

SPSS para Windows *SAMPLE SESSION*

Analysis of Cross-Sectional Data

Análise dos Dados Transversais

**Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise**

Utilizando SPSS para Windows 10

4º Edição

Revisão feita por

Jean-Charles Le Vallée, Visiting Research Specialist, MSU para versão 7.5

levallee@pilot.msu.edu

Jan Low, Visiting Associate Professor, MSU para versão 10.0

Feito inicialmente por o Agricultural Economics Computer Service em MSU

Departamento de Economia Agrária, Universidade Estatal de Michigan

East Lansing, Michigan, EUA

Janeiro de 2000

Material de Treinamento para Curso de Curta Duração

Secção 0 -Estrutura de ficheiros para SPSS 10.0 para Windows (janelas (“windows”) dentro do programa - o *Syntax Editor*, o *Data Editor* e o *Output Navigator*)

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

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

Secção 3 - Tabelas & Perguntas de Múltipla Escolha

Secção 4 - Teste Para A Diferença de Médias

Secção 5 - Análise de Variância: *Oneway*

Secção 6 - Correlação

ANEXO

- Três tabelas do 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.

Agradecimentos

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.

O Que É Aquela Palavra??
Uma Lista Breve de Terminologia no Pacote SPSS10.0: Data Editor

English	Secção	Português
File		Ficheiro
Edit		Editar
View		Olhar (Perspectiva)
Data		Dados
Transform		Transformar
Analyze		Analisar
Graphs		Gráficos
Utilities		Utilidades
Window		Janela
Help		Apoio (Ajudar)
New	File	Novo
Open	File	Abrir
Open Database	File	Abrir um base de dados
Read Text Data	File	Ler os dados do texto (*.txt)
Save	File	Gravar/ Salvar
Save As	File	Gravar Como/ Salvar como
Display Data Info	File	Apresentar visualmente a Informação sobre os dados
Apply Data Dictionary	File	Aplicar o Dicionário dos Dados
Cache Data	File	Depositar os dados numa copia temporária que permite o SPSS de utilizar o processador sem re-ler os dados de novo
Print	File	Imprimir
Print Preview	File	Ver o formato final antes de Imprimir
Switch Server	File	Trocar o computador central do <i>Network</i> (Rede)
Stop Processor	File	Para o Processador (Implementação dos comandos escritos no Sintaxe)
Recently Used Data	File	Dados Recentemente Utilizados
Recently Used Files	File	Ficheiros Recentemente Utilizados

English	Secção	Português
Exit	File	Sair do Programa do SPSS
Undo	Edit	Des-fazer a última acção feita Regressar ao ponto antes de última acção
Redo	Edit	Re-fazer uma acção eliminada. Se decidir o UNDO não foi apropriada, utilize o REDO para refazer
Cut	Edit	Cortar
Copy	Edit	Copiar
Paste	Edit	Pôr dentro
Clear	Edit	Eliminar o que está dentro
Find	Edit	Buscar/Descobrir
Options	Edit	Opções
Status Bars	View	Barras de Estado Operacional
Toolbars	View	Barras de “Ferramentas” --- Acções diferentes
Fonts	View	Família de Tipos - Estilo e Tamanho do Texto
Grid Lines	View	Linhas de Grelha
Value Labels	View	Etiquetas de Valores
Define Dates	Data	Definir as Datas
Insert Variable	Data	Inserir Variável
Insert Case	Data	Inserir Caso
Go to Case	Data	Vai ao Caso Específico
Sort Cases	Data	Ordenar os Casos
Transpose	Data	Transpor
Merge Files	Data	Unir os Ficheiros
Aggregate	Data	Agregar/Juntar
Split File	Data	Separar o Ficheiro em partes
Select Cases	Data	Seleccionar Casos
Weight Cases	Data	Pesar Casos
Compute	Transform	Calcular

English	Secção	Português
Random Number Seed	Transform	Semente de Número aleatório: Valor utilizado para iniciar a geração dos números aleatórios
Count	Transform	Contar
Recode	Transform	Re-codificar
Categorize Variables	Transform	Dividir as Variáveis nas categorias determinadas
Rank Cases	Transform	Ordenar Grupos
Automatic Recode	Transform	Re-codificar automaticamente
Create Time Series	Transform	Criar a Séries Temporais
Replace Missing Values	Transform	Sobrepôr Valores Omissos
Run Pending Transforms	Transform	“Correr” ou Processar as Transformações Pendentes
Reports	Analyze	Relatórios
Descriptive Statistics	Analyze	Estatísticas Descritivas
Custom Tables	Analyze	Tabelas Segundos as Preferências do Utilizador
Compare Means	Analyze	Comparar Médias
General Linear Model	Analyze	Modelo Linear Geral
Correlate	Analyze	Correlacionar
Regression	Analyze	Regressão
Loglinear	Analyze	Log-linear
Classify	Analyze	Classificar
Data Reduction	Analyze	Redução dos Dados
Scale	Analyze	Escala
Nonparametric Tests	Analyze	Testes não-paramétricos
Time Series	Analyze	Séries Temporais
Survival	Analyze	Sobrevivência
Multiple Response	Analyze	Respostas Múltiplas
Gallery	Graphs	Galeria
Interactive	Graphs	Inter-activo
Bar	Graphs	Barra

English	Secção	Português
Line	Graphs	Linha
Area	Graphs	Área
Pie	Graphs	Torta
High-Low	Graphs	Alto-Baixo
Pareto	Graphs	Pareto
Control	Graphs	Controlo
Boxplot	Graphs	Gráficos utilizando Caixas
Error Bar	Graphs	Barra de erro
Scatter	Graphs	Espalhar
Histogram	Graphs	Histograma
P-P	Graphs	Gráficos de Probabilidade
Q-Q	Graphs	Gráficos de Quantis de distribuição (contra a distribuição dos exemplos diferentes)
Sequence	Graphs	Sequência
ROC Curve	Graphs	Curva de ROC: Avaliação dos esquemas de classificação
Time Series	Graphs	Séries Temporais
Variables	Utilities	Variáveis
File Info	Utilities	Informação sobre o ficheiro
Define Sets	Utilities	Definir Conjuntos
Use Sets	Utilities	Utilizar Conjunto
Auto New Case	Utilities	Auto Caso Novo
Run Script	Utilities	Correr o <i>Script</i> (Comandos escritos)
Menu Editor	Utilities	Editor de Menu
Minimize all Windows	Window	Minimizar todas as janelas
Topics	Help	Tópicos
Tutorial	Help	Tutoria (Instruções a seguir)
SPSS Home Page	Help	Página de Casa: Endereço no WEB do SPSS
Syntax Guide	Help	Guião de Sintaxe
Statistics Coach	Help	Tutoria em Estatística
About	Help	Sobre....
Register Product	Help	Inscrever o Produto SPSS 10.

SPSS para Windows “SAMPLE SESSION”
SECCÃO 0 - Estrutura de ficheiros para SPSS 10.0 para Windows
(Janelas de Data, Sintaxe “Syntax”, Resultados “Output”)

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 10.0
4º Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Janeiro de 2000

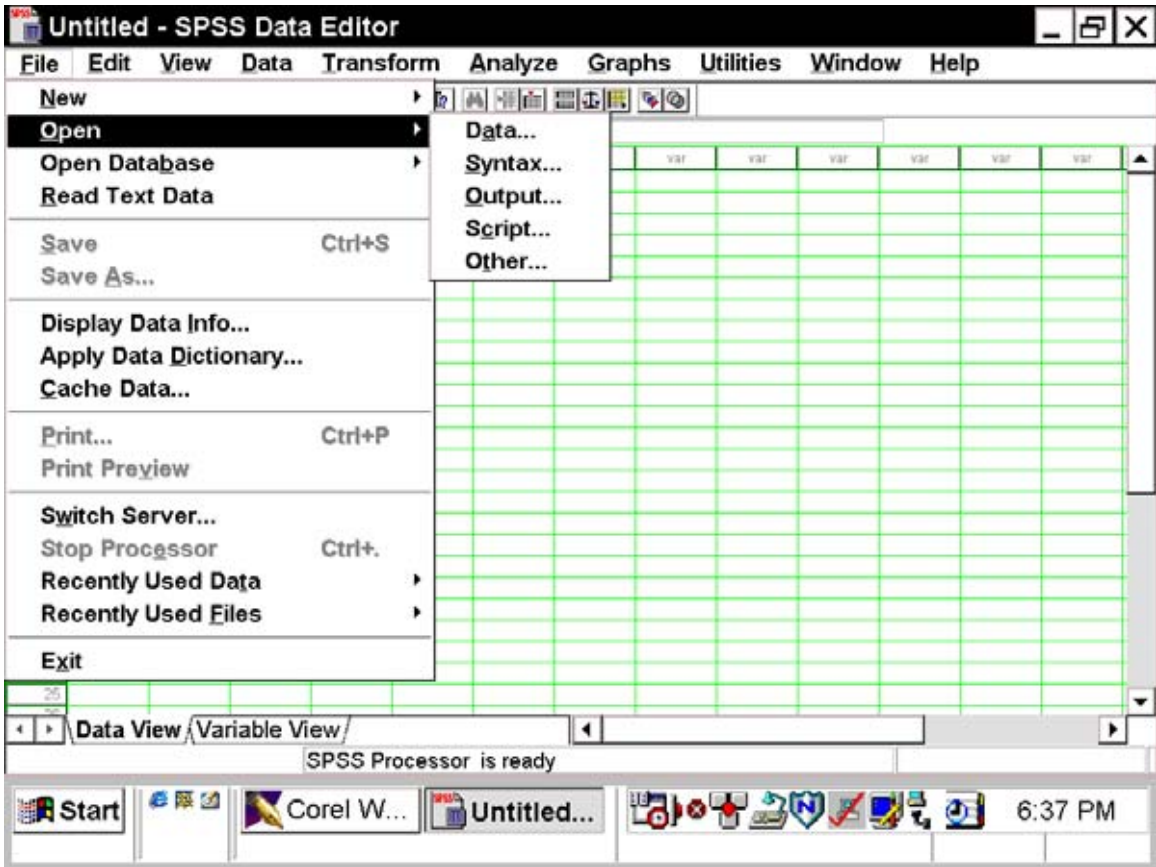
Este módulo introduz os conceitos básicos de níveis de dados e métodos de organização de dados. Também dá uma descrição breve sobre a estrutura dos ficheiros no SPSS 10.0 para Windows. É essencial ler este módulo antes de começar o Sample Session.

Ficheiros Utilizados no SPSS 10.0 para Windows

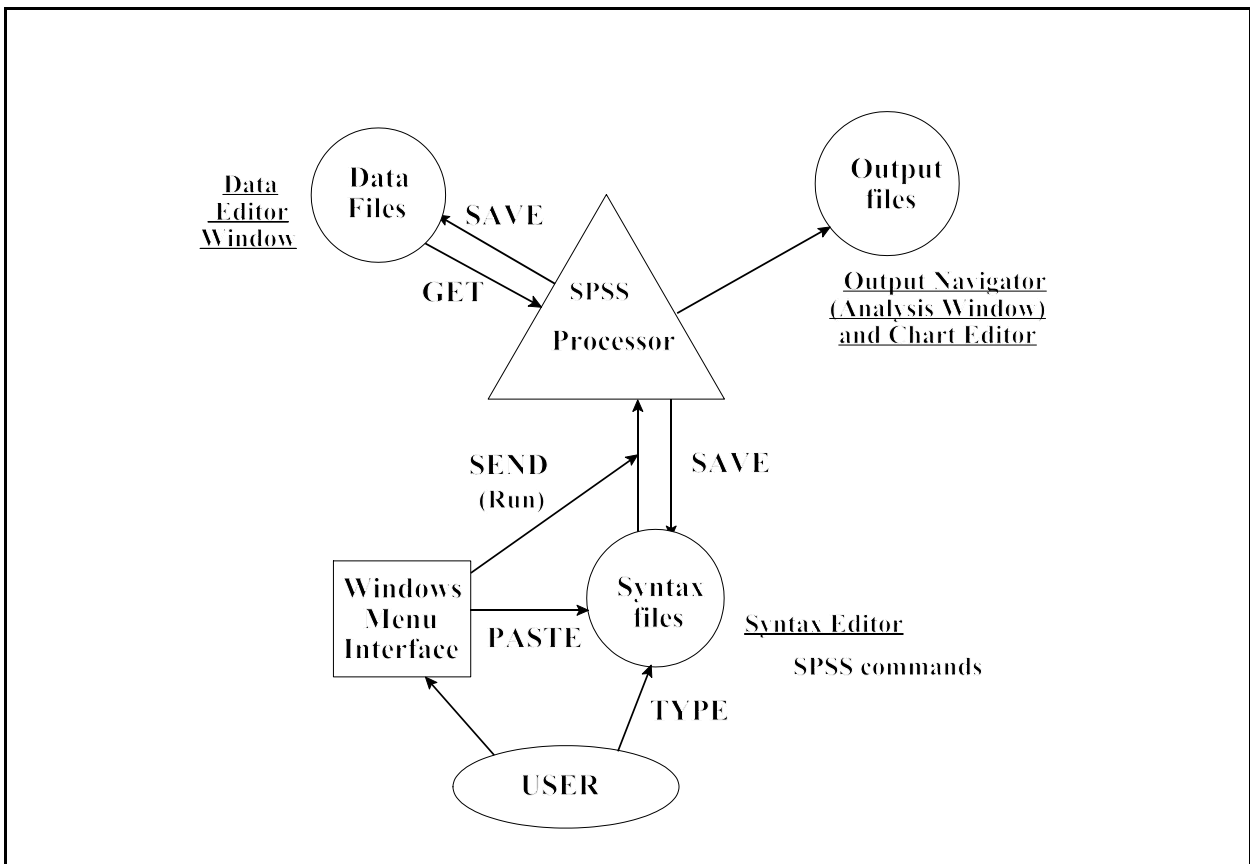
Ao utilizar o SPSS 10.0 para Windows da maneira ensinada neste “sample session”, está a usar quatro diferentes janelas (“windows”) dentro do programa - o *Syntax Editor*, o *Datas Editor* e o *Viewer*, e o *Scripts Facility*.

