

Guide de Formation au SPSS 10.0 pour Windows

Analyse transversale

Matériel du Cours de Formation Ponctuelle

**Comment Conceptualiser une Recherche Pertinente de Politique Générale et Assurer le
Traitement et l'Analyse des Données avec
SPSS 10.0 pour Windows
4e Édition**

**Département d'Économie Agricole, Michigan State University
East Lansing, Michigan
Octobre 2002**

Éléments du Matériel du Cours de Formation aux Techniques de l'Analyse Transversale

Module 0 - Introduction à la structure des fichiers sur SPSS 10.0 pour Windows (*Data et Syntax Editors, et Viewer - Output Navigator*; c'est-à-dire les éditeurs et les outils de visualisation des données et syntaxe). À lire avant de commencer la session .

Module 1 - Fonctions de base

Module 2 - Visualisation des tableaux, et agrégation

Module 3 - Tableaux, & questions à réponses multiples

Module 4 - Graphiques, tableaux, publications et présentations, comment les exporter au traitement de texte

Annexes

1. Présentation des filtres versus sélections des données de façon permanente. Mise en graphique et options graphiques des données sur SPSS 10.0 pour Windows
2. Six pages extraites de l'enquête socio-économique sur les petits producteurs de la province du Nampula au Mozambique (NDAE Working Paper 3, 1992).
3. Analyse par ordinateur des données d'enquête - Organisation des fichiers de données à niveaux multiples par Chris Wolf, MSU, au département d'Économie Agricole. Ce fichier est téléchargeable et disponible en anglais et français à l'adresse internet suivante :
<http://www.aec.msu.edu/agecon/fs2/survey/index.htm>

Remerciements

Cette recherche a été financée par l'accord de coopération Sécurité Alimentaire II entre le Département d'Économie Agricole de Michigan State University et l'Agence Américaine pour le Développement International (USAID), Bureau Mondial, Branche de l'agriculture et de la sécurité alimentaire.

Guide de Formation au SPSS pour Windows
MODULE 0 - Niveaux, séries chronologiques et structure des fichiers sur SPSS 10.0 pour Windows
(Fenêtres pour Données, Syntaxes et Résultats)

**Guide de Formation pour la recherche orientée
vers le développement de politiques
et pour le traitement et analyse de données par
SPSS 10.0 pour Windows
4e Édition**

**Matériel de Cours de Formation Ponctuelle
Comment Conceptualiser une Recherche Pertinente de Politiques Générales et
Entreprendre le Traitement et l'Analyse des Données avec SPSS 10.0 pour Windows
4e Édition**

**Département d'Économie Agricole, Michigan State University
East Lansing, Michigan
Octobre 2002**

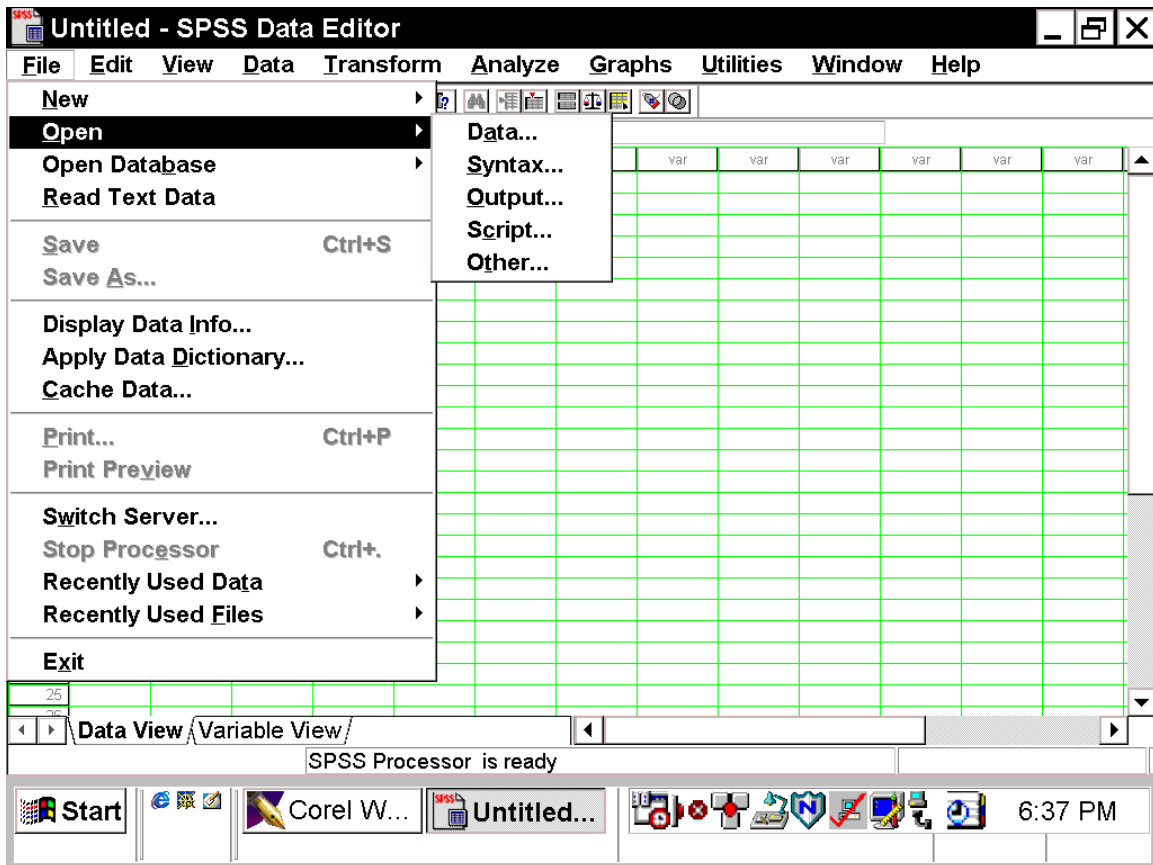
Ce module est une introduction des concepts de base tels que les niveaux, les analyses transversales, et par conséquent l'organisation des données. Ce module fait également une brève description de la structure des fichiers SPSS pour Windows version 10.0. Il est essentiel de lire ce module avant de commencer la session sur l'analyse transversale.

Fichiers utilisés dans SPSS 10.0 pour Windows

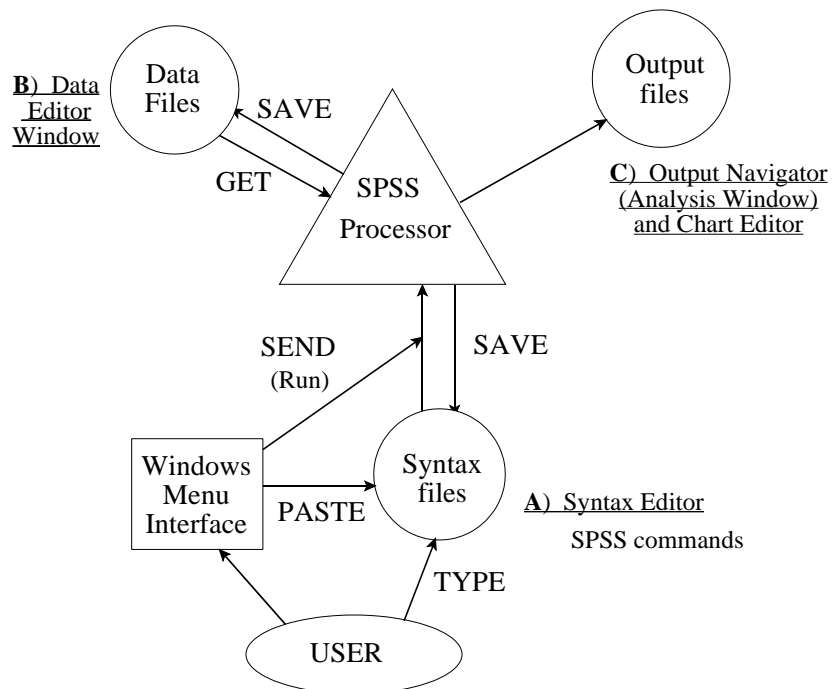
Lorsque vous utilisez SPSS 10.0 pour Windows d'après ce manuel d'instruction, vous aurez à travailler avec trois différentes fenêtres du programme : le *Syntax Editor*, le *Data Editor* et le *Viewer* (ce dernier inclut les graphiques). Le contenu de chacun (e.g. la programmation, les données et les résultats) peut être sauvegardé dans des fichiers SPSS 10.0 pour Windows du type approprié.

Lorsque vous ouvrez SPSS 10.0, dans le coin supérieur gauche de l'écran, choisissez **File**, suivi de **Open** et vous aurez à choisir une des 5 options de types de fichiers :

Data	fichiers de données	(Extension *.sav)
Syntax	syntaxe ou fichiers des commandes/programmation	(Extension *.sps)
Output	fichiers des résultats/graphiques/sortie	(Extension *.spo)
Script	programmation avancée pour utilisation avec Sax BASIC qui est créé automatiquement à chaque fois que Output est lancé	(Extension *.sbs)
Other	tous les fichiers	(*.*)



Il est important de reconnaître les différents types de fichiers et de comprendre les différentes commandes à utiliser pour créer et accéder à ces fichiers.



A) Le *Syntax Editor* (Éditeur pour la programmation)

Le Syntax Editor est la fenêtre où la syntaxe ou les commandes sont écrites avant qu'elles ne soient soumises au processeur SPSS. Pour écrire des commandes dans le Syntax Editor, vous pouvez **taper** les commandes directement dans le Syntax Editor, ou utiliser les menus pour sélectionner **Paste** quand vous avez fini de modifier la commande. Il y a quatre utilisations pour le Syntax Editor :

- Pour écrire des commandes directement ou coller des commandes du Data Editor pour être traitées plus tard par SPSS 10.0 pour Windows,
- Pour envoyer ces commandes à SPSS 10.0 pour Windows pour traitement,
- Pour écrire ou sauver ces commandes dans un fichier pour utilisation ultérieure et
- Pour retirer des fichiers de commandes que vous avez sauvés antérieurement.

Il est important de comprendre que les commandes que vous inscrivez dans le Syntax Editor ne sont exécutées (aucun fichier de sortie ne sera produit) que quand vous envoyiez les commandes au processeur. Le Syntax Editor est simplement un outil qui aide à préparer la programmation. Pour envoyer les commandes au processeur, vous devez utiliser le bouton **Run** dans la barre à outils de la fenêtre du Syntax Editor de SPSS 10.0 (ou sélectionner **Run ... Current** à partir des Menus). Une fois que vous cliquez sur le bouton Run, l'ordinateur envoie la(es) commande(s) au processeur, qui lira les commandes et les exécutera. Lorsque toutes les commandes sont traitées, SPSS ouvre le Viewer pour vous permettre d'examiner les résultats de votre(os) commande(s). Vous pouvez alors retourner au Syntax Editor pour ajouter ou éditer des commandes, faire exécuter ces changements à nouveaux pour observer les différents résultats.

Lorsque vous créez une commande dans le Data Editor, il est recommandé de voir la syntaxe des

commandes en utilisant l'option **Paste** au lieu de l'option **OK**. Ce dernier exécutera les commandes directement sans coller celles-ci dans le fichier de syntaxe. Si vous voulez faire apparaître les commandes dans le fichier des résultats (*Output*), faites ce qui suit :

Du Menu (peut importe le fichier SPSS), choisissez

Edit

Options ...

1. Choisissez l'onglet **Viewer**
2. Cliquez dans la boîte à côté de **Display commands in log**
3. Cliquez sur **OK** ou **Apply**

Lorsque chacune des étapes de l'analyse a été effectuée avec succès (ou lorsque vous êtes prêts à terminer la session SPSS 10.0 pour Windows) vous devez sauver les commandes dans un fichier pour utilisation ultérieure. Pour sauver les commandes, activez le Syntax Editor et sélectionnez **S**ave du menu File. Le fichier créé à partir du Syntax Editor s'appelle le fichier de syntaxe ou de commandes. C'est un fichier qui ne contient que des commandes; il ne contient jamais des données que vous analysez utilisant ces commandes. Vous devez sauvegarder vos données séparément comme décrit dans la section B. Nous vous suggérons d'utiliser les extensions de fichier offerts par défaut .SPS, par exemple SECTION1.SPS.

En écrivant vos commandes à un fichier de syntaxe, vous pouvez récupérer, voir, ou modifier des commandes pour les re-exécuter. Vous pouvez récupérer un fichier au menu **F**ile de n'importe quel fichier de SPSS pour Windows en choisissant **O**pen. Sélectionnez **S**yntax et choisissez le nom du fichier de travail qui vous intéresse. Une fois ouvert, vous pouvez ré-utiliser les commandes sans avoir à les retaper ou recréer à nouveau. Si vous faites des changements aux commandes que vous voulez garder, n'oubliez pas de sauvegarder à nouveau votre travail.

B) Le *Data Editor* (Éditeur de données)

SPSS 10.0 pour Windows garde en mémoire vos données sous forme de fichier de données, ce qui inclut non seulement les valeurs, mais aussi des *variable labels* et *value labels* (étiquettes), des informations sur le format, les valeurs manquantes (*missing-value specifications*), etc. Avant de faire des analyses avec SPSS 10.0 pour Windows, vous devez tout d'abord ouvrir ou créer un fichier de données. Sélectionnez **F**ile du menu, et puis **O**pen, **D**ata et cliquez sur un des fichiers. Vous avez deux choix : 1) cliquez sur **P**aste pour coller la commande au Syntax Editor et exécuter la commande, ou 2) exécutez (*run*) la commande directement de la fenêtre en cliquant sur **O**pen. Après avoir exécuté la commande, les données dans le fichier choisi sont disponibles dans l'éditeur des données (*Data Editor*) de SPSS 10.0 pour Windows.

Il y a deux visualisations possibles dans le Data Editor : **D**ata **V**iew affiche les valeurs actuelles des variables qui existent dans le fichier de données. **V**ariable **V**iew, affiche le dictionnaire sur les données qui inclut les étiquettes et autres informations. Pour passer d'une visualisation à l'autre, cliquez sur l'onglet au bas de l'écran.

Souvent vous prendrez un fichier SPSS, vous ferez de nouveaux calculs, vous créerez des variables, vous ferez des transformations, et vous sauvegarderez les nouvelles données pour une utilisation ultérieure. Par exemple, vous pourriez récupérer un fichier avec des données sur les superficies par culture, y ajouter la production par culture disponible dans un autre fichier pour calculer le rendement. Si vous voulez utiliser ces nouvelles variables plus tard, vous devez vous assurer que les nouvelles données sont sauvegardées. Pour sauvegarder un fichier de données, à partir du Data Editor, sélectionnez **Save As...** du menu **File** et donnez un nom au nouveau fichier. Notez que vous **devez** être dans la fenêtre du Data Editor pour sauvegarder vos données à moins que vous exécutiez la commande **SAVE OUTFILE** du fichier du Syntax Editor. Vous pouvez également choisir d'écrire par dessus l'ancien fichier s'il porte le même nom.

C) Le *Viewer* (l'outil de visualisation des résultats/graphiques/sorties)

SPSS 10.0 pour Windows écrit automatiquement tous les messages et les résultats résultant d'une exécution de commandes, à la fenêtre du Viewer. Par exemple, si vous exécutez la commande *frequency*, vous produirez alors un tableau de fréquences au Viewer. Pareillement, si vous générez un graphique, celui-ci apparaîtra au Viewer. Pour sauvegarder les résultats dans le Viewer à un fichier, assurez-vous d'avoir la fenêtre du Viewer active, sélectionnez **File** du menu et en suite **Save As...** Lorsque vous donnez un nom au fichier des résultats, SPSS rajoutera automatiquement l'extension *.SPO*. Il est important de sauvegarder le fichier des résultats. Ce fichier *Output* vous donne accès à vos résultats lorsque votre analyse avec le SPSS 10.0 pour Windows sera terminée. Par exemple, vous pouvez imprimer les sorties de votre travail de session pour examiner les résultats et vérifier les erreurs. Dans ce guide de formation, vous verrez comment sauver les sorties au Viewer et donner un nom de fichier pour chacun des modules. Veuillez noter que vous pouvez manipuler et gérer les résultats tout comme si vous étiez entrain d'utiliser un fichier de gestion (appelé Windows Explorer). Dans le Viewer, il y a deux sous-fenêtres : celle située à la droite, contient les résultats, et celle de gauche donne un aperçu des grandes lignes des matières. Opérant à l'intérieur de cette dernière, vous pouvez gérer les résultats en copiant, en déplaçant ou en supprimant différents éléments en refermant les tableaux ou graphiques, en renommant les noms et titres, en insérant du texte, etc.

Sommaire des différents types de fichiers de base

Syntax files (fichiers de syntaxe, programmation). Commandes sauvées dans un fichier du Syntax Editor. Ils ne contiennent pas de sorties ou de données. Les fichiers *Syntax* sont accessible à SPSS pour Windows par la commande **Open..Syntax** des menus. Tout comme avec SPSS 6.1.3., les extensions sont les mêmes : **.SPS* (était **.LOG* avec SPSS/PC+ pour DOS).

Output files (fichiers de sorties/résultats/graphiques et tableaux). Contient également soient des statistiques, des informations sur les variables, des différentes formes de présentation des résultats, selon les commandes envoyées au processeur de SPSS 10.0 pour Windows. Ces fichiers ne contiennent pas de données. Vous pouvez ouvrir les fichiers *Output* en utilisant la commande **Open..Output**. Les nouvelles extensions sont : **.SPO* (étaient **.LIS* avec SPSS/PC+ pour DOS, et **.LST* dans SPSS 6.1.3).

