans GLM - Mesures répétées. Numérateur et dénominateur doivent tous deux être multipliés par epsilon et la signification du rapport F doit être évaluée avec les nouveaux degrés de liberté.

L'approche multivariée considère les mesures sur un sujet comme un échantillon d'une distribution normale multivariée et que les matrices variance/covariance sont les mêmes sur les cellules formées par les effets inter-sujets. Pour tester si les matrices variance/covariance sur les cellules sont les mêmes, vous pouvez utiliser le test M .

Procédures apparentées : Utilisez la procédure Explorer pour étudier les données avant de mener une analyse de variance. S'il n'y a *pas* de mesures répétées sur chaque sujet, utilisez GLM – Univarié ou GLM

- Multivarié. S'il n'y a que deux mesures par sujet, (par exemple pré-test et post-test) et qu'il n'y a pas de facteurs inter-sujets, utilisez le test T pour échantillons appariés.

Obtention de GLM - Mesures répétées

1. A partir des menus, sélectionnez :

Analyse > Modèle linéaire général > Mesures répétées...

2. Saisissez le nom d'un facteur intrasujets et son nombre de niveaux.

3. Cliquez sur **Ajouter**.

4. Répétez ces opérations pour chaque facteur intrasujets.

Pour définir des facteurs de mesure pour un plan des mesures répétées multivariées :

5. Entrez le nom de la mesure.

6. Cliquez sur **Ajouter**.

Lorsque tous les facteurs et toutes les mesures sont définis :

7. Cliquez sur **Définir**.

8. Sélectionnez une variable dépendante qui corresponde à chaque combinaison des facteurs intra-sujets (et éventuellement des mesures) dans la liste.

Pour changer les positions des variables, utilisez les flèches Haut et Bas.

Pour effectuer des modifications des facteurs intrasujets, vous pouvez ouvrir de nouveau la boîte de dialogue GLM Mesures Répétées : Définir des facteurs sans pour autant fermer la boîte principale. Vous pouvez également spécifier des facteurs intersujets et des covariables.

GLM - Mesures répétées : Définir des facteurs

GLM - Mesures répétées analyse des groupes de variables dépendantes liées qui représentent différentes mesures d'un même attribut. La boîte de dialogue vous laisse définir un ou plusieurs facteurs intra-sujets pour GLM - Mesures répétées. Notez que l'ordre de facteurs intra-sujets est très important. Chaque facteur constitue un niveau dans un facteur précédent.

Pour utiliser des mesures répétées, vous devez définir vos données correctement. Vous devez définir des facteurs intra-sujets dans cette boîte de dialogue. Notez que ces facteurs ne correspondent pas à des variables existantes dans vos données mais à des facteurs que vous définissez ici.

Exemple : Dans une étude sur la perte pondérale, on suppose que la pondération de plusieurs individus est mesurée chaque semaine pendant 5 semaines. Dans le fichier de données, chaque individu constitue un sujet ou une observation. Les pondérations des différentes semaines sont enregistrées dans les variables *pondération1*, *pondération2*, etc. Le sexe de la personne est enregistré dans une autre variable. Les pondérations, mesurées plusieurs fois pour chaque individu, peuvent être regroupées en définissant un facteur intra-sujets. Ce facteur peut être appelé *semaine*, défini avec cinq niveaux. Dans la boîte de dialogue principale, les variables *pondération1*, ..., *pondération5* sont utilisées pour affecter les cinq niveaux différents du paramètre *SEMAINE*. La variable du fichier de données qui regroupe les hommes et les femmes (*sexe*) peut être spécifiée comme facteur inter-sujets pour étudier les différences entre les hommes et les femmes.

Mesures : Si les sujets sont testés sur plusieurs mesures à chaque fois, sélectionnez ces mesures. Par exemple, le pouls et la respiration peuvent être mesurés sur chaque sujet chaque jour pendant une semaine. Ces mesures n'existent pas en tant que variables dans le fichier de données mais sont définies ici. Un modèle disposant de plusieurs mesures est généralement appelé modèle des mesures répétées multivariées.

GLM - Mesures répétées : Modèles

Spécifier le modèle : Un modèle factoriel général contient tous les effets principaux des facteurs, des covariables et toutes les interactions facteur/facteur. Il ne contient pas de d'interactions de covariable. Sélectionnez **Autre** pour indiquer un sous-ensemble d'interactions ou des interactions facteur/covariable. Vous devez indiquer tous les termes à inclure dans le modèle.

Inter-sujets : Les facteurs inter-sujets et les covariables sont répertoriés.

Modèle : Le modèle dépend de la nature de vos données. Après avoir sélectionné **Autre**, vous pouvez choisir les effets et les interactions intra-sujets, ainsi que les effets et les interactions inter-sujets qui présentent un intérêt pour votre analyse.

Somme des carrés : Méthode de calcul des sommes des carrés pour le modèle inter-sujets. Pour les modèles inter-sujets, équilibrés ou non, auxquels il ne manque aucune cellule, la somme des carrés de type III est la méthode la plus fréquemment utilisée.

Termes construits

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau de toutes les variables sélectionnées. Il s'agit de la valeur par défaut.

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.

Somme des carrés

Pour ce modèle, vous pouvez choisir un type de sommes des carrés. Le type III est le plus courant et c'est la valeur par défaut.

Type I : Cette méthode est également appelée décomposition hiérarchique de la somme des carrés. Chaque terme est ajusté uniquement pour le terme qui le précède dans le modèle. La somme des carrés de type I est généralement utilisée pour :

- Une analyse de la variance équilibrée dans laquelle tout effet principal est spécifié avant les effets d'interaction de premier ordre, et chaque effet de premier ordre spécifié avant ceux de second ordre, et ainsi de suite.
- Un modèle de régression polynomial dans lequel les termes d'ordre inférieur sont spécifiés avant ceux d'ordre supérieur.
- Un modèle par imbrication pur dans lequel le premier effet spécifié est imbriqué dans le second et le second spécifié dans le troisième, etc. (Cette forme d'imbrication peut être spécifiée par la syntaxe uniquement.)

Type II : Cette méthode calcule les sommes des carrés d'un effet dans le modèle ajusté pour tous les autres effets « appropriés ». Un effet approprié est un effet qui correspond à tous les effets qui ne contiennent pas l'effet à étudier. La méthode des sommes des carrés de type II sert généralement pour :

- Une analyse de la variance équilibrée.
- Tout modèle qui contient des effets factoriels principaux uniquement.

- Tout modèle de régression.
- Un plan par imbrication pur. (Cette forme d'imbrication peut être spécifiée par la syntaxe.)

Type III : Valeur par défaut. Cette méthode calcule les sommes des carrés d'un effet dans le plan comme les sommes des carrés, ajustée pour tout autre effet qui ne contient pas le plan et orthogonale pour tous les effets qui le contiennent. Les sommes de carrés de type III présentent l'avantage essentiel qu'elles ne varient pas avec les fréquences de cellule tant que la forme générale d'estimabilité reste constante. Ce type de somme des carrés est donc souvent considéré comme utile pour les modèles déséquilibrés auxquels aucune cellule ne manque. Dans le plan factoriel sans cellule manquante, cette méthode est équivalente à la technique de Yates des carrés moyens pondérés. La méthode des sommes des carrés de type III sert généralement pour :

- Tous les modèles énumérés dans les types I et II.
- Tous les modèles équilibrés ou non qui ne contiennent pas de cellules vides.

Type IV : Cette méthode est conçue pour une situation dans laquelle il manque des cellules. Pour chaque effet F dans le plan, si F n'est inclus dans aucun autre effet, Type IV = Type III = Type II. Si F est inclus dans d'autres effets, le Type IV distribue les contrastes à effectuer parmi les paramètres dans F sur tous les effets de niveau supérieur de façon équitable. La méthode des sommes des carrés de type IV sert généralement pour :

- Tous les modèles énumérés dans les types I et II.
- Tous les modèles équilibrés ou non qui contiennent des cellules vides.

GLM - Mesures répétées : Contrastes

Les contrastes servent à tester les différences entre les niveaux d'un facteur inter-sujets. Vous pouvez spécifier un contraste pour chaque facteur inter-sujets dans un modèle. Les contrastes représentent des combinaisons linéaires des paramètres.

Le test des hypothèses est fondé sur l'hypothèse nulle $\mathbf{LBM} = 0$, où \mathbf{L} représente la matrice des coefficients des contrastes, \mathbf{B} le vecteur de paramètre et \mathbf{M} la matrice moyenne qui correspond à la transformation moyenne de la variable dépendante. Vous pouvez afficher la matrice de transformation en sélectionnant **Matrice de transformation** dans la boîte de dialogue GLM - Mesures répétées : Options. Par exemple, s'il y a quatre variables dépendantes et qu'on utilise un facteur intra-sujets de quatre niveaux et des contrastes polynomiaux (par défaut) pour les facteurs intra-sujets, la matrice \mathbf{M} sera $(0.5 \ 0.5 \ 0.5 \ 0.5)'$. Si un contraste est spécifié, une matrice \mathbf{L} est créée de telle sorte que les colonnes correspondant au facteur inter-sujets correspondent au contraste. Les colonnes restantes sont ajustées de telle sorte que la matrice \mathbf{L} puisse être estimée.

Les contrastes fournis sont déviation, simple, différence, Helmert, répétée et modèle polynomial. Pour les contrastes d'écart et simple, vous pouvez choisir si la catégorie de référence est la première ou la dernière.

Un contraste autre que **Aucun** doit être sélectionné pour les facteurs intra-sujets.

Types de contraste

Déviatiion : Compare la moyenne de chaque niveau (hormis une catégorie de référence) à la moyenne de tous les niveaux (grande moyenne). Les niveaux du facteur peuvent être de n'importe quel ordre.

Simple : Compare la moyenne de chaque niveau à celle d'un niveau donné. Ce type de contraste est utile lorsqu'il y a un groupe de contrôle. Vous pouvez prendre la première ou la dernière catégorie en référence.

Différence : Compare la moyenne de chaque niveau (hormis le premier) à la moyenne des niveaux précédents. (Parfois appelé contrastes d'Helmert inversé.)

Helmert : Compare la moyenne de chaque niveau de facteur (hormis le dernier) à la moyenne des niveaux suivants.

Répété : Compare la moyenne de chaque niveau (hormis le premier) à la moyenne du niveau suivant.

Polynomial : Compare l'effet linéaire, l'effet quadratique, l'effet cubique etc. Le premier degré de liberté contient l'effet linéaire sur toutes les catégories, le second degré l'effet quadratique, etc. Ces contrastes servent souvent à estimer les tendances polynomiales.

GLM - Mesures répétées : Tracés de profil

Les tracés de profil (tracés d'interaction) sont utiles pour comparer les moyennes marginales dans votre modèle. Un tracé de profil est un tracé en ligne dont chaque point indique la moyenne marginale estimée d'une variable dépendante (ajustée pour les covariables) à un niveau du facteur. Les niveaux d'un second facteur peuvent servir à dessiner des courbes distinctes. Chaque niveau dans un troisième facteur peut servir à créer un tracé distinct. Tous les facteurs sont disponibles pour les tracés. Les tracés de profil sont créés pour chaque variable dépendante. Les facteurs inter-sujets comme intra-sujets peuvent être utilisés dans les tracés de profil.

Un tracé de profil pour un facteur montre si la moyenne marginale estimée est croissante ou décroissante sur les niveaux. Pour au moins deux facteurs, des courbes parallèles indiquent qu'il n'y a pas d'interaction entre les facteurs, ce qui signifie que vous recherchez les niveaux d'un seul facteur. Les courbes non parallèles indiquent une interaction.

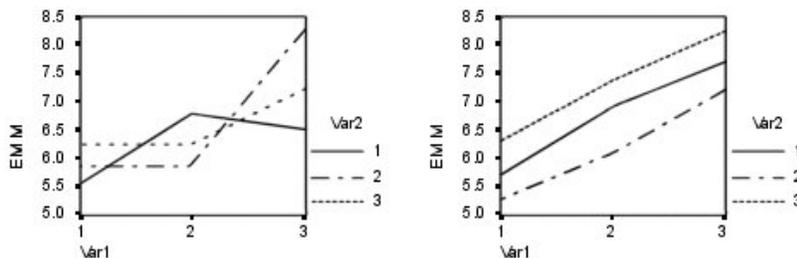


Figure 2. Tracé non parallèle (gauche) et tracé parallèle (droite)

Après avoir sélectionné des facteurs pour l'axe horizontal afin de spécifier un tracé et, éventuellement, des facteurs pour des courbes ou des tracés distincts, vous devez ajouter le tracé à la liste Tracés.

GLM - Mesures répétées : Comparaisons post hoc

Test de comparaison multiple post hoc : Lorsque vous avez déterminé qu'il existe des différences parmi les moyennes, les tests de plages post hoc et de comparaisons multiples appariées peuvent déterminer les moyennes qui diffèrent. Les comparaisons sont effectuées sur des valeurs non-ajustées. Ces tests ne sont pas disponibles s'il n'y a pas de facteurs inter-sujets. Les tests de comparaisons multiples post hoc sont effectués pour la moyenne de tous les niveaux des facteurs intra-sujets.

Les tests de différence significative de Bonferroni et Tukey servent généralement comme tests de comparaison multiples. Le **test de Bonferroni**, fondé sur la statistique t de Student, ajuste le niveau de signification observé en fonction du nombre de comparaisons multiples qui sont effectuées. Le **test t de Sidak** ajuste également le niveau de signification et fournit des limites plus strictes que le test de Bonferroni. Le **test de Tukey** utilise la statistique de plage de Student pour effectuer des comparaisons appariées entre les groupes et fixe le taux d'erreur empirique au taux d'erreur du regroupement de toutes les comparaisons appariées. Lorsque vous testez un grand nombre de paires de moyennes, le test de Tukey est plus efficace que celui de Bonferroni. Lorsqu'il y a peu de paires, Bonferroni est plus efficace.

Le **GT2 de Hochberg** est similaire au test de Tukey mais il utilise un modulo maximum selon Student. Le test de Tukey est généralement plus efficace. Le **test de comparaison appariée de Gabriel** utilise également le modulo maximum selon Student. Il est plus efficace que le GT2 de Hochberg lorsque les tailles des cellules sont inégales. Le test de Gabriel offre plus de souplesse lorsque les tailles des cellules divergent beaucoup.

Le **test de comparaison multiple appariée de Dunnett** compare un ensemble de traitements à une simple moyenne de contrôle. La dernière catégorie est la catégorie de contrôle par défaut. Vous pouvez également choisir la première catégorie. Vous pouvez également choisir un test unilatéral ou bilatéral. Pour tester que la moyenne à un certain niveau (hormis la catégorie de contrôle) du facteur n'est pas égale à celle de la catégorie de contrôle, utilisez le test double-face. Pour tester si la moyenne est inférieure, à un certain niveau du facteur, à celle de la catégorie de contrôle, sélectionnez **< Contrôle**. Pour tester si la moyenne est supérieure, à un certain niveau du facteur, à celle de la catégorie de contrôle, sélectionnez **> Contrôle**.

Ryan, Einot, Gabriel et Welsch (R-E-G-W) ont développé deux tests de plages multiples descendants. Les procédures multiples descendantes testent d'abord que toutes les moyennes sont égales. Si toutes les moyennes ne sont pas égales, l'égalité est testée sur des sous-ensembles de moyennes. Le **F de R-E-G-W** est fondé sur le test *F* et le **Q de R-E-G-W** est fondé sur la plage de Student. Ces tests sont plus efficaces que le test de plages multiples de Duncan et Student-Newman-Keuls (procédures multiples descendantes), mais ils sont conseillés lorsque les cellules sont de taille inégale.

Lorsque les variances sont inégales, utilisez le **T2 de Tamhane** (test de comparaisons appariées conservatif fondé sur un test *t*), le **T3 de Dunnett** (test de comparaison appariée fondé sur le modulo maximal de Student), le **test de comparaison appariée de Games-Howell** (parfois flexible) ou le **C de Dunnett** (test de comparaison appariée fondé sur la plage de Student).

Le **test de plages multiples de Duncan**, Student-Newman-Keuls (**S-N-K**) et le **b de Tukey** sont des tests de plage qui classifient les moyennes de groupe et calculent une valeur de plage. Ces tests ne sont pas utilisés aussi souvent que les tests évoqués précédemment.

Le **test t de Waller-Duncan** utilise une approche de Bayes. Ce test de plage utilise la moyenne harmonique de la taille de l'échantillon lorsque les échantillons sont de tailles différentes.

Le niveau de signification du **test de Scheffé** est conçu pour permettre toutes les combinaisons linéaires possibles des moyennes de groupe à tester, pas seulement appariée, disponibles dans cette fonction. Il en résulte que le test de Scheffé est souvent plus strict que les autres, ce qui signifie qu'une plus grande différence de moyenne est nécessaire pour être significative.

Le test de comparaison multiple appariée de différence la moins significative (**LSD**) est équivalent aux divers tests *t* individuels entre toutes les paires des groupes. L'inconvénient de ce test est qu'il n'essaie pas d'ajuster le niveau d'importance observée pour les comparaisons multiples.

Tests affichés : Les comparaisons appariées sont proposées pour LSD, Sidak, Bonferroni, Games et Howell, T2 et T3 de Tamhane, C et T3 de Dunnett. Des sous-ensembles homogènes pour les tests de plage sont proposés pour S-N-K, *b* de Tukey, Duncan, *F* et *Q* de R-E-G-W et Waller. Le test de Tukey, le GT2 de Hochberg, le test de Gabriel et le test de Scheffé sont à la fois des tests de comparaison multiple et des tests de plage.

GLM - Mesures répétées : Enregistrer

Vous pouvez enregistrer les prévisions par le modèle, les résidus et les mesures associées sous forme de nouvelles variables dans l'éditeur de données. La plupart de ces variables peuvent servir à étudier les hypothèses relatives aux données. Pour enregistrer les valeurs afin de les utiliser dans une autre session IBM SPSS Statistics, vous devez enregistrer le fichier de données en cours.

Prévisions : Valeurs que le modèle estime pour chaque observation.

- *Non standardisés.* Valeur prévue par le modèle pour la variable dépendante.
- *Erreur standard.* Estimation de l'écart type de la valeur moyenne de la variable expliquée pour les unités statistiques qui ont les mêmes valeurs pour les valeurs explicatives.

Diagnostics : Mesures permettant d'identifier les observations avec des combinaisons inhabituelles de valeurs pour les variables indépendantes et les observations qui peuvent avoir un impact important sur le modèle. Vous disposez de la distance de Cook et des valeurs influentes excentrées.

- *Distance 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. Si la distance de Cook est élevée, l'exclusion d'une observation changerait substantiellement la valeur des coefficients.
- *Valeurs influentes.* Valeurs influentes non centrées. Mesure de l'influence d'un point sur l'ajustement de la régression.

Résidus : Un résidu non standardisé correspond à la valeur réelle de la variable dépendante moins la valeur estimée par le modèle. Les résidus standardisés, de Student et supprimés sont également disponibles.

- *Non standardisés.* Différence entre la valeur observée et la valeur prévue par le modèle.
- *Standardisés.* Résidu, divisé par une estimation de son écart type. Également appelés résiduels de Pearson, les résiduels standardisés ont une moyenne de 0 et un écart type de 1.
- *De Student.* Résidu, divisé par une estimation de son écart type, qui varie d'une observation à l'autre, selon la distance entre les valeurs et la moyenne des variables indépendantes pour chaque observation.
- *Supprimées.* Résidu d'une observation lorsque celle-ci est exclue du calcul des coefficients de régression. Il s'agit de la différence entre la valeur de la variable dépendante et la prévision ajustée.

Statistiques à coefficients : Enregistre une matrice variance-covariance des estimations de paramètre dans un jeu de données ou fichier de données. D'autre part, pour chaque variable dépendante, il y aura une ligne d'estimations, une ligne de valeurs de signification pour les statistiques *t* correspondant aux estimations et une ligne de degrés de liberté résiduels. Pour un modèle multivarié, il y a les mêmes lignes pour chaque variable dépendante. Vous pouvez utiliser ces données de matrice dans les autres procédures qui lisent des fichiers de matrice. Les jeux de données sont disponibles pour utilisation ultérieure dans la même session mais ne sont pas enregistrés en tant que fichiers sauf si vous le faites explicitement avant la fin de la session. Le nom des jeux de données doit être conforme aux règles de dénomination de variables.

Options de GLM - Mesures répétées

Des statistiques facultatives sont disponibles à partir de cette boîte de dialogue. Ces statistiques sont calculées à l'aide de modèle à effets fixes.

Moyenne marginale estimée : Sélectionnez les facteurs et les interactions pour lesquels vous souhaitez obtenir des estimations de la moyenne marginale de la population dans les cellules. Ces moyennes sont ajustées pour les covariables, si elles existent. Vous pouvez sélectionner facteurs intra-sujets et inter-sujets.

- **Comparer les effets principaux :** Propose des comparaisons appariées non corrigées des moyennes marginales estimées pour tout effet principal dans le modèle, à la fois pour les facteurs intersujets et intrasujets. Ceci n'est valable que si les effets principaux sont sélectionnés dans la liste Afficher les moyennes.
- **Ajustement intervalle de confiance :** Sélectionnez l'ajustement aux intervalles de confiance et à la significativité des intervalles en adoptant l'une des méthodes suivantes : la différence de moindre signification (LSD), l'ajustement Bonferroni ou l'ajustement de Sidak. Cet élément est disponible uniquement si **Comparer les effets principaux** est sélectionné.

Afficher : Sélectionnez **Statistiques descriptives** pour produire des moyennes, des écarts-types et des effectifs pour toutes les variables dépendantes de toutes les cellules. L'option **Estimation d'effet de taille** fournit une valeur partielle de Eta carré pour chaque effet et chaque estimation. La statistique d'Eta carré décrit la proportion de la variabilité totale imputable au facteur. Sélectionnez **Puissance observée** pour obtenir la puissance du test lorsque l'autre hypothèse est définie sur la base de la valeur observée. Sélectionnez **Estimation des paramètres** pour produire des estimations de paramètres, des erreurs standard, des tests *t*, des intervalles de confiance et la puissance observée de chaque test. Vous pouvez afficher les **matrices SSCP** d'hypothèse et d'erreur, la **matrice SSCP des résidus** et le test de Bartlett de sphéricité de la matrice de covariance des résidus.

L'option des **tests d'homogénéité** produit le test de Levene d'homogénéité de la variance pour chaque variable dépendante sur toutes les combinaisons de niveaux des facteurs inter-sujets, uniquement pour les facteurs inter-sujets. Les tests d'homogénéité incluent également le test *M* de Box d'homogénéité des matrices de covariance des variables dépendantes sur toutes les combinaisons de niveaux des facteurs inter-sujets. Les options des tracés de dispersion par niveau et résiduels sont utiles pour vérifier les hypothèses sur les données. Ceci n'est pas valable s'il n'y a pas de facteurs. Sélectionnez **Tracés résiduels** pour produire un tracé résiduel observé/estimé/standardisé pour chaque variable dépendante. Ces tracés sont utiles pour vérifier l'hypothèse de variance égale. Sélectionnez **Manque d'ajustement** pour vérifier si la relation entre la variable dépendante et les variables indépendantes peut être convenablement décrite par le modèle. **Fonction générale estimée** vous permet de construire des tests d'hypothèses personnalisés basés sur la fonction générale estimée. Les lignes de n'importe quelle matrice des coefficients de contraste sont des combinaisons linéaires de la fonction générale estimée.

Niveau de signification : Vous souhaitez peut-être ajuster le niveau de signification utilisé dans les tests post hoc et le niveau de confiance utilisé pour construire des intervalles de confiance. La valeur spécifiée est également utilisée pour calculer l'intensité observée pour le test. Lorsque vous spécifiez un niveau de signification, le niveau associé des intervalles de confiance est affiché dans la boîte de dialogue.

Fonctions supplémentaires de la commande GLM

Ces fonctions s'appliquent aux analyses univariée, multivariée ou de mesures répétées. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Spécifier les effets imbriqués dans un plan (à l'aide de la sous-commande DESIGN).
- Spécifier les tests d'effets par rapport à une combinaison linéaire d'effets ou une valeur (à l'aide de la sous-commande TEST).
- Spécifier de multiples contrastes (à l'aide de la sous-commande CONTRAST).
- Inclure les valeurs manquantes de l'utilisateur (à l'aide de la sous-commande MISSING).
- Spécifier les critères EPS (à l'aide de la sous-commande CRITERIA).
- Construire une matrice **L** personnalisée, une matrice **M** ou une matrice **K** (à l'aide des sous-commandes LMATRIX, MMATRIX et KMATRIX).
- Pour les contrastes simples ou d'écart, spécifier une catégorie de référence intermédiaire (à l'aide de la sous-commande CONTRAST).
- Spécifier les mesures pour les contrastes polynomiaux (à l'aide de la sous-commande CONTRAST).
- Spécifier des termes d'erreur pour les comparaisons post hoc (à l'aide de la sous-commande POSTHOC).
- Calculer les moyennes marginales estimées pour chaque facteur ou interaction entre facteurs parmi les facteurs de la liste (à l'aide de la sous-commande EMMEANS).
- Attribuer des noms aux variables temporaires (à l'aide de la sous-commande SAVE).
- Construire un fichier de matrice de corrélation (à l'aide de la sous-commande OUTFILE).
- Construire un fichier de type matrice de données qui contient les statistiques provenant de la table ANOVA inter-sujets (à l'aide de la sous-commande OUTFILE).
- Enregistrer la matrice du plan dans un nouveau fichier de données (à l'aide de la sous-commande OUTFILE).

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

Chapitre 4. Analyse des composantes de variance

L'analyse des Composantes de variance, relative aux modèles à effets mixtes, estime la contribution de chaque effet aléatoire sur la variance de la variable dépendante. Cette procédure est particulièrement intéressante pour l'analyse des modèles mixtes, tels que tracé partiel, mesures répétées univariées et plans de blocs aléatoires. Le calcul des composantes de variance vous permet de déterminer les éléments sur lesquels se concentrer pour réduire la variance.

Il existe quatre méthodes différentes d'estimation des composantes de variance : estimation partielle quadratique des normes minimales (MINQUE), analyse de variance, maximum de vraisemblance (ML) et maximum de vraisemblance restreint (REML). Ces différentes méthodes disposent de nombreuses spécifications.

La sortie par défaut pour toutes les méthodes inclut les estimations des composantes de variance. Si la méthode ML ou REML est utilisée, une matrice de covariance asymptotique est également affichée. Les autres sorties possibles comprennent un tableau ANOVA et les carrés moyens attendus pour la méthode ANOVA et un historique des itérations pour les méthodes ML et REML. L'analyse des composantes de variance est totalement compatible avec la procédure GLM - Univarié.

La pondération WLS vous permet de spécifier une variable utilisée pour pondérer les observations pour une analyse pondérée, par exemple pour compenser les variations de précision des mesures.

Exemple : Un collègue agricole mesure, pendant un mois, la prise de poids de cochons élevés par 6 soues différentes. La variable Soue est un facteur aléatoire avec six niveaux. (Les six soues étudiées constituent un échantillon aléatoire d'une population plus vaste de soues.) Le chercheur découvre que la variance dans la prise de poids est plus liée aux différences de soue qu'aux différences entre les cochons dans une soue.

Remarques sur les données des composantes de variance

Données : La variable dépendante est quantitative. Les facteurs sont qualitatifs. Il peut s'agir de valeurs numériques ou alphanumériques de huit octets au maximum. L'un des facteurs au moins doit être aléatoire. Cela signifie que les niveaux du facteur doivent constituer un échantillon aléatoire des niveaux possibles. Les covariables sont des variables quantitatives liées à la variable dépendante.

Hypothèses : Toutes les méthodes partent du principe que les paramètres d'un effet aléatoire ont une moyenne nulle, des variances constantes finies et ne sont pas corrélées deux à deux. Les paramètres de différents effets aléatoires ne sont pas corrélés non plus.

Le terme résiduel a également une moyenne nulle et une variance constante finie. Il n'est pas corrélé avec les paramètres d'un effet aléatoire quelconque. On part du principe que les termes résiduels des différentes observations ne sont pas corrélés.

A partir de ces hypothèses, les observations de même niveau d'un facteur aléatoire sont corrélées. Cela différencie l'analyse des composantes de variance d'un modèle linéaire général.

ANOVA et MINQUE ne nécessitent pas d'hypothèses de normalité. Ces deux méthodes ont la puissance de modérer les abandons de l'hypothèse de normalité.

ML et REML requièrent que le paramètre de modèle et le terme résiduel soient distribués normalement.

Procédures apparentées : Utilisez la procédure Explorer pour étudier les données avant de mener l'analyse des composantes de variance. Pour tester les hypothèses, utilisez GLM – Univarié, GLM - Multivarié et GLM - Mesures répétées.

Obtenir les tableaux des composantes de variance

1. A partir des menus, sélectionnez :

Analyse > Modèle linéaire général > Composantes de variance...

2. Sélectionnez une variable dépendante.

3. Sélectionnez des variables pour Facteurs fixes, Facteurs aléatoires et Covariables en fonction de vos données. Pour spécifier une variable de pondération, utilisez Pondération WLS.

Modèle des composantes de variance

Spécifier le modèle : Un modèle factoriel général contient tous les effets principaux des facteurs, des covariables et toutes les interactions facteur/facteur. Il ne contient pas de d'interactions de covariable. Sélectionnez **Autre** pour indiquer un sous-ensemble d'interactions ou des interactions facteur/covariable. Vous devez indiquer tous les termes à inclure dans le modèle.

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

Modèle : Le modèle dépend de la nature de vos données. Après avoir sélectionné **Autre**, vous pouvez choisir les effets principaux et les interactions qui présentent un intérêt pour votre analyse. Le modèle doit contenir un facteur aléatoire.

Inclure une constante au modèle : L'ordonnée est généralement incluse dans le modèle. Si vous partez du principe que les données passent par l'origine, vous pouvez exclure la constante.

Termes construits

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau de toutes les variables sélectionnées. Il s'agit de la valeur par défaut.

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.

Options des composantes de variance

Méthode : Vous pouvez sélectionner l'une des quatre méthodes permettant d'estimer les composantes de variance.

- **MINQUE** (estimateur impartial quadratique des normes minimales) produit des estimations qui ne varient pas en fonction des effets fixes. Si les données sont distribuées normalement et si les estimations sont correctes, cette méthode produit la variance la plus faible des estimateurs impartiaux. Vous pouvez choisir une méthode pour effet aléatoire avant pondération.

- **ANOVA** (analyse de variance) calcule des estimations impartiales à l'aide des sommes de carrés de type I ou de type III pour chaque effet. La méthode ANOVA produit parfois des estimations de variance négatives, ce qui peut indiquer un modèle faux, une méthode d'estimation inadaptée ou un manque de données.
- Le **maximum de vraisemblance** (ML) produit, à l'aide d'itérations, les estimations les plus cohérentes par rapport aux données effectivement observées. Ces estimations peuvent être faussées. Cette méthode est normale sur le plan asymptotique. Les estimations ML et REML ne varient pas en cas de conversion. Cette méthode ne tient pas compte des degrés de liberté utilisés pour estimer les effets fixés.
- Les estimations du **maximum de vraisemblance restreint** (REML) réduisent les estimations ANOVA pour de nombreuses observations de données équilibrées. Cette méthode étant ajustée pour les effets fixés, ses erreurs standard sont plus petites que par la méthode ML. Cette méthode tient compte des degrés de liberté utilisés pour estimer les effets fixés.

Effet aléatoire a priori : Uniforme implique que tous les effets aléatoires et le terme résiduel ont le même impact sur les observations. Le schéma **Zéro** équivaut à assumer des variances d'effet aléatoire zéro. Cela n'est valable que pour la méthode MINQUE.

Somme des carrés : Les sommes des carrés de **type I** sont utilisées pour le modèle hiérarchique, souvent employé dans la documentation sur les composantes de variance. Si vous sélectionnez le **type III**, valeur par défaut de GLM, les estimations de variance peuvent servir dans la procédure GLM - Univarié pour le test des hypothèses avec les sommes de carrés de type III. Valable uniquement pour la méthode ANOVA.

Critères : Vous pouvez spécifier le critère de convergence et le nombre maximal d'itérations. Cela est disponible uniquement pour les méthodes ML ou REML.

Afficher : En ce qui concerne la méthode ANOVA, vous pouvez choisir d'afficher les sommes des carrés et les moyennes des carrés théoriques. Si vous sélectionnez **Maximum de vraisemblance** ou **Maximum de vraisemblance restreint**, vous pouvez afficher un historique des itérations.

Somme des carrés (Composantes de variance)

Pour ce modèle, vous pouvez choisir un type de sommes des carrés. Le type III est le plus courant et c'est la valeur par défaut.

Type I : Cette méthode est également appelée décomposition hiérarchique de la somme des carrés. Chaque terme est ajusté uniquement pour le terme qui le précède dans le modèle. La méthode de sommes des carrés de type I sert généralement pour :

- Une analyse de la variance équilibrée dans laquelle tout effet principal est spécifié avant les effets d'interaction de premier ordre, et chaque effet de premier ordre spécifié avant ceux de second ordre, et ainsi de suite.
- Un modèle de régression polynomial dans lequel les termes d'ordre inférieur sont spécifiés avant ceux d'ordre supérieur.
- Un modèle par imbrication pur dans lequel le premier effet spécifié est imbriqué dans le second et le second spécifié dans le troisième, etc. (Cette forme d'imbrication peut être spécifiée par la syntaxe uniquement.)

