



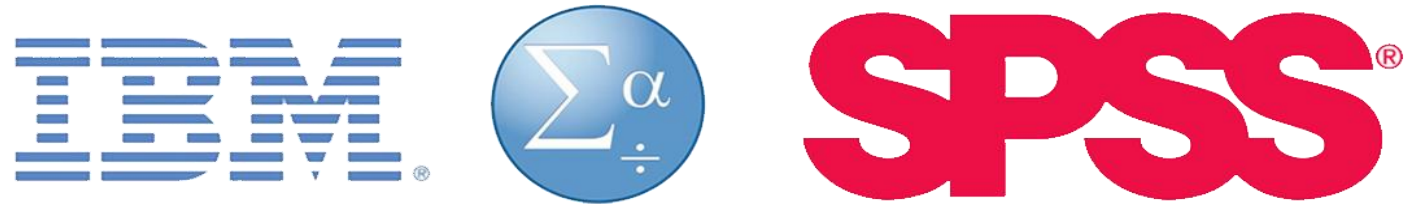
Module : Analyse de données avec SPSS

CERTIFICAT EXÉCUTIF - Analyse des données avec Excel avancé, SPSS, IA/Python & PowerBI

Premier pas

Qu'est-ce que SPSS ?

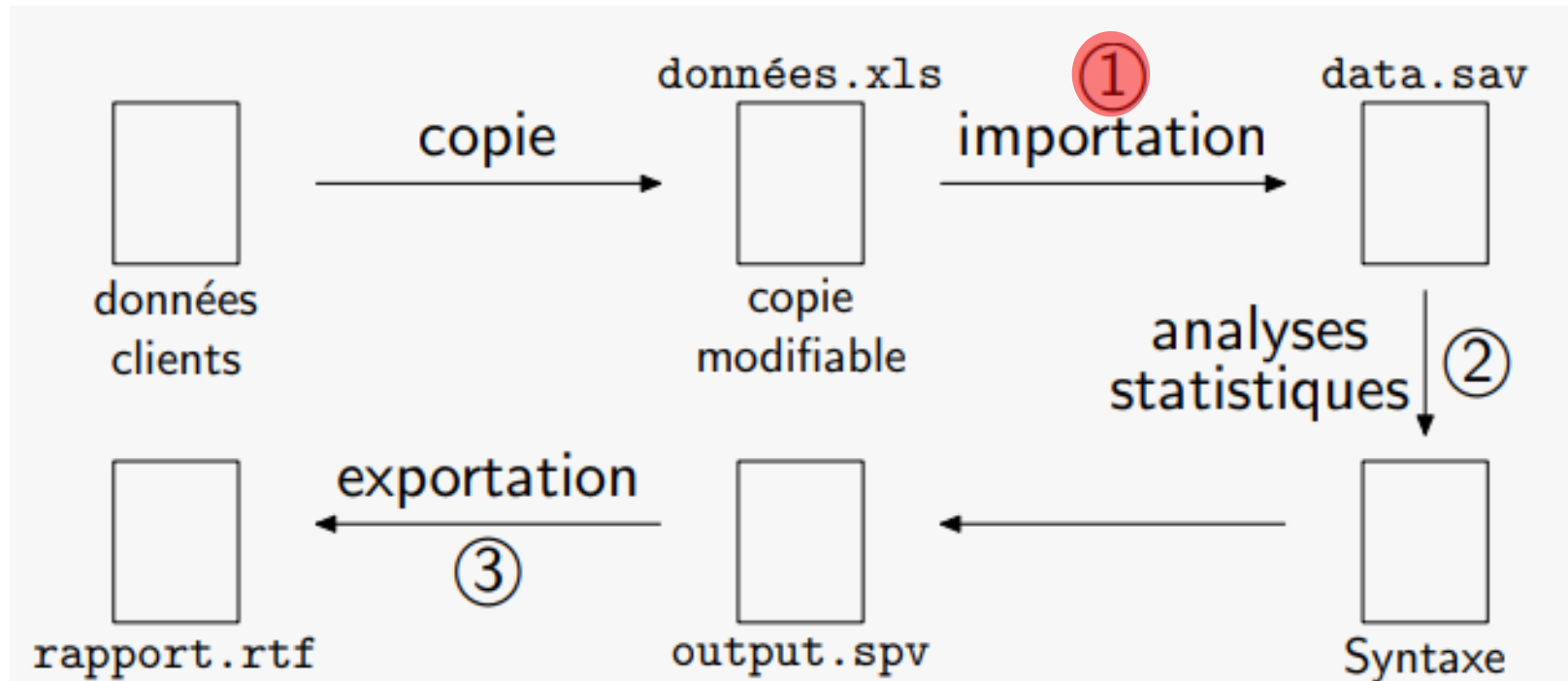
- ▶ SPSS : Statistical Package for the Social Sciences
- ▶ Logiciel pour **l'analyse statistique**.
- ▶ SPSS offre une large gamme de possibilité allant de la **préparation des données**, au **traitement** et à **l'analyse des données**. Il permet de répondre à la plupart des problèmes statistiques et son interface est facile d'utilisation :
 - ▶ Analyse descriptive ;
 - ▶ Analyse graphique ;
 - ▶ Inférence statistique ;
 - ▶ Modélisation ;
 - ▶ Analyse prédictive ;
 - ▶ Traitement et analyse des données d'enquêtes complexes ;
 - ▶ Etc.



Pourquoi utiliser SPSS ?

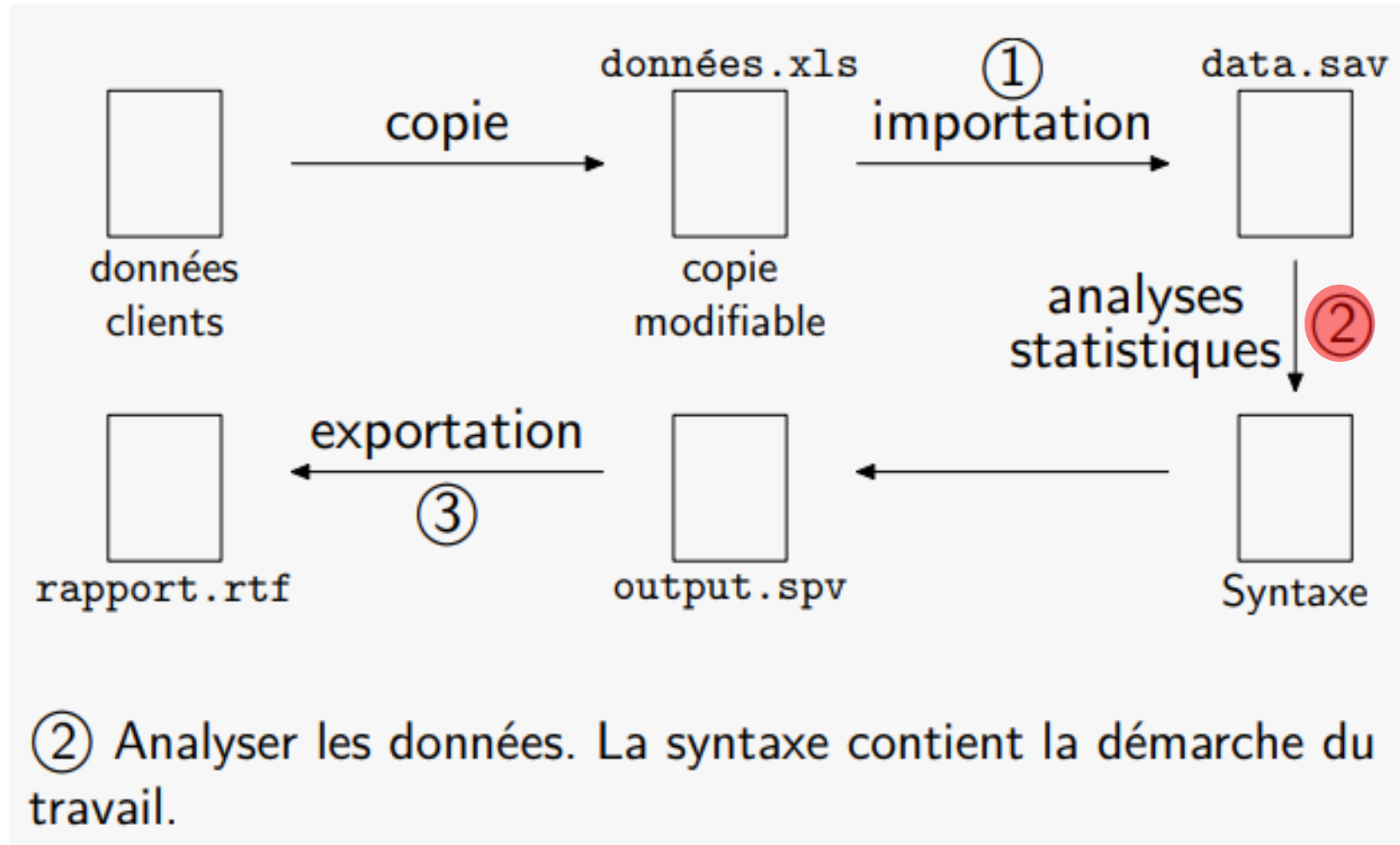
- ▶ Interface facile à utiliser
- ▶ Presque **tout peut se faire en cliquant**
 - ▶ Coller dans Syntaxe
 - ▶ Très utilisé sur le campus
- ▶ Important de connaître plus d'un langage

