

*IBM SPSS Statistics 24 - Guide abrégé*

**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 95.

**Notice d'édition**

La présente édition s'applique à la version 24.0.0 d'IBM SPSS Statistics et à toutes les éditions et modifications ultérieures sauf mention contraire dans les nouvelles éditions.

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

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

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

Vous pouvez également consulter les serveurs Internet suivants :

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

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

© Copyright IBM France 2015. Tous droits réservés.

---

# Table des matières

<b>Avis aux lecteurs canadiens . . . . .</b>	<b>v</b>	Modification des formats d'affichage des données	44
<b>Chapitre 1. Introduction . . . . .</b>	<b>1</b>	Modèles de tableaux . . . . .	45
Fichiers d'exemple . . . . .	1	Utilisation de formats prédéfinis . . . . .	46
Ouverture d'un fichier de données . . . . .	1	Personnalisation des styles Modèles de tableau	46
Exécution d'une analyse . . . . .	3	Modification des formats par défaut d'un tableau	49
Création de graphiques . . . . .	4	Personnalisation des paramètres de l'affichage initial . . . . .	49
<b>Chapitre 2. Lecture des données . . . . .</b>	<b>7</b>	Affichage de libellés de variable et de valeur . . . . .	50
Structure de base d'un fichier de données IBM SPSS Statistics . . . . .	7	Utilisation de résultats dans d'autres applications . . . . .	52
Lecture des fichiers de données IBM SPSS Statistics . . . . .	7	Collage des résultats sous forme de tableaux Word . . . . .	52
Lecture des données Excel . . . . .	8	Collage des résultats sous forme de texte . . . . .	52
Lecture de données à partir d'une base de données	11	Exportation des résultats vers un fichier Microsoft Word, PowerPoint ou Excel . . . . .	53
Lecture de données à partir d'un fichier texte . . . . .	14	Exportation des résultats vers PDF . . . . .	59
<b>Chapitre 3. Utilisation de l'éditeur de données . . . . .</b>	<b>19</b>	Exportation des résultats vers HTML . . . . .	61
Saisie de données numériques . . . . .	19	<b>Chapitre 7. Utilisation de la syntaxe . . . . .</b>	<b>63</b>
Saisie de données chaîne . . . . .	20	Collage d'une syntaxe . . . . .	63
Définition de données . . . . .	21	Modification d'une syntaxe . . . . .	64
Ajout des libellés de variables . . . . .	21	Ouverture et exécution d'un fichier de syntaxe . . . . .	65
Modification du type et du format de variable . . . . .	22	Utilisation des points de rupture . . . . .	65
Ajout de libellés de valeurs . . . . .	23	<b>Chapitre 8. Modification des valeurs de données . . . . .</b>	<b>67</b>
Gestion des données manquantes . . . . .	23	Création d'une variable catégorielle à partir d'une variable d'échelle . . . . .	67
Valeurs manquantes des variables numériques . . . . .	24	Calcul de nouvelles variables . . . . .	69
Valeurs manquantes d'une variable de chaîne . . . . .	24	Utilisation de fonctions dans les expressions . . . . .	70
<b>Chapitre 4. Examen des statistiques récapitulatives pour chaque variable . . . . .</b>	<b>27</b>	Utilisation d'expressions conditionnelles . . . . .	71
Niveau de mesure . . . . .	27	Utilisation des dates et heures . . . . .	72
Mesures récapitulatives pour données catégorielles	27	Calcul du temps écoulé entre deux dates . . . . .	73
Graphiques pour données catégorielles . . . . .	28	Ajout d'une durée à une date . . . . .	74
Mesures récapitulatives pour variables d'échelle . . . . .	29	<b>Chapitre 9. Tri et sélection de données</b>	<b>77</b>
Histogrammes pour variables d'échelle . . . . .	30	Tri des données . . . . .	77
<b>Chapitre 5. Création et modification de graphiques . . . . .</b>	<b>33</b>	Traitement d'un fichier scindé . . . . .	78
Création de graphiques - Notions de base . . . . .	33	Tri d'observations pour le traitement d'un fichier scindé . . . . .	79
Utilisation de la galerie Générateur de graphiques . . . . .	34	Activation et désactivation du traitement d'un fichier scindé . . . . .	79
Définition des variables et des statistiques . . . . .	34	Sélection de sous-groupes d'observations . . . . .	79
Ajout de texte . . . . .	36	Sélection des sous-ensembles d'observations à l'aide d'une expression conditionnelle . . . . .	80
Création du graphique . . . . .	37	Sélection d'un échantillon aléatoire . . . . .	81
<b>Chapitre 6. Utilisation de la sortie . . . . .</b>	<b>39</b>	Sélection d'une plage de temps ou d'observations	82
Utilisation du visualiseur . . . . .	39	Traitement des observations exclues . . . . .	83
Utilisation de l'éditeur de tableau croisé dynamique	40	Statut de la sélection d'une observation . . . . .	83
Accès aux définitions de sortie . . . . .	40	<b>Chapitre 10. Fichiers d'exemple . . . . .</b>	<b>85</b>
Tableaux croisés dynamiques . . . . .	41	<b>Remarques . . . . .</b>	<b>95</b>
Création et affichage de couches . . . . .	42	Marques . . . . .	97
Modification de tableaux . . . . .	43		
Masquage de lignes et de colonnes . . . . .	44		

**Index . . . . . 99**

---

## Avis aux lecteurs canadiens

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

### Illustrations

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

### Terminologie

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

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

### Claviers

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

### OS/2 et Windows - Paramètres canadiens

Au Canada, on utilise :

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

### Nomenclature

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

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

## Brevets

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

## Assistance téléphonique

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

---

## Chapitre 1. Introduction

Le présent guide va vous montrer comment utiliser la plupart des fonctions disponibles. Il est conçu pour vous apporter une aide étape par étape. Tous les fichiers mentionnés dans les exemples sont installés avec l'application pour vous permettre de suivre la démonstration tout en effectuant les mêmes analyses et en obtenant les mêmes résultats que ceux qui apparaissent ici.

Si vous souhaitez obtenir des exemples détaillés de méthodes d'analyse statistique, utilisez les Etudes de cas expliquées étape par étape, disponibles dans le menu d'aide.

---

### Fichiers d'exemple

La plupart des exemples présentés ici utilisent le fichier de données *demo.sav*. Ce fichier de données est une enquête factice menée auprès de plusieurs milliers de personnes, contenant des informations démographiques et relatives à la consommation.

Si vous utilisez la version Student, votre version du fichier *demo.sav* est un échantillon représentatif du fichier de données d'origine, qui a été réduit pour respecter la limite de 1 500 observations. Les résultats obtenus avec ce fichier de données seront différents des résultats présentés ici.

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.

---

### Ouverture d'un fichier de données

Ouvrir un fichier de données :

1. A partir des menus, sélectionnez :

**Fichier > Ouvrir > Données...**

Une boîte de dialogue d'ouverture des fichiers apparaît.

Par défaut, les fichiers de données IBM® SPSS Statistics (extension *.sav*) apparaissent.

Cet exemple utilise le fichier *demo.sav*.

	age	marital	address	income	inccat	car
1	55	Marital status	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

Figure 1. Fichier demo.sav dans Editeur de données

Le fichier de données apparaît dans l'Editeur de données. Dans l'onglet Vue de données, si vous placez le curseur de la souris sur le nom d'une variable (en-tête de colonne), un libellé de variable plus détaillée apparaît, si il a été défini pour cette variable.

Par défaut, les valeurs de données réelles sont affichées. Pour afficher les libellés :

2. A partir des menus, sélectionnez :

#### Affichage > Libellés de valeurs



Figure 2. Bouton Libellés de valeurs

Vous pouvez également utiliser le bouton Libellés de valeurs dans la barre d'outils.

	age	marital	address	income	inccat	car
1	55	Married	12	72.00	\$50 - \$74	36.
2	56	Unmarried	29	153.00	\$75+	76.
3	28	Married	9	28.00	\$25 - \$49	13.
4	24	Married	4	26.00	\$25 - \$49	12.
5	25	Unmarried	2	23.00	Under \$25	11.
6	45	Married	9	76.00	\$75+	37.
7	42	Unmarried	19	40.00	\$25 - \$49	19.
8	35	Unmarried	15	57.00	\$50 - \$74	28.
9	46	Unmarried	26	24.00	Under \$25	12.
10	34	Married	0	89.00	\$75+	46.
11	55	Married	17	72.00	\$50 - \$74	35.

Figure 3. Libellés de valeurs affichés dans l'éditeur de données

Des libellés de valeurs descriptifs s'affichent pour faciliter l'interprétation des réponses.



## Exécution d'une analyse

Si vous disposez de modules complémentaires, le menu Analyse contient une liste de catégories d'analyse de rapport et d'analyse statistique.

Nous commencerons par créer une simple table de fréquences. Cette exemple nécessite l'option Statistiques de base.

1. A partir des menus, sélectionnez :

**Analyse > Statistiques descriptives > Effectifs...**

La boîte de dialogue Effectifs s'affiche.

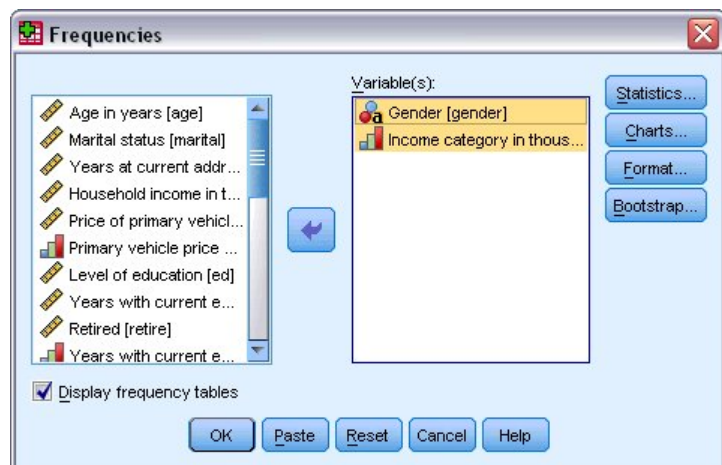


Figure 4. Boîte de dialogue Effectifs

Une icône à côté de chaque variable fournit des informations sur le type de données et le niveau de mesure.

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

Si le libellé et/ou le nom de la variable apparaît tronqué dans la liste, le libellé ou le nom complet est affiché lorsque vous placez le curseur dessus. Le nom de variable *inccat* est affiché entre crochets à la suite du libellé de variable descriptif. *Catégories de revenu en milliers* est le libellé de variable. S'il n'y avait pas de libellé de variable, seul le nom de la variable apparaîtrait dans la zone de liste.

A l'image des fenêtres, vous pouvez redimensionner les boîtes de dialogue en cliquant sur les bords externes ou sur les angles, et en les faisant glisser. Par exemple, si vous élargissez la boîte de dialogue, vous élargissez également les listes de variables.

Dans la boîte de dialogue, choisissez les variables à analyser dans la liste source à gauche et faites-les glisser dans la liste des variables à droite. Le bouton **OK**, qui exécute l'analyse, est désactivé jusqu'à ce qu'une variable soit placée dans la liste Variable(s).

Vous pouvez obtenir des informations supplémentaires en cliquant avec le bouton droit de la souris sur un nom de variable de la liste et en sélectionnant **Informations sur les variables** dans le menu contextuel.

2. Cliquez sur *Sexe [gender]* dans la liste des variables source, puis faites glisser la variable dans la liste cible Variable(s).
3. Cliquez sur *Catégorie de revenus en milliers [rev\_dis]* dans la liste source, puis faites-la glisser vers la liste cible.

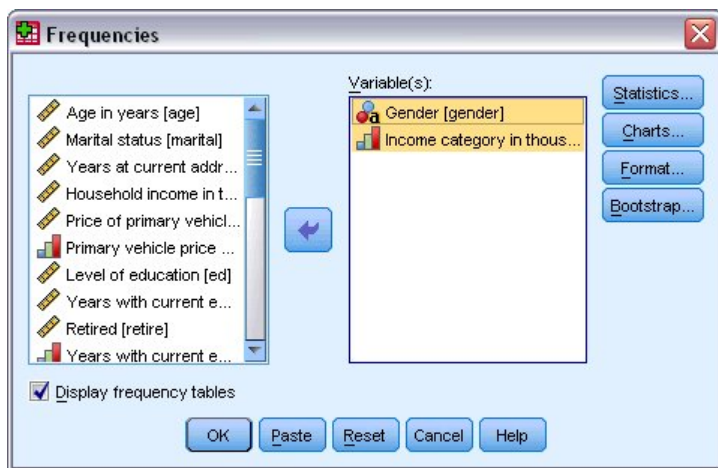


Figure 5. Variables sélectionnées pour l'analyse

4. Cliquez sur **OK** pour exécuter la procédure.

Les résultats sont affichés dans la fenêtre du visualiseur.

		Frequency	Percent	Valid Percent	Cumul: Percent
Valid	Female	3179	49.7	49.7	
	Male	3221	50.3	50.3	
	Total	6400	100.0	100.0	

		Frequency	Percent	Valid Percent	Cumul: Percent
Valid	Under \$25	1174	18.3	18.3	
	\$25 - \$49	2388	37.3	37.3	
	\$50 - \$74	1120	17.5	17.5	
	\$75+	1718	26.8	26.8	
	Total	6400	100.0	100.0	

Figure 6. Table de fréquences des catégories de revenus

## Création de graphiques

Bien que certaines procédures statistiques puissent créer des graphiques, vous pouvez également utiliser le menu Graphes pour la création de graphiques.

Vous pouvez, par exemple, créer un graphique montrant la relation entre l'utilisation de services de téléphonie sans fil et la propriété d'un agenda électronique.

1. A partir des menus, sélectionnez :

**Graphes > Générateur de graphiques...**

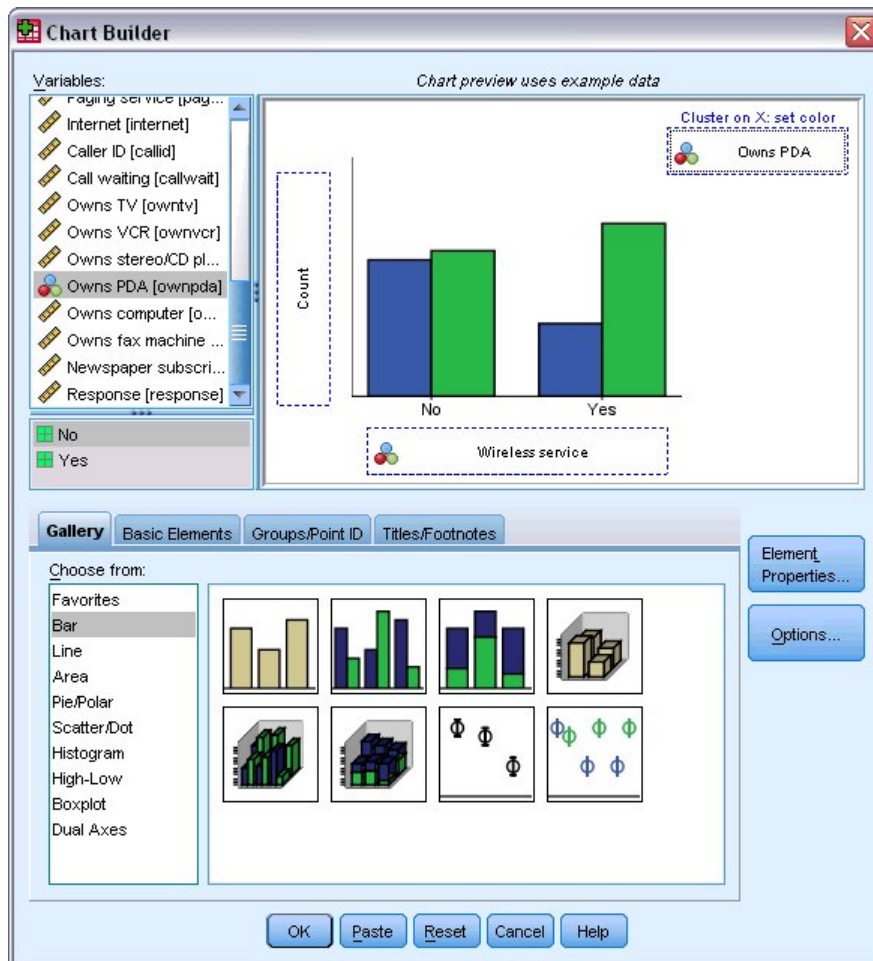


Figure 7. Boîte de dialogue Générateur de graphiques avec des zones d'insertion complétées

2. Cliquez sur l'onglet **Galerie** (s'il n'est pas sélectionné).
3. Cliquez sur **Barre** (s'il n'est pas sélectionné).
4. Faites glisser l'icône Bâtons juxtaposés sur le canevas, qui est la zone étendue au-dessus de la galerie.
5. Faites défiler la liste Variables et cliquez avec le bouton droit de la souris sur *Service sans fil [WiFi]* et sélectionnez **Nominales** comme niveau de mesure.
6. Faites glisser la variable *Service sans fil [WiFi]* dans l'axe des X.
7. Cliquez avec le bouton droit de la souris sur *Possède un agenda électronique [pda]* et sélectionnez **Nominales** comme niveau de mesure.
8. Faites glisser la variable *Possède un agenda électronique [pda]* dans la zone d'insertion de cluster dans le coin supérieur droit du canevas.
9. Cliquez sur **OK** pour créer le graphique.

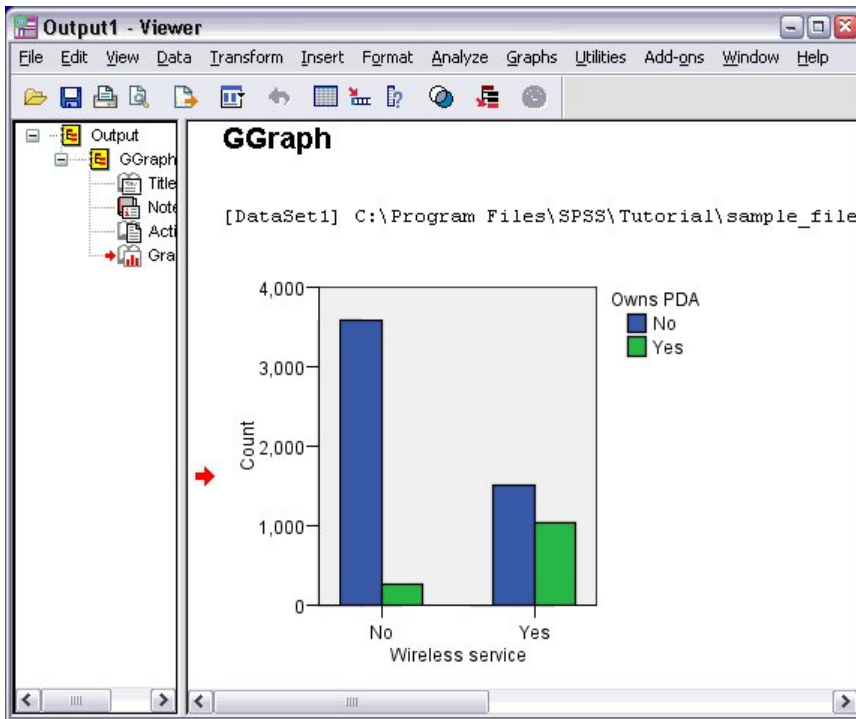


Figure 8. Graphique à barres affiché dans la fenêtre du visualiseur

Le graphique à barres est affiché dans le visualiseur. Le graphique indique que la proportion de personnes possédant un agenda électronique est plus importante chez les personnes utilisant un service de téléphonie sans fil.

Vous pouvez modifier les graphiques et les tableaux en double-cliquant dessus dans le panneau de contenu de la fenêtre visualiseur et vous pouvez copier-coller vos résultats dans d'autres applications. Ces sujets seront traités ultérieurement.

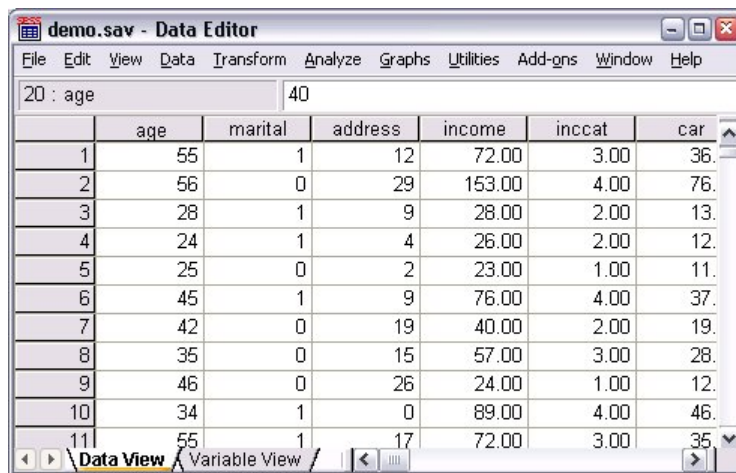
---

## Chapitre 2. Lecture des données

Les données peuvent être saisies directement ou importées à partir de sources différentes. Ce chapitre traite des processus de lecture de données stockées dans les fichiers de données IBM SPSS Statistics, dans des applications de feuilles de calcul telles que Microsoft Excel, dans les applications de base de données telles que Microsoft Access et dans des fichiers texte.

---

### Structure de base d'un fichier de données IBM SPSS Statistics



The screenshot shows the 'demo.sav - Data Editor' window. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The current view is 'Data View'. The data table has the following structure:

	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

Figure 9. Éditeur de données

Les fichiers de données IBM SPSS Statistics sont organisés par observations (lignes) et par variables (colonnes). Dans ce fichier de données, les observations représentent des répondants dans le cadre d'une enquête. Les variables représentent des réponses à chaque question posée au cours de l'enquête.

---

### Lecture des fichiers de données IBM SPSS Statistics

Les fichiers de données IBM SPSS Statistics, portant l'extension de fichier *.sav*, contiennent les données que vous avez enregistrées.

1. A partir des menus, sélectionnez :  
**Fichier > Ouvrir > Données...**
2. Recherchez le fichier *demo.sav* et ouvrez-le. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.

Les données sont affichées dans l'éditeur de données.

	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

Figure 10. Fichier de données ouvert

## Lecture des données Excel

Au lieu de saisir toutes vos données directement dans l'éditeur de données, vous pouvez les lire à partir d'applications telles que Microsoft Excel. Vous pouvez également lire des en-têtes de colonne comme noms de variable.

1. A partir des menus, sélectionnez :

**Fichier > Importer des données > Excel**

2. Accédez au dossier Samples\English et sélectionnez demo.xlsx.

La boîte de dialogue Lire le fichier Excel contient un aperçu du fichier de données. Le contenu de la première feuille du fichier est affiché. Si le fichier contient plusieurs feuilles, vous pouvez sélectionner la feuille dans la liste.

Vous constatez que certaines valeurs de chaîne de *Sexe* sont précédées par des espaces. Certaines des valeurs de *Situationfamiliale* sont affichées sous forme de points (.).

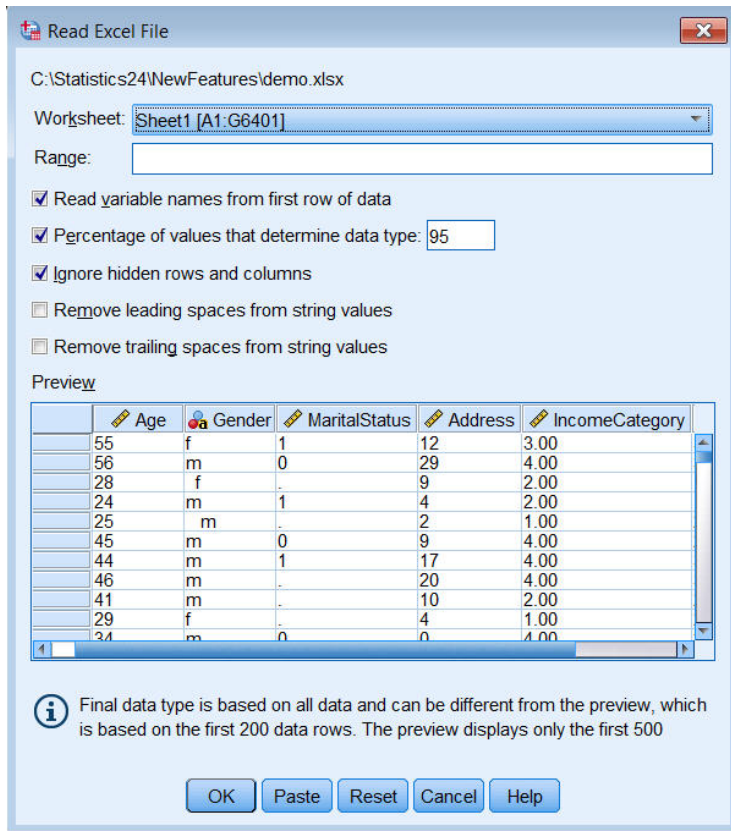
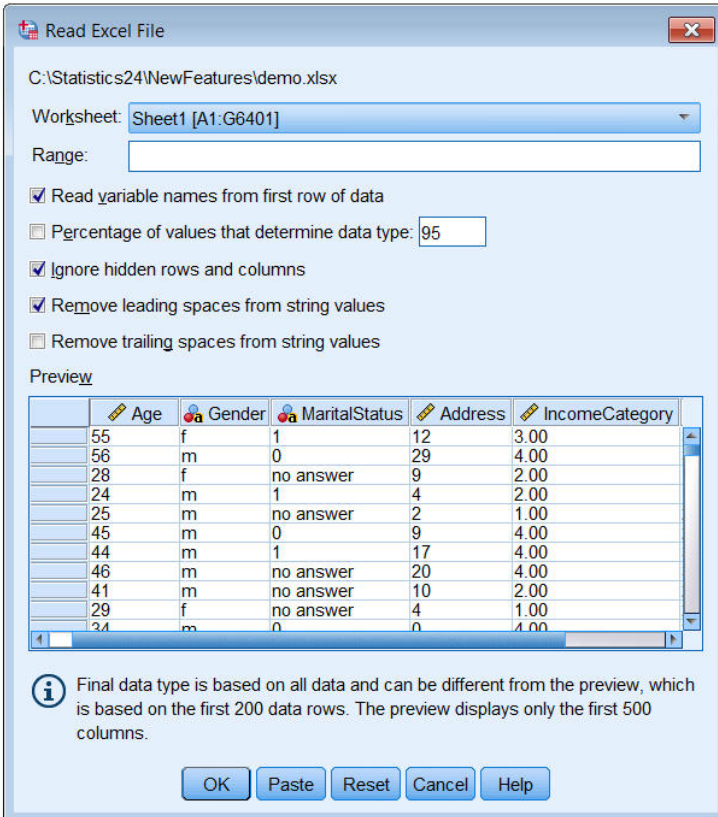


Figure 11. Boîte de dialogue Lire le fichier Excel

3. Vérifiez que l'option **Lire les noms de variable à partir de la première ligne de données** est sélectionnée. Les en-têtes de colonne qui ne sont pas conformes aux règles de dénomination de variables sont convertis en noms de variable valides. Les en-têtes de colonne d'origine sont enregistrés en tant que libellés de variable.
4. Sélectionnez **Retirer les espaces au début des valeurs de type chaîne**.
5. Désélectionnez **Pourcentage de valeurs qui détermine le type de données**.





La valeur de chaîne "no answer" figure maintenant dans les cellules identifiées comme valeurs système manquantes. S'il n'y a pas de paramètre de pourcentage de valeur est si la colonne contient plusieurs types de données, la variable est lue comme une chaîne. Toutes les valeurs sont conservées, mais les valeurs numériques sont traitées comme des chaînes.

6. Sélectionnez (cochez) **Pourcentage de valeurs qui détermine le type de données** pour traiter *Situationfamiliale* comme une variable numérique.
7. Cliquez sur **OK** pour lire le fichier Excel.

Les données apparaissent à présent dans l'Editeur de données, les en-têtes de colonne étant utilisés comme noms de variable. Les noms de variable ne pouvant pas contenir des espaces, les espaces des en-têtes de colonne d'origine sont supprimés. Ainsi, l'en-tête de colonne "Situation familiale" devient *Situationfamiliale* lorsqu'il est converti en variable. L'en-tête de colonne d'origine est conservé comme libellé de variable.



The screenshot shows the IBM SPSS Statistics Data Editor interface. The title bar reads '\*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Extensions, Window, and Help. The toolbar contains various icons for file operations and data manipulation. The main window displays a data grid with 10 rows and 6 columns. The columns are labeled Age, Gender, MaritalStatus, Address, IncomeCategory, and JobCategory. The data is as follows:

	Age	Gender	MaritalStatus	Address	IncomeCategory	JobCategory
1	55	f	1	12	3.00	3
2	56	m	0	29	4.00	3
3	28	f	.	9	2.00	1
4	24	m	1	4	2.00	1
5	25	m	.	2	1.00	2
6	45	m	0	9	4.00	2
7	44	m	1	17	4.00	3
8	46	m	.	20	4.00	3
9	41	m	.	10	2.00	2
10	29	f	.	4	1.00	2

At the bottom of the window, there are two tabs: 'Data View' (selected) and 'Variable View'.

Figure 12. Données Excel importées

**Information associée:**

Chapitre 10, «Fichiers d'exemple», à la page 85

---

## Lecture de données à partir d'une base de données

Les données issues de sources de bases de données peuvent facilement être importées à l'aide de l'Assistant de base de données. Toute base de données utilisant des pilotes ODBC (Open Database Connectivity) peut être lue directement une fois ces pilotes installés. Les pilotes ODBC de nombreux formats de base de données sont fournis sur le CD d'installation. Vous pouvez obtenir des pilotes supplémentaires auprès de distributeurs indépendants. L'une des applications de base de données les plus courantes, Microsoft Access, est traitée dans cet exemple.

*Remarque :* Cet exemple est propre à Microsoft Windows et nécessite un pilote ODBC pour Access. Le pilote ODBC Microsoft Access fonctionne uniquement avec la version 32 bits d'IBM SPSS Statistics. Les étapes sont semblables sur les autres plateformes mais peuvent nécessiter un pilote ODBC tiers pour Access.