Type III : Valeur par défaut. Cette méthode calcule les sommes des carrés d'un effet dans le plan comme les sommes des carrés ajustée pour tout autre effet qui ne le contient pas et orthogonal à chaque effet qui le contient. Les sommes de carrés de type III présentent l'avantage essentiel qu'elles ne varient pas avec les fréquences de cellule tant que la forme générale d'estimabilité reste constante. Ce type est donc souvent considéré comme utile pour les modèles déséquilibrés auxquels aucune cellule ne manque. Dans le plan factoriel sans cellule manquante, cette méthode est équivalente à la technique de Yates des carrés moyens pondérés. La méthode des sommes des carrés de type III sert généralement pour :

- Les modèles énumérés pour le type I.

- Tous les modèles équilibrés ou non qui ne contiennent aucune cellule vide.

Composantes de variance : Enregistrer dans un nouveau fichier

Vous pouvez enregistrer certains résultats de cette procédure dans un nouveau fichier de données IBM SPSS Statistics.

Estimations des composantes de covariance : Enregistre les estimations des composantes de variance et les libellés d'estimation dans un fichier de données ou un jeu de données. Elles peuvent servir au calcul d'autres statistiques ou pour de plus amples analyses de la procédure GML. Par exemple, vous pouvez les utiliser pour calculer des intervalles de confiance ou tester des hypothèses.

Composants de covariance : Enregistre la matrice variance-covariance ou une matrice de corrélation dans un fichier ou un jeu de données. Valable uniquement si vous avez indiqué **Maximum de vraisemblance** ou **Maximum de vraisemblance restreint**.

Destination des valeurs créées : Vous permet de spécifier le nom du jeu de données ou du fichier externe contenant les estimations et/ou la matrice des composantes de covariance. Les jeux de données sont disponibles pour utilisation ultérieure dans la même session mais ne sont pas enregistrés en tant que fichiers sauf si vous le faites explicitement avant la fin de la session. Le nom des jeux de données doit être conforme aux règles de dénomination de variables.

Vous pouvez utiliser la commande MATRIX pour extraire du fichier de données les données dont vous avez besoin, puis calculer les intervalles de confiance ou effectuer des tests.

Fonctions supplémentaires de la commande VARCOMP

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

- Spécifier les effets imbriqués dans un plan (à l'aide de la sous-commande DESIGN).
- Inclure les valeurs manquantes de l'utilisateur (à l'aide de la sous-commande MISSING).
- Spécifier les critères EPS (à l'aide de la sous-commande CRITERIA).

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

Chapitre 5. Modèles mixtes linéaires

La procédure des modèles mixtes linéaires développe le modèle linéaire général pour permettre aux données de présenter des variabilités en corrélation et des variabilités non constantes. Le modèle linéaire mixte offre donc une flexibilité pour modéliser non seulement les moyennes des données, mais également leurs variances et covariances.

La procédure des modèles mixtes linéaires constitue également un outil flexible pour l'adaptation d'autres modèles pouvant être considérés comme des modèles linéaires mixtes. De tels modèles comportent des modèles à plusieurs niveaux, des modèles linéaires hiérarchiques et des modèles de coefficients aléatoires.

Exemple : Une chaîne d'épicerie souhaiterait connaître les conséquences de divers bons de réduction sur les dépenses des consommateurs. En prenant un échantillon aléatoire des clients réguliers, elle suit les dépenses de chaque client pendant 10 semaines. Chaque semaine, un bon de réduction différent est envoyé aux clients. Modèles mixtes linéaires est utilisé pour évaluer les conséquences de chaque bon de réduction sur les dépenses lors de l'ajustement des corrélations liées aux observations répétées de chaque sujet pendant ces 10 semaines.

Méthodes : Estimation du maximum de vraisemblance (ML) et du maximum de vraisemblance restreint (REML).

Statistiques : Statistiques descriptives : taille des échantillons, moyennes et écarts types de la variable dépendante et des covariables de chaque combinaison distincte des niveaux de facteur. Informations relatives au niveau de facteur : tri des valeurs de niveau de chaque facteur et de leurs fréquences. Les estimations de paramètres et les intervalles de confiance des effets fixes, les tests Wald et les intervalles de confiance des paramètres des matrices de covariance sont également inclus. Les sommes des carrés de type I et II peuvent servir à évaluer différentes hypothèses. Le type III est la valeur par défaut.

Commentaires relatifs aux données des modèles mixtes linéaires

Données : Les variables dépendantes doivent être quantitatives. Les facteurs doivent être qualitatifs, et peuvent comporter des valeurs numériques ou alphanumériques. Les covariables et la variable de pondération doivent être quantitatives. Les variables de sujets et répétées peuvent être d'un type quelconque.

Hypothèses : La variable dépendante est considérée comme étant linéairement reliée aux facteurs fixes, aux facteurs aléatoires et aux covariables. Les effets fixes créent un modèle de la moyenne de la variable dépendante. Les effets aléatoires créent un modèle de la structure de covariance de la variable dépendante. Les effets aléatoires multiples sont considérés comme étant indépendants les uns des autres et les matrices de covariance séparées sont calculées pour chacun des effets ; toutefois, les termes du modèle indiqués sur le même effet aléatoire peuvent être en corrélation. Les mesures répétées créent un modèle de la structure de covariance des résidus. La variable dépendante est également supposée provenir d'une distribution normale.

Procédures apparentées : Utilisez la procédure d'exploration pour examiner les données avant l'exécution d'une analyse. S'il n'existe aucune variabilité en corrélation ou non constante, vous pouvez utiliser la procédure GLM - Univarié ou GLM - Mesures répétées. Vous pouvez également utiliser la procédure d'analyse des composantes de la variance si les effets aléatoires ont une structure de covariance des composantes de variance et n'ont aucune mesure répétée.

Obtention d'une analyse des modèles mixtes linéaires

1. A partir des menus, sélectionnez :

Analyse > Modèles mixtes > Linéaire...

2. Vous pouvez également sélectionner une ou plusieurs variables de sujet.
3. Sélectionnez une ou plusieurs variables répétées, si vous le souhaitez.
4. Sélectionnez une structure de covariance de résidus, si vous le souhaitez.
5. Cliquez sur **Poursuivre**.
6. Sélectionnez une variable dépendante.
7. Sélectionnez au moins un facteur ou une covariable.
8. Cliquez sur **Fixe** ou sur **Aléatoire**, puis indiquez au moins un modèle à effets fixes ou un modèle à effets aléatoires.

Eventuellement, choisissez une variable de pondération.

Modèles mixtes linéaires : Sélectionner des variables répétées/des sujets

Cette boîte de dialogue permet de sélectionner les variables qui définissent les sujets et les observations répétées, et de choisir une structure de covariance pour les résidus.

Sujets : Un sujet est une unité d'observation considérée comme indépendante des autres sujets. Par exemple, lors d'une étude médicale, la mesure de la tension artérielle d'un patient peut être considérée comme étant indépendante des résultats des autres patients. La définition des sujets devient particulièrement importante lorsqu'il existe deux évaluations répétées par sujet et que vous souhaitez modéliser la corrélation entre ces observations. Par exemple, vous pouvez vous attendre à ce que les mesures de tension artérielle d'un patient unique lors de visites consécutives chez le docteur soient en corrélation.

Les sujets peuvent également être définis par la combinaison des niveaux de facteur des variables multiples ; par exemple, vous pouvez indiquer les catégories *Sexe* et *Age* comme variables de sujet pour indiquer que les *hommes de plus de 65 ans* sont identiques les uns aux autres, mais indépendants de la catégorie des *hommes et femmes de moins de 65 ans*.

Toutes les variables spécifiées dans la liste Sujets sont utilisées pour définir les sujets d'une structure de covariance de résidus. Vous pouvez utiliser toutes les variables, ou certaines seulement, pour définir les sujets de la structure de covariance des effets aléatoires.

Répété : Les variables spécifiées dans cette liste sont utilisées pour identifier les observations répétées. Par exemple, une variable unique *Semaine* peut identifier les 10 semaines d'observations dans une étude médicale, ou les variables *Mois* et *Jour* peuvent être utilisées ensemble pour identifier les observations quotidiennes sur une année entière.

Type de covariance répétée : Indique la structure de covariance des résidus. Les structures disponibles sont les suivantes :

- Antédépendance : Premier ordre
- AR(1)
- AR(1) : Hétérogène
- ARMA(1,1)
- Symétrie composée
- Symétrie Composée : Métrique de corrélation
- Symétrie Composée : Hétérogène
- Diagonale
- Analytique Factorielle : Premier ordre

- Analytique Factorielle : Premier Ordre, hétérogène
- Huynh-Feldt
- Identité mise à l'échelle
- Toeplitz
- Toeplitz : Hétérogène
- Sans structure
- Corrélations non structurées

Pour plus d'informations, voir Chapitre 17, «Structures de covariance», à la page 105.

Modèles mixtes linéaires : Effets fixes

Effets fixes : Il n'existe aucun modèle par défaut ; par conséquent, vous devez indiquer les effets fixes de façon explicite. Vous pouvez aussi construire des termes imbriqués ou non imbriqués.

Inclure ordonnée à l'origine : La constante est généralement incluse dans le modèle. Si vous partez du principe que les données passent par l'origine, vous pouvez exclure la constante.

Somme des carrés : Méthode de calcul des sommes des carrés. La méthode de type III est la plus utilisée pour les modèles dont aucune cellule ne manque.

Termes de construction non imbriquée

Pour les facteurs et covariables sélectionnés :

Factoriel : Crée toutes les interactions et les effets principaux possibles pour les variables sélectionnées. Il s'agit de la valeur par défaut.

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.

Construction de termes imbriqués

Dans cette procédure, vous pouvez construire des termes imbriqués pour votre modèle. 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 dépenses de ses clients à divers emplacements de magasin. 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 les effets d'interaction ou ajouter plusieurs niveaux d'imbrication au terme imbriqué.

Limites : 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.

Somme des carrés

Pour ce modèle, vous pouvez choisir un type de sommes des carrés. Le type III est le plus courant et c'est la valeur par défaut.

Type I : Cette méthode est également appelée décomposition hiérarchique de la somme des carrés. Chaque terme est ajusté uniquement pour les termes qui le précèdent dans le modèle. La somme des carrés de type I est généralement utilisée pour :

- Une analyse de la variance équilibrée dans laquelle tout effet principal est spécifié avant les effets d'interaction de premier ordre, et chaque effet de premier ordre spécifié avant ceux de second ordre, et ainsi de suite.
- Un modèle de régression polynomial dans lequel les termes d'ordre inférieur sont spécifiés avant ceux d'ordre supérieur.
- Un modèle par imbrication pur dans lequel le premier effet spécifié est imbriqué dans le second et le second spécifié dans le troisième, etc. (Cette forme d'imbrication peut être spécifiée par la syntaxe uniquement.)

Type III : Valeur par défaut. Cette méthode calcule les sommes des carrés d'un effet dans le plan comme les sommes des carrés ajustée pour tout autre effet qui ne le contient pas et orthogonal à chaque effet qui le contient. Les sommes de carrés de type III présentent l'avantage essentiel qu'elles ne varient pas avec les fréquences de cellule tant que la forme générale d'estimabilité reste constante. Ce type de somme des carrés est donc souvent considéré comme utile pour les modèles déséquilibrés auxquels aucune cellule ne manque. Dans le plan factoriel sans cellule manquante, cette méthode est équivalente à la technique de Yates des carrés moyens pondérés. La méthode des sommes des carrés de type III sert généralement pour :

- Les modèles énumérés pour le type I.
- Tous les modèles équilibrés ou non qui ne contiennent aucune cellule vide.

Modèles mixtes linéaires : Effets aléatoires

Type de covariance : Permet d'indiquer la structure de covariance du modèle à effets aléatoires. Une matrice de covariance séparée est évaluée pour chaque effet aléatoire. Les structures disponibles sont les suivantes :

- Antédépendance : Premier ordre
- AR(1)
- AR(1) : Hétérogène
- ARMA(1,1)
- Symétrie composée
- Symétrie Composée : Métrique de corrélation
- Symétrie Composée : Hétérogène
- Diagonale
- Analytique Factorielle : Premier ordre
- Analytique Factorielle : Premier Ordre, hétérogène
- Huynh-Feldt
- Identité mise à l'échelle
- Toeplitz

- Toeplitz : Hétérogène
- Sans structure
- Non structuré : Métrique de corrélation
- Composantes de variance

Pour plus d'informations, voir Chapitre 17, «Structures de covariance», à la page 105.

Effets aléatoires : Il n'existe aucun modèle par défaut. Vous devez donc indiquer les effets aléatoires de façon explicite. Vous pouvez aussi construire des termes imbriqués ou non imbriqués. Vous pouvez également choisir d'inclure une constante dans le modèle à effets aléatoires.

Vous pouvez indiquer plusieurs modèles à effets aléatoires. Après la construction du premier modèle, cliquez sur **Suivant** pour construire le modèle suivant. Cliquez sur **Précédent** pour accéder aux modèles précédents. Les modèles à effets aléatoires sont considérés comme étant indépendants les uns des autres et les matrices de covariance séparées sont calculées pour chacun d'eux. Les termes indiqués sur le même effet aléatoire peuvent être en corrélation.

Regroupements de sujets : Les variables répertoriées sont celles que vous avez sélectionnées comme variables de sujet dans la boîte de dialogue Sélectionner sujets/variables répétées. Sélectionnez toutes les variables, ou certaines seulement, afin de définir les sujets du modèle à effets aléatoires.

Estimation des modèles mixtes linéaires

Méthode : Sélectionnez les estimations du maximum de vraisemblance et du maximum de vraisemblance restreint.

Itérations : Les options suivantes sont disponibles :

- **Nombre maximum d'itérations :** Spécifiez un nombre entier non négatif.
- **Nombre maximum de dichotomies :** A chaque itération, la taille de l'étape est réduite par un facteur de 0,5 jusqu'à ce que les augmentations de log de vraisemblance ou le nombre maximum de dichotomie soient atteints. Spécifiez un nombre entier positif.
- **Historique des itérations d'impression tous les n étapes :** Affiche un tableau contenant la valeur de la fonction de log de vraisemblance et les estimations de paramètres toutes les n itérations en commençant par l'itération 0 (les estimations initiales). Si vous imprimez l'historique des itérations, la dernière itération est toujours imprimée, indépendamment de la valeur de n .

Convergence de log de vraisemblance : La convergence est prise en compte si la modification absolue ou relative de la fonction log de vraisemblance est inférieure à la valeur spécifiée, qui doit être non négative. Le critère n'est pas utilisé si la valeur spécifiée est égale à 0.

Convergence des paramètres : La convergence est prise en compte si la modification absolue ou relative maximum des estimations de paramètres est inférieure à la valeur spécifiée, qui doit être non négative. Le critère n'est pas utilisé si la valeur spécifiée est égale à 0.

Convergence de Hess : Pour la spécification **Absolu**, la convergence est prise en compte si une statistique basée sur la matrice de Hess est inférieure à la valeur indiquée. Pour la spécification **Relatif**, la convergence est prise en compte si la statistique est inférieure au produit de la valeur indiquée et de la valeur absolue du log de vraisemblance. Le critère n'est pas utilisé si la valeur spécifiée est égale à 0.

Etapes pour évaluation max : Nécessite l'utilisation de l'algorithme des coordonnées de Fisher jusqu'au numéro n . Indiquez un entier non négatif.

Tolérance de singularité : Cette valeur est utilisée comme valeur de tolérance lors du contrôle des singularités. Indiquez une valeur positive.

Modèles mixtes linéaires : Statistiques

Statistiques récapitulatives Génère des tableaux pour :

- **Statistiques descriptives** : Affiche la taille des échantillons, les moyennes et les écarts-types des variables et des covariables dépendantes (le cas échéant). Ces statistiques sont affichées pour chaque combinaison de niveaux distincts des facteurs.
- **Récapitulatif de traitement des observations** : Affiche les valeurs triées des facteurs, les variables de mesure répétée, les sujets de mesure répétée, et les sujets avec des effets aléatoires et leurs fréquences.

Estimation des paramètres : Génère des tableaux pour :

- **Estimations des paramètres** : Affiche les estimations de paramètres à effets fixes et à effets aléatoires, ainsi que leurs erreurs standard approximatives.
- **Tests des paramètres de covariance** : Affiche les erreurs standard asymptotiques et les tests Wald des paramètres de covariance.
- **Corrélations des estimations de paramètres** : Affiche la matrice de corrélation asymptotique des estimations de paramètres à effets fixes.
- **Covariances des estimations de paramètres** : Affiche la matrice de covariance asymptotique des estimations de paramètres à effets fixes.
- **Covariances des effets aléatoires** : Affiche la matrice de covariance estimée des effets aléatoires. Cette option n'est disponible que si un effet aléatoire au minimum est spécifié. Si une variable de sujet est spécifiée pour un effet aléatoire, le bloc commun apparaît.
- **Covariances des résidus** : Affiche la matrice de covariance des résidus estimés. Cette option n'est disponible que lorsqu'une variable répétée est spécifiée. Si une variable de sujet est spécifiée, le bloc commun apparaît.
- **Matrice des coefficients de contraste** : Affiche les fonctions estimées utilisées pour tester les effets fixes et les hypothèses personnalisées.

Intervalle de confiance : Cette valeur est utilisée chaque fois qu'un intervalle de confiance est construit. Spécifiez une valeur supérieure ou égale à 0 et inférieure à 100. La valeur par défaut est 95.

Moyennes EM des modèles mixtes linéaires

Moyennes marginales estimées de modèles ajustés : Ce groupe vous permet de demander les moyennes marginales estimées du modèle de la variable dépendante dans les cellules, ainsi que de leurs erreurs standard pour les facteurs indiqués. Vous pouvez également demander la comparaison des niveaux de facteur des effets principaux.

- **Facteurs et interactions entre facteurs** : Cette liste contient les facteurs et interactions entre facteurs indiqués dans la boîte de dialogue Fixe, ainsi qu'un terme OVERALL. Les termes du modèle construits à partir des covariables sont exclus de cette liste.
- **Afficher moyennes pour** : La procédure calcule les moyennes marginales estimées des interactions entre facteurs et des facteurs sélectionnés dans cette liste. Si vous sélectionnez OVERALL, les moyennes marginales estimées de la variable dépendante apparaissent, fusionnant tous les facteurs. Notez que les facteurs et les interactions entre facteurs restent sélectionnés jusqu'à ce que vous supprimiez une variable associée de la liste des facteurs dans la boîte de dialogue principale.
- **Comparer les effets principaux** : Cette option vous permet de demander des comparaisons appariées des niveaux des effets principaux sélectionnés. La fonction Ajustement intervalle de confiance vous permet d'appliquer un ajustement aux intervalles de confiance et aux valeurs de signification pour rendre compte des comparaisons multiples. Les méthodes disponibles sont les suivantes : LSD (aucun ajustement), Bonferroni et Sidak. Enfin, pour chaque facteur, vous pouvez sélectionner la catégorie de référence en fonction de laquelle les comparaisons sont effectuées. Si vous ne sélectionnez aucune catégorie de référence, toutes les comparaisons appariées sont construites. La catégorie de référence peut être la première, la dernière ou une catégorie personnalisée (auquel cas vous entrez la valeur de la catégorie de référence).

Modèles mixtes linéaires : Enregistrer

Cette boîte de dialogue permet d'enregistrer divers résultats de modèle dans le fichier actuel.

Prévisions fixes : Enregistre les variables liées aux moyennes de régression sans les effets.

- **Prévisions :** Moyennes de régression sans effet aléatoire.
- **Erreurs standard :** Erreurs standard des estimations.
- **Degrés de liberté :** Degrés de liberté associés aux estimations.

Valeurs prévues et résiduelles : Enregistre les variables liées à la valeur ajustée au modèle.

- **Prévisions :** Valeur ajustée au modèle.
- **Erreurs standard :** Erreurs standard des estimations.
- **Degrés de liberté :** Degrés de liberté associés aux estimations.
- **Résidus :** Valeur des données moins la prévision.

Fonctions supplémentaires de la commande MIXED

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

- Spécifier les tests d'effets par rapport à une combinaison linéaire d'effets ou une valeur (à l'aide de la sous-commande TEST).
- Inclure les valeurs manquantes de l'utilisateur (à l'aide de la sous-commande MISSING).
- Calculez les moyennes marginales estimées des valeurs spécifiées pour les covariables (à l'aide du mot-clé WITH de la sous-commande EMMEANS).
- Comparez les effets principaux simples des interactions (à l'aide de la sous-commande EMMEANS).

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

Chapitre 6. Modèles linéaires généralisés

La procédure Modèles linéaires généralisés développe le modèle linéaire général de sorte que la variable dépendante soit linéairement reliée aux facteurs et covariables via une fonction de lien précise. En outre, le modèle permet à la variable dépendante de suivre une distribution non normale. Il couvre des modèles statistiques largement utilisés : la régression linéaire pour les réponses distribuées normalement, les modèles logistiques pour les données binaires, les modèles log-linéaires pour les données d'effectif, les modèles log-log complémentaires pour les données de survie censurées par intervalle. Il concerne également de nombreux autres modèles statistiques grâce à sa formulation de modèle très générale.

Exemples : Une compagnie de navigation peut utiliser des modèles linéaires généralisés pour ajuster une régression de Poisson au nombre de détériorations subies par plusieurs types de bateaux construits à des périodes différentes. Le modèle qui en résulte peut permettre de déterminer quels types de bateaux sont plus enclins aux détériorations.

Une compagnie d'assurance automobile peut utiliser des modèles linéaires généralisés pour ajuster une régression gamma à des actions en indemnisation pour des voitures. Le modèle qui en résulte peut permettre de déterminer les facteurs qui ont le plus d'influence sur le nombre de déclarations.

Des chercheurs en médecine peuvent utiliser des modèles linéaires généralisés pour ajuster une régression log-log complémentaire des données de survie censurées par intervalle pour prévoir la récurrence d'un état médical.

Remarques sur les données des modèles linéaires généralisés

Données : La réponse peut être de type échelle, effectif, binaire ou événements au cours des essais. On suppose que les facteurs sont de type catégorie. On suppose que les covariables, la pondération d'échelle et le décalage sont de type échelle.

Hypothèses : On suppose que les observations sont indépendantes.

Pour obtenir un modèle linéaire généralisé

A partir des menus, sélectionnez :

Analyse > Modèles linéaires généralisés > Modèles linéaires généralisés...

1. Indiquez une distribution et une fonction de lien (consultez ci-dessous les détails concernant les différentes options).
2. Dans l'onglet Réponse, sélectionnez une variable dépendante.
3. Dans l'onglet Prédicteurs, sélectionnez les facteurs et les covariables à utiliser pour prévoir la variable dépendante.
4. Dans l'onglet Modèle, indiquez les effets du modèle à l'aide des facteurs et covariables sélectionnés.

L'onglet Type de modèle permet d'indiquer la fonction de distribution et de lien pour votre modèle en fournissant des raccourcis pour plusieurs modèles communs qui sont classés par type de réponse.

Types de modèle

Réponse d'échelle : Les options suivantes sont disponibles :

- **Linéaire :** Indique la distribution Normale et la fonction de lien Identité.
- **Gamma avec lien log :** Indique la distribution Gamma et la fonction de lien Log.

Réponse ordinale : Les options suivantes sont disponibles :

- **Logistique ordinale** : Indique la distribution Multinomial (ordinal) et la fonction de lien Logit cumulé.
- **Probit ordinal** : Indique la distribution Multinomial (ordinal) et la fonction de lien Probit cumulé.

Effectifs : Les options suivantes sont disponibles :

- **Log-linéaire de Poisson** : Indique la distribution Poisson et la fonction de lien Log.
- **Binomial négatif avec lien log** : Indique la distribution Binomiale négative (avec la valeur 1 pour le paramètre auxiliaire) et la fonction de lien Log. Pour que la procédure soit estimée à la valeur du paramètre auxiliaire, spécifiez un modèle personnalisé avec une distribution binomiale négative et sélectionnez la **Valeur d'estimation** dans le groupe de paramètre.

Réponse binaire ou données événements/tentatives : Les options suivantes sont disponibles :

- **Logistique binaire** : Indique la distribution Binomiale et la fonction de lien Logit.
- **Probit binaire** : Indique la distribution Binomiale et la fonction de lien Probit.
- **Intervalle - Données de survie censurées** : Indique la distribution Binomiale et la fonction de lien Log-log complémentaire.

Mélange : Les options suivantes sont disponibles :

- **Tweedie avec lien log** : Indique la distribution Tweedie et la fonction de lien Log.
- **Tweedie avec lien d'identité** : Indique la distribution Tweedie et la fonction de lien Identité.

Personnalisé : Indique votre propre combinaison de distribution et de fonction de lien.

Distribution

Cette sélection indique la distribution de la variable dépendante. La possibilité de spécifier une distribution non normale et une fonction de lien sans identité constitue la principale amélioration du modèle linéaire généralisé par rapport au modèle linéaire général. Il existe de nombreuses combinaisons distribution/fonction de lien possibles, et plusieurs peuvent convenir à un jeu de données particulier. Votre choix peut être guidé par des considérations théoriques a priori ou en fonction de la combinaison qui vous semble la plus adaptée.

- **Binomiale** : Cette distribution convient uniquement aux variables représentant une réponse binaire ou un nombre d'événements.
- **Gamma** : Cette distribution convient aux variables avec des valeurs d'échelle positives arrondies à des valeurs positives supérieures. Si une valeur de données est inférieure ou égale à 0, ou manquante, l'observation correspondante n'est pas utilisée dans l'analyse.
- **Gaussienne inverse** : Cette distribution convient aux variables avec des valeurs d'échelle positives arrondies à des valeurs positives supérieures. Si une valeur de données est inférieure ou égale à 0, ou manquante, l'observation correspondante n'est pas utilisée dans l'analyse.
- **Binomiale négative** : Cette distribution peut être résumée comme étant le nombre d'essais nécessaires pour observer k succès. Elle convient aux variables comportant des entiers non négatifs. Si une valeur de données n'est pas un entier, est inférieure à 0 ou est manquante, l'observation correspondante n'est pas utilisée dans l'analyse. La valeur du paramètre auxiliaire de la distribution binomiale négative peut être n'importe quel nombre supérieur ou égal à 0. Vous pouvez lui attribuer une valeur fixe ou lui permettre d'être estimé par la procédure. Lorsque le paramètre auxiliaire a pour valeur 0, l'utilisation de cette distribution équivaut à celle de la distribution de Poisson.
- **Normale** : Cette distribution convient aux variables d'échelle dont les valeurs suivent une distribution symétrique, en cloche, autour d'une valeur centrale (moyenne). La variable dépendante doit être numérique.
- **Poisson** : Vous pouvez considérer cette distribution comme étant le nombre d'occurrences d'un événement qui vous intéresse au cours d'une période fixe. Elle convient aux variables comportant des

entiers non négatifs. Si une valeur de données n'est pas un entier, est inférieure à 0 ou est manquante, l'observation correspondante n'est pas utilisée dans l'analyse.

- **Tweedie** : Cette distribution convient aux variables pouvant être représentées par des mélanges de Poisson de distributions gamma. La distribution est "mélangée" dans la mesure où elle combine les propriétés de distributions continue (prend des valeurs réelles non négatives) et discrète (masse de probabilité positive à une valeur unique de 0). La variable dépendante doit être numérique, avec des valeurs de données supérieures ou égales à zéro. Si une valeur de données n'est pas entière, est inférieure à 0 ou est manquante, l'observation correspondante n'est pas utilisée dans l'analyse. La valeur fixe du paramètre de distribution de Tweedie peut être tout nombre supérieur à un et inférieur à deux.
- **Multinomiale** : Cette distribution convient aux variables représentant une réponse ordinale. La variable dépendante peut être numérique ou chaîne, et elle doit comporter au moins deux valeurs de données valides distinctes.

Fonctions de lien

La fonction de lien est une transformation de la variable dépendante qui permet d'estimer le modèle. Les fonctions suivantes sont disponibles :

- **Identité** : $f(x)=x$. La variable dépendante n'est pas transformée. Ce lien peut être utilisé avec toute distribution.
- **Log-log complémentaire** : $f(x)=\log(-\log(1-x))$. Cette fonction convient uniquement à la distribution binomiale.
- **Cauchit cumulé** : $f(x) = \tan(\pi (x - 0.5))$, appliqué à la probabilité cumulée de chaque catégorie de la réponse. Cette fonction convient uniquement à la distribution binomiale.
- **Log-log complémentaire cumulé** : $f(x)=\ln(-\ln(1-x))$, appliqué à la probabilité cumulée de chaque catégorie de la réponse. Cette fonction convient uniquement à la distribution binomiale.
- **Logit cumulé**. $f(x)=\ln(x / (1-x))$, appliqué à la probabilité cumulée de chaque catégorie de la réponse. Cette fonction convient uniquement à la distribution binomiale.
- **Log-log négatif cumulé** : $f(x)=-\ln(-\ln(x))$, appliqué à la probabilité cumulée de chaque catégorie de la réponse. Cette fonction convient uniquement à la distribution binomiale.
- **Probit cumulé**. $f(x)=\Phi^{-1}(x)$, appliqué à la probabilité cumulée de chaque catégorie de la réponse, où Φ^{-1} est la fonction de distribution cumulée normale standard inverse. Cette fonction convient uniquement à la distribution binomiale.
- **Journal** : $f(x)=\log(x)$. Ce lien peut être utilisé avec toute distribution.
- **Log complémentaire** : $f(x)=\log(1-x)$. Cette fonction convient uniquement à la distribution binomiale.
- **Logit** : $f(x)=\log(x / (1-x))$. Cette fonction convient uniquement à la distribution binomiale.
- **Binomiale négative** : $f(x)=\log(x / (x+k^{-1}))$, où k est le paramètre auxiliaire de la distribution binomiale négative. Cette fonction convient uniquement à la distribution binomiale négative.
- **Log-log négatif** : $f(x)=-\log(-\log(x))$. Cette fonction convient uniquement à la distribution binomiale.
- **Puissance du rapport des cotes** : $f(x)=[(x/(1-x))^\alpha - 1]/\alpha$, si $\alpha \neq 0$. $f(x)=\log(x)$, si $\alpha=0$. α représente le nombre requis et doit être un nombre réel. Cette fonction convient uniquement à la distribution binomiale.
- **Probit** : $f(x)=\Phi^{-1}(x)$, où Φ^{-1} représente la fonction de distribution cumulée normale inverse. Cette fonction convient uniquement à la distribution binomiale.
- **Puissance** : $f(x)=x^\alpha$, si $\alpha \neq 0$. $f(x)=\log(x)$, si $\alpha=0$. α représente le nombre requis et doit être un nombre réel. Ce lien peut être utilisé avec toute distribution.

Modèles linéaires généralisés - Réponse

Dans de nombreux cas, vous pouvez simplement indiquer une variable dépendante. Toutefois, les variables n'utilisant que deux valeurs et les réponses qui enregistrent les événements produits au cours d'essais exigent une attention accrue.

- **Réponse binaire** : Lorsque la variable dépendante accepte deux valeurs uniquement, vous pouvez spécifier la catégorie de référence pour l'estimation des paramètres. Une variable de réponse binaire peut être de type chaîne ou numérique.
- **Nombre d'événements se produisant dans un ensemble d'essais** : Lorsque la réponse se compose d'événements se produisant lors de plusieurs essais, la variable dépendante inclut cette valeur et vous pouvez sélectionner une autre variable comportant le nombre d'essais. Si le nombre d'essais est identique dans tous les sujets, vous pouvez également indiquer les essais à l'aide d'une valeur fixe. Le nombre d'essais doit être supérieur ou égal au nombre d'événements pour chaque observation. Les événements doivent être des entiers non négatifs et les essais, des entiers positifs.

Pour les modèles ordinaux multinomiaux, vous pouvez indiquer l'ordre des catégories de la réponse : croissant, décroissant ou des données (l'ordre des données signifie que la première valeur rencontrée dans les données définit la première catégorie et que la dernière valeur rencontrée définit la dernière catégorie).

Pondération d'échelle : Le paramètre d'échelle est un paramètre de modèle estimé lié à la variance de la réponse. Les pondérations d'échelle sont des "valeurs connues", susceptibles de varier d'une observation à l'autre. Si la variable de pondération d'échelle est spécifiée, le paramètre d'échelle, qui est lié à la variance de la réponse, est divisé par cette variable pour chaque observation. Les observations avec des valeurs de pondération d'échelle inférieures ou égales à 0, ou qui sont manquantes, ne sont pas utilisées dans l'analyse.

