

IBM SPSS Custom Tables 23

IBM

Important

Avant d'utiliser le présent document et le produit associé, prenez connaissance des informations générales figurant à la section «Remarques», à la page 105.

Informations sur le produit

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

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

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

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

Vous pouvez également consulter les serveurs Internet suivants :

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

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

Table des matières

Avis aux lecteurs canadiens v

Chapitre 1. Interface de génération de tableaux. 1

Interface de génération de tableaux	1
Construction de tableaux	1
Procédure de création d'un tableau	4
Empilement de variables	4
Imbrication de variables	4
Couches	5
Affichage et masquage des noms et/ou libellés de variable	6
Statistiques récapitulatives	6
Catégories et totaux	12
Catégories calculées	14
Tableaux de variables avec catégories partagées (Comperimeter Tables).	15
Personnalisation du générateur de tableaux.	16
Tableaux personnalisés : onglet Options	16
Tableaux personnalisés : onglet Titres.	17
Tableaux personnalisés : onglet Tests statistiques	17

Chapitre 2. Tableaux simples pour variables catégorielles 21

Tableaux simples pour variables catégorielles	21
Une seule variable catégorielle	21
Pourcentages	22
Totaux	23
Tableau croisé	23
Pourcentages dans des tableaux croisés	24
Contrôle du format d'affichage	24
Totaux marginaux	25
Tri et exclusion de catégories	26

Chapitre 3. Empilement, imbrication et couches avec des variables catégorielles 29

Empilement de variables catégorielles	29
Empilement et tableau croisé	29
Imbrication de variables catégorielles	30
Suppression de libellés de variable	31
Tableau croisé imbriqué	32
Couches	33
Deux variables catégorielles de couche empilées	34
Deux variables catégorielles de couche imbriquées	35

Chapitre 4. Totaux et sous-totaux pour les variables catégorielles 37

Total simple pour une seule variable	37
Tel écran, tel totalisé	37
Position d'affichage des totaux	38
Totaux pour tableaux imbriqués	38
Totaux des variables de couche	40

Sous-totaux	40
Tel écran, tel sous-totalisé.	41
Masquage de catégories de sous-total.	41
Sous-totaux d'une variable de couche.	42

Chapitre 5. Catégories calculées pour les variables catégorielles 43

Catégorie calculée simple.	43
Masquage des catégories dans une catégorie calculée.	44
Référencement des sous-totaux dans une catégorie calculée.	44
Utilisation des catégories calculées pour afficher des sous-totaux non-exhaustifs	46

Chapitre 6. Tableaux pour variables avec catégories partagées 49

Tableau d'effectifs	49
Tableau de pourcentages	50
Contrôle des totaux et des catégories	51
Imbrication dans des tableaux comportant des catégories partagées	52

Chapitre 7. Statistiques récapitulatives 53

Variable source des statistiques récapitulatives.	53
Source des statistiques récapitulatives pour les variables catégorielles	54
Source des statistiques récapitulatives pour les variables d'échelle	54
Variables empilées	55
Autres statistiques récapitulatives en total pour les variables catégorielles	56
Affichage des valeurs des catégories	57

Chapitre 8. Récapitulatif des variables d'échelle 59

Récapitulatif des variables d'échelle	59
Variables d'échelle empilées	59
Statistiques récapitulatives multiples	60
Effectif, N valide et valeurs manquantes.	60
Différents récapitulatifs pour différentes variables	61
Récapitulatifs de groupe dans des catégories	62
Variables de regroupement multiples	63
Imbrication de variables catégorielles dans des variables d'échelle	63

Chapitre 9. Statistiques de test 65

Statistiques de test	65
Tests d'indépendance (khi-deux)	65
Effets de l'imbrication et de l'empilement sur les tests d'indépendance	66
Comparaison des moyennes de colonne	68
Effets de l'imbrication et de l'empilement sur les tests des moyennes de colonne	69

Comparaison des proportions de colonne	71
Effets de l'imbrication et de l'empilement sur les tests des proportions de colonne	74
Une remarque sur les pondérations et les jeux de réponses multiples	76

Chapitre 10. Jeux de réponses multiples 77

Effectifs, réponses, pourcentages et totaux	77
Utilisation de jeux de réponses multiples avec d'autres variables	79
Variable source de statistiques et statistiques récapitulatives disponibles	80
Jeu de catégories multiples et réponses en double	80
Test de signification pour les jeux de réponses multiples	82
Tests d'indépendance avec des jeux de réponses multiples	82
Comparaison des moyennes de colonne avec des jeux de réponses multiples	83

Chapitre 11. Valeurs manquantes 85

Tableaux dépourvus de valeurs manquantes	85
Inclusion de valeurs manquantes dans les tableaux	86

Chapitre 12. Formatage et personnalisation de tableaux 89

Formatage et personnalisation de tableaux	89
Formats d'affichage des statistiques récapitulatives	89
Affichage des libellés pour les statistiques récapitulatives	90
Largeur des colonnes	91
Affichage d'une valeur dans les cellules vides	92
Affichage d'une valeur pour les statistiques manquantes	92

Chapitre 13. Fichiers d'exemple 95

Remarques 105

Marques	107
-------------------	-----

Index 109

Avis aux lecteurs canadiens

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

Illustrations

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

Terminologie

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

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

Claviers

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

OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

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

Nomenclature

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

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

Brevets

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

Assistance téléphonique

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

Chapitre 1. Interface de génération de tableaux

Interface de génération de tableaux

L'option Tableaux personnalisés utilise une interface de génération de tableaux simple, de type "glisser-déplacer", qui vous permet d'obtenir un aperçu de votre tableau lorsque vous sélectionnez des variables et des options. Elle offre également un niveau de flexibilité introuvable dans les boîtes de dialogue classiques, comme la possibilité de modifier la taille de la fenêtre et des panneaux au sein de la fenêtre.

Construction de tableaux



Figure 1. Boîte de dialogue Tableaux personnalisés, onglet Tableau

L'onglet Tableau du générateur de tableaux vous permet de sélectionner les variables et les mesures récapitulatives qui apparaîtront dans vos tableaux.

Liste des variables : Les variables du fichier de données apparaissent dans le panneau supérieur gauche de la fenêtre. L'option Tableaux personnalisés opère une distinction entre deux niveaux de mesure différents pour les variables. Par conséquent, les variables sont traitées différemment selon leur niveau de mesure :

Qualitative : Données ayant un nombre limité de valeurs ou de catégories distinctes (par exemple, sexe ou religion). Les variables catégorielles peuvent être des données chaîne (alphanumérique) ou des variables numériques qui utilisent des codes chiffrés pour représenter les catégories (par exemple, 0 = *Masculin* et 1 = *Féminin*). Elles sont parfois également qualifiées de données catégorielles. Les variables catégorielles peuvent être **nominales** ou **ordinales**

- *Nominal*. Une variable peut être traitée comme étant nominale si ses valeurs représentent des catégories sans classement intrinsèque (par exemple, le service de la société dans lequel travaille un employé). La région, le code postal ou l'appartenance religieuse sont des exemples de variables nominales.
- *Ordinal*. Une variable peut être traitée comme étant ordinale si ses valeurs représentent des catégories associées à un classement intrinsèque (par exemple, des niveaux de satisfaction allant de Très mécontent 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 : Données mesurées sur une échelle d'intervalle ou de rapport, où les valeurs de données indiquent à la fois l'ordre des valeurs et la distance qui les sépare. Par exemple, un salaire de 58 160 € est supérieur à un salaire de 42 212 € et la distance entre les deux valeurs est de 15 948 €. Ces données sont aussi appelées données quantitatives ou données continues.

Les variables catégorielles définissent des catégories (ligne, colonnes et couches) dans le tableau et la statistique récapitulative par défaut est l'effectif (nombre d'observations dans chaque catégorie). Par exemple, le tableau par défaut d'une variable catégorielle telle que le sexe afficherait simplement le nombre d'hommes et le nombre de femmes.

Les variables d'échelle sont généralement récapitulées au sein des catégories des variables catégorielles et la statistique récapitulative par défaut est la moyenne. Par exemple, le tableau par défaut présentant la variable revenu au sein des catégories de la variable Sexe afficherait le revenu moyen des hommes et le revenu moyen des femmes.

Vous pouvez également récapituler les variables d'échelle seules, sans utiliser de variable catégorielle pour définir des groupes. Ceci est essentiellement utile pour l'**empilement** des récapitulatifs de plusieurs variables d'échelle. Pour plus d'informations, voir «Empilement de variables», à la page 4.

Jeux de réponses multiples

L'option Tableaux personnalisés prend également en charge un type spécifique de variable nommé **jeu de réponses multiples**. Les jeux de réponses multiples ne sont pas réellement des variables au sens habituel du terme. En effet, vous ne pouvez pas les voir dans l'éditeur de données et ils ne sont pas reconnus par les autres procédures. Les jeux de réponses multiples utilisent plusieurs variables pour enregistrer les réponses à des questions auxquelles le répondant peut donner plusieurs réponses. Les jeux de réponses multiples sont traités comme des variables catégorielles, et la plupart des actions appliquées à ces dernières peuvent également l'être aux jeux de réponses multiples. Pour plus d'informations, voir Chapitre 10, «Jeux de réponses multiples», à la page 77.

Dans la liste des variables, une icône indique le type de chaque variable.

Tableau 1. Icônes de niveau de mesure

	Numérique	Chaîne	Date	Heure
Echelle (continue).		n/a		
Ordinal				
Nominal				

Tableau 2. Icônes de jeu de réponses multiples

Type du jeu de réponses multiples	Icône
Jeu de réponses multiples - catégories multiples	
Jeu de réponses multiples - dichotomies multiples	

Pour modifier le niveau de mesure d'une variable dans le générateur de tableaux, cliquez avec le bouton droit de la souris sur la variable dans la liste et sélectionnez **Nominales** ou **Echelle** dans le menu contextuel. Vous pouvez modifier de façon permanente le niveau de mesure d'une variable dans la vue de variable de l'éditeur de données. Les variables définies comme étant **nominales** ou **ordinales** sont traitées comme étant catégorielles par l'option Tableaux personnalisés.

Catégories : Lorsque vous sélectionnez une variable catégorielle dans la liste des variables, les catégories définies pour la variable apparaissent dans la liste Catégories. Ces catégories apparaissent également dans le panneau de canevas lorsque vous utilisez la variable dans un tableau. Si la variable ne comporte aucune catégorie définie, la liste Catégories et le panneau de canevas affichent deux catégories de substitution : *Catégorie 1* et *Catégorie 2*.

Les catégories définies affichées dans le générateur de tableaux sont basées sur des **libellés de valeurs**, des libellés descriptifs attribués à différentes valeurs de données (par exemple, les valeurs numériques 0 et 1, avec les libellés de valeurs *homme* et *femme*). Vous pouvez définir des libellés de valeurs dans la vue de variable de l'éditeur de données ou via l'option Définir les propriétés de variable du menu Données de la fenêtre de l'éditeur de données.

Panneau de canevas : Pour construire un tableau, faites glisser des variables vers les lignes et colonnes du panneau de canevas. Le panneau de canevas affiche un aperçu du tableau qui sera créé. Le panneau de canevas n'affiche pas les valeurs de donnée réelles dans les cellules, mais fournit un aperçu relativement précis de la présentation du tableau final. En ce qui concerne les variables catégorielles, le tableau final peut contenir davantage de catégories que l'aperçu si le fichier de données contient des valeurs uniques pour lesquelles aucun libellé de valeur n'a été défini.

- Le mode **Normal** affiche toutes les lignes et colonnes qui seront incluses dans le tableau, y compris les lignes et/ou colonnes pour les statistiques récapitulatives et les catégories des variables catégorielles.
- Le mode **Compact** affiche uniquement les variables qui apparaîtront dans le tableau, sans aperçu des lignes et colonnes qui composeront le tableau.

Règles élémentaires et limites relatives à la construction d'un tableau

- Pour les variables catégorielles, les statistiques récapitulatives sont basées sur la variable la plus interne dans la dimension source.
- La dimension source par défaut (ligne ou colonne) des statistiques pour les variables catégorielles est basée sur l'ordre dans lequel vous avez déplacé les variables vers le panneau de canevas. Par exemple, si vous faites glisser d'abord une variable vers la zone Lignes, la dimension de ligne est la dimension source par défaut des statistiques.
- Les variables d'échelle peuvent être récapitulées uniquement au sein des catégories de la variable la plus interne dans la dimension de ligne ou de colonne. (Vous pouvez positionner la variable d'échelle à tout niveau du tableau, mais elle est récapitulée au niveau le plus interne.)
- Les variables d'échelle ne peuvent pas être récapitulées au sein d'autres variables d'échelle. Vous pouvez empiler les récapitulatifs de plusieurs variables d'échelle ou récapituler des variables d'échelle

au sein des catégories de variables catégorielles. Vous ne pouvez pas imbriquer une variable d'échelle dans une autre, ou placer une variable d'échelle dans la dimension de ligne et une autre dans la dimension de colonne.

- Si une variable du jeu de données actif comprend plus de 12 000 libellés de valeurs définis, vous ne pouvez pas utiliser le générateur de tableaux pour créer des tableaux. Si vous n'avez pas besoin d'inclure dans vos tableaux des variables dépassant cette limite, vous pouvez définir et appliquer des jeux de variables qui les excluent. Si vous devez inclure des variables comportant plus de 12 000 libellés de valeurs définis, utilisez la syntaxe de commande CTABLES pour générer les tableaux.

Procédure de création d'un tableau

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Faites glisser des variables vers les zones Lignes et/ou Colonnes du panneau de canevas.
3. Cliquez sur **OK** pour créer le tableau.
4. Dans le panneau de canevas, sélectionnez la variable (en cliquant dessus).
5. Faites glisser la variable vers l'extérieur du panneau de canevas ou appuyez sur la touche Suppr.
Pour modifier le niveau de mesure d'une variable :
6. Dans la liste des variables, cliquez avec le bouton droit de la souris sur la variable (vous ne pouvez effectuer cette opération que dans la liste des variables, pas dans le canevas).
7. Dans le menu contextuel, sélectionnez **Nominales** ou **Echelle**.

Empilement de variables

Empiler revient à prendre des tableaux séparés et à les coller ensemble dans le même affichage. Vous pouvez, par exemple, afficher des informations sur les variables *sexe* et *tranche d'âge* dans des sections séparées du même tableau.

Pour empiler des variables

1. Dans la liste des variables, sélectionnez toutes les variables à empiler, puis faites-les glisser simultanément vers les lignes ou colonnes du panneau de canevas.
ou
2. Faites glisser les variables séparément, en positionnant chacune d'elles au-dessus ou au-dessous des variables existantes dans les lignes, ou à droite ou à gauche des variables existantes dans les colonnes.

Tableau 3. Variables catégorielles empilées.

Variables	Catégories	Statistiques récapitulatives
Variable 1	Catégorie 1	123
	Catégorie 2	456
Variable 2	Catégorie 1	123
	Catégorie 2	456
	Catégorie 3	789

Pour plus d'informations, voir «Empilement de variables catégorielles», à la page 29.

Imbrication de variables

L'imbrication, comme la technique des tableaux croisés, permet de présenter la relation entre deux variables catégorielles ; la seule différence est qu'une variable est imbriquée dans l'autre dans la même dimension. Vous pouvez, par exemple, imbriquer la variable *Sexe* dans la variable *Tranche d'âge* dans la dimension de ligne, ce qui a pour effet de présenter le nombre d'hommes et de femmes dans chaque tranche d'âge.

Vous pouvez également imbriquer une variable d'échelle dans une variable catégorielle. Vous pouvez, par exemple, imbriquer la variable *Revenu* dans la variable *Sexe*, ce qui a pour effet de présenter des valeurs de revenu moyen (ou la médiane ou toute autre mesure récapitulative) distinctes pour les hommes et les femmes.

Pour imbriquer des variables

1. Faites glisser une variable catégorielle vers la zone Lignes ou Colonnes du panneau de canevas.
2. Faites glisser une variable catégorielle ou une variable d'échelle vers la gauche ou la droite de la variable catégorielle de ligne, ou au-dessus ou au-dessous de la variable catégorielle de colonne.

Tableau 4. Variables catégorielles imbriquées.

Variable 1	Variable 2	Statistiques récapitulatives
Catégorie 1	Catégorie 1	12
	Catégorie 2	34
	Catégorie 3	56
Catégorie 2	Catégorie 1	12
	Catégorie 2	34
	Catégorie 3	56

Pour plus d'informations, voir «Imbrication de variables catégorielles», à la page 30.

Remarque : Les tableaux personnalisés ne prennent pas en charge le traitement de fichiers scindés en couches. Pour obtenir le même résultat que les fichiers scindés en couches, placez les variables de fichiers scindés dans les couches imbriquées les plus à l'extérieur du tableau.

Couches

Vous pouvez utiliser des couches pour ajouter une dimension de profondeur à vos tableaux et créer ainsi des « cubes » tridimensionnels. Les couches sont similaires à l'imbrication ou à l'empilement ; la différence principale réside dans le fait qu'une seule catégorie de couche est visible à la fois. Par exemple, l'utilisation de la variable *Tranche d'âge* comme variable de ligne et de la variable *Sexe* comme variable de couche a pour effet de produire un tableau dans lequel les informations relatives aux hommes et aux femmes apparaissent dans différentes couches.

Pour créer des couches

1. Dans l'onglet Tableau du générateur de tableaux, cliquez sur **Couches** pour afficher la liste Couches.
2. Faites glisser les variables d'échelle ou catégorielles qui définiront les couches vers la liste Couches.

Vous ne pouvez pas mélanger des variables d'échelle et des variables catégorielles dans la liste Couches. Toutes les variables doivent être du même type. Les jeux de réponses multiples sont traités comme des variables catégorielles pour la liste Couches. Les variables d'échelle dans les couches sont toujours empilées.

Si vous disposez de plusieurs variables catégorielles en couches, les couches peuvent être empilées ou imbriquées.

- L'**affichage d'une catégorie en couches** équivaut à l'empilement. Une couche distincte sera affichée pour chaque catégorie de chaque variable de couche. Le nombre total de couches est simplement la *somme* du nombre de catégories pour chaque variable de couche. Par exemple, si vous disposez de trois variables de couche, chacune comportant trois catégories, le tableau sera composé de neuf couches.

- L'**affichage de chaque combinaison de catégories en couche** équivaut à l'imbrication ou au croisement de couches. Le nombre total de couches est le *produit* du nombre de catégories pour chaque variable de couche. Par exemple, si vous disposez de trois variables, chacune comportant trois catégories, le tableau sera composé de 27 couches.

Affichage et masquage des noms et/ou libellés de variable

Les options suivantes sont disponibles pour l'affichage des noms et libellés de variable :

- Affichage des libellés de variable uniquement. Pour toute variable sans libellé de variable défini, le nom de la variable apparaît. Il s'agit du paramètre par défaut.
- Affichage des noms de variable uniquement.
- Affichage des libellés et des noms de variable.
- Masquage des noms ou des libellés de variable. Bien que la colonne/ligne qui contient le libellé ou le nom de variable reste affichée dans l'aperçu du tableau dans le panneau de canevas, cette colonne/ligne n'apparaîtra pas dans le tableau final.

Pour afficher ou masquer les libellés ou noms de variable :

1. Dans l'aperçu du tableau dans le panneau de canevas, cliquez avec le bouton droit de la souris sur la variable.
2. Sélectionnez **Afficher le libellé de la variable** ou **Afficher le nom de la variable** dans le menu contextuel pour activer ou désactiver l'affichage des libellés ou des noms. Une coche en regard de la sélection indique que ces éléments seront affichés.

Statistiques récapitulatives

La boîte de dialogue Statistiques récapitulatives vous permet d'effectuer les opérations suivantes :

- Ajouter et supprimer des statistiques récapitulatives dans un tableau.
- Modifier les libellés des statistiques.
- Modifier l'ordre des statistiques.
- Modifier le format des statistiques, comme le nombre de décimales.

Les statistiques récapitulatives (et autres options) disponibles ici dépendent du niveau de mesure de leur variable source, comme affiché en haut de la boîte de dialogue. La source des statistiques récapitulatives (variable sur laquelle sont basées les statistiques) est déterminée par les éléments suivants :

- **Niveau de mesure** : Si un tableau (ou une section dans un tableau de type empilé) contient une variable d'échelle, les statistiques récapitulatives sont basées sur cette dernière.
- **Ordre de sélection des variables**. La dimension source par défaut (ligne ou colonne) des statistiques pour les variables catégorielles est basée sur l'ordre dans lequel vous avez glissé les variables vers le panneau de canevas. Par exemple, si vous faites glisser d'abord une variable vers la zone Lignes, la dimension de ligne est la dimension source par défaut des statistiques.
- **Imbrication** : Pour les variables catégorielles, les statistiques récapitulatives sont basées sur la variable la plus interne dans la dimension source.

Un tableau empilé peut comporter plusieurs variables sources de statistiques récapitulatives (variable d'échelle et variable catégorielle), mais chaque section du tableau comporte une seule source de statistiques récapitulatives.

Pour modifier la dimension source des statistiques récapitulatives

1. Dans le groupe Statistiques récapitulatives de l'onglet Tableau, sélectionnez la dimension (ligne, colonne ou couche) dans la liste déroulante **Source**.

Pour contrôler les statistiques récapitulatives affichées dans un tableau

1. Dans le panneau de canevas de l'onglet Tableau, sélectionnez la variable source des statistiques récapitulatives (cliquez dessus).
2. Dans le groupe Définir de l'onglet Tableau, cliquez sur **Statistiques récapitulatives**.
ou
3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur la variable source des statistiques récapitulatives et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
4. Sélectionnez les statistiques récapitulatives à inclure dans le tableau. Pour déplacer les statistiques sélectionnées de la liste Statistiques vers la liste Afficher, vous pouvez utiliser la flèche ou faire glisser les statistiques.
5. Cliquez sur les flèches Haut ou Bas pour modifier la position d'affichage de la statistique récapitulative sélectionnée.
6. Dans la liste déroulante Format, sélectionnez un format d'affichage pour la statistique récapitulative sélectionnée.
7. Entrez le nombre de décimales à afficher dans la cellule Décimales pour la statistique récapitulative sélectionnée.
8. Cliquez sur **Appliquer à la sélection** pour inclure les statistiques récapitulatives sélectionnées pour les variables sélectionnées dans le panneau de canevas.
9. Cliquez sur **Appliquer à tout** pour inclure les statistiques récapitulatives définies pour toutes les variables empilées du même type dans le panneau de canevas.

Remarque : L'option **Appliquer à tout** diffère de l'option **Appliquer à la sélection** uniquement pour les variables empilées du même type déjà présentes dans le panneau de canevas. Dans les deux cas, les statistiques récapitulatives sélectionnées sont automatiquement incluses pour tout ajout de variables empilées du même type au tableau.

Statistiques récapitulatives pour les variables catégorielles

Les statistiques de base disponibles pour les variables catégorielles sont les effectifs et les pourcentages. Vous pouvez également indiquer d'autres statistiques récapitulatives pour les totaux et sous-totaux. Ces autres statistiques récapitulatives incluent les mesures de tendance centrale (comme la moyenne et la médiane) et de dispersion (comme l'écart-type) qui peuvent convenir à certaines variables catégorielles ordinales. Pour plus d'informations, voir «Autres statistiques récapitulatives en total pour les variables catégorielles», à la page 10.

Effectif : Nombre d'observations dans chaque cellule du tableau ou nombre de réponses pour les jeux de réponses multiples.

Occurrence non pondérée : Nombre non pondéré d'observations dans chaque cellule du tableau. Ceci diffère de l'effectif uniquement si la pondération est activée.

Pourcentages en colonne : Pourcentages dans chaque colonne. La somme des pourcentages dans chaque colonne d'une sous-table (pour les pourcentages simples) est égale à 100 %. Les pourcentages en colonne sont généralement utiles uniquement si vous disposez d'une variable catégorielle de *ligne*.

Pourcentages en ligne : Pourcentages dans chaque ligne. La somme des pourcentages dans chaque ligne d'une sous-table (pour les pourcentages simples) est égale à 100 %. Les pourcentages en ligne sont généralement utiles uniquement si vous disposez d'une variable catégorielle de *colonne*.

Pourcentages en ligne et en colonne de couche : La somme des pourcentages en ligne ou en colonne (pour les pourcentages simples) est égale à 100 % sur l'ensemble des sous-tables d'un tableau imbriqué. Si le tableau contient des couches, la somme des pourcentages en ligne ou en colonne est égale à 100 % sur l'ensemble des sous-tables imbriqués dans chaque couche.

Pourcentages en couche : Pourcentages dans chaque couche. Pour les pourcentages simples, la somme des pourcentages en cellule au sein de la couche actuellement visible est égale à 100 %. Si vous ne disposez d'aucune variable de couche, il s'agit de l'équivalent des pourcentages en tableau.

Pourcentages en tableau : Les pourcentages pour chaque cellule sont basés sur l'intégralité du tableau. Tous les pourcentages en cellule sont basés sur le même nombre total d'observations et leur somme est égale à 100 % (pour les pourcentages simples) dans l'intégralité du tableau.

Pourcentages en sous-table : Les pourcentages dans chaque cellule sont basés sur la sous-table. Tous les pourcentages en cellule de la sous-table sont basés sur le même nombre total d'observations et leur somme est égale à 100 % dans la sous-table. Dans les tableaux imbriqués, la variable qui précède le niveau d'imbrication le plus interne définit les sous-tables. Par exemple, dans un tableau dans lequel la variable *Situation familiale* est imbriquée dans la variable *Sexe*, elle-même imbriquée dans la variable *Tranche d'âge*, la variable *Sexe* définit les sous-tables.

Les pourcentages des jeux de réponses multiples peuvent être basés sur des observations, des réponses ou des occurrences. Pour plus d'informations, voir «Statistiques récapitulatives pour jeux de réponses multiples».

Tables empilées

Pour les calculs de pourcentage, chaque section de tableau définie par une variable d'empilement est traitée comme un tableau distinct. La somme des pourcentages en ligne de couche, en colonne de couche et en tableau est égale à 100 % (pour les pourcentages simples) dans chaque section du tableau empilé. La base de pourcentage pour différents calculs se fonde sur les observations dans chaque section du tableau empilé.

Base de pourcentage

Il est possible de calculer les pourcentages de trois façons différentes, selon le traitement des valeurs manquantes dans la base de calcul :

Pourcentage simple : Les pourcentages sont basés sur le nombre d'observations utilisées dans le tableau et leur somme est toujours égale à 100 %. Si une catégorie est exclue du tableau, les observations de cette catégorie sont exclues de la base. Les observations avec valeurs système manquantes sont toujours exclues de la base. Les observations avec valeurs manquantes de l'utilisateur sont exclues si des catégories manquantes de l'utilisateur sont exclues du tableau (par défaut) ou incluses si des catégories manquantes de l'utilisateur sont incluses dans le tableau. Tout pourcentage dont le nom ne comporte pas la mention *N valide* ou *N total* est un pourcentage simple.

Pourcentage N total : Les observations avec valeurs système manquantes et valeurs manquantes de l'utilisateur sont ajoutées à la base de pourcentage simple. La somme des pourcentages peut être inférieure à 100 %.

Pourcentage N valide : Les observations avec valeurs manquantes de l'utilisateur sont supprimées de la base de pourcentage simple, même si des catégories manquantes de l'utilisateur sont incluses dans le tableau.

Remarque : Les observations des catégories exclues manuellement, autres que les catégories manquantes spécifiées par l'utilisateur, sont toujours exclues de la base.

Statistiques récapitulatives pour jeux de réponses multiples

Les statistiques récapitulatives supplémentaires suivantes sont disponibles pour les jeux de réponses multiples.

% réponses colonne/ligne/strate : Pourcentage basé sur les réponses.

% réponses colonne/ligne/strate (base : Effectif) : Les réponses sont le numérateur et le total est le dénominateur.

Effectif % de colonne/ligne/strate (base : Réponses) : L'effectif est le numérateur et le total des réponses le dénominateur.

% réponses colonne/ligne de strate : Pourcentage dans les sous-tables. Pourcentage basé sur les réponses.

% réponses colonne/ligne/strate (base : Effectif) : Pourcentages dans les sous-tables. Les réponses sont le numérateur et le total est le dénominateur.

% réponses colonne/ligne de strate (base : Réponses) : Pourcentages dans les sous-tables. L'effectif est le numérateur et le total des réponses le dénominateur.

Réponses : Effectif des réponses.

% réponses table/sous-table : Pourcentage basé sur les réponses.

% réponses table/sous-table (base : Effectif) : Les réponses sont le numérateur et le total est le dénominateur.

% réponses table/sous-table (base : Réponses) : L'effectif est le numérateur et le total des réponses le dénominateur.

Statistiques récapitulatives pour les variables d'échelle et autres statistiques récapitulatives en total pour les variables catégorielles

Outre les effectifs et les pourcentages disponibles pour les variables catégorielles, les statistiques récapitulatives suivantes sont disponibles pour les variables d'échelle, et sont également disponibles en tant qu'autres statistiques récapitulatives en total et sous-total pour les variables catégorielles. Ces statistiques récapitulatives ne sont pas disponibles pour les jeux de réponses multiples ou les variables de chaîne (alphanumériques).

Moyenne : Moyenne arithmétique ; somme divisée par le nombre d'observations.

Médiane : Valeur au-dessus ou au-dessous de laquelle se trouve la moitié des observations ; 50e percentile.

Mode : Valeur la plus fréquente. Si deux valeurs sont à égalité, la plus petite valeur est affichée.

Minimum : Plus petite valeur (valeur la moins élevée).

Maximum : Plus grande valeur (valeur la plus élevée).

Manquant : Effectif des valeurs manquantes (valeurs système manquantes et valeurs manquantes spécifiées par l'utilisateur).

Percentile : Vous pouvez inclure le 5e, le 25e, le 75e, le 95e et/ou le 99e percentile.

Plage : Différence entre les valeurs minimale et maximale.

Erreur standard de la moyenne : Mesure du taux de variation de la valeur de la moyenne sur des échantillons provenant de la 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).

Ecart 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 l'âge moyen est de 45 ans, avec un écart-type de 10, 95 % des observations se situent entre 25 et 65 dans une distribution normale (racine carrée de la variance).

Somme : Somme des valeurs.

Pourcentage de somme : Pourcentages basés sur des sommes. Disponible pour les lignes et colonnes (au sein de sous-tables), l'intégralité des lignes et colonnes (parmi les sous-tables), les couches, les sous-tables et l'intégralité des tableaux.

N total : Effectif des valeurs non manquantes, des valeurs manquantes de l'utilisateur et des valeurs système manquantes. N'inclut pas les observations des catégories exclues manuellement autres que les catégories manquantes spécifiées par l'utilisateur.

N valide : Effectif des valeurs non manquantes. N'inclut pas les observations des catégories exclues manuellement autres que les catégories manquantes spécifiées par l'utilisateur.

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 est mesurée en unités qui représentent le carré des unités de la variable (le carré de l'écart-type).

Tableaux empilés

Chaque section de tableau définie par une variable d'empilement est traitée comme un tableau distinct et les statistiques récapitulatives sont calculées en conséquence.

Autres statistiques récapitulatives en total pour les variables catégorielles

En ce qui concerne les tableaux de variables catégorielles qui contiennent des totaux et sous-totaux, vous pouvez obtenir des statistiques récapitulatives différentes des récapitulatifs affichés pour chaque catégorie. Vous pouvez, par exemple, afficher les effectifs et les pourcentages en colonne pour une variable de ligne catégorielle ordinale, et afficher la médiane de la statistique de total.