1. A partir des menus, sélectionnez :

**Fichier > Importer des données > Base de données > Nouvelle requête...**

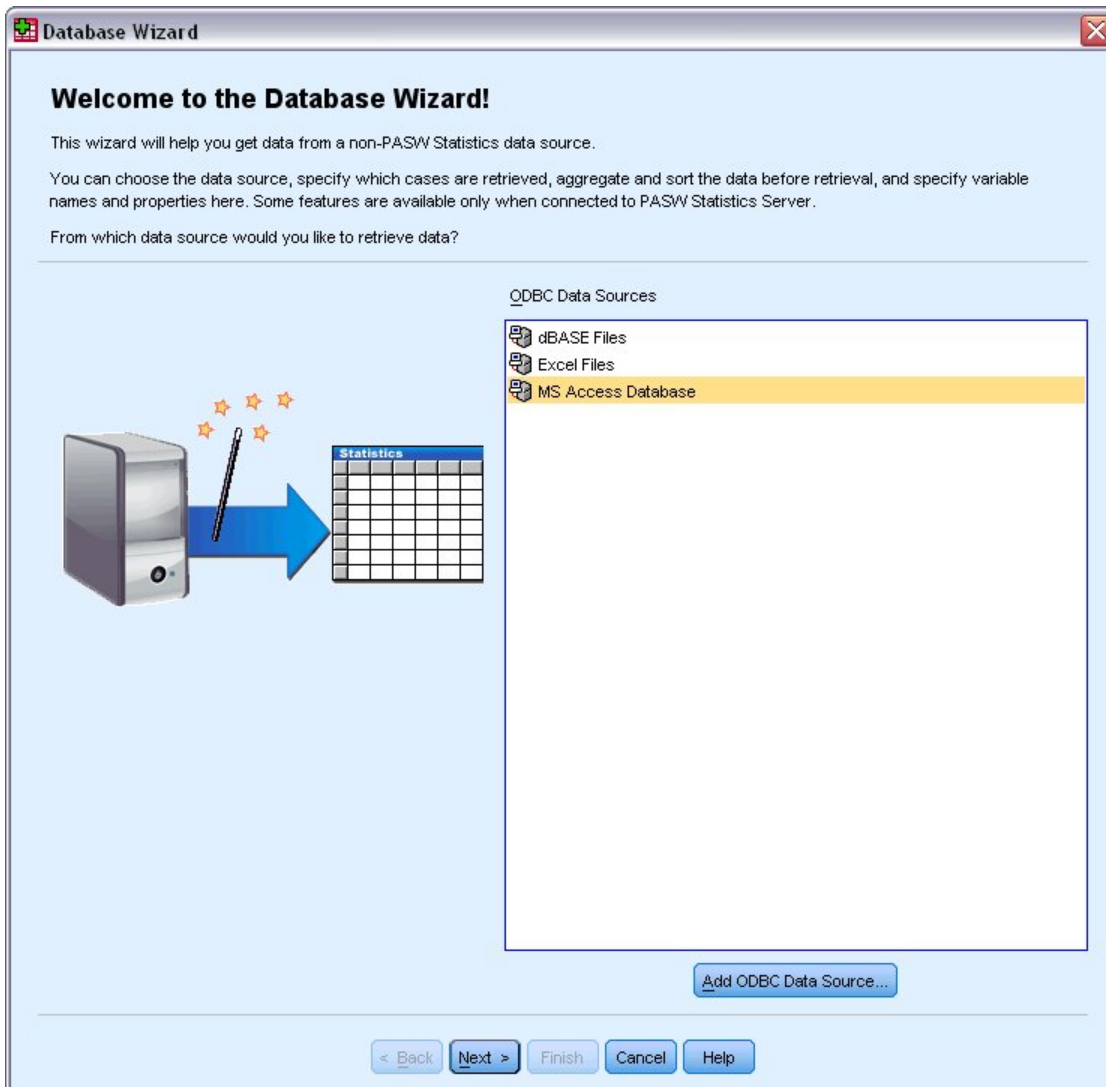


Figure 13. Boîte de dialogue Bienvenue dans l'assistant de base de données

2. Sélectionnez **Base de données MS Access** dans la liste des sources de données et cliquez sur **Suivant**.

*Remarque :* En fonction de votre installation, vous pouvez voir une liste des sources de données OLEDB à gauche de l'assistant (systèmes d'exploitation Windows uniquement), mais cet exemple utilise la liste de sources de données ODBC affichée sur la droite.

3. Cliquez sur **Parcourir** pour accéder au fichier de base de données Access à ouvrir.
4. Ouvrez le fichier *demo.mdb*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.
5. Cliquez sur **OK** dans la boîte de dialogue de connexion.

A l'étape suivante, vous pourrez indiquer les tables et les variables que vous souhaitez importer.

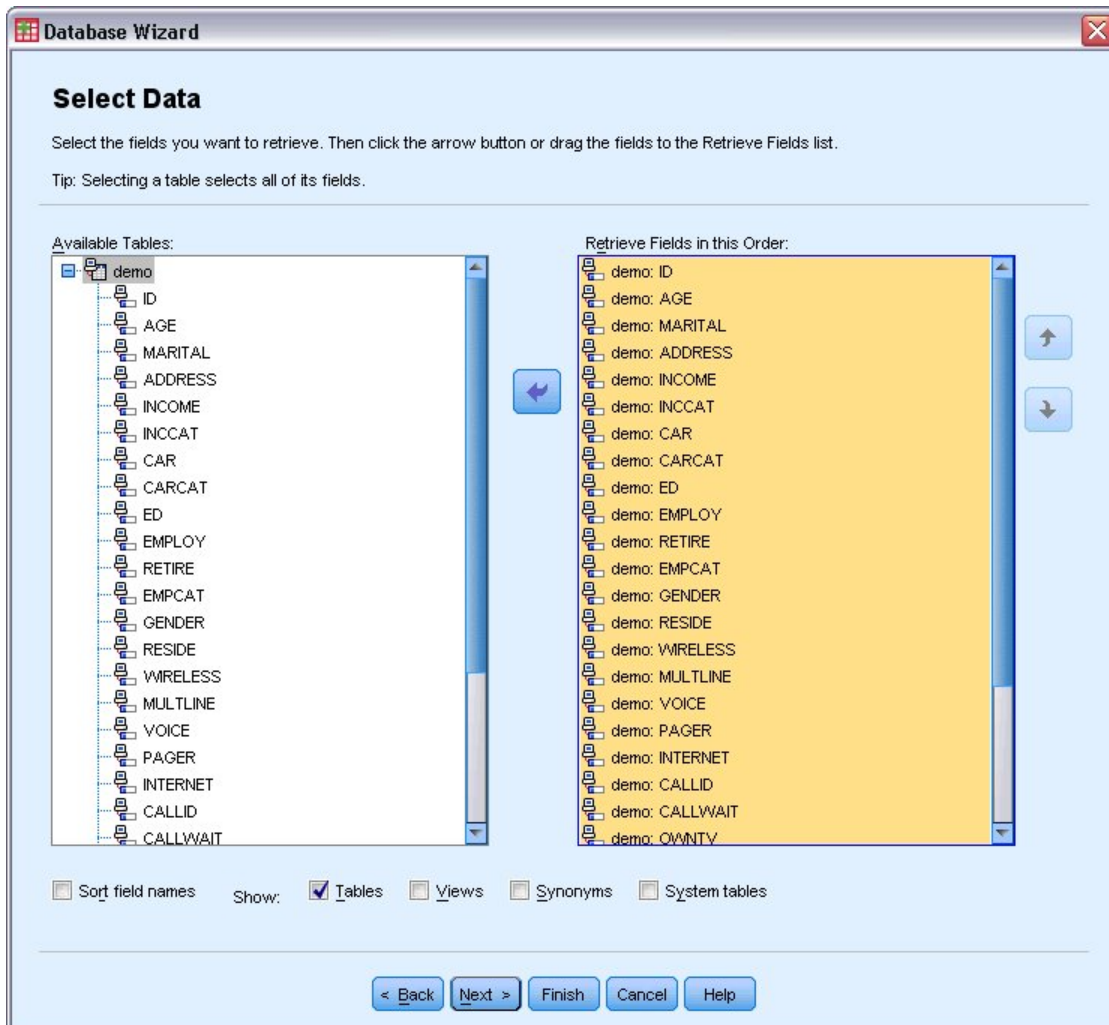


Figure 14. Étape Sélectionner des données

6. Faites glisser le tableau **demo** vers la liste Rétablir les champs dans cet ordre.
7. Cliquez sur **Suivant**.

A l'étape suivante, vous pouvez sélectionner les enregistrements (observations) à importer.

Si vous ne souhaitez pas importer toutes les observations, vous pouvez sélectionner un sous-ensemble d'observations (par exemple, les hommes de plus de 30 ans) ou un échantillon aléatoire d'observations dans la source de données. Pour les sources de données volumineuses, vous pouvez limiter le nombre d'observations à un échantillon restreint et représentatif afin de réduire la durée du traitement.

8. Cliquez sur **Suivant** pour continuer.

Les noms de champ permettent de créer des noms de variable. Si nécessaire, ils sont convertis en noms de variable valides. Les noms de champ d'origine sont conservés en tant que libellés de variable. Vous pouvez également modifier les noms de variable avant d'importer la base de données.

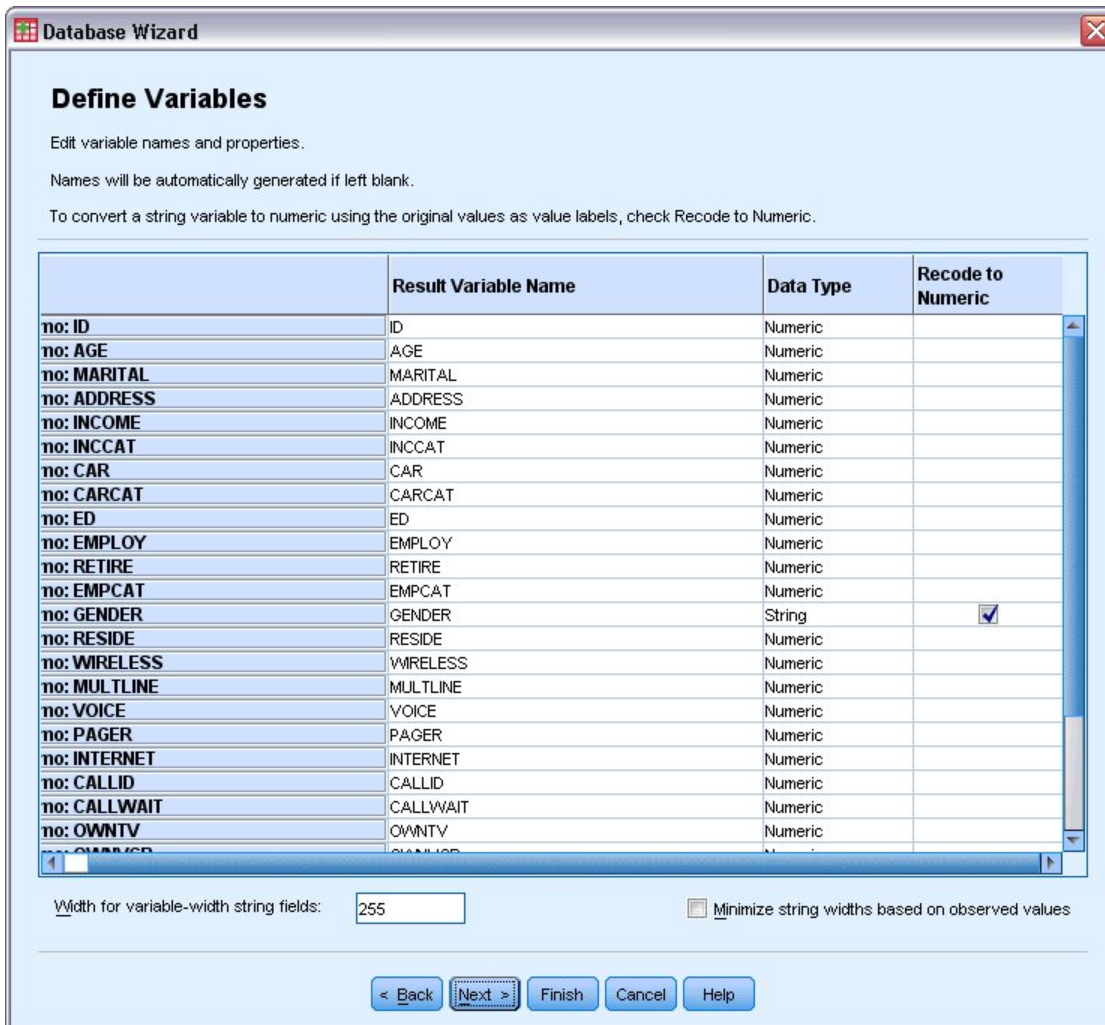


Figure 15. Etape de définition des variables

9. Cliquez sur la cellule **Recoder en numérique** dans le champ Sexe. Cette option convertit les variables de chaîne en variables numériques (entiers) et conserve la valeur d'origine comme libellé de valeur de la nouvelle variable.
10. Cliquez sur **Suivant** pour continuer.  
L'instruction SQL créée à partir de vos sélections dans l'Assistant de base de données apparaît à l'étape Résultats. Cette instruction peut être exécutée immédiatement ou enregistrée dans un fichier pour une utilisation ultérieure.
11. Cliquez sur **Terminer** pour importer les données.

Toutes les données de la base de données Access sélectionnées pour l'importation sont à présent disponibles dans l'éditeur de données.

## Lecture de données à partir d'un fichier texte

Les fichiers texte représentent une autre source commune de données. De nombreux tableurs et bases de données peuvent enregistrer leur contenu dans l'un des nombreux formats de fichier texte. Les fichiers délimités par une virgule ou une tabulation se rapportent aux lignes de données utilisant des virgules ou des tabulations pour indiquer chaque variable. Dans cet exemple, les données sont délimitées par des tabulations.

1. A partir des menus, sélectionnez :

**Fichier > Importer des données > Données texte**

2. Accédez au dossier Samples\English et sélectionnez demo.txt.

L'assistant d'importation de texte vous guide dans le processus de définition de l'interprétation du fichier texte indiqué.

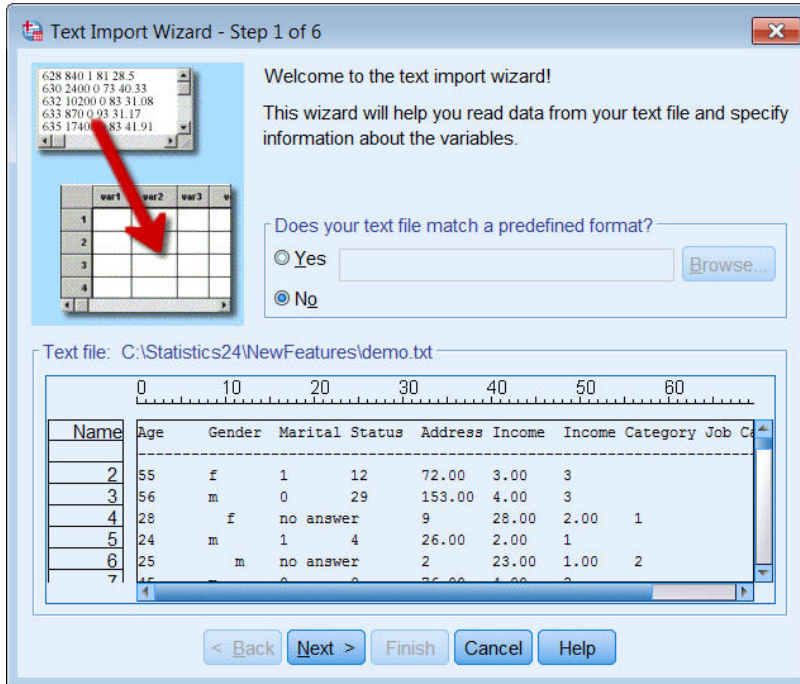


Figure 16. Assistant d'importation de texte - Etape 1 sur 6

3. A l'étape 1, vous pourrez sélectionner un format prédéfini ou créer un format dans l'Assistant. Sélectionnez **Non**.
4. Cliquez sur **Suivant** pour continuer.  
Comme indiqué précédemment, ce fichier utilise un format délimité par des tabulations. En outre, les noms de variable sont définis sur la première ligne de ce fichier.
5. A l'étape 2 de l'assistant, sélectionnez **Délimité** pour indiquer que le format du fichier est structuré.
6. Sélectionnez **Oui** pour indiquer que des noms de variable figurent en haut du fichier.
7. Cliquez sur **Suivant** pour continuer.
8. A l'étape 3, saisissez 2 pour le numéro de la ligne à partir de laquelle commence la première observation de données (compte tenu que les noms de variable se trouvent sur la première ligne).
9. Conservez les valeurs par défaut des autres champs de cette étape et cliquez sur **Suivant** pour continuer.  
L'aperçu des données à l'étape 4 permet de vérifier que le fichier peut être lu correctement.
10. Sélectionnez **Tab** et décochez les autres options de délimiteur. **Espace** est sélectionné par défaut car le fichier contient des espaces. Pour ce fichier, les espaces font partie des valeurs de données, et non des délimiteurs. Vous devez désélectionner **Espace** pour lire le fichier correctement.
11. Sélectionnez **Retirer les espaces au début des valeurs de type chaîne**. Les espaces situés au début des valeurs de chaîne affectent la manière dont les chaînes sont évaluées dans les expressions. Dans ce fichier, certaines valeurs de *Sexe* sont précédées d'espaces qui ne font pas partie de la valeur. Si vous ne les retirez pas, la valeur " f" est traitée comme autre chose que "f".

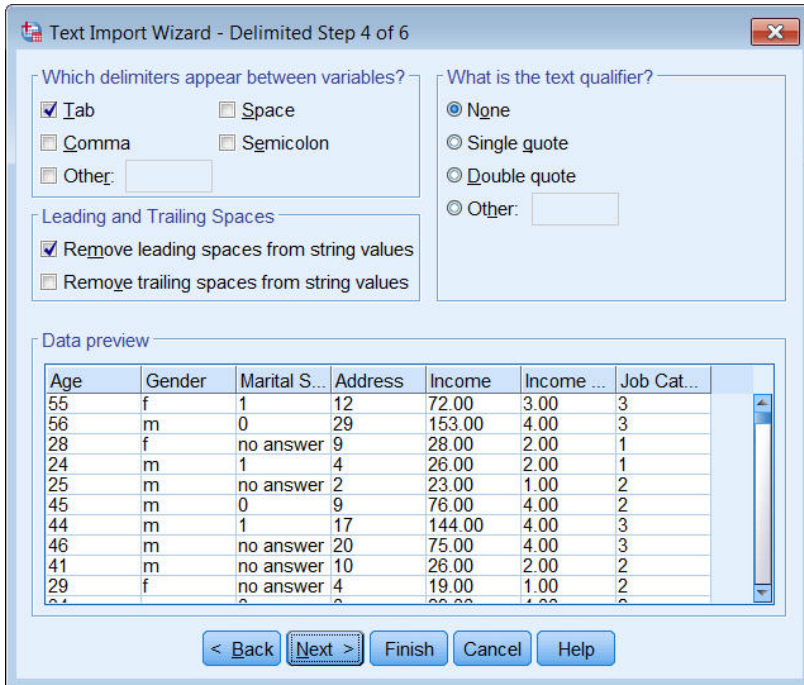


Figure 17. Assistant d'importation de texte - Etape 4 sur 6

12. Cliquez sur **Suivant** pour continuer.

Parce que les noms de variable sont modifiés de manière à respecter les règles de dénomination, l'étape 5 vous donne la possibilité de modifier les noms de votre choix.

Vous pouvez également définir les types de données dans cette boîte de dialogue. Par exemple, vous pouvez remplacer *Revenu* par le format de la devise dollar.

Pour modifier un type de données :

13. Dans **Aperçu des données**, sélectionnez *Revenu*.

14. Sélectionnez **Dollar** dans la liste déroulante Format des données.



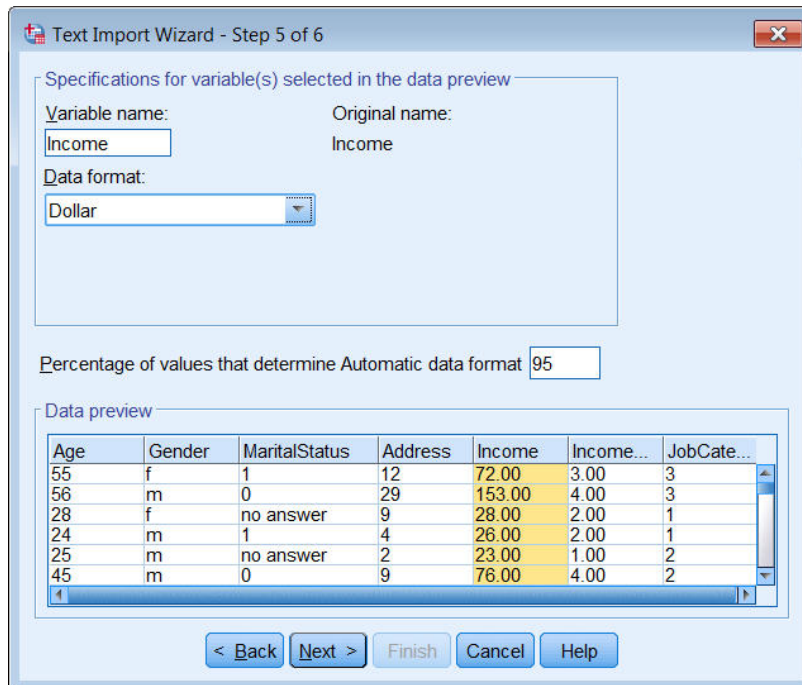


Figure 18. Sélectionnez le type de données.

La variable *Situationfamiliale* contient à la fois des valeurs de chaîne et des valeurs numériques. Moins de cinq pourcent des valeurs sont des chaînes. Avec la valeur de 95 % pour le paramètre **Pourcentage de valeurs qui détermine le format de données automatique**, la variable est traitée comme une valeur numérique, et les valeurs de chaîne sont définies comme Valeur système manquante. Si aucun format de données n'atteint ce pourcentage, la variable est traitée comme une chaîne. Si vous remplacez cette valeur par 100, toutes les valeurs sont conservées, mais toutes les valeurs numériques sont traitées comme des chaînes.

15. Cliquez sur **Suivant** pour continuer.
16. Conservez les sélections par défaut à la dernière étape, puis cliquez sur **Terminer** pour importer les données.





---

## Chapitre 3. Utilisation de l'éditeur de données

Cette fenêtre affiche le contenu du fichier de données actif. Les informations contenues dans l'éditeur de données se composent de variables et d'observations.

- Dans Vue de données, les colonnes représentent des variables et les lignes des observations.
- Dans la vue de variable, chaque ligne est une variable et chaque colonne est un attribut associé à cette variable.

Les variables sont utilisées pour représenter les différents types de données que vous avez rassemblées. Une analogie courante est celle de l'enquête. La réponse à chaque question d'une enquête est équivalente à une variable. Il existe de nombreux types de variable : nombres, caractères, monnaies et dates.

---

### Saisie de données numériques

Les données peuvent être saisies dans l'éditeur de données, qui peut s'avérer utile pour traiter les fichiers de données peu volumineux ou pour apporter de légères modifications à des fichiers de données plus volumineux.

1. Cliquez sur l'onglet **Vue de variable** en bas de la fenêtre de l'éditeur de données.  
Vous devez définir les variables qui seront utilisées. Dans cet exemple, seules trois variables sont nécessaires : *âge*, *situation familiale* et *revenu*.

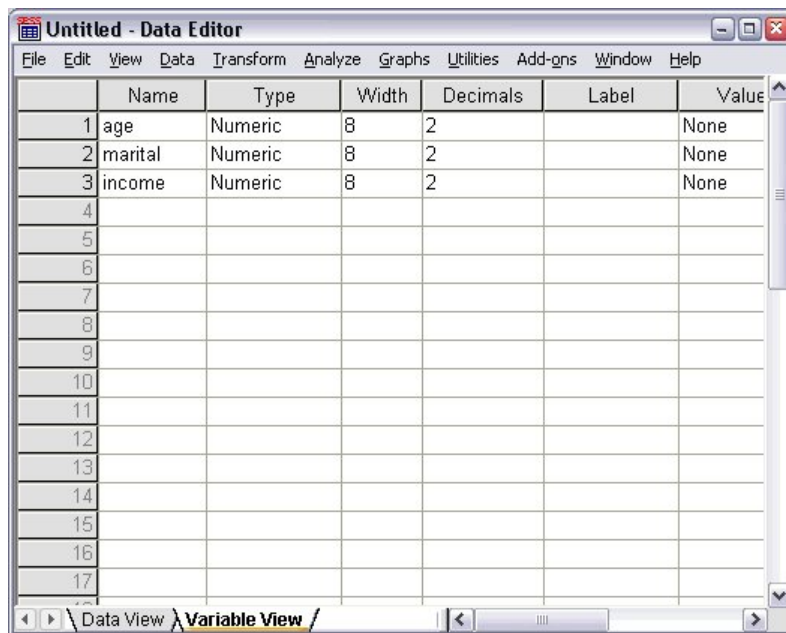


Figure 19. Noms de variable dans la vue de variable

2. Dans la première ligne de la première colonne, saisissez *age*.
3. Dans la deuxième ligne, saisissez *situation familiale*.
4. Dans la troisième ligne, saisissez *revenu*.

Un type de données numérique est automatiquement attribué aux nouvelles variables.

Si vous ne saisissez pas de noms de variable, des noms uniques sont automatiquement créés. Cependant, ces noms ne sont pas descriptifs et ne sont pas recommandés pour les fichiers de données volumineux.

5. Cliquez sur l'onglet **Vue de données** pour continuer à saisir des données.  
Les noms saisis dans la vue de variable sont à présent les en-têtes des trois premières colonnes dans Vue de données.  
Commencez à saisir des données dans la première ligne, en commençant par la première colonne.

	age	marital	income	var	var	var
1	55.00	1.00	72000.00			
2	53.00	.00	153000.0			
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						

Figure 20. Valeurs entrées dans Vue de données

6. Dans la colonne *âge*, saisissez 55.
7. Dans la colonne *situation familiale*, saisissez 1.
8. Dans la colonne *revenu*, saisissez 72000.
9. Placez le curseur dans la seconde ligne de la première colonne pour ajouter les données du sujet suivant.
10. Dans la colonne *âge*, saisissez 53.
11. Dans la colonne *situation familiale*, saisissez 0.
12. Dans la colonne *revenu*, saisissez 153 000.  
Les colonnes *âge* et *situation familiale* affichent actuellement un séparateur décimal alors que les valeurs sont des entiers. Pour masquer le séparateur décimal de ces variables :
13. Cliquez sur l'onglet **Vue de variable** en bas de la fenêtre de l'éditeur de données.
14. Dans la colonne *Décimales* de la ligne *âge*, saisissez 0 pour masquer la décimale.
15. Dans la colonne *Décimales* de la ligne *situation familiale*, saisissez 0 pour masquer la décimale.

## Saisie de données chaîne

Des données non numériques, telles que des chaînes de texte, peuvent également être saisies dans l'éditeur de données.

1. Cliquez sur l'onglet **Vue de variable** en bas de la fenêtre de l'éditeur de données.
2. Dans la première cellule de la première ligne vide, saisissez *sexe* comme nom de variable.
3. Cliquez sur la cellule *Type* à côté de votre entrée.
4. Cliquez sur le bouton à droite de la cellule *Type* pour ouvrir la boîte de dialogue Type de variable.
5. Sélectionnez **Chaîne** pour indiquer le type de variable.
6. Cliquez sur **OK** pour enregistrer votre sélection et revenir dans Editeur de données.

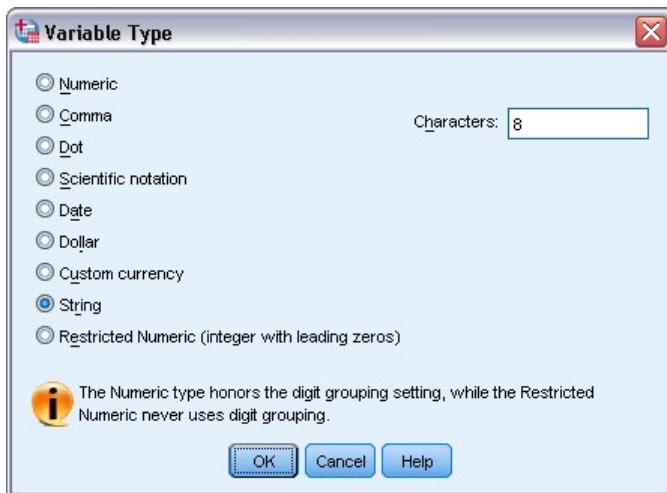


Figure 21. Boîte de dialogue Type de variable

---

## Définition de données

Outre les données, vous pouvez aussi définir des libellés de variables descriptifs et des libellés de valeurs pour les noms de variables et les valeurs de données. Ces libellés descriptifs sont utilisés dans les rapports statistiques et les graphiques.

### Ajout des libellés de variables

Les libellés fournissent une description des variables. Ces descriptions correspondent souvent à une version plus longue des noms de variable. Les libellés peuvent s'élever jusqu'à 255 octets. Ces libellés sont utilisés dans les sorties pour identifier les variables différentes.

1. Cliquez sur l'onglet **Vue de variable** en bas de la fenêtre de l'éditeur de données.
2. Dans la colonne *Libellé* de la ligne *age*, saisissez l'âge du répondant.
3. Dans la colonne *Libellé* de la ligne *situation maritale*, saisissez la situation familiale.
4. Dans la colonne *Libellé* de la ligne *revenu*, saisissez revenu du ménage.
5. Dans la colonne *Libellé* de la ligne *sexe*, saisissez le sexe.

