

**Communauté Economique et Monétaire de l'Afrique Centrale  
(CEMAC)**



---

*Institut Sous-régional de Statistique et d'Economie Appliquée*

---

# Initiation à SPSS

---

Ngahane Cédric  
Ingénieur Statisticien Economiste  
2015

# Table de Matière

---

L'environnement SPSS .....	4
La fenêtre Editeur de données.....	4
La fenêtre résultat .....	6
La fenêtre éditeur de syntaxe.....	7
Lecture des données dans SPSS. ....	7
La lecture à partir d'une source interne .....	7
La lecture à partir d'une source externe. ....	8
Importation depuis un fichier texte .....	8
Importation depuis une feuille de calcul Excel.....	8
Préparation de la base de données .....	10
Définition et modification des propriétés des variables .....	10
Définition et modification des propriétés des variables via la grille des variables.....	10
Définir des propriétés à partir des Menus .....	12
Gestion des données manquantes .....	13
Analyse univariée.....	15
Variable qualitative .....	15
Tri à plat .....	15
Diagrammes pour données qualitatives .....	16
Variable quantitative .....	17
Description à partir des statistiques calculées .....	17
Description à partir des graphiques .....	19
Les graphiques dans SPSS.....	20
Construction des graphiques.....	21
Modification des graphiques .....	25
Les options de Diagramme.....	31
Manipulation des données .....	32
Trier les données.....	32
Filtrer les données - Sous population.....	33
Recodifier les variables.....	35
Créer de nouvelles variables.....	36
Création d'une variable par recodification d'une autre. ....	37
Création d'une variable par combinaison d'une ou plusieurs autres.....	38
Fusionner les données (merge).....	39

Ajout des observations.....	39
Ajouter des variables .....	41
Agréger les données.....	45
Gérer des données pondérées .....	46
Création et modification des tableaux .....	47
Création des tableaux.....	47
Tableaux de Contingence .....	48
Tableaux présentant les statistiques d'une variable quantitative selon les classes d'une variable qualitative.....	51
Empilement des variables .....	53
Tableau pour variables avec modalités partagées .....	54
Tableaux avec utilisation de vecteurs de réponses multiples .....	55
Tableaux avec variables emboîtées .....	58
Tableau avec strate .....	61
Tableau avec variable d'échelles source des statistiques récapitulatives.....	63
Modification des tableaux.....	65
Modification de la structure d'un tableau.....	65
Modification et mise en forme des étiquettes des tableaux .....	66
Modification des formats d'affichage des données.....	67
Masquer des lignes ou des colonnes.....	68
Modèle de tableau .....	69
Utilisation des résultats .....	71
Eléments d'informations supplémentaires sur le Viewer.....	71
Exportations des résultats .....	72
Copier/coller vers Excel et Word par l'intermédiaire du presse-papier .....	72
Exportation des résultats vers un fichier Microsoft Word, PowerPoint ou Excel.....	72
Programmer avec SPSS .....	73
Règles de la syntaxe SPSS .....	74
La fenêtre Editeur de syntaxe.....	75
Ouverture d'un fichier de syntaxe.....	75
Utilisation de la fenêtre syntaxe .....	75
Exécution interactive d'un fichier de commandes. ....	76
Création d'un fichier de syntaxe .....	78

# Introduction

---

SPSS (Statistical Package for the Social Sciences), de par son ancienneté est l'un des logiciels les plus connus dans le monde de la statistique, particulièrement dans le domaine des sciences. En effet, les premières versions du logiciel datent des années 60. Initialement produit par SPSS Inc, SPSS a été acquis en 2009 par la société IBM et porte maintenant le nom officiel de IBM SPSS Statistics. La dernière version datant de mars 2015, est la version 23.

L'objet de ce document est d'initier l'utilisateur au traitement de données et à l'analyse statistique à l'aide du logiciel SPSS. Ce support est un document vraiment basique, s'adressant à large audience, principalement à ceux qui ne sont pas familiers à l'environnement SPSS.

SPSS dispose d'une littérature très fournie. En outre le présent document est fortement inspiré du Guide de l'utilisateur. Pour différents approfondissements, le lecteur est prié de s'y référer ou de consulter d'autres se focalisant sur des thématiques bien définies. En effet, la maîtrise du logiciel passera par l'étude de documents abordant de manière plus approfondie différents aspects et fonctionnalités du logiciel. Comme pour tout logiciel statistique l'apprentissage nécessitera beaucoup d'exercices pratiques et beaucoup de temps consacrés à l'appropriation du logiciel.

Les fichiers d'exemples utilisés dans ce document sont ceux fournis lors de l'installation du logiciel, trouvant dans le dossier *sample* du répertoire d'installation. Plus spécifiquement les illustrations s'appuieront sur deux de ces fichiers :

***demo.sav***: Ce fichier est une base de données client provenant d'une enquête fictive contenant des informations démographiques et relatives à la consommation.

***customer\_dbase.sav***: ce fichier contient également des données relatives à la clientèle d'une société. Cette dernière utilise ces informations pour étudier les réponses des clients à des offres spéciales susceptibles de les intéresser.

En plus de ces fichiers, le document s'appuie également sur les données issues troisième Enquête Camerounaise auprès des ménages (**ECAM 3**), réalisé par l'Institut National de la Statistique.

Ce support de cours est basé sur la version 21.0 du logiciel.

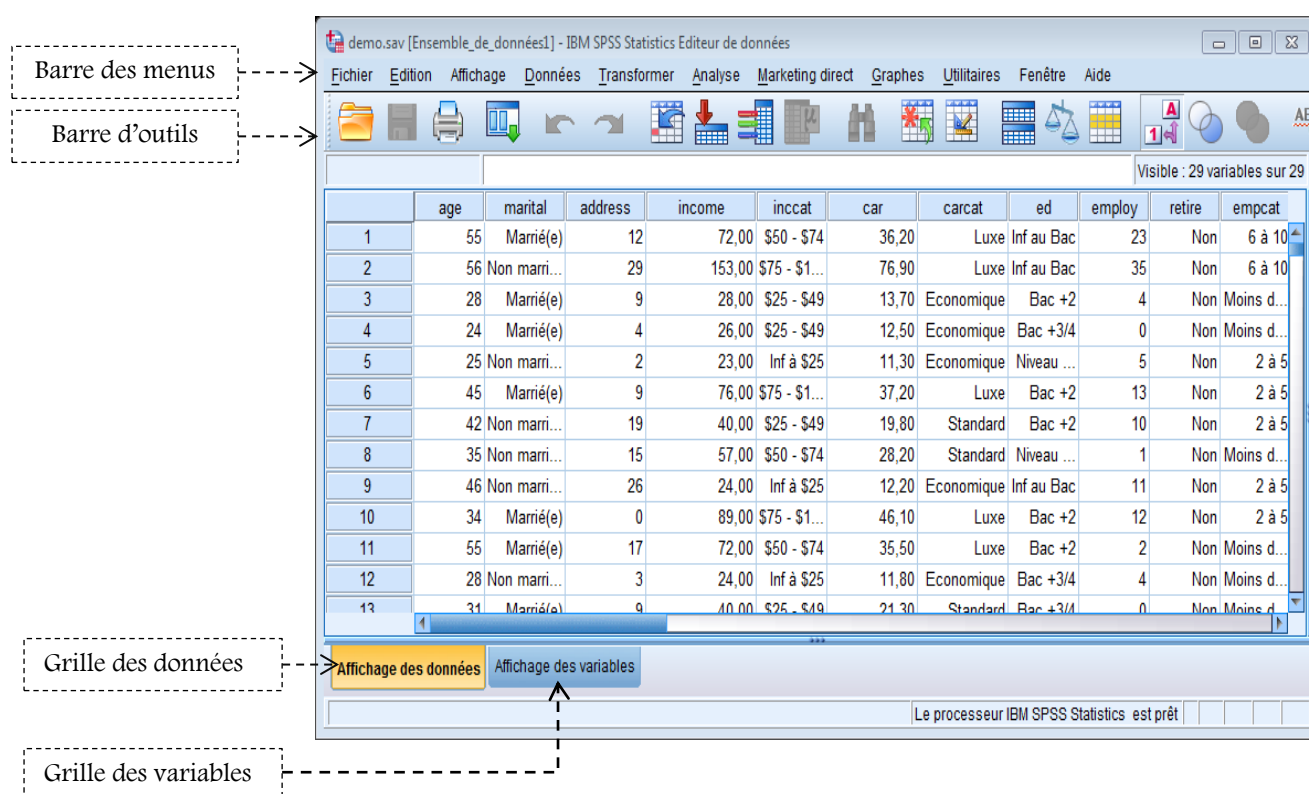
## 1 L'environnement SPSS

L'environnement SPSS comprend 3 principales. On distingue ainsi :

- ✓ La Fenêtre Éditeur De Données
- ✓ La Fenêtre Résultats
- ✓ La Fenêtre Syntaxe

### 1.1 La fenêtre Editeur de données

La fenêtre Editeur de données est la principale fenêtre de SPSS. Elle apparaît après le lancement de SPSS. Elle permet d'entrer (saisir) des données, de modifier, de supprimer et d'afficher ces derniers. Il est toutefois rare lors de l'utilisation courante du logiciel de saisir les données directement les données dans l'éditeur. On aura pour usage d'importer les données.



Dans cette fenêtre, on retrouve la barre des Menu qui permet via ses différents menus (Fichier Edition Données ...) d'accéder aux différentes fonctionnalités de SPSS, notamment, la gestion des fichiers, la réalisation des traitements et analyses, les réglages d'affichage, l'aide, etc.

La barre d'outils, quant à elle, fournit des raccourcis pour les fonctionnalités les plus couramment utilisées de la barre des menus.

L'éditeur des données possèdent deux onglets qui permettent d'accéder à deux différentes grilles : celle des données et la grilles des variables.

**La grille des Données** s'obtient via l'onglet « Affichage des données ». C'est elle l'on retrouve dans la vue précédente. Les données y sont affichées sous forme de tableau où chaque

colonne représente une variable et chaque ligne un individu. Les données peuvent y être directement saisies ou directement modifiées (bien que ce ne soit pas conseillé).

**La grille des variables** présente toutes les variables de la base ainsi que leurs propriétés. Elle s'obtient via l'onglet « Affichage des variables ». Elle constitue en quelque sorte le « dictionnaire des variables ».

	Nom	Type	Largeur	Décimales	Etiquette	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	age	Numérique	4	0	Age en années	Aucun	Aucun	6	Droite	Echelle	Entrée
2	marital	Numérique	4	0	Statut marital	{0, Non mar...	Aucun	7	Droite	Nominales	Entrée
3	address	Numérique	4	0	Nombre d'anné...	Aucun	Aucun	7	Droite	Ordinales	Entrée
4	income	Numérique	8	2	Revenu du foye...	Aucun	Aucun	10	Droite	Echelle	Entrée
5	inccat	Numérique	4	0	Catégories de r...	{1, Inf à \$25...	Aucun	6	Droite	Ordinales	Entrée
6	car	Numérique	8	2	Price of primary...	Aucun	Aucun	8	Droite	Echelle	Entrée
7	carcat	Numérique	8	2	Catégorie de pri...	{1,00, Econ...	Aucun	8	Droite	Ordinales	Entrée
8	ed	Numérique	4	0	Nombre d'anné...	{1, Inf au Ba...	Aucun	6	Droite	Ordinales	Entrée
9	employ	Numérique	4	0	Nombre d'anné...	Aucun	Aucun	6	Droite	Ordinales	Entrée
10	retire	Numérique	4	0	Retraité(e)	{0, Non}...	Aucun	6	Droite	Nominales	Entrée
11	empcat	Numérique	4	0	Nombre d'anné...	{1, Moins d...	Aucun	6	Droite	Ordinales	Entrée
12	jobsat	Numérique	4	0	Satisfaction E...	{1, Très ins...	Aucun	6	Droite	Ordinales	Entrée
13	gender	Chaîne	1	0	Sexe	{f, Femme}...	Aucun	8	Gauche	Nominales	Entrée
14	reside	Numérique	4	0	Nombre de per...	Aucun	Aucun	6	Droite	Echelle	Entrée

Dans cette grille, chaque ligne représente une variable de la base de données et chaque colonne une propriété de la variable. Les principales propriétés dans cette grille sont :

- ❖ **Le Nom** : c'est tout simplement de la variable. Chacune des variables doit posséder un nom unique. Celui-ci ne peut ni commencer par un nombre, ni contenir des caractères spéciaux tels que - / espace et bien d'autres. Par ailleurs, pour des versions antérieures du logiciel, le nombre maximal de caractères pour un nom était limité à 8. Il pourrait être avisé de toujours respecter cette limitation.
- ❖ **Le Type** : Il s'agit du type de la variable. **Cependant, il ne s'agit pas du type au sens statistique du terme.** En effet, le type dans ce contexte a plutôt trait au format (d'affichage et de stockage) des données. Les types les plus courants sont :

Le type numérique : il se rapporte aux variables quantitatives et à celles codifiées à l'aide de caractère numérique. Par exemple, si la variable « sexe » est codifiée en 1 (pour masculin) et 2 (pour féminin), elle sera déclarée de type numérique.

Le type chaîne : il concerne les variables ayant comme valeurs des chaînes de caractères. Si par exemple la variable sexe est codifiée en « f » et « m », elle sera déclarée de type chaîne.

Les types « points » et « virgule » : ces types se rapportent en quelque sorte au séparateur de milliers. Ainsi, l'affichage des valeurs de la variable dans la grille des données, se fera avec comme séparateur de milliers le point ou la virgule selon le type choisi.

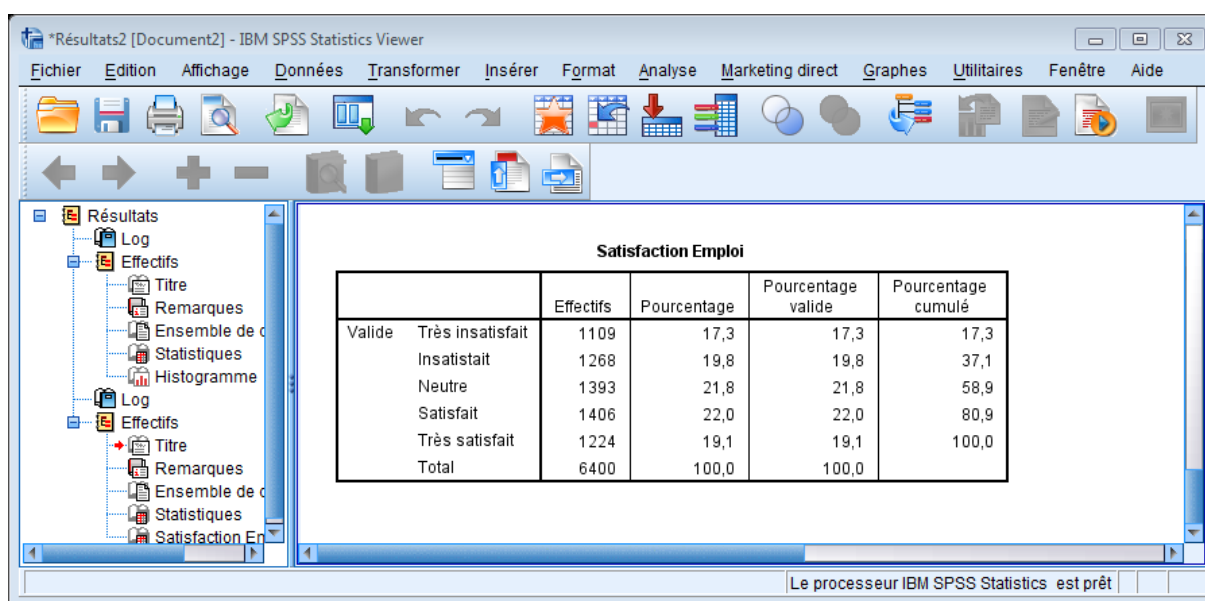
- ❖ **Etiquette** : elle permet de décrire la variable. Le nom obéissant à un certain nombre de contraintes (par exemple le caractère espace proscrit), il arrive que celui-ci ne

permette pas de traduire explicitement le contenu de la variable, d'où l'importance de l'étiquette.

- ❖ **Valeur** : elle permet d'attribuer des étiquettes aux valeurs prises par la variable, donnant ainsi la possibilité de les décrire celles-ci. Par exemple en codifiant la variable sexe en 1 et 2, il est important de préciser ce que représentent respectivement ces valeurs notamment, « masculin » et « féminin » ou bien « féminin » et « masculin »
- ❖ **Manquant** : cette propriété permet de définir les valeurs qui seront considérées comme donnée manquante en plus de celles qui le sont déjà par défaut.
- ❖ **Décimales** : c'est le nombre de décimales pour les variables de type numérique
- ❖ **Largeur** : lorsque la variable est de type chaîne, elle définit le nombre maximal de caractère.
- ❖ **Mesure** : elle correspond à la nature de la variable au sens statistique du terme. On distingue ainsi les mesures d'échelles (variables quantitatives), les mesures ordinales (variables qualitatives ordinales), mesures nominales (variables catégorielles).

## 1.2 La fenêtre résultat

Encore appelée Viewer, elle est automatiquement générée par SPSS et apparaît après l'exécution de tâches. Comme son nom le laisse sous-entendre, SPSS l'utilise pour l'affichage des résultats obtenus après exécution de la tâche.

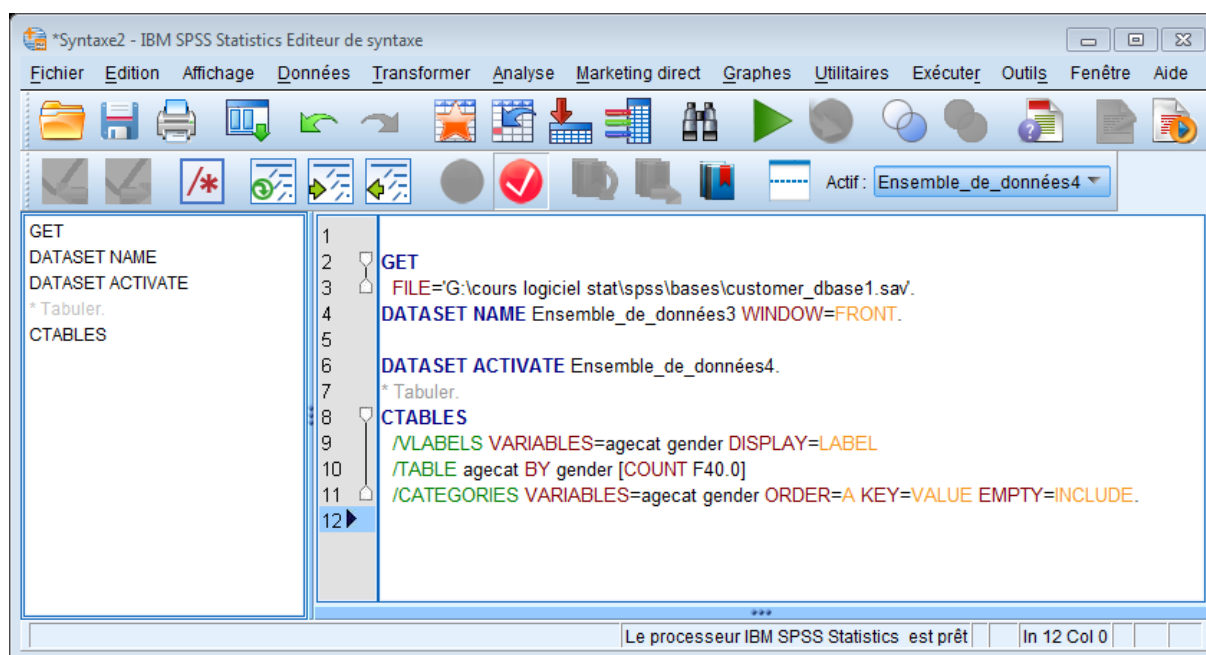


La fenêtre du Viewer possède deux principaux champs. On a ainsi un panneau de navigation qui permet de naviguer entre les différents résultats produits. Il constitue en quelque sorte un sommaire pour les résultats. Le second champ est un panneau où sont affichés les résultats (tableaux, graphiques, ...) proprement-dit. Toutefois, ce second panneau est dynamique, il ne joue pas un simple rôle d'affichage. En effet, une fois les résultats produits et affichés, ils peuvent encore être modifiés à partir du Viewer.

### 1.3 La fenêtre éditeur de syntaxe

Cette fenêtre permet de pourvoir réaliser des tâches dans SPSS via l'exécution de ligne de commandes. Elle permet ainsi de travailler en mode programmation et d'automatiser un certain nombre de tâches. En effet, l'ensemble des opérations effectuées sur SPSS peuvent également en utilisant le langage de commande.

La fenêtre syntaxe a donc pour rôle de permettre d'éditer et d'exécuter un fichier de syntaxe.



## 2 Lecture des données dans SPSS.

Les données utilisées dans SPSS peuvent être soit directement saisies via la fenêtre Editeur de données, soit lues depuis un format ou une source bien défini. Pour cette dernière option, l'étape de saisie des données est une tâche préliminaire qui a déjà été réalisée. C'est ce cas de figure qu'on rencontre le plus souvent dans l'usage courant du logiciel. En effet, la saisie des données dans l'éditeur de données présente un certain nombre d'inconvénients. De plus, les opérations de saisie sont parfois réalisées par de tierces personnes autres que l'analyste.

La lecture de données peut s'effectuer à partir d'une source interne ou externe.

### 2.1 La lecture à partir d'une source interne

On parle de lecture à partir d'une source interne lorsque les données sont disponibles dans un fichier ayant un format propre à SPSS. Le fichier porte alors l'extension « .sav ». C'est celle que portent les fichiers créés à partir de la fenêtre Editeur de données.

La lecture des fichiers d'extension « .sav » est tout ce qu'il y a de plus simple, ceux-ci étant déjà dans le format de SPSS. Pour ce faire :

- ✓ A partir de la barre des menus de la fenêtre Editeur de Données d'exécuter :

**Fichier> Ouvrir>Données**



- ✓ Ensuite, rechercher et sélectionner le fichier à ouvrir.
- ✓ Cliquer sur ouvrir

## 2.2 La lecture à partir d'une source externe.

Lorsque la lecture s'effectue à partir d'une source externe, on parle d'importation de données. L'importation peut se faire à partir de source très variées (fichiers textes, fichiers excel, base de données,...). Dans le cadre de ce document, nous nous limiterons uniquement au cas des fichiers textes et feuilles de calcul Excel.

### 2.2.1 Importation depuis un fichier texte

Dans le domaine de la Statistique, les fichiers textes du fait de leur souplesse sont des formats très couramment utilisés comme support de stockage des données. Ils présentent en outre l'avantage de pouvoir être utilisés par la plupart des logiciels statistiques. Dans ces fichiers les champs (variables) sont généralement délimités par des tabulations, des virgules ou encore des espaces.

- ✓ A partir du menu de l'Editeur de données sélectionner :

Fichier>lire les données textes...

- ✓ Sélectionner le fichier texte à ouvrir.

Après la sélection du fichier, l'assistant d'importation de texte guide l'utilisateur tout au long du processus à travers 6 étapes. Il est important pour celui-ci, de bien définir les paramètres d'importation qui sont présentés dans les différentes boîtes de dialogue relatives à chacune des étapes. Ces étapes se présentent succinctement de la manière suivante

Etape 1 : indiquer s'il existe un fichier pouvant être utilisé comme référence afin de réaliser l'importation

Etape 2 : (i) disposition des variables : indiquer s'il existe un caractère particulier (virgule, espace,...) qui délimite les variables ; (ii) indiquer si les noms des variables sont contenus dans la première ligne du fichier

Etape 3 : (i) indiquer la ligne à laquelle commence la première observation ; (ii) indiquer le nombre d'observations à importer

Etape 4 : indiquer quel séparateur est utilisé, au cas échéant, pour délimiter les variables

Etape 5 : définir les différents types des variables

Etape 6 : indiquer si l'on veut sauvegarder les paramètres d'importation.

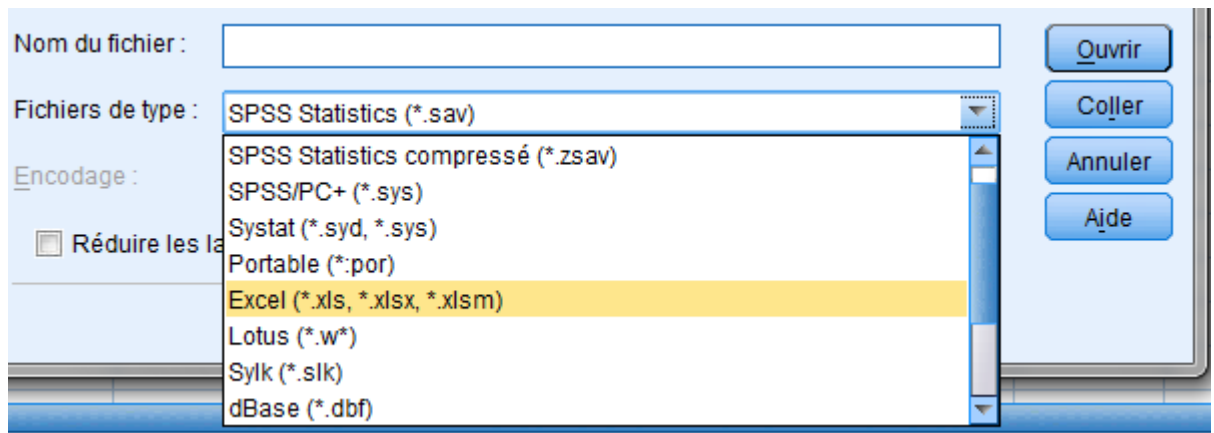
### 2.2.2 Importation depuis une feuille de calcul Excel.

La procédure de la lecture des données à partir d'une feuille Excel débute exactement comme la lecture d'un fichier interne.

- ✓ A partir de la barre des menus de la fenêtre Editeur de Données, exécuter :

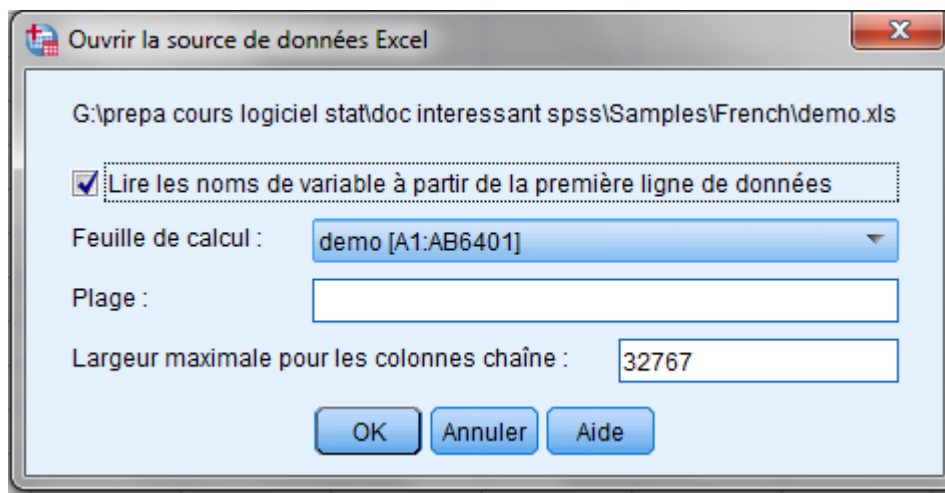
**Fichier> Ouvrir>Données**

- ✓ Indiquez dans l'onglet "Fichiers de type" le format du fichier à ouvrir. Dans le cas d'espèce, il s'agit des fichiers Excel d'extensions : « .xls », « .xlsx », « .xlsm ».



- ✓ Ensuite, sélectionnez le fichier et cliquez sur ouvrir.

Une fois cela effectué apparaît la boîte de dialogue *Ouvrir la source de données Excel*



Le champ *Feuille de calcul* de cette boîte de dialogue permet de spécifier laquelle des feuilles du classeur Excel qui contient les données à lire.

Le champ *Plage* permet de délimiter la plage des cellules de cette feuille qui sera lue.

Par ailleurs, Il est **important** de noter que par défaut, la case « lire les noms de variable à partir de la première ligne de données » est cochée. C'est d'ailleurs le cas de figure que l'on a coutume de rencontrer dans la pratique. Toutefois, dans des circonstances différentes, il est important de s'en souvenir et de la désactiver.

- ✓ Validez en cliquant sur le bouton OK. Pour lancer l'importation.

### 3 Préparation de la base de données

Après avoir importé, les données dans SPSS, il est souvent nécessaire d'effectuer des traitements préalables avant la phase d'analyse. On peut citer entre autres la définition des propriétés des variables ou encore la gestion des données manquantes. En effet, plusieurs tâches telles que la production des graphiques ou la construction des tableaux requièrent que les propriétés des variables (notamment celle relative à la mesure) soient convenablement définies.

#### 3.1 Définition et modification des propriétés des variables

Les propriétés des variables sont celles présentées dans la grille des variables de l'éditeur des données. Pour chacune des variables de la base, ces propriétés peuvent être définies ou modifiées directement à partir de cette grille, ou bien, via les menus de l'éditeur de données

##### 3.1.1 Définition et modification des propriétés des variables via la grille des variables

Cette section ne s'étendra pas sur toutes les propriétés des variables mais se plutôt focalisera sur quelques propriétés clés.

###### 3.1.1.1 Mesure ou typologie des variables

Comme nous souligner précédemment, suivant la nature statistique des variables on distingue trois types de mesures. Les mesures d'échelles (variables quantitatives), les mesures nominales (variables catégorielles) et les mesures ordinales (variables qualitatives ordinales). La définition ou la modification des mesures des variables s'effectue via la colonne Mesure de l'éditeur données. Lorsqu'on clique sur une cellule de cette colonne, cela fait apparaître une liste déroulante dans laquelle l'utilisateur peut choisir la mesure appropriée.

###### 3.1.1.2 Etiquetage des variables et des modalités

Pour ajouter une étiquette à une variable donnée, il suffit, à partir de la grille des variables, de saisir dans la colonne Etiquette la description de la variable. Par exemple dans le fichier de données *demo.sav*, l'étiquette de la variable *situatio* est « la situation familiale ».

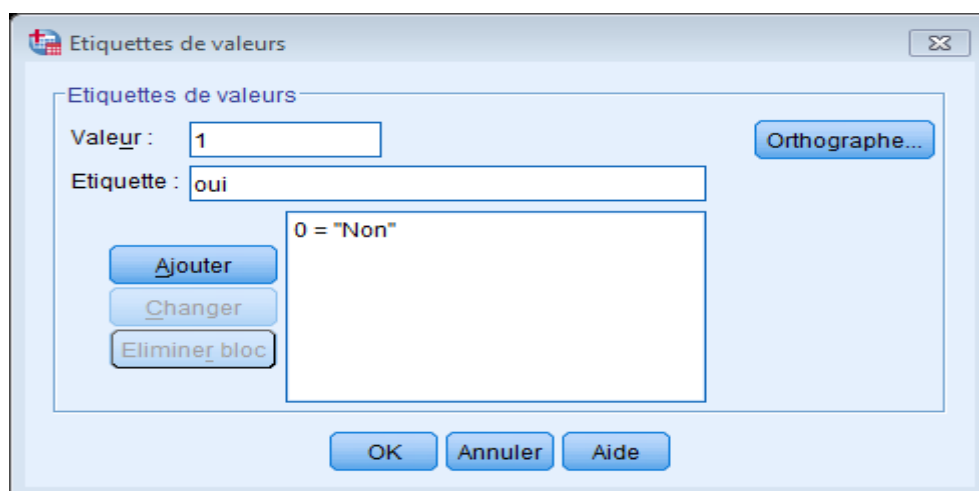
Pour ce qui est de l'attribution d'étiquettes aux valeurs prises par une variable, il convient de rappeler que cette fonctionnalité permet de donner des indications précises sur le sens concret de ces valeurs.

Dans base *demo*, obtenu après importation du fichier excel *demo.xls*, la variable *owntv* (relative à la possession d'une télé) prend la valeur 0 ou 1 indiquant respectivement que le sujet possède ou ne possède pas une télé.

Pour définir les étiquettes pour des valeurs :

A partir de la ligne de variable concernée dans la grille des variables, Cliquez sur la cellule de la colonne Valeurs, puis cliquez sur le bouton à droite de la cellule.

La boîte de dialogue « Etiquettes de valeur », présentée ci-dessous, s'ouvre alors pour permettre de définir ou de modifier les étiquettes de valeurs.



Pour chaque nouvelle étiquette définie, cliquez sur ajouter. Une fois l'opération terminée, validez en cliquant sur le bouton OK.

### 3.1.1.3 Définition des données manquantes

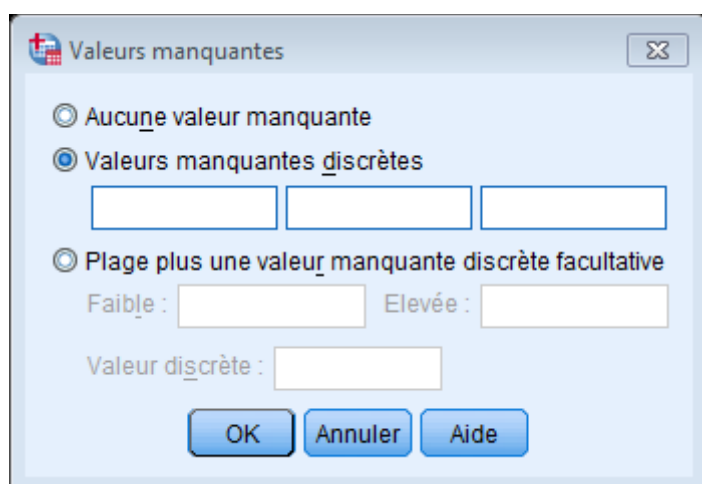
Pour une variable dans la base, il existe deux principaux types de valeurs manquantes à savoir celles identifiées par défaut par SPSS comme telles et celles spécifiées par l'utilisateur (comme valeurs manquantes).