Modèles linéaires généralisés - Catégorie de référence

Pour la réponse binaire, vous pouvez choisir la catégorie de référence pour la variable dépendante. Ceci peut affecter la sortie, comme les estimations de paramètres et les valeurs enregistrées, mais cela ne doit pas modifier l'ajustement au modèle. Par exemple, si votre réponse binaire prend les valeurs 0 et 1 :

- Par défaut, la procédure crée la dernière catégorie (la valeur la plus élevée), ou 1, la catégorie de référence. Dans cette situation, les probabilités enregistrées par le modèle estiment la chance qu'une observation donnée prenne la valeur 0, et les estimations de paramètres doivent être interprétées comme se rapportant à la probabilité de catégorie 0.
- Si vous indiquez la première catégorie (la valeur la plus basse), ou 0, comme catégorie de référence, les probabilités enregistrées par le modèle estiment la chance qu'une observation donnée prenne la valeur 1.
- Si vous indiquez la catégorie personnalisée et que votre variable a des libellés définis, vous pouvez définir la catégorie de référence en sélectionnant une valeur dans la liste. Ceci peut être pratique, si à mi-chemin de préciser un modèle, vous n'êtes pas sûr de la méthode de codage d'une variable spécifique.

Prédicteurs de modèles linéaires généralisés

L'onglet Prédicteurs vous permet de spécifier les covariables et les facteurs utilisés pour construire des effets de modèle et de spécifier éventuellement un décalage.

Facteurs : Les facteurs sont des prédicteurs catégoriels de type numérique ou chaîne.

Covariables : Les covariables sont des prédicteurs d'échelle ; elles doivent être numériques.

Remarque : Dans le cas d'une réponse binomiale au format binaire, la procédure calcule les statistiques de qualité d'ajustement du khi-deux et de la déviance par sous-population, en fonction de la classification croisée des valeurs observées des covariables et facteurs sélectionnés. Vous devez garder le même groupe de prédicteurs sur plusieurs exécutions de la procédure pour garantir un nombre constant de sous-populations.

Décalage : Le terme de décalage est un prédicteur « structurel ». Son coefficient n'est pas estimé par le modèle mais est supposé être égal à 1. Ainsi, les valeurs du décalage sont simplement ajoutées au prédicteur linéaire de la cible. Ceci est particulièrement utile dans les modèles de régression de Poisson, où chaque observation peut avoir différents niveaux d'exposition à l'événement étudié.

Par exemple, dans le cas de la modélisation du taux d'accident chez les conducteurs, il existe une énorme différence entre un conducteur responsable d'un accident avec 3 ans de permis et un conducteur responsable d'un accident en 25 ans. Le nombre d'accidents peut être modélisé sous la forme d'une réponse de Poisson ou d'une réponse binomiale négative avec un lien log si le log naturel de l'expérience du conducteur est inclus en tant que terme de décalage.

D'autres combinaisons de types de distribution et de lien nécessiteraient d'autres transformations de la variable de décalage.

Modèles linéaires généralisés - Options

Ces options sont appliquées à tous les facteurs indiqués dans l'onglet Prédicteurs.

Valeurs manquantes de l'utilisateur : Les facteurs doivent avoir des valeurs valides pour qu'une observation puisse être incluse dans l'analyse. Ces contrôles vous permettent d'indiquer si les valeurs manquantes de l'utilisateur sont considérées comme valides parmi les facteurs.

Ordre des catégories : Cette option permet de déterminer le dernier niveau d'un facteur, qui peut être associé à un paramètre redondant dans l'algorithme d'estimation. La modification de l'ordre des catégories peut entraîner celle des valeurs des effets de niveau de facteur, car ces estimations des paramètres sont calculées en fonction du "dernier" niveau. Les facteurs peuvent être triés dans l'ordre croissant (de la valeur inférieure à la valeur supérieure), dans l'ordre décroissant (de la valeur supérieure à la valeur inférieure) ou dans l'« ordre des données ». Autrement dit, la première valeur rencontrée dans les données définit la première catégorie et la dernière valeur unique trouvée, la dernière catégorie.

Modèle des modèles linéaires généralisés

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.

Termes non imbriqués

Pour les facteurs et covariables sélectionnés :

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.

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.

Termes imbriqués

Dans cette procédure, vous pouvez construire des termes imbriqués pour votre modèle. 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 à divers emplacements de magasin. 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.

Limites : 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.

Constante : La constante est généralement incluse dans le modèle. Si vous partez du principe que les données passent par l'origine, vous pouvez exclure la constante.

Les modèles avec une distribution ordinale multinomiale n'ont pas une seule constante ; à la place, il y a des paramètres de seuil qui définissent les points de transition entre les catégories adjacentes. Les seuils sont toujours inclus dans le modèle.

Estimation des modèles linéaires généralisés

Estimation des paramètres : Les contrôles de ce groupe vous permettent d'indiquer des méthodes d'estimation et de fournir les valeurs initiales des estimations des paramètres.

- **Méthode** : Vous pouvez sélectionner la méthode d'estimation des paramètres. Vous avez le choix entre la méthode de Newton-Raphson, celle des évaluations de Fisher ou une méthode hybride où des itérations des évaluations de Fisher sont effectuées avant de passer à la méthode de Newton-Raphson. En cas de convergence durant la phase des évaluations de Fisher de la méthode hybride, avant que le nombre maximal d'itérations de Fisher soit atteint, l'algorithme passe à la méthode de Newton-Raphson.
- **Méthode des paramètres d'échelle** : Vous pouvez sélectionner la méthode d'estimation des paramètres d'échelle. Le maximum de vraisemblance estime conjointement le paramètre d'échelle avec les effets du modèle. Cette option n'est pas valide si la réponse suit une distribution binomiale négative, une distribution de Poisson, une distribution binomiale ou une distribution multinomiale. Les options de déviance et du khi-deux de Pearson estiment le paramètre d'échelle à partir de la valeur de ces statistiques. Vous pouvez également spécifier une valeur fixe pour le paramètre d'échelle.
- **Valeurs initiales** : La procédure calcule automatiquement les valeurs initiales des paramètres. Vous pouvez également indiquer les valeurs initiales pour les estimations de paramètres.
- **Matrice de covariance** : L'estimateur basé sur un modèle est le négatif de l'inverse généralisé de la matrice de Hess. L'estimateur-M (ou Huber/White/sandwich) est un estimateur « corrigé » basé sur un modèle qui fournit une estimation cohérente de la covariance, même lorsque les spécifications de la variance et les fonctions de lien sont incorrectes.

Itérations : Les options suivantes sont disponibles :

- **Nombre maximum d'itérations** : Nombre maximal d'itérations exécutées par l'algorithme. Spécifiez un nombre entier non négatif.
- **Nombre maximum de dichotomie** : A chaque itération, la taille de l'étape est réduite par un facteur de 0,5 jusqu'à ce que les augmentations de log de vraisemblance ou le nombre maximum de dichotomie soient atteints. Spécifiez un nombre entier positif.

- **Vérifier la séparation des points de données** : Lorsque cette option est sélectionnée, l'algorithme effectue les tests permettant de s'assurer que les estimations de paramètre contiennent des valeurs uniques. La séparation s'opère lorsque la procédure peut produire un modèle capable de classer correctement chaque observation. Cette option est disponible pour les réponses multinomiales et binomiales au format binaire.

Critères de convergence : Les options suivantes sont disponibles :

- **Convergence des paramètres** : Lorsque cette option est sélectionnée, l'algorithme s'interrompt après une itération dans laquelle la modification relative ou absolue apportée aux estimations de paramètre est inférieure à la valeur spécifiée, qui doit être positive.
- **Convergence de log de vraisemblance** : Lorsque cette option est sélectionnée, l'algorithme s'interrompt après une itération dans laquelle la modification relative ou absolue apportée à la fonction de log de vraisemblance est inférieure à la valeur spécifiée, qui doit être positive.
- **Convergence de Hess** : Pour l'option Absolu, la convergence est prise en compte si une statistique basée sur la convergence de Hess est inférieure à la valeur positive spécifiée. Pour l'option Relatif, la convergence est prise en compte si la statistique est inférieure au produit de la valeur positive spécifiée et à la valeur absolue du log de vraisemblance.

Tolérance de singularité : Les matrices singulières (ou non inversables) comportent des colonnes linéairement dépendantes, ce qui peut provoquer de graves problèmes pour l'algorithme d'estimation. Même les matrices presque singulières peuvent générer des résultats médiocres. C'est pourquoi la procédure traite une matrice dont le déterminant est inférieur à la tolérance en tant que matrice singulière. Indiquez une valeur positive.

Modèles linéaires généralisés - Valeurs initiales

Si les valeurs initiales sont spécifiées, elles doivent être fournies pour tous les paramètres (y compris les paramètres redondants) du modèle. Dans le jeu de données, In the dataset, l'ordre des variables de gauche à droite doit être : *RowType_*, *VarName_*, *P1*, *P2*, etc., où *RowType_* et *VarName_* sont des variables de chaîne et *P1*, *P2*, etc. sont des variables numériques correspondant à une liste ordonnée des paramètres.

- Les valeurs initiales sont fournies sur un enregistrement dont la variable *RowType_* est *EST* ; les valeurs initiales réelles sont données dans les variables *P1*, *P2*, etc. La procédure ignore tous les enregistrements pour lesquels *RowType_* a une valeur différente de *EST*, ainsi que tous les enregistrements au-delà de la première occurrence de *RowType_* égaux à *EST*.
- Les paramètres de constante, si elle est incluse dans le modèle, ou de seuil, si la réponse comporte une distribution multinomiale, doivent être les premières valeurs initiales répertoriées.
- Le paramètre d'échelle et, si la réponse comporte une distribution binomiale négative, le paramètre binomial négatif, doivent être les dernières valeurs initiales indiquées.
- Si l'option Fichier scindé est activée, les variables doivent commencer par la ou les variables du fichier scindé, dans l'ordre indiqué lors de la création du fichier scindé, suivies de *RowType_*, *VarName_*, *P1*, *P2*, etc. comme ci-dessus. Les scissions doivent avoir lieu dans le jeu de données spécifié, dans le même ordre que celui du jeu de données d'origine.

Remarque : Les noms de variables *P1*, *P2*, etc. ne sont pas obligatoires. La procédure accepte tout nom de variable valide pour les paramètres. En effet, le mappage des variables à des paramètres est fonction de la position de la variable et non de son nom. Toute variable au-delà du dernier paramètre est ignorée.

La structure de fichier des valeurs initiales est identique à celle utilisée lors de l'export du modèle en tant que données. Vous pouvez donc utiliser les valeurs finales d'une séquence de la procédure comme entrées d'une séquence suivante.

Statistiques des modèles linéaires généralisés

Effets de modèle : Les options suivantes sont disponibles :

- **Type d'analyse** : Indiquez le type d'analyse à effectuer. L'analyse de type I est généralement appropriée lorsque vous avez des raisons a priori pour classer les prédicteurs dans le modèle, alors que celle de type III s'applique de manière plus générale. Les statistiques de Wald ou du rapport de vraisemblance sont calculées en fonction de la sélection dans le groupe de statistiques du khi-deux.
- **Intervalle de confiance** : Indiquez un niveau de confiance compris entre 50 et 100. Les intervalles de Wald sont basés sur l'hypothèse selon laquelle les paramètres suivent une distribution normale asymptotique. Les intervalles de vraisemblance de profil sont plus précis mais peuvent nécessiter beaucoup de calculs. Le niveau de tolérance pour les intervalles de vraisemblance de profil représente les critères employés pour arrêter l'algorithme itératif utilisé pour calculer les intervalles.
- **Fonction de log de vraisemblance** : Contrôle le format d'affichage de la fonction log de vraisemblance. La fonction complète comprend un terme supplémentaire cohérent par rapport aux estimations des paramètres. Ce terme n'a aucun effet sur l'estimation des paramètres et est exclu de l'affichage dans certains logiciels.

Imprimer : La sortie suivante est disponible :

- **Récapitulatif du traitement des observations** : Affiche le nombre et le pourcentage d'observations incluses dans l'analyse ou exclues de celle-ci, ainsi que la table récapitulative des données corrélées.
- **Statistiques descriptives** : Affiche les statistiques descriptives et les informations récapitulatives sur la variable dépendante, les covariables et les facteurs.
- **Informations sur le modèle** : Affiche le nom du jeu de données, la variable dépendante ou les variables d'événement et d'essai, la variable de décalage, la variable de pondération d'échelle, la distribution de probabilité et la fonction de lien.
- **Qualité d'ajustement** : Affiche la déviance et la déviance mise à l'échelle, le khi-deux de Pearson et le khi-deux de Pearson mis à l'échelle, le log de vraisemblance, le critère d'information d'Akaike (AIC), l'AIC corrigé d'échantillon fini (AICC), le critère d'information bayésien (BIC) et l'AIC cohérent (CAIC).
- **Statistiques du récapitulatif du modèle** : Affiche des tests d'ajustement du modèle, y compris des statistiques du rapport de vraisemblance pour le test composite d'ajustement du modèle, et des statistiques pour les contrastes de type I ou III pour chaque effet.
- **Estimations des 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.
- **Matrices (L) des coefficients de contraste** : Affiche les coefficients de contraste pour les effets par défaut et pour les moyennes marginales estimées, si l'onglet Moyennes EM en a besoin.
- **Fonctions estimées générales** : Affiche les matrices pour générer les matrices L des coefficients de contraste.
- **Historique des itérations** : Affiche l'historique des itérations pour les estimations des paramètres et le log de vraisemblance, et imprime la dernière évaluation du vecteur de gradient et la matrice de Hess. Le tableau de l'historique des itérations affiche les estimations de paramètres de toutes les $n^{\text{ièmes}}$ itérations, en commençant par l'itération 0 (les estimations initiales), n étant la valeur de l'intervalle d'impression. Si l'historique des itérations est requis, la dernière itération est toujours affichée quel que soit n .
- **Test du multiplicateur de Lagrange** : Affiche les statistiques du test du multiplicateur de Lagrange pour évaluer la validité d'un paramètre d'échelle calculé en utilisant la déviance ou le khi-deux de Pearson, ou défini à un nombre fixe, pour les distributions normale, gamma, gaussienne inverse et Tweedie. Pour la distribution binomiale négative, cette option teste le paramètre auxiliaire fixe.

Moyennes EM des modèles linéaires généralisés

Cet onglet vous permet d'afficher les moyennes marginales estimées des niveaux et interactions entre facteurs. Vous pouvez également demander l'affichage de la moyenne estimée globale. Les moyennes marginales estimées ne sont pas disponibles pour les modèles ordinaux multinomiaux.

Facteurs et interactions : Cette liste contient les facteurs et interactions entre facteurs spécifiés respectivement dans les onglets Prédicteurs et Modèle. Les covariables sont exclues de cette liste. Vous pouvez sélectionner des termes directement dans cette liste ou en combiner plusieurs dans un terme d'interaction à l'aide du bouton **Par ***.

Afficher les moyennes pour : Les moyennes estimées sont calculées pour les interactions entre facteurs et les facteurs sélectionnés. Le contraste détermine le mode de définition des tests d'hypothèse pour la comparaison des moyennes estimées. Le contraste simple nécessite une catégorie de référence ou un niveau de facteur servant de base de comparaison avec les autres.

- **Seulement composantes non valides :** Les comparaisons appariées sont calculées pour les combinaisons de tous les niveaux des facteurs indiqués ou implicites. C'est le seul contraste disponible pour les interactions entre facteurs.
- *Simple.* Compare la moyenne de chaque niveau à celle d'un niveau donné. Ce type de contraste est utile lorsqu'il y a un groupe de contrôle.
- **Déviatiion :** Chaque niveau du facteur est comparé à la moyenne générale. Les contrastes de déviation ne sont pas orthogonaux.
- *Différence.* Compare la moyenne de chaque niveau (hormis le premier) à la moyenne des niveaux précédents. (Parfois appelé contraste de Helmert inversé.)
- *Helmert.* Compare la moyenne de chaque niveau de facteur (hormis le dernier) à la moyenne des niveaux suivants.
- *Répété.* Compare la moyenne de chaque niveau (hormis le premier) à la moyenne du niveau suivant.
- *Polynomial.* Compare l'effet linéaire, l'effet quadratique, l'effet cubique etc. Le premier degré de liberté contient l'effet linéaire sur toutes les catégories, le second degré l'effet quadratique, etc. Ces contrastes servent souvent à estimer les tendances polynomiales.

Echelle : Les moyennes marginales estimées peuvent être calculées pour la réponse, selon l'échelle d'origine de la variable dépendante, ou pour le prédicteur linéaire, selon la variable dépendante transformée par la fonction de lien.

Ajustement pour les comparaisons multiples : Lors de l'exécution de tests d'hypothèse avec plusieurs contrastes, vous pouvez ajuster le niveau global de signification à partir des niveaux de signification des contrastes inclus. Ce groupe vous permet de choisir la méthode d'ajustement.

- **Différence la moins significative :** Cette méthode ne contrôle pas l'intégralité de la probabilité de rejet des hypothèses qui présentent des contrastes linéaires différents des valeurs d'hypothèse nulles.
- *Bonferroni.* Cette méthode ajuste le niveau de signification observé au fait que plusieurs contrastes sont en cours de test.
- *Bonferroni séquentiel.* Il s'agit d'une procédure descendante de rejet séquentiel de Bonferroni beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Sidak.* Cette méthode fournit des limites plus strictes que l'approche de Bonferroni.
- *Sidak séquentiel.* Il s'agit d'une procédure descendante de rejet séquentiel de Sidak beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.

Enregistrement des modèles linéaires généralisés

Les éléments sélectionnés sont enregistrés avec le nom spécifié ; vous pouvez choisir de remplacer les variables existantes ayant le même nom que les nouvelles variables, ou d'éviter les conflits en ajoutant des suffixes pour que les nouveaux noms de variables soient uniques.

- **Prévision de la moyenne de la réponse** : Enregistre les valeurs prévues par le modèle pour chaque observation dans la métrique de réponse d'origine. Lorsque la distribution de la réponse est binomiale et que la variable dépendante est binaire, la procédure enregistre les probabilités prévues. Lorsque la distribution de la réponse est multinomiale, le libellé d'élément devient **Cumulative predicted probability**, et la procédure enregistre la probabilité prévue cumulée pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.
- **Borne inférieure de l'intervalle de confiance pour la moyenne de la réponse** : Enregistre les bornes supérieure et inférieure de l'intervalle de confiance de la moyenne de la réponse. Lorsque la distribution de la réponse est multinomiale, le libellé d'élément devient **Lower bound of confidence interval for cumulative predicted probability**, et la procédure enregistre la borne inférieure pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.
- **Borne supérieure de l'intervalle de confiance pour la moyenne de la réponse** : Enregistre les bornes supérieure et inférieure de l'intervalle de confiance de la moyenne de la réponse. Lorsque la distribution de la réponse est multinomiale, le libellé d'élément devient **Upper bound of confidence interval for cumulative predicted probability**, et la procédure enregistre la borne supérieure pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.
- **Catégorie estimée** : Pour les modèles avec une distribution binomiale et une variable dépendante binaire, ou une distribution multinomiale, cette option enregistre la catégorie de réponse prévue pour chaque observation. Cette option n'est pas disponible pour les autres distributions de la réponse.
- **Prévision du prédicteur linéaire** : Enregistre les valeurs prévues par le modèle pour chaque observation dans la métrique du prédicteur linéaire (réponse transformée via la fonction de lien spécifiée). Lorsque la distribution de la réponse est multinomiale, la procédure enregistre la valeur prévue pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.
- **Erreur standard estimée de la prévision du prédicteur linéaire** : Lorsque la distribution de la réponse est multinomiale, la procédure enregistre l'estimation de l'erreur standard pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.

Les éléments suivants ne sont pas disponibles lorsque la distribution de la réponse est multinomiale.

- *Distance 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. Si la distance de Cook est élevée, l'exclusion d'une observation changerait substantiellement la valeur des coefficients.
- *Valeur influente*. Mesures de l'influence d'un point sur l'ajustement de la régression. La valeur influente centrée varie de 0 (aucune influence sur la qualité de l'ajustement) à $(N-1)/N$.
- *Résidu brut*. Différence entre la valeur observée et la valeur prévue par le modèle.
- **Résiduel de Pearson** : Racine carrée de la contribution d'une observation à la statistique du khi-deux de Pearson, avec le signe du résidu brut.
- **Résiduel de Pearson standardisé** : Résiduel de Pearson multiplié par la racine carrée de l'inverse du produit du paramètre d'échelle et valeur influente 1 pour l'observation.
- **Résidus au sens déviance** : Racine carrée de la contribution d'une observation à la statistique de déviance, avec le signe du résidu brut.
- **Résidu au sens déviance standardisé** : Résidu au sens déviance multiplié par la racine carrée de l'inverse du produit du paramètre d'échelle et valeur influente 1 pour l'observation.

- **Résidu de vraisemblance** : Racine carrée d'une moyenne pondérée (basée sur la valeur influente de l'observation) des carrés des résidus de Pearson et au sens déviance standardisés, avec le signe du résidu brut.

Export de modèles linéaires généralisés

Exporter le modèle en tant que données : Écrit un jeu de données dans un format IBM SPSS Statistics contenant la corrélation des paramètres ou la matrice de covariance avec les estimations des paramètres, les erreurs standard, les valeurs de signification et les degrés de liberté. L'ordre des variables dans le fichier de matrice est le suivant.

- **Variables de scission** : En cas d'utilisation, toutes les variables définissant des scissions.
- **RowType_** : Prend les valeurs (et libellés de valeurs) suivantes *COV* (covariances), *CORR* (corrélations), *EST* (estimations de paramètres), *SE* (erreurs standard), *SIG* (niveau de signification) et *DF* (degrés de liberté du plan d'échantillonnage). Il existe une observation distincte avec le type de ligne *COV* (ou *CORR*) pour chaque paramètre de modèle et une observation distincte pour chacun des autres types de ligne.
- **VarName_** : Prend les valeurs *P1*, *P2*, ..., correspondant à une liste ordonnée de tous les paramètres de modèle estimés (hormis les paramètres d'échelle et les paramètres binomial négatif) pour les types de ligne *COV* ou *CORR*, avec des libellés de valeur correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres. Les cellules sont vides pour les autres types de ligne.
- **P1, P2, ..** : Ces variables correspondent à une liste triée de toutes les paramètres de modèle (y compris les paramètres binomial négatif et d'échelle, si nécessaire) avec des libellés de variable correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres, et prennent des valeurs en fonction du type de ligne.

Pour les paramètres redondants, toutes les covariances et les estimations de paramètres sont définies sur zéro, et l'ensemble des corrélations, erreurs standard, niveaux de signification et degrés de liberté résiduels sont définis sur la valeur système manquante.

Pour le paramètre d'échelle, les covariances, les corrélations, le niveau de signification et les degrés de liberté sont définis sur la valeur système manquante. Si le paramètre d'échelle est estimé par le maximum de vraisemblance, l'erreur standard est donnée. Dans le cas contraire, il est défini sur la valeur système manquante.

Pour le paramètre binomial négatif, les covariances, les corrélations, le niveau de signification et les degrés de liberté sont définis sur la valeur système manquante. Si le paramètre binomial négatif est estimé par le maximum de vraisemblance, l'erreur standard est donnée. Dans le cas contraire, il est défini sur la valeur système manquante.

S'il existe des scissions, la liste de paramètres doit être cumulée pour toutes les scissions. Dans une scission donnée, certains paramètres peuvent ne pas être significatifs, ce qui est différent de redondant. Pour les paramètres non significatifs, l'ensemble des covariances ou corrélations, estimations de paramètres, erreurs standard, niveaux de signification et degrés de liberté sont définis sur la valeur système manquante.

Vous pouvez utiliser ce fichier de matrice comme valeur par défaut pour des estimations de modèle à venir. Notez que ce fichier n'est pas immédiatement utilisable pour d'autres analyses dans d'autres procédures que la lecture d'un fichier de matrice, sauf si ces procédures acceptent tous les types de ligne exportés ici. Même si c'est le cas, vous devez faire attention à ce que tous les paramètres de ce fichier de matrice aient la même signification pour la procédure lisant le fichier.

Exporter le modèle au format XML : Enregistre les estimations et la matrice de covariance des paramètres, si vous l'avez sélectionnée, 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 GENLIN

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

- Spécifier les valeurs initiales des estimations de paramètres sous forme de liste de nombres (à l'aide de la sous-commande CRITERIA).
- Donner aux covariables des valeurs autres que leur moyenne lors du calcul de la moyenne marginale estimée (à l'aide de la sous-commande EMMEANS).
- Spécifier des contrastes polynomiaux personnalisés pour les moyennes marginales estimées (à l'aide de la sous-commande EMMEANS).
- Spécifier un sous-ensemble de facteurs pour lesquels les moyennes marginales estimées sont affichées pour comparaison à l'aide du type de contraste spécifié (via les mots-clés TABLES et COMPARE de la sous-commande EMMEANS).

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

Chapitre 7. Equations d'estimation généralisées

La procédure Equations d'estimation généralisées étend le modèle linéaire généralisé pour permettre l'analyse des mesures répétées ou d'autres observations corrélées, par exemple des données juxtaposées.

Exemple : Les responsables de la santé publique peuvent utiliser des équations d'estimation généralisées pour ajuster une régression logistique à mesures répétées afin d'étudier les effets de la pollution de l'air sur la santé des enfants.

Remarques sur les données des équations d'estimation généralisées

Données : La réponse peut être de type échelle, effectif, binaire ou événements au cours des essais. On suppose que les facteurs sont de type catégorie. On suppose que les covariables, la pondération d'échelle et le décalage sont de type échelle. Les variables utilisées pour définir les sujets ou les mesures répétées intra-sujets ne permettent pas de définir la réponse mais peuvent avoir d'autres rôles dans le modèle.

Hypothèses : Les observations sont supposées être dépendantes à l'intérieur des sujets et indépendantes entre les sujets. La matrice de corrélation représentant les dépendances intra-sujets est estimée dans le cadre du modèle.

Obtention des équations d'estimation généralisées

A partir des menus, sélectionnez :

Analyse > Modèles linéaires généralisés > Equations d'estimation généralisées...

1. Sélectionnez une ou plusieurs variables sujets (reportez-vous aux options supplémentaires ci-dessous).
La combinaison des valeurs des variables spécifiées doit définir de manière unique les **sujets** à l'intérieur du fichier de données. Par exemple, une seule variable *ID du patient* suffit à définir les sujets dans un hôpital unique, mais la combinaison des variables *ID de l'hôpital* et *ID du patient* peut être nécessaire si les numéros d'identification des patients ne sont pas uniques entre les hôpitaux.
Dans le cas de mesures répétées, plusieurs observations sont enregistrées pour chaque sujet, de sorte que chaque sujet peut occuper plusieurs observations dans le jeu de données.
2. Dans l'onglet Type de modèle, indiquez une distribution et une fonction de lien.
3. Dans l'onglet Réponse, sélectionnez une variable dépendante.
4. Dans l'onglet Prédicteurs, sélectionnez les facteurs et les covariables à utiliser pour prévoir la variable dépendante.
5. Dans l'onglet Modèle, indiquez les effets du modèle à l'aide des facteurs et covariables sélectionnés.

Dans l'onglet Mesures répétées, vous pouvez éventuellement indiquer :

Variables intra-sujets : La combinaison des valeurs des variables intra-sujets définit l'ordre des mesures à l'intérieur des sujets. Ainsi, la combinaison des variables sujets et intra-sujets définit de façon unique chaque mesure. Par exemple, la combinaison des variables *Période*, *ID de l'hôpital* et *ID du patient* définit pour chaque observation une consultation particulière pour un patient donné dans un hôpital déterminé.

Si le jeu de données est déjà trié de sorte que les mesures répétées pour chaque sujet se produisent dans un bloc d'observations contiguës et dans l'ordre correct, il n'est pas nécessaire de spécifier une variable intra-sujets. Vous pouvez désélectionner l'option **Trier les observations par variables sujets et intra-sujets** et économiser le temps de traitement nécessaire au tri (temporaire). Il est généralement judicieux d'utiliser des variables intra-sujets pour garantir l'ordre correct des mesures.

Les variables sujets et intra-sujets ne peuvent pas être utilisées pour définir la réponse, mais elles peuvent avoir d'autres fonctions dans le modèle. Par exemple, la variable *ID de l'hôpital* peut être utilisée comme facteur dans le modèle.

Matrice de covariance : L'estimateur basé sur un modèle est le négatif de l'inverse généralisé de la matrice de Hess. L'estimateur-M (ou Huber/White/sandwich) est un estimateur « corrigé » basé sur un modèle qui fournit une estimation cohérente de la covariance, même lorsque les spécifications de la matrice de corrélation de travail sont incorrectes. Cette spécification s'applique aux paramètres de la partie modèle linéaire des équations d'estimation généralisées, tandis que la spécification de l'onglet Estimation s'applique uniquement au modèle linéaire généralisé initial.

Matrice de corrélation de travail : Cette matrice de corrélation représente les dépendances intra-sujets. Sa taille est déterminée par le nombre de mesures et donc par la combinaison des valeurs des variables intra-sujets. Vous pouvez choisir l'une des structures suivantes :

- **Indépendante :** Les mesures répétées ne sont pas corrélées.
- **AR(1) :** Les mesures répétées ont une relation auto-régressive d'ordre un. La corrélation entre deux éléments est égale à ρ pour les éléments adjacents, à ρ^2 pour les éléments séparés par un troisième, etc. La valeur est restreinte pour être comprise entre $-1 < \rho < 1$.
- **Echangeable :** Les corrélations entre les éléments de cette structure sont homogènes. Elle est également appelée symétrie composée.
- **Dépendante de M :** Les mesures consécutives ont un coefficient de corrélation commun, les paires de mesures séparées par une troisième ont un coefficient de corrélation commun, etc., jusqu'aux paires de mesures séparées par $m-1$ autres mesures. Supposons que vous fournissiez chaque année à vos étudiants des tests standardisés pour les niveaux 3 à 7. Cette structure suppose que les résultats des grades 3 et 4, 4 et 5, 5 et 6, et 6 et 7 posséderont la même corrélation ; que ceux des grades 3 et 5, 4 et 6, et 5 et 7 posséderont la même corrélation et que ceux des grades 3 et 6, et 4 et 7 posséderont la même corrélation. On part du principe que les mesures dont la séparation est supérieure à m ne sont pas corrélées. Lorsque vous choisissez cette structure, indiquez une valeur de m inférieure à l'ordre de la matrice de corrélation de travail.
- **Sans structure :** Il s'agit d'une matrice de corrélation générale complète.

Par défaut, la procédure ajuste les estimations de corrélation en fonction du nombre de paramètres non redondants. Il peut être souhaitable de supprimer cet ajustement si vous voulez que les estimations ne varient pas en fonction des modifications de la réplification de niveau sujet dans les données.

- **Nombre maximum d'itérations :** Nombre maximal d'itérations exécutées par l'algorithme des équations d'estimation généralisées. Spécifiez un nombre entier non négatif. Cette spécification s'applique aux paramètres de la partie modèle linéaire des équations d'estimation généralisées, tandis que la spécification de l'onglet Estimation s'applique uniquement au modèle linéaire généralisé initial.
- **Mettre à jour la matrice :** Les éléments de la matrice de corrélation de travail sont estimés en fonction des estimations de paramètres, lesquels sont mis à jour à chaque itération de l'algorithme. Si la matrice de corrélation de travail n'est jamais mise à jour, la matrice initiale est utilisée tout au long de l'estimation. Si la matrice est mise à jour, vous pouvez spécifier l'intervalle d'itération auquel la mise à jour des éléments de la matrice de corrélation de travail doit s'effectuer. Spécifier une valeur supérieure à 1 peut réduire le temps de traitement.

Critères de convergence : Ces spécifications s'appliquent aux paramètres de la partie modèle linéaire des équations d'estimation généralisées, tandis que la spécification de l'onglet Estimation s'applique uniquement au modèle linéaire généralisé initial.

- **Convergence des paramètres :** Lorsque cette option est sélectionnée, l'algorithme s'interrompt après une itération dans laquelle la modification relative ou absolue apportée aux estimations de paramètre est inférieure à la valeur spécifiée, qui doit être positive.
- **Convergence de Hess :** La convergence est prise en compte si une statistique basée sur la matrice de Hess est inférieure à la valeur spécifiée, qui doit être positive.

Equations d'estimation généralisées : type de modèle

L'onglet Type de modèle permet d'indiquer la fonction de distribution et de lien pour votre modèle en fournissant des raccourcis pour plusieurs modèles communs qui sont classés par type de réponse.

Types de modèle

Réponse d'échelle : Les options suivantes sont disponibles :

- **Linéaire** : Indique la distribution Normale et la fonction de lien Identité.
- **Gamma avec lien log** : Indique la distribution Gamma et la fonction de lien Log.

Réponse ordinale : Les options suivantes sont disponibles :

- **Logistique ordinale** : Indique la distribution Multinomial (ordinal) et la fonction de lien Logit cumulé.
- **Probit ordinal** : Indique la distribution Multinomial (ordinal) et la fonction de lien Probit cumulé.

Effectifs : Les options suivantes sont disponibles :

- **Log-linéaire de Poisson** : Indique la distribution Poisson et la fonction de lien Log.
- **Binomial négatif avec lien log** : Indique la distribution Binomiale négative (avec la valeur 1 pour le paramètre auxiliaire) et la fonction de lien Log. Pour que la procédure soit estimée à la valeur du paramètre auxiliaire, spécifiez un modèle personnalisé avec une distribution binomiale négative et sélectionnez la **Valeur d'estimation** dans le groupe de paramètre.

