IBM SPSS Statistics Base 32

Remarque

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 329.

Notice d'édition

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

© Copyright International Business Machines Corporation .

Table des matières

Chapitre 1. Fonctions de base 1
Analyse de puissance 1
Moyennes 2
Proportions 17
Corrélations 28
Régression 36
Analyse de l'alimentation: Précision 39
Analyse de puissance : Valeurs de grille 40
Méta-analyse 40
Méta-analyse continue 41
Méta-analyse : Taille d'effet continue 49
Méta-analyse binaire 58
Méta-analyse : Taille d'effet binaire 67
Régression de la méta-analyse 75
Options de traçage de la méta-analyse 80
Livre de codes 86
Onglet Sortie du livre de codes 87
Onglet Statistiques du livre de codes 89
Fréquences 89
Statistiques des effectifs 90
Graphiques des effectifs 91
Format des effectifs 92
Descriptives 92
Options Descriptives 93
Fonctions supplémentaires de la commande DESCRIPTIVES 93
Percentiles 94
Percentiles Valeurs manquantes 94
Explorer 95
Statistiques d'Explorer 96
Tracés d'Explorer 96
Options d'Explorer 97
Fonctions supplémentaires de la commande EXAMINE 97
tableaux croisés 98
Couches de tableaux croisés 99
Graphiques à barres en cluster de tableaux croisés 99
Tableaux croisés affichant les variables de couche dans des couches de tableau 99
Statistiques de tableaux croisés 99
Affichage de cellules (cases) de tableaux croisés 101
Format de tableau croisé 102
Récapituler 102
Options de Récapituler 103
Statistiques de Récapituler 103
Moyennes 105
Options de Moyennes 105
Cubes OLAP 107
Statistiques des Cubes OLAP 108
Différences des Cubes OLAP 109
Titre des Cubes OLAP 110
Proportions 110
Introduction aux proportions 110
Proportions d'échantillons uniques 110

Proportions d'échantillons appariés....113
Proportions d'échantillons indépendants....116
Tests T....119
Tests T....119
Test T pour échantillons indépendants....119
Test T pour échantillons appariés....121
Test T pour échantillon unique....123
Fonctions supplémentaires de la commande T TEST....124
ANOVA à 1 facteur....124
Contrastes ANOVA à 1 facteur....125
Tests Post Hoc ANOVA à 1 facteur....126
Options ANOVA à 1 facteur....128
Fonctions supplémentaires de la commande ONEWAY....129
Analyse GLM – Univarié....129
Modèle GLM....130
Contrastes GLM....132
Tracés de profil GLM....132
Comparaisons post hoc GLM....135
Enregistrement GLM....137
Moyenne marginale estimée GLM....138
Options GLM....139
Fonctions supplémentaires de la commande UNIANOVA....140
Bland-Altman....141
Bland Altman : Critères....141
Bland Altman : Imprimer....142
Bland Altman : Plot....142
Test du khi-carré....142
Test du chi carré : Exact....142
Test du chi carré : Cellules....143
Test du chi carré : Format....143
Corrélations bivariées....143
Options de corrélations bivariées....145
Intervalle de confiance des corrélations bivariées....145
Fonctions supplémentaires des commandes CORRELATIONS et NONPAR CORR....146
Mappage de proximité....146
Cartographie de proximité : Données....148
Cartographie de proximité : Poids....149
Cartographie de proximité : Initiale....149
Cartographie de proximité : Modèle....150
Cartographie de proximité : Restrictions....152
Cartographie de proximité : Complémentaire....153
Cartographie de proximité : Critères....153
Cartographie de proximité : Sortie....155
Cartographie de proximité : Sauvegarder....160
Corrélations partielles....161
Options Corrélations partielles....162
Fonctions supplémentaires de la commande PARTIAL CORR....162
Distances....163
Distances : Mesures de dissimilarité....163
Indices : Mesures de similarité....164
Fonctions supplémentaires de la commande PROXIMITIES....164
Corrélation de distance....164
Corrélation de distance : Critères....165
Corrélation de distance : Imprimer....166
Corrélation de distance : Tracé....167
Modèles linéaires....167
Obtention d'un modèle linéaire....168
Objectifs....168

Bases ... 169
Choix du modèle ... 169
Ensembles ... 170
Avancé ... 170
Options de modèle ... 170
Récapitulatif de modèle ... 171
Préparation automatique des données ... 171
Importance des prédicteurs ... 171
Valeurs prévues en fonction des valeurs observées ... 172
Résidus ... 172
Valeurs extrêmes ... 172
Effets ... 172
Coefficients ... 173
Moyennes estimées ... 173
Récapitulatif de génération de modèle ... 173
Régression linéaire ... 174
Méthodes de sélection des variables de régression linéaire ... 175
Régression linéaire : Définir la règle ... 175
Tracés de régression linéaire ... 176
Régression linéaire : Enregistrement des nouvelles variables ... 176
Statistiques de régression linéaire ... 178
Régression linéaire : Options ... 178
Fonctions supplémentaires de la commande REGRESSION ... 179
Régression ordinale ... 179
Régression ordinale : Options ... 180
Régression ordinale : Sortie ... 181
Régression ordinale : Emplacement ... 182
Régression ordinale : Echelle ... 182
Fonctions supplémentaires de la commande PLUM ... 183
Régression nette élastique linéaire ... 183
Régression linéaire élastique nette: Options ... 184
Régression Lasso linéaire ... 187
Régression lasso linéaire: Options ... 187
Régression Ridge linéaire ... 190
Régression linéaire de crête: Options ... 190
Analyse de la médiation ... 193
Type de modèle ... 198
Variables dépendantes ... 199
Prédicteur ... 200
Critères ... 201
Imprimer ... 203
Tracés ... 204
Filtres de séries temporelles ... 205
Filtres de séries temporelles : Options ... 206
Estimation de courbe ... 207
Modèles d'estimation de courbe ... 208
Enregistrement de l'estimation de courbe ... 208
Régression aux moindres carrés partiels ... 209
Modèle ... 210
Options ... 210
Analyse du voisin le plus proche ... 211
Voisins ... 213
Fonctions ... 214
Partitions ... 214
Enregistrement ... 215
Sortie ... 215
Options ... 216
Vue du modèle ... 216

Analyse discriminante ... 219
Définition de plages pour l'analyse discriminante ... 220
Sélection des observations pour l'analyse discriminante ... 220
Statistiques de l'analyse discriminante ... 221
Méthode détaillée étape par étape de l'analyse discriminante ... 221
Classement de l'analyse discriminante ... 222
Enregistrement de l'analyse discriminante ... 223
Fonctions supplémentaires de la commande DISCRIMINANT ... 223
Analyse factorielle ... 223
Sélection des observations pour l'analyse factorielle ... 224
Caractéristiques d'analyse factorielle ... 224
Extraction d'analyse factorielle ... 225
Rotation d'analyse factorielle ... 226
Scores d'analyse factorielle ... 226
Options d'analyse factorielle ... 227
Fonctions supplémentaires de la commande FACTOR ... 227
Choix d'une procédure de classification ... 227
Analyse de cluster TwoStep ... 228
Options de la procédure d'analyse de cluster TwoStep ... 229
Sortie de l'analyse de cluster TwoStep ... 231
Visualiseur de clusters ... 231
Analyse de cluster hiérarchique ... 236
Méthode d'analyse de cluster hiérarchique ... 237
Statistiques de l'analyse de cluster hiérarchique ... 238
Tracés (graphiques) de l'analyse de cluster hiérarchique ... 238
Sauvegarde des nouvelles variables de l'analyse de cluster hiérarchique ... 238
Fonctions supplémentaires de la syntaxe de commande CLUSTER ... 238
Analyse de cluster de nuées dynamiques ... 238
Efficacité de l'analyse de cluster de nuées dynamiques ... 239
Itération de l'analyse de cluster de nuées dynamiques ... 240
Enregistrement des analyses de cluster de nuées dynamiques ... 240
Options d'analyses de cluster de nuées dynamiques ... 240
Fonctions supplémentaires de la commande QUICK CLUSTER ... 241
Tests non paramétriques ... 241
Tests non paramétriques à un échantillon ... 241
Test de Wilcoxon pour échantillon unique ... 245
tests non paramétriques pour échantillons indépendants ... 246
Tests non paramétriques pour échantillons liés ... 248
Fonctions supplémentaires de la commande NPTESTS ... 251
Boîtes de dialogue ancienne version ... 251
Analyse des réponses multiples ... 263
Analyse des réponses multiples ... 263
Définition de jeux de réponses multiples ... 263
Tableaux de fréquences des réponses multiples ... 264
Tableaux croisés des réponses multiples ... 265
Rapports de résultats ... 267
Rapports de résultats ... 267
Récapitulatifs du rapport en lignes ... 267
Récapitulatifs du rapport en colonnes ... 270
Fonctions supplémentaires de la commande REPORT ... 272
Analyse de fiabilité ... 272
Analyse de fiabilité : Statistiques ... 274
Fonctions supplémentaires de la commande RELIABILITY ... 276
Kappa pondéré ... 276
Kappa pondéré : Critères ... 277
Kappa pondéré : Imprimer ... 278
Positionnement multidimensionnel ... 278
Forme des données du positionnement multidimensionnel ... 280

Positionnement multidimensionnel : créer une mesure....280
Modèle de positionnement multidimensionnel....280
Positionnement multidimensionnel : Options....281
Fonctions supplémentaires de la commande ALSCAL....281
Comparaison entre ALSCAL et PROXMAP....281
Statistiques de rapport....287
Statistiques de rapport : Statistiques....287
Tracés P-P....289
Tracé Q-Q....291
Analyse ROC....294
Analyse ROC : Options....295
Analyse ROC : Affichage....296
Analyse ROC : Définir des groupes (chaîne)....297
Analyse ROC : Définir des groupes (numérique)....297
Courbe ROC....297
Options de la courbe ROC....298
Simulation....298
Conception d'une simulation basée sur un fichier de modèle....299
Pour concevoir une simulation basée sur des équations personnalisées....300
Pour concevoir une simulation sans modèle prédictif....300
Exécution d'une simulation à partir d'un plan de simulation....301
Générateur de simulation....301
Boîte de dialogue Exécuter la simulation....313
Utilisation de la sortie graphique créée par la simulation....315
Modélisation géospatiale....317
Sélection de cartes....317
Sources de données....319
Règles d'association géospatiales....320
Prévision spatio-temporelle....324
Terminer....328

Remarques....329
Marques....330

Index....333

Chapitre 1. Fonctions de base

Les fonctions fondamentales suivantes sont incluses dans IBM SPSS Statistiques Base Edition.

Analyse de puissance

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

IBM SPSS Statistiques fournit les procédures d'analyse de puissance suivantes :

Test T sur échantillon unique
Dans les analyses à un échantillon, les données observées sont collectées sous forme d'échantillon aléatoire unique. On considère que les données de chaque échantillon suivent une distribution normale de façon indépendante et identique avec une moyenne et une variance fixes, et établissent l'inférence statistique sur le paramètre moyen.

Test T pour échantillon apparié
Dans une analyse à échantillon apparié, les données observées contiennent deux échantillons appariés et corrélés, et chaque observation possède deux mesures. On considère que les données de chaque échantillon suivent une distribution normale de façon indépendante et identique avec une moyenne une variance fixes, et établissent l'inférence statistique sur la différence des deux moyennes.

Test T pour échantillon indépendant
Dans une analyse à échantillon indépendant, les données observées contiennent deux échantillons indépendants. On considère que les données de chaque échantillon suivent une distribution normale de façon indépendante et identique avec une moyenne une variance fixes, et établissent l'inférence statistique sur la différence des deux moyennes.

ANOVA unidirectionnelle
L'analyse de variance (ANOVA) est une méthode statistique utilisée pour évaluer les moyennes de plusieurs populations qui sont souvent considérées comme étant distribuées suivant une loi normale. L'ANOVA à 1 facteur, qui est un type de variance très commun, est en réalité une extension du test *t* à deux échantillons.

Exemple. La puissance d'un test d'hypothèse pour détecter une hypothèse alternative correcte est la probabilité selon laquelle le test va rejeter l'hypothèse du test. Sachant que la probabilité d'une erreur de type II est la probabilité d'accepter l'hypothèse nulle alors que l'hypothèse alternative est vraie, la puissance peut être exprimée sous la forme (1-probabilité d'une erreur de type II), qui représente la probabilité que l'on rejette l'hypothèse nulle alors que l'hypothèse alternative est vraie.

Tracés et statistiques : Test unilatéral, test bilatéral, niveau d'importance, taux d'erreur de type I, hypothèses de test, écart type de population, moyenne de population sous test, valeur hypothétique, puissance bidimensionnelle à l'aide d'une taille d'échantillon, puissance bidimensionnelle à l'aide d'une taille d'effet, puissance tridimensionnelle à l'aide d'une taille d'échantillon, puissance tridimensionnelle à l'aide d'une taille d'effet, degrés de rotation, paires de groupes, coefficient de corrélation produit-moment de Pearson, différence moyenne, plage de tracés de la taille d'effet, écart type de la population combinée, différences par paires et contrastes, coefficients de contraste, test de contraste, BONFERRONI, SIDAK, LSD, puissance à l'aide de la taille de l'échantillon complet, puissance bidimensionnelle à l'aide d'un écart type combiné, puissance tridimensionnelle à l'aide de l'échantillon complet, puissance tridimensionnelle à l'aide de la taille de l'échantillon complet, écart type combiné, distribution t de Student, distribution t non centrale.

Remarques sur les données dans l'analyse de puissance

Données

Dans les analyses à un échantillon, les données observées sont collectées sous forme d'échantillon aléatoire unique.

Dans une analyse à échantillon apparié, les données observées contiennent deux échantillons appariés et corrélés, et chaque observation possède deux mesures.

Dans une analyse à échantillon indépendant, les données observées contiennent deux échantillons indépendants.

L'analyse de variance (ANOVA) est une méthode statistique utilisée pour évaluer les moyennes de plusieurs populations qui sont souvent considérées comme étant distribuées suivant une loi normale.

Hypothèses

Dans une analyse à un échantillon, on considère que les données d'échantillon suivent une distribution normale de façon indépendante et identique avec une moyenne et une variance fixes, et établissent l'inférence statistique sur le paramètre moyen.

Dans une analyse à échantillons appariés, on considère que les données de chaque échantillon suivent une distribution normale de façon indépendante et identique avec une moyenne et une variance fixes, et établissent l'inférence statistique sur les deux moyennes.

Dans une analyse à échantillons indépendants, on considère que les données de chaque échantillon suivent une distribution normale de façon indépendante et identique avec une moyenne et une variance fixes, et établissent l'inférence statistique sur la différence des deux moyennes.

Dans une analyse ANOVA à 1 facteur, la méthode statistique utilisée pour estimée les moyennes de plusieurs populations sont souvent considérées comme étant distribuées suivant une loi normale.

Obtention d'une analyse de puissance

1. A partir des menus, sélectionnez :

   Analyse > **Analyse de puissance** > **Comparer les moyennes** > **Test T sur échantillon unique** ou **Test T pour échantillons appariés** ou **Test T pour échantillons indépendants** ou **ANOVA à 1 facteur**
2. Définissez les hypothèses de test requises.
3. Cliquez sur **OK**.

Moyennes

Les fonctions statistiques ci-après sont incluses dans IBM SPSS Statistiques Base Edition.

Analyse de puissance du test T pour échantillon unique

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

Dans les analyses à un échantillon, les données observées sont collectées sous forme d'échantillon aléatoire unique. On considère que les données de chaque échantillon suivent une distribution normale de façon indépendante et identique avec une moyenne et une variance fixes, et établissent l'inférence statistique sur le paramètre moyen.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Moyennes** > **Test T pour échantillon unique**

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

2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.
4. Lorsque **Puissance** est sélectionné comme paramètre **Estimation** de l'hypothèse de test, entrez une valeur d'estimation **Taille de l'échantillon** pour la puissance. La valeur doit être un entier supérieur à 1.
5. Vous pouvez également sélectionner une option dans le champ **Spécifier**.

   Valeurs hypothétiques
   Le paramètre par défaut fournit les paramètres **Moyenne de population** et **Valeur NULL**.

   Moyenne de population
   Entrez une valeur indiquant la moyenne de la population à tester. La valeur doit être une valeur numérique unique.

   Valeur NULL
   Si vous le souhaitez, entrez une valeur qui spécifie la valeur d'hypothèse NULL à tester. La valeur doit être une valeur numérique unique.

   Taille d'effet
   Le f de Cohen est utilisé pour estimer la taille de l'effet comme donnée d'entrée pour l'estimation de la puissance ou de la taille de l'échantillon. La **Valeur** de la taille d'effet définie est passée à l'étape intermédiaire de la procédure et calcule la puissance souhaitée ou la taille de l'échantillon.
6. Saisissez une valeur **Ecart type de population**. La valeur doit être une valeur numérique unique supérieure à 0.
7. Indiquez si le test est unilatéral ou bilatéral.

   Analyse non directionnelle (bilatérale)
   Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

   Analyse directionnelle (unilatérale)
   Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.
8. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
9. Vous pouvez éventuellement cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test T pour échantillon unique : tracé», à la page 4 (sortie de graphique, paramètres de tracé bidimensionnel et paramètres de tracé tridimensionnel).

   Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné comme hypothèse de test.
10. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode **Estimation** de l'hypothèse de test, **Valeurs hypothétiques** est sélectionné dans la liste **Spécifier** et l'analyse **Non directionnelle (bilatérale)** est sélectionnée comme **Direction de test**.

Analyse de puissance du test T pour échantillon unique : tracé

La boîte de dialogue **Tracé** fournit des options permettant de contrôler les tracés qui sont produits pour illustrer les graphiques de puissance par taille d'échantillon et d'effet bidimensionnels et tridimensionnels. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel

Estimation de puissance vs taille de l'échantillon
Si ce paramètre facultatif est activé, il propose des options permettant de contrôler le graphique bidimensionnel de puissance par taille d'échantillon. Il est désactivé par défaut.

Plage de la taille de l'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'échantillon est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de puissance par taille d'échantillon. Cette valeur doit être supérieure à 1, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Estimation de puissance vs taille d'effet
Par défaut, ce paramètre facultatif est désactivé. Lorsqu'il est activé, le graphique s'affiche dans la sortie. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut de la taille d'effet est utilisée.

Plage de la taille d'effet
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles.

Limite inférieure
Contrôle la limite inférieure de la puissance bidimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure ou égale à -5.0, sans dépasser la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure de la puissance bidimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**, sans dépasser 5.0.

Tracé tridimensionnel

Estimation de puissance vs
Propose des options permettant de contrôler le graphique tridimensionnel de puissance par taille d'échantillon (axe des X) et taille d'effet (axe des Y), les paramètres de rotation verticale/horizontale, ainsi que la courbe spécifiée par l'utilisateur pour la taille d'échantillon/d'effet. Ce paramètre est désactivé par défaut.

Taille d'effet sur l'axe des X et taille d'échantillon sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des X) et taille d'effet (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Taille d'effet sur l'axe des Y et taille d'échantillon sur l'axe des X
Paramètre facultatif contrôlant la puissance tridimensionnelle à l'aide d'un graphique de taille d'échantillon (axe des Y) et d'un graphique de taille d'effet (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Plage de la taille de l'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'échantillon est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de puissance par taille d'échantillon. Cette valeur doit être supérieure à 1, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Plage de la taille d'effet
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'effet est utilisée.

Limite inférieure
Contrôle la limite inférieure de la puissance tridimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure ou égale à -5.0, sans dépasser la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure de la puissance tridimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**, sans dépasser 5.0.

Analyse de puissance du test T pour échantillons appariés

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

Dans une analyse à échantillons appariés, les données observées contiennent deux échantillons appariés et corrélés et chaque observation possède deux mesures. On considère que les données de chaque échantillon suivent une distribution normale de façon indépendante et identique avec une moyenne une variance fixes, et établissent l'inférence statistique sur la différence des deux moyennes.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Moyennes** > **Test T pour échantillons appariés**

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

2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.
4. Lorsque **Puissance** est sélectionné comme méthode d'**Estimation** de l'hypothèse de test, entrez la valeur **Taille de l'échantillon** appropriée. Il doit s'agir d'un entier positif supérieur ou égal à 2.
5. Vous pouvez également sélectionner une option dans le champ **Spécifier**.

   Valeurs hypothétiques
   Le paramètre par défaut fournit les paramètres **Moyenne** et **Écart type**.

   Différence moyenne de population
   Lorsqu'une seule moyenne de population est requise, entrez la valeur **Différence moyenne de la population**. Lorsqu'une valeur unique est renseignée, elle représente la différence moyenne de la population μ_d.

   Remarque : La valeur ne peut pas être 0 lorsque **Taille de l'échantillon** est sélectionné.

   Moyenne de population pour le groupe 1 et le groupe 2
   Lorsque plusieurs moyennes de population sont requises pour les paires de groupes spécifiées, entrez des valeurs pour **Moyenne de population pour le groupe 1** et **Moyenne de population pour le groupe 2**. Lorsque plusieurs valeurs sont spécifiées, elles indiquent la différence moyenne de population μ_1 et μ_2.

   Remarque : Les deux valeurs ne peuvent pas être identiques lorsque **Taille de l'échantillon** est sélectionné.

   Écart type de population pour la différence moyenne
   Lorsqu'une seule moyenne de population est spécifiée, entrez l'écart type de population pour la valeur de différence moyenne. Lorsqu'une valeur unique est renseignée, elle représente l'écart type de population de la différence de groupe σ_d. La valeur doit être une valeur numérique unique supérieure à 0.

   Écart type de population pour le groupe 1 et le groupe 2
   Lorsque plusieurs moyennes de population sont spécifiées, entrez l'écart type de population pour les valeurs du groupe 1 et du groupe 2. Lorsque des valeurs multiples sont renseignées, elles représentent l'écart type de population de la différence de groupe σ_1 et σ_2. Les valeurs soivent être des valeurs numériques uniques supérieures à 0.

   Coefficient de corrélation produit-moment de Pearson
   Le cas échéant, vous pouvez entrer une valeur qui indique le coefficient de corrélation produit-moment de Pearson ρ. La valeur doit être une valeur numérique unique comprise entre -1 et 1. Cette valeur ne peut pas être 0.

   Remarque : Lorsqu'une valeur **Ecart type de population pour la différence moyenne** unique est spécifiée, ce paramètre est ignoré. Sinon, les valeurs de l'**Écart type de population pour le groupe 1 et le groupe 2** sont utilisées pour calculer σ_d.

   Taille d'effet
   Estime la taille d'effet comme une entrée de l'estimation de la puissance ou de la taille de l'échantillon. La **Valeur** de la taille d'effet définie est passée à l'étape intermédiaire de la procédure et calcule la puissance souhaitée ou la taille de l'échantillon.
6. Indiquez si le test est unilatéral ou bilatéral.

   Analyse non directionnelle (bilatérale)
   Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

Analyse directionnelle (unilatérale)
Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.

7. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
8. Vous pouvez également cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test T pour échantillons appariés : tracé», à la page 7 (sortie de graphique, paramètres de tracé en deux dimensions, paramètres de tracé en trois dimensions et infobulles).

   Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné comme hypothèse de test.
9. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode d'**Estimation** de l'hypothèse de test et que **Valeurs hypothétiques** est sélectionné dans la liste **Spécifier**.

Analyse de puissance du test T pour échantillons appariés : tracé

La boîte de dialogue **Tracé** fournit des options permettant de contrôler les tracés qui sont produits pour illustrer les graphiques de puissance par taille d'échantillon et d'effet bidimensionnels et tridimensionnels. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel

Estimation de puissance vs taille de l'échantillon
Si ce paramètre facultatif est activé, il propose des options permettant de contrôler le graphique bidimensionnel de puissance par taille d'échantillon. Il est désactivé par défaut.

Plage de la taille de l'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'échantillon est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de puissance par taille d'échantillon. Cette valeur doit être supérieure à 1, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Estimation de puissance vs taille d'effet
Par défaut, ce paramètre facultatif est désactivé. Lorsqu'il est activé, le graphique s'affiche dans la sortie. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut de la taille d'effet est utilisée.

Plage de la taille d'effet
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles.

Limite inférieure
Contrôle la limite inférieure de la puissance bidimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure ou égale à -5.0, sans dépasser la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure de la puissance bidimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**, sans dépasser 5.0.

Tracé tridimensionnel

Estimation de puissance vs
Propose des options permettant de contrôler le graphique tridimensionnel de puissance par taille d'échantillon (axe des X) et taille d'effet (axe des Y), les paramètres de rotation verticale/horizontale, ainsi que la courbe spécifiée par l'utilisateur pour la taille d'échantillon/d'effet. Ce paramètre est désactivé par défaut.

Taille d'effet sur l'axe des X et taille d'échantillon sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des X) et taille d'effet (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Taille d'effet sur l'axe des Y et taille d'échantillon sur l'axe des X
Paramètre facultatif contrôlant la puissance tridimensionnelle à l'aide d'un graphique de taille d'échantillon (axe des Y) et d'un graphique de taille d'effet (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Plage de la taille de l'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'échantillon est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de puissance par taille d'échantillon. Cette valeur doit être supérieure à 1, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Plage de la taille d'effet
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'effet est utilisée.

Limite inférieure
Contrôle la limite inférieure de la puissance tridimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure ou égale à -5.0, sans dépasser la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure de la puissance tridimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**, sans dépasser 5.0.

Analyse de puissance du test T pour échantillons indépendants

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

Dans une analyse à échantillons indépendants, les données observées contiennent deux échantillons indépendants. On considère que les données de chaque échantillon suivent une distribution normale de façon indépendante et identique avec une moyenne une variance fixes, et établissent l'inférence statistique sur la différence des deux moyennes.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Moyennes** > **Test T pour échantillons indépendants**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.

   Entrez une valeur **Rapport de taille de groupe** pour spécifier le rapport des tailles d'échantillon (il doit s'agir d'une valeur unique comprise entre 0,01 et 100).
4. Lorsque **Puissance** est sélectionné comme méthode **Estimation** de l'hypothèse de test, entrez des valeurs pour spécifier la taille de l'échantillon pour **Taille de l'échantillon pour le groupe 1 et le groupe 2** à des fins de comparaison. Les valeurs doivent être des entiers supérieurs à 1.
5. Vous pouvez également sélectionner une option dans le champ **Spécifier**.

   Valeurs hypothétiques
   Le paramètre par défaut fournit les paramètres **Moyenne de population** et les **Écarts types de population**.

   Différence moyenne de population
   Lorsqu'une seule moyenne de population est spécifiée, entrez la valeur de la différence de moyenne de population. Lorsqu'une valeur unique est spécifiée, elle indique la population de la différence moyenne de groupe σ_D. La valeur doit être une valeur numérique unique supérieure à 0.

   Moyenne de population pour le groupe 1 et le groupe 2
   Lorsque plusieurs moyennes de population sont spécifiées, entrez la moyenne de population pour les valeurs du groupe 1 et du groupe 2. Lorsque plusieurs valeurs sont spécifiées, elles indiquent la moyenne de population de la différence de groupe σ_1 et σ_2. Les valeurs soivent être des valeurs numériques uniques supérieures à 0.

   Les écarts types de la population sont
   Indiquez sir les écarts types de population, sont **Identiques pour les deux groupes** ou **Différents pour les deux groupes**.

- Lorsque les écarts types de population sont identiques pour les deux groupes, saisissez une valeur **Ecart type combiné** qui représente σ et suppose que les variances des deux groupes sont identiques, ou $\sigma_1 = \sigma_2 = \sigma$.
- Lorsque les écarts types de population ne sont pas égaux pour deux groupes, entrez des valeurs pour **Écart type pour le groupe 1 et le groupe 2** qui indiquent σ_1 et σ_2.

  Remarque : Lorsque les valeurs d'**Écart type pour le groupe 1 et le groupe 2** sont identiques, elles sont traitées comme une valeur unique.

Taille d'effet
Estime la taille d'effet comme une entrée de l'estimation de la puissance ou de la taille de l'échantillon. La **Valeur** de la taille d'effet définie est passée à l'étape intermédiaire de la procédure et calcule la puissance souhaitée ou la taille de l'échantillon.

6. Indiquez si le test est unilatéral ou bilatéral.

   Analyse non directionnelle (bilatérale)
   Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

   Analyse directionnelle (unilatérale)
   Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.
7. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
8. Vous pouvez également cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test T pour échantillons indépendants : tracé», à la page 10 (sortie de graphique, paramètres de tracé en deux dimensions, paramètres de tracé en trois dimensions et infobulles).

   Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné comme hypothèse de test.
9. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode **Estimation** de l'hypothèse de test, **Valeurs hypothétiques** est sélectionné dans la liste **Spécifier** et l'analyse **Non directionnelle (bilatérale)** est sélectionnée comme **Direction de test**.

Analyse de puissance du test T pour échantillons indépendants : tracé

La boîte de dialogue **Tracé** fournit des options permettant de contrôler les tracés qui sont produits pour illustrer les graphiques bidimensionnels et tridimensionnels de puissance par rapport d'échantillon, taille d'effet ou différence moyenne. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel

Estimation de puissance vs rapport de taille de l'échantillon/Estimation de puissance vs taille de l'échantillon
Si ce paramètre facultatif est activé, il propose des options permettant de contrôler le graphique bidimensionnel de puissance par ration de taille d'échantillon. Les paramètres supplémentaires ne sont pas disponibles lorsque plusieurs valeurs de puissance ont été spécifiées (**Estimation de puissance vs taille de l'échantillon**). Il est désactivé par défaut.

Plage du rapport de taille de l'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'échantillon est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de puissance par ratio de taille d'échantillon. Cette valeur doit être comprise entre 0,01 et 100, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de puissance par ratio de taille d'échantillon. Cette valeur doit être comprise entre 0,01 et 100, mais supérieure à la valeur de la zone **Limite inférieure**.

Estimation de puissance vs taille d'effet (ou différence moyenne)
Par défaut, ce paramètre facultatif est désactivé. Lorsqu'il est activé, le graphique s'affiche dans la sortie. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut de la taille d'effet (ou différence moyenne) est utilisée.

Plage de la taille d'effet (ou différence moyenne)
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles.

Limite inférieure
Contrôle la limite inférieure de la puissance bidimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure ou égale à -5.0, sans dépasser la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure de la puissance bidimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**, sans dépasser 5.0.

Tracé tridimensionnel

Estimation de puissance vs
Propose des options permettant de contrôler le graphique tridimensionnel de puissance par ratio de taille d'échantillon (axe des X) et taille d'effet (axe des Y), les paramètres de rotation verticale/horizontale, ainsi que la courbe spécifiée par l'utilisateur pour la taille d'échantillon/d'effet. Ce paramètre est désactivé par défaut.

Taille d'effet (ou différence moyenne) sur l'axe des X et taille d'échantillon sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par ratio de taille d'échantillon (axe des X) et taille d'effet (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Taille d'effet (ou différence moyenne) sur l'axe des X et taille d'échantillon sur l'axe des Y
Paramètre facultatif contrôlant la puissance tridimensionnelle à l'aide d'un graphique de taille d'échantillon (axe des Y) et d'un graphique de taille d'effet (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Plage du rapport de taille de l'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'échantillon est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de puissance par ratio de taille d'échantillon. Cette valeur doit être comprise entre 0,01 et 100, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de puissance par ratio de taille d'échantillon. Cette valeur doit être comprise entre 0,01 et 100, mais supérieure à la valeur de la zone **Limite inférieure**.

Plage de la taille d'effet (ou différence moyenne)
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'effet est utilisée.

Limite inférieure
Contrôle la limite inférieure de la puissance tridimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure ou égale à -5.0, sans dépasser la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure de la puissance tridimensionnelle à l'aide d'un graphique de taille d'effet. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Analyse de puissance d'une ANOVA à 1 facteur

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

L'analyse de variance (ANOVA) est une méthode statistique utilisée pour évaluer les moyennes de plusieurs populations qui sont souvent considérées comme étant distribuées suivant une loi normale. L'ANOVA à 1 facteur, qui est un type de variance très commun, est en réalité une extension du test *t* à deux échantillons. La procédure fournit des approches permettant d'estimer la puissance de deux types d'hypothèses pour comparer les moyennes de groupes multiples, le test général et le test associé à des contrastes spécifiques. Le test général se concentre sur l'hypothèse nulle selon laquelle toutes les moyennes de groupes sont égales. Le test avec des contrastes divise les hypothèses ANOVA générales en petits groupes de moyennes plus descriptibles et plus utiles.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Moyennes** > **ANOVA unidirectionnel**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.
4. Vous pouvez également sélectionner une option dans le champ **Spécifier**.

   Valeurs hypothétiques
   Le paramètre par défaut fournit le paramètre **Écart type de la population regroupée en pool**. Entrez une valeur numérique unique supérieure à 0. La valeur par défaut est 1.

Taille de l'effet : f de Cohen
Le f de Cohen est utilisé pour estimer la taille de l'effet comme donnée d'entrée pour l'estimation de la puissance ou de la taille de l'échantillon. La **Valeur** de la taille d'effet définie est passée à l'étape intermédiaire de la procédure et calcule la puissance souhaitée ou la taille de l'échantillon.

Taille de l'effet : Eta-squared
Eta-squared (η^2) est utilisé pour estimer la taille de l'effet comme donnée d'entrée pour l'estimation de la puissance ou de la taille de l'échantillon. La **Valeur** de la taille d'effet définie est passée à l'étape intermédiaire de la procédure et calcule la puissance souhaitée ou la taille de l'échantillon.

5. Lorsque **Puissance** est sélectionné comme paramètre **Estimation** de l'hypothèse de test, indiquez les valeurs **Tailles de groupe** et **Moyennes de groupe**. Au moins deux valeurs de taille de groupe doivent être spécifiées, et chacune des valeurs de taille de groupe doit être supérieure ou égale à 2 (mais ne peut pas dépasser 2 147 483 647). Au moins deux valeurs de moyenne de groupes doivent également être spécifiées (le nombre de valeurs spécifiées doit être égal aux valeurs de taille de groupe).

   Lorsque **Taille de l'échantillon** est sélectionné comme paramètre **Estimation** de l'hypothèse de test, indiquez les valeurs **Pondération de groupe** (facultatif) et **Moyennes de groupe**. Les pondérations de groupe affectent les pondérations de taille de groupe lorsque **Puissance** est sélectionné.

   Remarque : Les paramètres **Pondérations de groupe** sont ignorés si des valeurs **Tailles de groupe** sont spécifiées.

   Cliquez sur **Ajouter** pour inclure des valeurs **Tailles de groupe**, **Pondération de groupe**et **Moyennes de groupe** supplémentaires. Cliquez sur **Supprimer** pour supprimer les valeurs existantes.
6. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
7. Vous pouvez également cliquer sur **Contraste** pour spécifier les paramètres «Analyse de puissance d'une ANOVA à 1 facteur : Contraste», à la page 13 (test de contraste, différences par paires, taille des effets et estimation de la taille de l'échantillon en fonction des intervalles de confiance).
8. Vous pouvez également cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance d'une ANOVA à 1 facteur : tracé», à la page 14 (sortie de graphique, paramètres de tracé en deux dimensions, paramètres de tracé en trois dimensions et infobulles).

   Remarque : Tracé est disponible uniquement lorsque les valeurs **Tailles de groupe** sont spécifiées et que **Puissance** est sélectionné comme paramètre **Estimation**.
9. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode **Estimation** de l'hypothèse de test.

Analyse de puissance d'une ANOVA à 1 facteur : Contraste

La boîte de dialogue **Contraste** fournit des options permettant de spécifier le contraste, le coefficient, la taille de l'effet, les différences par paire et les paramètres de demi-largeur de l'intervalle de confiance pour l'analyse de puissance de l'analyse ANOVA à 1 facteur.

Test de contraste

Test avec contrastes linéaires
Lorsque cette option est activée, les paramètres de contraste et de coefficient sont disponibles.

Direction du test

Analyse non directionnelle (bilatérale)
Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

Analyse directionnelle (unilatérale)
Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.

Coefficients
Utilisez le tableau pour indiquer les coefficients de contraste et demander un test de contraste. Les valeurs de tableau sont facultatives. Le nombre de valeurs spécifiées doit être identiques dans **Tailles de groupe** et **Moyennes de groupe**. La somme de l'ensemble des valeurs spécifiées doit être égale à 0, faute de quoi la dernière valeur est automatiquement ajustée.

Taille d'effet
Indiquez la taille d'effet du test de contraste mesurée par δ. Vous devez indiquer une valeur numérique unique. Lorsque **Direction du test** est défini sur **Analyse directionnelle (unilatérale)**, la valeur spécifiée doit être ≥ 0. Lorsque **Puissance** est spécifié comme hypothèse de test **Estimation**, la valeur spécifiée ne peut pas être 0.

Différences par paire

Estimer la puissance du test pour les différences par paire
Indique si la puissance de test pour les différences par paire doit être estimée ou non. Par défaut, le paramètre facultatif est désactivé, ce qui supprime la sortie des différences par paire.

Correction de Bonferroni
Utilise la correction de Bonferroni pour estimer la puissance des différences par paire. Il s'agit du paramètre par défaut.

Correction de Sidak
Utilise la correction Sidak pour estimer la puissance des différences par paire.

Différence la moins significative (LSD)
Utilise la correction LSD pour estimer la puissance des différences par paire.

Indiquez la demi-largeur de l'intervalle de confiance
Estime la taille de l'échantillon en fonction de la valeur de demi-largeur de l'intervalle de confiance. Entrez une valeur comprise entre 0 et 1. Pour le test binomial One-Sample, la valeur doit être comprise entre 0 et 0,5.

Remarque : Les valeurs en double sont ignorées.

- Cliquez sur **Ajouter** pour ajouter la valeur de demi-largeur spécifiée à la liste.
- Mettez en évidence une valeur de demi largeur existante et cliquez sur **Changement** pour mettre à jour la valeur.
- Mettez en évidence une valeur de demi-largeur existante et cliquez sur **Supprimer** pour supprimer la valeur de la liste.

Analyse de puissance d'une ANOVA à 1 facteur : tracé

La boîte de dialogue **Tracé** contrôle les tracés qui sont produits pour illustrer la puissance bidimensionnelle et tridimensionnelle par des graphiques de taille d'échantillon et d'effet. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel

Estimation de puissance vs taille de l'échantillon complet
Si ce paramètre facultatif est activé, il propose des options permettant de contrôler le graphique bidimensionnel de puissance par taille d'échantillon complet. Il est désactivé par défaut.

Plage de tailles de l'échantillon complet
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut de la taille de l'échantillon complet est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de puissance par taille d'échantillon complet. Cette valeur doit être supérieure ou égale à :

- 2 x le nombre d'entiers saisi dans le champ **Tailles de groupe**
- 2 x la somme des entiers saisie dans le champ **Tailles de groupe** / par la plus petite valeur entière spécifiée dans le champ **Tailles de groupe**

Cette valeur ne doit pas être supérieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de puissance par taille d'échantillon complet. Cette valeur doit être inférieure ou égale à :

- 5000 / par la plus grande valeur entière spécifiée dans le champ **Tailles de groupe** x la somme des entiers spécifiée dans le champ **Tailles de groupe**

Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 2147483647.

Estimation de puissance vs écart type combiné
Par défaut, ce paramètre facultatif est désactivé. Il contrôle le graphique bidimensionnel de puissance par écart type combiné. Lorsqu'il est activé, le graphique s'affiche dans la sortie. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut de l'écart type combiné est utilisée.

Remarque :

Le tracé est désactivé lorsque les valeurs **Moyennes de groupes** sont toutes identiques.

Plage de l'écart type combiné
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de puissance par écart type combiné. Cette valeur doit être supérieure à 0, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de puissance par écart type combiné. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**.

Tracé tridimensionnel

Estimation de puissance vs
Fournit des options pour le contrôle de la puissance tridimensionnelle par un graphique de taille totale d'échantillon (axe des x) et taille d'effet (axe des y), les paramètres de rotation verticale et horizontale, et la plage de tracé spécifiée par l'utilisateur de taille d'échantillon et d'effet. Ce paramètre est désactivé par défaut.

Remarque :

Le tracé est désactivé lorsque les valeurs **Moyennes de groupes** sont toutes identiques.

Ecart type combiné sur l'axe des X et taille de l'échantillon complet sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon complet (axe des X) et écart type combiné (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Ecart type combiné sur l'axe des Y et taille de l'échantillon complet sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon complet (axe des Y) et écart type combiné (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Plage de tailles de l'échantillon complet
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'échantillon est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de puissance par taille d'échantillon complet. Cette valeur doit être supérieure à 0, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de puissance par taille d'échantillon complet. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**.

Plage de l'écart type combiné
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'effet est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de puissance par écart type combiné. Cette valeur doit être supérieure à 0, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de puissance par écart type combiné. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**.

Estimation de puissance vs taille de l'échantillon
Fournit des options permettant de contrôler la puissance tridimensionnelle par taille d'effet et la plage de tracé spécifiée par l'utilisateur de la taille d'échantillon et d'effet. Ce paramètre est désactivé par défaut.

Remarque :

Le tracé est désactivé lorsque les valeurs **Moyennes de groupes** sont toutes identiques.

Ecart type combiné sur l'axe des X et taille de l'échantillon complet sur l'axe des Y
Le paramètre facultatif contrôle la puissance tridimensionnelle par la taille totale d'échantillon (axe des x) et la taille d'effet (axe des y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Ecart type combiné sur l'axe des Y et taille de l'échantillon complet sur l'axe des X
Le paramètre facultatif contrôle la puissance tridimensionnelle par la taille totale d'échantillon (axe des y) et la taille d'effet (axe des x). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Plage de tailles de l'échantillon complet
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'échantillon est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de puissance par taille d'échantillon complet. Cette valeur doit être supérieure à 0, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de puissance par taille d'échantillon complet. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**.

Plage de la taille d'effet
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les champs **Limite inférieure** ou **Limite supérieure**, la plage de tracés par défaut de la taille d'effet est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de puissance par taille d'échantillon complet. Cette valeur doit être supérieure à 0, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de puissance par taille d'échantillon complet. Cette valeur doit être supérieure à la valeur du champ **Limite inférieure**.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Proportions

Les fonctions statistiques ci-après sont incluses dans IBM SPSS Statistiques Base Edition.

Analyse de puissance du test binomial pour échantillon unique

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

La distribution binomiale est basée sur une séquence d'essais de Bernoulli. Elle permet de modéliser des expérimentations et notamment un nombre fixe d'essais totaux, supposés être indépendants les uns des autres. Chaque essai génère un résultat dichotomique, avec la même probabilité de réussite.

Le test binomial pour échantillon unique établit l'inférence statistique sur le paramètre de proportion en le comparant à une valeur hypothétique. Les méthodes d'estimation de la puissance d'un tel test sont l'approximation normale ou l'énumération binomiale.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Proportions** > **Test binomial pour échantillon unique**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et

1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.

4. Lorsque **Puissance** est sélectionné comme paramètre **Estimation** de l'hypothèse de test, entrez la valeur **Nombre total d'essais** appropriée. Il doit s'agir d'un entier supérieur ou égal à 1.
5. Saisissez une valeur qui spécifie la valeur d'hypothèse alternative du paramètre de proportion dans la zone **Proportion de la population**. La valeur doit être une valeur numérique unique.

   Remarque : Lorsqu'une valeur **Puissance** est spécifiée, la valeur **Proportion de la population** ne peut pas être égale à **Valeur NULL**.
6. Le cas échéant, vous pouvez entrer une valeur qui spécifie la valeur d'hypothèse nulle du paramètre de proportion à tester dans la zone **Valeur NULL**. La valeur doit être une valeur numérique unique comprise entre 0 et 1. La valeur par défaut est 0,50.
7. Sélectionnez une méthode d'estimation de la puissance.

   Approximation normale
   Active l'approximation normale. Il s'agit du paramètre par défaut.

   Appliquer la correction de continuité
   Détermine si la correction de continuité est utilisée ou non pour la méthode d'approximation normale.

   Enumération binomiale
   Active la méthode d'énumération binomiale. Vous pouvez également utiliser la zone **Limite de temps** pour indiquer le nombre maximum de minutes autorisées pour estimer la taille d'échantillon. Lorsque la limite de temps est atteinte, l'analyse se termine et un message d'avertissement s'affiche. Si vous indiquez cette valeur, il doit s'agir d'un entier positif unique indiquant le nombre de minutes. Le paramètre par défaut est 5 minutes.

   Remarque : L'hypothèse d'estimation de puissance sélectionnée n'a aucun effet lorsque la valeur **Nombre total d'essais** dépasse 500.
8. Indiquez si le test est unilatéral ou bilatéral.

   Analyse non directionnelle (bilatérale)
   Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

   Analyse directionnelle (unilatérale)
   Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.
9. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
10. Vous pouvez éventuellement cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test binomial pour échantillon unique : tracé», à la page 19 (sortie de graphique, paramètres de tracé bidimensionnel et paramètres de tracé tridimensionnel).

    Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné car l'hypothèse de test **Estimation** et **Énumération binomiale** n'est pas sélectionnée.
11. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode **Estimation** de l'hypothèse de test et que l'analyse **Non directionnel (bilatéral)** est sélectionnée comme **Direction de test**.

Analyse de puissance du test binomial pour échantillon unique : tracé

La boîte de dialogue **Tracé** fournit des options permettant de contrôler les tracés qui sont produits pour illustrer la puissance bidimensionnelle et tridimensionnelle par des graphiques. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel
Fournit des options permettant de contrôler les graphiques bidimensionnels de l'estimation de puissance. Ce paramètre est désactivé par défaut.

Estimation de puissance vs valeur d'hypothèse nulle
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par valeur Null. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs valeur d'hypothèse alternative
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par valeur alternative. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs différence entre les valeurs hypothétiques
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par différence entre les valeurs hypothétiques. Il est désactivé par défaut.

Estimation de puissance vs nombre total d'essais
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par nombre total d'essais. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe du nombre total d'essais
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par nombre total d'essais. Cette valeur doit être supérieure à 0, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par nombre total d'essais. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Tracé tridimensionnel
Fournit des options permettant de contrôler les graphiques tridimensionnels de l'estimation de la puissance. Ce paramètre est désactivé par défaut.

Estimation de puissance vs nombre total d'essais
Si ce paramètre est sélectionné, il active les options ci-après.

sur l'axe des X et différence entre les valeurs hypothétiques sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par nombre total d'essais (axe des X) et différence entre les valeurs hypothétiques (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

sur l'axe des Y et différence entre les valeurs hypothétiques sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par nombre total d'essais (axe des Y) et différence entre les valeurs hypothétiques (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Courbe du nombre total d'essais
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de l'estimation de puissance par nombre total d'essais. Cette valeur doit être supérieure à 0, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de l'estimation de puissance par nombre total d'essais. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Estimation de puissance vs valeur d'hypothèse nulle
Si ce paramètre est sélectionné, il active les options ci-après.

sur l'axe des X et valeur d'hypothèse alternative sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par valeur Null (axe des X) et valeur alternative (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

sur l'axe des Y et valeur d'hypothèse alternative sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par valeur Null (axe des Y) et valeur alternative (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Analyse de puissance du test binomial pour échantillons liés

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

La distribution binomiale est basée sur une séquence d'essais de Bernoulli. Elle permet de modéliser des expérimentations et notamment un nombre fixe d'essais totaux, supposés être indépendants les uns des autres. Chaque essai génère un résultat dichotomique, avec la même probabilité de réussite.

Le test binomial pour échantillons liés évalue la puissance du test de McNemar pour comparer deux paramètres de proportion en fonction des sujets des paires appariées échantillonnés à partir de deux populations binomiales liées.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Proportions** > **Test binomial pour échantillons**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).

3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.

4. Lorsque **Puissance** est sélectionné comme méthode **Estimation** de l'hypothèse de test, entrez la valeur **Nombre total de paires** appropriée.
5. Choisissez de spécifier des valeurs de test pour **Proportions** ou **Effectifs**.
   - Si **Proportions** est sélectionné, saisissez des valeurs dans les champs **Proportion 1** et **Proportion 2**. Les valeurs doivent être comprises entre 0 et 1.
   - Si **Effectifs** est sélectionné, saisissez des valeurs dans les champs **Comptage 1** et **Comptage 2**. Ces valeurs doivent être comprises entre 0 et la valeur spécifiée pour **Nombre total de paires**.

   Remarques sur Proportions :

   - L'option **Proportions** est la seule disponible si la valeur **Puissance** est spécifiée.
   - Lorsque **Les valeurs de test sont marginales** n'est pas sélectionné : 0 < **Proportion 1** + **Proportion 2** ≤ 1
   - Si **Les valeurs de test sont marginales** est sélectionné :
     - **Proportion 1** * **Proportion 2** > 0
     - **Proportion 1** < 1
     - **Proportion 2** < 1
     - Les valeurs de **Proportion 1** et **Proportion 2** ne peuvent pas être identiques.

   Remarques sur les effectifs :

   - Les paramètres **d'effectifs** sont disponibles uniquement lorsque **Puissance** est sélectionné comme paramètre **Estimation** de l'hypothèse de test.
   - Lorsque **Les valeurs de test sont marginales** n'est pas sélectionné : 0 < **Comptage 1** + **Comptage 2** ≤ **Nombre total de paires**
   - Si **Les valeurs de test sont marginales** est sélectionné :
     - **Comptage 1** * **Comptage 2** > 0
     - **Comptage 1** < **Nombre total de paires**
     - **Comptage 2** < **Nombre total de paires**
6. Vous pouvez éventuellement sélectionner **Les valeurs de test sont marginales** pour contrôler si les proportions ou les effectifs spécifiés sont marginaux. Si **Les valeurs de test sont marginales** est activé, vous devez spécifier une valeur **Corrélation entre paires appariées**. Il doit s'agir d'une valeur unique comprise entre -1 et 1.
7. Sélectionnez une méthode d'estimation de la puissance.

   Approximation normale
   Active l'approximation normale. Il s'agit du paramètre par défaut.

   Enumération binomiale
   Active la méthode d'énumération binomiale. Vous pouvez également utiliser la zone **Limite de temps** pour indiquer le nombre maximum de minutes autorisées pour estimer la taille d'échantillon. Lorsque la limite de temps est atteinte, l'analyse se termine et un message d'avertissement s'affiche. Si vous indiquez cette valeur, il doit s'agir d'un entier positif unique indiquant le nombre de minutes. Le paramètre par défaut est 5 minutes.

8. Indiquez si le test est unilatéral ou bilatéral.

   Analyse non directionnelle (bilatérale)
   Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

Analyse directionnelle (unilatérale)
Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.

9. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
10. Vous pouvez éventuellement cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test binomial pour échantillons liés», à la page 20 (sortie de graphique, paramètres de tracé bidimensionnel et paramètres de tracé tridimensionnel).

    Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné car l'hypothèse de test **Estimation** et **Énumération binomiale** n'est pas sélectionnée.
11. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode **Estimation** de l'hypothèse de test et que l'analyse **Non directionnel (bilatéral)** est sélectionnée comme **Direction de test**.

Analyse de puissance du test binomial pour échantillons liés : tracé

La boîte de dialogue **Tracé** fournit des options permettant de contrôler les tracés qui sont produits pour illustrer les graphiques d'estimation de puissance bidimensionnels et tridimensionnels. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel
Fournit des options permettant de contrôler les graphiques bidimensionnels de l'estimation de puissance. Ce paramètre est désactivé par défaut.

Estimation de puissance vs nombre total de paires
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par nombre total de paires. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe du nombre total de paires
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par nombre total de paires. Cette valeur doit être supérieure à 1, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par nombre total de paires. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 2 500.

Estimation de puissance vs différence de risque
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par différence de risque. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs rapport de risque
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par rapport de risque. Il est désactivé par défaut.

Courbe du rapport de risque
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par rapport de risque. Cette valeur ne doit pas être supérieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par rapport de risque. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 10.

Estimation de puissance vs rapport des cotes
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par rapport des cotes. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe du rapport des cotes
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par rapport des cotes. Cette valeur ne doit pas être supérieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par rapport des cotes. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 10.

Estimation de puissance vs corrélation entre paires appariées
Contrôle la puissance estimée Ω par rapport à la corrélation entre le graphique des paires appariées. Le graphique n'est créé que lorsque les proportions ou les nombres marginaux sont spécifiés (plutôt que les proportions discordantes).

Tracé tridimensionnel
Fournit des options permettant de contrôler les graphiques tridimensionnels de l'estimation de la puissance. Ce paramètre est désactivé par défaut.

Estimation de puissance vs proportions discordantes
Si ce paramètre facultatif est activé, il contrôle le graphique tridimensionnel de puissance par proportions discordantes. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs proportions marginales
Si ce paramètre facultatif est activé, il contrôle le graphique tridimensionnel de puissance par proportions marginales. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Remarque : Ce paramètre n'est disponible que si **Les valeurs de test sont marginales** est sélectionné.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Analyse de puissance du test binomial pour échantillons indépendants

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

La distribution binomiale est basée sur une séquence d'essais de Bernoulli. Elle permet de modéliser ces expérimentations et notamment un nombre fixe d'essais totaux, supposés être indépendants les uns des autres. Chaque essai génère un résultat dichotomique, avec la même probabilité de "réussite". Le test binomial d'échantillons indépendants compare deux paramètres de proportion indépendants.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Proportions** > **Test binomial pour échantillons indépendants**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.

   Vous pouvez également indiquer une valeur **Rapport de taille du groupe**. La valeur par défaut est 1.
4. Lorsque **Puissance** est sélectionné comme paramètre **Estimation** de l'hypothèse de test, entrez des valeurs pour spécifier **Nombre total d'essais pour le groupe 1 et le groupe 2**. Les valeurs doivent être des entiers supérieurs à 1.
5. Spécifiez les paramètres de proportion des deux groupes. Les deux valeurs doivent être comprises entre 0 et 1.

   Remarque : Les deux valeurs ne peuvent pas être identiques si une valeur**Puissance** est spécifiée.
6. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
7. Sélectionnez la méthode de test souhaitée :

   Test de khi-carré
   Estime la puissance à l'aide du test de khi-carré de Pearson. Il s'agit du paramètre par défaut.

   L'écart type est combiné
   Paramètre facultatif contrôlant si l'estimation de l'écart type est combiné ou non. Ce paramètre est activé par défaut.

   Appliquer la correction de continuité
   Paramètre facultatif contrôlant si la correction de continuité est utilisée ou non. Il est désactivé par défaut.

   Test T
   Estime la puissance à l'aide du test T de Student.

   L'écart type est combiné
   Paramètre facultatif contrôlant si l'estimation de l'écart type est combiné ou non. Ce paramètre est activé par défaut.

Test du rapport de vraisemblance
Estime la puissance à l'aide du test du rapport de vraisemblance.

Test exact de Fisher
Estime la puissance à l'aide du test exact de Fisher.

Remarques :

- Dans certains cas, l'exécution du test exact de Fisher peut être assez longue.
- Tous les tracés sont bloqués lorsque le test exact de Fisher est sélectionné.

8. Sélectionnez une méthode d'estimation de la puissance.

Approximation normale
Active l'approximation normale. Il s'agit du paramètre par défaut.

Enumération binomiale
Active la méthode d'énumération binomiale. Vous pouvez également utiliser la zone **Limite de temps** pour indiquer le nombre maximum de minutes autorisées pour estimer la taille d'échantillon. Lorsque la limite de temps est atteinte, l'analyse se termine et un message d'avertissement s'affiche. Si vous indiquez cette valeur, il doit s'agir d'un entier positif unique indiquant le nombre de minutes. Le paramètre par défaut est 5 minutes.

9. Indiquez si le test est unilatéral ou bilatéral.

Analyse non directionnelle (bilatérale)
Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

Analyse directionnelle (unilatérale)
Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.

10. Vous pouvez éventuellement cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test binomial pour échantillons indépendants : tracé», à la page 25 (sortie de graphique, paramètres de tracé bidimensionnel et paramètres de tracé tridimensionnel).

Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné car l'hypothèse de test **Estimation** et **Énumération binomiale** n'est pas sélectionnée.

11. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode **Estimation** de l'hypothèse de test et que l'analyse **Non directionnel (bilatéral)** est sélectionnée comme **Direction de test**.

Analyse de puissance du test binomial pour échantillons indépendants : tracé

La boîte de dialogue **Tracé** fournit des options permettant de contrôler les tracés qui sont produits pour illustrer les graphiques d'estimation de puissance bidimensionnels et tridimensionnels. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel
Fournit des options permettant de contrôler les graphiques bidimensionnels de l'estimation de puissance. Ce paramètre est désactivé par défaut.

Estimation de puissance vs rapport de taille de groupe/Estimation de puissance vs taille de groupe
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par rapport de taille de groupe. Si ce paramètre est sélectionné, il affiche le graphique. Les paramètres supplémentaires ne sont pas disponibles lorsque plusieurs valeurs de puissance ont été spécifiées (**Estimation de puissance vs taille de groupe**). Il est désactivé par défaut.

Courbe du rapport de taille de groupe
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par nombre total de paires. Cette valeur doit être supérieure à 0,01, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par nombre total de paires. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 100.

Estimation de puissance vs différence de risque
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par différence de risque. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs rapport de risque
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par rapport de risque. Il est désactivé par défaut.

Courbe du rapport de risque
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par rapport de risque. Cette valeur ne doit pas être supérieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par rapport de risque. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 10.

Estimation de puissance vs rapport des cotes
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par rapport des cotes. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe du rapport des cotes
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par rapport des cotes. Cette valeur ne doit pas être supérieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par rapport des cotes. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 10.

Tracé tridimensionnel
Fournit des options permettant de contrôler les graphiques tridimensionnels de l'estimation de la puissance. Ce paramètre est désactivé par défaut.

Estimation de puissance vs proportions
Si ce paramètre facultatif est sélectionné, il fournit les options de puissance par proportion suivantes :

Proportion du groupe 1 sur l'axe des X et proportion du groupe 2 sur l'axe des Y
Contrôle le graphique tridimensionnel de la puissance par proportion du groupe 1 (axe des X) et proportion du groupe 2 (axe des Y). Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Proportion du groupe 1 sur l'axe des Y et proportion du groupe 2 sur l'axe des X
Contrôle le graphique tridimensionnel de la puissance par proportion du groupe 2 (axe des X) et proportion du groupe 1 (axe des Y). Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs tailles de groupe
Si ce paramètre facultatif est sélectionné, il fournit les options de puissance par tailles de groupe suivantes :

Taille du groupe 1 sur l'axe des X et taille du groupe 2 sur l'axe des Y
Contrôle le graphique tridimensionnel de la puissance par nombre d'essais dans le groupe 1 (axe des X) et nombre d'essais dans le groupe 2 (axe des Y). Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Taille du groupe 1 sur l'axe des Y et taille du groupe 2 sur l'axe des X
Contrôle le graphique tridimensionnel de la puissance par nombre d'essais dans le groupe 2 (axe des X) et nombre d'essais dans le groupe 1 (axe des Y). Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe de taille du groupe 1
Si cette option est sélectionnée, les limites inférieure et supérieure de la courbe du groupe 1 sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par rapport des cotes. Cette valeur doit être supérieure ou égale à 2, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par rapport des cotes. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 2 500.

Courbe de taille du groupe 2
Si cette option est sélectionnée, les limites inférieure et supérieure de la courbe du groupe 2 sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par rapport des cotes. Cette valeur doit être supérieure ou égale à 2, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par rapport des cotes. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 2 500.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Corrélations

Les fonctions statistiques ci-après sont incluses dans IBM SPSS Statistiques Base Edition.

Analyse de puissance du test de corrélation de Pearson à un échantillon

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

Le coefficient de corrélation produit-moment de Pearson mesure la force de l'association linéaire entre deux variables d'échelle aléatoires censées suivre une distribution normale bivariée. Par convention, il s'agit d'une quantité sans dimension qui s'obtient en standardisant la covariance entre deux variables continues et qui est donc comprise entre -1 et 1.

Le test utilise la méthode asymptotique de Fisher pour estimer la puissance de la corrélation de Pearson à un échantillon.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Corrélations** > **Produit-Moment de Pearson**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.
4. Lorsque **Puissance** est sélectionné comme paramètre **Estimation** de l'hypothèse de test, entrez la valeur **Taille de l'échantillon par paires** appropriée. Il doit s'agir d'un entier unique supérieur à 3.
5. Saisissez une valeur qui spécifie la valeur d'hypothèse alternative du paramètre de corrélation dans la zone **Paramètre de corrélation de Pearson**. Il doit s'agir d'une valeur numérique unique comprise entre -1 et 1.

   Remarque : Lorsque **Alimentation** est spécifié, la valeur **Paramètre de corrélation de Pearson** ne peut pas être -1 ou 1 et ne peut pas être égale à **Valeur NULL**.
6. Le cas échéant, vous pouvez entrer une valeur qui spécifie la valeur d'hypothèse nulle du paramètre de corrélation à tester dans la zone **Valeur NULL**. Il doit s'agir d'une valeur numérique unique comprise entre -1 et 1. La valeur par défaut est 0.

   Remarque : Lorsque **Puissance** est spécifié, **Valeur NULL** ne peut pas être -1 ou 1.
7. Sélectionnez éventuellement **Utiliser la formule de correction de biais dans l'estimation de puissance** pour indiquer si l'ajustement de biais est impliqué ou ignoré. Ce paramètre est activé

par défaut et inclut la condition d'ajustement de biais dans l'estimation de puissance. Si ce paramètre n'est pas sélectionné, la condition d'ajustement de biais est ignorée.

8. Indiquez si le test est unilatéral ou bilatéral.

   Analyse non directionnelle (bilatérale)
   Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

   Analyse directionnelle (unilatérale)
   Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.

9. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
10. Vous pouvez éventuellement cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test de corrélation de Pearson à un échantillon : tracé», à la page 29 (sortie de graphique, paramètres de tracé bidimensionnel et paramètres de tracé tridimensionnel).

    Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné comme hypothèse de test.
11. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode **Estimation** de l'hypothèse de test et que l'analyse **Non directionnel (bilatéral)** est sélectionnée comme **Direction de test**.

Analyse de puissance du test de corrélation de Pearson à un échantillon : tracé

La boîte de dialogue **Tracé** fournit des options permettant de contrôler les tracés qui sont produits pour illustrer la puissance bidimensionnelle et tridimensionnelle par des graphiques. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel
Fournit des options permettant de contrôler les graphiques bidimensionnels de l'estimation de puissance. Ces paramètres sont désactivés par défaut.

Estimation de puissance vs valeur d'hypothèse nulle
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par valeur Null. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs valeur d'hypothèse alternative
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par valeur alternative. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs différence entre les valeurs hypothétiques
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par différence entre les valeurs hypothétiques. Il est désactivé par défaut.

Estimation de puissance vs taille d'échantillon (par paires)
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par taille d'échantillon. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe de taille d'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure ou égale à 4, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Tracé tridimensionnel
Fournit des options permettant de contrôler les graphiques tridimensionnels de l'estimation de la puissance. Ce paramètre est désactivé par défaut.

Estimation de puissance vs taille de l'échantillon
Si ce paramètre est sélectionné, il active les options ci-après.

sur l'axe des X et différence entre les valeurs hypothétiques sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des X) et différence entre les valeurs hypothétiques (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

sur l'axe des Y et différence entre les valeurs hypothétiques sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des Y) et différence entre les valeurs hypothétiques (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Courbe de taille d'échantillon (par paires)
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure ou égale à 4, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Estimation de puissance vs valeur d'hypothèse nulle
Si ce paramètre est sélectionné, il active les options ci-après.

sur l'axe des X et valeur d'hypothèse alternative sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par valeur Null (axe des X) et valeur alternative (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

sur l'axe des Y et valeur d'hypothèse alternative sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par valeur Null (axe des Y) et valeur alternative (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé

tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Analyse de puissance du test de corrélation de Spearman à un échantillon

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

Le coefficient de corrélation d'ordre des rangs de Spearman est une statistique d'ordre des rangs non paramétrique permettant de mesurer la relation monotone entre deux variables généralement censurées et non distribuées suivant une loi normale. La corrélation d'ordre des rangs de Spearman est égale à la corrélation de Pearson entre les valeurs de rang des deux variables, de ce fait aussi comprises entre -1 et 1. La détection de la puissance du test de corrélation de rang de Spearman est un sujet important dans l'analyse des données sur les séries chronologiques hydrologiques.

Le test utilise la méthode asymptotique de Fisher pour estimer la puissance de la corrélation d'ordre des rangs de Spearman à un échantillon.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Corrélations** > **Ordre des rangs de Spearman**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.
4. Lorsque **Puissance** est sélectionné comme paramètre **Estimation** de l'hypothèse de test, entrez la valeur **Taille de l'échantillon par paires** appropriée. Il doit s'agir d'un entier unique supérieur à 3.
5. Saisissez une valeur qui spécifie la valeur d'hypothèse alternative du paramètre de corrélation dans la zone **Paramètre de corrélation de Spearman**. Il doit s'agir d'une valeur numérique unique comprise entre -1 et 1.

   Remarque : Lorsque **Puissance** est spécifié, la valeur **Paramètre de corrélation de Spearman** ne peut pas être -1 ou 1 et ne peut pas être égale à la **Valeur NULL**.
6. Le cas échéant, vous pouvez entrer une valeur qui spécifie la valeur d'hypothèse nulle du paramètre de corrélation à tester dans la zone **Valeur NULL**. Il doit s'agir d'une valeur numérique unique comprise entre -1 et 1. La valeur par défaut est 0.

   Remarque : Lorsque **Puissance** est spécifié, **Valeur NULL** ne peut pas être -1 ou 1.
7. Sélectionnez éventuellement une option qui détermine la manière dont la variance asymptotique est évaluée pour l'analyse de puissance.

   Bonett et Wright
   Estime la variance recommandée par Bonett et Wright. Il s'agit du paramètre par défaut.

   Fieller, Hartley et Pearson
   Estime la variance recommandée par Fieller, Hartley et Pearson.

   Caruso et Cliff
   Estime la variance recommandée par Caruso et Cliff.

8. Indiquez si le test est unilatéral ou bilatéral.

   Analyse non directionnelle (bilatérale)
   Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

   Analyse directionnelle (unilatérale)
   Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.

9. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
10. Vous pouvez éventuellement cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test de corrélation de Spearman à un échantillon : tracé», à la page 32 (sortie de graphique, paramètres de tracé bidimensionnel et paramètres de tracé tridimensionnel).

    Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné comme hypothèse de test.
11. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode **Estimation** de l'hypothèse de test et que l'analyse **Non directionnel (bilatéral)** est sélectionnée comme **Direction de test**.

Analyse de puissance du test de corrélation de Spearman à un échantillon : tracé

La boîte de dialogue **Tracé** fournit des options permettant de contrôler les tracés qui sont produits pour illustrer la puissance bidimensionnelle et tridimensionnelle par des graphiques. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel
Fournit des options permettant de contrôler les graphiques bidimensionnels de l'estimation de puissance. Ces paramètres sont désactivés par défaut.

Estimation de puissance vs valeur d'hypothèse nulle
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par valeur Null. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs valeur d'hypothèse alternative
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par valeur alternative. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs différence entre les valeurs hypothétiques
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par différence entre les valeurs hypothétiques. Il est désactivé par défaut.

Estimation de puissance vs taille d'échantillon (par paires)
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par taille d'échantillon. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe de taille d'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure ou égale à 4, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Tracé tridimensionnel
Fournit des options permettant de contrôler les graphiques tridimensionnels de l'estimation de la puissance. Ce paramètre est désactivé par défaut.

Estimation de puissance vs taille de l'échantillon
Si ce paramètre est sélectionné, il active les options ci-après.

sur l'axe des X et différence entre les valeurs hypothétiques sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des X) et différence entre les valeurs hypothétiques (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

sur l'axe des Y et différence entre les valeurs hypothétiques sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des Y) et différence entre les valeurs hypothétiques (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Courbe de taille d'échantillon (par paires)
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure ou égale à 4, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Estimation de puissance vs valeur d'hypothèse nulle
Si ce paramètre est sélectionné, il active les options ci-après.

sur l'axe des X et valeur d'hypothèse alternative sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par valeur Null (axe des X) et valeur alternative (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

sur l'axe des Y et valeur d'hypothèse alternative sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par valeur Null (axe des Y) et valeur alternative (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Analyse de puissance du test de corrélation partielle de Pearson

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons, sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

La corrélation partielle peut être définie comme l'association entre deux variables aléatoires une fois que l'effet d'une ou plusieurs autres variables a été éliminé. Cette mesure s'avère utile en cas de confusion. Semblable au Coefficient de corrélation de Pearson, le coefficient de corrélation partielle est également une quantité sans dimension comprise entre -1 et 1.

Le test utilise la méthode asymptotique de Fisher pour estimer la puissance de la corrélation de Pearson à un échantillon.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Corrélations** > **Partiel**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.
4. Lorsque **Puissance** est sélectionné comme paramètre **Estimation** de l'hypothèse de test, entrez la valeur **Taille de l'échantillon** appropriée. La valeur doit être un entier unique supérieur à 1.
5. Entrez une valeur qui spécifie **Nombre de variables supposées être biaisées**. Il doit s'agir d'un entier unique supérieur ou égal à 0.
6. Entrez une valeur qui spécifie la valeur d'hypothèse alternative de **Paramètre de corrélation partielle**. Il doit s'agir d'une valeur numérique unique comprise entre -1 et 1.

   Remarque : Lorsque **Puissance** est spécifié, la valeur **Paramètre de corrélation partielle** ne peut pas être -1 ou 1 et ne peut pas être égale à **Valeur NULL**.
7. Le cas échéant, vous pouvez entrer une valeur qui spécifie la valeur d'hypothèse nulle du paramètre de corrélation partielle à tester dans la zone **Valeur NULL**. Il doit s'agir d'une valeur numérique unique comprise entre -1 et 1. La valeur par défaut est 0.

   Remarque : Lorsque **Puissance** est spécifié, **Valeur NULL** ne peut pas être -1 ou 1.
8. Indiquez si le test est unilatéral ou bilatéral.

   Analyse non directionnelle (bilatérale)
   Lorsque cette option est sélectionnée, un test bilatéral est utilisé. Il s'agit du paramètre par défaut.

   Analyse directionnelle (unilatérale)
   Lorsque cette option est sélectionnée, la puissance est calculée pour un test unilatéral.
9. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.

10. Vous pouvez éventuellement cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test de corrélation partielle de Pearson : tracé», à la page 35 (sortie de graphique, paramètres de tracé bidimensionnel et paramètres de tracé tridimensionnel).

    Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné comme hypothèse de test.

11. Vous pouvez également cliquer sur **Précision** pour estimer la taille de l'échantillon en fonction des intervalles de confiance en spécifiant les valeurs des demi-largeurs de l'intervalle de confiance. Pour plus d'informations, voir «Analyse de l'alimentation: Précision», à la page 39.

Remarque : Précision est disponible uniquement lorsque **Taille de l'échantillon** est sélectionné comme méthode **Estimation** de l'hypothèse de test et que l'analyse **Non directionnel (bilatéral)** est sélectionnée comme **Direction de test**.

Analyse de puissance du test de corrélation partielle de Pearson : tracé

La boîte de dialogue **Tracé** fournit des options permettant de contrôler les tracés qui sont produits pour illustrer la puissance bidimensionnelle et tridimensionnelle par des graphiques. La boîte de dialogue contrôle également les degrés de rotation verticale et horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel
Fournit des options permettant de contrôler les graphiques bidimensionnels de l'estimation de puissance. Ces paramètres sont désactivés par défaut.

Estimation de puissance vs valeur d'hypothèse nulle
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par valeur Null. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs valeur d'hypothèse alternative
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par valeur alternative. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs nombre de variables biaisées
Lorsque cette option est activée, ce paramètre facultatif contrôle le graphique de puissance bidimensionnelle par nombre de variables biaisées. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs différence entre les valeurs hypothétiques
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par différence entre les valeurs hypothétiques. Il est désactivé par défaut.

Estimation de puissance vs taille de l'échantillon
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par taille d'échantillon. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe de taille d'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure ou égale à 4, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Tracé tridimensionnel
Fournit des options permettant de contrôler les graphiques tridimensionnels de l'estimation de la puissance. Ce paramètre est désactivé par défaut.

Estimation de puissance vs taille de l'échantillon
Si ce paramètre est sélectionné, il active les options ci-après.

sur l'axe des X et différence entre les valeurs hypothétiques sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des X) et différence entre les valeurs hypothétiques (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

sur l'axe des Y et différence entre les valeurs hypothétiques sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des Y) et différence entre les valeurs hypothétiques (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Courbe de taille d'échantillon (par paires)
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique tridimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure ou égale à 4, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique tridimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Estimation de puissance vs valeur d'hypothèse nulle
Si ce paramètre est sélectionné, il active les options ci-après.

sur l'axe des X et valeur d'hypothèse alternative sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par valeur Null (axe des X) et valeur alternative (axe des Y). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

sur l'axe des Y et valeur d'hypothèse alternative sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par valeur Null (axe des Y) et valeur alternative (axe des X). Par défaut, le graphique est supprimé. Vous pouvez le faire apparaître si besoin.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Régression

Les fonctions statistiques ci-après sont incluses dans IBM SPSS Statistiques Base Edition.

Analyse de puissance du test de régression linéaire univariée

Cette fonction requiert IBM SPSS Statistiques Base Edition.

L'analyse de puissance joue un rôle essentiel dans la planification, la conception et le déroulement d'une étude. Le calcul de la puissance est généralement réalisé avant la collecte de données d'échantillons,

sauf dans le cas d'un projet pilote par exemple. L'estimation précise de la puissance peut nous dire qu'elles sont les chances de détecter une différence statistiquement significative en fonction de la taille de l'échantillon finie dans le cadre d'une hypothèse alternative vraie. Si la puissance est trop faible, on a peu de chance de détecter une différence significative et les résultats ne seront probablement pas significatifs si de réelles différences existent.

La régression linéaire univariée est une approche statistique de base standard où les chercheurs utilisent les valeurs de plusieurs variables pour expliquer ou prédire les valeurs d'une échelle de résultats.

L'analyse de puissance du test de régression linéaire univariée estime la puissance du test *F* de type III dans les modèles de régression linéaire multiple univariée. Avec la taille d'effet représentée par plusieurs corrélations (partielles), les approches des prédicteur fixes et aléatoires sont fournies. Pour les prédicteurs fixes, l'estimation de la puissance repose sur la distribution *F* non centrale. Pour les prédicteurs aléatoires, il est supposé que la variable cible et les prédicteurs suivent conjointement une distribution normale multivariée. Dans ce cas, l'estimation de la puissance repose sur la distribution de l'échantillon de coefficient de corrélation multiple.

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de puissance** > **Régression** > **Linéaire univariée**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un paramètre **Estimation** d'hypothèse de test (**Taille de l'échantillon** ou **Puissance**).
3. Lorsque **Taille de l'échantillon** est sélectionné, entrez un **Valeur de puissance unique** pour la valeur d'estimation de la taille de l'échantillon (la valeur doit être une valeur unique comprise entre 0 et 1), ou sélectionnez **Valeurs de puissance de grille**, puis cliquez sur **Grille** pour afficher les tailles d'échantillon projetées pour une plage de valeurs de puissance spécifiques.

   Pour plus d'informations, voir «Analyse de puissance : Valeurs de grille», à la page 40.
4. Si vous sélectionnez **Estimer la puissance**, saisissez une **Taille d'échantillon** appropriée pour la valeur d'estimation de la puissance. Il doit s'agir d'un entier unique supérieur ou égal nombre total de prédicteurs de modèle +2 (si le paramètre **Inclure le terme de la constante dans le modèle** est activé). Sinon, la valeur doit correspondre à un entier unique supérieur ou égal nombre total de prédicteurs de modèle +1.
5. Spécifiez la valeur du coefficient de corrélation partielle dans la zone **Corrélation partielle multiple de la population**. Il doit s'agir d'une valeur unique comprise entre -1 et 1.

   Remarque : Si une valeur **Puissance** est spécifiée, la valeur **Corrélation partielle multiple de la population** ne doit pas être nulle.

   Les paramètres suivants sont activés si la valeur **Corrélation partielle multiple de la population** est sélectionnée :

   Nombre total de prédicteurs dans le modèle
   Spécifiez le nombre total de prédicteurs ou le nombre de prédicteurs dans le modèle complet (sans la constante, si applicable). Il doit s'agir d'un entier unique supérieur ou égal à 1.

   Nombre de prédicteurs de test
   Spécifiez le nombre de prédicteurs de test ou le nombre de prédicteurs dans le modèle imbriqué (sans la constante, si applicable). Cette valeur doit être supérieure ou égale à 1, mais inférieure à la valeur **Nombre total de prédicteurs dans le modèle**.
6. Spécifiez **Valeurs R carré pour** plusieurs coefficients de corrélation pour **Modèle complet** et **Modèle imbriqué**. Il doit s'agir de valeurs uniques comprises entre 0 et 1.

   Remarque : Si une valeur **Puissance** est spécifiée, la valeur **Modèle complet** doit être supérieure à la valeur **Modèle imbriqué**.

   Les paramètres suivants sont activés si la valeur **Valeurs R carré pour** est sélectionnée :

   Nombre total de prédicteurs - Modèle complet
   Indiquez le nombre total de prédicteurs pour le modèle complet (sans la constante, le cas échéant). Il doit s'agir d'un entier unique supérieur ou égal à 1.

Nombre total de prédicteurs - Modèle imbriqué
Indiquez le nombre total de prédicteurs pour le modèle imbriqué (sans la constante, le cas échéant). La valeur doit être supérieure ou égale à 1, mais inférieure à la valeur **Nombre total de prédicteurs - Modèle complet**.

7. Spécifiez éventuellement le niveau d'importance du taux d'erreur Type I du test dans la zone **Niveau d'importance**. La valeur doit être une double unique comprise entre 0 et 1. La valeur par défaut est 0,05.
8. Vous pouvez éventuellement sélectionner le paramètre **Inclure le terme de la constante dans le modèle**. Ce paramètre est activé par défaut. Si cette option n'est pas sélectionnée, le terme de la constante est exclu de l'analyse de puissance.
9. Vous pouvez éventuellement indiquer si les prédicteurs du modèle sont de type **Fixe** ou **Aléatoire**. **Fixe** est le paramètre par défaut.
10. Vous pouvez éventuellement cliquer sur **Tracé** pour spécifier les paramètres «Analyse de puissance du test de régression linéaire univariée : tracé», à la page 38 (sortie de graphique, paramètres de tracé bidimensionnel et paramètres de tracé tridimensionnel).

    Remarque : Tracé est disponible uniquement lorsque **Puissance** est sélectionné comme hypothèse de test.

Analyse de puissance du test de régression linéaire univariée : tracé

Vous pouvez contrôler les tracés générés en sortie pour illustrer les graphiques bidimensionnels et tridimensionnels de puissance. Vous pouvez également contrôler l'affichage des infobulles et des degrés de rotation verticale/horizontale pour les graphiques tridimensionnels.

Tracé bidimensionnel
Fournit des options permettant de contrôler les graphiques bidimensionnels de l'estimation de puissance. Ces paramètres sont désactivés par défaut.

Estimation de puissance vs corrélation partielle multiple
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par coefficient de corrélation partielle multiple. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Estimation de puissance vs taille de l'échantillon
Si ce paramètre facultatif est activé, il contrôle le graphique bidimensionnel de puissance par taille d'échantillon. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe de taille d'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure ou égale à 4, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Tracé tridimensionnel
Propose des options permettant de contrôler le graphique tridimensionnel d'estimation de puissance, les paramètres de rotation verticale/horizontale, ainsi que la courbe spécifiée par l'utilisateur pour la taille d'échantillon. Ce paramètre est désactivé par défaut.

Estimation de puissance vs taille de l'échantillon
Si ce paramètre facultatif est activé, il contrôle les graphiques bidimensionnels de puissance par taille d'échantillon. Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

sur l'axe des X et corrélation partielle multiple sur l'axe des Y
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des X) et corrélation partielle multiple (axe des Y). Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

sur l'axe des Y et corrélation partielle multiple sur l'axe des X
Paramètre facultatif contrôlant le graphique tridimensionnel de puissance par taille d'échantillon (axe des Y) et corrélation partielle multiple (axe des X). Il est désactivé par défaut. Si ce paramètre est sélectionné, il affiche le graphique.

Courbe de taille d'échantillon
Lorsque cette option est sélectionnée, les limites inférieure et supérieure sont disponibles. Si aucune valeur entière n'est spécifiée pour les zones **Limite inférieure** ou **Limite supérieure**, la courbe par défaut est utilisée.

Limite inférieure
Contrôle la limite inférieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure ou égale à 4, mais inférieure à la valeur de la zone **Limite supérieure**.

Limite supérieure
Contrôle la limite supérieure du graphique bidimensionnel de l'estimation de puissance par taille d'échantillon. Cette valeur doit être supérieure à la valeur de la zone **Limite inférieure**, mais inférieure à 5 000.

Rotation verticale
Paramètre facultatif définissant les degrés de rotation verticale (dans le sens des aiguilles d'une montre, à partir de la gauche) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique verticalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 10.

Rotation horizontale
Paramètre facultatif définissant les degrés de rotation horizontale (dans le sens des aiguilles d'une montre, de face) pour le graphique tridimensionnel. Vous pouvez utiliser la souris pour faire pivoter le graphique horizontalement. Ce paramètre est appliqué lorsque le tracé tridimensionnel est nécessaire. La valeur doit être une valeur entière unique inférieure ou égale à 359. La valeur par défaut est 325.

Analyse de l'alimentation: Précision

La boîte de dialogue **Précision** est disponible pour chaque procédure d'analyse de puissance (à l'exception de la régression linéaire unidimensionnelle) et fournit des options permettant d'estimer la taille de l'échantillon en fonction des demi-largeurs de l'intervalle de confiance spécifié. La boîte de dialogue est disponible lorsque **Valeur d'alimentation unique** ou **Valeurs d'alimentation de la grille** sont spécifiés pour l'estimation **Taille de l'échantillon** et lorsque le paramètre **Direction de test** est défini sur **Non directionnel (analyse à deux côtés)**.

Remarque : Le paramètre **Direction de test** ne s'applique pas à One-Way ANOVA.

Lorsque les options **Estimation de la taille des échantillons** et **Valeurs d'alimentation de la grille** sont sélectionnées (cliquez sur le contrôle **Grille** pour afficher la boîte de dialogue).

Type(s) d'intervalle de confiance
Sélectionnez les types d'intervalles de confiance appropriés et si la correction de continuité est appliquée.

Remarque : Type (s) d'intervalle de confiance ne sont disponibles que pour les procédures de proportion d'analyse de puissance (proportion One-Sample, Related-Samples Binomial Test et Independent-Samples Binomial Test).

Indiquez la demi-largeur de l'intervalle de confiance
Estime la taille de l'échantillon en fonction de la valeur de demi-largeur de l'intervalle de confiance. Entrez une valeur comprise entre 0 et 1. Pour le test binomial One-Sample, la valeur doit être comprise entre 0 et 0,5.

Remarque : Les valeurs en double sont ignorées.

- Cliquez sur **Ajouter** pour ajouter la valeur de demi-largeur spécifiée à la liste.
- Mettez en évidence une valeur de demi largeur existante et cliquez sur **Changement** pour mettre à jour la valeur.
- Mettez en évidence une valeur de demi-largeur existante et cliquez sur **Supprimer** pour supprimer la valeur de la liste.

Analyse de puissance : Valeurs de grille

La boîte de dialogue **Valeurs de grille** propose des options permettant de spécifier une plage de valeurs **POWER** afin de visualiser les tailles d'échantillon prévisionnelles dans un format de grille pour chaque valeur de plage **POWER** spécifiée.

La boîte de dialogue **Valeurs de grille** est disponible pour chaque procédure d'analyse de puissance lorsque les options **Estimer la taille d'échantillon** et **Valeurs d'exposant de grille** sont sélectionnées (cliquez sur la commande **Grille** pour afficher la boîte de dialogue).

Spécifier un exposant unique
Si cette option est sélectionnée, vous devez entrer au moins une valeur pour exécuter l'analyse. Les valeurs multiples sont admises et chaque valeur doit être comprise dans l'intervalle `[0, 1]`. Vous pouvez spécifier plusieurs valeurs en les séparant par un ou plusieurs espaces. Utilisez les commandes **Ajouter**, **Modifier** et **Supprimer** pour utiliser les valeurs de la liste des valeurs de puissance.

Toutes les valeurs doivent être uniques (les valeurs en double sont interdites).

Spécifier une variation de l'exposant
Lorsque cette option est sélectionnée, il est possible de spécifier une gamme de valeurs de puissance allant d'une valeur de **départ** (`value1`) à une valeur de **fin** (`value2`) avec l'incrément de **par** (`value3`). Si elle est spécifiée, un seul ensemble valide de `[value1 TO value2 BY value3]` est autorisé. Il doit satisfaire la condition $0 \leq value1 \leq value2 \leq 1$. Si $value1 = value2$, cela revient à spécifier une valeur `value1` unique, quelle que soit la valeur `value3`.

Remarque : Les options **Spécifier un exposant unique** et **Spécifier une variation de l'exposant** sont indépendantes ; vous pouvez sélectionner une seule option ou les deux.

Méta-analyse

La méta-analyse est l'analyse des données obtenues à partir d'un ensemble d'études qui répondent à des questions d'études similaires. Ces études sont connues sous le nom d'études primaires. La méta-analyse utilise des méthodes statistiques pour produire une estimation globale d'un effet, explorer l'hétérogénéité entre les études et étudier l'impact du biais de publication ou, plus généralement, les effets de petite étude sur les résultats définitifs.

IBM SPSS Statistiques prend en charge les tailles d'effet standard et les tailles d'effet génériques (pré-calculées) pour les données binaires (telles que le log du rapport de cotes) et pour les données continues (telles que le g de Hedges). Les informations de méta-analyse (telles que les tailles d'effets spécifiques à l'étude et les erreurs standard correspondantes ainsi que le modèle et la méthode de la méta-analyse) sont spécifiées lors de l'étape de déclaration de la méta-analyse. Les informations sont automatiquement utilisées par toutes les méta-analyses ultérieures.

Les modèles d'analyse à effets aléatoires, à effet commun et à effets fixes sont pris en charge. Selon le modèle de méta-analyse choisi, les différentes méthodes d'estimation (par exemple, variance inverse et Mantel-Haenszel) sont disponibles pour les modèles à effet commun et à effets fixes. Plusieurs estimateurs différents sont également disponibles pour le paramètre de variance entre les études pour le modèle à effets aléatoires.

Méta-analyse continue

La procédure Méta-analyse continue effectue une méta-analyse avec des résultats continus sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Exemple

Plusieurs travaux de recherche ont été réalisés tout au long de l'histoire pour étudier un médicament à la mode mais discutable pour traiter le diabète de type II. Le médicament oral était censé réduire le taux de glucose dans le sang après les repas. Des données ont été recueillies sur différents sites de recherche de 1979 à 1986.

Une chercheuse principale souhaite établir des statistiques sur l'effet du médicament oral. Etant donné que les données ont été générées à partir d'études différentes, elle a proposé l'idée de synthétiser les résultats dans toutes les études afin de comprendre globalement l'effet et d'identifier les sources sous-jacentes de variation dans les résultats.

Statistiques

Niveau de confiance, méthode itérative, dichotomie, tolérance de convergence, moyennes d'échantillons, variance d'échantillon, écart type, taille d'effet estimée, *d* de Cohen, *g* de Hedges, delta de Glass, différence moyenne, analyse cumulée, méthode d'estimation, troncature et remplissage, test de régression, modèle à effets aléatoires, modèle à effets fixes, estimateur de vraisemblance maximale restreinte, estimateur Bayes empirique, estimateur Hedges, estimateur Hunter-Schmidt, estimateur DerSimonian-Laird, estimateur Sidik-Jonkman, ajustement d'erreur standard Knapp-Hartung, ajustement d'erreur standard Knapp-Hartung tronqué, coefficients, test de régression d'EGGER, constante, modèle multiplicatif, paramètre de dispersion multiplicative, estimateur quadratique, test d'homogénéité, mesures d'hétérogénéité, intervalle de prévision, erreur standard estimée, valeur *p* estimée, taille d'effet globale cumulée, pondération d'étude estimée.

Obtention d'une analyse continue de méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Données brutes...**

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

2. Dans la section **Groupe qui suit le traitement**, sélectionnez une variable **Taille de l'étude** pour représenter la taille d'échantillon pour le groupe qui suit le traitement. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
3. Sélectionnez une variable **Moyenne** pour représenter les moyennes d'échantillons pour le groupe qui suit le traitement. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
4. Sélectionnez **Ecart type** pour spécifier l'écart type de l'échantillon, ou **Variance** pour spécifier la variance de l'échantillon, puis sélectionnez une variable pour représenter l'écart type/variance pour le groupe qui suit le traitement.
5. Dans la section **Groupe de contrôle**, sélectionnez une variable **Taille de l'étude** pour représenter la taille d'échantillon pour le groupe de contrôle. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
6. Sélectionnez une variable **Moyenne** pour représenter les moyennes d'échantillons pour le groupe de contrôle. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).

7. Sélectionnez **Ecart type** pour spécifier l'écart type de l'échantillon, ou **Variance** pour spécifier la variance de l'échantillon, puis sélectionnez une variable pour représenter l'écart type/variance pour le groupe de contrôle.
8. (Facultatif) Sélectionnez les variables **ID étude** et/ou **Libellé de l'étude**. La variable **ID étude** sélectionnée ne peut pas être identique à la variable **Libellé de l'étude** sélectionnée.
9. (Facultatif) Sélectionnez un paramètre **Taille d'effet**.

   d de Cohen
   Le paramètre par défaut estime le *d* de Cohen. Lorsque l' **erreur standard ajustée** est sélectionnée, le paramètre estime le *d de* Cohen et sa variance à l'aide d'une formule alternative divisée par $2(N_{treatment} + N_{control} - 2)$.

   g de Hedges
   Estime le *g* de Hedges. Lorsque l' **erreur standard ajustée** est sélectionnée, le paramètre estime le *g de* Hedges et sa variance à l'aide d'une formule alternative divisée par $2(N_{treatment} + N_{control} - 3.94)$.

   Delta de Glass
   Estime le delta de Glass en fonction du groupe de contrôle. Lorsque l'option **Normalisé selon le groupe qui suit le traitement** est sélectionnée, le delta de Glass est normalisé en fonction de l'écart type du groupe qui suit le traitement.

   Différence moyenne non standardisée
   Estime la différence moyenne en supposant que les deux écarts types de la population sont égaux. Lorsque l'option **Variances de groupe inégales** est sélectionnée, la différence moyenne est estimée en supposant que les deux écarts types de la population ne sont pas égaux.
10. (Facultatif) Sélectionnez un paramètre **Modèle**. Lorsque les paramètres **Tronquer et remplir** sont activés, ce paramètre contrôle également le modèle utilisé par le regroupement dans l'analyse de troncature et de remplissage. Lorsque les paramètres **Biais** sont activés, ce paramètre contrôle également le modèle utilisé par le regroupement dans le test de régression.

    Effets aléatoires
    La valeur par défaut permet de générer le modèle à effets aléatoires.

    Effets fixes
    Génère le modèle à effets fixes.
11. (En option) Vous avez également ces possibilités :
    - Cliquez sur **Critères...** pour spécifier les critères généraux.
    - Cliquez sur **Analyse** pour spécifier le sous-groupe et l'analyse cumulée.
    - Cliquer sur **Inférence** pour spécifier les méthodes d'estimation.
    - Cliquez sur **Contraste** pour contrôler le test de contraste.
    - Cliquez sur **Biais** pour accéder au biais de publication en effectuant le test de régression d'EGGER.
    - Cliquez sur **Tronquer et remplir** pour mettre en œuvre l'analyse de troncature et de remplissage du biais de publication.
    - Cliquez sur **Imprimer** pour contrôler les sorties de tableau.
    - Cliquez sur **Enregistrer** pour enregistrer les statistiques estimées dans le jeu de données actif.
    - Cliquer sur **Tracé** pour spécifier les tracés à inclure dans la sortie.
12. Cliquez sur **OK**.

Méta-analyse continue : Critères

La boîte de dialogue **Critères** propose des paramètres permettant de spécifier des critères pour une méta-analyse avec des résultats continus sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Intervalle de confiance
Ce paramètre facultatif spécifie le niveau de confiance. Il doit s'agir d'une valeur numérique comprise entre 0 et 100. La valeur par défaut est 95.

La portée des données est manquante
Les paramètres facultatifs contrôlent la façon dont la procédure traite les données manquantes.

Exclure les observations analyse par analyse
Le paramètre par défaut inclut toutes les observations ayant suffisamment de données sur les variables utilisées dans chaque analyse particulière.

Exclure toute observation incomplète
Inclut toutes les observations ayant suffisamment de données sur toutes les variables utilisées dans toutes les analyses spécifiées par la procédure.

Valeurs utilisateur manquantes
Les paramètres facultatifs contrôlent la façon dont les valeurs utilisateur manquantes sont traitées.

Exclure
Le paramètre par défaut considère les valeurs utilisateur manquantes comme étant valides.

Inclure
Ignore les désignations de valeurs utilisateur manquantes et considère ces dernières comme étant valides.

Itérations

Nombre maximum d'itérations
Ce paramètre facultatif spécifie le nombre maximal d'itérations dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 100. La valeur 0 indique qu'aucune itération n'est effectuée.

Découpage maximal d'étape en deux
Ce paramètre facultatif spécifie le nombre maximal de dichotomies dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 5. Une valeur de 0 signifie que le step-halving n'est pas appliqué.

Convergence
Ce paramètre facultatif spécifie la tolérance de convergence. La valeur doit correspondre à une valeur positive unique. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur par défaut est 1E-6.

Définition des critères de la méta-analyse continue

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse continue**, cliquez sur **Critères**.
3. Sélectionnez et définissez les paramètres de critères appropriés.
4. Cliquez sur **Continuer**

Méta-analyse continue : Analyse

La boîte de dialogue **Analyse** propose des paramètres permettant de spécifier des analyses cumulées et de sous-groupe pour une méta-analyse avec des résultats continus sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Analyse de sous-groupe
Sélectionnez une variable évoquant l'analyse de sous-groupe. La variable ne peut pas être identique à la variable définie pour **Analyse cumulée**.

Analyse cumulée
Sélectionnez une variable évoquant l'analyse cumulée et en fonction de laquelle la méta-analyse cumulée est effectuée. La variable ne peut pas être identique à la variable définie pour **Analyse de sous-groupe**. Lorsque l'option **Croissant** est sélectionnée, l'analyse cumulée est basée sur la variable spécifiée dans l'ordre croissant. Lorsque l'option **Décroissant** est sélectionnée, l'analyse cumulée est basée sur la variable spécifiée dans l'ordre décroissant.

Statistiques cumulées
Propose des options permettant d'enregistrer la taille d'effet globale cumulée estimée. Les paramètres ne sont disponibles que lorsqu'une variable **Analyse cumulée** est sélectionnée.

Taille d'effet cumulée
Enregistre la taille d'effet globale cumulée estimée.

Erreur standard
Enregistre l'erreur standard estimée de la taille d'effet globale cumulée.

Limite inférieure de l'intervalle de confiance
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée.

Limite supérieure de l'intervalle de confiance
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée.

Valeur P
Enregistre la valeur *p* estimée de la taille d'effet globale cumulée.

Destination
Fournit des options permettant de spécifier un fichier ou jeu de données de sauvegarde. Lorsque **Jeu de données** est sélectionné, vous pouvez spécifier un nouveau nom de jeu de données (vous pouvez conserver le nom de jeu de données par défaut). Lorsque le **fichier de données** est sélectionné, cliquez sur **Parcourir...** pour sélectionner le nom et l'emplacement du fichier d'enregistrement.

Définition des paramètres d'analyse de la méta-analyse continue

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse continue**, cliquez sur **Analyse**.
3. Sélectionnez et définissez les paramètres d'analyse appropriés.
4. Cliquez sur **Continuer**

Méta-analyse continue : Inférence

La boîte de dialogue **Inférence** propose des paramètres permettant de spécifier les méthodes d'estimation pour une méta-analyse avec des résultats continus sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Remarque : La boîte de dialogue **Inférence** n'est disponible que lorsqu'un modèle à **effets aléatoires** est sélectionné.

Estimateur
Fournit des paramètres permettant de spécifier l'estimateur.

Remarques :

- Lorsque les paramètres Tronquer et remplir sont spécifiés, ce paramètre contrôle également l'estimateur qui est utilisé par le regroupement dans l'analyse de troncature et de remplissage.
- Lorsque les paramètres de biais sont spécifiés, ce paramètre contrôle également l'estimateur qui est utilisé par le test de régression.

Maximum de vraisemblance restreint (REML)
Le paramètre par défaut applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance restreint.

Maximum de vraisemblance (ML)
Applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance.

Bayes empirique
Applique la méthode itérative et calcule l'estimateur Bayes empirique.

Hedges
Applique la méthode non itérative et calcule l'estimateur Hedges.

Hunter-Schmidt
Applique la méthode non itérative et calcule l'estimateur Hunter-Schmidt.

DerSimonian-Laird
Applique la méthode non itérative et calcule l'estimateur DerSimonian-Laird.

Sidik-Jonkman
Applique la méthode non itérative et calcule l'estimateur Sidik-Jonkman.

Ajustement de l'erreur standard
Fournit des paramètres qui déterminent si l'ajustement d'erreur standard Knapp-Hartung doit être appliqué ou non.

Remarques :

- Lorsque les paramètres Tronquer et remplir sont spécifiés, ce paramètre contrôle également l'ajustement d'erreur standard qui est utilisé par le regroupement dans l'analyse de troncature et de remplissage.
- Lorsque les paramètres de biais sont spécifiés, ce paramètre contrôle également l'ajustement d'erreur standard qui est utilisé par le test de régression.

Aucun ajustement
Le paramètre par défaut n'applique pas l'ajustement.

Appliquer l'ajustement de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung.

Appliquer l'ajustement tronqué de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung et tronque la valeur si elle est inférieure à 1 lors de l'estimation de la matrice variance-covariance.

Définition des paramètres d'inférence de la méta-analyse continue

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse continue**, cliquez sur **Inférence**.
3. Sélectionnez et définissez les paramètres d'inférence appropriés.
4. Cliquez sur **Continuer**

Méta-analyse continue : Contraste

La boîte de dialogue **Contraste** propose des paramètres permettant de contrôler le test de contraste pour une méta-analyse avec des résultats continus sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Variables
La liste affiche toutes les variables de jeu de données disponibles. Sélectionnez des variables dans la liste et déplacez-les vers la liste **Contraste(s)**.

Contraste(s)
La liste identifie les coefficients qui sont stockés sous forme de variables dans le jeu de données actif. Plusieurs variables sont autorisées. Les variables de chaîne ne sont pas prises en charge.

Valeurs de coefficient entrées par l'utilisateur
Fournit des paramètres permettant de spécifier les coefficients d'un contraste fourni par l'utilisateur. Seules les valeurs numériques sont admises. Pour formuler un test de contraste valide, le nombre de valeurs spécifiées doit correspondre à celui des études valides.

Définition des paramètres de contraste continus pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse continue**, cliquez sur **Contraste**.
3. Sélectionnez et définissez les paramètres de contraste appropriés.
4. Cliquez sur **Continuer**

Méta-analyse continue : Biais

La boîte de dialogue **Biais** propose des paramètres permettant d'activer le biais de publication en effectuant le test de régression d'Egger pour une méta-analyse avec des résultats continus sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Test de régression d'Egger
Sélectionnez ce paramètre pour activer le biais de publication en effectuant le test de régression d'Egger.

Variables
La liste fournit toutes les variables de jeu de données disponibles.

Covariable(s)
Les variables sélectionnées dans la liste **Variables** sont traitées sous forme de covariables. Plusieurs covariables sont autorisées.

Facteur(s)
Les variables sélectionnées dans la liste **Variables** sont traitées sous forme de facteurs. Plusieurs facteurs sont autorisés.

Inclure une constante dans la régression
Contrôle le terme de la constante dans le test de régression.

Inclure un paramètre de dispersion dans le modèle à effets fixes
Contrôle le paramètre de modèle multiplicatif et introduit le paramètre de dispersion multiplicative dans l'analyse. Ce paramètre n'est disponible que lorsqu'un modèle à effets fixes est sélectionné.

Estimer les statistiques sur la base de la distribution t
Contrôle la distribution utilisée dans les tests de régression. Ce paramètre est activé par défaut, et permet d'estimer les statistiques en fonction de la distribution *t*. Lorsqu'il n'est pas sélectionné, les statistiques sont estimées en fonction de la distribution normale.

Définition des paramètres de biais continus pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse continue**, cliquez sur **Biais**.
3. Sélectionnez et définissez les paramètres de biais appropriés.
4. Cliquez sur **Continuer**

Méta-analyse continue : Tronquer et remplir

La boîte de dialogue **Tronquer et remplir** propose des paramètres permettant de mettre en oeuvre l'analyse de troncature et de remplissage du biais de publication pour une méta-analyse avec des résultats continus sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Nombre d'études manquantes estimé
Contrôle l'analyse de troncature et de remplissage du biais de publication. En sélectionnant ce paramètre, vous activez les autres paramètres de boîte de dialogue.

Côté où imputer les études
Propose des options permettant de spécifier le côté du graphique en entonnoir sur lequel les études manquantes sont imputées.

Déterminé par les pentes du test d'Egger
Le paramètre par défaut détermine le côté en fonction de la pente estimée du test d'Egger.

Gauche
Impute le côté gauche du graphique en entonnoir.

Droite
Impute le côté droit du graphique en entonnoir.

Méthode
Indique la méthode permettant d'estimer le nombre d'études manquantes.

Linéaire
Le paramètre par défaut calcule l'estimateur linéaire.

Exécuter
Calcule l'estimateur d'exécution.

Quadratique
Calcule l'estimateur quadratique.

Processus d'itération
Fournit des paramètres permettant de spécifier l'estimateur d'itération et l'ajustement d'erreur standard.

Modèle à effets fixes
Si ce paramètre est sélectionné, un modèle à effets fixes est utilisé et les options d'estimation d'itération et d'ajustement d'erreur standard ne sont pas disponibles.

Modèle à effets aléatoires
Si ce paramètre est sélectionné, un modèle à effets aléatoires est utilisé et les paramètres suivants sont disponibles.

Estimateur
Fournit des paramètres permettant de spécifier l'estimateur d'itération.

Maximum de vraisemblance restreint (REML)
Le paramètre par défaut applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance restreint.

Maximum de vraisemblance (ML)
Applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance.

Bayes empirique
Applique la méthode itérative et calcule l'estimateur Bayes empirique.

Hedges
Applique la méthode non itérative et calcule l'estimateur Hedges.

Hunter-Schmidt
Applique la méthode non itérative et calcule l'estimateur Hunter-Schmidt.

DerSimonian-Laird
Applique la méthode non itérative et calcule l'estimateur DerSimonian-Laird.

Sidik-Jonkman
Applique la méthode non itérative et calcule l'estimateur Sidik-Jonkman.

Ajustement de l'erreur standard
Fournit des paramètres qui déterminent si l'ajustement d'erreur standard Knapp-Hartung doit être appliqué ou non aux itérations de l'algorithme de troncature et remplissage.

Aucun ajustement
Le paramètre par défaut n'applique pas l'ajustement.

Appliquer l'ajustement de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung.

Appliquer l'ajustement tronqué de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung et tronque la valeur si elle est inférieure à 1 lors de l'estimation de la matrice variance-covariance.

Définition des paramètres de troncature et remplissage continus pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse continue**, cliquez sur **Tronquer et remplir**.
3. Sélectionnez et définissez les paramètres de troncature et de remplissage appropriés.
4. Cliquez sur **Continuer**

Méta-analyse continue : Imprimer

La boîte de dialogue **Imprimer** propose des paramètres permettant de contrôler les sorties de tableau pour une méta-analyse avec des résultats continus sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Homogénéité/hétérogénéité
Fournit des paramètres permettant de contrôler les tests d'homogénéité et d'hétérogénéité.

Test d'homogénéité
Lorsque ce paramètre est sélectionné, le test d'homogénéité correspondant est fourni dans la sortie.

Mesures d'hétérogénéité
Lorsque ce paramètre est sélectionné, les mesures d'hétérogénéité sont fournies dans la sortie.

Tailles d'effet
Fournit les paramètres de taille d'effet ci-dessous.

Etudes individuelles
Contrôle l'affichage des études individuelles. Si ce paramètre est sélectionné, la sortie correspondante est fournie.

Tailles d'effet cumulées
Contrôle l'affichage de l'analyse cumulée. Si ce paramètre est sélectionné, la sortie correspondante est fournie. Ce paramètre n'est disponible que lorsqu'une variable **Analyse cumulée** est sélectionnée dans la boîte de dialogue **Analyse**.

Intervalle de prévision dans un modèle à effets aléatoires
Contrôle l'affichage de l'intervalle de prévision. Si ce paramètre est sélectionné, la sortie correspondante est fournie. Ce paramètre n'est disponible que lorsqu'un modèle à effets aléatoires est sélectionné.

Définition des paramètres d'impression de la méta-analyse continue

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse continue**, cliquez sur **Imprimer**.
3. Sélectionnez et définissez les paramètres d'impression appropriés.
4. Cliquez sur **Continuer**

Méta-analyse continue : Enregistrer

La boîte de dialogue **Enregistrer** propose des paramètres permettant de sauvegarder les statistiques estimées dans le jeu de données actif pour une méta-analyse avec des résultats continus sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Etudes individuelles
Fournit des options permettant d'enregistrer la taille d'effet estimée.

Taille d'effet individuelle
Enregistre la taille d'effet estimée.

Erreur standard
Enregistre l'erreur standard estimée de la taille d'effet.

Limite inférieure de l'intervalle de confiance
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet.

Limite supérieure de l'intervalle de confiance
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet.

Valeur P
Enregistre la valeur *p* estimée de la taille d'effet.

Pondération de l'étude
Enregistre la pondération estimée de l'étude.

Pourcentage de la pondération de l'étude
Enregistre la pondération d'étude normalisée sous forme de pourcentage.

Définition des paramètres d'enregistrement de la méta-analyse continue

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse continue**, cliquez sur **Enregistrer**.
3. Sélectionnez et définissez les paramètres d'enregistrement de statistiques estimées appropriés.
4. Cliquez sur **Continuer**

Méta-analyse continue : Tracé

La boîte de dialogue **Tracé** propose des paramètres des types de tracé suivants :

- «Graphique en forêt», à la page 80
- «Graphique en forêt cumulée», à la page 81
- «Graphique à bulles», à la page 82
- «Graphique en entonnoir», à la page 83
- «Graphique de Galbraith», à la page 84

Méta-analyse : Taille d'effet continue

La procédure Méta-analyse : Taille d'effet continue effectue une méta-analyse avec des résultats continus lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Exemple
Plusieurs travaux de recherche ont été réalisés tout au long de l'histoire pour étudier un médicament à la mode mais discutable pour traiter le diabète de type II. Le médicament oral était censé réduire le taux de glucose dans le sang après les repas. Des données ont été recueillies sur différents sites de recherche de 1979 à 1986.

Une chercheuse principale souhaite établir des statistiques sur l'effet du médicament oral. Etant donné que les données ont été générées à partir d'études différentes, elle a proposé l'idée de

synthétiser les résultats dans toutes les études afin de comprendre globalement l'effet et d'identifier les sources sous-jacentes de variation dans les résultats.

Statistiques
Niveau de confiance, méthode itérative, dichotomie, tolérance de convergence, moyennes d'échantillons, variance d'échantillon, écart type, taille d'effet estimée, *d* de Cohen, *g* de Hedges, delta de Glass, différence moyenne, analyse cumulée, méthode d'estimation, troncature et remplissage, test de régression, modèle à effets aléatoires, modèle à effets fixes, estimateur de vraisemblance maximale restreinte, estimateur Bayes empirique, estimateur Hedges, estimateur Hunter-Schmidt, estimateur DerSimonian-Laird, estimateur Sidik-Jonkman, ajustement d'erreur standard Knapp-Hartung, ajustement d'erreur standard Knapp-Hartung tronqué, coefficients, test de régression d'EGGER, constante, modèle multiplicatif, paramètre de dispersion multiplicative, estimateur quadratique, test d'homogénéité, mesures d'hétérogénéité, intervalle de prévision, erreur standard estimée, valeur *p* estimée, taille d'effet globale cumulée, pondération d'étude estimée.

Obtention d'une analyse de taille d'effet continue pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Taille d'effet précalculée**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une variable **Taille d'effet** qui indique la taille d'un effet. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
3. Sélectionnez **Ecart type** pour spécifier l'écart type de taille d'effet ou **Variance** pour spécifier la variance de taille d'effet, puis sélectionnez une variable pour représenter l'écart type/variance pour la taille d'effet.
4. (Facultatif) Sélectionnez les variables **ID étude** et/ou **Libellé de l'étude**. La variable **ID étude** sélectionnée ne peut pas être identique à la variable **Libellé de l'étude** sélectionnée.
5. (Facultatif) Sélectionnez un paramètre **Modèle**. Lorsque les paramètres **Tronquer et remplir** sont activés, ce paramètre contrôle également le modèle utilisé par le regroupement dans l'analyse de troncature et de remplissage. Lorsque les paramètres **Biais** sont activés, ce paramètre contrôle également le modèle utilisé par le regroupement dans le test de régression.

   Effets aléatoires
   La valeur par défaut permet de générer le modèle à effets aléatoires.

   Effets fixes
   Génère le modèle à effets fixes.
6. (En option) Vous avez également ces possibilités :
   - Cliquez sur **Critères...** pour spécifier les critères généraux.
   - Cliquez sur **Analyse** pour spécifier le sous-groupe et l'analyse cumulée.
   - Cliquer sur **Inférence** pour spécifier les méthodes d'estimation.
   - Cliquez sur **Contraste** pour contrôler le test de contraste.
   - Cliquez sur **Biais** pour accéder au biais de publication en effectuant le test de régression d'EGGER.
   - Cliquez sur **Tronquer et remplir** pour mettre en œuvre l'analyse de troncature et de remplissage du biais de publication.
   - Cliquez sur **Imprimer** pour contrôler les sorties de tableau.
   - Cliquez sur **Enregistrer** pour enregistrer les statistiques estimées dans le jeu de données actif.
   - Cliquer sur **Tracé** pour spécifier les tracés à inclure dans la sortie.
7. Cliquez sur **OK**.

Méta-analyse : Taille d'effet continue : Critères

La boîte de dialogue **Critères** propose des paramètres permettant de spécifier des critères pour une méta-analyse avec des résultats continus lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Intervalle de confiance
Ce paramètre facultatif spécifie le niveau de confiance. Il doit s'agir d'une valeur numérique comprise entre 0 et 100. La valeur par défaut est 95.

La portée des données est manquante
Les paramètres facultatifs contrôlent la façon dont la procédure traite les données manquantes.

Exclure les observations analyse par analyse
Le paramètre par défaut inclut toutes les observations ayant suffisamment de données sur les variables utilisées dans chaque analyse particulière.

Exclure toute observation incomplète
Inclut toutes les observations ayant suffisamment de données sur toutes les variables utilisées dans toutes les analyses spécifiées par la procédure.

Valeurs utilisateur manquantes
Les paramètres facultatifs contrôlent la façon dont les valeurs utilisateur manquantes sont traitées.

Exclure
Le paramètre par défaut considère les valeurs utilisateur manquantes comme étant valides.

Inclure
Ignore les désignations de valeurs utilisateur manquantes et considère ces dernières comme étant valides.

Itérations

Nombre maximum d'itérations
Ce paramètre facultatif spécifie le nombre maximal d'itérations dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 100. La valeur 0 indique qu'aucune itération n'est effectuée.

Découpage maximal d'étape en deux
Ce paramètre facultatif spécifie le nombre maximal de dichotomies dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 5. Une valeur de 0 signifie que le step-halving n'est pas appliqué.

Convergence
Ce paramètre facultatif spécifie la tolérance de convergence. La valeur doit correspondre à une valeur positive unique. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur par défaut est 1E-6.

Définition des critères de taille d'effet continue pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Taille d'effet précalculée**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet continue**, cliquez sur **Critères**.
3. Sélectionnez et définissez les paramètres de critères appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet continue : Analyse

La boîte de dialogue **Analyse** propose des paramètres permettant de spécifier des analyses cumulées et de sous-groupe pour une méta-analyse avec des résultats continus lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Analyse de sous-groupe
Sélectionnez une variable évoquant l'analyse de sous-groupe. La variable ne peut pas être identique à la variable définie pour **Analyse cumulée**.

Analyse cumulée
Sélectionnez une variable évoquant l'analyse cumulée et en fonction de laquelle la méta-analyse cumulée est effectuée. La variable ne peut pas être identique à la variable définie pour **Analyse de sous-groupe**. Lorsque l'option **Croissant** est sélectionnée, l'analyse cumulée est basée sur la variable spécifiée dans l'ordre croissant. Lorsque l'option **Décroissant** est sélectionnée, l'analyse cumulée est basée sur la variable spécifiée dans l'ordre décroissant.

Statistiques cumulées
Propose des options permettant d'enregistrer la taille d'effet globale cumulée estimée. Les paramètres ne sont disponibles que lorsqu'une variable **Analyse cumulée** est sélectionnée.

Taille d'effet cumulée
Enregistre la taille d'effet globale cumulée estimée.

Erreur standard
Enregistre l'erreur standard estimée de la taille d'effet globale cumulée.

Limite inférieure de l'intervalle de confiance
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée.

Limite supérieure de l'intervalle de confiance
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée.

Valeur P
Enregistre la valeur *p* estimée de la taille d'effet globale cumulée.

Destination
Fournit des options permettant de spécifier un fichier ou jeu de données de sauvegarde. Lorsque **Jeu de données** est sélectionné, vous pouvez spécifier un nouveau nom de jeu de données (vous pouvez conserver le nom de jeu de données par défaut). Lorsque le **fichier de données** est sélectionné, cliquez sur **Parcourir...** pour sélectionner le nom et l'emplacement du fichier d'enregistrement.

Définition des paramètres d'analyse de taille d'effet continue pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Taille d'effet précalculée**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet continue**, cliquez sur **Analyse**.
3. Sélectionnez et définissez les paramètres d'analyse appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet continue : Inférence

La boîte de dialogue **Inférence** propose des paramètres permettant de spécifier les méthodes d'estimation pour une méta-analyse avec des résultats continus lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Remarque : La boîte de dialogue **Inférence** n'est disponible que lorsqu'un modèle à **effets aléatoires** est sélectionné.

Estimateur
Fournit des paramètres permettant de spécifier l'estimateur.

Remarques :

- Lorsque les paramètres Tronquer et remplir sont spécifiés, ce paramètre contrôle également l'estimateur qui est utilisé par le regroupement dans l'analyse de troncature et de remplissage.

- Lorsque les paramètres de biais sont spécifiés, ce paramètre contrôle également l'estimateur qui est utilisé par le test de régression.

Maximum de vraisemblance restreint (REML)
Le paramètre par défaut applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance restreint.

Maximum de vraisemblance (ML)
Applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance.

Bayes empirique
Applique la méthode itérative et calcule l'estimateur Bayes empirique.

Hedges
Applique la méthode non itérative et calcule l'estimateur Hedges.

Hunter-Schmidt
Applique la méthode non itérative et calcule l'estimateur Hunter-Schmidt.

DerSimonian-Laird
Applique la méthode non itérative et calcule l'estimateur DerSimonian-Laird.

Sidik-Jonkman
Applique la méthode non itérative et calcule l'estimateur Sidik-Jonkman.

Ajustement de l'erreur standard
Fournit des paramètres qui déterminent si l'ajustement d'erreur standard Knapp-Hartung doit être appliqué ou non.

Remarques :

- Lorsque les paramètres Tronquer et remplir sont spécifiés, ce paramètre contrôle également l'ajustement d'erreur standard qui est utilisé par le regroupement dans l'analyse de troncature et de remplissage.
- Lorsque les paramètres de biais sont spécifiés, ce paramètre contrôle également l'ajustement d'erreur standard qui est utilisé par le test de régression.

Aucun ajustement
Le paramètre par défaut n'applique pas l'ajustement.

Appliquer l'ajustement de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung.

Appliquer l'ajustement tronqué de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung et tronque la valeur si elle est inférieure à 1 lors de l'estimation de la matrice variance-covariance.

Définition des paramètres d'inférence de taille d'effet continue pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Taille d'effet précalculée**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet continue**, cliquez sur **Inférence**.
3. Sélectionnez et définissez les paramètres d'inférence appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet continue : Contraste

La boîte de dialogue **Contraste** propose des paramètres permettant de contrôler le test de contraste pour une méta-analyse avec des résultats continus lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Variables
La liste affiche toutes les variables de jeu de données disponibles. Sélectionnez des variables dans la liste et déplacez-les vers la liste **Contraste(s)**.

Contraste(s)
La liste identifie les coefficients qui sont stockés sous forme de variables dans le jeu de données actif. Plusieurs variables sont autorisées. Les variables de chaîne ne sont pas prises en charge.

Valeurs de coefficient entrées par l'utilisateur
Fournit des paramètres permettant de spécifier les coefficients d'un contraste fourni par l'utilisateur. Seules les valeurs numériques sont admises. Pour formuler un test de contraste valide, le nombre de valeurs spécifiées doit correspondre à celui des études valides.

Définition des paramètres de contraste de taille d'effet continue pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Taille d'effet précalculée**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet continue**, cliquez sur **Contraste**.
3. Sélectionnez et définissez les paramètres de contraste appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet continue : Biais

La boîte de dialogue **Biais** propose des paramètres permettant d'activer le biais de publication en effectuant le test de régression d'Egger pour une méta-analyse avec des résultats continus lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Test de régression d'Egger
Sélectionnez ce paramètre pour activer le biais de publication en effectuant le test de régression d'Egger.

Variables
La liste fournit toutes les variables de jeu de données disponibles.

Covariable(s)
Les variables sélectionnées dans la liste **Variables** sont traitées sous forme de covariables. Plusieurs covariables sont autorisées.

Facteur(s)
Les variables sélectionnées dans la liste **Variables** sont traitées sous forme de facteurs. Plusieurs facteurs sont autorisés.

Inclure une constante dans la régression
Contrôle le terme de la constante dans le test de régression.

Inclure un paramètre de dispersion dans le modèle à effets fixes
Contrôle le paramètre de modèle multiplicatif et introduit le paramètre de dispersion multiplicative dans l'analyse. Ce paramètre n'est disponible que lorsqu'un modèle à effets fixes est sélectionné.

Estimer les statistiques sur la base de la distribution t
Contrôle la distribution utilisée dans les tests de régression. Ce paramètre est activé par défaut, et permet d'estimer les statistiques en fonction de la distribution *t*. Lorsqu'il n'est pas sélectionné, les statistiques sont estimées en fonction de la distribution normale.

Définition des paramètres de biais de taille d'effet continue pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Taille d'effet précalculée**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet continue**, cliquez sur **Biais**.
3. Sélectionnez et définissez les paramètres de biais appropriés.

4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet continue : Tronquer et remplir

La boîte de dialogue **Tronquer et remplir** propose des paramètres permettant de mettre en oeuvre l'analyse de troncature et de remplissage du biais de publication pour une méta-analyse avec des résultats continus lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Nombre d'études manquantes estimé
Contrôle l'analyse de troncature et de remplissage du biais de publication. En sélectionnant ce paramètre, vous activez les autres paramètres de boîte de dialogue.

Côté où imputer les études
Propose des options permettant de spécifier le côté du graphique en entonnoir sur lequel les études manquantes sont imputées.

Déterminé par les pentes du test d'Egger
Le paramètre par défaut détermine le côté en fonction de la pente estimée du test d'Egger.

Gauche
Impute le côté gauche du graphique en entonnoir.

Droite
Impute le côté droit du graphique en entonnoir.

Méthode
Indique la méthode permettant d'estimer le nombre d'études manquantes.

Linéaire
Le paramètre par défaut calcule l'estimateur linéaire.

Exécuter
Calcule l'estimateur d'exécution.

Quadratique
Calcule l'estimateur quadratique.

Processus d'itération
Fournit des paramètres permettant de spécifier l'estimateur d'itération et l'ajustement d'erreur standard.

Modèle à effets fixes
Si ce paramètre est sélectionné, un modèle à effets fixes est utilisé et les options d'estimation d'itération et d'ajustement d'erreur standard ne sont pas disponibles.

Modèle à effets aléatoires
Si ce paramètre est sélectionné, un modèle à effets aléatoires est utilisé et les paramètres suivants sont disponibles.

Estimateur
Fournit des paramètres permettant de spécifier l'estimateur d'itération.

Maximum de vraisemblance restreint (REML)
Le paramètre par défaut applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance restreint.

Maximum de vraisemblance (ML)
Applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance.

Bayes empirique
Applique la méthode itérative et calcule l'estimateur Bayes empirique.

Hedges
Applique la méthode non itérative et calcule l'estimateur Hedges.

Hunter-Schmidt
Applique la méthode non itérative et calcule l'estimateur Hunter-Schmidt.

DerSimonian-Laird
Applique la méthode non itérative et calcule l'estimateur DerSimonian-Laird.

Sidik-Jonkman
Applique la méthode non itérative et calcule l'estimateur Sidik-Jonkman.

Ajustement de l'erreur standard
Fournit des paramètres qui déterminent si l'ajustement d'erreur standard Knapp-Hartung doit être appliqué ou non aux itérations de l'algorithme de troncature et remplissage.

Aucun ajustement
Le paramètre par défaut n'applique pas l'ajustement.

Appliquer l'ajustement de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung.

Appliquer l'ajustement tronqué de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung et tronque la valeur si elle est inférieure à 1 lors de l'estimation de la matrice variance-covariance.

Définition des paramètres de troncature et remplissage de taille d'effet continue pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Taille d'effet précalculée**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet continue**, cliquez sur **Tronquer et remplir**.
3. Sélectionnez et définissez les paramètres de troncature et de remplissage appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet continue : Imprimer

La boîte de dialogue **Imprimer** propose des paramètres permettant de contrôler les sorties de tableau pour une méta-analyse avec des résultats continus lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Homogénéité/hétérogénéité
Fournit des paramètres permettant de contrôler les tests d'homogénéité et d'hétérogénéité.

Test d'homogénéité
Lorsque ce paramètre est sélectionné, le test d'homogénéité correspondant est fourni dans la sortie.

Mesures d'hétérogénéité
Lorsque ce paramètre est sélectionné, les mesures d'hétérogénéité sont fournies dans la sortie.

Tailles d'effet
Fournit les paramètres de taille d'effet ci-dessous.

Etudes individuelles
Contrôle l'affichage des études individuelles. Si ce paramètre est sélectionné, la sortie correspondante est fournie.

Tailles d'effet cumulées
Contrôle l'affichage de l'analyse cumulée. Si ce paramètre est sélectionné, la sortie correspondante est fournie. Ce paramètre n'est disponible que lorsqu'une variable **Analyse cumulée** est sélectionnée dans la boîte de dialogue **Analyse**.

Intervalle de prévision dans un modèle à effets aléatoires
Contrôle l'affichage de l'intervalle de prévision. Si ce paramètre est sélectionné, la sortie correspondante est fournie. Ce paramètre n'est disponible que lorsqu'un modèle à effets aléatoires est sélectionné.

Définition des paramètres d'impression de taille d'effet continue pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Taille d'effet précalculée**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet continue**, cliquez sur **Imprimer**.
3. Sélectionnez et définissez les paramètres d'impression appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet continue : Enregistrer

La boîte de dialogue **Enregistrer** propose des paramètres permettant d'enregistrer les statistiques estimées dans le jeu de données actif pour une méta-analyse avec des résultats continus lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Etudes individuelles
Fournit des options permettant d'enregistrer la taille d'effet estimée.

Erreur standard
Enregistre l'erreur standard estimée de la taille d'effet. Ce paramètre n'est disponible que lorsqu'une variable **Taille d'effet** n'est pas spécifiée.

Limite inférieure de l'intervalle de confiance
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet.

Limite supérieure de l'intervalle de confiance
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet.

Valeur P
Enregistre la valeur *p* estimée de la taille d'effet.

Pondération de l'étude
Enregistre la pondération estimée de l'étude.

Pourcentage de la pondération de l'étude
Enregistre la pondération d'étude normalisée sous forme de pourcentage.