Pour créer le tableau d'une variable catégorielle avec une autre statistique récapitulative en total :

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
 Le générateur de tableaux s'ouvre.
2. Faites glisser une variable catégorielle vers la zone Lignes ou Colonnes du canevas.
3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur le canevas et sélectionnez **Catégories et totaux** dans le menu contextuel.
4. Cochez la case **Total**, puis cliquez sur **Appliquer**.
5. Dans le panneau de canevas, cliquez de nouveau avec le bouton droit de la souris sur la variable et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
6. Cochez la case **Personnaliser les statistiques récapitulatives pour les totaux et les sous-totaux**, puis sélectionnez les autres statistiques récapitulatives de votre choix.

Par défaut, toutes les statistiques récapitulatives, y compris les autres statistiques, apparaissent dans la dimension opposée à celle contenant la variable catégorielle. Par exemple, si vous disposez d'une variable de ligne catégorielle, les statistiques récapitulatives définissent des colonnes dans le tableau :

Tableau 5. Catégories de variables ordinales sur les lignes, effectifs et moyenne des statistiques récapitulatives sur les colonnes.

Variabes	Catégories	Effectif	Moyenne
Variable 1	1 Accord	196	

Tableau 5. Catégories de variables ordinales sur les lignes, effectifs et moyenne des statistiques récapitulatives sur les colonnes (suite).

Variabes	Catégories	Effectif	Moyenne
	2 Neutre	936	
	3 Pas d'accord	744	
	Total	1 876	2,29

Pour afficher les statistiques récapitulatives dans la même dimension que la variable catégorielle :

7. Dans l'onglet Tableau du générateur de tableaux, dans le groupe Statistiques récapitulatives, sélectionnez la dimension dans la liste déroulante Position.

Par exemple, si la variable catégorielle est affichée en lignes, sélectionnez **Lignes** dans la liste déroulante.

Formats d'affichage des statistiques récapitulatives

Les options de format d'affichage suivantes sont disponibles :

nnnn : Format numérique simple.

nnnn% : Le signe % est ajouté à la fin de la valeur.

Auto : Format d'affichage de variable défini, y compris le nombre de décimales.

N=nnnn : Affiche N= devant la valeur. Ceci peut être utile pour les effectifs, le N valide et le N total des tableaux dans lesquels les libellés des statistiques récapitulatives ne sont pas affichés.

(nnnn) : Toutes les valeurs sont mises entre parenthèses.

(nnnn)(valeur nég.) : Seules les valeurs négatives sont mises entre parenthèses.

(nnnn%) : Toutes les valeurs sont mises entre parenthèses et le signe % est ajouté à la fin des valeurs.

n,nnn.n : Format virgule. La virgule est utilisée comme séparateur de groupe et le point comme indicateur décimal, quels que soient les paramètres régionaux.

n.nnn,n : Format point. Le point est utilisé comme séparateur de groupe et la virgule comme indicateur décimal, quels que soient les paramètres régionaux.

\$n,nnn.n : Format dollar. Le signe dollar est affiché devant la valeur ; la virgule est utilisée comme séparateur de groupe et le point comme indicateur décimal, quels que soient les paramètres régionaux.

CCA, CCB, CCC, CCD, CCE : Formats de devise personnalisés. Le format actuellement défini pour chaque devise est affiché dans la liste. Ces formats sont définis dans l'onglet Devise de la boîte de dialogue Options (menu Edition, Options).

Règles générales et limites

- Sauf pour le format Auto, le nombre de décimales est déterminé par le paramètre de la colonne Décimales.
- Sauf pour les formats virgule, dollar et point, l'indicateur décimal utilisé est celui défini pour les paramètres régionaux dans la zone Options régionales du Panneau de configuration Windows.
- Avec les formats virgule, dollar et point, le séparateur de groupe est soit une virgule, soit un point. Cependant, aucun format d'affichage n'est disponible lors de la création pour afficher un séparateur de groupe sur la base des paramètres régionaux en cours (définis dans la zone Options régionales du Panneau de configuration Windows).

Catégories et totaux

La boîte de dialogue Catégories et totaux vous permet d'effectuer les opérations suivantes :

- Réorganiser et exclure des catégories.
- Insérer des sous-totaux et des totaux.
- Insérer des catégories calculées
- Inclure ou exclure des catégories vides.
- Inclure ou exclure des catégories définies comme contenant des valeurs manquantes.
- Inclure ou exclure des catégories qui ne présentent aucun libellé de valeur défini.
- Cette boîte de dialogue est disponible uniquement pour les variables catégorielles et les jeux de réponses multiples. Elle n'est pas disponible pour les variables d'échelle.
- Dans le cas de la sélection de plusieurs variables avec différentes catégories, vous ne pouvez pas insérer des sous-totaux, ni des catégories calculées, exclure des catégories ou réorganiser manuellement les catégories. Ceci se produit uniquement si vous sélectionnez plusieurs variables dans l'aperçu du canevas et accédez à cette boîte de dialogue pour toutes les variables sélectionnées simultanément. Vous pouvez toujours exécuter ces actions séparément pour chaque variable.
- En ce qui concerne les variables ne présentant aucun libellé de valeur défini, vous pouvez uniquement trier les catégories et insérer des totaux.

Pour accéder à la boîte de dialogue Catégories et totaux

1. Faites glisser une variable catégorielle ou un jeu de réponses multiples vers le panneau de canevas.
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur la variable et choisissez **Catégories et totaux** dans le menu contextuel
ou
3. Dans le panneau de canevas, sélectionnez la variable et cliquez sur **Catégories et totaux** dans le groupe Définir de l'onglet Tableau.
Vous pouvez également, dans le panneau de canevas, sélectionner plusieurs variables catégorielles dans la même dimension :
4. Dans panneau de canevas, cliquez sur chaque variable tout en maintenant la touche Ctrl enfoncée.
ou
5. Cliquez hors de l'aperçu du tableau dans le panneau de canevas, puis cliquez et faites glisser le pointeur pour sélectionner la zone incluant les variables à sélectionner.
ou
6. Cliquez avec le bouton droit de la souris sur une variable dans une dimension et choisissez **Sélectionner toutes les variables de la dimension** pour sélectionner toutes les variables de cette dimension.

Pour réorganiser les catégories

Pour réorganiser manuellement les catégories :

1. Sélectionnez une catégorie dans la liste (cliquez dessus).
2. Cliquez sur la flèche Haut ou Bas pour déplacer la catégorie vers le haut ou vers le bas dans la liste.
ou
3. Cliquez dans la colonne Valeur(s) correspondant à la catégorie, puis faites glisser cette dernière vers un autre emplacement.

Pour exclure des catégories

1. Sélectionnez une catégorie dans la liste (cliquez dessus).
2. Cliquez sur la flèche en regard de la liste Exclure.

ou

3. Cliquez dans la colonne Valeur(s) correspondant à la catégorie, puis faites glisser cette dernière hors de la liste.

Si vous excluez des catégories, toutes les catégories sans libellé de valeur défini seront également exclues.

Pour trier les catégories

Vous pouvez trier les catégories par valeur de donnée, libellé de valeur, effectif de cellule ou statistiques récapitulatives, dans l'ordre croissant ou décroissant.

1. Dans le groupe Tri des catégories, cliquez sur la liste déroulante Par et sélectionnez le critère de tri à utiliser : valeur, libellé, effectif ou statistique récapitulative (par exemple, la moyenne, la médiane ou le mode). Les statistiques récapitulatives disponibles pour le tri dépendent des statistiques récapitulatives que vous avez choisi d'afficher dans le tableau.
2. Cliquez sur la liste déroulante Tri pour sélectionner l'ordre de tri (croissant ou décroissant).

Le tri des catégories n'est pas disponible si vous avez exclu des catégories.

Sous-totaux

1. Dans la liste, sélectionnez la dernière catégorie de la plage à inclure dans le sous-total.
2. Cliquez sur **Ajouter un sous-total...**
3. Dans la boîte de dialogue Définir le sous-total, modifiez le texte de libellé du sous-total, si besoin.
4. Pour afficher uniquement un sous-total et supprimer l'affichage des catégories définissant le sous-total, sélectionnez **Masquer les catégories des sous-totaux dans le tableau**.
5. Cliquez sur **Poursuivre**, pour ajouter le sous-total.

Totaux

1. Cochez la case **Total**. Vous pouvez également modifier le texte du libellé du total.

Si la variable sélectionnée est imbriquée dans une autre, des totaux seront insérés pour chaque sous-table.

Position d'affichage pour les totaux et sous-totaux

Les totaux et sous-totaux peuvent être affichés au-dessus ou au-dessous des catégories incluses dans chaque total.

- Si l'option **En dessous** est sélectionnée dans le groupe Affichage des totaux et des sous-totaux, les totaux apparaissent au-dessus de chaque sous-table, et toutes les catégories situées au-dessus, y compris la catégorie sélectionnée (mais au-dessous des sous-totaux qui précèdent), sont incluses dans chaque sous-total.
- Si l'option **Au-dessus** est sélectionnée dans le groupe Affichage des totaux et des sous-totaux, les totaux apparaissent au-dessous de chaque sous-table, et toutes les catégories situées au-dessous, y compris la catégorie sélectionnée (mais au-dessus des sous-totaux qui précèdent), sont incluses dans chaque sous-total.

Important : Vous devez sélectionner la position d'affichage des sous-totaux avant de définir ces derniers. La modification de la position d'affichage affecte tous les sous-totaux (et non pas uniquement le sous-total sélectionné) ; en outre, elle *modifie les catégories incluses dans les sous-totaux*.

Catégories calculées

Les catégories calculées peuvent être affichées depuis les statistiques récapitulatives; les totaux, les sous-totaux et/ou les constantes. Pour plus d'informations, voir «Catégories calculées», à la page 14.

Autres statistiques récapitulatives en total et sous-total

Vous pouvez, à l'aide de la boîte de dialogue Statistiques récapitulatives, afficher des statistiques autres que des totaux dans les zones Totaux et Sous-totaux du tableau. Pour plus d'informations, voir «Statistiques récapitulatives pour les variables catégorielles», à la page 7.

Remarque : Si vous sélectionnez des statistiques totales personnalisées multiples qui sont également dans le corps du tableau et que vous masquez les libellés des statistiques, alors les totaux seront retriés dans le même ordre que dans le corps du tableau - et puisque les libellés ne sont pas affichés, vous ne pouvez pas savoir ce que chaque statistique totale représente. En général, le choix de statistiques multiples et le fait de masquer les libellés des statistiques ne sont pas conseillés.

Totaux, sous-totaux et catégories exclues

Les observations des catégories exclues ne sont pas incluses dans le calcul des totaux.

Valeurs manquantes, catégories vides et valeurs sans libellés de valeur

Valeurs manquantes : Cette option contrôle l'affichage des valeurs **manquantes spécifiées par l'utilisateur** ou des valeurs définies comme contenant des valeurs manquantes (par exemple, le code 99 représente la réponse "non applicable" pour la grossesse chez les hommes). Par défaut, les valeurs manquantes de l'utilisateur sont exclues. Sélectionnez cette option pour inclure, dans les tableaux, les catégories manquantes spécifiées par l'utilisateur. Bien que la variable puisse contenir plusieurs catégories avec valeurs manquantes, l'aperçu du tableau dans le canevas affiche uniquement une catégorie générique avec valeurs manquantes. Toutes les catégories manquantes spécifiées par l'utilisateur et définies seront incluses dans le tableau. Les **valeurs système manquantes** (cellules vides pour les variables numériques dans l'éditeur de données) sont toujours exclues.

catégories vides : Les catégories vides sont des catégories avec libellés de valeurs définis, mais qui ne comprennent aucune observation pour un tableau ou une sous-table donnée. Par défaut, les catégories vides sont incluses dans les tableaux. Désélectionnez cette option pour exclure les catégories manquantes du tableau.

Autres valeurs lorsque les données sont analysées : Par défaut, les valeurs de catégorie du fichier de données qui ne présentent aucun libellé de valeur défini sont automatiquement incluses dans les tableaux. Désélectionnez cette option pour exclure du tableau les valeurs sans libellé de valeur défini. Si vous excluez des catégories présentant des libellés de valeurs définis, les catégories sans libellé de valeur défini sont également exclues.

Catégories calculées

En plus d'afficher les résultats agrégés des statistiques récapitulatives, un tableau peut afficher une ou plusieurs catégories calculées de ces résultats agrégés, à partir de valeurs constantes, de totaux et sous-totaux ou d'une combinaison de ces derniers. Les résultats sont affichés sous la forme de catégories ou de post-calculs. Une catégorie calculée agit comme une catégorie dans une variable unique, et présente les similarités et différences suivantes :

- Une catégorie calculée est positionnée comme les autres catégories.
- Une catégorie calculée utilise les mêmes statistiques que les autres catégories.
- Les catégories calculées n'ont pas d'incidence sur les sous-totaux, totaux ou les tests de signification.
- Par défaut, les valeurs des catégories calculées utilisent le même format de statistiques récapitulatives que les autres catégories. Ce format peut être supprimé et remplacé lors de la définition de la catégorie calculée.

Les catégories calculées étant similaires aux sous-totaux, elles peuvent être utilisées pour calculer le total des résultats agrégés. Toutefois, elles présentent les avantages suivants par rapport aux sous-totaux :

- Les catégories calculées peuvent être calculées à partir des résultats d'autres sous-totaux.
- Elles peuvent se chevaucher entre elles, et opérer sur les mêmes (ou une partie des mêmes) catégories.
- Les catégories calculées ne doivent pas inclure de valeurs provenant des autres catégories supérieures ou inférieures à la catégorie calculée. Cela signifie que les catégories calculées ne sont pas exhaustives.
- Les catégories calculées peuvent inclure des valeurs provenant de catégories non-adjacentes.

A la différence des totaux et des sous-totaux, le calcul des catégories calculées est effectué à partir des données agrégées et non à partir des données d'origine. Ainsi, il est possible que les valeurs des catégories calculées ne correspondent pas aux résultats des totaux et des sous-totaux. En outre, à cause de l'option qui permet de masquer les catégories source lors de la définition de la catégorie calculée, l'interprétation des sous-totaux peut être rendue difficile dans le tableau des résultats. Si vous utilisez des catégories calculées, nous vous recommandons de spécifier des libellés personnalisés pour les sous-totaux.

Pour définir une catégorie calculée

Les catégories calculées sont ajoutées dans la boîte de dialogue Catégories et totaux. Pour en savoir plus sur la manière d'accéder à cette boîte de dialogue, voir «Catégories et totaux», à la page 12.

1. Dans la boîte de dialogue Catégories et totaux, sélectionnez **Ajouter une catégorie...**
2. Dans le champ **Libellé de la catégorie calculée**, spécifiez un libellé pour la catégorie calculée. Vous pouvez faire glisser des catégories depuis la liste Catégories et inclure les libellés de ces catégories.
3. Construisez une expression en sélectionnant des catégories et/ou des totaux ou des sous-totaux et en utilisant des opérateurs pour définir les catégories calculées. Vous pouvez aussi saisir des valeurs constantes (par exemple, 500) à inclure dans l'expression.
4. Pour afficher uniquement une catégorie calculée et supprimer l'affichage des catégories la définissant, sélectionnez **Masquer les catégories utilisées dans l'expression dans le tableau**.
5. Cliquez sur l'onglet **Formats d'affichage** pour modifier le format d'affichage et le nombre de décimales de la catégorie calculée. Pour plus d'informations, voir «Formats d'affichage des catégories calculées».
6. Cliquez sur **Poursuivre**, pour ajouter la catégorie calculée.

Formats d'affichage des catégories calculées

Par défaut, une catégorie calculée utilise le même format d'affichage et le même nombre de décimales que les autres catégories de la variable. Vous pouvez les annuler dans l'onglet Formats d'affichage de la boîte de dialogue Catégorie calculée. L'onglet Formats d'affichage répertorie les statistiques récapitulatives actuelles sur lesquelles est basée la catégorie calculée en plus des formats d'affichage et du nombre de décimales de ces statistiques.

Pour chaque statistique récapitulative, vous pouvez :

1. Sélectionner un format d'affichage pour la statistique récapitulative dans la liste déroulante Format. Pour obtenir une liste complète des formats d'affichage, voir «Formats d'affichage des statistiques récapitulatives», à la page 11.
2. Entrer le nombre de décimales à afficher dans la cellule Décimales pour la statistique récapitulative sélectionnée.

Tableaux de variables avec catégories partagées (Comperimeter Tables)

Les enquêtes contiennent souvent de nombreuses questions caractérisées par un ensemble commun de réponses possibles. Vous pouvez utiliser la technique de l'empilement pour afficher ces variables liées dans le même tableau et afficher les catégories de réponses partagées dans les colonnes du tableau.

Pour créer un tableau pour plusieurs variables avec catégories partagées

Tableau 6. Variables empilées avec catégories de réponses partagées dans les colonnes

Variabes	Catégorie 1	Catégorie 2	Catégorie 3
Variable 1	12	34	56
Variable 2	56	12	34
Variable 3	34	56	12

Pour plus d'informations, voir Chapitre 6, «Tableaux pour variables avec catégories partagées», à la page 49.

Personnalisation du générateur de tableaux

Contrairement aux boîtes de dialogue standard, vous pouvez modifier la taille du générateur de tableaux de la même manière que vous pouvez modifier la taille d'une fenêtre classique :

1. Cliquez sur un des bords ou des coins du générateur de tableaux, puis faites-le glisser pour augmenter ou diminuer sa taille.
Dans l'onglet Tableau, vous pouvez également modifier la taille de la liste des variables, de la liste Catégories et du panneau de canevas.
2. Cliquez sur la barre horizontale située entre la liste des variables et la liste Catégories, puis faites-la glisser pour agrandir ou réduire les listes. Le fait de la déplacer vers le bas agrandit la liste des variables et réduit la liste Catégories. Le fait de la déplacer vers le haut produit l'effet inverse.
3. Cliquez sur la barre verticale qui sépare la liste des variables et la liste Catégories du panneau de canevas, puis faites-la glisser afin d'élargir ou de rétrécir les listes. Le canevas est automatiquement redimensionné pour s'adapter à l'espace restant.

Tableaux personnalisés : onglet Options

L'onglet Options vous permet d'effectuer les opérations suivantes :

- Spécifier ce qui doit apparaître dans les cellules vides et les cellules pour lesquelles aucune statistique ne peut être calculée.
- Contrôler le traitement des valeurs manquantes dans le calcul des statistiques de variable d'échelle.
- Définir la largeur minimum et/ou maximum des colonnes de données.
- Contrôler le traitement des réponses en double dans les jeux de catégories multiples.

Apparence des cellules de données : Contrôlez ce qui doit apparaître dans les cellules vides et les cellules pour lesquelles aucune statistique ne peut être calculée.

- **Cellules vides :** Pour les cellules du tableau qui ne contiennent aucune observation (effectif de cellule 0), vous pouvez sélectionner l'une des trois options d'affichage suivantes : zéro, blanc ou une valeur de texte que vous indiquez. La valeur de texte peut comporter jusqu'à 255 caractères.
- **Statistiques impossibles à calculer :** Texte affiché si une statistique ne peut pas être calculée (par exemple, la moyenne pour une catégorie sans observation). La valeur de texte peut comporter jusqu'à 255 caractères. La valeur par défaut est un point (.).

Largeur des colonnes de données : Contrôlez la largeur maximum et minimum des colonnes de données. Ce paramètre n'affecte pas la largeur des colonnes pour les libellés de ligne.

- **Paramètres du modèle de tableau :** Utilisez la spécification de la largeur des colonnes de données du modèle de tableau par défaut en cours. Vous pouvez créer votre propre modèle de tableau par défaut personnalisé qui sera utilisé lors de la création de tableaux, et vous pouvez contrôler, à l'aide d'un modèle de tableau, la largeur des colonnes de données et de libellés de ligne.
- **Personnalisée :** Remplacez les paramètres du modèle de tableau par défaut pour la largeur des colonnes de données. Spécifiez la largeur maximum et minimum des colonnes de données pour le tableau, ainsi que l'unité de mesure (points, pouces ou centimètres).

Valeurs manquantes pour les variables d'échelle : Pour les tableaux comportant au moins deux variables d'échelle, cette option contrôle le traitement des données manquantes pour les statistiques de variable d'échelle.

- **Optimiser l'utilisation des données disponibles (suppression variable par variable) :** Toutes les observations avec valeurs valides pour chaque variable d'échelle sont incluses dans les statistiques récapitulatives pour cette variable d'échelle.
- **Utiliser une base d'observations cohérente d'une variable d'échelle à l'autre (suppression liste par liste) :** Les observations avec valeurs manquantes pour toute variable d'échelle dans le tableau sont exclues des statistiques récapitulatives pour toutes les variables d'échelle dans le tableau.

Dénombrer les réponses en double pour les jeux de catégories multiples : Une réponse en double est la même réponse pour au moins deux variables dans le jeu de catégories multiples. Par défaut, les réponses en double ne sont pas comptées, mais vous pouvez parfaitement les inclure dans l'effectif (comme un jeu de catégories multiples représentant le fabricant des trois dernières voitures achetées par un répondant).

Masquer les petits effectifs : Vous pouvez choisir de masquer les effectifs inférieurs à un entier spécifique. Les valeurs masquées seront affichées en tant que $<N$, où N est l'entier spécifié. L'entier spécifié doit être supérieur ou égal à 2.

Tableaux personnalisés : onglet Titres

L'onglet Titres contrôle l'affichage des titres, légendes et libellés de coins.

Titre : Texte affiché au-dessus du tableau.

Légende : Texte affiché au-dessous du tableau et au-dessus des notes de bas de page.

Coin : Texte affiché dans le coin supérieur gauche du tableau. Le texte de coin est affiché uniquement si le tableau contient des variables en ligne et si la propriété du libellé de dimension de ligne du tableau croisé dynamique est paramétrée sur **Imbriquées**. Il ne s'agit *pas* du paramètre de modèle de tableau par défaut.

Vous pouvez inclure les valeurs suivantes, générées automatiquement, dans le titre tableau, la légende ou le libellé de coin du tableau :

Date : Année, mois et jour en cours, affichés au format basé sur le paramétrage de vos options régionales Windows.

Heure : Heure, minutes et secondes en cours, au format basé sur le paramétrage de vos options régionales Windows.

Expression de table : Variables utilisées et façon dont elles sont utilisées dans le tableau. Si une variable présente un libellé de variable définie, le libellé est affiché. Dans le tableau généré, les symboles suivants indiquent comment les variables sont utilisées :

- + indique des variables empilées.
- > indique un imbrication.
- BY indique un tableau croisé ou des couches.

Tableaux personnalisés : onglet Tests statistiques

L'onglet Statistiques de test vous permet de demander divers tests de signification pour vos tableaux personnalisés, notamment :

- Tests du khi-deux pour tester l'indépendance.
- Tests d'égalité des moyennes de colonnes.

- Tests d'égalité des proportions des colonnes.
- Tests de signification pour les jeux de réponses multiples et les sous-totaux. (Pour obtenir des informations sur le test de signification des jeux de réponses multiples, consultez «Test de signification pour les jeux de réponses multiples», à la page 82.)

Ces tests ne sont pas disponibles pour les tableaux dans lesquels les libellés de catégorie sont déplacés hors de leur dimension de tableau par défaut ou pour les catégories calculées.

Comparer les moyennes de colonne (tests-t) : Cette option produit des tests par paire de l'égalité des moyennes de colonnes pour les tableaux dans lesquels au moins une variable catégorielle existe dans les colonnes et au moins une variable d'échelle existe dans les lignes. Le tableau doit inclure la moyenne en tant que statistique récapitulative. Vous pouvez indiquer si les valeurs p des tests sont ajustées à l'aide de la méthode Bonferroni. Vous pouvez également spécifier le niveau alpha du test qui doit être une valeur supérieure à 0 et inférieure à 1. Enfin, alors que la variance pour le test de moyennes est toujours basée uniquement sur les catégories comparées pour les tests de multiréponses, elle peut être estimée uniquement à partir des catégories comparées ou de toutes les catégories pour les variables de catégories ordinaires.

Comparer les proportions de colonne (tests-z) : Cette option produit des tests par paire de l'égalité des proportions des colonnes pour les tableaux dans lesquels au moins une variable catégorielle existe à la fois dans les lignes et les colonnes. Le tableau doit inclure les effectifs ou les pourcentages simples en colonne. Vous pouvez indiquer si les valeurs p des tests sont ajustées à l'aide de la méthode Bonferroni. Vous pouvez également indiquer le niveau alpha du test, qui doit être une valeur supérieure à 0 et inférieure à 1.

Identifier les différences significatives : Si vous sélectionnez **Comparer les moyennes de colonne** ou **Comparer les proportions de colonne**, vous pouvez choisir la manière d'indiquer des différences significatives.

- **Dans un tableau séparé :** Les résultats des tests de signification sont affichés dans un tableau distinct. Si deux valeurs sont différentes de manière significative, la cellule correspondant à la valeur la plus élevée affiche une clé identifiant la colonne de la valeur la plus basse. En voici un exemple.
- **Dans le tableau principal à l'aide des indices de style APA :** Le tableau principal identifie les différences significatives dans un format de style APA à l'aide d'indices. Si deux valeurs sont différentes de manière significative, elles sont affichées avec des indices *différents*. Ces indices ne sont pas des notes de bas de page. Si cette option est activée, le style de notes en bas de page du modèle de tableau actuel est annulé et les notes en bas de page seront affichées sous la forme d'exposants. En voici un exemple.

Pour voir l'exemple complet sur la façon de créer et d'interpréter ces tableaux, voir «Comparaison des proportions de colonne», à la page 71.

Remarque : Le tableau de style APA comprend une légende qui explique comment interpréter les indices du tableau. Si vous spécifiez une légende sur l'onglet Titres, la légende spécifiée apparaît à la place de la légende d'explication des indices.

Tests d'indépendance (khi-deux) : Cette option produit un test d'indépendance (khi-deux) pour les tableaux dans lesquels au moins une variable catégorielle existe à la fois dans les lignes et les colonnes. Vous pouvez également indiquer le niveau alpha du test, qui doit être une valeur supérieure à 0 et inférieure à 1.

Utiliser les sous-totaux à la place des catégories de sous-totaux : Si l'option est sélectionnée, chaque sous-total remplace ses catégories pour le test de signification. Dans le cas contraire, seuls les sous-totaux pour lesquels les catégories de sous-totaux sont masquées remplacent leurs catégories pour le test.

Inclure les variables multiréponses : Si l'option est sélectionnée, les tests sont réalisés en utilisant les catégories de jeux de réponses multiples. Dans le cas contraire, les jeux de réponses multiples sont ignorés lors de la réalisation des tests.

Chapitre 2. Tableaux simples pour variables catégorielles

Tableaux simples pour variables catégorielles

La plupart des tableaux que vous souhaitez créer incluent probablement au moins une **variable catégorielle**. Une variable catégorielle est une variable ayant un nombre limité de valeurs ou catégories distinctes (par exemple, sexe ou religion). Les variables catégorielles peuvent être **nominales** ou **ordinales**.

- *Nominal*. Une variable peut être traitée comme étant nominale si ses valeurs représentent des catégories sans classement intrinsèque (par exemple, le service de la société dans lequel travaille un employé). La région, le code postal ou l'appartenance religieuse sont des exemples de variables nominales.
- *Ordinal*. Une variable peut être traitée comme étant ordinale si ses valeurs représentent des catégories associées à un classement intrinsèque (par exemple, des niveaux de satisfaction allant de Très mécontent 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.

Dans la liste des variables, une icône indique le type de chaque variable.

Tableau 7. Icônes de niveau de mesure

	Numérique	Chaîne	Date	Heure
Echelle (continue).		n/a		
Ordinal				
Nominal				

Les tableaux personnalisés ont été optimisés pour être utilisés avec les variables catégorielles présentant des **libellés de valeurs** définis. Pour plus d'informations, voir «Construction de tableaux», à la page 1.

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Tous les exemples fournis ici affichent des libellés de variable dans des boîtes de dialogue ; ces libellés sont triés par ordre alphabétique. Les propriétés d'affichage de la liste des variables sont paramétrées dans l'onglet Général de la boîte de dialogue Options (menu Modifier, Options).

Une seule variable catégorielle

Bien qu'un tableau comportant une seule variable catégorielle soit l'un des tableaux les plus simples que vous puissiez créer, il correspondra le plus souvent à ce que vous souhaitez ou à ce dont vous avez besoin.

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.

Un aperçu du tableau apparaît dans le panneau de canevas. L'aperçu n'affiche pas les valeurs de données réelles ; il présente des valeurs de substitution à l'emplacement où des données apparaîtront.

3. Cliquez sur **OK** pour créer le tableau.

Le tableau apparaît dans la fenêtre du visualiseur.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Figure 2. Variable catégorielle unique dans les lignes

Dans ce tableau simple, l'en-tête de colonne *Effectif* n'est pas vraiment nécessaire et vous pouvez créer le tableau sans cet en-tête.

4. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
5. Dans le groupe Statistiques récapitulatives, sélectionnez (cliquez sur) **Masquer**.
6. Cliquez sur **OK** pour créer le tableau.

Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Figure 3. Une seule variable catégorielle sans libellé de colonne de statistiques récapitulatives

Pourcentages

Outre les effectifs, vous pouvez également afficher les pourcentages. Pour un tableau simple composé d'une seule variable catégorielle, si la variable est affichée en lignes, vous souhaitez probablement examiner les pourcentages en colonne. Inversement, pour une variable affichée en colonnes, vous souhaitez probablement examiner les pourcentages en ligne.

1. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le groupe Statistiques récapitulatives, désélectionnez **Masquer** pour Position. Puisque ce tableau présentera deux colonnes, vous souhaitez afficher les libellés de colonne afin de savoir ce que représente chaque colonne.
3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
4. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **N % colonne** dans la liste Statistiques et cliquez sur la flèche pour ajouter l'élément à la liste Afficher.
5. Dans la cellule Libellé de la liste Afficher, supprimez le libellé par défaut et tapez Pourcentage.
6. Dans le générateur de tableaux, cliquez sur **Appliquer à la sélection**, puis sur **OK** pour créer le tableau.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%

Figure 4. Effectifs et pourcentages de colonne

Totaux

Les totaux ne sont pas automatiquement inclus dans les tableaux personnalisés, mais vous pouvez les ajouter facilement.

1. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et sélectionnez **Catégories et totaux** dans le menu contextuel.
3. Dans la boîte de dialogue Catégories et totaux, sélectionnez **Total**.
4. Dans le générateur de tableaux, cliquez sur **Appliquer**, puis sur **OK** pour créer le tableau.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	Total	2828	100.0%

Figure 5. Effectifs, pourcentages de colonne et totaux

Pour plus d'informations, voir Chapitre 4, «Totaux et sous-totaux pour les variables catégorielles», à la page 37.

Tableau croisé

La technique des tableaux croisés est une technique de base qui permet d'examiner la relation entre deux variables catégorielles. Par exemple, en utilisant la variable *Tranche d'âge* comme variable de ligne et la variable *Sexe* comme variable de colonne, vous pouvez créer un tableau croisé bidimensionnel présentant le nombre d'hommes et de femmes dans chaque tranche d'âge.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Cliquez sur **Restaurer** pour supprimer les sélections précédentes dans le générateur de tableaux.
3. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
4. Faites glisser la variable *Sexe* de la liste des variables vers la zone Colonnes du panneau de canevas. (Vous pouvez être amené à parcourir la liste pour accéder à cette variable.)
5. Cliquez sur **OK** pour créer le tableau.

		Gender	
		Male	Female
		Count	Count
Age category	Less than 25	108	134
	25 to 34	276	351
	35 to 44	309	370
	45 to 54	221	260
	55 to 64	136	184
	65 or older	178	301

Figure 6. Tableau croisé avec les variables Tranche d'âge et Sexe

Pourcentages dans des tableaux croisés

Dans un tableau croisé bidimensionnel, les pourcentages en ligne et en colonne peuvent fournir des informations utiles.

- Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
- Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Sexe*.
Vous pouvez remarquer que l'option **Statistiques récapitulatives** est désactivée dans le menu contextuel. En effet, vous ne pouvez sélectionner l'option Statistiques récapitulatives que pour la variable la plus interne dans la dimension source des statistiques. La dimension source par défaut (ligne ou colonne) des statistiques pour les variables catégorielles est basée sur l'ordre dans lequel vous avez glissé les variables vers le panneau de canevas. Dans cet exemple, nous avons d'abord déplacé la variable *Tranche d'âge* vers la dimension de ligne ; étant donné que la dimension de ligne ne comprend pas d'autres variables, la variable *Tranche d'âge* est la variable source des statistiques. Vous pouvez modifier la dimension source des statistiques, mais ceci n'est pas nécessaire dans cet exemple. Pour plus d'informations, voir «Statistiques récapitulatives», à la page 6.
- Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
- Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **N % colonne** dans la liste Statistiques et cliquez sur la flèche pour ajouter l'élément à la liste Afficher.
- Dans la liste Statistiques, sélectionnez **N % ligne** et cliquez sur la flèche pour ajouter cet élément à la liste Afficher.
- Dans le générateur de tableaux, cliquez sur **Appliquer à la sélection** , puis sur **OK** pour créer le tableau.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.8%	44.6%	134	8.4%	55.4%
	25 to 34	276	22.5%	44.0%	351	21.9%	56.0%
	35 to 44	309	25.2%	45.5%	370	23.1%	54.5%
	45 to 54	221	18.0%	45.9%	260	16.3%	54.1%
	55 to 64	136	11.1%	42.5%	184	11.5%	57.5%
	65 or older	178	14.5%	37.2%	301	18.8%	62.8%

Figure 7. Tableau croisé avec pourcentages de ligne et de colonne

Contrôle du format d'affichage

Vous pouvez contrôler le format d'affichage, y compris le nombre de décimales affichées dans les statistiques récapitulatives. Par exemple, les pourcentages sont affichés par défaut avec une décimale et le signe %. Si vous souhaitez que les valeurs des cellules affichent deux décimales sans signe de pourcentage, procédez de la façon suivante.

- Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).

2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
3. Pour les deux statistiques récapitulatives de pourcentage sélectionnées (**N % colonne** et **N % ligne**), sélectionnez **nnnn,n** dans la liste déroulante Format et tapez 2 dans la cellule Décimales pour les deux statistiques.
4. Cliquez sur **OK** pour créer le tableau.

		Gender					
		Male			Female		
		Count	Column N %	Row N %	Count	Column N %	Row N %
Age category	Less than 25	108	8.79	44.63	134	8.38	55.37
	25 to 34	276	22.48	44.02	351	21.94	55.98
	35 to 44	309	25.16	45.51	370	23.13	54.49
	45 to 54	221	18.00	45.95	260	16.25	54.05
	55 to 64	136	11.07	42.50	184	11.50	57.50
	65 or older	178	14.50	37.16	301	18.81	62.84

Figure 8. Affichage de cellule formaté pour les pourcentages de ligne et de colonne

Totaux marginaux

Dans les tableaux croisés, il est très courant d'afficher des **totaux marginaux**, c'est-à-dire des totaux pour chaque ligne et colonne. Etant donné que ces derniers ne sont pas inclus par défaut dans les tableaux personnalisés, vous devez les ajouter explicitement aux tableaux.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Cliquez sur **Restaurer** pour supprimer les sélections précédentes dans le générateur de tableaux.
3. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
4. Faites glisser la variable *Sexe* de la liste des variables vers la zone Colonnes du panneau de canevas. (Vous pouvez être amené à parcourir la liste pour accéder à cette variable.)
5. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et sélectionnez **Catégories et totaux** dans le menu contextuel.
6. Dans la boîte de dialogue Catégories et totaux, sélectionnez **Total**, puis cliquez sur **Appliquer**.
7. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Sexe* et choisissez **Catégories et totaux** dans le menu contextuel.
8. Dans la boîte de dialogue Catégories et totaux, sélectionnez **Total**, puis cliquez sur **Appliquer**.
9. Dans le groupe Statistiques récapitulatives, sélectionnez (cliquez sur) **Masquer**. (Puisque vous affichez uniquement les effectifs, vous n'avez pas besoin d'identifier la « statistique » affichée dans les cellules de données du tableau.)
10. Cliquez sur **OK** pour créer le tableau.

		Gender		
		Male	Female	Total
Age category	Less than 25	108	134	242
	25 to 34	276	351	627
	35 to 44	309	370	679
	45 to 54	221	260	481
	55 to 64	136	184	320
	65 or older	178	301	479
	Total	1228	1600	2828

Figure 9. Tableau croisé avec totaux marginaux

Tri et exclusion de catégories

Par défaut, les catégories apparaissent dans l'ordre croissant des valeurs de donnée représentées par les libellés de valeurs des catégories. Par exemple, bien que les libellés de valeurs de *Moins de 25*, *Entre 25 et 34*, *Entre 35 et 44* etc. apparaissent pour les tranches d'âge, les valeurs de données sous-jacentes réelles sont 1, 2, 3, etc. et contrôlent l'ordre d'affichage par défaut des catégories.

Vous pouvez facilement modifier l'ordre des catégories et exclure des catégories que vous ne souhaitez pas voir apparaître dans le tableau.

Tri des catégories

Vous pouvez réorganiser manuellement ou trier des catégories dans l'ordre croissant ou décroissant pour :

- Les valeurs de données
 - Libellés de valeurs.
 - Les effectifs de cellule
 - Statistiques récapitulatives : Les statistiques récapitulatives disponibles pour le tri dépendent des statistiques récapitulatives que vous avez choisi d'afficher dans le tableau.
1. Ouvrez le générateur de tableaux (menu *Analyse*, *Tableaux*, *Tableaux personnalisés*).
 2. Si la variable *Tranche d'âge* n'est pas affichée dans la zone Lignes du panneau de canevas, faites-la glisser pour l'y amener.
 3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et sélectionnez **Catégories et totaux** dans le menu contextuel.

Les valeurs de données et les libellés de valeurs associés apparaissent dans l'ordre de tri actuel ; dans ce cas, il s'agit toujours de l'ordre croissant des valeurs de données.

4. Dans le groupe Trier les catégories, sélectionnez **Décroissant** dans la liste déroulante *Ordre*. L'ordre de tri est désormais inversé.
5. Dans la liste déroulante *Par*, sélectionnez **Libellés**.

Les catégories sont désormais triées dans l'ordre alphabétique décroissant des libellés de valeurs.

Notez que la catégorie portant le libellé *Moins de 25* est située au sommet de la liste. Dans le tri alphabétique, les lettres arrivent après les nombres. Puisqu'il s'agit de le seul libellé qui commence par une lettre et que la liste est triée dans l'ordre décroissant (inverse), cette catégorie arrive en début de liste.

Si vous souhaitez qu'une catégorie donnée apparaisse à un emplacement différent dans la liste, vous pouvez facilement la déplacer.

6. Dans la liste *Libellé*, cliquez sur la catégorie portant le libellé *Moins de 25*.
7. Cliquez sur la flèche vers le bas, située à droite de la liste. La catégorie descend d'une ligne dans la liste.
8. Continuez à cliquer sur la flèche vers le bas jusqu'à ce que la catégorie se retrouve au bas de la liste.

Exclusion de catégories

Si vous souhaitez que certaines catégories n'apparaissent pas dans le tableau, vous pouvez les exclure.

1. Dans la liste *Libellé*, cliquez sur la catégorie portant le libellé *Moins de 25*.
2. Cliquez sur le bouton flèche situé à gauche de la liste *Exclure*.
3. Dans la liste *Libellé*, cliquez sur la catégorie portant le libellé *65 ans et plus*.
4. Cliquez de nouveau sur le bouton flèche situé à gauche de la liste *Exclure*.

Les deux catégories sont déplacées de la liste *Afficher* vers la liste *Exclure*. Si vous changez d'avis, vous pouvez facilement les renvoyer vers la liste *Afficher*.

5. Dans le générateur de tableaux, cliquez sur **Appliquer** , puis sur **OK** pour créer le tableau.

		Gender		
		Male	Female	Total
Age category	55 to 64	136	184	320
	45 to 54	221	260	481
	35 to 44	309	370	679
	25 to 34	276	351	627
	Total	942	1165	2107

Figure 10. Tableau trié par libellé de valeur (ordre décroissant), avec exclusion de certaines catégories

Notez que les totaux étaient supérieurs avant l'exclusion des deux catégories. En effet, les totaux sont basés sur les catégories incluses dans le tableau. Toute catégorie exclue est éliminée du calcul des totaux. Pour plus d'informations, voir Chapitre 4, «Totaux et sous-totaux pour les variables catégorielles», à la page 37.

Chapitre 3. Empilement, imbrication et couches avec des variables catégorielles

L'empilement, l'imbrication et les couches sont des méthodes permettant l'affichage de plusieurs variables dans le même tableau. Ce chapitre présente l'utilisation de ces techniques avec des variables catégorielles, bien qu'elles puissent également être utilisées avec des variables d'échelle.

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Tous les exemples fournis ici affichent des libellés de variable dans des boîtes de dialogue ; ces libellés sont triés par ordre alphabétique. Les propriétés d'affichage de la liste des variables sont paramétrées dans l'onglet Général de la boîte de dialogue Options (menu Modifier, Options).

Empilement de variables catégorielles

Empiler revient à prendre des tableaux séparés et à les coller ensemble dans le même affichage. Vous pouvez, par exemple, afficher des informations sur les variables *sexe* et *tranche d'âge* dans des sections séparées du même tableau.

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Dans le générateur de tableaux, faites glisser la variable *Sexe* de la liste des variables vers la zone Lignes du panneau de canevas.
3. Faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes sous *Sexe*.
Les deux variables sont désormais empilées dans la dimension de ligne.
4. Cliquez sur **OK** pour créer le tableau.

		Count
Gender	Male	1232
	Female	1600
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

Figure 11. Tableau des variables catégorielles empilées en lignes

Vous pouvez également empiler de la même façon des variables en colonnes.

Empilement et tableau croisé

Un tableau empilé peut inclure d'autres variables dans d'autres dimensions. Par exemple, vous pouvez croiser deux variables empilées dans les lignes avec une troisième variable affichée dans la dimension de colonne.

1. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Si les variables *Tranche d'âge* et *Sexe* ne sont pas encore empilées dans les lignes, suivez les instructions précédentes pour les empiler.

- Faites glisser la variable *Informations sur Internet* de la liste des variables vers la zone Colonnes du panneau de canevas.
- Cliquez sur **OK** pour créer le tableau.

		Get news from internet	
		No	Yes
		Count	Count
Gender	Male	873	359
	Female	1092	508
Age category	Less than 25	146	96
	25 to 34	368	259
	35 to 44	435	244
	45 to 54	346	135
	55 to 64	252	68
	65 or older	416	63

Figure 12. Tableau croisé de deux variables de ligne empilées avec une variable de colonne

Remarque : Plusieurs variables ayant des libellés qui commencent par *Informations...*, il peut s'avérer difficile de les distinguer dans la liste des variables (les libellés peuvent en effet être trop longs pour être affichés complètement dans la liste des variables). Deux méthodes permettent de visualiser l'intégralité du libellé de variable :

- Positionnez le pointeur de la souris sur une variable dans la liste afin d'afficher l'intégralité du libellé dans une infobulle.
- Dans le panneau de canevas, faites glisser la barre verticale qui sépare la liste des variables et la liste des catégories afin d'élargir les listes.

Imbrication de variables catégorielles

L'imbrication, comme la technique des tableaux croisés, permet de présenter la relation entre deux variables catégorielles ; la seule différence est qu'une variable est imbriquée dans l'autre dans la même dimension. Vous pouvez, par exemple, imbriquer la variable *Sexe* dans la variable *Tranche d'âge* dans la dimension de ligne, ce qui a pour effet de présenter le nombre d'hommes et de femmes dans chaque tranche d'âge.

- Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
- Cliquez sur **Restaurer** pour supprimer les sélections précédentes dans le générateur de tableaux.
- Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
- Faites glisser la variable *Sexe* de la liste des variables vers la droite de la variable *Tranche d'âge* dans la zone Lignes.

Dans le panneau de canevas, l'aperçu indique désormais que le tableau imbriqué sera composé d'une seule colonne d'effectifs, chaque cellule contenant le nombre d'hommes et de femmes dans chaque tranche d'âge.

Vous pouvez remarquer que le libellé de variable *Sexe* apparaît plusieurs fois, pour chaque tranche d'âge. Vous pouvez minimiser ce type de répétition en plaçant la variable contenant le plus petit nombre de catégories dans le niveau d'imbrication le plus externe.

- Dans le panneau de canevas, cliquez sur le libellé de variable *Sexe*.
- Faites glisser la variable le plus à gauche possible vers la zone Lignes.
Désormais, au lieu que la variable *Sexe* soit répétée six fois, la variable *Tranche d'âge* est répétée deux fois. Le tableau est alors moins confus et produira essentiellement les mêmes résultats.
- Cliquez sur **OK** pour créer le tableau.

				Count
Gender	Male	Age category	Less than 25	108
			25 to 34	276
			35 to 44	309
			45 to 54	221
			55 to 64	136
			65 or older	178
	Female	Age category	Less than 25	134
			25 to 34	351
			35 to 44	370
			45 to 54	260
			55 to 64	184
			65 or older	301

Figure 13. Tableau présentant la variable Tranche d'âge imbriquée dans la variable Sexe

Remarque : Les tableaux personnalisés ne prennent pas en charge le traitement de fichiers scindés en couches. Pour obtenir le même résultat que les fichiers scindés en couches, placez les variables de fichiers scindés dans les couches imbriquées les plus à l'extérieur du tableau.

Suppression de libellés de variable

Une autre solution au problème de la redondance des libellés de variable dans les tableaux imbriqués consiste simplement à supprimer l'affichage des noms ou libellés de variable. Les libellés de valeurs des variables *Sexe* et *Tranche d'âge* étant suffisamment descriptifs sans les libellés de variable, nous pouvons éliminer les libellés pour les deux variables.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et désélectionnez **Afficher le libellé de la variable** dans le menu contextuel.
3. Répétez l'opération pour la variable *Sexe*.

Les libellés de variable restent affichés dans l'aperçu du tableau, mais ne seront pas inclus dans le tableau.

4. Cliquez sur **OK** pour créer le tableau.

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Figure 14. Tableau imbriqué sans libellés de variable

Si vous souhaitez que les libellés de variable soient inclus quelque part dans le tableau, sans toutefois les afficher plusieurs fois dans le corps du tableau, vous pouvez les inclure dans le titre ou le libellé de coin du tableau.

5. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
6. Cliquez sur l'onglet **Titres**.
7. Cliquez n'importe où dans la zone de texte Titre.
8. Cliquez sur **Expression de table**. Le texte $\&[Expression\ de\ table]$ apparaît dans la zone de texte Titre. Ceci génère un titre de tableau incluant les libellés des variables utilisés dans le tableau.

9. Cliquez sur **OK** pour créer le tableau.

Gender > Age category

		Count
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Figure 15. Libellés de variable dans le titre du tableau

Le signe "supérieur à" (>) dans le titre indique que la variable *Tranche d'âge* est imbriquée dans la variable *Sexe*.

Tableau croisé imbriqué

Un tableau imbriqué peut contenir d'autres variables dans d'autres dimensions. Par exemple, vous pouvez imbriquer la variable *Tranche d'âge* dans la variable *Sexe* dans les lignes et croiser les lignes imbriquées avec une troisième variable dans la dimension de colonne.

1. Ouvrez le générateur de tableaux (menu *Analyse, Tableaux, Tableaux personnalisés*).
2. Si la variable *Tranche d'âge* n'est pas encore imbriquée dans la variable *Sexe* dans les lignes, suivez les instructions précédentes pour les imbriquer.
3. Faites glisser la variable *Informations sur Internet* de la liste des variables vers la zone Colonnes du panneau de canevas.

Vous pouvez remarquer que le tableau est trop grand pour être affiché entièrement dans le panneau de canevas. Vous pouvez utiliser le défilement vertical ou horizontal dans le panneau de canevas pour visualiser davantage l'aperçu du tableau ou vous pouvez :

- Dans le générateur de tableaux, cliquez sur **Compact** pour obtenir le mode Compact. Ceci affiche uniquement les libellés de variable, sans aucune information sur les catégories ou les statistiques récapitulatives incluses dans le tableau.
 - Augmentez la taille du générateur de tableaux en faisant glisser ses bords ou ses coins.
4. Cliquez sur **OK** pour créer le tableau.

				Get news from internet	
				No	Yes
				Count	Count
Gender	Male	Age category	Less than 25	59	49
			25 to 34	159	117
35 to 44			217	92	
45 to 54			169	52	
55 to 64			112	24	
65 or older			155	23	
Female	Age category	Less than 25	87	47	
		25 to 34	209	142	
		35 to 44	218	152	
		45 to 54	177	83	
		55 to 64	140	44	
		65 or older	261	40	

Figure 16. Tableau croisé imbriqué

Permutation des lignes et des colonnes

Après avoir passé un certain temps à élaborer un tableau complexe, vous jugez que ce dernier est absolument parfait, si ce n'est que vous souhaitez en modifier l'orientation, c'est-à-dire placer toutes les variables en ligne dans les colonnes et vice versa. Par exemple, vous avez créé un tableau croisé imbriqué avec les variables *Tranche d'âge* et *Sexe* imbriquées dans les lignes, mais vous souhaitez désormais que ces deux variables démographiques soient plutôt imbriquées dans les colonnes.

- Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
- Dans le panneau de canevas, cliquez n'importe où avec le bouton droit de la souris et sélectionnez **Permuter les variables de ligne et de colonne** dans le menu contextuel.
Les variables en ligne et en colonne ont été permutées.
Avant de créer le tableau, procédons à quelques modifications afin de rendre l'affichage moins confus.
- Sélectionnez **Masquer** pour supprimer l'affichage du libellé de colonne des statistiques récapitulatives.
- Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Sexe* et désélectionnez **Afficher le libellé de la variable**.
- Cliquez sur **OK** pour créer le tableau.

		Male						Female					
		Age category						Age category					
		Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older	Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older
Get news from internet	No	59	159	217	169	112	155	87	209	218	177	140	261
	Yes	49	117	92	52	24	23	47	142	152	83	44	40

Figure 17. Tableau croisé dans lequel les variables démographiques sont imbriquées dans les colonnes

Couches

Vous pouvez utiliser des couches pour ajouter une dimension de profondeur à vos tableaux et créer ainsi des « cubes » tridimensionnels. Les couches sont en fait relativement similaires à l'imbrication ou à l'empilement ; la différence principale réside dans le fait qu'une seule catégorie de couche est visible à la fois. Par exemple, l'utilisation de la variable *Tranche d'âge* comme variable de ligne et de la variable *Sexe* comme variable de couche a pour effet de produire un tableau dans lequel les informations relatives aux hommes et aux femmes apparaissent dans différentes couches.

- Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
- Cliquez sur **Restaurer** pour supprimer les sélections précédentes dans le générateur de tableaux.

3. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
4. Cliquez sur **Couches** au sommet du générateur de tableaux pour afficher la liste Couches.
5. Faites glisser la variable *Sexe* de la liste des variables vers la liste Couches.
A ce stade, vous pouvez remarquer que l'ajout d'une variable de couche n'a pas d'effet visible sur l'aperçu affiché dans le panneau de canevas. Les variables de couche n'affectent pas l'aperçu du panneau de canevas, sauf si la variable de couche est la variable source des statistiques et si vous modifiez les statistiques récapitulatives.
6. Cliquez sur **OK** pour créer le tableau.

		Count
Age category	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178

Figure 18. Tableau simple en couches

Au premier abord, ce tableau ne semble pas différent d'un tableau simple comportant une seule variable catégorielle. La seule différence réside dans la présence du libellé *Sexe - Masculin* au sommet du tableau.

7. Dans la fenêtre du visualiseur, double-cliquez sur le tableau pour l'activer.
8. Vous pouvez désormais remarquer que le libellé *Sexe - Masculin* représente en fait un choix dans la liste déroulante.
9. Cliquez sur la flèche vers le bas de la liste déroulante pour afficher la liste complète des couches.
Pour ce tableau, un seul autre choix existe dans la liste.
10. Dans la liste déroulante, sélectionnez *Sexe - Féminin*.

		Count
Age category	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

Figure 19. Tableau simple en couches dans lequel une autre couche est affichée

Deux variables catégorielles de couche empilées

Si plusieurs variables catégorielles sont présentes dans les couches, vous pouvez les empiler ou les imbriquer. Par défaut, les variables de couche sont empilées. (*Remarque* : Si vous disposez de variables d'échelle en couches, les variables de couche peuvent uniquement être empilées.)

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Si la variable *Tranche d'âge* ne se situe pas déjà dans les lignes et la variable *Sexe* dans les couches, suivez les instructions précédentes pour créer un tableau en couches.
3. Faites glisser la variable *Diplôme le plus important* de la liste des variables vers la liste Couche, sous *Sexe*.

Les deux boutons radio situés sous la liste Couche dans le groupe Sortie couche sont désormais activés. La sélection par défaut est **Afficher chaque catégorie en couche**. Ceci équivaut à l'empilement.

4. Cliquez sur **OK** pour créer le tableau.
5. Dans la fenêtre du visualiseur. double-cliquez sur le tableau pour l'activer.
6. Cliquez sur la flèche vers le bas de la liste déroulante pour afficher la liste complète des couches.

Le tableau comporte sept couches : deux couches pour les deux catégories *Sexe* et cinq couches pour les cinq catégories *Diplôme le plus important*. En ce qui concerne les couches empilées, le nombre total de couches correspond au nombre de catégories pour les variables de couche (y compris les catégories de total et sous-total demandées pour les variables de couche).

Deux variables catégorielles de couche imbriquées

L'imbrication de variables catégorielles de couche crée une couche distincte pour chaque combinaison de catégories de la variable de couche.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Si vous ne l'avez déjà fait, suivez les instructions précédentes pour créer un tableau de couches empilées.
3. Dans le groupe Sortie couche, sélectionnez **Afficher chaque combinaison de catégories en couche**. Ceci équivaut à l'imbrication.
4. Cliquez sur **OK** pour créer le tableau.
5. Dans la fenêtre du visualiseur. double-cliquez sur le tableau pour l'activer.
6. Cliquez sur la flèche vers le bas de la liste déroulante pour afficher la liste complète des couches.

Le tableau comporte 10 couches (vous devez parcourir la liste pour les visualiser toutes), une pour chaque combinaison des variables *Sexe* et *Diplôme le plus important*. En ce qui concerne les couches imbriquées, le nombre total de couches correspond au *produit* du nombre de catégories pour chaque variable de couche (dans cet exemple, $5 \times 2 = 10$).

Impression de tableaux en couches

Par défaut, seule la couche actuellement visible est imprimée. Pour imprimer toutes les couches d'un tableau :

1. Dans la fenêtre du visualiseur. double-cliquez sur le tableau pour l'activer.
2. Dans les menus de la fenêtre des résultats, sélectionnez :
Format > Propriétés du tableau...
3. Cliquez sur l'onglet **Impression**.
4. Sélectionnez **Imprimer toutes les couches**.

Vous pouvez également enregistrer ce paramètre comme partie constituante d'un modèle de tableau, y compris le modèle de tableau par défaut.

Chapitre 4. Totaux et sous-totaux pour les variables catégorielles

Vous pouvez inclure à la fois des totaux et des sous-totaux dans les tableaux personnalisés. Vous pouvez appliquer des totaux et des sous-totaux à des variables catégorielles à tout niveau d'imbrication, dans toute dimension (ligne, colonne et couche).

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Tous les exemples fournis ici affichent des libellés de variable dans des boîtes de dialogue ; ces libellés sont triés par ordre alphabétique. Les propriétés d'affichage de la liste des variables sont paramétrées dans l'onglet Général de la boîte de dialogue Options (menu Edition, Options).

Total simple pour une seule variable

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Statistiques récapitulatives** dans le menu contextuel.
4. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **N % colonne** dans la liste Statistiques et cliquez sur la flèche pour ajouter l'élément à la liste Afficher.
5. Dans la cellule Libellé de la liste Afficher, supprimez le libellé par défaut et tapez Pourcentage.
6. Cliquez sur **Appliquer à la sélection**.
7. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Catégories et totaux** dans le menu contextuel.
8. Dans la boîte de dialogue Catégories et totaux, sélectionnez **Total**.
9. Dans le générateur de tableaux, cliquez sur **Appliquer** , puis sur **OK** pour créer le tableau.

		Count	Percent
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	Total	2828	100.0%

Figure 20. Total simple pour une seule variable catégorielle

Tel écran, tel totalisé

Les totaux sont basés sur les catégories affichées dans le tableau. Si vous choisissez d'exclure certaines catégories d'un tableau, les observations issues de ces catégories ne sont pas incluses dans les calculs des totaux.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).

2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Catégories et totaux** dans le menu contextuel.
3. Dans la liste Libellé, cliquez sur la catégorie portant le libellé *Moins de 25*.
4. Cliquez sur le bouton flèche situé à gauche de la liste Exclure.
5. Dans la liste Libellé, cliquez sur la catégorie portant le libellé *65 ans et plus*.
6. Cliquez de nouveau sur le bouton flèche situé à gauche de la liste Exclure.
Les deux catégories sont déplacées de la liste Afficher vers la liste Exclure.
7. Dans le générateur de tableaux, cliquez sur **Appliquer** , puis sur **OK** pour créer le tableau.

		Count	Percent
Age category	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%
	Total	2107	100.0%

Figure 21. Total dans un tableau dans lequel des catégories ont été exclues

Le total dans ce tableau est seulement de 2 107, alors qu'il est de 2 828 lorsque toutes les catégories sont incluses. Seules les catégories utilisées dans le tableau sont incluses dans le total. (Le total des pourcentages est tout de même de 100 % car tous les pourcentages sont basés sur le nombre total d'observations utilisées dans le tableau, et non sur le nombre total d'observations dans le fichier de données.)

Position d'affichage des totaux

Par défaut, les totaux sont affichés sous les catégories totalisées. Vous pouvez modifier la position d'affichage des totaux et les placer au-dessus des catégories totalisées.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Catégories et totaux** dans le menu contextuel.
3. Dans le groupe Affichage des totaux et des sous-totaux, sélectionnez **Catégories ci-dessus auxquelles ils s'appliquent**.
4. Dans le générateur de tableaux, cliquez sur **Appliquer** , puis sur **OK** pour créer le tableau.

		Count	Percent
Age category	Total	2107	100.0%
	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%

Figure 22. Total affiché au-dessus des catégories totalisées

Totaux pour tableaux imbriqués

Puisque les totaux peuvent être appliqués aux variables catégorielles à tout niveau d'imbrication, vous pouvez créer des tableaux qui contiennent des totaux de groupe à plusieurs niveaux d'imbrication.

Totaux de groupe

Les totaux des variables catégorielles imbriquées dans d'autres variables catégorielles représentent des totaux de groupe.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Faites glisser la variable *Sexe* vers la gauche de la variable *Tranche d'âge* dans le panneau de canevas.
3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Catégories et totaux** dans le menu contextuel.
Avant de créer le tableau, repositionnons les totaux sous les catégories totalisées.
4. Dans le groupe Affichage des totaux et des sous-totaux, sélectionnez **Catégories ci-dessous auxquelles ils s'appliquent**.
5. Cliquez sur **Appliquer** pour enregistrer le paramètre et revenir au générateur de tableaux.
6. Cliquez sur **OK** pour créer le tableau.

				Count	Percent
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			Total	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			Total	1165	100.0%

Figure 23. Totaux de la variable *Tranche d'âge* au sein des catégories de la variable *Sexe*

Le tableau comporte désormais deux totaux de groupe : un pour les hommes et un autre pour les femmes.

Totaux généraux

Les totaux appliqués aux variables imbriquées sont toujours des totaux de groupe et non des totaux généraux. Si vous souhaitez obtenir des totaux pour l'intégralité du tableau, vous pouvez appliquer des totaux à la variable située au niveau d'imbrication le plus externe.

1. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Sexe* et choisissez **Catégories et totaux** dans le menu contextuel.
3. Dans la boîte de dialogue *Catégories et totaux*, sélectionnez **Total**.
4. Dans le générateur de tableaux, cliquez sur **Appliquer**, puis sur **OK** pour créer le tableau.

				Count	Percent
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			Total	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			Total	1165	100.0%
	Total	Age category	25 to 34	627	29.8%
			35 to 44	679	32.2%
			45 to 54	481	22.8%
			55 to 64	320	15.2%
			Total	2107	100.0%

Figure 24. Totaux généraux pour un tableau imbriqué

Notez que le total général n'est que de 2 107, et non de 2 828. Deux tranches (catégories) d'âge étant toujours exclues du tableau, les observations de ces catégories sont donc exclues de tous les totaux.

Totaux des variables de couche

Les totaux des variables de couche apparaissent sous la forme de couches distinctes dans le tableau.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le générateur de tableaux, cliquez sur **Couches** pour afficher la liste Couches.
3. Faites glisser la variable *Sexe* de la zone Lignes du panneau de canevas vers la liste Couches.
Remarque : Puisque vous avez déjà spécifié des totaux pour la variable *Sexe*, il n'est pas nécessaire de le faire à ce stade. Le fait de déplacer la variable entre les dimensions n'affecte aucun des paramètres de cette variable.
4. Cliquez sur **OK** pour créer le tableau.
5. Dans le visualiseur, double-cliquez sur le tableau pour l'activer.
6. Cliquez sur la flèche vers le bas dans la liste déroulante Couche pour afficher une liste de toutes les couches du tableau.

Le tableau comporte trois couches : *Sexe - Masculin*, *Sexe - Féminin* et *Sexe - Total*.

Position d'affichage des totaux de couche

En ce qui concerne les totaux de variable de couche, la position d'affichage (au-dessus ou au-dessous) des totaux détermine la position de la couche pour les totaux. Par exemple, si vous spécifiez **Catégories ci-dessus auxquelles ils s'appliquent** pour un total de variable de couche, la couche du total est la première couche affichée.

Sous-totaux

Vous pouvez inclure des sous-totaux pour des sous-groupes de catégories d'une variable. Par exemple, vous pouvez inclure des sous-totaux pour les tranches (catégories) d'âge qui représentent tous les répondants âgés de moins et de plus de 45 ans dans l'échantillon d'enquête.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents dans le générateur de tableaux.
3. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
4. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Catégories et totaux** dans le menu contextuel.
5. Sélectionnez **3,00** dans la liste des valeurs.
6. Cliquez sur **Ajouter un sous-total** pour afficher la boîte de dialogue Définir le sous-total.
7. Saisissez Sous-total < 45 dans la zone de texte Libellé.
8. Cliquez ensuite sur **Poursuivre**.
Cette action insère une ligne contenant le sous-total des trois premières tranches d'âge.
9. Sélectionnez **6,00** dans la liste des valeurs.
10. Cliquez sur **Ajouter un sous-total** pour afficher la boîte de dialogue Définir le sous-total.
11. Saisissez Sous-total 45+ dans la zone de texte Libellé.
12. Cliquez ensuite sur **Poursuivre**.

Remarque importante : Vous devez sélectionner la position d'affichage des totaux et des sous-totaux (**Catégories ci-dessus auxquelles ils s'appliquent** ou **Catégories ci-dessous auxquelles ils s'appliquent**) avant de définir un sous-total. La modification de la position d'affichage affecte tous les sous-totaux (et non pas uniquement le sous-total sélectionné) ; en outre, elle *modifie les catégories incluses dans les sous-totaux*.

