

Application statistique sur ordinateur

Introduction à SPSS

1 Fichiers et fenêtres SPSS Statistics

Comme la plupart des logiciels de statistiques, SPSS Statistics utilise deux types de fichier : un fichier de données (extension .sav) et un fichier de résultats (.spv). Ces deux fichiers sont indépendants et doivent être enregistrés séparément.

En général, on ouvre un fichier de données (.sav) via Fichier > Ouvrir > Données... ou en double-cliquant sur le fichier. Ce fichier est visible dans la fenêtre **Editeur de données**. Lors du premier traitement statistique, une fenêtre **Viewer** est créée. Elle affiche les résultats des traitements statistiques. Il est possible d'enregistrer le contenu de cette fenêtre dans un fichier de résultats (.spv) via Fichier > Enregistrer.

On passe d'une fenêtre à l'autre, via la barre des tâches de Windows.

1.1 La fenêtre Editeur de données

Cette fenêtre affiche le fichier de données sur lequel les traitements statistiques seront effectués. Il s'agit d'un tableau dont les lignes correspondent aux **observations** (individus) et les colonnes aux **variables** (caractères).

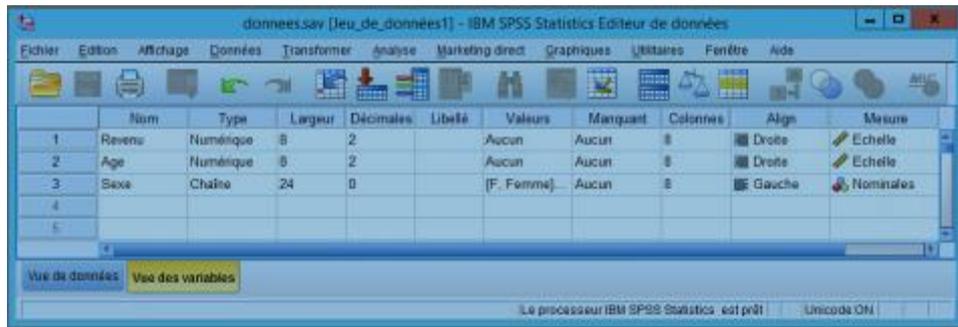
	Revenu	Age	Sexe	var	var	var	var	var	v
1	1203,00	24,00	M						
2	1459,00	43,00	F						
3	1123,00	23,00	F						
4	1796,00	54,00	M						
5	2439,00	55,00	M						
6	1997,00	39,00	F						
7									

Il est possible d'utiliser cette fenêtre pour saisir des données, même si en général, on préfère les saisir dans Excel pour les importer ensuite dans SPSS Statistics via Fichier > Ouvrir > Données | Fichiers de type > Excel (*.xls, *.xlsx, *.xlsm).

En bas en gauche de la fenêtre, deux onglets Vue des données et Vue des variables permettent de passer de l'affichage des données au paramétrage des variables.

1.2 L'onglet Vue des variables

Cet onglet permet de définir précisément la nature des variables et la façon dont elles seront affichées.



Chaque ligne correspond à une variable ; les colonnes représentent les différents attributs de chaque variable :

Nom : nom interne de la variable utilisé par SPSS Statistics. Ce nom doit être simple, c'est-à-dire, sans accent, sans espace ni caractère exotique. Se limiter aux 26 lettres de l'alphabet complétées par des chiffres et le tiret bas (_). En outre, le nom d'une variable est limité à 8 caractères.

Type : type (informatique) de données contenues dans la variable. Les deux types principaux sont **Numérique** (un nombre, éventuellement décimal, par exemple 123,45) et **Chaîne** (une chaîne de caractères, par exemple Limoges)

Largeur/Décimales : précise le format du type de données : nombre de caractères/chiffres (largeur) et nombre de chiffres après la virgule (décimales)

Libellé : nom de la variable utilisé pour l'affichage des résultats. Aucune limitation pour ce champ. Toutefois, il est préférable d'éviter les noms trop longs qui encombreraient l'affichage des résultats.

Valeurs : permet de définir des libellés de valeurs utilisées lors de l'affichage des résultats. Par exemple, il est fréquent de coder 1/0, la réponse à une question dichotomique (Oui / Non).