Data files (fichiers de données) incluant des variables d'enquête d'origine, et des nouvelles variables créées par des commandes de SPSS 10.0 pour Windows comme *COMPUTE* ou *AGGREGATE*. Vous ouvrez les fichiers *Data* de SPSS pour Windows par la commande **Open..Data**. Dans SPSS 6.1.3. et 10.0., les extensions sont : **.SAV* (était **.SYS* pour SPSS/PC+).

Guide de Formation au SPSS pour Windows

MODULE 1 - Fonctions de base: Fichiers SPSS, statistiques de base, et transformation des données

Introduction

Ce guide est un outil pour autodidacte développé dans le but d' aider tout utilisateur à se familiariser avec les commandes nécessaires pour réaliser des analyses statistiques typiques basées sur des d'enquêtes, avec le logiciel **SPSS 10.0 pour Windows**. Pour maximiser son utilité, et pour en retirer le plus de bénéfices, il est fortement recommandé de demander l'avis d'un technicien/utilisateur de SPSS pour Windows afin qu'il/elle puisse vous assister à bien démarrer et répondre à vos questions au fur et a mesure que vous travaillez de façon autonome à travers les différents modules. Néanmoins, ce guide de formation t peut aussi servir comme outil d'enseignement pour un cours, dans une classe, pour un groupe.

Une copie du questionnaire sur laquelle les données sont basées est disponible dans le rapport du projet au Mozambique **NDAE Working Paper 3: A Socio-economic survey of the smallholder survey in the province of Nampula: Research Methods** (1992). Des Copies des trois tableaux utilisés sont disponibles en annexe (pour de plus amples informations, veuillez contacter le professeur Michael Weber au webermi@msu.edu). Quatre sections du questionnaire sont présentées ci-dessous , chacune avec son fichier SPSS pour Windows respectif. Deux autres fichiers de données SPSS sont également cités, nécessaires pour faire les conversions des différentes unités.

Section du questionnaire	Fichiers des données de SPSS pour Windows
Section principale du ménage	C-HH.SAV
Tableau IA: Caractéristiques des membres du ménage	C-Q1A.SAV
Tableau IV: Détails sur la production	C-Q4.SAV
Tableau V: Ventes des produits agricoles	C-Q5.SAV
Facteurs de conversion pour calculer les kg	CONVER.SAV
Facteurs de conversion pour calculer les kcal	CALORIES.SAV

Le guide est divisé en quatre modules, prenant chacun environ deux heures. Nous vous recommandons de compléter entièrement chacun des modules lors d'une séance de travail. D'emblée, il est supposé que :

- l'utilisateur connaît Windows et sait se servir d'une souris
- que les six fichiers de données se trouvent sous c:\sample sur le disque dur de l'utilisateur. Si vous ne l'avez pas fait jusqu'à présent, vous devez copier les fichiers du fichier sample.zip dans ce dossier
- que sous **Options** du menu **Edit**, les éléments suivants sont choisis :
 - list variables* sont dans le même ordre que dans le fichier
 - list commands* dans la fenêtre *Output*
 - que les noms des variables sont affichés et non les étiquettes
 - qu'une fenêtre du Syntax Editor ne s'ouvre pas au début d'une session de travail de SPSS.

Vous pouvez modifier de cette fenêtre toutes les options de SPSS.

Important : *Souvenez-vous toujours de sauvegarder les changements aux données suite aux exercices et divers modules, utilisez à chaque fois un **nouveau** nom de fichier. Sauvegardez également les fichiers de syntaxe et des sorties que vous créez pour chacun des modules. Choisissez des noms logiques, qui vous paraissent naturels, comme *module1.sps* ou *session1.spo*. Si vous n'êtes pas sûr, n'hésitez pas à vous communiquer avec la personne qui vous assiste avec SPSS pour Windows.*

Ouvrez votre logiciel SPSS. Si vous n'avez pas lu ou complété le MODULE 0, veuillez le faire maintenant, ce qui vous permettra de clarifier les concepts de l'éditeur de syntaxe (*Syntax Editor*), où vous collez ou écrivez vos commandes, l'outil de visualisation des résultats (*Viewer*) où SPSS pour Windows affiche les résultats des commandes exécutées, et l'éditeur de données (*Data Editor*) où se trouvent les données et l'information sur les variables.

Les fichiers de données et le fichier de travail

Les données obtenues des questionnaires qui ont été saisies dans SPSS 10.0 pour Windows sont sauvegardées dans des fichiers de données (*data files*). Si vous voulez travailler avec certaines données, vous devez ouvrir le fichier de données correspondant pour qu'il soit disponible au processeur de SPSS.

Lorsqu'un fichier de données est ouvert, il est transféré du disque à la mémoire vive de l'ordinateur (la mémoire RAM), et devient le fichier de travail. Conséquemment, les données de ce fichier sont maintenant prêtes pour modification ou analyse. Commençons avec le questionnaire pour le Tableau IA: les caractéristiques des membres du ménage. Le fichier correspondant est le **C-Q1A.SAV**. Pour ouvrir ce fichier, suivez ces étapes :

1. Du menu **File**, choisissez **Open...**, et puis **Data**
Ceci ouvrira la fenêtre Open File.
2. Changez le dossier où se trouvent les données pour ce guide pratique et sélectionnez le fichier **c-q1a.sav**.
3. Cliquez sur **Paste** pour placer la commande dans le **Syntax Editor**.
Le Syntax Editor deviendra la fenêtre active et vous verrez le texte

```
GET  
FILE='C:\sample\C-Q1A.SAV'.
```

dans le Syntax Editor.
4. Cliquez sur **Run** la barre à outil, pour exécuter la commande.

Notez que la commande GET FILE que vous venez d'exécuter apparaîtra dans les sorties du Viewer.

Une fois la commande lancée, l'éditeur de données devient la fenêtre active et le fichier de données des membres du ménage est en mémoire de SPSS.

Une des informations clés à connaître sur un fichier concerne les variables que le-dit fichier contient. Ceci peut se faire en utilisant la commande **Variables...** du menu **Utilities**, soit dans le *Syntax Editor* ou avec le *Data Editor*. Cette commande vous permettra de visualiser la définition et les étiquettes des diverses variables. Alors, pour ce faire, suivez ces étapes :

1. Du menu **Utilities**, choisissez **Variables...**
2. Sélectionnez le nom de la variable et puis l'information de cette variable apparaîtra à la droite.

Cette fenêtre affiche les informations sur chacune des variables. Vous pouvez voir les noms des variables, **district, vil, ca1, ca2, ca4, ca5, ca6, et univ**, leurs étiquettes, leurs types (numérique, date, chaîne, etc.), la largeur en caractères (*display width*), le nombre de décimales (si la variable est de type numérique), et les valeurs manquantes définies par l'utilisateur.

Cliquez sur **Close** une fois que vous avez terminé.

Pour écrire toutes ces informations dans une sortie dans le *Viewer* pour une étude ultérieure, faites:

Du menu **Utilities**, cliquez sur **File Info**.


Cette commande s'exécutera immédiatement. Le Viewer deviendra actif en s'affichant à l'écran et contiendra une liste des variables avec leurs définitions respectives.

Vous pouvez aussi y voir le nom de chaque variable, leurs étiquettes, et les différents formats, e.g. F8.2 signifie largeur de 8 avec deux décimales. Cet outil est une bonne façon de cataloguer les données du fichier de travail. Par exemple, cette documentation pourrait commencer en copiant collant ces informations dans un traitement de texte.

Pour regarder la structure de chacune des variables, il y a une nouvelle façon avec SPSS 10.0. Dans la fenêtre du *Data Editor*, sélectionnez **Variable View** en bas à gauche de l'écran, plutôt que *Data View*. Vous pouvez modifier directement les caractéristiques des variables ici. Tout comme vous pouvez changer la valeur des données dans la fenêtre *Data View*. Un exemple pour la variable DISTRICT est disponible sur la page qui suit dans le tableau 1.1, accompagné d'une explication des choix dans chacune des colonnes.

Si vous voulez modifier un des paramètres de cette variable, cliquez sur sa cellule. S'il y a des choix à faire, une petite boîte ombrée apparaîtra dans le coin droit de cette cellule. Cliquez sur cette boîte pour voir les différents choix, ajouter des nouvelles options, ou voir les autres options. Dans certains cas, comme pour la largeur, les décimales, et les colonnes, au lieu d'une boîte, des flèches sont affichées pour hausser/diminuer la taille.

Exemple: Pour la variable **DISTRICT**, cliquez sur la colonne **Values**. Cliquez dans sa cellule (DISTRICT).

Vous verrez une petite boîte grise  Cliquez sur cette boîte.

Une fenêtre apparaîtra avec pour titre : **Value Labels**

Pour ajouter une nouvelle étiquette (*label*) de 4, associé à Nampula,

- entrez **4** dans la boîte de **Value** et appuyez sur la touche <tab>,
- écrivez Nampula dans la boîte de **Value Label**,
- cliquez sur **ADD**.
- D'habitude , vous auriez à sélectionner OK, mais comme nous ne voulons pas garder ce changement, cliquez sur Cancel.

Vous pouvez suivre ces étapes pour modifier ou enlever des étiquettes existantes. Vous choisissez les étiquettes et cliquez par la suite sur **Change** ou **Remove**.

Tableau 1.1. Structure de base utilisant *Variable View* avec SPSS 10.0

Numéro de la variable	Nom	Type	Largeur	Décimales	Étiquette	Valeurs	Manquants	Colonnes	Justification	Mesure
Explication	Nom de la variable	Numérique ou alpha-numérique	Espace requise pour inscrire la variable à la base de données	Nombre de chiffres à la droite de la décimale	Étiquette de la variable	Étiquettes pour les valeurs, e.g. pour des valeurs catégoriques	Valeurs manquantes établies par l'utilisateur (exemple: -99), fixent les valeurs qui seront exclues des calculs	Affiche la largeur de la variable dans <i>Data View</i>	Alignement des données dans <i>Data View</i> seulement : gauche, droite, centre	niveau de mesure de la variable : à l'échelle, nominal, ordinal
<i>Exemple:</i>										
1	District	Numérique	1	0	DISTRICT	1= MONAPO 2= RIBAUE 3= ANGOCHE	Aucun	8	Droite	à l'échelle

¹ Il y a trois catégories de mesures (niveaux) :

À l'échelle : Variables avec des valeurs qui sont généralement continues ou en intervalles (e.g. : rendement ou l'âge);

Ordinal : Valeurs ou variables alpha-numériques qui sont de type catégorique avec ordre intrinsèque (e.g. 1= court; 2= médium; 3= grand);

Nominal : Valeurs or variables alpha-numériques qui sont de type catégoriques sans ordre intrinsèque (e.g. 1= homme; 2= femme).


Statistiques de base - avec une variable

Une première étape à faire en entamant une analyse est d'obtenir des statistiques de base pour toutes les variables (e.g. les moyennes, maximum, minimum, et écarts-type). Ce type d'analyse vous aidera à trouver des erreurs de saisie, vous offrira une première impression des résultats, et vous permettra de savoir si les valeurs manquantes ont bien été définies, etc. Il est tentant parfois d'ignorer cette étape mais il reste que celle-ci vous fera gagner du temps plus tard et améliorera les analyses. Par exemple, apprendre que l'âge moyen des répondants n'est peut-être pas quelque chose qui vous intéressera pour votre analyse mais si cet âge moyen était de 91.3 ans, ceci vous alarmerait qu'il y a quelque chose d'anormal.

Les statistiques descriptives de base peuvent être obtenues à partir de deux commandes de SPSS : **Descriptives** et **Frequencies**. **Descriptives** est utilisé pour des variables continues, et **Frequencies** pour les variables catégoriques.

Une variable continue est une variable qui n'a pas de valeurs fixes tandis qu'une variable catégorique est limitée à un certain nombre de valeurs qui forment des catégories. Par exemple, dans le Tableau IA dans l'annexe : questionnaire sur le membres du ménage, la variable **ca3** (âge) est continue car l'âge peut prendre beaucoup de valeurs différentes; cependant la variable **ca2** (relation au chef du ménage) est catégorique comme les valeurs sont limitées aux catégories 1 à 6.




Commençons par examiner les données du fichier. Utilisez la fenêtre du *Data Editor* pour parcourir le fichier de données. Pour ce faire, suivez ces étapes :



1. Cliquez sur le bouton *Go To Data Editor*  sur la barre d'outil.
2. Parcourez le fichier des données.
*Notez qu'un point dans une cellule indique qu'il y a une valeur manquante soit un **sysmis**.*

Ce type d'analyse vous donnera une idée de la structure des données. Ceci pourrait aussi aider à déceler des erreurs évidentes; par exemple, une variable dont les valeurs manquent pour chaque cas. Comment savoir lesquelles des données sont continues ou catégoriques? Normalement, vous pouvez déceler cette information en regardant le questionnaire. Une fois que vous avez fait la distinction, vous pouvez exécuter les commandes appropriées pour chacune des variables. Si vous vous trompez en exécutant la commande **Frequencies** sur une variable continue, vous aurez beaucoup plus de résultats que voulus, facilement plusieurs dizaines et centaines de catégories, une pour chacune des valeurs. Si vous exécutez la commande **Descriptives** sur une variable catégorique, vous obtiendrez des résultats non significatifs comme la valeur moyenne d'une série de catégories n'explique rien.

Les Descriptives (moyennes, min, max, écarts-type)

En examinant les données, vous vous êtes rendu compte que la variable **ca3** (âge) est continue et que le reste des variables sont catégoriques. Pour exécuter les statistiques descriptives sur **ca3**, faites:

1. Du menu **Analyze**, choisissez **Descriptive Statistics...Descriptives**
Ceci ouvrira une nouvelle fenêtre
2. Sélectionnez **ca3** de la liste de gauche et cliquez sur .
*ca3 se déplacera à la boîte **Variable(s)**: à la droite*
3. Cliquez sur  pour coller la commande au *Syntax*. Si la fenêtre de l'éditeur de syntaxe ne s'est pas activée, vous pouvez y accéder en cliquant sur l'option de syntaxe qui se trouve sur la barre à outil de Windows située habituellement au bas de l'écran.)
4. Exécutez la commande en cliquant sur *Run*  situé sur la barre à outil de SPSS. (Notez que cette fois ci, il n'était pas nécessaire de déplacer le curseur comme il était déjà positionné sur une des lignes de la commande **Descriptives**.)
Le *Viewer* deviendra la fenêtre active et les résultats suivant l'exécution de la commande y apparaîtront. Vous verrez que la moyenne de l'âge (**ca3**) est de 21.34 ans (*years*).



- Autre façon utile d'examiner des variables continues, est d'exécuter la commande Frequency pour obtenir un histogramme, une distribution de la variable. De **Analyze...Descriptive Statistics...** Sélectionnez **Frequencies**.
- Choisissez **ca3** (âge) de la liste à gauche et cliquez sur .
- Cliquez sur **Charts** et choisissez Histograms, cliquez en suite sur **Continue**
- Cliquez à nouveau sur **Paste** pour mettre la commande dans le *Syntax Editor* et rendez le actif.
- Exécutez la commande en cliquant sur  sur la barre à outil de SPSS. Maintenant, observez la distribution des âges.

Sauvegarder le fichier de sortie et les résultats

Maintenant que vous avez obtenu des résultats dans le Viewer, il est toujours bon de les sauvegarder. Dans la fenêtre du Viewer, cliquez sur **File...Save as...** sur la barre à outil de SPSS à la droite en haut. Dans la boîte *File Name*, tapez **Session1** et assurez-vous que vous vous trouvez dans le dossier de votre choix où vous voulez sauvegarder ces fichiers. SPSS ajoutera automatiquement l'extension **.SPO** pour indiquer que ce fichier est un fichier de sortie/résultats.

Les Fréquences

Comme les variables **ca1** (travail a la ferme ou non), **ca2** (relation avec le chef du ménage), **ca4** (sexe), **ca5** (niveau d'éducation) et **ca6** (statut marital) sont catégoriques, nous allons exécuter la commande **Frequencies**.


- Analyze...Descriptive Statistics...** choisissez **Frequencies...**
Ceci ouvrira la fenêtre de la commande Frequencies
- Cliquez sur **Reset** pour vider la boîte de variables
- Sélectionnez **ca1** de la liste de gauche et cliquez sur 
ca1 se positionnera dans la boîte Variable(s): de droite
- Répétez l'étape 3 jusqu'à ce que les variables **ca2, ca4, ca5** et **ca6** se retrouvent tous dans la boîte **Variable(s):**
- Cliquez sur **Paste** pour inscrire à nouveau les commandes dans l'éditeur de syntaxe
- Exécutez la commande en cliquant sur  situé sur la barre à outil de SPSS

Le *Viewer* deviendra à nouveau actif et vous verrez que les résultats pour **ca1** montrent que 70.7% des membres de ménage travaillent sur une ferme. Les résultats pour **ca6** montre que 38.0% des répondants font partis des mariages monogames.


Pour une description complète des sorties provenant des commandes **Descriptive** et **Frequencies** référez-vous au manuel de *SPSS for Windows Base System User's Guide Release 10.0*, pages 213-221.