Quando abrir o SPSS10.0 na esquina esquerda em cima do quadro, tecele **File** (ficheiro), depois **Open** (abrir) e terá 5 opções de ficheiros de diferentes tipos a designar:

Data	ficheiros de dados	(Extensão *.sav)
Syntax	ficheiros de sintaxe ou comandos de programas	(Extensão *.sps)
Output	ficheiros de resultados obtidos	(Extensão *.spo)
Script	ficheiros avançados de programação em Sax BASIC que são criados cada vez que o Output é gerado	(Extensão *.sbs)



É importante conhecer o significado 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 ou sintaxe 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** (incluir) 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 10.0. para Windows para processamento,
- Para *write* ou guardar estes comandos a um ficheiro para uso futuro, e
- Para *retrieve* ou re-activar ficheiros de comandos que guardou anteriormente.

É 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 (Run...Current)** no SPSS 10.0 *Syntax Editor* para Windows **Toolbar**. Depois de pressionar o Run button, o computador envia os comandos para o processador, o qual le os comandos contidos no *Syntax Editor* e os executa. Depois de todos os comandos serem processados, o SPSS apresenta o *Viewer* 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.

É bom começar por compreender o sintaxe actual para gerar resultados mesmo que prefira utilizar sempre o sistema de *Paste* do menu. Para assegurar que vai ver os comandos criados quando usar o sistema de menu, faça o seguinte:

- ◆ Do sistema de menu escolha:
 - Edit**
 - Options...**
- ◆ Tecele o *Viewer*
- ◆ Seleccione *Display commands in the log*
- ◆ Tecele *OK* ou *Apply* (Aplicar)

Depois de completar com sucesso cada passo na sua análise (ou quando estiver pronto para terminar uma sessão de SPSS para Windows 10.0, embora não tivesse êxito 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** (re-abrir), 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 10.0 armazena os dados num ficheiro de dados ou *data file*. Para além dos valores em si, um *data file* contém informação tal como as etiquetas de variáveis (*variable labels*) e as etiquetas de valores actuais (*value labels*), informação sobre o formato das variáveis especificações, por exemplo sobre os valores omissos (*missing values*), etc. Para fazer qualquer análise de dados no SPSS para Windows 10.0, 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 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 Viewer

SPSS para Windows 10.0 automaticamente escreve no *Viewer* todos os resultados da execução dos comandos. Por exemplo, se corre um comando **frequency**, a tabela de frequência que se especifica é escrita no *Viewer*. A semelhança disto, se executas comandos que criam uma tabela ou gráfico, estes também serão enviados para o *Viewer*. Para guardar o conteúdo do *Viewer* para um ficheiro, active o *Viewer*, abra o **File menu** e seleccione **Save As...** Ao nomear o ficheiro, SPSS automaticamente colocará a extensão *.SPO*. É muito importante guardar o *output file*. Este lhe proporciona acesso aos resultados depois da sessão de SPSS para Windows 10.0 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 *Viewer* 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 *Viewer*, 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.

O Scripting Facility

A janela de *Script* é um meio-ambiente de programação feita na linguagem dos computadores que se chama Sax BASIC. A facilidade de *script* permite alguém automatizar tarefas e é útil quando tiver tarefas repetidas a fazer nos ficheiros dos dados distintos. Um ficheiro de *script* é criado cada vez que gera resultados no ficheiro do *output*. Depois estes ficheiros de *script* podem ser modificado pelo utilizador. As tarefas tipicamente automatizadas utilizando a facilidade de *script* são:

- ◆ Modificar segundo as preferências do utilizador a maneira como o **output** está a aparecer no *Viewer*.
- ◆ Abrir e Gravar os ficheiros
- ◆ Apresentar visualmente (*Display*) e manipular as caixas de diálogo (*Dialog boxes*)
- ◆ Efectivar (*Run*) as transformações dos dados e procedimentos estatísticos utilizando o sintaxe dos comandos

- ◆ Exportar diagramas e gráficos como ficheiros de gráficos nos formatos distintos

Alguns ficheiros de *scripts* já criados pela companhia de SPSS são incluídos. Quando criar os seus próprios *scripts* pode utilizar estes ficheiros como base e modificá-los segundo as suas preferências. O uso de *scripts* é tipicamente para os utilizadores de SPSS mais avançados que compreendem bem inglês. A maioria do nosso trabalho vai ser concentrado na utilização das janelas de **Data**, **Syntax**, e **Output**.

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 em SPSS 10.0 e SPSS 7.5 (*.LOG em SPSS/PC+ (DOS)).

Output files contêm resultados estatísticos, informação sobre dados, e tabelas, gráficos, etc. criados pelo processador de SPSS for Windows 10.0. Estes ficheiros não contêm dados. A nova extensão destes ficheiros é *.SPO (*.LIS em SPSS/PC+ (DOS) e *.LST em SPSS 6.1.3).

Datas files contêm dados, incluindo variáveis originais e variáveis criadas através dos comandos do SPSS para Windows 10.0. Abre-se estes ficheiros com o comando **Open...** Para SPSS 7.5 e 10.0., a extensão para estes ficheiros é *.SAV (*.SYS para SPSS/PC+ (DOS)).

Script files contêm comandos na linguagem de Sax BASIC, que incluem comandos para descrever objectos também para efectivar comandos estatísticos. Abrem-se estes ficheiros com o comando **Open...** Para SPSS 10.0., a extensão para estes ficheiros é *.SBS.

Localização dos Ficheiros

O Programa do SPSS 10.0 tem os directórios criados para os dados dentro do mesmo local que os ficheiros de programa. Devemos guardar todos os ficheiros de dados criados no sub-directório C:\Data\SPSS\..... para facilitar o “backup”-- gravação sistemática--- só dos ficheiros dos dados.

Para fazer isso, siga os seguintes comandos:

Utilizando o Windows **Explorer**, crie o sub-directório SPSS no directório de C:\Data. No C:\Data\SPSS, crie o sub-directório de C:\Data\SPSS\Sample. Vai receber os ficheiros do “Sample Session” para pôr neste directório.

Utilizando o Windows **Explorer**, vá para o c:\Program Files\SPSS10, corte o sub-directório de *Scripts* e transfira este sub-directório ao c:\Data\SPSS\Scripts.

Entre no programa SPSS10. No bar em cima, seleccione *Edit*, depois *Options* (Opções). Seleccione *Scripts*. Vai receber uma mensagem que indica um erro que o ficheiro não existe, mas deve continuar. Nas caixas abertas para corrigir o problema tecle o seguinte:

Global Procedure file:	C:\Data\SPSS\Scripts\global.sbs
Autoscript file:	C:\Data\SPSS\Scripts\autoscript.sbs

Tecle *OK* no fim do quadro para gravar esta mudança.

SPSS para Windows “SAMPLE SESSION”
ANÁLISE DOS DADOS TRANSVERSAIS (CROSS-SECTIONAL)
Departamento de Economia Agrária e Serviço de Informática
Michigan State University
Janeiro de 2000

Este “sample session” é desenhado para apresentar os comandos necessários para análises típicas de dados de inquéritos através do SPSS para Windows 10. 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 quilogramas	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:\Data\SPSS\Sample no se 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 o 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”
Secção 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 10.0
4º Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Janeiro de 2000

Inicia sua SPSS para Windows 10.0. Asegura-se de leer a secção 0 para esclarecer o conceito do Syntax window, onde faz o **Paste** dos comandos ou onde os escreve, do Viewer, 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 10.0 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 o 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:\Data\SPSS\SAMPLE\C-Q1A.SAV'.
No Syntax Editor.
4. Coloque o cursor em qualquer parte da linha do comando "GET" e *click* no **Run** button no Toolbar; depois *click* **Current**.

Notará que o comando GET FILE que você acaba de correr será escrito no Viewer.

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

Uma coisa que frequentemente queremos saber são as variáveis que um *data file* contém. 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 as etiquetas de variáveis (variable labels). Para fazer isto, faça o seguinte:

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

Este “display” mostra informação adicional, incluindo as etiquetas de valores (value labels) para variáveis como **district**, **vil**, **ca1**, **ca2**, **ca4**, **ca5**, **ca6**, e **univ**, o tipo de variável, a largura de apresentação (o display width) das variáveis em termos de número de caracteres reservado, o número de decimais (se Type for Numeric) que aparece depois do ponto final (.), e quaisquer valores omissos (missing values).

Click no botão do **Close** quando acabas

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

Seleccione o **Utilities** menu e seleccione **File Info**.
Este comando será executado imediatamente. O Viewer virá ser activo e terá uma lista de todas as variáveis e as suas respectivas definições.

Esta informação está incluída no Viewer. Pode ver o nome de cada variável, as suas etiquetas (*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 ficheiros dos dados (*data files*).

Se quiser ver a estrutura de cada variável, pode alternativamente ir ao visto de variável (*Variable View*) que está ao lado de *Data View* na esquina esquerda do quadro ao fundo do ecrã. Pode modificar as características de estrutura das variáveis também que como os dados aparecem no *Data View* facilmente dentro de *Variable View*. O exemplo para a primeira variável **DISTRICT** é mostrado na Tabela 1.1 na página seguinte com uma breve explicação de quais são as escolas dentro de cada coluna.

Quando quiser modificar o valor mostrado na célula, *click* na célula e uma pequena caixa vai aparecer na esquina direita da célula específica. *Click* aquela caixa. Nas colunas de **Width**, **Decimals**, e **Column** as setas para aumentar ou reduzir os números vão aparecer. Nas outras colunas, uma caixa de diálogo vai aparecer que inclui mais opções específicas.

Para a variável **district**, *click* na coluna de *Value*.

A caixa de diálogo aparece:

Value Labels

Value: tecle 4
 Value Label: tecle **NAMPULA**

OK Aceitar
 Cancel Cancelar
 Help Ajuda

Click **Add** (Adicionar) e uma nova etiqueta deve aparecer na lista de caixa principal. Alternativamente pode editar um valor existente (**Edit**) ou eliminar um valor (**Remove**). Não queremos guardar esta mudança; Assim, tecler **Cancel**.

Tabela 1.1. Estrutura Básica de *Variable View* no SPSS 10.0

Número de Variável	Name	Ty	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
Tradução	Nome	Tipo	Largura	Decimal	Etiqueta	Valores	Valores Omissos	Colunas	Alinhar	Medida
Explicação	Nome de variável	Númerico ou Alfanumérico (<i>String</i>)	Espaço requisit o para escrever todo	Espaços requisitos para os decimais depois de ponto final(.)	Descrição de variável	Descrição de conteúdo de valores das variáveis categóricas	Valores (exemplo- 99) que vai ser excluído de cálculos	Espaços reservados para apresentar o valor da variável no <i>Data View</i>	Onde aparece o valor na célula -- <i>Left</i> (esquerda) <i>Right</i> (direita) <i>Center</i> (no centro)	Nível de Medida da variável: ¹ <i>Scale</i> (Escala) <i>Nominal</i> (Nominal) <i>Ordinal</i> (Ordinal)
<i>Exemplo:</i>										
1	district	Numeric	1	0	DISTRICT	1= MONAPO 2= RIBAUE 3= ANGOCHE	None	8	Right	Scale

¹ Há três categorias de nível de medida:

Escala (*Scale*): Só valores numéricos que tipicamente são contínuos ou em intervalos (por exemplo, rendimento e idade)

Ordinal: Valores numéricos ou alfanuméricos (*String*) que consiste de categorias *com* uma ordenação intrínseca (por exemplo, 1= baixo, 2= médio, 3= alto)

Nominal: Valores numéricos ou alfanuméricos (*String*) que consiste de categorias *sem* uma ordenação intrínseca (por exemplo, 1= homem, 2= mulher)

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 valores omissos 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 você não tenha interesse neste dado, e o resultado indica que esta idade média é 91.3 anos, você 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 e é possível formar categorias. Por exemplo, Examine Tabela Apêndice I: variável **ca3**, idade (*age*), é uma variável contínua porque idade pode ter muitos diferentes valores. Variável **ca2**, relação ao chefe (*Relation to Head*), é uma variável categórica porque os seus valores se limitam ao intervalo 1-6.




Comece 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 o valor omissos (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 valores omissos 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** é contínua e que todos os outros são categóricas. Para correr um **Descriptives** na variável **ca3**, faça o seguinte:




1. Desde o **Analyze** menu seleccione **Descriptive Statistics...**
*Verá o **Descriptives** dialog box.*
2. Seleccione **ca3** da lista a mão esquerda e *click* no botão de .
ca3 irá para o **Variable(s):** caixa ao lado direito
3. *Click* no  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 Viewer virá ser a janela activa e os resultados do comando aparecerão lá. Verá que a média da ca3 é 21.34.

Quando ver o primeiro *output*, no SPSS Viewer, antes de continuar é importante gravar os resultados obtidos até agora. *Click* no File de menu, seleccione Save as.. (gravar como) e na caixa aberta, escreva o nome do ficheiro, por exemplo desc1 que vai automaticamente ter a extensão de .SPO.





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

O Viewer virá ser activo e os resultados estarão lá. Você 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 10.0, páginas Outro comando utilizado para produzir vários tipos de estatística descritivas é o comando Explore. Uma das estatísticas mais úteis que produz é os valores extremos (*outliers*). O comando Explore pode produzir grandes quantidades de output se for utilizado com os seus valores de *defaults*. Nos vamos limitar a quantidade dos resultados produzidos, seleccionando só estatísticas.

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

1. Desde o **Analyze** menu seleccione **Descriptive Statistics/Explore...**
2. Seleccione ca3 da lista a mão esquerda e *click* no  próximo ao **Dependent List**.
3. Na esquina sul-occidental do *dialog box* verá o **Display box**. *Click* no botão circular próximo a **Statistics**.
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 Outliers para colocar um "X" no quadrado.
*Verá que já existe um "X" no quadrado ao lado de **Descriptives**.*
6. *Click* no botão de .
*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 valores extremos (*extreme values or outliers*). Estes valores extremos (os *outliers*) serão identificados pelos seus números do caso *case numbers*. Faça referência às páginas do SPSS para Windows Base System User's Guide Release 10.0 para ver uma explicação do comando Explore.

É uma boa prática sempre gravar o seu ficheiro dos comandos no *Syntax Editor*. Frequentemente é necessário voltar a correr o mesmo programa depois de descobrir erros nos dados originais ou fazer pequenas mudanças no programa original e voltar a correr. No *Syntax Editor*, *click* no **File**, depois seleccione **Save as...** e grave este ficheiro dos comandos com o nome desc1.sps. É útil utilizar o mesmo nome para os ficheiros de comandos e o *output* correspondente para evitar confusão.

Aplique 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 variáveis categóricas.

