

SPSS para Windows SAMPLE SESSION

Cross-Sectional

**Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 7.5
3e Edição**

Revisão feita por

Jean-Charles Le Vallée, Visiting Research Specialist, MSU

levallee@pilot.msu.edu

Feito inicialmente por o Agricultural Economics Computer Service em MSU

Departamento de Economia Agrária, Universidade Estadual de Michigan

East Lansing, Michigan, USA

Fevereiro de 1999

Material de Treinamento para Curso de Curta Duração

Seccão 0 -Estrutura de ficheiros para SPSS 7.5. para Windows (janelas (“windows”) dentro do programa - o *Syntax Editor*, o *Datas Editor* e o *Output Navigator*)

Seccão 1 - Ficheiros de SPSS, Descritivos e Transformação de Dados

Seccão 2 - Restruturação dos Data Files - Table Lookup & Aggregation

Seccão 3 - Tabelas & Perguntas de Multipla Escolha

ANEXO

- Tres tabelas do relatório preliminar de pesquisa No. 3. Um inquérito socio-económico do sector familiar na provincia de Nampula: Observações metodológicas. Direcção Nacional de Economia Agrária, MAP.

Acknowledgments

Os fundos para este treinamento foram providenciados pela Food Security II Cooperative Agreement entre AID/Global Bureau, Office of Agriculture and Food Security, e o Department of Agricultural Economics at Michigan State University.

SPSS para Windows SAMPLE SESSION
SECCÃO 0 - Estrutura de ficheiros para SPSS 7.5. para Windows
(Janelas de Data, Syntax Output)

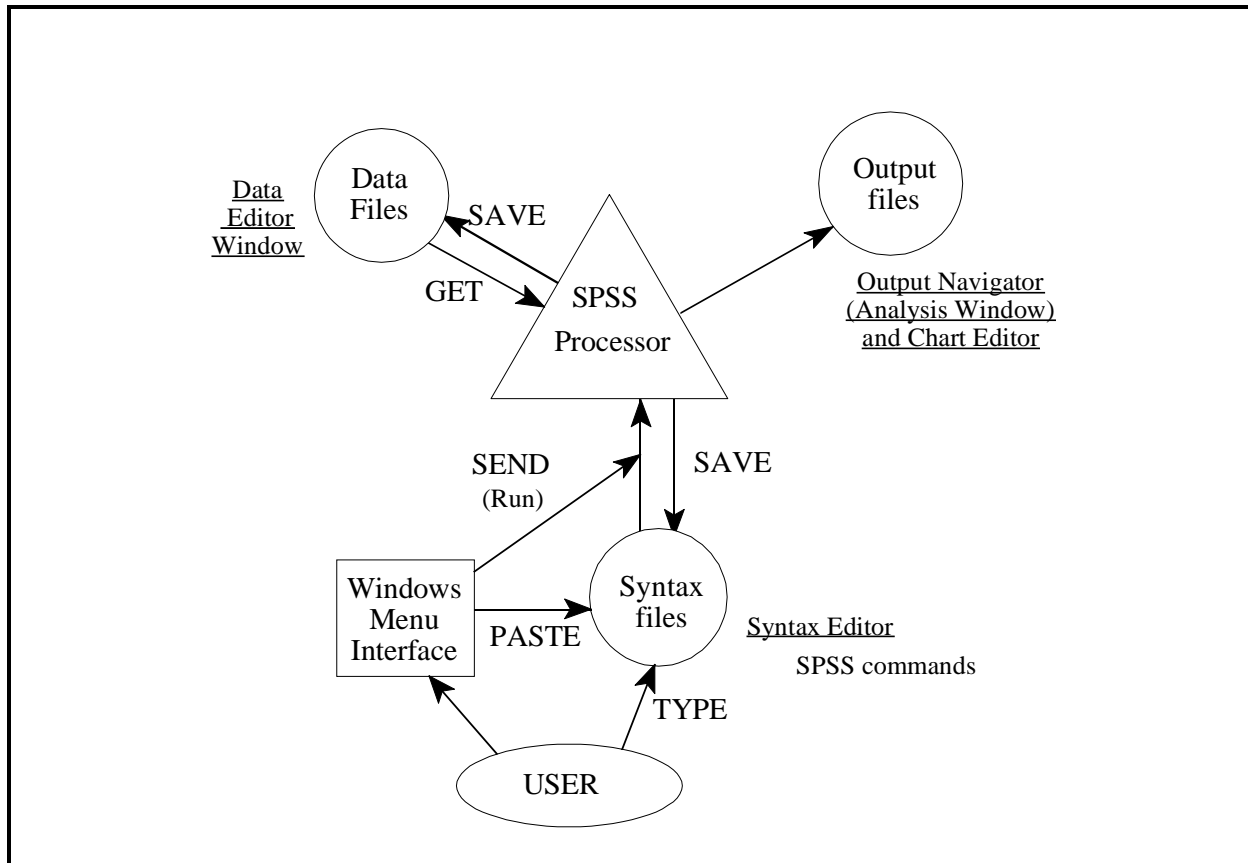
Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 7.5
3e Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Fevereiro de 1999

Este módulo introduz os conceitos básicos de níveis de dados e métodos de organização de dados. Também dá uma descrição breve sobre a estrutura dos ficheiros no SPSS 7.5 para Windows. É essencial ler este módulo antes de começar o Sample Session.

Ficheiros Utilizados no SPSS 7.5 para Windows

Ao utilizar o SPSS 7.5. para Windows na maneira ensinada neste “sample session”, está a usar tres diferentes janelas (“windows”) dentro do programa - o *Syntax Editor*, o *Das Editor* e o *Output Navigator*. E importante conhecer a significancia dos diferentes tipos de ficheiros e saber os diferentes comandos que se utiliza para criar e utilizar os ficheiros.



O *Syntax Editor*

O *Syntax Editor* é a janela onde se escreve os comandos antes destes serem submetidos ao processador de SPSS. Para entrar comandos no *Syntax Editor*, pode-se escrever os comandos directamente no *Syntax Editor*, ou pode-se utilizar o *pull-down menus* e seleccionar **Paste** depois de afinar o comando. Ha quatro usos principais do *Syntax Editor*:

- Para escrever ou *paste* comandos para posterior processamento pelo SPSS,
- Para enviar estes comandos ao SPSS 7.5. para Windows para processamento,
- Para *write* ou guardar estes comandos a um ficheiro para uso futuro, e
- Para *retrieve* (buscar) ficheiros de comandos que guardou anteriormente.

E importante entender que os comandos que se colocam no *Syntax Editor* não serão executados (quer dizer, não haverá produção de resultados) até enviar os comandos para o processador. O *Syntax Editor* é simplesmente uma área que lhe permite preparar os comandos. Para enviar os comandos para o processador, utiliza-se o **Run button** no SPSS 7.5. para Windows **Toolbar**. Depois de pressionar o Run button, o computador envia os comandos para o processador, o qual lee os comandos contidos no *Syntax Editor* e os executa. Depois de todos os comandos serem processados, o SPSS apresenta o *Output Navigator* para examinar os resultados dos comandos. Então é possível voltar para o *Syntax Editor*, acrescentar mais comandos ou editar os comandos

antigos e executar estas mudanças (manda-las para o processador) para observar os resultados diferentes.

Depois de completar com sucesso cada passo na sua análise (ou quando estiver pronto para terminar uma sessão de SPSS para Windows 7.5 session, embora não tivesse sucesso completo), deve guardar os comandos num ficheiro para uso futuro. Para guardar os comandos, active o *Syntax Editor* e seleccione **Save** no **File menu**. Um ficheiro criado no *Syntax Editor* chama-se um *syntax (or command) file*. Este é um ficheiro que contem só comandos; nunca contem os dados que estão a ser analisados. E necessário guardar os dados aparte, conforme se explica na secção a seguir. Ao nomear os *syntax files*, sugerimos que se use a extensão automática de “.SPS”. Exemplos são REP7.SPS, DEM-ALL.SPS, e SECTION1.SPS.

Escrevendo os comandos para um *syntax file*, é possível fazer o **retrieve** (buscar), examinar ou modificar comandos e processa-los de novo. Pode-se fazer o **retrieve** dum *syntax file* acessando o **File menu** em qualquer das janelas de SPSS e seleccionando **Open**. No menu nomeado **Files of type**, seleccione **Syntax(*.sps)** e faça o retrieve do ficheiro utilizando o nome no qual guardou o ficheiro a última vez. Depois de abrir um ficheiro específico, pode-se utilizar os comandos no ficheiro sem a necessidade de escreve-los outra vez. Se alteras os comandos no *syntax file* e quer preservar estas mudanças, não esqueça de guardar de novo o ficheiro.

O Data Editor Window

SPSS for Windows 7.5 armazena os dados num *data file*. Para além dos valores em si, um *data file* contem informação tal como *variable labels* e *value labels*, informação sobre o formato das variáveis especificações sobre *missing values*, etc. Para fazer qualquer análise de dados no SPSS para Windows 7.5, tem primeiro que abrir um *data file*. Para fazer isto, primeiro abra o **File menu** e seleccione **Open** (depois disto, pode fazer o **paste** do comando no *Syntax Editor* e correr o comando, ou pode correr o comando directamente fazendo o “click” no **Open** button). Depois de correr o comando, os dados no ficheiro são disponíveis no *Data Editor window*.

Frequentemente abrirá um *data file*, calculará novas variáveis e fará outras transformações para finalmente guardar o ficheiro modificado para uso posterior. Por exemplo, poderia abrir um *data file* contendo superfície por cultura, acrescentar produção por cultura vinda de outro ficheiro, e então calcular o rendimento. Se se quer utilizar posteriormente estas novas variáveis de produção e rendimento, é necessário guardar o *data file* com as novas variáveis. Para fazer isto, tem que activar o *Data Editor Window* e seleccionar **Save As...** no **File menu**, dando ao ficheiro um novo nome. Alternativamente, pode substituir o ficheiro antigo com o ficheiro novo, guardando o novo ficheiro com o mesmo nome que o antigo.

O Output Navigator

SPSS para Windows 7.5 automaticamente escreve no *Output Navigator* todos os resultados da execução dos comandos. Por exemplo, se se corre um comando **frequency**, a tabela de frequência que se especifica é escrita no *Output Navigator*. A semelhança disto, se executas comandos que criam uma tabela ou gráfico, estes também serão enviados para o *Output Navigator*. Para guardar o conteúdo do *Output Navigator* para um ficheiro, active o *Output Navigator*, abra o **File menu** e seleccione **Save As...** Ao nomear o ficheiro, SPSS automaticamente colocará a extensão .SPO. E muito importante guardar o *output file*. Este lhe proporciona acesso aos resultados depois da sessão de SPSS para Windows 7.5 acabar. Por exemplo, pode imprimir os resultados da sessão para examinar os resultados e verificar erros. No sample session, verá como guardar o conteúdo do

Output Navigator e dar nomes diferentes aos ficheiros de cada sessão. Finalmente, é possível manipular os resultados exactamente como se estivesse a usar um *file manager* (chamado Explorer em Windows 95). No *Output Navigator*, ha dois áreas; a área a mão direita produz os resultados enquanto a área à esquerda lhe permite manipular os resultados, por exemplo apagando dados, mudando nomes de títulos, movendo resultados, etc.

Sumário dos Tipos de Ficheiros

Syntax files (ou command files) contêm comandos guardados no *Syntax Editor*. Estes ficheiros não contêm resultados ou dados, só comandos. A semelhança do SPSS 6.1.3, a extensão para estes ficheiros é *.SPS (*.LOG em SPSS/PC+).

Output files contêm resultados estatísticos, informação sobre dados, e tabelas, gráficos, etc. criados pelo processador de SPSS for Windows 7.5. Estes ficheiros não contêm dados. A nova extensão destes ficheiros é *.SPO (*.LIS em SPSS/PC+ e *.LST em SPSS 6.1.3).

Datas files contêm dados, incluindo variáveis originais e variáveis criadas através dos comandos do SPSS para Windows 7.5. Abre-se estes ficheiros com o comando **Open...** Para SPSS 6.1.3. e 7.5., a extensão para estes ficheiros é *.SAV (*.SYS para SPSS/PC+).

