

Introduction à l'analyse exploratoire des données avec SPSS

A. Découverte de SPSS

1. Qu'est-ce que SPSS

SPSS signifie « Statistical Package for the Social Sciences ». Son objectif est d'offrir un logiciel permettant de réaliser la totalité des analyses statistiques habituellement utilisées en sciences humaines. C'est un logiciel très complet et dans ce cours, nous ne verrons qu'une très faible partie de ses possibilités. Il existe bien d'autres logiciels comme S-Plus, R ou SAS qui permettent d'atteindre les mêmes buts, c'est-à-dire faire des analyses statistiques.

2. Comment démarrer SPSS

Aller sous « **Démarrer** » (en bas à gauche de l'écran), puis sous « **Programme** », puis sous « **SPSS for Windows** » et finalement sous « **SPSS 13.0 for Windows** » (cf. Figure 1).

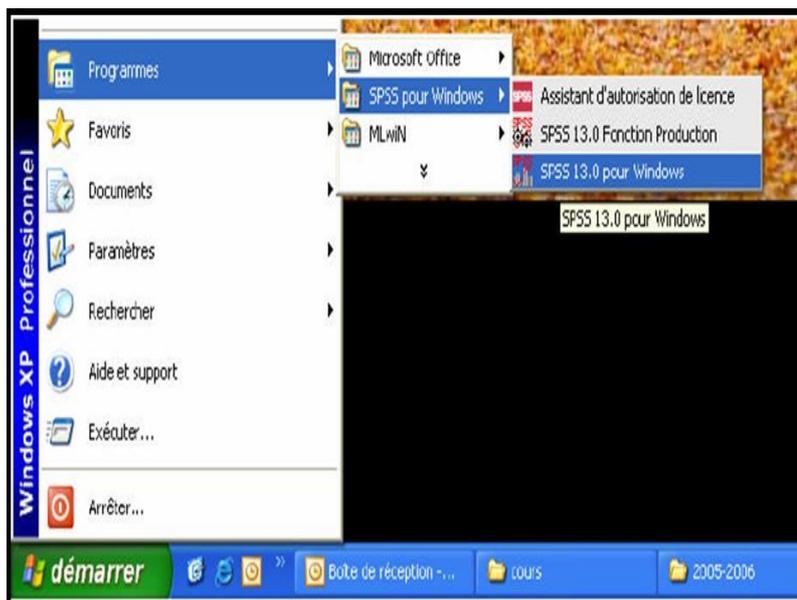


Figure 1

Lorsque SPSS démarre, il ouvre une fenêtre principale qui ressemble un peu à une feuille Excel. Il ouvre devant cette grille, une boîte de dialogue (Figure 2) permettant : d'ouvrir un jeu de données SPSS ou dans un autre format, de créer un nouveau jeu de données, etc. On peut soit utiliser cette boîte de dialogue, soit juste cliquer sur « **Cancel** », car nous pouvons faire exactement les mêmes manipulations directement à l'intérieur de SPSS.