Définition des paramètres d'enregistrement de taille d'effet continue pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats continus** > **Taille d'effet précalculée**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet continue**, cliquez sur **Enregistrer**.
3. Sélectionnez et définissez les paramètres d'enregistrement de statistiques estimées appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet continue : Tracé

La boîte de dialogue **Tracé** propose des paramètres des types de tracé suivants :

- «Graphique en forêt», à la page 80
- «Graphique en forêt cumulée», à la page 81
- «Graphique à bulles», à la page 82
- «Graphique en entonnoir», à la page 83
- «Graphique de Galbraith», à la page 84

Méta-analyse binaire

La procédure Méta-analyse binaire effectue une méta-analyse avec des résultats binaires sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Exemple

Plusieurs travaux de recherche ont été réalisés tout au long de l'histoire pour étudier un médicament à la mode mais discutable pour traiter le diabète de type II. Le médicament oral était censé réduire le taux de glucose dans le sang après les repas. Des données ont été recueillies sur différents sites de recherche de 1979 à 1986.

Une chercheuse principale souhaite établir des statistiques sur l'effet du médicament oral. Etant donné que les données ont été générées à partir d'études différentes, elle a proposé l'idée de synthétiser les résultats dans toutes les études afin de comprendre globalement l'effet et d'identifier les sources sous-jacentes de variation dans les résultats.

Statistiques

Intervalle de confiance, rapport des cotes de log, rapport des cotes de log de Peto, rapport des risques de log, différence de risque, effets aléatoires, effets fixes, variance inverse, Mantel-Haenszel, itérations, dichotomie, convergence, statistiques cumulées, taille d'effet cumulée, maximum de vraisemblance restreint (REML), maximum de vraisemblance (ML), Bayes empirique, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, test d'Egger, test de Harbord, test de Peters, constante dans la régression, paramètre de dispersion, homogénéité, hétérogénéité, statistiques mises en exposant, erreur standard, valeur *p*, pondération de l'étude.

Obtention d'une analyse binaire de méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Données brutes...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Dans la section **Groupe qui suit le traitement**, sélectionnez une variable **Succès** pour représenter le nombre de "succès" pour le groupe qui suit le traitement. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
3. Sélectionnez une variable **Echec** pour représenter le nombre "d'échecs" pour le groupe qui suit le traitement. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
4. Dans la section **Groupe de contrôle**, sélectionnez une variable **Succès** pour représenter le nombre de "succès" pour le groupe de contrôle. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
5. Sélectionnez une variable **Echec** pour représenter le nombre "d'échecs" pour le groupe de contrôle. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
6. (Facultatif) Sélectionnez les variables **ID étude** et/ou **Libellé de l'étude**. La variable **ID étude** sélectionnée ne peut pas être identique à la variable **Libellé de l'étude** sélectionnée.
7. (Facultatif) Sélectionnez un paramètre **Taille d'effet**. Les options disponibles sont **Rapport des cotes de log**, **Rapport des cotes de log de Peto**, **Rapport des risques de log** et **Différence de risque**.
8. (Facultatif) Sélectionnez un paramètre **Modèle**. Lorsque les paramètres **Tronquer et remplir** sont activés, ce paramètre contrôle également le modèle utilisé par le regroupement dans l'analyse de troncature et de remplissage. Lorsque les paramètres **Biais** sont activés, ce paramètre contrôle également le modèle utilisé par le regroupement dans le test de régression.

   Effets aléatoires
   La valeur par défaut permet de générer le modèle à effets aléatoires.

   Effets fixes
   Génère le modèle à effets fixes. **Variance-inverse** estime la pondération de la variance inverse. **Mantel-Haenszel** estime la pondération Mantel-Haenszel.

9. (En option) Vous avez également ces possibilités :
   - Cliquez sur **Critères...** pour spécifier les critères généraux.
   - Cliquez sur **Analyse** pour spécifier le sous-groupe et l'analyse cumulée.
   - Cliquer sur **Inférence** pour spécifier les méthodes d'estimation.
   - Cliquez sur **Contraste** pour contrôler le test de contraste.
   - Cliquez sur **Biais** pour accéder au biais de publication en effectuant le test de régression d'EGGER.
   - Cliquez sur **Tronquer et remplir** pour mettre en œuvre l'analyse de troncature et de remplissage du biais de publication.
   - Cliquez sur **Imprimer** pour contrôler les sorties de tableau.
   - Cliquez sur **Enregistrer** pour enregistrer les statistiques estimées dans le jeu de données actif.
   - Cliquer sur **Tracé** pour spécifier les tracés à inclure dans la sortie.
10. Cliquez sur **OK**.

Méta-analyse binaire : Critères

La boîte de dialogue **Critères** propose des paramètres permettant de spécifier des critères pour une méta-analyse avec des résultats binaires sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Intervalle de confiance
Ce paramètre facultatif spécifie le niveau de confiance. Il doit s'agir d'une valeur numérique comprise entre 0 et 100. La valeur par défaut est 95.

La portée des données est manquante
Les paramètres facultatifs contrôlent la façon dont la procédure traite les données manquantes.

Exclure les observations analyse par analyse
Le paramètre par défaut inclut toutes les observations ayant suffisamment de données sur les variables utilisées dans chaque analyse particulière.

Exclure toute observation incomplète
Inclut toutes les observations ayant suffisamment de données sur toutes les variables utilisées dans toutes les analyses spécifiées par la procédure.

Valeurs utilisateur manquantes
Les paramètres facultatifs contrôlent la façon dont les valeurs utilisateur manquantes sont traitées.

Exclure
Le paramètre par défaut considère les valeurs utilisateur manquantes comme étant valides.

Inclure
Ignore les désignations de valeurs utilisateur manquantes et considère ces dernières comme étant valides.

Itérations

Nombre maximum d'itérations
Ce paramètre facultatif spécifie le nombre maximal d'itérations dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 100. La valeur 0 indique qu'aucune itération n'est effectuée.

Découpage maximal d'étape en deux
Ce paramètre facultatif spécifie le nombre maximal de dichotomies dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 5. Une valeur de 0 signifie que le step-halving n'est pas appliqué.

Convergence
Ce paramètre facultatif spécifie la tolérance de convergence. La valeur doit correspondre à une valeur positive unique. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur par défaut est 1E-6.

Ajuster les effectifs nuls
Les paramètres facultatifs contrôlent la façon d'ajuster les données d'effectifs nuls, le cas échéant, dans l'estimation de la taille d'effet.

Ajuster chaque valeur des seules études contenant au moins un zéro
Le paramètre par défaut ajuste uniquement les données d'effectifs nuls.

Ajuster chaque valeur de toutes les études uniquement s'il existe au moins une étude contenant une valeur nulle
Ajuste toutes les fréquences lorsqu'au moins un effectif nul existe dans une étude.

Ajuster chaque valeur de toutes les études
Ajuste toutes les données indépendamment de l'existence d'effectifs nuls.

N'effectuer aucun ajustement et conserver la ou les valeurs nulles
N'effectue aucun ajustement de données.

Valeur ajoutée
Ce paramètre facultatif spécifie la valeur ajoutée aux données d'effectifs nuls. La valeur par défaut est 0,5. La valeur spécifiée doit correspondre à une valeur numérique unique supérieure à 0 et inférieure ou égale à 1.

Définition des critères de la méta-analyse binaire

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse binaire**, cliquez sur **Critères**.
3. Sélectionnez et définissez les paramètres de critères appropriés.
4. Cliquez sur **Continuer**

Méta-analyse binaire : Analyse

La boîte de dialogue **Analyse** propose des paramètres permettant de spécifier des analyses cumulées et de sous-groupe pour une méta-analyse avec des résultats binaires sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Analyse de sous-groupe
Sélectionnez une variable évoquant l'analyse de sous-groupe. La variable ne peut pas être identique à la variable définie pour **Analyse cumulée**.

Analyse cumulée
Sélectionnez une variable évoquant l'analyse cumulée et en fonction de laquelle la méta-analyse cumulée est effectuée. La variable ne peut pas être identique à la variable définie pour **Analyse de sous-groupe**. Lorsque l'option **Croissant** est sélectionnée, l'analyse cumulée est basée sur la variable spécifiée dans l'ordre croissant. Lorsque l'option **Décroissant** est sélectionnée, l'analyse cumulée est basée sur la variable spécifiée dans l'ordre décroissant.

Statistiques cumulées
Propose des options permettant d'enregistrer la taille d'effet globale cumulée estimée. Les paramètres ne sont disponibles que lorsqu'une variable **Analyse cumulée** est sélectionnée.

Taille d'effet cumulée
Enregistre la taille d'effet globale cumulée estimée.

Taille d'effet cumulée (sous forme d'exposant)
Enregistre la taille d'effet globale cumulée estimée sous forme d'exposant.

Erreur standard
Enregistre l'erreur standard estimée de la taille d'effet globale cumulée.

Limite inférieure de l'intervalle de confiance
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée.

Limite supérieure de l'intervalle de confiance
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée.

Limite inférieure de l'intervalle de confiance (sous forme d'exposant)
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée sous forme d'exposant.

Limite supérieure de l'intervalle de confiance (sous forme d'exposant)
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée sous forme d'exposant.

Valeur P
Enregistre la valeur *p* estimée de la taille d'effet globale cumulée.

Destination
Fournit des options permettant de spécifier un fichier ou jeu de données de sauvegarde. Lorsque **Jeu de données** est sélectionné, vous pouvez spécifier un nouveau nom de jeu de données (vous pouvez conserver le nom de jeu de données par défaut). Lorsque le **fichier de données** est sélectionné, cliquez sur **Parcourir...** pour sélectionner le nom et l'emplacement du fichier d'enregistrement.

Définition des paramètres d'analyse de la méta-analyse binaire

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse binaire**, cliquez sur **Analyse**.
3. Sélectionnez et définissez les paramètres d'analyse appropriés.
4. Cliquez sur **Continuer**

Méta-analyse binaire : Inférence

La boîte de dialogue **Inférence** propose des paramètres permettant de spécifier les méthodes d'estimation pour une méta-analyse avec des résultats binaires sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Remarque : La boîte de dialogue **Inférence** n'est disponible que lorsqu'un modèle à **effets aléatoires** est sélectionné.

Estimateur
Fournit des paramètres permettant de spécifier l'estimateur.

Remarques :

- Lorsque les paramètres Tronquer et remplir sont spécifiés, ce paramètre contrôle également l'estimateur qui est utilisé par le regroupement dans l'analyse de troncature et de remplissage.
- Lorsque les paramètres de biais sont spécifiés, ce paramètre contrôle également l'estimateur qui est utilisé par le test de régression.

Maximum de vraisemblance restreint (REML)
Le paramètre par défaut applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance restreint.

Maximum de vraisemblance (ML)
Applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance.

Bayes empirique
Applique la méthode itérative et calcule l'estimateur Bayes empirique.

Hedges
Applique la méthode non itérative et calcule l'estimateur Hedges.

Hunter-Schmidt
Applique la méthode non itérative et calcule l'estimateur Hunter-Schmidt.

DerSimonian-Laird
Applique la méthode non itérative et calcule l'estimateur DerSimonian-Laird.

Sidik-Jonkman
Applique la méthode non itérative et calcule l'estimateur Sidik-Jonkman.

Ajustement de l'erreur standard
Fournit des paramètres qui déterminent si l'ajustement d'erreur standard Knapp-Hartung doit être appliqué ou non.

Remarques :

- Lorsque les paramètres Tronquer et remplir sont spécifiés, ce paramètre contrôle également l'ajustement d'erreur standard qui est utilisé par le regroupement dans l'analyse de troncature et de remplissage.
- Lorsque les paramètres de biais sont spécifiés, ce paramètre contrôle également l'ajustement d'erreur standard qui est utilisé par le test de régression.

Aucun ajustement
Le paramètre par défaut n'applique pas l'ajustement.

Appliquer l'ajustement de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung.

Appliquer l'ajustement tronqué de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung et tronque la valeur si elle est inférieure à 1 lors de l'estimation de la matrice variance-covariance.

Définition des paramètres d'inférence de la méta-analyse binaire

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse binaire**, cliquez sur **Inférence**.
3. Sélectionnez et définissez les paramètres d'inférence appropriés.
4. Cliquez sur **Continuer**

Méta-analyse binaire : Contraste

La boîte de dialogue **Contraste** propose des paramètres permettant de contrôler le test de contraste pour une méta-analyse avec des résultats binaires sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Variables
La liste affiche toutes les variables de jeu de données disponibles. Sélectionnez des variables dans la liste et déplacez-les vers la liste **Contraste(s)**.

Contraste(s)
La liste identifie les coefficients qui sont stockés sous forme de variables dans le jeu de données actif. Plusieurs variables sont autorisées. Les variables de chaîne ne sont pas prises en charge.

Valeurs de coefficient entrées par l'utilisateur
Fournit des paramètres permettant de spécifier les coefficients d'un contraste fourni par l'utilisateur. Seules les valeurs numériques sont admises. Pour formuler un test de contraste valide, le nombre de valeurs spécifiées doit correspondre à celui des études valides.

Afficher les statistiques mises en exposant
Contrôle l'inclusion des statistiques sous forme d'exposant. Lorsque ce paramètre est sélectionné, la sortie inclut des statistiques sous forme d'exposant, y compris la taille d'effet et les limites de l'intervalle de confiance sous forme d'exposant. Le paramètre est disponible lorsque l'option **Taille**

d'effet est spécifiée comme **Rapport des cotes de log**, **Rapport des cotes de log de Peto** ou **Rapport des risques de log**.

Définition des paramètres de contraste de la méta-analyse binaire

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse binaire**, cliquez sur **Contraste**.
3. Sélectionnez et définissez les paramètres de contraste appropriés.
4. Cliquez sur **Continuer**

Méta-analyse binaire : Biais

La boîte de dialogue **Biais** propose des paramètres permettant d'activer le biais de publication en effectuant des tests de régression pour une méta-analyse avec des résultats binaires sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Tests de régression
Propose des options permettant de spécifier les tests de régression. Plusieurs tests peuvent être sélectionnés.

Test d'Egger
Si cette option est sélectionnée, le test d'Egger est réalisé.

Test de Harbord
Si cette option est sélectionnée, le test de Harbord est réalisé. Le test est disponible lorsque l'option **Taille d'effet** est spécifiée comme **Rapport des cotes de log** ou **Rapport des risques de log**.

Test de Peters
Si cette option est sélectionnée, le test de Peters est réalisé. Le test est disponible lorsque l'option **Taille d'effet** est spécifiée comme **Rapport des cotes de log**.

Variables
La liste fournit toutes les variables de jeu de données disponibles.

Covariable(s)
Les variables sélectionnées dans la liste **Variables** sont traitées sous forme de covariables. Plusieurs covariables sont autorisées.

Facteur(s)
Les variables sélectionnées dans la liste **Variables** sont traitées sous forme de facteurs. Plusieurs facteurs sont autorisés.

Inclure une constante dans la régression
Contrôle le terme de la constante dans le test de régression.

Inclure un paramètre de dispersion dans le modèle à effets fixes
Contrôle le paramètre de modèle multiplicatif et introduit le paramètre de dispersion multiplicative dans l'analyse. Ce paramètre n'est disponible que lorsqu'un modèle à effets fixes est sélectionné.

Estimer les statistiques sur la base de la distribution t
Contrôle la distribution utilisée dans les tests de régression. Ce paramètre est activé par défaut, et permet d'estimer les statistiques en fonction de la distribution *t*. Lorsqu'il n'est pas sélectionné, les statistiques sont estimées en fonction de la distribution normale.

Définition des paramètres de biais de la méta-analyse binaire

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse binaire**, cliquez sur **Biais**.

3. Sélectionnez et définissez les paramètres de biais appropriés.
4. Cliquez sur **Continuer**

Méta-analyse binaire : Tronquer et remplir

La boîte de dialogue **Tronquer et remplir** propose des paramètres permettant de mettre en oeuvre l'analyse de troncature et de remplissage du biais de publication pour une méta-analyse avec des résultats binaires sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Nombre d'études manquantes estimé
Contrôle l'analyse de troncature et de remplissage du biais de publication. En sélectionnant ce paramètre, vous activez les autres paramètres de boîte de dialogue.

Côté où imputer les études
Propose des options permettant de spécifier le côté du graphique en entonnoir sur lequel les études manquantes sont imputées.

Déterminé par les pentes du test d'Egger
Le paramètre par défaut détermine le côté en fonction de la pente estimée du test d'Egger.

Gauche
Impute le côté gauche du graphique en entonnoir.

Droite
Impute le côté droit du graphique en entonnoir.

Méthode
Indique la méthode permettant d'estimer le nombre d'études manquantes.

Linéaire
Le paramètre par défaut calcule l'estimateur linéaire.

Exécuter
Calcule l'estimateur d'exécution.

Quadratique
Calcule l'estimateur quadratique.

Processus d'itération
Fournit des paramètres permettant de spécifier l'estimateur d'itération et l'ajustement d'erreur standard.

Modèle à effets fixes
Si ce paramètre est sélectionné, un modèle à effets fixes est utilisé et les options d'estimation d'itération et d'ajustement d'erreur standard ne sont pas disponibles.

Modèle à effets aléatoires
Si ce paramètre est sélectionné, un modèle à effets aléatoires est utilisé et les paramètres suivants sont disponibles.

Estimateur
Fournit des paramètres permettant de spécifier l'estimateur d'itération.

Maximum de vraisemblance restreint (REML)
Le paramètre par défaut applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance restreint.

Maximum de vraisemblance (ML)
Applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance.

Bayes empirique
Applique la méthode itérative et calcule l'estimateur Bayes empirique.

Hedges
Applique la méthode non itérative et calcule l'estimateur Hedges.

Hunter-Schmidt
Applique la méthode non itérative et calcule l'estimateur Hunter-Schmidt.

DerSimonian-Laird
Applique la méthode non itérative et calcule l'estimateur DerSimonian-Laird.

Sidik-Jonkman
Applique la méthode non itérative et calcule l'estimateur Sidik-Jonkman.

Ajustement de l'erreur standard
Fournit des paramètres qui déterminent si l'ajustement d'erreur standard Knapp-Hartung doit être appliqué ou non aux itérations de l'algorithme de troncature et remplissage.

Aucun ajustement
Le paramètre par défaut n'applique pas l'ajustement.

Appliquer l'ajustement de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung.

Appliquer l'ajustement tronqué de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung et tronque la valeur si elle est inférieure à 1 lors de l'estimation de la matrice variance-covariance.

Définition des paramètres de troncature et de remplissage de la méta-analyse binaire

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse binaire**, cliquez sur **Tronquer et remplir**.
3. Sélectionnez et définissez les paramètres de troncature et de remplissage appropriés.
4. Cliquez sur **Continuer**

Méta-analyse binaire : Imprimer

La boîte de dialogue **Imprimer** propose des paramètres permettant de contrôler les sorties de tableau pour une méta-analyse avec des résultats binaires sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Homogénéité/hétérogénéité
Fournit des paramètres permettant de contrôler les tests d'homogénéité et d'hétérogénéité.

Test d'homogénéité
Lorsque ce paramètre est sélectionné, le test d'homogénéité correspondant est fourni dans la sortie.

Mesures d'hétérogénéité
Lorsque ce paramètre est sélectionné, les mesures d'hétérogénéité sont fournies dans la sortie.

Tailles d'effet
Fournit les paramètres de taille d'effet ci-dessous.

Etudes individuelles
Contrôle l'affichage des études individuelles. Si ce paramètre est sélectionné, la sortie correspondante est fournie.

Tailles d'effet cumulées
Contrôle l'affichage de l'analyse cumulée. Si ce paramètre est sélectionné, la sortie correspondante est fournie. Ce paramètre n'est disponible que lorsqu'une variable **Analyse cumulée** est sélectionnée dans la boîte de dialogue **Analyse**.

Intervalle de prévision dans un modèle à effets aléatoires
Contrôle l'affichage de l'intervalle de prévision. Si ce paramètre est sélectionné, la sortie correspondante est fournie. Ce paramètre n'est disponible que lorsqu'un modèle à effets aléatoires est sélectionné.

Afficher les statistiques mises en exposant
Contrôle l'inclusion des statistiques sous forme d'exposant. Lorsque ce paramètre est sélectionné, les statistiques sous forme d'exposant, y compris la taille d'effet et les limites de l'intervalle de confiance sous forme d'exposant sont générées. Le processus s'applique aux tables d'estimation et de prévision de taille d'effet obligatoires et facultatives. Le paramètre est disponible lorsque l'option **Taille d'effet** est spécifiée comme **Rapport des cotes de log**, **Rapport des cotes de log de Peto** ou **Rapport des risques de log**.

Définition des paramètres d'impression de la méta-analyse binaire

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse binaire**, cliquez sur **Imprimer**.
3. Sélectionnez et définissez les paramètres d'impression appropriés.
4. Cliquez sur **Continuer**

Méta-analyse binaire : Enregistrer

La boîte de dialogue **Enregistrer** propose des paramètres permettant de sauvegarder les statistiques estimées dans le jeu de données actif pour une méta-analyse avec des résultats binaires sur des données brutes fournies dans le jeu de données actif pour l'estimation de la taille d'effet.

Etudes individuelles
Fournit des options permettant d'enregistrer la taille d'effet estimée.

Taille d'effet individuelle
Enregistre la taille d'effet estimée.

Taille d'effet individuelle (sous forme d'exposant)
Enregistre la taille d'effet estimée sous forme d'exposant.

Erreur standard
Enregistre l'erreur standard estimée de la taille d'effet.

Limite inférieure de l'intervalle de confiance
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet.

Limite supérieure de l'intervalle de confiance
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet.

Limite inférieure de l'intervalle de confiance (sous forme d'exposant)
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet sous forme d'exposant.

Limite supérieure de l'intervalle de confiance (sous forme d'exposant)
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet sous forme d'exposant.

Valeur P
Enregistre la valeur *p* estimée de la taille d'effet.

Pondération de l'étude
Enregistre la pondération estimée de l'étude.

Pourcentage de la pondération de l'étude
Enregistre la pondération d'étude normalisée sous forme de pourcentage.

Définition des paramètres d'enregistrement de la méta-analyse binaire

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse binaire**, cliquez sur **Enregistrer**.

3. Sélectionnez et définissez les paramètres d'enregistrement de statistiques estimées appropriés.
4. Cliquez sur **Continuer**

Méta-analyse binaire : Tracé

La boîte de dialogue **Tracé** propose des paramètres des types de tracé suivants :

- «Graphique en forêt», à la page 80
- «Graphique en forêt cumulée», à la page 81
- «Graphique à bulles», à la page 82
- «Graphique en entonnoir», à la page 83
- «Graphique de Galbraith», à la page 84
- «Graphique de l'Abbé», à la page 85

Méta-analyse : Taille d'effet binaire

La procédure Méta-analyse : Taille d'effet binaire effectue une méta-analyse avec des résultats binaires lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Exemple

Plusieurs travaux de recherche ont été réalisés tout au long de l'histoire pour étudier un médicament à la mode mais discutable pour traiter le diabète de type II. Le médicament oral était censé réduire le taux de glucose dans le sang après les repas. Des données ont été recueillies sur différents sites de recherche de 1979 à 1986.

Une chercheuse principale souhaite établir des statistiques sur l'effet du médicament oral. Etant donné que les données ont été générées à partir d'études différentes, elle a proposé l'idée de synthétiser les résultats dans toutes les études afin de comprendre globalement l'effet et d'identifier les sources sous-jacentes de variation dans les résultats.

Statistiques

Intervalle de confiance, rapport des cotes de log, rapport des cotes de log de Peto, rapport des risques de log, différence de risque, effets aléatoires, effets fixes, variance inverse, Mantel-Haenszel, itérations, dichotomie, convergence, statistiques cumulées, taille d'effet cumulée, maximum de vraisemblance restreint (REML), maximum de vraisemblance (ML), Bayes empirique, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, test d'Egger, test de Harbord, test de Peters, constante dans la régression, paramètre de dispersion, homogénéité, hétérogénéité, statistiques mises en exposant, erreur standard, valeur *p*, pondération de l'étude.

Obtention d'une analyse de taille d'effet binaire pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Taille d'effet précalculée...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une variable **Taille d'effet** qui indique la taille d'un effet. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
3. Sélectionnez **Ecart type** pour spécifier l'écart type de taille d'effet ou **Variance** pour spécifier la variance de taille d'effet, puis sélectionnez une variable pour représenter l'écart type/variance pour la taille d'effet.
4. (Facultatif) Sélectionnez les variables **ID étude** et/ou **Libellé de l'étude**. La variable **ID étude** sélectionnée ne peut pas être identique à la variable **Libellé de l'étude** sélectionnée.
5. (Facultatif) Sélectionnez un paramètre **Taille d'effet**. Les options disponibles sont **Rapport des cotes de log**, **Rapport des cotes de log de Peto**, **Rapport des risques de log** et **Différence de risque**.
6. (Facultatif) Sélectionnez un paramètre **Modèle**. Lorsque les paramètres **Tronquer et remplir** sont activés, ce paramètre contrôle également le modèle utilisé par le regroupement dans l'analyse de

troncature et de remplissage. Lorsque les paramètres **Biais** sont activés, ce paramètre contrôle également le modèle utilisé par le regroupement dans le test de régression.

Effets aléatoires
La valeur par défaut permet de générer le modèle à effets aléatoires.

Effets fixes
Génère le modèle à effets fixes.

7. (En option) Vous avez également ces possibilités :
   - Cliquez sur **Critères...** pour spécifier les critères généraux.
   - Cliquez sur **Analyse** pour spécifier le sous-groupe et l'analyse cumulée.
   - Cliquer sur **Inférence** pour spécifier les méthodes d'estimation.
   - Cliquez sur **Contraste** pour contrôler le test de contraste.
   - Cliquez sur **Biais** pour accéder au biais de publication en effectuant le test de régression d'EGGER.
   - Cliquez sur **Tronquer et remplir** pour mettre en œuvre l'analyse de troncature et de remplissage du biais de publication.
   - Cliquez sur **Imprimer** pour contrôler les sorties de tableau.
   - Cliquez sur **Enregistrer** pour enregistrer les statistiques estimées dans le jeu de données actif.
   - Cliquer sur **Tracé** pour spécifier les tracés à inclure dans la sortie.
8. Cliquez sur **OK**.

Méta-analyse : Taille d'effet binaire : Critères

La boîte de dialogue **Critères** propose des paramètres permettant de spécifier des critères pour une méta-analyse avec des résultats binaires lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Intervalle de confiance
Ce paramètre facultatif spécifie le niveau de confiance. Il doit s'agir d'une valeur numérique comprise entre 0 et 100. La valeur par défaut est 95.

La portée des données est manquante
Les paramètres facultatifs contrôlent la façon dont la procédure traite les données manquantes.

Exclure les observations analyse par analyse
Le paramètre par défaut inclut toutes les observations ayant suffisamment de données sur les variables utilisées dans chaque analyse particulière.

Exclure toute observation incomplète
Inclut toutes les observations ayant suffisamment de données sur toutes les variables utilisées dans toutes les analyses spécifiées par la procédure.

Valeurs utilisateur manquantes
Les paramètres facultatifs contrôlent la façon dont les valeurs utilisateur manquantes sont traitées.

Exclure
Le paramètre par défaut considère les valeurs utilisateur manquantes comme étant valides.

Inclure
Ignore les désignations de valeurs utilisateur manquantes et considère ces dernières comme étant valides.

Itérations

Nombre maximum d'itérations
Ce paramètre facultatif spécifie le nombre maximal d'itérations dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 100. La valeur 0 indique qu'aucune itération n'est effectuée.

Découpage maximal d'étape en deux
Ce paramètre facultatif spécifie le nombre maximal de dichotomies dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 5. Une valeur de 0 signifie que le step-halving n'est pas appliqué.

Convergence
Ce paramètre facultatif spécifie la tolérance de convergence. La valeur doit correspondre à une valeur positive unique. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur par défaut est 1E-6.

Définition des critères de taille d'effet binaire pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet binaire**, cliquez sur **Critères**.
3. Sélectionnez et définissez les paramètres de critères appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet binaire : Analyse

La boîte de dialogue **Analyse** propose des paramètres permettant de spécifier des analyses cumulées et de sous-groupe pour une méta-analyse avec des résultats binaires lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Analyse de sous-groupe
Sélectionnez une variable évoquant l'analyse de sous-groupe. La variable ne peut pas être identique à la variable définie pour **Analyse cumulée**.

Analyse cumulée
Sélectionnez une variable évoquant l'analyse cumulée et en fonction de laquelle la méta-analyse cumulée est effectuée. La variable ne peut pas être identique à la variable définie pour **Analyse de sous-groupe**. Lorsque l'option **Croissant** est sélectionnée, l'analyse cumulée est basée sur la variable spécifiée dans l'ordre croissant. Lorsque l'option **Décroissant** est sélectionnée, l'analyse cumulée est basée sur la variable spécifiée dans l'ordre décroissant.

Statistiques cumulées
Propose des options permettant d'enregistrer la taille d'effet globale cumulée estimée. Les paramètres ne sont disponibles que lorsqu'une variable **Analyse cumulée** est sélectionnée.

Taille d'effet cumulée
Enregistre la taille d'effet globale cumulée estimée.

Taille d'effet cumulée (sous forme d'exposant)
Enregistre la taille d'effet globale cumulée estimée sous forme d'exposant.

Erreur standard
Enregistre l'erreur standard estimée de la taille d'effet globale cumulée.

Limite inférieure de l'intervalle de confiance
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée.

Limite supérieure de l'intervalle de confiance
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée.

Limite inférieure de l'intervalle de confiance (sous forme d'exposant)
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée sous forme d'exposant.

Limite supérieure de l'intervalle de confiance (sous forme d'exposant)
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet globale cumulée sous forme d'exposant.

Valeur P
Enregistre la valeur *p* estimée de la taille d'effet globale cumulée.

Destination
Fournit des options permettant de spécifier un fichier ou jeu de données de sauvegarde. Lorsque **Jeu de données** est sélectionné, vous pouvez spécifier un nouveau nom de jeu de données (vous pouvez conserver le nom de jeu de données par défaut). Lorsque le **fichier de données** est sélectionné, cliquez sur **Parcourir...** pour sélectionner le nom et l'emplacement du fichier d'enregistrement.

Définition des paramètres d'analyse de taille d'effet binaire pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet binaire**, cliquez sur **Analyse**.
3. Sélectionnez et définissez les paramètres d'analyse appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet binaire : Inférence

La boîte de dialogue **Inférence** propose des paramètres permettant de spécifier les méthodes d'estimation pour une méta-analyse avec des résultats binaires lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Estimateur
Fournit des paramètres permettant de spécifier l'estimateur.

Remarques :

- Lorsque les paramètres Tronquer et remplir sont spécifiés, ce paramètre contrôle également l'estimateur qui est utilisé par le regroupement dans l'analyse de troncature et de remplissage.
- Lorsque les paramètres de biais sont spécifiés, ce paramètre contrôle également l'estimateur qui est utilisé par le test de régression.

Maximum de vraisemblance restreint (REML)
Le paramètre par défaut applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance restreint.

Maximum de vraisemblance (ML)
Applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance.

Bayes empirique
Applique la méthode itérative et calcule l'estimateur Bayes empirique.

Hedges
Applique la méthode non itérative et calcule l'estimateur Hedges.

Hunter-Schmidt
Applique la méthode non itérative et calcule l'estimateur Hunter-Schmidt.

DerSimonian-Laird
Applique la méthode non itérative et calcule l'estimateur DerSimonian-Laird.

Sidik-Jonkman
Applique la méthode non itérative et calcule l'estimateur Sidik-Jonkman.

Ajustement de l'erreur standard
Fournit des paramètres qui déterminent si l'ajustement d'erreur standard Knapp-Hartung doit être appliqué ou non.

Remarques :

- Lorsque les paramètres Tronquer et remplir sont spécifiés, ce paramètre contrôle également l'ajustement d'erreur standard qui est utilisé par le regroupement dans l'analyse de troncature et de remplissage.
- Lorsque les paramètres de biais sont spécifiés, ce paramètre contrôle également l'ajustement d'erreur standard qui est utilisé par le test de régression.

Aucun ajustement
Le paramètre par défaut n'applique pas l'ajustement.

Appliquer l'ajustement de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung.

Appliquer l'ajustement tronqué de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung et tronque la valeur si elle est inférieure à 1 lors de l'estimation de la matrice variance-covariance.

Définition des paramètres d'inférence de taille d'effet binaire pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet binaire**, cliquez sur **Inférence**.
3. Sélectionnez et définissez les paramètres d'inférence appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet binaire : Contraste

La boîte de dialogue **Contraste** propose des paramètres permettant de contrôler le test de contraste pour une méta-analyse avec des résultats binaires lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Variables
La liste affiche toutes les variables de jeu de données disponibles. Sélectionnez des variables dans la liste et déplacez-les vers la liste **Contraste(s)**.

Contraste(s)
La liste identifie les coefficients qui sont stockés sous forme de variables dans le jeu de données actif. Plusieurs variables sont autorisées. Les variables de chaîne ne sont pas prises en charge.

Valeurs de coefficient entrées par l'utilisateur
Fournit des paramètres permettant de spécifier les coefficients d'un contraste fourni par l'utilisateur. Seules les valeurs numériques sont admises. Pour formuler un test de contraste valide, le nombre de valeurs spécifiées doit correspondre à celui des études valides.

Afficher les statistiques mises en exposant
Contrôle l'inclusion des statistiques sous forme d'exposant. Lorsque ce paramètre est sélectionné, la sortie inclut des statistiques sous forme d'exposant, y compris la taille d'effet et les limites de l'intervalle de confiance sous forme d'exposant. Le paramètre est disponible lorsque l'option **Taille d'effet** est spécifiée comme **Rapport des cotes de log**, **Rapport des cotes de log de Peto** ou **Rapport des risques de log**.

Définition des paramètres de contraste de taille d'effet binaire pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet binaire**, cliquez sur **Contraste**.
3. Sélectionnez et définissez les paramètres de contraste appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet binaire : Biais

La boîte de dialogue **Biais** propose des paramètres permettant d'activer le biais de publication en effectuant des tests de régression pour une méta-analyse avec des résultats binaires lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Test de régression d'Egger
Sélectionnez ce paramètre pour activer le biais de publication en effectuant le test de régression d'Egger.

Variables
La liste fournit toutes les variables de jeu de données disponibles.

Covariable(s)
Les variables sélectionnées dans la liste **Variables** sont traitées sous forme de covariables. Plusieurs covariables sont autorisées.

Facteur(s)
Les variables sélectionnées dans la liste **Variables** sont traitées sous forme de facteurs. Plusieurs facteurs sont autorisés.

Inclure une constante dans la régression
Contrôle le terme de la constante dans le test de régression.

Inclure un paramètre de dispersion dans le modèle à effets fixes
Contrôle le paramètre de modèle multiplicatif et introduit le paramètre de dispersion multiplicative dans l'analyse. Ce paramètre n'est disponible que lorsqu'un modèle à effets fixes est sélectionné.

Estimer les statistiques sur la base de la distribution t
Contrôle la distribution utilisée dans les tests de régression. Ce paramètre est activé par défaut, et permet d'estimer les statistiques en fonction de la distribution *t*. Lorsqu'il n'est pas sélectionné, les statistiques sont estimées en fonction de la distribution normale.

Définition des paramètres de biais de taille d'effet binaire pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet binaire**, cliquez sur **Biais**.
3. Sélectionnez et définissez les paramètres de biais appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet binaire : Tronquer et remplir

La boîte de dialogue **Tronquer et remplir** propose des paramètres permettant de mettre en oeuvre l'analyse de troncature et de remplissage du biais de publication pour une méta-analyse avec des résultats binaires lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Nombre d'études manquantes estimé
Contrôle l'analyse de troncature et de remplissage du biais de publication. En sélectionnant ce paramètre, vous activez les autres paramètres de boîte de dialogue.

Côté où imputer les études
Propose des options permettant de spécifier le côté du graphique en entonnoir sur lequel les études manquantes sont imputées.

Déterminé par les pentes du test d'Egger
Le paramètre par défaut détermine le côté en fonction de la pente estimée du test d'Egger.

Gauche
Impute le côté gauche du graphique en entonnoir.

Droite
Impute le côté droit du graphique en entonnoir.

Méthode
Indique la méthode permettant d'estimer le nombre d'études manquantes.

Linéaire
Le paramètre par défaut calcule l'estimateur linéaire.

Exécuter
Calcule l'estimateur d'exécution.

Quadratique
Calcule l'estimateur quadratique.

Processus d'itération
Fournit des paramètres permettant de spécifier l'estimateur d'itération et l'ajustement d'erreur standard.

Modèle à effets fixes
Si ce paramètre est sélectionné, un modèle à effets fixes est utilisé et les options d'estimation d'itération et d'ajustement d'erreur standard ne sont pas disponibles.

Modèle à effets aléatoires
Si ce paramètre est sélectionné, un modèle à effets aléatoires est utilisé et les paramètres suivants sont disponibles.

Estimateur
Fournit des paramètres permettant de spécifier l'estimateur d'itération.

Maximum de vraisemblance restreint (REML)
Le paramètre par défaut applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance restreint.

Maximum de vraisemblance (ML)
Applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance.

Bayes empirique
Applique la méthode itérative et calcule l'estimateur Bayes empirique.

Hedges
Applique la méthode non itérative et calcule l'estimateur Hedges.

Hunter-Schmidt
Applique la méthode non itérative et calcule l'estimateur Hunter-Schmidt.

DerSimonian-Laird
Applique la méthode non itérative et calcule l'estimateur DerSimonian-Laird.

Sidik-Jonkman
Applique la méthode non itérative et calcule l'estimateur Sidik-Jonkman.

Ajustement de l'erreur standard
Fournit des paramètres qui déterminent si l'ajustement d'erreur standard Knapp-Hartung doit être appliqué ou non aux itérations de l'algorithme de troncature et remplissage.

Aucun ajustement
Le paramètre par défaut n'applique pas l'ajustement.

Appliquer l'ajustement de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung.

Appliquer l'ajustement tronqué de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung et tronque la valeur si elle est inférieure à 1 lors de l'estimation de la matrice variance-covariance.

Définition des paramètres de troncature et remplissage de taille d'effet binaire pour la méta-analyse

1. A partir des menus, sélectionnez :

Analyse > **Méta-analyse** > **Résultats binaires** > **Taille d'effet précalculée...**

2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet binaire**, cliquez sur **Tronquer et remplir**.
3. Sélectionnez et définissez les paramètres de troncature et de remplissage appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet binaire : Imprimer

La boîte de dialogue **Imprimer** propose des paramètres permettant de contrôler les sorties de tableau pour une méta-analyse avec des résultats binaires lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Homogénéité/hétérogénéité
Fournit des paramètres permettant de contrôler les tests d'homogénéité et d'hétérogénéité.

Test d'homogénéité
Lorsque ce paramètre est sélectionné, le test d'homogénéité correspondant est fourni dans la sortie.

Mesures d'hétérogénéité
Lorsque ce paramètre est sélectionné, les mesures d'hétérogénéité sont fournies dans la sortie.

Tailles d'effet
Fournit les paramètres de taille d'effet ci-dessous.

Etudes individuelles
Contrôle l'affichage des études individuelles. Si ce paramètre est sélectionné, la sortie correspondante est fournie.

Tailles d'effet cumulées
Contrôle l'affichage de l'analyse cumulée. Si ce paramètre est sélectionné, la sortie correspondante est fournie. Ce paramètre n'est disponible que lorsqu'une variable **Analyse cumulée** est sélectionnée dans la boîte de dialogue **Analyse**.

Intervalle de prévision dans un modèle à effets aléatoires
Contrôle l'affichage de l'intervalle de prévision. Si ce paramètre est sélectionné, la sortie correspondante est fournie. Ce paramètre n'est disponible que lorsqu'un modèle à effets aléatoires est sélectionné.

Afficher les statistiques mises en exposant
Contrôle l'inclusion des statistiques sous forme d'exposant. Lorsque ce paramètre est sélectionné, les statistiques sous forme d'exposant, y compris la taille d'effet et les limites de l'intervalle de confiance sous forme d'exposant sont générées. Le processus s'applique aux tables d'estimation et de prévision de taille d'effet obligatoires et facultatives. Le paramètre est disponible lorsque l'option **Taille d'effet** est spécifiée comme **Rapport des cotes de log**, **Rapport des cotes de log de Peto** ou **Rapport des risques de log**.

Définition des paramètres d'impression de taille d'effet binaire pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet binaire**, cliquez sur **Imprimer**.
3. Sélectionnez et définissez les paramètres d'impression appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet binaire : Enregistrer

La boîte de dialogue **Enregistrer** propose des paramètres permettant d'enregistrer les statistiques estimées dans le jeu de données actif pour une méta-analyse avec des résultats binaires lorsque les données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Etudes individuelles
Fournit des options permettant d'enregistrer la taille d'effet estimée.

Taille d'effet individuelle (sous forme d'exposant)
Enregistre la taille d'effet estimée sous forme d'exposant.

Erreur standard
Enregistre l'erreur standard estimée de la taille d'effet.

Limite inférieure de l'intervalle de confiance
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet.

Limite supérieure de l'intervalle de confiance
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet.

Limite inférieure de l'intervalle de confiance (sous forme d'exposant)
Enregistre la limite inférieure de l'intervalle de confiance estimé de la taille d'effet sous forme d'exposant.

Limite supérieure de l'intervalle de confiance (sous forme d'exposant)
Enregistre la limite supérieure de l'intervalle de confiance estimé de la taille d'effet sous forme d'exposant.

Valeur P
Enregistre la valeur *p* estimée de la taille d'effet.

Pondération de l'étude
Enregistre la pondération estimée de l'étude.

Pourcentage de la pondération de l'étude
Enregistre la pondération d'étude normalisée sous forme de pourcentage.

Définition des paramètres d'enregistrement de taille d'effet binaire pour la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Résultats binaires** > **Taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse : Taille d'effet binaire**, cliquez sur **Enregistrer**.
3. Sélectionnez et définissez les paramètres d'enregistrement de statistiques estimées appropriés.
4. Cliquez sur **Continuer**

Méta-analyse : Taille d'effet binaire : Tracé

La boîte de dialogue **Tracé** propose des paramètres des types de tracé suivants :

- «Graphique en forêt», à la page 80
- «Graphique en forêt cumulée», à la page 81
- «Graphique à bulles», à la page 82
- «Graphique en entonnoir», à la page 83
- «Graphique de Galbraith», à la page 84
- «Graphique de l'Abbé», à la page 85

Régression de la méta-analyse

La procédure de régression de la méta-analyse effectue une analyse de méta-régression.

Exemple
Plusieurs travaux de recherche ont été réalisés tout au long de l'histoire pour étudier un médicament à la mode mais discutable pour traiter le diabète de type II. Le médicament oral était censé réduire le taux de glucose dans le sang après les repas. Des données ont été recueillies sur différents sites de recherche de 1979 à 1986.

Une chercheuse principale souhaite établir des statistiques sur l'effet du médicament oral. Etant donné que les données ont été générées à partir d'études différentes, elle a proposé l'idée de synthétiser les résultats dans toutes les études afin de comprendre globalement l'effet et d'identifier les sources sous-jacentes de variation dans les résultats.

Statistiques
Niveau de confiance, méthode itérative, dichotomie, tolérance de convergence, moyennes d'échantillons, variance d'échantillon, écart type, taille d'effet estimée, méthode d'estimation, test de régression, modèle à effets aléatoires, modèle à effets fixes, paramètre de dispersion, estimateur de vraisemblance maximale restreinte, estimateur Bayes empirique, estimateur Hedges, estimateur Hunter-Schmidt, estimateur DerSimonian-Laird, estimateur Sidik-Jonkman, ajustement d'erreur standard Knapp-Hartung, ajustement d'erreur standard Knapp-Hartung tronqué, coefficients, test de coefficient de modèle, statistiques sous forme d'exposant.

Obtention d'une analyse de régression de méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Méta-régression**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une seule variable **Taille d'effet** dépendante qui indique la taille d'un effet. La variable sélectionnée doit être numérique (les variables de chaîne ne sont pas prises en charge).
3. Sélectionnez l'un des paramètres suivants, puis sélectionnez une seule variable numérique correspondante :

   Erreur standard
   Sélectionnez une variable qui spécifie l'erreur standard convertie en pondération. Il s'agit du paramètre par défaut.

   Ecart
   Sélectionnez une variable qui spécifie la variance convertie en pondération.

   Pondération
   Sélectionnez une variable qui spécifie la pondération.
4. Vous avez également la possibilité d'ajouter des variables de facteur dans la liste **Facteur(s)**. Pour chaque variable de facteur sélectionnée, une valeur facultative peut être spécifiée pour chaque variable afin de désigner des **Dernières catégories** personnalisées. Vous pouvez cliquer sur **Réinitialiser** pour restaurer les valeurs par défaut des **Dernières catégories**.

   Remarque : Si aucune observation ne correspond aux valeurs des **Dernières catégories** spécifiées, les dernières valeurs obtenues sont traitées comme les dernières catégories.
5. (Facultatif) Sélectionnez des covariables numériques.
6. (Facultatif) Sélectionnez un paramètre **Modèle**.

   Effets aléatoires
   La valeur par défaut permet de générer le modèle à effets aléatoires.

   Effets fixes
   Génère le modèle à effets fixes. Vous avez également la possibilité de sélectionner le paramètre **Inclure un paramètre de dispersion**.
7. (En option) Vous avez également ces possibilités :
   - Cliquez sur **Critères...** pour spécifier les critères généraux.
   - Cliquer sur **Inférence** pour spécifier les méthodes d'estimation.
   - Cliquez sur **Imprimer** pour contrôler les sorties de tableau.
   - Cliquez sur **Enregistrer** pour enregistrer les statistiques estimées dans le jeu de données actif.
   - Cliquer sur **Tracé** pour spécifier les tracés à inclure dans la sortie.
8. Cliquez sur **OK**.

Régression de la méta-analyse : Critères

La boîte de dialogue **Critères** propose des paramètres permettant de spécifier des critères pour la régression de la méta-analyse.

Intervalle de confiance
Ce paramètre facultatif spécifie le niveau de confiance. Il doit s'agir d'une valeur numérique comprise entre 0 et 100. La valeur par défaut est 95.

Valeurs utilisateur manquantes
Les paramètres facultatifs contrôlent la façon dont les valeurs utilisateur manquantes sont traitées.

Exclure
Le paramètre par défaut considère les valeurs utilisateur manquantes comme étant valides.

Inclure
Ignore les désignations de valeurs utilisateur manquantes et considère ces dernières comme étant valides.

Itérations

Nombre maximum d'itérations
Ce paramètre facultatif spécifie le nombre maximal d'itérations dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 100. La valeur 0 indique qu'aucune itération n'est effectuée.

Découpage maximal d'étape en deux
Ce paramètre facultatif spécifie le nombre maximal de dichotomies dans les méthodes itératives. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur doit correspondre à un entier positif unique. La valeur par défaut est 5. Une valeur de 0 signifie que le step-halving n'est pas appliqué.

Convergence
Ce paramètre facultatif spécifie la tolérance de convergence. La valeur doit correspondre à une valeur positive unique. Ce paramètre est disponible lorsque des méthodes itératives sont utilisées. La valeur par défaut est 1E-6.

Définition des critères de régression de la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Méta-régression**
2. Dans la boîte de dialogue **Régression de la méta-analyse**, cliquez sur **Critères**.
3. Sélectionnez et définissez les paramètres de critères appropriés.
4. Cliquez sur **Continuer**

Régression de la méta-analyse : Inférence

La boîte de dialogue **Inférence** propose des paramètres permettant de spécifier les méthodes d'estimation pour la régression de la méta-analyse.

Inclure une constante dans la régression
Ce paramètre est activé par défaut.

Estimer les statistiques sur la base de la distribution t
Contrôle la distribution utilisée dans les tests de régression. Ce paramètre est activé par défaut, et permet d'estimer les statistiques en fonction de la distribution *t*. Lorsqu'il n'est pas sélectionné, les statistiques sont estimées en fonction de la distribution normale.

Estimateur
Fournit des paramètres permettant de spécifier l'estimateur.

Maximum de vraisemblance restreint (REML)
Le paramètre par défaut applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance restreint.

Maximum de vraisemblance (ML)
Applique la méthode itérative et calcule l'estimateur du maximum de vraisemblance.

Bayes empirique
Applique la méthode itérative et calcule l'estimateur Bayes empirique.

Hedges
Applique la méthode non itérative et calcule l'estimateur Hedges.

Hunter-Schmidt
Applique la méthode non itérative et calcule l'estimateur Hunter-Schmidt.

DerSimonian-Laird
Applique la méthode non itérative et calcule l'estimateur DerSimonian-Laird.

Sidik-Jonkman
Applique la méthode non itérative et calcule l'estimateur Sidik-Jonkman.

Ajustement de l'erreur standard
Fournit des paramètres qui déterminent si l'ajustement d'erreur standard Knapp-Hartung doit être appliqué ou non.

Aucun ajustement
Le paramètre par défaut n'applique pas l'ajustement.

Appliquer l'ajustement de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung.

Appliquer l'ajustement tronqué de Knapp-Hartung
Applique la méthode d'ajustement Knapp-Hartung et tronque la valeur si elle est inférieure à 1 lors de l'estimation de la matrice variance-covariance.

Définition des paramètres d'inférence de la régression de la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Méta-régression**
2. Dans la boîte de dialogue **Régression de la méta-analyse**, cliquez sur **Inférence**.
3. Sélectionnez et définissez les paramètres d'inférence appropriés.
4. Cliquez sur **Continuer**

Régression de la méta-analyse : Imprimer

La boîte de dialogue **Imprimer** contient des paramètres qui contrôlent les sorties de tableau de la régression de la méta-analyse.

Test du coefficient de modèle
Contrôle le test du coefficient de modèle. Ce paramètre n'est pas spécifié par défaut, ce qui supprime le test. Lorsque le paramètre est spécifié, le test est fourni dans la sortie.

Afficher les statistiques mises en exposant
Contrôle les estimations de paramètres. Ce paramètre n'est pas spécifié par défaut, ce qui supprime les estimations de paramètres. Lorsque le paramètre est spécifié, les estimations de paramètres sont fournies dans la sortie.

Définition des paramètres d'impression de la régression de la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Méta-régression**
2. Dans la boîte de dialogue **Régression de la méta-analyse**, cliquez sur **Imprimer**.

3. Sélectionnez et définissez les paramètres d'impression appropriés.
4. Cliquez sur **Continuer**

Régression de la méta-analyse : Enregistrer

La boîte de dialogue **Enregistrer** fournit des paramètres permettant d'enregistrer les statistiques estimées dans le jeu de données actif pour la régression de la méta-analyse.

Etudes individuelles
Fournit des options permettant d'enregistrer la taille d'effet estimée.

Valeurs prédites
Prédit et enregistre les valeurs ajustées.

Erreur standard de la prévision
Prédit et enregistre l'erreur standard des valeurs ajustées.

Limite inférieure de l'intervalle de confiance
Enregistre la limite inférieure de l'intervalle de confiance estimé des valeurs ajustées.

Limite supérieure de l'intervalle de confiance
Enregistre la limite supérieure de l'intervalle de confiance estimé des valeurs ajustées.

Résidus
Prédit et enregistre les résidus.

Erreur standard des résiduels
Prédit et enregistre l'erreur standard des résidus.

Bras de levier
Prédit et enregistre les bras de levier.

Statistiques cumulées
Fournit des options permettant d'enregistrer les prévisions d'options linéaires.

Remarque : Les paramètres de statistiques cumulées sont ignorés lorsque l'option **Effets fixes** est sélectionnée comme paramètre du **Modèle**.

Prédictions linéaires fixes
Prédit et enregistre la prédiction linéaire.

Erreur standard des prédictions linéaires fixes
Prédit et enregistre l'erreur standard de la prédiction linéaire.

Meilleures prédictions linéaires non biaisées
Prédit et enregistre la meilleure prédiction linéaire non biaisée des effets aléatoires.

Erreur standard des meilleures prédictions linéaires non biaisées
Prédit et enregistre l'erreur standard de la meilleure prédiction linéaire non biaisée des effets aléatoires.

Définition des paramètres d'enregistrement de la régression de la méta-analyse

1. A partir des menus, sélectionnez :

   Analyse > **Méta-analyse** > **Méta-régression**
2. Dans la boîte de dialogue **Régression de la méta-analyse**, cliquez sur **Enregistrer**.
3. Sélectionnez et définissez les paramètres d'enregistrement appropriés.
4. Cliquez sur **Continuer**

Régression de la méta-analyse : Tracé

La boîte de dialogue **Tracé** fournit des paramètres pour la définition des graphiques à bulles. Pour plus d'informations, voir «Graphique à bulles», à la page 82.

Options de traçage de la méta-analyse

Les options de traçage suivantes sont disponibles pour les procédures de méta-analyse et sont accessibles dans les boîtes de dialogue **Tracé**.

Graphique en forêt

L'onglet **Graphique en forêt** propose des paramètres permettant de contrôler les graphiques en forêt qui s'affichent dans la sortie pour une méta-analyse avec des résultats continus et binaires lorsque des données brutes et des données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Graphique en forêt
Ce paramètre facultatif active et désactive la sortie de graphique en forêt. Il est désactivé par défaut.

Colonnes d'affichage
Les paramètres facultatifs spécifient les statistiques estimées qui sont mises en tableau.

Taille d'effet
Inclut la taille d'effet estimée.

Erreur standard
Inclut l'erreur standard de la taille d'effet estimée.

Limites de l'intervalle de confiance
Inclut les limites supérieure et inférieure de l'intervalle de confiance de la taille d'effet estimée.

Valeur P
Inclut la valeur *p* de la taille d'effet estimée pour chaque étude individuelle.

Pondération
Inclut la pondération des études primaires.

Afficher sous forme d'exposant

Remarque : Ce paramètre ne s'applique qu'à la méta-analyse avec des résultats binaires.

Contrôle la colonne de graphique et l'affichage de la taille d'effet et de l'intervalle de confiance. Lorsque ce paramètre est sélectionné, il trace le rapport des cotes, le rapport des cotes de Peto ou le rapport des risques, puis met en tableau les statistiques sous forme d'exposant. Si ce paramètre n'est pas sélectionné, le rapport des cotes de log, le rapport des cotes de log de Peto ou le rapport des risques de log sont tracés et les statistiques transformées log sont mises en tableau.

Variables
Répertorie les variables de jeux de données disponibles.

Colonnes additionnelles
Vous avez la possibilité de sélectionner des variables à afficher sous forme de colonnes additionnelles. L'ordre des variables détermine l'ordre d'affichage des colonnes.

Tri
Vous avez la possibilité de spécifier une variable selon laquelle le graphique en forêt est trié. Lorsqu'une variable est spécifiée, le graphique en forêt est trié dans l'ordre **Croissant** (valeur par défaut). Si l'option **Décroissant** est sélectionnée, le graphique en forêt est trié dans l'ordre décroissant.

Position de la colonne du graphique
Contrôle le positionnement de la colonne du graphique.

Droite
Le paramètre par défaut place la colonne du graphique à droite des autres colonnes du tableau.

Gauche
Place la colonne du graphique à gauche des autres colonnes du tableau.

Lignes de référence
Contrôle les lignes de référence qui sont ajoutées au graphique en forêt. **Taille d'effet globale** ajoute la ligne pour indiquer la taille d'effet globale estimée. **Taille d'effet nulle** ajoute la ligne pour indiquer la taille d'effet nulle.

Annotations
Contrôle l'affichage des annotations.

Homogénéité
Imprime les statistiques du test d'homogénéité.

Hétérogénéité
Imprime les statistiques du test d'hétérogénéité.

Tester
Imprime le test de la taille d'effet globale et le test d'homogénéité entre les sous-groupes en présence de l'analyse de sous-groupe.

Plage de rognage
Spécifie la plage de rognage. Lorsque le réglage est sélectionné, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound`.

Pour les méta-analyses avec des résultats binaires, lorsque le paramètre **Taille de l'effet** est défini sur **Différence de risque**, il doit également satisfaire à `Upper bound ≤ 1` et `Lower bound ≥ −1`. Lorsque le paramètre **Afficher la forme exponentielle** est sélectionné, il doit également satisfaire à `Lower bound ≥ 0`.

Définition des paramètres de graphique en forêt

1. A partir des menus, sélectionnez :

   Analyser > **Méta-analyse** > **Résultats continus ou résultats binaires** > **Données brutes...** ou **taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse**, cliquez sur **Tracer** puis sur l'onglet **Graphique en forêt**.
3. Sélectionnez et définissez les paramètres de graphique en forêt appropriés.
4. Cliquez sur **Continuer**

Graphique en forêt cumulée

L'onglet **Graphique en forêt cumulée** propose des paramètres permettant de contrôler les graphiques en forêt cumulée qui s'affichent dans la sortie pour une méta-analyse avec des résultats continus et binaires lorsque des données brutes et des données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Graphique en forêt cumulée
Ce paramètre facultatif active et désactive la sortie de graphique en forêt. Il est désactivé par défaut.

Colonnes d'affichage
Les paramètres facultatifs spécifient les statistiques estimées qui sont mises en tableau.

Taille d'effet cumulée
Inclut la taille d'effet globale cumulée.

Erreur standard
Inclut l'erreur standard de la taille d'effet globale cumulée.

Limites de l'intervalle de confiance
Inclut les limites supérieure et inférieure de l'intervalle de confiance de la taille d'effet globale cumulée.

Valeur P
Inclut la valeur *p* de la taille d'effet globale cumulée pour chaque étude individuelle.

Afficher sous forme d'exposant

Remarque : Ce paramètre ne s'applique qu'à la méta-analyse avec des résultats binaires.

Contrôle la colonne de graphique et l'affichage de la taille d'effet et de l'intervalle de confiance. Lorsque ce paramètre est sélectionné, il trace le rapport des cotes, le rapport des cotes de Peto ou le rapport des risques, puis met en tableau les statistiques sous forme d'exposant. Si ce paramètre n'est pas activé, le rapport des cotes de log, le rapport des cotes de log de Peto ou le rapport des risques de log sont tracés et les statistiques transformées log sont mises en tableau.

Variables
Répertorie les variables de jeux de données disponibles.

Colonnes additionnelles
Vous avez la possibilité de sélectionner des variables à afficher sous forme de colonnes additionnelles. L'ordre des variables détermine l'ordre d'affichage des colonnes.

Position de la colonne du graphique
Contrôle le positionnement de la colonne du graphique.

Droite
Le paramètre par défaut place la colonne du graphique à droite des autres colonnes du tableau.

Gauche
Place la colonne du graphique à gauche des autres colonnes du tableau.

Plage de rognage
Spécifie la plage de rognage. Lorsque le réglage est sélectionné, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound`.

Pour les méta-analyses avec des résultats binaires, lorsque le paramètre **Taille de l'effet** est défini sur **Différence de risque**, il doit également satisfaire à `Upper bound ≤ 1` et `Lower bound ≥ −1`. Lorsque le paramètre **Afficher la forme exponentielle** est sélectionné, il doit également satisfaire à `Lower bound ≥ 0`.

Définition des paramètres de graphique en forêt cumulée

1. A partir des menus, sélectionnez :

   Analyser > **Méta-analyse** > **Résultats continus ou résultats binaires** > **Données brutes...** ou **taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse**, cliquez sur **Tracer** puis sur l'onglet **Graphique en forêt cumulée**.
3. Sélectionnez et définissez les paramètres de graphique en forêt cumulée appropriés.
4. Cliquez sur **Continuer**

Graphique à bulles

L'onglet **Graphique à bulles** propose des paramètres permettant de contrôler les graphiques à bulles qui s'affichent dans la sortie pour une méta-analyse avec des résultats continus et binaires lorsque des données brutes et des données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Remarque : Le graphique à bulles est le seul type de graphique disponible pour l'analyse de régression de la méta-analyse.

Graphique à bulles
Ce paramètre facultatif active et désactive la sortie de graphique à bulles. Il est désactivé par défaut.

Remarque : Ce paramètre n'est pas disponible pour l'analyse de régression de la méta-analyse.

Variables
Répertorie les variables de jeux de données disponibles.

Prédicteur(s)
Spécifiez des variables traitées comme des prédicteurs continus sur l'axe des X. Plusieurs variables sont admises, chacune créant un graphique à bulles distinct.

Remarque : Pour l'analyse de régression de la méta-analyse, ce paramètre est intitulé **Prédicteur(s) de graphique à bulles** et permet de sélectionner des covariables pour lesquelles créer des graphiques à bulles.

Prédicteur(s) centré(s) sur la moyenne
Ce paramètre facultatif contrôle le prédicteur continu sur l'axe des X. Il est désactivé par défaut, ce qui ne permet pas le centrage du prédicteur. Lorsque ce paramètre est activé, le prédicteur est centré sur la moyenne.

Libellé
Vous avez également la possibilité de spécifier une variable sur laquelle le graphique à bulles est libellé. La liste **Placement** fournit des options permettant de placer les étiquettes automatiquement (**Auto**) ou à **droite**, à **gauche**, dans la partie **supérieure** et en **bas**.

Dessiner des bulles proportionnelles aux pondérations
Contrôle la façon dont les bulles sont tracées par rapport aux pondérations. Ce paramètre est activé par défaut, ce qui permet de tracer les bulles proportionnellement aux pondérations. Lorsque ce paramètre est désactivé, toutes les bulles sont tracées à la même taille.

Afficher la ligne ajustée
Contrôle l'affichage de la courbe de régression ajustée. Ce paramètre est activé par défaut, ce qui permet d'ajouter la courbe de régression ajustée dans le graphique à bulles. Lorsque ce paramètre est désactivé, la courbe de régression ajustée n'est pas ajoutée dans le graphique à bulles.

Afficher les limites de l'intervalle de confiance
Contrôle l'affichage des limites de l'intervalle de confiance. Ce paramètre est activé par défaut, ce qui permet d'ajouter les limites de l'intervalle de confiance dans le graphique à bulles. Lorsque ce paramètre est désactivé, les limites de l'intervalle de confiance ne s'ajoutent pas dans le graphique à bulles.

Plage de l'axe des X
Spécifie la courbe de l'axe des X. Lorsque le réglage est activé, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound`. Il est à noter que les valeurs spécifiées s'appliquent à tous les graphiques à bulles.

Plage de l'axe des Y
Spécifie la courbe de l'axe des Y. Lorsque le réglage est activé, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound`. Il est à noter que les valeurs spécifiées s'appliquent à tous les graphiques à bulles.

Pour les méta-analyses avec des résultats binaires, lorsque le paramètre **Taille de l'effet** est défini sur **Différence de risque**, il doit également satisfaire à `Upper bound ≤ 1` et `Lower bound ≥ −1`.

Définition des paramètres de graphique à bulles

1. A partir des menus, sélectionnez :

   Analyser > **Méta-analyse** > **Résultats continus ou résultats binaires** > **Données brutes...** ou **taille d'effet précalculée...**

   ou

   Analyse > **Méta-analyse** > **Méta-régression**
2. Dans la boîte de dialogue **Méta-analyse**, cliquez sur **Tracer** puis sur l'onglet **Graphique à bulles**.
3. Sélectionnez et définissez les paramètres de graphique à bulles appropriés.
4. Cliquez sur **Continuer**

Graphique en entonnoir

L'onglet **Graphique en entonnoir** propose des paramètres permettant de contrôler les graphiques en entonnoir qui s'affichent dans la sortie pour une méta-analyse avec des résultats continus et binaires lorsque des données brutes et des données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Graphique en entonnoir
Ce paramètre facultatif active et désactive la sortie de graphique en entonnoir. Il est désactivé par défaut.

Valeurs de l'axe des Y
Contrôle les valeurs de l'axe des Y.

Erreur standard
Trace l'erreur standard.

Erreur standard inverse
Trace l'erreur standard inverse.

Ecart
Trace la variance.

Variance inverse
Trace la variance inverse.

Plage de l'axe des X
Spécifie la courbe de l'axe des X. Lorsque le réglage est activé, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound`. Il est à noter que les valeurs spécifiées s'appliquent à tous les graphiques en entonnoir.

Pour les méta-analyses avec des résultats binaires, lorsque le paramètre **Taille de l'effet** est défini sur **Différence de risque**, il doit également satisfaire à `Upper bound ≤ 1` et `Lower bound ≥ −1`.

Plage de l'axe des Y
Spécifie la courbe de l'axe des Y. Lorsque le réglage est activé, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound ≥ 0`. Il est à noter que les valeurs spécifiées s'appliquent à tous les graphiques en entonnoir.

Inclure les études imputées avec Tronquer et remplir
Contrôle si les études imputées sont incluses dans le graphique en entonnoir lorsque le paramètre **Nombre d'études manquantes estimé** est activé dans la boîte de dialogue **Tronquer et remplir**. Le paramètre est désactivé par défaut, ce qui exclut les études imputées. Lorsque l'option **Afficher la taille d'effet globale en fonction des études observées** est sélectionnée, une ligne de référence verticale indiquant la taille d'effet globale estimée est ajoutée. Ce paramètre tient compte des études observées et imputées.

Variables
Répertorie les variables de jeux de données disponibles.

Libellé
Vous avez également la possibilité de spécifier une variable sur laquelle le graphique en entonnoir est libellé. La liste **Placement** fournit des options permettant de placer les étiquettes automatiquement (**Auto**) ou à **droite**, à **gauche**, dans la partie **supérieure** et en **bas**.

Définition des paramètres de graphique en entonnoir

1. A partir des menus, sélectionnez :

   Analyser > **Méta-analyse** > **Résultats continus ou résultats binaires** > **Données brutes...** ou **taille d'effet précalculée...**
2. Dans la boîte de dialogue **Méta-analyse**, cliquez sur **Tracer** puis sur l'onglet **Graphique en entonnoir**.
3. Sélectionnez et définissez les paramètres de graphique en entonnoir appropriés.
4. Cliquez sur **Continuer**

Graphique de Galbraith

L'onglet **Graphique de Galbraith** propose des paramètres permettant de contrôler les graphiques de Galbraith qui s'affichent dans la sortie pour une méta-analyse avec des résultats continus et binaires lorsque des données brutes et des données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Graphique de Galbraith
Ce paramètre facultatif active et désactive la sortie de graphique à bulles. Il est désactivé par défaut.

Afficher les limites de l'intervalle de confiance
Contrôle l'affichage des limites de l'intervalle de confiance. Ce paramètre est activé par défaut, ce qui permet d'ajouter les limites de l'intervalle de confiance dans le graphique à bulles. Lorsque ce paramètre est désactivé, les limites de l'intervalle de confiance ne s'ajoutent pas dans le graphique à bulles.

Plage de l'axe des X
Spécifie la courbe de l'axe des X. Lorsque le réglage est activé, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound ≥ 0`. Il est à noter que les valeurs spécifiées s'appliquent à tous les graphiques de Galbraith.

Plage de l'axe des Y
Spécifie la courbe de l'axe des Y. Lorsque le réglage est activé, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound`. Il est à noter que les valeurs spécifiées s'appliquent à tous les graphiques de Galbraith.

Variables
Répertorie les variables de jeux de données disponibles.

Libellé
Vous avez également la possibilité de spécifier une variable sur laquelle le graphique à bulles est libellé. La liste **Placement** fournit des options permettant de placer les étiquettes automatiquement (**Auto**) ou à **droite**, à **gauche**, dans la partie **supérieure** et en **bas**.

Définition des paramètres de graphique de Galbraith

1. A partir des menus, sélectionnez :

   Analyser > **Méta-analyse** > **Résultats continus ou résultats binaires** > **Données brutes...** ou **taille d'effet pré-calculée...**
2. Dans la boîte de dialogue **Méta-analyse**, cliquez sur **Tracer** puis sur l'onglet **Graphique de Galbraith**.
3. Sélectionnez et définissez les paramètres de graphique de Galbraith appropriés.
4. Cliquez sur **Continuer**

Graphique de l'Abbé

L'onglet **Graphique de l'Abbé** propose des paramètres permettant de contrôler les graphiques de l'Abbé qui s'affichent dans la sortie pour une méta-analyse avec des résultats continus et binaires lorsque des données brutes et des données de taille d'effet précalculées sont fournies dans le jeu de données actif.

Graphique de l'Abbé
Ce paramètre facultatif active et désactive la sortie de graphique de l'Abbé. Il est désactivé par défaut.

Dessiner des points proportionnels aux pondérations
Contrôle la façon dont les points sont traces par rapport aux pondérations. Ce paramètre est activé par défaut, ce qui permet de tracer les points proportionnellement aux pondérations. Lorsque ce paramètre est désactivé, tous les points sont tracés à la même taille.

Lignes de référence

Taille d'effet globale
Contrôle l'affichage de la ligne de référence qui représente la taille d'effet globale estimée. Ce paramètre est désactivé par défaut, ce qui supprime la ligne de référence. la ligne de référence s'affiche lorsque le paramètre est activé.

Taille d'effet nulle
Contrôle l'affichage de la ligne de référence qui représente un effet nul. Ce paramètre est activé par défaut, ce qui affiche la ligne de référence. La ligne de référence est supprimée lorsque e paramètre n'est pas activé.

Plage de l'axe des X
Spécifie la courbe de l'axe des X. Lorsque le réglage est activé, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound`. Il est à noter que les valeurs spécifiées s'appliquent à tous les graphiques de Galbraith.

Pour les méta-analyses avec des résultats binaires, lorsque le paramètre **Taille de l'effet** est défini sur **Différence de risque**, il doit également satisfaire à `Lower bound ≥ 0` et `Upper bound ≤ 1`. Lorsque le paramètre **Taille de l'effet** est réglé sur **Log Risk Ratio**, il doit satisfaire à `Upper bound ≤ 0`.

Plage de l'axe des Y
Spécifie la courbe de l'axe des Y. Lorsque le réglage est activé, deux valeurs numériques doivent être spécifiées et satisfaire à `Upper bound > Lower bound`. Il est à noter que les valeurs spécifiées s'appliquent à tous les graphiques de Galbraith.

Pour les méta-analyses avec des résultats binaires, lorsque le paramètre **Taille de l'effet** est défini sur **Différence de risque**, il doit également satisfaire à `Lower bound ≥ 0` et `Upper bound ≤ 1`. Lorsque le paramètre **Taille de l'effet** est réglé sur **Log Risk Ratio**, il doit satisfaire à `Upper bound ≤ 0`.

Variables
Répertorie les variables de jeux de données disponibles.

Libellé
Vous avez également la possibilité de spécifier une variable sur laquelle le graphique à bulles est libellé. La liste **Placement** fournit des options permettant de placer les étiquettes automatiquement (**Auto**) ou à **droite**, à **gauche**, dans la partie **supérieure** et en **bas**.

Définition des paramètres de graphique de l'Abbé

1. A partir des menus, sélectionnez :

   Analyser > **Méta-analyse** > **Résultats continus ou résultats binaires** > **Données brutes...**
2. Dans la boîte de dialogue **Méta-analyse**, cliquez sur **Tracer** puis sur l'onglet **Graphique de l'Abbé**.
3. Sélectionnez et définissez les paramètres de graphique de l'Abbé appropriés.
4. Cliquez sur **Continuer**

Livre de codes

Le livre de codes indique les informations du dictionnaire, telles que les noms de variables, les libellés de variables, les libellés de valeurs, les valeurs manquantes, ainsi que les statistiques récapitulatives de toutes les variables (ou celles spécifiées) et les jeux de réponses multiples dans le jeu de données actif. Pour les variables ordinales et nominales ainsi que pour les jeux de réponses multiples, les statistiques récapitulatives comprennent les effectifs et les pourcentages. Pour les variables d'échelle, les statistiques récapitulatives comprennent la moyenne, l'écart type et les quartiles.

Remarque : Le livre de codes ignore le statut du fichier scindé. Ceci comprend les groupes de fichiers scindés crées pour l'imputation multiple de valeurs manquantes (disponible dans l'option complémentaire Valeurs manquantes).

Pour obtenir un livre de codes

1. A partir des menus, sélectionnez :

   Analyse > **Rapports** > **Livre de codes**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Cliquez sur l'onglet Variables.
3. Sélectionnez une ou plusieurs variables et/ou des jeux de réponses multiples.

(En option) Vous avez également ces possibilités :

- Contrôler les informations de variables affichées.
- Contrôler les statistiques affichées (ou exclure toutes les statistiques récapitulatives).
- Contrôler l'ordre d'affichage des variables et des jeux de réponses multiples.
- Modifier le niveau de mesure de toute variable dans la liste source afin de modifier les statistiques récapitulatives affichées. Pour plus d'informations, voir «Onglet Statistiques du livre de codes», à la page 89.

Modification des niveaux de mesure

Vous pouvez modifier temporairement le niveau de mesure des variables. (Vous ne pouvez pas modifier celui des jeux de réponses multiples. Ils sont toujours traités comme nominaux.)

1. Cliquez avec le bouton droit de la souris sur une variable de la liste source.
2. Sélectionnez un niveau de mesure dans le menu contextuel.

Ceci permet de modifier temporairement le niveau de mesure. Concrètement, cela n'est utile que pour les variables numériques. Le niveau de mesure des variables de chaîne est limité aux variables nominales ou ordinales qui sont traitées de la même façon par la procédure du livre de codes.

Onglet Sortie du livre de codes

L'onglet Sortie contrôle les informations de variables disponibles pour chaque variable et jeu de réponses multiples, leur ordre d'affichage et le contenu de la table d'informations des fichiers en option.

Informations sur les variables

Ceci permet de contrôler les informations du dictionnaire affichées pour chaque variable.

Position : Nombre entier qui représente la position de la variable dans l'ordre des fichiers. Non disponible pour les jeux de réponses multiples.

Libellé Libellé descriptif associé à la variable ou au jeu de réponses multiples.

Type : Type de données fondamental. Les valeurs admises sont *Numérique*, *Chaîne* ou *Jeu de réponses multiples*.

Format : Format d'affichage de la variable, tel que *A4*, *F8.2* ou *DATE11*. Non disponible pour les jeux de réponses multiples.

Niveau de mesure. Les valeurs possibles sont *Nominale*, *Ordinale*, *Echelle* et *Inconnue*. La valeur affichée est le niveau de mesure stocké dans le dictionnaire et elle n'est pas affectée par tout remplacement de niveau de mesure temporaire spécifié en changeant le niveau de mesure dans la liste de variable source de l'onglet Variables. Non disponible pour les jeux de réponses multiples.

Remarque : Le niveau de mesure des variables numériques peut être "inconnu" avant le premier passage de données lorsque le niveau de mesure n'a pas été explicitement défini, par exemple pour la lecture de données à partir d'une source externe ou des variables récemment créées. Pour plus d'informations, voir .

Rôle : Certaines boîtes de dialogue prennent en charge la fonction de présélection de variables pour une analyse basée sur des rôles définis.

Libellés de valeurs : Libellés descriptifs associés à des valeurs de données spécifiques.

- Si l'option Effectif ou Pourcentage est sélectionnée dans l'onglet Statistiques, les libellés de valeurs définis sont compris dans les sorties même si vous ne sélectionnez pas Libellés de valeur ici.
- Pour les jeux de dichotomies multiples, les "libellés de valeur" sont les libellés des variables élémentaires du jeu ou les libellés des valeurs comptées, selon la définition du jeu. Pour plus d'informations, voir .

Valeurs manquantes : Valeurs manquantes définies par l'utilisateur. Si l'option Effectif ou Pourcentage est sélectionnée dans l'onglet Statistiques, les libellés de valeurs définis sont compris dans les sorties même si vous ne sélectionnez pas Valeurs manquantes ici. Non disponible pour les jeux de réponses multiples.

Attributs personnalisés : Attributs de variable personnalisés. Les sorties comprennent à la fois les noms et les valeurs pour tout attribut de variables personnalisé associé à chaque variable. Pour plus d'informations, voir . Non disponible pour les jeux de réponses multiples.

Attributs réservés : Attributs de variables système réservés. Vous pouvez afficher les attributs système, mais vous ne devez pas les modifier. Les noms d'attribut système commencent par un signe dollar ($). Les attributs non affichés, dont les noms commencent par « @ » ou « $@ », ne sont pas inclus. Les sorties comprennent à la fois les noms et les valeurs pour tout attribut système associé à chaque variable. Non disponible pour les jeux de réponses multiples.

Informations sur les fichiers

La table d'informations de fichiers en option peut comprendre l'un des attributs de fichiers suivants :

Nom de fichier : Nom du fichier de données IBM SPSS Statistiques. Si le jeu de données n'a jamais été enregistré au format IBM SPSS Statistiques, aucun nom de fichier de données n'est disponible. (Si aucun nom de fichier n'est affiché dans la barre de titre de la fenêtre Editeur de données, le jeu de données actif ne comporte pas de nom de fichier.)

Emplacement : Emplacement du répertoire (dossier) du fichier de données IBM SPSS Statistiques. Si le jeu de données n'a jamais été enregistré au format IBM SPSS Statistiques, aucun n'emplacement n'est disponible.

Nombre d'observations : Nombre d'observations dans le jeu de données actif. Ceci est le nombre total d'observations, y compris celles qui peuvent être exclues des statistiques récapitulatives en raison des conditions de filtrage.

Libellé Il s'agit de l'étiquette du fichier (s'il y en a une) définie par la commande `FILE LABEL` .

Documents : Texte de document de fichier de données.

Statut de la pondération : Si la pondération est activée, le nom de la variable de pondération est affiché. Pour plus d'informations, voir .

Attributs personnalisés : Attributs de fichiers de données personnalisés définis par l'utilisateur. Les attributs de fichiers de données sont définis avec la commande `DATAFILE ATTRIBUTE`.

Attributs réservés : Attributs de fichiers de données système réservés. Vous pouvez afficher les attributs système, mais vous ne devez pas les modifier. Les noms d'attribut système commencent par un signe dollar ($). Les attributs non affichés, dont les noms commencent par « @ » ou « $@ », ne sont pas inclus. Les sorties incluent les noms et les valeurs pour tout attribut de fichiers de données système.

Ordre d'affichage des variables

Les alternatives suivantes sont disponibles pour contrôler l'ordre d'affichage des variables et des jeux de réponses multiples :

Alphabétique : Ordre alphabétique par nom de variable.

Fichier : Ordre d'affichage des variables dans le jeu de données (leur ordre d'affichage dans l'Editeur de données). Dans l'ordre croissant, les jeux de réponses multiples sont affichés en dernier, après toutes les variables sélectionnées.

Niveau de mesure : Tri par niveau de mesure. Ceci crée quatre groupes de tri (nominal, ordinal, échelle et inconnu). Les jeux de réponses multiples sont traités comme nominaux.

Remarque : Le niveau de mesure des variables numériques peut être "inconnu" avant le premier passage de données lorsque le niveau de mesure n'a pas été explicitement défini, par exemple pour la lecture de données à partir d'une source externe ou des variables récemment créées.

Liste des variables : Ordre d'affichage des variables et des jeux de réponses multiples dans la liste des variables sélectionnées de l'onglet Variables.

Nom d'attribut personnalisé : La liste des options d'ordre de tri comprend aussi le nom des attributs de variables personnalisés définis par l'utilisateur. Dans l'ordre croissant, les variables dont le tri des attributs ne figure pas en haut, puis celles dont la valeur n'est pas définie pour l'attribut, puis celles avec des valeurs définies pour l'attribut dans l'ordre alphabétique des valeurs.

Nombre maximal de catégories

Si les sorties comprennent les libellés de valeurs, les effectifs, ou les pourcentages pour chaque valeur unique, vous pouvez supprimer ces informations de la table si le nombre de valeurs dépasse la valeur indiquée. Par défaut, ces informations sont supprimées si le nombre de valeurs uniques de la variable dépasse 200.

Onglet Statistiques du livre de codes

L'onglet Statistiques permet de contrôler les statistiques récapitulatives comprises dans les sorties, ou de supprimer entièrement l'affichage des statistiques récapitulatives.

Nombres et pourcentages

Pour les variables nominales et ordinales, les jeux de réponses multiples et les valeurs de libellé des variables d'échelle, les statistiques disponibles sont :

Comptage. Effectif ou nombre d'observations possédant chaque valeur (ou plage de valeurs) d'une variable.

Pour cent. Pourcentage d'observations ayant une valeur spécifique.

Tendance et dispersion centrales

Pour les variables d'échelle, les statistiques disponibles sont :

moyenne. Mesure de la tendance centrale. Moyenne arithmétique ; somme divisée par le nombre d'observations.

Écart type. Mesure de la dispersion des valeurs autour de la moyenne. Dans le cas d'une distribution normale, 68 % des observations se situent à l'intérieur d'un écart type de la moyenne et 95 % se situent à l'intérieur de deux écarts types. Par exemple, si la moyenne d'âge est de 45 avec un écart type égal à 10, une distribution normale verra 95 % des observations se situer entre 25 et 65.

quartiles. Valeurs correspondant aux 25ème, 50ème et 75ème percentiles.

Remarque : Vous pouvez modifier temporairement le niveau de mesure associé à une variable (et par conséquent modifier les statistiques récapitulatives affichées pour cette variable) dans la liste de variables source de l'onglet Variables.

Fréquences

La procédure Effectifs permet d'obtenir des affichages statistiques et graphiques qui servent à décrire de nombreux types de variables. Elle peut jouer un rôle lorsque vous prenez connaissance de vos données.

Pour obtenir un rapport de fréquences et un graphique à barres, vous pouvez trier les différentes valeurs par ordre croissant ou décroissant, ou bien classer les catégories en fonction de leurs fréquences. Le rapport de fréquences peut être supprimé lorsqu'une variable a plusieurs valeurs distinctes. Vous pouvez libeller les graphiques avec des fréquences (par défaut) ou des pourcentages.

Exemple

Quelle est la répartition de la clientèle d'une société selon le type d'industrie dont elle fait partie ? La sortie pourrait vous apprendre que votre clientèle est composée à 37,5 % d'organismes d'état, à 24,9 % de sociétés commerciales, à 28,1 % d'établissements universitaires et à 9,4 % du secteur de la santé. Pour des données continues et quantitatives, comme par exemple les chiffres d'affaires, vous pourriez constater que la moyenne de vente par produit est de 3 576 € avec un écart type de 1 078 €.

Statistiques et tracés

Effectifs de fréquences, pourcentages, pourcentages cumulés, moyenne, médiane, mode, somme, écart type, variance, plage, valeurs minimale et maximale, erreur standard de la moyenne, asymétrie et kurtosis (avec leurs erreurs standard), quartiles, percentiles choisis par l'utilisateur, graphiques à barres, graphiques circulaires et histogrammes.

Remarques sur les données

Données
Utilisez des codes numériques ou alphanumériques pour coder les variables catégorielles (mesures de niveau nominal ou ordinal).

Hypothèses
Les tabulations et les pourcentages fournissent une description utile sur les données de n'importe quelle distribution, particulièrement pour les variables disposant de catégories triées ou non. Certaines des statistiques récapitulatives facultatives, telles que la moyenne et l'écart type, sont fondées sur la théorie de normalité et sont appropriées pour des variables quantitatives avec une distribution symétrique. Les statistiques de base, telles que la médiane, les quartiles et les percentiles, sont appropriées pour les variables quantitatives, qu'elles répondent ou non au critère de normalité.

Obtention de tables de fréquences

1. A partir des menus, sélectionnez :

   Analyser > **Statistiques descriptives** > **Fréquences ...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables catégorielles ou quantitatives.
3. Sélectionnez éventuellement **Créer des tables de style APA** pour créer des tables de sortie qui respectent les instructions de style APA.
4. (En option) Vous avez également ces possibilités :
   - Cliquer sur **Statistiques** pour obtenir des statistiques descriptives pour des variables quantitatives.
   - Cliquer sur **Graphiques** pour obtenir des graphiques à barres, des graphiques circulaires ou des histogrammes.
   - Cliquer sur **Format** pour définir l'ordre de présentation des résultats.
   - Cliquer sur **Style** pour spécifier des conditions de modification automatique des propriétés des tableaux croisés dynamiques en fonction de certaines conditions.
   - Cliquer sur **Bootstrap** pour déduire des estimations fiables des erreurs standard et des intervalles de confiance pour des estimations telles que la moyenne, la médiane, la proportion, le rapport des cotes, le coefficient de corrélation ou le coefficient de régression. Elle peut aussi être utilisée pour construire des tests d'hypothèse.

Statistiques des effectifs

Percentiles : Valeurs d'une variable quantitative qui divisent les données triées en groupes par centième. Les quartiles (25e, 50e et 75e percentiles) divisent les observations en quatre groupes de taille égale. Si vous souhaitez un nombre égal de groupes différent de quatre, sélectionnez **Partition en n groupes égaux**. Vous pouvez également spécifier des percentiles particuliers (par exemple, le 95e percentile, valeur au-dessus de 95 % des observations).

Tendance centrale : Les statistiques décrivant la position de la distribution comprennent la Moyenne, la Médiane, le Mode et la Somme de toutes les valeurs.

- *moyenne*. Mesure de la tendance centrale. Moyenne arithmétique ; somme divisée par le nombre d'observations.
- *Valeur médiane*. La valeur au-dessus et en dessous de laquelle se situent la moitié des cas, c'est-à-dire le 50e centile. Si le nombre d'observations est pair, la médiane correspond à la moyenne des deux observations du milieu lorsqu'elles sont triées dans l'ordre croissant ou décroissant. La médiane est une mesure de tendance centrale et elle n'est pas, à l'inverse de la moyenne, sensible aux valeurs éloignées.

- *Mode*. Valeur qui revient le plus fréquemment. Si plusieurs valeurs partagent la plus grande fréquence d'occurrence, chacune d'elles constitue un mode. La procédure Effectifs ne rend compte que du plus petit mode.
- *somme*. Somme ou total des valeurs, pour toutes les observations n'ayant pas de valeur manquante.

Dispersion : Les statistiques mesurant la variance ou la dispersion dans les données, comprennent l'écart type, la variance, la plage, le minimum, le maximum et l'erreur standard (ES) de la moyenne.

- *Erreur déviation*. Mesure de la dispersion des valeurs autour de la moyenne. Dans le cas d'une distribution normale, 68 % des observations se situent à l'intérieur d'un écart type de la moyenne et 95 % se situent à l'intérieur de deux écarts types. Par exemple, si la moyenne d'âge est de 45 avec un écart type égal à 10, une distribution normale verra 95 % des observations se situer entre 25 et 65.
- *Variance*. Mesure de la dispersion des valeurs autour de la moyenne, égale à la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d'observations moins un. La variance se mesure en unités, qui sont égales au carré des unités de la variable.
- *Intervalle*. Différence entre la valeur maximale et la valeur minimale d'une variable numérique (maximum–minimum).
- *Minimum*. La plus petite valeur d'une variable numérique.
- *maximum*. La plus grande valeur d'une variable numérique.
- *S. E. signifie*. Variation de la valeur de la moyenne d'un échantillon à l'autre issus d'une même distribution. Cette mesure permet de comparer approximativement la moyenne observée avec une valeur hypothétique (autrement dit, vous pouvez conclure que ces deux valeurs sont différentes si le rapport de la différence avec l'erreur standard est inférieur à -2 ou supérieur à +2).
- *CV*. Rapport entre l'écart type et la moyenne. Les valeurs sont exprimées en pourcentage.

Distribution : L'asymétrie et kurtosis sont des statistiques qui décrivent la forme et la symétrie de la distribution. Ces statistiques sont présentées avec leurs erreurs standard.

- *asymétrie* . Mesure de l'asymétrie d'une distribution. La distribution normale est symétrique et a une valeur d'asymétrie égale à 0. Une distribution caractérisée par une importante asymétrie positive présente une partie droite plus allongée. Une distribution caractérisée par une importante asymétrie négative présente une extrémité gauche plus allongée. Pour simplifier, une valeur d'asymétrie deux fois supérieure à l'erreur standard correspond à une absence de symétrie.
- *Kurtosis*. Mesure de l'étendue de la présence de valeurs extrêmes. Dans le cas d'une distribution normale, la valeur de la statistique kurtosis est égale à zéro. Un kurtosis positif indique qu'on observe dans les données plus de valeurs extrêmes que dans une distribution normale. Un kurtosis négatif indique que les données présentent moins de valeurs extrêmes par rapport à une distribution normale. La définition de kurtosis utilisée, où la valeur est 0 pour une distribution normale, est parfois appelée kurtosis excessif. Certains logiciels peuvent signaler un kurtosis de telle sorte que la valeur est 3 pour une distribution normale.

Valeurs sont des centres de groupes : Si les valeurs dans vos données représentent des centres de groupes (par exemple, les âges des individus trentenaires sont représentés par le code 35), sélectionnez cette option pour estimer la médiane et les percentiles des données originales, non regroupées.

Graphiques des effectifs

Remarque : Aucun graphique n'est généré dans la sortie si l'option **Effectuer un bootstrap** est activée dans la boîte de dialogue Bootstrap.

Type de graphique

Un graphique circulaire montre la participation de chaque partie à l'ensemble. Chaque tranche du graphique circulaire correspond à un groupe défini par une simple variable de regroupement. Un graphique à barres montre l'effectif de chaque valeur ou de chaque catégorie sous la forme d'une barre distincte, ce qui vous permet de comparer les catégories visuellement. Un histogramme est également constitué de barres mais ils sont répartis à intervalles égaux. La hauteur de chaque barre représente l'effectif des valeurs d'une variable quantitative appartenant à l'intervalle. Un histogramme

montre la forme, le centre et la dispersion de la distribution. Si vous superposez une courbe normale sur l'histogramme, vous pouvez déterminer si les données sont distribuées normalement.

Valeurs du graphique
Dans les graphiques à barres, l'axe peut être libellé par effectifs ou pourcentages de fréquences.

Format des effectifs

Ordre d'affichage : La table de fréquences peut être affichée en fonction des valeurs réelles des données ou de l'effectif (fréquence d'occurrence) de ces valeurs, et organisée par valeurs croissantes ou décroissantes. Cependant, si vous demandez un histogramme ou des percentiles, le format des effectifs part du principe que la variable est quantitative et affiche ses valeurs par ordre croissant.

Variables multiples : Si vous créez des tableaux statistiques pour des variables multiples, vous pouvez afficher toutes les variables dans un tableau unique (**Comparer variables**) ou bien afficher un tableau statistique séparé pour chaque variable (**Séparer la sortie par variables**).

Supprimer les tableaux comportant plusieurs catégories : Cette option évite l'affichage des tableaux ayant plus que le nombre spécifié de valeurs.

Descriptives

La procédure Descriptives affiche les statistiques récapitulatives univariées pour plusieurs variables en un seul tableau et calcule les valeurs standardisées (scores *z*). Les variables peuvent être ordonnées en fonction de la taille de leurs moyennes (en ordre ascendant ou descendant), alphabétiquement ou selon l'ordre dans lequel vous avez sélectionné les variables (par défaut).

Lorsque les scores *z* sont enregistrés, ils sont ajoutés aux données dans l'éditeur de données et sont disponibles pour les graphiques, les listes de données et les analyses. Lorsque les variables sont enregistrées avec des unités différentes (par exemple, produit domestique brut par personne et pourcentage de la population sachant lire et écrire), une transformation en score *z* place les variables sur une échelle commune pour que la comparaison soit plus facile.

Exemple. Si chaque observation dans vos données contient les totaux des ventes quotidiennes pour chacun des membres du personnel commercial (par exemple, une entrée pour Bob, une pour Kim et une pour Brian) rapportés chaque jour pendant plusieurs mois, la procédure Descriptives peut calculer les ventes quotidiennes moyennes pour chacun des membres du personnel et ordonner les résultats de la moyenne des ventes la plus élevée à la plus basse.

Statistiques : Taille de l'échantillon, moyenne, minimum, maximum, écart type, variance, plage, somme, erreur standard de la moyenne, et kurtosis et asymétrie avec leurs erreurs standard (ES).

Remarques sur les données de Descriptives

Données. Utilisez des variables numériques après les avoir visualisées graphiquement en cherchant des erreurs d'enregistrement, les valeurs extrêmes et les anomalies de distribution. La procédure Descriptives est très efficace pour les gros fichiers (milliers d'observations).

Hypothèses : La plupart des statistiques disponibles (y compris les écarts *z*) sont basées sur une théorie normale et conviennent pour des variables continues (mesures de niveau d'intervalle ou de rapport) avec distribution symétrique. Evitez les variables avec des catégories désordonnées ou des répartitions asymétriques. La distribution des écarts *z* a la même forme que celle des données d'origine. Ainsi, le calcul des écarts *z* n'est pas une solution aux données posant des problèmes.

Pour obtenir des statistiques descriptives

1. A partir des menus, sélectionnez :

   Analyse > **Statistiques descriptives** > **Descriptives...**

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

(En option) Vous avez également ces possibilités :

- Cliquez sur **Enregistrer des valeurs standardisées dans des variables** pour enregistrer les écarts *z* comme nouvelles variables.
- Cliquez sur **Options** pour les statistiques optionnelles et l'ordre d'affichage.

Options Descriptives

Moyenne et somme : La moyenne ou moyenne arithmétique s'affiche par défaut.

Dispersion : Les statistiques qui mesurent l'étendue ou les variations dans les données comprennent l'écart type, la variance, la plage, le minimum, le maximum, et l'erreur standard (ES) de la moyenne.

- *Erreur déviation*. Mesure de la dispersion des valeurs autour de la moyenne. Dans le cas d'une distribution normale, 68 % des observations se situent à l'intérieur d'un écart type de la moyenne et 95 % se situent à l'intérieur de deux écarts types. Par exemple, si la moyenne d'âge est de 45 avec un écart type égal à 10, une distribution normale verra 95 % des observations se situer entre 25 et 65.
- *Variance*. Mesure de la dispersion des valeurs autour de la moyenne, égale à la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d'observations moins un. La variance se mesure en unités, qui sont égales au carré des unités de la variable.
- *Intervalle*. Différence entre la valeur maximale et la valeur minimale d'une variable numérique (maximum–minimum).
- *Minimum*. La plus petite valeur d'une variable numérique.
- *maximum*. La plus grande valeur d'une variable numérique.
- *Moyenne S.E.*. Variation de la valeur de la moyenne d'un échantillon à l'autre issus d'une même distribution. Cette mesure permet de comparer approximativement la moyenne observée avec une valeur hypothétique (autrement dit, vous pouvez conclure que ces deux valeurs sont différentes si le rapport de la différence avec l'erreur standard est inférieur à -2 ou supérieur à +2).
- *CV*. Rapport entre l'écart type et la moyenne. Les valeurs sont exprimées en pourcentage.

Distribution : Kurtosis et l'asymétrie sont des statistiques qui caractérisent la forme et la symétrie de la distribution. Ces statistiques sont présentées avec leurs erreurs standard.

- *Kurtosis*. Mesure de l'étendue de la présence de valeurs extrêmes. Dans le cas d'une distribution normale, la valeur de la statistique kurtosis est égale à zéro. Un kurtosis positif indique qu'on observe dans les données plus de valeurs extrêmes que dans une distribution normale. Un kurtosis négatif indique que les données présentent moins de valeurs extrêmes par rapport à une distribution normale. La définition de kurtosis utilisée, où la valeur est 0 pour une distribution normale, est parfois appelée kurtosis excessif. Certains logiciels peuvent signaler un kurtosis de telle sorte que la valeur est 3 pour une distribution normale.
- *asymétrie* . Mesure de l'asymétrie d'une distribution. La distribution normale est symétrique et a une valeur d'asymétrie égale à 0. Une distribution caractérisée par une importante asymétrie positive présente une partie droite plus allongée. Une distribution caractérisée par une importante asymétrie négative présente une extrémité gauche plus allongée. Pour simplifier, une valeur d'asymétrie deux fois supérieure à l'erreur standard correspond à une absence de symétrie.

Ordre d'affichage : Par défaut, les variables s'affichent dans l'ordre dans lequel vous les avez sélectionnées. En option, vous pouvez afficher les variables alphabétiquement, par moyennes croissantes ou par moyennes décroissantes.

Fonctions supplémentaires de la commande DESCRIPTIVES

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

- Enregistrer les scores standardisés (écarts *z*) pour certaines variables uniquement (à l'aide de la sous-commande VARIABLES).
- Spécifier le nom des nouvelles variables contenant des scores standardisés (à l'aide de la sous-commande VARIABLES).

- Exclure de l'analyse les observations ayant des valeurs manquantes pour n'importe quelle variable (à l'aide de la sous-commande MISSING).
- Trier les variables affichées en utilisant la valeur d'une statistique, et pas uniquement la moyenne (à l'aide de la sous-commande SORT).

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

Percentiles

La procédure Percentiles affiche des tableaux de percentiles.

Valeurs des percentiles
Lorsque vous sélectionnez **Valeurs centiles**, les valeurs par défaut affichées sont celles des 5e, 10e, 25e, 50e, 75e, 90e et 95e centiles. Valeurs d'une variable quantitative qui divisent les données triées en groupes par centième.

Quartiles
Les quartiles (25e, 50e et 75e percentiles) divisent les observations en quatre groupes de taille égale.

Personnalisé
Lorsque vous sélectionnez **Personnalisé**, entrez au moins une valeur pour exécuter l'analyse. Les valeurs d'entrée doivent être un nombre compris entre 0 et 100. Utilisez les boutons **Ajouter**, **Modifier**et **Supprimer** pour utiliser les valeurs de la liste des valeurs de percentiles.

Méthode des percentiles
Par défaut, la méthode HAVERAGE est sélectionnée pour le calcul des percentiles.

Obtention de tables de percentiles

1. A partir des menus, sélectionnez :

   Analyser > **Statistiques descriptives** > **Percentiles ..**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables numériques, puis placez-les dans le champ Variables.
3. Sélectionnez **Valeurs des percentiles** pour spécifier les valeurs de calcul du percentile.
4. Sélectionnez **Méthode des percentiles** pour sélectionner la méthode de calcul des percentiles.
5. (En option) Vous avez également ces possibilités :
   - Cliquez sur **Manquant** pour contrôler le traitement des valeurs manquantes.
   - Cliquez sur **Bootstrap** pour calculer des estimations robustes des erreurs standard et des intervalles de confiance pour les estimations des percentiles.

Percentiles Valeurs manquantes

Valeurs manquantes
Contrôle le traitement des valeurs manquantes.

Exclure toute observation incomplète

Les observations avec des valeurs manquantes pour une variable spécifiée sont exclues de toutes les analyses. Il s'agit de la valeur par défaut.

Par Variable
Exclure les observations avec des valeurs manquantes variable par variable.

Explorer

La procédure Explorer produit des statistiques récapitulatives et des affichages graphiques pour toutes vos observations ou séparément pour des groupes d'observations. Il existe plusieurs raisons pour utiliser la procédure Explorer : le balayage de données, l'identification des valeurs extrêmes, la description, la vérification d'hypothèses et la caractérisation des différences parmi les sous populations (groupes d'observations). Le balayage de données peut vous indiquer les valeurs inhabituelles, les valeurs extrêmes, les trous dans les données ou d'autres particularités. L'exploration des données peut vous aider à déterminer si les techniques statistiques que vous envisagez d'utiliser pour l'analyse de vos données sont appropriées. L'exploration peut indiquer que vous avez besoin de transformer les données si la technique nécessite une répartition gaussienne. Vous pouvez également choisir d'utiliser des tests non paramétriques.

Exemple. Examinez la distribution des temps d'apprentissage pour les souris dans un labyrinthe avec quatre programmes de renforcement. Pour chacun des quatre groupes, vous pouvez voir si la répartition des temps est approximativement gaussienne et si les quatre variances sont égales. Vous pouvez aussi identifier les observations avec les cinq plus grands et les cinq plus petits temps. Les boîtes à moustaches et les tracés tige et feuille résument graphiquement la répartition des temps d'apprentissage pour chacun des groupes.

Tracés et statistiques : Moyenne, médiane, moyenne tronquée à 5 %, erreur standard, variance, écart type, minimum, maximum, plage, plage interquartile, asymétrie et kurtosis avec leurs erreurs standard, intervalle de confiance pour la moyenne (et niveaux de confiance spécifiés), percentiles, M-estimateur de Huber, Andrews, Hampel, Tukey, les cinq plus grandes et cinq plus petites valeurs, la statistique de Kolmogorov-Smirnov avec un niveau de signification Lilliefors pour tester la normalité, et la statistique Shapiro-Wilk. Boîtes à moustaches, tracés tige et feuille, histogrammes, tracés de répartition gaussienne, et dispersion/niveaux avec le test de Levene et les transformations.

Remarques sur les Données d'Explorer

Données. La procédure d'Explorer peut être utilisée pour les variables quantitatives (Mesures de niveaux d'intervalle ou de rapport). Un facteur (utilisé pour répartir les données en groupes d'observations) doit avoir un nombre raisonnable de valeurs distinctes (catégories). Ces valeurs peuvent être des chaînes de caractères courtes ou numériques. La variable de libellé par observation, utilisée pour libeller les valeurs extrêmes dans les boîtes à moustaches, peut être de courtes chaînes de caractères, de longues chaînes de caractères (15 premiers octets) ou numériques.

Hypothèses : La distribution de vos données ne doit pas obligatoirement être symétrique ou gaussienne.

Pour explorer vos données

1. A partir des menus, sélectionnez :

   Analyse > **Statistiques descriptives** > **Explorer...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins une variable dépendante.

(En option) Vous avez également ces possibilités :

- Sélectionner un ou plusieurs facteurs, dont les valeurs définiront les groupes d'observations.
- Sélectionner une variable d'identification pour libeller les observations.
- Cliquer sur **Statistiques** pour les M-estimateurs, les Valeurs extrêmes, les Percentiles et les tables de fréquences.
- Cliquer sur **Tracés** pour les histogrammes, les tracés de probabilités gaussiens avec tests et la dispersion/niveau avec test de Levene.
- Cliquer sur **Options** pour le traitement des valeurs manquantes.

Statistiques d'Explorer

Descriptives : Ces mesures de tendance centrale et de dispersion s'affichent par défaut. Les mesures de tendance centrale indiquent la position de la répartition. On y trouve la moyenne, la médiane et la moyenne tronquée à 5 %. Les mesures de dispersion montrent la dissimilarité des valeurs ; on y trouve l'erreur standard, la variance, l'écart type, le minimum, le maximum, la plage, et la plage interquartile. Les statistiques descriptives comprennent aussi les mesures de la forme des répartitions. L'asymétrie et kurtosis s'affichent avec leurs erreurs standard. L'intervalle du niveau de confiance à 95 % pour la moyenne s'affiche aussi. Vous pouvez spécifier un niveau de confiance différent.

Moyennes pondérées : Estimations de la moyenne et de la médiane de l'échantillon pour estimer la localisation. Les estimateurs qui sont calculés diffèrent dans les pondérations qu'ils appliquent aux observations. M-estimateur de Huber, Andrew, Hampel, et Tukey apparaissent.

Valeurs extrêmes : Affiche les cinq plus grandes et cinq plus petites valeurs avec les libellés d'observations.

Percentiles

La procédure Percentiles affiche des tableaux de percentiles. Lorsque vous sélectionnez **« Percentile »**, les valeurs des 5e, 10e, 25e, 50e, 75e, 90e et 95e centiles s'affichent par défaut. Valeurs d'une variable quantitative qui divisent les données triées en groupes par centième.

Quartiles

Les quartiles (25e, 50e et 75e percentiles) divisent les observations en quatre groupes de taille égale.

Personnalisé

Lorsque vous sélectionnez **Personnalisé**, entrez au moins une valeur pour exécuter l'analyse. Les valeurs d'entrée doivent être un nombre compris entre 0 et 100. Utilisez les boutons **Ajouter**, **Modifier**et **Supprimer** pour utiliser les valeurs de la liste des valeurs de percentiles.

Méthode

Par défaut, la méthode HAVERAGE est sélectionnée pour le calcul des percentiles.

Tracés d'Explorer

Boîtes à moustaches : Ces alternatives contrôlent l'affichage de boîtes à moustaches quand vous avez plus d'une variable dépendante. L'option **Niveaux de facteur** génère un affichage séparé pour chaque variable dépendante. Dans un affichage, les boîtes à moustaches sont données pour chacun des groupes définis par un facteur. L'option **Dépendantes** génère un affichage séparé pour chaque groupe défini par un facteur. Dans un affichage, les boîtes à moustaches s'affichent côte à côte pour chaque variable dépendante. Cet affichage est particulièrement utile lorsque les différentes variables représentent une seule caractéristique mesurée à des moments différents.

Caractéristique : Le groupe caractéristiques vous permet de choisir les tracés tige et feuille et les histogrammes.

Tracés de répartition gaussiens avec tests : Affiche les tracés de probabilités gaussiens et les résidus. La statistique de Kolmogorov-Smirnov avec un niveau de signification Lilliefors pour le test de normalité s'affiche. Si des pondérations non entières sont spécifiées, la statistique de Shapiro-Wilk est calculée lorsque la taille de l'échantillon pondéré se situe entre 3 et 50. Pour les pondérations nulles ou entières, la statistique est calculée lorsque la taille de l'échantillon pondéré se situe entre 3 et 5 000.

Dispersion/niveau avec test de Levene : Contrôle les transformations de données pour les tracés de dispersion par niveau. Pour tous les tracés de dispersion par niveau, la pente de la ligne de régression et les tests de Levene portant sur l'homogénéité de la variance s'affichent. Si vous sélectionnez une transformation, les tests de Levene sont basés sur les données transformées. Si aucun facteur n'est sélectionné, les tracés de dispersion par niveau ne sont pas produits. L'option **Estimation d'exposants** produit un tracé des logs naturels des plages interquartile opposés au logs naturels des médianes pour toutes les cellules, en même temps qu'une estimation de la transformation de l'exposant pour arriver à des variances égales dans les cellules. Un tracé de dispersion par niveau aide à déterminer l'exposant pour qu'une transformation stabilise (rende plus égales) les variances entre groupes. L'option **Transformation** vous permet de sélectionner une des alternatives de l'exposant, en suivant

éventuellement les recommandations de l'estimation de l'exposant et de produire les tracés des données transformées. La plage interquartile et la médiane des données transformées sont dessinés. L'option **Sans transformation** produit des tracés de données brutes. Ceci est équivalent à une transformation avec une puissance de 1.

Transformations de l'exposant d'Explorer

Voici les transformations de l'exposant pour les tracés de dispersion par niveau. Pour transformer les données, vous devez sélectionner un exposant pour la transformation. Vous avez le choix entre les options suivantes :

- **Log népérien :** Transformation par log naturel. Il s'agit de la valeur par défaut.
- **1/racine carrée :** L'inverse de la racine carrée est calculée pour chaque valeur des données.
- **Inverse :** L'inverse de chaque valeur des données est calculée.
- **Racine carrée :** La racine carrée de chaque valeur des données est calculée.
- **Carré :** Chaque valeur des données est élevée au carré.
- **Cube.** Chaque valeur des données est élevée au cube.

Options d'Explorer

Valeurs manquantes : Contrôle le traitement des valeurs manquantes.

- **Exclure toute observation incomplète :** Les observations avec des valeurs manquantes pour l'une ou l'autre des variables dépendantes ou facteurs sont exclues de toutes les analyses. Il s'agit de la valeur par défaut.
- **Exclure seulement les composantes non valides :** Les observations sans valeur manquante pour une variable dans un groupe (cellule) sont inclues dans l'analyse de ce groupe. L'observation peut avoir des valeurs manquantes pour les variables utilisées dans d'autres groupes.
- **Signaler les valeurs manquantes :** Les valeurs manquantes pour les facteurs sont traitées comme une catégorie séparée. Toute sortie est produite pour cette catégorie supplémentaire. Les tables de fréquences contiennent les catégories pour les valeurs manquantes. Les valeurs manquantes pour un facteur sont inclues, mais libellées comme manquantes.

Fonctions supplémentaires de la commande EXAMINE

La procédure Explorer utilise la syntaxe de la commande EXAMINE. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- demander le total des sorties et des tracés en complément des sorties et tracés relatifs aux groupes définis par les facteurs (au moyen de la sous-commande TOTAL) ;
- spécifier une échelle commune pour un groupe de boîtes à moustaches (au moyen de la sous-commande SCALE) ;
- préciser les interactions des facteurs (au moyen de la sous-commande VARIABLES) ;
- spécifier les percentiles autres que ceux par défaut (au moyen de la sous-commande PERCENTILES) ;
- calculer les percentiles à l'aide de l'une des cinq méthodes possibles (au moyen de la sous-commande PERCENTILES) ;
- spécifier toute transformation de l'exposant pour les tracés de dispersion par niveau (au moyen de la sous-commande PLOT) ;
- préciser le nombre de valeurs extrêmes à afficher (au moyen de la sous-commande STATISTICS) ;
- indiquer les paramètres pour les M-estimateurs d'un emplacement (au moyen de la sous-commande MESTIMATORS).

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

tableaux croisés

La procédure Tableaux croisés établit des tableaux à deux entrées ou à entrées multiples et propose une variété de tests et de mesures d'associations pour les tableaux à deux entrées. La structure du tableau et l'ordre des catégories déterminent les tests ou mesures à effectuer.

A l'exception des coefficients gamma partiels, les statistiques et mesures d'associations des tableaux croisés sont calculées séparément pour chaque tableau à deux entrées. Si vous spécifiez une ligne, une colonne et une couche de facteur (variable de contrôle), la procédure Tableaux croisés forme un tableau de statistiques et de mesures pour chaque valeur de la couche de facteur (ou une combinaison de valeurs pour deux variables de contrôle ou plus). Par exemple, si le *sexe* est un facteur de couche pour un tableau *marié* (oui, non) face à la *vie* (est excitante, routinière ou ennuyeuse), les résultats d'un tableau à deux entrées pour les femmes sont calculés séparément de ceux des hommes et affichés sous forme de tableaux consécutifs.

Exemple. Les clients de PME ont-ils plus de probabilités d'être rentables en ventes de services (par exemple, formation et conseil) que ceux de grandes sociétés ? A partir d'une tabulation croisée, vous pourriez apprendre que la majorité des PME (moins de 500 salariés) génèrent des bénéfices de services élevés, alors que la majorité des grandes sociétés (plus de 2 500 salariés) rapportent des bénéfices de services bas.

Statistiques et mesures d'association : khi-carré de Pearson, khi-carré du rapport de vraisemblance, test d'association linéaire par linéaire, test exact de Fisher, khi-carré corrigé de Yates, *r* de Pearson, rho de Spearman, coefficient de contingence, phi, *V* de Cramér, lambdas symétriques et asymétriques, tau de Goodman et Kruskal, coefficient d'incertitude, gamma, *d* de Somers, tau-*b* de Kendall, tau-*c* de Kendall, coefficient êta, kappa de Cohen, estimations de risque relatif, rapport de cotes, test de McNemar, statistiques de Cochran et de Mantel-Haenszel et statistiques des proportions de colonne.

Remarques sur les Données pour tableau croisé

Données. Pour définir les catégories de chaque variable de table, utilisez les valeurs d'une variable numérique ou chaîne. Par exemple, pour *gender*, vous pouvez codifier les données avec 1 et 2, ou avec *homme* et *femme*.

Hypothèses : Des statistiques et des mesures partent du principe de catégories ordonnées (données ordinales) ou de valeurs quantitatives (données d'intervalle ou données de type rapport), tel que décrit dans la section sur les statistiques. D'autres sont valides lorsque les variables du tableau ont des catégories désordonnées (données nominales). Pour les statistiques basées sur le khi-carré (phi, *V* de Cramér et coefficient de contingence), les données doivent provenir d'un échantillon aléatoire avec une distribution multinomiale.

Remarque : Les variables ordinales peuvent être des codes numériques représentant des catégories (par exemple, 1 = *faible*, 2 = *moyen*, 3 = *élevé*) ou des valeurs de chaîne. Toutefois, l'ordre alphabétique des valeurs de chaîne est supposé refléter l'ordre réel des catégories. Par exemple, pour une variable de chaîne comportant des valeurs *Faible*, *Moyen*, *Elevé*, l'ordre des catégories est interprété comme *Elevé*, *Faible* ou *Moyen*, ce qui ne correspond pas à l'ordre correct. En règle générale, il est recommandé d'utiliser les codes numériques pour représenter les données ordinales.

Pour obtenir des tableaux croisés

1. A partir des menus, sélectionnez :

   Analyse > **Statistiques descriptives** > **Tableaux croisés...**

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

(En option) Vous avez également ces possibilités :

- Sélectionner des variables de contrôle.
- Cliquer sur **Statistiques** pour les tests et les mesures d'association pour les tableaux à deux entrées ou les sous-tables.

- Cliquer sur **Cellules** pour les valeurs observées et théoriques, les pourcentages et les résidus.
- Cliquer sur **Format** pour contrôler l'ordre des catégories.

Couches de tableaux croisés

Si vous sélectionnez des variables de couche, un tableau croisé séparé est produit pour chacune des catégories de variable de couche (variable de contrôle). Par exemple, si vous avez une variable de ligne, une variable de colonne, et une variable de couche avec deux catégories, vous obtenez un tableau à deux entrées pour chacune des catégories de la variable de couche. Pour créer une autre couche de variables de contrôle, cliquez sur **Suivant**. Les sous-tables sont produites pour chaque combinaison de catégories pour chaque variable de premier niveau avec chaque variable de second niveau, etc. Si les statistiques et les mesures d'association sont requises, elles ne s'appliquent qu'aux sous-tables à deux entrées.

Graphiques à barres en cluster de tableaux croisés

Affichage de graphiques à barres en cluster : Un graphique à barres en cluster vous permet de résumer vos données pour des groupes d'observations. Il y a un cluster de barres pour chaque valeur de la variable que vous avez spécifiée dans Ligne(s). La variable qui définit les barres dans chaque cluster est la variable que vous avez spécifiée dans Colonne(s). Il y a un ensemble de barres de couleurs ou de motifs différents pour chaque valeur de cette variable. Si vous spécifiez plus d'une variable dans Colonnes ou Lignes, un graphique à barres en cluster est produit pour chaque combinaison de deux variables.

Tableaux croisés affichant les variables de couche dans des couches de tableau

Afficher les variables de couche dans les couches du tableau : Vous pouvez choisir d'afficher les variables de couche (variables de contrôle) sous forme de couches de tableau dans le tableau croisé. Cela vous permet de créer des vues qui montrent les statistiques globales des variables de ligne et de colonne, et de faire défiler les catégories des variables de couche.

Vous trouverez ci-dessous un exemple utilisant le fichier de données *demo.sav* (disponible dans le dossier Echantillons du répertoire d'installation) ayant été obtenu comme suit :

1. Sélectionnez *Catégorie de revenu en milliers (rev_dis)* comme variable de ligne, *Possède un agenda électronique (PDA)* comme variable de colonne et *Niveau d'éducation (educ)* comme variable de couche.
2. Sélectionnez l'option **Afficher les variables de couche dans des couches de tableau**.
3. Sélectionnez **Colonne** dans la sous-boîte de dialogue Contenu des cases.
4. Exécutez la procédure Tableaux croisés, double-cliquez sur le tableau croisé et sélectionnez **Diplôme universitaire** dans la liste déroulante Niveau d'éducation.

La vue sélectionnée du tableau croisé montre les statistiques relatives aux répondants qui possèdent un diplôme universitaire.

Statistiques de tableaux croisés

Khi-deux. Pour les tableaux avec deux lignes et deux colonnes, sélectionnez **Khi-carré** pour calculer le khi-carré de Pearson, le khi-carré du rapport de vraisemblance, le test exact de Fisher et le test du khi-carré de Yates corrigé (correction de continuité). Pour les tableaux 2 × 2, le test exact de Fisher est calculé lorsqu'un tableau qui ne résulte pas de lignes ou de colonnes manquantes dans un tableau plus grand possède une cellule dont la fréquence attendue est inférieure à 5. Le khi-carré corrigé de Yates est calculé pour tous les autres tableaux 2 × 2. Pour les tableaux avec n'importe quel nombre de lignes ou de colonnes, sélectionnez **Khi-carré** pour calculer le khi-carré de Pearson et le rapport de vraisemblance du khi-carré. Lorsque les deux variables du tableau sont quantitatives, le **Khi-carré** donne le test d'association linéaire par linéaire.

Corrélations : Pour les tableaux dans lesquels les lignes et les colonnes contiennent des valeurs ordonnées, les **corrélations** donnent le coefficient de corrélation de Spearman, rho (données numériques

seulement). Le Spearman rho est une mesure d'association entre les ordres de rang. Lorsque les deux variables (facteurs) du tableau sont quantitatives, les **corrélations** donnent le coefficient de corrélation de Pearson, *r*, une mesure de l'association linéaire entre les variables.

Nominal : Pour les données nominales (sans ordre intrinsèque, comme Catholique, Protestant et Juif, vous pouvez sélectionner le **Coefficient de contingence**, le coefficient **Phi** et **V de Cramér**, **Lambda** (lambdas symétriques et asymétriques, et tau de Goodman et Kruskal) et le **Coefficient d'incertitude**.

- *Coefficient d'urgence*. Mesure d'association basée sur le khi-deux. Les valeurs sont toujours comprises entre 0 et 1, 0 indiquant l'absence d'association entre les variables de ligne et de colonne, et les valeurs proches de 1 indiquant un degré d'association élevé entre les variables. La valeur maximale possible dépend du nombre de lignes et de colonnes dans le tableau.
- *Phi et V de Cramer*. Mesure d'association calculée à partir du khi-carré. Elle est obtenue en divisant la statistique du khi-carré par la taille de l'échantillon, puis en prenant la racine carrée du résultat. Le V de Cramer est également une mesure d'association basée sur le khi-carré.
- *lambda*. Mesure d'association qui reflète la réduction proportionnelle de l'erreur lorsque les valeurs de la variable indépendante sont utilisées pour prévoir les valeurs de la variable dépendante. La valeur 1 signifie que la variable indépendante prévoit parfaitement la variable dépendante. La valeur 0 signifie que la variable indépendante ne prévoit pas du tout la variable dépendante.
- *Coefficient d'incertitude*. Mesure d'association qui indique la réduction proportionnelle de l'erreur lorsque les valeurs d'une variable sont utilisées pour prévoir les valeurs de l'autre variable. Par exemple, la valeur 0,83 indique que la connaissance d'une variable réduit de 83 % l'erreur dans les prévisions de l'autre variable. Le programme calcule à la fois des versions symétriques et asymétriques de ce coefficient.

Ordinal : Pour les tableaux dont les lignes et les colonnes contiennent des valeurs ordonnées, sélectionnez **Gamma** (ordre zéro pour les tableaux à 2 entrées et conditionnel pour les tableaux de 3 à 10 entrées), le **tau-b de Kendall** et le **tau-c de Kendall**. Pour prévoir les catégories de colonnes à partir des catégories de lignes, sélectionnez le **d de Somers**.

- *Gamma*. Mesure symétrique de l'association entre deux variables ordinales comprises entre -1 et 1. Les valeurs proches d'une valeur absolue de 1 indiquent une relation forte entre les deux variables. Les valeurs proches de 0 indiquent une relation faible ou inexistante. Pour les tableaux d'ordre 2, les gammas d'ordre 0 (zéro) apparaissent. Pour les tableaux d'ordre 3 et les tableaux d'ordre n, les gammas conditionnels apparaissent.
- *d de Somers*. Mesure d'association entre deux variables ordinales allant de -1 à 1. Les valeurs proches d'une valeur absolue de 1 indiquent une relation forte entre les deux variables, et les valeurs proches de 0 indiquent peu ou pas de relation entre les variables. Le d de Somer est une extension asymétrique du gamma, qui ne diffère de celui-ci que par l'inclusion du nombre de paires non liées à la variable indépendante. Le programme calcule également une version symétrique de cette statistique.
- *tau-b de Kendall*. Mesure de corrélation non paramétrique pour variables ordinales ou classées qui prend en compte les ex aequo. Le signe du coefficient indique la direction de la relation et sa valeur absolue indique sa force, les valeurs absolues les plus grandes indiquant les relations les plus fortes. Les valeurs peuvent varier de -1 à +1 mais une valeur de -1 ou de +1 ne peut toutefois être obtenue que dans des tableaux carrés.
- *tau-c de Kendall*. Mesure d'association non paramétrique pour les variables ordinales qui ignore les ex aequo. Le signe du coefficient indique la direction de la relation et sa valeur absolue indique sa force, les valeurs absolues les plus grandes indiquant les relations les plus fortes. Les valeurs peuvent varier de -1 à +1 mais une valeur de -1 ou de +1 ne peut toutefois être obtenue que dans des tableaux carrés.

Données nominales / intervalle : Lorsqu'une variable est catégorielle et l'autre quantitative, sélectionnez **Eta**. La variable catégorielle doit être codée numériquement.

- *Eta*. Mesure d'association qui varie de 0 à 1, 0 indiquant l'absence d'association entre les variables de ligne et de colonne et les valeurs proches de 1 indiquant un degré d'association élevé. Eta convient à une variable dépendante continue mesurée sur une échelle d'intervalle (par exemple, le revenu) et une variable indépendante ayant un nombre limité de catégories (par exemple, le sexe). Deux valeurs êta sont calculées : l'une traite la variable de ligne comme variable d'intervalle et l'autre traite la variable de colonne comme variable d'intervalle.

Kappa. Le kappa de Cohen mesure la concordance entre les évaluations de deux indicateurs utilisés pour évaluer un même objet. La valeur 1 indique une concordance parfaite. La valeur 0 indique que la concordance ne dépasse pas celle due au hasard. Le Kappa est basé sur un tableau carré dans lequel les valeurs des lignes et des colonnes représentent la même échelle. Toute cellule ayant observé des valeurs pour une variable, mais pas l'autre, se voit attribuer un nombre de 0. Kappa n'est pas calculé si le type de stockage de données (chaîne ou numérique) n'est pas le même pour les deux variables. Les deux variables de chaîne d'une paire doivent être de même longueur.

Risque. Pour les tables 2 x 2, mesure de la force de l'association entre la présence d'un facteur et la réalisation d'un événement. Si l'intervalle de confiance de la statistique inclut une valeur de 1, il n'existe aucune association entre le facteur et l'événement. Le rapport des cotes peut être utilisé comme estimation du risque relatif dans le cas où la réalisation du facteur est rare.

McNemar. Test non paramétrique pour deux variables dichotomiques liées. Il recherche les changements de réponse en utilisant la répartition khi-deux. Ce test est utile pour détecter les changements avant-après dans les réponses dus à une intervention expérimentale dans les plans. Pour les tableaux carrés plus volumineux, le test McNemar-Bowker de symétrie est reporté.

Statistiques de Cochran et de Mantel-Haenszel. Les statistiques de Cochran et de Mantel-Haenszel servent à tester l'indépendance entre une variable de facteur dichotomique et une variable de réponse dichotomique, selon les modèles de covariable définis par des variables de couche (contrôle). Remarque : alors que les autres statistiques sont calculées couche par couche, les statistiques de Cochran et de Mantel-Haenszel sont calculées une seule fois pour toutes les couches.

Affichage de cellules (cases) de tableaux croisés

Pour vous aider à découvrir des motifs dans les données qui contribuent à un test du khi-deux significatif, la procédure de Tableaux croisés affiche les fréquences attendues et trois types de résidus (déviations) qui mesurent la différence entre les fréquences observées et les fréquences attendues. Chaque cellule du tableau peut contenir toute combinaison d'effectifs, de pourcentages et de résidus sélectionnés.

Effectifs : Nombre d'observations effectivement observées et nombre d'observations attendues si les variables de ligne et de colonne sont indépendantes l'une de l'autre. Vous pouvez choisir de masquer les effectifs inférieurs à un entier spécifique. Les valeurs masquées s'affichent sous la forme **<N**, où **N** est l'entier spécifié. L'entier spécifié doit être supérieur ou égal à 2, bien que la valeur 0 soit permise et indique qu'aucun effectif n'est masqué.

Comparer les proportions de colonne : Cette option calcule les comparaisons appariées des proportions de colonne et indique quelles paires de colonnes (pour une ligne donnée) sont significativement différentes. Les différences significatives sont indiquées dans le tableau croisé dans un format de style APA à l'aide d'indices et sont calculées au niveau de signification 0,05. *Remarque* : Si cette option est spécifiée sans sélectionner les effectifs observés ou les pourcentages de colonne, alors les effectifs observés sont inclus dans le tableau croisé, les indices de style APA indiquant les résultats des tests de proportion de colonne.

- **Ajustement des valeurs p (méthode Bonferroni) :** Les comparaisons appariées des proportions de colonne utilisent la correction Bonferroni, qui ajuste le taux de signification observé pour les comparaisons multiples.

Pourcentages : Les pourcentages peuvent s'additionner par ligne ou par colonne. Les pourcentages du nombre total d'observations représentées dans le tableau (une couche) sont également disponibles.

Remarque : Si l'option **Masquer les petits effectifs** est sélectionnée dans le groupe Effectifs, les pourcentages associés aux effectifs masqués sont aussi masqués.