Marche à suivre

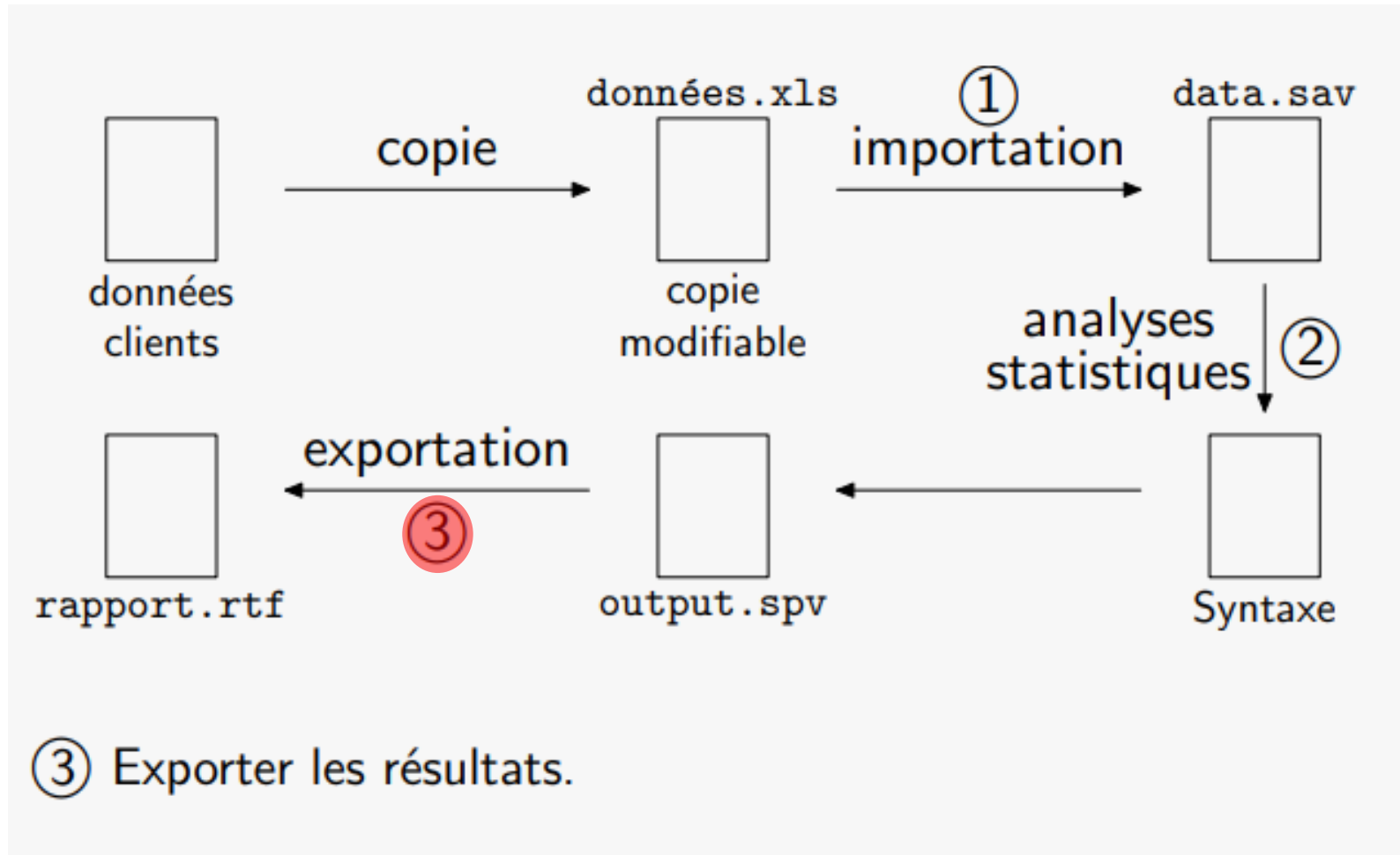


① Ajuster (le moins possible) la copie pour permettre l'importation.

Marche à suivre



Marche à suivre



GENERALITES SUR LE LOGICIEL SPSS

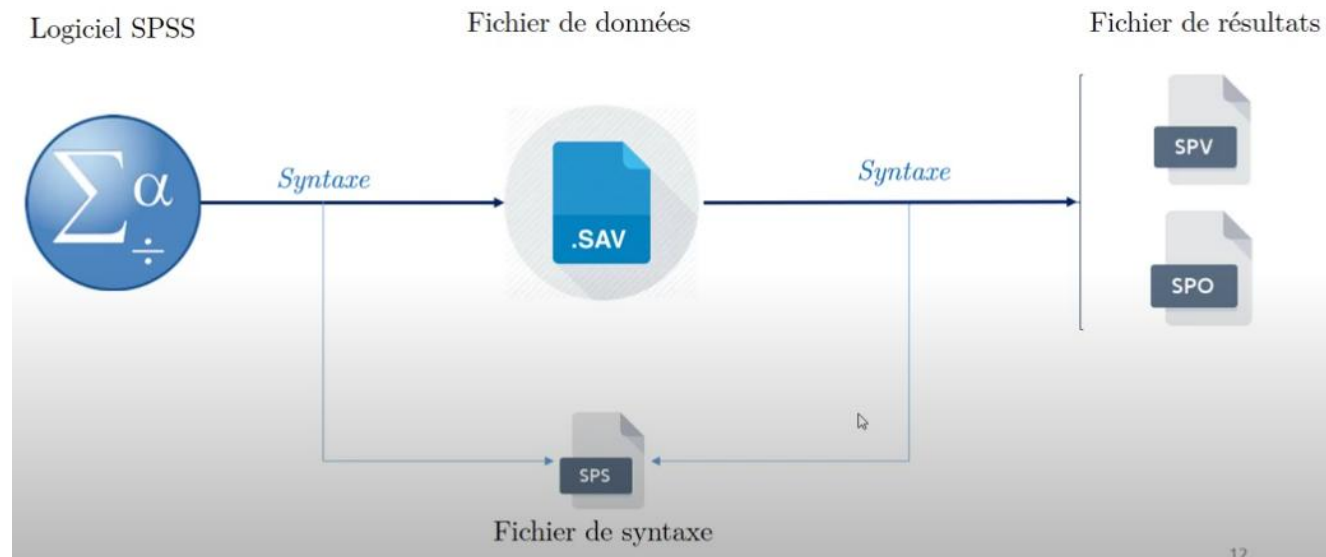
Fichier/File	<ul style="list-style-type: none"> ✓ Ouvrir les fichiers (données, syntaxes, résultats, script) ; ✓ Importer des données ; ✓ Enregistrer les fichiers ; ✓ Etc.
Edition/Edit	<ul style="list-style-type: none"> ✓ Copier les variables ;
Affichage/View	<ul style="list-style-type: none"> ✓ Afficher les variables ; ✓ Afficher les données ; ✓ Personnaliser l'affichage des variables ; ✓ Etc.
Données/Data	<ul style="list-style-type: none"> ✓ Définir les rôles des variables ; ✓ Trier les données ; ✓ Trier les variables ; ✓ Sélectionner les données ; ✓ Agréger les données ; ✓ Pondérer les observations ✓ Etc.
Transformer/Transform	<ul style="list-style-type: none"> ✓ Transformer les variables ; ✓ Créer de nouvelles variables ; ✓ Recoder les variables ; ✓ Etc.
Analyse/Analyze	<ul style="list-style-type: none"> ✓ Réaliser les tableaux ; ✓ Modéliser les données ; ✓ Analyser les données d'enquêtes ; ✓ Etc.

Graphiques/Graphs	<ul style="list-style-type: none"> ✓ Créer plusieurs types de graphiques <ul style="list-style-type: none"> ○ histogrammes, ○ boîtes à moustaches, ○ courbes, ○ etc.
Utilitaires/Utilities	<ul style="list-style-type: none"> ✓ Afficher les informations sur les variables du fichier de données ; ✓ Configurer les identificateurs OMS ; ✓ Etc.
Extensions/Extensions	
Fenêtre/Window	<ul style="list-style-type: none"> ✓ Fixer les volets (découper la fenêtre en plusieurs cadrants ; ✓
Aide/Help	<ul style="list-style-type: none"> ✓ Accéder : <ul style="list-style-type: none"> ○ Au support et forum de SPSS ; ○ A la documentation en ligne ○ au manuel de référence des commandes SPSS (en local) ✓

LES TYPES DE FICHIERS

Le logiciel SPSS gère quatre principaux fichiers. Il s'agit des :

- ▶ **Fichiers de données/ Data** : Les formats de **données** SPSS sont sauvegardés avec l'extension « **.sav** »
- ▶ **Fichiers de syntaxe/ Syntax** : Les listes d'instructions (**commandes**) sont gérées en SPSS par des fichiers d'extension « **.sps** » ;
- ▶ **Fichiers de résultats/Output** : Les sorties de SPSS peuvent être **exportées** ou copiées vers divers formats (MS Excel, MS Word, etc.). Toutefois, pour l'exploitation de ces **résultats** sous SPSS, ils devront être enregistrés avec l'extension appropriée « **.spv** »
- ▶ **Scripts** : Depuis la version 14 l'intégration du plug in Python permet d'exécuter des codes en codes sous SPSS. Les **scripts** sont des programmes en Python d'extension « **.py** » ; « **.pyc** » ou « **.pyo** ».



LES TYPES DE FICHIERS

- ▶ Pour créer l'un des quatre types de fichiers sous SPSS, il faut aller au menu « **Fichier** » puis « **Nouveau** » et sélectionner le type de fichier à créer.
- ▶ Exemple : **Fichier/Nouveau/Données** va ouvrir une fenêtre de données. Tandis que **Fichier/Nouveau/Syntaxe** ouvrira une nouvelle de l'éditeur de syntaxe



OUVRIR, IMPORTER ET EXPORTER DES FICHIERS DE DONNEES

- ▶ Pour un travail organisé, nous allons créer des fichiers de travail. Pour ce faire, Créez un dossier projet contenant tous les fichiers liés :

```
📁 Projet_SatisfactionClients/  
├── donnees_clients.sav  
├── analyse_satisfaction.sps  
├── resultats_satisfaction.spv  
└── rapport_final.rtf
```

FICHER DE DONNEES SOUS SPSS

- ▶ Le logiciel SPSS offre deux fenêtres pour la visualisation des données (« **Vue des données** » et « **Vue des variables** »).

Onglet Affichage des variables

Editeur de données

Onglet Affichage des données

Dans l'éditeur de données, il existe 2 onglets, en bas de page, affichage des données et affichage de variables.

FICHER DE DONNEES SOUS SPSS

La « **Vue des variables** » donne un aperçu des attributs de chaque variable du fichier de données :

- ▶ **Nom** : C'est le nom de la variable tel qu'il est affiché dans l'onglet « Vue des données »
- ▶ **Type** : C'est la caractéristique de votre variable, en général, on se limite à « **Chaîne** » lorsqu'il s'agit de données qualitatives et à « **Numérique** » lorsque l'on travaille avec des chiffres ou **Date**.
- ▶ **Largeur** : Ce champ contient le nombre de caractères de la donnée.
- ▶ **Décimales** : nombre de décimales désirées pour les données numériques
- ▶ **Libellé** : c'est le nom de la variable telle que vous voulez qu'il apparaisse dans le tableau de **résultats** (ex : pour la variable « Sit », vous voulez afficher Situation). Ce champ très important est le descriptif de la variable.
- ▶ **Valeur** : cette fonction donne la possibilité de labelliser des modalités (**catégorielle**) (ex : pour la variable sexe, il va être affiché 1 ou 0, il suffit de définir que 1 correspond à Homme et 0 à Femme). Contient la déclaration des valeurs possibles d'une variable catégorielle.
- ▶ **Manquant** : il s'agit ici de définir les valeurs que vous ne souhaitez pas intégrer dans vos calculs. Ex : une variable de satisfaction notée de 1 à 4, la modalité 5 correspondant aux « sans opinion », vous ne voulez pas retenir les sans opinion dans vos calculs alors entrez le chiffre 5 dans cette caractéristique « Manquant »
- ▶ **Colonnes** : Décrit la largeur à l'écran de la colonne (nombre de caractères)
- ▶ **Align** : Justification de la donnée dans la colonne (droite, gauche, etc.).