Réponse binaire ou données événements/tentatives : Les options suivantes sont disponibles :

- **Logistique binaire** : Indique la distribution Binomiale et la fonction de lien Logit.
- **Probit binaire** : Indique la distribution Binomiale et la fonction de lien Probit.
- **Intervalle - Données de survie censurées** : Indique la distribution Binomiale et la fonction de lien Log-log complémentaire.

Mélange : Les options suivantes sont disponibles :

- **Tweedie avec lien log** : Indique la distribution Tweedie et la fonction de lien Log.
- **Tweedie avec lien d'identité** : Indique la distribution Tweedie et la fonction de lien Identité.

Personnalisé : Indique votre propre combinaison de distribution et de fonction de lien.

Distribution

Cette sélection indique la distribution de la variable dépendante. La possibilité de spécifier une distribution non normale et une fonction de lien sans identité constitue la principale amélioration du modèle linéaire généralisé par rapport au modèle linéaire général. Il existe de nombreuses combinaisons distribution/fonction de lien possibles, et plusieurs peuvent convenir à un jeu de données particulier. Votre choix peut être guidé par des considérations théoriques a priori ou en fonction de la combinaison qui vous semble la plus adaptée.

- **Binomiale** : Cette distribution convient uniquement aux variables représentant une réponse binaire ou un nombre d'événements.
- **Gamma** : Cette distribution convient aux variables avec des valeurs d'échelle positives arrondies à des valeurs positives supérieures. Si une valeur de données est inférieure ou égale à 0, ou manquante, l'observation correspondante n'est pas utilisée dans l'analyse.
- **Gaussienne inverse** : Cette distribution convient aux variables avec des valeurs d'échelle positives arrondies à des valeurs positives supérieures. Si une valeur de données est inférieure ou égale à 0, ou manquante, l'observation correspondante n'est pas utilisée dans l'analyse.
- **Binomiale négative** : Cette distribution peut être résumée comme étant le nombre d'essais nécessaires pour observer k succès. Elle convient aux variables comportant des entiers non négatifs. Si une valeur

de données n'est pas un entier, est inférieure à 0 ou est manquante, l'observation correspondante n'est pas utilisée dans l'analyse. La valeur du paramètre auxiliaire de la distribution binomiale négative peut être n'importe quel nombre supérieur ou égal à 0. Vous pouvez lui attribuer une valeur fixe ou lui permettre d'être estimé par la procédure. Lorsque le paramètre auxiliaire a pour valeur 0, l'utilisation de cette distribution équivaut à celle de la distribution de Poisson.

- **Normale** : Cette distribution convient aux variables d'échelle dont les valeurs suivent une distribution symétrique, en cloche, autour d'une valeur centrale (moyenne). La variable dépendante doit être numérique.
- **Poisson** : Vous pouvez considérer cette distribution comme étant le nombre d'occurrences d'un événement qui vous intéresse au cours d'une période fixe. Elle convient aux variables comportant des entiers non négatifs. Si une valeur de données n'est pas un entier, est inférieure à 0 ou est manquante, l'observation correspondante n'est pas utilisée dans l'analyse.
- **Tweedie** : Cette distribution convient aux variables pouvant être représentées par des mélanges de Poisson de distributions gamma. La distribution est "mélangée" dans la mesure où elle combine les propriétés de distributions continue (prend des valeurs réelles non négatives) et discrète (masse de probabilité positive à une valeur unique de 0). La variable dépendante doit être numérique, avec des valeurs de données supérieures ou égales à zéro. Si une valeur de données n'est pas entière, est inférieure à 0 ou est manquante, l'observation correspondante n'est pas utilisée dans l'analyse. La valeur fixe du paramètre de distribution de Tweedie peut être tout nombre supérieur à un et inférieur à deux.
- **Multinomiale** : Cette distribution convient aux variables représentant une réponse ordinale. La variable dépendante peut être numérique ou chaîne, et elle doit comporter au moins deux valeurs de données valides distinctes.

Fonction de lien

La fonction de lien est une transformation de la variable dépendante qui permet d'estimer le modèle. Les fonctions suivantes sont disponibles :

- **Identité** : $f(x)=x$. La variable dépendante n'est pas transformée. Ce lien peut être utilisé avec toute distribution.
- **Log-log complémentaire** : $f(x)=\log(-\log(1-x))$. Cette fonction convient uniquement à la distribution binomiale.
- **Cauchit cumulé** : $f(x) = \tan(\pi (x - 0.5))$, appliqué à la probabilité cumulée de chaque catégorie de la réponse. Cette fonction convient uniquement à la distribution binomiale.
- **Log-log complémentaire cumulé** : $f(x)=\ln(-\ln(1-x))$, appliqué à la probabilité cumulée de chaque catégorie de la réponse. Cette fonction convient uniquement à la distribution binomiale.
- **Logit cumulé**. $f(x)=\ln(x / (1-x))$, appliqué à la probabilité cumulée de chaque catégorie de la réponse. Cette fonction convient uniquement à la distribution binomiale.
- **Log-log négatif cumulé** : $f(x)=-\ln(-\ln(x))$, appliqué à la probabilité cumulée de chaque catégorie de la réponse. Cette fonction convient uniquement à la distribution binomiale.
- **Probit cumulé**. $f(x)=\Phi^{-1}(x)$, appliqué à la probabilité cumulée de chaque catégorie de la réponse, où Φ^{-1} est la fonction de distribution cumulée normale standard inverse. Cette fonction convient uniquement à la distribution binomiale.
- **Journal** : $f(x)=\log(x)$. Ce lien peut être utilisé avec toute distribution.
- **Log complémentaire** : $f(x)=\log(1-x)$. Cette fonction convient uniquement à la distribution binomiale.
- **Logit** : $f(x)=\log(x / (1-x))$. Cette fonction convient uniquement à la distribution binomiale.
- **Binomiale négative** : $f(x)=\log(x / (x+k^{-1}))$, où k est le paramètre auxiliaire de la distribution binomiale négative. Cette fonction convient uniquement à la distribution binomiale négative.
- **Log-log négatif** : $f(x)=-\log(-\log(x))$. Cette fonction convient uniquement à la distribution binomiale.
- **Puissance du rapport des cotes** : $f(x)=[(x/(1-x))^\alpha-1]/\alpha$, si $\alpha \neq 0$. $f(x)=\log(x)$, si $\alpha=0$. α représente le nombre requis et doit être un nombre réel. Cette fonction convient uniquement à la distribution binomiale.

- **Probit** : $f(x)=\Phi^{-1}(x)$, où Φ^{-1} représente la fonction de distribution cumulée normale inverse. Cette fonction convient uniquement à la distribution binomiale.
- **Puissance** : $f(x)=x^\alpha$, si $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. α représente le nombre requis et doit être un nombre réel. Ce lien peut être utilisé avec toute distribution.

Réponse des équations d'estimation généralisées

Dans de nombreux cas, vous pouvez simplement indiquer une variable dépendante. Toutefois, les variables n'utilisant que deux valeurs et les réponses qui enregistrent les événements produits au cours d'essais exigent une attention accrue.

- **Réponse binaire** : Lorsque la variable dépendante accepte deux valeurs uniquement, vous pouvez spécifier la catégorie de référence pour l'estimation des paramètres. Une variable de réponse binaire peut être de type chaîne ou numérique.
- **Nombre d'événements se produisant dans un ensemble d'essais** : Lorsque la réponse se compose d'événements se produisant lors de plusieurs essais, la variable dépendante inclut cette valeur et vous pouvez sélectionner une autre variable comportant le nombre d'essais. Si le nombre d'essais est identique dans tous les sujets, vous pouvez également indiquer les essais à l'aide d'une valeur fixe. Le nombre d'essais doit être supérieur ou égal au nombre d'événements pour chaque observation. Les événements doivent être des entiers non négatifs et les essais, des entiers positifs.

Pour les modèles ordinaux multinomiaux, vous pouvez indiquer l'ordre des catégories de la réponse : croissant, décroissant ou des données (l'ordre des données signifie que la première valeur rencontrée dans les données définit la première catégorie et que la dernière valeur rencontrée définit la dernière catégorie).

Pondération d'échelle : Le paramètre d'échelle est un paramètre de modèle estimé lié à la variance de la réponse. Les pondérations d'échelle sont des "valeurs connues", susceptibles de varier d'une observation à l'autre. Si la variable de pondération d'échelle est spécifiée, le paramètre d'échelle, qui est lié à la variance de la réponse, est divisé par cette variable pour chaque observation. Les observations avec des valeurs de pondération d'échelle inférieures ou égales à 0, ou qui sont manquantes, ne sont pas utilisées dans l'analyse.

Catégorie de référence des équations d'estimation généralisées

Pour la réponse binaire, vous pouvez choisir la catégorie de référence pour la variable dépendante. Ceci peut affecter la sortie, comme les estimations de paramètres et les valeurs enregistrées, mais cela ne doit pas modifier l'ajustement au modèle. Par exemple, si votre réponse binaire prend les valeurs 0 et 1 :

- Par défaut, la procédure crée la dernière catégorie (la valeur la plus élevée), ou 1, la catégorie de référence. Dans cette situation, les probabilités enregistrées par le modèle estiment la chance qu'une observation donnée prenne la valeur 0, et les estimations de paramètres doivent être interprétées comme se rapportant à la probabilité de catégorie 0.
- Si vous indiquez la première catégorie (la valeur la plus basse), ou 0, comme catégorie de référence, les probabilités enregistrées par le modèle estiment la chance qu'une observation donnée prenne la valeur 1.
- Si vous indiquez la catégorie personnalisée et que votre variable a des libellés définis, vous pouvez définir la catégorie de référence en sélectionnant une valeur dans la liste. Ceci peut être pratique, si à mi-chemin de préciser un modèle, vous n'êtes pas sûr de la méthode de codage d'une variable spécifique.

Prédicteurs des équations d'estimation généralisées

L'onglet Prédicteurs vous permet de spécifier les covariables et les facteurs utilisés pour construire des effets de modèle et de spécifier éventuellement un décalage.

Facteurs : Les facteurs sont des prédicteurs catégoriels de type numérique ou chaîne.

Covariables : Les covariables sont des prédicteurs d'échelle ; elles doivent être numériques.

Remarque : Dans le cas d'une réponse binomiale au format binaire, la procédure calcule les statistiques de qualité d'ajustement du khi-deux et de la déviance par sous-population, en fonction de la classification croisée des valeurs observées des covariables et facteurs sélectionnés. Vous devez garder le même groupe de prédicteurs sur plusieurs exécutions de la procédure pour garantir un nombre constant de sous-populations.

Décalage : Le terme de décalage est un prédicteur « structurel ». Son coefficient n'est pas estimé par le modèle mais est supposé être égal à 1. Ainsi, les valeurs du décalage sont simplement ajoutées au prédicteur linéaire de la cible. Ceci est particulièrement utile dans les modèles de régression de Poisson, où chaque observation peut avoir différents niveaux d'exposition à l'événement étudié.

Par exemple, dans le cas de la modélisation du taux d'accident chez les conducteurs, il existe une énorme différence entre un conducteur responsable d'un accident avec 3 ans de permis et un conducteur responsable d'un accident en 25 ans. Le nombre d'accidents peut être modélisé sous la forme d'une réponse de Poisson ou d'une réponse binomiale négative avec un lien log si le log naturel de l'expérience du conducteur est inclus en tant que terme de décalage.

D'autres combinaisons de types de distribution et de lien nécessiteraient d'autres transformations de la variable de décalage.

Options des équations d'estimation généralisées

Ces options sont appliquées à tous les facteurs indiqués dans l'onglet Prédicteurs.

Valeurs manquantes de l'utilisateur : Les facteurs doivent avoir des valeurs valides pour qu'une observation puisse être incluse dans l'analyse. Ces contrôles vous permettent d'indiquer si les valeurs manquantes de l'utilisateur sont considérées comme valides parmi les facteurs.

Ordre des catégories : Cette option permet de déterminer le dernier niveau d'un facteur, qui peut être associé à un paramètre redondant dans l'algorithme d'estimation. La modification de l'ordre des catégories peut entraîner celle des valeurs des effets de niveau de facteur, car ces estimations des paramètres sont calculées en fonction du "dernier" niveau. Les facteurs peuvent être triés dans l'ordre croissant (de la valeur inférieure à la valeur supérieure), dans l'ordre décroissant (de la valeur supérieure à la valeur inférieure) ou dans l'« ordre des données ». Autrement dit, la première valeur rencontrée dans les données définit la première catégorie et la dernière valeur unique trouvée, la dernière catégorie.

Modèle des équations d'estimation généralisées

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.

Termes non imbriqués

Pour les facteurs et covariables sélectionnés :

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.

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.

Termes imbriqués

Dans cette procédure, vous pouvez construire des termes imbriqués pour votre modèle. 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 à divers emplacements de magasin. 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 les effets d'interaction ou ajouter plusieurs niveaux d'imbrication au terme imbriqué.

Limites : 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.

Constante : La constante est généralement incluse dans le modèle. Si vous partez du principe que les données passent par l'origine, vous pouvez exclure la constante.

Les modèles avec une distribution ordinaire multinomiale n'ont pas une seule constante ; à la place, il y a des paramètres de seuil qui définissent les points de transition entre les catégories adjacentes. Les seuils sont toujours inclus dans le modèle.

Estimation des équations d'estimation généralisées

Estimation des paramètres : Les contrôles de ce groupe vous permettent d'indiquer des méthodes d'estimation et de fournir les valeurs initiales des estimations des paramètres.

- **Méthode** : Vous pouvez sélectionner une méthode d'estimation de paramètre. Vous avez le choix entre la méthode de Newton-Raphson, les évaluations de Fisher ou une méthode hybride dans laquelle les itérations des évaluations de Fisher sont effectuées avant le passage à la méthode de Newton-Raphson. En cas de convergence durant la phase des évaluations de Fisher de la méthode hybride, avant que le nombre maximal d'itérations de Fisher soit atteint, l'algorithme passe à la méthode de Newton-Raphson.
- **Méthode des paramètres d'échelle** : Vous pouvez sélectionner la méthode d'estimation des paramètres d'échelle.

Le maximum de vraisemblance estime conjointement le paramètre d'échelle avec les effets du modèle. Cette option n'est pas valide si la réponse suit une distribution binomiale négative, une distribution de Poisson ou une distribution binomiale. Puisque le concept de vraisemblance ne s'applique pas aux

équations d'estimation généralisées, cette spécification ne concerne que le modèle linéaire généralisé initial. Cette estimation du paramètre d'échelle est ensuite transmise aux équations d'estimation généralisées, qui mettent à jour le paramètre d'échelle selon le khi-deux de Pearson divisé par ses degrés de liberté.

Les options de la déviance et du khi-deux de Pearson estiment le paramètre d'échelle à partir de la valeur de ces statistiques dans le modèle linéaire généralisé initial. Cette estimation du paramètre d'échelle est ensuite transmise aux équations d'estimation généralisées, qui la traitent comme étant fixe.

Vous pouvez également spécifier une valeur fixe pour le paramètre d'échelle. Elle sera traitée comme étant fixe lors de l'estimation du modèle linéaire généralisé initial et des équations d'estimation généralisées.

- **Valeurs initiales** : La procédure calcule automatiquement les valeurs initiales des paramètres. Vous pouvez également indiquer les valeurs initiales pour les estimations de paramètres.

Les itérations et les critères de convergence spécifiés dans cet onglet ne sont applicables qu'au modèle linéaire généralisé initial. Pour plus d'informations sur les critères d'estimation utilisés pour ajuster les équations d'estimation généralisées, voir l'onglet Répété.

Itérations : Les options suivantes sont disponibles :

- **Nombre maximum d'itérations** : Nombre maximal d'itérations exécutées par l'algorithme. Spécifiez un nombre entier non négatif.
- **Nombre maximum de dichotomie** : A chaque itération, la taille de l'étape est réduite par un facteur de 0,5 jusqu'à ce que les augmentations de log de vraisemblance ou le nombre maximum de dichotomie soient atteints. Spécifiez un nombre entier positif.
- **Vérifier la séparation des points de données** : Lorsque cette option est sélectionnée, l'algorithme effectue les tests permettant de s'assurer que les estimations de paramètre contiennent des valeurs uniques. La séparation s'opère lorsque la procédure peut produire un modèle capable de classer correctement chaque observation. Cette option est disponible pour les réponses multinomiales et binomiales au format binaire.

Critères de convergence : Les options suivantes sont disponibles :

- **Convergence des paramètres** : Lorsque cette option est sélectionnée, l'algorithme s'interrompt après une itération dans laquelle la modification relative ou absolue apportée aux estimations de paramètre est inférieure à la valeur spécifiée, qui doit être positive.
- **Convergence de log de vraisemblance** : Lorsque cette option est sélectionnée, l'algorithme s'interrompt après une itération dans laquelle la modification relative ou absolue apportée à la fonction de log de vraisemblance est inférieure à la valeur spécifiée, qui doit être positive.
- **Convergence de Hess** : Pour l'option Absolu, la convergence est prise en compte si une statistique basée sur la convergence de Hess est inférieure à la valeur positive spécifiée. Pour l'option Relatif, la convergence est prise en compte si la statistique est inférieure au produit de la valeur positive spécifiée et à la valeur absolue du log de vraisemblance.

Tolérance de singularité : Les matrices singulières (ou non inversables) comportent des colonnes linéairement dépendantes, ce qui peut provoquer de graves problèmes pour l'algorithme d'estimation. Même les matrices presque singulières peuvent générer des résultats médiocres. C'est pourquoi la procédure traite une matrice dont le déterminant est inférieur à la tolérance en tant que matrice singulière. Indiquez une valeur positive.

Valeurs initiales des équations d'estimation généralisées

La procédure estime un modèle linéaire généralisé initial, puis les estimations issues de ce modèle sont utilisées comme valeurs initiales des estimations de paramètres dans la partie modèle linéaire des équations d'estimation généralisées. Les valeurs initiales ne sont pas nécessaires pour la matrice de corrélation de travail car les éléments de la matrice sont basés sur les estimations de paramètres. Les

valeurs initiales spécifiées dans cette boîte de dialogue sont utilisées comme point de départ du modèle linéaire généralisé initial, et non des équations d'estimation généralisées, sauf si le nombre maximal d'itérations dans l'onglet Estimation est défini sur 0.

Si les valeurs initiales sont spécifiées, elles doivent être fournies pour tous les paramètres (y compris les paramètres redondants) du modèle. Dans le jeu de données, In the dataset, l'ordre des variables de gauche à droite doit être : *RowType_*, *VarName_*, *P1*, *P2*, etc., où *RowType_* et *VarName_* sont des variables de chaîne et *P1*, *P2*, etc. sont des variables numériques correspondant à une liste ordonnée des paramètres.

- Les valeurs initiales sont fournies sur un enregistrement dont la variable *RowType_* est *EST* ; les valeurs initiales réelles sont données dans les variables *P1*, *P2*, etc. La procédure ignore tous les enregistrements pour lesquels *RowType_* a une valeur différente de *EST*, ainsi que tous les enregistrements au-delà de la première occurrence de *RowType_* égaux à *EST*.
- Les paramètres de constante, si elle est incluse dans le modèle, ou de seuil, si la réponse comporte une distribution multinomiale, doivent être les premières valeurs initiales répertoriées.
- Le paramètre d'échelle et, si la réponse comporte une distribution binomiale négative, le paramètre binomial négatif, doivent être les dernières valeurs initiales indiquées.
- Si l'option Fichier scindé est activée, les variables doivent commencer par la ou les variables du fichier scindé, dans l'ordre indiqué lors de la création du fichier scindé, suivies de *RowType_*, *VarName_*, *P1*, *P2*, etc. comme ci-dessus. Les scissions doivent avoir lieu dans le jeu de données spécifié, dans le même ordre que celui du jeu de données d'origine.

Remarque : Les noms de variables *P1*, *P2*, etc. ne sont pas obligatoires. La procédure accepte tout nom de variable valide pour les paramètres. En effet, le mappage des variables à des paramètres est fonction de la position de la variable et non de son nom. Toute variable au-delà du dernier paramètre est ignorée.

La structure de fichier des valeurs initiales est identique à celle utilisée lors de l'export du modèle en tant que données. Vous pouvez donc utiliser les valeurs finales d'une séquence de la procédure comme entrées d'une séquence suivante.

Statistiques des équations d'estimation généralisées

Effets de modèle : Les options suivantes sont disponibles :

- **Type d'analyse** : Indiquez le type d'analyse à effectuer pour tester les effets du modèle. L'analyse de type I est généralement appropriée lorsque vous avez des raisons a priori pour classer les prédicteurs dans le modèle, alors que celle de type III s'applique de manière plus générale. Les statistiques de score généralisé ou de Wald sont calculées en fonction de la sélection dans le groupe Statistiques du khi-deux.
- **Intervalle de confiance** : Indiquez un niveau de confiance supérieur à 50 et inférieur à 100. Les intervalles de Wald sont toujours obtenus quel que soit le type de statistique du khi-deux sélectionné, et sont basés sur l'hypothèse selon laquelle les paramètres suivent une distribution normale asymptotique.
- **Fonction de log-quasi-vraisemblance** : Contrôle le format d'affichage de la fonction correspondante. La fonction complète comprend un terme supplémentaire cohérent par rapport aux estimations des paramètres. Ce terme n'a aucun effet sur l'estimation des paramètres et est exclu de l'affichage dans certains logiciels.

Imprimer : La sortie suivante est disponible.

- **Récapitulatif du traitement des observations** : Affiche le nombre et le pourcentage d'observations incluses dans l'analyse ou exclues de celle-ci, ainsi que la table récapitulative des données corrélées.
- **Statistiques descriptives** : Affiche les statistiques descriptives et les informations récapitulatives sur la variable dépendante, les covariables et les facteurs.

- **Informations sur le modèle** : Affiche le nom du jeu de données, la variable dépendante ou les variables d'événement et d'essai, la variable de décalage, la variable de pondération d'échelle, la distribution de probabilité et la fonction de lien.
- **Qualité d'ajustement** : Affiche deux extensions du critère d'information d'Akaike pour la sélection du modèle : quasi-vraisemblance sous un critère de modèle d'indépendance (QIC) pour choisir la meilleure structure de corrélation et une autre mesure QIC pour choisir le meilleur sous-ensemble de prédicteurs.
- **Statistiques du récapitulatif du modèle** : Affiche des tests d'ajustement du modèle, y compris des statistiques du rapport de vraisemblance pour le test composite d'ajustement du modèle, et des statistiques pour les contrastes de type I ou III pour chaque effet.
- **Estimations des 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.
- **Matrices (L) des coefficients de contraste** : Affiche les coefficients de contraste pour les effets par défaut et pour les moyennes marginales estimées, si l'onglet Moyennes EM en a besoin.
- **Fonctions estimées générales** : Affiche les matrices pour générer les matrices L des coefficients de contraste.
- **Historique des itérations** : Affiche l'historique des itérations pour les estimations des paramètres et le log de vraisemblance, et imprime la dernière évaluation du vecteur de gradient et la matrice de Hess. Le tableau de l'historique des itérations affiche les estimations de paramètres de toutes les n ièmes itérations, en commençant par l'itération 0 (les estimations initiales), n étant la valeur de l'intervalle d'impression. Si l'historique des itérations est requis, la dernière itération est toujours affichée quel que soit n .
- **Matrice de corrélation de travail** : Affiche les valeurs de la matrice représentant les dépendances intra-sujets. Sa structure dépend des spécifications dans l'onglet Répété.

Moyennes EM des équations d'estimation généralisées

Cet onglet vous permet d'afficher les moyennes marginales estimées des niveaux et interactions entre facteurs. Vous pouvez également demander l'affichage de la moyenne estimée globale. Les moyennes marginales estimées ne sont pas disponibles pour les modèles ordinaux multinomiaux.

Facteurs et interactions : Cette liste contient les facteurs et interactions entre facteurs spécifiés respectivement dans les onglets Prédicteurs et Modèle. Les covariables sont exclues de cette liste. Vous pouvez sélectionner des termes directement dans cette liste ou en combiner plusieurs dans un terme d'interaction à l'aide du bouton **Par ***.

Afficher les moyennes pour : Les moyennes estimées sont calculées pour les interactions entre facteurs et les facteurs sélectionnés. Le contraste détermine le mode de définition des tests d'hypothèse pour la comparaison des moyennes estimées. Le contraste simple nécessite une catégorie de référence ou un niveau de facteur servant de base de comparaison avec les autres.

- **Seulement composantes non valides** : Les comparaisons appariées sont calculées pour les combinaisons de tous les niveaux des facteurs indiqués ou implicites. C'est le seul contraste disponible pour les interactions entre facteurs.
- *Simple*. Compare la moyenne de chaque niveau à celle d'un niveau donné. Ce type de contraste est utile lorsqu'il y a un groupe de contrôle.
- **Déviation** : Chaque niveau du facteur est comparé à la moyenne générale. Les contrastes de déviation ne sont pas orthogonaux.

- *Différence*. Compare la moyenne de chaque niveau (hormis le premier) à la moyenne des niveaux précédents. (Parfois appelé contraste de Helmert inversé.)
- *Helmert*. Compare la moyenne de chaque niveau de facteur (hormis le dernier) à la moyenne des niveaux suivants.
- *Répété*. Compare la moyenne de chaque niveau (hormis le premier) à la moyenne du niveau suivant.
- *Polynomial*. Compare l'effet linéaire, l'effet quadratique, l'effet cubique etc. Le premier degré de liberté contient l'effet linéaire sur toutes les catégories, le second degré l'effet quadratique, etc. Ces contrastes servent souvent à estimer les tendances polynomiales.

Echelle : Les moyennes marginales estimées peuvent être calculées pour la réponse, selon l'échelle d'origine de la variable dépendante, ou pour le prédicteur linéaire, selon la variable dépendante transformée par la fonction de lien.

Ajustement pour les comparaisons multiples : Lors de l'exécution de tests d'hypothèse avec plusieurs contrastes, vous pouvez ajuster le niveau global de signification à partir des niveaux de signification des contrastes inclus. Ce groupe vous permet de choisir la méthode d'ajustement.

- **Différence la moins significative** : Cette méthode ne contrôle pas l'intégralité de la probabilité de rejet des hypothèses qui présentent des contrastes linéaires différents des valeurs d'hypothèse nulles.
- *Bonferroni*. Cette méthode ajuste le niveau de signification observé au fait que plusieurs contrastes sont en cours de test.
- *Bonferroni séquentiel*. Il s'agit d'une procédure descendante de rejet séquentiel de Bonferroni beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Sidak*. Cette méthode fournit des limites plus strictes que l'approche de Bonferroni.
- *Sidak séquentiel*. Il s'agit d'une procédure descendante de rejet séquentiel de Sidak beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.

Enregistrement des équations d'estimation généralisées

Les éléments sélectionnés sont enregistrés avec le nom spécifié ; vous pouvez choisir de remplacer les variables existantes ayant le même nom que les nouvelles variables, ou d'éviter les conflits en ajoutant des suffixes pour que les nouveaux noms de variables soient uniques.

- **Prévision de la moyenne de la réponse** : Enregistre les valeurs prévues par le modèle pour chaque observation dans la métrique de réponse d'origine. Lorsque la distribution de la réponse est binomiale et que la variable dépendante est binaire, la procédure enregistre les probabilités prévues. Lorsque la distribution de la réponse est multinomiale, le libellé d'élément devient **Probabilité prévue cumulée**, et la procédure enregistre la probabilité prévue cumulée pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.
- **Borne inférieure de l'intervalle de confiance pour la moyenne de la réponse** : Enregistre les bornes supérieure et inférieure de l'intervalle de confiance de la moyenne de la réponse. Lorsque la distribution de la réponse est multinomiale, le libellé d'élément devient **Borne inférieure de l'intervalle de confiance pour la probabilité prévue cumulée**, et la procédure enregistre la borne inférieure pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.
- **Borne supérieure de l'intervalle de confiance pour la moyenne de la réponse** : Enregistre les bornes supérieure et inférieure de l'intervalle de confiance de la moyenne de la réponse. Lorsque la distribution de la réponse est multinomiale, le libellé d'élément devient **Borne supérieure de l'intervalle de confiance pour la probabilité prévue cumulée**, et la procédure enregistre la borne supérieure pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.

- **Catégorie estimée** : Pour les modèles avec une distribution binomiale et une variable dépendante binaire, ou une distribution multinomiale, cette option enregistre la catégorie de réponse prévue pour chaque observation. Cette option n'est pas disponible pour les autres distributions de la réponse.
- **Prévision du prédicteur linéaire** : Enregistre les valeurs prévues par le modèle pour chaque observation dans la métrique du prédicteur linéaire (réponse transformée via la fonction de lien spécifiée). Lorsque la distribution de la réponse est multinomiale, la procédure enregistre la valeur prévue pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.
- **Erreur standard estimée de la prévision du prédicteur linéaire** : Lorsque la distribution de la réponse est multinomiale, la procédure enregistre l'estimation de l'erreur standard pour chaque catégorie de la réponse, hormis la dernière, jusqu'au nombre de catégories indiquées à enregistrer.

Les éléments suivants ne sont pas disponibles lorsque la distribution de la réponse est multinomiale.

- *Résidu brut*. Différence entre la valeur observée et la valeur prévue par le modèle.
- **Résiduel de Pearson** : Racine carrée de la contribution d'une observation à la statistique du khi-deux de Pearson, avec le signe du résidu brut.

Export d'équations d'estimation généralisées

Exporter le modèle en tant que données : Écrit un jeu de données dans un format IBM SPSS Statistics contenant la corrélation des paramètres ou la matrice de covariance avec les estimations des paramètres, les erreurs standard, les valeurs de signification et les degrés de liberté. L'ordre des variables dans le fichier de matrice est le suivant.

- **Variables de scission** : En cas d'utilisation, toutes les variables définissant des scissions.
- **RowType_** : Prend les valeurs (et libellés de valeurs) suivantes *COV* (covariances), *CORR* (corrélations), *EST* (estimations de paramètres), *SE* (erreurs standard), *SIG* (niveau de signification) et *DF* (degrés de liberté du plan d'échantillonnage). Il existe une observation distincte avec le type de ligne *COV* (ou *CORR*) pour chaque paramètre de modèle et une observation distincte pour chacun des autres types de ligne.
- **VarName_** : Prend les valeurs *P1*, *P2*, ..., correspondant à une liste ordonnée de tous les paramètres de modèle estimés (hormis les paramètres d'échelle et les paramètres binomial négatif) pour les types de ligne *COV* ou *CORR*, avec des libellés de valeur correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres. Les cellules sont vides pour les autres types de ligne.
- **P1, P2, ..** : Ces variables correspondent à une liste triée de toutes les paramètres de modèle (y compris les paramètres binomial négatif et d'échelle, si nécessaire) avec des libellés de variable correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres, et prennent des valeurs en fonction du type de ligne.

Pour les paramètres redondants, toutes les covariances et les estimations de paramètres sont définies sur zéro, et l'ensemble des corrélations, erreurs standard, niveaux de signification et degrés de liberté résiduels sont définis sur la valeur système manquante.

Pour le paramètre d'échelle, les covariances, les corrélations, le niveau de signification et les degrés de liberté sont définis sur la valeur système manquante. Si le paramètre d'échelle est estimé par le maximum de vraisemblance, l'erreur standard est donnée. Dans le cas contraire, il est défini sur la valeur système manquante.

Pour le paramètre binomial négatif, les covariances, les corrélations, le niveau de signification et les degrés de liberté sont définis sur la valeur système manquante. Si le paramètre binomial négatif est estimé par le maximum de vraisemblance, l'erreur standard est donnée. Dans le cas contraire, il est défini sur la valeur système manquante.

S'il existe des scissions, la liste de paramètres doit être cumulée pour toutes les scissions. Dans une scission donnée, certains paramètres peuvent ne pas être significatifs, ce qui est différent de redondant. Pour les paramètres non significatifs, l'ensemble des covariances ou corrélations, estimations de paramètres, erreurs standard, niveaux de signification et degrés de liberté sont définis sur la valeur système manquante.

Vous pouvez utiliser ce fichier de matrice comme valeur par défaut pour des estimations de modèle à venir. Notez que ce fichier n'est pas immédiatement utilisable pour d'autres analyses dans d'autres procédures que la lecture d'un fichier de matrice, sauf si ces procédures acceptent tous les types de ligne exportés ici. Même si c'est le cas, vous devez faire attention à ce que tous les paramètres de ce fichier de matrice aient la même signification pour la procédure lisant le fichier.

Exporter le modèle au format XML : Enregistre les estimations et la matrice de covariance des paramètres, si vous l'avez sélectionnée, 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 GENLIN

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

- Spécifier les valeurs initiales des estimations de paramètres sous forme de liste de nombres (à l'aide de la sous-commande CRITERIA).
- Spécifier une matrice de corrélation de travail fixe (à l'aide de la sous-commande REPEATED).
- Donner aux covariables des valeurs autres que leur moyenne lors du calcul de la moyenne marginale estimée (à l'aide de la sous-commande EMMEANS).
- Spécifier des contrastes polynomiaux personnalisés pour les moyennes marginales estimées (à l'aide de la sous-commande EMMEANS).
- Spécifier un sous-ensemble de facteurs pour lesquels les moyennes marginales estimées sont affichées pour comparaison à l'aide du type de contraste spécifié (via les mots-clés TABLES et COMPARE de la sous-commande EMMEANS).