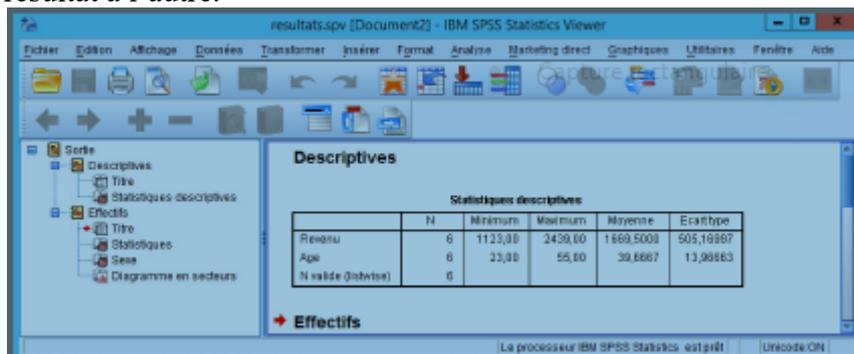
Manquant : permet de spécifier les valeurs utilisées pour représenter les données manquantes (données non renseignées). En général, 9, 99, 999, etc.

Colonnes/Alignx : permet de définir la largeur de la colonne et l'alignement des données dans la colonne.

Mesure : précise l'échelle de mesure utilisée : **Echelle** (variable quantitative), **Ordinales** (variable qualitative ordinale), **Nominales** (variable qualitative nominale).

1.3 La fenêtre Viewer

Cette fenêtre contient la suite chronologique des traitements statistiques effectués. Le navigateur de résultat, dans la partie gauche de la fenêtre, permet de passer rapidement d'un résultat à l'autre.

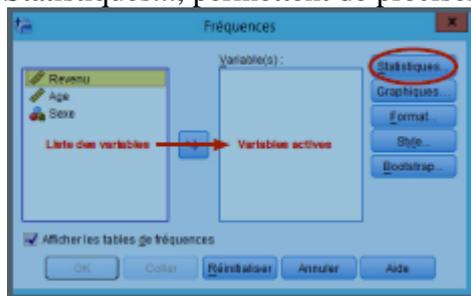


Il est possible de modifier l'affichage d'un résultat (tableau, graphique, etc) en double-cliquant dessus. Cette fenêtre peut être enregistrée dans un fichier de résultats (.spv) via Fichier > Enregistrer.

1.4 Utilisation des fonctions

La plupart des fonctions SPSS Statistics décrites dans la suite de ce document porte sur les variables définies dans le fichier de données. Les dialogues permettant d'accéder à ces fonctions ont souvent une apparence proche de celle ci en dessous.

Le cadre de gauche affiche la liste de toutes les variables. A l'aide de la flèche, on fait passer dans le cadre de droite les variables qu'on souhaite étudier. Des boutons, du type Statistiques..., permettent de préciser le traitement à effectuer.



2 Transformation des données

A partir de la fenêtre Editeur des données, il est possible de modifier les variables et les observations initiales.

2.1 Transformation des variables

Transformer > Calculer la variable : permet de définir une nouvelle variable à partir des variables initiales, en utilisant une formule de type Excel. Par exemple, si un prix est exprimé en francs, il est possible de définir une variable P_euros via la formule $P_francs / 6,559$.

Transformer > Recoder automatiquement : permet de transformer une variable qualitative (Mariée, Célibataire, . . .) en une variable numérique (1,2, . . .). De plus, des libellés de valeurs seront automatiquement définies pour la nouvelle variable.

Transformer > Regroupement visuel : permet de transformer une variable quantitative en une variable qualitative ordinaire en définissant des classes de valeurs.

Transformer > Création de variables : version paramétrable des deux procédures précédentes.

2.2 Manipulation des observations

Données > Sélectionner des observations : permet de préciser les observations (lignes) qui doivent être prises en compte dans les traitements statistiques, en se basant sur les valeurs d'une variable. Par exemple, il est possible de ne sélectionner que les observations dont la variable age est supérieure à 60 pour étudier uniquement la population de plus de 60 ans. Les autres observations ne sont pas effacées mais désactivées. Il est nécessaire de refaire cette opération à chaque ouverture du fichier de données.

Données > Pondérer les observations : permet d'associer un poids à chaque observation, soit en cas d'un échantillon redressé, soit pour indiquer un effectif à chaque type d'observation.

3 Statistiques descriptives univariées

L'ensemble des traitements statistiques et des graphiques est regroupé dans les menus Analyse et Graphiques.