FICHER DE DONNEES SOUS SPSS

- ▶ **Mesure** : Définit la **nature de la variable**. Elle correspond à l'unité de mesure que vous voulez retenir : Echelle, Nominale ou Ordinale. (Echelle par défaut pour les variables de type numérique, Nominale par défaut pour les variables de type chaîne)
 - **Échelle** : variable continue (intervalle ou de rapport) (ex : l'âge)
 - **Ordinale** : variable catégorielle ordinale (ex : satisfaction)
 - **Nominale** : variable catégorielle nominale (ex : sexe)
- ▶ **Rôle** : Il est possible de préciser le **rôle que la variable jouera dans les analyses** (ex. : entrée = variable indépendante).
 - « **entrée/Input** » dans les modèles. Comme des variables **indépendantes**.
 - « **Cible/Target** » ce sont les variables **dépendantes**
 - « **les deux/Both** » lorsqu'elles jouent à la fois le rôle d'entrée et de cible ;
 - « **aucun/None** » la variable n'a pas de rôle prédéfinie
 - « **partition/Partition** » variable de partition des données pour tester, entraîner ou valider le modèle
 - « **scindée/Split** » : Pour les versions de « IBM SPSS » les variables avec ce rôle ne sont pas utilisées comme telles . ce rôle a été introduit pour « IBM SPSS Modeler ».
 - **Par défaut toutes les variables jouent le rôle d'« entrée ».**

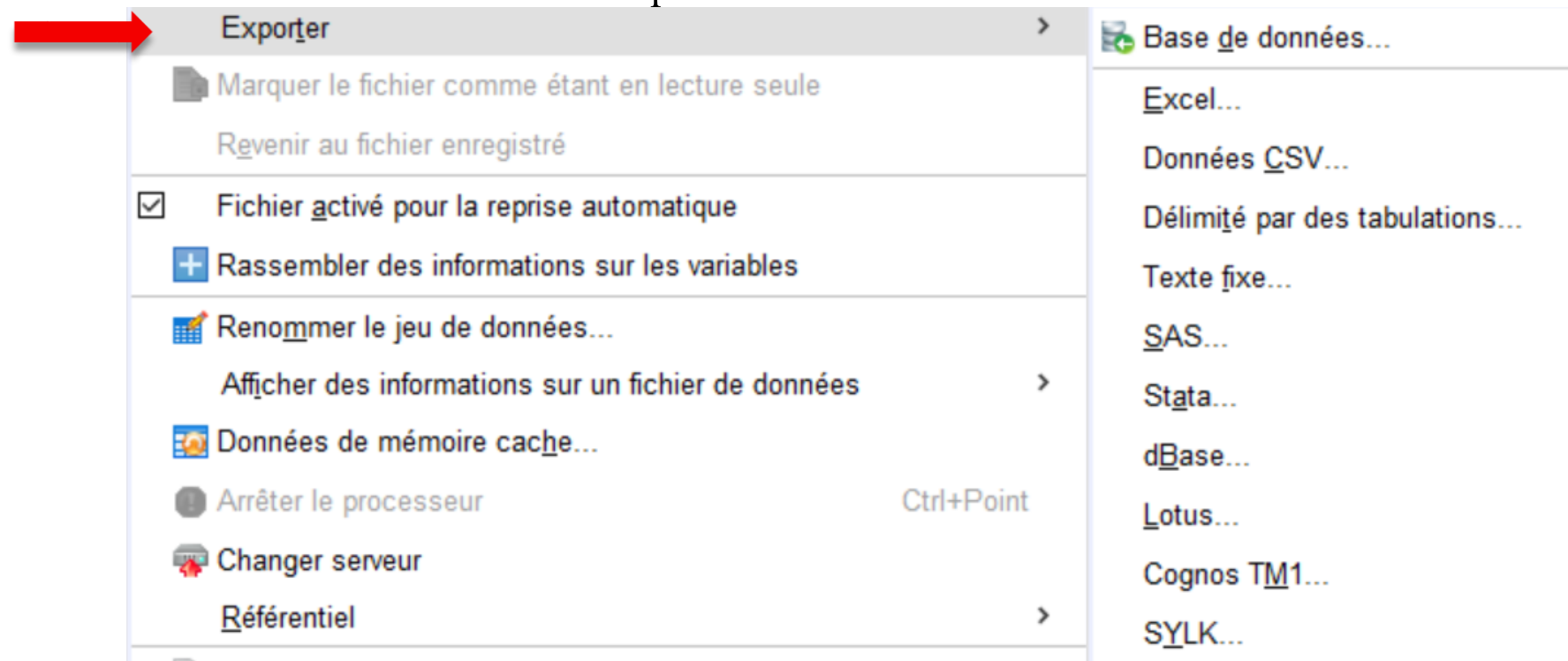
IMPORTER DES DONNEES SOUS SPSS

- ▶ SPSS permet d'ouvrir/importer des données de plusieurs formats en l'occurrence les données (Stata, SAS, Excel etc.). Pour ce faire, aller dans le menu « **Fichier** » puis « **Ouvrir** » et sélectionner le format du fichier à ouvrir dans le menu déroulant « **Fichiers de type** ».
- ▶ La même procédure peut être exécutée à partir du menu « **Fichier** » puis « **Importer des données** ».



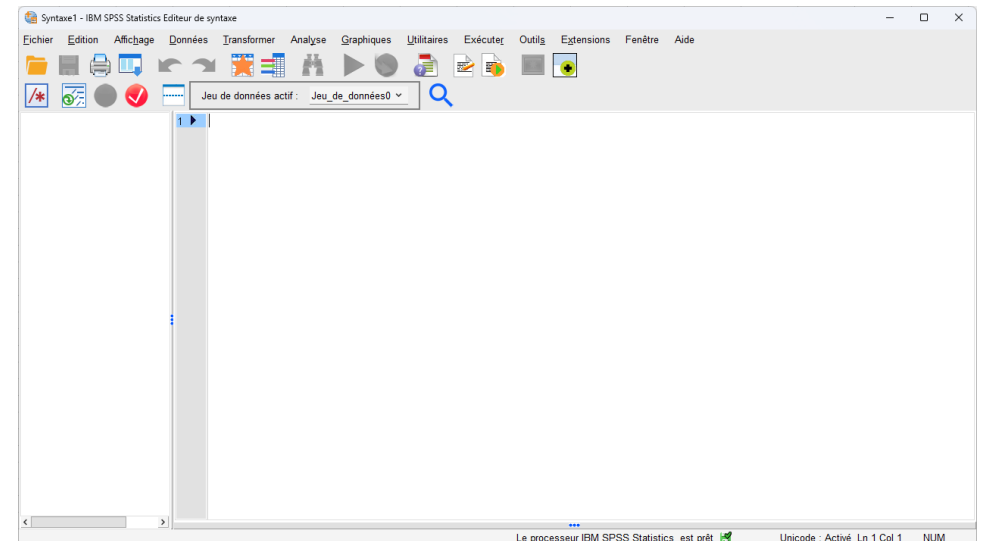
EXPORTER DES FICHIERS

- ▶ Pour exporter des données sous SPSS vers d'autres formats, il faut aller au menu « **Fichier** », puis « **Exporter** » et sélectionner le format d'exportation souhaitée (Excel, CSV, Délimité, Texte fixe, stata, etc.).
- ▶ Les fenêtres, qui suivent, permettent de définir ce qui doit être exporté (Noms ou libellés, Valeurs ou étiquettes, etc.) pour certains formats d'exportation comme MS Excel ou ASCII qui ne peuvent conserver à la fois les valeurs et les étiquettes.



EDITEUR DE SYNTAXE

- ▶ Le logiciel SPSS offre une interface conviviale. Toutefois pour un travail organisé, la liste des instructions peut et devrait être structurée dans un format que le logiciel peut lire et interpréter.
- ▶ En outre, travailler avec les lignes de commandes permet d'économiser un temps précieux, revisiter son travail pour apporter des corrections éventuelles et surtout de gérer des travaux collaboratifs.
- ▶ SPSS à l'instar des autres logiciels offre un éditeur de commandes. Pour l'ouvrir, il faut aller à « **Fichier** » puis « **Nouveau** » et choisir « **Syntaxe** ». Une page blanche s'ouvre et est prête à recevoir vos instructions
- ▶ Pour une commande ou un ensemble d'instructions, il suffit de sélectionner les lignes correspondantes et avec un clic droit sélectionner « **Exécuter la sélection** ». Un raccourci est généralement disponible sous la forme d'un triangle ▶ sur lequel il faut juste cliquer après avoir fait la sélection.



GERER ET TRANSFORMER LES VARIABLES

Étiquetage des variables

Dans SPSS, l'**étiquetage** des variables consiste à attribuer :

- ▶ **Un nom de variable** : un identifiant court utilisé en interne dans les analyses (ex : age, sexe, revenu).
- ▶ **Une étiquette de variable** (variable label) : une description plus longue et compréhensible de la variable (ex : « Âge du répondant », « Sexe de l'individu », « Revenu mensuel en dirhams »).
- ▶ **Des étiquettes de valeur** (value labels) : pour les variables **catégorielles**, on attribue un libellé à chaque modalité (ex : 1 = Homme, 2 = Femme)
- ▶ Exemple
 - ▶ Variable : **sexe**
 - ▶ Nom de variable : **sexe**
 - ▶ Étiquette de variable : **Sexe de l'enquêté**
 - ▶ Étiquettes de valeur :
 - ▶ 1 → **Homme**
 - ▶ 2 → **Femme**

Etiquetage des variables

Etiquette de variable

- ▶ Dans l'exemple précédent le libellé de la variable **sexe** est « **Sexe de l'enquêté** ».
- ▶ SPSS permet avec une grande flexibilité de décrire les variables.
- ▶ Pour ce faire, aller dans la « **vue des variables** », se positionner dans la colonne « **Libellé** » puis saisir l'étiquette de la variable sur la ligne correspondante à son nom dans la colonne « **Nom** ».

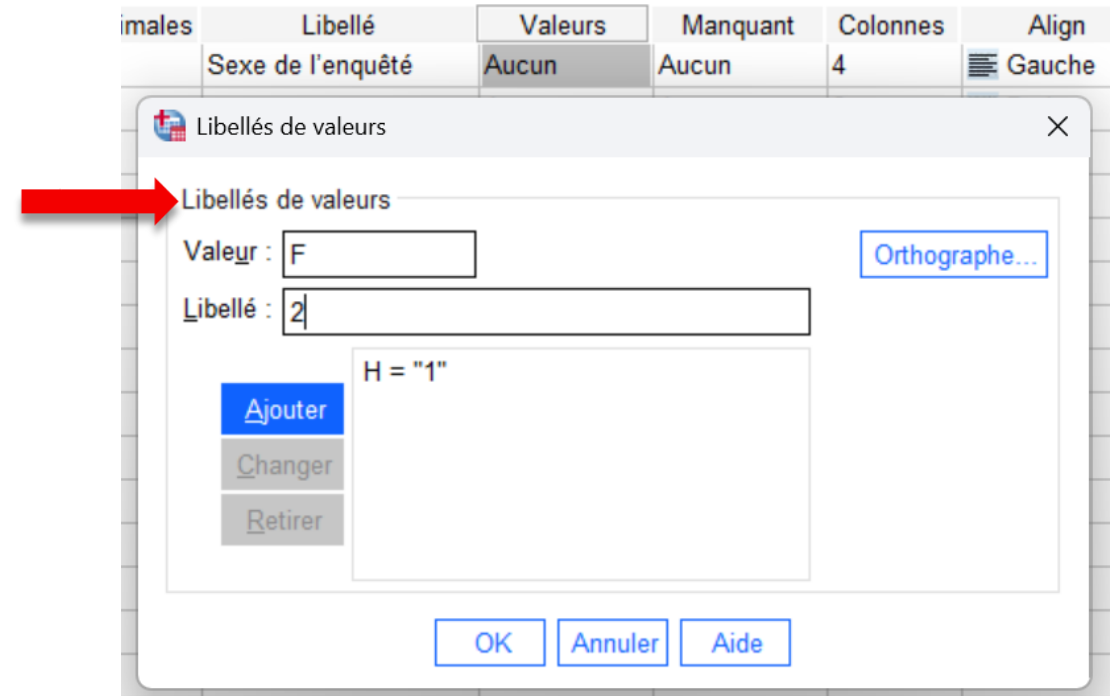
Nom	Type	Largeur	Décimales	Libellé
sexe	Chaîne	4	0	Sexe de l'enquêté

Ajout d'étiquettes de valeur

- ▶ Pour les étiquettes de valeurs, il faut plutôt se positionner la colonne « **Valeur** ». Une fenêtre apparaît comme ci-après dans la Figure. **Saisir** les valeurs et les étiquettes correspondantes à chaque valeur en cliquant sur le bouton « **Ajouter** ». Lorsque les étiquettes auront été définies pour toutes les valeurs, valider le processus en cliquant sur le bouton « **OK** »

Attributs « Type, Largeur, Décimales... » :

- ▶ Tout comme les étiquettes de variable et de valeurs, les autres propriétés des variables présentes dans la « **vue des variables** » peuvent être définies directement via l'interface.