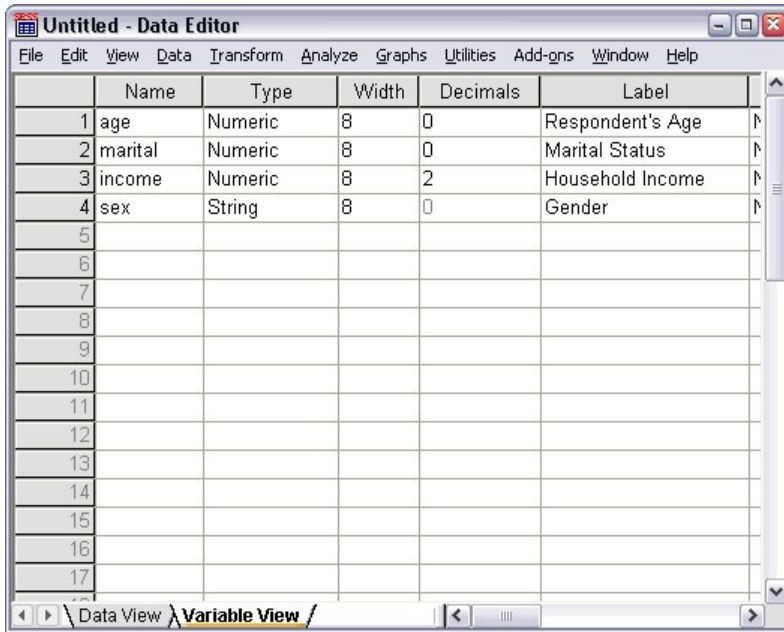


Figure 22. Libellés de variable entrées dans la vue de variable

## Modification du type et du format de variable

La colonne *Type* affiche le type de données actuel de chaque variable. Les types de données les plus courants sont les formats numériques et chaînes, mais de nombreux autres formats sont pris en charge. Dans le fichier de données actuel, la variable *revenu* est définie comme étant un type numérique.

1. Cliquez sur la cellule *Type* de la ligne *revenu*, puis sur le bouton à droite de la cellule pour ouvrir la boîte de dialogue *Type de variable*.
2. Sélectionnez **Dollar**.

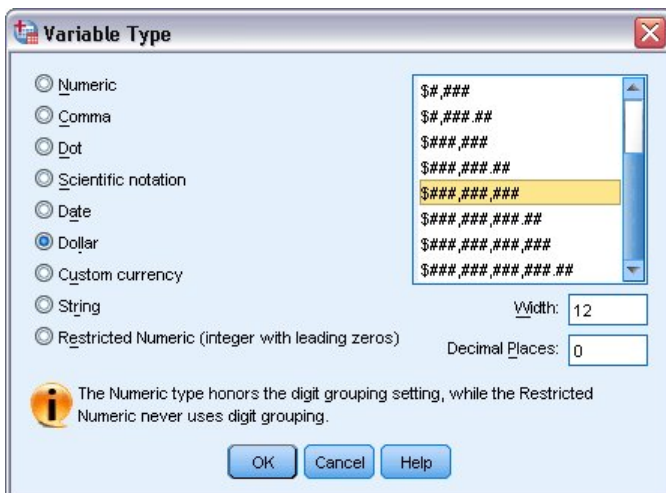


Figure 23. Boîte de dialogue *Type de variable*

Les options de format du type de données sélectionné apparaissent.

3. Dans cet exemple, sélectionnez **\$###,###,###** comme format de devise.
4. Cliquez sur **OK** pour appliquer vos modifications.

## Ajout de libellés de valeurs

Les libellés de valeur correspondent à une façon d'associer aux valeurs de variable des libellés de chaîne. Dans cet exemple, il y a deux valeurs possibles pour la variable *situation familiale*. La valeur 0 signifie que le sujet est célibataire et la valeur 1 qu'il est marié.

1. Cliquez sur la cellule *Valeurs* de la ligne *situation*, puis sur le bouton à droite de la cellule pour ouvrir la boîte de dialogue Libellés de valeur.

Le champ **Valeur** correspond à la valeur numérique réelle.

Le champ **Libellé de valeur** correspond au libellé de chaîne appliqué à la valeur numérique indiquée.

2. Saisissez 0 dans le champ Valeur.
3. Saisissez Célibataire dans le champ Libellé.
4. Cliquez sur **Ajouter** pour ajouter le libellé à la liste.

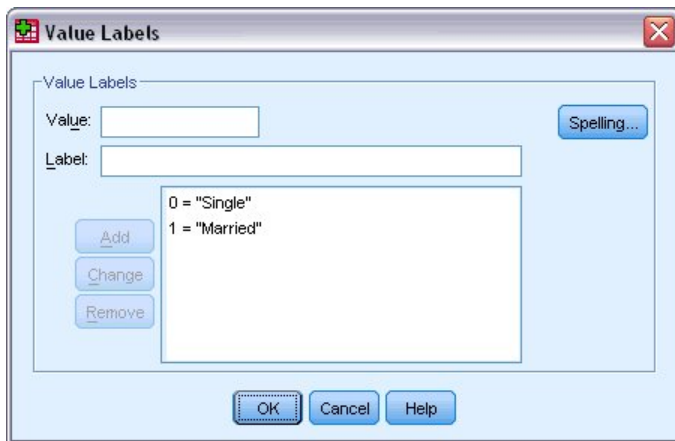


Figure 24. Boîte de dialogue Libellés de valeurs

5. Saisissez 1 dans le champ Valeur, puis Marié dans le champ Libellé.
6. Cliquez sur **Ajouter**, puis sur **OK** pour enregistrer vos modifications et revenir dans l'éditeur de données.

Ces libellés peuvent également apparaître dans Vue de données pour faciliter la lecture de vos données.

7. Cliquez sur l'onglet **Vue de données** en bas de la fenêtre de l'éditeur de données.
8. A partir des menus, sélectionnez :

### Affichage > Libellés de valeurs

Les libellés figurent à présent dans une liste lorsque vous entrez des valeurs dans Editeur de données. Cette configuration a l'avantage de proposer des réponses valides et plus descriptives.

Si l'élément de menu Libellés de valeurs est déjà activé (lorsqu'il est indiqué par une coche), sélectionner à nouveau **Libellés de valeurs** désactive l'affichage des libellés de valeurs.

## Gestion des données manquantes

Les données manquantes ou non valides sont trop fréquentes pour être ignorées. Les répondants peuvent refuser de répondre à certaines questions, ne pas connaître la réponse ou donner une réponse dont le format est inattendu. Si vous ne filtrez ou n'identifiez pas ces données, les résultats de votre analyse risquent d'être imprécis.

En ce qui concerne les données numériques, les champs de données vides ou contenant des entrées non valides sont convertis en données manquantes par défaut, signalées par une virgule (Windows français) ou un point (Windows anglais).

La raison pour laquelle une valeur est manquante peut être importante pour votre analyse. Par exemple, vous pouvez juger utile de distinguer les personnes qui ont refusé de répondre à une question de celles qui n'ont pas répondu car cette question ne les concernait pas.

## Valeurs manquantes des variables numériques

1. Cliquez sur l'onglet **Vue de variable** en bas de la fenêtre de l'éditeur de données.
2. Cliquez sur la cellule *Manquante* de la ligne *âge*, puis sur le bouton à droite de la cellule pour ouvrir la boîte de dialogue Valeurs manquantes.

Dans cette boîte de dialogue, vous pouvez indiquer jusqu'à trois valeurs manquantes ou une plage de valeurs et une valeur discrète supplémentaire.



Figure 25. Boîte de dialogue Valeurs manquantes

3. Sélectionnez **Valeurs manquantes discrètes**.
4. Saisissez 999 dans la première zone de texte. Ne remplissez pas les deux autres zones de texte.
5. Cliquez sur **OK** pour enregistrer vos modifications et revenir dans l'éditeur de données.  
La valeur des données manquantes ayant été ajoutée, vous pouvez appliquer un libellé à cette valeur.
6. Cliquez sur la cellule *Valeurs* de la ligne *âge*, puis sur le bouton à droite de la cellule pour ouvrir la boîte de dialogue Libellés de valeur.
7. Saisissez 999 dans le champ Valeur.
8. Saisissez Non répondu dans le champ Libellé.
9. Cliquez sur **Ajouter** pour ajouter le libellé au fichier de données.
10. Cliquez sur **OK** pour enregistrer vos modifications et revenir dans l'éditeur de données.

## Valeurs manquantes d'une variable de chaîne

Les valeurs manquantes des variables de chaîne sont gérées de la même façon que les valeurs manquantes des variables numériques. Cependant, contrairement aux valeurs numériques, les champs vides dans les variables de chaîne ne sont pas désignés comme données manquantes par défaut. Ils sont interprétés comme des chaînes de caractères vides.

1. Cliquez sur l'onglet **Vue de variable** en bas de la fenêtre de l'éditeur de données.
2. Cliquez sur la cellule *Manquante* de la ligne *sexe*, puis sur le bouton à droite de la cellule pour ouvrir la boîte de dialogue Valeurs manquantes.
3. Sélectionnez **Valeurs manquantes discrètes**.
4. Saisissez NR dans la première zone de texte.

Les valeurs manquantes des variables de chaîne distinguent les majuscules des minuscules. Par conséquent, la valeur *nr* n'est pas traitée comme une valeur manquante.

5. Cliquez sur **OK** pour enregistrer vos modifications et revenir dans l'éditeur de données.  
Vous pouvez à présent ajouter un libellé pour la valeur manquante.
6. Cliquez sur la cellule *Valeurs* de la ligne *sexe*, puis sur le bouton à droite de la cellule pour ouvrir la boîte de dialogue Libellés de valeur.
7. Saisissez NR dans le champ Valeur.
8. Saisissez Non répondu dans le champ Libellé.
9. Cliquez sur **Ajouter** pour ajouter le libellé au projet.
10. Cliquez sur **OK** pour enregistrer vos modifications et revenir dans l'éditeur de données.





---

## Chapitre 4. Examen des statistiques récapitulatives pour chaque variable

La présente section traite des mesures récapitulatives simples et de la façon dont le niveau de mesure d'une variable influence le type de statistiques devant être utilisé. Nous utiliserons le fichier de données *demo.sav*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.

---

### Niveau de mesure

Différentes mesures récapitulatives sont adaptées à différents types de données, selon le niveau de mesure :

**Nominales** : Données ayant un nombre limité de valeurs ou de catégories distinctes (par exemple, sexe ou situation de famille). Elles sont parfois également qualifiées de **données qualitatives**. 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 = *Célibataire* et 1 = *Marié*). Il existe deux types essentiels de données catégorielles :

- **Nominal** : Données catégorielles dont les catégories n'ont aucun ordre inhérent. Par exemple, une catégorie d'emploi de type *ventes* n'est pas supérieure ou inférieure à une catégorie d'emploi de type *marketing* ou *étude*.
- **Ordinal** : Données catégorielles dont les catégories possèdent un ordre significatif, mais pour lesquelles il n'existe aucune distance mesurable entre les catégories. Par exemple, les valeurs *élevée*, *moyenne* et *faible*, mais il est impossible de calculer la "distance" entre ces valeurs.

**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**.

---

### Mesures récapitulatives pour données catégorielles

Pour les données catégorielles, la mesure récapulative la plus courante est le nombre ou le pourcentage d'observations dans chaque catégorie. Le **mode** est la catégorie ayant le plus grand nombre d'observations. Pour les données ordinales, la **médiane** (valeur au-dessus ou au-dessous de laquelle se trouve la moitié des observations) peut également être une mesure récapulative utile s'il existe un grand nombre de catégories.

La procédure Fréquences produit des tables de fréquences qui affichent le nombre et le pourcentage d'observations pour chaque valeur observée d'une variable.

1. A partir des menus, sélectionnez :

**Analyse > Statistiques descriptives > Effectifs...**

Remarque : Cette fonction nécessite l'option Statistiques de base.

2. Sélectionnez *Possède un agenda électronique [pda]* et *Possède un téléviseur [tv]*, et déplacez-les vers la liste Variable(s).

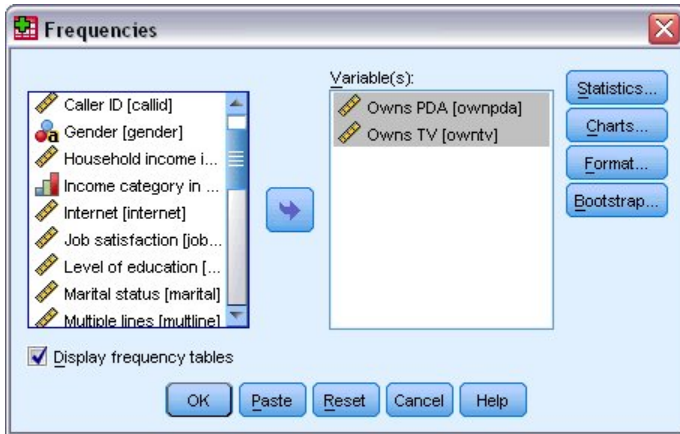


Figure 26. Variables catégorielles sélectionnées pour l'analyse

3. Cliquez sur OK pour exécuter la procédure.

**Frequency Table**

**Owns PDA**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	5093	79.6	79.6	79.6
	Yes	1307	20.4	20.4	100.0
Total		6400	100.0	100.0	

**Owns TV**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	63	1.0	1.0	1.0
	Yes	6337	99.0	99.0	100.0
Total		6400	100.0	100.0	

Figure 27. Tables de fréquences

Les tables de fréquences apparaissent dans la fenêtre du visualiseur. Les tables de fréquences révèlent que seuls 20,4 % des personnes possèdent un agenda électronique, mais que la quasi-totalité possèdent une télévision (99,0 %). Ces informations ne semblent pas vraiment pertinentes, mais il peut être intéressant d'en savoir plus sur le petit groupe de personnes qui ne possèdent pas de télévision.

## Graphiques pour données catégorielles

Vous pouvez afficher graphiquement les informations dans une table de fréquences avec un graphique à barres ou un graphique circulaire.

1. Ouvrez à nouveau la boîte de dialogue Fréquences. (Les deux variables doivent toujours être sélectionnées.)

Vous pouvez utiliser le bouton Rappeler boîte de dialogue de la barre d'outils pour revenir rapidement aux dernières procédures utilisées.



Figure 28. Bouton Rappeler boîte de dialogue

2. Cliquez sur **Graphiques**.
3. Cliquez sur **Graphiques à barres**, puis sur **Poursuivre**.
4. Cliquez sur **OK** dans la boîte de dialogue principale pour exécuter la procédure.

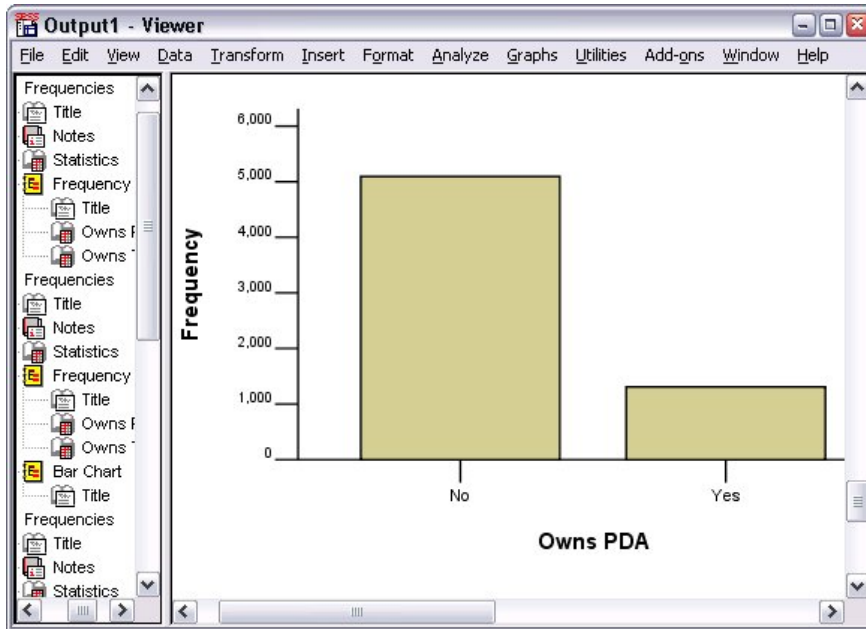


Figure 29. Graphique à barres

Outre les tables de fréquences, les mêmes informations sont à présent affichées sous forme de graphiques à barres ; vous pouvez ainsi voir que la plupart des personnes n'ont pas d'agenda électronique alors que la quasi-totalité d'entre elles possèdent une télévision.

## Mesures récapitulatives pour variables d'échelle

De nombreuses mesures récapitulatives sont disponibles pour les variables d'échelle, dont :

- **Mesures de la tendance centrale** : Les mesures les plus courantes de la tendance centrale sont la **moyenne** (moyenne arithmétique) et la **médiane** (valeur au-dessus ou au-dessous de laquelle se trouve la moitié des observations).
- **Mesures de la dispersion** : Les statistiques qui mesurent la quantité de variation ou de dispersion dans les données comprennent l'écart type, minimal et maximal.

1. Ouvrez à nouveau la boîte de dialogue Fréquences.
2. Cliquez sur **Réinitialiser** pour effacer les paramètres précédents.
3. Sélectionnez la variable *Revenu du ménage en milliers [revenu]* et déplacez-la dans la liste Variable(s).
4. Cliquez sur **Statistiques**.
5. Sélectionnez **Moyenne**, **Médiane**, **Ecart type**, **Minimum** et **Maximum**.
6. Cliquez sur **Poursuivre**.

7. Désélectionnez **Afficher les tables de fréquences** dans la boîte de dialogue principale Effectifs. (En général, les tables de fréquences ne sont pas très utiles pour les variables d'échelle car il peut exister presque autant de valeurs distinctes que d'observations dans le fichier de données.)
8. Cliquez sur **OK** pour exécuter la procédure.

Le tableau statistique des fréquences est affiché dans la fenêtre du visualiseur.

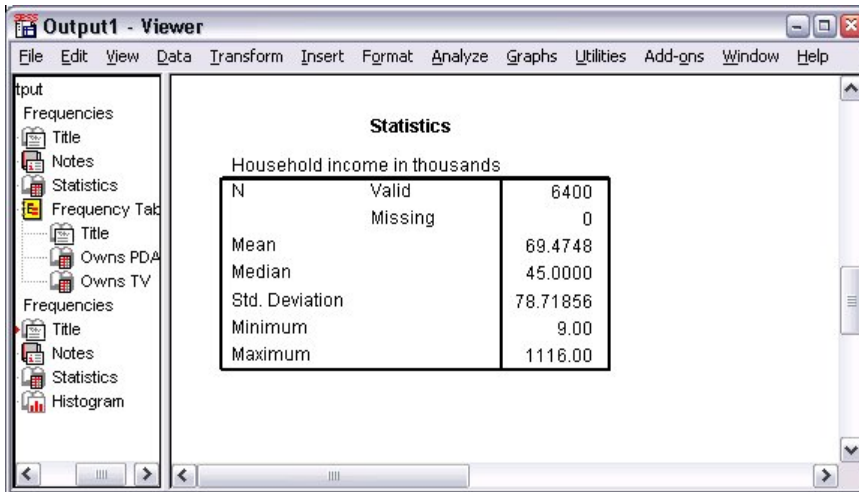


Figure 30. Tableau statistique des fréquences

Dans cet exemple, la différence entre la moyenne et la médiane est importante. La moyenne est plus importante que la médiane de quasiment 25 000, ce qui indique que les valeurs ne sont pas distribuées normalement. Vous pouvez vérifier visuellement la distribution grâce à un histogramme.

## Histogrammes pour variables d'échelle

1. Ouvrez à nouveau la boîte de dialogue Fréquences.
2. Cliquez sur **Graphiques**.
3. Cliquez sur **Histogrammes** et **Avec courbe gaussienne**.
4. Cliquez sur **Poursuivre**, puis sur **OK** dans la boîte de dialogue principale pour exécuter la procédure.

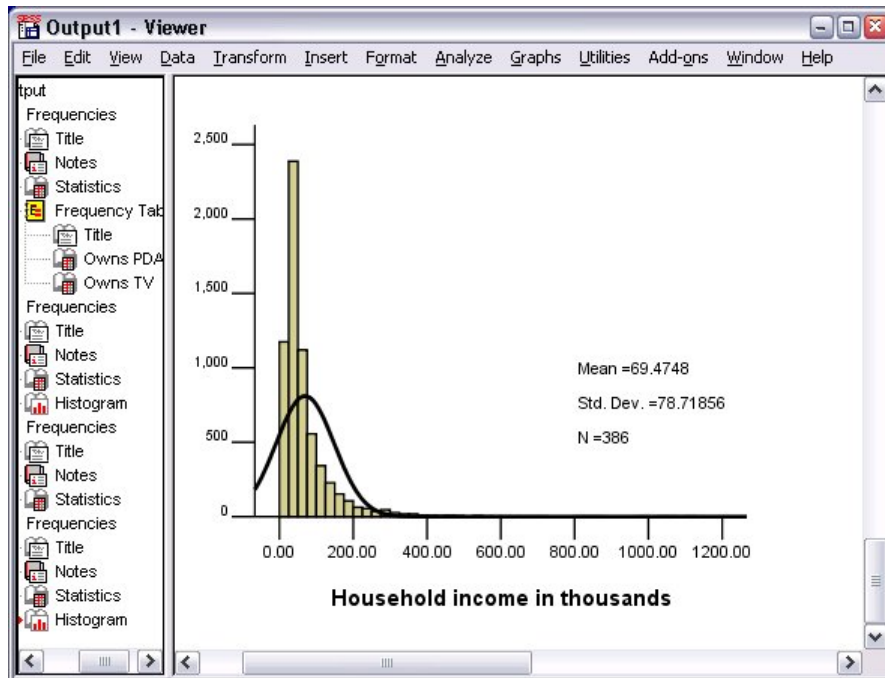


Figure 31. Histogramme

La grande majorité des observations est regroupée au bas de l'échelle, la plupart se trouvant au-dessous de 100 000. Quelques observations, cependant, se trouvent dans la plage 500 000 et au-delà (elles sont si peu nombreuses que vous devez modifier l'histogramme pour les voir). Ces valeurs très élevées pour quelques observations seulement ont un effet significatif sur la moyenne mais peu d'effet, voire aucun, sur la médiane ; cela signifie que, dans cet exemple, la médiane est un meilleur indicateur de la tendance centrale.



---

## Chapitre 5. Création et modification de graphiques

Vous pouvez créer et modifier des types de graphique divers et variés. Dans ce chapitre, nous allons créer et modifier des graphiques à barres. Vous pouvez appliquer les principes à n'importe quel type de graphique.

---

### Création de graphiques - Notions de base

Pour illustrer les notions de base de la création de graphiques, nous allons créer un graphique à barres du revenu moyen pour plusieurs niveaux de satisfaction professionnelle. Cet exemple utilise le fichier de données *demo.sav*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.

1. A partir des menus, sélectionnez :

**Graphes > Générateur de graphiques...**

La boîte de dialogue Générateur de graphiques est une fenêtre interactive qui vous permet d'obtenir l'aperçu d'un graphique avant que vous ne le génériez.

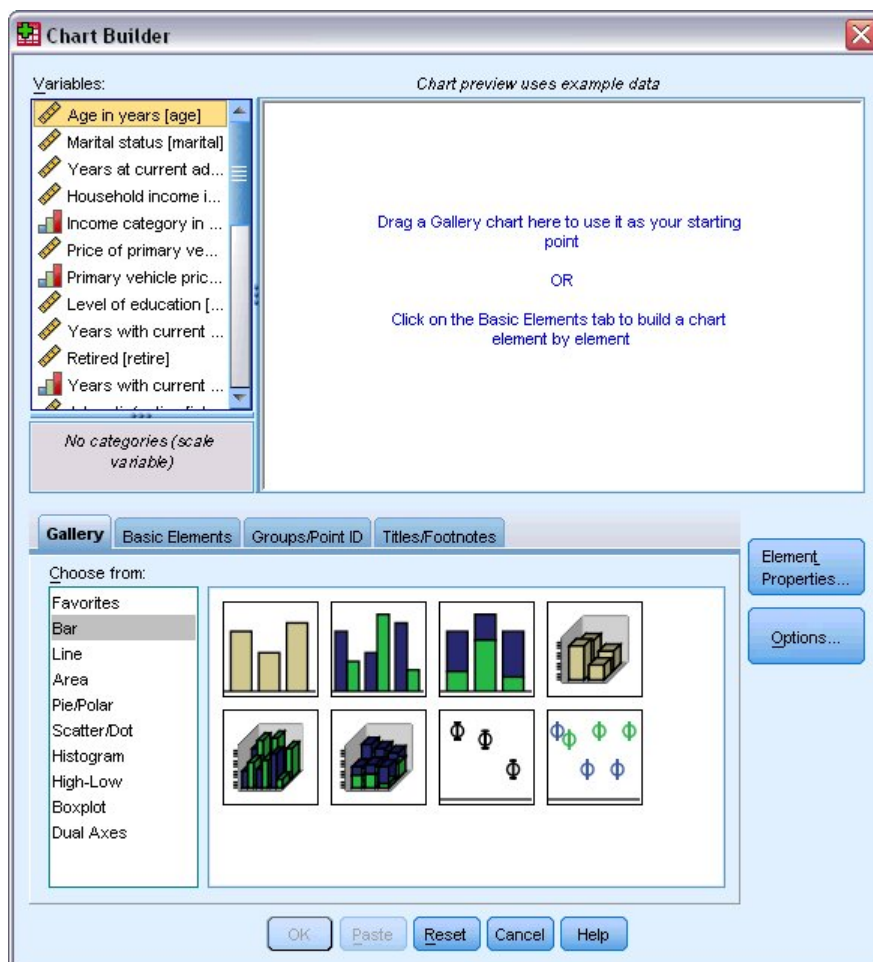


Figure 32. Boîte de dialogue Générateur de graphiques

## Utilisation de la galerie Générateur de graphiques

1. Cliquez sur l'onglet **Galerie** s'il n'est pas sélectionné.

La galerie inclut plusieurs graphiques différents prédéfinis, qui sont organisés par type de graphique. L'onglet Eléments de base fournit également des éléments de base (comme les axes et les éléments graphiques) pour créer des graphiques en partant de zéro, mais il est plus facile d'utiliser la galerie.

2. Cliquez sur **Barre** s'il n'est pas sélectionné.

Les icônes représentant les graphiques à barres disponibles dans la galerie apparaissent dans la boîte de dialogue. Les images doivent fournir suffisamment d'informations pour identifier le type de graphique spécifique. Si vous avez besoin de plus d'informations, vous pouvez également afficher une infobulle du graphique en posant votre curseur sur une icône.

3. Faites glisser l'icône du graphique à barres simples sur le « canevas », qui est en fait la zone étendue au-dessus de la galerie. Le Générateur de graphiques affiche un aperçu du graphique sur le canevas. Notez que les données utilisées pour dessiner le graphique ne sont pas vos données en cours. Il s'agit d'exemples.

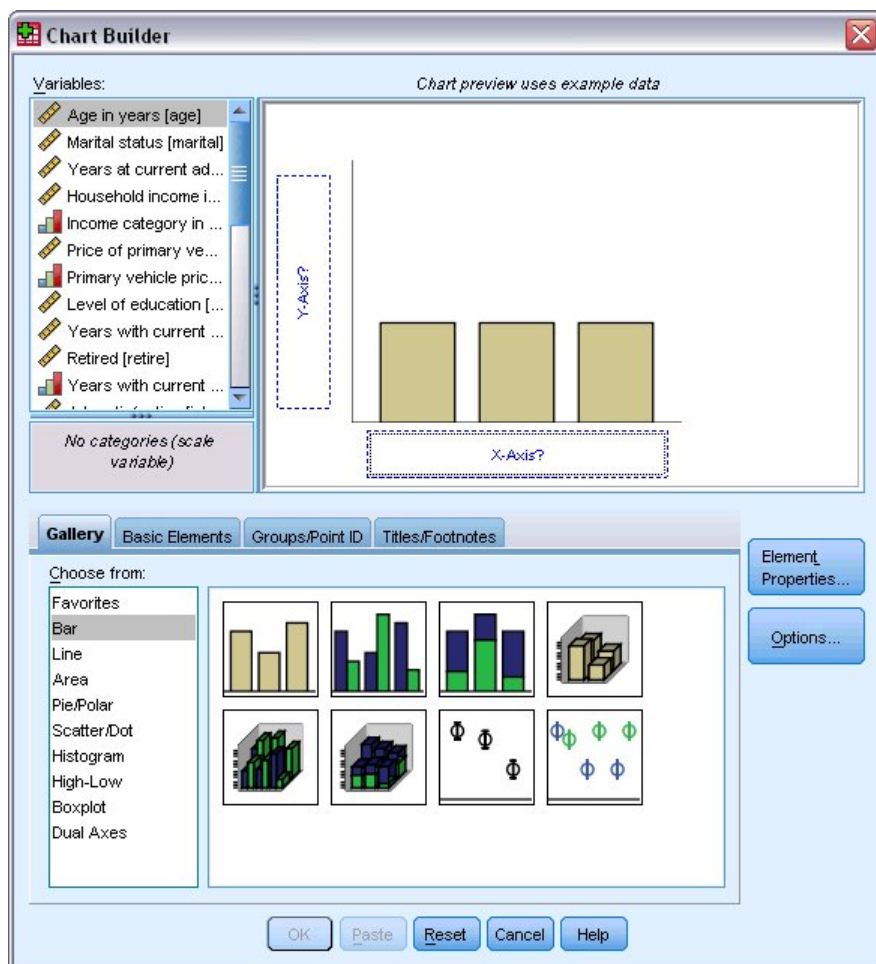


Figure 33. Graphique à barres sur le canevas du Générateur de graphiques

## Définition des variables et des statistiques

Bien qu'il y ait un graphique sur le canevas, il n'est pas complet car il n'y a aucune variable ou statistique pour contrôler la hauteur des barres et pour spécifier la catégorie de variable correspondant à chaque barre. Vous ne pouvez pas avoir de graphique sans variable ni statistique. Vous pouvez ajouter des variables en les glissant de la liste Variables qui se trouve à gauche du canevas.



Le niveau de mesure d'une variable est important dans le Générateur de graphiques. Vous allez utiliser la variable *Satisfaction professionnelle* de l'axe des X. Cependant, l'icône (qui ressemble à une règle) à côté de la variable indique que son niveau de mesure est défini en tant que variable d'échelle. Pour créer le graphique correct, vous devez utiliser un niveau de mesure catégoriel. Plutôt que de revenir et de modifier le niveau de mesure dans la vue de variable, vous pouvez le modifier temporairement dans le Générateur de graphiques.