Os resultados que deveria ver através destes comandos no Viewer 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, Tabela Apêndice I (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 sítio 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 **Descriptive Statistics** do **Analysis** 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 **Analyze** 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 **Paste** 
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** (Recodificar). 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** (Recodificar na mesma variável) 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** (Recodificar na variável diferente). Vamos fazer o **Recode** numa nova variável chamada **idade_gp**.

1. Seleccione **Recode** do **Transform** menu
2. Seleccione **Into Different Variables...**
3. Seleccione **ca3** da lista a esquerda
4. Click no  next to Input **V**ariable -> 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 **idade_gp** na caixa.
 7. Click uma vez na próxima caixa vazia para Label: na secção Output Variable.
 8. Escreva *Grupo Etário em Anos* na caixa.
 9. Click no **Change** (Mudar) para ter o nome da variável e a etiqueta mudados tomar efeito.
 10. Click no **Old and New Values...** (Valores Antigos e Novos)
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 “I” para subir até a caixa Value:
 15. Escreva 1 para o primeiro grupo de **idade_gp**.
 16. Click uma vez no **Add**
 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 **Continue**
 21. Click no **Paste**
 22. Seleccione 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 idade_gp .
  VARIABLE LABELS age 'Grupo Etário'.
EXECUTE .
```
 23. Corra o comando




Recode muda os valores de **idade_gp** 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. Seleccione **c:\data\spss\sample\q1a.sav** do **Window** menu.
2. Mova através do window com o scroll bars.





O formato padronizado 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 **idade_gp** para o mesmo formato como as nossas outras variáveis use **Variable View** na esquina esquerda ao fundo.


1. Passe para o *Data Editor* window se você ainda não está lá.
2. **Click uma vez na barra ao fundo onde aparece Variable View**
3. **Click** uma vez na caixa que na coluna do **Type** na linha 12 da variável **idade_gp**.

- Neste momento **Numeric** aparece dentro de caixa.
4. Click em cima de pequena caixa ao lado directa na caixa de Numeric 
 5. O *Variable Type* dialog box aparecerá para a variável **age**. Click no 
 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, seleccione-o.
 9. Click no 

Nós completaremos este procedimento no próximo passo dos passos.

Isto diz ao SPSS para Windows para mostrar a variável **idade_gp** com uma largura de 1 dígito e sem lugares decimais. Quando você faz o **Recode** da nova variável, ela não tem as etiquetas dos valores (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 **ca1**). 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 *Variable View* do ultimo passo dos passos.
1. Pode ver que a caixa na linha 12 de **idade_gp** na coluna de  já tem o etiqueta de variável definida durante o exercício anterior “Grupo Etário”
 2. *Se não há texto no Label: box, entre o texto “Grupo Etário” lá.* Devera estar lá o texto, uma vez que nós criamos a etiqueta quando nós fizemos o recode.
 3. Vá para a caixa na coluna de  e click
 4. Click em cima de pequena caixa ao lado directa na caixa de *None* 
 5. A caixa de diálogo *Value Labels* aparece. Tecele 1 dentro de primeira caixa vazia que se chama **Value**.
 6. Pressione <Tab> uma vez e escreva 0 até 10 na caixa de **Value Labels**
 7. Click no 
 8. Repita os passos 5 até 7 usando a seguinte informação:

Value:	Value Label:
2	11 até 19
3	20 até 60
4	61 e mais
 9. Click no 

No Data window você verá que a **idade_gp** é agora mostrada com um dígito singular.
 10. Volta a *Data View*. Seleccione **Variables...** do **Utilities** menu.

11. *Click* no **idade_gp** para verificar as mudanças que acaba de fazer.
12. *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 inclusa 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** (Gravar como...) de **File** de menu 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. Quando teclar em cima desta area qualquer texto anterior será apagado.
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 **idade_gp** será inclusa.

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 **Analyze/ Descriptive Statistics/ Crosstabs...** dos menus.
2. Use **idade_gp** dos **Rows** e **ca2** das **Columns**.
3. Confirme a selecção própria na secção **Cells:** Seleccione **ROW** e **COLUMN** na área de **Percentages..**
4. Paste o comando e corra-o.

Disto, você pode ver que 12% dos chefes dos agregados familiares têm 61 anos de idade ou mais. Também, no grupo etário de pessoas com 61 anos ou mais, 83.7% são chefes dos seus agregados familiares.

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 **Viewer**:

1. Do **Window** menu seleccione **Output1 - SPSS Viewer**.
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.

Deve notar que é também possível mudar de uma janela para outra fazendo o click no SPSS icons no Toolbar do Data Editor, encontrado por defeito no botão do ecrã (o taskbar pode ser movido em alguns lados do ecrã). De **View** de Menu, seleccione **Toolbars**; depois seleccione **Customize** (Fazer a sua própria selecção das coisas a incluir no *Toolbar*). A caixa de diálogo de *Customize Toolbars* aparecerá. No lado esquerdo a primeira escolha é do Toolbar de *File* (Ficheiro); ao lado direito na caixa de **Items**, aparecem botões diferentes. Para adicionar um botão no Toolbar de Data Editor, *click* em cima de botão desejado (por exemplo, *Open Data*) e não levando o seu dedo do rato (*mouse*), *drag* (empurrar) o item em cima de Toolbar cujos elementos estão a aparecer ao fundo da caixa de diálogo. Faça a mesma transferência de botões de *Open Syntax* e *Open Output*.

Aplique 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), descobre 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:
- a. Use o ficheiro C-HH.SAV.
 - b. Recode **h1** em **hsize** usando o seguinte grupo: (1 thru 4) (5 thru 7) (8 thru Highest).
 - c. Adicione variavel label e value labels.
 - d. 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 **Viewer**. 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 Viewer o active window usando o seu "icon" no Toolbar do Data Editor.
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

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 ,
Um dialog box irá comanda-lo para salvar os conteúdos do Syntax Editor para o Sintex1.
3. Carregue no dê um nome ao ficheiro, como por exemplo Module1.sps, dai o SPSS para Windows 10.0 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 10.0
4º Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Janeiro de 2000

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 estão no nível mais generalizado ou no nível mais alto. Os outros três contém dados mais detalhados, que são considerados como sendo do nível 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 Secção 1 foi feita separadamente em cada nível, usando so as variáveis 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.

Suponha 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	Produção de Calorias por Quartil			
	1	2	3	4
Monapo				
Ribaue				
Angoche				

Dados os dados na forma actual, numerosas transformações serão necessárias para a produção desta tabela. 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-PT.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 adulto equivalente por agregado.

- C-Q4-PT.SAV: Este ficheiro 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 padronizadas. 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 padronizada que é o quilograma.

- CONVER-PT.SAV: Este é um "*table-lookup file*" (ficheiro de tabela de referência) de factores de conversão especialmente criado para lidar com os problemas de unidades não padronizadas. 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 variável refer-se ao produto.
 - b. **unidade**: Esta variável refer-se a unidade.
 - c. **conver**: Este é o factor de conversão e é igual ao número de quilogramas correspondentes a cada combinação de **prod** e **unidade**.

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

- CALORIAS-PT.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 identificação do produto
 - b. **calorias**, 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 ficheiro de produção "production file", o C-Q4-PT.SAV, e dois "table-lookup files", CONVER-PT.SAV e CALORIAS-PT.SAV.
2. Precisamos de saber quantos adultos equivalentes cada agregado contém. Podemos criar um ficheiro com esta informação, usando dados do membros de agregado "member file", C-Q1A-PT.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: Criar um novo ficheiro a nível do agregado familiar que contém o número de calorias produzido pelo agregado.

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

Primeiro, toda a produção é actualmente medida em unidades não-padronizadas 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 quiilogramas, 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-PT.SAV.

1. Seleccione a opção **File/Open...**
2. Seleccione o ficheiro C-Q4-PT.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** (Seleccionar os Casos). 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 actual no mesmo nome igual ao do ficheiro original. Se filtrar o caso não seleccionado, (que e o que iremos fazer já que e o metodo mais seguro) terá a possibilidade de *disfiltrar* os dados que por seu lado irão activar todos os casos do ficheiro.

1. Do **Data** menu escolha **Select Cases...**
Deverá ver o **Select Cases dialog box** (caixa de diálogo).
2. Escolha o botão circular perto do **If condition is satisfied**
(Se a condição forá satisfeita)
3. Carregue **If...** sobre **If condition is satisfied**
4. Carregue **dentro** do box, a direita do , **não** no próprio botão.
5. Tecele 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 OU (*OR* em inglês). Estamos a dizer ao SPSS para seleccionar todos os casos com **prod** iguais a 47 ou 30 ou 31...
6. Carregue **Continue**
7. Escolha o botão circular perto do **Filtered** (Filtrado, quer dizer os casos **não**-seleccionados não são incluídos no procedimento actual, mais não são eliminados permanentemente)
8. Faça o **Paste** do comando
9. Escolha o texto no **Syntax Editor** de `USE ALL. até EXECUTE .` e faça o **Run..** do comando, seguido pelo **Selection..** (o texto marcado)








Agora, so os casos com este codigo de produto iram ser usados. Este sub-amostra de dados estará em efeito até abrimos outro ficheiro ou utilizarmos o **Select All cases** (anular o comando "*Filtered*" ou disfiltrar).

Vamos em primeiro lugar converter toda a produção destes sete produtos alimentares em quilogramas. Para encontrar o factor de conversão apropriado para cada caso no ficheiro de produção (C-Q4-PT.SAV), precisamos de ver as variáveis *produto* e *unidade* no file CONVER-PT.SAV. Iremos criar um novo ficheiro em que cada caso tem os dados do ficheiro de produção e uma variável contendo o factor de conversão para a combinacao produto-unidade. No SPSS para Windows, o comando que tem essa função e o **Merge Files /Add Variables** (Juntar os Ficheiros /Adicionar Variáveis).

Os ficheiros indicados para ser juntados (*merged*) têm que ser seleccionados pelos *key variable(s)* (variáveis chaves). As variáveis chaves são aquelas variáveis que está a usar para combinar os casos. Já que temos um único factor de conversão para cada combinação de produto-unidade, tanto a variável do produto assim como a variável de unidades são variáveis chaves. O ficheiro CONVER-PT.SAV já está escolhido por **prod** e **unidade**. Devemos escolher os actuais files de produção da mesma maneira, tendo em conta o facto de que a variável unidade se chama **p1a** e não **unidade**.

1. Do **Data** menu escolha **Sort Cases...**
O Sort Cases dialog box irá aparecer.
2. Escolha **prod** e carregue no 
3. Escolha **p1a** e carregue no 
4. Faça o **Paste** e **Run** do comando.

Agora os ficheiros estão prontos a serem juntados. **Merge Files** precisa de pelo menos dois ficheiros a ser combinados. Neste caso, os dois ficheiros são o CONVER-PT.SAV e o ficheiro que já está activo, o C-Q4-PT.SAV. O ficheiro criado por **Merge Files** tornar-se-a o novo ficheiro activo (*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 irão aparecer.
2. Escolha o filename CONVER-PT.SAV
3. Carregue 
4. Escolha **p1a** da lista sobre o **New Working Data File:** e carregue 
5. Carregue 
Permitir-lhe a fazer o rename de **p1a** para **unidade** para combinar com o ficheiro de conversão.
6. Próximo de **New Name:** escreva **unidade**
7. Carregue no 
8. *Click* na caixa próxima do Match cases on key variables in sorted files
(Juntar os casos baseados nas variáveis chaves nos ficheiros ordenados).
9. Carregue no botão circular próximo do **External file is keyed table**
(O ficheiro external (não activo neste momento) é o ficheiro de referência)
10. Escolha **prod** do **Excluded Variables List** (Lista de variáveis excluídas)
11. Carregue no  próximo do **Key Variables:** (A baixo do direita)
12. Repita os passos 10 e 11 para **unidade**
13. Faça o **Paste** do comando
Uma alerta apareceu dizendo-lhe que os ficheiros de dados devem ser ordenados. Já que os ficheiros estão ordenados com o **Sort...**
14. Carregue 
Um dialog box irá perguntar se pretende salvar o conteúdo dos dados no window. Não queremos salva-lo, o novo ficheiro pode tomar o seu lugar, então...
15. Carregue 
16. Escolha e faça o **Run** do comando. Tenha a certeza de incluir o EXECUTE.

Os anteriores passos vão indicar ao SPSS para Windows para combinar o ficheiro activo (*o working data file*) e o ficheiro de CONVER-PT.SAV, (usando CONVER-PT.SAV como uma tabela de referência) para adicionar a variável **conver** para o nosso ficheiro activo (*working data file*). Já que as variáveis chaves necessitam de ter os mesmos nomes em ambos os ficheiros, voltamos a nomear **p1a** (a variável unidade para o nosso ficheiro activo) para **unidade** (**p1a** irá permanecer **p1a** em C-Q4-PT.SAV).

As variáveis chaves são necessárias em qualquer *Merge* quando um dos ficheiros está a ser utilizado como uma tabela de referência (*keyed table*). As nossas variáveis chaves especificam que a procura seja por produto e unidade, porque temos diferentes factores de conversão para



cada combinação de produto-unidade. Se tivéssemos usado so **prod**, SPSS esperaria que cada produto tivesse só **um** factor de conversão, com o mesmo valor independentemente da unidade de medida utilizada. Por exemplo, o SPSS esperaria que o factor de conversão fosse o mesmo para o arroz, tanto num saco de 100 kg ou numa lata de 20 liter. Isto estaria errado.

O novo ficheiro activo (*working file*) produzido pelo *Merge* contém actualmente a variável do factor de conversao necessaria, **conver**. Para toda a combinação de produto-unidade, **conver** e igual ao numero de kilogramas naquela unidade.

Sempre é necessário verificar que o **Merge** fosse completado com êxito. Se não, todo o seu trabalho seguinte vai ter grandes erros. Volta a *Data View* dentro de Data Editor e verifica os valores de último coluna (**conver**) quando tiver os valores seguintes para o **prod** e a **unidade**:

prod (Produto)	unidade (unidade)	conver (factor de conversão)
5	7	1
5	8	15
30	8	17
30	24	47

Podemos agora calcular os quilogramas totais multiplicando o número de unidades (**p1b**) por este factor de conversão.

1. Do **Transform** menu escolha **Compute...**
2. No Target Variable: entre **qprod_tt** (para kilogramas totais)
3. **Type&Label**
3. *Click* no **Type&Label**
4. Tecla *Quilogramas Totais Produzidos* na caixa vazia ao lado de **Label** (etiqueta).
5. Carrega no **Continue**
4. Da lista a direita escolha **p1b** e carrega no 
5. Na caixa de **Numeric Expression** (Fórmula matemática), **p1b** aparecerá
5. Tecla * ou *click* em cima de o botão de * na caixa de diálogo com o calculador
6. Da lista a esquerda, escolha **conver** e carrega no 
7. Faça 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 de referência (*table-lookup*) CALORIAS-PT.SAV. O ficheiro possui duas variáveis—**produto** e número de **calorias** por quilograma. A variável chave é produto. Para adicionar a variável **calorias** (conversão de calorias) para o ficheiro de dados precisamos de fazer uma outra junção ‘*Merge*’ com o ficheiro de referência. 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 ordenar novamente.

1. Do menu **Data** selecciona **Merge Files** e a seguir **Add Variables...**
2. Selecciona o ficheiro CALORIAS-PT.SAV, carrega **Open**
3. *Click* na caixa de **Match cases...** (Juntar casos)
4. *Click* na caixa de **External file is keyed table** (Ficheiro external é a tabela de

- referência)
5. Coloca **prod** na caixa de **Key Variables**: (Variáveis chaves)
 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, **calorias**. Podemos agora fazer a computação da quantidade total de calorias produzidas.

1. Usa **Transform/Compute...**
2. Usa **cprod_tt** como **Target Variable**: (para calorias totais)

Type&Label

3. *Click* no
4. Tecla *Calorias Totais Produzidas* na caixa vazia ao lado de **Label** (etiqueta).
5. Carrega no **Continue**

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 **cprod_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** (agregar). **Aggregate** irá criar um novo ficheiro de dados com um *único* caso por cada agregado e com **cprod_tt** somado em cada. Serve-se sempre deste ficheiro activo (*the working file*) como ficheiro a ser agregado. Temos já aberto o ficheiro sobre a produção, então estamos prontos a agregar.

1. Do menu **Data** selecciona **Aggregate...**
Aggregate Data window irá aparecer.
2. Selecciona **distrito, ald, e af**, respectivamente, para **Break Variable(s)**:
3. Selecciona **cprod_tt** como **Aggregate Variable(s)**:
4. Pressione **Name & Label...** (Nome e Etiqueta)
5. Mude o nome **cprod__1** para **cprod_tt**
6. Entra as seguintes legendas!: **Calorias Produzidas nos Alimentos Básicos**
7. Pressione **Continue**
8. Pressione **Function...**
9. Selecciona **Sum of values** e selecciona **Continue**
10. Selecciona **Replace working data file** (Sobrepôr o ficheiro activo)
11. *Click* **OK**
12. Pressione **No** para não guardar o conteúdo no data window NewData
13. **Run** do comando. Se tivessémos 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 abrí-lo para termos o acesso.

As variáveis ‘*The Break Variable(s)*’ (as variáveis para partir-se) 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,—**distrito**, **ald**, e **af**.

Aggregate Variable(s) cria uma nova variável **cprod_tt**, que calculamos somando **cprod_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. **cprod_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 **cprod_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. Dá nome ao ficheiro de **AF-FICH1**
4. Faça o **Paste**, escolha e faça 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-PT.SAV**.

1. Seleccione **File/Open...**
2. Seleccione o nome do ficheiro **C-Q1A-PT.sav**
3. Faça o **Paste**, escolha e faça o **Run** do comando.

As regras que iremos usar para 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 necessitara. Assim por exemplo, uma criança (homem ou mulher) abaixo de 10 anos de idade e contada como 0,60 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 **Transform** seleccione **Compute...**
O **Compute Variable window** irá aparecer.
2. Para **Target Variable**: entre **ae**
3. **Click** no **Type&Label**
4. Tecla **Adulto Equivalente** na caixa vazia ao lado de **Label** (etiqueta).
4. Na caixa de **Numeric Expression** entra o **1**
5. Pressione **If...**
6. Seleccione o botão circular para **Include if case satisfies the condition**
(Incluir se o caso satisfazer a condição)
7. Entre as condições: **ca4 = 1 & ca3 >= 10**
8. Pressione **Continue**
9. Use o comando **Paste**, mas **não** corra ainda *'don't run it yet'*.
10. Repete os passos 1, e 3-8 substituindo a informação anterior pelo seguinte. Não é obrigatório usar os menus dentro SPSS. Uma vez que tem 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, simplesmente repetem os passos acima indicados.

Numeric Expression	If... Statement
0,84	ca4 = 2 & ca3 >= 10 & ca3 <= 19
0,72	ca4 = 2 & ca3 >= 20
0,60	ca3 < 10

DITA: Parece que quando está a teclar decimais no Sintaxe, utilize pontos . mais nas caixas de diálogo pode utilizar virgulas , segundo os seus *settings* no sistema de operação de Windows.

Pode verificar dentro de ficheiro de sintaxe os seguintes comandos:

```
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 'ADULTO EQUIVALENTE' .
EXECUTE .
```

10. Põe o cursor no início de primeiro acontecimento de frase de IF e dentro de **Run**, seleccione a opção **To End**. Todas as frases do If até ao fim vão ser corridas.

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

1. Precisarás de seleccionar **Analyze/Descriptive Statistics/Frequencies...**
2. Usa **ae**
3. Faça o **Paste**, escolha e faça 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 omissos. Pode 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. **Analyze/Descriptive Statistics/Descriptives...**
2. A variável é **ae**
3. Não se esqueça 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 indivíduos 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 omissos, o uso descuidado desta técnica pode dar resultados incorrectos. Estamos a usar isto como ilustração dos comandos do SPSS, não como recomendação a usar para compensar os dados em falta).

1. **Transform/Recode/Into Same Variables...** (Na Mesma Variáveis)
A caixa de diálogo Recode into Same Variables (Recodificar na mesma variáveis irá aparecer.
2. Tranfera **ae** a Variables:
3. Pressione **Old and New Values...** (Antigos e Novos Valores)
4. Selecciona o botão circular próximo de **System-missing** (Omisso -- determinado pelo programa de SPSS)
5. Selecciona **Value**: na secção de **New Value** (Valor Novo) e entra **0,79** no box
6. Pressione **Add** (Adicionar)
7. **Continue**
8. Faça o **Paste**, escolha e faça 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 **Data** selecciona **Aggregate...**
2. Transfere **distrito**, **ald**, e **af** para **Break Variable(s)**:
3. Transfere **ae** to **Aggregate Variable(s)**:
4. Pressione **Name & Label...** (Nome e etiqueta)
5. Na caixa de **Name**: entre **ae_tt**
6. Na caixa de **Label**: entre **Adulto Equivalentes**
7. **Continue**
8. **Function...**
9. Selecciona **Sum of values** (Somar os valores individuais)
10. **Continue**
11. Selecciona **Replace working data file** (Sobrepôr o ficheiro activo)
12. **Paste**, Limpe 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. **Analyze/Descriptive Statistics/Descriptives.....**
2. Faça o **Paste**, escolha e faça 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 Af-FICH2.SAV.

1. Assegure que **Data Editor window** esta activo
2. **File/Save As...**
3. Filename Af-FICH2
4. Faça o **paste**, escolha e faça 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 Af-FICH1.SAV com dados sobre produção-calorias e Af-FICH2.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 uma tabela de referência (*keyed tables*).

Vimos antes que variáveis chaves são necessárias para uma junção que inclua uma tabela de referência (*keyed table lookup*). Quando estiver a juntar dois ficheiros do mesmo nível, como estamos a fazer agora, pode não parecer importante incluir as variáveis chaves, mas é. As variáveis chaves determinam que casos devem ser combinados. *Nunca deve 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: AF-FICH2.SAV é ainda o ficheiro em uso

1. **Data/Merge Files/Add Variables...**
2. Usa o ficheiro AF-FICH1.SAV para File Name
3. **Open**
4. Seleccione Match cases nas variáveis chaves 'key variables...'
5. Seleccione Both files provide cases (Todos os dois ficheiros contribuem casos)
6. Key Variables: são **distrito**, **ald**, e **af**, respectivamente
7. **Paste**, limpa os avisos, seleccione o **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 (**cprod_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 (**cprod_ae**).

1. **Transform/Compute...**
2. Target Variable: **cprod_ae**
3. **Type & Label...**
4. Label: **Calorias produzidas per adulto equivalente**
5. **Continue**
6. Numeric Expression: entra **cprod_tt/ae_tt/365**
7. Faça o **Paste**, escolha e faça o **Run** do comando

Antes de podermos produzir a 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 produção de calorias. Uma nova variável contendo os valores 1 a 4 irá indicar a que quartil cada agregado pertence.

1. **Transform/Rank Cases...** (Transformar/Ordenar em grupos)
2. Transfere **cprod_ae** to Variable(s):
3. Transfere **distrito** to By: (Por)
4. **Rank Types...** (Ordenar em grupos os Tipos)
5. **Deseleccione Rank**
6. Seleccione Ntiles: (Quartis)...
7. **Continue**
8. Faça o **Paste**, escolha e faça o **Run** do comando
9. Nota que o nome da nova variável no Output window; deve ser **NCPROD_A**

A primeira coisa a especificar é a variável contendo os valores a usar para o ranking—neste caso **cprod_ae**. A seguir precisamos da variável By para especificar as variáveis que definem os grupos—neste caso **distrito**. **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. **Analyze/Compare Means/Means...**
2. Transfere **cprod_ae** para **D**ependent List:
3. Transfere **ncprod_a** para **I**ndependent list: layer 1 of 1 (Primeira camada)
ncprod_a came from the **Rank Cases** procedure.
4. **Next**
5. Mova **distrito** para **I**ndependent List: layer 2 of 2 (Segunda camada)
6. Faça o **Paste**, escolha e faça o **Run** do comando

Você pode notar que a média da população é 4014,5183 e a média para o segundo quartil em Ribeau 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 Af-FICH3.SAV.

1. Active a janela de Data Editor
2. **File/Save As...**
3. O nome do ficheiro é af-FICH3
4. Faça o **Paste**, escolha e faça 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 **sess2**
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 corridoos 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 Sintaxe previamente salvo, sem ser necessário ter que voltar a fazer de novo os comandos.

Agora vamos ver como podemos rever 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 Viewer.
2. Salve o Viewer com o nome de sess2

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; selecciona o ficheiro sess2.sps.
3.  **OK**
A janela do Syntax c:\sample\sess2.sps será activada.

Agora você pode reexecutar os mesmos comandos ou editá-los, segundo o seu desejo. A sua SESS2.SPS deve ser idêntica à seguinte:

GET

```
FILE='C:\DATA\SPSS\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) p1a (A) .
MATCH FILES /FILE=*
/RENAME p1a=unidade
/TABLE='C:\DATA\SPSS\SAMPLE\CONVER-PT.SAV'
/BY prod unidade.
EXECUTE.
COMPUTE qprod_tt = p1b * conver .
EXECUTE .
MATCH FILES /FILE=*
/TABLE='C:\DATA\SPSS\SAMPLE\CALORIAS-PT.SAV'
/BY prod.
EXECUTE.
COMPUTE cprod_tt = qprod_tt * calories .
EXECUTE .
AGGREGATE
/OUTFILE=*
/BREAK=distrito ald af
/cprod_tt 'Calorias Produzidas nos Alimentos Básicos' = SUM(cprod_tt).
DESCRIPTIVES
VARIABLES=cprod_tt
/FORMAT=LABELS NOINDEX
```



```

/STATISTICS=MEAN STDDEV MIN MAX
/SORT=MEAN (A) .
SAVE OUTFILE='C:\DATA\SPSS\SAMPLE\Af-FICH1.SAV'
/COMPRESSED.
GET
FILE='C:\DATA\SPSS\SAMPLE\C-Q1A-PT.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=distrito ald af
/ae_tt 'Adulto Equivalentes' = SUM(ae).
DESCRIPTIVES
VARIABLES=ae_tt
/FORMAT=LABELS NOINDEX
/STATISTICS=MEAN STDDEV MIN MAX
/SORT=MEAN (A) .
SAVE OUTFILE='C:\DATA\SPSS\SAMPLE\Af-FICH2.SAV'
/COMPRESSED.
MATCH FILES /FILE=*
/FILE='C:\DATA\SPSS\SAMPLE\Af-FICH1.SAV'
/BY distrito ald af.
EXECUTE.
COMPUTE cprod_ae = cprod_tt/ae_tt/365 .
VARIABLE LABELS cprod_ae 'Calories produced per adult equivalent' .
EXECUTE .
RANK
VARIABLES=cprod_ae (A) BY distrito /NTILES (4) /PRINT=YES
/TIES=MEAN .
MEANS
TABLES=cprod_ae BY ncprod_a BY distrito
/CELLS MEAN STDDEV COUNT

```

```
/FORMAT= LABELS .  
SAVE OUTFILE='C:\DATA\SPSS\SAMPLE\Af-FICH3.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.

Etapas:

- 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-PT.sav.
- c. Procure o ficheiro que contém as variáveis necessárias sobre as quantidades de produção vendidas. Nota-se que os códigos dos produtos são os mesmos de C-Q4-PT.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, Af-FICH3.sav.
- g. Lembre que apenas 256 famílias venderam produtos, mas todas as 343 famílias produziram e retiveram calorias. Se a variável **calorias-vendidas** 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 calorias retidas= calorias produzidas - calorias vendidas.
- i. Sub-divida em quartís.
- j. Use o comando **Compare Means** para ver as calorias retidas por **distrito** e por **quartile**.
- k. Salve o ficheiro de dados.
- l. Salve o Syntax Editor deste exercício com o nom Sess2_ex.
- 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:

Case Processing Summary

	Cases				Total	
	Included		Excluded			
	N	Percent	N	Percent	N	Percent
Calorias Retidas por Adulto Equivalente POR DIA * NTILES of CRET_AE by DISTRITO * DISTRICTO	343	100,0%	0	,0%	343	100,0%

Report

Calorias Retidas por Adulto Equivalente POR DIA

NTILES of CRET_AE by DISTRITO	DISTRITO	Mean	N	Std. Deviation
1	MONAPO	1158,7447	27	397,2175
	RIBAUE	1254,6048	29	349,5677
	ANGOICHE	918,7610	28	379,9903
	Total	1111,8447	84	397,5339
2	MONAPO	2257,1590	27	216,6665
	RIBAUE	2181,8827	30	211,4371
	ANGOICHE	1706,6587	29	169,8261
	Total	2045,2660	86	314,7687
3	MONAPO	3319,9480	28	482,8342
	RIBAUE	3169,8478	30	316,7824
	ANGOICHE	2442,9271	29	341,9849
	Total	2975,8490	87	540,8855
4	MONAPO	7639,0532	27	3555,5798
	RIBAUE	5819,6257	30	1620,8821
	ANGOICHE	5018,5639	29	2410,6923
	Total	6120,7158	86	2802,9356
Total	MONAPO	3591,2145	109	3031,5885
	RIBAUE	3122,0523	119	1908,6007
	ANGOICHE	2535,6665	115	1970,9785
	Total	3074,5430	343	2374,3545

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

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 10.0
4º 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 concatenações
- ▶ 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 usando o procedimento do **Crosstabs**, e o do **Tables**.