Transformer les variables

- ▶ L'analyste des données est amené à **créer de nouvelles variables** qui correspondent à ses besoins à partir des variables déjà existantes. La création de nouvelles variables se fait soit à travers une **transformation fonctionnelle** ou simplement par **regroupement de certaines modalités** d'une variable, et ce, dans le but de répondre à une question d'analyse que les variables présentes ne permettent pas

Créer des variables

- ▶ Pour créer/générer une nouvelle variable, aller au menu « **Transformer** » puis « **Calculer la variable** ». Une fenêtre apparaît comme à la Figure permettant de :

- a) saisir le nom de la nouvelle variable (1) ;
- b) définir le format (Chaîne ou numérique) de la variable (2) ;
- c) définir une condition si la transformation ne s'applique pas à toutes les observations du fichier (3) ;
- d) entrer l'expression de la variable comme transformée des variables existantes (4) par combinaison des opérateurs (6) et/ou des fonctions (8).

The screenshot shows the 'Calculer la variable' dialog box with the following components and callouts:

- (1) Saisir le nom de la nouvelle variable: Points to the 'Variable cible' field containing 'LnAgeCar'.
- (2) Définir le format de la variable et son libellé: Points to the 'Type et libellé...' button.
- (3) Liste des variables du fichier actif: Points to the list of variables (REGION, hh1, hh2, m1, M2, M3, M4, M5, M6, M7, MBA, M8B, M9, M10, M11, M12, M13, M14, M15).
- (4) Définir un filtre conditionnel: Points to the 'Si' field containing 'M4 >= 20'.
- (5) Saisir l'expression numérique de la nouvelle variable: Points to the 'Expression numérique' field containing 'Ln(M4) ** 2'.
- (6) Pavé des opérateurs: Points to the central numeric keypad.
- (7) Groupe de fonctions (Ensemble de fonctions rangées): Points to the 'Groupe de fonctions' list (Tous, Arithmétique, CDF et CDF non centré, Conversion, Date/Heure actuelle, Calcul de date).
- (8) Pavé des fonctions disponibles: Points to the 'Fonctions et variables spéciales' list (Abs, Arsin, Artan, Cos, Exp, Lg10, Ln, Lngamma, Mod, Rnd(1)).

Transformer les variables

- ▶ Quelques opérateurs logiques et relationnels sous SPSS

Opérateurs relationnels			Rôle	Opérateurs logiques			Rôle
>	ou	GT	Supérieur	~	ou	NOT	non
<	ou	LT	Inférieur		ou	OR	ou
=	ou	EQ	Egal	&	ou	AND	et
>=	ou	GE	Supérieur ou égal				
<=	ou	LE	Inférieur ou égal				
≠	ou	NE	Pas égal ou différent de				

Transformer les variables

Fonctions sous SPSS

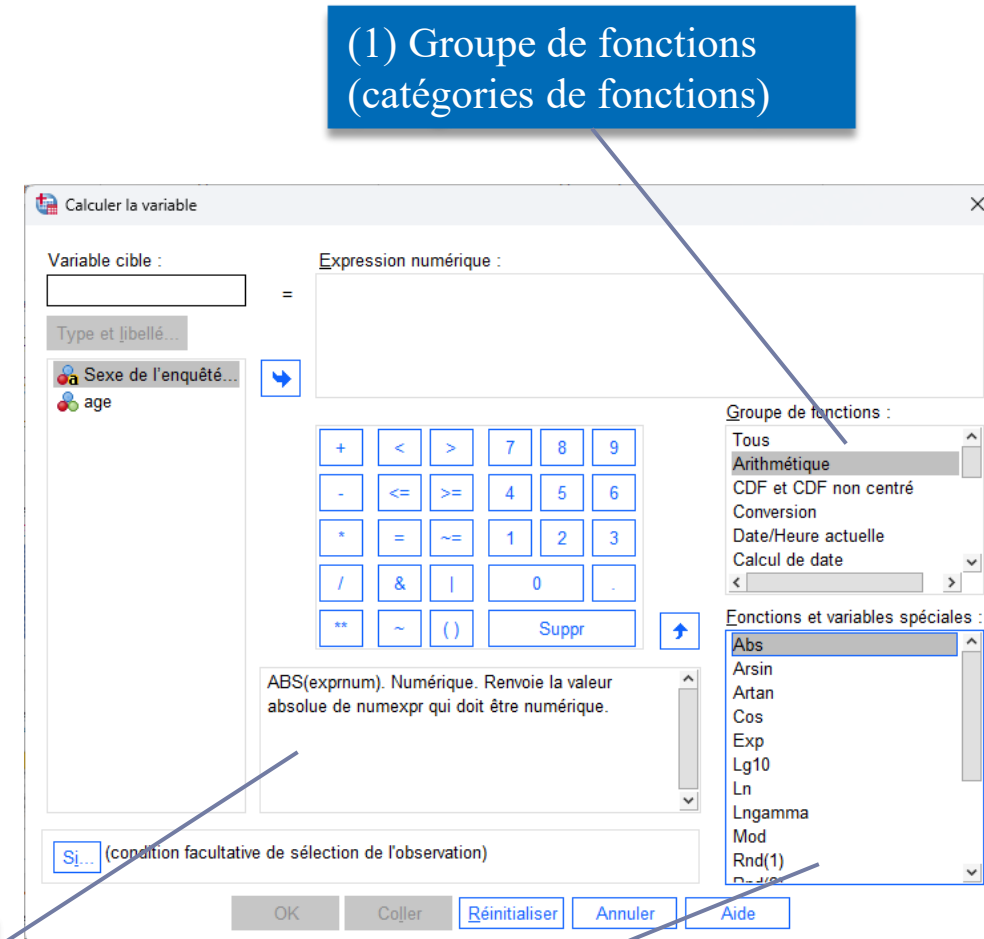
SPSS dispose d'une batterie de fonctions prédéfinies classées par groupe de fonctions dans le quadrant (1). La sélection d'un groupe de fonction au quadrant (1) fait apparaître systématiquement la liste des fonctions disponibles pour ce groupe dans le quadrant (8).

- ▶ a) **Arithmétique** : Fonctions **ABS** (Valeur absolue), **Ln** (Logarithme népérien), **Exp** (Fonction exponentielle), les fonctions trigonométriques et réciproques (Cos, Sin, Tan etc.), **SQRT** (racine) etc.
- ▶ b) **Statistiques** : Pour le calcul de certaines statistiques. **Mean** (moyenne), **Sum** (Somme), **Median** (médiane), **Max** (Maximum) etc.
- ▶ c) **Chaîne** : Il s'agit des fonctions qui s'appliquent aux variables de type alphanumérique. **Lower** (Minuscule), **Ucase** (Majuscule), **Length** (Longueur), **Char.index** (pour rechercher un caractère dans un texte), **Char.substr** (Extraire des caractères d'un texte à partir d'une position donnée); etc.
- ▶ d) **Conversion** : **STRING** (pour convertir une expression numérique en texte) ou **NUMBER** (pour convertir une expression de type chaîne en nombre).

Certaines fonctions sont définies par le système. Elles sont précédées par le signe « \$ »

- ▶ a) **\$Sysmis** : valeur manquante du système.
- ▶ b) **\$Casenum** : Renvoie le numéro séquentiel de chaque observation dans la base de données active.
- ▶ c) **\$Date** : Date courante au format international avec l'année deux chiffres (jj-mmm-aa) de largeur 9 (A9). Exemple : 20-SEP-21, pour 20 septembre 2021.

SPSS fournit la définition et le mode d'emploi (3) de chaque fonction dans une fenêtre contextuelle c'est-à-dire lorsque celle-ci est sélectionnée



(1) Groupe de fonctions (catégories de fonctions)

(3) mode d'emploi

(8) Pavé des fonctions disponibles

Transformer les variables

Recoder des variables

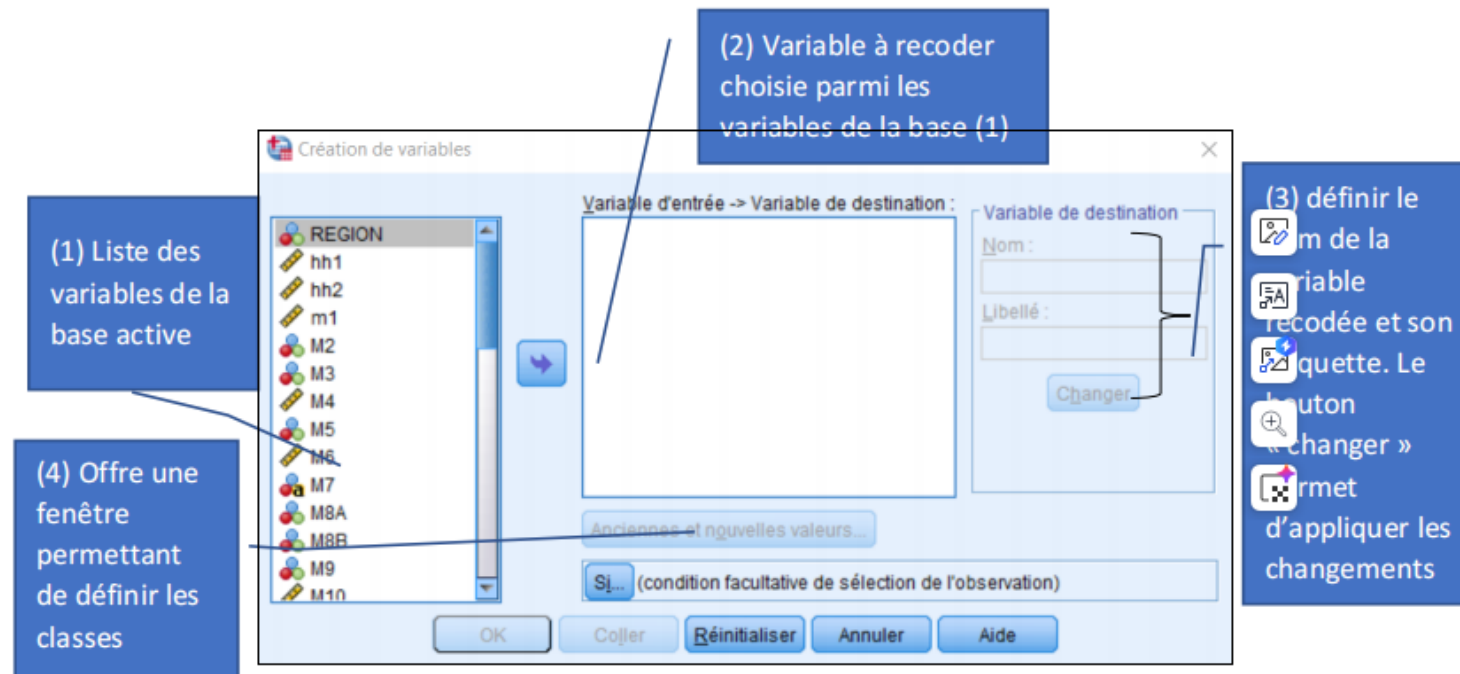
- ▶ Les besoins de l'analyste peuvent l'amener à préférer un **regroupement** de modalités pour une variable donnée. Par exemple, au lieu de travailler avec l'âge, l'on décide d'analyser les classes d'âge.
- ▶ Considérons une variable « M4 » qui correspond à l'âge de l'individu en années révolues. On souhaite étudier certaines caractéristiques des membres de ménages en considérant 7 classes d'âge :
 - ▶ 1-Moins de 15 ans
 - ▶ 2-Entre 15 et 24 ans
 - ▶ 3-Entre 25 et 34 ans
 - ▶ 4-Entre 35 et 44 ans
 - ▶ 5-Entre 45 et 54 ans
 - ▶ 6-Entre 55 et 64 ans
 - ▶ 7-65 ans ou plus
- ▶ SPSS permet de recoder une variable **sous le même** ou créer d'une **nouvelle** par recodage d'une ancienne.
- ▶ Pour créer une nouvelle variable par recodage d'une autre variable sous SPSS, aller au menu « **Transformer** » puis « **Création de variables** ». Une fenêtre s'ouvre (voir Figure(V) ci-après)