1. Cliquez avec le bouton droit de la souris sur *Satisfaction professionnelle* dans la liste Variables et choisissez **Ordinal**. Ordinal correspond à un niveau de mesure approprié car les catégories dans *Satisfaction professionnelle* peuvent être ordonnées par niveau de satisfaction. Notez que l'icône change une fois le niveau de mesure modifié.
2. Faites glisser *Satisfaction professionnelle* de la liste Variables dans la zone d'insertion de l'axe des X. La zone d'insertion de l'axe des Y prend par défaut la statistique *Effectif*. Si vous souhaitez utiliser une autre statistique (comme pourcentage ou moyenne), vous pouvez facilement en changer. Vous n'utiliserez aucune de ces statistiques dans cet exemple, mais nous allons revoir le processus au cas où vous devez changer cette statistique plus tard.
3. Cliquez sur **Propriété des éléments** pour afficher la fenêtre Propriété des éléments.



Figure 34. Fenêtre Propriété des éléments

La fenêtre Propriété des éléments vous permet de modifier les propriétés des différents éléments du graphique. Ces éléments comprennent les éléments graphiques (comme les barres du graphique) et les axes du graphique. Sélectionnez un des éléments dans Modifier les propriétés de liste pour modifier les propriétés associées à cet élément. Notez également le X rouge situé à droite de la liste. Ce bouton

supprime un élément graphique du canevas. Puisque **Bâton1** est sélectionné, les propriétés affichées s'appliquent aux éléments graphiques, en particulier les éléments de barres.

La liste déroulante Statistiques affiche les statistiques spécifiques qui sont disponibles. Les mêmes statistiques sont habituellement disponibles pour chaque type de graphique. Notez que certaines statistiques requièrent que la zone d'insertion de l'axe des Y contienne une variable.

- Revenez à la boîte de dialogue Générateur de graphiques et faites glisser *Revenu du ménage en milliers* de la liste Variables vers la zone d'insertion de l'axe des Y. Puisque la variable sur l'axe des Y est sous forme d'échelle et que la variable de l'axe des X est catégorielle (ordinal est un type de niveau de mesure catégoriel), la zone d'insertion de l'axe des Y prend par défaut la statistique *Moyenne*. Il s'agit des variables et des statistiques souhaitées, il n'y a donc aucun besoin de modifier les propriétés de l'élément.

## Ajout de texte

Vous pouvez également ajouter des titres et des notes de bas de page au graphique.

- Cliquez sur l'onglet **Titres/Notes de bas de page**.
- Sélectionnez **Titre 1**.

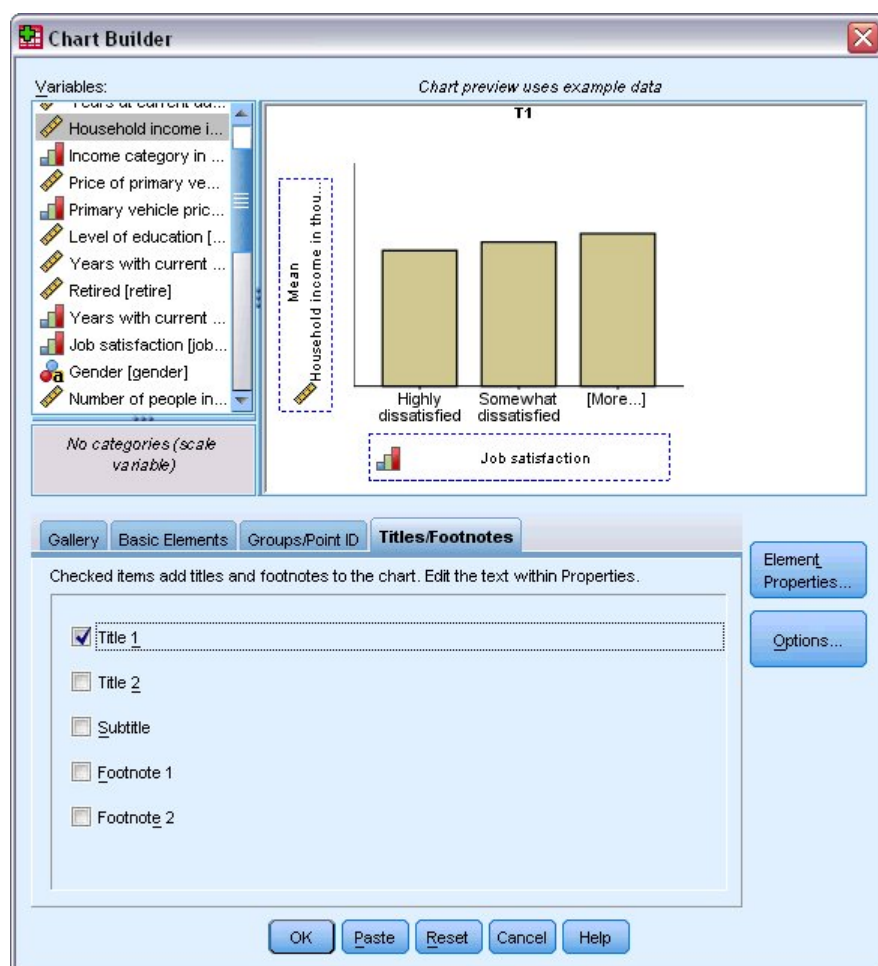


Figure 35. Titre 1 affiché sur le canevas

Le titre apparaît sur le canevas avec le libellé **T1**.

- Dans la fenêtre Propriété des éléments, sélectionnez **Titre 1** dans Modifier les propriétés de liste.
- Dans la zone de texte Contenu, saisissez Revenu par satisfaction professionnelle. Il s'agit du texte que le titre affichera.

5. Cliquez sur **Appliquer** pour enregistrer le texte. Bien que le texte ne soit pas affiché dans le Générateur de graphiques, il apparaît lorsque vous générez le graphique.

## Création du graphique

1. Cliquez sur **OK** pour créer le graphique à barres.

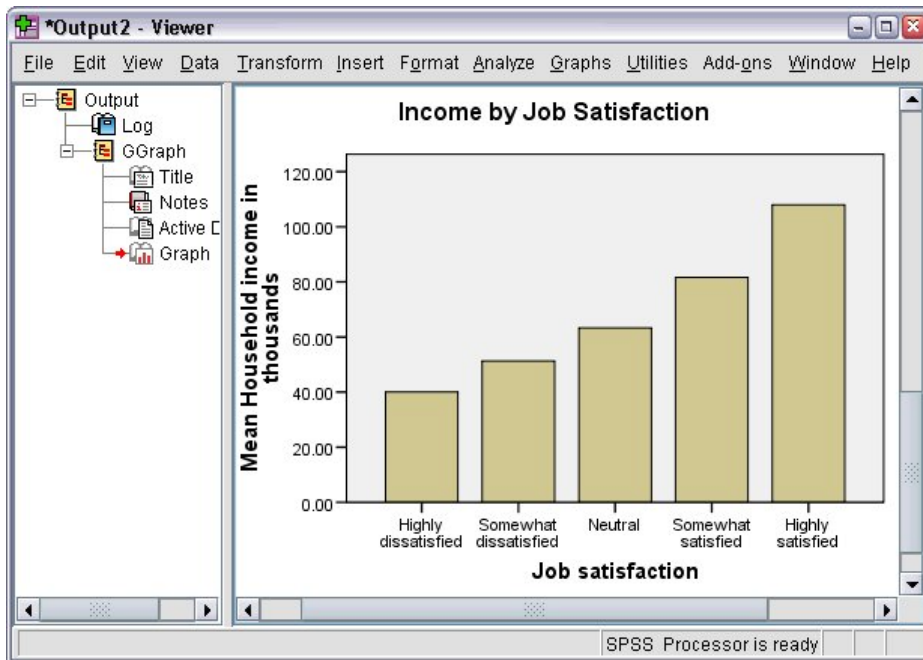


Figure 36. Graphique à barres

Le graphique à barres indique que les répondants les plus satisfaits de leur travaux ont tendance à avoir des revenus plus élevés.



---

## Chapitre 6. Utilisation de la sortie

Les résultats des procédures statistiques apparaissent dans le visualiseur. En fonction des choix effectués lors de l'exécution de la procédure, ces sorties peuvent prendre la forme d'un texte, ou de tableaux, de graphiques ou de graphiques statistiques. Cette section utilise les fichiers *viewertut.spv* et *demo.sav*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.

---

### Utilisation du visualiseur

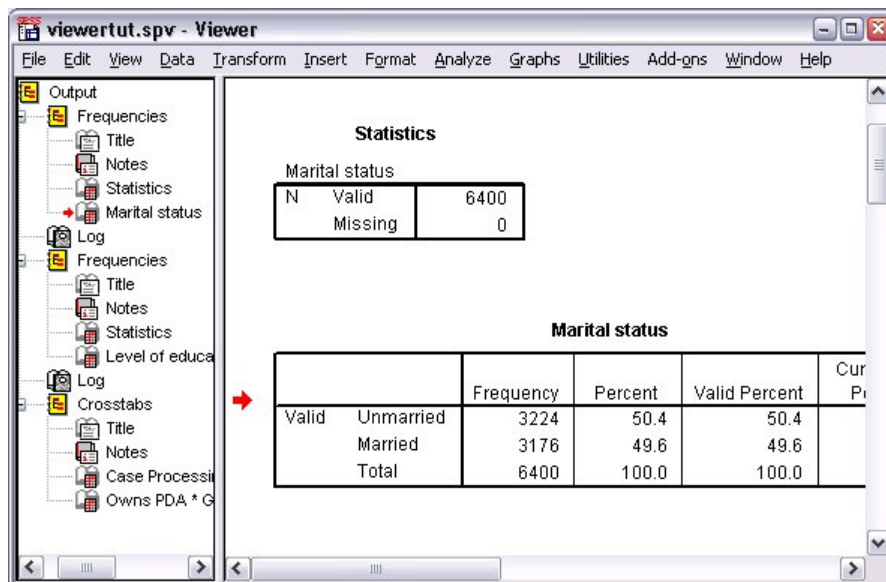


Figure 37. Visualiseur

La fenêtre du visualiseur est divisée en deux panneaux. Le **panneau de légende** contient la légende de toutes les informations stockées dans le visualiseur. Le **panneau de contenu** comporte les tableaux statistiques, les graphiques et les textes.

A l'aide des barres de défilement horizontale et verticale, parcourez le contenu de la fenêtre. Pour faciliter la navigation, cliquez sur un élément dans le panneau de légende pour l'afficher dans le panneau de contenu.

1. Cliquez sur la bordure droite du panneau de légende et faites-la glisser pour modifier sa largeur.  
Dans le panneau de légende, lorsqu'une icône en regard d'un élément représente un livre ouvert, cela indique qu'il est actuellement visible dans le visualiseur. même s'il n'est pas visible dans le panneau de contenu.
2. Pour masquer un tableau ou un graphique, double-cliquez sur l'icône du livre dans le panneau de légende.  
L'icône représente maintenant un livre fermé, ce qui signifie que les informations associées sont masquées.
3. Pour que la sortie apparaisse à nouveau, double-cliquez sur l'icône du livre fermé.  
Vous pouvez également masquer toutes les sorties d'une procédure statistique donnée ou l'intégralité des sorties dans le visualiseur.

4. Cliquez sur la case comportant le signe moins (-), à gauche de la procédure dont vous souhaitez masquer les résultats, ou cochez la case située en regard de l'élément figurant tout en haut du panneau de légende pour masquer la totalité des sorties.  
L'arborescence de la légende est réduite, indiquant visuellement que les résultats sont masqués.  
Vous pouvez également modifier l'ordre de l'affichage des sorties.
5. Dans le panneau de légende, cliquez sur les éléments à déplacer.
6. Faites glisser les éléments sélectionnés vers un nouvel emplacement de la légende.

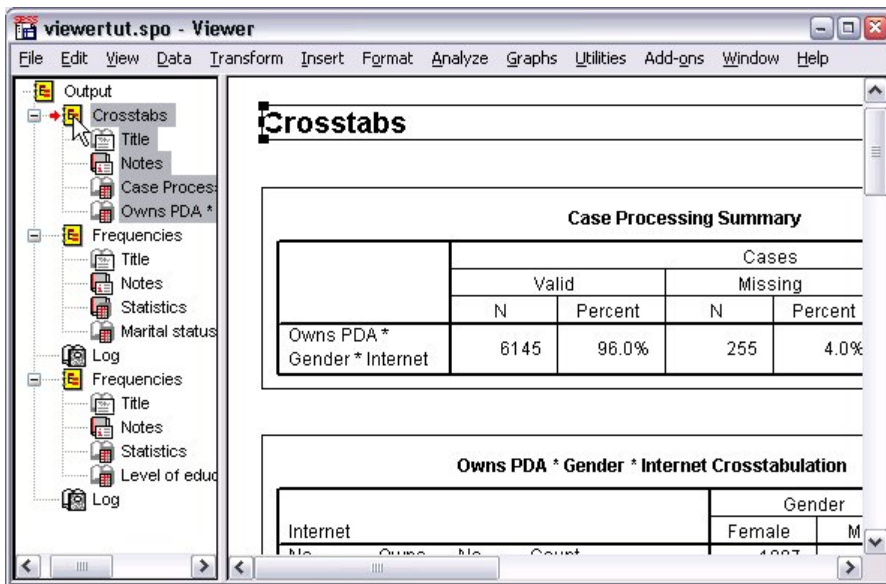


Figure 38. Réorganisation de la sortie dans le visualiseur

Vous pouvez également déplacer les éléments de sortie en cliquant dessus et en les faisant glisser dans le panneau de contenu.

## Utilisation de l'éditeur de tableau croisé dynamique

Les résultats de la plupart des procédures statistiques sont affichés dans des **tableaux croisés dynamiques**.

### Accès aux définitions de sortie

De nombreux termes statistiques sont affichés dans la sortie. Vous pouvez directement accéder aux définitions de ces termes dans le visualiseur.

1. Activez le tableau croisé (double-cliquez dessus) *Possède un agenda électronique \* Sexe \* Internet*.
2. Cliquez avec le bouton droit de la souris sur *Effectif théorique* et choisissez **Qu'est-ce que c'est ?** dans le menu contextuel.

La définition apparaît dans une fenêtre contextuelle.

The screenshot shows a window titled 'viewertut.spv - Viewer' with a menu bar (File, Edit, View, Insert, Pivot, Format, Analyze, Graphs, Utilities, Add-ons, Window, Help). The main content is a crosstabulation table titled 'Owns PDA \* Gender \* Internet Crosstabulation'. The table has columns for 'Internet' (No, Yes), 'Owns PDA' (No, Yes), 'Gender' (Female, Male), and 'Total'. A tooltip points to the 'Expected Count' column, stating: 'The number of cases that would be expected in the cell if the row and column variables are statistically independent or unrelated to one another.' The table data is as follows:

Internet			Gender		Total	
			Female	Male		
No	Owns PDA	No	Count	1897	1962	3859
			Expected Count	1903.4	1955.6	3859.0
		% within Owns PDA	50.3%	49.7%	100.0%	
	Total	Count	2224	2285	4509	
		Expected Count	2224.0	2285.0	4509.0	
	% within Owns PDA	49.3%	50.7%	100.0%		
Yes	Owns PDA	No	Count	513	512	1025
			Expected Count	512.5	512.5	1025.0
		% within Owns PDA	50.0%	50.0%	100.0%	
	Yes	Count	305	306	611	
		Expected Count	305.5	305.5	611.0	
	% within Owns PDA	49.9%	50.1%	100.0%		

Figure 39. Définition contextuelle

## Tableaux croisés dynamiques

Les tableaux produits par défaut risquent de ne pas présenter les informations aussi clairement que vous le voudriez. Avec les tableaux croisés dynamiques, vous pouvez transposer des lignes et des colonnes ("retourner" le tableau), ajuster l'ordre des données d'un tableau et modifier le tableau de nombreuses autres façons. Vous pouvez, par exemple, changer un tableau court et large en un tableau long et étroit, en transposant les lignes et les colonnes. La modification de la mise en page d'un tableau n'affecte pas les résultats. C'est, au contraire, une façon d'afficher vos informations d'une manière différente, plus attrayante.

1. S'il n'est pas déjà activé, double-cliquez sur le tableau croisé *Possède un agenda électronique \* Sexe \* Internet* pour l'activer.
2. Si la fenêtre Structures pivotantes n'est pas visible, à partir des menus, sélectionnez :

**Tableau croisé dynamique > Structure pivotante**

Les structures pivotantes vous permet de déplacer les données entre les colonnes, les lignes et les couches.

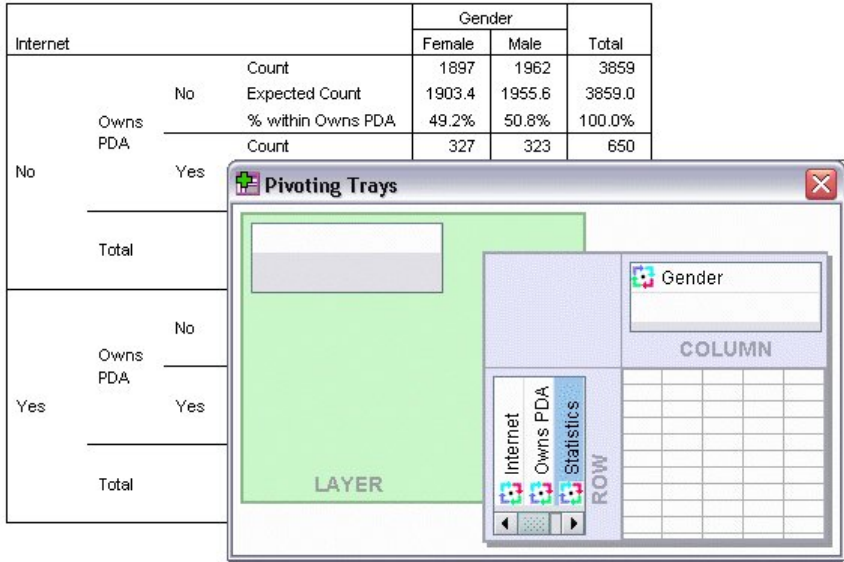


Figure 40. Structure pivotante

- Faites glisser l'élément *Statistiques* de la dimension de ligne à la dimension de colonne, sous *Sexe*. Le tableau est immédiatement reconfiguré pour refléter vos modifications.  
L'ordre des éléments dans la structure pivotante reflète l'ordre d'affichage des éléments dans le tableau.
- Faites glisser l'élément *Possède un agenda électronique* avant l'élément *Internet* dans la dimension de ligne pour inverser l'ordre de ces deux lignes.

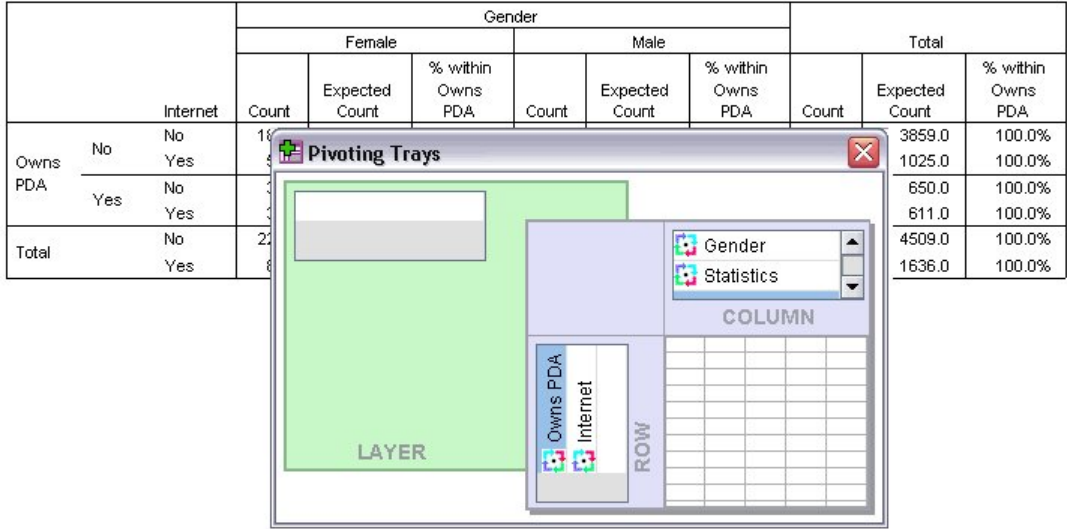


Figure 41. Permutation de lignes

### Création et affichage de couches

Les couches s'avèrent utiles pour les grands tableaux ayant des catégories d'informations imbriquées. En créant des couches, vous simplifiez le tableau qui devient plus facile à lire.

- Faites glisser l'élément *Sexe* de la dimension de colonne à la dimension de couche.



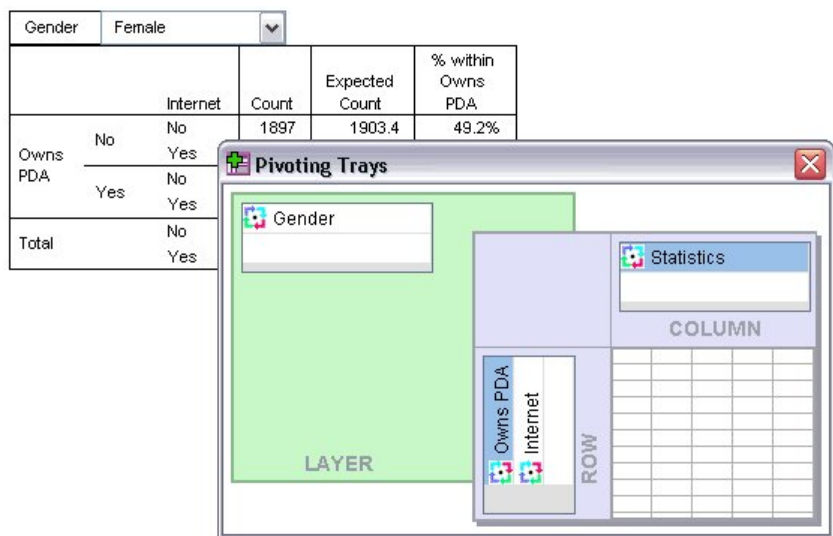


Figure 42. Icône de rotation Sexe dans la dimension Couche

Pour afficher une autre couche, sélectionnez une catégorie dans la liste déroulante du tableau.

## Modification de tableaux

A moins que vous ayez pris le temps de créer un modèle de tableaux personnalisé, les tableaux croisés dynamiques sont créés dans un format standard. Vous pouvez modifier le format d'un texte à l'intérieur d'un tableau. Les formats modifiables sont entre autres le nom, la taille, le style (italique ou gras) et la couleur de la police.

1. Double-cliquez sur le tableau *Niveau d'éducation*.
2. Si la barre d'outils de formatage n'est pas visible, à partir des menus, sélectionnez :  
**Affichage > Barre d'outils**
3. Cliquez sur le texte du titre, *Niveau d'éducation*.
4. Dans la liste déroulante des tailles de police de la barre d'outils, sélectionnez **12**.
5. Pour modifier la couleur du texte du titre, cliquez sur l'outil Couleur du texte et sélectionnez une couleur.

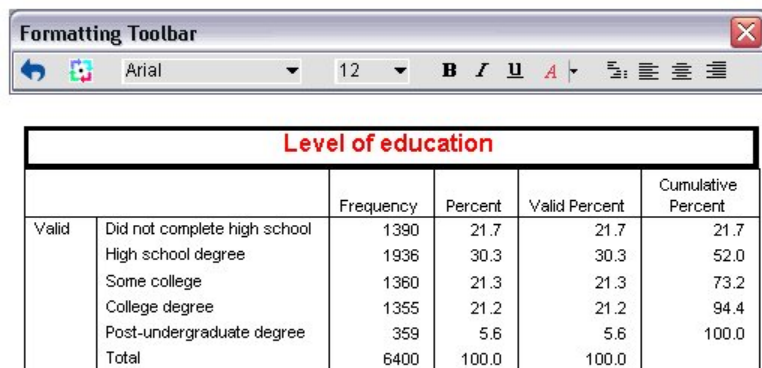


Figure 43. Texte du titre reformaté dans le tableau croisé dynamique

Vous pouvez également modifier le contenu des tableaux et des libellés. Vous pouvez, par exemple, modifier le titre de ce tableau.

6. Double-cliquez sur le titre.

7. Saisissez Niveau d'enseignement dans le nouveau libellé.

*Remarque* : Si vous modifiez les valeurs d'un tableau, les totaux et les autres statistiques ne sont pas recalculés.

## Masquage de lignes et de colonnes

Certaines données affichées dans un tableau ne sont pas forcément utiles ou rendent le tableau trop complexe. Vous pouvez masquer des lignes et des colonnes entières sans perdre de données.

1. S'il n'est pas déjà activé, double-cliquez sur le tableau *Niveau d'éducation* pour l'activer.
2. Cliquez sur la colonne *Pourcentage valide* pour la sélectionner.
3. Dans le menu Edition ou le menu contextuel, choisissez :  
**Sélectionner > Cellules de données et de libellés**
4. Choisissez **Masquer** dans le menu Affichage ou **Masquer la catégorie** dans le menu contextuel.  
La colonne est masquée, mais non supprimée.

**Education Level**

		Frequency	Percent	Cumulative Percent
Valid	Did not complete high school	1390	21.7	21.7
	High school degree	1936	30.3	52.0
	Some college	1360	21.3	73.2
	College degree	1355	21.2	94.4
	Post-undergraduate degree	359	5.6	100.0
	Total	6400	100.0	

Figure 44. Colonne *Pourcentage valide* masquée dans le tableau

Pour afficher de nouveau la colonne :

5. A partir des menus, sélectionnez :

**Affichage > Montrer Tout**

A l'instar des colonnes, les lignes peuvent être masquées et affichées.

## Modification des formats d'affichage des données

Vous pouvez modifier le format d'affichage des données dans les tableaux croisés dynamiques en toute facilité.

1. S'il n'est pas déjà activé, double-cliquez sur le tableau *Niveau d'éducation* pour l'activer.
2. Cliquez sur le libellé de colonne *Pourcentage* pour le sélectionner.
3. Dans le menu Edition ou le menu contextuel, choisissez :  
**Sélectionner > Cellules de Données**
4. Dans le menu Format ou le menu contextuel, choisissez **Propriétés de la cellule**.
5. Cliquez sur l'onglet **Valeur de format**.
6. Saisissez 0 dans le champ Décimales pour masquer tous les séparateurs décimaux dans cette colonne.

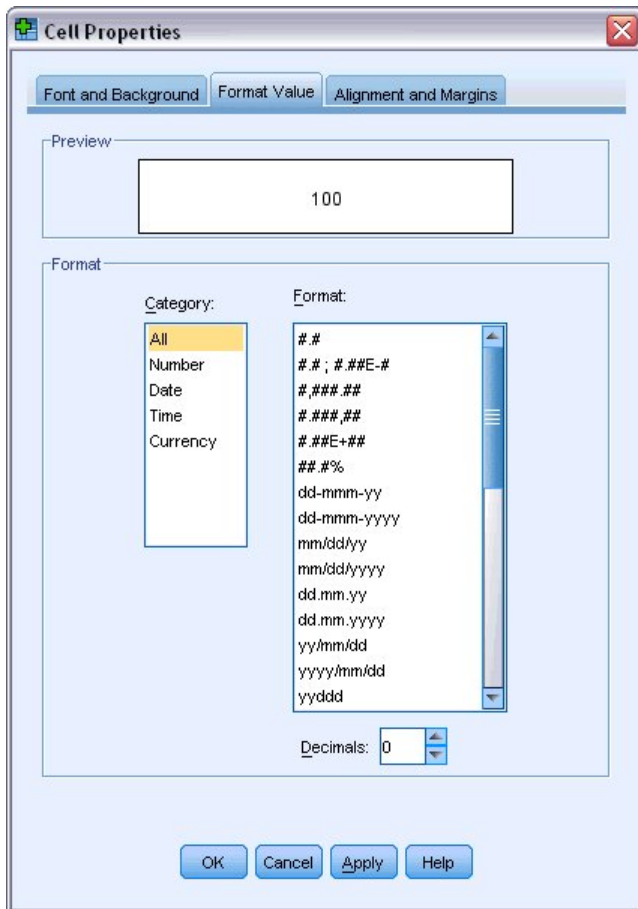


Figure 45. Propriétés des Cellules, onglet Valeur de format

Vous pouvez également modifier le type de données et le formatage dans cette boîte de dialogue.

7. Sélectionnez le type souhaité dans la liste Catégorie et le format de ce type dans la liste Format.
8. Cliquez sur **OK** ou **Appliquer** pour appliquer les modifications que vous avez apportées.

### Education Level

		Frequency	Percent	Cumulative Percent
Valid	Did not complete high school	1390	22	21.7
	High school degree	1936	30	52.0
	Some college	1360	21	73.2
	College degree	1355	21	94.4
	Post-undergraduate degree	359	6	100.0
	Total	6400	100	

Figure 46. Décimales masquées dans la colonne Pour cent

Les décimales sont à présent masquées dans la colonne *Pourcentage*.

## Modèles de tableaux

Le format de vos tableaux est primordial pour que les résultats soient clairs, précis et significatifs. Si votre tableau est complexe, les informations qu'il contient risquent d'être difficiles à comprendre.

## Utilisation de formats prédéfinis

1. Double-cliquez sur le tableau *Situation familiale*.
2. A partir des menus, sélectionnez :

**Format > Modèles de tableaux...**

La boîte de dialogue Modèles de tableaux répertorie différents styles prédéfinis. Sélectionnez un style dans la liste pour afficher son aperçu dans la fenêtre Exemple à droite.

