

SPSS para Windows *SAMPLE SESSION*

Analysis of Time-Series Data

Análise dos Dados em Séries Temporais

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

Cynthia Donovan, Assistant Professor
Departamento de Economia Agrária, Universidade Estatal de Michigan
East Lansing, Michigan, EUA

donovanc@msu.edu

Janeiro de 2000

Material de Treinamento para Curso de Curta Duração

Módulo 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*)

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

Módulo 2 - Restruturação dos Data Files - Table Lookup e Aggregation

Módulo 3 - Importação dos Dados de Outros Formatos

Módulo 4 - Introdução --Séries Temporais

Módulo 5 - Análises Básicas -- Séries Temporais

Módulo 6 - Análises Sazonais

Módulo 7 - Tendências

Módulo 8 - Preços Reais, Apresentação Gráfica de Preços e Tabelas

Módulo 9 - Análises Marginais

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”

Módulo 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*)

**Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 10.0
3ª Edição**

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

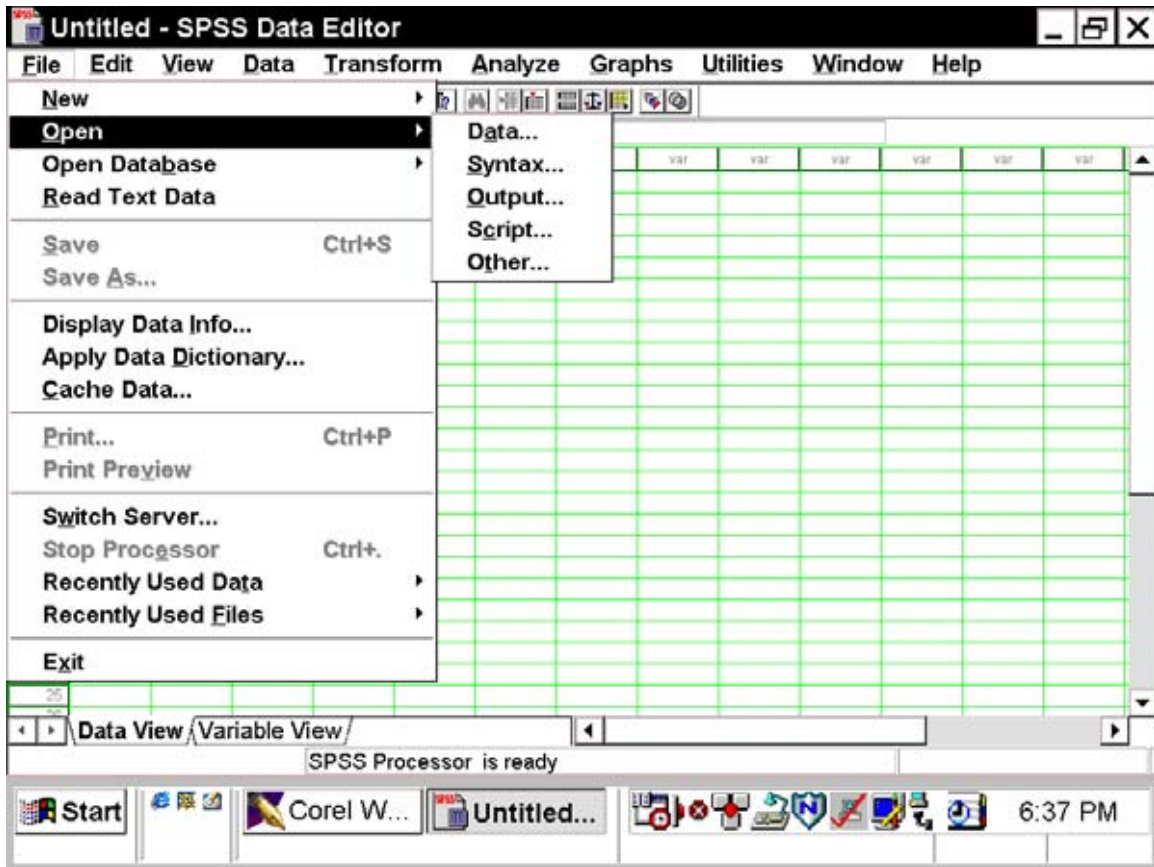
Este módulo introduz os conceitos básicos de níveis de dados e métodos de organização de dados. Também dá uma descrição breve sobre a estrutura dos ficheiros no SPSS 10.0 para Windows. É essencial ler este módulo antes de começar o Sample Session.

Ficheiros Utilizados no SPSS 10.0 para Windows

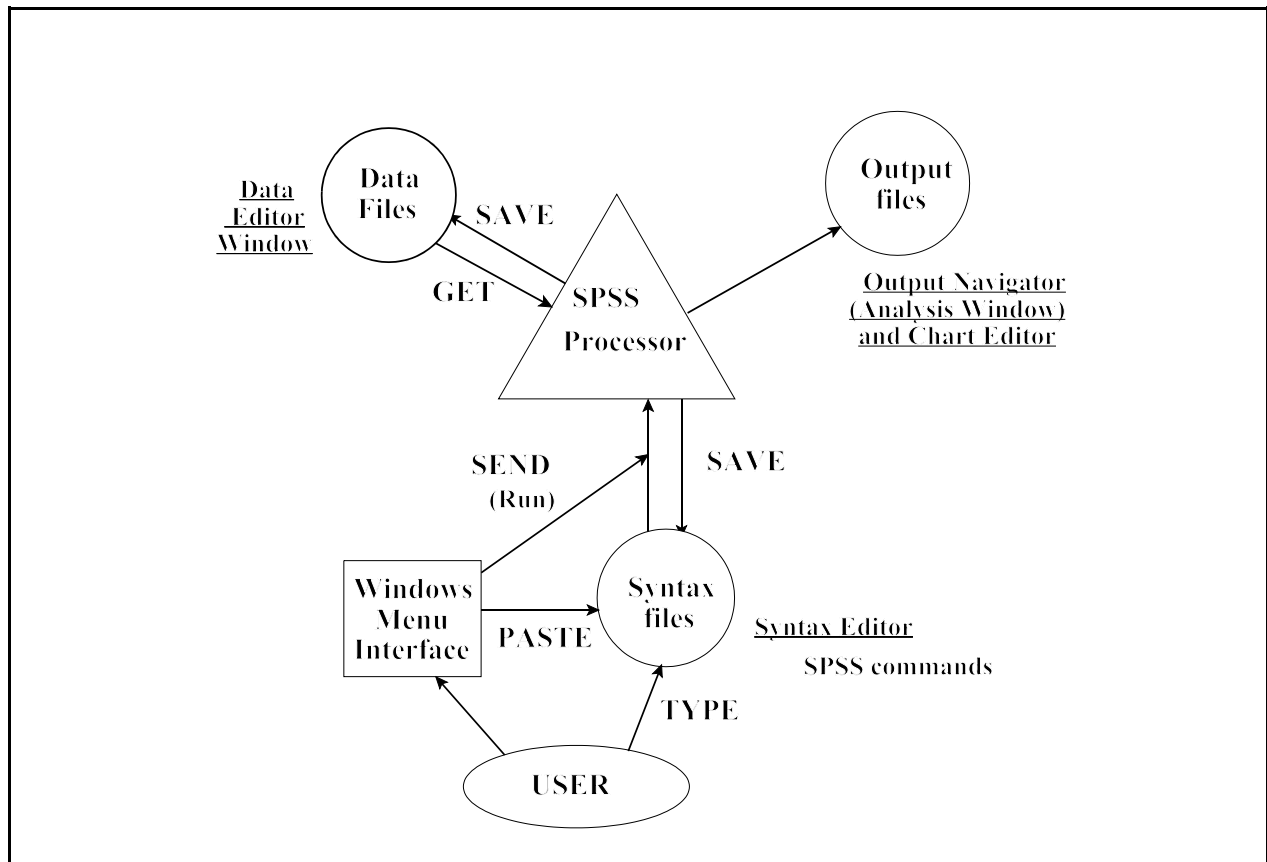
Ao utilizar o SPSS 10.0 para Windows da maneira ensinada neste “sample session”, está a usar quatro diferentes janelas (“windows”) dentro do programa - o *Syntax Editor*, o *Data Editor* e o *Viewer*, e o *Scripts Facility*.

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...**
- ◆ Tecle o *Viewer*
- ◆ Seleccione *Display commands in the log*
- ◆ Tecle *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* contem 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**, va 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 DE SÉRIES TEMPORAIS
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, MADER. 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”
Módulo 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* contem. Podemos averiguar isto, junto com outras informações, utilizando o comando **Variables...** no **Utilities** menu, tanto no **Syntax Editor** como no **Data Editor**. Isto lhe permite observar as definições das variáveis e 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		OK	Aceitar	Cancel	Cancelar	Ajuda
Value:	tecle 4			Help		
Value Label:	tecle <i>NAMPULA</i>					

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	Numérico ou Alfanumérico (<i>String</i>)	Espaço requisito 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= RIBAU 3= ANGO 4= CHE	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. Decida 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 Statistics... 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 Continue.*
voltará para o Explore dialog box.
7. *Click no Paste 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.