Résidus : Les résidus non standardisés fournissent la différence entre les valeurs observées et les valeurs théoriques. Les résidus standardisés et standardisés ajustés sont également disponibles.

- *Non normalisé*. Différence entre une valeur observée et la valeur attendue. La valeur théorique correspond au nombre d'observations attendues dans la cellule quand il n'existe pas de relation entre les deux variables. Un résidu positif indique que la cellule contient plus d'observations que si les variables de ligne et de colonne étaient indépendantes.

- *Standardisés*. Résidu, divisé par une estimation de son écart type. Egalement appelés résiduels de Pearson, les résiduels standardisés ont une moyenne de 0 et un écart type de 1.
- *Standardisé ajusté*. Résidu d'une cellule (valeur observée moins valeur théorique) divisé par une estimation de son erreur standard. Le résidu standardisé qui en résulte est exprimé en écarts par rapport à la moyenne.

Créer une table de style APA : Crée des tables de sortie qui respectent les instructions de style APA.

Remarque : Les options **Observé**, **Attendu**, **Ligne**, **Colonne** et **Total** ne sont pas disponibles si **Créer une table de style APA** est sélectionné.

Pondérations non entières : En général, les effectifs de cellules sont des valeurs entières, car ils représentent le nombre d'observations figurant dans chaque cellule. Toutefois, si le fichier de données est pondéré par une variable de pondération avec des fractions (par exemple, 1,25), les effectifs de cellules peuvent également être des fractions. Vous pouvez tronquer ou arrondir les valeurs avant ou après le calcul des effectifs de cellules, ou utiliser des effectifs de cellules non entiers pour l'affichage des tableaux et les calculs statistiques.

- *Effectifs de cellules arrondis*. Les pondérations d'observation sont utilisées telles quelles, mais les pondérations cumulées dans les cellules sont arrondies avant le calcul de toute statistique.
- *tronquer effectifs de cellules*. Les pondérations d'observation sont utilisées telles quelles, mais les pondérations cumulées dans les cellules sont arrondies avant le calcul de statistiques.
- *Pondérations des observations arrondies*. Les pondérations d'observation sont arrondies avant leur utilisation.
- *Pondérations des observations tronquées*. Les pondérations d'observation sont arrondies avant utilisation.
- *Aucun ajustement*. Les pondérations des observations sont utilisées telles quelles et les effectifs de cellules fractionnés sont utilisés. Toutefois, lorsque des statistiques exactes (disponibles uniquement avec Echantillonnage et test) sont demandées, les pondérations cumulées dans les cellules sont tronquées ou arrondies avant le calcul des statistiques du test exact.

Format de tableau croisé

Vous pouvez arranger les lignes par ordre croissant ou décroissant de valeur de la variable de ligne.

Récapituler

La procédure Récapituler calcule les statistiques de sous-groupes pour les variables à l'intérieur des catégories de variables de regroupement. Tous les niveaux de variables de regroupement sont à tabulation croisée. Vous pouvez choisir l'ordre dans lequel les statistiques sont affichées. Les statistiques récapitulatives sont affichées pour chaque variable à travers toutes les catégories. Les valeurs des données dans chaque catégorie peuvent être listées ou supprimées. Avec d'importants fichiers de données, vous pouvez choisir de lister seulement les premières observations *n*.

Exemple. Quel est le montant moyen de ventes de produits par région et par secteur de clientèle ? Vous pouvez découvrir que le montant moyen des ventes est légèrement plus élevé dans la région Ouest que dans les autres régions, avec des sociétés commerciales dans la région Ouest apportant le montant moyen de ventes le plus élevé.

Statistiques : Somme, nombre d'observations, moyenne, médiane, médiane groupée, erreur standard pour la moyenne, minimum, maximum, plage, valeur de la variable pour la première catégorie de la variable de regroupement, valeur de la variable pour la dernière catégorie de la variable de regroupement, écart type, variance, kurtosis, erreur standard de kurtosis, asymétrie, erreur standard d'asymétrie, pourcentage de la somme totale, pourcentage de *N* total, pourcentage de la somme dans, pourcentage de *N* dans, moyennes géométrique et harmonique.

Récapituler les commentaires de données

Données. Les variables de regroupement sont des variables catégorielles dont les valeurs peuvent être de type numérique ou chaîne. Le nombre de catégories doit être raisonnablement limité. Les autres variables doivent pouvoir être classées.

Hypothèses : Certains des sous-groupes statistiques optionnels, tels que la moyenne et l'écart type sont basés sur la théorie normale et conviennent aux variables quantitatives ayant une distribution symétrique. Les statistiques robustes telles que la médiane et la plage, conviennent aux variables quantitatives qui confirment ou infirment l'hypothèse de normalité.

Obtenir des récapitulatifs des observations

1. A partir des menus, sélectionnez :

   Analyse > **Rapports** > **Récapitulatif des observations...**

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

(En option) Vous avez également ces possibilités :

- Sélectionner au moins une variable de regroupement afin de diviser vos données en sous-groupes.
- Cliquer sur **Options** afin de modifier le titre de la sortie, ajouter une légende au-dessous de la sortie, ou exclure les observations ayant des valeurs manquantes.
- Cliquer sur **Statistiques** pour obtenir des statistiques facultatives.
- Sélectionner **Afficher les observations** afin de répertorier les observations dans chaque sous-groupe. Par défaut, le système ne liste que les 100 premières observations de votre fichier. Vous pouvez augmenter ou diminuer la valeur de l'option **Limiter les observations aux premières** *n* ou désélectionner cet élément pour répertorier toutes les observations.

Options de Récapituler

Récapituler vous permet de modifier le titre de votre sortie ou d'ajouter une légende qui apparaîtra en dessous du tableau de sortie. Vous pouvez contrôler les sauts de ligne dans les titres et légendes en tapant \n à tous les endroits où vous voulez insérer un retour à la ligne dans le texte.

Vous pouvez également choisir d'afficher ou de supprimer les sous-en-têtes des totaux et d'inclure ou d'exclure les observations ayant des valeurs manquantes pour toute variable prise en compte dans toute analyse. Il est souvent souhaitable de marquer les observations avec valeurs manquantes dans la sortie par un point ou un astérisque. Saisissez un caractère, une phrase, ou un code que vous souhaitez voir apparaître lorsqu'une valeur manque, sinon, aucun traitement spécial ne s'applique aux observations avec valeurs manquantes dans la sortie.

Statistiques de Récapituler

Vous pouvez choisir l'une des statistiques de sous-groupe suivantes pour les variables à l'intérieur de chaque catégorie de chacune des variables de regroupement : somme, nombre d'observations, moyenne, médiane, médiane groupée, erreur standard pour la moyenne, minimum, maximum, plage, valeur de la variable pour la première catégorie de la variable de regroupement, valeur de la variable pour la dernière catégorie de la variable de regroupement, écart type, variance, kurtosis, erreur standard de kurtosis, asymétrie, erreur standard d'asymétrie, pourcentage de la somme totale, pourcentage de *N* total, pourcentage de la somme dans, pourcentage de *N* dans, moyenne géométrique et moyenne harmonique. L'ordre dans lequel les statistiques apparaissent dans la liste Variables correspond à celui dans lequel elles seront affichées dans la sortie. Les statistiques récapitulatives sont aussi affichées pour chaque variable à travers toutes les catégories.

Premier. Affiche la première valeur de données rencontrée dans le fichier de données.

moyenne géométrique. Racine nième du produit des valeurs de données, n représentant le nombre d'observations.

médiane de groupes. Médiane calculée pour les données codées dans des groupes. Par exemple, pour les données d'âge, si chaque valeur de la trentaine est codée 35, chaque valeur de la quarantaine est codée 45, etc., la médiane de groupes est la médiane calculée à partir des données codées.

moyenne harmonique. Utilisé pour estimer la taille moyenne d'un groupe lorsque les tailles des échantillons dans les groupes ne sont pas égales. La moyenne harmonique correspond au nombre total d'échantillons divisé par la somme des inverses des tailles de l'échantillon.

Kurtosis. Mesure de l'étendue de la présence de valeurs extrêmes. Dans le cas d'une distribution normale, la valeur de la statistique kurtosis est égale à zéro. Un kurtosis positif indique qu'on observe dans les données plus de valeurs extrêmes que dans une distribution normale. Un kurtosis négatif indique que les données présentent moins de valeurs extrêmes par rapport à une distribution normale. La définition de kurtosis utilisée, où la valeur est 0 pour une distribution normale, est parfois appelée kurtosis excessif. Certains logiciels peuvent signaler un kurtosis de telle sorte que la valeur est 3 pour une distribution normale.

dernière. Affiche la première valeur de données rencontrée dans le fichier de données.

maximum. La plus grande valeur d'une variable numérique.

moyenne. Mesure de la tendance centrale. Moyenne arithmétique ; somme divisée par le nombre d'observations.

Valeur médiane. La valeur au-dessus et en dessous de laquelle se situent la moitié des cas, c'est-à-dire le 50e centile. Si le nombre d'observations est pair, la médiane correspond à la moyenne des deux observations du milieu lorsqu'elles sont triées dans l'ordre croissant ou décroissant. La médiane est une mesure de tendance centrale et elle n'est pas, à l'inverse de la moyenne, sensible aux valeurs éloignées.

Minimum. La plus petite valeur d'une variable numérique.

N. Nombre d'observations (observations ou enregistrements).

Pourcentage de N total. Pourcentage du nombre total d'observations dans chaque catégorie.

Pourcentage de la somme totale. Pourcentage de la somme totale dans chaque catégorie.

Intervalle. Différence entre la valeur maximale et la valeur minimale d'une variable numérique (maximum–minimum).

asymétrie . Mesure de l'asymétrie d'une distribution. La distribution normale est symétrique et a une valeur d'asymétrie égale à 0. Une distribution caractérisée par une importante asymétrie positive présente une partie droite plus allongée. Une distribution caractérisée par une importante asymétrie négative présente une extrémité gauche plus allongée. Pour simplifier, une valeur d'asymétrie deux fois supérieure à l'erreur standard correspond à une absence de symétrie.

Écart type. Mesure de la dispersion des valeurs autour de la moyenne. Dans le cas d'une distribution normale, 68 % des observations se situent à l'intérieur d'un écart type de la moyenne et 95 % se situent à l'intérieur de deux écarts types. Par exemple, si la moyenne d'âge est de 45 avec un écart type égal à 10, une distribution normale verra 95 % des observations se situer entre 25 et 65.

erreur standard de Kurtosis. Le rapport entre l'aplatissement et son erreur type peut être utilisé comme test de normalité (c'est-à-dire que vous pouvez rejeter la normalité si le rapport est inférieur à -2 ou supérieur à +2). Une valeur kurtosis positive importante indique que les extrémités de la distribution sont plus allongées que celles d'une distribution normale ; une valeur kurtosis négative présente des extrémités plus courtes (semblables à celles d'une distribution uniforme sous forme de boîtes).

erreur standard de la moyenne. Variation de la valeur de la moyenne d'un échantillon à l'autre issus d'une même distribution. Cette mesure permet de comparer approximativement la moyenne observée avec une valeur hypothétique (autrement dit, vous pouvez conclure que ces deux valeurs sont différentes si le rapport de la différence avec l'erreur standard est inférieur à -2 ou supérieur à +2).

erreur standard d'asymétrie. Le rapport entre l'asymétrie et son erreur standard peut être utilisé comme test de normalité (c'est-à-dire que vous pouvez rejeter la normalité si le rapport est inférieur à -2 ou supérieur à +2). Une valeur d'asymétrie positive importante indique une extrémité allongée vers la droite ; une valeur négative extrême produit une extrémité allongée vers la gauche.

somme. Somme ou total des valeurs, pour toutes les observations n'ayant pas de valeur manquante.

Variance. Mesure de la dispersion des valeurs autour de la moyenne, égale à la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d'observations moins un. La variance se mesure en unités, qui sont égales au carré des unités de la variable.

Moyennes

La procédure Moyennes calcule les moyennes de sous-groupes et les statistiques univariées correspondantes pour des variables dépendantes au sein des catégories d'une ou de plusieurs variables indépendantes. Vous pouvez également obtenir une analyse de variance à un facteur, un coefficient êta et des tests de linéarité.

Exemple. Mesurez la quantité moyenne de lipides absorbée par trois différents types d'huile alimentaire et effectuez une analyse de variance à un facteur pour voir si les moyennes divergent.

Statistiques : Somme, nombre d'observations, moyenne, médiane, médiane groupée, erreur standard pour la moyenne, minimum, maximum, plage, valeur de la variable pour la première catégorie de la variable de regroupement, valeur de la variable pour la dernière catégorie de la variable de regroupement, écart type, variance, kurtosis, erreur standard de kurtosis, asymétrie, erreur standard d'asymétrie, pourcentage de la somme totale, pourcentage de *N* total, pourcentage de la somme dans, pourcentage de *N* dans, moyennes géométrique et harmonique. Les options comportent une analyse de variance, un coefficient êta, un coefficient êta-carré, ainsi que des tests de linéarité *R* et R^2.

Remarques sur les données des moyennes

Données. Les variables dépendantes sont quantitatives et les variables indépendantes sont catégorielles. Les valeurs des variables catégorielles peuvent être soit numériques, soit des chaînes.

Hypothèses : Certains des sous-groupes statistiques optionnels, tels que la moyenne et l'écart type sont basés sur la théorie normale et conviennent aux variables quantitatives ayant une distribution symétrique. Les statistiques robustes, telles que la médiane, conviennent aux variables continues qui confirment ou infirment l'hypothèse de normalité. L'analyse de variance résiste aux écarts par rapport à la normalité, à condition que les données de chaque cellule soient symétriques. L'analyse de variance part également du principe que les groupes sont issus de populations ayant la même variance. Pour vérifier cette hypothèse, utilisez le test d'homogénéité de la variance de Levene, disponible dans la procédure ANOVA à un facteur.

Pour obtenir des moyennes de sous-groupes

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Moyennes...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins une variable dépendante.
3. Utilisez l'une des méthodes suivantes pour sélectionner des variables indépendantes catégorielles:
   - Sélectionnez une ou plusieurs variables indépendantes. Des résultats distincts sont présentés pour chaque variable indépendante.
   - Sélectionnez une ou plusieurs couches des variables indépendantes. Chaque couche divise une nouvelle fois l'échantillon. Si vous avez une variable indépendante dans la couche 1 et une dans la couche 2, les résultats sont présentés dans un tableau croisé, par opposition aux tableaux séparés pour chaque variable indépendante.
4. Cliquez sur **Options** pour obtenir des statistiques facultatives, telles que la table d'analyse de variance, êta, êta-carré, *R* et R^2.

Options de Moyennes

Vous pouvez choisir l'une des statistiques de sous-groupe suivantes pour les variables à l'intérieur de chaque catégorie de chacune des variables de regroupement : somme, nombre d'observations, moyenne, médiane, médiane groupée, erreur standard pour la moyenne, minimum, maximum, plage, valeur de

la variable pour la première catégorie de la variable de regroupement, valeur de la variable pour la dernière catégorie de la variable de regroupement, écart type, variance, kurtosis, erreur standard de kurtosis, asymétrie, erreur standard d'asymétrie, pourcentage de la somme totale, pourcentage de *N* total, pourcentage de la somme dans, pourcentage de *N* dans, moyenne géométrique et moyenne harmonique. Vous pouvez changer l'ordre de présentation des statistiques des sous-groupes. L'ordre dans lequel les statistiques apparaissent dans la liste Cellule Statistiques correspond à celui dans lequel elles seront affichées dans la sortie. Les statistiques récapitulatives sont aussi affichées pour chaque variable à travers toutes les catégories.

Premier. Affiche la première valeur de données rencontrée dans le fichier de données.

moyenne géométrique. Racine nième du produit des valeurs de données, n représentant le nombre d'observations.

médiane de groupes. Médiane calculée pour les données codées dans des groupes. Par exemple, pour les données d'âge, si chaque valeur de la trentaine est codée 35, chaque valeur de la quarantaine est codée 45, etc., la médiane de groupes est la médiane calculée à partir des données codées.

moyenne harmonique. Utilisé pour estimer la taille moyenne d'un groupe lorsque les tailles des échantillons dans les groupes ne sont pas égales. La moyenne harmonique correspond au nombre total d'échantillons divisé par la somme des inverses des tailles de l'échantillon.

Kurtosis. Mesure de l'étendue de la présence de valeurs extrêmes. Dans le cas d'une distribution normale, la valeur de la statistique kurtosis est égale à zéro. Un kurtosis positif indique qu'on observe dans les données plus de valeurs extrêmes que dans une distribution normale. Un kurtosis négatif indique que les données présentent moins de valeurs extrêmes par rapport à une distribution normale. La définition de kurtosis utilisée, où la valeur est 0 pour une distribution normale, est parfois appelée kurtosis excessif. Certains logiciels peuvent signaler un kurtosis de telle sorte que la valeur est 3 pour une distribution normale.

dernière. Affiche la première valeur de données rencontrée dans le fichier de données.

maximum. La plus grande valeur d'une variable numérique.

moyenne. Mesure de la tendance centrale. Moyenne arithmétique ; somme divisée par le nombre d'observations.

Valeur médiane. La valeur au-dessus et en dessous de laquelle se situent la moitié des cas, c'est-à-dire le 50e centile. Si le nombre d'observations est pair, la médiane correspond à la moyenne des deux observations du milieu lorsqu'elles sont triées dans l'ordre croissant ou décroissant. La médiane est une mesure de tendance centrale et elle n'est pas, à l'inverse de la moyenne, sensible aux valeurs éloignées.

Minimum. La plus petite valeur d'une variable numérique.

N. Nombre d'observations (observations ou enregistrements).

pourcentage de n total. Pourcentage du nombre total d'observations dans chaque catégorie.

pourcentage de la somme totale. Pourcentage de la somme totale dans chaque catégorie.

Intervalle. Différence entre la valeur maximale et la valeur minimale d'une variable numérique (maximum–minimum).

asymétrie . Mesure de l'asymétrie d'une distribution. La distribution normale est symétrique et a une valeur d'asymétrie égale à 0. Une distribution caractérisée par une importante asymétrie positive présente une partie droite plus allongée. Une distribution caractérisée par une importante asymétrie négative présente une extrémité gauche plus allongée. Pour simplifier, une valeur d'asymétrie deux fois supérieure à l'erreur standard correspond à une absence de symétrie.

Écart type. Mesure de la dispersion des valeurs autour de la moyenne. Dans le cas d'une distribution normale, 68 % des observations se situent à l'intérieur d'un écart type de la moyenne et 95 % se situent à l'intérieur de deux écarts types. Par exemple, si la moyenne d'âge est de 45 avec un écart type égal à 10, une distribution normale verra 95 % des observations se situer entre 25 et 65.

erreur standard de Kurtosis. Le rapport entre l'aplatissement et son erreur type peut être utilisé comme test de normalité (c'est-à-dire que vous pouvez rejeter la normalité si le rapport est inférieur à -2 ou

supérieur à +2). Une valeur kurtosis positive importante indique que les extrémités de la distribution sont plus allongées que celles d'une distribution normale ; une valeur kurtosis négative présente des extrémités plus courtes (semblables à celles d'une distribution uniforme sous forme de boîtes).

erreur standard de la moyenne. Variation de la valeur de la moyenne d'un échantillon à l'autre issus d'une même distribution. Cette mesure permet de comparer approximativement la moyenne observée avec une valeur hypothétique (autrement dit, vous pouvez conclure que ces deux valeurs sont différentes si le rapport de la différence avec l'erreur standard est inférieur à -2 ou supérieur à +2).

erreur standard d'asymétrie. Le rapport entre l'asymétrie et son erreur standard peut être utilisé comme test de normalité (c'est-à-dire que vous pouvez rejeter la normalité si le rapport est inférieur à -2 ou supérieur à +2). Une valeur d'asymétrie positive importante indique une extrémité allongée vers la droite ; une valeur négative extrême produit une extrémité allongée vers la gauche.

somme. Somme ou total des valeurs, pour toutes les observations n'ayant pas de valeur manquante.

Variance. Mesure de la dispersion des valeurs autour de la moyenne, égale à la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d'observations moins un. La variance se mesure en unités, qui sont égales au carré des unités de la variable.

Statistiques pour première couche

Table anova et êta. Affiche un tableau d'analyse unifactorielle de la variance et calcule êta et êta carré (mesures d'association) pour chaque variable indépendante de la première couche.

Test de linéarité. Calcule la somme des carrés, les degrés de liberté et le carré moyen associés aux composants linéaires et non linéaires, ainsi que le rapport F, le R et le R-deux. La linéarité n'est pas calculée si la variable indépendante est une chaîne courte.

Cubes OLAP

La procédure Cubes OLAP (Online Analytical Processing) calcule les totaux, les moyennes et autres statistiques univariées pour des variables récapitulatives continues à l'intérieur de catégories d'une ou de plusieurs variables de regroupement catégorielles. Une couche séparée dans le tableau est créée pour chaque catégorie de chaque variable de regroupement.

Exemple. Ventes totales et moyennes pour différentes régions et lignes de produits à l'intérieur de chaque région.

Statistiques : Somme, nombre d'observations, moyenne, médiane, médiane groupée, erreur standard pour la moyenne, minimum, maximum, plage, valeur de la variable pour la première catégorie de la variable de regroupement, valeur de la variable pour la dernière catégorie de la variable de regroupement, écart type, variance, kurtosis, erreur standard de kurtosis, asymétrie, erreur standard d'asymétrie, pourcentage des observations totales, pourcentage de somme totale, pourcentage des observations totales dans les variables de regroupement, pourcentage de la somme totale dans les variables de regroupement, moyenne géométrique et moyenne harmonique.

Remarques sur les données des cubes OLAP

Données. Les variables récapitulatives sont quantitatives (variables continues mesurées sur une échelle d'intervalle ou de rapport) et les variables de regroupement sont catégorielles. Les valeurs des variables catégorielles peuvent être soit numériques, soit des chaînes.

Hypothèses : Certains des sous-groupes statistiques optionnels, tels que la moyenne et l'écart type sont basés sur la théorie normale et conviennent aux variables quantitatives ayant une distribution symétrique. Les statistiques robustes telles que la médiane et la plage, conviennent aux variables quantitatives qui confirment ou infirment l'hypothèse de normalité.

Pour obtenir des cubes OLAP

1. A partir des menus, sélectionnez :

   Analyse > **Rapports** > **Cubes OLAP...**

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

2. Sélectionnez une ou plusieurs variables récapitulatives continues.
3. Sélectionnez une ou plusieurs variables de regroupement catégorielles.

Eventuellement :

- Sélectionnez d'autres statistiques récapitulatives (cliquez sur **Statistiques...**). Vous devez sélectionner un ou plusieurs critères de regroupement pour pouvoir sélectionner les statistiques récapitulatives.
- Calculez les différences entre des paires de variables et des paires de groupes définies par un critère de regroupement (cliquez sur **Différences**).
- Créez des titres de tableaux personnalisés (cliquez sur **Titre**).
- Masquez les effectifs inférieurs à un entier spécifique. Les valeurs masquées s'affichent sous la forme **<N**, où **N** est l'entier spécifié. L'entier spécifié doit être supérieur ou égal à 2.

Statistiques des Cubes OLAP

Vous pouvez choisir l'une des statistiques de sous-groupe suivantes pour les variables récapitulatives à l'intérieur de chaque catégorie de chacune des variables de regroupement : somme, nombre d'observations, moyenne, médiane, médiane groupée, erreur standard pour la moyenne, minimum, maximum, plage, valeur de la variable pour la première catégorie de la variable de regroupement, valeur de la variable pour la dernière catégorie de la variable de regroupement, écart type, variance, kurtosis, erreur standard de kurtosis, asymétrie, erreur standard d'asymétrie, pourcentage total des observations, pourcentage de la somme totale, pourcentage total des observations dans les variables de regroupement, pourcentage de la somme totale dans les variables de regroupement, moyenne géométrique et moyenne harmonique.

Vous pouvez changer l'ordre de présentation des statistiques des sous-groupes. L'ordre dans lequel les statistiques apparaissent dans la liste Cellule Statistiques correspond à celui dans lequel elles seront affichées dans la sortie. Les statistiques récapitulatives sont aussi affichées pour chaque variable à travers toutes les catégories.

Premier. Affiche la première valeur de données rencontrée dans le fichier de données.

moyenne géométrique. Racine nième du produit des valeurs de données, n représentant le nombre d'observations.

médiane de groupes. Médiane calculée pour les données codées dans des groupes. Par exemple, pour les données d'âge, si chaque valeur de la trentaine est codée 35, chaque valeur de la quarantaine est codée 45, etc., la médiane de groupes est la médiane calculée à partir des données codées.

moyenne harmonique. Utilisé pour estimer la taille moyenne d'un groupe lorsque les tailles des échantillons dans les groupes ne sont pas égales. La moyenne harmonique correspond au nombre total d'échantillons divisé par la somme des inverses des tailles de l'échantillon.

Kurtosis. Mesure de l'étendue de la présence de valeurs extrêmes. Dans le cas d'une distribution normale, la valeur de la statistique kurtosis est égale à zéro. Un kurtosis positif indique qu'on observe dans les données plus de valeurs extrêmes que dans une distribution normale. Un kurtosis négatif indique que les données présentent moins de valeurs extrêmes par rapport à une distribution normale. La définition de kurtosis utilisée, où la valeur est 0 pour une distribution normale, est parfois appelée kurtosis excessif. Certains logiciels peuvent signaler un kurtosis de telle sorte que la valeur est 3 pour une distribution normale.

dernière. Affiche la première valeur de données rencontrée dans le fichier de données.

maximum. La plus grande valeur d'une variable numérique.

moyenne. Mesure de la tendance centrale. Moyenne arithmétique ; somme divisée par le nombre d'observations.

Valeur médiane. La valeur au-dessus et en dessous de laquelle se situent la moitié des cas, c'est-à-dire le 50e centile. Si le nombre d'observations est pair, la médiane correspond à la moyenne des deux

observations du milieu lorsqu'elles sont triées dans l'ordre croissant ou décroissant. La médiane est une mesure de tendance centrale et elle n'est pas, à l'inverse de la moyenne, sensible aux valeurs éloignées.

Minimum. La plus petite valeur d'une variable numérique.

N. Nombre d'observations (observations ou enregistrements).

pourcentage de n dans. Pourcentage du nombre d'observations pour la variable de regroupement spécifiée dans les catégories des autres variables de regroupement. Si vous n'avez qu'un seul critère de regroupement, cette valeur est identique au pourcentage du nombre total d'observations.

Pourcentage de la somme dans. Pourcentage de la somme de la variable de regroupement spécifiée dans les catégories des autres variables de regroupement. Si vous n'avez qu'un seul critère de regroupement, cette valeur est identique au pourcentage de la somme totale.

Pourcentage de N total. Pourcentage du nombre total d'observations dans chaque catégorie.

Pourcentage de la somme totale. Pourcentage de la somme totale dans chaque catégorie.

Intervalle. Différence entre la valeur maximale et la valeur minimale d'une variable numérique (maximum–minimum).

asymétrie . Mesure de l'asymétrie d'une distribution. La distribution normale est symétrique et a une valeur d'asymétrie égale à 0. Une distribution caractérisée par une importante asymétrie positive présente une partie droite plus allongée. Une distribution caractérisée par une importante asymétrie négative présente une extrémité gauche plus allongée. Pour simplifier, une valeur d'asymétrie deux fois supérieure à l'erreur standard correspond à une absence de symétrie.

Écart type. Mesure de la dispersion des valeurs autour de la moyenne. Dans le cas d'une distribution normale, 68 % des observations se situent à l'intérieur d'un écart type de la moyenne et 95 % se situent à l'intérieur de deux écarts types. Par exemple, si la moyenne d'âge est de 45 avec un écart type égal à 10, une distribution normale verra 95 % des observations se situer entre 25 et 65.

erreur standard de Kurtosis. Le rapport entre l'aplatissement et son erreur type peut être utilisé comme test de normalité (c'est-à-dire que vous pouvez rejeter la normalité si le rapport est inférieur à -2 ou supérieur à +2). Une valeur kurtosis positive importante indique que les extrémités de la distribution sont plus allongées que celles d'une distribution normale ; une valeur kurtosis négative présente des extrémités plus courtes (semblables à celles d'une distribution uniforme sous forme de boîtes).

erreur standard de la moyenne. Variation de la valeur de la moyenne d'un échantillon à l'autre issus d'une même distribution. Cette mesure permet de comparer approximativement la moyenne observée avec une valeur hypothétique (autrement dit, vous pouvez conclure que ces deux valeurs sont différentes si le rapport de la différence avec l'erreur standard est inférieur à -2 ou supérieur à +2).

erreur standard d'asymétrie. Le rapport entre l'asymétrie et son erreur standard peut être utilisé comme test de normalité (c'est-à-dire que vous pouvez rejeter la normalité si le rapport est inférieur à -2 ou supérieur à +2). Une valeur d'asymétrie positive importante indique une extrémité allongée vers la droite ; une valeur négative extrême produit une extrémité allongée vers la gauche.

somme. Somme ou total des valeurs, pour toutes les observations n'ayant pas de valeur manquante.

Variance. Mesure de la dispersion des valeurs autour de la moyenne, égale à la somme des carrés des écarts par rapport à la moyenne, divisée par le nombre d'observations moins un. La variance se mesure en unités, qui sont égales au carré des unités de la variable.

Différences des Cubes OLAP

Cette boîte de dialogue vous permet de calculer les différences arithmétiques et de pourcentage qui existent entre des variables récapitulatives ou entre des groupes définis par un critère de regroupement. Les différences sont calculées pour toutes les mesures sélectionnées dans la boîte de dialogue Cubes OLAP : Statistiques.

Différences entre les variables : Calcule les différences existant entre des paires de variables. Les valeurs des statistiques récapitulatives de la seconde variable (variable moins) de chaque paire sont soustraites des valeurs des statistiques récapitulatives de la première variable de la paire. Pour les

différences de pourcentage, la valeur de la caractéristique de la variable moins est utilisée en tant que dénominateur. Vous devez sélectionner plusieurs variables récapitulatives dans la boîte de dialogue principale avant d'indiquer les différences entre les variables.

Différences entre les groupes d'observations : Calcule les différences existant entre les paires de groupes définies par une variable de regroupement. Les valeurs des statistiques récapitulatives de la seconde catégorie (catégorie moins) dans chaque paire sont soustraites des valeurs des statistiques récapitulatives de la première catégorie de la paire. Les différences de pourcentage utilisent la valeur de la statistique récapitulative pour la catégorie moins en tant que dénominateur. Vous devez sélectionner au moins une variable de regroupement dans la boîte de dialogue principale avant d'indiquer les différences entre groupes.

Titre des Cubes OLAP

Il vous est possible de modifier le titre de votre sortie ou d'ajouter une légende qui apparaîtra au dessous du tableau de sortie. Vous pouvez également contrôler la répartition des titres et légendes sur plusieurs lignes en tapant \n partout où vous souhaitez insérer un retour à la ligne dans le texte.

Proportions

Introduction aux proportions

Les proportions calculent des tests et des intervalles de confiance pour des proportions binomiales ou des différences de proportions. Des statistiques sont disponibles pour les proportions d'échantillons uniques (testées par rapport à une valeur spécifiée), les échantillons appariés (variables différentes) ou les échantillons indépendants (différents groupes d'observations). Diverses options sont offertes pour les statistiques de test et les intervalles de confiance. Parmi les autres procédures qui fournissent certaines des mêmes fonctionnalités figurent **CROSSTABS**, **NPAR TESTS** et **NPTESTS**.

Proportions d'échantillons uniques
Tests pour échantillon unique et intervalles de confiance. La sortie comporte la proportion observée, l'estimation de la différence entre la proportion de population et la proportion de population supposée, les erreurs standard asymptotiques sous les hypothèses nulles et alternatives, les statistiques de test spécifiées avec des probabilités bilatérales et les intervalles de confiance spécifiés pour les proportions.

Proportions d'échantillons appariés
Tests à échantillons appariés et intervalles de confiance pour les différences entre les proportions. La sortie comporte les proportions observées, les estimations des différences entre les proportions de population, les erreurs standard asymptotiques des différences de population sous les hypothèses nulles et les hypothèses alternatives, les statistiques de test spécifiées avec des probabilités bilatérales et des intervalles de confiance spécifiés pour les différences entre les proportions.

Proportions d'échantillons indépendants
Tests pour échantillons indépendants et intervalles de confiance. La sortie comporte les proportions observées, les estimations des différences entre les proportions de population, les erreurs standard asymptotiques des différences de population sous les hypothèses nulles et les hypothèses alternatives, les statistiques de test spécifiées avec des probabilités bilatérales et des intervalles de confiance spécifiés pour les différences entre les proportions.

Proportions d'échantillons uniques

La procédure Proportions d'échantillons uniques propose des tests et des intervalles de confiance pour les proportions binomiales individuelles. Les données sont supposées provenir d'une échantillon aléatoire simple, et chaque test d'hypothèse ou intervalle de confiance est un test distinct ou un intervalle de confiance individuel, basé sur une proportion binomiale. La sortie comporte la proportion observée, l'estimation de la différence entre la proportion de population et la proportion de population supposée, les erreurs standard asymptotiques sous les hypothèses nulles et alternatives, les statistiques de test spécifiées avec des probabilités bilatérales et les intervalles de confiance spécifiés pour les proportions.

Exemple

Statistiques
Agresti-Coull, Anscombe, Clopper-Pearson (Exact), Jeffreys, Logit, Wald, Wald (continuité corrigée), Score de Wilson, Score de Wilson (continuité corrigée), Binomial exact, Binomial ajusté p moyen, Score, Score (continuité corrigée).

Remarques sur les données

Données
La procédure affiche les statistiques de test demandées et les probabilités bilatérales, les intervalles de confiance pour les différences des proportions, ainsi que les proportions, les erreurs standard et les effectifs pour chaque groupe ou variable. La procédure est limitée au maximum d'une valeur de test.

Hypothèses

Obtention de tests de proportions à un échantillon

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons uniques...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins une variable test quantitative.
3. (En option) Vous avez également ces possibilités :
   - Sélectionnez les critères de succès sous la section **Définir le succès** :

     Dernière valeur
     La dernière valeur ou la valeur la plus élevée, parmi les valeurs distinctes triées dans les données, est utilisée. Ce paramètre s'applique aux variables numériques ou aux variables de chaîne. Il s'agit du paramètre par défaut.

     Première valeur
     La première valeur ou la valeur la plus faible, parmi les valeurs distinctes triées dans les données, est utilisée. Ce paramètre s'applique aux variables numériques ou aux variables de chaîne.

     Valeur(s)
     Une ou plusieurs valeurs spécifiques entre parenthèses. Les valeurs multiples doivent être séparées par un espace. Ce paramètre s'applique aux variables numériques ou aux variables de chaîne. Les valeurs de variable de chaîne doivent être placées entre guillemets simples.

     Centre
     Valeurs au milieu ou légèrement au-dessus de la plage des valeurs observées dans les données. Ce paramètre s'applique uniquement aux données numériques.

     Césure
     Valeurs supérieures ou égales à une valeur spécifiée. Ce paramètre s'applique uniquement aux données numériques.
   - Cliquez sur **Intervalles de confiance...** pour spécifier les types d'intervalles de confiance à afficher ou pour supprimer tous les intervalles de confiance.
   - Cliquez sur **Tests...** pour spécifier les types de statistiques de test à afficher ou pour supprimer tous les tests.
   - Cliquez sur **Valeurs manquantes...** pour contrôler le traitement des données manquantes.
   - Cliquez sur **Bootstrap...** pour obtenir des estimations robustes des erreurs types et des intervalles de confiance pour des estimations telles que la moyenne, la médiane, la proportion, le rapport de cotes, le coefficient de corrélation ou le coefficient de régression.
4. Cliquez sur **OK**.

Proportions d'échantillons uniques : Intervalles de confiance

La boîte de dialogue Intervalles de confiance propose des options permettant de définir le niveau de couverture et de sélectionner les types d'intervalles de confiance à afficher.

Niveau de couverture
Indique le pourcentage d'intervalle de confiance. Une valeur numérique comprise entre 0 et 100 doit être spécifiée. La valeur par défaut est 95.

Type(s) d'intervalle
Propose des options permettant de spécifier les types d'intervalles de confiance à afficher. Les options disponibles sont les suivantes :

- Agresti-Coull
- Anscombe
- Clopper-Pearson (Exact)
- Jeffreys
- Logit
- Wald
- Wald (continuité corrigée)
- Score de Wilson
- Score de Wilson (continuité corrigée)

Spécification des intervalles de confiance pour les proportions d'échantillons uniques

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons uniques...**
2. Cliquez sur **Intervalles de confiance** pour spécifier les types d'intervalle de confiance à afficher, ou pour supprimer tous les intervalles de confiance.

Proportions à un échantillon : Tests

La boîte de dialogue Tests propose des options permettant de spécifier les types de statistiques de test à afficher.

Tous
Toutes les statistiques de test sont affichées dans la sortie.

Aucun
Aucune statistique de test n'est affichée dans la sortie.

Binominal exact
Affiche les probabilités binomiales exactes.

Binomial ajusté p moyen
Affiche les probabilités binomiales ajustées p moyen. Il s'agit du paramètre par défaut.

Indice
Affiche les statistiques de test du score Z. Il s'agit du paramètre par défaut.

Score (continuité corrigée)
Affiche les statistiques de test du score Z avec correction de continuité.

Wald
Affiche les statistiques de test Z de Wald.

Wald (continuité corrigée)
Affiche les statistiques de test Z de Wald avec correction de continuité.

Valeur de test
Indique une valeur de test comprise entre 0 et 1. La valeur par défaut est 0,5.

Obtention de tests de proportions à un échantillon

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons uniques...**
2. Dans la boîte de dialogue Proportions d'échantillons uniques, cliquez sur **Tests**.
3. Sélectionnez un ou plusieurs des tests disponibles.

Proportions d'échantillons uniques : Valeurs manquantes

La boîte de dialogue Valeurs manquantes propose des options permettant de traiter les valeurs manquantes.

La portée des données est manquante

Exclure les observations analyse par analyse
Inclut toutes les observations ayant suffisamment de données sur les variables utilisées dans chaque analyse particulière. Il s'agit du paramètre par défaut.

Exclure toute observation incomplète
Inclut toutes les observations ayant suffisamment de données sur toutes les variables utilisées dans toutes les analyses.

Valeurs utilisateur manquantes
L'option **Exclure** considère les valeurs utilisateur manquantes comme étant manquantes. L'option **Inclure** ignore les désignations de valeurs utilisateur manquantes et considère ces dernières comme étant valides.

Définition des paramètres de valeurs manquantes pour les proportions d'échantillons uniques

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons uniques...**
2. Dans la boîte de dialogue Proportions d'échantillons uniques, cliquez sur **Valeurs manquantes**.
3. Sélectionnez les paramètres souhaités pour les valeurs manquantes.

Proportions d'échantillons appariés

La procédure Proportions d'échantillons appariés propose des tests et des intervalles de confiance pour la différence entre deux proportions binomiales appariées ou liées. Les données sont supposées provenir d'un échantillon aléatoire simple, et chaque test d'hypothèse ou intervalle de confiance est un test distinct ou un intervalle individuel. La sortie comporte les proportions observées, les estimations des différences entre les proportions de population, les erreurs standard asymptotiques des différences de population sous les hypothèses nulles et les hypothèses alternatives, les statistiques de test spécifiées avec des probabilités bilatérales et des intervalles de confiance spécifiés pour les différences entre les proportions.

Exemple

Statistiques
Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuité corrigée), Binomial exact, Binomial ajusté p moyen, McNemar, McNemar (continuité corrigée).

Analyse des données

Données

- Une liste de variables contenant au moins deux variables est requise.
- Si une liste unique de variables est spécifiée, chaque membre de la liste est apparié avec chaque autre membre de la liste.

Hypothèses

- Si deux listes de variables sont séparées par **WITH** sans le mot-clé (**PAIRED**), chaque membre de la première liste est apparié avec chaque membre de l'autre liste.
- Si deux listes de variables sont séparées par **WITH** et que la seconde liste est suivie par (**PAIRED**), les membres des deux listes sont appariés dans l'ordre : le premier membre de la première liste est apparié avec le premier membre de la deuxième liste, les deuxièmes membres de chaque liste sont appariés, etc. Les variables sans correspondance sont ignorées et un message d'avertissement est émis.

Obtention des tests de proportions à échantillons appariés

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons appariés...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins une variable test quantitative.
3. (En option) Vous avez également ces possibilités :
   - Sélectionnez les critères de succès sous la section **Définir le succès** :

     Dernière valeur
     La dernière valeur ou la valeur la plus élevée, parmi les valeurs distinctes triées dans les données, est utilisée. Ce paramètre s'applique aux variables numériques ou aux variables de chaîne. Il s'agit du paramètre par défaut.

     Première valeur
     La première valeur ou la valeur la plus faible, parmi les valeurs distinctes triées dans les données, est utilisée. Ce paramètre s'applique aux variables numériques ou aux variables de chaîne.

     Valeur(s)
     Une ou plusieurs valeurs spécifiques entre parenthèses. Les valeurs multiples doivent être séparées par un espace. Ce paramètre s'applique aux variables numériques ou aux variables de chaîne. Les valeurs de variable de chaîne doivent être placées entre guillemets simples.

     Centre
     Valeurs au milieu ou légèrement au-dessus de la plage des valeurs observées dans les données. Ce paramètre s'applique uniquement aux données numériques.

     Césure
     Valeurs supérieures ou égales à une valeur spécifiée. Ce paramètre s'applique uniquement aux données numériques.
   - Cliquez sur **Intervalles de confiance...** pour spécifier les types d'intervalles de confiance à afficher ou pour supprimer tous les intervalles de confiance.
   - Cliquez sur **Tests...** pour spécifier les types de statistiques de test à afficher ou pour supprimer tous les tests.
   - Cliquez sur **Valeurs manquantes...** pour contrôler le traitement des données manquantes.
   - Cliquez sur **Bootstrap...** pour obtenir des estimations robustes des erreurs types et des intervalles de confiance pour des estimations telles que la moyenne, la médiane, la proportion, le rapport de cotes, le coefficient de corrélation ou le coefficient de régression.
4. Cliquez sur **OK**.

Proportions d'échantillons appariés : Intervalles de confiance

La boîte de dialogue Intervalles de confiance propose des options permettant de définir le niveau de couverture et de sélectionner les types d'intervalles de confiance à afficher.

Niveau de couverture
Indique le pourcentage d'intervalle de confiance. Une valeur numérique comprise entre 0 et 100 doit être spécifiée. La valeur par défaut est 95.

Type(s) d'intervalle
Propose des options permettant de spécifier les types d'intervalles de confiance à afficher. Les options disponibles sont les suivantes :

- Agresti-Min
- Bonett-Price
- Newcombe
- Wald
- Wald (continuité corrigée)

Spécification des intervalles de confiance pour les proportions d'échantillons appariés

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons appariés...**
2. Cliquez sur **Intervalles de confiance** pour spécifier les types d'intervalle de confiance à afficher, ou pour supprimer tous les intervalles de confiance.

Proportions d'échantillons appariés : Tests

La boîte de dialogue Tests propose des options permettant de spécifier les types de statistiques de test à afficher.

Tous
Toutes les statistiques de test sont affichées dans la sortie.

Aucun
Aucune statistique de test n'est affichée dans la sortie.

Binominal exact
Affiche les probabilités binomiales exactes.

Binomial ajusté p moyen
Affiche les probabilités binomiales ajustées p moyen. Il s'agit du paramètre par défaut.

McNemar
Affiche les statistiques de test Z de McNemar. Il s'agit du paramètre par défaut.

McNemar (continuité corrigée)
Affiche les statistiques de test Z de McNemar avec correction de continuité.

Wald
Affiche les statistiques de test Z de Wald.

Wald (continuité corrigée)
Affiche les statistiques de test Z de Wald avec correction de continuité.

Obtention des tests de proportions à échantillons appariés

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons appariés...**
2. Dans la boîte de dialogue Proportions d'échantillons appariés, cliquez sur **Tests**.
3. Sélectionnez un ou plusieurs des tests disponibles.

Proportions d'échantillons appariés : Valeurs manquantes

La boîte de dialogue Valeurs manquantes propose des options permettant de traiter les valeurs manquantes.

La portée des données est manquante

Exclure les observations analyse par analyse
Inclut toutes les observations ayant suffisamment de données sur les variables utilisées dans chaque analyse particulière. Il s'agit du paramètre par défaut.

Exclure toute observation incomplète
Inclut toutes les observations ayant suffisamment de données sur toutes les variables utilisées dans toutes les analyses.

Valeurs utilisateur manquantes
L'option **Exclure** considère les valeurs utilisateur manquantes comme étant manquantes. L'option **Inclure** ignore les désignations de valeurs utilisateur manquantes et considère ces dernières comme étant valides.

Définition des paramètres de valeurs manquantes pour les proportions d'échantillons appariés

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons appariés...**
2. Dans la boîte de dialogue Proportions d'échantillons appariés, cliquez sur **Valeurs manquantes**.
3. Sélectionnez les paramètres souhaités pour les valeurs manquantes.

Proportions d'échantillons indépendants

La procédure Proportions d'échantillons indépendants propose des tests et des intervalles de confiance pour la différence entre deux proportions binomiales indépendantes. Les données sont supposées provenir d'un échantillon aléatoire simple, et chaque test d'hypothèse ou intervalle de confiance est un test distinct ou un intervalle individuel. La sortie comporte les proportions observées, les estimations des différences entre les proportions de population, les erreurs standard asymptotiques des différences de population sous les hypothèses nulles et les hypothèses alternatives, les statistiques de test spécifiées avec des probabilités bilatérales et des intervalles de confiance spécifiés pour les différences entre les proportions.

Exemple

Statistiques
Agresti-Min, Bonett-Price, Newcombe, Wald, Wald (continuité corrigée), Binomial exact, Binomial ajusté p moyen, McNemar, McNemar (continuité corrigée).

Analyse des données

Données

- Au moins une variable dépendante et une variable unique sont requises pour identifier les deux groupes à comparer.
- La variable de regroupement peut être une variable numérique ou une variable de chaîne.

Hypothèses

Obtention des tests de proportions à échantillons indépendants

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons indépendants...**

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

2. Sélectionnez au moins une variable test quantitative.
3. Sélectionnez une **Variable de regroupement** unique permettant d'identifier les deux groupes à comparer.
4. Vous pouvez également spécifier les paramètres de la **Variable de regroupement** sélectionnée.
   - Lorsque l'option **Valeur(s)** est sélectionnée, vous pouvez indiquer entre parenthèses deux valeurs à comparer, de type numérique ou chaîne. Les valeurs de chaîne doivent être placées entre guillemets simples. Les observations comportant d'autres valeurs sont ignorées.
   - Le paramètre **Centre** s'applique uniquement aux variables numériques. Les observations supérieures ou égales au centre de la distribution des variables de regroupement sont affectées au second groupe, tandis que celles situées en dessous sont affectées au premier groupe.
   - Le paramètre **Césure** s'applique uniquement aux variables numériques et permet d'indiquer une valeur numérique unique entre parenthèses. Les observations dont les valeurs sont supérieures ou égales à la césure sur la variable de regroupement sont affectées au second groupe, tandis que celles inférieures à la césure sont affectées au premier groupe.
5. (En option) Vous avez également ces possibilités :
   - Sélectionnez les critères de succès sous la section **Définir le succès** :

     Dernière valeur
     La dernière valeur ou la valeur la plus élevée, parmi les valeurs distinctes triées dans les données, est utilisée. Ce paramètre s'applique aux variables numériques ou aux variables de chaîne. Il s'agit du paramètre par défaut.

     Première valeur
     La première valeur ou la valeur la plus faible, parmi les valeurs distinctes triées dans les données, est utilisée. Ce paramètre s'applique aux variables numériques ou aux variables de chaîne.

     Valeur(s)
     Une ou plusieurs valeurs spécifiques entre parenthèses. Les valeurs multiples doivent être séparées par un espace. Ce paramètre s'applique aux variables numériques ou aux variables de chaîne. Les valeurs de variable de chaîne doivent être placées entre guillemets simples.

     Centre
     Valeurs au milieu ou légèrement au-dessus de la plage des valeurs observées dans les données. Ce paramètre s'applique uniquement aux données numériques.

     Césure
     Valeurs supérieures ou égales à une valeur spécifiée. Ce paramètre s'applique uniquement aux données numériques.
   - Cliquez sur **Intervalles de confiance...** pour spécifier les types d'intervalles de confiance à afficher ou pour supprimer tous les intervalles de confiance.
   - Cliquez sur **Tests...** pour spécifier les types de statistiques de test à afficher ou pour supprimer tous les tests.
   - Cliquez sur **Valeurs manquantes...** pour contrôler le traitement des données manquantes.
   - Cliquez sur **Bootstrap...** pour obtenir des estimations robustes des erreurs types et des intervalles de confiance pour des estimations telles que la moyenne, la médiane, la proportion, le rapport de cotes, le coefficient de corrélation ou le coefficient de régression.
6. Cliquez sur **OK**.

Proportions d'échantillons indépendants : Intervalles de confiance

La boîte de dialogue Intervalles de confiance propose des options permettant de définir le niveau de couverture et de sélectionner les types d'intervalles de confiance à afficher.

Niveau de couverture
Indique le pourcentage d'intervalle de confiance. Une valeur numérique comprise entre 0 et 100 doit être spécifiée. La valeur par défaut est 95.

Type(s) d'intervalle
Propose des options permettant de spécifier les types d'intervalles de confiance à afficher. Les options disponibles sont les suivantes :

- Agresti-Caffo
- Brown-Li-Jeffreys
- Hauck-Anderson
- Newcombe
- Newcombe (continuité corrigée)
- Wald
- Wald (continuité corrigée)

Spécification des intervalles de confiance pour les proportions d'échantillons indépendants

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons indépendants...**
2. Cliquez sur **Intervalles de confiance** pour spécifier les types d'intervalle de confiance à afficher, ou pour supprimer tous les intervalles de confiance.

Proportions d'échantillons indépendants : Tests

La boîte de dialogue Tests propose des options permettant de spécifier les types de statistiques de test à afficher.

Tous
Toutes les statistiques de test sont affichées dans la sortie.

Aucun
Aucune statistique de test n'est affichée dans la sortie.

Hauck-Anderson
Affiche les statistiques de test Z de Hauck-Anderson.

Wald
Affiche les statistiques de test Z de Wald.

Wald (continuité corrigée)
Affiche les statistiques de test Z de Wald avec correction de continuité.

Wald H0
Affiche les statistiques de test Z de Wald à l'aide des estimations de variance sous $_0$.

Wald H0 (continuité corrigée)
Affiche les statistiques de test Z de Wald avec correction de continuité à l'aide des estimations de variance sous H_0.

Obtention des tests de proportions à échantillons indépendants

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons indépendants...**
2. Dans la boîte de dialogue Proportions d'échantillons indépendants, cliquez sur **Tests**.
3. Sélectionnez un ou plusieurs des tests disponibles.

Proportions d'échantillons indépendants : Valeurs manquantes

La boîte de dialogue Valeurs manquantes propose des options permettant de traiter les valeurs manquantes.

La portée des données est manquante

Exclure les observations analyse par analyse
Inclut toutes les observations ayant suffisamment de données sur les variables utilisées dans chaque analyse particulière. Il s'agit du paramètre par défaut.

Exclure toute observation incomplète
Inclut toutes les observations ayant suffisamment de données sur toutes les variables utilisées dans toutes les analyses.

Valeurs utilisateur manquantes
L'option **Exclure** considère les valeurs utilisateur manquantes comme étant manquantes. L'option **Inclure** ignore les désignations de valeurs utilisateur manquantes et considère ces dernières comme étant valides.

Définition des paramètres de valeurs manquantes pour les proportions d'échantillons indépendants

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Proportions d'échantillons indépendants...**
2. Dans la boîte de dialogue Proportions d'échantillons indépendants, cliquez sur **Valeurs manquantes**.
3. Sélectionnez les paramètres souhaités pour les valeurs manquantes.

Tests T

Tests T

Il existe trois types de test *t* :

Test T pour échantillons indépendants (test t pour deux échantillons) : Permet de comparer la moyenne d'une variable de deux groupes d'observations. Les statistiques descriptives pour chaque groupe et le test de Levene permettant d'obtenir l'égalité des variances sont disponibles ainsi que les valeurs *t* de variance égale et inégale, et qu'un intervalle de confiance de 95 % pour la différence des moyennes.

Test T pour échantillons appariés (test t dépendant) : Permet de comparer la moyenne de deux variables pour un seul groupe. Ce test sert aussi pour les plans d'études appariées ou de contrôle d'observation. Les sorties incluent les statistiques descriptives pour les variables de test, leurs corrélations, les statistiques descriptives pour les différences appariées, le test *t* et un intervalle de confiance de 95 %.

Test T pour échantillon unique : Permet de comparer la moyenne d'une variable avec une valeur connue ou supposée. Les statistiques descriptives des variables tests sont affichées avec le test *t*. Un intervalle de confiance de 95 % pour la différence entre la moyenne de la variable test et la valeur test supposée fait partie de la sortie par défaut.

Test T pour échantillons indépendants

La procédure Test T sur échantillons indépendants compare la moyenne de deux groupes d'observations et automatise le calcul de la taille d'effet de test *t*. Idéalement, pour ce test, les sujets devraient être répartis de manière aléatoire en deux groupes afin que toute différence de réponse soit due au traitement (ou à l'absence de traitement) et non à d'autres facteurs. Ceci n'est pas le cas si l'on compare un revenu moyen pour les hommes et les femmes. Une personne n'est pas aléatoirement désignée comme devant être un homme ou une femme. Dans de telles situations, il faut s'assurer que les différences dans les autres facteurs ne cachent pas ou n'augmentent de différence significative dans les moyennes. Les

différences de revenu moyen peuvent être influencées par des facteurs tels que l'éducation et non par le sexe seul.

Exemple

Les patients souffrant d'hypertension se voient assignés de façon aléatoire à un groupe placebo et à un groupe auquel on donne un traitement. Les sujets du groupe placebo reçoivent une pilule inactive et les sujets du groupe auquel on donne un traitement reçoivent un nouveau médicament supposé réduire l'hypertension. Après que les sujets ont suivi le traitement pendant deux mois, le test *t* pour deux échantillons est utilisé pour comparer la tension artérielle moyenne du groupe placebo à celle du groupe qui suit le traitement. Chaque patient est examiné une fois et appartient à un groupe.

Statistiques

Pour chaque variable : taille d'échantillon, moyenne, écart type, erreur standard de la moyenne et estimation de la taille d'effet pour test *t*. Pour la différence de moyennes : moyenne, erreur standard et niveau de confiance (vous pouvez préciser le niveau de confiance). Tests : test de Levene sur l'égalité des variances et tests *t* des variances regroupées en pool et séparées pour l'égalité des moyennes.

Analyse des données

Données

Les valeurs de la variable quantitative qui vous intéresse se trouvent dans une seule colonne du fichier de données. La procédure utilise une variable de regroupement à deux valeurs pour séparer les observations en deux groupes. Le critère de regroupement peut être numérique (on peut avoir des valeurs telles que 1 et 2, ou 6,25 et 12,5) ou alphanumérique (telles que *oui* et *non*). Vous pouvez utiliser également une variable quantitative, telle que l'*âge*, pour séparer les observations en deux groupes en précisant une césure (la césure 21 provoque un groupe dont l'*âge* est inférieur à 21 ans et un groupe dont l'âge est supérieur à 21 ans).

Hypothèses

Pour le test *t* de variance égale, les observations doivent être indépendantes, et les échantillons aléatoires de distribution normale doivent avoir la même variance de population. Pour le test *t* de variance égale, les observations doivent être indépendantes, et les échantillons aléatoires doivent avoir une distribution normale. Le test *t* pour deux échantillons est assez robuste pour se départir de la normalité. Lors de la vérification graphique des distributions, vérifiez qu'elles sont symétriques et n'ont pas de valeurs extrêmes.

Obtention d'un test T pour échantillons indépendants

1. A partir des menus, sélectionnez :

   Analyse > **Comparer les moyennes** > **Test T pour échantillons indépendants...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins une variable test quantitative. Un test *t* distinct est alors calculé pour chaque variable.
3. Sélectionnez un seul critère de regroupement et cliquez sur **Définir groupes** pour spécifier deux codes pour les groupes à comparer.
4. Les paramètres suivants sont facultatifs.
   - Sélectionnez l'option **Estimer les tailles d'effet** pour contrôler l'estimation de la taille d'effet de test *t*.
   - Sélectionnez le **test d'homogénéité de la variance** pour calculer la statistique de Levene afin de tester l'égalité des variances des groupes. Vous pouvez obtenir les résultats dans un tableau séparé.
   - Cliquer sur **Options** pour contrôler le traitement des données manquantes et le niveau de l'intervalle de confiance.

- Cliquer sur **Bootstrap** pour dériver des estimations robustes des erreurs standard et des intervalles de confiance pour des estimations telles que la moyenne, la médiane, le calcul de la proportion, le rapport des cotes, le coefficient de corrélation ou de régression.

Définition de Groupes pour test T avec échantillons indépendants

Pour les variables de regroupement numérique, définissez les deux groupes du test *t* en spécifiant deux valeurs ou une césure :

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

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

Options de Test T pour échantillons indépendants

Intervalle de confiance. Par défaut, un intervalle de confiance de 95 % pour la différence dans les moyennes est affiché. Saisir une valeur comprise entre 1 et 99 pour demander un niveau de confiance différent.

Valeurs manquantes : Quand vous testez plusieurs variables et que des données sont manquantes pour au moins une variable, vous pouvez indiquer à la procédure les observations à inclure (ou exclure).

- **Exclure les observations analyse par analyse :** Chaque test *t* utilise toutes les observations qui ont des données valides pour les variables testées. La taille des échantillons peut varier d'un test à l'autre.
- **Exclure toute observation incomplète :** Chaque test *t* utilise seulement les observations qui ont des données valides pour toutes les variables utilisées dans les tests *t* requis. La taille des échantillons est constante durant les tests.

Test T pour échantillons appariés

La procédure Test T pour échantillons appariés compare la moyenne de deux variables pour un seul groupe. Elle permet de calculer les différences entre les valeurs des deux variables pour chaque observation et de tester si la moyenne diffère de 0. La procédure automatise également le calcul de la taille de l'effet de test *T*.

Exemple

Dans le cadre d'une étude sur l'hypertension, des mesures sont prises sur tous les patients au début de l'étude, un traitement est administré, puis on procède à une nouvelle mesure. Par conséquent, chaque sujet est l'objet de deux mesures, souvent nommées mesures *avant* et *après*. Il existe une alternative à ce test, il s'agit d'une étude appariée ou de contrôle d'observation dans laquelle chaque déclaration dans le fichier de données contient la réponse du patient ainsi que celle de son sujet de contrôle apparié. Dans le cadre d'une étude sur la tension artérielle, les patients et les contrôles peuvent être appariés selon l'âge (un patient âgé de 75 ans avec un membre du groupe de contrôle âgé de 75 ans).

Statistiques

Pour chaque variable : moyenne, taille d'échantillon, écart type et erreur standard de la moyenne. Pour chaque paire de variables : corrélation, différence moyenne de moyennes, test *t*, intervalle de confiance pour la différence moyenne (vous pouvez préciser le niveau de confiance) et estimation de la taille d'effet pour test *t*. Ecart type et erreur standard de la différence moyenne.

Remarques sur les données

Données

Pour chaque test apparié, précisez deux variables continues (niveau d'intervalle de mesure ou niveau de rapport de mesure). Dans le cadre d'une étude appariée ou de contrôle d'observation, la réponse pour chaque sujet test et son sujet de contrôle apparié doit être dans la même observation du fichier de données.

Hypothèses

Les observations pour chaque paire devraient être réalisées dans les mêmes conditions. Les différences moyennes devraient suivre une distribution normale. Les variances de chaque variable peuvent être égales ou inégales.

Obtention d'un test T pour échantillons appariés

1. A partir des menus, sélectionnez :

   Analyse > **>Comparer les moyennes** > **Test T pour échantillons appariés...**

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

2. Sélectionnez une ou plusieurs paires de variables.
3. (Facultatif) Changez/sélectionnez une option **Estimer les tailles d'effet**. Les paramètres contrôlent la façon dont l'utilisation de programme de standardisation est calculée lors de l'estimation de la correction *d* de Cohen et de la correction de Hedges pour chaque paire de variables.

   Ecart type de la différence

   Dénominateur utilisé pour estimer la taille d'effet. Le *d* de Cohen utilise l'écart type échantillon de la différence moyenne. La correction de Hedges se sert de l'écart type échantillon de la différence moyenne ajustée via un facteur de corrélation.

   Ecart type corrigé de la différence

   Dénominateur utilisé pour estimer la taille d'effet. Le *d* de Cohen utilise l'écart type échantillon de la différence moyenne ajustée par la corrélation entre les mesures. La correction de Hedges se sert de l'écart type échantillon de la différence moyenne ajustée par la corrélation entre les mesures, plus un facteur de correction.

   Moyenne des variances

   Dénominateur utilisé pour estimer la taille d'effet. Le *d* de Cohen utilise la racine carrée de la variance moyenne des mesures. La correction de Hedges se sert de la racine carrée de la variance moyenne des mesures, plus un facteur de correction.

4. (En option) Vous avez également ces possibilités :

   - Sélectionnez l'option **Estimer les tailles d'effet** pour contrôler l'estimation de la taille d'effet de test *t*. Si cette option est sélectionnée, vous pouvez contrôler davantage le mode de traitement du programme de standardisation lors de l'estimation de la correction *d* de Cohen et de la correction de Hedges pour chaque paire de variables.
   - Cliquer sur **Options** pour contrôler le traitement des données manquantes et le niveau de l'intervalle de confiance.
   - Cliquer sur **Bootstrap** pour dériver des estimations robustes des erreurs standard et des intervalles de confiance pour des estimations telles que la moyenne, la médiane, le calcul de la proportion, le rapport des cotes, le coefficient de corrélation ou de régression.

Options de Test T pour échantillons appariés

Intervalle de confiance. Par défaut, un intervalle de confiance de 95 % pour la différence dans les moyennes est affiché. Saisir une valeur comprise entre 1 et 99 pour demander un niveau de confiance différent.

Valeurs manquantes : Quand vous testez plusieurs variables et que des données sont manquantes pour au moins une variable, vous pouvez indiquer à la procédure les observations à inclure (ou exclure) :

- **Exclure les observations analyse par analyse :** Chaque test *t* utilise toutes les observations qui ont des données valides pour la paire de variables testées. La taille des échantillons peut varier d'un test à l'autre.
- **Exclure toute observation incomplète :** Chaque test *t* utilise seulement les observations qui ont des données valides pour toutes les paires de variables testées. La taille des échantillons est constante durant les tests.

Fonctions supplémentaires de la commande T TEST

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

- Produire à la fois des tests t pour un échantillon et pour des échantillons indépendants en exécutant une commande unique.
- Tester une variable avec chacune des variables d'une liste dans un test t apparié (avec la sous-commande PAIRS).
- Contrôler l'estimation de la taille d'effet de test *t* (avec la sous-commande ES).

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

Test T pour échantillon unique

La procédure Test T sur échantillon unique permet de tester si la moyenne d'une variable unique diffère d'une constante spécifiée et automatise le calcul de la taille d'effet de test *t*.

Exemples
Un chercheur pourrait vouloir vérifier si le score moyen de QI pour un groupe d'étudiants diffère de 100. Ou un fabricant de céréales peut prélever un échantillon de boîtes de la chaîne de production et vérifier si le poids moyen des échantillons diffère de 1,3 livre au niveau de confiance de 95%.

Statistiques
Pour chaque variable de test : moyenne, écart type et erreur standard de la moyenne et estimation de la taille d'effet pour test *t*. Différence moyenne entre chaque valeur de donnée et la valeur test supposée, le test *t* vérifie que cette différence est égale à 0 et vérifie également l'intervalle de confiance pour cette différence (vous pouvez préciser le niveau de confiance).

Analyse des données

Données
Afin de tester les valeurs d'une variable quantitative par rapport à une valeur test supposée, choisissez une variable quantitative et saisissez une valeur test supposée.

Hypothèses
Ce test suppose que les données sont distribuées normalement ; cependant, ce test résiste convenablement à la normalité.

Obtention d'un test T pour échantillon unique

1. A partir des menus, sélectionnez :

   Analyser > **Comparer les moyennes** > **Test T à échantillon unique ...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins une variable à tester par rapport à la même valeur supposée.
3. Entrez une valeur test numérique à laquelle vous souhaitez comparer chaque moyenne d'échantillon.
4. (En option) Vous avez également ces possibilités :
   - Sélectionnez l'option **Estimer les tailles d'effet** pour contrôler l'estimation de la taille d'effet de test *t*.

- Cliquer sur **Options** pour contrôler le traitement des données manquantes et le niveau de l'intervalle de confiance.

Options de Test T pour échantillon unique

Intervalle de confiance. Par défaut, un intervalle de confiance de 95 % pour la différence entre la moyenne et la valeur de test supposée est affiché. Saisir une valeur comprise entre 1 et 99 pour demander un niveau de confiance différent.

Valeurs manquantes : Quand vous testez plusieurs variables et que des données sont manquantes pour au moins une variable, vous pouvez indiquer à la procédure les observations à inclure (ou exclure).

- **Exclure les observations analyse par analyse :** Chaque test *t* utilise toutes les observations qui ont des données valides pour les variables testées. La taille des échantillons peut varier d'un test à l'autre.
- **Exclure toute observation incomplète :** Chaque test *t* utilise seulement les observations qui ont des données valides pour toutes les variables utilisées dans n'importe lequel des tests *t* requis. La taille des échantillons est constante durant les tests.

Fonctions supplémentaires de la commande T TEST

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

- Produire à la fois des tests t pour un échantillon et pour des échantillons indépendants en exécutant une commande unique.
- Tester une variable avec chacune des variables d'une liste dans un test t apparié (avec la sous-commande PAIRS).
- Contrôler l'estimation de la taille d'effet de test *t* (avec la sous-commande ES).

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

Fonctions supplémentaires de la commande T TEST

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

- Produire à la fois des tests t pour un échantillon et pour des échantillons indépendants en exécutant une commande unique.
- Tester une variable avec chacune des variables d'une liste dans un test t apparié (avec la sous-commande PAIRS).
- Contrôler l'estimation de la taille d'effet de test *t* (avec la sous-commande ES).

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

ANOVA à 1 facteur

La procédure ANOVA à 1 facteur effectue une analyse de variance à un facteur pour une variable dépendante quantitative d'après une variable (indépendante) à un seul facteur et estime la taille d'effet en ANOVA à 1 facteur. L'analyse de variance sert à tester l'hypothèse d'égalité des moyennes. Cette technique est une extension du test *t* pour deux échantillons.

Déterminer que des différences existent parmi les moyennes ne vous suffit peut-être pas. Vous voulez éventuellement savoir quelles sont les moyennes qui diffèrent. Il existe deux types de tests pour comparer les moyennes : les contrastes a priori et les tests post hoc. Les contrastes sont des tests définis *avant* l'expérience, et les tests post hoc sont effectués *après* l'expérience. Vous pouvez aussi tester les tendances à travers les catégories.

Exemple

Les beignets absorbent la graisse dans des proportions variées lorsqu'ils sont cuisinés. Une expérience est conduite à partir de l'utilisation de trois types de graisse : huile d'arachide, huile de maïs, et saindoux. L'huile d'arachide et l'huile de maïs sont des graisses non saturées, et le saindoux une graisse saturée. En plus du fait de déterminer si la quantité de graisse absorbée dépend du type

de graisse utilisée, vous pouvez créer un contraste a priori afin de déterminer si le degré d'absorption de graisse diffère pour les graisses saturées et non saturées.

Statistiques
Pour chaque groupe : nombre d'observations, moyenne, écart type, erreur standard pour la moyenne, minimum, maximum, intervalle de confiance à 95 % pour la moyenne et estimation de la taille d'effet pour ANOVA à 1 facteur. Tests de Levene pour l'homogénéité de la variance, tableau d'analyse de variance et tests d'égalité des moyennes pour chaque variable dépendante, contrastes a priori spécifiés par l'utilisateur et tests de plage et comparaisons multiples post hoc : Bonferroni, Sidak, test de Tukey, GT2 de Hochberg, Gabriel, Dunnett, test *F* de Ryan-Einot-Gabriel-Welsch (*F* de R-E-G-W), test de plage de Ryan-Einot-Gabriel-Welsch (*Q* de R-E-G-W), T2 de Tamhane, T3 de Dunnett, Games-Howell, *C* de Dunnett, test de plage multiple de Duncan, Student-Newman-Keuls (S-N-K), *b* de Tukey, Waller-Duncan, Scheffé et différence la moins significative.

Remarques sur les données

Données
La variable dépendante devrait être quantitative (niveau d'intervalle de mesures).

Hypothèses
Chaque groupe est un échantillon aléatoire indépendant extrait d'une population normale. L'analyse de variance supporte les écarts à la normalité, bien que les données doivent être symétriques.
Les groupes devraient être composés de populations à variance égale. Pour tester cette hypothèse, utiliser le test d'homogénéité de variance de Levene.

Obtention d'une analyse de variance à un facteur

1. A partir des menus, sélectionnez :

   Analyser > **Comparer les moyennes** > **ANOVA à 1 facteur...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins une variable dépendante.
3. Sélectionnez un facteur indépendant simple.

(En option) Vous avez également ces possibilités :

- Sélectionner **Estimer la taille d'effet pour les tests généraux** pour contrôler le calcul de la taille d'effet du test général. Si cette option est sélectionnée, le tableau d'effets de taille ANOVA est affiché dans la sortie.
- Cliquer sur **Contrastes** pour subdiviser les sommes des carrés inter-groupes en tendances composants ou pour spécifier des contrastes a priori.
- Cliquer sur **Post Hoc** pour utiliser les tests de plage post hoc range et les comparaisons multiples appariées pour déterminer les moyennes qui diffèrent.
- Cliquer sur **Options** pour contrôler le traitement des données manquantes et le niveau de l'intervalle de confiance.
- Cliquer sur **Bootstrap** pour dériver des estimations robustes des erreurs standard et des intervalles de confiance pour des estimations telles que la moyenne, la médiane, le calcul de la proportion, le rapport des cotes, le coefficient de corrélation ou de régression.

Contrastes ANOVA à 1 facteur

Vous pouvez diviser les sommes des carrés inter-groupes en tendances composants ou spécifier les contrastes a priori.

Polynomial
Diviser les sommes des carrés inter-groupes en tendances composants. Vous pouvez tester la tendance d'une variable dépendante à travers les niveaux ordonnés du facteur. Par exemple, vous

pourriez tester la tendance linéaire (croissante ou décroissante) des salaires perçus les plus élevés à travers les niveaux ordonnés.

- **Degré :** Vous pouvez choisir un polynôme de premier, deuxième, troisième, quatrième ou cinquième degré.

Coefficients
Contrastes a priori spécifiés à tester par la statistique *t*. Saisissez un coefficient pour chaque groupe (catégorie) du facteur et cliquez sur **Ajouter** après chaque saisie. Chaque nouvelle valeur s'ajoute au bas de la liste des coefficients. Pour spécifier des groupes de contrastes supplémentaires, cliquez sur **Suivant**. Utilisez **Suivant** et **Précédent** pour vous déplacer entre les groupes de contrastes.

Estimer la taille d'effet pour les contrastes
Contrôle le calcul de la taille d'effet pour le test général. Quand ce paramètre est activé, au moins une des options suivantes doit être sélectionnée pour calculer les tailles d'effet. Ce paramètre est activé quand au moins un contraste est spécifié et qu'il génère un tableau d'effets de taille ANOVA dans la sortie.

Utiliser l'écart type combiné pour tous les groupes comme programme de standardisation
Utilise l'écart type combiné pour tous les groupes comme programme de standardisation dans l'estimation de la taille d'effet. Il s'agit du paramètre par défaut qui est disponible quand l'option **Estimer la taille d'effet pour les contrastes** est sélectionnée.

Utiliser l'écart type combiné pour les groupes impliqués dans le contraste comme programme de standardisation
Utilise l'écart type combiné pour les groupes impliqués dans le contraste comme programme de standardisation. Ce paramètre est disponible est disponible quand l'option **Estimer la taille d'effet pour les contrastes** est sélectionnée.

L'ordre des coefficients est important car il correspond à l'ordre croissant des valeurs de catégorie du facteur. Le premier coefficient de la liste correspond à la valeur la plus petite du facteur, et le dernier coefficient correspond à la valeur la plus élevée. Par exemple, s'il y a six catégories de facteurs, les coefficients -1, 0, 0, 0, 0,5 et 0,5 mettent en contraste le premier groupe avec les cinquième et sixième groupes. Pour la plupart des applications, les coefficients doivent être de 0. Les ensembles dont la somme n'est pas égale à 0 peuvent également être utilisés, mais un message d'avertissement s'affiche.

Tests Post Hoc ANOVA à 1 facteur

Lorsque vous avez déterminé qu'il existe des différences parmi les moyennes, les tests de plages post hoc et de comparaisons multiples appariées peuvent déterminer les moyennes qui diffèrent. Les tests de plage identifient les sous-groupes homogènes de moyennes qui ne diffèrent pas les uns des autres. Les comparaisons multiples appariées testent la différence entre les moyennes appariées et engendrent une matrice pour laquelle les astérisques indiquent les moyennes de groupes significativement différentes au niveau alpha 0.05.

Hypothèse de variances égales

Le test de Tukey, le GT2 de Hochberg, le test de Gabriel et le test de Scheffé sont des tests de comparaisons multiples et de plage. Il existe d'autres tests de plage, tels que le test *B* de Tukey, le S-N-K (Student-Newman-Keuls), le Duncan, le R-E-G-W *F* (*F* de Ryan-Einot-Gabriel-Welsch), le R-E-G-W *Q* (test de plage de Ryan-Einot-Gabriel-Welsch) et le Waller-Duncan. Les tests de comparaison multiple disponibles sont les suivants : Bonferroni, Tukey, Sidak, Gabriel, Hochberg, Dunnett, Scheffé et LSD (Différence la moins significative).

- *LSD*. Utilise les tests t pour effectuer toutes les comparaisons appariées entre les moyennes de groupe. Le taux d'erreur n'est pas corrigé dans le cas de comparaisons multiples.
- *Bonferroni*. Utilise des tests *t* pour effectuer des comparaisons appariées entre les moyennes de groupes, mais contrôle le taux d'erreur global en spécifiant comme taux d'erreur pour chaque test le taux d'erreur empirique divisé par le nombre total de tests. Le niveau de signification observé est ainsi ajusté en raison des comparaisons multiples réalisées.

- *Sidak*. Test de comparaisons multiples appariées basé sur les statistiques t. Le test de Sidak ajuste le niveau de signification en fonction des comparaisons multiples et fournit des bornes plus étroites que le test de Bonferroni.
- *Scheffe*. Exécute des comparaisons appariées simultanées pour toutes les combinaisons de paires de moyennes possibles. Utilise la distribution d'échantillonnage F. Peut servir à examiner toutes les combinaisons linéaires possibles de moyennes de groupe, et pas seulement des comparaisons appariées.
- *F de R-E-G-W (Ryan-Einot-Gabriel-Welsch)*. Procédure multiple descendante de Ryan-Einot-Gabriel-Welsch basée sur un test F.
- *Q de R-E-G-W (Ryan-Einot-Gabriel-Welsch)*. Procédure multiple descendante de Ryan-Einot-Gabriel-Welsch basée sur la plage de Student.
- *S-N-K*. Ce test effectue toutes les comparaisons appariées de moyennes à l'aide de la distribution des plages de Student. Lorsque la taille des échantillons est égale, il compare aussi les moyennes par paire dans les sous-ensembles homogènes, en utilisant une procédure étape par étape. Les moyennes sont triées dans l'ordre décroissant et les différences extrêmes sont testées en premier.
- *Tukey*. Utilise les statistiques de plage de Student pour effectuer des comparaisons appariées de groupes. Fixe le taux d'erreur expérimental au niveau du taux d'erreur de l'ensemble pour toutes des comparaisons appariées.
- *B de Tukey*. Utilise la distribution des plages de Student pour effectuer des comparaisons appariées de groupes. La valeur critique est la moyenne de la valeur correspondante du test de Tukey et du test de Student-Newman-Keuls.
- *Duncan*. Réalise des comparaisons appariées en suivant un ordre pas à pas identique à celui utilisé dans le test de Student-Newman-Keuls, mais établit un niveau de protection du taux d'erreur pour l'ensemble des tests, plutôt que pour chaque test en particulier. Utilise la statistique de plage de Student.
- *GT2 de Hochberg*. Test de multiples comparaisons et intervalles appariés utilisant le module maximal de Student. Similaire au test de Tukey.
- *Gabriel*. Test de comparaison appariée qui utilise le module maximal de Student. Il est plus puissant que le GT2 de Hochberg lorsque les tailles des cellules sont inégales. Le test de Gabriel offre plus de souplesse lorsque les tailles des cellules divergent beaucoup.
- *Waller-Duncan*. Test de comparaisons multiples reposant sur une statistique t et utilisant une approche bayésienne.
- *Dunnett*. Test *t* de plusieurs comparaisons appariées comparant un ensemble de traitements à une moyenne de contrôle unique.La dernière catégorie est la catégorie de contrôle par défaut. Vous pouvez également choisir la première catégorie. Tests **bilatéraux** qui vérifient que la moyenne à n'importe quel niveau (sauf la catégorie de contrôle) du facteur n'est pas égale à celle de la catégorie de contrôle. **< Tests de contrôle** si la moyenne à n'importe quel niveau du facteur est inférieure à celle de la catégorie de contrôle. **> Tests de contrôle** si la moyenne à n'importe quel niveau du facteur est supérieure à celle de la catégorie de contrôle.

Hypothèse de variances inégales

Les tests de comparaison multiple qui ne supposent pas des variances égales sont les suivants : T2 de Tamhane, T3 de Dunnett, Games-Howell et *C* de Dunnett.

- *Tamhane's T2*. Test des comparaisons appariées basé sur le test t. Ce test est opportun lorsque les variances sont inégales.
- *Dunnett's T3*. Test des comparaisons appariées basé sur le module maximal de Student. Ce test est opportun lorsque les variances sont inégales.
- *Games-Howell*. Test de comparaison appariée qui est parfois libéral. Ce test est opportun lorsque les variances sont inégales.
- *Dunnett's C*. Test des comparaisons appariées basé sur la plage de Student. Ce test est opportun lorsque les variances sont inégales.

Remarque : Il peut être plus facile d'interpréter la sortie de tests post hoc si vous désélectionnez l'option **Masquer les lignes et les colonnes vides** dans la boîte de dialogue Propriétés du tableau (dans un tableau croisé dynamique, sélectionnez **Propriétés du tableau** dans le menu Format).

Test d'hypothèse nulle

Indique le mode de traitement du niveau d'importance (alpha) pour le test post hoc.

Utiliser le même niveau d'importance (alpha) que celui défini dans Options
Si cette option est sélectionnée, le niveau utilisé est celui spécifié dans la boîte de dialogue Options.

Spécifier le niveau d'importance (alpha) pour le test post hoc
Si cette option est sélectionnée, vous pouvez spécifier le niveau d'importance (alpha) dans la zone **Niveau**.

Obtention de tests post hoc pour ANOVA à 1 facteur

Options ANOVA à 1 facteur

Statistiques
Choisissez une ou plusieurs des options suivantes :

Descriptive
La procédure calcule le nombre d'observations, la moyenne, l'écart type, l'erreur standard de la moyenne, le minimum, le maximum, et les intervalles de confiance à 95 % pour chaque variable dépendante de chaque groupe.

Effets fixes et aléatoires
La procédure affiche l'écart type, l'erreur standard et l'intervalle de confiance à 95 % pour le modèle à effets fixes, ainsi que l'erreur standard, l'intervalle de confiance à 95 % et l'estimation de la variance inter-composants pour le modèle à effets aléatoires.

Test d'homogénéité de variance
La procédure calcule la statistique de Levene pour tester l'égalité des variances de groupe. Ce test ne dépend pas de l'hypothèse de normalité.

Brown-Forsythe
Calcule la statistique de Brown-Forsythe pour tester l'égalité des moyennes de groupe. Il est préférable d'utiliser cette statistique (au lieu de la statistique *F*) lorsque l'hypothèse d'égalité des variances n'est pas satisfaite.

Welch
Calcule la statistique de Welch pour tester l'égalité des moyennes de groupe. Il est préférable d'utiliser cette statistique (au lieu de la statistique *F*) lorsque l'hypothèse d'égalité des variances n'est pas satisfaite.

Valeurs manquantes
Contrôle le traitement des valeurs manquantes.

Exclure les observations analyse par analyse
Aucune observation avec valeur manquante n'est utilisée, que ce soit pour la variable dépendante ou pour la variable de facteur d'une analyse donnée. De même, on n'utilise pas d'observation en dehors de la plage spécifiée pour le facteur.

Exclure toute observation incomplète
Les observations ayant des valeurs manquantes pour le facteur ou pour toute variable dépendante contenue dans la liste dépendante de la boîte de dialogue principale sont exclues de toutes les analyses. Si vous n'avez pas spécifié de variables multiples dépendantes, cela est sans effet.

Intervalle de confiance
Par défaut, un intervalle de confiance de 95 % pour la différence entre la moyenne et la valeur de test supposée est affiché. Saisir une valeur comprise entre 1 et 99 pour demander un niveau de confiance différent.

Tracé des moyennes
Affiche un graphique qui trace les moyennes de sous-groupes (moyennes de chaque groupe défini par la variable de facteur).

Pour spécifier les options pour ANOVA à 1 facteur

Fonctions supplémentaires de la commande ONEWAY

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

- Obtenir des statistiques à effets fixes et aléatoires. Ecart type, erreur standard de la moyenne et intervalles de confiance de 95 % pour le modèle à effets fixes. Erreur standard, intervalles de confiance de 95 % et estimation de la variance inter-composants pour le modèle à effets aléatoires (en utilisant `STATISTICS=EFFECTS`).
- Spécifier les niveaux alpha pour la différence de moindre signification, les tests de comparaison multiple Bonferroni, Duncan et Scheffé (avec la sous-commande `RANGES`).
- Ecrire une matrice des moyennes, des écarts types et des fréquences ou lire une matrice des moyennes, des fréquences, des variances regroupées en pool et des degrés de liberté des variances regroupées en pool. Ces matrices peuvent être utilisées à la place des données brutes pour obtenir une analyse de variance à un facteur (avec la sous-commande `MATRIX`).

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

Analyse GLM – Univarié

La procédure GLM – Univarié fournit un modèle de régression et une analyse de variance pour plusieurs variables dépendantes par un ou plusieurs facteurs ou variables. Les facteurs divisent la population en groupes. Cette procédure de régression linéaire généralisée vous permet de tester les hypothèses nulles à propos des effets des autres variables sur la moyenne de différents regroupements de la variable dépendante. Vous pouvez rechercher les interactions entre les facteurs ainsi que les effets des différents facteurs, certains d'entre eux étant aléatoires. En outre, les effets et les interactions des covariables avec les facteurs peuvent être inclus. Pour l'analyse de la régression, les variables indépendantes (prédicteur) sont spécifiées comme covariables.

Vous pouvez tester les modèles équilibrés comme déséquilibrés. Un plan est équilibré si chaque cellule de ce modèle contient le même nombre d'observations. L'analyse GLM – Univarié teste non seulement les hypothèses mais elle produit également des estimations.

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

Les résidus, les prévisions, la distance de Cook et les valeurs influentes peuvent être enregistrées sous forme de nouvelles variables dans votre fichier de données pour vérifier les hypothèses.

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

Exemple : Des données sont collectées sur les différents participants au Marathon de Paris sur plusieurs années. Le temps effectué par chaque participant est la variable dépendante. Les autres facteurs comprennent le temps (froid, modéré, chaud), le nombre de mois d'entraînement, le nombre de marathons précédemment effectués et le sexe. L'âge est considéré comme co-variable. Vous devez trouver que le sexe a un effet significatif et que l'interaction du sexe avec le temps est significatif.

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

Statistiques : Tests de plage post hoc et comparaisons multiples : différence la moins significative, Bonferroni, Sidak, Scheffé, *test de F* multiple de Ryan-Einot-Gabriel-Welsch, plage multiple de Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls, test de Tukey, *b* de Tukey, Duncan, GT2 de Hochberg, test *t* de Gabriel, Waller-Duncan, Dunnett (unilatéral, bilatéral), T2 de Tamhane, T3 de Dunnett, Games-Howell et *C* de Dunnett. Statistiques descriptives : moyennes observées, écarts types et effectifs pour toutes les variables dépendantes de toutes les cellules. Tests de Levene pour l'homogénéité de la variance.

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

Remarques sur les données GLM - Univarié

Données. La variable dépendante est quantitative. Les facteurs sont qualitatifs. Il peut s'agir de valeurs numériques ou alphanumériques de 8 caractères au maximum. Les covariables sont des variables quantitatives liées à la variable dépendante.

Hypothèses : Les données forment un échantillon aléatoire d'une population normale ou gaussienne. Dans cette population, toutes les variances de cellule sont égales. L'analyse de variance supporte les écarts à la normalité, bien que les données doivent être symétriques. Pour vérifier les hypothèses, vous pouvez utiliser les tests d'homogénéité de la variance et les tracés de dispersion par niveau. Vous pouvez également étudier les résidus et les tracés résiduels.

Pour obtenir des tables GLM - Univarié

1. A partir des menus, sélectionnez :

   Analyse > **Modèle linéaire général** > **Univarié...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une variable dépendante.
3. Sélectionnez des variables pour Facteurs fixés, Facteurs aléatoires et Covariables en fonction de vos données.
4. En option, vous pouvez utiliser WLS Weight pour préciser une variable de pondération pour l'analyse des moindres carrés pondérés. Si la valeur de la variable de pondération est nulle, négative ou manquante, l'observation est exclue de l'analyse. Une variable déjà utilisée dans le modèle ne peut pas servir de variable de pondération.

Modèle GLM

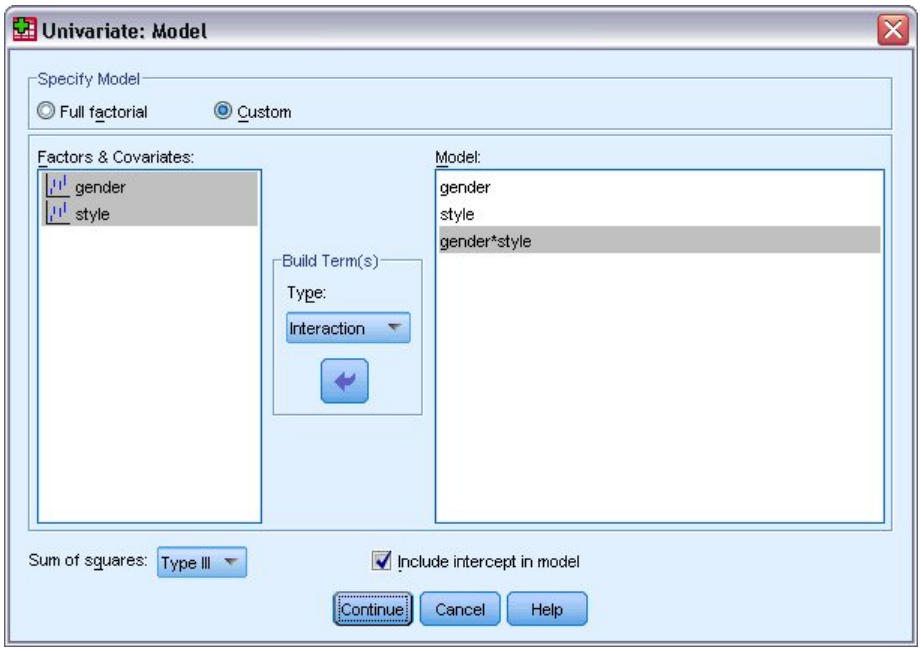

Figure 1. Boîte de dialogue Modèle univarié

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

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

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

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

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

Termes construits et termes personnalisés

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

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

Somme des carrés

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

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

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

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

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

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

équivalente à la technique de Yates des carrés moyens pondérés. La méthode des sommes des carrés de type III sert généralement pour :

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

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

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

Contrastes GLM

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

GLM - Univarié : Le test des hypothèses est fondé sur l'hypothèse nulle **LB** =0, **L** étant la matrice des coefficients de contraste et **B** le vecteur de paramètre. Lorsqu'un contraste est spécifié, une matrice **L** est créée. Les colonnes de la matrice **L** correspondantes au facteur concordent avec le contraste. Les colonnes restantes sont ajustées de telle sorte que la matrice **L** puisse être estimée.

La sortie reprend une statistique *F* pour chaque ensemble de contrastes. Pour les différences de contraste, le système affiche également les intervalles de confiance simultanés de type Bonferroni fondés sur la distribution *t* de Student.

Contrastes possibles

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

Types de contraste

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

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

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

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

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

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

Tracés de profil GLM

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

servir à créer un tracé distinct. Tous les facteurs fixés et aléatoire sont disponibles pour les tracés. Pour les analyses multivariées, les tracés de profil sont créés pour chaque variable dépendante. Dans une analyse à mesures répétées, à la fois les facteurs inter-sujets et intra-sujets peuvent être utilisés dans les tracés de profil. GLM - Multivarié et GLM - Mesures Répétées ne sont disponibles que si vous avez installé l'option Statistiques avancées.

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

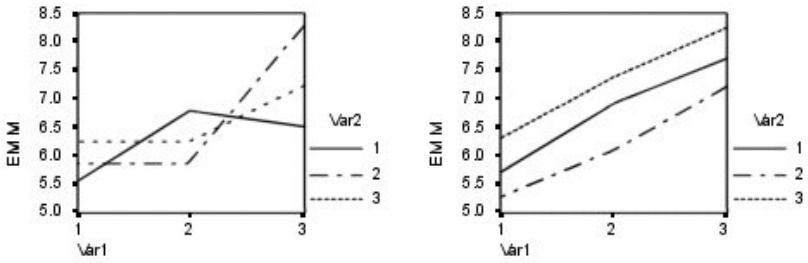

Figure 2. Tracé non parallèle (gauche) et tracé parallèle (droite)

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

Type de graphique
Il peut s'agir d'un graphique à courbes ou à barres.

Barres d'erreur
Vous pouvez inclure des barres d'erreur représentant l'intervalle de confiance ou le nombre d'erreurs standard. L'intervalle de confiance est basé sur le niveau de signification défini dans la boîte de dialogue Options.

Inclure la ligne de référence pour la moyenne générale
Inclut une ligne de référence qui représente la moyenne générale.

L'axe des Y commence à 0
Pour les graphiques en ligne avec toutes les valeurs positives ou toutes les valeurs négatives, force l'axe Y à commencer à 0. Les diagrammes à barres commencent toujours par 0 (ou incluent toujours 0).

Options GLM

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

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

L'option des **tests d'homogénéité** produit les tests de Levene d'homogénéité de la variance pour chaque variable dépendante sur toutes les combinaisons de niveaux des facteurs inter-sujets, uniquement pour les facteurs inter-sujets. . Les options des tracés de dispersion par niveau et résiduels sont utiles pour vérifier les hypothèses sur les données. Ceci n'est pas valable s'il n'y a pas de facteurs. Sélectionnez **Tracés résiduels** pour produire un tracé résiduel observé/estimé/standardisé pour chaque variable dépendante. Ces tracés sont utiles pour vérifier l'hypothèse de variance égale. Sélectionnez **Manque d'ajustement** pour vérifier si la relation entre la variable dépendante et les variables indépendantes peut

être convenablement décrite par le modèle. **Fonction(s) d'estimation générale(s)** permet de construire des tests d'hypothèses personnalisés basés sur la ou les fonctions d'estimation générale(s). Les lignes d'une matrice de coefficients de contraste sont des combinaisons linéaires de la ou des fonctions d'estimation générale(s).

Les **Tests d'hétéroscédasticité** permettent de tester si la variance des erreurs (pour chaque variable dépendante) dépend des valeurs des variables indépendantes. Pour les test de **Breusch-Pagan**, **Breusch-Pagan modifié** et **F**, vous pouvez spécifier le modèle sur lequel repose le test. Par défaut, le modèle est composé d'un terme constant, d'un terme linéaire dans les valeurs prédites, d'un terme quadratique dans les valeurs prédites et d'un terme d'erreur.

L'option **Estimations de paramètres avec erreurs standard robustes** affiche un tableau d'estimations de paramètres avec des erreurs standard robustes ou cohérentes avec l'hétéroscédasticité, ainsi que des statistiques *t*, des valeurs de signification et des intervalles de confiance qui utilisent les erreurs standard robustes. Il existe cinq méthodes différentes pour l'estimation de matrice de covariance robuste.

HC0
: Méthode basée sur l'estimation d'échantillons larges ou asymptotiques d'origine de type robuste, empirique ou sandwich de la matrice de covariance des estimations de paramètres. La partie centrale du sandwich contient les résidus OLS (moindres carrés ordinaires) ou WLS (moindres carrés pondérés).

HC1
: Variante de HC0 avec échantillons finis, impliquant sa multiplication par N/(N-p), N étant la taille de l'échantillon et p le nombre de paramètres non redondants dans le modèle.

HC2
: Variante de HC0, impliquant la division des carrés résiduels par 1-h, h étant la valeur influente pour l'observation.

HC3
: Variante de HC0 similaire à une estimation jackknife. Les carrés résiduels sont divisés par le carré de 1-h.

HC4
: Variante de HC0 impliquant la division des carrés résiduels par 1-h à une puissance différente en fonction des valeurs h, N et p, avec une borne supérieure de 4.

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

Fonctions supplémentaires de la commande UNIANOVA

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

- Spécifier les effets imbriqués dans un plan (à l'aide de la sous-commande `DESIGN`).
- Spécifier les tests d'effets par rapport à une combinaison linéaire d'effets ou une valeur (à l'aide de la sous-commande `TEST`).
- Spécifier de multiples contrastes (à l'aide de la sous-commande `CONTRAST`).
- Inclure les valeurs manquantes de l'utilisateur (à l'aide de la sous-commande `MISSING`).
- Spécifier les critères EPS (à l'aide de la sous-commande `CRITERIA`).
- Construisez une matrice **L** personnalisée, une matrice **M** ou une matrice **K** (à l'aide des sous-commandes `LMATRIX`, `MMATRIX` et `KMATRIX`).
- Pour les contrastes simples ou d'écart, spécifier une catégorie de référence intermédiaire (à l'aide de la sous-commande `CONTRAST`).
- Spécifier les mesures pour les contrastes polynomiaux (à l'aide de la sous-commande `CONTRAST`).
- Spécifier des termes d'erreur pour les comparaisons post hoc (à l'aide de la sous-commande `POSTHOC`).

- Calculer les moyennes marginales estimées pour chaque facteur ou interaction entre facteurs parmi les facteurs de la liste (à l'aide de la sous-commande EMMEANS).
- Attribuer des noms aux variables temporaires (à l'aide de la sous-commande SAVE).
- Construire un fichier de matrice de corrélation (à l'aide de la sous-commande OUTFILE).
- Construire un fichier de type matrice de données qui contient les statistiques provenant de la table ANOVA inter-sujets (à l'aide de la sous-commande OUTFILE).
- Enregistrer la matrice du plan dans un nouveau fichier de données (à l'aide de la sous-commande OUTFILE).

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

Comparaisons post hoc GLM

Test de comparaison multiple post hoc : Lorsque vous avez déterminé qu'il existe des différences parmi les moyennes, les tests de plages post hoc et de comparaisons multiples appariées peuvent déterminer les moyennes qui diffèrent. Les comparaisons sont effectuées sur des valeurs non-ajustées. Ces tests servent aux facteurs inter-sujets fixés seulement. Dans GLM - Mesures répétées, ces tests ne sont pas disponibles s'il n'y a pas de facteurs inter-sujets. Les tests de comparaisons multiples post hoc sont effectués pour la moyenne de tous les niveaux des facteurs intra-sujets. Pour GLM - Multivarié, les tests post hoc sont effectués séparément pour chaque variable dépendante. GLM - Multivarié et GLM - Mesures Répétées ne sont disponibles que si vous avez installé l'option Statistiques avancées.

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

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

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

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

Lorsque les variances sont inégales, utilisez le **T2 de Tamhane** (test de comparaisons appariées conservatif fondé sur un test *t*, le **T3 de Dunnett** (test de comparaison appariée fondé sur le modulo maximal de Student), le **test de comparaison appariée** de **Games-Howell** (parfois flexible) ou le **C de Dunnetts** (test de comparaison appariée fondé sur la plage de Student). Notez que s'il y a plusieurs facteurs dans le modèle, ces tests ne sont pas valides et ne seront pas produits.

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

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

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

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

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

Options GLM

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

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

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

Les **Tests d'hétéroscédasticité** permettent de tester si la variance des erreurs (pour chaque variable dépendante) dépend des valeurs des variables indépendantes. Pour les test de **Breusch-Pagan**, **Breusch-Pagan modifié** et **F**, vous pouvez spécifier le modèle sur lequel repose le test. Par défaut, le modèle est composé d'un terme constant, d'un terme linéaire dans les valeurs prédites, d'un terme quadratique dans les valeurs prédites et d'un terme d'erreur.

L'option **Estimations de paramètres avec erreurs standard robustes** affiche un tableau d'estimations de paramètres avec des erreurs standard robustes ou cohérentes avec l'hétéroscédasticité, ainsi que des statistiques *t*, des valeurs de signification et des intervalles de confiance qui utilisent les erreurs standard robustes. Il existe cinq méthodes différentes pour l'estimation de matrice de covariance robuste.

HC0
: Méthode basée sur l'estimation d'échantillons larges ou asymptotiques d'origine de type robuste, empirique ou sandwich de la matrice de covariance des estimations de paramètres. La partie

centrale du sandwich contient les résidus OLS (moindres carrés ordinaires) ou WLS (moindres carrés pondérés).

HC1
Variante de HC0 avec échantillons finis, impliquant sa multiplication par N/(N-p), N étant la taille de l'échantillon et p le nombre de paramètres non redondants dans le modèle.

HC2
Variante de HC0, impliquant la division des carrés résiduels par 1-h, h étant la valeur influente pour l'observation.

HC3
Variante de HC0 similaire à une estimation jackknife. Les carrés résiduels sont divisés par le carré de 1-h.

HC4
Variante de HC0 impliquant la division des carrés résiduels par 1-h à une puissance différente en fonction des valeurs h, N et p, avec une borne supérieure de 4.

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

Fonctions supplémentaires de la commande UNIANOVA

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

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

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

Enregistrement GLM

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

Prévisions : Valeurs que le modèle estime pour chaque observation.

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

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

- *Distance de Cook*. Mesure permettant de savoir de combien les résidus de toutes les observations seraient modifiés si une observation donnée était exclue du calcul des coefficients de régression. Si la distance de Cook est élevée, l'exclusion d'une observation changerait substantiellement la valeur des coefficients.
- *Valeurs de levier*. Valeurs influentes non centrées. Mesure de l'influence d'un point sur l'ajustement de la régression.

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

- *Non normalisé*. Différence entre une valeur observée et la valeur prévue par le modèle.
- *Pondéré*. Résidus non standardisés pondérés. Disponibles uniquement lorsqu'une variable WLS a été préalablement sélectionnée.
- *Standardisés*. Résidu, divisé par une estimation de son écart type. Egalement appelés résiduels de Pearson, les résiduels standardisés ont une moyenne de 0 et un écart type de 1.
- *Studentized*. Résidu, divisé par une estimation de son écart type, qui varie d'une observation à l'autre, selon la distance entre les valeurs et les moyennes des variables indépendantes de chaque observation. Parfois appelés résidus de Student interne.
- *Supprimé*. Résidu d'une observation lorsque cette dernière est exclue du calcul des coefficients de régression. Il s'agit de la différence entre la valeur de la variable dépendante et la prévision ajustée.

Statistiques à coefficients : Ecrit une matrice variance-covariance des estimations des paramètres du modèle dans un nouveau jeu de données de la session en cours ou dans un fichier de données externe au format IBM SPSS Statistiques. D'autre part, pour chaque variable dépendante, il y aura une ligne d'estimations de paramètres, une ligne d'erreurs standard des estimations de paramètres, une ligne de valeurs de signification pour les statistiques *t* correspondant aux estimations et une ligne de degrés de liberté résiduels. Pour un modèle multivarié, il y a les mêmes lignes pour chaque variable dépendante. Lorsque les statistiques cohérentes avec l'hétéroscédasticité sont sélectionnées (disponibles uniquement pour les modèles univariés), la matrice variance-covariance est calculée à l'aide d'un estimateur-M, la ligne des erreurs standard affiche les erreurs standard robustes et les valeurs de signification reflètent les erreurs robustes. Vous pouvez utiliser ces fichiers de matrice dans les autres procédures qui lisent des fichiers de matrice.

Moyenne marginale estimée GLM

Sélectionnez les facteurs et les interactions pour lesquels vous souhaitez obtenir des estimations de la moyenne marginale de la population dans les cellules. Ces moyennes sont ajustées pour les covariables, si elles existent.

Comparer les effets principaux

Propose des comparaisons appariées non corrigées des moyennes marginales estimées pour tout effet principal dans le modèle, à la fois pour les facteurs intersujets et intrasujets. Ceci n'est valable que si les effets principaux sont sélectionnés dans la liste Afficher les moyennes.

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

Ajustement intervalle de confiance
Sélectionnez l'ajustement aux intervalles de confiance et à la significativité des intervalles en adoptant l'une des méthodes suivantes : la différence de moindre signification (LSD), l'ajustement Bonferroni ou l'ajustement de Sidak. Cet élément n'est disponible que si **Comparer les effets principaux** et/ou **Comparer les effets principaux simples** est sélectionné.

Spécification de la moyenne marginale estimée

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

Options GLM

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

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

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

Les **Tests d'hétéroscédasticité** permettent de tester si la variance des erreurs (pour chaque variable dépendante) dépend des valeurs des variables indépendantes. Pour les test de **Breusch-Pagan**, **Breusch-Pagan modifié** et **F**, vous pouvez spécifier le modèle sur lequel repose le test. Par défaut, le modèle est composé d'un terme constant, d'un terme linéaire dans les valeurs prédites, d'un terme quadratique dans les valeurs prédites et d'un terme d'erreur.

L'option **Estimations de paramètres avec erreurs standard robustes** affiche un tableau d'estimations de paramètres avec des erreurs standard robustes ou cohérentes avec l'hétéroscédasticité, ainsi que des statistiques *t*, des valeurs de signification et des intervalles de confiance qui utilisent les erreurs standard robustes. Il existe cinq méthodes différentes pour l'estimation de matrice de covariance robuste.

HC0
Méthode basée sur l'estimation d'échantillons larges ou asymptotiques d'origine de type robuste, empirique ou sandwich de la matrice de covariance des estimations de paramètres. La partie centrale du sandwich contient les résidus OLS (moindres carrés ordinaires) ou WLS (moindres carrés pondérés).

HC1
Variante de HC0 avec échantillons finis, impliquant sa multiplication par N/(N-p), N étant la taille de l'échantillon et p le nombre de paramètres non redondants dans le modèle.

HC2
Variante de HC0, impliquant la division des carrés résiduels par 1-h, h étant la valeur influente pour l'observation.

HC3
Variante de HC0 similaire à une estimation jackknife. Les carrés résiduels sont divisés par le carré de 1-h.

HC4
Variante de HC0 impliquant la division des carrés résiduels par 1-h à une puissance différente en fonction des valeurs h, N et p, avec une borne supérieure de 4.

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

Modèle de régression auxiliaire GLM

La boîte de dialogue Modèle de régression auxiliaire définit le modèle utilisé pour le test d'hétéroscédasticité.

Utiliser les valeurs prédites
Utilise un modèle qui comporte un terme constant, un terme linéaire dans les valeurs prédites, un terme quadratique dans les valeurs prédites et un terme d'erreur.

Utiliser le modèle univarié
Utilise le modèle spécifié dans la sous-boîte de dialogue Modèle. Un terme de constante est inclus si le modèle spécifié n'en contient pas déjà.

Modèle personnalisé
Utilise le modèle que vous avez choisi.

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

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

Fonctions supplémentaires de la commande UNIANOVA

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

- Spécifier les effets imbriqués dans un plan (à l'aide de la sous-commande `DESIGN`).
- Spécifier les tests d'effets par rapport à une combinaison linéaire d'effets ou une valeur (à l'aide de la sous-commande `TEST`).
- Spécifier de multiples contrastes (à l'aide de la sous-commande `CONTRAST`).
- Inclure les valeurs manquantes de l'utilisateur (à l'aide de la sous-commande `MISSING`).
- Spécifier les critères EPS (à l'aide de la sous-commande `CRITERIA`).
- Construisez une matrice **L** personnalisée, une matrice **M** ou une matrice **K** (à l'aide des sous-commandes `LMATRIX`, `MMATRIX` et `KMATRIX`).
- Pour les contrastes simples ou d'écart, spécifier une catégorie de référence intermédiaire (à l'aide de la sous-commande `CONTRAST`).
- Spécifier les mesures pour les contrastes polynomiaux (à l'aide de la sous-commande `CONTRAST`).

- Spécifier des termes d'erreur pour les comparaisons post hoc (à l'aide de la sous-commande POSTHOC).
- Calculer les moyennes marginales estimées pour chaque facteur ou interaction entre facteurs parmi les facteurs de la liste (à l'aide de la sous-commande EMMEANS).
- Attribuer des noms aux variables temporaires (à l'aide de la sous-commande SAVE).
- Construire un fichier de matrice de corrélation (à l'aide de la sous-commande OUTFILE).
- Construire un fichier de type matrice de données qui contient les statistiques provenant de la table ANOVA inter-sujets (à l'aide de la sous-commande OUTFILE).
- Enregistrer la matrice du plan dans un nouveau fichier de données (à l'aide de la sous-commande OUTFILE).

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

Bland-Altman

L'analyse de Bland Altman est une technique graphique permettant d'évaluer le biais entre les différences moyennes. Altman et Bland ont proposé cette analyse en 1983. L'analyse de Bland Altman est basée sur la quantification de l'accord entre deux mesures quantitatives en explorant les différences moyennes et en construisant les limites de l'accord.

Exemple

Les laboratoires médicaux doivent souvent évaluer la concordance entre deux méthodes de mesure, choisir entre deux méthodes ou évaluer une méthode nouvelle ou alternative. De même, lorsqu'il existe un problème d'alignement entre deux instruments, il est nécessaire de disposer d'outils permettant de mesurer et d'évaluer les différences et leurs causes.

Cette procédure nécessite l'édition de base de Statistics.

Dans les menus, choisissez : **Analyser Statistiques descriptives** > **Analyse de Bland Altman**....

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

Considérations sur les données : Les variables sont numériques.

1. Sélectionnez les variables et ajoutez-les aux **Variables appariées:**.
2. Si vous sélectionnez **OK**, la procédure s'exécute en tenant compte des paramètres par défaut.
3. Vous pouvez choisir soit **Appliquer la transformation logarithmique** soit **Apparier seulement** soit les deux.
4. Cliquez sur **OK**.

Appliquer la transformation logarithmique : La **transformation logarithmique** est recommandée si les différences ne sont pas normalement distribuées.

Paired only : cette option est spécifiée lorsque la première variable de la première colonne est comparée à la première variable de la deuxième colonne.

Remarque : Parmi les variables appariées, l'analyse n'est effectuée que jusqu'à la dernière valeur de l'ensemble de variables contenant le plus petit nombre de valeurs. L'analyse est entièrement réalisée pour les ensembles de variables appariées contenant un nombre égal de valeurs.

Bland Altman : Critères

Vous pouvez modifier les paramètres par défaut dans l'onglet **Bland Atman : Criteria**.

1. Par défaut, l'option permettant d'afficher les ratios en pourcentage n'est pas sélectionnée. Si vous souhaitez l'ajouter, sélectionnez **Ratio en pourcentages**.
2. Vérifiez les paramètres de **Champ d'application** et **Limites de l'accord**, et modifiez-les si nécessaire.
3. Cliquez sur **Continuer**

Bland Altman : Imprimer

1. Cochez les options que vous souhaitez inclure dans le résultat.
2. Si vous sélectionnez **Imprimer les détails de l'analyse**, le nombre maximal de lignes à afficher est fixé par défaut à 10. Modifiez le nombre pour augmenter ou diminuer le nombre de lignes que vous souhaitez afficher. Le nombre maximum de lignes de données pouvant être affichées est de 1000. **Le tri** peut être **basé sur** les quatre options de liste et sur l'ordre **croissant ou décroissant**.

   Par défaut, **Tri** est défini sur la base de la première mesure et dans l'ordre décroissant.
3. Cliquez sur **Continuer**

Bland Altman : Plot

Vous pouvez modifier les paramètres par défaut dans l'onglet **Analyse de l'Atman glissant : Tracé**.

1. Choisissez les paramètres souhaités, notamment **Plans de rapport**, **Ligne de référence à la différence zéro**, **Limites d'accord** et **Intervalles de confiance des différences moyennes**.
2. Cliquez sur **Continuer**

Test du khi-carré

La procédure de test du chi carré permet de classer une variable en catégories et de calculer une statistique du chi carré. Le test du chi carré est un test d'hypothèse statistique utilisé pour déterminer s'il existe une association significative entre deux variables catégorielles. Le test évalue si les fréquences observées dans les données diffèrent des fréquences attendues sous l'hypothèse nulle, qui suppose qu'il n'y a pas d'association entre les variables.

Cette procédure nécessite l'édition de base de Statistics.

Exemple
Dans les études de marché, les analystes doivent comprendre les préférences des consommateurs. Pour ce faire, ils comparent le comportement observé des clients avec les modèles attendus afin d'élaborer des stratégies de marketing efficaces.

Statistiques
Le chi-carré de Pearson, le chi-carré du rapport de vraisemblance, le test exact de Fisher et le chi-carré corrigé de Yates (correction de continuité).

Considérations sur les données et hypothèses
Utilisez des variables catégorielles numériques ordonnées ou désordonnées (niveau de mesure ordinal ou nominal). Pour convertir des variables de chaîne en variables numériques, utilisez la procédure de recodage automatique, disponible dans le menu Transformer.

Les tests non paramétriques ne nécessitent pas d'hypothèses sur la forme de la distribution sous-jacente. On part du principe que les données constituent un échantillon aléatoire. Les fréquences prévues pour chaque catégorie devraient être d'au moins 1. Pas plus de 20% des catégories devraient avoir des fréquences attendues inférieures à 5.

Dans le menu, cliquez sur **Analyser les** > **statistiques** > **descriptivesChiSquare**.

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

1. Sélectionnez les variables et ajoutez-les à la **(aux) ligne(s)** et à la **(aux) colonne(s)**.
2. Cliquez sur **OK**. La procédure se déroule en tenant compte des paramètres par défaut.

Test du chi carré : Exact

- *Asymptotique*. Niveau d'importance basé sur la distribution asymptotique d'une statistique de test. En règle générale, une valeur inférieure à 0,05 est considérée comme significative. La signification asymptotique est basée sur l'hypothèse que le jeu de données est volumineux. Si le jeu de données est petit ou mal distribué, il ne peut être une bonne indication de la signification.

- *estimation de Monte Carlo*. Estimation non biaisée du niveau de signification exact, calculé par des échantillonnages répétés à partir d'un ensemble de tableaux de référence dont les dimensions, les lignes et les colonnes sont identiques à celles du tableau observé. La méthode de Monte Carlo permet d'estimer le niveau exact de signification sans abandonner les hypothèses requises pour la méthode asymptotique. Cette méthode est particulièrement efficace lorsqu'un jeu de données est trop volumineux pour que sa signification exacte soit calculée, mais que les données ne vérifient pas l'hypothèse de la méthode asymptotique.
- *exact*. La probabilité du résultat observé ou d'un résultat plus extrême est calculée de manière exacte. En général, un niveau de signification inférieur à 0,05 est considéré comme significatif, et indique qu'il existe une relation entre les variables de ligne et de colonne.

Test du chi carré : Cellules

La procédure du chi carré affiche les fréquences attendues. Chaque cellule du tableau peut contenir n'importe quelle combinaison d'effectifs et de pourcentages.

Effectifs
Le nombre de cas réellement observés et le nombre de cas attendus si les variables de la ligne et de la colonne sont indépendantes l'une de l'autre. Vous pouvez choisir de masquer les effectifs inférieurs à un entier spécifique. Les valeurs masquées s'affichent sous la forme **<N**, où **N** est l'entier spécifié. L'entier spécifié doit être supérieur ou égal à 2, bien que la valeur 0 soit permise et indique qu'aucun effectif n'est masqué.

Pourcentages
Les pourcentages peuvent s'additionner par ligne ou par colonne. Les pourcentages du nombre total d'observations représentées dans le tableau (une couche) sont également disponibles.

Remarque : Si l'option **Masquer les petits effectifs** est sélectionnée dans le groupe Effectifs, les pourcentages associés aux effectifs masqués sont aussi masqués.

Créer une table de style APA
Crée des tables de sortie qui respectent les instructions de style APA.

Remarque : Les options **Observé**, **Attendu**, **Ligne**, **Colonne** et **Total** ne sont pas disponibles si **Créer une table de style APA** est sélectionné.

Test du chi carré : Format

L'option format permet de spécifier l'ordre des variables affichées dans les résultats du test.

Accéder à **Analyser** > **les statistiques descriptives** > **ChiSquare**. Cliquez sur **Format** dans la boîte de dialogue du **test du chi carré**.

Vous pouvez arranger les lignes par ordre croissant ou décroissant de valeur de la variable de ligne. Dans l' **ordre des lignes**, sélectionnez **Ascendant** ou **Descendant**.

Corrélations bivariées

La procédure Corrélations bivariées calcule le coefficient de corrélation de Pearson, le rho de Spearman et le tau-*b* de Kendall avec leurs niveaux de signification. Les corrélations mesurent comment les variables ou les ordres de rang sont liés. Avant de calculer un coefficient de corrélation, parcourez vos données pour rechercher les valeurs extrêmes (qui peuvent provoquer des résultats erronés) et les traces d'une relation linéaire. Le coefficient de corrélation de Pearson est une mesure d'association linéaire. Deux variables peuvent être parfaitement liées, mais si la relation n'est pas linéaire, le coefficient de corrélation de Pearson n'est pas une statistique appropriée pour mesurer leur association.

Les paramètres de l'intervalle de confiance sont disponibles pour Pearson et Spearman.

Exemple
Le nombre de matchs de basket-ball remportés par une équipe est-il lié au nombre moyen de points marqués par match ? Un nuage de points indique qu'il existe une relation linéaire. L'analyse des données de la saison NBA 1994–1995 démontre que le coefficient de corrélation de Pearson (0.581)

est significatif au niveau 0.01. On peut penser que plus on a gagné de matchs dans une saison, moins l'adversaire a marqué de points. Ces variables sont liées négativement (–0.401) et la corrélation est significative au niveau 0.05.

Statistiques
Pour chaque variable : nombre d'observations avec des valeurs non manquantes, de moyenne et d'écart type. Pour chaque paire de : coefficient de corrélation de Pearson, rho de Spearman, tau-*b* de Kendall, déviation des produits en croix et covariance.

Remarques sur les données

Données
Utilisez des variables quantitatives symétriques pour le coefficient de corrélation de Pearson, et des variables quantitatives ou des variables avec des catégories ordonnées pour le rho de Spearman et le tau-*b* de Kendall.

Hypothèses
Le coefficient de corrélation de Pearson part du principe que chaque paire de variables est gaussienne bivariée.

Obtention de corrélations bivariées

A partir des menus, sélectionnez :

Analyse > **Corrélation** > **Bivariée...**

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

1. Sélectionnez plusieurs variables numériques.

   Les options suivantes sont également disponibles :

   Coefficients de corrélation
   Pour des variables quantitatives, normalement distribuées, choisissez le coefficient de corrélation de **Pearson**. Si vos données ne sont pas distribuées normalement ou si elles comportent des catégories ordonnées, choisissez le **Tau-b de Kendall** ou la corrélation de **Spearman**, qui mesure l'association entre les ordres de rangs. Les coefficients de corrélation vont de la valeur –1 (relation négative parfaite) à +1 (relation positive parfaite). La valeur 0 indique l'absence de relation linéaire. Lors de l'interprétation de vos résultats, vous ne pouvez pas, à partir de l'existence d'une corrélation significative, conclure en l'existence d'une relation de cause à effet.

   Test de signification
   Vous pouvez choisir des probabilités bilatérales ou unilatérales. Si la direction de l'association est connue à l'avance, choisissez **Unilatéral**. Sinon, sélectionnez **Bilatéral**.

   Repérer les corrélations significatives
   Les coefficients de corrélation significatifs au niveau 0,05 sont identifiés par un seul astérisque et ceux qui sont significatifs au niveau 0,01 sont identifiés par deux astérisques.

   Afficher uniquement le triangle inférieur
   Si cette option est sélectionnée, seul le triangle inférieur du tableau de la matrice de corrélation est présenté dans la sortie. Si elle n'est pas sélectionnée, l'intégralité du tableau de la matrice de corrélation est présenté dans la sortie. Ce paramètre permet à la sortie du tableau de respecter les instructions de style APA.

   Afficher la diagonale
   Si cette option est sélectionnée, le triangle inférieur du tableau de la matrice de corrélation et les valeurs diagonales sont présentés dans la sortie. Ce paramètre permet à la sortie du tableau de respecter les instructions de style APA.

2. Vous pouvez aussi sélectionner les options suivantes :

   - Cliquez sur **Options...** pour spécifier les statistiques de corrélation de Pearson et les paramètres des valeurs manquantes.

- Cliquez sur **Style...** pour spécifier des conditions permettant de modifier automatiquement les propriétés des tableaux croisés dynamiques en fonction de conditions spécifiques.
- Cliquez sur **Bootstrap...** pour obtenir des estimations robustes des erreurs types et des intervalles de confiance pour des estimations telles que la moyenne, la médiane, la proportion, le rapport de cotes, le coefficient de corrélation ou le coefficient de régression.
- Cliquez sur **Intervalle de confiance...** pour définir les options d'estimation des intervalles de confiance.

Options de corrélations bivariées

Statistiques
Pour les corrélations de Pearson, vous pouvez choisir l'une des options suivantes (ou les deux) :

Moyennes et déviations standard
Affichés pour chaque variable. Le nombre d'observations avec valeurs non manquantes est également affiché. Les valeurs manquantes sont examinées variable par variable quel que soit votre réglage des valeurs manquantes.

Ecarts des produits croisés et covariances
Indiqués pour chaque paire de variables. Le produit des déviations est égal à la somme des produits des variables moyennes corrigées. Ceci est le numérateur du coefficient de corrélation de Pearson. La covariance est une mesure non standardisée de la relation entre deux variables, égale à la déviation des produits en croix divisée par *N*–1.

Valeurs manquantes
Vous pouvez choisir l'un des éléments suivants :

Exclure seulement les composantes non valides
Les observations avec des valeurs manquantes pour l'une ou les deux variables d'une paire pour un coefficient de corrélation sont exclues de l'analyse. Etant donné que chaque coefficient est basé sur toutes les observations ayant des codes valides pour cette paire particulière de variables, la quantité maximale d'informations disponibles est utilisée dans chaque calcul. Ceci peut aboutir à un jeu de coefficients basé sur un nombre variable d'observations.

Exclure toute observation incomplète
Les observations avec des valeurs manquantes pour une variable sont exclues de toutes les analyses.

Intervalle de confiance des corrélations bivariées

La boîte de dialogue Intervalle de confiance propose des options permettant d'estimer des intervalles de confiance. Elle est disponible lorsque l'option **Pearson**, **Tau-b de Kendall** ou **Spearman** est sélectionnée dans la boîte de dialogue Corrélations bivariées.

Estimer l'intervalle de confiance du paramètre de corrélation bivariée
Contrôle l'estimation de l'intervalle de confiance du paramètre de corrélation bivariée. Si cette option est sélectionnée, l'intervalle de confiance est estimé.

Intervalle de confiance (%)
Spécifie le niveau de confiance pour tous les intervalles de confiance générés. Indiquer une valeur numérique comprise entre 0 et 100. La valeur par défaut est 95.

Corrélation de Pearson
Le paramètre **Appliquer l'ajustement de biais** contrôle si l'ajustement de biais est appliqué. Par défaut, ce paramètre n'est pas sélectionné, et le biais n'est pas pris en considération. Lorsqu'il est sélectionné, l'ajustement du biais par rapport à l'estimation des limites de confiance est appliqué. Ce paramètre est disponible lorsque l'option **Pearson** est sélectionnée dans la boîte de dialogue Corrélations bivariées.

Corrélation de Spearman
Ce paramètre est disponible lorsque l'option **Spearman** est sélectionnée dans la boîte de dialogue Corrélations bivariées, et propose des options permettant d'estimer la variance de corrélation de Spearman via les méthodes suivantes :

- **Fieller, Hartley et Pearson**
- **Bonett et Wright**
- **Coruso et Cliff**

Fonctions supplémentaires des commandes CORRELATIONS et NONPAR CORR

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

- Ecrire une f pour les corrélations de Pearson qui peut être utilisée à la place de données brutes pour obtenir d'autres analyses comme une analyse factorielle (avec la sous-commande `MATRIX`).
- Obtenir des corrélations de chaque variable dans une liste avec chaque variable d'une seconde liste (en utilisant le mot clé `WITH` avec la sous-commande `VARIABLES`).

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

Mappage de proximité

Cette documentation guide les utilisateurs dans la spécification des données d'entrée, la sélection du modèle, les fonctions de transformation, les paramètres de dimensionnalité et l'inclusion d'attributs et de propriétés dans la cartographie de proximité.

La cartographie de proximité est une technique de visualisation utilisée pour réduire la dimensionnalité des données multivariées et pour afficher les relations entre les objets (cas, articles ou autres entités) dans une configuration spatiale. Développée à l'origine sous le nom d'échelle multidimensionnelle (MDS) en psychologie, la méthode utilise les proximités (similitudes, dissimilitudes ou distances) pour positionner les objets de manière à ce que leur disposition spatiale reflète la structure des données. Les objets similaires apparaissent proches les uns des autres; les objets dissemblables sont placés plus loin les uns des autres. La cartographie de proximité facilite l'interprétation et la reconnaissance des modèles en transformant des données relationnelles complexes en graphiques intuitifs en deux ou trois dimensions.

Dans IBM SPSS Statistiques, la cartographie de proximité est mise en œuvre par le biais de la procédure PROXMAP. Contrairement à PROXSCAL, qui se limite à l'analyse des matrices de proximité, PROXMAP accepte de multiples formes d'entrée et offre une plus grande flexibilité. Il prend en charge plusieurs sources de proximité, intègre des variables supplémentaires (attributs et propriétés) et offre un large éventail d'options de transformation et de restriction. Ces caractéristiques font de PROXMAP un outil puissant pour l'exploration des structures multivariées et l'intégration de divers types de données dans une représentation spatiale unifiée.

Lorsque des informations sur la proximité sont disponibles à partir de sources multiples, soit directement spécifiées, soit dérivées de données multivariées, PROXMAP peut ajuster plusieurs modèles de différences individuelles en plus du *modèle d'identité* par défaut. Les modèles disponibles sont le modèle de l'identité pondérée (dilatation), le modèle euclidien pondéré (diagonale) et le modèle euclidien généralisé avec des restrictions de rang optionnelles.

Par rapport aux techniques traditionnelles de réduction des dimensions telles que l'analyse des composantes principales (ACP) et le MDS classique, PROXMAP fournit une représentation plus directe et plus souple des proximités. Le MDS classique peut être utilisé pour initialiser la configuration. Lorsqu'il est appliqué à des données multivariées, le MDS classique est équivalent à l'analyse des coordonnées principales, une variante de l'ACP.

PROXMAP accepte les types de données suivants.

Type d'entrée	Descriptif
Proximités brutes et proximités dérivées	Les proximités dérivées sont calculées à partir de variables numériques, ordinales ou nominales.

Type d'entrée	Descriptif
	Les deux types de proximités sont éventuellement transformés à l'aide de fonctions monotones (par exemple, ordinales ou basées sur des splines), avant d'être approximés par des distances.
Attributs	Variables qui contraignent la configuration et apparaissent comme des directions dans l'espace articulaire (cartographie supervisée).
Propriétés	Variables supplémentaires utilisées pour l'interprétation; elles sont affichées dans l'espace mais n'influencent pas la configuration (cartographie non supervisée).

