

IBM SPSS Complex Samples 23

IBM

Important

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 59.

Informations sur le produit

La présente édition s'applique à la version 23.0.0 d'IBM SPSS Statistics et à toutes les éditions et modifications ultérieures sauf mention contraire dans les nouvelles éditions.

LE PRESENT DOCUMENT EST LIVRE EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFACON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE.

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

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

Vous pouvez également consulter les serveurs Internet suivants :

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

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

© Copyright IBM France 2014. Tous droits réservés.

Table des matières

Avis aux lecteurs canadiens v

Chapitre 1. Introduction aux procédures d'échantillons complexes 1

Propriétés des échantillons complexes 1
Utilisation des procédures Echantillons complexes 2
 Fichiers de plan 2
Références supplémentaires 3

Chapitre 2. Echantillonnage depuis un plan complexe 5

Création d'un plan d'échantillonnage 5
Assistant d'échantillonnage : Variables du plan 5
 Contrôles d'arbre de navigation de l'assistant d'échantillonnage 6
Assistant d'échantillonnage : Méthode d'échantillonnage 6
Assistant d'échantillonnage : Taille de l'échantillon 7
 Définition de tailles inégales 8
Assistant d'échantillonnage : Variables destination 8
Assistant d'échantillonnage : Récapitulatif du plan 9
Assistant d'échantillonnage - Réaliser un échantillon : Options de sélection 9
Assistant d'échantillonnage - Réaliser un échantillon : Fichiers de sortie 9
Assistant d'échantillonnage : Fin 10
Modification d'un plan d'échantillonnage existant 10
Assistant d'échantillonnage : Récapitulatif du plan 10
Exécution d'un plan d'échantillonnage existant 10
Fonctions supplémentaires des commandes SPLAN et CSSELECT 11

Chapitre 3. Préparation d'un échantillon complexe en vue d'une analyse 13

Création d'un plan d'analyse 13
Assistant de préparation d'analyse : Variables du plan 13
 Contrôles d'arbre de navigation de l'assistant d'analyse 14
Assistant de préparation d'analyse : Méthode d'estimation 14
Assistant de préparation d'analyse : Taille 14
 Définition de tailles inégales 15
Assistant de préparation d'analyse : Récapitulatif du plan 15
Assistant de préparation d'analyse : Fin 15
Modification d'un plan d'analyse existant 16
Assistant de préparation d'analyse : Récapitulatif du plan 16

Chapitre 4. Plan d'échantillonnages complexes 17

Chapitre 5. Echantillons complexes - Effectifs 19

Statistiques de fréquences des échantillons complexes 19
Valeurs manquantes d'échantillons complexes 20
Options d'échantillons complexes 20

Chapitre 6. Echantillons complexes - Caractéristiques. 23

Statistiques des descriptifs des échantillons complexes 23
Valeurs manquantes des descriptifs d'échantillons complexes 24
Options d'échantillons complexes 24

Chapitre 7. Tableaux croisés des échantillons complexes 25

Statistiques de tableaux croisés d'échantillons complexes 25
Valeurs manquantes d'échantillons complexes 26
Options d'échantillons complexes 27

Chapitre 8. Echantillons complexes - Rapports 29

Echantillons complexes – Rapports : Statistiques 29
Echantillons complexes – Rapports : Valeurs manquantes 30
Options d'échantillons complexes 30

Chapitre 9. Modèle linéaire général des échantillons complexes 31

Modèle linéaire général des échantillons complexes 32
Modèle linéaire général des échantillons complexes - Statistiques 32
Tests d'hypothèse des échantillons complexes 33
Modèle linéaire général des échantillons complexes - Moyennes estimées 33
Modèle linéaire général des échantillons complexes - Enregistrement 34
Modèle linéaire général des échantillons complexes - Options 35
Fonctions supplémentaires de la commande CSGLM 35

Chapitre 10. Régression logistique des échantillons complexes 37

Régression logistique des échantillons complexes - Catégorie de référence 37
Modèle de régression logistique des échantillons complexes 38

Régression logistique des échantillons complexes - Statistiques	38
Tests d'hypothèse des échantillons complexes . . .	39
Régression logistique des échantillons complexes - Rapports des cotes	40
Régression logistique des échantillons complexes - Enregistrement	40
Régression logistique des échantillons complexes - Options.	41
Fonctions supplémentaires de la commande CSLOGISTIC	41

Chapitre 11. Régression ordinale des échantillons complexes 43

Probabilités des réponses de régression ordinale des échantillons complexes	44
Modèle de régression ordinale des échantillons complexes	44
Régression ordinale des échantillons complexes - Statistiques	45
Tests d'hypothèse des échantillons complexes . . .	46
Régression ordinale des échantillons complexes - Rapports des cotes	46
Régression ordinale des échantillons complexes - Enregistrement	46
Régression ordinale des échantillons complexes - Options.	47

Fonctions supplémentaires de la commande CSORDINAL	48
---	----

Chapitre 12. Régression de Cox des échantillons complexes 49

Définition d'un événement	50
Prédicteurs	51
Définition d'un prédicteur chronologique . . .	51
Sous-groupes	51
Modèle	52
Statistiques	52
Tracés	53
Tests d'hypothèse	54
Enregistrement	54
Exportation	55
Options.	56
Fonctions supplémentaires de la commande CSCOXREG	57

Remarques 59

Marques	61
-------------------	----

Index 63

Avis aux lecteurs canadiens

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

Illustrations

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

Terminologie

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

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

Claviers

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

OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

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

Nomenclature

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

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

Brevets

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

Assistance téléphonique

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

Chapitre 1. Introduction aux procédures d'échantillons complexes

Les procédures d'analyse effectuées dans les packs logiciel traditionnels se basent sur les observations d'un fichier de données qui ne représentent qu'un échantillon aléatoire simple de la population ciblée. De nombreuses entreprises et de nombreux chercheurs trouvent cela insuffisant et estiment qu'il est plus économique et pratique d'obtenir des échantillons de façon plus structurée.

L'option Echantillons complexes vous permet de sélectionner un échantillon en fonction d'un plan complexe et d'incorporer les spécifications de plan dans l'analyse de données, garantissant ainsi la validité des résultats.

Propriétés des échantillons complexes

Un échantillon complexe peut différer d'un échantillon aléatoire simple de plusieurs façons. Dans un échantillon aléatoire simple, des unités d'échantillonnage individuelles sont sélectionnées aléatoirement à probabilité égale et sans remplacement (WOR), directement dans la population entière. A l'opposé, un échantillon complexe spécifique peut avoir une partie ou la totalité des fonctions suivantes :

Stratification : Un échantillonnage stratifié nécessite de sélectionner des échantillons indépendamment dans des sous-groupes de population ou des strates qui ne se chevauchent pas. Par exemple, les strates peuvent être des groupes socioéconomiques, des catégories d'emploi, des groupes d'âge ou des groupes ethniques. La stratification vous garantit des tailles d'échantillons adéquates pour les sous-groupes d'intérêt, améliore la précision des estimations globales et permet d'utiliser différentes méthodes d'échantillonnage strate par strate.

Classification : L'échantillonnage en cluster nécessite la sélection de groupes d'unités d'échantillonnage ou de clusters. Par exemple, les clusters peuvent être des écoles, des hôpitaux ou des zones géographiques, et les unités d'échantillonnage peuvent être des étudiants, des patients ou des habitants. La classification est commune dans les plans à plusieurs phases et dans les échantillons de zones (géographiques).

Phases multiples : Dans un échantillonnage à plusieurs phases, vous sélectionnez un échantillon de première phase basé sur les clusters. Vous créez ensuite un échantillon de deuxième phase en créant des sous-échantillons à partir des clusters sélectionnés. Si l'échantillon de la deuxième phase se base sur les sous-classes, vous pouvez ajouter une troisième phase à l'échantillon. Par exemple, dans la première phase d'une enquête, vous pouvez créer un échantillon de villes. Vous pouvez ensuite échantillonner les ménages à partir des villes sélectionnées. Enfin, vous pouvez interroger les individus composant les ménages sélectionnés. Les assistants d'échantillonnage et de préparation d'analyse vous permettent de définir un plan en trois phases.

Echantillonnage non aléatoire : Lorsqu'il est difficile d'obtenir une sélection aléatoirement, les unités peuvent être échantillonnées systématiquement (à intervalles fixes) ou séquentiellement.

Probabilités de sélection inégales : Lors de l'échantillonnage de clusters contenant des nombres d'unités inégaux, vous pouvez utiliser l'échantillonnage de probabilité proportionnelle à la taille (probability proportional to size - PPS) pour qu'une probabilité de sélection de cluster soit égale à la proportion d'unités qu'elle contient. L'échantillonnage PPS peut également utiliser des schémas de pondération plus généraux pour sélectionner les unités.

Echantillonnage sans restriction : Un échantillonnage sans restriction sélectionne des unités avec remplacement (WR). Par conséquent, une unité individuelle peut être sélectionnée plusieurs fois pour l'échantillon.

Pondérations d'échantillonnage : Les pondérations d'échantillonnage sont automatiquement calculées lors de la création d'un plan complexe et correspondent idéalement à la « fréquence » de chaque unité d'échantillonnage dans la population cible. Par conséquent, la somme des pondérations de l'échantillon doit estimer la taille de la population. Les procédures d'analyse d'échantillons complexes nécessitent des pondérations d'échantillonnage pour analyser correctement un échantillon complexe. Ces pondérations doivent être entièrement utilisées dans le module Echantillonnage et ne doivent pas être utilisées avec d'autres procédures d'analyse via la procédure Observations pondérées, qui traite les pondérations comme des répliques d'observations.

Utilisation des procédures Echantillons complexes

Votre utilisation des procédures Echantillons complexes dépend de vos besoins. Les principaux « types » d'utilisateurs sont ceux qui :

- Planifient et mènent des enquêtes en fonction de plans complexes et qui analysent éventuellement l'échantillon ultérieurement. L'outil principal des sondeurs est l'Assistant d'échantillonnage.
- Analysent les fichiers de données d'échantillons précédemment obtenus en fonction de plans complexes. Avant d'utiliser des procédures d'analyse d'échantillons complexes, vous devez utiliser l'Assistant de préparation d'analyse.

Indépendamment du type d'utilisateur que vous êtes, vous devez fournir les informations sur le plan dans les procédures d'échantillons complexes. Ces informations sont stockées dans un **fichier de plan**, pour que vous puissiez les réutiliser rapidement.

Fichiers de plan

Un fichier de plan contient les spécifications des échantillons complexes. Il existe deux types de fichiers de plan :

Plan d'échantillonnage : Les spécifications fournies dans l'assistant d'échantillonnage définissent un plan d'échantillonnage utilisé pour réaliser un échantillon complexe. Le fichier de plan d'échantillonnage contient ces spécifications. Le fichier de plan d'échantillonnage contient également un plan d'analyse par défaut qui utilise les méthodes d'estimation appropriées pour le plan d'échantillonnage spécifié.

Plan d'analyse : Contient des informations nécessaires aux procédures d'analyse d'échantillons complexes. Ces informations permettent de calculer correctement les estimations de variance d'un échantillon complexe. Le plan contient la structure de l'échantillon, les méthodes d'estimation de chaque phase et les références aux variables requises, telles que les pondérations d'échantillon. L'assistant de préparation d'analyse vous permet de créer et de modifier les plans d'analyse.

L'enregistrement de vos spécifications dans un fichier de plan présente plusieurs avantages :

- Un sondeur peut indiquer la première phase d'un plan d'échantillonnage à plusieurs phases et créer les unités de la première phase immédiatement, collecter les informations sur les unités d'échantillonnage de la deuxième phase, puis modifier le plan d'échantillonnage afin d'y inclure la deuxième phase.
- Un analyste qui n'a pas accès au fichier de plan d'échantillonnage peut spécifier un plan d'analyse et faire référence à ce plan depuis chaque procédure d'analyse d'échantillons complexes.
- Le créateur d'échantillons d'usage public à grande échelle peut publier le fichier de plan d'échantillonnage ; ainsi, les instructions sont simplifiées pour les analystes et ceux-ci ne sont pas dans l'obligation de spécifier leurs propres plans d'analyse.

Références supplémentaires

Pour plus d'informations sur les techniques d'échantillonnage, reportez-vous aux textes suivants :

Cochran, W. G. 1977. *Sampling Techniques*, 3rd ed. New York: John Wiley and Sons.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.

Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.

Särndal, C., B. Swensson, and J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Chapitre 2. Echantillonnage depuis un plan complexe

L'assistant d'échantillonnage vous guide lors de la création, de la modification ou de l'exécution d'un fichier de plan d'échantillonnage. Avant d'utiliser l'assistant, pensez à définir précisément la population cible, la liste des unités d'échantillonnage et le plan d'échantillonnage.

Création d'un plan d'échantillonnage

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Sélectionner un échantillon...
2. Sélectionnez **Concevoir un échantillon** et choisissez le nom de fichier de plan sous lequel enregistrer le plan d'échantillonnage.
3. Cliquez sur **Suivant** pour poursuivre la procédure avec l'assistant.
4. A l'étape Variables du plan, vous pouvez éventuellement définir des strates, des clusters et des pondérations d'échantillon d'entrée. Une fois ces éléments définis, cliquez sur **Suivant**.
5. A l'étape Méthode d'échantillonnage, vous pouvez éventuellement choisir une méthode de sélection des éléments.
Si vous sélectionnez l'échantillonnage de **Brewer** ou de **Murthy** avec PPS, vous pouvez cliquer sur **Fin** pour réaliser l'échantillon. Sinon, cliquez sur **Suivant**, puis effectuez les opérations suivantes :
6. A l'étape Taille de l'échantillon, spécifiez le nombre ou la proportion d'unités à échantillonner.
7. Vous pouvez désormais cliquer sur **Fin** pour réaliser l'échantillon.

Les étapes suivantes sont facultatives. Elles permettent :

- Sélectionnez les variables destination à enregistrer.
- D'ajouter une deuxième ou une troisième étape au plan.
- Définissez les différentes options de sélection : phases auxquelles les échantillons sont réalisés, valeur de départ de nombre aléatoire. Indiquez également si les valeurs manquantes de l'utilisateur doivent être traitées comme valeurs valides des variables de plan.
- Sélectionnez l'emplacement d'enregistrement des données de sortie.
- De coller vos sélections en tant que syntaxe de commande.

Assistant d'échantillonnage : Variables du plan

Cette étape vous permet de sélectionner les variables de stratification et de classification, et de définir des pondérations d'échantillon d'entrée. Vous pouvez également spécifier un libellé pour cette étape.

Stratifier par : La classification croisée des variables de stratification définit des sous-populations distinctes, ou strates. Des échantillons distincts sont disponibles pour chaque strate. Pour améliorer la précision de vos estimations, les unités comprises dans les strates doivent être homogènes pour les caractéristiques d'intérêt.

Clusters : Les variables de cluster définissent les groupes d'unités d'observation, ou clusters. Les clusters sont utiles lorsque l'échantillonnage direct des unités d'observation de la population est coûteux ou impossible. Vous pouvez alors échantillonner des clusters de population, puis échantillonner des unités d'observation à partir des clusters sélectionnés. Cependant, l'utilisation de clusters peut créer des corrélations entre les unités d'échantillonnage. Il peut donc en résulter une perte de précision. Pour minimiser ce risque, les unités comprises dans les clusters doivent être hétérogènes pour les caractéristiques d'intérêt. Vous devez définir au moins une variable de cluster afin de concevoir un plan à

plusieurs phases. Les clusters sont également nécessaires pour utiliser plusieurs méthodes d'échantillonnage. Pour plus d'informations, voir «Assistant d'échantillonnage : Méthode d'échantillonnage».

Entrer une pondération d'échantillon : Si le plan d'échantillonnage actuel fait partie d'un plan d'échantillon plus important, vous devez disposer des pondérations d'échantillon du plan plus important (obtenues lors d'une phase précédente). Vous pouvez spécifier une variable numérique contenant ces pondérations dans la première phase du plan actuel. Les pondérations d'échantillon sont automatiquement calculées pour les étapes suivantes du plan actuel.

Libellé d'étape : Vous pouvez spécifier un libellé de type chaîne de caractères pour chaque étape. Elle est utilisée dans la sortie pour identifier les informations de chaque étape.

Remarque : La liste des variables source compile le contenu obtenu lors de l'exécution des étapes de l'assistant. En d'autres termes, les variables supprimées de la liste source à une étape sont supprimées de toutes les étapes dans la liste. Les variables renvoyées à la liste source apparaissent dans la liste de toutes les étapes.

Contrôles d'arbre de navigation de l'assistant d'échantillonnage

Chaque étape de l'assistant d'échantillonnage dispose à sa gauche d'une légende. Vous pouvez parcourir l'assistant en cliquant sur le nom d'une étape disponible dans la légende. Les étapes sont disponibles tant que les étapes précédentes sont valides. Cela signifie qu'elles le sont si chaque étape précédente a fourni les spécifications minimales requises pour cette étape. Pour plus d'informations sur les raisons de l'invalidité d'une étape, reportez-vous à l'aide de chaque étape.

Assistant d'échantillonnage : Méthode d'échantillonnage

Cette étape vous permet de spécifier le mode de sélection des observations depuis le jeu de données actif.