**SPSS para Windows SAMPLE SESSION
CROSS-SECTIONAL
Departamento de Economia Agrária e Serviço de Informática
Michigan State University
Fevereiro de 1999**

Este “sample session” é desenhado para apresentar os comandos necessários para análises típicas de dados de inquéritos através do SPSS para Windows 7.5. Você trabalhará independentemente na sessão, mas para ser mais efectivo, será necessário ter acesso a alguém que tenha bom conhecimento do SPSS para Windows para lhe ajudar no início e para responder às perguntas que possam surgir ao longo da sessão.

Uma cópia do questionário utilizado para recolher os dados que se utilizam neste sample session encontra-se no **Relatório Preliminar de Pesquisa No. 3: Um Inquérito Socio-Económico do Sector Familiar na Província de Nampula: Observações Metodológicas**, Direcção Nacional de Economia Agrária, MAP. Este relatório está disponível na secção do anexo até o fim do sample session. Si tem perguntas, comunica-se com o Dr. Weber (webermi@pilot.msu.edu). O sample session faz referência a quatro partes do questionário, cada uma das quais corresponde a um ficheiro de SPSS para Windows. Outros dois ficheiros de SPSS para Windows são necessários para a conversão de unidades de medida. Os ficheiros utilizados são:

Secção do Questionário	SPSS para Windows Data File
Secção Principal a Nível de Agregado Familiar	C-HH.SAV
Tabela IA: Características dos Membros do Agregado	C-Q1A.SAV
Tabela IV: Características da Produção	C-Q4.SAV
Tabela V: Vendas de Produtos da Machamba	C-Q5.SAV
Factores de conversão para cálculos de kilogramas	CONVER.SAV
Factores de conversão para cálculos de calorias	CALORIES.SAV

Nestas sessões assume-se que:

- Você sabe usar Windows com um mouse
- Os seis data files descritos acima estão situados no directório C:\SAMPLE no seu disco duro, si você não sabe como fazer, ver as instruções para a instalação.
- Dentro de **Preferences** no **Edit menu**, está indicado para SPSS para Windows listar variáveis na mesma ordem em que se alistam no ficheiro
- Também dentro de **Preferences** no **Edit menu**, está indicado para SPSS para Windows listar os comandos no **output window**
- O **Syntax window** não aparece no início da sessão

Nota: Se você tem alguma dúvida sobre estes aspectos, consulte com um orientador de SPSS ou de informática. Não esqueça nunca de guardar, ao acabar cada módulo, os *syntax files* e *output files* que se criaram durante a sessão. Utilize nomes lógicos como, por exemplo, módulo1.sps, output1.spo, etc.

SPSS para Windows SAMPLE SESSION
SECCÃO 1 - Ficheiros de SPSS, Descritivos e Transformação de Dados

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 7.5
3ª Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Fevereiro de 1999

Inicia sua SPSS para Windows 7.5. Asegura-se de ler a secção 0 para esclarecer o conceito do **Syntax window**, onde faz o **paste** dos comandos ou onde os escreve, do **Output Navigator**, onde o SPSS para Windows apresenta os resultados dos seus comandos, e do **Data Editor window** onde se observa os dados do *working data file*.

Data Files e o Working File

Dados dos questionários que foram entrados no SPSS para Windows 7.5 são armazenados em *data files*. Se se quer trabalhar com uma base de dados, é necessário primeiro abrir o *data file* correspondente, para que os dados sejam disponíveis ao programa.

Quando se abra um *data file*, é carregado do disco para a memória do computador, através do qual este *data file* chega ser o *working data file*. Isto significa que os dados deste ficheiro são disponíveis para ser trabalhados. Vamos começar com o questionário para Tabela IA: Características dos Membros de Agregado. O *data file* que corresponde a esta secção do questionário é C-Q1A.SAV. Para abrir este ficheiro, faça o seguinte:

1. Desde o **File** menu, seleccione **Open...**
Isto abrirá o Open File dialog box.
2. Vai para o directório onde se encontram os dados do sample session e seleccione of ficheiro **c-q1a.sav**.
3. Click no **Paste** button para colocar o comando no **Syntax Editor**.
O Syntax Editor agora será o "active window" e aparecerá o seguinte texto:

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

No Syntax Editor.
4. Coloque o cursor em qualquer parte da linha do comando "GET" e click no **Run** button no **Toolbar**.
Notará que o comando GET FILE que voce acaba de correr será escrito no Output Navigator.

O *data file* a nível de membro do agregado agora está na memória.

Uma coisa que frequentemente queremos saber é as variáveis que um *data file* contem. Podemos averiguar isto, junto com outras informações, utilizando o comando **Variables...** no **Utilities** menu, tanto no *Syntax Editor* como no *Data Editor*. Isto lhe permite observar as definições das variáveis

e os *variable labels*. Para fazer isto, faça o seguinte:

1. Desde o **Utilities** menu , eleccione **Variables...**
2. Selecciono o nome duma variável e a informação referente a essa variável aparecerá à mão direita.

Este “display” mostra informação adicional, incluindo os *value labels* para variáveis como **district**, **vil**, **ca1**, **ca2**, **ca4**, **ca5**, **ca6**, e **univ**, o tipo de variável, o *display width* das variáveis em termos de caracteres, o número de decimais (se **Type for Numeric**), e quaisquer *missing values*.

Click no  button quando acabas

Para escrever toda esta informação no Output Navigator para examinação posterior, faça o seguinte:

Selecciono o **Utilities** menu e selecciono **File Info**.

Este comando será executado imediatamente. O Output Navigator virá ser activo e terá uma lista de todas as variáveis e as suas respectivas definições.

Esta informação está incluída no Output Navigator. Pode ver o nome de cada variável, as suas *labels*, e os formatos, por exemplo 8.2 significa largura 8 com dois decimais. Fazendo o *Display* e guardando esta informação é uma maneira de documentar os seus *data files*.

Estatísticas Descritivas - envolvendo uma variável

A primeira coisa que se deve fazer ao começar uma análise é gerar estatísticas descritivas (por exemplo, médias, máximo, mínimo, desvio padrão) para todas as variáveis. Este tipo de análise ajuda em vários aspectos, tais como identificar erros de entrada de dados, proporcionar maior perspectiva e conhecimento sobre a base de dados, permitir confirmar que os *missing values* foram bem definidos, etc. As vezes é fácil pensar não fazer este passo para algumas bases de dados ou algumas variáveis, mas é um passo importante que quase sempre poupará tempo e melhorará análise no futuro. Por exemplo, se se calcula a idade média de todos os inquiridos, muito embora voce não tenha interesse neste dado, e o resultado indica que esta idade média é 91.3 anos, voce como analista sabe logo que provavelmente há algum problema com os dados de idade.

Estatísticas descritivas básicas podem ser geridas através de dois comandos muito comuns de SPSS para Windows -- **Descriptives** e **Frequencies**. **Descriptives** é utilizado para variáveis contínuas, enquanto **Frequencies** é utilizado para variáveis categóricas.




Uma variável contínua é uma variável que não tem um número fixo de valores. Uma variável categórica é uma variável que tem um número limitado de valores for formam categorias. Por exemplo, Examine Tabela IA: variável **ca3**, idade, é uma variável contínua porque idade pode ter muitos diferentes valores. Variável **ca2**, relação ao chefe, é uma variável categórica orque os serus valors se limitam ao intervalo 1-6.

Começe examinando os dados utilizando o Data Editor window. Para isto, faça o seguinte:

1. Click no **Go To Data Editor**  **button** no **Toolbar**.
2. Examine os dados
Um ponto numa célula indica "missing value" ("sysmis").




Este processo dar-lhe-á maior compreensão da base de dados. Poderá também identificar erros óbvios, por exemplo uma variável que tem *missing values* para todos os casos. Decide quais das variáveis são contínuas e quais categóricas (normalmente faria referência ao questionário para esta decisão). É necessário saber isto para seleccionar o procedimento apropriado para cada variável. Se se comete o erro de correr um **Frequencies** numa variável contínua, provavelmente receberá mais output do que queria, possivelmente com centenas de diferentes "categorias", uma para cada valor da variável. Se se corre um **Descriptives** numa variável categórica, geralmente receberá resultados sem nenhuma significado, já que o valor médio duma variável que tem categorias não tem significado.

Examinando os dados, devia ter averiguado que a variável **ca3** is é contínua e que todos os outros são categóricas. Para correr um **Descriptives** na variável **ca3**, faça o seguinte:

1. Desde o **Statistics** menu seleccione **Summarize/Descriptives....**
Verá o Descriptives dialog box.
2. Seleccione **ca3** da lista a mão esquerda e click no  **button**.
ca3 irá para o Variable(s): box a mão direita
3. Click no  **Paste** **button** para colocar o comando no Syntax window e activar o *Syntax Editor*
4. Execute o comando fazendo click no Run  **button** localizado no **Toolbar**. (Notará que esta vez não foi necessário mover o cursor, uma vez que já esteve posicionado numa das linhas do comando **Descriptives**)

O Output Navigator virá ser a janela activa e os resultados do comando aparecerão lá. Verá que a média da **ca3** é 21.34.

Já que as variáveis **ca1**, **ca2**, **ca4**, **ca5** e **ca6** são categóricas, vamos correr um **Frequencies** neles. Para correr **Frequencies**, faça o seguinte:






1. Desde o **Statistics** menu seleccione **Summarize/Frequencies....**
Verá o Frequencies dialog box.
2. Seleccione **ca1** da lista na mão esquerda e click no  **button**.
ca1 irá para o Variable(s): box a mão direita
3. Repete passo 2 até **ca2**, **ca4**, **ca5** e **ca6** forem todas movidas para o *Variable(s): box*.
4. Click no  **Paste** para colocar o comando no Syntax Editor e activa-lo.
5. Execute o comando fazendo click no Run  **button** localizado no **Toolbar**.

O Output Navigator virá ser activo e os resultados estarão lá. Voce verá, por exemplo, que **ca1** mostra que 70.7% dos membros trabalham e que **ca6** mostra que 38.0% dos inquiridos são monógamos.

Para uma descrição completa do output que se recebe através dos comandos **Descriptives** e **Frequencies**, faça referência ao SPSS para Windows Base System User's Guide Release 7.5, páginas 161-169. Outro comando utilizado para produzir vários tipos de estatística descritivas é o comando

Explore. Uma das estatísticas mais úteis que produz é *outliers*. O comando **Explore** pode produzir grandes quantidades de output se for utilizado com os seus *defaults*. Nos vamos limitar o output, seleccionando só estatísticas.

Corre o comando **Explore** na variável de idade (**ca3**) através dos seguintes passos:

1. Desde o **Statistics** menu seleccione **Summarize/Explore...**
2. Seleccione **ca3** da lista a mão esquerda e click no  próximo ao **D**ependent List.
3. Na esquina sul-occidental do *dialog box* verá o **D**isplay box. Click no *radio button* (círculo) próximo a **S**tatistics.
Isto nos dará só estatísticas e não gráficos.
4. Agora click no  button.
Verá o Explore:Statistics dialog box.
5. Click uma vez no quadrado ao lado de **O**utliers para colocar um “X” no quadrado.
Verá que já existe um “X” no quadrado ao lado de Descriptives.
6. Click no  button.
Voltará para o Explore dialog box.
7. Click no  para colocar o comando no Syntax Editor e activa-lo.
8. Click no  Run.

Este procedimento mostrar-lhe-á os cinco maiores e menores valores da **ca3**, para saber se tem alguns *outliers*. Estes valores extremos (os *outliers*) serão identificados pelos seus *case numbers*. Faça referência às páginas 171-177 do SPSS para Windows Base System User's Guide Release 7.5 para ver uma explicação do comando **Explore**.

Aplice o que acaba de aprender sobre estatísticas descritivas fazendo o seguinte exercício.

Exercício 1.1: Corra estatísticas descritivas em outro ficheiro. Utilize o questionário de produção - Tabela IV, cujos dados estão no *data file* C-Q4.SAV.