13. Dans le générateur de tableaux, cliquez sur **Appliquer** , puis sur **OK** pour créer le tableau.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Subtotal < 45	1548
	45 to 54	481
	55 to 64	320
	65 or older	479
	Subtotal 45+	1280

Figure 25. Sous-totaux pour la variable Tranche d'âge

Tel écran, tel sous-totalisé

Tout comme les totaux, les sous-totaux sont basés sur les catégories incluses dans le tableau.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Catégories et totaux** dans le menu contextuel.

Remarque : La valeur (non le libellé) affichée pour le premier sous-total est **1...3** ; elle indique que le sous-total inclut toutes les valeurs de la liste entre 1 et 3.

3. Sélectionnez **1** dans la liste des valeurs ou cliquez sur le libellé *Moins de 25*).
4. Cliquez sur le bouton flèche situé à gauche de la liste Exclure.

Désormais, la première tranche (catégorie) d'âge est exclue et la valeur affichée pour le premier sous-total passe à **2...3** ; elle indique que la catégorie exclue ne sera pas incluse dans le sous-total car les sous-totaux sont basés sur les catégories incluses dans le tableau. L'exclusion d'une catégorie entraîne son exclusion automatique des sous-totaux ; vous ne pouvez donc pas, par exemple, afficher uniquement des sous-totaux sans les catégories sur lesquelles ils se basent.

Masquage de catégories de sous-total

Vous pouvez supprimer l'affichage des catégories définissant un sous-total et afficher uniquement le sous-total, sans affecter les données sous-jacentes, en « fusionnant » efficacement les catégories.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents dans le générateur de tableaux.
3. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
4. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Catégories et totaux** dans le menu contextuel.
5. Sélectionnez **3,00** dans la liste des valeurs.
6. Cliquez sur **Ajouter un sous-total** pour afficher la boîte de dialogue Définir le sous-total.
7. Saisissez Inférieur à 45 dans la zone de texte Libellé.
8. Sélectionnez (cochez) **Masquer les catégories des sous-totaux dans le tableau**.
9. Cliquez ensuite sur **Poursuivre**.
Cette action insère une ligne contenant le sous-total des trois premières tranches d'âge.
10. Sélectionnez **6,00** dans la liste des valeurs.
11. Cliquez sur **Ajouter un sous-total** pour afficher la boîte de dialogue Définir le sous-total.
12. Saisissez 45 ou plus dans la zone de texte Libellé.
13. Sélectionnez (cochez) **Masquer les catégories de sous-total**.
14. Cliquez ensuite sur **Poursuivre**.
15. Pour inclure un total avec les sous-totaux, cochez la case **Total** dans le groupe Présenter.

16. Cliquez sur **Appliquer**.

Le canevas montre que les sous-totaux seront affichés, mais que les catégories définissant les sous-totaux seront exclues.

17. Cliquez sur **OK** pour générer ce tableau.

		Count
Age category	Less than 45	1548
	45 or older	1280
	Total	2828

Figure 26. Tableau contenant uniquement les sous-totaux et les totaux

Sous-totaux d'une variable de couche

Tout comme les totaux, les sous-totaux des variables de couche apparaissent sous la forme de couches distinctes dans le tableau. Les sous-totaux sont principalement traités comme des catégories. Chaque catégorie est une couche distincte dans le tableau et l'ordre d'affichage des catégories de couche est déterminé par l'ordre des catégories dans la boîte de dialogue Catégories et totaux, y compris la position d'affichage des catégories de sous-totaux.

Chapitre 5. Catégories calculées pour les variables catégorielles

Vous pouvez inclure des catégories calculées dans les tableaux personnalisés. Il s'agit de nouvelles catégories calculées à partir de catégories de la même variable au niveau de l'imbrication selon n'importe quelle dimension : ligne, colonne ou couche. Par exemple, vous pouvez inclure une catégorie calculée qui montre la différence entre deux catégories.

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Catégorie calculée simple

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux...
2. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Catégories et totaux** dans le menu contextuel.
4. Sélectionnez **3,00** dans la liste des valeurs.
5. Cliquez sur **Ajouter une catégorie** pour afficher la boîte de dialogue Définir la catégorie calculée.
6. Saisissez **Inférieur à 45** dans la zone de texte Libellé de la catégorie calculée.
7. Sélectionnez **Inférieur à 25 (1,00)** dans la liste des catégories et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée. [1] s'affiche dans l'expression.
8. Cliquez sur le bouton plus (+) du clavier numérique de la boîte de dialogue (ou appuyez sur la touche + sur votre clavier).
9. Sélectionnez **25 à 34 (2,00)** dans la liste des catégories et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
10. Cliquez sur le bouton plus (+) du clavier numérique de la boîte de dialogue (ou appuyez sur la touche + sur votre clavier).
11. Sélectionnez **35 à 44 (3,00)** dans la liste des catégories et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
12. Cliquez ensuite sur **Poursuivre**.
Cette action insère une ligne contenant le sous-total des trois premières tranches d'âge.
13. Sélectionnez **5,00** dans la liste des valeurs.
14. Cliquez sur **Ajouter un sous-total** pour afficher la boîte de dialogue Définir le sous-total.
15. Saisissez **Inférieur à 65** dans la zone de texte du libellé.
16. Cliquez ensuite sur **Poursuivre**.
Cette action insère une ligne contenant le sous-total des cinq premières catégories.
17. Dans le générateur de tableaux, cliquez sur **Appliquer**, puis sur **OK** pour créer le tableau.

		Count
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Less than 45	1548
	45 to 54	481
	55 to 64	320
	Less than 65	2349
	65 or older	479

Figure 27. Catégories calculées avec des sous-totaux

Le tableau comprend une catégorie calculée (*Inférieur à 45*) et un sous-total (*Inférieur à 65*). Le sous-total inclut les catégories comprises dans la catégorie calculée. Il n'est pas possible de créer le même tableau uniquement à l'aide des sous-totaux, car ces derniers ne peuvent pas partager les mêmes catégories.

Masquage des catégories dans une catégorie calculée

Comme pour les sous-totaux, vous avez la possibilité de supprimer l'affichage des catégories utilisées dans une expression de catégorie calculée, et de n'afficher que la catégorie calculée. L'exemple suivant est construit sur le précédent.

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et choisissez **Catégories et totaux** dans le menu contextuel.
3. Dans la liste des valeurs, sélectionnez la catégorie calculée *Inférieur à 45*.
4. Cliquez sur **Edition** pour afficher la boîte de dialogue Définir la catégorie calculée.
5. Sélectionnez **Masquer les catégories utilisées dans l'expression dans le tableau**.
6. Cliquez ensuite sur **Poursuivre**.
7. Dans la liste des valeurs, sélectionnez le sous-total *Inférieur à 65*.
8. Cliquez sur **Edition** pour afficher la boîte de dialogue Définir le sous-total.
9. Sélectionnez **Masquer les catégories des sous-totaux dans le tableau**.
10. Cliquez ensuite sur **Poursuivre**.
11. Dans le générateur de tableaux, cliquez sur **Appliquer**, puis sur **OK** pour créer le tableau.

		Count
Age category	Less than 45	1548
	Less than 65	2349
	65 or older	479

Figure 28. Catégories calculées avec sous-totaux et catégories masquées

Comme dans l'exemple précédent, le tableau comprend une catégorie calculée et un sous-total. Toutefois, dans ce cas les catégories sont masquées pour chacun d'entre eux et seuls les totaux sont affichés.

Référencement des sous-totaux dans une catégorie calculée

Vous pouvez inclure des sous-totaux dans une expression de catégorie calculée.

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux...
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents dans le générateur de tableaux.
3. Dans le générateur de tableaux, faites glisser la variable *Situation d'activité* de la liste des variables vers la zone Lignes du panneau de canevas.

4. Faites glisser la variable *Situation familiale* de la liste des variables vers la zone Colonnes.
5. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Situation d'activité* et choisissez **Catégories et totaux** dans le menu contextuel.
6. Sélectionnez **2** dans la liste des valeurs.
7. Cliquez sur **Ajouter un sous-total** pour afficher la boîte de dialogue Définir le sous-total.
8. Saisissez Travail dans la zone de texte Libellé.
9. Sélectionnez **Masquer les catégories des sous-totaux dans le tableau.**
10. Cliquez ensuite sur **Poursuivre.**
Cette action insère une ligne contenant le sous-total des deux premières catégories d'activité.
11. Sélectionnez **8** dans la liste des valeurs.
12. Cliquez sur **Ajouter un sous-total** pour afficher la boîte de dialogue Définir le sous-total.
13. Saisissez Sans travail dans la zone de texte Libellé.
14. Sélectionnez **Masquer les catégories des sous-totaux.**
15. Cliquez ensuite sur **Poursuivre.**
Cette action insère une ligne contenant le sous-total des autres catégories d'activité.
16. Dans la liste des valeurs, sélectionnez le sous-total *Sans travail*.
17. Cliquez sur **Ajouter une catégorie** pour afficher la boîte de dialogue Définir la catégorie calculée.
18. Saisissez Travail / Sans travail dans la zone de texte Libellé de la catégorie calculée.
19. Sélectionnez **Travail (Travail n°1)** dans la liste des totaux et sous-totaux et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
20. Cliquez sur le bouton de division (/) du clavier numérique de la boîte de dialogue (ou appuyez sur la touche / sur votre clavier).
21. Sélectionnez **Sans travail (Sans travail n°2)** dans la liste des totaux et sous-totaux et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
Par défaut, la catégorie calculée utilise le même format que la statistique de la variable, Effectif dans ce cas. Parce que nous voulons afficher les décimales résultant de la division dans l'expression de la catégorie calculée, et que le format par défaut de la statistique Effectif n'affiche pas les décimales, nous devons modifier le format.
22. Cliquez sur l'onglet Formats d'affichage.
23. Changez le paramètre Décimales d'Effectif sur **2**.
24. Cliquez ensuite sur **Poursuivre.**
25. Dans le générateur de tableaux, cliquez sur **Appliquer**, puis sur **OK** pour créer le tableau.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Count	Count	Count	Count	Count
Labor force status	Working	916	64	330	67	494
	Not Working	429	219	116	26	169
	Working / Not Working	2.14	.29	2.84	2.58	2.92

Figure 29. Catégorie calculée affichant le rapport des sous-totaux

Le tableau comprend deux sous-totaux et une catégorie calculée. La catégorie calculée affiche le rapport des sous-totaux, et vous permet de comparer facilement les groupes représentés par chaque sous-total. Le rapport entre les répondants veufs avec travail et les répondants veufs sans travail est plus faible que celui des autres groupes.. Il existe aussi un rapport légèrement plus faible de répondants mariés, résultant peut-être de conjoints qui ont arrêtés de travailler pour s'occuper des enfants.

Utilisation des catégories calculées pour afficher des sous-totaux non-exhaustifs

Les sous-totaux sont exhaustifs. Autrement dit, tous les sous-totaux d'un tableau comprennent toutes les valeurs au-dessus ou au-dessous de leur position dans le tableau. Les catégories calculées, elles, ne sont pas exhaustives et vous permettent d'ajouter et de mélanger les catégories d'un tableau.

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux...
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents dans le générateur de tableaux.
3. Dans le générateur de tableaux, faites glisser la variable *Se considère libéral ou conservateur* de la liste des variables vers la zone Lignes du panneau de canevas.
4. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur la variable *Se considère libéral ou conservateur* et choisissez **Catégories et totaux** dans le menu contextuel.
5. Sélectionnez 3 dans la liste des valeurs.
6. Cliquez sur **Ajouter une catégorie** pour afficher la boîte de dialogue Définir la catégorie calculée.
7. Saisissez Sous-total Libéraux dans la zone de texte Libellé de catégorie calculée. Remarque : il y a quatre espaces avant le texte. Ces espaces sont utilisés pour l'indentation du tableau des résultats.
8. Sélectionnez **Extrêmement libéral (1)** dans la liste des catégories et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
9. Cliquez sur le bouton plus (+) du clavier numérique de la boîte de dialogue (ou appuyez sur la touche + sur votre clavier).
10. Sélectionnez **Libéral (2)** dans la liste des catégories et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
11. Cliquez sur le bouton plus (+) du clavier numérique de la boîte de dialogue (ou appuyez sur la touche + sur votre clavier).
12. Sélectionnez **Légèrement libéral (3)** dans la liste des catégories et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
13. Cliquez sur **Poursuivre**.
Cette action insère une ligne contenant le sous-total des catégories Libéraux.
14. Sélectionnez 7 dans la liste des valeurs.
15. Cliquez sur **Ajouter une catégorie** pour afficher la boîte de dialogue Définir la catégorie calculée.
16. Saisissez Sous-total Conservateurs dans la zone de texte Libellé de la catégorie calculée.
Remarque : il y a quatre espaces avant le texte. Ces espaces sont utilisés pour l'indentation du tableau des résultats.
17. Sélectionnez **Légèrement conservateur (5)** dans la liste des catégories et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
18. Cliquez sur le bouton plus (+) du clavier numérique de la boîte de dialogue (ou appuyez sur la touche + sur votre clavier).
19. Sélectionnez **Conservateur (6)** dans la liste des catégories et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
20. Cliquez sur le bouton plus (+) du clavier numérique de la boîte de dialogue (ou appuyez sur la touche + sur votre clavier).
21. Sélectionnez **Extrêmement conservateur (7)** dans la liste des catégories et cliquez sur le bouton flèche pour copier cette valeur dans la zone de texte Expression pour la catégorie calculée.
22. Cliquez sur **Poursuivre**.
Cette action insère une ligne contenant le sous-total des catégories Conservateurs.
23. Dans le générateur de tableaux, cliquez sur **Appliquer**, puis sur **OK** pour créer le tableau.

		Count
Think of self as liberal or conservative	Extremely liberal	64
	Liberal	357
	Slightly liberal	351
	Liberal Subtotal	772
	Moderate	986
	Slightly conservative	432
	Conservative	415
	Extremely conservative	86
	Conservative Subtotal	933

Figure 30. Catégories calculées affichant les sous-totaux non-exhaustifs

Le tableau comprend deux catégories calculées qui n'incluent pas toutes les catégories du tableau. La catégorie *Modérés* n'est comprise dans aucune catégorie calculée. Il est impossible de créer le même tableau avec des sous-totaux car ils doivent être exhaustifs.

Chapitre 6. Tableaux pour variables avec catégories partagées

Les enquêtes contiennent souvent de nombreuses questions caractérisées par un ensemble commun de réponses possibles. Par exemple, notre échantillon d'enquête contient un certain nombre de variables relatives à la confiance accordée à divers services et institutions publics et privés, qui présentent toutes le même jeu de catégories de réponses suivant : 1 = *Beaucoup*, 2 = *Seulement certaines* et 3 = *Presque aucune*. Vous pouvez utiliser la technique de l'empilement pour afficher ces variables liées dans le même tableau et afficher les catégories de réponses partagées dans les colonnes du tableau. Ces fonctions sont également disponibles si vous utilisez des catégories calculées, à la condition que le libellé et l'expression de chaque catégorie calculée soient identiques dans toutes les variables.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Figure 31. Tableau de variables avec catégories partagées

Remarque : Dans la version précédente de la fonction Tableaux personnalisés, cet élément était nommé "tableau d'effectifs".

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Tous les exemples fournis ici affichent des libellés de variable dans des boîtes de dialogue ; ces libellés sont triés par ordre alphabétique. Les propriétés d'affichage de la liste des variables sont paramétrées dans l'onglet Général de la boîte de dialogue Options (menu Modifier, Options).

Tableau d'effectifs

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Dans la liste des variables du générateur de tableaux, cliquez sur *Confiance en les banques*, puis, tout en maintenant la touche Maj enfoncée, cliquez sur *Confiance en la télévision* pour sélectionner toutes les variables commençant par le terme "confiance". (*Remarque* : Ceci implique que les libellés de variable soient affichés dans l'ordre alphabétique, et non dans l'ordre du fichier, dans la liste des variables.)
3. Faites glisser les six variables de confiance vers la zone Lignes du panneau de canevas.
Les variables sont ainsi empilées dans la dimension de ligne. Par défaut, les libellés de catégorie de chaque variable apparaissent également dans les lignes, ce qui donne un tableau très long et étroit (6 variables x 3 catégories = 18 lignes). Cependant, puisque les six variables partagent les mêmes libellés de catégorie définis (libellés de valeurs), vous pouvez placer les libellés de catégorie dans la dimension de colonne.
4. Dans la liste déroulante Emplacement de la catégorie, sélectionnez **Libellés de ligne dans les colonnes**.

Le tableau comporte désormais seulement six lignes, une pour chaque variable empilée, et les catégories définies deviennent les colonnes du tableau.

5. Avant de créer le tableau, sélectionnez **Masquer** pour Position dans le groupe Statistiques récapitulatives, puisque le libellé de statistique récapitulative *Effectif* n'est pas vraiment nécessaire.
6. Cliquez sur **OK** pour créer le tableau.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	490	1068	306
Confidence in education	511	1055	315
Confidence in major companies	500	1078	243
Confidence in medicine	844	864	167
Confidence in press	176	878	808
Confidence in television	196	936	744

Figure 32. Tableau de variables de lignes empilées avec libellés de catégories partagées disposées en colonnes

Au lieu d'afficher les variables dans les lignes et les catégories dans les colonnes, vous pouvez créer un tableau dans lequel les variables sont empilées dans les colonnes et les catégories affichées dans les lignes. Cette solution est préférable si le nombre de catégories est supérieur au nombre de variables, ce qui n'est pas le cas dans notre exemple, puisque le nombre de variables est supérieur au nombre de catégories.

Tableau de pourcentages

Pour un tableau dans lequel les variables sont empilées dans les lignes et les catégories affichées dans les colonnes, il semble pertinent (ou du moins compréhensible) de faire apparaître les pourcentages en ligne. (Pour un tableau dans lequel les variables sont empilées dans les colonnes et les catégories affichées dans les lignes, les pourcentages en colonne peuvent sembler plus appropriés.)

1. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur l'une des variables de confiance dans l'aperçu du tableau et choisissez **Statistiques récapitulatives** dans le menu contextuel.
3. Dans la liste Statistiques, sélectionnez **N % ligne** et cliquez sur le bouton flèche pour ajouter cet élément à la liste Afficher.
4. Cliquez sur une cellule de la ligne *Effectif* dans la liste Afficher, puis sur le bouton flèche pour enlever cet élément de cette liste et le replacer dans la liste Statistiques.
5. Cliquez sur **Appliquer à tout** pour appliquer la modification de la statistique récapitulative à toutes les variables empilées du tableau.

Remarque : Si l'aperçu du tableau ne ressemble pas à la figure ci-dessus, vous avez probablement cliqué sur **Appliquer à la sélection** au lieu de cliquer sur **Appliquer à tout**, ce qui applique la nouvelle statistique récapitulative uniquement à la variable sélectionnée. Dans cet exemple, ceci donne deux colonnes pour chaque catégorie : une dans laquelle des valeurs de substitution apparaissent pour l'effectif pour toutes les autres variables et une dans laquelle une valeur de substitution apparaît pour le pourcentage de ligne de la variable sélectionnée. Il s'agit exactement du tableau qui serait généré, mais *pas* de celui souhaité dans cet exemple.

6. Cliquez sur **OK** pour créer le tableau.

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	26.3%	57.3%	16.4%
Confidence in education	27.2%	56.1%	16.7%
Confidence in major companies	27.5%	59.2%	13.3%
Confidence in medicine	45.0%	46.1%	8.9%
Confidence in press	9.5%	47.2%	43.4%
Confidence in television	10.4%	49.9%	39.7%

Figure 33. Tableau des pourcentages de ligne pour les variables empilées dans les lignes, avec affichage des catégories dans les colonnes

Remarque : Vous pouvez inclure un nombre quelconque de statistiques récapitulatives dans un tableau de variables avec catégories partagées. Par souci de simplicité, nos exemples illustrent une seule statistique à la fois.

Contrôle des totaux et des catégories

Vous pouvez créer des tableaux avec des catégories situées dans la dimension opposée à celle des variables uniquement si toutes les variables du tableau comportent les mêmes catégories, affichées dans le même ordre. Ceci inclut les totaux, les sous-totaux et tout autre ajustement de catégorie que vous effectuez. Par conséquent, toute modification apportée dans la boîte de dialogue Catégories et totaux doit être effectuée pour toutes les variables du tableau partageant les catégories.

1. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur la première variable de confiance dans l'aperçu du tableau et choisissez **Catégories et totaux** dans le menu contextuel.
3. Dans la boîte de dialogue Catégories et totaux, sélectionnez **Total**, puis cliquez sur **Appliquer**.
La première chose que vous pouvez remarquer est que les libellés de catégorie ont été déplacés des colonnes vers les lignes. Vous pouvez également noter que le contrôle Emplacement de la catégorie est désormais désactivé. En effet, les variables ne partagent plus exactement le même ensemble de "catégories". Une des variables comporte désormais une catégorie Total.
4. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur l'une des variables de confiance et choisissez l'option de menu **Sélectionner toutes les variables de ligne** ou, tout en maintenant la touche Ctrl enfoncée, cliquez sur chaque variable empilée dans le panneau de canevas jusqu'à ce qu'elles soient toutes sélectionnées (vous devrez peut-être parcourir le panneau ou agrandir la fenêtre du générateur de tableaux).
5. Cliquez sur **Catégories et totaux** dans le groupe Définir.
6. Si l'élément **Total** n'est pas déjà sélectionné dans la boîte de dialogue Catégories et totaux, sélectionnez-le, puis cliquez sur **Appliquer**.
7. La liste déroulante Emplacement de la catégorie doit de nouveau être activée, puisque toutes les variables comportent désormais la catégorie Total supplémentaire ; vous pouvez sélectionner **Libellés de ligne dans les colonnes**.
8. Cliquez sur **OK** pour créer le tableau.

	A great deal	Only some	Hardly any	Total
Confidence in banks & financial institutions	26.3%	57.3%	16.4%	100.0%
Confidence in education	27.2%	56.1%	16.7%	100.0%
Confidence in major companies	27.5%	59.2%	13.3%	100.0%
Confidence in medicine	45.0%	46.1%	8.9%	100.0%
Confidence in press	9.5%	47.2%	43.4%	100.0%
Confidence in television	10.4%	49.9%	39.7%	100.0%

Figure 34. Tableau des pourcentages de ligne pour les variables empilées dans les lignes, avec affichage des catégories et des totaux dans les colonnes

Imbrication dans des tableaux comportant des catégories partagées

Dans les tableaux imbriqués, les variables empilées comportant des catégories partagées doivent être situées au niveau d'imbrication le plus interne dans leur dimension si vous souhaitez afficher les libellés de catégorie dans la dimension opposée.

1. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Faites glisser la variable *Sexe* de la liste des variables vers la gauche de la zone Lignes.
Les variables empilées comportant des catégories partagées sont désormais imbriquées dans les catégories de la variable *Sexe* dans l'aperçu du tableau.
3. Faites glisser la variable *Sexe* vers la droite de l'une des variables de confiance empilées dans l'aperçu du tableau.

De nouveau, les libellés de catégorie ont été replacés dans la dimension de ligne et le contrôle Emplacement de la catégorie est désactivé. En outre, la variable *Sexe* est désormais imbriquée dans l'une des variables empilées, alors que les autres variables empilées ne contiennent aucune variable imbriquée. Vous pouvez ajouter la variable *Sexe* comme variable imbriquée à chaque variable empilée, mais le déplacement des libellés des lignes vers les colonnes aboutirait à l'affichage des libellés de catégorie de la variable *Sexe* dans les colonnes, au lieu des libellés de catégorie des variables empilées comportant des catégories partagées. En effet, la variable *Sexe* deviendrait la variable imbriquée la plus interne ; or, la modification de la position d'une catégorie s'applique systématiquement à la variable imbriquée la plus interne.

Chapitre 7. Statistiques récapitulatives

Les **statistiques récapitulatives** comprennent une large variété d'éléments, des effectifs simples pour les variables catégorielles aux mesures de dispersion, comme l'erreur standard de la moyenne pour les variables d'échelle. Elles n'incluent *pas* les tests de signification disponibles dans l'onglet Statistiques de test de la boîte de dialogue Tableaux personnalisés. Pour plus d'informations, voir «Statistiques de test», à la page 65.

Les statistiques récapitulatives pour les variables catégorielles et les jeux de réponses multiples incluent des effectifs et une grande variété de calculs de pourcentage, notamment :

- Pourcentages de ligne
- Pourcentages de colonne
- Pourcentages de sous-table
- Pourcentages de tableau
- Pourcentages de N valide

Outre les statistiques récapitulatives disponibles pour les variables catégorielles, les statistiques récapitulatives pour les variables d'échelle et d'autres statistiques en total pour les variables catégorielles incluent :

- Moyenne
- Médiane
- Percentiles
- Somme
- Ecart type
- Plage
- Valeurs minimale et maximale

Des statistiques récapitulatives supplémentaires sont disponibles pour les jeux de réponses multiples. Une liste complète des statistiques récapitulatives est également disponible. Pour plus d'informations, voir «Statistiques récapitulatives», à la page 6.

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Tous les exemples fournis ici affichent des libellés de variable dans des boîtes de dialogue ; ces libellés sont triés par ordre alphabétique. Les propriétés d'affichage de la liste des variables sont paramétrées dans l'onglet Général de la boîte de dialogue Options (menu Modifier, Options).

Variable source des statistiques récapitulatives

Les statistiques récapitulatives disponibles dépendent du niveau de mesure de leur variable source. La source des statistiques récapitulatives (variable sur laquelle sont basées les statistiques) est déterminée par les éléments suivants :

- **Niveau de mesure** : Si un tableau (ou une section dans un tableau de type empilé) contient une variable d'échelle, les statistiques récapitulatives sont basées sur cette dernière.
- **Ordre de sélection des variables** : La dimension source par défaut (ligne ou colonne) des statistiques pour les variables catégorielles est basée sur l'ordre dans lequel vous avez glissé les variables vers le

panneau de canevas. Par exemple, si vous faites glisser d'abord une variable vers la zone Lignes, la dimension de ligne est la dimension source par défaut des statistiques.

- **Imbrication** : Pour les variables catégorielles, les statistiques récapitulatives sont basées sur la variable la plus interne dans la dimension source.

Une table empilée peut comporter plusieurs variables sources de statistiques récapitulatives (variable d'échelle et variable catégorielle), mais chaque section du tableau comporte une seule source de statistiques récapitulatives.

Source des statistiques récapitulatives pour les variables catégorielles

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* et sélectionnez **Statistiques récapitulatives** dans le menu contextuel. (Puisqu'il s'agit de la seule variable du tableau, elle est la variable source des statistiques.)
4. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez *N % colonne* dans la liste Statistiques et cliquez sur la flèche pour ajouter l'élément à la liste Afficher.
5. Cliquez sur **Appliquer à la sélection**.
6. Dans le générateur de tableaux, faites glisser la variable *Informations sur Internet* vers la droite de la variable *Tranche d'âge* dans le panneau de canevas.
7. Dans le panneau de canevas, cliquez de nouveau avec le bouton droit de la souris sur *Tranche d'âge*. L'élément **Statistiques récapitulatives** du menu contextuel est désormais désactivé, car *Tranche d'âge* n'est pas la variable imbriquée la plus interne dans la dimension source des statistiques.
8. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Informations sur Internet*. L'élément **Statistiques récapitulatives** est activé car, s'agissant de la variable imbriquée la plus interne dans la dimension source, elle représente désormais la variable source des statistiques récapitulatives. (Le tableau ne comportant qu'une seule ligne Dimensions, il s'agit de la dimension source des statistiques.)
9. Faites glisser *Informations sur Internet* de la zone Lignes du panneau de canevas vers la zone Colonnes.
10. Dans le panneau de canevas, cliquez de nouveau sur le bouton droit de la souris sur *Informations sur Internet*. L'élément **Statistiques récapitulatives** du menu contextuel est désormais désactivé, car la variable ne se trouve plus dans la dimension source des statistiques.

Tranche d'âge est à nouveau la variable source des statistiques, car la dimension source par défaut pour les variables catégorielles est la première dimension dans laquelle vous placez des variables lors de la création du tableau. Dans cet exemple, la première opération a consisté à placer des variables dans la dimension de ligne. Par conséquent, la dimension de ligne est la dimension source par défaut des statistiques ; d'autre part, *Tranche d'âge* étant désormais la seule variable de cette dimension, il s'agit de la variable source des statistiques.

Source des statistiques récapitulatives pour les variables d'échelle

1. Faites glisser la variable d'échelle *Heures par jour devant la télé* vers la gauche de la variable *Tranche d'âge* dans la zone Lignes du panneau de canevas.
La première chose qui saute aux yeux est que les récapitulatifs *Effectif* et *N % colonne* ont été remplacés par *Moyenne*. En outre, si vous cliquez avec le bouton droit de la souris sur *Heures par jour devant la télé* dans le panneau de canevas, vous constatez qu'il s'agit désormais de la variable source des statistiques récapitulatives. Pour un tableau comportant une variable d'échelle, cette dernière est toujours la variable source des statistiques, quel que soit son niveau d'imbrication ou sa dimension, et la statistique récapitulative par défaut pour les variables d'échelle est la moyenne.

- Faites glisser *Heures par jour devant la télé* de la zone Lignes vers la zone Colonnes, au-dessus de *Informations sur Internet*.
- Cliquez avec le bouton droit de la souris sur *Heures par jour devant la télé* et sélectionnez **Statistiques récapitulatives** dans le menu contextuel. (Elle reste la variable source des statistiques, même lorsque vous la déplacez vers une autre dimension.)
- Dans la boîte de dialogue Statistiques récapitulatives, cliquez sur la cellule **Format** pour la moyenne dans la liste Afficher, puis sélectionnez **nnnn** dans la liste déroulante Format. (Vous pouvez être amené à parcourir la liste pour accéder à cet élément.)
- Dans la cellule Décimales, tapez 2.
- Cliquez sur **Appliquer à la sélection**.
L'aperçu du tableau dans le panneau de canevas indique désormais que les valeurs moyennes apparaîtront avec deux décimales.
- Cliquez sur **OK** pour créer le tableau.

		Hours per day watching TV	
		Get news from internet	
		No	Yes
		Mean	Mean
Age category	Less than 25	3.54	2.12
	25 to 34	3.42	2.14
	35 to 44	3.00	2.01
	45 to 54	2.83	2.06
	55 to 64	3.24	2.37
	65 or older	3.82	2.33

Figure 35. Variable d'échelle récapitulée avec des variables catégorielles de tableau croisé

Variables empilées

Etant donné qu'un tableau empilé peut contenir plusieurs variables source de statistiques et que vous pouvez indiquer différentes statistiques récapitulatives pour chacune de ces variables, vous devez prendre en considération certains éléments avant d'indiquer les statistiques récapitulatives dans les tableaux empilés.

- Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
- Cliquez sur **Réinitialiser** pour effacer les paramètres précédents dans le générateur de tableaux.
- Cliquez sur *Informations sur Internet* dans la liste des variables, puis sur *Informations à la télévision* dans la liste des variables tout en maintenant la touche Maj enfoncée afin de sélectionner toutes les variables relatives aux informations. (*Remarque* : Ceci implique que les libellés de variable soient affichés dans l'ordre alphabétique, et non dans l'ordre du fichier, dans la liste des variables.)
- Faites glisser les cinq variables relatives aux informations vers la zone Lignes du panneau de canevas.
Les cinq variables relatives aux informations sont empilées dans la dimension de ligne.
- Dans le panneau de canevas, cliquez sur *Informations sur Internet* de sorte que seule cette variable soit sélectionnée.
- Cliquez ensuite avec le bouton droit de la souris sur *Informations sur Internet* et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
- Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez *N % colonne* dans la liste Statistiques et cliquez sur la flèche pour ajouter l'élément à la liste Afficher. (Pour déplacer les statistiques sélectionnées de la liste Statistiques vers la liste Afficher, vous pouvez utiliser la flèche ou faire glisser les statistiques.)
- Cliquez ensuite sur **Appliquer à la sélection**.