Méthode : Les contrôles de ce groupe sont utilisées pour choisir une méthode de sélection. Certains types d'échantillonnage vous permettent de choisir entre un échantillonnage avec remplacement (WR) ou sans remplacement (WOR). Pour plus d'informations, reportez-vous aux descriptions des types d'échantillonnage. Certains types de probabilité proportionnelle à la taille (PPS) ne sont disponibles que lorsque les clusters ont été définies. Tous les types PPS ne sont disponibles que dans la première phase d'un plan. En outre, les méthodes WR ne sont disponibles que dans l'étape finale d'un plan.

- **Echantillonnage aléatoire simple :** Les unités sont sélectionnées à probabilité égale. Elles peuvent être sélectionnées avec ou sans remplacement.
- **Systématique simple :** Les unités sont sélectionnées à intervalles fixes dans le cadre d'échantillonnage (ou les strates, si celles-ci ont été spécifiées) et extraites sans remplacement. Une unité sélectionnée aléatoirement dans le premier intervalle est définie comme point de départ.
- **Séquentiel simple :** Les unités sont sélectionnées séquentiellement à probabilité égale et sans remplacement.
- **PPS :** Méthode de première phase qui effectue une sélection aléatoire avec probabilité proportionnelle à la taille. N'importe quelle unité peut être sélectionnée avec remplacement ; seules les clusters peuvent être échantillonnées sans remplacement.
- **Systématique avec PPS :** Méthode de première phase qui effectue une sélection systématique des unités avec probabilité proportionnelle à la taille. Elles sont sélectionnées sans remplacement.
- **Séquentiel avec PPS :** Méthode de première phase qui effectue une sélection séquentielle des unités avec probabilité proportionnelle à la taille du cluster et sans remplacement.
- **Brewer avec PPS :** Méthode de première phase qui sélectionne deux clusters dans chaque strate avec probabilité proportionnelle à la taille du cluster et sans remplacement. Une variable de cluster doit être spécifiée pour que cette méthode puisse être utilisée.

- **PPS Murthy** : Méthode de première phase qui sélectionne deux clusters dans chaque strate avec probabilité proportionnelle à la taille du cluster et sans remplacement. Une variable de cluster doit être spécifiée pour que cette méthode puisse être utilisée.
- **Sampford avec PPS** : Méthode de première phase qui sélectionne plus de deux clusters dans chaque strate avec probabilité proportionnelle à la taille du cluster et sans remplacement. Il s'agit d'une extension de la méthode Brewer. Une variable de cluster doit être spécifiée pour que cette méthode puisse être utilisée.
- **Utiliser une estimation WR pour l'analyse** : Par défaut, une méthode d'estimation est spécifiée dans le fichier de plan. Celle-ci est adaptée à la méthode d'échantillonnage sélectionnée. Cela vous permet d'utiliser une estimation avec remplacement, même si la méthode d'échantillonnage nécessite une estimation sans remplacement. Cette option n'est disponible que dans la phase 1.

Mesure de la taille (MOS) : Si une méthode PPS est sélectionnée, vous devez spécifier une mesure de taille permettant de définir la taille de chaque unité. Ces tailles peuvent être définies dans une variable, ou calculées à partir des données. Vous pouvez éventuellement définir des limites inférieure et supérieure dans la mesure de taille, qui ignorent les valeurs trouvées dans la variable de mesure de taille ou calculées à partir des données. Ces options ne sont disponibles que dans la phase 1.

Assistant d'échantillonnage : Taille de l'échantillon

Cette phase vous permet de spécifier le nombre ou la proportion d'unités à échantillonner dans la phase en cours. La taille de l'échantillon peut être fixe ou varier en fonction des strates. Pour spécifier la taille de l'échantillon, les clusters choisies dans les phases précédentes peuvent être utilisées pour définir les strates.

Unités : Vous pouvez spécifier une taille d'échantillon exacte ou une proportion d'unités à échantillonner.

- **Valeur** : Une valeur unique est appliquée à toutes les strates. Si vous sélectionnez **Effectifs** comme unité de mesure, vous devez saisir un entier positif. Si vous sélectionnez **Proportions**, vous devez saisir une valeur non négative. A moins qu'il ne s'agisse d'un échantillonnage avec remplacement, les valeurs de proportion ne doivent également pas être supérieures à 1.
- **Valeurs inégales pour les strates** : Vous permet de saisir les tailles pour chaque strate via la boîte de dialogue Définition de tailles inégales.
- **Lire les valeurs à partir de la variable** : Vous permet de sélectionner une variable numérique contenant les valeurs de taille des strates.

Si vous sélectionnez **Proportions**, vous pouvez éventuellement définir des limites inférieure et supérieure pour le nombre d'unités échantillonnées.

Définition de tailles inégales

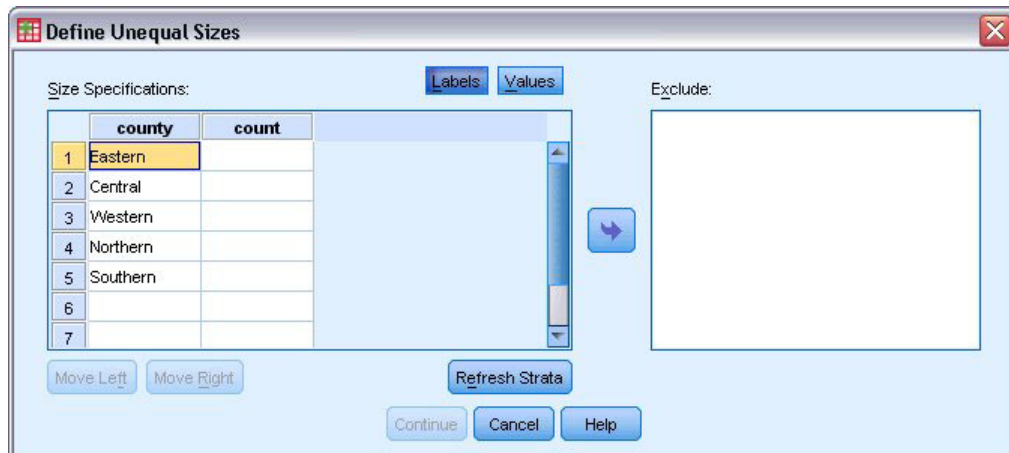


Figure 1. Boîte de dialogue Définition de tailles inégales

La boîte de dialogue Définition de tailles inégales vous permet de saisir des tailles par strate.

Grille des spécifications de taille : La grille affiche les classifications croisées de cinq variables ou de clusters maximum avec une combinaison strate/cluster par ligne. Les variables de grille sélectionnables incluent toutes les variables de stratification de la phase en cours et des phases précédentes, ainsi que les variables de clusters des phases précédentes. Les variables peuvent être réorganisées dans la grille ou déplacées dans la liste Exclure. Saisissez les tailles dans la colonne située à l'extrême droite. Cliquez sur **Libellés** ou sur **Valeurs** pour afficher les valeurs de libellés ou les valeurs de données des variables de stratification et de cluster dans les cellules de la grille. Les cellules contenant des valeurs non libellées affichent toujours des valeurs. Cliquez sur **Rafraîchir les strates** pour que la grille affiche chaque combinaison de valeurs de données libellées des variables dans la grille.

Exclure : Pour spécifier les tailles d'un sous-ensemble de combinaisons de strate/cluster, déplacez une ou plusieurs valeurs dans la liste Exclure. Ces variables ne sont pas utilisées pour définir les tailles des échantillons.

Assistant d'échantillonnage : Variables destination

Cette étape vous permet de choisir des variables à enregistrer lors de la réalisation de l'échantillon.

Taille de la population : Nombre d'unités estimées dans la population d'une phase spécifique. Le nom de racine de la variable enregistrée est *PopulationSize_*.

Proportion d'échantillon : Fréquence d'échantillonnage à une phase spécifique. Le nom de racine de la variable enregistrée est *SamplingRate_*.

Taille de l'échantillon : Nombre d'unités réalisées à une phase spécifique. Le nom de racine de la variable enregistrée est *SampleSize_*.

Pondération d'échantillon : L'inverse des probabilités d'inclusion. Le nom de racine de la variable enregistrée est *SampleWeight_*.

Certaines variables définies au cours des différentes étapes sont générées automatiquement. Parmi celles-ci, notons :

Probabilités d'inclusion : Proportion des unités réalisées à une phase spécifique. Le nom de racine de la variable enregistrée est *InclusionProbability_*.

Pondération cumulée : Pondération d'échantillon cumulée entre les phases précédentes et la phase actuelle incluse. Le nom de racine de la variable enregistrée est *SampleWeightCumulative_*.

Index : Identifie les unités sélectionnées plusieurs fois dans une phase spécifique. Le nom de racine de la variable enregistrée est *Index_*.

Remarque : Les noms de racine des variables enregistrées contiennent un suffixe entier faisant référence au numéro d'étape, par exemple *PopulationSize_1_* est la taille de population enregistrée de l'étape 1.

Assistant d'échantillonnage : Récapitulatif du plan

Il s'agit de la dernière étape de chaque phase ; elle fournit un récapitulatif des spécifications de plan d'échantillonnage dans la phase actuelle. Vous pouvez ensuite soit passer à la phase suivante (en la créant, si nécessaire), soit définir les options de réalisation de l'échantillon.

Assistant d'échantillonnage - Réaliser un échantillon : Options de sélection

Cette étape vous permet de choisir de réaliser un échantillon. Vous pouvez également contrôler d'autres options d'échantillonnage telles que la gestion de la valeur de départ aléatoire et des valeurs manquantes.

Réaliser un échantillon : Outre le fait de choisir de réaliser un échantillon, vous pouvez également choisir d'exécuter une partie du plan d'échantillonnage. Les étapes doivent être réalisées dans l'ordre, ce qui signifie que l'étape 2 ne peut pas être exécutée avant l'étape 1. Lors de la modification ou de l'exécution d'un plan, vous ne pouvez pas rééchantillonner les phases verrouillées.

Valeur de départ : Vous permet d'attribuer une valeur de départ utilisée pour les nombres aléatoires.

Inclure les valeurs manquantes de l'utilisateur : Permet de déterminer si les valeurs manquantes de l'utilisateur sont valides. Si c'est le cas, les valeurs manquantes de l'utilisateur sont traitées séparément.

Données déjà triées : Si votre cadre d'échantillon est prétrié par les valeurs des variables de stratification, cette option vous permet d'accélérer le processus de sélection.

Assistant d'échantillonnage - Réaliser un échantillon : Fichiers de sortie

Cette étape vous permet de choisir l'emplacement de redirection des observations échantillonnées et des variables de pondération, ainsi que des probabilités conjointes et des règles de sélection des observations.

Données exemple : Ces options vous permettent de déterminer l'emplacement d'enregistrement de l'échantillon de sortie. Celui-ci peut être ajouté à un nouveau jeu de données ou sauvegardé vers un fichier de données externe IBM® SPSS Statistics. Les jeux de données sont disponibles lors de la session en cours mais ne sont pas disponibles lors des sessions suivantes, sauf si vous les enregistrez clairement comme fichiers de données. Les noms des jeux de données doivent être conformes aux règles de dénomination de variables. Si vous spécifiez un fichier externe ou un nouveau jeu de données, les variables destination d'échantillonnage et les variables du jeu de données actif des observations sélectionnées sont enregistrées dans le fichier.

Probabilités conjointes : Ces options vous permettent de déterminer l'emplacement d'enregistrement des probabilités conjointes. Elles sont enregistrées dans un fichier de données externe IBM SPSS Statistics. Les

probabilités conjointes sont produites si vous spécifiez une méthode PPS WOR, PPS Brewer, PPS Sampford ou PSS Murthy et que vous ne définissez pas d'estimation avec remplacement.

Règles de sélection d'observation : Si vous réalisez votre échantillon phase par phase, vous souhaitez peut-être enregistrer les règles de sélection des observations dans un fichier texte. Elles sont utiles pour réaliser le sous-cadre des phases suivantes.

Assistant d'échantillonnage : Fin

Il s'agit de l'étape finale. Vous pouvez enregistrer le fichier de plan et réaliser l'échantillon maintenant, ou coller vos sélections dans une fenêtre de syntaxe.

Lorsque vous apportez des modifications aux phases du fichier de plan existant, vous pouvez enregistrer le plan modifié dans un nouveau fichier ou remplacer le fichier existant. Lorsque vous ajoutez des phases sans modifier les phases existantes, l'assistant remplace automatiquement le fichier de plan existant. Pour enregistrer le plan dans un nouveau fichier, sélectionnez **Coller la syntaxe générée par l'assistant dans une fenêtre de syntaxe** et modifiez le nom du fichier dans les commandes de la syntaxe.

Modification d'un plan d'échantillonnage existant

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Sélectionner un échantillon...
2. Sélectionnez **Modifier un plan d'échantillonnage** et choisissez le fichier de plan à modifier.
3. Cliquez sur **Suivant** pour poursuivre la procédure avec l'assistant.
4. Révisez le plan d'échantillonnage dans l'étape Récapitulatif du plan, puis cliquez sur **Suivant**.
Les étapes suivantes sont pratiquement les mêmes que celles effectuées pour un nouveau plan. Pour plus d'informations, reportez-vous à l'aide de chaque étape.
5. Allez jusqu'à l'étape Fin et spécifiez un nouveau nom pour le fichier de plan modifié, ou choisissez d'écraser le fichier de plan existant.

Sinon, vous pouvez :

- Spécifiez les phases qui ont déjà été échantillonnées.
- Supprimez les étapes du plan.

Assistant d'échantillonnage : Récapitulatif du plan

Cette phase vous permet de réviser le plan d'échantillonnage et d'indiquer les phases qui ont déjà été échantillonnées. Si vous modifiez un plan, vous pouvez également supprimer des phases du plan.

Étapes précédemment échantillonnées : Si un cadre d'échantillonnage étendu n'est pas disponible, vous devez exécuter un plan d'échantillonnage multiphase, phase par phase. Sélectionnez dans la liste déroulante les phases qui ont déjà été échantillonnées. Toutes les phases qui ont été exécutées sont verrouillées ; elles ne sont pas disponibles dans l'étape Réalisation de l'échantillon : Options de sélection et ne peuvent être interverties lors de la modification d'un plan.

Supprimer du plan les étapes : Vous pouvez supprimer les phases 2 et 3 d'un plan à plusieurs phases.

Exécution d'un plan d'échantillonnage existant

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Sélectionner un échantillon...
2. Sélectionnez **Réaliser un échantillon** et sélectionnez un fichier de plan à exécuter.
3. Cliquez sur **Suivant** pour poursuivre la procédure avec l'assistant.

4. Réviser le plan d'échantillonnage dans l'étape Récapitulatif du plan, puis cliquez sur **Suivant**.
5. Les étapes individuelles contenant des informations de phase sont ignorées lors de l'exécution d'un plan d'échantillonnage. Vous pouvez passer à l'étape Fin à tout moment.

Vous pouvez également spécifier les étapes qui ont déjà été échantillonnées.

Fonctions supplémentaires des commandes SPLAN et CSSELECT

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

- Spécifier les noms personnalisés des variables destination.
- Contrôler la sortie dans le visualiseur. Par exemple, vous pouvez supprimer le résumé de chaque phase du plan qui est affiché si un échantillon est réalisé ou modifié, supprimer le récapitulatif de la distribution des observations échantillonnées par strates qui apparaît si le plan d'échantillonnage est exécuté et demander un récapitulatif de traitement des observations.
- Choisir un sous-ensemble de variables dans le jeu de données actif pour effectuer l'enregistrement dans un fichier d'exemple externe ou dans un jeu de données différent.

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

Chapitre 3. Préparation d'un échantillon complexe en vue d'une analyse

L'assistant de préparation d'analyse vous guide dans la création ou la modification d'un plan d'analyse à utiliser avec les diverses procédures d'analyse des échantillons complexes. Avant d'utiliser l'assistant, vous devez réaliser un échantillon en fonction d'un plan spécifique.

La création d'un nouveau plan est principalement utile lorsque vous n'avez pas accès au fichier de plan d'échantillonnage utilisé pour réaliser l'échantillon (n'oubliez pas que le plan d'échantillonnage contient un plan d'analyse par défaut). Si vous avez accès au fichier de plan d'échantillonnage utilisé pour réaliser l'échantillon, vous pouvez utiliser le plan d'analyse par défaut contenu dans le fichier de plan d'échantillonnage ou ignorer les spécifications d'analyse par défaut et enregistrer vos modifications dans un nouveau fichier.

Création d'un plan d'analyse

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Préparer pour l'analyse...
2. Sélectionnez **Créer un fichier de plan** et attribuez un nom au fichier dans lequel enregistrer le plan d'analyse.
3. Cliquez sur **Suivant** pour poursuivre la procédure avec l'assistant.
4. A l'étape **Variables de plan**, spécifiez la variable contenant les pondérations d'échantillon et définissez éventuellement des strates et des clusters.
5. Vous pouvez alors cliquer sur **Fin** pour enregistrer le plan.

Les étapes suivantes sont facultatives. Elles permettent :

- De sélectionner la méthode d'estimation des erreurs standard à l'étape **Méthode d'estimation**.
- De spécifier le nombre d'unités échantillonnées ou la probabilité d'inclusion par unité à l'étape **Taille**.
- D'ajouter une deuxième ou une troisième étape au plan.
- De coller vos sélections en tant que syntaxe de commande.

Assistant de préparation d'analyse : Variables du plan

Cette étape vous permet d'identifier les variables pour la définition des couches et grappes, et pour celle des pondérations d'échantillon. Vous pouvez également associer un libellé à cette étape.