Pontos:

- a. Faça com que o C-Q4.SAV seja o *working data file*.
- b. Use o comando **Descriptives** para variáveis contínuas e **Frequencies** para variáveis categóricas.
- c. **Prod** é uma variável categórica.
- d. As quantidades (**p1b**, **p2b**, ...) são variáveis contínuas.
- e. As unidades (**p1a**, **p2a**, ...) são variáveis categóricas.
- f. **p4 & p6** são categóricas.

Uma pequena amostra do que deveria ver através destes comandos é o seguinte:

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	
Total	1693	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 - # OF UNITS	1231	,0	1460,0	15,612	86,104
STOCK FOR SEED - # OF UNITS	869	,0	100,0	4,938	6,876
Valid N (listwise)	151				



Estatística Descritiva - envolvendo duas ou mais variáveis

O comando **Crosstabs** produz tabelas mostrando a distribuição de casos de acordo com os seus valores para duas ou mais variáveis categóricas.

Olhe para o questionário dos membros do agregado familiar, TABLE IA (no anexo). Uma das coisas que poderia estar interessado em conhecer é como o sexo dos respondentes varia em relação com o chefe do agregado familiar. Isto poderia dizer-lhe, por exemplo, quantas mulheres são chefes dos agregados familiares. Esta espécie de sumário pode ser produzido usando o comando **Crosstabs**. Torne o ficheiro de membros do agregado familiar, C-Q1A.SAV, num ficheiro de dados de trabalho.

1. Do **File** menu, seleccione **Open...**
2. Seleccione o ficheiro c-q1a.sav.
3. Click no **Paste** para colocar o comando no Syntax Editor e fazê-lo activo.
4. Coloque o cursor num sitio qualquer da linha contendo o comando "GET" e faça o click no botão Run  do Toolbar.




Para usar o comando **Crosstabs** faça o seguinte:

1. Seleccione **Summarize** do **Statistics** menu
2. Seleccione **Crosstabs...**
Isto trará o Crosstabs dialog box.
3. Seleccione **ca4** da lista a esquerda e faça o click no  next to **Row(s)**:
4. Seleccione **ca2** da lista a esquerda e faça o click no  next to **Column(s)**:
5. Click no botão **Cells...**
Isto trará o Crosstabs:Cell Display dialog box
6. Na secção **Counts** click na próxima caixa para **Observed** para colocar um X nele, se não houver já um lá.
7. Na secção de **Percentages** click na próxima caixa para **Row** e **Column** para colocar um X lá.
8. Click no **Continue**
9. Click no **Paste**, e corra o comando no Syntax Editor.

The Crosstabs:Cell Display dialog box especifica que estatística você quer mostrar em cada célula da tabela—neste caso nos queríamos contagens, percentagens de linhas e percentagens de colunas. (A percentagem de linhas soma 100 entre todas as células numa linha, enquanto que a percentagem de colunas soma 100 entre todas as células numa coluna. Por defeito o comando **Crosstabs** simplesmente dá contagens.) A tabela produzida por este comando diz-lhe que há 21 chefes de agregados familiares femininos e que 6.1% do número total dos chefes dos agregados familiares são mulheres.

O comando **Compare Means** é algo similar ao **Crosstabs**, mas ele dá informação estatística acerca de variáveis contínuas. Ele mostra como a média e outras estatísticas para uma variável contínua diferem para os valores de uma ou mais variáveis categóricas. Outra forma de olhar para a relação entre **Crosstabs** e **Compare Means** é que, **Crosstabs** é uma forma de se obter **Frequencies**-tipo de resultado dividido por categorias de uma ou mais outras variáveis, enquanto **Compare Means** é uma forma de obter **Descriptives**-tipo de resultado dividido por categorias de uma ou mais outras variáveis.

Suponha que nós queremos saber como a idade dos respondentes varia pela sua relação com o chefe do agregado familiar. Se nós fizermos com o **Crosstabs** nós obteremos uma tabela com dezenas de células correspondentes a diferentes idades, o qual seria um formato impossível de usar. Em vez disso podemos usar o **Compare Means**.

1. Seleccione **Compare Means** do **Statistics** menu
2. Seleccione **Means...**
3. Seleccione **ca3** e faça click no  next to **Dependent List**:
4. Seleccione **ca2** e faça click no  next to **Independent List**:
5. Click no 
6. Corra o comando do **Syntax Editor**.

Este comando calculará médias de variáveis dependentes, o qual poderia ser normalmente uma variável contínua. A média poderá ser calculada separadamente para cada valor diferente da variável independente, o qual poderia ser uma variável categórica.

Deste resultado você encontra que a média de idades dos chefes dos agregados familiares é de 41.5 anos enquanto a média das suas esposas é de 33.2 anos.

Transformação de Dados

Depois de examinar os resultados da estatística descritiva você frequentemente deseja fazer transformação dos dados. Uma transformação dos dados é uma operação que toma as variáveis existentes e ou muda os seus valores numa forma sistemática ou usa os seus valores para calcular novas variáveis. O seguinte exemplo mostra uma transformação de dados comum; a conversão de uma variável contínua para uma variável categórica.







A informação recebida do comando **Means** é interessante, mas poderia ser também útil para ver a actual distribuição de idades por grupos ou categorias, assim nós podemos dizer, por exemplo, quantos chefes de agregados familiares são mais velhos do que 60. Desde que a variável idade, **ca3**, é contínua, nós não podemos fazer directamente—primeiro temos que transformá-la. Vamos supor que nós estamos interessados em 4 categorias: 0-10 anos, 11-19 anos, 20-60 anos, e mais de 60 anos de idade.

Para categorizar uma variável, você deve usar o comando **Recode**. Categorizando uma variável contínua torna a informação detalhada em informação mais geral. Se você deseja conservar a informação detalhada bem como a nova informação mais geral, você deve recodificar a variável em diferentes variáveis. Se você recodifica na mesma variável o valor original será perdido.

Neste ficheiro específico se você usa o comando **Recode Into Same Variable** para transformar **ca3**, **ca3** tomará o novo valor categórico assinado no **Recode** statement, e a idade original será perdida. Uma vez que nós queremos conservar a idade original e guardar os valores categorizados numa variável separada, nós devemos fazer **Recode Into A Different Variable**.

Vamos fazer o **Recode** numa nova variável chamada **age**.

1. Seleccione **Recode** do **Transform** menu
2. Seleccione **Into Different Variables...**
3. Seleccione **ca3** da lista a esquerda

4. Click no  next to Input Variable -> Output Variable: box
ca3 deve mover para o Input Variable->Output Variable: box e o nome do box mudará para Numeric Variable -> Output Variable.
5. Click uma vez na próxima caixa vazia para Name: na secção Output Variable para por o cursor lá.
6. Escreva **age** na caixa.
7. Click uma vez na próxima caixa vazia para Label: na secção Output Variable.
8. Escreva Age Group na caixa.
9. Click no  para ter o nome da variável e a etiqueta mudados tomar efeito.
10. Click no 
o Recode em Different Variables: Old e New Values dialog box aparecerá.
11. Na secção Old Value click no próximo círculo para Range: através
(O seu cursor deve estar na primeira caixa).
12. Escreva 0 na primeira caixa
13. Pressione <Tab> e escreva 10 na segunda caixa.
14. Pressione <Tab> duas vezes.
O seu cursor deverá estar na próxima caixa para Value: na secção New Value.
O Pedem pressionar a "Alt" no teclado mais a letra "l" para subir até a caixa Value:
15. Escreva 1 para o primeiro grupo de idade.
16. Click uma vez no 
17. Click na primeira caixa depois de Range: e repete passos 11 até 16 para recodificar idades 11 até 19 para 2 e idades 20 até 60 para 3.
18. Para recodificar idades 61 e acima de 4, click no próximo círculo para Range:
through highest
19. Entre 61 na caixa e repita os passos 14 até 16 usando 4 para o valor.
20. Click no 
21. Click no 
22. Selecciono o seguinte texto no Syntax Editor

```

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

```
23. Corra o comando

Recode muda os valores de **age** para os valores que queremos usar—1,2,3, e 4. Nós mudaremos para o Data Editor para ver que as mudanças foram feitas. Para mudar para o Data Editor window:

Desta vez nós usaremos um método diferente daquele que nós usamos antes.

1. Selecciono **c:\sample\q1a.sav** do **Window** menu.
2. Mova através do window com o scroll bars.

O formato estandardizado do SPSS para mostrar uma variável numérica inclui dois lugares decimais, o qual é inapropriado para uma variável que nós sabemos que terá sempre um valor inteiro. Para mudar o formato exposto de **age** para o mesmo formato como as nossas outras variáveis use **Data/Define Variable...**

1. Passe para o **Data Editor window** se você ainda não está lá.
2. Click uma vez na barra cinzenta onde o nome da variável **age** aparece.
3. Click no **Data** menu.
4. Selecciono **Define Variable...**
O Define Variable dialog box aparecerá para a variável **age**
5. Click no **Type...**
O Define Variable Type : age dialog box aparecerá.
6. Na próxima caixa para **Width**: escreva 1.
7. Na próxima caixa para **Decimal Places**: escreva 0
8. Se o próximo circulo para **Numeric** não está seleccionado, selecciono-o.
9. Click no **Continue**
Nós completaremos este procedimento no próximo passo dos passos.

Isto diz ao SPSS para Windows para mostrar a variável **age** com uma largura de 1 dígito e sem lugares decimais. Quando você faz o **Recode** da nova variável, ela não tem o *Value Labels*. O resultado estatístico do SPSS sempre inclui os nomes das variáveis que estão a ser analisadas, mas as vezes o nome duma variável não nos diz assim tanto como nós gostaríamos de saber. Como os nomes são limitados a oito caracteres, eles não poderão ser descritos o suficiente para nós nos lembrarmos da pergunta completa do questionário (exemplo da variável **ca4**). O nome também não nos diz quais os valores individuais a que se refere uma variável categorica. Para fazer o resultado mais compreensível acrescenta-se *Variable Labels* e *Value Labels*. Para evitar confusões e erros, você deve sempre acrescentar etiquetas para qualquer variável computada que você estará a salvar para posterior uso. O melhor momento para acrescentar etiquetas é imediatamente logo depois de você ter criado a nova variável, porque se você deixa para depois poderá esquecer. O comando recode facilita isto ao permitir -lhe adicionar o **Variable Label** quando você faz o recode. Par adicionar o **Value Labels** use os seguintes passos:

Você deveria estar ainda em Define Variable dialog box do ultimo passo dos passos.

1. Click no **Labels..**
O Define Labels: age dialog box aparecerá.
2. Se não há texto no **Variable Label:** box, entre o texto “Age Group” lá.
Devera estar lá o texto, uma vez que nós criamos a etiqueta quando nós fizemos o recode.
3. Vá para **Value:** na secção **Value Labels** e escreva 1
4. Pressione <Tab> uma vez e escreva 0 para 10
5. Click no **Add**
6. Repita os passos 3 até 5 usando a seguinte informação:

Value:	Value Label:
2	11 to 19
3	20 to 60
4	61 and older
7. Click no **Continue**
8. Click no **OK**
No Data window você verá que a idade é agora mostrada com um dígito singular.
9. Selecciono **Variables...** do **Utilities** menu.
10. Click no **age** para verificar as mudanças que acaba de fazer.
11. Click no **Close** quando você tiver terminado.

Esta nova variável ainda não é parte do ficheiro de dados armazenados no disco. Nós devemos salvar o ficheiro de modo para que esta variável seja incluída permanentemente no novo ficheiro de dados. É uma boa prática salvar o ficheiro com um nome diferente no caso de nós quisermos voltar atrás para termos a versão anterior do ficheiro. Por esta razão, nós usamos o comando **Save As** com o novo nome do ficheiro Q1A-AGE.SAV.

1. Tenha a certeza que o **Data Editor** window é o único em frente (o window activo).
2. Do **File** menu seleccione **Save As...**
O cursor deve estar na caixa sobre o File name: acima do Save as type: SPSS (.SAV) drop-down box. Typing while that area is highlighted will wipe out the current text.*
3. Escreva **q1a-age**
A extensão .sav será adicionada automaticamente.
4. Paste e corra o comando.