Une caractéristique unique de PROXMAP est la possibilité d'utiliser le même ensemble de variables pour différents rôles :

- Comme base de la dérivation de proximité
- En tant qu'attributs qui façonnent la configuration
- En tant que propriétés qui soutiennent l'interprétation

Il est également possible d'attribuer des ensembles distincts de variables à chaque rôle.

Outre la transformation linéaire, PROXMAP propose diverses options de transformation (y compris les transformations ordinales, splines monotones, splines non monotones et nominales) pour les attributs et les propriétés. Pour les proximités, les transformations doivent être monotones et inclure des transformations de puissance. Les résultats comprennent des diagnostics graphiques tels que des diagrammes de contraintes, des arbres de recouvrement minimum et des graphiques de voisinage pour aider à l'interprétation et à l'évaluation.

Exemple

La cartographie de proximité peut être appliquée à de nombreuses disciplines, notamment la psychologie et les sciences du comportement (perception, mémoire, préférences), le marketing (cartographie des marques, segmentation), la sociologie (cartographie des idéologies, réseaux sociaux), l'éducation et la psychométrie (diagnostics de tests, profilage des compétences), la linguistique (similarité sémantique, cartographie des dialectes), Géographie (distances perçues, analyse spatiale écologique), génomique et bioinformatique (modèles d'expression, structure de la population, exploration de l'hétérogénéité de l'expression d'une cellule unique), chimio-informatique (similarité moléculaire), informatique (visualisation de l'intégration), anthropologie et archéologie (similarité des artefacts, données crâniennes).

Données

Les données à analyser comprennent des variables qui représentent soit une matrice de proximité (ou des matrices), soit des données multivariées converties en matrice(s) de proximité.

Pour les données multivariées, les variables peuvent être un mélange de variables d'échelle, ordinales et nominales. Le nombre minimum de variables nécessaires est de 3

Pour les données de proximité, fournir le même nombre de variables que le nombre de cas.

Pour les données de proximité et les données multivariées, le nombre minimum de cas nécessaires est de 3.

La sortie comprend des configurations spatiales, divers diagrammes de diagnostic et des statistiques d'ajustement. Les biplots affichent des objets avec des variables, des attributs et des propriétés dans un espace commun.

Pour lancer l'analyse PROXMAP, cliquez sur **Analyser** > **Cartographie** > **Cartographie de proximité**.

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

Cartographie de proximité : Données

L'onglet **Données** permet de définir une ou plusieurs sources de proximité basées sur les variables de l'ensemble de données. Chaque source est créée en sélectionnant des variables dans l'ensemble de données actif et en spécifiant si elles représentent des matrices de proximité ou des données multivariées à partir desquelles les proximités seront dérivées.

Variables
La liste **Variables** affiche toutes les variables de l'ensemble de données actif. Les variables comprennent des variables numériques, ordinales et nominales pouvant être utilisées dans le calcul de proximité. Sélectionnez une ou plusieurs variables et déplacez-les dans **Variables in Source.**

Variables dans la source
La liste **Variables in Source** répertorie les variables actuellement sélectionnées pour la source de proximité active. Ces variables définissent la structure de la source et sont utilisées pour dériver les proximités (pour les données multivariées) ou pour spécifier une matrice de proximité (si les proximités sont fournies directement). Utilisez les boutons fléchés vers le haut ou vers le bas pour réorganiser les variables ou les supprimer si nécessaire.

Libellés d'objets
Avec **Étiquettes d'objets**, vous pouvez spécifier une variable qui fournit des étiquettes pour les objets (cas) à afficher dans les graphiques et les tableaux de sortie. Sélectionnez une variable dans la liste et déplacez-la vers **Object Labels** pour l'affecter.

définition de source
À droite du champ **Variables in Source**, les champs Source name et Source label définissent l'identifiant interne et l'étiquette descriptive de la proximité actuelle source.The Le **nom de la source** doit être un identifiant unique précédé de *$.*

Les boutons suivants permettent de gérer les sources.

- **Ajouter une source** crée une nouvelle source de proximité à partir des variables du panneau Variables dans la source.
- **Modifier la source** modifie la source actuellement sélectionnée en utilisant la sélection de variables mise à jour.
- **Remove Source (Supprimer la source**) supprime la source actuellement sélectionnée de la liste.

Type de données
Utilisez le sélecteur de **type de données** pour spécifier comment les variables sélectionnées sont interprétées. Le groupe **Type de données** permet d'indiquer comment les proximités sont dérivées des variables sélectionnées :

- **Données de proximité** Utilisez cette option si votre entrée est déjà une matrice de proximité.
- **Données multivariées** (par défaut) traite les variables sélectionnées comme des caractéristiques des objets. PROXMAP calcule les proximités en calculant les distances entre les objets sur la base de ces variables.
  - **Euclidienne** : Les proximités sont basées sur des distances euclidiennes non standardisées.
  - **Euclidienne normalisée** : Les variables sont normalisées avant de calculer les distances (par défaut).

Pondérations
Lorsque le **type de données** est défini sur les **données de proximité**, une liste optionnelle de **poids de proximité** est disponible. Vous pouvez ici spécifier des variables de poids supplémentaires à associer aux valeurs de proximité. Ces poids ajustent l'influence de chaque proximité dans le processus d'optimisation.

Cartographie de proximité : Poids

L'onglet **Pondérations** permet de spécifier les pondérations de proximité (cellule) pour chaque source et les pondérations d'objet (cas) pour les données de proximité. Ces poids influencent la manière dont les proximités contribuent au processus d'optimisation. Cet onglet n'est pertinent que lorsque le **type de données Données de proximité** est sélectionné dans l'onglet **Données.**

Variables
La liste **Variables** affiche toutes les variables de l'ensemble de données actif. Ces variables doivent contenir les poids de proximité. Sélectionnez une ou plusieurs variables et déplacez-les dans la rubrique **Pondérations dans la source.** Ces variables doivent contenir les poids de proximité pour une source spécifique.

Poids dans la source
Pondérations dans la source affiche les variables de pondération sélectionnées qui sont associées à la source active. Ces variables définissent une matrice de poids de proximité qui correspond à la structure de la matrice de proximité correspondante. Le nombre de variables de poids doit être égal au nombre de variables de la matrice source.

Utilisez **Ajouter des poids**, **Modifier des poids** ou **Supprimer des poids** pour gérer les définitions de poids pour la source sélectionnée dans **Sources**.

Source(s)
La liste **Source(s)** affiche toutes les sources de proximité définies auxquelles des pondérations peuvent être attribuées. Sélectionnez une source pour afficher ou modifier les poids qui lui sont associés.

Pondérations d'objet (casse)
Le champ Étiquettes d'objets, situé sous les panneaux de sélection, permet de définir une variable contenant des poids pour des cas individuels (objets). Ces poids ajustent l'influence de chaque objet dans la fonction de perte lorsque des données de proximité sont utilisées. Les pondérations de cas ne sont pas applicables aux données multivariées.

Fonction de pondération

La **fonction Pondérations** offre trois options pour des schémas de pondération de proximité spéciaux qui sont des fonctions des valeurs de proximité :

- **Aucune** : Aucune pondération particulière n'est appliquée. Seuls les poids de proximité et d'objet spécifiés par l'utilisateur (le cas échéant) sont utilisés.
- **Sammon** : applique des poids à la proximité entre les objets *i* et *j.* Ce poids met l'accent sur les petites proximités, comme dans la cartographie de Sammon.
- L' **énergie** : Applique des pondérations qui mettent encore plus l'accent sur les petites proximités.

Le poids total appliqué à chaque proximité est le produit des poids de proximité définis par l'utilisateur, des poids de cas et de la fonction de poids spéciale. Ces poids dérivés régissent l'importance relative de chaque proximité dans la fonction de perte qui est minimisée par PROXMAP.

Cartographie de proximité : Initiale

L'onglet **Initial** spécifie la configuration de départ du processus d'optimisation itérative utilisé dans PROXMAP. Une bonne configuration initiale peut accélérer la convergence et améliorer la qualité de la solution. Vous pouvez choisir entre une initialisation automatique utilisant une échelle classique sur les proximités brutes ou sur leur ordre de classement. Vous pouvez également fournir une configuration personnalisée à partir de l'ensemble de données.

Mise à l'échelle classique
Il s'agit de l'option par défaut. Une solution classique de mise à l'échelle (analyse des coordonnées principales) est calculée sur la base des proximités. Il fournit une configuration de départ dérivée analytiquement dans le nombre de dimensions spécifié.

Deux options supplémentaires permettent de contrôler la manière dont les proximités sont traitées lors de la mise à l'échelle classique :

- Avec la **valeur constante additive**, vous pouvez spécifier une constante fixe à ajouter à toutes les proximités avant la mise à l'échelle. Cela permet d'améliorer les propriétés métriques et d'éviter les valeurs propres négatives. La valeur ne doit pas donner lieu à des proximités négatives.
- Avec l'option **Estimer la constante additive**, PROXMAP estime automatiquement la constante optimale à ajouter. Cela garantit que la matrice de distance résultante est semi-définie positive.
- Avec **Use Rank Numbers for Proximities**, vous pouvez appliquer une échelle classique aux rangs des proximités plutôt qu'aux valeurs brutes. Cela permet d'obtenir une configuration de départ `ordinal` et de réduire la sensibilité aux valeurs aberrantes.

Configuration personnalisée

Avec la **configuration personnalisée**, vous pouvez fournir directement vos propres coordonnées initiales. Les variables de l'ensemble de données sont sélectionnées comme dimensions de la configuration de départ. Ces variables doivent contenir des valeurs numériques valides sans données manquantes.

La liste des **variables** contient toutes les variables disponibles. Sélectionnez les variables qui contiennent la configuration personnalisée et déplacez-les vers **Dimension(s).**

La liste des **dimensions** affiche les variables sélectionnées pour être utilisées comme coordonnées initiales. Le nombre de variables sélectionnées doit correspondre au nombre de dimensions. Si plusieurs variables sont fournies, les valeurs sont projetées sur les *p* premières composantes principales, où *p* est le nombre de dimensions.

Nombre actuel de dimensions affiche la dimensionnalité spécifiée pour la solution finale (spécifiée dans l'onglet **Critères**).

Cartographie de proximité : Modèle

L'onglet **Modèle** permet de spécifier le modèle de différences individuelles à appliquer lors de l'analyse de sources de proximité multiples. Il définit également la manière dont les transformations de proximité sont traitées entre les sources : soit de manière uniforme (inconditionnelle), soit de manière indépendante (conditionnelle à la matrice).

Modèle à trois voies

Les paramètres du **modèle à trois voies** déterminent la manière dont les proximités provenant de sources multiples sont mises en correspondance dans une représentation spatiale commune, en utilisant des pondérations spatiales. Le modèle détermine comment les pondérations spatiales sont appliquées à l'espace commun. Pour certains modèles, les pondérations d'espace sont équivalentes aux pondérations d'espace. Lorsqu'une seule source est définie, toutes les options, à l'exception de l' **identité**, sont désactivées.

Option	Descriptif
Identité	Suppose que toutes les sources partagent la même configuration et qu'il n'y a pas de mise à l'échelle spécifique à la source. Disponible uniquement lorsqu'une seule source est spécifiée.
Dilatation (identité pondérée)	Chaque source est modélisée avec un seul facteur d'échelle appliqué uniformément à toutes les dimensions. Grâce à cette option, chaque source peut percevoir l'espace commun à une échelle globale différente.
Euclidienne pondérée (diagonale)	Chaque source attribue un poids différent aux dimensions de l'espace commun. Cela permet aux sources de mettre l'accent sur différentes caractéristiques spatiales.

Option	Descriptif
Euclidien généralisé	Chaque source a droit à une version pivotée et repondérée (mise à l'échelle) de l'espace commun. Il s'agit du modèle le plus souple, qui permet de prendre en compte un large éventail de différences individuelles. • **Rang complet** : Chaque source utilise toute la dimension de l'espace commun. • **Rang réduit** : Lorsque vous appliquez une contrainte de rang réduit, la configuration de chaque source (espace individuel, avecrang = nombre de dimensions - 1) est représenté dans un sous-espace de dimension inférieure de l'espace commun. Cette approche implique une rotation orthogonale de la configuration commune, suivie d'une projection sur un sous-espace. Ici, la rotation et la projection forment ensemble ce que l'on appelle les poids spatiaux.

Conditionnalité/Transformation

Ce volet contrôle la façon dont les proximités sont transformées avant d'ajuster les distances dans la configuration.

- L'option **Inconditionnel** permet d'appliquer une seule fonction de transformation à toutes les sources. Cette option est appropriée lorsque les sources sont supposées refléter une structure commune.

  La transformation par défaut est affichée sous **Transformation** (ordinale avec des liens autorisés à être déliés).

  Cliquez sur **Définir la transformation** pour ouvrir une boîte de dialogue dans laquelle vous pouvez choisir entre tous les types de transformation (linéaire, puissance, monotone, spline, ordinale), la gestion des liens et d'autres paramètres.
- L'option **Matrice-Conditionnelle** permet de choisir différentes fonctions de transformation pour différentes sources ou groupes de sources. Lorsque cette option est sélectionnée, chaque source est répertoriée avec la fonction de transformation par défaut.

  Cliquez sur **Définir la transformation** pour choisir une autre fonction à appliquer aux sources sélectionnées. Cette option n'est activée que lorsque plusieurs sources sont définies.

Dans la section **Définir la transformation**, vous pouvez sélectionner la façon dont les proximités sont transformées de manière optimale avant d'être mises en correspondance avec les distances dans la configuration spatiale. Les options suivantes sont disponibles :

Option	Descriptif
Aucun	Aucune transformation n'est appliquée. Les proximités sont utilisées telles quelles.
Linéaire	Applique une transformation en ligne droite aux proximités. **Intercept** (case à cocher) permet d'ajouter une constante à la transformation, ce qui permet une modélisation à l'échelle de l'intervalle. Si la case n'est pas cochée, la transformation est mise à l'échelle (aucune constante n'est ajoutée).

Option	Descriptif
Puissance	Applique une transformation de puissance. **Intercept** (case à cocher) permet d'ajouter un décalage constant à la fonction puissance, si elle est activée.
Spline	Ajuste une fonction polynomiale lisse par morceaux (spline) aux proximités. L' **ordonnée à l'origine** inclut un terme constant dans la base spline. Le **degré** définit la complexité des segments polynomiaux (par exemple, 1 = linéaire, 2 = quadratique). **Nœuds intérieurs** contrôle le nombre de points intérieurs utilisés pour définir la forme de la spline.
Séquence de noeud	Intervalles égaux : Les nœuds sont régulièrement espacés le long de la zone de proximité. Percentiles : Les nœuds sont placés aux quantiles empiriques de la distribution de proximité.
Ordinal	Applique une transformation préservant le rang (monotone). **Permettre de dénouer les liens** : Les proximités liées peuvent être transformées en différentes valeurs. **Maintenir les liens attachés** : Les proximités liées restent liées après la transformation.

Cartographie de proximité : Restrictions

L'onglet **Restrictions** permet de restreindre l'espace de solution en spécifiant des variables d'attributs. Ces variables guident activement la configuration en limitant l'espace commun des objets à une combinaison linéaire des attributs. Cela permet une cartographie supervisée, qui améliore l'interprétabilité en ancrant les dimensions à des variables externes significatives.

Sélectionnez l'une des options suivantes.

Aucun
Aucune restriction n'est imposée. La configuration est déterminée uniquement par les proximités (dérivées).

Attributes
Permet la mise en correspondance supervisée en utilisant une ou plusieurs variables externes comme contraintes d'attributs. Lorsque cette option est sélectionnée, vous pouvez définir les variables qui agissent en tant qu'attributs et spécifier comment chacune d'entre elles est transformée.

Variables

La liste **Variables** sur la gauche affiche toutes les variables disponibles dans l'ensemble de données. Sélectionnez une ou plusieurs variables et cliquez sur le bouton flèche pour les déplacer vers **Attributs**.

Attributes

Le volet de droite affiche toutes les variables d'attributs sélectionnées ainsi que la fonction de transformation par défaut. Ces transformations contrôlent la manière dont les attributs sont mis à l'échelle ou transformés pour être utilisés dans la configuration restreinte.

Cliquez sur **Définir la transformation** pour attribuer une fonction de transformation à l'attribut sélectionné. Les options suivantes sont disponibles :

- **Linéaire** maintient l'attribut sur son échelle normalisée.
- **Spline** applique une transformation lisse, monotone ou non monotone, à l'aide d'une fonction polynomiale par morceaux.
- L' **ordinal** ne conserve que l'ordre des rangs.
- **Nominal** traite les catégories comme étant non ordonnées et les adapte de manière optimale. Donne une transformation unique (contrairement à la transformation nominale pour les propriétés).

Vous pouvez sélectionner plusieurs lignes dans le champ **Attributs** pour définir la même transformation pour plusieurs attributs.

L'espace commun est estimé sous la contrainte que les coordonnées de l'objet se trouvent dans le sous-espace couvert par les attributs transformés. Le champ **Nombre actuel de dimensions** affiche la dimensionnalité de la solution, qui est définie dans l'onglet **Critères**. Sélectionnez toujours un nombre d'attributs supérieur au nombre de dimensions.

Cartographie de proximité : Complémentaire

L'onglet **Supplémentaire** permet de définir des variables de propriété qui aident à l'interprétation de la configuration spatiale mais n'influencent pas son estimation (cartographie non supervisée).

Variables
La liste **Variables** sur la gauche affiche toutes les variables disponibles dans l'ensemble de données actif. Sélectionnez une ou plusieurs variables que vous souhaitez traiter comme des propriétés et déplacez-les dans la liste **Propriétés**. Il est possible de sélectionner un nombre quelconque de propriétés.

Propriétés

Le volet **Propriétés** répertorie toutes les variables désignées comme variables de propriété, ainsi que la transformation attribuée à chacune d'entre elles.

Cliquez sur **Définir la transformation** pour attribuer une fonction de transformation à une ou plusieurs variables de propriété sélectionnées. Voir les options disponibles suivantes :

- **Linear** maintient les propriétés sur son échelle standardisée.
- **Spline** applique une transformation lisse, monotone ou non monotone, à l'aide d'une fonction polynomiale par morceaux.
- L' **ordinal** ne conserve que l'ordre des rangs.
- Les catégories d'ajustements **nominaux** sont les centroïdes des points d'objets associés dans la configuration.

Cartographie de proximité : Critères

L'onglet **Critères** spécifie les paramètres algorithmiques qui contrôlent la dimensionnalité de la solution, le comportement de convergence du processus d'optimisation, ainsi que la normalisation et l'orientation post-solution de la configuration résultante.

Dimensionnalité
Définit le nombre de dimensions dans l'espace commun. La valeur doit être supérieure ou égale à 2 et inférieure à l'indicenombre d'objets - 1. L'augmentation du nombre de dimensions peut améliorer l'adéquation, mais elle peut réduire l'interprétabilité.

Convergence d'algorithme
Ces paramètres déterminent le moment où l'optimisation itérative s'arrête.

Zone	Descriptif	Valeur par défaut
Minimum de fonction	Définit le seuil de la valeur de la fonction de perte en dessous duquel la solution est acceptée. Si la valeur de la fonction de perte passe en dessous de ce minimum, l'itération s'arrête. La valeur admise doit être supérieure ou égale à 0.	1.0×10^{-6}
Différence de fonction	Définit le critère de convergence basé sur la variation de la fonction de perte entre les itérations. Si la différence absolue entre les valeurs des fonctions est inférieure à cette valeur, l'algorithme est considéré comme convergent. La valeur admise doit être supérieure ou égale à 0.	1.0×10^{-12}
Différence de configuration	Fixe le seuil de convergence pour les changements de configuration entre les itérations. Si le mouvement total des positions de l'objet est inférieur à cette valeur, l'algorithme s'arrête. La valeur admise doit être supérieure ou égale à 0.	1.0×10^{-6}
Nombre maximal d'itérations	Spécifie le nombre maximum d'itérations que l'algorithme peut effectuer. Si la convergence n'est pas atteinte avant ce nombre, l'itération s'arrête quand même. La valeur admise doit être supérieure ou égale à 0.	1 000

Standardisation

Spécifie comment les variables multivariées (si elles sont utilisées pour dériver les proximités) sont standardisées avant le calcul de la distance. Cette option s'applique également aux variables d'attributs et de propriétés.

- Par défaut, l'option **Somme des carrés = N** est sélectionnée.
- Sélectionnez **Somme des carrés = N–1** si vous souhaitez une cohérence avec d'autres procédures statistiques.

Dimension

Spécifie comment la configuration de l'espace commun est orientée dans l'espace des coordonnées.

Bouton radio	Descriptif
Aucun/En l'état	Aucune contrainte n'est appliquée. L'orientation est celle des axes principaux, les signes des dimensions étant arbitraires.

Bouton radio	Descriptif
Corrélation positive entre les dimensions	Définit la réflexion de la configuration pour maximiser la corrélation positive entre les dimensions et les données.
Quadrant positif pour la coordonnée absolue la plus élevée	Réfléchit les dimensions pour s'assurer que l'objet ayant la plus grande coordonnée apparaît dans le quadrant positif de l'espace.
Par dimension	Vous pouvez définir des réflexions individuellement pour chaque dimension. Cochez la case située à côté d'une dimension pour appliquer la réflexion à cet axe spécifique.

Cartographie de proximité : Sortie

Dans le sous-onglet **Sortie**, vous pouvez demander des tableaux et des graphiques qui résument les données et les résultats de l'analyse PROXMAP. Il contient des sous-onglets pour chacun des principaux onglets d'analyse (**Données**, **Initial**, **Modèle**, **Restrictions**, **Supplémentaire** et **Critères**)

Vous pouvez demander des résultats liés à n'importe quel aspect de l'analyse. Chaque sous-onglet comporte des cases à cocher permettant de demander des tableaux ou des graphiques spécifiques.

Données

Dans l'onglet **Données**, vous pouvez demander des résultats liés aux données : données multivariées, proximités et poids.

Tableaux (données)

Option	Descriptif
Données multivariées d'origine	Affiche les valeurs brutes des variables utilisées pour calculer les proximités.
Proximités d'origine	Affiche les proximités telles qu'elles ont été fournies directement par l'utilisateur (le cas échéant).
Proximités dérivées	Montre les proximités calculées à partir des données multivariées pour toutes les sources.
Proximités combinées	Affiche les proximités agrégées si plusieurs sources sont combinées.

Tableaux (poids)

Option	Descriptif
Pondérations d'origine	Liste les poids de proximité bruts (fournis par l'utilisateur) utilisés pour chaque source.
Pondérations dérivées	Fournit des listes de poids qui sont dérivés des poids bruts, des poids des cas d'objets, des données manquantes et de la fonction de poids appliquée aux proximités.
Poids des objets (cas)	Affiche les poids appliqués par objet (si spécifié).

Graphiques (données)

L' **histogramme** des proximités représente la distribution des valeurs de proximité pour l'ensemble des sources.

Graphiques (poids)

L' **histogramme** des poids représente la distribution des poids de proximité (si les poids sont fournis ou dérivés).

Initial

Le sous-onglet **Initial** permet de demander des résultats relatifs à la configuration initiale, à la décomposition de l'ajustement et aux mesures de diagnostic après l'estimation initiale.

Tableaux

Option	Descriptif
Coordonnées	Affiche les coordonnées des objets dans la configuration initiale.
Directions variables	Indique les directions vectorielles des variables originales dans l'espace de configuration (pour les données multivariées). Donne des coordonnées multiples pour les variables nominales (si elles sont présentes dans les données multivariées).
Directions des propriétés	Affiche les directions des variables de propriété (activée uniquement lorsque des propriétés sont définies). Donne des coordonnées multiples pour les variables nominales (si elles sont présentes dans les données multivariées).
Décomposition de la contrainte	Indique la contribution de chaque objet à la valeur globale de la perte (contrainte).
Diagnostics	Inclut des statistiques d'ajustement et des diagnostics pour évaluer la solution initiale. Il s'agit notamment des valeurs propres et des vecteurs propres pour le début de la mise à l'échelle classique.

Graphiques

Option	Descriptif
configuration	Affiche une représentation graphique de l'espace objet.
Tracé double multivarié	Combine les points de l'objet et les directions des vecteurs variables dans un tracé commun, disponible pour les données multivariées.
Tracé double de propriétés	Visualise la relation entre les objets et les variables des propriétés (disponible lorsque les propriétés sont définies).
Tracé d'effondrement	Graphique des valeurs propres de la solution classique de mise à l'échelle et des dimensions restantes.
Tracé de la contrainte	Un diagramme à bulles superposé à l'espace de configuration de l'objet. Chaque objet est

Option	Descriptif
	représenté par un point entouré d'un cercle. La taille de chaque cercle reflète la contribution relative de cet objet à la contrainte normalisée globale de la configuration.

Modèle

Le sous-onglet **Modèle** permet de demander des sorties graphiques et tabulaires relatives aux espaces communs et individuels, aux transformations spécifiques aux sources et aux mesures d'ajustement du modèle.

Tableaux

Option	Descriptif
Coordonnées	Liste les coordonnées des objets dans l'espace commun final.
Directions variables	Liste les directions des variables originales (pour les données multivariées). Donne des coordonnées multiples pour les variables nominales (si elles sont présentes dans les données multivariées).
Pondérations d'espace	Affiche les pondérations spatiales estimées pour chaque source dans le cadre de modèles tenant compte des différences individuelles.

Graphiques (Espace commun)

Option	Descriptif
Configuration (sélectionnée par défaut)	Affiche la configuration spatiale commune des objets. **La configuration** est un paramètre par défaut qui est inclus dans la sortie, sauf si vous sélectionnez d'autres options.
Tracé double multivarié	Indique que les données relatives à la configuration et aux directions sont multivariées.
Tracé double à rang réduit	Affiche les projections spécifiques à la source dans l'espace commun bidimensionnel si le modèle inclut des espaces individuels de rang réduit de rang 1. Ce graphique combine l'espace commun et les espaces individuels en un seul graphique.
Arbre couvrant minimum	Relie tous les objets de la configuration avec la longueur totale d'arête la plus courte possible. La TMS permet de visualiser la structure sous-jacente de la configuration, en mettant en évidence les groupements et les transitions basés sur la proximité.
Voisins	Un graphe de voisinage relie chaque objet de la configuration à ses k voisins les plus proches, sur la base des distances dans l'espace de configuration. Ce graphique met l'accent sur la structure locale et permet de détecter les

Option	Descriptif
	groupes ou les valeurs aberrantes dans la configuration. La valeur du nombre de voisins les plus proches peut être choisie.
Nombre maximal de pages	Relie les paires d'objets dans l' **espace de configuration** dont les proximités originales ou transformées sont inférieures ou égales à un seuil spécifié. La valeur du seuil peut être choisie.

Graphiques (espace individuel)

Option	Descriptif
configuration	Trace les configurations de l'espace individuel pour les sources sélectionnées (utilisé avec les modèles de différences individuelles).
Pondérations d'espace	Affiche les poids des dimensions pour le modèle de dilatation et le modèle euclidien pondéré.

Graphiques (mesures de transformation et d'ajustement)

Option	Descriptif
Transformations	Affiche les proximités transformées par rapport aux proximités originales pour les sources sélectionnées.
Résidus	Représente les différences entre les proximités transformées et les distances ajustées par rapport aux proximités transformées.
Ajustement	Affiche les distances ajustées par rapport aux distances transformées par source.
Shepard	Crée un diagramme de Shepard qui compare les proximités et les distances ajustées par rapport aux proximités originales pour les sources sélectionnées.

Restrictions

Le sous-onglet **Restrictions** fournit des données relatives à l'utilisation de variables d'attributs comme contraintes de configuration (mappage supervisé). Vous pouvez demander des tableaux et des visualisations pour les versions originales, transformées et ajustées de ces attributs.

Attributs de table

Option	Descriptif
Attributs d'origine	Affiche les valeurs brutes des variables d'attributs sélectionnées.
Attributs transformés	Liste les valeurs des attributs après application des transformations optimales.
Pondérations des attributs	Affiche les pondérations calculées pour les attributs pour la restriction de l'espace commun.

Option	Descriptif
Directions d'attribut	Liste les coordonnées vectorielles associées à chaque attribut dans l'espace commun.

Graphiques

Sélectionnez un ou plusieurs attributs dans le tableau suivant pour générer les graphiques suivants :

Option	Descriptif
Tracé double	Affiche la configuration de l'objet avec les vecteurs des variables d'attributs sélectionnées.
Transformations	Affiche la fonction de transformation appliquée à chaque attribut sélectionné.
Résidus	Affiche la différence entre les valeurs ajustées transformées et les valeurs transformées pour les attributs sélectionnés.

Le **graphique de projection** affiche une projection unidimensionnelle des objets de l'espace commun sur chacun des attributs séparément.

Supplémentaire

Le sous-onglet **Supplémentaire** fournit des données relatives aux variables de propriété, qui sont utilisées pour l'interprétation de la configuration mais n'affectent pas l'optimisation (cartographie non supervisée).

Propriétés des tableaux

Option	Descriptif
Propriétés d'origine	Affiche les valeurs brutes des variables de la propriété.
Propriétés transformées	Affiche les valeurs après application des transformations optimales.
Directions des propriétés	Liste les vecteurs directionnels des propriétés dans l'espace de configuration (le cas échéant).

Graphiques Propriétés

Sélectionnez une ou plusieurs propriétés dans la liste pour générer les graphiques suivants :

Option	Descriptif
Tracé double	Affiche la configuration de l'objet avec les vecteurs des variables de propriété sélectionnées.
Transformation	Affiche la transformation de chaque propriété sélectionnée.
Résidus	Trace la différence entre les valeurs transformées - ajustées et les valeurs transformées pour les propriétés sélectionnées.

Le **graphique de projection** affiche une projection unidimensionnelle des objets de l'espace commun sur chacune des propriétés séparément.

Critères

Le sous-onglet **Critères** fournit des informations de diagnostic et des visualisations liées au processus d'estimation itératif et à la convergence du modèle.

Tableaux

Option	Descriptif
Résumé (sélectionné par défaut)	Indique les valeurs de contrainte finales, le nombre d'itérations, le temps écoulé et les éventuels messages d'erreur. **Le résumé** est un paramètre par défaut qui est inclus dans la sortie, sauf si vous sélectionnez d'autres options.
Diagnostics	Liste des statistiques d'ajustement supplémentaires, y compris la formule de stress 1 de Kruskal, le coefficient d'aliénation, la dispersion prise en compte, le coefficient de congruence de Tucker, les corrélations de Pearson et de Spearman.
Historique des itérations	Affiche la contrainte normalisée pour chaque itération, la différence de contrainte entre les itérations et le changement de configuration (différence de coordonnées) entre les itérations. Pour limiter la taille du tableau, le tableau de l'historique des itérations ne rapporte qu'un sous-ensemble d'itérations, suivant les puissances de 2. Cela permet de saisir le modèle de convergence global sans énumérer toutes les itérations
Décomposition de la contrainte	Décompose la valeur globale de la contrainte en contributions provenant d'objets individuels.

Graphiques

Option	Descriptif
Historique des itérations	Affiche le tracé d'un sous-ensemble de valeurs pour les itérations, suivant les puissances de 2. Valeurs de la contrainte normalisée en fonction du nombre d'itérations.
Tracé de la contrainte	Un graphique à bulles montrant les contributions des objets dans l'espace commun en affichant des cercles autour des points de l'objet qui sont proportionnels à leur contribution à la contrainte normalisée.

Cartographie de proximité : Sauvegarder

L'onglet **Enregistrer** permet d'exporter les résultats sélectionnés vers un nouveau jeu de données.

Vous pouvez choisir de sauvegarder les éléments suivants :

Coordonnées initiales
Les coordonnées de la configuration initiale (espace initial).

Coordonnées finales
Les coordonnées de la configuration finale (espace commun).

Coordonnées du modèle
Les coordonnées des configurations individuelles (espaces individuels). L'option n'est activée que si le nombre de sources est supérieur à un.

Proximités dérivées
Les proximités dérivées des données multivariées, ou les proximités originales si elles sont fournies directement.

Proximités combinées
Les proximités qui sont agrégées sur toutes les sources (si plusieurs sources sont définies).

Proximités transformées
Les proximités après application des fonctions de transformation sélectionnées.

Attributs transformés
Les attributs transformés. L'option n'est activée que si des attributs sont spécifiés dans l'onglet **Restrictions**.

Attributs ajustés
Les valeurs d'attribut adaptées à la configuration. L'option n'est activée que si des attributs sont sélectionnés.

Propriétés transformées
Les valeurs transformées des propriétés supplémentaires. L'option n'est activée que si des propriétés sont définies sous l'onglet **Supplémentaire**.

Propriétés ajustées
Les valeurs des propriétés projetées sur la configuration. L'option n'est activée que si les **propriétés** sont sélectionnées.

Choisir un jeu de données

Les éléments sélectionnés sont enregistrés dans un nouvel ensemble de données. Cliquez sur le bouton d'ellipse (...) pour ouvrir la boîte de dialogue **Enregistrer les données sous**, dans laquelle vous pouvez spécifier le nom du fichier et le dossier de destination du nouveau jeu de données.

Corrélations partielles

La procédure Corrélations partielles calcule les coefficients de corrélation partielle qui décrivent le rapport linéaire entre deux variables tout en contrôlant les effets d'une ou de plusieurs autres variables. Les corrélations sont des mesures d'association linéaire. Deux variables peuvent être parfaitement liées mais, si leur rapport n'est pas linéaire, un coefficient de corrélation n'est pas une statistique adaptée pour mesurer leur association.

Exemple
Existe-t-il une relation entre le financement associé aux soins de santé et les taux d'attaque ? Contre toute attente, une étude fait état d'une corrélation *positive* : lorsque le financement associé aux soins de santé augmente, les taux d'attaque augmentent. Cependant, le contrôle du taux de visite aux fournisseurs de soins de santé supprime presque la corrélation positive observée. Le financement lié aux soins de santé et les taux d'attaque sont associés de manière positive car le nombre de personnes ayant accès aux soins de santé augmente en même temps que le financement. De ce fait, le nombre de maladies déclarées par les docteurs et les hôpitaux augmente également

Statistiques
Pour chaque variable : nombre d'observations avec des valeurs non manquantes, de moyenne et d'écart type. Matrices de corrélation partielle et simple, avec degrés de liberté et niveaux de signification.

Remarques sur les données

Données
Utiliser des variables quantitatives et symétriques.

Hypothèses
La procédure des Corrélations Partielles suppose que chaque paire de variables présente une corrélation normale.

Obtention de corrélations partielles

1. A partir des menus, sélectionnez :

   Analyser > **Corréler** > **Partiel...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins deux variables numériques pour lesquelles vous voulez calculer des corrélations partielles.
3. Sélectionnez une ou plusieurs variables numériques de contrôle.

Les options suivantes sont également disponibles :

Test de signification
Vous pouvez choisir des probabilités bilatérales ou unilatérales. Si la direction de l'association est connue à l'avance, choisissez **Unilatéral**. Sinon, sélectionnez **Bilatéral**.

Afficher le niveau de signification exact
La probabilité et les degrés de liberté sont affichés par défaut pour chaque coefficient de corrélation. Si vous désélectionnez cette option, les coefficients significatifs au niveau 0,05 sont identifiés par une astérisque, les coefficients significatifs au niveau de 0,01 par deux astérisques, et les degrés de liberté sont supprimés. Cette configuration affecte aussi bien les matrices de corrélation partielle que simple.

Options Corrélations partielles

Statistiques : Vous avez le choix entre les deux options suivantes :

- **Moyennes et écarts types :** Affichés pour chaque variable. Le nombre d'observations avec valeurs non manquantes est également affiché.
- **Corrélations simples :** Une matrice de corrélations simples entre toutes les variables, y compris les variables de contrôle, s'affiche.

Valeurs manquantes : Vous avez le choix entre les options suivantes :

- **Exclure toute observation incomplète :** Les observations ayant des valeurs manquantes pour une variable quelconque, y compris une variable de contrôle, sont exclues de tous les calculs.
- **Exclure seulement les composantes non valides :** Pour le calcul des corrélations simples sur lesquelles se basent les corrélations partielles, une observation ayant des valeurs manquantes pour une composante ou les deux composantes d'une paire de variables ne sera pas utilisée. La suppression des composantes non valides seulement utilise autant de données que possible. Le nombre d'observations peut toutefois différer selon les coefficients. Lorsque la suppression des composantes non valides seulement est sélectionnée, les degrés de liberté d'un coefficient partiel donné sont basés sur le plus petit nombre d'observations utilisées dans le calcul de l'une quelconque des corrélations d'ordre zéro.

Fonctions supplémentaires de la commande PARTIAL CORR

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

- Lire une matrice de corrélation d'ordre zéro ou écrire une matrice de corrélation d'ordre zéro (avec la sous-commande `MATRIX`).
- Obtenir des corrélations partielles entre deux listes de variables (en utilisant le mot-clé `WITH` dans la sous-commande `VARIABLES`).
- Obtenir des analyses multiples (avec les sous-commandes `VARIABLES`).

- Spécifier les valeurs des ordres à demander (par exemple, à la fois les corrélations partielles de premier et de second ordre) lorsque vous avez deux variables de contrôle (avec la sous-commande VARIABLES).
- Supprimer les coefficients redondants (avec la sous-commande FORMAT).
- Afficher une matrice de corrélations simples lorsque certains coefficients ne peuvent pas être calculés (avec la sous-commande STATISTICS).

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

Distances

La procédure Distances permet de calculer de très nombreuses statistiques mesurant les similitudes ou les différences (distances) entre des paires de variables ou d'observations. Vous pourrez ensuite utiliser ces mesures de similarité ou de dissimilarité avec d'autres procédures, comme l'analyse factorielle, l'analyse de cluster ou le positionnement multidimensionnel, afin de simplifier l'analyse des fichiers de données complexes.

Exemple. Est-il possible de mesurer les similarités entre des paires de voitures en fonction de certaines caractéristiques, comme la taille du moteur, la consommation et la puissance ? En calculant les similarités existant entre des voitures, vous pouvez déterminer les voitures qui sont semblables et celles qui sont différentes. Dans l'optique d'une analyse plus formelle, vous pouvez appliquer une analyse de cluster hiérarchique ou un positionnement multidimensionnel aux similarités afin d'examiner la structure sous-jacente.

Statistiques : Pour les données d'intervalle, les mesures de dissimilarité sont la distance Euclidienne, le carré de la distance Euclidienne, la distance de Tchebycheff, la distance de Manhattan (bloc), la distance de Minkowski ou une mesure personnalisée. Pour les données d'effectif, les mesures sont khi-deux et phi-deux. Pour les données binaires, les mesures de dissimilarité sont la distance Euclidienne, le carré de la distance Euclidienne, la différence de taille, la différence de motif, la variance, la forme, ou la mesure de Lance et Williams. Pour les données d'intervalles, les mesures de similarité sont la corrélation de Pearson ou cosinus. Pour les données binaires, il s'agit des mesures suivantes : Russel et Rao, indice de Sokal et Michener, Jaccard, Dice, Rogers et Tanimoto, Sokal et Sneath 1, Sokal et Sneath 2, Sokal et Sneath 3, Kulczynski 1, Kulczynski 2, Sokal et Sneath 4, Hamann, lambda, *D* d'Anderberg, *Y* de Yule, *Q* de Yule, Ochiai, Sokal et Sneath 5, corrélation phi tétrachorique ou dispersion.

Pour obtenir des matrices de distance

1. A partir des menus, sélectionnez :

   Analyse > **Corrélation** > **Distances...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au minimum une ou deux variables numériques pour calculer respectivement les distances existant entre des observations ou des variables.
3. Sélectionnez une possibilité dans le groupe Calculer les distances pour calculer les proximités existant entre des observations ou des variables.

Distances : Mesures de dissimilarité

Dans le groupe Mesure, sélectionnez la possibilité qui correspond au type de vos données (intervalle, effectif ou binaire). Ensuite, dans la liste déroulante, sélectionnez l'une des mesures correspondant à ce type de données. Les mesures disponibles sont, par type de données :

- **Intervalle :** Distance Euclidienne, Carré de la distance Euclidienne, Distance de Tchebycheff, Distance de Manhattan, Distance de Minkowski ou Autre.
- **Effectifs :** Distance du khi-deux ou Distance du phi-deux.
- **Binaire :** Distance Euclidienne, Carré de la distance Euclidienne, différence de taille, Différence de motif, Variance, Forme, ou Lance et Williams. (Entrez des valeurs dans les champs Présent et Absent

pour indiquer les deux valeurs significatives. Aucune autre valeur ne sera prise en compte dans Distances.)

Le groupe Transformer les valeurs vous permet de standardiser les valeurs des données pour les observations ou les variables *avant* le calcul des proximités. Ces transformations ne s'appliquent pas aux données binaires. Les méthodes de standardisation disponibles sont les scores *z*, la plage –1 à 1, la plage 0 à 1, l'amplitude maximale de 1, la moyenne de 1 ou l'écart type de 1.

Le groupe Transformer les mesures vous permet de transformer les valeurs générées par la mesure de distance. Elles sont appliquées après le calcul de la mesure d'indice. Les options disponibles sont Valeurs absolues, Inverser le signe et Rééchelonner entre 0 et 1.

Indices : Mesures de similarité

Dans le groupe Mesure, sélectionnez la possibilité qui correspond au type de vos données (intervalle ou binaire). Ensuite, dans la liste déroulante, sélectionnez l'une des mesures correspondant à ce type de données. Les mesures disponibles sont, par type de données :

- **Intervalle :** Corrélation de Pearson ou Cosinus.
- **Binaire :** Russel et Rao, Indice de Sokal et Michener, Jaccard, Dice, Rogers et Tanimoto, Sokal et Sneath 1, Sokal et Sneath 2, Sokal et Sneath 3, Kulczynski 1, Kulczynski 2, Sokal et Sneath 4, Hamann, Lambda, *D* d'Anderberg, *Y* de Yule, *Q* de Yule, Ochiai, Sokal et Sneath 5, Corrélation phi tétrachorique ou Dispersion. (Entrez des valeurs dans les champs Présent et Absent pour indiquer les deux valeurs significatives. Aucune autre valeur ne sera prise en compte dans Distances.)

Le groupe Transformer les valeurs vous permet de standardiser les valeurs des données pour les observations ou les variables avant le calcul des proximités. Ces transformations ne s'appliquent pas aux données binaires. Les méthodes de standardisation disponibles sont les scores *z*, la plage –1 à 1, la plage 0 à 1, l'amplitude maximale de 1, la moyenne de 1 et l'écart type de 1.

Le groupe Transformer les mesures vous permet de transformer les valeurs générées par la mesure de distance. Elles sont appliquées après le calcul de la mesure d'indice. Les options disponibles sont Valeurs absolues, Inverser le signe et Rééchelonner entre 0 et 1.

Fonctions supplémentaires de la commande PROXIMITIES

La procédure Distances utilise la syntaxe de la commande PROXIMITIES. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Indiquer un nombre entier comme la puissance pour la mesure de distance de Minkowski.
- Indiquer des nombres entiers comme la puissance et la racine pour une mesure de distance personnalisée.

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

Corrélation de distance

Cette fonction nécessite l'option Statistiques de base.

La procédure Distance Correlations calcule la dépendance entre deux variables aléatoires ou données multivariées. Elle évalue la relation non linéaire des variables, contrairement aux méthodes de corrélation traditionnelles.

Exemple

Une entreprise de vente au détail souhaite comprendre si les caractéristiques démographiques de ses clients (âge, revenu, éducation, état civil) influencent leur comportement d'achat. Les méthodes traditionnelles de corrélation peuvent échouer si les relations ne sont pas linéaires ou si elles impliquent des interactions entre plusieurs facteurs.

La procédure de corrélation de distance peut être appliquée dans le scénario avec les avantages suivants :

- Capture les dépendances non linéaires. Contrairement à la corrélation de Pearson, la corrélation de distance détecte les relations même si elles ne sont pas linéaires.
- Traite les relations multidimensionnelles. Il permet d'évaluer comment une combinaison de facteurs démographiques est liée aux habitudes d'achat.
- Fournit une mesure complète. Une corrélation de distance significative suggère que les caractéristiques démographiques des clients influencent fortement le comportement d'achat, ce qui oriente les stratégies de marketing ciblées.

Statistiques associées

- Corrélation de distance (*dCor(X,Y)) : Mesure la dépendance entre les caractéristiques démographiques du client (X*) et son comportement d'achat (*Y*).
- Valeur p basée sur la permutation : Évalue la signification statistique de la dépendance. Une faible valeur p (par exemple, p<0.05) indique une relation forte.
- Distance Covariance (*dCov(X,Y)) : Quantifie la force de la dépendance entre les variables.*
- Variance de distance (*dVar(X), dVar(Y)*) : mesure la variabilité au sein des variables individuelles pour normaliser la corrélation.

Remarques sur les données

Seules les variables continues peuvent être utilisées pour l'analyse. Un minimum de deux variables continues est nécessaire pour calculer la corrélation de distance. Un maximum de 30 variables est autorisé.

Obtention de la corrélation de distance

1. Dans le menu, cliquez sur : **Analyser** > **Corréler** > **Distance Corrélation**.

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez deux variables numériques ou plus pour lesquelles la corrélation de distance est calculée. Vous pouvez sélectionner jusqu'à 30 variables continues dans l'ensemble de données source pour l'analyse.
3. Sélectionnez une variable d'identification. Si vous optez pour une **variable ID**, les résultats affichent un tableau matriciel des distances avec les ID.
4. Cliquez sur **OK** pour exécuter la procédure avec les paramètres spécifiés et générer le résultat.

Corrélation de distance : Critères

La sous-boîte de dialogue Critères fournit des paramètres de contrôle pour la normalisation et d'autres options clés pour personnaliser la procédure de corrélation de distance.

Normalisation

Sélectionnez l'option permettant d'attribuer un type de normalisation à l'ensemble de données. La normalisation standardise les échelles des variables, mais elle risque de fausser par inadvertance les relations hautement non linéaires dans les données, qui sont cruciales pour évaluer avec précision la corrélation de distance. Il est donc inclus pour faciliter les relations linéaires.

- La normalisation **Min-Max** met à l'échelle chaque caractéristique dans une plage fixe, généralement [0, 1]. Cette transformation est utile lorsque votre ensemble de données contient des variables avec des unités ou des plages différentes. Cette normalisation permet une comparaison équitable entre les variables, en veillant à ce qu'aucune variable ne domine les calculs de distance en raison de son échelle. Il permet d'atténuer les effets des grandes plages numériques.
- La normalisation du **score Z** centre les données autour de la moyenne avec un écart-type de 1. Cette approche est utile pour les ensembles de données qui sont normalement distribués. Cette normalisation réduit l'influence des valeurs aberrantes, car elle donne à chaque variable une échelle normalisée. Elle facilite l'interprétation des résultats, en particulier lorsque vous comparez différentes variables, car toutes les variables sont traitées de la même manière par rapport à leur distribution.

- La **mise à l'échelle robuste** utilise la médiane et l'intervalle interquartile (IQR) pour mettre à l'échelle les données, ce qui les rend résistantes aux valeurs aberrantes. L'influence des valeurs aberrantes est minimisée, ce qui permet des calculs de corrélation de distance plus fiables en présence de valeurs extrêmes. Cette technique est utile lorsque l'ensemble de données contient des anomalies susceptibles de fausser les résultats.
- La **transformation logarithmique** est conçue pour réduire l'asymétrie des données, en particulier pour les distributions à asymétrie positive. Cela peut aider à stabiliser la variance dans l'ensemble des données. La transformation logarithmique comprime les grandes valeurs et augmente les petites valeurs, ce qui permet de normaliser la distribution. Cela peut améliorer la précision de l'analyse de corrélation de distance, en particulier lorsque vous traitez des ensembles de données fortement asymétriques.

Intervalle de confiance Pourcentage
Spécifiez le pourcentage de l'intervalle de confiance à utiliser dans l'analyse statistique.

Le pourcentage de l'intervalle de confiance indique le niveau de certitude associé aux estimations statistiques. Un pourcentage plus élevé reflète une plus grande confiance dans la précision de l'intervalle mais donne un intervalle plus large, tandis qu'un pourcentage plus faible donne un intervalle plus étroit avec moins de certitude

Permutations de la corrélation de distance

Dans la corrélation de distance, le test de permutation est utilisé pour évaluer la signification statistique de la corrélation de distance observée entre deux variables. L'objectif est d'évaluer si la corrélation observée est significativement différente de ce qui pourrait se produire par hasard. L'un des ensembles de données (ou les deux) est permuté de manière aléatoire, ce qui rompt toute association existante entre les variables. La corrélation de distance est ensuite recalculée pour chaque permutation, générant une distribution des valeurs de corrélation de distance sous l'hypothèse nulle d'absence d'association.

Sélectionnez **Set Custom Seed** pour activer les tests de permutation et pour spécifier le nombre d'itérations et le niveau de signification. Définir la valeur en **graines.** Vous pouvez éventuellement définir une graine pour le générateur de nombres aléatoires utilisé dans les tests de permutation. La graine doit être un nombre entier positif ou zéro. La valeur par défaut est 2000000.

Définir le **nombre maximum d'itérations**. Il doit s'agir d'un nombre entier positif. La valeur par défaut est fixée à 100 et les valeurs négatives ne sont pas autorisées.

En comparant la corrélation de distance observée avec la distribution des corrélations de distance des ensembles de données permutées, une valeur p est calculée. La valeur p représente la proportion de corrélations de distance permutées qui sont supérieures ou égales à la corrélation de distance observée. Un niveau de signification est utilisé comme seuil pour déterminer la signification statistique. Si la valeur p est inférieure au seuil de signification, la corrélation observée est considérée comme statistiquement significative. Définissez la valeur du **niveau de signification** entre 0 et 1. La valeur par défaut est 0,05.

Corrélation de distance : Imprimer

Imprimer détails
Sélectionnez **Imprimer les détails** pour afficher un tableau contenant les mesures essentielles suivantes : le nombre d'observations (N), la moyenne, l'écart-type (Std.), ainsi que les valeurs minimales et maximales pour chaque variable. Ces statistiques fournissent une vue d'ensemble de l'ensemble des données et vous permettent de mieux comprendre les distributions et les caractéristiques des variables comparées.

Coefficients de corrélation de distance
Sélectionnez **Coefficients de** corrélation de distance pour générer un tableau indiquant les coefficients de corrélation de distance calculés pour l'analyse. `dCor Value` indique le coefficient de corrélation à distance entre les variables. Si vous optez pour l'option **Set Custom Seed** dans la sous-boîte de dialogue **Criteria** et que vous définissez un niveau de signification, vous pouvez consulter la colonne `Sig.` qui indique le niveau de signification (valeur p) associé au coefficient de corrélation calculé.

`Number of Replicates (R)` indique le nombre de répétitions (ou de permutations si elles ont été choisies) utilisées dans l'analyse.

Estimations de la covariance des distances
Si les utilisateurs sélectionnent un niveau d'intervalle de confiance (IC), le tableau inclut les limites supérieures et inférieures des intervalles de confiance, offrant ainsi une vue plus complète des estimations.

Mesures de distance et estimations de la signification
Les estimations permettent d'évaluer les relations et les dépendances entre les variables analysées.

Matrice de distances
Sélectionnez **Matrice de distance** pour générer un tableau qui affiche la matrice de distance pour des données unidimensionnelles (par exemple. variable "âge"). Les distances entre chaque paire d'observations sont représentées. Les éléments diagonaux indiquent une distance nulle (distance par rapport à elle-même). Cette matrice fournit une vue claire des distances par paire afin que vous puissiez effectuer une analyse complète des relations entre les variables. Sélectionnez l'option permettant de configurer une matrice de distance. Dans la boîte de dialogue **Configurer la matrice de distance**, sélectionnez les variables dans la liste pour configurer la matrice de distance.

Enregistrer la matrice des distances sous forme de fichier
Sélectionnez l'option permettant d'enregistrer les matrices de distance dans un fichier. Indiquer le nom du fichier dans le **répertoire des fichiers sortants**. L'extension de fichier supportée est `.sav`.

Corrélation de distance : Tracé

Sélectionnez **Tracé** pour générer des tracés pour la corrélation de distance. Sélectionnez **Scatterplot** pour générer un diagramme de dispersion qui montre les distances ou les corrélations par paire entre les variables sélectionnées.

Sélectionnez dans la liste la variable à tracer sur l'axe des X. La variable doit être mutuellement exclusive avec les autres variables choisies pour le nuage de points.

Sélectionnez une autre variable dans la liste à tracer sur l'axe Y. Cette variable doit être mutuellement exclusive des autres variables choisies pour le nuage de points.

Modèles linéaires

Les modèles linéaires prédisent une cible continue en fonction de relations linéaires entre la cible et un ou plusieurs prédicteurs.

Les modèles linéaires sont relativement simples et produisent une formule mathématique pouvant facilement être évaluée. Les propriétés de ces modèles sont bien comprises et peuvent généralement être créées très rapidement par rapport à d'autres types de modèles (tels que les réseaux neuronaux ou les arbres de décisions) sur le même jeu de données.

Exemple. Une compagnie d'assurances disposant de ressources restreintes pour enquêter sur les demandes de remboursement des propriétaires de biens immobiliers, souhaite construire un modèle pour estimer les coûts des demandes d'indemnisation. Le déploiement de ce modèle dans les centres de service permettra aux représentants d'entrer les informations des demandes d'indemnisation alors qu'ils sont au téléphone avec les clients et de recevoir immédiatement le coût "prévu" de la demande d'indemnisation, sur la base de données anciennes.

Exigences concernant les champs. Il doit y avoir une cible et au moins une entrée. Par défaut, les champs avec les rôles prédéfinis Les deux ou Aucun ne sont pas utilisés.

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

La cible doit être continue (échelle). Il n'y a pas de restrictions sur les niveaux de mesure des prédicteurs (variables d'entrée) ; les champs catégoriels (nominaux et ordinaux) sont utilisés comme facteurs dans le modèle et les champs continus sont utilisés comme covariables.

Remarque : Si un champ catégoriel contient plus de 1000 catégories, la procédure ne s'exécute pas et aucun modèle n'est créé.

Obtention d'un modèle linéaire

Cette fonction nécessite l'option Statistiques de base.

A partir des menus, sélectionnez :

Analyse > **Régression** > **Modèles linéaires automatiques...**

1. Vérifiez qu'il existe au moins une cible et une entrée.
2. Cliquez sur **Options de création** pour spécifier les paramètres optionnels de création et de modèle.
3. Cliquez sur **Options du modèle** pour enregistrer les scores dans le jeu de données actif et exporter le modèle vers un fichier externe.
4. Cliquez sur **Exécuter** pour exécuter la procédure et créer les objets du modèle.

Objectifs

Quel est votre objectif principal ? Sélectionnez l'objectif de votre choix.

- **Créer un modèle standard.** Cette méthode permet de créer un modèle unique afin de prédire la cible à l'aide de prédicteurs. En général, les modèles standard sont plus faciles à interpréter et peuvent être plus rapidement évalués que des jeux de données boostés, de bagging ou de grande taille.
- **Améliorer la précision du modèle (boosting) :** Cette méthode permet de créer un modèle d'ensemble à l'aide du boosting, qui génère une séquence de modèles afin d'obtenir des prédictions plus précises. Les ensembles peuvent être plus longs à construire et l'obtention de leurs résultats peut être plus longue qu'avec un modèle standard.

  Le boosting produit une succession de "modèles de composant", chacun étant construit à partir de la totalité du jeu de données. Avant la création successive de chaque modèle de composant, les enregistrements sont pondérés en fonction des résidus des modèles de composant précédents. Les observations présentant de grands résidus se voient attribuées des pondérations d'analyse relativement plus élevées, de sorte que le modèle de composant suivant se concentre aussi sur la prédiction de ces enregistrements. Ensemble, ces modèles de composant forment un modèle d'ensemble. Le modèle d'ensemble évalue les nouveaux enregistrements à l'aide d'une règle de combinaison ; les règles disponibles dépendent du niveau de mesure de la cible.
- **Améliorer la stabilité du modèle (bagging) :** Cette méthode permet de créer un modèle d'ensemble à l'aide du bagging (agrégation par bootstrap), qui génère plusieurs modèles afin d'obtenir des prédictions plus fiables. Les ensembles peuvent être plus longs à construire et l'obtention de leurs résultats peut être plus longue qu'avec un modèle standard.

  L'agrégation de bootstrap (bagging) produit des réplicats du jeu de données d'apprentissage en effectuant un échantillonnage avec remplacement à partir du jeu de données d'origine. Ceci crée des échantillons de bootstrap de taille identique à celle du jeu de données d'origine. Ensuite, un modèle de composant est créé à partir de chaque réplicat. Ensemble, ces modèles de composant forment un modèle d'ensemble. Le modèle d'ensemble évalue les nouveaux enregistrements à l'aide d'une règle de combinaison ; les règles disponibles dépendent du niveau de mesure de la cible.
- **Créez un modèle pour de très grands jeux de données (nécessite IBM SPSS Statistiques Server).** Cette méthode permet de créer un modèle d'ensemble en scindant le jeu de données en blocs de données distincts. Choisissez cette option si votre jeu de données est trop important pour que vous puissiez créer l'un des modèles ci-dessus, ou pour la génération d'un modèle incrémental. La construction de cette option peut être moins longue, mais l'obtention des scores peut être plus longue qu'avec un modèle standard. Cette option nécessite une connectivité IBM SPSS Statistiques Serveur .

Voir «Ensembles », à la page 170 pour les paramètres liés au boosting, au bagging et aux jeux de données de très grande tailleboosting et bagging.

Bases

Préparer automatiquement les données. Cette option permet à la procédure de transformer la cible et les prédicteurs en interne afin de maximiser le pouvoir prédictif du modèle ; toutes les transformations sont enregistrées avec le modèle et appliquées aux nouvelles données pour l'évaluation. Les versions originales de champs transformés sont exclues du modèle. Par défaut, les préparations automatiques de données suivantes sont réalisées.

- **Gestion de la date et de l'heure :** Chaque prédicteur de date est transformé en une nouveau prédicteur continu qui contient la durée écoulée depuis une date de référence (01/01/1970). Chaque prédicteur d'heure est transformé en une nouveau prédicteur continu qui contient la durée écoulée depuis une heure de référence (00:00:00).
- **Régler le niveau de mesure :** Les prédicteurs continus ayant moins de 5 valeurs distinctes sont reconvertis en prédicteurs ordinales. Les prédicteurs ordinaux ayant plus de 10 valeurs distinctes sont reconvertis en prédicteurs continus.
- **Traitement des valeurs extrêmes :** Les valeurs de prédicteurs continus qui se trouvent au-delà d'une valeur de césure (écart type de 3 par rapport à la moyenne) sont définies sur la valeur de césure.
- **Traitement des valeurs manquantes.** Les valeurs manquantes de prédicteurs nominaux sont remplacées par le mode de la partition d'apprentissage. Les valeurs manquantes de prédicteurs ordinaux sont remplacées par la médiane de la partition d'apprentissage. Les valeurs manquantes de prédicteurs continus sont remplacées par la moyenne de la partition d'apprentissage.
- **Fusion supervisée :** Ceci crée un modèle plus petit en réduisant le nombre de champs à traiter en association avec la cible. Les catégories similaires sont identifiées en fonction de la relation entre l'entrée et la cible. Les catégories ne différant pas de manière significative (c'est-à-dire ayant une valeur p supérieure à 0,1), sont fusionnées. Si toutes les catégories sont fusionnées en une seule, les versions d'origine et dérivées du champ sont exclues du modèle car elles n'ont pas de valeur de prédicteur.

Niveau de confiance : Il s'agit du niveau de confiance utilisé pour calculer les estimations d'intervalle des coefficients de modèle dans la vue Coefficients. Spécifiez une valeur supérieure à 0 et inférieure à 100. La valeur par défaut est 95.

Choix du modèle

Méthode de sélection du modèle. Choisissez l'une des méthodes de sélection du modèle (détails ci-dessous) ou **Inclure tous les prédicteurs**, qui entre simplement tous les prédicteurs disponibles en tant que termes du modèle des effets principaux. Le modèle **Etape par étape ascendant** est utilisé par défaut.

Sélection de la méthode Pas à pas ascendante. Elle commence sans effet dans le modèle et ajoute et supprime des effets une étape à la fois jusqu'à ce qu'aucune autre ne puisse être ajoutée ou supprimée en fonction des critères étape par étape.

- **Critères d'entrée/suppression :** Il s'agit des statistiques utilisées pour savoir si un effet doit être ajouté ou supprimé du modèle. **Critère d'information (AICC)** est basé sur la vraisemblance du modèle fourni à l'ensemble d'apprentissage, et est ajusté afin de pénaliser des modèles trop complexes. **Statistiques F** est basé sur un test statistique de l'amélioration dans l'erreur d'un modèle. **R-carré ajusté** est basé sur l'adéquation de l'ensemble d'apprentissage, et est ajusté afin de pénaliser des modèles trop complexes. Le **Critère de prévention du surajustement (ASE)** est basé sur l'adéquation (carré de l'erreur moyenne, ou ASE) de l'ensemble de prévention du surajustement. L'ensemble de prévention du surajustement est un sous-échantillon aléatoire d'environ 30 % du jeu de données original qui n'est pas utilisé pour former le modèle.

  Si un autre critère que **Statistiques F** est sélectionné, à chaque étape l'effet qui correspond à l'accroissement positif le plus important dans le critère est ajouté au modèle. Tous les effets du modèle qui correspondent à une diminution du critère sont supprimés.

  Si **Statistiques F** est sélectionné en tant que critère, à chaque étape l'effet ayant la plus petite valeur *p*inférieure au seuil spécifié, **Inclure les effets avec des valeurs p inférieures à**, est ajouté au modèle. La valeur par défaut est 0.05. Tous les effets du modèle ayant une valeur *p* supérieure au seuil spécifié,

Supprimer les effets ayant des valeurs p supérieures à, sont supprimés. La valeur par défaut est 0.10.

- **Personnaliser le nombre maximum d'effets dans le modèle final.** Par défaut, tous les effets disponibles peuvent être entrés dans le modèle. Si l'algorithme étape par étape se termine à une étape avec le nombre spécifié d'effets, l'algorithme s'arrête à l'ensemble d'effets en cours.
- **Personnaliser le nombre maximal de pas :** L'algorithme étape par étape s'arrête après un certain nombre d'étapes. Par défaut, il s'agit de 3 fois le nombre d'effets disponibles. Vous pouvez également spécifier un nombre entier positif maximum d'étapes.

Sélection des meilleurs sous-ensembles. Ceci permet de vérifier "tous les modèles possibles" ou au moins un sous-ensemble plus important des modèles possibles qu'en étape par étape ascendant, pour choisir le meilleur en fonction du critère des meilleurs sous-ensembles. **Critère d'information (AICC)** est basé sur la vraisemblance de l'ensemble d'apprentissage compte tenu du modèle, et est ajusté pour pénaliser les modèles trop complexes. **R-carré ajusté** est basé sur l'adéquation de l'ensemble d'apprentissage, et est ajusté afin de pénaliser des modèles trop complexes. Le **Critère de prévention du surajustement (ASE)** est basé sur l'adéquation (carré de l'erreur moyenne, ou ASE) de l'ensemble de prévention du surajustement. L'ensemble de prévention du surajustement est un sous-échantillon aléatoire d'environ 30 % du jeu de données original qui n'est pas utilisé pour former le modèle.

Le modèle ayant la plus grande valeur de critère est sélectionné comme meilleur modèle.

Remarque : La sélection des meilleurs sous-ensembles demande plus de ressources de calcul que la sélection pas à pas ascendante. Lorsque la sélection des meilleurs sous-ensembles est effectuée en conjonction avec le boosting, le bagging ou le traitement d'ensembles très volumineux, elle peut être plus longue que la création d'un modèle standard à l'aide de la sélection pas à pas ascendante.

Ensembles

Ces paramètres déterminent le comportement d'assemblage qui se produit lors du boosting, du bagging ou lorsque que des ensembles volumineux de données sont requis dans les objectifs. Les options qui ne s'appliquent pas à l'objectif sélectionné sont ignorées.

Bagging et très grands jeux de données : Lors de l'évaluation d'un ensemble, il s'agit de la règle utilisée pour combiner les valeurs prédites à partir des modèles de base pour calculer la valeur de score d'un ensemble.

- **Règles de combinaison par défaut pour les cibles continues :** Des valeurs prédites d'ensemble pour des cibles continues peuvent être combinées à l'aide de la moyenne ou de la médiane des valeurs prédites à partir des modèles de base.

Veuillez noter que lorsque l'objectif consiste à améliorer la précision du modèle, les sélections de règles de combinaisons sont ignorées. Le boosting utilise toujours un vote majoritaire pondéré pour évaluer des cibles catégorielles et une médiane pondérée pour évaluer des cibles continues.

Boosting et bagging : Spécifiez le nombre de modèles de base à créer lorsque l'objectif est d'améliorer la précision ou la stabilité du modèle ; pour le bagging, il s'agit du nombre d'échantillons de bootstrap. Il doit s'agir d'un entier positif.

Avancé

Dupliquer les résultats. Définir une valeur de départ aléatoire vous permet de dupliquer des analyses. Le générateur de nombres aléatoires est utilisé pour choisir les enregistrements de l'ensemble de prévention du surajustement. Spécifiez un entier ou cliquez sur **Générer**, ce qui crée un entier pseudo-aléatoire compris entre 1 et 2147483647, inclus. La valeur par défaut est 54752075.

Options de modèle

Enregistrer les prévisions dans le jeu de données actif : Le nom par défaut de la variable est *PredictedValue*.

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

Récapitulatif de modèle

La vue Récapitulatif des modèles est un instantané permettant de consulter en un coup d'oeil le modèle et son ajustement.

Tableau : Le tableau identifie certains paramètres de modèle de haut niveau, dont :

- Le nom de la cible spécifié sur l'onglet Champs,
- Si la préparation automatique des données a été réalisée comme spécifié dans les paramètres de base,
- La méthode de sélection du modèle et le critère de sélection spécifiés dans les paramètres de sélection du modèle. La valeur du critère de sélection du modèle final est également affichée et elle est présentée en plus petit, disposant d'un meilleur format.

Graphique : Le graphique affiche la précision du modèle final, qui est présenté en plus grand, disposant d'un meilleur format. La valeur est 100 × le R^2 ajusté pour le modèle final.

Préparation automatique des données

Cette vue affiche des informations concernant les champs qui ont été exclus et la façon dont les champs transformés ont été dérivés dans l'étape de préparation automatique des données (ADP). Pour chaque champ transformé ou exclu, le tableau répertorie le nom du champ, son rôle au sein de l'analyse et l'action entreprise par l'étape ADP. Les champs sont triés selon l'ordre alphabétique croissant des noms de champ. Les actions possibles pour chaque champ comprennent :

- **Dériver les durées de date** calcule le temps écoulé, en mois, à partir des valeurs d'un champ contenant des dates jusqu'à la date système en cours.
- **Dériver les durées temporelles** calcule le temps écoulé, en heures, à partir des valeurs d'un champ contenant des heures jusqu'à l'heure système en cours.
- **Modifier le niveau de mesure de continu en ordinal** reconvertit les champs continus possédant moins de 5 valeurs uniques en champs ordinaux.
- **Modifier le niveau de mesure d'ordinal en continu** reconvertit les champs ordinaux possédant plus de 10 valeurs uniques en champs continus.
- **Tronquer les valeurs extrêmes** définit les valeurs des prédicteurs continus qui se trouvent au-delà d'une valeur de césure (écart type de 3 par rapport à la moyenne) sur la valeur de césure.
- **Remplacer les valeurs manquantes** remplace les valeurs manquantes des champs nominaux par le mode, celles des champs ordinaux par la médiane, et celles des champs continus par la moyenne.
- **Fusionner les catégories pour augmenter l'association avec la cible**identifie les catégories de prédicteurs similaires en fonction de la relation entre l'entrée et la cible. Les catégories ne différant pas de manière significative (c'est-à-dire ayant une valeur *p* supérieure à 0,05), sont fusionnées.
- **Exclure les prédicteurs constants / après le traitement des valeurs extrêmes / après la fusion des catégories** supprime les prédicteurs qui possèdent une valeur unique, éventuellement une fois les autres actions ADP effectuées.

Importance des prédicteurs

Généralement, vous souhaitez concentrer vos efforts de modélisation sur les champs prédicteurs les plus importants et vous envisagez d'exclure et d'ignorer les moins importants. Le graphique d'importance des prédicteurs peut vous y aider en indiquant l'importance relative de chaque prédicteur en estimant le modèle. Etant donné que les valeurs sont relatives, la somme des valeurs pour l'ensemble des prédicteurs affichés est 1,0. L'importance des prédicteurs n'a aucun rapport avec la précision du modèle. Elle est juste liée à l'importance de chaque prédicteur pour la réalisation d'une prévision, peu importe si cette dernière est précise ou non.

Valeurs prévues en fonction des valeurs observées

Ceci affiche un nuage de points mis en intervalles des valeurs prédites sur l'axe vertical par les valeurs observées sur l'axe horizontal. Idéalement, les points devraient se trouver sur une ligne de 45 degrés ; cette vue peut indiquer si des enregistrements sont particulièrement mal prédits par le modèle.

Résidus

Ceci affiche un graphique de diagnostique des résidus du modèle.

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

- **Histogramme :** Il s'agit d'un histogramme à intervalles des résidus de Student avec une superposition de la distribution normale. Les modèles linéaires supposent que les résidus ont une distribution normale, de sorte que l'histogramme doit, dans l'idéal, approcher étroitement la ligne de lissage.
- **Tracé P-P :** Il s'agit d'un tracé probabilité-probabilité mis en intervalles qui compare des résidus de Student à une distribution normale. Si la pente des points tracés est moins forte que la ligne normale, les résidus affichent une plus grande variabilité qu'une distribution normale ; si la pente est plus forte, les résidus affichent une moins grande variabilité qu'une distribution normale. Si les points tracés ont une courbe en S, la distribution des résidus est asymétrique.

Valeurs extrêmes

Ce tableau répertorie les enregistrements qui exercent une influence excessive sur le modèle et affiche l'ID d'enregistrement (s'il est spécifié dans l'onglet Champs), la valeur cible et la distance de Cook. La distance de Cook est une mesure du degré de modification des résidus de tous les enregistrements si un enregistrement donné est exclu des calculs des coefficients de modèle. Une distance de Cook importante signifie que l'exclusion d'un enregistrement modifie de manière importante les coefficients et doit donc être considérée comme ayant une influence.

Les enregistrements ayant une influence doivent être examinés soigneusement afin de déterminer si vous pouvez leur octroyer une pondération inférieure dans l'estimation du modèle, tronquer les valeurs éloignées à un seuil acceptable ou supprimer complètement les enregistrements ayant une influence.

Effets

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

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

- **Diagramme :** Il s'agit d'un graphique dans lequel les effets sont triés de haut en bas en diminuant l'importance du prédicteur. Les lignes de connexion du diagramme sont pondérées en fonction de la signification de l'effet, une largeur de ligne plus importante correspondant à des effets plus importants (valeurs *p* plus petites). Lorsque vous passez la souris sur une ligne de connexion, une infobulle affiche la valeur-*p* et l'importance de l'effet. Il s'agit de la valeur par défaut.
- **Tableau.** Il s'agit d'un tableau ANOVA pour le modèle général et les effets de modèle individuels. Il s'agit d'effets individuels triés de haut en bas en diminuant l'importance du prédicteur. Notez, que par défaut, le tableau est réduit et n'affiche que les résultats du modèle global. Pour consulter les résultats des effets du modèle individuel, cliquez sur la cellule **Modèle corrigé** dans le tableau.

Importance des prédicteurs : Il existe un curseur de l'importance du prédicteur qui contrôle ceux qui sont affichés dans la vue. Ceci ne modifie pas le modèle, mais vous permet simplement de vous concentrer sur les prédicteurs les plus importants. Par défaut, les 10 premiers effets sont affichés.

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

Coefficients

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

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

- **Diagramme :** Il s'agit d'un graphique qui affiche d'abord la constante, puis trie les effets de haut en bas en diminuant l'importance du prédicteur. Au sein des effets contenant des facteurs, les coefficients sont triés dans l'ordre croissant de la valeur des données. Les lignes de connexion du diagramme sont coloriées en fonction du signe du coefficient (voir le diagramme) et sont pondérées en fonction de la signification du coefficient, une largeur de ligne plus importante correspondant à des coefficient plus significatifs (valeurs-*p* plus petites). Lorsque vous passez la souris sur une ligne de connexion, une infobulle affiche la valeur du coefficient, sa valeur-*p* et l'importance de l'effet auquel est associé le paramètre. Il s'agit du style par défaut.
- **Tableau.** Affiche les valeurs, les tests de signification et les intervalles de confiance des coefficients de modèles individuels. Après la constante, les effets sont triés de haut en bas en diminuant l'importance du prédicteur. Au sein des effets contenant des facteurs, les coefficients sont triés dans l'ordre croissant de la valeur des données. Notez, que par défaut, le tableau est réduit et n'affiche que le coefficient, la signification et l'importance de chaque paramètre du modèle. Pour consulter l'erreur standard, la statistique *t* et l'intervalle de confiance, cliquez sur la cellule **Coefficient** dans le tableau. Lorsque vous passez la souris sur le nom d'un paramètre du modèle dans le tableau, une infobulle affiche le nom du paramètre, l'effet auquel il est associé, et (pour les prédicteurs catégoriels) les libellés des valeurs associées au paramètre du modèle. Ceci est particulièrement utile pour afficher les nouvelles catégories créées lorsque la préparation automatique des données fusionne les catégories similaires d'un prédicteur catégoriel.

Importance des prédicteurs : Il existe un curseur de l'importance du prédicteur qui contrôle ceux qui sont affichés dans la vue. Ceci ne modifie pas le modèle, mais vous permet simplement de vous concentrer sur les prédicteurs les plus importants. Par défaut, les 10 premiers effets sont affichés.

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

Moyennes estimées

Il s'agit de graphiques affichés pour des prédicteurs significatifs. Le graphique affiche la valeur estimée par le modèle de la cible sur l'axe vertical pour chaque valeur du prédicteur de l'axe horizontal en conservant tous les autres prédicteurs. Il offre une visualisation pratique des effets des coefficients de chaque prédicteur sur la cible.

Remarque : Si aucun prédicteur n'est significatif, aucune moyenne estimée n'est générée.

Récapitulatif de génération de modèle

Lorsqu'un algorithme de sélection de modèle différent de **Aucun** est sélectionné dans les paramètres de sélection de modèles, il propose certains détails concernant le processus de génération de modèle.

Pas à pas ascendante : Lorsque l'algorithme de sélection est pas à pas ascendante, le tableau affiche les 10 dernières étapes de l'algorithme pas à pas. Pour chaque pas, la valeur du critère de sélection et les effets du modèle à cette étape sont affichés. Ceci vous offre un aperçu de l'ampleur de la contribution de chaque étape au modèle. Chaque colonne vous permet de trier les lignes afin que vous puissiez voir plus facilement les effets qui se trouvent dans le modèle à chaque étape donnée.

Meilleurs sous-ensembles : Lorsque l'algorithme de sélection est Meilleurs sous-ensembles, le tableau affiche les 10 meilleurs modèles. Pour chaque modèle, la valeur du critère de sélection et les effets du modèle sont affichés. Ceci vous donne un aperçu de la stabilité des meilleurs modèles ; s'ils ont tendance à avoir des effets similaires avec quelques différences, vous pouvez alors avoir une confiance raisonnable dans le "meilleur" modèle ; s'ils ont tendance à avoir des effets très différents, certains des effets peuvent être trop similaires et doivent être associés (ou l'un d'entre eux doit être supprimé). Chaque colonne vous permet de trier les lignes afin que vous puissiez voir plus facilement les effets qui se trouvent dans le modèle à chaque étape donnée.

Régression linéaire

La procédure Régression linéaire estime les coefficients de l'équation linéaire, impliquant une ou plusieurs variables indépendantes, qui estiment le mieux la valeur de la variable dépendante. Par exemple, vous pouvez essayer d'estimer les ventes annuelles globales d'un commercial (la variable dépendante) à partir de variables indépendantes telles que l'âge, l'éducation et les années d'expérience.

Exemple. Le nombre de matches gagnés par une équipe de basket-ball au cours d'une saison est-il lié au nombre moyen de points marqués par l'équipe à chaque match ? Un nuage de points indique que ces variables ont un lien linéaire. Le nombre de matches gagnés et le nombre moyen de points marqué par l'équipe adverse ont également un lien linéaire. Ces variables ont une relation négative. Lorsque le nombre de matches gagnés augmente, le nombre moyen de points marqués par les adversaires diminue. A l'aide de la régression linéaire, vous pouvez modéliser la relation entre ces variables. Un bon modèle peut être utilisé pour prévoir combien de matches les équipes vont gagner.

Statistiques : Pour chaque : nombre d'observations valides, moyenne et écart type. Pour chaque modèle : coefficient de régression, matrice de corrélation, mesure et corrélations partielles, *R* multiple, R^2, R^2 ajusté, modification dans R^2, erreur standard de l'estimation, tableau d'analyse de variance, prévisions et résidus. En plus, intervalles de confiance à 95 % pour chaque coefficient de régression, matrice variances-covariances, facteur d'inflation de la variance, tolérance, test de Durbin-Watson, mesures de distances (Mahalanobis, Cook, et valeurs influentes), DfBêta, différence de prévision, intervalles d'estimation et informations de diagnostic des observations. Tracés : nuages de points, tracés partiels, histogrammes et tracés de probabilités gaussiens.

Régression linéaire : Remarques sur les données

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

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

Obtenir une analyse de régression linéaire

1. A partir des menus, sélectionnez :

   Analyse > **Régression** > **Linéaire...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Dans la boîte de dialogue Régression linéaire, sélectionnez une variable numérique dépendante.
3. Sélectionnez une ou plusieurs variables indépendantes.

(En option) Vous avez également ces possibilités :

- Grouper des variables indépendantes en blocs et spécifier différentes méthodes d'entrée pour différents sous-groupes de variables.
- Choisir une variable de sélection pour limiter l'analyse à un sous-groupe d'observations ayant une ou des valeurs particulières pour cette variable.

- Sélectionner une variable d'identification d'observations pour identifier des points sur les tracés.
- Sélectionnez une variable de pondération WLS numérique pour une analyse des moindres carrés pondérés.

WLS. Permet d'obtenir un modèle des moindres carrés pondérés. Les points de données sont pondérés par l'inverse de leur variance. Ainsi, les observations dont la variance est élevée ont moins d'impact sur l'analyse que celles dont la variance est faible. Si la valeur de la variable de pondération est nulle, négative ou manquante, l'observation est exclue de l'analyse.

Méthodes de sélection des variables de régression linéaire

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

- *introduire (régression)*. Procédure de sélection de variables au cours de laquelle toutes les variables d'un bloc sont introduites en une seule opération.
- *Stepwise* . A chaque étape, la variable indépendante qui ne figure pas dans l'équation ayant la plus petite probabilité de F est introduite, si cette probabilité est suffisamment faible. Les variables déjà comprises dans l'équation de régression sont éliminées si leur probabilité de F devient trop grande. Le processus s'arrête lorsqu'aucune variable ne peut plus être introduite ou supprimée.
- *Supprimer*. Procédure de sélection de variables au cours de laquelle toutes les variables d'un bloc sont supprimées en une seule opération.
- *élimination descendante*. Procédure de sélection de variables au cours de laquelle toutes les variables sont introduites dans l'équation, puis éliminées une à une. La variable ayant la plus petite corrélation partielle avec la variable dépendante est la variable dont la suppression est étudiée en premier. Si elle répond aux critères d'élimination, elle est supprimée. Une fois la première variable éliminée, l'élimination de la variable suivante restant dans l'équation et ayant le plus petit coefficient de corrélation partielle est étudiée. La procédure prend fin quand plus aucune variable de l'équation ne satisfait aux critères de suppression.
- *Sélection ascendante*. Procédure de sélection pas à pas de variables dans laquelle les variables sont introduites séquentiellement dans le modèle. La première variable considérée est celle qui a la plus forte corrélation positive ou négative avec la variable dépendante. Cette variable n'est introduite dans l'équation que si elle satisfait le critère d'introduction. Si la première variable est introduite dans l'équation, la variable indépendante externe à l'équation et qui présente la plus forte corrélation partielle est considérée ensuite. La procédure s'interrompt lorsqu'il ne reste plus de variables satisfaisant au critère d'introduction.

Les valeurs de signification dans vos sorties sont basées sur l'adéquation à un modèle unique. Par conséquent, les valeurs de signification ne sont généralement pas valables lorsqu'on utilise une méthode détaillée (étape par étape, ascendante ou descendante).

Toutes les variables doivent respecter le critère de tolérance pour être entrées dans l'équation, quelle que soit la méthode d'entrée spécifiée. Le niveau de tolérance par défaut est 0,0001. Une variable n'est pas entrée si elle fait passer la tolérance d'une autre variable déjà entrée dans le modèle en dessous du seuil de tolérance.

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

Régression linéaire : Définir la règle

Les observations définies par la règle de sélection sont incluses dans l'analyse. Par exemple, si vous sélectionnez une variable, choisissez **égale** et saisissez 5 pour la valeur, alors seules les observations pour lesquelles la variable sélectionnée a une valeur égale à 5 seront incluses dans l'analyse. Une valeur chaîne est également permise.

Tracés de régression linéaire

Les tracés peuvent aider à valider les hypothèses de normalité, linéarité et d'égalité des variances. Les tracés sont également utiles pour détecter les valeurs extrêmes, les observations éloignées et les observations influentes. Après avoir été enregistrés comme variables nouvelles, les prévisions, résidus et autres informations de diagnostic sont disponibles dans l'éditeur de données pour construire des tracés avec les variables indépendantes. Les tracés suivants sont disponibles :

Nuages de points : Vous pouvez tracer deux des éléments suivants : la variable dépendante, les prévisions standardisées, les résidus standardisés, les résidus supprimés, les prévisions ajustées, les résidus de Student et les résidus supprimés de Student. Tracer les résidus standardisés par rapport aux prévisions standardisées pour vérifier la linéarité et l'égalité des variances.

Liste des variables source. Répertorie la variable dépendante (DEPENDNT) et les variables résiduelles et prévues suivantes : prévisions standardisées (*ZPRED), résidus standardisés (*ZRESID), résidus supprimés (*DRESID), prévisions ajustées (*ADJPRED), résidus de Student (*SRESID), résidus supprimés de Student (*SDRESID).

Générer tous les tracés partiels : Affiche des nuages de points des résidus de chaque variable indépendante et les résidus de la variable dépendante lorsque les deux variables sont régressées séparément par rapport au reste des variables indépendantes. Au moins deux variables indépendantes doivent être dans l'équation pour produire un tracé partiel.

Tracés résiduels standardisés : Vous pouvez obtenir des histogrammes des résidus standardisés et des tracés de probabilités gaussiens en comparant la répartition des résidus standardisés à une répartition gaussienne.

Si vous demandez des tracés, des statistiques récapitulatives sont affichées pour les prévisions standardisées et les résidus standardisés (***ZPRED** et ***ZRESID**).

Régression linéaire : Enregistrement des nouvelles variables

Vous pouvez enregistrer les prévisions, les résidus et les autres statistiques utiles pour les informations de diagnostic. Chaque sélection ajoute une ou plusieurs variables à votre fichier de données actif.

Prévisions : Valeurs prévues par le modèle de régression pour chaque observation.

- *Non normalisé*. Valeur prévue par le modèle pour la variable dépendante.
- *Standardisés*. Transformation de chaque prévision en sa forme standardisée. La prévision moyenne est soustraite de la prévision, et la différence est divisée par l'écart type des prévisions. Les prévisions standardisées ont une moyenne de 0 et un écart type de 1.
- *Ajusté*. Prévision pour une observation lorsque cette dernière est exclue du calcul des coefficients de régression.
- *S.E. des prédictions moyennes*. Erreurs standard des prévisions. Estimation de l'écart type de la valeur moyenne de la variable expliquée pour les unités statistiques qui ont les mêmes valeurs pour les valeurs explicatives.

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

- *Mahalanobis*. Mesure du degré de différence entre les valeurs d'une observation sur les variables indépendantes et la moyenne de toutes les observations. Une distance de Mahalanobis importante identifie une observation qui a des valeurs extrêmes pour des variables indépendantes.
- *Cuisinier*. Mesure permettant de savoir de combien les résidus de toutes les observations seraient modifiés si une observation donnée était exclue du calcul des coefficients de régression. Si la distance de Cook est élevée, l'exclusion d'une observation changerait substantiellement la valeur des coefficients.
- *Valeurs de levier*. Mesure l'influence d'un point sur l'ajustement de la régression. La valeur influente centrée varie de 0 (aucune influence sur la qualité de l'ajustement) à (N-1)/N.

Intervalles de la prévision : Les limites supérieure et inférieure pour les intervalles de la prévision moyenne et individuelle.

- *moyenne*. Bornes inférieure et supérieure (deux variables) de l'intervalle de prévision de la réponse moyenne prévue.
- *Individuel*. Limites supérieure et inférieure (deux variables) de l'intervalle de prévision de la variable dépendante pour une observation particulière.
- *Intervalle de confiance*. Entrez une valeur comprise entre 1 et 99.99 pour spécifier le niveau de confiance des deux intervalles de la prévision. Vous devez sélectionner Moyenne ou Individuelle avant d'entrer cette valeur. Les niveaux d'intervalle de confiance typiques sont 90, 95 et 99.

Résidus : La valeur réelle de la variable indépendante moins la valeur prévue par l'équation de régression.

- *Non normalisé*. Différence entre une valeur observée et la valeur prévue par le modèle.
- *Standardisés*. Résidu, divisé par une estimation de son écart type. Egalement appelés résiduels de Pearson, les résiduels standardisés ont une moyenne de 0 et un écart type de 1.
- *Studentized*. Résidu, divisé par une estimation de son écart type, qui varie d'une observation à l'autre, selon la distance entre les valeurs et les moyennes des variables indépendantes de chaque observation. Parfois appelés résidus de Student interne.
- *Supprimé*. Résidu d'une observation lorsque cette dernière est exclue du calcul des coefficients de régression. Il s'agit de la différence entre la valeur de la variable dépendante et la prévision ajustée.
- *Suppression des études.*. Résidu supprimé pour une observation, divisé par son erreur standard. La différence entre le résidu supprimé de Student et le résidu de Student associé indique l'impact de l'élimination d'une observation sur sa propre prédiction. Parfois appelés résidus de Student externe.

Influences individuelles : Modification des coefficients de régression (DfBêta[s]) et des prévisions (différence de prévision) qui résulte de l'exclusion d'une observation particulière. Les valeurs DfBêtas et de différence de prévision standardisées sont également disponibles ainsi que le rapport de covariance.

- *DfBetas*. La différence de bêta correspond au changement du coefficient de régression qui résulte de l'exclusion d'une observation particulière. Une valeur est calculée pour chaque terme du modèle, y compris la constante.
- *DfBêta standardisée*. Différence normalisée de la valeur bêta. Modification du coefficient de régression, résultant de l'exclusion d'une observation donnée. Vous pouvez par exemple examiner les observations ayant des valeurs absolues supérieures à 2, divisées par la racine carrée de N, N représentant le nombre d'observations. Une valeur est calculée pour chaque terme du modèle, y compris la constante.
- *DfFit*. La différence de valeur ajustée correspond au changement de la valeur prévue résultant du retrait d'une observation donnée.
- *différence ajustée standardisée*. Différence standardisée de la valeur ajustée. Modification de la prévision qui résulte de l'exclusion d'une observation donnée. Vous pouvez par exemple examiner les valeurs standardisées dont la valeur absolue est supérieure à 2 fois la racine carrée de p/N, p correspondant au nombre de paramètres du modèle et N, au nombre d'observations.
- *Rapport de covariance*. Rapport entre le déterminant de la matrice de covariance avec une observation particulière exclue du calcul des coefficients de régression et le déterminant de la matrice de covariance avec toutes les observations incluses. Si le rapport est proche de 1, l'observation modifie peu la matrice de covariance.

Statistiques à coefficients : Enregistre les coefficients de régression dans un jeu de données ou dans un fichier de données. Les jeux de données sont disponibles pour utilisation ultérieure dans la même session mais ne sont pas enregistrés en tant que fichiers sauf si vous le faites explicitement avant la fin de la session. Le nom des jeux de données doit être conforme aux règles de dénomination de variables.

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

Statistiques de régression linéaire

Les statistiques suivantes sont disponibles :

Coefficients de régression- Estimations affiche le coefficient de régression *B*, l'erreur standard de *B*, le coefficient bêta normalisé, la valeur *t* pour *B*et le niveau de signification bilatérale de *t*. **Intervalles de confiance** affiche des intervalles de confiance avec le niveau de confiance spécifié pour chaque coefficient de régression ou matrice de covariance. L'option **Matrice de covariance** affiche la matrice de variance-covariance des coefficients de régression avec les covariances hors de la diagonale et les variances dans la diagonale. Une matrice de corrélation est également affichée.

Ajustement du modèle- Les variables introduites et éliminées dans le modèle sont répertoriées et les'statistiques de qualité d'ajustement'suivantes sont affichées: *R*multiple, R^2 et R^2ajusté, erreur standard de l'estimation et tableau d'analyse de variance.

Variation de R carré- Variation de la statistique R^2 qui est produite par l'ajout ou la suppression d'une variable indépendante. Si la variation du R^2 associée à une variable est importante, cela signifie que la variable est un bon prédicteur de la variable dépendante.

Descriptif- Indique le nombre d'observations valides, la moyenne et l'écart type de chaque variable de l'analyse. Une matrice de corrélations avec le niveau de signification unilatéral et le nombre d'observations pour chaque corrélation sont également affichés.

corrélation partielle. Il s'agit de la corrélation entre la variable dépendante et une variable indépendante lorsque les effets linéaires des autres variables indépendantes du modèle ont été éliminés de la variable indépendante. Elle est liée à la modification du R-deux lorsqu'une variable est ajoutée à une équation. (Parfois appelée corrélation semi-partielle.)

corrélation partielle. Corrélation qui reste entre deux variables après suppression de la corrélation due à leur association mutuelle avec les autres variables. Il s'agit de la corrélation entre la variable dépendante et une variable indépendante lorsque les effets linéaires des autres variables indépendantes du modèle ont été éliminés des deux variables.

Diagnostics de colinéarité- La colinéarité (ou multicolinéarité) est la situation indésirable où une variable indépendante est une fonction linéaire d'autres variables indépendantes. Les valeurs propres de la matrice des produits croisés dimensionnés et non centrés, les indices de condition et les proportions de décomposition de variance sont affichés ainsi que les facteurs d'inflation de la variance (VIF) et les tolérances pour les variables individuelles.

Critères de sélection- Inclut le critère d'information Akaike (AIC), le critère de prévision d'Ameniya (PC), l'erreur quadratique moyenne conditionnelle du critère de prévision (Cp) et le critère bayésien de Schwarz (SBC). Les statistiques sont affichées dans la table récapitulative du modèle.

Résidus- Vous pouvez sélectionner **'STATISTIQUES DE PRESSE'** à utiliser comme statistique de validation croisée pour comparer différents modèles. Le test **'Durban-Watson'** de corrélation en série des résidus est également affiché. Choisissez les informations de **diagnostic par observation** pour les observations qui répondent au critère de sélection (valeurs extrêmes supérieures à *n* écarts type).

Régression linéaire : Options

Les options suivantes sont disponibles :

Paramètres des méthodes progressives : Ces options s'appliquent lorsque la méthode de sélection des variables ascendante, descendante ou pas à pas est spécifiée. Les variables peuvent être introduites ou éliminées dans le modèle en fonction de la signification (probabilité) de la valeur *F* ou de la valeur *F* elle-même.

- *utiliser la probabilité de F*. Une variable est entrée dans le modèle si le niveau de signification de la valeur F est inférieur à la valeur Entrée et elle est éliminée si le niveau de signification est supérieur à la valeur Suppression. La valeur Entrée doit être inférieure à la valeur Suppression et toutes deux doivent être positives. Pour introduire davantage de variables dans le modèle, diminuez la valeur Entrée. Pour éliminer davantage de variables du modèle, réduisez la valeur Suppression.

- *utiliser valeur F*. Une variable est introduite dans un modèle si sa valeur F est supérieure à la valeur Entrée et elle est éliminée si la valeur F est inférieure à la valeur Suppression. La valeur Entrée doit être supérieure à la valeur Suppression et toutes deux doivent être positives. Pour introduire davantage de variables dans le modèle, réduisez la valeur du champ Entrée. Pour éliminer davantage de variables dans le modèle, augmentez la valeur du champ Suppression.

Tolérance: Par défaut, la valeur est .0001. La tolérance est la proportion de la variance d'une variable de l'équation qui n'est pas prise en compte par d'autres variables indépendantes de l'équation. La tolérance minimale pour toute variable de l'équation si la variable considérée a été incluse dans l'analyse est la tolérance minimale d'une variable qui n'est pas incluse dans l'équation. Les variables doivent réussir les tests de tolérance et de tolérance minimale pour entrer et rester dans une équation de régression. Si une variable satisfait aux critères de tolérance, elle peut être incluse en fonction de la méthode en vigueur.

Inclure terme constant dans l'équation : Par défaut, le modèle de régression inclut un terme constant. Désélectionner cette option force la régression jusqu'à l'origine, ce qui est rarement utilisé. Certains résultats de la régression jusqu'à l'origine ne sont pas comparables aux résultats de la régression incluant une constante. Par exemple, R^2 ne peut pas être interprété de la manière habituelle.

Valeurs manquantes : Vous pouvez choisir l'un des éléments suivants :

- **Exclure toute observation incomplète :** Seules les observations dont les valeurs sont valides pour toutes les variables sont incluses dans les analyses.
- **Exclure seulement les composantes non valides :** Les observations pour lesquelles les données sont complètes pour la paire de variables corrélées sont utilisées pour calculer le coefficient de corrélation sur lequel l'analyse de régression est basée. Les degrés de liberté sont basés sur le minimum *N* par paire.
- **Remplacer par la moyenne :** Toutes les observations sont utilisées pour les calculs, en substituant la moyenne de la variable aux observations manquantes.

Fonctions supplémentaires de la commande REGRESSION

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

- Ecrire une matrice de corrélation ou lire une matrice à la place de données brutes afin d'obtenir une analyse de régression (avec la sous-commande `MATRIX`).
- Spécifier des niveaux de tolérance (avec la sous-commande `CRITERIA`).
- Obtenir plusieurs modèles pour des variables dépendantes différentes ou identiques (avec les sous-commandes `METHOD` et `DEPENDENT`).
- Obtenir des statistiques supplémentaires (avec les sous-commandes `DESCRIPTIVES` et `STATISTICS`).

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

Régression ordinale

La procédure Régression ordinale vous permet d'effectuer un modèle dont la variable dépendante est une réponse ordinale sur un groupe de prédicteurs pouvant être soit des facteurs, soit des covariables. Le plan de régression ordinale repose sur la méthodologie de McCullagh (1980, 1998) ; vous trouverez cette procédure sous le nom de `PLUM` dans la syntaxe.

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 coefficients estimés reflètent le mode d'affectation de la réponse due aux modifications des prédicteurs. La réponse est numérique, dans le sens où les changements de niveau de réponse sont équivalents pour l'ensemble des plages de réponse. Par exemple, la différence de taille existant entre une personne de 150 cm et une de 140 cm est de 10 cm, ce qui correspond à la même différence existant entre une personne de 210 cm et une de 200 cm. Ces relations n'existent peut-être pas pour les variables ordinales, pour lesquelles le choix et le nombre de catégories de réponse peut être relativement arbitraire.

Exemple. La régression ordinale peut être utilisée en vue d'étudier la réaction des patients à certaines doses de médicament. §AA Les réactions possibles peuvent être classées en *aucune*, *légère*, *modérée* ou *grave*. La différence entre une réaction légère et une réaction modérée est difficile, voire impossible à quantifier car elle repose sur une perception. De plus, la différence entre une réponse légère et une réponse modérée peut être supérieure ou inférieure à la différence existant entre une réponse modérée et une réponse grave.

Tracés et statistiques : Fréquences observées et théoriques et fréquences cumulées, résiduels de Pearson pour les fréquences cumulées, probabilités observées et théoriques, probabilités observées et cumulées théoriques de chaque catégorie de réponse par motif de covariable, corrélation asymptotique et matrices de covariance des estimations des paramètres, khi-deux de Pearson et khi-deux du rapport de vraisemblance, statistiques de qualité d'ajustement, historique des itérations, test d'hypothèses de lignes parallèles, estimations des paramètres, erreurs standard, intervalles de confiance, statistiques de Cox et Snell, de Nagelkerke et R^2 de McFadden.

Régression ordinale : Remarques sur les données

Données. La variable dépendante est considérée comme ordinale, mais peut être soit numérique, soit chaîne. L'ordre est déterminé par le tri des valeurs de la variable dépendante par ordre croissant. La plus petite valeur définit la première catégorie. Les facteurs sont supposés être catégoriels. Les covariables doivent être numériques. Notez que l'utilisation de plusieurs covariables continues peut engendrer la création d'un tableau volumineux de probabilités par cellule.

Hypothèses : Seule une variable de réponse est autorisée et doit donc être spécifiée. De plus, pour chaque motif de valeurs distinct parmi les variables explicatives, les réponses sont supposées être des variables multinomiales explicatives.

Procédures apparentées : La régression logistique nominale utilise des modèles similaires pour les variables dépendantes nominales.

Obtention d'une régression ordinale

1. A partir des menus, sélectionnez :

   Analyse > **Régression** > **Ordinale...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une variable dépendante.
3. Cliquez sur **OK**.

Régression ordinale : Options

La boîte de dialogue Options vous permet d'ajuster des paramètres utilisés dans l'algorithme d'estimation itératif, de choisir un niveau de confiance pour vos estimations de paramètres et de sélectionner une fonction de lien.

Itérations : Vous pouvez personnaliser l'algorithme itératif.

- **Nombre maximum d'itérations :** Spécifiez un nombre entier non négatif. Si vous indiquez 0, la procédure renvoie aux estimations initiales.
- **Nombre maximum de dichotomie :** Spécifiez un nombre entier positif.
- **Convergence de log de vraisemblance :** L'algorithme s'interrompt si la modification absolue ou relative apportée au log de vraisemblance est inférieure à cette valeur. Le critère n'est pas utilisé si 0 est spécifié.
- **Convergence des paramètres :** L'algorithme s'interrompt si la modification absolue ou relative apportée à chaque estimation de paramètres est inférieure à cette valeur. Le critère n'est pas utilisé si 0 est spécifié.

Intervalle de confiance. Spécifiez une valeur supérieure ou égale à 0 et inférieure à 100.

Delta : Valeur ajoutée aux fréquences zéro par cellule. Spécifiez une valeur non négative, inférieure à 1.

Tolérance de singularité : Option utilisée pour vérifier les prédicteurs à haute dépendance. Sélectionnez une valeur dans la liste des options.

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

- **Logit :** $f(x)=\log(x/(1-x))$. Généralement utilisée pour les catégories uniformément réparties.
- **Log-log complémentaire :** $f(x)=\log(-\log(1-x))$. Cette option est généralement utilisée lorsque les catégories supérieures sont les plus probables.
- **Log-log négatif :** $f(x)=-\log(-\log(x))$. Cette option est généralement utilisée lorsque les catégories inférieures sont les plus probables.
- **Probit :** $f(x)=\Phi^{-1}(x)$. Cette option est généralement utilisée lorsque la variable de latence est normalement distribuée.
- **Cauchit (Cauchy inverse) :** $f(x)=\tan(\pi(x-0.5))$. Cette option est généralement utilisée lorsque la variable de latence possède un grand nombre de valeurs extrêmes.

Régression ordinale : Sortie

La boîte de dialogue Sortie vous permet de générer des tableaux d'affichage dans le visualiseur et d'enregistrer des variables dans le fichier de travail.

Afficher. Génère des tableaux pour :

- **Historique des itérations d'impression tous les n étapes :** Le log de vraisemblance et les estimations des paramètres sont imprimés en fonction de la fréquence des itérations d'impression spécifiée. La première et la dernière itération sont toujours imprimées.
- **Qualité d'ajustement :** Statistiques du khi-carré du rapport de vraisemblance et de Pearson. Ces éléments sont calculés en fonction du classement indiqué dans la liste des variables.
- **Statistiques récapitulatives :** Statistiques de Cox et Snell, de Nagelkerke et R^2 de McFadden.
- **Estimations des paramètres :** Estimations des paramètres, erreurs standard et intervalles de confiance.
- **Corrélation asymptotique des estimations :** Matrice de corrélations des estimations de paramètres.
- **Covariance asymptotique des estimations :** Matrice de covariances des estimations de paramètres.
- **Informations sur les cellules :** Fréquences observées et théoriques, et fréquences cumulées, résiduels de Pearson pour les fréquences cumulées, probabilités observées et théoriques, probabilités observées et cumulées de chaque catégorie de réponse par motif de covariable. Notez que pour les modèles comportant plusieurs motifs de covariable (par exemple, les modèles avec covariables continues), cette option peut générer un tableau très volumineux et donc, difficile à gérer.
- **Test de droites parallèles :** Test d'hypothèse selon laquelle les paramètres d'emplacement sont équivalents pour tous les niveaux de la variable dépendante. Cette option est uniquement disponible pour le modèle d'emplacement.

Variables enregistrées : Enregistre les variables suivantes dans le fichier de travail :

- **Probabilités des réponses estimées :** Probabilités estimées sur un motif de classement d'un motif de facteur/covariable dans les catégories de réponse. Il existe autant de probabilités que de nombre de catégories de réponse.
- **Catégorie estimée :** Catégorie de réponse contenant la probabilité estimée maximale pour un motif de facteur/covariable.
- **Probabilité de catégorie estimée :** Probabilité estimée de classement d'un motif de facteur/covariable au sein d'une catégorie prévue. Cette probabilité représente également le maximum de probabilités estimées du motif de facteur/covariable.
- **Probabilité de catégorie actuelle :** Probabilité estimée de classement d'un motif de facteur/covariable au sein de la catégorie actuelle.

Imprimer le log de vraisemblance : Contrôle l'affichage du log de vraisemblance. **Inclure la constante multinomiale** vous indique la valeur complète de la vraisemblance. Pour comparer vos résultats parmi les produits n'incluant pas de constante, vous pouvez choisir de l'exclure.

Régression ordinale : Emplacement

La boîte de dialogue Emplacement vous permet de spécifier le modèle d'emplacement de l'analyse.

Spécifier un modèle : Un modèle comportant des effets principaux contient des effets principaux de covariable et de facteur, mais aucun effet d'interaction. Vous pouvez créer un modèle personnalisé pour définir des sous-groupes d'interactions entre facteurs ou covariables.

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

Modèle d'emplacement : Le modèle dépend des effets principaux et des effets d'interaction que vous sélectionnez.

Pour les facteurs et covariables sélectionnés :

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

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

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

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

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

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

Termes construits et termes personnalisés

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

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

Régression ordinale : Echelle

La boîte de dialogue Echelle vous permet de spécifier le modèle d'échelle de l'analyse.

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

Modèle d'échelle : Le modèle dépend des effets principaux et des effets d'interaction que vous sélectionnez.

Pour les facteurs et covariables sélectionnés :

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

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

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

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

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

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

Termes construits et termes personnalisés

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

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

Fonctions supplémentaires de la commande PLUM

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

- Créer des tests d'hypothèse personnalisés en spécifiant des hypothèses nulles comme combinaisons linéaires de paramètres.

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

Régression nette élastique linéaire

Linear Elastic Net utilise la classe Python sklearn.linear_model.ElasticNet pour estimer les modèles de régression linéaire régularisés pour une variable dépendante sur une ou plusieurs variables indépendantes. La régularisation combine les pénalités L1 (Lasso) et L2 (Ridge). L'extension inclut des modes facultatifs permettant d'afficher des tracés de trace pour différentes valeurs d'alpha pour un rapport L1 donné, et de sélectionner le rapport L1 et les valeurs d'hyperparamètre alpha en fonction de la validation croisée. Lorsqu'un modèle unique est ajusté ou qu'une validation croisée est utilisée pour sélectionner le rapport de pénalité et / ou l'alpha, une partition des données restantes peut être utilisée pour estimer les performances hors de l'échantillon.

Outre l'ajustement d'un modèle avec les valeurs spécifiées du rapport du paramètre de pénalité L1 et de régularisation alpha, le réseau élastique linéaire peut afficher un tracé trace des valeurs de coefficient pour une plage de valeurs alpha pour un rapport donné, ou faciliter le choix de la valeur d'hyperparamètres via une validation croisée de niveau k sur des grilles de valeurs spécifiées. Si un modèle unique est ajusté ou si une sélection de ratio et / ou alpha est effectuée via une validation croisée, le modèle final peut être appliqué aux données retenues qui sont créées par une partition des données d'entrée pour obtenir une estimation valide des performances hors échantillon du modèle.

Obtention d'une analyse de régression linéaire élastique nette

1. A partir des menus, sélectionnez :

   Analyser > **Régression** > **Solutions de remplacement linéaires OLS** > **Elastic Net**

   La boîte de dialogue permet de spécifier une variable qui affecte chaque observation du jeu de données actif à l'échantillon d'apprentissage ou restant.

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

2. Sélectionnez une variable cible numérique. Une seule variable cible est requise pour exécuter une analyse.
3. Indiquez une valeur numérique dépendante.
4. Spécifiez au moins une variable de facteur catégorielle ou une variable de covariable numérique.

En option, **Partition** permet de créer un sous-ensemble restant ou de test des données d'entrée pour l'estimation des performances hors échantillon du modèle spécifié ou choisi. Tous les partitionnages sont effectués après la suppression des observations incomplètes de toutes les observations dont les données ne sont pas valides pour toute variable utilisée par la procédure. Notez que pour la validation croisée, les niveaux ou les partitions des données d'apprentissage sont créés dans Python. Les données restantes créées par la partition ne sont pas utilisées dans l'estimation, quel que soit le mode en vigueur.

La partition peut être définie soit en spécifiant le rapport des observations qui sont affectées de manière aléatoire à chaque échantillon (sous **Partitions d'apprentissage et de rétention**), soit par une variable qui affecte chaque observation à l'échantillon d'apprentissage ou de rétention. Vous ne pouvez pas spécifier à la fois l'apprentissage et les variables. Si la partition n'est pas spécifiée, un échantillon restant d'environ 30% des données d'entrée est créé.

Le **% d'apprentissage** spécifie le nombre relatif d'observations dans le jeu de données actif à affecter de manière aléatoire à l'échantillon d'apprentissage. L'entraînement par défaut est de 70%.

Régression linéaire élastique nette: Options

L'onglet Options fournit des options pour:

Mode

Cette sélection fournit des options permettant de spécifier l'un des modes suivants:

Ajuster au rapport L1 et alpha spécifiés
Lorsque vous sélectionnez ce mode, un modèle unique est ajusté aux données d'apprentissage à l'aide des valeurs de rapport L1 et de régularisation alpha spécifiées. Il s'agit de la valeur par défaut. Si une partition est spécifiée, le modèle unique ou final qui est ajusté est appliqué aux données de test retenues pour l'estimation des performances hors échantillon.

Sous **Tracé**, vous pouvez sélectionner les tracés des valeurs observées et / ou des résidus par rapport aux prévisions.

Sous **Enregistrer**, vous pouvez spécifier les valeurs prédites et les résidus à enregistrer.

Tracé de séries chronologiques
Lorsque vous sélectionnez ce mode, trois tracés pour les données d'apprentissage sont affichés en fonction de alpha pour l'ensemble de valeurs alpha spécifié:

- Tracé de trace des coefficients de régression.
- Un tracé de R^2.
- Tracé de l'erreur quadratique moyenne (MSE).

Bien que la partition soit respectée, aucun résultat pour les données de test retenues n'est fourni car aucun résultat de modèle final de ce mode.

Rapport L1 et/ou sélection alpha via la validation croisée
Lorsque vous sélectionnez ce mode, une recherche de grille avec validation croisée est effectuée pour évaluer les modèles, et les meilleures valeurs de rapport et alpha sont choisies en fonction de la meilleure moyenne R^2 sur les niveaux de validation. Le champ **Nombre de niveaux de validation croisée** peut être utilisé pour modifier la valeur par défaut de cinq divisions ou niveaux de validation croisée. Si une partition est spécifiée, le modèle unique ou final qui est ajusté est appliqué aux données de test retenues pour l'estimation des performances hors échantillon.

Sous **Afficher**, vous pouvez choisir d'afficher des informations de base uniquement sur le modèle avec la valeur de rapport et alpha choisie (**Best**), des informations de base sur tous les modèles comparés (**Comparer les modèles**) ou des informations complètes sur toutes les divisions ou tous les niveaux de tous les modèles (**Comparer les modèles et les niveaux**). **Best** est la valeur par défaut.

Sous **Tracé**, vous pouvez sélectionner des tracés de moyenne R^2 et/ou de MSE sur des niveaux de validation. Les tracés des valeurs observées et / ou des résidus par rapport aux valeurs prédites peuvent également être sélectionnés.

Sous **Enregistrer**, vous pouvez spécifier les valeurs prédites et les résidus à enregistrer.

Spécifier rapports L1 individuels
Lorsque vous sélectionnez ce mode pour le mode **Ajuster avec le rapport L1 spécifié et le mode alpha** ou **Tracé** , vous pouvez spécifier une seule valeur de rapport de pénalité L1 . Lorsque cette option est sélectionnée pour le rapport **L1 et/ou la sélection alpha via le mode de validation croisée** , vous pouvez spécifier plusieurs valeurs.

Spécifier la grille de valeurs alpha
Lorsque vous sélectionnez ce mode pour le rapport **L1 et/ou la sélection alpha via le mode de validation croisée** , une grille de valeurs alpha uniques peut être spécifiée à partir d'une valeur **Start** (`value1`) à une valeur **End** (`value2`) avec l'incrément de **By** (`value3`). S'il est spécifié, un seul ensemble valide de [`value1 TO value2 BY value3`] est autorisé. Il doit satisfaire la condition `0 ≤ value1 ≤ value2 ≤ 1`. Dans les cas où `value1 = value2`, cela revient à spécifier un seul `value1`, indépendamment de `value3`.

Les tracés sont affichés à l'aide de la métrique spécifiée pour les axes X horizontaux des valeurs alpha variables.

Spécifier alphas individuels
Lorsque vous sélectionnez ce mode pour le mode **Ajuster avec le rapport L1 spécifié et le mode alpha** , vous pouvez spécifier une valeur de régularisation alpha unique. Lorsque cette option est sélectionnée pour le **tracé de trace** ou la sélection de rapport et / ou d'alpha **L1 via le mode de validation croisée** , vous pouvez spécifier plusieurs valeurs.

Spécifier la grille de valeurs alpha
Lorsque vous sélectionnez ce mode pour le mode **Tracé** ou **RapportL1 et/ou sélection alpha via la validation croisée** , une grille de valeurs alpha uniques peut être spécifiée à partir d'une valeur **Début** (`value1`) à une valeur **Fin** (`value2`) avec l'incrément de **Par** (`value3`). S'il est spécifié, un seul ensemble valide de [`value1 TO value2 BY value3`] est autorisé. Il doit satisfaire ce `0 ≤ value1 ≤ value2 ≤ 1`. Si `value1 = value2`, cela revient à spécifier une valeur `value1` unique, quelle que soit la valeur `value3`.

La **métrique alpha** des plages de valeurs peut être soit **linéaire** , soit **logarithmique de base 10** (10 est élevé à la puissance des valeurs spécifiées).

Les tracés sont affichés à l'aide de la métrique spécifiée pour les axes X horizontaux des valeurs alpha variables.

Critères

Contrôle les analyses.

Inclure constante
Ce critère inclut une constante dans le (s) modèle (s) ajusté (s). Notez que la procédure d'extension ne centre pas ou ne normalise pas la variable dépendante et que la constante n'est pas pénalisée lors de l'estimation.

Normaliser les prédicteurs
Normalise toutes les variables indépendantes.

Nombre de niveaux de validation croisée
Nombre de divisions ou de niveaux pour l'évaluation de la validation croisée des modèles. Il doit s'agir d'un entier positif supérieur à 1. La valeur par défaut est 5.

Etat aléatoire Python
Valeur du paramètre random_state dans Python utilisée lors de l'évaluation de la validation croisée des modèles. Il permet la réplication des résultats qui impliquent des nombres pseudo-aléatoires. La valeur doit être un entier compris entre 0 et 2^{32}-1. La valeur par défaut est 0.

Limite de temps (minutes)
Nombre de minutes autorisé pour l'exécution des calculs de modèle. Si vous indiquez 0, le temporisateur est désactivé. La valeur par défaut est 5.

Afficher

Cette option indique le nombre de sorties à afficher pour le rapport **L1 et/ou la sélection alpha via le mode de validation croisée** .

Meilleures salutations,
Affiche uniquement les résultats de base pour le meilleur modèle choisi. Cette valeur est définie par défaut.

Comparer des modèles
Affiche les résultats de base pour tous les modèles évalués.

Comparer modèles et niveaux
Affiche les résultats prolixes complets pour chaque division ou pli pour chaque modèle évalué.

Tracé

Cette option spécifie les tracés des valeurs observées ou résiduelles par rapport aux prévisions et, avec la validation croisée, la spécification des tracés de l'erreur quadratique moyenne (MSE) et/ou de la moyenne R^2 sur les niveaux de validation croisée par rapport aux valeurs alpha.

Moyenne de l'erreur quadratique moyenne de validation croisée (MSE) par rapport à l'erreur alpha
Pour la sélection de rapport et / ou alpha **L1 via le mode** de validation croisée, affiche un tracé de courbes de MSE moyenne sur les niveaux de validation croisée par rapport à alpha pour la meilleure valeur de rapport L1 spécifiée ou sélectionnée. Pour le mode **Tracé de trace** , un tracé similaire est automatiquement généré en fonction des données d'apprentissage complètes.

Moyenne R-carré de la validation croisée par rapport à alpha
Pour la sélection de rapport et / ou d'alpha **L1 via le mode de validation croisée** , affiche un tracé en ligne de la moyenne R^2 sur les niveaux de validation croisée par rapport à alpha pour la meilleure valeur de rapport L1 spécifiée ou sélectionnée. Pour le mode **Tracé de trace** , un tracé similaire est automatiquement généré en fonction des données d'apprentissage complètes.

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

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

Sauvegarder

Indique les variables à enregistrer dans le jeu de données actif.

Valeurs prédites
Enregistre les valeurs prédites à partir du modèle spécifié ou du meilleur modèle dans le jeu de données actif. Vous pouvez également spécifier un **nom de variable personnalisé**.

Résidus
Enregistre les résidus des prévisions de modèle spécifiées ou des meilleures prévisions de modèle dans le jeu de données actif. Vous pouvez également spécifier un **nom de variable personnalisé**.

Régression Lasso linéaire

Linear Lasso utilise la classe Python sklearn.linear_model.Lasso pour estimer les modèles de régression linéaire régularisée de perte L1 pour une variable dépendante sur une ou plusieurs variables indépendantes, et inclut des modes facultatifs pour afficher les tracés de trace et sélectionner la valeur d'hyperparamètre alpha en fonction de la validation croisée. Lorsqu'un modèle unique est ajusté ou qu'une validation croisée est utilisée pour sélectionner alpha, une partition de données restantes peut être utilisée pour estimer les performances hors échantillon.

En plus de l'ajustement d'un modèle avec une valeur spécifiée du paramètre de régularisation alpha, le lasso linéaire peut afficher une trace des valeurs de coefficient pour une plage de valeurs alpha, ou faciliter le choix de la valeur d'hyperparamètre via une validation croisée d'un facteur k sur des grilles de valeurs spécifiées. Si un modèle unique est ajusté ou qu'une sélection alpha est effectuée via une validation croisée, le modèle final peut être appliqué aux données retenues qui sont créées par une partition des données d'entrée pour obtenir une estimation valide des performances hors échantillon du modèle.

Obtention d'une analyse de régression lasso linéaire

1. A partir des menus, sélectionnez :

   Analyse > **Régression** > **Alternatives OLS linéaires** > **Lasso**

   La boîte de dialogue permet de spécifier une variable qui affecte chaque observation du jeu de données actif à l'échantillon d'apprentissage ou restant.

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une variable cible numérique. Une seule variable cible est requise pour exécuter une analyse.
3. Indiquez une valeur numérique dépendante.
4. Spécifiez au moins une variable de facteur catégorielle ou une variable de covariable numérique.

En option, **Partition** permet de créer un sous-ensemble restant ou de test des données d'entrée pour l'estimation des performances hors échantillon du modèle spécifié ou choisi. Tous les partitionnages sont effectués après la suppression des observations incomplètes de toutes les observations dont les données ne sont pas valides pour toute variable utilisée par la procédure. Notez que pour la validation croisée, les niveaux ou les partitions des données d'apprentissage sont créés dans Python. Les données restantes créées par la partition ne sont pas utilisées pour l'estimation, quel que soit le mode en vigueur.

La partition peut être définie soit en spécifiant le rapport des observations affectées de manière aléatoire à chaque échantillon (sous **Partitions d'apprentissage et de rétention**), soit par une variable qui affecte chaque observation à l'échantillon d'apprentissage ou de rétention. Vous ne pouvez pas spécifier à la fois l'apprentissage et les variables. Si la partition n'est pas spécifiée, un échantillon restant d'environ 30% des données d'entrée est créé.

Le **% d'apprentissage** spécifie le nombre relatif d'observations dans le jeu de données actif à affecter de manière aléatoire à l'échantillon d'apprentissage. L'entraînement par défaut est de 70%.

Régression lasso linéaire: Options

L'onglet Options fournit des options pour:

Mode

Cette sélection fournit des options permettant de spécifier l'un des modes suivants:

Ajuster à l'alpha spécifié
Lorsque vous sélectionnez ce mode, un modèle unique est ajusté aux données d'apprentissage en utilisant une seule valeur de régularisation alpha. Cette valeur est définie par défaut. Si une partition

est spécifiée, le modèle unique ou final qui est ajusté est appliqué aux données de test retenues pour l'estimation des performances hors échantillon.

Sous **Tracé**, vous pouvez sélectionner les tracés des valeurs observées et / ou des résidus par rapport aux prévisions.

Sous **Enregistrer**, vous pouvez spécifier les valeurs prédites et les résidus à enregistrer.

Tracé de séries chronologiques
Lorsque vous sélectionnez ce mode, trois tracés pour les données d'apprentissage sont affichés en fonction de alpha pour l'ensemble de valeurs alpha spécifié:

- Tracé de trace des coefficients de régression.
- Un tracé de R^2.
- Tracé de l'erreur quadratique moyenne (MSE).

Bien que la partition soit respectée, aucun résultat pour les données de test retenues n'est fourni car aucun résultat de modèle final de ce mode.

Sélection alpha via la validation croisée

Lorsque vous sélectionnez une recherche de grille avec validation croisée pour évaluer les modèles et que vous choisissez la meilleure alpha basée sur la meilleure moyenne, R^2 sur les niveaux de validation. Le champ **Nombre de niveaux de validation croisée** peut être utilisé pour modifier la valeur par défaut de cinq divisions ou niveaux de validation croisée. Si une partition est spécifiée, le modèle unique ou final qui est ajusté est appliqué aux données de test retenues pour l'estimation des performances hors échantillon.

Sous **Afficher**, vous pouvez choisir d'afficher des informations de base uniquement sur le modèle avec la valeur alpha choisie (**Meilleur**), des informations de base sur tous les modèles comparés (**Comparer les modèles**) ou des informations complètes sur toutes les divisions ou tous les plis de tous les modèles (**Comparer les modèles et les plis**). **Best** est la valeur par défaut.

Sous **Tracé**, vous pouvez sélectionner des tracés de moyenne R^2 et/ou de MSE sur des niveaux de validation. Les tracés des valeurs observées et / ou des résidus par rapport aux valeurs prédites peuvent également être sélectionnés.

Sous **Enregistrer**, vous pouvez spécifier les valeurs prédites et les résidus à enregistrer.

Spécifier alphas individuels
Lorsque vous sélectionnez le mode **Ajuster avec alpha spécifié** , vous pouvez spécifier une valeur de régularisation alpha unique. Lorsque vous sélectionnez le mode **Tracé** ou **Sélection alpha via la validation croisée** , vous pouvez spécifier plusieurs valeurs.

Valeur(s)
Spécifiez une ou plusieurs valeurs de régularisation alpha positive. Plusieurs valeurs peuvent être spécifiées individuellement ou sous forme de plages. La valeur par défaut est 1.

Spécifier la grille de valeurs alpha
Lorsque vous sélectionnez le mode **Tracé** ou la sélection **Alpha via la validation croisée** , une grille de valeurs alpha uniques peut être spécifiée à partir d'une valeur **Début** (`value1`) à une valeur **Fin** (`value2`) avec l'incrément de **Par** (`value3`). S'il est spécifié, un seul ensemble valide de [`value1 TO value2 BY value3`] est autorisé. Il doit satisfaire la condition $0 \leq value1 \leq value2 \leq 1$. Dans les cas où `value1 = value2`, cela revient à spécifier un seul `value1`, indépendamment de `value3`.

La **métrique alpha** des plages de valeurs peut être soit **linéaire** , soit **logarithmique de base 10** (10 est élevé à la puissance des valeurs spécifiées).

Les tracés sont affichés à l'aide de la métrique spécifiée pour les axes X horizontaux des valeurs alpha variables.

Critères

Contrôle les analyses.

Inclure constante
Inclut une constante dans le (s) modèle (s) ajusté (s). Notez que la procédure d'extension ne centre pas ou ne normalise pas la variable dépendante et que la constante n'est pas pénalisée lors de l'estimation.

Normaliser les prédicteurs
Normalise toutes les variables indépendantes.

Nombre de niveaux de validation croisée
Nombre de divisions ou de niveaux pour l'évaluation de la validation croisée des modèles. Il doit s'agir d'un entier positif supérieur à 1. La valeur par défaut est 5.

Etat aléatoire Python
Valeur du paramètre random_state dans Python utilisée lors de l'évaluation de la validation croisée des modèles. Permet la réplication des résultats impliquant des nombres pseudo-aléatoires. Doit être un entier compris entre 0 et 2^{32}-1. La valeur par défaut est 0.

Limite de temps (minutes)
Nombre de minutes autorisé pour l'exécution des calculs de modèle. Si vous indiquez 0, le temporisateur est désactivé. La valeur par défaut est 5.

Afficher

Indique la quantité de sortie à afficher pour la **sélection alpha via le mode de validation croisée** .

Meilleures salutations,
Affiche uniquement les résultats de base pour le meilleur modèle choisi. Il s'agit de la valeur par défaut.

Comparer des modèles
Affiche les résultats de base pour tous les modèles évalués.

Comparer modèles et niveaux
Affiche les résultats prolixes complets pour chaque division ou pli pour chaque modèle évalué.

Tracé

Spécifie les tracés des valeurs observées ou résiduelles par rapport aux valeurs prédites, et avec la validation croisée, la spécification des tracés de l'erreur quadratique moyenne (MSE) et/ou de la moyenne R^2 sur les niveaux de validation croisée par rapport aux valeurs alpha.

Moyenne de l'erreur quadratique moyenne de validation croisée (MSE) par rapport à l'erreur alpha
Pour le mode **Sélection alpha via la validation croisée** , affiche un tracé en ligne de MSE moyenne sur des niveaux de validation croisée par rapport à alpha. Pour le mode **Tracé de trace** , un tracé similaire est automatiquement généré en fonction des données d'apprentissage complètes.

Moyenne R-carré de la validation croisée par rapport à alpha
Pour le mode **Sélection alpha via la validation croisée** , affiche un tracé en ligne représentant la moyenne de R^2 sur les niveaux de validation croisée par rapport à alpha. Pour le mode **Tracé de trace** , un tracé similaire est automatiquement généré en fonction des données d'apprentissage complètes.

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

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

Sauvegarder

Indique les variables à enregistrer dans le jeu de données actif.

Valeurs prédites
Enregistre les valeurs prédites à partir du modèle spécifié ou du meilleur modèle dans le jeu de données actif. Vous pouvez également spécifier un **nom de variable personnalisé**.

Résidus
Enregistre les résidus des prévisions de modèle spécifiées ou des meilleures prévisions de modèle dans le jeu de données actif. Vous pouvez également spécifier un **nom de variable personnalisé**.