Par défaut, pour les données numériques, les champs de données vides ou contenant des valeurs non valides sont convertis en données manquantes. Dans le cas des variables chaînes par contre, par défaut, un champ vide n'est pas considéré comme une donnée manquante. Il est interprété comme un champ contenant la chaîne de caractère vide.

L'utilisateur peut identifier d'autres valeurs comme donnée manquante, en plus des celles reconnues par défaut par SPSS. Cela se fait via la colonne **Manquant** de la grille des variables.

Pour ce faire, sélectionnez la cellule correspondante à la variable dans la colonne **Manquant** et cliquez sur le bouton apparaissant à droite de la cellule.

Alors la boîte de dialogue **Valeurs manquantes**, présentée ci-dessous s'ouvre pour permettre à l'utilisateur de spécifier les valeurs manquantes.



Cette boîte de dialogue à trois boutons d'option :

- Aucune valeur manquante  
Dans ce cas, seules les valeurs manquantes par défaut sont considérées.
- Valeurs manquantes discrètes  
Cette option offre à l'utilisateur de pouvoir définir jusqu'à trois valeurs distinctes comme manquantes. En effet, il pourrait par exemple être utile dans le cadre d'une étude de distinguer les données manquantes provenant de non réponses de celles où l'enquête n'avait pas la réponse à question. Dans ce cas d'espèce, on codera chacun de ces deux types de données manquantes par une valeur spécifique, et on attribuera leur attribuer des étiquettes propres à chacun des cas. En effet, **il est possible d'attribuer des étiquettes à des valeurs identifiées comme données manquantes**.  
  
Dans le cas particulier des variables chaines, pour déclarer les champs vides comme valeurs manquantes, il suffit d'entrer un seul espace dans l'un des champs.
- Plage plus une valeur manquante discrète facultative.  
Cette option permet de définir une plage de valeur comme manquante avec en plus la possibilité d'identifier une valeur discrète comme donnée manquante.

#### 3.1.1.4 Copie des propriétés d'une variable

Lors de la définition des propriétés des variables, il arrive très fréquemment que l'on rencontre dans la base des variables qui ont des propriétés communes. SPSS offre la possibilité de copier des propriétés d'une variable et de les appliquer à d'autres. Ce procédé peut permettre un gain en temps significatif lors l'exercice de définition des propriétés des variables qui peut se révéler fastidieux en fonction de la taille de la base de données.

Pour copier une propriété donnée, on se place dans cellule correspondante et on fait un clic droit et on sélectionne copier. Ensuite, **toujours dans la même colonne**, on sélectionne la plage de cellule dans laquelle on voudrait coller les propriétés (cliquez et faites glisser vers le haut/bas de la colonne). Puis on fait un clic droit et on sélectionne coller

Pour copier toutes les propriétés d'une variable on sélectionne la ligne entière. Après avoir fait un clic droit et choisi *copier* dans le menu contextuel, on sélectionne les lignes où seront collées ces propriétés et ensuite on les colle. Toutes les propriétés à l'exception du nom de la variable seront appliquées aux autres variables (**y compris, l'étiquette de variable**).

#### 3.1.2 Définir des propriétés à partir des Menus

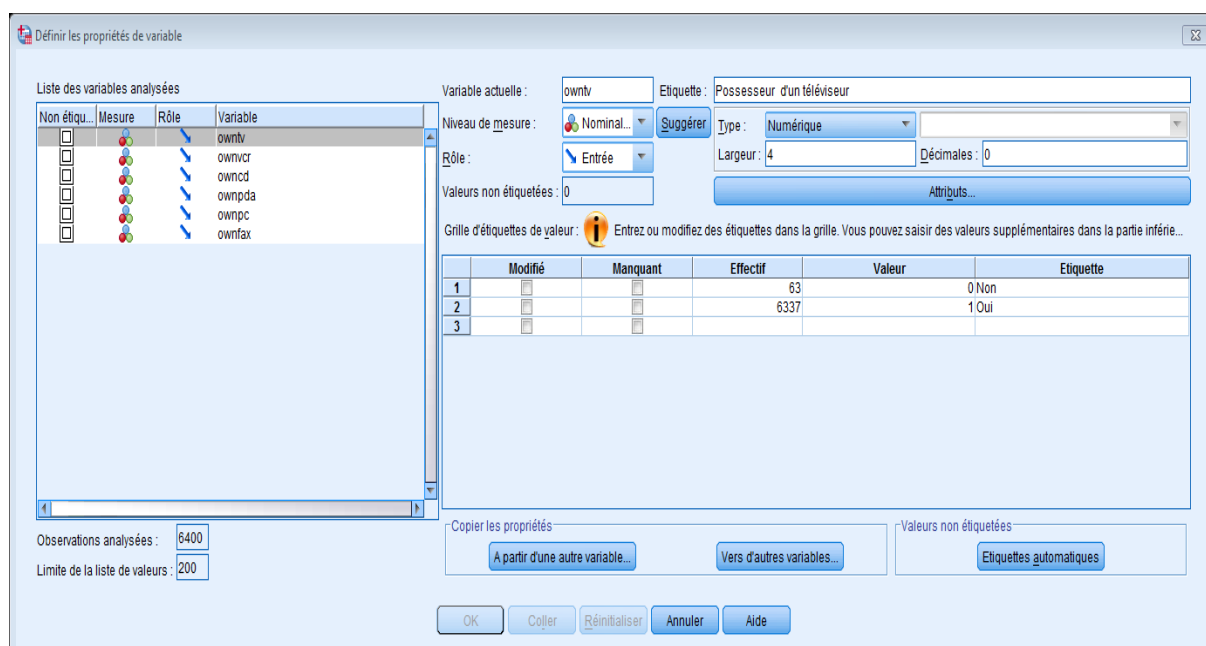
Il est également possible d'utiliser les menus de l'éditeur de données pour définir les propriétés des variables.

- ✓ A partir, des menus de l'Editeur de données, sélectionnez :

**Données>Définir les propriétés de variables**

Cela ouvrira une boîte de dialogue qui permettra de sélectionner les variables pour lesquelles on veut définir les propriétés.

Une fois les variables sélectionnées, s'ouvre alors la fenêtre définir les propriétés d'une variable.



Cette fenêtre contient un panneau avec la liste des variables précédemment sélectionnées. Pour définir ou modifier les propriétés d'une variable de cette liste, il suffit de la sélectionner. Les propriétés de la variable sélectionnée apparaissent dans les différents champs. Il suffit alors de les modifier. On peut ainsi définir l'étiquette de la variable, les étiquettes de valeurs, la mesure et le type de la variable. Pour déclarer, une valeur comme données manquantes, il suffit de cocher case manquant.

La fenêtre **Définir les propriétés de variable** offre également la possibilité de copier les propriétés d'une variable et de les appliquer à d'autres.

Pour cela, dans le cadre *Copier les propriétés*, cliquez sur le bouton *Vers d'autres variables*.

Cela entraîne l'ouverture d'une nouvelle boîte de dialogue. Dans celle-ci, toutes les variables analysées ayant le même type que la variable courante (numérique ou chaîne) sont alors affichées.

Après avoir sélectionné, les variables auxquelles appliquées les propriétés cliquez sur coller.

**Toutefois, les étiquettes de valeurs existantes et les catégories de valeurs manquantes des variables cibles ne sont pas remplacées.** Les étiquettes de valeurs et les catégories de valeurs manquantes des valeurs qui ne sont pas encore définies pour les variables cibles sont ajoutées au groupe d'étiquettes de valeurs et aux catégories de valeurs manquantes des variables cibles

Pour ce qui est des niveaux de mesure des variables cibles, ils sont toujours remplacés.

Ce mode de copie des propriétés présente l'avantage de ne pas copier les étiquettes de variables.

## 3.2 Gestion des données manquantes

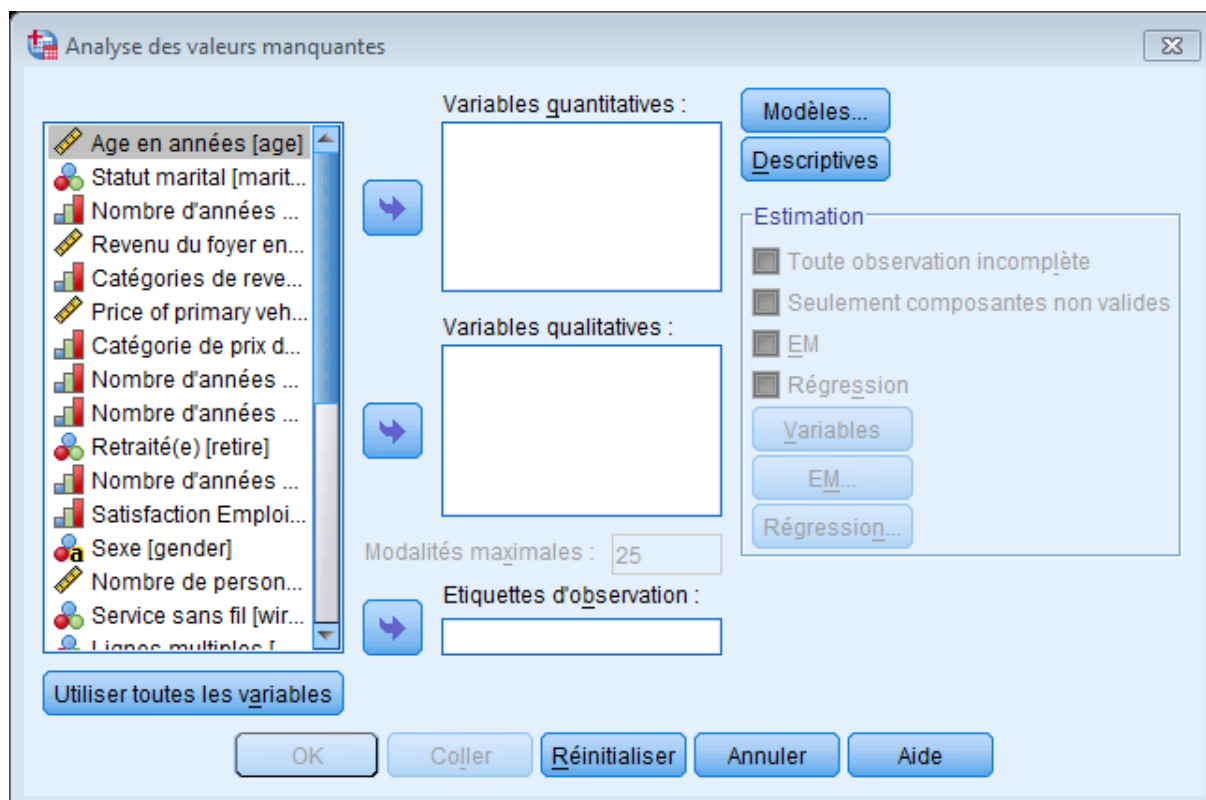
La question du traitement des données manquantes lors des analyses statistiques est une problématique importante. SPSS permet l'utilisation de nombreuses procédures de gestion des données manquantes notamment diverses techniques d'imputation. Toutefois dans le

cadre de ce document, nous ne nous étendrons pas sur l'utilisation de ces techniques. Nous nous limiterons à l'évaluation de l'étendue des données manquantes.

- ✓ Pour cela à partir du menu sélectionnez,

### Analyse>Analyse des valeurs manquantes

Cela entraîne l'ouverture de la boîte de dialogue Analyse des valeurs manquantes, ci-dessous présentée, qui permet sélectionner les variables à analyser.



Pour analyser toutes les variables de la base, cliquer sur le bouton de commande **Utiliser toutes les variables**.

Une fois les variables sélectionnées cliquez sur OK. Le résultat est présenté dans le Viewer sous forme d'un tableau qui donne pour chaque variable, le nombre de données manquantes et d'autres informations supplémentaires.

\*Résultats1 [Document1] - IBM SPSS Statistics Viewer

Fichier Edition Affichage Données Transformer Insérer Format Analyse Marketing direct Graphes Utilitaires Fenêtre Aide

Résultats

- Log
- MVA
- Titre
- Remarques
- Ensemble de don
- Statistiques univa

**Statistiques univariées**

	N	Moyenne	Ecart-type	Manquante		Nombre d'extrema <sup>a</sup>	
				Effectif	Pourcentage	Basse	Haute
age	6400	42,06	12,290	0	,0	0	174
marital	6400	,50	,500	0	,0	0	0
address	6400	11,56	9,938	0	,0	0	311
income	6400	69,4748	78,71856	0	,0	0	270
inccat	6400	2,53	1,074	0	,0	0	0
car	6400	30,1284	21,92692	0	,0	0	481



## 4 Analyse univariée

Il s'agit dans cette section de présenter des outils qui peuvent être utilisés pour décrire et analyser des variables prises de manière individuelle.

### 4.1 Variable qualitative

Généralement la description d'un variable qualitative s'appuie sur des mesures de tendance centrale telle que le mode, la réalisation de tris à plats et les représentations graphiques.

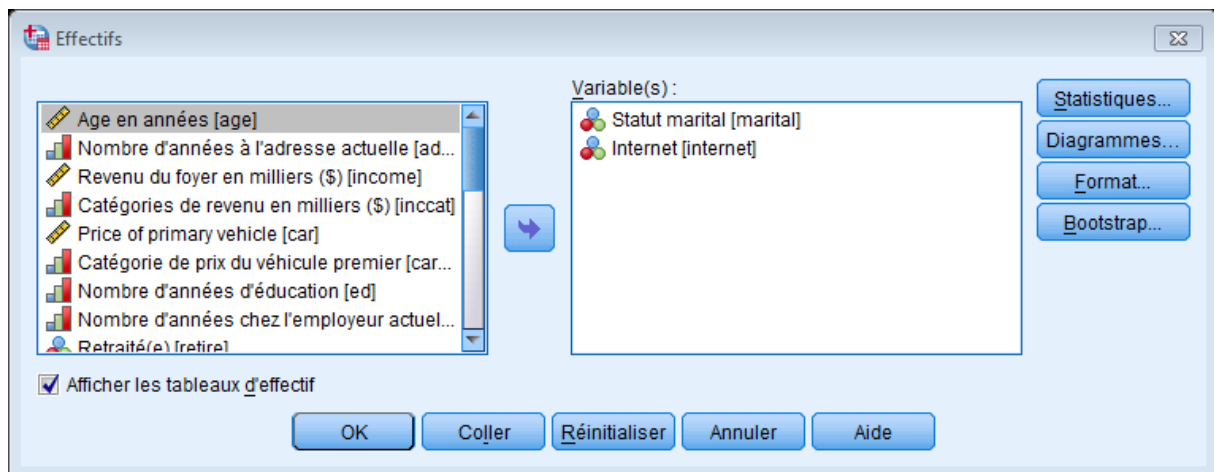
#### 4.1.1 Tri à plat

Effectuer un tri à plat pour une variable qualitative consiste à réaliser un tableau présentant les effectifs (ou les fréquences) pour chacune de ses modalités.

- ✓ A partir du menu, sélectionnez :

**Analyse > Statistiques descriptives > Effectifs**

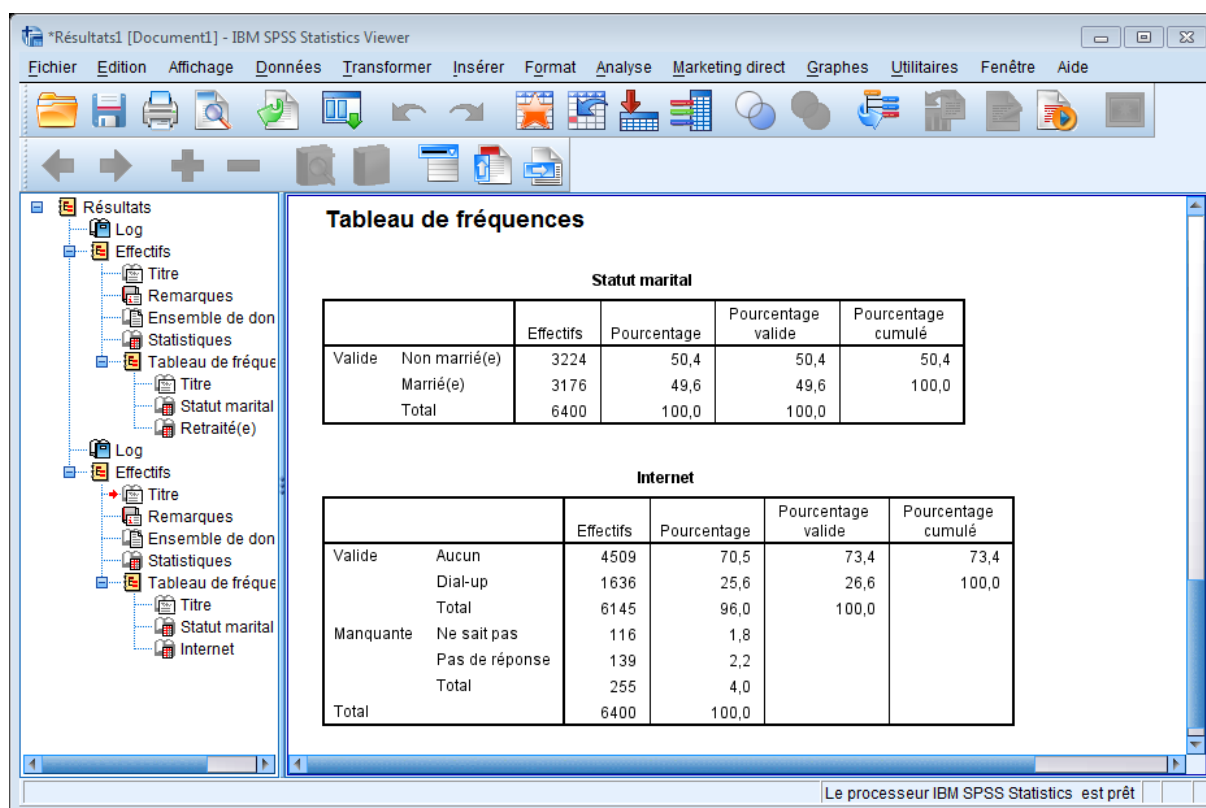
Cela entrainera l'ouverture d'une boîte de dialogue permettant de choisir les variables qu'on l'on désire étudier.



Il est important de s'assurer que la case *Afficher les tableaux d'effectif* est coché.

Par ailleurs, l'on note qu'il est possible de sélectionner plus d'une variable. Dans ce cas, un tri à plat sera réalisé pour chacune de ces variables.





Quand le tableau des effectifs est réalisé, celui-ci contient deux champs relatifs aux fréquences : « pourcentage » et « pourcentage valide ». En effet, pour la colonne « pourcentage » le calcul est effectué en considérant l'ensemble des individus, par contre pour « pourcentage valide » les individus ayant des données manquantes ne sont pas pris en compte.

#### 4.1.2 Diagrammes pour données qualitatives

Pour l'analyse d'une variable qualitative, l'on rencontre généralement deux types de graphiques : les diagrammes en secteurs (camemberts) et les diagrammes en bâton. Toutefois dans la pratique, on a tendance à apprécier ces derniers par rapport aux précédents, surtout dans les cas où la variable possède de nombreuses modalités. En effet, il se pourrait qu'il soit plus facile pour l'œil humain d'évaluer des longueurs que des surfaces. Ainsi, il est de plus en plus conseillé d'utiliser des diagrammes en bâton plutôt que des diagrammes en secteurs. Les camemberts sont plus indiqués dans les cas où l'on cherche à comparer chaque modalité à l'ensemble de la population.

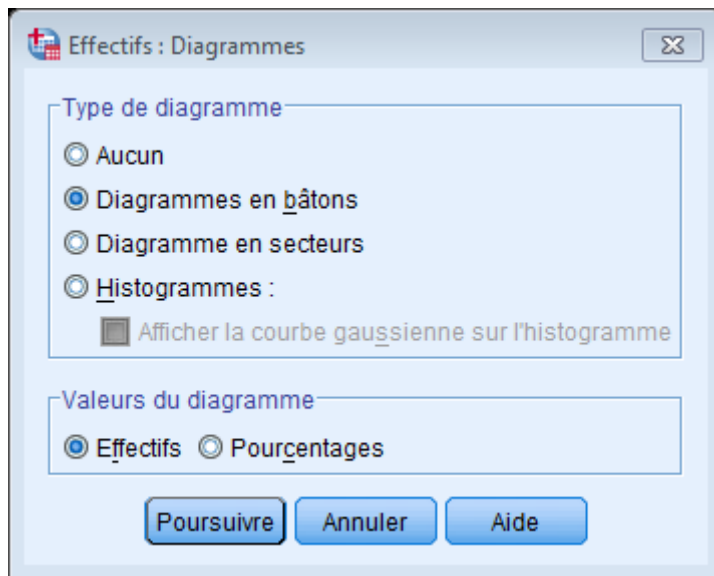
Les procédures pour la réalisation d'un diagramme en secteurs ou en bâtons, sont à peu près similaires.

- ✓ Tout comme pour le tri à plat, à partir du menu, sélectionnez :

**Analyse > Statistiques descriptives > Effectifs**

- ✓ Lorsque la fenêtre précédemment évoquée s'ouvre, cliquez sur diagramme après avoir préalablement sélectionné la (ou les) variable(s).

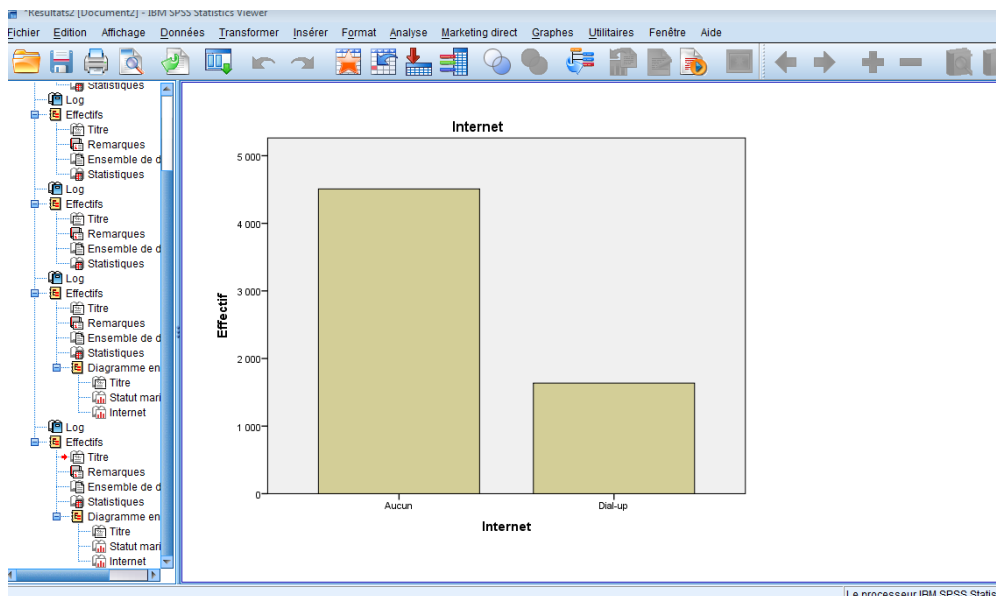
La fenêtre qui s'ouvre ensuite permet d'indiquer le graphique que l'on veut réaliser :



- ✓ Cliquez sur poursuivre, après avoir sélectionné le type de graphique

Pensez à décocher la case **Afficher les tableaux d'effectif** si vous ne voulez construire le tableau des effectifs.

Une fois la procédure achevée, le résultat apparaît dans le Viewer (fenêtre des résultats)



## 4.2 Variable quantitative

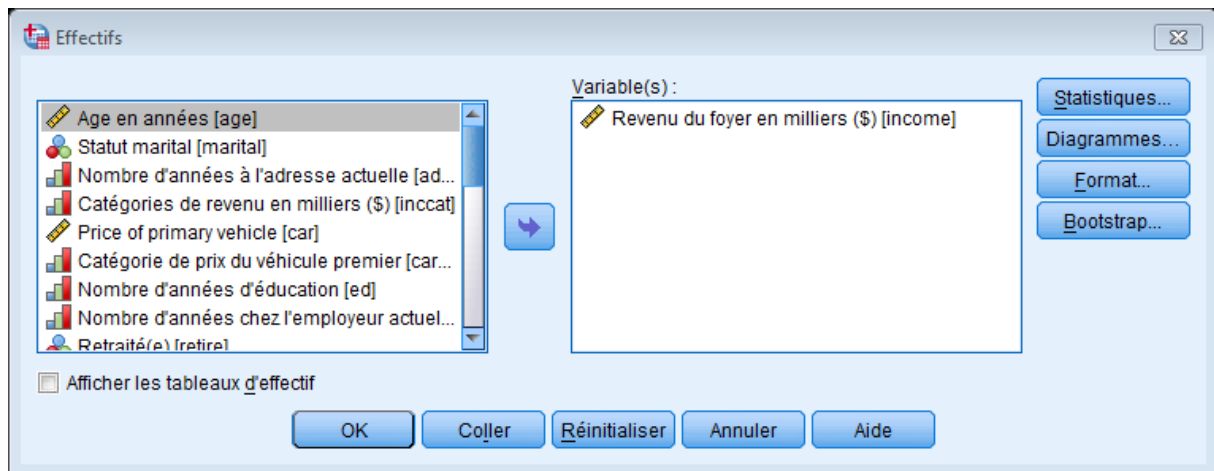
La description de variables quantitatives s'appuie généralement sur les mesures de tendances centrales (moyenne, médiane, mode), les mesures de dispersion (écart-type, variance, écart inter quartile, étendue), les fractiles et les graphiques.

### 4.2.1 Description à partir des statistiques calculées

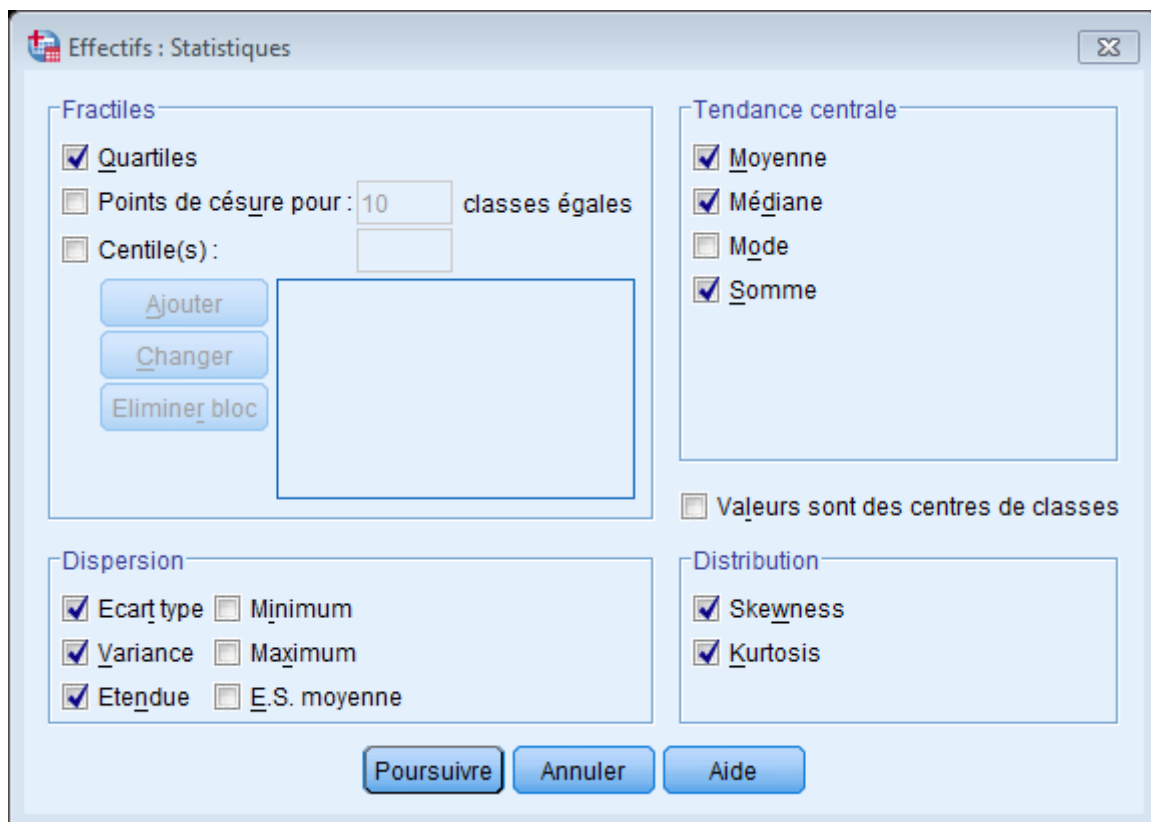
- ✓ À partir du menu, sélectionnez :

**Analyse > Statistiques descriptives > Effectifs**

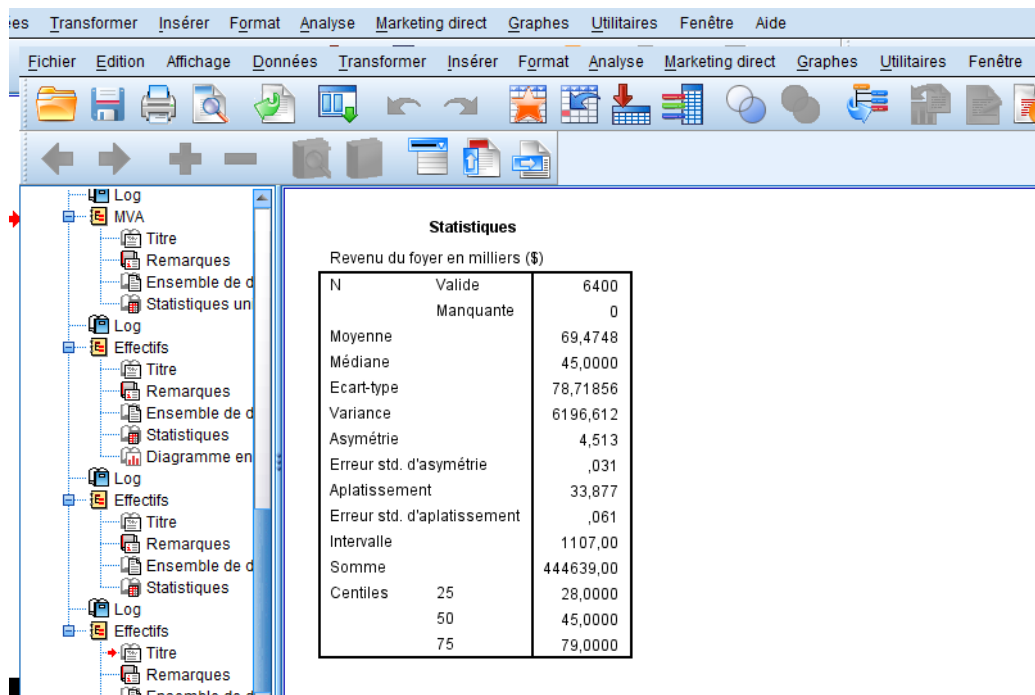
La boîte de dialogue Fréquences s'ouvrira pour permettre de spécifier les variables pour lesquelles les statistiques doivent être calculées.



- ✓ Assurez-vous que la case **Afficher les tableaux d'effectif**, dans le coin inférieur gauche est bel et bien désélectionnée
- ✓ Cliquez ensuite sur le bouton de commande **Statistiques**.



Une fois toutes les étapes validées, les résultats sont produits sous la forme d'un tableau dans le Viewer.



#### 4.2.2 Description à partir des graphiques

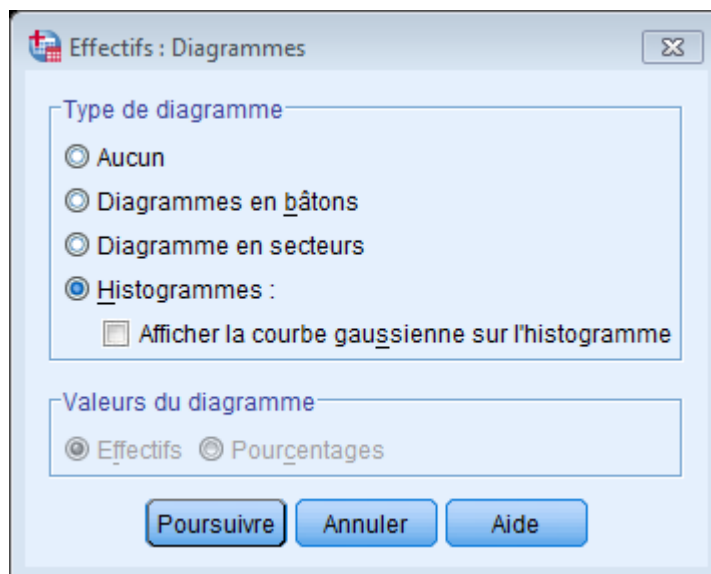
L'histogramme est un outil très couramment employé pour la description d'une variable numérique.

- ✓ Tout comme pour le calcul des statistiques, à partir du menu la boîte de dialogue Fréquences, pour indiquer la (les) variable(s) à analyser.