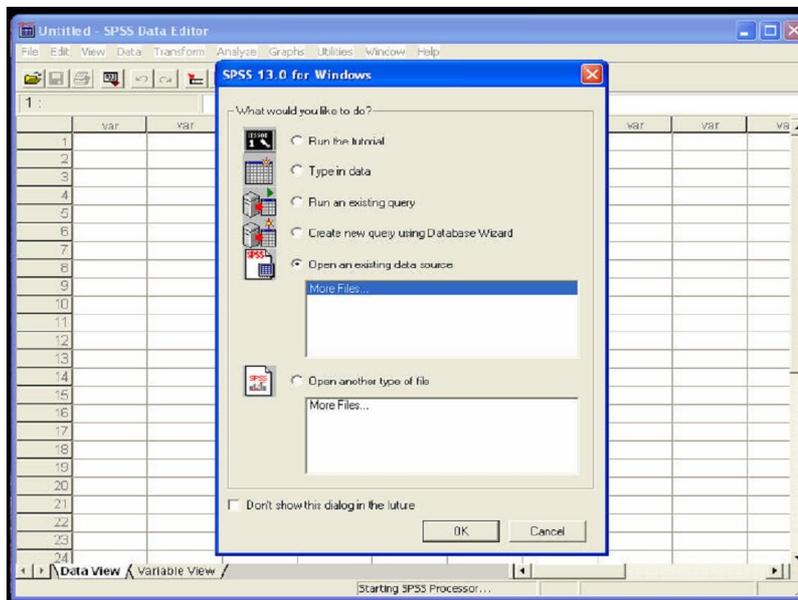


Figure 2

3. L'éditeur de données

Une fois cette boîte de dialogue fermée, nous voyons l'éditeur de donnée. L'éditeur de données contient la grille (matrice) de données (vue des données = « **Data View** ») et les descriptions des variables (vue des variables = « **Variable View** »).

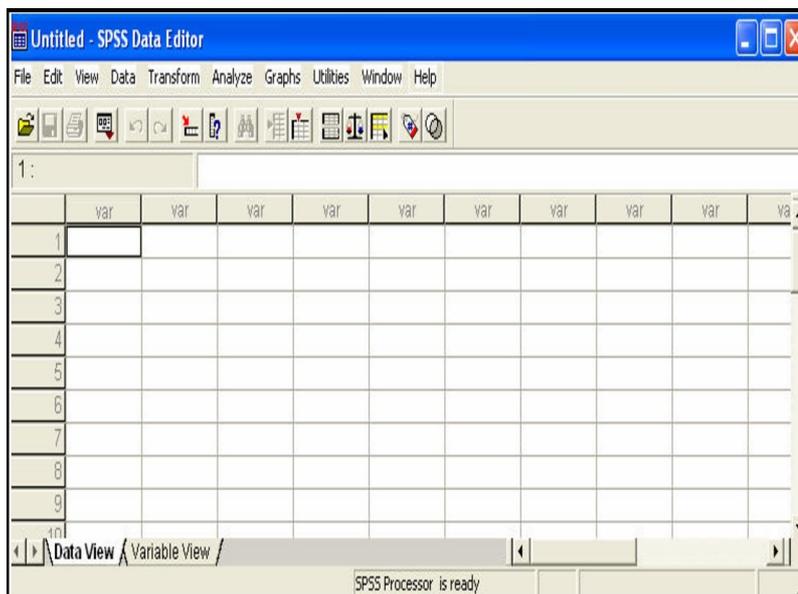


Figure 3

Dans la partie supérieure de l'éditeur de données nous avons, comme dans Word et Excel, des menus déroulants :



Figure 4

File:... **Edit:...** **View:...** **Data:...** **Transform:...** **Analyze:...** **Graphs:...** **Utilities:...** **Window:...** **Help:...**

4. Menu File : Ouvrir un fichier de données avec SPSS

Ouvrir un fichier de données avec SPSS

Pour ouvrir un fichier de données SPSS (extension **.sav**), nous allons utiliser les menus déroulant. Cliquer sur « **File** », puis sur « **Open** » et finalement sur « **Data...** ». Là, vous pouvez chercher votre fichier dans le répertoire où vous l'avez enregistré.

