

Tel: (226) 25 37 46 55 / 70 80 64 52 Email: cres.courriel@gmail; cres@iavs-edu.org

Site web: cres-edu.org

COURS D'INITIATION AU LOGICIEL SPSS

BAKO Dramane

Ingénieur Statisticien Economiste

Sommaire

Ir	itroduc	tion .		3		
1	L'er	nviro	nnement graphique SPSS Windows	3		
2	2 Démarrage					
3	Les	fenê	ètres et leurs contenus	4		
	3.1	La f	enêtre des données	5		
	3.2	La f	enêtre des résultats	5		
	3.3	La f	fenêtre des syntaxes	6		
4	Mar	nipula	ations des données d'un fichier	6		
	4.1	Car	actéristiques d'une variable	6		
	4.2	Inse	ertion et trier	8		
	4.2.	.1	Insertion de variables et d'observations :	8		
	4.2.	.2	Trier les données selon une ou plusieurs variables :	9		
	4.3	Agr	éger les observations	9		
	4.4	Por	ndérer ; calculer, recoder	11		
	4.4.	.1	Pondérer les observations :	11		
	4.4.	.2	La fonction calculer	12		
	4.4.	.3	Recoder une variable	14		
5	Laf	fusio	n des fichiers	16		
	5.1	Ajo	uter des observations	17		
	5.2	Ajo	uter des variables	19		
	5.3	Tab	ole de consultation	20		
۵	Loc	table	COLLY COLLE SDSS	22		

Introduction

Le logiciel SPSS est un instrument particulièrement adapté à la mise en œuvre de techniques d'analyse des données statistiques. Il favorise la gestion des données dans un environnement graphique convivial associant menu descriptif et boites de dialogue. En plus cet environnement offre un langage de commande qui permet d'écrire des programmes afin d'optimiser des taches de production. Il permet de traiter les données avec efficacité et d'effectuer des analyses sur de grandes bases de données. Il offre plusieurs possibilités pour organiser et synthétiser les informations statistiques.

L'exploitation et le traitement statistique des données d'enquête de terrain supposent toujours un travail préalable d'organisation et de restructuration des informations. SPSS présente toutes les fonctionnalités requises pour effectuer ce travail. Ce cours d'initiation porte sur la dernière version du logiciel : PASW Statistics 18.

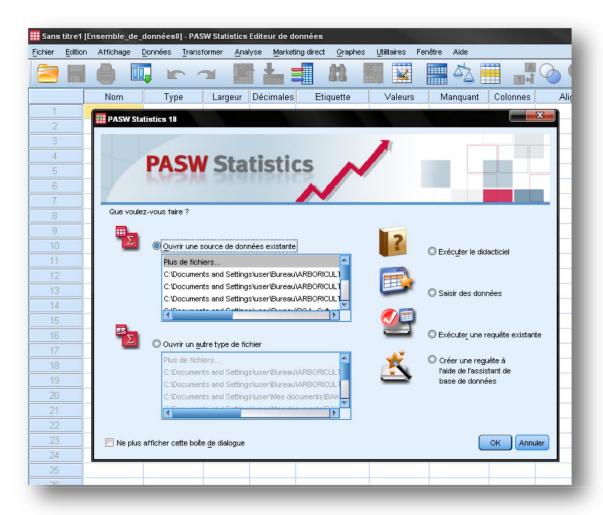
1 L'environnement graphique SPSS Windows

Le logiciel SPSS fonctionne dans un environnement graphique Windows, son utilisation repose largement sur la manipulation de fenêtres et de menus déroulants. Le mode de fonctionnement de ces menus et la gestion des fenêtres répond donc aux principes généraux du système d'exploitation Windows.