La commande Explore

Explore est une autre commande utilisée pour obtenir plusieurs types de statistiques de base. Parmi les nombreuses statistiques possibles avec cette commande, la plus utile ici est le choix de créer une liste des valeurs écartées (*outliers*). Exécutez la commande **Explore** sur la variable **ca3** (âge) en suivant les étapes suivantes:

- Du menu **Analyze...Descriptive Statistics**, choisissez **Explore...**
- Sélectionnez **ca3** de la liste de gauche et cliquez sur  à côté de **Dependent List**
- Dans le coin inférieur gauche de la fenêtre vous verrez **Display**. Cliquez sur le cercle à côté de **Statistics**.
Ceci nous donnera seulement les statistiques, pas de graphiques
- Cliquez par la suite sur **Statistics...**

Ceci affichera la fenêtre **Explore: Statistics**

5. Cliquez une fois sur le carré à côté de **Outliers**, vous posterez un ✓ dans la boîte. Vous verrez qu'il y a déjà un ✓ dans la boîte pour **Descriptives**.
6. Cliquez sur **Continue**
Vous retournerez à la fenêtre de la commande **Explore**
7. Cliquez sur **Paste** pour ajouter les commandes à l'éditeur de syntaxe et choisissez cette fenêtre
8. Sélectionnez la commande et exécutez-la en cliquant sur 

Vous pouvez voir le tableau de statistiques descriptives ce qui inclut le tableau des valeurs extrêmes, soient les 5 valeurs les plus élevées et le 5 plus basses pour la variable de l'âge (**ca3**). Vous pouvez maintenant déterminer si ces valeurs peuvent être considérées extrêmes. Les cas (*cases*) sont identifiés par leur numéro (*case number*). Référez-vous aux pages 223-230 du guide *SPSS for Windows Base System User's Guide Release 10.0* pour de plus amples explications sur la commande **Explore**.

Sauvegarder le fichier de syntaxe

Sauvegarder votre travail de syntaxe est toujours une bonne pratique. Vous pourrez être obligé d'exécuter à nouveau la même syntaxe après avoir corrigé des données, une erreur de transcription, ou dans le cas où votre ordinateur tombe en panne, ou a un problème avec SPSS ou un autre programme. Pour sauver ce fichier, ayez la fenêtre du *Syntax Editor* active, cliquez sur **File...Save as...**, dans la partie *File Name*, tapez le nom **Session1**. Souvent il est utile de sauver la syntaxe et les sorties correspondantes avec le même nom de fichier, la seule différence étant l'extension. SPSS ajoutera automatiquement l'extension **.SPS** pour le fichier de syntaxe. Vérifiez toujours que vous êtes dans le bon dossier. Enfin, vous pouvez seulement sauver les fichiers de syntaxe ou de programmation à partir du *Syntax Editor*.

Appliquez ce que vous avez appris jusqu'à présent sur les statistiques descriptives, en faisant l'exercice qui suit :

Exercice 1.1: Exécutez les statistiques sur un fichier différent. Utilisez le questionnaire du tableau IV sur la production situé dans le fichier **C-Q4.SAV**.

Indices:

- a. Faites que **C-Q4.SAV** soit votre fichier de travail
- b. Utilisez la commande **Descriptives** pour les variables continues, et **Frequencies** pour les variables catégoriques
- c. **Prod** est une variable catégorique
- d. Les quantités (**p1b, p2b, ...**) sont de variables continues
- e. Les unités (**p1a, p2a, ...**) sont catégoriques
- f. **p4** (mois où les réserves se sont épuisées l'année d'avant) & **p6** (mois où les réserves seront épuisées cette année) sont catégoriques

Voici un échantillon des résultats produits en ayant exécuté les commandes *Frequencies* et *Descriptives*:

PRODUCT

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid cotton	83	4.9	4.9	4.9
peanuts	144	8.5	8.5	13.4
rough rice	155	9.2	9.2	22.6
bananas	50	3.0	3.0	25.5
sweet potato	12	.7	.7	26.2
cashew liquor	24	1.4	1.4	27.6
sugar cane liquor	11	.6	.6	28.3
dried cashew	2	.1	.1	28.4
sugar cane	13	.8	.8	29.2
cashew nut	130	7.7	7.7	36.9
coconut	45	2.7	2.7	39.5
beans	279	16.5	16.5	56.0
manteiga beans	7	.4	.4	56.4
sunflower	5	.3	.3	56.7
oranges	13	.8	.8	57.5
cashew fruit	44	2.6	2.6	60.1
manioc	338	20.0	20.0	80.0
sorghum	124	7.3	7.3	87.4
maize	192	11.3	11.3	98.7
"ossura"	5	.3	.3	99.0
tobacco	4	.2	.2	99.2
tomato	13	.8	.8	100.0
Total	1693	100.0	100.0	

Descriptive Statistics



	N	Minimum	Maximum	Mean	Std. Deviation
PROD THIS YR - # OF UNITS	1670	.0	5000.0	26.353	163.436
PROD NORMAL YR - # OF UNITS	1598	.5	5000.0	22.815	159.510
STOCK ENTERING HARVEST - # OF UNITS	173	.0	30.0	2.523	4.575
STORED FOR CONS THIS YR - UNIT	1227	1	9	2.25	1.82
STOCK FOR SEED - # OF UNITS	869	.0	100.0	4.938	6.876
Valid N (listwise)	147				

Statistiques de base avec des variables multiples






La commande Crosstabs

La commande **Crosstabs** produit un tableau affichant la distribution des échantillons selon leurs valeurs, pour deux ou plusieurs variables catégoriques.

Regardez le questionnaire sur les membres du ménage dans l'annexe, Tableau IA. Une chose que vous aimeriez savoir est comment le sexe des répondants varie selon leurs relations avec le chef du ménage. Vous pouvez savoir combien de femmes sont chefs de ménage par exemple. La commande **Crosstabs** peut produire ces statistiques. Assurez-vous que le fichier **C-Q1A.SAV** sur les membres du ménage est votre fichier de travail :

1. Ouvrez le fichier en utilisant la barre d'outil du *Data Editor*
2. Sélectionnez **c-q1a.sav**
3. Cliquez sur  pour placer la commande dans l'éditeur de syntaxe et affichez la fenêtre de syntaxe
4. Placez le curseur n'importe où sur la ligne de la commande "GET" et cliquez sur  sur la barre à outil.

Maintenant pour exécuter la commande **Crosstabs** :

1. Sélectionnez **Analyze...Descriptive statistics** du menu
2. Choisissez **Crosstabs...**
3. Sélectionnez **ca2** (relation avec le chef du ménage) de la liste de gauche et cliquez sur  à côté de **Row(s)**:
4. Choisissez **ca4** (sexe) de la liste de gauche et cliquez sur  à côté de **Column(s)**:
5. Cliquez sur 
6. Dans la section **Counts**, cliquez sur la boîte à côté de **Observed** pour placer un ✓, s'il n'en existe pas encore
7. Dans la section **Percentages** cliquez sur les boîtes **Row** and **Column**
8. Cliquez sur 
9. Cliquez sur 
10. Exécutez la commande dans le *Syntax Editor*.




La fenêtre **Crosstabs:Cell Display** vous permet de spécifier quelles statistiques vous voulez afficher dans chacune des cellules du tableau — dans ce cas ci, nous voulions les chiffres (*counts*), les pourcentages des colonnes et pour les rangées (à travers une rangée et une colonne, la somme des pourcentages arrivent à 100%). Le tableau produit par la commande affiche qu'il y a 21 têtes de ménage qui sont des femmes, et que 6.1% de tous les chefs de ménages sont des femmes.

Les Moyennes

La commande **Compare Means** est similaire à la commande **Crosstabs**, mais pour les variables continues. Les statistiques ainsi obtenues montrent comment la moyenne pour une variable continue ou d'autre type de statistiques diffère, en termes de leurs valeurs, des variables catégoriques. Une autre comparaison pour mieux élucider les différences entre **Crosstabs** et **Compare Means** est que **Crosstabs** est un moyen d'obtenir des résultats de type **Frequencies** par catégorie de variables, tandis que **Compare Means** est un moyen d'obtenir des résultats de type **Descriptives** selon les catégories de variables.

Supposons que nous voulions savoir comment l'âge des répondants variait selon leurs relations avec chef du ménage. Si nous faisons cette analyse avec **Crosstabs**, nous obtiendrons une pléthore de cellules du tableau pour les différents groupes d'âge, ce qui serait dans un format pas approprié. Au lieu de cela, nous allons

utiliser la commande **Compare Means**.

1. Sélectionnez **Compare Means** du menu **Analyze**
2. Choisissez **Means...**
3. Sélectionnez **ca3** (âge) et cliquez sur  à côté de **Dependent List**:
4. Sélectionnez **ca2** (relation au chef) et cliquez sur  à côté de **Independent List**:
5. Cliquez sur 
6. Exécutez la commande dans le *Syntax Editor*

Cette commande calculera les moyennes de la variable dépendante (âge), qui devrait être normalement une variable continue. Les moyennes pour les variables indépendantes seront calculées séparément, variables qui sont normalement catégoriques, telle que la variable représentant la relation avec chef du ménage.

Selon les résultats, vous verrez que l'âge moyen des chefs de ménage (*average age of heads of households*) est de 41.5 ans (*years*) tandis que l'âge moyen des leurs épouses est de 33.2 ans.

Transformations des données

Après avoir examiné les résultats des statistiques descriptives, souvent les utilisateurs veulent faire certains changements aux données. La transformation des données est une opération qui prend les variables existantes soit pour en changer les valeurs ou pour créer de nouvelles variables. L'exemple qui suit démontre comment convertir une variable continue en variable catégorique.


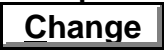
Bien que les résultats provenant de la commande **Means** soient intéressants, il est peut-être plus utile de connaître la distribution des groupes d'âges en créant des catégories. Comme la variable **ca3** est continue, il faut modifier les données par transformation. Supposons que nous voulons connaître la distribution en quatre catégories : 0-10 ans, 11-19 ans, 20-60 ans, et 60 ans et plus.

La commande Recode

Pour catégoriser cette variable, vous utiliserez la commande **Recode**. Catégoriser une variable continue rend les informations détaillées plus générales. Si vous voulez garder les informations détaillées d'origine aussi bien que la nouvelle information plus générale, vous devez utiliser la commande pour re-coder la variable en une nouvelle variable. Si vous re-codez une variable en termes de la variable elle-même, vous perdez l'information sur la variable d'origine.

Dans ce cas particulier, si vous utilisez la commande **Recode Into Same Variable** pour changer **ca3**, cette variable prendra des nouvelles valeurs catégoriques fixées par la commande, et les valeurs originales seront perdues. Comme nous voulons préserver les valeurs originales et créer une nouvelle variable, vous choisirez la commande **Recode Into A Different Variable**.

Re-codons la variable **ca3** en créant une nouvelle variable pour les groupes d'âges **age_gp**.

1. Choisissez **Recode** du menu **Transform**
2. Sélectionnez **Into Different Variables...**
3. Sélectionnez **ca3** de la liste à gauche
4. Cliquez sur  à côté de la boîte **Input Variable -> Output Variable**:
ca3 se déplacera dans la boîte Input Variable->Output Variable: et le nom de la boîte changera à Numeric Variable -> Output Variable
5. Cliquez une fois dans la boîte vide à côté de **Name**: dans la section **Output Variable** se qui positionnera le curseur
6. Tapez **age_gp** dans la boîte
7. Cliquez une fois dans la boîte vide près de **Label**: dans la section **Output Variable**
8. Tapez **Age Group**
9. Cliquez sur  pour que ces changements prennent effet

10. Cliquez sur **Old and New Values...**
Une nouvelle fenêtre s'affichera **Recode into Different Variables: Old and New Values**
11. Dans la section Old Value cliquez sur le cercle à côté de **Range:** **through**
Votre curseur devrait se situer dans la première boîte
12. Tapez **0** dans la première boîte
13. Appuyez sur <Tab> et tapez **10** dans la deuxième boîte
14. Appuyez sur <Tab> deux fois
*Votre curseur devrait se trouver dans la boîte à côté de **Value:** dans la section **New Value** OU vous pouvez appuyez sur la touche "Alt" et gardant votre doigt sur la touche, appuyez également sur la touche "l" ce qui vous amènera directement à la boîte "New Value"*
15. Tapez **1** pour le premier groupe d'âge
16. Cliquez une fois sur **Add**
17. Cliquez sur la première boîte après **Range:** et répétez les étapes 11 à 16 pour re-coder les âges **11 à 19** au groupe 2 et **20 à 60** au groupe 3
18. Pour re-coder les âges **61** et plus au groupe **4**, cliquez sur le cercle près de **Range:**
through highest
19. Entrez **61** dans la boîte et répétez les étapes 14 à 16 avec 4 comme valeur
20. Cliquez sur **Continue**
21. Cliquez sur **Paste**
22. Sélectionnez le texte qui suite dans le Syntax Editor

```

RECODE
  ca3
  (0 thru 10=1) (11 thru 19=2) (20 thru 60=3) (61 thru Highest=4) INTO
  age_gp .
VARIABLE LABELS age_gp 'age group'.
EXECUTE .

```
23. Exécutez la commande

La commande **Recode** change les valeurs de **age_gp** aux codes que nous voulons utiliser — 1,2,3, et 4. Allons voir ces transformations maintenant dans l'éditeur des données.

Pour passer à la fenêtre du **Data Editor** (*nous employons ici une nouvelle méthode*):

1. Cliquez sur **Window** des menus et sélectionnez **c:\sample\q1a.sav**.
2. Faites défiler les données en utilisant les barres de déplacements

Par défaut, le format standard qu'emploie SPSS pour afficher une variable numérique est avec deux décimales, ce qui n'est pas approprié pour toutes variables ayant une valeur entière. Pour changer le format de la variable **age_gp** empruntez la fenêtre Variable View.

1. Choisissez la fenêtre des données si ce n'est pas déjà le cas
2. Sélectionnez maintenant la fenêtre Variable View depuis la barre de gauche en bas
3. Cliquez une fois sur la boîte pour que la variable **age_gp** ait sur la ligne 12 dans la colonne pour Type. "Numeric" apparaîtra dans la boîte
4. Cliquez sur la boîte grise qui se trouve à la droite de la petite boîte et la boîte Variable Type pour que la **age_gp** apparaissent.
5. Dans la boîte à côté de **Width:** tapez 1
6. Dans celle à côté de **Decimal Places:** tapez 0
7. Si le cercle près de **Numeric** n'a pas été sélectionné, cliquez dessus
8. Cliquez sur **Continue**

*Vous n'avez pas les options de coller la commande par **Paste**. Les changements se font automatiquement.*

Vous auriez pu également changer la taille de la variable dans les colonnes “**Width**” et “**Decimals**”. Vous devez tout d’abord changer les décimales à 0 et puis seulement après la largeur à 1.

Ces changements indiquent au logiciel SPSS pour Windows d’afficher **age_gp** avec une taille 1 et sans décimales. Lorsque vous re-codez en une nouvelle variable, celle-ci n’a pas d’étiquettes de valeurs (*Value Labels*). Les sorties statistiques de SPSS incluent toujours les noms des variables d’analyse, mais parfois ce nom explique très peu. Vu que les noms sont limités à 8 caractères, ils ne sont pas habituellement très informatif et ne nous permet pas de bien comprendre la question du questionnaire (e.g. la variable **ca1** représente les réponses oui ou non, relatives au travail à la ferme). Le nom ne nous dit pas souvent non plus les valeurs individuelles des variables catégoriques (e.g. **ca4** est la variable pour le sexe, 1 pour les hommes et 2 pour les femmes). Pour rendre les données et la présentation des résultats plus compréhensible, nous ajoutons les étiquettes des variables (*Variable Labels*) et étiquettes des valeurs (*Value Labels*). Le meilleur moment pour rajouter ces étiquettes est immédiatement après la création d’une nouvelle variable, car si vous attendez, il est possible de les oublier. La commande **Recode** vous permet de le faire en vous permettant de rajouter des **Variable Labels** :

1. Vous devez être dans la fenêtre du *Variable viewer*
2. Dans la colonne **Label** pour la variable **age_gp**, vous devez pouvoir voir le texte “Age Group” car vous l’aviez ajouter dans la dernière commande
3. S’il n’y a pas de texte, ajoutez “Age Group”
4. Allez à la boîte pour **age_gp** dans la colonne de Values, où il est écrit “None”
5. Cliquez sur la boîte grise une fois pour afficher la fenêtre des étiquettes de valeurs
6. Tapez **1** dans la boîte *Value*, et appuyez sur **<Tab>** pour vous déplacer à la boîte de *Value Labels* et tapez ensuite **0 to 10**
7. Cliquez sur **Add**
Vous avez peut-être remarqué qu’il y avait t deux autres options possibles **Remove** pour enlever les valeurs et ses étiquettes, et **Change** pour modifier les étiquettes des valeurs
8. Répétez les étapes 6 et 7, suivant les informations ci-bas :

Valeur:	Étiquette :
2	11 to 19
3	20 to 60
4	61 and older
9. Cliquez sur **OK**
10. Allez à la fenêtre Data View et vous verrez que **age_gp** est maintenant affiché en un seul chiffre
11. Sélectionnez **Variables...** du menu **Utilities**
12. Cliquez sur **age_gp** pour vérifier les changements que vous venez de faire
13. Cliquez sur **Close** lorsque vous aurez terminé

Cette nouvelle variable ne fait toujours pas partie du fichier de données sur le disque dur. Vous devez sauvegarder votre fichier pour inclure vos changements. Un conseil est de toujours sauvegarder le fichier sous un autre nom pour ne pas éliminer le travail fait dans le fichier précédent. Pour cette raison, nous allons utiliser la commande **Save As** du menu **File** avec comme nouveau nom de fichier **Q1A-AGE.SAV**.