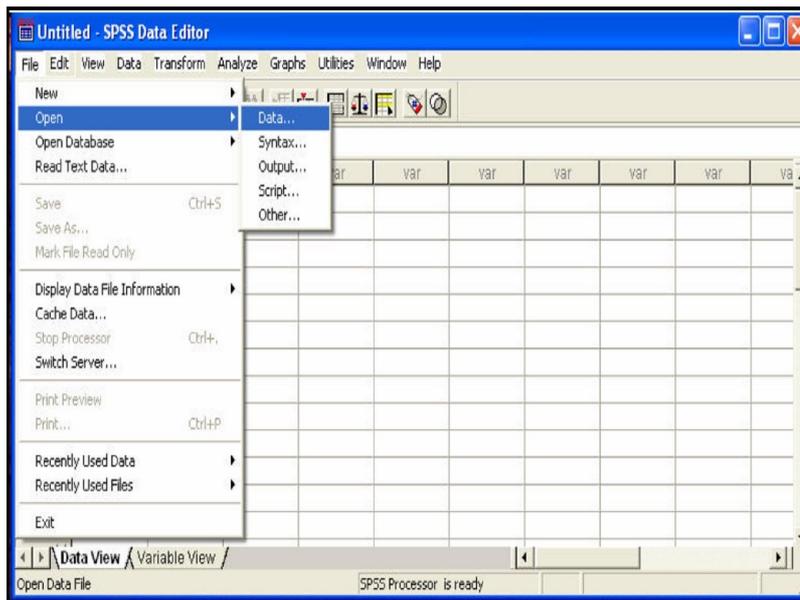


Figure 5

Une fois nos données ouvertes, nous pouvons explorer les deux affichages mentionnés en dessus : l'affichage des données (« **Data View** ») et l'affichage des variables (« **Variable View** ») :

Affichage des données :

- Chaque ligne représente un cas, par exemple un sujet (*case*)
- Chaque colonne représente une variable (*variable*)
- Chaque cellule contient une valeur d'un cas sur une variable

The screenshot shows the SPSS Data Editor window with the 'Data View' tab active. The window title is 'data.sav - SPSS Data Editor'. The table displayed has 13 rows and 11 columns. The first column is labeled 'id' and the subsequent columns are labeled 'exp01' through 'exp10'. The data is as follows:

	id	exp01	exp02	exp03	exp04	exp05	exp06	exp07	exp08	exp09
1	1	3	3	1	3	1	1	2	3	
2	2	3	4	2	3	2	2	4	4	
3	3	3	3	2	4	2	2	3	3	
4	4	3	4	2	4	2	4	4	4	
5	5	4	4	3	4	2	4	3	4	
6	6	4	4	1	2	2	2	3	4	
7	7	3	4	2	1	1	2	2	4	
8	8	3	4	2	2	1	2	2	3	
9	9	3	4	2	4	1	2	4	4	
10	10	4	3	2	2	2	3	3	3	
11	11	3	4	2	2	2	3	4	4	
12	12	3	3	3	3	2	3	4	3	
13	13	3	3	2	2	2	2	3	3	

Figure 6

Vue des variables :

- Chaque ligne représente une variable
- Les colonnes décrivent les caractéristiques des variables

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	3	0	Numéro d'ident	None	None	8	Right	Scale
2	exp01	Numeric	1	0	Affection	{1, jamais ou presque	None	8	Right	Scale
3	exp02	Numeric	1	0	Joie	{1, jamais ou presque	None	8	Right	Scale
4	exp03	Numeric	1	0	Peur	{1, jamais ou presque	None	8	Right	Scale
5	exp04	Numeric	1	0	Colère	{1, jamais ou presque	None	8	Right	Scale
6	exp05	Numeric	1	0	Honte	{1, jamais ou presque	None	8	Right	Scale
7	exp06	Numeric	1	0	Tristesse	{1, jamais ou presque	None	8	Right	Scale
8	exp07	Numeric	1	0	Amour	{1, jamais ou presque	None	8	Right	Scale
9	exp08	Numeric	1	0	Bonheur	{1, jamais ou presque	None	8	Right	Scale
10	exp09	Numeric	1	0	Inquiétude	{1, jamais ou presque	None	8	Right	Scale
11	exp10	Numeric	1	0	Irritation	{1, jamais ou presque	None	8	Right	Scale
12	exp11	Numeric	1	0	Culpabilité	{1, jamais ou presque	None	8	Right	Scale
13	exp12	Numeric	1	0	Dépression	{1, jamais ou presque	None	8	Right	Scale
14	bf01	Numeric	1	0	amica(e)	{1, pas du tout}...	None	8	Right	Scale

Figure 7

Name: nom de la variable

- le nom de la variable doit être unique
- max 8 caractères (à partir de la version SPSS 12, on peut utiliser plus de caractères : max. 64 !)
- débutant par une lettre
- les autres caractères peuvent être une lettre, un chiffre, un point ou les symbols @ _ # \$ (tous les autres symbols sont interdits)
- Le nom ne peut pas se terminer par un point
- les espaces vides ne sont pas possibles
- majuscules et minuscules ne sont pas différenciés dans l'appellation d'une variable ou dans celui d'un fichier
- les mots clé de SPSS ne peuvent pas être utilisés (all, and, by, eq, ge, gt, le, lt, ne, not, or, to, with)
- ne pas mettre de lettre accentuée même si c'est possible

Type : nature de la variable (numérique, date, dollar, etc.)