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

Chapitre 8. Modèles linéaires mixtes généralisés

Les modèles linéaires mixtes généralisés étendent le modèle linéaire de sorte que :

- La cible est linéairement reliée aux facteurs et aux covariables via une fonction de lien spécifiée.
- La cible peut avoir une distribution non normale.
- Les observations peuvent être corrélées.

Les modèles linéaires mixtes généralisés couvrent une grande variété de modèles, depuis des modèles de régression linéaire simples jusqu'à des modèles multiniveaux complexes pour des données longitudinales non normales.

Exemples : La commission scolaire du secteur peut utiliser un modèle mixte linéaire généralisé pour déterminer si une méthode d'apprentissage expérimentale est efficace pour l'amélioration des notes en mathématiques. Les élèves d'une même classe devraient être corrélés puisque le même enseignant leur dispense les cours, et les classes d'une même école devraient aussi être corrélées donc nous pouvons inclure des effets aléatoires aux niveaux de l'école et de la classe pour prendre en compte les différentes sources de variabilité.

Les chercheurs en médecine peuvent utiliser un modèle mixte linéaire généralisé pour déterminer si un médicament anti-convulsion peut réduire le taux des crises d'épilepsie d'un patient. Les mesures répétées chez le même patient sont normalement corrélées positivement, donc un modèle mixte avec des effets aléatoires devrait être adéquat. Le champ cible, le nombre de crises, prend des valeurs positives, donc un modèle mixte linéaire généralisé avec une distribution de Poisson et un lien de log devraient être appropriés.

Les dirigeants d'un fournisseur de services de télévision câblée, de service téléphoniques et Internet peuvent utiliser un modèle mixte linéaire généralisé pour en savoir plus sur les clients potentiels. Du fait que les réponses possibles ont des niveaux de mesure nominaux, l'analyste de la société utilise un modèle mixte logit généralisé avec une constante aléatoire pour capturer la corrélation entre les réponses aux questions d'utilisation du service selon différents types de services (tv, téléphone, Internet) et les réponses d'une enquête donnée.

L'onglet Structure des données vous permet de spécifier les relations structurelles existant entre les enregistrements de votre jeu de données lorsque les observations sont corrélées. Si les enregistrements du jeu de données représentent des observations indépendantes, vous n'avez pas besoin de spécifier quoi que ce soit sur cet onglet.

Sujets : La combinaison des valeurs des champs catégoriels spécifiés doit définir de manière unique les sujets à l'intérieur du jeu de données. Par exemple, un seul champ *ID du patient* doit suffire à définir les sujets dans un même hôpital, mais la combinaison des variables *ID de l'hôpital* et *ID du patient* peut être nécessaire si les numéros d'identification des patients ne sont pas uniques entre les hôpitaux. Dans le cas de mesures répétées, plusieurs observations sont enregistrées pour chaque sujet, de sorte que chaque sujet peut occuper plusieurs enregistrements dans le jeu de données.

Un **sujet** est une unité d'observation considérée comme indépendante des autres sujets. Par exemple, lors d'une étude médicale, la mesure de la tension artérielle d'un patient peut être considérée comme étant indépendante des résultats des autres patients. La définition des sujets devient particulièrement importante lorsqu'il existe deux évaluations répétées par sujet et que vous souhaitez modéliser la corrélation entre ces observations. Par exemple, vous pouvez vous attendre à ce que les mesures de tension artérielle d'un patient unique lors de visites consécutives chez le docteur soient en corrélation.

Tous les champs spécifiés comme Sujets dans l'onglet Structure des données sont utilisés pour définir les sujets dans le cadre de la structure de covariance des résidus, et constituent la liste des champs possibles à utiliser lors de la définition des sujets pour les structures de covariance à effets aléatoires sur le Bloc d'effet aléatoire.

Mesures répétées : Les champs spécifiés ici sont utilisés pour identifier des observations répétées. Par exemple, une variable unique *Semaine* peut identifier les 10 semaines d'observations dans une étude médicale, ou les variables *Mois* et *Jour* peuvent être utilisées ensemble pour identifier les observations quotidiennes sur une année entière.

Définir les groupes de covariance par : Les champs catégoriels spécifiés ici définissent des ensembles indépendants de paramètres de covariance d'effets répétés, un ensemble étant défini pour chaque catégorie par la classification croisée des champs de regroupement. Tous les sujets ont le même type de covariance ; les sujets à l'intérieur de la même classe de covariance auront les mêmes valeurs de paramètres.

Type de covariance répétée : Indique la structure de covariance des résidus. Les structures disponibles sont :

- Autorégressif de premier ordre (AR1)
- Moyenne mobile autorégressive (1,1) (ARMA11)
- Symétrie composée
- Diagonale
- Identité mise à l'échelle
- Toeplitz
- Sans structure
- Composantes de variance

Obtention d'un modèle linéaire mixte généralisé

Cette fonction nécessite l'option Statistiques avancées.

A partir des menus, sélectionnez :

Analyse > Modèles Mixtes > Linéaires généralisés...

1. Définissez la structure des sujets de votre jeu de données sur l'onglet **Structure des données**.
2. Sur l'onglet **Champs et effets**, une seule cible doit être présente. Celle-ci peut posséder tout niveau de mesure, ou une spécification d'événements/d'essais, auquel cas la spécification d'événements ou d'essais doit être continue. Spécifiez éventuellement sa distribution et sa fonction de lien, les effets fixes ainsi que les blocs d'effets aléatoires, le décalage ou les pondérations d'analyse.
3. Cliquez sur **Options de création** pour spécifier les paramètres optionnels de création.
4. Cliquez sur **Options du modèle** pour enregistrer les scores dans le jeu de données actif et exporter le modèle vers un fichier externe.
5. Cliquez sur **Exécuter** pour exécuter la procédure et créer les objets du modèle.

Cible

Ces paramètres définissent la cible, sa distribution et sa relation avec les prédicteurs via la fonction de lien.

Cible : La cible est requise. Elle peut avoir n'importe quel niveau de mesure, et le niveau de mesure de la cible limite les distributions et fonctions de lien appropriées.

- **Utiliser le nombre d'essais en tant que dénominateur** : Lorsque la réponse cible est un nombre d'événements se produisant lors de plusieurs essais, le champ cible contient ce nombre et vous pouvez sélectionner un autre champ comportant le nombre d'essais. Par exemple, lors du test d'un nouveau pesticide, vous devez exposer des échantillons de fourmis à différentes concentrations de pesticide et enregistrer le nombre de fourmis tuées et le nombre de fourmis exposées dans chaque échantillon. Dans ce cas, le champ enregistrant le nombre de fourmis tuées doit être spécifié comme le champ cible (d'événements), et le champ enregistrant le nombre de fourmis présentes dans chaque échantillon doit être spécifié comme le champ d'essais. Si le nombre de fourmis est identique dans tous les échantillons, alors le nombre d'essais peut être indiqué à l'aide d'une valeur fixe.
Le nombre d'essais doit être supérieur ou égal au nombre d'événements pour chaque enregistrement. Les événements doivent être des entiers non négatifs et les essais, des entiers positifs.
- **Personnaliser la catégorie de référence** : Pour une cible qualitative, vous pouvez choisir la catégorie de référence. Cela peut avoir des conséquences sur certaines sorties, comme les estimations de paramètre, mais ne devrait pas modifier l'ajustement du modèle. Par exemple, si votre cible prend les valeurs 0, 1 et 2 par défaut, la procédure fait de la dernière catégorie (la plus élevée), soit 2, la catégorie de référence. Dans cette situation, les estimations du paramètre doivent être interprétées comme concernant la probabilité de la catégorie 0 ou 1 *relativement* à la probabilité de la catégorie 2. Si vous spécifiez une catégorie personnalisée et que votre cible a des labels définis, vous pouvez stipuler la catégorie de référence en sélectionnant une valeur dans la liste. Ceci peut être pratique, si à mi-chemin de préciser un modèle, vous n'êtes pas sûr de la méthode de codage d'un champ spécifique.

Distribution et relation de la cible (lien) avec le modèle linéaire : D'après les valeurs des prédicteurs, le modèle s'attend à ce que la distribution des valeurs de la cible revête la forme spécifiée, et à ce que les valeurs cibles soient liées linéairement aux prédicteurs via la fonction de lien spécifiée. Des raccourcis sont fournis pour plusieurs modèles communs, ou vous pouvez choisir un paramètre **personnalisé** si vous voulez ajuster une distribution particulière et une combinaison de fonction de lien qui n'apparaît pas sur la liste des raccourcis.

- **Modèle linéaire** : Spécifie une distribution normale avec un lien d'identité, ce qui est utile lorsque la cible peut être prédite à l'aide d'une régression linéaire ou d'un modèle ANOVA.
- **Régression gamma** : Spécifie une distribution gamma avec un lien log, qui doit être utilisée lorsque la cible ne contient que des valeurs positives et est arrondie à des valeurs supérieures.
- **Log Linéaire** : Spécifie une distribution de Poisson avec un lien log, qui doit être utilisée lorsque la cible représente un nombre d'occurrences sur une période fixe.
- **Régression binomiale négative** : Spécifie une distribution binomiale négative avec un lien log, qui doit être utilisée lorsque la cible et le dénominateur représentent le nombre d'essais requis pour observer k succès.
- **Régression logistique multinomiale** : Spécifie une distribution multinomiale qui doit être utilisée lorsque la cible est une réponse multicatégorie. Elle utilise un lien logit cumulé (résultats ordinaux) ou un lien logit généralisé (réponses nominales multicatégorie).
- **Régression logistique binaire** : Spécifie une distribution binomiale avec un lien logit, qui doit être utilisée lorsque la cible est une réponse binaire prédite par un modèle de régression logistique.
- **Probit binaire** : Spécifie une distribution binomiale avec un lien probit, qui doit être utilisée lorsque la cible est une réponse binaire avec une distribution normale sous-jacente.
- **Intervalle - Données de survie censurées** : Spécifie une distribution binomiale avec un lien log-log complémentaire, ce qui est utile dans l'analyse de survie lorsque certaines observations n'ont pas d'événement d'arrêt.

Distribution

Cette sélection spécifie la distribution de la cible. La possibilité de spécifier une distribution non normale et une fonction de lien sans identité constitue la principale amélioration du modèle linéaire mixte généralisé par rapport au modèle linéaire mixte. Il existe de nombreuses combinaisons

distribution/fonction de lien possibles, et plusieurs peuvent convenir à un jeu de données particulier. Votre choix peut être guidé par des considérations théoriques a priori ou en fonction de la combinaison qui vous semble la plus adaptée.

- **Binomiale** : Cette distribution convient uniquement à une cible représentant une réponse binaire ou un nombre d'événements.
- **Gamma** : Cette distribution convient à une cible avec des valeurs d'échelle positives arrondies à des valeurs positives supérieures. Si une valeur de données est inférieure ou égale à 0, ou manquante, l'observation correspondante n'est pas utilisée dans l'analyse.
- **Gaussienne inverse** : Cette distribution convient à une cible avec des valeurs d'échelle positives arrondies à des valeurs positives supérieures. Si une valeur de données est inférieure ou égale à 0, ou manquante, l'observation correspondante n'est pas utilisée dans l'analyse.
- **Multinomiale** : Cette distribution convient à une cible représentant une réponse multicatégorique. Le format du modèle dépendra du niveau de mesure de la cible.

Une cible **nominale** engendrera un modèle multinominal nominal dans lequel un ensemble de paramètres de modèle distinct est estimé pour chaque catégorie de la cible (à l'exception de la catégorie de référence). Les estimations de paramètre pour un prédicteur indiquent la relation entre ce prédicteur et la probabilité de chaque catégorie de la cible, par rapport à la catégorie de référence.

Une cible **ordinaire** engendrera un modèle multinominal ordinal dans lequel la constante traditionnelle est remplacée par un ensemble de paramètres **deseuil** associés à la probabilité cumulée des catégories cibles.

- **Binomiale négative** : La régression binomiale négative utilise une distribution binomiale négative avec un lien log, qui doit être utilisée lorsque la cible représente un nombre d'occurrences de variance élevée.
- **Normale** : Cette distribution convient à une cible continue dont les valeurs suivent une distribution symétrique, en cloche, autour d'une valeur centrale (moyenne).
- **Poisson** : Vous pouvez considérer cette distribution comme étant le nombre d'occurrences d'un événement qui vous intéresse au cours d'une période fixe. Elle convient aux variables comportant des entiers non négatifs. Si une valeur de données n'est pas un entier, est inférieure à 0 ou est manquante, l'observation correspondante n'est pas utilisée dans l'analyse.

Fonctions de lien

La fonction de lien consiste en une transformation de la cible permettant d'estimer le modèle. Les fonctions suivantes sont disponibles :

- **Identité** : $f(x)=x$. La cible n'est pas transformée. Ce lien peut être utilisé avec n'importe quelle distribution, exceptée la distribution multinomiale.
- **Log-log complémentaire** : $f(x)=\log(-\log(1-x))$. Cette fonction convient uniquement à la distribution binomiale ou multinomiale.
- **Cauchit** : $f(x) = \tan(\pi (x - 0.5))$. Cette fonction convient uniquement à la distribution binomiale ou multinomiale.
- **Journal** : $f(x)=\log(x)$. Ce lien peut être utilisé avec n'importe quelle distribution, exceptée la distribution multinomiale.
- **Log complémentaire** : $f(x)=\log(1-x)$. Cette fonction convient uniquement à la distribution binomiale.
- **Logit** : $f(x)=\log(x / (1-x))$. Cette fonction convient uniquement à la distribution binomiale ou multinomiale.
- **Log-log négatif** : $f(x)=-\log(-\log(x))$. Cette fonction convient uniquement à la distribution binomiale ou multinomiale.
- **Probit** : $f(x)=\Phi^{-1}(x)$, où Φ^{-1} représente la fonction de distribution cumulée normale inverse. Cette fonction convient uniquement à la distribution binomiale ou multinomiale.
- **Puissance** : $f(x)=x^\alpha$, si $\alpha \neq 0$. $f(x)=\log(x)$, si $\alpha=0$. α représente le nombre requis et doit être un nombre réel. Ce lien peut être utilisé avec n'importe quelle distribution, exceptée la distribution multinomiale.

Effets fixes

Les facteurs d'effets fixes sont généralement considérés comme des champs dont les valeurs intéressantes sont toutes représentées dans le jeu de données, et peuvent être utilisés pour l'évaluation. Par défaut, les champs avec un rôle d'entrée prédéfini qui ne sont pas spécifiés ailleurs dans la boîte de dialogue sont entrés dans la partie effets fixes du modèle. Les champs qualitatifs (nominaux et ordinaux) sont utilisés comme facteurs dans le modèle et les champs continus sont utilisés comme covariables.

Entrez les effets dans le modèle en sélectionnant un ou plusieurs champs dans la liste source et en le(s) faisant glisser vers la liste des effets. Le type d'effet créé dépend de l'endroit où vous déposez la sélection.

- **Principaux** : Les champs déposés apparaissent sous forme d'effets principaux séparés au bas de la liste des effets.
- **Bidirectionnels** : Toutes les paires possibles des champs déposés apparaissent sous forme d'interactions bidirectionnelles au bas de la liste des effets.
- **Tridirectionnels** : Tous les triplets possibles des champs déposés apparaissent sous forme d'interactions tridirectionnelles au bas de la liste des effets.
- *. Combinaison de tous les champs déposés sous forme d'interaction simple au bas de la liste des effets.

Les boutons situés à droite du générateur d'effets vous permettent d'effectuer diverses actions.

Tableau 1. Description du bouton générateur d'effets.

Icône	Description
	Supprimer des termes du modèle à effets fixes, en sélectionnant ceux que vous souhaitez supprimer, puis en cliquant sur le bouton de suppression.
	Réordonner les termes dans le modèle à effets fixes, en sélectionnant ceux que vous souhaitez réorganiser, puis en cliquant sur les flèches vers le haut ou vers le bas.
	Ajouter des termes imbriqués au modèle à l'aide de la boîte de dialogue «Ajout d'un terme personnalisé», en cliquant sur le bouton d'ajout d'un terme personnalisé.

Inclure constante : La constante est généralement incluse dans le modèle. Si vous partez du principe que les données passent par l'origine, vous pouvez exclure la constante.

Ajout d'un terme personnalisé

Dans cette procédure, vous pouvez construire des termes imbriqués pour votre modèle. 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 à divers emplacements de magasin. 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.

Limites : 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.

Construction d'un terme imbriqué

1. Sélectionnez une covariable ou un facteur imbriqué dans un autre facteur, puis cliquez sur le bouton fléché.
2. Cliquez sur **(Dans)**.
3. Sélectionnez le facteur dans lequel la covariable ou le facteur précédent est imbriqué, puis cliquez sur le bouton en forme fléché.
4. Cliquez sur **Ajouter terme**.

Vous pouvez éventuellement inclure des effets d'interaction ou ajouter plusieurs niveaux d'imbrication au terme imbriqué.

Effets aléatoires

Les facteurs d'effets aléatoires sont des champs dont les valeurs dans le fichier de données peuvent être considérées comme un échantillon aléatoire d'un plus grand nombre de valeurs. Ils sont utiles pour expliquer la variabilité des excès dans la cible. Par défaut, si vous avez sélectionné plus d'un sujet dans l'onglet Structure des données, un bloc Effet aléatoire sera créé pour chaque sujet au-delà du sujet le plus interne. Par exemple, si vous avez sélectionné École, Classe et Étudiant comme sujets sur l'onglet Structure des données, les blocs d'effets aléatoires suivants sont automatiquement créés :

- Effet aléatoire 1 : le sujet est l'école (sans effets, constante seulement)
- Effet aléatoire 2 : le sujet est l'école * classe (sans effets, constante uniquement)

Vous pouvez travailler avec les blocs d'effets aléatoires comme suit :

1. Pour ajouter un nouveau bloc, cliquez sur **Ajouter un bloc...** La boîte de dialogue «Bloc d'effets aléatoires» s'ouvre.
2. Pour modifier un bloc existant, sélectionnez le bloc que vous souhaitez modifier et cliquez sur **Modifier un bloc...** La boîte de dialogue «Bloc d'effets aléatoires» s'ouvre.
3. Pour supprimer un ou plusieurs blocs, sélectionnez le(s) bloc(s) que vous souhaitez supprimer et cliquez sur le bouton Suppr.

Bloc d'effets aléatoires

Entrez les effets dans le modèle en sélectionnant un ou plusieurs champs dans la liste source et en le(s) faisant glisser vers la liste des effets. Le type d'effet créé dépend de l'endroit où vous déposez la sélection. Les champs qualitatifs (nominaux et ordinaux) sont utilisés comme facteurs dans le modèle et les champs continus sont utilisés comme covariables.

- **Principaux** : Les champs déposés apparaissent sous forme d'effets principaux séparés au bas de la liste des effets.
- **Bidirectionnels** : Toutes les paires possibles des champs déposés apparaissent sous forme d'interactions bidirectionnelles au bas de la liste des effets.
- **Tridirectionnels** : Tous les triplets possibles des champs déposés apparaissent sous forme d'interactions tridirectionnelles au bas de la liste des effets.
- *****. Combinaison de tous les champs déposés sous forme d'interaction simple au bas de la liste des effets.

Les boutons situés à droite du générateur d'effets vous permettent d'effectuer diverses actions.

Tableau 2. Description du bouton générateur d'effets.

Icône	Description
	Supprimer des termes du modèle en sélectionnant ceux que vous souhaitez supprimer, puis en cliquant sur le bouton de suppression.
	Réordonner les termes dans le modèle en sélectionnant ceux que vous souhaitez réorganiser, puis en cliquant sur les flèches vers le haut ou vers le bas.
	
	Ajouter des termes imbriqués au modèle à l'aide de la boîte de dialogue «Ajout d'un terme personnalisé», à la page 63, en cliquant sur le bouton d'ajout d'un terme personnalisé.

Inclure constante : La constante n'est pas incluse dans le modèle à effets aléatoires par défaut. Si vous partez du principe que les données passent par l'origine, vous pouvez exclure la constante.

Définir les groupes de covariance par : Les champs catégoriels spécifiés ici définissent des ensembles indépendants de paramètres de covariance d'effets aléatoires, un ensemble étant défini pour chaque catégorie par la classification croisée des champs de regroupement. Un ensemble de champs de regroupement différent peut être spécifié pour chaque bloc d'effet aléatoire. Tous les sujets ont le même type de covariance ; les sujets à l'intérieur de la même classe de covariance auront les mêmes valeurs de paramètres.

Combinaison de sujets : Cette option vous permet de spécifier des sujets à effets aléatoires à partir de combinaisons prédéfinies de sujets depuis l'onglet Structure des données. Par exemple, si *École*, *Classe* et *Étudiant* sont définis comme des sujets sur l'onglet Structure des données, et dans cet ordre, alors la liste déroulante Combinaison de sujets aura **Aucun**, **École**, **École * Classe**, et **École * Classe * Étudiant** comme options.

Type de covariance des effets aléatoires : Indique la structure de covariance des résidus. Les structures disponibles sont :

- Autorégressif de premier ordre (AR1)
- Moyenne mobile autorégressive (1,1) (ARMA11)
- Symétrie composée
- Diagonale
- Identité mise à l'échelle
- Toeplitz
- Sans structure
- Composantes de variance

Pondération et décalage

Pondération d'analyse : Le paramètre d'échelle est un paramètre de modèle estimé lié à la variance de la réponse. Les pondérations d'analyse sont des valeurs "connues", susceptibles de varier d'une observation à l'autre. Si le champ de pondération d'analyse est spécifié, le paramètre d'échelle, qui est lié à la variance de la réponse, est divisé par les valeurs de pondération d'analyse pour chaque observation. Les enregistrements avec des valeurs de pondération d'analyse inférieures ou égales à 0, ou manquantes, ne sont pas utilisées dans l'analyse.

Décalage : Le terme de décalage est un prédicteur « structurel ». Son coefficient n'est pas estimé par le modèle mais est supposé être égal à 1. Ainsi, les valeurs du décalage sont simplement ajoutées au

prédicteur linéaire de la cible. Ceci est particulièrement utile dans les modèles de régression de Poisson, où chaque observation peut avoir différents niveaux d'exposition à l'événement étudié.

Par exemple, dans le cas de la modélisation du taux d'accident chez les conducteurs, il existe une énorme différence entre un conducteur responsable d'un accident avec 3 ans de permis et un conducteur responsable d'un accident en 25 ans. Le nombre d'accidents peut être modélisé sous la forme d'une réponse de Poisson ou d'une réponse binomiale négative avec un lien log si le log naturel de l'expérience du conducteur est inclus en tant que terme de décalage.

D'autres combinaisons de types de distribution et de lien nécessiteraient d'autres transformations de la variable de décalage.

Options de création générales

Ces sélections spécifient des critères plus avancés utilisés pour créer le modèle.

Ordre de tri : Ces contrôles déterminent l'ordre des catégories pour la cible et les facteurs (entrées qualitatives) afin de déterminer la "dernière" catégorie. Le paramètre ordre de tri de la cible est ignoré si la cible n'est pas qualitative ou si une catégorie de référence personnalisée est spécifiée sur les paramètres «Cible», à la page 60.

Règles d'arrêt : Vous pouvez spécifier le nombre maximal d'itérations exécutées par l'algorithme. L'algorithme utilise un double processus itératif qui comporte une boucle interne et une boucle externe. La valeur maximale d'itérations s'applique aux deux boucles. Spécifiez un nombre entier non négatif. La valeur par défaut est 100.

Paramètres post-estimation : Ces paramètres déterminent la façon dont une partie de la sortie du modèle est calculée pour l'affichage.

- **Niveau de confiance :** Il s'agit du niveau de confiance utilisé pour calculer les estimations d'intervalle des coefficients de modèle. Spécifiez une valeur supérieure à 0 et inférieure à 100. La valeur par défaut est 95.
- **Degrés de liberté :** Cela spécifie la façon dont les degrés de liberté sont calculés pour les tests de signification. Choisissez **Fixes pour tous les tests (méthode des résidus)** si la taille de votre échantillon est suffisamment grande, si les données sont équilibrées ou si le modèle utilise un type de covariance plus simple ; par exemple, identité mise à l'échelle ou diagonale. Il s'agit de la valeur par défaut. Choisissez **Différents selon les tests (approximation Satterthwaite)** si la taille de votre échantillon est petite, si les données sont déséquilibrées ou si le modèle utilise un type de covariance compliqué ; par exemple, non structuré.
- **Tests des effets fixes et coefficients :** C'est la méthode de calcul de la matrice de covariance des estimations de paramètres. Choisissez l'estimation robuste si vous craignez que les hypothèses du modèle sont violées.

Estimation

L'algorithme de génération de modèle utilise un double processus itératif qui comporte une boucle interne et une boucle externe. Les paramètres ci-dessous s'appliquent à la boucle interne.

Convergence des paramètres.

La convergence est prise en compte si la modification absolue ou relative maximum des estimations de paramètres est inférieure à la valeur spécifiée, qui doit être non négative. Le critère n'est pas utilisé si la valeur spécifiée est égale à 0.

Convergence de log de vraisemblance.

La convergence est prise en compte si la modification absolue ou relative de la fonction log de vraisemblance est inférieure à la valeur spécifiée, qui doit être non négative. Le critère n'est pas utilisé si la valeur spécifiée est égale à 0.

Convergence de Hess.

Pour la spécification **Absolu**, la convergence est prise en compte si une statistique basée sur la matrice de Hess est inférieure à la valeur indiquée. Pour la spécification **Relatif**, la convergence est prise en compte si la statistique est inférieure au produit de la valeur indiquée et de la valeur absolue du log de vraisemblance. Le critère n'est pas utilisé si la valeur spécifiée est égale à 0.

Nombre maximal d'étapes de l'évaluation de Fisher.

Spécifiez un nombre entier non négatif. La valeur 0 spécifie la méthode Newton-Raphson. Les valeurs supérieures à 0 indiquent qu'il faut utiliser l'algorithme d'évaluation de Fisher jusqu'au numéro d'itération n , où n est l'entier spécifié, puis la méthode Newton-Raphson après.

Tolérance singularité.

Cette valeur est utilisée comme valeur de tolérance lors du contrôle des singularités. Indiquez une valeur positive.

Remarque : Par défaut, le paramètre Convergence des paramètres est utilisé lorsque le critère **Absolu** de modification maximum est contrôlé avec une tolérance 1E-6. Ce paramètre peut produire des résultats différents de ceux obtenus dans les versions antérieures à la version 22. Pour reproduire les résultats de ces versions, utilisez le critère de convergence des paramètres **Relatif** et conservez la valeur de tolérance 1E-6 par défaut.

Moyennes estimées

Cet onglet vous permet d'afficher les moyennes marginales estimées des niveaux et interactions entre facteurs. Les moyennes marginales estimées ne sont pas disponibles pour les modèles multinomiaux.

Termes : Les termes du modèle dans les effets fixes entièrement composés de champs catégoriels sont répertoriés ici. Cochez chaque terme pour lequel vous souhaitez que le modèle produise des moyennes marginales estimées.

- **Type de contraste :** Cela spécifie le type de contraste à utiliser pour le champ niveaux de contraste. Si **Aucun** est sélectionné, aucun contraste n'est produit. **Par paire** produit des comparaisons appariées pour toutes les combinaisons de niveaux des facteurs spécifiés. C'est le seul contraste disponible pour les interactions entre facteurs. Les contrastes de **Déviati**on comparent chaque niveau du facteur à la grande moyenne. Les contrastes **simples** comparent chaque niveau du facteur, sauf le dernier, au dernier niveau. Le "dernier" niveau est déterminé par l'ordre de tri des facteurs spécifiés dans Options de génération. Notez que tous ces types de contrastes ne sont pas orthogonaux.
- **Champ Contraste :** Cela spécifie un facteur, dont les niveaux sont comparés à l'aide du type de contraste sélectionné. Si **Aucun** est sélectionné comme type de contraste, aucun champ de contraste ne peut (ou ne doit) être sélectionné.

Champs continus : Les champs continus répertoriés sont extraits des termes situés dans les effets fixes utilisant des champs continus. Lors du calcul des moyennes marginales estimées, les covariables sont fixées aux valeurs spécifiées. Sélectionnez la moyenne ou spécifiez une valeur personnalisée.

Afficher les moyennes estimées en termes de : Cela spécifie s'il faut calculer les moyennes marginales estimées d'après l'échelle originale de la cible ou d'après la transformation de la fonction de lien. **Échelle de cible originale** calcule les moyennes marginales estimées pour la cible. Notez que lorsque la cible est spécifiée à l'aide de l'option événements/essais, cela donne les moyennes marginales estimées pour la proportion événements/essais plutôt que pour le nombre d'événements. **Transformation de la fonction de lien** calcule les moyennes marginales estimées pour le prédicteur linéaire.

Apportez des ajustements pour des comparaisons multiples à l'aide de : Lors de l'exécution de tests d'hypothèse avec plusieurs contrastes, vous pouvez ajuster le niveau global de signification à partir des niveaux de signification des contrastes inclus. Cela vous permet de choisir la méthode d'ajustement.

- **Différence la moins significative :** Cette méthode ne contrôle pas l'intégralité de la probabilité de rejet des hypothèses qui présentent des contrastes linéaires différents des valeurs d'hypothèse nulles.

- *Bonferroni séquentiel*. Il s'agit d'une procédure descendante de rejet séquentiel de Bonferroni beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Sidak séquentiel*. Il s'agit d'une procédure descendante de rejet séquentiel de Sidak beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.

La méthode de plus petite différence significative est moins stricte que la méthode de Sidak séquentielle, qui elle-même est moins stricte que la méthode séquentielle de Bonferroni. En d'autres termes, la plus petite différence significative va rejeter au moins autant d'hypothèses individuelles que la méthode séquentielle de Sidak, qui elle-même va rejeter au moins autant d'hypothèses individuelles que la méthode séquentielle de Bonferroni.

Enregistrement

Les éléments cochés sont enregistrés avec le nom spécifié ; les conflits avec les noms de champs existants ne sont pas autorisés.

Prévisions : Enregistre la valeur prédite de la cible. Le nom par défaut du champ est *PredictedValue*.

Probabilité prédite pour les cibles qualitatives : Si la cible est qualitative, ce mot-clé enregistre les probabilités prédites des n premières catégories, jusqu'à la valeur spécifiée comme **Catégories maximum à enregistrer**. Les valeurs calculées sont des probabilités cumulées pour les cibles ordinales. Le nom de racine par défaut est *PredictedProbability*. Pour enregistrer la probabilité prédite de la catégorie prédite, enregistrez la confiance (voir ci-dessous).

Intervalles de confiance : Enregistre les limites supérieure et inférieure de l'intervalle de confiance de la valeur ou de la probabilité prédite. Pour toutes les distributions sauf les multinomiales, cela crée deux variables et le nom de racine par défaut est *IC*, avec *_Inférieur* et *_Supérieur* comme suffixes.

Pour la distribution multinomiale et une cible nominale, un champ est créé pour chaque catégorie de variable dépendante. Il enregistre les limites inférieures et supérieures des probabilités prédites des n premières catégories, jusqu'à la valeur spécifiée comme **Catégories maximum à enregistrer**. Le nom de racine par défaut est *IC*, et les noms de champs par défaut sont *IC_Inférieur_1*, *IC_Supérieur_1*, *IC_Inférieur_2*, *IC_Supérieur_2*, etc., selon l'ordre des catégories de la cible.

Pour la distribution multinomiale et une cible ordinale, un champ est créé pour chaque catégorie de variable dépendante à l'exception de la dernière (pour plus d'informations, voir la rubrique «Options de création générales», à la page 66). Cela enregistre les limites inférieure et supérieure de la probabilité prédite cumulative pour les n premières catégories, jusqu'à l'avant-dernière et jusqu'à la valeur spécifiée comme **Catégories maximum à enregistrer**. Le nom de racine par défaut est *IC*, et les noms de champs par défaut sont *IC_Inférieur_1*, *IC_Supérieur_1*, *IC_Inférieur_2*, *IC_Supérieur_2*, etc., selon l'ordre des catégories de la cible.

Résiduels de Pearson : Enregistre le résiduel de Pearson pour chaque enregistrement, qui peut être utilisé après l'estimation comme diagnostic de l'ajustement du modèle. Le nom par défaut du champ est *PearsonResidual*.

Confiances : Enregistre la confiance dans la valeur prédite pour la cible qualitative. La confiance calculée peut être basée sur la probabilité de la valeur prédite (la probabilité prédite la plus élevée) ou sur la différence entre la probabilité prédite la plus élevée et la deuxième probabilité prédite la plus élevée. Le nom par défaut du champ est *Confiance*.

Exporter le modèle : Cette option écrit le modèle sur un fichier *.zip* externe. Vous pouvez utiliser ce fichier de modèle pour appliquer les informations du modèle aux autres fichiers de données à des fins d'évaluation. Spécifiez un nom de fichier valide et unique. Si la spécification du fichier pointe vers un fichier existant, le fichier est écrasé.

Vue Modèle

La procédure crée un objet Modèle dans le visualiseur. En activant cet objet par un double-clic, vous obtenez une vue interactive du modèle.

Par défaut, la vue Récapitulatif du modèle apparaît. Pour voir une autre vue de modèle, sélectionnez-la parmi les miniatures des vues.

