

IBM SPSS Advanced Statistics 32

IBM

Remarque

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section [«Remarques»](#), à la page 145.

Notice d'édition

La présente édition s'applique à la version 32.0.0 d'IBM® SPSS Statistiques et à toutes les éditions et modifications ultérieures sauf mention contraire dans les nouvelles éditions.

© **Copyright International Business Machines Corporation .**

Table des matières

Chapitre 1. Statistiques avancées.....	1
Introduction aux Statistiques avancées.....	1
Analyses GLM - Tests multivariés.....	1
Modèle d'analyse GLM - Multivariés.....	3
GLM - Tests multivariés : Contrastes.....	4
GLM - Tests multivariés : Tracés de profil.....	5
Comparaisons multivariées post hoc GLM.....	5
Moyenne marginale estimée GLM.....	6
Enregistrement GLM.....	7
GLM - Tests multivariés : Options.....	8
Fonctions supplémentaires de la commande GLM.....	9
GLM - Mesures répétées.....	9
GLM - Mesures répétées : Définir des facteurs.....	12
GLM - Mesures répétées : Modèles.....	12
GLM - Mesures répétées : Contrastes.....	13
GLM - Mesures répétées : Tracés de profil.....	14
GLM - Mesures répétées : Comparaisons post hoc.....	15
Moyenne marginale estimée GLM.....	16
GLM - Mesures répétées : Enregistrer.....	16
Options de GLM - Mesures répétées.....	17
Fonctions supplémentaires de la commande GLM.....	18
Analyse des composantes de variance.....	18
Modèle des composantes de variance.....	20
Options des composantes de variance.....	20
Composantes de variance : Enregistrer dans un nouveau fichier.....	21
Fonctions supplémentaires de la commande VARCOMP.....	22
Modèles mixtes linéaires.....	22
Modèles mixtes linéaires : Sujets et éléments répétés.....	23
Modèles mixtes linéaires : Effets fixes.....	25
Modèles mixtes linéaires : Effets aléatoires.....	26
Estimation des modèles mixtes linéaires.....	27
Modèles mixtes linéaires : Statistiques.....	28
Moyennes EM des modèles mixtes linéaires.....	28
Modèles mixtes linéaires : Enregistrer.....	29
Modèles mixtes linéaires-Exportation.....	29
Fonctions supplémentaires de la commande MIXED.....	29
modèles linéaires généralisés.....	30
Modèles linéaires généralisés - Réponse.....	32
Prédicteurs de modèles linéaires généralisés.....	33
Modèle des modèles linéaires généralisés.....	34
Estimation des modèles linéaires généralisés.....	35
Statistiques des modèles linéaires généralisés.....	36
Moyennes EM des modèles linéaires généralisés.....	37
Enregistrement des modèles linéaires généralisés.....	38
Export de modèles linéaires généralisés.....	39
Fonctions supplémentaires de la commande GENLIN.....	40
équations d'estimation généralisées.....	40
Equations d'estimation généralisées : type de modèle.....	42
Réponse des équations d'estimation généralisées.....	44
Prédicteurs des équations d'estimation généralisées.....	45
Modèle des équations d'estimation généralisées.....	46

Estimation des équations d'estimation généralisées.....	47
Statistiques des équations d'estimation généralisées.....	48
Moyennes EM des équations d'estimation généralisées.....	49
Enregistrement des équations d'estimation généralisées.....	50
Export d'équations d'estimation généralisées.....	51
Fonctions supplémentaires de la commande GENLIN.....	52
Modèles linéaires mixtes généralisés.....	52
Obtention d'un modèle linéaire mixte généralisé.....	54
Cible	55
Effets fixes	57
Effets aléatoires	58
Pondération et décalage	59
Options de création générales	60
Estimation	60
Moyennes estimées	61
Enregistrement	62
Exportation	63
Vue Modèle	63
Analyse log-linéaire Sélection du modèle.....	67
Analyse log-linéaire : Définir plage.....	68
Modèle d'analyse log-linéaire.....	68
Options de l'analyse log-linéaire de sélection de modèle.....	68
Fonctions supplémentaires de la commande HILOGLINEAR.....	69
Analyse log-linéaire généralisée.....	69
Modèle d'analyse log-linéaire généralisée.....	70
Options de l'analyse log-linéaire généralisée.....	71
Analyse log-linéaire généralisée : Enregistrer.....	71
Fonctions supplémentaires de la commande GENLOG.....	72
Analyse log-linéaire Logit.....	72
Modèle d'analyse log-linéaire Logit.....	73
Options de l'analyse log-linéaire Logit.....	74
Analyse log-linéaire Logit : Enregistrer.....	74
Fonctions supplémentaires de la commande GENLOG.....	75
Tables de survie.....	75
Tables de survie : Définir plage.....	76
Tables de survie : Options.....	76
Fonctions supplémentaires de la commande SURVIVAL.....	77
Modèles de temps d'échec accéléré paramétriques.....	77
Modèles de temps d'échec accéléré paramétriques: Critères.....	78
Modèles de temps d'échec accéléré paramétriques: Modèle.....	79
Modèles de temps d'échec accélérés paramétriques: estimation.....	79
Modèles de temps d'échec accéléré paramétriques: Imprimer.....	81
Modèles de temps d'échec accéléré paramétriques: Prévoir.....	81
Modèles de temps d'échec accéléré paramétriques: Tracé.....	82
Modèles de temps d'échec accéléré paramétriques: Exportation.....	83
Survie AFT Définir des événements pour les variables de statut.....	83
Modèles de temps d'échec accélérés paramétriques: Sélectionner une catégorie.....	83
Modèles de fragilité partagés paramétriques.....	84
Modèles de fragilité partagés paramétriques: Critères.....	85
Modèles de fragilité partagés paramétriques: Modèle.....	86
Modèles de fragilité partagés paramétriques: estimation.....	87
Modèles de fragilité partagés paramétriques: Imprimer.....	88
Modèles de fragilité partagés paramétriques: Prévoir.....	88
Modèles de fragilité partagés paramétriques: Tracé.....	89
Modèles de fragilité partagés paramétriques: Exportation.....	90
Modèles de fragilité partagés paramétriques: Définir des événements.....	90
Modèles de fragilité partagés paramétriques-Exemples.....	91
Modèles de fragilité partagés paramétriques-Étude de cas pour les données récurrentes.....	93

Analyse de survie de Kaplan-Meier.....	95
Kaplan-Meier : Définir un événement pour la variable de statut.....	96
Kaplan-Meier : Comparer les niveaux du facteur.....	96
Kaplan-Meier : Enregistrer les nouvelles variables.....	97
Kaplan-Meier : Options.....	97
Fonctions supplémentaires de la commande KM.....	97
Analyse de la régression de Cox.....	97
Modèle de Cox... Définir des variables qualitatives.....	98
Tracés de la régression de Cox.....	99
Modèle de Cox : Enregistrer les nouvelles variables.....	100
Options de la régression de Cox.....	100
Modèle de Cox : Définir un événement pour la variable de statut.....	101
Fonctions supplémentaires de la commande COXREG.....	101
Calcul des prédicteurs chronologiques.....	101
Calculer un prédicteur chronologique.....	102
Méthodes de codification des variables catégorielles.....	102
Déviation.....	102
Simple.....	103
Helmert.....	103
Différence.....	103
Polynomial.....	104
Répété.....	104
Spécial.....	105
Indicateur.....	105
Structures de covariance.....	105
Statistiques de Bayes.....	110
Inférence bayésienne à échantillon unique : Normale.....	111
Inférence bayésienne à échantillon unique : Distribution binomiale.....	115
Inférence bayésienne à échantillon unique : Poisson.....	117
Inférence bayésienne à échantillons liés : Distribution normale.....	119
Inférence bayésienne à échantillon indépendant.....	121
Inférence bayésienne sur corrélation de Pearson.....	125
Inférence bayésienne sur modèles de régression linéaire.....	128
ANOVA à 1 facteur de Bayes.....	133
Modèles log-linéaires bayésiens.....	136
Modèles d'analyse ANOVA bayésienne à un facteur avec mesures répétées.....	140
Kernel Ridge Regression.....	142
Paramètres de noyau.....	143
Régression de la crête du noyau: Options.....	144
Remarques.....	145
Marques.....	146
Index.....	149

Chapitre 1. Statistiques avancées

Les fonctions statistiques avancées suivantes sont incluses dans SPSS Statistics Standard Edition ou l'option Statistiques avancées.

Introduction aux Statistiques avancées

L'option Statistiques avancées inclut des procédures qui offrent des options de modélisation plus avancées que celles disponibles via 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.

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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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 la plus fréquemment utilisée.

Inclure ordonnée à l'origine dans le 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 et termes personnalisés

Termes construits

Utilisez cette option lorsque vous souhaitez inclure des termes non imbriqués d'un certain type (tels que des effets principaux) pour toutes les combinaisons d'un ensemble de facteurs et de covariables.

Termes construits personnalisés

Utilisez cette option lorsque vous souhaitez inclure des termes imbriqués ou lorsque vous souhaitez générer explicitement un terme variable par variable. Le processus de construction d'un terme imbriqué implique les étapes suivantes :

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.

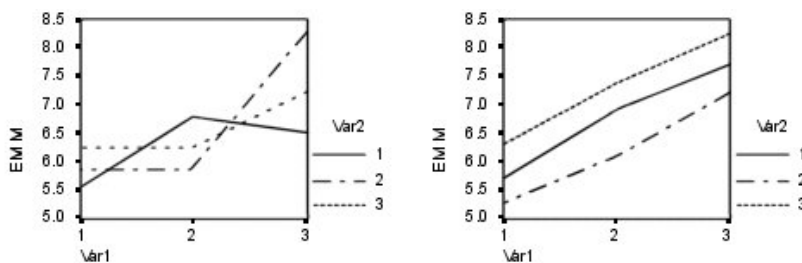


Figure 1. Tracé non parallèle (gauche) et tracé parallèle (droite)

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 à n'importe quel niveau du facteur est plus petite que celle de la catégorie de contrôle, sélectionnez **< Control**. De

même, pour vérifier si la moyenne à n'importe quel niveau du facteur est supérieure 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.

Moyenne marginale estimée GLM

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.

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.

Comparer les effets principaux simples

Ce paramètre est activé lorsque la liste cible contient un ou plusieurs effets de produit ou d'interaction (par exemple, $A \times B$, $A \times B \times C$). Ce paramètre permet de spécifier des comparaisons entre des effets principaux simples, qui sont des effets principaux imbriqués dans les niveaux d'autres facteurs.

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 n'est disponible que si **Comparer les effets principaux** et/ou **Comparer les effets principaux simples** est sélectionné.

Spécification de la moyenne marginale estimée

1. Dans les menus, choisissez l'une des procédures disponibles sous > **Analyse** > **Modèle linéaire général**.
2. Dans la boîte de dialogue principale, cliquez sur **Moyennes EM**.

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 Statistiques, vous devez enregistrer le fichier de données en cours.

Prévisions : Valeurs que le modèle estime pour chaque observation.

- *Non normalisé*. Valeur prévue par le modèle pour la variable dépendante.
- *Pondéré*. Valeurs prédites 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.

Diagnostiques : 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 de levier*. 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 normalisé*. Différence entre une valeur observée et la valeur prévue par le modèle.
- *Pondéré*. Résidus non standardisés pondérés. Disponibles uniquement lorsqu'une variable WLS a été préalablement sélectionnée.
- *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.
- *Studentized*. Résidu, divisé par une estimation de son écart type, qui varie d'une observation à l'autre, selon la distance entre les valeurs et les moyennes des variables indépendantes de chaque observation. Parfois appelés résidus de Student interne.
- *Supprimé*. Résidu d'une observation lorsque cette dernière 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 : Écrit 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 Statistiques. D'autre part, pour chaque variable dépendante, il y aura une ligne d'estimations de paramètres, une ligne d'erreurs standard des estimations de paramètres, 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. Lorsque les statistiques cohérentes avec l'hétéroscédasticité sont sélectionnées (disponibles uniquement pour les modèles univariés), la matrice variance-covariance est calculée à l'aide d'un estimateur-M, la ligne des erreurs standard affiche les erreurs standard robustes et les valeurs de signification reflètent les erreurs robustes. 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.

Affichage : 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(s) d'estimation générale(s)** permet de construire des tests d'hypothèses personnalisés basés sur la ou les fonctions d'estimation générale(s). Les lignes d'une matrice de coefficients de contraste sont des combinaisons linéaires de la ou des fonctions d'estimation générale(s).

Afficher

Statistiques descriptives

Produit les moyennes observées, les écarts types et dénombre toutes les variables dépendantes dans toutes les cellules.

Estimations d'effet de taille

Renvoie une valeur η^2 carré partielle pour chaque effet et estimation de paramètre. La statistique d'Eta carré décrit la proportion de la variabilité totale imputable au facteur.

Puissance observée

Renvoie la puissance du test lorsque l'hypothèse alternative n'est pas basée sur la valeur observée.

Estimations des paramètres

Renvoie les estimations de paramètres, les écarts type, les tests *t*, les intervalles de confiance et la puissance observée pour chaque test.

Matrices SSCP

Affiche l'hypothèse et les matrices SSCP d'erreur.

Matrice SSCP des résidus

Affiche l'hypothèse et les matrices SSCP des résidus.

Matrice de transformation

Affiche le test Bartlett de sphéricité de la matrice de covariance des résidus.

Tests d'homogénéité

Produit le test Levene d'homogénéité de la variance pour chaque variable dépendante sur toutes les combinaisons de niveaux des facteurs intra-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.

Tracé de propagation par niveau

Utile pour vérifier les hypothèses sur les données de l'examen de l'hypothèse de variance égale. Ceci n'est pas valable s'il n'y a pas de facteurs.

Tracé résiduel

Produit un tracé des résidus observé-prévu-standardisé pour chaque variable dépendante. Ce tracé est utile pour vérifier l'hypothèse de variance égale.

Manque d'ajustement

Vérifie si la relation entre la variable dépendante et les variables indépendantes peut être décrite adéquatement par le modèle.

Fonction(s) d'estimation générale(s)

Permet de construire des tests d'hypothèses personnalisés basés sur la ou les fonctions d'estimation générale(s). Les lignes d'une matrice de coefficients de contraste sont des combinaisons linéaires de la ou des fonctions d'estimation générale(s).

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.

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 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 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 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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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.

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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 *weight1*, ..., *weight5* sont utilisés pour attribuer les cinq niveaux de *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 et termes personnalisés

Termes construits

Utilisez cette option lorsque vous souhaitez inclure des termes non imbriqués d'un certain type (tels que des effets principaux) pour toutes les combinaisons d'un ensemble de facteurs et de covariables.

Termes construits personnalisés

Utilisez cette option lorsque vous souhaitez inclure des termes imbriqués ou lorsque vous souhaitez générer explicitement un terme variable par variable. Le processus de construction d'un terme imbriqué implique les étapes suivantes :

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

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 - 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.

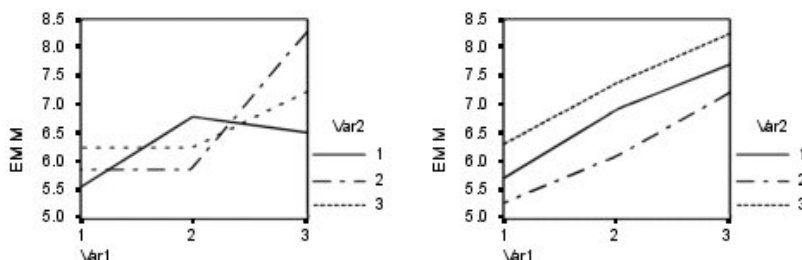


Figure 2. Tracé non parallèle (gauche) et tracé parallèle (droite)

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 à n'importe quel niveau du facteur est plus petite que celle de la catégorie de contrôle, sélectionnez **< Contrôle**. De même, pour vérifier si la moyenne à n'importe quel niveau du facteur est supérieure 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 Dunnetts** (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.

Moyenne marginale estimée GLM

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.

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.

Comparer les effets principaux simples

Ce paramètre est activé lorsque la liste cible contient un ou plusieurs effets de produit ou d'interaction (par exemple, $A \times B$, $A \times B \times C$). Ce paramètre permet de spécifier des comparaisons entre des effets principaux simples, qui sont des effets principaux imbriqués dans les niveaux d'autres facteurs.

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 n'est disponible que si **Comparer les effets principaux** et/ou **Comparer les effets principaux simples** est sélectionné.

Spécification de la moyenne marginale estimée

1. Dans les menus, choisissez l'une des procédures disponibles sous > **Analyse** > **Modèle linéaire général**.
2. Dans la boîte de dialogue principale, cliquez sur **Moyennes EM**.

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 Statistiques, vous devez enregistrer le fichier de données en cours.

Prévisions : Valeurs que le modèle estime pour chaque observation.

- *Non normalisé.* 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.

Diagnostiques : 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 de levier.* 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 normalisé*. Différence entre une valeur observée et la valeur prévue par le modèle.
- *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.
- *Studentized*. Résidu, divisé par une estimation de son écart type, qui varie d'une observation à l'autre, selon la distance entre les valeurs et les moyennes des variables indépendantes de chaque observation. Parfois appelés résidus de Student interne.
- *Supprimé*. Résidu d'une observation lorsque cette dernière 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.

Afficher

Statistiques descriptives

Produit les moyennes observées, les écarts types et dénombre toutes les variables dépendantes dans toutes les cellules.

Estimations d'effet de taille

Renvoie une valeur η^2 carré partielle pour chaque effet et estimation de paramètre. La statistique d' η^2 carré décrit la proportion de la variabilité totale imputable au facteur.

Puissance observée

Renvoie la puissance du test lorsque l'hypothèse alternative n'est pas basée sur la valeur observée.

Estimations des paramètres

Renvoie les estimations de paramètres, les écarts type, les tests t , les intervalles de confiance et la puissance observée pour chaque test.

Matrices SSCP

Affiche l'hypothèse et les matrices SSCP d'erreur.

Matrice SSCP résiduelle

Affiche la matrice SSCP des résidus.

Matrice de transformation

Affiche le test Bartlett de sphéricité de la matrice de covariance des résidus.

Tests d'homogénéité

Produit le test Levene d'homogénéité de la variance pour chaque variable dépendante sur toutes les combinaisons de niveaux des facteurs intra-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.

Tracé de propagation par niveau

Utile pour vérifier les hypothèses sur les données. Cette option est désactivée s'il n'existe pas de facteurs.

Tracé résiduel

Produit un tracé des résidus observé-prévu-standardisé pour chaque variable dépendante. Ces tracés sont utiles pour vérifier l'hypothèse de variance égale. Cette option est désactivée s'il n'existe pas de facteurs.

Manque d'ajustement

Vérifie si la relation entre la variable dépendante et les variables indépendantes peut être décrite adéquatement par le modèle.

Fonction(s) d'estimation générale(s)

Permet de construire des tests d'hypothèses personnalisés basés sur la ou les fonctions d'estimation générale(s). Les lignes d'une matrice de coefficients de contraste sont des combinaisons linéaires de la ou des fonctions d'estimation générale(s).

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 univariées, multivariées 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).
- Construisez 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.

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 collège 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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez une variable dépendante.

3. Sélectionnez des variables pour Facteurs fixés, 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 & 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.

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.

Tous d'ordre 2

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

Tous d'ordre 3

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

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.

Inclure ordonnée à l'origine dans le 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 et termes personnalisés

Termes construits

Utilisez cette option lorsque vous souhaitez inclure des termes non imbriqués d'un certain type (tels que des effets principaux) pour toutes les combinaisons d'un ensemble de facteurs et de covariables.

Termes construits personnalisés

Utilisez cette option lorsque vous souhaitez inclure des termes imbriqués ou lorsque vous souhaitez générer explicitement un terme variable par variable. Le processus de construction d'un terme imbriqué implique les étapes suivantes :

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 fixés. 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 Statistiques.

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.

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.

Remarques sur les 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 :

Analyser > 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. Si des variables répétées sont définies, sélectionnez un **Type de covariance répété** dans la liste déroulante.
4. Vous pouvez également sélectionner une ou plusieurs variables de mesure de Kronecker.
5. Sélectionnez une structure de covariance de résidus, si vous le souhaitez.
6. Cliquez sur **Continuer**
7. Sélectionnez une variable dépendante.
8. Sélectionnez au moins un facteur ou une covariable.
9. 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 : Sujets et éléments répétés

Cette boîte de dialogue permet de sélectionner les variables qui définissent les sujets, les observations répétées, ainsi que les mesures de Kronecker, et de choisir une structure 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 de la 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)
- Produit direct AR1 (UN_AR1)
- Produit direct non structuré (UN_UN)
- Symétrie composée de produit direct (UN_CS)
- 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
- Spatiale : Puissance
- Spatiale : Exponentielle
- Spatiale : de Gausse
- Spatiale : Linéaire
- Spatiale : Log linéaire
- Spatiale : Sphérique

Mesures de Kronecker

Sélectionnez des variables qui indiquent la structure des sujets des mesures de covariance de Kronecker et déterminent le mode de corrélation des erreurs de mesure. Ce champ est disponible uniquement lorsque l'une des options **Type de covariance répétée** suivante est sélectionnée :

- Produit direct AR1 (UN_AR1)
- Produit direct non structuré (UN_UN)
- Symétrie composée de produit direct (UN_CS)

Coordonnées de covariance spatiale

Les variables de cette liste précisent les coordonnées des observations répétées lorsqu'un type de covariance spatiale est sélectionné comme type de covariance répétée.

Pour plus d'informations, voir la rubrique [«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 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.

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 la rubrique «[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.

Afficher les prédictions de paramètres pour cet ensemble d'effets aléatoires : Affiche les estimations de paramètres à effets aléatoires.

Estimation des modèles mixtes linéaires

Méthode

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

Degrés de liberté

Fournit les options permettant de définir les degrés de liberté de tous les tests.

Méthode des résidus

La méthode des résidus indique des degrés de liberté fixes pour tous les tests. Elle s'avère utile 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).

Approximation de Satterthwaite

La méthode de Satterthwaite comporte un champ indiquant les degrés de liberté des tests. Elle s'avère utile 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).

Approximation de Kenward-Roger

La méthode de Kenward-Roger offre un estimateur de petit échantillon plus précis pour la variance-covariance des paramètres d'effets fixes et les degrés approximatifs de liberté du dénominateur de t-tests et F-tests. La méthode introduit un facteur d'échelle pour les statistiques F et les estime, ainsi que les degrés de liberté du dénominateur, en utilisant une extension de la série Taylor pour la structure aléatoire estimée dans les données.