2 Démarrage

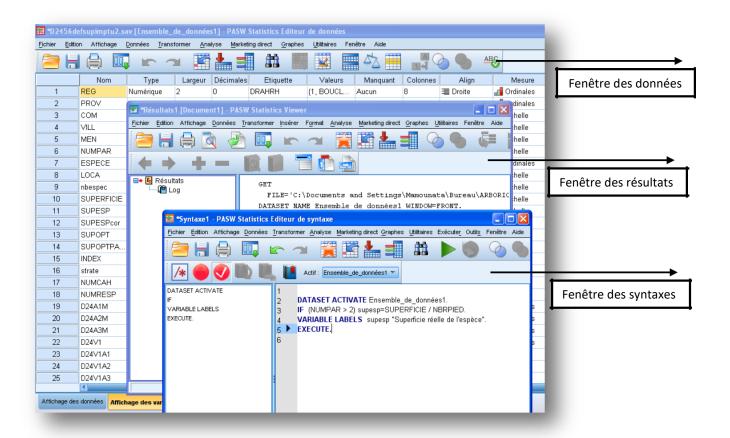
Après l'installation du logiciel, pour le démarrage il faut aller dans *Démarrer* + *Tous les programmes* + *SPSS Inc* + *PASW Statistics 18* et cliquer sur *PASW Statistics 18*. Toutefois il est conseillé de faire un click droit et créer un raccourci sur le bureau pour ne pas reprendre chaque fois ce long processus.

La fenêtre de démarrage se présente comme l'indique l'image ci-dessous. Cette fenêtre donne la liste des fichiers SPSS récemment ouverts. Pour ouvrir un fichier, il suffit de faire un double click sur son nom. La fenêtre de démarrage permet d'exécuter le didacticiel (aide), d'exécuter ou de créer un requête d'importation de données d'un fichier non SPSS.



3 Les fenêtres et leurs contenus

Le logiciel SPSS est composé de trois fenêtres : la fenêtre des données (ou fenêtre de l'éditeur), celle des résultats (souvent appelée viewer) et celle des syntaxes.



3.1 La fenêtre des données

La fenêtre des données permet d'entrer les données. Cette fenêtre présente deux onglets (en bas à gauche) : affichage des variables, et affichage des données. En cliquant sur affichage des variables, vous aurez l'ensemble de vos variables si vous avez déjà ouvert une base. Vous pouvez aussi ajouter ou supprimer des variables dans cette fenêtre.

3.2 La fenêtre des résultats

Elle affiche les résultats des commandes exécutées par SPSS. C'est dans cette fenêtre que nous observerons les sortis de tableaux, de graphiques, etc. Cette fenêtre s'ouvre automatiquement dès que vous lancez une procédure qui doit afficher des résultats.

3.3 La fenêtre des syntaxes

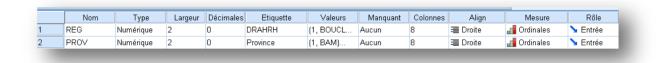
Elle permet de saisir les commandes SPSS. En effet, SPSS présente une configuration Windows avec menu déroulant (on peut donc travailler avec la souris) et une possibilité de saisir les programmes en langage SPSS. SPSS permet en effet d'obtenir la syntaxe de toute opération avec le bouton *coller*.

4 Manipulations des données d'un fichier

Dans cette section nous nous intéresserons à quelques manipulations courantes sur les variables ou sur les observations.

4.1 Caractéristiques d'une variable

On ne peut bien manipuler les données sans connaître parfaitement les caractéristiques des variables. Remarquez dans SPSS que la fenêtre des données, dans l'affichage des variables, les caractéristiques suivantes: *Nom, Type, Largeur, Décimales, Etiquette, Valeur, Manquante, Colonnes, Align, Mesure et Rôle.*



- Le **nom** de la variable est un abrégé (au maximum en 8 lettres) de la variable. Par exemple pour la question : quel est l'âge du chef de ménage ? On peut prendre comme **nom** de la variable : agecm (cm mit pour chef de ménage).
- Le *type* de la variable spécifie la nature numérique, alphanumérique, format date, etc. de la variable. Il est fortement conseiller de mettre les variables sous forme numérique (moyennant une codification) car SPSS se retrouve très facilement avec des chiffres. Par exemple, lorsque vous avez une variable telle que statut matrimonial qui présente cinq modalités : célibataire, marié, divorcé, veuf, autre, codifiez les modalités en 1, 2, 3, 4 et 5 et dès lors vous pouvez prendre le type numérique dans SPSS. Cependant il appartient aux statisticiens de savoir que cette variable n'est pas numérique au sens de la statistique descriptive, donc sa moyenne, par exemple, n'a aucun sens.



