

IBM SPSS Categories 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 57.

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 de codage optimal pour les données catégorielles 1

Définition du codage optimal	1
Raisons de l'utilisation du codage optimal	1
Niveau de codage optimal et niveau de mesure	2
Sélection du niveau de codage optimal	2
Tracés de transformation	3
Codes de catégorie	3
Utilisation de la procédure la plus adaptée à votre application	5
Régression catégorielle	6
Analyse des correspondances multiples	6
Analyse de corrélation canonique non linéaire	7
Analyse des correspondances	8
Analyse de correspondance multiple	9
Positionnement multidimensionnel	10
Dépliage multidimensionnel	10
Rapport d'aspect des graphiques de codage optimal	11

Chapitre 2. Régression catégorielle (CATREG) 13

Définir une échelle dans la régression catégorielle	14
Régression catégorielle : Discrétisation	14
Régression catégorielle : Valeurs manquantes	15
Régression catégorielle : Options	15
Régularisation de régression catégorielle	16
Régression catégorielle : Sortie	17
Régression catégorielle : Enregistrement	17
Régression catégorielle des tracés de transformation	18
Fonctions supplémentaires de la commande CATREG	18

Chapitre 3. Analyse des correspondances multiples (CATPCA) 19

Définir l'échelle et la pondération dans CATPCA	20
Analyse des correspondances multiples : Discrétisation	21
Analyse des correspondances multiples : Valeurs manquantes	21
Analyse des correspondances multiples : Options	22
Analyse des correspondances multiples : Sortie	23
Analyse des correspondances multiples : Enregistrer	24
Analyse des correspondances multiples : Tracés d'objets et de variables	24
Analyse des correspondances multiples : Tracés de catégories	25
Analyse des correspondances multiples : Tracés de chargement	25
Analyse des correspondances multiples : Bootstrap	25
Fonctions supplémentaires de la commande CATPCA	26

Chapitre 4. Analyse canonique non linéaire (OVERALS) 27

Définition d'une plage et d'une échelle	28
Définition d'une plage	29
Analyse de corrélation canonique non linéaire – Options	29
Fonctions supplémentaires de la commande OVERALS	30

Chapitre 5. Analyse des correspondances 31

Définition de la plage de ligne dans l'analyse des correspondances	32
Définition de la plage de colonne dans l'analyse des correspondances	32
Modèle d'analyse des correspondances	33
Statistiques de l'analyse des correspondances	34
Tracés de l'analyse des correspondances	34
Fonctions supplémentaires de la commande CORRESPONDENCE	35

Chapitre 6. Analyse de correspondance multiple 37

Définition d'une pondération de variable dans une analyse de correspondance multiple	38
Analyse des correspondances multiples : Discrétisation	38
Analyse des correspondances multiples : Valeurs manquantes	38
Analyse des correspondances multiples : Options	39
Analyse des correspondances multiples : Sortie	40
Analyse des correspondances multiples : Enregistrer	41
Analyse des correspondances multiples : Tracés d'objets	41
Analyse des correspondances multiples : Tracés de variables	41
Commande MULTIPLE CORRESPONDENCE - Fonctions additionnelles	41

Chapitre 7. Positionnement multidimensionnel (PROXSCAL) 43

Proximités dans des matrices sur plusieurs colonnes	44
Proximités sur plusieurs colonnes	44
Proximités dans une colonne	45
Création de proximités à partir des données	45
Création d'une mesure à partir des données	45
Définition d'un modèle de positionnement multidimensionnel	46
Positionnement multidimensionnel : Restrictions	47
Positionnement multidimensionnel : Options	47
Positionnement multidimensionnel : Tracés, Version 1	48
Positionnement multidimensionnel : Tracés, Version 2	48

Positionnement multidimensionnel : Sortie	49
Fonctions supplémentaires de la commande PROXSCAL	49

Chapitre 8. Dépliage multidimensionnel (PREFSCAL)	51
Définition d'un modèle de dépliage multidimensionnel	52
Restrictions du dépliage multidimensionnel	52
Options de dépliage multidimensionnel	53

Tracés de dépliage multidimensionnel	54
Sortie du dépliage multidimensionnel	55
Fonctions supplémentaires de la commande PREFSCAL	55

Remarques	57
Marques	59

Index	61
------------------------	-----------

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 de codage optimal pour les données catégorielles

Les procédures de catégorie font appel au codage optimal pour analyser les données dont l'analyse, par le biais des procédures statistiques standard, est complexe, voire impossible. Ce chapitre décrit le fonctionnement de chacune des procédures, les circonstances dans lesquelles leur utilisation est la plus favorable, les relations entre les différentes procédures et les relations de ces dernières avec les procédures statistiques standard.

Remarque : Ces procédures et leur mise en oeuvre dans IBM® SPSS Statistics ont été développées par le groupe DTSS (Data Theory Scaling System), composé de membres des départements d'enseignement et de psychologie de la Faculté des sciences sociales et du comportement de l'Université de Leyde (Pays-Bas).

Définition du codage optimal

Le codage optimal consiste à associer des quantifications numériques aux catégories de chaque variable. Ainsi, les procédures standard peuvent être utilisées pour obtenir une solution portant sur les variables quantifiées.

Les valeurs d'échelle optimale sont attribuées aux catégories de chaque variable selon le critère d'optimisation de la procédure utilisée. A la différence des libellés d'origine des variables nominales ou ordinales de l'analyse, ces valeurs d'échelle ont des propriétés métriques.

Dans la plupart des procédures de catégorie, la quantification optimale de chaque variable codée est obtenue via une méthode itérative appelée **moindres carrés alternés**. Dans cette méthode, les quantifications actuelles, une fois utilisées pour chercher une solution, sont mises à jour à l'aide de cette solution. Les quantifications mises à jour permettent alors de chercher une autre solution, utilisée pour mettre à jour ces quantifications, jusqu'à ce que le critère signalant la fin du processus soit satisfait.

Raisons de l'utilisation du codage optimal

En général, les données catégorielles sont utilisées dans le cadre d'une recherche commerciale, d'une enquête ou d'une recherche liée aux sciences sociales et du comportement. En fait, nombreux sont les chercheurs qui travaillent exclusivement avec ce type de données catégorielles.

Alors que les adaptations de la plupart des modèles standard sont disponibles notamment pour l'analyse des données catégorielles, leur utilisation ne convient pas aux jeux de données contenant :

- Un nombre d'observations insuffisant
- Un nombre de variables excessif
- Un nombre de valeurs par variable excessif

En quantifiant les catégories, les méthodes de codage optimal évitent tout problème dans ces cas-là. En outre, elles s'avèrent utiles même si des méthodes spécifiques sont appropriées.

Habituellement, l'interprétation de sortie de codage optimal repose sur des graphiques, plutôt que sur des estimations de paramètres. Les méthodes de codage optimal fournissent d'excellentes analyses exploratoires qui complètent bien les autres modèles IBM SPSS Statistics. Si vous affinez votre recherche, la visualisation des données codées de façon optimale peut servir de base à une analyse centrée sur l'interprétation de paramètres de modèle.

Niveau de codage optimal et niveau de mesure

Ce concept peut fortement prêter à confusion lorsque vous utilisez les procédures de catégorie pour la première fois. Si vous spécifiez le niveau, il ne s'agit pas du niveau de *mesure* des variables, mais de leur niveau de *codage*. L'idée est la suivante : les variables à quantifier peuvent avoir des relations non linéaires, quelle que soit la manière dont elles sont mesurées.

On distingue trois niveaux de mesure de base pour les catégories :

- Le niveau **nominal** signifie que les valeurs d'une variable représentent des catégories non classées. Voici quelques exemples de variables pouvant être considérées comme nominales : les catégories de région, de code postal, d'appartenance religieuse et à choix multiples.
- Le niveau **ordinal** signifie que les valeurs d'une variable représentent des catégories classées. En voici quelques exemples : les échelles d'attitude représentant le degré de satisfaction ou de confiance, et les échelles de classement des préférences.
- Le niveau **numérique** signifie que les valeurs d'une variable représentent des catégories classées avec une mesure significative, de sorte que les comparaisons de distance entre les catégories soient adéquates. L'âge en années et le revenu en milliers de dollars constituent des exemples.

Par exemple, supposons que les variables *région*, *travail* et *âge* sont codées comme l'indique le tableau suivant.

Tableau 1. Système de codification de la région, du travail et de l'âge

Code de région	Valeur de région.	Code de travail	Valeur de travail	Age
1	Nord	1	Stagiaire	20
2	Sud	2	Représentant	22
3	Est	3	Gestionnaire	25
4	Ouest			27

Les valeurs mentionnées représentent les catégories de chaque variable. *Région* est une variable nominale. On distingue quatre catégories de *région* sans ordre intrinsèque. Les valeurs 1 à 4 représentent simplement ces quatre catégories. Le système de codification est totalement arbitraire. En revanche, la variable *travail* peut être considérée comme une variable ordinale. Les catégories d'origine représentent une progression du statut de stagiaire à celui de responsable. Plus les codes sont élevés, plus ils font référence à fonction élevée dans la hiérarchie de l'entreprise. Toutefois, seules les informations relatives à l'ordre sont connues, mais aucun élément d'information ne peut être fourni concernant la distance entre les catégories adjacentes. En revanche, la variable *âge* peut être considérée comme une variable numérique. Dans le cas de la variable *âge*, les distances entre les valeurs sont intrinsèquement explicites. La distance entre 20 et 22 est identique à celle entre 25 et 27, alors que la distance entre 22 et 25 est supérieure à ces deux distances.

Sélection du niveau de codage optimal

Il est important de comprendre qu'aucune propriété intrinsèque de variable ne prédéfinit automatiquement le niveau de codage optimal que vous devez indiquer. Vous pouvez explorer les données de manière cohérente et simplifiant l'interprétation. En analysant une variable numérique au niveau ordinal, par exemple, une transformation non linéaire autorise une solution dans un nombre inférieur de dimensions.

Les deux exemples suivants illustrent le fait que le niveau de mesure "évident" n'est peut-être pas le niveau de codage optimal. Supposez qu'une variable répartit les objets dans les différents groupes d'âge. Bien que la variable *âge* puisse être codée en tant que variable numérique, il s'avère parfois que, pour les

jeunes de moins de 25 ans, la sécurité a un rapport positif avec l'âge alors que, pour les personnes de plus de 60 ans, ce rapport est négatif. Dans ce cas, mieux vaut peut-être considérer âge comme une variable nominale.

Autre exemple : une variable triant les personnes par préférence politique semble avant tout être nominale. Toutefois, si vous trie les partis politiques de gauche à droite, il se peut que leur quantification doive respecter cet ordre. Dans ce cas, vous devrez utiliser un niveau d'analyse ordinal.

Même s'il n'existe aucune propriété prédéfinie de variable la transformant exclusivement en tel ou tel niveau, l'utilisateur débutant peut s'aider des règles générales suivantes. Dans la quantification nominale simple, vous ne connaissez pas en général l'ordre des catégories, mais l'analyse doit en imposer un. Si l'ordre des catégories est connu, vous devez faire appel à la quantification ordinale. Si les catégories ne peuvent pas être classées, vous pouvez utiliser la quantification nominale multiple.

Tracés de transformation

Les différents niveaux auxquels chaque variable peut être codée imposent plusieurs restrictions dans les quantifications. Les tracés de transformation illustrent la relation entre les quantifications et les catégories d'origine résultant du niveau de codage optimal sélectionné. Par exemple, un tracé de transformation linéaire est obtenu lorsqu'une variable est considérée comme numérique. Les variables considérées comme ordinales entraînent la création d'un tracé de transformation non décroissant. Les tracés de transformation de variables considérées comme nominales, en forme de U (ou l'inverse), affichent une relation quadratique. Ces variables peuvent également créer des tracés de transformation sans tendance apparente en changeant complètement l'ordre des catégories. La figure suivante représente un exemple de tracé de transformation.

Les tracés de transformation conviennent particulièrement à la définition du mode de fonctionnement du niveau de codage optimal sélectionné. Si plusieurs catégories reçoivent des quantifications similaires, la fusion de ces catégories en une seule catégorie peut être garantie. Si une variable considérée comme nominale reçoit des quantifications affichant une tendance croissante, une transformation ordinale peut également entraîner un ajustement similaire. Si cette tendance est linéaire, il peut être approprié de considérer la variable comme numérique. Toutefois, si la fusion des catégories ou la modification des niveaux de codage est garantie, l'analyse ne varie pas de façon significative.

Codes de catégorie

Soyez vigilant lorsque vous codez des variables catégorielles, car certains systèmes de codification peuvent générer des sorties indésirables ou des analyses incomplètes. Les systèmes de codification applicables à la variable *travail* sont répertoriés dans le tableau suivant.

Tableau 2. Autres systèmes de codification de la variable travail

Catégorie	A	B	C	D
Stagiaire	1	1	5	1
Représentant	2	2	6	5
Gestionnaire	3	7	7	3

Certaines procédures de catégorie exigent que la plage de valeurs de chaque variable soit définie. Toute valeur en dehors de cette plage est considérée comme manquante. La valeur minimale de catégorie est toujours égale à 1. La valeur maximale de catégorie, quant à elle, est fournie par l'utilisateur. Cette valeur ne représente pas le nombre de catégories d'une variable. Il s'agit de la valeur *maximale* de catégorie. Par exemple, dans ce tableau, le système de codage A est doté d'une catégorie maximale égale à 3, et le système de codage B, d'une valeur maximale de catégorie égale à 7. Toutefois, ces deux systèmes codent les trois mêmes catégories.

La plage de variables détermine les catégories qui sont omises de l'analyse. Les catégories ayant des codes en dehors de la plage définie sont également omises de cette analyse. Cette méthode est certes simple pour omettre des catégories, mais elle peut entraîner des analyses indésirables. Une catégorie mal définie peut omettre des catégories *valides* de l'analyse. Par exemple, pour le système de codage B, définir la valeur maximale de catégorie sur 3 signifie que la variable *travail* possède des catégories codées de 1 à 3. La catégorie de *responsable* est considérée comme manquante. Aucune catégorie n'ayant été réellement codée 3, la troisième catégorie de l'analyse ne contient aucune observation. Si vous souhaitez omettre toutes les catégories de responsable, cette analyse est tout à fait appropriée. Toutefois, si des responsables doivent être ajoutés, la catégorie maximale doit être définie sur 7 et les valeurs manquantes doivent être codées avec des valeurs supérieures à 7 ou inférieures à 1.

Pour les variables considérées comme nominales ou ordinales, la plage des catégories n'a aucune incidence sur les résultats. Pour les valeurs nominales, seul le libellé, et non la valeur qui lui est associée, est déterminant. Pour les variables ordinales, l'ordre des catégories est conservé dans les quantifications. Les valeurs de catégorie proprement dites ne sont pas importantes. Tous les systèmes de codification aboutissant au même classement des catégories auront des résultats identiques. Par exemple, les trois premiers systèmes de codage du tableau sont fonctionnellement équivalents si la variable *travail* est analysée à un niveau ordinal. L'ordre des catégories est identique dans ces systèmes. En revanche, le système de codage D inverse les deuxième et troisième catégories, et génère des résultats différents de ceux des autres systèmes.

Bien que de nombreux systèmes de codification de variable soient fonctionnellement équivalents, on leur préfère l'utilisation d'autres systèmes présentant de légères différences entre les codes, car ces derniers influent sur le nombre de sorties générées par une procédure. Toutes les catégories codées dotées de valeurs comprises entre 1 et la valeur maximale définie par l'utilisateur sont valides. Si l'une de ces catégories est vide, les quantifications correspondantes seront manquantes par défaut ou nulles, selon la procédure utilisée. Bien qu'aucune de ces affectations n'ait d'incidence sur les analyses, des sorties sont créées pour ces catégories. Par conséquent, pour le système de codage B, la variable *travail* possède quatre catégories recevant des valeurs système manquantes. Pour le système de codage C, on distingue également quatre catégories recevant des indicateurs manquants par défaut. En revanche, pour le système de codage A, il n'existe aucune quantification manquante par défaut. Utiliser des entiers consécutifs en tant que codes pour les variables traitées comme des variables nominales ou ordinales génère beaucoup moins de sorties sans affecter pour autant les autres résultats.

Les systèmes de codification des variables considérées comme numériques sont plus restreints que l'observation ordinaire. Pour ces variables, les différences entre les catégories consécutives sont significatives. Le tableau suivant répertorie trois systèmes de codification pour la variable *âge*.

Tableau 3. Autres systèmes de codification d'âge

Catégorie	A	B	C
20	20	1	1
22	22	3	2
25	25	6	3
27	27	8	4

Tout recodage des variables numériques doit conserver les différences entre les catégories. Une méthode garantissant leur conservation consiste à utiliser les valeurs d'origine. Toutefois, nombreuses sont les catégories qui risquent d'avoir au final des indicateurs manquants par défaut. Par exemple, le système de codage A emploie les valeurs observées d'origine. Pour les procédures de catégorie, à l'exception de l'Analyse des correspondances, la valeur maximale de catégorie est égale à 27, et la valeur minimale de catégorie est définie sur 1. Les 19 premières catégories sont vides et reçoivent des indicateurs manquants par défaut. Le nombre de sorties peut devenir rapidement conséquent si la catégorie maximale est nettement supérieure à 1 et qu'il existe de nombreuses catégories vides comprises entre 1 et la valeur maximale.

Pour réduire le nombre de sorties, vous pouvez procéder à un recodage. Néanmoins, pour les variables numériques, vous ne devez pas utiliser la fonction de recodage automatique. La codification appliquée aux entiers consécutifs génère des différences de 1 entre toutes les catégories consécutives et, par conséquent, l'ensemble des quantifications est espacé de la même manière. Les caractéristiques métriques jugées primordiales lorsqu'une variable est considérée comme numérique sont supprimées par l'application d'un recodage aux entiers consécutifs. Par exemple, le système de codage C du tableau correspond au recodage automatique de la variable *âge*. La différence entre les catégories 22 et 25 passe de trois à un. Les quantifications reflètent ce changement.

Un autre système de recodage conservant les différences entre les catégories consiste à déduire de chaque catégorie la plus petite valeur de la catégorie et à ajouter 1 à chaque différence. Le système B constitue l'aboutissement de cette transformation. La plus petite valeur de catégorie, 20, a été déduite de chaque catégorie, et 1 a été ajouté à chaque résultat. Les codes transformés possèdent une valeur minimale, 1, et l'ensemble des différences est identique aux données d'origine. La valeur maximale de catégorie est désormais égale à 8. En outre, les quantifications nulles précédant la première quantification non nulle sont toutes supprimées. Toutefois, les quantifications non nulles qui correspondent à chaque catégorie issue du système B sont identiques aux quantifications du système A.

Utilisation de la procédure la plus adaptée à votre application

Les méthodes contenues dans quatre de ces procédures (analyse des correspondances, analyse de correspondance multiple, analyse des correspondances multiples et analyse de corrélation canonique non linéaire) font partie du processus général d'analyse des données multivariées, appelé **réduction des dimensions**. En termes plus précis, les relations entre les variables sont représentées dans plusieurs dimensions (deux ou trois), aussi souvent que possible. Vous pouvez ainsi décrire les structures ou les motifs des relations qu'il serait trop difficile de comprendre dans leur richesse et leur complexité originales. Dans les applications d'étude de marché, ces méthodes peuvent représenter un type de **configuration perceptuelle**. Ces procédures présentent un avantage majeur : elles adaptent les données à différents niveaux de codage optimal.

La régression catégorielle décrit la relation entre une variable de réponse catégorielle et une combinaison de variables de prédicteur catégoriel. L'influence de chaque variable de prédicteur sur la variable de réponse est signalée par la pondération de régression correspondante. Comme dans les autres procédures, les données peuvent être analysées avec plusieurs niveaux de codage optimal.

Le positionnement et le dépliage multidimensionnels décrivent les relations entre les objets dans un espace de petite dimension à l'aide des proximités entre les objets.

Voici quelques règles applicables à chaque procédure :

- Utilisez la régression catégorielle pour prévoir les valeurs d'une variable dépendante catégorielle issue d'une combinaison de variables indépendantes catégorielles.
- Utilisez l'analyse des correspondances multiples pour représenter les motifs de variation d'un ensemble de variables de niveaux de codage optimal mixtes.
- Utilisez l'analyse de corrélation canonique non linéaire pour évaluer l'importance de la corrélation de plusieurs ensembles de variables de niveaux de codage optimal mixtes.
- Utilisez l'analyse des correspondances pour analyser les tableaux de contingence à deux entrées ou les données pouvant être fournies en tant que tableau à deux entrées, comme les données de préférence de marque ou de choix sociométrique.
- Utilisez l'analyse de correspondance multiple pour analyser une matrice de données multivariées catégorielles si vous souhaitez simplement que les variables soient analysées au niveau nominal.
- Utilisez le positionnement multidimensionnel pour analyser des données de proximité. L'objectif est de trouver une représentation à moindres carrés d'un seul ensemble d'objets dans un espace de petite dimension.

- Utilisez le dépliage multidimensionnel pour analyser des données de proximité. L'objectif est de trouver une représentation à moindres carrés de deux ensembles d'objets dans un espace de petite dimension.

Régression catégorielle

La régression catégorielle convient le mieux si votre analyse a pour but de prévoir une variable (de réponse) dépendante issue d'un ensemble de variables indépendantes (prédicteur). Comme pour toutes les procédures de codage optimal, des valeurs d'échelle sont attribuées à chaque catégorie de chaque variable, afin que ces valeurs soient optimales par rapport à la régression. La solution d'une régression catégorielle optimise la corrélation carrée entre la réponse transformée et la combinaison pondérée de prédicteurs transformés.

Relation avec les autres procédures de catégorie. La régression catégorielle avec codage optimal est comparable à l'analyse de corrélation canonique avec codage optimal utilisant deux ensembles, dont l'un contient uniquement la variable dépendante. Dans la dernière méthode, la similitude des ensembles est calculée par comparaison de chaque ensemble à une variable inconnue située entre tous les ensembles. Dans la régression catégorielle, la similitude de la réponse transformée et de la combinaison linéaire de prédicteurs transformés est évaluée directement.