Transformer les variables

Dans cette nouvelle, procéder par les étapes

- Choisir la variable à recoder (variable d'entrée) dans le quadrant (1) vers le quadrant (2);
- Saisir le nom de la nouvelle variable, son étiquette au cas échéant et valider cette opération en cliquant sur le bouton « [changer](#) » ;
- Définir au cas échéant les observations auxquelles s'applique le recodage ;
- Le bouton « [Anciennes et nouvelles valeurs](#) » ouvre une **nouvelle fenêtre** qui permet véritablement de définir les différents codages.

Figure (V)



Transformer les variables

Recoder et créer de nouvelles variables : Anciennes et nouvelles valeurs ×

Ancienne valeur

Valeur :

Valeur système manquante

Valeurs système manquante ou valeur manquante de l'utilisateur

Plage :

à

Plage, du MINIMUM à la valeur :

Plage, de la valeur au MAXIMUM :

Toutes les autres valeurs

Nouvelle valeur

Valeur :

Valeur système manquante

Copier les anciennes valeurs

Ancienne --> Nouvelle :

Lowest thru 14 --> 1
15 thru 24 --> 2
25 thru 34 --> 3
35 thru 44 --> 4
45 thru 54 --> 5
55 thru 64 --> 6
65 thru Highest --> 7

Variables destination sont des chaînes Largeur :

Convertir les chaînes numériques en nombres ('5' -> 5)

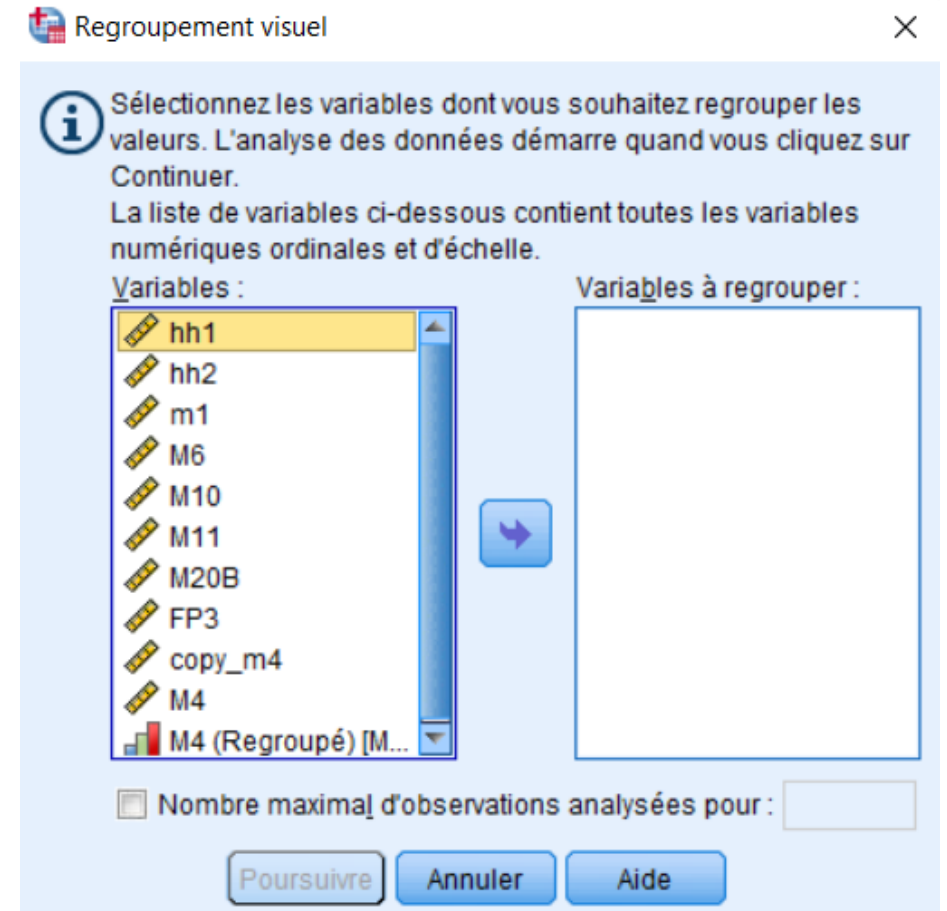
Transformer les variables

- ▶ En ce qui concerne le recodage d'une variable sous le même nom, aller au menu « **Transformer** » puis « **Recodage des variables** ». La fenêtre qui apparaît est de la Figure **V**, à l'exception du quadrant (**3**) qui permet de créer la nouvelle variable.

Transformer les variables

Assistant regroupement visuel

- ▶ Plutôt que l'**analyste** détermine les points de césure avant de procéder au recodage, il peut, en utilisant l'assistant « **Regroupement visuel/ Visual Binning** », laisser le soin à la machine de générer ces points à partir des données pendant le recodage
- ▶ Selon la **Documentation de IBM SPSS Statistics**, le regroupement visuel aide à la création de variables basées sur le regroupement des valeurs **contiguës** de variables dans un nombre distinct de catégories. Le Regroupement visuel peut être utilisé pour :
 - ▶ Créer des variables catégorielles à partir de variables d'échelle continues.
 - ▶ Fusionner un grand nombre de catégories ordinales en un jeu de catégories plus petit.
- ▶ Pour utiliser l'assistant « **Regroupement visuel** », aller au menu « **Transformer** » puis « **Regroupement visuel** ». La boîte de dialogue de la ci-contre s'affiche et permet de sélectionner parmi les variables d'échelle et ordinales la ou les variables à recoder.
- ▶ Il est donc important de se souvenir que cet assistant n'est utile que pour les variables d'**échelle ou ordinale**. Le recodage des variables nominales ou de type chaîne doit se faire selon la méthode « **RECODE** » précédemment expliquée



Transformer les variables

- ▶ La sélection d'une variable à regrouper permet d'accéder à la **boîte de dialogue** de la Figure qui permet de paramétrer le mode de regroupement. Le bouton « **Créer des divisions** » affiche les options disponibles :
 - ▶ Intervalles de longueur identique
 - ▶ Percentiles égaux fondés sur les observations analysées
 - ▶ Divisions au niveau de la moyenne et des écarts types, fondées sur les observations analysées
- ▶ La deuxième option permet alors en définissant **M4_Deciles** la variable « M4 » recodée suivant

- ▶ $M4_Deciles=1$ si $(M4 \leq D1)$
- ▶ $M4_Deciles=2$ si $(D1 < M4 \leq D2)$
- ▶ $M4_Deciles=3$ si $(D2 < M4 \leq D3)$
- ▶ $M4_Deciles=4$ si $(D3 < M4 \leq D4)$
- ▶ $M4_Deciles=5$ si $(D4 < M4 \leq D5)$
- ▶ $M4_Deciles=6$ si $(D5 < M4 \leq D6)$
- ▶ $M4_Deciles=7$ si $(D6 < M4 \leq D7)$
- ▶ $M4_Deciles=8$ si $(D7 < M4 \leq D8)$
- ▶ $M4_Deciles=9$ si $(D8 < M4 \leq D9)$
- ▶ $M4_Deciles=10$ si $(M4 > D9)$

Où $D1, D2, \dots, D9$ sont les déciles de la variable « M4 »

Regroupement visuel

Liste des variables analysées : M4

Nom : M4 Libellé :

Variable actuelle : M4

Variable regroupée : M4_Deciles Libellé : Déciles de la variable M4

Minimum : 0 Valeurs non manquantes Maximum : 98

Entrez les divisions des intervalles ou cliquez sur Créer des divisions pour définir automatiquement des intervalles. La division 10, par exemple, définit un intervalle qui commence juste au-dessus de l'intervalle précédent et se termine à 10.

	Valeur	Libellé
1	2,0	
2	4,0	
3	7,0	
4	9,0	
5	12,0	
6	16,0	
7	23,0	
8	30,0	

Observations analysées : 19534

Valeurs manquantes : 0

Copier les casiers

À partir d'une autre variable...

Vers d'autres variables...

Limites supérieures

Inclus (<=)

Exclus (<)

Créer des divisions...

Créer des libellés

Inverser l'échelle

OK Coller Réinitialiser Annuler Aide

Transformer les variables

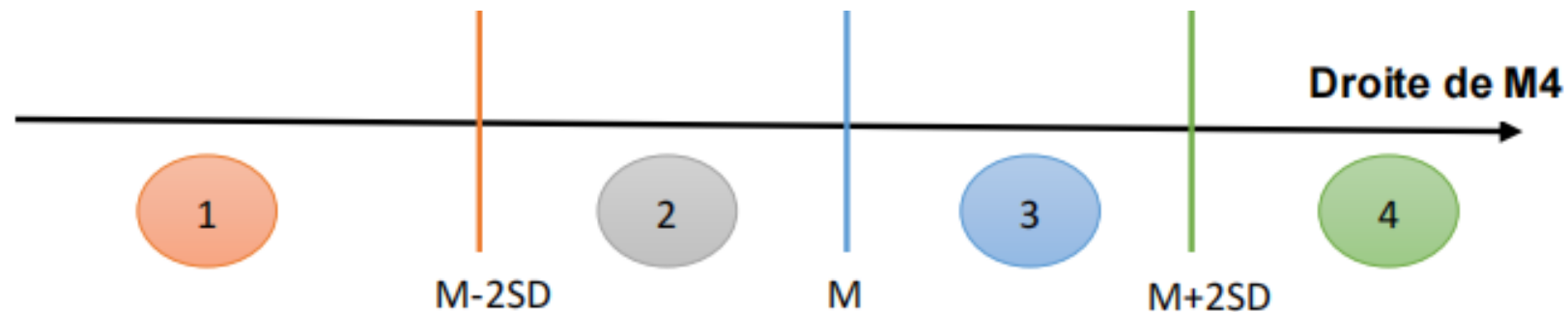
- ▶ Nous pouvons aussi utiliser cet assistant pour créer des groupes basés sur leur distance par rapport à la moyenne où les distances sont exprimées en termes d'écart-type.