Abra (**usando Paste**) o ficheiro membro que criamos, o ficheiro Q1A-IDADE.SAV, o qual contém a variável **idade_gp**.

1. **File/Open...**
2. Seleccione q1a-age.sav
3. **Paste**
4. 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. **Analyze/Descriptive Statistics/Crosstabs...**
2. Transfira a **ca2** para **Row(s)**:
3. Transfira **idade_gp** para **Column(s)**
4. **Cells...**
5. Seleccione **O**bserved na secção do Counts.
6. Seleccione **R**ow em Percentages
7. **Continue**
8. Faça o **Paste**, escolha e faça o Run do comando.

Produz a tabela seguinte:

RELAÇÃO AO CHEFE * Grupo Etário em ANOS Crosstabulation

			Grupo Etário em ANOS				Total
			0 até 10	11 até 19	20 até 60	61 e mais	
RELAÇÃO AO CHEFE	CHEFE	Count		6	296	41	343
		% within RELAÇÃO AO CHEFE		1,7%	86,3%	12,0%	100,0%
	ESPOSA(O)	Count		25	280	5	310
		% within RELAÇÃO AO CHEFE		8,1%	90,3%	1,6%	100,0%
	FILHO (A)	Count	503	184	31		718
		% within RELAÇÃO AO CHEFE	70,1%	25,6%	4,3%		100,0%
	MÃE/PAI	Count			5	1	6
		% within RELAÇÃO AO CHEFE			83,3%	16,7%	100,0%
	OUTRO PARENTE	Count	70	55	16	2	143
		% within RELAÇÃO AO CHEFE	49,0%	38,5%	11,2%	1,4%	100,0%
Total		Count	573	270	628	49	1520
		% within RELAÇÃO AO CHEFE	37,7%	17,8%	41,3%	3,2%	100,0%

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

1. **Analyze/Custom Tables/Basic Tables...**
2. Mova **ca2** para a opção **Down**
3. Transfira **idade_gp** para a opção **Across**
4. **Paste** e corra (Run) o programa.

		Grupo Etário em ANOS			
		0 até 10	11 até 19	20 até 60	61 e mais
RELAÇÃO AO CHEFE	CHEFE		6	296	41
	ESPOSA(O)		25	280	5
	FILHO (A)	503	184	31	
	MÃE/PAI			5	1
	OUTRO PARENTE	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 cabeçalhos das colunas são os cabeçalhos dos valores que você designou para a variável **idade_gp**. Se você quiser adaptar a tabela para as suas necessidades, o formato **Basic Tables** pode tornar-se muito mais complexo. Para um exemplo de uma tabela mais complexa:

1. **Analyze/Custom Tables/General Tables...**
2. Transfira **ca2** a **Rows**:
3. Transfira **idade_gp** a **Columns**:
4. Selecciona **idade_gp**, escolha o **Edit Statistics...**
5. Selecciona **Count** em **Cell Statistics**:
6. Cambia **Label**: to **N**
7. Cambia **Width**: to **5**
8. **Change**
9. Selecciona **Count Row %** do **Statistics**:
10. Cambia o **Label**: to **%**
11. Cambia o **Width**: to **5**
12. **Change**
13. **Continue**
14. **Insert Total**
15. Selecciona **ageTotal**, e clica **Edit Statistics...**
16. Selecciona o botão circular cerca do **Custom Total Statistics**
17. Selecciona **Count** em **Statistics**:
18. Cambia **Label**: a **N**
19. Cambia **Width**: a **5**
20. **Add**
21. **Continue**
22. Selecciona **ca2**, e clica **Insert Total**
23. **Format...**
24. ponha **Empty Cell Appearance** a **Zero**
25. **Continue**
26. **Titles...**
27. Em o **Title box**:
28. Escreva **Tabela 1. SPSS para Windows**:
29. Escreva **Grupos Etários por Relação ao Chefe**
30. Escreva **Fonte: O Inquérito do Sector de Família em Nampula, 1991**
31. Em o **Corner** boa: escreva **Relação ao Chefe**
32. **Continue**
33. **Paste** e corra (Run) o programa

Esta é a tabela produzida através daqueles comandos:

**Tabela 1. SPSS para Windows:
Grupos Etários por Relação ao Chefe**
Fonte: O Inquérito do Sector de Família em Nampula, 1991

	Grupo Etário em ANOS								Total
	0 até 10		11 até 19		20 até 60		61 e mais		N
	N	%	N	%	N	%	N	%	
RELAÇÃO CHEFE	0	0%	6	2%	296	86%	41	12%	343
AO CHEFE ESPOSA(O)	0	0%	25	8%	280	90%	5	2%	310
FILHO (A)	503	70%	184	26%	31	4%	0	0%	718
MÃE/PAI	0	0%	0	0%	5	83%	1	17%	6
OUTRO PAREN	70	49%	55	38%	16	11%	2	1%	143
Total	573	38%	270	18%	628	41%	49	3%	1520

Nota: Pode mudar as propriedades da tabela, formatos, *pivotear* 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 periódicos, 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 secção 2.

1. **File/Open...**
2. Seleccione AF-FICH3.sav
3. Cole (**Paste**) e corra (run) o programa
4. **Analyze/Compare Means/Means...**
5. Transfira **Cprod_ae** a Dependent List:
6. Transfira **nkprodae** a Independent list: layer 1 of 1
7. **Next**
8. Transfira **distrito** a Independent List: layer 2 of 2
9. **Paste** e corra (Run) o programa

Isto produz o seguinte resultado:

Report

CALORIAS PRODUZIDAS PER ADULTO EQUIVALENTE 1

NTILES of CPRODAE1	DISTRITO	Mean	N	Std. Deviation
1	MONAPO	1231,6631	27	405,5409
	RIBAUE	1510,1383	29	436,3345
	ANGOICHE	1285,9320	28	483,6820
	Total	1345,8930	84	454,6786
2	MONAPO	2531,5697	27	350,1131
	RIBAUE	2570,0926	30	356,5024
	ANGOICHE	2445,2457	29	304,5521
	Total	2515,8987	86	337,9757
3	MONAPO	3994,1629	28	611,6761
	RIBAUE	4081,3756	30	572,3731
	ANGOICHE	3705,8857	29	504,6422
	Total	3928,1439	87	580,6200
4	MONAPO	9170,2090	27	4678,5504
	RIBAUE	7549,1448	30	2155,2811
	ANGOICHE	8432,4336	29	4004,4919
	Total	8355,9368	86	3742,3294
Total	MONAPO	4229,7201	109	3810,0764
	RIBAUE	3948,0034	119	2558,0066
	ANGOICHE	3990,6912	115	3394,5911
	Total	4051,8408	343	3270,7882

Esta é a informação que precisávamos para preencher os números da nossa tabela da secção 2. Vamos usar “**Custom Tables**” para produzir resultados que se assemelhem a tabela que tenhamos 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 **Variable View/Labels...**

Variável	New Label
cprod_ae1	<i>eliminar a etiqueta</i>
distrito	Distrito
nkprodae	<i>eliminar a etiqueta</i>

2. **Analyze/Custom Tables/Basic Tables...**
3. Transfira **cprod_ae1** a Rows:
4. Transfira **nkprodae** a Columns:
5. Transfira **districto** a Layers:
6. **Statistics...**
7. Selecciona Mean, usa **MEAN**, Format: (**ddd.dd**), Width: **5** e Decimal: **0**
8. **Add**
9. Selecciona Maximum, usa **MAX**, Format: (**ddd.dd**), Width: **5** e Decimal: **0**

10. **Add**
11. Selecciona Minimum, usa o label **MIN**, Format: **(ddd.dd)**, Width: **5** e Decimal: **0**
12. **Add**
13. **Continue**
14. **Layout...**
15. Em a secção dos **Statistic Labels**, selecciona Down the Left side, **Continue**
16. **Titles...**
17. Escreva em o Title boa: **Tabela 1: Produção Alimentar em Calorias**
18. Escreva depois **Per adulto equivalente por dia**
19. Escreva em o Corner boa: **Quartil de Produção**
20. **Continue**, **Paste** e corra o programa

Isto produz seguinte tabela:

**Tabela 1: Produção Alimentar em Calorias
Per Adulto Equivalente por Dia**

		DISTRITO		
		MONAPO	RIBAUE	ANGOCHE
1,00	Mean	(1232)	(1510)	(1286)
	Maximum	(1973)	(2098)	(1952)
	Minimum	(294)	(429)	(362)
2,00	Mean	(2532)	(2570)	(2445)
	Maximum	(3169)	(3190)	(2961)
	Minimum	(1984)	(2102)	(2024)
3,00	Mean	(3994)	(4081)	(3706)
	Maximum	(5067)	(4984)	(4806)
	Minimum	(3176)	(3217)	(2996)
4,00	Mean	(9170)	(7549)	(8432)
	Maximum	(28466)	(13124)	(20485)
	Minimum	(5107)	(4993)	(4974)

A opção “**Rows**” 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 **cprod_ae1** é uma variável contínua para a qual os mínimos, médias, e máximos são calculados para cada categoria de **nkprodae**.

A opção “**Coluna**” determina como agrupar os casos em linhas e colunas da tabela. As variáveis designadas na opção “**Coluna**” devem ser sempre variáveis de categorias (variáveis que tem um número finito de valores). A opção **Edit 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 um a 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 écran, use o F3 para interromper os processamentos subsequentes. Se obtiver algum destes erros, verifique e certifique-se de que listou uma variável continua na opção “**Rows**”, 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 **Viewer** e imprima.

1. Selecciono o **Viewer**
2. Selecciono 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

Tabela 1: Calorias Retidas Por Adulto Equivalente^{es} Por Dia

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

no Exercício 2.1. **Inclua totais** (terá que usar a opção **General Tables**). A sua tabela deverá assemelhar-se a esta:

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 refere-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-PT.sav
3. **Paste**, seleccione e Run o programa

Para criar a tabela faça o seguinte:

1. **Analyze/Custom Tables/General Tables...**
2. **Mult Response Sets...**
3. Selecciona **h35a, h35b, h35c** e Transfira a Variables in Set:
4. Selecciona Categories em as Variables Are coded As boa
5. Name: **Cultura**
6. Label: **Cultura Produzida Principalmente para Ser Vendida**
7. **Add**, **Save**
8. Transfira **\$cultura**, em Mult Response: a Rows:
9. Transfira **distrito** a Columns:
10. Selecciona **\$cultura**
11. **Insert Total**
12. Selecciona **\$culturaTotal**
13. Cambia o nome a **Casos Totais** em Total Label:
14. **Paste** e corra o programa

O comando no Syntax Viewer é:

```
TABLES
  /FORMAT BLANK MISSING('.') /MRGROUP $cultura 'Cultura Produzida'+
  ' Principalmente para Ser Vendida' h35a h35b h35c
  /GBASE=CASES
  /FTOTAL= $t000001 "Casos Totais"
  /TABLE=$cultura + $t000001 BY distrito > (STATISTICS) .
```

A opção **Múltipla Resposta** cria um grupo da variável **\$cultura** 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 Tables** caixa de diálogo..

A tabela dos resultados deve ser apresentada da seguinte maneira:

		DISTRITO		
		MONAPO	RIBAUE	ANGOICHE
Cultura	ALGODAO	63	24	3
Produzida	FEIJAO	13	2	70
Principalmente	GERGELIM			3
para Ser	GIRASOL		1	
Vendida	ARROZ	5	2	78
	MILHO/FEIJAO	7	18	16
	BANANA		2	2
	MANDIOCA		2	5
	CANA ACUCAR	3	1	
	TABACO		1	
	BATATA DOCE			1
	CAJU	1		
Casos Totais		75	44	90

SPSS para Windows SAMPLE SESSION
SECCÃO 4 - Teste Para A Diferença de Médias

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 10.0
4º Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Fevereiro de 2000

Assunto: Há uma diferença significativa entre o nível médio de consumo real (per capita por dia) entres os agregados familiares afectados pela cheias de 1996 e os agregados familiares não afectados pelos cheias?

Base de dados a utilizar: CAJC_V03.SAV do Inquérito aos Agregados Familiares sobre as Condições da Vida (1996/97); Sob-amostra de ÁREAS RURAIS.

Pressupostos de teste para a diferença de médias (Teste de T)

1. A distribuição da média de amostra é aproximadamente normal. É mais provável que este pressuposto seja verdade quando a amostra é grande.
2. Cada observação é seleccionada aleatoriamente e independentemente da população. A verdade deste pressuposto depende do desenho de inquérito.
3. A variância de duas populações $\sigma_1^2 = \sigma_2^2$, mesmo se não conhecemos o valor dessa variâncias.

Hipótese da nulidade: **não** existe diferença significativa entra as médias das duas populações.

Estatística de teste:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/N_1 + S_2^2/N_2}}$$

onde:

X = média

S = variância

N = tamanho de amostra

1. Selecciona **Analyze**
2. Selecciona **Compare Means** (Comparar Médias)
3. Selecciona **Independent Samples T Test** (Teste de T: Amostra Independente)
4. Transfere a variável **consreal** (*Consumo real*) na caixa de **Test Variable** (Variável para ser testada)
5. Transfere a variável **cheia96** (*Afectado até 01/97 pela cheias*) na caixa de **Group Variable** (Variável de Grupo)
6. Selecciona **Define Groups** (Definir os Grupos)
7. Tecle no botão circular de **Use specified values** (Utilizar valores especificados)
8. Ponha o valor de **0** na caixa de **Group 1**
9. Ponha o valor de **1** na caixa de **Group 2**
10. Selecciona **Continue**
11. Selecciona **Options** (Opções)
12. Ponha o valor de **95%** na caixa de **Confidence Interval**
13. Selecciona **Continue**
14. Tecle **Paste**; depois **RUN (To End)** no Syntax Editor

O Comando é:

```
T-TEST
GROUPS=cheia96(0 1)
/MISSING=ANALYSIS
/VARIABLES=consreal
/CRITERIA=CIN(.95) .
```

Os resultados são:

Group Statistics

Afectado ate 01/97 pela cheias		N	Mean	Std. Deviation	Std. Error Mean
Consumo real	0	3949	5555,7507	4951,2996	78,7908
(rafvtdp/pindx_ne)	1	800	4616,0977	4211,4991	148,8990

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Consumo real (rafvtdp/pindx_ne)	Equal variances assumed	,365	,546	5,013	4747	,000	939,6530	187,4489	572,1662	1307,1397
	Equal variances not assumed			5,578	1288,647	,000	939,6530	168,4604	609,1663	1270,1397

1. A diferença das médias é grande, $5555,75 - 4616,09$ Meticais = $939,65$ Meticais

2. O teste de F (F Value) é o teste para a diferença de variâncias.

$$F = (4951,3)^2 / (4211,5)^2 = 0,365$$

Se as variâncias forem consideradas “iguais”, nós conseguimos usar a probabilidade do “Equal variances assumed (Variâncias iguais assumidas)” (2-Tail Prob=0.000). Se as variâncias forem consideradas não iguais, nós vamos usar a probabilidade do “Equal variances not assumed” (2-Tail Prob=0.000). Neste caso a significância não é diferente.