- par défaut, SPSS considérera la variable de type numérique;
- peut être changé dans la boîte de dialogue qui s'ouvre en appuyant sur  (symbole qui apparaît quand on sélectionne la cellule)

Width : nombre de chiffres accordés à la donnée (décimaux inclus)

Decimals : nombre de décimaux



Attention : **Width** et **Decimals** définissent le format des données dans la grille de données - la donnée apparaissant dans une cellule peut différer de la donnée exacte (qui est utilisée par SPSS pour les opérations de calcul et qui est visible dans la cellule blanche au-dessous la barre d'outils pour la cellule « active »)

Label : étiquette ou description de la variable

- max. 256 caractères

Values : valeurs définies et leur description

- p.ex. 1 = Femme, 2 = Homme
- peuvent être définis dans la boîte de dialogue qui s'ouvre en cliquant sur



- ce qui apparaît quand on choisit la cellule

Missing : attribution de certaines valeurs comme codes pour valeurs manquantes



Attention : « **none** » ne signifie pas qu'il n'y a pas de valeurs manquantes dans les données, ça signifie qu'il n'y a pas de valeurs manquantes définies par l'utilisateur. Par contre, il est possible d'avoir des valeurs **system missing** (correspond aux cellules vides dans la grille de données).

Columns: largeur des colonnes dans la vue de données (en augmentant/diminuant le défaut « 8 », plus/moins de caractères de la colonne seront visibles dans la vue de données)

Align : Alignement des valeurs des variables dans les cellules de la grille de données (à droite, à gauche, centrées)

Measure : Description de l'échelle de mesure (continue (**scale**), **ordinal** ou **nominal**)



Attention : n'a pas d'influence sur le type d'analyse qu'on peut effectuer avec SPSS !!!

Obtenir les informations sur les variables dans un format imprimable

Pour obtenir toutes les informations sur les variables dans un format imprimable, il faut cliquer sur **File | Display data file information | Working file**. Les informations sont transférées dans une nouvelle fenêtre (**SPSS Viewer**). Cette page des résultats peut être sauvegardée et/ou imprimée. Les éléments peuvent être copiés et collés dans un document géré par un autre logiciel (p.ex. Word). Le fichier des résultats (Output Editor) porte l'extension « **.spo** ».

Ouvrir et sauvegarder des fichiers de données

Dans le menu « **File** », on trouve toutes les commandes pour gérer les fichiers SPSS (ouvrir, sauvegarder, sauvegarder sous...) - pareil aux autres logiciels Windows (p.ex. Word)

5. Informations compactes sur les variables dans un fichier de données

Menu Utilities : Dans le menu **Utilities | Variables** on peut consulter les caractéristiques de toutes les variables. Quand vous travaillez avec un fichier avec beaucoup de variables, cette 10 option a l'avantage de trouver des variables en tapant les premières lettres du nom de la variable. En appuyant sur le bouton « **Go to** » on arrive à la variable cherchée dans la vue de données.

Menu Edit : en général, il est plus facile de trouver des variables dans les listes des boîtes dialogue quand elles sont affichées par leur nom (variable name) et pas par leur étiquette (**variable label**). En plus, il est recommandé de choisir le même ordre que dans le fichier (au lieu d'un ordre alphabétique). Pour cela, il faut aller sous **Edit | Options: General, Variable list : Display names** (au lieu de **labels**), **File** (au lieu de **alphabetical**).

Menu View : pour passer de la vue des données à la vue des variables et vice versa, il faut cliquer sur **Variable View** (respectivement « **Data View** ») dans le coin inférieur gauche de la grille des données ou via le menu **View | Variable** (respectivement **Data**)

6. Les différentes fenêtres de SPSS (données, résultats et syntaxe)

Fenêtre de données (Data Editor)

Cette fenêtre permet d'entrer des données, de les modifier ou de les effacer. Il est rare que l'on tape les données manuellement dans SPSS car il y a trop d'erreurs de saisie possibles (on reviendra sur ce point dans [la partie B](#)). On va plutôt ouvrir un fichier déjà existant.