Virgule et points désignent les types de séparateurs de milliers des nombres ; chaîne est relative aux chaînes de caractère non numériques

• La *largeur* de la variable indique

le nombre de positions occupées par les valeurs prises par cette variable. Par exemple pour une variable concernant l'âge d'une personne nous pouvons prendre comme largeur 3 car aucune personne ne peut dépasser 999 ans. Pour une variable nom de famille du chef de ménage nous pouvons prendre jusqu'à 40 par exemple. Attention à ne pas prendre des largeurs inutilement élevées car SPSS va réserver une grande mémoire pour la variable ; ce qui va contribuer à alourdir le fichier.

- **Décimal** désigne le nombre de décimal à *afficher* après la virgule.
- **Etiquette** de variable: ce n'est pas tout le monde qui peut décoder le nom de la variable *agecm*; vous pouvez spécifier carrément la phrase qui forme la variable dans la fenêtre étiquette, on peut donc préciser : *Age du chef de ménage*.
- Valeur: vous pouvez oublier les codifications que vous avez faites pour la variable statut matrimonial. Pour éviter des confusions, précisez dans la fenêtre valeur: 1 = " célibataire"; 2 = "marier"; 3 = "divorcé"; 4="veuf" 5 = "autre". Pour cela il suffit de taper 1 dans la fenêtre valeur puis célibataire dans la fenêtre étiquette de valeur, puis cliquez sur ajouter. Reprendre la même chose pour les autres modalités. Le bouton orthographe permet une vérification de l'orthographe des



- *Manquant*: cette partie permet de donner une valeur pour les individus qui n'ont pas de réponse pour une question donnée. Par exemple, pour la variable *agecm* nous pouvons prendre comme système manquant le chiffre 999. **Attention**: l'on doit s'assurer que la valeur choisie pour maquant ne peut en aucun cas se confondre à une vraie valeur qu'un individu a donnée comme réponse. Par exemple pour *agecm* on ne peut pas prendre 99 comme système manquant car un chef de ménage peut avoir 99 ans. Pour le *genre* coder en 1 = "masculin" et 2 = "femme" on peut prendre 9 comme système manquant.
- *Colonne* : indique la largeur de la colonne de la variable en question
- *Aligner*: indique si les valeurs de la variable doivent être centrées dans les cellules ou aligner à gauche ou à droite. Par défaut SPSS aligne les nombres à droite et les chaînes à gauche.
- **Mesure**: indique l'échelle de mesure de la variable. Il y'a 3 échelles : **nominal** pour les variables discrètes non ordonnées (comme le numéro de la région ou la province) ; **ordinal** pour les variables discrètes ordonnées et **échelle** désigne les variables continues. Il convient de bien renseigner la mesure car SPSS en tient compte dans certaines opérations comme la tabulation.
- **Rôle**: permet de donner le rôle de la variable dans les futures analyses (variable dépendante, indépendantes, discriminante).

Application : saisir une base de données de 10 élèves d'une école en donnant pour chaque élèves les informations suivantes : nom, prénom, classe, sexe, date de naissance, âge, montant de la bourse, moyenne, rang.

4.2 Insertion et trier

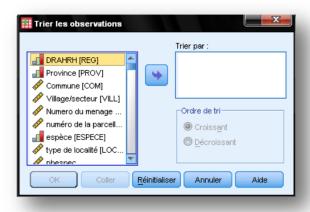
4.2.1 Insertion de variables et d'observations :

Pour insérer une variable placez vous dans la fenêtre affichage de variables, cliquez sur données, puis sur insérer une variable: SPSS insère la variable juste en haut de votre sélection. C'est à vous de préciser les bonnes caractéristiques de la variable après insertion.

Pour insérer une observation placez vous dans affichage observation, cliquez sur données puis sur insérer une observation.

4.2.2 Trier les données selon une ou plusieurs variables :

Pour trier votre fichier selon une ou plusieurs variables, placez vous dans affichage des données, cliquez sur données, puis sur trier les observations ; vous verrez le menu suivant qui contient trois zones : la zone dans laquelle vous avez les variables, la zones au dessus de laquelle il est écrit trier par : c'est dans cette zone que vous devez envoyer le critère du tri ; et la zone qui permet de spécifier l'ordre du tri (ordre croisant ou décroissant).



Vous pouvez trier selon plusieurs variables, telle variable en ordre croissant telle autre en ordre décroissant. Pour trier le fichier procédez comme suit :

- 1. Choisissez l'ordre du tri;
- 2. Sélectionner la variable selon laquelle (par ordre de tirage) vous voulez trier ;
- 3. Envoyer cette variable dans la zone trier par ; puis cliquez sur ok.