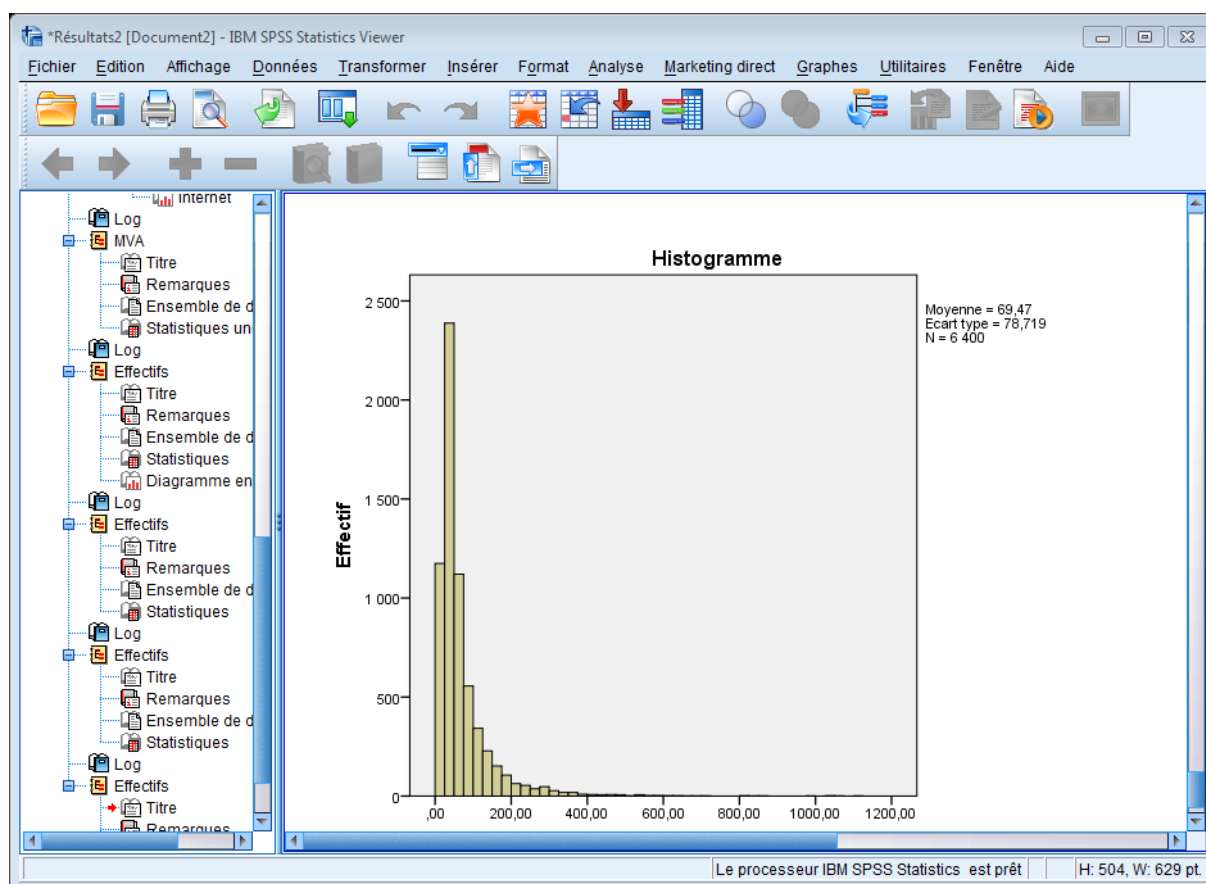
##### Analyse > Statistiques descriptives > Effectifs

- ✓ cliquez sur le bouton diagramme de la fenêtre *Effectifs*.

La fenêtre qui s'ouvre ensuite permet d'indiquer le graphique que l'on veut réaliser :



Une fois réalisé le graphique est présenté dans la fenêtre des résultats.



## 5 Les graphiques dans SPSS

SPSS offre aux utilisateurs la possibilité de pouvoir élaborer une gamme très variée de graphiques. L'ensemble de ces graphiques est accessible via le menu **Graphique** de l'éditeur de données. Les graphiques sont groupés par types de graphique, chaque type intégrant aussi bien les graphes pour les analyses univariées que ceux pour celles dites multivariées. Comme principaux types on peut citer :

- Diagrammes en bâtons
- Diagrammes circulaires/polaires : camembert
- Diagrammes curvilignes : pour la construction de courbes notamment les séries temporelles
- Diagrammes en aires : pour les représentations des surfaces
- Diagrammes de dispersion et diagrammes de points : pour les constructions des nuages de points.
- Boîtes à moustaches : intégrant les boxplots « simples » et les boxplots « juxtaposés ».

Les graphiques sont généralement élaborés par l'intermédiaire du Générateur de diagrammes accessible via le menu Graphique.

La prise en main du *générateur de diagramme* est assez naturelle. Ce faisant, cette section s'attardera sur nombre restreint d'exemples, étant donné que les principes sont pratiquement les mêmes pour les différents types de graphiques.

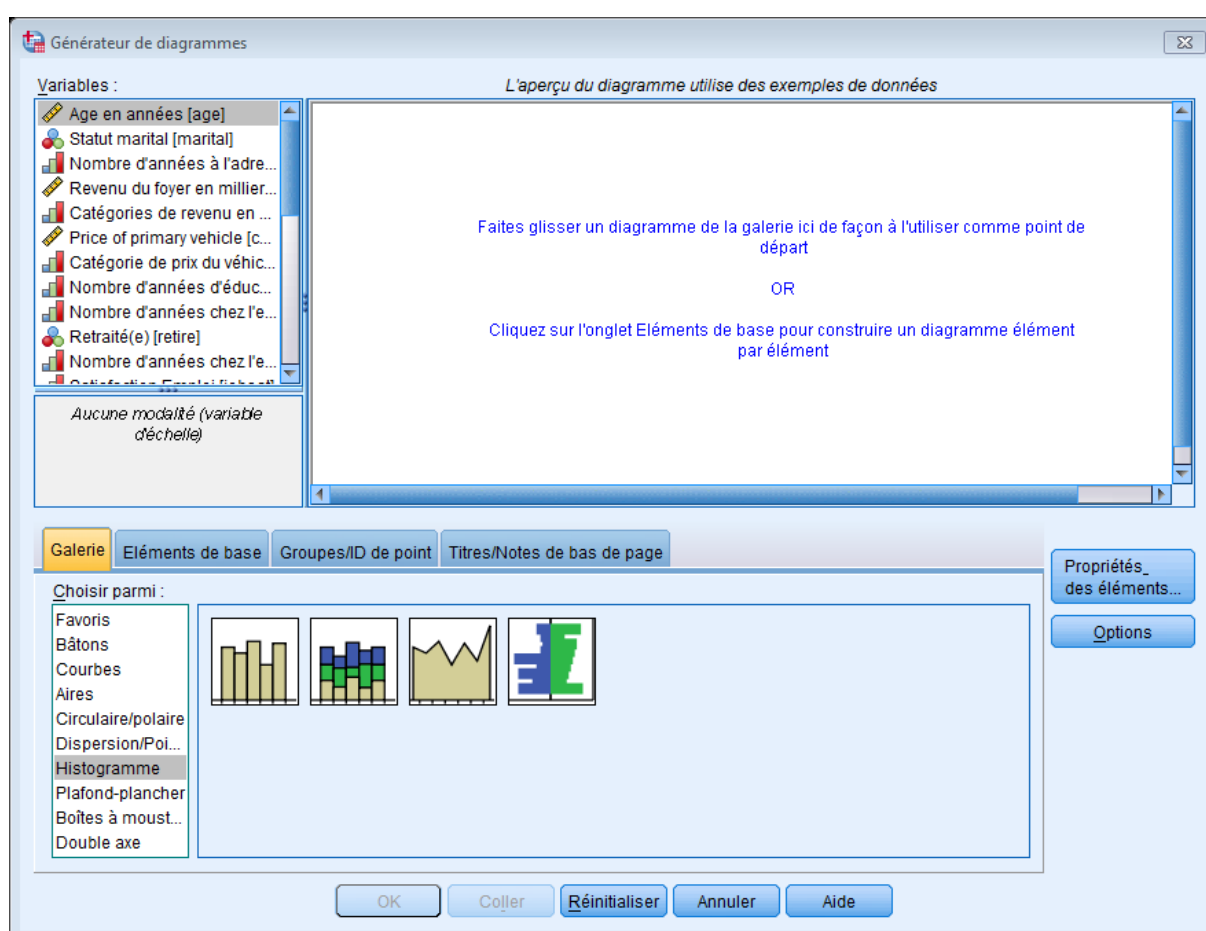
## 5.1 Construction des graphiques

La procédure pour la construction d'un graphique est la suivante :

- ✓ A partir des menus de l'éditeur de données, sélectionnez :

**Graphiques > Générateur de diagrammes**

Cela entrainera l'ouverture du *Générateur de diagrammes*. Ce dernier, présenté ci-contre, est une fenêtre permettant de construire de manière interactive les diagrammes.

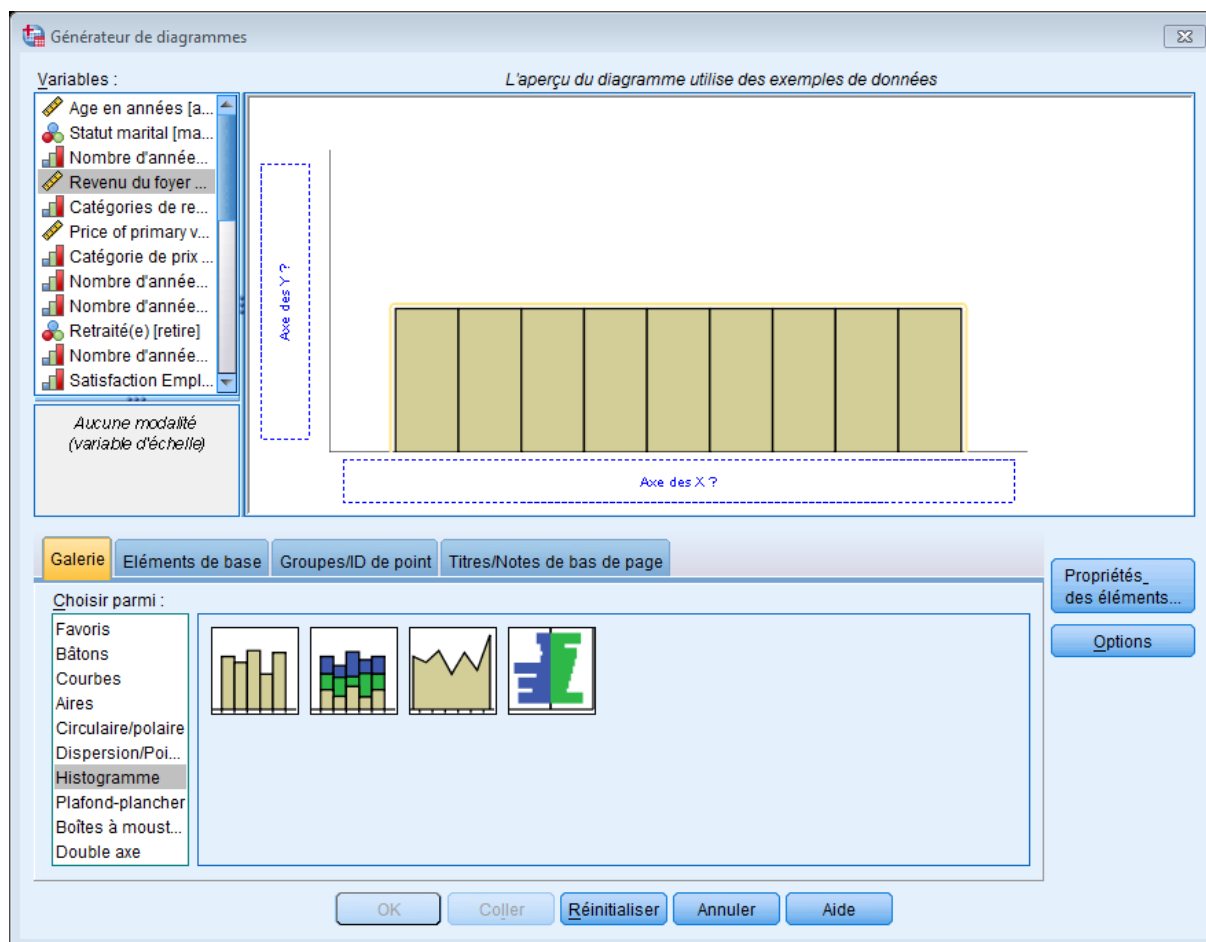


L'onglet galerie présente la panoplie de diagrammes prédéfinis et regroupés par type, mise à la disposition de l'utilisateur.

L'onglet **Eléments de base** permet la création des graphiques ex-nihilo avec notamment la construction des axes ou encore le choix des éléments de base du graphique. Ceci étant il sera extrêmement rare de recourir à cette procédure car l'éventail de diagrammes prédéfinis répond à l'essentiel des problématiques auxquels sera confronté l'utilisateur.

Une fois le graphe à construire identifié dans la galerie, l'utilisateur doit faire glisser l'icône de ce dernier vers le canevas permettant d'avoir un aperçu du graphique en cours d'élaboration.

A titre d'illustration, nous utiliserons une nouvelle fois la base demo.sav et nous construirons l'histogramme de la variable « income » (relative au revenu du foyer) en utilisant les fréquences et non les effectifs.



Après le choix du type de graphes, afin d'obtenir un graphique complet, il reste encore à indiquer la (les) variable(s) à utiliser et la statistique à calculer.

### Choix des variables

Pour indiquer les variables impliquées dans la construction du graphique, il suffit de faire glisser chacune de celles-ci de la liste des variables et de la déposer dans le champ concerné. Par exemple la variable devant figurer en abscisse doit être déposée dans le champ « axe de X ».

Aussi, il est nécessaire lors de la sélection des variables que leurs mesures (« nominal », « ordinal » ou « échelle ») soient convenablement définies. Par exemple, il ne sera pas possible de construire correctement un histogramme avec une variable ordinale ou nominale.

Toutefois, dans le cas où la mesure d'une variable n'est pas appropriée pour le rôle qu'elle aura à jouer dans la construction du graphe, il est possible de modifier **temporairement** le niveau de mesure dans le Générateur de diagrammes, plutôt que de revenir le modifier dans

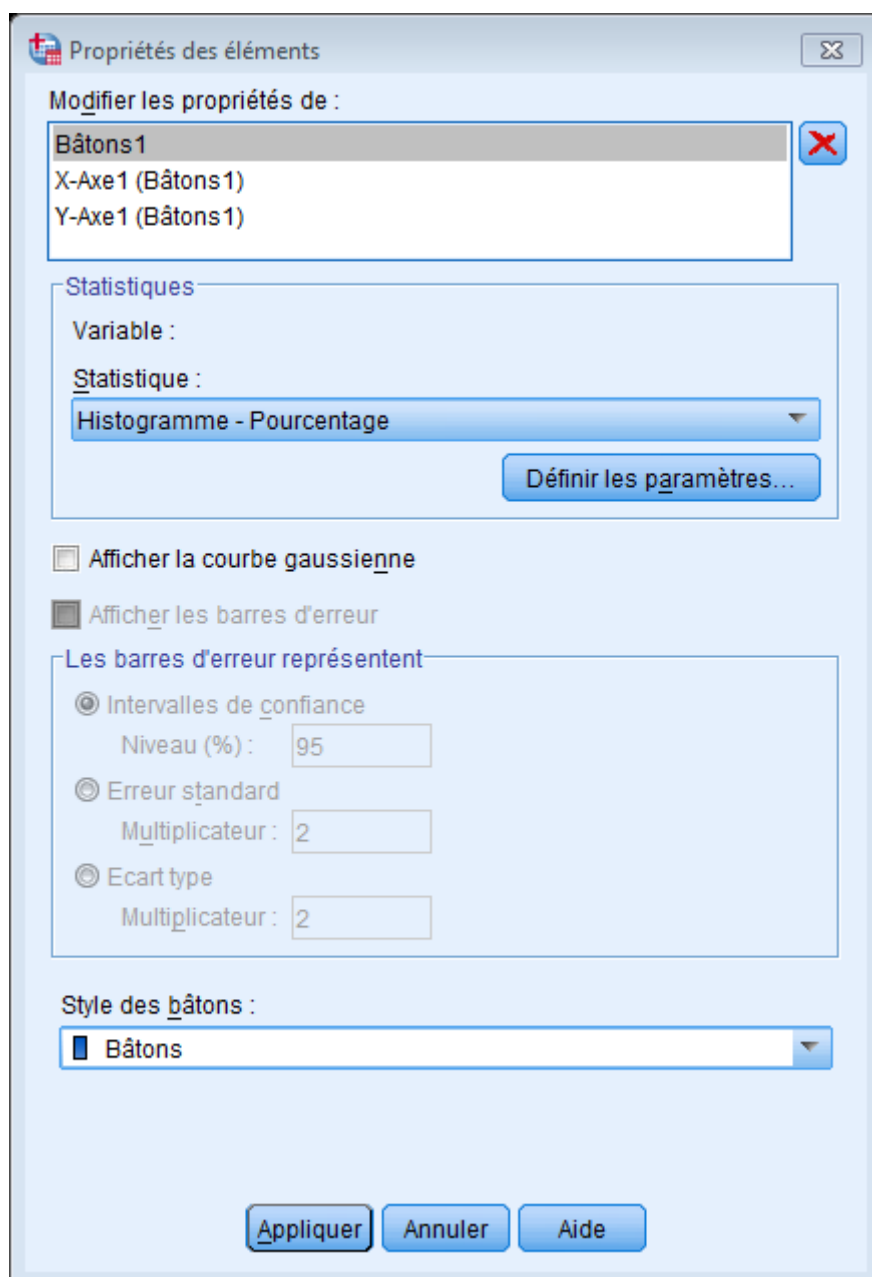
l’Affichage des variables. Pour ce faire, il suffit de faire un clic droit sur la variable concernée dans la liste des variables du Générateur de diagrammes et de modifier sa mesure. Une fois effectué, la variable peut alors être sélectionnée pour la construction du graphique.

### Définition des statistiques à calculer

Il arrive que l’on veuille lors de la construction des graphiques utilisées d’autres statistiques que celles proposées par défaut. Par exemple, pour la construction d’un histogramme, l’utilisateur peut souhaiter utiliser les pourcentages alors que la statistique par défaut c’est «les Effectifs ». On peut également vouloir un diagramme en construire un diagramme en bâton portant sur la variable sexe où la longueur des bâtons ne serait pas relatifs aux effectifs dans chaque classe mais plutôt au revenu moyen (ou même total) dans chacune de ces dernières. Ainsi, la définition des statistiques permet d’indiquer la nature précise de l’information à ressortir.

La définition des statistiques se fait par l’intermédiaire de la fenêtre Propriété des éléments. Elle peut s’ouvrir automatiquement, après la réalisation du choix de graphique ou manuellement en cliquant sur le bouton de commande **Propriété des éléments** du générateur de diagrammes.

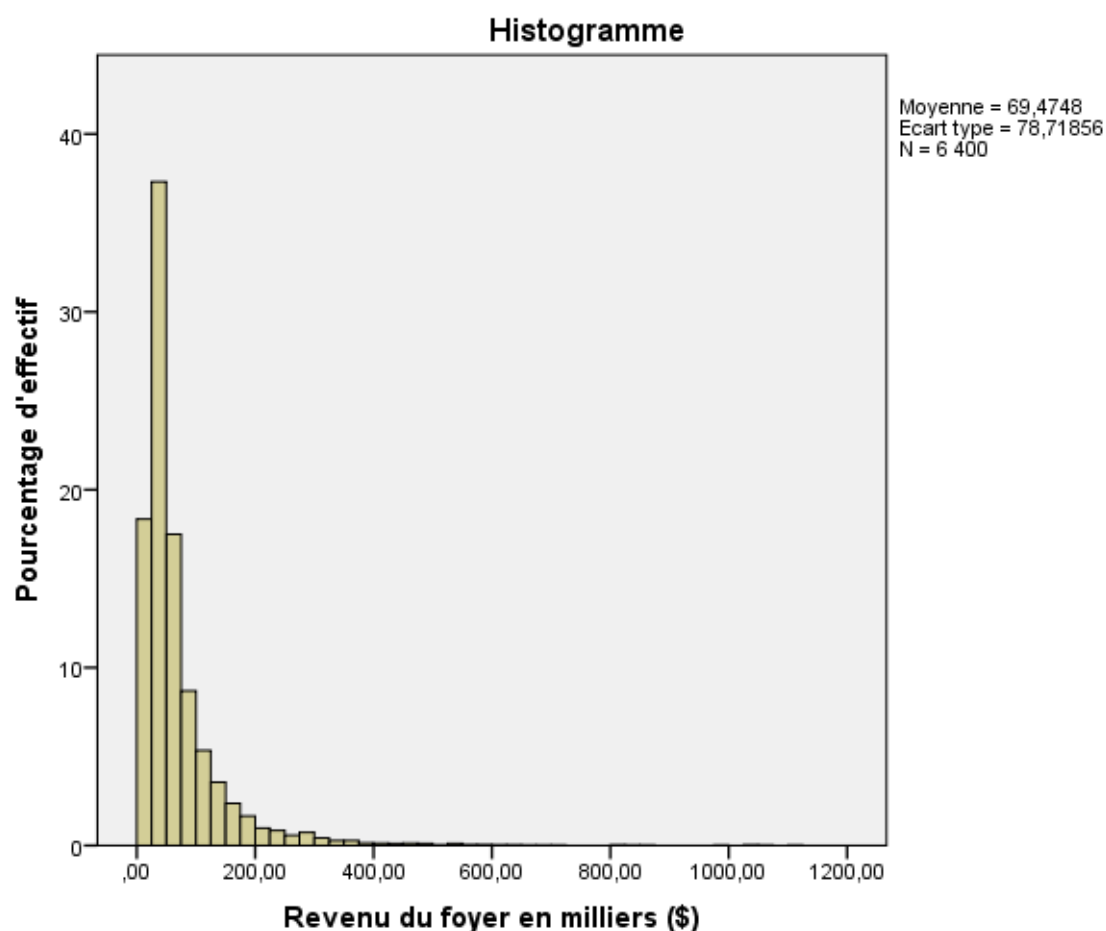




Cette fenêtre permet d'afficher et de modifier les propriétés des éléments du graphique. Elle permet également de supprimer ces éléments. Le premier cadre de cette fenêtre « *modifier les propriétés de :* » présente la liste des éléments du graphique. Les propriétés de l'élément sélectionné dans cette liste sont affichées plus bas dans la fenêtre et peuvent ainsi être modifiées.

Dans la figure ci-dessus « Bâtons1 » est sélectionné, l'on note qu'on a un cadre « Statistique » qui permet de définir la statistique à calculer. Pour avoir l'histogramme avec les pourcentages et non les effectifs l'on a sélectionné « Histogramme – Pourcentage ».

En sélectionnant « X-axe 1 », on a la possibilité de définir les étiquettes pour l'axe des abscisses dans le champ Etiquette de l'axe.



### Informations supplémentaires

L'onglet **Les titres/Note de bas de page** permettent d'ajouter des titres, sous-titres et note de bas de page au graphique. Une fois ajoutés, ceux-ci deviennent aussi des éléments du graphique dont les propriétés peuvent être modifiées dans la fenêtre propriété des éléments.

L'onglet **Groupes/ID point** permet l'utilisation du variable de panel dans la construction des graphiques. Par exemple, l'on peut être amené à comparer la distribution des revenus chez les hommes à celle des revenus chez les femmes. Pour cela, on peut faire construction l'histogramme des revenus pour chacune des classes. Au lieu de réaliser deux procédures distinctes, l'utilisation de la variable sexe comme variable de panel permettra d'avoir une figure constituée de deux graphes juxtaposés présentant chacun la distribution des revenus dans l'une des classes.

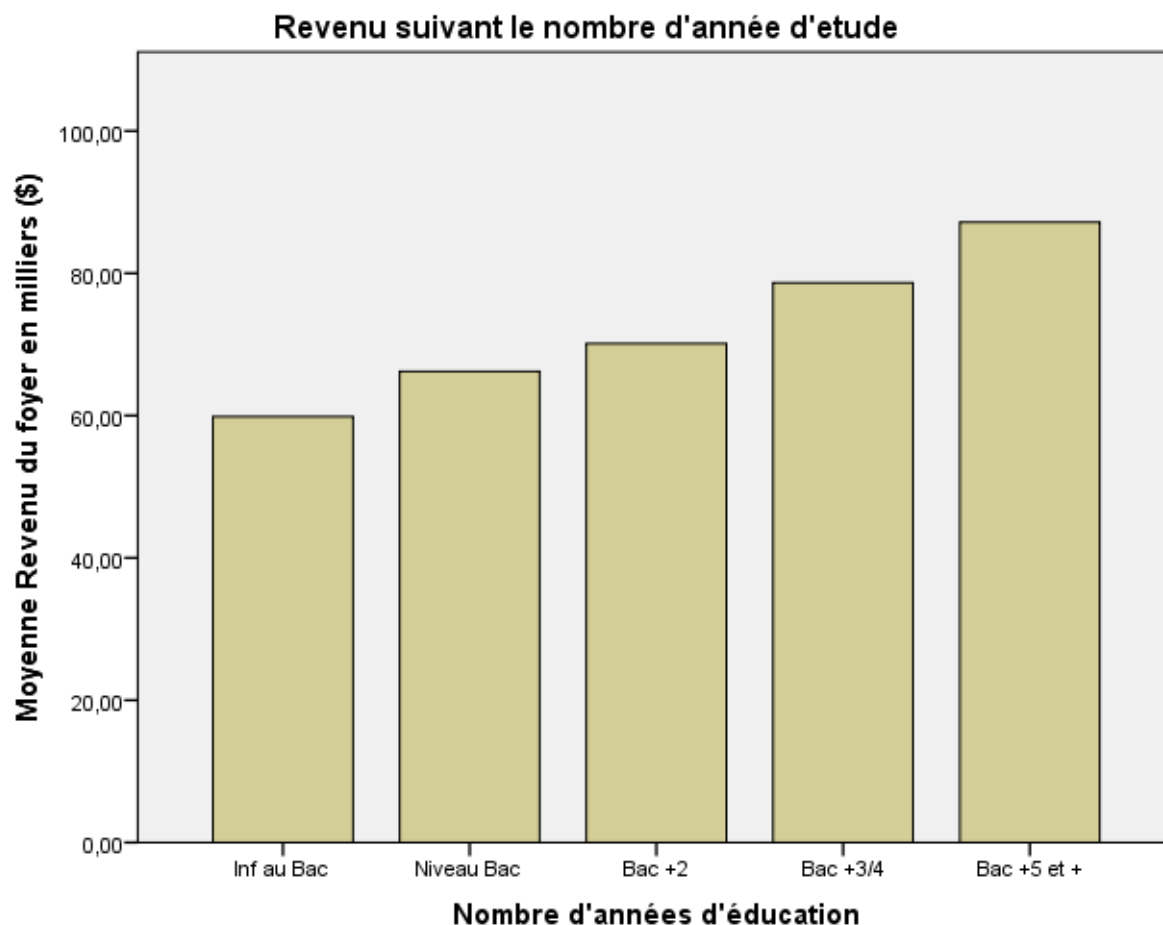
## 5.2 Modification des graphiques

Lors de la phase exploratoire des données, les graphiques réalisés ont coutume d'être très simplifiés. Une fois cette phase achevée, les graphes retenus pour figurer dans le rapport final se doivent d'être assez parlant et d'avoir un minimum d'esthétique.

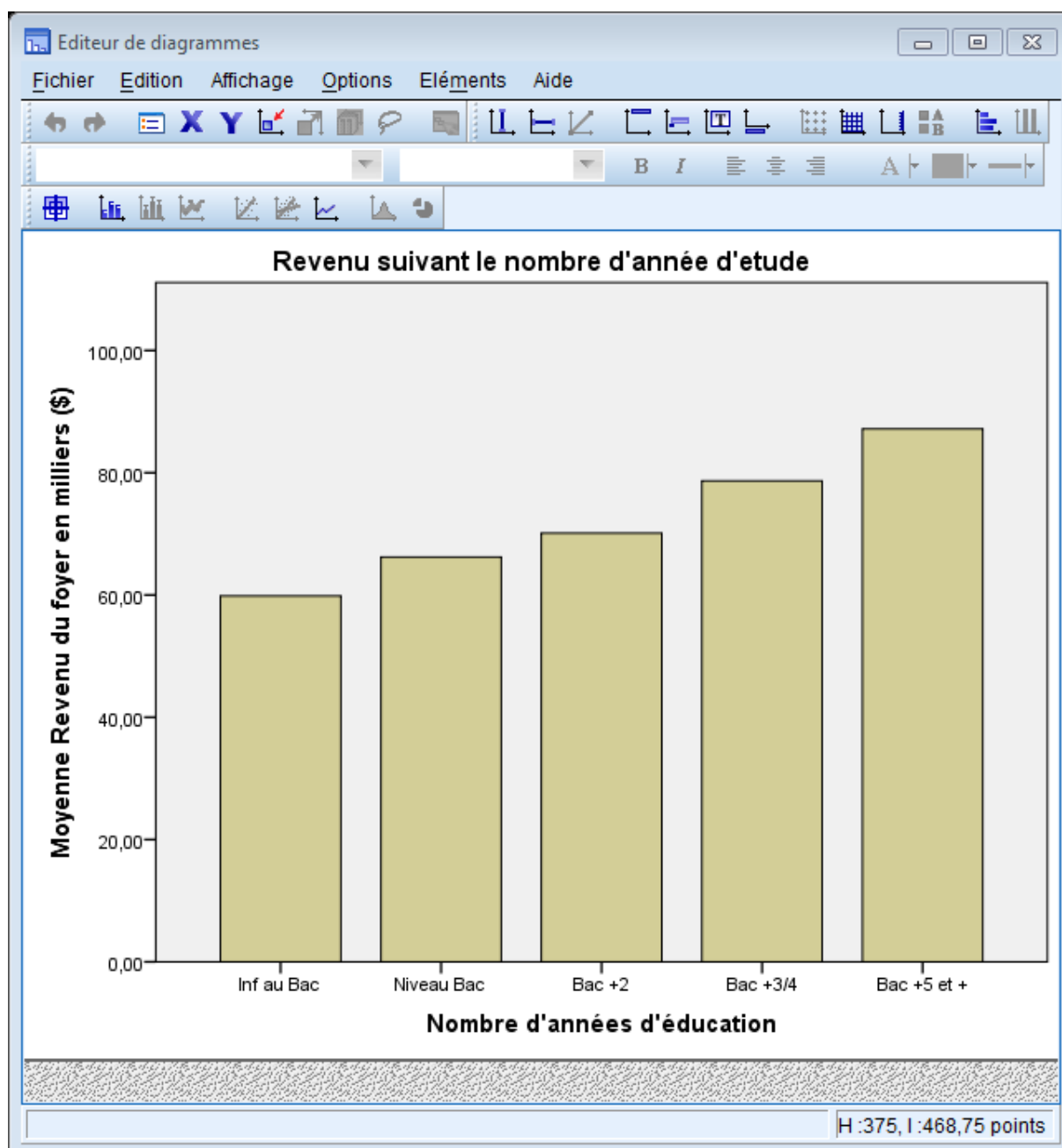
La modification permet de personnaliser un grand nombre éléments du graphique notamment les couleurs utilisés, l'affichage et le contenu des libellés des étiquettes, le format graduation des axes, etc.

Dans cette section, à partir du fichier *demo.sav*, nous allons construire un diagramme en bâton présentant le revenu le moyen (income) suivant le niveau d'éducation (ed). Ce graphique servira de cadre pour les illustrations.

Par défaut, le graphique présenté ci-dessous est affiché tel quel dans la fenêtre des résultats.



La modification des diagrammes se fera à partir de l'éditeur de diagramme accessible par la fenêtre Viewer. Pour ouvrir cet éditeur, il suffit de double cliquer sur le graphique affiché dans la fenêtre des résultats.



L'éditeur de diagramme permet de sélectionner les groupes d'éléments ou les éléments individuels afin d'y apporter des modifications. Dans le cas présent, un groupe d'éléments peut par exemple être l'ensemble des barres du diagramme en bâton.

Pour sélectionner un groupe d'éléments, il suffit de cliquer sur un élément du groupe. Pour sélectionner un élément du groupe, il faut après avoir préalablement sélectionné le groupe, cliquer sur celui-ci.

Pour désélectionner tous les éléments, appuyez sur la touche **Echap** du clavier.