Une colonne est ajoutée pour les pourcentages en colonne, mais l'aperçu du tableau dans le panneau de canevas indique que les pourcentages en colonne apparaîtront pour une seule variable. Ceci est dû au fait qu'un tableau empilé comporte plusieurs variables source de statistiques et que chacune d'elles peut présenter des statistiques récapitulatives différentes. Dans cet exemple, toutefois, nous souhaitons afficher les mêmes statistiques récapitulatives pour toutes les variables.

9. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Informations dans les journaux* et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
10. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez *N % colonne* dans la liste Statistiques et cliquez sur la flèche pour ajouter l'élément à la liste Afficher.
11. Cliquez ensuite sur **Appliquer à tout**.

L'aperçu du tableau indique désormais que les pourcentages en colonne apparaîtront pour toutes les variables empilées.

Autres statistiques récapitulatives en total pour les variables catégorielles

En ce qui concerne les variables source de statistiques catégorielles, vous pouvez inclure d'autres statistiques récapitulatives en total, différentes des statistiques affichées pour les catégories de la variable. Par exemple, pour une variable ordinale, vous pouvez afficher des pourcentages pour chaque catégorie, et la moyenne ou la médiane pour l'autre statistique récapitulative en total.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents dans le générateur de tableaux.
3. Cliquez sur *Confiance en la presse* dans la liste des variables, puis sur *Confiance en la télévision* tout en maintenant la touche Ctrl enfoncée afin de sélectionner les deux variables.
4. Faites glisser les deux variables vers la zone Lignes du panneau de canevas. Les deux variables sont ainsi empilées dans la dimension de ligne.
5. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur l'une des deux variables et choisissez l'option de menu **Sélectionner toutes les variables de ligne**. (Elles sont peut-être déjà sélectionnées toutes les deux, mais cette action permet de le garantir.)
6. Cliquez de nouveau avec le bouton droit de la souris sur la variable et sélectionnez **Catégories et totaux** dans le menu contextuel.
7. Dans la boîte de dialogue Catégories et totaux, cochez **Total**, puis cliquez sur **Appliquer**.
L'aperçu du tableau dans le panneau de canevas affiche désormais une ligne de total pour les deux variables. Pour afficher d'autres statistiques récapitulatives en total, vous devez indiquer des totaux et/ou des sous-totaux pour le tableau.
8. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur l'une des deux variables et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
9. Dans la boîte de dialogue Statistiques récapitulatives, cliquez sur *Effectif* dans la liste Afficher, puis sur la flèche pour déplacer l'élément vers la liste Statistiques (il disparaît de la liste Afficher).
10. Cliquez sur *N % colonne* dans la liste Statistiques et cliquez sur le bouton flèche pour déplacer cet élément vers la liste Afficher.
11. Cochez **Personnaliser les statistiques récapitulatives pour les totaux et les sous-totaux**.
12. Cliquez sur *Effectif* dans la liste Afficher relative aux autres statistiques récapitulatives, puis sur la flèche pour déplacer l'élément vers la liste Statistiques (il disparaît de la liste Afficher).
13. Cliquez sur *Moyenne* dans la liste Statistiques relative aux autres statistiques récapitulatives, puis sur la flèche pour déplacer l'élément vers la liste Afficher.
14. Cliquez sur la cellule **Format** pour la moyenne dans la liste Afficher et sélectionnez **nnnn** dans la liste déroulante des formats. (Vous pouvez être amené à parcourir la liste pour accéder à cet élément.)
15. Dans la cellule Décimales, tapez 2.

16. Cliquez sur **Appliquer à tout** pour appliquer ces paramètres aux deux variables dans le tableau.
Une nouvelle colonne a été ajoutée pour l'autre statistique récapitulative en total, mais peut-être n'est-ce pas ce que vous souhaitez, puisque l'aperçu du panneau de canevas indique clairement que ceci aboutira à un tableau contenant de nombreuses cellules vides.
17. Dans le générateur de tableaux, dans le groupe Statistiques récapitulatives, sélectionnez **Lignes** dans la liste déroulante Position.
Ceci déplace toutes les statistiques récapitulatives vers la dimension de ligne et les affiche dans une seule colonne du tableau.
18. Cliquez sur **OK** pour créer le tableau.

Confidence in press	A great deal	Column N %	9.5%
	Only some	Column N %	47.2%
	Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	A great deal	Column N %	10.4%
	Only some	Column N %	49.9%
	Hardly any	Column N %	39.7%
	Total	Mean	2.29

Figure 36. Variables catégorielles avec d'autres statistiques récapitulatives en total

Affichage des valeurs des catégories

Le tableau précédent pose un petit problème : il peut s'avérer difficile d'interpréter la valeur moyenne sans connaître les valeurs de catégorie sous-jacentes sur lesquelles elle est basée. Une moyenne de 2,34 se situe-t-elle quelque part entre *Beaucoup* et *Seulement certaines*, ou quelque part entre *Seulement certaines* et *Presque aucune* ?

Bien qu'il ne soit pas possible d'aborder ce problème directement dans Tableaux personnalisés, nous pouvons le traiter de façon plus générale.

1. A partir du menu, sélectionnez :
Edition > Options...
2. Dans la boîte de dialogue Options, cliquez sur l'onglet **Libellés de sortie**.
3. Dans le groupe Libellé tableau croisé dynamique, sélectionnez **Valeurs et libellés** dans la liste déroulante **Valeurs des variables en libellés montrés comme**.
4. Cliquez sur **OK** pour enregistrer ce paramètre.
5. Ouvrez le générateur de tableaux (menu Analyse, Tableaux (sélectionner tables et champs), Tableaux personnalisés), puis cliquez sur **OK** pour créer à nouveau le tableau.

Confidence in press	1 A great deal	Column N %	9.5%
	2 Only some	Column N %	47.2%
	3 Hardly any	Column N %	43.4%
	Total	Mean	2.34
Confidence in television	1 A great deal	Column N %	10.4%
	2 Only some	Column N %	49.9%
	3 Hardly any	Column N %	39.7%
	Total	Mean	2.29

Figure 37. Valeurs et libellés affichés pour les catégories de la variable

Les valeurs de catégorie indiquent clairement qu'une moyenne de 2,34 se situe quelque part entre *Seulement certaines* et *Presque aucune*. L'affichage des valeurs de catégorie dans le tableau facilite grandement l'interprétation de la valeur des autres statistiques récapitulatives en total, comme la moyenne.

Ce paramètre d'affichage est un paramètre global qui affecte toutes les sorties de tableau croisé dynamique de toutes les procédures, et qui est conservé lors des sessions suivantes jusqu'à ce que vous le modifiiez. Pour modifier le paramètre et afficher de nouveau uniquement les libellés de valeur :

6. A partir du menu, sélectionnez :
Edition > Options...
7. Dans la boîte de dialogue Options, cliquez sur l'onglet **Libellés de sortie**.
8. Dans le groupe Libellé tableau croisé dynamique, sélectionnez **Libellés** dans la liste déroulante **Valeurs des variables en libellés montrés comme**.
9. Cliquez sur **OK** pour enregistrer ce paramètre.

Chapitre 8. Récapitulatif des variables d'échelle

Récapitulatif des variables d'échelle

Une grande variété de statistiques récapitulatives sont disponibles pour les variables d'échelle. Outre les effectifs et les pourcentages disponibles pour les variables catégorielles, les statistiques récapitulatives pour les variables d'échelle incluent également :

- Moyenne
- Médiane
- Percentiles
- Somme
- Ecart type
- Plage
- Valeurs minimale et maximale

Pour plus d'informations, voir «Statistiques récapitulatives pour les variables d'échelle et autres statistiques récapitulatives en total pour les variables catégorielles», à la page 9.

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Tous les exemples fournis ici affichent des libellés de variable dans des boîtes de dialogue ; ces libellés sont triés par ordre alphabétique. Les propriétés d'affichage de la liste des variables sont spécifiées dans l'onglet Général de la boîte de dialogue Options (menu Modifier, Options).

Variables d'échelle empilées

Vous pouvez récapituler plusieurs variables d'échelle dans le même tableau en les empilant.

1. A partir du menu, sélectionnez :

Analyse > Tableaux > Tableaux personnalisés...

2. Dans le générateur de tableaux, cliquez sur *Age du répondant* dans la liste des variables, puis, tout en maintenant la touche Ctrl enfoncée, cliquez sur *Niveau d'étude le plus élevé* et sur *Heures par jour devant la télé* pour sélectionner les trois variables.

3. Faites glisser les trois variables sélectionnées vers la zone Lignes du panneau de canevas.

Les trois variables sont empilées dans la dimension de ligne. Les trois variables étant des variables d'échelle, aucune catégorie n'est affichée et la statistique récapitulative par défaut est la moyenne.

4. Cliquez sur **OK** pour créer le tableau.

	Mean
Age of respondent	46
Highest year of school completed	13
Hours per day watching TV	3

Figure 38. Tableau des valeurs moyennes des variables d'échelle empilées

Statistiques récapitulatives multiples

Par défaut, la moyenne apparaît pour les variables d'échelle ; toutefois, vous pouvez choisir d'autres statistiques récapitulatives pour les variables d'échelle et en afficher plusieurs.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur l'une des trois variables d'échelle dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
3. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez *Médiane* dans la liste Statistiques et cliquez sur la flèche pour ajouter cet élément à la liste Afficher. (Pour déplacer les statistiques sélectionnées de la liste Statistiques vers la liste Afficher, vous pouvez utiliser la flèche ou faire glisser les statistiques.)
4. Cliquez sur la cellule **Format** pour la médiane dans la liste Afficher et sélectionnez **nnnn** dans la liste déroulante des formats.
5. Dans la cellule Décimales, saisissez 1.
6. Apportez les mêmes modifications pour la moyenne dans la liste Afficher.
7. Cliquez sur **Appliquer à tout** pour appliquer ces modifications aux trois variables d'échelle.
8. Dans le générateur de tableaux, cliquez sur **OK** pour créer le tableau.

	Mean	Median
Age of respondent	45.6	42.0
Highest year of school completed	13.3	13.0
Hours per day watching TV	2.9	2.0

Figure 39. Moyenne et médiane affichées dans le tableau des variables d'échelle empilées

Effectif, N valide et valeurs manquantes

Il est souvent utile d'afficher le nombre d'observations utilisées pour calculer les statistiques récapitulatives, comme la moyenne, et vous pouvez raisonnablement penser que la statistique récapitulative *Effectif* est susceptible de fournir cette information. Toutefois, ceci n'offrira pas une base d'observations précise s'il existe des valeurs manquantes. Pour obtenir une base d'observations précise, utilisez la statistique *N valide*.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur l'une des trois variables d'échelle dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
3. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **Effectif** dans la liste Statistiques et cliquez sur la flèche pour ajouter cet élément à la liste Afficher.
4. Sélectionnez ensuite **N valide** dans la liste Statistiques et cliquez sur la flèche pour ajouter cet élément à la liste Afficher.
5. Cliquez sur **Appliquer à tout** pour appliquer ces modifications aux trois variables d'échelle.
6. Dans le générateur de tableaux, cliquez sur **OK** pour créer le tableau.

	Mean	Median	Count	Valid N
Age of respondent	45.6	42.0	2832	2828
Highest year of school completed	13.3	13.0	2832	2820
Hours per day watching TV	2.9	2.0	2832	2337

Figure 40. Effectif et N valide

Pour les trois variables, la statistique *Effectif* est identique (2 832). Il ne s'agit pas d'un hasard, puisqu'elle représente le nombre total d'observations dans le fichier de données. Puisque les variables d'échelle ne sont pas imbriquées dans des variables catégorielles, la statistique *Effectif* représente simplement le nombre total d'observations dans le fichier de données.

D'autre part, la statistique *N valide* est différente pour chaque variable et se détache de façon significative de la statistique *Effectif* pour *Heures par jour devant la télé*. En effet, il existe un grand nombre de **valeurs manquantes** pour cette variable, c'est-à-dire des observations sans valeur enregistrée pour cette variable ou des valeurs définies en tant que données manquantes (le code 99 peut représenter *Sans objet* dans le cas de grossesse pour hommes).

7. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
8. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur l'une des trois variables d'échelle dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
9. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **N valide** dans la liste Afficher, puis cliquez sur le bouton flèche pour renvoyer cet élément vers la liste Statistiques (il n'apparaît plus dans la liste Afficher).
10. Sélectionnez **Effectif** dans la liste Afficher, puis cliquez sur le bouton flèche pour renvoyer cet élément vers la liste Statistiques (il n'apparaît plus dans la liste Afficher).
11. Sélectionnez **Manquant** dans la liste Statistiques et cliquez sur le bouton flèche pour ajouter cet élément à la liste Afficher.
12. Cliquez sur **Appliquer à tout** pour appliquer ces modifications aux trois variables d'échelle.
13. Dans le générateur de tableaux, cliquez sur **OK** pour créer le tableau.

	Mean	Median	Missing
Age of respondent	45.6	42.0	4
Highest year of school completed	13.3	13.0	12
Hours per day watching TV	2.9	2.0	495

Figure 41. Nombre de valeurs manquantes affichées dans le tableau des statistiques récapitulatives de variable d'échelle

Le tableau indique désormais le nombre de valeurs manquantes pour chaque variable d'échelle. Il apparaît évident que la variable *Heures par jour devant la télé* présente un grand nombre de valeurs manquantes, tandis que les deux autres variables en présentent très peu. Il peut s'agir ici d'un facteur à prendre en compte avant d'accorder une grande confiance aux valeurs récapitulatives pour cette variable.

Différents récapitulatifs pour différentes variables

Outre l'affichage de plusieurs statistiques récapitulatives, vous pouvez afficher, dans un tableau empilé, différentes statistiques récapitulatives pour différentes variables d'échelle. Par exemple, le tableau précédent a montré qu'une seule des trois variables présentait un grand nombre de valeurs manquantes ; vous pouvez dès lors choisir d'afficher le nombre de valeurs manquantes uniquement pour cette variable.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez sur *Age du répondant* dans l'aperçu du tableau, puis cliquez sur *Niveau d'étude le plus élevé* tout en maintenant la touche Ctrl enfoncée pour sélectionner les deux variables.
3. Cliquez avec le bouton droit de la souris sur l'une des deux variables sélectionnées, puis choisissez **Statistiques récapitulatives** dans le menu contextuel.
4. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **Manquant** dans la liste Afficher, puis cliquez sur le bouton flèche pour renvoyer cet élément vers la liste Statistiques (il n'apparaît plus dans la liste Afficher).

5. Cliquez sur **Appliquer à la sélection** pour appliquer la modification uniquement aux deux variables sélectionnées.

Les valeurs de substitution dans les cellules de données du tableau indiquent que le nombre de valeurs manquantes apparaîtra uniquement pour la variable *Heures par jour devant la télé*.

6. Cliquez sur **OK** pour créer le tableau.

	Mean	Median	Missing
Age of respondent	45.6	42.0	
Highest year of school completed	13.3	13.0	
Hours per day watching TV	2.9	2.0	495

Figure 42. Tableau de différentes statistiques récapitulatives pour différentes variables

Bien que ce tableau fournisse les informations souhaitées, sa présentation peut compliquer son interprétation. Si vous visualisez le tableau, vous pouvez penser que les cellules vides de la colonne *Manquant* indique une absence de valeurs manquantes pour ces variables.

7. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
8. Dans le générateur de tableaux, dans le groupe Statistiques récapitulatives, sélectionnez **Lignes** dans la liste déroulante Position.
9. Cliquez sur **OK** pour créer le tableau.

Age of respondent	Mean	45.6
	Median	42.0
Highest year of school completed	Mean	13.3
	Median	13.0
Hours per day watching TV	Mean	2.9
	Median	2.0
	Missing	495

Figure 43. Statistiques récapitulatives et variables affichées dans la dimension de ligne

Il est désormais évident que le tableau fait état du nombre de valeurs manquantes pour une seule variable.

Récapitulatifs de groupe dans des catégories

Vous pouvez utiliser des variables catégorielles comme variables de regroupement afin d'afficher des récapitulatifs de variables d'échelle au sein de groupes définis par les catégories de la variable catégorielle.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Faites glisser la variable *Sexe* de la liste des variables vers la zone Colonnes du panneau de canevas.
Dans le panneau de canevas, si vous cliquez avec le bouton droit de la souris sur *Sexe* dans l'aperçu du tableau, vous remarquez que l'option **Statistiques récapitulatives** est désactivée dans le menu contextuel. En effet, dans un tableau comportant des variables d'échelle, ces variables sont toujours les variables source des statistiques.
3. Cliquez sur **OK** pour créer le tableau.

		Gender	
		Male	Female
Age of respondent	Mean	44.6	46.3
	Median	42.0	43.0
Highest year of school completed	Mean	13.4	13.2
	Median	13.0	13.0
Hours per day watching TV	Mean	2.8	2.9
	Median	2.0	2.0
	Missing	213	282

Figure 44. Récapitulatifs groupés de variable d'échelle avec une variable catégorielle de colonne

Ce tableau permet de comparer facilement les moyennes (moyenne et médiane) pour les hommes et les femmes, et indique clairement qu'il n'existe pas une grande différence entre elles (information peu pertinente au premier abord, mais susceptible d'être utile).

Variables de regroupement multiples

Vous pouvez diviser davantage les groupes en imbriquant, et/ou en utilisant des variables de regroupement catégorielles de ligne et de colonne.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Faites glisser *Informations sur Internet* de la liste des variables vers l'extrême gauche de la zone Lignes du panneau de canevas. Veillez à positionner cet élément de sorte que les trois variables d'échelle (et non pas seulement une) soient imbriquées dedans.

Bien que le second exemple présente un intérêt certain, ce n'est pas ce que nous allons choisir dans ce cas.

3. Cliquez sur **OK** pour créer le tableau.

				Gender	
				Male	Female
Get news from internet	No	Age of respondent	Mean	47.0	48.8
			Median	45.0	46.0
		Highest year of school completed	Mean	13.4	13.1
			Median	13.0	12.0
		Hours per day watching TV	Mean	3.2	3.4
			Median	2.0	3.0
	Yes	Age of respondent	Missing	213	282
			Mean	38.7	41.1
		Highest year of school completed	Median	35.0	38.0
			Mean	13.2	13.3
		Hours per day watching TV	Median	13.0	13.0
			Mean	2.1	2.1
		Median	2.0	2.0	
		Missing	0	0	

Figure 45. Récapitulatifs de variable d'échelle groupés par variables catégorielles de ligne et de colonne

Imbrication de variables catégorielles dans des variables d'échelle

Bien que le tableau précédent puisse fournir les informations souhaitées, le format de ces informations n'est pas le plus aisé pour l'interprétation. Par exemple, vous pouvez comparer l'âge moyen des hommes qui utilisent Internet pour s'informer et celui des hommes qui ne l'utilisent pas, mais cette comparaison serait facilitée si les valeurs se trouvaient côte à côte, plutôt que séparées. Vous pouvez améliorer le tableau en permutant les positions des deux variables de ligne et en imbriquant la variable de regroupement catégorielle dans les trois variables d'échelle. En ce qui concerne les variables d'échelle, le niveau d'imbrication n'a aucun effet sur la variable source des statistiques. La variable d'échelle est toujours la variable source des statistiques, quel que soit le niveau d'imbrication.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).

2. Dans le panneau de canevas, cliquez sur *Age du répondant* dans l'aperçu du tableau, puis, tout en maintenant la touche Ctrl enfoncée, cliquez sur *Niveau d'étude le plus élevé* et sur *Heures par jour devant la télé* pour sélectionner les trois variables d'échelle.
3. Faites glisser les trois variables d'échelle vers l'extrême gauche de la zone Lignes, ce qui a pour effet d'imbriquer la variable catégorielle *Informations sur Internet* dans chacune des trois variables d'échelle.
4. Cliquez sur **OK** pour créer le tableau.

				Gender	
				Male	Female
Age of respondent	Get news from internet	No	Mean	47.0	48.8
			Median	45.0	46.0
		Yes	Mean	38.7	41.1
			Median	35.0	38.0
Highest year of school completed	Get news from internet	No	Mean	13.4	13.1
			Median	13.0	12.0
		Yes	Mean	13.2	13.3
			Median	13.0	13.0
Hours per day watching TV	Get news from internet	No	Mean	3.2	3.4
			Median	2.0	3.0
			Missing	213	282
		Yes	Mean	2.1	2.1
			Median	2.0	2.0
			Missing	0	0

Figure 46. Variable de ligne catégorielle imbriquée dans les variables d'échelle empilées

Le choix de l'ordre d'imbrication dépend des relations ou des comparaisons à mettre en évidence dans le tableau. La modification de l'ordre d'imbrication des variables d'échelle n'implique pas la modification des valeurs des statistiques récapitulatives ; seule leur position relative dans le tableau est modifiée.

Chapitre 9. Statistiques de test

Statistiques de test

Trois différents tests de signification sont disponibles pour l'étude de la relation entre les variables en ligne et en colonne. Ce chapitre décrit la sortie de chacun de ces tests, notamment les effets de l'imbrication et de l'empilement. Pour plus d'informations, voir Chapitre 3, «Empilement, imbrication et couches avec des variables catégorielles», à la page 29.

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Tests d'indépendance (khi-deux)

Le test d'indépendance khi-deux sert à déterminer l'existence d'une relation entre deux variables catégorielles. Par exemple, vous pouvez déterminer si la variable *Situation d'activité* est liée à la variable *Situation familiale*.

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Dans le générateur de tableaux, faites glisser la variable *Situation d'activité* de la liste des variables vers la zone Lignes du panneau de canevas.
3. Faites glisser la variable *Situation familiale* de la liste des variables vers la zone Colonnes.
4. Sélectionnez **Lignes** comme position pour les statistiques récapitulatives.
5. Sélectionnez *Situation d'activité*, puis cliquez sur **Statistiques récapitulatives** dans le groupe Définir.
6. Dans la liste Statistiques, sélectionnez la statistique **N % colonne**, puis ajoutez-la à la liste Afficher.
7. Cliquez sur **Appliquer à la sélection**.
8. Dans la boîte de dialogue Tableaux personnalisés, cliquez sur l'onglet **Statistiques de test**.
9. Sélectionnez **Tests d'indépendance (Khi-deux)**.
10. Cliquez sur **OK** pour créer le tableau et obtenir le test du khi-deux.

			Marital status				
			Married	Widowed	Divorced	Separated	Never married
Labor force status	Working full time	Count	778	44	295	58	392
		Column %	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	Count	138	20	35	9	102
		Column %	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	Count	23	2	9	1	11
		Column %	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	Count	13	3	10	0	32
		Column %	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	Count	168	150	53	6	17
		Column %	12.5%	53.0%	11.9%	6.5%	2.6%
	School	Count	9	1	7	2	60
		Column %	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	Count	200	55	25	13	35
		Column %	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	Count	16	8	12	4	14
		Column %	1.2%	2.8%	2.7%	4.3%	2.1%

Figure 47. Variables Situation d'activité et Situation familiale

Ce tableau est un tableau croisé des variables *Situation d'activité* et *Situation familiale*, dans lequel les effectifs et les proportions de colonne constituent les statistiques récapitulatives. Les proportions de colonne sont calculées de sorte que leur somme soit égale à 100 % pour chaque colonne. Si ces deux variables ne sont pas liées, dans chaque ligne les proportions doivent être identiques d'une colonne à l'autre. Des différences semblent exister entre les proportions ; vous pouvez effectuer une vérification à l'aide du test du khi-deux.

		Marital status
Labor force status	Chi-square	729.242
	df	28
	Sig.	.000*

*. The Chi-square statistic is significant at the 0.05 level.

Figure 48. Test du khi-deux de Pearson

Le test d'indépendance part de l'hypothèse que les variables *Situation d'activité* et *Situation familiale* ne sont pas liées, c'est-à-dire que les proportions de colonne sont identiques d'une colonne à l'autre et que toute différence observée est due à une variation aléatoire. La statistique du khi-deux mesure la différence globale entre les effectifs de cellules observés et les effectifs attendus si les proportions de colonne étaient identiques d'une colonne à l'autre. Plus la valeur de la statistique du khi-deux est élevée, plus la différence entre les effectifs de cellules observés et théoriques est importante, et plus il apparaît que les proportions de colonne ne sont pas égales, que l'hypothèse d'indépendance est incorrecte et, par conséquent, que les variables *Situation d'activité* et *Situation familiale* sont liées.

La statistique du khi-deux donne la valeur 729,242. Pour déterminer si ceci suffit pour rejeter l'hypothèse d'indépendance, la valeur de signification de la statistique est calculée. La valeur de signification est la probabilité qu'une variable aléatoire issue d'une distribution khi-deux avec 28 degrés de liberté soit supérieure à 729,242. Dans la mesure où cette valeur est inférieure au niveau alpha spécifié dans l'onglet Statistiques de test, vous pouvez rejeter l'hypothèse d'indépendance au niveau 0,05. Par conséquent, les variables *Situation d'activité* et *Situation familiale* sont en fait liées.

Effets de l'imbrication et de l'empilement sur les tests d'indépendance

La règle suivante s'applique aux tests d'indépendance : un test distinct est exécuté pour chacune des sous-tables les plus internes. Pour voir l'impact de l'imbrication sur les tests, examinons l'exemple précédent, mais avec la variable *Situation familiale* imbriquée dans les niveaux de la variable *Sexe*.

1. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Faites glisser la variable *Sexe* de la liste des variables vers la zone Colonnes du panneau de canevas, au-dessus de la variable *Situation familiale*.
3. Cliquez sur **OK** pour créer le tableau.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^{*,1,2}	.000 ^{*,1,2}

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.
2. The minimum expected cell count in this sub-table is less than one.

Figure 49. Test du khi-deux de Pearson

Une fois la variable *Situation familiale* imbriquée dans les niveaux de la variable *Sexe*, deux tests sont exécutés (un pour chaque niveau de la variable *Sexe*). La valeur de signification de chaque test indique que vous pouvez rejeter l'hypothèse d'indépendance entre les variables *Situation familiale* et *Situation d'activité* pour les hommes et pour les femmes. Toutefois, le tableau indique que plus de 20 % des cellules de chaque tableau comportent des effectifs théoriques inférieurs à 5 et que l'effectif de cellule théorique minimum est inférieur à 1. Ces remarques montrent que les hypothèses du test du khi-deux ne sont pas vérifiées dans ces tableaux et que, par conséquent, les résultats des tests sont suspects.

Remarque : Les bordures des cellules peuvent masquer les notes de bas de page. Pour les afficher, vous pouvez modifier l'alignement des cellules concernées à l'aide de la boîte de dialogue Propriétés de la cellule.

Pour voir l'impact de l'empilement sur les tests :

4. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
5. Faites glisser la variable *Diplôme le plus important* de la liste des variables vers la zone Lignes, sous la variable *Situation d'activité*.
6. Cliquez sur **OK** pour créer le tableau.

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	Chi-square	246.637	542.589
	df	28	28
	Sig.	.000 ^{*,1,2}	.000 ^{*,1,2}
Highest degree	Chi-square	43.844	105.506
	df	16	16
	Sig.	.000*	.000*

*. The Chi-square statistic is significant at the 0.05 level.

1. More than 20% of cells in this sub-table have expected cell counts less than 5.
2. The minimum expected cell count in this sub-table is less than one.

Figure 50. Test du khi-deux de Pearson

Une fois la variable *Diplôme le plus important* empilée avec la variable *Situation d'activité*, quatre tests sont exécutés : un test d'indépendance des variables *Situation familiale* et *Situation d'activité*, et un test des variables *Situation familiale* et *Diplôme le plus important* pour chaque niveau de la variable *Sexe*. Les résultats des tests pour les variables *Situation familiale* et *Situation d'activité* sont les mêmes que précédemment. Les résultats des tests pour les variables *Situation familiale* et *Diplôme le plus important* indiquent que ces variables ne sont pas indépendantes.

Comparaison des moyennes de colonne

Les tests sur les moyennes de colonne permettent de déterminer l'existence d'une relation entre une variable catégorielle dans les colonnes et une variable continue dans les lignes. En outre, vous pouvez utiliser les résultats des tests pour déterminer le classement relatif des catégories de la variable catégorielle sur le plan de la valeur moyenne de la variable continue. Par exemple, supposons que vous souhaitiez déterminer si la variable *Heures par jour devant la télé* est liée à la variable *Informations dans les journaux*.

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Cliquez sur **Restaurer** pour rétablir les valeurs par défaut dans tous les onglets.
3. Dans le générateur de tableaux, faites glisser la variable *Heures par jour devant la télé* de la liste des variables vers la zone Lignes du panneau de canevas.
4. Faites glisser la variable *Informations dans les journaux* de la liste des variables vers la zone Colonnes.
5. Sélectionnez la variable *Heures par jour devant la télé*, puis cliquez sur **Statistiques récapitulatives** dans le groupe Définir.
6. Sélectionnez le format **nnnn**.
7. Sélectionnez **2** comme nombre de décimales à afficher. Suite à cette opération, le format devient **nnnn.nn**.
8. Cliquez sur **Appliquer à la sélection**.
9. Dans la boîte de dialogue Tableaux personnalisés, cliquez sur l'onglet **Statistiques de test**.
10. Sélectionnez **Comparer les moyennes de colonne (tests-t)**.
11. Cliquez sur **OK** pour créer le tableau et obtenir les tests des moyennes de colonne.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92	2.74

Figure 51. Variables *Informations dans les journaux* et *Heures par jour devant la télé*

Ce tableau indique la moyenne *Heures par jour devant la télé* pour les personnes qui s'informent ou non avec les journaux. La différence observée entre ces moyennes semble indiquer que les personnes qui ne s'informent pas avec les journaux passent environ 0,18 heure de plus devant leur poste de télévision que les autres. Pour voir si cette différence est due à une variation aléatoire, vérifiez les tests des moyennes de colonne.

	Get news from newspapers	
	No	Yes
	(A)	(B)
Hours per day watching TV		

Figure 52. Comparaisons des moyennes de colonne

Le tableau des tests des moyennes de colonne attribue une clé (lettre) à chaque catégorie de la variable en colonne. Pour la variable *Informations dans les journaux*, la catégorie *Non* se voit attribuer la lettre A, tandis que la lettre B est attribuée à la catégorie *Oui*. Pour chaque paire de colonnes, les moyennes de colonne sont comparées à l'aide d'un test *t*. Comme il n'y a que deux colonnes, un seul test est exécuté. Pour chaque paire significative, la clé de la catégorie dont la moyenne est la plus basse est placée sous la catégorie dont la moyenne est la plus élevée. Aucune clé n'étant reportée dans les cellules du tableau, les moyennes de colonne ne sont pas statistiquement différentes.

Résultats de signification dans la notation de style APA

Si vous ne souhaitez afficher les résultats de signification dans un tableau distinct, vous pouvez les afficher dans le tableau principal. Les résultats de signification sont identifiés à l'aide de la notation de style APA avec des indices. Réalisez les étapes précédentes pour comparer les moyennes des colonnes, mais effectuez la modification suivante sur l'onglet Statistiques de test.

1. Dans la zone Identifier les différences significatives, sélectionnez **Dans le tableau principal à l'aide d'indices de style APA**.
2. Cliquez sur **OK** pour créer le tableau et obtenir les tests des moyennes de colonne avec la notation de style APA.

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92 _a	2.74 _a

Figure 53. Comparaison des moyennes de colonnes à l'aide de la notation de style APA

Le tableau des tests des moyennes de colonne attribue un indice aux catégories de la variable de colonne. Pour chaque paire de colonnes, la moyenne des colonnes est comparée à l'aide d'un test *t*. Si une paire de valeurs est différente de manière significative, les valeurs se voient attribuées des indices *différents*. Comme il n'y a que deux colonnes, un seul test est exécuté. Les moyennes des colonnes de cet exemple partageant le même indice, il apparaît que les moyennes ne sont pas différentes.