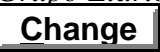

Transformação de Dados

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 Variable -> Output Variable: box **ca3** deve mover para o Input Variable->Output Variable: box e o nome do box mudará para Numeric Variable -> Output Variable.
5. Click uma vez na próxima caixa vazia para **Name**: na secção Output Variable para por o cursor lá.
6. Escreva **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. Selecciono o seguinte texto no Syntax Editor

```

RECODE
  ca3
  (0 thru 10=1) (11 thru 19=2) (20 thru 60=3) (61 thru Highest=4) INTO 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. Selecciono **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 funda 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

O *Variable Type* dialog box aparecerá para a variável **age**

5. Click no **Type...**
O *Variable Type : idade_gp* dialog box aparecerá.
6. Na próxima caixa para **Width**: escreva 1.
7. Na próxima caixa para **Decimal Places**: escreva 0
8. Se o próximo círculo para **Numeric** não está seleccionado, selecciono-o.
9. Click no **Continue**

Nós completaremos este procedimento no próximo passo dos passos.

Isto diz ao SPSS para Windows para mostrar a variável **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 de uma 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 categórica. 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**. Para adicionar o **Value Labels** use os seguintes passos:

- Você deveria estar ainda em *Variable View* do último passo dos passos.
1. Pode ver que a caixa na linha 12 de **idade_gp** na coluna de **Label..** 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 **Values..** e click
 4. Click em cima de pequena caixa ao lado directa na caixa de *None* 
 5. A caixa de diálogo *Value Labels* aparece. Tecla 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 **Add**
 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 **OK**
No Data window você verá que a **idade_gp** é agora mostrada com um dígito singular.
 10. Volta a *Data View*. Selecciona **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 querermos 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 selecciona **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 61anos 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:
- Use o ficheiro C-HH.SAV.
 - Recode **h1** em **hsize** usando o seguinte grupo: (1 thru 4) (5 thru 7) (8 thru Highest).
 - Adicione variavel label e value labels.
 - Corra **Crosstabs** nesta variavel por **district**.

Household size * DISTRICT Crosstabulation

			DISTRICT			Total
			MONAPO	RIBAUE	ANGOICHE	
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 seleccione a opção **Save As...**
3. Entre o nome do ficheiro "session1".
A extensão ".spo" será automaticamente adicionada ao nome.
4. Carregue no **Save**

Para sair do SPSS para o Windows:

1. Do **File** menu seleccione **Exit SPSS**
O dialog box dará a opção para salvar o conteúdo do
C:\sample\c-hh.sav
2. Carregue no **No**,
Um dialog box irá comanda-lo para salvar os conteúdos do Syntax
Editor para o Sintex1.
3. Carregue no **Save** dê um nome ao ficheiro, como por exemplo Module1.sps, daí
o SPSS para Windows 10.0 será terminado.

SPSS para Windows SAMPLE SESSION
Módulo 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 quiiogramas, 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. Selecciona a opção **File/Open...**
2. Selecciona 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  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 símbolos 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 
 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 código de produto iram ser usados. Este sub-amostra de dados estará em efeito até abrirmos 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 combinação 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. Carrege
4. Escolha **p1a** da lista sobre o New Working Data File: e carrege
5. Carrege
 Permir-lhe a fazer o rename de **p1a** para **unidade** para combinar com o ficheiro de conversão.
6. Próximo de New Name: escreva **unidade**
7. Carrege no
8. *Click* na caixa próxima do **M**atch cases on key variables in sorted files (Juntar os casos baseados nas variáveis chaves nos ficheiros ordenados).
9. Carrege no botão circular próximo do **E**xternal file is keyed table (O ficheiro external (não activo neste momento) é o ficheiro de referência)
10. Escolha **prod** do **E**xcluded Variables List (Lista de variáveis excluídas)
11. Carrege no proximo do **K**ey **V**ariables: (A baixo do direita)
12. Repita os passos 10 e 11 para **unidade**
13. Faça o **P**aste do comando
 Uma alerta apareceu dizendo-lhe que os ficheiros de dados devem ser ordenados. Já que os ficheiros estão ordenados com o **S**ort...
14. Carrege
 Um dialog box ira perguntar se pretende salvar o conteúdo dos dados no window. Não queremos salva-lo, o novo ficheiro pode tomar o seu lugar, entao...
15. Carrege
16. Escolha e faça o **R**un 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 necessarias 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 quilogramas 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 
- 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 **qprod_tt** como **Target Variable:** (para calorias totais)
3. **Type&Label**
3. *Click* no **Type&Label**
4. Tecla *Calorias Totais Produzidas* na caixa vazia ao lado de **Label** (etiqueta).
5. Carrega no **Continue**

3. Entra a equação **cprod_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 tivéssemos seleccionado **Create new data file** em vez de **Replace working data file**, o novo ficheiro resultante da agregação teria sido guardado no disco e não estaria a ser o nosso ficheiro em uso. Teríamos que 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. Selecciona **File/Open...**
2. Selecciona 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** seccione **Compute...**
O **Compute Variable window** irá aparecer.
2. Para **Target Variable**: entre **ae**
Type&Label
3. *Click no*
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. Seccione o botão circular para **Include if case satisfies the condition**
(Incluir se o caso satisfizer 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, simplismente repetem os passos acima indicados.

Numeric Expression	If... Statement
0.84	ca4 = 2 & ca3 >= 10 & ca3 <= 19
0.72	ca4 = 2 & ca3 >= 20
0.6	ca3 < 10

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. Transfere **ae** a Variables:
3. Pressione **Old and New Values...** (Antigos e Novos Valores)
4. Seleccione o botão circular próximo de **System-missing** (Omisso -- determinado pelo programa de SPSS)
5. Seleccione **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. Seleccione **Sum of values** (Somar os valores individuais)
10. **Continue**
11. Seleccione **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. Selecciona Match cases nas variáveis chaves 'key variables...'
5. Selecciona Both files provide cases (Todos os dois ficheiros contribuem casos)
6. Key Variables: são **distrito**, **ald**, e **af**, respectivamente
7. **Paste**, limpa os avisos, selecciona 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:** 4 (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 **Dependent List**:
3. Transfere **ncprod_a** para **Independent list: layer 1 of 1** (Primeira camada)
ncprod_a came from the **Rank Cases** procedure.
4. **Next**
5. Mova **distrito** para **Independent 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 Ribaue é 2517,4551. O resultado de **Compare Means** dá-lhe os números necessários para a tabela, mas contudo eles não estão formados exactamente como vimos na tabela do começo desta secção. Na Secção 3 aprenderá os procedimentos para conseguir os mesmos resultados, mas numa tabela com uma melhor apresentação.

Salve este ficheiro como **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 corrido 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. 
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 $\text{calorias retidas} = \text{calorias produzidas} - \text{calorias vendidas}$.
- 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		Excluded		Total	
	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”
Módulo 3: Importação dos dados de outros formatos --Séries Temporais**

**Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 10.0
3ª Edição**

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

Há vários métodos que podemos usar para introduzir dados no SPSS para Windows.

1. Pode-se digitar os dados directamente no programa.
2. Pode-se usar SPSS Data Entry para Windows, Versão 2.0. Um ficheiro de dados em SPSS é criado e podemos ler directamente em SPSS para Windows. As etiquetas e os valores das variáveis definidos no Data Entry são mantidos no ficheiro dos dados quando se abre o ficheiro em SPSS para Windows.
3. Os dados também podem ser lidos de uma variedade de diferentes formatos de ficheiros/ou pacotes. SPSS para Windows pode ler dados armazenados em Lotus 1-2-3 (*.w*) e Excel (*.xls) folha de cálculo (spreadsheets), base de dados em dBase (*.dbf), ficheiros em Systat (*.syd and *.sys), ficheiros em ASCII (*.txt) e ficheiros em versões antigas de SPSS para DOS (*.sys).

Ficheiros de Textos: Os dados armazenados nos ficheiros de texto num destes formatos – delimitado (tab, espaço, vírgula) “delimited (tab, space, comma)” ou largura fixada “fixed width” – podem ser lidos em SPSS usando o “Text Import Wizard”. O “wizard” vai guiar-lhe através das especificações necessárias para importar dados em formato de texto. Se se vai importar dados em texto regularmente, pode-se gravar o formato especificado para o nome de ficheiro para que não haja necessidade de especificar os parâmetros em cada vez. Pode-se colar os comandos a um ficheiro de sintaxe “syntax file” e reusar o syntax file sempre que necessário.

Nesta secção, lemos dados num formato *database* (*.dbf) porque é o formato usado por muitos sistemas de informação de mercado para digitar os dados brutos deles (um exemplo é SIM do Senegal (*Système d’informations de marché*)). Também lemos dados em folha de Lotus (*.wk1). Este ficheiro contem um índice de preços que usamos para calcular preços reais ou nominais na Secção 8 deste tutorial de séries temporais.

Abrir e gravar ficheiros de dados

O comando **Open /Data** do menu **File** vai abrir um ficheiro de SPSS existente como o ficheiro activo “working file” ou abrir o ficheiro noutra formato.

Para ler os dados de um texto ou ficheiro ASCII, seleccione **File /Read Text Data**. Apartir da caixa de diálogo (dialogue box) Open File, seleccione o ficheiro para abrir. O extensão por defeito para este tipo de ficheiro é *.txt. O Text Import Wizard dialog box vai aparecer para guiar-lhe através de 6 passos. Na parte inferior da dialog box, verá as primeiras filas do ficheiro. O primeiro ecrã vai pedir um nome de ficheiro dum formato predeterminado. Se não tiver nenhum (a resposta por defeito é não), faça um click do **Next**. O segundo ecrã vai pedir-lhe seleccionar um formato, Delimited or Fixed Width. A segunda pergunta vai pedir-lhe se há nomes de variáveis na primeira fila. A resposta por defeito é não. Os ecrãs seguintes guiam-lhe para as restantes especificações. Tem oportunidade de dar nome as variáveis ou usar

o *default*. Depois de alcançar o “passo 6” pode colar **Paste** o comando no ficheiro syntax para usa-lo posteriormente, ou gravar o “template” para um nome, ou pode apenas clicar em **OK**.

Se os dados estão no formato predeterminado que SPSS pode-se ler, seleccione **File /Open /Data**. No caixa de diálogo **Open File**, verá todos os ficheiro alistados com a extensão *.sav - o nome de ficheiro por defeito do dados SPSS para Windows. Pode-se seleccionar o ficheiro para abrir. Se o ficheiro está no directório diferente, pode-se ir a directório onde os ficheiros estão localizados. Use o Look in: ceta para baixo para seleccionar um “drive” /directório. Para mudar o tipo de ficheiro que quer abrir, faça um click na ceta para baixo” para Files of type:. SPSS não vai ler os dados correctamente a não tenha escolhido o tipo de ficheiro apropriado.

Como SPSS Lê dados de Folha de Cálculo (Spreadsheet)

O ficheiro de dados de SPSS é rectângulo. As fronteiras (ou dimenções) do ficheiro são determinados pelo número das observações (filas) e variáveis (colunas). Todas as celulas têm um valor, mesmo se o valor está em branco (o que chamamos valor omissio “system-missing”). As regras seguintes são aplicaveis para ler dados em formato de folha de cálculo.

- Filas são iguais as observações, colunas são iguais as variáveis.
- Os nomes das variáveis podem-se ler desde da primeira fila do ficheiro. Se o nome da variável for mais longo que 8 caracteres, o nome será cortado a 8 caracteres. Se os nomes das variáveis não forem unicos dentro dos 8 caracteres, SPSS vai criar um nome unico. Para Excel 5 ou os mais recentes ficheiros - o nome original é usado como etiqueta da variável (*variable name*) em SPSS.
- Etiqueta da variável - as versões mais antigas do Excel e Lotus – se a primerira fila não contiver nomes das variáveis, SPSS vai usar as letras das colunas (a, b, c, ...)
- Para Excel 5 ou ficheiros mais recentes, o tipo assim como a largura de cada variável são determinados pela largura e tipo de dado da coluna. Se a coluna contiver mais de um tipo de dado, o tipo de dado é posicionado no alfanumérico. Se a primeria fila não contiver nomes das variáveis, SPSS vai usar as letras da colunas (V1, V2, V3, ...)
- Para versões anteriores do Excel, o tipo do dado e largura de cada variável e determinado pela primeira celula na coluna. Valores de otros tipos são convertidos a valores omissos. Se a primeiro celula na coluna estiver em branco, o formato por defeito de tipo de dado para a folha (geralmente numérico) será usado.
- Para variáveis numericas, as celulas em branco são convertidas a valores omissos, indicado pelo ponto. Para variáveis alfanumericas, o espaço em branco é um valor alfanumérico valido, e as celulas em branco são tratadas como valores validos.

Exercicio 3.1

A informação contida neste folha de cálculo SNAP_PARTIAL.XLS é o que aparece no quadro de dados de SNAP sobre a produção de culturas ao longo de tempo em Mozambique, desde 1996/97 a 2001/02. Apenas o milho e mapira se encontram neste ficheiro. O ficheiro estão formatado de tal maneira que vai interferir com a importação e por tanto precisamos limpar o ficheiro.

Usando Explorer, localize o ficheiro SNAP_PARTIAL.XLS. Primeiro, tente abrir o ficheiro directamente no SPSS e note os problemas que incluem:

- 1) importar o intervalo inteiro da folha de cálculo inclui mensagens textuais (titulos, etc.) que não são dados;
- 2) dados para duas culturas estão localizados lado a lado de tal modo que a importação de ambos não podem-se percebidas e não há variável que indica a cultura;

- 3) olhando para a area do intervalo (A11:N24), há uma fila extra entre os nomes das variáveis e os dados;
- 4) nomes das variáveis que SPSS nao vai reconhecer como nome da variável (1996/97) por duas razões: a) não se pode ter um “/” simbolo num nome da variável, e b) não se pode começar o nome com um número;
- 5) a fila para a produção a nível nacional “National” é so um agregado dos dados provinciais e não está ao mesmo nível dos outros dados – portanto, não deve ser incluído;
- 6) há uma fila extra entre a ultima observação (National) e a obsevação anterior para a província de Maputo.

Precisa de limpar o ficheiro para cada um desses erros e gravar a **snap_partial_clean.xls**. Depois de limpar of ficheiro, repare no intervalo de dados, incluindo os nomes das variáveis modificadas dentro do intervalo. Introduza o ficheiro a SPSS. O ficheiro de syntax deve aparecer da seguinte maneira:

```
GET DATA /TYPE=XLS
/FILE='C:\Donovan\Moz\Docs\Training\snap_partial.xls'
/SHEET=name 'Sheet1'
/CELLRANGE=range 'A11:G23'
/READNAMES=on .
EXECUTE .
```

Na Secção 8, vamos importar um ficheiro de dados duma folha de cálculo para juntar a índice de preços e usar-los para como deflatores dos preços nominais.

Como SPSS Lê Ficheiros de dBase

Ficheiros de dBases

Os nomes das variáveis (“*field names*” no dBase) num ficheiro dBase são convertidos como nomes das variáveis. Estes nomes devem seguir as regras de como dar nomes de variáveis no SPSS. Se for mais longo que 8 caracteres, o nome será cortado. Se não for unico dentro dos primeiros 8 caracteres, o nome será apagado. Dois pontos são convertidos a “_”.

Observações marcados para apagar num ficheiro de dBase serão importados para SPSS. Uma variável é adicionado ao ficheiro chamado D_R. Se houver um “*” nesta observação, a observação é marcado para ser apagar.

Base de Dados - Microsoft Access


Para ler uma tabela de dados, deve-se instalar o “database driver” no computador para o tipo de base de dados que se vai usar. Pode-se instalar o “driver” apropriado do CD-ROM para instalações. O SPSS Data Access Pack contem uma variedade de formatos de base de dados. Microsoft Data Access Pack instala “drivers” para produtos de Microsoft.

Perguntas “Queries” são usados para ler dados de uma tabela. Pode-se gravar uma especificação de “*query*”. (SPSS vai usar a extensão .spq). Para acesar um base de dados, seleccione File /Open Database /New Query. O “Database Wizard” sai e vai guiar-lhe atraves do processamento para criar um “data source” e seleccionar o ficheiro, tabelas e variáveis que quer usar. SPSS vai modificar os nomes das variáveis de acordo com as regras de SPSS para nomes de variáveis – se forem validos os primeiros 8 caracteres, o nome será usado; caso não, SPSS vai dar um nome valido.

Exemplo usando dBase

Vamos ler um ficheiro em dBase. Usamos um ficheiro de SIM do Senegal (Sistema de Informaçao do Mercado) com dados de cereais de ano 1995.



1. Abra SPSS e click **File /Open...**
2. Verifique se está no directório C:\Sample. Click a ceta para baixo para *File of type:* e seleccione *dBase*.
*Deve se ver so os ficheiros que têm extensão *.dbf.*
3. Escolhe **cer95.dbf**
4. Click **Paste** e cole o comando no *Syntax Editor*. O *Syntax Editor* vai ser o ficheiro activo e verá