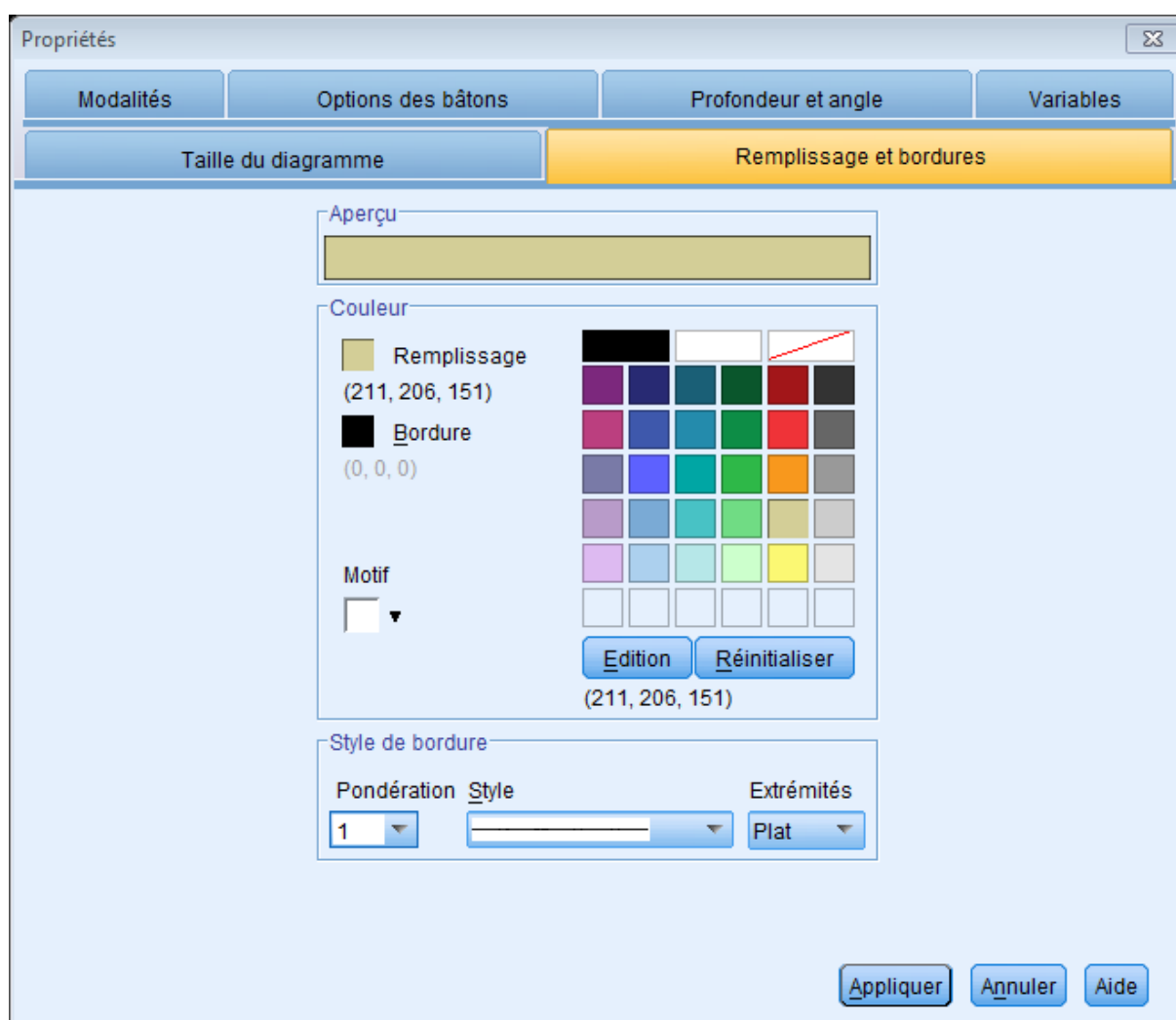
Modifier un élément ou un groupe d'éléments consistera généralement à modifier ses propriétés. Pour avoir accès aux propriétés d'un groupe d'éléments, après avoir sélectionné le groupe, à partir du menu de l'éditeur de graphique faites :

**Edition>Propriétés.**

Cela ouvre la fenêtre propriétés avec des onglets permettant d'accéder à différents types de propriétés.

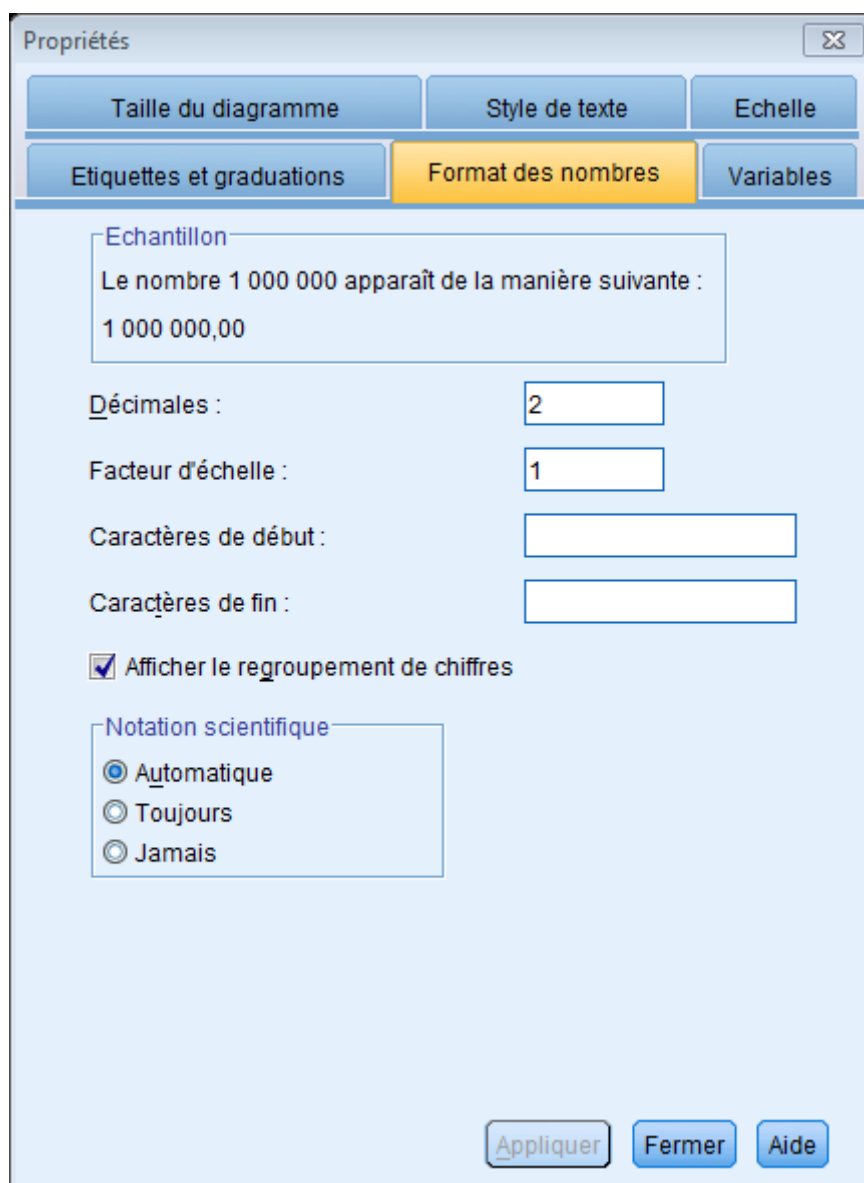
**Modifications des couleurs des barres**

Dans le cas d'exemple, si l'on veut modifier la couleur des bâtons, on ouvre la fenêtre propriété et l'on sélectionne ensuite l'onglet **Remplissage et bordures**. On peut alors modifier la couleur de remplissage.

**Modifications des formats d'affichage des graduations**

Il est possible modifier le format numérique des étiquettes de graduation sur les axes et de définir des échelles pour ceux-ci.

Dans notre exemple, on peut sélectionner les graduations sur l'axe des ordonnées et modifier, via l'onglet **format de nombre** de la fenêtre **Propriétés**, les propriétés telles que le nombre de décimal après la virgule, l'échelle. Dans le cas d'espèce, les revenus sont en milliers, on peut donc utiliser une échelle de 0,001 pour que les graduations soient affichées en valeur réelle.



### Affichage des étiquettes

L'éditeur de graphique offre la possibilité d'afficher les étiquettes de valeurs de données dans les graphiques. Cela offre permet à l'observation du graphique d'avoir des informations précises sur les différentes mesures.

Dans le cas de notre diagramme en bâton, pour afficher les étiquettes donnant la hauteur des bâtons :

- ✓ Sélectionnez les bâtons
- ✓ Ensuite à partir du menu de l'éditeur de graphique sélectionner :

**Eléments>Afficher les étiquettes de données.**

### Modification des titres et des étiquettes des axes

Les étiquettes des axes ainsi que les titres et sous titres des graphiques peuvent être également modifiés. Pour cela, il suffit de double-cliquer sur l'étiquette et de saisir les nouvelles étiquettes.

### Utilisation des Modèles

SPSS fournit une fonctionnalité très pratique dans la mise en forme des graphiques, à savoir l'utilisation des modèles. Celle-ci peut s'avérer utile lorsqu'on doit appliquer le même type de modifications à plusieurs graphiques. En effet, les différents attributs d'un graphique peuvent être enregistrés dans un package appelé modèle. Ces modèles sont des fichiers d'extension « .sgt ». Ils peuvent être appliqués par la suite, à de nouveaux graphiques qui présenteront les mêmes attributs que ceux se trouvant dans le modèle.

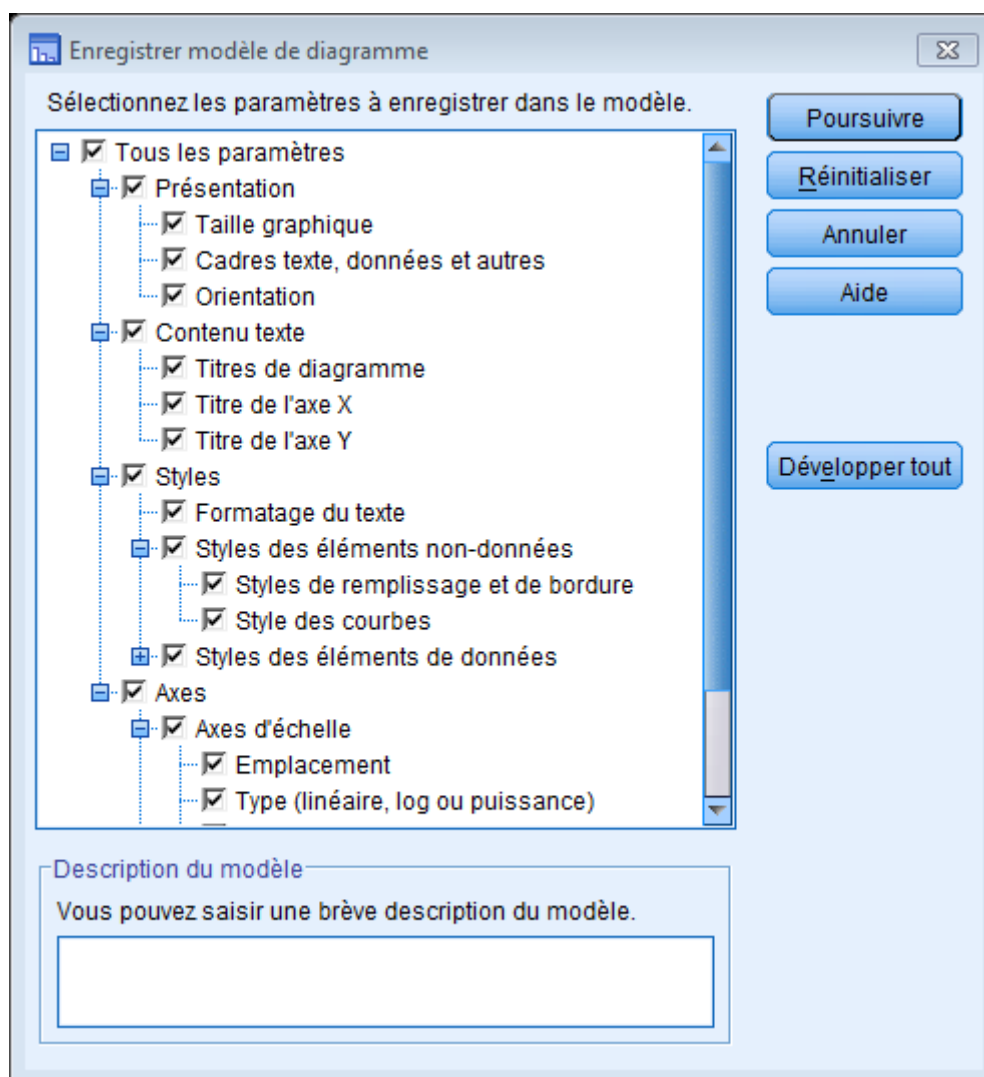
### Enregistrement d'un modèle

Pour enregistrer un modèle, à partir du menu de l'**Editeur de graphique** sélectionnez :

**Fichier > Enregistrer modèle de diagramme**

La fenêtre **Enregistrer modèle de diagramme** qui s'ouvre ensuite présente une arborescence permettant de sélectionner les différents attributs du graphique en cours qui seront intégrés au modèle.

L'on peut choisir de sélectionner tous les attributs. Toutefois, cela intègre même les attributs tels que les libellés des axes.



Une fois les attributs sélectionnés, l'on peut saisir une description du modèle dans le champ situé en bas dans la fenêtre.

Après avoir cliqué sur le bouton de commande **Poursuivre**, il ne reste plus qu'à spécifier le nom et l'emplacement du fichier.

### Application d'un modèle à un nouveau graphique

Pour appliquer un modèle à un nouveau graphique, lors de sa construction on peut cliquer sur le bouton de commande **option** du **Générateur de graphique**. La fenêtre qui s'ouvre ensuite présente un cadre relatif au modèle. Il suffit alors de cliquer sur le bouton **ajouter** et d'indiquer les modèles que l'on souhaite utilisé.

Il est également possible d'appliquer un modèle via l'**Editeur de graphique**.

A partir du menu, sélectionnez

**Fichier > Appliquer modèle de diagramme**

## 5.3 Les options de Diagramme

On peut modifier les attributs utilisés par défaut lors de la construction des graphiques. Ainsi, tous les graphes construits intégreront par défaut, ces nouveaux attributs. De plus, ces



attributs pourront être également modifiés lors de la construction d'un graphique spécifique.

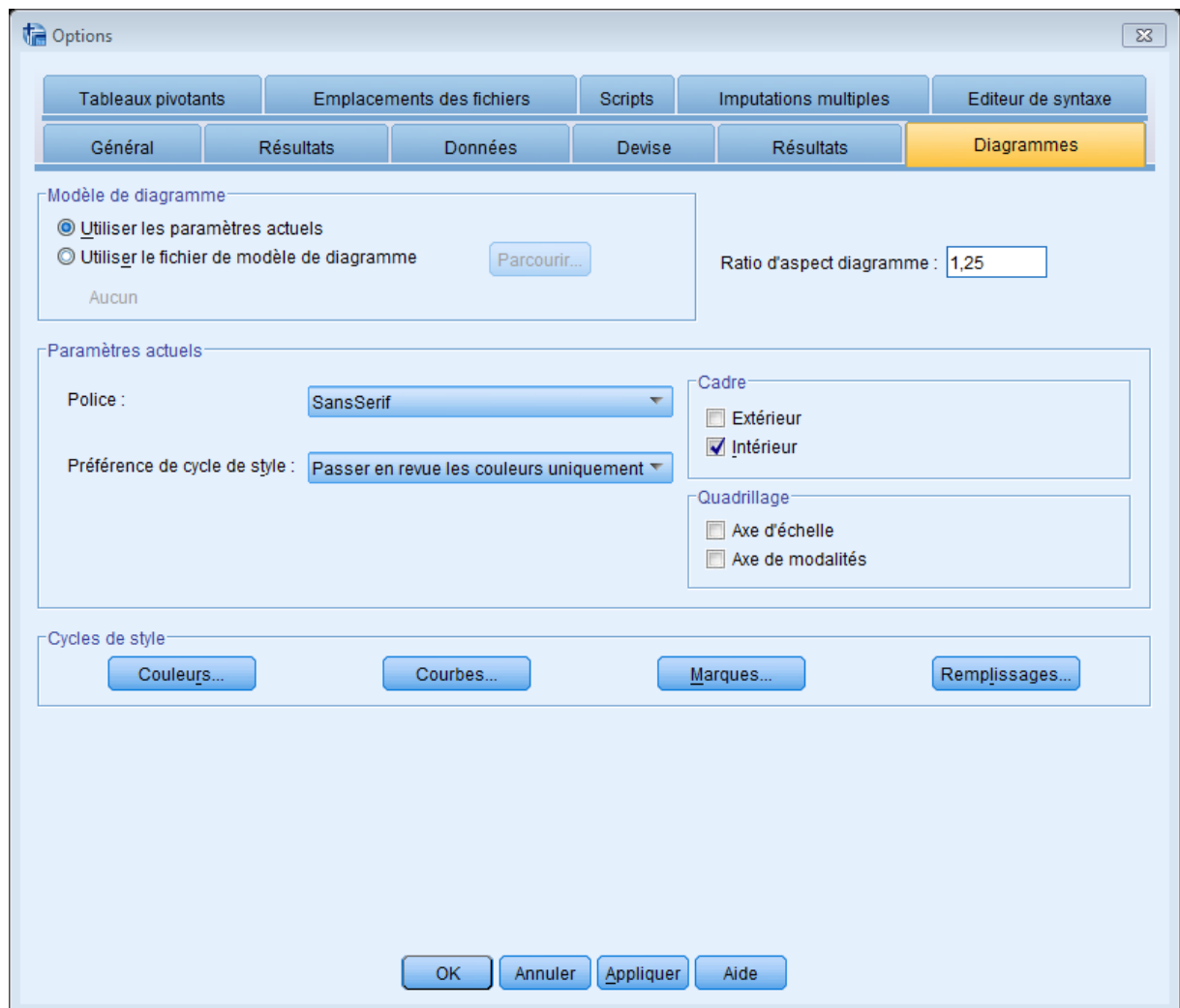
Pour modifier les options par défaut des diagrammes :

- ✓ A partir du menu de l'**Editeur de données** ou du **Viewer**, sélectionnez :

**Edition > Options**

La fenêtre qui s'ouvre permet de configurer de nombreux éléments.

- ✓ Cliquez sur l'onglet Diagrammes pour paramétrer les options relatifs aux graphiques.



## 6 Manipulation des données

### 6.1 Trier les données

Lors des analyses, il s'avère parfois utile de réorganiser les observations (de façon croissante ou décroissante) en fonction des valeurs prises par une ou plusieurs variables. On parle alors de tri. Lorsqu'il y a plusieurs variables de tri, un ordre de priorité est attribué aux variables. C'est le même principe lorsqu'on classe par exemple, par ordre alphabétique, une

liste comprenant les noms et les prénoms des individus. Le champ « nom » est utilisé en priorité en suite, en cas d'égalité, on se réfère au champ « prénom ».

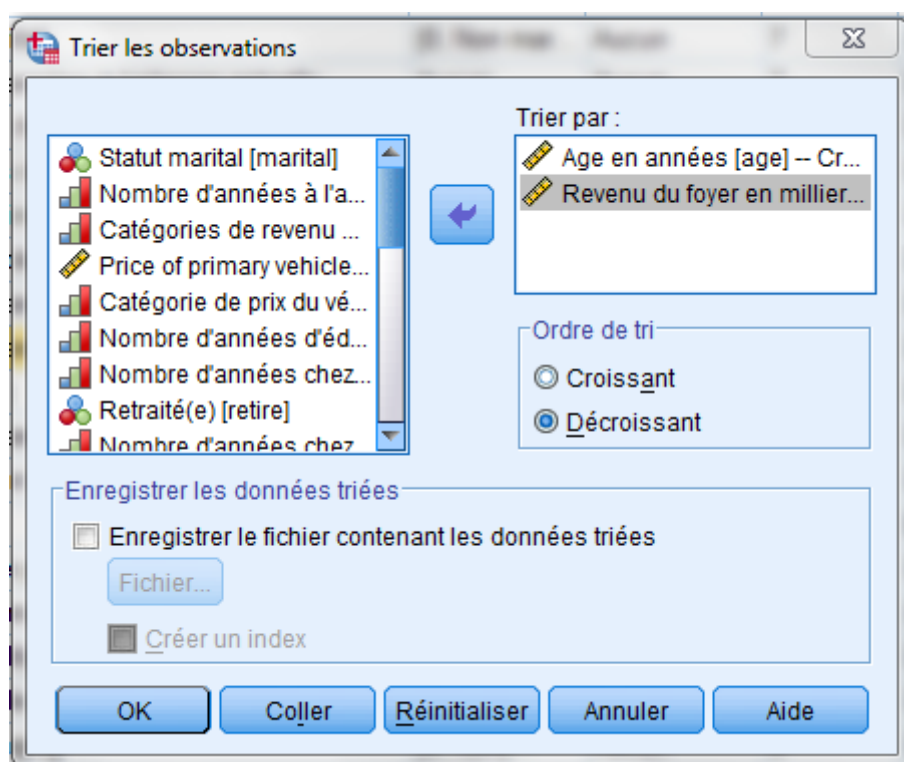
Pour trier les données, la démarche se présente comme suit :

- ✓ A partir de la barre des menus de la fenêtre Editeur de Données, exécuter :

#### Données > Trier les observations

Après cette étape apparaît une boîte de dialogue permettant de désigner les variables de tri et d'indiquer l'ordre du tri (croissant ou décroissant). Au cas on a plusieurs variables de tri, celles-ci sont classées par ordre de priorité.

En reprenant notre fichier *demo.sav*, la figure ci-dessous nous présente le cas d'un tri des individus par revenu familial décroissant et par Age décroissant.



Ainsi, les individus vivant dans les foyers plus nanties se trouvent dans les premières lignes du tableau, ensuite à revenu égale les plus âgés seront d'abord pris en compte.

## 6.2 Filtrer les données - Sous population

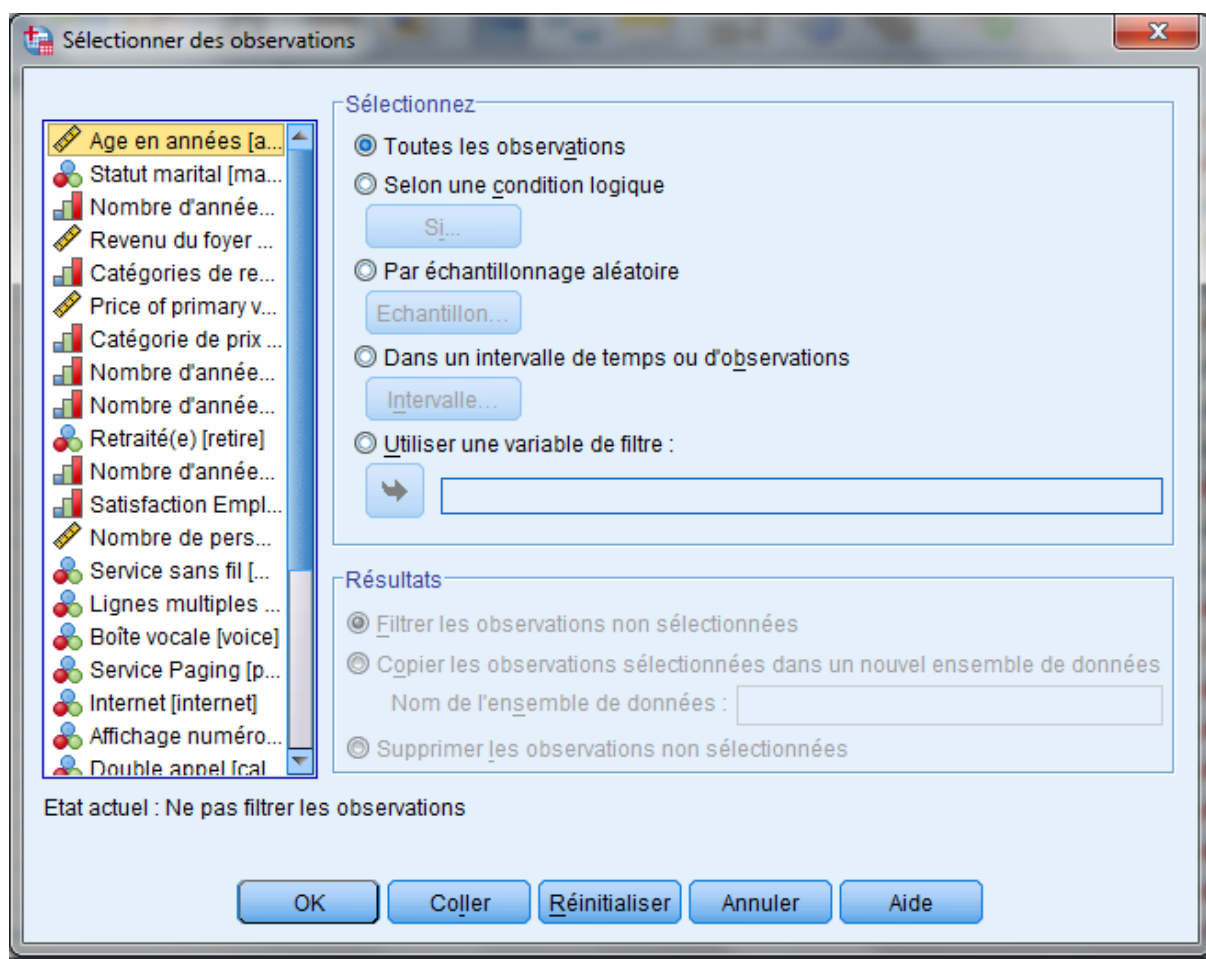
Cette fonctionnalité obéit aux besoins d'effectuer des traitements et/ou des analyses sur une partie de l'échantillon et non sur sa globalité. Par exemple, on peut être amené dans un échantillon comprenant des individus de tout âge, à travailler uniquement sur les enfants de 0 à 4 ans. Lorsqu'on sélectionne une partie des observations on dit qu'on filtre les données.

Pour appliquer un filtre aux données, on procède de la façon suivante :

- ✓ A partir des menus, sélectionnez :

#### Données > Sélectionner des observations

Cela permet d'ouvrir la boîte de dialogue ci-dessous.



Le cadre « sélectionner » de cette boîte de dialogue permet de définir le filtre à appliquer. Ainsi :

➤ **« Selon une condition logique »**

Permet de définir une expression conditionnelle qui déterminera si une ligne est sélectionnée ou non. Pour ce faire, après avoir choisi cette option, on clique sur le bouton « Si ». Cela ouvre une boîte de dialogue permet d'entrée l'expression conditionnelle.

Par exemple, on peut sélectionner dans notre fichier *demo.sav* les hommes âgés de moins de 30 ans. L'expression conditionnelle se présente comme suit :

$(age < 30) \& (gender = "m")$

Ou encore on peut sélectionner dans *demo.sav* les individus n'ayant pas de données manquantes pour la variable internet. On a donc

$MISSING(internet) \sim= 1$

Par ailleurs, il faut souligner qu'une expression conditionnelle a trois valeurs possibles : « vrai », « faux » ou « manquant ». **Une observation ne sera sélectionnée que si la valeur de l'expression conditionnelle est « vrai », les valeurs « faux » et « manquant » ne le sont pas.**

➤ **Par échantillon aléatoire**

Cette option permet d'extraire de la base de données originale, un sous-échantillon, de manière aléatoire. Pour ce faire, après avoir choisi cette option, cliquez sur échantillon. Cela ouvrira une boîte de dialogue qui permettra de paramétrer l'opération. A ce niveau deux choix s'offrent à l'utilisateur :

- (i) il peut définir un pourcentage. Alors l'échantillon généré aura un nombre d'individus qui correspondra approximativement au pourcentage spécifié
- (ii) il peut définir un effectif. Ce dernier correspondra à l'effectif exact de l'échantillon généré. Pour ce cas-ci ; il est également demandé à l'utilisateur d'indiquer le nombre d'observations à partir duquel l'échantillon sera généré. **Celui doit être inférieur ou égal au nombre total d'observations.**

➤ ***Dans un intervalle de temps ou d'observations***

Cette option permet de sélectionner observation à partir de leurs numéros de ligne ou à partir du temps, dans le cas de séries temporelles. Après avoir choisi cette option, cliquez sur intervalle. Une boîte de dialogue s'ouvre alors pour permettre d'indiquer l'intervalle désiré. On entre alors le numéro de ligne de la première observation (ou bien la date de départ) à retenir et ensuite le numéro de la dernière observation à retenir. Si le numéro de la première observation à retenir est spécifié et que la dernière ne l'est pas, la sélection s'étendra jusqu'à la dernière observation de l'échantillon initiale. De même si la dernière observation à retenir est spécifiée et que la première ne l'est pas alors la sélection commencera dès la première observation du fichier original et s'étendra jusqu'à l'observation indiquée.

➤ ***Utiliser une variable de Filtre***

Cette option consiste à utiliser une variable numérique déjà présente dans la base comme filtre. Les observations ayant une valeur autre que « manquante » ou « nulle » pour la variable spécifiée comme variable de filtre seront sélectionnées. Toutefois, il s'avère souvent pratique d'utiliser comme variables de filtre des variables qui prennent deux valeurs possibles 0 ou 1.

Le cadre « Résultats » permet définir le traitement à effectuer après application du filtre. Les observations non sélectionnées peuvent juste être « filtrées ». Dans ce cas, les numéros de ligne des observations non sélectionnées peuvent juste être barrés. L'utilisateur peut également créer un nouveau fichier contenant les observations sélectionnées. Dans ce cas, le fichier original reste non modifié, un nouveau fichier est créé et les observations sélectionnées y sont copiées. En fin, l'utilisateur peut tout simplement choisir de supprimer les observations non sélectionnées. Contrairement au cas précédent, un nouveau fichier n'est pas créé ; les modifications sont faites dans le fichier de données en cours d'utilisation.

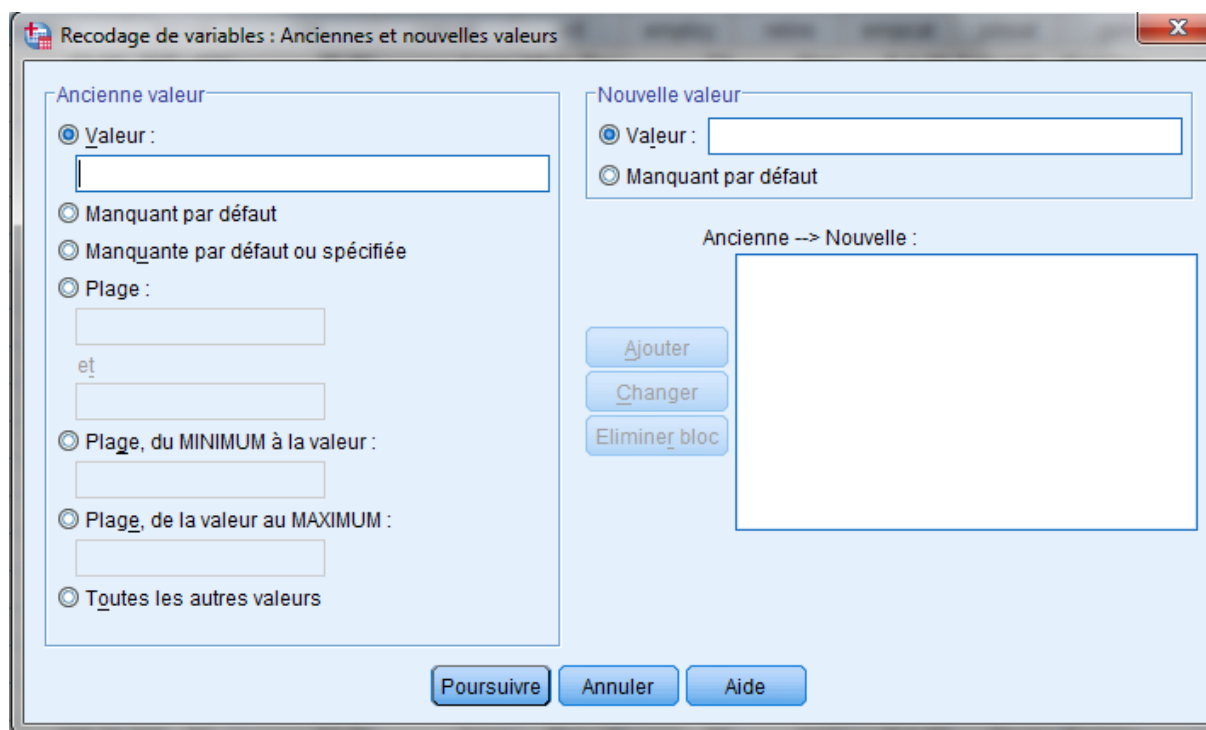
### 6.3 Recodifier les variables

La recodification est utilisée pour réaffecter les valeurs des variables présentes dans la base et regrouper les modalités de celles-ci suivant les choix de l'utilisateur. La procédure pour regrouper les valeurs d'une variable est la suivante.

- ✓ A partir des menus, sélectionnez :

**Transformer > Recoder des variables**

- ✓ La boîte de dialogue qui s'ouvre par la suite invite l'utilisateur à sélectionner la variable à recodifier. Sélectionnez-la.
- ✓ cliquez sur le bouton « Anciennes et nouvelles valeurs ». Cela fait apparaître la boîte de dialogue suivante.



- ✓ Dans le rectangle ancienne valeur, inscrivez les valeurs que vous désirez changer.
- ✓ Dans le rectangle de nouvelle valeur, indiquez la valeur qui sera utilisé pour remplacer les précédentes.
- ✓ cliquez sur « ajouter » puis reprenez les deux précédentes étapes jusqu'à avoir spécifié tous les changements que vous désirez effectuer.
- ✓ Cliquez enfin sur poursuivre.

**NB :** Il faut être prudent lors de la recodification d'une variable, parce qu'une fois recodifier il y a plus de retour en arrière possible. C'est pour ça qu'il est parfois conseillé de « recodifier dans une nouvelle variable ». Ainsi, aucune modification n'est apportée à la variable originale et la nouvelle variable est plutôt créée à partir de l'originale en intégrant les changements souhaités.

#### 6.4 Créer de nouvelles variables

Il existe plusieurs façons de créer des nouvelles variables. L'une des plus simple est d'insérer directement la variable à partir de l'Editeur de données sous l'onglet « affichage des variable ». La nouvelle variable peut être créée en ajoutant une nouvelle ligne et en y renseignant les attributs de la variable (nom, échelle,...). Cette procédure crée une variable « vide » dont les valeurs devront être remplies par la suite.

Toutefois, dans la pratique, les nouvelles variables sont généralement en réalisant des opérations sur celles déjà présentes dans la base. A cet effet on peut distinguer deux

principaux procédés : la création d'une variable par combinaison d'une ou plusieurs autres et la Création d'une variable par recodification d'une autre.