Effets de l'imbrication et de l'empilement sur les tests des moyennes de colonne

La règle suivante s'applique aux tests des moyennes de colonne : un groupe distinct de tests par paire est exécuté pour chacune des sous-tables les plus internes. Pour voir l'impact de l'imbrication sur les tests, examinez l'exemple précédent, mais dans lequel la variable *Heures par jour devant la télé* est imbriquée dans les niveaux de la variable *Situation d'activité*.

1. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Faites glisser la variable *Situation d'activité* de la liste des variables vers la zone Lignes du panneau de canevas.
3. Cliquez sur **OK** pour créer le tableau.

			Get news from newspapers	
			No	Yes
			(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B	
	Working part-time	Hours per day watching TV		
	Temporarily not working	Hours per day watching TV		
	Unemployed, laid off	Hours per day watching TV		
	Retired	Hours per day watching TV		
	School	Hours per day watching TV		
	Keeping house	Hours per day watching TV		
	Other	Hours per day watching TV		

Figure 54. Comparaisons des moyennes de colonne

Une fois la variable *Heures par jour devant la télé* imbriquée dans les niveaux de la variable *Situation d'activité*, sept groupes de tests des moyennes de colonne sont exécutés : un pour chaque niveau de la variable *Situation d'activité*. Les mêmes clés (lettres) sont attribuées aux catégories de la variable *Informations dans les journaux*. Pour les répondants *qui travaillent à temps plein*, la clé B apparaît dans la colonne A. Par conséquent, pour les salariés à temps plein, la valeur moyenne de la variable *Heures par jour devant la télé* est inférieure pour les personnes qui s'informent avec les journaux. L'absence de toute autre clé dans les colonnes indique qu'il n'existe aucune autre différence statistiquement significative dans les moyennes de colonne.

Ajustements Bonferroni : Lorsque plusieurs tests sont exécutés, l'ajustement Bonferroni est appliqué aux tests des moyennes de colonne pour garantir que le niveau alpha (ou test du faux positif) spécifié dans l'onglet Statistiques de test s'applique à chaque *groupe* de tests. Par conséquent, dans ce tableau, aucun ajustement Bonferroni n'a été appliqué car, bien que sept groupes de tests soient exécutés, une seule paire de colonnes est comparée au sein de chaque groupe.

Pour voir l'impact de l'empilement sur les tests :

4. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
5. Faites glisser la variable *Informations sur Internet* de la liste des variables vers la zone Colonnes, à gauche de la variable *Informations dans les journaux*.
6. Cliquez sur **OK** pour créer le tableau.

			Get news from internet		Get news from newspapers	
			No	Yes	No	Yes
			(A)	(B)	(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B		B	
	Working part-time	Hours per day watching TV	B			
	Temporarily not working	Hours per day watching TV				
	Unemployed, laid off	Hours per day watching TV	B			
	Retired	Hours per day watching TV	B			
	School	Hours per day watching TV	B			
	Keeping house	Hours per day watching TV	B			
	Other	Hours per day watching TV	B			

Figure 55. Comparaisons des moyennes de colonne

Une fois la variable *Informations sur Internet* empilée avec la variable *Informations dans les journaux*, 14 groupes de tests des moyennes de colonne sont exécutés : un pour chaque niveau de la variable *Situation d'activité*, pour les variables *Informations sur Internet* et *Informations dans les journaux*. Là encore, aucun ajustement Bonferroni n'est appliqué car, au sein de chaque groupe, une seule paire de colonnes est comparée. Les tests pour la variable *Informations dans les journaux* sont les mêmes que précédemment. Pour la variable *Informations sur Internet*, la lettre A est attribuée à la catégorie *Non*, tandis que la lettre B est attribuée à la catégorie *Oui*. La clé B est reportée dans la colonne A pour chaque groupe de tests des moyennes de colonne, sauf pour les répondants temporairement sans activité. Par conséquent, la valeur moyenne de la variable *Heures par jour devant la télé* est plus faible pour les personnes qui s'informent au moyen d'Internet que pour celles qui ne s'informent pas avec les journaux. Aucune clé n'étant reportée pour le groupe *Temporairement sans activité*, les moyennes de colonne ne sont pas statistiquement différentes pour ces répondants.

Comparaison des proportions de colonne

Les tests des proportions de colonne permettent de déterminer le classement relatif des catégories de la variable catégorielle Colonnes sur le plan des proportions de catégorie de la variable catégorielle Lignes. Par exemple, après avoir utilisé un test du khi-deux révélant que les variables *Situation d'activité* et *Situation familiale* ne sont pas indépendantes, vous pouvez déterminer les lignes et les colonnes responsables de cette relation.

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Cliquez sur **Restaurer** pour rétablir les valeurs par défaut dans tous les onglets.
3. Dans le générateur de tableaux, faites glisser la variable *Situation d'activité* de la liste des variables vers la zone Lignes du panneau de canevas.
4. Faites glisser la variable *Situation familiale* de la liste des variables vers la zone Colonnes.
5. Sélectionnez *Situation d'activité*, puis cliquez sur **Statistiques récapitulatives** dans le groupe Définir.
6. Dans la liste Statistiques, sélectionnez la statistique **N % colonne**, puis ajoutez-la à la liste Afficher.
7. Désélectionnez **Effectif** dans la liste Afficher.
8. Cliquez sur **Appliquer à la sélection**.
9. Dans la boîte de dialogue Tableaux personnalisés, cliquez sur l'onglet **Statistiques de test**.

10. Sélectionnez **Comparer les proportions de colonne (tests-z)**.

11. Cliquez sur **OK** pour créer le tableau et obtenir les tests des proportions de colonne.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Column %	Column %	Column %	Column %	Column %
Labor force status	Working full time	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	12.5%	53.0%	11.9%	6.5%	2.6%
	School	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

Figure 56. Variables Situation d'activité et Situation familiale

Ce tableau est un tableau croisé des variables *Situation d'activité* et *Situation familiale*, dans lequel les proportions de colonne constituent la statistique récapitulative.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(A)	(B)	(C)	(D)	(E)
Labor force status	Working full time	B		A B	B	B
	Working part-time					A B C
	Temporarily not working					A B
	Unemployed, laid off					A B
	Retired	E	A C D E	E		
	School					A B C
	Keeping house	C E	C E		C E	
	Other					

Figure 57. Comparaisons des proportions de colonne

Le tableau des tests sur les proportions de colonne attribue une clé (lettre) à chaque catégorie des variables en colonne. Pour la variable *Situation familiale*, la lettre A est attribuée à la catégorie *Marié*, la lettre B à la catégorie *Veuf*, et ainsi de suite jusqu'à la catégorie *N'a jamais été marié*, à laquelle est affectée la lettre E. Pour chaque paire de colonnes, les proportions de colonne sont comparées à l'aide d'un test z. Sept groupes de tests des proportions de colonne sont exécutés : un pour chaque niveau de la variable *Situation d'activité*. La variable *Situation familiale* comportant cinq niveaux, $(5*4/2) = 10$ paires de colonnes sont comparées dans chaque groupe de tests et des ajustements Bonferroni permettent d'ajuster les valeurs de signification. Pour chaque paire significative, la clé de la catégorie dont la proportion est la plus basse est placée sous la catégorie dont la proportion est la plus élevée.

Pour le groupe de tests associé à la réponse *Salarié à temps plein*, la clé B apparaît dans chacune des autres colonnes. En outre, la clé A apparaît dans la colonne C. Aucune autre clé n'est reportée dans les autres colonnes. Par conséquent, vous pouvez arriver à la conclusion que la proportion des personnes divorcées travaillant à temps plein est supérieure à celle des personnes mariées travaillant à temps plein, elle-même supérieure à celle des personnes veuves travaillant à temps plein. La proportion des personnes séparées ou n'ayant jamais été mariées, et travaillant à temps plein, ne peut pas être différenciée de celle des personnes divorcées ou mariées, et travaillant à temps plein ; toutefois, cette proportion est supérieure à celle des personnes veuves travaillant à temps plein.

Pour les tests associés aux réponses *Salarié à temps partiel* ou *Etudiant*, les clés A, B et C apparaissent dans la colonne E. Aucune autre clé n'est reportée dans les autres colonnes. Par conséquent, la proportion des personnes qui n'ont jamais été mariées, et qui fréquentent une école ou travaillent à temps partiel, est supérieure à la proportion des personnes mariées, veuves ou divorcées qui fréquentent une école ou travaillent à temps partiel.

Pour les tests associés aux réponses *Temporairement sans activité* ou *Autre*, aucune autre clé n'est reportée dans les colonnes. Par conséquent, il n'y a pas de différence perceptible entre les proportions des personnes mariées, veuves, divorcées, séparées ou qui n'ont jamais été mariées, et qui sont sans emploi ou dans une situation d'activité non répertoriée.

Les tests associés à la réponse *Retraité* indiquent que la proportion des personnes veuves retraitées est supérieure aux proportions des personnes retraitées relevant des autres situations familiales. En outre, les proportions des personnes retraitées mariées ou divorcées sont supérieures à celle des personnes retraitées n'ayant jamais été mariées.

La proportion des personnes ayant répondu Homme/Femme au foyer est plus importante parmi les personnes mariées, veuves ou séparées que parmi les personnes divorcées ou n'ayant jamais été mariées.

La proportion des personnes qui n'ont jamais été mariées et qui ont répondu *Sans emploi* est supérieure à celles des personnes mariées ou veuves sans emploi. En outre, la colonne *Séparé* comporte le symbole " . ", qui indique que la proportion observée des personnes séparées dans la ligne *Sans emploi* est égale à 0 ou 1 et, par conséquent, que cette colonne ne permet pas d'effectuer des comparaisons pour les répondants sans emploi.

Résultats de signification dans la notation de style APA

Si vous ne souhaitez afficher les résultats de signification dans un tableau distinct, vous pouvez les afficher dans le tableau principal. Les résultats de signification sont identifiés à l'aide de la notation de style APA avec des indices. Réalisez les étapes précédentes pour comparer les proportions des colonnes, mais effectuez la modification suivante sur l'onglet Statistiques de test :

1. Dans la zone Identifier les différences significatives, sélectionnez **Dans le tableau principal à l'aide d'indices de style APA**.
2. Cliquez sur **OK** pour créer le tableau et obtenir les tests des moyennes de colonne avec la notation de style APA.

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Column N %	Column N %	Column N %	Column N %	Column N %
Labor force status	Working full time	57.8% _a	15.5% _b	66.1% _c	62.4% _{a,c}	59.1% _{a,c}
	Working part-time	10.3% _a	7.1% _a	7.8% _a	9.7% _{a,b}	15.4% _b
	Temporarily not working	1.7% _a	.7% _a	2.0% _a	1.1% _a	1.7% _a
	Unemployed, laid off	1.0% _a	1.1% _a	2.2% _{a,b}	.0%	4.8% _b
	Retired	12.5% _a	53.0% _b	11.9% _a	6.5% _{a,c}	2.6% _c
	School	.7% _a	.4% _a	1.6% _a	2.2% _{a,b}	9.0% _b
	Keeping house	14.9% _a	19.4% _a	5.6% _b	14.0% _a	5.3% _b
	Other	1.2% _a	2.8% _a	2.7% _a	4.3% _a	2.1% _a

Figure 58. Comparaison des proportions de colonnes à l'aide de la notation de style APA

Le tableau des tests des proportions de colonne attribue un indice aux catégories de la variable de colonne. Pour chaque paire de colonnes, les proportions des colonnes sont comparées à l'aide d'un test z. Si une paire de valeurs est différente de manière significative, les valeurs se voient attribuées des indices *différents*.

Pour le groupe de tests associés à *Salarié à temps plein*, la catégorie *Veuf* a un indice qui n'est pas utilisé dans les autres colonnes et les catégories *Séparé* et *Jamais marié* partagent les mêmes deux indices. Grâce à ces indices et aux proportions affichées dans le tableau, vous pouvez effectuer la même analyse que pour le premier exemple à base de tableaux distincts. Par conséquent, vous pouvez arriver à la conclusion que la proportion des personnes divorcées travaillant à temps plein est supérieure à celle des personnes mariées travaillant à temps plein, elle-même supérieure à celle des personnes veuves travaillant à temps plein. La proportion des personnes séparées ou n'ayant jamais été mariées, et travaillant à temps plein, ne peut pas être différenciée de celle des personnes divorcées ou mariées, et travaillant à temps plein ; toutefois, cette proportion est supérieure à celle des personnes veuves travaillant à temps plein. Le reste de l'analyse provenant de l'exemple précédent s'applique.

Effets de l'imbrication et de l'empilement sur les tests des proportions de colonne

La règle suivante s'applique aux tests des proportions de colonne : un groupe distinct de tests par paire est exécuté pour chacune des sous-tables les plus internes. Pour voir l'impact de l'imbrication sur les tests, examinez l'exemple précédent, mais avec la variable *Situation d'activité* imbriquée dans les niveaux de la variable *Sexe*.

1. Ouvrez à nouveau le générateur de tableaux (menu *Analyse, Tableaux, Tableaux personnalisés*).
2. Faites glisser la variable *Sexe* de la liste des variables vers la zone *Lignes* du panneau de canevas.
3. Cliquez sur **OK** pour créer le tableau.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Other				A			
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
Keeping house			C E	C E		C		
Other								

Figure 59. Comparaisons des proportions de colonne

Une fois la variable *Situation d'activité* imbriquée dans les niveaux de la variable *Sexe*, 14 groupes de tests des proportions de colonne sont exécutés : un pour chaque niveau de la variable *Situation d'activité*, pour chaque niveau de la variable *Sexe*. Les mêmes clés (lettres) sont attribuées aux catégories de la variable *Situation familiale*.

Les résultats du tableau font apparaître les deux points suivants :

- L'exécution d'un nombre supérieur de tests génère davantage de colonnes avec une proportion de colonne égale à zéro. Ceci concerne essentiellement les répondants séparés et les hommes veufs.

- Les différences de colonne précédemment observées parmi les répondants *Homme/Femme au foyer* semblent être entièrement imputables aux personnes de sexe féminin.

Pour voir l'impact de l'empilement sur les tests :

4. Ouvrez à nouveau le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
5. Faites glisser la variable *Diplôme le plus important* de la liste des variables vers la zone Lignes, sous la variable *Sexe*.
6. Cliquez sur **OK** pour créer le tableau.

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	ACDE	E		
			School					AC
			Keeping house					
	Female	Labor force status	Working full time	B		AB	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	ACDE	E		
			School					ABC
			Keeping house	CE	CE		C	
Highest degree	LT High school		ACE					
	High school							
	Junior college	B		B		B		
	Bachelor	B				B		
	Graduate	B						

Figure 60. Comparaisons des proportions de colonne

Une fois la variable *Diplôme le plus important* empilée avec la variable *Sexe*, 19 groupes de tests des moyennes de colonne sont exécutés : les 14 tests indiqués précédemment, ainsi qu'un test pour chaque niveau de la variable *Diplôme le plus important*. Les mêmes clés (lettres) sont attribuées aux catégories de la variable *Situation familiale*.

Les résultats du tableau font apparaître les points suivants :

- Les résultats des 14 groupes de tests précédemment exécutés sont les mêmes.
- La proportion des personnes dont le niveau d'études est inférieur à un diplôme d'école secondaire est plus élevée chez les personnes veuves que parmi les répondants mariés, divorcés ou n'ayant jamais été mariés.
- La proportion des personnes ayant suivi un enseignement post-secondaire semble plus importante parmi les personnes mariées, divorcées et n'ayant jamais été mariées que parmi les personnes veuves.

Une remarque sur les pondérations et les jeux de réponses multiples

Les pondérations sont toujours basées sur l'effectif et non sur les réponses, et ce même si l'une des variables est une variable multiréponses.

Chapitre 10. Jeux de réponses multiples

Les options Tableaux personnalisés et le Générateur de graphiques prennent également en charge un type spécifique de variable nommé **jeu de réponses multiples**. Les jeux de réponses multiples ne sont pas réellement des "variables" au sens habituel du terme. En effet, vous ne pouvez pas les voir dans l'éditeur de données et les autres procédures ne les reconnaissent pas. Les jeux de réponses multiples utilisent plusieurs variables pour enregistrer les réponses à des questions auxquelles le répondant peut donner plusieurs réponses. Les jeux de réponses multiples sont traités comme des variables catégorielles, et la plupart des actions appliquées à ces dernières peuvent également l'être aux jeux de réponses multiples.

Les jeux de réponses multiples sont construits à partir de plusieurs variables dans le fichier de données. Un jeu de réponses multiples est une structure spéciale dans un fichier de données. Vous pouvez définir et enregistrer les jeux de réponses multiples dans un fichier de données IBM® SPSS Statistics, mais vous ne pouvez ni importer ni exporter des jeux de réponses multiples à partir/vers d'autres formats de fichiers. (Vous pouvez copier les jeux de réponses multiples d'autres fichiers de données IBM SPSS Statistics via l'option Copier des propriétés de données accessibles à partir du menu Données dans la fenêtre de l'Editeur de données.

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Tous les exemples fournis ici affichent des libellés de variable dans des boîtes de dialogue ; ces libellés sont triés par ordre alphabétique. Les propriétés d'affichage de la liste des variables sont spécifiées dans l'onglet Général de la boîte de dialogue Options (menu Modifier, Options).

Effectifs, réponses, pourcentages et totaux

Toutes les statistiques récapitulatives disponibles pour les variables catégorielles le sont également pour les jeux de réponses multiples. Des statistiques supplémentaires sont également disponibles pour les jeux de réponses multiples.

1. A partir du menu, sélectionnez :

Analyse > Tableaux > Tableaux personnalisés...

2. Faites glisser *Sources d'informations* (libellé descriptive du jeu de réponses multiples *\$mltnews*) de la liste des variables vers la zone Lignes du panneau de canevas.

Dans la liste de variables, l'icône figurant en regard de la " variable " identifie celle-ci comme étant un jeu de dichotomies multiples.



Figure 61. Icône du jeu de dichotomies multiples

Pour un jeu de dichotomies multiples, chaque " catégorie " est en fait une variable distincte et les libellés de catégorie sont les libellés de variable (ou noms de variable pour les variables ne comportant aucun libellé défini). Dans cet exemple, les effectifs qui seront affichés représentent le nombre d'observations recevant la réponse *Oui* pour chaque variable du vecteur.

3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Sources d'informations* dans l'aperçu du tableau et sélectionnez **Catégories et totaux** dans le menu contextuel.

4. Dans la boîte de dialogue Catégories et totaux, sélectionnez **Total** en cliquant dessus, puis cliquez sur **Appliquer**.
5. Cliquez de nouveau avec le bouton droit de la souris sur *Sources d'informations* et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
6. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **N % colonne** dans la liste Statistiques et cliquez sur la flèche pour ajouter l'élément à la liste Afficher.
7. Cliquez sur **Appliquer à la sélection**, puis sur **OK** pour créer le tableau.

		Count	Column N %
News sources	Get news from internet	867	41.7%
	Get news from radio	551	26.5%
	Get news from television	1077	51.8%
	Get news from news magazines	294	14.1%
	Get news from newspapers	805	38.7%
	Total	2081	100.0%

Figure 62. Dichotomies multiples - Effectifs et pourcentages de colonne

Totaux ne pouvant pas s'additionner

Si vous examinez les chiffres dans le tableau, vous pouvez remarquer qu'il existe une différence relativement importante entre les " totaux " et les valeurs soi-disant totalisées ; plus précisément, les totaux sont beaucoup plus bas qu'ils ne devraient l'être. Ceci est dû au fait que l'effectif de chaque " catégorie " du tableau correspond au nombre d'observations dont la valeur est 1 (c'est-à-dire dont la réponse est *Oui*) pour cette variable, et que le nombre total de réponses *Oui* pour les cinq variables du jeu de dichotomies multiples pourrait facilement excéder le nombre total d'observations dans le fichier de données.

Toutefois, l'" effectif " total correspond au nombre total d'observations dont la réponse est *Oui* pour au moins une variable du vecteur, ce nombre ne pouvant jamais dépasser le nombre total d'observations dans le fichier de données. Dans cet exemple, le total de 2 081 est inférieur de presque 800 unités au nombre total d'observations dans le fichier de données. Si aucune de ces variables ne comporte de valeurs manquantes, cela signifie que près de 800 répondants ont indiqué qu'ils ne tirent leurs informations d'aucune de ces sources. Le total est la base des pourcentages en colonne ; par conséquent, dans cet exemple, la somme des pourcentages en colonne est supérieure aux 100 % affichés pour le pourcentage en colonne total.

Totaux pouvant s'additionner

Bien que le terme " effectif " n'est pas ambigu, l'exemple ci-dessus montre qu'il peut être source de confusion s'agissant des totaux dans les jeux de réponses multiples, pour lesquels la *réponse* est souvent la statistique récapitulative que vous recherchez vraiment.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Sources d'informations* dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
3. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **Réponses** dans la liste Statistiques et cliquez sur la flèche pour ajouter cet élément à la liste Afficher.
4. Dans la liste Statistiques, sélectionnez **% réponses colonne** et cliquez sur la flèche pour ajouter cet élément à la liste Afficher.
5. Cliquez sur **Appliquer à la sélection**, puis sur **OK** pour créer le tableau.

		Count	Column N %	Responses	Column Responses %
News sources	Get news from internet	867	41.7%	867	24.1%
	Get news from radio	551	26.5%	551	15.3%
	Get news from television	1077	51.8%	1077	30.0%
	Get news from news magazines	294	14.1%	294	8.2%
	Get news from newspapers	805	38.7%	805	22.4%
	Total	2081	100.0%	3594	100.0%

Figure 63. Dichotomies multiples - Réponses et réponses de la colonne en pourcentage

Pour chaque " catégorie " du jeu de dichotomies multiples, *Réponses* est identique à *Effectif* ; il en sera toujours ainsi dans les jeux de dichotomie multiples. Cependant, les totaux sont très différents. Le nombre total de réponses est de 3 594 ; il est donc supérieur de plus de 1 500 unités au total et de plus de 700 unités au nombre total d'observations dans le fichier de données.

En ce qui concerne les pourcentages, les totaux de *N % colonne* et de *% réponses colonne* sont tous deux de 100 % ; cependant, les pourcentages pour chaque catégorie du jeu de dichotomies multiples sont bien inférieurs pour les réponses de la colonne en pourcentage. Ceci est dû au fait que la base de pourcentage pour les réponses de la colonne en pourcentage est le nombre total de réponses (dans ce cas 3 594), ce qui donne des pourcentages bien inférieurs à la base des pourcentages en colonne, qui est de 2 081.

Totaux de pourcentages supérieurs à 100 %

Les pourcentages en colonne et les réponses de la colonne en pourcentage génèrent des totaux de 100 % même si, dans notre exemple, la somme des valeurs individuelles de la colonne *N % colonne* est largement supérieure à 100 %. Par conséquent, que se passe-t-il si vous souhaitez afficher les pourcentages sur la base du total plutôt que sur le nombre total de réponses, mais que vous souhaitez également que le pourcentage " total " reflète précisément la somme des pourcentages de chaque catégorie ?

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Sources d'informations* dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
3. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **% réponses colonne (base : Effectif)** dans la liste Statistiques et cliquez sur la flèche pour ajouter l'élément à la liste d'affichage.
4. Cliquez sur **Appliquer à la sélection**, puis sur **OK** pour créer le tableau.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
News sources	Get news from internet	867	41.7%	867	24.1%	41.7%
	Get news from radio	551	26.5%	551	15.3%	26.5%
	Get news from television	1077	51.8%	1077	30.0%	51.8%
	Get news from news magazines	294	14.1%	294	8.2%	14.1%
	Get news from newspapers	805	38.7%	805	22.4%	38.7%
	Total	2081	100.0%	3594	100.0%	172.7%

Figure 64. Réponses de la colonne en pourcentage, dont la base est l'effectif

Utilisation de jeux de réponses multiples avec d'autres variables

En général, vous pouvez utiliser plusieurs jeux de réponses multiples à l'instar des variables catégorielles. Par exemple, vous pouvez croiser un jeu de réponses multiples avec une variable catégorielle, ou imbriquer un jeu de réponses multiples dans une variable catégorielle.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).

- Faites glisser la variable *Sexe* de la liste des variables vers la gauche de la zone Lignes dans le panneau d'aperçu, imbriquant ainsi le jeu de réponses multiples *Sources d'informations* dans les catégories de la variable *Sexe*.
- Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Sexe* dans l'aperçu du tableau et désélectionnez **Afficher le libellé de la variable** dans le menu contextuel.
- Répétez l'opération pour *Sources d'informations*.
Cette opération entraîne la suppression des colonnes comportant les libellés de variable du tableau (puisqu'elles ne sont pas vraiment nécessaires dans ce cas).
- Cliquez sur **OK** pour créer le tableau.

		Count	Column N %	Responses	Column Responses %	Column Responses % (Base: Count)
Male	Get news from internet	359	40.1%	359	23.3%	40.1%
	Get news from radio	233	26.0%	233	15.1%	26.0%
	Get news from television	451	50.3%	451	29.3%	50.3%
	Get news from news magazines	121	13.5%	121	7.9%	13.5%
	Get news from newspapers	375	41.9%	375	24.4%	41.9%
	Total	896	100.0%	1539	100.0%	171.8%
Female	Get news from internet	508	42.9%	508	24.7%	42.9%
	Get news from radio	318	26.8%	318	15.5%	26.8%
	Get news from television	626	52.8%	626	30.5%	52.8%
	Get news from news magazines	173	14.6%	173	8.4%	14.6%
	Get news from newspapers	430	36.3%	430	20.9%	36.3%
	Total	1185	100.0%	2055	100.0%	173.4%

Figure 65. Jeu de réponses multiples imbriqué dans une variable catégorielle

Variable source de statistiques et statistiques récapitulatives disponibles

En l'absence d'une variable d'échelle dans un tableau, les variables catégorielles et les jeux de réponses multiples sont traités de la même manière en ce qui concerne la variable source de statistiques : La variable imbriquée la plus interne dans la dimension source des statistiques est la variable source de statistiques. Dans la mesure où certaines statistiques récapitulatives peuvent être affectées uniquement à des jeux de réponses multiples, le jeu de réponses multiples doit être la variable imbriquée la plus interne dans la dimension source des statistiques si vous souhaitez obtenir l'une des statistiques récapitulatives multiréponses spéciales.

- Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
- Dans l'aperçu du tableau du panneau de canevas, faites glisser *Sources d'informations* vers la gauche de *Sexe* pour modifier l'ordre d'imbrication.

Toutes les statistiques récapitulatives multiréponses spéciales (réponses, réponses de la colonne en pourcentage) sont supprimées de l'aperçu du tableau car la variable catégorielle *Sexe* est désormais la variable imbriquée la plus interne et, par conséquent, la variable source des statistiques.

Heureusement, le générateur de tableaux " se souvient " de ces paramètres. Si vous remplacez la variable *Sources d'informations* à sa position précédente, c'est-à-dire imbriquée dans la variable *Sexe*, toutes les statistiques récapitulatives liées aux réponses sont restaurées dans l'aperçu du tableau.

Jeu de catégories multiples et réponses en double

Les jeux de catégories multiples comportent une fonction non disponible pour les jeux de dichotomies multiples, à savoir, la fonction permettant de compter les réponses en double. Dans de nombreux cas, les réponses en double dans les jeux de catégories multiples représentent probablement des erreurs de codification. Par exemple, dans le cadre de la question " Selon vous, quels sont les trois pays à l'origine des meilleures voitures ? ", une réponse du type *Suède, Allemagne et Suède* n'est probablement pas valide.

Cependant, dans d'autres cas, des réponses en double peuvent être parfaitement valides. Par exemple, pour la question " Dans quels pays vos trois dernières voitures ont-elles été fabriquées ? ", une réponse du type *Suède, Allemagne et Suède* est parfaitement logique.

Dans les tableaux personnalisés, vous pouvez choisir les réponses en double dans les jeux de catégories multiples. Par défaut, les réponses en double ne sont pas comptées, mais vous pouvez demander à les inclure.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents.
3. Faites glisser *Constructeur automobile, voitures les plus récentes* de la liste des variables vers la zone Lignes du panneau de canevas.

Dans la liste des variables, l'icône figurant en regard de la " variable " identifie celle-ci comme étant un jeu de catégories multiples.



Figure 66. Icône du jeu de catégories multiples

Pour les jeux de catégories multiples, les catégories affichées représentent l'ensemble commun de libellés de valeurs définis pour toutes les variables du jeu (pour les jeux de dichotomies multiples, les " catégories " sont en réalité les libellés de variable de chaque variable du jeu).

4. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Constructeur automobile, voitures les plus récentes* dans l'aperçu du tableau et sélectionnez **Catégories et totaux** dans le menu contextuel.
5. Dans la boîte de dialogue Catégories et totaux, sélectionnez **Total** en cliquant dessus, puis cliquez sur **Appliquer**.
6. Cliquez de nouveau avec le bouton droit de la souris sur *Constructeur automobile, voitures les plus récentes* et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
7. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **Réponses** dans la liste Statistiques et cliquez sur la flèche pour ajouter cet élément à la liste Afficher.
8. Cliquez sur **Appliquer à la sélection**, puis sur **OK** pour créer le tableau.

		Count	Responses
Car maker, most recent cars	American	1938	1938
	Japanese	1327	1327
	Korean	695	695
	German	693	693
	Swedish	360	360
	Other	343	343
	Total	2832	5356

Figure 67. Jeu de catégories multiples : Effectifs et réponses sans doublons

Par défaut, les réponses en double ne sont pas comptées ; par conséquent, dans ce tableau, les valeurs de chaque catégorie dans les colonnes *Effectif* et *Réponses* sont identiques. Seuls les totaux diffèrent.

9. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
10. Cliquez sur l'onglet **Options**.
11. Cliquez sur **Dénombrer les réponses en double pour les jeux de catégories multiples**.
12. Cliquez sur **OK** pour créer le tableau.

		Count	Responses
Car maker, most recent cars	American	1938	2797
	Japanese	1327	1717
	Korean	695	760
	German	693	754
	Swedish	360	383
	Other	343	359
	Total	2832	6770

Figure 68. Jeu de catégories multiples dans lesquels les réponses en double sont incluses

Dans ce tableau, on remarque une différence notable entre les valeurs des colonnes *Effectif* et *Réponses*, en particulier pour les voitures américaines, ce qui indique que de nombreux répondants ont possédé plusieurs voitures américaines.

Test de signification pour les jeux de réponses multiples

Vous pouvez utiliser des jeux de réponses multiples dans des tests de signification de la même façon que vous utiliseriez des variables catégorielles.

- Pour les tests d'indépendance (khi-deux) ou pour la comparaison des proportions de colonne (tests-z), les tests sont réalisés sur la base des effectifs et l'Effectif doit être une des statistiques récapitulatives affichées dans le tableau.
- Pour les jeux de catégories multiples, les tests de comparaison des proportions de colonne ou des moyennes de colonnes (t-tests) ne sont pas réalisés si l'option **Dénombrer les réponses en double pour les jeux de catégories multiples** est sélectionnée dans l'onglet Options. Pour plus d'informations, voir «Tableaux personnalisés : onglet Options», à la page 16.

Tests d'indépendance avec des jeux de réponses multiples

Cet exemple crée un tableau croisé de variables catégorielles et un jeu de réponses multiples exécute et effectue un test d'indépendance khi-deux sur le tableau croisé.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents.
3. Faites glisser *Sources d'informations* (libellé descriptif du jeu de dichotomies multiples \$mltnews) de la liste des variables vers la zone Colonnes du panneau de canevas.
4. Faites glisser la variable *Sexe* de la liste des variables vers la zone Lignes du panneau de canevas.
5. Cliquez sur l'onglet **Statistiques de test**.
6. Sélectionnez (cochez) **Tests d'indépendance (khi-deux)**.
7. Si cela n'est pas déjà fait, sélectionnez **Inclure des variables de réponses multiples dans le test**.
8. Cliquez sur **OK** pour exécuter la procédure.