Remarque : La méthode Kenward-Roger est utilisée dans la covariance basée sur un modèle (au lieu de la covariance robuste). Lorsque la méthode de Kenward-Roger et la covariance robuste sont sélectionnées, la méthode de Kenward-Roger est appliquée à la covariance basée sur les modèles et l'avertissement suivant s'affiche : "Puisque la méthode de Kenward-Roger est sélectionnée, la méthode de la covariance robuste devient méthode de covariance basée sur les modèles".

Itérations

Les options suivantes sont disponibles :

Nombre maximum d'itérations

Spécifiez un nombre entier non négatif.

Découpage maximal d'étape en deux

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.

Imprimer l'historique des itérations pour chaque étape

Affiche un tableau contenant la valeur de la fonction de log-vraisemblance et les estimations des paramètres toutes les n itérations en commençant par la 0^{e} itération (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.

Étapes maxi d'évaluation

Demandes d'utilisation de l'algorithme de coordonnées de Fisher jusqu'au numéro d'itération n .
Spécifiez un nombre 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 du 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 :

- **Estimation des paramètres pour les effets fixes** : Affiche les estimations de paramètres à effets fixes, 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.

Prévisions & Résidus. 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.

Modèles mixtes linéaires-Exportation

Cette boîte de dialogue vous permet d'exporter le contenu des tableaux de sortie EBLUP (Empirique Best Linear ???) vers des jeux de données ou des fichiers .sav . Le bouton Exporter de la boîte de dialogue Modèles mixtes linéaires est activé si au moins un effet aléatoire est spécifié via la boîte de dialogue Aléatoire et que la case " Afficher les prévisions des paramètres pour ce bloc est cochée.

Exporter EBLUPS

Sélectionnez une destination: **Fichier** ou **Fichier de données**

Vous pouvez éventuellement indiquer un nom.

Si plusieurs effets aléatoires avec EBLUP ont été spécifiés, décochez la case pour que chaque tableau de résultats apparaisse dans un jeu de données ou un fichier distinct.

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.

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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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 secondaire de la distribution binomiale négative peut être n'importe quel nombre supérieur ou égal à 0 ; vous pouvez la définir sur une valeur fixe ou lui permettre d'être estimée 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 est la spécification de nombre requise 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** : $f(x)=x^\alpha$, if $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. α est la spécification du 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-carré 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 β_0 structurel β_1 . 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.

Tous d'ordre 2

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

Tous d'ordre 3

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

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.

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 hessienne.** 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, l'ordre des variables de gauche à droite doit être : *RowType_*, *VarName_*, *P1*, *P2*, ..., où *RowType_* et *VarName_* sont des variables de chaîne et *P1*, *P2*, ... 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 Fichier scindé est en vigueur, les variables doivent commencer par la variable de fichier scindé ou les variables dans l'ordre indiqué lors de la création du fichier scindé, suivi de *RowType_*, *VarName_*, *P1*, *P2*, ... 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 variable *P1*, *P2*, ... ne sont pas requis ; la procédure accepte les noms de variables valides pour les paramètres car le mapping des variables aux paramètres est basé sur la position de la variable, et non sur le nom de la variable. 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.
- **Intervalles de confiance :** Indiquez un niveau de confiance supérieur à 50 et inférieur à 100. Les intervalles de Wald sont basés sur l'hypothèse selon laquelle les paramètres ont une distribution normale asymptotique. Les intervalles par vraisemblance de profil sont plus exacts mais peuvent s'avérer coûteux en termes 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.

Impression. 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-carré de Pearson et le khi-carré 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 des paramètres pour chaque n -ième itération à partir de la 0^e itération (les estimations initiales), où n correspond à la valeur de l'intervalle d'affichage. Si l'historique des itérations est requis, la dernière itération est toujours affichée quel que soit n .
- **Test du multiplicateur de Lagrange du paramètre d'échelle ou paramètre auxiliaire de distribution binomiale négative**. 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-carré 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.

- **Déviation** : Chaque niveau du facteur est comparé à la moyenne générale. Les contrastes de déviation ne sont pas orthogonaux.
- *Difference*. 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 d'importance 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 seuil global de signification.
- *Sidak*. Cette méthode fournit des limites (bornes) 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 seuil 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.

Valeurs influentes (ou bras de levier)

Mesure 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

Différence entre une 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-carré 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.

Résidu 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 : Ecrit un jeu de données dans un format IBM SPSS Statistiques 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.

- **Variation de scission :** En cas d'utilisation, toutes les variables définissant des scissions.
- **RowType_.** Prend des valeurs (et des libellés de valeur) *COV* (covariances), *CORR* (corrélations), *EST* (estimations de paramètres), *SE* (erreurs standard), *SIG* (niveaux de signification) et *DF* (degrés de liberté de conception 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 $P1, P2, \dots$, correspondant à une liste ordonnée de tous les paramètres estimés du modèle (à l'exception des paramètres d'échelle ou binomiaux négatifs), pour les types de ligne *COV* ou *CORR*, avec des étiquettes de valeur correspondant aux chaînes de paramètres indiquées 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 d'importance 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 d'importance 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.

équations 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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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 hessienne.** 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 secondaire de la distribution binomiale négative peut être n'importe quel nombre supérieur ou égal à 0 ; vous pouvez la définir sur une valeur fixe ou lui permettre d'être estimée 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 est la spécification de nombre requise 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** : $f(x)=x^\alpha$, if $\alpha \neq 0$. $f(x)=\log(x)$, if $\alpha=0$. α est la spécification du 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-carré 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 γ structurel β . 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'épiceries 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 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 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 hessienne.** 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, l'ordre des variables de gauche à droite doit être : *RowType_*, *VarName_*, *P1*, *P2*, ..., où *RowType_* et *VarName_* sont des variables de chaîne et *P1*, *P2*, ... 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 Fichier scindé est en vigueur, les variables doivent commencer par la variable de fichier scindé ou les variables dans l'ordre indiqué lors de la création du fichier scindé, suivi de *RowType_*, *VarName_*, *P1*, *P2*, ... 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 variable *P1*, *P2*, ... ne sont pas requis ; la procédure accepte les noms de variables valides pour les paramètres car le mapping des variables aux paramètres est basé sur la position de la variable, et non sur le nom de la variable. 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-carré sélectionné, et sont basés sur l'hypothèse selon laquelle les paramètres suivent une distribution normale asymptotique.
- **Fonction de log-quasi-vraisemblance** : Contrôlez 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.
- **Fonction(s) d'estimation générale(s)** : 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 des paramètres pour chaque $n^{\text{ième}}$ itération en commençant par la $0^{\text{ème}}$ itération (les estimations initiales), où n est 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.
- *Difference.* 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 d'importance 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 seuil global de signification.
- *Sidak.* Cette méthode fournit des limites (bornes) 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 seuil 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 **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.

- *Résiduel brut*. Différence entre une 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-carré de Pearson, avec le signe du résidu brut.

Export d'équations d'estimation généralisées

Exporter le modèle en tant que données : Ecrit un jeu de données dans un format IBM SPSS Statistiques 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.

- **Variation de scission** : En cas d'utilisation, toutes les variables définissant des scissions.
- **RowType_**. Prend des valeurs (et des libellés de valeur) *COV* (covariances), *CORR* (corrélations), *EST* (estimations de paramètres), *SE* (erreurs standard), *SIG* (niveaux de signification) et *DF* (degrés de liberté de conception 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 estimés du modèle (à l'exception des paramètres d'échelle ou binomiaux négatifs), pour les types de ligne *COV* ou *CORR*, avec des étiquettes de valeur correspondant aux chaînes de paramètres indiquées 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 d'importance 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 d'importance 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.

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.

Options d'effets

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 la boîte de dialogue Variables sont utilisés pour définir des sujets pour la structure de covariance des résidus et composent la liste des champs possibles pour définition de sujets pour des structures de covariance des 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.

Options de covariance

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 répété

Indique la structure de covariance des résidus. Les différentes options de covariance sont disponibles en fonction du **type de covariance répétée**. Les structures disponibles sont :

- Autorégressif de premier ordre (AR1)
- Produit direct AR1 (UN_AR1)
- Produit direct non structuré (UN_UN)
- Symétrie composée de produit direct (UN_CS)
- Symétrie composée hétérogène (CSH)
- Autorégressive hétérogène (ARH1)
- Moyenne mobile autorégressive (1,1) (ARMA11)
- Symétrie composée
- Diagonale
- Identité mise à l'échelle
- Toeplitz
- Sans structure
- Composantes de variance
- Spatiale : Puissance
- Spatiale : Exponentielle
- Spatiale : de Gausse
- Spatiale : Linéaire
- Spatiale : Log linéaire
- Spatiale : Sphérique

Mesures de Kronecker

Sélectionnez des variables qui indiquent la structure des sujets des mesures de covariance de Kronecker et déterminent le mode de corrélation des erreurs de mesure. Ce champ est disponible uniquement lorsque l'une des options **Type de covariance répété** suivante est sélectionnée :

- Produit direct AR1 (UN_AR1)
- Produit direct non structuré (UN_UN)
- Symétrie composée de produit direct (UN_CS)

Coordonnées de covariance spatiale

Les variables de cette liste précisent les coordonnées des observations répétées lorsqu'un type de covariance spatiale est sélectionné comme type de covariance répétée.

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

Pseudo R² mesures

Les mesures de pseudo-R² et le coefficient de corrélation intra-classe sont inclus dans la sortie GLMM (le cas échéant). Les mesures pseudo-R² sont entièrement basées sur les estimations finales et sont produites une fois l'estimation terminée. Le coefficient de détermination R² est une statistique couramment rapportée, car il représente la proportion de variance expliquée par un modèle linéaire. Le coefficient de corrélation intra-classe (ICC) est une statistique connexe qui quantifie la proportion de variance expliquée par un facteur de regroupement (aléatoire) dans les données multiniveaux / hiérarchiques.

Obtention d'un modèle linéaire mixte généralisé

Cette fonction requiert SPSS Statistics Standard Edition ou 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**. Cet onglet contient des options permettant d'indiquer les effets de modèle et le type de covariance répété.
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 multicatégorie. 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** : $f(x) = x^\alpha$, if $\alpha \neq 0$. $f(x) = \log(x)$, if $\alpha = 0$. α est la spécification du 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.
- *****. La combinaison de tous les champs déposés apparaît comme une seule interaction 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	Descriptif
	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», à la page 58, 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», à la page 58 s'ouvre.
2. Pour éditer un bloc existant, sélectionnez le bloc que vous souhaitez éditer et cliquez sur **Éditer le bloc...** La boîte de dialogue «Bloc d'effets aléatoires», à la page 58 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 les ajoutant à la liste **Générateur d'effets**.

Le type d'effet dépend du **Type** que vous sélectionnez. Les champs qualitatifs (nominaux et ordinaux) sont utilisés comme facteurs dans le modèle et les champs continus sont utilisés comme covariables.

Interaction

La combinaison de tous les effets apparaît comme une interaction unique au bas de la liste des effets.

Effets principaux

Les champs déposés apparaissent sous forme d'effets principaux séparés au bas de la liste des effets.

Tous d'ordre 2

Toutes les paires possibles des champs déposés apparaissent sous forme d'interactions bidirectionnelles au bas de la liste des effets.

Tous d'ordre 3

Tous les triplets possibles des champs déposés apparaissent sous forme d'interactions tridirectionnelles au bas de la liste des effets.

Toutes d'ordre 4

Tous les triplets possibles des champs déposés apparaissent sous forme d'interactions d'ordre 4 au bas de la liste des effets.

Toutes d'ordre 5

Tous les triplets possibles des champs déposés apparaissent sous forme d'interactions d'ordre 5 au bas de la liste des effets.

Inclure la 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.

Afficher les prévisions de paramètres pour ce bloc

Affiche les estimations de paramètres à effets aléatoires.

Combinaison de sujet

Permet de spécifier des sujets à effets aléatoires à partir de combinaisons prédéfinies de sujets dans la boîte de dialogue Variables. Par exemple, si *Ecole*, *Classe* et *Etudiant* sont définis comme sujets et dans cet ordre dans la boîte de dialogue Variables, la liste déroulante Combinaison de sujet proposera les options **Aucune**, **Ecole**, **Ecole * Classe** et **Ecole * Classe * Etudiant**.

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

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.

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 γ structurel β . 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 55.

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 **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 du paramètre par défaut. Choisissez **Approximation de 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). Choisissez **Approximation de Kenward-Roger** si la taille de votre échantillon est petite et que vous disposez d'un modèle Maximum de vraisemblance restreint (REML).

Test des effets fixes et des 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 **Déviation** 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 seuil 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 seuil 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.

Valeurs prédites- Enregistre la valeur prédite de la cible. Le nom par défaut du champ est *PredictedValue*.

Probabilité prédite pour les cibles catégorielles- Si la cible est catégorielle, 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 bornes supérieure et inférieure de l'intervalle de confiance pour la valeur prédite ou 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 60). 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.

Résidus de Pearson- Enregistre le résidu de Pearson pour chaque enregistrement, qui peut être utilisé dans les diagnostics de post-estimation de l'ajustement du modèle. Le nom par défaut du champ est *PearsonResidual*.

Confiances- Enregistre la confiance dans la valeur prédite pour la cible catégorielle. 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*.

Exportation

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. Pour plus d'informations, voir la rubrique [Assistant de scoring](#) . Spécifiez un nom de fichier valide et unique. Si la spécification du fichier pointe vers un fichier existant, le fichier est écrasé.

Exportation de sortie

Exporter EBLUPS. Cette section est activée si au moins un bloc d'effets aléatoires est créé via la boîte de dialogue de l'élément Effets aléatoires de l'onglet Champs et effets et que la case à cocher permettant d'afficher les prévisions de paramètres pour ce bloc est sélectionnée. Si plusieurs blocs de ce type ont été spécifiés, vous pouvez choisir d'enregistrer les résultats dans des jeux de données ou des fichiers distincts, ou de les combiner dans un jeu de données ou un fichier unique.

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 Modèle, vous pouvez générer des tableaux croisés dynamiques et des graphiques en sélectionnant **Tableaux croisés et graphiques** dans le groupe Affichage de sortie de l'onglet Sortie de la boîte de dialogue Options (Édition > 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.* Une mesure pour la sélection et la comparaison de modèles mixtes basée sur la log-vraisemblance (restreinte) -2 . 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.* Une mesure de sélection et de comparaison des modèles basée sur la log-vraisemblance -2 . Les petites valeurs indiquent de meilleurs modèles. Le critère BIC (Bayesian Information Criterion) "pénalise" lui aussi les modèles surparamétrés (modèles complexes avec un nombre élevé d'entrées, par exemple), mais de façon plus stricte que le critère AIC (Akaike's Information Criterion).

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** .

- **Diagramme :** 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** .

- **Diagramme :** 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 coloriées 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 corrogramme correspondent aux valeurs des cellules telles qu'affichées dans la clé. Il s'agit de la valeur par défaut.
- **Corrogramme :** 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 affichés 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** .

- **Diagramme :** 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.

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 catégoriels. 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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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 résultant contiendra l'interaction tridirectionnelle spécifiée $A*B*C$, les interactions bidirectionnelles $A*B$, $A*C$ et $B*C$, et les principaux effets pour *A*, *B* et *C*. Ne spécifiez pas les éléments liés à l'ordre inférieur dans la classe de génération.

Pour les facteurs 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.

Tous d'ordre 2

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

Tous d'ordre 3

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

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 construits et termes personnalisés

Termes construits

Utilisez cette option lorsque vous souhaitez inclure des termes non imbriqués d'un certain type (tels que des effets principaux) pour toutes les combinaisons d'un ensemble de facteurs et de covariables.

Termes construits personnalisés

Utilisez cette option lorsque vous souhaitez inclure des termes imbriqués ou lorsque vous souhaitez générer explicitement un terme variable par variable. Le processus de construction d'un terme imbriqué implique les étapes suivantes :

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.

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 la cellule a une valeur égale à 0 ou 1. N'utilisez pas de variable de structure de cellule pour pondérer les 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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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

(En option) Vous avez également ces possibilités :

- 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 & 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.

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.

Tous d'ordre 2

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

Tous d'ordre 3

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

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 construits et termes personnalisés

Termes construits

Utilisez cette option lorsque vous souhaitez inclure des termes non imbriqués d'un certain type (tels que des effets principaux) pour toutes les combinaisons d'un ensemble de facteurs et de covariables.

Termes construits personnalisés

Utilisez cette option lorsque vous souhaitez inclure des termes imbriqués ou lorsque vous souhaitez générer explicitement un terme variable par variable. Le processus de construction d'un terme imbriqué implique les étapes suivantes :

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 le nombre maximal d'itérations, les critères de convergence et le delta (constante ajoutée à toutes les cellules pour les 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.

- *Résiduels.* Appelé également résidu simple ou brut, il indique la différence entre l'effectif observé et l'effectif théorique de la cellule.
- *Résiduels standardisés.* Résidu, divisé par une estimation de son erreur standard. Les résiduels standardisés sont aussi appelés résiduels de Pearson.

- *Résiduels ajustés*. Résidu standardisé divisé par l'erreur standard estimée. Comme les résidus ajustés sont asymptotiquement gaussiens, ils sont préférés aux résidus standardisés pour étudier la normalité, lorsque le modèle est correct.
- *Résiduels de déviance*. Racine carrée signée d'une contribution individuelle à la statistique du khi-carré 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.

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 la cellule a une valeur égale à 0 ou 1. N'utilisez pas de variable de structure de cellule pour pondérer les 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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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 variables actives doit être inférieur ou égal à 10.

(En option) Vous avez également ces possibilités :

- 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 & 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.

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.

Tous d'ordre 2

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

Tous d'ordre 3

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

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.

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 et termes personnalisés

Termes construits

Utilisez cette option lorsque vous souhaitez inclure des termes non imbriqués d'un certain type (tels que des effets principaux) pour toutes les combinaisons d'un ensemble de facteurs et de covariables.

Termes construits personnalisés

Utilisez cette option lorsque vous souhaitez inclure des termes imbriqués ou lorsque vous souhaitez générer explicitement un terme variable par variable. Le processus de construction d'un terme imbriqué implique les étapes suivantes :

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 le nombre maximal d'itérations, les critères de convergence et le delta (constante ajoutée à toutes les cellules pour les 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.

- *Résiduels*. Appelé également résidu simple ou brut, il indique la différence entre l'effectif observé et l'effectif théorique de la cellule.
- *Résiduels standardisés*. Résidu, divisé par une estimation de son erreur standard. Les résiduels standardisés sont aussi appelés résiduels de Pearson.
- *Résiduels ajustés*. Résidu standardisé divisé par l'erreur standard estimée. Comme les résidus ajustés sont asymptotiquement gaussiens, ils sont préférés aux résidus standardisés pour étudier la normalité, lorsque le modèle est correct.
- *Résiduels de déviance*. Racine carrée signée d'une contribution individuelle à la statistique du khi-carré 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.

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 employés 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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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 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.
- *Risque*. Affiche la fonction de risque cumulée sur 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.

Modèles de temps d'échec accéléré paramétriques

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

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

1. A partir des menus, sélectionnez :

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

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez une variable source.

Durée

Survie

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

Début/Fin

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

Statut

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

Défaillance/Événement

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

Censure à droite

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

Censure à gauche

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

Censure par intervalle

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

Traitement valeurs non mappées

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

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

Covariable(s)

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

Facteur(s) fixe(s)

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

Troncature à gauche

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

Modèles de temps d'échec accéléré paramétriques: Critères

Critères

Panneau facultatif permettant de spécifier les critères généraux.

Intervalle de confiance

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

Valeurs manquantes

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

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

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

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

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

Traitement du statut

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

Supprimer l'enregistrement en conflit

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

Obtenir les informations temporelles en fonction du statut

Obtient les informations d'heure en fonction du statut.

Dériver le statut en fonction des informations temporelles

Modifie le statut en fonction des informations d'heure.

Modèles de temps d'échec accéléré paramétriques: Modèle

Modèle

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

Répartition du temps de survie

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

Weibull

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

Exponentiel

Indique la distribution exponentielle.

Loi log-normale

Indique la distribution log-normale.

Log logistique

Indique la distribution log-logistique.

Paramètres de covariable

Spécifiez les variables de covariable.

Paramètres de facteur

Spécifiez les variables de facteur.

Valeur initiale de la constante

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

Valeur initiale du paramètre d'échelle

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

Erreur standard de la régression correspondante des MCO

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

Inverse l'erreur standard de la régression des MCO correspondante

Utilise la réciproque de l'erreur standard.

Valeur fournie par l'utilisateur

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

Modèles de temps d'échec accélérés paramétriques: estimation

Estimation

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

Méthode de direction alternée ou multipinces (ADMM)

Rapide

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

Traditionnel

Applique l'algorithme ADMM traditionnel.

Appliquer la régularisation L-1

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

Critères de convergence des modèles

Convergence des paramètres

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

Convergence des fonctions d'objectif

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

Convergence de Hess

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

Critères de convergence des résidus

Option permettant de contrôler le processus d'optimisation.

Résidu primaire et double

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

Résidu primaire uniquement

Applique le critère de convergence des résidus primaux.

Résidu double uniquement

Applique le critère de convergence des résidus duaux.

Méthode

Paramètre facultatif permettant de spécifier la méthode d'estimation.

Automatique

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

Newton-Raphson

Applique la méthode de Newton-Raphson.

L-BFGS

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

Itération

Nombre maximum d'itérations

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

Découpage maximal en deux des étapes

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

Nombre maximal de recherches de ligne

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

Convergence absolue pour le processus d'itération

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

Convergence relative pour le processus d'itération

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

Modèles de temps d'échec accéléré paramétriques: Imprimer

Imprimer

Panneau facultatif permettant de contrôler les sorties de la table.

Détails du codage de facteur

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

Valeurs initiales affectées aux paramètres de régression

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

Historique des itérations de modèle

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

Résultats de sélection contenant

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

Variables sélectionnées et non sélectionnées

Affiche les variables sélectionnées et non sélectionnées dans le tableau.

Uniquement les variables sélectionnées

Affiche uniquement les variables sélectionnées.