Comme alternative à l'objet de modèle, vous pouvez également générer des tableaux croisés dynamiques et des graphiques en sélectionnant **Tableaux croisés dynamiques et graphiques** dans le groupe Affichage de la sortie situé sur l'onglet Sortie de la boîte de dialogue Options (Edition > Options). Les rubriques suivantes décrivent l'objet de modèle.

Récapitulatif du modèle

Cette vue du modèle est un instantané, permettant de consulter en un coup d'oeil le modèle et son ajustement.

Tableau : Le tableau identifie la cible, la distribution des probabilités et la fonction de lien spécifiée dans les Paramètres de la cible. Si la cible est définie par des événements et des essais, la cellule est divisée de façon à montrer le champ des événements et le champ des essais ou un nombre fixe d'essais. En outre, le critère d'information Akaike corrigé (AICC) et le critère d'information bayésien (BIC) de l'échantillon fini sont affichés.

- *Akaike corrigé.* Mesure de sélection et de comparaison des modèles mixtes basée sur la valeur Log de vraisemblance -2 (restreinte). Les petites valeurs indiquent de meilleurs modèles. L'AICC "corrige" l'AIC pour obtenir des tailles d'échantillon plus petites. Plus la taille de l'échantillon augmente, plus l'AICC converge vers l'AIC.
- *Bayésien.* Mesure de sélection et de comparaison des modèles basée sur le log de vraisemblance -2 . Les petites valeurs indiquent de meilleurs modèles. Le critère d'informations bayésien (BIC) "pénalise" les modèles surparamétrés (modèles complexes avec un nombre élevé d'entrées, par exemple), mais plus de manière plus stricte que le critère d'informations d'Akaike (AIC).

Graphique : Si la cible est qualitative, un graphique affiche la précision du modèle final, qui est le pourcentage de classifications correctes.

Structure des données

Cette vue offre un récapitulatif de la structure des données que vous avez spécifiée et vous aide à vérifier que les sujets et les mesures répétées ont été correctement spécifiés. Les informations observées pour le premier sujet sont affichées pour chaque champ de sujet et champ de mesures répétées, et la cible. En outre, le nombre de niveaux pour chaque champ de sujet et champ de mesures répétées est affiché.

Valeurs prédites en fonction des valeurs observées

Pour les cibles continues, y compris les cibles spécifiées comme événements/essais, ceci affiche un nuage de points mis en intervalles des valeurs prédites sur l'axe vertical par les valeurs observées sur l'axe horizontal. Idéalement, les points devraient se trouver sur une ligne de 45 degrés ; cette vue peut indiquer si des enregistrements sont particulièrement mal prédits par le modèle.

Classification

Pour les cibles qualitatives, cela affiche la classification croisée des valeurs observées par rapport aux valeurs prédites sur une carte thermique, ainsi que le pourcentage global correct.

Styles de tableaux : Il existe différents styles d'affichage, accessibles depuis la liste déroulante **Style** .

- **Pourcentages de ligne :** Cela affiche les pourcentages de ligne (les effectifs de cellules exprimés sous forme de pourcentage du nombre total de lignes) dans les cellules. Il s'agit de la valeur par défaut.
- **Effectifs de cellule :** Cela affiche les effectifs de cellules dans les cellules. L'ombrage de la carte thermique reste basé sur les pourcentages de ligne.
- **Carte thermique :** Cela n'affiche aucune valeur dans les cellules, juste l'ombrage.
- **Compressé :** Cela n'affiche aucun titre de ligne ou de colonne, ni valeur dans les cellules. Cela peut être utile lorsque la cible comporte beaucoup de catégories.

Manquant : Si des enregistrements ont des valeurs manquantes sur la cible, elles sont affichées sur une ligne (**Manquant**) sous les lignes valides. Les enregistrements avec des valeurs manquantes ne contribuent pas au pourcentage global correct.

Cibles multiples : S'il y a plusieurs cibles qualitatives, chaque cible est affichée dans un tableau séparé et il y a une liste déroulante **Cible** qui contrôle la cible à afficher.

Grands tableaux : Si la cible affichée comporte plus de 100 catégories, aucun tableau n'est affiché.

Effets fixes

Cette vue affiche la taille de chaque effet fixe dans le modèle.

Styles. Il existe différents styles d'affichage accessibles depuis la liste déroulante **Style** .

- **Graphique.** Il s'agit d'un graphique dans lequel les effets sont triés de haut en bas dans l'ordre où ils ont été spécifiés sur les paramètres des effets fixes. Les lignes de connexion du diagramme sont pondérées en fonction de la signification de l'effet, une largeur de ligne plus importante correspondant à des effets plus importants (valeurs p plus petites). Il s'agit de la valeur par défaut.
- **Tableau :** Il s'agit d'un tableau ANOVA pour le modèle général et les effets de modèle individuels. Les effets individuels sont triés de haut en bas dans l'ordre où ils ont été spécifiés sur les paramètres des effets fixes.

Signification. Il existe un curseur de signification qui contrôle les effets qui sont affichés dans la vue. Les effets ayant des valeurs de signification plus grandes que la valeur du curseur sont masqués. Ceci ne modifie pas le modèle, mais vous permet simplement de vous concentrer sur les effets les plus importants. Par défaut, la valeur est de 1,00, de sorte qu'aucun effet n'est filtré en fonction de la signification.

Coefficients fixes

Cette vue affiche la valeur de chaque coefficient fixe du modèle. Veuillez noter que les facteurs (prédicteurs catégoriels) sont codés par un indicateur dans le modèle, de sorte que les **effets** comportant des facteurs ont généralement plusieurs **coefficients** associés, un pour chaque catégorie exceptée la catégorie correspondant au coefficient redondant.

Styles : Il existe différents styles d'affichage accessibles depuis la liste déroulante **Style** .

- **Graphique :** Il s'agit d'un graphique qui affiche la constante d'abord, puis qui trie les effets de haut en bas dans l'ordre où ils ont été spécifiés sur les paramètres des effets fixes. Au sein des effets contenant des facteurs, les coefficients sont triés dans l'ordre croissant de la valeur des données. Les lignes de connexion du diagramme sont colorées et pondérées en fonction de la signification du coefficient, une largeur de ligne plus importante correspondant à des coefficients plus importants (valeurs p plus petites). Il s'agit du style par défaut.

- **Tableau** : Affiche les valeurs, les tests de signification et les intervalles de confiance des coefficients de modèles individuels. Après la constante, les effets individuels sont triés de haut en bas dans l'ordre où ils ont été spécifiés sur les paramètres des effets fixes. Au sein des effets contenant des facteurs, les coefficients sont triés dans l'ordre croissant de la valeur des données.

Multinomiale : Si la distribution multinomiale est activée, la liste déroulante Multinomiale contrôle quelle catégorie de cible à afficher. L'ordre de tri des valeurs dans la liste est déterminé par la spécification des paramètres des options de création.

Exponentiel : Cela affiche les estimations de coefficient exponentiel et les intervalles de confiance pour certains types de modèles, y compris la régression logistique binaire (distribution binomial et lien logit), la régression logistique nominale (distribution multinomial et lien logit), la régression binomiale négative (distribution binomiale négative et lien log) et le modèle Log-linéaire (distribution de Poisson et lien log).

Signification : Il existe un curseur de signification qui contrôle les coefficients qui sont affichés dans la vue. Les coefficients ayant des valeurs de signification plus grandes que la valeur du curseur sont masqués. Ceci ne modifie pas le modèle, mais vous permet simplement de vous concentrer sur les coefficients les plus importants. Par défaut, la valeur est de 1,00, de sorte qu'aucun coefficient n'est filtré en fonction de la signification.

Covariances des effets aléatoires

Cette vue affiche la matrice de covariance des effets aléatoires (**G**).

Styles : Il existe différents styles d'affichage accessibles depuis la liste déroulante **Style** .

- **Valeurs de covariance** : Il s'agit d'une carte thermique de la matrice de covariance dans laquelle les effets sont triés de haut en bas dans l'ordre où ils ont été spécifiés sur les paramètres des effets fixes. Les couleurs du corrgram correspondent aux valeurs des cellules telles qu'affichées dans la clé. Il s'agit de la valeur par défaut.
- **Corrgram** : Il s'agit d'une carte thermique de la matrice de covariance.
- **Compressé** : Il s'agit d'une carte thermique de la matrice de covariance sans les titres de ligne et de colonne.

Blocs : S'il y a plusieurs blocs d'effets aléatoires, il y a une liste déroulante Bloc pour sélectionner le bloc à afficher.

Groupes : Si un bloc d'effets aléatoires a une spécification de groupe, il y a une liste déroulante Groupe pour sélectionner le niveau de groupe à afficher.

Multinomiale : Si la distribution multinomiale est activée, la liste déroulante Multinomiale contrôle quelle catégorie de cible à afficher. L'ordre de tri des valeurs dans la liste est déterminé par la spécification des paramètres des options de création.

Paramètres de covariance

Cette vue affiche les estimations de paramètres de covariance et les statistiques associées pour les effets résiduels et aléatoires. Il s'agit de résultats élaborés, mais fondamentaux, qui indiquent si la structure de covariance est appropriée ou pas.

Table récapitulative : Il s'agit d'un aide-mémoire pour le nombre de paramètres dans les matrices de covariance des effets résiduels (**R**) et aléatoires (**G**), le rang (nombre de colonnes) dans les matrices de plan des effets fixes (**X**) et aléatoires (**Z**) et le nombre de sujets définis par les champs de sujet qui définissent la structure des données.

Tableau des paramètres de covariance : Pour l'effet sélectionné, l'estimation, l'erreur standard et l'intervalle de confiance sont affichés pour chaque paramètre de covariance. Le nombre de paramètres

affiché dépend de la structure de covariance pour l'effet et, pour les blocs d'effets aléatoires, du nombre d'effets dans le bloc. Si vous voyez que les paramètres hors diagonale ne sont pas significatifs, vous pouvez peut-être utiliser une structure de covariance plus simple.

Effets : S'il y a plusieurs blocs d'effets aléatoires, il y a une liste déroulante Effet pour sélectionner le bloc d'effets résiduels ou aléatoires à afficher. L'effet résiduel est toujours disponible.

Groupes : Si un bloc d'effets résiduels ou aléatoires a une spécification de groupe, il y a une liste déroulante Groupe pour sélectionner le niveau de groupe à afficher.

Multinomiale : Si la distribution multinomiale est activée, la liste déroulante Multinomiale contrôle quelle catégorie de cible à afficher. L'ordre de tri des valeurs dans la liste est déterminé par la spécification des paramètres des options de création.

Moyennes estimées : Effets significatifs

Il s'agit des graphiques affichés pour les 10 effets de tous les facteurs fixes "les plus significatifs", avec d'abord les interactions tridirectionnelles, puis les interactions bidirectionnelles et enfin les effets principaux. Le graphique affiche la valeur de la cible estimée par le modèle sur l'axe vertical pour chaque valeur de l'effet principal (ou l'effet listé en premier dans une interaction) sur l'axe horizontal ; une ligne séparée est produite pour chaque valeur du deuxième effet listé dans une interaction ; un graphique séparé est produit pour chaque valeur du troisième effet listé dans une interaction tridirectionnelle ; tous les autres prédicteurs restent constantes. Il offre une visualisation pratique des effets des coefficients de chaque prédicteur sur la cible. Notez que si aucun prédicteur n'est significatif, aucune moyenne estimée n'est générée.

Confiance : Cela affiche les limites de confiance supérieure et inférieure pour les moyennes marginales, en utilisant le niveau de confiance spécifié dans le cadre des Options de création.

Moyennes estimées : Effets personnalisés

Il s'agit de tableaux et de graphiques pour les effets de tous les facteurs fixes demandés par l'utilisateur.

Styles : Il existe différents styles d'affichage accessibles depuis la liste déroulante **Style** .

- **Graphique** : Ce style affiche un graphique curviligne de la valeur de la cible estimée par le modèle sur l'axe vertical pour chaque valeur de l'effet principal (ou l'effet listé en premier dans une interaction) sur l'axe horizontal ; une ligne séparée est produite pour chaque valeur du deuxième effet listé dans une interaction ; un graphique séparé est produit pour chaque valeur du troisième effet listé dans une interaction tridirectionnelle ; tous les autres prédicteurs restent constants.

Si des contrastes sont demandés, un autre graphique est affiché pour comparer les niveaux du champ contraste ; pour les interactions, un graphique est affiché pour chaque combinaison de niveaux des effets autres que le champ contraste. Pour les contrastes **par paire**, il s'agit d'un graphique de réseau des distances. C'est une représentation graphique du tableau des comparaisons dans lequel les distances entre les noeuds du réseau correspondent aux différences entre les échantillons. Les lignes jaunes correspondent aux différences statistiques significatives, alors que les lignes noires correspondent aux différences non significatives. Lorsque vous passez la souris sur une ligne du réseau, la signification ajustée de la différence entre les noeuds connectés par la ligne s'affiche dans une infobulle.

Pour les contrastes de **déviatio**n , un graphique à barres est affiché avec la valeur de la cible estimée par le modèle sur l'axe vertical et les valeurs du champ contraste sur l'axe horizontal ; pour les interactions, un graphique est affiché pour chaque combinaison de niveaux des effets autres que le champ contraste. Les barres montrent la différence entre chaque niveau du champ contraste et la moyenne globale, qui est représentée par une ligne horizontale noire.

Pour les contrastes **simples** , un graphique à barres est affiché avec la valeur de la cible estimée par le modèle sur l'axe vertical et les valeurs du champ contraste sur l'axe horizontal ; pour les interactions, un graphique est affiché pour chaque combinaison de niveaux des effets autres que le champ contraste.

Les barres montrent la différence entre chaque niveau du champ contraste (sauf le dernier) et le dernier niveau, qui est représenté par une ligne horizontale noire.

- **Tableau** : Ce style affiche un tableau de la valeur de la cible estimée par le modèle, son erreur standard et l'intervalle de confiance pour chaque combinaison de niveaux des champs de l'effet ; tous les autres prédicteurs restent constants.

Si des contrastes sont demandés, un autre tableau est affiché avec l'estimation, l'erreur standard, le test de signification et l'intervalle de confiance pour chaque contraste ; pour les interactions, il y a un ensemble séparé de lignes pour chaque combinaison de niveaux des effets autres que le champ contraste. En outre, un tableau avec les résultats des tests globaux est affiché ; pour les interactions, il y a un test global séparé pour chaque combinaison de niveaux des effets autres que le champ contraste.

Confiance : Cela affiche ou masque les limites de confiance supérieure et inférieure pour les moyennes marginales, en utilisant le niveau de confiance spécifié dans le cadre des Options de création.

Présentation : Cela affiche ou masque la présentation du diagramme des contrastes par paire. La présentation en cercle est moins révélatrice des contrastes que la présentation en réseau mais évite que les lignes ne se chevauchent.

Chapitre 9. Analyse log-linéaire Sélection du modèle

La procédure d'analyse log-linéaire de sélection de modèle analyse les multiples tableaux croisés (tableau de contingence). Elle adapte les modèles log-linéaires hiérarchiques aux tableaux croisés multidimensionnelles à l'aide d'un algorithme itératif d'adaptation proportionnelle. Cette procédure vous permet de trouver les variables qualitatives associées. Pour construire des modèles vous disposez des méthodes d'élimination descendante et d'introduction forcée. Pour les modèles saturés, vous pouvez demander des estimations des paramètres et des tests d'association partielle. Un modèle saturé ajoute 0.5 à toutes les cellules.

Exemple : Dans une étude de préférence pour des lessives, les chercheurs ont compté les individus de chaque groupe, en combinant plusieurs catégories de douceur de l'eau (douce, moyenne et dure), utilisation antérieure d'une des marques et température de lavage (à froid ou à chaud). Ils ont découvert que la température est liée à la douceur de l'eau et également à la préférence d'une marque.

Statistiques : Fréquences, résidus, estimations des paramètres, erreurs standard, intervalles de confiance et tests d'association partielle. Pour les modèles personnalisés, tracés des résidus et de probabilités gaussiens.

Remarques sur les données de l'analyse log-linéaire de la sélection de modèle

Données : Les facteurs sont qualitatifs. Toutes les variables à analyser doivent être numériques. Les variables de chaîne qualitatives doivent être enregistrées en tant que variables numériques avant le début de l'analyse de sélection de modèle.

Évitez de spécifier trop de variables ayant de nombreux niveaux. Cela peut conduire à une situation dans laquelle plusieurs cellules contiennent un petit nombre d'observations, et les valeurs du khi-deux ne servent à rien.

Procédures apparentées : La sélection de modèle peut vous permettre d'identifier les termes nécessaires dans le modèle. Vous pouvez alors poursuivre l'évaluation du modèle à l'aide de l'analyse log-linéaire généralisée ou de l'analyse log-linéaire Logit. Vous pouvez utiliser le recodage automatique pour recoder les variables de chaîne. Si une variable numérique contient des catégories vides, utilisez le recodage pour générer des valeurs entières consécutives.

Obtenir une analyse log-linéaire de sélection de modèle

A partir des menus, sélectionnez :

Analyse > Log Linéaire > Sélection de modèle...

1. Sélectionnez un ou plusieurs facteurs numériques qualitatifs.
2. Sélectionnez des facteurs dans la liste Facteur(s), puis cliquez sur **Définir plage**.
3. Définissez la plage de valeurs pour chaque facteur.
4. Sélectionnez une option dans le groupe Génération de modèle.

Vous pouvez éventuellement sélectionner une variable de pondération de cellule pour spécifier des zéros structurels.

Analyse log-linéaire : Définir plage

Vous devez indiquer la plage des catégories pour tout facteur. Les valeurs Minimum et Maximum correspondent aux catégories supérieures et inférieures du facteur. Ces deux valeurs doivent être des nombres entiers, et la valeur minimale doit être inférieure à la valeur maximale. Les observations dont les valeurs se trouvent à l'extérieur des limites sont exclues. Par exemple, si vous spécifiez une valeur minimum de 1 et une valeur maximum de 3, seules les valeurs 1, 2, et 3 sont utilisées. Répétez ce processus pour chaque facteur.

Modèle d'analyse log-linéaire

Spécifier le modèle : Un modèle saturé contient tous les effets factoriels principaux et toutes les interactions facteur par facteur. Sélectionnez **Autres** pour indiquer une classe génératrice pour un modèle non saturé.

Classe de génération : Une classe de génération est une liste des termes d'ordre supérieur dans lesquels les facteurs apparaissent. Un modèle hiérarchique contient les termes qui définissent la classe de génération et tous les éléments liés d'ordre inférieur. Supposons que vous sélectionniez les variables *A*, *B* et *C* dans la liste des facteurs, puis **Interaction** dans la liste déroulante Termes construits. Le modèle qui en résulte contient l'interaction triple $A*B*C$, les interactions doubles $A*B$, $A*C$ et $B*C$, et les effets principaux *A*, *B* et *C*. N'indiquez pas les éléments liés d'ordre inférieur dans la classe génératrice.

Termes construits

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau de toutes les variables sélectionnées. Il s'agit de la valeur par défaut.

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.

Options de l'analyse log-linéaire de sélection de modèle

Afficher : Vous pouvez choisir **Fréquences** ou **Résidus**, ou les deux. Dans un modèle saturé, les fréquences observées et théoriques sont égales ; les résidus sont nuls.

Tracé : Pour les modèles personnalisés, vous pouvez choisir le type de tracés **Résidus** ou **Graphique de répartition gaussien**, ou les deux. Ils vous permettent de déterminer si un modèle est adapté aux données.

Afficher pour modèle saturé : Pour obtenir un modèle saturé, vous pouvez sélectionner **Estimations des paramètres**. Les estimations peuvent vous permettre de déterminer les termes du modèle qui peuvent être abandonnés. Un tableau d'association, qui énumère les tests d'association partielle, est également à votre disposition. Cette option est fastidieuse du point de vue des calculs pour les tables contenant de nombreux facteurs.

Critères du modèle : Un algorithme itératif d'adaptation proportionnelle est utilisé pour obtenir les estimations de paramètres. Vous pouvez ignorer des critères d'estimation. Pour cela, indiquez **Maximum des itérations**, **Convergence** ou **Delta** (valeur ajoutée à toutes les fréquences de cellules pour les modèles saturés).

Fonctions supplémentaires de la commande HILOGLINEAR

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

- Spécifier la pondération des cellules sous forme de matrices (à l'aide de la sous-commande CWEIGHT).
- Générer des analyses de plusieurs modèles avec une seule commande (à l'aide de la sous-commande DESIGN).

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

Chapitre 10. Analyse log-linéaire généralisée

La procédure d'analyse log-linéaire généralisée analyse l'effectif de fréquences d'observations tombant dans chaque catégorie de la classification croisée dans un tableau croisé ou une table de contingence. Chaque classification croisée dans le tableau constitue une cellule et chaque variable qualitative est appelée un facteur. La variable dépendante est le nombre d'observations (fréquence) dans une cellule du tableau croisé et les variables explicatives sont des facteurs ou des covariables. Cette procédure estime les paramètres de vraisemblance maximum des modèles log-linéaires hiérarchiques ou non hiérarchiques à l'aide de la méthode de Newton-Raphson. La distribution de Poisson et la distribution multinomiale peuvent être analysées en même temps.

Vous pouvez sélectionner jusqu'à 10 facteurs pour définir les cellules d'une table. Une variable structurée en cellule vous permet de définir des zéros structurels pour les tables incomplètes, notamment un terme de décalage dans le modèle, adapter un modèle logarithmique ou mettre en oeuvre la méthode d'ajustement des tableaux marginaux. Les variables de contraste permettent le calcul de log rapports des cotes généralisés.

Les informations du modèle et les statistiques de qualité d'ajustement s'affichent automatiquement. Vous pouvez également afficher un grand nombre de statistiques et de tracés ou enregistrer les résidus et les prévisions dans le jeu de données actif.

Exemple : Les données d'un rapport sur les accidents de voiture en Floride servent à déterminer la relation entre le port de la ceinture de sécurité et la mortalité suite à des blessures. Le rapport des cotes permet de mettre à jour l'existence d'une relation.

Statistiques : Fréquences observées et théoriques, résidus naturels, ajustés et au sens déviance, matrice du plan, estimations des paramètres, rapport des cotes, log rapport des cotes généralisé, GLOR, statistiques de Wald et intervalles de confiance. Tracés : résidus ajustés, résidus au sens déviance et probabilité normale.

Remarques sur les données de l'analyse log-linéaire généralisée

Données : Les facteurs sont qualitatifs, et les covariables dans les cellules sont continues. Lorsqu'une covariable est dans le modèle, la valeur moyenne de la covariable pour les observations est appliquée à une cellule. Les variables de contraste sont continues. Elles servent à calculer des log rapports des cotes généralisés. Les valeurs de la variable de contraste sont les coefficients pour la combinaison linéaire des logs des effectifs de cellule attendus.

Une variable de structure de cellule affecte des pondérations. Par exemple, si certaines des cellules sont des zéros structurels, la variable de structure de cellule a la valeur 0 ou 1. N'utilisez pas une variable de structure de cellule pour pondérer des données agrégées. Choisissez plutôt **Pondérer les observations** dans le menu Données.

Hypothèses : Deux distributions sont disponibles dans l'analyse log-linéaire généralisée : Poisson et Multinomiale.

Dans le cas d'une distribution de Poisson :

- La taille de l'échantillon complet n'est pas fixée avant l'étude ou l'analyse n'est pas conditionnelle sur la taille de l'échantillon complet.
- L'événement d'une observation qui se trouve dans une cellule est indépendante d'un point de vue statistique de l'effectif des autres cellules.

Dans le cas d'une distribution multinomiale :

- La taille de l'échantillon complet est fixée ou l'analyse est conditionnelle sur tout l'échantillon.
- Les effectifs des cellules ne sont pas indépendants d'un point de vue statistique.

Procédures apparentées : Utilisez la procédure des tableaux croisés pour étudier les tableaux croisés. Utilisez l'analyse log-linéaire Logit lorsqu'il est naturel de considérer une ou plusieurs des variables qualitatives comme étant des variables de réponse et les autres des variables explicatives.

Obtenir une analyse log-linéaire généralisée

1. A partir des menus, sélectionnez :

Analyse > Log Linéaire > Généralisée...

2. Dans la boîte de dialogue Analyse log-linéaire généralisée, sélectionnez jusqu'à 10 facteurs.

Sinon, vous pouvez :

- Sélectionner des covariables de cellule.
- Sélectionner une variable de structure de cellule pour définir des zéros structurels ou inclure un terme de décalage.
- Sélectionner une variable de contraste.

Modèle d'analyse log-linéaire généralisée

Spécifier le modèle : Un modèle saturé contient tous les effets principaux et interactifs impliquant des facteurs. Il ne contient pas de termes de covariable. Sélectionnez **Autre** pour indiquer un sous-ensemble d'interactions ou des interactions facteur/covariable.

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

Termes du modèle : Le modèle dépend de la nature de vos données. Après avoir sélectionné **Autre**, vous pouvez choisir les effets principaux et les interactions qui présentent un intérêt pour votre analyse. Vous devez indiquer tous les termes à inclure dans 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. Il s'agit de la valeur par défaut.

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.

Options de l'analyse log-linéaire généralisée

L'analyse log-linéaire généralisée affiche les informations sur le modèle et les statistiques de qualité d'ajustement. Vous pouvez également sélectionner les options suivantes :

Afficher : Vous pouvez afficher plusieurs statistiques : Fréquences observées et théoriques par cellule, Résidus naturels et au sens déviance, Matrice du plan du modèle et Estimations des paramètres pour le modèle.

Tracé : Les tracés possibles pour les modèles personnalisés uniquement comprennent deux matrices de nuage de points (résidus ajustés ou résidus au sens déviance par rapport aux effectifs observés et théoriques). Vous pouvez également afficher les tracés de probabilité gaussiens et de répartition gaussiens des résidus pour les résidus ajustés ou les résidus de déviance.

Intervalle de confiance : L'intervalle de confiance des estimations peut être ajusté.

Critères : La méthode de Newton-Raphson permet d'obtenir les estimations selon le maximum de vraisemblance. Vous pouvez entrer de nouvelles valeurs pour les critères Maximum des Itérations, Convergence et Delta (constante ajoutée à toutes les cellules relatives aux approximations initiales). Delta reste dans les cellules pour les modèles saturés.

Analyse log-linéaire généralisée : Enregistrer

Sélectionnez les valeurs que vous souhaitez enregistrer en tant que nouvelles variables dans le jeu de données actif. Le suffixe *n* dans le nom des nouvelles variables est incrémenté pour rendre ce nom univoque pour chaque nouvelle variable enregistrée.

Les valeurs enregistrées font référence à des données agrégées (cellules dans la table de contingence), même si les données sont enregistrées dans différentes observations dans l'éditeur de données. Si vous enregistrez les résidus ou les prévisions pour les données agrégées, la valeur enregistrée pour une cellule dans le tableau de contingence est entrée dans l'éditeur de données pour chaque observation de la cellule. Pour rendre les valeurs enregistrées cohérentes, vous devriez agréger les données pour obtenir les effectifs de cellule.

Vous pouvez enregistrer quatre types de résidus : brut, standardisé, ajusté et déviance. Vous pouvez également enregistrer les prévisions.

- *Résidus*. Appelé également résidu simple ou brut, il indique la différence entre l'effectif observé et l'effectif théorique de la cellule.
- *Résidus standardisés*. Résidu, divisé par une estimation de son erreur standard. Les résiduels standardisés sont aussi appelés résiduels de Pearson.
- *Résidus ajustés*. Résidu normalisé divisé par l'estimation de son erreur standard. Comme les résidus ajustés sont asymptotiquement gaussiens, ils sont préférés aux résidus standardisés pour étudier la normalité, lorsque le modèle est correct.
- *Résidus au sens déviance*. Racine carrée signée de la contribution individuelle à la statistique du rapport de vraisemblance (G au carré), où le signe est celui du résidu (effectif observé moins effectif théorique). Les résidus au sens déviance ont une distribution normale standard asymptotique.

Fonctions supplémentaires de la commande GENLOG

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

- Calculer les combinaisons linéaires des fréquences de cellule observées et théoriques, puis imprimer les résidus standardisés ou les résidus ajustés, de cette combinaison (à l'aide de la sous-commande GERESID).
- Changer la valeur seuil par défaut pour le contrôle de redondance (à l'aide de la sous-commande CRITERIA).
- Afficher les résidus standardisés (à l'aide de la sous-commande PRINT).

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

Chapitre 11. Analyse log-linéaire Logit

L'analyse log-linéaire Logit analyse la relation entre des variables dépendantes (ou de réponses) et des variables indépendantes (ou explicatives). Les variables dépendantes doivent toujours être qualitatives tandis que les variables indépendantes peuvent l'être ou pas (facteur). Les autres variables indépendantes (covariables de cellule) peuvent être continues mais elles ne s'appliquent pas à une base par observation. La moyenne pondérée des covariables pour une cellule est appliquée à cette cellule. Le logarithme de la probabilité des variables dépendantes est exprimé sous la forme d'une combinaison linéaire des paramètres. On part automatiquement sur le principe d'une distribution multinomiale. Ces modèles sont parfois appelés modèles Logit multinomiaux. Cette procédure estime les paramètres des modèles log-linéaires Logit à l'aide de l'algorithme de Newton-Raphson.

Vous pouvez sélectionner de 1 à 10 variables dépendantes et facteurs combinés. Une variable structurée en cellule vous permet de définir des zéros structurels pour les tables incomplètes, notamment un terme de décalage dans le modèle, adapter un modèle logarithmique ou mettre en oeuvre la méthode d'ajustement des tableaux marginaux. Les variables de contraste permettent le calcul de log rapports des cotes généralisés. Les valeurs de la variable de contraste sont les coefficients pour la combinaison linéaire des logs des effectifs de cellule attendus.

Les informations du modèle et les statistiques de qualité d'ajustement s'affichent automatiquement. Vous pouvez également afficher un grand nombre de statistiques et de tracés ou enregistrer les résidus et les prévisions dans le jeu de données actif.

Exemple : Une étude menée en Floride porte sur 219 alligators. Dans quelle mesure le régime alimentaire des alligators varie-t-il avec leur taille et leur habitat (un des 4 lacs) ? L'étude a montré que la probabilité qu'un petit alligator préfère les reptiles au poisson est 0,70 fois moindre que pour les alligators plus gros. La probabilité de sélection des reptiles en premier lieu plutôt que le poisson est supérieure dans le lac 3.

Statistiques : Fréquences observées et théoriques, résidus naturels, ajustés et au sens déviance, matrice du plan, estimations des paramètres, log rapport des cotes généralisé, statistiques de Wald et intervalle de confiance. Tracés : résidus ajustés, résidus au sens déviance et tracés de probabilités gaussiens.

Remarques sur les données de l'analyse log-linéaire Logit

Données : Les variables dépendantes sont qualitatives. Les facteurs sont qualitatifs. Les covariables peuvent être continues, mais lorsqu'une covariable est dans le modèle, la valeur moyenne de la covariable pour les observations d'une cellule est appliquée à cette cellule. Les variables de contraste sont continues. Elles servent à calculer les log rapports des cotes généralisés (GLOR). Les valeurs de la variable de contraste sont les coefficients pour la combinaison linéaire des logs des effectifs de cellule attendus.

Une variable de structure de cellule affecte des pondérations. Par exemple, si certaines des cellules sont des zéros structurels, la variable de structure de cellule a la valeur 0 ou 1. N'utilisez pas une variable de structure de cellule pour pondérer des données agrégées. Utilisez plutôt Observations pondérées du menu Données.

Hypothèses : Les effectifs au sein de chaque combinaison de catégories de variables explicatives sont considérées comme ayant une distribution multinomiale. Dans le cas d'une distribution multinomiale :

- La taille de l'échantillon complet est fixée ou l'analyse est conditionnelle sur tout l'échantillon.
- Les effectifs des cellules ne sont pas indépendants d'un point de vue statistique.

Procédures apparentées : Utilisez la procédure de tableaux croisés pour afficher les tableaux de contingence. Utilisez l'analyse log-linéaire généralisée lorsque vous souhaitez analyser la relation entre l'effectif observé et un ensemble de variables explicatives.

Obtenir une analyse log-linéaire Logit

1. A partir des menus, sélectionnez :
Analyse > Log Linéaire > Logit...
2. Dans la boîte de dialogue Analyse log-linéaire Logit, sélectionnez une ou plusieurs variables dépendantes.
3. Sélectionnez un ou plusieurs facteurs.

Le nombre total de variables dépendantes et de facteurs doit être inférieur ou égal à 10.

Sinon, vous pouvez :

- Sélectionner des covariables de cellule.
- Sélectionner une variable de structure de cellule pour définir des zéros structurels ou inclure un terme de décalage.
- Sélectionner une ou plusieurs variables de contraste.

Modèle d'analyse log-linéaire Logit

Spécifier le modèle : Un modèle saturé contient tous les effets principaux et interactifs impliquant des facteurs. Il ne contient pas de termes de covariable. Sélectionnez **Autre** pour indiquer un sous-ensemble d'interactions ou des interactions facteur/covariable.