Strates : La classification croisée des variables de stratification définit des sous-populations distinctes, ou strates. La totalité de votre échantillon représente la combinaison des échantillons indépendants de chaque strate.

Clusters : Les variables de cluster définissent les groupes d'unités d'observation, ou clusters. Les échantillons réalisés en plusieurs étapes sélectionnent les clusters des étapes précédentes, puis sous-échantillonnent les unités à partir des clusters sélectionnées. Lors de l'analyse d'un fichier de données obtenu par échantillonnage en clusters avec remplacement, vous devez inclure l'index de duplication comme variable de cluster.

Pondération d'échantillon : Vous devez fournir des pondérations d'échantillon à la première étape. Les pondérations d'échantillon sont automatiquement calculées pour les étapes suivantes du plan actuel.

Libellé d'étape : Vous pouvez spécifier un libellé de type chaîne de caractères pour chaque étape. Elle est utilisée dans la sortie pour identifier les informations de chaque étape.

Remarque : La liste des variables source compile le contenu obtenu lors de l'exécution des étapes de l'assistant. En d'autres termes, les variables supprimées de la liste source à une étape sont supprimées de toutes les étapes dans la liste. Les variables renvoyées à la liste source apparaissent à toutes les étapes.

Contrôles d'arbre de navigation de l'assistant d'analyse

Chaque étape de l'assistant d'analyse dispose à sa gauche d'une légende. Vous pouvez parcourir l'assistant en cliquant sur le nom d'une étape disponible dans la légende. Les étapes sont disponibles tant que les étapes précédentes sont valides. Cela signifie qu'elles le sont si chaque étape précédente a fourni les spécifications minimales requises pour cette étape. Pour plus d'informations sur les raisons de l'invalidité d'une étape, reportez-vous à l'aide de chaque étape.

Assistant de préparation d'analyse : Méthode d'estimation

Cette étape vous permet d'indiquer une méthode d'estimation pour l'étape.

WR (échantillonnage avec remplacement) : L'estimation WR n'inclut pas de correction d'échantillonnage d'une population finie (FPC) lors de l'estimation de la variance dans le plan d'échantillonnage complexe. Vous pouvez choisir d'inclure ou d'exclure la correction FPC lors de l'estimation de la variance dans un échantillonnage aléatoire simple (SRS).

Il est recommandé de ne pas inclure la correction FPC pour l'estimation de la variance SRS lorsque les pondérations d'analyse ont été redimensionnées afin de ne pas s'ajouter à la taille de la population. L'estimation de la variance SRS est utilisée dans le calcul de statistiques telles que l'effet de plan. L'estimation WR ne peut être spécifiée que dans l'étape finale d'un plan. L'assistant ne vous laisse plus ajouter d'autre étape ensuite.

WOR égal (échantillonnage de probabilité égale sans remplacement) : L'estimation WOR égal inclut la correction de population finie et part du principe que les unités sont échantillonnées à probabilité égale. L'estimation WOR égal peut être spécifiée dans n'importe quelle étape d'un plan.

WOR inégal (échantillonnage de probabilité inégale sans remplacement) : Outre l'utilisation de la correction de population finie, l'estimation WOR inégal échantillonne les unités (généralement des clusters) sélectionnées à probabilité inégale. Cette méthode d'estimation n'est disponible qu'à la première étape.

Assistant de préparation d'analyse : Taille

Cette étape est utilisée pour spécifier les probabilités d'inclusion ou les tailles de population de l'étape en cours. Ces tailles peuvent être fixes ou varier suivant les strates. Les clusters spécifiées dans les étapes précédentes peuvent être utilisées pour définir les strates dans le cadre de la spécification des tailles. Notez que cette étape est nécessaire seulement si vous avez choisi la méthode d'estimation WOR égal.

Unités : Vous pouvez spécifier des tailles de population exactes ou les probabilités d'échantillonnage des unités.

- **Valeur** : Une valeur unique est appliquée à toutes les strates. Si vous définissez l'unité de mesure sur **Tailles de population**, vous devez saisir un entier non négatif. Si vous définissez l'unité de mesure sur **Probabilités d'inclusion**, vous devez saisir une valeur comprise entre 0 et 1 inclus.
- **Valeurs inégales pour les strates** : Vous permet de saisir les tailles pour chaque strate via la boîte de dialogue Définition de tailles inégales.
- **Lire les valeurs à partir de la variable** : Vous permet de sélectionner une variable numérique contenant les valeurs de taille des strates.

Définition de tailles inégales

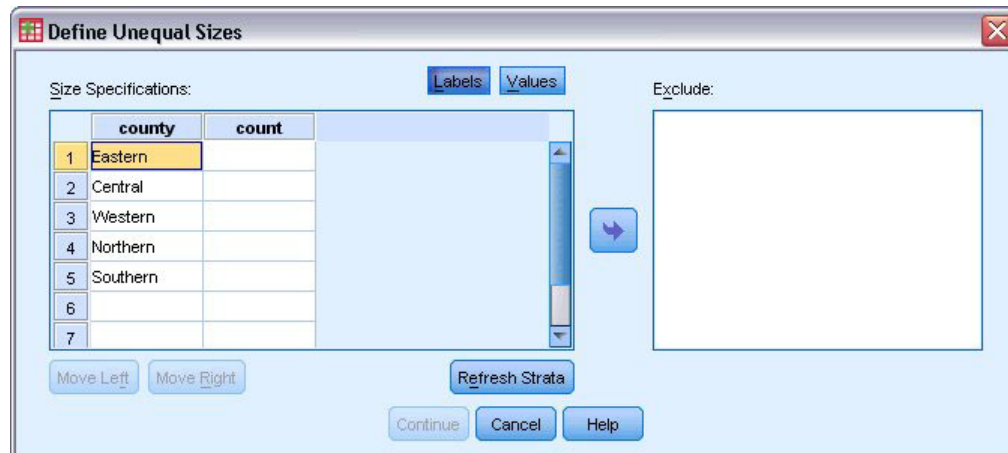


Figure 2. Boîte de dialogue Définition de tailles inégales

La boîte de dialogue Définition de tailles inégales vous permet de saisir des tailles par strate.

Grille des spécifications de taille : La grille affiche les classifications croisées de cinq variables ou de clusters maximum avec une combinaison strate/cluster par ligne. Les variables de grille sélectionnables incluent toutes les variables de stratification de la phase en cours et des phases précédentes, ainsi que les variables de clusters des phases précédentes. Les variables peuvent être réorganisées dans la grille ou déplacées dans la liste Exclure. Saisissez les tailles dans la colonne située à l'extrême droite. Cliquez sur **Libellés** ou sur **Valeurs** pour afficher les valeurs de libellés ou les valeurs de données des variables de stratification et de cluster dans les cellules de la grille. Les cellules contenant des valeurs non libellées affichent toujours des valeurs. Cliquez sur **Rafraîchir les strates** pour que la grille affiche chaque combinaison de valeurs de données libellées des variables dans la grille.

Exclure : Pour spécifier les tailles d'un sous-ensemble de combinaisons de strate/cluster, déplacez une ou plusieurs valeurs dans la liste Exclure. Ces variables ne sont pas utilisées pour définir les tailles des échantillons.

Assistant de préparation d'analyse : Récapitulatif du plan

Il s'agit de la dernière étape de chaque phase ; celle-ci fournit un récapitulatif des spécifications de plan d'analyse dans l'étape en cours. Vous pouvez ensuite soit passer à la phase suivante (en la créant, si nécessaire), soit enregistrer les spécifications d'analyse.

Les raisons pour lesquelles vous ne pouvez pas ajouter d'autre étape sont principalement les suivantes :

- A l'étape Variables de plan, aucune variable de cluster n'a été spécifiée.
- Vous avez sélectionné l'estimation WR à l'étape Méthode d'estimation.
- Il s'agit de la troisième étape de l'analyse et l'assistant n'en prend pas en charge davantage.

Assistant de préparation d'analyse : Fin

Il s'agit de l'étape finale. Vous pouvez enregistrer le fichier de plan maintenant ou collecter vos sélections dans une fenêtre de syntaxe.

Lorsque vous apportez des modifications aux phases du fichier de plan existant, vous pouvez enregistrer le plan modifié dans un nouveau fichier ou remplacer le fichier existant. Lorsque vous ajoutez des phases sans modifier les phases existantes, l'assistant remplace automatiquement le fichier de plan existant. Pour

enregistrer le plan dans un nouveau fichier, sélectionnez **Coller la syntaxe générée par l'assistant dans une fenêtre de syntaxe** et modifiez le nom du fichier dans les commandes de la syntaxe.

Modification d'un plan d'analyse existant

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Préparer pour l'analyse...
2. Sélectionnez **Modifier un fichier de plan** et attribuez un nom au fichier de plan dans lequel enregistrer le plan d'analyse.
3. Cliquez sur **Suivant** pour poursuivre la procédure avec l'assistant.
4. Révisez le plan d'analyse dans l'étape Récapitulatif du plan, puis cliquez sur **Suivant**.
Les étapes suivantes sont pratiquement les mêmes que celles effectuées pour un nouveau plan. Pour plus d'informations, reportez-vous à l'aide de chaque étape.
5. Allez jusqu'à l'étape Fin et spécifiez un nouveau nom pour le fichier de plan modifié ou choisissez d'écraser le fichier de plan existant.

Vous pouvez éventuellement supprimer des étapes du plan.

Assistant de préparation d'analyse : Récapitulatif du plan

Cette étape vous permet de passer en revue le plan d'analyse et de supprimer des phases du plan.

Supprimer du plan les étapes : Vous pouvez supprimer les phases 2 et 3 d'un plan à plusieurs phases. Comme un plan nécessite au moins une phase, vous pouvez modifier la phase 1 du plan, mais pas la supprimer.

Chapitre 4. Plan d'échantillonnages complexes

Les procédures d'analyse d'échantillons complexes nécessitent des spécifications d'analyse depuis un fichier de plan d'analyse ou d'échantillon, afin de disposer de résultats valides.

Plan : Spécifiez le chemin d'un fichier de plan d'analyse ou d'échantillon.

Probabilités conjointes : Pour utiliser l'estimation WOR inégal avec les clusters créées à l'aide d'une méthode PPS WOR, vous devez spécifier un fichier distinct ou un jeu de données ouvert contenant les probabilités conjointes. Le fichier ou le jeu de données est créé par l'assistant d'échantillonnage lors de l'échantillonnage.

Chapitre 5. Echantillons complexes - Effectifs

La procédure Echantillons complexes - Fréquences génère les tables de fréquences des variables sélectionnées et affiche des statistiques univariées. Vous pouvez éventuellement classer les statistiques par sous-groupes, définis par une ou plusieurs variables catégorielles.

Exemple : La procédure Echantillons complexes - Fréquences vous permet d'obtenir les statistiques tabulaires univariées de la consommation de vitamines des Américains, basées sur les résultats du NHIS (National Health Interview Survey) et dotées d'un plan d'analyse approprié pour ces données d'usage public.

Statistiques : La procédure génère pour chaque estimation les estimations des tailles de population et les pourcentages de tableaux, ainsi que les erreurs standard, les intervalles de confiance, les coefficients de variation, les effets de plan, les racines carrées d'effets de plan, les valeurs cumulées et les effectifs non pondérés. En outre, les statistiques khi-deux et de rapport de vraisemblance sont calculées pour le test d'uniformité des proportions des cellules.

Remarques concernant les données de fréquence des échantillons complexes

Données : Les variables pour lesquelles les tables de fréquences sont produites doivent être catégorielles. Les variables de sous-population peuvent être du type chaîne de caractères ou numérique, mais ne doivent pas être catégorielles.

Hypothèses : Les observations dans le fichier de données représentent un échantillon provenant d'un plan complexe et devant être analysées en fonction des spécifications du fichier sélectionné dans la Boîte de dialogue Plan d'échantillonnages complexes.

Obtention de fréquences d'échantillons complexes

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Effectifs...
2. Sélectionnez un fichier de plan. Vous pouvez également sélectionner un fichier personnalisé de probabilités conjointes.
3. Cliquez sur **Poursuivre**.
4. Dans la boîte de dialogue Echantillons complexes - Effectifs, sélectionnez au moins une variable d'effectif.

Vous pouvez éventuellement spécifier les variables définissant les sous-populations. Les statistiques sont calculées séparément pour chaque sous-population.

Statistiques de fréquences des échantillons complexes

Cellules : Ce groupe vous permet de demander les estimations des tailles de population de cellule et les pourcentages en tableau.

Statistiques : Ce groupe génère des statistiques associées à la taille de la population ou au pourcentage en tableau.

- **Erreur standard :** Erreur standard de l'estimation.
- **Intervalle de confiance :** Intervalle de confiance de l'estimation utilisant le niveau spécifié.
- **Coefficient de variation :** Rapport entre l'erreur standard de l'estimation et l'estimation.
- **Effectif non pondéré :** Nombre d'unités utilisées pour calculer l'estimation.

- **Effet de plan** : Rapport entre la variance de l'estimation et la variance, en partant du principe que l'échantillon est un échantillon aléatoire simple. Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.
- **Racine carrée de l'effet de plan** : Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.
- **Valeurs cumulées** : Estimation cumulée pour chaque valeur de la variable.

Test d'uniformité des proportions des cellules : Ce test génère des tests du khi-deux et de rapport de vraisemblance basés sur l'hypothèse que les catégories d'une variable possèdent des fréquences égales. Des tests distincts sont effectués pour chaque variable.

Valeurs manquantes d'échantillons complexes

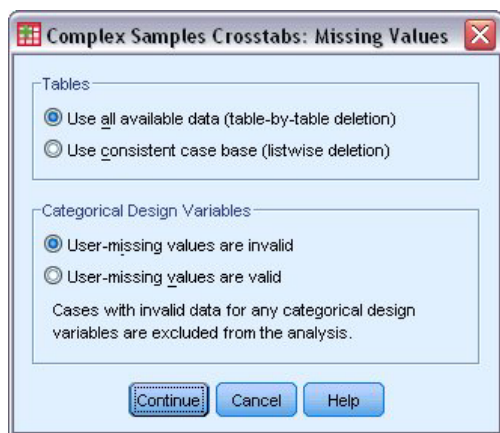


Figure 3. Boîte de dialogue Valeurs manquantes

Tableaux : Ce groupe détermine les observations utilisées dans l'analyse.

- **Utiliser toutes les données disponibles** : Les valeurs manquantes sont définies tableau par tableau. Ainsi, les observations utilisées pour calculer les statistiques peuvent varier selon la fréquence ou les tableaux croisés.
- **Utiliser une base d'observations cohérente** : Les valeurs manquantes sont déterminées dans toutes les variables. Par conséquent, les observations utilisées pour calculer les statistiques sont cohérentes dans tous les tableaux.

Variables de plan qualitatives : Ce groupe détermine si les valeurs manquantes de l'utilisateur sont valides ou non.

Options d'échantillons complexes

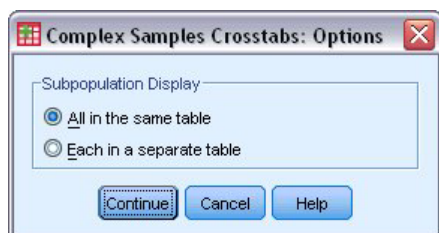


Figure 4. Boîte de dialogue Options

Affichage des sous-populations : Vous pouvez choisir d'afficher les sous-populations dans le même tableau ou dans des tableaux distincts.

Chapitre 6. Echantillons complexes - Caractéristiques

La procédure Echantillons complexes – Descriptives affiche les statistiques récapitulatives univariées de plusieurs variables. Vous pouvez éventuellement classer les statistiques par sous-groupes, définis par une ou plusieurs variables catégorielles.

Exemple : La procédure Echantillons complexes – Caractéristiques vous permet d'obtenir les statistiques descriptives univariées des niveaux d'activité des Américains, basées sur les résultats du NHIS (National Health Interview Survey) et dotées d'un plan d'analyse approprié à ces données d'usage public.

Statistiques : La procédure calcule pour chaque estimation la moyenne et la somme, ainsi que des tests t , les erreurs standard, les intervalles de confiance, les coefficients de variation, les effectifs non pondérés, les tailles de population, les effets de plan et les racines carrées d'effets de plan.

Remarques concernant les données descriptives des échantillons complexes

Données : Les mesures doivent être des variables d'échelle. Les variables de sous-population peuvent être du type chaîne de caractères ou numérique, mais ne doivent pas être catégorielles.

Hypothèses : Les observations dans le fichier de données représentent un échantillon provenant d'un plan complexe et devant être analysées en fonction des spécifications du fichier sélectionné dans la Boîte de dialogue Plan d'échantillonnages complexes.

Obtention de caractéristiques des échantillons complexes

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Descriptives...
2. Sélectionnez un fichier de plan. Vous pouvez également sélectionner un fichier personnalisé de probabilités conjointes.
3. Cliquez sur **Poursuivre**.
4. Dans la boîte de dialogue Echantillons complexes - Caractéristiques, sélectionnez au moins une variable de mesure.

Vous pouvez éventuellement spécifier les variables définissant les sous-populations. Les statistiques sont calculées séparément pour chaque sous-population.

Statistiques des descriptifs des échantillons complexes

Principales statistiques : Ce groupe vous permet de demander les estimations des moyennes et des sommes des variables de mesure. En outre, vous pouvez effectuer des tests t sur les estimations en vous basant sur une valeur spécifiée.