Uniquement les variables non sélectionnées

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

Modèles de temps d'échec accéléré paramétriques: Prévoir

Prévision

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

Valeurs de temps pour le scoring**Valeurs de temps définies par variable(s) dépendante(s)**

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

Intervalles ordinaires

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

Durée

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

Prévisions**Survie**

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

Risque

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

Risque cumulé

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

Survie conditionnelle

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

Modèles de temps d'échec accéléré paramétriques: Tracé

Tracé**Tracé résiduel de Cox-Snell**

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

Tracés de fonction

Option permettant de contrôler les tracés de fonction.

Type**Survie**

Crée le tracé des fonctions de survie.

Risque

Crée le tracé pour les fonctions de risque.

Densité

Crée un tracé pour les fonctions de densité.

Nombre de points à afficher

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

Valeurs de covariable pour le tracé

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

Valeurs de facteur pour le tracé

Facultatif permettant de spécifier les valeurs fournies par l'utilisateur et de les affecter aux prédicteurs. En présence de variables en double, celle spécifiée en premier est reconnue et le

reste est ignoré. Une variable valide doit être contenue dans un effet de modèle. L'omission d'une variable indique que la fréquence de la catégorie et la moyenne sont utilisées par défaut pour le facteur et la covariable, respectivement. Si une valeur non valide est affectée à une variable, le modèle demandé ne sera pas tracé.

Séparer les lignes pour

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

Nombre maximal de lignes dans un graphique

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

Modèles de temps d'échec accéléré paramétriques: Exportation

Exporter

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

Survie AFT Définir des événements pour les variables de statut

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

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

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

Sélection de la catégorie

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

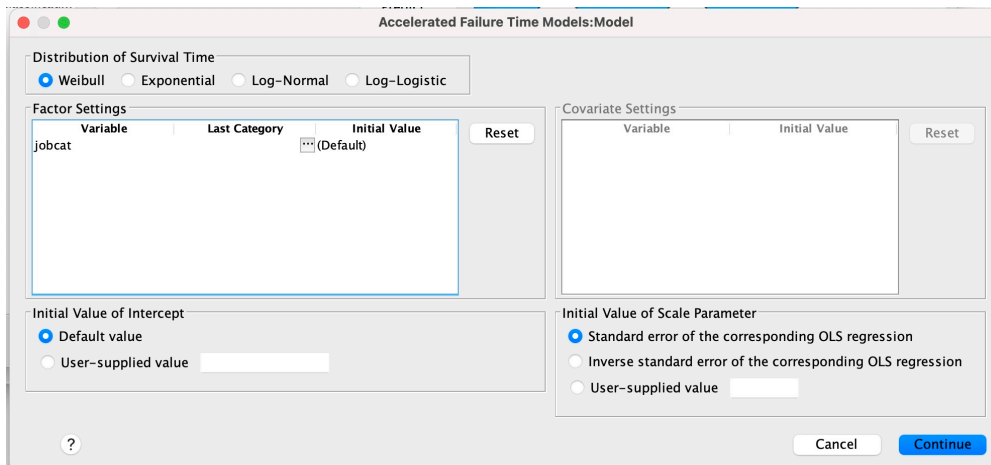


Figure 3. Modèles de durée de vie accélérée-Boîte de dialogue-Catégorie

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

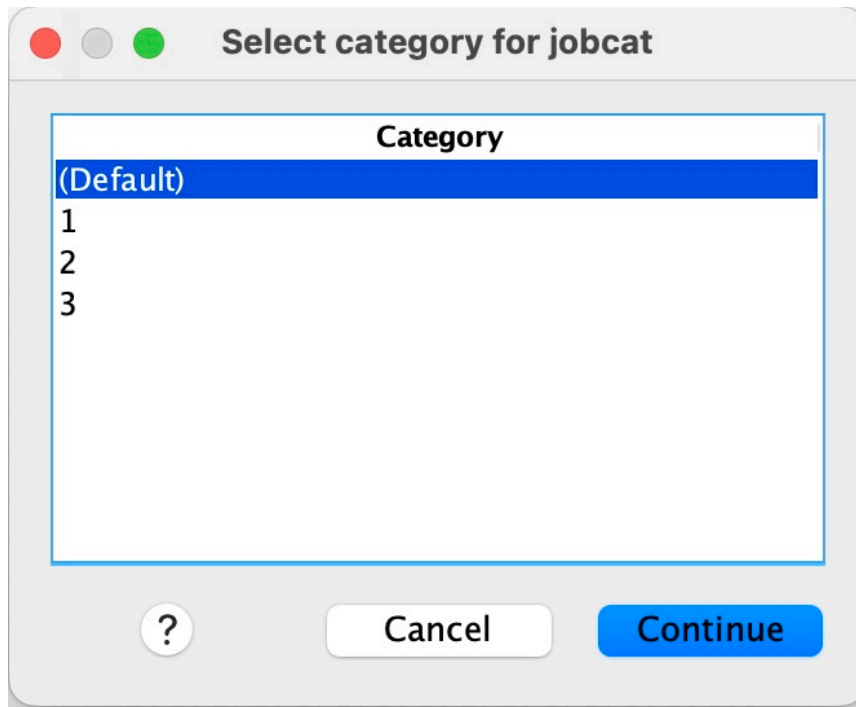


Figure 4. Modèles de durée de vie accélérée-Boîte de dialogue-Sélectionner une catégorie

Cliquez sur Continuer.

Modèles de fragilité partagés paramétriques

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

Obtention d'une analyse Parametric Shared Frailty Models

1. Dans les menus, sélectionnez :

Analyser > Survie > Modèles de fragilité partagés paramétriques

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez une variable source.

Durée

Survie

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

Début / Fin

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

Objet

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

Intervalle

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

Statut

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

Défaillance/Événement

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

Censure à droite

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

Traitement valeurs non mappées

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

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

Covariable(s)

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

Facteur(s) fixe(s)

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

Modèles de fragilité partagés paramétriques: Critères

Critères

Panneau facultatif permettant de spécifier les critères généraux.

Intervalle de confiance

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

Niveau d'importance

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

Valeurs manquantes

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

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

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

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

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

Traitement de l'intervalle

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

Supprimer les enregistrements en conflit

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

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

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

Modèles de fragilité partagés paramétriques: Modèle

Modèle

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

Répartition du temps de survie

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

Weibull

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

Exponentiel

Indique la distribution exponentielle.

Loi log-normale

Indique la distribution Log-normal.

Log logistique

Indique la distribution log-logistique.

Paramètres de covariable

Spécifiez les variables de covariable.

Paramètres de facteur

Spécifiez les variables de facteur.

Valeur initiale de la constante

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

Valeur initiale du paramètre d'échelle

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

Erreur standard de la régression correspondante des MCO

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

Inverse l'erreur standard de la régression des MCO correspondante

Utilise la réciproque de l'erreur standard.

Valeur fournie par l'utilisateur

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

Composant fragilité

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

Gamma

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

Inverse-Gaussien

Indique la distribution gaussienne inverse.

Valeur initiale de la variance

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

Modèles de fragilité partagés paramétriques: estimation

Estimation

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

Méthode de direction alternée ou multipinces (ADMM)

Rapide

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

Traditionnel

Applique l'algorithme ADMM traditionnel.

Appliquer la régularisation L-1

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

Critères de convergence des modèles

Convergence des paramètres

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

Convergence des fonctions d'objectif

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

Convergence de Hess

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

Critères de convergence des résidus

Option permettant de contrôler le processus d'optimisation.

Résidu primaire et double

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

Résidu primaire uniquement

Applique le critère de convergence des résidus primaux.

Résidu double uniquement

Applique le critère de convergence des résidus duaux.

Méthode

Paramètre facultatif permettant de spécifier la méthode d'estimation.

Automatique

Sélectionne automatiquement la méthode en fonction de l'exemple de jeu de données. Cette méthode est sélectionnée par défaut. Le champ **Nombre de prédicteurs de seuil** indique le seuil

du nombre de prédicteurs et doit être un entier unique supérieur à 1. La valeur par défaut est de 1000.

Newton-Raphson

Applique la méthode de Newton-Raphson.

L-BFGS

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

Itération

Nombre maximum d'itérations

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

Découpage maximal en deux des étapes

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

Nombre maximal de recherches de ligne

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

Convergence absolue pour le processus d'itération

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

Convergence relative pour le processus d'itération

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

Modèles de fragilité partagés paramétriques: Imprimer

Imprimer

Panneau facultatif qui contrôle les sorties de la table.

Détails du codage de facteur

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

Valeurs initiales affectées aux paramètres de régression

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

Historique des itérations de modèle

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

Modèles de fragilité partagés paramétriques: Prévoir

Prévision

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

Valeurs de temps pour le scoring

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

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

Intervalles ordinaires

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

Durée

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

Prévisions

Survie

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

Risque

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

Risque cumulé

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

Survie conditionnelle

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

Survie inconditionnelle

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

Risque inconditionnel

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

Risque de cumul inconditionnel

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

Modèles de fragilité partagés paramétriques: Tracé

Tracé

Tracés de fonction

Option permettant de contrôler les tracés de fonction.

Type

Survie

Crée le tracé pour les fonctions de survie inconditionnelles.

Risque

Crée le tracé des fonctions de risque inconditionnelles.

Densité

Crée un tracé pour les fonctions de densité.

Nombre de points à afficher

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

Valeurs de covariable pour le tracé

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

Valeurs de facteur pour le tracé

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

Séparer les lignes pour

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

Nombre maximal de lignes dans un graphique

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

Modèles de fragilité partagés paramétriques: Exportation

Exporter

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

Modèles de fragilité partagés paramétriques: Définir des événements

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

1. Dans le menu, choisissez:

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

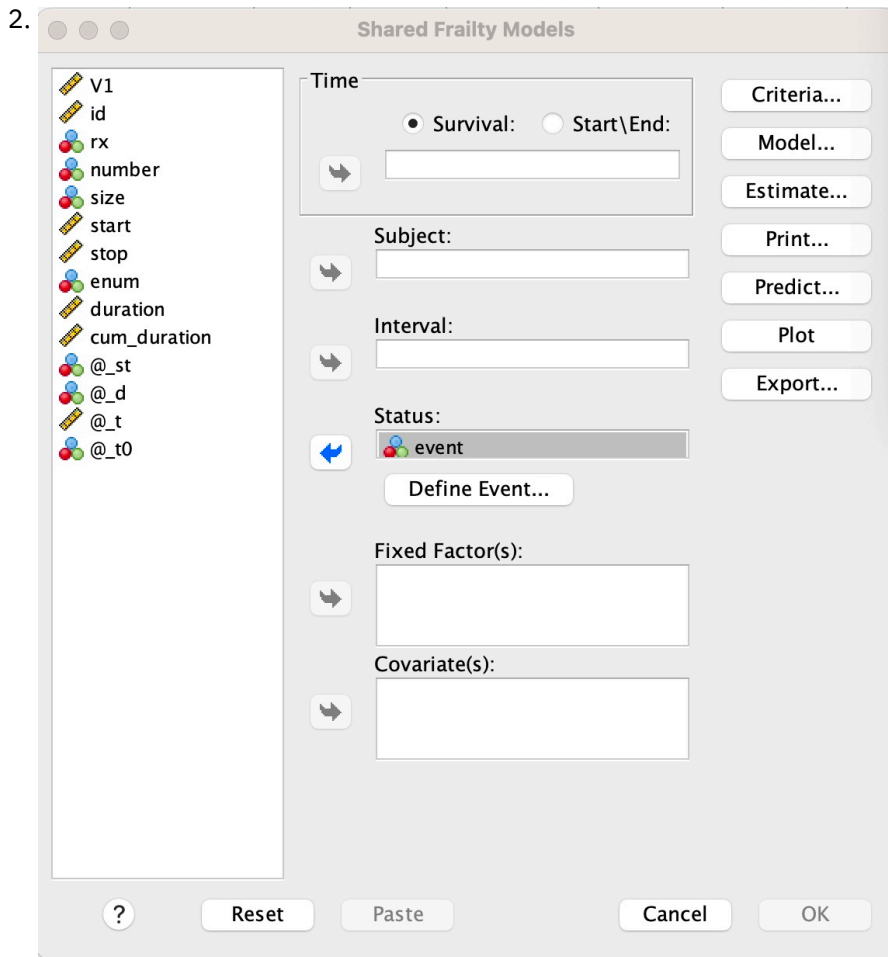


Figure 5. Modèles de fragilité partagés-boîte de dialogue-Statut

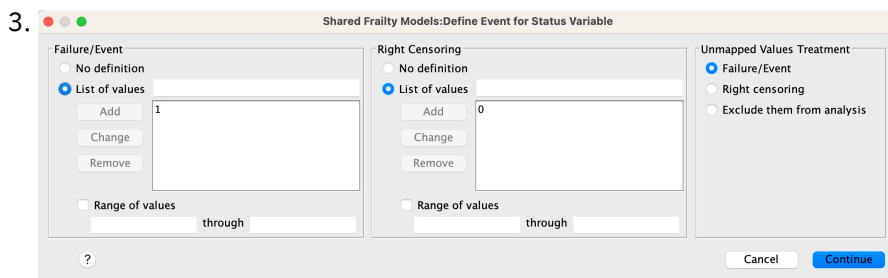


Figure 6. Modèles de fragilité partagés-boîte de dialogue-Statut-Définir un événement

Modèles de fragilité partagés paramétriques-Exemples

Exemple 1

SURVREG RECURRENT y AVEC x1 PAR x2

```
/MODEL SUBJECT = id FRAILTY=GAMMA DISTRIBUTION=WEIBULL.
```

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

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

Les sujets sont identifiés par l'ID variable.

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

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

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

Exemple 2

```
SURVREG RECURRENT y AVEC x1 PAR x2
```

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

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

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

Les sujets sont identifiés par l'ID variable.

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

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

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

Exemple 3

```
SURVREG RÉCURRENT y1 y2 AVEC x1 PAR x2(1 )
```

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

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

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

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

Les sujets sont identifiés par l'ID variable.

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

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

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

Exemple 4

```
SURVREG RECURRENT y AVEC x1 PAR x2
```

```
/MODEL SUBJECT = id
```

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

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

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

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

Les sujets sont identifiés par l'ID variable.

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

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

Exemple 5

SURVREG RECURRENT y AVEC x1 PAR x2

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

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

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

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

Les sujets sont identifiés par l'ID variable.

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

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

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

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

Exemple 6

SURVREG RECURRENT y AVEC x1 PAR x2

```
/MODEL SUBJECT = id
```

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

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

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

Les sujets sont identifiés par l'ID variable.

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

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

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

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

Acteurs-Chercheur et praticien en santé publique.

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

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

- patID: Numéro d'identification permettant d'identifier un participant unique.
- endTime: Durée de survie (en jours) de l'effet secondaire, à la suite d'un traitement, mesurée à partir du début du traitement jusqu'à la déclaration d'un effet secondaire ou la censure dans les 60 jours.
- sideEffect: Statut de l'effet secondaire, statut = 0 si censuré et statut = 1 si l'effet latéral léger est signalé.
- âge: âge du participant à la période de recherche.
- femelle: femelle = 0 si mâle et femelle = 1 si femelle.

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

Syntaxe-

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

Synopsis :

La syntaxe spécifiée par Patrick désigne endTime comme une variable de temps dépendante unique. La procédure suppose automatiquement que l'heure de début est 0 pour chaque enregistrement. Les variables âge et femme sont modélisées comme covariable et facteur, respectivement. Les temps de survie de récurrence sont supposés suivre une distribution de Weibull. Le terme de fragilité non observé est supposé suivre une distribution Gamma, et sa composante de variance est modélisée. En ce qui

concerne les sorties, le tableau récapitulatif du modèle fournit les informations sur la procédure et le modèle. Le tableau récapitulatif de traitement des cas fournit une liste complète du statut d'échec ou de censure, ainsi que les cas exclus de l'analyse.

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

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

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...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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.

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.

- *Rang du journal*. 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.
- *Regroupé en pool sur des strates*. Compare tous les niveaux de facteurs d'un test unique afin de tester l'égalité des courbes de survie.
- *Appariés sur des strates*. Compare chaque paire de niveaux de facteurs. Les tests de tendance par paires ne sont pas disponibles.
- *Pour chaque strate*. Exécute un test distinct d'égalité de tous les niveaux de facteur pour chaque strate. Si vous ne disposez pas d'une variable de stratification, les tests ne sont pas effectués.
- *Appariés pour 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.

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*. Estimation 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*.
- *Risque*. Estimation de la fonction de risque cumulé. 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*.
- *Événements cumulés*. Fréquence cumulée des événements lorsque les observations sont triées selon 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.
- *Risque*. Affiche la fonction de risque cumulée sur 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.

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-carré 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-carré 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 connu 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 :

Analyser > Survie > Régression de Cox...

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

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 inclure des termes d'interaction, sélectionnez toutes les variables impliquées dans l'interaction, puis cliquez sur **> 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 catégorielles.

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

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

- **Indicateur** : Les contrastes indiquent la présence ou l'absence d'appartenance à la catégorie. La catégorie de référence est représentée par la matrice de contraste sous la forme d'une ligne de zéros.
- **Simple** : Chaque catégorie de la variable de prédicteur (hormis la catégorie de référence) est comparée 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.
- **Déviaton** : Chaque catégorie de la variable de prédicteur (hormis la catégorie de référence) est comparée à l'effet global.

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

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

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.

Type de tracé

Survie

Affiche la fonction de survie cumulée d'après une échelle linéaire.

Risque

Affiche la fonction de risque cumulée sur 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.

Tracé des valeurs de covariable

Vous pouvez tracer une courbe distincte pour chaque valeur de la covariable catégorielle en déplaçant cette covariable vers la zone de texte **Séparer les lignes pour**. Cette option est disponible uniquement pour les covariables catégorielles, indiquées par **(Cat)** après leur nom dans la liste **Tracé des valeurs de covariable**.

Modifier une valeur

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

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 pour un temps donné. Elle correspond à la probabilité de survie à cette période.

Erreur standard de la fonction de survie

Erreur standard de l'estimation de survie cumulée.

Fonction de survie LN (-Logn)

Estimation de survie cumulée après application de la transformation $\ln(-\ln)$ à l'estimation.

Fonction de risque

Enregistre l'estimation de la fonction de risque cumulée (également appelée résidu de Cox-Snell).

Résidus partiels

Vous pouvez tracer les résidus 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.

DfBeta(s)

Variation estimée d'un coefficient si une observation est supprimée. 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 de prédicteur linéaire. Somme du produit des valeurs de covariable centrée sur la moyenne et des estimations de paramètre correspondantes, pour chaque observation.

Remarque : Seules les différences de bêta sont sauvegardées si vous exécutez Cox avec une covariable chronologique.

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 dans 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 Statistiques 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.

Calcul des prédicteurs chronologiques

Dans certaines situations, vous souhaiteriez calculer un modèle de régression Cox, mais l'hypothèse des risques proportionnels ne tient pas. 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 un modèle de régression étendu Cox, dans lequel vous pouvez spécifier des **covariables dépendant du temps**.

Pour analyser un tel modèle, vous devez d'abord définir la covariable avec contrainte horaire. Pour faciliter cela, une "variable système" représentant le temps est disponible. Cette variable est appelée $T_.$. Vous pouvez utiliser cette variable pour définir des covariables chronologiques de deux manières générales :

- Pour tester l'hypothèse des risques proportionnels en ce qui concerne une covariable particulière ou estimer un modèle de régression étendu Cox qui autorise les risques non proportionnels, vous devez définir la covariable dépendante du temps comme une fonction de la variable 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. Le test de la signification du coefficient de la covariable dépendante du temps indique si l'hypothèse des risques proportionnels est raisonnable.
- Certaines variables peuvent avoir des valeurs différentes à des périodes différentes, mais ne sont pas systématiquement liées au temps. Dans de tels cas, vous devez définir une **covariable dépendante du temps segmentée**, qui peut être effectuée à l'aide d' **expressions logiques**. 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 vous avez une tension artérielle mesurée une fois par semaine pour les quatre semaines de votre étude (identifiée comme $BP1$ à $BP4$), vous pouvez définir la covariable avec contrainte horaire comme $(T_ < 1) * BP1 + (T_ >= 1 \& T_ < 2) * BP2 + (T_ >= 2 \& T_ < 3) * BP3 + (T_ >= 3 \& T_ < 4) * BP4$. Vous remarquez qu'un seul des termes entre parenthèses est égal à 1 pour chaque cas spécifique et que le reste sera égal à 0. Pour résumer, cette fonction signifie que si le temps est inférieur à une semaine, utilisez $BP1$; s'il est supérieur à une semaine mais inférieur à deux semaines, utilisez $BP2$, etc.

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

Dans la boîte de dialogue Calculer la covariable dépendante du temps, vous pouvez utiliser les contrôles de génération de fonction pour générer l'expression pour une covariable dépendante du temps, ou vous pouvez l'entrer directement dans la zone de texte Nom. 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. Toutes les covariables dépendantes du temps qui en résultent doivent être incluses en tant que covariables dans votre modèle de régression Cox.

Calculer un prédicteur chronologique

1. Dans le menu, choisissez

Analyser > Survie > Cox av. prédicteur chronologique...

2. Entrez une expression pour le prédicteur chronologique.

3. Sélectionnez **Modèle** pour effectuer votre régression de Cox.

Remarque - Veillez à inclure les nouvelles variables ajoutées en tant que covariables dans votre modèle de régression Cox.

Méthodes de codification des variables catégorielles

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

Déviaton

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

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

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

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

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

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

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

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

Simple

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

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

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

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

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

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

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

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

Helmert

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

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

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

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

Différence

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

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

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

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

Polynomial

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

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

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

est identique à

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

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

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

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

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

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

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

Répété

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

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

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

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

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

Spécial

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

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

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

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

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

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

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

Pour LOGLINEAR, vous devez indiquer :

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

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

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

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

Indicateur

Codification des variables indicateur : Egalement appelé codification factice, ce type de codification n'est pas disponible dans LOGLINEAR ni MANOVA. Le numéro des nouvelles variables codées est $k-1$. Les observations de la catégorie de référence sont codées 0 pour toutes les variables $k-1$. Une observation dans la $n^{\text{ième}}$ catégorie est codée 0 pour toutes les variables indicateur, sauf la $n^{\text{ième}}$, codée 1.

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 < rho < 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.