Régression Ridge linéaire

Linear Ridge utilise la classe Python sklearn.linear_model.Ridge pour estimer les modèles de régression linéaire régularisée de type L2 ou carré des pertes pour une variable dépendante sur une ou plusieurs variables indépendantes, et inclut des modes facultatifs pour afficher les tracés de trace et sélectionner la valeur d'hyperparamètre alpha basée sur la validation croisée. Lorsqu'un modèle unique est ajusté ou qu'une validation croisée est utilisée pour sélectionner alpha, une partition de données restantes peut être utilisée pour estimer les performances hors échantillon.

En plus de l'ajustement d'un modèle avec une valeur spécifiée du paramètre de régularisation alpha, l'arête linéaire peut afficher une trace d'arête des valeurs de coefficient pour une plage de valeurs alpha, ou faciliter le choix de la valeur de l'hyperparamètre via une validation croisée d'un facteur k sur des grilles de valeurs spécifiées. Si un modèle unique est ajusté ou qu'une sélection alpha est effectuée via une validation croisée, le modèle final peut être appliqué aux données retenues qui sont créées par une partition des données d'entrée pour obtenir une estimation valide des performances hors échantillon du modèle.

Obtention d'une analyse de régression linéaire de crête

1. A partir des menus, sélectionnez :

   Analyser > **Régression** > **Solutions de rechange OLS linéaires** > **Ridge**

   La boîte de dialogue permet de spécifier une variable qui affecte chaque observation du jeu de données actif à l'échantillon d'apprentissage ou restant.
2. Sélectionnez une variable cible numérique. Une seule variable cible est requise pour exécuter une analyse.
3. Indiquez une valeur numérique dépendante.
4. Spécifiez au moins une variable de facteur catégorielle ou une variable de covariable numérique.

En option, **Partition** permet de créer un sous-ensemble restant ou de test des données d'entrée pour l'estimation des performances hors échantillon du modèle spécifié ou choisi. Tous les partitionnages sont effectués après la suppression des observations incomplètes de toutes les observations dont les données ne sont pas valides pour toute variable utilisée par la procédure. Notez que pour la validation croisée, les niveaux ou les partitions des données d'apprentissage sont créés dans Python. Les données restantes créées par la partition ne sont pas utilisées dans l'estimation, quel que soit le mode en vigueur.

La partition peut être définie soit en spécifiant le rapport des observations qui sont affectées de manière aléatoire à chaque échantillon (sous **Partitions d'apprentissage et de rétention**), soit par une variable qui affecte chaque observation à l'échantillon d'apprentissage ou de rétention. Vous ne pouvez pas spécifier à la fois l'apprentissage et les variables. Si la partition n'est pas spécifiée, un échantillon restant d'environ 30% des données d'entrée est créé.

Le **% d'apprentissage** spécifie le nombre relatif d'observations dans le jeu de données actif à affecter de manière aléatoire à l'échantillon d'apprentissage. L'entraînement par défaut est de 70%.

Régression linéaire de crête: Options

L'onglet Options fournit des options pour:

Mode

Cette sélection fournit des options permettant de spécifier l'un des modes suivants:

Ajuster à l'alpha spécifié
Lorsque vous sélectionnez cette option, un modèle unique est ajusté aux données d'apprentissage qui utilisent une seule valeur de régularisation alpha. Cette valeur est définie par défaut. Si une partition est spécifiée, le modèle unique ou final qui est ajusté est appliqué aux données de test retenues pour l'estimation des performances hors échantillon.

Sous **Tracé**, vous pouvez sélectionner les tracés des valeurs observées et / ou des résidus par rapport aux prévisions.

Sous **Enregistrer**, vous pouvez spécifier les valeurs prédites et les résidus à enregistrer.

Tracé de séries chronologiques
Lorsque vous sélectionnez cette option, trois tracés pour les données d'apprentissage sont affichés en tant que fonction alpha pour l'ensemble de valeurs alpha spécifié:

- Tracé de trace de crête des coefficients de régression.
- Un tracé de R^2.
- Tracé de l'erreur quadratique moyenne (MSE).

Bien que la partition soit respectée, aucun résultat pour les données de test retenues n'est fourni car aucun résultat de modèle final de ce mode.

Sélection alpha via la validation croisée

Lorsque vous sélectionnez cette option, une recherche de grille avec validation croisée est effectuée pour évaluer les modèles, et la meilleure alpha est choisie en fonction de la meilleure moyenne R^2 sur les niveaux de validation. Le champ **Nombre de niveaux de validation croisée** peut être utilisé pour modifier la valeur par défaut de cinq divisions ou niveaux de validation croisée. Si une partition est spécifiée, le modèle unique ou final qui est ajusté est appliqué aux données de test retenues pour l'estimation des performances hors échantillon.

Sous **Afficher**, vous pouvez choisir d'afficher des informations de base uniquement sur le modèle avec la valeur alpha choisie (**Meilleur**), des informations de base sur tous les modèles comparés (**Comparer les modèles**) ou des informations complètes sur toutes les divisions ou tous les plis de tous les modèles (**Comparer les modèles et les plis**). **Best** est la valeur par défaut.

Sous **Tracé**, vous pouvez sélectionner des tracés de moyenne R^2 et/ou de MSE sur des niveaux de validation. Les tracés des valeurs observées et / ou des résidus par rapport aux valeurs prédites peuvent également être sélectionnés.

Sous **Enregistrer**, vous pouvez spécifier les valeurs prédites et les résidus à enregistrer.

Spécifier alphas individuels
Lorsque vous sélectionnez le mode **Ajuster avec alpha spécifié** , vous pouvez spécifier une valeur de régularisation alpha unique. Lorsque vous sélectionnez le mode **Tracé** ou **Sélection alpha via la validation croisée** , vous pouvez spécifier plusieurs valeurs.

Valeur(s)
Spécifiez une ou plusieurs valeurs de régularisation alpha positive. Vous pouvez spécifier plusieurs valeurs individuellement ou sous forme de plages. La valeur par défaut est 1.

Spécifier la grille de valeurs alpha
Lorsque vous sélectionnez le mode **Tracé** ou la sélection **Alpha via la validation croisée** , une grille de valeurs alpha uniques peut être spécifiée à partir d'une valeur **Début** (`value1`) à une valeur **Fin** (`value2`) avec l'incrément de **Par** (`value3`). S'il est spécifié, un seul ensemble valide de [`value1 TO value2 BY value3`] est autorisé. Il doit satisfaire la condition $0 \leq value1 \leq value2 \leq 1$. Dans les cas où `value1 = value2`, cela revient à spécifier un seul `value1`, indépendamment de `value3`.

La **métrique alpha** des plages de valeurs peut être soit **linéaire** , soit **logarithmique de base 10** (10 est élevé à la puissance des valeurs spécifiées).

Les tracés sont affichés à l'aide de la métrique spécifiée pour les axes X horizontaux des valeurs alpha variables.

Critères

Contrôle les analyses.

Inclure constante
Ce critère inclut une constante dans un ou plusieurs modèles ajustés. Notez que la procédure d'extension ne centre pas ou ne normalise pas la variable dépendante et que la constante n'est pas pénalisée lors de l'estimation.

Normaliser les prédicteurs
Ce critère normalise toutes les variables indépendantes.

Nombre de niveaux de validation croisée
Utilisez ce critère pour définir le nombre de divisions ou de niveaux pour l'évaluation de la validation croisée des modèles. Le nombre doit être un entier positif supérieur à 1. La valeur par défaut est 5.

Etat aléatoire Python
La valeur du paramètre random_state dans Python est utilisée lors de l'évaluation de la validation croisée des modèles. Permet la réplication des résultats impliquant des nombres pseudo-aléatoires. Doit être un entier compris entre 0 et 2^{32}-1. La valeur par défaut est 0.

Limite de temps (minutes)
Nombre de minutes autorisé pour l'exécution des calculs de modèle. Si vous indiquez 0, le temporisateur est désactivé. La valeur par défaut est 5.

Afficher

Cette section indique le nombre de sorties à afficher pour la **sélection alpha via le mode de validation croisée** .

Meilleures salutations,
Il affiche uniquement les résultats de base pour le meilleur modèle choisi. Cette valeur est définie par défaut.

Comparer des modèles
Il affiche également les résultats de base de tous les modèles évalués.

Comparer modèles et niveaux
Enfin, il affiche les résultats complets en mode prolixe pour chaque division ou pli pour chaque modèle évalué.

Tracé

Spécifie les tracés des valeurs observées ou résiduelles par rapport aux valeurs prédites, avec validation croisée, la spécification des tracés de l'erreur quadratique moyenne (MSE) et/ou de la moyenne R^2 sur les niveaux de validation croisée par rapport aux valeurs alpha.

Moyenne de l'erreur quadratique moyenne de validation croisée (MSE) par rapport à l'erreur alpha
Le mode **Sélection alpha via validation croisée** affiche un tracé de courbes de MSE moyenne sur des niveaux de validation croisée par rapport à alpha. Pour le mode **Tracé de trace** , un tracé similaire est automatiquement généré en fonction des données d'apprentissage complètes.

Moyenne R-carré de la validation croisée par rapport à alpha
Le mode **Sélection alpha via la validation croisée** affiche un tracé de la moyenne de R^2 sur les niveaux de validation croisée par rapport à alpha. Pour le mode **Tracé de trace** , un tracé similaire est automatiquement généré en fonction des données d'apprentissage complètes.

Observé / Prévu
Il affiche un nuage de points des valeurs observées par rapport aux valeurs prédites pour le modèle spécifié ou le meilleur modèle.

Résidus par rapport aux prévisions
Il affiche également un nuage de points des résidus par rapport aux valeurs prédites pour le modèle spécifié ou le meilleur modèle.

Sauvegarder

Indique les variables à enregistrer dans le jeu de données actif.

Valeurs prédites
Enregistre les valeurs prédites à partir du modèle spécifié ou du meilleur modèle dans le jeu de données actif. Sinon, vous pouvez spécifier un **nom de variable personnalisée**.

Résidus
Enregistre les résidus des prévisions de modèle spécifiées ou des meilleures prévisions de modèle dans le jeu de données actif. Sinon, vous pouvez spécifier un **nom de variable personnalisée**.

Analyse de la médiation

L'analyse de médiation est une méthode statistique utilisée pour comprendre le mécanisme par lequel une variable indépendante influence une variable dépendante par l'intermédiaire d'une ou plusieurs variables médiatrices. Cette procédure vous aide à tester des hypothèses concernant les effets indirects et à comprendre les mécanismes d'influence au sein de vos données.

La procédure d'analyse de médiation dans IBM SPSS Statistiques offre les fonctionnalités suivantes :

- Analyser les effets directs et indirects entre les variables prédictives, les variables médiatrices et les variables de résultat
- Décomposer les effets globaux en leurs composantes pour comprendre les mécanismes causaux
- Gérer divers types de variables, notamment les données continues, binaires et de comptage
- Contrôle des variables de confusion par ajustement des covariables
- Réaliser des analyses par sous-groupes afin d'évaluer l'hétérogénéité des effets
- Représenter les relations de médiation à l'aide de diagrammes de cheminement et d'autres graphiques

Concepts clés

- **Variable prédictrice (A)** : la variable indépendante dont vous souhaitez examiner l'effet (par exemple, l'exposition au traitement, le niveau d'intervention).
- **Variable médiatrice (M)** : variable intermédiaire qui transmet l'effet de la variable prédictive à la variable dépendante (par exemple, état psychologique, marqueur biologique).
- **Variable de résultat (Y)** : variable dépendante influencée à la fois par la variable prédictive et par la variable médiatrice (par exemple, état de santé, indicateur de performance).
- **Effet direct** : effet de la variable prédictive sur la variable dépendante qui ne passe pas par la variable médiatrice.
- **Effet indirect** : effet de la variable prédictive sur la variable dépendante qui s'exerce par l'intermédiaire du médiateur.
- **Effet total** : l'effet global de la variable prédictive sur la variable dépendante, qui tient compte à la fois des voies directes et indirectes.

Applications
L'analyse de médiation est largement utilisée dans de nombreuses disciplines :

- **Sciences de la santé** : Comprendre comment les interventions influencent les résultats en matière de santé par le biais de mécanismes comportementaux ou biologiques
- **Psychologie** : Étude de l'influence des traitements sur la santé mentale par le biais de processus cognitifs ou émotionnels
- **Sciences sociales** : analyser l'impact des politiques sur les résultats sociaux par le biais de facteurs intermédiaires

- **Domaine d'activité** : Étudier l'influence des stratégies marketing sur les ventes à travers le prisme des attitudes des clients
- **Enseignement** : Étudier l'influence des méthodes pédagogiques sur les résultats d'apprentissage à travers l'implication des élèves

Cette procédure est particulièrement utile lorsque vous devez :

- Tester des modèles théoriques sur les mécanismes causaux
- Identifier les cibles d'intervention en comprenant les mécanismes de médiation
- Évaluer l'importance relative des différentes voies causales
- Évaluer si les effets varient selon les sous-groupes
- Prise en compte des variables de confusion dans les études observationnelles

Types d'analyse

La procédure prend en charge deux méthodes d'analyse :

- **Analyse de médiation causale** : calcule les effets causaux complets, notamment l'effet direct contrôlé (CDE), l'effet direct naturel (NDE), l'effet indirect naturel (NIE) et les composantes de décomposition. Cette approche tient compte des interactions entre prédicteurs et médiateurs et permet une décomposition détaillée des effets.
- **Analyse de médiation classique** : calcule des effets de médiation plus simples, sans termes d'interaction ni décomposition détaillée. Cette approche est indiquée lorsque l'on ne s'attend pas à des interactions entre la variable prédictive et la variable médiatrice.

Combinaisons de variables prises en charge

La procédure prend en charge différentes combinaisons de types de variables :

- Variable prédictive continue, variable médiatrice continue, variable de résultat continue
- Variable prédictive continue, médiateur binaire, variable de résultat continue
- Variable prédictive continue, variable médiatrice continue, résultat binaire
- Variable prédictive continue, médiateur binaire, résultat binaire
- Variable prédictive continue, variable médiatrice continue, variable de résultat discrète

Statistical Methods

Cette procédure utilise des méthodes basées sur la régression pour estimer les effets de médiation :

- Régression linéaire pour les variables continues
- Régression logistique pour les variables binaires
- Régression de Poisson ou régression binomiale négative pour les variables de type compté

Les erreurs-types et les intervalles de confiance peuvent être calculés soit à l'aide de la méthode delta (par défaut), soit par rééchantillonnage bootstrap. La méthode du delta permet un calcul rapide grâce à la théorie asymptotique, tandis que le bootstrapping offre une inférence plus robuste pour les distributions non normales ou les petits échantillons.

Statistiques

La procédure fournit des résultats statistiques complets, classés en plusieurs catégories :

Résumés des modèles de régression

Modèle de régression médiatisée :

- Coefficients (β_0, β_1, β_2) avec leurs erreurs-types et leurs valeurs p
- Pour les variables continues : $E(M|a,c) = \beta_0 + \beta_1 a + \beta_2 c$
- Pour les variables binaires : modèle de régression logistique

Modèle de régression des résultats :

- Coefficients (θ_0, θ_1, θ_2, θ_3, θ_4) avec leurs erreurs-types et leurs valeurs de p

- Pour les variables continues : $E(Y|a,m,c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 (a \times m) + \theta_4 c$
- Pour les résultats binaires : modèle de régression logistique
- Pour les données de type compté : régression de Poisson ou régression binomiale négative

Mesures de l'effet

Statistiques comparatives des niveaux d'exposition :

- **CDE (effet direct contrôlé)** : effet du prédicteur sur la variable dépendante à un niveau fixe du médiateur
- **NDE (effet direct naturel)** : effet direct permettant au médiateur de varier naturellement
- **NIE (effet indirect naturel)** : effet induit par l'intermédiaire d'un médiateur
- Présentés sous forme de différences pour les résultats continus, de rapports de cotes (RC) ou de excRR pour les résultats binaires

Résultats de l'analyse de décomposition :

- **TE (effet total)** : effet global du prédicteur sur la variable dépendante
- **PIE (effet indirect pur)** : composante de l'effet indirect
- **TDE (effet direct total)** : composante de l'effet direct
- **PAI (interaction pure)** : interaction attribuable au prédicteur
- **INTmed (Interaction médiatisée)** : Interaction par l'intermédiaire d'un médiateur
- **INTref (Interaction de référence)** : effet d'interaction de référence

Statistiques de cotes (pour les résultats binaires) :

- **PM (proportion médiatisée)** : proportion de l'effet total qui est médiatisée
- **Pourcentage (ENTIER)** : pourcentage dû à l'interaction
- **Pourcentage (ELIM)** : Pourcentage éliminé

Méthodes d'inférence

Méthode Delta (par défaut) :

- Erreurs-types calculées à partir des matrices de covariance ($\Sigma\beta$, $\Sigma\theta$)
- Intervalles de confiance basés sur la distribution normale
- Calcul rapide adapté à la plupart des analyses

Méthode Bootstrap :

- Erreurs-types et intervalles de confiance basés sur le rééchantillonnage
- Par défaut : 1 000 échantillons de bootstrap avec un niveau de confiance de 95 %
- Intervalles de percentile ou BCa (corrigés du biais et accélérés)
- Plus robuste pour les distributions non normales

Options de la matrice de covariance :

- **MODÈLE** : utilise la variance supposée du modèle (résultats optimaux lorsque les données s'adaptent bien au modèle)
- **ROBUST** : tient compte des valeurs aberrantes et de l'hétéroscédasticité (recommandé pour les variables binaires)

Options de visualisation

Diagramme de cheminement :

- Représentation visuelle des chemins de médiation
- Présente les effets directs et indirects avec l'ampleur de ces effets
- Affichage facultatif des effets des covariables

Nuage de points :

- Relations entre la variable prédictive, la variable médiatrice et la variable de résultat
- Lignes de tendance (linéaires ou logistiques) avec intervalles de confiance

Diagramme à barres :

- Nombres ou proportions pour les variables catégorielles
- Barres d'erreur et intervalles de confiance

Forest Plot :

- Résumé des mesures d'effet avec leurs intervalles de confiance
- Affiche les coefficients de cheminement ou les estimations des effets

Analyse des données

Tableau 1. Exigences variables

Variables	Condition requise
Variable de sortie	• Données continues, binaires ou discrètes • Pour les résultats binaires : indiquer la catégorie de référence (DERNIER, PREMIER ou une VALEUR spécifique) • Pour les variables de comptage : choisissez entre la distribution de Poisson et la distribution binomiale négative
Variable de médiation	• Données continues ou binaires • Pour les variables continues : précisez le niveau d'analyse (MOYENNE, MÉDIANE ou une VALEUR spécifique) • Pour les variables binaires : indiquer la catégorie de référence
Variable prédictive	• Données continues ou catégorielles • Indiquez le niveau de référence (BLEVEL) et le niveau d'exposition (ELEVEL) pour les comparaisons d'effets • Terme d'interaction prédicteur-médiateur facultatif (INTERACTION = VRAI/FAUX)

Spécifications de modèles

La procédure s'appuie sur le cadre des modèles linéaires généralisés (GLM), ce qui nécessite de spécifier à la fois la distribution et la fonction de liaison.

Pour le modèle du médiateur	Pour le modèle des résultats
• **Une distribution normale avec un lien identique** implique une régression linéaire pour les médiateurs continus • **Une distribution binomiale avec un lien logit** implique une régression logistique pour les médiateurs binaires	• **Une distribution normale avec un lien identitaire** implique une régression linéaire pour les résultats continus • **La distribution binomiale avec fonction de liaison logit** correspond à une régression logistique pour les résultats binaires

Pour le modèle du médiateur	Pour le modèle des résultats
	• **Une distribution de Poisson avec un lien logarithmique** implique une régression de Poisson pour les données de comptage (en supposant que la variance est égale à la moyenne) • **Modèle binomial négatif avec lien logarithmique** pour les données de comptage surdispersées (la variance dépasse la moyenne) – **NEGBIN(1)** : paramètre de dispersion fixe α = 1 – **NEGBIN(MLE)** : estime le paramètre de dispersion à partir des données

Le paramètre de dispersion α contrôle la surdispersion : $Var(Y) = \mu + \alpha\mu^2$. Lorsque $\alpha = 0$, le modèle se réduit au modèle de Poisson. Des valeurs α plus élevées indiquent une plus grande variabilité par rapport à la moyenne.

Taille de l'échantillon et convergence

Tableau 2. Paramètres d'algorithme

Paramètre	Descriptif	Plage
MAXITERATIONS	Nombre maximum d'itérations	de 10 à 1 000 Augmenter à 200-500 pour les modèles complexes
PCONVERGE	Tolérance de convergence	1E-8 sur 1E-3 Utilisez 1E-8 pour les publications et 1E-4 pour les analyses exploratoires
Singulières	Seuil de détection des singularités	valeur numérique comprise entre 1E-14 et 1E-6
MAXSTEPHALVING	Tentatives de réduction de moitié par paliers	1 à 20
PÉNALITÉ DE RIDGE	Régularisation des crêtes	1E-8 sur 1E-3

Considérations Qualité de données

Données manquantes

- Les cas présentant des valeurs manquantes au niveau des variables d'analyse sont exclus
- Rapports récapitulatifs sur le traitement des dossiers : cas valides et cas exclus

Pondérations

- La prise en charge des pondérations de fréquence s'effectue à l'aide de la variable « Scale Weight »
- Le rééchantillonnage par la méthode Bootstrap arrondit les pondérations à l'entier le plus proche

Covariables

- Inclure des covariables pour tenir compte des facteurs de confusion
- Tenir compte des facteurs susceptibles d'influencer à la fois le médiateur et le résultat

Analyse de sous-groupe

- Utilisez la clause BY pour effectuer des analyses distinctes en fonction de différents facteurs (par exemple, le sexe, la tranche d'âge)
- Utile pour examiner l'hétérogénéité des effets entre les groupes

Hypothèses

Hypothèses du modèle

- Aucun facteur de confusion non mesuré entre les relations prédicteur-médiateur, prédicteur-résultat et médiateur-résultat
- Spécification correcte des fonctions de distribution et de liaison
- Pour l'interprétation causale : ordre temporel (prédicteur → médiateur → résultat)
- Aucune erreur de mesure dans les variables

Hypothèses statistiques

- Taille d'échantillon appropriée pour obtenir des estimations stables des paramètres
- Pour le bootstrap : un nombre suffisant d'échantillons (au moins 1 000 recommandés)
- Pour les résultats binaires : nombre suffisant d'événements par variable (au moins 10 à 15 recommandés)
- Pour les modèles d'interaction : une puissance statistique suffisante pour détecter les effets d'interaction

Obtenir une analyse de médiation

1. A partir des menus, sélectionnez :

   Analyser > **Analyse de médiation**

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

Type de modèle

Dans l'onglet « **Type de modèle** », définissez la nature statistique de l'analyse de médiation en sélectionnant le **type de médiateur** et **le type de variable dépendante**.

Sélectionnez le type de modèle de médiation que vous souhaitez analyser. Les paramètres définis dans **les onglets « Type de médiateur** » et « **Type de résultat** » déterminent la distribution sous-jacente, la fonction de liaison et les variables autorisées dans les onglets suivants. Les choix que vous effectuez dans cet onglet déterminent les options disponibles dans les onglets « **Variables dépendantes** » et « **Critères** ».

Les options « Type de médiateur » et « Type de résultat » sont obligatoires.

Type de médiateur

Sélectionnez la manière dont la variable médiatrice est modélisée.

Continu
Suppose une distribution normale avec une fonction de liaison identitaire. Seules les variables continues sont autorisées pour le médiateur.

Utilisez cette option lorsque la variable médiatrice est mesurée sur une échelle continue.

Binaire

Suppose une distribution binomiale avec une fonction de liaison LOGIT. Seules les variables binaires sont autorisées pour le médiateur.

Utilisez cette option lorsque le médiateur peut prendre deux valeurs différentes.

Type de résultat

Choisissez la manière dont la variable de résultat est modélisée. Les tableaux de résultats sont générés lorsque la prévalence de la variable de résultat est d'au moins 0.30.

Continu
Suppose une distribution normale avec une fonction de liaison identitaire.

Sélectionnez l'option permettant de définir un type de résultat continu, ce qui implique une distribution NORMALE des données et une fonction de liaison IDENTITY. Si vous sélectionnez cette option, vous pouvez configurer la **variable de résultat** dans l'onglet « **Variables dépendantes** » afin de n'autoriser que les variables continues.

Binaire

Suppose une distribution binomiale. Vous pouvez choisir entre une fonction de liaison **LOGIT** ou **LOG**.

Sélectionnez l'option permettant de définir un résultat binaire avec une distribution binomiale et une fonction de liaison LOGIT. Si vous sélectionnez cette option, vous pouvez configurer le champ « **Variable de résultat** » dans l'onglet « **Variables dépendantes** » et activer l'option « Binaire » comme variable de résultat.

Dans la liste déroulante « Fonction de liaison », vous pouvez choisir « **Logit** » (par défaut) ou « Fonction de liaison **logarithmique** pour résultat binaire ». Par défaut, l'option « **Binary** » définit la distribution binomiale et la fonction de liaison LOGIT. Dans l'onglet « **Variables dépendantes** », la variable de résultat peut être définie comme « **Binaire** avec **catégorie de référence** » afin de choisir parmi des valeurs binaires.

Effectifs
Sélectionnez cette option pour configurer la variable « counts » comme **variable de résultat** dans l'onglet « **Variables dépendantes** ». Vous pouvez ajouter des distributions **de Poisson** ou **binomiales négatives** avec une fonction de liaison logarithmique. Des paramètres supplémentaires sont disponibles pour l'option « **Binomial négatif** ». Vous pouvez définir le paramètre de dispersion α qui détermine le degré de surdispersion dans le modèle. Des valeurs élevées de α indiquent une plus grande variabilité par rapport à la moyenne.

- **Indiquez une valeur** : saisissez un nombre entier positif fixe.
- **Valeur estimée** : sélectionnez cette option si vous souhaitez estimer le paramètre de dispersion α à l'aide de la méthode du maximum de vraisemblance.

Variables dépendantes

Dans l'onglet « **Variables dépendantes** », vous pouvez définir la variable de résultat, qui détermine le résultat de l'analyse de médiation.

Indiquez la variable de réponse (résultat), avec des options de personnalisation pour les variables continues, binaires (y compris la catégorie de référence) et de comptage. Elle définit également la variable médiatrice avec des options pour les variables continues et binaires (y compris la catégorie de référence). Vous pouvez définir des niveaux de contrôle pour l'analyse des effets et pour l'intégration d'une variable de pondération d'échelle.

Vous pouvez également indiquer la **variable** « Mediator » dans le champ « Mediator », qui définit le composant médiateur pour l'analyse de médiation. Toutes les variables doivent être uniques au sein de chaque catégorie et entre les variables dépendantes, les variables indépendantes et la variable de pondération de l'échelle.

Variable de sortie
Définit la variable de résultat pour l'analyse de médiation. Fournissez des variables continues, binaires et de type « nombre ». Cela dépend du type de résultat que vous avez spécifié dans l'onglet « **Type de modèle** »; la variable continue est sélectionnée par défaut. Si vous optez pour un résultat binaire, vous

pouvez choisir des variables binaires. Si vous optez pour la distribution de Poisson ou la distribution binomiale négative, vous pouvez également sélectionner des variables de comptage.

Variable à résultat binaire
Vous pouvez choisir « binaire » comme variable de résultat pour la régression. Si un résultat binaire est sélectionné, vous pouvez choisir une valeur de référence pour la variable binaire sélectionnée. Par défaut, une distribution binomiale et une fonction de liaison LOGIT/LOG.

Les tableaux de résultats sont générés lorsque la prévalence de la variable de résultat est d'au moins 0.30.

Catégorie de référence
Cette option vous permet de définir la catégorie de référence pour le type de résultat binaire que vous avez sélectionné dans l'onglet « **Type de modèle** ». Sélectionnez l'une des options suivantes :

- **Dernière valeur (valeur la plus élevée)** : définir la dernière valeur ou la valeur la plus élevée comme référence.
- **Première (valeur la plus basse)** : définir la première valeur ou la valeur la plus basse comme référence.
- **Personnalisé** : sélectionnez cette option pour définir une sélection de référence personnalisée. Vous pouvez sélectionner les niveaux de la variable binaire choisie dans la variable de résultat ou saisir une valeur, à condition que celle-ci figure dans la liste des variables.

Variable de médiation
Cette option vous permet de définir la **variable médiatrice** pour l'analyse de médiation. Le type de données doit être continu et binaire.

Cette option dépend du **type** de médiateur que vous avez spécifié dans l'onglet « **Type de modèle** »; par défaut, il s'agit d'une variable continue, d'une distribution normale et d'une fonction de liaison de type IDENTITY. Si l'on choisit un médiateur binaire, on peut sélectionner des variables binaires avec une distribution binomiale et une fonction de liaison de type LOGIT.

Définir le niveau du médiateur
Vous pouvez sélectionner un niveau de médiateur à utiliser pour le calcul de l'effet direct contrôlé. Sélectionnez l'une des options suivantes :

- **Moyenne** : sélectionnez cette option pour définir le niveau du médiateur sur la moyenne de la variable médiateur sélectionnée.
- **Médiane** : sélectionnez cette option pour définir le niveau du médiateur sur la médiane de la variable médiateur sélectionnée.
- **Personnalisé** : sélectionnez cette option pour définir un niveau personnalisé. Cette option ne peut être sélectionnée que pour les variables binaires. Entrez une valeur numérique correspondant au niveau du médiateur. La valeur doit figurer dans les valeurs observées binaires du Mediator. Si le médiateur est binaire, l'utilisateur peut sélectionner le niveau binaire (0 ou 1).

Pondération d'échelle
Sélectionnez l'option permettant d'attribuer une variable de pondération. Ce champ n'accepte que les variables de type numérique.

Prédicteur

Dans l'onglet « **Predictor** », vous pouvez définir la variable prédictive dans le champ « **Primary Predictor** », qui désigne la variable **prédictive** ou la variable **d'exposition**, notée *A*, dans l'analyse de médiation.

Spécifiez la variable prédictive principale ainsi que les facteurs et covariables supplémentaires, tout en ayant la possibilité de définir les valeurs de référence et d'exposition pour la variable prédictive afin de faciliter l'analyse des effets.

Les valeurs observées de la variable prédictive dans l'ensemble de données sont représentées par le vecteur a.

Facteur prédictif principal
Définit la variable prédictive principale ou variable d'exposition A pour l'analyse de médiation.

Seules les variables continues sont autorisées (les variables catégorielles sont exclues afin de garantir la clarté dans la définition de l'effet de l'exposition).

Définir le niveau de référence
Cette option vous permet de définir les niveaux de référence. Saisissez des valeurs dans les champs « **Définir le niveau de référence** » et « **Définir le nouveau niveau d'exposition** ». Les deux valeurs doivent figurer dans l'ensemble de données pour la variable choisie comme prédicteur.

- Définir le niveau de référence : le niveau le plus bas (ou la première valeur) de la variable prédictive dans son ordre naturel.
- Définir un nouveau niveau d'exposition : le niveau immédiatement supérieur (ou la deuxième valeur) de la variable prédictive, dans le même ordre.

Facteurs
Cette liste définit les facteurs F qui représentent les variables catégorielles dans le modèle. Chaque facteur est converti en un ensemble de variables virtuelles qui seront incluses dans la matrice C. Ces variables fictives sont incluses dans la régression des résultats afin de tenir compte des différences entre les groupes.

Les variables catégorielles sont autorisées (par exemple, les groupes de traitement, le sexe, etc.).

Ces facteurs sont facultatifs, mais si votre analyse porte sur des groupes catégoriels ou si vous évaluez des effets d'interaction, ces variables sont obligatoires.

Covariables
Cette liste définit les covariables *C* qui permettent de contrôler les effets de confusion potentiels dans le modèle. Chaque covariable est incluse dans le modèle de régression du résultat Y.

Si les covariables sont catégorielles, elles doivent être converties en variables fictives avant d'être intégrées au modèle afin d'éviter toute redondance.

Chaque covariable est incluse en tant qu'effet principal dans le modèle de régression des résultats Y.

Critères

Dans l'onglet « **Critères** », vous pouvez définir des options permettant de choisir les méthodes d'estimation pour la matrice des covariables, de calculer les estimations des paramètres et de spécifier les types d'inférence, le type d'analyse ainsi que les options de réglage de l'algorithme.

Matrice de covariance
Détermine la méthode utilisée pour estimer la matrice de covariance des estimations des paramètres dans l'analyse de médiation.

- **Estimateur basé sur un modèle** : calcule la matrice de covariance à partir des hypothèses du modèle statistique spécifié.
- **Estimateur robuste** : fournit une estimation robuste de la matrice de covariance qui tient compte des violations du modèle, telles que l'hétéroscédasticité ou la non-normalité.

Type d'inférence
Cette option permet de définir la méthode utilisée pour l'inférence dans l'analyse de médiation.

- **Delta** : cette méthode utilise la distribution asymptotique de l'estimateur pour calculer les erreurs-types et les intervalles de confiance.
- **Bootstrap** : cette méthode utilise le rééchantillonnage pour générer une distribution de l'estimateur à des fins d'inférence, y compris les intervalles de confiance.

  Options :

  – **Nombre d'échantillons** : pour les intervalles de percentile et les intervalles BCa générés, il est recommandé d'utiliser au moins 1 000 échantillons de bootstrap. Spécifiez un nombre entier positif. La valeur par défaut est 500.

- **Définir la graine pour Mersenne Twister** : la définition d'une graine vous permet de reproduire les analyses. L'utilisation de ce contrôle revient à définir le Mersenne Twister comme le générateur actif et à spécifier un point de départ fixe dans la boîte de dialogue Générateurs de nombres aléatoires. La différence notoire est que la définition de la valeur de départ dans cette boîte de dialogue conserve l'état actuel du générateur de nombres aléatoires et restaure cet état une fois l'analyse terminée.
- **Intervalles de confiance** : indiquez un niveau de confiance supérieur à 50 et inférieur à 100. Les intervalles de percentile utilisent seulement des valeurs de bootstrap ordonnées correspondant aux percentiles d'intervalle de confiance. Par exemple, un intervalle de confiance de percentile de 95 % utilise les 2,5e et 97,5e percentiles des valeurs de bootstrap comme bornes inférieure et supérieure de l'intervalle (en interpolant des valeurs de bootstrap si nécessaire). Les intervalles de biais corrigé et accéléré (BCa) sont des intervalles ajustés plus précis, toutefois ils requièrent plus de temps de calcul.

Type d'analyse

- **Causal** : Calcule les effets de médiation causale complets, notamment les CDE, les NDE, les NIE, les effets totaux, les rapports de cotes, les risques relatifs excédentaires et les décompositions d'interaction.
- **Traditionnel** : calcule des effets de médiation plus simples, sans termes d'interaction ni décomposition causale complète.

Options d'algorithme

Les options sont activées lorsque le type de résultat est défini sur « **Binaire** » ou « **Nombre** » dans l'onglet « **Type de** modèle ».

- **Nombre maximal d'itérations** : définit le nombre maximal d'itérations IRWLS avant que l'algorithme ne s'arrête. Vous pouvez saisir des nombres entiers compris entre 10 et and1000. La valeur par défaut est 100.
- **Tolérance de convergence** : définit le seuil de convergence des paramètres (variation absolue de la norme du vecteur des coefficients).

  Seuil de convergence des paramètres. La valeur par défaut est 1E-6. L'algorithme s'arrête lorsque la variation absolue des coefficients passe en dessous de cette valeur. Utilisez 1E-8 pour obtenir une précision digne d'une publication et 1E-4 pour des analyses exploratoires plus rapides.

  Entrez une valeur comprise entre 1E-8 et 1E-3.

  Ce champ accepte les formats de notation scientifique suivants : 1E-6, 1e-6, 0.000001
- **Tolérance à la singularité** : définit le seuil de détection des matrices hessiennes singulières (seuil du déterminant). Seuil de détection des matrices hessiennes singulières. La valeur par défaut est 1E-12. L'algorithme détecte une situation proche de la singularité lorsque le déterminant de l'hessien passe en dessous de cette valeur. Utilisez 1E-10 pour comparer vos résultats avec ceux de R/Stata. Consultez les conditions préalables suivantes pour la tolérance aux singularités :
  - Les valeurs plus faibles sont plus prudentes (moins de détections erronées de singularités).
  - Entrez une valeur numérique comprise entre 1E-14 et 1E-6
  - Ce champ accepte les formats de notation scientifique suivants : 1E-12, 1e-12, 0.000000000001
- **Réduction de moitié maximale** :

  Ce champ indique le nombre maximal de tentatives de division par deux de l'incrément par itération. La valeur par défaut est de 5. La méthode de réduction par demi-pas garantit que la vraisemblance logarithmique s'améliore à chaque itération. Augmentez cette valeur à 10-20 si vous rencontrez des échecs de réduction de pas. La valeur doit être un nombre entier compris entre 1 et 20.
- **Pénalité de Ridge** : définit le niveau de régularisation de Ridge pour les matrices hessiennes singulières.

La valeur par défaut est 1E-6. Lorsque le hessien est quasi-singulier (déterminant inférieur au seuil de tolérance de singularité), cette pénalité est appliquée afin de stabiliser l'inversion de la matrice. Utilisez des valeurs plus élevées (1E-4, 1E-3) pour les variables prédictives fortement colinéaires.

Entrez une valeur comprise entre 1E-8 et 1E-3. Ce champ accepte également les formats en notation scientifique : 1E-6, 1e-6, 0.000001

Imprimer

Dans l'onglet **Imprimer**, vous pouvez définir des options pour personnaliser la sortie afin qu'elle ne comprenne que les éléments pertinents que vous sélectionnez.

Sélectionnez l'une des options suivantes pour générer la section de sortie correspondante dans le cadre des résultats de l'analyse.

Récapitulatif du modèle de régression de la médiation
Cette option est sélectionnée par défaut en tant qu'option intégrée et ne peut pas être modifiée. Le résultat affiche un tableau qui résume les résultats du modèle de régression de médiation. Il comprend les coefficients estimés (β0, β1, β2), leurs erreurs-types et les niveaux de signification statistique. Ces résultats permettent de mieux comprendre la relation entre la variable prédictive (*A*), la variable médiatrice (*M*) et les covariables (*C*). Cette option est sélectionnée par défaut.

Récapitulatif du modèle de régression du résultat
Cette option est sélectionnée par défaut en tant qu'option intégrée et ne peut pas être modifiée. Le résultat est un tableau qui résume les résultats du modèle de régression des résultats. Il fournit les coefficients estimés (θ0, θ1, θ2, θ3, θ4), leurs erreurs-types et les niveaux de signification. Ces résultats évaluent les effets du prédicteur, du médiateur et des covariables sur la variable dépendante (*Y*). Cette option est sélectionnée par défaut.

Statistiques de comparaison des niveaux d'exposition
Cette option est sélectionnée par défaut en tant qu'option intégrée et ne peut pas être modifiée. Le résultat génère une section qui présente les différences entre les résultats prévus pour différents niveaux d'exposition de la variable prédictive. Ces statistiques comprennent des estimations telles que l'effet direct contrôlé (*CDE*), l'effet direct naturel (*NDE*) et l'effet indirect naturel (*NIE*), qui permettent de mieux comprendre comment les variations de l'exposition influencent le médiateur et le résultat.

Résultats de l'analyse de décomposition
Sélectionnez cette option pour générer un tableau présentant la décomposition de l'effet total (*ET*) en composantes directes et indirectes, notamment *l'effet commun direct (ECD)*, *l'effet commun indirect* (*ECI*) et l'effet individuel indirect (EII). Il présente également d'autres éléments, tels que la proportion médiée (*PM*) et les effets d'interaction.

Analyse des effets d'interaction
Sélectionnez cette option pour générer une section qui évalue les effets d'interaction tels que *INTmed* et *INTref*, qui mesurent les contributions des interactions médiateur-résultat et prédicteur-médiateur. Ces résultats permettent de mieux comprendre comment ces interactions influencent le résultat.

Statistiques de rapport
Sélectionnez cette option pour générer des statistiques de ratios telles que *r(CDE)*, *r(NDE)*, *r(NIE)* et *r(INTmed)*. Ces chiffres représentent la part de l'effet total (*ET*) attribuable à chaque composante, exprimée en pourcentage.

Matrice de covariance des estimations des paramètres
Cette option est sélectionnée par défaut. Vous pouvez générer un tableau présentant les covariances entre les estimations des paramètres, tels que β (issu de la régression de médiation) et ϑ (issu de la régression sur les résultats). Il est essentiel de calculer les erreurs-types des statistiques dérivées à l'aide de la méthode delta. Cette option est sélectionnée par défaut.

Récapitulatif de traitement des observations
Cette option est sélectionnée par défaut. Vous pouvez générer un résumé qui donne un aperçu des données utilisées dans l'analyse, notamment le nombre de cas, les cas exclus et les pondérations appliquées. Cela permet d'évaluer l'exhaustivité et la qualité de l'ensemble de données. Cette option est sélectionnée par défaut.

Statistiques descriptives
Sélectionnez cette option pour générer une section présentant les statistiques descriptives de toutes les variables utilisées dans l'analyse, telles que les moyennes, les écarts-types, les valeurs minimales et les valeurs maximales. Il donne un aperçu de la répartition et de la variabilité des données.

Tracés

Dans l'onglet « **Graphiques** », vous pouvez définir des options permettant de générer des visualisations qui aident à expliquer les relations entre la variable médiatrice, la variable prédictive et la variable de résultat.

Il est possible de définir quatre types d'options de visualisation : les diagrammes de chemin, les nuages de points, les histogrammes et les diagrammes en forêt. Chaque type de graphique offre une perspective différente sur le modèle de médiation et peut être personnalisé en fonction de vos besoins d'analyse.

Vous pouvez personnaliser chaque graphique en y ajoutant des tailles d'effet, des intervalles de confiance ou des ventilations par sous-groupes.

Diagramme des chemins
Représentez graphiquement les relations entre la variable médiatrice, les variables prédictives et les variables de résultat. Les diagrammes de cheminement sont utiles pour présenter le modèle de médiation dans son ensemble, y compris les effets directs et indirects.

- **Masquer les effets des covariables** : masque toutes les arêtes associées aux covariables. Les chemins prédicteur-médiateur, médiateur-résultat et prédicteur-résultat restent visibles.
- **Couleur des bords** : lorsque cette option est sélectionnée, les bords sont codés par couleur pour indiquer le type d'effet (par exemple, effets directs, indirects et effets de covariables). Le vert indique des effets plus marqués, tandis que le gris indique des effets plus faibles. Si cette option n'est pas sélectionnée, la couleur par défaut des contours est uniforme.

Nuage de points
Les nuages de points permettent de visualiser les relations entre les variables et aident à identifier les corrélations et les schémas de distribution.

Dans l'analyse de médiation, les nuages de points permettent de visualiser les relations entre la variable prédictive (X), la variable médiatrice (M) et la variable dépendante (Y). Ces relations correspondent aux chemins de médiation suivants :

- Chemin a : Effet de la variable prédictive (X) sur la variable médiatrice (M)
- Voie b : Effet du médiateur (M) sur la variable dépendante (Y)
- Chemin c′ : Effet direct de la variable prédictive (X) sur la variable dépendante (Y)

Un nuage de points à trois variables montre comment la variable prédictive influence à la fois le médiateur et le résultat, illustrant ainsi le rôle du médiateur dans la relation entre la variable prédictive et le résultat. La visualisation comprend généralement :

- Données illustrant la relation entre la variable prédictive et la variable médiatrice (chemin a)
- Données illustrant la relation entre le médiateur et le résultat (chemin b)
- La relation globale entre la variable prédictive et la variable dépendante, incluant à la fois les effets directs et indirects (chemin c′)

Les nuages de points vous permettent d'évaluer visuellement la force, la direction et la pertinence de ces relations. Les axes représentent généralement les variables médiatrices et les variables de résultat, tandis que la variable prédictive est représentée à l'aide de dégradés de couleurs ou de marqueurs.

Tracé des barres
Ils permettent d'afficher les proportions ou les fréquences des variables catégorielles et sont utiles pour visualiser les niveaux des variables médiatrices et des variables de résultat. Cette option n'est disponible que lorsque la variante de médiation sélectionnée nécessite un graphique à barres (par exemple, médiateur binaire, résultat binaire ou médiateur continu avec résultat binaire).

Cette option est disponible pour les variantes avec médiateur binaire, résultat binaire ou médiateur continu et résultat binaire.

Cette option est désactivée pour les variantes à variable intermédiaire continue et à variable de résultat continue, pour lesquelles on utilise plutôt des nuages de points.

- **Barres d'erreur** : ajoute des barres d'erreur aux graphiques à barres. Dans les graphiques à barres groupées, les barres d'erreur représentent l'incertitude des proportions (Y = 1 ou Y = 0) pour chaque niveau du médiateur.
- **Intervalles de confiance** : affiche les intervalles de confiance pour les barres d'erreur. Cette valeur est déterminée par celle que vous indiquez dans le champ « **Niveau d'intervalle de confiance %** » de l'onglet « **Critères** ». 95 % est la valeur par défaut pour le niveau de l'intervalle de confiance.

Graphique en forêt
Affichez les coefficients de cheminement et leurs intervalles de confiance, qui vous aident à évaluer l'ampleur et l'incertitude des effets de médiation.

Les graphiques en forêt présentent les ampleurs d'effet des chemins de médiation (par exemple, a, b et c′) ainsi que leurs intervalles de confiance correspondants.

Options :

- **Coefficients de chemin** : affiche les coefficients de chemin (a, b et c′) pour chaque chemin de médiation.
- **Intervalles de confiance** : affiche les intervalles de confiance pour chaque coefficient de chemin. Cette valeur est déterminée par celle que vous indiquez dans le champ « **Niveau d'intervalle de confiance %** » de l'onglet « **Critères** ». 95 % est la valeur par défaut pour le niveau de l'intervalle de confiance.

Filtres de séries temporelles

Dans l'analyse des séries temporelles, il est essentiel d'éliminer les valeurs qui sortent d'une fourchette décidée et qui provoquent des fluctuations inhabituelles dans la série temporelle. Un bon filtre doit être capable d'éliminer les racines unitaires et les composantes cycliques, ou le filtre doit être capable d'isoler les fluctuations des données à une certaine fréquence.

Le filtrage des séries temporelles est une technique essentielle en économétrie et en analyse des séries temporelles, souvent utilisée pour décomposer une série temporelle en composantes tendancielles et cycliques. Ceci est particulièrement important en macroéconomie et en finance, où il est essentiel d'identifier les fluctuations du cycle économique ou les tendances économiques à long terme.

Les filtres de séries temporelles les plus utilisés sont le filtre de Hodrick-Prescott (HP), le filtre de Baxter-King (BK) et le filtre de Christiano-Fitzgerald (CF).

Obtention de filtres de séries temporelles

Les filtres de séries temporelles sont disponibles en tant que procédure d'extension avec le plug-in Python. Il faut donc installer le module d'extension pour exécuter la procédure.

effectuez les étapes suivantes pour exécuter la procédure relative aux filtres de séries temporelles.

1. Allez dans **DataDefine** > **date and time** pour générer les variables de date.
2. Après avoir défini la date et l'heure, cliquez sur **Analyze** > **ForcastingTime** > **Series Filters (TSF).**
3. Sélectionnez le filtre de série temporelle que vous souhaitez appliquer à l'ensemble de données.

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
4. Sélectionnez les variables et déplacez-les dans la liste des variables du filtre de série temporelle approprié.
5. Définissez la valeur de **Lambda** pour le filtre de Hodrick-Prescott (HP) ou définissez les valeurs de **Low**, **High** et **K** pour le filtre de Baxter-King (BK) et le filtre de Christiano-Fitzgerald (CF).

En option, vous pouvez sélectionner **Dérive** pour le filtre Christiano-Fitzgerald (CF).

6. Cliquez sur **Exécuter**.

Filtres de séries temporelles : Options

La boîte de dialogue Filtres de séries temporelles contient les options suivantes.

Important : Avant de définir les options, allez dans **DonnéesDéfinir la** > **date et l'heure** pour générer les variables de date.

Hodrick-Prescott (HP)

Le filtre de **Hodrick-Prescott (HP)** est un outil utilisé en macroéconomie pour ajuster une tendance lisse aux séries temporelles.

Le filtre de Hodrick-Prescott sépare une série temporelle $_{yt}$ en une composante tendancielle $_{\tau t}$ et une composante cyclique $_{ct}$.

$y_t = \tau_t + c_t + \square_t$où $\square_{est}$ la composante d'erreur (ou de bruit).

Les composants sont déterminés en minimisant la fonction de perte quadratique et le filtre HP est explicitement donné par la formule,

$HP = [\ ^{\lambda L2-}\ 4\lambda L + (1 + 6\lambda) -\ ^{4\lambda L-1} + \ ^{\lambda L-2}\ ]^{-1}$

où λ représente le paramètre de lissage (valeur positive) et L représente l'opérateur de décalage, qui est dérivé de la condition de premier ordre pour le problème de minimisation.

Choisissez une variable pour le filtre HP dans la liste des variables et définissez une valeur pour **Lambda**.

Le paramètre de lissage *λ* détermine l'équilibre entre l'ajustement et le lissage. Plus la valeur de *λ* est importante, plus la pénalité est élevée. Lorsque vous analysez des données trimestrielles, la valeur par défaut de 1600 est recommandée pour *λ*. De même, *λ* devrait être $^{1600/44}$ = 6.25 pour les données annuelles, et 1600 ∗ ^{34} = 129 600 pour les données mensuelles. Lorsque *λ*= 0, cela signifie que la sortie de la série de filtres HP correspond exactement à la série d'entrée. Au contraire, des valeurs plus élevées de *λ* donnent des séries de résultats plus lisses.

Baxter-King (BK)

Le filtre de **Baxter-King (BK**) est un filtre passe-bande couramment utilisé dans l'analyse des séries temporelles pour isoler les composantes cycliques des données, souvent dans le cadre d'études macroéconomiques. Son objectif est d'éliminer le bruit à haute fréquence (court terme) et les composantes à basse fréquence (tendance à long terme), en se concentrant sur les cycles à moyen terme qui nous intéressent, généralement entre 1.5 et 8 ans. Le filtre BK suppose la stationnarité de la série temporelle sous-jacente. Dans la liste des variables, choisissez 1 ou 2 variables pour le filtre BK.

Réglez les valeurs appropriées dans **Low** (période minimale), **High** (période maximale) et **K** (longueur du décalage).

Christiano-Fitzgerald (CF)

Le filtre de **Christiano-Fitzgerald (CF)** est un autre filtre passe-bande utilisé pour séparer une série temporelle en composantes tendancielles et cycliques. Ce filtre est similaire au filtre de Baxter King (BK), qui fonctionne également sur la base d'un filtre passe-bande idéal. Le filtre BK suppose une moyenne mobile symétrique pour approximer le filtre passe-bande, mais le filtre CF suppose que la série temporelle suit un processus aléatoire sans dérive. Le filtrage CF est mis en œuvre en estimant un filtre optimal dans l'hypothèse où les données sont générées par une marche aléatoire ou un processus de type ARIMA.

Dans la liste des variables, choisissez 1 ou 2 variables pour le filtre CF. Réglez les valeurs appropriées pour **Low**, **High** et **K.** Sélectionnez **Drift** si nécessaire pour le filtre CF.

Lambda, **Low** et **High** peuvent être des valeurs flottantes quelconques. La valeur de **K** doit être un nombre entier. Voir les valeurs suivantes :

Faible	1.5 (annuel) et 6 (trimestriel)
Elevé	8 (annuel) et 32 (trimestriel)
k	3 (annuel) et 12 (trimestriel)

Estimation de courbe

La procédure Estimation de courbe produit des statistiques de régression d'estimation de courbe et les tracés relatifs pour 11 modèles différents de régression d'estimation de courbe. Un modèle différent est produit pour chaque variable dépendante. Vous pouvez aussi enregistrer les prévisions, les résidus et les intervalles de la prévision comme nouvelles variables.

Exemple. Un fournisseur de services Internet suit le pourcentage dans le temps du trafic de messages électroniques infectés par un virus sur ses réseaux. Un nuage de points révèle que la relation n'est pas linéaire. Vous pouvez ajuster un modèle quadratique ou cubique en fonction des données, vérifier la validité des hypothèses et la qualité d'ajustement du modèle.

Statistiques : Pour chaque modèle : coefficients de régression, *R* multiple, R^2, R^2 ajusté, erreur standard de l'estimation, tableau d'analyse de variance, prévisions, résidus et intervalles de la prévision. Modèles : linéaire, logarithmique, inverse, quadratique, cubique, de puissance, composé, courbe en S, logistique, de croissance et exponentiel.

Remarques sur les données de l'estimation de courbe

Données. Les variables dépendantes et indépendantes doivent être quantitatives. Si vous sélectionnez **Temps** à partir du jeu de données actif comme variable indépendante (au lieu de sélectionner une variable), la procédure Estimation de courbe génère une variable de temps où la durée entre les observations est uniforme. Si **Temps** est sélectionné, la variable dépendante doit être une mesure de séries temporelles. L'analyse des séries temporelles nécessite une structure de fichier de données dans lequel chaque observation (ligne) représente un ensemble d'observations à des moments différents et où la durée entre les observations est uniforme.

Hypothèses : Vérifiez vos données graphiquement pour déterminer comment sont reliées les variables indépendantes et dépendantes (de manière linéaire ou exponentielle, etc.). Les résidus d'un bon modèle doivent être répartis aléatoirement et doivent être normaux. Si un modèle linéaire est utilisé, les hypothèses suivantes doivent être satisfaites : pour chaque valeur de la variable indépendante, la distribution de la variable dépendante doit être normale. La variance de la distribution de la variable dépendante doit être constante pour toutes les valeurs de la variable indépendante. La relation entre la variable dépendante et la variable indépendante doit être linéaire, et toutes les observations doivent être indépendantes.

Pour obtenir une estimation de courbe

1. A partir des menus, sélectionnez :

   Analyse > **Régression** > **Estimation de courbe...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins une variable dépendante. Un modèle différent est produit pour chaque variable dépendante.
3. Sélectionnez une variable indépendante (une variable dans le jeu de données actif ou dans le jeu de données actif ou **Temps**).
4. Eventuellement :

- Sélectionner une variable pour libeller des observations dans les nuages de points. Pour chaque point du nuage de points, utilisez l'outil de sélection de points pour afficher la valeur de la variable avec Libellé d'observation :
- Cliquez sur **Enregistrer** pour enregistrer les prévisions, les résidus et les intervalles de prévision comme nouvelles variables.

Les options suivantes sont également disponibles :

- **Inclure terme constant dans l'équation :** Evalue un terme constant dans l'équation de régression. La constante est incluse par défaut.
- **Tracer sous forme graphique :** Trace graphiquement les valeurs de la variable dépendante et chaque modèle sélectionné face à la variable indépendante. Un graphique séparé est produit pour chaque variable dépendante.
- **Afficher le tableau ANOVA :** Affiche une table récapitulative de l'analyse de variance pour chaque modèle sélectionné.

Modèles d'estimation de courbe

Vous pouvez choisir des modèles de régression d'estimation de courbe. Pour déterminer quel modèle utiliser, tracez vos données sous forme graphique. Si vos variables semblent être liées linéairement, utilisez un modèle de régression linéaire simple. Lorsque vos variables ne sont pas liées linéairement, essayez de transformer vos données. Lorsque la transformation n'améliore pas les choses, vous devrez peut-être utiliser un modèle plus élaboré. Observez un nuage de points de vos données. Si le tracé ressemble à une fonction mathématique que vous reconnaissez, ajustez vos données en fonction de ce type de modèle. Par exemple, si vos données ressemblent à une fonction exponentielle, utilisez un modèle exponentiel.

Linéaire. Modèle dont l'équation est Y = b0 + (b1 * t). Les valeurs de la série sont modélisées comme fonction linéaire du temps.

Logarithmique. Modèle dont l'équation est Y = b0 + (b1 / t).

Inverse. Modèle dont l'équation est Y = b0 + (b1 / t).

Quadratique. Modèle dont l'équation est Y = b0 + (b1 * t) + (b2 * t**2). Le modèle quadratique peut être utilisé pour modéliser une série qui "décolle" ou qui s'amortit.

Cubique. Modèle défini par l'équation Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3).

Puissance. Modèle dont l'équation est Y = b0 * (t**b1) ou ln(Y) = ln(b0) + (b1 * ln(t)).

Composé. Modèle dont l'équation est la suivante : Y = b0 * (b1**t) ou ln(Y) = ln(b0) + (ln(b1) * t).

Courbe S. Modèle dont l'équation est Y = e**(b0 + (b1/t)) ou ln(Y) = b0 + (b1/t).

Logistique. Modèle dont l'équation est Y = 1 / (1/u + (b0 * (b1**t))) ou ln (1/y-1/u) = ln (b0) + (ln (b1) * t) où u est la valeur limite supérieure. Après avoir sélectionné la logistique, précisez la valeur limite supérieure à utiliser dans l'équation de régression. La valeur doit être un nombre positif supérieur à la plus grande valeur de la variable dépendante.

Croissance. Modèle dont l'équation est Y = e**(b0 + (b1 * t)) ou ln(Y) = b0 + (b1 * t).

Exponentiel. Modèle dont l'équation est Y = b0 * (e**(b1 * t)) ou ln(Y) = ln(b0) + (b1 * t).

Enregistrement de l'estimation de courbe

Enregistrer les variables : Pour chaque modèle sélectionné, vous pouvez enregistrer les prévisions, les résidus (valeur observée de la variable dépendante moins la prévision du modèle) et les intervalles de prévision (limites supérieure et inférieure). Les nouveaux noms de variable et les libellés descriptifs s'affichent dans un tableau dans la fenêtre de sortie.

Calculer une prévision : Si, dans le jeu de données actif, vous sélectionnez **Temps** à la place d'une variable comme variable indépendante, vous pouvez spécifier une période de prévision au-delà de la fin de la série temporelle. Vous avez le choix entre les options suivantes :

- **A partir d'une estimation limitée à une période :** Prévoit les valeurs pour toutes les observations du fichier, à partir des observations de la période d'estimation. La période d'estimation qui s'affiche en bas de la boîte de dialogue est définie avec la boîte de sous dialogue Plage de l'option Sélectionner des observations du menu Données. Si aucune période d'estimation n'a été définie, toutes les observations sont utilisées pour prévoir les valeurs.

- **Jusqu'à :** Prévoit les valeurs jusqu'à la date, l'heure ou le numéro de l'observation spécifié, à partir des observations de la période d'estimation. Cette fonction peut être utilisée pour prévoir les valeurs au-delà de la dernière observation de la série temporelle. Les variables courantes de date définies déterminent les zones de texte disponibles pour la spécification de la fin de la période de prévision. Si aucune variable de date n'est définie, vous pouvez spécifier le numéro de l'observation finale.

Utilisez l'option Définir des dates... dans le menu Données pour créer des variables de date.

Régression aux moindres carrés partiels

La procédure Régression aux moindres carrés partiels estime les modèles de régression aux moindres carrés partiels (également connus sous le nom de "projection to latent structure", PLS). La technique de prévision PLS constitue une solution de remplacement par rapport à la régression par les moindres carrés classiques, à la corrélation canonique ou à la modélisation d'équation structurelle, particulièrement utile lorsque les variables de prédicteur présentent une forte corrélation ou lorsque le nombre de prédicteurs dépasse le nombre d'observations.

PLS combine des fonctions d'analyse des composants principaux et la régression multiple. Un ensemble de facteurs latents expliquant autant que possible la covariance entre les variables indépendantes et dépendantes est extrait. Ensuite, une étape de régression prévoit les valeurs des variables dépendantes à l'aide de la décomposition des variables indépendantes.

Tableaux
Proportion de variance expliquée (par facteur latent), pondérations de facteurs latents, chargements de facteurs latents, importance de la variable indépendante dans la projection (VIP), et estimations des paramètres de régression (par variable dépendante) sont tous générés par défaut.

Graphiques
Importance de la variable dans la projection (VIP), scores factoriels, pondération des trois premiers facteurs latents et distance au modèle sont tous générés depuis l'onglet Options.

Remarques sur les données

Niveau de mesure
Les variables dépendantes et indépendantes (prédicteur) peuvent être échelle, nominal ou ordinal. La procédure considère que le niveau de mesure approprié a été assigné à toutes les variables, bien que vous puissiez changer provisoirement le niveau de mesure d'une variable en cliquant avec le bouton droit de la souris sur la variable dans la liste des variables source, puis en sélectionnant un niveau de mesure dans le menu contextuel. Les variables catégorielles (nominales ou ordinales) sont traitées de manière équivalente par la procédure.

Codage de variable catégorielle
La procédure recode provisoirement les variables dépendantes catégorielles via la codification un-de-*c* pendant la durée de la procédure. S'il existe *c* catégories d'une variable, la variable est stockée en tant que vecteurs *c* , avec la première catégorie notée (1,0, ..., 0), la catégorie suivante (0,1,0, ..., 0), ..., et la catégorie finale (0,0, ..., 0, 1). Les variables dépendantes catégorielles sont représentées à l'aide de la codification de façon fictive, c'est-à-dire, elles omettent simplement l'indicateur correspondant à la catégorie de référence.

Pondérations de fréquence
Les valeurs de pondération sont arrondies au nombre entier le plus près avant utilisation. Les observations avec des pondérations manquantes ou des pondérations inférieures à 0,5, ne sont pas utilisées dans les analyses.

Valeur manquante
Les valeurs manquantes spécifiées par l'utilisateur et par le système sont traitées comme non valides.

Rééchelonnement
Tous les variables de modèle sont centrées et standardisées, dont les variables indicateur représentant les variables catégorielles.

Obtention de la régression des moindres carrés partiels

A partir des menus, sélectionnez :

Analyze > **Régression** > **Moindres carrés partiels...**

1. Sélectionnez au moins une variable dépendante.
2. Sélectionnez au moins une variable indépendante.

(En option) Vous avez également ces possibilités :

- Indiquer une catégorie de référence pour les variables catégorielles dépendantes (nominale ou ordinale).
- Indiquer une variable à utiliser comme identificateur unique pour les jeux de données enregistrés et les sorties par observation.
- Indiquer une limite supérieure sur le nombre de facteurs latents à extraire.

Modèle

Spécifier les effets du modèle : Un modèle comportant des effets principaux contient tous les effets principaux de covariable et de facteur. Sélectionnez **Personnalisé** pour préciser les interactions. Vous devez indiquer tous les termes à inclure dans le modèle.

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

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

Termes construits

Pour les facteurs et covariables sélectionnés :

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

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

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

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

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

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

Options

L'onglet Options permet d'enregistrer et de tracer des estimations de modèles pour des observations individuelles, des facteurs latents, et des prédicteurs.

Pour chaque type de données, indiquez le nom d'un jeu de données. Les noms de jeu de données doivent être uniques. Si vous spécifiez le nom d'un jeu de données existant, son contenu est remplacé ; sinon, un jeu de données est créé.

- **Enregistrer les estimations concernant les observations individuelles :** Enregistre les estimations de modèle par observation suivantes : prévisions, résidus, distance au modèle de facteur latent et scores factoriels latents. Elle permet également de tracer les scores factoriels latents.
- **Enregistrer les estimations concernant les facteurs latents :** Enregistre les chargements de facteurs latents et les pondérations de facteurs latents. Elle permet également de tracer les pondérations de facteurs latents.
- **Enregistrer les estimations concernant les variables indépendantes :** Enregistre les estimations des paramètres de régression et l'importance des variables dans la projection (VIP). Elle permet également de tracer l'importance des variables dans la projection par facteur latent.

Analyse du voisin le plus proche

L'analyse du voisin le plus proche est une méthode de classification d'observations en fonction de leur similarité avec les autres observations. En apprentissage automatique, elle a été développée comme une façon de reconnaître les configurations de données sans avoir à recourir à une correspondance exacte avec d'autres configurations ou observations stockées. Les observations semblables sont proches l'une de l'autre et les observations dissemblables sont éloignées l'une de l'autre. Par conséquent, la distance entre deux observations est une mesure de leur dissemblance.

Les observations proches l'une de l'autre sont "voisines". Lorsqu'une observation (restante) est présentée, sa distance par rapport à chacune des observations du modèle est calculée. Les classifications des observations les plus semblables - les voisins les plus proches - sont comptées et la nouvelle observation est placée dans la catégorie qui contient le plus grand nombre de voisins les plus proches.

Vous pouvez indiquer le nombre de voisins les plus proches à examiner ; cette valeur s'appelle *K*.

L'analyse du voisin le plus proche peut également être utilisée pour calculer des valeurs pour une cible continue. Dans cette situation, la valeur cible de la médiane ou de la moyenne des voisins les plus proches est utilisée pour obtenir la valeur prédite de la nouvelle observation.

Remarques sur l'analyse du voisin le plus proche

Cible et fonctions : La cible et les fonctions peuvent être :

- *Nominal*. Une variable peut être traitée comme nominale lorsque ses valeurs représentent des catégories sans classement intrinsèque (par exemple, le service de la société dans laquelle travaille un employé). La région, le code postal ou l'appartenance religieuse sont des exemples de variables nominales.
- *Ordinaux*. Une variable peut être traitée comme ordinale lorsque ses valeurs représentent des catégories avec un classement intrinsèque (par exemple, des niveaux de satisfaction allant de très insatisfait à très satisfait). Exemples de variable ordinale : des scores d'attitude représentant le degré de satisfaction ou de confiance, et des scores de classement des préférences.
- *Echelle*. Une variable peut être traitée comme une variable d'échelle (continue) lorsque ses valeurs représentent des catégories ordonnées avec une mesure significative, de sorte que les comparaisons de distance entre les valeurs soient appropriées. L'âge en années et le revenu en milliers de dollars sont des exemples de variable d'échelle.

  Les variables catégorielles et ordinales sont traitées de manière équivalente par l'analyse du voisin le plus proche. La procédure considère que le niveau de mesure approprié a été assigné à chaque variable, bien que vous puissiez changer provisoirement le niveau de mesure d'une variable en cliquant avec le bouton droit de la souris sur la variable dans la liste des variables sources, puis en sélectionnant un niveau de mesure dans le menu contextuel.

Dans la liste des variables, une icône indique le niveau de mesure et le type de données :

Tableau 3. Icônes de niveau de mesure

	Numérique	**Chaîne**	**Date :**	**Durée**
Echelle (continue).		n/a		
Ordinal				
Nominal				

Codification des variables indicatrices : La procédure recode provisoirement les variables de prédicteur catégorielles et les variables dépendantes via la codification un-de-*c* pour la durée de la procédure. S'il

existe *c* catégories d'une variable, celle-ci est stockée sous forme de c vecteurs, la première catégorie étant notée (1,0,...,0), la suivante (0,1,0,...,0),..., et la catégorie finale (0,0,...,0,1).

Ce schéma de codification augmente la dimensionnalité de l'espace des fonctions. Plus particulièrement, le nombre total de dimensions correspond au nombre de prédicteurs d'échelle plus le nombre de catégories sur l'ensemble des prédicteurs catégoriels. En conséquence, ce système de codification peut provoquer un ralentissement de l'entraînement. Si l'entraînement des voisins les plus proches s'effectue très lentement, vous pouvez essayer de réduire le nombre de catégories dans vos prédicteurs catégoriels en combinant des catégories similaires ou en supprimant les observations comportant des catégories extrêmement rares avant de lancer la procédure.

Toute codification un-de-*c* repose sur les données d'apprentissage, même si un échantillon restant est défini (voir «Partitions», à la page 214). Ainsi, si l'échantillon restant contient des observations avec des catégories de prédicteurs absentes des données d'apprentissage, ces observations ne seront pas évaluées. Si l'échantillon restant contient des observations avec des catégories de variable dépendantes absentes des données d'apprentissage, ces observations seront évaluées.

Rééchelonnement : Les fonctions d'échelle sont normalisées par défaut. Le rééchelonnement repose entièrement sur les données d'apprentissage, même si un échantillon restant est défini (voir «Partitions», à la page 214). Si vous spécifiez une variable pour définir des partitions, il est important que ces fonctions présentent des distributions similaires à travers les échantillons d'apprentissage et les échantillons restants. Par exemple, utilisez la procédure Explorer pour examiner les distributions à travers les partitions.

Pondérations de fréquence : Cette procédure ignore les pondérations de fréquence.

Réplication de résultats : La procédure utilise la génération de nombres aléatoires pendant l'affectation aléatoire des partitions et les niveaux de validation croisée. Si vous souhaitez répliquer vos résultats exactement, en plus d'utiliser les mêmes paramètres de procédure, définissez une valeur de départ pour le Mersenne Twister (voir «Partitions», à la page 214) ou utilisez des variables pour définir les partitions et les niveaux de validation croisée.

Obtention d'une analyse du voisin le plus proche

A partir des menus, sélectionnez :

Analyse > **Classifier** > **Voisin le plus proche...**

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

1. Spécifiez une ou plusieurs fonctions, qui peuvent être considérées comme des prédicteurs si une cible existe.

   Cible (facultative) : Si aucune cible (variable dépendante ou réponse) n'est spécifiée, la procédure cherche uniquement les voisins les plus proches de *k* : aucune classification ni prévision n'est effectuée.

   Normaliser les fonctions d'échelle : Les fonctions normalisées possèdent la même plage de valeurs, ce qui permet d'améliorer les performances de l'algorithme d'estimation. La normalisation ajustée [2*(x–min)/(max–min)]–1, est utilisée. Les valeurs normalisées ajustées sont comprises entre –1 et 1.

   Identificateur d'observations focales (facultatif) : Cela vous permet de marquer les observations présentant un intérêt particulier. Par exemple, un chercheur veut déterminer si les résultats scolaires d'un certain quartier (l'observation focale) sont comparables à ceux de quartiers similaires. Il utilise l'analyse du voisin le plus proche pour connaître les districts scolaires les plus identiques selon un ensemble de fonctions donné. Il compare ensuite les scores de l'examen du district focal à ceux des voisins les plus proches.

   Les observations focales pourraient être également appliquées à des études cliniques pour sélectionner les observations de contrôle similaires aux observations cliniques. Les observations focales sont affichées dans le tableau des *k* voisins les plus proches et des distances, sur le graphique de l'espace des fonctions, dans le graphique des homologues et sur la carte des quadrants. Les informations sur les observations locales sont enregistrées dans les fichiers spécifiés sur l'onglet Sortie.

Les observations à valeur positive sur la variable spécifiée sont traitées comme des observations focales. Spécifier une variable sans valeur positive n'est pas valide.

Libellé d'observation (facultatif) : Les observations sont libellées à l'aide de ces valeurs sur le graphique de l'espace des fonctions, dans le graphique des homologues et sur la carte des quadrants.

Champs avec un niveau de mesure inconnu

L'alerte du niveau de mesure apparaît lorsque le niveau de mesure d'une ou plusieurs variables (champs) du jeu de données est inconnu. Le niveau de mesure ayant une incidence sur le calcul des résultats de cette procédure, toutes les variables doivent avoir un niveau de mesure défini.

Analyser les données : Lit les données dans le jeu de données actifs et attribue le niveau de mesure par défaut à tous les champs ayant un niveau de mesure inconnu. Si le jeu de données est important, cette action peut prendre un certain temps.

Affecter manuellement : Ouvre une boîte de dialogue qui répertorie tous les champs ayant un niveau de mesure inconnu. Vous pouvez utiliser cette boîte de dialogue pour attribuer un niveau de mesure à ces champs. Vous pouvez également attribuer un niveau de mesure dans la vue de variable de l'éditeur de données.

Le niveau de mesure étant important pour cette procédure, vous ne pouvez pas accéder à la boîte de dialogue d'exécution de cette procédure avant que tous les champs n'aient des niveaux de mesure définis.

Voisins

Nombre de voisins les plus proches (k) : Spécifiez le nombre de voisins les plus proches. Remarque : l'utilisation d'un nombre élevé de voisins ne garantit pas forcément un modèle plus précis.

Si une cible est spécifiée sur l'onglet Variables, vous pouvez également indiquer une plage de valeurs et permettre à la procédure de choisir le nombre "optimal" de voisins au sein de cette plage. La méthode pour déterminer le nombre de voisins les plus proches dépend si la sélection des fonctions est requise par l'onglet Fonctions ou non.

- Si oui, la sélection des fonctions sera alors exécutée pour chaque valeur de *k* dans la plage requise, et le *k* ainsi que l'ensemble des fonctions l'accompagnant, avec le taux d'erreur le plus faible (ou l'erreur de la somme des carrés la plus faible si la cible est une échelle), seront sélectionnés.
- Si la sélection des fonctions n'est pas activée, la validation croisée de niveau *V* est utilisée pour sélectionner le nombre "optimal" de voisins. Reportez-vous à l'onglet Partitions pour contrôler l'affectation de niveaux.

Calcul de la distance : Il s'agit de la métrique employée pour spécifier la distance métrique utilisée dans la mesure de la similarité des observations.

- **Métrique euclidienne :** La distance entre deux observations, x et y, est la racine carrée de la somme, sur toutes les dimensions, des carrés des différences entre les valeurs de ces observations.
- **Mesure de la distance de Manhattan :** La distance entre deux observations est la somme, sur toutes les dimensions, des différences absolues entre les valeurs de ces observations. Appelée également distance City Block.

Si une cible est spécifiée dans l'onglet Variables, vous pouvez également choisir de pondérer les fonctions selon leur importance normalisée lors du calcul des distances. L'importance des fonctions pour un prédicteur est calculée par le rapport du taux d'erreur ou l'erreur de la somme des carrés du modèle avec le prédicteur supprimé du modèle vers le taux d'erreur ou l'erreur de la somme des carrés pour le modèle entier. L'importance normalisée est calculée par nouvelle pondération des valeurs d'importance des fonctions de sorte que leur somme soit égale à 1.

Prévisions pour cible d'échelle : Lorsqu'une cible d'échelle est spécifiée sur l'onglet Variables, elle détermine si la valeur prévue est calculée à partir de la valeur moyenne ou la médiane des voisins les plus proches ou non.

Fonctions

L'onglet Fonctions vous permet de demander et de spécifier des options pour la sélection des fonctions lorsqu'une cible est spécifiée dans l'onglet Variables. Par défaut, toutes les fonctions sont prises en compte pour la sélection de fonctions, mais vous pouvez également sélectionner un sous-ensemble de fonctions à introduire de force dans le modèle.

Critère d'arrêt : A chaque étape, la fonction dont l'addition au modèle entraîne l'erreur la plus faible (calculée comme le taux d'erreur pour une cible catégorielle et l'erreur de la somme des carrés pour une cible d'échelle) est prise en compte afin d'être incluse dans l'ensemble de modèle. La sélection ascendante se poursuit jusqu'à la rencontre de la condition spécifiée.

- **Nombre de fonctions spécifié :** L'algorithme ajoute un nombre fixe de fonctions en plus de celles introduites de force dans le modèle. Spécifiez un nombre entier positif. La diminution des valeurs du nombre à sélectionner produit un modèle plus réduit, au risque d'un manque de fonctions importantes. L'augmentation des valeurs du nombre à sélectionner capturera toutes les fonctions importantes, au risque d'ajouter des fonctions qui en réalité alimentent l'erreur du modèle.
- **Changement minimal dans le rapport d'erreur absolue :** L'algorithme prend fin lorsque le changement dans le rapport d'erreur absolue indique que le modèle ne peut pas être davantage amélioré par l'ajout de nouvelles fonctions. Indiquez un nombre positif. La diminution des valeurs pour le changement minimal aura tendance à inclure davantage de fonctions, au risque d'en inclure certaines qui n'apportent pas beaucoup de valeur au modèle. L'augmentation de la valeur du changement minimal aura tendance à exclure davantage de fonctions, au risque de perdre des fonctions importantes pour le modèle. La valeur "optimale" du changement minimal dépendra de vos données et de l'application. Reportez-vous au Journal d'erreur de sélection des fonctions pour pouvoir déterminer quelles sont les fonctions les plus importantes. Pour plus d'informations«Journal des erreurs de sélection des fonctions », à la page 218, voir la rubrique .

Partitions

L'onglet Partitions vous permet de diviser le jeu de données en un ensemble d'apprentissage et un ensemble traité, et lorsque cela s'applique, il vous permet d'affecter des observations aux niveaux de validation croisée.

Partition d'apprentissage et partition traitée : Ce groupe indique la méthode de partitionnement du jeu de données actif en échantillons d'apprentissage et restants. L'**échantillon d'apprentissage** comprend les enregistrements de données utilisés pour former le modèle Voisin le plus proche. Un certain pourcentage d'observations contenues dans le jeu de données doit être affecté à l'échantillon d'apprentissage pour l'obtention d'un modèle. L'**échantillon restant** est un ensemble indépendant d'enregistrements de données utilisé pour évaluer le modèle final ; l'erreur pour l'échantillon restant donne une estimation "honnête" de la capacité de prévision du modèle parce que les observations restantes n'ont pas été utilisées pour construire le modèle.

- **Affecter aléatoirement des observations aux partitions :** Spécifiez le pourcentage d'observations à affecter à l'échantillon d'apprentissage. Le reste est affecté à l'échantillon restant.
- **Utiliser une variable pour affecter des observations :** Indiquez une variable numérique qui affecte chaque observation du jeu de données actif à l'échantillon d'apprentissage et restant. Les observations contenant une valeur positive sur la variable sont affectées à l'échantillon d'apprentissage, celles contenant une valeur égale à 0 ou une valeur négative sont affectées à l'échantillon restant. Les observations contenant des valeurs système manquantes sont exclues de l'analyse. Les valeurs manquantes de l'utilisateur pour la variable de partitionnement sont toujours considérées comme étant valides.

Niveaux de validation croisée : Le Niveau *V* de validation croisée est utilisé pour déterminer le "meilleur" nombre de voisins. Il n'est pas disponible en association avec la sélection des fonctions pour des raisons de performance.

La validation croisée divise l'échantillon en plusieurs sous échantillons, ou niveaux. Les modèles du voisin le plus proche 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. L'erreur est estimée pour chaque modèle en appliquant le modèle au sous-échantillon exclu lors de la génération du modèle. Le "meilleur" nombre des voisins les plus proches est celui qui produit l'erreur la plus faible sur les sous-échantillons.

- **Affecter aléatoirement des observations aux niveaux :** Spécifiez le nombre de niveaux à utiliser pour la validation croisée. Cette procédure affecte aléatoirement des observations aux sous-échantillons, numérotés de 1 à *V*, le nombre de sous-échantillons.
- **Utiliser une variable pour affecter des observations :** Indiquez une variable numérique qui affecte chaque observation du jeu de données actif à un niveau. La variable doit être numérique et prendre les valeurs de 1 à *V*. Si des valeurs de cette plage sont manquantes et sur toutes les divisions si des fichiers de fractionnement sont en vigueur, une erreur s'affiche.

Définir la valeur de départ pour Mersenne Twister : Définir une valeur de départ vous permet de reproduire les analyses. L'utilisation de ce contrôle revient à définir le Mersenne Twister comme le générateur actif et à spécifier un point de départ fixe dans la boîte de dialogue Générateurs de nombres aléatoires. La différence notoire est que la définition de la valeur de départ dans cette boîte de dialogue conserve l'état actuel du générateur de nombres aléatoires et restaure cet état une fois l'analyse terminée.

Enregistrement

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

Variables à enregistrer

- **Valeur ou catégorie prévue :** Cette option enregistre la valeur prévue pour une cible d'échelle ou la catégorie prévue pour une cible catégorielle.
- **Probabilité prédite :** Enregistre les probabilités prévues pour une cible catégorielle. Une variable distincte est enregistrée pour chacune des *n* premières catégories, *n* étant spécifié dans le contrôle **Catégories maximales à enregistrer pour la cible catégorielle**.
- **Variable de partition de formation/traitement :** Si des observations sont affectées aléatoirement aux échantillons d'apprentissage et aux échantillons restants dans l'onglet Partitions, cela enregistre la valeur de la partition (d'apprentissage ou traitée) à laquelle l'observation a été affectée.
- **Variable du niveau de validation croisée :** Si des observations ont été affectées aléatoirement à des niveaux de validation croisée dans l'onglet Partitions, cela enregistre la valeur du niveau auquel l'observation a été affectée.

Sortie

Sortie du visualiseur

- **Récapitulatif du traitement des observations :** Affiche la table récapitulative de traitement des observations, qui récapitule le nombre d'observations incluses et exclues dans l'analyse, au total et par échantillon de formation et par échantillon restant.
- **Graphiques et tableaux :** Affiche les sorties liées au modèle, y compris les tableaux et les graphiques. Les tables du modèle incluent les *k* voisins les plus proches et les distances pour observations focales, classement des variables de réponse catégorielle, ainsi qu'un récapitulatif d'erreur. Les sorties graphiques dans l'affichage du modèle incluent un journal d'erreur de sélection, un graphique d'importance des fonctions, un graphique d'espace des fonctions, un graphique des homologues et une carte des quadrants. Pour plus d'informations, voir «Vue du modèle », à la page 216.

Fichiers

- **Exporter le modèle dans un fichier XML :** Vous pouvez utiliser ce fichier de modèle pour appliquer les informations du modèle aux autres fichiers de données à des fins d'évaluation. Cette option n'est pas disponible si des fichiers scindés ont été définis.

- **Exporter les distances entre les observations focales et les voisins les plus proches de k :** Pour chaque observation focale, une variable distincte est créée pour chacun des *k* voisins les plus proches des observations focales (à partir de l'échantillon d'apprentissage et les *k* distances les plus proches correspondantes.

Options

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

Les valeurs système manquantes et les valeurs manquantes pour les variables d'échelle sont toujours considérées comme non valides.

Vue du modèle

Lorsque vous sélectionnez **Graphiques et tableaux** dans l'onglet Sortie, la procédure produit un objet de Voisin le plus proche dans le visualiseur. En activant cet objet par un double-clic, vous obtenez une vue interactive du modèle. Le modèle présente une fenêtre à double panels :

- Le premier affiche une présentation du modèle, appelée vue principale.
- Le second affiche un des deux types de vues :

  Une vue de modèle auxiliaire affiche davantage d'informations sur le modèle, mais n'est pas focalisée sur le modèle lui-même.

  Un vue liée est un affichage montrant les détails d'une fonction du modèle lorsque l'utilisateur fait défiler une partie de la vue principale.

Par défaut, le premier panel affiche l'espace des fonctions et le second le graphique d'importance de variable. Si ce dernier n'est pas disponible, c'est-à-dire lorsque **Pondérer les fonctions par importance** n'a pas été sélectionné dans l'onglet Fonctions, la première vue disponible dans la la vue déroulante est affichée.

Lorsqu'une vue n'a aucune information disponible, son élément texte dans la vue déroulante est désactivé.

Espace des fonctions

Le graphique d'espace des fonctions est un graphique interactif de l'espace des fonctions (ou un sous-espace, s'il existe plus de 3 fonctions). Chaque axe représente une fonction du modèle, et l'emplacement des points dans le graphique montre la valeur de ces fonctions pour des observations dans les partitions Apprentissage et Restant.

Clés. En plus des valeurs de fonctions, les points du tracé contiennent d'autres informations.

- La forme indique la partition à laquelle appartient un point : Apprentissage ou Restant.
- La couleur/l'ombrage d'un point indique la valeur de la cible pour cette observation, avec les valeurs de couleur distinctes correspondant aux catégories d'une cible catégorielle, et les ombres indiquant la plage des valeurs d'une cible continue. La valeur indiquée pour la partition Apprentissage est la valeur observée. Pour la partition Restant, il s'agit de la valeur prévue. Si aucune cible n'est spécifiée, la clé ne s'affiche pas.
- Les contours plus épais indiquent que l'observation est focale. Les observations focales sont affichées avec un lien vers les *k* voisins les plus proches.

Contrôles et interactivité : Un certain nombre de contrôles dans le graphique vous permettent d'explorer l'espace des fonctions.

- Vous pouvez choisir quel sous-ensemble de fonctions vous souhaitez afficher dans le graphique et modifier les fonctions à représenter dans les dimensions.

- Les "observations focales" sont tout simplement des points sélectionnés dans le graphique de l'espace des fonctions. Si vous avez spécifié une variable d'observation focale, les points représentant les observations focales seront sélectionnées dès le début. Cependant, tous les points peuvent devenir temporairement une observation focale si vous les sélectionnez. Les "contrôles" habituels pour la sélection des points sont appliqués. Cliquer sur un point permet de sélectionner ce point et de désélectionner tous les autres. Cliquer sur un point avec la touche Ctrl enfoncée ajoute ce point à l'ensemble des points sélectionnés. Les vues liées, tels que le graphique des homologues, sont automatiquement mis à jour en fonction des observations sélectionnées dans l'espace des fonctions.
- Vous pouvez modifier le nombre de voisins les plus proches (*k*) à afficher pour les observations focales.
- Positionner le curseur sur un point du graphique affiche une infobulle avec la valeur du libellé d'observation, ou le nombre d'observations si les libellés d'observation ne sont pas définis, et les valeurs des cibles observées et prévues.
- Le bouton "Réinitialiser" permet de restaurer l'état d'origine de l'espace des fonctions.

Ajout et suppression des champs et des variables

Vous pouvez ajouter de nouveaux champs ou de nouvelles variables à l'espace des fonctions, ou supprimer ceux qui y sont déjà affichés.

Palette des variables

La palette des variables doit être affichée afin de pouvoir ajouter ou supprimer des variables. Pour faire apparaître la palette des variables, le visualiseur de modèles doit être en mode Edition et une observation doit être sélectionnée dans l'espace des fonctions.

1. Pour mettre le visualiseur de modèle en mode Edition, sélectionnez à partir des menus :

   Afficher > **Mode édition**
2. Une fois le mode Edition sélectionné, cliquez sur l'une des observations de l'espace des fonctions.
3. Pour afficher la palette des variables, dans les menus choisissez :

   Afficher > **Palettes** > **Variables**

   La palette des variables répertorie toutes les variables de l'espace des fonctions. L'icône en regard du nom de la variable indique le niveau de mesure de celle-ci.
4. Vous pouvez modifier le niveau de mesure d'une variable de façon temporaire. Pour cela, cliquez sur la variable avec le bouton droit de la souris dans la palette des variables et choisissez une option.

Zones des variables

Les variables sont ajoutées à des zones dans l'espace des fonctions. Pour afficher les zones, faites glisser une variable depuis la palette des variables ou sélectionnez **Afficher les zones**.

L'espace des fonctions comprend des zones pour les axes *x*, *y*, et *z*.

Déplacement des variables vers les zones

Voici quelques règles et conseils généraux permettant de déplacer les variables vers les zones :

- Pour déplacer une variable vers une zone, cliquez dessus et faites-la glisser de la palette des variables à la zone. Si vous sélectionnez **Afficher les zones**, vous pouvez cliquer avec le bouton droit sur une zone puis choisir la variable à ajouter à la zone.
- Si vous faites glisser une variable depuis la palette variables vers une zone déjà occupée par une autre variable, l'ancienne variable est remplacée par la nouvelle.
- Si vous faites glisser une variable depuis une zone vers une autre zone déjà occupée par une autre variable, les variables échangent leurs positions.
- En cliquant sur le signe X d'une zone, vous supprimez la variable située dans cette zone.
- Si la visualisation comporte plusieurs éléments graphiques, chaque élément peut posséder ses propres zones de variables associées. Sélectionnez d'abord l'élément graphique.

Tracé d'importance des variables

Généralement, vous souhaitez concentrer vos efforts de modélisation sur les variables les plus importantes et vous envisagez d'exclure et d'ignorer les moins importantes. Le graphique d'importance des variables peut vous y aider en indiquant l'importance relative de chaque variable en estimant le modèle. Etant donné que les valeurs sont relatives, la somme des valeurs pour l'ensemble des variables affichée est 1.0. L'importance des variables n'a aucun rapport avec la précision du modèle. Elle est juste rattachée à l'importance de chaque variable pour la réalisation d'une prévision, peu importe si cette dernière est précise ou non.

Homologues

Ce graphique affiche les observations focales et leurs *k* voisins les plus proches sur chaque fonction et sur la cible. Il est disponible si une observation focale est sélectionnée dans l'espace des fonctions.

Comportement de lien : Le graphique des homologues est relié à l'espace des fonctions de deux manières différentes.

- Les observations sélectionnées (focales) dans l'espace des fonctions sont affichées dans le graphique des homologues, ainsi que leurs *k* voisins les plus proches.
- La valeur de *k* sélectionnée dans l'espace des fonctions est utilisée dans le graphique des homologues.

Distances du voisin le plus proche

Ce tableau affiche les *k* voisins les plus proches et les distances pour les observations focales uniquement. Il est disponible si un identificateur d'observations focale est spécifié dans l'onglet Variable, et il n'affiche que les observations focales identifiées par cette variable.

Chaque ligne de :

- La colonne **Observation focale** contient la valeur de la variable de libellé des observations pour l'observation focale. Si les libellés d'observations ne sont pas définis, cette colonne contient le nombre d'observations de l'observation focale.
- La *i*ème colonne sous le groupe des voisins les plus proches contient la valeur de la variable de libellé d'observation pour le *i*ème voisin le plus proche de l'observation focale. Si les libellés d'observations ne sont pas définis, cette colonne contient le nombre d'observation du *i*ème voisin le plus proche de l'observation focale.
- La *i*ème colonne sous le groupe Distances les plus proches contient la distance du *i*ème voisin le plus proche de l'observation focale.

Carte des quadrants

Ce graphique affiche les observations focales et leur *k* voisins les plus proches sur un nuage de points (ou tracé de points, selon le niveau de mesure de la cible) avec la cible sur l'axe *y* et une fonction d'échelle sur l'axe *x*, affichés sous forme de panel par fonction. Il est disponible si une cible existe et si une observation focale est sélectionnée dans l'espace des fonctions.

- Les lignes de référence sont tracées pour des variables continues, aux moyennes variables dans la partition Apprentissage.

Journal des erreurs de sélection des fonctions

Les points du graphique affichent l'erreur (taux d'erreur ou l'erreur de la somme des carrés, selon le niveau de mesure de la cible) sur l'axe *y* pour le modèle avec la fonction sur l'axe *x* (plus toutes les fonctions à gauche sur l'axe *x*). Ce graphique est disponible si une cible existe et si la sélection des fonctions est activée.

Journal des erreurs de sélection de k

Les points du graphique affichent l'erreur (taux d'erreur ou l'erreur de la somme des carrés, selon le niveau de mesure de la cible) sur l'axe *y* pour le modèle avec le nombre de voisins les plus proches (*k*) sur l'axe *x*. Ce graphique est disponible si une cible existe et si la sélection de *k* est activée.

Journal des erreurs de sélection de k et de fonction

Il s'agit de graphiques de sélection de fonctions (voir «Journal des erreurs de sélection des fonctions », à la page 218), en panneaux par *k*. Ce graphique est disponible s'il existe une cible et si *k* et la sélection de la fonction sont tous deux en vigueur.

Tableau de classification

Cette table affiche par partition la classification croisée des valeurs prévues de la cible observées contre celles prévues. Elle est disponible si une cible existe et si elle est catégorielle.

- La ligne **(Manquante)** de la partition Restant contient des observations restantes contenant des valeurs manquantes sur la cible. Ces observations contribuent à l'échantillon restant : les valeurs de pourcentage global mais pas les valeurs de pourcentage correct.

Récapitulatif d'erreur

Ce tableau est disponible si une variable cible existe. Il affiche l'erreur associée au modèle, la somme des carrés pour une cible continue et le taux d'erreur (100% - pourcentage général correct) pour une cible catégorielle.

Analyse discriminante

L'analyse discriminante crée un modèle de prévision de groupe d'affectation. Le modèle est composé d'une fonction discriminante (ou, pour plus de deux groupes, un ensemble de fonctions discriminantes) basée sur les combinaisons linéaires des variables de prédicteur qui donnent la meilleure discrimination entre groupes. Les fonctions sont générées à partir d'un échantillon d'observations pour lesquelles le groupe d'affectation est connu. Les fonctions peuvent alors être appliquées aux nouvelles observations avec des mesures de variables de prédicteur, mais de groupe d'affectation inconnu.

Remarque : La variable de groupe peut avoir plus de deux valeurs. Les codes de la variable de regroupement doivent cependant être des nombres entiers, et vous devez spécifier leur valeur minimale et maximale. Les observations dont les valeurs se situent hors des limites sont exclues de l'analyse.

Exemple. En moyenne, les habitants des pays des zones tempérées consomment plus de calories par jour que ceux des tropiques, et une plus grande proportion de ces habitants vit en ville. Un chercheur veut combiner ces informations en une fonction pour déterminer comment un individu peut être différencié selon les deux groupes de pays. Le chercheur pense que la taille de la population et des informations économiques peuvent aussi être importantes. L'analyse discriminante vous permet d'estimer les coefficients de la fonction discriminante linéaire, qui ressemble à la partie droite d'une équation de régression linéaire multiple. Ainsi, en utilisant les coefficients *a*, *b*, *c* et *d*, la fonction est :

```
D = a * climate + b * urban + c * population + d * gross domestic product per capita
```

Si ces variables sont utiles pour établir la différence entre les deux zones climatiques, les valeurs de *D* seront différentes pour les pays tempérés et les pays tropicaux. Si vous utilisez une méthode de sélection des variables étape par étape, vous pouvez découvrir que vous n'avez pas forcément besoin d'inclure les quatre variables dans la fonction.

Statistiques : Pour chaque variable : moyennes, écarts types, ANOVA à 1 facteur. Pour chaque analyse : *Test M* de Box, matrice de corrélations intra-groupe, matrice de covariance intra-groupe, matrice de covariance de chaque groupe, matrice de covariance totale. Pour chaque fonction discriminante canonique : valeur propre, pourcentage de la variance, corrélation canonique, lambda de Wilks, khi-deux. Pour chaque

pas : probabilités a priori,coefficients de fonction de Fisher, coefficients de fonction non standardisés, lambda de Wilks pour chaque fonction canonique.

Remarques sur les données de l'analyse discriminante

Données. La variable de regroupement doit avoir un nombre limité de catégories distinctes, codifiées sous forme de nombres entiers. Les variables indépendantes nominales doivent être recodées en variables muettes ou de contraste.

Hypothèses : Les observations doivent être indépendantes. Les variables de prédicteur doivent avoir une distribution gaussienne multivariée, et les matrices de variance-covariance intra-groupes doivent être égales entre groupes. On part de l'hypothèse que les groupes d'affectation sont mutuellement exclusifs (c'est-à-dire qu'aucune observation n'est affectée à plus d'un groupe) et collectivement exhaustifs (c'est-à-dire que toutes les observations sont affectées à un groupe). La procédure est la plus efficace lorsque l'affectation à un groupe est une variable réellement catégorielle. Si l'affectation à un groupe est basée sur les valeurs d'une variable continue (par exemple, QI élevé contre QI bas), vous devez envisager d'utiliser la régression linéaire pour exploiter les informations plus riches données par la variable continue elle-même.

Pour obtenir une analyse discriminante

1. A partir des menus, sélectionnez :

   Analyse > **Classifier** > **Discriminant...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une variable de regroupement à valeur entière et cliquez sur **Définir plage** pour spécifier les catégories à considérer.
3. Sélectionnez les variables indépendantes, ou de prédicteur. (Si votre variable de regroupement n'a pas de valeurs entières, la procédure de recodification automatique du menu Transformer permettra d'en créer un avec des valeurs entières.)
4. Sélectionnez la méthode de saisie des variables indépendantes.
   - **Entrer les variables simultanément :** Entre simultanément toutes les variables indépendantes qui satisfont aux critères de tolérance.
   - **Utiliser la méthode détaillée étape par étape :** Utilise l'analyse étape par étape pour contrôler l'entrée et la suppression de variables.
5. Vous pouvez également sélectionner les observations avec une variable de sélection.

Définition de plages pour l'analyse discriminante

Spécifiez la valeur minimum et maximum de la variable de regroupement pour l'analyse. Les observations avec des valeurs hors de cette plage ne sont pas utilisées dans l'analyse discriminante mais elles sont classées dans un des groupes existants en fonction des résultats de l'analyse. Les valeurs minimum et maximum doivent être des entiers.

Sélection des observations pour l'analyse discriminante

Pour sélectionner les observations pour votre analyse :

1. Dans la boîte de dialogue Analyse discriminante, sélectionnez une variable de sélection.
2. Cliquez sur **Valeur** pour entrer un entier comme valeur de sélection.

Seules les observations avec la valeur spécifiée pour la variable de sélection sont utilisées pour dériver les fonctions discriminantes. Les résultats des statistiques et de classification sont générés pour les observations sélectionnées et celles qui ne le sont pas. Ce processus fournit une méthode de classification des nouvelles observations reposant sur des données existantes ou de partitionnement de vos données dans un sous-ensemble de test ou de formation en vue d'effectuer une validation sur le modèle créé.

Statistiques de l'analyse discriminante

Descriptives : Les options disponibles sont moyennes (y compris les écarts types), ANOVA à 1 facteur et Test *M* de Box.

- *moyennes*. Affiche les moyennes totales et les moyennes de groupe, ainsi que l'écart type des variables indépendantes.
- *ANOVA univariées*. Effectue pour chacune des variables indépendantes une analyse de variance à 1 facteur pour tester l'égalité des moyennes de groupe.
- *Test de Box*. Test d'égalité des matrices de covariance de groupes. Pour les échantillons de taille suffisamment importante, une valeur p non significative indique qu'il n'est pas démontré que les matrices diffèrent. Ce test est sensible aux déviations par rapport à la distribution gaussienne multivariée.

Coefficients de la fonction : Les options disponibles sont les coefficients de la classification de Fisher et les coefficients non standardisés.

- *de Fisher*. Affiche les coefficients de la fonction de classification de Fisher qui peuvent être directement utilisés pour la classification. Un groupe séparé de coefficients de fonctions de classification est obtenu pour chaque groupe et une observation est affectée au groupe qui a le plus grand score discriminant (valeur de fonction de classification).
- *non standardisés*. Affiche les coefficients de fonction discriminante non standardisés.

Matrices : Les matrices de coefficients pour variables indépendantes disponibles sont la matrice de corrélation intra-groupe, la matrice de covariance intra-groupe, la matrice de covariance de chaque groupe et la matrice de covariance totale.

- *Corrélation intra-groupe*. Affiche une matrice de corrélation intra-groupes regroupée en pool, obtenue en calculant la moyenne des matrices de covariance distinctes pour tous les groupes avant de calculer les corrélations.
- *Covariance intra-groupe*. Affiche une matrice de covariance intragroupes regroupée en pool qui peut différer de la matrice de covariance totale. Cette matrice est obtenue en calculant la moyenne des matrices de covariances distinctes de tous les groupes.
- *Covariance de chaque groupe*. Affiche des matrices de covariance par groupes distincts pour chaque groupe.
- *Covariance totale*. Affiche la matrice de covariance de toutes les observations comme si elles provenaient d'un seul échantillon.

Méthode détaillée étape par étape de l'analyse discriminante

Méthode. Sélectionnez la statistique à utiliser pour introduire ou éliminer de nouvelles variables. Les alternatives disponibles sont la lambda de Wilks, la variance inexpliquée, la distance de Mahalanobis, le plus petit rapport *F* et le *V*de Rao. Avec le *V*de Rao, vous pouvez spécifier l'augmentation minimale de *V* pour une variable à entrer.

- *Lambda de Wilks*. Méthode de sélection des variables pour une analyse discriminante étape par étape qui sélectionne les variables à introduire dans l'équation d'après leur capacité à faire baisser le lambda de Wilks. A chaque pas, la variable qui minimise le lambda de Wilks global est introduite.
- *Variance inexpliquée*. A chaque étape, la variable qui minimise la somme des variations inexpliquées (résiduelles) entre les groupes est saisie.
- *Mahalanobis distance*. Mesure du degré de différence entre les valeurs d'une observation sur les variables indépendantes et la moyenne de toutes les observations. Une distance de Mahalanobis importante identifie une observation qui a des valeurs extrêmes pour des variables indépendantes.
- *Plus petit rapport F*. Méthode de sélection des variables en analyse pas à pas, fondée sur la maximisation d'un rapport F calculé à partir de la distance de Mahalanobis entre des groupes.
- *V de Rao*. Mesure des différences entre les moyennes des groupes. Egalement appelée trace de Lawley-Hotelling. A chaque étape, la variable qui maximise l'augmentation du V de RAO est entrée. Après

avoir sélectionné cette option, entrez la valeur minimale que doit avoir une variable pour entrer dans l'analyse.

Critères : Les alternatives disponibles sont **Utiliser la valeur F** et **Utiliser la probabilité de F**. Entrez des valeurs pour entrer et supprimer des variables.

- *Utiliser la valeur F*. Une variable est introduite dans un modèle si sa valeur F est supérieure à la valeur Entrée et elle est éliminée si la valeur F est inférieure à la valeur Suppression. La valeur Entrée doit être supérieure à la valeur Suppression et toutes deux doivent être positives. Pour introduire davantage de variables dans le modèle, réduisez la valeur du champ Entrée. Pour éliminer davantage de variables dans le modèle, augmentez la valeur du champ Suppression.
- *Choisir la probabilité de F*. Une variable est entrée dans le modèle si le niveau de signification de la valeur F est inférieur à la valeur Entrée et elle est éliminée si le niveau de signification est supérieur à la valeur Suppression. La valeur Entrée doit être inférieure à la valeur Suppression et toutes deux doivent être positives. Pour introduire davantage de variables dans le modèle, diminuez la valeur Entrée. Pour éliminer davantage de variables du modèle, réduisez la valeur Suppression.

Afficher. Récapitulatif des étapes affiche des statistiques pour toutes les variables après chaque étape ; **F pour les distances par paires** affiche une matrice des rapports *F* par paires pour chaque paire de groupes.

Classement de l'analyse discriminante

Probabilités a priori : Cette option permet de déterminer si les coefficients de classification sont ajustés pour une connaissance a priori de l'appartenance à un groupe.

- **Egales pour tous les groupes :** Des probabilités a priori égales sont supposées pour tous les groupes; ceci n'a aucune incidence sur les coefficients.
- **A calculer selon les effectifs :** Les tailles de groupes observés dans votre échantillon déterminent les probabilités a priori du groupe d'appartenance. Par exemple, si 50% des observations comprises dans l'analyse appartiennent au premier groupe, 25 % au deuxième, et 25 % au troisième, les coefficients de classification sont ajustés pour accroître la probabilité d'affectation du premier groupe par rapport aux deux autres.

Afficher. Les options d'affichage disponibles sont les résultats par observation, la table récapitulative et la classification par élimination.

- *Résultats par observation*. Les codes du groupe actuel, du groupe prévu, des probabilités à posteriori et des scores discriminants sont affichés pour chaque observation.
- *Table récapitulative*. Nombre d'observations correctement et incorrectement affectées à chacun des groupes sur la base de l'analyse discriminante. Parfois appelés "matrice de confusion".
- *Classification par élimination*. Classement de chaque observation de l'analyse par les fonctions dérivées de l'ensemble des observations autres que cette observation. Cette classification est également appelée "méthode U".

Remplacer les valeurs manquantes par la moyenne : Sélectionnez cette option pour remplacer la valeur manquante d'une variable indépendante par la moyenne de cette variable, mais seulement durant la phase de classification.

Utiliser la matrice de covariance : Vous pouvez choisir de classer les observations en utilisant une matrice de covariance intra-groupe ou une matrice de covariance groupe par groupe.

- *Sous-groupes*. La matrice de covariances intra-groupes regroupée en pool est utilisée pour classifier les observations.
- *Groupes distincts*. Les matrices de covariance de chaque groupe sont utilisées pour la classification. Comme la classification repose sur les fonctions discriminantes et pas sur les variables d'origine, cette option n'est pas toujours équivalente à la discrimination quadratique.

Tracés : Les options de tracé disponibles sont groupes combinés, groupe par groupe, et carte territoriale d'affectation.

- *Groupes combinés*. Crée un nuage de points de tous les groupes, des deux premières valeurs de fonction discriminante. S'il n'y a qu'une seule fonction, un histogramme est tracé à la place.
- *Groupes distincts*. Crée des nuages de points par groupes distincts des deux premières valeurs de fonction discriminante. Lorsqu'il n'y a qu'une seule fonction, des histogrammes sont affichés à la place.
- *Carte territoriale*. Tracé des limites servant à classifier les observations en fonction de valeurs de fonction. Les numéros correspondent aux groupes auxquels les observations ont été affectées. La moyenne de chaque groupe est indiquée par un astérisque à l'intérieur de ses limites. La carte n'est pas affichée s'il n'existe qu'une seule fonction discriminante.

Enregistrement de l'analyse discriminante

Vous pouvez ajouter de nouvelles variables à votre fichier de données actif. Les options disponibles sont classe(s) d'affectation (une seule variable), valeurs du facteur discriminant (une variable pour chaque fonction discriminante dans la solution), et probabilités d'affectation à un groupe en fonction des valeurs du facteur discriminant (une variable pour chaque groupe).

Vous pouvez également exporter les informations du modèle vers le fichier spécifié au format XML. Vous pouvez utiliser ce fichier de modèle pour appliquer les informations du modèle aux autres fichiers de données à des fins d'évaluation.

Fonctions supplémentaires de la commande DISCRIMINANT

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

- effectuer plusieurs analyses discriminantes avec une seule commande et contrôler l'ordre d'entrée des variables (au moyen de la sous-commande `ANALYSIS`) ;
- spécifier des probabilités a priori pour la classification (au moyen de la sous-commande `PRIORS`) ;
- afficher les matrices des coordonnées factorielles et les matrices des corrélations structurelles après rotation (au moyen de la sous-commande `ROTATE`) ;
- limiter le nombre de fonctions discriminantes extraites (au moyen de la sous-commande `FUNCTIONS`) ;
- restreindre la classification aux observations sélectionnées (ou non sélectionnées) pour l'analyse (au moyen de la sous-commande `SELECT`) ;
- lire et analyser une matrice de corrélation (au moyen de la sous-commande `MATRIX`) ;
- créer une matrice de corrélation pour une analyse ultérieure (au moyen de la sous-commande `MATRIX`) ;

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

Analyse factorielle

L'analyse factorielle essaie d'identifier des variables sous-jacentes, ou **facteurs**, qui permettent d'expliquer le motif des corrélations à l'intérieur d'un ensemble de variables observées. L'analyse factorielle est souvent utilisée pour réduire un jeu de données. L'analyse factorielle est souvent utilisée dans la réduction de données, en identifiant un petit nombre de facteurs qui expliquent la plupart des variances observées dans le plus grand nombre de variables manifestes. On peut également utiliser l'analyse factorielle pour générer des hypothèses concernant des mécanismes de causalité ou pour afficher des variables pour une analyse ultérieure (par exemple, pour identifier la colinéarité avant une analyse de régression linéaire).

La procédure d'analyse factorielle offre une très grande flexibilité :

- Il existe sept méthodes d'extraction de facteur.
- Il existe cinq méthodes de rotation, dont directe Oblimin et Promax pour les rotations non orthogonales.
- Il existe trois méthodes pour calculer les scores factoriels, et ces facteurs peuvent être enregistrés en tant que variables pour des analyses ultérieures.

Exemple. Quelle est l'attitude sous-jacente qui pousse les personnes à répondre d'une certaine manière aux questions concernant une enquête politique ? L'examen des corrélations parmi les éléments d'une enquête révèle qu'il y a des recouvrements significatifs parmi divers sous-groupes d'éléments. Les questions sur les impôts ont tendance à être en corrélation, de même que les questions à thèmes militaires, etc. Avec l'analyse factorielle, vous pouvez enquêter sur le nombre de facteurs sous-jacents, et, dans de nombreux cas, vous pouvez identifier le concept représenté par ces facteurs. De plus, vous pouvez calculer les scores factoriels pour chaque répondant, facteurs que vous pouvez utiliser pour des analyses ultérieures. Par exemple, sur la base des scores factoriels, vous pouvez développer un modèle logistique de régression pour prévoir le comportement de vote.

Statistiques : Pour chaque : nombre d'observations valides, moyenne et écart type. Pour chaque analyse factorielle : matrice de corrélation des variables, incluant des niveaux de signification, déterminant, inverse ; les matrices des corrélations reconstituées, incluant l'anti-image ; les solutions initiales (qualités de représentation, valeurs propres et pourcentage de variance expliqué) ; mesure d'adéquation d'échantillonnage de Kaiser-Meyer-Olkin et le test de sphéricité de Bartlett ; structure avant rotation, incluant les chargements des facteurs, la qualité de représentation, et les valeurs propres; structure après rotation, incluant une matrice de forme après rotation et une matrice de transformation. Pour les rotations obliques : motif et matrices de structure après rotation ; matrice de coefficients de scores factoriels et matrice de facteur de covariance. Tracés : tracé d'effondrement et carte de chargement du premier, du deuxième et du troisième facteur.

Remarques sur les données d'analyse factorielle

Données. Les variables doivent être quantitatives au niveau de l'*intervalle* ou du *rapport*. Les données catégorielles (comme la religion ou le pays d'origine) ne conviennent pas pour l'analyse factorielle. Les données pour lesquelles la corrélation de Pearson calculée a un sens conviennent pour l'analyse factorielle.

Hypothèses : Les données doivent posséder une distribution gaussienne bivariée pour chaque paire de variables et les observations doivent être indépendantes. Le modèle d'analyse factorielle spécifie que les variables sont déterminées par des facteurs communs (les facteurs estimés par le modèle) et des facteurs uniques (qui ne sont pas corrélés entre les variables observées); les estimations calculées se basent sur l'hypothèse que tous les facteurs uniques ne sont pas en corrélation entre eux ainsi qu'avec les facteurs communs.

Pour obtenir une analyse factorielle

1. A partir des menus, sélectionnez :

   Analyse > **Réduction des dimensions** > **Analyse factorielle...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez les variables pour l'analyse factorielle.

Sélection des observations pour l'analyse factorielle

Pour sélectionner les observations pour votre analyse :

1. Sélectionnez une variable de sélection.
2. Cliquez sur **Valeur** pour entrer un entier comme valeur de sélection.

Seules les observations ayant cette valeur pour la variable de sélection sont utilisées dans l'analyse factorielle.

Caractéristiques d'analyse factorielle

Statistiques. Caractéristiques univariées inclut la moyenne, l'écart type et le nombre de cas valides pour chaque variable. La **structure initiale** affiche la qualité de représentation initiale, les valeurs propres et le pourcentage de variance expliqué.

Matrice de corrélation : Les options disponibles sont les coefficients, les niveaux de signification, les déterminants, les inverses, les reproduits, l'anti-image et l'indice KMO et le test de sphéricité de Bartlett.

- *KMO et le test de sphéricité de Bartlett*. Mesure d'adéquation de l'échantillonnage de Kaiser-Meyer-Olkin qui teste si les corrélations partielles entre les variables sont faibles. Le test de sphéricité de Bartlett teste si la matrice des corrélations est une matrice d'identité, ce qui indiquerait que le modèle de facteur n'est pas adapté.
- *Reproduit*. Matrice de corrélation estimée à partir de la solution factorielle. Les résidus (différence entre les corrélations estimées et observées) sont également affichés.
- *Anti-image*. La matrice de corrélation des anti-images contient les négatifs des coefficients de corrélation partielle et la matrice de covariance des anti-images contient les négatifs des covariances partielles. Dans un bon modèle factoriel, la plupart des éléments hors diagonale doivent être petits. La mesure d'adéquation d'échantillonnage pour une variable est affichée sur la diagonale de la matrice de corrélation des anti-images.

Extraction d'analyse factorielle

Méthode. Vous permet de spécifier la méthode d'extraction de facteur. Les méthodes disponibles sont les Composantes principales, les Moindres carrés non pondérés, les Moindres carrés généralisés, le Maximum de vraisemblance, la Factorisation en axes principaux, l'Alpha-maximisation et la Factorisation en projections.

- *Analyse des composantes principales*. Méthode d'extraction de facteur utilisée pour former des combinaisons linéaires non corrélées des variables observées. La première composante principale a une variance maximale. Les autres composantes expliquent progressivement des portions plus petites de la variance sans être corrélées les unes aux autres. L'analyse des composantes principales est utilisée pour obtenir la solution factorielle initiale. Elle peut être utilisée quand la matrice des corrélations est singulière.
- *méthode des moindres carrés non pondérés*. Méthode d'extraction de facteur qui minimise la somme des carrés des différences entre les matrices de corrélations observées et reconstituées, en ignorant les diagonales.
- *méthode des moindres carrés généralisés*. Méthode d'extraction de facteur qui minimise la somme des carrés des différences entre les matrices de corrélations observées et reconstituées. Les corrélations sont pondérées par l'inverse de leur unicité, de façon à ce que les variables présentant une forte unicité reçoivent une pondération inférieure à celles présentant une faible unicité.
- *Méthode du maximum de vraisemblance*. Méthode d'extraction de facteur qui fournit les estimations de paramètres les plus susceptibles d'avoir généré la matrice de corrélation observée si l'échantillon est issu d'une distribution normale multivariée. Les corrélations sont pondérées par l'inverse de l'unicité des variables et un algorithme itératif est utilisé.
- *Factorisation en axes principaux*. Méthode d'extraction de facteurs à partir de la matrice des corrélations initiales où les coefficients de corrélation multiple au carré sont placés sur la diagonale comme estimation initiale des qualités de représentation. Ces chargements factoriel sont utilisés pour une nouvelle estimation des qualités de représentation qui remplace alors l'ancienne sur la diagonale. Les itérations se poursuivent jusqu'à ce que les variations des qualités de représentation d'une itération à l'autre satisfassent le critère de convergence de l'extraction.
- *Factorisation alpha*. Méthode d'extraction de facteur qui considère les variables de l'analyse comme étant un échantillon de l'univers des variables de potentiel. Cette méthode maximise l'alpha de Cronbach des facteurs.
- *factorisation en projections*. Méthode d'extraction de facteur développée par Guttman et basée sur la théorie des images. La partie commune de la variable, appelée image partielle, est définie comme sa régression linéaire sur les autres variables, plutôt qu'une fonction de facteurs hypothétiques.

Analyser : Vous permet de spécifier si l'analyse porte sur une matrice de corrélation ou sur une matrice de covariance.

- **Matrice de corrélation :** Utile si les variables de votre analyse sont mesurées selon des échelles différentes.
- **Matrice de covariance :** Utile lorsque vous souhaitez appliquer l'analyse factorielle à plusieurs groupes avec des variances différentes pour chaque variable.

Extraire : Vous pouvez retenir tous les facteurs dont les valeurs propres dépassent une valeur spécifique ou retenir un nombre spécifique de facteurs.

Afficher. Vous permet de demander la solution factorielle avant rotation et un tracé d'effondrement.

- *solution factorielle sans rotation*. Affiche les corrélations factorielles sans rotation (matrice de la structure factorielle), les qualités de représentation et les valeurs propres de la solution factorielle.
- *Tracé d'effondrement*. Tracé de la variance associée à chaque facteur. Permet de déterminer le nombre de facteurs à conserver. Généralement, le tracé montre une rupture franche entre la forte pente des facteurs élevés et la traîne graduelle du reste (valeurs propres).

Maximum des itérations pour converger : Vous permet de spécifier le nombre maximum d'étapes que l'algorithme peut utiliser pour estimer la solution.

Rotation d'analyse factorielle

Méthode. Permet de sélectionner la méthode de rotation des facteurs et d'appliquer ou non la normalisation Kaiser. Les méthodes disponibles sont Varimax, Oblimin directe, Quartimax, Equamax ou Promax.

- *méthode varimax*. Méthode de rotation orthogonale qui minimise le nombre de variables ayant de fortes corrélations sur chaque facteur. Simplifie l'interprétation des facteurs.
- *Méthode Oblimin directe*. Méthode de rotation oblique (non orthogonale). Lorsque delta est nul (valeur par défaut), les solutions sont les plus obliques. Plus la valeur de delta est négative, moins les facteurs sont obliques. Pour remplacer la valeur nulle par défaut de delta, entrez un nombre inférieur ou égal à 0,8.
- *Méthode Quartimax*. Méthode de rotation qui réduit le nombre de facteurs requis pour expliquer chaque variable. Simplifie l'interprétation des variables observées.
- *méthode Equamax*. Méthode de rotation qui est une combinaison de la méthode Varimax (qui simplifie les facteurs) et de la méthode Quartimax (qui simplifie les variables). Le nombre de variables pesant sur un facteur et le nombre de facteurs nécessaires pour expliquer une variable sont minimisés.
- *Rotation Promax*. Rotation oblique qui permet la corrélation des facteurs. Peut être calculée plus rapidement qu'une rotation oblimin directe, aussi est-elle utile pour les vastes jeux de données.
- *Appliquer la normalisation Kaiser*. Par défaut, permet d'appliquer la normalisation Kaiser lorsqu'une rotation est spécifiée.

Afficher. Vous permet d'inclure la sortie de la structure après rotation, et également d'afficher les tracés de chargement sur le premier, le second et le troisième facteur (Cartes factorielles).

- *Structure après rotation*. Vous devez sélectionner une méthode de rotation pour obtenir une solution après rotation. Pour les rotations orthogonales, la matrice de forme après rotation et la matrice de transformation factorielle sont affichées. Pour les rotations obliques, le programme affiche la matrice des projections factorielles, la matrice de structure et la matrice des corrélations de facteurs.
- *Tracé des saturations de facteurs*. Tracé de chargement de facteur tridimensionnel pour les trois premiers facteurs. Pour une solution à deux facteurs, un tracé en deux dimensions est affiché. Le tracé n'est pas affiché si un seul facteur est extrait. Les tracés affichent des solutions ayant subi une rotation si cette dernière est demandée.

Maximum des itérations pour converger : Vous permet de spécifier le nombre maximum d'étapes que l'algorithme peut utiliser pour réaliser la rotation.

Scores d'analyse factorielle

Enregistrer dans des variables : Vous permet de créer une nouvelle variable pour chaque facteur selon la structure finale.

Méthode. Les méthodes alternatives pour calculer les scores factoriels sont la Régression, Bartlett, et Anderson-Rubin.

- *Méthode de régression*. Méthode d'estimation des coefficients de score factoriel. Les écarts obtenus ont une moyenne de 0 et une variance égale au carré de la corrélation multiple entre les scores factoriels estimés et les vraies valeurs du facteur. Les écarts peuvent être corrélés même lorsque les facteurs sont orthogonaux.
- *Scores Bartlett*. Méthode d'estimation des coefficients de scores factoriels. Les scores produits ont une moyenne de 0. La somme des carrés des facteurs uniques sur la gamme des variables est minimisée.
- *méthode d'Anderson-Rubin*. Méthode d'estimation des coefficients factoriels ; variante de la méthode de Bartlett qui garantit l'orthogonalité des facteurs estimés. Les scores obtenus affichent une moyenne de 0 et un écart type de 1 et ne sont pas corrélés.

Afficher la matrice des coefficients des scores factoriels : Vous permet de montrer les coefficients par lesquels les variables sont multipliées pour obtenir les scores factoriels. Cela permet également de montrer les corrélations entre les scores factoriels.

Options d'analyse factorielle

Valeurs manquantes : Vous permet de spécifier comment traiter les valeurs manquantes. Les options disponibles sont d'exclure toute observation *incomplète*, d'exclure seulement les composantes *non valides*, ou de les remplacer par la moyenne.

Affichage des projections : Vous permet de contrôler le format des matrices de sortie. Triez les coefficients par leur taille et supprimez les coefficients dont la valeur absolue est inférieure à la valeur spécifiée.

Fonctions supplémentaires de la commande FACTOR

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

- spécifier des critères de convergence pour les itérations lors de l'extraction et de la rotation ;
- définir des tracés factoriels individuels après rotation ;
- indiquer le nombre de scores factoriels à enregistrer ;
- spécifier les valeurs des diagonales pour la méthode de factorisation en axes principaux ;
- créer des matrices de corrélation ou des matrices de chargement factoriel sur un disque pour une analyse ultérieure ;
- lire et analyser des matrices de corrélation ou des matrices de chargement factoriel.

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

Choix d'une procédure de classification

Vous pouvez effectuer des analyses de cluster à l'aide de la procédure TwoStep, hiérarchique ou de nuées dynamiques. Chaque procédure utilise un algorithme différent pour la création des clusters, et chacune d'elles comporte des options qui ne sont pas disponibles dans les autres procédures.

Analyse de cluster TwoStep : La procédure Analyse de cluster TwoStep est la méthode privilégiée pour de nombreuses applications. Elle offre les fonctions spécifiques suivantes :

- Sélection automatique du meilleur nombre de clusters, en plus des mesures de sélection parmi des modèles de cluster.
- Possibilité de créer simultanément des modèles de cluster sur la base de variables catégorielles et continues.
- Possibilité d'enregistrer le modèle de cluster dans un fichier XML externe, puis de lire ce fichier et de mettre à jour le modèle de cluster à l'aide des données les plus récentes.

En outre, la procédure Analyse de cluster TwoStep permet d'analyser des fichiers de données volumineux.

Analyse de cluster hiérarchique : La procédure d'analyse de cluster hiérarchique est limitée à des fichiers de données plus petits (centaines d'objets à classer), mais offre les fonctions spécifiques suivantes :

- Possibilité de classer des observations ou des variables.
- Possibilité de calculer plusieurs solutions possibles et d'enregistrer des clusters d'affectation pour chacune de ces solutions.
- Plusieurs méthodes de formation de clusters, de transformation de variables et de mesure de la dissimilarité entre les clusters.

Tant que toutes les variables sont du même type, la procédure d'analyse de cluster hiérarchique peut analyser des variables d'intervalle (continues), d'effectif ou binaires.

Analyse de cluster des nuées dynamiques : La procédure d'analyse de cluster de nuées dynamiques est limitée aux données continues et exige que vous indiquiez au préalable le nombre de clusters. Elle offre néanmoins les fonctions spécifiques suivantes :

- Possibilité d'enregistrer les distances à partir des centres de clusters pour chaque objet.
- Possibilité de lire les centres de clusters initiaux à partir d'un fichier IBM SPSS Statistiques externe et d'enregistrer les centres de clusters finaux dans ce même fichier.

En outre, la procédure d'analyse de cluster de nuées dynamiques permet d'analyser des fichiers de données volumineux.

Analyse de cluster TwoStep

L'analyse de cluster TwoStep est un outil d'exploration conçu pour révéler des groupements naturels (ou clusters) au sein d'un jeu de données. L'algorithme utilisé par cette procédure possède plusieurs fonctions qui le distinguent des techniques de classification standard :

- **Gestion des données catégorielles et continues :** En supposant que les variables soient indépendantes, une distribution jointe multinomiale-normale peut être placée sur des variables catégorielles et continues.
- **Sélection automatique du nombre de clusters :** En comparant les valeurs d'un critère de modèle-choix dans différentes solutions de classification, la procédure peut déterminer automatiquement le nombre optimal de clusters.
- **Evolutivité :** En construisant une arborescence de fonctions de cluster (CF) qui récapitule les enregistrements, l'algorithme TwoStep vous permet d'analyser des fichiers de données volumineux.

Exemple. Les entreprises de vente au détail et de produits de consommation appliquent régulièrement des techniques de regroupement à des données qui décrivent les habitudes d'achat de leurs clients, leur sexe, leur âge, leur niveau de revenu, etc. Ces entreprises adaptent leur marketing et leurs stratégies de développement de produits pour accroître leurs ventes et fidéliser leur clientèle.

Mesure de distance : Cette sélection détermine la façon dont la similarité entre deux clusters est calculée.

- **Log de vraisemblance :** La mesure de vraisemblance place une distribution de probabilité sur les variables. Les variables continues sont considérées comme étant distribuées normalement alors que les variables catégorielles sont considérées comme étant multinomiales. Toutes les variables sont considérées comme étant indépendantes.
- **Euclidienne :** La mesure euclidienne est la distance « en ligne droite » entre deux clusters. Elle peut être utilisée uniquement lorsque toutes les variables sont continues.

Nombre de clusters : Cette sélection vous permet d'indiquer la façon dont le nombre de clusters doit être déterminé.

- **Déterminer automatiquement :** Cette procédure déterminera automatiquement le « meilleur » nombre de clusters en utilisant le critère défini dans le groupe Critère de classification. Vous pouvez également entrer un nombre entier positif qui définit le nombre maximal de clusters que la procédure doit prendre en compte.

- **Indiquer une valeur fixe :** Vous permet d'indiquer le nombre de clusters (valeur fixe) dans la solution. Entrez un entier positif.

Nombre de variables continues : Ce groupe fournit un récapitulatif des spécifications de standardisation des variables continues qui sont définies dans la boîte de dialogue Options. Pour plus d'informations, reportez-vous à la rubrique «Options de la procédure d'analyse de cluster TwoStep», à la page 229.