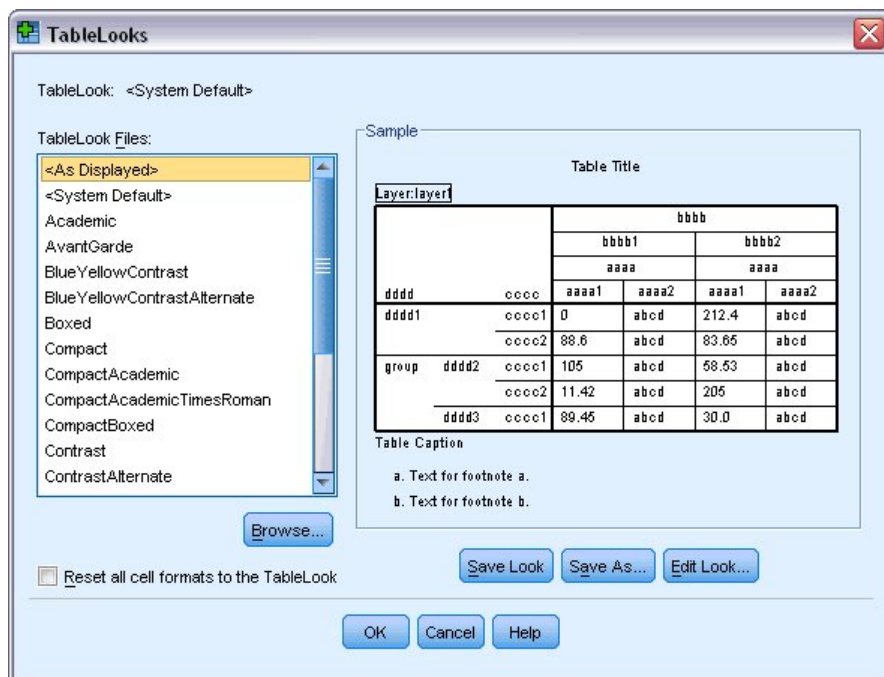


Figure 47. Boîte de dialogue Modèles de tableaux

Vous pouvez utiliser un style tel qu'il est ou le modifier pour l'adapter à vos besoins.

3. Pour utiliser un style existant, sélectionnez-le et cliquez sur **OK**.

## Personnalisation des styles Modèles de tableau

Vous pouvez personnaliser un format pour l'adapter à vos besoins. La quasi-totalité des aspects d'un tableau peut être personnalisée, de la couleur d'arrière-plan aux styles de bordure.

1. Double-cliquez sur le tableau *Situation familiale*.
2. A partir des menus, sélectionnez :  
**Format > Modèles de tableaux...**
3. Sélectionnez le style qui correspond le mieux au format de votre choix et cliquez sur **Modifier modèle**.
4. Cliquez sur l'onglet **Formats de cellule** pour afficher les options de formatage.

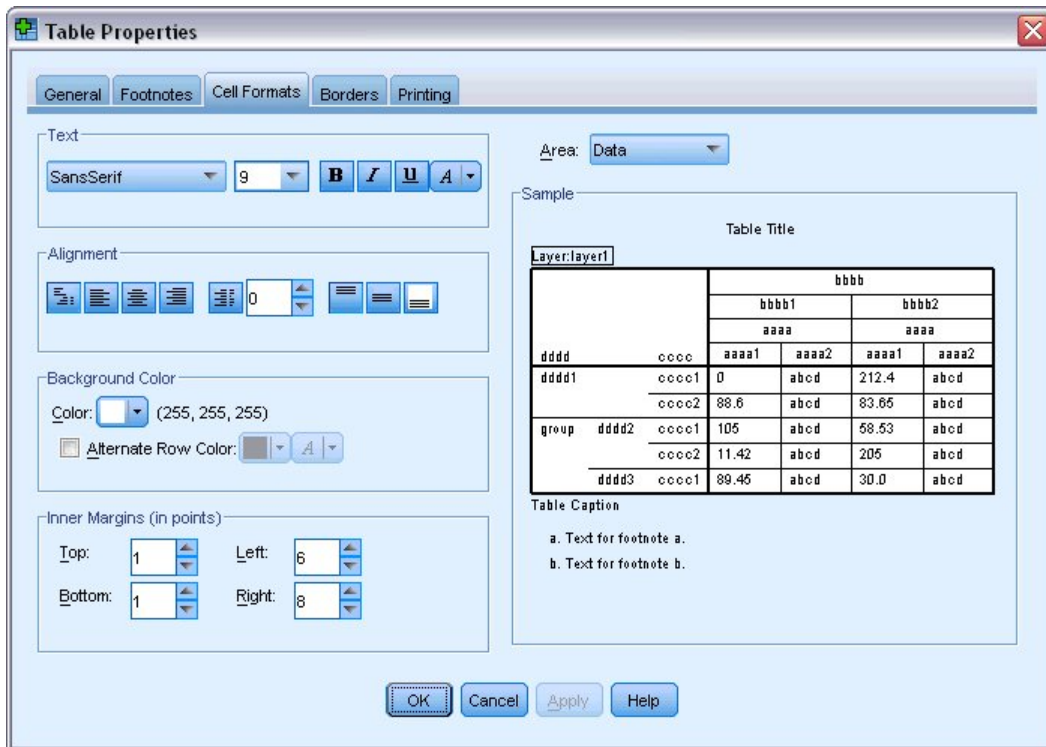


Figure 48. Boîte de dialogue Propriétés du tableau

Les options de formatage comprennent le nom, la taille, le style et la couleur de police. Elles incluent également l'alignement, la couleur du texte et de l'arrière-plan, et la taille des marges.

La fenêtre Echantillon située à droite fournit un aperçu de la modification de l'aspect du tableau. Chaque zone du tableau peut avoir des styles de format différents. Vous pouvez, par exemple, ne pas appliquer le même style au titre et aux données. Pour sélectionner une zone de tableau à modifier, sélectionnez le nom de la zone dans la liste déroulante Surfaces ou cliquez sur la zone dans la fenêtre Echantillon.

5. Sélectionnez **Données** dans la liste déroulante Surfaces.
6. Sélectionnez une nouvelle couleur dans la palette déroulante Arrière-plan.
7. Ensuite, sélectionnez une nouvelle couleur de texte.

La fenêtre Exemple illustre le nouveau style.

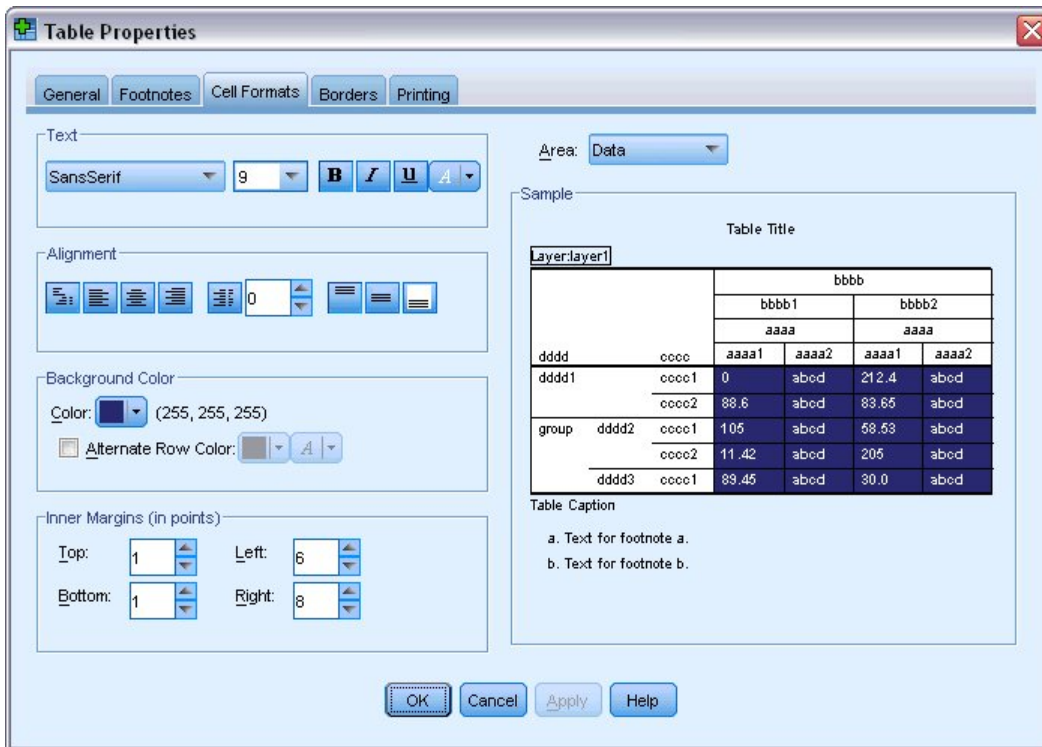


Figure 49. Modification des formats de cellule de tableau

8. Cliquez sur **OK** pour revenir à la boîte de dialogue Modèles de tableaux.
- Vous pouvez enregistrer le nouveau style pour l'utiliser ultérieurement dans d'autres tableaux.
9. Cliquez sur **Enregistrer sous**.
10. Accédez au répertoire cible et saisissez un nom pour le nouveau style dans la zone Nom du fichier.
11. Cliquez sur **Enregistrer**.
12. Cliquez sur **OK** pour appliquer vos modifications et revenir dans le visualiseur.

Le tableau contient à présent le format personnalisé que vous avez spécifié.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unmarried	3224	50.4	50.4	50.4
	Married	3176	49.6	49.6	100.0
	Total	6400	100.0	100.0	

Figure 50. Modèle de tableau personnalisé

## Modification des formats par défaut d'un tableau

Bien que vous puissiez modifier le format des tableaux créés, il peut s'avérer plus efficace de modifier une fois pour toutes le modèle de tableau par défaut.

Pour modifier le style du modèle de tableau par défaut pour vos tableaux croisés dynamiques, à partir des menus, sélectionnez :

**Edition > Options...**

1. Cliquez sur l'onglet **Tableaux croisés dynamiques** de la boîte de dialogue Options.

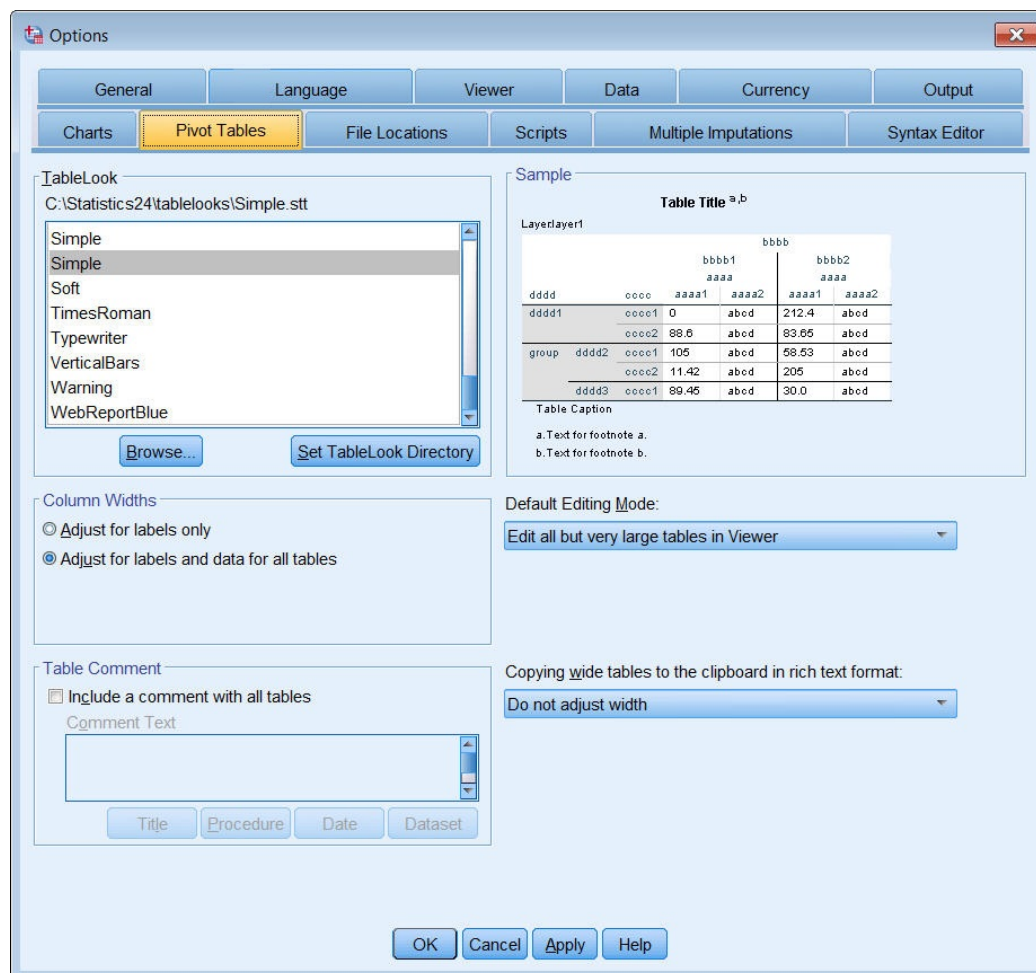


Figure 51. Boîte de dialogue Options

2. Sélectionnez le style du modèle de tableau à utiliser pour tous les nouveaux tableaux.

La fenêtre Echantillon à droite affiche un aperçu de chaque modèle de tableau.

3. Cliquez sur **OK** pour enregistrer les paramètres et fermer la boîte de dialogue.

Tous les tableaux créés après la modification du modèle de tableau par défaut suivent automatiquement les nouvelles règles de formatage.

## Personnalisation des paramètres de l'affichage initial

Les paramètres de l'affichage initial comprennent l'alignement des objets dans le visualiseur, qu'ils soient affichés ou masqués par défaut, et la largeur de la fenêtre du visualiseur. Pour modifier ces paramètres :



1. A partir des menus, sélectionnez :  
**Edition > Options...**
2. Cliquez sur l'onglet **Visualiseur**.

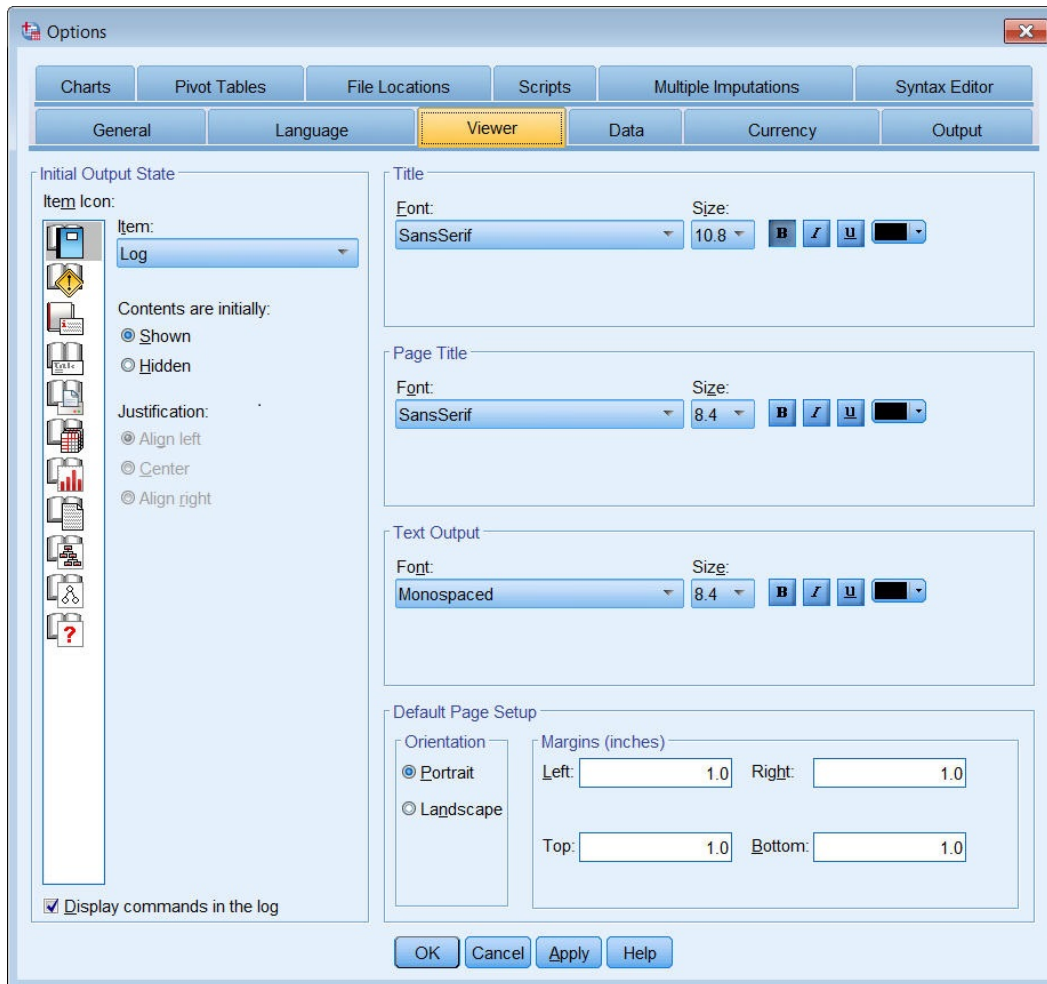


Figure 52. Options du visualiseur

Vos paramètres sont appliqués objet par objet. Vous pouvez, par exemple, personnaliser le mode d'affichage des graphiques sans modifier celui des tableaux. Il suffit de sélectionner l'objet à personnaliser et d'apporter les modifications.

3. Cliquez sur l'icône **Titre** pour afficher ses paramètres.
4. Cliquez sur **Centre** pour afficher tous les titres au centre (horizontal) du visualiseur.  
Vous pouvez également masquer des éléments, tels que le journal et les messages d'avertissement, qui ont tendance à rendre les sorties confuses. Cliquez deux fois sur une icône pour modifier automatiquement la propriété d'affichage de cet objet.
5. Double-cliquez sur l'icône **Avertissement** pour masquer les messages d'avertissement dans les sorties.
6. Cliquez sur **OK** pour enregistrer vos modifications et fermer la boîte de dialogue.

## Affichage de libellés de variable et de valeur

Dans la plupart des cas, il est plus judicieux d'afficher les libellés des variables et des valeurs plutôt que le nom de la variable et la valeur réelle des données. Toutefois, il peut arriver que vous vouliez afficher les noms et les libellés.

1. A partir des menus, sélectionnez :



## Edition > Options...

2. Cliquez sur l'onglet **Libellés de sortie**.

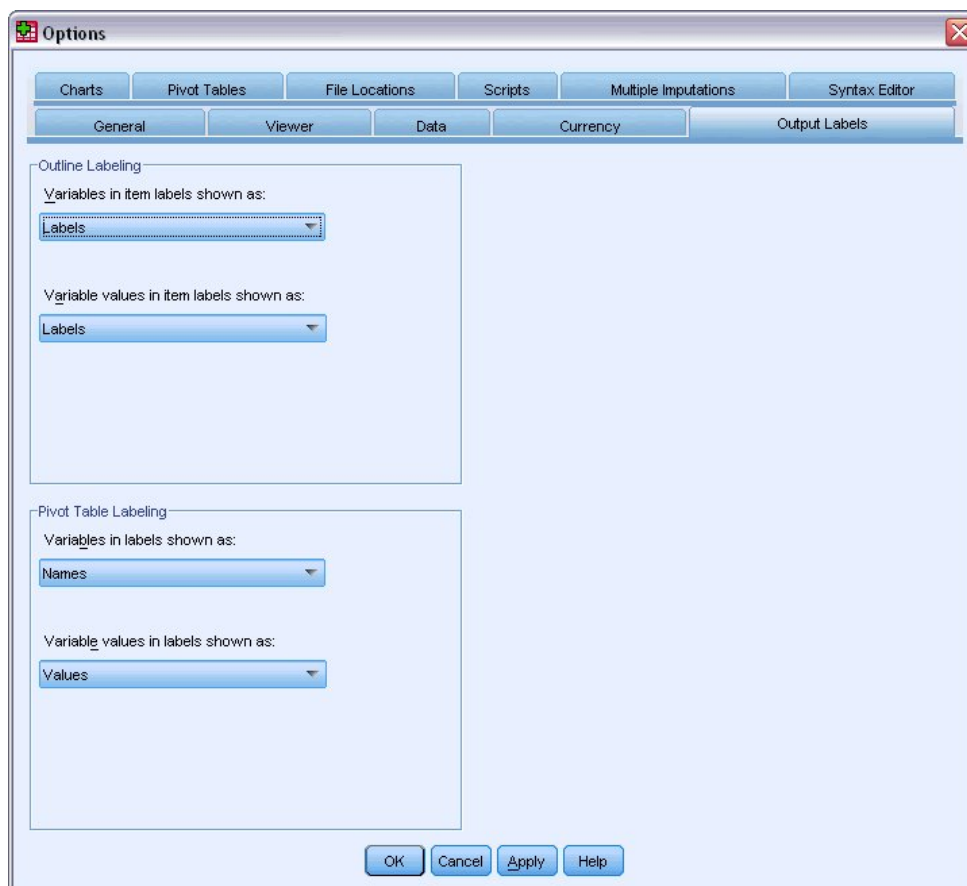


Figure 53. Paramètres Libellé tableau croisé dynamique

Vous pouvez indiquer des paramètres différents pour les panneaux de légende et de contenu. Par exemple, pour afficher des libellés dans la légende, et des noms de variable et des valeurs de données dans le contenu :

3. Dans le groupe Libellé tableau croisé dynamique, sélectionnez **Noms** dans la liste déroulante Noms des variables en libellés pour afficher des noms de variable à la place des libellés.
4. Sélectionnez ensuite **Valeurs** dans la liste déroulante Valeurs des variables en libellés pour afficher des valeurs de données à la place des libellés.

Les tableaux suivants générés au cours de la session refléteront ces modifications.

**marital**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	3224	50.4	50.4	50.4
	1	3176	49.6	49.6	100.0
	Total	6400	100.0	100.0	

Figure 54. Affichage des noms et valeurs de variable

---

## Utilisation de résultats dans d'autres applications

Vos résultats peuvent être utilisés dans de nombreuses applications. Vous pouvez, par exemple, inclure un graphique ou un tableau dans une présentation ou un rapport.

Les exemples suivants sont propres à Microsoft Word, mais ils peuvent fonctionner de la même façon dans d'autres applications de traitement de texte.

### Collage des résultats sous forme de tableaux Word

Vous pouvez coller des tableaux croisés dynamiques dans Word en tant que tableaux natifs Word. Tous les attributs de tableau, comme la taille et la couleur de la police, sont conservés. Du fait que le tableau est collé dans le format de tableau Word, vous pouvez le modifier dans Word comme tout autre tableau.

1. Cliquez sur un tableau dans le visualiseur pour le sélectionner.
2. A partir des menus, sélectionnez :  
**Edition > Copier**
3. Ouvrez l'application de traitement de texte.
4. A partir des menus du traitement de texte, sélectionnez :  
**Edition > Collage spécial...**
5. Sélectionnez **Texte mis en forme (RTF)** dans la boîte de dialogue Collage spécial.
6. Cliquez sur **OK** pour coller vos résultats dans le document actuel.

Le tableau est à présent affiché dans votre document. Vous pouvez appliquer des formats personnalisés, modifier les données et redimensionner le tableau en fonction de vos besoins.

### Collage des résultats sous forme de texte

Les tableaux croisés dynamiques peuvent être copiés dans d'autres applications sous forme de texte normal. Avec cette méthode, les styles de format ne sont pas conservés, mais vous pouvez modifier les données du tableau après l'avoir collé dans l'application cible.

1. Cliquez sur un tableau dans le visualiseur pour le sélectionner.
2. A partir des menus, sélectionnez :  
**Edition > Copier**
3. Ouvrez l'application de traitement de texte.
4. A partir des menus du traitement de texte, sélectionnez :  
**Edition > Collage spécial...**
5. Sélectionnez **Texte sans mise en forme** dans la boîte de dialogue Collage spécial.
6. Cliquez sur **OK** pour coller vos résultats dans le document actuel.

The screenshot shows a Microsoft Word window titled 'Document1 - Microsoft Word'. The main content area displays a dynamic pivot table with the following data:

Marital status		Frequency	Percent	Valid Percent	Cumulative
Valid	Unmarried	3224	50.4	50.4	50.4
	Married	3176	49.6	49.6	100.0
Total		6400	100.0	100.0	

The table is displayed with tabular alignment in the Word document. The status bar at the bottom indicates 'Page 1', 'Sec 1', '1/1', 'At 3.5"', 'Ln 13', 'Col 1', and 'REC TRK EXT OVR'.

Figure 55. Tableau croisé dynamique affiché dans Word

Chaque colonne du tableau est séparée par des tabulations. Vous pouvez modifier la largeur des colonnes en ajustant les marques de tabulation dans votre application de traitement de texte.

## Exportation des résultats vers un fichier Microsoft Word, PowerPoint ou Excel

Vous pouvez exporter les résultats vers un fichier Microsoft Word, PowerPoint ou Excel. Vous pouvez exporter les éléments sélectionnés ou tous les éléments du visualiseur. Cette section utilise les fichiers *msouttut.spv* et *demo.sav*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.

*Remarque* : L'exportation vers PowerPoint n'est disponible que sous les systèmes d'exploitation Windows et n'est pas disponible dans la version Student.

Dans le panneau de légende du visualiseur, vous pouvez sélectionner les éléments spécifiques que vous souhaitez exporter ou exporter tous les éléments ou tous les éléments visibles.

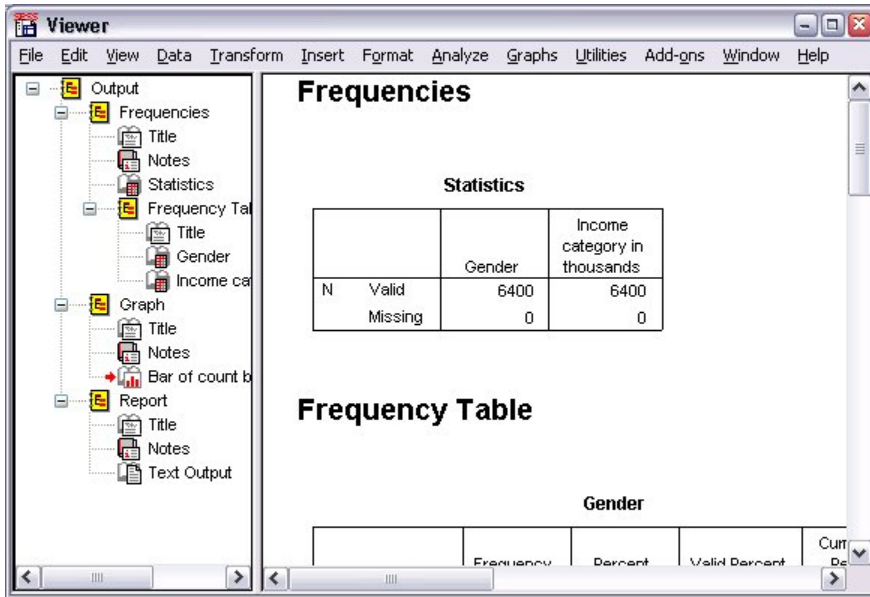


Figure 56. Visualiseur

1. A partir du menu du visualiseur, sélectionnez :

**Fichier > Exporter...**

Au lieu d'exporter tous les objets du visualiseur, vous pouvez n'exporter que les objets visibles (livres ouverts dans le panneau de légende) ou ceux sélectionnés dans le panneau de légende. Si vous n'avez sélectionné aucun élément du panneau de légende, vous n'avez pas la possibilité d'exporter les objets sélectionnés.

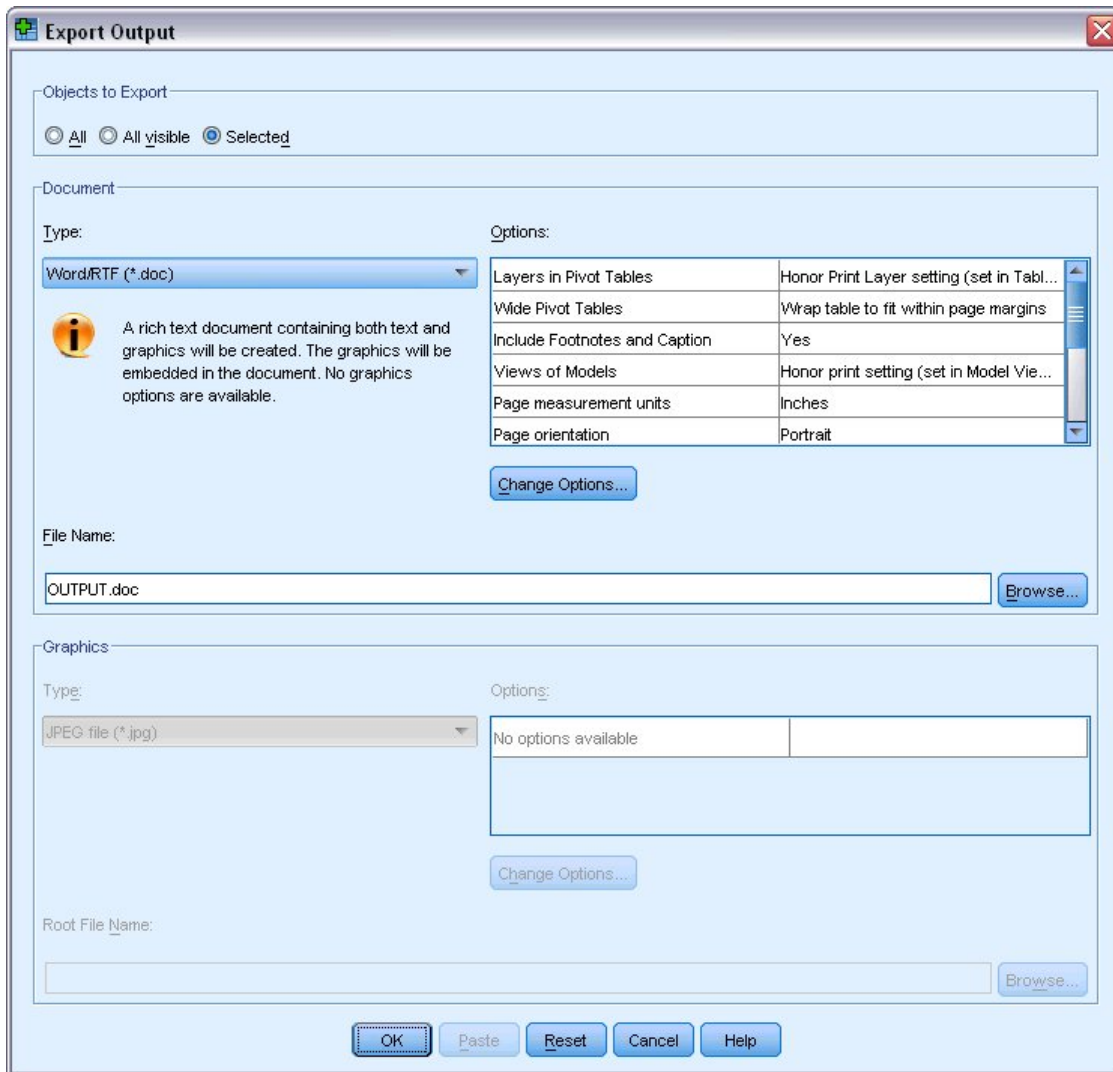


Figure 57. Boîte de dialogue Exporter sortie

2. Dans le groupe Objets à exporter, sélectionnez **Tous**.
3. Dans la liste déroulante Type, sélectionnez **Word/RTF (\*.doc)**.
4. Cliquez sur **OK** pour générer le fichier Word.