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

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

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	$\sigma_4\sigma_3\rho)$
$(\sigma_4\sigma_1\rho$	$\sigma_4\sigma_2\rho$	$\sigma_4\sigma_3\rho$	$\sigma_4^2)$

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	$\sigma_4^2)$

Produit direct AR1 (UN_AR1). Spécifie le produit de Kronecker d'une matrice non structurée et d'une autre matrice de covariance autorégressive de premier ordre. La première matrice non structurée modélise l'observation multivariée et la seconde structure de covariance autorégressive de premier ordre modélise la covariance des donnée dans le temps ou selon un autre facteur.

Produit direct non structuré (UN_UN). Spécifie le produit de Kronecker de deux matrices non structurées, la première modélisant l'observation multivariée et la seconde modélisant la covariance des données dans le temps ou selon un autre facteur.

Symétrie composée de produit direct (UN_CS). Spécifie le produit de Kronecker d'une matrice non structurée et d'une autre matrice de covariance à symétrie composée avec une variance et une covariance constantes. La première matrice non structurée modélise l'observation multivariée et la seconde structure de covariance à symétrie composée modélise la covariance des données dans le temps ou selon un autre facteur.

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 Γ ; circulaire Γ ; 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.

$$\begin{pmatrix} \sigma_1^2 & [\sigma_1^2 + \sigma_2^2]/2 - \lambda & [\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_1^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_2^2]/2 - \lambda & \sigma_2^2 & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & \sigma_3^2 & [\sigma_3^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_4^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda & [\sigma_3^2 + \sigma_4^2]/2 - \lambda & \sigma_4^2 \end{pmatrix}$$

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

$$\begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

Spatial : Puissance. Cette structure de covariance comporte des variances et des corrélations hétérogènes entre les éléments. d_{ij} est la distance euclidienne estimée entre la $i^{\text{ème}}$ et la $j^{\text{ème}}$ mesure.

$$\begin{pmatrix} \sigma^2 & \sigma^2 \rho^{d_{12}} & \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{14}} \\ \sigma^2 \rho^{d_{12}} & \sigma^2 & \sigma^2 \rho^{d_{23}} & \sigma^2 \rho^{d_{24}} \\ \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{23}} & \sigma^2 & \sigma^2 \rho^{d_{34}} \\ \sigma^2 \rho^{d_{14}} & \sigma^2 \rho^{d_{24}} & \sigma^2 \rho^{d_{34}} & \sigma^2 \end{pmatrix}$$

Spatial : Exponentiel. Cette structure de covariance comporte des variances et des corrélations hétérogènes entre les éléments. d_{ij} est la distance euclidienne estimée entre la $i^{\text{ème}}$ et la $j^{\text{ème}}$ mesure.

$$\begin{pmatrix} \sigma^2 & \sigma^2 \exp\{-d_{12}/\theta\} & \sigma^2 \exp\{-d_{13}/\theta\} & \sigma^2 \exp\{-d_{14}/\theta\} \\ \sigma^2 \exp\{-d_{12}/\theta\} & \sigma^2 & \sigma^2 \exp\{-d_{23}/\theta\} & \sigma^2 \exp\{-d_{24}/\theta\} \\ \sigma^2 \exp\{-d_{13}/\theta\} & \sigma^2 \exp\{-d_{23}/\theta\} & \sigma^2 & \sigma^2 \exp\{-d_{34}/\theta\} \\ \sigma^2 \exp\{-d_{14}/\theta\} & \sigma^2 \exp\{-d_{24}/\theta\} & \sigma^2 \exp\{-d_{34}/\theta\} & \sigma^2 \end{pmatrix}$$

Spatial : Gaussien. Cette structure de covariance comporte des variances et des corrélations hétérogènes entre les éléments. d_{ij} est la distance euclidienne estimée entre la $i^{\text{ème}}$ et la $j^{\text{ème}}$ mesure.

$$\begin{pmatrix} \sigma^2 & \sigma^2 \exp\{-d_{12}/\rho^2\} & \sigma^2 \exp\{-d_{13}/\rho^2\} & \sigma^2 \exp\{-d_{14}/\rho^2\} \\ \sigma^2 \exp\{-d_{12}/\rho^2\} & \sigma^2 & \sigma^2 \exp\{-d_{23}/\rho^2\} & \sigma^2 \exp\{-d_{24}/\rho^2\} \\ \sigma^2 \exp\{-d_{13}/\rho^2\} & \sigma^2 \exp\{-d_{23}/\rho^2\} & \sigma^2 & \sigma^2 \exp\{-d_{34}/\rho^2\} \\ \sigma^2 \exp\{-d_{14}/\rho^2\} & \sigma^2 \exp\{-d_{24}/\rho^2\} & \sigma^2 \exp\{-d_{34}/\rho^2\} & \sigma^2 \end{pmatrix}$$

Spatial : Linéaire. Cette structure de covariance comporte des variances et des corrélations hétérogènes entre les éléments. d_{ij} est la distance euclidienne estimée entre la $i^{\text{ème}}$ et la $j^{\text{ème}}$ mesure, et 1_{ij} est une fonction indicatrice qui vaut 1 si $\rho d_{ij} \leq 0$ et 0 sinon.

$$\begin{pmatrix} \sigma^2 & \sigma^2(1 - \rho d_{12}) 1_{12} & \sigma^2(1 - \rho d_{13}) 1_{13} & \sigma^2(1 - \rho d_{14}) 1_{14} \\ \sigma^2(1 - \rho d_{12}) 1_{12} & \sigma^2 & \sigma^2(1 - \rho d_{23}) 1_{23} & \sigma^2(1 - \rho d_{24}) 1_{24} \\ \sigma^2(1 - \rho d_{13}) 1_{13} & \sigma^2(1 - \rho d_{23}) 1_{23} & \sigma^2 & \sigma^2(1 - \rho d_{34}) 1_{34} \\ \sigma^2(1 - \rho d_{14}) 1_{14} & \sigma^2(1 - \rho d_{24}) 1_{24} & \sigma^2(1 - \rho d_{34}) 1_{34} & \sigma^2 \end{pmatrix}$$

Spatial : Log linéaire. Cette structure de covariance comporte des variances et des corrélations hétérogènes entre les éléments. d_{ij} est la distance euclidienne estimée entre la $i^{\text{ème}}$ et la $j^{\text{ème}}$ mesure, et 1_{ij} est une fonction indicatrice qui vaut 1 si $\rho \log(d_{ij}) \leq 0$ et 0 sinon.

$$\begin{pmatrix} \sigma^2 & \sigma^2(1 - \rho \log(d_{12})) 1_{12} & \sigma^2(1 - \rho \log(d_{13})) 1_{13} & \sigma^2(1 - \rho \log(d_{14})) 1_{14} \\ \sigma^2(1 - \rho \log(d_{12})) 1_{12} & \sigma^2 & \sigma^2(1 - \rho \log(d_{23})) 1_{23} & \sigma^2(1 - \rho \log(d_{24})) 1_{24} \\ \sigma^2(1 - \rho \log(d_{13})) 1_{13} & \sigma^2(1 - \rho \log(d_{23})) 1_{23} & \sigma^2 & \sigma^2(1 - \rho \log(d_{34})) 1_{34} \\ \sigma^2(1 - \rho \log(d_{14})) 1_{14} & \sigma^2(1 - \rho \log(d_{24})) 1_{24} & \sigma^2(1 - \rho \log(d_{34})) 1_{34} & \sigma^2 \end{pmatrix}$$

Spatial : Sphérique. Cette structure de covariance comporte des variances et des corrélations hétérogènes entre les éléments. $r_{ij} = d_{ij}/\rho$, où d_{ij} est la distance euclidienne estimée entre la $i^{\text{ème}}$ et la $j^{\text{ème}}$ mesure. 1_{ij} représente une fonction d'indicateur égale à 1 si $d_{ij} \leq \rho$ et à 0 dans le cas contraire.

$$\begin{pmatrix} \sigma^2 & \sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12} & \sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13} & \sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14} \\ \sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12} & \sigma^2 & \sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23} & \sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24} \\ \sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13} & \sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23} & \sigma^2 & \sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34} \\ \sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14} & \sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24} & \sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34} & \sigma^2 \end{pmatrix}$$

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.

$$\begin{pmatrix} \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 & \sigma^2\rho_3 \\ \sigma^2\rho_1 & \sigma^2 & \sigma^2\rho_1 & \sigma^2\rho_2 \\ \sigma^2\rho_2 & \sigma^2\rho_1 & \sigma^2 & \sigma^2\rho_1 \\ \sigma^2\rho_3 & \sigma^2\rho_2 & \sigma^2\rho_1 & \sigma^2 \end{pmatrix}$$

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.

$$\begin{pmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_2 & \sigma_4\sigma_1\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_1 & \sigma_4\sigma_2\rho_2 \\ \sigma_3\sigma_1\rho_2 & \sigma_3\sigma_2\rho_1 & \sigma_3^2 & \sigma_4\sigma_3\rho_1 \\ \sigma_4\sigma_1\rho_3 & \sigma_4\sigma_2\rho_2 & \sigma_4\sigma_3\rho_1 & \sigma_4^2 \end{pmatrix}$$

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

$$\begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

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	$\sigma_4\sigma_3\rho_{43})$
$(\sigma_4\sigma_1\rho_{41}$	$\sigma_4\sigma_2\rho_{42}$	$\sigma_4\sigma_3\rho_{43}$	$\sigma_4^2)$

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

Statistiques de Bayes

IBM SPSS Statistiques prend en charge les statistiques bayésiennes suivantes.

Tests T à 1 échantillon et à échantillons appariés

La procédure Inférence bayésienne à 1 échantillon propose des options pour réaliser une inférence bayésienne sur test T à 1 échantillon et test T à échantillons appariés en caractérisant les distributions a posteriori. Lorsque vos données relèvent d'une distribution normale, vous pouvez utiliser une distribution a priori normale pour obtenir une distribution a posteriori normale.

Tests de proportion binomiale

La procédure Inférence bayésienne à 1 échantillon : distribution binomiale propose des options pour réaliser une inférence bayésienne à 1 échantillon sur une distribution binomiale. Le paramètre étudié est π , lequel dénote la probabilité de succès dans un nombre d'essais fixe pouvant conduire à un succès ou à un échec. Notez que chaque essai est indépendant des autres et que la probabilité π demeure identique dans chaque essai. Une variable aléatoire binomiale peut être considérée comme étant la somme d'un nombre fixé d'essais de Bernoulli indépendants.

Analyse de la distribution de Poisson

La procédure Inférence bayésienne à 1 échantillon : distribution de Poisson propose des options pour réaliser une inférence bayésienne à 1 échantillon sur une distribution de Poisson. La distribution de Poisson, qui constitue un modèle utile pour des événements rares, suppose qu'à de courts intervalles de temps, la probabilité qu'un événement se produise est proportionnelle à la longueur de l'attente. Une probabilité a priori conjuguée de la famille de distribution Gamma est utilisée lors du calcul de l'inférence statistique de Bayes sur la distribution de Poisson.

Echantillons associés

La conception de l'inférence d'échantillon associé à Bayes est assez similaire à celle de l'inférence de Bayes à un seul échantillon en termes de traitement des échantillons de paires. Vous pouvez spécifier les noms de variable par paires et exécuter l'analyse de Bayes sur la différence moyenne.

Tests T d'échantillons indépendants

La procédure d'inférence de Bayes d'échantillon indépendant propose des options pour utiliser une variable de groupe afin de définir deux groupes sans lien et pour dégager une inférence de Bayes sur la différence entre les moyennes des deux groupes. Vous pouvez estimer les facteurs de Bayes en utilisant différentes approches, et aussi caractériser la distribution a posteriori désirée en supposant que les variances sont connues ou inconnues.

Corrélation par paires (Pearson)

L'inférence bayésienne sur le coefficient de corrélation de Pearson mesure la relation linéaire entre deux variables d'échelle suivant conjointement une distribution normale bivariée. L'inférence statistique classique sur le coefficient de corrélation a été largement discutée et son utilisation est pratiquée de longue date dans IBM SPSS Statistics. La conception de l'inférence bayésienne sur le coefficient de corrélation de Pearson vous permet de générer une inférence bayésienne en estimant les facteurs de Bayes et en caractérisant des distributions a posteriori.

Régression linéaire

L'inférence bayésienne sur régression linéaire est une méthode statistique couramment utilisée en modélisation quantitative. La régression linéaire est une approche de base standard où les chercheurs utilisent les valeurs de plusieurs variables pour expliquer ou prédire les valeurs d'une échelle de

résultats. La régression linéaire univariée de Bayes est une approche de la régression linéaire où l'analyse statistique est réalisée dans le contexte de l'inférence bayésienne.

ANOVA unidirectionnelle

La procédure ANOVA à 1 facteur de Bayes génère une analyse de variance à un facteur pour une variable dépendante quantitative par une seule variable de facteur (indépendante). L'analyse de variance sert à tester l'hypothèse d'égalité des moyennes. SPSS Statistics prend en charge les facteurs de Bayes, les distributions a priori conjuguées et les distributions a priori non informatives.

Modèles de régression Log Linéaire

La conception du test d'indépendance de deux facteurs nécessite deux variables catégorielles pour la construction d'un tableau de contingence et génère une inférence bayésienne sur l'association ligne-colonne. Vous pouvez estimer les facteurs de Bayes en supposant différents modèles et en caractérisant la distribution a posteriori désirée en simulant l'intervalle crédible simultané pour les termes d'interaction.

ANOVA à 1 facteur avec mesures répétées

La procédure ANOVA bayésienne à un facteur avec mesures répétées mesure un seul facteur sur un même sujet à chaque point temporel ou condition et permet de croiser des sujets à l'intérieur des niveaux. Il est supposé que chaque sujet est associé à une observation unique pour chaque point temporel ou condition (dans ce cas, l'interaction sujet-traitement d'interaction n'est pas prise en compte).

Inférence bayésienne à échantillon unique : Normale

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

La procédure Inférence bayésienne à échantillon unique : distribution normale propose des options pour réaliser une inférence bayésienne sur test T à 1 échantillon et test T à échantillons appariés en caractérisant les distributions a posteriori. Lorsque vos données relèvent d'une distribution normale, vous pouvez utiliser une distribution a priori normale pour obtenir une distribution a posteriori normale.

1. A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > Distribution normale à échantillons indépendants

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez les **variables de test** appropriées dans la liste **Variables disponibles**. Vous devez sélectionner au moins une variable.

Remarque : La liste des variables disponibles recense toutes celles disponibles, excepté celles de type Date et Chaîne.

3. Sélectionnez l'**Analyse de Bayes** voulue :

- **Caractériser la distribution a posteriori** : lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.
- **Estimer le facteur de Bayes** : lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

Tableau 2. Seuils couramment utilisés pour définir la signification des constatations

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H0 : hypothèse nulle

H1 : hypothèse alternative

1

2

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.
4. Sélectionnez et/ou entrez les paramètres **Valeurs de variance de données et d'hypothèse** appropriés. Le tableau reflète les variables figurant actuellement dans la liste **Variables de test**. Lorsque des variables sont ajoutées ou retirées de la liste **Variables à tester**, le tableau ajoute ou retire automatiquement ces mêmes variables de ses colonnes de variables.

- Lorsqu'une ou plusieurs variables sont présentes dans la liste **Variables de test**, les colonnes **Variable connue** et **Valeur de la variance** sont activées.

Variance connue

Sélectionnez cette option pour chaque variable dont la variance est connue.

Valeur de la variance

Paramètre facultatif qui spécifie la valeur de la variance (si elle est connue) des données des observations.

- Lorsqu'une ou plusieurs variables sont présentes dans la liste **Variables de test** et que l'option **Caractériser la distribution a posteriori** n'est pas sélectionnée, les colonnes **Valeur de test Null** et **Valeur g** sont activées.

Valeur de test nulle.

Paramètre obligatoire spécifiant la valeur Null dans l'estimation du facteur de Bayes. Une seule valeur est autorisée et 0 est la valeur par défaut.

Valeur g

Spécifie la valeur pour définir $\psi^2 = g\sigma^2_x$ dans l'estimation du facteur de Bayes. Lorsque la **Valeur de variance** est spécifiée, la valeur par défaut de la **Valeur g** est 1. Lorsque la **Valeur de variance** n'est pas spécifiée, vous pouvez spécifier une valeur g fixe ou omettre la valeur pour ne pas l'intégrer.

¹ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

² Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

5. Vous pouvez éventuellement cliquer sur **Critères** pour spécifier des paramètres pour «Inférence bayésienne à échantillon unique : Critères», à la page 113 (pourcentage d'intervalle crédible, options pour valeurs manquantes et paramètres de méthode numérique) ou sur **Distributions a priori** pour spécifier des paramètres pour «Inférence bayésienne à échantillon unique : Probabilités a priori pour distribution normale», à la page 114 (types de distributions a priori, tels que paramètres d'inférence, variance de moyenne indiquée ou précision).

Inférence bayésienne à échantillon unique : Critères

Vous pouvez définir les critères d'analyse suivants pour votre inférence bayésienne à échantillon unique :

Pourcentage d'intervalle crédible

Spécifiez le niveau de signification pour les intervalles de calcul crédibles. Niveau par défaut : 95 %.

Valeurs manquantes

Spécifiez la méthode de contrôle des valeurs manquantes.

Exclure seulement les composantes non valides

Il s'agit du paramètre par défaut, lequel exclut les enregistrements avec des valeurs manquantes sur une base analyse par analyse. Les enregistrements qui incluent des valeurs manquantes pour une zone utilisée pour un test spécifique sont omis du test.

Exclure toute observation incomplète

Ce paramètre exclut les enregistrements incluant des valeurs manquantes sur la base d'une liste. Les enregistrements qui incluent des valeurs manquantes pour une zone quelconque nommée dans une sous-commande sont exclus de toutes les analyses.

Remarque : Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour **Analyse de Bayes**.

Méthode numérique

Spécifiez la méthode numérique à utiliser pour estimation de l'intégrale.

Quadrature adaptative de Gauss-Lobatto

Il s'agit du paramètre par défaut, lequel invoque l'approche Quadrature adaptative de Gauss-Lobatto.

Tolérance

Spécifiez la valeur de tolérance pour les méthodes numériques. Valeur par défaut : 0.000001. Cette option n'est disponible que si le paramètre **Quadrature adaptative de Gauss-Lobatto** est sélectionné.

Nombre maximum d'itérations

Spécifiez le nombre maximal d'itérations de la méthode Quadrature adaptative de Gauss-Lobatto. Il doit s'agir d'un nombre entier positif. Valeur par défaut : 2000. Cette option n'est disponible que si le paramètre **Quadrature adaptative de Gauss-Lobatto** est sélectionné.

Approximation de Monte Carlo

Cette option invoque l'approche Monte Carlo.

Définir une valeur de départ personnalisée

Lorsque cette option est sélectionnée, vous pouvez définir une valeur de départ personnalisée dans la zone **Valeur de départ**.

Valeur de départ

Spécifiez une valeur de départ aléatoire pour la méthode Approximation de Monte Carlo. Il doit s'agir d'un nombre entier positif. Par défaut, une valeur de départ aléatoire est affectée.

Nombre d'échantillons de Monte Carlo

Spécifiez le nombre de points d'échantillonnage à utiliser pour l'approximation de Monte Carlo. Il doit s'agir d'un nombre entier positif. Valeur par défaut : 1000000. Cette option n'est disponible que si le paramètre **Approximation de Monte Carlo** est sélectionné.

Inférence bayésienne à échantillon unique : Probabilités a priori pour distribution normale

Vous pouvez spécifier les critères de distribution a priori suivantes pour votre inférence de Bayes à un seul échantillon :

Remarque : De nombreux chercheurs peuvent contester la nécessité d'indiquer une distribution a priori. Les distributions a priori de référence minimisent la préoccupation que cette distribution soit submergée lorsque les données augmentent. Lorsqu'une distribution a priori informative est spécifiée, les méthodes bayésiennes peuvent utiliser efficacement les informations. L'exigence d'utilisation d'une distribution a priori ne doit pas vous dissuader d'utiliser une analyse bayésienne.

Distribution a priori pour la variance/précision

Fournit des options pour définir les valeurs de variance et de précision.

Ecart

Sélectionnez cette option pour spécifier la distribution a priori pour le paramètre de variance.

Lorsque cette option est sélectionnée, la liste **Distribution a priori** propose les options suivantes :

Remarque : Lorsque la variance des données est déjà spécifiée pour certaines variables, les paramètres suivants sont ignorés pour ces variables.

- **Diffuse** - Valeur par défaut. Spécifie la distribution à priori diffuse.
- **Khi-carré inverse** - Spécifie la distribution et les paramètres pour inverse- $\chi^2(v_0, \sigma^2_0)$, où $v_0 > 0$ est le degré de liberté et $\sigma^2_0 > 0$, le paramètre d'échelle.
- **Gamma inverse** - Spécifie la distribution et les paramètres pour inverse-Gamma(α_0, β_0), où $\alpha_0 > 0$ est le paramètre de forme et $\beta_0 > 0$, le paramètre d'échelle.
- **Jeffreys S2** - Spécifie la distribution a priori non informative $\propto 1/\sigma^2_0$.
- **Jeffreys S4** - Spécifie la distribution a priori non informative $\propto 1/\sigma^4_0$.

Précision

Sélectionnez cette option pour spécifier la distribution a priori pour le paramètre de précision.

Lorsque cette option est sélectionnée, la liste **Distribution a priori** propose les options suivantes :

- **Gamma** - Spécifie la distribution et les paramètres pour Gamma (α_0, β_0), où $\alpha_0 > 0$ est le paramètre de forme et $\beta_0 > 0$, le paramètre d'échelle.
- **Khi-carré** - Spécifie la distribution et les paramètres pour $\chi^2(v_0)$, où $v_0 > 0$ est le degré de liberté.

Paramètre de forme

Spécifiez le paramètre de forme a_0 pour la distribution Inverse-Gamma. Vous devez entrer une seule valeur, supérieure à 0.

Paramètre d'échelle

Spécifiez le paramètre d'échelle b_0 pour la distribution Inverse-Gamma. Vous devez entrer une seule valeur, supérieure à 0. Plus le paramètre d'échelle est grand, plus la distribution est étendue.

Distribution a priori sur la moyenne étant donné la variance/précision

Spécifiez la distribution a priori pour le paramètre de moyenne qui est conditionnel de la variance ou de la précision.

Normal

Spécifie la distribution et les paramètres pour Normal($\mu_0, K^{-1}_0\sigma^2_0$) sur la variance ou Normal($\mu_0, K_0/\sigma^2_0$) sur la précision, où $\mu_0 \in (-\infty, \infty)$ et $\sigma^2 > 0$.