Agora cada vez que o ficheiro de dados Q1A-AGE.SAV é aberto, a variável **age** será incluída.

Você poderia querer analisar esta nova variável categórica usando o comando **Crosstabs** para determinar quantas pessoas em cada grupo de idades são chefes dos agregados familiares, esposas ou crianças.

1. Use **Statistics/Summarize/Crosstabs...** dos menus.
2. Use **age** dos Rows e **ca2** das Columns.
3. Confirme a selecção própria na secção **Cell**.
4. Paste o comando e corra-o.

Disto, você pode ver que 12% dos chefes dos agregados familiares têm 61anos de idade ou mais. Também, years of age or older. Also, that of the people 61 years or older, 83.7% are heads of households.

Compare a informação que você obteve deste análise de **Crosstabs** com a informação do comando **Compare Means** realizada antes no **ca3**. Para fazer isto, nós exploraremos a habilidade do SPSS para mudar entre o **Syntax**, **Output** e **Data** windows.

Para mudar para o **Output Navigator**:

1. Do **Window** menu seleccione **Output1 - SPSS Output Navigator**.
2. Mova para trás através do window com o scroll bars.
3. Encontra o **Crosstabs table** e compare com o **Compare Means table**.

Para mudar para o **Syntax Editor**:

1. Do **Window** menu seleccione o **Syntax1 - SPSS Syntax Editor**.
2. Mova através do window com o scroll bars.

Para mudar para o **Data Editor**:

1. Do **Window** menu seleccione o **C-q1a - SPSS Data Editor**.
2. Mova através do window com o scroll bars.

Por favor note que é também possível mudar de uma janela para outra fazendo o click no SPSS icons no Windows 95 taskbar, encontrado por defeito no botão do ecrã (o taskbar pode ser movido em alguns lados do ecrã).

Aplice o que você aprendeu sobre transformação de dados e estatística descritiva fazendo o seguinte exercício.

Exercício 1.2: Usando o Questionario do agregado familiar (no anexo mais os dados), descubra o número de agregados familiares que tem 1-4, 5-7, e mais do que 7 pessoas por agregado familiar. Uma forma para descobrir esta informação é criar a seguinte tabela.

- Sugestão:
- Use o ficheiro **C-HH.SAV**.
 - Recode **h1** em **hsize** usando o seguinte grupo: (1 thru 4) (5 thru 7) (8 thru Highest).
 - Adicione variavel label e value labels.
 - Corra **Crosstabs** nesta variavel por **district**.

Household size * DISTRICT Crosstabulation

			DISTRICT			Total
			MONAPO	RIBAUE	ANGOCHE	
Household size	1 to 4 members	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%
	5 to 7 members	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%
	8 or more members	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%	

Olhando para os resultados, poderá ver que por exemplo em Monapo 34,8% desde o membro 1 a 4 do agregado familiar(group 1) são encontrados dentro de Monapo e que 60,7% de todos os agregados familiares em Monapo tem 1 a 4 membros.

Antes de sair do SPSS para o Windows 7.5, devemos salvar os conteúdos do **Output Navigator**. O resultado do windows contém todos os comandos e os resultados desses mesmos comandos. Será de extrema ajuda manter este resultado num ficheiro para poder reve-lo mais tarde, imprimi-lo ou inclui-lo num documento.

1. Faça do Output Navigator o active window usando o seu "icon" no Windows 95 taskbar.
2. Do **File** menu selecione a opção **Save As...**
3. Entre o nome do ficheiro "session1".
A extensão ".spo" será automaticamente adicionada ao nome.
4. Carregue no **Save**

Para sair do SPSS para o Windows:

1. Do **File** menu selecione **Exit SPSS**
O dialog box dará a opção para salvar o conteúdo do C:\sample\c-hh.sav
2. Carregue no **No**,
Um dialog box irá comanda-lo para salvar os conteúdos do Syntax Editor para o Sintex1.
3. Carregue no **Save** dê um nome ao ficheiro, como por exemplo Module1.sps, dai o SPSS para Windows 7.5 será terminado.

SPSS para Windows SAMPLE SESSION
SECCÃO 2 - Restruturação dos Data Files - Table Lookup & Aggregation

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 7.5
3e Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Fevereiro de 1999

Alguns tipos de análise de dados iram necessitar de uma restruturação dos ficheiros de dados .Os dados dos quatro questionários ao agregado familiar, membro, produção e vendas estão separadamente organizados em quatro ficheiros de dados, porque os dados encontram-se em diferentes níveis. Os dados do agregado familiar estao no nivel mais generalizado ou no nivel mais alto. Os outros três contém dados mais detalhados, que são considerados como sendo do nivel mais baixo. Se não esta familiarizado com o conceito de níveis de dados, leia o manual "Computer Analysis of Survey Data -- File Organization for Multi-Level Data" por Chris Wolf, antes de continuar.

A análise que fizemos na Seccão 1 foi feita separadamente em cado nivel, usando so as variaveis de um simples file de cada vez. Contudo, outros tipos de análise necessitam uma combinação de dados de mais de um file. Olhemos para um exemplo.

Supanha que queiramos criar uma tabela de calorias por cada adulto equivalente ao que se produz por dia do principal produto alimentar. Mais ainda, queremos ver a variação por distrito e por caloria-produção num “quartil”.

TABLE:1 Produção de alimentos em calorias por adulto
equivalente por dia

Distritos	Producao de Calorias por Quartil			
	1	2	3	4
Monapo				
Ribaue				
Angoche				

Dados os dados na forma actual, numerosas transformações serão necessarias para a produção desta tablela. Este é um exemplo concreto das complicações que irá encontrar no mundo real de análise de dados. Toda esta secção, será dedicada a criação desta mesma tabela.

Para começar, vamos em primeiro lugar analisar alguns ficheiros que já temos e as respectivas variáveis que precisamos em cada tabela:

- C-Q1A.SAV: Este ficheiro contém dados sobre características do membro do agregado familiar. Esta a nível do membro do agregado. As variáveis **ca3** (idade) e **ca4** (sexo), são usadas neste exercício para computar o equivalente ao número de adulto por agregado.
- C-Q4.SAV: Este file contém variáveis ao nível *agregado-produto*, algumas das quais iremos precisar para este exercício. As variáveis que iremos precisar são as seguintes:
 - a. **prod**, contendo códigos para as culturas produzidas.
 - b. **p1a**, contendo códigos para a unidade em que a produção foi medida (saco de 100 kg, saco de 50 kg, etc).
 - c. **p1b**, o número de *unidades produzidas este ano*.

Nota que as unidades de produção não são unidades standartizadas. Por exemplo, um "saco de 100 kg", como se usa o termo em Moçambique, pesa 100 kg só quando está cheio de milho. Quando o saco é enchido por exemplo de mandioca, neste caso pesa menos de 100 kg. Por isso, precisamos de factores de conversão *para podermos converter* cada uma das unidades em que a produção foi realmente medida na nossa unidade standartizada. Que e o kilograma.

- CONVER.SAV: Este e um "*table-lookup file*" de factores de conversão especialmente criado para lidar com os problemas de unidades não standartizadas. Este ficheiro, contém cada combinação de produto-unidade. Por exemplo, o factor de conversão para um saco de 50 kg de arroz é diferente ao de um saco de 50 kg de algodão, que é também diferente ao factor de conversão de um saco de mandioca. As variáveis neste ficheiro são:
 - a. **prod**: Esta variavel refer-se ao produto.
 - b. **unit**: Esta variavel refer-se a unidade.
 - c. **conver**: Este e o factor de conversão e é igual ao número de quilogramas correspondentes a cada combinação de **prod** e **unit**.

Em baixo, um exemplo de dados do CONVER.SAV mostra-nos que, uma lata de 20 litros (**unit**=8) de arroz (**prod**=7) pesa 19 kg; um saco de 50 kg (**unit**=24) de arroz pesa 53 kg; uma lata de 20 litros de feijão (**prod**=30) pesa 17 kg; e um saco de 50 kg de feijão pesa 47 kg.

prod (Produto)	unit (unidade)	conver (factor de conversao)
...
...
7	8	19
7	24	53
...
...
30	8	17
30	24	47
...
...

- CALORIES.SAV: Este é também um ficheiro do tipo "table-lookup", criado para a conversão de quilogramas de alimento para calorias de alimento. Contém duas variáveis:
 - a. **prod**, é variável de produto
 - b. **calories**, é igual ao número de calorias por quilograma de cada alimento registado na pesquisa.

Com esta informação, podemos então pensar nos passos específicos que devemos seguir para criar a tabela necessária. Há logicamente três passos a seguir:

1. Precisamos de saber quantas calorias cada agregado produziu por o ano. Podemos criar um ficheiro com esta informação usando dados que guardamos em três lugares o "production file", o C-Q4.SAV, e dois "table-lookup files", CONVER.SAV e CALORIES.SAV.
2. Precisamos de saber quantos adultos equivalentes cada agregado contém. Podemos criar um ficheiro com esta informação, usando dados do "member file", C-Q1A.SAV.
3. Precisamos de obter os resultados dos passos 2 no mesmo ficheiro para podermos computar calorias produzidas por adulto equivalente por dia.

Passo 1: Generate a household level file containing the number of calories produced per household.

Ao seguirmos este passo, devemos tomar atenção a três promenores:

Primeiro, toda a produção e actualmente medida em unidades não standartizadas em que o peso é diferente para cada produto. Por isso devemos em primeiro lugar converter toda a produção em quilos.

Segundo, queremos saber as calorias produzidas por cada agregado, que não estão em quilogramas. Por isso, depois de converter toda a produção para quilogramas, devemos converter novamente para calorias.

Terceiro, um exame da Tabela IV mostra que temos dados para cada produto produzido pelo agregado. Mas nós queremos saber das calorias totais produzidas pelo agregado, e não as calorias totais de cada produto. Dai, depois de convertermos toda a produção para calorias, devemos para cada agregado, somar as calorias de cada produto para chegarmos ao agregado total.

Com estes pontos consolidados, vamos começar por abrir o C-Q4.SAV.

1. Selecione a opção **File/Open...**
2. Selecione o ficheiro c-q4.sav
3. Faça o **paste** e **run** .

Interesam-nos as sete amostras de culturas de alimentos (milho, feijão nhemba, feijão manteiga, mandioca, arroz, mapira, e amendoim). Olhando para o **prod** no questionário, vemos que estes produtos tem códigos que são, 47, 30, 31, 41, 6, 44 e 5. Para activar só estes casos, usamos a opção **Select Cases**. O **Select Cases** selecciona um *subset* de casos baseados num certo critério. O **Select Cases** pode filtrar os casos não seleccionados assim também como pode elimina-los. Se eliminar os casos não seleccionados, pode voltar aos ficheiros originais desde que não salve o ficheiro atual no mesmo nome igual ao do ficheiro original. Se filtrar o caso não seleccionado, (que e o que iremos fazer

ja que e o metodo mais seguro) tera a possibilidade de disfiltrar os dados que por seu lado irao ativar todos os casos do ficheiro.

1. Do **Data** menu escolha **Select Cases...**
Devera ver o Select Cases dialog box.
2. Escolha o botao radio perto do **If condition is satisfied**
3. Carrege **If...** sobre **If condition is satisfied**
4. Carrege **dentro** do box, a direita do **▶**, **nao** no proprio botao.
5. Enter o seguinte texto (sem carregar no return ou enter):

```
PROD = 47 | PROD = 30 | PROD = 31 | PROD = 41 | PROD = 6 | PROD = 44 | PROD = 5
```

*Os "/" são simbolos para a palavra OR. Estamos dizendo ao SPSS para selecionar todos os casos com **prod** iguais a 47 ou 30 ou 31...*
6. Carrege **Continue**
7. Escolha o botao radio perto do **Filtered**
8. Fassa o paste do comando
9. Escolha o texto no **Syntax Editor** de `USE ALL TO EXECUTE` e fassa o run do comando.

Agora, so os casos com este codico de product iram ser usados. Este subset de dados estara em efeito ate abrimos outro file ou o Select All cases (disfiltrar os casos).