Relation avec les méthodes standard. Dans la régression linéaire standard, les variables catégorielles peuvent être soit recodées en tant que variables indicateur, soit traitées de la même manière que les variables de niveau d'intervalle. Dans la première approche, le modèle inclut une constante et une pente différentes pour chaque combinaison de niveaux des variables catégorielles. Un grand nombre de paramètres à interpréter est ainsi généré. Dans la seconde approche, un seul paramètre est estimé pour chaque variable. Toutefois, la nature arbitraire des codages de catégorie rend toute généralisation impossible.

Si une partie des variables n'est pas continue, d'autres types d'analyse sont disponibles. Si la réponse est continue et les prédicteurs catégoriels, l'analyse de variance est généralement utilisée. Si la réponse est catégorielle et les prédicteurs continus, la régression logistique ou l'analyse discriminante peut convenir. Si la réponse et les prédicteurs sont catégoriels, les modèles log-linéaires sont généralement utilisés.

La régression avec codage optimal fournit trois niveaux de codage pour chaque variable. Les combinaisons de ces niveaux peuvent représenter des relations non linéaires très diverses auxquelles une méthode « standard » n'est pas du tout adaptée. Par conséquent, le codage optimal s'avère une solution beaucoup plus souple que les approches standard un peu plus complexes.

En outre, les transformations non linéaires des prédicteurs réduisent habituellement les dépendances des uns par rapport aux autres. Si vous comparez les valeurs propres de la matrice de corrélation des prédicteurs avec celles de la matrice de corrélation des prédicteurs codés de façon optimale, ces dernières sont généralement moins variables que les autres. En d'autres termes, dans la régression catégorielle, le codage optimal réduit les valeurs propres supérieures de la matrice de corrélation des prédicteurs et incrémente les valeurs propres inférieures.

Analyse des correspondances multiples

L'utilisation de l'analyse des correspondances multiples convient le mieux pour représenter les motifs de variation d'un ensemble de variables de niveaux de codage optimal mixtes. Cette méthode tente de réduire la dimensionnalité d'un ensemble de variables et de représenter cette variation dans la mesure du possible. Des valeurs d'échelle sont attribuées à chaque catégorie des variables afin que ces valeurs soient optimales par rapport à la solution en composantes principales. Les objets utilisés pour l'analyse reçoivent les scores des composantes basés sur les données quantifiées. Les tracés de scores des composantes révèlent les motifs figurant parmi les objets de l'analyse, ainsi que les objets inhabituels contenus dans les données. La solution d'une analyse des correspondances multiples optimise les corrélations de scores des objets avec chaque variable quantifiée pour le nombre de composantes (dimensions) indiqué.

Une fonction importante des composantes principales catégorielles consiste à vérifier les données de préférence, où les répondants classent ou évaluent un nombre d'éléments par rapport à la préférence. Dans la configuration habituelle des données IBM SPSS Statistics, les lignes correspondent aux individus, les colonnes, aux mesures des éléments et les scores figurant sur les lignes, aux scores de préférence (sur une échelle de 0 à 10, par exemple), ce qui rend les données dépendantes des lignes. Pour les données de préférence, vous pouvez considérer les individus comme des variables. Grâce à la procédure de transposition, vous pouvez transposer ces données. Les indicateurs sont les variables et toutes les variables sont déclarées ordinales. Si vous le souhaitez, vous pouvez utiliser plus de variables que d'objets pour la procédure CATPCA.

Relation avec les autres procédures de catégorie. Si toutes les variables sont déclarées nominales multiples, l'analyse des correspondances multiples génère une analyse équivalant à une analyse de correspondance multiple exécutée sur les mêmes variables. Par conséquent, l'analyse des correspondances multiples peut être considérée comme un type d'analyse de correspondance multiple dans lequel certaines variables sont déclarées ordinales ou numériques.

Relation avec les méthodes standard. Si toutes les variables sont codées au niveau numérique, l'analyse des correspondances multiples équivaut à l'analyse en composantes principales standard.

Plus généralement, l'analyse des correspondances multiples représente un autre moyen de calculer les corrélations entre les échelles non numériques, et de leur appliquer une analyse factorielle ou en composantes principales standard. Toute utilisation simpliste du coefficient de corrélation de Pearson habituel comme mesure d'association de données ordinales peut avoir une incidence significative sur l'estimation des corrélations.

Analyse de corrélation canonique non linéaire

L'Analyse de corrélation canonique non linéaire est une procédure très générale comportant de nombreuses tâches. Ce type d'analyse a pour but d'analyser les relations entre plusieurs ensembles de variables, au lieu des variables proprement dites, comme dans l'analyse en composantes principales. Par exemple, vous pouvez utiliser deux ensembles de variables : l'un peut inclure des éléments d'ordre démographique concernant un groupe de répondants, alors que l'autre peut contenir les réponses à un ensemble d'éléments d'attitude. Les niveaux de codage de l'analyse peuvent représenter une combinaison de niveaux nominal, ordinal et numérique. L'analyse de corrélation canonique avec codage optimal détermine la similitude entre les ensembles en comparant simultanément les variables canoniques de chaque ensemble à un groupe de scores de compromis associé aux objets.

Relation avec les autres procédures de catégorie. Si plusieurs ensembles de variables contiennent chacun une seule variable, l'analyse de corrélation canonique avec codage optimal équivaut à l'analyse en composantes principales avec codage optimal. Si toutes les variables d'une analyse de type « une variable par ensemble » sont nominales multiples, l'analyse de corrélation canonique avec codage optimal équivaut à l'analyse de correspondance multiple. Dans le cas de deux ensembles de variables, dont l'un comprend une seule variable, l'analyse de corrélation canonique avec codage optimal équivaut à la régression catégorielle avec codage optimal.

Relation avec les méthodes standard. L'analyse de corrélation canonique standard est une méthode statistique qui recherche une combinaison linéaire d'un premier ensemble de variables et celle d'un second ensemble de variables corrélées de façon optimale. Du fait de ces combinaisons linéaires, l'analyse de corrélation canonique peut rechercher les ensembles indépendants de combinaisons linéaires suivants, appelés variables canoniques. Le nombre maximal d'ensembles doit être égal au nombre de variables contenues dans le plus petit ensemble.

Si deux ensembles de variables sont utilisés dans l'analyse et toutes les variables définies comme étant numériques, l'analyse de corrélation canonique avec codage optimal équivaut à une analyse de

corrélation canonique standard. Bien qu'IBM SPSS Statistics ne propose aucune procédure d'analyse de corrélation canonique, vous pouvez obtenir une bonne partie des statistiques concernées par le biais de l'analyse de variance multivariée.

L'analyse de corrélation canonique avec codage optimal fournit de nombreuses fonctions. Si vous utilisez deux ensembles de variables et que l'un d'eux contient une variable nominale déclarée nominale simple, les résultats de l'analyse de corrélation canonique avec codage optimal peuvent être interprétés d'une manière similaire à ceux de l'analyse de régression. Si vous considérez que cette variable est nominale multiple, l'analyse avec codage multiple constitue une alternative à l'analyse discriminante. Regrouper les variables dans plus de deux ensembles vous permet d'analyser les données de différentes manières.

Analyse des correspondances

L'analyse des correspondances a pour but de créer des tracés doubles pour les tables de correspondances. Dans une table de correspondances, les variables de ligne et de colonne sont supposées représenter les catégories non classées. Par conséquent, le niveau de codage optimal nominal est systématiquement utilisé. Seules les données nominales sont recherchées dans ces deux types de variable. Il s'agit en réalité de tenir compte du fait que certains objets se trouvent dans la même catégorie, alors que ce n'est pas le cas pour d'autres. Aucune hypothèse n'est avancée concernant la distance ou l'ordre entre les catégories de la même variable.

L'analyse des correspondances peut notamment servir à analyser les tableaux de contingence à deux entrées. Si un tableau possède r lignes actives et c colonnes actives, le nombre de dimensions de la solution d'analyse des correspondances correspond au nombre minimal de r moins 1 ou de c moins 1, selon la valeur la plus faible. En d'autres termes, vous pouvez parfaitement représenter les catégories de ligne ou de colonne d'un tableau de contingence dans un espace de dimensions. En pratique, vous pouvez néanmoins représenter les catégories de ligne et de colonne d'un tableau à deux entrées dans un espace comportant peu de dimensions, plus précisément deux, pour la simple raison que la compréhension des tracés bidimensionnels est bien plus facile que celle des représentations spatiales multidimensionnelles.

Lorsqu'un nombre de dimensions inférieur au nombre maximal de dimensions possibles est utilisé, les statistiques créées lors de l'analyse décrivent la manière dont les catégories de ligne et de colonne sont reproduites dans la représentation comportant peu de dimensions. Si la qualité de la représentation de la solution bidimensionnelle est satisfaisante, vous pouvez vérifier les tracés des points de ligne et de colonne pour déterminer les catégories similaires de la variable de ligne et de la variable de colonne, et les catégories de ligne et de colonne similaires les unes aux autres.

Relation avec les autres procédures de catégorie. L'analyse simple des correspondances se limite aux tableaux à deux entrées. Si plusieurs variables vous intéressent, vous pouvez en combiner certaines pour créer des variables d'interaction. Par exemple, pour les variables *région*, *travail* et *âge*, vous pouvez combiner *région* et *travail* afin de créer une variable *rétrav* possédant les 12 catégories répertoriées dans le tableau suivant. Cette variable crée un tableau à deux entrées avec la variable *âge* (12 lignes, 4 colonnes), qui peut faire l'objet d'une analyse de correspondances.

Tableau 4. Combinaisons des variables *région* et *travail*

Code de catégorie	Définition de catégorie	Code de catégorie	Définition de catégorie
1	Nord, stagiaire	7	Est, stagiaire
2	Nord, représentant	8	Est, représentant
3	Nord, responsable	9	Est, responsable
4	Sud, stagiaire	10	Ouest, stagiaire
5	Sud, représentant	11	Ouest, représentant
6	Sud, responsable	12	Ouest, responsable

Cette approche présente un défaut, à savoir que toute paire de variables peut être combinée. Nous pouvons combiner *travail* et *âge*, et ainsi obtenir une autre variable de 12 catégories. Nous pouvons également combiner *région* et *âge*, ce qui entraîne la création d'une variable de 16 catégories. Chacune de ces variables d'interaction génère un tableau à deux entrées avec l'autre variable. Les analyses des correspondances de ces trois tableaux donnent des résultats différents, même si chaque résultat est valide. En outre, dans le cas de quatre variables au moins, vous pouvez créer des tableaux à deux entrées comparant une variable d'interaction avec une autre. Le nombre de tableaux possibles à analyser peut devenir très important, même pour quelques variables seulement. Vous pouvez combiner l'un de ces tableaux pour l'analyse ou les analyser tous. Vous pouvez également utiliser la procédure d'analyse de correspondance multiple pour vérifier toutes les variables à la fois sans avoir à créer de variables d'interaction.

Relation avec les méthodes standard. En outre, la procédure de tableau croisé permet d'analyser les tableaux de contingence, avec l'indépendance comme valeur commune aux différentes analyses. Toutefois, même dans les petits tableaux, déterminer l'origine d'un départ à partir de la valeur d'indépendance peut s'avérer complexe. L'analyse des correspondances est utile car elle analyse ces motifs pour les tableaux à deux entrées, quelle que soit leur taille. En cas d'association entre les variables de ligne et de colonne (c'est-à-dire si la valeur khi-deux est significative), l'analyse des correspondances peut contribuer à révéler la nature de la relation.

Analyse de correspondance multiple

L'Analyse de correspondance multiple tente de créer une solution dans laquelle les objets faisant partie de la même catégorie sont tracés proches les uns des autres, et les objets de catégories différentes, éloignés les uns des autres. Chaque objet se trouve aussi près que possible des points de catégorie qui s'appliquent. Ainsi, les catégories divisent les objets en sous-groupes homogènes. Les variables sont considérées comme homogènes lorsqu'elles classent les objets des mêmes catégories dans les mêmes sous-groupes.

Pour une solution en une dimension, l'analyse de correspondance multiple attribue des valeurs d'échelle optimale (quantifications de catégorie) à chaque catégorie de chaque variable si bien que, dans l'ensemble, les catégories ont en moyenne une étendue maximale. Pour une solution en deux dimensions, l'analyse de correspondance multiple recherche un second ensemble de quantifications des catégories de chaque variable non lié au premier ensemble, en réessayant d'optimiser l'étendue, etc. Les catégories recevant autant d'évaluations qu'il existe de dimensions, les variables de l'analyse sont censées être nominales multiples au niveau de codage optimal.

L'analyse de correspondance multiple affecte également des scores aux objets de l'analyse, afin que les quantifications de catégorie représentent les moyennes, ou centroïdes, des scores des objets de la catégorie.

Relation avec les autres procédures de catégorie. L'analyse de correspondance multiple est également appelée analyse d'homogénéité ou double codage. Elle fournit des résultats, certes comparables mais pas identiques, à ceux de l'analyse des correspondances lorsque seules deux variables sont utilisées. L'analyse des correspondances génère des sorties uniques récapitulant l'ajustement et la qualité de la représentation de la solution, y compris les informations de stabilité. Par conséquent, dans le cas de deux variables, il vaut mieux généralement préférer l'analyse des correspondances à l'analyse de correspondance multiple. Ces deux procédures présentent une autre différence : l'entrée de l'analyse de correspondance multiple est une matrice de données, dans laquelle les lignes sont des objets et les colonnes sont des variables. Quant au point de départ de l'analyse des correspondances, il peut être la même matrice de données, une matrice de proximité générale ou un tableau de contingence joint, qui est une matrice récapitulative où les lignes et les colonnes représentent des catégories de variables. L'analyse de correspondance multiple peut également être assimilée à l'analyse en composantes principales de données codées au niveau nominal multiple.

Relation avec les méthodes standard. L'analyse de correspondance multiple peut être considérée comme étant l'analyse d'un tableau de contingence à entrées multiples. Un tableau de contingence à entrées multiples peut également être analysé avec la procédure de tableaux croisés, mais celle-ci fournit des statistiques récapitulatives distinctes pour chaque catégorie de chaque variable de contrôle. Grâce à l'analyse de correspondance multiple, il est généralement possible de récapituler la relation entre toutes les variables à l'aide d'un tracé bidimensionnel. Un mode d'utilisation avancé de ce type d'analyse consiste à remplacer les valeurs de catégorie d'origine par les valeurs d'échelle optimale de la première dimension, puis à effectuer une analyse multivariée secondaire. Puisque l'analyse de correspondance multiple remplace les libellés de catégorie par des valeurs d'échelle numérique, de nombreuses procédures nécessitant des données numériques peuvent être appliquées lorsqu'elle est terminée. Par exemple, la procédure d'analyse factorielle crée une première composante principale équivalant à la première dimension de l'analyse de correspondance multiple. Les scores des composantes de la première dimension sont identiques à ceux des objets et les chargements entre composantes, aux mesures de discrimination. Néanmoins, la deuxième dimension de l'analyse de correspondance multiple est différente de celle de l'analyse factorielle.

Positionnement multidimensionnel

Le positionnement multidimensionnel convient le mieux si votre analyse a pour but de rechercher une structure dans un ensemble de mesures de distance entre un ensemble d'objets ou d'observations unique. Pour cela, il affecte les observations à des positions particulières dans un espace conceptuel de petite dimension de telle sorte que les distances entre les points dans l'espace correspondent le mieux possible aux dissimilarités données. Le résultat est une représentation à moindres carrés des objets dans cet espace de petite dimension, qui vous aidera, dans certains cas, à mieux comprendre vos données.

Relation avec les autres procédures de catégorie. Lorsque vous utilisez des données multivariées à partir desquelles vous créez des distances et que vous analysez ensuite avec le positionnement multidimensionnel, les résultats s'avèrent similaires à ceux de l'analyse des données utilisant une analyse des correspondances multiples, impliquant la standardisation principale des objets. Ce type d'analyse en composantes principales est également appelé analyse des coordonnées principales.

Relation avec les méthodes standard. La procédure de positionnement multidimensionnel qualitatif (PROXSCAL) apporte des améliorations à la procédure de codage disponible dans l'option Statistiques de base (ALSCAL). PROXSCAL fournit un algorithme accéléré pour certains modèles et vous permet d'appliquer des restrictions à l'espace commun. En outre, PROXSCAL tente de minimiser la contrainte brute normalisée plutôt que le S-stress (également appelé **pression**). En général, on dénote une certaine préférence pour la contrainte brute normalisée, car cette mesure est basée sur les distances, alors que le S-stress est basé sur leur carré.

Dépliage multidimensionnel

Le Dépliage multidimensionnel convient mieux si votre analyse a pour but de rechercher une structure dans un ensemble de mesures de distance entre deux ensembles d'objets (appelés objets de ligne et de colonne). Pour cela, il affecte les observations à des positions particulières dans un espace conceptuel de petite dimension de telle sorte que les distances entre les points dans l'espace correspondent le mieux possible aux dissimilarités données. Le résultat est une représentation à moindres carrés des objets de ligne et de colonne dans cet espace de petite dimension, qui vous aidera, dans certains cas, à mieux comprendre vos données.

Relation avec les autres procédures de catégorie. Si vos données sont constituées de distances entre un ensemble unique d'objets (une matrice carrée, symétrique), utilisez Positionnement multidimensionnel.

Relation avec les méthodes standard. La procédure de dépliage multidimensionnel des catégories (PREFSCAL) apporte des améliorations à la fonctionnalité de dépliage disponible dans l'option Statistiques de base (avec ALSCAL). PREFSCAL vous permet d'instaurer des restrictions sur l'espace commun. En outre, PREFSCAL tente de minimiser une mesure de contrainte pénalisée, l'aidant ainsi à éviter de dégénérer des solutions (auxquels les algorithmes précédents sont enclins).

Rapport d'aspect des graphiques de codage optimal

Le rapport d'aspect des tracés de codage optimal est isotrope. Dans un tracé bidimensionnel, la distance représentant une unité de la dimension 1 est égale à celle représentant une unité de la dimension 2. Si, dans ce type de tracé, vous modifiez l'étendue d'une dimension, le système modifie la taille de l'autre dimension pour que les distances physiques restent égales. Il est impossible de remplacer un rapport d'aspect isotrope pour les procédures de codage optimal.

Chapitre 2. Régression catégorielle (CATREG)

La **régression catégorielle** quantifie les données catégorielles en affectant des valeurs numériques aux catégories ; une équation de régression linéaire optimale est ainsi créée pour les variables transformées. La régression catégorielle est également appelée CATREG, acronyme de *categorical regression*.

L'analyse de la régression linéaire standard implique la réduction des différences de sommes des carrés entre une variable de réponse (dépendante) et une combinaison pondérée des prédicteurs (variables indépendantes). Les variables sont habituellement quantitatives, les données catégorielles nominales étant recodées en variables binaires ou de contraste. En conséquence, les variables catégorielles servent à séparer les groupes d'observations et cette technique estime des séries de paramètres distinctes pour chaque groupe. Les coefficients estimés reflètent le mode d'affectation de la réponse due aux modifications des prédicteurs. Il est possible de prévoir la réponse pour n'importe quelle combinaison de valeurs de prédicteur.

Une autre approche consiste à effectuer la régression de la réponse sur les valeurs de prédicteur catégoriel proprement dites. Dans ce cas, un seul coefficient est estimé pour chaque variable. Toutefois, pour les variables catégorielles, les valeurs des catégories sont arbitraires. La codification des catégories selon plusieurs méthodes produit différents coefficients, ce qui complique les comparaisons d'analyses portant sur les mêmes variables.

CATREG constitue une extension de l'approche standard en codant simultanément les variables catégorielles, ordinales et numériques. Cette procédure quantifie les variables catégorielles afin que les valeurs affectées reflètent les caractéristiques des catégories d'origine. La procédure traite les variables catégorielles quantifiées de la même façon que les variables numériques. L'utilisation de transformations non linéaires permet d'analyser les variables à différents niveaux afin de déterminer le modèle correspondant au meilleur ajustement possible.

Exemple : La régression catégorielle peut être utilisée pour décrire dans quelle mesure la satisfaction professionnelle dépend de la catégorie d'emploi, de la région et de la durée du transport. Vous pourriez ainsi déterminer que les plus hauts niveaux de satisfaction professionnelle correspondent aux postes de direction et aux temps de transport les plus faibles. Vous avez ainsi la possibilité d'utiliser l'équation de régression résultante pour prévoir la satisfaction professionnelle relative à n'importe quelle combinaison de ces trois variables indépendantes.

Tracés et statistiques : Fréquences, coefficients de régression, tableau ANOVA, historique des itérations, valeurs affectées aux catégories, corrélations entre prédicteurs non transformés, corrélations entre prédicteurs transformés, les tracés résiduels et de transformation.

Considérations sur les données de la régression catégorielle

Données : CATREG traite les variables indicateur de catégories. Les indicateurs de catégories doivent être des nombres entiers positifs. Vous pouvez utiliser la boîte de dialogue Discretisation pour convertir les variables fractionnées et les variables de chaîne en nombres entiers positifs.

Hypothèses : Une seule variable de réponse est autorisée, mais le nombre maximal de variables de prédicteur est de 200. Les données doivent comporter au moins trois observations valides, le nombre d'observations valides ne devant pas dépasser le nombre de variables de prédicteur plus un.

Procédures apparentées : La procédure CATREG équivaut à la procédure d'analyse de corrélation canonique nominale avec codage optimal (OVERALS) avec deux groupes, dont l'un ne comporte qu'une seule variable. Le codage de toutes les variables au niveau numérique correspond à l'analyse de régression multiple standard.

Pour obtenir une régression catégorielle

1. A partir des menus, sélectionnez :
Analyse > Régression > Codage optimal (CATREG)...
2. Sélectionnez la variable dépendante ainsi que la ou les variables indépendantes.
3. Cliquez sur **OK**.

Sinon, modifiez le niveau de codage de chaque variable.

Définir une échelle dans la régression catégorielle

Vous pouvez définir le niveau de codage optimal des variables dépendantes et indépendantes. Par défaut, elles sont codées comme des splines monotones de second degré (ordinales) avec deux noeuds intérieurs. En outre, vous pouvez également définir la pondération pour les variables d'analyse.