Lorsque vous ouvrez le fichier obtenu dans Word, vous pouvez observer le mode d'exportation des résultats. Les notes, qui ne sont pas des objets visibles, apparaissent dans Word car vous avez choisi d'exporter tous les objets.

Les tableaux croisés dynamiques deviennent des tableaux Word et conservent l'intégralité de leur formatage d'origine (polices, couleurs, bordures, etc.).

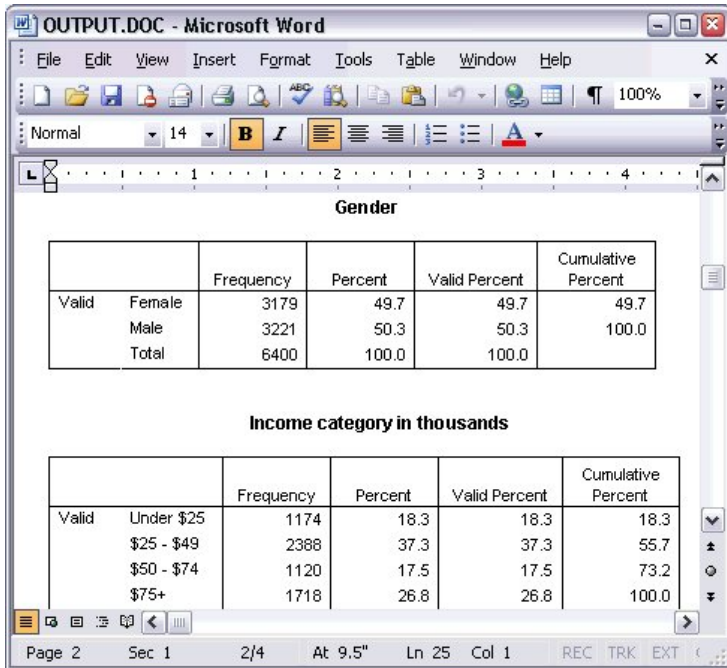


Figure 58. Tableaux croisés dynamiques dans Word

Les graphiques sont inclus dans le document Word sous la forme d'images graphiques.

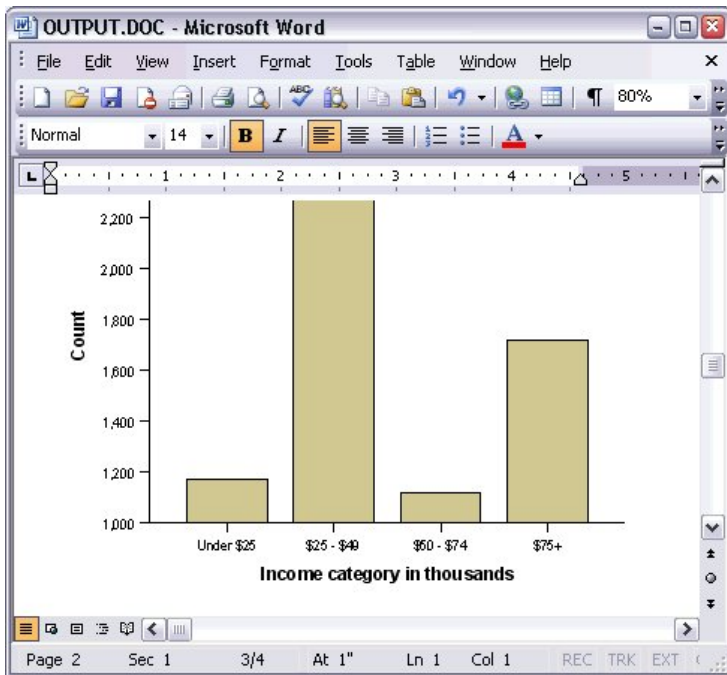


Figure 59. Graphiques dans Word

Les sorties sous forme de texte sont affichés avec la même police que celle utilisée pour l'objet texte dans le visualiseur. Toutes les sorties de texte nécessitent une police à point fixe (espacement unique) pour un alignement correct.

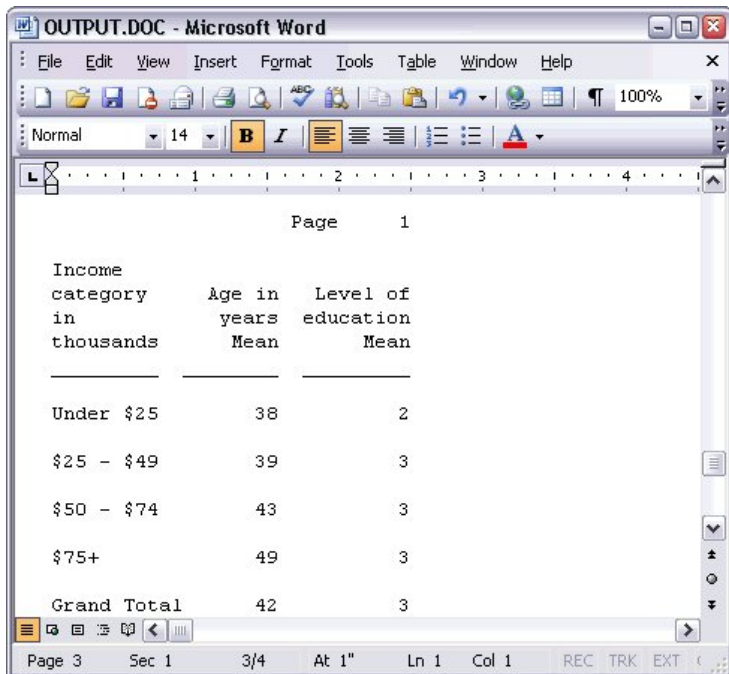


Figure 60. Sortie texte dans Word

Si vous exportez des données vers un fichier PowerPoint, chaque élément exporté est placé sur une diapositive distincte. Les tableaux croisés dynamiques exportés dans PowerPoint deviennent des tableaux Word et conservent l'intégralité du formatage d'origine (polices, couleurs, bordures, etc.).

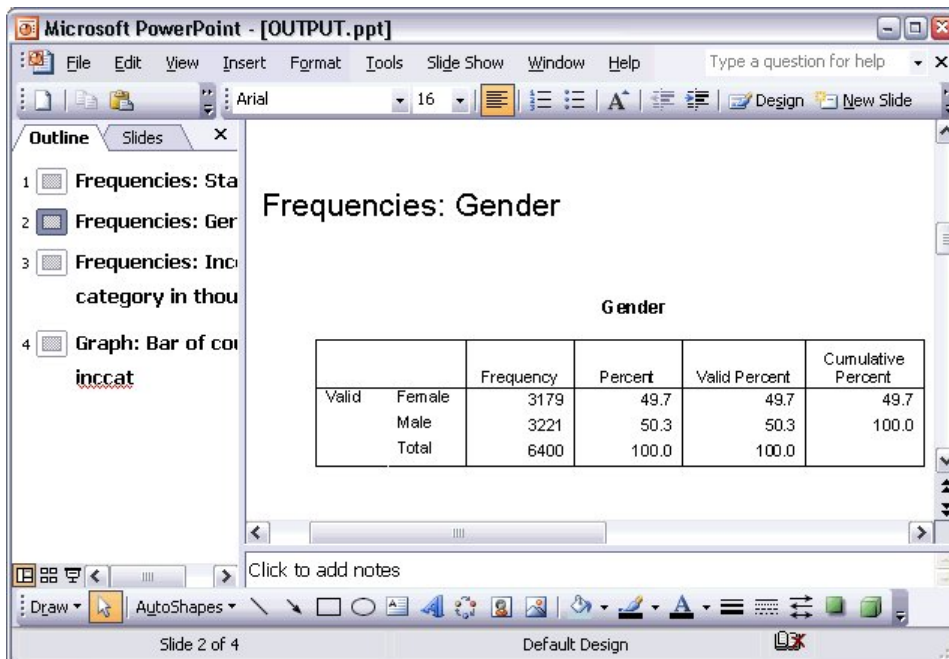


Figure 61. Tableaux croisés dynamiques dans PowerPoint

Les graphiques sélectionnés pour l'exportation vers PowerPoint sont intégrés dans le fichier PowerPoint.

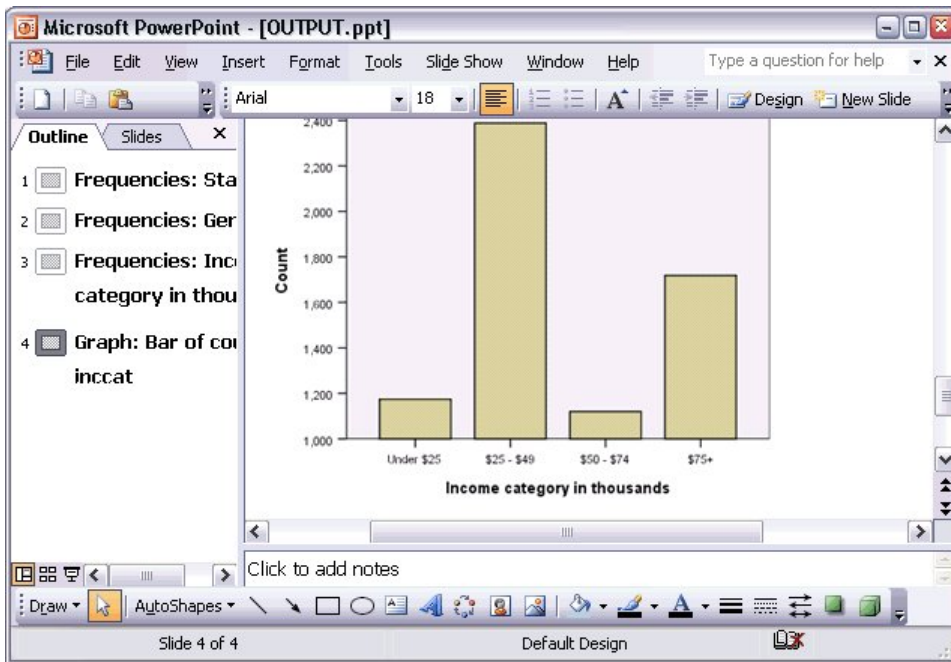


Figure 62. Graphiques dans PowerPoint

Remarque : L'exportation vers PowerPoint n'est disponible que sous les systèmes d'exploitation Windows et n'est pas disponible dans la version Student.

Si vous exportez des résultats vers un fichier Excel, l'exportation est réalisée de manière différente.

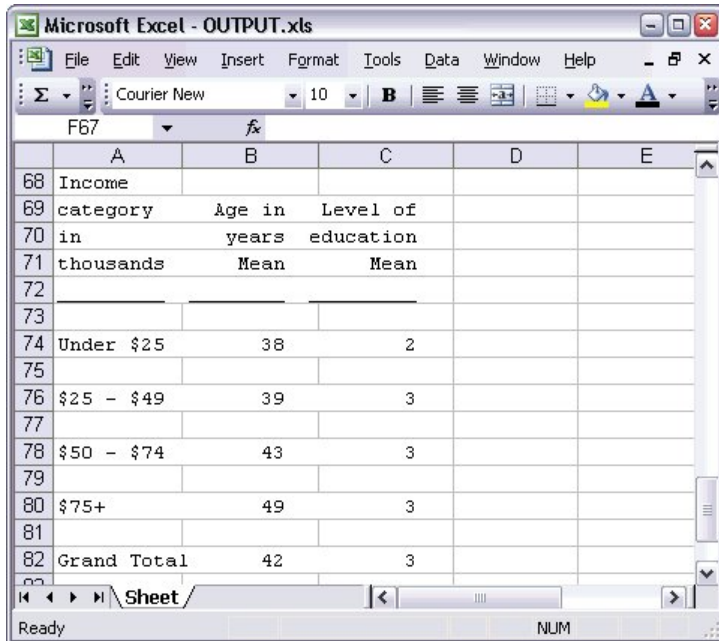
		Frequency	Percent	Valid Percent	Cumulative Percent
<b>Gender</b>					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	3,179	49.7	49.7	49.7
	Male	3,221	50.3	50.3	100.0
	Total	6,400	100.0	100.0	
<b>Income category in thousands</b>					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under \$25	1,174	18.3	18.3	18.3
	\$25 - \$49	2,368	37.3	37.3	55.7
	\$50 - \$74	1,120	17.5	17.5	73.2
	\$75+	1,718	26.8	26.8	100.0
	Total	6,400	100.0	100.0	

Figure 63. Output.xls dans Excel

Les lignes, les colonnes et les cellules des tableaux croisés dynamiques deviennent des lignes, des colonnes et des cellules Excel.



Chaque ligne de la sortie texte est une ligne dans le fichier Excel, avec le contenu de toute la ligne dans une seule cellule.



	A	B	C	D	E
68	Income				
69	category	Age in	Level of		
70	in	years	education		
71	thousands	Mean	Mean		
72					
73					
74	Under \$25	38	2		
75					
76	\$25 - \$49	39	3		
77					
78	\$50 - \$74	43	3		
79					
80	\$75+	49	3		
81					
82	Grand Total	42	3		

Figure 64. Sortie texte dans Excel

## Exportation des résultats vers PDF

Vous pouvez exporter les éléments sélectionnés ou tous les éléments du visualiseur dans un fichier PDF (Portable Document Format).

1. A partir des menus de la fenêtre du visualiseur contenant le résultat à exporter vers un fichier PDF, sélectionnez :

**Fichier > Exporter...**

2. Dans la boîte de dialogue Exporter la sortie, choisissez **Portable Document Format** dans la liste déroulante du format d'exportation.

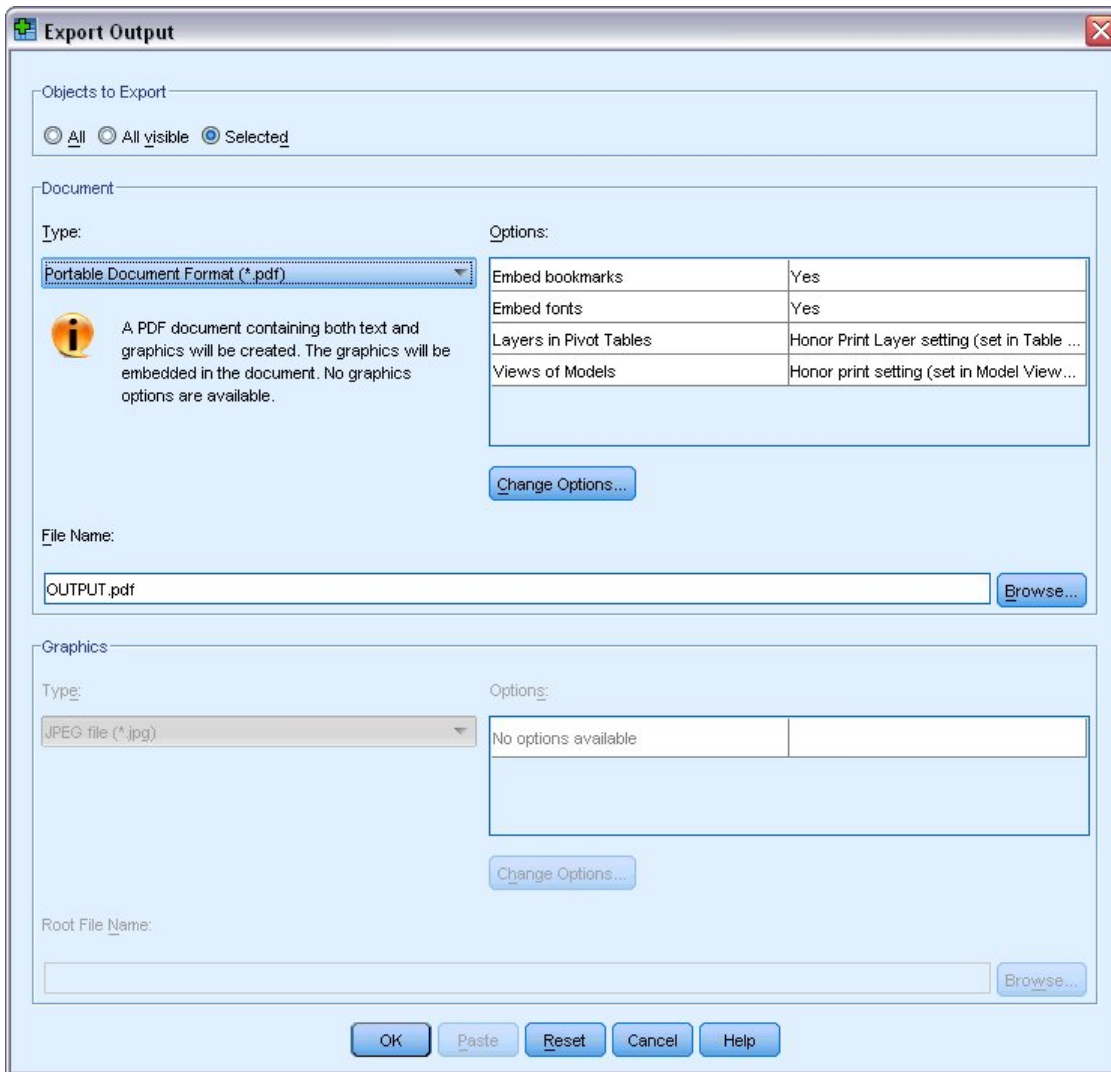


Figure 65. Boîte de dialogue Exporter sortie

- Le panneau de légende du document du visualiseur est converti en signets dans le fichier PDF pour une navigation plus facile.
- La taille de la page et son orientation, les marges, le contenu et l'affichage des en-têtes et des pieds de page, ainsi que la taille des graphiques imprimés d'un document PDF sont gérés par les options de mise en page (menu Fichier, option Mise en page de la fenêtre du visualiseur).
- La résolution (PPP) du document PDF correspond au paramètre de résolution en cours de l'imprimante par défaut ou sélectionnée (qui peut être changée à l'aide de l'option Mise en page). La résolution maximale est de 1 200 ppp. Si le paramètre de résolution de l'imprimante est supérieur, la résolution du document PDF sera de 1 200 ppp. *Remarque* : Une haute résolution peut générer des résultats médiocres si les documents sont imprimés sur des imprimantes de résolution inférieure.

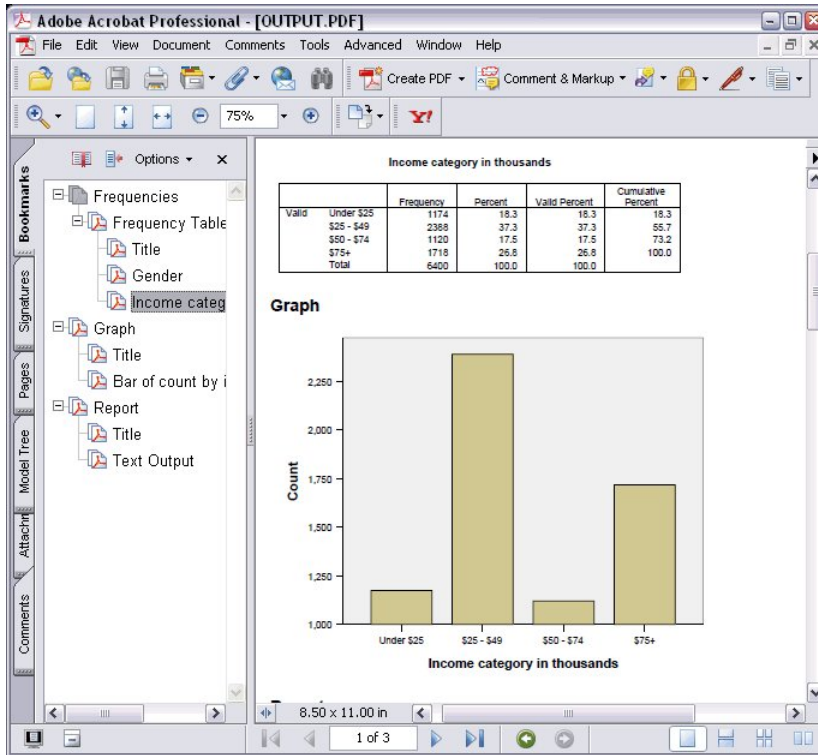


Figure 66. Fichier PDF avec signets

## Exportation des résultats vers HTML

Vous pouvez également exporter des résultats vers HTML. Lorsque vous enregistrez les sorties au format HTML, toutes les sorties non graphiques sont exportées dans un fichier HTML unique.

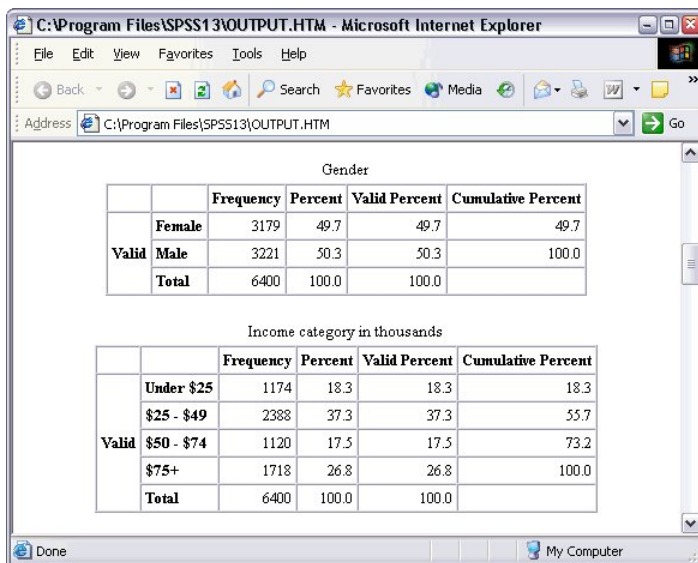


Figure 67. Output.htm dans un navigateur Web

Lorsque vous exportez les résultats vers le format HTML, vous pouvez également exporter des graphiques, mais pas vers un fichier unique.

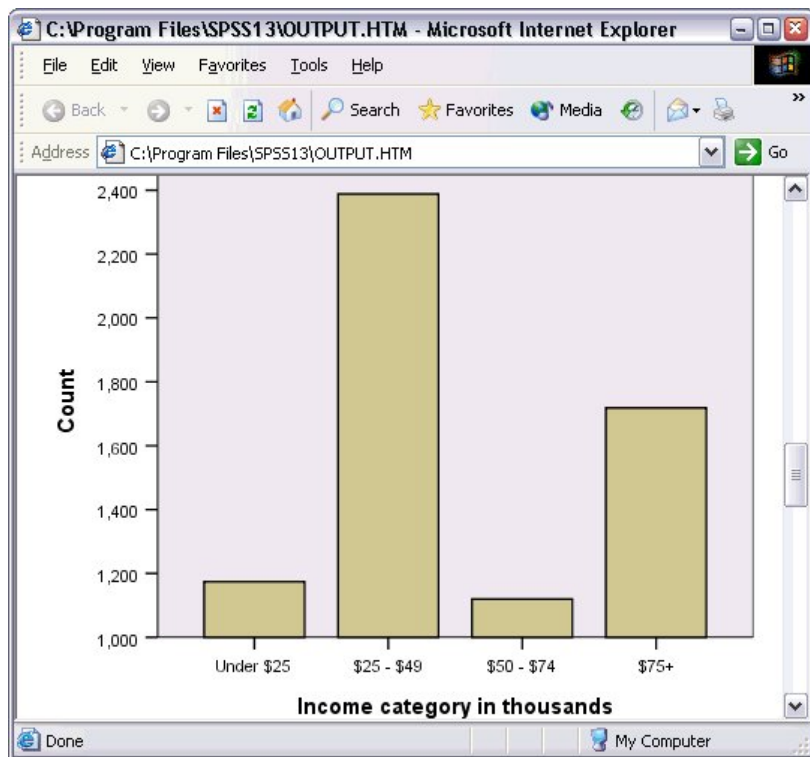


Figure 68. Graphique au format HTML

Chaque graphique sera enregistré sous la forme d'un fichier au format que vous spécifiez, et les références à ces fichiers graphiques seront placées dans le document HTML. Une option vous permet également d'exporter tous les graphiques (ou seulement certains) vers des fichiers graphiques distincts.

---

## Chapitre 7. Utilisation de la syntaxe

Vous pouvez enregistrer et automatiser de nombreuses tâches courantes grâce au puissant langage de commande. Il fournit également des fonctionnalités qui ne se trouvent ni dans les menus ni dans les boîtes de dialogue. La plupart des commandes sont accessibles depuis les menus et boîtes de dialogue. Cependant, certaines options et commandes ne sont disponibles qu'en utilisant le langage de commande. Le langage de commande permet également d'enregistrer vos travaux dans un fichier de syntaxe afin de vous permettre de relancer votre analyse à une date ultérieure.

Un fichier de syntaxe de commande est un fichier texte simple contenant des commandes de syntaxe IBM SPSS Statistics. Vous pouvez ouvrir une fenêtre de syntaxe et y entrer des commandes directement, mais il est souvent plus simple de laisser la boîte de dialogue faire une partie ou tout le travail pour vous.

Les exemples de ce chapitre utilisent le fichier de données *demo.sav*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.

*Remarque* : La syntaxe de commande n'est pas disponible dans la version Student.

---

### Collage d'une syntaxe

Le bouton **Coller**, affiché dans la plupart des boîtes de dialogue, constitue le moyen le plus simple de créer une syntaxe.

1. Ouvrez le fichier de données *demo.sav*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.
2. A partir des menus, sélectionnez :  
**Analyse > Statistiques descriptives > Effectifs...**
3. Sélectionnez la variable *Situation familiale [situation]* et déplacez-la vers la liste Variable(s).
4. Cliquez sur **Graphiques**.
5. Dans la boîte de dialogue Graphiques, sélectionnez **Graphiques à barres**.
6. Dans le groupe Valeurs du graphique, sélectionnez **Pourcentages**.
7. Cliquez sur **Poursuivre**. Cliquez ensuite sur **Coller** pour copier la syntaxe créée grâce aux sélections effectuées dans la boîte de dialogue de l'Editeur de syntaxe.

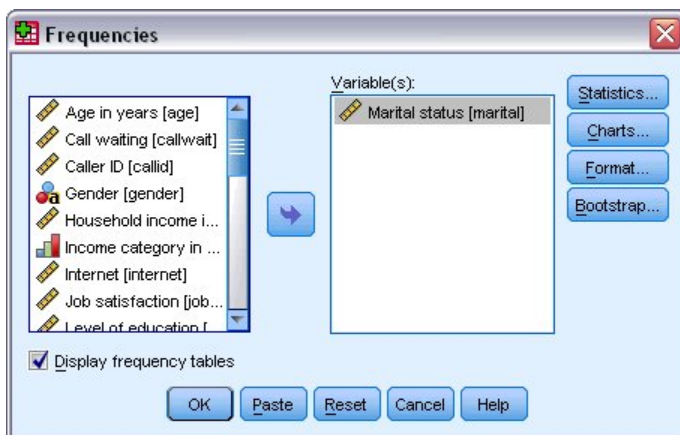


Figure 69. Boîte de dialogue Effectifs

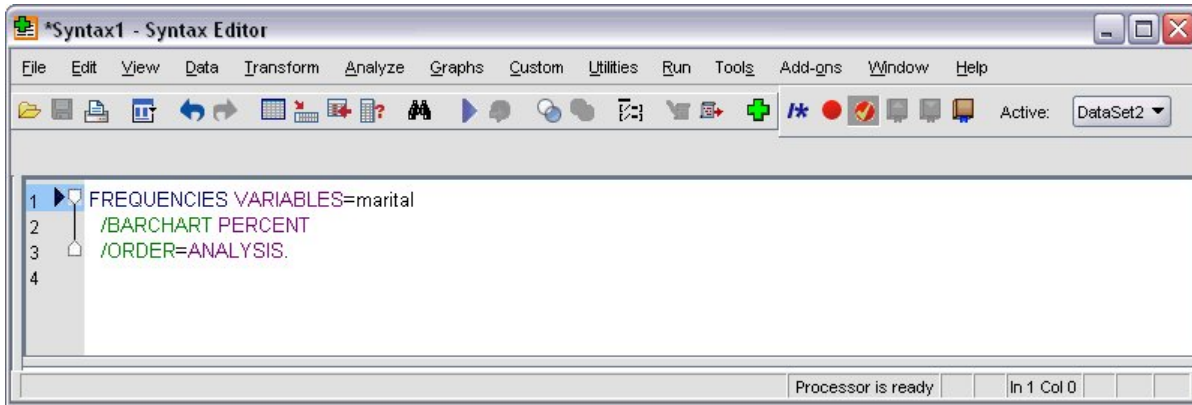


Figure 70. Syntaxe d'effectifs

8. Pour exécuter la syntaxe actuellement affichée, à partir des menus, sélectionnez :  
Exécuter > Sélection

## Modification d'une syntaxe

Dans la fenêtre de syntaxe, vous pouvez modifier la syntaxe. Par exemple, vous pouvez modifier la sous-commande /BARCHART pour afficher des effectifs à la place des pourcentages. (Une sous-commande est indiquée par une barre oblique.) Si vous connaissez le mot clé pour afficher les effectifs, vous pouvez le saisir directement. Si vous ne le connaissez pas, vous pouvez obtenir une liste des mots clés disponibles pour la sous-commande en positionnant le curseur n'importe où après le nom de la sous-commande et en appuyant simultanément sur Ctrl+barre d'espace. Ceci affiche le contrôle de saisie semi-automatique pour la sous-commande.

Supprimez le mot clé PERCENT de la sous-commande BARCHART.

Appuyez sur Ctrl+ barre d'espace.