Vamos em primeiro lugar converter toda a producao destes sete produtos alimentares em kilogramas. Para encontrar o factor de conversao apropriado para cada caso no production file(C-Q4.SAV), Precisamos de ver o produto(product) e a unidade(unit) no file CONVER.SAV. Iremos criar um novo file em que cada caso tem os dados do file production e uma variavel contendo o factor de conversao para a combinacao produto-unidade No SPSS para Windows, o comando que tem essa funcao e o **Merge Files /Add Variables**.



Os input files para um merge Tem que ser seleccionados pelos key variable(s) (variaveis chaves) (Aqueles variaveis que esta usando para combinar os casos). Ja que temos um unico factor de conversao para cada combinacao de produto-unidade, tanto a variavel do produto assim como a variavel de unidades sao variaveis chaves. O file CONVER.SAV ja esta escolhido por **prod** e **unit**. Devemos escolher os actuais files de producao da mesma maneira, tendo em conta o facto de que a variavel unidade se chama **p1a** e **nao unit**.


1. Do **Data** menu escolha **Sort Cases...**
O Sort Cases dialog box ira aparecer
2. Escolha **prod** e carrege no **▶**
3. Escolha **p1a** e carrege no **▶**
4. Fassa o paste e run do comando.


Agora os file estao prontos a serem emergidos. **Merge Files** precisa de pelo menos dois files como inputs. Neste caso, os dois files sao o CONVER.SAV e o working data file. O file criado por **Merge Files** tornar-se-a o working data file, substituindo o actual.

1. Do **Data** menu escolha o **Merge Files**, e em seguida o **Add Variables...**
Os Add Variables: Read File dialog box iram aparecer.
2. Escolha o filename `conver.sav`
3. Carrege **Open**

4. Escolha **p1a** da lista sobre o New Working Data File: e carregue 
5. Carregue 

*Permita-lhe a fazer o rename de **p1a** para **unit** para combinar com o file de conversao.*
6. Proximo de New Name: escreva **unit**
7. Carregue no 
8. Repare no box proximo do Match cases on key variables nos sorted files
9. Carregue no botao radio proximo do External file is keyed table
10. Escolha **prod** do Excluded Variables: list
11. Carregue no  proximo do Key Variables: (A baixo do direita)
12. Repita os passos 10 e 11 para **unit**
13. Fassa o paste do comando



Uma alerta apareceu dizendo-lhe que os data files devem ser seleccionados. Ja que os files estao seleccionados...
14. Carregue 

Um dialog box ira perguntar se pretende salvar o conteudo dos dados no window. Não queremos salva-lo, o novo file pode tomar o seu lugar, entao...
15. Carregue 
16. Escolha e fassa o run do comando. Tenha a certeza de incluir o EXECUTE.

Os anteriores passos vao indicar ao SPSS para Windows para combinar o working data file e o CONVER.SAV file, (usando CONVER.SAV como uma tabela) para adicionar a variavel unit para o nosso working data file. Ja que os key variables necessitam de ter os mesmos nomes em ambos os files, voltamos a nomear **p1a** (a variavel unit para o nosso working file) para **unit** (**p1a** ira permanecer **p1a** em c-q4.sav).

Os Key Variables são necessarios em qualquer *Merge* quando um dos files esta sendo utilizado como um keyed table. Os nosso key variables especifica que a procura seja por produto e unidade, porque temos diferentes factores de conversao para cada combinacao de produto-unidade. Se tivessemos usado so **prod**, SPSS esperaria que cada produto tivesse um so factor de conversao, com o mesmo valor independentemente da unidade de medida utilizada. Por exemplo, o SPSS esperaria que o factor de conversao fosse o mesmo para o arroz, tanto num saco de 100 kg ou numa lata de 20 liter. Isto estaria errado.

O novo working file produzido pelo merge contem actualmente a variavel do factor de conversao necessaria, **conver**. Para toda a combinacao de produto-unidade, **conver** e igual ao numero de kilogramas naquela unidade. Podemos agora calcular os kilogramas totais multiplicando o numero de unidades (**p1b**) por este factor de conversao.

1. Do Transform menu escolha **Compute...**
2. No Target Variable: entre **qprod_tt** (para kilogramas totais)
3. Marque **qprod_tt** aqui se quiser.
4. Da lista a direita escolha **p1b** e carregue no 
5. Type * ou escolha o botão no dialog box
6. Da lista a esquerda, escolha **conver** e carregue no 
7. Fassa o paste, escolha e faça o run do comando

A seguir, precisamos de ver quantas calorias por quilograma cada produto contém. Esta informação está no ficheiro 'table-lookup' CALORIES.SAV. O ficheiro possui duas variáveis—produto e número de calorias por quilograma. A variável chave é produto. Para adicionar a variável 'calorie-conversion' para o ficheiro de dados precisamos de fazer uma outra junção 'merge' com o ficheiro table lookup. Neste momento a variável chave deve ser somente produto. O ficheiro de dados deve estar já sorteado pela variável produto (veja na junção anterior 'merge'), pelo que não precisamos de sortear novamente.

1. Do menu **Data** seleciona **Merge Files** e a seguir **Add Variables...**
2. Seleciona o ficheiro calories.sav, **Continue**
3. Check the **Match cases...** box
4. Check the **External file is keyed table** box
5. Coloca **prod** no **Key Variables:** box
6. Usa o comando Paste
7. Limpa os avisos 'the warnings' quando necessário
8. Escolha e faça o run do comando

O novo ficheiro, produzido pela junção, possui a variável calorias, **calories**. Podemos agora fazer a computação da quantidade total de calorias produzidas.

1. Usa **Transform/Compute...**
2. Usa **kprod_tt** como **Target Variable:** (para calorias totais)
3. Entra a equação **qprod_tt * calories**
4. Faça o paste, escolha e faça o run do comando

Isto produz um ficheiro de dados com total de calorias produzidas por cada agregado familiar. Agora precisamos de saber quantas calorias foram produzidas pelo agregado familiar em todos os produtos alimentares combinados. Para isto, precisamos de somar em cada agregado familiar os valores de **kprod_tt** de todas as culturas alimentares produzidas pelo agregado. Noutras palavras, precisamos de criar um novo nível de ficheiro do agregado a partir do actual nível do ficheiro produto com um único caso por agregado.

Para criar um novo nível de ficheiro do agregado, usamos **Aggregate**. **Aggregate** irá criar um novo ficheiro de dados com um caso por cada agregado e com **kprod_tt** somado em cada. Serve-se sempre deste ficheiro, como ficheiro a ser agregado. Temos já aberto o ficheiro sobre a produção, então estamos prontos a agregar.

1. Do menu **Data** seleciona **Aggregate...**
Aggregate Data window irá aparecer.
2. Seleciona **district**, **vil**, e **hh**, respectivamente, para **Break Variable(s):**
3. Seleciona **kprod_tt** como **Aggregate Variable(s):**
4. Pressiona **Name & Label...**
5. Muda o nome kprod__1 para **kprod_tt**
6. Entra as seguintes legendas: **Calories Produced in Staple Foods**
7. Pressiona **Continue**
8. Pressiona **Function...**
9. Seleciona **Sum of values** e seleciona **Continue**

10. Selecciona **Replace working data file**
11. Cole do comando
12. Pressiona **No** para não guardar o conteúdo no data window NewData
13. Run do comando. Se tivéssemos seleccionado **Create new data file** em vez de **Replace working data file**, o novo ficheiro resultante da agregação teria sido guardado no disco e não estaria a ser o nosso ficheiro em uso. Teríamos que abri-lo para termos o acesso.

As variáveis 'The Break Variable(s)' especificam as variáveis a serem utilizadas para a combinação de casos no ficheiro da agregação. Qualquer caso a partir do ficheiro original que tenha valores idênticos para todas 'break variables' irá ser combinado num único caso no ficheiro resultante da agregação. Queremos agregar ficheiros para ter um caso por agregado familiar. Temos as variáveis que identificam os agregados familiares no nosso inquérito,—**district, vil, e hh**.

Aggregate Variable(s) cria uma nova variável **kprod_tt**, que calculamos somando **kprod_tt**, total de calorias produzidas, para cada agregado familiar. As únicas variáveis contidas no ficheiro resultante da agregação são as 'break variables' e uma nova variável de agregação criada (exemplo. **kprod_tt**).

O novo ficheiro em uso contém o que queremos, número total de calorias produzidas por agregado familiar. Para estar seguro de que esta variável existe, faça **Descriptives** na variável **kprod_tt**. Irás encontrar que o valor médio de calorias produzidas por agregado familiar por ano é 4,483,964.7.

Guarda este ficheiro usando o comando **Save As...**

1. Activa **Data Editor window**.
2. Usa **Save As...** a partir do menu **File**
3. De nome ao ficheiro **hh-file1**
4. Fassa o paste, escolha e fassa o run do comando.

Passo 2: Produzir um ficheiro do nível do agregado familiar contendo o número de adultos equivalente por cada agregado.

Os dados necessários para calcular adultos equivalente por cada agregado familiar estão no ficheiro dos membros, **C-Q1A.SAV**.

1. Selecciona **File/Open...**
2. Selecciona o nome do ficheiro **c-q1a.sav**
3. Fassa o paste, escolha e fassa o run do comando.

As regras que iremos usar pra calcular adultos equivalente para este inquérito são:

Homens, idade de 10 anos ou superior	= 1.0
Mulheres , idade entre 10 e 19 anos de idade	= 0.84
Mulheres, idade de 20 anos ou superior	= 0.72
Crianças, idade inferior a 10 anos	= 0.60

Isto quer dizer, em média, uma mulher entre 10 e 19 anos de idade necessita apenas de 84% de calorias do que um homem de 10 anos de idade ou superior, e que uma criança inferior a 10 anos necessita de apenas 60% de calorias que um homem com idade superior a 10 anos necessitaria. Assim por exemplo, uma criança (homem ou mulher) abaixo de 10 anos de idade e contada como .60 adulto equivalente. Por cada pessoa (caso) no ficheiro de membro precisamos de olhar para o sexo, **ca4**, e idade, **ca3**, para calcular o respectivo adulto equivalente.

Compute.../If... permite-nos fazer isto. A variável adulto equivalente a ser criada é **ae**.

1. Do menu **T**ransform seleciona **C**ompute...
O Compute Variable *window* irá aparecer.
2. Para **T**arget Variable: entra **ae**
3. Na Numeric Expression: box entra o **1**
4. Pressiona **I**f...
5. Seleciona o 'radio button' para Include if case satisfaz a condição:
6. Entra a proposicao **ca4 = 1 & ca3 >= 10**
7. Pressiona **C**ontinue
8. Use o comando Paste, mas não corra ainda 'don't run it yet'.
9. Repete os passos 1, e 3-8 substituindo a informação anterior pelo seguinte. Não és obrigado a usar os menus dentro SPSS. Uma vez tendo, o conjunto de comandos que passastes para Syntax editor torna-se muito mais fácil nesta fase simplesmente copiar e fazer o paste dos mesmos comandos dentro do Syntax editor do que mudar os nomes das variáveis. É rápido. Para aqueles que não conseguem realizar as manobras copy/paste dentro do Syntax editor, simplismente repetem os passos acima indicados.

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

10. Seleciona todos os **IF** statements e corra.

Para verificar se a nova variavel adulto equivalente, **ae**, foi calculada, faça uma tabela de frequencias.

1. Precisarás de selecionar **S**tatistics/**S**ummarize/**F**requencies...
2. Usa **ae**
3. Fassa o paste, ecolha e fassa o run do comando

Verás que existem 1524 casos no total. O ideal seria ter quatro valores representados na tabela —1, .72, .84, e .60— e não devia haver casos em falta. Podes ver que temos nove casos em falta. Isto indica-nos que no nosso ficheiro de dados está a faltar, em nove pessoas, a idade ou o sexo. Isto deveria ter sido identificado logo durante o processo de limpeza. Numa situação destas o investigador deve voltar aos questionários originais para resolver este problema. Como não podemos fazer isso agora, usaremos um método alternativo.