Niveau du codage optimal : Vous pouvez également sélectionner le niveau de codage pour la quantification de chaque variable.

- **Spline ordinale :** L'ordre des catégories de la variable observée est conservé dans la variable codée de façon optimale. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. La transformation résultante est un modèle polynomial monotone lissé du degré choisi. Ses différents éléments dépendent du nombre de noeuds intérieurs défini par l'utilisateur ainsi que du positionnement de ces derniers, déterminé par la procédure.
- **Spline nominale :** La seule information de la variable observée qui est conservée dans la variable codée de façon optimale est le groupe des objets dans les catégories. L'ordre des catégories de la variable observée n'est pas conservé. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. La transformation résultante est un modèle polynomial lissé, peut-être non monotone, du degré choisi. Ses différents éléments dépendent du nombre de noeuds intérieurs défini par l'utilisateur ainsi que du positionnement de ces derniers, déterminé par la procédure.
- **Ordinal :** L'ordre des catégories de la variable observée est conservé dans la variable codée de façon optimale. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. La transformation du résultat convient mieux que la transformation ordinale spline, mais s'avère moins lissée.
- **Nominal :** La seule information de la variable observée qui est conservée dans la variable codée de façon optimale est le groupe des objets dans les catégories. L'ordre des catégories de la variable observée n'est pas conservé. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. La transformation du résultat convient mieux que la transformation nominale spline mais s'avère moins lissée.
- **Numérique :** Les catégories sont considérées comme triées et espacées régulièrement (niveau d'intervalle). L'ordre des catégories ainsi que les distances égales entre les nombres de catégories de la variable sont conservées dans la variable codée de façon optimale. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. Lorsque toutes les variables sont au niveau numérique, l'analyse est analogue à celle en composantes principales standard.

Régression catégorielle : Discrétisation

La boîte de dialogue Discrétisation vous permet de choisir une méthode de recodage des variables. Les valeurs fractionnées sont regroupées en sept catégories (ou en nombre de valeurs distinctes de variables si le nombre est inférieur à sept) avec une distribution normale approximative, à moins qu'une autre configuration ne soit spécifiée. Les variables de chaîne sont toujours converties en nombres entiers positifs en affectant des indicateurs de catégories selon l'ordre croissant alphanumérique. La discrétisation des variables de chaîne s'applique à ces nombres entiers. Par défaut, d'autres variables sont laissées inutilisées. Les variables discrétisées sont ensuite utilisées dans l'analyse.

Méthode : Choisissez entre Regroupement, Classement et Multiplier.

- **Regroupement** : Recodez en un nombre spécifié de catégories ou par intervalle.
- **Classement** : La variable est discrétisée via le classement des observations.
- **Multiplier** : Les valeurs courantes de la variable sont standardisées, multipliées par 10 et arrondies, et possèdent une constante ajoutée de sorte que la valeur discrétisée la plus faible soit égale à 1.

Regroupement : Les options suivantes sont disponibles lorsque vous discrétisez des variables par groupe :

- **Nombre de catégories** : Indiquez un nombre de catégories et définissez si les valeurs de la variable doivent faire l'objet d'une distribution approximativement gaussienne ou uniforme entre ces catégories.
- **Intervalles égaux** : Les variables sont recodées en catégories définies par ces intervalles de taille égale. N'oubliez pas de spécifier la longueur des intervalles.

Régression catégorielle : Valeurs manquantes

La boîte de dialogue Valeurs manquantes vous permet de choisir la stratégie de gestion des valeurs manquantes pour les variables de l'analyse et supplémentaires.

Stratégie : Vous pouvez exclure des objets contenant des valeurs manquantes (suppression par liste) ou inclure des valeurs manquantes (traitement actif).

- **Exclure les objets contenant des valeurs manquantes sur cette variable** : Les objets contenant des valeurs manquantes dans la variable sélectionnée sont retirés de l'analyse. Cette option n'est pas disponible pour les variables supplémentaires.
- **Inclure les valeurs manquantes** : Des valeurs sont prises en compte pour les objets contenant des valeurs manquantes sur la variable sélectionnée. Vous pouvez choisir la méthode de calcul : Sélectionnez **Mode** pour remplacer les valeurs manquantes par la catégorie la plus fréquente. S'il existe plusieurs modes, utilisez celui dont l'indicateur de catégories est le plus petit. Sélectionnez **Catégorie supplémentaire** pour remplacer les valeurs manquantes par la valeur affectée à une catégorie supplémentaire. Cela suppose que les objets contenant une valeur manquante pour cette variable sont considérés comme appartenant à la même catégorie (supplémentaire).

Régression catégorielle : Options

La boîte de dialogue Options permet de sélectionner le style de configuration initiale, de spécifier les critères d'itération et de convergence, de sélectionner les objets supplémentaires et de définir le libellé des tracés.

Objets supplémentaires : Cette option permet de définir les objets à traiter comme objets supplémentaires. Entrez simplement le numéro d'un objet supplémentaire (ou spécifiez une plage d'observations), puis cliquez sur **Ajouter**. Vous ne pouvez pas pondérer des objets supplémentaires (les pondérations indiquées sont ignorées).

Configuration initiale : Si aucune variable n'est considérée comme nominale, sélectionnez la configuration **Numérique**. Si une variable au moins est considérée comme nominale, choisissez la configuration **Aléatoire**.

Si au moins une variable a un niveau d'échelle ordinal ou Spline ordinal, l'algorithme habituel pour les modèles peut également générer une solution moins optimale. Choisir les **Départs multiples systématiques** avec tous les motifs de signes possibles permettra toujours de trouver la solution optimale, mais la durée d'exécution requise augmente rapidement en même temps que le nombre de variables ordinales et Spline ordinales dans le jeu de données. Vous pouvez réduire le nombre de motifs de test en spécifiant un pourcentage de perte de seuil de variance, pour lequel plus le seuil est élevé, plus le nombre de motifs de signes exclus augmente. Cette option ne permet pas de garantir l'obtention de la solution optimale, mais elle réduit le risque d'obtenir une solution moins optimale. De plus, si la solution optimale n'est pas trouvée, il y a moins de chances que la solution moins optimale soit très différente de

la solution optimale. Lorsque des départs multiples systématiques sont demandés, les signes des coefficients de régression pour chaque départ sont écrits dans un fichier de données IBM SPSS Statistics externe ou dans un jeu de données de la session en cours. Pour plus d'informations, voir «Régression catégorielle : Enregistrement», à la page 17.

Les résultats d'une exécution précédente avec départs multiples systématiques vous permettent d' **Utiliser des signes fixes pour les coefficients de régression**. Les signes (indiqués par 1 et -1) doivent se trouver dans une ligne du jeu de données ou du fichier spécifiés. Le chiffre de départ à valeur entière est le numéro d'observation de la ligne de ce fichier qui contient les signes à utiliser.

Critères : Vous pouvez spécifier le nombre maximal d'itérations que la régression peut prendre en charge dans ses calculs. Vous avez également la possibilité de sélectionner une valeur de critère de convergence. La régression interrompt son itération dès que la différence du total ajusté entre les deux dernières itérations est inférieur à la valeur de la convergence, ou dès que le nombre maximal d'itérations est atteint.

Libeller les tracés par : Vous permet de préciser si les libellés de variable et de valeurs ou les noms ou valeurs de variables sont utilisés dans les tracés. Vous pouvez également spécifier une longueur maximale pour les libellés.

Régularisation de régression catégorielle

Méthode : Les méthodes de régularisation peuvent améliorer l'erreur de prédiction du modèle en réduisant la variabilité des estimations du coefficient de régression à l'aide d'une réduction des estimations tendant vers 0. Le Lasso et Elastic Net réduiront certaines estimations de coefficient à 0 exactement, permettant ainsi une forme de sélection de variables. Lorsqu'une méthode de régularisation est demandée, le modèle et les coefficients régularisés pour chaque valeur de coefficient de pénalité sont écrits dans un fichier de données IBM SPSS Statistics externe ou un jeu de données de la session en cours. Pour plus d'informations, voir «Régression catégorielle : Enregistrement», à la page 17.

- **Régression de crête** : La régression de crête réduit les coefficients en introduisant un terme de pénalité égal à la somme des coefficients au carré multipliée par un **coefficient de pénalité**. Ce coefficient peut être compris entre 0 (aucune pénalité) et 1 ; cette procédure recherchera la "meilleure" valeur de pénalité si vous spécifiez une plage et un incrément.
- **Lasso** : Le terme de pénalité de Lasso est basé sur la somme des coefficients absolus et la spécification d'un coefficient de pénalité est semblable à celle d'une régression pseudo-orthogonale. Néanmoins, le Lasso nécessite beaucoup plus de calculs.
- **Elastic net** : Elastic Net regroupe simplement les pénalités de régression Lasso et de crête et effectuera une recherche dans la grille des valeurs spécifiées pour trouver les "meilleurs" coefficients de pénalité de régression Lasso et de crête. Pour une paire de pénalités de régression Lasso et de crête donnée, Elastic Net ne nécessite pas plus de calculs que le Lasso.

Afficher les tracés de régularisation : Il s'agit de tracés comparant les coefficients de régression et la pénalité de régularisation. Pendant que ce graphique recherche une plage de valeurs pour le "meilleur" coefficient de pénalité, il affiche les modifications des coefficients de régression dans cette plage.

Tracés Elastic Net : Pour la méthode Elastic Net, des tracés de régularisation séparés sont générés par les valeurs de la pénalité de régression de crête. La fonction **Tous les tracés possibles** utilise chaque valeur de la plage déterminée par les valeurs minimum et maximum de pénalité de régression de crête spécifiées. La fonction **Pour certaines pénalités de crête** permet de spécifier un sous-ensemble des valeurs dans la plage déterminée par le minimum et le maximum. Entrez simplement le chiffre d'une valeur de pénalité (ou spécifiez une plage de valeurs), puis cliquez sur **Ajouter**.

Régression catégorielle : Sortie

La boîte de dialogue Sortie permet de sélectionner les statistiques à afficher dans la sortie.

Tableaux : Génère des tableaux pour :

- **R multiple :** Comprend R^2 et R^2 ajusté, R^2 ajusté prend en compte le codage optimal.
- **ANOVA :** Cette option présente les sommes des carrés de régression et des résidus, la moyenne des carrés ainsi qu'un test-F. Deux tableaux ANOVA sont affichés : l'un avec des degrés de liberté pour la régression égaux au nombre de variables de prédicteur, et l'autre avec les degrés de liberté pour la régression prenant en compte le codage optimal.
- **Coefficients :** Cette option propose trois tableaux : un tableau Coefficients qui comporte des bêtas, l'erreur standard des bêtas, des valeurs t et la signification ; un tableau Coefficients : Codage optimal qui contient l'erreur standard des bêtas prenant en compte les degrés de liberté du codage optimal et un tableau des corrélations simples et partielles qui comporte les mesures d'importance relative de Pratt pour les prédicteurs transformés, ainsi que la tolérance avant et après transformation.
- **Historique des itérations :** Pour chaque itération, y comprises les valeurs de départ de l'algorithme, le R multiple et l'erreur de régression apparaissent. L'augmentation dans le R multiple est répertoriée en commençant à partir de la première itération.
- **Corrélations des variables d'origine :** Une matrice affichant les corrélations entre les variables sans transformation apparaît.
- **Corrélations des variables transformées :** Une matrice affichant les corrélations entre les variables transformées apparaît.
- **Modèles et coefficients régularisés :** Affiche les valeurs de pénalité, le R-deux et les coefficients de régression pour chaque modèle régularisé. Si une méthode de rééchantillonnage est spécifiée ou si des objets supplémentaires (observations de test) sont spécifiés, l'erreur de prévision ou la MSE de test sont également affichées.

Rééchantillonnage : Les méthodes de rééchantillonnage offrent une estimation de l'erreur de prédiction du modèle.

- **Validation croisée :** La validation croisée divise l'échantillon en plusieurs sous-échantillons ou niveaux. Les modèles de régression catégorielle sont générés en excluant à tour de rôle les données de chaque sous-échantillon. Le premier modèle est basé sur toutes les observations à l'exception de celles du premier sous-échantillon, le deuxième modèle est basé sur toutes les observations à l'exception de celles du deuxième sous-échantillon, etc. Pour chaque modèle, l'erreur de prédiction est estimée en appliquant le modèle au sous-échantillon exclu lors de sa génération.
- **Bootstrap .632 :** Avec le bootstrap, les observations sont extraites aléatoirement à partir des données avec remplacement. Ce processus se répète autant de fois que nécessaire pour obtenir un nombre d'échantillons du bootstrap. Un modèle est ajusté pour chaque échantillon de bootstrap. L'erreur de prévision pour chaque modèle est estimée en appliquant le modèle ajusté aux observations ne faisant pas partie de l'échantillon de bootstrap.

Valeurs affectées aux catégories : Les tableaux des valeurs transformées des variables sélectionnées apparaissent.

Statistiques descriptives : Les tableaux affichant les fréquences, les valeurs manquantes et les modes des variables sélectionnées apparaissent.

Régression catégorielle : Enregistrement

La boîte de dialogue Enregistrer vous permet d'enregistrer des prévisions, des résidus et des valeurs transformées dans le jeu de données actif et/ou d'enregistrer les données discrétisées, les valeurs transformées, les modèles et coefficients régularisés ainsi que les signes des coefficients de régression dans un fichier de données externe IBM SPSS Statistics ou un jeu de données de la session en cours.

- 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.
- Les noms de fichiers ou les noms du jeu de données doivent être différents pour chaque type de données enregistrées.

Les modèles et coefficients régularisés sont enregistrés chaque fois qu'une méthode de régularisation est sélectionnée dans la boîte de dialogue Régularisation. Par défaut, la procédure crée un nouveau jeu de données avec un nom unique, mais vous pouvez spécifier le nom de votre choix ou écrire dans un fichier externe.

Les signes de coefficients de régression sont enregistrés chaque fois que des départs multiples systématiques sont utilisés comme configuration initiale dans la boîte de dialogue Options. Par défaut, la procédure crée un nouveau jeu de données avec un nom unique, mais vous pouvez spécifier le nom de votre choix ou écrire dans un fichier externe.

Régression catégorielle des tracés de transformation

La boîte de dialogue Tracés vous permet de définir les variables qui produiront des tracés résiduels et de transformation.

Tracés de transformation : Pour chacune de ces variables, les valeurs affectées aux catégories sont tracées par rapport aux valeurs des catégories d'origine. Les catégories vides apparaissent sur l'axe horizontal mais n'affectent pas les calculs. Ces catégories sont identifiées par des ruptures dans la courbe reliant les valeurs affectées.

Tracés résiduels : Pour chacune de ces variables, les résidus (calculés pour la variable dépendante à partir de toutes les variables de prédicteur exceptée la variable de prédicteur en question) sont appliqués aux indicateurs de catégories et aux valeurs affectées aux catégories optimales multipliées par bêta par rapport aux indicateurs de catégories.

Fonctions supplémentaires de la commande CATREG

Vous pouvez personnaliser la régression catégorielle en collant vos sélections dans une fenêtre de syntaxe et en modifiant la syntaxe de commande CATREG. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Spécifier les noms de racine des variables transformées lorsque vous les enregistrez dans le jeu de données actif (avec la sous-commande SAVE).

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

Chapitre 3. Analyse des correspondances multiples (CATPCA)

Cette procédure quantifie simultanément des variables catégorielles en réduisant la dimensionnalité des données. L'analyse des correspondances multiples est également appelée CATPCA, acronyme de CATegorical Principal Components Analysis.

Le but d'une telle analyse est de réduire un groupe original de variables en un groupe plus petit de composantes non corrélées représentant la plupart des informations rencontrées dans les variables d'origine. Cette technique est d'une grande utilité lorsqu'un grand nombre de variables empêche d'interpréter efficacement les relations entre les objets (sous-objets et unités). En réduisant la dimensionnalité, vous pouvez interpréter plusieurs composantes et non plus un grand nombre de variables.

L'analyse en composantes principales standard comporte des relations linéaires entre les variables numériques. D'un autre côté, l'approche du codage optimal permet aux variables d'être codées à différents niveaux. Les variables catégorielles sont quantifiées de façon optimale par rapport à la dimensionnalité spécifiée. En conséquence, des relations non linéaires entre les variables peuvent être spécifiées.

Exemple : L'analyse des correspondances multiples peut être utilisée afin de représenter sur un graphique les relations entre la catégorie d'emploi, la division d'emploi, la région, le temps de transport (élevé, moyen ou faible), et la satisfaction professionnelle. Vous constatez peut-être que deux dimensions représentent une part importante de la variance. La première dimension peut séparer les catégories d'emploi par région, alors que la seconde sépare les divisions d'emploi en fonction du temps de transport. Notez également que la satisfaction professionnelle est liée au temps moyen de transport.

Tracés et statistiques : Effectifs, valeurs manquantes, niveau de codage optimal, mode, variance représentée par les coordonnées du centroïde, coordonnées vectorielles, total par variable et par dimension, chargements entre composantes et variables initiales pour variables quantifiées par vecteur, valeurs affectées aux catégories et coordonnées, historique des itérations, corrélations des variables transformées et des valeurs propres de la matrice de corrélation, corrélations des variables d'origine et des valeurs propres de la matrice de corrélation, scores des objets, tracés de catégories, tracés de catégories joints, tracés de transformation, tracés résiduels, tracés de centroïdes projetés, tracés d'objets, tracés doubles, tracés triples et tracés des chargements entre composantes et variables initiales.

Remarques sur les données de l'analyse des correspondances multiples (CATPCA)

Données : Les variables de chaîne sont toujours converties en nombres entiers positifs par ordre croissant alphanumérique. Les valeurs manquantes définies par l'utilisateur, les valeurs système manquantes et les valeurs inférieures à 1 sont considérées comme manquantes ; vous pouvez donc recoder ou ajouter une constante aux variables contenant des valeurs inférieures à 1 pour les définir comme non manquantes.

Hypothèses : Les données doivent contenir au moins trois observations valides. L'analyse repose sur des données sous forme de nombres entiers positifs. L'option de discrétisation classe automatiquement une variable fractionnée en regroupant ses valeurs en catégories avec une distribution "normale" et convertit automatiquement les valeurs des variables de chaîne en nombre entiers positifs. Vous pouvez en outre, spécifier d'autres schémas de discrétisation.

Procédures apparentées : Le codage de toutes les variables au niveau numérique correspond à l'analyse en composantes principales standard. Les fonctions de traçage alterné sont disponibles en utilisant les variables transformées dans une analyse en composantes principales linéaires standard. Si toutes les variables possèdent des niveaux de codage nominal multiple, l'analyse des correspondances multiples est

identique à l'analyse des correspondances. Si des groupes de variables sont intéressants, vous devez utiliser une analyse des corrélations canoniques nominales (non linéaires).

Obtenir une analyse des correspondances multiples

1. A partir des menus, sélectionnez :
Analyse > Réduction des dimensions > Codage optimal...
2. Sélectionnez **Certaines variables non nominales multiples**.
3. Sélectionnez **Un groupe**.
4. Cliquez sur **Définir**.
5. Sélectionnez au moins deux variables d'analyse et spécifiez le nombre de dimensions de la solution.
6. Cliquez sur **OK**.

Vous pouvez peut-être spécifier des variables supplémentaires qui sont ajustées à la solution trouvée, ou des variables de libellé pour les tracés.

Définir l'échelle et la pondération dans CATPCA

Vous pouvez définir le niveau de codage optimal des variables d'analyse et des variables supplémentaires. Par défaut, elles sont codées comme des splines monotones de second degré (ordinales) avec deux noeuds intérieurs. En outre, vous pouvez également définir la pondération pour les variables d'analyse.

Pondération de la variable : Vous pouvez choisir une pondération pour chaque variable. La valeur spécifiée doit être un nombre entier positif. La valeur par défaut est 1.

Niveau du codage optimal : Vous pouvez également sélectionner le niveau de codage à utiliser pour quantifier chaque variable.

- **Spline ordinale :** L'ordre des catégories de la variable observée est conservé dans la variable codée de façon optimale. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. La transformation résultante est un modèle polynomial monotone lissé du degré choisi. Ses différents éléments dépendent du nombre de noeuds intérieurs défini par l'utilisateur ainsi que du positionnement de ces derniers, déterminé par la procédure.
- **Spline nominale :** La seule information de la variable observée qui est conservée dans la variable codée de façon optimale est le groupe des objets dans les catégories. L'ordre des catégories de la variable observée n'est pas conservé. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. La transformation résultante est un modèle polynomial lissé, peut-être non monotone, du degré choisi. Ses différents éléments dépendent du nombre de noeuds intérieurs défini par l'utilisateur ainsi que du positionnement de ces derniers, déterminé par la procédure.
- **Nominal multiple :** La seule information de la variable observée qui est conservée dans la variable codée de façon optimale est le groupe des objets dans les catégories. L'ordre des catégories de la variable observée n'est pas conservé. Les points des catégories se trouvent sur les centroïdes des objets dans les catégories particulières. L'option *Multiple* indique que divers groupes de valeurs affectées sont obtenus pour chaque dimension.
- **Ordinal :** L'ordre des catégories de la variable observée est conservé dans la variable codée de façon optimale. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. La transformation du résultat convient mieux que la transformation ordinale spline, mais s'avère moins lissée.
- **Nominal :** La seule information de la variable observée qui est conservée dans la variable codée de façon optimale est le groupe des objets dans les catégories. L'ordre des catégories de la variable observée n'est pas conservé. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. La transformation du résultat convient mieux que la transformation nominale spline mais s'avère moins lissée.

- **Numérique** : Les catégories sont considérées comme triées et espacées régulièrement (niveau d'intervalle). L'ordre des catégories ainsi que les distances égales entre les nombres de catégories de la variable sont conservées dans la variable codée de façon optimale. Les points des catégories se trouvent sur une ligne droite (vecteur) passant par l'origine. Lorsque toutes les variables sont au niveau numérique, l'analyse est analogue à celle en composantes principales standard.