1. Assurez-vous que la fenêtre du Data Editor est active
2. Du menu **File** sélectionnez **Save As...**
*Le curseur devrait se retrouver dans la ligne File name:
Save as type: SPSS (*.SAV)*
3. Tapez **q1a-age** (l’extension .sav sera ajouté automatiquement)

4. Collez et exécutez la commande

À chaque fois que vous ouvrirez le fichier Q1A-AGE.SAV, la variable **age_gp** apparaîtra.

Vous pourrez analyser cette nouvelle variable catégorique en utilisant la commande **Crosstabs** pour déterminer combien de gens de chaque groupe sont chefs de ménage, des époux ou bien des enfants.

1. Utilisez la commande **Analyze...Descriptive Statistics... Crosstabs...** par les menus
2. Prenez **age_gp** pour Rows et **ca2** pour Columns
3. Vérifiez les sélections des choix du bas pour les Cells, car nous voulons les pourcentages pour les rangées et les colonnes.
4. Collez et exécutez la commande

De cette analyse, vous verrez que 12% des chefs de ménage ont 61 ans et plus. Aussi, que parmi les gens de 61 ans et plus, 83.7% sont chefs de ménage.

Comparez cette information ci-dessus avec les statistiques venant de **Compare Means** sur la variable **ca3**. Pour ce faire, nous vous rappelons qu'il existe plusieurs façons de naviguer entre les fenêtres de SPSS.

Pour afficher le Viewer:

1. Du menu **Window**, sélectionnez Session1 - SPSS.Viewer
2. Défilez en arrière à travers la fenêtre en utilisant les barres de déplacements.
3. Trouvez le tableau *Crosstabs* et comparez avec le tableau *Compare Means*.

Pour afficher la fenêtre de l'éditeur de syntaxe (le *Syntax Editor*) :

1. Du menu **Window** sélectionnez Session1 - SPSS Syntax Editor.
2. Défilez en haut et en bas de la fenêtre

Pour afficher le *Data Editor* :

1. Du menu **Window** sélectionnez **q1a** - SPSS Data Editor.
2. Défilez à nouveau pour observer les données

Veillez noter qu'il est aussi possible de choisir la fenêtre SPSS voulue en cliquant sur les boutons apparaissant sur la barre d'outil de Windows.

Appliquez vos connaissances sur les transformations et les statistiques de base avec le prochain exercice :

Exercice 1.2: En utilisant le questionnaire et données sur les ménages (disponible en annexe), trouvez le nombre de ménages dans chacun des districts qui ont 1-4, 5-7, et plus de 7 personnes par ménage. Essayez-vous en créant un tableau.

- a. Indices Utilisez le fichier C-HH.SAV
- b. Re-codez **h1** pour **hhsiz** en utilisant les groupes suivants : (1 thru 4) (5 thru 7) (8 thru Highest)
- c. Ajoutez des étiquettes pour les variables ainsi que pour les valeurs
- d. Exécutez la commande **Crosstabs** sur cette nouvelle variable par **district**.

Household size * DISTRICT Crosstabulation

			DISTRICT			Total
			MONAPO	RIBAUE	ANGOCHÉ	
Household size	1.00	Count	65	48	74	187
		% within Household size	34.8%	25.7%	39.6%	100.0%
		% within DISTRICT	60.7%	40.3%	64.3%	54.8%
		% of Total	19.1%	14.1%	21.7%	54.8%
	2.00	Count	39	56	36	131
		% within Household size	29.8%	42.7%	27.5%	100.0%
		% within DISTRICT	36.4%	47.1%	31.3%	38.4%
		% of Total	11.4%	16.4%	10.6%	38.4%
	3.00	Count	3	15	5	23
		% within Household size	13.0%	65.2%	21.7%	100.0%
		% within DISTRICT	2.8%	12.6%	4.3%	6.7%
		% of Total	.9%	4.4%	1.5%	6.7%
Total	Count	107	119	115	341	
	% within Household size	31.4%	34.9%	33.7%	100.0%	
	% within DISTRICT	100.0%	100.0%	100.0%	100.0%	
	% of Total	31.4%	34.9%	33.7%	100.0%	

En regardant les résultats, vous verrez pour Monapo par exemple, 34,8% du groupe de 1 à 4 membres (group 1) se retrouvent à l'intérieur de Monapo et que 60,7% de tous les ménages à Monapo ont 1 à 4 membres dans un ménage.

Avant de sortir du logiciel SPSS pour Windows, sauvegardez vos résultats du Viewer. Ces sorties contiennent non seulement les résultats mais aussi les commandes. Il est utile de sauvegarder ces sorties pour une étude ultérieure, afin de les imprimer ou les inclure dans un rapport.

1. Affichez le Viewer à l'écran
2. Du menu **File** sélectionnez **Save As...**
3. Entrez un nouveau nom de fichier **session1**
l'extension .spo sera ajoutée automatiquement
4. Cliquez sur **Save**

Pour sortir de SPSS pour Windows :

1. Du menu **File** sélectionnez **Exit SPSS**
Une fenêtre vous demandera de sauvegarder le contenu C:\sample\c-hh.sav
2. Cliquez sur **No**
Une fenêtre vous demandera de sauvegarder le contenu du Syntax Editor à Syntax1
3. Cliquez sur **Save** et donnez-lui un nom de fichier comme Module1.sps
Le logiciel SPSS pour Windows fermera par la suite.

Guide de Formation au SPSS pour Windows
MODULE 2 - Manipulation des bases de données:
Visualisation des tableaux, et agrégation

Pour certains types d'analyses, les données nécessitent une restructuration de niveau. Les données des quatre questionnaires (le ménage, les membres du ménage, la production et les ventes) ont été transcrites dans quatre fichiers séparés car elles représentent différents niveaux de données. Les données sur les ménages sont les plus générales et ont le plus haut niveau,-- c'est-à-dire un cas par ménage. Les trois autres fichiers contiennent des données plus détaillées, considéré comme des niveaux inférieurs, c'est-à-dire plusieurs cas possibles par ménage. Si vous n'êtes pas familier avec le concept de niveaux, lisez le document "Analyse de données d'enquêtes sur ordinateur : Organisation d'un fichier pour des données à niveaux multiples" par Chris Wolf, avant de poursuivre avec le module 2.

Les analyses faites dans le module 1 ont été faites à chaque niveau séparément, en utilisant uniquement les variables disponibles dans un même fichier. Cependant, autres types d'analyses demandent des combinaisons de données de plusieurs fichiers. Regardons un exemple.

Supposons que nous voulions créer un tableau de calories par adulte-équivalent produites par jour et provenant des principales cultures. De plus, nous voulons voir comment ceci varie par district et en quartile de calorie produite.

TABLEAU 1: Production alimentaire en calories
équivalent adulte par jour

Districts	Quartile des calories produites			
	1	2	3	4
Monapo				
Ribaue				
Angoche				

La manière dont les données sont présentées ci-haut ne peut répondre à la question posée, plusieurs transformations seront nécessaires pour produire ce tableau. Ceci est l'illustration d'une situation réelle que vous pourrez rencontrer en faisant des analyses de ce genre. Pour commencer, regardons les fichiers que nous avons et les variables dont nous aurons besoins :

- C-Q1A.SAV : Ce fichier contient des données sur les membres des ménages - signifiant que nous sommes au niveau membre-du-ménage. Les variables **ca3**(âge) et **ca4**(sexe), nous sont utiles car nous allons calculer le nombre de calories équivalent- adulte pour chaque ménage.
- C-Q4.SAV : Ce fichier contient des variables sur la production et se trouve au niveau de la production du ménage. Certaines variables dont nous aurons besoins sont :
 - a. **prod** - des codes sur la culture produite
 - b. **p1a** - des codes pour les unités pour lesquelles la production a été mesurée (sac de 100 kg, de 50 kg, etc.)
 - c. **p1b** - le nombre d'unités produites cette année.

Notez que les unités de production ne sont pas standardisées pour chaque culture. Par exemple, un sac de 100 kg, pèse 100 kg rempli de maïs mais beaucoup moins que 100 kg pour le manioc. Ainsi

nous avons besoin de *facteurs de conversion* pour nous permettre de convertir la production en une unité standard, soit le kilogramme.

- **CONVER.SAV** : Ce fichier est notre fichier d'appariement de facteurs de conversion (*table-lookup file*). Ce fichier a été créé spécialement pour aider à résoudre le problème de convertir les unités non-standards à des unités standards. Ces facteurs sont équivalents aux poids en kilogrammes pour chacun des produits observés. Par exemple, le facteur de conversion pour un sac de 50 kg de riz est ____; pour 50 kg de coton _____, tandis que 50 kg de manioc est _____. Les variables dans ce fichier sont :
 - a. **prod**: Code du produit.
 - b. **unit**: L'unité
 - c. **conver**: Le facteur de conversion, correspondant au nombre de kg pour chaque combinaison de **prod** et **unit**.

Le tableau ci-après montre un exemple du contenu du fichier CONVER.SAV, une boîte de 20 litres (**unit**=8) de riz (**prod**=7) pèse 19 kg; un sac de 50 kg (**unit**=24) de riz pèse 53 kg; une boîte de 20 litres d'haricots (**prod**=30) pèse 17 kg; et un sac de 50 kg de haricots pèse 47 kg.

prod (Produit)	unit (unité)	conver (facteur de conversion)
...
7	8	19
7	24	53
...
30	8	17
30	24	47
...

- **CALORIES.SAV** : Un second fichier d'appariement, créé pour convertir les kilogrammes d'aliments en calories. Ce fichier contient deux variables:
 - a. **prod** - le produit
 - b. **calories** - le nombre de calories par kilogramme de chaque produit alimentaire enregistré par l'enquête

Avec ces informations, nous pouvons maintenant penser aux étapes qu'il faut suivre pour arriver à produire le tableau que l'on cherche. Logiquement, il y en a trois :

1. Il faut savoir combien de calories chaque ménage a produit cette année. Ce qui peut se faire avec ces trois fichiers : sur la production C-Q4.SAV, et les deux fichiers d'appariement CONVER.SAV et CALORIES.SAV.
2. Il faut connaître combien d'équivalentadultes il y a dans chaque ménage. Nous pouvons partir du fichier sur les membres du ménage C-Q1A.SAV.
3. Il faut réunir les résultats des étapes 1 et 2 dans un seul fichier pour calculer les calories produites par équivalent- adulte par jour.

Première étape : Générer un fichier au niveau du ménage, avec le nombre de calories produites par ménage.

Pour ce faire, vous devez vous rappeler que :

1) toute la production est présentement en unités non-standards; alors il faut convertir la production en kilogramme

2) on veut savoir les calories produites par ménage mais pas par kilogramme, alors une fois la production convertie en kg, il faut les reconvertir en calories.


3) le contenu du fichier nous montre que nous avons les données pour les cultures produites par le ménage. Ce que nous voulons savoir est le nombre total de calories produites par le ménage, pas le total des calories pour chacun des produits. Une fois la production convertie en calories, nous devons faire la somme des calories pour chacun des ménages, des produits pour avoir le nombre total de calories par ménage.

Tenant compte de ces trois points ouvrez le fichier C-Q4.SAV.

1. Choisissez **File/Open/Data...**
2. Sélectionnez le nom du fichier c-q4.sav
3. Collez et exécutez la commande

Convertissons la production de ces sept cultures en kg. La procédure pour pouvoir combiner des données de deux fichiers est le fusionnement. Nous utiliserons le fichier de production C-Q4.SAV et notre fichier d'appariement que nous allons chercher pour la conversion CONVER.SAV. Nous allons créer un nouveau fichier où chaque cas contient à la fois les données du fichier de production et une variable pour le facteur de conversion pour combinaison produit-unité. Dans SPSS pour Windows, la commande utilisée est **Data/ Merge Files/ Add Variables**.



Le fichier intrant pour le fusionnement doit être trié par les variables clés, c'est-à-dire celles utilisées pour correspondre et appareiller les deux fichiers. Comme nous avons un facteur de conversion unique pour chaque combinaison de produit et d'unité, les variables "produit" et "unité" sont des variables clés. Le fichier CONVER.SAV est déjà trié par **prod** et **unit**. Nous devons donc trier le fichier de travail sur la production de la même façon; vous noterez que la variable s'appelle **p1a** et non **unit**. Il faut changer le nom pour que l'on puisse fusionner les deux fichiers.

1. Du menu **Data** sélectionnez **Sort Cases...**
2. Choisissez **prod** et appuyez sur 
3. Faites de même pour **p1a**
4. Collez et exécutez la commande

Les fichiers sont maintenant prêts pour être fusionnés. La commande **Merge Files** exige l'existence d'au moins deux fichiers. Dans notre cas, ces deux fichiers sont le fichier de travail et CONVER.SAV. Nous faisons un appariement fichier-tableau (*File - Table*) où le deuxième fichier est notre fichier de référence (*Lookup Table*). Le nouveau fichier créé par la commande **Merge Files** deviendra le nouveau fichier de travail, remplaçant le fichier de travail actuel.

1. Du menu **Data** sélectionnez **Merge Files**, et puis **Add Variables...**
2. Sélectionnez le fichier conver.sav
3. Cliquez sur 

Les variables utilisées pour jumeler les cas doivent avoir les mêmes noms. Nous devons sélectionner **p1a** du *New Working Data File* (nouveau fichier de travail) et le transférer à la boîte des *excluded variables* (variables exclues). Nous allons le renommer **unit** et nous pourrons les jumeler par la suite.


4. Sélectionnez **p1a** de la liste sous *New Working Data File:* et cliquez sur 
5. Cliquez sur 

Ceci vous permettra de renommer p1a en unit pour jumeler le fichier de conversion.

6. À côté de *New Name:* tapez **unit**

7. Cliquez sur **Continue**

Nous ne pouvons pas sélectionner les variables à fusionner tant que nous n'indiquons pas que nous voulons le faire par variables clés (*match cases on key variables*).



8. Cochez la boîte à côté de **Match cases on key variables in sorted files**
9. Cliquez sur le cercle à côté de **External file is keyed table**
10. Choisissez **prod** de la liste **Excluded Variables:**
11. Cliquez sur  près de **Key Variables:** (en bas, côté droit)
12. Répétez les étapes 10 et 11 pour la variable **unit**
13. Collez la commande
Un avis s'affichera vous informant que le fichier de données doit être classé. Comme nous l'avons fait...
14. Cliquez sur **OK**
Une fenêtre apparaîtra vous demandant si vous voulez sauvegarder le contenu "données" (data window). Ce n'est pas nécessaire ici, le nouveau fichier prendra sa place, alors...
15. Cliquez sur **NO**
16. Sélectionnez et exécutez la commande. Assurez-vous d'inclure EXECUTE.

La commande dit à SPSS de fusionner le fichier de travail (actif dans le Data Editor) et le tableau de consultation du fichier CONVER.SAV pour ajouter une variable *unité* au fichier de travail. Comme il est nécessaire que les variables clés aient le même nom, nous avons changé le nom **p1a** (la variable unité pour le fichier de travail) à **unit (p1a** restera **p1a** dans le fichier c-q4.sav).

Les variables clés sont nécessaires pour tout fusionnement lorsqu'un des deux fichiers est utilisé comme tableau de consultation. Les variables clés sont spécifiques pour chaque combinaison produit et unité possible représentant un différent facteur de conversion. Si nous avons utilisé uniquement **prod**, SPSS exigerait que chaque produit ait un seul facteur de conversion, avec une même valeur peut importe l'unité de mesure. Par exemple, on aurait un même facteur de conversion pour un sac de 100 kg de riz ou un contenant de 20 litres, ce qui serait erroné.

Le nouveau fichier de travail produit par le fusionnement contient maintenant la variable de conversion nécessaire, **conver**. Pour chaque combinaison produit-unité, cette variable est égale au nombre de kilogrammes de cette unité. Il est toujours important de vérifier si la commande Merge a été exécutée correctement. Retournez au Data Editor et regardez divers cas pour s'assurer que les facteurs de conversion sont bien assortis aux produits. Par exemple, un contenant de 20 litres rempli de maïs en grain pèse 18 kilogrammes, ainsi, vérifier que lorsque PROD=47 et UNIT=8, que CONVER=18.

Nous pouvons maintenant calculer le poids total en multipliant le nombre d'unités (**p1b**) par ce facteur de conversion.

1. Du menu **Transform** sélectionnez **Compute...**
2. Sous **Target Variable:** entrez **qprod_tt** (pour quantité produite en kg)
3. Cliquez sur **Type & Label** pour rajouter une étiquette pour **qprod_tt** si vous le voulez et cliquez sur **Continue**.
4. De la liste à gauche de la fenêtre Compute Variable, sélectionnez **p1b** et cliquez sur  pour la mettre dans la fenêtre de droite, la boîte *numeric expression*.
5. Tapez * ou sélectionnez le bouton dans la fenêtre pour rajouter le signe de multiplication à côté de **p1b**.
6. De la liste de gauche, choisissez **conver** et cliquez sur 
7. Collez, sélectionnez et exécutez la commande.