Si vous avez plusieurs variables critères, répétez les étapes 1 ; 2 ; 3 par ordre de priorité dans le tri (c'est-à-dire la machine va trier d'abord selon la première variable choisie, ensuite selon la deuxième, ensuite....).

4.3 Agréger des variables

Supposons que nous avons un fichier contenant comme individus des ménages, et d'autres variables telles que la province du ménage, la superficie agricole du ménage, etc.

Nous savons que la superficie agricole totale d'une province est la somme des superficies agricoles des ménages de la province en question ; la superficie agricole

moyenne par ménage d'une province est la moyenne de la variable superficie dans la province en question.

La commande agréger permet de calculer ces nouvelles variables.

Procédez comme suit : dans affichage des données, cliquez sur donnée puis sur agréger, le menu suivant apparaît

Fenêtre d'agrégation des variables



 Envoyez dans la fenêtre critère (s) d'agrégation la variable province car nous voulons ici (par exemple) la superficie agricole totale de chaque province (Ajouter ou non la région n'a pas d'impact sur les résultats car la province est incluse dans la région).

- Envoyez dans la fenêtre Récapitulatifs de variable(s) la variable superficie (car c'est elle qu'on veut agréger).
- 3. En bas de votre fenêtre, vous avez 3 options :
 - Ajouter les variables agrégées dans votre fichier de travail
 - Créer un nouveau fichier contenant seulement les variables agrégées : dans ce cas le logiciel créé le fichier et l'ouvre automatiquement
 - Créer un nouveau fichier contenant seulement les variables agrégées : dans ce cas le logiciel créé le fichier avec un nom et un emplacement que vous indiquer avec le bouton Fichier.
- 4. Par défaut SPSS vous donnera *la valeur moyenne* alors que nous voulons (par exemple) le revenu total par province.
- 5. Cliquez sur fonction, la fenêtre du graphique d'en haut (à droite) apparaît. Dans cette fenêtre vous avez une panoplie de fonctions, nous choisissons ici : somme des valeurs. Faites encore entrer la variable superficie et garder la fonction moyenne.
- 6. Puis changer les noms et étiquettes de vos nouvelles variables avec le bouton Nom et étiquette puis faire OK.

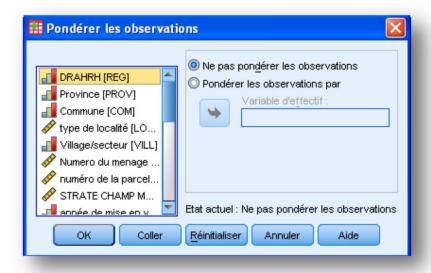
4.4 Pondérer; calculer, recoder

4.4.1 Pondérer les observations :

Il arrive souvent qu'après une enquête qu'on cherche à extrapoler les résultats de l'enquête sur toute la population. Pour cela, le statisticien doit calculer ce qu'on peut appeler un coefficient de pondération (ou plus précisément un coefficient d'extrapolation). Donc à chaque individu sera rattachée une valeur (de la variable coefficient d'extrapolation).

Ainsi, avant la sortir les résultats, il convient d'indiquer à SPSS cette variable d'extrapolation. Pour ce faire : cliquez sur donnée, puis sur pondérer les observations, vous aurez la fenêtre suivante.

Présentation de la fenêtre de pondération

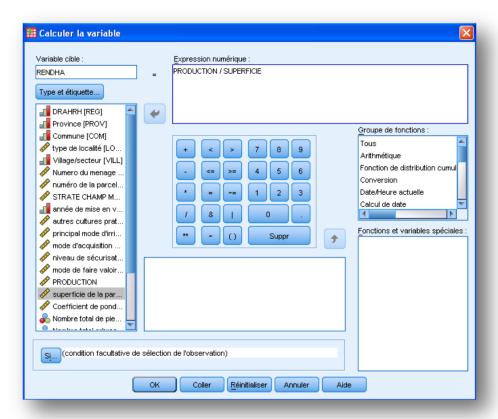


Vous avez deux options: **ne pas pondérer les observations** (donc un individu vaut un individu) ; **pondérer les observations par**. C'est cette dernière option qui nous intéresse. Faites passer du côté droit la variable de pondération. Puis cliquez sur ok.