Se deixarmos estes valores em falta, o tamanho do nosso agregado familiar irá aparecer ligeiramente pequeno do que realmente é, o que irá distorcer os nossos resultados. Poderíamos evitar este problema eliminando os agregados dos nove indivíduos da nossa análise, mas não podemos usar a informação acerca da produção alimentar desses agregados. Em vez, iremos fazer uma suposição razoável acerca desses nove membros em falta. Sabemos que os valores de adulto equivalente variam entre o mais baixo .6 para as crianças e o mais alto 1.0 para homens adultos, que não é um intervalo grande. Para saber a média do valor de adulto equivalente na nossa amostra.

1. **Statistics/Sumarize/Descriptives...**
2. A variavel e **ae**
3. Não te esqueças de fazer o paste antes do comando run

Isto mostra que a média dos valores de **ae** para todos os indivíduos é .79, com o desvio padrão de apenas .17. Iremos assumir que os nove individuos com dados em falta para o sexo ou idade são todos indivíduos médios, e atribuí-los um valor de adulto equivalente de .79 . (Aviso: seja muito cuidadoso ao preencher os dados em falta, o uso descuidado desta técnica pode dar resultados incorretos. Estamos a usar isto como ilustração dos comandos do SPSS, não como recomendação a usar para compessar os dados em falta).

1. **Transform/Recode/Into Same Variables...**
Recode into Same Variables dialog box irá aparecer.
2. Mova **ae** a **V**ariables:
3. Pressione **Old and New Values...**
4. Selecciona **S**ystem-missing
5. Selecciona **V**alue: no **N**ew Value section e entra **.79** no box
6. Pressiona **A**dd
7. **C**ontinue
8. Fassa o paste, escolha e fassa o run do comando

Agora precisamos de calcular o número de adultos equivalente para cada agregado familiar. O ficheiro em uso está ao nível de membro, mas os valores que precisamos estão ao nível do agregado familiar. Novamente, usamos **Aggregate** para ir do nível de membros ao nível de agregado familiar. A nova variável **ae_tt** será calculada somando **ae** de todos os membros dum agregado familiar.

1. Do menu **D**ata selecciona **A**ggregate...
2. Move **d**istrict, **v**il, e **h**h para **B**reak Variable(s):
3. Move **ae** to **A**ggregate Variable(s):
4. Pressiona **N**ame & Label...
5. No **N**ame: box entra **ae_tt**
6. No **L**abel: box entra **Adult Equivalents**
7. **C**ontinue
8. **F**unction...
9. Selecciona **S**um of values
10. **C**ontinue
11. Selecciona **R**eplace working data file
12. Paste, Limpa os avisos e run.

Aggregate cria um novo ficheiro de trabalho. O novo ficheiro de dados está no nível do agregado familiar, com um caso por agregado. A variável **ae_tt** é o total de adultos equivalente por agregado. Para verificar se a variável foi criada faça **Descriptives** na variável **ae_tt**.

1. **Statistics/Sumarize/Descriptives...**
2. Fassa o paste, escolha e fassa o run do comando.

Deverias ter a média de adultos equivalente nos agregados como 3.49.

Isto completa o passo 2. Guarde o ficheiro como HH-FILE2.SAV.

1. Assegure que Data Editor window esta activo
2. **File/Save As...**
3. Filename hh-file2
4. Fassa o paste, escolha e fassa o run do comando.

Passo 3: Precisámos de juntar dois ficheiros criados no passo 1 e no passo 2 de modo a calcular as calorias produzidas por adulto equivalente.

Agora temos HH-FILE1.SAV com dados sobre produção-calorias e HH-FILE2.SAV com dados sobre adultos equivalente para todos os agregados familiares. Precisamos de combinar estes ficheiros caso por caso para ter os conjuntos de dados num único ficheiro. Para isto, usamos **Merge Files**, mas desta vez nenhum dos ficheiros é 'keyed tables'.

Vimos antes que variáveis chaves são necessárias para uma junção que inclua 'a keyed table lookup'. Quando estiveres a juntar dois ficheiros do mesmo nível, como estamos para fazer agora, pode não parecer importante incluir as variáveis chaves, mas é. As variáveis chaves determinam que casos devem ser combinados. *Nunca debes usar **Merge Files** sem as variáveis chave 'Key Variables' porque sem elas não há garantia que SPSS irá combinar os casos desejados.* O comando irá executar sem nenhum aviso ou mesmo sem uma mensagem de erro, mas os resultados podem estar incorrectos.

Nota: hh-file2.sav é ainda o ficheiro em uso

1. **Data/Merge Files/Add Variables...**
2. *Usa o ficheiro hh-file1.sav para Read File*
3. **Open**
4. *Selecciona Match cases nas variaveis chaves 'key variables...'*
5. *Selecciona Both files provide cases*
6. *Key Variables: sao **district**, **vil**, e **hh** respectivamente*
7. *Paste, limpa os avisos, selecciona and run.*

Merge Files criou um novo ficheiro de dados em uso. As duas variáveis necessárias para calcular as calorias produzidas por adulto equivalente estão agora no ficheiro em uso. Total de calorias produzidas (**kprod_tt**) por agregado familiar num ano dividido pelo total de adultos equivalente por agregado (**ae_tt**) dividido por 365 dias por ano dá-nos calorias produzidas por adulto equivalente por dia (**kprod_ae**).

1. **Transform/Compute...**
2. Target Variable: **kprod_ae**
3. **Type & Label...**
4. Label: **Calories produced per adult equivalent**
5. **Continue**
6. Numeric Expression: entra **kprod_tt/ae_tt/365**
7. Fassa o paste, escolha e fassa o run do comando

Antes de podermos produzir uma tabela que queremos , temos que criar mais uma variável, indicando em que quartil de produção-caloria cada agregado cai dentro do seu distrito. **Rank Cases** pode fazer isto. **Rank Cases** calcula uma nova variável para cada caso, mostrando onde esse caso se situa 'ranks' dentro do grupo de acordo com os valores das outras variáveis. Neste caso, queremos classificar cada agregado onde se situa 'ranks' dentro do seu distrito em termos de calorias produzidas por **ae**. Especificamente, por cada distrito, queremos quebrar os agregados em quatro grupos de igual tamanho (quartiles), do mais baixo para o mais alto na producao de calorias. Uma nova variável contendo os valores 1 a 4 irá indicar a que quartil cada agregado pertence.

1. **Transform/Rank Cases...**
2. Move **kprod_ae** to Variable(s):
3. Move **district** to By:
4. **Rank Types...**
5. Unselect Rank
6. Select Ntiles:
7. **Continue**
8. Fassa o paste, escolha e fassa o run do comando
9. Nota que o nome da nova variável no Output window; deve ser **NKPROD_A**

A primeira coisa a especificar é a variável contendo os valores a usar para o ranking—neste caso **kprod_ae**. A seguir precisamos da variável By para especificar as variáveis que definem os grupos—neste caso **district**. **Rank Cases** tem um número diferente de métodos de fazer o ranking. Estamos a usar o método mais simples—/NTILES(4) instrue SPSS for Windows para quebrar a variável em quartis. Deste comando o SPSS for Windows irá criar uma nova variável e atribuirá um nome que irá conter os rankings. Agora podemos usar o **Means** para obter os números para a nossa Tabela.

1. **Statistics/Compare Means/Means...**
2. Mova **kprod_ae** para Dependent List:
3. Mova **nkprod_a** para Independent list: layer 1 of 1
nkprod_a came from the Rank Cases procedure.
4. **Next**
5. Mova **district** para Independent List: layer 2 of 2
6. Fassa o paste, escolha e fassa o run do comando

Você pode notar que a média da população é 4014.5183 e a média para o segundo quartil em Ribaue é 2517.4551. O resultado de **Compare Means** dá-lhe os números necessários para a Tabela, mas contudo eles não estão formados exactamente como vimos na Tabela do começo desta secção. Na Secção 3 aprenderá os procedimentos para conseguir os mesmos resultados, mas numa Tabela com uma melhor apresentação.

Salve este ficheiro como HH-FILE3.SAV.

1. Active a janela de Data Editor
2. **File/Save As...**
3. O nome do ficheiro é hh-file3
4. Fassa o paste, escolha e fassa o run do comando

Agora pode salvar o conteúdo do Syntax Editor com um nome específico para o uso posterior. .

1. Active o Syntax Editor
2. **File/Save As...**
3. Salve o ficheiro com o nome `session2`
A extensão .sps ser-lhe-á adicionada automaticamente.

Este ficheiro já contém todos os comandos a partir do Syntax Editor. *Sempre que você fizer um trabalho substancial deverá salvar o conteúdo do Syntax Editor para um nome específico. Você já deve ter notado que , ao longo da sessão, nós poderíamos ter corridos comandos (podíamos ter feito o run) digitando o comando **OK** em vez de **Paste** . Fazendo o Paste dos comandos para o Syntax Editor e correndo-os posteriormente, em vez de os correr directamente através do **OK**, você fica com o registo de todo o seu trabalho e permite-lhe poder voltar a correr esses comandos para análises posteriores. Por exemplo, se você constatar que alguns dos dados foram mal entrados, você pode corrigí-los e voltar a correr facilmente os comandos e fazer as análises pretendidas com os dados já corrigidos, bastando, para isso, o uso do ficheiro Syntax previamente salvo, sem ser necessário ter que voltar a fazer de novo os comandos.*

Agora vamos ver como podemos reaver o ficheiro de comandos já por nós criado. Para sairmos do SPSS for Windows vamos fazer o seguinte:

1. **File/Exit SPSS**
O SPSS vai induzir-nos a salvar o conteúdo das janelas que ainda não foram salvas; neste caso, a janela do Output.
2. Salve o Output Navigator com o nome de `session2`

Recomece o SPSS for Windows. Para abrir os nossos comandos, vamos fazer o seguinte:

1. **File/Open...**
2. Selecciona Syntax(*.sps) em Files of type; seleccione o ficheiro session2.sps.
3. **OK**
A janela do Syntax !c:\sample\session2.sps será activada

Agora você pode reexecutar os mesmos comandos ou editá-los, segundo o seu desejo.

A sua SESSION2.SPS deve ser idêntica à seguinte:

```
GET
  FILE='C:\SAMPLE\C-Q4.SAV'.
EXECUTE .
USE ALL.
COMPUTE filter_$=(prod=47 or prod=30 or prod=31 or prod=41 or prod=6 or
  prod=44 or prod=5).
VARIABLE LABEL filter_$ 'prod=47 or 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 .
SORT CASES BY
  prod (A) pla (A) .
MATCH FILES /FILE=*
  /RENAME pla=unit
  /TABLE='C:\SAMPLE\CONVER.SAV'
  /BY prod unit.
EXECUTE.
COMPUTE qprod_tt = plb * conver .
EXECUTE .
MATCH FILES /FILE=*
  /TABLE='C:\SAMPLE\CALORIES.SAV'
  /BY prod.
EXECUTE.
COMPUTE kprod_tt = qprod_tt * calories .
EXECUTE .
AGGREGATE
  /OUTFILE=*
  /BREAK=district vil hh
  /kprod_tt 'Calories Produced in Staple Foods' = SUM(kprod_tt).
DESCRIPTIVES
  VARIABLES=kprod_tt
  /FORMAT=LABELS NOINDEX
  /STATISTICS=MEAN STDDEV MIN MAX
  /SORT=MEAN (A) .
SAVE OUTFILE='C:\SAMPLE\HH-FILE1.SAV'
  /COMPRESSED.
GET
  FILE='C:\SAMPLE-Q1A.SAV'.
EXECUTE .
IF (ca4 = 1 & ca3 >= 10) ae = 1 .
EXECUTE .
IF (ca4 = 2 & ca3 >= 10 & ca3 <= 19) ae = .84 .
EXECUTE .
IF (ca4 = 2 & ca3 >=20) ae = .72 .
EXECUTE .
IF (ca3 < 10) ae = .60 .
EXECUTE .
FREQUENCIES
  VARIABLES=ae .
DESCRIPTIVES
  VARIABLES=ae
  /FORMAT=LABELS NOINDEX
  /STATISTICS=MEAN STDDEV MIN MAX
  /SORT=MEAN (A) .
RECODE
  ae (SYSMIS=.79) .
EXECUTE .
AGGREGATE
  /OUTFILE=*
  /BREAK=district vil hh
  /ae_tt 'Adult Equivalents' = SUM(ae).
DESCRIPTIVES
  VARIABLES=ae_tt
  /FORMAT=LABELS NOINDEX
  /STATISTICS=MEAN STDDEV MIN MAX
  /SORT=MEAN (A) .
SAVE OUTFILE='C:\SAMPLE\HH-FILE2.SAV'
  /COMPRESSED.
MATCH FILES /FILE=*
  /FILE='C:\SAMPLE\HH-FILE1.SAV'
  /BY district vil hh.
EXECUTE.
COMPUTE kprod_ae = kprod_tt/ae_tt/365 .
VARIABLE LABELS kprod_ae 'Calories produced per adult equivalent' .
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 STDDEV COUNT
  /FORMAT= LABELS .
SAVE OUTFILE='C:\SAMPLE\HH-FILE3.SAV'
  /COMPRESSED.
```


Exercício 2.1: Produza um resultado semelhante usando as calorias retidas (produção menos vendas) em vez das calorias produzidas. Isto vai indicar as calorias retidas por adulto equivalente por dia, do total das 6 culturas alimentares. O resultado pode ser sub-dividido por distrito e por quartís de produção de calorias.