		News sources
Gender	Chi-square	10.266
	df	5
	Sig.	.068

Figure 69. Résultats khi-deux

Le niveau de signification de 0,068 pour le test khi-deux indique que les hommes et les femmes ne diffèrent probablement pas de manière significative dans leurs choix de nouvelles sources (en considérant que vous utilisez une valeur de signification de 0,05 ou inférieur comme critère de détermination de la signification statistique)

Comparaison des moyennes de colonne avec des jeux de réponses multiples

Cet exemple calcule les moyennes d'une variable d'échelle au sein des catégories définies par un jeu de réponses multiples et compare la moyenne de chaque catégorie avec celle des autres catégories afin de déterminer des différences significatives.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents.
3. Faites glisser *Sources d'informations* (libellé descriptif du jeu de dichotomies multiples \$mltnews) de la liste des variables vers la zone Colonnes du panneau de canevas.
4. Faites glisser *Age du répondant* vers la zone Lignes du panneau de canevas.
5. Cliquez sur l'onglet **Statistiques de test**.
6. Sélectionnez (cochez) **Comparer les moyennes de colonne (tests-t)**.
7. Si cela n'est pas déjà fait, sélectionnez **Inclure des variables de réponses multiples dans le test**.
8. Cliquez sur **OK** pour exécuter la procédure.

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	Mean	Mean	Mean	Mean	Mean
Age of respondent	52	40	48	40	40

Comparisons of Column Means

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	(A)	(B)	(C)	(D)	(E)
Age of respondent	B C D E		B D E		

Results are based on two-sided tests assuming equal variances with significance level 0.05. For each significant pair, the key of the smaller category appears under the category with larger mean.

Figure 70. Résultats du test de signification

- Chaque catégorie de jeu de réponses multiples est identifiée par une lettre (A, B, C, D, E) ; pour chaque catégorie pour laquelle la moyenne d'une autre catégorie est à la fois inférieure et diffère de manière significative de la moyenne de la catégorie en question, la lettre représentant la catégorie comportant la moyenne la plus basse est affichée.
- *Informations dans les journaux* (A) comprend l'âge moyen le plus élevé et toutes les autres catégories en diffèrent de manière significative.
- *Informations à la télévision* (C) comprend l'âge moyen suivant le plus élevé et toutes les autres moyennes de catégories (B, D et E) en diffèrent de manière significative. (C diffère également de manière significative de A, comme précédemment indiqué.)
- Les âges moyens pour *Informations dans les magazines* (B), *Informations à la radio* (D) et *Informations sur Internet* (E) ne diffèrent pas de manière significative les uns des autres.

Chapitre 11. Valeurs manquantes

De nombreux fichiers de données contiennent une certaine quantité de données manquantes. De nombreux facteurs peuvent être à l'origine des données manquantes. Par exemple, il arrive que des répondants ne répondent pas à chaque question, que certaines variables ne soient pas applicables à certaines observations et que des erreurs de codification aboutissent à la suppression de certaines valeurs.

Il existe deux sortes de valeur manquante dans IBM SPSS Statistics :

- **Valeur manquante de l'utilisateur** : Valeurs définies comme contenant des données manquantes. Des libellés de valeurs peuvent être affectés à ces valeurs pour identifier la raison de l'absence des données (par exemple, le code 99 et le libellé de valeur *Sans objet* pour la grossesse chez les hommes).
- **Valeur système manquante** : Si aucune valeur n'est présente pour une variable numérique, la valeur système manquante lui est affectée. Ceci est indiqué par un point dans la vue de données de l'éditeur de données.

Un ensemble de fonctionnalités permet de compenser les effets des données manquantes et d'analyser le motif de ces dernières. Toutefois, l'objet de ce chapitre est beaucoup plus simple : il s'agit de décrire la manière dont la fonction Tableaux personnalisés gère les données manquantes et la manière dont ces dernières affectent le calcul des statistiques récapitulatives.

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier *missing_values.sav*. Pour plus d'informations, voir Fichiers de données. Il s'agit d'un fichier extrêmement simple et complètement artificiel, comportant une seule variable et dix observations, conçu pour illustrer les concepts de base des valeurs manquantes.

Tableaux dépourvus de valeurs manquantes

Par défaut, les catégories manquantes de l'utilisateur n'apparaissent pas dans les tableaux personnalisés (et les valeurs système manquantes n'apparaissent jamais).

1. A partir du menu, sélectionnez :

Analyse > Tableaux > Tableaux personnalisés...

2. Dans le générateur de tableaux, faites glisser la variable *Variable avec valeurs manquantes* (la seule variable du fichier) de la liste des variables vers la zone Lignes du panneau de canevas.
3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur la variable et sélectionnez **Catégories et totaux** dans le menu contextuel.
4. Dans la boîte de dialogue Catégories et totaux, cliquez sur **Total**, puis sur **Appliquer**.
5. Dans le panneau de canevas, cliquez de nouveau avec le bouton droit de la souris sur *Variable avec valeurs manquantes* dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
6. Dans la boîte de dialogue Statistiques récapitulatives, sélectionnez **N % colonne** dans la liste Statistiques et cliquez sur la flèche pour ajouter l'élément à la liste Afficher.
7. Cliquez sur **Appliquer à la sélection**.

Vous pouvez noter une légère différence entre les catégories affichées dans l'aperçu du tableau dans le panneau de canevas et celles affichées dans la liste Catégories (sous la liste des variables, dans la partie gauche du générateur de tableaux). La liste Catégories contient une catégorie portant le libellé *Valeurs manquantes* qui n'est pas incluse dans l'aperçu du tableau, car les catégories de valeurs manquantes sont exclues par défaut. Puisque le terme " valeur " apparaît au pluriel dans le libellé, la variable comporte au moins deux catégories manquantes spécifiées par l'utilisateur.
8. Cliquez sur **OK** pour créer le tableau.

		Count	Column N %
Variable with missing values	Low	2	28.6%
	Medium	3	42.9%
	High	2	28.6%
	Total	7	100.0%

Figure 71. Tableau dépourvu de valeurs manquantes

Le contenu de ce tableau est parfait. Les valeurs des catégories s'additionnent pour donner des totaux et les pourcentages reflètent précisément les valeurs que vous obtiendriez si vous utilisiez le total comme base de pourcentage (par exemple, $3/7 = 0,429$ ou 42,9 %). Toutefois, le total n'est pas le nombre total d'observations dans le fichier de données, mais celui des observations comportant des valeurs **non manquantes**, ou des observations dépourvues de valeurs manquantes spécifiées par l'utilisateur ou de valeurs système manquantes pour cette variable.

Inclusion de valeurs manquantes dans les tableaux

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Variable avec valeurs manquantes* dans l'aperçu du tableau et sélectionnez **Catégories et totaux** dans le menu contextuel.
3. Dans la boîte de dialogue Catégories et totaux, cliquez sur **Valeurs manquantes**, puis sur **Appliquer**. L'aperçu du tableau comporte désormais une catégorie *Valeurs manquantes*. Bien que l'aperçu du tableau ne fasse apparaître qu'une catégorie pour les valeurs manquantes, toutes les catégories manquantes de l'utilisateur seront affichées dans le tableau.
4. Dans le panneau de canevas, cliquez de nouveau avec le bouton droit de la souris sur *Variable avec valeurs manquantes* dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
5. Dans la boîte de dialogue Statistiques récapitulatives, cliquez sur **Personnaliser les statistiques récapitulatives pour les totaux et les sous-totaux**.
6. Dans la liste Statistiques relative aux autres statistiques récapitulatives, sélectionnez **N valide**, puis cliquez sur la flèche pour ajouter cet élément à la liste Afficher.
7. Répétez l'opération pour la statistique **N total**.
8. Dans le générateur de tableaux, cliquez sur **Appliquer à la sélection**, puis sur **OK** pour créer le tableau.

		Count	Column N %	Valid N	Total N
Variable with missing values	Low	2	22.2%		
	Medium	3	33.3%		
	High	2	22.2%		
	Don't know	1	11.1%		
	Not applicable	1	11.1%		
	Total	9	100.0%	7	10

Figure 72. Tableau avec valeurs manquantes

Les deux catégories manquantes de l'utilisateur définies, à savoir *Ne sais pas* et *Sans objet*, apparaissent désormais dans le tableau, et le total est de 9 au lieu de 7, ce qui reflète l'ajout des deux observations comportant des valeurs manquantes de l'utilisateur (une par catégorie manquante de l'utilisateur). Les pourcentages en colonne sont également différents car ils sont désormais basés sur le nombre de valeurs non manquantes et de valeurs manquantes de l'utilisateur. Seules les valeurs système manquantes ne sont pas incluses dans le calcul des pourcentages.

La statistique *N valide* indique le nombre total d'observations non manquantes (7), tandis que la statistique *N total* indique le nombre total d'observations, y compris les observations manquantes spécifiées par l'utilisateur et les observations manquantes par défaut. Le nombre total d'observations

est 10, soit une de plus que le total des valeurs non manquantes et des valeurs manquantes de l'utilisateur affiché dans la colonne *Effectif*. En effet, une observation comporte une valeur système manquante.

9. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
10. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Variable avec valeurs manquantes* dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
11. Sélectionnez **N valide % colonne** dans la liste Statistiques supérieure (pas les récapitulatifs personnalisés des totaux et sous-totaux) et cliquez sur la flèche pour ajouter cet élément à la liste Afficher.
12. Répétez l'opération pour la statistique **N total % colonne**.
13. Vous pouvez également les ajouter à la liste des statistiques récapitulatives personnalisées des totaux et sous-totaux.
14. Cliquez sur **Appliquer à la sélection**, puis sur **OK** pour créer le tableau.

		Count	Column N %	Column Valid N %	Column Total N %	Valid N	Total N
Variable with missing values	Low	2	22.2%	28.6%	20.0%		
	Medium	3	33.3%	42.9%	30.0%		
	High	2	22.2%	28.6%	20.0%		
	Don't know	1	11.1%	.0%	10.0%		
	Not applicable	1	11.1%	.0%	10.0%		
	Total	9	100.0%	100.0%	100.0%	7	10

Figure 73. Tableau comportant des valeurs manquantes et des pourcentages totaux valides

- La statistique *N % colonne* est le pourcentage dans chaque catégorie sur la base du nombre de valeurs non manquantes et de valeurs manquantes de l'utilisateur (dans la mesure où les valeurs manquantes de l'utilisateur ont été explicitement incluses dans le tableau).
- La statistique *N valide % colonne* est le pourcentage dans chaque catégorie uniquement sur la base des observations valides non manquantes. Ces valeurs sont les mêmes que les pourcentages en colonne du tableau d'origine qui n'incluait pas les valeurs manquantes de l'utilisateur.
- La statistique *N total % colonne* est le pourcentage dans chaque catégorie sur la base de toutes les observations, y compris les observations manquantes spécifiées par l'utilisateur et les observations manquantes par défaut. Si vous ajoutez les différents pourcentages dans cette catégorie, vous obtenez seulement 90 % ; en effet, sur les 10 observations, une seule (soit 10 %) comporte la valeur système manquante. Bien que cette observation soit incluse dans la base du calcul des pourcentages, aucune catégorie n'est fournie dans le tableau pour les observations comportant des valeurs système manquantes.

Chapitre 12. Formatage et personnalisation de tableaux

Formatage et personnalisation de tableaux

La fonction Tableaux personnalisés permet de contrôler un certain nombre de propriétés de formatage de tableau dans le cadre de la génération de tableaux, notamment :

- Format d'affichage et libellés des statistiques récapitulatives
- Largeur minimum et maximum des colonnes de données
- Texte ou valeur apparaissant dans les cellules vides

Ces paramètres persistent dans l'interface de génération de tableaux (jusqu'à leur modification, la réinitialisation des paramètres du générateur de tableaux ou l'ouverture d'un autre fichier de données), ce qui vous permet de créer plusieurs tableaux dotés des mêmes propriétés de formatage sans devoir les modifier manuellement après leur création. Vous pouvez aussi enregistrer ces paramètres de formatage, ainsi que tous les autres paramètres du tableau, à l'aide du bouton Coller de l'interface de génération de tableaux pour coller la syntaxe de commande dans une fenêtre de syntaxe, que vous pouvez ensuite enregistrer sous la forme d'un fichier.

De plus, vous pouvez modifier de nombreuses propriétés de formatage d'un tableau nouvellement créé à l'aide de toutes les fonctionnalités de formatage disponibles dans le visualiseur pour les tableaux croisés dynamiques. Toutefois, ce chapitre est consacré au contrôle des propriétés de formatage d'un tableau avant sa création. Pour plus d'informations sur les tableaux croisés dynamiques, dans l'onglet Index du système d'aide, saisissez tableaux croisés dynamiques comme mot-clé.

Fichier de données d'exemple

Les exemples de ce chapitre utilisent le fichier de données *survey_sample.sav*. Pour plus d'informations, voir Fichiers de données.

Tous les exemples fournis ici affichent des libellés de variable dans des boîtes de dialogue ; ces libellés sont triés par ordre alphabétique. Les propriétés d'affichage de la liste des variables sont paramétrées dans l'onglet Général de la boîte de dialogue Options (menu Modifier, Options).

Formats d'affichage des statistiques récapitulatives

La fonction Tableaux personnalisés tente d'appliquer des formats par défaut relativement intelligents aux statistiques récapitulatives, mais vous pouvez ignorer ces valeurs par défaut.

1. A partir du menu, sélectionnez :
Analyse > Tableaux > Tableaux personnalisés...
2. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
3. Faites glisser la variable *Confiance en la télévision* sous la variable *Tranche d'âge* dans la zone Lignes, ce qui a pour effet d'empiler les deux variables dans la dimension de ligne.
4. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* dans l'aperçu du tableau et sélectionnez **Sélectionner toutes les variables de ligne** dans le menu contextuel.
5. Cliquez de nouveau avec le bouton droit de la souris sur *Tranche d'âge* et sélectionnez **Catégories et totaux** dans le menu contextuel.
6. Dans la boîte de dialogue Catégories et totaux, sélectionnez (cochez) **Total**, puis cliquez sur **Appliquer**.

7. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur une variable dans l'aperçu du tableau, puis sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
8. Sélectionnez % **colonne** dans la liste Statistiques et cliquez sur le bouton flèche pour ajouter cet élément à la liste Afficher.
9. Sélectionnez (cochez) **Personnaliser les statistiques récapitulatives pour les totaux et les sous-totaux**.
10. Dans la liste Statistiques des autres statistiques récapitulatives, sélectionnez **N % colonne** et cliquez sur la flèche pour ajouter cet élément à la liste Afficher.
11. Répétez l'opération pour la **moyenne**.
12. Cliquez ensuite sur **Appliquer à tout**.

Dans l'aperçu du tableau, les valeurs de substitution reflètent le format par défaut de chaque statistique récapitulative.

- Pour les effectifs, le format d'affichage par défaut est **nxxx**, c'est-à-dire des valeurs entières sans décimales.
- Pour les pourcentages, le format d'affichage par défaut est **nxxx.n%**, c'est-à-dire des nombres avec une décimale, suivis du signe %.
- Pour la moyenne, le format d'affichage par défaut est *différent* pour les deux variables.

Pour les statistiques récapitulatives qui ne sont ni un effectif (notamment N valide et N total), ni un pourcentage, le format d'affichage par défaut est le format défini pour la variable dans l'éditeur de données. Si vous observez les variables dans la vue de variable de l'éditeur de données, vous pouvez constater que la variable *Tranche d'âge (agecat)* est définie comme ayant deux décimales, tandis que la variable *Confiance en la télévision (variable contv)* est définie comme n'ayant aucune décimale.

Il s'agit de l'un des cas dans lesquels le format par défaut ne vous convient probablement pas ; il serait sûrement préférable que les deux valeurs de moyenne comportent le même nombre de décimales.

13. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur une variable dans l'aperçu du tableau, puis sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
Pour la moyenne, la cellule Format de la liste Afficher indique le format *Auto*, ce qui signifie que le format d'affichage défini pour la variable est utilisé, tandis que la cellule Décimales est désactivée. Pour spécifier le nombre de décimales, vous devez d'abord sélectionner un format différent.
14. Dans la liste Afficher des statistiques récapitulatives personnalisées, cliquez sur la cellule Format de la moyenne et sélectionnez **nxxx** dans la liste déroulante des formats.
15. Dans la cellule Décimales, saisissez la valeur 1.
16. Cliquez ensuite sur **Appliquer à tout** pour appliquer ce paramètre aux deux variables.

L'aperçu du tableau indique désormais que les deux valeurs de moyenne apparaîtront avec une décimale. (Vous pourriez continuer et créer ce tableau maintenant, mais la valeur " moyenne " de la variable *Tranche d'âge* peut vous sembler un peu difficile à interpréter car les codes numériques réels de cette variable sont uniquement compris entre 1 et 6.)

Affichage des libellés pour les statistiques récapitulatives

Outre les formats d'affichage des statistiques récapitulatives, vous pouvez contrôler les libellés descriptifs de chacune d'elles.

1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents dans le générateur de tableaux.
3. Dans le générateur de tableaux, faites glisser la variable *Tranche d'âge* de la liste des variables vers la zone Lignes du panneau de canevas.
4. Faites glisser la variable *Mode de paiement la semaine dernière* de la liste des variables vers la zone Colonnes du panneau de canevas.

- Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur *Tranche d'âge* dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
 - Sélectionnez % **colonne** dans la liste Statistiques et cliquez sur le bouton flèche pour ajouter cet élément à la liste Afficher.
 - Double-cliquez n'importe où sur le mot *Colonne* dans la cellule Libellé de la liste Afficher pour modifier le contenu de la cellule. Supprimez le mot *Colonne* du libellé afin que ce dernier se limite à %.
 - Modifiez la cellule Libellé de la statistique d'*effectif* de la même façon, de sorte que son libellé se limite à N.
- Nous allons modifier le format de la statistique N % colonne afin de supprimer le signe % superflu (puisque le libellé de la colonne indique que celle-ci contient des pourcentages).
- Cliquez sur la cellule Format de la statistique N % *colonne* et sélectionnez **nnnn.n** dans la liste déroulante des formats.
 - Cliquez ensuite sur **Appliquer à la sélection**.
- L'aperçu du tableau affiche les libellés et le format d'affichage modifiés.
- Cliquez sur **OK** pour créer le tableau.

		How get paid last week													
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate			
		N	%	N	%	N	%	N	%	N	%	N	%		
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7		
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1		
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7		
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7		
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5		
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3		

Figure 74. Tableau dans lequel les libellés des statistiques récapitulatives ont été modifiés

Largeur des colonnes

Vous avez peut-être remarqué que le tableau de l'exemple ci-dessus est assez volumineux. Une solution à ce problème consiste tout simplement à permuter les variables en ligne et les variables en colonne. Une autre solution consiste à réduire la largeur des colonnes, car celles-ci semblent beaucoup trop larges. (De fait, nous avons raccourci les libellés des statistiques récapitulatives afin de pouvoir réduire la largeur des colonnes.)

- Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
- Cliquez sur l'onglet **Options**.
- Dans le groupe Largeur des colonnes de données, sélectionnez **Personnalisé**.
- Dans la zone Maximum, saisissez 36. (Vérifiez que le paramètre Unités a pour valeur **Points**.)
- Cliquez sur **OK** pour créer le tableau.

		How get paid last week													
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate			
		N	%	N	%	N	%	N	%	N	%	N	%		
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7		
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1		
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7		
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7		
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5		
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3		

Figure 75. Tableau dans lequel la largeur des colonnes a été réduite

Le tableau est désormais beaucoup plus compact.

Affichage d'une valeur dans les cellules vides

Par défaut, un 0 apparaît dans les cellules vides (cellules dépourvues d'observations). Vous pouvez choisir de ne rien afficher dans ces cellules (les laisser vierges) ou indiquer la chaîne de caractères à afficher.

1. Ouvrez le générateur de tableaux (menu *Analyse, Tableaux, Tableaux personnalisés*).
2. Cliquez sur l'onglet **Options**.
3. Dans le groupe Apparence des cellules de données, pour Cellules vides, sélectionnez **Texte** et tapez *Aucune*.
4. Cliquez sur **OK** pour créer le tableau.

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	None	None	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	None	None	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

Figure 76. Tableau dont les cellules vides contiennent la mention "Aucune".

Les quatre cellules vides du tableau contiennent désormais le texte *Aucune* à la place de la valeur 0.

Affichage d'une valeur pour les statistiques manquantes

Si une statistique ne peut pas être calculée, la valeur apparaissant par défaut est un point (.), symbole utilisé pour indiquer la valeur système manquante. Ceci est différent d'une cellule "vide" ; par conséquent, la valeur apparaissant pour les statistiques manquantes est contrôlée indépendamment de celle des cellules dépourvues d'observations.

1. Ouvrez le générateur de tableaux (menu *Analyse, Tableaux, Tableaux personnalisés*).
2. Faites glisser la variable *Heures par jour devant la télé* de la liste de variables vers le sommet de la zone Colonnes dans le canevas, au-dessus de la variable *Mode de paiement la semaine dernière*.
La variable *Heures par jour devant la télé* étant une variable d'échelle, elle devient automatiquement la variable source des statistiques et la moyenne devient la statistique récapitulative.
3. Dans le panneau de canevas, cliquez avec le bouton droit de la souris sur la variable *Heures par jour devant la télé* dans l'aperçu du tableau et sélectionnez **Statistiques récapitulatives** dans le menu contextuel.
4. Dans la liste Statistiques, sélectionnez **N valide**, puis cliquez sur le bouton flèche pour ajouter cet élément à la liste Afficher.
5. Cliquez sur **Appliquer à la sélection**.
6. Cliquez sur l'onglet **Options**.
7. Dans le champ de texte correspondant à l'élément Statistiques impossibles à calculer, saisissez *N/A*.
8. Cliquez sur **OK** pour créer le tableau.

		Hours per day watching TV											
		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N	Mean	Valid N
Age category	Less than 25	3	71	NA	None	3	10	2	3	2	6	2	8
	25 to 34	3	134	5	2	2	30	2	29	2	52	2	22
	35 to 44	3	136	2	5	3	30	2	34	2	47	3	46
	45 to 54	2	90	2	4	2	22	2	36	2	45	2	34
	55 to 64	3	40	NA	None	3	7	2	15	2	23	3	15
	65 or older	3	18	2	2	1	1	NA	0	1	2	3	11

Figure 77. Tableau dans lequel la mention "N/A" apparaît pour les statistiques manquantes

Le texte N/A apparaît pour la moyenne dans trois cellules du tableau. Dans chaque cas, la valeur *N valide* correspondante explique pourquoi : Aucune observation ne permet de calculer la moyenne.

Toutefois, vous pouvez remarquer une légère différence ; l'une des trois valeurs N valides est exprimée par 0, et non par le libellé *Aucune* supposé apparaître dans les cellules dépourvues d'observations. En effet, malgré l'absence d'observations permettant de calculer la moyenne, la catégorie n'est pas vraiment vide. Si vous revenez au tableau d'origine comportant uniquement les deux variables catégorielles, vous pouvez constater que cette catégorie croisée comporte en réalité trois observations. Toutefois, aucune des observations n'est valide, car toutes trois comportent des valeurs manquantes pour la variable d'échelle *Heures par jour devant la télé*.

Chapitre 13. Fichiers d'exemple

Les fichiers d'exemple installés avec le produit figurent dans le sous-répertoire *Echantillons* du répertoire d'installation. Il existe un dossier distinct au sein du sous-répertoire *Echantillons* pour chacune des langues suivantes : anglais, français, allemand, italien, japonais, coréen, polonais, russe, chinois simplifié, espagnol et chinois traditionnel.

Seuls quelques fichiers d'exemples sont disponibles dans toutes les langues. Si un fichier d'exemple n'est pas disponible dans une langue, le dossier de langue contient la version anglaise du fichier d'exemple.

Descriptions

Voici de brèves descriptions des fichiers d'exemple utilisés dans divers exemples à travers la documentation.

- **accidents.sav** : Ce fichier de données d'hypothèse concerne une société d'assurance qui étudie les facteurs de risque liés à l'âge et au sexe dans les accidents de la route survenant dans une région donnée. Chaque observation correspond à une classification croisée de la catégorie d'âge et du sexe.
- **adl.sav** : Ce fichier de données d'hypothèse concerne les mesures entreprises pour identifier les avantages d'un type de thérapie proposé aux patients qui ont subi une attaque cardiaque. Les médecins ont assigné de manière aléatoire les patients du sexe féminin ayant subi une attaque cardiaque à un groupe parmi deux groupes possibles. Le premier groupe a fait l'objet de la thérapie standard tandis que le second a bénéficié en plus d'une thérapie émotionnelle. Trois mois après les traitements, les capacités de chaque patient à effectuer les tâches ordinaires de la vie quotidienne ont été notées en tant que variables ordinales.
- **advert.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un détaillant pour examiner la relation existant entre l'argent dépensé dans la publicité et les ventes résultantes. Pour ce faire, il collecte les chiffres des ventes passées et les coûts associés à la publicité.
- **aflatoxin.sav** : Ce fichier de données d'hypothèse concerne le test de l'aflatoxine dans des récoltes de maïs. La concentration de ce poison varie largement d'une récolte à l'autre et au sein de chaque récolte. Un processeur de grain a reçu 16 échantillons issus de 8 récoltes de maïs et a mesuré les niveaux d'aflatoxine en parties par milliard (PPB).
- **anorectic.sav** : En cherchant à développer une symptomatologie standardisée du comportement anorexique/boulimique, des chercheurs¹ ont examiné 55 adolescents souffrant de troubles alimentaires. Chaque patient a été observé quatre fois sur une période de quatre années, soit un total de 220 observations. A chaque observation, les patients ont été notés pour chacun des 16 symptômes. En raison de l'absence de scores de symptôme pour le patient 71/visite 2, le patient 76/visite 2 et le patient 47/visite 3, le nombre d'observations valides est de 217.
- **bankloan.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une banque pour réduire le taux de défaut de paiement. Il contient des informations financières et démographiques sur 850 clients existants et éventuels. Les premières 700 observations concernent des clients auxquels des prêts ont été octroyés. Les 150 dernières observations correspondant aux clients éventuels que la banque doit classer comme bons ou mauvais risques de crédit.
- **bankloan_binning.sav** : Ce fichier de données d'hypothèse concerne des informations financières et démographiques sur 5 000 clients existants.
- **behavior.sav** : Dans un exemple classique², il a été demandé à 52 étudiants de noter les combinaisons établies à partir de 15 situations et de 15 comportements sur une échelle de 0 à 9, où

1. Van der Ham, T., J. J. Meulman, D. C. Van Strien, and H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

2. Price, R. H., and D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579-586.

0 = "extrêmement approprié" et 9 = "extrêmement inapproprié". En effectuant la moyenne des résultats de l'ensemble des individus, on constate une certaine différence entre les valeurs.

- **behavior_ini.sav** : Ce fichier de données contient la configuration initiale d'une solution bidimensionnelle pour *behavior.sav*.
- **brakes.sav** : Ce fichier de données d'hypothèse concerne le contrôle qualité effectué dans une usine qui fabrique des freins à disque pour des voitures haut de gamme. Le fichier de données contient les mesures de diamètre de 16 disques de 8 machines de production. Le diamètre cible des freins est de 322 millimètres.
- **breakfast.sav** : Au cours d'une étude classique ³, on a demandé à 21 étudiants en MBA (Master of Business Administration) de l'école de Wharton et à leurs conjoints de classer 15 aliments du petit-déjeuner selon leurs préférences, de 1= « aliment préféré » à 15= « aliment le moins apprécié ». Leurs préférences ont été enregistrées dans six scénarios différents, allant de "Préférence générale" à "En-cas avec boisson uniquement".
- **breakfast-overall.sav** : Ce fichier de données contient les préférences de petit-déjeuner du premier scénario uniquement, "Préférence générale".
- **broadband_1.sav** : Ce fichier de données d'hypothèse concerne le nombre d'abonnés, par région, à un service haut débit. Le fichier de données contient le nombre d'abonnés mensuels de 85 régions sur une période de quatre ans.
- **broadband_2.sav** : Ce fichier de données est identique au fichier *broadband_1.sav* mais contient les données relatives à trois mois supplémentaires.
- **car_insurance_claims.sav** : Il s'agit d'un jeu de données présenté et analysé ailleurs ⁴ qui concerne des actions en indemnisation pour des voitures. Le montant d'action en indemnisation moyen peut être modélisé comme présentant une distribution gamma, à l'aide d'une fonction de lien inverse pour associer la moyenne de la variable dépendante à une combinaison linéaire de l'âge de l'assuré, du type de véhicule et de l'âge du véhicule. Le nombre d'actions entreprises peut être utilisé comme pondération de positionnement.
- **car_sales.sav** : Ce fichier de données contient des estimations de ventes hypothétiques, des barèmes de prix et des spécifications physiques concernant divers modèles et marques de véhicule. Les barèmes de prix et les spécifications physiques proviennent tour à tour de *edmunds.com* et des sites des constructeurs.
- **car_sales_upprepared.sav** : Il s'agit d'une version modifiée de *car_sales.sav* qui n'inclut aucune version transformée des champs.
- **carpet.sav** : Dans un exemple courant ⁵, une société intéressée par la commercialisation d'une nouvelle shampouineuse souhaite examiner l'influence de cinq facteurs sur la préférence du consommateur : le plan du pack, la marque, le prix, une étiquette *Economique* et une garantie Satisfait ou remboursé. Il existe trois niveaux de facteurs pour le plan du pack, suivant l'emplacement de l'applicateur, trois marques (*K2R*, *Glory* et *Bissell*), trois niveaux de prix et deux niveaux (non ou oui) pour chacun des deux derniers facteurs. Dix consommateurs classent 22 profils définis par ces facteurs. La variable *Préférence* indique le classement des classements moyens de chaque profil. Un classement faible correspond à une préférence élevée. Cette variable reflète une mesure globale de préférence pour chaque profil.
- **carpet_prefs.sav** : Ce fichier de données repose sur le même exemple que celui décrit pour *carpet.sav*, mais contient les classements réels issus de chacun des 10 clients. On a demandé aux consommateurs de classer les 22 profils de produits, du préféré au moins intéressant. Les variables *PREF1* à *PREF22* contiennent les identificateurs des profils associés, tels qu'ils sont définis dans *carpet_plan.sav*.
- **catalog.sav** : Ce fichier de données contient des chiffres de ventes mensuelles hypothétiques relatifs à trois produits vendus par une entreprise de vente par correspondance. Les données relatives à cinq variables de prédicteur possibles sont également incluses.