Quando a probabilidade do valor do F é menor do que 0.05, a hipótese de que as variâncias de população são iguais é rejeitada e vamos usar as variâncias separadas (não iguais).

4. Neste caso, o valor de t é:

$$t = \frac{5555,7507 - 4161,0977}{\sqrt{4951,2996^2 / 3949 + 4211,4991^2 / 800}}$$

5. O valor de t usando as variâncias iguais tem uma probabilidade de 0.000. Quando esta probabilidade é maior do 0.05 nós não vamos rejeitar o hipótese a testar de que as médias são iguais. Assim, nós concluimos que **há** a uma diferença estatisticamente significativa entre o consumo real per capita por dia dependente se o agregado familiar foi afectado pelas cheias em 1996.

Segundo Assunto: Há uma diferença significativa entre o nível médio de consumo real (per capita por dia) que depende da quantidade de homens (adultos) em comparação às mulheres (adultos) nos os agregados familiares?

Esta vez vamos utilizar uma variável continua **masc_fem** que é a razão (quociente) de masculino por feminino no agregado familiar. Queremos testar o hipótese de que quando não houver muitos homens dentro de agregado familiar, o nível de consumo real per capita per dia será mais baixo.

Faça uma **Frequency** de variável **masc_fem**.. Deve ver que 26,3% de amostra tem uma razão de menos 0,80 (masculino/feminino), que indica que o agregado familiar tem mais mão-de-obra feminina disponível do que mão-de-obra masculina. Vamos utilizar **0,80** como nosso ponto de decisão, em inglês **cut-off point**, para dividir a amostra nos dois grupos -- o primeiro é todos os casos com a razão masculino/feminino menos de que 0,80; o segundo com todos os casos com a razão masculino $\geq 0,80$.

1. Seleccione **Analyze**
2. Seleccione **Compare Means** (Comparar Médias)
3. Seleccione **Independent Samples T Test** (Teste de T: Amostra Independente)
4. Transfere a variável **consreal** (*Consumo real*) na caixa de **Test Variable** (Variável para ser testada)
5. Transfere a variável **masc_fem** (*Razão de masculino por feminino*) na caixa de **Group Variable** (Variável de Grupo)
6. Seleccione **Define Groups** (Definir os Grupos)
7. Tecele no botão circular de **Cut Point**
8. Ponha **0,80** na caixa de **Cut Point**
9. Tecele **Continue**
10. Seleccione **Options** (Opções)
11. Ponha o valor de **95%** na caixa de **Confidence Interval**
12. Seleccione **Continue**
13. Tecele **Paste**; depois **RUN (To End)** no Syntax Editor

Os comandos:

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

```
T-TEST  
  GROUPS=masc_fem(.80)  
  /MISSING=ANALYSIS  
  /VARIABLES=consreal  
  /CRITERIA=CIN(.95) .
```


Os resultados são:

Razao de masculinos por feminin					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.67	26	.5	.5	.5
	.71	23	.5	.5	1.0
	.71	157	3.3	3.3	4.3
	.71	27	.6	.6	4.9
	.72	101	2.1	2.1	7.0
	.72	69	1.5	1.5	8.5
	.72	41	.9	.9	9.3
	.74	119	2.5	2.5	11.9
	.74	24	.5	.5	12.4
	.75	46	1.0	1.0	13.3
	.76	98	2.1	2.1	15.4
	.77	26	.5	.5	15.9
	.77	24	.5	.5	16.4
	.78	54	1.1	1.1	17.6
	.78	96	2.0	2.0	19.6
	.78	31	.7	.7	20.3
	.79	70	1.5	1.5	21.7
	.79	177	3.7	3.7	25.5
	.79	41	.9	.9	26.3
	.80	27	.6	.6	26.9
	.81	16	.3	.3	27.2
	.81	40	.8	.8	28.1
	.82	18	.4	.4	28.4
	.82	26	.5	.5	29.0
	.83	21	.4	.4	29.4
	.84	30	.6	.6	30.1
	.85	21	.4	.4	30.5
	.85	27	.6	.6	31.1
	.85	72	1.5	1.5	32.6
	.86	38	.8	.8	33.4
	.86	68	1.4	1.4	34.8
	.87	90	1.9	1.9	36.7
	.87	27	.6	.6	37.3
	.87	33	.7	.7	38.0
	.87	27	.6	.6	38.6
	.87	39	.8	.8	39.4
	.87	35	.7	.7	40.1
	.88	17	.4	.4	40.5
	.88	5	.1	.1	40.6
	.88	54	1.1	1.1	41.7
	.88	12	.3	.3	42.0
	.89	49	1.0	1.0	43.0
	.89	69	1.5	1.5	44.5
	.89	36	.8	.8	45.2
	.89	36	.8	.8	46.0
	.89	24	.5	.5	46.5
	.90	15	.3	.3	46.8
	.90	135	2.8	2.8	49.6
	.90	41	.9	.9	50.5
	.91	18	.4	.4	50.9
	.91	53	1.1	1.1	52.0
	.91	42	.9	.9	52.9
	.91	51	1.1	1.1	53.9
	.91	24	.5	.5	54.5
	.92	23	.5	.5	54.9
	.92	121	2.5	2.5	57.5
	.92	162	3.4	3.4	60.9
	.92	32	.7	.7	61.6
	.93	47	1.0	1.0	62.6
	.93	123	2.6	2.6	65.2
	.93	114	2.4	2.4	67.6
	.94	23	.5	.5	68.0
	.94	31	.7	.7	68.7
	.94	32	.7	.7	69.4
	.94	41	.9	.9	70.2
	.94	60	1.3	1.3	71.5
	.95	63	1.3	1.3	72.8
	.95	146	3.1	3.1	75.9
	.95	43	.9	.9	76.8
	.95	99	2.1	2.1	78.9
	.95	32	.7	.7	79.6
	.95	107	2.3	2.3	81.8
	.96	27	.6	.6	82.4
	.96	25	.5	.5	82.9
	.96	77	1.6	1.6	84.5
	.96	39	.8	.8	85.3
	.96	98	2.1	2.1	87.4
	.96	32	.7	.7	88.1
	.96	52	1.1	1.1	89.2
	.97	63	1.3	1.3	90.5
	.97	69	1.5	1.5	92.0
	.98	81	1.7	1.7	93.7
	.98	43	.9	.9	94.6
	.99	61	1.3	1.3	95.9
	1,00	167	3.5	3.5	99.4
	1,02	30	.6	.6	100.0
Total		4749	100.0	100.0	

Group Statistics

Razao de masculinos por feminin		N	Mean	Std. Deviation	Std. Error Mean
Consumo real (rafvtdp/pindx_ne) >= ,80		3499	5388,2806	4713,1483	79,6781
< ,80		1250	5423,1550	5205,2204	147,2259

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Consumo real (rafvtdp/pindx_ne)	Equal variances assumed	2,124	,145	-,218	4747	,827	-34,8744	159,7308	-348,0208	278,2721
	Equal variances not assumed			-,208	2025,745	,835	-34,8744	167,4039	-363,1761	293,4274

O valor de t usando as variâncias iguais tem uma probabilidade de 0.827. Quando esta probabilidade é maior do 0.05 nós vamos **aceitar** o hipótese a testar de que as médias são iguais. Assim, nós concluímos que não há uma diferença estatisticamente significativa entre o consumo real per capita por dia dependente de razão masculino-feminino em 1996/97.

Exercício 4.

1. Refazer o teste de t com a variável de grupo **masc_fem**, utilizando o **cut-off point** de 0,85.
2. Refazer o teste de t com a variável de grupo **masc_fem**, utilizando o **cut-off point** de 0,75.
3. Quais são as suas conclusões? São as conclusões esperadas?
4. Seleccionar só os casos das províncias de norte e centro do país e repetir etapas #1 e #2.
5. O que podemos concluir sobre a nossa hipótese geral que as famílias com razões masculinas/femininas mais baixas têm níveis médias de consumo real mais baixos?

SPSS para Windows SAMPLE SESSION
SECCÃO 5 - Análise de Variância: Oneway

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 10.0
4º Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Março de 2000

Assunto: Tendências nos Padrões de Consumo Alimentar em Malawi

I. Introdução

O avanço na tecnologia agrícola e processamento durante as últimas quatro décadas aumentou em geral a disponibilidade de certos alimentos (trigo, arroz, milho). Por outro lado, devido ao seu quase exclusivo ênfase no trigo e arroz, este avanço contribuiu indirectamente para a distorção do padrão de disponibilidade relativa de outros cereais e, para o declínio da disponibilidade per capita de soja e outras leguminosas. O resultado é que a soja está agora longe do alcance económico das camadas mais pobres da população (devido ao baixo poder de compra) nos países em desenvolvimento, resultando na deterioração de qualidade protéica das dietas nas famílias pobres (Gopalan, 1994).

Análise do Problema

Nome do ficheiro: C:\...\Data\SPSS\Sample\CARACT_P.SAV
Abre o ficheiro, utilizando o PASTE

Objectivos:

- ▶ **Calcular a média de calorias consumidas dos diferentes alimentos pelos agregados familiares (AFs) de acordo com seu rendimento.**
- ▶ **Testar se a média de calorias consumidas dos diferentes alimentos é significativamente diferente entre os diferentes níveis de rendimento dos AFs, através do teste *ONEWAY* (*Análise de Variância Simples*) e o teste da *Mínima Diferença Significativa* (ou *homogeneidade das variâncias*).**
- ▶ **Produzir os resultados na forma final de tabela**

II. Descrição dos Dados

As variáveis usadas para analisar este problema são: **PREND** (rendimento por capita), **NPREND** (quartis de rendimento per capita) e porção de calorias consumidas dos vários grupos de alimentos (por exemplo, **PCALMILH**). Mas, algumas destas variáveis precisam de ser criadas, notadamente as porções de calorias consumidas dos vários grupos de alimentos (**PCALMILH**, etc.). Assim, as porções de calorias dos principais grupos de alimentos serão calculadas dividindo as calorias consumidas de cada grupo de alimentos pelas calorias totais consumidas. Criar estas variáveis utilizando as seguintes etiquetas para

os 8 grupos: **PCALMILH** (milho), **PCALCER** (outros cereais), **PCALTBER** (raízes e tubérculos e plátanos), **PCALCPOV** (carne/peixe/ovos), **PCALLEIT** (leite), **PCALVEG** (vegetais e frutas), **PCALGOR** (óleos/gorduras), e **PCALSOJA** (soja).

O rendimento por capita (**PREND**) pode ser obtido dividindo o rendimento total do AF pelo seu tamanho. Os quartis de rendimento serão obtidos dividindo a amostra em quatro sub-amostras baseando-se no rendimento anual per capita, através do comando RANK. Estes serão classificados em **4 quartis sócio-económicos**.

III. Método de Análise

1. Para realizar as análises necessárias, precisamos de definir as novas variáveis atrás mencionadas e criar os grupos ou quartis de rendimento sócio-económicos. Assim, podemos proceder na seguinte ordem:

- 1.a. **Definição das novas variáveis para grupos de alimentos e cálculo da porção de calorias consumidas de vários grupos de alimentos.**

*Utilize o comando **COMPUTE**, e escreva as fórmulas para calcular as porções de calorias de cada grupo de alimentos em percentagens.*

- 1.b. **Tradução das variáveis anteriormente criadas.**

*Utilize o comando **VARIABLE LABELS**.*

- 1.c. **Criação de novas variáveis de rendimento per capita e dos quartis de rendimento per capita --- 4 quartis de rendimento.**

*Utilize os comandos **COMPUTE** e **RANK***

1. A partir de **Transform** seleccione **Compute...**

2. Para a **Target Variable**: escreve **PREND**

3. *Click* 


4. Na caixa do **Label**: insira o texto “**Rendimento por Capita**”.

Click 


5. Na caixa de Numeric Expression, tecle: *rendtot/tamaf*.

6. A partir de **Transform** seleccione **Rank Cases**

7. Transfira **PREND** na caixa de Variable.

8. *Click* na caixa de 

9. Seleccione a caixa de **NTILES**, para criar quartis (4)

10. *Click* 

11. **Paste** e **Run** o comando.

2. Depois precisamos de conhecer as médias de calorias consumidas de cada grupo de alimentos em cada um dos grupos sócio-económicos (ou quartis de rendimento per capita).

- 2.1. **Cálculo da média de porção de calorias consumidas de cada grupo de alimentos por quartis de rendimento.**

Utilize os comandos apropriados para o cálculo das médias para produzir a seguinte tabela.