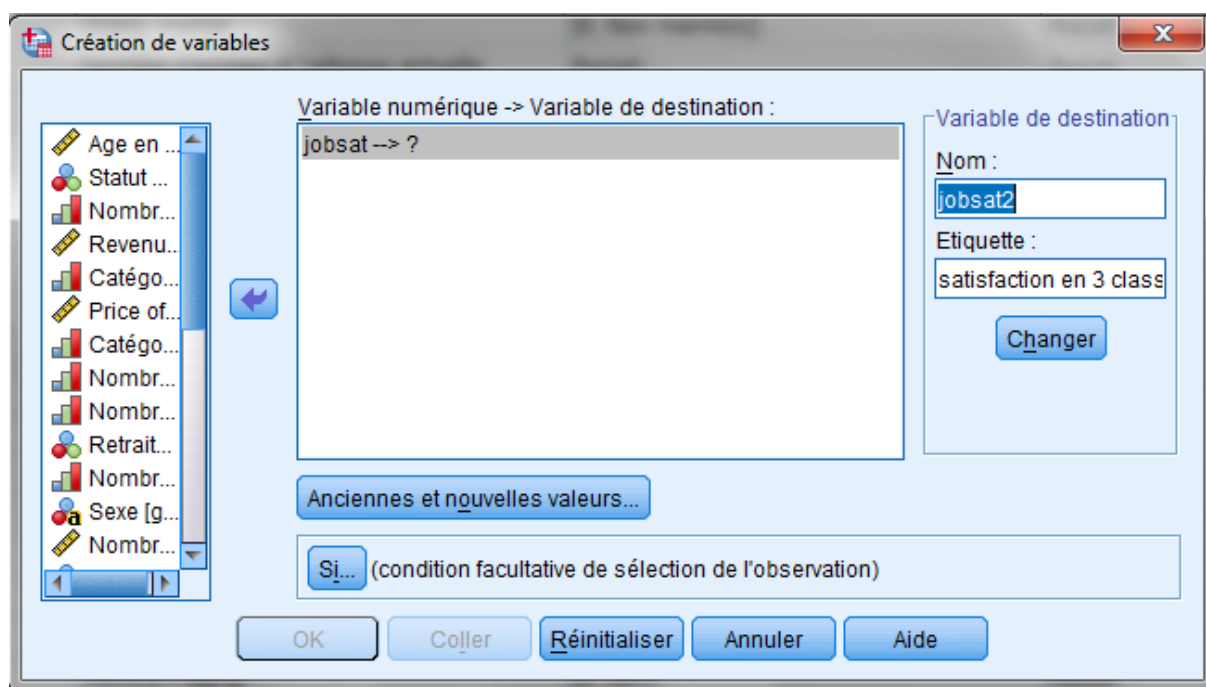
#### 6.4.1 Création d'une variable par recodification d'une autre.

Comme signaler précédemment, l'avantage de cette méthode est de conserver telle quelle, la variable d'origine sans y apporter de modification. De plus, cela permettra de pouvoir faire des vérifications pour s'assurer de la conformité des résultats par rapport aux attentes.

La démarche est la suivante :

- ✓ A partir du menu, sélectionnez

**Transformer > Création des variables**



- ✓ Sélectionnez la variable à transformer dans le rectangle de gauche ;
- ✓ indiquez ensuite le nom et l'étiquette de la nouvelle variables (validez en cliquant sur changer) ;
- ✓ cliquez enfin sur le bouton « Anciennes et nouvelles valeurs ».

La boîte de dialogue qui s'ouvre par la suite est la même celle vue dans la section **Recodifier les variables**. Pour plus d'informations, se reporter cette section.

Une fois ces étapes achevées, la nouvelle variable est créée dans la dernière colonne du tableau de données.

**NB** : il est toujours important de faire des vérifications immédiatement après le recodage afin de s'assurer que les transformations telles qu'escomptées ont bien eu lieu. Les tableaux de contingences peuvent être utilisés à cet effet, dans le cadre des variables catégorielles.

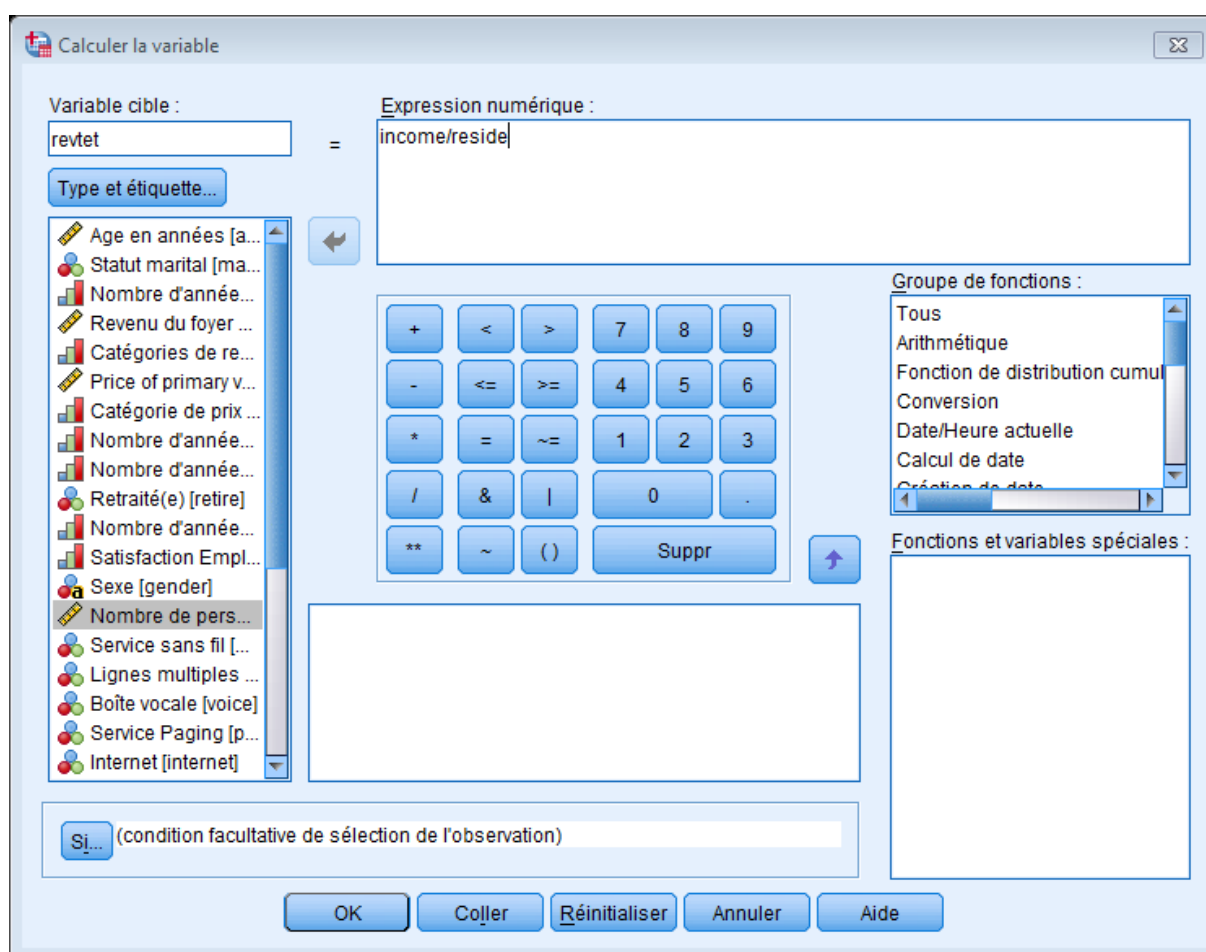
### 6.4.2 Création d'une variable par combinaison d'une ou plusieurs autres

La création d'une nouvelle variable calculée à partir d'autres est une opération très courante dans le traitement et l'analyse statistique.

Dans le fichier *demo.sav* qui contient pour chaque individu, le revenu du foyer et le nombre de résident dans le foyer, il peut par exemple être intéressant de connaître le revenu par tête dans le foyer. Cette variable peut être créée en divisant la variable *revenu du foyer* par la variable *nombre de résident dans le foyer*.

- ✓ A partir du menu, sélectionnez :

**Transformer > Calculer la variable**



Le champ Variable cible permet définir le nom de la nouvelle variable,

Le champ expression numérique permet d'indiquer la formule de calcul à utiliser.

Les champs « Groupe de fonctions » et « Fonctions et variables spéciales » permettent d'accéder à des fonctions qui peuvent être utiles lors de l'élaboration de la formule de calcul.

L'encadré si dessous présente les opérateurs et quelques fonctions utiles



<b>Opérateurs arithmétiques</b>			
+	:addition ;	- :soustraction,	* : multiplication / :division ** :puissance
<b>Opérateurs logiques</b>			
&	: et	: ou	~ négation
<b>Quelques fonctions</b>			
<b>Abs:</b> valeur absolue <b>Mean:</b> moyenne <b>Sd :</b> écart type <b>Sqrt:</b> racine carrée <b>Rnd:</b> arrondir au nombre entier <b>Sum:</b> faire la somme <b>Yrmoda:</b> calculer la période entre deux dates			

## 6.5 Fusionner les données (merge)

Lors du traitement de grosses enquêtes, il arrive souvent que les données soient réparties dans plusieurs fichiers. Il est dans ce cas nécessaire de rechercher les informations dans d'autres tableaux de données autres celui sur lequel on travaille.

Ce genre de problématique fait généralement appel à ce qu'on appelle la fusion des données. Elle consiste à regrouper les informations provenant de différentes sources (ou fichiers). Suivant la configuration dans laquelle on se trouve, on distingue deux types de fusions :

- **L'ajout des observations.** Dans cas de figure, les fichiers ont les mêmes variables mais pas les mêmes observations. Par exemple, dans la pratique on peut avoir une enquête qui couvre une large zone géographique. On pourrait se retrouver à l'issue de celle-ci avec deux fichiers l'un concernant la région 1 et l'autre région 2. Lors de la phase d'analyse l'on sera amené à fusionner ces bases en une seule.
- **L'ajout des variables.** Dans ce cas d'espèce, les fichiers portent sur les mêmes individus mais les variables sont différentes. Dans la pratique, on peut ce cas avec les enquêtes menées en deux phases (périodes) à l'instar des enquêtes d'opinion relatives aux élections où les répondants sont interviewés avant et après le vote. Tout comme dans le cas précédent l'étape de l'analyse peut nécessiter de regrouper les deux bases en une seule.

### 6.5.1 Ajout des observations

Si l'on se met dans le cas figure précédemment évoqué où l'on a une enquête réalisée dans deux régions et dont les données sont disponibles dans deux fichiers distincts relatifs à chacune des régions.



Pour illustrer la procédure, nous allons utiliser la base *customer\_dbase.sav*. Tout d'abord, nous allons partitionner cette base en deux fichiers. La variable *region* (indicateur géographique) fait état de cinq zones dans lesquelles a été conduite la collecte. A partir de celle-ci on créera les deux fichiers. Le premier fichier, *customer\_dbase1.sav*, contiendra les informations relatives aux zones 1, 2 et 3. Le second fichier, *customer\_dbase2.sav*, aura les données pour les zones 4 et 5.

Par ailleurs, pour des raisons didactiques, dans ce second fichier nous allons : (i) modifier le nom de la variable *custid* (identifiant client) et la renommer en « id » ; (ii) modifier le type la variable *zone* est lui assigner la valeur « chaîne » au lieu de « numérique » et (iii) supprimer la variable *ed* (années d'éducation).

Pour fusionner les fichiers *customer\_dbase1.sav* et *customer\_dbase2.sav*, la procédure est la suivante

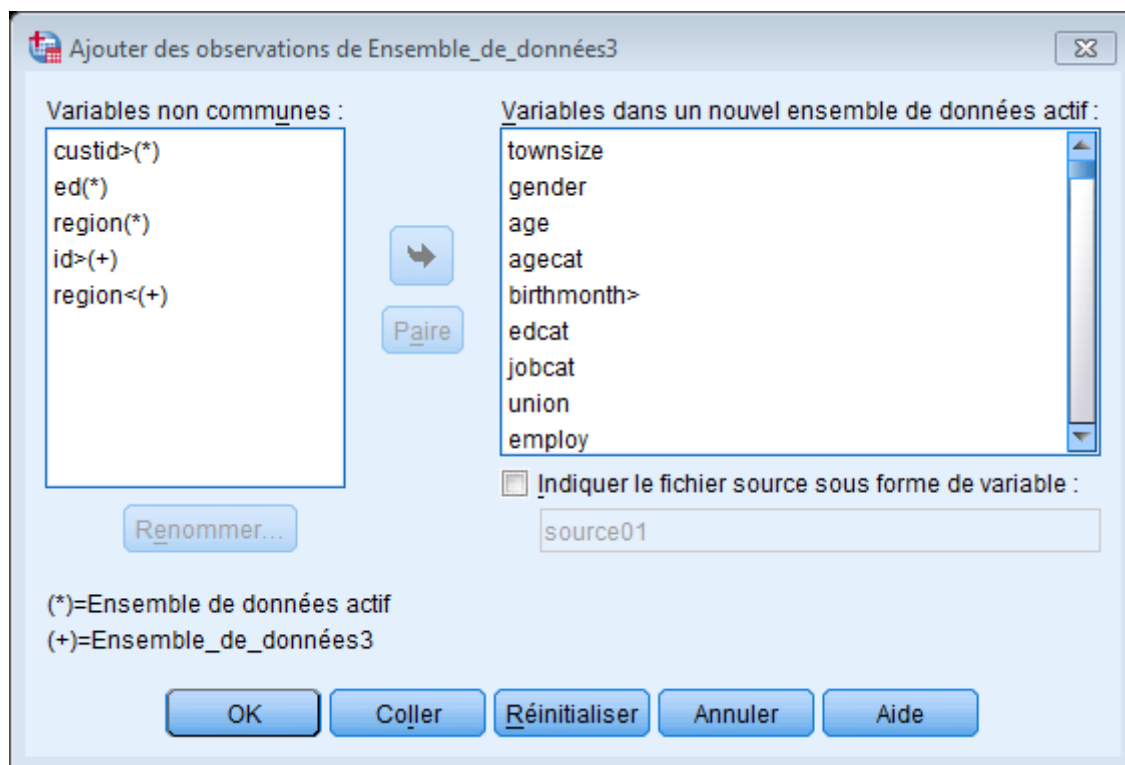
- ✓ Ouvrir l'un des fichiers (*customer\_dbase1.sav*)
- ✓ A partir du menu, sélectionnez :

**Données> Fusionner des fichiers> Ajouter des observations.**

Une fenêtre apparaît pour permettre d'indiquer le fichier contenant les observations supplémentaires.

- ✓ Désigner le fichier *customer\_dbase2.sav* comme étant celui contenant les informations à ajouter

Une nouvelle boîte de dialogue « Ajouter des observations de Ensemble\_de\_données2 » s'ouvre



Dans cette boîte de dialogue, on retrouve deux listes : « Variables non communes » et « Variables dans un nouvel ensemble de données actifs ».

**Variables non communes** c'est l'ensemble de variables qui n'apparaîtront pas dans le nouveau fichier fusionné. Les variables issues de l'ensemble de données actif (celui à partir duquel on a initié procédure) sont identifiées par un astérisque (\*). Les variables de l'autre ensemble de données (le second fichier) sont identifiées par un signe (+).

Par défaut liste des variables non communes s'apparente à la notion intuitive qu'on peut avoir d'elle. Ainsi, elle comprend :

- Des variables sans correspondance dans l'autre fichier. C'est-à-dire les variables qui se retrouvent uniquement dans l'un des fichiers. Dans notre exemple on a les variables *ed* (années d'éducation), *region* (localisation géographique) et *zone* (localisation géographique). Pour ces deux dernières, bien que soient des mêmes variables le logiciel ne peut les identifier telles quelles car elles portent des noms différents dans les deux fichiers.
- Des variables contenant des données numériques dans l'un des fichiers et des données alphanumériques dans l'autre. C'est le cas de la variable *region*
- Des variables alphanumériques de longueurs différentes

**Variables dans un nouvel ensemble de données actif** : c'est l'ensemble des variables qui seront inclus dans le nouveau fichier. Par défaut, ce sont toutes les variables dont les noms et les types correspondent dans les deux fichiers.

Il est possible d'exclure du fichier fusionné une variable dont le nom et le type correspond dans les deux fichiers. Pour cela il suffit de la sélectionner et de l'envoyer dans liste des variables non communes.

Réciproquement, l'utilisateur peut repositionner une variable se trouvant par défaut dans la liste des variables non conformes dans celles des *Variables dans un nouvel ensemble de données actif*. Toutefois dans le fichier fusionné, ces variables contiendront des données manquantes pour les observations du fichier qui ne les contient pas.

Par ailleurs, il arrive qu'on ait des cas des variables communes aux deux fichiers et enregistrées sous des noms différents. Ce cas de figure est rencontré dans notre exemple avec la variable *custid* du fichier actif et *id* du second fichier. Pour inclure la variable dans le fichier fusionné, la démarche consiste à sélectionner les deux variables dans la liste des variables non communes et de cliquer sur le bouton **Paire** alors actif.

La case à cocher « Indiquer le fichier source sous forme de variable » permet la création d'une variable dans le fichier fusionné indiquant la provenance chaque observation. Elle prend la valeur de 0 pour les observations de l'ensemble de données actif et la valeur de 1 pour les observations du second fichier.

## 6.5.2 Ajouter des variables

Pour illustrer cette section, en plus du fichier *customer\_dbase.sav* nous utiliserons également les données de la troisième Enquête Camerounaise Auprès des Ménages (ECAM3). Ces données sont disponibles dans 2 fichiers données, dont l'un est relatif des ménages aux ménages enquêtés et l'autres aux individus interviewés dans ces ménages.

Dans un premier temps, partant du fichier *customer\_dbase.sav*, nous allons créer deux fichiers de données. (i) le premier fichier, *customer\_dbase1v.sav*, contiendra uniquement les 120 premières variables de la base c'est-à-dire celles comprises entre *custid* et *hourstv*; (ii) le second, *customer\_dbase2v.sav*, contiendra 13 variables à savoir *custid* ainsi que celles précédemment exclues lors de la construction de la base précédente.

Pour reprendre l'un des exemples précédemment évoqués, l'on va considérer que l'enquête a été réalisée en 2 phases (périodes). Lors de la première phase les informations contenues dans *customer\_dbase1v.sav* ont été collectées. La seconde phase a quant à elle porté sur *customer\_dbase2v.sav*. La variable *custid* constitue l'identifiant de chaque enquêté et prend une valeur unique pour chacun de ceux-ci. La problématique est donc de constituer un fichier unique à l'issue de l'enquête.

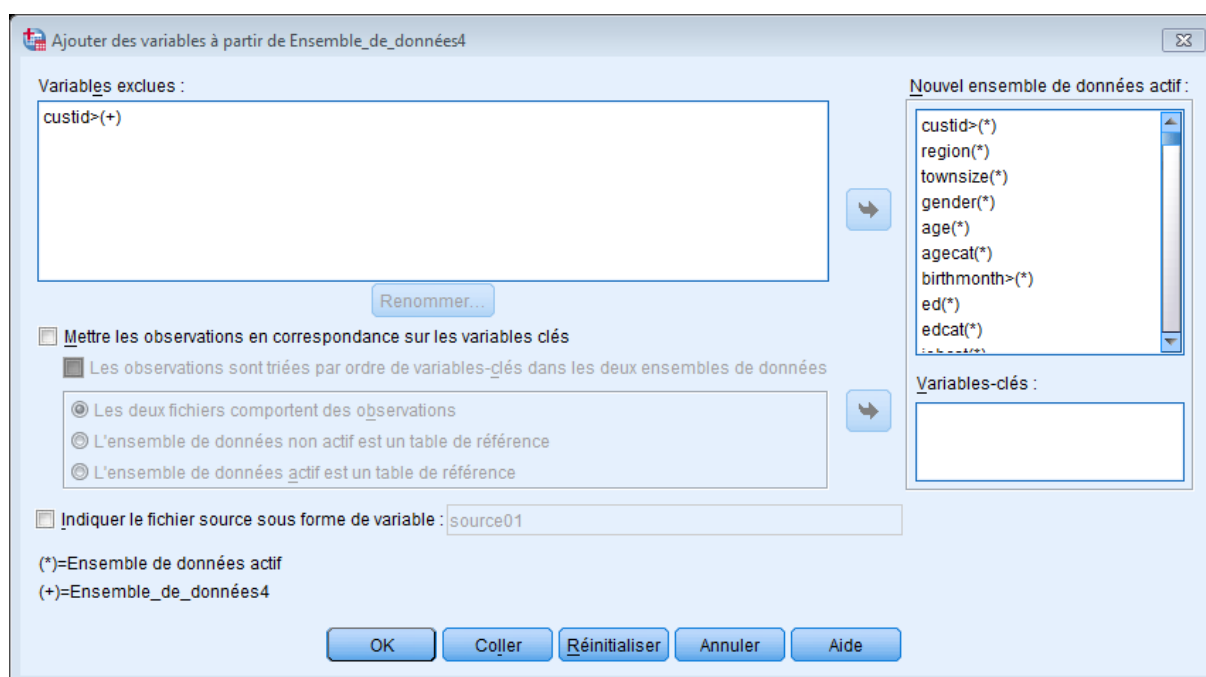
- ✓ Ouvrir l'un des fichiers (*customer\_dbase1v.sav*)
- ✓ A partir du menu, sélectionnez :

**Données> Fusionner des fichiers> Ajouter des variables.**

Une fenêtre apparaît pour indiquer le fichier contenant les variables supplémentaires.

- ✓ Désigner le fichier *customer\_dbase2.sav* comme étant celui contenant les informations à ajouter

Une nouvelle boîte de dialogue « Ajouter des variables de ... » s'ouvre



Le Cadre *Variables exclues* contient la liste de toutes les variables à exclure du fichier fusionné. Par défaut, cette liste contient les noms de variable qui apparaissent dans les deux fichiers. Comme dans le cas de l'ajout des observations les symboles (\*) et (+) indiquent la provenance de la variable. (\*) désigne les variables de l'ensemble des données actif et (+) celles de l'autre fichier. Ainsi, dans notre exemple seule la variable *custid* du second fichier a été exclue.

La listbox **Nouvel ensemble de données actif** contient toutes les variables à inclure dans le fichier fusionné. Par défaut, elle contient les noms de variable qui n'apparaissent pas dans la liste *variables exclues*.

La case à cocher **Mettre les observations en correspondance sur les variables Clés** constitue l'un des paramétrages les plus importants de l'opération de fusion des deux fichiers. En effet, on distingue deux cas de figure selon que l'on utilise ou non une clé d'appariement.

### ❖ La fusion des fichiers sans utilisation d'une clé d'appariement

Il correspond à la situation où la case à cocher *Mettre les observations en correspondance sur les variables Clés* n'est pas sélectionnée. Dans ce cas, l'ordre du fichier détermine le mode d'appariement des observations. En d'autres termes, les individus dans les deux fichiers doivent être ordonnés de la même façon.

### ❖ La fusion des fichiers avec utilisation d'une clé d'appariement

Dans ce cas, pour chaque observation, la clé sera utilisée pour pouvoir identifier les lignes à appairer dans les deux fichiers. Cocher la case *Mettre les observations en correspondance sur les variables Clés* permet de définir des clés d'appariement

**La clé permet d'identifier de façon unique un individu de la base.** Elle peut être constituée par une variable ou d'une combinaison de plusieurs variables. Les clés d'appariement doivent avoir le même dans les deux fichiers de données. Dans le cas contraire, le bouton *renommer* permet à l'utilisateur de renommer l'une des variables de façon à avoir des noms identiques.

Lorsqu'on indique que les observations sont triées par ordre de variables-clés dans les deux ensembles de données cela signifie que les deux fichiers doivent être triés suivant ces clés d'appariement par ordre croissant. **Il est fortement conseillé d'utiliser cette option.** De plus, des versions précédentes de SPSS, imposait comme préalable pour la fusion avec utilisation d'une clé d'appariement que les fichiers de données soient triés suivant la clé.

### Cas spécifique où l'un des fichiers à fusionner est une table codée

La fonctionnalité « fusion de fichiers par ajout de variables » permet également de traiter un cas de figure plus subtil que le cadre d'application classique présenté plus haut, à savoir, « les deux fichiers portent sur les mêmes individus mais les variables ne correspondent pas ». En effet, l'on rencontre aussi un second cas d'espèce où l'un des fichiers est une table codée.

Un fichier est appelé « table codée » ou encore « fichier de consultation de table » lorsque chacune de ses observations peut s'appliquer à plusieurs observations de l'autre fichier. Par exemple, on peut avoir à l'issue d'une enquête un fichier contenant les informations sur les ménages enquêtés et un second portant sur les individus se trouvant dans ces ménages. Pour chaque ménage, les informations générales contenues dans la base de données (telles que le revenu global, la taille) s'appliquent à toutes les observations du fichier « individus » appartenant à celui-ci.

Pour illustrer cela nous allons utiliser la base de données ECAM3. Lors de cette opération l'échantillonnage s'est effectué en 4 étapes

- Premièrement douze régions ont été identifiées comme étant des strates dans le plan de sondage (1<sup>er</sup> niveau de stratification)
- Deuxièmement, toutes les strates ont été subdivisées en Zones de Dénombrement (2<sup>e</sup> niveau de stratification)
- Dans chacune des Zones de Dénombrement (ZD) a été tirés aléatoirement les ménages à interviewer (échantillon des ménages)
- Pour chaque ménage sélectionné, les informations ont été collectées sur tous les individus le constituant.

Ainsi un ménage de l'échantillon est identifié de façon unique par : le numéro de sa région (SOQ1), le numéro de sa Zone de Dénombrement (SOQ2) et son propre numéro dans la ZD (SOQ4). La clé pour les ménages est donc constituée la combinaison de ces trois variables.

Dans cet exemple, il s'agit pour nous, d'ajouter dans la base de données relative aux individus, une variable, qui donne pour chaque observation, le montant dépenses de logement du ménage de cette dernière.

Concrètement, la variable *deplog* de la base ménages donne le cout des dépenses de logement pour chaque ménage. Il est donc question de fusionner la base « individus » avec la base « ménage », en ajoutant à celui-ci la variable *deplog*. On note ici que les natures des observations dans les deux fichiers ne sont pas exactement les mêmes. On a d'un côté des individus et de l'autre des ménages. Cependant, la base de données « ménage » est une table codée pour fichier relatif aux individus.

Pour faire la fusion :

- ✓ Ouvrez les deux bases de données et triez les dans l'ordre croissant suivant les variables : SOQ1, SOQ2 et SOQ4.
- ✓ A partir du menu de la fenêtre de la base de données « individus », sélectionnez :  
**Données> Fusionner des fichiers> Ajouter des variables.**
- ✓ Désigner la base de données « ménages » comme étant celle contenant les informations à ajouter
- ✓ cochez la case mettre les observations en correspondance sur les variables clés
- ✓ sélectionnez comme clé d'appariement les variables (**la hiérarchie des variables est importante**)
- ✓ cochez la case *les observations sont triées par ordre de variables clés dans les deux ensembles de données*
- ✓ dans la liste *nouvel ensemble de données actif*, parmi les variables provenant du fichier ménage (celles avec le symbole(+)), ne conservez que la variable *deplog*
- ✓ sélectionnez le bouton d'option *l'ensemble de données non actif est une table de référence*
- ✓ validez en cliquant sur OK

## 6.6 Agréger les données

Cette fonctionnalité permet d'agréger les données par groupes définis à l'aide de critère d'agrégation. Pour un groupe précis, l'action « d'agréger » peut consister à un simple dénombrement ou encore à calculer une statistique pour une variable donnée sur les ensemble des individus constituant celui-ci. Par exemple dans un fichier dans données issues d'une enquête portant sur les individus, on peut avoir pour chaque observation de la base un identifiant ménage permettant de définir de façon unique le ménage auquel celle-ci appartient. En utilisant, comme variable d'agrégation l'identifiant ménage, il est possible de déterminer pour chaque ménage figurant dans la base, le nombre d'individus total enquêtés appartenant à celui ou encore de calculer le revenu moyen des personnes de ce ménage ayant été enquêtées. Pour ce dernier, cas il est toutefois nécessaire d'avoir une variable dans la base une variable relative au revenu des enquêtés.

L'option Agréger les données propose deux traitements possibles pour répondre aux besoins de l'utilisateur. Ainsi, il est possible de créer un nouveau fichier de données ou chaque regroupement constitue une observation ou bien de créer une nouvelle variable dans la base existante pour laquelle chaque individu aura la valeur de la statistique agrégée du groupe dans lequel il se trouve.

En reprenant à l'exemple précédent, la création d'un nouveau fichier consiste ici à avoir une nouvelle base de données où chaque ligne représentera un ménage. Cette nouvelle base contiendra une variable qui identifiera le ménage et une seconde qui donnera pour chaque ménage le nombre d'individu enquêtés.

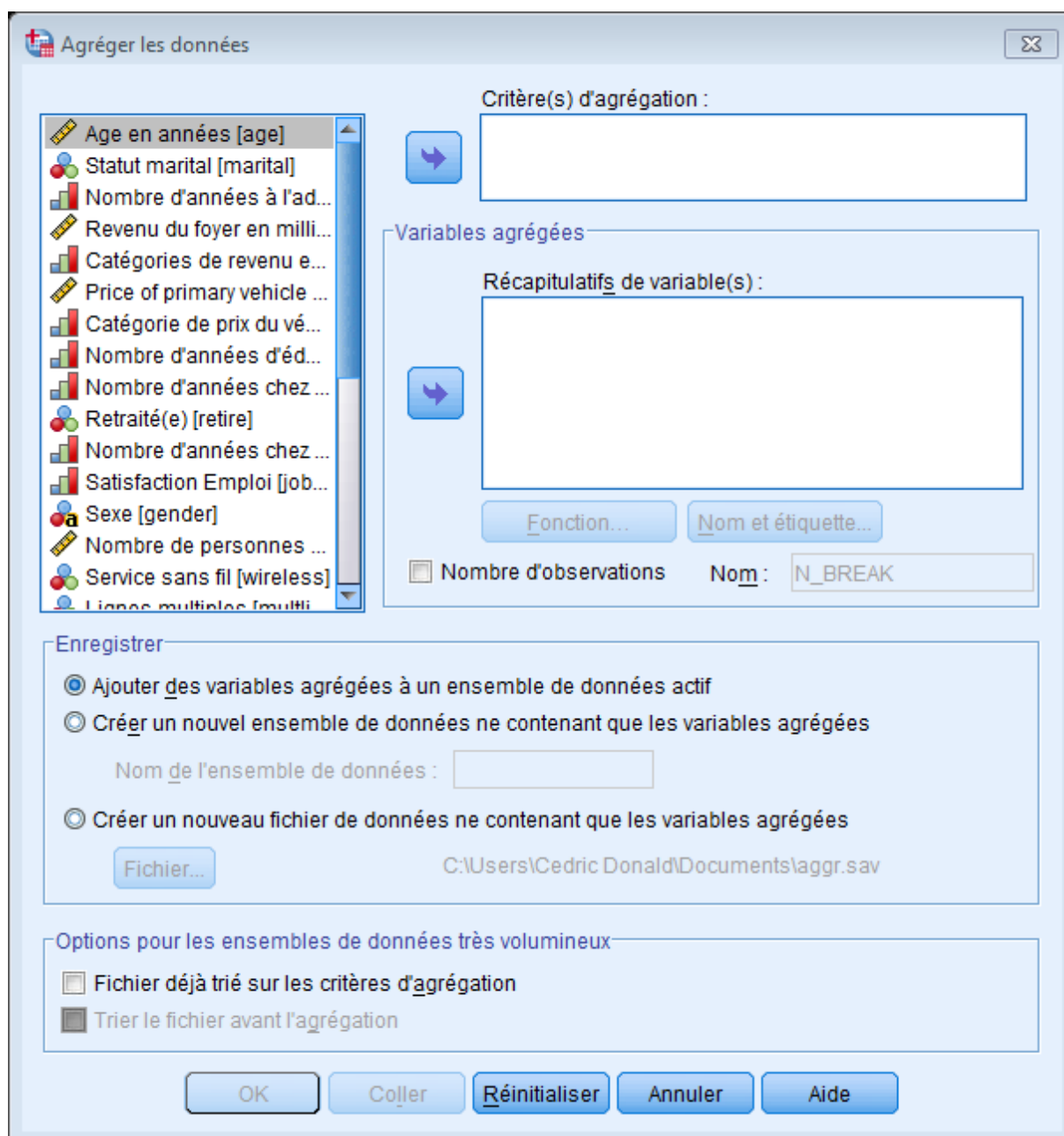
La seconde option, quant à elle, pour cet exemple va consister à créer une nouvelle variable dans la base. Pour chaque individu de la base, elle indiquera le total de personnes du même ménage figurant dans la base.

La procédure est la suivante :

- ✓ A partir du menu, sélectionnez :

**Données>Agréger les données.**

La boîte de dialogue suivante s'ouvrira :



Sélectionner une ou plusieurs variables d'agrégation et utiliser les fonctions quelle statistique agréger vous voulez obtenir.

## 6.7 Gérer des données pondérées

Lors des enquêtes de grandes envergures il arrive souvent que les données le plan de sondage conduisent à avoir des pondérations (non identiques) pour chaque individu de la population. Dans ce cas, une analyse ignorant ces pondérations peut être potentiellement biaisé.