Statistiques : Ce groupe génère des statistiques associées à la moyenne ou à la somme.

- **Erreur standard :** Erreur standard de l'estimation.
- **Intervalle de confiance :** Intervalle de confiance de l'estimation utilisant le niveau spécifié.
- **Coefficient de variation :** Rapport entre l'erreur standard de l'estimation et l'estimation.
- **Effectif non pondéré :** Nombre d'unités utilisées pour calculer l'estimation.
- **Taille de la population :** Nombre d'unités estimées composant la population.
- **Effet de plan :** Rapport entre la variance de l'estimation et la variance, en partant du principe que l'échantillon est un échantillon aléatoire simple. Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.

- **Racine carrée de l'effet de plan** : Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.

Valeurs manquantes des descriptifs d'échantillons complexes

Statistiques des variables de mesure : Ce groupe détermine les observations utilisées dans l'analyse.

- **Utilisez toutes les données disponibles** : Les valeurs manquantes sont déterminées variable par variable ; par conséquent, les observations utilisées pour calculer les statistiques peuvent varier suivant les variables de mesure.
- **Assurez-vous de disposer d'une base d'observations cohérente** : Les valeurs manquantes sont déterminées dans toutes les variables ; par conséquent, les observations utilisées pour calculer les statistiques sont cohérentes.

Variables de plan qualitatives : Ce groupe détermine si les valeurs manquantes de l'utilisateur sont valides ou non.

Options d'échantillons complexes

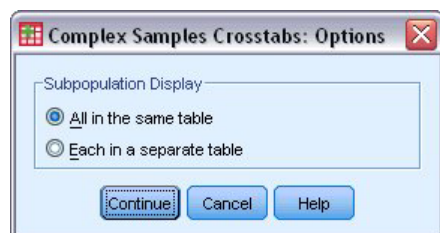


Figure 5. Boîte de dialogue Options

Affichage des sous-populations : Vous pouvez choisir d'afficher les sous-populations dans le même tableau ou dans des tableaux distincts.

Chapitre 7. Tableaux croisés des échantillons complexes

La procédure Echantillons complexes - Tableaux croisés génère les tableaux croisés des paires de variables sélectionnées et affiche des statistiques à deux entrées. Vous pouvez éventuellement classer les statistiques par sous-groupes, définis par une ou plusieurs variables catégorielles.

Exemple : La procédure Echantillons complexes - Tableaux croisés vous permet d'obtenir les statistiques de classification croisée fréquence consommation de cigarette/consommation de vitamines des Américains, basées sur les résultats du NHIS (National Health Interview Survey) et dotées d'un plan d'analyse pour ces données d'usage public.

Statistiques : La procédure génère pour chaque estimation des tailles de population, des pourcentages en lignes, colonnes et tableaux, des erreurs standard, des intervalles de confiance, des coefficients de variation, des valeurs théoriques, des effets de plan, des racines carrées d'effets de plan, des résidus, des résidus ajustés et des effectifs non pondérés. Le rapport des cotes, le risque relatif et la différence de risque sont calculés pour les tableaux 2*2. En outre, les statistiques de Pearson et de rapport de vraisemblance sont calculées pour le test d'indépendance des variables de ligne et de colonne.

Remarques concernant les données de la procédure Echantillons complexes - Tableaux croisés

Données : Les variables de ligne et de colonne doivent être catégorielles. Les variables de sous-population peuvent être du type chaîne de caractères ou numérique, mais ne doivent pas être catégorielles.

Hypothèses : Les observations dans le fichier de données représentent un échantillon provenant d'un plan complexe et devant être analysées en fonction des spécifications du fichier sélectionné dans la Boîte de dialogue Plan d'échantillonnages complexes.

Obtention de tableaux croisés des échantillons complexes

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Tableaux croisés...
2. Sélectionnez un fichier de plan. Vous pouvez également sélectionner un fichier personnalisé de probabilités conjointes.
3. Cliquez sur **Poursuivre**.
4. Dans la boîte de dialogue Tableaux croisés des échantillons complexes, sélectionnez au moins une variable de ligne et une variable de colonne.

Vous pouvez éventuellement spécifier les variables définissant les sous-populations. Les statistiques sont calculées séparément pour chaque sous-population.

Statistiques de tableaux croisés d'échantillons complexes

Cellules : Ce groupe vous permet de demander les estimations des tailles de population de cellule et les pourcentages de ligne, de colonne et de tableau.

Statistiques : Ce groupe génère des statistiques associées à la taille de la population et au pourcentage de ligne, de colonne et de tableau.

- **Erreur standard :** Erreur standard de l'estimation.
- **Intervalle de confiance :** Intervalle de confiance de l'estimation utilisant le niveau spécifié.
- **Coefficient de variation :** Rapport entre l'erreur standard de l'estimation et l'estimation.

- **Effectifs théoriques** : Valeur théorique de l'estimation, sous l'hypothèse de l'indépendance de la variable en ligne et en colonne.
- **Effectif non pondéré** : Nombre d'unités utilisées pour calculer l'estimation.
- **Effet de plan** : Rapport entre la variance de l'estimation et la variance, en partant du principe que l'échantillon est un échantillon aléatoire simple. Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.
- **Racine carrée de l'effet de plan** : Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.
- **Résidus** : La valeur théorique correspond au nombre d'observations attendues dans la cellule quand il n'existe pas de relation entre les deux variables. Un résidu positif indique que la cellule contient plus d'observations que si les variables de ligne et de colonne étaient indépendantes.
- **Résidus ajustés** : Résidu d'une cellule (valeur observée moins valeur théorique) divisé par une estimation de son erreur standard. Le résidu standardisé qui en résulte est exprimé en écarts par rapport à la moyenne.

Récapitulatifs pour les tableaux 2*2 : Ce groupe génère les statistiques des tableaux dans lesquels les variables de ligne et de colonne possèdent chacune 2 catégories. Chaque groupe est une mesure de la force de l'association entre la présence d'un facteur et la réalisation d'un événement.

- **Rapport des cotes** : Peut être utilisé comme estimation du risque relatif dans le cas où la réalisation du facteur est rare.
- **Risque relatif** : Rapport entre le risque que survienne un événement en présence du facteur et le risque que survienne l'événement en l'absence du facteur.
- **Différence de risque** : Différence entre le risque que survienne un événement en présence du facteur et le risque que survienne l'événement en l'absence du facteur.

Test d'indépendance des lignes et des colonnes : Ce test génère des tests du khi-deux et de rapport de vraisemblance sur l'hypothèse de l'indépendance entre une ligne et une colonne. Des tests distincts sont effectués pour chaque paire de variables.

Valeurs manquantes d'échantillons complexes

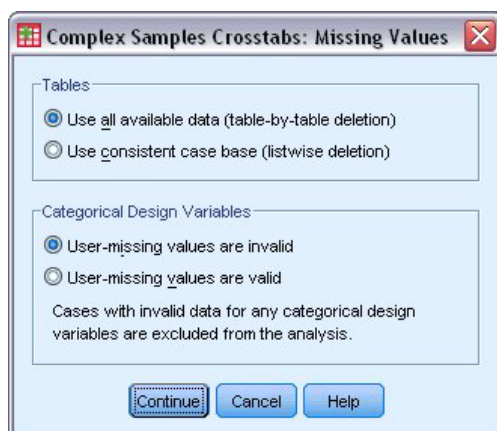


Figure 6. Boîte de dialogue Valeurs manquantes

Tableaux : Ce groupe détermine les observations utilisées dans l'analyse.

- **Utiliser toutes les données disponibles** : Les valeurs manquantes sont définies tableau par tableau. Ainsi, les observations utilisées pour calculer les statistiques peuvent varier selon la fréquence ou les tableaux croisés.

- **Utiliser une base d'observations cohérente** : Les valeurs manquantes sont déterminées dans toutes les variables. Par conséquent, les observations utilisées pour calculer les statistiques sont cohérentes dans tous les tableaux.

Variables de plan qualitatives : Ce groupe détermine si les valeurs manquantes de l'utilisateur sont valides ou non.

Options d'échantillons complexes

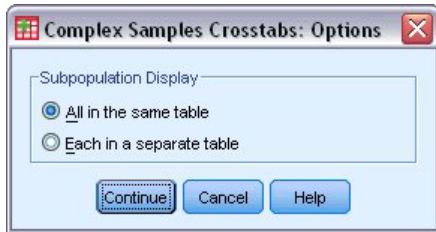


Figure 7. Boîte de dialogue Options

Affichage des sous-populations : Vous pouvez choisir d'afficher les sous-populations dans le même tableau ou dans des tableaux distincts.

Chapitre 8. Echantillons complexes - Rapports

La procédure Echantillons complexes – Rapports affiche les statistiques récapitulatives univariées des rapports de variables. Vous pouvez éventuellement classer les statistiques par sous-groupes, définis par une ou plusieurs variables catégorielles.

Exemple : La procédure Echantillons complexes – Rapports vous permet d'obtenir les statistiques descriptives du rapport entre la valeur de propriété en cours et la dernière valeur estimée, basée sur les résultats d'une enquête nationale menée en fonction d'un plan complexe et dotée d'un plan d'analyse approprié pour les données.

Statistiques : La procédure génère des estimations de rapport, des tests *t*, des erreurs standard, des intervalles de confiance, des coefficients de variation, des effectifs non pondérés, des tailles de population, des effets de plan et des racines carrées d'effets de plan.

Remarques concernant les données de rapports d'échantillons complexes

Données : Les numérateurs et les dénominateurs doivent être des variables d'échelle positives. Les variables de sous-population peuvent être du type chaîne de caractères ou numérique, mais ne doivent pas être catégorielles.

Hypothèses : Les observations dans le fichier de données représentent un échantillon provenant d'un plan complexe et devant être analysées en fonction des spécifications du fichier sélectionné dans la Boîte de dialogue Plan d'échantillonnages complexes.

Obtention de rapports d'échantillons complexes

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Ratios...
2. Sélectionnez un fichier de plan. Vous pouvez également sélectionner un fichier personnalisé de probabilités conjointes.
3. Cliquez sur **Poursuivre**.
4. Dans la boîte de dialogue Echantillons complexes - Rapports, sélectionnez au moins une variable de numérateur et une variable de dénominateur.

Vous pouvez également indiquer des variables pour définir les sous-groupes pour lesquels les statistiques sont produites.

Echantillons complexes – Rapports : Statistiques

Statistiques : Ce groupe produit les statistiques associées à l'estimation du rapport.

- **Erreur standard :** Erreur standard de l'estimation.
- **Intervalle de confiance :** Intervalle de confiance de l'estimation utilisant le niveau spécifié.
- **Coefficient de variation :** Rapport entre l'erreur standard de l'estimation et l'estimation.
- **Effectif non pondéré :** Nombre d'unités utilisées pour calculer l'estimation.
- **Taille de la population :** Nombre d'unités estimées composant la population.
- **Effet de plan :** Rapport entre la variance de l'estimation et la variance, en partant du principe que l'échantillon est un échantillon aléatoire simple. Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.
- **Racine carrée de l'effet de plan :** Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.

Test t : Vous pouvez demander d'effectuer des tests t sur les estimations basés sur une valeur spécifiée.

Echantillons complexes – Rapports : Valeurs manquantes

Ratios : Ce groupe détermine les observations utilisées dans l'analyse.

- **Utiliser toutes les données disponibles :** Les valeurs manquantes sont définies rapport par rapport. Ainsi, les observations utilisées pour calculer les statistiques peuvent varier d'une paire numérateur-dénominateur à l'autre.
- **Assurez-vous de disposer d'une base d'observations cohérente :** Les valeurs manquantes sont déterminées dans toutes les variables. Par conséquent, les observations utilisées pour le calcul des statistiques sont cohérentes.

Variables de plan qualitatives : Ce groupe détermine si les valeurs manquantes de l'utilisateur sont valides ou non.

Options d'échantillons complexes

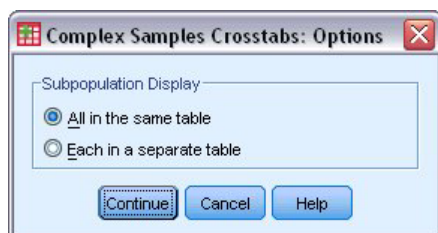


Figure 8. Boîte de dialogue Options

Affichage des sous-populations : Vous pouvez choisir d'afficher les sous-populations dans le même tableau ou dans des tableaux distincts.

Chapitre 9. Modèle linéaire général des échantillons complexes

La procédure relative au modèle linéaire général des échantillons complexes effectue une analyse de régression linéaire, ainsi qu'une analyse de variance et de la covariance, pour les échantillons réalisés à l'aide de méthodes d'échantillonnage complexes. Vous pouvez également demander une analyse pour une sous-population.

Exemple : Une chaîne d'épicerie a interrogé, en fonction d'un plan complexe, un groupe de clients au sujet de leurs habitudes de consommation. Compte tenu des résultats de l'enquête et de la somme dépensée par les clients au cours du mois précédent, l'enseigne souhaite voir si la fréquence des achats est liée aux dépenses mensuelles, et ce en prenant en compte le sexe du client et en intégrant le plan d'échantillonnage.

Statistiques : La procédure produit des estimations, des erreurs standard, des intervalles de confiance, des tests t , des effets de plan, des racines carrées d'effets de plan pour les paramètres du modèle ; elle fournit également les corrélations et les covariances des estimations des paramètres. Les mesures de qualité d'ajustement et les statistiques descriptives pour les variables dépendantes et indépendantes sont également disponibles. Vous pouvez également demander la moyenne marginale estimée pour les niveaux de facteurs de modèles et les interactions entre facteurs

Remarques sur les données du modèle linéaire général des échantillons complexes

Données : La variable dépendante est quantitative. Les facteurs sont qualitatifs. Les covariables sont des variables quantitatives liées à la variable dépendante. Les variables de sous-population peuvent être du type chaîne de caractères ou numérique, mais ne doivent pas être catégorielles.

Hypothèses : Les observations dans le fichier de données représentent un échantillon provenant d'un plan complexe et devant être analysées en fonction des spécifications du fichier sélectionné dans la Boîte de dialogue Plan d'échantillonnages complexes.

Obtention d'un modèle linéaire général des échantillons complexes

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Modèle linéaire général...
2. Sélectionnez un fichier de plan. Vous pouvez également sélectionner un fichier personnalisé de probabilités conjointes.
3. Cliquez sur **Poursuivre**.
4. Dans la boîte de dialogue Modèle linéaire général des échantillons complexes, sélectionnez une variable dépendante.

Sinon, vous pouvez :

- Sélectionner des variables pour les facteurs et covariables, en fonction de vos données.
- Indiquer une variable pour définir une sous-population. L'analyse est effectuée uniquement pour la catégorie sélectionnée de la variable de sous-population.

Modèle linéaire général des échantillons complexes

Spécifier les effets du modèle : Par défaut, la procédure élabore un modèle contenant des effets principaux à l'aide des covariables et facteurs spécifiés dans la boîte de dialogue principale. Vous pouvez également créer un modèle personnalisé contenant des effets d'interaction et des termes imbriqués.

Termes non imbriqués

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau pour 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.

Termes imbriqués

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

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

Limites : Les termes imbriqués comportent les restrictions suivantes :

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

Constante : La constante est généralement incluse dans le modèle. Si vous partez du principe que les données passent par l'origine, vous pouvez exclure la constante. Même si vous incluez la constante dans le modèle, vous pouvez supprimer les statistiques qui lui sont associées.

Modèle linéaire général des échantillons complexes - Statistiques

Paramètres du modèle : Ce groupe permet de contrôler l'affichage des statistiques associées aux paramètres du modèle.

- **Estimation :** Affiche des estimations des coefficients.
- **Erreur standard :** Affiche l'erreur standard pour chaque estimation de coefficient.
- **Intervalle de confiance :** Affiche l'intervalle de confiance pour chaque estimation de coefficient. Le niveau de confiance de l'intervalle est défini dans la boîte de dialogue Options.

- **Test t** : Affiche un test t pour chaque estimation de coefficient. L'hypothèse nulle de chaque test correspond au cas où la valeur du coefficient est 0.
- **Covariances des estimations de paramètres** : Affiche une estimation de la matrice de covariance pour les coefficients du modèle.
- **Corrélations des estimations de paramètres** : Affiche une estimation de la matrice de corrélation pour les coefficients du modèle.
- **Effet de plan** : Rapport entre la variance de l'estimation et la variance, en partant du principe que l'échantillon est un échantillon aléatoire simple. Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.
- **Racine carrée de l'effet de plan** : Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.

Qualité de l'ajustement : Affiche les statistiques de R^2 et d'erreur quadratique moyenne.

Moyennes des populations pour les variables et covariables dépendantes : Affiche les informations récapitulatives sur la variable dépendante, les covariables et les facteurs.

Informations sur le plan d'échantillonnage : Affiche les informations récapitulatives relatives à l'échantillon, y compris les effectifs non pondérés et la taille de la population.

Tests d'hypothèse des échantillons complexes

Statistique de test : Ce groupe vous permet de sélectionner le type de statistiques à utiliser pour tester les hypothèses. Vous avez le choix entre F , F ajusté, khi-deux et khi-deux ajusté.

Degrés de liberté de l'échantillonnage : Ce groupe permet de contrôler les degrés de liberté du plan d'échantillonnage utilisés pour calculer les valeurs p pour toutes les statistiques de test. Si elle est basée sur le plan d'échantillonnage, cette valeur correspond à la différence entre le nombre d'unités d'échantillonnage principales et le nombre de strates présentes à la première étape de l'échantillonnage. Vous pouvez également définir un degré de liberté personnalisé en indiquant un entier positif.