- Pontos:
- a. O procedimento é muito semelhante ao que foi feito no trabalho que acabamos de concluir.
 - b. Os dados sobre as vendas podem ser encontrados em `c-q5.sav`.
 - c. Procure o ficheiro que contém as variáveis necessárias sobre as quantidades de produção vendidas. Repare que os códigos dos produtos são os mesmos de `c-q4.sav`. Procure também as variáveis pelas quais fará o sort.
 - d. Reveja os comandos que foram usados para produzir a Tabela anterior, e verifique todos os passos para fazer as mudanças necessárias. Haverá mudanças nos códigos dos produtos, nos nomes do ficheiro e das variáveis.
 - e. A computação das calorias vendidas requer os mesmos passos que foram usados na computação das calorias produzidas. (Passo 1)
 - f. Faça o Merge do ficheiro recém criado, (o ficheiro que contém as calorias vendidas), com o ficheiro que contém as calorias produzidas, `hh-file3.sav`.
 - g. Tenha em mente que apenas 256 famílias venderam produtos, mas todas as 343 famílias produziram e retiveram calorias. Se a variável `calories-sold` está em falta, isso significa que a família não produziu alimentos, pelo que nesse caso a variável pode ter o código zero.
 - h. Crie a variável `calories retained = calories produced - calories sold`.
 - i. Sub-divida em quartís.
 - j. Use o comando **Compare Means** para ver as calorias retidas por **district** e por **quartile**.
 - k. Salve o ficheiro de dados.
 - l. Não é necessário salvar o Syntax Editor deste exercício.
 - m. Execute o ficheiro Syntax recém criado, faça o select all e corra-o, usando o comando `run`.

Um exemplo do resultado que se obtém é o seguinte:

- - Description of Subpopulations - -

Summaries of KRET_AE Calories Retained per adult equivalent
 By levels of NKRET_AE NTILES of KRET_AE by DISTRICT
 DISTRICT DISTRICT

Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			3044.2336	2370.1465	343
NKRET_AE	1		1098.8770	401.0378	84
DISTRICT	1	MONAPO	1148.0448	409.6144	27
DISTRICT	2	RIBAUE	1232.8030	350.2260	29
DISTRICT	3	ANGOCHE	912.7559	384.7468	28
NKRET_AE	2		2015.5753	297.9913	86
DISTRICT	1	MONAPO	2211.3833	205.7199	27
DISTRICT	2	RIBAUE	2145.8446	202.8158	30
DISTRICT	3	ANGOCHE	1698.5099	168.4997	29
NKRET_AE	3		2946.5741	547.1454	87
DISTRICT	1	MONAPO	3314.8568	477.1234	28
DISTRICT	2	RIBAUE	3126.3578	329.8936	30
DISTRICT	3	ANGOCHE	2405.0077	336.4856	29
NKRET_AE	4		6071.8027	2821.2709	86
DISTRICT	1	MONAPO	7619.1018	3557.1354	27
DISTRICT	2	RIBAUE	5759.0391	1649.5839	30
DISTRICT	3	ANGOCHE	4954.7625	2426.8245	29

Total Cases = 343

SPSS para Windows SAMPLE SESSION
SECCÃO 3 - Tabelas & Perguntas de Multipla Escolha

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 7.5
3e Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Fevereiro de 1999

Tabelas

Usando o **Tables** você pode calcular várias estatísticas e apresentá-las de formas diferentes sob o seu domínio. Ao contrário de outros procedimentos do SPSS for Windows, o procedimento **Tables** permite-lhe fazer o seguinte:

- ▶ seleccionar o arranjo que você gostaria de dar às variáveis e estatísticas, apresentando-as em linhas, colunas, por escalão, nestings e/ou concatenacoes
- ▶ manipular a estrutura, conteúdo e apresentação da Tabela
- ▶ ter múltiplas Tabelas na mesma mostra (concatenar), e criar sub-Tabelas múltiplas em qualquer dimensão.
- ▶ incluir percentagens flexíveis, especificando a sua base (o seu denominador), podendo somar 100% em linha, coluna, em sub-Tabelas, ou em toda a Tabela.
- ▶ apresentar mais de 60 caracteres para as etiquetas das variáveis e dos valores.

Vamos agora comparar as Tabelas com dados cruzados usandoo procedimento do **Crosstabs**, e o do **Tables**.

Abra o ficheiro member que criamos, o ficheiro Q1A-AGE.SAV, o qual contém a variável age.

1. **File/Open...**
2. Seleccione q1a-age.sav
3. Paste; seleccione e corra o ficheiro, através do comando run.

Primeiro vamos fazer uma simples Tabela de cruzamento de dados através do procedimento de **Crosstabs**.

1. **Statistcs/**Summarize/**Crosstabs...******
2. Mova **ca2** para **R**ow(s):
3. Mova **age** para **C**olumn(s)
4. **Cells...**
5. Seleccione **O**bserved na secção do Counts.
6. Seleccione **R**ow em Percentages
7. **Continue**
8. Fassa o paste, escolha e fassa o run do comando.

Produz a tabela seguinte:

RELATION TO HEAD * age group Crosstabulation

			age group				Total
			0 to 10	11 to 19	20 to 60	61 and older	
RELATION TO HEAD	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/DAUGHER		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%

Produz um tipo similar de tabela de tabulacao cruzada “crosstabulation” usando o menu **Basic Tables**.

1. **Statistics/Custom Tables/Basic Tables...**
2. Mova **ca2** para a opção **Down**
3. Mova **age** para a opção **Across**
4. Cole (paste) e corra (run) o programa.

		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/DAUGHER	503	184	31	
	MOTHER/FATHER			5	1
	OTHER RELATIVE	70	55	16	2

Esta é uma “**Basic Table**” usando os formatos iniciais do programa. Os cabeçalhos das linhas correspondem aos cabeçalhos dos valores da variável **ca2**. Os cabecarios das colunas sao os cabeçalhos dos valores que voce designou para a variable **age**. Se voce quiser adaptar a tabela paras as suas necessidades, o formato **Basic Tables** pode tornar-se muito mais complexo. Para um exemplo de uma tabela mais complexa...

1. Selecciona **ca2** em o Data Editor
2. **Data/Define Variable...**
3. **Labels...**
4. Delete the text in the Variable Label: box
5. **Continue**
6. **OK**
7. **Statistics/Custom Tables/General Tables...**
8. Mova **ca2** a Rows:
9. Mova **age** a Columns:
10. Selecciona **age**, escolha o **Edit Statistics...**
11. Selecciona **Count** em Cell Statistics:
12. Cambia Label: to **N**
13. Cambia Width: to **5**
14. **Change**
15. Selecciona Count Row % do Statistics:
16. Cambia the Label: to **%**
17. Cambia the Width: to **5**
18. **Add**
19. **Continue**
20. **Insert Total**
21. Selecciona **ageTotal**, e clica **Edit Statistics...**
22. Selecciona o radio button cerca do Custom Total Statistics
23. Selecciona **Count** em Statistics:
24. Cambia Label: a **N**
25. Cambia Width: a **5**
26. **Add**
27. **Continue**
28. Selecciona **ca2**, e clica **Insert Total**
29. **Format...**
30. Pone Empty Cell Appearance a Zero
31. **Continue**
32. **Titles...**
33. Em o Title box:
34. Escribe **Table 1: SPSS for Windows Sample Session**
35. Escribe **Age breakdown by Relation to Head**
36. Escribe **Source: Nampula family sector household survey, 1991.**
37. Em o Corner box: escribe **Relation to Head**
38. **Continue**
39. Cole (paste) e corra (run) o programa

Esta é a tabela produzida através daqueles comandos:

Table 1: SPSS for Windows Sample Session
Age breakdown by relation to Head
 Source: Nampula family sector hh survey, 1991.

		age group								Total
		0 to 10		11 to 19		20 to 60		61 and older		N
		N	%	N	%	N	%	N	%	
RELATION	HEAD	0	,0%	6	1,7%	296	86,3%	41	12,0%	343
TO HEAD	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

Nota: Pode mudar as propriedades da tabela, formatos, *pivotar* as tabelas,, etc.

Isto pode parecer muito esforço para produzir uma única tabela. Para uma única aplicação pode parecer fácil criar a tabela num processador de texto. Contudo se isto é alguma coisa que vai fazer repetidamente, o benefício de ter o **SPSS para Windows** a produzir a tabela “maravilhosa” de longe ultrapassa o esforço de criar a tabela. Isto torna-se importante com dados periodicos, tais como preços mensais, em que em cada mês a tabela deve ser actualizada.

A seguir é uma comparação da computação de médias usando a opção “**Compare Means**” usando “**Custom Tables**”, baseada no exemplo da seccao 2.

1. **File/Open...**
2. Seleccione hh-file3.sav
3. Cole (paste) e corra (run) o programa
4. **Statistics/Compare Means/Means...**
5. Mover **kprod_ae** a Dependent List:
6. Mover **nkprod_a** a Independent list: layer 1 of 1
7.

Next

8. Mover **district** a Independent List: layer 2 of 2
9. Cole (paste) e corra (run) o programa

Isto produz o seguinte resultado (a direita):

Esta é a informação que precisavamos para preencher os números da nossa tabela da seccao 2. Vamos usar “**Custom Tables**” para produzir resultados que se assemelhem a tabela que tinhamos em vista ao longo da secção 2. Para mais informações, vamos adicionar à tabela o Mínimo e o Máximo.