Analyse des correspondances multiples : Discrétisation

La boîte de dialogue Discrétisation vous permet de choisir une méthode de recodage des variables. Les valeurs fractionnées sont regroupées en sept catégories (ou en nombre de valeurs distinctes de variables si le nombre est inférieur à sept) avec une distribution normale approximative, à moins qu'une autre configuration ne soit spécifiée. Les variables de chaîne sont toujours converties en nombres entiers positifs en affectant des indicateurs de catégories selon l'ordre croissant alphanumérique. La discrétisation des variables de chaîne s'applique à ces nombres entiers. Par défaut, d'autres variables sont laissées inutilisées. Les variables discrétisées sont ensuite utilisées dans l'analyse.

Méthode : Choisissez entre Regroupement, Classement et Multiplier.

- **Regroupement** : Recodez en un nombre spécifié de catégories ou par intervalle.
- **Classement** : La variable est discrétisée via le classement des observations.
- **Multiplier** : Les valeurs courantes de la variable sont standardisées, multipliées par 10 et arrondies, et possèdent une constante ajoutée de sorte que la valeur discrétisée la plus faible soit égale à 1.

Regroupement : Les options suivantes sont disponibles lorsque vous discrétisez des variables par groupe :

- **Nombre de catégories** : Indiquez un nombre de catégories et définissez si les valeurs de la variable doivent faire l'objet d'une distribution approximativement gaussienne ou uniforme entre ces catégories.
- **Intervalles égaux** : Les variables sont recodées en catégories définies par ces intervalles de taille égale. N'oubliez pas de spécifier la longueur des intervalles.

Analyse des correspondances multiples : Valeurs manquantes

La boîte de dialogue Valeurs manquantes vous permet de choisir la stratégie de gestion des valeurs manquantes pour les variables de l'analyse et les variables supplémentaires.

Stratégie : Choisissez d'exclure les valeurs manquantes (traitement passif), d'affecter des valeurs (traitement actif) ou d'exclure les objets contenant des valeurs manquantes (suppression des observations incomplètes).

- **Exclure les valeurs manquantes pour affectation de corrélations après quantification** : Les objets contenant des valeurs manquantes sur la variable sélectionnée ne contribuent pas à l'analyse de cette variable. Si un traitement passif est effectué sur toutes les variables, les objets dont les variables comportent des valeurs manquantes sont traités comme étant supplémentaires. Si les corrélations sont spécifiées dans la boîte de dialogue Sortie, les valeurs manquantes après analyse sont alors prises en compte avec la catégorie la plus fréquente ou le mode de la variable pour les corrélations des variables d'origine. Pour corrélérer des variables codées de façon optimale, vous devez choisir une méthode d'imputation.
 - **Mode** : Permet de remplacer les valeurs manquantes par le mode de la variable codée de façon optimale.
 - **Catégorie supplémentaire** : Permet de remplacer les valeurs manquantes par la quantification d'une catégorie supplémentaire. Ce paramètre suppose que les objets contenant une valeur manquante pour cette variable sont considérés comme appartenant à la même catégorie (supplémentaire).
 - **Catégorie aléatoire** : Permet d'attribuer à chaque valeur manquante d'une variable la valeur quantifiée d'un nombre de catégories aléatoires différentes en fonction des effectifs marginaux des catégories de la variable.

- **Inclure les valeurs manquantes** : Des valeurs sont prises en compte pour les objets contenant des valeurs manquantes sur la variable sélectionnée. Vous pouvez choisir la méthode de calcul :
 - **Mode** : Permet de remplacer les valeurs manquantes par la catégorie la plus fréquente. S'il existe plusieurs modes, utilisez celui dont l'indicateur de catégories est le plus petit.
 - **Catégorie supplémentaire** : Permet de remplacer les valeurs manquantes par la même quantification d'une catégorie supplémentaire. Ce paramètre suppose que les objets contenant une valeur manquante pour cette variable sont considérés comme appartenant à la même catégorie (supplémentaire).
 - **Catégorie aléatoire** : Permet de remplacer chaque valeur manquante d'une variable par un nombre de catégories aléatoires différentes en fonction des effectifs marginaux des catégories.
- **Exclure les objets contenant des valeurs manquantes sur cette variable** : Les objets contenant des valeurs manquantes dans la variable sélectionnée sont retirés de l'analyse. Cette option n'est pas disponible pour les variables supplémentaires.

Analyse des correspondances multiples : Options

La boîte de dialogue Options permet de sélectionner la configuration initiale, de spécifier les itérations et les critères de convergence, de sélectionner une méthode de standardisation, de sélectionner une méthode de libellé des tracés et, enfin, de spécifier des objets supplémentaires.

Objets supplémentaires : Indiquez le numéro d'observation de l'objet, ou les premier et dernier numéros d'observation d'une plage d'objets que vous souhaitez définir comme objet supplémentaire, puis cliquez sur **Ajouter**. Si un objet est spécifié comme supplémentaire, alors les pondérations d'observation est ignorée pour cet objet.

Méthode de standardisation : Vous pouvez spécifier l'une des cinq options de standardisation des scores des objets et des variables. Une seule méthode de standardisation peut être utilisée dans chaque analyse.

- **Variable principale** : Cette option optimise l'association entre les variables. Les coordonnées des variables dans l'espace objet correspondent aux chargements entre composants et variables initiales (corrélations comportant des composantes principales telles que des dimensions et des scores des objets). Cette méthode est utile si vous êtes avant tout intéressé par les corrélations entre variables.
- **Objet principal** : Cette option optimise les distances entre les objets. Cette méthode est utile si vous êtes avant tout intéressé par les différences ou similitudes entre objets.
- **Symétrique** : Utilisez cette option de standardisation si vous êtes avant tout intéressé par la relation entre les objets et les variables.
- **Indépendant** : Utilisez cette option de standardisation si vous souhaitez examiner les distances entre les objets ainsi que les corrélations entre variables séparément.
- **Personnalisée** : Vous pouvez spécifier toute valeur réelle comprise dans l'intervalle $[-1, 1]$. La valeur 1 équivaut à la méthode Principal d'objet. La valeur 0 équivaut à la méthode Symétrique. La valeur -1 équivaut à la méthode Principal de variable. En spécifiant une valeur comprise entre -1 et 1 , la valeur propre peut comprendre à la fois les objets et les variables. Cette méthode est utile pour effectuer des tracés doubles ou triples.

Critères : Vous pouvez spécifier le nombre maximum d'itérations que la procédure peut prendre en charge dans ses calculs. Vous avez également la possibilité de sélectionner une valeur de critère de convergence. L'algorithme interrompt son itération dès que la différence du total ajusté entre les deux dernières itérations est inférieur à la valeur de la convergence ou dès que le nombre maximum d'itérations est atteint.

Libeller les tracés par : Vous pouvez préciser si les libellés de variable et de valeurs ou les noms ou valeurs de variables sont utilisés dans les tracés. Vous pouvez également spécifier une longueur maximale pour les libellés.

Dimension des tracés : Vous pouvez contrôler les dimensions affichées dans la sortie.

- **Afficher toutes les dimensions dans la solution** : Toutes les dimensions de la solution apparaissent dans une matrice de nuage de points.
- **Limiter le nombre de dimensions** : Les dimensions affichées sont limitées à des paires de dimensions tracées. Si vous restreignez ces dimensions, vous devez sélectionner la plus petite et la plus grande à tracer. La plus petite dimension peut être comprise entre 1 et le nombre de dimensions contenues dans la solution moins 1. En outre, elle est tracée par rapport aux dimensions plus grandes. La valeur de dimension la plus élevée peut être comprise entre 2 et le nombre de dimensions contenues dans la solution. Par ailleurs, elle indique la plus grande dimension à utiliser pour le traçage des paires de dimensions. Cette spécification s'applique à l'ensemble des tracés multidimensionnelles demandés.

Rotation : Vous pouvez sélectionner cette méthode pour obtenir une rotation des résultats.

Remarque : Ces méthodes de rotation ne sont disponibles que si vous sélectionnez l'option **Effectuer un bootstrap** dans la boîte de dialogue **Bootstrap**.

- **Varimax** : Méthode de rotation orthogonale qui minimise le nombre de variables ayant de forts chargements sur chaque composant. Cette méthode simplifie l'interprétation des composants.
- **Quartimax** : Méthode de rotation qui réduit le nombre de composants requis pour expliquer chaque variable. Cette méthode simplifie l'interprétation des variables observées.
- **Equamax** : Méthode de rotation qui est une combinaison de la méthode Varimax (qui simplifie les composants) et de la méthode Quartimax (qui simplifie les variables). Le nombre de variables pesant sur un composant et le nombre de composants nécessaires pour expliquer une variable sont minimisés.
- **Oblimin** : Méthode de rotation oblique (non orthogonale). Lorsque delta équivaut à 0, les composants sont les plus obliques. Plus la valeur de delta est négative, moins les composants sont obliques. Les valeurs positives autorisent une corrélation des composants supplémentaires. La valeur **Delta** doit être inférieure ou égale à 0,8.
- **Promax** : Méthode de rotation oblique (non orthogonale), qui permet aux composants d'être corrélés. Elle peut être calculée plus rapidement qu'une rotation Oblimin, aussi est-elle utile pour les vastes jeux de données. La quantité de corrélation (obliquité) autorisée est contrôlée par le paramètre kappa. La valeur **Kappa** doit être supérieure ou égale à 1 et inférieure à 10 000.

Configuration : Vous pouvez lire les données d'un fichier contenant les coordonnées de la configuration. La première variable du fichier contient les coordonnées de la première dimension. La deuxième contient les coordonnées de la deuxième dimension et ainsi de suite.

- **Initiale** : La configuration spécifiée dans le fichier est utilisée comme point de départ de l'analyse.
- **Fixe** : La configuration spécifiée dans le fichier est utilisée pour ajuster les variables. Les variables ainsi ajustées doivent être sélectionnées comme des variables d'analyse, mais la configuration étant fixe, elles doivent être considérées comme des variables supplémentaires (il est donc inutile de les sélectionner comme telles).

Analyse des correspondances multiples : Sortie

La boîte de dialogue Sortie contrôle l'affichage des résultats.

Scores d'objet : Affiche les scores des objets avec les options suivantes :

- **Inclure les catégories de** : Présente les indicateurs de catégories des variables d'analyse sélectionnées.
- **Libeller les scores des objets par** : Vous pouvez sélectionner l'une des variables spécifiées dans la liste de variables de libellé pour libeller les objets.

Chargements entre composantes et variables initiales : Affiche les chargements entre composantes et variables initiales pour toutes les variables n'ayant pas reçu de niveau de codage nominal multiple. Vous pouvez trier les chargements de composantes par taille.

Historique des itérations : Pour chaque itération, la variance représentée, la perte et l'augmentation de la variance représentée sont affichées.

Corrélations des variables d'origine : Affiche la matrice de corrélation des variables d'origine ainsi que les valeurs propres de cette matrice.

Corrélations des variables transformées : Affiche la matrice de corrélation des variables transformées (codées de façon optimale) ainsi que les valeurs propres de cette matrice.

Variance expliquée par : Affiche le nombre de variances représentées par les coordonnées du centroïde, les coordonnées vectorielles et le total (coordonnées du centroïde et vectorielles combinées) par variable et par dimension.

Valeurs affectées aux catégories : Indique les valeurs affectées aux catégories et les coordonnées pour chaque dimension des variables sélectionnées.

Statistiques descriptives : Affiche les effectifs, le nombre de valeurs manquantes et le mode des variables sélectionnées.

Analyse des correspondances multiples : Enregistrer

Les contrôles de la boîte de dialogue Enregistrer permettent de sauvegarder des données discrétisées, des scores d'objets, des valeurs transformées et d'autres résultats dans le jeu de données actif, dans un nouveau jeu de données de la session en cours ou dans un fichier externe.

- 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.
- Les noms de fichiers ou les noms du jeu de données doivent être différents pour chaque type de données enregistrées.
- Si vous enregistrez les scores des objets ou les valeurs transformées dans le jeu de données actif, vous pouvez indiquer le nombre des dimensions nominales multiples.
- Les options du groupe **Ellipses de confiance de bootstrap** ne sont disponibles que si vous avez sélectionné l'option **Effectuer un bootstrap** dans la boîte de dialogue Bootstrap.

Analyse des correspondances multiples : Tracés d'objets et de variables

La boîte de dialogue Tracés d'objets et de variables vous permet de spécifier les types de tracés de votre choix ainsi que les variables pour lesquelles des tracés seront représentés.

Points des objets : Un tracé des points des objets s'affiche.

Objets et variables (tracé double) : Les points des objets sont tracés avec les coordonnées de variables de votre choix : chargements entre composants et variables initiales ou centroïdes de variables.

Objets, chargements et centroïdes (triplot) : Les points des objets sont tracés avec les centroïdes des variables de niveau de codage nominal multiple et avec les chargements entre composants et variables initiales des autres variables.

Variables des tracés doubles et triples : Vous pouvez choisir d'utiliser toutes les variables des tracés doubles et triples ou de sélectionner un sous-groupe.

Etiqueter les objets : Vous pouvez choisir de libeller des objets avec les catégories des variables sélectionnées (choisissez les valeurs des indicateurs de catégories ou les libellés de valeurs dans la boîte de dialogue Options) ou avec le nombre d'observations. Si vous avez sélectionné **Variables**, un seul tracé est créé par variable.

Analyse des correspondances multiples : Tracés de catégories

La boîte de dialogue Tracés de catégories vous permet de spécifier les types de tracés de votre choix ainsi que les variables pour lesquelles des tracés seront représentés.

Tracés de catégories : Pour chaque variable sélectionnée, un tracé des coordonnées du centroïde et vectorielles est représenté. Pour les variables contenant des niveaux de codage nominal multiple, les catégories figurent dans les centroïdes des objets des catégories particulières. Pour les autres niveaux de codage, les catégories figurent dans un vecteur passant par l'origine.

Joindre les tracés de catégories : Il s'agit d'un tracé simple des coordonnées du centroïde et des coordonnées vectorielles de chaque variable sélectionnée.

Tracés de transformation : Affiche un tracé des valeurs affectées aux catégories optimales contre les indicateurs de catégories. Vous pouvez spécifier le nombre de dimensions pour les variables contenant des niveaux de codage multiples. Un tracé sera alors généré pour chaque dimension. Il vous est également possible de choisir d'afficher des tracés résiduels pour chaque variable sélectionnée.

Projeter les centroïdes de : Vous pouvez choisir une variable et projeter ses centroïdes sur les variables sélectionnées. Les variables comportant un niveau de codage nominal multiple ne peuvent pas être sélectionnées pour être projetées. Lorsque vous lancez ce tracé, un tableau doté des coordonnées des centroïdes projetés est également affiché.

Analyse des correspondances multiples : Tracés de chargement

La boîte de dialogue Tracés de chargement contrôle les variables incluses dans le tracé, l'affichage des centroïdes dans le tracé de chargement et l'affichage de la variance représentée.

Variance représentée : Pour chaque dimension, affiche le tracé de la variance représentée.

Afficher les chargements entre composants : Si cette option est sélectionnée, un tracé des chargements entre composants apparaît.

Variables de chargement : Vous pouvez choisir d'utiliser toutes les variables d'un tracé des chargements entre composants ou de sélectionner un sous-groupe.

Inclure les centroïdes : Les variables de niveau de codage nominal multiple ne possèdent pas de chargements mais vous pouvez choisir d'inclure leurs centroïdes dans le tracé. Vous pouvez utiliser toutes les variables catégorielles multiples ou sélectionner un sous-groupe.

Analyse des correspondances multiples : Bootstrap

La boîte de dialogue Bootstrap permet de définir les paramètres nécessaires à l'analyse par bootstrap.

Effectuer un bootstrap : Effectue un rééchantillonnage par bootstrap.

- Si des tracés de chargements, de catégories ou de scores de composants sont requis, des tracés supplémentaires s'affichent. Ils comportent les points des échantillons de données et les estimations de bootstrap. Les tracés de transformation incluent des régions de confiance. Un tracé des valeurs propres est également fourni.

- Si une solution à deux dimensions est définie, des tracés d'ellipse de confiance s'affichent pour les valeurs propres, les chargements de composantes, les points de catégories et les points d'objets.
- Le rééchantillonnage par bootstrap n'est pas disponible si vous spécifiez une méthode de rotation dans la boîte de dialogue Options.

Vous pouvez choisir un bootstrap de type **Équilibré** ou **Déséquilibré**.

Nombre d'échantillons : Nombre d'échantillons de bootstrap utilisés pour calculer les estimations de bootstrap. Il doit s'agir d'un nombre entier positif.

Niveau de confiance : Niveau de confiance des estimations de bootstrap, exprimé en pourcentage. Il doit s'agir d'un nombre positif inférieur à 100.

Méthode de correspondance : **Analyse procustéenne** ou **Réflexion**.

Ellipses de confiance : Contrôle la zone de seuil des ellipses de confiance dans les tracés. Pour les zones supérieures à (opérateur >) la valeur spécifiée, le nombre d'ellipses diminue au fur et à mesure que la valeur de seuil augmente.

- Les paramètres de ce groupe ne sont disponibles que si le nombre de dimensions spécifiées dans la boîte de dialogue principale est 2.
- Les ellipses de confiance destinées aux tracés de chargements ne sont disponibles que si l'option **Afficher les chargements de composants** est sélectionnée dans la boîte de dialogue Tracés de chargements.
- Les ellipses de confiance destinées aux tracés d'objets ne sont disponibles que si l'option **Points des objets** est sélectionnée dans la boîte de dialogue Tracés d'objets et de variables.
- Les ellipses de confiance destinées aux tracés de catégories ne sont disponibles que si une ou plusieurs variables sont définies dans la liste **Tracés de catégories** de la boîte de dialogue du même nom.

Nombre de points de contour de l'ellipse : Nombre de points de tracés utilisés pour représenter chaque ellipse de confiance. Les valeurs les plus grandes génèrent de meilleures ellipses. La valeur doit être un entier positif inférieur ou égal à 100.

Fonctions supplémentaires de la commande CATPCA

Vous pouvez personnaliser l'analyse des correspondances multiples si vous collez vos sélections dans une fenêtre de syntaxe et modifiez la syntaxe de commande CATPCA. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Spécifier les noms de racine des variables transformées, les scores des objets et les approximations lorsque vous les enregistrez dans le jeu de données actif (avec la sous-commande SAVE).
- Spécifier la longueur maximale pour les libellés de chaque tracé séparément (avec la sous-commande PLOT).
- Spécifier une liste de variables distincte pour les tracés résiduels (avec la sous-commande PLOT).

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

Chapitre 4. Analyse canonique non linéaire (OVERALS)

L'analyse de corrélation canonique non linéaire correspond à l'analyse de corrélation canonique nominale avec codage optimal. Le but de cette procédure est de déterminer la similitude entre les groupes de variables catégorielles et les autres. Cette analyse est également connue sous l'acronyme OVERALS.

L'analyse de corrélation canonique standard est une extension de la régression multiple, dans laquelle le second groupe ne contient pas de variable de réponse unique, mais contient des variables de réponses multiples à la place. Elle sert à expliquer autant que possible la variance tirée des relations entre deux groupes de variables numériques dans un espace de petite dimension. Initialement, les variables de chaque groupe sont combinées de façon linéaire de sorte que les combinaisons comportent une corrélation maximale. Compte tenu de ces combinaisons, celles qui sont linéaires sont déterminées par celles qui ne le sont pas avec les combinaisons précédentes et par celles ayant la plus importante corrélation.

L'approche de codage optimal développe l'analyse standard de trois façons différentes. D'abord, OVERALS vous permet d'avoir plus de deux groupes de variables. Deuxièmement, les variables peuvent être codées soit de façon nominale, soit ordinale, soit numérique. En conséquence, des relations non linéaires entre les variables peuvent être analysées. Enfin, au lieu d'optimiser les corrélations entre les jeux de variable, ceux-ci sont comparés à un groupe de compromis inconnu défini par les scores des objets.

Exemple : L'analyse de corrélation canonique nominale avec codage optimal peut être utilisée pour afficher graphiquement la relation entre un groupe de variables contenant une catégorie d'emploi et les années d'étude, et un autre groupe de variables contenant la zone de résidence et le sexe. Il est possible que vous trouviez que les années d'étude et la zone de résidence établissent une différence plus importante que les autres variables. Mais, vous pouvez considérer que les années d'étude établissent une différence fondamentale sur la première dimension.

Tracés et statistiques : Effectifs, centroïdes, historique des itérations, scores des objets, valeurs affectées aux catégories, pondérations et chargements entre composantes et variables initiales, ajustement unique et multiple, tracés de coordonnées des objets, tracés de coordonnées des catégories, tracés de chargements entre composantes et variables initiales, tracés de centroïdes de catégorie et tracés de transformation.

Informations sur les données d'analyse de corrélation canonique non linéaire

Données : Utilisez des entiers pour coder les variables catégorielles (niveau de codage nominal ou ordinal). Pour réduire le nombre de sorties, utilisez des entiers consécutifs commençant par 1 pour coder les variables. Les variables codées à un niveau numérique ne doivent pas être recodées en entiers consécutifs. Pour réduire le nombre de sorties, pour chaque variable codée à un niveau numérique, soustrayez la plus petite valeur observée de chaque valeur et ajoutez-lui 1. Les valeurs fractionnelles sont tronquées après la décimale.

Hypothèses : Les variables peuvent être classées en deux groupes ou plus. Les variables dans l'analyse sont codées sous forme nominale multiple, nominale simple, ordinale ou numérique. Le nombre maximum de dimensions utilisées dans la procédure dépend du niveau de codage optimal des variables. Si toutes les variables sont indiquées comme étant ordinales, nominales simples ou numériques, le nombre maximum de dimensions est le plus petit des deux valeurs suivantes : le nombre d'observations moins 1 ou le nombre total des variables. Cependant, si seuls les deux groupes de variables sont définis, le nombre maximum de dimensions correspond au nombre de variables du plus petit groupe. Si plusieurs variables sont nominales multiples, le nombre maximum de dimensions correspond au nombre total de catégories nominales multiples plus le nombre de variables catégorielles non multiples et moins le nombre de variables catégorielles multiples. Par exemple, si l'analyse implique cinq variables et si l'une

d'elles est nominale multiple avec quatre catégories, le nombre maximum de dimensions est $(4 + 4 - 1)$ ou 7. Si vous spécifiez un nombre supérieur au maximum, la valeur maximale est alors utilisée.