1. Cliquez sur l'élément libellé **FREQ** pour les effectifs. En cliquant sur un élément dans le contrôle de saisie semi-automatique, il sera inséré à l'emplacement actuel du curseur.  
Par défaut, le contrôle de saisie semi-automatique vous proposera une liste des termes disponibles au fur et à mesure que vous saisissez. Supposons par exemple que vous souhaiteriez inclure un graphique circulaire avec un graphique à barres. Le graphique circulaire est spécifié par une sous-commande distincte.
2. Appuyez sur Entrée après le mot clé **FREQ** et tapez une barre oblique pour indiquer le début de la sous-commande.

L'Editeur de syntaxe vous propose une liste des sous-commandes pour la commande en cours.

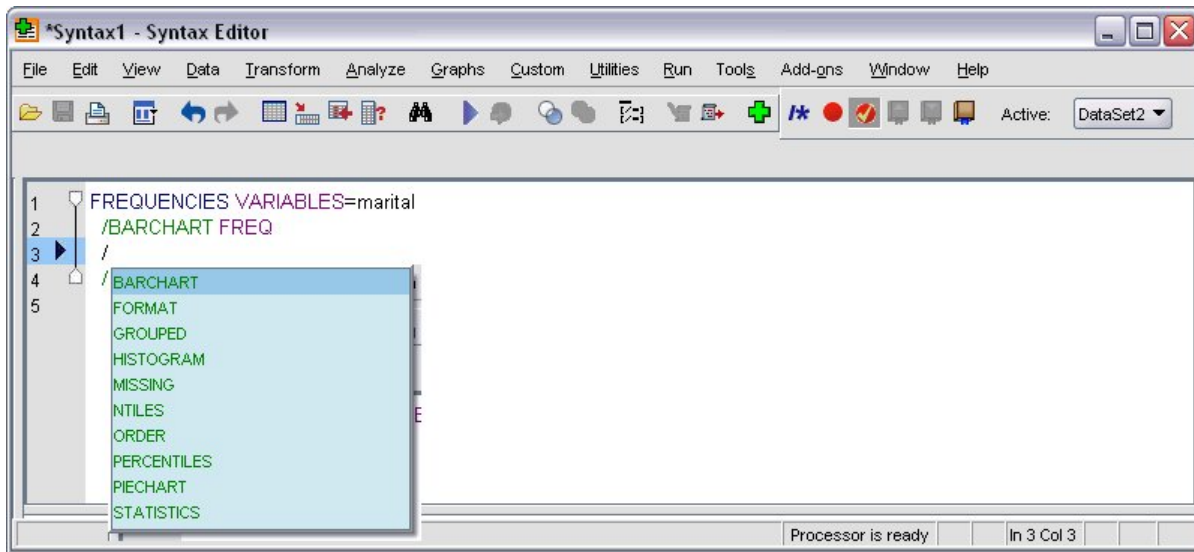


Figure 71. Contrôle de saisie semi-automatique affichant les sous-commandes

Pour obtenir plus d'aide sur la commande en cours, appuyez sur la touche F1. Vous êtes directement dirigé vers un guide de référence détaillé de la syntaxe de commande en cours.

Vous avez peut-être remarqué que le texte affiché dans la fenêtre de syntaxe est en couleur. La codification par couleur vous permet d'identifier rapidement les termes non reconnus, puisque que seuls les termes reconnus sont en couleur. Supposons par exemple que vous ayez mal orthographié la sous-commande FORMAT en FRMAT. Les sous-commandes sont en vert par défaut, mais le texte FRMAT n'apparaîtra pas en couleur, puisqu'il n'est pas reconnu.

## Ouverture et exécution d'un fichier de syntaxe

1. Pour ouvrir un fichier de syntaxe enregistré, à partir du menu, sélectionnez :  
**Fichier > Ouvrir > Syntaxe...**  
 Une boîte de dialogue standard d'ouverture de fichiers apparaît.
2. Sélectionnez un fichier de syntaxe. Si aucun fichier de syntaxe n'apparaît, assurez-vous que l'option **Syntaxe (\*.sps)** est sélectionnée en tant que type de fichier à afficher.
3. Cliquez sur **Ouvrir**.
4. Utilisez le menu Exécuter de la fenêtre de syntaxe pour exécuter les commandes.

Si les commandes s'appliquent à un fichier de données particulier, vous devez ouvrir ce dernier avant d'exécuter les commandes ou inclure une commande qui ouvre le fichier de données. Vous pouvez coller ce type de commande à partir des boîtes de dialogue permettant d'ouvrir les fichiers de données.

## Utilisation des points de rupture

Les points de rupture vous permettent d'interrompre l'exécution d'une syntaxe de commande à des points spécifiés à l'intérieur de la syntaxe et de poursuivre l'exécution lorsque vous êtes prêt. Ceci vous permet de visualiser les sorties ou les données à un moment intermédiaire du travail de syntaxe, ou d'exécuter la syntaxe de commande affichant les informations sur l'état actuel des données, telles que FREQUENCIES. Les points de rupture ne peuvent être réglés qu'au niveau d'une commande, et non sur les lignes spécifiques au sein d'une commande.

Pour insérer un point de rupture dans une commande :

1. Cliquez n'importe où dans la zone à gauche du texte associé à la commande.

Le point de rupture est représenté par un cercle rouge dans la zone à gauche du texte de la commande et sur la même ligne que le nom de la commande, peu importe où vous cliquez.

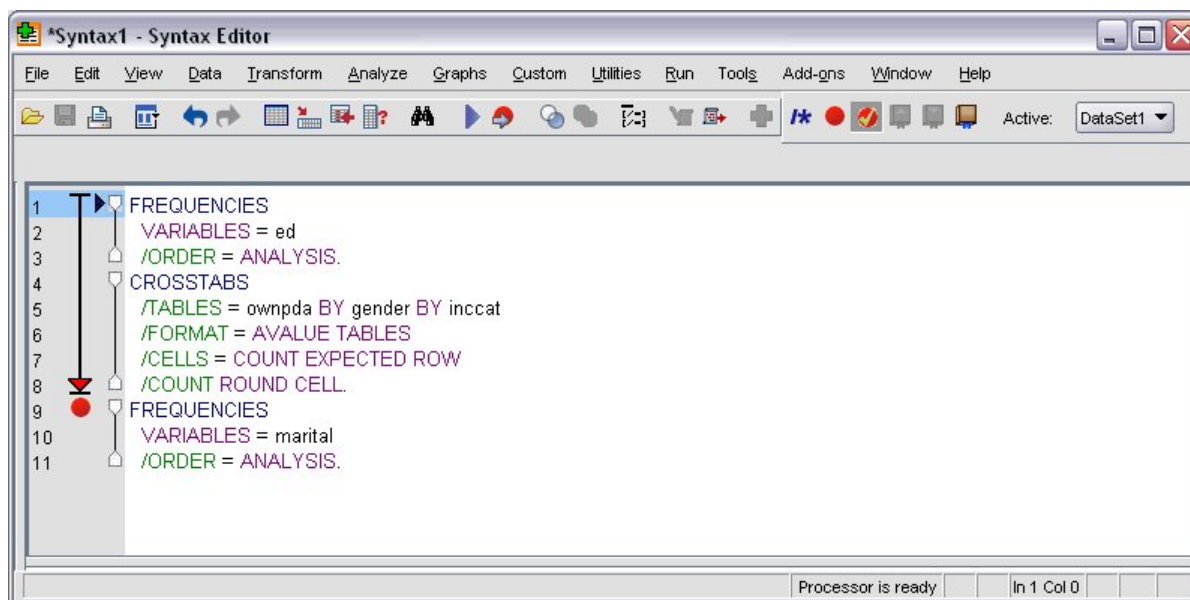


Figure 72. Exécution interrompue à un point de rupture

Lorsque vous exécutez une syntaxe de commande contenant des points de rupture, l'exécution s'interrompt avant chaque commande contenant un point de rupture.

La flèche pointant vers le bas à gauche du texte de commande présente la progression de l'exécution de la syntaxe. Elle couvre la zone s'étendant de la première exécution de commande à la dernière exécution de commande. Cela s'avère très pratique lorsque l'exécution d'une syntaxe de commande contient des points de rupture.

Pour reprendre l'exécution après un point de rupture :

2. Dans le menu de la fenêtre de l'Editeur de syntaxe, sélectionnez :

**Exécuter > Poursuivre**



---

## Chapitre 8. Modification des valeurs de données

L'organisation des données de départ ne répond pas systématiquement à vos besoins en matière de création de rapport ou d'analyse. Par exemple, vous pouvez :

- Créer une variable catégorielle à partir d'une variable d'échelle.
- Combiner plusieurs catégories de réponses en une seule catégorie.
- Créer une nouvelle variable qui représente la différence calculée entre deux variables existantes.
- Calculer le temps écoulé entre deux dates.

Dans ce chapitre, nous utilisons le fichier de données *demo.sav*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.

---

### Création d'une variable catégorielle à partir d'une variable d'échelle

Plusieurs variables catégorielles du fichier de données *demo.sav* sont dérivées de variables d'échelle de ce même fichier. Par exemple, la variable *inccat* est simplement *revenu*, réparti en quatre catégories. Cette variable catégorielle utilise les valeurs entières comprises entre 1 et 4 pour représenter les catégories de revenus suivantes (en milliers) : inférieur à 25 €, 25–49 €, 50–74 €, 75 € ou plus.

Pour créer la variable catégorielle *inccat* :

1. A partir des menus de la fenêtre de l'éditeur de données, sélectionnez :

**Transformer > Regroupement visuel...**

Dans la boîte de dialogue initiale Regroupement visuel, sélectionnez les variables d'échelle et/ou ordinales pour lesquelles vous souhaitez créer des variables regroupées. Le **regroupement** consiste à prendre plusieurs valeurs contiguës et à les regrouper dans une même catégorie.

Etant donné que le regroupement visuel se base sur les valeurs réelles du fichier de données pour vous proposer les meilleurs choix de regroupement, il doit tout d'abord lire ce fichier. Si le fichier de données contient un nombre d'observations important, cette opération peut prendre un certain temps. Par conséquent, la boîte de dialogue initiale vous permet également de limiter le nombre d'observations à lire (« analyser »). Cette limitation est inutile avec notre fichier de données exemple. Il contient plus de 6 000 observations, mais l'analyse d'un tel nombre d'observations est rapide.

2. Faites glisser *Revenu du foyer en milliers (revenu)* de la liste Variables vers la liste Variables à regrouper, puis cliquez sur **Poursuivre**.

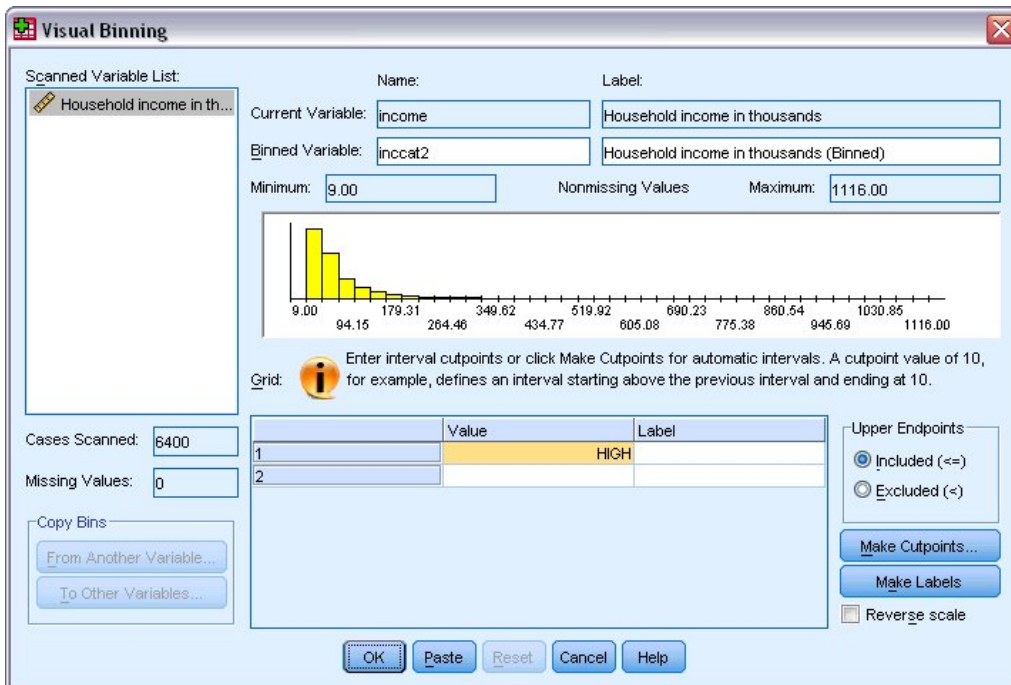


Figure 73. Boîte de dialogue principale Regroupement visuel

3. Dans la liste des variables analysées de la boîte de dialogue principale Regroupement visuel, sélectionnez *Revenu du foyer en milliers (revenu)*.

Un histogramme affiche la distribution de la variable sélectionnée (très asymétrique dans le cas qui nous occupe).

4. Affectez le nom *inccat2* à la nouvelle variable regroupée et sélectionnez le libellé de variable *Catégorie de revenu en milliers*.
5. Cliquez sur **Créer des divisions**.
6. Sélectionnez **Intervalle de longueur identique**.
7. Saisissez 25 pour l'emplacement de la première division, 3 pour le nombre de divisions et 25 pour la largeur.

Le nombre de catégories regroupées correspond au nombre de divisions, plus 1. Ainsi, dans cet exemple, la nouvelle variable regroupée comportera quatre catégories, les trois premières couvrant une plage de 25 (milliers) et la dernière toutes les valeurs supérieures à celle de la dernière division (75).

8. Cliquez sur **Appliquer**.

Les valeurs qui figurent à présent dans la grille représentent les divisions définies, à savoir les points finaux supérieurs de chaque catégorie. Par ailleurs, les lignes verticales de l'histogramme indiquent l'emplacement des divisions.

Par défaut, les valeurs de division sont incluses dans les catégories correspondantes. Par exemple, la première valeur (25) inclurait toutes les valeurs inférieures ou égales à 25. Dans cet exemple toutefois, nous voulons définir les catégories suivantes : inférieur à 25, 25–49, 50–74 et 75 ou plus.

9. Dans le groupe Points finaux supérieurs, sélectionnez **Exclus (<)**.
10. Cliquez ensuite sur **Créer des libellés**.

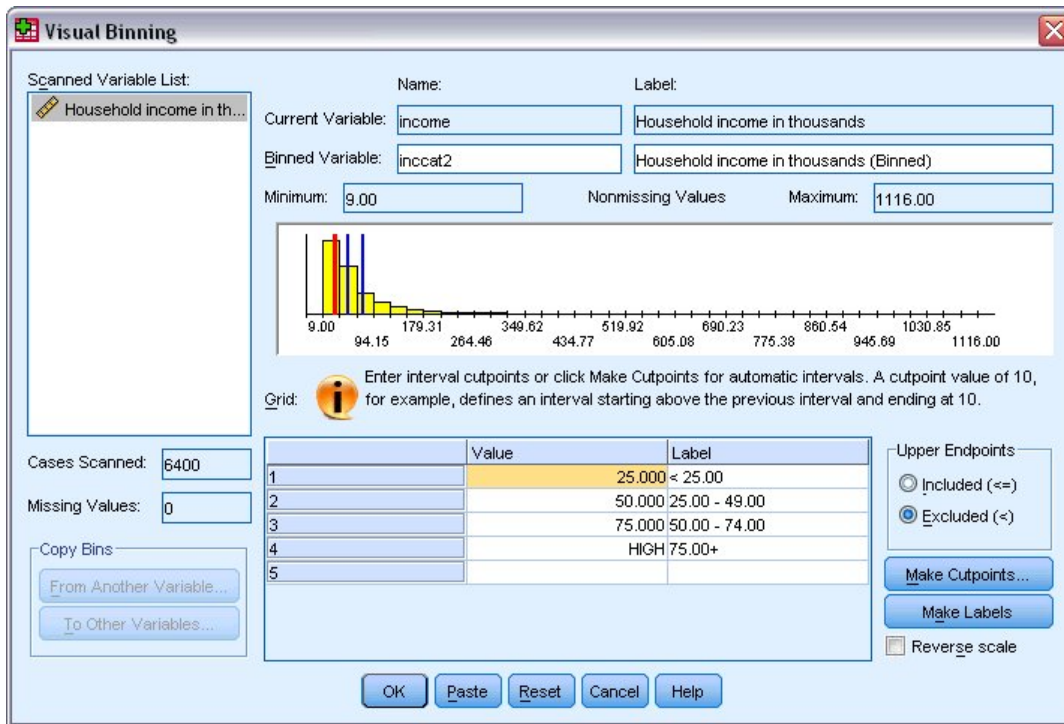


Figure 74. Libellés de valeur générés automatiquement

Cette opération permet de générer automatiquement des libellés de valeur descriptifs pour chaque catégorie. Étant donné que les valeurs réelles affectées à la nouvelle variable regroupée sont de simples entiers séquentiels (le premier étant 1), les libellés de valeur peuvent s'avérer très utiles.

Vous pouvez également procéder manuellement pour entrer ou modifier des divisions et des libellés dans la grille, pour modifier des emplacements de division en faisant glisser les lignes de division de l'histogramme, et pour supprimer des divisions en faisant glisser les lignes de division hors de l'histogramme.

11. Cliquez sur **OK** pour créer la variable regroupée.

La nouvelle variable est affichée dans l'éditeur de données. Étant donné que la variable est ajoutée à la fin du fichier, elle apparaît dans la colonne la plus à droite dans Vue de données et dans la dernière ligne de la vue de variable.

## Calcul de nouvelles variables

A l'aide d'une grande variété de fonctions mathématiques, vous pouvez calculer de nouvelles variables en fonction d'équations extrêmement complexes. Dans cet exemple, cependant, nous calculerons simplement une nouvelle variable qui représente la différence entre les valeurs de deux variables existantes.

Le fichier de données *demo.sav* contient une variable pour l'âge actuel du répondant et une variable pour le nombre d'années d'occupation du poste actuel. En revanche, il ne contient pas de variable indiquant l'âge du répondant au moment où il a été embauché à ce poste. Nous pouvons créer une nouvelle variable qui représenterait la différence calculée entre l'âge actuel et le nombre d'années d'occupation du poste actuel afin d'obtenir l'âge approximatif du répondant au moment de l'embauche.

1. A partir des menus de la fenêtre de l'éditeur de données, sélectionnez :

**Transformer > Calculer la variable...**

2. Pour Variable destination, saisissez `débutempl`.

3. Sélectionnez la variable *Age en années [âge]* dans la liste des variables source et cliquez sur le bouton fléché pour la copier dans la zone Expression numérique.

4. Cliquez sur le bouton moins (-) du clavier du pavé numérique de la boîte de dialogue (ou appuyez sur la touche moins du clavier).
5. Sélectionnez *Nb d'années avec l'employeur actuel [emploi]* et cliquez sur le bouton fléché pour la copier dans l'expression.

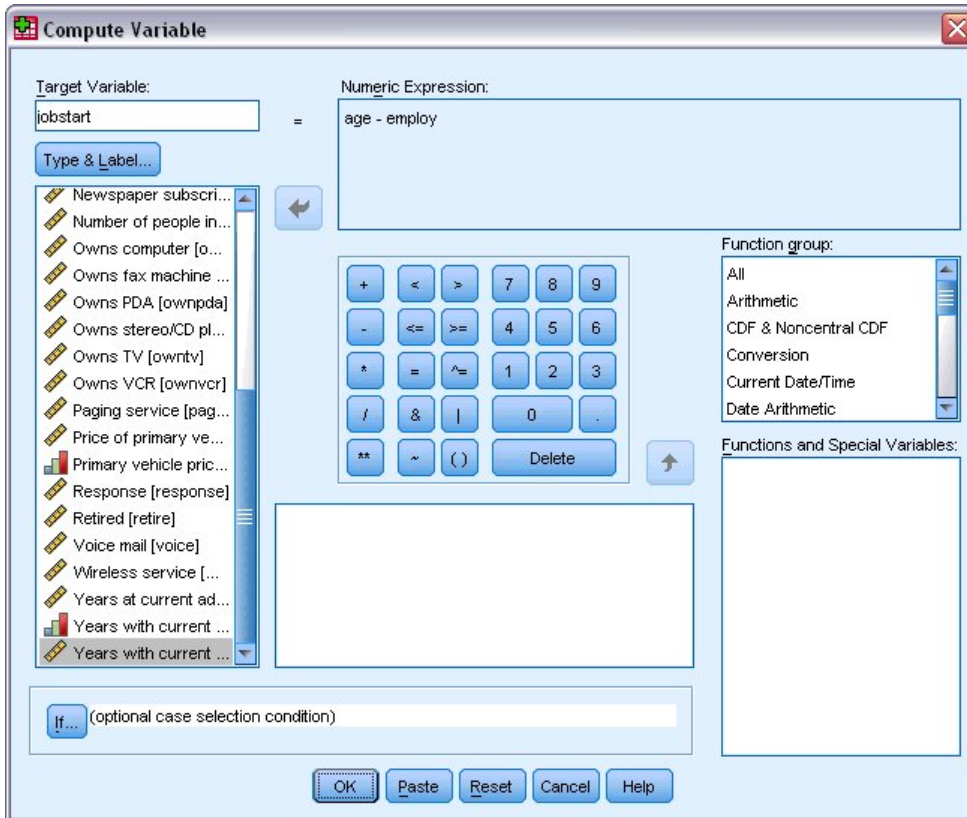


Figure 75. Boîte de dialogue Calculer la variable

*Remarque* : Veillez à sélectionner la variable d'emploi appropriée. Il existe également une version catégorielle recodée de la variable, mais ce n'est *pas* ce que vous recherchez. L'expression numérique doit être *age-employ* et non *age-empcat*.

6. Cliquez sur **OK** pour calculer la nouvelle variable.

La nouvelle variable est affichée dans l'éditeur de données. Etant donné que la variable est ajoutée à la fin du fichier, elle apparaît dans la colonne la plus à droite dans Vue de données et dans la dernière ligne de la vue de variable.

## Utilisation de fonctions dans les expressions

Vous pouvez également utiliser des fonctions prédéfinies dans les expressions. Vous disposez de plus de 70 fonctions intégrées, dont les suivantes :

- Fonctions arithmétiques
- Fonctions statistiques
- Fonctions de distribution
- Fonctions logiques
- Fonctions d'agrégation et d'extraction relatives à la date et à l'heure
- Fonctions de valeur manquante
- Fonctions d'observations croisées

- Fonctions sur chaînes

Les fonctions sont organisées en groupes logiques, tels qu'un groupe pour les opérations arithmétiques et un autre pour le calcul des métriques statistiques. Dans un souci de commodité, une série de variables système couramment utilisées, telles que la variable *\$TIME* (date et heure actuelles), figurent également dans des groupes de fonctions appropriés.

Collage d'une fonction dans une expression

Pour coller une fonction dans une expression :

1. Dans l'expression, positionnez le curseur là où vous souhaitez que la fonction apparaisse.
2. Sélectionnez le groupe approprié dans la liste Groupe de fonctions. Le groupe libellé **Tous** répertorie toutes les fonctions et variables système disponibles.
3. Double-cliquez sur la fonction dans la liste Fonctions et variables spéciales (ou sélectionnez-la, puis cliquez sur la flèche adjacente à la liste Groupe de fonctions).

La fonction est insérée dans l'expression. Si vous sélectionnez une partie de l'expression, puis insérez la fonction, la partie sélectionnée de l'expression est utilisée comme premier argument dans la fonction.

Modification d'une fonction dans une expression

La fonction collée n'est complète que lorsque vous avez saisi les arguments, symbolisés par des points d'interrogation. Le nombre de points d'interrogation indique le nombre minimum d'arguments requis pour compléter la fonction.

1. Sélectionnez les points d'interrogation dans la fonction collée.
2. Saisissez les arguments. Si les arguments sont des noms de variable, vous pouvez les coller à partir de la liste des variables.

## Utilisation d'expressions conditionnelles

Vous pouvez utiliser des expressions conditionnelles (également appelées expressions logiques) pour appliquer des transformations aux sous-groupes d'observations sélectionnés. Une expression conditionnelle renvoie la valeur vrai, faux ou manquant pour chaque observation. Si le résultat d'une expression conditionnelle est True (vrai), la transformation est appliquée à cette observation. Si le résultat est False (faux) ou Manquant, la transformation n'est pas appliquée à l'observation.

Pour indiquer une expression conditionnelle :

1. Cliquez sur **Si** dans la boîte de dialogue Calculer la variable. Cette opération ouvre la boîte de dialogue Si les observations.
2. Sélectionnez **Inclure lorsque l'observation remplit la condition** :
3. Entrez l'expression conditionnelle.

La plupart des expressions conditionnelles contiennent au moins un opérateur relationnel, comme dans :

`âge>=21`

ou

`revenu*3<100`

Dans le premier exemple, seules les observations dont la valeur *Age [âge]* est supérieure ou égale à 21 sont sélectionnées. Dans le deuxième exemple, la valeur *Revenu du foyer en milliers [revenu]* multipliée par 3 doit être inférieure à 100 pour qu'une observation soit sélectionnée.

Vous pouvez aussi lier au moins deux expressions conditionnelles à l'aide d'opérateurs logiques, comme dans :

```
âge>=21 | ed>=4
```

ou

```
revenu*3<100 & ed=5
```

Dans le premier exemple, les observations qui remplissent la condition *Age [âge]* ou *Niveau d'éducation [ed]* sont sélectionnées. Dans le deuxième exemple, les observations doivent remplir les conditions *Revenu du foyer en milliers [revenu]* et *Nombre d'années d'éducation [ed]* pour être sélectionnées.

---

## Utilisation des dates et heures

L'Assistant Date et heure vous permet de réaliser de nombreuses tâches courantes portant sur les dates et les heures. A l'aide de cet assistant, vous pouvez effectuer les opérations suivantes :

- Créer une variable date/heure à partir d'une variable de chaîne contenant une date ou une heure.
- Construire une variable date/heure en fusionnant des variables contenant différentes parties de la date ou de l'heure.
- Ajouter ou soustraire des valeurs aux variables date/heure, notamment ajouter ou soustraire deux variables date/heure.
- Extraire une partie d'une variable date ou heure, telle que le jour du mois d'une variable date/heure de format mm/jj/aaaa.

Les exemples de cette section utilisent le fichier de données *upgrade.sav*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.

Pour utiliser l'Assistant Date et heure :

1. A partir des menus, sélectionnez :

**Transformer > Assistant Date et heure...**



Figure 76. Ecran de présentation de l'Assistant Date et heure

L'écran de présentation de l'Assistant Date et heure met à votre disposition une série de tâches générales. Les tâches inapplicables aux données utilisées sont désactivées. Par exemple, le fichier de données *upgrade.sav* ne contient aucune variable de chaîne. Par conséquent, la tâche permettant de créer une variable date à partir d'une chaîne est désactivée.

Si vous ne savez pas comment manipuler les dates et les heures dans IBM SPSS Statistics, vous pouvez sélectionner l'option **Connaître la représentation des dates et heures**, puis cliquer sur le bouton **Suivant**. Cette opération fait apparaître un écran contenant une brève présentation des variables date/heure et un lien, par le biais du bouton Aide, vers des informations complémentaires.

## Calcul du temps écoulé entre deux dates

L'une des tâches les plus courantes portant sur les dates est le calcul du temps s'écoulant entre deux dates. A titre d'exemple, imaginez un éditeur de logiciels qui souhaite analyser les achats de licence de mise à niveau en déterminant le nombre d'années écoulées depuis le dernier achat d'une mise à niveau par chaque client. Le fichier de données *upgrade.sav* contient une variable pour la date du dernier achat d'une mise à niveau par client, mais pas pour le nombre d'années écoulées depuis cet achat. Une nouvelle variable représentant le temps écoulé, en années, entre la date de la dernière mise à niveau et la date de la version suivante du produit fournira une mesure de cette quantité.

Pour calculer le temps écoulé entre deux dates :

1. Sélectionnez l'option **Calculer avec des dates et heures** dans l'écran de présentation de l'Assistant Date et heure, puis cliquez sur le bouton **Suivant**.
2. Sélectionnez l'option **Calculez le nombre d'unités de temps séparant deux dates**, puis cliquez sur le bouton **Suivant**.

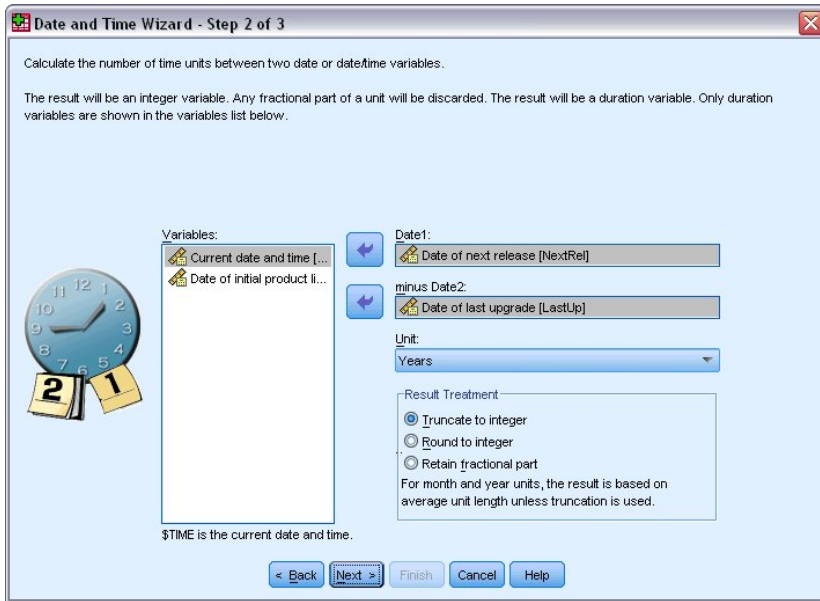


Figure 77. Calcul du temps écoulé entre deux dates : Etape 2

3. A l'étape 2, sélectionnez *Date de la prochaine version* pour la variable Date1.
4. Sélectionnez l'option *Date de la dernière mise à niveau* pour la variable Date2.
5. Sélectionnez **années** pour l'unité et **Tronquer à l'entier** pour le traitement du résultat. (Ce sont des sélections par défaut.)
6. Cliquez sur **Suivant**.
7. A l'étape 3, entrez *NbreAnsDepMAN* comme nom de la variable de résultat. Les variables de résultat ne peuvent pas porter le même nom qu'une variable existante.
8. Entrez *Nombre d'années écoulées depuis la dernière mise à niveau* comme libellé de la variable de résultat. Les libellés des variables de résultat sont facultatifs.
9. Conservez la sélection par défaut de l'option **Créer la variable maintenant**, puis cliquez sur le bouton **Terminer** pour créer la nouvelle variable.