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.
- *Sidak séquentiel*. Il s'agit d'une procédure descendante de rejet séquentiel de Sidak beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Bonferroni séquentiel*. Il s'agit d'une procédure descendante de rejet séquentiel de Bonferroni beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Sidak*. Cette méthode fournit des limites plus strictes que l'approche de Bonferroni.
- *Bonferroni*. Cette méthode ajuste le niveau de signification observé au fait que plusieurs contrastes sont en cours de test.

Modèle linéaire général des échantillons complexes - Moyennes estimées

La boîte de dialogue Moyennes estimées permet d'afficher les moyennes marginales estimées du modèle pour les niveaux de facteurs et les interactions entre facteurs indiqués dans la sous-boîte de dialogue Modèle. Vous pouvez également demander l'affichage de la moyenne de population globale.

Terme : Les moyennes estimées sont calculées pour les interactions entre facteurs et les facteurs sélectionnés.

Contraste : Le contraste détermine le mode de définition des tests d'hypothèse pour la comparaison des moyennes estimées.

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

Catégorie de référence : Les contrastes simple et d'écart nécessitent une catégorie de référence ou un niveau de facteur servant de base de comparaison avec les autres.

Modèle linéaire général des échantillons complexes - Enregistrement

Enregistrer les variables : Ce groupe permet d'enregistrer les prévisions et les résidus du modèle en tant que nouvelles variables dans le fichier de travail.

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

- **rowtype_ :** Prend les valeurs (et libellés de valeurs) suivantes : COV (covariances), CORR (corrélations), EST (estimations des paramètres), SE (erreurs standard), SIG (niveau de signification) et DF (degrés de liberté du plan d'échantillonnage). Il existe une observation distincte avec le type de ligne COV (ou CORR) pour chaque paramètre de modèle et une observation distincte pour chacun des autres types de ligne.
- **varname_ :** Prend les valeurs P1, P2, etc., correspondant à une liste triée de tous les paramètres de modèle pour les types de ligne COV ou CORR, avec des libellés de valeur correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres. Les cellules sont vides pour les autres types de ligne.
- **P1, P2, .. :** Ces variables correspondent à une liste triée de tous les paramètres de modèle, avec des libellés de variable correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres, et prennent leurs 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.

Remarque : 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.

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.

Modèle linéaire général des échantillons complexes - Options

Valeurs manquantes de l'utilisateur : Toutes les variables de plan, ainsi que la variable dépendante et les covariables, doivent contenir des données valides. Les observations comportant des données non valides pour l'une de ces variables sont supprimées de l'analyse. Vous pouvez ainsi décider si les valeurs manquantes de l'utilisateur sont traitées comme étant valides dans la strate, le cluster, la sous-population et les facteurs.

Intervalle de confiance : Il s'agit du niveau d'intervalle de confiance pour les estimations de coefficient et les moyennes marginales estimées. Spécifiez une valeur supérieure ou égale à 50 et inférieure à 100.

Fonctions supplémentaires de la commande CSGLM

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 CUSTOM).
- 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 les mesures pour les contrastes polynomiaux (à l'aide de la sous-commande EMMEANS).
- Indiquer une valeur de tolérance pour le contrôle des singularités (à l'aide de la sous-commande CRITERIA).
- Créer des noms définis par l'utilisateur pour les variables enregistrées (à l'aide de la sous-commande SAVE).
- Générer un tableau des fonctions générales estimées (à l'aide de la sous-commande PRINT).

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

Chapitre 10. Régression logistique des échantillons complexes

La procédure de régression logistique des échantillons complexes effectue une analyse de régression logistique sur une variable dépendante binaire ou multinomiale, pour les échantillons réalisés à l'aide de méthodes d'échantillonnage complexes. Vous pouvez également demander une analyse pour une sous-population.

Exemple : Un responsable des prêts a recueilli l'historique des prêts octroyés aux clients à différents guichets, en fonction d'un plan complexe. Lors de l'intégration du plan d'échantillonnage, il souhaite voir si la probabilité de défaut de paiement est liée à l'âge, au parcours professionnel et au montant de la dette.

Statistiques : La procédure produit des estimations, des estimations exponentielles, des erreurs standard, des intervalles de confiance, des tests t , des effets de plan des racines carrées d'effets de plan pour les paramètres du modèle ; elle fournit également les corrélations et les covariances des estimations des paramètres. Les statistiques de pseudo R^2 , les tables de classification et les statistiques descriptives des variables dépendantes et indépendantes sont également disponibles.

Remarques sur les données de régression logistique des échantillons complexes

Données : La variable dépendante est qualitative. Les facteurs sont qualitatifs. Les covariables sont des variables quantitatives liées à la variable dépendante. Les variables de sous-population peuvent être du type chaîne de caractères ou numérique, mais ne doivent pas être catégorielles.

Hypothèses : Les observations dans le fichier de données représentent un échantillon provenant d'un plan complexe et devant être analysées en fonction des spécifications du fichier sélectionné dans la Boîte de dialogue Plan d'échantillonnages complexes.

Obtention de la régression logistique des échantillons complexes

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Régression logistique...
2. Sélectionnez un fichier de plan. Vous pouvez également sélectionner un fichier personnalisé de probabilités conjointes.
3. Cliquez sur **Poursuivre**.
4. Dans la boîte de dialogue Régression logistique des échantillons complexes, sélectionnez une variable dépendante.

Sinon, vous pouvez :

- Sélectionner des variables pour les facteurs et covariables, en fonction de vos données.
- Indiquer une variable pour définir une sous-population. L'analyse est effectuée uniquement pour la catégorie sélectionnée de la variable de sous-population.

Régression logistique des échantillons complexes - Catégorie de référence

Par défaut, la procédure de régression logistique des échantillons complexes fait de la catégorie ayant la valeur la plus élevée la catégorie de référence. Cette boîte de dialogue vous permet de spécifier la valeur la plus élevée, la plus basse ou une catégorie personnalisée en tant que catégorie de référence.

Modèle de régression logistique des échantillons complexes

Spécifier les effets du modèle : Par défaut, la procédure élabore un modèle contenant des effets principaux à l'aide des covariables et facteurs spécifiés dans la boîte de dialogue principale. Vous pouvez également créer un modèle personnalisé contenant des effets d'interaction et des termes imbriqués.

Termes non imbriqués

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau pour 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.

Termes imbriqués

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

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

Limites : Les termes imbriqués comportent les restrictions suivantes :

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

Constante : La constante est généralement incluse dans le modèle. Si vous partez du principe que les données passent par l'origine, vous pouvez exclure la constante. Même si vous incluez la constante dans le modèle, vous pouvez supprimer les statistiques qui lui sont associées.

Régression logistique des échantillons complexes - Statistiques

Qualité de l'ajustement : Contrôle l'affichage des statistiques qui mesurent les performances globales du modèle.

- **Pseudo R-deux :** La statistique R^2 de la régression linéaire n'a pas de véritable équivalent dans les modèles de régression logistique. On trouve, à la place, plusieurs mesures qui essaient de reproduire les propriétés de la statistique R^2 .
- **Table de classification :** Affiche les classifications croisées mises en tableau des catégories observées, classées selon les catégories prévues par le modèle en fonction de la variable dépendante.

Paramètres : Ce groupe permet de contrôler l'affichage des statistiques associées aux paramètres du modèle.

- **Estimation** : Affiche des estimations des coefficients.
- **Estimation mise en exposant** : Affiche la base du logarithme népérien élevée à la puissance des estimations des coefficients. L'estimation présente des propriétés parfaitement adaptées aux tests statistiques ; l'estimation exponentielle, ou $\exp(B)$, est, quant à elle, plus facile à interpréter.
- **Erreur standard** : Affiche l'erreur standard pour chaque estimation de coefficient.
- **Intervalle de confiance** : Affiche l'intervalle de confiance pour chaque estimation de coefficient. Le niveau de confiance de l'intervalle est défini dans la boîte de dialogue Options.
- **Test t** : Affiche un test t pour chaque estimation de coefficient. L'hypothèse nulle de chaque test correspond au cas où la valeur du coefficient est 0.
- **Covariances des estimations de paramètres** : Affiche une estimation de la matrice de covariance pour les coefficients du modèle.
- **Corrélations des estimations de paramètres** : Affiche une estimation de la matrice de corrélation pour les coefficients du modèle.
- **Effet de plan** : Rapport entre la variance de l'estimation et la variance, en partant du principe que l'échantillon est un échantillon aléatoire simple. Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.
- **Racine carrée de l'effet de plan** : Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.

Statistiques récapitulatives des variables de modèle : Affiche les informations récapitulatives sur la variable dépendante, les covariables et les facteurs.

Informations sur le plan d'échantillonnage : Affiche les informations récapitulatives relatives à l'échantillon, y compris les effectifs non pondérés et la taille de la population.

Tests d'hypothèse des échantillons complexes

Statistique de test : Ce groupe vous permet de sélectionner le type de statistiques à utiliser pour tester les hypothèses. Vous avez le choix entre F , F ajusté, khi-deux et khi-deux ajusté.

Degrés de liberté de l'échantillonnage : Ce groupe permet de contrôler les degrés de liberté du plan d'échantillonnage utilisés pour calculer les valeurs p pour toutes les statistiques de test. Si elle est basée sur le plan d'échantillonnage, cette valeur correspond à la différence entre le nombre d'unités d'échantillonnage principales et le nombre de strates présentes à la première étape de l'échantillonnage. Vous pouvez également définir un degré de liberté personnalisé en indiquant un entier positif.

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.
- *Sidak séquentiel*. Il s'agit d'une procédure descendante de rejet séquentiel de Sidak beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Bonferroni séquentiel*. Il s'agit d'une procédure descendante de rejet séquentiel de Bonferroni beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Sidak*. Cette méthode fournit des limites plus strictes que l'approche de Bonferroni.
- *Bonferroni*. Cette méthode ajuste le niveau de signification observé au fait que plusieurs contrastes sont en cours de test.

Régression logistique des échantillons complexes - Rapports des cotes

La boîte de dialogue Rapports des cotes permet d'afficher les rapports des cotes estimés par modèle pour les covariables et les facteurs indiqués. Un groupe de rapports des cotes distinct est calculé pour chaque catégorie de la variable dépendante, à l'exception de la catégorie de référence.

Facteurs : Pour chaque facteur sélectionné, affiche le rapport des cotes de chaque catégorie du facteur par rapport à celui de la catégorie de référence indiquée.

Covariables : Pour chaque covariable sélectionnée, affiche le rapport des cotes au niveau de la valeur moyenne de la covariable et des unités de modification indiquées, par rapport à celui de la moyenne.

Lors du calcul des rapports des cotes d'un facteur ou d'une covariable, la procédure paramètre tous les autres facteurs à leur niveau le plus haut et toutes les autres covariables à leur moyenne. Si un facteur ou une covariable interagit avec les autres prédicteurs du modèle, tous les rapports des cotes dépendent alors non seulement de la modification apportée à la variable indiquée, mais aussi des valeurs des variables avec lesquelles il interagit. Si une covariable spécifiée interagit avec elle-même dans le modèle (par exemple, $\text{âge}^*\text{âge}$), les rapports des cotes dépendent alors de la modification de la covariable et de la valeur covariable.

Régression logistique des échantillons complexes - Enregistrement

Enregistrer les variables : Ce groupe permet d'enregistrer les catégories estimées et les probabilités prévues du modèle en tant que nouvelles variables dans le jeu de données actif.

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

- **rowtype_ :** Prend les valeurs (et libellés de valeurs) suivantes : COV (covariances), CORR (corrélations), EST (estimations des paramètres), SE (erreurs standard), SIG (niveau de signification) et DF (degrés de liberté du plan d'échantillonnage). Il existe une observation distincte avec le type de ligne COV (ou CORR) pour chaque paramètre de modèle et une observation distincte pour chacun des autres types de ligne.
- **varname_ :** Prend les valeurs P1, P2, etc., correspondant à une liste triée de tous les paramètres de modèle pour les types de ligne COV ou CORR, avec des libellés de valeur correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres. Les cellules sont vides pour les autres types de ligne.
- **P1, P2, .. :** Ces variables correspondent à une liste triée de tous les paramètres de modèle, avec des libellés de variable correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres, et prennent leurs 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.

Remarque : 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.

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.

Régression logistique des échantillons complexes - Options

Estimation : Ce groupe permet de contrôler plusieurs critères utilisés dans l'estimation du modèle.

- **Maximum des itérations :** Nombre maximal d'itérations exécutées par l'algorithme. Spécifiez un nombre entier non négatif.
- **Nombre maximum de dichotomies :** A chaque itération, la taille de l'étape est réduite par un facteur de 0,5 jusqu'à ce que les augmentations de log de vraisemblance ou le nombre maximum de dichotomie soient atteints. Spécifiez un nombre entier positif.
- **Limiter les itérations basées sur les changements dans les estimations de 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 ne doit pas être négative.
- **Limiter les itérations basées sur les changements dans les rapports 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 ne doit pas être négative.
- **Vérifier la séparation complète 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.
- **Afficher l'historique des itérations :** Affiche les estimations de paramètre et les statistiques toutes les n itérations, en commençant par l'itération 0 (les estimations initiales). Si vous imprimez l'historique des itérations, la dernière itération est toujours imprimée, indépendamment de la valeur de n .

Valeurs manquantes de l'utilisateur : Toutes les variables de plan, ainsi que la variable dépendante et les covariables, doivent contenir des données valides. Les observations comportant des données non valides pour l'une de ces variables sont supprimées de l'analyse. Vous pouvez ainsi décider si les valeurs manquantes de l'utilisateur sont traitées comme étant valides dans la strate, le cluster, la sous-population et les facteurs.

Intervalle de confiance : Il s'agit du niveau d'intervalle de confiance pour les estimations de coefficient, les estimations de coefficient exponentielles et les rapports des cotes. Spécifiez une valeur supérieure ou égale à 50 et inférieure à 100.

Fonctions supplémentaires de la commande CSLOGISTIC

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 CUSTOM).
- Définir des valeurs pour les autres variables de modèle lors du calcul des rapports des cotes des facteurs et covariables (à l'aide de la sous-commande ODDS RATIOS).
- Indiquer une valeur de tolérance pour le contrôle des singularités (à l'aide de la sous-commande CRITERIA).
- Créer des noms définis par l'utilisateur pour les variables enregistrées (à l'aide de la sous-commande SAVE).
- Générer un tableau des fonctions générales estimées (à l'aide de la sous-commande PRINT).

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

Chapitre 11. Régression ordinale des échantillons complexes

La procédure de régression ordinale des échantillons complexes effectue une analyse de régression sur une variable dépendante binaire ou ordinale, pour les échantillons réalisés à l'aide de méthodes d'échantillonnage complexes. Vous pouvez également demander une analyse pour une sous-population.

Exemple : Des élus étudiant un projet de loi devant l'assemblée législative souhaitent savoir si ce projet est populaire auprès des électeurs et déterminer le lien existant entre cette popularité et la répartition démographique des électeurs. Les enquêteurs conçoivent et mènent des entretiens en fonction d'un plan d'échantillonnage complexe. Avec la procédure de régression ordinale des échantillons complexes, vous pouvez ajuster un modèle concernant la cote de popularité du projet de loi en fonction de la répartition démographique des électeurs.

Remarques sur les données de régression ordinale des échantillons complexes

Données : La variable dépendante est ordinale. Les facteurs sont qualitatifs. Les covariables sont des variables quantitatives liées à la variable dépendante. Les variables de sous-population peuvent être du type chaîne de caractères ou numérique, mais ne doivent pas être catégorielles.

Hypothèses : Les observations dans le fichier de données représentent un échantillon provenant d'un plan complexe et devant être analysées en fonction des spécifications du fichier sélectionné dans la Boîte de dialogue Plan d'échantillonnages complexes.

Obtention de la régression ordinale des échantillons complexes

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Régression ordinale...
2. Sélectionnez un fichier de plan. Vous pouvez également sélectionner un fichier personnalisé de probabilités conjointes.
3. Cliquez sur **Poursuivre**.
4. Dans la boîte de dialogue Régression ordinale des échantillons complexes, sélectionnez une variable dépendante.

Sinon, vous pouvez :

- Sélectionnez des variables pour les facteurs et covariables, en fonction de vos données.
- Indiquez une variable pour définir une sous-population. L'analyse n'est effectuée que pour la catégorie sélectionnée de la variable de sous-population, bien que les variances soient correctement estimées sur le jeu de données entier.
- Sélectionnez une fonction de lien.

Fonction de lien : La fonction de lien consiste en une transformation des probabilités cumulées permettant d'estimer le modèle. Les cinq fonctions de lien suivantes sont disponibles.

- **Logit :** $f(x)=\log(x/(1-x))$. Cette option est généralement utilisée pour les catégories distribuées de façon égale.
- **Log-log complémentaire :** $f(x)=\log(-\log(1-x))$. Cette option est généralement utilisée lorsque les catégories supérieures sont les plus probables.
- **Log-log négatif :** $f(x)=-\log(-\log(x))$. Cette option est généralement utilisée lorsque les catégories inférieures sont les plus probables.
- **Probit :** $f(x)=\Phi^{-1}(x)$. Cette option est généralement utilisée lorsque la variable de latence est normalement distribuée.