⊙ Divisions au niveau de la moyenne et des écarts types sélectionnés, fondées sur les observations analysées

+/- 1 écart type

+/- 2 écarts types

+/- 3 écarts types



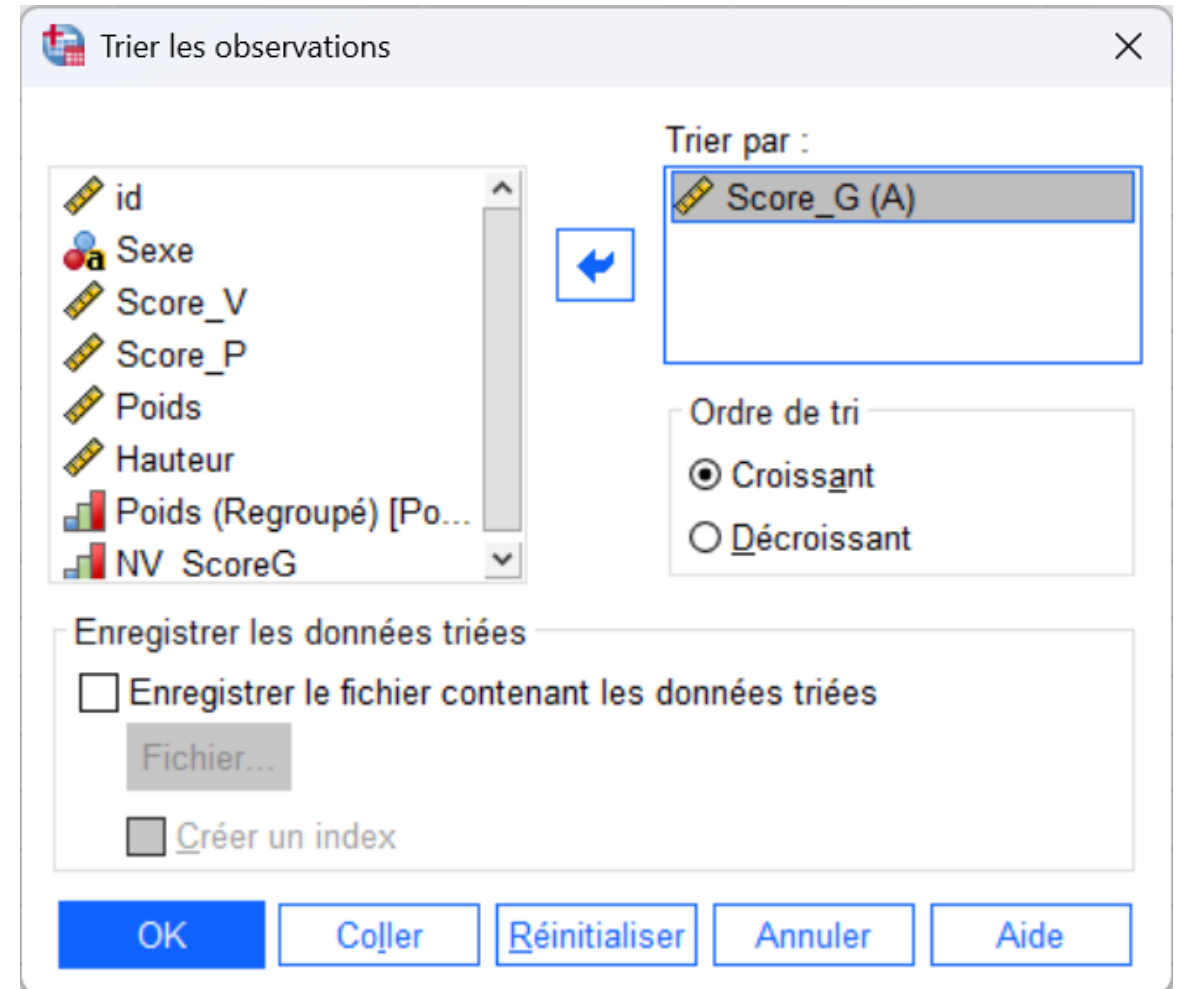
Droite de M4 et partition suivant la distance de ± 2 Ecart-Type par rapport à la moyenne de M4

MANIPULER LES FICHIERS DE DONNEES

OPERATIONS USUELLES SUR LES DONNEES : TRI, SELECTION, ET EDITION DES DONNEES

Tri des observation

- ▶ Pour ordonner les observations d'un fichier suivant une ou plusieurs variables, aller au menu « **Données** » puis « **Trier les observation** »



OPERATIONS USUELLES SUR LES DONNEES : TRI, SELECTION, ET EDITION DES DONNEES

Sélection des observations

Il arrive souvent que les besoins de l'analyste portent une partie des données et non l'intégralité.

4 options de sélection dans la fenêtre "Sélectionner des observations" de SPSS :

1. Toutes les observations

- ▶ Aucune sélection n'est faite.
- ▶ Tous les cas de votre base de données seront pris en compte dans les analyses.

2. Selon une condition logique

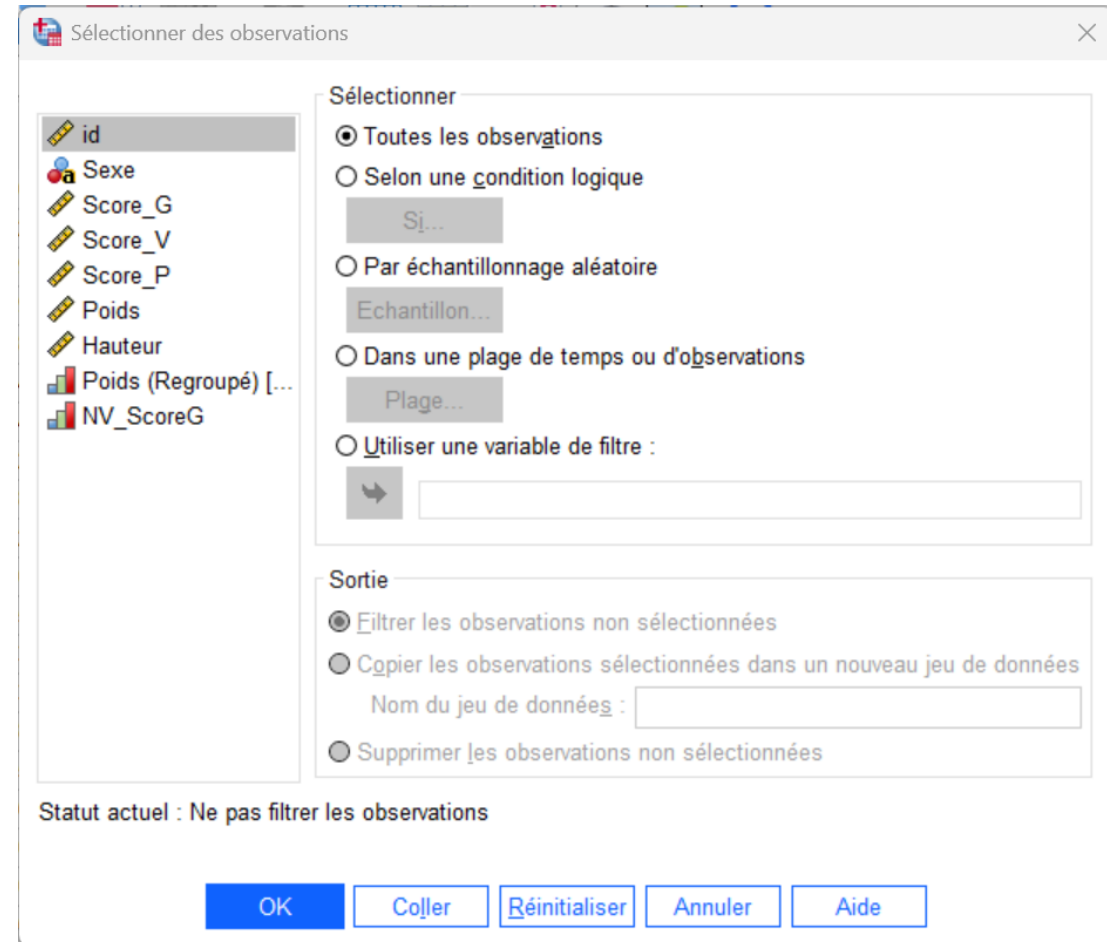
- ▶ Permet de filtrer les cas selon un critère (ex. : Sexe = 1 ou Score_G > 10).
- ▶ Très utile pour analyser un sous-groupe spécifique (hommes, jeunes, etc.).

3. Par échantillonnage aléatoire

- ▶ SPSS sélectionne un sous-ensemble aléatoire de vos observations.
- ▶ le nombre ou le pourcentage de cas à extraire.
- ▶ Pratique pour tester un modèle ou faire des simulations.

4. Dans une plage de temps ou d'observations

- ▶ Permet de sélectionner un intervalle précis (ex. : les 100 premiers cas).
- ▶ Très utile dans les analyses de séries temporelles ou fichiers ordonnés



OPERATIONS USUELLES SUR LES DONNEES : TRI, SELECTION, ET EDITION DES DONNEES

Édition des données

- ▶ L'édition des données est indispensable lorsqu'il s'agit d'apurer les données.
- ▶ L'édition des données dans SPSS consiste à **modifier**, **corriger**, ou **ajuster** les valeurs contenues dans l'éditeur de données, soit en mode manuel, soit en utilisant des outils de transformation.

1. **Édition manuelle des données** - Dans la « **Vue Données** » (Data View), qui ressemble à un tableau **Excel**.

2. Édition automatique avec transformation

« **Transformer** »

- ▶ Compute Variable : créer ou recalculer une variable (ex. : $\text{score_total} = \text{score1} + \text{score2}$)
- ▶ Recode : transformer les valeurs d'une variable (ex. : regrouper des âges ou recoder des modalités).
- ▶ Replace Missing Values : remplacer les valeurs manquantes par une moyenne ou médiane.
- ▶ ...

3. Ajout et modification des métadonnées

« **Vue des Variables** » :

- ▶ Modifier le nom, type, libellé, valeurs ou format d'une variable.
- ▶ Définir des valeurs manquantes.
- ▶ Modifier l'échelle de mesure (Nominale, Ordinale, Échelle).