Report

NTILES of PREND		Porção de Calorias do Milho	Porção de Calorias do Outros Cereais	Porção de Calorias dos Raízes e Tubérculos	Porção de Calorias dos Carne, Peixe e Ovos	Porção de Calorias da Leite	Porção de Calorias das Vegetais e Frutas	Porção de Calorias das Óleos e Gorduras	Porção de Calorias da Soja
1	Mean	90,7508	,6079	1,9691	1,2167	5,534E-02	,9371	,5975	3,3039
	N	127	127	127	127	127	127	127	127
	Std. Deviation	11,5667	2,7708	6,1522	1,8517	,3976	1,5839	2,8116	8,2459
2	Mean	91,6801	,5697	,8333	1,5649	5,351E-02	1,0713	,6985	3,0958
	N	169	170	170	169	169	169	169	170
	Std. Deviation	10,8803	3,5700	2,5981	2,7333	,2458	1,7466	2,0336	7,1667
3	Mean	88,0085	,4583	1,5227	2,1419	,1689	3,5417	,9223	2,7872
	N	171	171	171	171	171	171	171	171
	Std. Deviation	13,4552	2,2358	5,5866	3,9422	,9504	8,5869	4,3000	4,2394
4	Mean	88,8975	,8584	1,3042	2,6573	,1988	,7343	1,2544	3,9933
	N	170	170	170	170	170	170	170	170
	Std. Deviation	11,6763	3,0732	3,0953	3,9618	,6705	1,5791	2,3103	6,0948
Total	Mean	89,7666	,6244	1,3696	1,9419	,1236	1,6178	,8868	3,2936
	N	637	638	638	637	637	637	637	638
	Std. Deviation	12,0248	2,9602	4,5060	3,3568	,6422	4,8010	3,0136	6,4759

3. A seguir iremos comparar as médias de porções de calorias consumidas de cada grupo de alimentos, entre os quatro grupos sócio-económicos. Como a variável dependente (porção de calorias consumidas) é uma variável de intervalo e a variável independente (os quatro grupos sócio-económicos) é uma variável ordinal, **a análise de variância (oneway) (*)** é a mais apropriada para comparar as médias entre os quatro grupos sócio-económicos. A análise envolve comparação da variação na variável dependente com a variação observada entre os grupos formados pela variável independente. Juntamente com a análise de variância, precisamos de incluir o comando para o teste da Mínima Diferença Significativa. Assim:

3.1. Comparação das porções de calorias consumidas de vários grupos de alimentos entre os quatro grupos sócio-económicos.

1. Seleccione **Analyze**
2. Seleccione **Compare Means** (Comparar Médias)
3. Seleccione **One-Way ANOVA** (Análise de Variância Simples)
4. Transfere as variáveis **PCALMILH** até **PCALSOJA** na caixa de **Dependent List** (Listas das Variáveis Dependentes)
5. Transfere a variável **NPREND** na caixa de **Factor**
6. *Click* na caixa de Post-Hoc.. (Comparações Múltiplas Post-Hoc)
7. Na caixa de **Equal Variances Assumed** (Variâncias Iguais Assumidas), seleccione **LSD**, o teste *Least Significant Difference* (Mínima Diferença Significativa)
8. Seleccione **Continue**
9. Tecele **Paste**; depois **RUN (To End)** no Syntax Editor

A Análise de Variância (*Oneway*) é usada para comparar as médias de três ou mais sub-amostras. A variável dependente deve ser contínua ou numa escala de índices. A análise envolve comparações da variação na variável dependente que ocorre com a variação que é observada entre os grupos formados pela variável independente.

A análise de variância usa o *F Teste* para testar a hipótese nula de que as médias de percentagem de calorias de cada fonte particular são idênticas entre os quatro quartis de rendimento dos AFs.

Para o exemplo de Porção de calorias do milho:

$$\begin{aligned} \text{O proporção de } F \text{ (} F \text{ Ratio)} &= \frac{\text{Between Groups Mean Square}}{\text{Within Groups Mean Square}} \\ &= \frac{466,252}{143,070} = 3,259 \quad \text{onde significância} = 0,021 \end{aligned}$$

onde Between Groups Mean Square é uma medição de variabilidade **entre** os grupos e Within Groups Mean Square é uma medição de variabilidade **dentro de** grupos. Quando a

probabilidade de F (*F Prob.*) ≤ 0.05 , nós concluímos que as médias de população de grupos diferentes **não** são provavelmente iguais. Para saber ONDE as médias são significativamente diferentes, precisamos usar os testes de comparação múltipla, como LSD ou Scheffe. O *output* vai mostrar todos as estatísticas descritivas. Nós podemos ver que o Grupo 2 tem a média mais alta (91,7%) e Grupo 3 tem a média mais baixa (88,0). Suponhamos que haja uma boa probabilidade de que pelo menos estas duas médias serão significativamente diferentes, mas isto vai depender do nível de desvio padrão e o número total de casos. Os resultados de LSD diz que diferenças significativas entre as médias são vistas entre as seguintes grupos: 1 e 3 (sig. = 0,051), 2 e 3 (sig. = 0,005), e 2 e 4 (0,033).

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Porção de Calorias do Milho	Between Groups	1398,755	3	466,252	3,259	,021
	Within Groups	90563,469	633	143,070		
	Total	91962,224	636			
Porção de Calorias dos Outros Cereais	Between Groups	14,571	3	4,857	,553	,646
	Within Groups	5567,169	634	8,781		
	Total	5581,741	637			
Porção de Calorias dos Raízes e Tubérculos	Between Groups	99,271	3	33,090	1,635	,180
	Within Groups	12834,566	634	20,244		
	Total	12933,837	637			
Porção de Calorias dos Carne, Peixe e Ovos	Between Groups	184,659	3	61,553	5,581	,001
	Within Groups	6981,731	633	11,030		
	Total	7166,390	636			
Porção de Calorias da Leite	Between Groups	2,735	3	,912	2,223	,084
	Within Groups	259,590	633	,410		
	Total	262,325	636			
Porção de Calorias das Vegetais e Frutas	Between Groups	874,928	3	291,643	13,392	,000
	Within Groups	13784,865	633	21,777		
	Total	14659,793	636			
Porção de Calorias das Óleos e Gorduras	Between Groups	39,801	3	13,267	1,464	,223
	Within Groups	5736,082	633	9,062		
	Total	5775,884	636			
Porção de Calorias da Soja	Between Groups	133,741	3	44,580	1,063	,364
	Within Groups	26580,434	634	41,925		
	Total	26714,175	637			

Multiple Comparisons

Dependent Variable: Porção de Calorias do Milho

LSD

(I) NTILES of PREND	(J) NTILES of PREND	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-,9293	1,4047	,508	-3,6877	1,8291
	3	2,7423	1,4011	,051	-9,1416E-03	5,4938
	4	1,8533	1,4029	,187	-,9016	4,6082
2	1	,9293	1,4047	,508	-1,8291	3,6877
	3	3,6716*	1,2974	,005	1,1239	6,2193
	4	2,7826*	1,2993	,033	,2311	5,3340
3	1	-2,7423	1,4011	,051	-5,4938	9,142E-03
	2	-3,6716*	1,2974	,005	-6,2193	-1,1239
	4	-,8890	1,2955	,493	-3,4330	1,6549
4	1	-1,8533	1,4029	,187	-4,6082	,9016
	2	-2,7826*	1,2993	,033	-5,3340	-,2311
	3	,8890	1,2955	,493	-1,6549	3,4330

*. The mean difference is significant at the .05 level.

Tipicamente, quando há uma diferença significativa entre duas médias, nos indicamos esta diferença com um par de letras. No caso de milho, a diferença entre 1 e 3 é indicada pelo *a*, a diferença entre 2 e 3 com a letra *b*, entre 2 e 4 pela letra *c*.

Exercício 5.1. Analisar os resultados para os outros grupos de alimentos e quando existe uma diferença significativa entre um par de médias, indique esta diferença na coluna apropriada na Tabela seguinte.

Report

NTILE S of PREND		Porção de Calorias do Milho	Porção de Calorias do Outros Cereais	Porção de Calorias dos Raízes e Tubérculos	Porção de Calorias dos Carne, Peixe e Ovos	Porção de Calorias da Leite	Porção de Calorias das Vegetais e Frutas	Porção de Calorias das Óleos e Gorduras	Porção de Calorias da Soja
1	Mean	90,8 ^a	0,6	2,0	1,2	0,1	0,9	0,6	3,3
	N	127,0	127,0	127,0	127,0	127,0	127,0	127,0	127,0
	Std.	11,6	2,8	6,2	1,9	0,4	1,6	2,8	8,2
2	Mean	91,7 ^{bc}	0,6	0,8	1,6	0,1	1,1	0,7	3,1
	N	169,0	170,0	170,0	169,0	169,0	169,0	169,0	170,0
	Std.	10,9	3,6	2,6	2,7	0,2	1,7	2,0	7,2
3	Mean	88,0 ^{ab}	0,5	1,5	2,1	0,2	3,5	0,9	2,8
	N	171,0	171,0	171,0	171,0	171,0	171,0	171,0	171,0
	Std.	13,5	2,2	5,6	3,9	1,0	8,6	4,3	4,2
4	Mean	88,9 ^c	0,9	1,3	2,7	0,2	0,7	1,3	4,0
	N	170,0	170,0	170,0	170,0	170,0	170,0	170,0	170,0
	Std.	11,7	3,1	3,1	4,0	0,7	1,6	2,3	6,1
Total	Mean	89,8	0,6	1,4	1,9	0,1	1,6	0,9	3,3
	N	637,0	638,0	638,0	637,0	637,0	637,0	637,0	638,0
	Std.	12,0	3,0	4,5	3,4	0,6	4,8	3,0	6,5

Exercício 5.2. Quais são as implicações destes resultados para a formulação de políticas em Malawi?

Respostas:

NTILES of PREN D		Porção de Calorias do Milho	Porção de Calorias do Outros Cereais	Porção de Calorias dos Raízes e Tubérculos	Porção de Calorias dos Carne, Peixe e Ovos	Porção de Calorias da Leite	Porção de Calorias das Vegetais e Frutas	Porção de Calorias das Óleos e Gorduras	Porção de Calorias da Soja
1	Mean	90,8 ^a	0,6	2,0 ^a	1,2 ^{ac}	0,1	0,9 ^a	0,6	3,3
2	Mean	91,7 ^{bc}	0,6	0,8 ^a	1,6 ^b	0,1 ^a	1,1 ^b	0,7	3,1
3	Mean	88,0 ^{ab}	0,5	1,5	2,1 ^c	0,2	3,5 ^{abc}	0,9	2,8
4	Mean	88,9 ^c	0,9	1,3	2,7 ^{ab}	0,2 ^a	0,7 ^c	1,3	40,0
Total	Mean	89,8	0,6	1,4	1,9	0,1	1,6	0,9	3,3

SPSS para Windows SAMPLE SESSION
SECCÃO 6 - Correlação

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 10.0
4º Edição

Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
Março de 2000

ASSOCIAÇÃO EM RELAÇÃO À CAUSALIDADE

I. O que é uma **associação** entre as variáveis?

O coeficiente de correlação linear (ρ) permite-nos medir o grau de associação entre duas variáveis. Dizemos que há uma correlação **positiva** quando os registos com os valores baixos para uma variável correspondem com os valores baixos na outra variável e os registos com os valores altos também têm tendência de ter os valores altos na outra. As variáveis que têm uma correlação **negativa** mostra uma relação contrária: os valores mais altos numa variável correspondem aos mais baixos na outra variável.

O coeficiente de Correlação linear é dado pela fórmula:

$$\rho_{XY} = \frac{\text{Covariância (X,Y)}}{\sqrt{\text{Variância (X)*Variância(Y)}}} \quad \text{onde X e Y são variáveis diferentes}$$

Os valores de ρ variam entre -1 e 1. A correlação positiva é perfeita quando $\rho=1$, e a correlação negativa é perfeita quando $\rho=-1$. Normalmente o símbolo ρ é para a correlação a nível de população e r é utilizado para as amostras. Os resultados do comando CORRELATE no SPSS vai dizer-nos se o valor de r é diferente estatisticamente de zero (indicado com o * (.05) ou ** (.01)).

B. Tipos de associação

Quando dois factores (variáveis) não têm associação estatística, há 3 razões possíveis:

1. É verdadeiro que não exista uma associação entre as variáveis
2. A medição das variáveis foi mal feita
3. Há um problema de confundimento negativo-- um outro factor está a esconder a relação genuína

Por exemplo, em muitos estudos, há uma relação entre a quantidade ingerida de comida por dia e ZPI (peso para idade), um indicador de estado nutricional. Temos uma hipótese que há

uma relação causal entre estas variáveis, quer dizer que uma criança que come muito cada dia vai crescer bem e ter um peso adequado:

Quantidade Ingerida Por dia \Rightarrow ZPI

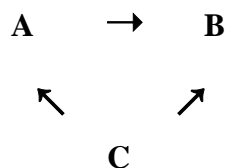
Se o r so é 0,65, temos uma associação fraca que pode ser devida a chance. Há varias razões para este resultado insignificativo. Pode ser:

1. Um erro de escrita
2. As idades verdadeiras não estão disponíveis
3. O sistema de alimentar à criança mudou com a presença do inquérito
4. Os dias escolhidos para medir a quantidade ingerida de comida não foram representativos

Quando dois factores (variáveis) **estão** associados, há dois tipos de associação:

1. Há uma associação causal, quer dizer que uma variável *independente* pode explicar (causar) o movimento duma outra variável (a variável *dependente*). Esta relação causal pode ser *directa* ou *indirecta*.
2. Não há uma associação causal, mas estamos a observar um outro exemplo de confundimento.

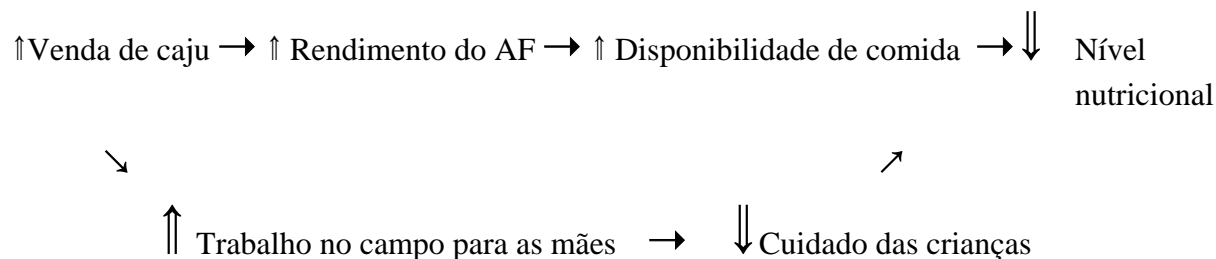
Quando há confundimento positivo, um outro factor desconhecido (C) está a causar esta associação observada entre as variáveis A e B:



Quando há um confundimento negativo, uma associação existe, mas um outro factor (ou factores) estão a eliminar o efeito desejado. Por exemplo, temos uma hipótese que a venda significativa de caju vai aumentar o rendimento do agregado familiar e a família vai comprar mais comida. Esta comida vai ser dado à criança e o nível nutricional da criança vai ser melhorado.