Fenêtre des résultats (Output Editor)

Cette fenêtre apparaît après qu'une commande d'analyse a été effectuée, et contient les résultats de cette analyse. Les résultats apparaissent à droite dans la fenêtre. A gauche, figure une table des matières des résultats générés par SPSS.

Les résultats peuvent être imprimés tels quels (mais avec le risque de voir un tableau s'imprimer sur plusieurs pages. Il est également possible de copier les tableaux qui nous intéressent pour les coller ensuite dans Word, Excel ou dans un autre logiciel. Il est possible de copier un tableau de deux manières. En cliquant sur le tableau en appuyant sur le bouton de droite de la souris, SPSS vous propose de copier (**copy**) ou de copier l'objet (**copy object**). Copier correspond à copier les valeurs, mais lorsqu'il est collé il peut perdre son format (utile pour copier les résultats dans une feuille Excel par exemple). Copier les objets correspond à copier les valeurs et le format du tableau : une fois collé, impossible de modifier les cellules du tableau (utile pour copier les résultats dans Word).

Parfois, le résultat est très long et SPSS n'en montre qu'une partie, suivi d'un triangle rouge. Pour visualiser la suite, il faut double-cliquer sur le texte, puis étendre la taille de la zone de texte vers le bas jusqu'à ce que vous ne voyez plus de texte.

Fenêtre de syntaxe (Syntax Editor)

Jusqu'à maintenant, nous avons vu comment travailler avec les menus déroulant. Il existe une autre manière de lancer des analyse : passer par la fenêtre de syntaxe. Cette fenêtre permet d'écrire les commandes d'analyses statistiques. Elle fonctionne comme un traitement de texte simple.

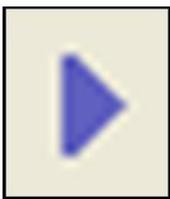
Mais pourquoi s'amuser à taper les analyses quand il suffit de faire défiler des menus et cliquer sur des icônes ? Parce que lorsqu'on veut refaire une analyse qu'on a déjà effectuée et juste changer un paramètre, il suffit de reprendre la feuille de syntaxe de l'analyse précédente, changer un mot et renvoyer l'analyse. Sinon, il faudrait recommencer toute l'analyse de A à Z, vu qu'on ne peut pas enregistrer la séquence de clics. De plus, les menus ne donnent pas accès à toutes les commandes SPSS. Finalement, c'est aussi utile pour faire la même analyse sur plusieurs fichiers de données.

Ainsi, une fois la syntaxe faite pour une opération, il est facile d'enregistrer les commandes et de les réutiliser pour différents fichiers de données.

Voici les règles générales pour écrire des commandes dans SPSS :

- Chaque nouvelle commande se trouve en tête de ligne, précédée d'aucun espace.
- Les options qui suivent une commande débutent sur la ligne suivante et sont précédées d'au moins un espace et d'une barre oblique (/).
- Chaque commande doit ABSOLUMENT se terminer par un point.
- Lorsqu'on spécifie un nom de fichier, il doit être "entre guillemets".
- SPSS ne fait pas de différence entre les lettres majuscules et minuscules. Vous pouvez taper les commandes autant d'une manière ou de l'autre.
- De plus, entre les commandes, vous pouvez insérer des lignes vides. SPSS les ignore, mais elles peuvent améliorer la lisibilité des commandes quand il y en a plusieurs dans une fenêtre.

Pour exécuter une commande, il faut sélectionner la commande à exécuter puis choisir dans le menu « **Run : Current** » ou encore Ctrl-R ou encore utiliser le bouton 'run'



Vous pouvez enregistrer votre script pour le modifier plus tard (menu **File | Save**). Vous pouvez aussi enregistrer la fenêtre de résultats bien que très souvent, il en résulte un fichier énorme (que vous ne pourrez sauvegarder sur une disquette). Cependant, tant et aussi longtemps que vous avez votre syntaxe et vos données, vous pouvez toujours exécuter la syntaxe à nouveau pour revoir les résultats.