OPERATIONS USUELLES SUR LES DONNEES : TRI, SELECTION, ET EDITION DES DONNEES

Agrégation des donnée

- ▶ L'agrégation des données consiste à **résumer plusieurs lignes d'observations en une seule**, par groupe, en appliquant des fonctions comme : moyenne, somme, minimum, maximum, nombre de cas, etc.
- ▶ Par exemple, pour chaque ménage :
 - ▶ L'âge moyen des membres ;
 - ▶ L'âge maximum des membres ;
 - ▶ La taille (nombre de membres) ;
 - ▶ Etc.
- ▶ Chaque observation du nouveau fichier ainsi constitué à **renvoie à un groupe d'observations** du fichier initial.
- ▶ La procédure d'agrégation des données sous SPSS se réalise à l'aide de l'assistant « **Agréger** » disponible dans le menu « **Données** ».
 - ▶ Le **quadrant (1)** contient la liste de toutes les variables du fichier.
 - ▶ Le **quadrant (2)** permet sélectionner les variables d'agrégation à partir de la liste du quadrant (1).
 - ▶ Le **quadrant (3)** est réservé aux variables à agréger. Celles-ci sont sélectionnées à partir du quadrant (1).
 - ▶ Le **quadrant (4)** concerne les boutons « **Fonction** » et « **Libellé** ». Le bouton « **Fonction** » donne accès à une boîte des fonctions prédéfinies.
 - ▶ Le **quadrant (5)** permet d'ajouter aux données le nombre d'observations du fichier initial ayant servi à constituer chaque observation du fichier final. Il suffit pour cela de cocher la case « **Nombre d'observations** ». La variable « Nombre d'observations » sera créée dans le fusion agrégée sous le nom renseigné dans la partie « **N_BREAK** ». Par défaut, cette variable aura pour nom « **N_BREAK** »
 - ▶ Le **quadrant (6)** est très explicite. Il permet de choisir comment sont gérer les nouvelles variables issues de l'agrégation. On peut soit sauvegarder ces variables dans un nouveau fichier, soit les ajouter à l'ensemble de données actif ou dans un nouvel ensemble de données.



FUSION DEUX OU PLUSIEURS ENSEMBLES DE DONNEES

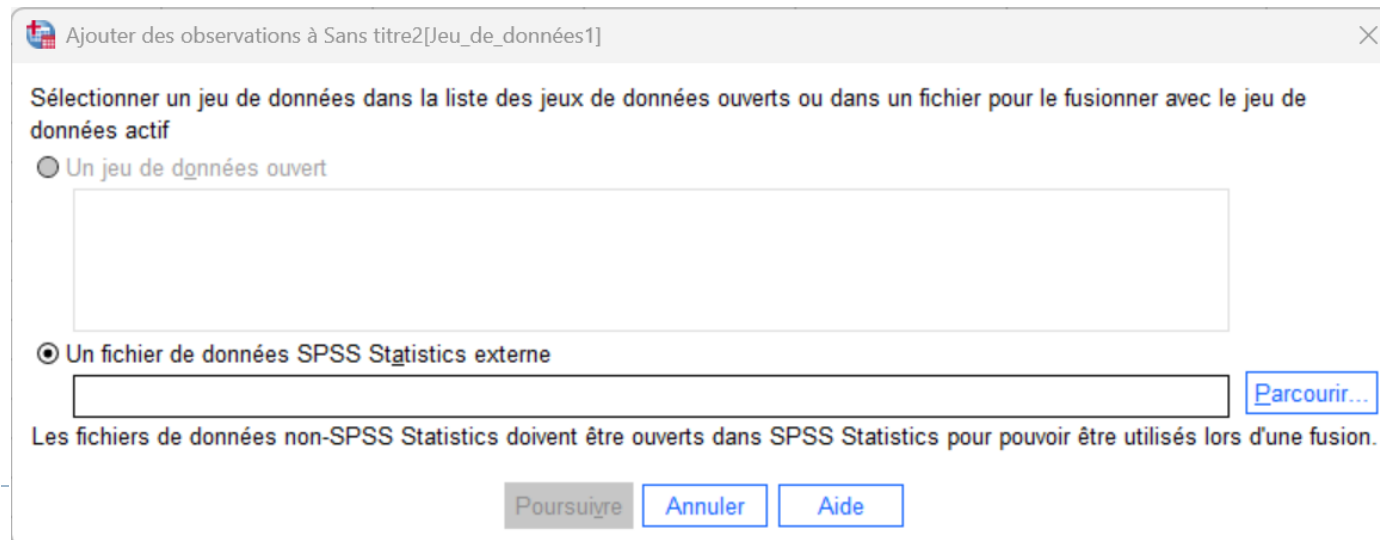
- ▶ On distingue **deux** types de **fusion** de données. Lorsqu'on a une partition de données en deux ou plusieurs fichiers que l'on veut assembler dans un seul fichier de données, la **fusion** est appelée **ajout de cas**. Il s'agit dans ce cas d'empiler verticalement les enregistrements (observations).
- ▶ En revanche, lorsque pour deux fichiers de données, des informations (en termes de variables) sont disponibles sur l'un et pas dans l'autre, on parle d'**ajout de variables**.



FUSION DEUX OU PLUSIEURS ENSEMBLES DE DONNEES

Ajout de cas

- ▶ Supposons que les données sur les caractéristiques des ménages ont été saisies séparément selon le milieu de résidence (**urbain/rural**). A l'issue de la saisie, le statisticien d'enquête dispose par conséquent de deux fichiers soit « **MenageUrbain.sav** » et « **MenageRural.sav** » avec des variables similaires.
- ▶ Pour constituer un fichier de ménage unique avec les deux fichiers, il doit les **superposer** ou empiler, c'est-à-dire ajouter les observations du fichier « **MenageRural.sav** » à la suite de celles du fichier « **MenageUrbain.sav** » (vice versa). Ce type de superposition de fichiers s'appelle **fusion par ajout de cas**. Sa mise en œuvre sous SPSS se fait dans le menu « **Données** » puis « **Fusionner des fichiers** » et « **Ajout des observations** ». Une boîte de dialogue vous permet alors de sélectionner les fichiers à fusionner.



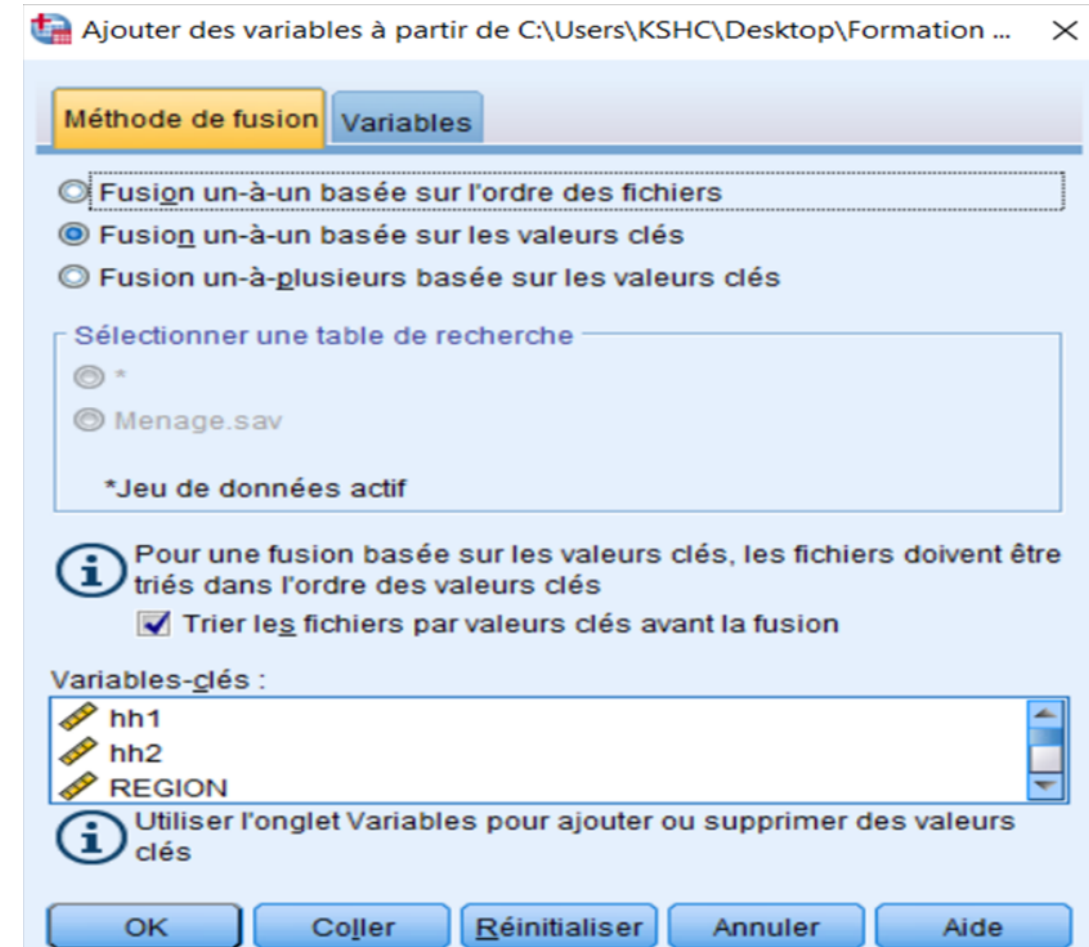
FUSION DEUX OU PLUSIEURS ENSEMBLES DE DONNEES

Ajout de variable

- ▶ Considérons le fichier « Individu.sav » contenant les caractéristiques des membres de ménages et le fichier « Menage.sav » les caractéristiques telles que le logement des ménages. L'on souhaite produire des statistiques du genre « pourcentage de la population ayant accès à l'électricité ». Pour ce faire, nous avons besoin d'un fichier d'analyse comprenant à la fois les individus et les informations sur leurs logements.
- ▶ Le statisticien d'enquête doit procéder soit à l'ajout des variables portant sur les individus dans le fichier « [Menage.sav](#) », soit ajouter les variables portant sur le logement dans la base « [Individu.sav](#) ». Ce type de fusion s'appelle « [ajout de variables](#) ». Il est mis en œuvre sous SPSS, dans en allant dans le menu « [Données](#) » puis « [Fusionner des fichiers](#) » et « [Ajout des variables](#) ».
- ▶ cette opération exige que les observations des deux fichiers :
 - ▶ a) aient une **clé de fusion**. Il s'agit d'une ou plusieurs variables communes aux deux fichiers à fusionner et qui permettent d'identifier les observations de part et d'autre,
 - ▶ b) soient préalablement **triées** suivant la clé de fusion.