Critère de classification : Cette sélection détermine la façon dont l'algorithme de classification automatique détermine le nombre de clusters. Vous pouvez spécifier le critère d'information bayésien (BIC) ou le critère d'information d'Akaike (AIC).

Remarques sur les données de la procédure d'analyse de cluster TwoStep

Données. Cette procédure fonctionne avec des variables continues et catégorielles. Les observations représentent les objets à classer et les variables représentent les attributs sur lesquels est basée la classification.

Tri par observation : Remarque : l'arborescence des fonctions de cluster et la solution finale peuvent dépendre de l'ordre des observations. Pour réduire les effets de tri, classez les observations de manière aléatoire. Vous pouvez obtenir différentes solutions pour lesquelles les observations ont été triées de manière aléatoire, afin de vérifier la stabilité d'une solution donnée. Lorsque cela s'avère difficile en raison de fichiers très volumineux, vous pouvez effectuer plusieurs fois l'opération sur un échantillon des observations triées de différentes manières aléatoires.

Hypothèses : La mesure de la distance de vraisemblance considère que les variables du modèle de cluster sont indépendantes. De plus, chaque variable continue est considérée comme ayant une distribution normale (gaussienne) et chaque variable catégorielle comme ayant une distribution multinomiale. Des tests internes empiriques indiquent que la procédure est assez résistante aux violations de l'hypothèse d'indépendance et des hypothèses de distribution, mais vous devez savoir comment ces hypothèses sont vérifiées.

Utilisez la procédure Corrélations bivariées pour tester l'indépendance des deux variables continues. Utilisez la procédure Tableaux croisés pour tester l'indépendance de deux variables catégorielles. Utilisez la procédure Moyennes pour tester l'indépendance entre une variable continue et une variable catégorielle. Utilisez la procédure Explorer pour tester la normalité d'une variable continue. Utilisez la procédure Test du Khi-deux pour tester si une variable catégorielle possède une distribution multinomiale spécifique.

Pour effectuer une procédure d'analyse de cluster TwoStep

1. A partir des menus, sélectionnez :

   Analyse > **Classifier** > **Cluster TwoStep...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables catégorielles ou continues.

(En option) Vous avez également ces possibilités :

- Ajuster les critères sur lesquels est basée la construction des clusters.
- Sélectionner les paramètres de gestion du bruit, d'affectation de mémoire, de standardisation de variable et d'entrée de modèle de cluster.
- Demander les sorties du visualiseur de modèles.
- Enregistrer les résultats de modèle dans le fichier de travail ou dans un fichier XML externe.

Options de la procédure d'analyse de cluster TwoStep

Traitement des valeurs extrêmes : Ce groupe permet de traiter les valeurs extrêmes, notamment lors de la classification, si l'arborescence des fonctions de cluster (CF) est saturée. L'arborescence CF est saturée si elle ne peut plus accepter d'autres observations dans un noeud feuille et qu'aucun noeud feuille ne peut être divisé.

- Si vous sélectionnez la gestion du bruit et que l'arborescence CF est saturée, l'arborescence est reconstruite lorsque vous placez des observations de feuilles éclatées dans une feuille « bruit ». Une feuille est éclatée si elle contient un pourcentage inférieur au pourcentage d'observations correspondant à la taille maximale de la feuille. Une fois que l'arborescence est reconstruite, les valeurs extrêmes sont placées dans l'arborescence CF si cela est possible. Sinon, les valeurs extrêmes sont supprimées.
- Si vous ne sélectionnez pas la gestion du bruit et que l'arborescence CF est saturée, elle sera reconstruite à l'aide d'un seuil de changement de distance supérieur. Après la classification finale, les valeurs ne pouvant être affectées à un cluster deviennent des valeurs extrêmes libellées. Un numéro d'identification de -1 est affecté au cluster de valeur extrême et cette dernière n'est pas prise en compte dans le nombre de clusters.

Allocation de mémoire : Ce groupe vous permet d'indiquer la quantité de mémoire maximale en mégaoctets (Mo) que l'algorithme de cluster doit utiliser. Si la procédure dépasse cette limite, elle utilisera le disque pour stocker les informations ne pouvant pas être enregistrées en mémoire. Spécifiez une valeur supérieure ou égale à 4.

- Consultez l'administrateur système pour connaître la plus grande valeur que vous pouvez spécifier sur votre système.
- L'algorithme risque de ne pas trouver le nombre correct ou spécifié de clusters si cette valeur est trop basse.

Standardisation de variable : L'algorithme de classification fonctionne avec des variables continues standardisées. Les variables continues non standardisées doivent être laissées en tant que variables dans la liste À standardiser . Pour gagner du temps et éviter trop de calculs, vous pouvez sélectionner une variable continue déjà standardisée comme variable dans la liste Standardisée.

Options avancées

Critères de réglage de l'arborescence CF : Les paramètres d'algorithme de classification suivants s'appliquent de façon spécifique à l'arborescence CF et doivent être modifiés avec le plus grand soin :

- **Seuil de modification de distance initiale :** Il s'agit du seuil initial utilisé pour construire l'arborescence CF. Si l'insertion d'une observation dans une feuille de l'arborescence CF provoque une étroitesse inférieure au seuil, la feuille n'est pas divisée. Si l'étroitesse dépasse le seuil, la feuille est divisée.
- **Nombre maximum de branches (par noeud feuille) :** Nombre maximum de noeuds enfant qu'un noeud feuille peut contenir.
- **Profondeur maximum de l'arborescence :** Nombre maximum de niveaux que l'arborescence CF peut contenir.
- **Nombre maximal de noeuds :** Indique le nombre maximal de noeuds de l'arborescence CF pouvant être générés par la procédure, d'après la fonction $(b^{d+1} - 1) / (b - 1)$, où b correspond au nombre maximal de branches et d la profondeur maximale de l'arbre. Notez qu'une arborescence CF trop volumineuse risque d'épuiser les ressources du système et d'avoir des effets défavorables sur les performances de la procédure. Chaque noeud exige au moins 16 octets.

Mettre à jour le modèle de cluster : Ce groupe vous permet d'importer et de mettre à jour un modèle de cluster généré dans une analyse précédente. Le fichier d'entrée contient l'arborescence CF au format XML. Le modèle est ensuite mis à jour avec les données du fichier actif. Vous devez sélectionner les noms des variables dans la boîte de dialogue principale dans le même ordre que celui dans lequel ils ont été spécifiés dans l'analyse précédente. Le fichier XML demeure inchangé, sauf si vous enregistrez les informations du nouveau modèle en utilisant le même nom de fichier. Pour plus d'informations«Sortie de l'analyse de cluster TwoStep», à la page 231, voir la rubrique .

Si vous avez indiqué la mise à jour d'un modèle de cluster, les options relatives à la génération de l'arborescence CF spécifiées pour le modèle d'origine sont utilisées. Plus précisément, les paramètres de mesure de la distance, de gestion du bruit, d'affectation de mémoire ou les critères de réglage de l'arborescence CF pour le modèle enregistré sont utilisés et tous les paramètres de ces options dans les boîtes de dialogue sont ignorés.

Remarque : Lorsque vous effectuez une mise à jour d'un modèle de cluster, la procédure considère qu'aucune des observations sélectionnées dans le jeu de données actif n'a été utilisée pour créer le modèle de cluster d'origine. La procédure considère également que les observations utilisées dans la mise à jour du modèle sont issues de la même population d'observations utilisée pour créer le modèle d'origine. En d'autres termes, les moyennes et les variances des variables continues et les niveaux des variables catégorielles sont considérés comme étant identiques dans les deux groupes d'observations. Si votre « nouveau » groupe d'observations ne provient pas de la même population que votre « ancien » groupe, exécutez la procédure Analyse de cluster TwoStep sur les groupes d'observations combinés pour obtenir de meilleurs résultats.

Sortie de l'analyse de cluster TwoStep

Sortie : Ce groupe fournit des options d'affichage pour les résultats de la classification.

- **Tableaux croisés dynamiques :** Les résultats sont affichés dans des tableaux croisés dynamiques.
- **Tableaux et graphiques dans le visualiseur de modèle :** Les résultats sont affichés dans le visualiseur de modèle.
- **Champs d'évaluation :** Calcule les données de cluster pour les variables non utilisées dans la création des clusters. Les champs d'évaluation peuvent être affichés en même temps que les fonctions d'entrée du visualiseur de modèles en les sélectionnant dans la sous-boîte de dialogue Affichage. Les champs avec valeurs manquantes sont ignorés.

Fichier de travail : Ce groupe vous permet d'enregistrer des variables dans le jeu de données actif.

- **Créer une variable d'appartenance à un cluster :** Cette variable contient un numéro d'identification de cluster pour chaque observation. Le nom de cette variable est *tsc_n*, *n* étant un nombre entier positif qui indique l'ordinal de l'opération d'enregistrement du jeu de données actif effectuée par cette procédure au cours d'une session.

Fichiers XML : Le modèle de cluster final et l'arborescence CF représentent deux types de fichiers de sortie pouvant être exportés au format XML.

- **Exporter le modèle final :** Le modèle de cluster final est exporté vers le fichier indiqué au format XML (PMML). Vous pouvez utiliser ce fichier de modèle pour appliquer les informations du modèle aux autres fichiers de données à des fins d'évaluation.
- **Exporter l'arborescence CF :** Cette option vous permet de sauvegarder l'état en cours de l'arborescence des clusters et de la mettre à jour ultérieurement à l'aide de nouvelles données.

Visualiseur de clusters

Les modèles de cluster sont généralement utilisés pour trouver des groupes (ou des clusters) d'enregistrements similaires en fonction des variables examinées, où la similarité entre les membres d'un même groupe est élevée et où la similarité entre les membres de différents groupes est faible. Les résultats peuvent être utilisés pour identifier des associations qui ne seraient pas évidentes autrement. Par exemple, grâce à l'analyse de cluster des préférences des clients, du niveau de revenu et des habitudes d'achat, il peut être possible d'identifier les types de clients les plus susceptibles de répondre à une campagne de marketing particulière.

Il existe deux approches pour interpréter les résultats d'un affichage de cluster :

- Examiner les clusters afin de déterminer les caractéristiques uniques de ce cluster. *Est-ce qu'un cluster contient tous les emprunteurs à revenus élevés ? Est-ce que ce cluster contient davantage d'enregistrements que les autres ?*
- Examiner les champs des clusters afin de déterminer comment les valeurs sont distribuées parmi les clusters. *Est-ce que le niveau d'éducation détermine l'appartenance à un cluster ? Une cote de solvabilité élevée permet-elle de distinguer l'appartenance à un cluster spécifique ?*

Grâce à l'utilisation des vues principales et des différentes vues liées dans le visualiseur de clusters, vous pouvez avoir un bon aperçu qui vous aidera à répondre à ces questions.

Pour consulter les informations concernant un modèle de cluster, activez (double-cliquez) sur l'objet visualiseur de modèle dans le visualiseur.

Visualiseur de clusters

Le visualiseur de cluster est constitué de deux panneaux, l'affichage principal à gauche et la vue liée, ou auxiliaire, à droite. Il existe deux vues principales :

- Récapitulatif du modèle (par défaut). Pour plus d'informations, voir «Vue récapitulative du modèle», à la page 232.
- Clusters. Pour plus d'informations, voir «Vue des clusters», à la page 232.

Il existe quatre vues liées/auxiliaires :

- Importance des prédicteurs. Pour plus d'informations, voir «Vue de l'importance des prédicteurs de cluster», à la page 234.
- Taille des clusters (par défaut). Pour plus d'informations, voir «Vue de la taille des clusters», à la page 234.
- Distribution des cellules. Pour plus d'informations, voir «Vue de la distribution des cellules», à la page 234.
- Comparaison des clusters. Pour plus d'informations, voir «Vue de comparaison des clusters», à la page 235.

Vue récapitulative du modèle

La vue récapitulative du modèle affiche un instantané, ou un récapitulatif, du modèle de cluster, y compris une mesure par silhouette de la cohésion et de la séparation des clusters qui est ombrée pour indiquer des résultats faibles, moyens ou bons. Cet instantané vous permet de vérifier rapidement si la qualité est faible, auquel cas vous pouvez décider de revenir au noeud de modélisation afin de corriger les paramètres du modèle de cluster pour obtenir un meilleur résultat.

Les résultats faibles, moyens et bons sont basés sur le travail de Kaufman et Rousseeuw (1990) concernant l'interprétation des structures du cluster. Dans la vue récapitulative du modèle, d'après le classement de Kaufman et Rousseeuw, un bon résultat équivaut à des données qui indiquent une preuve raisonnable ou forte de la structure du cluster, un résultat moyen signifie une preuve faible et un résultat mauvais reflète une absence de preuve significative.

La mesure par silhouette établit la moyenne, de tous les enregistrements, (B A) / max(A,B), où A correspond à la distance de l'enregistrement au centre de son cluster et B correspond à la distance de l'enregistrement au centre du cluster le plus proche auquel il n'appartient pas. Un coefficient de silhouette de 1 signifie que toutes les observations sont situées directement au centre de leurs clusters. Une valeur de –1 signifie que toutes les observations sont situées au centre de cluster d'autres clusters. Une valeur de 0 signifie, en moyenne, que les observations sont équidistantes du centre de leur propre cluster et du centre de l'autre cluster la plus proche.

Le récapitulatif inclut un tableau qui contient les informations suivantes :

- **Algorithme.** L'algorithme de classification utilisé, par exemple "TwoStep".
- **Caractéristiques d'entrée.** Le nombre de champs, ou d'**entrées** ou de **prédicteurs**.
- **Clusters :** Le nombre de clusters dans la solution.

Vue des clusters

La vue des clusters contient une grille présentant les clusters par fonctions, qui inclut les noms des clusters, leurs tailles et les profils de chaque cluster.

Les colonnes de la grille contiennent les informations suivantes :

- **Cluster.** Les nombres de clusters créées par l'algorithme.

- **Libellé.** Tous les libellés appliqués à chaque cluster (celle-ci est vierge par défaut). Double-cliquez dans la cellule pour saisir un libellé qui décrit les contenus de cluster ; par exemple "Acheteurs de voitures de luxe".
- **Description.** Toutes les descriptions du contenu de cluster (celle-ci est vierge par défaut). Double-cliquez dans la cellule pour saisir une description du cluster ; par exemple "professionnels, plus de 55 ans, gagnant plus de 100 000 $".
- **Taille.** La taille de chaque cluster sous la forme d'un pourcentage de l'échantillon général des clusters. Chaque cellule de taille de la grille affiche une barre verticale qui indique le pourcentage de taille au sein du cluster, un pourcentage de taille au format numérique et les effectifs d'observation du cluster.
- **Fonctions :** Les entrées ou prédicteurs individuels, triés par importance générale par défaut. Si des colonnes ont la même taille, elles sont affichées par ordre croissant des numéros de cluster.

  L'importance des fonctions générales est indiquée par la couleur d'ombrage de l'arrière-plan de la cellule ; la fonction la plus importante étant plus sombre et la moins importante n'étant pas ombrée. Un guide situé au-dessus du tableau indique l'importance correspondant à chaque couleur de cellule de fonction.

Lorsque vous passez la souris sur une cellule, le nom complet/le libellé de la fonction et la valeur de l'importance de la cellule s'affichent. De plus amples informations peuvent être affichées selon le type de vue et de fonction. Dans la vue Centres de clusters, cela inclut les statistiques et la valeur de la cellule ; par exemple : "Moyenne : 4,32". Pour les fonctions catégorielles, la cellule affiche le nom de la catégorie la plus fréquente (modale) et son pourcentage.

Dans la vue des clusters, vous pouvez sélectionner plusieurs manières d'afficher les informations des clusters :

- Transposer les clusters et les fonctions. Pour plus d'informations, voir «Transposition des clusters et des fonctions», à la page 233.
- Trier les fonctions. Pour plus d'informations, voir «Tri des fonctions», à la page 233.
- Trier les clusters. Pour plus d'informations, voir «Tri des clusters», à la page 234.
- Sélectionner le contenu des cellules. Pour plus d'informations, voir «Contenu des cellules», à la page 234.

Transposition des clusters et des fonctions

Par défaut, les clusters sont affichés en tant que colonnes et les fonctions sont affichées en tant que lignes. Pour inverser cet affichage, cliquez sur le bouton **Transposer les clusters et les fonctions** à gauche des boutons **Trier les fonctions par**. Par exemple, vous pouvez réaliser ceci lorsque de nombreux clusters sont affichés, afin de réduire le défilement horizontal nécessaire pour visualiser les données.

Tri des fonctions

Le bouton **Trier les fonctions par** vous permet de sélectionner la façon dont les cellules de fonctions sont affichées :

- **Importance générale.** Il s'agit de l'ordre de tri par défaut. Les fonctions sont triées en ordre décroissant de l'importance générale, et l'ordre de tri est le même pour tous les clusters. Si des fonctions ont des valeurs d'importance liées, les fonctions liées sont répertoriées par ordre croissant des noms de fonctions.
- **Importance intra-cluster.** Les fonctions sont triées par rapport à leur importance pour chaque cluster. Si des fonctions ont des valeurs d'importance liées, les fonctions liées sont répertoriées par ordre croissant des noms de fonctions. Lorsque cette option est sélectionnée, l'ordre de tri varie généralement d'un clusters à l'autre.
- **Nom.** Les fonctions sont triées par nom dans l'ordre alphabétique.
- **Ordre des données.** Les fonctions sont triées selon leur ordre dans le jeu de données.

Tri des clusters

Par défaut, les clusters sont triées en ordre de taille décroissante. Les boutons **Trier les clusters par** vous permettent de les trier par nom dans l'ordre alphabétique, ou, si vous avez créé des libellés uniques, par ordre alphabétique des libellés.

Les fonctions ayant le même libellé sont triées selon le nom de cluster. Si des clusters sont triées par libellé et que vous modifiez le libellé d'un cluster, l'ordre de tri est automatiquement mis à jour.

Contenu des cellules

Les boutons **Cellules** vous permettent de modifier l'affichage du contenu des cellules pour les champs de fonctions et d'évaluation.

- **Centre de cluster :** Par défaut, les cellules affichent les noms/libellés des fonctions et la tendance centrale pour chaque combinaison de cluster/fonction. La moyenne est affichée pour des champs continus et le mode (de la catégorie se présentant le plus fréquemment) avec le pourcentage de catégorie des champs catégoriels.
- **Distributions absolues :** Affiche des noms/libellés de fonctions et des distributions absolues des fonctions dans chaque cluster. Pour les fonctions catégorielles, l'affichage montre des graphiques à barres où sont superposées des catégories en ordre croissant de la valeur des données. Pour les fonctions continues, l'affichage montre un tracé de densité lissé qui utilise les mêmes points finaux et intervalles pour chaque cluster.

  L'affichage en rouge uni montre la distribution des clusters, alors qu'un affichage plus pâle représente les données générales.

- **Distributions relatives :** Affiche les noms/libellés des fonctions et les distributions relatives dans les cellules. En général, les affichages sont similaires à ceux des distributions absolues, excepté les distributions relatives qui sont affichées à la place.

  L'affichage en rouge uni montre la distribution des clusters, alors qu'un affichage plus pâle représente les données générales.

- **Vue de base :** Là où il y a beaucoup de clusters, il peut être difficile de distinguer tous les détails sans procéder à un défilement. Afin de réduire le défilement, sélectionnez cette vue pour modifier l'affichage en une version plus compacte du tableau.

Vue de l'importance des prédicteurs de cluster

La vue de l'importance des prédicteurs affiche l'importance relative de chaque champ dans l'estimation du modèle.

Vue de la taille des clusters

La vue de la taille des clusters affiche un graphique circulaire qui contient chaque cluster. La taille en pourcentage de chaque cluster est affichée sur chaque tranche ; passez la souris sur chaque tranche pour afficher l'effectif de celle-ci.

En dessous du graphique, un tableau répertorie les informations suivantes sur la taille :

- La taille du cluster le plus petit (en effectif et en pourcentage de l'ensemble).
- La taille du cluster le plus grand (en effectif et en pourcentage de l'ensemble).
- Le rapport de taille du cluster le plus grand par rapport au cluster le plus petit.

Vue de la distribution des cellules

La vue de la distribution des cellules affiche un tracé étendu et plus détaillé de la distribution des données pour toutes les cellules de fonction que vous sélectionnez dans le tableau du panneau principal des clusters.

Vue de comparaison des clusters

La vue de comparaison des clusters se compose d'une présentation sous forme de grille avec des fonctions dans les lignes et des clusters sélectionnés dans les colonnes. Cette vue vous aide à mieux comprendre les facteurs qui composent les clusters ; elle vous permet aussi de voir les différences entre les clusters non seulement par comparaison avec les données générales mais aussi en comparant les classes les unes aux autres.

Pour sélectionner des clusters à afficher, cliquez en haut de la colonne de cluster sur le panneau principal des clusters. Cliquez en maintenant la touche Ctrl ou Maj enfoncée pour sélectionner ou désélectionner plus d'un cluster pour la comparaison.

Remarque : Vous pouvez sélectionner jusqu'à cinq clusters à afficher.

Les clusters sont affichés dans l'ordre où ils ont été sélectionnés, alors que l'ordre des champs est déterminé par l'option **Trier les fonctions par**. Lorsque vous sélectionnez **Importance intra-cluster**, les champs sont toujours triés par ordre d'importance générale.

Les tracés d'arrière-plan affichent les distributions générales de chaque fonction :

- Les fonctions catégorielles sont affichées sous forme de tracés de points, où la taille des points indique la catégorie la plus fréquente/modale pour chaque cluster (par fonction).
- Les fonctions continues seront affichées sous forme de boîtes à moustaches, qui affichent les médianes générales et les plages interquartiles.

Des boîtes à moustaches des clusters sélectionnés recouvrent ces vues d'arrière-plan :

- Pour les fonctions continues, des marqueurs en points carrés et des lignes horizontales indiquent la médiane et la plage interquartile de chaque cluster.
- Chaque cluster est représenté par une couleur différente, affichée en haut de la vue.

Navigation dans le visualiseur de cluster

Le visualiseur de cluster est un affichage interactif. Vous pouvez :

- Sélectionner un champ ou un cluster pour afficher plus de détails.
- comparer des clusters afin de sélectionner des éléments intéressants ;
- modifier l'affichage ;
- Transposer des axes.

Utilisation des barres d'outils

Vous contrôlez les informations affichées dans les panneaux de gauche et de droite à l'aide des options des barres d'outils. Vous pouvez modifier l'orientation de l'affichage (haut-bas, gauche-droite ou droite-gauche) à l'aide des contrôles des barres d'outils. En outre, vous pouvez aussi réinitialiser le visualiseur à ses paramètres par défaut et ouvrir une boîte de dialogue pour spécifier le contenu de la vue des clusters dans le panneau principal.

Les options **Trier les fonctions par**, **Trier les clusters par**, **Cellules** et **Affichage** ne sont disponibles que lorsque vous sélectionnez la vue **Clusters** du panneau principal. Pour plus d'informations, voir «Vue des clusters», à la page 232.

Tableau 4. Icônes de la barre d'outils

Icône	Rubrique
	Voir Transposer les clusters et les fonctions.
	Voir Trier les fonctions par.
	Voir Trier les clusters par.

Tableau 4. Icônes de la barre d'outils (suite)	
Icône	**Rubrique**
	Voir Cellules.

Contrôler l'affichage de la vue des clusters

Pour contrôler ce qui est affiché dans la vue des clusters sur le panneau principal, cliquez sur le bouton **Afficher**. La boîte de dialogue Afficher s'ouvre.

Fonctions : Sélectionné par défaut. Pour masquer toutes les fonctions entrées, décochez la case.

Champs d'évaluation. Sélectionnez les champs d'évaluation à afficher (les champs qui ne sont pas utilisés pour créer le modèle de cluster, mais envoyés au visualiseur de modèle pour évaluer les clusters) ; par défaut, aucun n'est affiché. *Remarque :* Le champ d'évaluation doit être une chaîne comportant plusieurs valeurs. Cette case à cocher n'est pas disponible si aucun champ d'évaluation n'est disponible.

Descriptions des clusters. Sélectionné par défaut. Pour masquer toutes les cellules de description des clusters, décochez la case.

Tailles des clusters. Sélectionné par défaut. Pour masquer toutes les cellules de taille des clusters, décochez la case.

Nombre maximal de catégories. Spécifiez le nombre maximal de catégorie à afficher dans les graphiques de fonctions catégorielles ; la valeur par défaut est de 20.

Filtrage des enregistrements

Si vous souhaitez en savoir plus sur les observations d'un cluster particulière ou d'un groupe de clusters, vous pouvez sélectionner un sous-ensemble d'enregistrements afin de poursuivre l'analyse en fonction des clusters sélectionnées.

1. Sélectionnez les clusters dans la vue des clusters du visualiseur de cluster. Pour sélectionner plusieurs clusters, cliquez tout en maintenant la touche Ctrl enfoncée.
2. A partir des menus, sélectionnez :

   Générer > **Filtrer les enregistrements...**
3. Entrez le nom d'une variable de filtre. Les enregistrements des clusters sélectionnés reçoivent une valeur de 1 pour ce champ. Tous les autres enregistrements reçoivent une valeur de 0 et sont exclus des analyses ultérieures jusqu'à ce que vous modifiiez le statut du filtre.
4. Cliquez sur **OK**.

Analyse de cluster hiérarchique

L'analyse de cluster hiérarchique tente d'identifier les groupes d'observations (ou de variables) relativement homogènes basées sur des caractéristiques sélectionnées, en utilisant un algorithme qui débute avec chaque observation (ou variable) dans un cluster séparée et qui combine les clusters jusqu'à ce qu'il n'en reste qu'une. Vous pouvez analyser des variables non normées ou vous pouvez choisir parmi un assortiment de transformations standardisées. Les mesures de distance ou de similarité sont générées par la procédure Proximities (Proximités). Les statistiques s'affichent à chaque étape pour vous aider à choisir la meilleure solution.

Exemple. Y a-t-il des groupes identifiables de spectacles télévisuels qui attirent des audiences similaires à l'intérieur de chaque groupe ? Avec une analyse de cluster hiérarchique, vous pouvez reclasser les spectacles télévisuels (observations) en groupes homogènes basées sur les caractéristiques du spectateur. Cette méthode peut être utilisée pour identifier des segments à des fins commerciales. Vous pouvez aussi classer les villes (observations) en groupes homogènes pour permettre la sélection de villes comparables afin de tester diverses stratégies commerciales.

Statistiques : Planning des agglomérations, matrice de distances (ou des similarités) et cluster d'affectation pour une seule solution ou un ensemble de solutions. Tracés : dendrogrammes et tracés en stalactite.

Remarques sur les données de l'analyse de cluster hiérarchique

Données. Les variables peuvent être des données quantitatives, binaires ou d'effectif. L'échelle des variables est un élément important : des différences d'échelle qui peuvent affecter votre (vos) solution(s) en clusters. Si vos variables sont d'échelles très différentes (par exemple, une variable est mesurée en dollars et l'autre est mesurée en années), vous devez envisager de les standardiser (ceci peut être fait automatiquement avec la procédure de l'analyse de cluster hiérarchique).

Tri par observation : Si des distances ex aequo ou des similitudes se présentent dans les données d'entrée ou entre les clusters mis à jour au cours de l'opération de jointure, la solution de cluster qui en résulte risque de dépendre de l'ordre des observations dans le fichier. Vous pouvez obtenir différentes solutions pour lesquelles les observations ont été triées de manière aléatoire, afin de vérifier la stabilité d'une solution donnée.

Hypothèses : Les mesures de distance ou de similarité utilisées doivent convenir aux données analysées (Voir la procédure Proximities (proximités) pour plus de renseignements sur le choix des mesures de distances et de similarité). Vous devez aussi inclure toutes les variables appropriées dans votre analyse. L'omission de variables influentes peut aboutir à une solution erronée. Parce que l'analyse de cluster hiérarchique est une méthode d'exploration, les résultats doivent être considérés comme provisoires tant qu'ils ne sont pas confirmés avec un échantillon indépendant.

Obtenir une analyse de cluster hiérarchique

1. A partir des menus, sélectionnez :

   Analyse > **Classifier** > **Cluster hiérarchique...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Si vous classez des observations, sélectionnez au moins une variable numérique. Si vous classez des variables, sélectionnez au moins trois variables numériques.

Vous avez la possibilité de sélectionner une variable d'identification pour libeller les observations.

Méthode d'analyse de cluster hiérarchique

Méthode d'agrégation : Les choix disponibles sont : la Distance moyenne entre groupes, la Distance moyenne dans les groupes, l'Agrégation suivant le saut minimum, l'Agrégation suivant le diamètre, les centroïdes, la Médiane et la Méthode de Ward.

Mesure : Il permet de spécifier la mesure de distance ou de similarité devant être utilisée pour la classification. Sélectionnez le type de données et la mesure appropriée de distance ou de similarité :

- **Intervalle :** Les choix possibles sont la Distance Euclidienne, le Carré de la distance Euclidienne, le Cosinus, la Corrélation de Pearson, la Distance de Tchebycheff, la Distance de Manhattan (bloc), la Distance de Minkowski, et Autre.
- **Effectifs :** Les choix possibles sont la Distance du khi-carré et la Distance du phi-carré.
- **Binaire :** Les choix possibles sont la Distance Euclidienne, le Carré de la distance Euclidienne, la différence de taille, la Différence de motif, la Variance, la Dispersion, la Forme, l'Indice de Sokal et Michener, la Corrélation phi tétrachorique, le Lambda, le *D* d'Anderberg, Dice, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Lance et Williams, Ochiai, Rogers et Tanimoto, Russel et Rao, Sokal et Sneath 1, Sokal et Sneath 2, Sokal et Sneath 3, Sokal et Sneath 4, Sokal et Sneath 5, le *Y* de Yule, et le *Q* de Yule.

Transformer les valeurs : Vous permet de standardiser les valeurs des données pour les observations ou les valeurs avant le calcul des proximités (non disponible pour les données binaires). Les méthodes de standardisation disponibles sont les scores *z*, la plage -1 à 1, la plage 0 à 1, l'amplitude maximale de 1, la moyenne de 1 et l'écart type de 1.

Transformer les mesures : Vous permet de transformer les valeurs générées par la mesure de distance. Elles sont appliquées après le calcul de la mesure d'indice. Les choix possibles sont Valeurs absolues, Inverser le signe et Rééchelonner entre 0 à 1.

Statistiques de l'analyse de cluster hiérarchique

Planning des agglomérations : Affiche les observations ou les clusters combinés à chaque étape, les distances entre les observations ou les clusters en cours de combinaison, et le dernier niveau de cluster auquel une observation (ou une variable) a rejoint le cluster.

Matrice des distances : Indique les distances ou les similarités entre éléments.

Cluster(s) d'affectation : Affiche le cluster auquel chaque observation appartient lors d'une ou de plusieurs étapes de la combinaison de clusters. Les options disponibles sont Une seule partition, Plusieurs partitions ou Aucune.

Tracés (graphiques) de l'analyse de cluster hiérarchique

Dendrogramme : Affiche un *dendrogramme*. Les dendrogrammes peuvent être utilisés pour évaluer la cohésion des clusters formés et ils fournissent des renseignements sur le nombre approprié de clusters à conserver.

Stalactites : Affiche un *tracé en stalactite*, incluant tous les clusters ou une plage de clusters spécifiée. Les tracés en stalactite affichent des informations sur la façon dont les observations sont regroupées à chaque itération de l'analyse. Orientation vous permet de sélectionner un tracé vertical ou horizontal.

Sauvegarde des nouvelles variables de l'analyse de cluster hiérarchique

Cluster(s) d'affectation : Vous permet de sauvegarder les clusters d'affectation pour une ou plusieurs ou aucune partition(s). Les variables sauvegardées peuvent alors être utilisées pour des analyses ultérieures pour explorer d'autres différences entre groupes.

Fonctions supplémentaires de la syntaxe de commande CLUSTER

La procédure de classification hiérarchique utilise la syntaxe de commande CLUSTER. Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Utiliser plusieurs méthodes de classification dans une seule analyse.
- Lire et analyser une matrice de proximité.
- Ecrire une matrice de proximité à analyser ultérieurement.
- Indiquer des valeurs pour la puissance et la racine dans la mesure de distance personnalisée (Puissance).
- Spécifier les noms des variables enregistrées.

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

Analyse de cluster de nuées dynamiques

Cette procédure cherche à identifier des groupes d'observations relativement homogènes d'après des caractéristiques sélectionnées, au moyen d'un algorithme qui peut traiter de grands nombres d'observations. L'algorithme vous demande toutefois d'indiquer le nombre de clusters. Vous pouvez indiquer les centres de cluster initiaux si vous connaissez cette information. Vous pouvez choisir entre deux méthodes de classement des observations, soit la mise à jour des centres de cluster de façon itérative, soit la classification seule. Vous pouvez enregistrer l'appartenance à un cluster, les informations de distance et les centres de clusters finaux. Vous pouvez éventuellement indiquer une variable dont les valeurs servent à libeller les sorties par observations. Vous pouvez également demander des statistiques *F* d'analyse de variance. Bien que ces statistiques soient opportunistes (la procédure cherche à former des groupes qui diffèrent), la taille relative des statistiques fournit des informations sur la contribution de chaque variable à la séparation des groupes.

Exemple. Quels sont les groupes de programmes de télévision identifiables qui attirent des publics similaires au sein de chaque groupe ? Grâce à l'analyse de *cluster de nuées dynamiques*, vous pouvez classer les programmes de télévision (observations) en *k* groupes homogènes d'après les caractéristiques des téléspectateurs. Cette méthode peut être utilisée pour identifier des segments à des fins commerciales. Vous pouvez aussi classer les villes (observations) en groupes homogènes pour permettre la sélection de villes comparables afin de tester diverses stratégies commerciales.

Statistiques : Solution complète : centres de clusters initiaux, tableau ANOVA. Chaque observation : informations de cluster, distance au centre de cluster.

Considérations de données sur l'analyse de cluster de nuées dynamiques

Données. Les variables doivent être quantitatives au niveau intervalle ou rapport. Si vos variables sont binaires ou sont des effectifs, utilisez la procédure d'analyse de cluster hiérarchique.

Ordre des observations et des centres de cluster initiaux : L'algorithme par défaut permettant de choisir les centres de cluster initiaux varie en fonction du tri par observation. L'option **Utiliser les nouveaux centres** de la boîte de dialogue Itérer rend la solution résultante potentiellement dépendante du tri par observation, quel que soit le mode de sélection des centres de cluster initiaux. Si vous utilisez l'une de ces méthodes, vous pouvez obtenir différentes solutions pour lesquelles les observations ont été triées de manière aléatoire, afin de vérifier la stabilité d'une solution donnée. Si vous indiquez les centres de cluster initiaux et que vous n'utilisez pas l'option **Utiliser les nouveaux centres**, vous évitez tout problème lié au tri par observation. Toutefois, le tri des centres de cluster initiaux risque d'affecter la solution s'il existe des distances ex aequo entre les observations et les centres de cluster. Pour évaluer la stabilité d'une solution donnée, vous pouvez comparer les résultats des analyses pour lesquelles les valeurs des centres initiaux ont été permutées de différentes manières.

Hypothèses : Les distances sont calculées à l'aide de la distance euclidienne simple. Si vous souhaitez utiliser une autre distance ou une mesure de similarité, utilisez la procédure d'analyse de cluster hiérarchique. Il est important de prendre en compte la mise à l'échelle des variables. Si vos variables sont mesurées selon des échelles différentes (une variable est exprimée en dollars par exemple et une autre en années), vos résultats risquent d'être erronés. Dans ces cas, vous pouvez envisager de standardiser vos variables avant d'effectuer l'analyse de cluster de *nuées dynamiques* (cela peut être fait dans la procédure Descriptives). La procédure suppose que vous avez sélectionné le nombre voulu de clusters et que vous avez inclus toutes les variables pertinentes. Si vous avez choisi un nombre de clusters inadéquat ou omis de variables importantes, vos résultats risquent d'être erronés.

Obtenir une analyse de cluster de nuées dynamiques

1. A partir des menus, sélectionnez :

   Analyse > **Classifier** > **Cluster de nuées dynamiques...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez les variables à utiliser dans l'analyse de cluster.
3. Spécifiez le nombre de clusters. Le nombre de clusters doit être au moins de deux et ne doit pas être supérieur au nombre d'observations contenues dans le fichier de données.
4. Sélectionnez soit la méthode **Itérer et classer** soit la méthode **Classer seulement**.
5. Vous avez la possibilité de sélectionner une variable d'identification pour libeller les observations.

Efficacité de l'analyse de cluster de nuées dynamiques

La commande d'analyse de cluster de *nuées dynamiques* est efficace essentiellement parce qu'elle ne calcule pas les distances entre toutes les paires d'observations, comme c'est le cas dans de nombreux algorithmes de classification, y compris celui utilisé par la commande de classification hiérarchique.

Pour plus d'efficacité, prenez un échantillon d'observations et utilisez la méthode **Itérer et classer** pour déterminer les centres de cluster. Sélectionnez **Ecrire les centres finaux dans Fichier**. Ensuite, restaurez la totalité du fichier de données et sélectionnez la méthode**Classer seulement** puis sélectionnez **Lire les fichiers initiaux à partir de** pour classer tout le fichier en utilisant les centres qui sont estimés à partir

de l'échantillon. Vous pouvez écrire et lire depuis un fichier ou un jeu de données. Les jeux de données sont disponibles pour utilisation ultérieure dans la même session mais ne sont pas enregistrés en tant que fichiers sauf si vous le faites explicitement avant la fin de la session. Le nom des jeux de données doit être conforme aux règles de dénomination de variables. Pour plus d'informations, voir .

Itération de l'analyse de cluster de nuées dynamiques

Remarque : Ces options sont disponibles uniquement si vous avez sélectionné la méthode **Itérer et classer** dans la boîte de dialogue Analyse de cluster de nuées dynamiques.

Maximum des itérations : Limite le nombre des itérations dans l'algorithme des *nuées dynamiques*. L'itération s'arrête après ce nombre d'itérations même si le critère de convergence n'est pas satisfait. Ce nombre doit être compris entre 1 et 999.

Pour reproduire l'algorithme utilisé par la commande Quick Cluster antérieure à la version 5.0, définissez **Maximum des itérations** sur 1.

Critère de convergence : Détermine le moment où l'itération s'arrête. Il représente une proportion de la distance minimale entre les centres de cluster initiaux, de sorte qu'il doit être supérieur à 0 mais pas supérieur à 1. Si le critère est égal à 0,02, par exemple, l'itération cesse lorsqu'une itération complète ne déplace aucun des centres de cluster à une distance de plus de 2 % de la plus petite distance entre les centres de cluster initiaux.

Utiliser les nouveaux centres : Vous permet de demander la mise à jour des centres de cluster après l'affectation de chaque observation. Si vous ne sélectionnez pas cette option, les nouveaux centres de cluster seront calculés lorsque toutes les observations auront été affectées.

Enregistrement des analyses de cluster de nuées dynamiques

Vous pouvez enregistrer des informations sur la solution relatives aux nouvelles variables à utiliser dans les analyses ultérieures :

Cluster(s) d'affectation : Crée une nouvelle variable indiquant le cluster d'affectation finale de chaque observation. Les valeurs de la nouvelle variable vont de 1 au nombre de clusters.

Distance au centre de cluster : Crée une nouvelle variable indiquant la distance euclidienne entre chaque nouvelle variable et son centre de classification.

Options d'analyses de cluster de nuées dynamiques

Statistiques : Les statistiques suivantes sont disponibles : Centres de cluster initiaux, Tableau ANOVA et Affectations et distances au centre.

- *Centres de cluster initiaux*. Première estimation des moyennes de variable pour chacun des clusters. Par défaut, le nombre d'observations assez espacées sélectionné dans les données est égal au nombre de clusters. Les centres de cluster initiaux sont utilisés pour une première classification et sont ensuite mis à jour.
- *tableau ANOVA*. Affiche un tableau d'analyse de variance qui inclut les tests F univariés pour chaque variable de classification. Les tests F sont uniquement descriptifs et les probabilités qui en résultent ne doivent pas être interprétées. Le tableau ANOVA n'apparaît pas si toutes les observations sont affectées à un seul cluster.
- *Informations de cluster pour chaque observation*. Affiche pour chaque observation l'affectation de cluster finale et la distance euclidienne entre l'observation et le centre de cluster utilisé pour classer l'observation. Affiche également la distance euclidienne entre les centres de cluster finaux.

Valeurs manquantes : Les options disponibles sont **Exclure toute observation incomplète** ou **Exclure seulement les composantes non valides**.

- **Exclure toute observation incomplète :** Exclut de l'analyse les observations qui ont des valeurs manquantes pour une variable de grappe.

- **Exclure seulement les composantes non valides :** Affecte des observations aux clusters basés sur des distances calculées à partir de toutes les variables n'ayant pas de valeur manquante.

Fonctions supplémentaires de la commande QUICK CLUSTER

La procédure K-Means Cluster utilise la syntaxe de la commande `QUICK CLUSTER` . Le langage de syntaxe de commande vous permet également d'effectuer les actions suivantes :

- Accepter les premières observations *k* comme centres de cluster initiaux, évitant ainsi le passage de données normalement utilisé pour les estimer.
- Spécifier les centres de cluster initiaux directement comme faisant partie de la syntaxe de commande.
- Spécifier les noms des variables enregistrées.

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

Tests non paramétriques

Les tests non paramétriques effectuent des hypothèses minimales concernant la distribution sous-jacente des données. Les tests disponibles dans ces boîtes de dialogue peuvent être groupés en trois grandes catégories basées sur la façon dont les données sont organisées :

- Un test à un échantillon analyse un champ.
- Un test pour échantillons liés compare deux champs ou plus pour le même ensemble d'observations.
- Un test pour échantillons indépendants analyse un champ qui est regroupé par catégories d'un autre champ.

Tests non paramétriques à un échantillon

Les tests non paramétriques à un échantillon identifient les différences dans les champs uniques en utilisant un ou plusieurs tests non paramétriques. Les tests non paramétriques ne supposent pas que vos données suivent la distribution normale.

Quel est votre objectif ? Les objectifs servent à spécifier rapidement des paramètres de test différents mais fréquemment utilisés.

- **Comparer automatiquement les données observées à des données hypothétiques :** Cet objectif applique le test binomial aux champs catégoriels avec seulement deux catégories, le test du khi-carré à tous les autres champs catégoriels et le test de Kolmogorov-Smirnov aux champs continus.
- **Tester le caractère aléatoire de la séquence :** Cet objectif utilise les suites en séquences pour tester la séquence de valeurs de données observées pour le caractère aléatoire.
- **Analyse personnalisée.** Lorsque vous souhaitez modifier manuellement les paramètres du test dans l'onglet Paramètres, sélectionnez cette option. Veuillez noter que ce paramètre est automatiquement sélectionné si vous modifiez ensuite des options dans l'onglet Paramètres qui ne sont pas compatibles avec l'objectif actuellement sélectionné.

Obtention de tests non paramétriques à un échantillon

A partir des menus, sélectionnez :

Analyse > **Tests non paramétriques** > **Un échantillon...**

1. Cliquez sur **Exécuter**.

(En option) Vous avez également ces possibilités :

- Spécifier un objectif dans l'onglet Objectif.
- Spécifier les affectations de champ dans l'onglet Champs.
- Spécifiez les paramètres experts dans l'onglet Paramètres.

Onglet Champs

L'onglet Champs indique les champs à tester.

Utiliser des rôles prédéfinis : Cette option utilise des informations sur des champs existants. Tous les champs avec un rôle prédéfini d'Entrée, Cible ou Les deux seront utilisés comme champs de test. Au moins un champ de test est requis.

Utiliser des affectations de champ personnalisées : Cette option permet de modifier les rôles des champs. Après avoir sélectionné cette option, spécifiez les champs ci-dessous :

- **Champs de test :** Sélectionnez un ou plusieurs champs.

Onglet Paramètres

L'onglet Paramètres contient plusieurs groupes de paramètres différents que vous pouvez modifier pour affiner le traitement des données par l'algorithme. Si vous modifiez les paramètres par défaut et que ces modifications sont incompatibles avec l'objectif actuellement sélectionné, l'onglet Objectif est automatiquement mis à jour pour sélectionner l'option **Personnaliser l'analyse**.

Choisir des tests

Ces paramètres indiquent les tests à effectuer sur les champs spécifiés dans l'onglet Champs.

Choisir automatiquement les tests en fonction des données : Ce paramètre applique le test binomial aux champs catégoriels avec seulement deux catégories valides (non manquantes), le test du khi-deux à tous les autres champs catégoriels et le test de Kolmogorov-Smirnov aux champs continus.

Personnaliser les tests : Ce paramètre permet de choisir des tests spécifiques à exécuter.

- **Comparer la probabilité binaire observée à la probabilité hypothétique (test binomial) :** Le test binomial peut s'appliquer à tous les champs. Cela produit un test à un échantillon qui évalue si la distribution observée d'un champ indicateur (un champ catégoriel avec seulement deux catégories) est semblable à ce qui est attendu d'une distribution binomiale spécifiée. De plus, vous pouvez demander des intervalles de confiance. Pour en savoir plus sur les paramètres de test, voir «Test binomial : Options », à la page 242.
- **Comparer les probabilités observées aux probabilités hypothétiques (test khi-deux) :** Le test du khi-deux s'applique aux champs nominaux et ordinaux. Cela produit un test à un échantillon qui calcule une statistique du khi-deux basée sur les différences entre les fréquences observées et attendues des catégories d'un champ. Pour en savoir plus sur les paramètres de test, voir «Test du Khi-carré : options », à la page 243.
- **Tester la distribution observée par rapport à la distribution hypothétique (test de Kolmogorov-Smirnov) :** Le test de Kolmogorov-Smirnov s'applique aux champs continus et ordinaux. Cela produit un test à un échantillon évaluant si l'échantillon de la fonction de distribution cumulée d'un champ est homogène avec une distribution uniforme, normale, de Poisson ou exponentielle. Pour en savoir plus sur les paramètres de test, voir «Options Kolmogorov-Smirnov », à la page 243.
- **Comparer la médiane à la valeur hypothétique (test de Wilcoxon) :** Le test de Wilcoxon s'applique aux champs continus et ordinaux. Cela produit un test à un échantillon de la valeur de la médiane d'un champ. Spécifier un nombre comme médiane hypothétique.
- **Tester le caractère aléatoire de la séquence (suite en séquence) :** La suite en séquence s'applique à tous les champs. Cela produit un test à un échantillon évaluant si la séquence des valeurs d'un champ dichotomisé est aléatoire. Pour en savoir plus sur les paramètres de test, voir «Options Suites en séquence », à la page 244.

Test binomial : Options

Le test binomial est destiné aux champs indicateurs (champs catégoriels avec seulement deux catégories) mais peut s'appliquer à tous les champs en utilisant des règles pour définir le "succès".

Proportion hypothétique : Indique la proportion attendue d'enregistrements définis comme « réussites » ou *p*. Indiquez une valeur supérieure à 0 et inférieure à 1. La valeur par défaut est 0,5.

Intervalle de confiance. Les méthodes de calculs d'intervalles de confiance pour les données binaires suivantes sont disponibles :

- **Clopper-Pearson (exact) :** Un intervalle exact basé sur la distribution binomiale cumulée.
- **Jeffreys :** Un intervalle bayésien basé sur la distribution postérieure de *p* utilisant la loi a priori de Jeffreys.
- **Rapport de vraisemblance :** Un intervalle basé sur la fonction de vraisemblance pour *p*.

Définir le succès pour des champs catégoriels : Indique comment le "succès, à savoir les valeurs de données testées par rapport à la proportion hypothétique, est défini pour les champs catégoriels.

- **Utiliser la première catégorie trouvée dans les données** effectue le test binomial en utilisant la première valeur trouvée dans l'échantillon pour définir le "succès". Cette option s'applique uniquement aux champs nominaux ou ordinaux avec seulement deux valeurs ; tous les autres champs catégoriels spécifiés dans l'onglet Champs où cette option est utilisée, ne seront pas testés. Il s'agit de la valeur par défaut.
- **Spécifier les valeurs de succès** effectue le test binomial à l'aide d'une liste de valeurs spécifiée pour définir le "succès". Spécifier une liste de chaîne ou des valeurs numériques. Les valeurs de la liste n'ont pas besoin d'être présentes dans l'échantillon.

Définir le succès pour des champs continus : Indique comment le "succès", à savoir les valeurs de données testées para rapport à la valeur de test, est défini pour les champs continus. Le succès est défini comme des valeurs inférieures ou égales à la césure.

- **Centre de l'échantillon** définit la césure à la moyenne des valeurs minimales et maximales.
- **Césure personnalisée** permet de spécifier une valeur de césure.

Test du Khi-carré : options

Toutes les catégories ont la même probabilité : Produit des fréquences égales pour toutes les catégories de l'échantillon. Il s'agit de la valeur par défaut.

Personnaliser la probabilité prévue : Permet de spécifier les fréquences inégales pour une liste de catégories spécifiée. Spécifier une liste de chaîne ou des valeurs numériques. Les valeurs de la liste n'ont pas besoin d'être présentes dans l'échantillon. Dans la colonne **Catégorie** , spécifiez les valeurs des catégories. Dans la colonne **Fréquence relative** , spécifiez une valeur supérieure à 0 pour chaque catégorie. Les fréquences personnalisées sont traitées comme des rapports. Par exemple, spécifier des fréquences 1, 2 et 3 est l'équivalent de spécifier des fréquences 10, 20 et 30 et les deux spécifient que 1/6 des enregistrements doit se trouver dans la première catégorie, 1/3 dans la seconde et 1/2 dans la troisième. Lorsque les probabilités attendues personnalisées sont spécifiées, les valeurs de catégories personnalisées doivent inclure toutes les valeurs de champ dans les données ; sinon, le test n'est pas exécuté pour ce champ.

Options Kolmogorov-Smirnov

Cette boîte de dialogue indique les distributions à tester et les paramètres des distributions hypothétiques.

Si certains paramètres de la distribution doivent être estimés à partir de l'échantillon, le test de Kolmogorov-Smirnov ne s'applique plus. Dans ces cas-là, les statistiques du test de Lilliefors peuvent être utilisées pour estimer la valeur *p* en utilisant l'échantillon de Monte Carlo pour tester la normalité alors que la moyenne et la variance sont inconnues. Le test de Lilliefors s'applique aux trois distributions continues (**Normale**, **Exponentielle** et **Uniforme**). Notez que le test ne s'applique pas si la distribution sous-jacente est discrète (**Poisson**). Le test n'est défini que pour une inférence à un échantillon si les paramètres de distribution correspondants ne sont pas spécifiés.

Normal

Utiliser des données d'échantillon utilise la moyenne observée et l'écart type et propose des options permettant de sélectionner les résultats existants du **Test asymptotique** ou d'utiliser **Test de Lilliefors basé sur l'échantillonnage de Monte Carlo**. **Personnalisé** permet de spécifier des valeurs.

Uniforme
Utiliser des données d'échantillon s'appuie sur le minimum et le maximum observés et utilise le test de Lilliefors basé sur l'échantillonnage de Monte Carlo. **Personnalisé** permet de spécifier les valeurs minimale et maximale.

Exponentiel
Moyenne d'échantillon s'appuie sur la moyenne observée et utilise le test de Lilliefors basé sur l'échantillonnage de Monte Carlo. **Personnalisé** permet de spécifier une valeur moyenne observée.

Poisson
Moyenne permet de spécifier une valeur moyenne observée.

Options Suites en séquence

Les suites en séquence sont destinées aux champs indicateurs (champs catégoriels avec seulement deux catégories) mais peuvent s'appliquer à tous les champs en utilisant des règles pour définir les groupes.

Définir des groupes pour des champs catégoriels : Les options suivantes sont disponibles :

- **Il n'existe que 2 catégories dans l'échantillon** effectue les suites en séquence en utilisant des valeurs trouvées dans l'échantillon pour définir les groupes. Cette option s'applique uniquement aux champs nominaux ou ordinaux avec seulement deux valeurs ; tous les autres champs catégoriels spécifiés dans l'onglet Champs où cette option est utilisée, ne seront pas testés.
- **Recoder les données en 2 catégories** effectue les suites en séquence à l'aide de la liste de valeurs spécifiée pour définir un des groupes. Toutes les autres valeurs de l'échantillon définissent l'autre groupe. Les valeurs de la liste n'ont pas besoin d'être présentes dans l'échantillon, mais au moins un enregistrement doit se trouver dans chaque groupe.

Définir une césure pour des champs continus : Indique la façon dont les groupes sont définis pour les champs continus. Le premier groupe est défini comme comprenant des valeurs inférieures ou égales à la césure.

- **Médiane d'échantillon** définit la césure à la médiane d'échantillon.
- **Moyenne d'échantillon** définit la césure à la moyenne d'échantillon.
- **Personnalisé** permet de spécifier une valeur de césure.

Options de test

Niveau de signification
Indique le niveau d'importance (alpha) pour tous les tests. Spécifiez une valeur numérique comprise entre 0 et 1. La valeur par défaut est 0.05.

Intervalle de confiance (%)
Indique le niveau de confiance pour tous les intervalles de confiance générés. Indiquez une valeur numérique comprise entre 0 et 100. La valeur par défaut est 95.

Observations exclues
Indique comment déterminer la base des observations pour les tests.

Exclure les observations test par test
Les enregistrements qui comportent des valeurs manquantes pour un champ utilisé pour un test spécifique sont omis du test. Lorsque plusieurs tests sont spécifiés dans l'analyse, chaque test est évalué séparément.

Exclure toute observation incomplète
Les enregistrements comportant des valeurs manquantes pour un champ quelconque nommé dans l'onglet Champs sont exclus de toutes les analyses.

Echantillonnage de Monte Carlo

Si certains paramètres de la distribution doivent être estimés à partir de l'échantillon, le test de Kolmogorov-Smirnov ne s'applique plus. Dans ces cas-là, les statistiques du test de Lilliefors peuvent être utilisées pour estimer la valeur *p* en utilisant l'échantillon de Monte Carlo pour tester la normalité alors que la moyenne et la variance sont inconnues. Le test de Lilliefors s'applique aux trois distributions continues (**Normale**, **Exponentielle** et **Uniforme**). Notez que le test ne s'applique

pas si la distribution sous-jacente est discrète (**Poisson**). Le test n'est défini que pour une inférence à un échantillon si les paramètres de distribution correspondants ne sont pas spécifiés.

Définir une valeur de départ personnalisée
Si ce paramètre est activé, il permet de réinitialiser la **Valeur de départ** aléatoire utilisée pour l'échantillonnage de Monte Carlo. Cette valeur doit être un entier unique compris entre 1 et 2 147 483 647 (la valeur par défaut est 2 000 000). Par défaut, ce paramètre n'est pas activé, ce qui génère une valeur de départ aléatoire.

Nombre d'échantillons
Réinitialise le nombre de réplicats de l'échantillonnage de Monte Carlo utilisés par le test de Lilliefors. Cette valeur doit être un entier unique compris entre 100 et l'entier le plus grand. La valeur par défaut est 10 000.

Niveau de confiance de simulation (%)
Réinitialise le niveau d'intervalle de confiance estimé du test de Kolmogorov-Smirnov. Il doit s'agir d'une valeur unique comprise entre 0 et 100. La valeur par défaut est 99.

Valeurs manquantes de l'utilisateur

Valeurs manquantes de l'utilisateur pour les champs catégoriels : Les champs catégoriels doivent avoir des valeurs valides pour qu'un enregistrement puisse être inclus dans l'analyse. Ces contrôles vous permettent d'indiquer si les valeurs manquantes de l'utilisateur sont considérées comme valides parmi les champs catégoriels. Les valeurs système manquantes et les valeurs manquantes pour les champs continus sont toujours considérées comme non valides.

Fonctions supplémentaires de la commande NPTESTS

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

- spécifier des tests à un échantillon, pour échantillon indépendants et pour échantillons liés en n'exécutant la procédure qu'une seule fois.

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

Test de Wilcoxon pour échantillon unique

Le test de Wilcoxon est un test statistique non paramétrique qui est utilisé pour tester l'emplacement d'une population sur la base de données d'échantillon ou pour comparer les emplacements de deux populations. Le test de rang signé de Wilcoxon à un échantillon est une alternative non paramétrique au test t de l'échantillon lorsque les données ne suivent pas une distribution normale. Il permet de déterminer s'il existe une différence significative entre la médiane d'un échantillon et une valeur médiane hypothétique. Ici, les statistiques de test sont calculées en fonction des rangs de la différence entre les valeurs observées et la valeur hypothétique.

Le test de rang signé de Wilcoxon d'un échantillon suppose que la population parent est symétrique par rapport à la médiane. Comme les échantillons aléatoires proviennent d'une population ayant une fonction de distribution cumulative continue et sont symétriques par rapport à la médiane, sous l'hypothèse nulle, les différences entre les valeurs observées et les valeurs hypothétiques sont distribuées symétriquement autour de zéro de sorte que les différences positives et négatives de magnitude absolue égale ont la même probabilité d'occurrence.

Exemple

Supposons qu'une nouvelle société de démarrage soit intéressée par l'embauche d'ingénieurs logiciels mais qu'elle souhaite s'assurer qu'elle offre des salaires compétitifs. Supposons qu'ils ont collecté des données sur les salaires auprès d'un échantillon de 30 ingénieurs en logiciel qui sont actuellement employés dans diverses entreprises de la même industrie. Si la société souhaite déterminer si le salaire médian qu'elle offre à ses employés potentiels est sensiblement différent du salaire médian des ingénieurs en logiciel de l'industrie, un échantillon de rang de Wilcoxon peut être utilisé à cette fin. Ici, nous supposons l'hypothèse nulle que le salaire médian offert par la startup n'est pas significativement différent du salaire médian dans l'industrie, c'est-à-dire que la différence de salaire médian est nulle.

Pour exécuter le test de Wilcoxon à un échantillon

(Cette fonction requiert l'option Statistiques de base.)

Dans le menu, sélectionnez:

1. **Analyse > Tests non paramétriques > Un échantillon ...**
2. Cliquez sur **'Paramètres'**.
3. Sous **'Sélectionner un élément'**, sélectionnez **'Sélectionner des tests'**.
4. Sélectionnez **'Tests personnalisés'**.
5. Sélectionnez **'Comparer la médiane à l'hypothèse (test de Wilcoxon)'**.
6. Entrez la valeur de la **'valeur médiane hypothétisée'** et cliquez sur **'Exécuter'**.

Pour plus d'informations, voir https://www.ibm.com/support/pages/can-spss-statistics-produce-one-sample-wilcoxon-test

tests non paramétriques pour échantillons indépendants

Les tests non paramétriques pour échantillons indépendants identifient les différences entre deux groupes ou plus à l'aide d'un ou de plusieurs tests non paramétriques. Les tests non paramétriques ne supposent pas que vos données suivent la distribution normale.

Quel est votre objectif ? Les objectifs servent à spécifier rapidement des paramètres de test différents mais fréquemment utilisés.

- **Comparer automatiquement les distributions entre les groupes :** Cet objectif applique le test U de Mann-Whitney aux données avec 2 groupes ou le Kruskal-Wallis ANOVA à un facteur aux données avec *k* groupes.
- **Comparer les médianes entre les groupes :** Cet objectif utilise le test de la médiane pour comparer les médianes observées entre les groupes.
- **Analyse personnalisée.** Lorsque vous souhaitez modifier manuellement les paramètres du test dans l'onglet Paramètres, sélectionnez cette option. Veuillez noter que ce paramètre est automatiquement sélectionné si vous modifiez ensuite des options dans l'onglet Paramètres qui ne sont pas compatibles avec l'objectif actuellement sélectionné.

Obtention des tests non paramétriques pour échantillons indépendants

A partir des menus, sélectionnez :

Analyse > **Tests non paramétriques** > **Echantillons indépendants...**

1. Cliquez sur **Exécuter**.

(En option) Vous avez également ces possibilités :

- Spécifier un objectif dans l'onglet Objectif.
- Spécifier les affectations de champ dans l'onglet Champs.
- Spécifier les paramètres d'expert dans l'onglet Paramètres.

Onglet Champs

L'onglet Champs indique les champs à tester et le champ utilisé pour définir les groupes.

Utiliser des rôles prédéfinis : Cette option utilise des informations sur des champs existants. Tous les champs continus et ordinaux avec un rôle prédéfini de Cible ou Les deux seront utilisés comme champs de test. Si un champ unique catégoriel avec un rôle prédéfini d'Entrée est disponible, il sera utilisé comme champ de regroupement. Sinon, il n'y aura pas d'utilisation par défaut de champ de regroupement et vous devrez utiliser des affectations de champs personnalisées. Au moins un champ de test et un champ de regroupement sont requis.

Utiliser des affectations de champ personnalisées : Cette option permet de modifier les rôles des champs. Après avoir sélectionné cette option, spécifiez les champs ci-dessous :

- **Champs de test :** Sélectionnez un ou plusieurs champs continus ou ordinaux.
- **Groupes :** Sélectionnez un champ catégoriel.

Onglet Paramètres

L'onglet Paramètres contient plusieurs groupes de paramètres différents que vous pouvez modifier pour affiner le traitement des données par l'algorithme. Si vous modifiez les paramètres par défaut et que ces modifications sont incompatibles avec l'objectif actuellement sélectionné, l'onglet Objectif est automatiquement mis à jour pour sélectionner l'option **Personnaliser l'analyse**.

Choisir des tests

Ces paramètres indiquent les tests à effectuer sur les champs spécifiés dans l'onglet Champs.

Choisir automatiquement les tests en fonction des données : Ce paramètre applique le test U de Mann-Whitney aux données avec 2 groupes ou le Kruskal-Wallis ANOVA à un facteur aux données avec *k* groupes.

Personnaliser les tests : Ce paramètre permet de choisir des tests spécifiques à exécuter.

- **Comparer les distributions entre les groupes :** Ces paramètres produisent des tests pour échantillons indépendants indiquant si les échantillons sont issus de la même population.

  Mann-Whitney U (2 échantillons) utilise le rang de chaque observation pour tester si les groupes sont issus de la même population. La première valeur dans l'ordre croissant du champ de regroupement définit le premier groupe et la deuxième valeur définit le deuxième groupe. Si le champ de regroupement a plus de deux valeurs, ce test n'est pas généré.

  Kolmogorov-Smirnov (2 échantillons) est sensible aux différences de médiane, de dispersion, d'asymétrie, etc. entre les deux distributions. Si le champ de regroupement a plus de deux valeurs, ce test n'est pas généré.

  Tester le caractère aléatoire de la séquence (Wald-Wolfowitz pour 2 échantillons) génère des suites en séquence avec l'appartenance au groupe comme critère. Si le champ de regroupement a plus de deux valeurs, ce test n'est pas généré.

  Kruskal-Wallis ANOVA à un facteur (k échantillons) est une extension du test U de Mann-Whitney et l'équivalent non paramétrique de l'analyse de variance à un facteur. En option, vous pouvez demander des comparaisons multiples des *k* échantillons, soit toutes les comparaisons multiples **par paire** soit les comparaisons **étape par étape descendantes** .

  Tester les possibilités ordonnées (Jonckheere-Terpstra pour k échantillons) est une alternative à Kruskal-Wallis plus puissante lorsque les *k* échantillons ont un ordre naturel. Par exemple, les *k* populations peuvent représenter *k* températures croissantes. L'hypothèse selon laquelle différentes températures produisent la même distribution des réponses est testée contre l'hypothèse alternative selon laquelle l'accroissement de température fait augmenter l'ampleur de la réponse. Ici, l'hypothèse alternative est ordonnée ; le test de Jonckheere-Terpstra est donc le plus approprié. **Du plus petit au plus grand** spécifie l'hypothèse alternative suivante : le paramètre d'emplacement du premier groupe est inférieur ou égal au deuxième, lui-même inférieur ou égal au troisième, et ainsi de suite. **Du plus grand au plus petit** spécifie l'hypothèse alternative suivante : le paramètre d'emplacement du premier groupe est supérieur ou égal au deuxième, lui-même supérieur ou égal au troisième, et ainsi de suite. Pour ces deux options, l'hypothèse alternative suppose aussi que les emplacements ne sont pas tous égaux. En option, vous pouvez demander des comparaisons multiples des *k* échantillons, soit toutes les comparaisons multiples **par paire** soit les comparaisons **étape par étape descendantes** .
- **Comparer les plages entre les groupes :** Cela génère des tests pour échantillons indépendants indiquant si les échantillons ont la même plage. La **Réaction extrême de Moses (2 échantillons)** teste un groupe de contrôles par rapport à un groupe de comparaisons. La première valeur dans l'ordre croissant du champ de regroupement définit le groupe de contrôles et la deuxième valeur définit le groupe de comparaisons. Si le champ de regroupement a plus de deux valeurs, ce test n'est pas généré.

- **Comparer les médianes entre les groupes :** Cela génère des tests pour échantillons indépendants indiquant si les échantillons ont la même médiane. Le **Test de la médiane (k échantillons)** peut utiliser soit la médiane d'échantillon regroupé en pool (calculée à partir de tous les enregistrements du jeu de données) ou une valeur personnalisée comme la médiane hypothétique. En option, vous pouvez demander des comparaisons multiples des *k* échantillons, soit toutes les comparaisons multiples **par paire** soit les comparaisons **étape par étape descendantes** .
- **Estimer les intervalles de confiance entre les groupes :** L'**estimation de Hodges-Lehman (2 échantillons)** génère une estimation d'échantillons indépendants et un intervalle de confiance pour la différence entre les médianes des deux groupes. Si le champ de regroupement a plus de deux valeurs, ce test n'est pas généré.

Options de test

Niveau de signification : Indique le niveau d'importance (alpha) pour tous les tests. Spécifiez une valeur numérique comprise entre 0 et 1. La valeur par défaut est 0.05.

Intervalle de confiance (%). Indique le niveau de fiabilité pour tous les intervalles de confiance générés. Indiquez une valeur numérique comprise entre 0 et 100. La valeur par défaut est 95.

Observations exclues : Indique comment déterminer la base des observations pour les tests. **Exclure toute observation incomplète** signifie que les enregistrements avec des valeurs manquantes dans les champs nommés dans une sous-commande sont exclus de toutes les analyses. **Exclure les observations test par test** signifie que les enregistrements avec des valeurs manquantes dans un champ utilisé pour un test spécifique sont ignorés pendant ce test. Lorsque plusieurs tests sont spécifiés dans l'analyse, chaque test est évalué séparément.

Valeurs manquantes de l'utilisateur

Valeurs manquantes de l'utilisateur pour les champs catégoriels : Les champs catégoriels doivent avoir des valeurs valides pour qu'un enregistrement puisse être inclus dans l'analyse. Ces contrôles vous permettent d'indiquer si les valeurs manquantes de l'utilisateur sont considérées comme valides parmi les champs catégoriels. Les valeurs système manquantes et les valeurs manquantes pour les champs continus sont toujours considérées comme non valides.

Fonctions supplémentaires de la commande NPTESTS

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

- spécifier des tests à un échantillon, pour échantillon indépendants et pour échantillons liés en n'exécutant la procédure qu'une seule fois.

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

Tests non paramétriques pour échantillons liés

Identifie les différences entre deux champs liés ou plus à l'aide d'un ou de plusieurs tests non paramétriques. Les tests non paramétriques ne supposent pas que vos données suivent la distribution normale.

Analyse des données : Chaque enregistrement correspond à un sujet donné pour lequel deux mesures associées ou plus sont stockées dans des champs distincts du jeu de données. Par exemple, une étude concernant l'efficacité d'un régime peut être analysée à l'aide de tests d'échantillons liés non paramétriques, si la pondération de chaque sujet est mesurée à intervalles réguliers et stockée dans des champs tels que *Pondération avant le régime*, *Pondération intermédiaire*, et *Pondération après le régime*. Ces champs sont "liés".

Quel est votre objectif ? Les objectifs servent à spécifier rapidement des paramètres de test différents mais fréquemment utilisés.

- **Comparer automatiquement les données observées à des données hypothétiques :** Cet objectif applique le test de McNemar aux donnés catégorielles lorsque 2 champs sont spécifiés, le Q de Cochran aux données catégorielles lorsque plus de 2 champs sont spécifiés, le test de Wilcoxon en séries

appariées aux données continues lorsque 2 champs sont spécifiés et le test de Friedman ANOVA à deux facteurs par classement aux données continues lorsque plus de 2 champs sont spécifiés.

- **Analyse personnalisée.** Lorsque vous souhaitez modifier manuellement les paramètres du test dans l'onglet Paramètres, sélectionnez cette option. Veuillez noter que ce paramètre est automatiquement sélectionné si vous modifiez ensuite des options dans l'onglet Paramètres qui ne sont pas compatibles avec l'objectif actuellement sélectionné.

Lorsque des champs de niveaux de mesure différents sont spécifiés, ils sont séparés en fonction de leur niveau de mesure puis le test approprié est appliqué à chaque groupe. Par exemple, si vous choisissez **Comparer automatiquement les données observées à des données hypothétiques** comme objectif et spécifiez 3 champs continus et 2 champs nominaux, le test de Friedman est appliqué aux champs continus et le test de McNemar est appliqué aux champs nominaux.

Obtention des tests non paramétriques pour échantillons liés

A partir des menus, sélectionnez :

Analyse > **Tests non paramétriques** > **Echantillons liés...**

1. Cliquez sur **Exécuter**.

(En option) Vous avez également ces possibilités :

- Spécifier un objectif dans l'onglet Objectif.
- Spécifier les affectations de champ dans l'onglet Champs.
- Spécifier les paramètres d'expert dans l'onglet Paramètres.

Onglet Champs

L'onglet Champs indique les champs à tester.

Utiliser des rôles prédéfinis : Cette option utilise des informations sur des champs existants. Tous les champs avec un rôle prédéfini de Cible ou Les deux seront utilisés comme champs de test. Au moins deux champs de test sont requis.

Utiliser des affectations de champ personnalisées : Cette option permet de modifier les rôles des champs. Après avoir sélectionné cette option, spécifiez les champs ci-dessous :

- **Champs de test :** Sélectionnez deux ou plusieurs champs. Chaque champ correspond à un échantillon lié séparé.

Onglet Paramètres

L'onglet Paramètres contient plusieurs groupes de paramètres différents que vous pouvez modifier pour affiner le traitement des données par la procédure. Si vous modifiez les paramètres par défaut et que ces modifications sont incompatibles avec les autres objectifs, l'onglet Objectif est automatiquement mis à jour pour sélectionner l'option **Personnaliser l'analyse**.

Choisir des tests

Ces paramètres indiquent les tests à effectuer sur les champs spécifiés dans l'onglet Champs.

Choisir automatiquement les tests en fonction des données : Ce paramètre applique le test de McNemar aux donnés catégorielles lorsque 2 champs sont spécifiés, le Q de Cochran aux données catégorielles lorsque plus de 2 champs sont spécifiés, le test de Wilcoxon en séries appariées aux données continues lorsque 2 champs sont spécifiés et le test de Friedman ANOVA à deux facteurs par classement aux données continues lorsque plus de 2 champs sont spécifiés.

Personnaliser les tests : Ce paramètre permet de choisir des tests spécifiques à exécuter.

- **Test de modification des données binaires. Test de McNemar (2 échantillons)** peut être appliqué à des champs catégoriels. Cela produit un test pour échantillons liés évaluant si les combinaisons de valeurs entre deux champs indicateurs (champs catégoriels avec seulement deux valeurs) sont

aussi probables l'une que l'autre. S'il existe plus de deux champs spécifiés dans l'onglet Champs, ce test n'est pas effectué. Pour en savoir plus sur les paramètres de test, voir «Test de McNemar : Définir le succès », à la page 250. Le **Q de Cochran (k échantillons)** peut être appliqué aux champs catégoriels. Cela produit un test pour échantillons liés évaluant si les combinaisons de valeurs entre *k* champs indicateurs (champs catégoriels avec seulement deux valeurs) sont aussi probables l'une que l'autre. En option, vous pouvez demander des comparaisons multiples des *k* échantillons, soit toutes les comparaisons multiples **par paire** soit les comparaisons **étape par étape descendantes** . Pour en savoir plus sur les paramètres de test, voir «Q de Cochran : Définir le succès », à la page 250.

- **Test de modification des données multinomiales. Test d'homogénéité marginale (2 échantillons)** produit un test d'échantillons liés pour déterminer si les combinaisons de valeurs entre deux champs ordinaux appariées sont également probables. Le test d'homogénéité marginale est généralement utilisé dans les cas avec des mesures répétées. Ce test est un développement du test de McNemar d'une réponse binaire à une réponse multinomiale. S'il existe plus de deux champs spécifiés dans l'onglet Champs, ce test n'est pas effectué.
- **Comparer la différence de médiane à la différence de médiane hypothétique :** Ces tests produisent chacun un test d'échantillons liés pour déterminer si la différence médiane entre deux champs est différente de 0. Le test s'applique aux champs ordinaux et continus. S'il y a plus de deux champs spécifiés dans l'onglet Champs, ces tests ne sont pas effectués.
- **Estimer l'intervalle de confiance :** Cela génère une estimation et un intervalle de confiance d'échantillons liés pour la différence de médiane entre deux champs appariés. Le test s'applique aux champs ordinaux et continus. S'il y a plus de deux champs spécifiés dans l'onglet Champs, ce test n'est pas effectué.
- **Quantifier les associations :** Le **coefficient de concordance de Kendall (k échantillons)** génère une mesure d'accord entre les juges ou les observateurs, où chaque enregistrement est le classement par un juge de plusieurs éléments (champs). En option, vous pouvez demander des comparaisons multiples des *k* échantillons, soit toutes les comparaisons multiples **par paire** soit les comparaisons **étape par étape descendantes** .
- **Comparer les distributions. ANOVA bidirectionnel par rangs de Friedman (échantillons k)** génère un test d'échantillons liés pour déterminer si des échantillons liés à *K* ont été tirés de la même population. En option, vous pouvez demander des comparaisons multiples des *k* échantillons, soit toutes les comparaisons multiples **par paire** soit les comparaisons **étape par étape descendantes** .

Test de McNemar : Définir le succès

Le test de McNemar concerne les champs indicateurs (champs catégoriels avec seulement deux catégories), mais peut s'appliquer à tous les champs catégoriels en utilisant des règles de définition du "succès".

Définir le succès pour des champs catégoriels : Indique la procédure de définition du "succès" pour les champs catégoriels.

- **Utiliser la première catégorie trouvée dans les données** effectue le test en utilisant la première valeur trouvée dans l'échantillon pour définir le "succès". Cette option s'applique uniquement aux champs nominaux ou ordinaux avec seulement deux valeurs ; tous les autres champs catégoriels spécifiés dans l'onglet Champs où cette option est utilisée, ne seront pas testés. Il s'agit de la valeur par défaut.
- **Spécifier les valeurs de succès** effectue le test à l'aide d'une liste de valeurs spécifiée pour définir le "succès". Spécifier une liste de chaîne ou des valeurs numériques. Les valeurs de la liste n'ont pas besoin d'être présentes dans l'échantillon.

Q de Cochran : Définir le succès

Le test Q de Cochran concerne les champs indicateurs (champs catégoriels avec seulement deux catégories), mais peut s'appliquer à tous les champs catégoriels en utilisant des règles de définition du "succès".

Définir le succès pour des champs catégoriels : Indique la procédure de définition du "succès" pour les champs catégoriels.

- **Utiliser la première catégorie trouvée dans les données** effectue le test en utilisant la première valeur trouvée dans l'échantillon pour définir le "succès". Cette option s'applique uniquement aux champs nominaux ou ordinaux avec seulement deux valeurs ; tous les autres champs catégoriels spécifiés dans l'onglet Champs où cette option est utilisée, ne seront pas testés. Il s'agit de la valeur par défaut.
- **Spécifier les valeurs de succès** effectue le test à l'aide d'une liste de valeurs spécifiée pour définir le "succès". Spécifier une liste de chaîne ou des valeurs numériques. Les valeurs de la liste n'ont pas besoin d'être présentes dans l'échantillon.

Options de test

Niveau de signification : Indique le niveau d'importance (alpha) pour tous les tests. Spécifiez une valeur numérique comprise entre 0 et 1. La valeur par défaut est 0.05.

Intervalle de confiance (%). Indique le niveau de fiabilité pour tous les intervalles de confiance générés. Indiquez une valeur numérique comprise entre 0 et 100. La valeur par défaut est 95.

Observations exclues : Indique comment déterminer la base des observations pour les tests.

- **Exclure toute observation incomplète** signifie que les enregistrements avec des valeurs manquantes dans les champs nommés dans une sous-commande sont exclus de toutes les analyses.
- **Exclure les observations test par test** signifie que les enregistrements avec des valeurs manquantes dans un champ utilisé pour un test spécifique sont ignorés pendant ce test. Lorsque plusieurs tests sont spécifiés dans l'analyse, chaque test est évalué séparément.

Valeurs manquantes de l'utilisateur

Valeurs manquantes de l'utilisateur pour les champs catégoriels : Les champs catégoriels doivent avoir des valeurs valides pour qu'un enregistrement puisse être inclus dans l'analyse. Ces contrôles vous permettent d'indiquer si les valeurs manquantes de l'utilisateur sont considérées comme valides parmi les champs catégoriels. Les valeurs système manquantes et les valeurs manquantes pour les champs continus sont toujours considérées comme non valides.

Fonctions supplémentaires de la commande NPTESTS

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

- spécifier des tests à un échantillon, pour échantillon indépendants et pour échantillons liés en n'exécutant la procédure qu'une seule fois.

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

Fonctions supplémentaires de la commande NPTESTS

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

- spécifier des tests à un échantillon, pour échantillon indépendants et pour échantillons liés en n'exécutant la procédure qu'une seule fois.

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

Boîtes de dialogue ancienne version

Il existe un certain nombre de boîtes de dialogue "ancienne version" qui effectuent également des tests non paramétriques. Ces boîtes de dialogue prennent en charge la fonctionnalité fournie par l'option Tests exacts.

Test khi-deux : Tabule une variable en catégories et calcule une statistique khi-carré basée sur les différences entre les fréquences observées et les fréquences attendues.

Test binomial : Compare la fréquence observée dans chaque catégorie d'une variable dichotomique avec les fréquences attendues de la distribution binomiale.

Suites en séquence : Teste si l'ordre d'occurrence de deux valeurs d'une variable est aléatoire.

Test Kolmogorov-Smirnov pour un échantillon : Compare la fonction de distribution cumulée observée pour une variable avec une distribution théorique spécifiée, qui peut être normale, uniforme, exponentielle ou Poisson.

Tests de deux échantillons indépendants : Compare deux groupes d'observations d'une variable. Le test *U* de Mann-Whitney, le test de Kolmogorov-Smirnov pour deux échantillons, le test de réactions extrêmes Moses et les suites en séquences Wald-Wolfowitz sont disponibles.

Tests de deux échantillons liés : Compare les distributions de deux variables. Le test de Wilcoxon, le test des signes et le test de McNemar sont disponibles.

Tests de plusieurs échantillons indépendants : Compare deux groupes d'observations ou plus d'une variable. Le test de Kruskal-Wallis, le test de la médiane et le test de Jonckheere-Terpstra sont disponibles.

Tests de plusieurs échantillons liés : Compare les distributions de deux variables ou plus. Le test de Friedman, le test *W* de Kendall et le test *Q* de Cochran sont disponibles.

Les quartiles et la moyenne, l'écart type, le minimum, le maximum, et le nombre d'observations sont disponibles pour tous les tests ci-dessus.

Test du khi-deux

La procédure de test du khi-deux (ou khi-carré) tabule une variable en catégories et calcule une statistique khi-deux. Ce test de qualité de l'ajustement compare les fréquences observées et attendues dans chaque catégorie pour vérifier si toutes les catégories contiennent la même proportion de valeurs ou si chaque catégorie contient une proportion de valeurs spécifiées par l'utilisateur.

Exemples : Le test du khi-deux peut être utilisé pour déterminer si un sac de bonbons contient les mêmes proportions de bonbons bleus, marrons, verts, oranges, rouges et jaunes. Vous pouvez aussi tester si le sac de bonbons contient 5 % de bonbons bleus, 30 % de bonbons marrons, 10 % de bonbons verts, 20 % bonbons oranges, 15 % de bonbons rouges et 15 % de bonbons jaunes.

Statistiques : Moyenne, écart type, minimum, maximum et quartiles. Le nombre et le pourcentage d'observations manquantes et non manquantes, le nombre de cas observés et attendus pour chaque catégorie, les résidus et la statistique du khi-carré.

Test du khi-carré : remarques sur les données

Données. Utilisez des variables catégorielles numériques ordonnées ou désordonnées (niveau de mesure ordinal ou nominal). Pour convertir des variables de chaîne en variables numériques, utilisez la procédure de recodage automatique, disponible dans le menu Transformer.

Hypothèses : Les tests non paramétriques ne nécessitent pas d'hypothèses sur la forme de la distribution sous-jacente. On part du principe que les données constituent un échantillon aléatoire. Les fréquences prévues pour chaque catégorie devraient être d'au moins 1. Pas plus de 20% des catégories devraient avoir des fréquences attendues inférieures à 5.

Pour obtenir un test khi-carré

1. A partir des menus, sélectionnez :

   Analyse > **Tests non paramétriques** > **Boîtes de dialogue ancienne version** > **Khi-carré...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez des variables de test. Chaque variable produit un test distinct.
3. Vous pouvez également cliquer sur **Options** pour les statistiques descriptives, les quartiles et le contrôle du traitement des données manquantes.

Valeurs et plages théoriques du test du khi-deux

Plage théorique : Par défaut, chaque valeur distincte de la variable est définie comme catégorie. Pour établir des catégories dans une plage spécifique, sélectionnez **Utiliser la plage indiquée** et fournissez des valeurs entières pour les limites inférieure et supérieure. Des catégories sont établies pour chaque

valeur entière comprise dans la plage, et les observations à l'extérieur des limites sont exclues. Par exemple, si vous spécifiez une valeur de 1 pour la limite inférieure et une valeur de 4 pour la limite supérieure, seules les valeurs entières comprises entre 1 et 4 sont utilisées pour le test du khi-deux.

Valeurs théoriques : Par défaut, toutes les catégories ont des valeurs théoriques égales. Les catégories peuvent avoir des proportions attendues définies par l'utilisateur. Sélectionnez **Valeurs**, indiquez une valeur supérieure à 0 pour chaque catégorie de variable de test et cliquez ensuite sur **Ajouter**. Chaque fois que vous ajoutez une valeur, celle-ci apparaît au bas de la liste des valeurs. L'ordre des valeurs est important. Il correspond à l'ordre croissant des valeurs des catégories de la variable de test. La première valeur de la liste correspond à la valeur de groupe la plus basse de la variable de test et la dernière valeur correspond à la valeur la plus élevée. Les éléments de la liste des valeurs sont additionnés et chaque valeur est ensuite divisée par cette somme pour calculer la proportion d'observations attendues dans la catégorie correspondante. Par exemple, une liste de valeurs de 3, 4, 5, 4 indique les proportions attendues de 3/16, 4/16, 5/16 et 4/16.

Options du test du khi-carré

Statistiques : Vous pouvez choisir l'une des statistiques récapitulatives, ou bien les deux.

- **Caractéristique :** Affiche la moyenne, l'écart type, le minimum, le maximum, et le nombre d'observations non manquantes.
- **Quartiles :** Indique les valeurs correspondant au 25e, 50e et 75e percentiles.

Valeurs manquantes : Contrôle le traitement des valeurs manquantes.

- **Exclure les observations test par test :** Lorsque plusieurs tests sont indiqués, chaque test est effectué séparément selon le nombre des valeurs manquantes.
- **Exclure toute observation incomplète :** Les observations avec des valeurs manquantes pour l'une ou l'autre variable sont exclues de toutes les analyses.

Fonctions supplémentaires de la commande NPAR TESTS (test du khi-deux)

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

- Spécifier des valeurs minimale et maximale différentes, ou des fréquences attendues pour différentes variables (avec la sous-commande CHISQUARE).
- Tester la même variable avec différentes fréquences attendues ou utiliser différentes plages (avec la sous-commande EXPECTED).

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

test binomial

La procédure de test binomial compare les fréquences observées des deux catégories d'une variable dichotomique avec les fréquences que l'on peut attendre d'une distribution binomiale avec un paramètre de probabilité spécifié. Par défaut, le paramètre de probabilité pour les deux groupes est de 0,5. Pour modifier les probabilités, vous pouvez entrer un test de proportion pour le premier groupe. La probabilité pour le second groupe sera de 1 moins la probabilité spécifiée pour le premier groupe.

Exemple. Quand vous jetez une pièce, la probabilité qu'elle tombe côté face est égale à 1/2. Sur la base de cette hypothèse, une pièce est lancée 40 fois et les résultats sont enregistrés (pile ou face). Du test binomial, il se peut que vous observiez que les 3/4 des lancements sont tombés sur le côté face et que le niveau de signification observé est bas (0,0027). Ces résultats indiquent qu'il est peu probable que la probabilité pour que la pièce tombe sur le côté face soit égale à 1/2. La pièce est probablement truquée.

Statistiques : Moyenne, écart type, minimum, maximum, nombre d'observations avec des valeurs non manquantes et quartiles.

Test binomial : remarques sur les données

Données. Les variables testées doivent être numériques et dichotomiques. Pour convertir des variables de chaîne en variables numériques, utilisez la procédure de recodage automatique, disponible dans le menu Transformer. Une **variable dichotomique** est une variable qui accepte deux valeurs uniquement :

oui ou *non*, *true* ou *false*, 0 ou 1, etc. La première valeur rencontrée dans le jeu de données définit le premier groupe, et l'autre valeur définit le deuxième groupe. Si les variables ne sont pas dichotomiques, vous devez spécifier une césure. La césure affecte les observations avec les valeurs qui sont inférieures ou égales au premier groupe et le reste des observations à un deuxième groupe.

Hypothèses : Les tests non paramétriques ne nécessitent pas d'hypothèses sur la forme de la distribution sous-jacente. On part du principe que les données constituent un échantillon aléatoire.

Pour obtenir un test binomial

1. A partir des menus, sélectionnez :

   Analyse > **Tests non paramétriques** > **Boîtes de dialogue ancienne version** > **Binomiale...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables numériques à tester.
3. Vous pouvez également cliquer sur **Options** pour les statistiques descriptives, les quartiles et le contrôle du traitement des données manquantes.

Options du test binomial

Statistiques : Vous pouvez choisir l'une des statistiques récapitulatives, ou bien les deux.

- **Caractéristique :** Affiche la moyenne, l'écart type, le minimum, le maximum, et le nombre d'observations non manquantes.
- **Quartiles :** Indique les valeurs correspondant au 25e, 50e et 75e percentiles.

Valeurs manquantes : Contrôle le traitement des valeurs manquantes.

- **Exclure les observations test par test :** Lorsque plusieurs tests sont indiqués, chaque test est effectué séparément selon le nombre des valeurs manquantes.
- **Exclure toute observation incomplète :** Les observations avec des valeurs manquantes pour toutes les variables testées sont exclues de toutes les analyses.

Fonctions supplémentaires de la commande NPAR TESTS (Test binomial)

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

- Sélectionner des groupes spécifiques (et en exclure d'autres) lorsqu'une variable comporte plus de deux catégories (avec la sous-commande `BINOMIAL`).
- Spécifier différentes césures ou probabilités pour différentes variables (avec la sous-commande `BINOMIAL`).
- Tester la même variable avec différentes césures ou probabilités (avec la sous-commande `EXPECTED`).

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

Suites en séquences

La procédure Suites en séquences teste si l'ordre d'occurrence de deux valeurs d'une variable est aléatoire. Une séquence est une suite d'observations semblables. Un échantillon comportant trop ou trop peu de séquences suggère que l'échantillon n'est pas aléatoire.

Exemples : Supposons que 20 personnes soient sondées pour déterminer si elles achèteraient un produit donné. On peut douter que l'échantillon soit aléatoire si toutes les personnes sont du même sexe. Les suites en séquences peuvent être utilisées pour déterminer si l'échantillon a été tiré au hasard.

Statistiques : Moyenne, écart type, minimum, maximum, nombre d'observations avec des valeurs non manquantes et quartiles.

Suites en séquences : Remarques sur les Données

Données. Les variables doivent être numériques. Pour convertir des variables de chaîne en variables numériques, utilisez la procédure de recodage automatique, disponible dans le menu Transformer.

Hypothèses : Les tests non paramétriques ne nécessitent pas d'hypothèses sur la forme de la distribution sous-jacente. Utilisez des échantillons à distribution de probabilité continue.

Pour obtenir un test de suites

1. A partir des menus, sélectionnez :

   Analyse > **Tests non paramétriques** > **Boîtes de dialogue ancienne version** > **Séquences...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables numériques à tester.
3. Vous pouvez également cliquer sur **Options** pour les statistiques descriptives, les quartiles et le contrôle du traitement des données manquantes.

Césure des Suites en séquences

Césure : Spécifie une césure pour dichotomiser les variables que vous avez choisies. Vous pouvez utiliser soit la moyenne, la médiane ou le mode observés, soit une valeur spécifiée comme césure. Les observations dont les valeurs sont inférieures à la césure sont assignées à un groupe et les observations dont les valeurs sont supérieures à la césure sont assignées à l'autre groupe. Un test est réalisé pour chaque césure choisie.

Options des Suites en séquences

Statistiques : Vous pouvez choisir l'une des statistiques récapitulatives, ou bien les deux.

- **Caractéristique :** Affiche la moyenne, l'écart type, le minimum, le maximum, et le nombre d'observations non manquantes.
- **Quartiles :** Indique les valeurs correspondant au 25e, 50e et 75e percentiles.

Valeurs manquantes : Contrôle le traitement des valeurs manquantes.

- **Exclure les observations test par test :** Lorsque plusieurs tests sont indiqués, chaque test est effectué séparément selon le nombre des valeurs manquantes.
- **Exclure toute observation incomplète :** Les observations avec des valeurs manquantes pour l'une ou l'autre variable sont exclues de toutes les analyses.

Fonctions supplémentaires de la commande NPAR TESTS (Suites en séquences)

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

- Spécifier différentes césures pour différentes variables (à l'aide de la sous-commande RUNS).
- Tester la même variable par rapport à différentes césures personnalisées (à l'aide de la sous-commande RUNS).

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

Test Kolmogorov-Smirnov pour un échantillon

Le test de Kolmogorov-Smirnov pour un échantillon compare la fonction de distribution cumulée observée d'une variable avec une distribution théorique spécifiée, qui peut être normale, uniforme, de Poisson ou exponentielle. Le *Z* de Kolmogorov-Smirnov est calculé à partir de la plus grande différence (en valeur absolue) entre les fonctions de distribution cumulées observées et théoriques. Le test de qualité de l'ajustement contrôle si les observations peuvent avoir été raisonnablement déduites de la distribution spécifiée.

A partir de la version 27.0, les statistiques du test de Lilliefors peuvent être utilisées pour estimer la valeur *p* en utilisant l'échantillon de Monte Carlo pour le tester par rapport à une distribution normale avec les paramètres estimés (cette fonctionnalité n'était auparavant possible que par l'intermédiaire de la procédure Explorer).

Exemple
La plupart des tests paramétriques nécessitent des variables distribuées normalement. Le test de Kolmogorov-Smirnov pour un échantillon permet de vérifier qu'une variable (par exemple *Revenu*) est distribuée normalement.

Statistiques
Moyenne, écart type, minimum, maximum, nombre d'observations sans valeurs manquantes, quartiles, test de Lilliefors et simulation de Monte Carlo.

Remarques sur les données du test de Kolmogorov-Smirnov pour un échantillon

Données
Utilisez des variables quantitatives (mesure d'intervalle ou de rapport).

Hypothèses
Le test de Kolmogorov-Smirnov part du principe que les paramètres de la distribution à tester sont précisés a priori. Cette procédure estime les paramètres à partir d'un échantillon. L'échantillon de moyenne et l'échantillon d'écart type sont les paramètres pour une distribution normale. Les valeurs minimum et maximum de l'échantillon définissent la plage de la distribution uniforme, l'échantillon de moyenne est le paramètre pour la distribution de Poisson et l'échantillon de l'écart type est le paramètre pour la distribution exponentielle. La puissance du test à détecter les abandons de la distribution hypothétique peut être sérieusement diminuée.

Si certains paramètres de la distribution doivent être estimés à partir de l'échantillon, le test de Kolmogorov-Smirnov ne s'applique plus. Dans ces cas-là, les statistiques du test de Lilliefors peuvent être utilisées pour estimer la valeur *p* en utilisant l'échantillon de Monte Carlo pour tester la normalité alors que la moyenne et la variance sont inconnues. Le test de Lilliefors s'applique aux trois distributions continues (**Normale**, **Exponentielle** et **Uniforme**). Notez que le test ne s'applique pas si la distribution sous-jacente est discrète (**Poisson**). Le test n'est défini que pour une inférence à un échantillon si les paramètres de distribution correspondants ne sont pas spécifiés.

Obtention d'un test de Kolmogorov-Smirnov pour un échantillon

1. A partir des menus, sélectionnez :

   Analyse > **Tests non paramétriques** > **Boîtes de dialogue ancienne version** > **K-S à 1 échantillon...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables numériques à tester. Chaque variable produit un test distinct.
3. Sélectionnez éventuellement une méthode de distribution de test :

   Normal
   Si cette option est sélectionnée, vous pouvez indiquer si le ou les paramètres de distribution sont estimés à partir des données d'échantillon (paramètre par défaut) ou à partir de paramètres personnalisé. Si **Utiliser des données d'échantillon** est sélectionné, les résultats asymptotiques existants et la correction de signification de Lilliefors basée sur l'échantillonnage de Monte Carlo sont tous deux utilisés. Si **Personnalisé** est sélectionné, spécifiez des valeurs pour **Moyenne** et **Ecart type**.

   Uniforme
   Si cette option est sélectionnée, vous pouvez indiquer si le ou les paramètres de distribution sont estimés à partir des données d'échantillon (paramètre par défaut) ou à partir de paramètres personnalisé. Si **Utiliser des données d'échantillon** est sélectionné, le test de Lilliefors est utilisé. Si **Personnalisé** est sélectionné, spécifiez des valeurs pour **Min** et **Max**.

   Poisson
   Si cette option est sélectionnée, spécifiez une valeur de paramètre **Moyenne**.

   Exponentiel
   Si cette option est sélectionnée, vous pouvez indiquer si le ou les paramètres de distribution sont estimés à partir de la moyenne de l'échantillon (paramètre par défaut) ou à partir de paramètres

personnalisé. Si **Utiliser des données d'échantillon** est sélectionné, le test de Lilliefors est utilisé. Si **Personnalisé** est sélectionné, spécifiez une valeur de paramètre **Moyenne**.

4. Cliquez éventuellement sur **Simulation** pour spécifier les paramètres de simulation de Monte Carlo, cliquez sur **Exact** pour spécifier les paramètres de test exacts ou cliquez sur **Options** pour les statistiques descriptives, les quartiles et le contrôle du traitement des données manquantes.

Test de Kolmogorov-Smirnov pour un échantillon : Simulation

Si certains paramètres de la distribution doivent être estimés à partir de l'échantillon, le test de Kolmogorov-Smirnov ne s'applique plus. Dans ces cas-là, les statistiques du test de Lilliefors peuvent être utilisées pour estimer la valeur *p* en utilisant l'échantillon de Monte Carlo pour tester la normalité alors que la moyenne et la variance sont inconnues. Le test de Lilliefors s'applique aux trois distributions continues (**Normale**, **Exponentielle** et **Uniforme**). Notez que le test ne s'applique pas si la distribution sous-jacente est discrète (**Poisson**). Le test n'est défini que pour une inférence à un échantillon si les paramètres de distribution correspondants ne sont pas spécifiés.

Paramètres de simulation de Monte Carlo

Niveau de fiabilité
Ce paramètre facultatif réinitialise le niveau d'intervalle de confiance estimé par le test de Kolmogorov-Smirnov si les simulations de Monte Carlo sont utilisées. La valeur doit être comprise entre 0 et 100. La valeur par défaut est 99.

Nombre d'échantillons
Ce paramètre facultatif réinitialise le nombre de réplicats utilisés par le test de Lilliefors pour l'échantillonnage de Monte Carlo. Cette valeur doit être un entier unique compris entre 10 000 et le nombre le plus élevé de valeurs d'échantillon. La valeur par défaut est 10 000.

Supprimer les résultats de Monte Carlo pour la distribution normale
Ce paramètre facultatif supprime l'échantillonnage de Monte Carlo pour les résultats de la distribution normale. Par défaut, ce paramètre n'est pas sélectionné (ce qui signifie que les résultats asymptotiques et les résultats du test de Lilliefors, basés sur l'échantillonnage de Monte Carlo, sont tous deux présentés).

Test de Kolmogorov-Smirnov pour un échantillon : Options

Statistiques
Vous pouvez choisir l'une des statistiques récapitulatives, ou bien les deux.

Descriptive
Affiche la moyenne, l'écart type, le minimum, le maximum, et le nombre d'observations non manquantes.

Quartiles
Indique les valeurs correspondant au 25e, 50e et 75e percentiles.

Valeurs manquantes
Contrôle le traitement des valeurs manquantes.

Exclure les observations test par test
Lorsque plusieurs tests sont indiqués, chaque test est effectué séparément selon le nombre des valeurs manquantes.

Exclure toute observation incomplète
Les observations avec des valeurs manquantes pour l'une ou l'autre variable sont exclues de toutes les analyses.

Fonctions supplémentaires de commandes NPAR TESTS (test de Kolmogorov-Smirnov pour un échantillon)

Le langage de syntaxe de commande vous permet également de spécifier des paramètres pour la distribution du test (avec la sous-commande K-S).

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

Tests à deux échantillons indépendants

La procédure des tests à deux échantillons indépendants compare deux groupes d'observations en fonction d'une variable.

Exemple. De nouveaux appareils dentaires qui sont censés être plus confortables, avoir une apparence plus agréable et provoquer des progrès plus rapides pour le redressage des dents ont été développés. Pour savoir si les nouveaux appareils doivent être portés aussi longtemps que les anciens, 10 enfants sont choisis de façon aléatoire pour porter les anciens appareils et 10 autres pour porter les nouveaux appareils. Le test *U* de Mann-Whitney peut, en moyenne, vous montrer que les sujets portant les nouveaux appareils ne doivent pas les porter aussi longtemps que les sujets utilisant les anciens appareils.

Statistiques : Moyenne, écart type, minimum, maximum, nombre d'observations avec des valeurs non manquantes et quartiles. Tests : *U* de Mann-Whitney, réactions extrêmes de Moses, *Z* de Kolmogorov-Smirnov et suites de Wald-Wolfowitz.

Tests pour deux échantillons indépendants : remarques sur les données

Données. Utilisez des variables numériques qui peuvent être ordonnées.

Hypothèses : Utilisez des échantillons indépendants, aléatoires. Le test *U* de Mann-Whitney teste l'égalité de deux distributions. Afin de l'utiliser pour tester les différences entre deux distributions, vous devez supposer que les distributions sont de la même forme.

Pour effectuer les tests pour deux échantillons indépendants

1. A partir des menus, sélectionnez :

   Analyse > **Tests non paramétriques** > **Boîtes de dialogue ancienne version** > **2 échantillons indépendants...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables numériques.
3. Sélectionnez une variable de regroupement et cliquez sur **Définir groupes...** pour scinder le fichier en deux groupes ou échantillons.

Types de tests pour deux échantillons indépendants

Type de test : Quatre tests sont disponibles pour tester si deux échantillons (groupes) proviennent de la même population.

Le **test U de Mann-Whitney** est le plus populaire des tests pour deux échantillons indépendants. Il équivaut au test de Wilcoxon et au test de Kruskal-Wallis pour deux groupes. Les tests de Mann-Whitney servent à vérifier que deux échantillons d'une population ont une position équivalente. Les observations des deux groupes sont combinées et ordonnées, et il leur est attribué un rang moyen en cas d'ex aequo. Le nombre d'ex aequo doit être petit par rapport au nombre total d'observations. Si les populations ont une position identique, les rangs doivent être attribués de façon aléatoire entre les deux échantillons. Le test calcule le nombre de fois qu'un score du groupe 1 précède un score du groupe 2, ainsi que le nombre de fois qu'un score du groupe 2 précède un score du groupe 1. La statistique *U* de Mann-Whitney est le plus petit de ces deux nombres. La statistique de la somme des rangs de Wilcoxon *W* est également affichée. *M* est la somme des rangs pour le groupe avec le rang moyen plus petit, à moins que les groupes aient le même rang moyen, auquel cas il s'agit de la somme de rang du groupe qui est nommé en dernier dans la boîte de dialogue Deux échantillons indépendants Définir groupes.

Le **test Z de Kolmogorov-Smirnov** et les **suites en séquences de Wald-Wolfowitz** sont des tests plus généraux qui détectent les différences de position et la forme des distributions. Le test Z de Kolmogorov-Smirnov est basé sur la différence absolue maximum entre les fonctions de distribution cumulées observées pour les deux échantillons. Lorsque cette différence est significative, on considère que les deux distributions sont différentes. Le test des suites en séquences de Wald-Wolfowitz combine et ordonne les observations des deux groupes. Si les deux échantillons proviennent de la même population, les deux groupes doivent être répartis au hasard dans le classement.

Le **test des réactions extrêmes de Moses** part du principe que la variable expérimentale influence certains sujets dans une direction et d'autres sujets dans la direction opposée. Le test vérifie les réponses extrêmes par rapport à un groupe de contrôle. Ce test permet d'étudier l'intervalle du groupe de contrôle et de mesurer à quel point les valeurs extrêmes du groupe expérimental influencent l'amplitude lorsque ce test est associé au groupe de contrôle. Le groupe de contrôle est défini par la valeur du groupe 1 dans la boîte de dialogue Définition des deux groupes d'échantillons indépendants. Les observations des deux groupes sont combinées et ordonnées. L'intervalle du groupe de contrôle est calculé comme étant la différence entre les rangs des valeurs les plus grandes et les plus petites dans le groupe de contrôle plus 1. Comme les valeurs aberrantes peuvent facilement déformer la plage de l'intervalle, 5 % des cas de contrôle sont automatiquement tronqués à chaque extrémité.

Définition de deux groupes d'échantillons indépendants

Pour diviser le fichier en deux groupes ou exemples, entrez une valeur entière pour le groupe 1 et une autre pour le groupe 2. Les observations avec d'autres valeurs sont exclues de l'analyse.

Options des tests pour deux échantillons indépendants

Statistiques : Vous pouvez choisir l'une des statistiques récapitulatives, ou bien les deux.

- **Caractéristique :** Indique la moyenne, l'écart type, le minimum, le maximum, et le nombre d'observations sans valeurs manquantes.
- **Quartiles :** Indique les valeurs correspondant au 25e, 50e et 75e percentiles.

Valeurs manquantes : Contrôle le traitement des valeurs manquantes.

- **Exclure les observations test par test :** Lorsque plusieurs tests sont indiqués, chaque test est effectué séparément selon le nombre des valeurs manquantes.
- **Exclure toute observation incomplète :** Les observations avec des valeurs manquantes pour l'une ou l'autre variable sont exclues de toutes les analyses.

Fonctions supplémentaires de la commande NPAR TESTS (tests pour deux échantillons indépendants)

Le langage de syntaxe de commande vous permet également de spécifier le nombre d'observations devant être tronquées pour le test de Moses (avec la sous-commande `MOSES`).

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

tests à deux échantillons liés

La procédure des tests à deux échantillons liés compare les distributions pour deux variables.

Exemple. En général, une famille qui vend sa maison perçoit-elle le prix demandé ? En appliquant le test de Wilcoxon aux données de 10 foyers, vous apprendrez que sept familles perçoivent moins que le prix demandé, qu'une famille perçoit plus que le prix demandé et que deux familles perçoivent le prix demandé.

Statistiques : Moyenne, écart type, minimum, maximum, nombre d'observations avec des valeurs non manquantes et quartiles. Tests : classement Wilcoxon, signe, McNemar. Si l'option Tests exacts est installée (disponible uniquement sous les systèmes d'exploitation Windows), le test d'Homogénéité marginale est alors disponible.

Tests pour deux échantillons liés : remarques sur les données

Données. Utilisez des variables numériques qui peuvent être ordonnées.

Hypothèses : Bien qu'aucune distribution particulière ne soit supposée pour les deux variables, on part du principe que la distribution de la population des différences liées est symétrique.

Pour obtenir des tests pour deux échantillons liés

1. A partir des menus, sélectionnez :

Analyse > **Tests non paramétriques** > **Boîtes de dialogue ancienne version** > **2 échantillons liés...**

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

2. Sélectionnez une ou plusieurs paires de variables.

Types de tests pour deux échantillons liés

les tests de cette section comparent les distributions pour deux variables liées. Le test qu'il convient d'utiliser dépend du type de données.

Si vos données sont continues, utilisez le test de Signe ou le test de Wilcoxon. Le **test de signe** calcule les différences entre les deux variables pour toutes les observations, et classe les différences comme étant positives, négatives ou liées. Si les deux variables sont réparties de la même manière, le nombre de différences positives et le nombre de différences négatives ne diffèrent pas de façon significative. Le **test de Wilcoxon** prend en compte les informations relatives au signe des différences, ainsi qu'à l'amplitude des différences entre paires. Comme le test de Wilcoxon intègre plus de renseignements sur les données, il est plus puissant que le test des signes.

Si vos données sont binaires, utilisez le **test de McNemar**. Ce test s'utilise fréquemment lors de situations de mesures répétées, au cours desquelles la réponse du sujet est provoquée deux fois, une fois avant qu'un événement spécifié se produise et une fois après qu'un événement spécifié s'est produit. Le test de McNemar détermine si le taux de réponses initial (avant l'événement) est égal au taux de réponse final (après l'événement). Ce test est utile pour détecter les changements dans les réponses dues à une intervention expérimentale dans les plans avant et après.

Si vos données sont catégorielles, utilisez le **test d'Homogénéité marginale**. Ce test est un développement du test de McNemar d'une réponse binaire à une réponse multinomiale. Il recherche les changements de réponse en utilisant la distribution khi-carré et permet de détecter les changements de réponse dus à une intervention expérimentale dans les plans avant et après. Le test d'homogénéité marginale n'est disponible que si vous avez installé Exact Tests.

Options des tests pour deux échantillons liés

Statistiques : Vous pouvez choisir l'une des statistiques récapitulatives, ou bien les deux.

- **Caractéristique :** Indique la moyenne, l'écart type, le minimum, le maximum, et le nombre d'observations sans valeurs manquantes.
- **Quartiles :** Indique les valeurs correspondant au 25e, 50e et 75e percentiles.

Valeurs manquantes : Contrôle le traitement des valeurs manquantes.

- **Exclure les observations test par test :** Lorsque plusieurs tests sont indiqués, chaque test est effectué séparément selon le nombre des valeurs manquantes.
- **Exclure toute observation incomplète :** Les observations avec des valeurs manquantes pour l'une ou l'autre variable sont exclues de toutes les analyses.

Fonctions supplémentaires de la commande NPAR (Deux échantillons liés)

Le langage de syntaxe de commande vous permet également de tester une variable avec chaque variable d'une liste.

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

Tests pour plusieurs échantillons indépendants

La procédure de Tests pour Plusieurs Echantillons Indépendants compare deux groupes d'observations ou plus sur une variable.

Exemple. Trois marques d'ampoules 100 watts diffèrent-elles par leur durée moyenne de fonctionnement ? A partir de l'analyse de variance d'ordre 1 de Kruskal-Wallis, vous apprendrez peut-être que les trois marques diffèrent par leur durée de vie moyenne.

Statistiques : Moyenne, écart type, minimum, maximum, nombre d'observations avec des valeurs non manquantes et quartiles. Tests : *H* de Kruskal-Wallis, médiane.

Tests pour Plusieurs Echantillons Indépendants : Remarques sur les Données

Données. Utilisez des variables numériques qui peuvent être ordonnées.

Hypothèses : Utilisez des échantillons indépendants, aléatoires. Le test du *H* de Kruskal-Wallis nécessite que les échantillons testés soient de forme similaire.

Pour obtenir des tests pour plusieurs échantillons indépendants

1. A partir des menus, sélectionnez :

   Analyse > **Tests non paramétriques** > **Boîtes de dialogue ancienne version** > **K échantillons indépendants...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables numériques.
3. Sélectionnez une variable de regroupement et cliquez sur **Définir plage** pour spécifier les valeurs entières minimale et maximale pour la variable de regroupement.

Tests pour Plusieurs Echantillons Indépendants : Types de tests

Trois tests permettent de déterminer si plusieurs échantillons indépendants proviennent de la même population. Le test du *H* de Kruskal-Wallis, le test de la Médiane et le test de Jonckheere-Terpstra testent tous si plusieurs échantillons indépendants proviennent de la même population.

Le test **H de Kruskal-Wallis**, extension du test du *U* de Mann-Whitney, est l'équivalent non paramétrique de l'analyse de variance d'ordre 1 et détecte les différences dans la position de la distribution. Le **test de la médiane**, test plus général mais moins puissant, détecte les différences de position et de forme des distributions. Le test du *H* de Kruskal-Wallis et le test de la médiane supposent qu'il n'existe aucun classement *a priori* des *k* populations à partir desquelles les échantillons sont tirés.

Lorsqu'il *existe* un classement naturel *a priori* (ascendant ou descendant) des *k* populations, le test de **Jonckheere-Terpstra** est plus puissant. Par exemple, les *k* populations peuvent représenter *k* températures croissantes. L'hypothèse selon laquelle différentes températures produisent la même distribution des réponses est testée contre l'hypothèse alternative selon laquelle l'accroissement de température fait augmenter l'ampleur de la réponse. Ici, l'hypothèse alternative est ordonnée ; le test de Jonckheere-Terpstra est donc le plus approprié. Le test de Jonckheere-Terpstra n'est disponible que si vous avez installé le module complémentaire Tests Exacts.

Tests pour Plusieurs Echantillons Indépendants : Définir Plage

Pour définir la plage, entrez des valeurs entières pour **Minimum** et **Maximum** qui correspondent à la catégorie la plus basse et à la plus haute du critère de regroupement. Les observations dont les valeurs se trouvent à l'extérieur des limites sont exclues. Par exemple, si vous spécifiez une valeur minimale de 1 et une valeur maximale de 3, seules les valeurs entières comprises entre 1 et 3 seront utilisées. La valeur minimale doit être inférieure à la valeur maximale, et les deux valeurs doivent être spécifiées.

Options des Tests pour Plusieurs Echantillons Indépendants

Statistiques : Vous pouvez choisir l'une des statistiques récapitulatives, ou bien les deux.

- **Caractéristique :** Indique la moyenne, l'écart type, le minimum, le maximum, et le nombre d'observations sans valeurs manquantes.
- **Quartiles :** Indique les valeurs correspondant au 25e, 50e et 75e percentiles.

Valeurs manquantes : Contrôle le traitement des valeurs manquantes.

- **Exclure les observations test par test :** Lorsque plusieurs tests sont indiqués, chaque test est effectué séparément selon le nombre des valeurs manquantes.

- **Exclure toute observation incomplète :** Les observations avec des valeurs manquantes pour l'une ou l'autre variable sont exclues de toutes les analyses.

Fonctions supplémentaires de la commande NPAR TESTS (K échantillons indépendants)

Le langage de syntaxe de commande vous permet également de spécifier une valeur différente de la médiane observée pour le test de la médiane (avec la sous-commande `MEDIAN`).

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

Tests pour plusieurs échantillons liés

La procédure de Tests pour Plusieurs Echantillons Liés compare les distributions de deux variables ou plus.

Exemple. Le public associe-t-il différents niveaux de prestige à un docteur, un avocat, un officier de police et un enseignant ? On demande à dix personnes de classer ces quatre métiers par ordre de prestige. Le test de Friedman indique que le public associe effectivement différents niveaux de prestige à ces quatre professions.

Statistiques : Moyenne, écart type, minimum, maximum, nombre d'observations avec des valeurs non manquantes et quartiles. Tests : Friedman, *W* de Kendall et *Q* de Cochran.

Tests pour plusieurs échantillons liés de considérations de données

Données. Utilisez des variables numériques qui peuvent être ordonnées.

Hypothèses : Les tests non paramétriques ne nécessitent pas d'hypothèses sur la forme de la distribution sous-jacente. Utilisez des échantillons dépendants, aléatoires.

Pour obtenir des tests pour plusieurs échantillons liés

1. A partir des menus, sélectionnez :

   Analyse > **Tests non paramétriques** > **Boîtes de dialogue ancienne version** > **K échantillons liés...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez deux variables numériques ou plus à tester.

Tests pour plusieurs échantillons liés de types de tests

Trois tests sont disponibles pour comparer les distributions de plusieurs variables liées.

Le **test** de **Friedman** est l'équivalent non paramétrique d'un plan de mesures répétées sur un échantillon ou d'une analyse de variance à deux facteurs avec une observation par cellule. Le test de Friedman teste l'hypothèse nulle selon laquelle *k* variables liées proviennent de la même population. Pour chaque cas, les variables *k* sont classées de 1 à *k*. La statistique de test est basée sur ces rangs.

Le test **W de Kendall** est une standardisation de la statistique de Friedman. Le test *W* de Kendall peut être interprété comme le coefficient de concordance, qui est une mesure de l'accord entre les évaluateurs. Chaque observation est un juge ou un indicateur, et chaque variable est une personne ou un élément jugé. Pour chaque variable, la somme des rangs est calculée. Le *W* de Kendall se situe entre 0 (pas d'accord) et 1 (accord total).

Le **Q de Cochran** est identique au test de Friedman mais s'applique lorsque toutes les réponses sont binaires. Ce test est une extension du test de McNemar à *K* échantillons. Le *Q* de Cochran teste l'hypothèse nulle selon laquelle plusieurs variables dichotomiques liées ont la même moyenne. Les variables sont mesurées sur le même individu ou sur des individus comparables.

Statistiques des tests pour plusieurs échantillons liés

Vous pouvez choisir les statistiques.

- **Caractéristique :** Indique la moyenne, l'écart type, le minimum, le maximum, et le nombre d'observations sans valeurs manquantes.
- **Quartiles :** Indique les valeurs correspondant au 25e, 50e et 75e percentiles.

Fonctions supplémentaires de la commande NPAR TESTS (K échantillons liés)

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

Analyse des réponses multiples

Analyse des réponses multiples

Deux procédures sont proposées pour l'analyse des jeux comportant plusieurs variables dichotomiques ou plusieurs catégories. La procédure Fréquences multiréponses affiche les tables de fréquences. La procédure des Tableaux croisés multiréponses affiche des tableaux croisés à deux ou trois dimensions. Avant d'utiliser l'une de ces procédures, vous devez définir des jeux de réponses multiples.

Exemple. Cet exemple illustre l'utilisation des éléments multiréponses dans une étude de marché. Ces données sont fictives et ne doivent pas être considérées comme réelles. Une compagnie aérienne est parfois amenée à interroger les passagers d'un trajet donné pour évaluer la concurrence. Dans cet exemple, American Airlines veut savoir si ses passagers utilisent d'autres compagnies aériennes pour couvrir le trajet Chicago-New York et connaître l'importance relative des horaires et du service dans le choix d'une compagnie aérienne. L'hôtesse distribue à chaque passager un court questionnaire lors de l'embarquement. La première question est la suivante : Entourez toutes les compagnies aériennes par lesquelles vous avez effectué au moins un vol au cours des six derniers mois parmi American, United, TWA, USAir et d'autres. Il s'agit d'une question à réponses multiples, car le passager peut entourer plus d'une réponse. Cependant, cette question ne peut pas être codée directement, parce qu'une variable ne peut avoir qu'une valeur pour chaque cas. Vous devez utiliser plusieurs variables pour mapper les réponses à chaque question. Ceci peut être fait de deux manières. L'une consiste à définir une variable correspondant à chaque choix possible (par exemple, American, United, TWA, USAir et d'autres). Si le passager entoure United, le code 1 est attribué à la variable *united* ; dans le cas contraire, le code 0 est attribué. Il s'agit d'une **méthode de dichotomies multiples** pour le mappage des variables. L'autre méthode permettant de mapper les réponses est la **méthode des catégories multiples**, où vous devez estimer le nombre maximal de réponses possibles à la question et définir le même nombre de variables, avec des codes correspondant à la compagnie aérienne empruntée. En utilisant un échantillon de questionnaires, vous vous apercevrez peut-être que personne n'a emprunté plus de trois compagnies différentes pour ce trajet. Qui plus est, vous vous rendrez compte que, du fait de la déréglementation des compagnies aériennes, 10 autres compagnies figurent dans la catégorie Autre. A l'aide de la méthode multiréponses, vous pouvez définir trois variables, codées comme suit : 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta*, etc. Si un passager entoure American et TWA, la première variable porte le code 1, la seconde le code 3, et la troisième un code sans valeur. Un autre passager a peut-être entouré American et ajouté Delta. Ainsi, la première variable porte le code 1, la seconde le code 5, et la troisième un code sans valeur. Si vous utilisez la méthode des dichotomies multiples, d'un autre côté, vous finissez par vous retrouver avec 14 variables différentes. Les deux méthodes de codage sont possibles dans le cadre de cette enquête. Cependant, votre choix dépendra de la répartition des réponses.

Définition de jeux de réponses multiples

La procédure de définition de jeux de réponses multiples regroupe des variables élémentaires dans des jeux de dichotomies ou de catégories multiples, pour lesquels vous pouvez obtenir des tables de fréquences et des tableaux croisés. Vous pouvez définir jusqu'à 20 jeux de réponses multiples. Chaque vecteur doit avoir un nom unique. Pour éliminer un jeu, sélectionnez-le dans la liste des jeux de réponses multiples et cliquez sur **Eliminer**. Pour modifier un vecteur, sélectionnez-le dans la liste, modifiez-en les caractéristiques et cliquez sur **Changer**.

Vous pouvez coder vos variables élémentaires sous forme de dichotomies ou de catégories. Pour utiliser des variables dichotomiques, sélectionnez **Variables dichotomiques** afin de créer un jeu de dichotomies multiples. Entrez une valeur entière dans Valeur comptée. Chaque variable ayant au moins

une occurrence de la valeur comptée devient une catégorie du jeu de dichotomies multiples. Sélectionnez **Catégories** pour créer un jeu de catégories multiples ayant la même plage de valeurs que les variables qui le composent. Entrez des nombres entiers pour le minimum et le maximum de la plage des catégories du jeu de catégories multiples. La procédure totalise chaque valeur entière contenue dans la plage pour toutes les variables qui le composent. Les catégories vides ne sont pas mises en tableau.

A chaque jeu de réponses multiples doit être attribué un nom unique de 7 caractères maximum. La procédure ajoute un signe dollar ($) devant le nom que vous avez attribué. Vous ne pouvez pas utiliser les noms réservés suivants : *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* et *width*. Le nom du jeu de réponses multiples doit uniquement être utilisé dans les procédures multiréponses. Vous ne pouvez pas faire référence aux noms d'ensemble de réponses multiples dans les autres procédures. A titre facultatif, vous pouvez entrer un libellé de variable décrivant le jeu de réponses multiples. Ce libellé peut comporter jusqu'à 40 caractères.

Définir des jeux de réponses multiples

1. A partir des menus, sélectionnez :

   Analyse > **Réponses multiples** > **Définir des jeux de variables...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez au moins deux variables.
3. Si vos variables sont codées comme dichotomies, indiquez la valeur que vous souhaitez calculer. Si elles sont codées comme catégories, définissez leur plage.
4. Entrez un nom unique pour chaque jeu de réponses multiples.
5. Cliquez sur **Ajouter** pour ajouter le jeu de réponses multiples à la liste des jeux définis.

Tableaux de fréquences des réponses multiples

La procédure Tableaux d'effectifs des réponses multiples produit des tables de fréquences pour les jeux de réponses multiples. Vous devez d'abord définir un ou plusieurs jeux de réponses multiples (voir " Définir les jeux de réponses multiples ").

Pour les jeux de dichotomies multiples, les noms de catégorie apparaissant dans la sortie proviennent de libellés de variable définis pour les variables élémentaires du groupe. Si les libellés de variable ne sont pas définis, les noms de variables servent de libellés. Pour les jeux de catégories multiples, les libellés des catégories proviennent des libellés de valeurs de la première variable du groupe. Si les catégories manquantes de la première variable sont présentes pour d'autres variables du groupe, définissez un libellé de valeurs pour les catégories manquantes.

Valeurs manquantes : Les cas de valeurs manquantes sont exclus tableau par tableau. Vous pouvez donc choisir l'une ou les deux solutions suivantes :

- **Exclure les observations ayant une information incomplète à l'intérieur des dichotomies :** Ceci permet d'exclure les observations ayant des valeurs manquantes pour toute variable issue du tableau croisé du jeu de dichotomies multiples. Ceci s'applique seulement aux jeux de réponses multiples définis comme jeux de dichotomies. Par défaut, une observation est considérée manquante pour un jeu de dichotomies multiples si aucune de ses variables composantes ne contient de valeur comptée. Les cas de valeurs manquantes pour certaines variables, mais pas toutes, sont inclus dans les tabulations du groupe si au moins une variable contient la valeur comptée.
- **Exclure toute observation ayant une information incomplète à l'intérieur des catégories :** Cela permet d'exclure les observations ayant des valeurs manquantes pour toute variable provenant du tableau croisé du jeu des catégories multiples. Ceci s'applique seulement aux jeux de réponses multiples définis comme des jeux de catégories. Par défaut, une observation est considérée manquante pour un jeu de catégories multiples si aucune de ses composantes n'a de valeurs valides à l'intérieur de la plage définie.

Exemple. Chaque variable créée à partir d'une question de l'enquête est une variable élémentaire. Pour analyser un élément multiréponses, vous devez combiner les variables dans l'un des deux types de

jeux de réponses multiples : jeu de catégories multiples ou jeu de dichotomies multiples. Par exemple, si dans une enquête, une compagnie aérienne vous demande la compagnie (American Airlines, United Airlines ou TWA) que vous avez empruntée au cours des six derniers mois, si vous utilisez des variables dichotomiques et avez défini un **jeu de dichotomies multiples**, chacune des trois variables du jeu devient une catégorie de la variable de groupe. Les effectifs et les pourcentages correspondant aux trois compagnies aériennes s'affichent dans une table de fréquences. Si vous découvrez qu'aucun des répondants n'a mentionné plus de deux compagnies, vous pouvez créer deux variables, chacune ayant trois codes, un par compagnie aérienne. Si vous définissez un **jeu de catégories multiples**, les valeurs sont mises en tableau et les mêmes codes sont ajoutés dans toutes les variables élémentaires. Le vecteur de valeurs résultant est le même que pour chacune des variables élémentaires. Par exemple, 30 réponses pour United sont la somme des cinq réponses United pour la compagnie 1 et des 25 réponses United pour la compagnie 2. Les effectifs et les pourcentages correspondant aux trois compagnies aériennes s'affichent dans une table de fréquences.

Statistiques : Tables de fréquences contenant des effectifs, des pourcentages de réponses, des pourcentages de cas, le nombre de cas valables, et le nombre de cas manquants.

Gestion des données des fréquences de réponses multiples

Données. Utilisez des jeux de réponses multiples.

Hypothèses : Les effectifs et pourcentages représentent une description utile des données de n'importe quelle distribution.

Procédures apparentées : La procédure Définir Jeux de réponses multiples vous permet de définir des jeux de réponses multiples.

Pour obtenir des tableaux de fréquences de réponses multiples

1. A partir des menus, sélectionnez :

   Analyse > **Réponses multiples** > **Fréquences...**
2. Sélectionnez un ou plusieurs jeux de réponses multiples.

Tableaux croisés des réponses multiples

La procédure Tableaux croisés de réponses multiples classe, par tableaux croisés, des jeux de réponses multiples définis, des variables élémentaires ou une combinaison. Vous pouvez également obtenir des pourcentages de cellules basés sur des observations ou des réponses, modifier la gestion des valeurs manquantes ou obtenir des tableaux croisés appariés. Vous devez d'abord définir un ou plusieurs jeux de réponses multiples (veuillez consulter " Pour Définir des jeux de réponses multiples ").

Pour les jeux de dichotomies multiples, les noms de catégorie apparaissant dans la sortie proviennent de libellés de variable définis pour les variables élémentaires du groupe. Si les libellés de variable ne sont pas définis, les noms de variables servent de libellés. Pour les jeux de catégories multiples, les libellés des catégories proviennent des libellés de valeurs de la première variable du groupe. Si les catégories manquantes de la première variable sont présentes pour d'autres variables du groupe, définissez un libellé de valeurs pour les catégories manquantes. La procédure affiche les libellés de catégorie des colonnes sur trois lignes, avec jusqu'à huit caractères par ligne. Pour éviter de scinder les mots, vous pouvez inverser les éléments lignes et les éléments colonnes ou redéfinir les libellés.