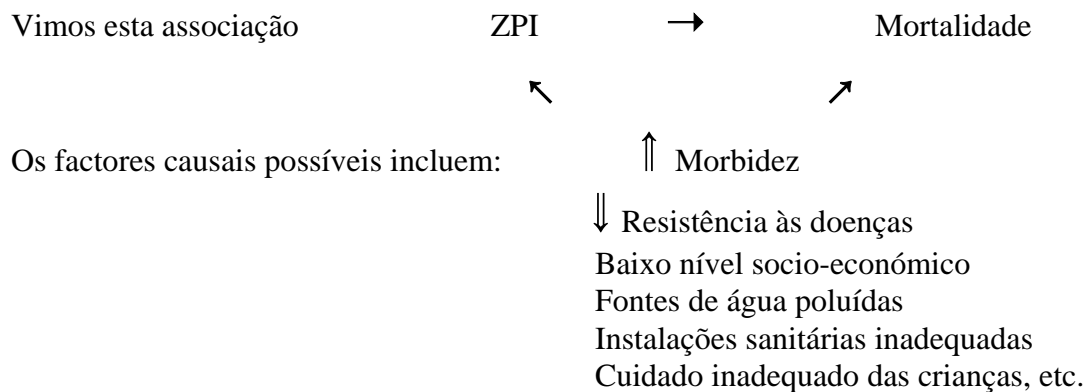
↑ Venda de caju \rightarrow ↑ Rendimento do AF \rightarrow ↑ Disponibilidade de comida \rightarrow ↑ Nível nutricional

Mas, não vimos o aumento do nível nutricional porque no mesmo tempo, o aumento da produção de caju causa o aumento do trabalho das mães de crianças e o tempo dado ao cuidado das crianças é reduzido.



O benefício do aumento de rendimento do AF é eliminado pela diminuição do cuidado das crianças.

É muito importante examinar as associações hipotéticas criticamente e compreender todos os factores que podem influenciar uma relação. Por exemplo, vários estudos em África e em Ásia encontraram que as crianças com um Peso para Idade (ZPI) muito baixo têm probabilidades muito mais altas de morrer nos seguintes dois anos que as crianças com um peso para idade adequado. Antes de decidir em investir muito dinheiro para reduzir desnutrição esperando que a taxa de mortalidade vai descer consideravelmente, é preciso ter a certeza que esta relação causal é verdadeira.



Muitos destes factores têm uma associação com ZPI e Mortalidade e podem ser as variáveis confundidas. Se todas estas variáveis foram incluídos no inquérito, é possível controlar para o efeito destes factores mais tarde na análise de regressão. Quer dizer que a presença dos factores confundidos aumenta a quantidade de informação recolhida para compreender quais são os verdadeiros factores responsáveis para a mortalidade das crianças numa situação determinada.

C. Quais são os critérios para estabelecer causalidade:

1. Deve ser um efeito ordenado cronologicamente: CAUSA → EFEITO

2. O grau e a consistência da associação:
Quer dizer, esta associação mantém-se nas áreas diferentes, para homens e mulheres?
Se puder ver alguma reacção quando há uma intervenção?
3. Os resultados estão de acordo com o conhecimento já existente?
4. Pode replicar os resultados em outro sítio ou com os dados de outro inquérito?

III. Calcular as Correlações Utilizando SPSS 10.0

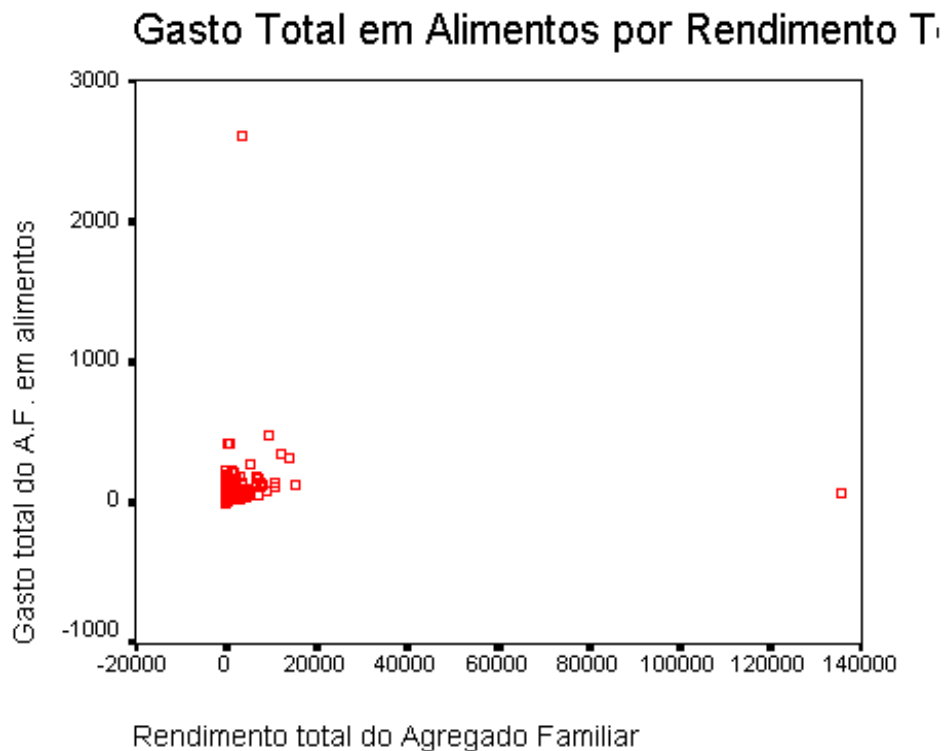
Quando pensamos que uma associação é positiva ou negativa ANTES de calcular a correlação, usamos o teste de significância duma aba. Por exemplo, nos esperamos que quando o rendimento total aumento, a quantidade de dinheiro gasto em alimentos também aumenta. Estamos a assumir que a associação é uni-direccional, no sentido positivo.

Abre o ficheiro CARACT_P.SAV, utilizando **Paste**.

Como sempre, queremos VER os nossos dados antes de analisar.

1. Seleccione **Graphs** do menu
2. Seleccione **Scatter** (espalhar)
3. Seleccione **Bivariate**
4. Tecele em cima de caixa de **Simple**
5. Seleccione **Define**
6. Transfira **gatotal** (gasto total do AF em alimentos) na caixa de **Y-Axis**
7. Transfira **rendtot** (rendimento total do AF) na caixa de **X-Axis**
8. Seleccione **Titles**
9. Na caixa de primeira linha, tecele “Gasto Total em Alimentos por Rendimento Total do AF”
10. Seleccione **Continue**
11. Seleccione **Paste e Run** no SYNTAX EDITOR.

No SPSS Viewer, temos o gráfico seguinte que parece tem 2 *outliers* (valores extremos) muito influentes.



Utiliza **Select Cases** para incluir só os casos onde o **gatotal** é menos que 2000 e o rendimento total (**rendtot**) é menos que 120000. Refazer o *scatter plot* (gráfico bivariate) e ver a diferença em termos de distribuição.

Agora, podemos calcular a nossa correlação de PEARSON's, uma medição de uma relação linear.

1. Selecciona **Analyze**
2. Selecciona **Correlate** (Fazer uma Correlação)
3. Selecciona **Bivariate** (uma correlação entre duas variáveis)
4. Transfira duas variáveis **gatotal** e **rendtot** na caixa de **Variables**
5. Na caixa de **Correlation Coefficients** (Coeficientes de Correlação), selecciona **PEARSON** que é apropriada para comparar duas variáveis contínuas.
Quando pelo menos uma de duas variáveis é categórica, precisa de utilizar SPEARMAN's ou Kendall's.
6. Na caixa de **Test of Significance** (Teste de Significância), Selecciona **One-Tailed** (uma aba) porque nos esperamos que a relação é positiva.
Se não tem um pressuposto sobre a direcção, utiliza Two-Tailed.
7. Selecciona **Flag Significant Correlations** (Indicar as correlações significativas com um *) na esquina esquerda.
8. Selecciona **Options**.

9. Seleccione Means and Standard Deviations
10. Teclé **Continue**
11. Seleccione **Paste** e **Run** no SYNTAX EDITOR.

Correlations

		Gasto total do A.F. em alimentos	Rendimento total do Agregado Familiar
Gasto total do A.F. em alimentos	Pearson Correlation	1,000	,445**
	Sig. (1-tailed)	,	,000
	N	680	680
Rendimento total do Agregado Familiar	Pearson Correlation	,445**	1,000
	Sig. (1-tailed)	,000	,
	N	680	680

** . Correlation is significant at the 0.01 level (1-tailed).

Como nos esperamos, a uma relação positiva significativa (a nível de 0,01) entre estas duas variáveis.

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 10

**Departamento de Economia Agrária, Universidade Estadual de Michigan
East Lansing, Michigan, USA
2000**

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, queremos 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						

IV. PRODUÇÃO

Quadro IV: Características de Produção

Produto	Quantidade colhida nesta campanha		Quantidade colhida num ano normal		Na altura da colheita, quanto havia de estoque		Em que mes, os estoques se acabaram	Quanto desta campanha pensa guardar para o consumo da família?		Até que mes os estoques actuais vão durar	Quanto desta campanha foi guardado para semente?		
	Unidade	Qt	Unidade	Qt	Unidade	Qt		Unidade	Qt		Unidade	Qt	
1=milho	8=amendoim	Unidade	Qt	Unidade	Qt	Unidade	Qt	Entrar o mes	Unidade	Qt	Entrar o mes ou "ano todo" se for o caso	Unidade	Qt
2=feijões	9=castanha	1=saco 100		1=saco 100		1=saco 100			1=saco 100			1=saco 100	
3=feijão	de caju	2=saco 50		2=saco 50		2=saco 50			2=saco 50			2=saco 50	
4=mantiega	10=bebida	3=kilo		3=kilo		3=kilo			3=kilo			3=kilo	
5=mandioca	de caju	4=litro		4=litro		4=litro			4=litro			4=litro	
6=arroz	11=bebida	5=lata 20		5=lata 20		5=lata 20			5=lata 20			5=lata 20	
7=mapira	de cana	outro		outro		outro			outro			outro	
12=coco													
13=ossura													
outros													
PROD		P1A	P1	P2A	P2	P3A	P3	P4	P5A	P5B	P6	P7A	P7

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 machamba	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=época 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=AGRICO M 4=ambulante 5=brigada 6=companhia outro	1=na machamba/casa 2=aldeia 3=localidade 4=distrito 5=província outro	entrar os quilómetros entra a machamba 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												
9												

Observações metodológicas

(Secção disponível só em inglês)

Filters Versus Permanent Selections.

You can filter or delete cases that don't meet the selection criteria. In Section 2 of the cross-sectional training, we filtered the data but we did not delete any cases. When you set a filter from the **Data/Select cases** command, unselected cases are filtered by default. Filtered cases remain in the data file but are excluded from analysis. You can see which cases are filtered out by looking at the far left column of the Data View window, where the case numbers are given. Numbers with a slash through them have been filtered and will not be included in an analysis or reporting. SPSS creates a filter variable, FILTER_\$, to indicate filter status. Selected cases have a value of 1; filtered cases have a value of 0. To turn filtering off and include all cases in your analysis, select All cases in the **Data/Select cases** command. If you want to delete specific cases from the data set, use the **Data/Select cases** command, complete an IF statement for those cases that you want to keep, and then select the choice **deleted** in the **Unselected Cases Are** box. Be sure to save this file under a new name or you will permanently delete the cases from the data file.

The Three Line Charts and Three Data in Charts Options.

The **Graph/Line** command allows you to make selections that determine the type of chart you obtain, simple, multiple and drop-line. In the menu, select the icon for the chart type you want, and select the option under **Data in Chart Are** that best describes your data. You can see a description of the three available **Data in Chart** types below. A category axis on a chart is an axis that displays values individually, without necessarily arranging them to scale. (A scale axis, in contrast, displays numerical values to scale.) Bar charts, line charts, and area charts usually have one category axis and at least one scale axis. Scatterplots and histograms do not have a category axis.

The **Missing Values** options are available only when the new chart will display or summarize more than one variable (not including variables that define groups):

- **Exclude cases listwise** excludes a case from the entire chart if it has a missing value for any of the variables summarized.
- **Exclude cases variable by variable** excludes a case separately from each summary statistic calculated. Different chart elements may be based on different groups of cases.

Display groups defined by missing values is available only when you use a categorical variable to define groups for a new chart. If selected, each missing value for the grouping variable (including the system-missing value) will appear as a separate group in the chart. If not, cases with system-missing or user-missing values for the grouping variable are excluded from the chart. It is recommended to always uncheck this box as it is not of interest to show on a graph the missing values or sysmisses.

Simple lines

Summaries for Groups of Cases

Categories of a single variable are summarized. The y-height of the points is determined by the **Line Represents** option.

A single **Category Axis** variable.

Summaries of Separate Variables

Two or more variables are summarized. Each point represents one of the variables.
Two or more Line Represents variables.

Values of Individual Cases

A single variable is summarized. Each point represents an individual case.
A single Line Represents variable.

Multiple lines

Summaries for Groups of Cases

Categories of one variable are summarized within categories of another variable. The y-height of the points is determined by the Lines Represent option.

A Category Axis variable (Category Variable 1).

A Define Lines by variable (Category Variable 2).

Summaries of Separate Variables

Two or more variables are summarized within categories of another variable.

Two or more Lines Represent variables (Var 1, Var 2).

A Category Axis variable (Category Variable).

Values of Individual Cases

Two or more variables are summarized for each case.

Two or more Lines Represent variables (Var 1, Var 2).

Manipulating Output in SPSS 10.0 for Windows.

Numerous modules could be dedicated to working with the Output navigator. Section 4 only discussed simple cutting and pasting of results. One suggestion would be to follow the tutorial within SPSS to learn about the countless possibilities and options which are available to the SPSS user in the Output navigator. Your results have never looked this good! Easier and faster data exploration and to ability to drag icons in the navigator outline and content panes on the left, expand and collapse the outline - see the output you want; multi-dimensional pivot tables, swapping and hiding rows and columns, new and numerous styles for charts and tables, colors, fonts, line styles, text attributes; no loss of any custom formatting, dragging output from SPSS to a word processor (in windows metafile format); change a title directly within the output, right click for pop-up menus as shortcuts, and much more.

You may have trouble viewing the complete output following a SPSS command like **Frequencies** or **Tables**. It may run hundred and thousands of cases but will only show the first 50 for example. To view all of the specific output in this case, simply double click or right click on the selected output and choose **Open**. This will open a separate window called a pivot table. Then scroll down to see the output in whole. You may also edit the table here as well. Enjoy using the various options given to you to modify the styles, formats, colors, text attributes and so on.