Procédures apparentées : Si chaque groupe contient une variable, l'analyse de corrélation canonique non linéaire équivaut à l'analyse des composantes principales avec codage optimal. Si chacune de ces variables est nominale multiple, l'analyse correspond à l'analyse de correspondance multiple. Si deux groupes de variables sont impliqués et que l'un d'eux contient une seule variable, l'analyse correspond à une régression catégorielle avec codage optimal.

Obtenir une analyse de corrélation canonique non linéaire

1. A partir des menus, sélectionnez :

Analyse > Réduction des dimensions > Codage optimal...

2. Sélectionnez soit **Toutes les variables catégorielles multiples**, soit **Certaines variables non nominales multiples**.

3. Sélectionnez **Plusieurs groupes**.

4. Cliquez sur **Définir**.

5. Définissez au moins deux groupes de variables. Sélectionnez les variables que vous souhaitez inclure dans le premier groupe. Pour atteindre le dernier groupe, cliquez sur **Suivant** et sélectionnez les variables à inclure dans le second. Vous pouvez également ajouter des groupes supplémentaires. Cliquez sur **Précédent** pour revenir au jeu de variables défini précédemment.

6. Définissez la plage de valeurs et l'échelle de mesure (niveau de codage optimal) pour chaque variable sélectionnée.

7. Cliquez sur **OK**.

8. Eventuellement :

- Sélectionnez une ou plusieurs variables pour fournir les libellés de point aux tracés de scores des objets. Chaque variable produit un tracé séparé, avec les points libellés par ses valeurs. Vous devez définir une plage pour chacune de ces variables de libellé de tracés. Lorsque vous utilisez la boîte de dialogue, une variable unique ne peut pas être utilisée à la fois dans l'analyse et sous forme de variable de libellé. Si vous souhaitez libeller un tracé de scores d'objets avec une variable utilisée dans l'analyse, utilisez le sous-menu Calculer (disponible depuis le menu Transformer) pour créer une copie de cette variable. Utilisez la nouvelle variable pour libeller le tracé. Il vous est également possible d'utiliser la syntaxe de commande.
- Indiquez le nombre de dimensions souhaitées dans la solution. En général, choisissez autant de dimensions que nécessaires pour expliquer le maximum de la variation. Si l'analyse implique plusieurs dimensions, des tracés 3D des trois premières dimensions sont créés. D'autres dimensions peuvent également être affichées en éditant le graphique.

Définition d'une plage et d'une échelle

Vous devez définir une plage pour chaque variable. La valeur maximale indiquée doit être un nombre entier. Les valeurs des données fractionnelles sont tronquées dans l'analyse. Une valeur de catégorie située en dehors de la plage spécifiée est ignorée dans l'analyse. Pour réduire le nombre de sorties, utilisez le sous-menu Recoder automatiquement (disponible depuis le menu Transformer) pour créer des catégories consécutives commençant par 1 pour des variables considérées comme nominales ou ordinales. Recoder en entiers consécutifs n'est pas recommandé pour les variables codées à un niveau numérique. Pour réduire le nombre de sorties pour les variables traitées comme numériques, pour chaque variable, soustrayez la valeur minimale de chaque valeur et ajoutez-lui 1.

Vous pouvez également sélectionner le codage à utiliser pour quantifier chaque variable.

- **Ordinal :** L'ordre des catégories de la variable observée est conservé dans la variable quantifiée.
- **Nominal simple :** Dans la variable quantifiée, les objets d'une même catégorie reçoivent les mêmes scores.

- **Nominal multiple** : Les quantifications peuvent différer pour chaque dimension.
- **Numérique discret** : Les catégories sont considérées comme triées et espacées régulièrement. Les différences entre le nombre des catégories et l'ordre de celles de la variable observée sont conservées dans la variable quantifiée.

Définition d'une plage

Vous devez définir une plage pour chaque variable. La valeur maximale indiquée doit être un nombre entier. Les valeurs des données fractionnelles sont tronquées dans l'analyse. Une valeur de catégorie située en dehors de la plage spécifiée est ignorée dans l'analyse. Pour réduire le nombre de sorties, utilisez le sous-menu Recoder automatiquement (disponible depuis le menu Transformer) pour créer des catégories consécutives commençant par 1.

Vous devez également définir une plage pour chaque variable utilisée pour libeller les tracés de scores des objets. Cependant, les libellés des catégories comportant des valeurs de données situées en dehors de la plage définie pour la variable apparaissent sur les tracés.

Analyse de corrélation canonique non linéaire – Options

La boîte de dialogue Options vous permet de sélectionner des statistiques et des tracés facultatifs, d'enregistrer les scores des objets en tant que nouvelles variables dans le jeu de données actif, de spécifier les critères d'itérations et de convergence et d'indiquer une configuration initiale pour l'analyse.

Afficher : Les statistiques disponibles incluent les effectifs marginaux, les centroïdes, l'historique des itérations, les pondérations et chargements entre composantes et variables initiales, les valeurs affectées aux catégories, les scores des objets et l'ajustement unique et multiple.

- *Centroïdes*. Quantifications des catégories, et moyennes projetées et réelles des scores des objets (observations) inclus dans chaque ensemble pour les observations qui appartiennent à la même catégorie de la variable.
- *Pondérations et chargements*. Coefficients de régression dans chaque dimension pour chaque variable quantifiée d'un groupe. Les scores des objets sont régressés sur les variables quantifiées et la projection de la variable est quantifiée dans l'espace d'objet. Fournit une indication de la contribution que chaque variable apporte à la dimension dans chaque classe.
- *Ajustement simple et multiple*. Mesure la qualité de l'ajustement des coordonnées simples et multiples/quantifications de catégories par rapport aux objets.
- *Valeurs affectées aux catégories*. Les valeurs de codage optimal sont affectées aux catégories d'une variable.
- *Scores des objets*. Quantification optimale affectée à un objet (observation) dans une dimension particulière.

Tracé : Vous pouvez générer des tracés de coordonnées des catégories, de scores des objets, de chargements entre composantes et variables initiales, de centroïdes de catégorie et de transformation.

Enregistrer les scores des objets : Il est possible d'enregistrer les scores des objets en tant que nouvelles variables dans le jeu de données actif. Ces scores des objets sont enregistrés en fonction du nombre de dimensions spécifiées dans la boîte de dialogue principale.

Utiliser configuration initiale aléatoire : Une configuration initiale aléatoire doit être utilisée si une partie ou la totalité des variables est nominale simple. Si cette case n'est pas cochée, une configuration initiale imbriquée est utilisée.

Critères : Vous pouvez spécifier le nombre maximum d'itérations que l'analyse canonique non linéaire peut prendre en charge dans ses calculs. Vous avez également la possibilité de sélectionner une valeur de

critère de convergence. L'analyse interrompt son itération dès que la différence de l'ajustement total entre les deux dernières itérations est inférieure à la valeur de la convergence, ou dès que le nombre maximum d'itérations est atteint.

Fonctions supplémentaires de la commande OVERALS

Vous pouvez personnaliser l'analyse canonique non linéaire en collant vos sélections dans une fenêtre de syntaxe et en modifiant la syntaxe de commande OVERALS. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Spécifier les paires de dimensions à tracer, plutôt que tracer toutes les dimensions extraites (à l'aide du mot-clé NDIM de la sous-commande PLOT).
- Indiquer le nombre de caractères composant les libellés de valeurs utilisés pour libeller des points sur les tracés (avec la sous-commande PLOT).
- Désigner plus de cinq variables sous forme de variables de libellé pour les tracés de scores des objets (avec la sous-commande PLOT).
- Sélectionner les variables utilisées dans l'analyse en tant que variables de libellé pour les tracés de scores des objets (avec la sous-commande PLOT).
- Sélectionner les variables à fournir aux libellés de points pour le tracé de score de quantification (avec la sous-commande PLOT).
- Indiquer le nombre d'observations à inclure dans l'analyse si vous ne souhaitez pas utiliser toutes les observations dans le jeu de données actif (avec la sous-commande NOBSERVATIONS).
- Spécifier les noms de racine des variables créées en enregistrant les scores des objets (avec la sous-commande SAVE).
- Spécifier le nombre de dimensions à enregistrer, plutôt que de sauvegarder toutes les dimensions extraites (avec la sous-commande SAVE).
- Ecrire les valeurs affectées aux catégories dans un fichier de matrice (avec la sous-commande MATRIX).
- Produire des tracés à faible résolution pouvant être plus faciles à lire que des tracés à haute résolution (avec la sous-commande SET).
- Produire des tracés de centroïdes et de transformations uniquement pour les variables spécifiées (avec la sous-commande PLOT).

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

Chapitre 5. Analyse des correspondances

L'une des fonctions de l'analyse des correspondances consiste à décrire les relations existant entre deux variables catégorielles dans une table de correspondances pour un espace comportant peu de dimensions, tout en décrivant simultanément les relations entre les catégories de chaque variable. Pour chacune des variables, les distances séparant les points des catégories d'un tracé reflètent les relations existant entre ces catégories : plus les catégories sont similaires, plus elles sont proches les unes des autres. Les points de projection d'une variable du vecteur situés entre l'origine et l'un des points de catégorie de l'autre variable décrivent les relations entre les deux variables.

Une analyse des tableaux de contingence implique fréquemment l'examen des profils des lignes et des colonnes ainsi qu'un test d'indépendance au moyen de la statistique khi-deux. Toutefois, le nombre de profils peut s'avérer assez élevé et le test du khi-deux n'indique pas la structure des dépendances. La procédure Tableaux croisés offre plusieurs mesures d'association et tests d'association mais ne permet pas de représenter graphiquement les relations existant entre les variables.

L'analyse factorielle constitue une technique standard de description des relations entre les variables d'un espace comportant peu de dimensions. Toutefois, l'analyse factorielle nécessite des données d'intervalle et le nombre d'observations doit être égal au nombre de variables multiplié par cinq. L'analyse des correspondances, en revanche, met en jeu des variables catégorielles et peut décrire les relations entre les catégories de chaque variable, ainsi que les relations entre les variables. En outre, l'analyse des correspondances permet d'analyser n'importe quel tableau de mesures de correspondances positives.

Exemple : L'analyse des correspondances peut être utilisée pour représenter graphiquement les relations existant entre la catégorie socio-professionnelle et le nombre de cigarettes consommées. Vous pourriez ainsi déterminer que la consommation de tabac diffère entre les jeunes cadres et les secrétaires, mais est similaire entre les secrétaires et les cadres supérieurs. Il vous serait également possible de déduire que les grands fumeurs sont principalement de jeunes cadres, alors que les fumeurs occasionnels sont généralement des secrétaires.

Tracés et statistiques : Mesures de correspondances, profils de lignes et de colonnes, valeurs singulières, scores de lignes et de colonnes, inertie, masse, statistiques de confiance des scores de lignes et de colonnes, statistiques de confiance des valeurs singulières, tracés de transformation, tracés de point de ligne, tracés de point de colonne et tracés doubles.

Remarques sur les données de l'analyse des correspondances

Données : Les variables catégorielles à analyser sont codées de façon nominale. Pour les données agrégées ou pour les mesures de correspondances autres que les effectifs, utilisez une variable de pondération présentant des valeurs de similarité positives. Pour les données de tableau, utilisez la syntaxe pour lire le tableau.

Hypothèses : Le nombre maximal de dimensions utilisé dans la procédure dépend du nombre de catégories de ligne et de colonne actives et du nombre de contraintes d'égalité. Si aucune contrainte d'égalité n'est appliquée et que toutes les catégories sont actives, la dimensionnalité maximale est inférieure de un au nombre de catégories de la variable présentant le plus petit nombre de catégories. Par exemple, si l'une des variables comporte cinq catégories et l'autre quatre, le nombre maximal de dimensions sera de trois. Les catégories supplémentaires ne sont pas actives. Par exemple, si une variable comporte cinq catégories, dont deux supplémentaires, et que l'autre variable possède quatre catégories, le nombre maximal de dimensions sera égal à deux. Tous les groupes de catégories faisant l'objet d'une contrainte d'égalité doivent être considérés comme une seule catégorie. Ainsi, si une variable comporte cinq catégories, dont trois doivent être égales, vous devrez considérer cette variable comme ne possédant que trois catégories pour déterminer la dimensionnalité maximale. Deux de ces catégories sont non

contraintes, et la troisième correspond aux trois catégories contraintes. Si vous définissez un nombre de dimensions supérieur au nombre maximal autorisé, la valeur maximale sera appliquée par défaut.

Procédures apparentées : Si vous travaillez avec plus de deux variables, procédez à une analyse de correspondance multiple. Si les variables doivent être codées de façon ordinale, utilisez l'analyse des correspondances multiples.

Pour obtenir une analyse des correspondances

1. A partir des menus, sélectionnez :
Analyse > Réduction des dimensions > Analyse des correspondances...
2. Sélectionnez une variable de ligne.
3. Sélectionnez une variable de colonne.
4. Définir les plages des variables.
5. Cliquez sur **OK**.

Définition de la plage de ligne dans l'analyse des correspondances

Vous devez définir une plage pour la variable en ligne. Les valeurs minimale et maximale spécifiées doivent être des nombres entiers. Les valeurs des données fractionnelles sont tronquées dans l'analyse. Une valeur de catégorie située en dehors de la plage spécifiée est ignorée dans l'analyse.

A l'origine, toutes les catégories sont non contraintes et actives. Vous pouvez par la suite contraindre certaines catégories de ligne à être égales à d'autres catégories de ligne, ou définir une catégorie de ligne comme supplémentaire.

- **Les catégories doivent être égales :** Les catégories doivent présenter des scores identiques. Appliquez des contraintes d'égalité si l'ordre obtenu pour les catégories n'est pas souhaitable ou est contraire à l'intuition. Le nombre maximal de catégories de ligne pouvant faire l'objet d'une contrainte d'égalité correspond au nombre total de catégories de ligne actives moins 1. Pour imposer différentes contraintes d'égalité aux groupes de catégories, utilisez la syntaxe. Par exemple, utilisez la syntaxe pour contraindre les catégories 1 et 2 à être égales, puis pour appliquer la même contrainte aux catégories 3 et 4.
- **La catégorie est un supplément :** Les catégories supplémentaires n'influencent pas l'analyse, mais sont représentées dans l'espace défini par les catégories actives. Les catégories supplémentaires ne jouent aucun rôle dans la définition des dimensions. Le nombre maximal de catégories de ligne supplémentaires correspond au nombre total de catégories de ligne moins 2.

Définition de la plage de colonne dans l'analyse des correspondances

Vous devez définir une plage pour la variable en colonne. Les valeurs minimale et maximale spécifiées doivent être des nombres entiers. Les valeurs des données fractionnelles sont tronquées dans l'analyse. Une valeur de catégorie située en dehors de la plage spécifiée est ignorée dans l'analyse.

A l'origine, toutes les catégories sont non contraintes et actives. Vous pouvez par la suite contraindre certaines catégories de colonne à être égales à d'autres catégories de colonne ou définir une catégorie de colonne comme supplémentaire.

- **Les catégories doivent être égales :** Les catégories doivent présenter des scores identiques. Appliquez des contraintes d'égalité si l'ordre obtenu pour les catégories n'est pas souhaitable ou est contraire à l'intuition. Le nombre maximal de catégories de colonne pouvant faire l'objet d'une contrainte d'égalité correspond au nombre total de catégories de colonne actives moins 1. Pour imposer différentes contraintes d'égalité aux groupes de catégories, utilisez la syntaxe. Par exemple, utilisez la syntaxe pour contraindre les catégories 1 et 2 à être égales, puis pour appliquer la même contrainte aux catégories 3 et 4.

- **La catégorie est un supplément** : Les catégories supplémentaires n'influencent pas l'analyse, mais sont représentées dans l'espace défini par les catégories actives. Les catégories supplémentaires ne jouent aucun rôle dans la définition des dimensions. Le nombre maximal de catégories de colonne supplémentaires correspond au nombre total de catégories de colonne moins 2.

Modèle d'analyse des correspondances

La boîte de dialogue Modèle vous permet de définir le nombre de dimensions, la mesure de distance, la méthode de standardisation et la méthode de standardisation.

Dimensions de la solution : Spécifiez le nombre de dimensions. En général, choisissez autant de dimensions que nécessaires pour expliquer le maximum de la variation. Le nombre maximal de dimensions dépend du nombre de catégories actives utilisé dans l'analyse et des contraintes d'égalité. Le nombre maximal de dimensions est égal au plus petit d'entre ces deux nombres :

- Le nombre de catégories de ligne actives moins le nombre de catégories de ligne faisant l'objet d'une contrainte d'égalité, plus le nombre de groupes de catégories de ligne avec contrainte ;
- Le nombre de catégories de colonne actives moins le nombre de catégories de colonne faisant l'objet d'une contrainte d'égalité, plus le nombre de groupes de catégories de colonne avec contrainte.

Mesure de distance : Vous pouvez sélectionner la mesure de la distance entre les lignes et les colonnes de la table des correspondances. Choisissez l'une des options suivantes :

- **Khi-deux** : Utilisez une distance de profil pondérée, la pondération correspondant à la masse des lignes ou des colonnes. Cette mesure est requise pour l'analyse des correspondances standard.
- **Euclidienne** : Utilisez la racine carrée de la somme des différences entre paires de lignes et paires de colonnes élevées au carré.

Méthode de standardisation : Choisissez l'une des options suivantes :

- **Moyennes de lignes et de colonnes éliminées** : Les lignes et les colonnes sont centrées. Cette méthode est requise pour l'analyse des correspondances standard.
- **Moyennes de lignes éliminées** : Seules les lignes sont centrées.
- **Moyennes de colonnes éliminées** : Seules les colonnes sont centrées.
- **Les Totaux de lignes sont égalisés et les moyennes éliminées** : Les marges des lignes sont égalisées avant que les lignes ne soient centrées.
- **Les totaux de colonnes sont égalisés et les moyennes éliminées** : Les marges des colonnes sont égalisées avant que les colonnes soient centrées.

Méthode de standardisation : Choisissez l'une des options suivantes :

- **Symétrique** : Pour chaque dimension, les scores des lignes représentent la moyenne pondérée des scores des colonnes, divisée par la valeur singulière correspondante ; les scores des colonnes représentent la moyenne pondérée des scores des lignes, divisée par la valeur singulière correspondante. Utilisez cette méthode si vous souhaitez examiner les différences ou les similitudes existant entre les catégories des deux variables.
- **Principale** : Les distances entre les points des lignes et des colonnes sont des approximations des distances de la table des correspondances en fonction de la mesure de distance sélectionnée. Appliquez cette méthode si vous souhaitez examiner les différences existant entre les catégories de l'une ou des deux variables, plutôt que les différences entre ces deux variables.
- **Principale en ligne** : Les distances entre les points des lignes sont des approximations des distances de la table des correspondances en fonction de la mesure de distance sélectionnée. Les scores des lignes correspondent à la moyenne pondérée des scores des colonnes. Utilisez cette méthode si vous souhaitez examiner les différences ou les similitudes existant entre les catégories de la variable de ligne.
- **Principale en colonne** : Les distances entre les points des colonnes sont des approximations des distances de la table des correspondances en fonction de la mesure de distance sélectionnée. Les scores

des colonnes correspondent à la moyenne pondérée des scores des lignes. Utilisez cette méthode si vous souhaitez examiner les différences ou les similitudes existant entre les catégories de la variable en colonne.

- **Personnalisée** : Vous devez définir une valeur comprise entre -1 et 1. La valeur -1 correspond à la méthode principale en colonne. La valeur 1 correspond à la méthode de standardisation principale en ligne. La valeur 0 correspond à la méthode de standardisation symétrique. Toutes les autres valeurs propagent l'inertie sur les scores des lignes et des colonnes à différents degrés. Cette méthode s'avère utile pour la création de tracés doubles adaptés à vos besoins.

Statistiques de l'analyse des correspondances

La boîte de dialogue Statistiques vous permet de définir les sorties numériques que vous souhaitez obtenir.

Table des correspondances : Tableau croisé des variables d'entrée incluant les totaux marginaux de ligne et de colonne.

Caractéristiques des points lignes : Pour chaque catégorie de ligne, indique les scores, la masse, l'inertie, la contribution du point à l'inertie de la dimension ainsi que la contribution de la dimension à l'inertie du point.

Caractéristiques des points colonnes : Pour chaque catégorie de colonne, indique les scores, la masse, l'inertie, la contribution du point à l'inertie de la dimension ainsi que la contribution de la dimension à l'inertie du point.

Profils lignes : Pour chaque catégorie de ligne, indique la distribution entre les catégories de la variable en colonne.

Profils colonnes : Pour chaque catégorie de colonne, indique la distribution entre les catégories de la variable en ligne.

Permutations de la table des correspondances : Réorganisation de la table des correspondances afin que les lignes et les colonnes apparaissent dans l'ordre croissant en fonction des scores de la première dimension. Une option vous permet de définir le nombre maximal de dimensions pour lequel vous souhaitez créer des tableaux permutés. Un tableau permuté sera alors généré pour chaque dimension comprise entre 1 et le nombre défini par vous.

Statistiques de confiance pour points lignes : Ecart type et corrélations de tous les points de ligne non supplémentaires.

Statistiques de confiance pour points colonnes : Ecart type et corrélations de tous les points de colonne non supplémentaires.

Tracés de l'analyse des correspondances

La boîte de dialogue Tracés vous permet de définir les tracés que vous souhaitez créer.

Nuages de points : Génère une matrice de tous les tracés présentant les dimensions par paire. Les nuages de points disponibles sont les suivants :

- **Tracé double** : Crée une matrice des tracés joints représentant les points des lignes et des colonnes. Si vous avez sélectionné la méthode de standardisation principale, l'option tracé double ne sera pas disponible.
- **Points lignes** : Crée une matrice des tracés représentant les points des lignes.
- **Points colonnes** : Crée une matrice des tracés représentant les points des colonnes.

Une option vous permet de définir le nombre de caractères composant les libellés de valeurs utilisés pour le libellé des points. Cette valeur doit être un nombre entier positif inférieur ou égal à 20.