4.4.2 La fonction calculer

Cette fonction permet de créer une nouvelle variable à l'aide notamment des anciennes variables. Par exemple pour avoir une variable *rendement à l'hectare*, il suffit de diviser la *production* par la *superficie*. On dispose dans les options de cette fonction des fonctions élémentaires tel que : le logarithme, l'exponentiel, la valeur absolue, et même la fonction **concaténer** qui permet de créer les identifiants.

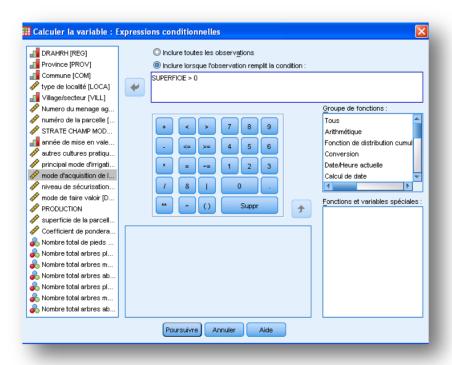
Pour utiliser cette fonction cliquer sur **Transformer**, puis sur **calculer la variable**, vous aurez la fenêtre suivante :



Suivez les étapes :

- 1. saisissez dans la zone variable de destination le nom de la nouvelle variable que vous voulez calculer (par exemple RENDHA : pour dire *rendement à l'hectare*).
- 2. Et posez vous tout de suite la question de savoir le type de la variable que vous venez de nommer ; donc juste en bas de la zone Variable cible, cliquez sur type et étiquette pour précisez surtout le type de la variable. Cette étape ne doit pas être sautée car certaines fonctions ne sont valables que pour certains types de variables.
- 3. Maintenant, à l'aide des flèches et de la calculette écrire l'expression de la nouvelle variable dans la zone Expression numérique.

L'option **SI** permet de définir une condition pour laquelle cette variable doit être calculée; par exemple si la superficie est non nulle.



Exemple : création d'un identifiant après une enquête.

- 1. Supposons que dans l'enquête nous avons le numéro de la province, le numéro du village, le numéro du quartier. Et supposons que notre identifiant s'obtient en écrivant cote à cote ces différents numéros.
- 2. En utilisant la fonction calculer, procédez comme suit :
 - 2.1 Saisissez comme nom de la variable les lettres : indent ;
 - 2.2 Dans type et étiquette prenez chaîne et prenez 30 pour largeur.
 - 2.3. Dans fonction choisissez la fonction **concat** pour l'envoyer dans la fenêtre d'expression numérique. Et écrivez dans cette dernière fenêtre concat (depart, vilag, quartier). Puis cliquez sur ok.

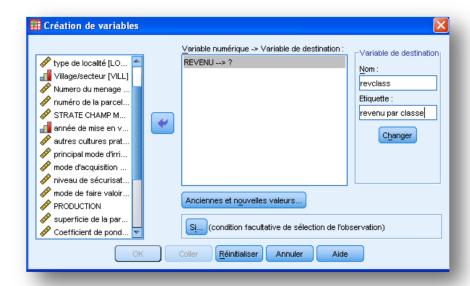
4.4.3 Recoder une variable

Il arrive très souvent de vouloir recoder une variable en classe de valeur. Par exemple lorsqu'on a le revenu des ménages, on peut vouloir recoder cette variable en cinq

groupe de 1 à 5. Par exemple si revenu < 50 000 alors revclass = 1 ; si 50 000 <= revenu < 100 000 alors revclass = 2 ; etc.

Exemple de recodage de la variable revenu :

- 1. Cliquez sur **transformer**; vous aurez deux options de recodage: **recodage des** variables et création de variables.
- 2. Prenez l'option création de variables que permet de préserver la variable source, contrairement à l'option recodage de variable qui écrase la variable source. Après un clic sur l'option création de variable vous obtenez la fenêtre suivante :



- Faire passez la variable que voulez recoder dans la fenêtre variable numérique variable.
- 4. Dans la zone nom écrivez revclass ; comme étiquette on peut mettre : revenu par classe.
- Maintenant cliquez sur anciennes et nouvelles valeurs. Vous aurez la fenêtre ci dessous. C'est dans cette fenêtre que vous devez spécifier le mode de recodage.