Paramètre d'emplacement

Entrez une valeur numérique qui spécifie le paramètre d'emplacement de la distribution.

Paramètre d'échelle

Spécifiez le paramètre d'échelle b_0 pour la distribution Inverse-Gamma. Vous devez entrer une seule valeur, supérieure à 0.

Kappa

Spécifiez la valeur de K_0 dans $\text{Normal}(\mu_0, K_0^{-1}\sigma_0^2)$ ou $\text{Normal}(\mu_0, K_0/\sigma_0^2)$. Vous devez entrer une seule valeur, supérieure à 0 (la valeur par défaut est 1).

Diffuse

Valeur par défaut qui spécifie la distribution a priori diffuse 1.

Inférence bayésienne à échantillon unique : Distribution binomiale

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

La procédure Inférence bayésienne à 1 échantillon : distribution binomiale propose des options pour réaliser une inférence bayésienne à 1 échantillon sur une distribution binomiale. Le paramètre étudié est π , lequel dénote la probabilité de succès dans un nombre d'essais fixe pouvant conduire à un succès ou à un échec. Notez que chaque essai est indépendant des autres et que la probabilité π demeure identique dans chaque essai. Une variable aléatoire binomiale peut être considérée comme étant la somme d'un nombre fixé d'essais de Bernoulli indépendants.

Bien que ceci ne soit pas requis, une distribution de la famille Bêta est généralement choisie pour l'estimation d'un paramètre binomial. La famille Bêta est un conjugué pour la famille binomiale et comme telle débouche sur une distribution a posteriori sous un format encore proche de la famille de distribution Bêta.

1. A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > Binomial à échantillons indépendants

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez les **variables de test** appropriées dans la liste **Variables disponibles**. Vous devez sélectionner au moins une variable.

Remarque : La liste des variables disponibles recense toutes celles disponibles, excepté celles de type Date et Chaîne.

3. Sélectionnez l'**Analyse de Bayes** voulue :

- **Caractériser la distribution a posteriori :** lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.
- **Estimer le facteur de Bayes :** lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

Tableau 3. Seuils couramment utilisés pour définir la signification des constatations

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1

Tableau 3. Seuils couramment utilisés pour définir la signification des constatations (suite)

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H0 : hypothèse nulle

H1 : hypothèse alternative

3

4

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.
4. Sélectionnez et/ou entrez les paramètres **Catégories de succès et valeurs d'hypothèse** appropriés. Le tableau reflète les variables figurant actuellement dans la liste **Variables de test**. Au fur et à mesure que des variables sont ajoutées ou retirées de la liste **Variables de test**, le tableau les ajoute ou les retire automatiquement de ses colonnes de paires de variables.
- Lorsque **Caractériser la distribution a posteriori** est sélectionné pour **Analyse de Bayes**, la colonne **Catégories de succès** est activée.
 - Lorsque **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** sont sélectionnés pour **analyse de Bayes**, toutes les colonnes modifiables sont activées.

Point Null

Active et désactive l'option **Proportion Null**. Lorsque ce paramètre est activé, les options **Forme de distribution a priori avec hypothèse nulle** et **Echelle de distribution a priori avec hypothèse nulle** sont désactivées.

Forme de distribution a priori avec hypothèse nulle

Spécifie le paramètre de forme α_0 sous l'hypothèse nulle d'inférence binomiale.

Echelle de distribution a priori avec hypothèse nulle

Spécifie le paramètre d'échelle b_0 sous l'hypothèse nulle d'inférence binomiale.

Proportion Null

Indique le paramètre de forme α_0 et le paramètre d'échelle b_0 sous l'hypothèse nulle pour une distribution a priori conjuguée (pour une adaptation aux distributions a priori Bêta et Haldane). La plage valide comprend les valeurs numériques entre 0 et 1.

Forme de distribution a priori alternative

Paramètre obligatoire pour spécifier α_0 sous l'hypothèse alternative d'inférence binomiale si le facteur de Bayes doit être estimé.

Echelle de distribution a priori alternative

Paramètre obligatoire pour spécifier b_0 sous l'hypothèse alternative d'inférence binomiale si le facteur de Bayes doit être estimé.

Catégories de succès

Fournit des options pour définir des distributions à priori conjuguées. Ces options spécifient comment le succès est défini pour les variables numériques et chaînes quand les valeurs de données sont testées vis à vis de la valeur de test.

³ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

⁴ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Dernière catégorie

Paramètre par défaut qui effectue le test binomial en utilisant la dernière valeur numérique présente dans la catégorie après son tri en ordre ascendant.

Première catégorie

Effectue le test binomial en utilisant la première valeur numérique présente dans la catégorie après son tri en ordre ascendant.

Centre

Utilise les valeurs numériques \geq au centre comme observations. Une valeur centre est la moyenne des données d'échantillon minimum et maximum.

Césure

Utilise les valeurs numériques \geq la valeur de césure spécifiée comme observations. Ce paramètre doit être une seule valeur numérique.

Niveau

Traite les valeurs de chaîne spécifiées par l'utilisateur (éventuellement plus d'une) comme des observations. Utilisez des virgules pour séparer les différentes valeurs.

5. Vous pouvez éventuellement cliquer sur **Critères** pour spécifier des paramètres pour «Inférence bayésienne à échantillon unique : Critères», à la page 113 (pourcentage d'intervalle crédible, options pour valeurs manquantes et paramètres de méthode numérique) ou sur **Distributions à priori** pour spécifier des paramètres pour «Inférence bayésienne à échantillon unique : Probabilités a priori binomiales/Poisson», à la page 117 (distributions à priori conjuguées ou personnalisées).

Inférence bayésienne à échantillon unique : Probabilités a priori binomiales/Poisson

Vous pouvez spécifier les critères de distribution a priori suivantes pour votre inférence de Bayes à un seul échantillon :

Remarque : De nombreux chercheurs peuvent contester la nécessité d'indiquer une distribution a priori. Les distributions a priori de référence minimisent la préoccupation que cette distribution soit submergée lorsque les données augmentent. Lorsqu'une distribution a priori informative est spécifiée, les méthodes bayésiennes peuvent utiliser efficacement les informations. L'exigence d'utilisation d'une distribution a priori ne doit pas vous dissuader d'utiliser une analyse bayésienne.

Paramètre de forme

Pour les distributions a priori binomiales, spécifiez comme paramètre de forme a_0 pour la distribution bêta.

Pour les distributions de Poisson, spécifiez comme paramètre de forme a_0 pour la distribution gamma.

Vous devez entrer une seule valeur, supérieure à 0.

Paramètre d'échelle

Pour les distributions a priori binomiales, spécifiez comme paramètre d'échelle b_0 pour la distribution bêta.

Pour les distributions de Poisson, spécifiez comme paramètre d'échelle b_0 pour la distribution gamma.

Vous devez entrer une seule valeur, supérieure à 0.

Inférence bayésienne à échantillon unique : Poisson

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

La procédure Inférence bayésienne à 1 échantillon : distribution de Poisson propose des options pour réaliser une inférence bayésienne à 1 échantillon sur une distribution de Poisson. La distribution de Poisson, qui constitue un modèle utile pour des événements rares, suppose qu'à de courts intervalles de temps, la probabilité qu'un événement se produise est proportionnelle à la longueur de l'attente. Une probabilité a priori conjuguée de la famille de distribution Gamma est utilisée lors du calcul de l'inférence statistique de Bayes sur la distribution de Poisson.

1. A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > Poisson à un échantillon

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez les **variables de test** appropriées dans la liste **Variables disponibles**. Vous devez sélectionner au moins une variable.

Remarque : La liste des variables disponibles recense toutes celles disponibles, excepté celles de type Date et Chaîne.

3. Sélectionnez l'**Analyse de Bayes** voulue :

- **Caractériser la distribution a posteriori** : lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.
- **Estimer le facteur de Bayes** : lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

Tableau 4. Seuils couramment utilisés pour définir la signification des constatations

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H0 : hypothèse nulle

H1 : hypothèse alternative

5

6

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.

4. Sélectionnez et/ou entrez les paramètres **Valeurs d'hypothèse** appropriés. Le tableau reflète les variables figurant actuellement dans la liste **Variables de test**. Au fur et à mesure que des variables sont ajoutées ou retirées de la liste **Variables de test**, le tableau les ajoute ou les retire automatiquement de ses colonnes de paires de variables.

⁵ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

⁶ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

- Lorsque **Caractériser la distribution a posteriori** est sélectionnée pour **Analyse de Bayes**, aucune colonne n'est activée.
- Lorsque **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** sont sélectionnés pour **analyse de Bayes**, toutes les colonnes modifiables sont activées.

Point Null

Active et désactive l'option **Taux Null**. Lorsque ce paramètre est activé, les options **Forme de distribution a priori avec hypothèse nulle** et **Echelle de distribution a priori avec hypothèse nulle** sont désactivées.

Forme de distribution a priori avec hypothèse nulle

Spécifie le paramètre de forme a_0 sous l'hypothèse nulle d'inférence de Poisson.

Echelle de distribution a priori avec hypothèse nulle

Spécifie le paramètre d'échelle b_0 sous l'hypothèse nulle d'inférence de Poisson.

Taux Null

Indique le paramètre de forme a_0 et le paramètre d'échelle b_0 sous l'hypothèse nulle pour une distribution a priori conjuguée (pour une adaptation à la relation Poisson-Gamma). La valeur minimale doit être une valeur numérique supérieure à 0 ; la valeur maximale doit être une valeur de type double maximale.

Forme de distribution a priori alternative

Paramètre obligatoire pour spécifier a_1 sous l'hypothèse alternative d'inférence de Poisson si le facteur de Bayes doit être estimé.

Echelle de distribution a priori alternative

Paramètre obligatoire pour spécifier b_1 sous l'hypothèse d'inférence de Poisson si le facteur de Bayes doit être estimé.

5. Vous pouvez éventuellement cliquer sur **Critères** pour spécifier des paramètres pour «Inférence bayésienne à échantillon unique : Critères», à la page 113 (pourcentage d'intervalle crédible, options pour valeurs manquantes et paramètres de méthode numérique) ou sur **Distributions à priori** pour spécifier des paramètres pour «Inférence bayésienne à échantillon unique : Probabilités a priori binomiales/Poisson», à la page 117 (distributions à priori conjuguées ou personnalisées).

Inférence bayésienne à échantillons liés : Distribution normale

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

La procédure Inférence bayésienne à échantillon lié : Distribution normale fournit des options d'inférence bayésienne pour des échantillons appariés. Vous pouvez spécifier les noms de variable par paires et exécuter l'analyse de Bayes sur la différence moyenne.

1. A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > Échantillons liés Distribution normale

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez les **Variables appariées** appropriées dans la liste **Variables disponibles**. Au moins une paire de variables source doit être sélectionnée et pas plus de deux variables source peuvent être sélectionnées pour chaque paire mentionnée.

Remarque : La liste des variables disponibles recense toutes celles disponibles, excepté celles de type Chaîne.

3. Sélectionnez l'**Analyse de Bayes** voulue :

- **Caractériser la distribution a posteriori** : lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.

- **Estimer le facteur de Bayes** : lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

Tableau 5. Seuils couramment utilisés pour définir la signification des constatations

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H0 : hypothèse nulle

H1 : hypothèse alternative

7

8

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.
4. Sélectionnez et/ou entrez les paramètres **Valeurs de variance de données et d'hypothèse** appropriés. Le tableau reflète les paires de variables figurant actuellement dans la liste **Variables appariées**. Au fur et à mesure que des paires de variables sont ajoutées ou retirées de la liste **Variables appariées**, le tableau les ajoute ou les retire automatiquement des colonnes de paires de variables.
- Lorsqu'une ou plusieurs variables sont présentes dans la liste **Variables appariées**, les colonnes **Variance connue** et **Valeur de la variance** sont activées.
 - Variance connue**
Sélectionnez cette option pour chaque variable dont la variance est connue.
 - Valeur de la variance**
Paramètre facultatif qui spécifie la valeur de la variance (si elle est connue) des données des observations.
 - Lorsqu'une ou plusieurs paires de variables sont présentes dans la liste **Variables appariées** et que l'option **Caractériser la distribution a posteriori** n'est pas sélectionnée, les colonnes **Valeur de test nulle** et **Valeur g** sont activées.
 - Valeur de test nulle.**
Paramètre obligatoire spécifiant la valeur Null dans l'estimation du facteur de Bayes. Une seule valeur est autorisée et 0 est la valeur par défaut.

⁷ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

⁸ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Valeur g

Spécifie la valeur pour définir $\psi^2 = g\sigma^2_x$ dans l'estimation du facteur de Bayes. Lorsque la **Valeur de variance** est spécifiée, la valeur par défaut de la **Valeur g** est 1. Lorsque la **Valeur de variance** n'est pas spécifiée, vous pouvez spécifier une valeur g fixe ou omettre la valeur pour ne pas l'intégrer.

- Vous pouvez éventuellement cliquer sur **Critères** pour spécifier des paramètres pour «Inférence bayésienne à échantillon unique : Critères», à la page 113 (pourcentage d'intervalle crédible, options pour valeurs manquantes et paramètres de méthode numérique) ou sur **Distributions à priori** pour spécifier des paramètres pour «Inférence bayésienne à échantillon unique : Probabilités a priori binomiales/Poisson», à la page 117 (distributions à priori conjuguées ou personnalisées).

Inférence bayésienne à échantillon indépendant

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

La procédure Inférence bayésienne à échantillon indépendant fournit des options pour utiliser une variable de groupe afin de définir deux groupes sans lien et de générer une inférence sur la différence des moyennes des deux groupes. Vous pouvez estimer les facteurs de Bayes en utilisant différentes approches, et aussi caractériser la distribution a posteriori désirée en supposant que les variances sont connues ou inconnues.

- A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > Distribution normale à échantillons indépendants

- Sélectionnez les **Variables de test** appropriées dans la liste des variables source. Vous devez sélectionner au moins une variable source.

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

- Sélectionnez la **Variable de regroupement** appropriées dans la liste **Variables disponibles**. Une variable de regroupement définit deux groupes pour le test t non apparié. Il peut s'agir d'une variable numérique ou chaîne.
- Sélectionnez l'**Analyse de Bayes** voulue :
 - **Caractériser la distribution a posteriori** : lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.
 - **Estimer le facteur de Bayes** : lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H0 : hypothèse nulle

H1 : hypothèse alternative

9

10

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.
5. Utilisez les options de **Définir groupes** pour définir deux groupes pour le test *t* en spécifiant deux valeurs (pour les variables chaîne), ou deux valeurs, un centre, ou une césure (pour les variables numériques).

Remarque : Les valeurs spécifiées doivent exister dans la variable, faute de quoi un message d'erreur s'affiche en vous avisant qu'au moins l'un des groupes est vide.

Pour les variables numériques :

- **Utiliser les valeurs spécifiées** : Entrez une valeur pour le groupe 1 et une autre valeur pour le groupe 2. Les observations avec d'autres valeurs sont exclues de l'analyse. Il n'est pas nécessaire que les nombres soient des entiers (par exemple, 6,25 et 12,5 sont valides).
- **Utiliser la valeur du centre**. Lorsque cette option est sélectionnée, les groupes sont séparés en valeurs $<$ et \geq au centre.
- **Utiliser la césure**.
 - **Césure** : Vous avez également la possibilité de saisir un nombre qui sépare les valeurs de la variable de regroupement en deux groupes. Toutes les observations ayant des valeurs inférieures à la césure constituent un groupe et les observations ayant des valeurs supérieures ou égales à la césure constituent l'autre groupe.

Pour les variables de regroupement de chaînes, entrez une chaîne pour le groupe 1 et une autre pour le groupe 2, par exemple *Oui* et *Non*. Les observations avec d'autres chaînes sont exclues de l'analyse.

6. Vous pouvez éventuellement cliquer sur **Critères** pour spécifier des paramètres «Inférence bayésienne à échantillon indépendant : critères», à la page 123 (pourcentage d'intervalle crédible, options pour valeurs manquantes et paramètres de méthode de quadrature adaptative), sur **Distributions a priori** pour spécifier des paramètres «Inférence bayésienne à échantillon indépendant : distribution a priori», à la page 123 (variance des données, distribution a priori sur la variance et distribution a priori sur la moyenne conditionnelle à la variance), ou sur **Estimer le facteur de Bayes** pour spécifier des paramètres «Analyse bayésienne indépendante - Inférence d'échantillon : Estimation du facteur de Bayes», à la page 125.

Inférence bayésienne à échantillon indépendant - Définition de groupes (numérique)

Pour les variables de groupement numérique, définissez les deux groupes pour le test *t* en spécifiant deux valeurs, un centre ou un point de césure.

⁹ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹⁰ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Remarque : Les valeurs spécifiées doivent exister dans la variable, faute de quoi un message d'erreur s'affiche en vous avisant qu'au moins l'un des groupes est vide.

- **Utiliser les valeurs spécifiées :** Entrez une valeur pour le groupe 1 et une autre valeur pour le groupe 2. Les observations avec d'autres valeurs sont exclues de l'analyse. Il n'est pas nécessaire que les nombres soient des entiers (par exemple, 6,25 et 12,5 sont valides).
- **Utiliser la valeur du centre.** Lorsque cette option est sélectionnée, les groupes sont séparés en valeurs $<$ et \geq au centre.
- **Utiliser la césure.**
 - **Césure :** Vous avez également la possibilité de saisir un nombre qui sépare les valeurs de la variable de regroupement en deux groupes. Toutes les observations ayant des valeurs inférieures à la césure constituent un groupe et les observations ayant des valeurs supérieures ou égales à la césure constituent l'autre groupe.

Inférence bayésienne à échantillon indépendant - Définition de groupes (chaîne)

Pour les variables de regroupement de chaînes, entrez une chaîne pour le groupe 1 et une autre pour le groupe 2, par exemple *Oui* et *Non*. Les observations avec d'autres chaînes sont exclues de l'analyse.

Remarque : Les valeurs spécifiées doivent exister dans la variable, faute de quoi un message d'erreur s'affiche en vous avisant qu'au moins l'un des groupes est vide.

Inférence bayésienne à échantillon indépendant : critères

Vous pouvez définir les critères d'analyse suivants pour votre inférence bayésienne à échantillon indépendant :

Pourcentage d'intervalle crédible

Spécifiez le niveau de signification pour les intervalles de calcul crédibles. Niveau par défaut : 95 %.

Valeurs manquantes

Spécifiez la méthode de contrôle des valeurs manquantes.

Exclure seulement les composantes non valides

Il s'agit du paramètre par défaut, lequel exclut les enregistrements avec des valeurs manquantes sur une base analyse par analyse. Les enregistrements qui incluent des valeurs manquantes pour une zone utilisée pour un test spécifique sont omis du test.

Exclure toute observation incomplète

Ce paramètre exclut les enregistrements incluant des valeurs manquantes sur la base d'une liste. Les enregistrements qui incluent des valeurs manquantes pour une zone quelconque nommée dans une sous-commande sont exclus de toutes les analyses.

Remarque : Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour **Analyse de Bayes**.

Méthode de quadrature adaptative

Spécifiez les valeurs de tolérance et de nombre maximal d'itérations pour la méthode de quadrature adaptative.

Tolérance

Spécifiez la valeur de tolérance pour les méthodes numériques. Valeur par défaut : 0.000001.

Nombre maximum d'itérations

Spécifiez le nombre maximal d'itérations de la méthode Quadrature adaptative. Il doit s'agir d'un nombre entier positif. Valeur par défaut : 500.

Inférence bayésienne à échantillon indépendant : distribution a priori

Vous pouvez spécifier les critères de distribution a priori suivants pour votre Inférence bayésienne à échantillon indépendant :

Remarque : De nombreux chercheurs peuvent contester la nécessité d'indiquer une distribution a priori. Les distributions a priori de référence minimisent la préoccupation que cette distribution soit submergée lorsque les données augmentent. Lorsqu'une distribution a priori informative est spécifiée, les méthodes bayésiennes peuvent utiliser efficacement les informations. L'exigence d'utilisation d'une distribution a priori ne doit pas vous dissuader d'utiliser une analyse bayésienne.

Variance des données

Fournit des options pour définir les paramètres de variance des données.

Variance connue

Lorsque cette option est sélectionnée, vous pouvez entrer deux variances de groupe connues. Les deux valeurs doivent être supérieures à 0.

Variance du groupe 1

Entrez la première variance de groupe connue.

Variance du groupe 2

Entrez la seconde variance de groupe connue.

Supposer une variance égale

Détermine si les variances des deux groupes sont supposées être égales. Par défaut, il est supposé que les variances de groupe sont inégales. Ce paramètre est ignoré lorsque des valeurs sont saisies pour les deux variances de groupe.

Supposer une variance inégale

Détermine si les variances des deux groupes sont supposées être inégales. Par défaut, il est supposé que les variances de groupe sont inégales. Ce paramètre est ignoré lorsque des valeurs sont saisies pour les deux variances de groupe.

Distribution a priori pour la variance

Spécifiez la distribution a priori pour les deux variances égales.

Jeffreys

Lorsque cette option est sélectionnée, une distribution a priori non informative (objective) pour un espace de paramètre est utilisée.

Inverse-ChiSquare

Spécifie la distribution de probabilité continue d'une variable aléatoire à valeur positive et les paramètres pour inverse- $\chi^2(v_0, \sigma^2_0)$, où $v_0 > 0$ est le degré de liberté et $\sigma^2_0 > 0$ le paramètre d'échelle.

Degrés de liberté

Spécifiez une valeur pour le nombre de valeurs dans le calcul final libres de varier.

Paramètre d'échelle

Spécifiez le paramètre d'échelle $\sigma^2_0 > 0$ pour inverse- $\chi^2(v_0, \sigma^2_0)$. Vous devez entrer une seule valeur, supérieure à 0. Plus le paramètre d'échelle est grand, plus la distribution est étendue.

Probabilités a priori sur la moyenne conditionnelle à la variance

Fournit des options pour spécifier la distribution a priori de deux moyennes de groupe.

Remarque : Les options **Diffuse** et **Normale** ne sont disponibles que lorsque l'option **Variance connue** est sélectionnée.

Diffuse

Valeur par défaut. Spécifie la distribution à priori diffuse.

Normal

Lorsque cette option est sélectionnée, vous devez spécifier les paramètres d'emplacement et d'échelle pour les moyennes des groupes définis.

Paramètre d'emplacement

Entrez une valeur numérique spécifiant les paramètres d'emplacement des paramètres pour les distributions de groupe.