Courbes : Crée un tracé pour chaque dimension de la variable sélectionnée. Les tracés en ligne disponibles sont les suivants :

- **Catégories de ligne transformées** : Produit un tracé représentant les valeurs des catégories de ligne d'origine par rapport aux scores des lignes qui leur correspondent.
- **Catégories de colonnes transformées** : Produit un tracé représentant les valeurs des catégories de colonne d'origine par rapport aux scores des colonnes qui leur correspondent.

Une option vous permet de définir le nombre de caractères composant les libellés de valeurs utilisés pour le libellé de l'axe des catégories. Cette valeur doit être un nombre entier positif inférieur ou égal à 20.

Dimension des tracés : Permet de contrôler les dimensions contenues dans la sortie.

- **Afficher toutes les dimensions dans la solution** : Toutes les dimensions de la solution apparaissent dans une matrice de nuage de points.
- **Limiter le nombre de dimensions** : Les dimensions affichées sont limitées à des paires de dimensions tracées. Si vous restreignez ces dimensions, vous devez sélectionner la plus petite et la plus grande à tracer. La plus petite dimension peut être comprise entre 1 et le nombre de dimensions contenues dans la solution moins 1. En outre, elle est tracée par rapport aux dimensions plus grandes. La valeur de dimension la plus élevée peut être comprise entre 2 et le nombre de dimensions contenues dans la solution. Par ailleurs, elle indique la plus grande dimension à utiliser pour le traçage des paires de dimensions. Cette spécification s'applique à l'ensemble des tracés multidimensionnelles demandés.

Fonctions supplémentaires de la commande CORRESPONDENCE

Vous pouvez personnaliser votre analyse des correspondances en collant vos sélections dans une fenêtre de syntaxe, puis en modifiant la syntaxe de la commande CORRESPONDENCE. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Indiquer les données des tableaux comme entrées au lieu d'utiliser les données d'observation (au moyen de la sous-commande TABLE = ALL).
- Spécifier le nombre de caractères composant les libellés de valeurs utilisés pour le libellé des points de chaque type de matrice de nuage de points ou de tracé double (au moyen de la sous-commande PLOT).
- Indiquer le nombre de caractères composant les libellés de valeurs utilisés pour le libellé des points de chaque type de tracé en ligne (au moyen de la sous-commande PLOT).
- Créer une matrice des scores des lignes et des colonnes dans un fichier de données de matrice (avec la sous-commande OUTFILE).
- Créer une matrice des statistiques de confiance (variances et covariances) pour les valeurs singulières et les scores dans un fichier de données de matrice (avec la sous-commande OUTFILE).
- Appliquer une contrainte d'égalité à plusieurs groupes de catégories (au moyen de la sous-commande EQUAL).

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

Chapitre 6. Analyse de correspondance multiple

L'analyse de correspondance multiple quantifie les données (catégorielles) nominales en attribuant des valeurs numériques aux observations (objets) et aux catégories, pour que les objets faisant partie de la même catégorie soient proches les uns des autres et ceux de différentes catégories, éloignés les uns des autres. Chaque objet se trouve aussi près que possible des points de catégorie qui s'appliquent. Ainsi, les catégories divisent les objets en sous-groupes homogènes. Les variables sont considérées comme homogènes lorsqu'elles classent les objets des mêmes catégories dans les mêmes sous-groupes.

Exemple : L'analyse de correspondance multiple peut être utilisée pour afficher graphiquement la relation entre la catégorie d'emploi, la classification des minorités et le sexe. Vous pouvez trouver que la classification par minorités et le sexe sont discriminant pour les personnes, mais que la catégorie d'emploi ne l'est pas. Vous avez également la possibilité de constater que les catégories Latino et Afro-Américaines sont similaires les unes des autres.

Tracés et statistiques : Scores des objets, mesures de discrimination, historique des itérations, corrélations des variables d'origine et des variables transformées, quantifications des catégories, statistiques descriptives, tracés de points des objets, tracés doubles, tracés de catégories, tracés de catégories joints, tracés de transformation et tracés de mesures de discrimination.

Remarques sur les données de l'analyse de correspondance multiple

Données : Les variables de chaîne sont toujours converties en nombres entiers positifs par ordre croissant alphanumérique. Les valeurs manquantes définies par l'utilisateur, les valeurs système manquantes et les valeurs inférieures à 1 sont considérées comme manquantes ; vous pouvez donc recoder ou ajouter une constante aux variables contenant des valeurs inférieures à 1 pour les définir comme non manquantes.

Hypothèses : Toutes les variables contiennent le niveau de codage nominal multiple. Les données doivent contenir au moins trois observations valides. L'analyse repose sur des données sous forme de nombres entiers positifs. L'option de discrétisation classe automatiquement une variable fractionnée en regroupant ses valeurs en catégories avec une distribution quasi normale et convertit automatiquement les valeurs des variables de chaîne en nombre entiers positifs. Vous pouvez en outre, spécifier d'autres schémas de discrétisation.

Procédures apparentées : Pour deux variables, l'analyse de correspondance multiple est identique à l'analyse des correspondances. Si vous pensez que ces variables possèdent des propriétés ordinales ou numériques, vous devez utiliser l'analyse des correspondances multiples. Si des groupes de variables sont intéressants, vous devez utiliser une analyse des corrélations canoniques (non linéaires).

Pour obtenir une analyse de correspondance multiple

1. A partir des menus, sélectionnez :
Analyse > Réduction des dimensions > Codage optimal...
2. Sélectionnez **Toutes les variables nominales multiples**.
3. Sélectionnez **Un groupe**.
4. Cliquez sur **Définir**.
5. Sélectionnez au moins deux variables d'analyse et spécifiez le nombre de dimensions de la solution.
6. Cliquez sur **OK**.

Vous pouvez peut-être spécifier des variables supplémentaires qui sont ajustées à la solution trouvée, ou des variables de libellé pour les tracés.

Définition d'une pondération de variable dans une analyse de correspondance multiple

Vous pouvez définir la pondération pour les variables d'analyse.

Pondération de la variable : Vous pouvez choisir une pondération pour chaque variable. La valeur spécifiée doit être un nombre entier positif. La valeur par défaut est 1.

Analyse des correspondances multiples : Discrétisation

La boîte de dialogue Discrétisation vous permet de choisir une méthode de recodage des variables. Les valeurs fractionnées sont regroupées en sept catégories (ou en nombre de valeurs distinctes de variables si le nombre est inférieur à sept) avec une distribution normale approximative, à moins qu'une autre configuration ne soit spécifiée. Les variables de chaîne sont toujours converties en nombres entiers positifs en affectant des indicateurs de catégories selon l'ordre croissant alphanumérique. La discrétisation des variables de chaîne s'applique à ces nombres entiers. Par défaut, d'autres variables sont laissées inutilisées. Les variables discrétisées sont ensuite utilisées dans l'analyse.

Méthode : Choisissez entre Regroupement, Classement et Multiplier.

- **Regroupement :** Recodez en un nombre spécifié de catégories ou par intervalle.
- **Classement :** La variable est discrétisée via le classement des observations.
- **Multiplier :** Les valeurs courantes de la variable sont standardisées, multipliées par 10 et arrondies, et possèdent une constante ajoutée de sorte que la valeur discrétisée la plus faible soit égale à 1.

Regroupement : Les options suivantes sont disponibles lorsque vous discrétisez des variables par groupe :

- **Nombre de catégories :** Indiquez un nombre de catégories et définissez si les valeurs de la variable doivent faire l'objet d'une distribution approximativement gaussienne ou uniforme entre ces catégories.
- **Intervalles égaux :** Les variables sont recodées en catégories définies par ces intervalles de taille égale. N'oubliez pas de spécifier la longueur des intervalles.

Analyse des correspondances multiples : Valeurs manquantes

La boîte de dialogue Valeurs manquantes vous permet de choisir la stratégie de gestion des valeurs manquantes pour les variables de l'analyse et supplémentaires.

Stratégie de la valeur manquante : Choisissez d'exclure les valeurs manquantes (traitement passif), d'affecter des valeurs (traitement actif) ou d'exclure les objets contenant des valeurs manquantes (suppression des observations incomplètes).

- **Exclure les valeurs manquantes pour affectation de corrélations après quantification :** Les objets contenant des valeurs manquantes sur la variable sélectionnée ne contribuent pas à l'analyse de cette variable. Si un traitement passif est effectué sur toutes les variables, les objets dont les variables comportent des valeurs manquantes sont traités comme étant supplémentaires. Si les corrélations sont spécifiées dans la boîte de dialogue Sortie, les valeurs manquantes après analyse sont alors prises en compte avec la catégorie la plus fréquente ou le mode de la variable pour les corrélations des variables d'origine. Pour corrélérer des variables codées de façon optimale, vous devez choisir une méthode d'imputation. Sélectionnez **Mode** pour remplacer les valeurs manquantes par le mode de la variable codée de façon optimale. Sélectionnez **Catégorie supplémentaire** pour remplacer les valeurs manquantes par la valeur affectée à une catégorie supplémentaire. Cela suppose que les objets contenant une valeur manquante pour cette variable sont considérés comme appartenant à la même catégorie (supplémentaire).
- **Inclure les valeurs manquantes :** Des valeurs sont prises en compte pour les objets contenant des valeurs manquantes sur la variable sélectionnée. Vous pouvez choisir la méthode de calcul :

Sélectionnez **Mode** pour remplacer les valeurs manquantes par la catégorie la plus fréquente. S'il existe plusieurs modes, utilisez celui dont l'indicateur de catégories est le plus petit. Sélectionnez **Catégorie supplémentaire** pour remplacer les valeurs manquantes par la valeur affectée à une catégorie supplémentaire. Cela suppose que les objets contenant une valeur manquante pour cette variable sont considérés comme appartenant à la même catégorie (supplémentaire).

- **Exclure les objets contenant des valeurs manquantes sur cette variable** : Les objets contenant des valeurs manquantes dans la variable sélectionnée sont retirés de l'analyse. Cette option n'est pas disponible pour les variables supplémentaires.

Analyse des correspondances multiples : Options

La boîte de dialogue Options vous permet de sélectionner la configuration initiale, de spécifier les itérations et les critères de convergence, de sélectionner une méthode de standardisation, de sélectionner une méthode de libellé des tracés et, enfin, de spécifier des objets supplémentaires.

Objets supplémentaires : Indiquez le numéro d'observation de l'objet (ou les premier et dernier numéros d'observation d'une plage d'objets) que vous souhaitez définir comme objet supplémentaire, puis cliquez sur **Ajouter**. Poursuivez jusqu'à ce que vous ayez indiqué tous les objets supplémentaires. Si un objet est spécifié comme supplémentaire, alors les pondérations d'observation est ignorée pour cet objet.

Méthode de standardisation : Vous pouvez spécifier l'une des cinq options de standardisation des scores des objets et des variables. Une seule méthode de standardisation peut être utilisée dans une analyse donnée.

- **Variable principale** : Cette option optimise l'association entre les variables. Les coordonnées des variables dans l'espace objet correspondent aux chargements entre composants et variables initiales (corrélations comportant des composantes principales telles que des dimensions et des scores des objets). Cela est utile si vous êtes avant tout intéressé par les corrélations entre variables.
- **Objet principal** : Cette option optimise les distances entre les objets. Cela est utile si vous êtes avant tout intéressé par les différences ou similitudes entre objets.
- **Symétrique** : Utilisez cette option de standardisation si vous êtes avant tout intéressé par la relation entre les objets et les variables.
- **Indépendant** : Utilisez cette option de standardisation si vous souhaitez examiner les distances entre les objets ainsi que les corrélations entre variables séparément.
- **Personnalisée** : Vous pouvez spécifier toute valeur réelle comprise dans l'intervalle $[-1, 1]$. La valeur 1 correspond à la méthode Objet principal, la valeur 0 correspond à la méthode Symétrique et la valeur -1 correspond à la méthode Variable principale. En spécifiant une valeur comprise entre -1 et 1 , la valeur propre peut comprendre à la fois les objets et les variables. Cette méthode est utile pour effectuer des tracés doubles ou triples.

Critères : Vous pouvez spécifier le nombre maximum d'itérations que la procédure peut prendre en charge dans ses calculs. Vous avez également la possibilité de sélectionner une valeur de critère de convergence. L'algorithme interrompt son itération dès que la différence du total ajusté entre les deux dernières itérations est inférieur à la valeur de la convergence ou dès que le nombre maximum d'itérations est atteint.

Libeller les tracés par : Vous permet de préciser si les libellés de variable et de valeurs ou les noms ou valeurs de variables sont utilisés dans les tracés. Vous pouvez également spécifier une longueur maximale pour les libellés.

Dimension des tracés : Permet de contrôler les dimensions contenues dans la sortie.

- **Afficher toutes les dimensions dans la solution** : Toutes les dimensions de la solution apparaissent dans une matrice de nuage de points.
- **Limiter le nombre de dimensions** : Les dimensions affichées sont limitées à des paires de dimensions tracées. Si vous restreignez ces dimensions, vous devez sélectionner la plus petite et la plus grande à

tracer. La plus petite dimension peut être comprise entre 1 et le nombre de dimensions contenues dans la solution moins 1. En outre, elle est tracée par rapport aux dimensions plus grandes. La valeur de dimension la plus élevée peut être comprise entre 2 et le nombre de dimensions contenues dans la solution. Par ailleurs, elle indique la plus grande dimension à utiliser pour le traçage des paires de dimensions. Cette spécification s'applique à l'ensemble des tracés multidimensionnelles demandés.

Configuration : Vous pouvez lire les données d'un fichier contenant les coordonnées de la configuration. La première variable du fichier doit contenir les coordonnées de la première dimension, la deuxième variable, celles de la deuxième dimension, et ainsi de suite.

- **Initiale :** La configuration du fichier spécifié sera utilisée comme point de départ de l'analyse.
- **Fixe :** La configuration du fichier spécifié sera utilisée pour ajuster les variables. Les variables ainsi ajustées doivent être sélectionnées comme des variables d'analyse, mais la configuration étant fixe, elles doivent être considérées comme des variables supplémentaires (il est donc inutile de les sélectionner comme telles).

Analyse des correspondances multiples : Sortie

La boîte de dialogue Sortie vous permet de créer des tableaux pour les scores des objets, les mesures de discrimination, l'historique des itérations, les corrélations des variables d'origine et des variables transformées, ainsi que les quantifications des catégories et statistiques descriptives des variables sélectionnées.

Scores d'objet : Affiche les scores des objets, y compris la masse, l'inertie et les contributions, ainsi que les options suivantes :

- **Inclure les catégories de :** Présente les indicateurs de catégories des variables d'analyse sélectionnées.
- **Libeller les scores des objets par :** Vous pouvez sélectionner l'une des variables spécifiées dans la liste de variables de libellé pour libeller les objets.

Mesures de discrimination : Affiche les mesures de discrimination par variable et par dimension.

Historique des itérations : Pour chaque itération, la variance représentée, la perte et l'augmentation de la variance représentée sont affichées.

Corrélations des variables d'origine : Affiche la matrice de corrélation des variables d'origine ainsi que les valeurs propres de cette matrice.

Corrélations des variables transformées : Affiche la matrice de corrélation des variables transformées (codées de façon optimale) ainsi que les valeurs propres de cette matrice.

Valeurs affectées aux catégories et contributions : Indique les valeurs affectées aux catégories (coordonnées), y compris la masse, l'inertie et les contributions pour chaque dimension de la ou des variables sélectionnées.

Remarque : Les coordonnées et les contributions (dont la masse et l'inertie) sont affichées dans des couches distinctes des sorties du tableau croisé dynamique, les coordonnées étant affichées par défaut. Pour afficher les contributions, double-cliquez sur le tableau et sélectionnez Contributions dans la liste déroulante Couche.

Statistiques descriptives : Affiche les effectifs, le nombre de valeurs manquantes et le mode de la ou des variables sélectionnées.

Analyse des correspondances multiples : Enregistrer

La boîte de dialogue Enregistrer vous permet d'enregistrer les données discrétisées, les scores des objets et les valeurs transformées dans un fichier de données externe IBM SPSS Statistics ou un jeu de données dans la session en cours. Vous pouvez également enregistrer les valeurs transformées et les scores des objets dans le jeu de données actif.

- 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.
- Les noms de fichiers ou les noms du jeu de données doivent être différents pour chaque type de données enregistrées.
- Si vous enregistrez les scores des objets ou les valeurs transformées dans le jeu de données actif, vous pouvez indiquer le nombre des dimensions nominales multiples.

Analyse des correspondances multiples : Tracés d'objets

La boîte de dialogue Tracés d'objets vous permet d'indiquer les types de tracés de votre choix ainsi que les variables à tracer.

Points des objets : Un tracé des points des objets s'affiche.

Objets et centroïdes (tracés doubles) : Les points des objets sont tracés avec les centroïdes de variable.

Variables de tracé double : Vous pouvez choisir d'utiliser toutes les variables des tracés doubles ou de sélectionner un sous-groupe.

Libeller objets : Vous pouvez choisir de libeller des objets avec les catégories des variables sélectionnées (choisissez les valeurs des indicateurs de catégories ou les libellés de valeurs dans la boîte de dialogue Options) ou avec le nombre d'observations. Si vous avez sélectionné **Variables**, un seul tracé est créé par variable.

Analyse des correspondances multiples : Tracés de variables

La boîte de dialogue Tracés de variables vous permet d'indiquer les types de tracés de votre choix ainsi que les variables à tracer.

Tracés de catégories : Pour chaque variable sélectionnée, un tracé des coordonnées du centroïde est tracé. Les catégories se trouvent dans les centroïdes des objets des catégories concernées.

Joindre les tracés de catégories : Il s'agit d'un tracé simple représentant les coordonnées du centroïde de chaque variable sélectionnée.

Tracés de transformation : Affiche un tracé des valeurs affectées aux catégories optimales contre les indicateurs de catégories. Vous pouvez spécifier le nombre de dimensions, un tracé sera créé pour chaque dimension. Il vous est également possible de choisir d'afficher des tracés résiduels pour chaque variable sélectionnée.

Mesures de discrimination : Crée un tracé des mesures de discrimination pour les variables sélectionnées.

Commande MULTIPLE CORRESPONDENCE - Fonctions additionnelles

Vous pouvez personnaliser votre analyse de correspondance multiple en collant vos sélections dans une fenêtre de syntaxe, puis en modifiant la syntaxe de la commande MULTIPLE CORRESPONDENCE. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Spécifier les noms de racine des variables transformées, les scores des objets et les approximations lorsque vous les enregistrez dans le jeu de données actif (avec la sous-commande SAVE).
- Spécifier la longueur maximale pour les libellés de chaque tracé séparément (avec la sous-commande PLOT).
- Spécifier une liste de variables distincte pour les tracés résiduels (avec la sous-commande PLOT).

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

Chapitre 7. Positionnement multidimensionnel (PROXSCAL)

Le positionnement multidimensionnel tente de déterminer la structure d'un groupe de mesures de proximité entre les objets. Ce procédé est effectué en affectant des observations à des positions particulières dans un espace conceptuel de petite dimension de telle sorte que les distances entre les points dans l'espace correspondent le mieux possible aux (dis)similarités données. Le résultat est une représentation à moindres carrés des objets dans cet espace de petite dimension, qui vous aidera, dans certains cas, à mieux comprendre vos données.

Exemple : Le positionnement multidimensionnel peut être très utile pour déterminer les relations perceptuelles. Par exemple, en considérant l'image de votre produit, vous pouvez mener une enquête en vue d'obtenir un jeu de données décrivant la similarité distinguée (ou proximité) de votre produit comparée à celle de vos concurrents. En utilisant ces variables de proximité et indépendantes (un prix, par exemple), vous pouvez essayer de déterminer quelles variables sont importantes suivant le mode d'affichage de ces produits et vous pouvez ajuster votre image en fonction.

Tracés et statistiques : Historique des itérations, mesures de contrainte, décomposition de la contrainte, coordonnées de l'espace commun, distances des objets dans la configuration finale, pondérations des espaces individuels, espaces individuels, proximités transformées, variables indépendantes transformées, tracés de contrainte, nuages de points des espaces communs, nuages de points de pondération des espaces individuels, nuages de points des espaces individuels, tracés de transformation, tracés résiduels de Shepard et tracés de transformation des variables explicatives.

Remarques sur les données du positionnement multidimensionnel

Données : Les données peuvent être indiquées dans le formulaire des matrices de proximité ou des variables qui sont converties en matrice de proximité. Les matrices peuvent être formatées en colonnes ou entre les colonnes. Les proximités peuvent être traitées par niveaux de codage rapport, intervalle, ordinal ou spline.

Hypothèses : Trois variables au moins doivent être spécifiées. Le nombre de dimensions ne doit pas dépasser le nombre d'objets moins un. La réduction de la dimensionnalité est omise si elle est combinée aux départs aléatoires multiples. Si vous indiquez une source seulement, tous les modèles équivalent au modèle d'identité, puis l'analyse sélectionne par défaut le modèle d'identité.

Procédures apparentées : Le codage de toutes les variables à un niveau numérique correspond à l'analyse de positionnement multidimensionnel standard.

Obtenir un positionnement multidimensionnel

1. A partir des menus, sélectionnez :

Analyse > Echelle > Positionnement multidimensionnel (PROXSCAL)...

Cette opération ouvre la boîte de dialogue Format des données.

2. Spécifiez le format des données :

Format des données : Indiquez si vos données constituent des mesures de proximité ou si vous souhaitez créer des proximités à partir des données.

Nombre de sources : Si vos données sont des proximités, spécifiez si vous avez des sources uniques ou multiples de mesures de proximité.

Une source : S'il existe une source de proximité, spécifiez si votre jeu de données est formaté avec les proximités d'une matrice sur des colonnes ou sur une colonne unique avec deux variables séparées pour identifier les lignes et colonnes de chaque proximité.