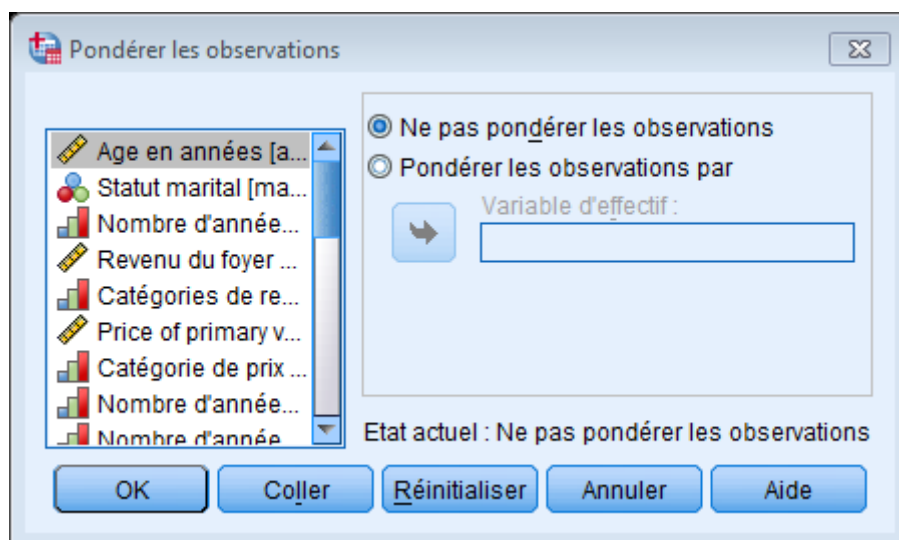
Dans SPSS, la prise en compte des pondérations est assez simple. En effet, il suffit de spécifier la variable comprenant les pondérations pour que ces dernières soient prises en compte dans différents outils d'analyse (tableaux, graphes ...). La procédure pour designer la variable des pondérations est la suivante :

- ✓ A partir du menu, sélectionnez :



**Données>Pondérer les observations**

Une fois exécuter la boîte de dialogue suivante s'ouvre, invitant l'utilisateur à indiquer la variable contenant les pondérations pour chaque observation.



Il est conseillé d'utiliser des pondérations absolues plutôt que des poids relatifs (proportions). En effet, SPSS gère les pondérations par réplication simulée. Ainsi, les pondérations doivent indiquer le nombre individus de la population totale représentées par des observations uniques de la base de données. Les observations avec zéro, une valeur négative ainsi que les variables manquantes pour la variable de pondération sont exclues de l'analyse.

Par ailleurs, certaines procédures, telles que les effectifs et les tableaux croisés peuvent utiliser les valeurs fractionnelles.

## 7 Création et modification des tableaux

Il existe différentes façon de créer des tableaux dans SPSS. Toutefois la fonctionnalité de création des tableaux personnalisés est assez complète et offre une très grande souplesse dans la construction. Cette section se focalisera donc sur la création des tableaux personnalisés.

### 7.1 Création des tableaux

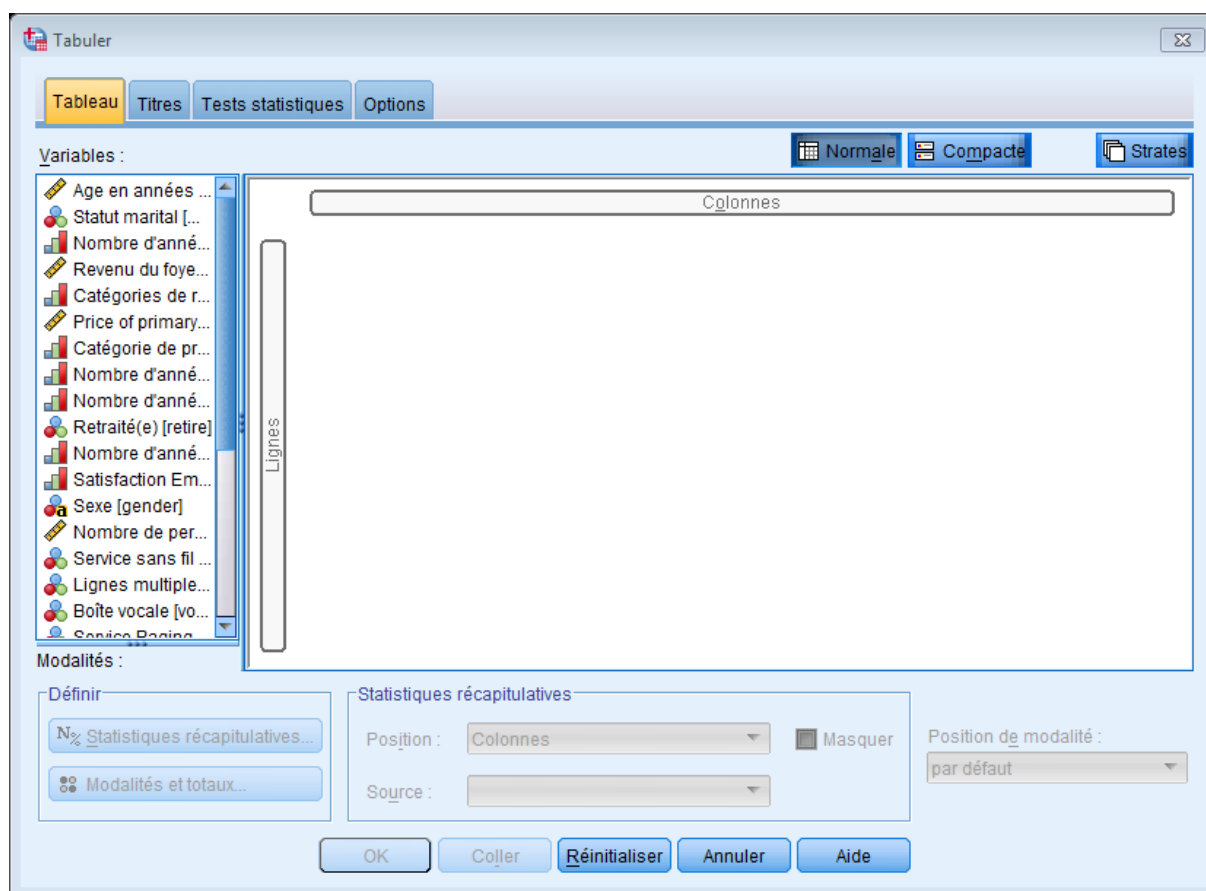
Pour créer un tableau personnalisé :

- ✓ A partir du menu de l'éditeur de données, sélectionnez :

**Analyse>Tableau>Tabuler**

Cela entraine l'ouverture du générateur de tableau.



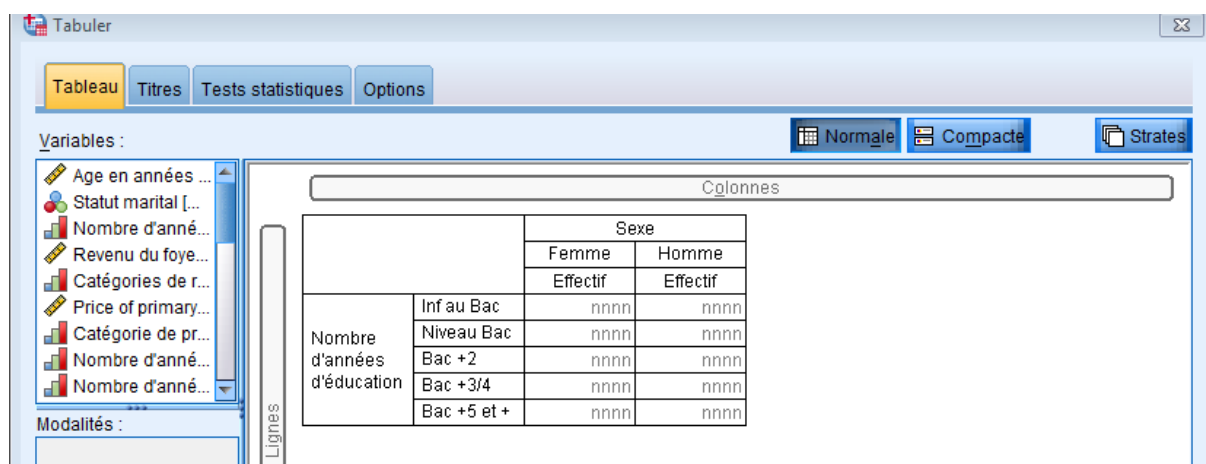


L'interface du générateur est très intuitive. La sélection des variables se fait l'aide de « glisser-déposer » dans le panneau de canevas. Ce panneau permet d'avoir un aperçu du tableau qui sera implémenté.

### 7.1.1 Tableaux de Contingence

Pour créer un tableau de contingence, il suffit juste d'indiquer les variables devant figurer en ligne et en colonne.

Comme exemple, nous allons croiser les variables catégorielles « nombre d'années d'éducatons » et « sexe » du fichier *demo.sav*.



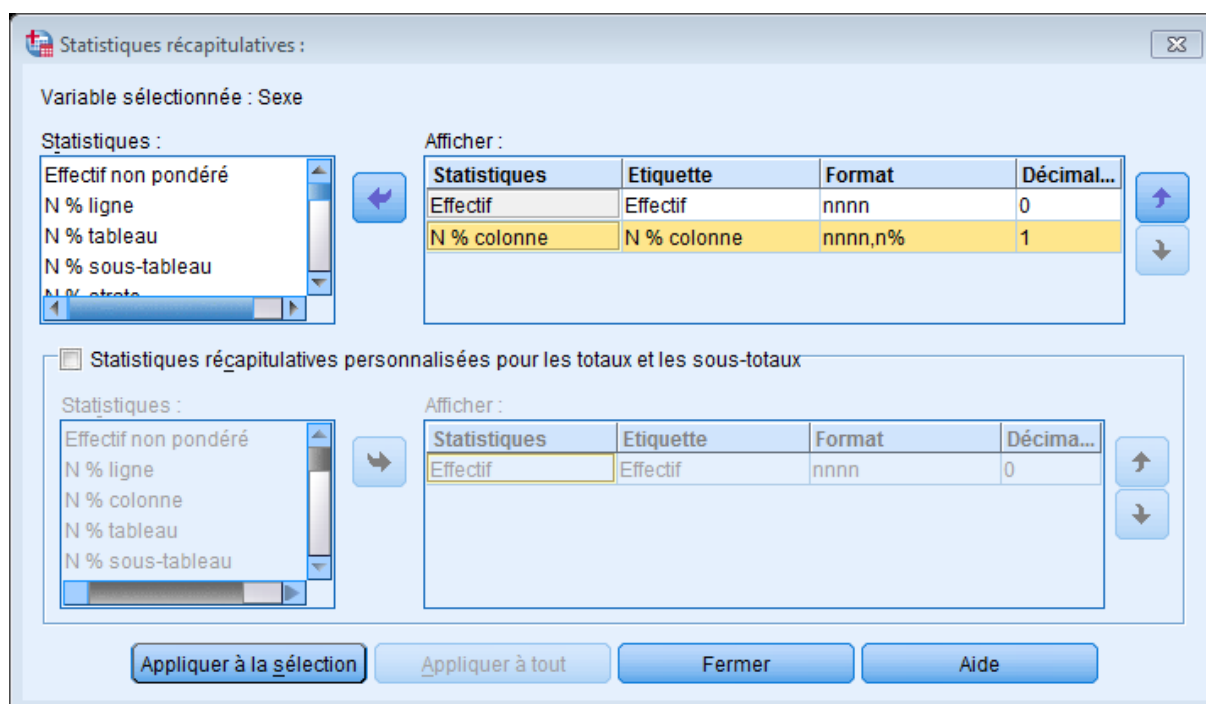
Une fois les variables sélectionnées, on valide en cliquant sur OK. Le tableau est alors généré.

Par défaut, lorsqu'il s'agit des variables qualitatives, la statistique affichée dans les cellules c'est « les effectifs ». Cependant, Il y a des cas où l'on souhaiterait avoir des pourcentages (pourcentages en ligne, pourcentages en colonne ou encore pourcentage totale).

Dans notre exemple où nous avons croisé les variables catégorielles « nombres d'années d'éductions » et « sexe » nous allons présenter les pourcentages par colonne. Pour cela :

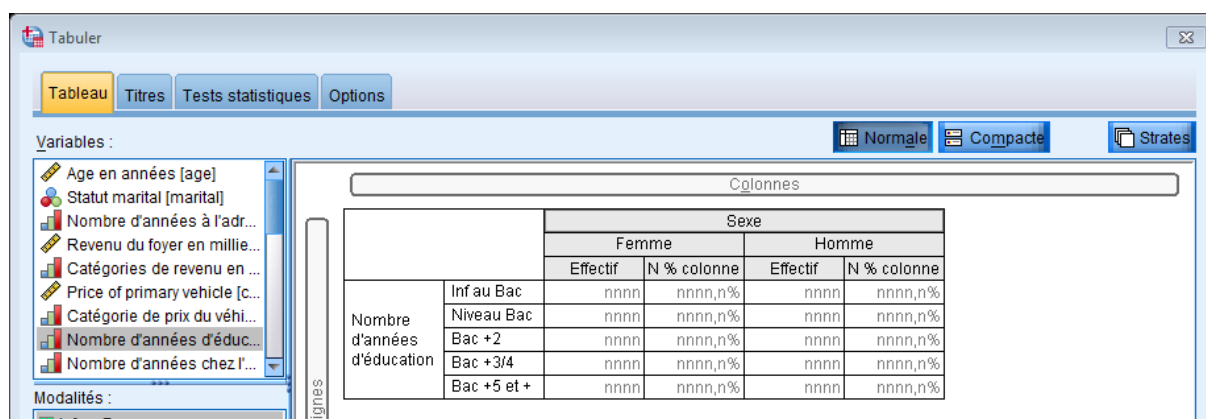
- ✓ revenez au générateur de tableau
- ✓ Sélectionnez la variable « sexe » et cliquer sur le bouton de commande statistique **N% Statistiques récapitulatives**.

Cela ouvre une boîte de dialogue qui permet de sélectionner les statistiques que l'on voudrait afficher dans le tableau.



Dans notre exemple, on a sélectionné « N% colonne » de la liste de statistique et on l'a envoyé dans le cadre Afficher. Il y a également la possibilité de supprimer la statique « Effectif » et de ne laisser que les pourcentages colonnes.

- ✓ Pour valider ces paramètres, cliquez sur Appliquer à la sélection pour enregistrer ces paramètres.



Il est souvent commun de présenter tableaux les informations relatifs aux sous- totaux et totaux. Par défaut, celles-ci ne sont pas incluses lors de la création des tableaux.

Pour ajouter les totaux en ligne, on va sélectionner la variable « nombre d'années d'éducatons » et faire un clic-droit. Ensuite dans menu contextuel, on va choisir **Modalités et totaux**.

Dans la boîte de dialogue **Modalités et totaux** qui s'ouvre par la suite, on coche la case **Total** qui se trouve dans le cadre **Afficher** situé du côté droit de la boîte de dialogue.

On peut aussi insérer les sous totaux « inférieur ou égale au bac » et « supérieur au bac ». Pour cela on va sélectionner la modalité « niveau BAC » et cliquer sur le bouton **Ajouter un sous-total**. Une fois effectué, on définit l'étiquette du sous total. Après cela, on sélectionne la modalité « Bac +5 et + » et on clique sur sous total.

Modalités et totaux

Variable sélectionnée : Nombre d'années d'éducation

**Affichage**

Valeurs

Valeur(s)	Etiquette
1	Inf au Bac
2	Niveau Bac
1...2	inferieur ou égale au BAC
3	Bac +2
4	Bac +3/4
5	Bac +5 et +
3...5	superieur au BAC

Catégories de sous-totaux et calculées

Ajouter un sous-total... Ajouter une catégorie... Modifier... Supprimer

Modalités omises de tous les sous-totaux : 0

Tri des modalités

Par :  Ordre : Croissant

Exclure :

Afficher

☒ Total

Etiquette : Total

☐ Valeurs manquantes

☒ Modalités vides

☒ Autres valeurs lorsque les données sont analysées

Totaux et sous-totaux

☐ Modalités ci-dessus auxquelles ils s'appliquent

☒ Modalités ci-dessous auxquelles ils s'appliquent

Appliquer Annuler Aide

On note également que cette boîte permet aussi de sélectionner les modalités à exclure lors de la construction du tableau et de définir l'ordre d'apparition des modalités dans le tableau.

Une fois terminé, on clique sur le bouton **Appliquer** puis sur le bouton **OK** dans la fenêtre tabuler.

		Sexe			
		Femme		Homme	
		Effectif	N % colonne	Effectif	N % colonne
Nombre d'années d'éducation	Inf au Bac	674	21,2%	716	22,2%
	Niveau Bac	957	30,1%	979	30,4%
	inferieur ou égale au BAC	1631	51,3%	1695	52,6%
	Bac +2	689	21,7%	671	20,8%
	Bac +3/4	676	21,3%	679	21,1%
	Bac +5 et +	183	5,8%	176	5,5%
	superieur au BAC	1548	48,7%	1526	47,4%
	Total	3179	100,0%	3221	100,0%

### 7.1.2 Tableaux présentant les statistiques d'une variable quantitative selon les classes d'une variable qualitative.

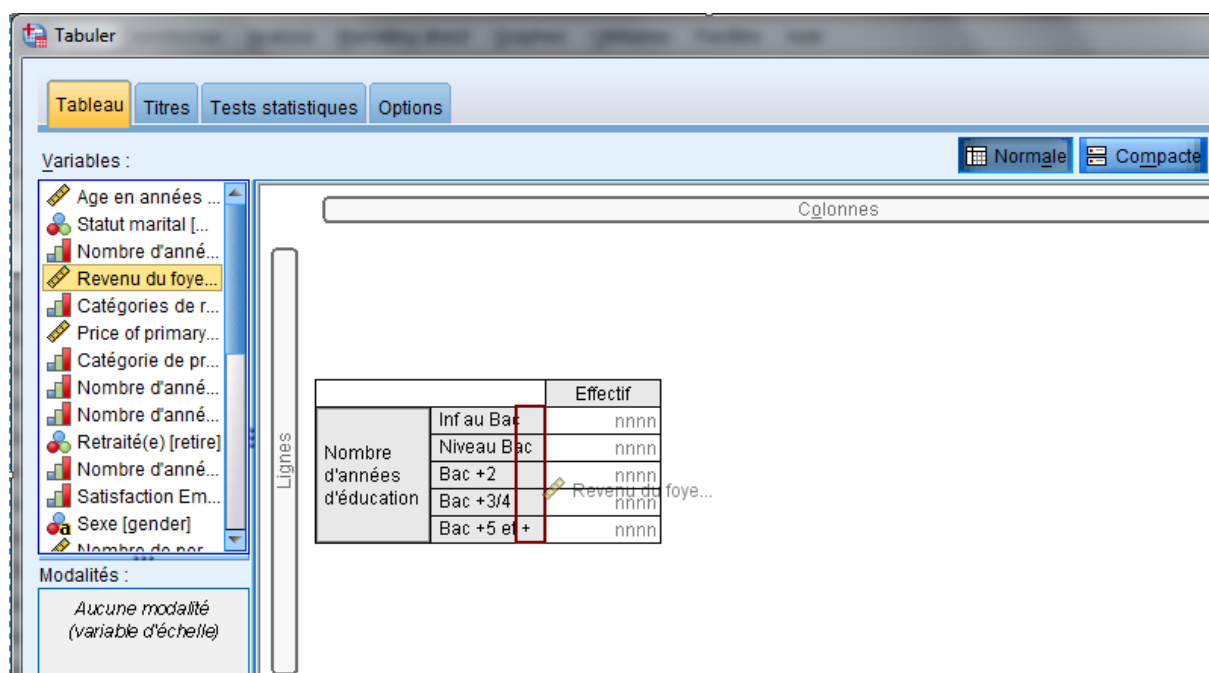
Ce sont des tableaux qui croisent en quelques sortes une variable qualitative et une variable quantitative. Pour chaque groupe constitué des individus ayant la même de modalité de variable catégorielle, on calcule les statistiques de la variable quantitative dans ce groupe.

A titre d'illustration, nous allons calculer les revenus moyen et médian, en fonction du nombre d'année d'étude.

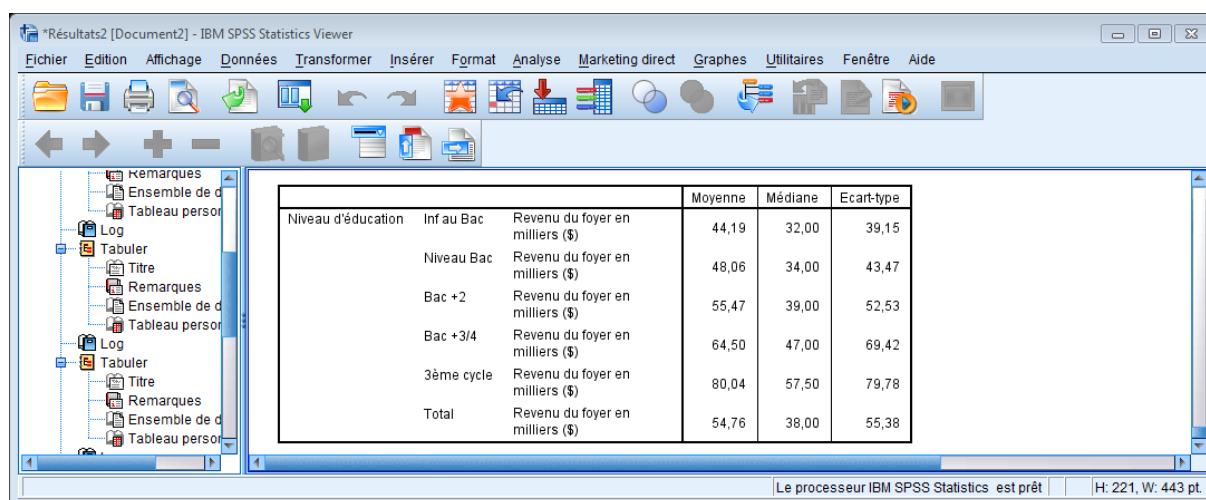
- ✓ On commence par ouvrir le générateur de tableau à partir du Menu :

**Analyse>Tableaux>Tabuler**

- ✓ Une fois dans le générateur de graphique, on sélectionne le niveau d'éducation comme variable ligne.
- ✓ Une fois cela effectué, on sélectionne la variable revenu et on effectue un cliquer-glisser pour la déposer à **droite** de variable éducation (et non pas en colonne) comme indiqué ci-dessous



Après avoir déposé la variable **revenu** à son emplacement, on définit les statistiques que l'on veut calculer. La statistique par défaut pour une variable quantitative, c'est la moyenne. Pour définir d'autres statistiques dans notre exemple, on sélectionne la variable **revenu** puis on clique sur le bouton **N% Statistiques récapitulatives** et ensuite l'on choisit les statistiques que l'on veut voir apparaître dans le tableau.



### 7.1.3 Empilement des variables

L'empilement consiste en quelque à prendre des tableaux séparés et à les coller en un seul tableau. Pour empiler des variables deux options s'offrent à l'utilisateur.

- Ainsi, on peut sélectionner toutes les variables à empiler et faire glisser-déposer pour les poser ensemble en ligne ou en colonne.
- L'autre option consiste à sélectionner les variables les unes après les autres et de les disposer, dans le cas d'un empilement en colonne, à au gauche ou à droite de celles précédemment sélectionnées ; ou bien, au-dessus ou en dessous de celles précédemment sélectionnées dans le cas d'un empilement en ligne.

A titre d'illustration nous allons empiler les variables **sexe** et **catégorie de revenus**.

Colonnes

Lignes

		Effectif
Sexe	Femme	nnnn
	Homme	nnnn
Catégories de revenu en milliers (\$)	Inf à \$25	nnnn
	\$25 - \$49	nnnn
	\$50 - \$74	nnnn
	\$75 - \$124	nnnn
	\$125+	nnnn

Le résultat obtenu est le suivant.

		Effectif
Sexe	Femme	3179
	Homme	3221
Catégories de revenu en milliers (\$)	Inf à \$25	1174
	\$25 - \$49	2388
	\$50 - \$74	1120
	\$75 - \$124	1718
	\$125+	0

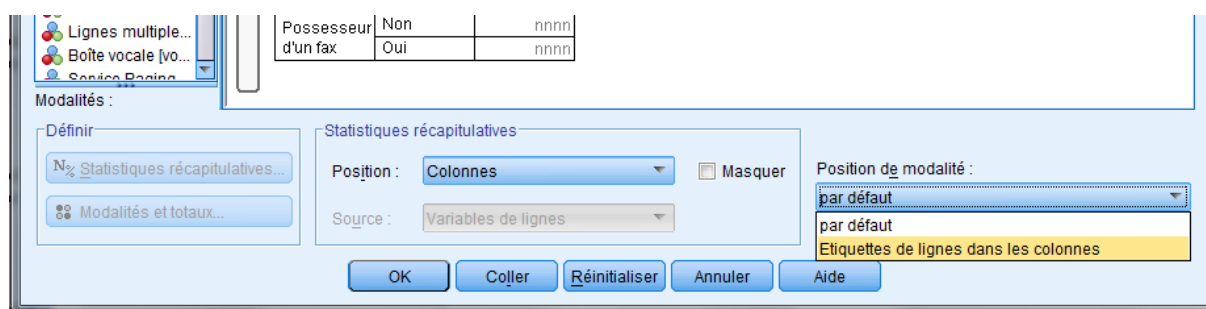
#### 7.1.4 Tableau pour variables avec modalités partagées

L'emboîtement peut être utilisé pour la construction de tableau pour des variables avec des modalités partagées. En effet, dans la base donnée, il arrive qu'on ait des variables qui possèdent les mêmes modalités. Très souvent ces variables relèvent d'une même thématique. Il peut donc être intéressant de les présenter dans un tableau unique avec en ligne les variables et en colonne ces modalités ou bien vice versa.

Dans le cas de notre base *demo.sav* on a par exemple, les variables *owntv*, *ownvcr*, *owncd*, *ownpda*, *ownpc*, *ownfax* relatives à la possession de certains biens d'équipement (Télé, magnétoscope, ordinateur,...) qui ont toutes deux modalités « oui » et « non ».

Pour construire un tableau avec modalités partagées :

- ✓ Dans le générateur de graphique, on commence par positionner les variables à empiler en ligne.
- ✓ Ensuite, Dans la liste déroulante **Position de la modalité**, on sélectionne « Etiquettes de ligne dans les colonnes »



- ✓ on valide en cliquant sur OK

Le tableau ainsi obtenu se présente comme suit :

	Non	Oui
	Effectif	Effectif
Possesseur d'un téléviseur	63	6337
Possesseur d'un magnétoscope	255	6145
Possesseur de matériel HiFi	194	6206
Possesseur d'un agenda électronique	5093	1307
Possesseur d'un ordinateur	3589	2811
Possesseur d'un fax	5198	1202

Pour que l'étiquette « Effectif » ne soit pas affichée, il aurait fallu cocher la case *masquer* dans le cadre *statistiques récapitulatives*.

Pour cet exemple précis, l'utilisation des vecteurs de réponse multiple peut également permettre d'avoir de des résultats similaires (voire même meilleurs).

### 7.1.5 Tableaux avec utilisation de vecteurs de réponses multiples

Les vecteurs de réponses multiples sont utilisés pour enregistrer des réponses des questions invitant le répondant à faire des sélections multiples dans une liste de choix. Par exemple lors d'une enquête l'on peut demander aux enquêtés « lesquels de ces équipements possédez-vous ? » et dans liste des choix on propose : téléviseur, magnétoscope, matériel HiFi, agenda électronique, ordinateur, fax. Le répondant sélectionnera ou cochera les biens qu'il possède.

Dans la base de données, plusieurs variables sont cependant utilisées pour enregistrer ces différents informations comme c'est le cas dans notre fichier *demo.sav* (avec les variables *owntv*, *ownpc*, ...). En effet, les vecteurs multiples ne sont pas des variables au sens propre du terme. Ils peuvent être considérés comme des « variables spécifiques » qui sont pris en charge par certaines procédures telles que : les Tableaux personnalisés et le Générateur de diagrammes. Les vecteurs multiples n'apparaissent pas dans l'éditeur de données. Cependant, ils sont enregistrés dans un fichier de données « .sav ».

#### ❖ Définition d'un vecteur multiple

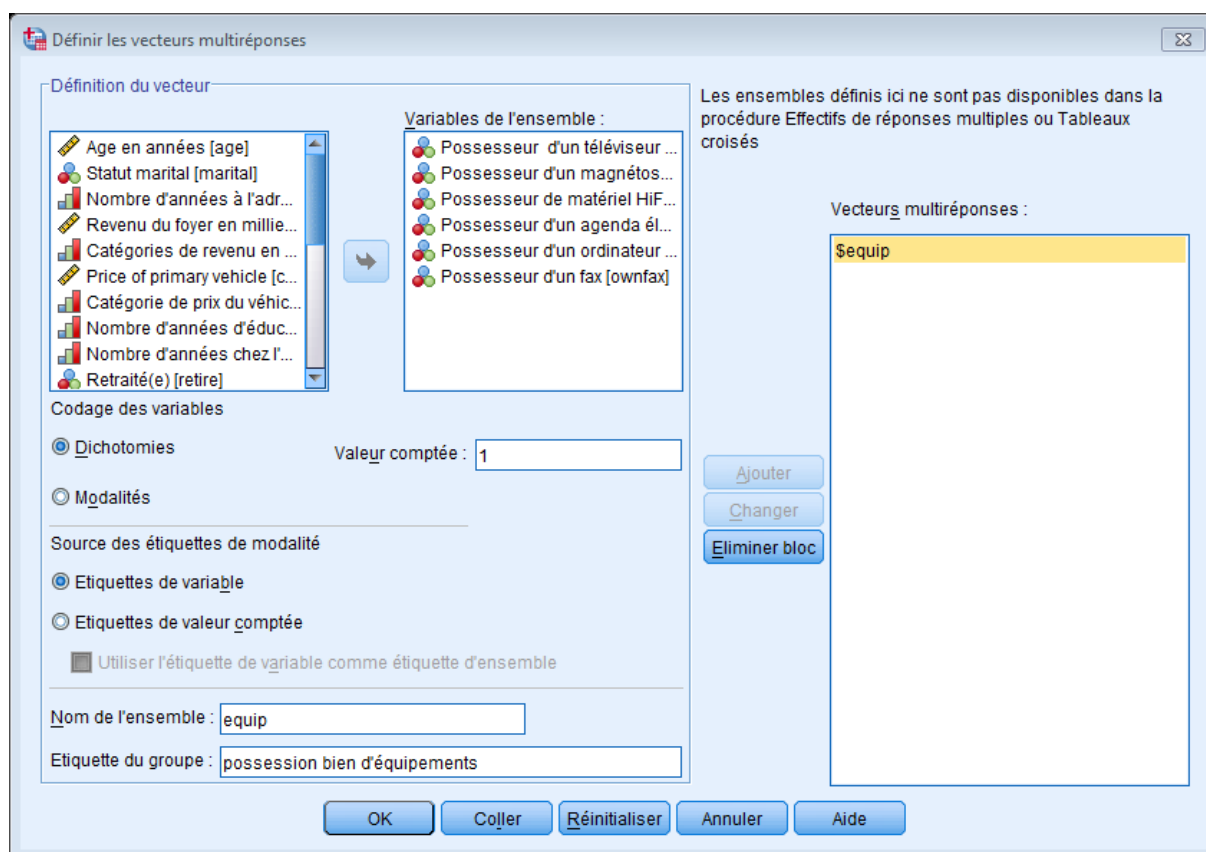
Dans le fichier *demo.sav*, nous allons créer un vecteur multiple relatif à la possession des équipements suivant : téléviseur, magnétoscope, matériel HiFi, agenda électronique, ordinateur et fax. Pour cela :

- ✓ A partir des menus de l'éditeur de données, sélectionnez :

**Données > Définir des vecteurs multiréponses**

La fenêtre Définir les vecteurs de multiréponses apparaît :





- ✓ Sélectionnez les variables qui seront utilisées pour construire le vecteur de réponse multiple.
- ✓ Cochez l'option dichotomies puisque dans notre exemple les variables utilisées sont binaires.
- ✓ Indiquez la valeur (la modalité) qui sera comptabilisée
- ✓ Indiquez le nom et l'étiquette (facultatif) du vecteur de réponse multiple à créer
- ✓ Cliquez sur Ajouter pour ajouter le vecteur multiréponses à la liste des de ceux déjà présents dans la base.
- ✓ Validez en cliquant sur OK.

#### ❖ Utilisation des vecteurs de réponse multiples

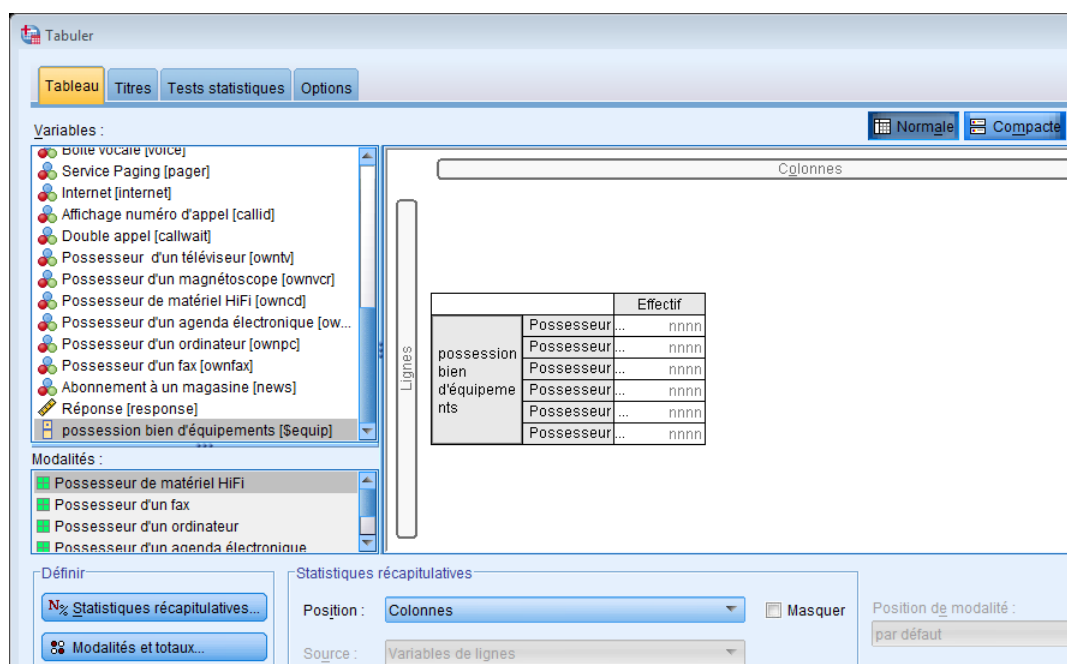
Une fois créés, les vecteurs de réponse multiples peuvent être utilisés dans la création de tableaux personnalisés. En effet, ils sont considérés comme des variables qualitatives.