- **Cauchit (Cauchy inverse)** : $f(x)=\tan(\pi(x-0.5))$. Cette option est généralement utilisée lorsque la variable de latence possède un grand nombre de valeurs extrêmes.

Probabilités des réponses de régression ordinale des échantillons complexes

La boîte de dialogue Probabilités de réponses permet d'indiquer si la probabilité cumulée d'une réponse (à savoir la probabilité d'appartenir à une catégorie particulière de la variable dépendante et d'inclure cette catégorie) augmente avec les valeurs croissantes ou décroissantes de la variable dépendante.

Modèle de régression ordinale des échantillons complexes

Spécifier les effets du modèle : Par défaut, la procédure élabore un modèle contenant des effets principaux à l'aide des covariables et facteurs spécifiés dans la boîte de dialogue principale. Vous pouvez également créer un modèle personnalisé contenant des effets d'interaction et des termes imbriqués.

Termes non imbriqués

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau pour 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.

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.

Régression ordinale des échantillons complexes - Statistiques

Qualité de l'ajustement : Contrôle l'affichage des statistiques qui mesurent les performances globales du modèle.

- **Pseudo R-deux :** La statistique R^2 de la régression linéaire n'a pas de véritable équivalent dans les modèles de régression ordinale. On trouve, à la place, plusieurs mesures qui essaient de reproduire les propriétés de la statistique R^2 .
- **Table de classification :** Affiche les classifications croisées mises en tableau des catégories observées, classées selon les catégories prévues par le modèle en fonction de la variable dépendante.

Paramètres : Ce groupe permet de contrôler l'affichage des statistiques associées aux paramètres du modèle.

- **Estimation :** Affiche des estimations des coefficients.
- **Estimation mise en exposant :** Affiche la base du logarithme népérien élevée à la puissance des estimations des coefficients. L'estimation présente des propriétés parfaitement adaptées aux tests statistiques ; l'estimation exponentielle, ou $\exp(B)$, est, quant à elle, plus facile à interpréter.
- **Erreur standard :** Affiche l'erreur standard pour chaque estimation de coefficient.
- **Intervalle de confiance :** Affiche l'intervalle de confiance pour chaque estimation de coefficient. Le niveau de confiance de l'intervalle est défini dans la boîte de dialogue Options.
- **Test t :** Affiche un test t pour chaque estimation de coefficient. L'hypothèse nulle de chaque test correspond au cas où la valeur du coefficient est 0.
- **Covariances des estimations de paramètres :** Affiche une estimation de la matrice de covariance pour les coefficients du modèle.
- **Corrélations des estimations de paramètres :** Affiche une estimation de la matrice de corrélation pour les coefficients du modèle.
- **Effet de plan :** Rapport entre la variance de l'estimation et la variance, en partant du principe que l'échantillon est un échantillon aléatoire simple. Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.
- **Racine carrée de l'effet de plan :** Il s'agit d'une mesure, exprimée en unités comparables à celles de l'erreur standard, de l'effet de spécification d'un plan complexe où les valeurs éloignées de 1 indiquent des effets importants.

Droites parallèles : Ce groupe permet de demander des statistiques associées à un modèle avec courbes non parallèles où une courbe de régression distincte est ajustée pour chaque catégorie de réponse (sauf la dernière).

- **Test de Wald :** Produit un test de l'hypothèse nulle selon laquelle les paramètres de régression sont égaux pour toutes les réponses cumulées. Le modèle avec courbes non parallèles est estimé et le test de Wald des paramètres égaux est appliqué.
- **Estimations des paramètres :** Affiche les estimations des coefficients et des erreurs standard du modèle avec courbes non parallèles.
- **Covariances des estimations de paramètres :** Affiche une estimation de la matrice de covariance pour les coefficients du modèle avec courbes non parallèles.

Statistiques récapitulatives des variables de modèle : Affiche les informations récapitulatives sur la variable dépendante, les covariables et les facteurs.

Informations sur le plan d'échantillonnage : Affiche les informations récapitulatives relatives à l'échantillon, y compris les effectifs non pondérés et la taille de la population.

Tests d'hypothèse des échantillons complexes

Statistique de test : Ce groupe vous permet de sélectionner le type de statistiques à utiliser pour tester les hypothèses. Vous avez le choix entre F , F ajusté, khi-deux et khi-deux ajusté.

Degrés de liberté de l'échantillonnage : Ce groupe permet de contrôler les degrés de liberté du plan d'échantillonnage utilisés pour calculer les valeurs p pour toutes les statistiques de test. Si elle est basée sur le plan d'échantillonnage, cette valeur correspond à la différence entre le nombre d'unités d'échantillonnage principales et le nombre de strates présentes à la première étape de l'échantillonnage. Vous pouvez également définir un degré de liberté personnalisé en indiquant un entier positif.

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.
- *Sidak séquentiel.* Il s'agit d'une procédure descendante de rejet séquentiel de Sidak beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Bonferroni séquentiel.* Il s'agit d'une procédure descendante de rejet séquentiel de Bonferroni beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Sidak.* Cette méthode fournit des limites plus strictes que l'approche de Bonferroni.
- *Bonferroni.* Cette méthode ajuste le niveau de signification observé au fait que plusieurs contrastes sont en cours de test.

Régression ordinale des échantillons complexes - Rapports des cotes

La boîte de dialogue Rapports des cotes permet d'afficher les rapports des cotes cumulées estimés par modèle pour les covariables et les facteurs indiqués. Cette fonction n'est disponible que pour les modèles utilisant la fonction de lien logit. Un rapport des cotes cumulées unique est calculé pour toutes les catégories de la variable dépendante, à l'exception de la dernière. Le modèle d'odds proportionnel émet le postulat selon lequel toutes les catégories sont égales.

Facteurs : Pour chaque facteur sélectionné, affiche le rapport des cotes cumulées de chaque catégorie du facteur par rapport à celui de la catégorie de référence indiquée.

Covariables : Pour chaque covariable sélectionnée, affiche le rapport des cotes cumulées au niveau de la valeur moyenne de la covariable et des unités de modification indiquées, par rapport à celui de la moyenne.

Lors du calcul des rapports des cotes d'un facteur ou d'une covariable, la procédure paramètre tous les autres facteurs à leur niveau le plus haut et toutes les autres covariables à leur moyenne. Si un facteur ou une covariable interagit avec les autres prédicteurs du modèle, tous les rapports des cotes dépendent alors non seulement de la modification apportée à la variable indiquée, mais aussi des valeurs des variables avec lesquelles il interagit. Si une covariable spécifiée interagit avec elle-même dans le modèle (par exemple, $\text{âge}^*\text{âge}$), les rapports des cotes dépendent alors de la modification de la covariable et de la valeur covariable.

Régression ordinale des échantillons complexes - Enregistrement

Enregistrer les variables : Ce groupe permet d'enregistrer les probabilités prévues du modèle, la probabilité de catégorie prévue, la probabilité de catégorie observée, les probabilités cumulées et les probabilités prévues en tant que nouvelles variables dans le jeu de données actif.

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

- **rowtype_ :** Prend les valeurs (et libellés de valeurs) suivantes : COV (covariances), CORR (corrélations), EST (estimations des paramètres), SE (erreurs standard), SIG (niveau de signification) et DF (degrés de liberté du plan d'échantillonnage). Il existe une observation distincte avec le type de ligne COV (ou CORR) pour chaque paramètre de modèle et une observation distincte pour chacun des autres types de ligne.
- **varname_ :** Prend les valeurs P1, P2, etc., correspondant à une liste triée de tous les paramètres de modèle pour les types de ligne COV ou CORR, avec des libellés de valeur correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres. Les cellules sont vides pour les autres types de ligne.
- **P1, P2, .. :** Ces variables correspondent à une liste triée de tous les paramètres de modèle, avec des libellés de variable correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres, et prennent leurs 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.

Remarque : 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.

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.

Régression ordinale des échantillons complexes - Options

Méthode d'estimation : 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.

Estimation : Ce groupe permet de contrôler plusieurs critères utilisés dans l'estimation du modèle.

- **Maximum des itérations :** Nombre maximal d'itérations exécutées par l'algorithme. Spécifiez un nombre entier non négatif.
- **Nombre maximum de dichotomies :** A chaque itération, la taille de l'étape est réduite par un facteur de 0,5 jusqu'à ce que les augmentations de log de vraisemblance ou le nombre maximum de dichotomie soient atteints. Spécifiez un nombre entier positif.
- **Limiter les itérations basées sur les changements dans les estimations de 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 ne doit pas être négative.
- **Limiter les itérations basées sur les changements dans les rapports 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 ne doit pas être négative.
- **Vérifier la séparation complète 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.

- **Afficher l'historique des itérations** : Affiche les estimations de paramètre et les statistiques toutes les n itérations, en commençant par l'itération 0 (les estimations initiales). Si vous imprimez l'historique des itérations, la dernière itération est toujours imprimée, indépendamment de la valeur de n .

Valeurs manquantes de l'utilisateur : Les variables de plan d'échelle, ainsi que la variable dépendante et les covariables, doivent contenir des données valides. Les observations comportant des données non valides pour l'une de ces variables sont supprimées de l'analyse. Vous pouvez ainsi décider si les valeurs manquantes de l'utilisateur sont traitées comme étant valides dans la strate, le cluster, la sous-population et les facteurs.

Intervalle de confiance : Il s'agit du niveau d'intervalle de confiance pour les estimations de coefficient, les estimations de coefficient exponentielles et les rapports des cotes. Spécifiez une valeur supérieure ou égale à 50 et inférieure à 100.

Fonctions supplémentaires de la commande CSORDINAL

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 CUSTOM).
- Donner aux autres variables de modèle des valeurs autres que leur moyenne lors du calcul des rapports des cotes cumulées des facteurs et covariables (à l'aide de la sous-commande ODDS RATIOS).
- Utiliser des valeurs non libellées comme catégories de référence personnalisées pour les facteurs lorsque les rapports des cotes sont demandés (à l'aide de la sous-commande ODDS RATIOS).
- Indiquer une valeur de tolérance pour le contrôle des singularités (à l'aide de la sous-commande CRITERIA).
- Générer un tableau des fonctions générales estimées (à l'aide de la sous-commande PRINT).
- Enregistrer plus de 25 variables de probabilité (à l'aide de la sous-commande SAVE).

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

Chapitre 12. Régression de Cox des échantillons complexes

La procédure de la régression de Cox des échantillons complexes effectue une analyse de survie pour les échantillons réalisés à l'aide de méthodes d'échantillonnage complexes. Vous pouvez également demander une analyse pour une sous-population.

Exemples : Une administration chargée de l'application de la loi s'inquiète des taux de récidive dans sa juridiction. L'une des mesures de récidive est le temps qui s'écoule avant la deuxième arrestation des délinquants. L'agence souhaite modéliser le temps s'écoulant jusqu'à la deuxième arrestation à l'aide d'une régression de Cox, mais craint que l'hypothèse des risques proportionnels ne soit pas valide sur l'ensemble des tranches d'âge.

Des chercheurs en médecine étudient les durées de survie des patients qui quittent un programme de rééducation à la suite d'un accident ischémique. Il est possible d'avoir plusieurs observations par sujet, étant donné que les antécédents des patients changent à mesure que des événements non mortels significatifs sont constatés et que l'heure de ces événements est enregistrée. L'échantillon est également tronqué à gauche dans le sens où les durées de survie observées sont "augmentées" par la durée de la rééducation. En effet, si le risque commence au moment de l'accident ischémique, seuls les patients survivant au programme de rééducation sont inclus dans l'échantillon.

Remarques sur les données de régression de Cox des échantillons complexes

Durée de survie : La procédure applique la régression de Cox à l'analyse des durées de survie, c'est-à-dire le temps qui s'écoule avant qu'un événement se produise. Il existe deux façons d'indiquer la durée de survie, en fonction de l'heure de début de l'intervalle :

- **Temps=0 :** En général, vous aurez des informations complètes concernant le début de l'intervalle pour chaque sujet et vous aurez simplement une variable contenant les heures de fin (ou créez une variable unique avec des heures de fin à partir des variables de date et d'heure (voir ci-dessous).
- **Varie en fonction du sujet :** Cette solution est adaptée lorsque vous avez une **troncation à gauche**, également appelée **entrée différée**. Par exemple, si vous analysez les durées de survie de patients quittant un programme de rééducation suite à une attaque, vous pouvez considérer que le risque commence au moment de l'attaque. Cependant, si votre échantillon ne comprend que des patients ayant survécu au programme de rééducation, il est tronqué à gauche car les durées de survie observées sont "augmentées" par la durée de la rééducation. Pour représenter cette information, indiquez l'heure à laquelle ils ont quitté le programme de rééducation comme heure d'entrée dans l'étude.

Variables de date et d'heure : Vous ne pouvez pas utiliser les variables de date et d'heure pour définir directement le début et la fin de l'intervalle. Si vous disposez de variables de ce type, vous devez les utiliser pour créer des variables contenant des durées de survie. En l'absence de troncation à gauche, il suffit de créer une variable contenant des heures de fin, basée sur la différence entre la date d'entrée dans l'étude et la date d'observation. En présence d'une troncation à gauche, créez une variable contenant des heures de début, basée sur la différence entre la date de début de l'étude et la date d'entrée, et une variable contenant des heures de fin, basée sur la différence entre la date de début de l'étude et la date d'observation.

Statut des événements : Vous avez besoin d'une variable pour enregistrer si le sujet a vécu l'événement étudié dans l'intervalle. Les sujets qui n'ont pas connu cet événement sont censurés à droite.

Identificateur de sujet : Vous pouvez facilement incorporer des prédicteurs chronologiques de constante "par morceau" en scindant les observations d'un sujet sur plusieurs observations. Par exemple, si vous analysez les durées de survie des patients suite à l'attaque, les variables représentant leurs antécédents médicaux doivent être utiles en tant que prédicteurs. Au fil du temps, ils peuvent vivre des événements

médicaux majeurs modifiant leurs antécédents médicaux. Le tableau ci-dessous montre comment structurer un tel jeu de données : *ID patient* est l'identificateur de sujet, *Heure de fin* définit les intervalles observés, *Statut* enregistre des événements médicaux majeurs, et *Antécédents d'attaques cardiaques* et *Antécédents d'hémorragies* sont des prédicteurs chronologiques de constance "par morceau".

Tableau 1. Structure de données permettant d'intégrer des prédicteurs chronologiques de constance "par morceau".

<i>ID patient</i>	<i>Fin du programme</i>	<i>Statut</i>	<i>Antécédents d'attaques cardiaques</i>	<i>Antécédents d'hémorragies</i>
1	5	Attaque cardiaque	Non	Non
1	7	Hémorragie	Oui	Non
1	8	Décédé	Oui	Oui
2	24	Décédé	Non	Non
3	8	Attaque cardiaque	Non	Non
3	15	Décédé	Oui	Non

Hypothèses : Les observations dans le fichier de données représentent un échantillon provenant d'un plan complexe et devant être analysées en fonction des spécifications du fichier sélectionné dans la Boîte de dialogue Plan d'échantillonnages complexes.

En général, les modèles de régression de Cox partent du principe que les risques sont proportionnels, c'est-à-dire que le rapport des risques d'une observation à une autre ne doit pas varier dans le temps. Si cette hypothèse n'est pas valable, il peut être nécessaire d'ajouter au modèle des prédicteurs chronologiques.

Analyse Kaplan-Meier : Si vous ne sélectionnez aucun prédicteur (ou que vous n'entrez aucun prédicteur sélectionné dans le modèle) et que vous choisissez la méthode de limite du produit pour calculer la courbe de survie de la ligne de base dans l'onglet Options, la procédure effectue une analyse de survie de type Kaplan-Meier.

Pour obtenir une régression de Cox des échantillons complexes

1. A partir des menus, sélectionnez :
Analyse > Echantillons complexes > Régression de Cox...
2. Sélectionnez un fichier de plan. Vous pouvez également sélectionner un fichier personnalisé de probabilités conjointes.
3. Cliquez sur **Poursuivre**.
4. Spécifiez la durée de survie en sélectionnant les heures d'entrée et de sortie de l'étude.
5. Sélectionnez une variable de statut de l'événement.
6. Cliquez sur Définir un événement et précisez au moins une valeur.

Vous pouvez également sélectionner un identificateur de sujet.

Définition d'un événement

Spécifiez les valeurs indiquant qu'un événement final s'est produit.

- **Valeur(s) individuelle(s) :** Spécifiez une ou plusieurs valeurs en les entrant dans la grille ou en les sélectionnant à partir d'une liste de valeurs contenant des libellés de valeur définis.
- **Plage de valeurs :** Spécifiez une plage de valeurs en entrant les valeurs minimum et maximum, ou en sélectionnant des valeurs à partir d'une liste contenant des libellés de valeur définis.

Prédicteurs

L'onglet Prédicteurs vous permet d'indiquer les covariables et facteurs utilisés pour élaborer des effets de modèle.

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

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

Prédicteurs chronologiques : Dans certains cas, l'hypothèse des risques proportionnels n'est pas valable. Les taux de probabilité des risques varient dans le temps. Les valeurs de l'un ou plusieurs de vos prédicteurs sont différentes à différentes dates. Dans de tels cas, vous devez spécifier des prédicteurs chronologiques. Pour plus d'informations, voir «Définition d'un prédicteur chronologique». Les prédicteurs chronologiques peuvent être sélectionnés en tant que facteurs ou covariables.