```
GET TRANSLATE
FILE='c:\sample\cer95.dbf'
/TYPER=DBF /MAP .
```
5. Click o botão  na barra de ferramentas.

Volte a *Viewer* e reveja a informação que saio no *Log*. Repare que muitas das variáveis em *dBase* eram mais longo que 8 caracteres e SPSS cortou-os a 8 caracteres. Repare também que duas variáveis foram dadas nomes das variáveis para SPSS (V27 e V28). A razão dada por esta mudança é o nome original foi uma palavra reservada o duplicada. Nestes casos, pode-se verificar que o nome foi duplicado. A variável *maz_impqt* foi cortado para *maz_imp_*. A variável *maz_imp_br* não pode-se cortar a *map_imp_* porque ja exitiu, portanto SPSS deu lhe o nome V26. A variável *maz_imp_en* tem o mesmo problema e foi dado o nome V27.

Agora, volte a *Data Editor* para ver os dados. Faça um click de *Variables View* no canto inferior esquerdo do extrã. Muitas das variáveis tem uma *Label* mas alguns não tem. Apenas as variáveis cujos nomes foram cortadas recembem um nome. O *Label* é o nome original da variável. Não há etiquetas de valores para muitas das variáveis. Valores omissoos “missing values” também não são definidos.

Agora, faça um click da *Data View* tab. A primeira variável (*d_r*) foi discutida antes como sendo uma variável que *dBase* usa para indicar se uma obsevarção e marcado para apagar. Quaisquer observações com “*” devem ser apagados deste ficheiro. Para verificar se existem observações nesta variável, corra **Frequencies** com esta variável.

1. **Analyze...Descriptive Statistics...** seleccione **Frequencies ...**
2. Seleccione **d_r** da lista na esquerad e click o  botão.
3. Click **Paste** para colar o comando no *Syntax Editor*. Volte a *Syntax Editor*.
4. Executa o comando clickando no botão **Run** .

Se não forem observações com “*”, podemos usar o **Select Cases** comando para seleccionnar so as observações onde *d_r* contém espaços em brancos

1. Apartir do *Data Editor*, seleccione **Data /Select Cases**
2. Seleccione **If condition is satisfied**
3. Click o **If...** botão directamente sob **If condition is satisfied**
4. Seleccione a variável **d_r** na caixa esquerda and mova para a caixa a direita.
5. Click dentro da caixa a direita do nome da variável e escreva
= " "
6. Click **Continue**
7. Seleccione o botão redondo ao lado do **Deleted** (para apagar as observações não escolhidos).
8. Click **Paste**
9. Escolhe o texto no *Syntax Editor* desde a linha **FILTER OFF** até a linha **EXECUTE** e corra o comando.

Syntax:

```
FILTER OFF .
USE ALL .
SELECT IF (d_r = " ") .
EXECUTE .
```

Quando um ficheiro é introduzido em SPSS que não é um ficheiro para SPSS para Windows, o nome do ficheiro não aparece em cima do ecrã na barra de título. Grave o ficheiro em formato do SPSS.

1. Asegure que o *Data Editor* é activo
2. Click **File /Save As...**
3. **Filename cer95**
4. Faça **Paste** e corra o comando.

Agora o nome do ficheiro vai aparecer na barra de título.

Logo que tirarmos as observações marcados para apagar, podemos apagar a variável *d_r* porque ocupa espaço no disco duro. Há dois métodos para apagar uma variável – manualmente e com syntax.

Método manual – a partir do *Data Editor*, click uma vez o nome da variável para seleccionar a coluna inteira e depois, tecla o botão “Delete” no teclado ou Selecciona **Edit /Clear** do menu.

Método do sintaxe –

1. **Data /Merge Files /Add Variables...**
2. Selecciona o ficheiro *cer95.sav* para Read File
3. **Open**
Todas as variáveis no segundo ficheiro serão apagados porque os nomes são iguais.
4. Selecciona *d_r* da lista das variáveis na *New Working Data File* caixa e mova para a caixa **Excluded Variables:**
Repare que não precisamos de variáveis chaves porque estamos a juntar o ficheiro a si proprio.
5. Faça **Paste** no *Syntax Editor* e corra o comando.

O syntax vai aparecer como:

```
MATCH FILES /FILE=*
/RENAME (d r = d0)
/FILE='C:\sample\cer95.sav'
/RENAME (d r date enqueteu mai imp mai loc mais dem mais det mais imp
mais pro marche mil deta mil gros mil imp mil prod mil quan paddy quantite
region riz imp riz loc riz loc riz quan sol demi sol deta sol prod
sor imp sor loc sorgo i v26 v27 = d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12
d13 d14 d15 d16 d17 d18 d19 d20 d21 d22 d23 d24 d25 d26 d27 d28 d29 d30 d31)
/DROP= d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10 d11 d12 d13 d14 d15 d16 d17 d18 d19
d20 d21 d22 d23 d24 d25 d26 d27 d28 d29 d30 d31.
EXECUTE.
```

Usando o menu o comando *MATCH FILE* acima é criado. Podemos simplificar isto eliminando o */RENAME* subcomando e adicionando o nome real da variável ao subcomando */DROP* subcomando, e.g.

```
MATCH FILES /FILE=*
/FILE='C:\sample\cer95.sav'
/DROP= d_r.
EXECUTE.
```

Vamos ver os dados de data na *Data Editor*. Vemos *****. Os asteriscos significam que a largura para mostrar os dados não é suficientemente grande para mostrar a data. Podemos aumentar a largura fazendo um click na linha a direita entre os nomes das variáveis e colocando a linha a direita. O formato da data é apresentado da forma dia-mes-ano. O formato de apresentação pode ser modificado para outros formatos. Faça um click da coluna “Type” na fila 4. Faça um click da caixa cinzeta a direita. Há muitos formatos disponíveis para data.

SPSS 10 ja tem funções para manipular datas que não eram disponíveis nas versões antigas do programa. Podemos calcular uma nova variável que contem so o dia, semana, mes ou ano para facilitar a reorganização de dados para solucionar os nossos diferentes necessidades analíticas.

Converter Variáveis de Formato Alfanumérico/Numérico para Variáveis Numéricas Numerados Consecutivamente

Variáveis alfanuméricas

Para especificos tipos de análise com SPSS (i.e., ONEWAY, ANOVA, MANOVA and DISCRIMINANT), variáveis alfanuméricas não podem ser usadas como um agrupamento independente da variável. O comando **Automatic Recode** vai criar uma variável numérica que agrupa os valores alfanuméricos e usa o valor alfanumérico como etiqueta do valor da nova variável numérica. Outra razão para usar **Automatic Recode** é que SPSS vai mostrar so os primeiros 8 caracteres duma variável alfanumérica. Se se quiser ver a informação completa, pode-se usar este comando para criar uma nova variável que vai conter os valores alfanuméricos como etiquetas dos valores e usar a nova variável no comando TABLES.

Valores alfanuméricos são recodificado em ordem alfabetico. Letras maiúsculas precedem letras minúsculas.

Variáveis Numéricas

A análise da MANOVA requer que os valores da variável numérica estejam em ordem consecutivo. Podemos usar **Automatic Recode** para recodificar os valores em ordem consecutivo. As vantagens da recodificação das variáveis em ordem sequencial é que os requerimentos da memoria serão reduzidos e o processamento melhorará. Quaisquer valores que não têm etiquetas dos valores definidos na variável anterior são dados os valores anteriores como etiquetas das valores da nova variável.

Os valores omissos são recodificados para novos valores omissos que são maiores que quaisquer valores existentes. Por exemplo, se a variável original tiver 10 valores não-omissos, o minimo valor para o valor omissos será 11, e o valor 11 seria um valor omissos para a nova variável.

Exercicio 3.2.

Para recodificar valores numéricos ou alfanuméricos em valores consecutivos inteiros, do menu, escolhe

1. **Transform /Automatic Recode...**
2. Seleccione uma variável na lista na esquerda, i.e. **marche** (Market) e mova a *Variables->New Name* caixa.
3. Na caixa a direita do **New name** botão, escreva o nome da nova variável que quer criar. Depois, click **New name**. O nome vai sair na caixa *Variables->New Name*. *O ordem por defeito é ascendente. Se quiser mudar o ordem para descendente, click o botão ao lado do Highest value na area de Recode Starting from.*
4. Cole a corra o comando.

Volte a Viewer e veja o ficheiro **Log**. Uma lista do valor anterior, valor novo e etiqueta estão mostrados. Ao ver o ficheiro, pode-se ver algumas das desvantegems em usar strings. Puxa para baixo a letra "T".

TABACOUNDA	55	TABACOUNDA
TAMABOUNDA	56	TAMABOUNDA
TAMBA	57	TAMBA
TAMBACOUNDA	58	TAMBACOUNDA

É possível que os 4 valores sejam o mesmo mercado. Se correr **Frequencies** a nova variável, qualquer valores menores que 10 serão possivelmente erros tipográficos.

Exercício 3.3 (Opcional)

Abra o ficheiro IPCMOZ.XLS e veja os dados. Tire as filas brancas, ponha os nomes das variáveis so numa fila. Salve o ficheiro com o nome IPMOZ_clean.XLS. Feche o Excel e abra o SPSS. Introduza o ficheiro XLS a SPSS e estude os resultados. Ainda vai encontrar problemas. Por exemplo, abra a Variable View na Data Editor. Há variáveis alfanuméricas onde devem ser variáveis numéricas. Repare que SPSS formata uma variável de acordo com o valor do primeiro dado na coluna. Portanto IPC89, com “na” seria considerado como string e não numérico. Mude o TYPE manualmente, seleccionando um formato com largura de 8 caracteres e dois decimais. Estude o resultado. O que aconteceu com “na”?

Uma pista: No DATA EDITOR, se uma variável está justificada a esquerda, é uma indicação que a variável está formatada como uma variável alfanumérica. Variáveis numéricas estão justificada a direita.

SPSS para Windows “SAMPLE SESSION”
Módulo 4 - Limpeza e Verificação dos Dados (Séries Temporais)

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 10.0
3ª Edição

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

Assim como os módulos anteriores, os módulos de séries temporais que se seguem referem-se a um auto-treinamento para introduzir os comandos necessários para um processo de análises de estudos em SPSS para Windows 10.0. De modo a utilizar-lo efectivamente, será necessário ter acesso a alguém que tenha bom conhecimento do SPSS para Windows, para lhe ajudar no início e para responder as perguntas que possam surgir ao longo da sessão enquanto você trabalha independentemente. Contudo, pretende-se parcialmente que este seja um instrumento completo de auto-treinamento, não obstante poder também ser usado como um guião para treinamento na sala de aulas.

Os dados usados nesta sessão de SPSS para Windows 10.0 foram retirados de dados de séries temporais semanais recolhidos pelo Sistema de Informação de Mercados Agrícolas (SIMA) no Mali. Foram escolhidos mercados múltiplos para representar vários tipos de mercados tais como o mercado do produtor, mercado por armazenista e mercado do consumidor (ou a retalho). Estão disponíveis dados de vários cereais mas para o propósito desta sessão, a ênfase será dada a mapira (*sorghum*).

Secção	SPSS para Windows Data File
Dados sobre a Quantidade e Preço para as épocas de 1989 á 1995	PQ_8995P.SAV
Dados sobre o Preço para as épocas de 1982 á 1989	P_8289P.SAV
QuattroPro File com dados sobre índice de preços para 1995	INDEXP.WK1
QuattroPro File com dados sobre a produção para taxas de crescimento	MALIPROD_P.WK1

Estas sessões também assume que:

- Você sabe como usar o Windows com o “rato”
- Os data files acima descritos estão armazenados no directório C:\DATA\SAMPLE no seu disco duro
- O programa é accionado para fazer uma lista de variáveis na mesma ordem que aparecem apresentadas no ficheiro (EDIT/OPTIONS/GENERAL/Variable List -- FILE)
- O programa é também accionado para fazer uma lista de comandos no janela (*window*) de VIEWER (EDIT/OPTIONS/VIEWER/Display commands in the log)


- Você tem um conhecimento de estatística básica (R^2 , estimativa de como o modelo se ajusta aos dados ou proporções explicadas neste modelo; **F** teste para verificar como a regressão se ajusta aos dados (Teste **t** para os estudantes e distribuição na raiz quadrada de **F** - permite testar a relação linear), variabilidade **ANOVA** e relação linear, e assim em diante).

P.S. Você pode modificar qualquer uma das preferências para ver as janelas diferentes usando **Edit/Options...** do sistema de *menu*. Lembre-se sempre de gravar as mudanças efectuadas nos dados após cada exercício e módulo, usando um novo nome do ficheiro. E, recorde-se de gravar o seu **OUTPUT** pois iremos usa-lo no módulo 10. É altamente recomendado gravar a listagem do **Syntax Editor window** e os resultados no **SPSS VIEWER**. Isto irá permitir que você faça a revisão do documento e interprete todas as análises efectuadas.

Ficheiro de Dados de Preços e Quantidades e o Ficheiro Activo

Os dados no SPSS para Windows 10.0 são armazenados em nos **ficheiros de dados**. Conforme visto em módulos anteriores, se pretendemos trabalhar com um conjunto de dados, devemos abrir o ficheiro de dados correspondente ou torna-lo activo, de modo que este esteja disponível para o programa. De novo, quando um ficheiro de dados está aberto, é transportado do disco para a memória, constituindo o ficheiro activo. Começemos com o ficheiro de dados que corresponde aos dados de preço e quantidade **PQ_8995P.SAV**. Para abrir este ficheiro de dados e constituir o ficheiro activo:

1. A partir do **File** menu, seleccionar **Open...**
2. Vai para o directório no qual se encontram os seus dados do *sample session* e seleccione o ficheiro **PQ_8995P.SAV**.
3. *Click* no **Paste** button para colocar o comando no Syntax window.
 O *Syntax window* irá tornar-se a "janela activa" e irá aparecer o seguinte texto