Paramètre d'échelle

Spécifiez le paramètre d'échelle $\sigma^2_0 > 0$ pour inverse- $\chi^2(v_0, \sigma^2_0)$. Pour chaque groupe, vous devez entrer une valeur unique supérieure à 0. Plus le paramètre d'échelle est grand, plus la distribution est étendue.

Analyse bayésienne indépendante - Inférence d'échantillon : Estimation du facteur de Bayes

Vous pouvez spécifier la méthode à utiliser pour estimer le facteur de Bayes.

Méthode de Rouder

Lorsque cette option est sélectionnée, elle appelle l'approche de Rouder. Il s'agit du paramètre par défaut.

Méthode de Gonen

Lorsque cette option est sélectionnée, elle appelle l'approche de Gonen et vous devez spécifier les paramètres de taille d'effet suivants :

Moyenne pour taille d'effet

Entrez une valeur spécifiant la différence de la moyenne entre les deux groupes.

Variance pour effet de taille

Entrez une valeur spécifiant la variance pour les deux groupes. Cette valeur doit être supérieure à 0.

Méthode hyper à priori

Lorsque cette option est sélectionnée, elle appelle l'approche hyper-g dans laquelle vous devez spécifier une seule valeur. Entrez une valeur comprise entre -1 et -0.5 dans le champ **Paramètre de forme**. Valeur par défaut : -0.75.

Inférence bayésienne sur corrélation de Pearson

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

Le coefficient de corrélation de Pearson mesure la relation linéaire entre deux variables d'échelle suivant ensemble une distribution normale bivariée. L'inférence statistique classique sur le coefficient de corrélation a été largement discutée et son utilisation est pratiquée de longue date dans IBM SPSS Statistiques. La conception de l'inférence bayésienne sur le coefficient de corrélation de Pearson permet aux utilisateurs de générer une inférence bayésienne en estimant les facteurs de Bayes et en caractérisant des distributions a posteriori.

1. A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > Corrélation de Pearson

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez les **Variables de test** appropriées à utiliser pour l'inférence de corrélation par paires dans la liste **Variables disponibles**. Au moins deux variables source doivent être spécifiées. Lorsque plus de deux variables sont sélectionnées, l'analyse est exécutée sur toutes les combinaisons des paires de variables sélectionnées.

3. Sélectionnez l'**Analyse de Bayes** voulue :

- **Caractériser la distribution a posteriori** : lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.
- **Estimer le facteur de Bayes** : lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

Tableau 7. Seuils couramment utilisés pour définir la signification des constatations

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H0 : hypothèse nulle

H1 : hypothèse alternative

11

12

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.
- Spécifiez le **Nombre maximal de tracés** à afficher dans la sortie. L'ensemble de tracés peut comporter 3 tracés dans le même panneau. Les tracés sont générés dans l'ordre, de la première variable par rapport aux variables restantes, puis de la deuxième variable aux variables restantes, et ainsi de suite. La valeur d'entier stipulée doit être comprise entre 0 et 50. Par défaut, 10 jeux de tracés sont générés pour rendre compte de cinq variables. Cette option n'est pas disponible quand **Estimer le facteur de Bayes** est sélectionné.
 - Vous pouvez éventuellement cliquer sur **Critères** pour spécifier les paramètres de «Corrélation bayésienne Pearson : critères», à la page 126 (pourcentage d'intervalle crédible, options de valeurs manquantes et paramètres de la méthode numérique), cliquer sur **Valeurs a priori** pour spécifier les paramètres de «Corrélation bayésienne de Pearson : distribution a priori», à la page 127 (valeur C pour la valeur a priori $p(\rho) \propto (1 - \rho^2)^C$), ou cliquer sur **Facteur de Bayes** pour spécifier les paramètres de «Analyse bayésienne indépendante - Inférence d'échantillon : Estimation du facteur de Bayes», à la page 125.

Corrélation bayésienne Pearson : critères

Vous pouvez spécifier les critères d'analyse suivants pour la corrélation bayésienne de Pearson (avec appariement).

Pourcentage d'intervalle crédible

Spécifiez le niveau de signification pour les intervalles de calcul crédibles. Niveau par défaut : 95 %.

Valeurs manquantes

Spécifiez la méthode de contrôle des valeurs manquantes.

Exclure seulement les composantes non valides

Ce paramètre exclut les enregistrements avec des valeurs appariées manquantes.

¹¹ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹² Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Exclure toute observation incomplète

Ce paramètre exclut les enregistrements incluant des valeurs manquantes sur la base d'une liste. Les enregistrements qui incluent des valeurs manquantes pour une zone quelconque nommée dans une sous-commande sont exclus de toutes les analyses.

Remarque : Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour **Analyse de Bayes**.

Méthode numérique

Spécifiez la méthode numérique à utiliser pour estimation de l'intégrale.

Définir une valeur de départ personnalisée

Lorsque cette option est sélectionnée, vous pouvez définir une valeur de départ personnalisée dans la zone **Valeur de départ**.

Tolérance

Spécifiez la valeur de tolérance pour les méthodes numériques. Valeur par défaut : 0.000001.

Nombre maximum d'itérations

Spécifiez le nombre maximal d'itérations de la méthode. Il doit s'agir d'un nombre entier positif. Valeur par défaut : 2000.

Nombre d'échantillons de Monte Carlo

Spécifiez le nombre de points d'échantillonnage à utiliser pour l'approximation de Monte Carlo. Il doit s'agir d'un nombre entier positif. La valeur par défaut est 10 000.

Echantillons simulés à partir de distribution a posteriori

Spécifiez le nombre d'échantillons utilisé pour composer la distribution a posteriori désirée. La valeur par défaut est 10 000.

Corrélation bayésienne de Pearson : distribution a priori

Vous pouvez spécifier la valeur c pour la distribution a priori $p(\rho) \propto (1-\rho^2)^c$.

Remarque : De nombreux chercheurs peuvent contester la nécessité d'indiquer une distribution a priori. Les distributions a priori de référence minimisent la préoccupation que cette distribution soit submergée lorsque les données augmentent. Lorsqu'une distribution a priori informative est spécifiée, les méthodes bayésiennes peuvent utiliser efficacement les informations. L'exigence d'utilisation d'une distribution a priori ne doit pas vous dissuader d'utiliser une analyse bayésienne.

Uniforme (c = 0)

Lorsque cette option est sélectionnée, la distribution a priori uniforme est utilisée.

Jeffreys (c = -1.5)

Lorsque cette option est sélectionnée, une distribution a priori non uniforme est utilisée.

Définir valeur c personnalisée

Lorsque cette option est sélectionnée, vous pouvez spécifier une **valeur c** personnalisée. N'importe quel nombre réel unique est autorisé.

Corrélation bayésienne Pearson : facteur de Bayes

Vous pouvez spécifier la méthode à utiliser pour estimer le facteur de Bayes. Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour l'analyse de Bayes.

Facteur JZS Bayes

Lorsque cette option est sélectionnée, elle appelle l'approche de Zellner-Siow. Il s'agit du paramètre par défaut.

Facteur de Bayes fractionnel

Lorsque cette option est sélectionnée, vous pouvez spécifier le facteur Bayes fractionnel et la valeur d'hypothèse nulle. Pour le facteur Bayes fractionnel, vous devez spécifier une valeur ϵ (0,1). La valeur par défaut est 0,5.

Inférence bayésienne sur modèles de régression linéaire

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

La régression est une méthode statistique couramment utilisée en modélisation quantitative. La régression linéaire est une approche de base standard où les chercheurs utilisent les valeurs de plusieurs variables pour expliquer ou prédire les valeurs d'une échelle de résultats. La régression linéaire univariée de Bayes est une approche de la régression linéaire où l'analyse statistique est réalisée dans le contexte de l'inférence bayésienne.

Vous pouvez appeler la procédure de régression et définir un modèle complet.

1. A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > Régression linéaire

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez une seule variable dépendante, non chaîne, dans la liste **Variables disponibles**. Vous devez sélectionner une seule variable, non chaîne.

3. Sélectionnez une ou plusieurs variables de facteur catégoriel pour le modèle dans la liste **Variables disponibles**.

4. Sélectionnez une ou plusieurs variables d'échelle de covariable, non chaîne, dans la liste **Variables disponibles**.

Remarque : La liste **Facteur(s)** et la liste **Covariable(s)** ne doivent pas être vides. Vous devez sélectionner au moins une variable **Facteur(s)** ou **Covariable(s)**.

5. Vous pouvez éventuellement sélectionner une seule variable, non chaîne, pour faire office de pondération de régression dans la liste **Variables disponibles**.

6. Sélectionnez l'**Analyse de Bayes** voulue :

- **Caractériser la distribution a posteriori** : lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.
- **Estimer le facteur de Bayes** : lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

<i>Tableau 8. Seuils couramment utilisés pour définir la signification des constatations</i>					
Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H_0 : hypothèse nulle

H_1 : hypothèse alternative

13

14

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.

(En option) Vous avez également ces possibilités :

- Cliquez sur **Critères** pour spécifier le pourcentage d'intervalle crédible et les paramètres de méthode numérique.
- Cliquez sur **Distributions à priori** pour définir les paramètres de référence et de distribution à priori conjuguée.
- Cliquez sur **Facteur de Bayes** pour spécifier des paramètres de facteur de Bayes.
- Cliquez sur **Enregistrer** pour identifier les éléments à sauvegarder et enregistrer les informations du modèle dans un fichier XML.
- Cliquez sur **Prédire** pour spécifier des régresseurs pour la prévision de Bayes.
- Cliquez sur **Tracés** pour tracer les distributions a posteriori des paramètres de régression, la variance des termes d'erreur et les valeurs prédites.
- Cliquez sur **Tests F** pour comparer des modèles statistiques afin d'identifier celui correspondant le mieux à la population dont est issu l'échantillon.

Modèles de régression linéaire de Bayes : critères

Vous pouvez spécifier les critères d'analyse suivants pour les modèles de régression linéaire de Bayes.

Pourcentage d'intervalle crédible

Spécifiez le niveau de signification pour les intervalles de calcul crédibles. Niveau par défaut : 95 %.

Remarque : Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour **Analyse de Bayes**.

Méthode numérique

Spécifiez la méthode numérique à utiliser pour estimation de l'intégrale.

Tolérance

Spécifiez la valeur de tolérance pour les méthodes numériques. Valeur par défaut : 0.000001.

Nombre maximum d'itérations

Spécifiez le nombre maximal d'itérations de la méthode. Il doit s'agir d'un nombre entier positif. Valeur par défaut : 2000.

Modèles de régression linéaire bayésienne : Distributions a priori

Vous pouvez spécifier les paramètres de distribution a priori suivants pour les paramètres de régression et la variance des erreurs. Les options suivantes ne sont disponibles que si l'option **Caractériser la distribution a posteriori** est sélectionnée pour **Analyse de Bayes**.

Remarque : De nombreux chercheurs peuvent contester la nécessité d'indiquer une distribution a priori. Les distributions a priori de référence minimisent la préoccupation que cette distribution soit submergée lorsque les données augmentent. Lorsqu'une distribution a priori informative est spécifiée, les méthodes bayésiennes peuvent utiliser efficacement les informations. L'exigence d'utilisation d'une distribution a priori ne doit pas vous dissuader d'utiliser une analyse bayésienne.

¹³ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹⁴ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Référence

Lorsque cette option est sélectionnée, l'analyse des références génère une inférence bayésienne objective. Les instructions inférentielles ne dépendent que du modèle supposé et des données disponibles, et la distribution a priori utilisée pour générer une inférence est la moins informative. Il s'agit du paramètre par défaut.

Conjuguer

Fournit des options pour définir des distributions à priori conjuguées. Les distributions à priori conjuguées supposent une distribution jointe Normale-Inverse-Gamma. Bien que des distributions à priori conjuguées ne soient pas obligatoires pour effectuer des mises à jour de Bayes, elles facilitent les processus de calcul.

Remarque : Pour spécifier des distributions a priori conjuguées pour un modèle de régression linéaire, définissez votre moyenne attendue des paramètres de régression dans la table **Distributions a priori sur variance des erreurs**. Vous pouvez également utiliser les paramètres de **Matrice de variance de covariance** pour spécifier la variance-covariance de la distribution a priori.

Distributions à priori sur variance des erreurs

Paramètre de forme

Spécifiez le paramètre de forme α_0 pour la distribution Inverse-Gamma. Vous devez entrer une seule valeur, supérieure à 0.

Paramètre d'échelle

Spécifiez le paramètre d'échelle b_0 pour la distribution Inverse-Gamma. Vous devez entrer une seule valeur, supérieure à 0. Plus le paramètre d'échelle est grand, plus la distribution est étendue.

Le tableau indique la moyenne des paramètres de régression (y-compris la constante), laquelle spécifie le vecteur de moyenne ϑ_0 pour les paramètres de régression spécifiés. Le nombre de valeurs doit correspondre à celui des paramètres de régression, y-compris le terme de la constante.

Le premier nom de variable est toujours INTERCEPT. A partir de la seconde ligne, la colonne **Variables** est automatiquement alimentée avec les variables spécifiées par Facteur(s) et Covariable(s). La colonne **Moyenne** n'inclut pas de valeurs par défaut.

Cliquez sur **Réinitialiser** pour effacer les valeurs.

Matrice de variance de la covariance : $\sigma^2\mathbf{x}$

Spécifiez les valeurs V_0 dans le triangle inférieur de la matrice de variance-covariance pour la distribution à priori normale multivariée. Notez que V_0 doit être définie comme semi-positive. La dernière valeur de chaque ligne doit être positive. La ligne suivante doit comporter une valeur de plus que la précédente. Aucune valeur n'est spécifiée pour les catégories de référence (s'il en existe).

Cliquez sur **Réinitialiser** pour effacer les valeurs.

Utiliser la matrice d'identité

Lorsque cette option est sélectionnée, la matrice d'identité mise à l'échelle est utilisée. Vous ne pouvez pas spécifier de valeurs V_0 dans le triangle inférieur de la matrice de variance-covariance pour la distribution a priori normale multivariée.

Modèles de régression linéaire de Bayes : facteur de Bayes

Vous pouvez spécifier la conception de modèle pour l'analyse, y-compris l'approche utilisée pour estimer le facteur de Bayes pour les modèles de régression linéaire de Bayes. Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour l'analyse de Bayes.

Modèle Nul

Lorsque cette option est sélectionnée, les facteurs de Bayes estimés sont basés sur le modèle nul. Il s'agit du paramètre par défaut.

Modèle complet

Lorsque cette option est sélectionnée, les facteurs de Bayes estimés sont basés sur le modèle complet et vous pouvez sélectionner des variables à utiliser et d'autres facteurs et covariables.

Variables

Répertorie toutes les variables disponibles pour le modèle complet.

Facteur(s) supplémentaire(s)

Sélectionnez dans la liste **Variables** des variables à utiliser comme facteurs supplémentaires.

Covariable(s) supplémentaire(s)

Sélectionnez dans la liste **Variables** des variables à utiliser comme covariables supplémentaires.

Calcul

Spécifiez l'approche à utiliser pour estimer les facteurs de Bayes. Par défaut, il s'agit de la méthode JZS.

Méthode JZS

Lorsque cette option est sélectionnée, elle appelle l'approche de Zellner-Siow. Il s'agit du paramètre par défaut.

Méthode Zellner

Lorsque cette option est sélectionnée, elle appelle l'approche de Zellner et vous devez spécifier une valeur a priori unique $g > 0$ (il n'existe pas de valeur par défaut).

Méthode hyper à priori

Lorsque cette option est sélectionnée, elle appelle l'approche hyper- g et vous devez spécifier un paramètre de forme a_0 pour la distribution Inverse-Gamma. Vous devez spécifier une seule valeur > 0 (valeur par défaut : 3).

Méthode de Rouder

Lorsque cette option est sélectionnée, elle appelle l'approche de Rouder et vous devez spécifier un paramètre d'échelle b_0 pour la distribution Inverse-Gamma. Vous devez spécifier une seule valeur > 0 (valeur par défaut : 1).

Modèles de régression linéaire de Bayes : enregistrement

Cette boîte de dialogue vous permet de spécifier quelles statistiques évaluer pour la distribution de prévision de Bayes et pour exporter les résultats du modèle dans un fichier XML.

Statistiques prédictives à postériori

Vous pouvez évaluer les statistiques suivantes dérivées des prévisions de Bayes.

Moyennes

Moyenne de la distribution prédictive a posteriori.

Variances

Variance de la distribution prédictive a posteriori.

Modes

Mode de la distribution prédictive a posteriori.

Limite inférieure de l'intervalle crédible

Valeur plancher de l'intervalle crédible de la distribution prédictive a posteriori.

Limite supérieure d'intervalle crédible

Valeur plafond de l'intervalle crédible de la distribution prédictive a posteriori.

Remarque : Vous pouvez affecter des noms de variable correspondants à chaque statistique.

Exporter les informations du modèle dans un fichier XML

Entrez un nom de fichier XML et un emplacement où exporter la matrice de variance-covariance des paramètres évalués.

Modèles de régression linéaire de Bayes : prévision

Vous pouvez spécifier les régresseurs pour générer des distributions prédictives.

Régresseurs pour prévision de Bayes

Le tableau recense tous les régresseurs disponibles. La colonne **Régresseurs** est alimentée automatiquement avec certaines variables de facteur et de covariable. Spécifiez les vecteurs observés avec les valeurs pour les régresseurs. Une seule valeur ou chaîne peut être affectée à un régresseur et celui-ci ne peut prédire qu'une observation. Pour les facteurs, les valeurs et les chaînes sont admises.

Toutes les valeurs de régresseurs (ou bien aucune) doivent être spécifiées pour lancer la prévision (en cliquant sur **Exécuter l'analyse**).

Lorsqu'une variable de facteur ou de covariable est retirée, la ligne de régresseur correspondante est retirée du tableau.

Pour les covariables, seules des valeurs numériques peuvent être spécifiées. Pour les facteurs, les valeurs numériques et les chaînes sont toutes deux autorisées.

Remarque : Cliquez sur **Réinitialiser** pour effacer les valeurs définies.

Modèles de régression linéaire de Bayes : tracés

Vous pouvez contrôler les tracés générés en sortie.

Covariables

Répertorie les covariables définies actuellement.

Tracer les covariables

Sélectionnez dans la liste **Covariables** celles à tracer et ajoutez-les à liste **Tracer les covariables**.

Facteurs

Répertorie les facteurs définis actuellement.

Tracé de facteur(s)

Sélectionnez dans la liste **Facteurs** ceux pour lesquels générer un tracé et ajoutez-les à la liste **Tracé de facteurs**.

Nombre maximal de catégories à tracer

Sélectionnez le nombre maximal de catégories pour lesquelles générer un tracé (entier unique positif). Ce paramètre s'applique à tous les facteurs. Par défaut, les 2 premiers niveaux sont tracés pour chaque facteur.

Inclure les tracés de

Terme Constante

Lorsque cette option est sélectionnée, le tracé du terme de constante est généré. Ce paramètre est désélectionné par défaut.

Variance des termes d'erreur

Si ce paramètre est sélectionné, la variance des erreurs figure sur le tracé. Ce paramètre est désélectionné par défaut.

Distribution Bayes prévue

Lorsque cette option est sélectionnée, le tracé de distribution prédictive est généré. Ce paramètre est désélectionné par défaut. Ce paramètre ne peut être sélectionné que si des valeurs de régresseur valides ont été sélectionnées.

Modèles de régression linéaire de Bayes : tests F

Vous pouvez créer un ou plusieurs tests F partiels. Un test F est un test statistique quelconque dans lequel la statistique de test suit une distribution F sous l'hypothèse nulle. Les tests F sont couramment utilisés lors de la comparaison de modèles statistiques ajustés à un ensemble de données afin d'identifier celui correspondant le mieux à la population dont est issu l'échantillon.

Variables disponibles

Répertorie les variables de facteurs et de covariables sélectionnées dans la boîte de dialogue Régression linéaire de Bayes principale. Lorsque des variables de facteur et de covariable sont ajoutées ou retirées depuis la boîte de dialogue principale, la liste est mise à jour en conséquence.

Variable(s) à tester

Sélectionnez dans la liste **Variables disponibles** les variables de facteur/covariable à tester et ajoutez-les à la liste **Variables à tester**.

Remarque : L'option **Inclure le terme constante** doit être sélectionnée si aucun facteur ou covariable de test n'ont été sélectionnés.

Variable(s) à tester et valeur(s)

Spécifiez les valeurs à tester. Le nombre de valeurs doit correspondre au nombre de paramètres du modèle d'origine. Lorsque des valeurs sont spécifiées, la première doit se référer au terme de la constante (en supposant que les valeurs sont 0 si elles n'ont pas été définies explicitement).

Terme Inclure constante

Lorsque cette option est sélectionnée, les termes de constante sont inclus dans le test. Par défaut, ce paramètre n'est pas sélectionné.

Si cette option est activée, utilisez la zone **Valeur à tester** pour spécifier une valeur.

Libellé du test (facultatif)

Vous pouvez spécifier un libellé pour chaque test. Vous pouvez spécifier une valeur chaîne d'une longueur maximale de 255 octets. Un seul libellé est autorisé pour chaque test F.

ANOVA à 1 facteur de Bayes

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

La procédure ANOVA à 1 facteur permet d'effectuer une analyse de variance univariée sur une variable quantitative dépendante par une variable critère simple (indépendant). L'analyse de variance sert à tester l'hypothèse d'égalité des moyennes. SPSS Statistics prend en charge les facteurs de Bayes, les distributions a priori conjuguées et les distributions a priori non informatives.

1. A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > ANOVA unidirectionnel

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez une seule variable numérique **dépendante** dans la liste **Variables**. Vous devez sélectionner au moins une variable.

3. Sélectionnez une seule variable **Facteur** pour le modèle dans la liste **Variables disponibles**. Vous devez sélectionner au moins une variable **Facteur**.

4. Sélectionnez dans la liste **Variables disponibles** une variable unique, non chaîne, pour faire office de **Pondération** de régression. La zone **Pondération** de la variable peut être vide.

5. Sélectionnez l'**Analyse de Bayes** voulue :

- **Caractériser la distribution a posteriori** : lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.
- **Estimer le facteur de Bayes** : lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

Tableau 9. Seuils couramment utilisés pour définir la signification des constatations

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H0 : hypothèse nulle

H1 : hypothèse alternative

15

16

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.