La nouvelle variable *NbreAnsDepMAN*, affichée dans l'éditeur de données, est le nombre entier d'années entre les deux dates. Les parties fractionnelles d'une année ont été tronquées.

## Ajout d'une durée à une date

Vous pouvez ajouter ou soustraire des durées, telles que 10 jours ou 12 mois, à une date. A partir de l'exemple de l'éditeur de logiciels utilisé dans la section précédente, vous pouvez envisager de déterminer la date à laquelle expire le contrat d'assistance technique d'origine de chaque client. Le fichier de données *upgrade.sav* contient une variable pour le nombre d'années d'assistance sous contrat et une variable pour la date d'achat initiale. Vous pouvez donc déterminer la date d'expiration de l'assistance initiale en ajoutant le nombre d'années d'assistance à la date achat.

Pour ajouter une durée à une date :

1. Sélectionnez l'option **Calculer avec des dates et heures** dans l'écran de présentation de l'Assistant Date et heure, puis cliquez sur le bouton **Suivant**.
2. A l'étape 1, sélectionnez **Ajoutez une durée à une date ou soustrayez-la de la date** et cliquez sur **Suivant**.



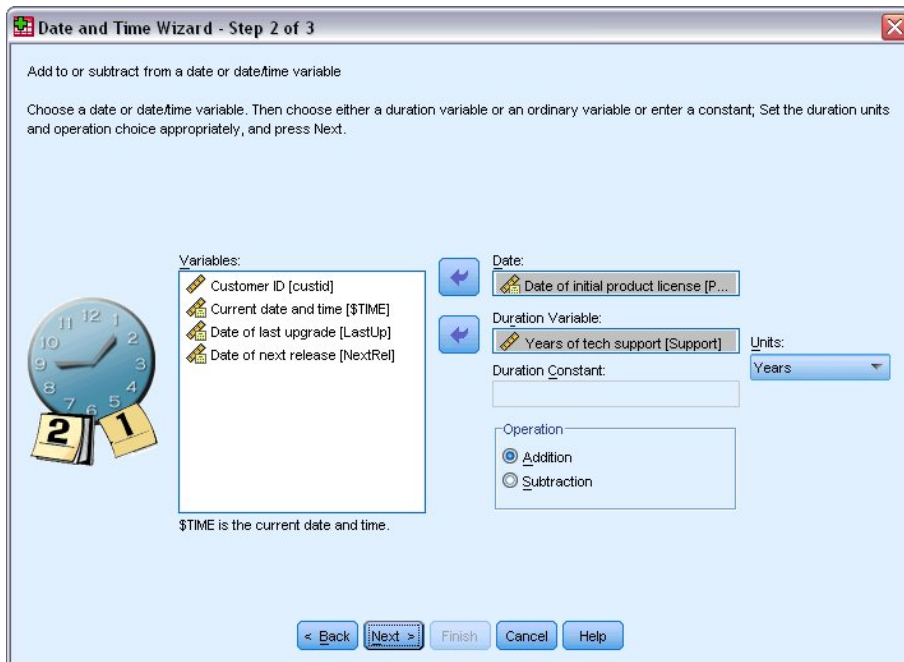


Figure 78. Ajout d'une durée à une date : Etape 2

3. Sélectionnez l'option *Date de la première licence produit* pour la variable Date.
4. A l'étape 2, sélectionnez *Années d'assistance technique* pour la variable de durée.  
Dans la mesure où la variable *Années d'assistance technique* est uniquement une variable numérique, vous devez indiquer l'unité dans laquelle cette variable doit être ajoutée en tant que durée.
5. Sélectionnez **Années** dans la liste déroulante Unités.
6. Cliquez sur **Suivant**.
7. A l'étape 3, entrez *DateExpAss* comme nom de la variable de résultat. Les variables de résultat ne peuvent pas porter le même nom qu'une variable existante.
8. Entrez *Date d'expiration de l'assistance* comme libellé de la variable résultat. Les libellés des variables de résultat sont facultatifs.
9. Cliquez sur le bouton **Terminer** pour créer la nouvelle variable.

La nouvelle variable est affichée dans l'éditeur de données.



---

## Chapitre 9. Tri et sélection de données

L'organisation des fichiers de données ne répond pas toujours idéalement à vos besoins. Pour préparer des données pour une analyse, vous avez le choix entre un grand nombre de transformations de fichiers. Vous pouvez effectuer les opérations suivantes :

- **Trier les données** : Vous pouvez trier les observations en fonction de la valeur de certaines variables.
- **Sélectionner des sous-ensembles d'observations** : Vous pouvez limiter votre analyse à un sous-ensemble d'observations ou effectuer des analyses simultanées sur différents sous-ensembles.

Les exemples de ce chapitre utilisent le fichier de données *demo.sav*. Pour plus d'informations, voir Chapitre 10, «Fichiers d'exemple», à la page 85.

---

### Tri des données

Il est souvent utile (et parfois nécessaire) de trier les observations (tri des lignes du fichier de données) pour certains types d'analyse.

Pour réorganiser la séquence des observations dans le fichier de données en fonction de la valeur de certaines variables de tri, procédez comme suit :

1. A partir des menus, sélectionnez :  
**Données > Trier les observations...**  
La boîte de dialogue Trier les observations est affichée.

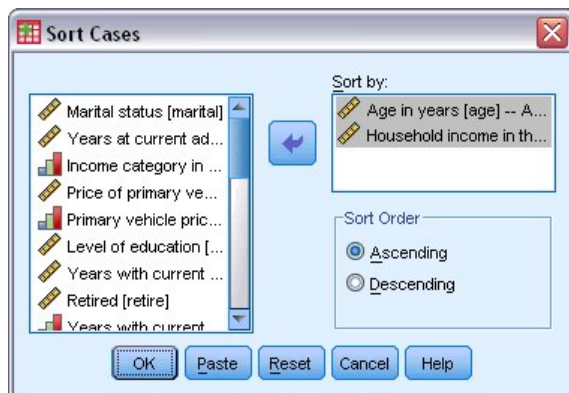


Figure 79. Boîte de dialogue Trier les observations

2. Ajoutez les variables *Age en années [âge]* et *Revenu du ménage en milliers [revenu]* à la liste Trier par.

Si vous sélectionnez plusieurs variables de tri, l'ordre dans lequel elles apparaissent dans la liste Trier par détermine celui dans lequel les observations sont triées. Dans cet exemple, basé sur les entrées de la liste Trier par, les observations seront triées en fonction de la valeur de la variable *Revenu du ménage en milliers [revenu]* dans les catégories de la variable *Age en années [âge]*. Pour les variables de chaîne, les majuscules précèdent leurs équivalents minuscules dans l'ordre de tri (par exemple, la variable de chaîne *Oui* précède la variable de chaîne *oui* dans l'ordre de tri).

## Traitement d'un fichier scindé

Pour scinder le fichier de données en groupes distincts pour l'analyse :

1. A partir des menus, sélectionnez :

**Données > Scinder un fichier...**

La boîte de dialogue Scinder un fichier est affichée.

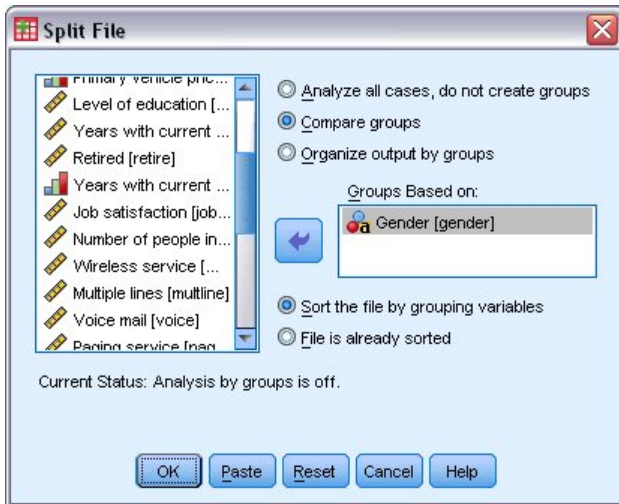


Figure 80. Boîte de dialogue Scinder un fichier

2. Sélectionnez **Comparer les groupes** ou **Séparer la sortie par groupes**. (Les exemples qui suivent ces étapes montrent les différences entre les deux options.)
3. Sélectionnez *Sexe [sexe]* afin de scinder le fichier en groupes distincts pour ces variables.

Vous pouvez utiliser des variables numériques, des variables de chaîne courte et des variables de chaîne longue comme variables de regroupement. Une analyse distincte est effectuée pour chaque sous-groupe défini par les variables de regroupement. Si vous sélectionnez plusieurs variables de regroupement, l'ordre dans lequel elles apparaissent dans la liste Critères de regroupement détermine la façon dont les observations sont regroupées.

Si vous sélectionnez **Comparer les groupes**, les résultats de tous les groupes du fichier scindé seront inclus dans le(s) même(s) tableau(x), comme dans le tableau des statistiques récapitulatives suivant généré par la procédure Fréquences.

Statistics			
Household income in thousands			
Female	N	Valid	3179
		Missing	0
	Mean		68.7798
	Median		44.0000
	Std. Deviation		75.73510
Male	N	Valid	3221
		Missing	0
	Mean		70.1608
	Median		45.0000
	Std. Deviation		81.56216

Figure 81. Sortie Scinder un fichier contenant un seul tableau croisé dynamique

Si vous sélectionnez **Séparer sorties par groupes** et que vous exécutez la procédure Effectifs, deux tableaux croisés dynamiques sont créés : un tableau pour les femmes et un pour les hommes.

**Statistics<sup>a</sup>**

Household income in thousands

N	Valid	3179
	Missing	0
Mean		68.7798
Median		44.0000
Std. Deviation		75.73510

a. Gender = Female

Figure 82. Sortie Scinder un fichier contenant un tableau croisé dynamique pour les femmes

**Statistics<sup>a</sup>**

Household income in thousands

N	Valid	3221
	Missing	0
Mean		70.1608
Median		45.0000
Std. Deviation		81.56216

a. Gender = Male

Figure 83. Sortie Scinder un fichier contenant un tableau croisé dynamique pour les personnes de sexe masculin

## Tri d'observations pour le traitement d'un fichier scindé

La procédure Scinder un fichier crée un nouveau sous-groupe chaque fois qu'elle rencontre une valeur différente pour l'une des variables de regroupement. Par conséquent, il est important de trier les observations en fonction des valeurs des variables de regroupement avant de lancer le traitement du fichier scindé.

Par défaut, la procédure Scinder un fichier trie automatiquement le fichier de données en fonction des valeurs des variables de regroupement. Si le fichier est déjà trié dans l'ordre correct, vous pouvez gagner du temps lors du traitement en sélectionnant **Le fichier est déjà trié**.

## Activation et désactivation du traitement d'un fichier scindé

Une fois le traitement d'un fichier scindé lancé, il reste actif pour le reste de la session, à moins que vous ne le désactiviez.

- **Analyser toutes les observations** : Cette option désactive le traitement du fichier scindé.
- **Comparer les groupes et Séparer la sortie par groupes** : Cette option active le traitement du fichier scindé.

Si le traitement du fichier scindé est activé, le message **Séparer fichier actif** apparaît sur la barre de statut située au bas de la fenêtre de l'application.

---

## Sélection de sous-groupes d'observations

Vous pouvez limiter votre analyse à un sous-groupe donné en fonction de critères contenant des variables et des expressions complexes. Vous pouvez également sélectionner un échantillon aléatoire d'observations. Les critères utilisés pour définir un sous-groupe comprennent :

- Plages et valeurs de variables
- Plages de dates et d'heures
- Nombres d'observations (lignes)

- Expressions arithmétiques
- Expressions logiques
- Fonctions

Pour sélectionner un sous-ensemble d'observations à analyser, procédez comme suit :

1. A partir des menus, sélectionnez :

**Données > Sélectionner des observations...**

Cette opération ouvre la boîte de dialogue Sélectionner des observations.

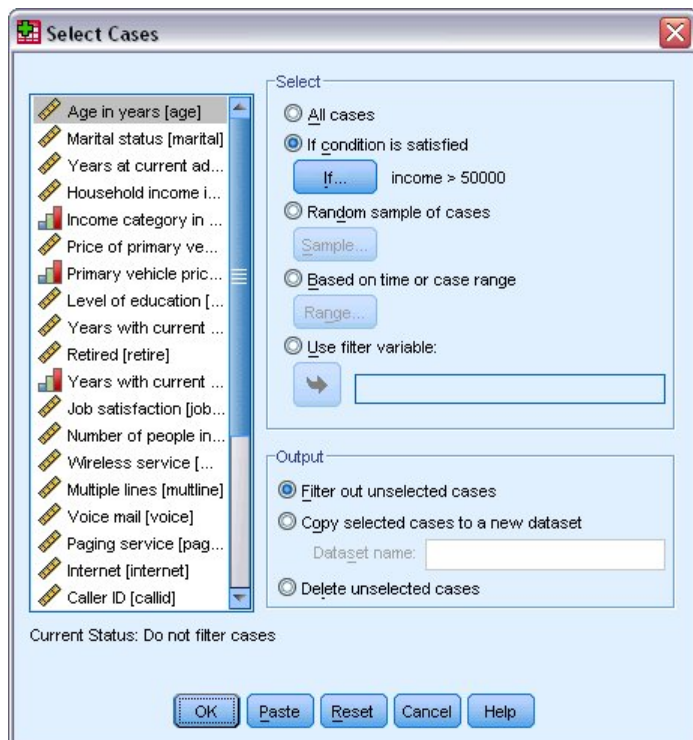


Figure 84. Boîte de dialogue Sélectionner des observations

## Sélection des sous-ensembles d'observations à l'aide d'une expression conditionnelle

Pour sélectionner des observations sur la base d'une expression conditionnelle, procédez comme suit :

1. Sélectionnez **Selon une condition logique**, puis cliquez sur **Si** dans la boîte de dialogue Sélectionner des observations.

Cette opération ouvre la boîte de dialogue Sélectionner des observations - Si.

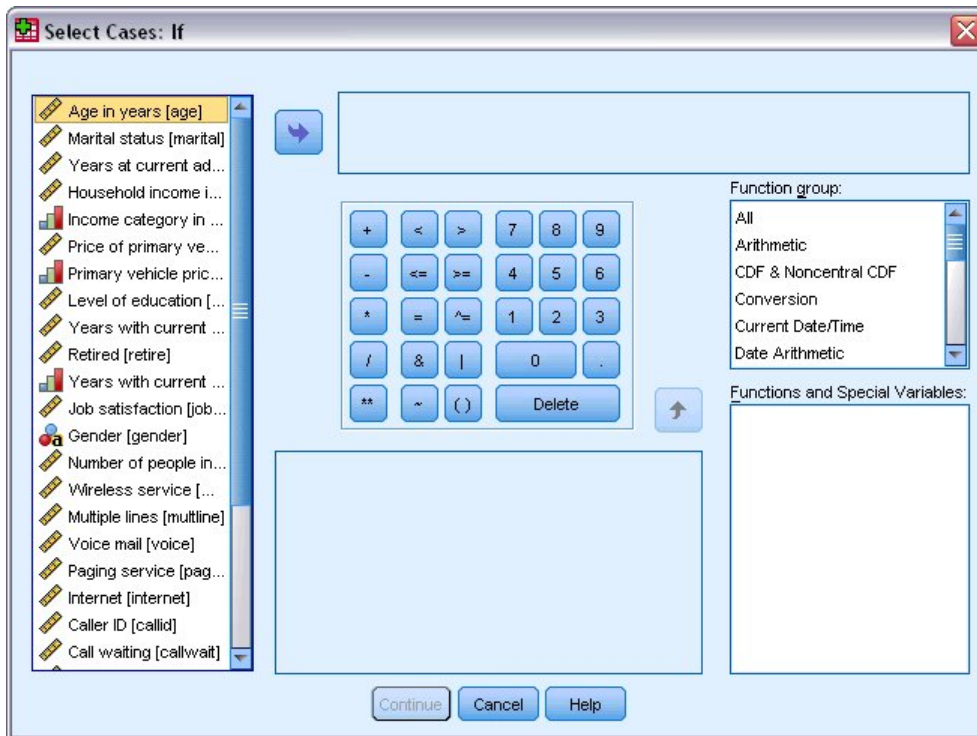


Figure 85. Boîte de dialogue Sélectionner observations : Si

L'expression conditionnelle peut utiliser des noms de variables existantes, des constantes, des opérateurs arithmétiques, des opérateurs logiques, des opérateurs relationnels et des fonctions. Vous pouvez saisir et modifier l'expression dans la zone de texte comme n'importe quel texte dans une fenêtre de sortie. Vous pouvez également utiliser le pavé numérique de la boîte de dialogue, la liste des variables et la liste des fonctions pour coller des éléments dans l'expression. Pour plus d'informations, voir «Utilisation d'expressions conditionnelles», à la page 71.

## Sélection d'un échantillon aléatoire

Pour obtenir un échantillon aléatoire, procédez comme suit :

1. Sélectionnez **Par échantillonnage aléatoire** dans la boîte de dialogue Sélectionner des observations.
2. Cliquez sur **Echantillon**.

Cette opération ouvre la boîte de dialogue Sélectionner observations : Echantillon aléatoire.

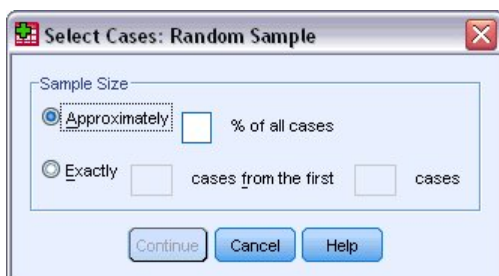


Figure 86. Boîte de dialogue Sélectionner observations : Echantillon aléatoire

Pour la taille de l'échantillon, deux options sont disponibles :

- **Environ** : Pourcentage défini par l'utilisateur. Cette option génère un échantillon aléatoire d'observations dont le nombre correspond approximativement au pourcentage indiqué.

- **Exactement** : Nombre d'observations spécifié par l'utilisateur. Vous devez également indiquer le nombre d'observations à partir duquel l'échantillon sera généré. Ce deuxième nombre doit être inférieur ou égal au nombre total d'observations dans le fichier de données. Si ce nombre dépasse le nombre total d'observations dans le fichier de données, l'échantillon contiendra proportionnellement moins d'observations que le nombre demandé.

## Sélection d'une plage de temps ou d'observations

Pour sélectionner une plage d'observations sur la base de dates, d'heures ou de numéros (lignes) d'observations, procédez comme suit :

1. Sélectionnez **Dans une plage de temps ou d'observations**, puis cliquez sur **Plage** dans la boîte de dialogue Sélectionner des observations.

Cette opération ouvre la boîte de dialogue Sélectionner observations : Plage, dans laquelle vous pouvez sélectionner une plage de numéros (lignes) d'observations.

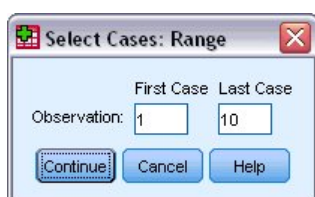


Figure 87. Boîtes de dialogue Sélectionner observations : Plage

- **Première observation** : Entrez la date de début et/ou les valeurs de temps de la plage. Si aucune variable de date n'est définie, entrez le numéro d'observation de départ (numéro de ligne dans l'éditeur de données, sauf si l'option Scinder un fichier est activée). Si vous n'indiquez aucune valeur pour la zone Dernière, toutes les observations à partir de la date/l'heure de début jusqu'à la fin de la série chronologique sont sélectionnées.
- **Dernière observation** : Entrez la date de fin et/ou les valeurs de temps pour la plage. Si aucune variable de date n'est définie, entrez le numéro d'observation de fin (numéro de ligne dans l'éditeur de données, sauf si l'option Scinder un fichier est activée). Si vous n'indiquez aucune valeur pour la zone Première, toutes les observations à partir du début de la série chronologique jusqu'à la date/l'heure de fin sont sélectionnées.

Pour les données de séries chronologiques comportant des variables de date définies, vous pouvez sélectionner une plage de dates et/ou d'heures sur la base des variables de date définies. Chaque observation représente des constatations effectuées à différents moments et le fichier est trié dans l'ordre chronologique.



Figure 88. Boîte de dialogue Sélectionner observations : Plage

Pour générer des variables de date pour des données de séries chronologiques, procédez comme suit :

2. A partir des menus, sélectionnez :  
**Données > Définir des dates...**



## Traitement des observations exclues

Vous pouvez choisir l'une des options suivantes pour traiter les observations exclues :

- **Filtrez les observations exclues** : Les observations exclues ne sont pas incluses dans l'analyse, mais restent dans le jeu de données. Vous pouvez utiliser les observations exclues ultérieurement dans la session si vous désactivez le filtrage. Si vous sélectionnez un échantillon aléatoire ou si vous sélectionnez des observations sur la base d'une expression conditionnelle, une variable nommée *filter\_\$* est générée ; elle comporte la valeur 1 pour les observations sélectionnées et la valeur 0 pour les observations exclues.
- **Copiez les observations sélectionnées dans le jeu de données** : Les observations sélectionnées sont copiées dans un nouveau jeu de données ; le jeu de données d'origine reste inchangé. Les observations exclues ne sont pas incluses dans le nouveau jeu de données et sont conservées dans leur état d'origine dans le jeu de données d'origine.
- **Supprimez les observations exclues** : Les observations exclues sont supprimées du jeu de données. Vous pouvez récupérer les observations supprimées uniquement en fermant le fichier sans enregistrer les modifications et en l'ouvrant à nouveau. La suppression des observations est définitive si vous enregistrez les modifications apportées au fichier de données.

*Remarque* : Si vous supprimez des observations exclues et enregistrez le fichier, vous ne pouvez pas récupérer ces observations.

---

## Statut de la sélection d'une observation

Si vous avez sélectionné un sous-ensemble d'observations mais n'avez pas écarté les observations non sélectionnées, celles-ci sont identifiées dans l'éditeur de données par une ligne verticale dans le numéro de ligne.

1
2
3
4
5
6
7
8
9
10
11

Figure 89. Statut de la sélection d'une observation



---

## Chapitre 10. 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 chercheurs<sup>1</sup> 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 classique<sup>2</sup>, 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 <sup>3</sup>, 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 <sup>4</sup> 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\_uprepared.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 <sup>5</sup>, 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 frappé<sup>6</sup>. 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.
- **diabetes\_costs.sav** : Il s'agit d'un fichier de données d'hypothèse contenant des informations conservées par une compagnie d'assurance sur les assurés qui ont du diabète. Chaque observation correspond à un assuré différent.
- **dietstudy.sav** : Ce fichier de données d'hypothèse contient les résultats d'une étude portant sur le régime de Stillman<sup>7</sup>. 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.
- **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.
- **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 Databases<sup>8</sup> 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** : Bell<sup>9</sup> a présenté un tableau pour illustrer les groupes sociaux possibles. Guttman<sup>10</sup> 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

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/~mlern/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.

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 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 <sup>11</sup> 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 Etats-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/](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/). Accès en 2003.

---

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.

- **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<sup>12, 13</sup>, ont décelé parmi ces variables des non-linéarités qui pénalisent les approches standard de la régression.
- **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

---

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.



- 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.<sup>14</sup>
  - **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 ailleurs<sup>15</sup> 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 Validation 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.

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.

- **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.
- **tcm\_kpi.sav**. Il s'agit d'un fichier de données d'hypothèse contenant les indicateurs de performance clés hebdomadaires d'une activité commerciale. Il contient aussi les données hebdomadaires d'un certain nombre d'indicateurs contrôlables sur la même période.
- **tcm\_kpi\_upd.sav** : Ce fichier de données est identique au fichier *tcm\_kpi.sav* mais contient les données relatives à quatre semaines supplémentaires.
- **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 <sup>16</sup>.

---

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

- **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 <sup>17</sup>.
- **verd1985.sav** : Ce fichier de données concerne une enquête <sup>18</sup>. 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.
- **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 <sup>19</sup>. 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.

---

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.

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 concerne des produits et des services disponibles dans différents pays. Il peut être mis à disposition par IBM dans d'autres langues. Toutefois, il peut être nécessaire de posséder une copie du produit ou de la version du produit dans cette langue pour pouvoir y accéder.

Le présent document peut contenir des informations ou des références concernant certains produits, logiciels ou services IBM non annoncés dans ce pays. 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, MD-NC119  
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.  
19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokyo 103-8510, Japan*

LE PRESENT DOCUMENT EST LIVRE EN L'ETAT SANS AUCUNE GARANTIE EXPLICITE OU IMPLICITE. IBM DECLINE NOTAMMENT TOUTE RESPONSABILITE RELATIVE A CES INFORMATIONS EN CAS DE CONTREFAÇON AINSI QU'EN CAS DE DEFAUT D'APTITUDE A L'EXECUTION D'UN TRAVAIL DONNE. Certaines juridictions n'autorisent pas l'exclusion des garanties implicites, auquel cas l'exclusion ci-dessus ne vous sera pas applicable.

Le présent document peut contenir des inexactitudes ou des coquilles. Ce document est mis à jour périodiquement. Chaque nouvelle édition inclut les mises à jour. IBM peut, à tout moment et sans préavis, modifier les produits et 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 Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
U.S.A*

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

Le logiciel sous licence décrit dans ce document et tous les éléments sous licence disponibles s'y rapportant sont fournis par IBM conformément aux dispositions 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 performances et les exemples de clients sont fournis à titre d'exemple uniquement. Les performances réelles peuvent varier en fonction des configurations et des conditions d'exploitation spécifiques.

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

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

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

#### LICENCE DE COPYRIGHT :

Le présent logiciel contient des exemples de programmes d'application en langage source destinés à illustrer les techniques de programmation sur différentes plateformes d'exploitation. Vous avez le droit de copier, de modifier et de distribuer ces 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](http://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](http://www.ibm.com/legal/copytrade.shtml).

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

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

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

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

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

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





---

# Index

## A

Access (Microsoft) 11  
assistant d'importation de texte 14  
Assistant Date et heure 72  
Assistant de base de données 11

## C

Calcul de nouvelles variables 69  
Collage d'une syntaxe  
Boîte de dialogue 63  
couches  
Création dans des tableaux croisés dynamiques 42  
création de libellés de variable 21

## D

déplacement  
Éléments dans des tableaux croisés dynamiques 41  
éléments dans le visualiseur 39  
données catégorielles  
Mesures récapitulatives 27  
Données catégorielles 27  
données chaîne  
Saisie de données 20  
Données continues 27  
Données d'échelle 27  
Données d'intervalle 27  
Données de type rapport 27  
Données nominales 27  
Données numériques 19  
Données ordinales 27  
Données qualitatives 27  
Données quantitatives 27

## E

Editeur de données  
Saisie de données non numériques 20  
Saisie de données numériques 19  
Edition de tableaux croisés dynamiques 43  
effectifs  
Tableaux d'effectifs 27  
Excel (Microsoft)  
exportation des résultats 53  
exportation des résultats  
dans Excel 53  
dans PowerPoint 53  
dans Word 53  
HTML 61  
Expressions conditionnelles 71

## F

Fenêtres de syntaxe  
codification par couleur 64  
Collage de commandes 63  
Exécution de commandes 63, 65  
Modification de commandes 64  
points de rupture 65  
saisie semi-automatique 64  
fichiers d'exemple  
emplacement 85  
Fichiers de base de données  
Lecture 11  
fichiers de données texte  
lecture 14  
fichiers de feuille de calcul  
Lecture 8  
lecture des noms de variable 8  
Fichiers de syntaxe  
Ouverture 65  
fichiers Excel  
lecture 8  
fonctions dans les expressions 70

## G

graphiques  
Bâtons 28, 33  
création de graphiques 33  
Histogrammes 30  
Graphiques à barres 28

## H

Histogrammes 30  
HTML  
exportation des résultats 61

## L

libellés de valeurs  
affectation 23  
contrôle de l'affichage dans le visualiseur 23  
Variables numériques 23  
libellés de variable  
Création 21

## M

masquage des lignes et colonnes dans des tableaux croisés dynamiques 44  
mesures récapitulatives  
Données catégorielles 27  
Variables d'échelle 29

## N

niveau de mesure 27  
Niveau de mesure 27

## O

observations  
Sélection 79  
Tri 77, 79  
Outil Aide syntaxe 64

## P

PowerPoint (Microsoft)  
exportation des résultats 53

## R

Recodage de valeurs 67

## S

Saisie de données 19, 20  
Non numérique 20  
numérique 19  
Sélection d'observations 79  
Sous-groupes d'observations  
Dates et heures 82  
Echantillon aléatoire 81  
Expressions conditionnelles 80  
Filtrage des observations exclues 83  
Sélection 79  
Selon une condition logique 80  
Suppression d'observations  
exclues 83  
Syntaxe 63

## T

table de fréquences 27  
tableaux croisés dynamiques  
Accès aux définitions 40  
couches 42  
Formatage 43  
Formats de cellule 44  
Masquage de lignes et de colonnes 44  
Masquage du séparateur décimal 44  
modification 43  
Structure pivotante 41  
Transposition de lignes et de colonnes 41  
Types de données des cellules 44  
traitement d'un fichier scindé 78  
Transposition de lignes et de colonnes dans des tableaux croisés dynamiques 41  
Tri d'observations 77

Types de données  
Pour les variables 22

## V

valeurs manquantes  
manquantes par défaut 23  
Variables non numériques 24  
Variables numériques 24  
Valeurs système manquantes 23  
variables 19  
libellés 21  
Types de données 22  
variables d'échelle  
Mesures récapitulatives 29  
Variables de date et d'heure 72  
visualiseur  
Déplacement de la sortie 39  
Masquage et affichage de la sortie 39

## W

Word (Microsoft)  
exportation des résultats 53