- *Les proximités se trouvent sur plusieurs colonnes d'une matrice*. La matrice de proximité s'étend à des colonnes dont le nombre est égal au nombre d'objets. Vous accédez ensuite à la boîte de dialogue Proximités sur plusieurs colonnes de matrices.
- *Les proximités se trouvent dans une seule colonne*. La matrice de proximité est réduite dans une seule colonne, ou variable. Deux variables supplémentaires identifiant la ligne et la colonne de chaque cellule sont nécessaires. Vous accédez ensuite à la boîte de dialogue Proximités sur une seule colonne.

Plusieurs sources : S'il existe plusieurs sources de proximités, spécifiez si le jeu de données est formaté avec les proximités des matrices empilées sur plusieurs colonnes, sur des colonnes multiples contenant une source par colonne ou sur une colonne simple.

- *Les proximités se trouvent dans des matrices empilées sur des colonnes*. Les matrices de proximité s'étalent sur un nombre de colonnes équivalent au nombre d'objets et sont empilées les unes sur les autres sur un nombre de lignes équivalent au produit du nombre d'objets et du nombre de sources. Vous accédez ensuite à la boîte de dialogue Proximités sur plusieurs colonnes de matrices.
- *Proximités sur plusieurs colonnes, une source par colonne*. Les matrices de proximité sont réduites dans plusieurs colonnes, ou variables. Deux variables supplémentaires identifiant la ligne et la colonne de chaque cellule sont nécessaires. Vous accédez ensuite à la boîte de dialogue Proximités sur des colonnes.
- *Les proximités sont empilées dans une seule colonne*. Les matrices de proximité sont réduites dans une seule colonne, ou variable. Trois variables supplémentaires identifiant la ligne, la colonne et la source de chaque cellule, sont nécessaires. Vous accédez ensuite à la boîte de dialogue Proximités sur une seule colonne.

3. Cliquez sur **Définir**.

Proximités dans des matrices sur plusieurs colonnes

Si vous sélectionnez les proximités dans un modèle de matrice des données pour une ou plusieurs sources dans la boîte de dialogue Format des données, procédez comme suit :

1. Sélectionnez deux ou plusieurs variables de proximité. (Veuillez vous assurer que l'ordre des variables dans la liste correspond à l'ordre des colonnes des proximités.)
2. Sélectionnez éventuellement un nombre de variables de pondération égal au nombre des variables de proximité. (Veuillez vous assurer que l'ordre des pondérations correspond à celui des proximités qu'elles pondèrent.)
3. S'il existe plusieurs sources, vous pouvez également sélectionner une variable de sources. (Le nombre d'observations dans chaque variable de proximité doit être égal au nombre de variables de proximité multiplié par le nombre de sources.)

De plus, vous pouvez définir un modèle pour un positionnement multidimensionnel, placer les restrictions dans l'espace commun, définir les critères de convergence, spécifier la configuration initiale à utiliser et enfin, choisir des tracés et des sorties.

Proximités sur plusieurs colonnes

Si vous sélectionnez le modèle de colonnes multiples pour plusieurs sources dans la boîte de dialogue Format des données, procédez comme suit :

1. Sélectionnez deux ou plusieurs variables. (Chaque variable est considérée comme étant une matrice de proximité provenant d'une source distincte.)
2. Sélectionnez une variable de lignes pour définir les positions de lignes pour les proximités dans chaque variable de proximités.
3. Sélectionnez une variable de colonnes pour définir les emplacements de colonne pour les proximités dans la variable des proximités. (Les cellules de la matrice de proximité n'ayant pas de désignation lignes/colonnes sont considérées comme manquantes.)

4. Sélectionnez éventuellement un nombre de variables de pondération égal au nombre des variables de proximité.

De plus, vous pouvez définir un modèle pour un positionnement multidimensionnel, placer les restrictions dans l'espace commun, définir les critères de convergence, spécifier la configuration initiale à utiliser et enfin, choisir des tracés et des sorties.

Proximités dans une colonne

Si vous sélectionnez le modèle de colonne unique pour une ou plusieurs sources dans la boîte de dialogue Format des données, procédez comme suit :

1. Sélectionnez une variable de proximité. (On considère qu'il existe une ou plusieurs matrices des proximités.)
2. Sélectionnez une variable de lignes pour définir les positions de lignes pour les proximités dans la variable des proximités.
3. Sélectionnez une variable de colonnes pour définir les emplacements de colonne pour les proximités dans la variable des proximités.
4. S'il existe plusieurs sources, sélectionnez une variable de sources. (Pour chaque source, les cellules de la matrice de proximité n'ayant pas de désignation lignes/colonnes sont considérées comme manquantes.)
5. Eventuellement, choisissez une variable de pondération.

De plus, vous pouvez définir un modèle pour un positionnement multidimensionnel, placer les restrictions dans l'espace commun, définir les critères de convergence, spécifier la configuration initiale à utiliser et enfin, choisir des tracés et des sorties.

Création de proximités à partir des données

Si vous choisissez de créer des proximités à partir des données de la boîte de dialogue Format des données, procédez comme suit :

1. Si vous créez des distances entre les variables (voir la boîte de dialogue Créer une mesure à partir des données), sélectionnez au moins trois variables. Ces variables seront utilisées pour créer la matrice de proximité (ou les matrices, s'il existe plusieurs sources). Si vous créez des distances entre les observations, seule une variable est requise.
2. S'il existe plusieurs sources, sélectionnez une variable de sources.
3. Choisissez éventuellement une mesure de création de proximités.

De plus, vous pouvez définir un modèle pour un positionnement multidimensionnel, placer les restrictions dans l'espace commun, définir les critères de convergence, spécifier la configuration initiale à utiliser et enfin, choisir des tracés et des sorties.

Création d'une mesure à partir des données

Le positionnement multidimensionnel utilise les données de dissimilarité pour créer une solution de codage. Si vos données sont multivariées (valeurs des variables mesurées), vous devez créer des données de dissimilarité afin de calculer une solution de positionnement multidimensionnel. Vous pouvez spécifier les détails de création de mesures de dissimilarité à partir de vos données.

Mesure : Vous permet de spécifier la mesure de dissimilarité adaptée à votre analyse. Sélectionnez une possibilité dans le groupe Mesure correspondant à votre type de données, puis sélectionnez l'une des mesures dans la liste déroulante correspondant à ce type de mesure. Les possibilités sont :

- **Intervalle :** Distance Euclidienne, Carré de la distance Euclidienne, Distance de Tchebycheff, Distance de Manhattan, Distance de Minkowski ou Autre.

- **Effectif** : Distance du khi-deux ou Distance du phi-deux.
- **Binaire** : Distance Euclidienne, Carré de la distance Euclidienne, Différence de taille, Différence de motif, Variance ou Lance et Williams.

Créer une matrice de distances : Vous permet de choisir l'unité d'analyse. Les possibilités sont Par variables ou Par observations.

Transformer les valeurs : Dans certains cas, comme lorsque les variables sont mesurées selon des échelles très différentes, vous voudrez standardiser des valeurs avant de calculer les proximités (ne s'applique pas aux données binaires). Sélectionnez une méthode de standardisation dans la liste déroulante Standardiser (si la standardisation n'est pas nécessaire, sélectionnez **Aucune**).

Définition d'un modèle de positionnement multidimensionnel

La boîte de dialogue Modèle vous permet d'indiquer un modèle de positionnement, son nombre minimum et maximum de dimensions, la structure de la matrice de proximité, la transformation à utiliser sur les proximités, et de déterminer si les proximités sont transformées dans chaque source séparément ou sans condition sur la source.

Modèle de positionnement : Choisissez parmi les options suivantes :

- **Identité** : Toutes les sources ont la même configuration.
- **Euclidien pondéré** : Ce modèle est un modèle des différences individuelles. Chaque source comporte un espace individuel dans lequel chaque dimension de l'espace commun est pondérée de façon différentielle.
- **Euclidien généralisé** : Ce modèle est un modèle des différences individuelles. Chaque source comporte un espace individuel qui est égal à une rotation de l'espace commun, suivie d'une pondération différenciée des dimensions.
- **Rang réduit** : Il s'agit d'un modèle euclidien généralisé pour lequel vous pouvez spécifier le rang de l'espace individuel. Vous devez spécifier un rang supérieur ou égal à 1 et inférieur au nombre maximum de dimensions.

Forme : Spécifiez si les proximités doivent être extraites des parties triangulaires inférieure ou supérieure de la matrice de proximité. Vous pouvez indiquer que la totalité de la matrice est utilisée, auquel cas la somme pondérée des parties triangulaires supérieure et inférieure sera analysée. Dans tous les cas, la matrice complète doit être spécifiée, y compris la diagonale, bien que seules les parties spécifiées seront utilisées.

Proximités : Spécifiez si votre matrice de proximité contient des mesures de similarité ou de dissimilarité.

Transformations de proximité : Choisissez parmi les options suivantes :

- **Rapport** : Les proximités transformées sont proportionnelles aux proximités originales. Uniquement disponible pour les proximités à valeurs positives.
- **Intervalle** : Les proximités transformées sont proportionnelles aux proximités originales et à une constante. Cette constante fait en sorte que les proximités transformées soient positives.
- **Ordinal** : Les proximités transformées ont le même ordre que les originales. Vous spécifiez si les proximités liées doivent être gardées liées ou autorisées à ne plus l'être.
- **Spline** : Les proximités transformées représentent une transformation polynomiale non décroissante lissée des proximités originales. Vous spécifiez le degré de la fonction polynomiale ainsi que le nombre de noeuds intérieurs.

Appliquer les transformations : Spécifiez si seules les proximités de chaque source sont comparées entre elles ou si les comparaisons sont sans condition sur la source.

Dimensions : Par défaut, une solution est calculée dans deux dimensions (minimum =2, maximum =2). Vous choisissez un entier minimum et maximum depuis 1 jusqu'au nombre d'objets moins 1 (tant que le minimum reste inférieur ou égal au maximum.) La procédure calcule une solution des dimensions maximales, puis réduit la dimensionnalité en matière d'étapes, jusqu'à ce que la plus petite soit atteinte.

Positionnement multidimensionnel : Restrictions

La boîte de dialogue Restrictions vous permet de placer les restrictions sur l'espace commun.

Restrictions sur l'espace commun : Spécifiez le type de restriction de votre choix.

- **Pas de restrictions :** Aucune restriction n'est placée sur l'espace commun.
- **Certaines coordonnées fixées :** La première variable sélectionnée contient les coordonnées des objets sur la première dimension ; la seconde correspond aux coordonnées des objets sur la deuxième dimension, et ainsi de suite. Une valeur manquante indique qu'une coordonnée sur une dimension est libre. Le nombre de variables sélectionnées doit être égal au nombre maximum de dimensions requis.
- **Combinaison linéaire de variables indépendantes :** L'espace commun se réduit à une combinaison linéaire des variables sélectionnées.

Variables de restriction : Sélectionnez les variables qui définissent les restrictions sur l'espace commun. Si vous spécifiez une combinaison linéaire, vous spécifiez une transformation d'intervalle, nominale, ordinale ou spline pour des variables de restriction. Dans tous les cas, le nombre d'observations pour chaque variable doit être égal au nombre d'objets.

Positionnement multidimensionnel : Options

La boîte de dialogue Options vous permet de sélectionner le style de configuration initiale, de spécifier les critères d'itération et de convergence, et de sélectionner des mises à jour standard ou relaxées.

Configuration initiale : Choisissez l'une des options suivantes :

- **Simplex :** Les objets sont placés à la même distance les uns des autres dans la dimension maximale. Une itération est prise pour améliorer cette configuration à haute dimension, suivie d'une réduction de dimension en vue d'obtenir une configuration initiale comportant le nombre maximum de dimensions spécifié dans la boîte de dialogue Modèle.
- **Torgerson :** Une solution de codage classique est utilisée comme configuration initiale.
- **Départ aléatoire unique :** Une configuration est choisie de façon aléatoire.
- **Départs aléatoires multiples :** Plusieurs configurations sont choisies de façon aléatoire, et celle ayant la contrainte brute la moins standardisée est utilisée comme configuration initiale.
- **Personnalisée :** Vous sélectionnez les variables contenant les coordonnées de votre configuration initiale. Le nombre de variables sélectionnées doit être égal au nombre de dimensions spécifié, avec la première variable correspondant aux coordonnées sur la dimension 1, la seconde correspondant aux coordonnées sur la dimension 2, et ainsi de suite. Le nombre d'observations dans chaque variable doit être égal au nombre d'objets.

Critères d'itération : Spécifiez les valeurs des critères d'itération.

- **Convergence de la contrainte :** L'algorithme interrompt son itération lorsque la différence des valeurs de la contrainte brute standardisée consécutif est inférieure au nombre spécifié ici, lequel doit être compris entre 0,0 et 1,0.
- **Contrainte minimum :** L'algorithme s'interrompt lorsque la contrainte brute standardisée tombe en dessous du nombre spécifié ici, lequel doit être compris entre 0,0 et 1,0.
- **Nombre maximum d'itérations :** L'algorithme exécute le nombre d'itérations spécifiées ici, à moins que l'un des critères ci-dessus ne soit déjà satisfait.
- **Utiliser mises à jour relaxées :** Ces mises à jour accélèrent l'algorithme ; elles ne peuvent être utilisées ni avec les modèles autres que le modèle d'identité, ni avec des restrictions.

Positionnement multidimensionnel : Tracés, Version 1

La boîte de dialogue Tracés vous permet de spécifier quels tracés doivent être produits. La présente rubrique décrit la boîte de dialogue Tracés lorsque vous sélectionnez le format de données Proximités dans les colonnes. Pour les tracés **Pondérations des espaces individuels**, **Proximités originales contre transformées** et **Proximités contre distances transformées**, il vous est possible de spécifier les sources pour lesquelles les tracés doivent être générés. La liste des sources disponibles constitue la liste des variables de proximité dans la boîte de dialogue principale.

Contrainte : Un tracé est produit à partir de la contrainte brute standardisée par opposition aux dimensions. Ce tracé est uniquement généré si le nombre maximum de dimensions est supérieur au nombre minimum.

Espace commun : Une matrice de nuage de points des coordonnées de l'espace commun est affiché.

Espaces individuels : Pour chaque source, les coordonnées des espaces individuels sont affichées dans les matrices de nuage de points. Cela est uniquement possible si l'un des modèles de différences individuels est spécifié dans la boîte de dialogue Modèle.

Pondérations des espaces individuels : Un nuage de points est produit à partir des pondérations des espaces individuels. Cela est uniquement possible si l'un des modèles de différences individuels est spécifié dans la boîte de dialogue Modèle. Pour le modèle Euclidien pondéré, les pondérations sont imprimées dans des tracés dont une dimension sur chaque axe. Pour le modèle Euclidien généralisé, un tracé est produit par dimension, indiquant à la fois la rotation et sa pondération. Le modèle Rang réduit génère le même tracé que le modèle Euclidien généralisé, mais réduit le nombre de dimensions des espaces individuels.

Proximités originales contre transformées : Les tracés sont générés à partir des proximités originales par opposition aux proximités transformées.

Proximités et distances transformées : Les proximités et distances transformées sont tracées sous la forme d'un graphique.

Variables indépendantes transformées : Les tracés de transformation sont produits pour les variables indépendantes.

Corrélations de variables et dimensions : Un tracé de corrélation entre les variables indépendantes et les dimensions de l'espace commun s'affiche.

Positionnement multidimensionnel : Tracés, Version 2

La boîte de dialogue Tracés vous permet de spécifier quels tracés doivent être produits. La présente rubrique décrit la boîte de dialogue Tracés lorsque vous sélectionnez un format de données autre que Proximités dans les colonnes. Pour les tracés **Pondérations des espaces individuels**, **Proximités originales contre transformées** et **Proximités contre distances transformées**, il vous est possible de spécifier les sources pour lesquelles les tracés doivent être générés. Les numéros de source entrés doivent être des valeurs de la variable de sources spécifiée dans la boîte de dialogue principale et être classés de 1 jusqu'au nombre de sources.

Positionnement multidimensionnel : Sortie

La boîte de dialogue Sortie vous permet de contrôler les sorties affichées et d'en enregistrer certains pour séparer des fichiers.

Afficher : Sélectionnez l'un des items suivants à afficher :

- **Coordonnées de l'espace commun :** Affiche les coordonnées de l'espace commun.
- **Coordonnées de l'espace individuel :** Les coordonnées de l'espace individuel sont affichées uniquement si le modèle n'est pas le modèle d'identité.
- **Pondérations des espaces individuels :** Affiche les pondérations des espaces individuels, uniquement si l'un des modèles de différences individuelles est spécifié. En fonction du modèle, les pondérations des espaces sont décomposées en pondérations de rotation et de dimension, lesquelles sont également affichées.
- **Distances :** Affiche les distances entre les objets de la configuration.
- **Transformations transformées :** Affiche les proximités transformées entre les objets de la configuration.
- **Données d'entrée :** Inclut les proximités originales, et si elles existent, les pondérations de données, la configuration initiale et les coordonnées fixées des variables indépendantes.
- **Contrainte pour les départs aléatoires :** Affiche la valeur de départ de nombre aléatoire et la valeur de la contrainte brute standardisée de chaque départ aléatoire.
- **Historique des itérations :** Affiche l'historique des itérations de l'algorithme principal.
- **Mesures de contraintes multiples :** Affiche les différentes valeurs de contrainte. Le tableau contient des valeurs pour la contrainte brute standardisée, le stress-I, le stress-II, le S-stress, la dispersion représentée (DAF) et enfin le coefficient de congruence de Tucker.
- **Décomposition de la contrainte :** Affiche la décomposition par objet et par source de la contrainte brute finale standardisée, notamment la moyenne par objet et par source.
- **Variables indépendantes transformées :** Si une restriction de combinaisons linéaires a été sélectionnée, les variables indépendantes transformées et les pondérations de régression correspondantes sont affichées.
- **Corrélations de variables et dimensions :** Si une restriction de combinaisons linéaires a été sélectionnée, les corrélations entre les variables indépendantes et les dimensions de l'espace commun sont affichées.

Enregistrer dans nouveau fichier : Vous pouvez enregistrer les coordonnées de l'espace commun, les pondérations des espaces individuels, les distances, les proximités transformées et les variables indépendantes transformées pour séparer les fichiers de données IBM SPSS Statistics.

Fonctions supplémentaires de la commande PROXSCAL

Vous pouvez personnaliser l'analyse de votre positionnement multidimensionnel de proximité en collant vos sélections dans une fenêtre de syntaxe et en modifiant la syntaxe de commande PROXSCAL. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Spécifier des listes de variables distinctes pour les tracés de transformation et résiduels (avec la sous-commande PLOT).
- Spécifier des listes de sources distinctes pour les tracés de pondération des espaces individuels, de transformation et résiduels (avec la sous-commande PLOT).
- Spécifier un sous-groupe des tracés de transformation de variables indépendantes à afficher (avec la sous-commande PLOT).

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

Chapitre 8. Dépliage multidimensionnel (PREFSCAL)

La procédure de dépliage multidimensionnel tente de trouver une échelle quantitative commune vous permettant d'examiner les relations entre deux ensembles d'objets de manière visuelle.

Exemples : Vous demandez à 21 personnes de classer 15 aliments constituant un petit-déjeuner selon leurs préférences, de 1 à 15. Le dépliage multidimensionnel vous permet de déterminer que la logique discriminatoire des individus suit deux schémas primaires : entre les pains mous et les pains durs et entre les aliments gras et allégés.

Autre exemple : vous demandez à un groupe de conducteurs de noter 26 modèles de voitures sur 10 critères selon une échelle de 6 points, allant de 1="pas vrai du tout" à 6="tout à fait vrai". En effectuant la moyenne des résultats de l'ensemble des individus, on constate une certaine similarité des valeurs. Le dépliage multidimensionnel vous permet cependant de distinguer des regroupements de modèles similaires et les critères avec lesquels ils sont le plus souvent associés.

Tracés et statistiques : La procédure de dépliage multidimensionnel permet de produire un historique des itérations, les mesures de contrainte, la décomposition de la contrainte, les coordonnées de l'espace commun, les distances des objets dans la configuration finale, les pondérations des espaces individuels, les espaces individuels, les proximités transformées, les tracés de contrainte, les nuages de points des espaces communs, les nuages de points de pondération des espaces individuels, les nuages de points des espaces individuels, les tracés de transformation et les tracés résiduels de Shepard.

Remarques sur les données du dépliage multidimensionnel

Données : Les données sont fournies sous forme de matrices de proximité rectangulaires. Chaque colonne est considérée comme un objet de colonne distinct. Chaque ligne d'une matrice de proximité est considérée comme un objet de ligne distinct. Lorsqu'il existe plusieurs sources de proximités, les matrices sont empilées.

Hypothèses : Deux variables au moins doivent être spécifiées. Le nombre de dimensions de la solution ne doit pas dépasser le nombre d'objets moins un. Si vous indiquez une source seulement, tous les modèles équivalent au modèle d'identité, puis l'analyse sélectionne par défaut le modèle d'identité.

Obtenir un dépliage multidimensionnel

1. A partir des menus, sélectionnez :
Analyse > Echelle > Dépliage multidimensionnel (PREFSCAL)...
2. Sélectionnez deux variables ou plus identifiant les colonnes dans la matrice de proximité rectangulaire. Chaque variable représente un objet de colonne distinct.
3. Sélectionnez éventuellement un nombre de variables de pondération égal au nombre de variables d'objets de colonnes. L'ordre des variables de pondération doit être le même que celui des objets de colonnes qu'elles pondèrent.
4. Eventuellement, choisissez une variable de ligne. Les valeurs (ou libellés de valeur) de cette variable sont utilisées pour libeller les objets de lignes de la sortie.
5. S'il existe plusieurs sources, sélectionnez éventuellement une variable de sources. Le nombre d'observations dans chaque fichier de données doit être égal au nombre d'objets de lignes multiplié par le nombre de sources.

De plus, vous pouvez définir un modèle pour un dépliage multidimensionnel, placer les restrictions dans l'espace commun, définir les critères de convergence, spécifier la configuration initiale à utiliser et enfin, choisir des tracés et des sorties.

Définition d'un modèle de dépliage multidimensionnel

La boîte de dialogue Modèle vous permet de spécifier un modèle de positionnement, un nombre minimum et maximum de dimensions, la structure de la matrice de proximité, la transformation à utiliser sur les proximités, et de déterminer si les proximités sont transformées avec condition sur la ligne, avec condition sur la source ou sans condition sur la source.