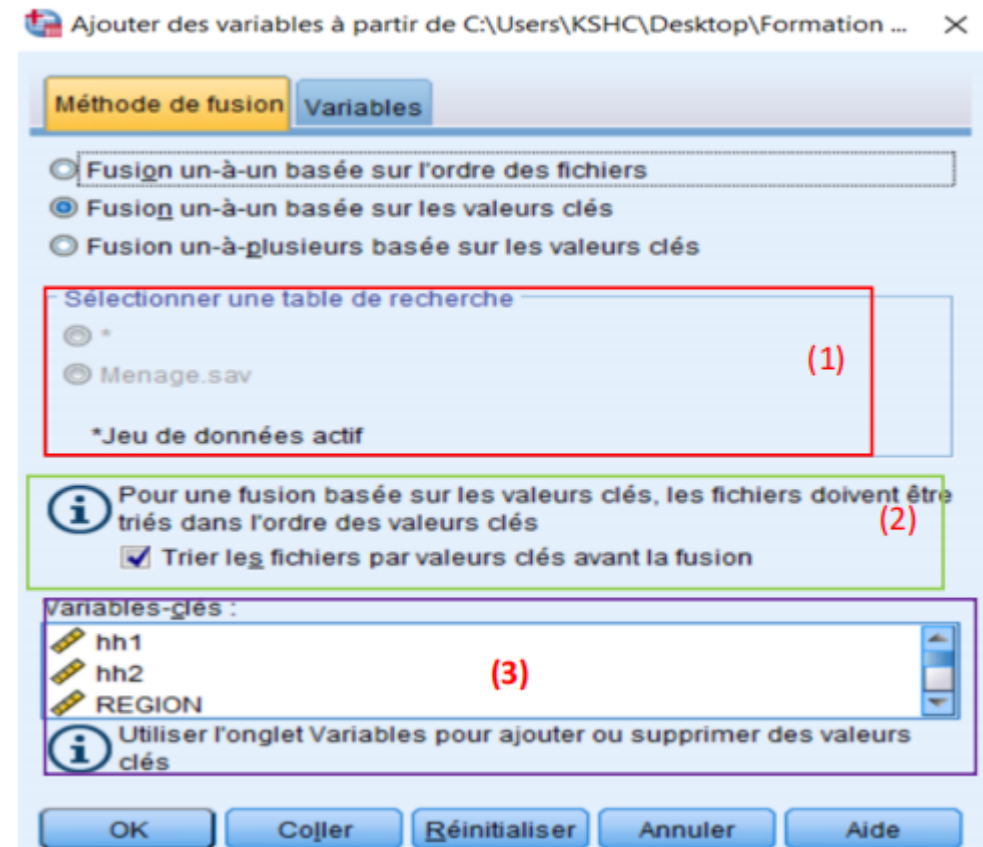
FUSION DEUX OU PLUSIEURS ENSEMBLES DE DONNEES

- ▶ La boîte de dialogue offre deux onglets « **Méthode de fusion** » et l'onglet « **Variables** ». L'onglet « **Méthode de fusion** » propose trois méthodes de fusion :
 - ▶ a) **Fusion un-à-un basé sur l'ordre des fichiers** : Ce type de fusion ne nécessite pas une clé de fusion. Les variables du second fichier sont ajoutées à la suite de celles du premier fichier.
 - ▶ b) **Fusion un-à-un basée sur les valeurs clés** : Chaque observation de la base de données de départ sera reliée par une seule observation de la seconde base de données (analogie avec la relation de **1 :1** dans les bases de données relationnelles)
 - ▶ c) **Fusion un-à-plusieurs basée sur les valeurs clés** : A une observation de la base de données de départ peut correspondre une ou plusieurs observations de la seconde base de données (relation **1 :m** ou **m :1**)



FUSION DEUX OU PLUSIEURS ENSEMBLES DE DONNEES

- ▶ **Le quadrant (1)** invite à la sélection d'une « [table de recherche](#) ».
- ▶ **Le quadrant (2)** permet en cochant la case « [Trier les observations par valeurs clés avant la fusion](#) » permet à SPSS de réaliser les tris des données avant de procéder à la fusion des données. Les versions antérieures n'offraient pas cette possibilité et il fallait nécessairement faire les tris avant la fusion.
- ▶ **Le quadrant (3)** présente les variables communes aux deux fichiers comme « [variables-clés](#) ». Un avertissement vous renvoie cependant à aller au deuxième onglet « [Variables](#) » pour ajouter ou supprimer des valeurs clés



DataSet

- ▶ **.CSV :**

- ▶ Dataset : Supermarket Sales

- ▶ Lien : <https://www.kaggle.com/datasets/faresashraf1001/supermarket-sales>

- ▶ **.SAV :**

- ▶ Dataset : **Enquête sur la population active, décembre 2023 [Canada]**
EPA_decembre_2023.sav

- ▶ **Lien :** https://odesi.ca/fr/d%C3%A9tails?id=/odesi/doi_10-5683_SP3_GAOXM4.xml

ANALYSES DES DONNEES

Introduction

- ▶ Cette section s'intéresse à l'analyse statistique sous SPSS, nous traiterons essentiellement des statistiques simples et des graphiques.
- ▶ L'ensemble des outils d'analyses statistiques sont disponibles dans le menu « **Analyse** » de SPSS

Analyse de puissance	>
Rapports	>
Statistiques descriptives	>
Statistiques de Bayes	>
Tableaux	>
Comparer les moyennes	>
Modèle linéaire général	>
Modèles linéaires généralisés	>
Modèles Mixtes	>
Corrélation	>
Régression	>
Log Linéaire	>
Réseaux neuronaux	>
Classifier	>
Réduction des dimensions	>
Echelle	>
Tests non paramétriques	>
Prévisions	>
Survie	>
Réponses multiples	>
 Analyse des valeurs manquantes	>
Imputation multiple	>
Echantillons complexes	>
 Simulation...	>
Contrôle de qualité	>
Modélisation spatio-temporelle...	>
Marketing direct	>

LES STATISTIQUES SIMPLES (UNIVARIEES, BIVARIEES)

- ▶ SPSS offre de nombreuses commandes pour l'analyse des données, autant sous forme de **tableaux** que de **graphiques**.
- ▶ La forme des tableaux dépend du nombre de variables croisées.
 - ▶ Lorsqu'une **seule variable** est décrite, on parle de distribution **univariée**, ou de **fréquence simple**.
 - ▶ Pour **deux variables**, on parle de distribution **bivariée**, et, au-delà de deux variables, de distribution **multivariée**.
- ▶ Plus le nombre de variables est élevé, plus la lecture sera complexe : un tableau croisant plus de **quatre variables** est souvent incompréhensible pour une personne normalement constituée (y compris un statisticien).

LES STATISTIQUES SIMPLES (UNIVARIEES, BIVARIEES)

Résumé des variables

- ▶ La première chose à faire avant de travailler sur un fichier est d'examiner **l'ensemble** des variables, pour détecter d'éventuelles erreurs à la saisie ou lors du transfert des données, et surtout pour **se familiariser avec les données**.
- ▶ Dans le menu « **Analyse** », la boîte d'outils « **Rapports** » permet de générer des informations **résumées** sur une ou plusieurs variables

«**Analyse**» → «**Rapports**» → «**Livre de codes**»

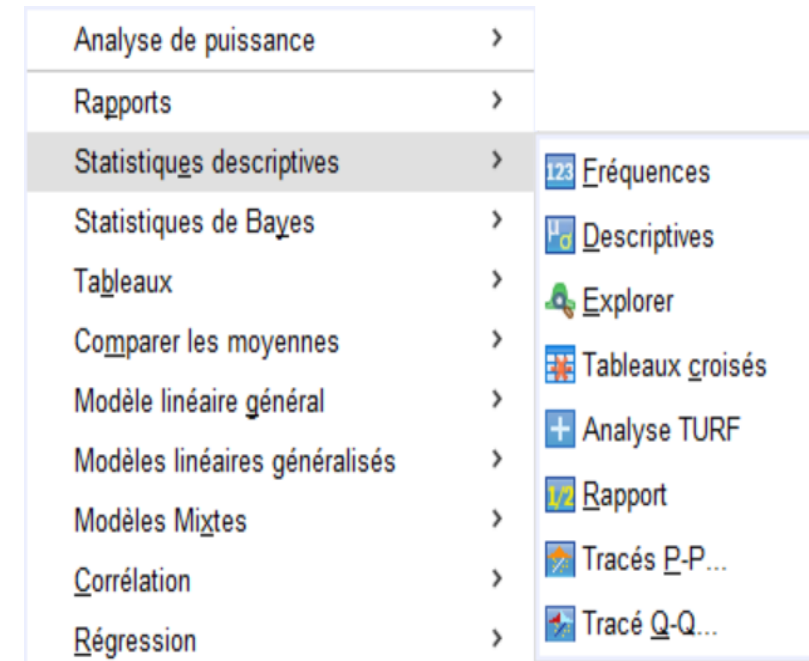
Livre de codes

Score_G		Valeur
Attributs standard	Position	3
	Libellé	<sans>
	Type	Numérique
	Format	F3
	Mesure	Echelle
	Rôle	Entrée
N	Valide	995
	Manquant	0
Tendance et dispersion centrales	Moyenne	96,43
	Ecart type	25,784
	25ème percentile	76,00
	50ème percentile	95,00
	75ème percentile	118,00

LES STATISTIQUES SIMPLES (UNIVARIEES, BIVARIEES)

Statistiques univariées

- ▶ SPSS dispose d'une boîte d'outils variés pour la production des statistiques descriptives (fréquences, tableaux croisés, etc.). Pour accéder à cette boîte d'outils aller dans le menu « **Analyse** » puis « **Statistiques descriptives** ».
- ▶ Pour réaliser les fréquences d'une variable, on utilise son assistant accessible via : « **Analyse** » → « **Statistiques Descriptives** » → « **Fréquences** ».
- ▶ Pour les variables de mesure « **Echelle/Scale** » autrement dit les variables **continues**, on peut produire les paramètres de tendance centrale et dispersion.





LES STATISTIQUES SIMPLES (UNIVARIEES, BIVARIEES)


► **Statistiques univariées**


► « Statistiques descriptives »


1. Fréquences
2. Descriptives
3. Explorer
4. Tableaux croisés
5. Analyse TURF (Total Unduplicated Reach and Frequency)
6. Rapport
7. Tracés P-P (probabilité-probabilité)
8. Tracé Q-Q (quantile-quantile)

 Fréquences


 Descriptives


 Explorer

 Tableaux croisés

 Analyse TURF

 Rapport

 Tracés P-P...

 Tracé Q-Q...

LES STATISTIQUES SIMPLES (UNIVARIEES, BIVARIEES)

Statistiques bivariées

Variables qualitatives

- ▶ Pour l'analyse de deux variables de mesure **nominale** on utilise généralement **un tableau croisé**.
- ▶ Le test d'indépendance de **Khi-deux** permet par ailleurs de se prononcer sur la présence éventuelle d'une **liaison** entre les deux variables.
- ▶ Sous SPSS, les tableaux croisés sont réalisés grâce à la boîte d'outils « **Statistiques descriptives** » à travers l'assistant « **Tableaux croisés** » accessible via : « **Analyse** » → « **Statistiques Descriptives** » → « **Tableaux croisés** »

LES STATISTIQUES SIMPLES (UNIVARIEES, BIVARIEES)

Statistiques bivariées

Variables quantitatives

- ▶ Lorsqu'on a à faire deux variables **continues**, les outils d'analyse descriptives sont en général **le coefficient de corrélation** qui détermine s'il y a une liaison **linéaire** entre les deux variables.
- ▶ Sous SPSS, pour calculer le coefficient de corrélation entre deux variables **quantitatives** on utilise l'assistant disponible dans la boîte à outils « **Corrélation** » accessible : « **Analyse** » → « **Corrélation** » → « **Bivariée** »
- ▶ Il existe plusieurs coefficients de corrélation dans SPSS :
 - ▶ **Pearson**: Permet d'étudier la linéarité entre deux variables continues ;
 - ▶ **Spearman** (Pearson basé sur les rangs) : Permet de décrire dans quelle mesure la liaison entre deux variables est monotone. Il est utile pour les variables quantitatives non normales ou les variables qualitatives ordinales.
 - ▶ **Kendall tau-b** (basé sur le nombre de concordances et discordances des rangs) : pour des variables ordinales

Coefficients de corrélation

Pearson Tau-b de Kendall Spearman

Corrélations

Statistiques descriptives

	Moyenne	Ecart type	N
Score_G	95,62	25,664	495
Poids	69,64	16,701	495

Corrélations

		Score_G	Poids
Score_G	Corrélation de Pearson	1	,040
	Sig. (bilatérale)		,378
	N	495	495
Poids	Corrélation de Pearson	,040	1
	Sig. (bilatérale)	,378	
	N	495	495



Travaux Pratiques