7. Menu Graphs

Le menu graphique permet de générer en quelques cliques de souris toute une panoplie de graphiques. Nous reviendrons plus en détail lors de [la partie D](#).

Le menu **Graphs** permet en quelque clique de souris de reproduire le graphique. Pour cela, allez chercher votre base de données (exemple : data.sav). Enregistrez-la sur votre répertoire. Puis ouvrez SPSS par **Démarrer | Programme | SPSS for Windows | SPSS 13.0 for Windows**. Allez sous **File | Open | Data**. Cherchez votre base de données que vous désirez traiter. Une fois la base de données ouvertes, allez sous **Graphs**.

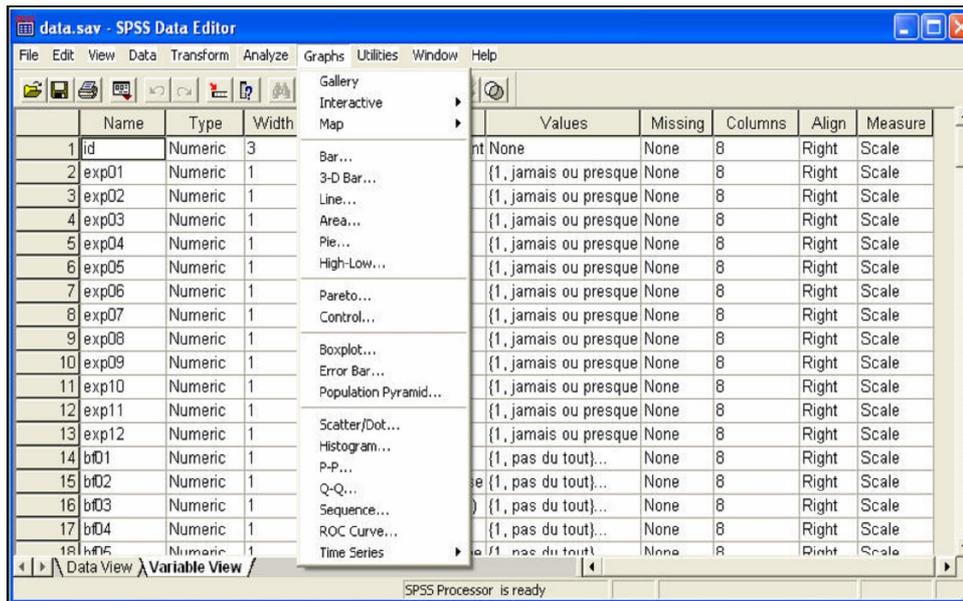


Figure 8

Voici un exemple d'un **Scatter Plot** : sélectionnez **Scatter/Dot...| Simple Scatter**. Dans **Y Axis** introduisez la première variable avec des valeurs continues, par exemple : **sexfreq** (fréquence des relations sexuelles) et dans **X Axis**, la deuxième variable avec de nouveau des valeurs continues, par exemple **satisf** (satisfaction de la relation avec la mère). Cliquez sur **Ok** et vous obtenez le graphique suivant :