- 6. Les anciennes valeurs peuvent être des valeurs ponctuelles, spécifiées en intervalle ou des valeurs manquantes. Pour notre cas c'est des intervalles qui nous intéressent.
- 7. Saisissez dans la zone intervalle : 0 à 50000 ; puis tapez 1 dans la zone valeur en haut à droite, puis cliquez sur ajouter. Encore tapez dans intervalle 50000 à 100000 puis dans valeur mettez 2 puis cliquez sur ajouter, ainsi de suite jusqu'à la fin.

Remarque ne pensez pas qu'il aura une ambiguïté sur les bornes : tous les 50000 seront coder en 1, les 100 000 en 2.

8. Cliquez sur poursuivre, ainsi vous revenez à la première fenêtre, cliquez sur remplacer puis sur ok. Et voila votre variable codifiée.

5 La fusion des fichiers

Il arrive fréquemment de vouloir récupérer une information dans un autre fichier. Par exemple on peut ajouter des observations, ou bien des variables à notre fichier de travail. La chose qui nous semble géniale est ce qu'on appelle table de consultation. Les lignes qui suivent donnent les procédures pour effectuer ces manipulations.

5.1 Ajouter des observations

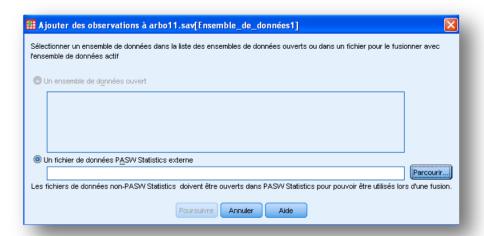
L'idée est la suivante : nous avons effectué une enquête dans les départements de Tchériba et de Saponé. Après saisie dans les deux centres nous obtenons deux fichiers l'un pour les individus de Tchériba (par exemple 500 individus) et l'autre pour les individus de Saponé (par exemple 600 individus), donc les deux fichiers ont les mêmes variables mais les individus diffèrent. Or nous voulons avoir des résultats qui concernent l'ensemble. Nous devons donc fusionner les deux fichiers, et nous aurons ainsi 500 + 600 = 1 100 individus. Les tableaux suivants schématisent la situation.

Fusion des deux fichiers : Tchériba + Saponé

individus	V1	V2	 		VJ-1	٧J
1						
2						
500						
501						
1 100	I			I		

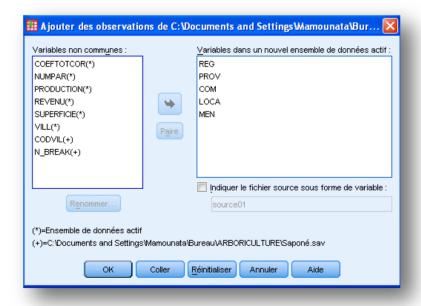
Mise en œuvre : procédez comme suit : ouvrez le fichier Tchériba, cliquez dans données, placez vous sur fusionner les fichiers, puis cliquez sur ajouter des observations. Vous aurez la fenêtre suivante.

Première fenêtre de l'ajout des observations



Si le fichier de Saponé est déjà ouvert, il apparaitra automatiquement dans le carré en dessous de l'onglet **Un ensemble de données ouvert** et il suffit de cliquer sur son nom puis sur **poursuivre**. Si le fichier n'est pas ouvert, cliquer sur **parcourir**. La machine vous demande de préciser l'autre fichier. Donc ouvrir le fichier de Saponé, vous aurez la figure suivante qui contient deux grande zone : **variables non communes** (cette contiendra les variables qui ne figure pas à la foi dans les deux fichiers). La zone : **variable dans un nouvel ensemble de données actif** (c'est les variables que SPSS vous propose pour la construction du fichier concaténé). Cliquez sur ok, et le tour est joué, le fichier Tchériba contient maintenant toutes les observations.

Deuxième fenêtre de l'ajout des observations



5.2 Ajouter des variables

L'ajout de variables a des ressemblances avec l'ajout des observations. Cette fois-ci on a un fichier des ménages (par exemple 100 ménages) avec des variables démographiques et un autre fichier des mêmes ménages avec des variables économiques. On veut fusionner ces deux fichiers.