Nous allons construire un tableau illustratif, qui présente pour les différents biens d'équipement retenus les effectifs et les pourcentages de personnes les possédant.

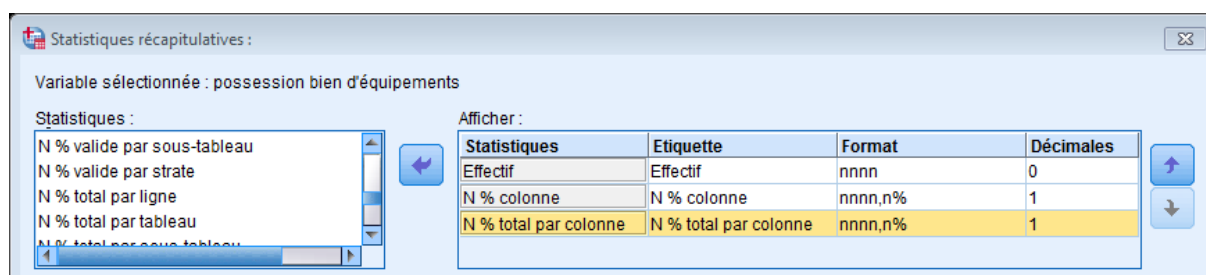
- ✓ A partir du menu, sélectionnez :

**Analyse > Tableaux > Tabuler...**

- ✓ Faites glisser le vecteur de réponse multiples \$equip (son étiquette c'est possession des biens d'équipement) de la liste des variables et positionnez le dans la zone Lignes du panneau de canevas.



- ✓ Cliquez sur Modalités et totaux, et cochez Total dans la boîte de dialogue, puis cliquez sur Appliquer.
- ✓ Cliquez sur le bouton *N% statistiques récapitulatives*
- ✓ Sélectionnez et ajouter les statistiques « % N colonne » et « N total % colonne »



- ✓ Cliquez sur Appliquer à la sélection, puis sur OK pour créer le tableau.

Le tableau est le suivant :

		Effectif	N % colonne	N total % colonne
possession bien d'équipements	Posesseur d'un téléviseur	6337	99,3%	99,0%
	Posesseur d'un magnétoscope	6145	96,3%	96,0%
	Posesseur de matériel HiFi	6206	97,3%	97,0%
	Posesseur d'un agenda électronique	1307	20,5%	20,4%
	Posesseur d'un ordinateur	2811	44,1%	43,9%
	Posesseur d'un fax	1202	18,8%	18,8%
	Total	6379	100,0%	100,0%

ce tableau présente les statistiques demandées pour chaque variable. La statistique « Effectif » correspond nombre de réponse « oui ». Elle ne présente aucune ambiguïté. Ce n'est pas le cas pour les statistiques *Total* et *pourcentage*.

**Il faut faire très attention à l'interprétation des totaux et des pourcentages dans les tableaux construits à partir des vecteurs de réponse multiple.**

Dans notre exemple, dans la colonne effectif, on note que la cellule *Total* ne correspond pas à la somme des valeurs de la colonne. En fait, la valeur présentée dans cette cellule correspond à l'effectif des personnes ayant au moins une réponse « oui » pour l'ensemble de nos six variables binaires. En d'autres termes, c'est le nombre de personnes ayant au moins un des biens figurant dans la liste.

Dans la colonne ***N% colonne***, les pourcentages sont calculés à partir des données de la colonne effectif **avec comme base la valeur affichée dans la cellule total** (6379). Ce résultat n'est pas a priori celui que l'on voudrait présenter.

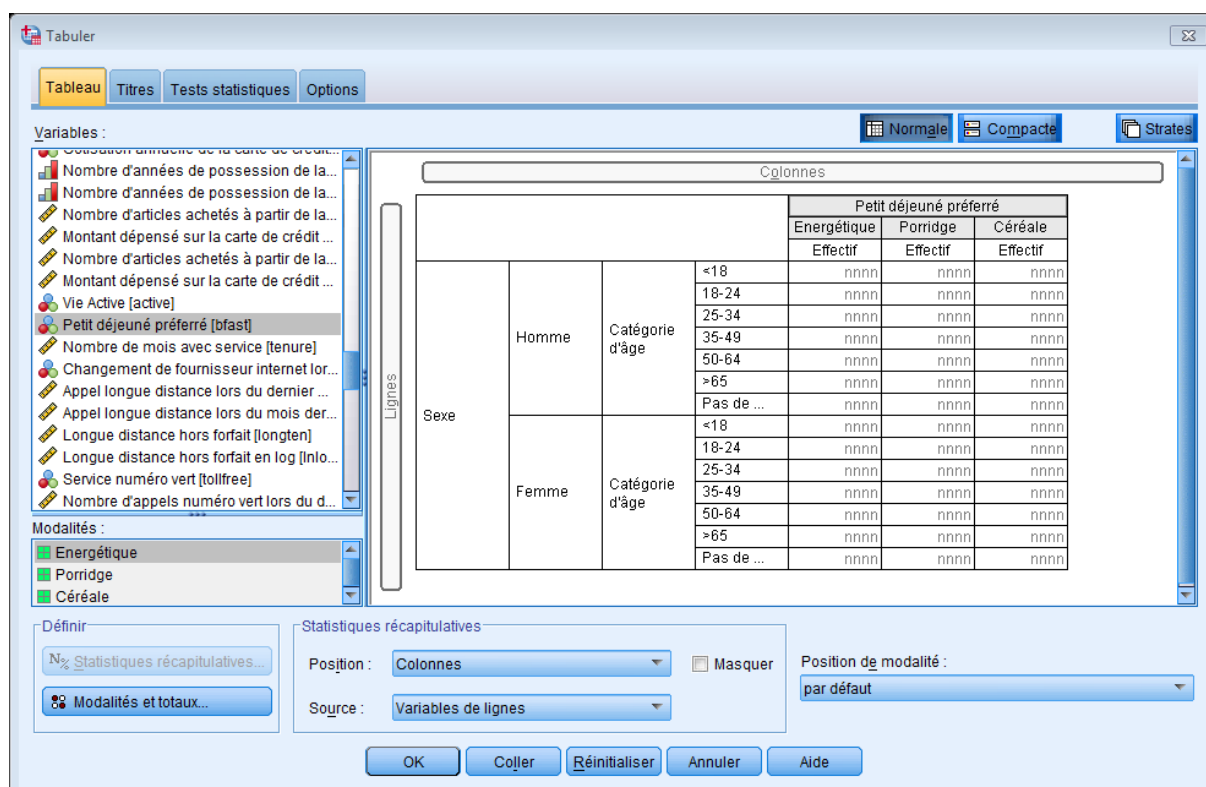
Dans la colonne ***N total % colonne***, pour chaque variable, les pourcentages sont obtenus en faisant le rapport entre le nombre de réponses « oui » (effectif) et le nombre total d'individu dans la base.

### 7.1.6 Tableaux avec variables emboîtées

Ces tableaux sont ceux où l'on a des variables qui sont imbriquées en colonne ou en ligne. Pour avoir des variables emboîtées en ligne, il suffit de les aligner de gauche à droite dans le générateur de graphique. Pour avoir des variables emboîtées en colonne, on les aligne de haut en bas.

A partir du fichier *customer\_dbase.sav* on va construire un tableau présentant les goûts en matière de petit déjeuner (breafst) suivant le sexe et la catégorie d'âge. Pour cela, nous allons emboîter les variables *sexe* et *catégorie d'âge* que nous allons croiser avec la variable *breafst* relative aux préférences au petit-déjeuner.

Ainsi, dans le générateur de graphique on dispose les variables de la manière suivante :



### Gestion des étiquettes

Lors de la construction des tableaux avec variables emboîtées, par défaut, les étiquettes de variables sont répétées avant la déclinaison des modalités. Ceci peut avoir un caractère redondant comme c'est le cas dans la vue précédente avec l'étiquette « catégorie d'âge ». En effet, les modalités de cette variable sont suffisamment descriptives si bien que l'ajout de l'étiquette variable devient superflu. Dans ce cas, il est donc opportun de supprimer cette étiquette. Pour cela :

- ✓ A partir du générateur de tableau, on sélectionne les cellules catégories d'âge,
- ✓ On fait un clic droit et dans le menu contextuel, on décoche la case affichée l'étiquette de la variable.

On peut également supprimer l'étiquette de la variable sexe pour les mêmes raisons.

- ✓ Après on valide en cliquant sur OK,

Le résultat obtenu est le suivant :

		Petit déjeuner préféré		
		Energétique	Porridge	Céréale
		Effectif	Effectif	Effectif
Homme	<18	0	0	0
	18-24	195	2	123
	25-34	218	12	184
	35-49	231	110	274
	50-64	119	261	215
	>65	38	382	118
	Pas de réponse	0	0	0
Femme	<18	0	0	0
	18-24	167	3	132
	25-34	231	28	220
	35-49	213	99	295
	50-64	121	261	218
	>65	49	385	96
	Pas de réponse	0	0	0

L'utilisateur peut cependant vouloir néanmoins afficher les étiquettes de variables dans le tableau sans toutefois qu'ils apparaissent à plusieurs endroits. Dans ce cas il peut les inclure dans le titre du graphique ou encore, dans l'étiquette de coin du tableau.

Dans le générateur de graphique,

- ✓ Sélectionnez l'onglet Titre
- ✓ Placez le curseur dans la zone de texte Coin
- ✓ Cliquez sur Expression de tableau. cela fait apparaître &[Expression de tableau] dans la zone de texte
- ✓ Cliquez sur OK pour créer le tableau

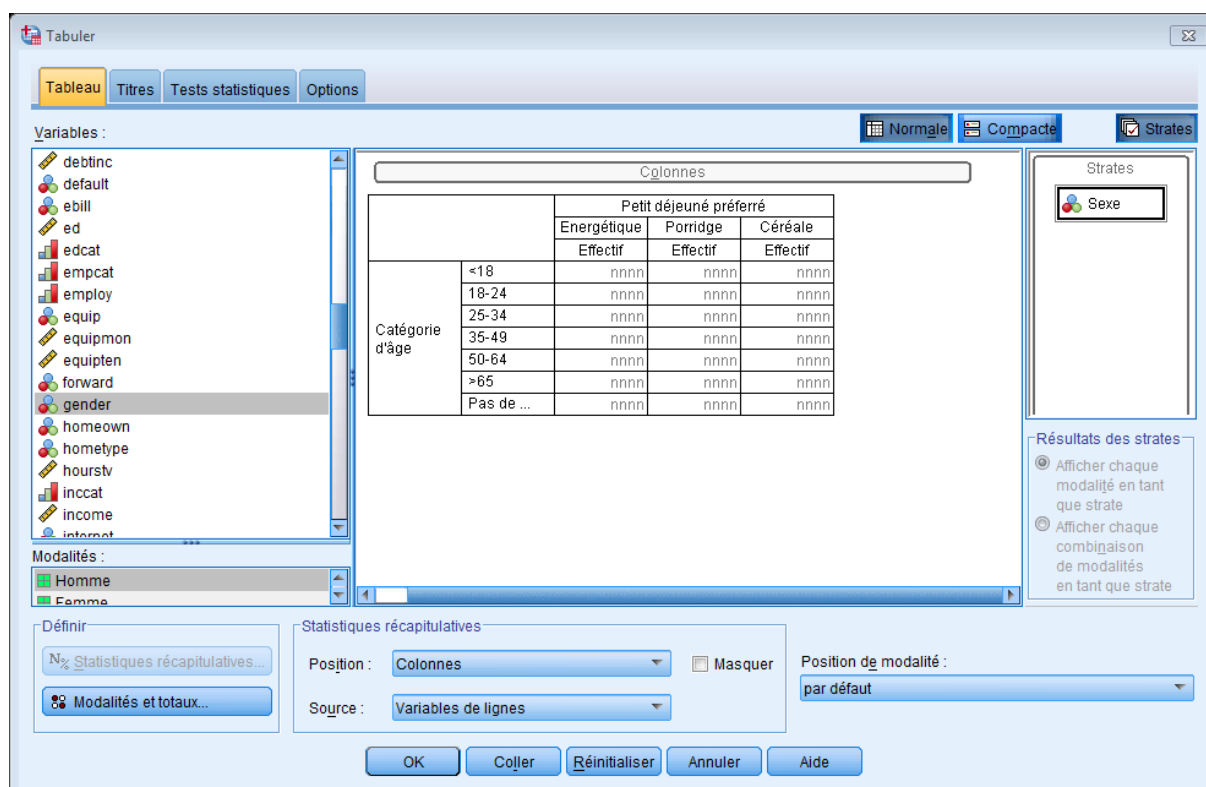
Sexe > Catégorie d'âge BY Petit déjeuner préféré		Petit déjeuner préféré		
		Energétique	Porridge	Céréale
		Effectif	Effectif	Effectif
Homme	<18	0	0	0
	18-24	195	2	123
	25-34	218	12	184
	35-49	231	110	274
	50-64	119	261	215
	>65	38	382	118
	Pas de réponse	0	0	0
Femme	<18	0	0	0
	18-24	167	3	132
	25-34	231	28	220
	35-49	213	99	295
	50-64	121	261	218
	>65	49	385	96
	Pas de réponse	0	0	0

### 7.1.7 Tableau avec strate

Les strates tout comme l'emboîtement permet d'ajouter une dimension supplémentaire au tableau. La différence entre ces deux approches se situe essentiellement au niveau de l'affichage. En effet, alors que dans le cas de l'empilement le tableau est entièrement présenté, dans le second cas, le tableau est présenté suivant l'occurrence de la variable strate.

A titre d'illustration nous allons reprendre l'exemple précédent où nous avons construit le tableau en croisant les variables sexe, catégorie d'âge et petit déjeuner à la seule différence que cette fois ci la variable *sexe* sera utilisée comme strate.

- ✓ Initialisez tous les onglets du générateur de tableau, en cliquant sur **Réinitialiser** puis, sur **Tous les onglets**.
- ✓ Sélectionnez en ligne la variable catégorie d'âges et en colonne la variable petit déjeuner préféré
- ✓ Cliquez sur Strates dans le coin supérieur droit du générateur de tableaux. Cela permet de pouvoir spécifier les paramètres de création des strates.
- ✓ Sélectionnez la variable sexe et faites la glissez vers le cadre strates.



✓ Cliquez sur OK pour créer le tableau.

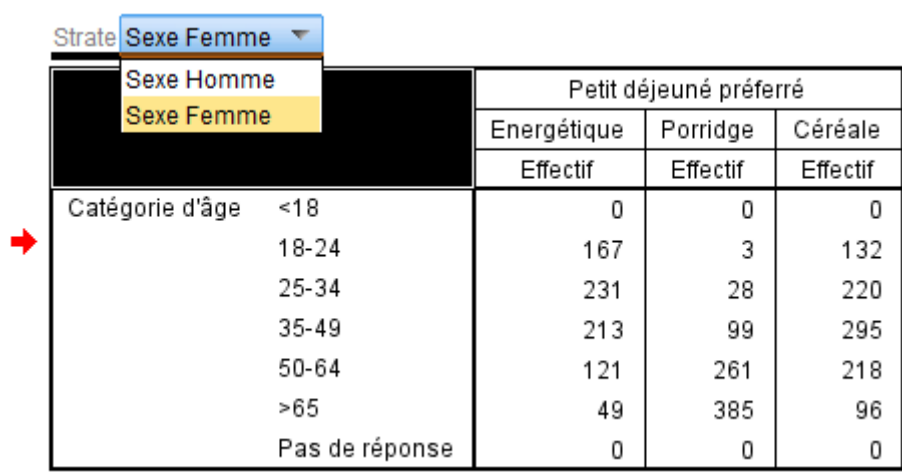
Le résultat obtenu dans le Viewer est le tableau suivant :

Sexe Homme		Petit déjeuner préféré		
		Energétique	Porridge	Céréale
		Effectif	Effectif	Effectif
Catégorie d'âge	<18	0	0	0
	18-24	195	2	123
	25-34	218	12	184
	35-49	231	110	274
	50-64	119	261	215
	>65	38	382	118
	Pas de réponse	0	0	0

On remarque que ce tableau correspond à la première partie du tableau précédent obtenue dans la section emboîtement. En effet, l'étiquette Sexe Homme au-dessus du tableau indique que les données présentées dans le tableau ne concernent que les hommes.

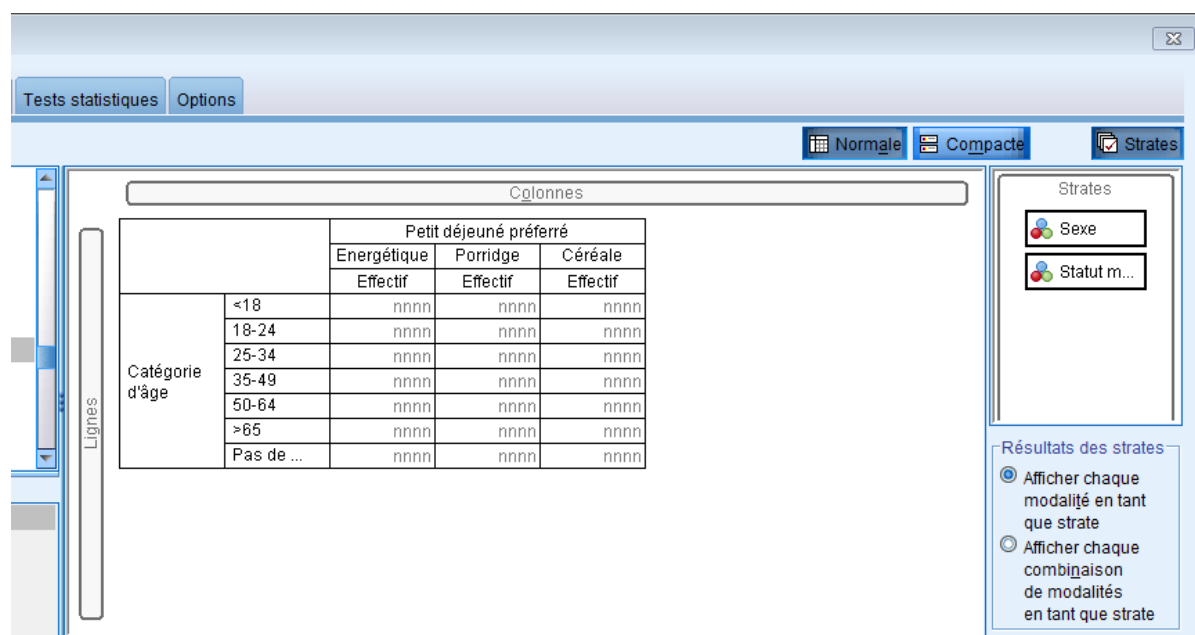
Pour avoir accès aux différentes strates, à partir du Viewer, double cliquez sur le tableau pour activer ce dernier.

On note alors que le champ strate prend alors l'aspect d'une liste déroulante qui permet à l'utilisateur de choisir la strate à afficher.



		Petit déjeuner préféré		
		Energétique	Porridge	Céréale
		Effectif	Effectif	Effectif
Catégorie d'âge	<18	0	0	0
	18-24	167	3	132
	25-34	231	28	220
	35-49	213	99	295
	50-64	121	261	218
	>65	49	385	96
	Pas de réponse	0	0	0

Par ailleurs, il est possible d'utiliser conjointement plusieurs variables de strate. L'utilisateur peut alors créer des strates empilées ou des strates emboîtées. Dans la vue suivante, on voit un cas d'exemple où les variables *sexe* et *statut marital* sont utilisées comme variables de strate.



Le cadre **Résultats des strates** permet distinguer les cas des strates empilées de ceux des strates emboîtées.

L'option « *Afficher chaque modalité en strate* », qui se trouve être la sélection par défaut, correspond à l'empilement des strates tandis que la seconde « *Afficher chaque combinaison de modalités en strate* » équivaut à l'emboîtement.

### 7.1.8 Tableau avec variable d'échelles source des statistiques récapitulatives

Dans cette section, nous allons nous intéresser au cas particulier où l'on cherche à calculer les statistiques d'une variable continue pour différentes sous-populations définies à partir de variables catégorielles. Par exemple, on peut calculer les revenus médian et moyen suivant la tranche d'âge et le sexe. Dans ce cas, les variables *sexe* et *tranche d'âge* sont utilisées pour



partitionner l'ensemble des individus en sous-population. Puis, les statistiques portant sur la moyenne et la médiane du revenu sont calculées dans chacune des sous-populations.

A partir du générateur de tableau:

- ✓ Sélectionnez la variable *tranche d'âge* et positionnez-la dans la zone ligne
- ✓ Faites glisser la variable *sexe* et déposez la zone colonne
- ✓ Sélectionnez la variable **revenu**, faites un glisser-déposer et positionnez la sous la variable sexe.

Tabuler















Tableau

Titres

Tests statistiques

Options

Variables :

-  Catégorie d'em...
-  Membre d'un sy...
-  Nombre d'anné...
-  Nombre d'anné...
-  Retraité(e) [re...
-  Revenu du foye...
-  Revenu en log [...]
-  Catégories de r...
-  Ratio dette / rev...
-  Débit carte de c...
-  Débit carte de c...
-  Autres dettes e...
-  Autres dettes e...
-  Refus de prêt b...

Lignes

Colonnes

		Sexe	
		Homme	Femme
		Revenu du ..	Revenu du ..
		Moyenne	Moyenne
Catégorie d'âge	<18	nnnn,nn	nnnn,nn
	18-24	nnnn,nn	nnnn,nn
	25-34	nnnn,nn	nnnn,nn
	35-49	nnnn,nn	nnnn,nn
	50-64	nnnn,nn	nnnn,nn
	>65	nnnn,nn	nnnn,nn
	Pas de ...	nnnn,nn	nnnn,nn

Modalités :

- ✓ Dans le panneau de canevas, sélectionnez la variable revenu puis faites un clic-droit et choisissez **Statistiques récapitulatives**.
- ✓ Dans liste des statistiques, choisissez médiane et ajouter à la liste des statistiques à afficher.

Variable sélectionnée : Revenu du foyer en milliers (\$)

Statistiques :

Statistiques	Etiquette	Format	Décima...
Moyenne	Moyenne	Auto	
Médiane	Médiane	Auto	

Buttons: Appliquer à la sélection, Appliquer à tout, Fermer, Aide

- ✓ Cliquez sur le bouton **Appliquer à la sélection** pour valider vos choix, puis sur le bouton OK dans le générateur de tableau pour implémenter celui-ci.

		Sexe			
		Homme		Femme	
		Revenu du foyer en milliers (\$)		Revenu du foyer en milliers (\$)	
		Moyenne	Médiane	Moyenne	Médiane
Catégorie d'âge	<18	.	.	.	.
	18-24	23,99	20,00	23,37	21,00
	25-34	39,17	31,00	38,44	31,00
	35-49	59,23	48,00	58,68	47,00
	50-64	85,42	62,00	79,49	62,00
	>65	48,61	28,00	52,32	29,00
	Pas de réponse	.	.	.	.

## 7.2 Modification des tableaux

Les tableaux produits (ou plus généralement les résultats de procédures statistiques) dans le Viewer sont affichés dans les tableaux pivotants. Ceux-ci permettent d'avoir accès à de nombreuses fonctionnalités. Par exemple en laissant le pointeur de la souris sur les termes statistiques affichés tableau, il est possible d'avoir une définition de ces derniers via une fenêtre contextuelle.

Les tableaux pivotant offrent également de grandes possibilités de modification des tableaux. C'est cet aspect qui fera l'objet de cette section.

### 7.2.1 Modification de la structure d'un tableau

Il arrive parfois après la construction d'un tableau que l'on se rende compte que la présentation des données n'a pas la forme escomptée. On peut par exemple trouver le tableau trop long ou trop. Les tableaux pivotants permettent de changer les structures des données en modifiant par exemple le positionnement des variables.

Considérons le tableau suivant où l'on croise les variables *catégories d'âge* et *sexe*. Prenons en ligne la variable *sexe* et en colonne *catégorie d'âge*. Le tableau généré dans le Viewer se présente comme suit :

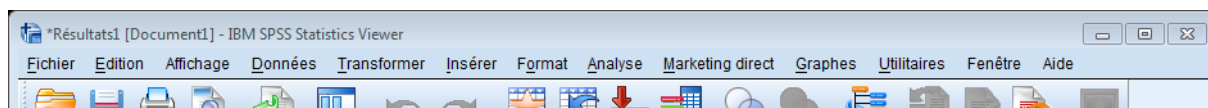
		Catégorie d'âge						
		<18	18-24	25-34	35-49	50-64	>65	Pas de réponse
		Effectif	Effectif	Effectif	Effectif	Effectif	Effectif	Effectif
Sexe	Homme	0	320	414	615	595	538	0
	Femme	0	302	479	607	600	530	0

L'utilisateur pourrait trouver ce tableau un peu trop court et décider de transposer les lignes et les colonnes.

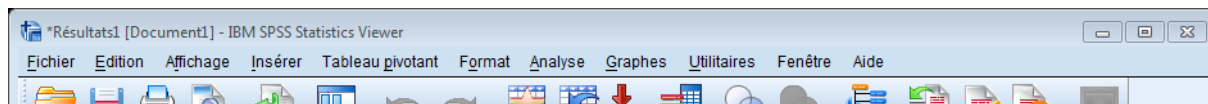
Pour cela activez le tableau, en double-cliquant sur ce dernier.

La modification du tableau peut se faire directement dans le Viewer ou bien dans une autre fenêtre indépendante. Cette dernière option est celle par défaut, dans le cas des grands tableaux.

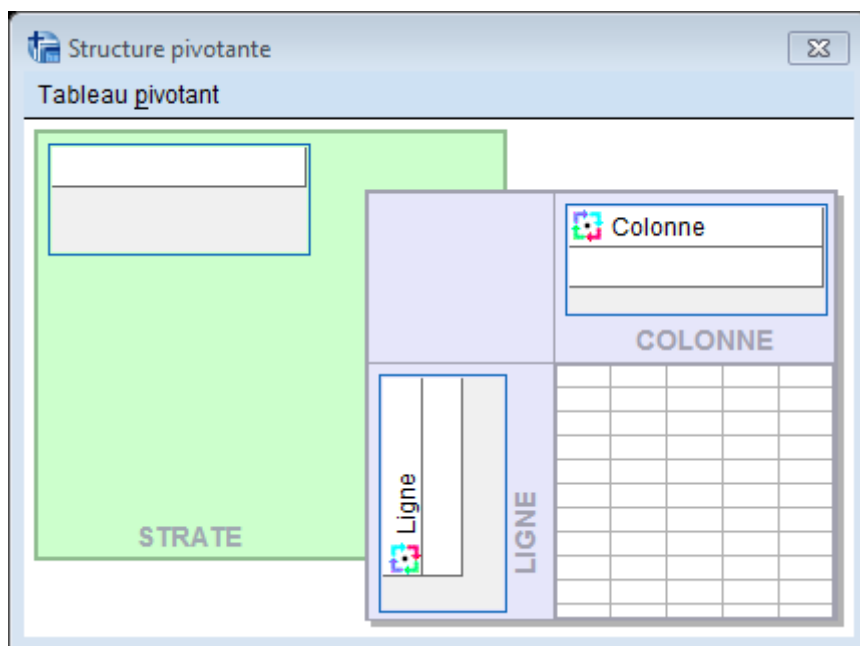
Lorsque la modification se fait dans le Viewer, on constate que, la barre des menus du Viewer change. En effet, la barre des menus initiale ci-dessous :



Après activation du tableau, se transforme comme suit, avec notamment l'apparition du menu Tableau pivotant :



Afin de pouvoir transposer les lignes et les colonnes assurez-vous que la fenêtre Structure pivotante est visible. Dans le cas contraire, à partir du menu, sélectionnez Tableau pivotant et cochez la case structure pivotante.



En faisant glisser les éléments de ligne vers la colonne et les éléments de colonne vers la ligne, ceux-ci seront alors transposés.

### 7.2.2 Modification et mise en forme des étiquettes des tableaux

Pour modifier ou mettre en forme les étiquettes, il faut tout d'abord activer le tableau.

- ✓ Faites apparaître, la barre d'outils de formatage. Pour cela, à partir des menus sélectionnez :

**Affichage>Barre d'outils**

- ✓ Sélectionnez le texte à modifier et utilisez la barre d'outils pour apporter les modifications, notamment la police, la taille de la police, la couleur et l'alignement du texte.

Pour modifier un titre ou une étiquette, double-cliquez sur l'objet et saisissez le nouveau texte.

### 7.2.3 Modification des formats d'affichage des données

Il s'agit ici de modifier les formats des données affichées dans un tableau pivotant. Pour illustrer la démarche, nous allons construire le tableau suivant où l'on croise les variables *catégorie d'âge* et *sexe* en ressortant les statistiques effectifs et pourcentage ligne.

		Sexe			
		Homme		Femme	
		Effectif	N % ligne	Effectif	N % ligne
Catégorie d'âge	<18	0	0,0%	0	0,0%
	18-24	320	51,4%	302	48,6%
	25-34	414	46,4%	479	53,6%
	35-49	615	50,3%	607	49,7%
	50-64	595	49,8%	600	50,2%
	>65	538	50,4%	530	49,6%
	Pas de réponse	0	0,0%	0	0,0%

Nous allons modifier le format d'affichage des pourcentages dans la colonne « Homme ». Pour cela :

- ✓ Activez le tableau, en double-cliquant sur celui-ci.
- ✓ Cliquez sur la cellule *N% ligne* de la colonne « Homme »
- ✓ A partir du menu sélectionnez :

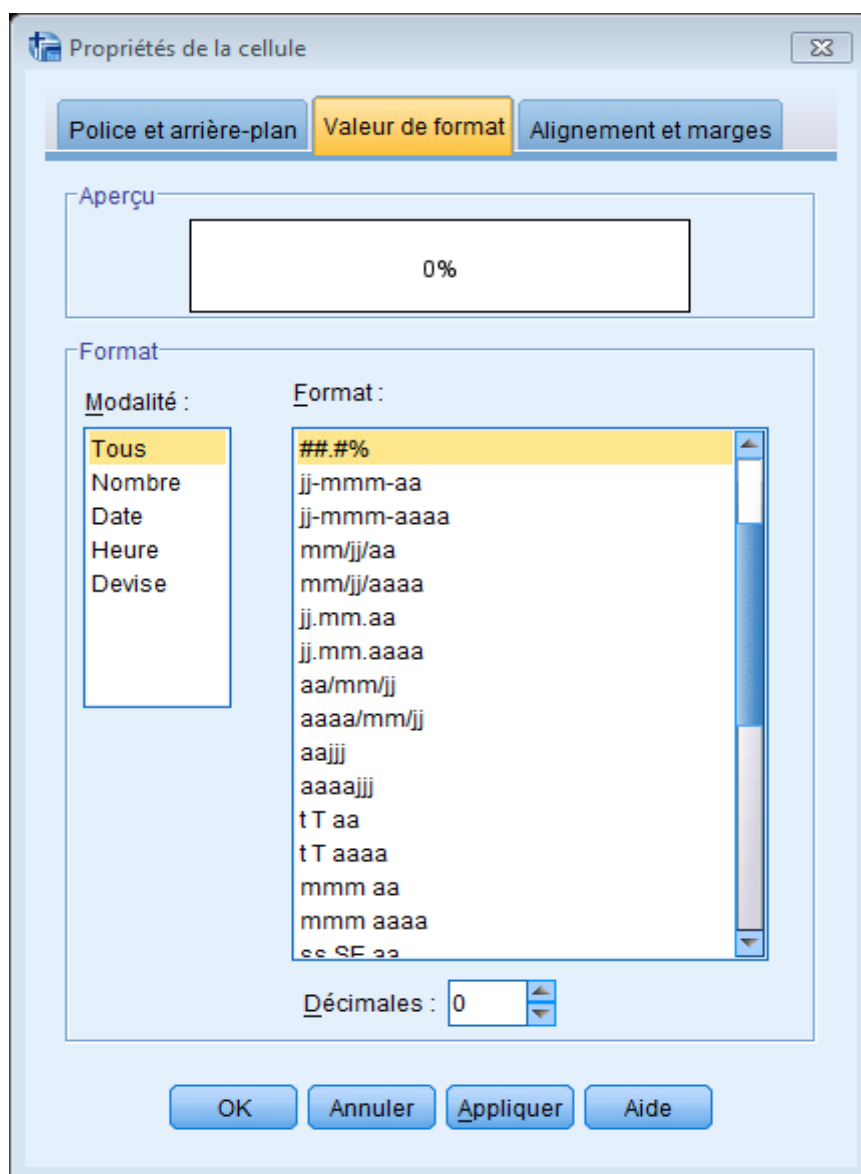
**Edition> Sélectionnez > Cellules de Données**

- ✓ Dans le menu Format, choisissez Propriétés de la cellule

La fenêtre Propriétés de la Cellule dispose de trois onglets police et arrière-plan ; valeur de format ; ainsi que Alignement et marges

- ✓ Sélectionnez l'onglet Valeur de format.

On retrouve dans cet onglet, la liste de formats pouvant être appliqués aux données des cellules sélectionnées. Dans le cas d'exemple, on se contentera juste de supprimer les décimales en mettant comme nombre décimales 0.



✓ Validez avec OK.

### 7.2.4 Masquer des lignes ou des colonnes

Il arrive après élaboration du tableau que l'on se rende compte que certaines lignes ou colonnes sont superflues. On peut donc décider de les masquer sans perdre de données.