Analyse > Rapport > Récapitulatif des observations : permet de visualiser les données, en affichant l'ensemble des observations. Il est aussi possible d'afficher des statistiques descriptives (moyenne, minimum, maximum, . . .) et même de demander l'affichage par groupe, suivant une variable qualitative (critère de regroupement).

3.1 Variables quantitatives

Analyse > Statistiques descriptives > Descriptives : statistiques descriptives pour des variables quantitatives : moyenne, minimum, maximum, écart-type.

Graphiques > Boîtes... > Histogramme : représentation graphique, sous forme d'histogramme d'une variable quantitative.

3.2 Variables qualitatives

Analyse > Statistiques descriptives > Fréquences : statistiques descriptives pour des variables qualitatives : tris à plat, mode, médiane, quartiles, centiles, . . . (à préciser dans le dialogue Statistiques). Il est aussi possible d'obtenir des graphiques via le dialogue Diagrammes (voir ci-dessous)

Graphiques > Boîtes... > Barres > Simple | Récapitulatifs pour groupes d'observations : diagramme en bâtons des modalités d'une variable qualitative. Possibilité d'afficher les effectifs ou les pourcentages. Effectifs préférables. Adapté aux variables qualitatives ordinales

Graphiques > Boîtes... > Circulaire > Récapitulatifs pour groupes d'observations : diagramme en secteur (camembert) des modalités d'une variable qualitative . Possibilité d'afficher les effectifs ou les pourcentages. Pourcentages préférables. Adapté aux variables qualitatives nominales.

4 Autres statistiques univariées

Graphiques > Boîtes... > Boîte à moustaches > Simple | Récapitulatifs pour variables distinctes : construit à une boîte à moustache (box-plot) permettant de déceler les valeurs aberrantes d'une variable quantitative.

Analyse > Statistiques descriptives > Tracés P-P ou Q-Q : graphique Probabilité-Probabilité ou Quantile-Quantile permettant d'évaluer la normalité d'une variable.

Analyse > Tests non paramétriques > Boîtes... > K-S à 1 échantillon : Test de Kolmogorov-Smirnov permettant de tester la normalité d'une variable.

5 Statistiques descriptives bivariées

5.1 Variables quantitatives / qualitatives ordinales

Analyse > Corrélation > Bivariée : Calcule le coefficient de corrélation de Pearson (variables quantitatives) ou de Spearman (variables qualitatives ordinales) entre deux (ou plus) variables.

Graphiques > Boîtes... > Dispersion/Points > Dispersion simple : affiche le diagramme de dispersion de 2 variables quantitatives ; chaque observation est représentée dans un diagramme à 2 dimensions dont les axes correspondent aux deux variables sélectionnées.

5.2 Variables qualitatives nominales ou ordinales

Analyse > Statistiques descriptives > Tableaux croisés : affiche un tri croisé de 2 variables qualitatives. Possibilité d'indiquer dans le dialogue Cellules le contenu des cases : effectifs (observé), pourcentages (par ligne, par colonne, total). Le dialogue Statistiques permet d'obtenir le Khi-deux et le V de Cramer.

Graphiques > Boîtes... > Barres > En cluster | Récapitulatifs pour groupes d'observations : affiche le diagramme en bâtons juxtaposés ou empilés de deux variables qualitatives.

6 Tests sur une moyenne

Analyse > Comparer les moyennes > Test T pour échantillon unique : permet de tester la valeur de la moyenne d'une variable avec une valeur spécifiée.

Analyse > Comparer les moyennes > Test T pour échantillons indépendants : permet de tester l'égalité de la moyenne d'une variable sur deux groupes d'observations définis par une variable qualitative.

Analyse > Comparer les moyennes > Test T pour échantillons appariés : permet de tester l'égalité des moyennes de deux variables sur un même échantillon.

7 Statistiques multivariés

Analyse > Réduction des dimensions > Analyse factorielle : analyse en composantes principales de plusieurs variables quantitatives.

Analyse > Réduction des dimensions > Analyse des correspondances : analyse factorielle des correspondances (simples) entre deux variables qualitatives.

Analyse > Réduction des dimensions > Codage optimal : analyse factorielle des correspondances multiples entre plusieurs variables qualitatives.

Analyse > Classification > Analyse discriminante : analyse factorielle discriminante expliquant une variable qualitative par plusieurs variables quantitatives.

Analyse > Classification > Cluster hiérarchique : classification hiérarchique ascendante.

Analyse > Classification > Cluster de nuées dynamiques : classification en nuées dynamiques.