Fichier démographique

individus	D1	D2	 DJ-1	DJ
1				
2				
100				

Fichier économique

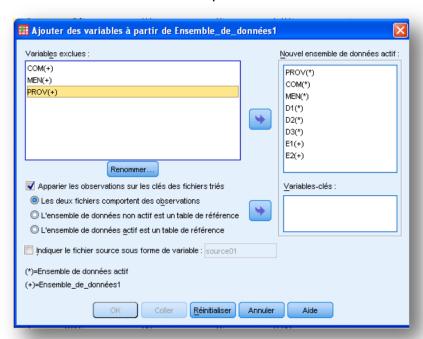
_	individus	E1	E2	 EK-1	EK
ı	1				
	2				
	100				

Le fichier fusionner donne

individus	D1	D2	 DJ-1	DJ	E1	E2	 EK-1	EK
1								
2								
•								
100								

Mise en pratique :

- trier d'abord les deux fichiers suivant la variable clé : individus, cette étape est plus que obligatoire (il peut avoir plusieurs variables clés, ne pas oublier de les trier toutes à l'ordre croissant).
- ouvrez le fichier démographique, cliquez sur **données**, placez vous sur **fusionner les fichiers**, puis cliquez sur **ajouter des variables**, la machine vous oriente dans le choix de l'autre fichier (il s'agit ici du fichier économique), ouvrez donc le fichier économique. Vous aurez la fenêtre suivante :



A gauche nous avons les Variables exclues (variables appartenant aux deux fichiers); à droite nous avons l'ensemble des variables du fichier fusionné. Lorsqu'il y'a plus d'une variable exclue, il faut choisir les variables d'appariement. Pour cela cochez Apparier les observations sur les clés des fichiers triés. Si les fichiers sont au même niveau (par exemple niveau ménage, on peut choisir l'option Les deux fichiers comportent des observations sinon il faudra choisir une table de référence.

5.3 Table de consultation

La table de consultation est légèrement distincte des deux cas précédents. Le schéma est le suivant : nous avons deux fichiers concernant deux enquêtes. Une sur les ménages et l'autre les des individus tiré dans les ménages. Donc nous avons un fichier ménages et un fichier individus. Généralement le fichier individus est largement

supérieur (en terme de taille) au fichier ménage car dans un ménage donné on peut tirer plusieurs personnes.

Supposons que nous travaillons pour le moment sur le fichier des individus (c'est-à-dire des personnes tiré dans les ménages). Supposons qu'on a besoin du quartier des individus personne dans leur fichier. Donc il s'agit des créer une variable quartier dans le fichier des individus, toute les personnes issue d'un même ménage auront certes le même quartier. Pour cela on demande à la machine le travail suivant : prendre la variable quartier dans le fichier ménages et de l'amener dans le fichier individus de la façon suivante : dès que deux personnes sont dans même ménage alors ils auront même quartier. Regarder le schéma suivant pour voir claire :

Le fichier ménage (contient 5 individus statistiques)

Num.	quartier	V2	V3	V4
01	Cissin			
02	Ouidi			
03	Wayalguin			
04	Cissin			
05	Ouidi			

Le fichier des individus (contient 13 individus statistiques)

Num.ind	ménage	T2	quartier
01	01		?
02	01	i	?
03	01	ē	?
04	02		?
05	02		?
06	03		?
07	03		?
08	03		?
09	04		?
10	04		?
11	04		?
12	05		?
13	05		?

Le fichier construit à partir des deux fichiers

Num. ind	ménage	T2	quartier
01	01		Cissin
02	01		Cissin
03	01		Cissin
04	02		Ouidi
05	02		Ouidi
06	03		Wayalguin
07	03		Wayalguin
08	03		Wayalguin
09	04		Cissin
10	04		Cissin
11	04		Cissin
12	05		Ouidi
13	05		Ouidi

Mise en pratique :

On reprend le même processus comme celui dans le cas de l'ajout de variable mais en cochant cette fois l'option « *l'ensemble des données actif est une table de référence* » ou « *l'ensemble des données non actif est une table de référence* » selon le fichier sur lequel on travaille est le fichier ménage ou non.

6 Les tableaux sous SPSS

Remarque importante : avant de commencer la tabulation, il convient de s'assurer que l'échelle de mesure de la variable est bien définie. Les variables qualitatives doivent notamment être définies avec une échelle nominale ou ordinale.

Pour réaliser un tableau, il suffit d'aller dans le menu analyse et cliquer sur tableaux. La fenetre ci-dessous s'ouvre :