Par exemple, dans le tableau précédemment utilisé à titre d'illustration, les lignes **<18** et **Pas de réponse** sont superflues.

		Sexe	
		Homme	Femme
		Effectif	Effectif
Catégorie d'âges	<18	0	0
	18-24	320	302
	25-34	414	479
	35-49	615	607
	50-64	595	600
	>65	538	530
	Pas de réponse	0	0

On va donc masquer la ligne <18. Pour cela :

- ✓ Activer le tableau ;
- ✓ Sélectionnez la cellule <18
- ✓ A partir du menu, faites

**Edition > Sélectionnez > Cellules de données et d'étiquettes**

- ✓ A partir du menu,

**Affichage > Masquer**

On obtient donc le résultat suivant :

		Sexe	
		Homme	Femme
		Effectif	Effectif
Catégorie d'âges	18-24	320	302
	25-34	414	479
	35-49	615	607
	50-64	595	600
	>65	538	530
	Pas de réponse	0	0

### 7.2.5 Modèle de tableau

Les modèles de tableaux peuvent être utilisés pour modifier l'aspect général. Cela peut s'avérer utile pour une présentation optimale des données.

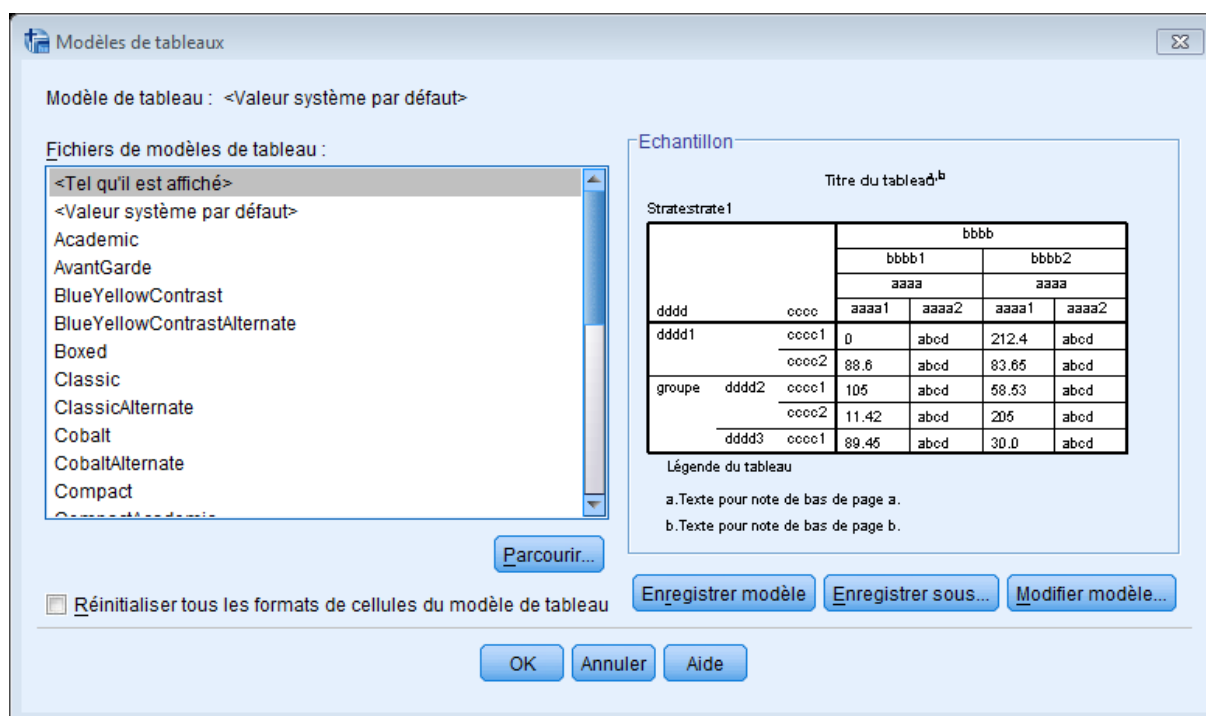
#### 7.2.5.1 Utilisation de modèles prédéfinis

SPSS fournit un grand nombre de modèles qui peuvent être utilisés sur différents tableaux. Pour appliquer un modèle prédéfini à un tableau :

- ✓ Activez le tableau
- ✓ A partir des menus, sélectionnez :

**Format > Modèles de tableaux**

La fenêtre Modèles de tableaux présente



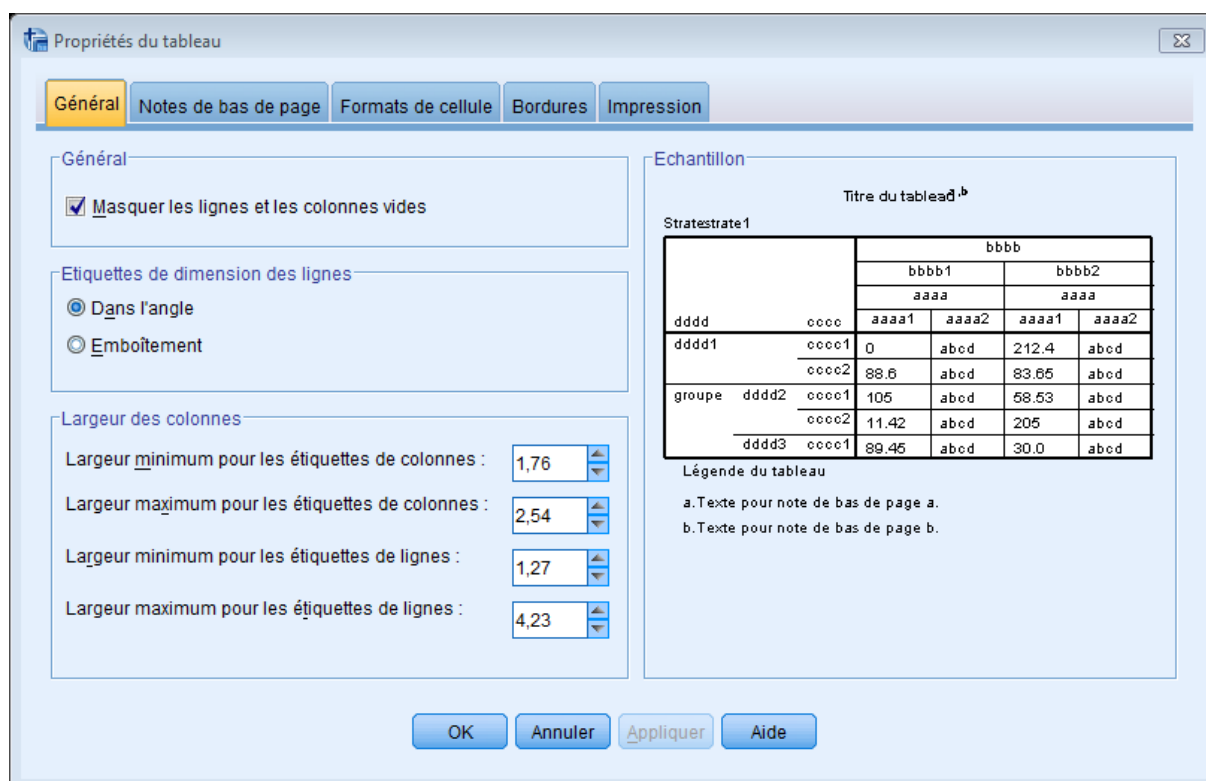
- ✓ Sélectionnez le style à appliquer et cliquer sur OK.

### 7.2.5.2 Utilisation des styles personnalisés

Pour utiliser un style personnalisé, il suffit de sélectionner le style prédéfini se rapprochant le plus possible du style désiré et modifier le.

Pour modifier un style prédéfini, dans la fenêtre modèle de tableaux, cliquez sur le bouton Modifier modèle

La fenêtre propriété du tableau apparait pour permettre à l'utilisateur de définir les différentes propriétés qui seront intégrées au modèle.



### 7.2.5.3 Définir les styles par défaut

Il est également possible de définir les styles qui seront appliqués par défaut lors de la construction des tableaux.

- ✓ A partir du menu du Viewer, faites :

**Edition > options**

Les onglets tableaux pivotants et Résultats de la fenêtre options permettent de définir les styles qui seront utilisés par défaut

## 8 Utilisation des résultats

### 8.1 Éléments d'informations supplémentaires sur le Viewer

A titre de rappel, le Viewer présente deux panneaux. Un panneau de navigation et un panneau de contenu où sont affichés les résultats (tableaux, graphique,...).

Le panneau de navigation présente une arborescence qui permet à l'utilisateur de pouvoir parcourir les différents résultats produits. Dans cette arborescence, les résultats élémentaires possèdent une icône représentant un livre ouvert. Pour masquer un résultat il suffit de double-cliquer sur livre. L'icône devient alors un livre fermé.

Les signes – dans l'arborescence permettent également de réduire l'arborescence. En cliquant sur ce signe, il se transforme en signe +, la branche dépendante est réduite et les résultats de la procédure concernée sont masqués. Pour développer à nouveau la branche et afficher les résultats il suffit de cliquer sur +.



Il est également possible de modifier l'ordre d'affichage des résultats à partir du panneau de navigation. Il suffit de faire un cliquer glisser vers l'emplacement de destination.

Après avoir édité les résultats de manière appropriée, la page du Viewer peut être imprimée ou enregistrée sous un fichier d'extension « .spv » pour consultation ultérieure.

### 8.2 Exportations des résultats

Dans le cadre de production des rapports statistiques et de l'élaboration des présentations on est généralement amené à exploiter les résultats sous d'autres formats. En d'autres termes, l'on sera souvent amené à utiliser les résultats produits par SPSS dans d'autres applications. Dans cette section, nous nous concentrerons principalement sur les logiciels de Microsoft office particulièrement sur Word et Excel.

#### 8.2.1 Copier/coller vers Excel et Word par l'intermédiaire du presse-papier

Cette méthode est l'une des plus intuitives. En effet, il suffit de :

- ✓ Sélectionner le résultat que l'on veut exporter
- ✓ Dans le menu Edition ou à partir de menu contextuel, sélectionner **copier**
- ✓ Dans l'application de destination, coller à l'endroit approprié.

Dans les cas où on l'on a comme résultats des tableaux et comme logiciel de destination Microsoft Word, trois options de collage spécial sont disponibles : texte mis en forme (RTF), Texte sans mise en forme ; image. Toutefois la 1<sup>ère</sup> option se trouve être la plus fréquente.

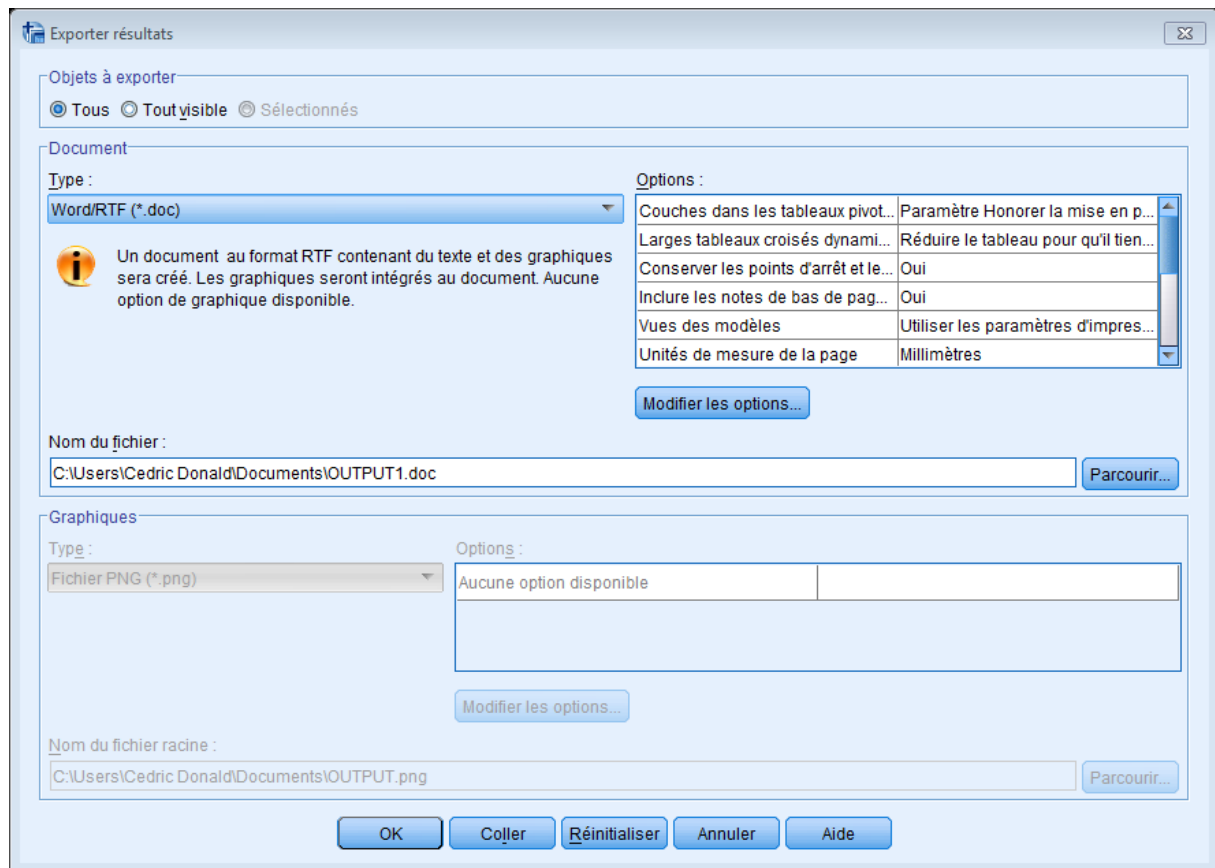
#### 8.2.2 Exportation des résultats vers un fichier Microsoft Word, PowerPoint ou Excel

A la différence de la démarche précédente, il ne s'agit plus de copier les données un à un mais de copier le contenu du Viewer vers d'autres applications. Pour ce faire :

- ✓ A partir du menu du Viewer, sélectionnez :

**Fichier > Exporter...**

Cela ouvre la fenêtre Exporter résultats, qui permet de définir les paramètres d'exportation.



Le cadre **Objets à exporter** permet de sélectionner les résultats à exporter. Trois choix sont proposés à l'utilisateur. Il peut exporter soit (i) tous les objets du Viewer, soit (ii) n'exporter que les objets visibles, ou encore (iii) exporter uniquement les objets sélectionnés dans le panneau de légende.

La liste déroulante **Type** permet de spécifier vers quel type de fichier l'on veut exporter les résultats. (On peut sélectionner des fichiers Word ou Excel et bien d'autres encore).

#### Cas particulier des graphiques

Il est possible d'exporter les graphiques dans des fichiers images. Pour cela, il suffit de choisir dans la fenêtre d'exportation l'option « Aucun (diagrammes uniquement) » au niveau de la liste déroulante **Type** du Cadre **document**.

Une fois ce choix effectué, l'on sélectionne le format des fichiers images vers lesquels seront exportés les graphiques. Cette sélection s'effectue via la liste déroulante **Type** du cadre **Graphiques**.

## 9 Programmer avec SPSS

SPSS offre la possibilité de pouvoir automatiser différentes tâches et procédures à partir du la fenêtre éditeur de syntaxe précédemment évoquée. Le langage de programmation très développé de SPSS donne de pouvoir mettre en œuvre toutes les opérations accessibles via les menus et fenêtre. De plus, il fournit également des fonctionnalités qui ne se trouvent ni dans les menus ni dans les boîtes de dialogue.

Outre cet atout, l'utilisation des fichiers de syntaxe présente l'avantage de permettre de consigner par écrit, dans les moindres détails de toutes les étapes du travail réalisé. L'utilisation des fichiers est plus que conseillé pour toute personne désirant se perfectionner dans l'utilisation de SPSS.

Un fichier de commande SPSS c'est juste un fichier texte contenant des instructions dans le langage de programmation de SPSS. L'extension couramment utilisée pour les fichiers de syntaxe SPSS est « .spo ». C'est également ce type de fichier qui est généré par défaut par l'éditeur de syntaxe.

### 9.1 Règles de la syntaxe SPSS

Les fichiers de syntaxe SPSS peuvent être exécutés de deux façons :

- Execution en mode interactif : ici le fichier de syntaxe est ouvert et l'utilisateur choisit les instructions qu'il veut exécuter
- Exécution via la commande *INCLUDE* : ici le fichier de syntaxe est entièrement lu et toutes les instructions qui s'y trouvent sont exécutées.

Les règles de la syntaxe peuvent légèrement varier suivant le mode utilisé. Toutefois les règles présentées dans ce document se veulent plus restrictives de façon afin que la syntaxe puisse être utilisée dans les deux modes voire même avec des versions antérieures du logiciel qui auraient des règles plus strictes. **Ainsi, pour connaître les règles de programmation telles que définies pour la version 21 du logiciel, le lecteur peut se référer au Guide de l'utilisateur ou le manuel de référence de la syntaxe SPSS.**

Le fichier de syntaxe SPSS doit vérifier les règles suivantes :

- Chaque commande débute sur une nouvelle ligne et ne doit être précédé d'aucun espace (elle débute à la première colonne)<sup>1</sup>.
- Chaque commande doit se terminer par un point (.) servant de caractère de fin de commande<sup>2</sup>.
- Les sous-commandes ou options sont séparées par des barres obliques (/).
- Si plusieurs lignes sont utilisées pour une commande, la colonne 1 de chaque ligne suivante doit être vide.
- Chaque ligne de syntaxe de commande ne doit pas dépasser 256 octets.
- Le nom des variables doit être écrit en toute lettre
- Le texte inclus dans les apostrophes ou les guillemets doit tenir sur une seule ligne.
- Les noms de fichier et les adresses doivent être entre guillemets
- L'indicateur de décimal c'est le point (.), quelques soient les paramètres régionaux ou locaux

---

<sup>11</sup> En mode interactif et pour la version 21 de SPSS, il n'est pas obligatoire de débiter à la première colonne, excepté pour la commande END DATA

<sup>2</sup> Dans le mode interactif de la version 21 de SPSS, en l'absence d'un point servant de caractère de fin de commande, toute ligne vide est interprétée comme un caractère de fin de commande.

- Le nom des variables finissant par un point peut engendrer des erreurs dans les commandes créées par les boîtes de dialogue<sup>3</sup>

## 9.2 La fenêtre Editeur de syntaxe

### 9.2.1 Ouverture d'un fichier de syntaxe

L'ouverture d'un fichier de syntaxe peut s'effectuer à partir des trois principales fenêtres de SPSS (Editeur de données, Viewer, fenêtre syntaxe).

Dans le cas de l'ouverture d'un nouveau fichier :

- ✓ A partir des menus, sélectionnez :

**Fichier>Nouveau>Syntaxe**

Dans le cas de l'ouverture d'un fichier existant :

- ✓ A partir des menus, sélectionnez :

**Fichier>Ouvrir>Syntaxe**

Les fichiers syntaxes dans SPSS portent l'extension «.sps ».

### 9.2.2 Utilisation de la fenêtre syntaxe

L'éditeur de syntaxe SPSS a grandement été amélioré au fil des versions. Elle offre de nombreuses fonctionnalités facilitant la création de syntaxe. Nous présenterons quelques-unes :

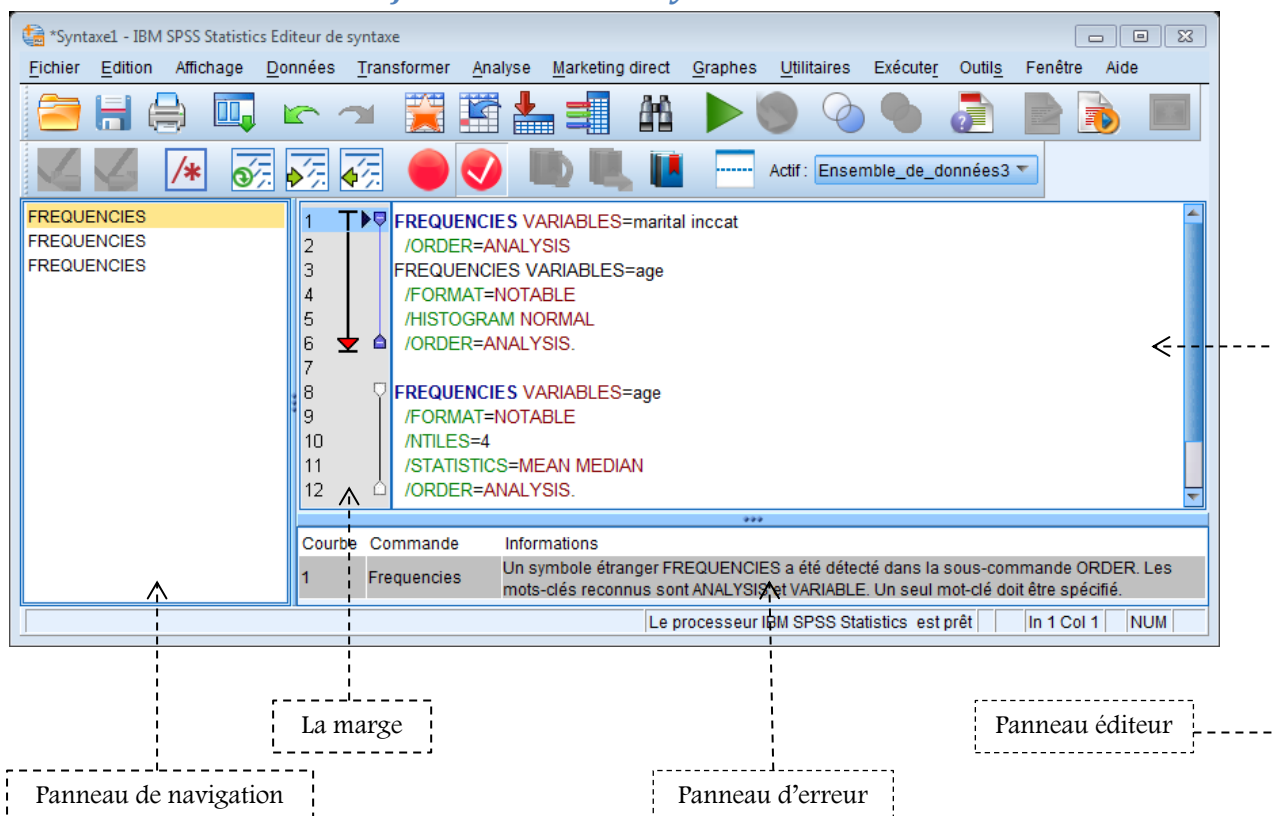
*Saisie semi-automatique* : lorsqu'on commence à saisir le nom d'une commande, d'une sous commande ou un mot clé SPSS propose une liste contextuelle de termes qui pourraient correspondre à ce que l'on a l'intention de saisir

*Le Codage par couleur* : Les éléments reconnus de la syntaxe de commande (commandes, sous-commandes, mots-clés et valeurs de mot-clé) ou encore certaines erreurs telles que les « les guillemets non fermés » sont mis en exergue via les couleurs de leur police. Cela s'avère très utile, particulièrement lorsque l'utilisateur saisit directement la syntaxe ou modifie une syntaxe existante.

---

<sup>3</sup> Dans des versions antérieures de SPSS l'utilisation du caractère point (.) dans le nom des variables est tout simplement proscrite.

### 9.2.2.1 Présentation de la fenêtre Editeur de syntaxe



La fenêtre syntaxe est divisée en quatre zones :

- **Le panneau éditeur** : c'est la partie principale de la fenêtre Editeur de syntaxe. C'est dans cette partie que figurent les instructions de commande.
- **Le panneau de navigation** : il référence toutes les commandes se trouvant dans le panneau éditeur et permet d'aller d'une commande à l'autre par simple clic. Les commandes y sont listées dans l'ordre avec lequel elles apparaissent dans le panneau éditeur. Le premier mot de chaque ligne de texte non reconnu y figure en gris.
- **Le panneau d'erreur** : invisible par défaut, il apparaît dès lors qu'une erreur se produit. Il affiche les erreurs d'exécution survenues lors de l'exécution la plus récente. Il contient des informations très détaillées qui sont très souvent utiles pour effectuer des corrections. Il indique notamment le numéro de la première de la commande à l'origine de l'erreur, le nom de la commande et d'autres informations sur la cause de l'erreur.  
Il est possible de cliquer sur une entrée du panneau d'erreur. Cela positionne le curseur sur la première ligne de la commande qui a généré l'erreur correspondante à l'entrée sélectionnée.
- **La marge** : elle contient des informations telles que le numéro de ligne, la position des points d'insertion ou encore la délimitation des commandes.

### 9.2.3 Exécution interactive d'un fichier de commandes.

Le menu Exécuter (Run) de la fenêtre de l'éditeur de syntaxe offre quatre possibilités concernant l'exécution d'un fichier de commandes :

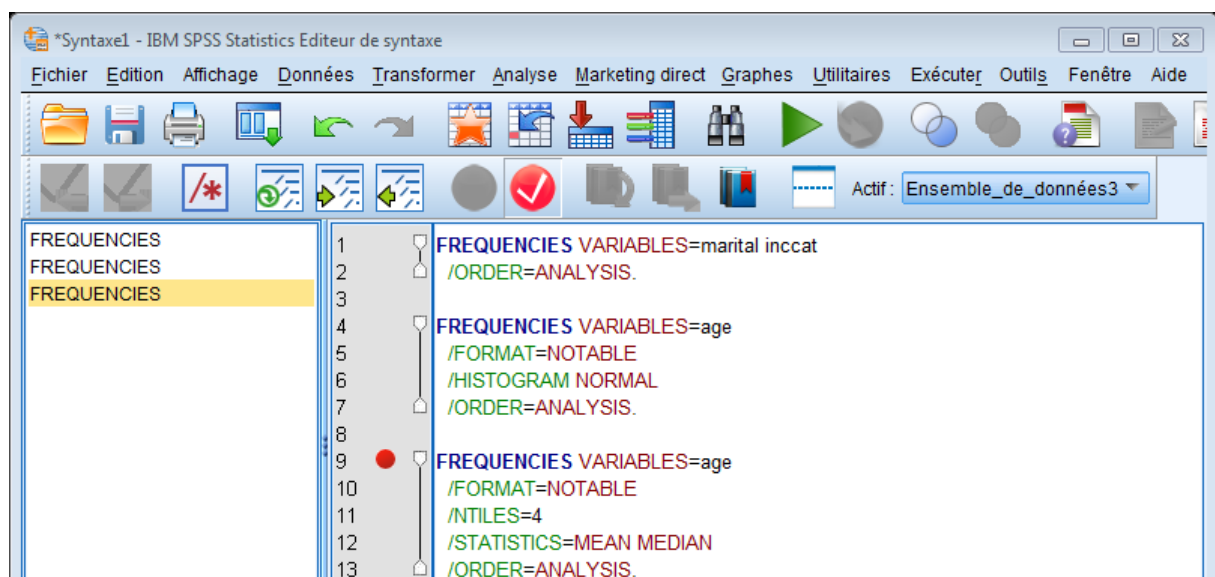
- Exécuter toutes les commandes du fichier
- Exécuter les commandes sélectionnées. Cette option figure également dans la barre d'outils de la fenêtre éditeur de syntaxe. Elle correspond au bouton de commande qui portant le symbole flèche ou triangle (on retrouve également ce symbole sur les boutons « Play » des lecteurs de musique)
- Exécuter depuis l'emplacement du curseur jusqu'à la fin du fichier
- Exécuter à partir de début ou bien à partir de l'emplacement du curseur jusqu'à l'arrêt. l'arrêt peut correspondre à la fin du fichier de commande ou à un point précis du fichier, spécifié par l'utilisateur. Dans ce dernier cas on parle de point d'arrêt

Lorsqu'on exécute une syntaxe une flèche orientée vers le bas apparait dans la marge de la fenêtre. Celle-ci représente la progression de l'exécution.

### *Utilisation des points d'arrêts.*

Le point d'arrêt permet d'interrompre l'exécution du fichier de syntaxe juste avant la commande contenant le point d'arrêt. En faisant usage, de cette fonctionnalité, l'utilisateur peut interrompre l'exécution de son fichier à différents points spécifiés. Cela permet d'effectuer des contrôles en vérifiant les résultats intermédiaires obtenus avant de poursuivre l'exécution.

Pour insérer un point d'arrêt, il suffit de cliquer sur dans la marge au niveau de la ligne où l'on veut l'insérer.



Pour poursuivre l'exécution après un point d'arrêt :

- ✓ A partir du menu de la fenêtre de l'Editeur de syntaxe, faites :

**Exécuter > Poursuivre**

### 9.3 Création d'un fichier de syntaxe

Pour créer un fichier de syntaxe, l'utilisateur ayant déjà une parfaite connaissance du langage de programmation SPSS n'aura qu'à saisir les instructions qu'il souhaite exécuter, dans le panneau de saisie de la fenêtre syntaxe.

Concernant l'utilisateur n'ayant pas une maîtrise avancée du langage de programmation, deux options s'offre à lui : (i) la copie de syntaxe à partir des fichiers résultats ou (ii) la création de syntaxe à partir des boîtes de dialogue.

Concernant la première option, il s'agit tout simplement de copier les instructions de commande apparaissant dans le Viewer et de les coller dans le fichier de syntaxe. En effet, lorsqu'un résultat est produit dans le Viewer, la syntaxe permettant produire ce résultat est également éditée dans le Viewer<sup>4</sup>.

La deuxième option relative à l'utilisation des boîtes de dialogue, constitue le moyen le plus simple de créer une syntaxe. En effet, la plupart des boîtes de dialogue, propose via le bouton **Coller** la fonctionnalité de « collage de syntaxe ». Pour une boîte de dialogue donnée, cela consiste à coller directement dans le fichier de syntaxe, l'instruction correspondant à l'analyse à effectuer.

A titre d'illustration, nous allons coller la syntaxe permettant d'effectuer des tris à plat (tableaux des effectifs) pour les variables *statut marital* et *catégorie de revenu* du fichier *demo.sav*.

- ✓ A partir du menu, sélectionnez :

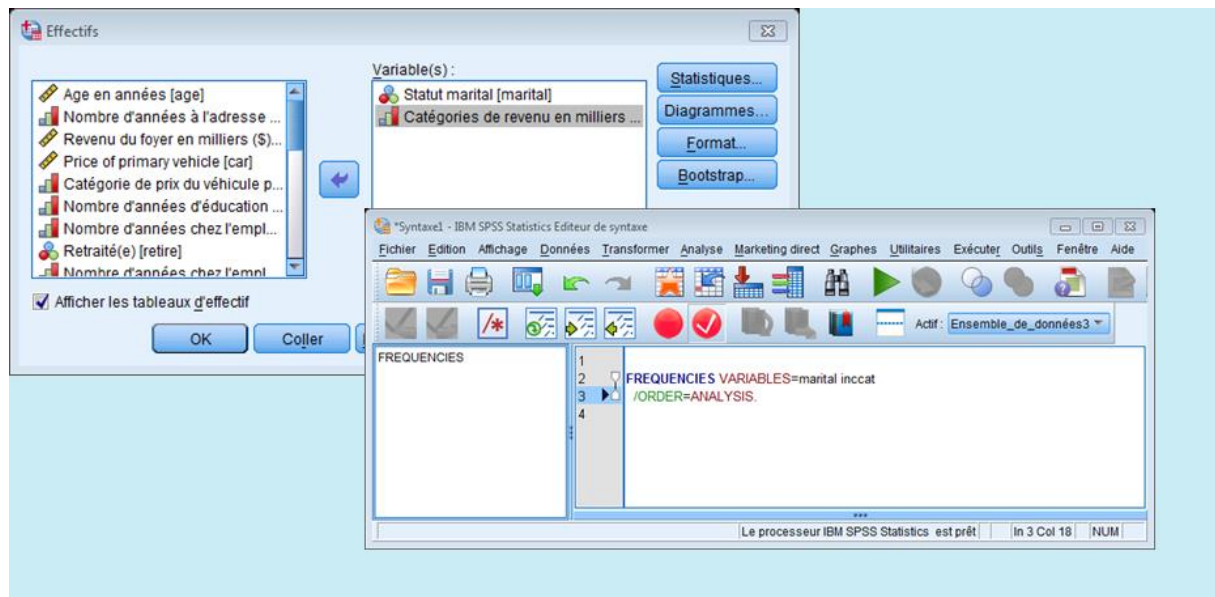
**Analyse>Statistiques descriptives>Effectifs**

- ✓ Sélectionnez les variables *statut marital* et *catégorie de revenu*
- ✓ Cliquez sur **coller**

La syntaxe correspondante est alors collée dans le fichier de syntaxe comme on peut le voir dans la vue ci-dessous.

---

<sup>4</sup> si ce n'est pas le cas, pensez à cocher l'option "Afficher commandes dans les paramètres du Viewer" (dans le menu Edition>Options, onglet Viewer) avant de lancer les analyses



Pour les utilisateurs n'ayant pas une parfaite maîtrise du langage, le processus d'élaboration d'un fichier de syntaxe se fera en deux étapes. La première consistera à créer la syntaxe à partir de la boîte de dialogue, la seconde consistera à modifier la syntaxe produite de façon à épouser parfaitement les desiderata l'utilisateur ou encore à avoir une meilleure lisibilité. Cette démarche constitue également un bon cadre d'apprentissage du langage de programmation.

La modification des syntaxes créées à partir des boîtes dialogues doit se faire en préservant les règles du langage de programmation présentées précédemment. L'utilisateur s'appuiera fortement sur l'aide pour la modification d'une syntaxe.

Pour obtenir l'aide sur une commande, il suffit de placer le curseur sur celle-ci et d'appuyer la touche F1 du clavier. Il est également possible d'aller rechercher la commande via le menu aide.