Définition d'un prédicteur chronologique

La boîte de dialogue Définir un prédicteur chronologique vous permet de créer un prédicteur dépendant de la variable de temps intégrée T_* . Vous pouvez l'utiliser pour définir des prédicteurs chronologiques de deux façons :

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

Remarque : Si votre prédicteur chronologique segmenté est constant dans les segments, comme dans l'exemple de pression artérielle ci-dessus, il vous sera plus facile d'indiquer le prédicteur chronologique de constante "par morceau" en scindant les sujets sur plusieurs observations. Pour plus d'informations, voir la section relative aux identificateurs de sujets dans Chapitre 12, «Régression de Cox des échantillons complexes», à la page 49.

Dans la boîte de dialogue Définir un prédicteur chronologique, vous pouvez utiliser des contrôles de construction de fonction pour construire l'expression pour le prédicteur chronologique ou bien vous pouvez la saisir directement dans la zone de texte Expression numérique. 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. Le nom attribué à la variable est celui que vous spécifiez. Cette variable doit être incluse comme facteur ou covariable dans l'onglet Prédicteurs.

Sous-groupes

Strates de la ligne de base : Une fonction de risque de base et de survie distincte est calculée pour chaque valeur de cette variable, alors qu'un ensemble unique de coefficients du modèle est estimé sur l'ensemble des strates.

Variable de sous-population : Indiquez une variable pour définir une sous-population. L'analyse est effectuée uniquement pour la catégorie sélectionnée de la variable de sous-population.

Modèle

Spécifier les effets du modèle : Par défaut, la procédure élabore un modèle contenant des effets principaux à l'aide des covariables et facteurs spécifiés dans la boîte de dialogue principale. Vous pouvez également créer un modèle personnalisé contenant des effets d'interaction et des termes imbriqués.

Termes non imbriqués

Pour les facteurs et covariables sélectionnés :

Interaction : Crée le terme d'interaction du plus haut niveau pour 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.

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.

Statistiques

Informations sur le plan d'échantillonnage : Affiche les informations récapitulatives relatives à l'échantillon, y compris les effectifs non pondérés et la taille de la population.

Récapitulatif de l'événement et de la censure : Affiche des informations récapitulatives relatives au nombre et au pourcentage d'observations censurées.

Ensemble des risques au moment des événements : Affiche le nombre d'événements et le nombre à risque au moment de chaque événement dans chacune des strates de la ligne de base.

Paramètres : Ce groupe permet de contrôler l'affichage des statistiques associées aux paramètres du modèle.

- **Estimation** : Affiche des estimations des coefficients.
- **Estimation mise en exposant** : Affiche la base du logarithme népérien élevée à la puissance des estimations des coefficients. L'estimation présente des propriétés parfaitement adaptées aux tests statistiques ; l'estimation exponentielle, ou $\exp(B)$, est, quant à elle, plus facile à interpréter.
- **Erreur standard** : Affiche l'erreur standard pour chaque estimation de coefficient.
- **Intervalle de confiance** : Affiche l'intervalle de confiance pour chaque estimation de coefficient. Le niveau de confiance de l'intervalle est défini dans la boîte de dialogue Options.
- **Test t** : Affiche un test t pour chaque estimation de coefficient. L'hypothèse nulle de chaque test correspond au cas où la valeur du coefficient est 0.
- **Covariances des estimations de paramètres** : Affiche une estimation de la matrice de covariance pour les coefficients du modèle.
- **Corrélations des estimations de paramètres** : Affiche une estimation de la matrice de corrélation pour les coefficients du modèle.
- **Effet de plan** : Rapport entre la variance de l'estimation et la variance, en partant du principe que l'échantillon est un échantillon aléatoire simple. Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.
- **Racine carrée de l'effet de plan** : Il s'agit d'une mesure de l'effet de spécification d'un plan complexe, où les valeurs éloignées de 1 indiquent des effets importants.

Hypothèses du modèle : Ce groupe vous permet de générer un test de l'hypothèse des risques proportionnels. Le test compare le modèle ajusté à un modèle alternatif incluant des prédicteurs chronologiques x_{TF} pour chaque prédicteur x , TF étant la fonction d'heure indiquée.

- **Fonction d'heure** : Indique la forme de TF pour le modèle alternatif. Pour la fonction d'**identité**, $TF=T$. Pour la fonction **log**, $TF=\log(T)$. Pour **Kaplan-Meier**, $TF=1-S_{KM}(T)$, où $S_{KM}(\cdot)$ correspond à l'estimation Kaplan-Meier de la fonction de survie. Pour le **rang**, TF est le rang de T parmi les heures de fin observées.
- **Estimations des paramètres pour modèle alternatif** : Affiche l'estimation, l'erreur standard et l'intervalle de confiance pour chaque paramètre du modèle alternatif.
- **Matrice de covariance pour modèle alternatif** : Affiche la matrice de covariances estimées entre les paramètres du modèle alternatif.

Fonctions de survie et de risque cumulées de la ligne de base : Affiche la fonction de survie de la ligne de base et la fonction de risques cumulés de la ligne de base, ainsi que leurs erreurs standard.

Remarque : Si des prédicteurs chronologiques définis dans l'onglet Prédicteurs sont inclus dans le modèle, cette option n'est pas disponible.

Tracés

L'onglet Tracés vous permet de demander des tracés des fonctions de risque, de survie, LN (-Logn) de la fonction de survie et un moins survie. Vous pouvez également choisir de tracer les intervalles de confiance le long des fonctions spécifiées. Le niveau de confiance est défini dans l'onglet Options.

Motifs de prédicteurs : Vous pouvez indiquer un motif de valeurs de prédicteur à utiliser pour les tracés demandés et le fichier de survie exporté dans l'onglet Exporter. Notez que ces options ne sont pas disponibles si les prédicteurs chronologiques définis dans l'onglet Prédicteurs sont inclus dans le modèle.

- **Tracer les facteurs à** : Par défaut, chaque facteur est évalué à son niveau maximum. Entrez ou sélectionnez un autre niveau, si vous le souhaitez. Vous pouvez également choisir de tracer des courbes distinctes pour chaque niveau d'un facteur unique en cochant la case de ce facteur.

- **Tracer les covariables à :** Chaque covariable est évaluée à sa moyenne. Entrez ou sélectionnez une autre valeur, si vous le souhaitez.

Tests d'hypothèse

Statistique de test : Ce groupe vous permet de sélectionner le type de statistiques à utiliser pour tester les hypothèses. Vous avez le choix entre F , F ajusté, khi-deux et khi-deux ajusté.

Degrés de liberté de l'échantillonnage : Ce groupe permet de contrôler les degrés de liberté du plan d'échantillonnage utilisés pour calculer les valeurs p pour toutes les statistiques de test. Si elle est basée sur le plan d'échantillonnage, cette valeur correspond à la différence entre le nombre d'unités d'échantillonnage principales et le nombre de strates présentes à la première étape de l'échantillonnage. Vous pouvez également définir un degré de liberté personnalisé en indiquant un entier positif.

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.
- *Sidak séquentiel.* Il s'agit d'une procédure descendante de rejet séquentiel de Sidak beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Bonferroni séquentiel.* Il s'agit d'une procédure descendante de rejet séquentiel de Bonferroni beaucoup moins stricte en ce qui concerne le rejet des différentes hypothèses mais qui conserve le même niveau global de signification.
- *Sidak.* Cette méthode fournit des limites plus strictes que l'approche de Bonferroni.
- *Bonferroni.* Cette méthode ajuste le niveau de signification observé au fait que plusieurs contrastes sont en cours de test.

Enregistrement

Enregistrer les variables : Ce groupe vous permet d'enregistrer les variables liées au modèle dans le jeu de données actif pour les utiliser ultérieurement dans les diagnostics et les rapports des résultats. Notez qu'elles ne sont pas disponibles quand des prédicteurs chronologiques sont inclus dans le modèle.

- **Fonction de survie :** Enregistre la probabilité de survie (valeur de la fonction de survie) au moment observé et les valeurs des prédicteurs pour chaque observation.
- **Borne inférieure de l'intervalle de confiance pour la fonction de survie :** Enregistre la borne inférieure de l'intervalle de confiance pour la fonction de survie au moment observé et les valeurs des prédicteurs pour chaque observation.
- **Borne supérieure de l'intervalle de confiance pour la fonction de survie :** Enregistre la borne supérieure de l'intervalle de confiance pour la fonction de survie au moment observé et les valeurs des prédicteurs pour chaque observation.
- **Fonction de risque cumulée :** Enregistre le risque cumulé, ou $-\ln(\text{survie})$, au moment observé et les valeurs des prédicteurs pour chaque observation.
- **Borne inférieure de l'intervalle de confiance pour la fonction de risque cumulée :** Enregistre la borne inférieure de l'intervalle de confiance pour la fonction de risque cumulée au moment observé et les valeurs des prédicteurs pour chaque observation.
- **Borne supérieure de l'intervalle de confiance pour la fonction de risque cumulée :** Enregistre la borne supérieure de l'intervalle de confiance pour la fonction de risque cumulée au moment observé et les valeurs des prédicteurs pour chaque observation.

- **Prévision du prédicteur linéaire** : Enregistre la combinaison linéaire des coefficients de régression pour les heures des prédicteurs corrigés de la valeur de référence. Le prédicteur linéaire correspond au rapport entre la fonction de risque et de risque de base. Dans le modèle des risques proportionnels, cette valeur est constante dans le temps.
- **Résidu de Schoenfeld** : Pour chaque observation non censurée et chaque paramètre non redondant du modèle, le résidu de Schoenfeld correspond à la différence entre la valeur observée du prédicteur associé au paramètre du modèle et la valeur théorique du prédicteur pour les observations de l'ensemble des risques au moment où les événements sont observés. Les résidus de Schoenfeld peuvent être utilisés pour évaluer l'hypothèse des risques proportionnels. Par exemple, pour un prédicteur x , les tracés des résidus de Schoenfeld pour le prédicteur chronologique $x \cdot \ln(T_)$ par rapport au temps doivent montrer une ligne horizontale au niveau de 0 si les risques proportionnels sont valables. Une variable distincte est enregistrée pour chaque paramètre non redondant du modèle. Les résidus de Schoenfeld sont uniquement calculés pour les observations non censurées.
- **Résidu de Martingale** : Pour chaque observation, le résidu de Martingale correspond à la différence entre la censure observée (0 si elle est censurée, 1 si ce n'est pas le cas) et la prévision d'un événement pendant le temps d'observation.
- **Résidus au sens déviance** : Les résidus au sens déviance sont des résidus de Martingale "ajustés" pour paraître plus symétriques au niveau de 0. Les tracés des résidus au sens déviance par rapport aux prédicteurs ne doivent révéler aucun motif.
- **Résidu de Cox-Snell** : Pour chaque observation, le résidu de Cox-Snell correspond à la prévision d'un événement pendant la période d'observation, ou la censure observée moins le résidu de Martingale.
- **Résidu de score** : Pour chaque observation et chaque paramètre non redondant du modèle, le résidu de score est la contribution de l'observation à la première dérivée de la pseudo-vraisemblance. Une variable distincte est enregistrée pour chaque paramètre non redondant du modèle.
- **Résidu de la différence de Bêta** : Pour chaque observation et chaque paramètre non redondant du modèle, le résidu de la différence de Bêta se rapproche du changement de la valeur de l'estimation des paramètres lorsque l'observation est supprimée du modèle. Il est possible que les observations ayant des résidus de la différence de Bêta relativement élevés aient un effet d'annulation sur l'analyse. Une variable distincte est enregistrée pour chaque paramètre non redondant du modèle.
- **Résidus agrégés** : Quand plusieurs observations représentent un sujet unique, le résidu agrégé d'un sujet est simplement la somme des résidus de l'observation correspondante sur l'ensemble des observations appartenant au même sujet. Pour le résidu de Schoenfeld, la version agrégée est semblable à celle de la version non agrégée parce que le résidu de Schoenfeld est défini uniquement pour les observations non censurées. Ces résidus ne sont disponibles que si un identificateur de sujet est spécifié dans l'onglet Heure et événement.

Noms des variables enregistrées : Grâce à la génération automatique de nom, vous conservez l'ensemble de votre travail. Les noms personnalisés vous permettent de supprimer/remplacer les résultats d'exécutions précédentes sans supprimer d'abord les variables enregistrées dans l'éditeur de données.

Exportation

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

- **rowtype_** : Prend les valeurs (et libellés de valeurs) suivantes : COV (covariances), CORR (corrélations), EST (estimations des paramètres), SE (erreurs standard), SIG (niveau de signification) et DF (degrés de liberté du plan d'échantillonnage). Il existe une observation distincte avec le type de ligne COV (ou CORR) pour chaque paramètre de modèle et une observation distincte pour chacun des autres types de ligne.

- **varname_ :** Prend les valeurs P1, P2, etc., correspondant à une liste triée de tous les paramètres de modèle pour les types de ligne COV ou CORR, avec des libellés de valeur correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres. Les cellules sont vides pour les autres types de ligne.
- **P1, P2, .. :** Ces variables correspondent à une liste triée de tous les paramètres de modèle, avec des libellés de variable correspondant aux chaînes de paramètres affichées dans le tableau Estimations des paramètres, et prennent leurs 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.

Remarque : 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.

Exporter la fonction de survie sous forme de données IBM SPSS Statistics : Ecrit un jeu de données dans un format IBM SPSS Statistics contenant la fonction de survie, l'erreur standard de la fonction de survie, les bornes supérieure et inférieure de l'intervalle de confiance de la fonction de survie, et la fonction des risques cumulés pour chaque échec ou moment de l'événement, évaluées à la ligne de base et aux motifs de prédicteurs spécifiés dans l'onglet Tracé. L'ordre des variables dans le fichier de matrice est le suivant.

- **Variable Strate de base :** Des tableaux de survie distincts sont générés pour chaque valeur de la variable de strates.
- **Variable Durée de survie :** Moment de l'événement. Une observation distincte est créée à chaque moment de l'événement.
- **Sur_0, LCL_Sur_0, UCL_Sur_0 :** Fonction de survie de la ligne de base et bornes supérieure et inférieure de son intervalle de confiance.
- **Sur_R, LCL_Sur_R, UCL_Sur_R :** Fonction de survie évaluée au motif de "référence" (voir le tableau des valeurs de motif dans les sorties) et bornes supérieure et inférieure de son intervalle de confiance.
- **Sur_##, LCL_Sur_##, UCL_Sur_##, ...** Fonction de survie évaluée à chaque modèle de prédicteurs spécifié sur l'onglet Tracés et bornes supérieure et inférieure de leur intervalle de confiance. Consultez le tableau des valeurs de motif dans les sorties pour faire correspondre les motifs au nombre ##.
- **Haz_0, LCL_Haz_0, UCL_Haz_0 :** Fonction de risque cumulée de la ligne de base et bornes supérieure et inférieure de son intervalle de confiance.
- **Haz_R, LCL_Haz_R, UCL_Haz_R :** Fonction de risque cumulée évaluée au motif de "référence" (voir le tableau des valeurs de motif dans les sorties) et bornes supérieure et inférieure de son intervalle de confiance.
- **Haz_##, LCL_Haz_##, UCL_Haz_##, ...** Fonction de risque cumulée évaluée à chaque modèle de prédicteurs spécifié sur l'onglet Tracés et bornes supérieure et inférieure de leur intervalle de confiance. Consultez le tableau des valeurs de motif dans les sorties pour faire correspondre les motifs au nombre ##.

Exporter le modèle au format XML : Enregistre toutes les informations nécessaires pour prévoir la fonction de survie, y compris les estimations des paramètres et la fonction de survie de la ligne de base, au format XML (PMML). Vous pouvez utiliser ce fichier de modèle pour appliquer les informations du modèle aux autres fichiers de données à des fins d'évaluation.

Options

Estimation : Ces contrôles spécifient des critères pour l'estimation des coefficients de régression.

- **Maximum des itérations :** Nombre maximal d'itérations exécutées par l'algorithme. Spécifiez un nombre entier non négatif.

- **Nombre maximum de dichotomies** : A chaque itération, la taille de l'étape est réduite par un facteur de 0,5 jusqu'à ce que les augmentations de log de vraisemblance ou le nombre maximum de dichotomie soient atteints. Spécifiez un nombre entier positif.
- **Limiter les itérations basées sur les changements dans les estimations de 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.
- **Limiter les itérations basées sur les changements dans les rapports 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.
- **Afficher l'historique des itérations** : Affiche l'historique des itérations pour les estimations des paramètres et le pseudo log de vraisemblance, et imprime la dernière évaluation du changement des estimations des paramètres et de la pseudo log de vraisemblance. Le tableau de l'historique des itérations imprime toutes les n itérations en commençant par l'itération 0 (les estimations initiales), n étant la valeur de l'incrément. Si l'historique des itérations est requis, la dernière itération est toujours affichée quel que soit n .
- **Méthode de départage pour l'estimation des paramètres** : Lorsque vous observez des heures d'échec ex aequo, l'une de ces méthodes est utilisée pour départager les ex aequo. La méthode d'Efron nécessite plus de calculs.

Fonctions de survie : Ces contrôles spécifient des critères pour les calculs impliquant la fonction de survie.