Exemple. Les jeux de dichotomies multiples et les jeux de catégories multiples peuvent être croisés avec d'autres variables dans cette procédure. Dans le cadre d'une enquête menée auprès de passagers de compagnies aériennes, voici ce qui leur a été demandé : Parmi les compagnies aériennes suivantes, entourez toutes celles avec lesquelles vous avez voyagé au moins une fois durant les six derniers mois (American, United, TWA). Est-il plus important de privilégier l'horaire ou le service ? Choisissez une seule réponse. Après avoir saisi les données en tant que dichotomies ou catégories multiples, et après les avoir combinées dans un vecteur, vous pouvez croiser les choix de compagnie aérienne déclarés avec la question relative au service ou aux horaires.

Statistiques : Tableau croisé avec cellule, ligne, colonne, et effectif total, et avec les pourcentages ligne, colonne, et effectif total. Les pourcentages cellule peuvent être basés sur les observations ou les réponses.

Remarques sur les Données de Tableaux croisés de réponses multiples

Données. Utilisez des jeux de réponses multiples ou des variables catégorielles numériques.

Hypothèses : Les effectifs et pourcentages offrent une description utile des données qui suivent tout type de distribution.

Procédures apparentées : La procédure Définir Jeux de réponses multiples vous permet de définir des jeux de réponses multiples.

Pour obtenir des tableaux croisés des réponses multiples

1. A partir des menus, sélectionnez :

   Analyse > **Réponses multiples** > **Tableaux croisés...**
2. Sélectionnez une ou plusieurs variables numériques ou jeux de réponses multiples pour chaque dimension de tableau croisé.
3. Définissez la plage de chaque variable élémentaire.

Sinon, vous pouvez obtenir un tableau croisé bilatéral pour chaque catégorie de variable de contrôle ou chaque jeu de réponses multiples. Sélectionnez un ou plusieurs éléments pour la liste de couche(s).

Définir Plages Tableaux croisés De réponses multiples

Les plages des valeurs doivent être définies pour toute variable élémentaire de tableaux croisés. Entrez les valeurs entières de catégories minimum et maximum que vous souhaitez tabuler. Les catégories se situant en dehors de la plage sont exclues de l'analyse. Les valeurs se situant à l'intérieur de la plage inclusive sont supposées être des nombres entiers (les nombres non entiers sont tronqués).

Options Tableaux croisés de réponses multiples

Pourcentages de cellule : Les effectifs des cellules sont toujours affichés. Vous pouvez choisir d'afficher les pourcentages lignes, les pourcentages colonnes, et les pourcentages tableau bilatéral (total).

Pourcentages basés sur : Vous pouvez baser les pourcentages cellules sur les observations (ou répondants). Ceci n'est pas possible si vous sélectionnez la fonction qui permet d'apparier les variables entre les jeux de catégories multiples. Vous pouvez aussi baser les pourcentages cellules sur les réponses. Pour les jeux de dichotomies multiples, le nombre de réponses est égal au nombre de valeurs comptées à travers les observations. Pour les jeux de catégories multiples, le nombre de réponses correspond au nombre de valeurs comprises dans la plage défini.

Valeurs manquantes : Vous avez le choix entre les deux options suivantes :

- **Exclure les observations ayant une information incomplète à l'intérieur des dichotomies :** Ceci permet d'exclure les observations ayant des valeurs manquantes pour toute variable issue du tableau croisé du jeu de dichotomies multiples. Ceci s'applique seulement aux jeux de réponses multiples définis comme jeux de dichotomies. Par défaut, une observation est considérée manquante pour un jeu de dichotomies multiples si aucune de ses variables composantes ne contient de valeur comptée. Les observations ayant des valeurs manquantes pour certaines, mais pas toutes, les variables sont incluses dans les tableaux croisés du groupe si au moins une variable contient la valeur comptée.
- **Exclure toute observation ayant une information incomplète à l'intérieur des catégories :** Cela permet d'exclure les observations ayant des valeurs manquantes pour toute variable provenant du tableau croisé du jeu des catégories multiples. Ceci s'applique seulement aux jeux de réponses multiples définis comme des jeux de catégories. Par défaut, une observation est considérée manquante pour un jeu de catégories multiples si aucune de ses composantes n'a de valeurs valides à l'intérieur de la plage définie.

Par défaut, lorsque vous croisez deux jeux de catégories multiples, la procédure tabule chaque variable du premier groupe avec chaque variable du second groupe et additionne les effectifs de chaque cellule. Par conséquent, certaines réponses peuvent apparaître plus d'une fois dans un tableau. Vous pouvez choisir l'option suivante :

Apparier les variables entre les vecteurs réponses : Cela permet d'apparier la première variable du premier groupe avec la première variable du second groupe, etc. Si vous sélectionnez cette option, la procédure basera les pourcentages cellules sur les réponses plutôt que sur les répondants. On ne peut apparier les jeux de dichotomies multiples ou les variables élémentaires.

Fonctions supplémentaires de la commande MULT RESPONSE

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

- Obtenir des tableaux croisés ayant jusqu'à cinq dimensions (avec la sous-commande BY).
- Modifier les options de formatage de la sortie, y compris la suppression des libellés de valeurs (avec la sous-commande FORMAT).

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

Rapports de résultats

Rapports de résultats

Les listes d'observations et les statistiques descriptives sont des outils de base permettant d'étudier et de présenter des données. Vous pouvez obtenir les listes d'observations à l'aide de l'éditeur de Données ou de la procédure Récapituler, les effectifs de fréquences et les statistiques descriptives à l'aide de la procédure Fréquences, et les statistiques de sous-population à l'aide de la procédure Moyennes. Chacune de ces procédures utilise un format destiné à rendre les informations claires. Si vous souhaitez afficher les informations dans un format différent, les procédures Rapport en lignes et Rapport en colonnes vous permettent de contrôler la présentation des données.

Récapitulatifs du rapport en lignes

Rapport en lignes produit des rapports dans lesquels différentes statistiques récapitulatives sont disposées en lignes. Les listes d'observations sont également disponibles, avec ou sans statistiques récapitulatives.

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

Exemple. Une société possédant une chaîne de magasins conserve des dossiers sur les employés comprenant le salaire, l'ancienneté, le magasin et la division où chaque employé travaille. Vous pourriez générer un rapport fournissant les informations individuelles sur les employés (liste) divisées par magasin et par division (variables d'agrégation), avec les statistiques récapitulatives (par exemple, salaire moyen) pour chaque magasin, division et par division dans chaque magasin.

Colonnes de données : Donne la liste des variables de rapport pour lesquelles vous voulez obtenir des listes d'observations ou des statistiques récapitulatives, et contrôle le format d'affichage des colonnes de données.

Colonne de rupture : Donne la liste des variables d'agrégation optionnels qui divisent le rapport en groupes et contrôle les statistiques récapitulatives et les formats d'affichage des colonnes de rupture. Pour les variables d'agrégation multiples, il y aura un groupe séparé pour chaque catégorie de chaque variable d'agrégation à l'intérieur des catégories de la variable d'agrégation précédent dans la liste. Les variables d'agrégation doivent être des variables catégorielles discrètes qui divisent les observations en un nombre limité de catégories significatives. Les valeurs individuelles de chaque variable d'agrégation apparaissent, triées, dans une colonne séparée à gauche des colonnes de données.

Rapport : Contrôle les caractéristiques globales du rapport, y compris les statistiques récapitulatives globales, l'affichage des valeurs manquantes, la numérotation des pages et les titres.

Afficher les observations : Affiche les valeurs réelles (ou les libellés de valeurs) des variables de Variables en colonnes pour chaque observation. Cela produit une liste , qui peut être nettement plus longue qu'un rapport.

Aperçu : N'affiche que la première page du rapport. Cette option est utile pour avoir un aperçu du format de votre rapport sans traiter le rapport entier.

Les données sont déjà triées : Pour les rapports avec variables d'agrégation, le fichier de données doit être trié par valeur des variables d'agrégation avant de générer le rapport. Si votre fichier de données est déjà trié par valeur des variables d'agrégation, vous pouvez gagner du temps de traitement en sélectionnant cette option. Cette option est particulièrement utile après avoir vu un aperçu du rapport.

Obtention d'un rapport récapitulatif : Récapitulatifs en lignes

1. A partir des menus, sélectionnez :

   Analyser > **Rapports** > **Résumés des rapports dans les lignes...**
2. Sélectionnez une ou plusieurs variables pour les Colonnes de données. Une colonne est générée dans le rapport pour chaque variable sélectionnée.
3. Pour les rapports triés et affichés par sous-groupe, sélectionnez une ou plusieurs variables pour les critères d'agrégation.
4. Pour les rapports avec statistiques récapitulatives de sous-groupe définies par des variables d'agrégation, sélectionnez la variable d'agrégation dans la liste Variables de rupture et cliquez sur **Récapitulatif** dans le groupe Variables de rupture pour spécifier les mesures récapitulatives.
5. Pour les rapports avec statistiques récapitulatives globales, cliquez sur **Récapitulatif** pour spécifier les mesures récapitulatives.

Format des Colonnes de données/Rupture des rapports

Les boîtes de dialogue Format contrôlent les titres et largeurs des colonnes, l'alignement du texte et l'affichage des valeurs de données ou des libellés de valeurs. Format colonne de données contrôle le format des colonnes de données du côté droit de la page du rapport. Format colonne de rupture contrôle le format des Colonnes de rupture du côté gauche.

Titre de la colonne : Pour la variable sélectionnée, contrôle le titre de la colonne. Les titres longs sont automatiquement ajustés dans la colonne. Utilisez la touche Entrée pour insérer manuellement des retours à la ligne aux endroits où vous voulez ajuster les titres.

Position des valeurs dans la colonne : Pour la variable sélectionnée, contrôle l'alignement des valeurs de données ou des libellés de données dans la colonne. L'alignement des valeurs ou des libellés n'affecte pas l'alignement des titres de colonnes. Vous pouvez soit indenter le contenu de la colonne d'un nombre de caractères donné, soit centrer le contenu de la colonne.

Contenu de la colonne : Pour la variable sélectionnée, contrôle l'affichage soit des valeurs de données, soit des libellés de valeurs définis. Les valeurs de données sont affichées pour toutes les valeurs qui ne possèdent pas de libellé de valeur défini. (Non disponible pour les colonnes de données dans les rapports en colonnes)

Lignes récapitulatives des rapports pour/Lignes récapitulatives Finales

Les deux boîtes de dialogue Lignes récapitulatives contrôlent l'affichage des statistiques récapitulatives pour les groupes de rupture et pour l'ensemble du rapport. Lignes récapitulatives contrôle les statistiques de sous-groupe pour chaque catégorie définie par la ou les variables d'agrégation. Lignes récapitulatives Finales contrôle les statistiques globales affichées à la fin du rapport.

Les statistiques récapitulatives disponibles sont la somme des valeurs, la moyenne des valeurs, la valeur minimale, la valeur maximale, le nombre d'observations, le pourcentage d'observations situées au-dessus ou en dessous d'une valeur spécifiée, le pourcentage d'observations comprises à l'intérieur d'une plage donnée de valeurs, l'écart type, kurtosis, la variance et l'asymétrie.

Options de rupture de rapport

Les options de rupture contrôle l'espacement et la pagination des informations de catégorie de rupture.

Contrôle de page : Contrôle l'espacement et la pagination des catégories de la variable d'agrégation sélectionné. Vous pouvez spécifier un nombre de lignes vides entre les catégories de rupture ou commencer chaque catégorie de rupture sur une nouvelle page.

Lignes à sauter avant fonctions élémentaires : Contrôle le nombre de lignes vides entre les libellés ou les données des catégories de rupture et les statistiques récapitulatives. Cette option est particulièrement utile pour les rapports combinés incluant des listes d'observations individuelles et des statistiques récapitulatives pour les catégories de rupture ; dans ces rapports, vous pouvez insérer des espaces entre les listes d'observations et les statistiques récapitulatives.

Options du rapport

Options du rapport contrôle le traitement et l'affichage des valeurs manquantes et la numérotation des pages du rapport.

Exclure les observations avec des valeurs manquantes : Elimine (du rapport) toute observation avec des valeurs manquantes pour l'une des variables du rapport.

Représentation des valeurs manquantes : Vous permet de spécifier le symbole représentant les valeurs manquantes dans le fichier de données. Ce symbole ne peut comporter qu'un seul caractère et sert à représenter les *valeurs système manquantes* et les *valeurs manquantes de l'utilisateur*.

Paginer à partir de : Vous permet de spécifier un numéro pour la première page du rapport.

Présentation du rapport

Présentation du rapport contrôle la largeur et la longueur de chaque page du rapport, l'emplacement du rapport sur la page et l'insertion de lignes vides et de libellés.

Mise en page : Contrôle les marges de page exprimées en lignes (haut et bas) et en caractères (gauche et droite), et reporte l'alignement à l'intérieur des marges.

Titres et bas de page : Contrôle le nombre de lignes séparant les titres et les pieds de page du corps du rapport.

Colonne de rupture : Contrôle l'affichage des colonnes de rupture. Si des variables d'agrégation multiples sont spécifiés, ils peuvent être affichés en colonnes séparées ou dans la première colonne. Placer toutes les variables d'agrégation dans la première colonne produit un rapport plus étroit.

Titres des colonnes : Contrôle l'affichage des titres de colonnes, y compris le soulignement des titres, l'espacement entre les titres et le corps du rapport, et l'alignement vertical des titres de colonnes.

Lignes de variables en colonnes et libellés de rupture : Contrôle l'emplacement des informations de Variables en colonnes (valeurs de données et/ou statistiques récapitulatives) par rapport aux libellés de rupture au début de chaque catégorie de rupture. La première ligne des informations de Variables en colonnes peut commencer soit sur la même ligne que le libellé de catégorie de rupture, soit un nombre donné de lignes après ce libellé. (Non disponible pour les rapports en colonnes)

Titres du rapport

Titres du rapport contrôle le contenu et l'emplacement des titres et pieds de page du rapport. Vous pouvez spécifier jusqu'à dix lignes de titre et jusqu'à dix lignes de pieds de page, avec des composants justifiés à gauche, centrés et justifiés à droite sur chaque ligne.

Si vous insérez des variables dans les titres ou les pieds de page, le libellé de valeur actuel ou la valeur de la variable est affichée dans le titre ou le pied de page. Dans les titres, le libellé de valeur correspondant à la valeur de la variable au début de la page est affiché. Dans les pieds de page, le libellé de valeur correspondant à la valeur de la variable à la fin de la page est affiché. S'il n'y a aucun libellé de valeur, la valeur réelle est affichée.

Variables spéciales : Les variables spéciales *DATE* et *PAGE* vous permettent d'insérer la date actuelle ou le numéro de page dans l'une des lignes d'un en-tête ou d'un pied de page. Si votre fichier de données contient des variables nommées *DATE* ou *PAGE*, vous ne pouvez pas utiliser ces variables dans les titres ou les pieds de page des rapports.

Récapitulatifs du rapport en colonnes

Rapport en colonnes produit des rapports dans lesquels différentes statistiques récapitulatives apparaissent en colonnes séparées.

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

Exemple : Une société possédant une chaîne de magasins conserve des dossiers sur les employés comprenant le salaire, l'ancienneté et la division où chaque employé travaille. Vous pourriez générer un rapport fournissant des statistiques récapitulatives sur les salaires (par exemple moyenne, minimum, maximum) pour chaque division.

Colonnes de données : Fournit la liste des variables du rapport pour lesquelles vous voulez des statistiques récapitulatives et contrôle le format d'affichage et les statistiques récapitulatives affichées pour chaque variable.

Variables de ventilation : Fournit la liste des variables d'agrégation optionnelles qui divisent le rapport en groupes et contrôle les formats d'affichage des colonnes de rupture. Pour les variables d'agrégation multiples, il y aura un groupe séparé pour chaque catégorie de chaque variable d'agrégation à l'intérieur des catégories de la variable d'agrégation précédent dans la liste. Les variables d'agrégation doivent être des variables catégorielles discrètes qui divisent les observations en un nombre limité de catégories significatives.

Rapport : Contrôle les caractéristiques globales du rapport, y compris l'affichage des valeurs manquantes, la numérotation des pages et les titres.

Aperçu : N'affiche que la première page du rapport. Cette option est utile pour avoir un aperçu du format de votre rapport sans traiter le rapport entier.

Les données sont déjà triées : Pour les rapports avec variables d'agrégation, le fichier de données doit être trié par valeur des variables d'agrégation avant de générer le rapport. Si votre fichier de données est déjà trié par valeur des variables d'agrégation, vous pouvez gagner du temps de traitement en sélectionnant cette option. Cette option est particulièrement utile après avoir vu un aperçu du rapport.

Obtention d'un rapport récapitulatif : Récapitulatifs en colonnes

1. A partir des menus, sélectionnez :

   Analyser > **Rapports** > **Résumés des rapports dans les colonnes ...**
2. Sélectionnez une ou plusieurs variables pour les Colonnes de données. Une colonne est générée dans le rapport pour chaque variable sélectionnée.
3. Pour modifier la mesure récapitulative d'une variable, sélectionnez la variable dans la liste Variables en colonnes et cliquez sur **Récapitulatif**.
4. Pour obtenir plus d'une mesure récapitulative pour une variable, sélectionnez la variable dans la liste source et déplacez-la dans la liste Variables en colonnes plusieurs fois, une fois pour chaque mesure récapitulative que vous souhaitez.
5. Pour afficher une colonne contenant la somme, la moyenne, le rapport ou une autre fonction de colonnes existantes, cliquez sur **Insérer le total**. Une variable appelée *total* est alors placée dans la liste des colonnes de donnée.
6. Pour les rapports triés et affichés par sous-groupe, sélectionnez une ou plusieurs variables pour les critères d'agrégation.

Fonction récapitulative des colonnes de données

Lignes récapitulatives contrôle les statistiques récapitulatives affichées pour la variable de colonne de données sélectionnée.

Les statistiques récapitulatives disponibles sont la somme des valeurs, la moyenne des valeurs, la valeur minimale, la valeur maximale, le nombre d'observations, le pourcentage d'observations situées au-dessus

ou en dessous d'une valeur spécifiée, le pourcentage d'observations comprises à l'intérieur d'une plage donnée de valeurs, l'écart type, kurtosis, la variance et l'asymétrie.

Fonction élémentaire des colonnes de données pour colonne de total

Variables à récapituler contrôle les statistiques récapitulatives totales qui récapitulent deux ou plusieurs colonnes de données.

Les statistiques récapitulatives totales sont la somme des colonnes, la moyenne des colonnes, le minimum, le maximum, la différence entre les valeurs de deux colonnes, le quotient des valeurs d'une colonne divisées par les valeurs d'une autre colonne et le produit des valeurs de colonnes multipliées.

Somme des colonnes : La colonne *total* représente la somme des colonnes de la liste Variables à récapituler.

Moyenne des colonnes : La colonne *total* représente la moyenne des colonnes de la liste Variables à récapituler.

Minimum des colonnes : La colonne *total* représente la somme minimale des colonnes de la liste Variables à récapituler.

Maximum des colonnes : La colonne *total* représente la somme maximale des colonnes de la liste Variables à récapituler.

1re colonne - 2e colonne : La colonne *total* représente la différence des colonnes de la liste Variables à récapituler. La liste Variables à récapituler doit contenir exactement deux colonnes.

1re colonne / 2e colonne : La colonne *total* représente le quotient des colonnes de la liste Variables à récapituler. La liste Variables à récapituler doit contenir exactement deux colonnes.

% 1re colonne / 2e colonne : La colonne *total* représente le pourcentage de la première colonne par rapport à la seconde colonne de la liste Variables à récapituler. La liste Variables à récapituler doit contenir exactement deux colonnes.

Produit des colonnes : La colonne *total* représente le produit des colonnes de la liste Variables à récapituler.

Format des colonnes du rapport

Les options de format des colonnes de données et de rupture pour les Rapports en colonnes sont identiques à celles décrites pour les Rapports en lignes.

Rapport en Colonnes : Options de rupture

Options de rupture contrôle l'affichage des sous-totaux, l'espacement et la pagination des catégories de rupture.

Sous-total : Contrôle l'affichage des sous-totaux pour les catégories de rupture.

Contrôle de page : Contrôle l'espacement et la pagination des catégories de la variable d'agrégation sélectionné. Vous pouvez spécifier un nombre de lignes vides entre les catégories de rupture ou commencer chaque catégorie de rupture sur une nouvelle page.

Lignes à sauter avant sous-total : Contrôle le nombre de lignes vides entre les données des catégories de rupture et les sous-totaux.

Options des Rapports en Colonnes

Options contrôle l'affichage des totaux généraux, l'affichage des valeurs manquantes et la pagination dans les rapports en colonnes.

Total général : Affiche et libelle un total général pour chaque colonne ; affiché au bas de la colonne.

Valeurs manquantes : Vous pouvez exclure les valeurs manquantes du rapport ou sélectionner un caractère unique indiquant les valeurs manquantes dans le rapport.

Présentation du rapport en colonnes

Les options de présentation pour les Rapports en colonnes sont identiques à celles présentées pour les Rapports en lignes.

Fonctions supplémentaires de la commande REPORT

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

- Afficher différentes fonctions récapitulatives dans les colonnes d'une ligne récapitulative unique.
- Insérer des lignes récapitulatives dans les colonnes de données pour des variables autres que la variable de la colonne de données, ou pour diverses combinaisons (fonctions composites) de fonctions récapitulatives.
- Utiliser la Médiane, le Mode, la Fréquence et le Pourcentage comme des fonctions récapitulatives.
- Contrôler plus précisément le format d'affichage des statistiques récapitulatives.
- Insérer des lignes vides à divers emplacements du rapport.
- Insérer des lignes vides toutes les *n* observations dans les listes.

Du fait de la complexité de la syntaxe de la commande REPORT, vous trouverez peut-être utile, lorsque vous construirez un nouveau rapport avec syntaxe, d'approcher le rapport généré à partir des boîtes de dialogue, de copier et coller la syntaxe correspondante, puis de préciser cette syntaxe afin d'obtenir le rapport exact que vous souhaitez.

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

Analyse de fiabilité

L'analyse de fiabilité vous permet d'étudier les propriétés des échelles de mesure et des éléments qui les constituent. La procédure d'analyse de fiabilité calcule plusieurs mesures fréquemment utilisées de la fiabilité de l'échelle et propose également des informations sur les relations entre les différents éléments de l'échelle. Les coefficients de corrélation intra-classe peuvent être utilisés pour calculer les estimations de fiabilité entre les observateurs.

L'analyse de fiabilité fournit les statistiques du kappa de Fleiss à plusieurs observateurs qui évaluent la concordance entre les observateurs afin de déterminer la fiabilité des différents observateurs. Plus la concordance est élevée, plus les circonstances vraies reflétées par les juges peuvent être considérées comme fiables. Les options du kappa de Fleiss à juges multiples sont disponibles dans la boîte de dialogue «Analyse de fiabilité : Statistiques», à la page 274.

Exemple
Mon questionnaire mesure-t-il de façon fidèle la satisfaction de la clientèle ? L'analyse de la fiabilité vous permet de déterminer dans quelle mesure les éléments de votre questionnaire sont liés les uns aux autres et vous procure un indice général de la consistance ou de la cohérence interne de l'échelle dans son ensemble. Elle vous permet enfin d'identifier les éléments qui posent problème et qu'il faudrait exclure de l'échelle.

Statistiques
Descriptions de chaque variable et pour l'échelle, statistiques récapitulatives sur les éléments, corrélations et covariances entre éléments, prévisions de fiabilité, tableau d'ANOVA, coefficients de corrélation intra-classe, T^2 d'Hotelling, test d'additivité de Tukey et kappa de Fleiss à plusieurs observateurs.

Modèles
Les modèles suivants de fiabilité sont disponibles :

Alpha (Cronbach)
Ce modèle est une mesure de cohérence interne basée sur la corrélation moyenne entre éléments.

Coefficient Omega (McDonald)

Ce modèle suppose que le modèle est unidimensionnel, avec un seul facteur sans dépendance d'élément local sous forme de covariances d'erreur. Le modèle implique que la covariance des deux éléments différents est le produit de leurs chargements.

Bipartition
Ce modèle fractionne l'échelle en deux et examine la corrélation entre les deux parties.

Guttman
Ce modèle calcule les limites minimales de Guttman pour une fiabilité vraie.

Parallèle
Ce modèle part de l'hypothèse que tous les éléments ont des variances égales et des variances d'erreur égales en cas de réplication.

Parallèle strict
Ce modèle se fonde sur les mêmes hypothèses que le modèle parallèle mais envisage également que tous les éléments ont la même moyenne.

Remarques sur les données dans l'analyse de fiabilité

Données
Les données peuvent être dichotomiques, ordinales ou constituer des intervalles, mais elles doivent être codées en numérique.

Hypothèses
Les observations doivent être indépendantes, les erreurs ne doivent pas être corrélées entre éléments. Chaque paire d'éléments doit avoir une distribution gaussienne bivariée. Les échelles doivent être additives, de sorte que chaque élément est linéairement relié au total. Les suppositions suivantes s'appliquent aux statistiques du kappa de Fleiss à plusieurs observateurs :

- Vous devez sélectionner au moins deux variables d'élément pour exécuter n'importe quelle statistique de fiabilité.
- Lorsqu'au moins deux variables d'évaluation sont sélectionnées, la syntaxe du kappa de Fleiss à plusieurs observateurs est collée.
- Il n'existe aucun lien entre les observateurs.
- Le nombre d'observateurs est une constante.
- Chaque sujet est évalué par le même groupe contenant un seul observateur.
- Aucune pondération ne peut être affectée aux différences.

Procédures apparentées
Si vous souhaitez explorer la dimensionnalité des éléments de votre échelle (pour voir si plusieurs éléments de base sont nécessaires au motif des calculs), utilisez Analyse factorielle ou Positionnement multidimensionnel. Pour identifier des groupes homogènes de variables, utilisez l'analyse de cluster hiérarchique pour classer les variables.

Obtention d'une analyse de fiabilité

1. A partir des menus, sélectionnez :

   Analyser > **Echelle** > **Analyse de fiabilité...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez deux variables (éléments) au moins en tant que composants potentiels d'une échelle additive.
3. Sélectionnez un modèle dans la liste déroulante **Modèle**.
4. Si vous le souhaitez, vous pouvez cliquer sur **Statistiques** pour sélectionner différentes statistiques décrivant vos éléments d'échelle ou la concordance entre les observateurs.

Analyse de fiabilité : Statistiques

Vous pouvez sélectionner différentes statistiques décrivant votre échelle, vos éléments et l'accord entre observateurs pour déterminer la fiabilité entre les différents observateurs. Les statistiques émises par défaut regroupent le nombre d'observations, le nombre d'éléments et les prévisions de fiabilité de la façon suivante :

Modèles alpha
Coefficient alpha ; pour les données dichotomiques, il s'agit d'un équivalent du coefficient Kuder-Richardson 20 (KR20).

Modèles Oméga
Estimation du coefficient Omega de McDonald pour évaluer la fiabilité.

Modèles de bipartition
Corrélation entre les sous-échelles, fiabilité Split-half de Guttman, fiabilité de Spearman-Brown (longueur égale ou inégale) et coefficient alpha pour chaque moitié.

Modèles de Guttman
Coefficients de fiabilité lambda 1 à lambda 6.

Modèles Parallèle et Parallèle strict
Test de qualité de l'ajustement du modèle, estimation de la variance de l'erreur, variance commune et réelle, estimation de la corrélation commune entre éléments, estimation de la fiabilité et estimation de la fiabilité non biaisée.

Caractéristiques pour
Produit des statistiques descriptives pour les échelles ou les éléments sur les observations.

Item
Produit des statistiques descriptives pour les éléments sur les observations.

Echelle
Produit des statistiques descriptives pour les échelles.

Echelle sans l'élément
Affiche les statistiques récapitulatives en comparant chaque élément à l'échelle composée des autres éléments. Les statistiques incluent la moyenne et la variance de l'échelle si l'élément a été supprimé de l'échelle, la corrélation entre l'élément et l'échelle composée des autres éléments, et l'alpha de Cronbach si l'élément a été supprimé de l'échelle.

Récapitulatifs
Fournit des statistiques descriptives de la distribution des éléments sur l'ensemble des éléments dans l'échelle.

moyennes
Statistiques récapitulatives pour les moyennes d'élément. Les moyennes d'élément minimale, maximale et moyenne sont affichées, ainsi que le rapport de la moyenne maximale à la moyenne minimale.

Variances
Statistiques récapitulatives pour les variances d'élément. Les valeurs de variance maximale, minimale et moyenne sont affichées, ainsi que la plage et la variance des variances d'élément, et le rapport entre les variances d'élément maximale et minimale.

Corrélations
Statistiques récapitulatives pour les corrélations inter éléments. Les corrélations entre éléments minimale, maximale et moyenne sont affichées, ainsi que la plage et la variance des corrélations entre éléments, et le rapport de la corrélation entre éléments maximale à la corrélation minimale.

Covariances
Statistiques récapitulatives pour les covariances inter éléments. Les covariances entre éléments minimale, maximale et moyenne sont affichées, ainsi que la plage et la variance des covariances entre éléments, et le rapport de la covariance entre éléments maximale à la covariance minimale.

Cohérence inter-éléments
Produit des matrices de corrélations et de covariances entre éléments.

Tableau ANOVA

Produit des tests de moyennes égales.

Test F

Affiche un tableau d'analyse de variance des mesures répétées.

Khi-deux de Friedman

Affiche le khi-carré de Friedman et le coefficient de concordance de Kendall. Cette option convient aux données organisées sous forme de rangs. Le test du khi-deux remplace le test F habituel dans le tableau ANOVA.

Khi-deux de Cochran

Affiche le Q de Cochran. Cette option est appropriée pour les données dichotomiques. Le Q de Cochran remplace le test F habituel dans le tableau ANOVA.

Accord inter-juges : Kappa de Fleiss

Evalue l'accord inter-observateurs pour déterminer la fiabilité des différents observateurs. Plus la concordance est élevée, plus les circonstances vraies reflétées par les juges peuvent être considérées comme fiables. Les statistiques du kappa sans pondération généralisées mesurent la concordance entre n'importe quel nombre constant d'observateurs dans les conditions suivantes :

- Vous devez spécifier au moins deux variables d'élément pour exécuter n'importe quelle statistique de fiabilité.
- Vous devez spécifier au moins deux variables d'évaluation.
- Les variables sélectionnées comme éléments peuvent également être sélectionnées comme évaluations.
- Il n'existe aucun lien entre les observateurs.
- Le nombre d'observateurs est une constante.
- Chaque sujet est évalué par le même groupe contenant un seul observateur.
- Aucune pondération ne peut être affectée aux différences.

Affichage de l'accord sur des catégories individuelles

Indique si la sortie de l'accord doit être affiché sur les catégories individuelles. Par défaut, la sortie supprime l'estimation sur toutes les catégories individuelles. Lorsque cette option est activée, plusieurs tables s'affichent dans la sortie.

Ignorer les cas de chaînes

Contrôle si les variables de chaîne respectent la distinction majuscules-minuscules. Par défaut, les valeurs d'évaluation de chaîne respectent cette distinction.

Les libellés de catégorie de chaîne s'affichent en majuscules

Contrôle si les libellés de catégorie dans les tables de sortie s'affichent en majuscules ou en minuscules. Ce paramètre est activé par défaut et les libellés de catégorie de chaîne s'affichent en majuscules.

Niveau d'importance asymptotique (%)

Indique le niveau d'importance des intervalles de confiance asymptotique. La valeur par défaut est 95.

Manquant

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

Contrôle l'exclusion des valeurs manquantes de l'utilisateur et des valeurs système manquantes. Par défaut, les valeurs manquantes de l'utilisateur et les valeurs système manquantes sont exclues.

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

Lorsque cette option est activée, les valeurs manquantes de l'utilisateur et les valeurs système manquantes sont considérées comme des données valides. Il est désactivé par défaut.

T-carré de Hotelling

Produit un test multivarié basé sur l'hypothèse nulle que tous les éléments sur l'échelle ont la même moyenne.

Test d'additivité de Tukey
Produit un test basé sur l'hypothèse d'absence d'interaction multiplicative entre les éléments.

Coefficient de corrélation intra-groupe
Produit des mesures d'homogénéité ou de cohérence des valeurs par observation.

Modèle
Sélectionnez le modèle de calcul du coefficient de corrélation intra-classe. Les modèles disponibles sont Mixte à deux facteurs, Aléatoire à deux facteurs et Aléatoire à un facteur. Sélectionnez **Mixte à deux facteurs** lorsque les effets de population sont aléatoires et les effets d'éléments sont fixes, sélectionnez **Aléatoire à deux facteurs** lorsque les effets de population et les effets d'éléments sont aléatoires, ou sélectionnez **Aléatoire à un facteur** lorsque les gens effectuent un facteur.

Type
Sélectionnez le type d'index. Les types disponibles sont Homogénéité et Cohérence absolue.

Intervalle de confiance (%)
Spécifiez le niveau de l'intervalle de confiance. La valeur par défaut est 95 %.

Valeur test
Spécifiez la valeur hypothétique du coefficient pour le test d'hypothèse. Il s'agit de la valeur par rapport à laquelle la valeur observée est comparée. La valeur par défaut est 0.

Fonctions supplémentaires de la commande RELIABILITY

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

- Lire et analyser une matrice de corrélation.
- Enregistrer une matrice de corrélation à analyser ultérieurement.
- Spécifier un fractionnement autre qu'en deux moitiés égales quant au nombre d'éléments pour la méthode de la bipartition.

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

Kappa pondéré

Le Kappa pondéré de Cohen est largement utilisé dans la classification croisée comme consensus entre les observateurs. Il s'agit d'un indice d'accord approprié si les évaluations sont des échelles nominales sans structure d'ordre. Le développement du kappa pondéré de Cohen a été motivé par le fait que la gravité de certaines affectations d'un tableau de contingence pouvait être supérieure à celle d'autres affectations. Ces statistiques s'appuient sur les pondérations de cellule prédéfinies qui reflètent un accord ou un désaccord.

La procédure du kappa pondéré fournit des options permettant d'estimer le kappa pondéré de Cohen, une généralisation importante de la statistique du kappa qui mesure l'accord de deux sujets ordinaux avec des catégories identiques.

Remarque : La procédure Kappa pondéré remplace la fonctionnalité précédemment fournie par l'extension `STATS WEIGHTED KAPPA.spe`.

Exemple
Dans certains cas, les différences entre les observateurs ne doivent pas être traitées de manière égale. C'est notamment le cas dans le secteur de la santé où de nombreuses personnes collectent des données de recherche ou des données cliniques. Dans ces cas là, la fiabilité des données peut être remise en question étant donnée la variabilité entre les personnes qui les collectent.

Statistiques
Kappa pondéré de Cohen, échelle linéaire, échelle quadratique, intervalle de confiance asymptotique

Remarques sur les données du Kappa pondéré

Données

Un tableau à deux entrées basé sur un jeu de données actif est requis pour évaluer la statistique du kappa pondéré de Cohen.

Les variables d'évaluation doivent être de même type (chaîne ou valeur numérique).

L'estimation du kappa pondéré de Cohen n'a d'intérêt que si les catégories des deux variables d'évaluation, représentées par la ligne et la colonne dans le tableau, sont classées correctement (pour une paire de variables numériques, l'ordre numérique est appliqué ; pour une paire de variables de chaîne, l'ordre alphabétique est appliqué).

Hypothèses

Si des paires de variables mixtes sont sélectionnées, le kappa pondéré de Cohen n'est pas évalué.

Les variables d'évaluation sont supposées partager le même ensemble de catégories.

Obtention d'une analyse du Kappa pondéré

1. A partir des menus, sélectionnez :

   Analyser > **Échelle** > **Kappa pondéré...**

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

2. Sélectionnez plusieurs variables de chaîne ou numériques à spécifier comme **Indicateurs appariés**.

   Remarque : Vous ne devez sélectionner que des variables de chaîne ou que des variables numériques.

3. Activez éventuellement le paramètre **Spécifier des indicateurs pour des lignes et des colonnes** pour contrôler l'affichage des indicateurs appariés ou des indicateurs de ligne/colonne.

   - Si ce paramètre est activé, les indicateurs appariés sont supprimés et les indicateurs de ligne/colonne sont affichés. L'interface utilisateur est mise à jour pour afficher les zones **Indicateur(s) de ligne** et **Indicateur(s) de colonne** (qui remplacent la zone **Indicateurs appariés**).
   - Si ce paramètre est désactivé, les indicateurs de ligne/colonne sont supprimés et les indicateurs appariés sont affichés (valeur par défaut)

   Si **Spécifier des indicateurs pour des lignes et des colonnes** est activé, spécifiez au moins une variable pour **Indicateur(s) de ligne** et **Indicateur(s) de colonne**.

   Remarque : Si **Indicateur(s) de ligne** et **Indicateur(s) de colonne** ne contiennent tous deux qu'une seule variable, les variables sélectionnées ne peuvent pas être identiques.

4. Cliquez éventuellement sur **Critères** pour spécifier les paramètres d'échelle de pondération et de valeurs manquantes ou sur **Imprimer** pour spécifier les paramètres de format d'affichage et de tableau croisé.

Kappa pondéré : Critères

La boîte de dialogue Critères propose des options pour spécifier l'estimation de la statistique du kappa pondéré de Cohen.

Echelle de pondération

Propose des options permettant de spécifier une pondération linéaire ou quadratique pour l'accord. L'utilisation de la pondération linéaire (également connue sous le nom de Cicchetti-Allison) est le paramètre par défaut.

Valeurs manquantes

Propose des options permettant de retirer les observations avec des valeurs manquantes par paire et de traiter les valeurs manquantes de l'utilisateur comme valides.

Les variables d'évaluation de chaîne sont sensibles à la casse

Si cette option est sélectionnée, les variables de chaîne sont traitées comme étant sensibles à la casse.

Intervalle de confiance asymptotique (%)
Ce paramètre facultatif spécifie le niveau de confiance de l'estimation des intervalles de confiance asymptotiques. Il doit s'agir d'une valeur double unique comprise entre 0 et 100 (la valeur par défaut est 95).

Kappa pondéré : Imprimer

La boîte de dialogue Imprimer propose des options pour contrôler les tableaux croisés.

Afficher et formater
Propose des options permettant de contrôler l'affichage et le format des tableaux croisés.

Les catégories d'évaluation sont affichées par ordre croissant
Si ce paramètre est sélectionné, les catégories d'évaluation des tableaux croisés sont affichées suivant l'ordre croissant. Si ce paramètre n'est pas sélectionné, les catégories d'évaluation des tableaux croisés sont affichées suivant l'ordre décroissant. Ce paramètre est activé par défaut.

Les libellés de catégorie de chaîne s'affichent en majuscules
Si ce paramètre est sélectionné, les tableaux croisés sont affichés en majuscules. Si ce paramètre n'est pas sélectionné, les tableaux croisés sont affichés en minuscules. Ce paramètre est activé par défaut.

Tableau croisé
Propose des options permettant de spécifier les variables d'évaluation utilisées dans les tableaux croisés. Par défaut, les paramètres des tableaux croisés ne sont pas activés, ce qui supprime le tableau croisé des éventuelles variables d'évaluation.

Afficher le tableau croisé des variables d'évaluation
Si ce paramètre est sélectionné, il active le tableau croisé de toutes les variables d'évaluation ou de celles spécifiées par l'utilisateur.

Inclure toutes les variables d'évaluation
Si ce paramètre est sélectionné, les tableaux croisés sont imprimés pour toutes les paires de variables d'évaluation définies.

Inclure des variables d'évaluation spécifiées par l'utilisateur
Si ce paramètre est sélectionné, utilisez les zones **Variables disponibles**, **Indicateur(s) de ligne** et **Indicateur(s) de colonne** pour sélectionner les variables d'évaluation incluses dans les tableaux croisés.

Positionnement multidimensionnel

Le positionnement multidimensionnel tente de déterminer une structure dans un ensemble de mesures de distance entre objets ou observations. Pour cela, il affecte les observations à des positions particulières dans un espace conceptuel (à deux ou trois dimensions généralement) de telle sorte que les distances entre les points dans l'espace correspondent le mieux possible aux dissimilarités données. Dans la plupart des cas, les dimensions de cet espace conceptuel peuvent être interprétées et utilisées pour mieux comprendre les données.

Si vous avez mesuré objectivement les variables, vous pouvez utiliser le positionnement multidimensionnel comme technique de réduction de données (le Positionnement multidimensionnel calcule pour vous les distances à partir des données multivariées, le cas échéant). Le positionnement multidimensionnel peut également s'appliquer à des classements subjectives de dissimilarité entre objets ou concepts. D'autre part, le positionnement multidimensionnel peut gérer les données de dissimilarité provenant de plusieurs sources, comme c'est le cas lorsqu'il y a plusieurs indicateurs ou plusieurs répondants au questionnaire.

Exemple. Comment les gens perçoivent-ils les relations entre différentes voitures ? Si les données que vous obtenez de vos répondants indiquent des classements de similarité entre différentes modèles, le positionnement multidimensionnel peut servir à identifier les dimensions qui décrivent les perceptions des consommateurs. Vous pouvez trouver, par exemple, que le prix et la taille du véhicule définissent un espace à deux dimensions qui tient compte des similarités reportées par les répondants.

Statistiques : Pour chaque modèle, vous disposez de la matrice des données, du positionnement optimisé des données de la matrice, de la contrainte S (de Young), de la contrainte S (de Kruskal), de RSQ, des coordonnées de stimulus, de la contrainte moyenne et de RSQ pour chaque stimulus (modèles RMDS). Pour les modèles des différences individuelles (INDSCAL), vous disposez des pondérations de sujet et de l'indice de singularité pour chaque objet. Pour les matrices situées dans les modèles de positionnement multidimensionnel répliqués, vous disposez de la contrainte et de RSQ pour chaque stimulus. Pour les tracés, vous disposez des coordonnées de stimulus (à deux ou trois dimensions), du nuage de points des disparités par rapport aux distances.

Remarques sur les données du positionnement multidimensionnel

Données. Si vos données sont dissemblables, toutes les dissemblances doivent être quantitatives et mesurées avec les mêmes unités et échelles. Si vos données sont multivariées, les variables peuvent être quantitatives, binaires mais peuvent aussi être des données d'effectif. Le positionnement des variables est un enjeu de taille : les différences de positionnement peuvent affecter votre solution. Si vos variables présentent de grandes différences de positionnement (par exemple, si une variable est mesurée en dollar et l'autre en année), vous devez envisager de les standardiser (et cela automatiquement par la procédure de positionnement multidimensionnel).

Hypothèses : La procédure de positionnement multidimensionnel est relativement indépendante de toute hypothèse de distribution. Assurez-vous que vous avez sélectionné le niveau de mesure approprié (ordinal, intervalle ou rapport) dans la boîte de dialogue Positionnement multidimensionnel : Options afin de garantir la justesse des résultats.

Procédures apparentées : Si votre but est la réduction de données, vous pouvez également envisager l'analyse factorielle, plus particulièrement si vos données sont quantitatives. Si vous souhaitez identifier des groupes d'observations similaires, envisagez de compléter votre analyse par positionnement multidimensionnel avec une analyse de cluster de *nuées dynamiques* ou une analyse de la classification hiérarchique.

Obtenir une analyse par positionnement multidimensionnel

1. A partir des menus, sélectionnez :

   Analyse > **Echelle** > **Echelle multidimensionnelle...**
2. Sélectionnez au moins quatre variables numériques pour l'analyse.
3. Dans le groupe Distances, sélectionnez **Données en matrice(s)** ou **Calculées à partir des données**.
4. Si vous avez sélectionné **Calculées à partir des données**, vous pouvez également sélectionner une variable de regroupement pour les matrices individuelles. La variable de regroupement peut être numérique ou être une variable de chaîne.

Eventuellement, vous pouvez aussi :

- Indiquez la forme de la matrice lorsque les données sont des distances.
- Spécifiez la mesure de la distance à utiliser lors de la création de distances à partir des données.

Utilisation de la cartographie de proximité pour la mise à l'échelle multidimensionnelle

Outre ALSCAL, [IBM SPSS Statistiques ...], propose **Proximity Mapping (PROXMAP)** comme approche connexe à la mise à l'échelle multidimensionnelle. ALSCAL et PROXMAP diffèrent dans la manière dont l'ajustement du modèle est calculé. ALSCAL ajuste les distances au carré aux dissimilarités au carré, tandis que PROXMAP ajuste directement les distances aux dissimilarités. Par conséquent, ces deux procédures peuvent donner des valeurs numériques différentes, même lorsque les configurations spatiales obtenues semblent similaires. Ces différences peuvent être plus perceptibles lorsque vous travaillez avec des données pour lesquelles il n'existe pas de solution parfaite.

Alors qu'ALSCAL propose un scalage multidimensionnel de base avec des résultats textuels limités et un minimum d'informations diagnostiques, PROXMAP offre des solutions modernes et complètes comprenant des visualisations relativement précises (notamment des diagrammes de Shepard, des graphiques de transformation et une analyse des résidus), des résultats statistiques améliorés (mesures

multiples de tension, valeurs R^2 et indices d'ajustement détaillés), ainsi que des fonctionnalités avancées telles que l'ajustement des propriétés.

Consultez les liens suivants pour vous lancer dans la cartographie de proximité :

- «Comparaison entre ALSCAL et PROXMAP», à la page 281
- Mappage de proximité
-
-
-

Forme des données du positionnement multidimensionnel

Si votre jeu de données actif représente les distances au sein d'un ensemble d'objets ou entre deux ensembles d'objets, vous devez indiquer la forme de votre matrice de données afin d'obtenir des résultats corrects.

Remarque : Vous pouvez sélectionner **Carré symétrique** si la boîte de dialogue Modèle indique une conditionnalité de ligne.

Positionnement multidimensionnel : créer une mesure

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-carré ou Distance du phi-carré.
- **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 peut-être standardiser des valeurs avant de calculer les proximités (ne s'applique pas aux données binaires). Sélectionnez une méthode de standardisation dans la liste déroulante. Si aucune standardisation n'est requise, choisissez **Aucune**.

Modèle de positionnement multidimensionnel

Une estimation correcte d'un modèle de positionnement multidimensionnel dépend des aspects des données et du modèle lui-même.

Niveau de mesure : Vous permet de spécifier le niveau de vos données. Les possibilités sont Ordinales, Intervalle ou Rapport. Si vos variables sont ordinales, la sélection de l'option **Délier des observations liées** demande qu'elles soient traitées en tant que variables continues, de telle sorte que les liens (mêmes valeurs pour des observations différentes) soient résolus de manière optimale.

Conditionnalité : Vous permet de spécifier les comparaisons pertinentes. Les possibilités sont Matrice, Ligne et Inconditionnel.

Dimensions : Vous permet de spécifier la dimensionnalité de la ou des solutions de positionnement. Une seule solution est calculée pour chaque nombre de la plage. Indiquez des nombres entiers entre 1 et 6. La valeur minimale 1 n'est autorisée que si vous sélectionnez l'option **Distance Euclidienne** comme modèle

de positionnement. Pour n'obtenir qu'une seule solution, indiquez le même nombre en tant que minimum et maximum.

Modèle de positionnement : Vous permet de spécifier les hypothèses sous lesquelles le positionnement est effectué. Les possibilités existantes sont Distance Euclidienne ou Distance Euclidienne des différences individuelles (connue également en tant que INDSCAL). Pour le modèle Distance Euclidienne des différences individuelles, vous pouvez sélectionner l'option **Permet la pondération de sujet négative**, si cela convient à vos données.

Positionnement multidimensionnel : Options

Vous pouvez spécifier les options de votre analyse par positionnement multidimensionnel.

Afficher. Vous permet d'afficher les différents types d'affichage. Les options possibles sont Tracés des stimuli, Tracés des sujets, Matrice des données et Récapitulatif des options du modèle.

Critères : Vous permet de déterminer quand l'itération doit s'interrompre. Pour modifier les valeurs par défaut, entrez des valeurs pour la **Convergence du S-stress**, le **Minimum du s-stress** et le **Maximum des itérations**.

Traiter les dissimilarités inférieures à n comme des valeurs manquantes : Ces distances sont exclues de l'analyse.

Fonctions supplémentaires de la commande ALSCAL

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

- Utiliser trois types de modèle supplémentaires, ASCAL, AINDS et GEMSCAL dans la documentation relative au Positionnement multidimensionnel.
- Effectuer des transformations polynomiales sur l'intervalle et les données de type rapport.
- Analyser les similarités (plutôt que les distances) avec des données ordinales.
- Analyser les données nominales.
- Enregistrer diverses matrices de coordonnées et de pondération dans des fichiers et les relire pour l'analyse.
- Contraindre le dépliage multidimensionnel.

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

Comparaison entre ALSCAL et PROXMAP

ALSCAL et PROXMAP effectuent tous deux un recourbement multidimensionnel (MDS) afin de représenter les données de proximité dans un nombre donné de dimensions. Dans PROXMAP, ces résultats sont présentés à l'aide d'une terminologie actualisée et de graphiques de diagnostic qui permettent une comparaison directe avec les résultats d'ALSCAL.

ALSCAL et PROXMAP estiment la disposition spatiale des objets à partir de données de proximité. PROXMAP offre un cadre basé sur des boîtes de dialogue qui prend en charge les mêmes résultats conceptuels qu'ALSCAL, en utilisant une terminologie actualisée et des graphiques de diagnostic harmonisés.

Cartographie conceptuelle

Le tableau suivant présente la correspondance entre les principaux termes et résultats d'ALSCAL et la terminologie de PROXMAP. La signification analytique des résultats reste la même quelle que soit la méthode utilisée.

Tableau 5. Correspondance terminologique entre ALSCAL et PROXMAP

Terme ou résultat ALSCAL	**Équivalent de PROXMAP**
Configuration stimulus dérivé	Espace d'objet commun

Tableau 5. Correspondance terminologique entre ALSCAL et PROXMAP (suite)	
Terme ou résultat ALSCAL	**Équivalent de PROXMAP**
Nuage de points d'ajustement linéaire Disparités Distances	**Fit Source <Étiquette de la source>** Proximités transformées Distances
Nuage de points d'ajustement non linéaire Observations Distances	**Tracé de Shepard[1]Source <Étiquette de source>** Proximités d'origine Distances + Proximités transformées
Nuage de points de transformation Observations Disparités	**Source de transformation <Étiquette de la source>** Proximités d'origine Proximités transformées

Passage d'ALSCAL à PROXMAP

Le tutoriel suivant montre comment reproduire une analyse de scalage multidimensionnel (MDS) ALSCAL à l'aide de PROXMAP, et explique comment interpréter les similitudes et les différences dans les résultats.

Présentation

Ce tutoriel montre comment configurer PROXMAP pour reproduire une analyse ALSCAL et comment comparer les configurations obtenues.

Cet exemple se concentre sur un seul ensemble de données de proximité et utilise les paramètres par défaut ou des paramètres équivalents, de sorte que les différences entre ALSCAL et PROXMAP se limitent à la mise à l'échelle et à l'orientation, plutôt qu'à des modifications substantielles de la solution.

Données utilisées dans l'analyse

L'analyse s'appuie sur fichier Dist_Cities_LowerTr.sav, un ensemble de données de proximité représentant les distances de vol entre douze villes des États-Unis. Les données sont stockées sous la forme d'une matrice de proximité symétrique en forme triangulaire inférieure.

Dans Data View, l'ensemble de données comprend une matrice de proximité symétrique de type « 12x12 » stockée sous forme triangulaire inférieure, ainsi qu'une variable d'identifiant d'objet nominal utilisée pour nommer la configuration.

Dans la vue « Variable View », l'ensemble de données se compose de douze variables de niveau d'échelle contenant la matrice de proximité et d'une variable nominale définissant les étiquettes des objets.

La variable *CITY* est définie comme variable d'étiquette d'objet.

Outre la matrice de proximité, l'ensemble de données comprend une variable d'identifiant d'objet nominal (*CITY*) qui attribue des étiquettes aux objets de la configuration.

Les valeurs de *CITY* vont de 1 à 12. Des étiquettes de valeur sont attribuées à ces nombres entiers, où :

- 1 = ATLANTA
- 2 = BOSTON

[1] Le terme « diagramme de Shepard » fait référence à Roger Shepard, l'un des pionniers de l'analyse factorielle multidimensionnelle non métrique. Il combine le graphique « Ajustement non linéaire » et le graphique « Transformation » d'ALSCAL.

- 3 = CINCINATI
- 4 = COLUMBUS
- 5 = DALLAS
- 6 = INDNPLIS
- 7 = LITTROCK
- 8 = LOSANGEL
- 9 = MEMPHIS
- 10 = Saint-Louis
- 11 = SPOKANE
- 12 = TAMPA

PROXMAP utilise ces étiquettes de valeur pour identifier les objets dans les tableaux et les représentations graphiques, garantissant ainsi que ce sont les noms des objets qui s'affichent à la place des codes numériques.

Configuration de l'analyse PROXMAP

Accéder à **la cartographie** > **de** > **proximité d'AnalyzeMapping**

Les variables de la matrice de proximité sont définies comme une source de proximité unique dans la boîte de dialogue « Proximity Mapping ». PROXMAP accepte les données de proximité stockées soit sous forme de matrice triangulaire inférieure (diagonale comprise), soit sous forme de matrice symétrique complète, à condition que la matrice soit correctement définie.

Onglet « Données » (Définition de la matrice de proximité et des étiquettes d'objets)
Définir la source de proximité :

1. Sélectionnez les variables *ATLANTA* à *TAMPA* dans la liste des variables et déplacez-les vers **la zone « Variable(s) » dans la source**.
2. Indiquez un nom dans le champ « **Étiquette source** » (par exemple, « *Distances de vol entre les villes* »).
3. Cliquez sur « **Ajouter une source** » pour créer une source (par exemple, *$SRC001*).
4. Dans la section « **Type de données** », sélectionnez « **Données de proximité** ».
5. Déplacer la variable *CITY* vers **les étiquettes d'objet**

Le nombre actuel d'objets = 12 correspond au nombre d'objets définis par la matrice de proximité et confirme que douze villes seront représentées dans la configuration.

Onglet Modèle (Définition du modèle PROXMAP)
L'analyse repose sur une seule source de proximité ($SRC001); veuillez donc sélectionner les options suivantes :

1. Dans **le modèle à trois voies**, sélectionnez l'option **Identité**.
2. Dans la section « **Conditionnalité / Transformation** », sélectionnez **« Sans condition »**
3. Cliquez sur « **Définir la transformation** » et sélectionnez les options « **Ordinal** » et « **Autoriser la résolution des ex aequo** ».

   Cela signifie que les proximités sont considérées comme des dissemblances ordinales. L'ordre de classement des proximités est conservé, et les valeurs identiques dans les données restent à égalité dans la transformation monotone.

   Si nécessaire, vous pouvez modifier le type de transformation (par exemple, linéaire, puissance, spline, etc.).

Le nombre actuel de dimensions = 2 indique qu'une configuration bidimensionnelle sera estimée.

Cette configuration correspond aux spécifications utilisées pour la comparaison avec ALSCAL.

Onglet « Critères » (Définition de la dimensionnalité, de la convergence et de la normalisation)

1. **Dimensionnalité** : le **nombre de dimensions** est défini sur *2*

   Cela indique qu'une configuration bidimensionnelle sera estimée.

2. **Convergence de l'algorithme** : les critères de convergence régissent le processus d'optimisation itératif. Les paramètres suivants sont définis :

   - **Fonction Minimum :** 0.000001
   - **Différence de fonctionnalités :** 0.000000000001
   - **Différence de configuration :** 0.000001
   - **Nombre maximal d'itérations :** 1 000

   Ces valeurs déterminent à quel moment l'algorithme itératif s'arrête. La convergence est atteinte lorsque l'amélioration de la fonction de perte et les variations de la configuration tombent en dessous des tolérances spécifiées, ou lorsque le nombre maximal d'itérations est atteint.

3. **Normalisation** : choisir « **Somme des carrés = N** ». Cette option est particulièrement utile pour la normalisation des données multivariées.

4. **Dimension** : sélectionner **par dimension**. Cette option vous permet d'afficher les dimensions individuelles. Cochez la case située sous « **Reflection** » pour la dimension 2.

   Les réflexions n'affectent que l'orientation (le signe) des axes et ne modifient ni les distances ni l'adéquation du modèle. Ils servent généralement à aligner la solution sur une configuration de référence (par exemple, lors d'une comparaison avec une solution ALSCAL).

Onglet « Résultats » : sous-onglet « Modèle » (Demande de tableaux et de graphiques)

1. Dans la section « Tables », sélectionnez **« Coordonnées »** afin qu'un tableau contenant les coordonnées de l'objet dans l'espace commun soit généré dans le fichier de sortie.

   Les options « **Directions variables** » et « **Pondérations spatiales** » ne sont pas sélectionnées, car cette analyse repose sur une seule source de proximité et ne fait pas intervenir de sources supplémentaires ni de variables multivariées.

2. Dans la rubrique « Graphiques », sous la section « Espace commun », sélectionnez « **Configuration** ».

   Cela génère un graphique représentant la configuration de l'objet pour l' $SRC001 de source (par exemple, les distances de vol entre les villes), avec des étiquettes correspondant à la variable spécifiée dans l'onglet « **Données** » (par exemple, *CITY*).

3. **Sorties supplémentaires** : dans la liste **Sources**, déplacez « *$SRC001* » vers **Transformation**, **Fit** et **Shepard**.

   Les graphiques suivants sont générés :

   - Un graphique représentant **les proximités transformées** sur l'axe vertical et les proximités d'origine sur l'axe horizontal.
   - Un nuage de points représentant les **distances ajustées** dans la configuration en fonction des **proximités transformées**.
   - Un nuage de points de diagnostic qui affiche les distances ajustées, ainsi que les proximités transformées par rapport aux proximités d'origine, afin de visualiser la qualité de l'ajustement dans le cadre d'un recouvrement multidimensionnel non métrique (cartographie de proximité).

On demande une solution bidimensionnelle, correspondant à la dimensionnalité utilisée dans l'analyse ALSCAL.

Syntaxe de commande

Consultez également la syntaxe de commande suivante, utilisée pour l'analyse à l'aide de ALSCAL et PROXMAP.

Syntaxe de commande utilisée pour l'analyse avec ALSCAL

```
ALSCAL
  VARIABLES=ATLANTA BOSTON CINCNATI COLUMBUS DALLAS INDNPLIS LITTROCK LOSANGEL MEMPHIS
STLOUIS
    SPOKANE TAMPA
  /SHAPE=SYMMETRIC
  /LEVEL=ORDINAL
  /CONDITION=MATRIX
  /MODEL=EUCLID
  /CRITERIA=CONVERGE(0.001) STRESSMIN(0.005) ITER(30) CUTOFF(0) DIMENS(2,2)
  /PLOT.
```

Syntaxe des commandes utilisée pour l'analyse avec PROXMAP

```
PROXMAP
  /SOURCEID NAME=$SRC001 LABEL='Cities flying distances'  DATA=ATLANTA BOSTON CINCNATI
COLUMBUS DALLAS INDNPLIS
    LITTROCK LOSANGEL MEMPHIS STLOUIS SPOKANE TAMPA
  /DATA PROXIMITIES
  /OBJECTID LABELS=CITY
  /CONDITION UNCONDITIONAL
  /TRANSFORMATIONS SOURCES=$SRC001 FUNCTION=ORDINAL APPROACH=KEEPTIES
  /MODEL IDENTITY
  /CRITERIA DIMENSIONS=2 MAXIMUM=1000 MINIMUM=0.000001 RELATIVE=0.000000001
ABSOLUTE=0.000001
    STANDARDIZATION='N' REFLECTION=DIMENSIONS NUMBERS=1 -1
  /PRINT DATA(RAW) MODEL (COORDINATES) CRITERIA (SUMMARY ITERATIONS)
  /PLOT MODEL (CONFIGURATION TRANSFORMATION ($SRC001) FIT ($SRC001) SHEPARD ($SRC001)).
```

Graphiques de sortie et de diagnostic

Le résultat comprend un tableau des coordonnées des objets dans l'espace commun des objets ainsi qu'un graphique de configuration illustrant la représentation spatiale des villes.

Des graphiques de diagnostic supplémentaires sont demandés, notamment des graphiques de transformation, des graphiques d'ajustement et des graphiques de Shepard. Ces graphiques illustrent les relations entre les proximités initiales, les proximités transformées et les distances ajustées, et fournissent des informations visuelles sur la qualité de l'ajustement de la solution.

La terminologie PROXMAP diffère de celle d'ALSCAL. Par exemple, les stimuli ALSCAL sont appelés « objets » dans PROXMAP, et les disparités ALSCAL sont appelées « proximités transformées ». Le graphique de Shepard dans PROXMAP combine les informations présentées dans les graphiques d'ajustement non linéaire et de transformation d'ALSCAL.

Résultats et interprétation

Lorsque PROXMAP est exécuté avec des paramètres comparables à ceux d'ALSCAL, les configurations obtenues sont très similaires. Les différences entre les solutions se limitent à un facteur d'échelle global et à d'éventuelles symétries par rapport aux axes.

Les valeurs des coordonnées peuvent varier numériquement, mais elles représentent la même configuration sous-jacente à l'échelle. PROXMAP indique les coordonnées en fonction de l'échelle d'origine des proximités.

Une comparaison visuelle des graphiques de configuration montre que les positions relatives des objets sont conservées, ce qui confirme que PROXMAP reproduit fidèlement la solution d'ALSCAL.

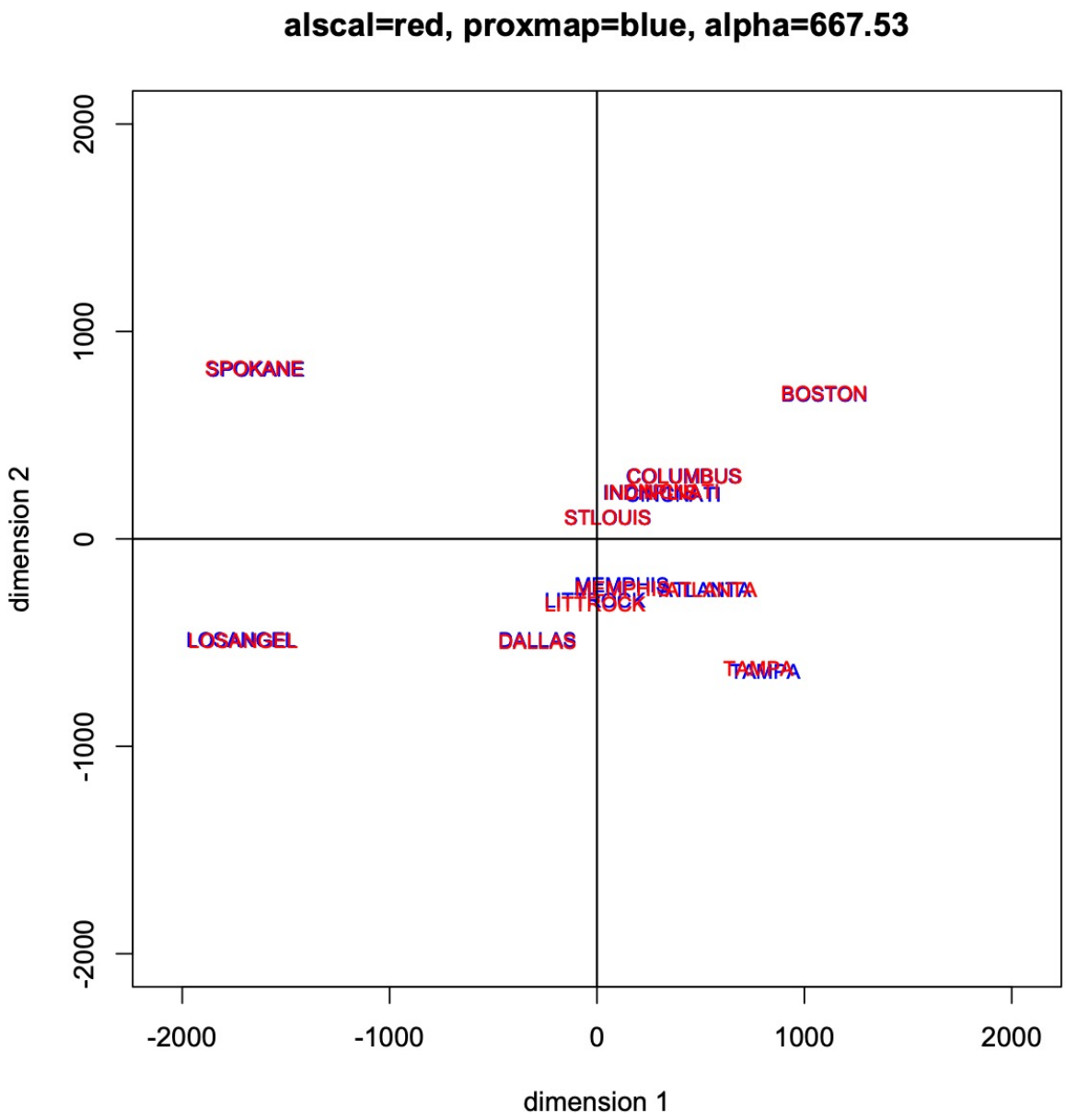

Figure 3. Configuration de l'ALSCAL Stimulus - Coordonnées PROXMAP - Espace commun

Points clés à retenir pour les utilisateurs d'ALSCAL

- Les données de proximité existantes pour ALSCAL peuvent être réutilisées dans PROXMAP, y compris celles stockées sous forme de matrice triangulaire inférieure.
- Une variable d'étiquette d'objet supplémentaire est nécessaire pour identifier les objets dans les résultats et les graphiques de PROXMAP.
- Lorsque vous comparez les résultats de PROXMAP avec ceux d'ALSCAL, des différences d'orientation des coordonnées peuvent apparaître. Les rotations n'affectent que l'orientation des axes et ne modifient ni les distances ni la qualité de l'ajustement.
- PROXMAP utilise une terminologie actualisée et des méthodes modernes de réduction multidimensionnelle, tout en préservant la continuité conceptuelle avec ALSCAL.

Statistiques de rapport

La procédure Statistiques de rapport permet d'obtenir la liste exhaustive des statistiques récapitulatives qui servent à décrire le rapport entre deux variables d'échelle.

Vous pouvez trier la sortie sur la base des valeurs d'une variable de regroupement, dans l'ordre croissant ou décroissant. Vous pouvez supprimer le rapport des statistiques de rapport dans le document de sortie et enregistrer les résultats dans un fichier externe.

Exemple
Le rapport existant entre le prix estimatif et le prix de vente des maisons est-il uniforme dans chacun de ces cinq comtés ? D'après les sorties, vous pouvez conclure que la distribution des rapports varie considérablement d'un comté à l'autre.

Statistiques
Médiane, moyenne pondérée, intervalles de confiance, coefficient de dispersion (COD), coefficient de variation (COV), coefficient de variation centré sur la médiane, coefficient de variation centré sur la moyenne, biais lié au prix (PRB), écart-type, écart-type, écart moyen absolu (AAD), valeurs minimale et maximale, et indice de concentration calculé pour une fourchette ou un pourcentage spécifié par l'utilisateur dans le rapport médian.

Remarques sur les données

Données
Utilisez des codes numériques ou alphanumériques pour coder les variables de regroupement (mesures de niveau nominal ou ordinal).

Hypothèses
Vous devez utiliser des variables d'échelle acceptant les valeurs positives pour les variables qui définissent le numérateur et le dénominateur du rapport.

Obtention de statistiques de rapport

1. A partir des menus, sélectionnez :

   Analyse > **Statistiques descriptives** > **Rapport...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez la variable du numérateur.
3. Sélectionnez la variable du dénominateur.

Eventuellement :

- Sélectionner une variable de regroupement et préciser l'ordre de présentation des groupes dans le résultat.
- Choisissez si vous souhaitez afficher les résultats dans le visualiseur de résultats.
- Choisissez d'ajouter un mot clé **N** à la sous-commande **PRINT**. Ce paramètre ajoute les tailles d'échantillon à la table de statistiques de la sortie.
- Choisissez ou non d'enregistrer les résultats dans un fichier externe en vue d'une utilisation ultérieure, et précisez le nom du fichier dans lequel les résultats sont enregistrés.

Statistiques de rapport : Statistiques

Tendance centrale
Les mesures de tendance centrale sont des statistiques qui décrivent la distribution des ratios.

Médiane
La valeur telle que le nombre de rapports inférieurs à cette valeur est identique au nombre de rapports supérieurs à cette valeur.

Moyenne
Résultat de la somme des rapports divisée par le nombre total de rapports.

Moyenne pondérée
Résultat de la division de la moyenne du numérateur par la moyenne du dénominateur. La moyenne pondérée correspond également à la moyenne des rapports pondérée par le dénominateur.

Intervalles de confiance pour PRB et les mesures de tendance centrales
Affiche les intervalles de confiance pour PRB, la moyenne, la médiane et la moyenne pondérée (si demandée). Affectez une valeur supérieure ou égale à 0 et inférieure à 100 au niveau de confiance. Le paramétrage par défaut est 95 %.

Dispersion
Ces statistiques mesurent le degré de variation, ou de dispersion, dans les valeurs observées.

AAD
La déviation absolue moyenne est égale à la somme des déviations absolues des rapports relatifs à la médiane, divisée par le nombre total de rapports.

COD
Le coefficient de dispersion résulte de l'expression de la déviation moyenne absolue en pourcentage de la médiane.

écart de coût
CV désigne le coefficient de variation. Il s'agit du rapport entre l'écart-type et la moyenne. Les valeurs sont exprimées en pourcentage.

PRB
Biais lié au prix. Index indiquant si les rapports évaluation/prix sont systématiquement supérieurs ou inférieurs pour les biens à prix élevé. Régresse les différences de pourcentage des ratios d'évaluation par rapport au ratio médian sur les logarithmes en base 2 des mesures approximatives de valeur calculées comme les moyennes des prix de vente et les rapports entre les valeurs évaluées et le ratio médian. Fournit la variation en pourcentage des ratios d'évaluation pour une variation de 100 % de la valeur.

PRD
Le différentiel lié au prix, ou index de régressivité, résulte de la division de la moyenne par la moyenne pondérée.

Ecart type
L'écart type est la racine carrée positive de la somme des carrés des écarts des rapports relatifs à la moyenne divisée par le nombre total des rapports moins un.

Plage
Résultat de la soustraction du ratio minimal au ratio maximal.

Minimum
Le minimum est le plus petit ratio.

Maximum
Le maximum est le plus grand ratio.

Index de concentration
Le coefficient de concentration mesure le pourcentage des rapports compris dans un intervalle. Vous pouvez le calculer de deux manières :

Entre les proportions
Dans ce cas, vous définissez l'intervalle de manière explicite en précisant les valeurs minimale et maximale. Entrez les valeurs des proportions inférieure et supérieure, puis cliquez sur **Ajouter** pour obtenir un intervalle.

Dans le pourcentage de la médiane
Dans ce cas, vous définissez l'intervalle de manière implicite en indiquant le pourcentage de la médiane. Entrez une valeur comprise entre 0 et 100, et cliquez sur **Ajouter**. La limite inférieure de l'intervalle est égale à (1 – 0.01 × valeur) × médiane, et la limite supérieure est égale à (1 + 0.01 × valeur) × médiane.

Tracés P-P

La procédure Tracés P-P produit des tracés de probabilité d'une ou plusieurs variables de série temporelle ou de séquence. Les variables peuvent être normalisées, différenciées et transformées avant de le traçage.

Les distributions test disponibles sont les suivantes : Bêta, khi-deux, Exponentielle, Gamma, Semi-normale, Laplace, Logistique, Lognormal, Gaussienne, Pareto, t de Student, Weibull et Uniforme. En fonction de la distribution sélectionnée, vous pouvez spécifier des degrés de liberté et d'autres paramètres.

- Vous obtenez des tracés de probabilités pour les valeurs transformées. Les options de transformation sont les suivantes : Log naturel, Standardiser les valeurs, Différence et Différence saisonnière.
- Vous pouvez spécifier la méthode de calcul des distributions prévues et de résolution des "égalités" (plusieurs observations ayant la même valeur).

Distribution du test
Spécifiez un type de distribution pour vos données. La liste déroulante propose les options suivantes :

Bêta
Distribution bêta. Les paramètres shape1 et shape2 *a* et *b* doivent être positifs. S'ils ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon. Toutes les observations doivent être comprises entre 0 et 1, inclusivement.

Khi-carré
Distribution khi-carré. Vous devez spécifier les degrés de liberté *(df)*. Les observations négatives ne sont pas autorisées.

Exponentiel
Distribution exponentielle. Le paramètre d'échelle *a* doit être positif. Si le paramètre n'est pas spécifié, DISTRIBUTION l'estime à partir de la moyenne de l'échantillon. Les observations négatives ne sont pas autorisées.

Gamma
Distribution gamma. Les paramètres de forme et d'échelle *a* et *b* doivent être positifs. S'ils ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon. Les observations négatives ne sont pas autorisées.

Semi-normale
Distribution semi-normale. On considère que les données ne sont pas localisées ou qu'elles sont centralisées. (Paramètre d'emplacement = 0.) Vous pouvez spécifier le paramètre d'échelle *a* ou laisser DISTRIBUTION l'estimer à l'aide de la méthode d'estimation du maximum de vraisemblance.

Laplace
Distribution Laplace ou exponentielle double. LAPLACE prend un paramètre d'emplacement et un paramètre d'échelle (*a* et *b).* Le paramètre d'échelle *(b)* doit être positif. Si les paramètres ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon.

Logistique
Distribution logistique. LOGISTIC prend un paramètre d'emplacement et un paramètre d'échelle (*a* et *b).* Le paramètre d'échelle *(b)* doit être positif. Si les paramètres ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon.

Lognormal
Distribution logarithmique normale. Les paramètres d'échelle et de forme *a* et *b* doivent être positifs. S'ils ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne et de l'écart type du logarithme naturel des données d'échantillon. Les observations négatives ne sont pas autorisées.

Normal
Distribution normale. Le paramètre d'emplacement *a* peut être une valeur numérique, tandis que le paramètre d'échelle *b* doit être positif. S'ils ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon.

Pareto
Distribution de type Pareto. Les paramètres de seuil et de forme *a* et *b* doivent être positifs. Si elles ne sont pas spécifiées, DISTRIBUTION suppose que *a* est égal à l'observation minimale et estime *b* à l'aide de la méthode d'estimation du maximum de vraisemblance. Les observations négatives ne sont pas autorisées.

Student t
Distribution t *de Student.* Vous devez spécifier les degrés de liberté *(df)*.

Uniforme
Distribution uniforme. UNIFORM prend un paramètre minimum et un paramètre maximum *(a* et *b).* Le paramètre *A* doit être égal ou supérieur à *B.* Si les paramètres ne sont pas spécifiés, DISTRIBUTION les suppose à partir des données exemple.

Weibull
Distribution Weibull. Les paramètres d'échelle et de forme *a* et *b* doivent être positifs. Si elles ne sont pas spécifiées, DISTRIBUTION les estime à l'aide de la méthode des moindres carrés. Les observations négatives ne sont pas autorisées.

Paramètres de distribution
Fournit des options de paramètres et de stratégies de distribution.

Estimer à partir des données
Lorsque cette option est sélectionnée, ce paramètre estime les paramètres de distribution en fonction des données et du type de distribution sélectionné.

Indiquer
Lorsque cette option est sélectionnée, vous pouvez spécifier les paramètres de distribution du type de distribution sélectionné.

Remarque : Les paramètres disponibles varient en fonction du type de distribution sélectionné.

Transformer
Les options fournies définissent les paramètres de transformation et de périodicité.

Transformée log népérien
Transforme les données en utilisant le logarithme naturel (base *e*) pour éliminer l'amplitude variable.

Standardiser les valeurs
Transforme les variables de séquence ou de série temporelle dans un échantillon avec une moyenne de 0 et un écart type de 1.

Différence
Spécifie le degré de différentiation qui est utilisé avant le tracé pour convertir une variable non stationnaire en une variable stationnaire avec une moyenne et une variance constantes. Entrez une valeur appropriée dans la zone.

Différence saisonnière
Si la variable présente un schéma saisonnier ou périodique, vous pouvez utiliser ce paramètre pour différencier la variable selon les saisons avant de la tracer.

Remarque : Ce paramètre est activé uniquement lorsqu'une variable de séquence ou de série temporelle, avec une périodicité définie, est sélectionnée comme une des variables quantitatives.

Formule d'estimation de proportion
Les options fournies établissent la formule utilisée pour estimer les proportions.

Blom
Crée une variable de classement basée sur des estimations de proportion qui utilise la formule `(r-3/8) / (w+1/4)`, où `w` est la somme des pondérations d'observation et `r` le rang.

Rankit
Utilise la formule `(r-1/2) / w`, où `w` est le nombre d'observations et `r` le rang, compris entre 1 et `w`.

Tukey
Utilise la formule `(r-1/3) / (w+1/3)`, où `r` est le rang et `w` la somme des pondérations d'observation.

Van der Waerden
Transformation de Van der Waerden, définie par la formule `r/(w+1)`, où `w` est la somme des pondérations d'observation et `r` le rang, compris entre 1 et `w`.

Rang attribué aux ex aequo
Les options fournies contrôlent la méthode permettant de déterminer comment traiter les valeurs égales. Le tableau suivant montre comment les différentes méthodes attribuent des rangs aux valeurs égales.

Tableau 6. Méthodes de classement et résultats

Valeur	Moyenne	Faible	Elevée	Départager les ex aequo arbitrairement
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Obtention de tracés de probabilités Q-Q

1. A partir des menus, sélectionnez :

   Analyser > **Statistiques descriptives** > **Tracés Q-Q...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables numériques, puis placez-les dans le champ Variables.
3. Sélectionnez une distribution test.

Si vous le souhaitez, vous pouvez sélectionner des options de transformation pour obtenir des tracés de probabilités pour des valeurs transformées. Vous pouvez également spécifier la méthode de calcul des distributions prévues.

Tracé Q-Q

La procédure de tracés Q-Q produit des tracés de probabilités pour les valeurs transformées. Les distributions test disponibles sont les suivantes : Bêta, khi-deux, Exponentielle, Gamma, Semi-normale, Laplace, Logistique, Lognormal, Gaussienne, Pareto, t de Student, Weibull et Uniforme. En fonction de la distribution sélectionnée, vous pouvez spécifier des degrés de liberté et d'autres paramètres.

- Vous obtenez des tracés de probabilités pour les valeurs transformées. Les options de transformation sont les suivantes : Log naturel, Standardiser les valeurs, Différence et Différence saisonnière.
- Vous pouvez spécifier la méthode de calcul des distributions prévues et de résolution des "égalités" (plusieurs observations ayant la même valeur).

Distribution du test
Spécifiez un type de distribution pour vos données. La liste déroulante propose les options suivantes :

Bêta
Distribution bêta. Les paramètres shape1 et shape2 *a* et *b* doivent être positifs. S'ils ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon. Toutes les observations doivent être comprises entre 0 et 1, inclusivement.

Khi-carré
Distribution khi-carré. Vous devez spécifier les degrés de liberté *(df)*. Les observations négatives ne sont pas autorisées.

Exponentiel
Distribution exponentielle. Le paramètre d'échelle *a* doit être positif. Si le paramètre n'est pas spécifié, DISTRIBUTION l'estime à partir de la moyenne de l'échantillon. Les observations négatives ne sont pas autorisées.

Gamma
Distribution gamma. Les paramètres de forme et d'échelle *a* et *b* doivent être positifs. S'ils ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon. Les observations négatives ne sont pas autorisées.

Semi-normale
Distribution semi-normale. On considère que les données ne sont pas localisées ou qu'elles sont centralisées. (Paramètre d'emplacement = 0.) Vous pouvez spécifier le paramètre d'échelle *a* ou laisser DISTRIBUTION l'estimer à l'aide de la méthode d'estimation du maximum de vraisemblance.

Laplace
Distribution Laplace ou exponentielle double. LAPLACE prend un paramètre d'emplacement et un paramètre d'échelle (*a* et *b).* Le paramètre d'échelle *(b)* doit être positif. Si les paramètres ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon.

Logistique
Distribution logistique. LOGISTIC prend un paramètre d'emplacement et un paramètre d'échelle (*a* et *b).* Le paramètre d'échelle *(b)* doit être positif. Si les paramètres ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon.

Lognormal
Distribution logarithmique normale. Les paramètres d'échelle et de forme *a* et *b* doivent être positifs. S'ils ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne et de l'écart type du logarithme naturel des données d'échantillon. Les observations négatives ne sont pas autorisées.

Normal
Distribution normale. Le paramètre d'emplacement *a* peut être une valeur numérique, tandis que le paramètre d'échelle *b* doit être positif. S'ils ne sont pas spécifiés, DISTRIBUTION les estime à partir de la moyenne de l'échantillon et de l'écart type de l'échantillon.

Pareto
Distribution de type Pareto. Les paramètres de seuil et de forme *a* et *b* doivent être positifs. Si elles ne sont pas spécifiées, DISTRIBUTION suppose que *a* est égal à l'observation minimale et estime *b* à l'aide de la méthode d'estimation du maximum de vraisemblance. Les observations négatives ne sont pas autorisées.

Student t
Distribution t *de Student.* Vous devez spécifier les degrés de liberté *(df)*.

Uniforme
Distribution uniforme. UNIFORM prend un paramètre minimum et un paramètre maximum *(a* et *b).* Le paramètre *A* doit être égal ou supérieur à *B.* Si les paramètres ne sont pas spécifiés, DISTRIBUTION les suppose à partir des données exemple.

Weibull
Distribution Weibull. Les paramètres d'échelle et de forme *a* et *b* doivent être positifs. Si elles ne sont pas spécifiées, DISTRIBUTION les estime à l'aide de la méthode des moindres carrés. Les observations négatives ne sont pas autorisées.

Paramètres de distribution
Fournit des options de paramètres et de stratégies de distribution.

Estimer à partir des données
Lorsque cette option est sélectionnée, ce paramètre estime les paramètres de distribution en fonction des données et du type de distribution sélectionné.

Indiquer
Lorsque cette option est sélectionnée, vous pouvez spécifier les paramètres de distribution du type de distribution sélectionné.

Remarque : Les paramètres disponibles varient en fonction du type de distribution sélectionné.

Transformer
Les options fournies définissent les paramètres de transformation et de périodicité.

Transformée log népérien
Transforme les données en utilisant le logarithme naturel (base *e*) pour éliminer l'amplitude variable.

Standardiser les valeurs
Transforme les variables de séquence ou de série temporelle dans un échantillon avec une moyenne de 0 et un écart type de 1.

Différence
Spécifie le degré de différentiation qui est utilisé avant le tracé pour convertir une variable non stationnaire en une variable stationnaire avec une moyenne et une variance constantes. Entrez une valeur appropriée dans la zone.

Différence saisonnière
Si la variable présente un schéma saisonnier ou périodique, vous pouvez utiliser ce paramètre pour différencier la variable selon les saisons avant de la tracer.

Remarque : Ce paramètre est activé uniquement lorsqu'une variable de séquence ou de série temporelle, avec une périodicité définie, est sélectionnée comme une des variables quantitatives.

Formule d'estimation de proportion
Les options fournies établissent la formule utilisée pour estimer les proportions.

Blom
Crée une variable de classement basée sur des estimations de proportion qui utilise la formule `(r-3/8) / (w+1/4)`, où `w` est la somme des pondérations d'observation et `r` le rang.

Rankit
Utilise la formule `(r-1/2) / w`, où `w` est le nombre d'observations et `r` le rang, compris entre 1 et `w`.

Tukey
Utilise la formule `(r-1/3) / (w+1/3)`, où `r` est le rang et `w` la somme des pondérations d'observation.

Van der Waerden
Transformation de Van der Waerden, définie par la formule `r/(w+1)`, où `w` est la somme des pondérations d'observation et `r` le rang, compris entre 1 et `w`.

Rang attribué aux ex aequo
Les options fournies contrôlent la méthode permettant de déterminer comment traiter les valeurs égales. Le tableau suivant montre comment les différentes méthodes attribuent des rangs aux valeurs égales.

Tableau 7. Méthodes de classement et résultats

Valeur	Moyenne	Faible	Elevée	Départager les ex aequo arbitrairement
10	1	1	1	1
15	3	2	4	2

Tableau 7. Méthodes de classement et résultats (suite)

Valeur	Moyenne	Faible	Elevée	Départager les ex aequo arbitrairement
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Obtention de tracés de probabilités Q-Q

1. A partir des menus, sélectionnez :

   Analyser > **Statistiques descriptives** > **Tracés Q-Q...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez une ou plusieurs variables numériques, puis placez-les dans le champ Variables.
3. Sélectionnez une distribution test.

Si vous le souhaitez, vous pouvez sélectionner des options de transformation pour obtenir des tracés de probabilités pour des valeurs transformées. Vous pouvez également spécifier la méthode de calcul des distributions prévues.

Analyse ROC

L'analyse ROC (Receiver Operating Characteristic) permet d'évaluer l'exactitude des prévisions d'un modèle en traçant la sensibilité par rapport au taux de faux positifs (1-spécificité) d'un test de classification (étant donné que le seuil varie sur toute une plage de résultats de tests de diagnostic). L'aire complète sous une courbe ROC donnée, ou AUC, formule une statistique importante qui représente la probabilité que la prédiction soit dans le bon ordre lorsqu'une variable de test est observée (pour un sujet sélectionné au hasard dans le groupe de cas, et pour l'autre sélectionné au hasard dans le groupe de contrôle). L'analyse ROC prend en charge l'inférence concernant une aire AUC unique et des courbes PR (précision-rappel). Elle propose des options pour comparer deux courbes ROC générées à partir de groupes indépendants ou de sujets appariés.

L'ancienne procédure Courbe ROC prend en charge l'inférence statistique sur une courbe ROC unique. Cela peut également être récupéré par la nouvelle procédure d'analyse ROC. En outre, la nouvelle procédure d'analyse ROC peut comparer deux courbes ROC générées à partir de groupes indépendants ou de sujets appariés.

Les courbes PR comparent la précision du tracé au rappel, ont tendance à être plus informatives lorsque les échantillons de données observées sont très asymétriques et offrent une alternative aux courbes ROC pour les données dont la distribution des classes est très asymétrique.

Exemple
Une banque envisage de classer correctement ses clients en catégories, à savoir ceux qui assumeront ou non le remboursement de leur prêt. Des modèles particuliers sont développés afin de supporter la prise de décision. L'analyse ROC peut être utilisée pour déterminer et évaluer l'exactitude des prévisions d'un modèle.

Statistiques
Aire sous la courbe (AUC), groupe négatif, valeurs manquantes, classification positive, valeur limite, force de conviction, intervalle de confiance asymptotique bilatéral, distribution, erreur standard, plan à groupes indépendants ; plan à échantillons appariés, hypothèse non paramétrique, hypothèse de distribution exponentielle bi-négative, centre, césure, courbe PR, interpolation progressive, signification asymptotique (bilatérale), sensibilité et taux de faux positifs (1-spécificité), précision et rappel (PR).

Méthodes
La comparaison porte sur les aires situées sous deux courbes ROC et générées à partir de groupes indépendants ou de sujets appariés. La comparaison de deux courbes ROC peut fournir de plus amples informations sur l'exactitude de deux approches de diagnostic comparatives.

Remarques sur les données d'analyse ROC

Données
Les courbes PR comparent la précision du tracé au rappel et ont tendance à être plus informatives lorsque les échantillons de données observées sont très asymétriques. Une simple interpolation linéaire peut générer une erreur d'estimation trop optimiste d'une courbe PR.

Hypothèses
La prévision se trouve dans l'ordre approprié lorsqu'une variable de test est observée (pour un sujet sélectionné au hasard dans le groupe d'observations et un autre sujet sélectionné au hasard dans le groupe de contrôle). Chaque groupe défini doit contenir au moins une observation valide. Une seule variable de regroupement est utilisée par procédure.

Obtention d'une analyse ROC

1. A partir des menus, sélectionnez :

   Analyse > **Classifier** > **Analyse ROC...**

   Remarque : Les champs surlignés en rouge sont obligatoires. Les boutons **Coller** et **OK** sont activés après avoir saisi des valeurs valides dans tous les champs obligatoires.
2. Sélectionnez un ou plusieurs **Variable de test** tester les variables de probabilité.
3. Sélectionnez-en un **État variable** .
4. Identifiez le *positif* **Valeur de la variable d'état** .
5. Sélectionnez l'option **Plan à échantillons appariés** ou une variable de regroupement unique (vous ne pouvez pas sélectionner les deux options).
   - Le paramètre **Plan à échantillons appariés** permet de demander le plan à échantillons appariés pour la ou les variables à tester. Le plan à échantillons appariés compare deux courbes ROC dans un scénario de test à échantillons appariés lorsque plusieurs valeurs de test sont mesurées sur les mêmes sujets associés à une variable d'état.

     Remarque : Lorsque **Plan à échantillons appariés** est sélectionné, les options **Variable de regroupement** et **Hypothèse de distribution** (dans la boîte de dialogue Options) sont désactivées.
   - Lorsqu'une variable de regroupement numérique est sélectionnée, vous pouvez cliquer sur **Définir groupes ...** pour demander le plan de groupe indépendant pour la ou les variables de test et pour spécifier deux valeurs, un point milieu ou un point de césure.
6. Vous pouvez également cliquer sur Options pour définir la classification, la direction du test, les paramètres d'erreur standard et les paramètres de valeur manquante.
7. Si vous le souhaitez, cliquez sur Affichage pour définir les paramètres de traçage et d'impression (dont la courbe ROC, la courbe PR et les paramètres de qualité du modèle).
8. Cliquez sur **OK**.

Analyse ROC : Options

Vous pouvez indiquer les options suivantes pour votre analyse ROC :

Classification
Permet de spécifier si la valeur de césure doit être incluse ou exclue lors d'une classification *positive*. Ce paramètre n'a pas de conséquence sur la sortie.

Direction du test
Propose des options pour préciser la direction de la variable de résultat de test qui indique une conviction croissante que le sujet est diagnostiqué positif.

Paramètres pour une erreur standard de zone

Vous permet de spécifier la méthode utilisée pour estimer l'erreur standard de la zone située sous la courbe. Les méthodes disponibles sont des valeurs exponentielles non paramétriques et bi-négatives. La valeur par défaut du paramètre **Non paramétrique** fournit des estimations sous l'hypothèse non paramétrique. Le paramètre **Exponentielle bi-négative** fournit des estimations sous l'hypothèse de distribution exponentielle bi-négative.

La section vous permet également d'indiquer le niveau de confiance de l'intervalle de confiance asymptotique bilatéral de l'aire AUC. La plage disponible est comprise entre 0,0 % et 100,0 % (la valeur par défaut est 95 %).

Remarque : Le paramètre s'applique uniquement au plan à groupes indépendants et est sans incidence sur le plan à échantillons appariés.

Valeurs manquantes

Vous permet de spécifier comment traiter les valeurs manquantes. Lorsque ce paramètre n'est pas sélectionné, les valeurs manquantes (utilisateur et système) sont exclues. Lorsque ce paramètre est sélectionné, les valeurs utilisateur manquantes sont considérées comme valides, les valeurs système manquantes sont exclues. Les observations avec des valeurs système manquantes, dans la variable à tester ou la variable d'état, sont toujours exclues de l'analyse.

Analyse ROC : Affichage

Vous pouvez indiquer les paramètres d'affichage suivants pour votre analyse ROC :

Tracé

Propose des options pour tracer des courbes ROC et PR (précision-rappel).

Courbe ROC

Lorsque cette option est sélectionnée, un graphique Courbe ROC s'affiche dans la sortie. Sélectionnez **Avec diagonale de référence** pour dessiner une diagonale de référence sur le graphique Courbe ROC.

Courbe PR (précision-rappel)

Lorsque cette option est sélectionnée, un graphique Courbe PR (précision-rappel) s'affiche dans la sortie. Les courbes PR ont tendance à être plus informatives lorsque les échantillons de données observées sont très asymétriques et offrent une alternative aux courbes ROC pour les données dont la distribution des classes est très asymétrique. Le paramètre par défaut (**Interpoler selon les vrais positifs**) effectue l'interpolation progressive selon les vrais positifs. Le paramètre **Interpoler selon les faux positifs** effectue l'interpolation progressive selon les faux positifs.

Qualité générale du modèle

Ce paramètre contrôle si un graphique à barres est créé pour afficher la valeur de la limite inférieure de l'intervalle de confiance de l'aire AUC estimée. Par défaut, il n'est pas sélectionné (le graphique à barres est supprimé).

Imprimer

Propose des options pour définir la sortie des statistiques correspondantes.

Erreur standard et intervalle de confiance

Ce paramètre contrôle les statistiques à afficher dans la table AUC ("Aire sous la courbe"). Lorsqu'il n'est pas sélectionné, seules les aires AUC estimées s'affichent. Lorsqu'il est sélectionné, des statistiques supplémentaires s'affichent, dont l'erreur standard de l'aire AUC, la signification asymptotique (bilatérale) et les limites de l'intervalle de confiance asymptotique sous l'hypothèse nulle.

Coordonnées de la courbe ROC

Ce paramètre contrôle les coordonnées de la courbe ROC, ainsi que les valeurs de césure. Lorsqu'il n'est pas sélectionné, la sortie des coordonnées est supprimée. Lorsque ce paramètre est sélectionné, les paires de valeurs "Sensibilité et 1-Spécificité" sont fournies avec les valeurs de césure pour chaque courbe ROC.

Indice de Youden

Affiche la valeur de l'index de Youden pour chaque valeur de césure de la courbe ROC.

Coordonnées de la courbe PR
Ce paramètre contrôle les coordonnées de la courbe PR (précision-rappel), ainsi que les valeurs de césure. Lorsqu'il n'est pas sélectionné, la sortie des coordonnées est supprimée. Lorsque ce paramètre est sélectionné, les paires de valeurs "précision et rappel" sont fournies avec les valeurs de césure pour chaque courbe PR.

Métriques d'évaluation du discriminant
Ce paramètre contrôle l'affichage du tableau Métriques d'évaluation du discriminant dans la sortie. Le tableau montre comment un modèle de classification s'ajuste aux données par rapport à une affectation aléatoire et fournit les informations suivantes :

- Variables à tester, spécifiées par l'utilisateur
- Informations de groupe
- Indice de Gini (correspond à 2*AUC - 1, où AUC (Area Under Curve) est la zone sous la courbe ROC)
- Valeurs K-S max et Césure

Analyse ROC : Définir des groupes (chaîne)

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

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

Analyse ROC : Définir des groupes (numérique)

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

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

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

Courbe ROC

Cette procédure constitue un moyen efficace d'évaluer les performances des méthodes de classement ne mettant en oeuvre qu'une seule variable à deux catégories et utilisées pour la classification des sujets.

Remarque : Version29.0.1 à partir de maintenant, les menus et boîtes de dialogue de la commande ROC sont supprimés et cette commande est disponible uniquement en tant que syntaxe.

Exemple. Une banque envisage de classer correctement ses clients en catégories, à savoir ceux qui assumeront ou non le remboursement de leur prêt. Des méthodes particulières sont développées afin de supporter la prise de décision. Les courbes ROC peuvent être utilisées pour évaluer le mode de fonctionnement optimal de ces méthodes.

Statistiques : La zone inférieure à la courbe ROC comporte un intervalle de confiance ainsi que les coordonnées de cette courbe. Tracés : courbe ROC.

Méthodes : L'estimation de la zone située sous la courbe ROC peut être calculée de façon paramétrique ou non à l'aide du modèle exponentiel binégatif.

Remarques sur les données de la courbe ROC

L'ancienne procédure Courbe ROC prend en charge l'inférence statistique sur une courbe ROC unique. Cela peut également être récupéré par la nouvelle procédure d'analyse ROC. En outre, la nouvelle procédure d'analyse ROC peut comparer deux courbes ROC générées à partir de groupes indépendants ou de sujets appariés.

Données. Les variables de test sont quantitatives. Les variables de test sont souvent composées des probabilités issues d'une analyse discriminante ou d'une régression logistique, ou bien des scores indiqués sur une échelle arbitraire et spécifiant la "force de conviction" d'un indicateur lorsqu'un sujet se rapporte à l'une ou l'autre des catégories. La variable d'état peut être d'un type quelconque et indique la véritable catégorie à laquelle un sujet appartient. La valeur de la variable d'état indique la catégorie à considérer comme *positive*.

Hypothèses : Les nombres croissants d'une échelle d'indicateurs confirment que le sujet appartient à une catégorie, tandis que les nombres décroissants d'une échelle confirment qu'il appartient à une autre catégorie. L'utilisateur doit choisir la direction *positive*. On suppose également que la *véritable* catégorie à laquelle chaque sujet appartient est connue.

Obtention d'une courbe ROC

1. Dans les menus, choisissez : **Analyser** > **Classer** > **Analyse ROC...**
2. Sélectionnez un ou plusieurs **Variable de test** .
3. Sélectionnez-en un **État variable** .
4. Identifiez la valeur *positive* de la variable d'état.
5. Cliquez sur **Afficher** , sélectionner **Courbe ROC** et **Avec ligne de référence diagonale** , si besoin.
6. Cliquez sur **Continuer**.

Options de la courbe ROC

Vous pouvez indiquer les options suivantes pour votre analyse ROC :

Classification
Permet de spécifier si la valeur de césure doit être incluse ou exclue lors d'une classification *positive*. Ce paramètre n'a pas de conséquence sur la sortie.

Direction du test
Permet de spécifier la direction de l'échelle en fonction de la catégorie *positive*.

Paramètres pour une erreur standard de zone
Vous permet de spécifier la méthode utilisée pour estimer l'erreur standard de la zone située sous la courbe. Les méthodes disponibles sont les exponentielles « non paramétriques » et « bi-négatives ». Vous permet également de définir le niveau de l'intervalle de confiance. Les valeurs de la plage se situent entre 50,1 % et 99,9 %.

Valeurs manquantes
Vous permet de spécifier comment traiter les valeurs manquantes.

Simulation

Les modèles prédictifs, tels que les modèles à régression linéaire, nécessitent un ensemble d'entrées connues afin de prédire un résultat ou une valeur cible. Toutefois, dans de nombreuses applications de la vie réelle, les valeurs des entrées sont incertaines. La simulation vous permet de prendre en compte l'incertitude relative aux entrées des modèles prédictifs et d'évaluer la probabilité de divers résultats du modèle en présence de cette incertitude. Par exemple, vous avez un modèle de profit qui comprend les coûts des matériaux en entrée, mais il existe une incertitude quant à ces coûts en raison de l'instabilité du marché. Dans ce cas, vous pouvez utiliser la simulation pour modéliser cette incertitude et déterminer les effets qu'elle a sur les profits.

La simulation offerte dans IBM SPSS Statistiques utilise la méthode de Monte Carlo. Les entrées incertaines sont modélisées à l'aide de distributions de probabilité (telle que la distribution triangulaire), et des valeurs simulées sont générées pour ces entrées d'après ces distributions. Les entrées dont les valeurs sont connues sont fixées selon les valeurs connues. Le modèle prédictif est évalué à l'aide d'une valeur simulée pour chaque entrée incertaine et des valeurs fixes des entrées connues, afin de calculer la cible (ou les cibles) du modèle. Ce processus est répété plusieurs fois (en général des dizaines ou des centaines de milliers de fois) et résulte en une distribution des valeurs cible qui peut être utilisée pour répondre à des questions de nature probabiliste. Dans le contexte d'IBM SPSS Statistiques, chaque répétition du processus génère une observation distincte (enregistrement) de données qui consiste en l'ensemble des valeurs simulées des entrées incertaines, des valeurs des entrées fixes et de la ou des cible(s) prédite(s) du modèle.

Vous pouvez également simuler des données en l'absence d'un modèle prédictif en indiquant les distributions de probabilité pour les variables que vous souhaitez simuler. Chaque observation de données générée consiste en l'ensemble des valeurs simulées pour les variables spécifiées.

Pour lancer une simulation, vous devez spécifier des informations telles que le modèle prédictif, les distributions de probabilité des entrées incertaines et les corrélations entre ces entrées et les valeurs des entrées fixes. Une fois toutes les informations requises pour la simulation spécifiées, vous pouvez l'exécuter et éventuellement enregistrer les spécifications dans un fichier de **plan de simulation**. Il est possible de partager le plan de simulation avec d'autres utilisateurs, qui peuvent à leur tour exécuter la simulation sans avoir besoin de savoir exactement comment elle a été créée.

Deux interfaces sont disponibles pour utiliser les simulations. Le Générateur de simulation est une interface avancée destinée aux utilisateurs qui conçoivent et exécutent des simulations. Il offre l'ensemble complet de fonctionnalités permettant la conception de simulations, l'enregistrement des spécifications dans un fichier de plan de simulation, la spécification de la sortie et l'exécution de la simulation. Il est possible de construire une simulation basée sur un fichier de modèle IBM SPSS ou sur un ensemble d'équations personnalisées que vous définissez dans le Générateur de simulation. Il est également possible de charger un plan de simulation existant dans le Générateur de simulation, de modifier ses paramètres et d'exécuter la simulation, ainsi que d'enregistrer le plan mis à jour au besoin. Pour les utilisateurs disposant d'un plan de simulation et qui souhaitent principalement exécuter une simulation, une interface simplifiée est disponible. Elle vous permet de modifier des paramètres afin d'exécuter la simulation dans différentes conditions, mais n'offre pas l'ensemble des fonctionnalités du Générateur de simulation qui permettent de concevoir des simulations.

Conception d'une simulation basée sur un fichier de modèle

1. A partir des menus, sélectionnez :

   Analyse > **Simulation...**
2. Cliquez sur **Sélectionner un fichier de modèle SPSS**, puis cliquez sur **Poursuivre**.
3. Ouvrez le fichier de modèle.

   Le fichier de modèle est un fichier XML qui contient le fichier PMML Model créé à partir d'IBM SPSS Statistiques ou d'IBM SPSS Modeler. Pour plus d'informations, voir «Onglet Modèle », à la page 302.
4. Sur l'onglet Simulation (dans le Générateur de simulation), spécifiez les distributions de probabilité des entrées simulées et les valeurs des entrées fixes. Si le jeu de données actif contient des données historiques relatives aux entrées simulées, cliquez sur **Ajuster tout** pour déterminer automatiquement la distribution la mieux adaptée à chaque entrée, ainsi que les corrélations entre celles-ci. Pour chaque entrée simulée non ajustée aux données historiques, vous devez explicitement indiquer une distribution en sélectionnant un type et en entrant les paramètres requis.
5. Cliquez sur **Exécuter** pour exécuter la simulation. Par défaut, le plan de simulation qui spécifie les détails de la simulation, est enregistré à l'emplacement indiqué dans les paramètres d'enregistrement.

Les options suivantes sont disponibles :

- Modifier l'emplacement du plan de simulation enregistré.
- Spécifier les corrélations connues existant entre les entrées simulées.

- Calculer automatiquement un tableau de contingence des associations entre les entrées catégorielles et utiliser ces associations lorsque les données sont générées pour ces entrées.
- Spécifier l'analyse de sensibilité pour vérifier l'effet de la variation de la valeur d'une entrée fixe ou d'un paramètre de distribution d'une entrée simulée.
- Spécifier des options avancées telles que le nombre maximum d'observations à générer ou un échantillonnage des extrémités.
- Personnaliser la sortie.
- Enregistrer les données simulées dans un fichier de données.

Pour concevoir une simulation basée sur des équations personnalisées

1. A partir des menus, sélectionnez :

   Analyse > **Simulation...**
2. Cliquez sur **Entrer les Equations** puis cliquez sur **Poursuivre**.
3. Cliquez sur **Nouvelle équation** sur l'onglet Modèle (du Générateur de simulation) afin de définir chaque équation dans votre modèle prédictif.
4. Cliquez sur l'onglet Simulation et spécifiez les distributions de probabilité des entrées simulées ainsi que les valeurs des entrées fixes. Si le jeu de données actif contient des données historiques relatives aux entrées simulées, cliquez sur **Ajuster tout** pour déterminer automatiquement la distribution la mieux adaptée à chaque entrée, ainsi que les corrélations entre celles-ci. Pour chaque entrée simulée non ajustée aux données historiques, vous devez explicitement indiquer une distribution en sélectionnant un type et en entrant les paramètres requis.
5. Cliquez sur **Exécuter** pour exécuter la simulation. Par défaut, le plan de simulation qui spécifie les détails de la simulation, est enregistré à l'emplacement indiqué dans les paramètres d'enregistrement.

Les options suivantes sont disponibles :

- Modifier l'emplacement du plan de simulation enregistré.
- Spécifier les corrélations connues existant entre les entrées simulées.
- Calculer automatiquement un tableau de contingence des associations entre les entrées catégorielles et utiliser ces associations lorsque les données sont générées pour ces entrées.
- Spécifier l'analyse de sensibilité pour vérifier l'effet de la variation de la valeur d'une entrée fixe ou d'un paramètre de distribution d'une entrée simulée.
- Spécifier des options avancées telles que le nombre maximum d'observations à générer ou un échantillonnage des extrémités.
- Personnaliser la sortie.
- Enregistrer les données simulées dans un fichier de données.

Pour concevoir une simulation sans modèle prédictif

1. A partir du menu, sélectionnez :

   Analyse > **Simulation...**
2. Cliquez sur **Créer des données simulées**, puis sur **Poursuivre**.
3. Sur l'onglet Modèle (dans le Générateur de simulation), sélectionnez les champs à simuler. Vous pouvez sélectionner des champs à partir du jeu de données actif ou définir de nouveaux champs en cliquant sur **Nouveau**.
4. Cliquez sur l'onglet Simulation et spécifiez les distributions de probabilité des champs à simuler. Si le jeu de données actif contient des données historiques relatives à ces champs, cliquez sur **Ajuster tout** pour déterminer automatiquement la distribution la mieux adaptée aux données, ainsi que les corrélations entre les champs. Pour les champs non ajustés aux données historiques, vous devez explicitement indiquer une distribution en sélectionnant un type de distribution et en entrant les paramètres requis.

5. Cliquez sur **Exécuter** pour exécuter la simulation. Par défaut, les données simulées sont sauvegardées dans le nouveau jeu de données défini dans les paramètres d'enregistrement. De plus, le plan de simulation, qui spécifie les détails de la simulation, est enregistré à l'emplacement indiqué dans les paramètres d'enregistrement.

Les options suivantes sont disponibles :

- Modifier l'emplacement des données simulées ou du plan de simulation enregistré.
- Spécifier les corrélations entre les champs simulés.
- Calculer automatiquement un tableau de contingence des associations entre les champs catégoriels et utiliser ces associations lorsque les données sont générées pour ces champs.
- Spécifier l'analyse de sensibilité pour vérifier l'effet de la variation d'un paramètre de distribution d'un champ simulé.
- Spécifier des options avancées telles que le nombre d'observations à générer.

Exécution d'une simulation à partir d'un plan de simulation

Deux options sont disponibles pour exécuter une simulation à partir d'un plan de simulation. Vous pouvez utiliser soit la boîte de dialogue Exécuter la simulation, conçue principalement pour exécuter une simulation à partir d'un plan de simulation, soit le Générateur de simulation.

Pour utiliser la boîte de dialogue Exécuter la simulation :

1. A partir des menus, sélectionnez :

   Analyse > **Simulation...**
2. Cliquez sur **Ouvrir un plan de simulation existant**.
3. Vérifiez que la case **Ouvrir dans le Générateur de simulation** n'est pas cochée et cliquez sur **Poursuivre**.
4. Ouvrez le plan de simulation.
5. Cliquez sur **Exécuter** dans la boîte de dialogue Exécuter la simulation.

Pour exécuter la simulation à partir du Générateur de simulation :

1. A partir des menus, sélectionnez :

   Analyse > **Simulation...**
2. Cliquez sur **Ouvrir un plan de simulation existant**.
3. Sélectionnez la case **Ouvrir dans le Générateur de simulation** et cliquez sur **Poursuivre**.
4. Ouvrez le plan de simulation.
5. Modifiez les paramètres requis dans l'onglet Simulation.
6. Cliquez sur **Exécuter** pour exécuter la simulation.

Vous pouvez au besoin procéder aux actions suivantes :

- Configurer ou modifier l'analyse de sensibilité pour vérifier l'effet de la variation de la valeur d'une entrée fixe ou d'un paramètre de distribution d'une entrée simulée.
- Réajuster les distributions et les corrélations des entrées simulées aux nouvelles données.
- Modifier la distribution d'une entrée simulée.
- Personnaliser la sortie.
- Enregistrer les données simulées dans un fichier de données.

Générateur de simulation

Le Générateur de simulation fournit l'ensemble complet des fonctionnalités permettant de concevoir et d'exécuter des simulations. Elle vous permet de réaliser les tâches générales suivantes :

- Concevoir et exécuter une simulation pour un modèle IBM SPSS défini dans un fichier de modèle PMML.

- Concevoir et exécuter une simulation pour un modèle prédictif défini par un ensemble d'équations personnalisées.
- Concevoir et exécuter une simulation qui génère des données en l'absence d'un modèle prédictif.
- Exécuter une simulation basée sur un plan de simulation existant, et éventuellement modifier des paramètres du plan.

Onglet Modèle

Pour les simulations basées sur un modèle prédictif, l'onglet Modèle indique la source du modèle. Pour les simulations qui n'incluent pas de modèle prédictif, l'onglet Modèle spécifie les champs à simuler.

Sélectionner un fichier de modèle SPSS : Cette option spécifie que le modèle prédictif est défini dans un fichier de modèle IBM SPSS. Un fichier de modèle IBM SPSS est un fichier XML ou une archive compressée (fichier .zip) qui contient le modèle PMML créé à partir d'IBM SPSS Statistiques ou d'IBM SPSS Modeler. Les modèles prédictifs sont créés par des procédures telles que la régression linéaire et les arbres de décisions dans IBM SPSS Statistiques, et ils peuvent être exportés vers un fichier de modèle. Vous pouvez sélectionne un fichier de modèle différent en cliquant sur **Parcourir** et en accédant à son emplacement.

Modèles PMML pris en charge par les simulations

- Régression linéaire
- Modèle linéaire automatique
- Modèle linéaire généralisé
- Modèles mixtes linéaires généralisés
- Modèle linéaire général
- Régression logistique binaire
- Régression logistique multinomiale
- Régression multinomiale ordinale
- Régression de Cox
- Arbre
- Arbre boosté (C5)
- Discriminant
- Cluster en deux étapes
- Nuées dynamiques
- Réseau neuronal
- Ensemble de règles (Liste de décision)

Remarque :

- Les modèles PMML qui possèdent plusieurs champs cible (variables) ou scissions ne sont pas pris en charge par les simulations.
- Les valeurs des entrées de chaîne dans les modèles de régression binaire sont limitées à 8 octets. Si vous ajustez de telles chaînes dans le jeu de données actif, veillez à ce que les valeurs dans les données ne dépassent pas 8 octets de longueur. Les valeurs de données qui dépassent 8 octets sont exclues de la distribution catégorielle associée pour l'entrée, et sont affichées comme n'ayant pas de correspondance dans la table de sortie Catégories sans correspondance.

Entrez les équations pour le modèle : Cette option spécifie que le modèle prédictif consiste en une ou plusieurs équations personnalisées que vous devez créer. Créez les équations en cliquant sur **Nouvelle équation**. Cette action ouvre l'Editeur d'équation. Vous pouvez modifier des équations existantes, les copier et les utiliser en tant que modèles pour de nouvelles équations, les réorganiser et les supprimer.

- Le Générateur de simulation ne prend pas en charge les systèmes d'équations simultanées ou les équations non linéaires dans la variable cible.

- Les équations personnalisées sont évaluées en fonction de leur ordre de spécification. Si l'équation d'une cible donnée dépend d'une autre cible, alors la seconde cible doit être définie par une équation précédente.

  Par exemple, considérons l'ensemble des trois équations suivantes, l'équation du *profit* dépend des valeurs des *revenus* et des *dépenses*, par conséquent les équations des *revenus* et des *dépenses* doivent précéder celle du *profit*.

```
revenue = price*volume
expenses = fixed + volume*(unit_cost_materials + unit_cost_labor)
profit = revenue - expenses
```

Créer des données simulées sans modèle : Sélectionnez cette option pour simuler des données sans modèle prédictif. Indiquez les champs à simuler en les sélectionnant dans le jeu de données actif ou en cliquant sur **Nouveau** pour définir de nouveaux champs.

Editeur d'équation

L'Editeur d'équation vous permet de créer ou de modifier une équation personnalisée de votre modèle prédictif.

- L'expression de l'équation peut contenir des champs provenant du jeu de données actif ou de nouveaux champs d'entrée que vous définissez dans l'Editeur d'équation.
- Vous pouvez spécifier les propriétés de la cible, par exemple son niveau de mesure et ses libellés de valeur, et choisir si une sortie est générée pour la cible.
- Vous pouvez utiliser des cibles provenant d'équations précédemment définies en tant qu'entrées de l'équation en cours, ce qui vous permet de créer des équations couplées.
- Vous pouvez aussi ajouter un commentaire descriptif à l'équation. Les commentaires s'affichent sur l'onglet Modèle en même temps que l'équation.

1. Saisissez le nom de la cible. Si vous le souhaitez, cliquez sur **Modifier** en dessous de la zone de texte Cible pour ouvrir la boîte de dialogue Entrées définies, qui vous permet de modifier les propriétés par défaut de la cible.
2. Pour construire une expression, vous pouvez soit coller les composants dans le champ Expression numérique, soit les saisir au clavier directement dans ce même champ.

- Il est possible de construire une expression en utilisant des champs provenant du jeu de données actif ou de définir de nouvelles entrées en cliquant sur le bouton **Nouveau**. Cette action ouvre la boîte de dialogue Entrées définies.
- Vous pouvez coller des fonctions en sélectionnant un groupe dans la liste Groupe de fonctions, puis en double-cliquant sur la fonction désirée dans la liste Fonctions (ou sélectionnez la fonction, puis cliquez sur la flèche adjacente à la liste Groupe de fonctions). Entrez tous les paramètres indiqués par un point d'interrogation. Le groupe de fonctions libellé **Tous** répertorie toutes les fonctions disponibles. Une brève description de la fonction sélectionnée apparaît dans une zone particulière de la boîte de dialogue.
- Les constantes chaîne doivent être présentées entre guillemets.
- Si des valeurs contiennent des chiffres décimaux, utilisez la virgule comme indicateur décimal.

Remarque : Les simulations ne prennent pas en charge les équations personnalisées contenant des cibles chaîne.

Entrées définies

La boîte de dialogue Entrées définies vous permet de définir de nouvelles entrées et de définir les propriétés des cibles.

- Si une entrée devant être utilisée dans une équation n'existe pas dans le jeu de données actif, vous devez la définir pour pouvoir l'utiliser dans l'équation.

- Si vous simulez des données sans modèle prédictif, vous devez définir l'ensemble des entrées simulées qui ne figurent pas dans le jeu de données actif.

Nom. Spécifiez le nom d'une cible ou d'une entrée.

Cible : Vous pouvez spécifier le niveau de mesure d'une cible. Le niveau de mesure par défaut est continu. Vous pouvez également indiquer si une sortie sera créée pour la cible. Par exemple, pour un ensemble d'équations couplées, il se peut que vous soyez intéressé uniquement par la génération de la sortie de la cible de l'équation finale, par conséquent dans un tel cas vous supprimerez la sortie des autres cibles.

Entrée à simuler : Cette option indique que les valeurs de l'entrée seront simulées en fonction d'une distribution de probabilité spécifique (cette distribution de probabilité est spécifiée sur l'onglet Simulation). Le n de mesure détermine l'ensemble par défaut des distributions prises en compte lors de la recherche de la distribution la mieux adaptée à l'entrée (cliquez sur **Ajustement** ou **Ajuster tout** dans l'onglet Simulation). Par exemple, si le niveau de mesure est continu, la distribution normale (appropriée aux données continues) sera prise en compte, mais pas la distribution binomiale.

Remarque : Sélectionnez un niveau de mesure de chaîne pour les entrées de chaîne. Les entrées de chaîne à simuler sont limitées à la distribution catégorielle.

Entrée de valeur fixe : Cette option spécifie que la valeur de l'entrée est connue et sera fixée à la valeur connue. Les entrées fixes peuvent être des valeurs numériques ou des chaînes. Spécifiez la valeur de l'entrée fixe. Les valeurs de chaîne ne doivent pas être entre guillemets.

Libellés de valeurs : Vous pouvez spécifier des libellés de valeurs pour les cibles, les entrées simulées et les entrées fixes. Les libellés de valeur sont utilisés dans les graphiques et les tableaux de sortie.

Onglet Simulation

L'onglet Simulation spécifie toutes les propriétés de la simulation, autres que le modèle prédictif associé. Vous pouvez réaliser les tâches générales suivantes sur l'onglet Simulation :

- Spécifier les distributions de probabilité des entrées simulées et les valeurs des entrées fixes.
- Spécifier les corrélations existant entre les entrées simulées. Pour les entrées catégorielles, vous pouvez spécifier que les associations qui existent entre ces entrées du jeu de données actif soient utilisées lorsque les données sont générées pour ces entrées.
- Spécifier les options avancées telles que l'échantillonnage des extrémités et les critères d'ajustement des distributions aux données historiques.
- Personnaliser la sortie.
- Spécifier l'emplacement d'enregistrement du plan de simulation et éventuellement enregistrer les données simulées.

Champs simulés

Pour exécuter une simulation, chaque zone d'entrée doit être spécifiée en tant que zone fixe ou simulée. Les entrées simulées sont celles dont les valeurs sont incertaines et seront générées à partir d'une distribution de probabilité spécifique. Lorsque les données historiques sont disponibles pour les entrées à simuler, les distributions les mieux adaptées peuvent être déterminées automatiquement, ainsi que les corrélations entre ces entrées. Il est également possible de spécifier manuellement des distributions ou des corrélations si aucune donnée historique n'est disponible, ou si vous avez besoin d'utiliser des distributions ou des corrélations spécifiques.

Les entrées fixes sont celles dont les valeurs sont connues. Elles restent constantes pour chaque observation générée dans la simulation. Par exemple, vous avez un modèle de régression linéaire concernant les ventes qui est fonction d'un certain nombre d'entrées, dont le prix, et vous souhaitez que le prix soit fixe et corresponde au prix du marché actuel. Vous spécifierez donc le prix en tant qu'entrée fixe.

Pour les simulations basées sur un modèle prédictif, chaque prédicteur du modèle représente un champ d'entrée pour la simulation. Pour les simulations qui n'incluent pas de modèle prédictif, les champs spécifiés sur l'onglet Modèle représentent les entrées de la simulation.

Ajustement automatique des distributions et calcul automatique des corrélations pour les entrées simulées : Si le jeu de données actif contient des données historiques relatives aux entrées à simuler, vous pouvez déterminer automatiquement les distributions les mieux adaptées à ces entrées, ainsi que les corrélations entre celles-ci. Les étapes à suivre sont les suivantes :

1. Vérifiez que chaque entrée que vous souhaitez simuler est mise en correspondance avec la zone appropriée dans le jeu de données actif. Les entrées sont répertoriées dans la colonne Entrée et la colonne Ajuster à affiche la zone correspondante dans le jeu de données actif. Vous pouvez mettre les entrées en correspondance avec un champ différent du jeu de données actif en sélectionnant un élément différent dans la liste déroulante Ajuster à.

   La valeur *-Aucun-* dans la colonne Ajuster à indique que l'entrée ne peut être automatiquement mise en correspondance avec un champ du jeu de données actif. Par défaut, les entrées sont mises en correspondance avec des champs du jeu de données dont le nom, le niveau de mesure et le type (numérique ou chaîne) correspondent. Si le jeu de données actif ne contient pas de données historiques pour une entrée particulière, déterminez manuellement la distribution à utiliser pour cette entrée, ou spécifiez que l'entrée est une entrée fixe.
2. Cliquez sur **Tout ajuster**.

La distribution la mieux adaptée et ses paramètres associés s'affichent alors dans la colonne Distribution en même temps qu'un tracé de la distribution superposé à un histogramme (graphique à barres) représentant les données historiques. Les corrélations entre les entrées simulées s'affichent dans les paramètres Corrélations. Vous pouvez examiner les résultats de l'ajustement et personnaliser l'ajustement automatique de la distribution d'une entrée particulière en sélectionnant la ligne de cette entrée et en cliquant sur **Détails de l'ajustement**. Pour plus d'informations, reportez-vous à la rubrique «Détails de l'ajustement », à la page 307.

Vous pouvez exécuter l'ajustement automatique de la distribution d'une entrée particulière en sélectionnant la ligne de cette entrée et en cliquant sur **Ajuster**. Les corrélations de toutes les entrées simulées mises en correspondance à des champs du jeu de données actif sont calculées automatiquement.

Remarque :

- Les observations comportant des valeurs manquantes pour n'importe quelle entrée simulée sont exclues de l'ajustement de la distribution, du calcul des corrélations et du calcul du tableau de contingence facultatif (pour les entrées avec une distribution Catégorielle). Vous pouvez également indiquer si les valeurs manquantes de l'utilisateur des entrées comportant une distribution catégorielle sont traitées comme étant valides. Par défaut, elles sont traitées comme étant manquantes. Pour plus d'informations, voir la rubrique «Options avancées », à la page 309.
- Pour les entrées continues et ordinales, s'il est impossible de trouver un ajustement acceptable pour une des distributions testées, la distribution empirique est alors utilisée. Pour les entrées continues, la distribution empirique est la fonction de distribution cumulée des données historiques. Pour les entrées ordinales, la distribution empirique est la distribution catégorielle des données historiques.

Spécification manuelle des distributions : Vous pouvez spécifier manuellement la distribution de probabilité de n'importe quelle entrée simulée en sélectionnant une distribution dans la liste déroulante **Type** et en saisissant les paramètres de la distribution dans la grille Paramètres. Une fois les paramètres d'une distribution définis, un tracé de la distribution s'appuyant sur les paramètres spécifiés s'affiche à côté de la grille Paramètres. Voici quelques remarques sur des distributions particulières :

- **Catégorielle :** La distribution catégorielle décrit un champ d'entrée comportant un nombre fixe de valeurs, appelées catégories. Chaque catégorie possède une probabilité associée, de sorte que la somme des probabilités de toutes les catégories est égale à 1. Pour entrer une catégorie, cliquez sur la colonne de gauche dans la grille Paramètres et indiquez la valeur de catégorie. Entrez la probabilité associée à la catégorie dans la colonne de droite.

  Remarque : Les entrées catégorielles d'un modèle PMML possèdent des catégories qui sont définies à partir du modèle et qui ne sont pas modifiables.

- **Binomiale négative - Echecs :** Décrit la distribution du nombre d'échecs dans une séquence d'essais avant qu'un nombre spécifique de succès ne soit observé. Le paramètre *seuil* est le nombre spécifique de succès et le paramètre *prob* est la probabilité de succès pour chaque essai donné.
- **Binomiale négative - Essais :** Décrit la distribution du nombre d'essais requis avant qu'un nombre spécifique de succès ne soit observé. Le paramètre *seuil* est le nombre spécifique de succès et le paramètre *prob* est la probabilité de succès pour chaque essai donné.
- **Plage :** Cette distribution consiste en un ensemble d'intervalles avec une probabilité attribuée à chaque intervalle de sorte que la somme des probabilités sur tous les intervalles est égale à 1. Les valeurs d'un intervalle donné sont tirées d'une distribution uniforme définie sur cet intervalle. Les intervalles sont spécifiés en entrant une valeur minimale, une valeur maximale et une probabilité.

  Par exemple, vous pensez que le coût des matières premières a 40 % de chance de chuter d'environ 10 - 15 $ par unité et une chance de 60 % de chuter d'environ 15 - 20 $ par unité. Vous procédez à la modélisation du coût à l'aide d'une distribution Plage consistant en deux intervalles [10 - 15] et [15 - 20], définissez la probabilité associée au premier intervalle à 0,4 et la probabilité associée au second intervalle à 0,6. Les intervalles n'ont pas besoin d'être contigus et ils peuvent même se chevaucher. Par exemple, vous pouvez spécifier les intervalles 10 - 15 $ et 20-25 $ ou 10-15 $ et 13-16 $.
- **Weibull :** Le paramètre *c* est un paramètre d'emplacement facultatif qui spécifie l'emplacement de l'origine de la distribution.