(En option) Vous avez également ces possibilités :

- Cliquez sur **Critères** pour spécifier le pourcentage d'intervalle crédible et les paramètres de méthode numérique.
- Cliquez sur **Distributions à priori** pour définir les paramètres de référence et de distribution à priori conjuguée.
- Cliquez sur **Facteur de Bayes** pour spécifier des paramètres de facteur de Bayes.
- Cliquez sur **Tracés** pour déterminer les tracés à générer en sortie.

ANOVA à 1 facteur de Bayes : critères

Vous pouvez spécifier les critères d'analyse suivants pour les modèles ANOVA à 1 facteur de Bayes.

Pourcentage d'intervalle crédible

Spécifiez le niveau de signification pour les intervalles de calcul crédibles. Niveau par défaut : 95 %.

Remarque : Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour **Analyse de Bayes**.

Méthode numérique

Spécifiez la méthode numérique à utiliser pour estimation de l'intégrale.

Tolérance

Spécifiez la valeur de tolérance pour les méthodes numériques. Valeur par défaut : 0.000001.

Nombre maximum d'itérations

Spécifiez le nombre maximal d'itérations de la méthode. Il doit s'agir d'un nombre entier positif. Valeur par défaut : 2000.

¹⁵ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹⁶ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

ANOVA à 1 facteur de Bayes : probabilités a priori

Vous pouvez spécifier les paramètres de distribution a priori suivants pour les paramètres de régression et la variance des erreurs. Les options suivantes ne sont disponibles que si l'option **Caractériser la distribution a posteriori** est sélectionnée pour **Analyse de Bayes**.

Remarque : De nombreux chercheurs peuvent contester la nécessité d'indiquer une distribution a priori. Les distributions a priori de référence minimisent la préoccupation que cette distribution soit submergée lorsque les données augmentent. Lorsqu'une distribution a priori informative est spécifiée, les méthodes bayésiennes peuvent utiliser efficacement les informations. L'exigence d'utilisation d'une distribution a priori ne doit pas vous dissuader d'utiliser une analyse bayésienne.

Référence

Lorsque cette option est sélectionnée, l'analyse des références génère une inférence bayésienne objective. Les instructions inférentielles ne dépendent que du modèle supposé et des données disponibles, et la distribution a priori utilisée pour générer une inférence est la moins informative. Il s'agit du paramètre par défaut.

Conjuguer

Fournit des options pour définir des distributions à priori conjuguées. Les distributions à priori conjuguées supposent une distribution jointe Normale-Inverse-Gamma. Bien que des distributions à priori conjuguées ne soient pas obligatoires pour effectuer des mises à jour de Bayes, elles facilitent les processus de calcul.

Distributions à priori sur variance des erreurs

Paramètre de forme

Spécifiez le paramètre de forme α_0 pour la distribution Inverse-Gamma. Vous devez entrer une seule valeur, supérieure à 0.

Paramètre d'échelle

Spécifiez le paramètre d'échelle b_0 pour la distribution Inverse-Gamma. Vous devez entrer une seule valeur, supérieure à 0. Plus le paramètre d'échelle est grand, plus la distribution est étendue.

Probabilités a priori sur paramètres de régression

Spécifiez le vecteur de moyenne β_0 pour les moyennes des groupes. Le nombre de valeurs doit correspondre à celui des paramètres de régression, y-compris le terme de la constante.

La colonne **Variables** est automatiquement alimentée avec les niveaux du facteur. La colonne **Moyenne** n'inclut pas de valeurs par défaut.

Cliquez sur **Réinitialiser** pour effacer les valeurs.

Matrice de variance de la covariance : $\sigma^2\mathbf{x}$

Spécifiez les valeurs V_0 dans le triangle inférieur de la matrice de variance-covariance pour la distribution à priori normale multivariée. Notez que V_0 doit être définie comme semi-positif. Seul le triangle inférieur de la table doit être renseigné.

Les lignes et les colonnes sont automatiquement alimentées avec les niveaux du facteur. Toutes les valeurs sur la diagonale indiquent 1 ; toutes les valeurs hors diagonale indiquent 0.

Cliquez sur **Réinitialiser** pour effacer les valeurs.

Utiliser la matrice d'identité

Lorsque cette option est sélectionnée, la matrice d'identité est utilisée. Vous ne pouvez pas spécifier de valeurs V_0 dans le triangle inférieur de la matrice de variance-covariance pour la distribution a priori normale multivariée.

ANOVA à 1 facteur de Bayes : facteur de Bayes

Vous pouvez spécifier l'approche à utiliser pour estimer le facteur de Bayes pour les modèles d'analyse ANOVA à 1 facteur de Bayes. Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour l'analyse de Bayes.

Calcul

Spécifiez l'approche à utiliser pour estimer les facteurs de Bayes. Par défaut, il s'agit de la méthode JZS.

Méthode JZS

Lorsque cette option est sélectionnée, elle appelle l'approche de Zellner-Siow. Il s'agit du paramètre par défaut.

Méthode Zellner

Lorsque cette option est sélectionnée, elle appelle l'approche de Zellner et vous devez spécifier une valeur a priori unique $g > 0$ (il n'existe pas de valeur par défaut).

Méthode hyper à priori

Lorsque cette option est sélectionnée, elle appelle l'approche hyper-g et vous devez spécifier un paramètre de forme α_0 pour la distribution Inverse-Gamma. Vous devez spécifier une seule valeur > 0 (valeur par défaut : 3).

Méthode de Rouder

Lorsque cette option est sélectionnée, elle appelle l'approche de Rouder et vous devez spécifier un paramètre d'échelle b_0 pour la distribution Inverse-Gamma. Vous devez spécifier une seule valeur > 0 (valeur par défaut : 1).

ANOVA à 1 facteur de Bayes : tracés

Vous pouvez contrôler les tracés générés en sortie.

Tracés des groupes

Spécifiez les sous-groupes pour lesquels générer un tracé. Tracez la vraisemblance, la distribution a priori et la distribution a posteriori pour les moyennes des groupes spécifiés. La liste **Groupes** est un sous-ensemble des catégories de la variable de facteur, le format doit donc être cohérent avec le type de données et les valeurs effectives du facteur.

Variance des termes d'erreur

Si ce paramètre est sélectionné, la variance des erreurs figure sur le tracé. Ce paramètre est désélectionné par défaut. Cette option n'est pas disponible lorsque **Estimer le facteur de Bayes** est sélectionné pour l'analyse de Bayes.

Modèles log-linéaires bayésiens

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

La conception du test d'indépendance de deux facteurs nécessite deux variables catégorielles pour la construction d'un tableau de contingence et génère une inférence bayésienne sur l'association ligne-colonne. Vous pouvez estimer les facteurs de Bayes en supposant différents modèles et en caractérisant la distribution a posteriori désirée en simulant l'intervalle crédible simultané pour les termes d'interaction.

1. A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > Modèles Log-linéaires

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

2. Sélectionnez une seule variable de ligne, non d'échelle, dans la liste **Variables disponibles**. Vous devez sélectionner au moins une variable non d'échelle.

3. Sélectionnez une seule variable de colonne, non d'échelle, dans la liste **Variables disponibles**. Vous devez sélectionner au moins une variable non d'échelle.

4. Sélectionnez l'**Analyse de Bayes** voulue :

- **Caractériser la distribution a posteriori** : lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant

les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.

- **Estimer le facteur de Bayes** : lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

Tableau 10. Seuils couramment utilisés pour définir la signification des constatations

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H0 : hypothèse nulle

H1 : hypothèse alternative

17

18

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.

(En option) Vous avez également ces possibilités :

- Cliquez sur **Critères** pour spécifier le pourcentage d'intervalle crédible et les paramètres de méthode numérique.
- Cliquez sur **Facteur de Bayes** pour spécifier des paramètres de facteur de Bayes.
- Cliquez sur **Impression** pour spécifier comment afficher le contenu dans les tableaux en sortie.

Modèles log-linéaires bayésiens : Critères

Vous pouvez spécifier les critères d'analyse suivants pour les modèles log-linéaires bayésiens.

Pourcentage d'intervalle crédible

Spécifiez le niveau de signification pour les intervalles de calcul crédibles. Niveau par défaut : 95 %.

Méthode numérique

Spécifiez la méthode numérique à utiliser pour estimation de l'intégrale.

Définir une valeur de départ personnalisée

Lorsque cette option est sélectionnée, vous pouvez définir une valeur de départ personnalisée dans la zone **Valeur de départ**. Spécifiez une valeur de départ aléatoire. Il doit s'agir d'un nombre entier positif. Par défaut, une valeur de départ aléatoire est affectée.

¹⁷ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

¹⁸ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Remarque : Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour **Analyse de Bayes**.

Tolérance

Spécifiez la valeur de tolérance pour les méthodes numériques. Valeur par défaut : 0.000001.

Nombre maximum d'itérations

Spécifiez le nombre maximal d'itérations de la méthode. Il doit s'agir d'un nombre entier positif. Valeur par défaut : 2000.

Echantillons simulés pour distribution a posteriori

Spécifiez le nombre d'échantillons utilisé pour composer la distribution a posteriori désirée. La valeur par défaut est 10 000.

Format

Sélectionnez les catégories à afficher en ordre **Ascendant** ou **Descendant**. L'ordre par défaut est l'ordre ascendant.

Modèles log-linéaires bayésiens : Facteur de Bayes

Vous pouvez spécifier le modèle supposé pour les données observées (Poisson, Multinomiale ou Non paramétrique). La distribution multinomiale est le paramètre par défaut. Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour l'analyse de Bayes.

Modèle de Poisson

Lorsque cette option est sélectionnée, le modèle de Poisson est supposé être celui pertinent pour les données observées.

Modèle multinomial

Lorsque cette option est sélectionnée, le modèle multinomial est supposé être celui pertinent pour les données observées. Il s'agit du paramètre par défaut.

Marges fixes

Sélectionnez **Total général**, **Somme ligne** ou **Somme colonne** pour spécifier les totaux marginaux fixes pour le tableau de contingence. **Total général** est le paramètre par défaut.

Distribution a priori

Spécifiez le type de distribution a priori lors de l'estimation du facteur de Bayes.

Conjuguer

Sélectionnez cette option pour spécifier une distribution a priori conjuguée. Utilisez le tableau **Paramètres de forme** pour spécifier les paramètres de forme a_{rs} pour la distribution Gamma. Vous devez spécifier les paramètres de forme lorsque **Conjugué** est sélectionné comme type de distribution a priori.

Lorsqu'une seule valeur est spécifiée, tous les a_{rs} sont supposés être égaux à cette valeur. $a_{rs} = 1$ est le paramètre par défaut. Si vous devez spécifier plusieurs valeurs, vous pouvez les séparer par des espaces.

Le nombre de valeurs numériques qui sont spécifiées dans chaque ligne et chaque colonne doit correspondre à la dimension du tableau de contingence. Toutes les valeurs spécifiées doivent être > 0 .

Cliquez sur **Réinitialiser** pour effacer les valeurs.

Paramètre d'échelle

Spécifiez le paramètre d'échelle b pour la distribution Gamma. Vous devez spécifier une seule valeur > 0 .

Mélange Dirichlet

Sélectionnez cette option pour spécifier une distribution a priori de mélange Dirichlet.

Intrinsèque

Sélectionnez cette option pour spécifier une distribution a priori intrinsèque.

Modèle non paramétrique

Lorsque cette option est sélectionnée, le modèle non paramétrique est supposé être celui pertinent pour les données observées.

Marges fixes

Sélectionnez **Somme ligne** ou **Somme colonne** pour spécifier les totaux marginaux fixes pour le tableau de contingence. **Somme ligne** est le paramètre par défaut.

Distribution a priori

Spécifiez les paramètres pour les distributions a priori de Dirichlet. Vous devez spécifier les paramètres **Distribution a priori** quand **Modèle non paramétrique** est sélectionné. Lorsqu'une seule valeur est spécifiée, tous les λ_s sont supposés être égaux à cette valeur. $\lambda_s = 1$ est le paramètre par défaut. Si vous devez spécifier plusieurs valeurs, vous pouvez les séparer par des espaces. Toutes les valeurs spécifiées doivent être > 0 . Le nombre de valeurs numériques spécifiées doit correspondre à la dimension de la ligne ou de la colonne, laquelle n'est pas fixe pour le tableau de contingence.

Cliquez sur **Réinitialiser** pour effacer les valeurs.

Modèles log-linéaires bayésiens : Imprimer

Vous pouvez spécifier comment afficher le contenu dans les tableaux en sortie.

Conception de table

Supprimer la table

Lorsque cette option est sélectionnée, le tableau de contingence n'est pas inclus dans la sortie. Ce paramètre n'est pas activé par défaut.

Remarque : Les paramètres suivants sont sans effet lorsque le paramètre **Supprimer le tableau** est activé.

Statistiques

Spécifiez les statistiques pour le test d'indépendance.

Khi-carré

Sélectionnez cette option pour calculer la statistique Khi-carré de Pearson, les degrés de liberté et la signification asymptotique bilatérale. Dans le cas d'un tableau de contingence 2x2, ce paramètre calcule également les statistiques de continuité Yates corrigées, les degrés de liberté et la signification asymptotique bilatérale associée. Dans le cas d'un tableau de contingence 2x2, avec au moins un nombre de cellules < 5 , ce paramètre calcule également la signification exacte bilatérale et unilatérale du test exact de Fisher.

Rapport de vraisemblance

Sélectionnez cette option pour calculer la statistique de test du ratio de vraisemblance, les degrés de liberté et la signification asymptotique bilatérale associée.

Effectifs

Spécifiez les types de comptage à inclure dans le tableau de contingence.

Observé

Sélectionnez cette option pour inclure les nombres de cellules observés dans le tableau de contingence.

Attendu

Sélectionnez cette option pour inclure les nombres de cellules attendus dans le tableau de contingence.

Pourcentages

Spécifiez les types de pourcentages à inclure dans le tableau de contingence.

Ligne

Sélectionnez cette option pour inclure les pourcentages de ligne dans le tableau de contingence.

Colonne

Sélectionnez cette option pour inclure les pourcentages de colonne dans le tableau de contingence.

Total

Sélectionnez cette option pour inclure les pourcentages totaux dans le tableau de contingence.

Modèles d'analyse ANOVA bayésienne à un facteur avec mesures répétées

Cette fonction requiert l'option SPSS Statistics Standard Edition ou Statistiques avancées.

Les modèles d'analyse ANOVA (analyse de variance) bayésienne à un facteur supposent qu'il existe une mesure unique par sujet. Cependant, cette supposition n'est pas toujours vraie. Il n'est pas rare qu'une enquête vise à étudier les réponses moyenne sur plusieurs points temporels ou conditions. La procédure ANOVA bayésienne à un facteur avec mesures répétées mesure un seul facteur sur un même sujet à chaque point temporel ou condition et permet de croiser des sujets à l'intérieur des niveaux. Il est supposé que chaque sujet est associé à une observation unique pour chaque point temporel ou condition (dans ce cas, l'interaction sujet-traitement d'interaction n'est pas prise en compte).

1. A partir des menus, sélectionnez :

Analyser > Statistiques bayésiennes > ANOVA à un facteur avec mesures répétées

Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.

- Sélectionnez au moins deux variables **Mesures répétées** dans la liste **Variables disponibles**.
- Vous pouvez également sélectionner dans la liste **Variables disponibles** une variable unique pour la **pondération** de régression. La zone **Pondération** de la variable peut être vide.

Remarque : La liste des variables disponibles recense toutes celles disponibles, excepté celles de type Chaîne.

4. Sélectionnez l'**Analyse de Bayes** voulue :

- **Caractériser la distribution a posteriori** : lorsque cette option est sélectionnée, l'inférence de Bayes est réalisée depuis une perspective de caractérisation de distributions a posteriori. Vous pouvez examiner la distribution à postériori marginale du ou des paramètres pertinents en éliminant les paramètres intempestifs et en construisant ensuite des intervalles crédibles pour induire une inférence directe. Il s'agit du paramètre par défaut.
- **Estimer le facteur de Bayes** : lorsque cette option est sélectionnée, l'estimation des facteurs de Bayes (une des méthodologies importantes dans l'inférence de Bayes) constitue un ratio naturel pour comparaison des probabilités marginales entre une hypothèse Null et une hypothèse alternative.

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
>100	Evidence extrême pour H0	1 à 3	Evidence anecdotique pour H0	1/30 à 1/10	Forte évidence pour H1
30 à 100	Très forte évidence pour H0	1	Pas d'évidence	1/100 à 1/30	Très forte évidence pour H1
10 à 30	Forte évidence pour H0	1/3 à 1	Evidence anecdotique pour H1	1/100	Evidence extrême pour H1

Tableau 11. Seuils couramment utilisés pour définir la signification des constatations (suite)

Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation	Facteur de Bayes	Catégorie de constatation
3 à 10	Evidence modérée pour H0	1/10 à 1/3	Evidence modérée pour H1		

H0 : hypothèse nulle

H1 : hypothèse alternative

19

20

- **Utiliser les deux méthodes** : lorsque cette option est sélectionnée, les méthodes d'inférence **Caractériser la distribution a posteriori** et **Estimer le facteur de Bayes** sont toutes deux utilisées.

(En option) Vous avez également ces possibilités :

- Cliquez sur **Critères** pour spécifier le pourcentage d'intervalle crédible et les paramètres de méthode numérique.
- Cliquez sur **Facteur de Bayes** pour spécifier des paramètres de facteur de Bayes.
- Cliquez sur **Tracés** pour tracer les distributions a posteriori des moyennes de groupe.

ANOVA bayésienne à 1 facteur avec mesures répétées : critères

Vous pouvez spécifier les critères d'analyse suivants pour les modèles d'analyse ANOVA bayésienne à 1 facteur avec mesures répétées.

Pourcentage d'intervalle crédible

Spécifiez le niveau de signification pour les intervalles de calcul crédibles. Niveau par défaut : 95 %.

Méthode numérique

Spécifiez la méthode numérique à utiliser pour estimation de l'intégrale.

Définir une valeur de départ personnalisée

Lorsque cette option est sélectionnée, vous pouvez définir une valeur de départ personnalisée dans la zone **Valeur de départ**. La valeur par défaut est 2 000 000. Cette valeur doit être un entier positif compris entre 1 et 2 147 483 647. Par défaut, une valeur de départ aléatoire est affectée.

Nombre d'échantillons de Monte Carlo

Spécifiez le nombre de points d'échantillonnage à utiliser pour l'approximation de Monte Carlo. Cette valeur doit être un entier positif compris entre 10^3 et 10^6 . La valeur par défaut est 30 000.

ANOVA bayésienne à 1 facteur avec mesures répétées : facteur de Bayes

Vous pouvez spécifier l'approche à utiliser pour estimer le facteur de Bayes pour les modèles d'analyse ANOVA bayésienne à 1 facteur avec mesures répétées. Les options suivantes ne sont disponibles que si l'option **Estimer le facteur de Bayes** ou **Utiliser les deux méthodes** est sélectionnée pour l'analyse de Bayes.

Critères d'informations bayésiens (BIC)

Ce paramètre applique une extension de l'approximation BIC au plan des mesures répétées pour estimer les facteurs de Bayes. Il dérive la taille effective de l'échantillon en tenant compte de la corrélation des mesures répétées et suggère un terme de pénalité amélioré lors de l'estimation des critères BIC pour faire un choix entre deux modèles concurrents. Il s'agit du paramètre par défaut.

¹⁹ Lee, M.D., et Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

²⁰ Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

Modèle mixte de Rouder

Utilise les généralisations multivariées de la distribution de Cauchy comme valeur préalable pour la taille d'effet standardisée et comme valeur préalable non informative pour la variance.

Remarque : Le paramètre de pondération de fréquence globale et la pondération de régression sont ignorés lorsque cette option est sélectionnée.

ANOVA bayésienne à 1 facteur avec mesures répétées : tracés

Vous pouvez contrôler les tracés générés en sortie pour illustrer les distributions a posteriori des moyennes de groupe. Le tableau répertorie toutes les variables sélectionnées comme mesures répétées dans la boîte de dialogue Variables. Sélectionnez les variables de mesures répétées à tracer.

Kernel Ridge Regression

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

Exemple

Statistiques

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

Remarques sur les données

Données

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

Hypothèses

Obtention d'une régression de la crête du noyau

1. A partir des menus, sélectionnez :

Analyser > Régression > Crête du noyau...

2. Sélectionnez une **variable** dépendante.

3. Sélectionnez une ou plusieurs variables **Indépendante(s)**.

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

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

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

- Cliquez sur le contrôle d'ajout (+) pour inclure des fonctions de noyau supplémentaires.
- Cliquez sur la cellule vide dans la colonne **Noyau** pour sélectionner une fonction de noyau.
- Cliquez deux fois sur une cellule de ligne de fonction du noyau pour spécifier les valeurs des paramètres de fonction du noyau pour la colonne correspondante (**Alpha, Gamma, Coef0, Degré**). Pour plus d'informations, voir «Paramètres de noyau», à la page 143. Les paramètres d'optimisation de la fonction de noyau par défaut sont répertoriés ci-dessous.

Additive_CHI2

ALPHA=1 GAMMA=1

CHI2

ALPHA=1 GAMMA=1

Cosinus

ALPHA=1

Laplacien

ALPHA=1 GAMMA=1/p

Linéaire

Fonction de noyau par défaut. ALPHA=1

Polynomial

ALPHA=1 GAMMA=1/p COEF0=1 DEGREE=3

RBF

ALPHA=1 GAMMA=1/p

Ogive de Galton

ALPHA=1 GAMMA=1/p COEF0=1

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

- Si vous le souhaitez, cliquez sur **Options** pour indiquer le nombre de plis de validation croisée, les options d'affichage, les paramètres de tracé et les éléments à sauvegarder. Pour plus d'informations, voir «Régression de la crête du noyau: Options», à la page 144.
- Cliquez sur **OK**.

Paramètres de noyau

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

Spécifier les paramètres uniques

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

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

Spécifier les paramètres de grille

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

Régression de la crête du noyau: Options

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

Nombre de plis de validation croisée

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

Afficher

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

Meilleures salutations,

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

Comparer

Affiche les résultats de base pour tous les modèles évalués.