- **Méthode d'estimation des fonctions de survie de la ligne de base** : La méthode de **Breslow** (ou de Nelson-Aalan ou empirique) estime le risque cumulé de la ligne de base par une fonction d'échelon non décroissante avec des échelons aux heures d'échec observées, puis calcule la survie de la ligne de base par la relation survie= $\exp(-\text{risque cumulé})$. La méthode d'**Efron** nécessite plus de calculs et se limite à la méthode de Breslow quand il n'y a pas d'ex aequo. La méthode de **limite du produit** estime la survie de la ligne de base par une fonction continue à droite non croissante. S'il n'y a pas de prédicteurs dans le modèle, cette méthode se limite à l'estimation de Kaplan-Meier.
- **Intervalle de confiance des fonctions de survie** : L'intervalle de confiance peut être calculé de trois manières : dans les unités d'origine, via une transformation Log ou une transformation Log moins log (-Logn). Seule la transformation LN (-Logn) garantit que les bornes de l'intervalle de confiance seront comprises entre 0 et 1, mais la transformation log semble généralement donner de "meilleurs" résultats.

Valeurs manquantes de l'utilisateur : Toutes les variables doivent avoir des valeurs valides pour qu'une observation puisse être incluse dans l'analyse. Ces contrôles vous permettent de décider si les valeurs manquantes de l'utilisateur sont considérées comme valides parmi les modèles qualitatifs (y compris les variables catégorielles, d'événement, de strates et de sous-population) et les variables du plan d'échantillonnage.

Intervalle de confiance (%) : Il s'agit du niveau d'intervalle de confiance utilisé pour les estimations de coefficient, les estimations de coefficient exponentielles, les estimations de fonction de survie et les estimations de fonction de risque cumulée. Spécifiez une valeur supérieure ou égale à 0, et inférieure à 100.

Fonctions supplémentaires de la commande CSCOXREG

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

- Effectuer des tests d'hypothèse personnalisés (à l'aide de la sous-commande CUSTOM et de /PRINT LMATRIX).
- Spécifier la tolérance (à l'aide de /CRITERIA SINGULAR).
- Tableau de la fonction générale estimée (à l'aide de /PRINT GEF).

- Motifs de prédicteurs multiples (à l'aide de plusieurs sous-commandes PATTERN).
- Nombre maximum de variables enregistrées quand un nom de racine est spécifié (à l'aide de la sous-commande SAVE). La boîte de dialogue prend en compte la valeur CSCOXREG par défaut de 25 variables.

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

Remarques

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

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

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

Pour le Canada, veuillez adresser votre courrier à :

IBM Director of Commercial Relations
IBM Canada Ltd
3600 Steeles Avenue East
Markham, Ontario
L3R 9Z7 Canada

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

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

Le paragraphe suivant ne s'applique ni au Royaume-Uni, ni dans aucun pays dans lequel il serait contraire aux lois locales : 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 logiciels décrits dans ce document.

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

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

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

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

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

Le logiciel sous licence décrit dans ce document et tous les éléments sous licence disponibles s'y rapportant sont fournis par IBM conformément aux dispositions du Livret Contractuel IBM, des Conditions Internationales d'Utilisation de Logiciels IBM, des Conditions d'Utilisation du Code Machine ou de tout autre contrat équivalent.

Les données de performance indiquées dans ce document ont été déterminées dans un environnement contrôlé. Par conséquent, les résultats peuvent varier de manière significative selon l'environnement d'exploitation utilisé. Certaines mesures évaluées sur des systèmes en cours de développement ne sont pas garanties sur tous les systèmes disponibles. En outre, elles peuvent résulter d'extrapolations. Les résultats peuvent donc varier. Il incombe aux utilisateurs de ce document de vérifier si ces données sont applicables à leur environnement d'exploitation.

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

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

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

LICENCE DE COPYRIGHT :

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

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

© (nom de votre société) (année). Des segments de code sont dérivés des Programmes exemples d'IBM Corp.

© Copyright IBM Corp. _entrez l'année ou les années_. Tous droits réservés.

Marques

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

Adobe, le logo Adobe, PostScript et le logo PostScript sont des marques d'Adobe Systems Incorporated aux Etats-Unis et/ou dans certains autres pays.

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

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

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

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

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

Index

B

- Bonferroni
 - dans Echantillons complexes 33, 39, 46
 - Régression de Cox des échantillons complexes 54

C

- catégorie de référence
 - dans le modèle linéaire général des échantillons complexes 33
 - dans une régression logistique des échantillons complexes 37
- catégories estimées
 - dans une régression logistique des échantillons complexes 40
 - dans une régression ordinale des échantillons complexes 46
- clusters
 - dans l'assistant d'échantillonnage 5
 - dans l'assistant de préparation d'analyse 13
- Coefficient de variation (COV)
 - dans Echantillons complexes - Fréquences 19
 - dans Echantillons complexes - Descriptives 23
 - dans Echantillons complexes - Rapports 29
 - dans les tableaux croisés d'échantillons complexes 25
- Contrastes
 - dans le modèle linéaire général des échantillons complexes 33
- Contrastes à la précédente
 - dans le modèle linéaire général des échantillons complexes 33
- contrastes de déviation
 - dans le modèle linéaire général des échantillons complexes 33
- Contrastes de Helmert
 - dans le modèle linéaire général des échantillons complexes 33
- Contrastes différence
 - dans le modèle linéaire général des échantillons complexes 33
- Contrastes polynomiaux
 - dans le modèle linéaire général des échantillons complexes 33
- Contrastes simples
 - dans le modèle linéaire général des échantillons complexes 33
- convergence de vraisemblance
 - dans une régression logistique des échantillons complexes 41
 - dans une régression ordinale des échantillons complexes 47
- Convergence des paramètres
 - dans une régression logistique des échantillons complexes 41
 - dans une régression ordinale des échantillons complexes 47
- Correction Bonferroni séquentiel
 - dans Echantillons complexes 33, 39, 46
 - Régression de Cox des échantillons complexes 54
- Correction Sidak
 - dans Echantillons complexes 33, 39, 46
 - Régression de Cox des échantillons complexes 54
- Correction Sidak séquentiel
 - dans Echantillons complexes 33, 39, 46
 - Régression de Cox des échantillons complexes 54
- corrélations des estimations de paramètres
 - dans le modèle linéaire général des échantillons complexes 32
 - dans une régression logistique des échantillons complexes 38
 - dans une régression ordinale des échantillons complexes 45
- covariances des estimations de paramètres
 - dans le modèle linéaire général des échantillons complexes 32
 - dans une régression logistique des échantillons complexes 38
 - dans une régression ordinale des échantillons complexes 45

D

- degrés de liberté
 - dans Echantillons complexes 33, 39, 46
 - Régression de Cox des échantillons complexes 54
- différence de risque
 - dans les tableaux croisés d'échantillons complexes 25
- Différence la moins significative
 - dans Echantillons complexes 33, 39, 46
 - Régression de Cox des échantillons complexes 54

E

- Echantillonnage
 - plan complexe 5
- échantillonnage aléatoire simple
 - dans l'assistant d'échantillonnage 6

- échantillonnage complexe
 - plan d'analyse 13
 - plan d'échantillonnage 5
- échantillonnage PPS
 - dans l'assistant d'échantillonnage 6
- échantillonnage séquentiel
 - dans l'assistant d'échantillonnage 6
- échantillonnage systématique
 - dans l'assistant d'échantillonnage 6
- Echantillons complexes
 - tests d'hypothèse 33, 39, 46
 - valeurs manquantes 20, 26
- échantillons complexes
 - options 20, 24, 27, 30
- Echantillons complexes -
 - Caractéristiques 23
 - statistiques 23
 - valeurs manquantes 24
- Echantillons complexes - Effectifs 19
 - statistiques 19
- Echantillons complexes - Rapports 29
 - statistiques 29
 - valeurs manquantes 30
- effectif non pondéré
 - dans Echantillons complexes - Fréquences 19
 - dans Echantillons complexes - Descriptives 23
 - dans Echantillons complexes - Rapports 29
 - dans les tableaux croisés d'échantillons complexes 25
- effet de plan
 - dans Echantillons complexes - Fréquences 19
 - dans Echantillons complexes - Descriptives 23
 - dans Echantillons complexes - Rapports 29
 - dans le modèle linéaire général des échantillons complexes 32
 - dans les tableaux croisés d'échantillons complexes 25
 - dans une régression logistique des échantillons complexes 38
 - dans une régression ordinale des échantillons complexes 45
 - Régression de Cox des échantillons complexes 52
- Erreur standard
 - dans Echantillons complexes - Fréquences 19
 - dans Echantillons complexes - Descriptives 23
 - dans Echantillons complexes - Rapports 29
 - dans le modèle linéaire général des échantillons complexes 32
 - dans les tableaux croisés d'échantillons complexes 25

Erreur standard (*suite*)
 dans une régression logistique des
 échantillons complexes 38
 dans une régression ordinale des
 échantillons complexes 45
 estimation d'échantillonnage
 dans l'assistant de préparation
 d'analyse 14
 Estimations des paramètres
 dans le modèle linéaire général des
 échantillons complexes 32
 dans une régression logistique des
 échantillons complexes 38
 dans une régression ordinale des
 échantillons complexes 45
 Régression de Cox des échantillons
 complexes 52
 Evaluation de Fisher
 dans une régression ordinale des
 échantillons complexes 47

F
 fichier de plan 2

H
 Historique des itérations
 dans une régression logistique des
 échantillons complexes 41
 dans une régression ordinale des
 échantillons complexes 47

I
 informations sur le plan
 d'échantillonnage
 Régression de Cox des échantillons
 complexes 52
 Intervalles de confiance
 dans Echantillons complexes -
 Fréquences 19
 dans Echantillons complexes -
 Descriptives 23
 dans Echantillons complexes -
 Rapports 29
 dans le modèle linéaire général des
 échantillons complexes 32, 35
 dans les tableaux croisés
 d'échantillons complexes 25
 dans une régression logistique des
 échantillons complexes 38
 dans une régression ordinale des
 échantillons complexes 45
 Itérations
 dans une régression logistique des
 échantillons complexes 41
 dans une régression ordinale des
 échantillons complexes 47

K
 khi-deux
 dans Echantillons complexes 33, 39,
 46

khi-deux (*suite*)
 Régression de Cox des échantillons
 complexes 54
 khi-deux ajusté
 dans Echantillons complexes 33, 39,
 46
 Régression de Cox des échantillons
 complexes 54

M
 mesure de la taille
 dans l'assistant d'échantillonnage 6
 méthode d'échantillonnage
 dans l'assistant d'échantillonnage 6
 Méthode d'échantillonnage Brewer
 dans l'assistant d'échantillonnage 6
 Méthode d'échantillonnage Murthy
 dans l'assistant d'échantillonnage 6
 Méthode d'échantillonnage Sampford
 dans l'assistant d'échantillonnage 6
 Méthode d'estimation d'Efron
 Régression de Cox des échantillons
 complexes 56
 Méthode d'estimation de Breslow
 Régression de Cox des échantillons
 complexes 56
 Méthode de Newton-Raphson
 dans une régression ordinale des
 échantillons complexes 47
 Modèle linéaire général des échantillons
 complexes 31
 Enregistrement de variables 34
 fonctions supplémentaires de la
 commande 35
 Modèle 32
 moyennes estimées 33
 Options 35
 statistiques 32
 Moyenne
 dans Echantillons complexes -
 Descriptives 23
 Moyennes marginales estimées
 dans le modèle linéaire général des
 échantillons complexes 33

N
 niveau de confiance
 dans une régression logistique des
 échantillons complexes 41
 dans une régression ordinale des
 échantillons complexes 47

P
 plan d'analyse 13
 plan d'échantillonnage 5
 pondérations d'échantillon
 dans l'assistant d'échantillonnage 8
 dans l'assistant de préparation
 d'analyse 13
 pondérations d'échantillon d'entrée
 dans l'assistant d'échantillonnage 5

Pourcentages en colonne
 dans les tableaux croisés
 d'échantillons complexes 25
 Pourcentages en ligne
 dans les tableaux croisés
 d'échantillons complexes 25
 Pourcentages en tableau
 dans Echantillons complexes -
 Fréquences 19
 dans les tableaux croisés
 d'échantillons complexes 25
 prédicteur chronologique
 Régression de Cox des échantillons
 complexes 51
 Prévisions
 dans le modèle linéaire général des
 échantillons complexes 34
 probabilité prédite
 dans une régression logistique des
 échantillons complexes 40
 dans une régression ordinale des
 échantillons complexes 46
 probabilités cumulées
 dans une régression ordinale des
 échantillons complexes 46
 probabilités d'inclusion
 dans l'assistant d'échantillonnage 8
 probabilités des réponses
 dans une régression ordinale des
 échantillons complexes 44
 proportion de l'échantillon
 dans l'assistant d'échantillonnage 8

R
 racine carrée d'effet de plan
 dans Echantillons complexes -
 Fréquences 19
 dans Echantillons complexes -
 Descriptives 23
 dans Echantillons complexes -
 Rapports 29
 dans le modèle linéaire général des
 échantillons complexes 32
 dans les tableaux croisés
 d'échantillons complexes 25
 dans une régression logistique des
 échantillons complexes 38
 dans une régression ordinale des
 échantillons complexes 45
 Régression de Cox des échantillons
 complexes 52
 rapport des cotes
 dans les tableaux croisés
 d'échantillons complexes 25
 dans une régression logistique des
 échantillons complexes 40
 dans une régression ordinale des
 échantillons complexes 46
 régression de Cox des échantillons
 complexes
 export de modèle 55
 Régression de Cox des échantillons
 complexes
 Analyse Kaplan-Meier 49
 Définition d'événements 50
 Enregistrement de variables 54

- Régression de Cox des échantillons complexes (*suite*)
 - Modèle 52
 - Options 56
 - prédicteur chronologique 51
 - prédicteurs 51
 - sous-groupes 51
 - statistiques 52
 - tests d'hypothèse 54
 - tracés 53
 - Variables de date et d'heure 49
- Régression logistique des échantillons complexes 37
 - catégorie de référence 37
 - Enregistrement de variables 40
 - fonctions supplémentaires de la commande 41
 - Modèle 38
 - Options 41
 - rapport des cotes 40
 - statistiques 38
- Régression ordinale des échantillons complexes 43
 - Enregistrement de variables 46
 - Modèle 44
 - Options 47
 - probabilités des réponses 44
 - rapport des cotes 46
 - statistiques 45
- Résiduels partiels de Schoenfeld
 - Régression de Cox des échantillons complexes 54
- Résidus
 - dans le modèle linéaire général des échantillons complexes 34
 - dans les tableaux croisés d'échantillons complexes 25
- résidus agrégés
 - Régression de Cox des échantillons complexes 54
- résidus ajustés
 - dans les tableaux croisés d'échantillons complexes 25
- résidus au sens déviance
 - Régression de Cox des échantillons complexes 54
- résidus de Cox-Snell
 - Régression de Cox des échantillons complexes 54
- résidus de Martingale
 - Régression de Cox des échantillons complexes 54
- résidus de score
 - Régression de Cox des échantillons complexes 54
- risque relatif
 - dans les tableaux croisés d'échantillons complexes 25

S

- Séparation
 - dans une régression logistique des échantillons complexes 41
 - dans une régression ordinale des échantillons complexes 47

- Somme
 - dans Echantillons complexes – Descriptives 23
- sous-population
 - Régression de Cox des échantillons complexes 51
- statistique F
 - dans Echantillons complexes 33, 39, 46
 - Régression de Cox des échantillons complexes 54
- statistique F ajustée
 - dans Echantillons complexes 33, 39, 46
 - Régression de Cox des échantillons complexes 54
- statistique R2
 - dans le modèle linéaire général des échantillons complexes 32
- statistiques de pseudo R2
 - dans une régression logistique des échantillons complexes 38
 - dans une régression ordinale des échantillons complexes 45
- Step-halving
 - dans une régression logistique des échantillons complexes 41
 - dans une régression ordinale des échantillons complexes 47
- strates de la ligne de base
 - Régression de Cox des échantillons complexes 51
- stratification
 - dans l'assistant d'échantillonnage 5
 - dans l'assistant de préparation d'analyse 13

T

- Table de classification
 - dans une régression logistique des échantillons complexes 38
 - dans une régression ordinale des échantillons complexes 45
- Tableaux croisés des échantillons complexes 25
 - statistiques 25
- taille de l'échantillon
 - dans l'assistant d'échantillonnage 7, 8
- taille de la population
 - dans Echantillons complexes – Fréquences 19
 - dans Echantillons complexes – Descriptives 23
 - dans Echantillons complexes – Rapports 29
 - dans l'assistant d'échantillonnage 8
 - dans les tableaux croisés d'échantillons complexes 25
- Test des droites parallèles
 - dans une régression ordinale des échantillons complexes 45
- test des risques proportionnels
 - Régression de Cox des échantillons complexes 52

- Test t :
 - dans le modèle linéaire général des échantillons complexes 32
 - dans une régression logistique des échantillons complexes 38
 - dans une régression ordinale des échantillons complexes 45

V

- valeurs cumulées
 - dans Echantillons complexes – Fréquences 19
- valeurs manquantes
 - dans Echantillons complexes 20, 26
 - dans Echantillons complexes – Descriptives 24
 - dans Echantillons complexes – Rapports 30
 - dans le modèle linéaire général des échantillons complexes 35
 - dans une régression logistique des échantillons complexes 41
 - dans une régression ordinale des échantillons complexes 47
- Valeurs théoriques
 - dans les tableaux croisés d'échantillons complexes 25