3. Green, P. E., and V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

4. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

5. Green, P. E., and Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

- **catalog_seasfac.sav** : Ce fichier de données est identique à *catalog.sav* mais contient en plus un ensemble de facteurs saisonniers calculés à partir de la procédure de désaisonnalisation, ainsi que les variables de date correspondantes.
- **cellular.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un opérateur téléphonique pour réduire les attritions. Des scores de propension à l'attrition sont attribués aux comptes, de 0 à 100. Les comptes ayant une évaluation égale ou supérieure à 50 sont susceptibles de changer de fournisseur.
- **ceramics.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un fabricant pour déterminer si un nouvel alliage haute qualité résiste mieux à la chaleur qu'un alliage standard. Chaque observation représente un test séparé de l'un des deux alliages ; le degré de chaleur auquel l'alliage ne résiste pas est enregistré.
- **cereal.sav** : Ce fichier de données d'hypothèse concerne un sondage de 880 personnes interrogées sur leurs préférences de petit-déjeuner et sur leur âge, leur sexe, leur situation familiale et leur mode de vie (actif ou non actif, selon qu'elles pratiquent une activité physique au moins deux fois par semaine). Chaque observation correspond à un répondant distinct.
- **clothing_defects.sav** : Ce fichier de données d'hypothèse concerne le processus de contrôle qualité observé dans une usine de textile. Dans chaque lot produit à l'usine, les inspecteurs prélèvent un échantillon de vêtements et comptent le nombre de vêtements qui ne sont pas acceptables.
- **coffee.sav** : Ce fichier de données concerne l'image perçue de six marques de café frappé⁶. Pour chacun des 23 attributs d'image de café frappé, les personnes sollicitées ont sélectionné toutes les marques décrites par l'attribut. Les six marques sont appelées AA, BB, CC, DD, EE et FF à des fins de confidentialité.
- **contacts.sav** : Ce fichier de données d'hypothèse concerne les listes de contacts d'un groupe de représentants en informatique d'entreprise. Chaque contact est classé selon le service de l'entreprise où il travaille et le classement de son entreprise. Sont également enregistrés le montant de la dernière vente effectuée, le temps passé depuis la dernière vente et la taille de l'entreprise du contact.
- **creditpromo.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un grand magasin pour évaluer l'efficacité d'une promotion récente de carte de crédit. A cette fin, 500 détenteurs de carte ont été sélectionnés au hasard. La moitié a reçu une publicité faisant la promotion d'un taux d'intérêt réduit sur les achats effectués dans les trois mois à venir. L'autre moitié a reçu une publicité saisonnière standard.
- **customer_dbase.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une société pour utiliser les informations figurant dans sa banque de données et proposer des offres spéciales aux clients susceptibles d'être intéressés. Un sous-groupe de la base de clients a été sélectionné au hasard et a reçu des offres spéciales. Les réponses des clients ont été enregistrées.
- **customer_information.sav** : Un fichier de données d'hypothèse qui contient les informations postales du client, telles que le nom et l'adresse.
- **customer_subset.sav** : Un sous-ensemble de 80 observations de *customer_dbase.sav*.
- **debate.sav** : Ce fichier de données d'hypothèse concerne des réponses appariées à une enquête donnée aux participants à un débat politique avant et après le débat. Chaque observation représente un répondant distinct.
- **debate_aggregate.sav** : Il s'agit d'un fichier de données d'hypothèse qui rassemble les réponses dans le fichier *debate.sav*. Chaque observation correspond à une classification croisée de préférence avant et après le débat.
- **demo.sav** : Ce fichier de données d'hypothèse concerne une base de données clients achetée en vue de diffuser des offres mensuelles. Les données indiquent si le client a répondu ou non à l'offre et contiennent diverses informations démographiques.

6. Kennedy, R., C. Riquier, and B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56-70.

- **demo_cs_1.sav** : Ce fichier de données d'hypothèse concerne la première mesure entreprise par une société pour compiler une base de données contenant des informations d'enquête. Chaque observation correspond à une ville différente. La région, la province, le quartier et la ville sont enregistrés.
- **demo_cs_2.sav** : Ce fichier de données d'hypothèse concerne la seconde mesure entreprise par une société pour compiler une base de données contenant des informations d'enquête. Chaque observation correspond à un ménage différent issu des villes sélectionnées à la première étape. La région, la province, le quartier, la ville, la sous-division et l'identification sont enregistrés. Les informations d'échantillonnage des deux premières étapes du plan sont également incluses.
- **demo_cs.sav** : Ce fichier de données d'hypothèse concerne des informations d'enquête collectées via un plan complexe d'échantillonnage. Chaque observation correspond à un ménage différent et diverses informations géographiques et d'échantillonnage sont enregistrées.
- **dmdata.sav** : Il s'agit d'un fichier de données d'hypothèse qui contient des informations démographiques et des informations concernant les achats pour une entreprise de marketing direct. *dmdata2.sav* contient les informations pour un sous-ensemble de contacts qui ont reçu un envoi d'essai, et *dmdata3.sav* contient des informations sur les contacts restants qui n'ont pas reçu l'envoi d'essai.
- **dietstudy.sav** : Ce fichier de données d'hypothèse contient les résultats d'une étude portant sur le régime de Stillman⁷. Chaque observation correspond à un sujet distinct et enregistre son poids en livres avant et après le régime, ainsi que ses niveaux de triglycérides en mg/100 ml.
- **dvdplayer.sav** : Ce fichier de données d'hypothèse concerne le développement d'un nouveau lecteur DVD. À l'aide d'un prototype, l'équipe de marketing a collecté des données de groupes spécifiques. Chaque observation correspond à un utilisateur interrogé et enregistre des informations démographiques sur cet utilisateur, ainsi que ses réponses aux questions portant sur le prototype.
- **german_credit.sav** : Ce fichier de données provient du jeu de données "German credit" figurant dans le référentiel Machine Learning Databases⁸ de l'université de Californie, Irvine.
- **grocery_1month.sav** : Ce fichier de données d'hypothèse est le fichier de données *grocery_coupons.sav* dans lequel les achats hebdomadaires sont organisés par client distinct. Certaines variables qui changeaient toutes les semaines disparaissent. En outre, le montant dépensé enregistré est à présent la somme des montants dépensés au cours des quatre semaines de l'enquête.
- **grocery_coupons.sav** : Il s'agit d'un fichier de données d'hypothèse qui contient des données d'enquête collectées par une chaîne de magasins d'alimentation qui cherche à déterminer les habitudes de consommation de ses clients. Chaque client est suivi pendant quatre semaines et chaque observation correspond à une semaine distincte. Les informations enregistrées concernent les endroits où le client effectue ses achats, la manière dont il les effectue, ainsi que les sommes dépensées en provisions au cours de cette semaine.
- **guttman.sav** : Bell⁹ a présenté un tableau pour illustrer les groupes sociaux possibles. Guttman¹⁰ a utilisé une partie de ce tableau, dans lequel cinq variables décrivant des éléments tels que l'interaction sociale, le sentiment d'appartenance à un groupe, la proximité physique des membres et la formalité de la relation, ont été croisées avec sept groupes sociaux théoriques, dont les foules (par exemple, le public d'un match de football), l'audience (par exemple, au cinéma ou dans une salle de classe), le public (par exemple, les journaux ou la télévision), les bandes (proche d'une foule, mais qui serait caractérisée par une interaction beaucoup plus intense), les groupes primaires (intimes), les groupes secondaires (volontaires) et la communauté moderne (groupement lâche issu d'une forte proximité physique et d'un besoin de services spécialisés).
- **health_funding.sav** : Ce fichier de données d'hypothèse concerne des données sur le financement des soins de santé (montant par groupe de 100 individus), les taux de maladie (taux par groupe de

7. Rickman, R., N. Mitchell, J. Dingman, and J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228:, 54-58.

8. Blake, C. L., and C. J. Merz. 1998. "UCI Repository of machine learning databases." Disponible sur la page <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

9. Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

10. Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469-506.

- 10 000 individus) et les visites chez les prestataires de soins de santé (taux par groupe de 10 000 individus). Chaque observation représente une ville différente.
- **hivassay.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un laboratoire pharmaceutique pour développer une analyse rapide de détection d'infection HIV. L'analyse a pour résultat huit nuances de rouge, les nuances les plus marquées indiquant une plus forte probabilité d'infection. Un test en laboratoire a été effectué sur 2 000 échantillons de sang, la moitié de ces échantillons étant infectée par le virus HIV et l'autre moitié étant saine.
 - **hourlywagedata.sav** : Ce fichier de données d'hypothèse concerne les salaires horaires d'infirmières occupant des postes administratifs et dans les services de soins, et affichant divers niveaux d'expérience.
 - **insurance_claims.sav**. Il s'agit d'un fichier de données hypothétiques qui concerne une compagnie d'assurance souhaitant développer un modèle pour signaler des réclamations suspectes, potentiellement frauduleuses. Chaque observation correspond à une réclamation distincte.
 - **insure.sav** : Ce fichier de données d'hypothèse concerne une compagnie d'assurance qui étudie les facteurs de risque indiquant si un client sera amené à déclarer un incident au cours d'un contrat d'assurance vie d'une durée de 10 ans. Chaque observation figurant dans le fichier de données représente deux contrats, l'un ayant enregistré une réclamation et l'autre non, appariés par âge et sexe.
 - **judges.sav** : Ce fichier de données d'hypothèse concerne les scores attribués par des juges expérimentés (plus un juge enthousiaste) à 300 performances de gymnastique. Chaque ligne représente une performance distincte ; les juges ont examiné les mêmes performances.
 - **kinship_dat.sav** : Rosenberg et Kim ¹¹ ont mis en place une analyse de 15 termes de parentalité (tante, frère, cousin, fille, père, petite-fille, grand-père, grand-mère, petit-fils, mère, neveu, nièce, soeur, oncle). Ils ont demandé à quatre groupes d'étudiants (deux groupes de femmes et deux groupes d'hommes) de trier ces termes en fonction des similarités. Deux groupes (un groupe de femmes et un groupe d'hommes) ont été invités à effectuer deux tris, en basant le second sur un autre critère que le premier. Ainsi, un total de six "sources" a été obtenu. Chaque source correspond à une matrice de proximité 15 x 15, dont le nombre de cellules est égal au nombre de personnes dans une source moins le nombre de fois où les objets ont été partitionnés dans cette source.
 - **kinship_ini.sav** : Ce fichier de données contient une configuration initiale d'une solution tridimensionnelle pour *kinship_dat.sav*.
 - **kinship_var.sav** : Ce fichier de données contient les variables indépendantes *sexe*, *génér(ation)* et *degré* (de séparation) permettant d'interpréter les dimensions d'une solution pour *kinship_dat.sav*. Elles permettent en particulier de réduire l'espace de la solution à une combinaison linéaire de ces variables.
 - **marketvalues.sav** : Ce fichier de données concerne les ventes de maisons dans un nouvel ensemble à Algonquin (Illinois) au cours des années 1999–2000. Ces ventes relèvent des archives publiques.
 - **nhis2000_subset.sav** : Le NHIS (National Health Interview Survey) est une enquête de grande envergure concernant la population des États-Unis. Des entretiens ont lieu avec un échantillon de ménages représentatifs de la population américaine. Des informations démographiques et des observations sur le statut de santé et le comportement sanitaire sont recueillies auprès des membres de chaque ménage. Ce fichier de données contient un sous-groupe d'informations issues de l'enquête de 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Fichier de données et documentation d'usage public. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/. Accès en 2003.
 - **ozone.sav** : Les données incluent 330 observations portant sur six variables météorologiques pour prévoir la concentration d'ozone à partir des variables restantes. Des chercheurs précédents ^{12, 13}, ont décelé parmi ces variables des non-linéarités qui pénalisent les approches standard de la régression.

11. Rosenberg, S., and M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

12. Breiman, L., and J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

13. Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

- **pain_medication.sav** : Ce fichier de données d'hypothèse contient les résultats d'un essai clinique d'un remède anti-inflammatoire traitant les douleurs de l'arthrite chronique. On cherche notamment à déterminer le temps nécessaire au médicament pour agir et les résultats qu'il permet d'obtenir par rapport à un médicament existant.
- **patient_los.sav** : Ce fichier de données d'hypothèse contient les dossiers médicaux de patients admis à l'hôpital pour suspicion d'infarctus du myocarde suspecté (ou "attaque cardiaque"). Chaque observation correspond à un patient distinct et enregistre de nombreuses variables liées à son séjour à l'hôpital.
- **patlos_sample.sav** : Ce fichier de données d'hypothèse contient les dossiers médicaux d'un échantillon de patients sous traitement thrombolytique après un infarctus du myocarde. Chaque observation correspond à un patient distinct et enregistre de nombreuses variables liées à son séjour à l'hôpital.
- **poll_cs.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un enquêteur pour déterminer le niveau de soutien du public pour un projet de loi avant législature. Les observations correspondent à des électeurs enregistrés. Chaque observation enregistre le comté, la ville et le quartier où habite l'électeur.
- **poll_cs_sample.sav** : Ce fichier de données d'hypothèse contient un échantillon des électeurs répertoriés dans le fichier *poll_cs.sav*. L'échantillon a été prélevé selon le plan spécifié dans le fichier de plan *poll_csplan* et ce fichier de données enregistre les probabilités d'inclusion et les pondérations d'échantillon. Toutefois, ce plan faisant appel à une méthode d'échantillonnage de probabilité proportionnelle à la taille (PPS – Probability-Proportional-to-Size), il existe également un fichier contenant les probabilités de sélection conjointes (*poll_jointprob.sav*). Les variables supplémentaires correspondant à la répartition démographique des électeurs et à leur opinion sur le projet de loi proposé ont été collectées et ajoutées au fichier de données une fois l'échantillon prélevé.
- **property_assess.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un contrôleur au niveau du comté pour maintenir les évaluations de valeur de propriété à jour sur des ressources limitées. Les observations correspondent à des propriétés vendues dans le comté au cours de l'année précédente. Chaque observation du fichier de données enregistre la ville où se trouve la propriété, l'évaluateur ayant visité la propriété pour la dernière fois, le temps écoulé depuis cette évaluation, l'évaluation effectuée à ce moment-là et la valeur de vente de la propriété.
- **property_assess_cs.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un contrôleur du gouvernement pour maintenir les évaluations de valeur de propriété à jour sur des ressources limitées. Les observations correspondent à des propriétés de l'état. Chaque observation du fichier de données enregistre le comté, la ville et le quartier où se trouve la propriété, le temps écoulé depuis la dernière évaluation et l'évaluation alors effectuée.
- **property_assess_cs_sample.sav** : Ce fichier de données d'hypothèse contient un échantillon des propriétés répertoriées dans le fichier *property_assess_cs.sav*. L'échantillon a été prélevé selon le plan spécifié dans le fichier de plan *property_assess_csplan* et ce fichier de données enregistre les probabilités d'inclusion et les pondérations d'échantillon. La variable supplémentaire *Valeur courante* a été collectée et ajoutée au fichier de données une fois l'échantillon prélevé.
- **recidivism.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une agence administrative d'application de la loi pour interpréter les taux de récidive dans la juridiction. Chaque observation correspond à un récidiviste et enregistre les informations démographiques qui lui sont propres, certains détails sur le premier délit commis, ainsi que le temps écoulé jusqu'à la seconde arrestation si elle s'est produite dans les deux années suivant la première.
- **recidivism_cs_sample.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une agence administrative d'application de la loi pour interpréter les taux de récidive dans la juridiction. Chaque observation correspond à un récidiviste libéré suite à la première arrestation en juin 2003 et enregistre les informations démographiques qui lui sont propres, certains détails sur le premier délit commis et les données relatives à la seconde arrestation, si elle a eu lieu avant fin juin 2006. Les récidivistes ont été choisis dans plusieurs départements échantillonnés conformément au plan d'échantillonnage spécifié dans *recidivism_cs.csplan*. Ce plan faisant appel à une méthode d'échantillonnage de probabilité proportionnelle à la taille (PPS - Probability proportional to size), il existe également un fichier contenant les probabilités de sélection conjointes (*recidivism_cs_jointprob.sav*).

- **rfm_transactions.sav** : Un fichier de données d'hypothèse qui contient les données de transaction d'achat, y compris la date d'achat, le/les élément(s) acheté(s) et le montant monétaire pour chaque transaction.
- **salesperformance.sav** : Ce fichier de données d'hypothèse concerne l'évaluation de deux nouveaux cours de formation en vente. Soixante employés, divisés en trois groupes, reçoivent chacun une formation standard. En outre, le groupe 2 suit une formation technique et le groupe 3 un didacticiel pratique. A l'issue du cours de formation, chaque employé est testé et sa note enregistrée. Chaque observation du fichier de données représente un stagiaire distinct et enregistre le groupe auquel il a été assigné et la note qu'il a obtenue au test.
- **satisf.sav** : Il s'agit d'un fichier de données d'hypothèse portant sur une enquête de satisfaction effectuée par une société de vente au détail au niveau de quatre magasins. Un total de 582 clients ont été interrogés et chaque observation représente la réponse d'un seul client.
- **screws.sav** : Ce fichier de données contient des informations sur les descriptives des vis, des boulons, des écrous et des clous.¹⁴
- **shampoo_ph.sav** : Ce fichier de données d'hypothèse concerne le processus de contrôle qualité observé dans une usine de produits capillaires. A intervalles réguliers, six lots de sortie distincts sont mesurés et leur pH enregistré. La plage cible est 4.5–5.5.
- **ships.sav** : Il s'agit d'un jeu de données présenté et analysé ailleurs¹⁵ et concernant les dommages causés à des cargos par les vagues. Les effectifs d'incidents peuvent être modélisés comme des incidents se produisant selon un taux de Poisson en fonction du type de navire, de la période de construction et de la période de service. Les mois de service totalisés pour chaque cellule du tableau formé par la classification croisée des facteurs fournissent les valeurs d'exposition au risque.
- **site.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une société pour choisir de nouveaux sites pour le développement de ses activités. L'entreprise a fait appel à deux consultants pour évaluer séparément les sites. Ces consultants, en plus de fournir un rapport approfondi, ont classé chaque site comme constituant une éventualité "bonne", "moyenne" ou "faible".
- **smokers.sav** : Ce fichier de données est extrait de l'enquête National Household Survey of Drug Abuse de 1998 et constitue un échantillon de probabilité des ménages américains. (<http://dx.doi.org/10.3886/ICPSR02934>) Ainsi, la première étape dans l'analyse de ce fichier doit consister à pondérer les données pour refléter les tendances de population.
- **stocks.sav** Ce fichier de données hypothétiques contient le cours et le volume des actions pour un an.
- **stroke_clean.sav** : Ce fichier de données d'hypothèse concerne l'état d'une base de données médicales une fois celle-ci purgée via des procédures de l'option Préparation de données.
- **stroke_invalid.sav** : Ce fichier de données d'hypothèse concerne l'état initial d'une base de données médicales et comporte plusieurs erreurs de saisie de données.
- **stroke_survival** : Ce fichier de données d'hypothèse concerne les durées de survie de patients qui quittent un programme de rééducation à la suite d'un accident ischémique et rencontrent un certain nombre de problèmes. Après l'attaque, l'occurrence d'infarctus du myocarde, d'accidents ischémiques ou hémorragiques est signalée, et le moment de l'événement enregistré. L'échantillon est tronqué à gauche car il n'inclut que les patients ayant survécu durant le programme de rééducation mis en place suite à une attaque.
- **stroke_valid.sav** : Ce fichier de données d'hypothèse concerne l'état d'une base de données médicales une fois les valeurs vérifiées via la procédure Validation de données. Elle contient encore des observations anormales potentielles.
- **survey_sample.sav** : Ce fichier de données concerne des informations d'enquête dont des données démographiques et des mesures comportementales. Il est basé sur un sous-ensemble de variables de la 1998 NORC General Social Survey, bien que certaines valeurs de données aient été modifiées et que des variables supplémentaires fictives aient été ajoutées à titre de démonstration.

14. Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

15. McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

- **telco.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une société de télécommunications pour réduire les attritions de sa base de clients. Chaque observation correspond à un client distinct et enregistre diverses informations démographiques et d'utilisation de service.
- **telco_extra.sav** : Ce fichier de données est semblable au fichier de données *telco.sav* mais les variables de permanence et de dépenses des consommateurs transformées log ont été supprimées et remplacées par des variables de dépenses des consommateurs transformées log standardisées.
- **telco_missing.sav** : Ce fichier de données est un sous-ensemble du fichier de données *telco.sav* mais certaines des valeurs de données démographiques ont été remplacées par des valeurs manquantes.
- **testmarket.sav** : Ce fichier de données d'hypothèse concerne une chaîne de fast foods et ses plans marketing visant à ajouter un nouveau plat à son menu. Trois campagnes étant possibles pour promouvoir le nouveau produit, le nouveau plat est introduit sur des sites sur plusieurs marchés sélectionnés au hasard. Une promotion différente est effectuée sur chaque site et les ventes hebdomadaires du nouveau plat sont enregistrées pour les quatre premières semaines. Chaque observation correspond à un site-semaine distinct.
- **testmarket_1month.sav** : Ce fichier de données d'hypothèse est le fichier de données *testmarket.sav* dans lequel les ventes hebdomadaires sont organisées par site distinct. Certaines variables qui changeaient toutes les semaines disparaissent. En outre, les ventes enregistrées sont à présent la somme des ventes réalisées au cours des quatre semaines de l'enquête.
- **tree_car.sav** : Ce fichier de données d'hypothèse concerne des données démographiques et de prix d'achat de véhicule.
- **tree_credit.sav** : Ce fichier de données d'hypothèse concerne des données démographiques et d'historique de prêt bancaire.
- **tree_missing_data.sav** Ce fichier de données d'hypothèse concerne des données démographiques et d'historique de prêt bancaire avec un grand nombre de valeurs manquantes.
- **tree_score_car.sav** : Ce fichier de données d'hypothèse concerne des données démographiques et de prix d'achat de véhicule.
- **tree_textdata.sav** : Ce fichier de données simples ne comporte que deux variables et vise essentiellement à indiquer l'état par défaut des variables avant affectation du niveau de mesure et des libellés de valeurs.
- **tv-survey.sav** : Ce fichier de données d'hypothèse concerne une enquête menée par un studio de télévision qui envisage de prolonger la diffusion d'un programme ou de l'arrêter. On a demandé à 906 répondants si ils regarderaient le programme dans diverses conditions. Chaque ligne représente un répondant distinct et chaque colonne une condition distincte.
- **ulcer_recurrence.sav** : Ce fichier contient des informations partielles d'une enquête visant à comparer l'efficacité de deux thérapies de prévention de la récurrence des ulcères. Il fournit un bon exemple de données censurées par intervalle et a été présenté et analysé ailleurs ¹⁶.
- **ulcer_recurrence_recoded.sav** : Ce fichier réorganise les informations figurant dans le fichier *ulcer_recurrence.sav* pour que vous puissiez modéliser la probabilité d'événement pour chaque intervalle de l'enquête plutôt que la probabilité d'événement de fin d'enquête. Il a été présenté et analysé ailleurs ¹⁷.
- **verd1985.sav** : Ce fichier de données concerne une enquête ¹⁸. Les réponses de 15 sujets à 8 variables ont été enregistrées. Les variables présentant un intérêt sont divisées en trois ensembles. Le groupe 1 comprend l'âge et la *situation familiale*, le groupe 2 les *animaux domestiques* et la *presse*, et le groupe 3 la *musique* et l'*habitat*. A la variable *animal domestique* est appliqué un codage nominal multiple et à *âge*, un codage ordinal ; toutes les autres variables ont un codage nominal simple.

16. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

17. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

18. Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in Dutch)*. Leiden: Department of Data Theory, University of Leiden.

- **virus.sav** : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend un fournisseur de services Internet pour déterminer les effets d'un virus sur ses réseaux. Il a suivi le pourcentage (approximatif) de trafic de messages électroniques infectés par un virus sur ses réseaux sur la durée, de la découverte à la circonscription de la menace.
- **wheeze_steubenville.sav** : Il s'agit d'un sous-ensemble d'une enquête longitudinale des effets de la pollution de l'air sur la santé des enfants ¹⁹. Les données contiennent des mesures binaires répétées de l'état asthmatique d'enfants de la ville de Steubenville (Ohio), âgés de 7, 8, 9 et 10 ans, et indiquent si la mère fumait au cours de la première année de l'enquête.
- **workprog.sav** : Ce fichier de données d'hypothèse concerne un programme de l'administration visant à proposer de meilleurs postes aux personnes défavorisées. Un échantillon de participants potentiels au programme a ensuite été prélevé. Certains de ces participants ont été sélectionnés au hasard pour participer au programme. Chaque observation représente un participant au programme distinct.
- **worldsales.sav** Ce fichier de données hypothétiques contient les revenus des ventes par continent et par produit.

19. Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, and B. G. Ferris Jr. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366-374.

Remarques

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

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

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

Pour le Canada, veuillez adresser votre courrier à :

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

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

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

Le paragraphe suivant ne s'applique ni au Royaume-Uni, ni dans aucun pays dans lequel il serait contraire aux lois locales : LE PRESENT DOCUMENT EST LIVRE "EN L'ETAT" SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFAÇON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Le présent document peut contenir des inexactitudes ou des coquilles. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, modifier les produits et logiciels décrits dans ce document.

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

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

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

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

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

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

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

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

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

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

LICENCE DE COPYRIGHT :

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

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

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

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

Marques

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

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

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

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

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

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

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

Index

A

affichage des valeurs de catégorie 57
autres statistiques récapitulatives en total 56

C

catégories calculées
à partir des sous-totaux 44
formats d'affichage 15
masquer des catégories dans une expression 44
tableaux personnalisés 14, 43
catégories post-calculées
tableaux personnalisés 14, 43
cellules vides
valeur affichée dans les tableaux personnalisés 16, 92
contrôle du nombre de décimales affichées 24

D

date
insertion de la date dans les tableaux personnalisés 17
décimales
contrôle du nombre de décimales affichées dans les tableaux personnalisés 6, 24, 89
différentes statistiques récapitulatives pour différentes variables
tables empilées 61

E

écart type
tableaux personnalisés 9
effectif
N valide 60
empilement des variables
empilement de variables de couche 34
plusieurs variables source de statistiques récapitulatives 55
tableaux personnalisés 29
exclusion de catégories
tableaux personnalisés 26
exclusion des catégories
tableaux personnalisés 12

F

fichiers d'exemple
emplacement 95
formats d'affichage 24
statistiques récapitulatives dans des tableaux personnalisés 11, 89

fusion de catégories
tableaux personnalisés 41

H

heure
insertion de l'heure dans les tableaux personnalisés 17

I

imbrication de variables
tableaux personnalisés 30, 32
imbrication des variables
variables d'échelle 63
impression de tableaux avec couches 35

J

jeux de réponses multiples 77
pourcentages 8
réponses en double dans les jeux de catégories multiples 16
test de signification 76, 82, 83

K

khi-deux
tableaux personnalisés 65

L

largeur de colonne
contrôle dans les tableaux personnalisés 16, 91
légendes
tableaux personnalisés 17
libellés
modification du texte de libellé des statistiques récapitulatives 90
libellés de coins
tableaux personnalisés 17
libellés de variable
suppression de l'affichage dans les tableaux personnalisés 6

M

masquage des libellés de statistiques dans des tableaux personnalisés 21
maximum
tableaux personnalisés 9
médiane 60
tableaux personnalisés 9
minimum
tableaux personnalisés 9
mode
tableaux personnalisés 9

moyenne 59
tableaux personnalisés 9

N

N total 86
N valide 60, 86
tableaux personnalisés 9
niveau de mesure
modification dans des tableaux personnalisés 1

O

omission de catégories
tableaux personnalisés 26

P

plage
tableaux personnalisés 9
pourcentages
dans les tableaux personnalisés 7, 8, 22, 24
jeux de réponses multiples 8
valeurs manquantes 86

R

récapitulatifs groupés
variables d'échelle 62
réorganisation des catégories
tableaux personnalisés 12

S

somme
tableaux personnalisés 9
sous-totaux 40
masquage des catégories de sous-total 41
tableaux personnalisés 12, 37
statistiques
autres statistiques récapitulatives en total 56
statistiques récapitulatives 53
tables empilées 55
statistiques de test
tableaux personnalisés 17, 65
statistiques des moyennes de colonne
tableaux personnalisés 68
statistiques des proportions de colonne
tableaux personnalisés 71
statistiques récapitulatives 53
autres statistiques récapitulatives en total 56
différents récapitulatifs pour différentes variables dans des tableaux empilés 61

- statistiques récapitulatives (*suite*)
 - dimension source 53
 - format d'affichage 89
 - modification du texte de libellé 90
 - tables empilées 55
 - variable source 53
- Suppression de catégories
 - tableaux personnalisés 26
- suppression des catégories
 - tableaux personnalisés 12

T

- tableau croisé
 - tableaux personnalisés 23
- tableau d'effectifs
 - tableaux personnalisés 15, 49
- tableaux
 - tableaux personnalisés 1
- tableaux avec catégories partagées 15, 49
- tableaux de moyenne et d'effectifs 10, 56
- tableaux personnalisés
 - affichage et masquage des noms et libellés de variable 6
 - catégories calculées 12, 14, 43
 - catégories post-calculées 14, 43
 - cellules vides 16
 - contrôle du nombre de décimales affichées 6
 - dimension source des statistiques 24
 - empilement des variables 29
 - exclusion de catégories 12, 26
 - exclusion des valeurs manquantes pour les récapitulatifs d'échelle 16
 - formats d'affichage 6
 - formats d'affichage des statistiques récapitulatives 11
 - fusion de catégories 41
 - imbrication de variables 30, 32
 - Imbrication de variables de couche 35
 - impression de tableaux en couches 35
 - jeux de catégories multiples 16
 - jeux de réponses multiples 1, 77
 - largeur de colonne 16
 - légendes 17
 - libellés de coins 17
 - libellés de valeurs pour variables catégorielles 1
 - masquage des catégories de sous-total 41
 - masquage des libellés de statistiques 21
 - mode compact 32
 - modification de la dimension des statistiques récapitulatives 10
 - Modification des libellés pour les statistiques récapitulatives 22
 - modification du niveau de mesure 1
 - permutation des variables de ligne et de colonne 33
 - pourcentages 7, 8, 22, 24
 - Pourcentages en ligne et pourcentages en colonne 22

- tableaux personnalisés (*suite*)
 - pourcentages pour jeux de réponses multiples 8
 - procédure de création d'un tableau 4
 - réorganisation des catégories 12
 - sous-totaux 12
 - Sous-totaux 37
 - statistiques de test 17, 65
 - statistiques récapitulatives 7, 8, 9
 - tableau croisé 23
 - tableau d'effectifs 15, 49
 - tableaux avec catégories partagées 15, 49
 - tableaux de moyenne et d'effectifs 10
 - tableaux de variables avec catégories partagées 15, 49
 - tableaux simples pour variables catégorielles 21
 - test de signification et multiréponses 76
 - titres 17
 - totaux 12, 23, 37
 - totaux dans des tableaux dans lesquels des catégories ont été exclues 26
 - totaux marginaux 25
 - totaux personnalisés 10
 - traitement d'un fichier scindé 4
 - tri de catégories 26
 - variables catégorielles 1
 - variables d'échelle 1
 - variables de couche 34
 - Variables de couche 33, 35
- test de signification
 - tableaux personnalisés 17
- tests de signification
 - jeux de réponses multiples 82, 83
- titres
 - tableaux personnalisés 17
- totaux
 - catégories exclues 37
 - couches 40
 - position d'affichage 38
 - tableaux imbriqués 38
 - tableaux personnalisés 12, 23, 37
 - totaux de groupe 38
 - totaux marginaux pour tableaux personnalisés 25
- totaux de groupe 38
- totaux de sous-groupe 38
- traitement d'un fichier scindé
 - tableaux personnalisés 4
- tri de catégories
 - tableaux personnalisés 26

V

- valeurs
 - affichage des libellés et des valeurs de catégorie 57
- valeurs et libellés de valeur 57
- valeurs manquantes 60, 85
 - effet sur le calcul des pourcentages 86
 - inclusion dans les tableaux personnalisés 86
- valeurs manquantes de l'utilisateur 85

- valeurs système manquantes 85
- variable source des statistiques récapitulatives
 - variables d'échelle 63
- variables d'échelle
 - empilement 59
 - imbrication 63
 - récapitulatifs groupés 62
 - récapitulatifs groupés par variables catégorielles de ligne et de colonne 63
 - statistiques récapitulatives 59
 - statistiques récapitulatives multiples 60
- variables de couche
 - empilement de variables de couche 34
 - impression de tableaux en couches 35
 - tableaux personnalisés 34
- Variables de couche
 - Imbrication de variables de couche 35
 - tableaux personnalisés 33, 35
- variables empilées
 - différentes statistiques récapitulatives pour différentes variables 61
 - variables d'échelle 59
- variance
 - tableaux personnalisés 9