Nous avons besoin maintenant de connaître le nombre de calories par kilogramme et par produit. Cette information est disponible dans le fichier CALORIES.SAV. Ce fichier contient deux variables: “produit” et “nombre de calories par kg”. La variable clé est “produit”. Pour ajouter la variable de conversion calorique au fichier de travail, il faut à nouveau fusionner avec un tableau de consultation (*keyed table lookup*). Cette fois, la variable clé est “produit”. Le tri du fichier a déjà été fait et n’est pas nécessaire.

1. Du menu **Data** sélectionnez **Merge Files** et puis **Add Variables...**
2. Choisissez le fichier calories.sav, **Open**
3. Cochez la boîte **Match cases...**
4. Cochez la boîte **External file is keyed table**
5. Mettez **prod** la boîte **Key Variables:**
6. Collez la commande
7. Dîtes non aux avis
8. Sélectionnez et exécutez la commande

Le nouveau fichier de travail produit par le fusionnement contient la variable “calorie” dont nous avons besoin. Le maïs en grain (PROD=47) doit avoir 3590 calories par kilogram pour la variable **calories**. Nous pouvons maintenant calculer le nombre total de calories produites.

1. Utilisez **Transform/Compute...**
2. Utilisez **kprod tt** comme **Target Variable:** (pour les calories produites)
3. Cliquez sur **Type & Label** pour rajouter une étiquette pour **kprod tt** si vous le désirez et cliquez sur **Continue**.
4. Cliquez sur la boîte **Numeric Expression** et entrez cette équation : **kprod tt * calories**
5. Collez, sélectionnez et exécutez la commande

Ceci nous donne un nouveau fichier de travail avec le nombre de calories produites par produit pour tous les ménages. Nous sommes uniquement intéressés aux 7 produits alimentaires de base suivant : corn (prod=47), nhemba bean (prod=30), manteiga bean (prod=31), manioc (prod=41), rice (prod=6), sorghum (prod=44), et peanuts (prod=5).

Vous pouvez voir les codes pour la variable **prod** dans le questionnaire. Comme nous nous intéressons seulement à ces produits, nous pouvons filtrer les données pour avoir seulement l’information que nous voulons pour l’analyse. Pour ce faire, nous utiliserons la commande **Select Cases**. Cette commande choisit une section des données selon les critères demandés. **Select Cases** peut soit filtrer les données non-choisies soit les éliminer complètement. Si vous les enlevez, vous pouvez toujours retourner au fichier original. Si vous filtrez les données (ce que nous allons faire ci-bas - méthode plus sécuritaire) vous pouvez toujours enlever le filtre des données à tout moment et réactiver toutes les données du fichier de travail.

1. De la fenêtre du Data Editor, sélectionnez **Data/Select Cases**
2. Sélectionnez le cercle à côté de **If condition is satisfied**
3. Cliquez sur **If...** sous **If condition is satisfied**
4. Cliquez dans la boîte, à la droite de **▶**, et **non** sur le bouton lui-même
5. Entrez le texte qui suit:

$$\text{PROD} = 47 \mid \text{PROD} = 30 \mid \text{PROD} = 31 \mid \text{PROD} = 41 \mid \text{PROD} = 6 \mid \text{PROD} = 44 \mid \text{PROD} = 5$$

*LE symbol “|” représente ou. Nous indiquons à SPSS de sélectionner les données **prod** égale à 47 ou 30 ou 31, etc.*

6. Cliquez sur **Continue**
7. Sélectionnez le cercle à côté de **Filtered** (pour que les données des autres produits ne soient pas inclus de façon temporaire dans les analyses)

8. Collez la commande
9. Sélectionnez le texte dans le Syntax Editor de ligne qui commence par USE ALL, jusqu'à la ligne EXECUTE, et exécutez la commande.

Seul les données dont les codes pour les produits choisis seront désormais utilisées. Ce sous-groupe de données restera filtré jusqu'au moment où un autre fichier est ouvert ou que le filtre soit enlevé.

Nous voulons savoir maintenant combien de calories ont été produites par ménage pour l'ensemble des aliments de base. Pour ce faire, nous allons faire la somme par ménage, des valeurs de **kprod_tt** pour toutes les cultures que les ménages produisent. En outre, nous allons créer un nouveau fichier au niveau du ménage à partir de ce fichier de travail au niveau de la production avec une observation par ménage. SPSS utilise le terme "AGGREGATE" pour comprimer les observations d'un niveau à l'autre. Nous allons faire la somme des observations produit-ménage pour une seule observation pour le ménage.

Pour créer ce fichier, nous utiliserons la procédure d'**Agrégation**. Elle va créer un nouveau fichier de travail avec une observation par ménage en faisant la somme de **kprod_tt** pour tous les produits pour chaque ménage. Elle prend toujours le fichier de travail comme celui qui sera agrégé. Nous avons déjà le fichier ouvert alors commençons.

1. Du menu **Data** sélectionnez **Aggregate...**
2. Sélectionnez **district**, **vil**, et **hh**, respectivement, pour les **Break Variable(s)**:
3. Sélectionnez **kprod_tt** comme la **Aggregate Variable(s)**:
4. Cliquez sur **Name & Label...**
5. Changez le nom par défaut de **kprod__1** à **kprod_tt**
6. Entrez l'étiquette suivante : **Calories Produced in Staple Foods**
7. Cliquez sur **Continue**
8. Cliquez sur **Function...**
9. Sélectionnez **Sum of values** et cliquez sur **Continue**
10. Sélectionnez **Replace working data file**, **OK**
11. Collez la commande
12. Cliquez sur **No** pour ne pas sauvegarder les nouvelles données
13. Exécutez la commande.

Si nous avions sélectionné **Create new data file** à la place de **Replace working data file**, le nouveau fichier agrégé aurait été sauvegardé sur le disque dur, et ne serait pas devenu le fichier de travail. Il aurait fallu ouvrir le fichier pour accéder aux données.

Les critères d'agrégation sous **Break Variable(s)** identifient les variables qui seront utilisées pour combiner les observations dans le nouveau fichier d'agrégation. Toutes les observations du fichier original qui ont des valeurs identiques seront combinées en une observation unique dans le fichier agrégé. Nous voulons que le fichier agrégé ait une observation par ménage alors nous utilisons les variables qui identifient le ménage de cette enquête — **district**, **vil**, et **hh**.

Les variables agrégées **Aggregate Variable(s)** ont créé la variable **kprod_tt**, que nous calculons en faisant la somme des **kprod_tt**, à travers l'ensemble des observations par ménage. Les seules variables qui apparaissent dans un fichier agrégé sont les critères d'agrégation et les nouvelles variables créées par l'agrégation (e.g. **kprod_tt**).

Le nouveau fichier de travail contient maintenant ce que nous voulions, le nombre total de calories produites par ménage. Faites des statistiques descriptives sur **kprod_tt**, vous devriez trouver que le nombre moyen de calories produites par ménage par année est de 4,483,964.7.

Sauvegardez votre fichier en utilisant la commande **Save As...**

1. Activez la fenêtre Data Editor
2. Utilisez **Save As...** du menu **File**
3. Nommez le fichier **hh-file1**
4. Collez et exécutez la commande

Deuxième étape : Générer un fichier au niveau des ménages qui contient le nombre d'équivalent-adultes par ménage.

Les données nécessaires pour calculer les équivalent-adultes par ménage se trouvent dans le fichier C-Q1A.SAV.

1. Cliquez sur le bouton pour ouvrir un fichier de la barre à outil de SPSS dans la fenêtre de l'éditeur de données
2. Sélectionnez le fichier **c-q1a.sav**
3. Collez et exécutez la commande

Les règles que nous emprunterons pour calculer les équivalent-adultes sont :

Hommes, 10 ans et plus	= 1.0
Femmes, 10 à 19 ans	= 0.84
Femmes, 20 ans et plus	= 0.72
Enfants, de moins de 10 ans	= 0.60

Ces chiffres veulent dire qu'en moyenne, les femmes âgées de 10 à 19 ans nécessitent seulement 84% des calories qu'un homme de plus de 10 ans, les enfants 60%, etc. Ce qui veut dire par exemple, qu'un enfant compte pour 0.60 équivalent-adulte. Pour chaque observation, dans le fichier des membres du ménage, nous avons besoin de voir leur sexe **ca4**, et leur âge **ca3**, pour calculer leur équivalent-adulte.

Nous allons créer une nouvelle variable **ae** par l'intermédiaire de la commande **Compute.../If...**

1. Du menu **Transform** sélectionnez **Compute...**
2. Pour la **Target Variable**: entrez **ae**
3. Sélectionnez la boîte **Type & Label** et entrez Adulte- équivalent pour Label. Cliquez sur **Continue**
3. Dans la boîte **Numeric Expression**: entrez **1**
4. Cliquez sur **If...**
5. Sélectionnez le bouton pour **Include if case satisfies condition**:
6. Entrez **ca4 = 1 & ca3 >= 10**
7. Cliquez sur **Continue**
8. Collez la commande sans l'exécuter
9. Répétez les étapes 1 et 3 à 8 en remplaçant les informations précédentes par ce qui suit : vous n'êtes pas obligé d'utiliser le système de menu. Une fois que des commandes sont collées à la fenêtre de syntaxe, il devient plus facile et rapide de copier et coller la même commande à l'intérieur de la fenêtre de syntaxe. Mais n'oubliez pas de changer les noms de variables. Pour ceux qui ne peuvent pas le faire encore, n'hésitez pas à suivre les instructions ci-haut.

Numeric Expression	If... Statement
0,84	ca4 = 2 & ca3 >= 10 & ca3 <= 19
0,72	ca4 = 2 & ca3 >= 20
0,6	ca3 < 10

10. Sélectionnez toutes les commandes `IF` et exécutez-les.

Votre syntaxe devrait ressembler à ceci :

```
IF (ca4 = 1 & ca3 >= 10) AE = 1 .  
IF (ca4 = 2 & ca3 >= 10 & ca3 <= 19) AE = 0.84 .  
IF (ca4 = 2 & ca3 >= 20) AE = 0.72 .  
IF (ca3 < 10) AE = 0.60 .  
VARIABLE LABELS AE 'ADULT- EQUIVALENT' .  
EXECUTE .
```

Pour vérifier que la nouvelle variable d'équivalent- adulte a été calculée, faites une analyse de fréquences comme vous l'aviez fait au module 1.

1. Utilisez la commande **Analyze/Descriptive Statistics/Frequencies...**
2. Choisissez **ae**
3. Collez et exécutez

Vous devriez voir 1524 observations. Normalement, on devrait voir 4 valeurs représentées dans le tableau —1, .72, .84, et .60— avec aucune observation manquante. Vous pouvez voir qu'il y a neuf observations manquantes. Ceci nous dit que le fichier a des données manquantes sur le sexe et l'âge de neuf individus. C'est quelque chose qui aurait dû être identifiée par un processus de nettoyage des données. À ce moment, un chercheur devrait consulter le questionnaire et corriger ces erreurs. Comme nous ne pouvons le faire, nous utiliserons une méthode alternative.

Si nous laissons ces valeurs comme manquantes nous aurons des problèmes d'analyse et les résultats seront faussés. On ne veut pas éliminer ces ménages comme ils contiennent d'autres informations et d'autres variables. Par contre, nous savons que les valeurs de la variable "équivalent- adulte" s'étend de 0.6 à 1.0 et nous pouvons calculer une moyenne par les statistiques descriptives -- Pour ce faire, faites :

1. **Analyze/Descriptive Statistics/Descriptives...**
2. Variable **ae**
3. Collez et exécutez la commande

Vous devriez obtenir une valeur de .79 avec un écart-type de .17. Nous allons supposer que les individus avec les données manquantes sont de la moyenne et nous allons les assigner la valeur moyenne. Il n'est pas très fortement recommandé de modifier les données ainsi, le but de cet exercice ici est simplement pour permettre aux participants de manipuler les outils de SPSS ainsi que les données.

1. **Transform/Recode/Into Same Variables...**
2. Déplacez **ae** à Variables:
3. Cliquez sur **Old and New Values...**
4. Sélectionnez System-missing
5. Choisissez **Value:** dans la section **New Value** et entrez **.79** dans la boîte
6. Cliquez sur **Add**
7. **Continue**
8. Collez, sélectionnez et exécutez

Maintenant il faut calculer le nombre d'équivalent- adulte par ménage. Le fichier ainsi créé se situe au niveau des membres du ménage mais les valeurs dont nous avons besoins sont du niveau des ménages. Nous pouvons utiliser la procédure d'**Agrégation** pour aller d'un niveau à l'autre. La

nouvelle variable **ae_tt** sera calculée en faisant la somme des **ae** à travers tous les membres du ménage.

1. De menu **Data** sélectionnez **Aggregate...**
2. Déplacez **district**, **vil**, et **hh** aux **Break Variable(s)**:
3. Déplacez **ae** au **Aggregate Variable(s)**:
4. Cliquez sur **Name & Label...**
5. Dans la boîte **Name**: entrez **ae_tt**
6. Pour l'étiquette **Label**: entrez **Adult Equivalents**
7. **Continue**
8. **Function...**
9. Choisissez **Sum of values**
10. **Continue**
11. Sélectionnez **Replace working data file**
12. Collez, refusez les avis et exécutez

Nous avons créé un fichier au niveau du ménage avec une observation par ménage. La variable **ae_tt** représente le nombre total d'équivalent- adultes par ménage. Pour vérifier que la variable a été créée, faites une statistique descriptive sur **ae_tt**, vous devriez trouver 3.49 comme moyenne.

1. **Analyze/Descriptive Statistics/Descriptives...**
2. Collez et exécutez

Ceci complète l'étape 2. Sauvegardez votre fichier de données sous le nom HH-FILE2.SAV.

1. Assurez-vous que la fenêtre du **Data Editor** est active
2. **File/Save As...**
3. Filename **hh-file2**
4. Collez et exécutez

Troisième étape : Il faut fusionner les deux fichiers que nous avons créés aux étapes 1 et 2, pour calculer les calories produites par équivalent- adulte.

Nous avons maintenant les deux fichiers, HH-FILE1.SAV avec les données de production et calories et le fichier HH-FILE2.SAV qui contient les données des équivalent- adultes, que nous voulons fusionner et obtenir un seul fichier. Pour ce faire nous allons prendre la procédure de fusionnement **Merge Files** mais cette fois ci, nous n'utiliserons pas de tableaux de consultation. Assurez-vous que votre fichier hh-file2.sav est toujours votre fichier de travail.

Nous avons noté plutôt que les variables clés étaient nécessaires pour tout fusionnement qui inclut des tableaux de consultation. Lorsque vous fusionnez deux fichiers du même niveau, il est toujours important de fusionner par variable clé. Ces variables identifient comment les observations seront combinées.

*Vous ne devrez jamais utiliser **Merge Files** sans **Key Variables** car vous n'aurez pas de garantie que SPSS fusionnera les bonnes observations.* La commande s'exécutera sans avis ou message d'erreur, mais les résultats seront erronés.

Notez que le fichier hh-file2.sav est toujours le fichier de travail

1. **Data/Merge Files/Add Variables...**
2. Utilisez le fichier hh-file1.sav comme **Read File**

3. **Open**
4. Sélectionnez Match cases on key variables...
5. Choisissez Both files provide cases
6. Key Variables: sont **district**, **vil**, et **hh** respectivement
7. Collez, refusez les avis, sélectionnez et exécutez

Le fusionnement des fichiers a créé un nouveau fichier avec les deux variables dont nous avons besoin pour calculer les calories produites par équivalent- adulte. Le nombre de calories produites (**kprod_tt**) par ménage pour l'année, divisé par le nombre total d'équivalent- adultes par ménage (**ae_tt**), celui-ci divisé par 365 jours par an, nous donne le nombre de calories produites par équivalent- adulte par jour (**kprod_ae**).

1. **Transform/Compute...**
2. Target Variable: **kprod_ae**
3. **Type & Label...**
4. Label: **Calories produced per adult equivalent per day**
5. **Continue**
6. Numeric Expression: entrez **kprod_tt/ae_tt/365**
7. Collez, sélectionnez et exécutez

Avant de produire le tableau voulu, il faut créer une autre variable qui identifie la production de calories par quartile dans lequel chaque ménage se situe dans son district. La procédure **Rank Cases** qui nous aidera à calculer les quartiles, permettra de calculer une nouvelle variable pour chaque observation nous informant du rang de celle-ci à l'intérieur d'un sous-groupe selon les valeurs des autres variables. Dans ce cas, nous voulons classer chaque ménage dans son district en termes de calories produites par équivalent- adulte **ae**. Spécifiquement, pour chacun des districts, nous voulons diviser les ménages en quatre groupes de la même taille (des quartiles), de la production la plus faible à la plus élevée. Une nouvelle variable avec les valeurs 1 à 4 nous indiquera auquel quartile appartiendra chacun des ménages.

1. **Transform/Rank Cases...**
2. Déplacez **kprod_ae** au Variable(s):
3. Déplacez **district** à By:
4. **Rank Types...**
5. Dé-sélectionnez Rank
6. Sélectionnez Ntiles:
7. **Continue**
8. Collez et exécutez
9. Notez la nouvelle variable dans le Viewer, qui devrait être **NKPROD_A**

La première chose que nous spécifions est la variable ayant les valeurs à utiliser pour classer les données, ici c'est **kprod_ae**. Ensuite nous avons besoin de la variable **By** pour spécifier les variables qui identifieront les groupes, ici **district**. **Rank Cases** a un certain nombre de méthodes pour classer les données. Nous avons pris une des plus faciles — /NTILES(4), les quartiles. Par cette commande, SPSS pour Windows va créer une nouvelle variable, lui assigner un nom et contiendra la classification des 4 sous-groupes.