Modèle de positionnement : Choisissez parmi les options suivantes :

- **Identité :** Toutes les sources ont la même configuration.
- **Euclidien pondéré :** Ce modèle est un modèle des différences individuelles. Chaque source comporte un espace individuel dans lequel chaque dimension de l'espace commun est pondérée de façon différentielle.
- **Euclidien généralisé :** Ce modèle est un modèle des différences individuelles. Chaque source comporte un espace individuel qui est égal à une rotation de l'espace commun, suivie d'une pondération différenciée des dimensions.

Proximités : Spécifiez si votre matrice de proximité contient des mesures de similarité ou de dissimilarité.

Dimensions : Par défaut, une solution est calculée dans deux dimensions (minimum =2, maximum =2). Vous pouvez choisir un entier minimum et maximum depuis 1 jusqu'au nombre d'objets moins 1, tant que le minimum reste inférieur ou égal au maximum. La procédure calcule une solution dans la dimensionnalité maximale, puis réduit la dimensionnalité en matière d'étapes, jusqu'à ce que la plus petite soit atteinte.

Transformations de proximité : Choisissez parmi les options suivantes :

- **Aucune :** Les proximités ne sont pas transformées. Vous pouvez éventuellement sélectionner **Inclure une constante** afin de décaler les proximités d'une constante définie.
- **Linéaire :** Les proximités transformées sont proportionnelles aux proximités d'origine : la fonction de transformation estime une pente et la constante est définie sur 0. C'est ce qu'on appelle une transformation de rapport. Vous pouvez éventuellement sélectionner **Inclure une constante** afin de décaler les proximités d'une constante définie. Ce procédé est également appelé transformation d'intervalle.
- **Spline :** Les proximités transformées représentent une transformation polynomiale non décroissante lissée des proximités originales. Vous spécifiez le degré de la fonction polynomiale ainsi que le nombre de noeuds intérieurs. Vous pouvez éventuellement sélectionner **Inclure une constante** afin de décaler les proximités d'une constante définie.
- **Lissé :** Les proximités transformées présentent le même ordre que les proximités d'origine, y compris la restriction qui prend en compte les différences entre les valeurs successives. Il en résulte une transformation « ordinale lissée ». Vous pouvez spécifier si les proximités liées doivent être gardées liées ou autorisées à ne plus l'être.
- **Ordinal :** Les proximités transformées ont le même ordre que les originales. Vous pouvez spécifier si les proximités liées doivent être gardées liées ou autorisées à ne plus l'être.

Appliquer les transformations : Spécifiez si les proximités sont comparées l'une à l'autre dans chaque ligne ou dans chaque source ou si les comparaisons sont sans condition sur la ligne ou sur la source, c'est à dire si les transformations sont effectuées par ligne, par source ou sur toutes les proximités en une fois.

Restrictions du dépliage multidimensionnel

La boîte de dialogue Restrictions vous permet de placer les restrictions sur l'espace commun.

Restrictions sur l'espace commun : Vous pouvez choisir de fixer les coordonnées des objets de lignes et/ou de colonnes dans l'espace commun.

Variables de restriction des lignes/colonnes : Sélectionnez le fichier contenant les restrictions et sélectionnez les variables définissant les restrictions de l'espace commun. La première variable sélectionnée contient les coordonnées des objets sur la première dimension ; la seconde correspond aux coordonnées des objets sur la deuxième dimension, et ainsi de suite. Une valeur manquante indique qu'une coordonnée sur une dimension est libre. Le nombre de variables sélectionnées doit être égal au nombre maximum de dimensions requis. Le nombre d'observations dans chaque variable doit être égal au nombre d'objets.

Options de dépliage multidimensionnel

La boîte de dialogue Options vous permet de sélectionner le style de configuration initiale, de spécifier les critères d'itération et de convergence et de configurer le terme de pénalité pour la contrainte.

Configuration initiale : Choisissez l'une des options suivantes :

- **Classique :** La matrice de proximité rectangulaire est utilisée pour compléter les valeurs intrablocs (valeurs entre les lignes et entre les colonnes) de la matrice MDS symétrique complète. Une fois la matrice complète formée, une solution de positionnement classique est utilisée pour la configuration initiale. Les valeurs intrablocs peuvent être calculées à l'aide de l'inégalité triangulaire ou des distances de Spearman.
- **Ross-Cliff :** Le départ de Ross-Cliff utilise les résultats de la décomposition d'une valeur singulière sur une matrice de proximité à double centre carrée comme valeurs initiales pour les objets de lignes et de colonnes.
- **Correspondance :** Le départ par correspondance utilise les résultats d'une analyse de correspondance sur les données inversées (similitudes au lieu des différences) avec une normalisation symétrique des scores des lignes et des colonnes.
- **Centroïdes :** La procédure démarre avec le positionnement des objets de lignes dans la configuration à l'aide de la décomposition de la valeur propre. Les objets de colonnes sont ensuite positionnés dans le centroïde des choix spécifiés. Pour le nombre de choix, spécifiez un entier positif entre 1 et le nombre de variables de proximité.
- **Départs aléatoires multiples :** Les solutions sont calculées pour plusieurs configurations initiales sélectionnées de manière aléatoire et celle présentant la mesure de contrainte pénalisée la plus basse représente la meilleure.
- **Personnalisée :** Vous sélectionnez les variables contenant les coordonnées de votre configuration initiale. Le nombre de variables sélectionnées doit être égal au nombre de dimensions spécifié, avec la première variable correspondant aux coordonnées sur la dimension 1, la seconde correspondant aux coordonnées sur la dimension 2, et ainsi de suite. Le nombre d'observations dans chaque variable doit être égal au nombre combiné d'objets de lignes et de colonnes. Les coordonnées des lignes et des colonnes doivent être empilées, avec les coordonnées des colonnes à la suite des coordonnées des lignes.

Critères d'itération : Spécifiez les valeurs des critères d'itération.

- **Convergence de la contrainte :** L'algorithme interrompt son itération lorsque la différence relative des valeurs des mesures de contrainte pénalisées consécutives est inférieure au nombre spécifié ici, lequel ne peut pas être négatif.
- **Contrainte minimum :** L'algorithme s'arrête lorsque la mesure de contrainte pénalisée est inférieure au nombre spécifié ici, qui ne peut pas être négatif.
- **Nombre maximum d'itérations :** L'algorithme exécute le nombre d'itérations spécifié ici, à moins que l'un des critères ci-dessus ne soit déjà satisfait.

Terme de pénalité : L'algorithme tente de minimiser la mesure de contrainte pénalisée, qui est une mesure de la qualité d'ajustement égale au produit du Stress-I de Kruskal par un terme de pénalité basé sur le coefficient de variation des proximités transformées. Ces contrôles vous permettent de définir l'intensité et la plage du terme de pénalité.

- **Intensité** : Plus la valeur du paramètre d'intensité est petite, plus la pénalité est intense. Spécifiez une valeur comprise entre 0,0 et 1,0.
- **Plage** : Ce paramètre définit le moment auquel la pénalité devient active. Si vous le définissez sur 0,0, la pénalité est inactive. L'augmentation de la valeur entraîne la recherche par l'algorithme d'une solution présentant une plus grande variation parmi les proximités transformées. Spécifiez une valeur non négative.

Tracés de dépliage multidimensionnel

La boîte de dialogue Tracés vous permet de spécifier quels tracés doivent être produits.

Tracés : Les tracés suivants sont disponibles :

- **Départs multiples** : Affiche un histogramme empilé de la mesure de contrainte pénalisée, affichant à la fois la contrainte et la pénalité.
- **Espace commun initial** : Affiche une matrice de nuage de points des coordonnées de l'espace commun initial.
- **Contrainte par dimension** : Produit un graphique en lignes de la mesure de la contrainte pénalisée en fonction de la dimensionnalité. Ce tracé est uniquement généré si le nombre maximum de dimensions est supérieur au nombre minimum.
- **Espace commun final** : Une matrice de nuage de points des coordonnées de l'espace commun est affiché.
- **Pondération de l'espace** : Un nuage de points est produit à partir des pondérations des espaces individuels. Cela est uniquement possible si l'un des modèles de différences individuels est spécifié dans la boîte de dialogue Modèle. Pour le modèle Euclidien pondéré, les pondérations de toutes les sources sont affichées dans un tracé avec une dimension sur chaque axe. Pour le modèle Euclidien généralisé, un tracé est produit par dimension, indiquant à la fois la rotation et sa pondération pour chaque source.
- **Espaces individuels** : Une matrice de nuage de points des coordonnées de l'espace individuel de chaque source est affichée. Cela est uniquement possible si l'un des modèles de différences individuels est spécifié dans la boîte de dialogue Modèle.
- **Tracés de transformation** : Un nuage de points est généré à partir des proximités d'origine par opposition aux proximités transformées. Selon l'application des transformations, une couleur distincte est assignée à chaque ligne ou source. Une transformation inconditionnelle génère une seule couleur.
- **Tracés de Shepard** : Les proximités d'origine en fonction des proximités transformées et des distances. Les distances sont représentées par des points et les proximités transformées par une ligne. Selon l'application des transformations, une ligne distincte est générée pour chaque ligne ou source. Une transformation inconditionnelle génère une seule ligne.
- **Nuage de points de l'ajustement** : Un nuage de points des proximités transformées en fonction des distances est affiché. Une couleur distincte est assignée à chaque source lorsque plusieurs sources sont spécifiées.
- **Tracés résiduels** : Un nuage de points des proximités transformées en fonction des résidus (proximités transformées moins les distances) est affiché. Une couleur distincte est assignée à chaque source lorsque plusieurs sources sont spécifiées.

Styles d'objets de lignes : Les styles vous apportent un contrôle supplémentaire pour l'affichage des objets de lignes dans les tracés. Les valeurs des variables de couleurs facultatives sont utilisées pour passer en revue toutes les couleurs. Les valeurs des variables de marques facultatives sont utilisées pour passer en revue toutes les marques possibles.

Tracés sources : Pour les **espaces individuels**, **nuages de points de l'ajustement** et **tracés résiduels**, ainsi que pour les **tracés de transformation** et les **tracés de Shepard** si les transformations sont appliquées par

la source, vous pouvez spécifier les sources pour lesquelles générer des tracés. Les numéros de source entrés doivent être des valeurs de la variable de sources spécifiée dans la boîte de dialogue principale et être classés de 1 jusqu'au nombre de sources.

Tracés des lignes : Si des transformations sont appliquées par lignes, vous pouvez spécifier la ligne pour laquelle les tracés doivent être générés pour les **Tracés de transformations** et les **Tracés de Shepard**. Les numéros de lignes doivent être compris entre 1 et le nombre de lignes.

Sortie du dépliage multidimensionnel

La boîte de dialogue Sortie vous permet de contrôler les sorties affichées et d'en enregistrer certains pour séparer des fichiers.

Afficher : Sélectionnez l'un des éléments suivants à afficher :

- **Données d'entrée :** Inclut les proximités d'origine et, si elles existent, les pondérations de données, la configuration initiale et les coordonnées fixées.
- **Départs multiples :** Affiche la valeur de départ de nombre aléatoire et la valeur de la contrainte pénalisée de chaque départ aléatoire.
- **Données initiales :** Affiche les coordonnées de l'espace commun initial.
- **Historique des itérations :** Affiche l'historique des itérations de l'algorithme principal.
- **Mesures d'ajustement :** Affiche différentes mesures. Le tableau contient plusieurs mesures de qualité de l'ajustement, de défaut de l'ajustement, de corrélation, de variation et de non-dégénérescence.
- **Décomposition de la contrainte :** Affiche la mesure de la contrainte pénalisée de la décomposition d'un objet, d'une ligne ou d'une source, y compris les moyennes et les écarts-types de la ligne, de la colonne ou de la source.
- **Transformations transformées :** Affiche les proximités transformées.
- **Espace commun final :** Affiche les coordonnées de l'espace commun.
- **Pondération de l'espace :** Affiche les pondérations de l'espace individuel. Cette option est uniquement disponible lorsque l'un des modèles de différences individuelles est spécifié. En fonction du modèle, les pondérations des espaces sont décomposées en pondérations de rotation et de dimension, lesquelles sont également affichées.
- **Espaces individuels :** Les coordonnées de l'espace individuel sont affichées. Cette option est uniquement disponible lorsque l'un des modèles de différences individuelles est spécifié.
- **Distances ajustées :** Affiche les distances entre les objets de la configuration.

Enregistrer dans nouveau fichier : Vous pouvez enregistrer les coordonnées de l'espace commun, les pondérations des espaces individuels, les distances et les proximités transformées dans des fichiers de données IBM SPSS Statistics distincts.

Fonctions supplémentaires de la commande PREFSCAL

Vous pouvez personnaliser l'analyse des proximités du dépliage multidimensionnel en collant vos sélections dans une fenêtre de syntaxe et en modifiant la syntaxe de commande PROXSCAL résultante. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Spécifier plusieurs listes source pour les Espaces individuels, les Nuages de points de l'ajustement et les Tracés résiduels, ainsi que pour les Tracés de transformations et les Tracés de Shepard dans le cas de transformations conditionnelles d'une matrice, lorsque plusieurs sources sont disponibles (avec la sous-commande PLOT).
- Spécifier plusieurs listes de lignes pour les Tracés de transformations et les Tracés de Shepard dans le cas de transformations conditionnelles par lignes (avec la sous-commande PLOT).
- Spécifier un nombre de lignes au lieu d'une variable ID de ligne (avec la commande INPUT).
- Spécifier un nombre de sources au lieu d'une variable ID de source (avec la commande INPUT).

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

A

- ajustement
 - dans l'analyse de corrélation canonique non linéaire 29
- analyse de corrélation canonique non linéaire 27
 - statistiques 27
 - tracés 27
- Analyse de corrélation canonique non linéaire 28, 29
 - fonctions supplémentaires de la commande 30
- Analyse de correspondance multiple
 - fonctions supplémentaires de la commande 41
- analyse des correspondances 31, 33, 34
 - statistiques 31
 - tracés 31
- Analyse des correspondances 32
 - fonctions supplémentaires de la commande 35
- analyse des correspondances multiples 19, 22, 37, 39
 - bootstrap 25
 - enregistrement de variables 24, 41
 - niveau de codage optimal 20, 38
- Analyse des correspondances multiples
 - fonctions supplémentaires de la commande 26
- ANOVA
 - dans la régression catégorielle 17

B

- bootstrap
 - analyse des correspondances multiples 25

C

- centroïdes
 - dans l'analyse de corrélation canonique non linéaire 29
- chargements entre composantes
 - dans l'analyse de corrélation canonique non linéaire 29
 - dans l'analyse des correspondances multiples 23
- coefficients de régression
 - dans la régression catégorielle 17
- configuration initiale
 - dans l'analyse de corrélation canonique non linéaire 29
 - dans la régression catégorielle 15
 - dans le dépliage
 - multidimensionnel 53
 - dans le positionnement
 - multidimensionnel 47

coordonnées de l'espace commun

- dans le dépliage
 - multidimensionnel 55
 - dans le positionnement
 - multidimensionnel 49
- coordonnées de l'espace individuel
- dans le dépliage
 - multidimensionnel 55

corrélations

- dans le positionnement
 - multidimensionnel 49

critères d'itération

- dans le dépliage
 - multidimensionnel 53
- dans le positionnement
 - multidimensionnel 47

D

- dépliage multidimensionnel 51
 - fonctions supplémentaires de la commande 55
 - modèle 52
 - options 53
 - restrictions sur l'espace commun. 52
 - sortie 55
 - statistiques 51
 - tracés 51, 54
- dimensions
 - dans l'analyse des correspondances 33
- discrétisation
 - dans l'analyse des correspondances multiples 21
 - dans la régression catégorielle 14
 - dans une analyse de correspondance multiple 38
- distances
 - dans le positionnement
 - multidimensionnel 49
- Distances
 - dans le dépliage
 - multidimensionnel 55

E

- elastic net
 - dans la régression catégorielle 16

G

- graphiques triples
 - dans l'analyse des correspondances multiples 24

H

- historique des itérations
 - dans l'analyse des correspondances multiples 23
 - dans le dépliage
 - multidimensionnel 55
 - dans le positionnement
 - multidimensionnel 49
 - dans une analyse de correspondance multiple 40

I

- inertie
 - dans l'analyse des correspondances 34

J

- joindre les tracés de catégories
 - dans l'analyse des correspondances multiples 25
 - dans une analyse de correspondance multiple 41

L

- lasso
 - dans la régression catégorielle 16

M

- matrice de corrélation
 - dans l'analyse des correspondances multiples 23
 - dans une analyse de correspondance multiple 40
- mesures de contrainte
 - dans le dépliage
 - multidimensionnel 55
 - dans le positionnement
 - multidimensionnel 49
- mesures de discrimination
 - dans une analyse de correspondance multiple 40
- mesures de distance
 - dans l'analyse des correspondances 33
- mises à jour relaxées
 - dans le positionnement
 - multidimensionnel 47
- modèle d'identité
 - dans le dépliage
 - multidimensionnel 52
- modèle Euclidien généralisé
 - dans le dépliage
 - multidimensionnel 52

- modèle Euclidien pondéré
 - dans le dépliage
 - multidimensionnel 52
- modèles de positionnement
 - dans le dépliage
 - multidimensionnel 52

N

- niveau de codage optimal
 - dans l'analyse des correspondances multiples 20
 - dans une analyse de correspondance multiple 38
- normalisation
 - dans l'analyse des correspondances 33
- nuage de points de l'ajustement
 - dans le dépliage
 - multidimensionnel 54

O

- objets supplémentaires
 - dans la régression catégorielle 15

P

- pondération des variables
 - dans l'analyse des correspondances multiples 20
 - dans une analyse de correspondance multiple 38
- pondérations
 - dans l'analyse de corrélation canonique non linéaire 29
- pondérations des espaces individuels
 - dans le dépliage
 - multidimensionnel 55
 - dans le positionnement
 - multidimensionnel 49
- positionnement multidimensionnel 43, 44, 45
 - modèle 46
 - options 47
 - restrictions 47
 - sortie 49
 - statistiques 43
 - tracés 43, 48
- Positionnement multidimensionnel
 - fonctions supplémentaires de la commande 49
- PREFSCAL 51
- proximités transformées
 - dans le dépliage
 - multidimensionnel 55
 - dans le positionnement
 - multidimensionnel 49

R

- R multiple
 - dans la régression catégorielle 17
- régression catégorielle 13
 - enregistrement 17

- régression catégorielle (*suite*)
 - niveau de codage optimal 14
 - régularisation 16
 - statistiques 13
 - tracés 13
- Régression catégorielle
 - fonctions supplémentaires de la commande 18
- régression de crête
 - dans la régression catégorielle 16
- restrictions
 - dans le positionnement
 - multidimensionnel 47
- restrictions sur l'espace commun.
 - dans le dépliage
 - multidimensionnel 52

S

- scores des objets
 - dans l'analyse de corrélation canonique non linéaire 29
 - dans l'analyse des correspondances multiples 23
 - dans une analyse de correspondance multiple 40
- standardisation
 - dans l'analyse des correspondances 33
- statistiques de confiance
 - dans l'analyse des correspondances 34
- statistiques descriptives
 - dans la régression catégorielle 17

T

- terme de pénalité
 - dans le dépliage
 - multidimensionnel 53
- tracés
 - dans l'analyse de corrélation canonique non linéaire 29
 - dans l'analyse des correspondances 34
 - Dans la régression catégorielle 18
 - dans le positionnement
 - multidimensionnel 48
- tracés à départs multiples
 - dans le dépliage
 - multidimensionnel 54
- tracés d'espace commun
 - dans le dépliage
 - multidimensionnel 54
 - dans le positionnement
 - multidimensionnel 48
- tracés d'espace individuel
 - dans le dépliage
 - multidimensionnel 54
 - dans le positionnement
 - multidimensionnel 48
- tracés de catégories
 - dans l'analyse des correspondances multiples 25
 - dans une analyse de correspondance multiple 41

- tracés de centroïdes projetés
 - dans l'analyse des correspondances multiples 25
- Tracés de chargements
 - Dans l'analyse des correspondances multiples 25
- tracés de contrainte
 - dans le dépliage
 - multidimensionnel 54
 - dans le positionnement
 - multidimensionnel 48
- tracés de corrélations
 - dans le positionnement
 - multidimensionnel 48
- tracés de l'espace commun final
 - dans le dépliage
 - multidimensionnel 54
- tracés de l'espace commun initial
 - dans le dépliage
 - multidimensionnel 54
- tracés de mesures de discrimination
 - dans une analyse de correspondance multiple 41
- tracés de points d'objet
 - dans l'analyse des correspondances multiples 24
 - dans une analyse de correspondance multiple 41
- tracés de pondération d'espace individuel
 - dans le dépliage
 - multidimensionnel 54
 - dans le positionnement
 - multidimensionnel 48
- tracés de pondération des espaces
 - dans le dépliage
 - multidimensionnel 54
- Tracés de Shepard
 - dans le dépliage
 - multidimensionnel 54
- tracés de transformation
 - dans l'analyse des correspondances multiples 25
 - dans le dépliage
 - multidimensionnel 54
 - dans le positionnement
 - multidimensionnel 48
 - dans une analyse de correspondance multiple 41
- tracés doubles
 - dans l'analyse des correspondances 34
 - dans l'analyse des correspondances multiples 24
 - dans une analyse de correspondance multiple 41
- tracés résiduels
 - dans le dépliage
 - multidimensionnel 54
- transformations de proximité
 - dans le dépliage
 - multidimensionnel 52

V

- valeurs affectées aux catégories
 - dans l'analyse de corrélation canonique non linéaire 29

valeurs affectées aux catégories (*suite*)
dans l'analyse des correspondances
multiples 23
dans la régression catégorielle 17
dans une analyse de correspondance
multiple 40
valeurs manquantes
dans l'analyse des correspondances
multiples 21
dans la régression catégorielle 15
dans une analyse de correspondance
multiple 38
variables indépendantes transformées
dans le positionnement
multidimensionnel 49
variance expliquée par
dans l'analyse des correspondances
multiples 23