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

Termes du modèle : Le modèle dépend de la nature de vos données. Après avoir sélectionné **Autre**, vous pouvez choisir les effets principaux et les interactions qui présentent un intérêt pour votre analyse. Vous devez indiquer tous les termes à inclure dans le modèle.

Les termes sont ajoutés plan. Pour cela, on prend toutes les combinaisons possibles de termes dépendants et on compare chaque combinaison avec chaque terme de la liste du modèle. Si vous sélectionnez **Inclure constante pour variable dépendante**, un terme unitaire (1) est ajouté à la liste du modèle.

Par exemple, supposons que les variables $D1$ et $D2$ soient dépendantes. La liste des termes dépendants est créée par l'analyse log-linéaire Logit ($D1, D2, D1*D2$). Si la liste Termes du modèle contient $M1$ et $M2$, et qu'une constante est incluse, la liste du modèle contient 1, $M1$ et $M2$. Le plan résultant inclut les combinaisons de chaque terme du modèle avec chaque terme dépendant :

$D1, D2, D1*D2$

$M1*D1, M1*D2, M1*D1*D2$

$M2*D1, M2*D2, M2*D1*D2$

Inclure terme constant pour variable dépendante : Inclut une constante pour la variable dépendante dans un modèle personnalisé.

Termes construits

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau de toutes les variables sélectionnées. Il s'agit de la valeur par défaut.

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.

Options de l'analyse log-linéaire Logit

L'analyse log-linéaire Logit affiche les informations sur le modèle et les statistiques de qualité d'ajustement. Vous pouvez également choisir une ou plusieurs des options suivantes :

Afficher : Vous pouvez afficher plusieurs statistiques : fréquences observées et théoriques par cellule, résidus bruts, ajustés et au sens déviance, matrice du plan du modèle et estimations des paramètres pour le modèle.

Tracé : Les tracés possibles comprennent deux matrices de nuage de points (Résidus ajustés ou Résidus au sens déviance par rapport aux effectifs observés et théoriques). Vous pouvez également afficher les tracés de probabilité gaussiens et de répartition gaussiens des résidus pour les résidus ajustés ou les résidus de déviance.

Intervalle de confiance : L'intervalle de confiance des estimations peut être ajusté.

Critères : La méthode de Newton-Raphson permet d'obtenir les estimations selon le maximum de vraisemblance. Vous pouvez entrer de nouvelles valeurs pour les critères Maximum des Itérations, Convergence et Delta (constante ajoutée à toutes les cellules relatives aux approximations initiales). Delta reste dans les cellules pour les modèles saturés.

Analyse log-linéaire Logit : Enregistrer

Sélectionnez les valeurs que vous souhaitez enregistrer en tant que nouvelles variables dans le jeu de données actif. Le suffixe *n* dans le nom des nouvelles variables est incrémenté pour rendre ce nom univoque pour chaque nouvelle variable enregistrée.

Les valeurs enregistrées se réfèrent à des données agrégées (cellules dans le tableau de contingence), même si les données sont enregistrées dans différentes observations dans l'éditeur de données. Si vous enregistrez les résidus ou les prévisions pour les données agrégées, la valeur enregistrée pour une cellule dans le tableau de contingence est entrée dans l'éditeur de données pour chaque observation de la cellule. Pour rendre les valeurs enregistrées cohérentes, vous devriez agréger les données pour obtenir les effectifs de cellule.

Vous pouvez enregistrer quatre types de résidus : brut, standardisé, ajusté et déviance. Vous pouvez également enregistrer les prévisions.

- *Résidus*. Appelé également résidu simple ou brut, il indique la différence entre l'effectif observé et l'effectif théorique de la cellule.
- *Résidus standardisés*. Résidu, divisé par une estimation de son erreur standard. Les résiduels standardisés sont aussi appelés résiduels de Pearson.
- *Résidus ajustés*. Résidu normalisé divisé par l'estimation de son erreur standard. Comme les résidus ajustés sont asymptotiquement gaussiens, ils sont préférés aux résidus standardisés pour étudier la normalité, lorsque le modèle est correct.

- *Résidus au sens déviance*. Racine carrée signée de la contribution individuelle à la statistique du rapport de vraisemblance (G au carré), où le signe est celui du résidu (effectif observé moins effectif théorique). Les résidus au sens déviance ont une distribution normale standard asymptotique.

Fonctions supplémentaires de la commande GENLOG

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

- Calculer les combinaisons linéaires des fréquences de cellule observées et théoriques, puis d'imprimer les résidus standardisés ou les résidus ajustés, de cette combinaison (à l'aide de la sous-commande GERESID).
- Changer la valeur seuil par défaut pour le contrôle de redondance (à l'aide de la sous-commande CRITERIA).
- Afficher les résidus standardisés (à l'aide de la sous-commande PRINT).

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

Chapitre 12. Tables de survie

Il arrive souvent que vous souhaitiez examiner la distribution des temps entre deux événements, telle que l'ancienneté (durée entre la date d'embauche et le départ de l'entreprise). Cependant ce type de données inclut généralement certaines observations pour lesquelles le second événement n'est pas enregistré, par exemple pour les employées toujours en activité. Il existe plusieurs raisons à cela : pour certaines observations, l'événement n'a simplement pas lieu avant la fin de l'étude tandis que pour d'autres, nous perdons la trace de leur statut avant la fin de l'étude. Certaines observations, enfin, ne peuvent plus être poursuivies pour des raisons indépendantes de l'étude (employés en congé maladie, etc.). Collectivement, on appelle ces observations des **observations censurées**. Elles rendent ce type d'étude inapproprié aux techniques traditionnelles comme les tests t et la régression linéaire.

Une technique statistique utile pour ce type de données est appelée une **table de survie** de suivi. L'idée de base de la table de survie est de subdiviser la période d'observation en intervalles de temps plus petits. Pour chaque intervalle, tous les individus observés au moins sur cette durée sont utilisés dans le calcul de la probabilité de l'événement final qui se produit au cours de cet intervalle. Les probabilités estimées sur chaque intervalle sont ensuite utilisées pour estimer la probabilité globale de l'événement se produisant à différentes dates.

Exemple : Un patch à la nicotine est-il plus efficace que le traitement par patch traditionnel pour aider des patients à arrêter de fumer ? Vous pouvez mener une étude à l'aide de deux groupes de patients fumeurs, l'un recevant une thérapie classique et l'autre le traitement expérimental. La construction de la table de survie vous permet de comparer les taux d'abstinence globale entre les deux groupes pour déterminer si le traitement expérimental constitue un progrès par rapport à la thérapie classique. Vous pouvez également tracer graphiquement la survie ou les fonctions de risque pour obtenir des informations plus détaillées.

Statistiques : Nombre d'entrants, nombre de sortants, nombre d'individus exposés au risque, nombre d'événements finals, proportion d'individus terminant, proportion d'individus survivants, proportion de survivants cumulés (et erreur standard), densité de probabilité (et erreur standard), taux de hasard (et erreur standard) pour chaque intervalle de temps et pour chaque groupe, durée de survie médiane pour chaque groupe et test de Wilcoxon (Gehan) pour comparer les répartitions de survie entre les groupes. Tracés : tracés de fonction pour survie, survie sur échelle log, densité, taux de hasard et un moins survie.

Remarques sur les données de Tables de survie

Données : Votre variable de temps doit être quantitative. Votre variable de statut doit être dichotomique ou qualitative, codée sous forme de nombres entiers, les événements étant codés sous la forme d'une valeur unique ou d'une plage de valeurs consécutives. Les facteurs doivent être qualitatifs, codés sous forme de nombres entiers.

Hypothèses : Les probabilités pour un événement donné doivent dépendre seulement du temps après l'événement initial. On considère qu'elles sont stables par rapport au temps absolu. Cela signifie que les observations entrant dans l'étude à des dates différentes (par exemple, les patients qui commencent le traitement à des dates différentes) doivent avoir le même comportement. Il ne doit pas y avoir non plus de différences systématiques entre les observations censurées et les autres. Si, par exemple, la plupart des observations censurées concernent des patients dont le cas est plus grave, les résultats seront faussés.

Procédures apparentées : La procédure de tables de survie utilise une approche actuarielle de ce type d'analyse (généralement connue sous le nom d'analyse de survie). La procédure d'analyse de survie de Kaplan-Meier utilise une méthode légèrement différente de calcul de la table de survie qui ne se fonde pas sur une partition de la période d'observation en intervalles de temps plus courts. Cette méthode est recommandée si vous ne disposez que d'un petit nombre d'observations, de telle sorte que chaque

intervalle de durée de survie ne contiendrait qu'un petit nombre d'observations. Si vous pensez que vos variables sont liées à la durée de survie ou à des variables que vous souhaitez contrôler (covariables), utilisez la procédure de la régression de Cox. Si vos covariables sont susceptibles d'avoir des valeurs différentes à différentes dates pour la même observation, utilisez la régression de Cox avec prédicteurs chronologiques.

Créer des tables de survie

1. A partir des menus, sélectionnez :
Analyse > Survie > Tables de survie...
2. Sélectionnez une variable de survie *numérique*.
3. Spécifiez les intervalles de temps à étudier.
4. Sélectionnez une variable de statut pour définir les observations pour lesquelles l'événement final a eu lieu.
5. Cliquez sur **Définir l'événement** pour indiquer la valeur de la variable de statut qui indique qu'un événement s'est produit.

Vous pouvez également sélectionner un facteur du premier ordre. Des tables actuarielles de la variable de survie sont générées pour chaque facteur.

Vous pouvez aussi sélectionner *un facteur* de deuxième ordre. Des tables actuarielles de la variable de survie sont générées pour chaque combinaison des facteurs du premier et du second ordre.

Tables de survie : Définir l'événement pour la variable de statut

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

Tables de survie : Définir plage

Les observations qui ont des valeurs pour le facteur dans la plage que vous avez spécifié sont incluses dans l'analyse. Des tables distinctes (ainsi que des tracés, si vous le demandez) sont générées pour chaque valeur de cette plage.

Tables de survie : Options

Vous pouvez contrôler plusieurs aspects de votre analyse des tables de survie.

Table(s) de survie : Pour supprimer l'affichage des tables de survie dans la sortie, désélectionnez **Table(s) de survie**.

Tracé : Vous permet de demander des tracés des fonctions de survie. Si vous avez défini un ou plusieurs facteurs, des tracés sont générés pour chaque sous-groupe défini par ces facteurs. Les tracés disponibles sont *Survie*, *Survie sur échelle log*, *Risque*, *Densité*, et *Un moins survie*.

- *Survie*. Affiche la fonction de survie cumulée d'après une échelle linéaire.
- *Survie sur échelle log*. Affiche la fonction de survie cumulée d'après une échelle logarithmique.
- *Hasard*. Affiche la fonction de risque cumulée d'après une échelle linéaire.
- *Densité*. Affiche la fonction de densité.
- *Un moins survie*. Trace un moins la fonction de survie sur une échelle linéaire.

Comparer les niveaux du premier facteur : Si vous disposez d'une variable de contrôle du premier ordre, vous pouvez sélectionner une des possibilités de ce groupe pour effectuer un test de Wilcoxon (Gehan),

qui compare la survie des sous-groupes. Les tests sont effectués sur le facteur de premier ordre. Si vous avez défini un facteur de second ordre, les tests sont effectués sur chaque niveau du facteur de second ordre.

Fonctions supplémentaires de la commande SURVIVAL

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

- Spécifier plus d'une variable dépendante.
- Spécifier des intervalles de tailles différentes.
- Spécifier plusieurs variables de statut.
- Spécifier les comparaisons qui ne tiennent pas compte de tous les facteurs et variables de contrôle.
- Calculer des comparaisons approximatives et non pas exactes.

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

Chapitre 13. Analyse de survie de Kaplan-Meier

Il arrive souvent que vous souhaitiez examiner la distribution des temps entre deux événements, telle que l'ancienneté (durée entre la date d'embauche et le départ de l'entreprise). Cependant, ce type de données inclut généralement certaines observations censurées. Les observations censurées constituent des observations pour lesquelles le second événement n'est pas enregistré (par exemple, les personnes encore en activité dans l'entreprise à la fin de l'étude). La procédure de Kaplan-Meier est une méthode d'estimation des modèles de durée à l'événement avec des observations censurées. Le modèle de Kaplan-Meier est fondé sur l'estimation de probabilités conditionnelles à chaque date à laquelle un événement se produit et sur l'utilisation de la limite du produit de ces probabilités pour estimer le taux de survie à chaque moment dans le temps.

Exemple : Un nouveau traitement contre le SIDA a-t-il des effets thérapeutiques permettant de prolonger la durée de vie des patients ? Vous pouvez mener une étude à l'aide de deux groupes de patients atteints par le SIDA, l'un recevant une thérapie classique et l'autre le traitement expérimental. La construction d'un modèle de Kaplan-Meier à partir des données vous permet de comparer tous les taux de survie entre les deux groupes pour déterminer si le traitement expérimental représente un progrès par rapport à la thérapie classique. Vous pouvez également tracer graphiquement la survie ou les fonctions de risque pour obtenir des informations plus détaillées.

Statistiques : Table de survie, incluant durée, statut, erreur standard et survie cumulée, événements cumulés et nombre restant ; moyenne et médiane de durée de survie, avec erreur standard et intervalle de confiance de 95 %. Tracés : Survie, Hasard, Survie sur échelle log et Un moins survie.

Remarques sur les données de Kaplan-Meier

Données : La variable de temps doit être continue, la variable de statut peut être qualitative ou continue, les facteurs et variables de strates doivent être qualitatifs.

Hypothèses : Les probabilités pour un événement donné doivent dépendre seulement du temps après l'événement initial. On considère qu'elles sont stables par rapport au temps absolu. Cela signifie que les observations entrant dans l'étude à des dates différentes (par exemple, les patients qui commencent le traitement à des dates différentes) doivent avoir le même comportement. Il ne doit pas y avoir non plus de différences systématiques entre les observations censurées et les autres. Si, par exemple, la plupart des observations censurées concernent des patients dont le cas est plus grave, les résultats seront faussés.

Procédures apparentées : La procédure de Kaplan-Meier utilise une méthode de calcul de la table de survie qui estime la survie ou la fonction de risque à la date de chaque événement. La procédure Tables de survie utilise une approche actuarielle de l'analyse de survie fondée sur la partition de la période d'observation en intervalles de temps plus petits. Elle peut donc être utile pour traiter de vastes échantillons. Si vous pensez que vos variables sont liées à la durée de survie ou à des variables que vous souhaitez contrôler (covariables), utilisez la procédure de la régression de Cox. Si vos covariables sont susceptibles d'avoir des valeurs différentes à différentes dates pour la même observation, utilisez la régression de Cox avec prédicteurs chronologiques.

Obtenir une analyse de survie de Kaplan-Meier

1. A partir des menus, sélectionnez :
Analyse > Survie > Kaplan-Meier...
2. Sélectionnez une variable de temps.
3. Sélectionnez une variable de statut pour identifier les observations pour lesquelles l'événement final a eu lieu. Cette variable peut être numérique ou être une *chaîne courte*. Cliquez ensuite sur **Définir l'événement**.

Vous pouvez éventuellement sélectionner un facteur pour étudier les différences entre les groupes. Vous pouvez enfin définir une variable de strates qui produira des analyses séparées pour chaque niveau (strate) de la variable.

Kaplan-Meier : Définir un événement pour la variable de statut

Entrez la ou les valeurs indiquant que l'événement final a eu lieu. Vous pouvez saisir une seule valeur, une plage ou une liste de valeurs. L'option Plage de valeurs n'est disponible que si votre variable de statut est numérique.

Kaplan-Meier : Comparer les niveaux du facteur

Vous pouvez demander des statistiques pour tester l'égalité des distributions de survie sur les différents niveaux du facteur. Les statistiques disponibles sont Log Rank (classement logarithmique), Breslow et Tarone-Ware. Sélectionnez l'une de ces options pour spécifier les comparaisons à effectuer : sur l'ensemble des strates, pour chaque strate, appariées sur les strates ou appariées sur chaque strate.

- *Log Rank*. Test de comparaison de l'égalité des distributions de survie. Tous les points temporels sont pondérés de façon égale dans ce test.
- *Breslow*. Test de comparaison de l'égalité des distributions de survie. Les points dans le temps sont pondérés par le nombre d'observations à risques à chaque point.
- *Tarone-Ware*. Test de comparaison de l'égalité des distributions de survie. Les points dans le temps sont pondérés par la racine carrée du nombre d'observations à risque, à chaque point.
- *Sur l'ensemble des strates*. Compare tous les niveaux de facteurs d'un test unique afin de tester l'égalité des courbes de survie.
- *Appariés sur les strates*. Compare chaque paire de niveaux de facteurs. Les tests de tendance par paires ne sont pas disponibles.
- *Pour chaque strate*. Effectue un test d'égalité distinct sur tous les niveaux de facteur de chaque strate. Si vous ne disposez pas d'une variable de stratification, les tests ne sont pas effectués.
- *Appariés sur chaque strate*. Compare chaque paire de niveaux de facteurs pour chaque strate. Les tests de tendance par paires ne sont pas disponibles. Si vous ne disposez pas d'une variable de stratification, les tests ne sont pas effectués.

Tendance linéaire des niveaux de facteur : Vous permet de tester la tendance linéaire des niveaux du facteur. Cette option n'est disponible que pour les comparaisons globales (et non pas appariée) des niveaux de facteur.

Kaplan-Meier : Enregistrer les nouvelles variables

Vous pouvez enregistrer les informations de votre table de Kaplan-Meier sous forme de nouvelles variables. Ces dernières peuvent ensuite servir à d'autres analyses pour tester des hypothèses ou les vérifier. Vous pouvez enregistrer la Survie, l'Erreur standard de survie, le Hasard et la Fréquence cumulée sous forme de nouvelles variables.

- *Survie*. Estimations de la probabilité de survie cumulée. Le nom de variable par défaut se compose du préfixe *sur_* suivi d'un numéro séquentiel. Si par exemple *sur_1* existe, Kaplan-Meier choisira le nom de variable *sur_2*.
- *Erreur standard de survie*. Erreur standard de l'estimation de survie cumulée. Le nom de variable par défaut se compose du préfixe *se_* suivi d'un numéro séquentiel. Si par exemple *se_1* existe, Kaplan-Meier choisira le nom de variable *se_2*.
- *Hasard*. Estimation de la fonction de risque cumulée. Le nom de variable par défaut se compose du préfixe *haz_* suivi d'un numéro séquentiel. Si par exemple *haz_1* existe, Kaplan-Meier choisira le nom de variable *haz_2*.

- *Fréquence cumulée.* Effectif cumulé des événements lorsque les observations sont triées par leur durée de survie et leurs codes de statut. Le nom de variable par défaut se compose du préfixe cum_ suivi d'un numéro séquentiel. Si par exemple cum_1 existe, Kaplan-Meier choisira le nom de variable cum_2.

Kaplan-Meier : Options

Vous pouvez demander différents types de sortie à partir de votre analyse Kaplan-Meier.

Statistiques : Vous pouvez sélectionner les statistiques affichées pour les fonctions de survie calculées, notamment les tables de survie, la moyenne et médiane de survie et les quartiles. Si vous avez inclus des facteurs, des statistiques distinctes sont générées pour chaque groupe.

Tracés : Les tracés vous permettent d'examiner visuellement les fonctions Survie, Risque, Survie sur échelle log, et Survie moins un. Si vous avez inclus des facteurs, les fonctions sont tracées graphiquement pour chaque groupe.

- *Survie.* Affiche la fonction de survie cumulée d'après une échelle linéaire.
- *Un moins survie.* Trace un moins la fonction de survie sur une échelle linéaire.
- *Hasard.* Affiche la fonction de risque cumulée d'après une échelle linéaire.
- *Survie sur échelle log.* Affiche la fonction de survie cumulée d'après une échelle logarithmique.

Fonctions supplémentaires de la commande KM

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

- Obtenir des tables de fréquences qui tiennent compte des observations hors du suivi en tant que catégorie distincte des observations censurées.
- Spécifier un espacement inégal pour le test de la tendance linéaire.
- Calculer les percentiles autres que les quartiles de la variable de durée de survie.

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

Chapitre 14. Analyse de la régression de Cox

La régression de Cox crée un modèle de prévision pour les données de la durée à l'événement. Le modèle génère une fonction de survie qui prévoit la probabilité d'occurrence de l'événement étudié à un instant t donné pour les valeurs fournies pour les variables de prédicteur. La forme de la fonction de survie et les coefficients de régression des prédicteurs sont estimés à partir des sujets observés. Le modèle peut alors être appliqué aux nouvelles observations présentant des mesures pour les variables de prédicteur. Les informations provenant de sujets censurés, à savoir celles qui ne rencontrent pas l'événement étudié au moment de l'observation, contribuent de manière utile à l'estimation du modèle.

Exemple : Les hommes et les femmes sont-ils égaux face au risque de développement du cancer du poumon lié à la consommation de tabac ? Lorsque vous construisez une régression de Cox, avec la consommation de tabac (nombre de cigarettes fumées par jour) et le sexe de l'individu entrés en tant que covariables, vous pouvez tester les hypothèses en fonction des effets du sexe et de la consommation sur le temps jusqu'à l'apparition d'un cancer du poumon.

Statistiques : Pour chaque modèle : $-2LL$, statistique du rapport de vraisemblance et khi-deux global. Pour les variables du modèle : estimations de paramètres, erreurs standard et statistiques de Wald. Pour les variables hors du modèle : statistiques de score et khi-deux résiduel.

Remarques sur les données de la régression de Cox

Données : La variable de temps doit être quantitative, et la variable de statut peut être qualitative ou continue. Les variables indépendantes (covariables) doivent être continues ou qualitatives. S'il s'agit de variables qualitatives, elles doivent être factices ou codées par indicateurs (il existe une option dans la procédure pour recoder automatiquement les variables qualitatives). Les variables de strates doivent être qualitatives, codées en tant que nombres entiers ou chaînes alphanumériques courtes.

Hypothèses : Les observations doivent être indépendantes et le taux de risque doit être constant dans le temps. Cela signifie que la proportionnalité des risques d'une observation à une autre ne doit pas varier dans le temps. Cette dernière hypothèse est connue sous le nom d'**hypothèse des risques proportionnels**.

Procédures apparentées : Si l'hypothèse des risques proportionnels n'est pas satisfaite, vous devez utiliser la régression de Cox avec prédicteur chronologique. Si vous n'utilisez pas de covariable, ou si vous ne disposez que d'une seule covariable catégorielle, vous pouvez employer les Tables de survie ou la procédure de Kaplan-Meier pour étudier les fonctions de survie ou de risque pour votre échantillon. Si vous n'avez pas de données censurées dans votre exemple (c'est-à-dire que chaque observation contient un événement final), vous pouvez utiliser la régression linéaire pour modéliser la relation entre les prédicteurs et la durée jusqu'à l'événement.

Obtenir une analyse par la régression de Cox

1. A partir des menus, sélectionnez :
Analyse > Survie > Modèle de Cox...
2. Sélectionnez une variable de temps. Les observations dont les valeurs de date sont négatives ne sont pas analysées.
3. Sélectionnez une variable de statut, puis cliquez sur **Définir l'événement**.
4. Sélectionnez une ou plusieurs covariables. Pour ajouter des termes d'interaction, sélectionnez toutes les variables impliquées dans l'interaction, puis sélectionnez **>a*b>**.

Vous pouvez également calculer des modèles distincts pour les différents groupes. Pour cela, définissez une variable de strates.

Modèle de Cox... Définir des variables qualitatives

Vous pouvez indiquer la façon dont la régression de Cox gèrera les variables qualitatives.

Covariables : Répertorie toutes les covariables indiquées dans la boîte de dialogue principale, soit seules, soit au sein d'une interaction, dans une couche. Si certaines de ces covariables sont des variables de chaîne, vous pouvez utiliser des covariables qualitatives.

Covariables qualitatives : Etablit la liste de toutes les variables identifiées comme étant qualitatives. 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 qualitatives. Sélectionnez n'importe quelle autre covariable qualitative à partir de la liste des covariables qualitatives.

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 avec 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 première catégorie) est comparée avec la catégorie précédente.
- **Polynomial :** Contraste polynomial orthogonal. On part de l'hypothèse que les catégories sont espacées de manière équivalente. Les contrastes polynomiaux sont utilisables pour les variables numériques seulement.
- **Déviation :** Chaque catégorie de la variable de prédicteur (hormis la catégorie de référence) est comparée à l'effet global.

Si vous sélectionnez **Déviation**, **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 qualitatives, vous devez supprimer tous les termes contenant cette variable de la liste des covariables de la boîte de dialogue principale.

Tracés de la régression de Cox

Les tracés permettent d'évaluer le modèle estimé et d'interpréter les résultats. Vous pouvez tracer les fonctions Survie, Risque, LN (-Logn), et Un moins survie.

- *Survie.* Affiche la fonction de survie cumulée d'après une échelle linéaire.
- *Hasard.* Affiche la fonction de risque cumulée d'après une échelle linéaire.
- **LN (-Logn) :** Estimation de survie cumulée après application de la transformation $\ln(-\ln)$ à l'estimation.
- *Un moins survie.* Trace un moins la fonction de survie sur une échelle linéaire.

Ces fonctions dépendant des valeurs des covariables, vous devez utiliser les valeurs constantes des covariables pour tracer graphiquement ces fonctions par rapport au temps. La valeur par défaut se sert

de la moyenne de chaque covariable en tant que constante mais vous pouvez également saisir vos propres valeurs pour le tracé à l'aide du groupe de contrôle Modifier la valeur.

Vous pouvez tracer une courbe distincte pour chaque valeur de la covariable qualitative. Il suffit pour cela de déplacer cette covariable dans la zone de texte Courbes distinctes. Cette option est disponible uniquement pour les covariables qualitatives indiquées par (Cat) après leur nom dans la liste Valeurs des covariables tracées à.

Modèle de Cox : Enregistrer les nouvelles variables

Vous pouvez enregistrer les différents résultats de votre analyse sous forme de nouvelles variables. Ces variables peuvent servir aux prochaines analyses pour tester ou vérifier les hypothèses.

Enregistrer les variables de modèle : Permet d'enregistrer la fonction de survie et son erreur standard, les estimations LN (-logn), la fonction de hasard, les résiduels partiels, les différences de bêta de la régression et le prédicteur linéaire $X \cdot \text{Beta}$ comme nouvelles variables.

- *Fonction de survie.* Valeur de la fonction de survie cumulée à un temps donné. Elle correspond à la probabilité de survie à cette période.
- *Fonction de survie log moins log.* Estimation de survie cumulée après application de la transformation $\ln(-\ln)$ à l'estimation.
- *Fonction de hasard.* Enregistre l'estimation de la fonction de risque cumulée (également désigné par le nom de résidu de Cox-Snell).
- *Résidus partiels.* Vous pouvez tracer les résiduels partiels en fonction de la durée de survie pour tester l'hypothèse de hasard proportionnel. Une variable est enregistrée pour chaque covariable du modèle final. Les résidus partiels ne sont utilisables qu'avec les modèles contenant au moins une covariable.
- *DfBêta(s).* Variation estimée d'un coefficient lorsqu'une observation est éliminée. Une variable est enregistrée pour chaque covariable du modèle final. Les DfBêtas ne sont utilisables qu'avec les modèles contenant au moins une covariable.
- *$X \cdot \text{Bêta}$.* Score prédicteur linéaire. Somme du produit des valeurs de covariable centrée sur la moyenne et des estimations de paramètre correspondantes, pour chaque observation.

Si vous exécutez la régression de Cox avec un prédicteur chronologique, les différences de bêta et la variable de prédicteur linéaire $X \cdot \text{Beta}$ sont les seules variables que vous pouvez enregistrer.

Exporter les informations du modèle dans un fichier XML : Les estimations de paramètres sont exportées vers le fichier spécifié au format XML. 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 la régression de Cox

Vous pouvez contrôler certains aspects de votre analyse et de sa sortie.

Estimation des paramètres : Vous pouvez obtenir des statistiques sur les paramètres du modèle, y compris des intervalles de confiance pour $\exp(B)$ et la corrélation des estimations. Vous pouvez afficher les statistiques du modèle à chaque étape ou uniquement à la dernière étape.

Probabilité pour la méthode détaillée étape par étape : Si vous avez sélectionné cette option, spécifiez la probabilité pour chaque entrée ou suppression du modèle. Une variable est entrée si le niveau de signification de son F à saisir est inférieure à la valeur Entrée ; une variable est supprimée si le niveau de signification est supérieur à la valeur Suppression. La valeur Entrée doit être inférieure à la valeur Suppression.

Maximum des itérations : Permet de spécifier le maximum d'itérations pour le modèle, ce qui contrôle la durée de la procédure pour la recherche d'une solution.

Afficher la fonction de la ligne de base : Permet d'afficher le risque de base et la survie cumulée à la moyenne des covariables. Cet affichage n'est possible que si vous avez spécifié des prédicteurs chronologiques.

Modèle de Cox : Définir un événement pour la variable de statut

Entrez la ou les valeurs indiquant que l'événement final a eu lieu. Vous pouvez saisir une seule valeur, une plage ou une liste de valeurs. L'option Plage de valeurs n'est disponible que si votre variable de statut est numérique.

Fonctions supplémentaires de la commande COXREG

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

- Obtenir des tables de fréquences qui tiennent compte des observations hors du suivi en tant que catégorie distincte des observations censurées.
- Sélectionner une catégorie de référence, autre que la première ou la dernière pour les méthodes de contraste déviation, Simple et Indicatrice.
- Spécifier des espacements inégaux entre les catégories pour la méthode de contraste polynomial.
- Spécifier les critères d'itération supplémentaires.
- Contrôler le traitement des valeurs manquantes.
- Spécifier les noms des variables enregistrées.
- Enregistrer la sortie dans un fichier de données IBM SPSS Statistics externe.
- Conserver les données pour chaque groupe partiel du fichier dans un fichier temporaire au cours du traitement. Cela peut vous permettre d'économiser des ressources mémoire lorsque vous effectuez les analyses avec de grandes masses de données. Cette fonction n'est pas disponible avec les prédicteurs chronologiques.

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

Chapitre 15. Calcul des prédicteurs chronologiques

Dans certaines situations vous souhaiteriez calculer un modèle de régression de Cox mais l'hypothèse des risques proportionnels n'est pas satisfaite. Les taux de risque varient dans le temps. Les valeurs de l'une ou plusieurs de vos covariables sont différentes à différentes dates. Dans ce cas, vous devez utiliser une version étendue du modèle de régression de Cox, vous permettant de spécifier des **prédicteurs chronologiques**.

Pour analyser un tel modèle, vous devez avant tout définir votre prédicteur chronologique. (Plusieurs prédicteurs chronologiques peuvent être spécifiés à l'aide de la syntaxe de commande.) Pour faciliter cela, vous disposez d'une variable système représentant le temps. Cette variable s'appelle $T_.$ Vous pouvez l'utiliser pour définir des prédicteurs chronologiques de deux façons :

- Si vous souhaitez tester l'hypothèse des risques proportionnels à propos d'une covariable spéciale, ou si vous souhaitez estimer un modèle de régression de Cox étendue qui autorise les risques non proportionnels, vous pouvez définir un prédicteur chronologique sous la forme d'une fonction de la variable de temps $T_.$ et de la covariable en question. Exemple : le simple produit de la variable de temps et de la covariable. Vous pouvez également définir des fonctions plus complexes. En testant la signification du coefficient du prédicteur chronologique, vous saurez si l'hypothèse des risques proportionnels est pertinente.
- Certaines variables peuvent avoir différentes valeurs à des périodes différentes sans pour autant être liées au temps (chronologiques). Dans ce cas, vous devez définir un **prédicteur chronologique segmenté** à l'aide d'une **expression logique**. Les expressions logiques prennent la valeur 1 si elles sont vraies, 0 si elles sont fausses. A l'aide d'une série d'expressions logiques, vous pouvez créer votre prédicteur chronologique à partir d'un ensemble de mesures. Par exemple, si votre pression artérielle est mesurée une fois par semaine pendant les quatre semaines de votre étude, (mesures identifiées par $PA1$ à $PA4$), vous pouvez définir votre prédicteur chronologique sous la forme $(T_ < 1) * PA1 + (T_ >= 1 \& T_ < 2) * PA2 + (T_ >= 2 \& T_ < 3) * PA3 + (T_ >= 3 \& T_ < 4) * PA4$. Notez qu'un et un seul des termes entre parenthèses est égal à 1 pour chaque cas, tandis que les autres termes sont égaux à 0. Cette fonction peut être interprétée ainsi : " Si le temps est inférieur à une semaine, utilisez $PA1$. S'il est supérieur à une semaine mais inférieur à deux, utilisez $PA2$, et ainsi de suite ".

Dans la boîte de dialogue Calculer la variable temporelle, vous pouvez utiliser des contrôles de construction de fonction pour construire l'expression pour le prédicteur chronologique ou bien vous pouvez l'entrer directement dans Expression pour la zone de texte $T_COV_.$ Notez que 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. La variable qui en résulte est nommée $T_COV_.$ et doit être incluse en tant que covariable dans votre modèle de régression de Cox.

Calculer un prédicteur chronologique

1. A partir des menus, sélectionnez :
 Analyse > Survie > Cox à prédicteur chron...
2. Entrez une expression pour le prédicteur chronologique.
3. Sélectionnez **Modèle** pour effectuer votre régression de Cox.