Nous pouvons maintenant employer la procédure **Means** pour obtenir les valeurs qui entreront dans le tableau.

1. **Analyze/Compare Means/Means...**
2. Déplacez **kprod_ae** à **Dependent List**:

3. Déplacez **nkprod_a** à Independent list: layer 1 of 1
nkprod_a vient de la procédure Rank Cases
4.
5. Déplacez **district** à Independent List: layer 2 of 2
6. Collez et exécutez

Vous trouverez que la moyenne pour la population entière est de 4014.5183 et que celle pour le 2ème quartile pour Ribaue est de 2517.4551. Les résultats de cette procédure nous donne les valeurs nécessaires pour le tableau que nous voulons créer cependant, elles ne sont pas structurées exactement comme le tableau que nous avons indiqué au début du module. Dans le module 3, vous verrez comment produire les mêmes résultats sous une meilleure présentation sous forme de tableau.

Sauvegardez ce fichier de travail sous le nom HH-FILE3.SAV.

1. Activez la fenêtre du Data Editor
2. **File/Save As...**
3. Le nom du fichier est hh-file3
4. Collez et exécutez

Vous devez maintenant sauvegarder votre syntaxe pour une utilisation ultérieure.

1. Activez l'éditeur de Syntaxe
2. **File/Save As...**
3. Utilisez le nom **session2**
L'extension .sps s'ajoutera automatiquement

Ce fichier contient maintenant toutes les commandes SPSS. Il est toujours recommandé de sauvegarder votre syntaxe. Vous avez dû remarquer que vous avez effectué vos commandes SPSS en les collant tout premièrement à votre fenêtre de syntaxe au lieu de les exécuter immédiatement. Coller les commandes à la fenêtre de syntaxe est toujours recommandé en tout temps, afin d'avoir une documentation sur la méthode empruntée, sur votre travail, vos collègues peuvent voir cette documentation pour reproduire les mêmes données, etc.

Regardons si vous avez bien enregistré votre session sur SPSS (la syntaxe). Sortez de SPSS, et ouvrez à nouveau le logiciel - ensuite ouvrez la syntaxe SPSS et cherchez votre fichier .sps

1. **File/Exit SPSS**
SPSS vous demandera si vous voulez sauvegarder vos résultats; dites oui
2. Sauvegardez vos sorties dans le Viewer sous le nom **session2**

Démarrez SPSS pour Windows, et ouvrez votre fichier de syntaxe pour le module 2 :

1. **File/Open/Syntax...**
2. Sélectionnez session2.sps
3.

Vous pouvez re-exécuter toutes les commandes soit les modifier au besoin. Votre syntaxe SESSION2.SPS devrait ressembler à ceci. L'exception étant que les commentaires de la documentation ont été ajoutés à cet exemple en utilisant le "*" pour indiquer les commentaires.

*session 2 - Pour produire une table sur la production alimentaire en calories par adulte équivalent par jour.

*M. Beaver - 6/6/2002.

```
GET
FILE='C:\sample\C-Q4.SAV'.
```


*****Première étape *****.

*Préparation pour la fusion du fichier conver avec ce fichier - il faut trier par des variables identiques .

```
SORT CASES BY
  prod (A) p1a (A) .
MATCH FILES /FILE=*
/RENAME p1a=unit
/TABLE='C:\sample\CONVER.SAV'
/BY prod unit.
EXECUTE.
```

*Calcul de la quantité totale produite en kgs.

```
COMPUTE qprod_tt = conver * p1b .
VARIABLE LABELS qprod_tt 'COMPUTE qprod_tt = conver * p1b (COMPUTE)' .
EXECUTE .
```

*Fusion en valeurs équivalentes en calories.

```
MATCH FILES /FILE=*
/TABLE='C:\sample\CALORIES.SAV'
/BY prod.
EXECUTE.
```

*Calcul des calories totales produites.

```
COMPUTE kprod_tt = qprod_tt * calories .
VARIABLE LABELS kprod_tt 'COMPUTE kprod_tt = qprod_tt * calories (COMPUTE)' .
EXECUTE .
```

*Mise en place de filtre pour sélectionner seulement les denrées alimentaires.

```
USE ALL.
COMPUTE filter_$=(prod = 47 | prod = 30 or prod = 31 or prod = 41 or prod = 6
  or prod = 44 or prod = 5).
VARIABLE LABEL filter_$ 'prod = 47 | prod = 30 or prod = 31 or prod = 41 or'+
' prod = 6 or prod = 44 or prod = 5 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

*Vérifiez pour s'assurer que les produits choisis ont été sélectionnés.

```
FREQUENCIES
  VARIABLES=prod
/ORDER= ANALYSIS .
```

*Agrégation au niveau des ménages pour additionner les calories totales produites.

```
AGGREGATE
/OUTFILE=*
/BREAK=district vil hh
/kprod_tt 'Calories produced in staple foods' = SUM(kprod_tt).
```

*Vérifiez que la variable a été créée et la valeur est raisonnable.

```
DESCRIPTIVES
  VARIABLES=kprod_tt
/STATISTICS=MEAN STDDEV MIN MAX .
```

*Sauvegardez le fichier de niveau des ménages.

```
SAVE OUTFILE='C:\sample\hh-file1.sav'
/COMPRESSED.
```

**** Deuxième étape - Calcul le nombre des adultes équivalents.

```
GET
  FILE='C:\sample\C-Q1A.SAV'.
```

```
IF (ca4 = 1 & ca3 >= 10) ae = 1 .
```

```
VARIABLE LABELS ae 'Adult equivalent' .
IF (ca4 = 2 & ca3 >= 10 &ca3 <= 19) ae = 0.84 .
IF (ca4 = 2 & ca3 >= 20 ) ae = 0.72 .
IF ( ca3 < 10 ) ae = 0.6 .
EXECUTE .
```

*Vérifiez pour voir si le calcul a été fait correctement.
list ca4 ca3 ae / cases=20.
freq ae.

*Calculez la moyenne de la population totale..

```
DESCRIPTIVES
  VARIABLES=ae
  /STATISTICS=MEAN STDDEV MIN MAX .
```

*Remplacez **sysmis** avec la moyenne de la population totale.

```
RECODE
  ae (SYSMIS=.79) .
EXECUTE .
freq ae.
```

*Aggregation au niveau des ménages en additionnant les adultes équivalents.

```
AGGREGATE
  /OUTFILE=*
  /BREAK=district vil hh
  /ae_tt 'Adult Equivalents' = SUM(ae).
DESCRIPTIVES
  VARIABLES=ae_tt
  /STATISTICS=MEAN STDDEV MIN MAX .
```

```
SAVE OUTFILE='C:\sample\hh-file2.sav'
  /COMPRESSED.
```

**** Troisième étape - Joindre les caloriers des ménages aux adultes équivalents des ménages. ****.

```
MATCH FILES /FILE=*
  /FILE='C:\sample\hh-file1.sav'
  /BY district vil hh.
EXECUTE.
```

* Calculez les calories produites par adulte équivalent par jour.

```
COMPUTE kprod_ae = kprod_tt / ae_tt / 365 .
VARIABLE LABELS kprod_ae 'Calories produced per adult equivalent per day' .
EXECUTE .
```

```
RANK
  VARIABLES=kprod_ae (A) BY district /NTILES (4) /PRINT=YES
  /TIES=MEAN .
```

```
MEANS
  TABLES=kprod_ae BY nkprod_a BY district
  /CELLS MEAN COUNT STDDEV .
```