Les paramètres des distributions suivantes ont la même signification que dans les fonctions de variable aléatoire associées disponibles dans la boîte de dialogue Calculer la variable : Bernoulli, Bêta, Binomiale, Exponentielle, Gamma, Lognormale, Binomiale négative (essais et échecs, Normale, Poisson et Uniforme.

Spécification des entrées fixes : Spécifiez une entrée fixe en sélectionnant Fixe dans la liste déroulante **Type** située dans la colonne Distribution et en saisissant la valeur de l'entrée fixe. Cette valeur peut être numérique ou alphanumérique selon que l'entrée est une entrée numérique ou chaîne. Les valeurs de chaîne ne doivent pas être entre guillemets.

Spécification des bornes sur les valeurs simulées : La plupart des distributions prennent en charge les bornes supérieure et inférieure sur les valeurs simulées. La borne inférieure peut être spécifiée en saisissant sa valeur dans la zone de texte **Min** et la borne supérieure peut être spécifiée en saisissant sa valeur dans la zone de texte **Max**.

Verrouillage des entrées : Le verrouillage d'une entrée, en cochant la case correspondante dans la colonne indiquée par une icône en forme de verrou, exclut cette entrée de l'ajustement automatique de la distribution. Cette option est très utile lorsque vous spécifiez manuellement une distribution ou une valeur fixe et souhaitez vous assurer qu'elle ne sera pas affectée par un ajustement automatique. Le verrouillage est également très utile si vous avez l'intention de partager votre plan de simulation avec des utilisateurs qui l'exécuteront dans la boîte de dialogue Exécuter la simulation, et que vous souhaitez empêcher ces utilisateurs de modifier certaines entrées. A cette fin, les spécifications d'entrées verrouillées ne peuvent être modifiées dans la boîte de dialogue Exécuter la simulation.

Analyse de sensibilité. L'analyse de sensibilité permet d'étudier l'effet des changements systématiques d'une entrée fixe ou d'un paramètre de distribution pour une entrée simulée en générant un ensemble indépendant de cas simulés - en fait, une simulation séparée - pour chaque valeur spécifiée. Pour spécifier l'analyse de sensibilité, sélectionnez une entrée simulée ou une entrée fixe et cliquez sur **Analyse de sensibilité**. L'analyse de sensibilité est limitée à une seule entrée fixe ou un seul paramètre de distribution associé à une entrée simulée. Pour plus d'informations, voir la rubrique «Analyse de sensibilité », à la page 308 .

Icônes de statut d'ajustement

Les icônes de la colonne Ajuster à indiquent le statut d'ajustement de chaque zone d'entrée.

Tableau 8. Icônes de statut

Icône	Descriptif
	Aucune distribution n'a été spécifiée pour l'entrée et l'entrée n'a pas été spécifiée en tant qu'entrée fixe. Pour exécuter la simulation, vous devez soit spécifier une distribution pour cette entrée, soit la définir en tant qu'entrée fixe et lui affecter une valeur.
	L'entrée a été précédemment ajustée à un champ qui n'existe pas dans le jeu de données actif. Aucune action n'est nécessaire sauf si vous souhaitez réajuster la distribution de l'entrée à le jeu de données actif.
	La distribution la mieux adaptée a été remplacée par une distribution automatique de la boîte de dialogue Ajuster les détails.
	L'entrée est définie sur la distribution la mieux adaptée.
	La distribution a été spécifiée manuellement ou des itérations de l'analyse de sensibilité ont été spécifiées pour cette entrée.

Détails de l'ajustement

La boîte de dialogue Détails de l'ajustement affiche les résultats de l'ajustement automatique de la distribution d'une entrée spécifique. Les distributions sont ordonnées par la qualité de l'ajustement, la distribution la plus appropriée étant répertoriée en tête de liste. Vous pouvez remplacer la distribution la mieux adaptée en sélectionnant le bouton radio correspondant à la distribution de votre choix dans la colonne Utilisation. En sélectionnant un bouton radio dans la colonne Utilisation, vous affichez également un tracé de la distribution superposé à un histogramme (graphique à barres) représentant les données historiques de l'entrée.

Statistiques de l'ajustement : Par défaut, pour les champs continus, le test d'Anderson-Darling est utilisé pour déterminer la qualité de l'ajustement. Pour les champs continus uniquement, vous pouvez aussi choisir le test de Kolmogorov-Smirnoff pour déterminer la qualité de l'ajustement. Ce choix peut être fait dans les paramètres des Options avancées. Pour les entrées continues, les résultats des deux tests s'affichent dans la colonne Statistiques de l'ajustement avec une indication du test choisi (A pour Anderson-Darling et K pour Kolmogorov-Smirnoff) pour l'ordonnancement des distributions. Pour les entrées ordinales et nominales, le test du khi-carré est utilisé. Les valeurs p associées aux tests sont également affichées.

Paramètres : Les paramètres de distribution associés à chaque distribution ajustée sont affichés dans la colonne Paramètres. Les paramètres des distributions suivantes ont la même signification que dans les fonctions de variable aléatoire associées disponibles dans la boîte de dialogue Calculer la variable : Bernoulli, Bêta, Binomiale, Exponentielle, Gamma, Lognormale, Binomiale négative (essais et échecs, Normale, Poisson et Uniforme. Pour la distribution catégorielle, les noms des paramètres sont les catégories et les valeurs de paramètres sont les probabilités associées.

Réajustement avec un ensemble de distribution personnalisé : Par défaut, le niveau de mesure de l'entrée est utilisé pour déterminer l'ensemble des distributions considéré pour l'ajustement automatique de la distribution. Par exemple, les distributions continues, telles que les distributions Lognormale et Gamma, sont prises en considération lors de l'ajustement d'une entrée continue, mais les distributions discrètes, telles que les distributions Poisson et Binomiale, ne le sont pas. Vous pouvez choisir un sous-ensemble des distributions par défaut en sélectionnant les distributions dans la colonne Réajuster. Vous pouvez également remplacer l'ensemble des distributions par défaut en sélectionnant un niveau de mesure différent dans la liste déroulante **Traiter comme (Mesure)**, puis en sélectionnant les distributions dans la colonne Réajuster. Cliquez sur **Exécuter le réajustement** pour réajuster l'ensemble de distributions personnalisé.

Remarque :

- Les observations comportant des valeurs manquantes pour n'importe quelle entrée simulée sont exclues de l'ajustement de la distribution, du calcul des corrélations et du calcul du tableau de contingence facultatif (pour les entrées avec une distribution Catégorielle). Vous pouvez également indiquer si les valeurs manquantes de l'utilisateur des entrées comportant une distribution catégorielle sont traitées comme étant valides. Par défaut, elles sont traitées comme étant manquantes. Pour plus d'informations, voir la rubrique «Options avancées », à la page 309.
- Pour les entrées continues et ordinales, s'il est impossible de trouver un ajustement acceptable pour une des distributions testées, la distribution empirique est alors utilisée. Pour les entrées continues, la distribution empirique est la fonction de distribution cumulée des données historiques. Pour les entrées ordinales, la distribution empirique est la distribution catégorielle des données historiques.

Analyse de sensibilité

L'analyse de sensibilité vous permet de vérifier l'effet de la variation d'une entrée fixe ou d'un paramètre de distribution associé à une entrée simulée, pour un ensemble spécifique de valeurs. Un ensemble indépendant d'observations simulées - en fait, une simulation distincte - est généré pour chaque valeur spécifiée, ce qui vous permet de vérifier les effets de la variation de l'entrée. Chaque ensemble d'observations simulées est appelé **itération**.

Itérer : Cette option vous permet d'indiquer l'ensemble des valeurs appliquées à l'entrée et en fonction desquelles l'entrée va varier.

- Si vous préférez faire varier la valeur d'un paramètre de distribution, sélectionnez ce paramètre dans la liste déroulante. Entrez l'ensemble des valeurs dans la grille représentant les valeurs de paramètres par itération. Cliquez sur **Poursuivre** pour ajouter les valeurs spécifiées à la grille Paramètres de l'entrée associée, avec un indice indiquant le nombre d'itérations de la valeur.
- Pour les distributions Catégorielle et Plage, les probabilités des catégories ou des intervalles, respectivement, peuvent être modifiées mais les valeurs des catégories et les points finaux des intervalles ne peuvent l'être. Sélectionnez une catégorie ou un intervalle dans la liste déroulante et spécifiez l'ensemble des probabilités dans la grille des valeurs de paramètres par itération. Les probabilités des autres catégories ou intervalles seront automatiquement ajustées en conséquence.

Aucune itération : Utilisez cette option pour annuler les itérations d'une entrée. Cliquez sur **Poursuivre** pour supprimer les itérations.

Corrélations

Les champs d'entrée à simuler sont souvent corrélés, comme par exemple, la taille et la pondération. Les corrélations existant entre les entrées qui seront simulées doivent être prises en compte afin d'assurer que les valeurs simulées conservent ces corrélations.

Recalculer les corrélations lors de l'ajustement : Cette option indique que les corrélations entre les entrées simulées sont automatiquement calculées lors de l'ajustement des distributions à le jeu de données actif à l'aide de l'action **Tout ajuster** ou **Ajuster** située dans les paramètres Champs simulés.

Ne pas recalculer les corrélations lors de l'ajustement : Choisissez cette option si vous souhaitez spécifier manuellement les corrélations et éviter qu'elles ne soient remplacées lors de l'ajustement automatique des distributions à le jeu de données actif. Les valeurs entrées dans la grille Corrélations doivent être comprises entre -1 et 1. La valeur 0 indique qu'il n'existe aucune corrélation entre la paire d'entrées associée.

Restaurer : Cette option rétablit toutes les corrélations sur 0.

Utiliser un tableau de contingence à entrées multiples pour les entrées comportant une distribution catégorielle : Pour les entrées comportant une distribution catégorielle, vous pouvez automatiquement calculer un tableau de contingence à entrées multiple à partir du jeu de données actif qui décrit les associations entre ces entrées. Le tableau de contingence est ensuite utilisé lorsque les données sont générées pour ces entrées. Si vous choisissez d'enregistrer le plan de simulation, le tableau de contingence est enregistré dans le fichier de plan et est utilisé lorsque vous exécutez le plan.

- **Calculer le tableau de contingence à partir du jeu de données actif :** Si vous travaillez avec un plan de simulation existant qui contient déjà un tableau de contingence, il vous est possible de recalculer

le tableau de contingence à partir du jeu de données actif. Cette action remplace le tableau de contingence du fichier de plan chargé.

- **Utiliser le tableau de contingence à partir du plan de simulation chargé :** Par défaut, lorsque vous chargez un plan de simulation qui contient un tableau de contingence, le tableau du plan est utilisé. Vous avez la possibilité de recalculer le tableau de contingence à partir du jeu de données actif à l'aide de l'option **Calculer le tableau de contingence à partir du jeu de données actif**.

Options avancées

Nombre maximum d'observations : Cette option indique le nombre maximum d'observations de données simulées et les valeurs cible associées à générer. Lorsque l'analyse de sensibilité est sélectionnée, cette option indique le nombre maximal d'observations pour chaque itération.

Cible des critères d'arrêt : Si votre modèle prédictif contient plusieurs cibles, vous pouvez choisir la cible à laquelle s'appliquent les critères d'arrêt.

Critères d'arrêt : Ces options permettent de spécifier les critères d'arrêt de la simulation, en général avant que le nombre maximal d'observations autorisé ait été généré.

- **Poursuivre jusqu'à atteindre le maximum :** Cette option indique que des observations simulées seront générées tant que le nombre maximal d'observations n'est pas atteint.
- **Arrêter une fois les extrémités échantillonnées :** Utilisez cette option si vous souhaitez vous assurer que l'une des extrémités d'une distribution cible spécifique a bien été échantillonnée. Les observations simulées seront générées tant que l'échantillonnage des extrémités spécifié n'est pas terminé ou que le nombre maximal d'observations n'est pas atteint. Si votre modèle prédictif contient plusieurs cibles, choisissez celle à laquelle s'appliquent ces critères d'arrêt dans la liste déroulante **Cible des critères d'arrêt**.

  Type : Vous pouvez définir la limite de la zone d'extrémité en spécifiant une valeur de cible, par ex. 10 000 000 ou un percentile, par ex. le 99e percentile. Si vous choisissez Valeur dans la liste déroulante **Type**, saisissez la valeur de la limite dans la zone de texte Valeur et choisissez dans la liste déroulante **Côté** le côté de la zone d'extrémité (Gauche ou Droite). Si vous choisissez Percentile dans la liste déroulante **Type**, saisissez une valeur dans la zone de texte Percentile.

  Effectif : Spécifiez ici le nombre de valeurs de la cible qui doivent se situer dans la zone d'extrémité afin de vous assurer que l'extrémité a été correctement échantillonnée. La génération d'observations va cesser une fois ce nombre atteint.
- **Arrêter lorsque l'intervalle de confiance de la moyenne atteint le seuil spécifié :** Utilisez cette option si vous souhaitez vous assurer que la moyenne d'une cible donnée est connue avec un degré de précision spécifique. Les observations simulées seront générées tant que le degré de précision spécifique ou le nombre maximal d'observations n'est pas atteint. Pour utiliser cette option, vous devez spécifier un niveau de confiance et un seuil. Les observations simulées seront générées tant que l'intervalle de confiance associé au niveau spécifique n'atteint pas le seuil. Par exemple, vous pouvez utiliser cette option pour spécifier que les observations doivent être générées tant que l'intervalle de confiance de la moyenne à un niveau de confiance de 95 % n'atteint pas 5 % de la valeur moyenne. Si votre modèle prédictif contient plusieurs cibles, choisissez celle à laquelle s'appliquent ces critères d'arrêt dans la liste déroulante **Cible des critères d'arrêt**.

  Type de seuil : Le seuil peut être spécifié en tant que valeur numérique ou en tant que pourcentage de la moyenne. Si vous choisissez Valeur dans la liste déroulante **Type de seuil**, saisissez le seuil dans la zone de texte Seuil sous forme de valeur. Si vous choisissez Pourcentage dans la liste déroulante **Type de seuil**, saisissez une valeur dans la zone de texte Seuil sous forme de pourcentage.

Nombre d'observations à échantillonner : Cette option permet de spécifier le nombre d'observations à utiliser lors de l'ajustement automatique des distributions des entrées simulées à le jeu de données actif. Si votre jeu de données est très grand, vous voudrez peut-être limiter le nombre d'observations à utiliser pour l'ajustement de la distribution. En sélectionnant **Limiter à N observations**, seules les N premières observations sont utilisées.

Critères de qualité de l'ajustement (continus) : Pour les entrées continues, vous pouvez utiliser le test d'Anderson-Darling ou le test de Kolmogorov-Smirnoff relatif à la qualité de l'ajustement afin de

classer les distributions des entrées simulées lors de leur ajustement à le jeu de données actif. Le test d'Anderson-Darling est sélectionné par défaut et est particulièrement recommandé si vous souhaitez assurer le meilleur ajustement possible dans les zones d'extrémité.

Distribution empirique : Pour les entrées continues, la distribution empirique est la fonction de distribution cumulée des données historiques. Cette option vous permet de spécifier le nombre de casiers utilisés pour calculer la distribution empirique des entrées continues. La valeur par défaut est 100 et la valeur maximale est 1000.

Dupliquer les résultats. Définir une valeur de départ aléatoire vous permet de dupliquer votre simulation. Spécifiez un entier ou cliquez sur **Générer**, ce qui crée un entier pseudo-aléatoire compris entre 1 et 2147483647, inclus. La valeur par défaut est 629111597.

Remarque : Pour une valeur de départ aléatoire donnée, les résultats seront dupliqués jusqu'à ce que le nombre d'unités d'exécution change. Sur un ordinateur particulier, le nombre d'unités d'exécution est constant à moins que vous ne le changiez en exécutant la commande SET THREADS. Le nombre d'unités d'exécution peut changer si vous exécutez la simulation sur un autre ordinateur car un algorithme interne est utilisé pour déterminer le nombre d'unités d'exécution pour chaque ordinateur.

Valeurs manquantes de l'utilisateur pour les entrées comportant une distribution catégorielle : Option indiquant si les valeurs manquantes de l'utilisateur pour les entrées comportant une distribution catégorielle sont traitées comme étant valides. Les valeurs système manquantes et les valeurs manquantes des utilisateurs pour tous les autres types d'entrée sont traitées comme étant non valides. Toutes les entrées doivent avoir des valeurs valides pour qu'une observation puisse être incluse dans l'ajustement de la distribution, le calcul des corrélations et le calcul d'un tableau de contingence optionnel.

Fonctions de densité

Ces paramètres vous permettent de personnaliser la sortie des fonctions de densité de probabilité et des fonctions de distribution cumulée pour les cibles continues, ainsi que les graphiques à barres des valeurs prédites pour les cibles catégorielles.

Fonction de densité de probabilité (FDP) : La fonction de densité de probabilité affiche la distribution des valeurs cible. Pour les cibles continues, elle vous permet de déterminer la probabilité que la cible se trouve dans une zone donnée. Pour les cibles catégorielles (cibles dont le niveau de mesure est nominal ou ordinal), un graphique à barres est généré, qui affiche le pourcentage d'observations situées dans chaque catégorie de la cible. Des options supplémentaires sont disponibles pour les cibles catégorielles des modèles PMML, dont le paramètre Valeurs de catégorie à rapporter, décrit ultérieurement.

Pour les modèles de cluster en deux étapes et les modèles de cluster en nuées dynamiques, un graphique à barres représentant l'appartenance aux clusters est généré.

Fonction de distribution cumulée (FDC) : La fonction de distribution cumulée affiche la probabilité qu'une valeur de la cible soit inférieure ou égale à une valeur donnée. Elle est disponible uniquement pour les variables continues.

Positions des curseurs : Vous pouvez spécifier les positions initiales des lignes de référence sur les graphiques FDP et FDC. Les valeurs saisies dans les lignes inférieure et supérieure indiquent les positions le long de l'axe horizontal, et non les percentiles. Vous pouvez supprimer la ligne inférieure en sélectionnant **-Infini** ou la ligne supérieure en sélectionnant **Infini**. Par défaut, les lignes sont positionnées sur les 5e et 95e percentiles. Lorsque plusieurs fonctions de distribution sont affichées sur un même graphique (en raison de cibles ou de résultats multiples provenant des itérations de l'analyse de sensibilité), la valeur par défaut indique la distribution associée à la première itération ou à la première cible.

Lignes de référence (continues) : Vous pouvez demander l'ajout de plusieurs lignes de référence verticales aux fonctions de densité de probabilité et aux fonctions de distribution cumulée pour les cibles continues.

- **Sigmas :** Des lignes de référence peuvent être ajoutées à plus et moins un nombre d'écarts types de la moyenne d'une cible.

- **Percentiles :** Des lignes de références peuvent être ajoutées à une ou deux valeurs de percentiles de la distribution d'une cible, en entrant des valeurs dans les zones de texte Haut et Bas. Par exemple, une valeur de 95 dans la zone de texte Haut représente le 95e percentile, qui est la valeur en dessous de laquelle se trouvent 95 % des observations. De même, une valeur de 5 dans la zone de texte Bas représente le 5e percentile, qui est la valeur en dessous de laquelle se trouvent 5 % des observations.
- **Lignes de référence personnalisées :** Des lignes de références peuvent être ajoutées à des valeurs spécifiques de la cible.

Remarque : Lorsque plusieurs fonctions de distribution sont affichées sur un même graphique (en raison de cibles ou de résultats multiples provenant des itérations de l'analyse de sensibilité), les lignes de référence sont appliquées à la distribution de la première itération ou de la première cible. Vous pouvez ajouter des lignes de référence à d'autres distributions depuis la boîte de dialogue Options de graphique, accessible à partir du graphique FDP ou FDC.

Superposer les résultats provenant de cibles continues distinctes : Dans le cas de cibles continues multiples, cette option spécifie si les fonctions de distribution de ces cibles s'affichent sur une seule représentation graphique, superposant le graphique des fonctions de densité de probabilité et celui des fonctions de distribution cumulée. Si cette option n'est pas sélectionnée, les résultats de chaque cible s'affichent sur un graphique distinct.

Valeurs de catégorie à rapporter : Pour les modèles PMML comportant des cibles catégorielles, le résultat du modèle est un ensemble de probabilités prédites, une pour chaque catégorie, que la valeur de la cible tombe dans chaque catégorie. La catégorie présentant la probabilité la plus élevée est considérée comme la catégorie prédite et est utilisée pour générer le graphique à barres décrit dans le paramètre **Fonction de densité de probabilité** sus-mentionné. Sélectionnez **Catégorie prédite** pour générer le graphique à barres. Sélectionnez **Probabilités prédites** pour générer les histogrammes de la distribution des probabilités prédites de chaque catégorie de la cible.

Regroupement pour l'analyse de sensibilité : Les simulations qui comprennent une analyse de sensibilité génèrent un ensemble indépendant de valeurs cible prédites pour chaque itération définie par l'analyse (une itération pour chaque valeur d'entrée soumise à une variation). Lorsque des itérations sont présentes, le graphique à barres de la catégorie prédite d'une cible catégorielle est représenté sous forme de graphique à barres en cluster qui comprend les résultats de toutes les itérations. Vous pouvez choisir de regrouper les catégories ou les itérations.

Sortie

Graphiques tornado : Les graphiques Tornado sont des graphiques à barres qui représentent les relations entre des cibles et des entrées simulées à l'aide de plusieurs mesures.

- **Corrélation de cible avec entrée :** Cette option crée un graphique tornado des coefficients de corrélation existant entre une cible donnée et chacune de ses entrées simulées. Ce type de graphique tornado ne prend pas en charge les cibles avec un niveau de mesure nominal ou ordinal ni les entrées simulées avec une distribution catégorielle.
- **Contribution à la variance :** Cette option crée un graphique tornado qui affiche la contribution à la variance d'une cible provenant de chacune de ses entrées simulées. Cela vous permet d'évaluer le degré de contribution de chaque entrée à l'incertitude globale de la cible. Ce type de graphique tornado ne prend pas en charge les cibles avec des niveaux de mesure ordinaux ou nominaux ni les entrées simulées avec l'une des distributions suivantes : catégorielle, Bernoulli, binomiale, Poisson ou binomiale négative.
- **Sensibilité de la cible à modifier :** Cette option crée un graphique tornado qui représente les effets sur la cible de la modification de chaque entrée simulée, en ajoutant ou en retirant un nombre spécifique d'écarts-type de la distribution associée à l'entrée. Ce type de graphique tornado ne prend pas en charge les cibles avec des niveaux de mesure ordinaux ou nominaux ni les entrées simulées avec l'une des distributions suivantes : catégorielle, Bernoulli, binomiale, Poisson ou binomiale négative.

Boîtes à moustaches des distributions cible : Les boîtes à moustaches sont disponibles pour les cibles continues. Sélectionnez **Superposer les résultats provenant de cibles distinctes** si votre modèle prédictif contient plusieurs cibles continues et que vous souhaitez afficher les boîtes à moustaches de toutes les cibles sur un même graphique.

Nuages de points comparant les cibles et les entrées : Les nuages de points comparant les cibles et les entrées simulées sont disponibles à la fois pour les cibles continues et les cibles catégorielles, et comprennent les dispersions des cibles ayant à la fois des entrées continues et catégorielles associées. Les nuages de points utilisant une cible ou une entrée catégorielle sont affichées sous la forme d'une carte thermique.

Créer un tableau des valeurs de percentiles : Pour les cibles continues, vous pouvez obtenir un tableau des percentiles des distributions cible spécifiés. Les quartiles (25e, 50e et 75e percentiles) divisent les observations en quatre groupes de taille égale. Si vous souhaitez un nombre égal de groupes différent de quatre, sélectionnez **Intervalles** et indiquez le nombre souhaité. Sélectionnez **Percentiles personnalisés** pour indiquer des percentiles personnalisés, par ex. le 99e percentile.

Statistiques descriptives des distributions cible : Cette option crée des tableaux de statistiques descriptives pour les cibles continues et catégorielles, ainsi que pour les entrées continues. Pour les cibles continues, le tableau comprend la moyenne, l'écart type, la médiane, les valeurs minimale et maximale, l'intervalle de confiance de la moyenne au niveau spécifié et les 5e et 95e percentiles de la distribution cible. Pour les cibles catégorielles, le tableau comprend le pourcentage d'observations qui se situent dans chaque catégorie de la cible. Pour les cibles catégorielles des modèles PMML, le tableau comprend également la probabilité moyenne de chaque catégorie de la cible. Pour les entrées continues, le tableau comprend la moyenne, l'écart type et les valeurs minimale et maximale.

Corrélations et tableau de contingence pour les entrées : Cette option affiche un tableau des coefficients de corrélation entre les entrées simulées. Lorsque des entrées comportant une distribution catégorielle sont générées à partir d'un tableau de contingence, le tableau de contingence des données générées pour ces entrées est également affiché.

Entrées simulées à inclure dans la sortie : Par défaut, toutes les entrées simulées sont incluses dans les sorties. Vous pouvez exclure certaines entrées simulées des sorties. Dans ce cas, elles seront exclues des graphiques tornado, des nuages de points et des sorties sous forme de tableau.

Limiter les plages des cibles continues : Vous pouvez déterminer la plage des valeurs vides d'une ou de plusieurs cibles continues. Les valeurs non comprises dans cette plage sont exclues de toutes les sorties et analyses associées aux cibles. Pour définir une limite inférieure, sélectionnez **Inférieur** dans la colonne Limite et saisissez une valeur dans la colonne Minimum. Pour définir une limité supérieure, sélectionnez **Supérieur** dans la colonne Limite et saisissez une valeur dans la colonne Maximum. Si vous souhaitez définir les deux limites, sélectionnez **Les deux** dans la colonne Limite et saisissez une valeur dans les colonnes Minimum et Maximum.

Formats d'affichage : Vous pouvez définir le format utilisé pour l'affichage des valeurs de cibles et d'entrées (aussi bien les entrées simulées que les entrées fixes).

Enregistrement

Enregistrer le plan de cette simulation : Vous pouvez enregistrer les spécifications de votre simulation dans un fichier de plan de simulation. Les fichiers de plan de simulation possèdent l'extension *.splan*. Vous pouvez rouvrir le plan dans le Générateur de simulation, éventuellement lui apporter des modifications et y exécuter la simulation. Il est possible de partager le plan de simulation avec d'autres utilisateurs, qui peuvent à leur tour exécuter la simulation dans la boîte de dialogue Exécuter la simulation. Les plans de simulation contiennent toutes les spécifications sauf les suivantes : paramètres des fonctions de densité, paramètres de sortie des graphiques et tableaux, paramètres d'options avancées pour l'ajustement, la distribution empirique et la valeur de départ aléatoire.

Enregistrer les données simulées en tant que nouveau fichier de données : Vous pouvez enregistrer les entrées simulées, les entrées fixes et les valeurs cible prédites dans un fichier de données SPSS Statistics, un nouveau jeu de données de la session en cours ou un fichier de données Excel. Chaque observation (ou ligne) du fichier de données contient les valeurs prédites des cibles ainsi que les entrées simulées et les entrées fixes qui ont généré les valeurs cible. Lorsque l'analyse de sensibilité est spécifiée, chaque itération génère un ensemble d'observations continues libellées par le numéro d'itération.

Boîte de dialogue Exécuter la simulation

La boîte de dialogue Exécuter la simulation est destinée aux utilisateurs disposant d'un plan de simulation et qui souhaitent principalement l'exécuter. Elle offre les fonctions nécessaires à l'exécution de la simulation dans différentes conditions. Elle vous permet de réaliser les tâches générales suivantes :

- Configurer ou modifier l'analyse de sensibilité pour vérifier l'effet de la variation de la valeur d'une entrée fixe ou d'un paramètre de distribution d'une entrée simulée.
- Réajuster les distributions de probabilité des entrées incertaines (et les corrélations existant entre ces entrées) aux nouvelles données.
- Modifier la distribution d'une entrée simulée.
- Personnaliser la sortie.
- Exécuter la simulation.

Onglet Simulation

L'onglet Simulation vous permet de spécifier l'analyse de sensibilité, de réajuster les distributions de probabilité des entrées simulées et des corrélations entre elles aux nouvelles données, et de modifier la distribution de probabilité associée à une entrée simulée particulière.

La grille Entrées simulées contient une entrée pour chaque zone d'entrée définie dans le plan de simulation. Ces données comprennent le nom de l'entrée et le type de distribution de probabilité associée à l'entrée, ainsi qu'un tracé de la courbe de distribution associée. Chaque entrée se voit également affecter une icône de statut (un cercle de couleur avec une coche), utile lors du réajustement des distributions aux nouvelles données. En outre, les entrées peuvent présenter une icône en forme de verrou qui indique que l'entrée est verrouillée et ne peut être modifiée ou réajustée aux nouvelles données dans la boîte de dialogue Exécuter la simulation. Pour modifier une entrée verrouillée, vous devez ouvrir le plan de simulation dans le Générateur de simulation.

Chaque entrée peut être soit simulée soit fixe. Les entrées simulées sont celles dont les valeurs sont incertaines et seront générées à partir d'une distribution de probabilité spécifique. Les entrées fixes sont celles dont les valeurs sont connues. Elles restent constantes pour chaque observation générée dans la simulation. Pour utiliser une entrée particulière, sélectionnez le jeu de données correspondant dans la grille Entrées simulées.

Spécification de l'analyse de sensibilité

L'analyse de sensibilité permet d'étudier l'effet des changements systématiques d'une entrée fixe ou d'un paramètre de distribution pour une entrée simulée en générant un ensemble indépendant de cas simulés - en fait, une simulation séparée - pour chaque valeur spécifiée. Pour spécifier l'analyse de sensibilité, sélectionnez une entrée simulée ou une entrée fixe et cliquez sur **Analyse de sensibilité**. L'analyse de sensibilité est limitée à une seule entrée fixe ou un seul paramètre de distribution associé à une entrée simulée. Pour plus d'informations, voir la rubrique «Analyse de sensibilité », à la page 308 .

Réajustement des distributions aux nouvelles données

Pour réajuster automatiquement les distributions de probabilité des entrées simulées (et les corrélations existant entre ces entrées) aux données du jeu de données actif :

1. Vérifiez que chaque entrée du modèle est mise en correspondance avec la zone appropriée dans le jeu de données actif. Chaque entrée simulée est ajustée au champ du jeu de données actif spécifié dans la liste déroulante **Champ** associée à cette entrée. Il est facile d'identifier les entrées non mises en correspondance, il suffit de rechercher les entrées dont l'icône de statut représente une coche et un point d'interrogation, tel qu'illustré ci-dessous.

2. Au besoin, modifiez les champs de mise en correspondance en sélectionnant l'option **Ajuster à un champ dans le jeu de données**, puis en choisissant la zone dans la liste.

3. Cliquez sur **Tout ajuster**.

Pour chaque entrée ajustée, la distribution la mieux adaptée s'affiche en même temps qu'un tracé de la distribution superposé à un histogramme (graphique à barres) représentant les données historiques de l'entrée. S'il est impossible de trouver un ajustement acceptable, la distribution empirique est alors utilisée. Pour les entrées ajustées par distribution empirique, seul un histogramme des données historiques s'affiche car la distribution empirique est en fait représentée par cet histogramme.

Remarque : Pour obtenir la liste complète des icônes de statut, voit «Champs simulés», à la page 304.

Modification des distributions de probabilité

Vous pouvez modifier la distribution de probabilité d'une entrée simulée et éventuellement changer une entrée simulée en entrée fixe, et vice-versa.

1. Sélectionnez l'entrée et choisissez **Définir manuellement la distribution**.
2. Sélectionnez le type de distribution et spécifiez les paramètres correspondants. Pour changer une entrée simulée en entrée fixe, sélectionnez Fixe dans la liste déroulante **Type**.

Une fois les paramètres d'une distribution définis, le tracé représentant la distribution (affiché dans le jeu de données de l'entrée dans la grille) sera mis à jour afin de refléter vos modifications. Pour plus d'informations sur la spécification manuelle des distributions de probabilité, voir «Champs simulés», à la page 304.

Inclure les valeurs manquantes de l'utilisateur pour les entrées catégorielles lors de l'ajustement : Option indiquant si les valeurs manquantes de l'utilisateur pour les entrées comportant une distribution catégorielle sont traitées comme étant valides lorsque vous réajustez les données dans le jeu de données actif. Les valeurs système manquantes et les valeurs manquantes des utilisateurs pour tous les autres types d'entrée sont traitées comme étant non valides. Toutes les entrées doivent avoir des valeurs valides pour qu'une observation puisse être incluse dans l'ajustement de la distribution et le calcul des corrélations.

Onglet Sortie

L'onglet Sortie vous permet de personnaliser les sorties générées par la simulation.

Fonctions de densité : Les fonctions de densité sont les principaux moyens permettant de sonder l'ensemble de résultats généré par votre simulation.

- **Fonction de densité de probabilité :** La fonction de densité de probabilité affiche la distribution des valeurs cible, vous permettant de déterminer la probabilité que la cible se situe dans une zone donnée. Pour les cibles dont le résultat est fixe, par ex. "niveau de service faible", "niveau de service correct", "bon niveau de service" et "excellent niveau de service", un graphique à barres est généré, qui affiche le pourcentage d'observations situées dans chaque catégorie de la cible.
- **Fonction de distribution cumulée :** La fonction de distribution cumulée affiche la probabilité qu'une valeur de la cible soit inférieure ou égale à une valeur donnée.

Graphiques tornado : Les graphiques Tornado sont des graphiques à barres qui représentent les relations entre des cibles et des entrées simulées à l'aide de plusieurs mesures.

- **Corrélation de cible avec entrée :** Cette option crée un graphique tornado des coefficients de corrélation existant entre une cible donnée et chacune de ses entrées simulées.
- **Contribution à la variance :** Cette option crée un graphique tornado qui affiche la contribution à la variance d'une cible provenant de chacune de ses entrées simulées. Cela vous permet d'évaluer le degré de contribution de chaque entrée à l'incertitude globale de la cible.
- **Sensibilité de la cible à modifier :** Cette option crée un graphique tornado qui représente les effets sur la cible de la modification de chaque entrée simulée, en ajoutant ou en retirant un écart type de la distribution associée à l'entrée.

Nuages de points comparant les cibles et les entrées : Cette option génère des nuages de points des cibles par rapport aux entrées simulées.

Boîtes à moustaches des distributions cible : Cette option génère des boîtes à moustaches des distributions cible.

Tableau de quartiles : Cette option génère un tableau des quartiles des distributions cible. Les quartiles d'une distribution sont les 25e, 50e et 75e percentiles de cette distribution et ils divisent les observations en quatre groupes de taille égale.

Corrélations et tableau de contingence pour les entrées : Cette option affiche un tableau des coefficients de corrélation entre les entrées simulées. Une tableau de contingence des associations entre les entrées comportant une distribution catégorielle est affiché lorsque le plan de simulation indique la génération des données catégorielles à partir d'un tableau de contingence.

Superposer les résultats provenant de cibles distinctes : Si le modèle prédictif que vous simulez contient plusieurs cibles, vous pouvez choisir d'afficher ou non les résultats des différentes cibles sur un même graphique. Ce paramètre s'applique aux graphiques des fonctions de densité de probabilité, des fonctions de distribution cumulée et des boîtes à moustaches. Par exemple, si vous sélectionnez cette option, les fonctions de densité de probabilité de toutes les cibles s'afficheront sur un même graphique.

Enregistrer le plan de cette simulation : Vous pouvez enregistrer les modifications apportées à votre simulation dans un fichier de plan de simulation. Les fichiers de plan de simulation possèdent l'extension *.splan*. Vous pouvez rouvrir le plan dans le Générateur de simulation ou dans la boîte de dialogue Exécuter la simulation. Les plans de simulation contiennent toutes les spécifications sauf les paramètres de sortie.

Enregistrer les données simulées en tant que nouveau fichier de données : Vous pouvez enregistrer les entrées simulées, les entrées fixes et les valeurs cible prédites dans un fichier de données SPSS Statistics, un nouveau jeu de données de la session en cours ou un fichier de données Excel. Chaque observation (ou ligne) du fichier de données contient les valeurs prédites des cibles ainsi que les entrées simulées et les entrées fixes qui ont généré les valeurs cible. Lorsque l'analyse de sensibilité est spécifiée, chaque itération génère un ensemble d'observations continues libellées par le numéro d'itération.

Si vous avez besoin de personnaliser davantage les sorties, exécutez plutôt votre simulation dans le Générateur de simulation. Pour plus d'informations, voir «Exécution d'une simulation à partir d'un plan de simulation», à la page 301.

Utilisation de la sortie graphique créée par la simulation

Plusieurs graphiques générés par la simulation offrent des fonctions interactives vous permettant de personnaliser leur affichage. Ces fonctions interactives sont disponibles en activant (par double-clic) l'objet de graphique dans le visualiseur de sortie. Tous les graphiques de simulation sont des visualisations graphiques.

Graphiques de fonction de densité de probabilité pour les cibles continues : Ce type de graphique présente deux lignes de référence verticales pouvant être déplacées par glissement, qui divisent le graphique en zones distinctes. Le tableau en dessous du graphique affiche la probabilité que la cible se trouve dans chacune des zones. Si plusieurs fonctions de densité sont affichées sur le même graphique, le tableau comporte une ligne distincte pour les probabilités associées à chaque fonction de densité. Chacune des lignes de référence présente un curseur (triangle inversé) qui vous permet de déplacer la ligne. D'autres fonctions supplémentaires sont disponibles en cliquant sur le bouton **Options de graphique** situé sur le graphique. Vous pouvez notamment définir explicitement les positions des curseurs, ajouter des lignes de référence fixes et modifier l'affichage du graphique, pour passer par exemple d'une courbe continue à un histogramme ou vice-versa. Pour plus d'informations, voir «Options de graphique », à la page 316.

Graphiques de fonction de distribution cumulée pour les cibles continues : Ce type de graphique présente les mêmes deux lignes de référence verticales et le même tableau associé décrits ci-dessus pour les graphiques de fonction de densité de probabilité. Il offre également un accès à la boîte de dialogue Options de graphique qui vous permet de définir explicitement les positions des curseurs, d'ajouter des lignes de référence fixes et de spécifier si oui ou non la fonction de distribution cumulée est affichée en tant que fonction croissante (par défaut) ou fonction décroissante. Pour plus d'informations, reportez-vous à la rubrique «Options de graphique », à la page 316.

Graphiques à barres pour les cibles catégorielles avec itérations d'analyse de sensibilité : Pour les cibles catégorielles avec itérations d'analyse de sensibilité, les résultats de la catégorie cible prédite sont affichés sous la forme d'un graphique à barres en cluster qui comprend les résultats de toutes les itérations. Le graphique comprend une liste déroulante qui vous permet de regrouper par catégorie ou par itération. Pour les modèles de cluster en deux étapes et les modèles de cluster en nuées dynamiques, il est possible de regrouper par cluster ou par itération.

Boîtes à moustaches pour les cibles multiples avec itérations d'analyse de sensibilité : S'applique aux modèles prédictifs avec plusieurs cibles continues et des itérations d'analyse de sensibilité. Permet d'afficher des boîtes à moustaches pour toutes les cibles sur un même graphique unique, sous la forme d'une boîte à moustache juxtaposée. Le graphique comprend une liste déroulante qui vous permet de regrouper par cible ou par itération.

Options de graphique

La boîte de dialogue Options de graphique vous permet de personnaliser l'affichage des graphiques de fonctions de densité de probabilité et de fonctions de distribution cumulée actifs, générés par une simulation.

Vue. La liste déroulante **Affichage** s'applique uniquement au graphique de fonction de densité de probabilité. Elle vous permet de faire basculer l'affichage d'une courbe continue à un histogramme. Cette fonction n'est pas disponible lorsque plusieurs fonctions de densité sont affichées sur le même graphique. Dans un tel cas, les fonctions de densité peuvent être affichées uniquement sous la forme de courbes continues.

Ordre : La liste déroulante **Ordre** s'applique uniquement au graphique de fonction de distribution cumulée. Elle indique si la fonction de distribution cumulée s'affiche sous la forme d'une fonction croissante (par défaut) ou d'une fonction décroissante. Si elle s'affiche sous la forme d'une fonction décroissante, la valeur de la fonction à un point donné sur l'axe horizontal est la probabilité que la cible se situe à la droite de ce point.

Positions des curseurs : Vous pouvez définir explicitement les positions des lignes de référence ajustables en entrant des valeurs dans les zones de texte Supérieur et Inférieur. Vous pouvez supprimer la ligne de gauche en sélectionnant **Infini -**, ce qui a pour effet de définir la position du curseur sur l'infini négatif, et vous pouvez supprimer la ligne de droite en sélectionnant **Infini**, ce qui défini sa position sur l'infini.

Lignes de référence : Vous pouvez ajouter différentes lignes de référence verticales fixes aux fonctions de densité de probabilité et aux fonctions de distribution cumulée. Lorsque plusieurs fonctions sont affichées sur un même graphique (en raison de cibles ou de résultats multiples provenant des itérations de l'analyse de sensibilité), vous pouvez indiquer les fonctions auxquelles les lignes sont appliquées.

- **Sigmas :** Des lignes de référence peuvent être ajoutées à plus et moins un nombre d'écarts types de la moyenne d'une cible.
- **Percentiles :** Des lignes de références peuvent être ajoutées à une ou deux valeurs de percentiles de la distribution d'une cible, en entrant des valeurs dans les zones de texte Haut et Bas. Par exemple, une valeur de 95 dans la zone de texte Haut représente le 95e percentile, qui est la valeur en dessous de laquelle se trouvent 95 % des observations. De même, une valeur de 5 dans la zone de texte Bas représente le 5e percentile, qui est la valeur en dessous de laquelle se trouvent 5 % des observations.
- **Positions personnalisées :** Des lignes de référence peuvent être ajoutées à des valeurs spécifiques le long de l'axe horizontal.

Lignes de référence des libellés : Cette option permet de contrôler si les libellés sont appliqués aux lignes de référence sélectionnées.

Les lignes de référence peuvent être supprimées en décochant l'option associée dans la boîte de dialogue Options de graphique avant de cliquer sur **Poursuivre**.

Modélisation géospatiale

Les techniques de modélisation géospatiales ont été conçues pour découvrir les motifs dans les données qui incluent un composant géospatial (carte). L'assistant de modélisation géospatiale fournit des méthodes d'analyse des données géospatiales avec et sans composant temporel.

Rechercher des associations en fonction de l'événement et des données géospatiales (Règles d'association géospatiales)
A l'aide des règles d'association géospatiales, vous pouvez rechercher des motifs dans les données en fonction des propriétés spatiales et non spatiales. Par exemple, vous pouvez identifier des motifs dans les données de crime par le biais de l'emplacement et des attributs démographiques. A partir de ces motifs, vous pouvez ensuite générer des règles qui établissent des prévisions quant à l'endroit supposé de certains crimes.

Effectuer des prévisions à l'aide des séries temporelles et des données géospatiales (prévision spatio-temporelle)
Une prévision spatio-temporelle utilise des données qui contiennent des données d'emplacement, des champs d'entrée pour la prévision (prédicteurs), un ou plusieurs champs Heure et un champ cible. Chaque emplacement comporte de nombreuses lignes dans les données qui représentent les valeurs de chaque prédicteur et la cible dans chaque intervalle de temps.

Utilisation de l'assistant de modélisation géospatiale

1. A partir des menus, sélectionnez :

   Analyser > **Modélisation spatiale et temporelle** > **Modélisation spatiale**

2. Suivez les instructions indiquées dans l'assistant.

Exemples

Des exemples détaillés sont disponibles dans le système d'aide.

- Règles d'association géospatiale : **Aide** > **Rubriques** > **Etudes de cas pas à pas** > **Statistics Base** > **Règles d'associations spatiales**
- Prévision spatio-temporelle : **Aide** > **Rubriques** > **Etudes de cas pas à pas** > **Statistics Base** > **Prévision spatio-temporelle**

Sélection de cartes

La modélisation géospatiale peut utiliser une ou plusieurs sources de données de carte. Ces dernières contiennent des informations qui définissent les zones géographiques et d'autres fonctions géographiques, telles que des routes ou des rivières. Un grand nombre de sources de cartes contiennent également des données démographiques ou descriptives, ainsi que des données d'événement, par exemple un rapport sur un crime ou un taux de chômage. Vous pouvez vous servir d'un fichier de spécification de carte défini précédemment ou alors définir des spécifications de carte et les enregistrer pour une utilisation ultérieure.

Charger une spécification de carte
Charge un fichier de spécification de carte (.mplan) précédemment défini. Les sources de données de carte que vous définissez ici peuvent être enregistrées dans un fichier de spécification de carte. Dans le cadre d'une prévision spatio-temporelle, si vous sélectionnez un fichier de spécification de carte qui identifie plusieurs cartes, le système vous invite à sélectionner celle de votre choix dans le fichier.

Ajouter un fichier de carte
Ajoutez un fichier de forme (.shp) ESRI ou une archive .zip qui contient un fichier de forme ESRI.

- Un fichier .dbf correspondant doit se trouver au même emplacement que le fichier .shp et tous deux doivent posséder le même nom racine.
- Si le fichier est une archive .zip, les fichiers .shp et .dbf doivent posséder le même nom racine que l'archive.

- S'il n'y a pas de fichier de projection (.prj) correspondant, le système vous invite à sélectionner un système de projection.

Relation
Pour les règles d'association géospatiales, cette colonne indique la manière dont les événements sont reliés aux fonctions dans la carte. Ce paramètre n'est pas disponible pour la prévision spatio-temporelle.

Déplacer vers le haut, Déplacer vers le bas
L'ordre des couches des éléments de la carte est déterminé par l'ordre dans lequel ils apparaissent dans la liste. La première carte de la liste correspond à la couche du bas.

Sélection d'une carte

Dans le cadre d'une prévision spatio-temporelle, si vous sélectionnez un fichier de spécification de carte qui identifie plusieurs cartes, le système vous invite à sélectionner celle de votre choix dans le fichier. La prévision temporelle spatiale ne prend pas en charge les cartes multiples.

Relation géospatiale

Pour les règles d'association géospatiales, la boîte de dialogue Relation géospatiale définit la manière dont les événements sont reliés aux fonctions dans la carte.

- Ce paramètre ne s'applique qu'aux règles d'association géospatiales.
- Il n'affecte que les sources de données associées à des cartes qui sont spécifiées en tant que données de contexte lors de l'étape de sélection des sources de données.

Relation

Fermer
L'événement se produit près d'une zone ou d'un point donné sur la carte.

Dans
L'événement se produit dans une zone donnée sur la carte.

Contient
La zone d'événement contient un objet de contexte de carte.

Intersection
Emplacement correspondant à un point d'intersection de lignes ou de régions provenant de différentes cartes.

Croix
Lorsque plusieurs cartes co-existent, points de croisement des différentes lignes (routes, rivières, chemins de fer).

Au nord de, Au sud de, A l'est de, A l'ouest de
L'événement se produit dans la zone nord, sud, est ou ouest d'un point donné sur la carte.

Définition d'un système de coordonnées

Si un fichier de projection (.prj) n'est pas associé à la carte ou si vous définissez deux champs d'une source de données comme ensemble de coordonnées, vous devez définir le système de coordonnées.

Système géographique par défaut (longitude et latitude)
Le système de coordonnées est celui de la longitude et latitude.

Système cartésien simple (X et Y)
Le système de coordonnées est celui des coordonnées X et Y simple.

Utiliser un ID connu (WKID)
"ID connu" pour les projections standard.

Utiliser un nom de système de coordonnées
Le système de coordonnées est basé sur la projection citée. Le nom est mentionné entre parenthèses.

Définition de projection

Si le système de projection ne peut être déterminé à partir des informations fournies avec la carte, vous devez le spécifier. La cause la plus courante de cette situation est le fait qu'il n'y a pas de fichier de projection (.prj) associé à la carte ou que le fichier de projection ne peut être utilisé.

- **Une ville, une région ou un pays (projection de Mercator)**
- **Un grand pays, plusieurs pays ou des continents (projection de Winkel Tripel)**
- **Une zone proche de l'équateur (projection de Mercator)**
- **Une zone proche de l'un des pôles (projection stéréographique)**

La projection de Mercator est une projection couramment utilisée dans beaucoup de cartes. Cette projection traite le globe comme un cylindre qui est roulé sur une surface plane. La projection de Mercator déforme la taille et la forme des grands objets. Cette distorsion augmente au fur et à mesure que vous vous éloignez de l'équateur et que vous vous rapprochez des pôles. Les projections de Winkel Tripel et stéréographique font des ajustements pour compenser le fait qu'une carte représente une portion de sphère en trois dimensions affichée en deux dimensions.

Projection et système de coordonnées

Si vous sélectionnez plusieurs cartes et que celles-ci possèdent différents systèmes de coordonnées et projections, vous devez sélectionner la carte avec le système de projection que vous souhaitez utiliser. Ce système de projection sera ensuite utilisé pour toutes les cartes lorsqu'elles sont associées dans la sortie.

Sources de données

Une source de données peut être un fichier dBase fourni avec le fichier de forme, un fichier de données IBM SPSS Statistiques ou un jeu de données ouvert dans la session en cours.

Données de contexte. Les données de contexte identifient les fonctions sur la carte. Elles peuvent aussi contenir des champs pouvant être utilisés comme entrées pour le modèle. Pour utiliser un fichier de contexte dBase (.dbf) qui est associé à un fichier de forme carte (.shp), le fichier de contexte dBase doit se trouver au même emplacement que le fichier de forme et posséder le même nom racine. Ainsi, si le fichier de forme s'appelle `geodata.shp`, le fichier dBase doit s'appeler `geodata.dbf`

Données d'événement. Les données d'événement comportent des informations sur les événements qui se produisent, notamment des crimes ou des accidents. Cette option est disponible uniquement pour les règles d'association géospatiales.

Densité de point : Intervalle de temps et données de coordonnées pour les estimations de la densité par noyau. Cette option n'est disponible que pour la prévision spatio-temporelle.

Ajouter : Ouvre une boîte de dialogue dans laquelle ajouter des sources de données. Une source de données peut être un fichier dBase fourni avec le fichier de forme, un fichier de données IBM SPSS Statistiques ou un jeu de données ouvert dans la session en cours.

Associer : Ouvre une boîte de dialogue permettant de spécifier les identificateurs (coordonnées ou clés) utilisés pour associer les données aux cartes. Chaque source de données doit contenir un ou plusieurs identificateurs permettant d'associer les données à la carte. Les fichiers dBase qui sont livrés avec le fichier de forme contiennent généralement un champ qui est automatiquement utilisé comme identificateur par défaut. Pour les autres sources de données, vous devez spécifier les champs qui sont utilisés comme identificateurs.

Valider la clé : Ouvre une boîte de dialogue permettant de valider la correspondance des clés entre la carte et la source de données.

Règles d'association géospatiales

- Au moins une source de données doit être une source de données d'événement.
- Toutes les sources de données d'événement doivent utiliser les mêmes identificateurs d'association de règle : coordonnées ou valeurs de clés.

- Si les sources de données d'événement sont associées aux cartes avec des valeurs clés, toutes les sources d'événement doivent utiliser le même type de fonction de carte (par exemple, polygones, points, lignes).

Prévision temporelle spatiale

- Une source de données contextuelle doit exister.
- S'il existe une seule source de données (un fichier de données sans carte associée), elle doit inclure les valeurs de coordonnées.
- Si vous disposez de deux sources de données, une source de données doit correspondre aux données contextuelles et l'autre, aux données de densité des points.
- Vous ne pouvez inclure plus de deux sources de données.

Ajout d'une source de données

Une source de données peut être un fichier dBase fourni avec les fichiers de forme et de contexte, un fichier de données IBM SPSS Statistiques ou un jeu de données ouvert dans la session en cours.

Vous pouvez ajouter plusieurs fois une même source de données si sous souhaitez utiliser une association spatiale différente avec chacune d'entre elle.

Association de carte et données

Chaque source de données doit contenir un ou plusieurs identificateurs permettant d'associer les données à la carte.

Coordonnées
: La source de données comporte des champs qui représentent des données cartésiennes. Sélectionnez ceux qui représentent les coordonnées X et Y. Pour les règles d'association géospatiales, vous pouvez également disposer de coordonnées Z.

Valeurs clés
: Les valeurs de clés figurant dans les champs de la source de données correspondent aux clés de cartes sélectionnées. Ainsi, une carte des régions peut avoir un identificateur de nom (clé de carte) fournissant un libellé à chaque région. L'identificateur correspond à un champ de données qui contient également le nom des régions (clé de données). Les champs sont mis en corrélation afin de mapper les clés en fonction de l'ordre dans lequel elles apparaissent dans les deux listes.

Valider les clés

La boîte de dialogue Valider les clés fournit un récapitulatif de la correspondance des enregistrements entre la carte et la source de données, basé sur les clés d'identificateur sélectionnées. Si des valeurs de clés de données ne correspondent pas, vous pouvez les faire correspondre manuellement aux valeurs de clés de cartes.

Règles d'association géospatiales

Pour les règles d'association géospatiales, après défini les cartes et les sources de données, les étapes restantes de l'assistant sont les suivantes :

- Si plusieurs sources de données d'événement existent, définissez leurs modalités de fusion.
- Sélectionnez les champs à utiliser en tant que conditions et prévisions dans l'analyse.

Eventuellement, vous pouvez aussi :

- Sélectionner des options de sortie différentes.
- Enregistrer un fichier de modèle d'évaluation.
- Créer de nouveaux champs pour les prévisions et les règles des sources de données utilisées dans le modèle.

- Personnaliser les paramètres de génération de règles d'association.
- Personnaliser les paramètres de regroupement et d'agrégation.

Définition de champs de données d'événement

Pour les règles d'association géospatiales, s'il existe plusieurs sources de données d'événement, celles-ci sont fusionnées.

- Par défaut, seuls les champs communs à toutes les sources de données d'événement sont inclus.
- Vous pouvez afficher la liste des champs communs, des champs spécifiques à une source de données ou des champs de toutes les sources de données, puis sélectionner ceux que vous voulez inclure.
- Pour les champs communs, les options **Type** et **Mesure** doivent être identiques pour toutes les sources de données. En cas de conflit, vous pouvez préciser le type et le niveau de mesure de chaque champ commun.

Sélection des champs

La liste des champs disponibles inclut les champs provenant des sources de données d'événement et ceux provenant des sources de données contextuelles.

- Vous pouvez contrôler la liste des champs affichés en sélectionnant une source de données dans la liste **Sources de données**.
- Vous devez sélectionner au moins deux champs, dont l'un doit être une condition et l'autre une prévision. Vous pouvez atteindre cet objectif de plusieurs manières, notamment en sélectionnant deux champs pour la liste **Les deux (Condition et Prévision)**.
- Les valeurs de prévision des règles d'association reposent sur les valeurs des champs de condition. Par exemple, dans la règle "Si x=1 et y=2, alors z=3", les valeurs x et y sont des conditions et la valeur z est la prévision.

Sortie

Tables de règles
Chaque table de règles affiche les règles et valeurs principales de confiance, support de règle, lift, support de condition et déployabilité. Chaque table est triée en fonction des valeurs du critère sélectionné. Vous pouvez afficher toutes les règles ou le plus grand **Nombre** de règles, en fonction du critère sélectionné.

Nuage de mots pour tri
Liste des règles principales en fonction des valeurs du critère sélectionné. La taille du texte indique l'importance relative de la règle. L'objet de sortie interactive contient les règles principales de confiance, de support de règle, de lift, de support de condition et de déployabilité. Le critère sélectionné détermine la liste de règles qui est affichée par défaut. Vous pouvez choisir un critère différent de façon interactive dans la sortie. L'option **Nombre maximal de règles à afficher** détermine le nombre de règles qui sont affichées dans la sortie.

Carte
Carte et diagramme à barres interactifs des principales règles, en fonction du critère sélectionné. Chaque objet de sortie interactive contient les règles principales de confiance, de support de règle, de lift, de support de condition et de déployabilité. Le critère sélectionné détermine la liste de règles qui est affichée par défaut. Vous pouvez choisir un critère différent de façon interactive dans la sortie. L'option **Nombre maximal de règles à afficher** détermine le nombre de règles qui sont affichées dans la sortie.

Tableaux d'informations sur le modèle

Transformations de champs (zones).
Décrit les transformations qui sont appliquées aux champs utilisés dans l'analyse.

Récapitulatif des enregistrements
Nombre et pourcentage d'enregistrements inclus et exclus.

Statistiques de règles
Statistiques récapitulatives sur le support de condition, la confiance, le support de règle, le lift et la déployabilité. Ces statistiques incluent la moyenne, les valeurs minimum et maximum et l'écart type.

Eléments les plus fréquents
Eléments qui se produisent le plus fréquemment. Un élément est inclus dans une condition ou une prévision d'une règle. Par exemple, âge < 18 ou genre=féminin.

Champs les plus fréquents
Champs qui figurent le plus fréquemment dans les règles.

Entrées exclues
Champs qui sont exclus de l'analyse et raison de l'exclusion.

Critère des tables de règles, nuages de mots et cartes

Confiance.
Pourcentage des prévisions de règles correctes.

Support de règle
Pourcentage des observations pour lesquelles la règle est true (vraie). Par exemple, si la règle est "Si x=1 et y=2, alors z=3," le support de règle correspond au pourcentage réel d'observations dans les données pour lesquelles x=1, y=2 et z=3.

Lift
Le lift détermine la mesure dans laquelle les prévisions définies par la règle sont plus efficaces que celles définies de manière aléatoire. Il s'agit du rapport entre les prévisions correctes et l'occurrence générale de la prévision. La valeur doit être supérieure à 1. Ainsi, si la prévision se produit dans 20 % des cas et que la confiance de la prévision est de 80 %, la valeur de lift est de 4.

Support de condition
Pourcentage des observations pour lequel la condition de règle existe. Par exemple, si la règle est "Si x=1 et y=2, alors z=3," le support de condition correspond à la proportion d'observations dans les données pour lesquelles x=1 et y=2.

Déployabilité
Pourcentage des prédictions incorrectes quand les conditions sont vérifiées. La déployabilité se calcule ainsi : (confiance à 1) multiplié par support de condition, ou support de condition moins support de règle.

Enregistrer

Enregistrer la carte et les données de contexte en tant que spécification de carte
Enregistrez les spécifications de carte dans un fichier externe (.mplan). Vous pouvez ensuite charger ce fichier dans l'assistant à des fins d'analyse ultérieure. Vous pouvez également l'utiliser avec la commande `SPATIAL ASSOCIATION RULES`.

Copier les fichiers de carte et de données dans la spécification
Les données issues des fichiers de forme carte, des fichiers de données externes et des jeux de données utilisés dans la spécification de carte sont enregistrées dans le fichier de spécification de carte.

Evaluation
Enregistre les meilleures valeurs de règle, valeurs de confiance des règles et valeurs d'ID numérique des règles en tant que nouveaux champs dans la source de données spécifiée.

Sources de données à évaluer
Sources de données dans lesquelles sont créés les nouveaux champs. La source de données est ouverte dans la session en cours, si ce n'est pas déjà fait. Vous devez explicitement enregistrer le fichier modifié pour pouvoir enregistrer les nouveaux champs.

Valeurs cible
Créez de nouveaux champs pour les champs cible sélectionnés (prévision).

- Deux nouveaux champs sont créés pour chaque cible : la valeur prévue et la valeur de confiance.

- Pour les champs cible continus (échelle), la prévision est une chaîne décrivant une plage de valeurs. Une valeur au format "(valeur1, valeur2]" signifie "supérieure à la valeur1 et inférieure ou égale à la valeur2".

Nombre de meilleures règles
Créez de nouveaux champs pour le nombre de meilleures règles spécifié. Trois nouveaux champs sont créés pour chaque règle, à savoir, la valeur de règle, la valeur de confiance et la valeur d'ID numérique de la règle.

Préfixe de nom
Préfixe à utiliser pour les nouveaux noms de champs.

Construction de règle

Les paramètres de construction de règle définissent les critères des règles d'association générées.

Eléments par règle
Nombre de valeurs de champs pouvant être incluses dans les prévisions et conditions de règles. Le nombre total d'éléments ne doit pas dépasser 10. Par exemple, la règle "Si x=1 et y=2, alors z=3" comporte deux éléments de condition et un élément de prévision.

Nombre maximal de prévisions
Nombre maximal de valeurs de champs pouvant se produire dans les prévisions d'une règle.

Nombre maximal de conditions
Nombre maximal de valeurs de champs qui peuvent se produire dans les conditions d'une règle.

Exclure la paire
Empêche les paires de champs indiquées d'être incluses dans la même règle.

Critères de règles

Confiance.
Niveau de confiance minimal qu'une règle doit avoir pour être incluse dans la sortie. La confiance correspond au pourcentage de prévisions correctes.

Support de règle
Support de règle minimal qu'une règle doit avoir pour être incluse dans la sortie. La valeur représente le pourcentage d'observations pour lequel la règle est vérifiée (true) dans les données observées. Par exemple, si la règle est "Si x=1 et y=2, alors z=3," le support de règle correspond au pourcentage réel d'observations dans les données pour lesquelles x=1, y=2 et z=3.

Support de condition
Support de condition minimal qu'une règle doit avoir pour être incluse dans la sortie. La valeur représente le pourcentage d'observations pour lequel la condition existe. Par exemple, si la règle est "Si x=1 et y=2, alors z=3", le support de condition correspond au pourcentage d'observations dans les données pour lesquelles x=1 et y=2.

Lift
Lift minimal qu'une règle doit avoir pour être incluse dans la sortie. Le lift détermine la mesure dans laquelle les prévisions définies par la règle sont plus efficaces que celles définies de manière aléatoire. Il s'agit du rapport entre les prévisions correctes et l'occurrence générale de la prévision. Ainsi, si la prévision se produit dans 20 % des cas et que la confiance de la prévision est de 80 %, la valeur de lift est de 4.

Traiter de manière identique
Identifie les paires de champs pouvant être traitées comme un même champ.

Regroupement et agrégation

- L'agrégation est nécessaire lorsque les enregistrements dans les données sont plus nombreux que les fonctions dans la carte. Par exemple, vous possédez des enregistrements de données pour des pays individuels, mais vous disposez d'une carte des Etats.
- Vous pouvez spécifier la méthode de mesure récapitulative d'agrégation pour les champs ordinaux et continus. Les champs nominaux sont agrégés en fonction de la valeur modale.

Continu
Pour les champs continus (échelle), la mesure récapitulative peut être moyenne, médiane ou somme.

Ordinal
Pour les champs ordinaux, la mesure récapitulative peut être médiane, mode, plus élevée ou plus faible.

Nombre de regroupements
Définit le nombre de regroupements pour les champs continus (échelle). Les champs continus sont toujours rassemblés ou "regroupés" dans des plages de valeurs. Par exemple : inférieur ou égal à 5, supérieur à 5 et inférieur ou égal à 10, ou supérieur à 10.

Agréger la carte
Applique l'agrégation à la fois aux données et aux cartes.

Paramètres personnalisés pour les champs spécifiques
Vous pouvez remplacer la mesure récapitulative par défaut et le nombre de regroupements pour des champs spécifiques.

- Cliquez sur l'icône permettant d'ouvrir la boîte de dialogue **Sélecteur de champs** et sélectionnez un champ à ajouter à la liste.
- Dans la colonne **Agrégation**, sélectionnez une mesure récapitulative.
- Pour les champs continus, cliquez sur le bouton dans la colonne **Regroupements** pour spécifier le nombre de regroupements de votre choix dans le champ de la boîte de dialogue **Regroupements**.

Prévision spatio-temporelle

Pour la prévision spatio-temporelle, après avoir défini les cartes et les sources de données, les étapes restantes de l'assistant sont les suivantes :

- Indiquez le champ cible, les champs Heure et les prédicteurs facultatifs.
- Définissez les intervalles de temps ou les périodes cycliques des champs Heure.

Eventuellement, vous pouvez aussi :

- Sélectionner des options de sortie différentes.
- Personnaliser les paramètres de génération de modèle.
- Personnaliser les paramètres d'agrégation.
- Enregistrer les prévisions dans un jeu de données de la session en cours ou dans un fichier de données au format IBM SPSS Statistiques.

Sélection des champs

La liste des champs disponibles inclut les champs provenant des sources de données sélectionnées. Vous pouvez contrôler la liste des champs affichés en sélectionnant une source de données dans la liste **Sources de données**.

Cible
Le champ cible est obligatoire. La cible est le champ pour lequel les valeurs sont prédites.

- Le champ cible doit être un champ numérique continu (échelle).
- S'il existe deux sources de données, la cible correspond aux estimations de densité du noyau et la "Densité" s'affiche comme nom de cible. Vous ne pouvez pas changer cette sélection.

Prédicteurs
Vous pouvez spécifier un ou plusieurs champs prédicteurs. Ce paramètre est facultatif.

Champs Heure
Vous devez sélectionner un ou plusieurs champs représentant des plages de temps ou sélectionner l'option **Périodes cycliques**.

- S'il existe deux sources de données, vous devez sélectionner les champs Heure de ces deux sources. Ces champs doivent représenter le même intervalle.

- Pour les périodes cycliques, vous devez spécifier les champs qui définissent des cycles périodiques dans le panneau **Intervalles de temps** de l'assistant.

Intervalles de temps

Les options de ce panneau dépendent du choix de l'option **Champs Heure** ou **Période cyclique** effectué à l'étape de sélection des champs.

Champs Heure

Champs Heure sélectionnés : Si vous sélectionnez un ou plusieurs champs Heure au cours de l'étape de sélection des champs, ceux-ci s'affichent dans cette liste.

Intervalle de temps : Sélectionnez l'intervalle de temps approprié dans la liste. En fonction de l'intervalle de temps, vous pouvez également indiquer d'autres paramètres, tels que l'intervalle entre les observations (incrément) et la valeur de début. Cet intervalle de temps s'applique à tous les champs horaires sélectionnés.

- La procédure suppose que toutes les observations (enregistrements) représentent des intervalles réguliers.
- Selon l'intervalle de temps sélectionné, la procédure peut détecter des observations manquantes ou plusieurs observations d'un même intervalle de temps à agréger. Par exemple, si l'intervalle de temps est exprimé en jours et que la date 2014-10-27 est suivie de la date 2014-10-29, il existe une observation manquante pour la date 2014-10-28. Si l'intervalle de temps est exprimé en mois, plusieurs dates du même mois sont agrégées.
- Pour certains intervalles de temps, le paramètre supplémentaire peut définir des ruptures dans les intervalles régulièrement espacés. Par exemple, si l'intervalle de temps est toujours exprimé en jours, mais que seuls les jours de la semaine sont valides, vous pouvez indiquer qu'une semaine comporte cinq jours et qu'elle commence le lundi.
- Si le champ Heure sélectionné ne représente pas un champ de format de date ni d'heure, l'intervalle de temps est automatiquement défini sur **Périodes** et ne peut pas être modifié.

Champs de cycle

Si vous sélectionnez **Période cyclique** à l'étape de sélection des champs, vous devez indiquer les champs qui définissent les périodes cycliques. Une période cyclique identifie une variation cyclique répétitive, telle que le nombre de mois dans une année ou le nombre de jours de la semaine.

- Vous pouvez spécifier jusqu'à trois champs qui définissent les périodes cycliques.
- Le premier champ de cycle représente le niveau le plus haut du cycle. Par exemple, s'il existe une variation par année, trimestre et mois, le champ qui représente l'année est le champ du premier cycle.
- La longueur du cycle des champs de premier et deuxième cycle correspond à la périodicité au niveau suivant. Par exemple, si les champs de cycle sont année, trimestre et mois, la longueur du premier cycle est 4 et celle du deuxième est 3.
- La valeur de début des champs des deuxième et troisième cycles correspond à la première valeur de chacune des périodes cycliques.
- La longueur du cycle et les valeurs de début doivent être des entiers positifs.

Agrégation

- Si vous sélectionnez des **Prédicteurs** à l'étape de sélection des champs, vous pouvez choisir la méthode récapitulative d'agrégation des prédicteurs.
- L'agrégation est nécessaire lorsqu'un intervalle de temps défini comporte plusieurs enregistrements. Par exemple, si l'intervalle de temps est exprimé en mois, plusieurs dates d'un même mois sont agrégées.
- Vous pouvez spécifier la méthode récapitulative d'agrégation pour les champs continus et ordinaux. Les champs nominaux sont agrégés en fonction de la valeur modale.

Continu
Pour les champs continus (échelle), la mesure récapitulative peut être moyenne, médiane ou somme.

Ordinal
Pour les champs ordinaux, la mesure récapitulative peut être médiane, mode, plus élevée ou plus faible.

Paramètres personnalisés pour les champs spécifiques
Vous pouvez remplacer la mesure récapitulative d'agrégation pour des prédicteurs spécifiques.

- Cliquez sur l'icône permettant d'ouvrir la boîte de dialogue **Sélecteur de champs** et sélectionnez un champ à ajouter à la liste.
- Dans la colonne **Agrégation**, sélectionnez une mesure récapitulative.

Sortie

Cartes

Valeurs cible
Carte de valeurs pour le champ cible sélectionné.

Corrélation
Carte des corrélations.

Clusters
Carte mettant en évidence les clusters d'emplacements similaires. Les cartes de clusters sont disponibles uniquement pour les modèles empiriques.

Seuil de similitude d'emplacement
Similitude qui est requise pour créer des clusters. La valeur doit être un nombre supérieur à zéro et inférieur à 1.

Spécifier le nombre maximal de clusters
Nombre maximal de clusters à afficher.

Tableaux et graphiques d'évaluation de modèle

Spécifications de modèles
Récapitulatif des spécifications qui sont utilisées pour exécuter l'analyse, y compris les champs cible, d'entrée et d'emplacement.

Récapitulatif d'informations temporelles
Identifie les champs d'heure et les intervalles de temps utilisés dans le modèle.

Test des effets dans la structure moyenne
La sortie contient la valeur de statistiques de test, les degrés de liberté et le niveau de confiance pour le modèle corrigé et pour chaque effet.

Structure de la moyenne des coefficients du modèle
La sortie inclut la valeur de coefficient, l'erreur standard, la valeur de statistiques de test, le niveau d'importance et les intervalles de confiance pour chaque terme de modèle.

Coefficients autorégressifs
La sortie inclut la valeur de coefficient, l'erreur standard, la valeur de statistiques de test, le niveau d'importance et les intervalles de confiance pour chaque décalage.

Tests de covariance spatiale
Pour les modèles paramétriques de variogramme, affiche les résultats du test de la qualité d'ajustement pour la structure de covariance spatiale. Ces résultats permettent de déterminer s'il convient de modéliser la structure de covariance spatiale de façon paramétrique ou d'utiliser un modèle non paramétrique.

Covariance spatiale paramétrique
Affiche les estimations de covariance spatiale paramétrique pour les modèles paramétriques de variogramme.

Options de modèle

Paramètres du modèle

Inclure automatiquement une constante
Incluez la constante dans le modèle.

Nombre maximal de décalages d'autorégression
Nombre maximal de décalages d'autorégression. Cette valeur doit être un entier compris entre 1 et 5.

Covariance spatiale
Indique la méthode d'estimation de la covariance spatiale.

Paramétrique
La méthode d'estimation est paramétrique. Elle peut être de type **Gaussien**, **Exponentiel** ou **Exponentiel de puissance**. Pour ce dernier type, vous pouvez spécifier la valeur **Puissance**.

Non paramétrique
La méthode d'estimation est non paramétrique.

Enregistrer

Enregistrer la carte et les données de contexte en tant que spécification de carte
Enregistrez les spécifications de carte dans un fichier externe (.mplan). Vous pouvez ensuite charger ce fichier dans l'assistant à des fins d'analyse ultérieure. Vous pouvez également l'utiliser avec la commande `SPATIAL TEMPORAL PREDICTION`.

Copier les fichiers de carte et de données dans la spécification
Les données issues des fichiers de forme carte, des fichiers de données externes et des jeux de données utilisés dans la spécification de carte sont enregistrées dans le fichier de spécification de carte.

Evaluation
Enregistre les prévisions, la variance et les bornes de confiance inférieure et supérieure dans le fichier de données sélectionné.

- Vous pouvez enregistrer les prévisions dans un jeu de données ouvert dans la session en cours ou dans un fichier de données au format IBM SPSS Statistiques.
- Le fichier de données ne peut pas être une source de données utilisée dans le modèle.
- Le fichier de données doit inclure tous les champs d'heure et les prédicteurs utilisés dans le modèle.
- Les valeurs d'heure doivent être supérieures aux valeurs d'heure utilisées dans le modèle.

Avancé

Nombre maximal d'observations avec valeurs manquantes (%)
Pourcentage maximal d'observations comportant des valeurs manquantes.

Niveau de signification
Niveau d'importance permettant de déterminer si un modèle paramétrique de variogramme est approprié. La valeur doit être supérieure à 0 et inférieure à 1. La valeur par défaut est 0,05. Le niveau d'importance est utilisé dans le test de la qualité d'ajustement pour la structure de covariance spatiale. La statistique de qualité d'ajustement permet de déterminer s'il faut utiliser un modèle paramétrique ou non paramétrique.

Facteur d'incertitude (%)
Le facteur d'incertitude est une valeur exprimée en pourcentage qui représente la croissance de l'incertitude pour les prévisions futures. Les bornes supérieure et inférieure de l'incertitude des prévisions augmentent d'après ce pourcentage à chaque étape dans le futur.

Terminer

Dans l'étape finale de l'assistant de modélisation géospatiale, vous pouvez soit exécuter le modèle, soit coller la syntaxe de commande générée dans une fenêtre de syntaxe. Vous pouvez ensuite modifier et enregistrer la syntaxe ainsi générée pour une utilisation ultérieure.

Remarques

Le présent document concerne des produits et des services disponibles dans différents pays. Il peut être mis à disposition par IBM dans d'autres langues. Toutefois, il peut être nécessaire de posséder une copie du produit ou de la version du produit dans cette langue pour pouvoir y accéder.

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. Consultez votre représentant IBM local pour obtenir des informations sur les produits et services actuellement disponibles dans votre région. Toute référence à un produit, logiciel ou service IBM n'implique pas que seul ce produit, logiciel ou service IBM puisse être utilisé. Tout produit, programme ou service fonctionnellement équivalent qui ne porte pas atteinte à un droit de propriété intellectuelle IBM peut être utilisé à la place. Il est de la responsabilité de l'utilisateur d'évaluer et de vérifier lui-même les installations et applications réalisées avec des produits, logiciels ou services non expressément référencés par IBM.

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

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
U.S.A.

Pour toute demande d'informations relatives au jeu de caractères codé sur deux octets, contactez le service de propriété intellectuelle IBM ou envoyez vos questions par courrier à l'adresse suivante :

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

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

Le présent document peut contenir des inexactitudes ou des coquilles. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, modifier les produits et/ou programmes décrits dans ce document.

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

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

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

IBM Director of Licensing
IBM Corporation

North Castle Drive, MD-NC119
Armonk, NY 10504-1785
U.S.A.

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

Le logiciel sous licence décrit dans ce document et tous les éléments sous licence disponibles s'y rapportant sont fournis par IBM conformément aux dispositions de l'ICA (IBM Customer Agreement), des Conditions internationales d'utilisation des logiciels IBM ou de tout autre accord équivalent.

Les données de performances et les exemples de clients sont fournis à titre d'exemple uniquement. Les performances réelles peuvent varier en fonction des configurations et des conditions d'exploitation spécifiques.

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

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

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

LICENCE DE COPYRIGHT :

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

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

© Copyright IBM Corp. 2021. Des parties de ce code sont proviennent d'IBM Corp. Exemples de programmes.

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

Marques

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

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

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

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

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

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

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

Index

A

accord entre observateurs 274
affectation de mémoire
  dans l'analyse de cluster TwoStep 229
Agresti-Caffo
  dans les proportions d'échantillons indépendants 117
Agresti-Coull
  dans les proportions d'échantillons uniques 112
Agresti-Min
  dans les proportions d'échantillons appariés 114
ajustement automatique de la distribution
  dans la simulation 304
ajustement de la distribution
  dans la simulation 304
Alpha de Cronbach
  dans l'analyse de fiabilité 272, 274
alpha-maximisation 225
analyse d'hypothèses
  dans la simulation 308
analyse de cluster
  Analyse de cluster de nuées dynamiques 238
  Analyse de cluster hiérarchique 236
  efficacité 239
Analyse de cluster de nuées dynamiques
  cluster d'affectation 240
  critères de convergence 240
  distances entre les clusters 240
  efficacité 239
  enregistrement des informations sur les clusters 240
  exemples 238
  fonctions supplémentaires de la commande 241
  itérations 240
  méthodes 238
  présentation 238
  Statistiques 238, 240
  valeur manquante 240
Analyse de cluster hiérarchique
  classification d'observations 236
  classification de variables 236
  cluster d'affectation 238
  dendrogrammes 238
  enregistrement des nouvelles variables 238
  exemple 236
  fonctions supplémentaires de la commande 238
  matrices de distance 238
  mesures de distance 237
  mesures de similarité 237
  méthodes de classification 237
  orientation du tracé 238
  plannings des agglomérations 238
  Statistiques 236, 238
  tracés en stalactite 238
  transformation de mesures 237
  Transformation de valeurs 237
Analyse de cluster TwoStep
  enregistrement dans le fichier de travail 231
Analyse de cluster TwoStep *(suite)*
  enregistrement dans le fichier externe 231
  Options 229
  Statistiques 231
Analyse de fiabilité
  coefficient de corrélation intra-classe 274
  corrélations et covariances entre éléments 274
  descriptives 274
  exemple 272
  fonctions supplémentaires de la commande 276
  Kuder-Richardson 20 274
  Statistiques 272, 274
  T2 de Hotelling 274
  tableau ANOVA 274
  Test d'additivité de Tukey 274
Analyse de puissance
  Statistiques 1
analyse de sensibilité
  dans la simulation 308
analyse de séries temporelles
  prévision 208
  prévision d'observations 208
analyse de variance
  Dans ANOVA à 1 facteur 124
  dans la régression linéaire 178
  dans Moyennes 105
  estimation de courbe 207
analyse des réponses multiples
  table de fréquences 264
  tableau croisé 265
  Tableaux croisés des réponses multiples 265
  Tableaux de fréquences des réponses multiples 264
Analyse discriminante
  coefficients de la fonction 221
  critères 221
  définition de plages 220
  Distance de Mahalanobis 221
  enregistrement des variables de classification 223
  exemple 219
  exportation des informations du modèle 223
  fonctions supplémentaires de la commande 223
  Lambda de Wilks 221
  matrice de covariance 222
  matrices 221
  méthodes de l'analyse discriminante 221
  méthodes détaillées étape par étape 219
  Options d'affichage 221, 222
  probabilités a priori 222
  sélection d'observations 220
  Statistiques 219, 221
  Statistiques descriptives 221
  tracés 222
  V de Rao 221
  valeur manquante 222
  variables de regroupement 219
  variables indépendantes 219
Analyse du voisin le plus proche

Analyse du voisin le plus proche *(suite)*
enregistrement de variables 215
Options 216
partitions 214
sélection des fonctions 214
sortie 215
voisins 213
vue du modèle 216
analyse en composantes principale 223, 225
Analyse factorielle
convergence 225, 226
descriptives 224
exemple 223
fonctions supplémentaires de la commande 227
format d'affichage des coefficients 227
méthodes d'extraction 225
méthodes de rotation 226
présentation 223
scores factoriels 226
sélection d'observations 224
Statistiques 223, 224
tracés de chargement 226
valeur manquante 227
Analyse PROXMAP 146, 149
Analyse ROC
Tracés et statistiques 295, 296
Analyse, boîte de dialogue
Méta-analyse : Taille d'effet binaire 69
Méta-analyse : Taille d'effet continue 51
Méta-analyse binaire 60
Méta-analyse continue 43
Andrew
dans Explorer 96
ANOVA
Dans ANOVA à 1 facteur 124
dans des modèles linéaires 172
dans GLM - Univarié 129
dans Moyennes 105
modèle 130
ANOVA à 1 facteur
Comparaisons multiples 126
Contrastes 125
contrastes polynomiaux 125
fonctions supplémentaires de la commande 129
Options 128
Statistiques 128
Tests post hoc 126
valeur manquante 128
variables de facteur 124
Anscombe
dans les proportions d'échantillons uniques 112
association linéaire par linéaire
Tableaux croisés 99
asymétrie
dans Descriptives 93
dans Explorer 96
dans les cubes OLAP 108
dans les effectifs 90
dans les rapports en colonnes 270
dans les rapports en lignes 268
dans Moyennes 105
dans Récapituler 103

B

B de Tukey
Dans ANOVA à 1 facteur 126
GLM 135
bagging
dans des modèles linéaires 168
Binomial ajusté p moyen
dans les proportions d'échantillons appariés 115
Binominal exact
dans les proportions d'échantillons appariés 115
Bland-Altman
analyse de l'accord 141, 142
boîte de dialogue Biais
Méta-analyse : Taille d'effet binaire 72
Méta-analyse : Taille d'effet continue 54
Méta-analyse binaire 63
Méta-analyse continue 46
boîte de dialogue Contraste
Méta-analyse : Taille d'effet binaire 71
Méta-analyse : Taille d'effet continue 53
Méta-analyse binaire 62
Méta-analyse continue 45
Boîte de dialogue Critères
Méta-analyse : Taille d'effet binaire 68
Méta-analyse binaire 59
Méta-analyse continue 42, 51
Régression de la méta-analyse 77
boîte de dialogue Enregistrer
Méta-analyse : Taille d'effet binaire 74
Méta-analyse : Taille d'effet continue 57
Méta-analyse binaire 66
Méta-analyse continue 49
Régression de la méta-analyse 79
boîte de dialogue Imprimer
Méta-analyse : Taille d'effet binaire 74
Méta-analyse : Taille d'effet continue 56
Méta-analyse binaire 65
Méta-analyse continue 48
Régression de la méta-analyse 78
boîte de dialogue Tracé
Méta-analyse : Taille d'effet binaire 75
Méta-analyse : Taille d'effet continue 57
Méta-analyse binaire 67
Méta-analyse continue 49
Régression de la méta-analyse 79
boîtes à moustaches
comparaison des niveaux de facteur 96
comparaison des variables 96
dans Explorer 96
dans la simulation 311
Bonett-Price
dans les proportions d'échantillons appariés 114
Bonferroni
Dans ANOVA à 1 facteur 126
GLM 135
boosting
dans des modèles linéaires 168
Brown-Li-Jeffreys
dans les proportions d'échantillons indépendants 117

C

C de Dunnett

C de Dunnett *(suite)*
Dans ANOVA à 1 facteur 126
GLM 135
carré de la distance euclidienne
distances 163
carte des quadrants
dans l'analyse du voisin le plus proche 218
CARTOGRAPHIE DE PROXIMITÉ 146, 149
catégorie de référence
GLM 132
ChiSquare 142, 143
Classification
dans Courbe ROC 297
dans l'analyse ROC 294
Clopper-Pearson (Exact)
dans les proportions d'échantillons uniques 112
coefficient alpha
dans l'analyse de fiabilité 272, 274
Coefficient d'incertitude
Tableaux croisés 99
coefficient de concordance de Kendall (W)
Tests non paramétriques pour échantillons liés 249
coefficient de contingence
Tableaux croisés 99
coefficient de corrélation de Spearman
Corrélations bivariées 143
Tableaux croisés 99
coefficient de corrélation intra-classe
dans l'analyse de fiabilité 274
coefficient de corrélation par rang
Corrélations bivariées 143
coefficient de corrélation r
Corrélations bivariées 143
Tableaux croisés 99
coefficient de dispersion (COD)
dans les statistiques de rapport 287
coefficient de variation (COV)
dans les statistiques de rapport 287
Coefficient Omega de McDonald
dans l'analyse de fiabilité 272, 274
coefficients bêta
dans la régression linéaire 178
coefficients de régression
dans la régression linéaire 178
colonne de total
dans les rapports 271
comparaison des groupes
dans les cubes OLAP 109
comparaison des méthodes
à Bland Altman 141
comparaison des variables
dans les cubes OLAP 109
Comparaisons multiples
Dans ANOVA à 1 facteur 126
Comparaisons multiples post hoc 126
conceptions pour les tests d'hétéroscédasticité
GLM 140
contrainte
Dans le positionnement multidimensionnel 278
contrainte S
Dans le positionnement multidimensionnel 278
Contrastes
Dans ANOVA à 1 facteur 125
GLM 132
contrastes de déviation
GLM 132
contrastes de différence
GLM 132
contrastes de Helmert
GLM 132
contrastes polynomiaux
Dans ANOVA à 1 facteur 125
GLM 132
contrastes répétés
GLM 132
contrastes simples
GLM 132
contrôle de page
Dans les rapports en colonnes 271
Dans les rapports en lignes 269
convergence
dans Analyse de cluster de nuées dynamiques 240
dans Analyse factorielle 225, 226
correction pour la continuité de Yates
Tableaux croisés 99
Corrélation de distance 164
corrélation de Pearson
Corrélations bivariées 143
Tableaux croisés 99
corrélations
Corrélations bivariées 143
dans Corrélations partielles 161
dans la simulation 308
ordre zéro 162
Tableaux croisés 99
Corrélations bivariées
coefficients de corrélation 143
fonctions supplémentaires de la commande 146
intervalle de confiance 143
intervalles de confiance 145
niveau de signification 143
Options 145
Statistiques 145
valeur manquante 145
Corrélations partielles
corrélations simples 162
dans la régression linéaire 178
fonctions supplémentaires de la commande 162
Options 162
Statistiques 162
valeur manquante 162
corrélations simples
dans Corrélations partielles 162
couches
Tableaux croisés 99
Courbe ROC
Tracés et statistiques 298
critère d'information d'Akaike
dans des modèles linéaires 169
critère de prévention du surajustement
dans des modèles linéaires 169
critères
à Bland Altman 141
critères d'informations
dans des modèles linéaires 169
critères de sélection
dans la régression linéaire 178
Cubes OLAP

Cubes OLAP *(suite)*
Statistiques 108
titres 110

D

d
Tableaux croisés 99
D de Somers
Tableaux croisés 99
de sécurité
colonne de total 271
comparaison des colonnes 271
division des valeurs de colonnes 271
multiplication des valeurs de colonnes 271
rapport en lignes 267
rapports en colonnes 270
totaux composites 271
décomposition hiérarchique 131
Définir les jeux de réponses multiples
catégories 263
définition des libellés 263
dichotomies 263
noms d'ensemble 263
dendrogrammes
dans Analyse de cluster hiérarchique 238
dernier
dans les cubes OLAP 108
dans Moyennes 105
dans Récapituler 103
Descriptives
enregistrement des scores z 92
fonctions supplémentaires de la commande 93
ordre d'affichage 93
Statistiques 93
déviation absolue moyenne (AAD)
dans les statistiques de rapport 287
DfBeta
dans la régression linéaire 176
DfFit
dans la régression linéaire 176
dictionnaire
Livre de codes 86
Différence la moins significative
Dans ANOVA à 1 facteur 126
GLM 135
Différence significative de Tukey
Dans ANOVA à 1 facteur 126
GLM 135
différences entre les groupes
dans les cubes OLAP 109
différences entre les variables
dans les cubes OLAP 109
différentiel lié au prix (PRD)
dans les statistiques de rapport 287
distance de Cook
dans la régression linéaire 176
GLM 137
Distance de Mahalanobis
dans Analyse discriminante 221
dans la régression linéaire 176
distance de Manhattan
dans l'analyse du voisin le plus proche 213
distances 163
Distance de Minkowski
distances 163
distance de Tchebycheff
distances 163
distance du khi-deux
distances 163
Distance euclidienne
dans l'analyse du voisin le plus proche 213
distances 163
distance Manhattan
dans l'analyse du voisin le plus proche 213
Distances
calcul des distances existant entre des observations 163
calcul des distances existant entre des variables 163
exemple 163
fonctions supplémentaires de la commande 164
mesures de dissimilarité 163
mesures de similarité 164
Statistiques 163
transformation de mesures 163, 164
Transformation de valeurs 163, 164
distances du voisin le plus proche
dans l'analyse du voisin le plus proche 218
division
division dans les colonnes de rapports 271

E

écart type
dans Descriptives 93
dans Explorer 96
dans les cubes OLAP 108
dans les effectifs 90
dans les rapports en colonnes 270
dans les rapports en lignes 268
dans les statistiques de rapport 287
dans Moyennes 105
dans Récapituler 103
échantillon d'apprentissage
dans l'analyse du voisin le plus proche 214
échantillon restant
dans l'analyse du voisin le plus proche 214
échantillons liés 259, 262
échelle
dans Kappa pondéré 276
dans l'analyse de fiabilité 272
Dans le positionnement multidimensionnel 278
effectif observé
Tableaux croisés 101
effectif théorique
Tableaux croisés 101
élimination descendante
dans la régression linéaire 175
ensembles
dans des modèles linéaires 170
erreur standard
dans Courbe ROC 298
dans Descriptives 93
dans Explorer 96
dans l'analyse ROC 295, 296
dans les effectifs 90
GLM 137
erreur standard d'asymétrie

erreur standard d'asymétrie *(suite)*
dans les cubes OLAP 108
dans Moyennes 105
dans Récapituler 103
erreur standard de kurtosis
dans les cubes OLAP 108
dans Moyennes 105
dans Récapituler 103
erreur standard de la moyenne
dans les cubes OLAP 108
dans Moyennes 105
dans Récapituler 103
Estimation de courbe
analyse de variance 207
enregistrement d'intervalles de prévision 208
enregistrement de prévisions 208
enregistrement de résidus 208
inclusion de la constante 207
modèles 208
prévision 208
estimations de Hodges-Lehman
Tests non paramétriques pour échantillons liés 249
estimations des paramètres
régression ordinale 181
êta
dans Moyennes 105
Tableaux croisés 99
êta carré
dans Moyennes 105
étude appariée
test T pour échantillons appariés 121
étude de contrôle d'observation
Test T pour échantillons appariés 121
Explorer
fonctions supplémentaires de la commande 97
Options 97
Statistiques 96
tracés 96
Transformations de l'exposant 97
valeur manquante 97

F

F de R-E-G-W (Ryan-Einot-Gabriel-Welsch)
Dans ANOVA à 1 facteur 126
GLM 135
F multiple de Ryan-Einot-Gabriel-Welsch
Dans ANOVA à 1 facteur 126
GLM 135
facteur d'inflation de la variance
dans la régression linéaire 178
factorisation en axes principaux 225
factorisation en projections 225
fiabilité de Spearman-Brown
dans l'analyse de fiabilité 274
fiabilité Split-half
dans l'analyse de fiabilité 272, 274
Filtres de séries temporelles
Christiano Fitzgerald Filtre 205, 206
Filtre Baxter king 205, 206
Filtre Hodrick Prescott 205, 206
fonctions de densité de probabilité
dans la simulation 310
fonctions de distribution cumulée
dans la simulation 310
fonctions de distribution cumulée *(suite)*
dans la simulation 310
Formatage
Colonnes dans les rapports 268
Fréquences
formats 92
graphiques 91
ordre d'affichage 92
Statistiques 90
suppression de tableaux 92
fréquences cumulées
régression ordinale 181
fréquences de cluster
dans l'analyse de cluster TwoStep 231
fréquences observées
régression ordinale 181
fréquences théoriques
régression ordinale 181

G

gamma
Tableaux croisés 99
Gamma de Goodman et Kruskal
Tableaux croisés 99
Générateur de simulation 301
gestion du bruit
dans l'analyse de cluster TwoStep 229
GLM
Enregistrement de matrices 137
enregistrement de variables 137
modèle 130
somme des carrés 130
Tests post hoc 135
tracés de profil 132
GLM - Univarié
Contrastes 132
Graphique à bulles
Régression de la méta-analyse 82
Graphique de Galbraith 80
Graphique de l'Abbé 80
Graphique de l'Abbé, onglet
Méta-analyse : Taille d'effet binaire 85
Méta-analyse : Taille d'effet continue 85
Méta-analyse binaire 85
Méta-analyse continue 85
graphique de l'espace des fonctions
dans l'analyse du voisin le plus proche 216
Graphique en entonnoir 80
Graphique en forêt 80
Graphique en forêt cumulée 80
graphiques
dans Courbe ROC 297
dans l'analyse ROC 294
libellés d'observations 207
tracés de probabilités 289, 291
Graphiques à barres
dans les effectifs 91
graphiques circulaires
dans les effectifs 91
graphiques tornado
dans la simulation 311
GT2 de Hochberg
Dans ANOVA à 1 facteur 126

GT2 de Hochberg *(suite)*
GLM 135

H

H de Kruskal-Wallis
Tests pour deux échantillons indépendants 260
Hauck-Anderson
dans les proportions d'échantillons indépendants 117, 118
histogrammes
dans Explorer 96
dans la régression linéaire 176
dans les effectifs 91
historique des itérations
régression ordinale 181

I

ICC. Voir Coefficient de corrélation intra-classe 274
importance des prédicteurs
modèles linéaires 171
importance des variables
dans l'analyse du voisin le plus proche 218
imprimer
à Bland Altman 142
index de concentration
dans les statistiques de rapport 287
Indice 112
Inférence, boîte de dialogue
Méta-analyse : Taille d'effet binaire 70
Méta-analyse : Taille d'effet continue 52
Méta-analyse binaire 61
Méta-analyse continue 44
Régression de la méta-analyse 77
informations de diagnostic de colinéarité
dans la régression linéaire 178
informations de diagnostic des observations
dans la régression linéaire 178
intervalle
dans Descriptives 93
dans les cubes OLAP 108
dans les effectifs 90
dans les statistiques de rapport 287
dans Moyennes 105
dans Récapituler 103
intervalle de confiance
Corrélations bivariées 143
intervalles de Clopper-Pearson
Tests non paramétriques à un échantillon 242
intervalles de confiance
Corrélations bivariées 145
Dans ANOVA à 1 facteur 128
dans Courbe ROC 298
dans Explorer 96
dans l'analyse ROC 295, 296
dans la régression linéaire 178
enregistrement dans la régression linéaire 176
GLM 132
test T pour échantillon unique 124
test T pour échantillons appariés 122
test T pour échantillons indépendants 121
intervalles de Jeffreys
intervalles de Jeffreys *(suite)*
Tests non paramétriques à un échantillon 242
intervalles de prévision
enregistrement dans l'estimation de courbe 208
enregistrement dans la régression linéaire 176
intervalles du rapport de vraisemblance
Tests non paramétriques à un échantillon 242
itérations
dans Analyse de cluster de nuées dynamiques 240
dans Analyse factorielle 225, 226

J

Jeffreys
dans les proportions d'échantillons uniques 112
jeux de réponses multiples
Livre de codes 86

K

k et sélection de fonction
dans l'analyse du voisin le plus proche 219
Kappa
Tableaux croisés 99
Kappa de Cohen
Tableaux croisés 99
kappa de Fleiss à plusieurs observateurs 272, 274
Kappa pondéré
critères 277
exemple 276
imprimer 278
Statistiques 276, 277
tableau croisé 278
Kappa pondéré de Cohen 276–278
khi-carré
association linéaire par linéaire 99
correction pour la continuité de Yates 99
indépendance 99
Options 253
Pearson 99
plage théorique 252
rapport de vraisemblance 99
Statistiques 253
Tableaux croisés 99
test à un échantillon 252
Test d'hypothèse statistique 142, 143
Test exact de Fisher 99
valeur manquante 253
valeurs théoriques 252
khi-carré de Pearson
régression ordinale 181
Tableaux croisés 99
khi-deux du rapport de vraisemblance
régression ordinale 181
Tableaux croisés 99
KR20
dans l'analyse de fiabilité 274
Kuder-Richardson 20 (KR20)
dans l'analyse de fiabilité 274
kurtosis
dans Descriptives 93
dans Explorer 96
dans les cubes OLAP 108

kurtosis *(suite)*
dans les effectifs 90
dans les rapports en colonnes 270
dans les rapports en lignes 268
dans Moyennes 105
dans Récapituler 103

L

lambda
Tableaux croisés 99
Lambda de Goodman et Kruskal
Tableaux croisés 99
Lambda de Wilks
dans Analyse discriminante 221
lien
régression ordinale 180
Linéaire 164
liste des observations 102
Livre de codes
sortie 87
Statistiques 89
Logit
dans les proportions d'échantillons uniques 112
LSD de Fisher
GLM 135

M

M-estimateur de Huber
dans Explorer 96
M-estimateur redescendant de Hampel
dans Explorer 96
M-estimateurs
dans Explorer 96
matrice de corrélation
dans Analyse discriminante 221
dans Analyse factorielle 223, 224
régression ordinale 181
matrice de covariance
dans Analyse discriminante 221, 222
dans la régression linéaire 178
GLM 137
régression ordinale 181
matrice de forme
dans Analyse factorielle 223
matrice de transformation
dans Analyse factorielle 223
maximum
comparaison des colonnes de rapports 271
dans Descriptives 93
dans Explorer 96
dans les cubes OLAP 108
dans les effectifs 90
dans les statistiques de rapport 287
dans Moyennes 105
dans Récapituler 103
maximum de vraisemblance
dans Analyse factorielle 225
McFadden R2
régression ordinale 181
McNemar
dans les proportions d'échantillons appariés 115
McNemar (continuité corrigée)
dans les proportions d'échantillons appariés 115
médiane
dans Explorer 96
dans les cubes OLAP 108
dans les effectifs 90
dans les statistiques de rapport 287
dans Moyennes 105
dans Récapituler 103
médiane de groupes
dans les cubes OLAP 108
dans Moyennes 105
dans Récapituler 103
meilleurs sous-ensembles
dans des modèles linéaires 169
mesure de différence de la taille
distances 163
Mesure de différence de structures
distances 163
mesure de dissimilarité de Lance et Williams
distances 163
mesure de distance du Phi-deux
distances 163
mesures de distance
dans Analyse de cluster hiérarchique 237
dans l'analyse du voisin le plus proche 213
distances 163
mesures de la dispersion
dans Descriptives 93
dans Explorer 96
dans les effectifs 90
dans les statistiques de rapport 287
mesures de la distribution
dans Descriptives 93
dans les effectifs 90
mesures de la tendance centrale
dans Explorer 96
dans les effectifs 90
dans les statistiques de rapport 287
mesures de similarité
dans Analyse de cluster hiérarchique 237
distances 164
Méta-analyse
régression 40
résultats binaires 40
résultats continus 40
Tracés 80
Méta-analyse : Taille d'effet binaire
Analyse, boîte de dialogue 69
boîte de dialogue Biais 72
boîte de dialogue Contraste 71
Boîte de dialogue Critères 68
boîte de dialogue Enregistrer 74
boîte de dialogue Imprimer 74
boîte de dialogue Tracé 75
Graphique de l'Abbé, onglet 85
Inférence, boîte de dialogue 70
onglet Graphique à bulles 82
onglet Graphique de Galbraith 84
onglet Graphique en entonnoir 83
onglet Graphique en forêt 80
onglet Graphique en forêt cumulée 81
Tronquer et remplir, boîte de dialogue 72
Méta-analyse : Taille d'effet continue

Méta-analyse : Taille d'effet continue *(suite)*
Analyse, boîte de dialogue 51
boîte de dialogue Biais 54
boîte de dialogue Contraste 53
boîte de dialogue Enregistrer 57
boîte de dialogue Imprimer 56
boîte de dialogue Tracé 57
Graphique de l'Abbé, onglet 85
Inférence, boîte de dialogue 52
onglet Graphique à bulles 82
onglet Graphique de Galbraith 84
onglet Graphique en entonnoir 83
onglet Graphique en forêt 80
onglet Graphique en forêt cumulée 81
Tronquer et remplir, boîte de dialogue 55
Méta-analyse binaire
Analyse, boîte de dialogue 60
boîte de dialogue Biais 63
boîte de dialogue Contraste 62
Boîte de dialogue Critères 59
boîte de dialogue Enregistrer 66
boîte de dialogue Imprimer 65
boîte de dialogue Tracé 67
Graphique de l'Abbé, onglet 85
Inférence, boîte de dialogue 61
onglet Graphique à bulles 82
onglet Graphique de Galbraith 84
onglet Graphique en entonnoir 83
onglet Graphique en forêt 80
onglet Graphique en forêt cumulée 81
Tronquer et remplir, boîte de dialogue 64
Méta-analyse continue
Analyse, boîte de dialogue 43
boîte de dialogue Biais 46
boîte de dialogue Contraste 45
Boîte de dialogue Critères 42, 51
boîte de dialogue Enregistrer 49
boîte de dialogue Imprimer 48
boîte de dialogue Tracé 49
Graphique de l'Abbé, onglet 85
Inférence, boîte de dialogue 44
onglet Graphique à bulles 82
onglet Graphique de Galbraith 84
onglet Graphique en entonnoir 83
onglet Graphique en forêt 80
onglet Graphique en forêt cumulée 81
Tronquer et remplir, boîte de dialogue 46
minimum
comparaison des colonnes de rapports 271
dans Descriptives 93
dans Explorer 96
dans les cubes OLAP 108
dans les effectifs 90
dans les statistiques de rapport 287
dans Moyennes 105
dans Récapituler 103
mise en cluster
affichage de clusters 232
affichage général 232
choix d'une procédure 227
mode
dans les effectifs 90
modèle composé
estimation de courbe 208
modèle cubique
estimation de courbe 208
modèle d'échelle
régression ordinale 182
modèle d'emplacement
régression ordinale 182
modèle de croissance
estimation de courbe 208
modèle de Guttman
dans l'analyse de fiabilité 272, 274
modèle de puissance
estimation de courbe 208
modèle de régression auxiliaire
GLM 140
modèle en S
estimation de courbe 208
modèle exponentiel
estimation de courbe 208
modèle inverse
estimation de courbe 208
modèle linéaire
estimation de courbe 208
modèle logarithmique
estimation de courbe 208
modèle logistique
estimation de courbe 208
modèle parallèle
dans l'analyse de fiabilité 272, 274
modèle parallèle strict
dans l'analyse de fiabilité 272, 274
modèle quadratique
estimation de courbe 208
modèles factoriels complets
GLM 130
modèles linéaires
choix du modèle 169
coefficients 173
critère d'informations 171
duplication des résultats 170
ensembles 170
importance des prédicteurs 171
moyennes estimées 173
niveau de fiabilité 169
objectifs 168
options de modèle 170
préparation automatique des données 169, 171
récapitulatif de génération de modèle 173
récapitulatif du modèle 171
règles de combinaison 170
Résidus 172
statistique R-deux 171
tableau ANOVA 172
valeurs extrêmes 172
valeurs prédites en fonction des valeurs observées 172
modèles personnalisés
GLM 130
modélisation géospatiale 317–328
modélisation spatiale 317
moindres carrés généralisés
dans Analyse factorielle 225
moindres carrés non pondérés
dans Analyse factorielle 225
moindres carrés pondérés
dans la régression linéaire 174

moyenne
Dans ANOVA à 1 facteur 128
dans Descriptives 93
dans Explorer 96
dans les cubes OLAP 108
dans les effectifs 90
dans les rapports en colonnes 270
dans les rapports en lignes 268
dans les statistiques de rapport 287
dans Moyennes 105
dans Récapituler 103
des colonnes de rapports multiples 271
sous-groupe 105, 107
moyenne géométrique
dans les cubes OLAP 108
dans Moyennes 105
dans Récapituler 103
moyenne harmonique
dans les cubes OLAP 108
dans Moyennes 105
dans Récapituler 103
moyenne pondérée
dans les statistiques de rapport 287
moyenne tronquée
dans Explorer 96
Moyennes
Options 105
Statistiques 105
moyennes des groupes 105, 107
moyennes des sous-groupes 105, 107
multiplication
multiplication dans les colonnes de rapports 271

N

Newcombe
dans les proportions d'échantillons appariés 114
dans les proportions d'échantillons indépendants 117
Newcombe (continuité corrigée)
dans les proportions d'échantillons indépendants 117
Newman-Keuls
GLM 135
niveau d'importance asymptotique 274
nombre d'observations
dans les cubes OLAP 108
dans Moyennes 105
dans Récapituler 103
nombre maximal de branches
dans l'analyse de cluster TwoStep 229
non linéaire 164
Normalisation Kaiser
dans Analyse factorielle 226
nuage de points
dans la simulation 311
nuages de points
dans la régression linéaire 176
Numérotation des pages
Dans les rapports en colonnes 271
Dans les rapports en lignes 269

O

onglet Graphique à bulles
onglet Graphique à bulles *(suite)*
Méta-analyse : Taille d'effet binaire 82
Méta-analyse : Taille d'effet continue 82
Méta-analyse binaire 82
Méta-analyse continue 82
onglet Graphique de Galbraith
Méta-analyse : Taille d'effet binaire 84
Méta-analyse : Taille d'effet continue 84
Méta-analyse binaire 84
Méta-analyse continue 84
onglet Graphique en entonnoir
Méta-analyse : Taille d'effet binaire 83
Méta-analyse : Taille d'effet continue 83
Méta-analyse binaire 83
Méta-analyse continue 83
onglet Graphique en forêt
Méta-analyse : Taille d'effet binaire 80
Méta-analyse : Taille d'effet continue 80
Méta-analyse binaire 80
Méta-analyse continue 80
onglet Graphique en forêt cumulée
Méta-analyse : Taille d'effet binaire 81
Méta-analyse : Taille d'effet continue 81
Méta-analyse binaire 81
Méta-analyse continue 81

P

pairs
dans l'analyse du voisin le plus proche 218
pas à pas ascendante
dans des modèles linéaires 169
percentiles
dans Explorer 96
dans la simulation 311
dans les effectifs 90
Percentiles 94
phi
Tableaux croisés 99
Plage multiple de Ryan-Einot-Gabriel-Welsch
Dans ANOVA à 1 facteur 126
GLM 135
Plum
régression ordinale 179
Positionnement multidimensionnel
conditionnalité 280
création de matrices de distance 280
critères 281
Définition de la forme des données 280
dimensions 280
exemple 278
fonctions supplémentaires de la commande 281
mesures de distance 280
modèles de positionnement 280
niveaux de mesure 280
Options d'affichage 281
Statistiques 278
Transformation de valeurs 280
pourcentages
Tableaux croisés 101
pourcentages en colonne
Tableaux croisés 101
Pourcentages en ligne
Tableaux croisés 101

pourcentages totaux
Tableaux croisés 101
premier
dans les cubes OLAP 108
dans Moyennes 105
dans Récapituler 103
préparation automatique des données
dans des modèles linéaires 171
prévision
estimation de courbe 208
Prévisions
enregistrement dans l'estimation de courbe 208
enregistrement dans la régression linéaire 176
Prévisions pondérées
GLM 137
profondeur d'arbre
dans l'analyse de cluster TwoStep 229
Proportions d'échantillons appariés 113
Proportions d'échantillons indépendants 116, 117
proportions d'échantillons uniques 114
Proportions d'échantillons uniques 110, 112
Proximité 146
Proximités
dans Analyse de cluster hiérarchique 236

Q

Q de Cochran
dans les tests pour plusieurs échantillons liés 262
Q de R-E-G-W (Ryan-Einot-Gabriel-Welsch)
Dans ANOVA à 1 facteur 126
GLM 135
Qualité de l'ajustement
régression ordinale 181
quartiles
dans les effectifs 90

R

R 2
dans la régression linéaire 178
dans Moyennes 105
modification R 2 178
R multiple
dans la régression linéaire 178
R-carré
dans des modèles linéaires 171
R-deux ajusté
dans des modèles linéaires 169
R2 ajusté
dans la régression linéaire 178
R2 de Cox et Snell
régression ordinale 181
R2 de Nagelkerke
régression ordinale 181
rapport de covariance
dans la régression linéaire 176
rapports en colonnes 270
récapitulatif d'erreur
dans l'analyse du voisin le plus proche 219
Récapitulatifs du rapport en colonnes
colonne de total 271
contrôle de page 271
Récapitulatifs du rapport en colonnes *(suite)*
fonctions supplémentaires de la commande 272
Format de colonne 268
Mise en page 269
Numérotation des pages 271
sous-totaux 271
Total général 271
valeur manquante 271
Récapitulatifs du rapport en lignes
colonnes de données 267
contrôle de page 268
critères d'agrégation 267
Espacement de rupture 268
fonctions supplémentaires de la commande 272
Format de colonne 268
Mise en page 269
Numérotation des pages 269
Pieds de page 269
séquences de tri 267
titres 269
valeur manquante 269
Variables dans les titres 269
Récapituler
Options 103
Statistiques 103
Rechercher 142, 143
règles de combinaison
dans des modèles linéaires 170
régression
Méta-analyse 40
Régression linéaire 174
régression multiple 174
tracés 176
Régression aux moindres carrés partiels
exporter des variables 210
modèle 210
Régression de la méta-analyse
Boîte de dialogue Critères 77
boîte de dialogue Enregistrer 79
boîte de dialogue Imprimer 78
boîte de dialogue Tracé 79
Graphique à bulles 82
Inférence, boîte de dialogue 77
Régression Lasso linéaire
Options 187
partitionner 187
variables 187
Régression linéaire
blocs 174
enregistrement des nouvelles variables 176
exportation des informations du modèle 176
fonctions supplémentaires de la commande 179
Méthodes de sélection des variables 175, 178
pondérations 174
Résidus 176
Statistiques 178
tracés 176
valeur manquante 178
variable de sélection 175
régression multiple
dans la régression linéaire 174
Régression nette élastique linéaire
Options 184
partitionner 183

Régression nette élastique linéaire *(suite)*
variables 183
Régression ordinale
fonctions supplémentaires de la commande 183
lien 180
modèle d'échelle 182
modèle d'emplacement 182
Options 180
Statistiques 179
Régression Ridge linéaire
Options 190
partitionner 190
variables 190
relation spatiale 146
Réponses multiples
fonctions supplémentaires de la commande 267
Résidu non standardisé
GLM 137
résiduels de Pearson
régression ordinale 181
Résidus
enregistrement dans l'estimation de courbe 208
enregistrement dans la régression linéaire 176
Tableaux croisés 101
résidus de Student
dans la régression linéaire 176
Résidus standardisés
dans la régression linéaire 176
GLM 137
Résidus supprimés
dans la régression linéaire 176
GLM 137
résultats binaires
données de tailles d'effet brutes 40
données de tailles d'effet brutes précalculées 40
résultats continus
données de tailles d'effet brutes 40
données de tailles d'effet brutes précalculées 40
Méta-analyse 40
rho
Corrélations bivariées 143
Tableaux croisés 99
Risque
Tableaux croisés 99
risque relatif
Tableaux croisés 99
rotation equamax
dans Analyse factorielle 226
rotation oblimin directe
dans Analyse factorielle 226
rotation quartimax
dans Analyse factorielle 226
rotation Varimax
dans Analyse factorielle 226

S

Score (continuité corrigée) 112
Score de Wilson
dans les proportions d'échantillons uniques 112
Score de Wilson (continuité corrigée)
dans les proportions d'échantillons uniques 112
scores factoriels 226
scores factoriels d'Anderson-Rubin 226
scores factoriels de Bartlett 226
scores z
dans Descriptives 92
enregistrement sous forme de variables 92
sélection ascendante
dans l'analyse du voisin le plus proche 214
dans la régression linéaire 175
sélection de k
dans l'analyse du voisin le plus proche 219
sélection des fonctions
dans l'analyse du voisin le plus proche 218
sélection progressive
dans la régression linéaire 175
seuil initial
dans l'analyse de cluster TwoStep 229
simulation
ajustement de la distribution 304
analyse d'hypothèses 308
analyse de sensibilité 308
boîtes à moustaches 311
corrélations entre entrées 308
création d'un plan de simulation 299, 300
création de nouvelles entrées 303
critères d'arrêt 309
échantillonnage des extrémités 309
éditeur d'équation 303
enregistrer les données simulées 312
enregistrer un plan de simulation 312
exécution d'un plan de simulation 301, 313
fonction de densité de probabilité 310
fonction de distribution cumulée 310
formats d'affichage des cibles et des entrées 311
Générateur de simulation 301
graphiques interactifs 315
graphiques tornado 311
modèles pris en charge 302
options de graphique 316
percentiles des distributions cible 311
personnalisation de l'ajustement de la distribution 307
réajustement des distributions aux nouvelles données 313
résultats de l'ajustement de la distribution 307
sortie 310, 311
spécification du modèle 302
tracés de dispersion 311
Simulation de Monte Carlo 298
Somme
dans Descriptives 93
dans les cubes OLAP 108
dans les effectifs 90
dans Moyennes 105
dans Récapituler 103
somme des carrés
GLM 130
sous-totaux
Dans les rapports en colonnes 271
standardisation
dans l'analyse de cluster TwoStep 229
statistique de Brown-Forsythe
Dans ANOVA à 1 facteur 128
Statistique de Cochran
Tableaux croisés 99
Statistique de Durbin-Watson
dans la régression linéaire 178

Statistique de Mantel-Haenszel
Tableaux croisés 99
statistique de Welch
Dans ANOVA à 1 facteur 128
statistique F
dans des modèles linéaires 169
statistique R
dans la régression linéaire 178
dans Moyennes 105
Statistiques de rapport
Statistiques 287
statistiques des proportions de colonne
Tableaux croisés 101
Statistiques descriptives
dans Descriptives 92
dans Explorer 96
dans l'analyse de cluster TwoStep 231
dans les effectifs 90
dans les statistiques de rapport 287
dans Récapituler 103
Student-Newman-Keuls
Dans ANOVA à 1 facteur 126
GLM 135
Suites de Wald-Wolfowitz
Tests pour deux échantillons indépendants 258
suites en séquences
Tests non paramétriques à un échantillon 242, 244
Suites en séquences
césures 254, 255
fonctions supplémentaires de la commande 255
Options 255
Statistiques 255
valeur manquante 255

T

T2 de Hotelling
dans l'analyse de fiabilité 272, 274
T2 de Tamhane
Dans ANOVA à 1 facteur 126
GLM 135
T3 de Dunnett
Dans ANOVA à 1 facteur 126
GLM 135
table de classification
dans l'analyse du voisin le plus proche 219
table de fréquences
dans Explorer 96
dans les effectifs 89
tableau croisé
multiréponses 265
Tableaux croisés 98
tableaux croisés
affichage de cellules 101
couches 99
formats 102
graphiques à barres en cluster 99
Statistiques 99
suppression de tableaux 98
Variables de contrôle 99
Tableaux croisés des réponses multiples
appariement des variables entre les vecteurs 266
Définition des plages de valeurs 266
pourcentages basés sur les observations 266
Tableaux croisés des réponses multiples *(suite)*
pourcentages basés sur les réponses 266
pourcentages de cellules 266
valeur manquante 266
tableaux de contingence 98
Tableaux de fréquences des réponses multiples
valeur manquante 264
taille d'effet
dans test T sur échantillons appariés 121
test T pour échantillons indépendants 119
Tau de Goodman et Kruskal
Tableaux croisés 99
tau de Kruskal
Tableaux croisés 99
tau-b
Tableaux croisés 99
Tau-b de Kendall
Corrélations bivariées 143
Tableaux croisés 99
tau-c
Tableaux croisés 99
Tau-c de Kendall
Tableaux croisés 99
termes construits 131, 182, 183
Termes d'interaction 131, 182, 183
test binomial
dichotomies 253
fonctions supplémentaires de la commande 254
Options 254
Statistiques 254
Tests non paramétriques à un échantillon 242
valeur manquante 254
Test d'additivité de Tukey
dans l'analyse de fiabilité 272, 274
test d'homogénéité marginale
Tests non paramétriques pour échantillons liés 249
tests pour deux échantillons liés 259
Test de comparaison appariée de Gabriel
Dans ANOVA à 1 facteur 126
GLM 135
Test de comparaison appariée de Games et Howell
Dans ANOVA à 1 facteur 126
GLM 135
test de Friedman
dans les tests pour plusieurs échantillons liés 262
Tests non paramétriques pour échantillons liés 249
test de Kolmogorov-Smirnov
Tests non paramétriques à un échantillon 242, 243
Test de Kolmogorov-Smirnov
test de Lilliefors 243
test de la médiane
Tests pour deux échantillons indépendants 260
test de Levene
Dans ANOVA à 1 facteur 128
dans Explorer 96
test de Lilliefors
dans Explorer 96
test de McNemar
Tableaux croisés 99
Tests non paramétriques pour échantillons liés 249, 250
tests pour deux échantillons liés 259
Test de plage multiple de Duncan
Dans ANOVA à 1 facteur 126

Test de plage multiple de Duncan *(suite)*
GLM 135
test de Scheffé
Dans ANOVA à 1 facteur 126
GLM 135
test de Shapiro-Wilks
dans Explorer 96
test de sphéricité de Bartlett
dans Analyse factorielle 224
test de Wilcoxon
Tests non paramétriques à un échantillon 242
Tests non paramétriques pour échantillons liés 249
tests pour deux échantillons liés 259
Test des droites parallèles
régression ordinale 181
Test des réactions extrêmes de Moses
Tests pour deux échantillons indépendants 258
test des signes
Tests non paramétriques pour échantillons liés 249
tests pour deux échantillons liés 259
test du khi-carré
Tests non paramétriques à un échantillon 242, 243
Test exact de Fisher
Tableaux croisés 99
Test Kolmogorov-Smirnov pour un échantillon
distribution du test 255
fonctions supplémentaires de la commande 257
Options 257
Statistiques 257
test de Lilliefors 255, 257
valeur manquante 257
Test M de Box
dans Analyse discriminante 221
test Q de Cochran
Tests non paramétriques pour échantillons liés 249, 250
test t
test T pour échantillon unique 123
test T pour échantillons appariés 121
test T pour échantillons indépendants 119
Test *t* d'échantillons appariés 121
Test t de Dunnett
Dans ANOVA à 1 facteur 126
GLM 135
Test t de Sidak
Dans ANOVA à 1 facteur 126
GLM 135
test t de Student 119
Test t de Waller-Duncan
Dans ANOVA à 1 facteur 126
GLM 135
test t dépendant
test T pour échantillons appariés 121
test t pour deux échantillons
test T pour échantillons indépendants 119
Test T pour échantillon unique
fonctions supplémentaires de la commande 123, 124
intervalles de confiance 124
Options 124
valeur manquante 124
Test T pour échantillons appariés
Options 122
sélection de variables appariées 121
valeur manquante 122
Test T pour échantillons indépendants
Test T pour échantillons indépendants *(suite)*
intervalles de confiance 121
Options 121
valeur manquante 121
Test T sur échantillons indépendants
Définition de groupes 121
variables de chaîne 121
variables de regroupement 121
Tests à deux échantillons indépendants
Définition de groupes 259
fonctions supplémentaires de la commande 259
Options 259
Statistiques 259
types de test 258
valeur manquante 259
variables de regroupement 259
tests à deux échantillons liés
fonctions supplémentaires de la commande 260
Options 260
Statistiques 260
types de test 260
valeur manquante 260
Tests d'homogénéité de la variance
Dans ANOVA à 1 facteur 128
Tests de l'indépendance
khi-carré 99
tests de linéarité
dans Moyennes 105
tests de normalité
dans Explorer 96
Tests de Wilcoxon sur échantillon unique 245
tests non paramétriques
khi-carré 252
Suites en séquences 254
Test Kolmogorov-Smirnov pour un échantillon 255
Tests à deux échantillons indépendants 258
tests à deux échantillons liés 259
Tests pour plusieurs échantillons indépendants 260
Tests pour plusieurs échantillons liés 262
Tests non paramétriques à un échantillon
suites en séquences 244
test binomial 242
test de Kolmogorov-Smirnov 243
test du khi-carré 243
Zones 242
tests non paramétriques pour échantillons indépendants
Champs, onglet 246
Tests non paramétriques pour échantillons liés
test de McNemar 250
test Q de Cochran 250
Zones 249
Tests pour plusieurs échantillons indépendants
définition de la plage 261
fonctions supplémentaires de la commande 262
Options 261
Statistiques 261
types de test 261
valeur manquante 261
variables de regroupement 261
Tests pour plusieurs échantillons liés
fonctions supplémentaires de la commande 263
Statistiques 262
types de test 262
titres

titres *(suite)*
dans les cubes OLAP 110
tolérance
dans la régression linéaire 178
Tolérance
Régression linéaire 178
Totaux généraux
Dans les rapports en colonnes 271
tracé de probabilités gaussien
dans Explorer 96
dans la régression linéaire 176
P-P 289
Q-Q 291
Tracé de probabilités P-P 289
tracé de répartition gaussien des résidus
dans Explorer 96
tracer
à Bland Altman 142
Tracés
Méta-analyse 80
tracés de chargement
dans Analyse factorielle 226
tracés de probabilités
P-P 289
Q-Q 291
Tracés de probabilités Q-Q 291
tracés de profil
GLM 132
tracés dispersion/niveau
dans Explorer 96
tracés en stalactite
dans Analyse de cluster hiérarchique 238
tracés partiels
dans la régression linéaire 176
tracés tige et feuille
dans Explorer 96
Tronquer et remplir, boîte de dialogue
Méta-analyse : Taille d'effet binaire 72
Méta-analyse : Taille d'effet continue 55
Méta-analyse binaire 64
Méta-analyse continue 46
TSF 205, 206
Tukey
dans Explorer 96

U

U de Mann-Whitney
Tests pour deux échantillons indépendants 258

V

V
Tableaux croisés 99
V de Cramér
Tableaux croisés 99
V de Rao
dans Analyse discriminante 221
valeur manquante
Corrélations bivariées 145
dans Analyse factorielle 227
Dans ANOVA à 1 facteur 128
dans Corrélations partielles 162
valeur manquante *(suite)*
dans Courbe ROC 298
dans Explorer 97
dans l'analyse du voisin le plus proche 216
dans l'analyse ROC 295, 296
dans la régression linéaire 178
dans les proportions d'échantillons appariés 116
dans les proportions d'échantillons indépendants 119
dans les proportions d'échantillons uniques 113
Dans les rapports en colonnes 271
dans les rapports en lignes 269
dans les tableaux croisés des réponses multiples 266
dans les tableaux de fréquences des réponses multiples 264
dans Test Kolmogorov-Smirnov pour un échantillon 257
sans les suites en séquences 255
test binomial 254
test du khi-carré 253
test T pour échantillon unique 124
test T pour échantillons appariés 122
test T pour échantillons indépendants 121
Tests pour deux échantillons indépendants 259
tests pour deux échantillons liés 260
tests pour plusieurs échantillons indépendants 261
valeurs extrêmes
dans Explorer 96
dans l'analyse de cluster TwoStep 229
dans la régression linéaire 176
valeurs influentes
dans la régression linéaire 176
GLM 137
Valeurs manquantes
dans les percentiles 94
valeurs propres
dans Analyse factorielle 224, 225
dans la régression linéaire 178
valeurs standardisées
dans Descriptives 92
variable de sélection
dans la régression linéaire 175
Variables de contrôle
Tableaux croisés 99
variance
dans Descriptives 93
dans Explorer 96
dans les cubes OLAP 108
dans les effectifs 90
dans les rapports en colonnes 270
dans les rapports en lignes 268
dans Moyennes 105
dans Récapituler 103
visualisation
modèles de classification 232
visualiseur de clusters
à propos des modèles de cluster 231
affichage du contenu des cellules 234
comparaison des clusters 235
distribution des cellules 234
faire basculer les clusters et les fonctions 233
filtrage des enregistrements 236
importance des prédicteurs 234
présentation 232
récapitulatif du modèle 232
taille des clusters 234

visualiseur de clusters *(suite)*
transposer les clusters et les fonctions 233
trier l'affichage des clusters 234
trier l'affichage des fonctions 233
trier le contenu des cellules 234
trier les clusters 234
trier les fonctions 233
Utilisation 235
vue de base 234
vue de comparaison des clusters 235
vue de l'importance des prédicteurs de cluster 234
vue de la distribution des cellules 234
vue de la taille des clusters 234
vue des centres de clusters 232
vue des clusters 232
vue récapitulative 232
vue du modèle
dans l'analyse du voisin le plus proche 216

W

W de Kendall
dans les tests pour plusieurs échantillons liés 262
Wald
dans les proportions d'échantillons appariés 114, 115
dans les proportions d'échantillons indépendants 117, 118
dans les proportions d'échantillons uniques 112
Wald (continuité corrigée)
dans les proportions d'échantillons appariés 114, 115
dans les proportions d'échantillons indépendants 117, 118
dans les proportions d'échantillons uniques 112
Wald H0
dans les proportions d'échantillons indépendants 118
Wald H0 (continuité corrigée)
dans les proportions d'échantillons indépendants 118

X

X^2 142, 143

Z

Z de Kolmogorov-Smirnov
dans Test Kolmogorov-Smirnov pour un échantillon 255
Tests pour deux échantillons indépendants 258