Remarque : Vérifiez que vous incluez bien la nouvelle variable $T_COV_.$ en tant que covariable du modèle de régression de Cox.

Pour plus d'informations, voir Chapitre 14, «Analyse de la régression de Cox», à la page 95.

Régression de Cox munie de fonctions supplémentaires pour les prédictors chronologiques

Le langage de syntaxe de commande vous permet également de spécifier plusieurs prédictors chronologiques. D'autres fonctions de la syntaxe de commande sont disponibles pour la régression de Cox avec ou sans prédictor chronologique.

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

Chapitre 16. Méthodes de codification des variables qualitatives

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.

Déviatation

Déviatation 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 déviatation 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(1,20,3)

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 que 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(1,20,3)

Toutefois, si le dosage administré au deuxième groupe est le quadruple de celui administré au premier groupe, et que 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 :

$$\begin{array}{l} \text{mean} \quad (1/k \quad 1/k \quad 1/k \quad \dots \quad 1/k \quad 1/k) \\ \text{df}(1) \quad (\quad 1 \quad -1 \quad 0 \quad \dots \quad 0 \quad 0) \\ \text{df}(2) \quad (\quad 0 \quad 1 \quad -1 \quad \dots \quad 0 \quad 0) \\ \quad \quad \cdot \\ \quad \quad \cdot \\ \text{df}(k-1) \quad (\quad 0 \quad 0 \quad 0 \quad \dots \quad 1 \quad -1) \end{array}$$

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 :

$$\begin{array}{l} (1/4 \quad 1/4 \quad 1/4 \quad 1/4) \\ (\quad 1 \quad -1 \quad 0 \quad 0) \\ (\quad 0 \quad 1 \quad -1 \quad 0) \\ (\quad 0 \quad 0 \quad 1 \quad -1) \end{array}$$

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 )
```

La somme de chaque ligne, à l'exception de la ligne des moyennes, est égale à 0, de même que celle des produits de chaque paire de lignes disjointes :

```
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.

Chapitre 17. Structures de covariance

Cette section fournit des informations supplémentaires sur les structures de covariance.

Antédépendance : Premier ordre : Cette structure de covariance comporte des variances et des corrélations hétérogènes entre les éléments adjacents. La corrélation entre deux éléments non adjacents est le produit des corrélations entre les éléments compris entre les éléments étudiés.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_1$	$\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_4\sigma_1\rho_1\rho_2\rho_3$)
$(\sigma_2\sigma_1\rho_1$	σ_2^2	$\sigma_3\sigma_2\rho_2$	$\sigma_4\sigma_2\rho_2\rho_3$)
$(\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_3\sigma_2\rho_2$	σ_3^2	$\sigma_4\sigma_3\rho_3$)
$(\sigma_4\sigma_1\rho_1\rho_2\rho_3$	$\sigma_4\sigma_2\rho_2\rho_3$	$\sigma_4\sigma_3\rho_3$	σ_4^2)

AR(1) : Il s'agit d'une structure autorégressive d'ordre 1 comportant des variances homogènes. La corrélation entre deux éléments est égale à rho pour les éléments adjacents, à rho² pour les éléments séparés par un troisième, etc. La valeur est restreinte pour être comprise entre -1 << 1.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$	$\sigma^2\rho^3$)
$(\sigma^2\rho$	σ^2	$\sigma^2\rho$	$\sigma^2\rho^2$)
$(\sigma^2\rho^2$	$\sigma^2\rho$	σ^2	$\sigma^2\rho$)
$(\sigma^2\rho^3$	$\sigma^2\rho^2$	$\sigma^2\rho$	σ^2)

AR(1) : Hétérogène : Il s'agit d'une structure autorégressive d'ordre 1 comportant des variances hétérogènes. La corrélation entre deux éléments est égale à r pour les éléments adjacents, r² pour les éléments séparés par un troisième, etc. La valeur est restreinte pour être comprise entre -1 et 1.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho^2$	$\sigma_4\sigma_1\rho^3$)
$(\sigma_2\sigma_1\rho$	σ_2^2	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho^2$)
$(\sigma_3\sigma_1\rho^2$	$\sigma_3\sigma_2\rho$	σ_3^2	$\sigma_4\sigma_3\rho$)
$(\sigma_4\sigma_1\rho^3$	$\sigma_4\sigma_2\rho^2$	$\sigma_4\sigma_3\rho$	σ_4^2)

ARMA(1,1) : Il s'agit d'une structure de moyenne mobile autorégressive d'ordre 1. Elle comporte des variances homogènes. La corrélation entre deux éléments est égale à * pour les éléments adjacents, à *(²) pour les éléments séparés par un troisième, etc. Leurs valeurs sont comprises dans un intervalle compris entre -1 et 1.

$(\sigma^2$	$\sigma^2\phi\rho$	$\sigma^2\phi\rho^2$	$\sigma^2\phi\rho^3$)
$(\sigma^2\phi\rho$	σ^2	$\sigma^2\phi\rho$	$\sigma^2\phi\rho^2$)
$(\sigma^2\phi\rho^2$	$\sigma^2\phi\rho$	σ^2	$\sigma^2\phi\rho$)
$(\sigma^2\phi\rho^3$	$\sigma^2\phi\rho^2$	$\sigma^2\phi\rho$	σ^2)

Symétrie composée : Cette structure comporte une variance et une covariance constante.

$(\sigma^2 + \sigma_1^2$	σ_1	σ_1	σ_1)
$(\sigma_1$	$\sigma^2 + \sigma_1^2$	σ_1	σ_1)
$(\sigma_1$	σ_1	$\sigma^2 + \sigma_1^2$	σ_1)
$(\sigma_1$	σ_1	σ_1	$\sigma^2 + \sigma_1^2$)

Symétrie Composée : Métrique de corrélation : Cette structure de covariance comporte des variances et des corrélations homogènes entre les éléments.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$
$(\sigma^2\rho$	σ^2	$\sigma^2\rho$	$\sigma^2\rho$
$(\sigma^2\rho$	$\sigma^2\rho$	σ^2	$\sigma^2\rho$
$(\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$	σ^2)

Symétrie Composée : Hétérogène : Cette structure de covariance comporte des variances hétérogènes et des corrélations constantes entre les éléments.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho$	$\sigma_4\sigma_1\rho$)
$(\sigma_2\sigma_1\rho$	σ_2^2	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho$)
$(\sigma_3\sigma_1\rho$	$\sigma_3\sigma_2\rho$	σ_3^2	$\sigma_4\sigma_3\rho$)
$(\sigma_4\sigma_1\rho$	$\sigma_4\sigma_2\rho$	$\sigma_4\sigma_3\rho$	σ_4^2)

Diagonale : Cette structure de covariance comporte des variances hétérogènes entre les éléments, mais aucune corrélation.

$(\sigma_1^2$	0	0	0)
(0	σ_2^2	0	0)
(0	0	σ_3^2	0)
(0	0	0	σ_4^2)

Analytique Factorielle : Premier ordre : Cette structure de covariance possède des variances hétérogènes qui sont composées d'un terme hétérogène entre les éléments et d'un terme homogène entre les éléments. La covariance entre deux éléments est égale à la racine carrée du produit de leurs termes de variance hétérogène.

$(\lambda_1^2 + d$	$\lambda_2\lambda_1$	$\lambda_3\lambda_1$	$\lambda_4\lambda_1$)
$(\lambda_2\lambda_1$	$\lambda_2^2 + d$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2$)
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d$	$\lambda_4\lambda_3$)
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d$)

Analytique Factorielle : Premier Ordre, hétérogène : Cette structure de covariance comporte des variances hétérogènes composées de deux termes hétérogènes pour tous les éléments. La covariance entre deux éléments est égale à la racine carrée du produit du premier de leurs termes de variance hétérogène.

$(\lambda_1^2 + d_1$	$\lambda_2\lambda_1$	$\lambda_3\lambda_1$	$\lambda_4\lambda_1$)
$(\lambda_2\lambda_1$	$\lambda_2^2 + d_2$	$\lambda_3\lambda_2$	$\lambda_4\lambda_2$)
$(\lambda_3\lambda_1$	$\lambda_3\lambda_2$	$\lambda_3^2 + d_3$	$\lambda_4\lambda_3$)
$(\lambda_4\lambda_1$	$\lambda_4\lambda_2$	$\lambda_4\lambda_3$	$\lambda_4^2 + d_4$)

Huynh-Feldt : Il s'agit d'une matrice « circulaire » dans laquelle la covariance entre deux éléments est égale à la moyenne de leurs variances moins une constante. Ni les variances, ni les covariances ne sont constantes.

$(\sigma_1^2$	$[\sigma_1^2 + \sigma_2^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_1^2 + \sigma_4^2]/2 - \lambda$)
$([\sigma_1^2 + \sigma_2^2]/2 - \lambda$	σ_2^2	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda$)
$([\sigma_1^2 + \sigma_3^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_3^2]/2 - \lambda$	σ_3^2	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda$)
$([\sigma_1^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_2^2 + \sigma_4^2]/2 - \lambda$	$[\sigma_3^2 + \sigma_4^2]/2 - \lambda$	σ_4^2)

Identité codée : Cette structure comporte une variance constante. On considère qu'aucune corrélation n'existe entre les éléments.

$(\sigma^2$	0	0	0)
(0	σ^2	0	0)
(0	0	σ^2	0)
(0	0	0	σ^2)

Toeplitz : Cette structure de covariance comporte des variances et des corrélations hétérogènes entre les éléments. La corrélation entre les éléments adjacents est homogène pour toutes les paires d'éléments adjacents. La corrélation entre deux éléments séparés par un troisième est également homogène, etc.

$(\sigma^2$	$\sigma^2\rho_1$	$\sigma^2\rho_2$	$\sigma^2\rho_3$)
$(\sigma^2\rho_1$	σ^2	$\sigma^2\rho_1$	$\sigma^2\rho_2$)
$(\sigma^2\rho_2$	$\sigma^2\rho_1$	σ^2	$\sigma^2\rho_1$)
$(\sigma^2\rho_3$	$\sigma^2\rho_2$	$\sigma^2\rho_1$	σ^2)

Toeplitz : Hétérogène : Cette structure de covariance comporte des variances et des corrélations hétérogènes entre les éléments. La corrélation entre les éléments adjacents est homogène pour toutes les paires d'éléments adjacents. La corrélation entre deux éléments séparés par un troisième est également homogène, etc.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_1$	$\sigma_3\sigma_1\rho_2$	$\sigma_4\sigma_1\rho_3$)
$(\sigma_2\sigma_1\rho_1$	σ_2^2	$\sigma_3\sigma_2\rho_1$	$\sigma_4\sigma_2\rho_2$)
$(\sigma_3\sigma_1\rho_2$	$\sigma_3\sigma_2\rho_1$	σ_3^2	$\sigma_4\sigma_3\rho_1$)
$(\sigma_4\sigma_1\rho_3$	$\sigma_4\sigma_2\rho_2$	$\sigma_4\sigma_3\rho_1$	σ_4^2)

Sans structure : Il s'agit d'une matrice de covariance générale complète.

$(\sigma_1^2$	$\sigma_{2\ 1}$	σ_{31}	σ_{41})
$(\sigma_{2\ 1}$	σ_2^2	σ_{32}	$\sigma_{4\ 2}$)
$(\sigma_{31}$	σ_{32}	σ_3^2	$\sigma_{4\ 3}$)
$(\sigma_{41}$	$\sigma_{4\ 2}$	$\sigma_{4\ 3}$	σ_4^2)

Non structuré : Métrique de corrélation : Cette structure de covariance comporte des variances et des corrélations hétérogènes.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_{21}$	$\sigma_3\sigma_1\rho_{31}$	$\sigma_4\sigma_1\rho_{41}$)
$(\sigma_2\sigma_1\rho_{21}$	σ_2^2	$\sigma_3\sigma_2\rho_{32}$	$\sigma_4\sigma_2\rho_{42}$)
$(\sigma_3\sigma_1\rho_{31}$	$\sigma_3\sigma_2\rho_{32}$	σ_3^2	$\sigma_4\sigma_3\rho_{43}$)
$(\sigma_4\sigma_1\rho_{41}$	$\sigma_4\sigma_2\rho_{42}$	$\sigma_4\sigma_3\rho_{43}$	σ_4^2)

Composantes de variance : Cette structure affecte une structure d'identité mise à l'échelle (ID) à chacun des effets aléatoires indiqués.

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. Toutefois, il peut être nécessaire de posséder une copie du produit ou de la 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. Pour plus de détails, référez-vous aux documents d'annonce disponibles dans votre pays, ou adressez-vous à votre partenaire commercial IBM. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service 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 mentionnés dans le présent document. La remise de ce document ne vous donne 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-NC119
Armonk, NY 10504-1785
U.S.A*

Pour le Canada, veuillez adresser votre courrier à :

*IBM Director of Commercial Relations
IBM Canada Ltd.
3600 Steeles Avenue East
Markham, Ontario
L3R 9Z7 Canada*

Les informations sur les licences concernant les produits 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-ku
Tokyo 103-8510, Japan*

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. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, modifier les produits et 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-NC119
Armonk, NY 10504-1785
U.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 s'y rapportant sont fournis par IBM conformément aux dispositions du Livret Contractuel IBM, des Conditions Internationales d'Utilisation de Logiciels IBM, des Conditions d'Utilisation du Code Machine ou de tout autre contrat équivalent.

Les données de performances et les exemples de clients sont fournis à titre d'exemple uniquement. Les performances réelles peuvent varier en fonction des configurations et des conditions d'exploitation spécifiques.

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 illustratives ou explicatives uniquement. Toute ressemblance avec des noms de personnes, de sociétés ou des données réelles 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 lesquels ils ont été écrits ou aux interfaces de programmation IBM. Ces exemples de programmes n'ont pas été rigoureusement testés dans toutes les conditions. Par conséquent, IBM ne peut garantir expressément ou implicitement la fiabilité, la maintenabilité ou le fonctionnement de ces programmes. Les exemples de programmes sont fournis "EN L'ETAT", sans garantie d'aucune sorte. IBM ne sera en aucun cas responsable des dommages liés à l'utilisation des exemples de 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 :

© (nom de votre société) (année). Des segments de code sont dérivés des Programmes exemples d'IBM Corp.

© Copyright IBM Corp. _entrez l'année ou les années_. Tous droits réservés.

Marques

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines Corp. dans de nombreux pays. Les autres noms de produits et de services 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 l'adresse 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 certains 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/ou 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 toutes les marques et tous les logos incluant Java sont des marques d'Oracle et/ou de ses sociétés affiliées.

Index

A

Analyse de la covariance
GLM - Tests multivariés 3
analyse de survie
Kaplan-Meier 91
Analyse de survie
Modèle de Cox chronologique 99
Régression de Cox 95
Tables de survie 87
analyse de variance
dans les modèles linéaires mixtes
généralisés 59
Analyse de variance
dans Composantes de la Variance 22
analyse log-linéaire
dans les modèles linéaires mixtes
généralisés 59
Analyse log-linéaire 75
Analyse log-linéaire généralisée 79
Analyse log-linéaire Logit 83
analyse log linéaire généralisée
enregistrement de prévisions 81
enregistrement de variables 81
résidus 81
spécification du modèle 80
Analyse log-linéaire généralisée
Contrastes 79
Covariables de cellule 79
Critères 80
Distribution des effectifs des
cellules 79
Facteurs 79
fonctions supplémentaires de la
commande 81
Intervalles de confiance 80
Options d'affichage 80
Structures de cellule 79
tracés 80
analyse log-linéaire Logit
enregistrement de variables 85
prévisions 85
résidus 85
spécification du modèle 84
Analyse log-linéaire Logit 83
Contrastes 83
Covariables de cellule 83
Critères 85
Distribution des effectifs des
cellules 83
facteurs 83
Intervalles de confiance 85
Options d'affichage 85
Structures de cellule 83
tracés 85
analyse log-linéaire Sélection du modèle
définition des plages de facteurs 76
modèles 76
Analyse log-linéaire Sélection du
modèle 75
fonctions supplémentaires de la
commande 77

Analyse log-linéaire Sélection du modèle
(suite)
Options 76
analyse par la méthode des probits
modèles linéaires mixtes
généralisés 59
ANOVA
GLM - Mesures répétées 11
GLM - Tests multivariés 3
ANOVA multivariée 3

B

B de Tukey
GLM - Mesures répétées 16
GLM - Tests multivariés 7
Bonferroni
GLM - Mesures répétées 16
GLM - Tests multivariés 7

C

C de Dunnett
GLM - Mesures répétées 16
GLM - Tests multivariés 7
catégorie de référence
dans Modèles linéaires
généralisés 36
équations d'estimation
généralisées 49, 50
classe génératrice
analyse log-linéaire : Sélection du
modèle 76
composantes de variance 21
enregistrement des résultats 24
modèle 22
Composantes de variance
fonctions supplémentaires de la
commande 24
Options 22
Contrastes
Analyse log-linéaire généralisée 79
Analyse log-linéaire Logit 83
Régression de Cox 96
Convergence de Hess
dans Modèles linéaires
généralisés 38
équations d'estimation
généralisées 51
Convergence de log de vraisemblance
dans Modèles linéaires
généralisés 38
équations d'estimation
généralisées 51
Modèles mixtes linéaires 29
Convergence des paramètres
dans Modèles linéaires
généralisés 38
équations d'estimation
généralisées 51

Convergence des paramètres (suite)
Modèles mixtes linéaires 29
Covariables
Régression de Cox 96
covariables de chaîne
Régression de Cox 96

D

décomposition hiérarchique 5, 14
dans Composantes de la Variance 23
Différence la moins significative
GLM - Mesures répétées 16
GLM - Tests multivariés 7
Différence significative de Tukey
GLM - Mesures répétées 16
GLM - Tests multivariés 7
Distance de Cook
dans Modèles linéaires
généralisés 42
GLM 8
GLM - Mesures répétées 17
distribution binomiale
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33
distribution binomiale négative
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33
Distribution de Poisson
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33
distribution gamma
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33
distribution gaussienne inverse
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33
distribution multinomiale
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33
distribution normale
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33
distribution Tweedie
dans les équations d'estimation
généralisées 47

distribution Tweedie (*suite*)
dans les modèles linéaires
généralisés 33

E

Ecart-type
GLM - Mesures répétées 18
GLM - Tests multivariés 9

Effet aléatoire a priori
dans Composantes de la Variance 22

Effets aléatoires
Modèles mixtes linéaires 28

effets fixes
modèles mixtes linéaires 27

Elimination descendante
Analyse log-linéaire : Sélection du
modèle 75

Equations d'estimation généralisées
catégorie de référence pour les
réponses binaires 49
Critères d'estimation 51
options des facteurs catégoriels 50
statistiques 53
type de modèle 47

équations d'estimation généralisées 45
enregistrement des variables dans le
jeu de données actif 55
export de modèle 56
moyennes marginales estimées 54
prédicteurs 50
réponse 49
spécification du modèle 50
valeurs initiales 52

Erreur standard
GLM 8
GLM - Mesures répétées 17, 18
GLM - Tests multivariés 9

Estimation de l'intensité des effets
GLM - Mesures répétées 18
GLM - Tests multivariés 9

Estimation du maximum de
vraisemblance
dans Composantes de la Variance 22

Estimation du maximum de
vraisemblance restreint
dans Composantes de la Variance 22

Estimations de puissance
GLM - Mesures répétées 18
GLM - Tests multivariés 9

Estimations des paramètres
Analyse log-linéaire : Sélection du
modèle 76
Analyse log-linéaire généralisée 79
Analyse log-linéaire Logit 83
dans Modèles linéaires
généralisés 39
équations d'estimation
généralisées 53
GLM - Mesures répétées 18
GLM - Tests multivariés 9
Modèles mixtes linéaires 30

Eta carré
GLM - Mesures répétées 18
GLM - Tests multivariés 9

évaluation
Modèles mixtes linéaires 29

Evaluation de Fisher
Modèles mixtes linéaires 29

F

F de R-E-G-W (Ryan-Einot-Gabriel-
Welsch)
GLM - Mesures répétées 16
GLM - Tests multivariés 7

F multiple de Ryan-Einot-Gabriel-Welsch
GLM - Mesures répétées 16
GLM - Tests multivariés 7

Facteurs
GLM - Mesures répétées 13

fonction de lien binomiale négative
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien Cauchit cumulée
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien complément log
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien d'identité
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien log
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien log-log complémentaire
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien log-log complémentaire
cumulée
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien log-log négative
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien log-log négative cumulée
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien logit
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien logit cumulée
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien probit
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien probit cumulée
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien puissance
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

fonction de lien puissance du rapport des
cotes
dans les équations d'estimation
généralisées 47
dans les modèles linéaires
généralisés 33

Fonction de survie
Tables de survie 87

fonction générale estimée
dans Modèles linéaires
généralisés 39
équations d'estimation
généralisées 53

fonctions de lien
modèles linéaires mixtes
généralisés 60

fréquences
Analyse log-linéaire : Sélection du
modèle 76

G

GLM
Enregistrement de matrices 8
enregistrement de variables 8

GLM - Mesures répétées 11

Affichage 18
définir des facteurs 13

Diagnostics 18
enregistrement de variables 17
fonctions supplémentaires de la
commande 19

Modèle 14

Moyennes marginales estimées 18

Options 18
Tests post hoc 16
tracés de profil 16

GLM - Multivarié 3, 10

Affichage 9
Covariables 3
Diagnostics 9
facteurs 3
Moyennes marginales estimées 9
Options 9
Tests post hoc 7
tracés de profil 6
Variable dépendante 3

GLM – Tests multivariés 3
 GLOR
 Analyse log-linéaire généralisée 79
 GT2 de Hochberg
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

H

Historique des itérations
 dans Modèles linéaires
 généralisés 39
 équations d'estimation
 généralisées 53
 Modèles mixtes linéaires 29

I

Informations relatives au niveau de
 facteur
 Modèles mixtes linéaires 30
 informations sur le modèle
 dans Modèles linéaires
 généralisés 39
 équations d'estimation
 généralisées 53
 Intervalles de confiance
 Analyse log-linéaire généralisée 80
 Analyse log-linéaire Logit 85
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9
 Modèles mixtes linéaires 30
 Itérations
 Analyse log-linéaire : Sélection du
 modèle 76
 dans Modèles linéaires
 généralisés 38
 équations d'estimation
 généralisées 51

K

Kaplan-Meier 91
 comparaison des niveaux de
 facteur 92
 définition d'événements 92
 durée de survie de la moyenne et de
 la médiane 93
 enregistrement de nouvelles
 variables 92
 exemple 91
 fonctions supplémentaires de la
 commande 93
 quartiles 93
 statistiques 91, 93
 tables de survie 93
 tendance linéaire des niveaux de
 facteur 92
 tracés 93
 variables de statut de survie 92

L

Log rapport des cotes généralisé
 Analyse log-linéaire généralisée 79

LSD de Fisher
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

M

Matrice de corrélation
 dans Modèles linéaires
 généralisés 39
 équations d'estimation
 généralisées 53
 Modèles mixtes linéaires 30
 Matrice de covariance
 dans Modèles linéaires
 généralisés 38, 39
 équations d'estimation
 généralisées 51, 53
 GLM 8
 Modèles mixtes linéaires 30
 Matrice de covariance à effet aléatoire
 Modèles mixtes linéaires 30
 Matrice de covariance des paramètres
 Modèles mixtes linéaires 30
 Matrice de covariance des résidus
 Modèles mixtes linéaires 30
 matrice des coefficients de contraste
 dans Modèles linéaires
 généralisés 39
 équations d'estimation
 généralisées 53
 Matrice L
 dans Modèles linéaires
 généralisés 39
 équations d'estimation
 généralisées 53
 Méthode de Newton-Raphson
 Analyse log-linéaire généralisée 79
 Analyse log-linéaire Logit 83
 MINQUE
 dans Composantes de la Variance 22
 Modèle des risques proportionnels
 Régression de Cox 95
 modèle linéaire généralisé
 modèles linéaires mixtes
 généralisés 59
 modèle linéaire généralisé
 dans les modèles linéaires mixtes
 généralisés 59
 modèles factoriels complets
 dans Composantes de la Variance 22
 Modèles factoriels complets
 GLM - Mesures répétées 14
 modèles hiérarchiques
 modèles linéaires mixtes
 généralisés 59
 modèles linéaires généralisés 33
 catégorie de référence pour les
 réponses binaires 36
 distribution 33
 export de modèle 43
 fonctions de lien 33
 moyennes marginales estimées 41
 options des facteurs catégoriels 37
 prédicteurs 36
 réponse 35
 spécification du modèle 37
 types de modèle 33

modèles linéaires généralisés (*suite*)
 valeurs initiales 39
 Modèles linéaires généralisés
 Critères d'estimation 38
 enregistrement des variables dans le
 jeu de données actif 42
 statistiques 39
 modèles linéaires mixtes généralisés 59
 bloc d'effets aléatoires 64
 coefficients fixes 70
 covariances à effet aléatoire 71
 décalage 65
 distribution de la cible 60
 effets aléatoires 64
 effets fixes 63
 Effets fixes 70
 enregistrer champs 68
 export de modèle 68
 fonctions de lien 60
 moyennes estimées 72
 moyennes marginales estimées 67
 paramètres de covariance 71
 pondération d'analyse 65
 récapitulatif du modèle 69
 structure des données 69
 table de classification 70
 termes personnalisés 63
 valeurs prédites en fonction des
 valeurs observées 69
 vue du modèle 69
 Modèles log-linéaires hiérarchiques 75
 Modèles logit multinomiaux 83
 modèles longitudinaux
 modèles linéaires mixtes
 généralisés 59
 modèles mixtes
 linéaire 25
 modèles linéaires mixtes
 généralisés 59
 modèles mixtes linéaires 25, 105
 effets fixes 27
 enregistrement de variables 31
 moyennes marginales estimées 30
 Structure de covariance 105
 Modèles mixtes linéaires
 Critères d'estimation 29
 Effets aléatoires 28
 fonctions supplémentaires de la
 commande 31
 Modèle 30
 Termes construits 27
 Termes d'interaction 27
 modèles multi-niveaux
 modèles linéaires mixtes
 généralisés 59
 modèles personnalisés
 analyse log-linéaire : Sélection du
 modèle 76
 dans Composantes de la Variance 22
 Modèles personnalisés
 GLM - Mesures répétées 14
 modèles saturés
 analyse log-linéaire : Sélection du
 modèle 76
 moyennes marginales estimées
 dans Modèles linéaires
 généralisés 41

moyennes marginales estimées (*suite*)
 équations d'estimation
 généralisées 54
 modèles mixtes linéaires 30
 Moyennes marginales estimées
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9
 Moyennes observées
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9

N

Newman-Keuls
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

O

observations censurées
 Kaplan-Meier 91
 Observations censurées
 Régression de Cox 95
 Tables de survie 87

P

paramètre d'échelle
 dans Modèles linéaires
 généralisés 38
 équations d'estimation
 généralisées 51
 Plage multiple de Ryan-Einot-Gabriel-
 Welsch
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7
 Prédicteurs chronologiques segmentés
 Régression de Cox 99
 prévisions
 analyse log linéaire généralisée 81
 analyse log-linéaire Logit 85
 modèles mixtes linéaires 31
 prévisions fixes
 modèles mixtes linéaires 31
 Prévisions pondérées
 GLM 8
 GLM - Mesures répétées 17
 Produits croisés
 Matrices d'hypothèse et d'erreur 9

Q

Q de R-E-G-W (Ryan-Einot-Gabriel-
 Welsch)
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7
 Qualité de l'ajustement
 dans Modèles linéaires
 généralisés 39
 équations d'estimation
 généralisées 53

R

rapport des cotes
 Analyse log-linéaire généralisée 79
 récapitulatif de traitement des
 observations
 dans Modèles linéaires
 généralisés 39
 équations d'estimation
 généralisées 53
 régression de Cox
 définition d'événements 98
 DfBêta(s) 97
 enregistrement de nouvelles
 variables 97
 fFonctions de la ligne de base 97
 fonction de hasard 97
 fonction de survie 97
 introduction et élimination pas à
 pas 97
 itérations 97
 résiduels partiels 97
 statistiques 97
 tracés 96
 variable de statut de survie 98
 Régression de Cox 95
 Contrastes 96
 Covariables 95
 covariables de chaîne 96
 covariables qualitatives 96
 exemple 95
 fonctions supplémentaires de la
 commande 98
 Prédicteurs chronologiques 99, 100
 statistiques 95
 régression de Poisson
 modèles linéaires mixtes
 généralisés 59
 Régression de Poisson
 Analyse log-linéaire généralisée 79
 Régression logistique
 modèles linéaires mixtes
 généralisés 59
 régression logistique multinomiale
 modèles linéaires mixtes
 généralisés 59
 Régression multivariée 3
 Résidu non standardisé
 GLM 8
 GLM - Mesures répétées 17
 résiduels de Pearson
 équations d'estimation
 généralisées 55
 Résiduels de Pearson
 dans Modèles linéaires
 généralisés 42
 résidus
 analyse log linéaire généralisée 81
 analyse log-linéaire Logit 85
 équations d'estimation
 généralisées 55
 modèles mixtes linéaires 31
 Résidus
 Analyse log-linéaire : Sélection du
 modèle 76
 dans Modèles linéaires
 généralisés 42

résidus au sens déviance
 dans Modèles linéaires
 généralisés 42
 résidus de vraisemblance
 dans Modèles linéaires
 généralisés 42
 Résidus standardisés
 GLM 8
 GLM - Mesures répétées 17
 Résidus supprimés
 GLM 8
 GLM - Mesures répétées 17

S

Séparation
 dans Modèles linéaires
 généralisés 38
 équations d'estimation
 généralisées 51
 somme des carrés 5, 14
 dans Composantes de la Variance 23
 modèles mixtes linéaires 28
 Somme des carrés
 Matrices d'hypothèse et d'erreur 9
 Somme des carrés et produits croisés des
 résidus
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9
 SSCP
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9
 Statistique de Wald
 Analyse log-linéaire généralisée 79
 Analyse log-linéaire Logit 83
 Statistiques descriptives
 dans Modèles linéaires
 généralisés 39
 équations d'estimation
 généralisées 53
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9
 Modèles mixtes linéaires 30
 Step-halving
 dans Modèles linéaires
 généralisés 38
 équations d'estimation
 généralisées 51
 Modèles mixtes linéaires 29
 structures de covariance 105
 modèles mixtes linéaires 105
 Student-Newman-Keuls
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

T

T2 de Tamhane
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7
 T3 de Dunnett
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7
 Tableau croisé
 Analyse log-linéaire : Sélection du
 modèle 75

Tableaux de contingence
 Analyse log-linéaire généralisée 79

tables de survie
 variables de facteur 88
 variables de statut de survie 88

Tables de survie 87
 Comparaison des niveaux de
 facteur 88
 exemple 87
 Fonction de survie 87
 fonctions supplémentaires de la
 commande 89
 statistiques 87
 Suppression de l'affichage de la
 table 88
 Taux de risque 87
 Test de Wilcoxon (Gehan) 88
 tracés 88

Taux de risque
 Tables de survie 87

termes construits 4, 14, 22, 76, 80, 84
 termes d'interaction 4, 14, 22, 76, 80, 84

Termes d'interaction
 Modèles mixtes linéaires 27

termes imbriqués
 dans Modèles linéaires
 généralisés 37
 équations d'estimation
 généralisées 50

Termes imbriqués
 Modèles mixtes linéaires 27

test de Breslow
 Kaplan-Meier 92

Test de comparaison appariée de Gabriel
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

Test de comparaison appariée de Games
 et Howell
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

Test de Gehan
 Tables de survie 88

Test de Levene
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9

Test de plage multiple de Duncan
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

test de Scheffé
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

Test de sphéricité de Bartlett
 GLM - Tests multivariés 9

Test de sphéricité de Mauchly
 GLM - Mesures répétées 18

test de Tarone-Ware
 Kaplan-Meier 92

Test de Wilcoxon
 Tables de survie 88

Test des paramètres de covariance
 Modèles mixtes linéaires 30

Test du multiplicateur de Lagrange
 dans Modèles linéaires
 généralisés 39

test log rank
 Kaplan-Meier 92

Test M de Box
 GLM - Tests multivariés 9

Test t :
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9

Test t de Dunnett
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

Test t de Sidak
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

Test t de Waller-Duncan
 GLM - Mesures répétées 16
 GLM - Tests multivariés 7

Tests d'homogénéité de la variance
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9

Tolérance de singularité
 Modèles mixtes linéaires 29

tracé de probabilités gaussien
 Analyse log-linéaire : Sélection du
 modèle 76

tracés
 Analyse log-linéaire généralisée 80
 Analyse log-linéaire Logit 85

tracés de profil
 GLM - Mesures répétées 16
 GLM - Tests multivariés 6

tracés dispersion/niveau
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9

Tracés résiduels
 GLM - Mesures répétées 18
 GLM - Tests multivariés 9

V

Valeurs influentes
 dans Modèles linéaires
 généralisés 42
 GLM 8
 GLM - Mesures répétées 17

variables de mesures répétées
 modèles mixtes linéaires 26

variables de sujets
 modèles mixtes linéaires 26

vue du modèle
 dans les modèles linéaires mixtes
 généralisés 69