```
SAVE OUTFILE='C:\sample\hh-file3.sav'
  /COMPRESSED.
```

Exercice 2.1:

Produisez des résultats similaires en utilisant les calories retenues (la production moins les ventes) au lieu des calories produites. Vous aurez les calories retenues par équivalent- adulte par jour (pour la même quantité totale provenant des six aliments de base). Les résultats devront être ordonnés par district et par quartile de calories produites.

- Conseils :
- a. La procédure est similaire à celui que vous venez d'effectuer.
 - b. Les ventes viennent du fichier **c-q5.sav**.
 - c. Cherchez dans ce fichier la variable appropriée pour la quantité de la production qui a été vendue. Notez que les codes pour les produits sont les mêmes qu'avec le fichier **c-q4.sav**. N'oubliez pas de bien trier vos données.
 - d. Récupérer le syntaxe du module deux où nous avons généré le tableau précédent et vérifiez chacune des étapes pour les changements nécessaires : il y aura des changements pour les codes des produits, le nom des fichiers et des variables.
 - e. Calculer les calories vendues implique les mêmes étapes de base que celles à suivre pour calculer les calories produites. (Step 1)
 - f. Fusionnez ce nouveau fichier avec les calories vendues, avec le fichier des calories produites **hh-file3.sav**.
 - g. Retenez que seulement 256 ménages ont vendu des produits, mais 343 ménages ont produit et retenu des calories. Si la variable représentant les calories vendues est manquante, cela veut dire que le ménage n'a pas vendu d'aliments et ces valeurs devront être re-codées à 0.
 - h. Calculez : calories retenues = calories produites - calories vendues.
 - i. Ordonnez les données en quartiles.
 - j. Utilisez **Compare Means** pour montrer les calories retenues par **district** et **quartile**.
 - k. Sauvegardez vos données.
 - l. Il n'est pas nécessaire de sauvegarder votre syntaxe pour cet exercice.
 - m. Exécutez les commandes et regardez les résultats.

Voici un exemple des résultats que vous deviez produire :

Report

KRET_AE

NKRET_AE NTILES of KRET_AE by DISTRICT	DISTRICT DISTRICT	Mean	N	Std. Deviation
1	1 MONAPO	1148.045	27	409.61445
	2 RIBAUE	1232.803	29	350.22596
	3 ANGOCHE	912.7559	28	384.74681
	Total	1098.877	84	401.03778
2	1 MONAPO	2211.383	27	205.71992
	2 RIBAUE	2145.845	30	202.81580
	3 ANGOCHE	1698.510	29	168.49973
	Total	2015.575	86	297.99128
3	1 MONAPO	3314.857	28	477.12339
	2 RIBAUE	3126.358	30	329.89358
	3 ANGOCHE	2405.008	29	336.48560
	Total	2946.574	87	547.14537
4	1 MONAPO	7619.102	27	3557.135
	2 RIBAUE	5759.039	30	1649.584
	3 ANGOCHE	4954.763	29	2426.824
	Total	6071.803	86	2821.271
Total	1 MONAPO	3570.975	109	3032.696
	2 RIBAUE	3081.416	119	1902.739
	3 ANGOCHE	2506.498	115	1957.991
	Total	3044.234	343	2370.146

Guide de Formation au SPSS 10.0 pour Windows Module 3 - Tableaux, & questions à réponses multiples

La commande Tables

En utilisant **Tables** vous pouvez calculer différents types de statistiques et les présenter de différentes façons selon votre choix. **Tables** vous permet de faire les choses suivantes :

- choisir comment vous désirez rassembler les données et les statistiques pour les présenter en lignes, colonnes ou couches. Les variables peuvent être empilées (*stacked*) ou emboîtées (*nested*) : le fait de les empiler signifie que plus d'une variable peuvent être présentées les unes sous les autres, ou en colonnes une à côté de l'autre; les mettre en niches signifie que toutes les valeurs pour une variable seront présentées sous les valeurs individuelles d'une autre variable.
- manipuler la structure d'un tableau, son contenu et le format de la présentation.
- inclure des pourcentages flexibles, en spécifiant la base des pourcentages (leur dénominateur) de façon à ce que leurs sommes à travers les lignes, colonnes, sous-tableaux ou tableaux fassent 100%.
- monter jusqu'à 60 caractères pour les étiquettes des variables et des valeurs.

Il y a 4 types de tables sous la commande Custom Tables :

Basic tables - tous les outils sous Tables sont appliqués de façon uniforme sur toutes les variables

General tables - les statistiques en niche ou superposées et les totaux peuvent être appliqués de façon différente sur différentes variables

Multiple response tables - les variables pour lesquelles le répondant peut avoir donné plus d'une réponse à une question

Table of Frequencies (special purpose table) - fréquence de variables catégoriques qui ont les mêmes catégories, i.e. le plus important ... le moins important

Comparons la procédure **Crosstabs** avec la procédure **Tables** pour *crosstabulation*.

Ouvrez le fichier que nous avons créé et qui contient la variable âge, Q1A-AGE.SAV.

1. **File/Open/Data...**
2. Sélectionnez q1a-age.sav
3. Collez, sélectionnez et exécutez

Dans un premier temps faites une *crosstabulation* en utilisant **Crosstabs**.

1. **Analyze/Descriptive Statistics/Crosstabs...**
2. Déplacez **ca2** à Row(s):
3. Déplacez **age_gp** à Column(s):
4. **Cells...**
5. Sélectionnez **Observed** dans la section Counts
6. Sélectionnez **Row** dans Percentages
7. **Continue**
8. Collez et exécutez

Ci-dessous le résultat :

RELATION TO HEAD * age group Crosstabulation

		age group				Total
		0 to 10	11 to 19	20 to 60	61 and older	
RELATION TO HEAD	Count		6	296	41	343
	% within RELATION TO HEAD		1.7%	86.3%	12.0%	100.0%
WIFE/HUSBAND	Count		25	280	5	310
	% within RELATION TO HEAD		8.1%	90.3%	1.6%	100.0%
SON/DAUGHTER	Count	503	184	31		718
	% within RELATION TO HEAD	70.1%	25.6%	4.3%		100.0%
MOTHER/FATHER	Count			5	1	6
	% within RELATION TO HEAD			83.3%	16.7%	100.0%
OTHER RELATIVE	Count	70	55	16	2	143
	% within RELATION TO HEAD	49.0%	38.5%	11.2%	1.4%	100.0%
Total	Count	573	270	628	49	1520
	% within RELATION TO HEAD	37.7%	17.8%	41.3%	3.2%	100.0%

Utilisons la commande **Basic Tables** pour produire le tableau suivant :

1. **Analyze/Custom Tables/Basic Tables...**
2. Déplacez **ca2** à Down:
3. Déplacez **age_gp** à Across:
4. Collez et exécutez

		age group			
		0 to 10	11 to 19	20 to 60	61 and older
RELATION TO HEAD	HEAD		6	296	41
	WIFE/HUSBAND		25	280	5
	SON/DAUGHTER	503	184	31	
	MOTHER/FATHER			5	1
	OTHER RELATIVE	70	55	16	2

Voici un tableau simple (**Basic Table**), avec le réglage par défaut. Les étiquettes de rangée correspondent aux étiquettes des observations pour la variable **ca2** (lien au chef du ménage). Les étiquettes pour les colonnes sont celles que vous avez assignées pour la variable **age_gp**. Si vous désirez personnaliser votre tableau, la commande **Basic Tables** peut devenir plus difficile à construire. Voyons comment personnaliser un tableau :

Modifier les étiquettes

1. Sélectionnez **Variable View** dans **Data Editor**
2. Dans la colonne **Label**, cliquez sur la cellule pour **ca2**.
3. Enlever le texte

*Vous faites cela pour que l'étiquette de **ca2** n'apparaissent pas dans la première colonne.*

- Créer un tableau
1. **Analyze//Custom Tables/General Tables...**
 2. Déplacez **ca2** à **Rows**:
 3. Déplacez **age_gp** à **Columns**:
 4. Choisissez **age_gp**, et cliquez sur **Edit Statistics...**
- Changer les statistiques
1. Sélectionnez **Count** sous **Cell Statistics**:
 2. Changez **Label**: à **N**
 3. Vérifiez que **Width**: = **5**
 4. Cliquez sur **Change**
 5. Sélectionnez **Row %** sous **Statistics**:
 6. Changez l'étiquette **Label**: à **%**
 7. Vérifiez que **Width**: = **5**
 8. **Add**
 9. **Continue**
- Ajouter le total
1. **Insert Total**
 2. Sélectionnez **age_gpTotal**, et cliquez sur **Edit Statistics...**
 3. Sélectionnez le cercle près de **Custom Total Statistics**
 4. Sélectionnez **Count** sous **Statistics**:
 5. Changez **Label**: à **N**
 6. Changez **Width**: à **5**
 7. **Add**
 8. **Continue**
 9. Sélectionnez **ca2**, cliquez sur **Insert Total**
- Modifier le format
1. **Format...**
 2. Mettez **Empty Cell Appearance** à **Zero**
 3. **Continue**
- Ajouter des titres
1. **Titles...**
 2. Dans la boîte **Title** :
 3. Tapez **Table 1: SPSS for Windows Sample Session**
 4. Tapez **Age Breakdown by Relation to Head**
 5. Dans la boîte **Caption Box**, tapez **Source: Nampula family sector household survey, 1991.**
 6. Dans la boîte **Corner**: tapez **Relation to Head**
 7. **Continue**

Collez et exécutez la commande.

**Table 1: SPSS for Windows Sample Session
Age-Breakdown by Relation to Head**

Relation to Head	age group								Total
	0 to 10		11 to 19		20 to 60		61 and older		N
	N	%	N	%	N	%	N	%	
HEAD	0	.0%	6	1.7%	296	86.3%	41	12.0%	343
WIFE/HUSBAND	0	.0%	25	8.1%	280	90.3%	5	1.6%	310
SON/DAUGHTER	503	70.1%	184	25.6%	31	4.3%	0	.0%	718
MOTHER/FATHER	0	.0%	0	.0%	5	83.3%	1	16.7%	6
OTHER RELATIVE	70	49.0%	55	38.5%	16	11.2%	2	1.4%	143
Total	573	37.7%	270	17.8%	628	41.3%	49	3.2%	1520

Source: Nampula family sector household survey, 1991

Voici le tableau résultant de ces commandes. Notez que vous pouvez toujours changer les propriétés du tableau, son format, son orientation, etc.

L'étiquette du coin ne s'affiche pas automatiquement. Pour qu'elle s'affiche, sélectionnez le tableau en double-cliquant sur le tableau dans le Viewer. Cliquez du côté droit de la souris sur le tableau et sélectionnez **Table Properties...** du menu. Cochez le cercle près de *Nested* dans la boîte *Row Dimension Labels*. Cliquez sur OK. Et puis cliquez à l'extérieur du tableau pour continuer votre travail avec SPSS. Vous devriez voir une étiquette qui apparaît dans le coin maintenant.

Sauvegardez votre fichier de sorties du Viewer avec la commande **Save As...**

1. Choisissez **Save As...** du menu **File**
2. Nommez le fichier **session3.spo** et sauvez.

Ceci peut paraître long pour produire un seul tableau. Comme ces manipulations du logiciel SPSS et des données seront fréquentes, il y a des bénéfices à bien comprendre et à se familiariser avec la commande pour créer des tableaux. Par exemple, lorsque vous utilisez des données périodiques tels que les prix mensuels, il faut mettre à jour les tableaux du rapport de façon mensuelle également.

Ce qui suit est une comparaison des moyennes calculées utilisant les commandes **Compare Means** et **Custom Tables**, basée sur un exemple du module 2.

1. **File/Open/Data...**
2. Sélectionnez hh-file3.sav
3. Collez et exécutez
4. **Analyze/Compare Means/Means...**
5. Déplacez **kprod_ae** à **D**ependent List:
6. Déplacez **nkprod_a** à **I**ndependent list: layer 1 of 1
7. **Next**
8. Déplacez **district** à **I**ndependent List: layer 2 of 2
9. Collez et exécutez

Vous produisez le tableau qui suit :

Report

Calories produced per adult equivalent

NTILES of KPROD_AE	DISTRICT	Mean	N	Std. Deviation
1	MONAPO	1221.7281	27	416.1286
	RIBAUE	1484.0298	29	422.1161
	ANGOICHE	1272.0519	28	486.2593
	Total	1329.0592	84	452.2224
2	MONAPO	2494.8048	27	377.1214
	RIBAUE	2517.4551	30	366.0805
	ANGOICHE	2431.9673	29	296.8005
	Total	2481.5167	86	345.8224
3	MONAPO	3968.1419	28	621.3403
	RIBAUE	4000.8905	30	549.8340
	ANGOICHE	3640.3535	29	453.2870
	Total	3870.1717	87	562.9770
4	MONAPO	9150.0222	27	4686.2114
	RIBAUE	7520.2527	30	2158.8635
	ANGOICHE	8364.3191	29	4054.9027
	Total	8316.5516	86	3764.1698
Total	MONAPO	4206.4675	109	3813.5641
	RIBAUE	3900.7967	119	2559.3106
	ANGOICHE	3950.2610	115	3390.5114
	Total	4014.5183	343	3271.4011

Souvenez-vous que nous avons utilisé cette information pour remplacer les valeurs manquantes du module 2. Exécutons la commande **Custom Tables** pour produire un tableau qui ressemble à celui du module 2. Ajoutons cette fois-ci, les valeurs Minimum et Maximum au tableau.

1. Changez les étiquettes pour les variables suivantes par la fenêtre de Variable Views dans l'éditeur de données. Dans la colonne pour Label, cliquez deux fois de suite sur la boîte pour les variables indiquées ci-dessous et changez le texte comme suit :

Variable	Nouvelle étiquette (Label)
kprod_ae	delete existing label and leave blank
district	type in District
nkprod_a	delete existing label and leave blank

2. **Analyze/Custom Tables/Basic Tables...**
3. Déplacez **kprod_ae** à Summaries:
4. Déplacez **nkprod_a** à Down:
5. Déplacez **district** à Across:
6. **Statistics...**
7. Sélectionnez Mean, utilisez l'étiquette **MEAN**, Format: **ddd.dd**, Width: **5** et Decimals: **0**
8. **Add**
9. Sélectionnez Maximum, utilisez **MAX**, Format: **ddd.dd**, Width: **5** et Decimals: **0**
10. **Add**
11. Sélectionnez Minimum, avec étiquette **MIN**, Format: **ddd.dd**, Width: **5** et Decimals: **0**

12. **Add**
13. **Continue**
14. **Layout...**
15. Dans la section *Statistic Labels*, sélectionnez *Down the Left side*, **Continue**
21. **Titles...**
22. Tapez dans la boîte Title : **Table 1: Food Production in Calories**
23. Tapez sur <Enter>, et tapez par la suite **per Adult Equivalent per Day**
24. Entrez pour le Corner box: **Production Quartile**.
25. **Continue**, collez et exécutez

Ces commandes produisent le tableaux suivant :

**Table 1: Food Production in Calories
Per Adult Equivalent per Day**

Production Quartile		District		
		MONAPO	RIBAUE	ANGOCHE
1.00	Mean	1222	1484	1272
	Max	1956	1938	1952
	Min	294	429	354
2.00	Mean	2495	2517	2432
	Max	3169	3120	2961
	Min	1973	2030	2024
3.00	Mean	3968	4001	3640
	Max	5067	4834	4563
	Min	3176	3141	2996
4.00	Mean	9150	7520	8364
	Max	28466	13124	20485
	Min	5107	4984	4692

Il est possible que **Production Quartile** ne soit pas indiqué dans la boîte supérieure gauche (*Corner box*). Le *Corner box* peut être niché (*Nested*) grâce au *Pivot Table Editor*. Sélectionnez le tableau en le cliquant deux fois de suite dans le Viewer. Cliquez le côté droit du tableau et sélectionnez **Table Properties...** du menu. Sélectionnez le cercle **Nested** dans la boîte *Row Dimension Labels*. Cliquez sur OK. Et puis cliquez à l'extérieur du tableau pour poursuivre votre travail dans SPSS. Vous devriez voir le titre apparaître maintenant dans la boîte de coin.

Summaries nous indique quelles variables sont continues, normalement la variable la plus importante, celle dont nous obtiendrons des statistiques. Dans l'exemple ci-haut, des minimums, moyennes, et maximums seront calculés par catégorie **nkprod_a** pour la variable **kprod_ae**.

Subgroups détermine comment les observations seront regroupées en rangées et en colonnes. Les variables sous **Subgroups** doivent toujours être catégoriques. **Statistics** indique quelles statistiques seront calculées pour les variables continues (*Summaries*).

Si SPSS pour Windows vous avise d'une erreur pour la commande **Custom Table**, il s'agit habituellement d'une erreur de choix de variable selon la fonction demandée soit une largeur spécifiée qui est trop grande pour le tableau. Si le tableau est erroné à partir du premier écran, appuyez sur <F3> pour arrêter le processeur de SPSS. Si vous obtenez un avis d'erreur, vérifiez que vous avez spécifié avoir une variable continue sous *Summaries* et que les largeurs des variables et colonnes sont adéquates.

Une manière simple d'imprimer la table ainsi créée est de sélectionner les tables dans le **Viewer** et de les imprimer.

1. Activez le Viewer
2. Sélectionnez le tableau que vous voulez imprimer
3. **File/Print.../**. Le bouton **Selection** doit être choisi. Cliquez sur **OK**.

Exercice 3.1: Bâtissez un tableau de format similaire utilisant les calories retenues comme vous l'avez fait dans l'exercice 2.1. Assurez-vous d'**inclure (totals)** les sommes par quartile de production (vous utiliserez la commande **Basic Tables**).

Table 1: Food retention in calories
Per adult equivalent per day

Production Quartile	District			Total
	MONAPO	RIBAUE	ANGOICHE	
1				
MEAN	1148	1233	913	109
MAX	1806	1783	1391	180
MIN	224	429	208	20
2				
MEAN	2211	2146	1699	201
MAX	2544	2556	1936	255
MIN	1807	1790	1396	139
3				
MEAN	3315	3126	2405	294
MAX	4303	3730	3055	430
MIN	2555	2566	1984	198
4				
MEAN	7619	5759	4955	607
MAX	20874	9465	12675	2087
MIN	4360	3731	3064	306

Questions aux réponses multiples

Les informations recherchées requièrent souvent des questions qui permettent aux répondants d'avoir le choix des réponses. Une variable dans SPSS ne peut bien capturer les réponses étant donné que la variable ne peut avoir qu'une seule valeur par observation. La solution est d'enregistrer chacune des réponses sous une variable séparée. Les réponses peuvent être analysées séparément par des commandes que vous avez déjà vues (e.g. **Frequencies**, **Crosstabs**), mais l'idéal serait d'analyser ces variables de façon conjointe; ce qui peut se faire par une option de regroupement de la commande **Custom Tables/General Tables**. SPSS pour Windows offre deux types de méthodes selon la manière dont la question a été posée.

Si une question du sondage demande au répondant “Cochez toutes les réponses qui vous concernent” Selon un choix de 10 réponses, alors 10 variables seront nécessairement codées dans le fichier de données. Il faudra donc créer une variable pour chacune des 10 réponses possible. Chacune d’elles auraient une valeur pour indiquer si la réponse a été cochée (1), ou non (0). Ces variables sont des variables à dichotomie multiple.

Si par exemple, l’enquête demande de "lister jusqu’à 4 choix" dans une liste de 10 choix possibles, seulement 4 variables doivent être codées selon les réponses. Il y aura 10 choix de réponses de 1 à 10, et 4 d’entre eux seront enregistrer au fichier de données. Il y aura donc seulement 4 variables et celles-ci s’appellent des variables à réponse multiple.

Question 35 du questionnaire sur les ménages est un bon exemple d’une question à réponse multiple. La question est de savoir parmi les cultures produites, lesquelles sont vendues. Il est demande a chaque ménage d’indiquer jusqu’à trois cultures qui sont codées sous les noms de variable **h35a**, **h35b**, et **h35c**. Les codes pour les cultures sont fournis pour les 5 cultures les plus populaires. La question est ouverte, c.à.d. qu’un sixième code est permis afin d’inscrire d’autres cultures.

Comme cette question est ouverte, des catégories supplémentaires ont été rajoutées à ces variables. Ceci est fait une fois l’enquête terminée. Il faut trier les questionnaires pour assigner un code à chacune des nouvelles cultures ayant un code 6. Les codes et étiquettes sont alors établis pour chacune de ces cultures. Comme vous le verrez ci-dessous , il y a eut 11 codes de rajoutés pour la question 35.

La meilleure façon pour étudier cette question est d’utiliser la commande **Custom Tables** avec l’option des réponses multiples (*Multiple Response Sets*). Vous pourrez utiliser la commande **Frequencies** pour chacune des variables mais il faudra en faire la somme à la main. La commande **Custom Tables** peut calculer ces statistiques de base en créant une variable de groupe grâce à l’option des réponses multiples. Ouvrez le fichier de données sur le ménage :

1. **File/Open/Data...**
2. Sélectionnez c-hh.sav
3. Collez et exécutez

Pour créer le tableau, suivez les instructions :

1. **Analyze/Custom Tables/General Tables...**
2. **Mult Response Sets...**
3. Sélectionnez **h35a**, **h35b**, **h35c** et déplacez- les à *Variables in Set:*
4. Sélectionnez *Categories* dans la section *Variables Are Coded As*
5. **Name: crops**
6. **Label: Crops grown principally to be sold**
7. **Add**, **Save**
8. Déplacez **\$crops**, de Mult Response: à **Rows:**
9. Déplacez **district** aux **Columns:**
10. Sélectionnez **\$crops**
11. **Insert Total**
12. Sélectionnez **\$cropsTotal**
13. Changez l’étiquette pour **Total Cases** dans *Total Label:*
14. Collez et exécutez.

Vous devez obtenir le texte suivant dans l'éditeur de syntaxe :

TABLES