Comparer modèles et niveaux

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

Tracé

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

Observations versus Prévisions

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

Résidus versus Prévisions

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

Sauvegarder

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

Valeurs prédites

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

Résidus

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

Coefficients doubles

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

Remarques

Le présent document concerne des produits et des services disponibles dans différents pays. Il 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. Consultez votre représentant IBM local pour obtenir des informations sur les produits et services actuellement disponibles dans votre région. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service IBM puisse être utilisé. Tout produit, programme ou service fonctionnellement équivalent qui ne porte pas atteinte à un droit de propriété intellectuelle IBM peut être utilisé à la place. 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 toute demande d'informations relatives au jeu de caractères codé sur deux octets, contactez le service de propriété intellectuelle IBM ou envoyez vos questions par courrier 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 CONTREFAÇON 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/ou programmes 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 de l'ICA (IBM Customer Agreement), des Conditions internationales d'utilisation des logiciels IBM ou de tout autre accord 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 programmes exemples 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 programmes exemples 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 n'est en aucun cas responsable des dommages liés à l'utilisation de ces 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 :

© Copyright IBM Corp. 2021. Des parties de ce code sont proviennent d'IBM Corp. Exemples de programmes.

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

Marques

IBM, le logo IBM et ibm.com sont des marques d'International Business Machines Corporation 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 dans certains autres pays.

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

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

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

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

Index

A

AFT de survie
Boîte de dialogue de survie-Variables de catégorie [83](#)

Analyse de la covariance
GLM - Tests multivariés [1](#)

Analyse de survie
dans la régression d'arête de noyau [142](#)
Kaplan-Meier [95](#)
Modèle de Cox chronologique [101](#)
Régression de Cox [97](#)
Tables de survie [75](#)

analyse de variance
dans Composantes de la Variance [20](#)
dans les modèles linéaires mixtes généralisés [52](#)

Analyse log-linéaire
Analyse log-linéaire généralisée [69](#)
Analyse log-linéaire Logit [72](#)
dans les modèles linéaires mixtes généralisés [52](#)

Analyse log-linéaire généralisée
Contrastes [69](#)
Covariables de cellule [69](#)
critères [71](#)
Distribution des effectifs des cellules [69](#)
enregistrement de prévisions [71](#)
enregistrement de variables [71](#)
facteurs [69](#)
fonctions supplémentaires de la commande [72](#)
intervalles de confiance [71](#)
Options d'affichage [71](#)
Résidus [71](#)
spécification du modèle [70](#)
Structures de cellule [69](#)
tracés [71](#)

Analyse log-linéaire Logit
Contrastes [72](#)
Covariables de cellule [72](#)
critères [74](#)
Distribution des effectifs des cellules [72](#)
enregistrement de variables [74](#)
facteurs [72](#)
intervalles de confiance [74](#)
Options d'affichage [74](#)
Prévisions [74](#)
Résidus [74](#)
spécification du modèle [73](#)
Structures de cellule [72](#)
tracés [74](#)

Analyse log-linéaire Sélection du modèle
définition des plages de facteurs [68](#)
fonctions supplémentaires de la commande [69](#)
modèles [68](#)
Options [68](#)

analyse par la méthode des probits
modèles linéaires mixtes généralisés [52](#)

ANOVA

ANOVA (*suite*)
GLM - Mesures répétées [9](#)
GLM - Tests multivariés [1](#)

ANOVA multivariée [1](#)

Arête de noyau
alpha [142](#)
choix du modèle [142](#)
coef0 [142](#)
degré [142](#)
gamma [142](#)
Modèle unique [142](#)

B

B de Tukey
GLM - Mesures répétées [15](#)
GLM - Tests multivariés [5](#)

Bonferroni
GLM - Mesures répétées [15](#)
GLM - Tests multivariés [5](#)

C

C de Dunnett
GLM - Mesures répétées [15](#)
GLM - Tests multivariés [5](#)

catégorie de référence
dans Modèles linéaires généralisés [33](#)
équations d'estimation généralisées [45](#), [46](#)

classe génératrice
Analyse log-linéaire : Sélection du modèle [68](#)

Composantes de variance
enregistrement des résultats [21](#)
fonctions supplémentaires de la commande [22](#)
modèle [20](#)
Options [20](#)

Contrastes
dans Analyse log-linéaire généralisée [69](#)
dans Analyse log-linéaire Logit [72](#)
Régression de Cox [98](#)

Convergence de Hess
dans Modèles linéaires généralisés [35](#)
équations d'estimation généralisées [47](#)

Convergence de log de vraisemblance
dans Modèles linéaires généralisés [35](#)
équations d'estimation généralisées [47](#)
Modèles mixtes linéaires [27](#)

Convergence des paramètres
dans Modèles linéaires généralisés [35](#)
équations d'estimation généralisées [47](#)
Modèles mixtes linéaires [27](#)

Covariables
Régression de Cox [98](#)

covariables de chaîne
Régression de Cox [98](#)

D

- décomposition hiérarchique
 - dans Composantes de la Variance [21](#)
- Différence la moins significative
 - GLM - Mesures répétées [15](#)
 - GLM - Tests multivariés [5](#)
- Différence significative de Tukey
 - GLM - Mesures répétées [15](#)
 - GLM - Tests multivariés [5](#)
- distance de Cook
 - dans Modèles linéaires généralisés [38](#)
 - GLM [7](#)
 - GLM - Mesures répétées [16](#)
- distribution binomiale
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- distribution binomiale négative
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- Distribution de Poisson
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- distribution gamma
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- distribution gaussienne inverse
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- distribution multinomiale
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- distribution normale
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- distribution Tweedie
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)

E

- Effet aléatoire a priori
 - dans Composantes de la Variance [20](#)
- Effets aléatoires
 - Modèles mixtes linéaires [26](#), [29](#)
- effets fixes
 - Modèles mixtes linéaires [25](#)
- élimination descendante
 - Analyse log-linéaire : Sélection du modèle [67](#)
- équations d'estimation généralisées
 - catégorie de référence pour les réponses binaires [45](#)
 - Critères d'estimation [47](#)
 - enregistrement des variables dans le jeu de données actif [50](#)
 - exporter le modèle [51](#)
 - Moyennes marginales estimées [49](#)
 - options des facteurs catégoriels [46](#)
 - prédicteurs [45](#)
 - réponse [44](#)
 - spécification du modèle [46](#)
 - Statistiques [48](#)
 - type de modèle [42](#)
 - valeurs initiales [48](#)
- erreur standard

- erreur standard (*suite*)
 - GLM [7](#)
 - GLM - Mesures répétées [16](#)
- Estimation du maximum de vraisemblance
 - dans Composantes de la Variance [20](#)
- Estimation du maximum de vraisemblance restreint
 - dans Composantes de la Variance [20](#)
- estimations des paramètres
 - Analyse log-linéaire : Sélection du modèle [68](#)
 - dans Analyse log-linéaire généralisée [69](#)
 - dans Analyse log-linéaire Logit [72](#)
 - dans Modèles linéaires généralisés [36](#)
 - équations d'estimation généralisées [48](#)
 - Modèles mixtes linéaires [28](#)
- évaluation
 - Modèles mixtes linéaires [27](#)
- Evaluation de Fisher
 - Modèles mixtes linéaires [27](#)

F

- F de R-E-G-W (Ryan-Einot-Gabriel-Welsch)
 - GLM - Mesures répétées [15](#)
 - GLM - Tests multivariés [5](#)
- F multiple de Ryan-Einot-Gabriel-Welsch
 - GLM - Mesures répétées [15](#)
 - GLM - Tests multivariés [5](#)
- facteurs
 - GLM - Mesures répétées [12](#)
- fonction de lien binomiale négative
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien Cauchit cumulée
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien complément log
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien d'identité
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien log
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien log-log complémentaire
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien log-log complémentaire cumulée
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien log-log négative
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien log-log négative cumulée
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien logit
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien logit cumulée
 - dans les équations d'estimation généralisées [42](#)
 - dans les modèles linéaires généralisés [30](#)
- fonction de lien probit
 - dans les équations d'estimation généralisées [42](#)

fonction de lien probit (*suite*)
dans les modèles linéaires généralisés [30](#)
fonction de lien probit cumulée
dans les équations d'estimation généralisées [42](#)
dans les modèles linéaires généralisés [30](#)
fonction de lien puissance
dans les équations d'estimation généralisées [42](#)
dans les modèles linéaires généralisés [30](#)
fonction de lien puissance du rapport des cotes
dans les équations d'estimation généralisées [42](#)
dans les modèles linéaires généralisés [30](#)
fonction de survie
Tables de survie [75](#)
fonction générale estimée
équations d'estimation généralisées [48](#)
fonctions d'estimation générale(s)
dans Modèles linéaires généralisés [36](#)
fonctions de lien
modèles linéaires mixtes généralisés [55](#)
frequencies
Analyse log-linéaire : Sélection du modèle [68](#)

G

GLM
Enregistrement de matrices [7](#)
enregistrement de variables [7](#)
GLM - Mesures répétées
définir des facteurs [12](#)
enregistrement de variables [16](#)
fonctions supplémentaires de la commande [18](#)
modèle [12](#)
Tests post hoc [15](#)
tracés de profil [14](#)
GLM - Multivarié
Covariables [1](#)
facteurs [1](#)
Tests post hoc [5](#)
tracés de profil [5](#)
Variable dépendante [1](#)
GLM – Tests multivariés [1](#)
GLOR
dans Analyse log-linéaire généralisée [69](#)
GT2 de Hochberg
GLM - Mesures répétées [15](#)
GLM - Tests multivariés [5](#)

H

historique des itérations
dans Modèles linéaires généralisés [36](#)
équations d'estimation généralisées [48](#)
Modèles mixtes linéaires [27](#)

I

Informations relatives au niveau de facteur
Modèles mixtes linéaires [28](#)
informations sur le modèle
dans Modèles linéaires généralisés [36](#)
équations d'estimation généralisées [48](#)
intervalles de confiance
dans Analyse log-linéaire généralisée [71](#)

intervalles de confiance (*suite*)
dans Analyse log-linéaire Logit [74](#)
Modèles mixtes linéaires [28](#)
itérations
Analyse log-linéaire : Sélection du modèle [68](#)
dans Modèles linéaires généralisés [35](#)
équations d'estimation généralisées [47](#)

K

Kaplan-Meier
comparaison des niveaux de facteur [96](#)
définition d'événements [96](#)
durée de survie de la moyenne et de la médiane [97](#)
enregistrement des nouvelles variables [97](#)
exemple [95](#)
fonctions supplémentaires de la commande [97](#)
quartiles [97](#)
Statistiques [95](#), [97](#)
tables de survie [97](#)
tendance linéaire des niveaux de facteur [96](#)
tracés [97](#)
variables de statut de survie [96](#)
Kernel Ridge Regression
affichage [144](#)
enregistrer [144](#)
paramètres [143](#)
paramètres de grille [143](#)
Plis de validation croisée [144](#)
tracés [144](#)

L

Log rapport des cotes généralisé
dans Analyse log-linéaire généralisée [69](#)
LSD de Fisher
GLM - Mesures répétées [15](#)
GLM - Tests multivariés [5](#)

M

matrice de corrélation
dans Modèles linéaires généralisés [36](#)
équations d'estimation généralisées [48](#)
Modèles mixtes linéaires [28](#)
matrice de covariance
dans Modèles linéaires généralisés [35](#), [36](#)
équations d'estimation généralisées [47](#), [48](#)
GLM [7](#)
Modèles mixtes linéaires [28](#)
Matrice de covariance à effet aléatoire
Modèles mixtes linéaires [28](#)
Matrice de covariance des paramètres
Modèles mixtes linéaires [28](#)
Matrice de covariance des résidus
Modèles mixtes linéaires [28](#)
matrice des coefficients de contraste
dans Modèles linéaires généralisés [36](#)
équations d'estimation généralisées [48](#)
Matrice L
dans Modèles linéaires généralisés [36](#)
équations d'estimation généralisées [48](#)
Méthode de Newton-Raphson

Méthode de Newton-Raphson (*suite*)
dans Analyse log-linéaire généralisée [69](#)
dans Analyse log-linéaire Logit [72](#)

MINQUE
dans Composantes de la Variance [20](#)

Modèle des risques proportionnels
Régression de Cox [97](#)

modèle linéaire général
modèles linéaires mixtes généralisés [52](#)

modèle linéaire généralisé
dans les modèles linéaires mixtes généralisés [52](#)

Modèles de fragilité paramétriques
variables de statut de survie [90](#)

Modèles de fragilité partagés paramétriques
analyse [84](#)
critères [85](#)
estimation [87](#)
exporter [90](#)
imprimer [88](#)
modèle [86](#)
prévoir [88](#)
tracer [89](#)

Modèles de temps d'échec accéléré paramétriques
analyse [77](#)
critères [78](#)
estimation [79](#)
exporter [83](#)
imprimer [81](#)
modèle [79](#)
prévoir [81](#)
tracer [82](#)

modèles factoriels complets
dans Composantes de la Variance [20](#)
GLM - Mesures répétées [12](#)

modèles hiérarchiques
modèles linéaires mixtes généralisés [52](#)

modèles linéaires généralisés
catégorie de référence pour les réponses binaires [33](#)
Critères d'estimation [35](#)
distribution [30](#)
enregistrement des variables dans le jeu de données
actif [38](#)
exporter le modèle [39](#)
fonctions de lien [30](#)
Moyennes marginales estimées [37](#)
options des facteurs catégoriels [34](#)
prédicteurs [33](#)
réponse [32](#)
spécification du modèle [34](#)
Statistiques [36](#)
types de modèle [30](#)
valeurs initiales [36](#)

modèles linéaires mixtes généralisés
bloc d'effets aléatoires [58](#)
coefficients fixes [64](#)
covariances à effet aléatoire [65](#)
décalage [59](#)
distribution de la cible [55](#)
Effets aléatoires [58](#)
effets fixes [57](#), [64](#)
enregistrer champs [62](#), [63](#)
exporter le modèle [62](#), [63](#)
fonctions de lien [55](#)
moyennes estimées [66](#)

modèles linéaires mixtes généralisés (*suite*)
Moyennes marginales estimées [61](#)
paramètres de covariance [65](#)
pondération d'analyse [59](#)
récapitulatif du modèle [63](#)
structure des données [63](#)
table de classification [64](#)
termes personnalisés [58](#)
valeurs prédites en fonction des valeurs observées [64](#)
vue du modèle [63](#)

Modèles log-linéaires hiérarchiques [67](#)

Modèles logit multinomiaux [72](#)

modèles longitudinaux
modèles linéaires mixtes généralisés [52](#)

modèles mixtes
linéaire [22](#)
modèles linéaires mixtes généralisés [52](#)

Modèles mixtes linéaires
Critères d'estimation [27](#)
Effets aléatoires [26](#), [29](#)
effets fixes [25](#)
enregistrement de variables [29](#)
fonctions supplémentaires de la commande [29](#)
modèle [28](#)
Moyennes marginales estimées [28](#)
Structure de covariance [105](#)
termes construits [25](#)
Termes d'interaction [25](#)

modèles multi-niveaux
modèles linéaires mixtes généralisés [52](#)

modèles personnalisés
Analyse log-linéaire : Sélection du modèle [68](#)
dans Composantes de la Variance [20](#)
GLM - Mesures répétées [12](#)

modèles saturés
Analyse log-linéaire : Sélection du modèle [68](#)

Moyennes marginales estimées
dans Modèles linéaires généralisés [37](#)
équations d'estimation généralisées [49](#)
Modèles mixtes linéaires [28](#)

N

Newman-Keuls
GLM - Mesures répétées [15](#)
GLM - Tests multivariés [5](#)

O

Observations censurées
Kaplan-Meier [95](#)
Régression de Cox [97](#)
Tables de survie [75](#)

P

paramètre d'échelle
dans Modèles linéaires généralisés [35](#)
équations d'estimation généralisées [47](#)

Plage multiple de Ryan-Einot-Gabriel-Welsch
GLM - Mesures répétées [15](#)
GLM - Tests multivariés [5](#)

Prédicteurs chronologiques segmentés

Prédicteurs chronologiques segmentés (*suite*)

Régression de Cox [101](#)

Prévisions

dans Analyse log-linéaire généralisée [71](#)

dans Analyse log-linéaire Logit [74](#)

Modèles mixtes linéaires [29](#)

prévisions fixes

Modèles mixtes linéaires [29](#)

Prévisions pondérées

GLM [7](#)

GLM - Mesures répétées [16](#)

Q

Q de R-E-G-W (Ryan-Einot-Gabriel-Welsch)

GLM - Mesures répétées [15](#)

GLM - Tests multivariés [5](#)

Qualité de l'ajustement

dans Modèles linéaires généralisés [36](#)

équations d'estimation généralisées [48](#)

R

rapport des cotes

dans Analyse log-linéaire généralisée [69](#)

récapitulatif de traitement des observations

dans Modèles linéaires généralisés [36](#)

équations d'estimation généralisées [48](#)

Régression de Cox

Contrastes [98](#)

Covariables [97](#)

covariables catégorielles [98](#)

covariables de chaîne [98](#)

définition d'événements [101](#)

DfBeta(s) [100](#)

enregistrement des nouvelles variables [100](#)

exemple [97](#)

fonction de hasard [100](#)

fonction de survie [100](#)

Fonctions de la ligne de base [100](#)

fonctions supplémentaires de la commande [101](#)

introduction et élimination pas à pas [100](#)

itérations [100](#)

Prédicteurs chronologiques [101](#)

résiduels partiels [100](#)

Statistiques [97](#), [100](#)

tracés [99](#)

variable de statut de survie [101](#)

régression de Poisson

dans Analyse log-linéaire généralisée [69](#)

modèles linéaires mixtes généralisés [52](#)

régression logistique

modèles linéaires mixtes généralisés [52](#)

régression logistique multinomiale

modèles linéaires mixtes généralisés [52](#)

Régression multivariée [1](#)

Résidu non standardisé

GLM [7](#)

GLM - Mesures répétées [16](#)

résiduels de Pearson

dans Modèles linéaires généralisés [38](#)

équations d'estimation généralisées [50](#)

Résidus

Résidus (*suite*)

Analyse log-linéaire : Sélection du modèle [68](#)

dans Analyse log-linéaire généralisée [71](#)

dans Analyse log-linéaire Logit [74](#)

dans Modèles linéaires généralisés [38](#)

équations d'estimation généralisées [50](#)

Modèles mixtes linéaires [29](#)

résidus au sens déviance

dans Modèles linéaires généralisés [38](#)

résidus de vraisemblance

dans Modèles linéaires généralisés [38](#)

Résidus standardisés

GLM [7](#)

GLM - Mesures répétées [16](#)

Résidus supprimés

GLM [7](#)

GLM - Mesures répétées [16](#)

S

Séparation

dans Modèles linéaires généralisés [35](#)

équations d'estimation généralisées [47](#)

somme des carrés

dans Composantes de la Variance [21](#)

Modèles mixtes linéaires [25](#)

Statistique de Wald

dans Analyse log-linéaire généralisée [69](#)

dans Analyse log-linéaire Logit [72](#)

Statistiques descriptives

dans Modèles linéaires généralisés [36](#)

équations d'estimation généralisées [48](#)

Modèles mixtes linéaires [28](#)

Step-halving

dans Modèles linéaires généralisés [35](#)

équations d'estimation généralisées [47](#)

Modèles mixtes linéaires [27](#)

structures de covariance

Modèles mixtes linéaires [105](#)

Student-Newman-Keuls

GLM - Mesures répétées [15](#)

GLM - Tests multivariés [5](#)

T

T2 de Tamhane

GLM - Mesures répétées [15](#)

GLM - Tests multivariés [5](#)

T3 de Dunnett

GLM - Mesures répétées [15](#)

GLM - Tests multivariés [5](#)

tableau croisé

Analyse log-linéaire : Sélection du modèle [67](#)

tableaux de contingence

dans Analyse log-linéaire généralisée [69](#)

Tables de survie

comparaison des niveaux de facteur [76](#)

exemple [75](#)

fonction de survie [75](#)

fonctions supplémentaires de la commande [77](#)

Statistiques [75](#)

Suppression de l'affichage de la table [76](#)

Taux de risque [75](#)

Tables de survie (*suite*)
 Test de Wilcoxon (Gehan) [76](#)
 tracés [76](#)
 variables de facteur [76](#)
 variables de statut de survie [83](#)

Taux de risque
 Tables de survie [75](#)

termes construits [3](#), [12](#), [20](#), [68](#), [71](#), [74](#)

Termes d'interaction
 Modèles mixtes linéaires [25](#)

Termes imbriqués
 dans Modèles linéaires généralisés [34](#)
 équations d'estimation généralisées [46](#)
 Modèles mixtes linéaires [25](#)

test de Breslow
 Kaplan-Meier [96](#)

Test de comparaison appariée de Gabriel
 GLM - Mesures répétées [15](#)
 GLM - Tests multivariés [5](#)

Test de comparaison appariée de Games et Howell
 GLM - Mesures répétées [15](#)
 GLM - Tests multivariés [5](#)

Test de Gehan
 Tables de survie [76](#)

Test de plage multiple de Duncan
 GLM - Mesures répétées [15](#)
 GLM - Tests multivariés [5](#)

test de Scheffé
 GLM - Mesures répétées [15](#)
 GLM - Tests multivariés [5](#)

test de Tarone-Ware
 Kaplan-Meier [96](#)

Test de Wilcoxon
 Tables de survie [76](#)

Test des paramètres de covariance
 Modèles mixtes linéaires [28](#)

Test du multiplicateur de Lagrange
 dans Modèles linéaires généralisés [36](#)

test log rank
 Kaplan-Meier [96](#)

Test t de Dunnett
 GLM - Mesures répétées [15](#)
 GLM - Tests multivariés [5](#)

Test t de Sidak
 GLM - Mesures répétées [15](#)
 GLM - Tests multivariés [5](#)

Test t de Waller-Duncan
 GLM - Mesures répétées [15](#)
 GLM - Tests multivariés [5](#)

Tolérance de singularité
 Modèles mixtes linéaires [27](#)

tracé de probabilités gaussien
 Analyse log-linéaire : Sélection du modèle [68](#)

tracés
 dans Analyse log-linéaire généralisée [71](#)
 dans Analyse log-linéaire Logit [74](#)

tracés de profil
 GLM - Mesures répétées [14](#)
 GLM - Tests multivariés [5](#)

valeurs influentes (*suite*)
 GLM [7](#)
 GLM - Mesures répétées [16](#)
 variables de mesures répétées
 Modèles mixtes linéaires [23](#)
 variables de sujets
 Modèles mixtes linéaires [23](#)
 vue du modèle
 dans les modèles linéaires mixtes généralisés [63](#)

V

valeurs influentes
 dans Modèles linéaires généralisés [38](#)