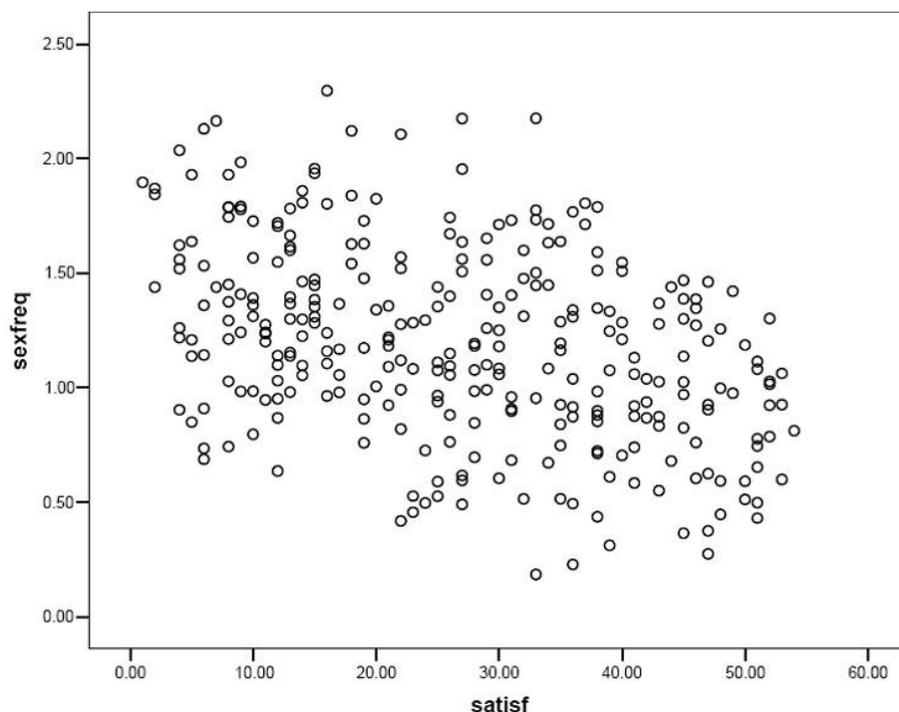


Figure 9

8. Menu Help

SPSS offre plusieurs niveaux d'aide (extrait tiré de l'aide de SPSS pour Windows 13.0 en français):

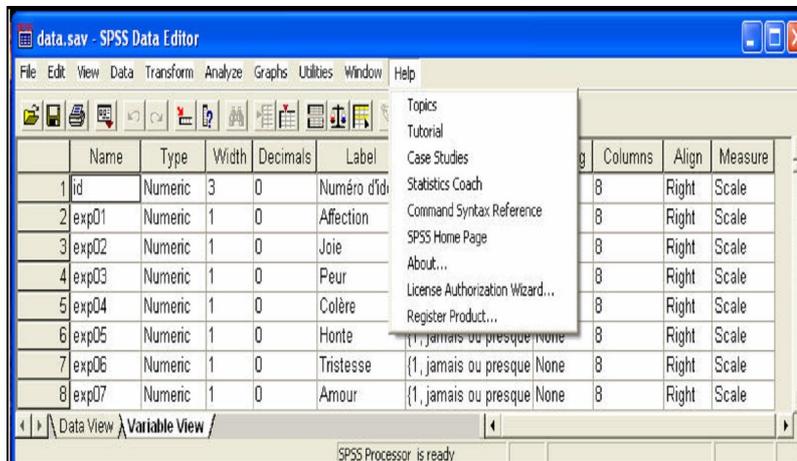


Figure 10

L'aide apparaît sous plusieurs formes :

Menu Aide : Le menu Aide de la plupart des fenêtres SPSS permet d'accéder au système d'aide principal, ainsi qu'aux didacticiels et aux informations de référence technique.

- Rubriques (Topics) : Les rubriques permettent d'accéder aux onglets Sommaire, Index et Rechercher, que vous pouvez utiliser pour chercher des rubriques d'aide particulières.
- Didacticiel (Tutorial) : Instructions illustrées et détaillées étape par étape vous expliquant comment utiliser de nombreuses fonctions de base de SPSS. Il n'est pas nécessaire de visualiser le didacticiel du début à la fin. Vous pouvez choisir les rubriques à visualiser, ignorer et visualiser des rubriques dans l'ordre de votre choix, et utiliser l'index ou le sommaire pour rechercher des rubriques données.
- Etudes de cas (Case Studies) : Exemples pratiques indiquant comment créer différents types d'analyse statistique et comment interpréter les résultats. Les fichiers de données d'exemple utilisés dans ces cas pratiques vous sont également fournis afin que vous puissiez voir exactement comment les résultats ont été générés. Vous pouvez choisir à partir du sommaire les procédures sur lesquelles vous souhaitez obtenir des informations ou rechercher des rubriques appropriées dans l'index.
- Assistant statistique (Statistics Coach) : Vous aide à rechercher la procédure que vous souhaitez utiliser. Une fois vos sélections effectuées, l'Assistant statistique ouvre la boîte de dialogue de la procédure statistique, de rapport ou de diagramme qui correspond aux critères sélectionnés. L'Assistant statistique permet d'accéder à la plupart des procédures statistiques et de rapport du système de base, et à de nombreuses procédures de diagramme.