1. Mude os cabeçalho para as seguintes variáveis seleccionando as variáveis no **Data Editor** e usando a sequência de opções **Data/Define Variables/Labels...**

Variable	New Label
kprod_ae	delete label
district	District
nkprod_a	delete label

2. **Statistics/Custom Tables/Basic Tables...**
3. Mover **kprod_ae** a **Summaries**:
4. Mover **nkprod_a** a **Down**:
5. Mover **district** a **Across**:
6. **Statistics...**
7. Selecciona **Mean**, usa **MEAN**, **Format: ddd.dd**, **Width: 5** e **Decimals: 0**
8. **Add**
9. Selecciona **Maximum**, usa **MAX**, **Format: ddd.dd**, **Width: 5** e **Decimals: 0**
10. **Add**
11. Selecciona **Minimum**, usa the label **MIN**, **Format: ddd.dd**, **Width: 5** e **Decimals: 0**
12. **Add**
13. **Continue**
14. **Layout...**
15. Em os **Statistic Labels** section, selecciona **Down the Left side**, **Continue**
21. **Titles...**
22. Escribe em o **Title box: Table 1: Food Production in calories**
23. Escribe depois **per adult equivalent per day**
24. Escribe em o **Corner box: Production Quartile.**
25. **Continue**, Cole e corra o programa

Report

Calories produced per adult equivalent

1	MONAPO	Mean	1221,7281
		N	27
		Std. Deviation	416,1286
	RIBAUE	Mean	1484,0298
		N	29
		Std. Deviation	422,1161
	ANGOICHE	Mean	1272,0519
		N	28
		Std. Deviation	486,2593
	Total	Mean	1329,0592
		N	84
		Std. Deviation	452,2224
2	MONAPO	Mean	2494,8048
		N	27
		Std. Deviation	377,1214
	RIBAUE	Mean	2517,4551
		N	30
		Std. Deviation	366,0805
	ANGOICHE	Mean	2431,9673
		N	29
		Std. Deviation	296,8005
	Total	Mean	2481,5167
		N	86
		Std. Deviation	345,8224
3	MONAPO	Mean	3968,1419
		N	28
		Std. Deviation	621,3403
	RIBAUE	Mean	4000,8905
		N	30
		Std. Deviation	549,8340
	ANGOICHE	Mean	3640,3535
		N	29
		Std. Deviation	453,2870
	Total	Mean	3870,1717
		N	87
		Std. Deviation	562,9770
4	MONAPO	Mean	9150,0222
		N	27
		Std. Deviation	4686,2114
	RIBAUE	Mean	7520,2527
		N	30
		Std. Deviation	2158,8635
	ANGOICHE	Mean	8364,3191
		N	29
		Std. Deviation	4054,9027
	Total	Mean	8316,5516
		N	86
		Std. Deviation	3764,1698
Total	MONAPO	Mean	4206,4675
		N	109
		Std. Deviation	3813,5641
	RIBAUE	Mean	3900,7967
		N	119
		Std. Deviation	2559,3106
	ANGOICHE	Mean	3950,2610
		N	115
		Std. Deviation	3390,5114
	Total	Mean	4014,5183
		N	343
		Std. Deviation	3271,4011

Isto produz seguinte tabela:

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

Production Quartile		District		
		MONAPO	RIBAUE	ANGOCHE
1	Mean	(1222)	(1484)	(1272)
	Maximum	(1956)	(1938)	(1952)
	Minimum	(294)	(429)	(354)
2	Mean	(2495)	(2517)	(2432)
	Maximum	(3169)	(3120)	(2961)
	Minimum	(1973)	(2030)	(2024)
3	Mean	(3968)	(4001)	(3640)
	Maximum	(5067)	(4834)	(4563)
	Minimum	(3176)	(3141)	(2996)
4	Mean	(9150)	(7520)	(8364)
	Maximum	(28466)	(13124)	(20485)
	Minimum	(5107)	(4984)	(4692)

A opção “**Summaries**” especifica quais são as variáveis que são contínuas. Geralmente esta é a variável mais importante, aquela para a qual serão computadas as estatísticas. No exemplo acima, a variável **kprod_ae** é uma variável contínua para a qual os mínimos, médias, e máximos são calculados para cada categoria de **nkprod_a**.

A opção **Subgroups** determina como agrupar os casos em linhas e colunas da tabela. As variáveis designadas na opção **Subgroups** devem ser sempre variáveis de categorias (variáveis que tem um número finito de valores). A opção **Statistics** indica que estatísticas devem ser computadas usando as variáveis contínuas (Summaries).

Se o SPSS for Windows reporta um erro no menu “**Custom Table**”, tal muitas vezes tem a ver com o uso de uma variável incorrecta para a função solicitada ou com o facto de a largura da variável ser muito grande para a tabela. Se a tabela é obviamente errada à primeira aparição no ecrã, use o F3 para interromper os processamentos subsequentes. Se obtiver algum destes erros, verifique e certifique-se de que listou uma variável contínua na opção “**Summaries**”, e que a largura das variáveis e a largura das colunas são adequadas.

Uma maneira simples de imprimir a tabela que acabou de criar, é seleccionar a(s) tabela(s) no menu **Output Navigator** e imprimir.

1. Seleccionne o **Output Navigator**
2. Seleccionne depois a tabela que deseja imprimir
3. **File/Print.../OK**

Exercício 3.1: Produza uma tabela formatada de modo similar usando calorias retidas como fez no Exercício 2.1. **Inclua totais** (tera' que usar a opcao **General Tables**). A sua tabela deverá assemelhar-se a esta:

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

Production Quartile	District			Total
	MONAPO	RIBAUE	ANGOCHE	
1				
MEAN	1148	1233	913	1099
MAX	1806	1783	1391	1806
MIN	224	429	208	208
2				
MEAN	2211	2146	1699	2016
MAX	2544	2556	1936	2556
MIN	1807	1790	1396	1396
3				
MEAN	3315	3126	2405	2947
MAX	4303	3730	3055	4303
MIN	2555	2566	1984	1984
4				
MEAN	7619	5759	4955	6072
MAX	20874	9465	12675	20874
MIN	4360	3731	3064	3064

Perguntas de Múltipla Escolha

A informação escolhida através de uma análise de perguntas que permitem ao respondente seleccionar respostas de múltipla escolha. Uma simples variável de SPSS para Windows não pode registar as respostas para este tipo de perguntas adequadamente, porque a variável só pode ter um valor para cada caso. A solução é de registar cada resposta numa variável diferente. As respostas podem ser analisadas separadamente usando comandos já vistos (**Frequencies**, **Crosstables**), mas a nossa ideia é de analisar essas variáveis relacionadas no conjunto. Isto pode ser feito com agrupamento especial de **“Custom Tables”**. O SPSS para Windows permite dois métodos diferentes de agrupamento, para lidar com duas maneiras diferentes de fazer este tipo de perguntas.

Se a pergunta de pesquisa requer do respondente que “verifique tudo que aplica” de um conjunto de dez escolhas, dez variáveis deverão ser usadas para codificar as respostas. Haveria de precisar de criar uma variável para cada uma das dez possíveis respostas. Cada variável poderia ter um valor se a resposta marcada for (1) ou não marcada (0). Estas variáveis são chamadas de “multiple dichotomy”. Para mais detalhes consulte o **Manual de Tabelas**.

Noutros casos se a pergunta de pesquisa para o respondente for para “alistar até 4 escolhas” de um conjunto de 10, quatro variáveis devem ser usadas para codificar as respostas. O conjunto das possíveis respostas seria enumerada de 1 a 10, e o utilizador poderia entrar 4 daquelas variáveis. Estes valores seriam entrados para 4 das variáveis em SPSS para Windows. Estas variáveis são chamadas de “multiple response”. Para mais detalhes refira-se outra vez o **Manual de Tabelas**.

A pergunta 35 do questionário do agregado familiar é um exemplo de perguntas de múltipla escolha. Faz perguntas sobre as culturas cultivadas principalmente para a venda. Cada agregado é perguntado para especificar até três culturas principais que são codificadas em variáveis **h35a**, **h35b** e **h35c**. Os códigos das culturas são providenciados para cinco (5) das culturas mais comuns. A pergunta era deixada em aberto já que um código de seis (6) é permitido para outras culturas que são identificadas pelo seu nome.

Pelo facto da pergunta ser deixada em aberto isso permitiu adicionar mais categorias a essas variáveis. Isto é feito depois de toda a pesquisa estar completa. Alguém analisa as pesquisas manualmente e indica um código para cada uma das diferentes culturas correspondentes e saiu o número 6 para a variável “outros”. Os códigos e os cabeçalhos dos valores são indicados para cada cultura. Como pode ver com os seguintes comandos, onze diferentes culturas acabaram sendo codificadas para a pergunta 35.

A melhor maneira de analisar esta pergunta é de usar o **“Custom Tables”** com um conjunto de múltiplas respostas. Poderia usar a opção **“frequencies”** em cada uma das variáveis individualmente mas depois teria que somar os resultados a mão. O **“Customs Tables”** irá fazer o cálculo deste sumário estatístico para si se criar um grupo de variáveis usando o conjunto de **“Múltiplas Respostas”**.

Abra o ficheiro de dados do agregado familiar.

1. **File/Open...**
2. **Seleccione c-hh.sav**
3. **Cole, seleccione e corra o programa**

Para criar a tabela faça o seguinte:

1. **Statistics/Custom Tables/General Tables...**
2. **Mult Response Sets...**
3. Selecciona **h35a, h35b, h35c** e mova a **V**ariables in Set:
4. Selecciona **C**ategories em as **V**ariables Are coded As box
5. **N**ame: **Crops**
6. **L**abel: **Crops grown principally to be sold**
7. **A**dd, **S**ave
8. Mova **\$crops**, em **M**ult Response: a **R**ows:
9. Mova **district** a **C**olumns:
10. Selecciona **\$crops**
11. **I**nsert Total
12. Selecciona **\$cropsTotal**
13. Cambia o nome a **Total Cases** em **T**otal Label:
14. Cole e corra o programa

A opção **Múltipla Resposta** cria um grupo da variável **\$crops** das três variáveis **h35a, h35b e h35c**. Chame a opção **Insert Total** para produzir o sumário de estatísticas. Como foi discutido na secção anterior o formato para a tabela pode ser adaptado usando as opções no “Custom Tables”: **General** dialog box..

A tabela dos resultados deve ser apresentada da seguinte maneira:

		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

SPSS para Windows SAMPLE SESSION

ANEXO

Material de Treinamento para Curso de Curta Duração Desenho de Investigação, Processamento de Dados e Análise Utilizando SPSS para Windows 7.5 3e Edição

**Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Fevereiro de 1999**

Tres tabelas do relatório preliminar de pesquisa No. 3.
Um inquérito socio-económico do sector familiar na provincia de Nampula:
Observações metodológicas.
Direcção Nacional de Economia Agrária, MAP.

I. CARACTERÍSTICAS DA FAMÍLIA

- H1** 1. Quantas pessoas compoem este agregado familiar?
Agora, queríamos fazer algumas perguntas sobre estas pessoas.

Quadro IA: Características da Família

Nome	No. da Pessoa	Esta pessoa trabalha na machamba ou na casa ou fora 1=sim 2=não	Relação ao chefe 1=chefe 2=esposo(a) 3=filho(a) 4=pai/mãe 5=outro parente 6=outro	Idade	Sexo 1=m 2=f	Nível de escolaridade entrar o número da última classe completa 0=analfabeto 12=nível superior 98=nao estudou, mas sabe ler/escrever	Estado civil 1=casado(a)--monógamo 2=casado(a)--polígamo 3=solteiro(a) 4=viuvo(a) 5=divorciado(a) 6=mulher de emigrante (marido fora mais de seis meses)
	MEM	CA1	CA2	CA3	CA4	CA5	CA6
	1		Chefe				
	2						
	3						
	4						
	5						
	6						
	7						
	8						
	9						
	10						
	11						
	12						

V. VENDAS AGRÍCOLAS

Quadro V: Vendas de Produtos Agrícolas

No. da venda	Cultura	Quantidade vendida		Época de venda	Porque vendeu durante esta época?	Quem Comprou	Ponto de venda	Distância da machamb a	Porque vendeu a este comprador?	Valor da venda		Quem da família decidiu vender
		Unidade	Qt							meticais	Unidade	
	1=milho 2=feijão manteiga 3=feijões 4=mandioca 5=arroz 6=algodão 7=amendoim 8=castanha de caju 9=bebida de caju 10=cocos outros	Unidade 1=saco 100 2=saco 50 3=kilo 4=litro 5=lata 20 outro		1=sementeira (ago.- dez.) 2=epoca de fome (jan.- abril) 3=colheita desta campanha 4=várias épocas	1=precisava do dinheiro 2=havia comprador 3=havia bens de consumo 4=preço estava bom outro	1=lojista 2=armazenista 3=AGRICOM 4=ambulante 5=brigada 6=companhia outro	1=na machamba/ casa 2=aldeia 3=localidade 4=distrito 5=província outro	entrar os kilometros entra a machamb a e o ponto de venda	1=era o único 2=sempre vendeu a ele 3=melhor preço 4=transporta o produto 5=vende os bens de consumo outro		Unidade 1=preço unitário 2=valor total	1=marido 2=mulher outro
VEN	V1	V2A	V2	V3	V4	V5	V6	V7	V8	V9A	V9B	V10
1												
2												
3												
4												
5												
6												
7												
8												
9												