```
/FORMAT BLANK MISSING('.') /MRGROUP $crops 'Crops grown principally to be' +  
'sold' h35a h35b h35c  
/GBASE=CASES  
/FTOTAL= $t000001 "Total Cases"  
/TABLE=$crops + $t000001 BY district > (STATISTICS) .
```

La commande des réponses multiples (*Mult Response Sets*) crée une variable de groupe **\$crops** à partir des trois variables **h35a**, **h35b**, et **h35c**. La commande *Insert Total* permet d'avoir des statistiques de base. Comme expliqué au module 2, le format du tableau peut être personnalisé selon les options disponibles dans la boîte *Custom Tables: General Tables*.

Sauvegardez le fichier de sorties incluant tous vos tableaux par la commande **Save As...** :

1. Activez la fenêtre du Viewer
2. Utilisez **Save** du menu **File**
3. Nommez le fichier Session 3.spo

Le tableau devrait ressembler à celui-ci :

		DISTRICT		
		MONAPO	RIBAUE	ANGOICHE
Crops grown principally to be sold	COTTON	63	24	3
	PEANUTS	13	2	70
	SESAME			3
	SUNFLOWER		1	
	RICE	5	2	78
	MAIZE, BEANS	7	18	16
	BANANA		2	2
	MANIOC		2	5
	SUGAR CANE	3	1	
	TOBACCO		1	
	SWEET POTATO			1
	CASHEW NUT	1		
Total Cases	75	44	90	

Guide de Formation au SPSS pour Windows
MODULE 4 - Graphiques, tableaux, publications et présentations,
comment les exporter au traitement de texte

L'objectif de ce module est de vous donner les outils nécessaires pour préparer des rapports et apprendre comment déplacer des résultats de SPSS à d'autres logiciels tels que les traitements de texte. Tandis qu'il est possible de transférer des résultats sous forme de texte, de tableaux, de graphiques et autres formes de données, à des traitements de texte et des chiffriers, pour ce module, nous allons nous concentrer que sur des graphiques, format populaire de présentation notamment pour des bulletins d'information de marché. Par contre, les méthodes utilisées dans ce module sont similaires pour les autres types et présentations de résultats.

Ce module ne servira pas de guide sur les méthodes de préparations de rapports ou publications. Un module supplémentaire devrait être développé pour examiner cette question. Il inclurait des concepts tels que les titres, boîtes de texte et d'images, insertion d'image, etc. Néanmoins, les outils présentés dans ce module pour le transfert des résultats de SPSS à un traitement de texte est clé pour la diffusion et vulgarisation des informations de marché.

La méthode est simple : une fois que des résultats sont produits par SPSS et qu'ils sont enregistrés en mémoire, tel qu'un graphique, ces résultats peuvent être imprimés soient importés et introduits dans des rapports utilisant un traitement de texte et autres logiciels de publication. Introduire des graphiques de SPSS peut être réaliser par une simple procédure pour copier/coller (la commande **Copy** doit se faire depuis le fichier des sorties dans le Viewer). Entamez la procédure avec un tableau disponible du fichier Session3.spo. Trouvez le tableau qui suit dans ce fichier :

1. Suivez **File/Open/Output...**
2. Sélectionnez Session3.spo dans le dossier *folder* où vous avez sauvegardé vos résultats de (avec l'extension *.spo)
3. Cliquez sur **O**pen
4. Trouvez le tableau ci-dessous dans le Viewer. Cliquez une fois sur le tableau pour le sélectionner
5. Faites **E**dit/**C**opy à travers le menu (soit vous pouvez également cliquer côté droit de la souris sur le tableau pour copier)
6. Ouvrez votre traitement de texte si ce n'est pas déjà fait.

		DISTRICT		
		1 MONAPO	2 RIBAUE	3 ANGOCHE
Crops grown principally to be sold	1 COTTON	63	24	3
	2 PEANUTS	13	2	70
	3 SESAME			3
	4 SUNFLOWER		1	
	5 RICE	5	2	78
	6 MAIZE, BEANS	7	18	16
	7 BANANA		2	2
	8 MANIOC		2	5
	9 SUGAR CANE	3	1	
	10 TOBACCO		1	
	11 SWEET POTATO			1
	12 CASHEW NUT	1		
Total Cases	75	44	90	

Les instructions qui suivent sont pour les logiciels WordPerfect 8.0 et Word 97.

Avec votre traitement de texte, sélectionnez **Edit**

A) Si vous cliquez sur **Paste**, le tableau sera collé sous le format RTF - *rich text format*. Vous pouvez l'éditer ainsi comme tout autre tableau. Si vous rencontrez des problèmes, essayez B)

B) Si vous choisissez **Paste Special** du menu **Edit**, vous avez trois choix :

- 1) le format RTF - comme avec **Paste**
- 2) *unformatted text* - ce qui collera le dernier texte que vous avez copié
- 3) *picture* - collera le tableau comme une image

De préférence, choisissez *picture* (image). Par contre, l'image ne peut être modifiée dans le traitement de texte néanmoins elle retient l'image du tableau comme dans le Viewer de SPSS. Si vous voulez modifier le tableau, vous devez retourner à SPSS, faire les changements, et puis le recopier à nouveau depuis le Viewer pour le coller par la suite dans le traitement de texte.

Si vous avez copié un graphique, il faudra alors choisir **Paste Special** et sélectionner *bitmap* si vous êtes en WordPerfect, sinon les formats *bitmap* ou *Enhanced metafile* pour Word.

La commande **Copy Object** depuis le Viewer marche bien pour Word mais pas du tout pour WordPerfect.

Pour faire les changements au tableau comme rajouter un titre, il est préférable que vous les fassiez dans SPSS et non pas avec le traitement de texte.

Vous pouvez aussi copier et coller des tableaux de SPSS dans un chiffrier si vous le voulez.

Il existe plusieurs outils pour les divers formats SPSS. Vérifiez le manuel sur les tableaux de SPSS pour les instructions sur l'utilisation des *Pivot Tables* et autres options de modifications des tableaux.

Le processus est le même pour les graphiques, tels que les histogrammes ou les diagrammes à pointe. A l'aide d'un histogramme, nous allons illustrer la distribution de la propriété des arbres de cajous parmi les données sur les ménages mozambicains. .

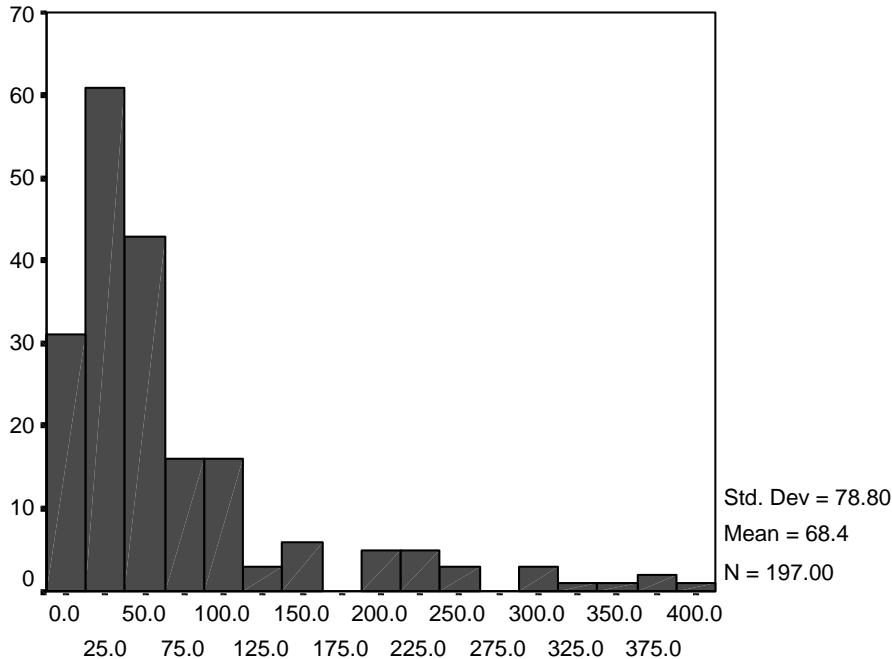
Ouvrez le fichier C-HH.SAV :

1. **File/Open/Data...**
2. Sélectionnez C-HH.sav
3. Collez et exécutez la commande

Créez l'histogramme avec la variable H57 (number of trees owned) :

4. Sélectionnez **Graphs/Histogram....**
5. Trouvez H57 dans la liste des variables et déplacez-la dans la boîte des Variables
6. Collez et exécutez

Vous devriez obtenir un histogramme comme le suivant :



NUMBER OF CASHEW TREES

Maintenant, copiez le graphique pour le coller dans un traitement de texte. Encore une fois, notez qu'il y a des différences entre les logiciels de traitement de texte. L'exemple qui suit est pour Word Perfect :

7. Activez le Viewer et cliquez une fois sur le graphique pour le sélectionner. Si vous voulez faire certaines modifications comme rajouter un titre, cliquez deux fois pour obtenir l'éditeur. Une fois terminé, cliquez à l'extérieur du graphique pour sortir de l'éditeur et cliquez à nouveau sur le graphique pour le sélectionner.
8. Utilisez CTRL-C (appuyez sur les deux touches en même temps) pour copier le graphique
9. Activez Word Perfect et placez le curseur où vous aimeriez mettre le graphique
10. Utilisez les touches CTRL-V pour coller maintenant le graphique dans Word Perfect

Une fois dans le traitement de texte, vous ne pouvez plus éditer le graphique mais vous pouvez simplement changer son placement, sa taille, son association au texte, etc. Dans SPSS, vous pouvez aussi sauvegarder un graphique comme fichier à lui seul. Cliquez deux fois sur le graphique dans le Viewer, sélectionnez **File/Export Chart** et puis BMP comme *file type*. Utilisez **Insert/file** dans Word Perfect pour insérer le graphique BMP. Pour Word, nos essais nous indiquent qu'utiliser **Edit/Copy object** fonctionne le mieux à partir du Viewer, et **Edit/Paste special/Picture** dans Word pour coller le graphique. Référez-vous au *SPSS Base 10.0 Users Guide pp.144-155* pour les autres options.

Exercice 4.1.

Sélectionnez un autre tableau de votre fichier de résultats Session3.SPO. Répétez les étapes 1 à 9 pour créer un nouveau document en Word Perfect ou Word. Exercez -vous à faire des changements dans les tableaux dans SPSS et copiez-les dans un traitement de texte.

Annexes

Les annexes qui suivent servent de références pour les usagers du guide de formation , pour expliquer diverses fonctions des commandes de SPSS les plus souvent utilisées dans cette formation, pour décrire les diverses options disponibles à l'usager à travers les menus, et pour mieux naviguer et manipuler les résultats dans le *Output navigator*.

Sélectionner un sous-ensemble des données par filtrage ou sélection temporaire des données

Vous pouvez filtrer ou éliminer les observations selon les critères de sélection. Au module 2, nous avons filtré des données sans les effacer. Lorsque vous établissez un filtre par la commande **Data/Select cases**, les observations non-sélectionnées ne sont pas incluses dans l'analyse mais elles apparaissent toujours dans le fichier de travail. Vous pouvez voir quelles observations ont été filtrées en regardant dans la colonne de gauche de la fenêtre Dat View, où les numéros sont inscrits. Les numéros avec une barre à travers sont filtrés et sont exclus de l'analyse. SPSS crée une variable filtre, *FILTER_\$*, pour indiquer ce statut. Les observations choisies ont une valeur de 1; celles filtrées ont 0. Pour terminer le filtre pour inclure toutes les observations, choisissez *Select All cases* de la commande **Data/Select cases**. Si vous voulez enlever certaines observations, utilisez la même commande **Data/Select cases**, complétez un *IF statement* pour celles que vous voulez garder, et puis cochez le choix **deleted** dans la boîte **Unselected Cases Are**. Sauvegardez ce nouveau fichier de travail sous un autre nom pour ne pas perdre les données originales.

Les diagrammes curvilignes et les trois données du diagramme

La commande **Graph/Line** vous permet de faire des sélections qui détermine le type de graphique, soit simple, multiple ou en lignes de chute. Dans le menu, choisissez l'icône pour le choix de graphique et sélectionnez par la suite votre option de données du diagramme qui décrit mieux vos données. Voyez une petite description des choix **Data in Chart** plus bas. Dans un diagramme, l'**axe de modalités** (*category axis*) est l'axe qui affiche les valeurs individuellement, sans nécessairement tenir compte de l'échelle (l'axe de graduations, en revanche, affiche les valeurs numériques à l'échelle). Les diagrammes en bâtons et les diagrammes curvilignes ont généralement un axe de modalités et au moins un axe de graduations. Les diagrammes de dispersion et les histogrammes ne présentent aucun axe de modalités.

Les Valeurs manquantes (*Missing Values*) : ces options ne sont accessibles que lorsque le nouveau diagramme affiche ou décrit plusieurs variables (variables définissant les groupes non comprises) :

- **Exclude cases listwise** exclut une observation du diagramme s'il existe une valeur manquante pour l'une quelconque des variables décrites.
- **Exclude cases variable by variable** exclut séparément une observation de chaque statistique principale calculée. Différents éléments du diagramme peuvent être basés sur différents groupes d'observations.

Display groups defined by missing values n'est disponible que lorsque vous utilisez une variable qualitative pour définir les groupes d'un nouveau diagramme. Si vous sélectionnez cette option, chaque valeur manquante de cette variable/critère de regroupement (y compris la valeur manquante par défaut) apparaîtra comme un groupe séparé dans le diagramme. Dans le cas contraire, les observations comprenant des valeurs manquantes par défaut ou des valeurs manquantes spécifiées pour cette variable/critère de regroupement sont exclues du diagramme.

Diagramme curviligne simple

Summaries for Groups of Cases

Crée un diagramme qui récapitule les modalités d'une seule variable. La hauteur-“ y” des points est déterminée par l'option Line Represents
Une seule variable Category Axis

Summaries of Separate Variables

Deux ou plusieurs variables sont récapitulées. Chaque point représente l'une des variables.
Deux ou plusieurs variables Line Represents.

Values of Individual Cases

Crée un diagramme qui récapitule une seule variable par observation individuelle.
Une seule variable Line Represents

Diagramme curviligne multiple

Summaries for Groups of Cases

Crée un diagramme qui récapitule les modalités d'une variable dans des modalités d'une autre variable. La hauteur y des points est déterminé par l'option Line Represents
Une variable Category Axis (Category Variable 1).
Une variable Define Lines By (Category Variable 2).

Summaries of Separate Variables

Crée un diagramme qui récapitule deux ou plusieurs variables dans des modalités d'une autre variable.
Deux variables Lines Represent ou plus (Var 1, Var 2).
Une variable Category Axis (Category Variable).

Values of Individual Cases

Crée un diagramme qui récapitule deux ou plusieurs variables pour chaque observation.
Deux variables Lines Represent ou plus (Var 1, Var 2).

Manipuler les résultats dans SPSS 10.0 pour Window

Plusieurs modules pourraient être écrits sur ce thème, le module 4 n'a discuté que des procédures à suivre pour copier et coller les résultats. Il est recommandé de suivre le *tutorial* offert par le logiciel SPSS pour mieux connaître les capacités et options du logiciel.

Vous aurez peut-être des difficultés pour visualiser tous les résultats provenant des commandes telles que **Frequencies** ou **Tables**. Vous pourrez obtenir des centaines soit des milliers d'observations mais SPSS vous affichera que la première cinquantaine. Pour voir toutes les données, cliquez deux fois ou cliquez sur le côté droit de la souris sur les résultats sélectionnés et choisissez **Open**. Ceci ouvrira une nouvelle fenêtre qui s'appelle *pivot table*. Vous pouvez modifier votre tableau dans cette fenêtre également.

ANNEXE 2

Socio-Economic Survey of Family Sector Farms in the Province of Nampula (Angoche, Monapo e Ribaúe)

July/August 1991

Departamento de Preços e Mercados
Food Security Project

Name of Household Head _____

Household Number _____ HH

Aldeia _____ VIL

Distrito _____ DIST

(Subset of questions from original questionnaire)

I. HOUSEHOLD CHARACTERISTICS

- H1** _____ 1. How many persons are in this household?
- H4** _____ 4. Has your family always lived in this village?
1=yes 2=no
- H8** _____ 8. Is your family registered as "deslocada"?
1=yes 2=no
- H19** _____ 19. Do you presently have lands in fallow?
1=yes 2=no
- H21** _____ 21. What is the total area of these fallowed parcels? (hectares)
- H24** _____ 24. Do you have lands that you have completely abandoned?
1=yes --> question 25 2=no --> question 27
- H25** _____ 25. What is the total area of these abandoned lands? (hectares)
- H26** _____ 26. What was the principal motive for abandoning these lands?
1=no security
2=lands lost fertility
3=lack of labor
4=insect attacks
5=other

[We would like to ask you about the food crops you grow.]

- H29** _____ 29. Over the last five years, have you increased or decreased the amount of land in food crops?
1=increased 2=decreased 3=no change
- H31** _____ 31. During a normal year, is your farm production sufficient to feed your entire family?
1=yes 2=no

[We would like to ask you about the cash crops you grow on your farm?]

H34 _____ 34. Do you grow any crops that are principally destined for the market?
1=yes 2=no

35. Which crops are grow principally to be sold? (List the three most important)

H35A _____ 1=cotton 4=sunflower

H35B _____ 2=peanuts 5=rice

H35C _____ 3=sesame 6=other

H36 _____ 36. Over the last five years, have you changed the area grown in these cash crops?
1=increased
2=decreased
3=no change

H39 _____ 39. Do you normally grow cotton?
1=yes 2=no

H52 _____ 52. Since your involvement with the cotton companies, have you reduced your area dedicated to food crops, such as maize and manioc?
1=yes 2=no

IV. PRODUCTION

H56 _____ 56. Do you have cashew trees?
1=yes 2=no

H57 _____ 57. How many trees do you presently have? (number)

H57A _____ 57A. Of these trees, from how many did you harvest during the last year?
(number)

V. AGRICULTURAL SALES

We would like to ask about the marketing of your agricultural products since August of 1990.

64. Over the last five years, have you increased the quantities marketed of the following crops:

H64A _____ a. maize 1=yes 2=no

H64B _____ b. manioc 1=yes 2=no

H64C _____ c. rice 1=yes 2=no

H64D _____ d. cotton 1=yes 2=no

H64E _____ e. peanuts 1=yes 2=no

H64F _____ f. beans 1=yes 2=no

H64G _____ g. sorghum 1=yes 2=no

H64H _____ h. cashew nuts 1=yes 2=no

H65 _____ 65. Compared with five years ago, has the marketing of these products been more difficult or easier?
1=more difficult --> question 66
2=easier --> question 67

H66 _____ 66. If more difficult, why?
1=fewer buyers
2=transportation problems
3=security problems
4=low prices
5=lack of consumer goods
6=other _____

H67 _____ 67. If easier, why?
1=more buyers
2=better transportation
3=better security
4=attractive prices
5=more consumer goods
6=other_____

H83 _____ 83. Does your family usually receive traditional gifts or participate in exchange relations?
1=yes 2=no

H84 _____ 84. If yes, how often?
1=only when there is a lack of food
2=only during feasts and rituals
3=frequently

XI. TYPICAL CONSUMPTION PATTERNS.

H86 _____ 86. How many meals did these people have yesterday? (Number of meals)

H89 _____ 89. Do you consider these meals adequate to maintain the health of all the household members?
1=yes 2=no

We would also like to ask you about your diet during the hungry period (January to May).

H91 _____ 91. How meals do you customarily prepare daily during hungry period?

H92 _____ 92. In general, are these hungry period meals adequate to maintain the health of all household members?
1=yes 2=no

H96 _____ 96. During the hungry period, was there always food available to purchase from the market or from your neighbors?
1=yes 2=no

I. CARACTÉRISTIQUES DU MÉNAGE

Table IA: Household Characteristics

Name	Family Member Number	This person works on-farm or off-farm 1=yes 2=no	Relation to Head 1=head 2=spouse 3=child 4=parent 5=other kin 6=other	Age	Sex 1=m 2=f	Level of Schooling (enter the last completed year) 0=illiterate 12=post-high school 98=no formal schooling but literate	Marital Status 1=monogamous 2=polygamous 3=single 4=widowed 5=divorced 6=emigrant wife (husband out longer than six months)
	MEM	CA1	CA2	CA3	CA4	CA5	CA6
	1		Head				
	2						
	3						
	4						
	5						
	6						
	7						
	8						
	9						
	10						
	11						

IV. PRODUCTION

V. VENTES AGRICOLES

Table V: Sales of Farm Products

Sale	Crop	Quantity sold		Period of sale	Motive for sale at this time	Buyer	Locale of sale	Distance from the farm	Why sold to this buyer	Value of Sales		Who in the household is responsible for the sale
		Units	No. of Unit							meticais	Unit	
	1=corn 2=manteiga bean 3=beans 4=manioc 5=rice 6=cotton 7=peanuts 8=cashew nut 9=cashew drink 10=cocos others	1=sack 100 2=sack 50 3=kilo 4=liter 5=can 20		1= planting (Aug-Dec.) 2= hungry period (Jan-April) 3=this year's harvest 4= various times	1=needed money 2=buyers available 3=consumer goods available 4=attractive price	1=lojista 2=wholesaler 3=AGRICOM 4=ambulante 5=brigada 6=company	1=farmgate/ house 2=village 3=locality 4=district 5=province	(enter the kms between farmer and point of sale)	1=the only one available 2=always sell to this one 3=best price 4=transportation provided 5=carries consumer goods		1=unit price 2=total value	1=husband 2=wife
VE	V1	V2A	V2B	V3	V4	V5	V6	V7	V8	V9A	V9B	V10
1												
2												
3												
4												
5												
6												
7												
8												
9												

N.B. Pas toutes les variables apparaissant ici sont incluses dans le fichier C-Q5.sav. Seules VEN, V2a, V2b, V9a et V9b ont été retenues pour cet entraînement. La variable PROD remplace V1.

ANNEXE 3

Analyse par ordinateur des données d'enquête - Organisation de fichiers de données à niveaux multiples, par Chris Wolf, MSU, Département d'agro-économie

Vous pouvez télécharger le document au <http://www.aec.msu.edu/agecon/fs2/survey/index.htm> disponible en version française et anglaise.