- **Command Syntax Reference** : Vous trouverez des informations détaillées de référence de syntaxe des commandes dans le manuel SPSS Command Syntax Reference, accessible via le menu Aide.

Aide sensible au contexte : Plusieurs emplacements de l'interface utilisateur vous permettent d'accéder à l'aide contextuelle.

- Boutons Aide de boîte de dialogue : La plupart des boîtes de dialogue disposent d'un bouton Help qui vous conduit directement à la rubrique d'aide relative à la boîte de dialogue. La rubrique d'aide fournit des informations générales et propose des liens vers les rubriques apparentées.
- Aide du menu contextuel de la boîte de dialogue : De nombreuses boîtes de dialogue fournissent une aide contextuelle pour les commandes et les fonctions. Cliquez avec le bouton droit sur n'importe quel contrôle d'une boîte de dialogue et sélectionnez **What's this ?** dans le menu contextuel. Une description du contrôle s'affiche alors, ainsi que son mode d'utilisation. (Si **What's this ?** n'apparaît pas dans le menu contextuel, ce type d'aide n'est pas disponible pour cette boîte de dialogue.)
- Aide du menu contextuel du tableau pivotant : Cliquez sur le bouton droit de la souris sur les termes du tableau pivotant dans le Viewer et sélectionnez **What's this ?** dans le menu contextuel afin d'afficher les définitions de ces termes.
- Etudes de cas : Cliquez avec le bouton droit de la souris sur un tableau pivotant, puis sélectionnez **Case Studies** dans le menu contextuel pour accéder directement à un exemple détaillé relatif à la procédure qui a généré ce tableau. (Si Case Studies n'apparaît pas dans le menu contextuel, ce type d'aide n'est pas disponible pour cette procédure.)
- Tableaux de syntaxe de commande : Dans une fenêtre de syntaxe de commande, positionnez le curseur n'importe où dans un bloc de syntaxe de commande, puis appuyez sur la touche **F1**. Vous voyez alors apparaître un tableau de syntaxe de commande complet. (L'Aide français de SPSS 13.0 pour Windows)

Exercice 1 : Questions à choix multiple

Une question à la fois

1. Quelle est l'extension que porte le fichier de données?

- A. ? .sav
- B. ? .spo
- C. ? .sps
- D. ? .txt

2. Où les données sont-elles affichées ?

- A. ? Variable view
- B. ? File
- C. ? Output
- D. ? Data View

3. Dans Variable View, à quoi chaque ligne correspond-elle ?

- A. ? un item
- B. ? un chiffre
- C. ? une variable
- D. ? un menu

4. Quelle est l'extension que porte le fichier des résultats ?

- A. ? .sav
- B. ? .spo
- C. ? .sps
- D. ? .txt

5. Dans quelle fenêtre peut-on entrer les données, de les modifier ou de les effacer ?

- A. ? Output Editor
- B. ? Syntax Editor
- C. ? Graphs Editor
- D. ? Data Editor

6. Quelle est l'une des règles générales pour écrire des commandes dans SPSS ?

- A. ? Chaque commande peut se terminer par une virgule.
- B. ? Lorsqu'on spécifie un nom de fichier, il doit être "entre guillemets".
- C. ? Spss fait une différence entre minuscule et majuscule.
- D. ? Entre les commandes, vous ne pouvez pas insérer des lignes vides.

7. Pour obtenir un ou des graphiques, où devons nous cliquer ?

- A. ? menu File
- B. ? menu Data
- C. ? menu Graphs
- D. ? menu Utilities

8. Comment peut-on accéder au didacticiel SPSS?

- A. ? avec le menu Help | Tutoriel
- B. ? avec le menu Graphs | Scatterplot
- C. ? avec le menu File | Open
- D. ? avec le menu Help | Topics