```
GET
FILE='C:\SAMPLE\PQ_8995P.SAV'.
EXECUTE .
```
4. Lembre-se que você poderá colocar o cursor em qualquer parte na linha que contem o comando "GET" e depois *click* o Run button  no **Toolbar**.

O ficheiro de dados de preços e quantidades para os anos de 1989 e 1995 agora está na memória.

Tal como no primeiro módulo, é importante saber em que ficheiro de dados está a trabalhar e as variáveis que contem. Podemos descobrir isso olhando para o ficheiro de dados assim como usando o comando **Variables...** no menu **Utilities**, seguindo os seguintes passos:

1. A partir do menu **Utilities** seleccione **Variables...**
2. Seleccione o nome da variável, e a informação acerca da variável irá aparecer no lado direito

Esta caixa mostra no lado esquerdo, a lista das variáveis encontradas no ficheiro de dados que cobre dados sobre a localização, os mercados, o tempo, os preços e a quantidade. E, ao se escolher uma variável específica, a caixa direita também mostra os valores das etiquetas (*labels*) para cada uma das variáveis, o tipo de variável, o tipo e largura de variável, o número de espaços decimais (se o tipo for Numérico), e qualquer valor omisso. Leve o tempo que quiser para verificar o número de mercados, de bens, de tipos de preços, por aí em diante.

Escolha a variável **p4c** que representa o preço do consumidor calculado. Irá verificar na caixa do lado direito, uma área de valores omissos de 997 à 999. Estes valores não são preços mas representam códigos para razões explicando quaisquer dados omissos de preços! As razões são abaixo demonstradas na caixa. É importante saber as razões da omissão dos dados. Novamente, estes códigos não são dados mas foram postos na categoria dos valores omissos da variável para algum propósito: então o SPSS para Windows irá evitar incluir e usar estes valores em qualquer análise que for efectuada.

Se desejar ter toda esta informação por escrito para a sua janela de VIEWER para uma posterior examinação, faça o seguinte:

Seleccione o menu **Utilities** menu e seleccione **File Info**.

Este comando será executado imediatamente. O Output window irá tonar-se activo e irá conter uma listagem de todas as variáveis com as suas definições

Esta informação está agora incluída no SPSS VIEWER. Fazendo o **File Info** e gravando esta informação acerca dos seus ficheiros é uma maneira de documentar os seus ficheiros de dados.

Click no botão de  quando acabar

Outros tipos de erros podem ser encontrados nos dados. Seleccione o ficheiro e leve alguns minutos para ver os dados. Para ver a natureza (estrutura) de qualquer variável específica, poderá fazer o *click* duas vezes no nome da variável na parte de cima da tabela na janela de DATA VIEW. Ao olhar para os dados, irá reparar muitos valores omissos que podem ser repostos por um período. Estes referem-se a valores omissos do sistemas, também denominados por SYSMIS. É importante compreender correctamente os dados. Contudo, a razão pela qual a maioria dos dados aparentam estar omissos é pelo facto de o ficheiro trazer consigo dados de diferentes tipos de mercados, por exemplo, os mercados do produtor não podem incluir dados do mercado por armazenista, por aí em diante. Os dados podem parecer omissos, mas concretamente, não existem nenhuns dados omissos dado não existirem nenhuns preços para mercados não existentes! Assim, é possível encontrar um SYSMIS nos dados quando os valores foram recolhidos e inseridos. Este é um verdadeiro erro e não é permissível. Porque haveriam então de existir valores omissos e onde estará o valor real? Isto deve ser explicado e não é permitido.





Antes de efectuar alguma análise nos dados apresentados, é sempre bom verificar se os dados contêm erros. Estes erros podem ser originários de dados mal inseridos ou observações mal efectuadas do campo ou do mercado. Outro tipo de erro é lógico para o mercado de preços. Os exercícios de limpeza (*cleaning exercises*) permitem-nos fazer uma verificação dos dados e para esta sessão, iremos verificar o aspecto lógico do mercado de preços da mapira.

A Lógica do Mercado dos Preços - Produtor, Por Armazenista e Retalhista (Consumidor).

O preço do produtor (**p4p**) significa o preço do bem que foi pago pelos comerciantes ou técnicos de montagem, para o produtor num dado mercado. O preço por armazenista (**pr8**) significa o preço recebido por um comerciante por um dado bem por um comprador do bem no mercado (denominado nível de montagem nos dados). O preço retalhista (**p4c**) significa o preço pago pelos consumidores a comerciantes retalhistas (por exemplo por quilograma). As unidades locais de medida podem variar em tamanho e de local a local, e poderia ser necessário converter estas medidas para quilogramas ou unidades métricas. Como já terminou esse exercício no módulo anterior, nesta sessão as unidades locais já foram convertidas em unidades métricas. E em Mali,

os inquiridores fazem esta conversão por forma a minimizar o comprimento dos transmissores de dados via rádio.




Como pretendemos limpar os dados sobre a mapira, devemos escolher os valores relativos a este bem. Para fazer isso, podemos usar o comando **Select**.

1. Primeiro, active a janela de DATA EDITOR. Caso contrário, poderá fazê-lo com o *click* no botão  do Toolbar. O ficheiro PQ_8995P.SAV será aberto.
2. A partir do **Data** menu seleccione **Select Cases** e seleccione o botão circular mais perto de If **C**ondition is Satisfied
3. *Click* no  no If **c**ondition is satisfied
4. *Click* **cer** na coluna à esquerda, e depois , *click* o sinal de igual “=” e depois *click* uma vez no valor **8**.
5. O texto na caixa deverá ser parecido com: CER = 8
6. *Click* no 
7. Seleccione o radio button perto de **F**iltered
8. Faça o paste, active a janela de SYNTAX , Ponha o cursor em frente de USE ALL e **R**un (corra) o documento **T**o End.

Vários erros lógicos do mercado de preços podem ocorrer. Para identificar erros de limpeza (*cleaning errors*) específicos devemos primeiro criar uma nova variável. Para fazer isso:

1. Vá para Data View window e a partir do **T**ransform menu seleccione **C**ompute...
2. Por baixo de Target Variable: tecla **limpa** na caixa.
3. Insira o valor 0 na caixa (*box*) **N**umeric **E**xpression
4. Seleccione o **P**aste, e **R**un o comando.

Devemos agora identificar os tipos de erros de limpeza possíveis. Existem três grupos de preços então devemos calcular três diferentes erros, (1) o preço do produtor é superior ao preço do consumidor, (2) o preço do produtor é superior ao preço por armazenista e (3) o preço por armazenista é superior ao preço do consumidor. Para programar isto no SPSS, deverá:

1. A partir de **T**ransform seleccione **C**ompute...
2. Para a **T**arget Variable: a palavra **limpa** já deverá aparecer, caso contrário, deverá inseri-la.
3. *Click* 
*Uma caixa de diálogo (dialog box) irá aparecer. Poderíamos ter teclado a etiqueta da variável quando calculamos o **limpa** como 0 mas acabamos de aprender o que a variável representa. Devemos também adicionar o value labels nos três tipos de erros que iremos cometer mais adiante.*
4. Na caixa do **L**abel: insira o texto “Tipos de Erros Lógicos de Preços”.
Click 
5. Na caixa **N**umeric **E**xpression, insira **1**
6. *Click* no 
7. Seleccione o botão circular para Incluir **i**f case satisfies condition:

8. Insira a declaração **p4p > 0 & p4c > 0 & p4p > p4c**
9. *Click* no
10. Faça o **Paste** do comando mas **NÃO** corra o comando, ainda.
*Talvez receba uma mensagem que disse "Change existing variable?" (Muda a informação que já está lá). Se receber, responde **OK** para guardar estas novas definições.*
11. Não precisa usar sempre o sistema de *menu*, poderá fazer o copy e fazer o paste do sintaxe dentro do Syntax editor e depois repor os nomes das variáveis. Ou, se estiver familiarizado com o sintaxe poderá tecla-lo directamente. Assim, numa nova linha, mesmo por baixo da declaração **If (...) limpa=1** no Syntax editor, escreva mais dois comandos **IF** com as seguintes variáveis e expressões numéricas (*numeric expression*):

Numeric Expression	If... Statement
2	p4p > 0 & pr8 > 0 & p4p > pr8
3	p4c > 0 & pr8 > 0 & pr8 > p4c

12. Queremos adicionar **Value Labels** (etiquetas de valores) para cada um dos erros. No Data Editor vai o **VARIABLE VIEW** ou fazer o *click* na esquina direita de caixa para a variável limpa no coluna de **Values**.
13. Vá para **Value**: na secção **Value Labels** e tecla "1"
14. Pressione <Tab> uma vez e tecla "Preço de Produtor é superior ao preço de consumidor" na caixa de **Value Labels**.
15. *Click* no , ou pressione as teclas "Alt" e "A" no teclado.
16. Repita os passos 13 até 15 usando a informação seguinte:

Value:	Value Label:
2	O preço do produtor é superior ao preço de montagem
3	O preço de montagem é superior ao preço do consumidor
17. *Click* no .

Agora pretendemos ver se existem alguns erros nos dados. Para fazer isso, poderá accionar o comando **Frequency** (frequências) na variável **limpa**.




1. Será necessário seleccionar **Analyze/Descriptive Statistics/Frequencies...**
2. Selecciona a variável **limpa** e tecla
3. Faça o **Paste** e **Run**.

Vá para o SPSS Viewer e olhe para os resultados: foram encontrados 8 erros! Em todos eles, não existiam erros onde o preço do produtor é superior ao preço do consumidor, 7 erros onde o preço do produtor é superior ao preço de montagem e 1 erro onde o preço de montagem é maior que o preço do consumidor. Um passo lógico é recuar e olhar para o questionário e também falar com os inquiridores.

Existem muitas razões possíveis para a presença desses erros e não podemos modificar os dados para se adequarem ---a coisa mais simples a fazer é evitar usar esses valores em análises posteriores. Uma das formas para fazer isso é instalar um filtro (**Filter**) que irá excluir todos os

valores das variáveis **limpa** acima de 0. O método para programar o filtro é o mesmo que usamos quando seleccionamos a mapira para as análises de limpeza.

Vejam agora com maior detalhe esses erros. Usando o comando **Case summaries**, podemos ver os preços, mercados e os números dos casos para cada um dos erros de limpeza. Mas só pretendemos ver os erros de limpeza então devemos especificar não fazer a listagem de todos os casos. Podemos fazer isso repondo o filter com um novo para ver só os erros de limpeza. Para fazer isso, deve:

1. A partir do **Data** menu seleccione **Select Cases** e seleccione o botão circular perto de **If condition is satisfied**
2. *Click* no  no **If condition is satisfied**
3. *Click* primeiro o **&** e depois o **limpa** na coluna á esquerda, depois *click* no , *click* no sinal de maior “>” e depois *click* uma vez no value 0. O texto na caixa deverá ser como: CER = 8 & Limpa > 0
4. *Click* no 
5. Seleccione o botão circular perto de **Filtered**
6. Faça o paste do comando.
7. Seleccione **Reports** do **Analyze** menu
8. Seleccione **Case Summaries...**
9. Seleccione **mer cer ano mes dia p4p pr8 p4c limpa** da lista à esquerda e *click* no perto de **Variable(s)**: (Sugestão: Pode manter pressionada a tecla “Ctrl” para seleccionar todas as variáveis ao mesmo tempo).
10. *Click* na caixa perto de **Show case numbers**
11. Faça o **Paste** e **Run** ambos comandos.
- 12.

Agora poderá ver (na página seguinte) os erros que cometeu, os preços, em que mercados e em que números de caso (case numbers). Poderá verificar que os dois últimos valores para o preço do consumidor (**p4c**) estão em falta. Isto é explicado pelo facto de as observações terem sido feitas in Zangasso, mercado número 75, que não é um mercado do consumidor --- e não foram tirados nenhuns preços. Isto não influencia de nenhuma forma os erros de limpeza (cleaning errors). Estes erros são baseados em preços observados. O problema com estes erros poderá também estar entre **pr8** e **p4p**, o mais recente sendo maior que o antigo.

Case Summaries^a

	Case Number	Mercado	Cereal	Ano	Mês	Dia	Preço médio calculado de produtor	Préco de montagem (armazenista) pesado	Preço médio calculado do consumidor	Tipos de Erros Lógicos de Preços
1	25691	KOUTIAL A	Mapira	1993	Fevereiro	18	40,00	39,70	48,00	Preço do produtor é superior ao preço de montagem
2	25693	KOUTIAL A	Mapira	1993	Março	4	40,00	39,80	48,00	Preço do produtor é superior ao preço de montagem
3	25705	KOUTIAL A	Mapira	1993	Maió	27	48,00	47,60	57,00	Preço do produtor é superior ao preço de montagem
4	25739	KOUTIAL A	Mapira	1994	Fevereiro	10	38,00	43,40	41,00	Preço de montagem é superior ao preço do consumidor
5	25758	KOUTIAL A	Mapira	1994	Junho	23	45,00	44,90	60,00	Preço do produtor é superior ao preço de montagem
6	25775	KOUTIAL A	Mapira	1994	Outobre	20	44,00	43,73	54,00	Preço do produtor é superior ao preço de montagem
7	26994	ZANGASS O	Mapira	1992	Novembr o	6	37,00	,	,00	Preços negativos!!!
8	27020	ZANGASS O	Mapira	1993	Maió	14	40,00	32,00	,	Preço do produtor é superior ao preço de montagem
9	27086	ZANGASS O	Mapira	1994	Agosto	19	40,00	33,20	,	Preço do produtor é superior ao preço de montagem
Total	N	9	9	9	9	9	9	8	7	9






a. Limited to first 100 cases.

Exercício 4.1.

Repita o mesmo exercício de limpeza usando o milho como produto. Use todos os casos e depois use os filtros para seleccionar o milho. Também precisa calcular uma nova variável para o milho. Corra **Frequencies** e faça uma lista de todos os erros. Deverá obter 14 erros ao todo: 9 erros do tipo número 2 e 5 erros do número 3.

Outliers (Valores Extremos)

Podemos usar o comando **Frequencies** para ver *outliers* mas este produz listas longas e não obtemos uma boa distribuição dos dados. A outra forma de explorar os dados para preços extraordinários ou valores extremos é **olhar** para os *outliers*, tal como no primeiro módulo. O uso desta estatística descritiva irá ajudar-nos a encontrar os maiores e menores valores nos dados. Também podemos ver a distribuição dos dados assim que poderemos produzir um **Stem and Plot** (Caula e Folha) usando o mesmo comando.

1. Reponha o presente **Filter** seleccionando a **mapira** como o produto para análise e seleccione o comando **Explore** nos três preços da variável (**p4p**, **pr8** e **p4c**) usando os seguintes passos:
2. A partir do **Analyze** menu seleccione **Descriptive Statistics/Explore...**
3. Seleccione **p4c** da lista á esquerda e click no botão  perto do **Dependent List**.
4. No canto inferior esquerdo do caixa de diálogo (*dialog box*) existe uma caixa denominada **Display**. *Click* no botão circular perto do **Both**.
Isto irá dar-nos estatísticas e gráficos (plots).
5. Depois *click* no .
Isto irá abrir o Explore: Statistics dialog box.
6. *Click* uma vez no quadrado perto do **Outliers** e ponha um **X** na caixa .
Irá verificar que já existe um X na caixa perto de Descriptives.
7. *Click* no  button.
Isto irá leva-lo de volta para o Explore dialog box.
8. Depois *click* no  button.
9. *Click* no botão circular **None** no Boxplots box display.
10. *Click* no  button.
11. *Click* no **Paste** para pôr o comando no Syntax Editor e Run o comando.
12. Repita separadamente para os preços do produtor e por armazenista.

Se olhar para o SPSS Viewer, irá ver estatísticas descritivas e na distribuição do **Stem-and-Leaf** dos preços da mapira assim como os cinco valores mais altos e mais baixos para cada preço. Assim poderá dizer se tem alguns *outliers* extremos. Eles serão identificados pelos seus números de casos. Para a mapira foi encontrado um outlier com o valor 775. Como todos os outros valores estão abaixo de 150 FCFA, é provável que a pessoa que inseriu os dados teclou o 7 duas vezes, então podemos ignorar este valor. Um outro procedimento recomendado é verificar os questionários e ver que o valor real é 75 FCFA, e depois poderá escrever 75, repondo 775.

Extreme Values

			Case Number	Value
Preço médio calculado do consumidor	Highest	1	1057	775,00
		2	13295	150,00
		3	1177	150,00
		4	18562	150,00
		5	1175	, ^a
	Lowest	1	26994	,00
		2	26844	30,00
		3	26845	30,00
		4	25528	34,10
		5	26846	, ^b

a. Only a partial list of cases with the value 150 are shown in the table of upper extremes.

b. Only a partial list of cases with the value 35 are shown in the table of lower extremes.

1. Uma forma para eliminar esses preços extremos poderá ser declarando os Valores Omissos ou **Missing Values** fazendo o *click* nas esquina direita de caixa da variável na tabela (**P4C**) na coluna de Missing no Variable View do Data editor.
2. Aí poderá escolher inserir **Discrete (especifico) missing values**, um Range plus one optional discrete missing value (Série mais um valor omissos discreto, seleccione esta opção e insira 997 para *Low*: 999 para *High*: e 775 para **Discrete value**.
3. *Click Ok*. Esta é a forma pela qual os códigos 997 to 999 foram previamente inseridos. Isto irá permitir não incluir esses valores extremos em análises posteriores.
4. **GRAVE** os seus novos dados agrupados para o PQ_8995P2.sav e o sintaxe para o seriet4.sps.

Exercício 4.2.

Abra o mesmo exercício para ver os valores extremos para o milho. Poderá também correr o comando **Frequencies** para fazer a comparação. Irá encontrar um valor extremo. Tente inserir uma área de valores omissos. (Sugestão: procure por um preço do consumidor igual a 380 FCFA e agrupe os valores omissos para 380 e acima). Uma vez inseridos os valores omissos, accione outra vez o comando **Explore**: irá verificar que a média do preço alterou! Não obstante não se encontrarem nenhuns *outliers*, poderá também accionar estes exercícios nos preços do produtor e por armazenista. Certifique-se de gravar as alterações feitas aos dados no novo ficheiro PQ_8995P2.sav.

SPSS para Windows “SAMPLE SESSION”
Módulo 5 - Análises Básicas
(Séries Temporais)

Material de Treinamento para Curso de Curta Duração
Desenho de Investigação, Processamento de Dados e Análise
Utilizando SPSS para Windows 10.0
3ª Edição


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

As análises básicas proporcionam-nos os instrumentos para interpretar os dados de forma simples e rápida. Usando comandos elementares podemos preparar gráficos e tabelas que nos permitem representar, analisar e difundir informação sobre mercados de forma rápida e efectiva. O propósito deste módulo é de ensinar como usar esses instrumentos e como preparar relatórios.



Os exercícios que se seguem são alguns exemplos de como a informação sobre o mercado pode ser representada e preparada com base em análises básicas. Não obstante existirem muitas formas para representar os dados e análises estatísticas, os exercícios que se seguem incluem: uma tabela e um gráfico sobre percentagem semanais de mudanças, agregação aos níveis semanais e mensais, criar o gráfico de preço médio, e ajustar os valores para desenvolver índices sazonais agregados por forma a elaborar um gráfico por época agrícola ou por ano de produção.

Mudança percentual da semana anterior

Por forma a se conseguir observar mudanças entre semanas, devemos primeiro calcular a diferença em preços e depois calcular a percentagem. Podemos depois representar essas mudanças num gráfico ou tabela. Para calcular a diferença e percentagem para o preço do consumidor, seguem-se os seguintes passos:

1. Se a sua janela de SYNTAX EDITOR do último módulo ainda estiver aberto, feche-o de modo a iniciar um novo para este módulo. Se a janela de DATA EDITOR ainda não estiver aberto, torne-o activo seleccionando o ficheiro **PQ_8995P2.SAV**, faça o **Paste** e **Run**. Pretendemos trabalhar sobre a mapira, então deve organizar um filtro tal como no módulo 4 (Accione **Select Cases...** do **Data** menu, **If... cer = 8**, e por aí em diante).
2. Crie uma nova variável usando o comando **Transform/Create time series....**
3. Selecciona **p4c** da lista á esquerda e **Click** no botão  perto da caixa de **New Variables(s)**. Irá verificar que **p4c_1=DIFF(p4c 1)** aparecerá na caixa. Isto irá calcular a diferença nos preços entre dois casos consecutivos.
4. Faça o **Paste** e **Run** o comando.
5. Se olhar para a janela de DATA EDITOR irá verificar que os valores são desligados um de cada vez. Isto significa que devemos ajustar os valores ás semanas correspondentes. A forma para fazer isso é usar o mesmo comando **Create time series** (criar uma série temporal) usando a função **LEAD** para repor o valor correcto temporalmente (para cima um caso em termos de época--veja a explicação abaixo para LAG e LEAD).

Abra a caixa de diálogo (*dialog box*). Altere a função fazendo o *Click* na seta virada para baixo perto do Function *dialog box* com a palavra **Difference** (Diferença calculada) dentro. Escolha **Lead**. Selecciono o primeiro **p4c_1** para regressar a coluna á esquerda. Depois selecciono **p4c_1** da lista á esquerda





e *Click* no botão  perto da caixa de New Variables(s) outra vez (isto irá alterara a função). *Click* em cima de frase **p4c_1_1=LEAD(p4c 1)**. Altere o nome da variável de **p4c_1_1** para **p4c_d** (para diferenciar) na caixa de Name e *Click* no botão de  **Change** (Mudança).

6. Faça o **Paste** e **Run** o comando.

Lag (retardamento). Valor de um caso anterior, baseado numa ordem específica do lag. A ordem é o número de casos anterior ao caso corrente no qual o valor é obtido. O número de casos com o sistema de valores omissos no início das séries é igual ao valor da ordem.

Lead (dianteira). Valor de um caso subsequente, baseado numa ordem específica principal. A ordem refer-se ao número de casos após o caso corrente no qual o valor foi obtido. O número de casos com o sistema de valores omissos no fim das séries é igual ao valor da ordem.

Agora queremos calcular a percentagem de mudança entre duas semanas. Devemos depois calcular uma nova variável.

1. A partir de **Transform** selecciono **Compute...**
2. Se ainda tem a informação das operações, *Click* no  **Reset**
3. Para o Target Variable: insira (enter) **pctchg**
4. *Click* no  **Type&Label...** e insira a etiqueta “Mudança percentual”.
5. *Click* no  **Continue**
6. Na Numeric Expression box, insira **(p4c_d / p4c) * 100**
7. Se existir algo perto de botão de IF... abaixo, a condição do módulo anterior ainda existe e você deve repô-la. *Click* no botão de IF..., selecciono o botão circular perto do **Include all cases**, *Click* no  **Continue**.
8. Faça o **Paste** e **Run** o comando.





Vejamos agora estudar mudanças percentuais para três mercados durante quatro semanas. Outra forma de seleccionar dados específicos, sem somente usar o comando **Select** (para filtros (selecção temporária) ou para eliminar dados não-seleccionados permanentemente) é usar ao mesmo tempo os comandos **Temporary** e **Select** . O comando **Temporary** sinaliza o inicio das transformações temporárias que somente serão levadas a cabo no procedimento seguinte. Serão levadas a cabo até ao comando seguinte que leia os dados. Assim, poderá incluir a manipulação de dados tal como no exemplo abaixo (isto é, a transformação será aplicada para seleccionar dados e para outro comando tal como uma tabela, um gráfico, ou regressão e assim em diante -- quer dizer os comandos que actualmente lêem os dados e não somente modificam-os; veja o anexo para mais informações). Este comando não está disponível através dos menus então devemos tecla-lo no syntax editor. Active o syntax editor e tecla e accione o seguinte comando:

```
Temporary.  
Select if (sem = 20 or sem =21 or sem =22 or sem =23).  
Select if (mer = 3 or mer = 10 or mer = 41).
```

Selecionamos temporariamente os dados que iremos representar na tabela. Quando a tabela for elaborada o comando **Temporary** já não será mais efectivo. Usando **Tables** poderá calcular várias estatísticas e apresenta-las numa variedade de formas e estejam completamente sob o seu controle. Diferente de outros procedimentos do SPSS para Windows, o **Tables** permite fazer o seguinte:

- escolher como pretende que estejam as variáveis de manutenção e estatísticas para serem apresentadas em linhas, colunas, camadas e/ou concatenações.
- manipular estruturas das tabelas, conteúdo, e apresentação do formato.
- apresentar tabelas múltiplas na mesma apresentação (concatenar) e aninhar-se (*nest*) sub-tabelas múltiplas em qualquer dimensão.
- incluir percentagens flexíveis, especificando a base para as percentagens (seu denominador) de modo a serem possível atingir 100% quando somar entre linhas, colunas, sub-tabelas, ou todas as tabelas.
- fazer uma apresentação até 60 caracteres para etiquetas de variáveis e etiquetas de valores.

Para criar a tabela com as suas preferências (*customize*) usando o comando **Basic Table** (tabela básica) vá para:


1. **Analyze/Custom Tables/Basic Tables...**
2. Altere **pctchg** para **S**ummaries:
3. Altere **mer** para **D**own:
4. Altere **sem** para **A**cross:
5. *Click* no 
6. Na caixa de Summary Variable Labels, *Click* o botão circular perto do **A**cross the top
7. 
8. 
9. Na caixa de **T**itle (título) tecle: **Mudanças semanais para três mercados do Mali ao longa de quatro semanas**
10. 
11. Faça o **Paste** e **Run**.

Olha para as variáveis na tabela! Podemos ver as mudanças nos preços de uma semana para outra. Veja como as mudanças variam de mercado a mercado! Vamos agora analisar os mercados do produtor usando o gráfico. Podemos seleccionar dois mercados do produtor diferentes teclando no **Syntax EDITOR**:

```
Temporary.
Select if mer = 41 or mer = 75.
```

Agora você pode desenhar (*plot*) um gráfico seguindo os seguintes passos:

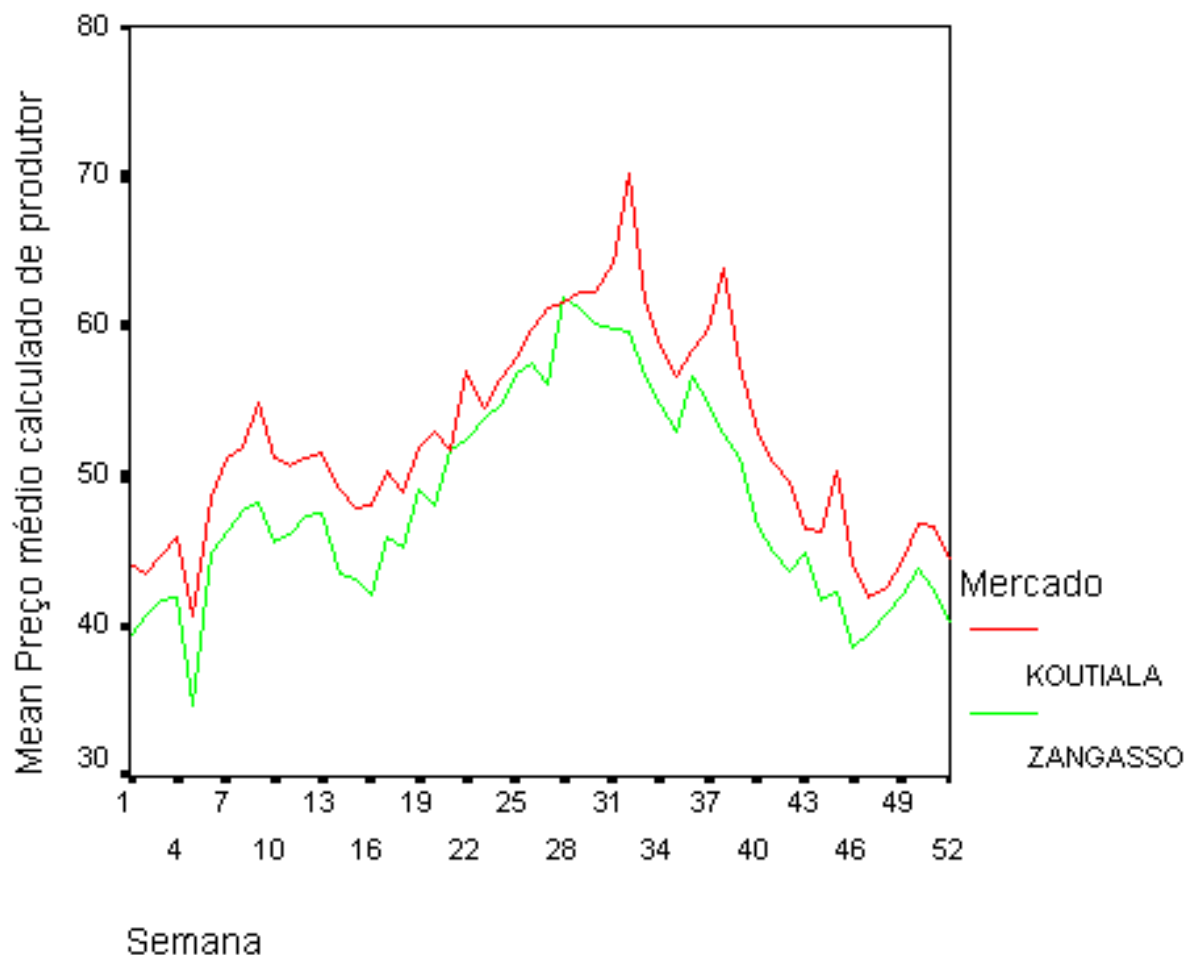
1. **Graphs/L**ine...
2. *Click* no quadrado perto do **M**ultiple

3. *Click* no **Define**
4. *Click* no botão circular perto de **Other summary function** na caixa de **Lines represents** box.
5. Agora escolha a variável **p4p** referente ao preço do produtor na coluna á esquerda e *Click* no  perto do **Variable:** e valor médio do preço do produtor será enquadrado no gráfico.
6. Escolha **sem** para **Category Axis:**
7. Escolha **mer** para **Define Lines** por:
Click no **Options...**
8. Assegure que a caixa circular perto do **Display groups defined by missing values** (Mostre os grupos definidos por valores omissos) **não** é seleccionado (*click* para deseleccionar), isto irá eliminar o valor do gráfico que representa o médio de todos os valores omissos.
9. **Continue**
10. Faça o **Paste** e **Run** o comando.

Observe o gráfico: veja como ambos mercados seguem as mesmas tendências mas os preços da mapira são altos para Koutiala! Guarde os seus resultados no SPSS Viewer usando como nome do ficheiro seriet5.spo

Mudanças semanais para três mercados do Mali ao longa de quatro semanas

		Semana			
		20,00	21,00	22,00	23,00
		Mudança percentual	Mudança percentual	Mudança percentual	Mudança percentual
Mercado	BADALA	,19	-1,27	3,87	1,66
	DIBIDA	-,84	2,37	1,07	1,74
	KOUTIALA	1,92	2,57	4,38	2,81



Exercício 5.1





Observe os mercados do consumidor. Para fazê-lo, poderá seleccionar qualquer um dos mercados do consumidor tecendo o comando **Temporary (Select if)** acima apresentado. Para este primeiro exercício, escolha os mercados 3-10-20-21-51-56-70 (não execute ainda o comando). Agora faça o **Paste** do comando **Graph** mas substitua a variável do **preço do produtor** com a variável do preço do consumidor **p4c**. Faça o **Paste** e depois **Run** ao mesmo tempo ambos comandos. O que observa?

Podemos verificar que os preços seguem a mesma tendência ao longo do tempo. Podemos também notar como todas as curvas estão muito próximas. Uma das formas para evitar que isso aconteça é separar as curvas usando o comando **Split file** (Dividir o ficheiro), como se segue:

1. Copie e faça o paste dos comandos para seleccionar os mercados.
2. Vá para **Data/Split File...**
3. **Click** no botão circular perto do **Compare groups** (Comparar grupos)
*Poderá também escolher **Organize output by groups** (organizar os resultados por grupo) que terá a mesma função que no caso de **Compare groups** diferenciando-se no resultado mostrado (output): neste exemplo, os grupos split-file são representados de forma conjunta de forma a permitir uma comparação. Ao se organizar o output por grupos irá permitir a apresentação de output separados para cada grupo de split-file, não como gráficos apresentados conjuntamente em um único output.*

4. Escolha **mer** da coluna á esquerda e coloque na caixa de **Groups based on:** (grupos baseado em)
5. *Click* no botão circular perto de **File is already sorted** (Ficheiro já está ordenado)
6. Faça o paste mas não accione o comando.

Agora faça o paste do comando **Graph**. Como pretendemos observar cada gráfico separadamente, certifique-se de seleccionar a caixa (*box*) perto de **Simple**. Use os seguintes passos:

1. **Graphs/Line...**
2. *Click* no quadrado perto de **Simple**
3. *Click* no 
4. *Click* no botão circular perto de **Other summary function** (Outra função sumária) no **Lines represents** (Linhas representam) box.
5. Agora escolha a variável **p4c** do preço do consumidor na coluna á esquerda e *Click* no  perto de **Variable:** e o valor médio dos preços do consumidor irá aparecer no gráfico.
6. Escolha o **sem** para **Category Axis:**
7. *Click* no 
8. Assegure que a caixa circular perto do **Display groups defined by missing values** (Mostre os grupos definidos por valores omissos) **não** é seleccionado, isto irá eliminar o valor do gráfico que representa o médio de todos os valores omissos.
9. 
10. Faça o **Paste** do comando.
11. Vá para o SYNTAX EDITOR e observe os comandos. O comando **Split file** irá permanecer activado até a função ser retirada. Como pretendemos direccionar este comando somente para a função do gráfico, podemos ou teclar **SPLIT FILE OFF.** numa nova linha que segue o fim do comando **graph** no SYNTAX EDITOR ou podemos seleccionar **Analyze all cases, do not create groups** do comando **Split file** e fazer o **Paste** para o SYNTAX EDITOR e Run..

Agregação em níveis mensais e apresentação gráfica do preço médio

Uma outra forma de apresentar os dados ou um grupo de dados para análise é agregando os dados para outros níveis. Por exemplo, se pretendemos observar um preço médio para Bamako que é composto por muitos mercados, podemos observar os preços por localidade (*localité* em francês) que agrupa os mercados permitindo-nos obter um valor médio para a área. Vejamos o preço médio de Bamako.

1. Vá para o Syntax editor e tecle:
Temporary.
Select if loc = 3.
2. Faça o **Paste** do comando para um gráfico de linha simples para preços do consumidor por semana.

Nota: Para aqueles que estão confortáveis a manipular o syntax, poderá copiar e fazer o paste de um comando **graph** anterior, directamente do Syntax editor. Poderá sempre assim fazer a qualquer momento para módulos posteriores do “sample session”. Irá verificar que assim se torna



mais rápido que se tornar familiar com o syntax. Lembre-se de alterar as variáveis no comando e seja cuidadoso em fazer o copy da escolha correcta das três opções possíveis do gráfico.

Exercício 5.2



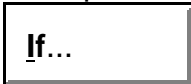
Repita os passos 1 e 2 mas substitua os gráficos semanais por gráficos mensais (i.e. **mes**). Pratique usando o comando **Split file** por localidade (**loc**). Lembre-se de teclar e accionar SPLIT FILE OFF. quando terminar de usar a função **Split file** (poderá também fazer o **Paste** através do menu). Poderá também praticar estes comandos usando outros produtos tais como, por exemplo, o milho. Guarde qualquer gráfico que seja do seu interesse. A parte /Missing Report do comando é desnecessária pelo facto de somente apresentar grupos definidos por valores omissos, sendo estes de pouco interesse.

Ajustamento de valores para índices sazonais agregados de modo apresentar graficamente por época agrícola ou ano de produção

A sub-secção que se segue irá permitir que modifiquemos a apresentação dos dados escolhendo o mês de Novembro como o primeiro do ano agrícola ou de produção ao invés do mês de Janeiro tal como o calendário anual. Para fazer isso, precisamos calcular novas variáveis para o mês e ano ajustados para época agrícola. Mas primeiro, devemos calcular a variável mensal e anual concatenada para identificar cada período de tempo representado nos gráficos. Siga os seguintes passos:

1. A partir de **Transform** seleccione **Compute...**
2. Para o **Target Variable:** tecla **data_m**
3. *Click* no  e insira a etiqueta “Mes_Ano”.
4. *Click* no 
5. A partir de caixa de **Functions:** seleccione a função **DATE.MOYR(month,year)** e coloque na caixa de **Numeric Expression.**
6. Substitua ambos **?,?** da função por **mes** e **ano**
7. Faça o **Paste** do comando.
8. Tecla no SYNTAX EDITOR **FORMATS data_m (MOYR8).**
9. **Run** os comandos **Compute** e **Formats** ao mesmo tempo.

Siga estes passos para calcular as variáveis mensais do mês e do ano de produção:

1. A partir de **Transform** no *menu* seleccione **Compute...**
2. Para o **Target Variable:** insira **c_mes**
3. *Click* no  e insira a etiqueta: “Mês de Produção”.
4. *Click* no 
5. No **Numeric Expression** box, insira **mes + 2**
6. *Click* no 
7. Seleccione o botão circular para **Include if case satisfies condition:**
8. Insira a declaração **mes <= 10**

9. *Click* no

Continue

10. Faça o **Paste** e **Run** o comando.
11. Repita os passos 2 até 10 substituindo a informação anterior com o seguinte.

Numeric Expression	If... Statement
mes - 10	mes >= 11

Quando comparar `c_mes` e `c_mes2` devia ver os mesmos valores.

12. Devemos criar uma variável também para o ano de produção **c_ano**. Repita os passos de 1 até 10 mas para **c_ano** (use a etiqueta 'Ano Agrícola de Produção').
13. Repita os passos de 2 á 10 para as duas declarações que se seguem:

Numeric Expression	If... Statement
ano + 1	c_mes <= 2
ano	c_mes >= 3

14. **Run** estes comandos. Agora queremos adicionar **Value Labels** (Etiquetas de valores) para cada um dos novos meses e ano de produção. Para fazer isso podemos usar o comando **Variable View** no DATA EDITOR ou neste caso, iremos teclar directamente no SYNTAX EDITOR o seguinte:

```
VALUE LABELS
c_mes 1 'NOV' 2 'DEZ' 3 'JAN' 4 'FEV' 5 'MAR' 6 'ABR' 7 'MAI' 8 'JUN' 9 'JUL'
      10 'AUG' 11 'SEP' 12 'OUT'
/c_ano 1982 '81-82' 1983 '82-83' 1984 '83-84' 1985 '84-85' 1986 '85-86'
      1987 '86-87' 1988 '87-88' 1989 '88-89' 1990 '89-90' 1991 '90-91' 1992 '91-92'
      1993 '92-93' 1994 '93-94' 1995 '94-95' 1996 '95-96'.
```

EXECUTE .

Lembra que uma etiqueta não pode ser mais longa de que 60 caracteres.

15. Agora **Run** este comando.
16. Faça o SAVE dos dados. É MUITO IMPORTANTE QUE NÃO SE ESQUEÇA DE GUARDAR OS DADOS POIS SERÃO USADOS NOS MÓDULOS QUE SE SEGUEM.! Use o nome do ficheiro PQ_8995P3.sav, faça o **Paste** e **Run**.
17. Grave o seu syntax file com o nome PQ_8995P3.sps e feche o documento.

Exercício 5.3

Agora, vejamos como se apresentam os novos dados. Apresente o gráfico do preço do consumidor por produção mensal (i.e. **c_mes**) e compare com os gráficos calculados antes de se ter usado o calendário mensal. Veja como os dois meses foram ajustados para incluir esses dois meses do anterior ano e deixar de lado os últimos dois anos para o ano de comercialização seguinte! Apresente o gráfico do preço do consumidor por ano agrícola (i.e. **c_ano**).

SPSS para Windows “SAMPLE SESSION”
Módulo 6 - Análises Sazonais (Séries Temporais)

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

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

As análises temporais de preços históricos e quantidade de dados das séries semanais, mensais ou anuais para produtos alimentares ou agrícolas constituem uma importante parte da análise do mercado (veja Goetz e Weber 1986, IDWP 29 Michigan State University, para mais detalhes sobre análises de séries temporais). Esta análise tem como objectivo fazer o agrupamento das variações sazonais repetitivas e intra-sazonais nas séries de dados, cujas inferências do seu comportamento pode ser delineada. Caso ocorra sazonalidade na produção e comercialização dos cereais, espera-se que os preços e quantidades variem de forma inversa ao longo do ano de produção e de comercialização.

“Merging” (Unificação) e Agregação de ficheiros e dados

Como pretendemos efectuar uma análise sazonal nos dados de 1982 a 1995 do Sistema de Informação de Mercados, precisamos de fazer o **Merge** dos dados de dois ficheiros diferentes: PQ_8995P3.sav (do último módulo) e P_8289P. sav. Para o propósito deste módulo, iremos seleccionar dados de Bamako para a mapira.

1. Inicie um novo janela de SYNTAX, e abra PQ_8995P3. sav. Faça o **Paste** e **Run**.
2. Verifique se todos os filtros e funções **Split Files** estão desactivadas.
3. Agora, tal como em módulos anteriores, seleccione os dados para Bamako e da mapira usando **Data/Select Cases. . .** e especificando **loc = 3** e **cer = 8**. Certifique-se de fazer o *Click* no botão circular perto de **Deleted** (Eliminado) na caixa para os casos não-seleccionados (**Unselected cases are**) e depois faça o **Paste** e **Run** o comando.

Contrariamente ao módulo passado, iremos usar o comando **Aggregate** para agregar os dados para um novo nível, neste caso a Média para Bamako. Para este efeito, siga os seguintes passos:

1. Vá para **Data/Aggregate. . .**
2. A partir da coluna à esquerda, seleccione, **ano mes data_m c_mes c_ano loc** e insira as variáveis na caixa de **Break Variable(s)**
3. Seleccione a variável do preço do consumidor **p4c** e insira na caixa de **Aggregate variable(s)**.
4. *Click* no **Name & Label. . .**, insira como etiqueta a expressão “Preço Médio de Consumidor” e altere o nome de **p4c_1** para **p4c**.
5. **Continue**
6. *Click* no **File. . .** e altere o nome do ficheiro para AGGBK_O1.sav
7. **Save**


8. Faça o **Paste e Run** o comando
9. Agora abra o data file p_8289P.sav usando **File/Open. . .**
10. Selecione a mapira e Bamako da mesma forma que acima fez para os dados de 1989 a 1995.
11. Não precisamos agregar os dados neste ficheiro pois podemos simplesmente guardar estes dados no seu formato actual. Para assim o fazer, deverá:
12. Ir para **File/Save As. . .**
13. Denominar o ficheiro por AGGBK_O2.sav e fazer o **Paste** do comando
14. Verifique no **Syntax Editor**, irá observar o comando **Save** escrito acrescido pela palavra /compressed (reduzido em tamanho). Mas de facto, o comando **Save** do syntax (que poderá observar ou procurar no SPSS Help Tools) apresenta-se do seguinte modo:

```
SAVE OUTFILE=file
  [/KEEP={ALL  }] [/DROP=varlist]
    {varlist}
  [/RENAME=(old varlist=new varlist). . . ]
  [/MAP] [/{COMPRESSED  }]
    {UNCOMPRESSED}
  [/UNSELECTED=[{RETAIN}]
    {DELETE}]
```

Verifique como o sintaxe aqui é diferente do efectuado no Syntax editor? Não obstante não poderemos obter muitas das funções de sintaxe possíveis dos *menus* para muitos dos comandos, poderemos teclar em qualquer caso as várias funções ou adicionar o sintaxe, directamente no **Syntax Editor** dado que o SPSS irá reconhecer e executar esses comandos. Para assim o fazer, primeiro:

15. Assim, por baixo da linha com /*compressed* no syntax file (depois da linha com **Save outfile. . .**), tecele o seguinte (**Keep** significa ficar com):
 /KEEP ano mes data_m c_ano c_mes loc p4c
16. **Run** o comando **Save outfile**.

Agora que ambos ficheiros são similares em relação aos dados (quer dizer o mesmo nível de agregação), estamos prontos para fazer o merge de ambos ficheiros.

1. Se ainda não estiver activado, faça o **Paste** do comando e abra o ficheiro aggbk_02.sav
2. Vá para **Data/Merge File/Add Cases. . .** (Data/Unir os Ficheiros, Adicionar Casos)
3. Escolha aggbko1.sav e *click* no 
4. Faça o **Paste e Run** o comando
5. Guarde o ficheiro com o nome aggbk_03.sav. Faça o **Paste e Run** o comando e depois guarde o seu novo syntax file para aggbk_03.sps

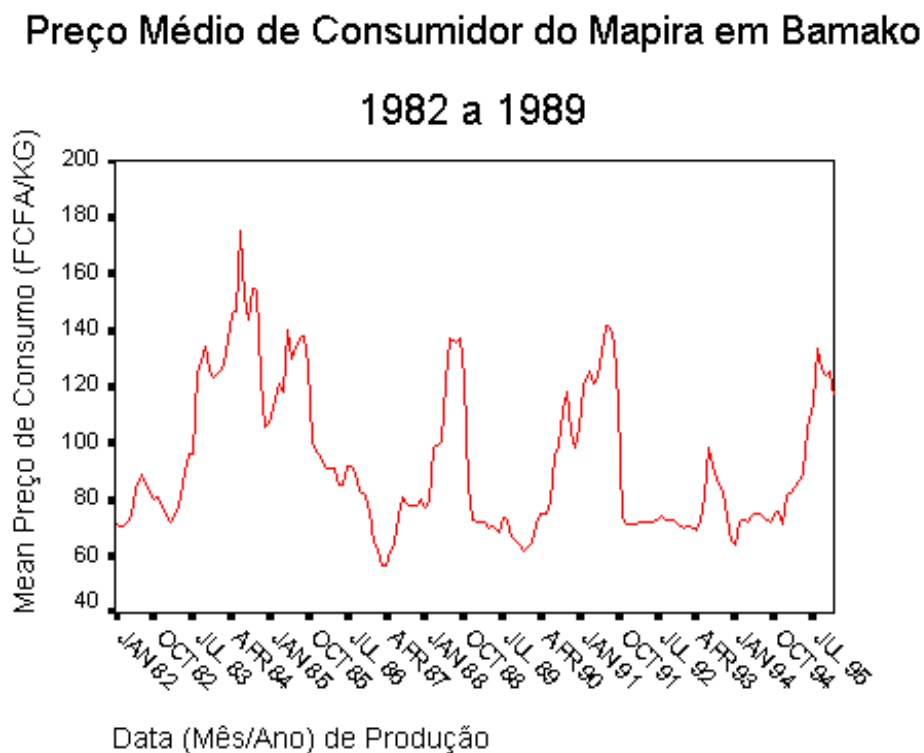
Observe os dados. Irá verificar que os dados de preços para os anos de 1982 a 1989 estão agora juntos a aqueles de 1989 a 1995! Observemos agora os dados apresentando graficamente os preços do consumidor durante todo o período e depois por ano de comercialização.

- a) Para todo o período, apresente graficamente uma curva simples para **p4c** por **data_m** com o título “Preço Médio de Consumidor do Mapira em Bamako”. Vá para o **SPSS**

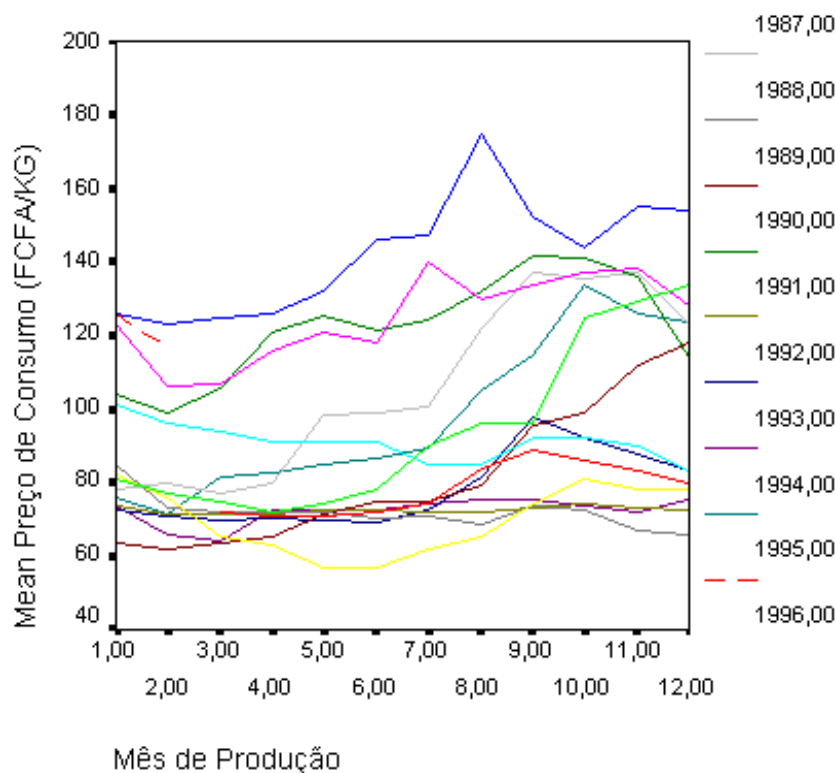
Viewer e veja como as tendências no gráfico variam ao longo do tempo. Para reduzir o Tamanho (*Font*) do Título no gráfico final, *click* duas vezes em cima de gráfico, entre no SPSS Chart Editor, **Click em cima do Título, vai ao Format/ Text**, e reduza o tamanho de 14 a 12.

- b) Apresente graficamente e ao mesmo tempo múltiplos anos de produção para **p4c** por **c_mes** e por **c_ano** simultaneamente. Veja como as curvas variam por ano de produção e comparativamente entre os anos.
- c) Para cada ano de produção individual, apresente graficamente curvas simples por **c_mes** usando o comando **Split file** por **c_ano**. Não se esqueça de teclar e accionar **SPLIT FILE OFF**. quando terminar. Guarde os resultados para a), b), e c) como **aggbk_03.spo** (poderá eliminar os resultados anteriores teclando o botão **Delete** dos dados seleccionados na coluna à esquerda).

Resultados do secção a):



Resultado de secção b):





Calculo de “moving average” (Média móvel) e apresentação gráfica de preços nominais

Uma **média móvel** ajuda a calcular o índice sazonal. Uma observação irá depender de alguns dos valores anteriores e subsequentes da mesma variável. Isto significa que se uma observação de um valor individual é maior ou menor, o procedimento de calculo da média irá aproximar este valor a outros valores das séries e as flutuações são assim eliminadas. Ao se remover estas flutuações de curto prazo nas séries temporais, o analista poderá centrar-se em padrões importantes de longo prazo tal como ciclos e tendências.


Para calcular a média móvel, precisamos de usar o comando **Create time series** (Criar séries temporais) usando os passos que se seguem. Este comando irá produzir uma nova série como uma função de uma série já existente, neste caso de preços.

1. Crie uma nova variável usando o **Transform/Create time series. . .** menu.
2. Precisamos usar a função **Moving average**. Poderá alterar a função fazendo o *click* na seta perto de caixa de **Function** (função) com a palavra **Difference** (diferencia) entre ela. Escolha **Centered moving average** (média móvel centrada -- veja explicação abaixo).

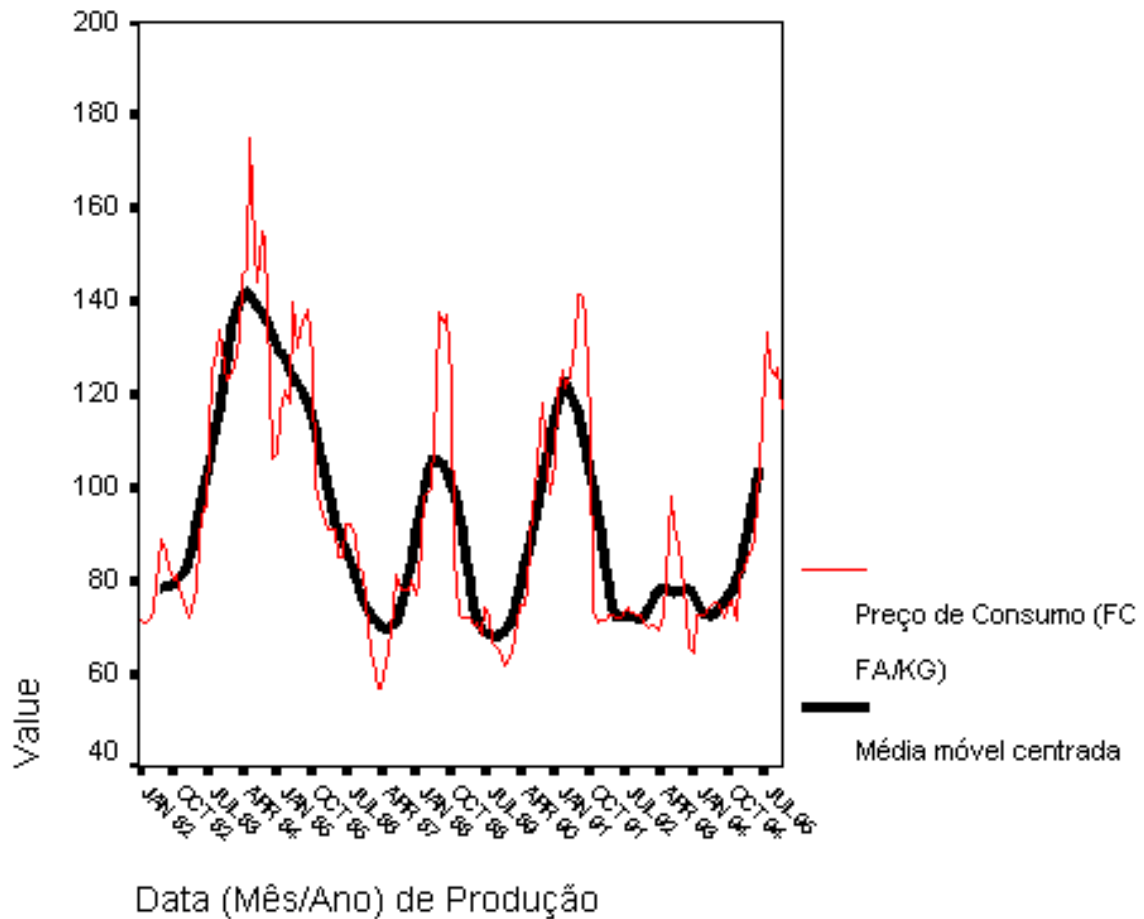
3. Depois, logo por baixo, deve inserir o número 12 na caixa de Span (Atravessar): esse 12 representa os 12 meses.
4. Selecciona **p4c** na coluna à esquerda e *click* no  perto de New Variables(s). Altere o nome da variável de **p4c_1** para **p4c_ma** (para *moving average*) e *click* no .
5. Faça o **Paste e Run** o comando.
6. Poderá também dar uma etiqueta a variável teclando no syntax:
VAR LAB **p4c_ma** "Média móvel centrada".
7. **Run** no Syntax Editor.

Centered moving average é uma média de amplitude ou *span* de valores duma séries de dados que estão a volta, incluindo o valor corrente. O *span* refere-se ao número de valores de séries usadas para calcular a média. Se o *span* é par (2, 4, 6...), a média móvel é calculado através do calculo da média de cada par de médias não-centradas. O número de casos com o valor omissos do sistema no princípio e no fim das séries para um *span* de **n** é igual para **n/2** para valores de *span* pares e para valores de *span* ímpares. Por exemplo, se o *span* é 5, o número de casos com o valor omissos do sistema no princípio e no fim da série é 2. Se o *span* é 4, o número de casos com o valor omissos do sistema no princípio e no fim da série também é 2.

Apresentemos graficamente os preços nominais com a média móvel.

1. **Graphs/Line. . .**
2. *Click* no quadrado perto de **Multiple**
3. Escolha o botão circular perto de Values of individual cases
4. *Click* no .
5. Insira **p4c** e **p4c_ma** na caixa de Lines represent. (Linhas representam)
6. *Click* no botão circular perto de Variable: na caixa de **Category Labels**
7. Selecciona **c_mes** ou **data_m** e insira-os como a escolha da variável na caixa de **Category Labels**, faça o **Paste e Run** o comando
8. Se o SPLIT FILE era ainda activo, vai ver erros.

Agora observe o gráfico: veja como a média móvel apresenta toda a tendência e elimina as flutuações de curto prazo!



Índice Sazonal

Um outro bom indicador para análises sazonais é o índice sazonal. Movimentos sazonais em séries temporais de preços e quantidades são particularmente prevalentes em culturas agrícolas.

Os processos e ciclos ecológicos exercem um forte influência nos processos de produção agrícola e são reflectidos através de quantidades produzidas e do comportamento dos preços ao longo do tempo. O movimento dos preços sazonais podem reflectir custos de armazenamento e poderão também afectar produtos não armazenáveis. O índice pode ser calculado a partir da média móvel: **o índice sazonal é o preço sazonal dividido pela média móvel**. Para calcular o índice sazonal, use os seguintes passos:

1. A partir de **Transform** menu seleccione **Compute. . .**
2. Para **Target Variable**: tecler época
- *Click* no **Type & Label. . .** e insira a etiqueta “Índice Sazonal”.
4. *Click* no **Continue**
5. A partir da caixa de **Numeric Expression**, insira **p4c / p4c_ma*100**
6. Faça o **Paste e Run** o comando

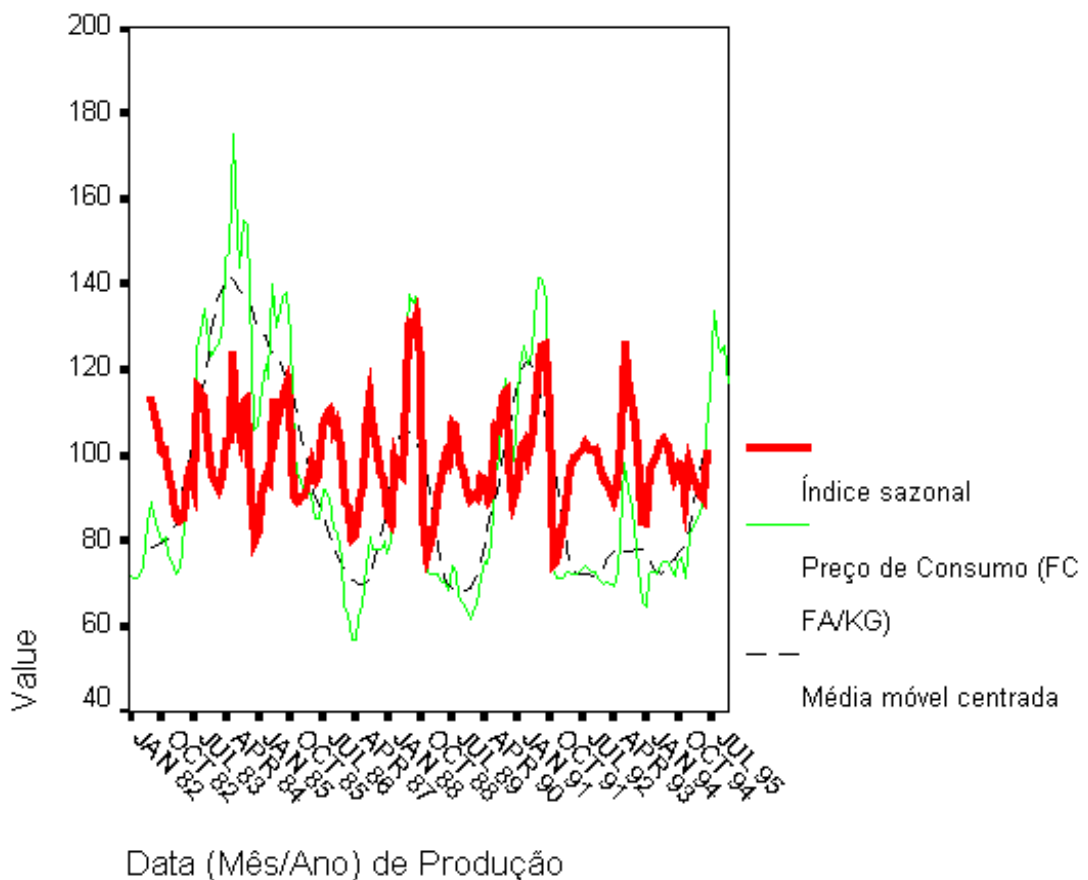
Agora, apresente graficamente o índice sazonal com os preços nominais na forma de mês de calendário (**date_m**) ou na forma de mês de produção ou de comercialização (**c_mes**) usando os mesmos passos para a média móvel acima apresentado graficamente. Por exemplo, o sintaxe para este comando deverá apresentar-se do seguinte modo:

```
GRAPH
/LINE(MULTIPLE)=VALUE(epoca p4c p4c_ma) BY date_m.
```

Observe o gráfico: veja como o índice encontra-se a volta de 100? Consegue observar uma tendência periódica (por exemplo, os pontos máximos anuais)? Pesquisas mais aprofundadas demonstrariam que as tendências sazonais dos preços começam a diminuir um pouco em Setembro e Outubro quando as pessoas têm uma ideia se é um bom ano ou não. Na época de colheita em Novembro, os preços de Dezembro decrescem. Os preços aumentam ao longo da época seca de Janeiro a Junho e alcançam o seu máximo na época chuvosa em Julho e Agosto. Poderá guardar o output como `aggbk_03.spo` (está a adicionar mais informação do ficheiro já existente do `aggbk_03.spo`).






Este seria um bom momento para também guardar os dados. É sempre bom dar um nome ao ficheiro que ajuda-nos a definir o tipo de dados no qual estamos a trabalhar ou que estão contidos no ficheiro.

1. Guarde o ficheiro dos dados como **epoca.sav**
2. Faça o **Paste e Run** o comando.



Agregando Índices Sazonais por Mês para Criar e Apresentar Graficamente os Índices Históricos Mensais Médios e Desvio Padrão

Conforme acima explicado, é importante agregar os dados a outros níveis para análises. Nesta sub-seção, iremos agregar de um nível semanal para mensal e iremos observar a distribuição do desvio padrão. Primeiro agreguemos em meses de produção usando os seguintes passos:






1. Vá para **Data/Aggregate. . .**
2. A partir da coluna á esquerda, seleccione **c_mes** e insira na caixa de **Break Variable(s)**.
3. Seleccione a variável **epoca** e insira na caixa de **Aggregate variable(s)**, depois atribua uma etiqueta como “Índice sazonal médio”.
4. Seleccione a variável **epoca** e insira por uma segunda vez na caixa de **Aggregate variable(s)**. *Click* no **epoca_2 = MEAN (epoca)** dentro da caixa, logo que tenha atribuido a etiqueta “Desvio Padrão de Índice Sazonal”, *click* no 
5. *Click* no  e seleccione o botão circular perto de **Standard deviation** (o desvio padrão).
6. 
7. *Click* no  e altere o nome do ficheiro para AGG_EPO1.SAV
8. 
9. Faça o **Paste e Run** o comando.

Agora, abra o ficheiro AGG_EPO1.sav (faça o **Paste e Run** o comando) e observe a janela de **Data Editor**: irá ver um valor em cada um dos 12 meses! Para observar a tendência do índice sazonal médio (epoca_1) ao longo do ano de produção, apresente graficamente uma simples curva por **c_mes** para **Values of individual cases** (Valores dos casos individuais). Para apresentar graficamente o índice sazonal agregado com o desvio padrão devemos primeiro calcular as variáveis. Use o comando **Compute** de modo a obter as seguintes declarações do syntax:


```
COMPUTE DP_MAIS = epoca_1 + epoca_2 .  
EXECUTE.  
COMPUTE DP_MENOS = epoca_1 - epoca_2 .  
EXECUTE .
```

Faça o **Paste e Run** os comandos. Guarde os novos dados em AGG_EPO2.SAV, faça o **Paste e Run** o comando. Para apresentar graficamente as curvas, use linhas múltiplas representando **epoca_1**, **dp_mais** e **dp_menos**. Apresente graficamente por **c_mes** para **Summaries of separate variables** (Sumários das variáveis individuais). Observando o gráfico, podemos verificar como as tendências para todos os três são próximas umas das outras.

Deixando de lado o desvio padrão, é de interesse apresentar graficamente o índice sazonal agregado como os preços nominais para uma época agrícola particular ou ano de produção. Tomemos o ano de produção de 1994 para este exemplo:

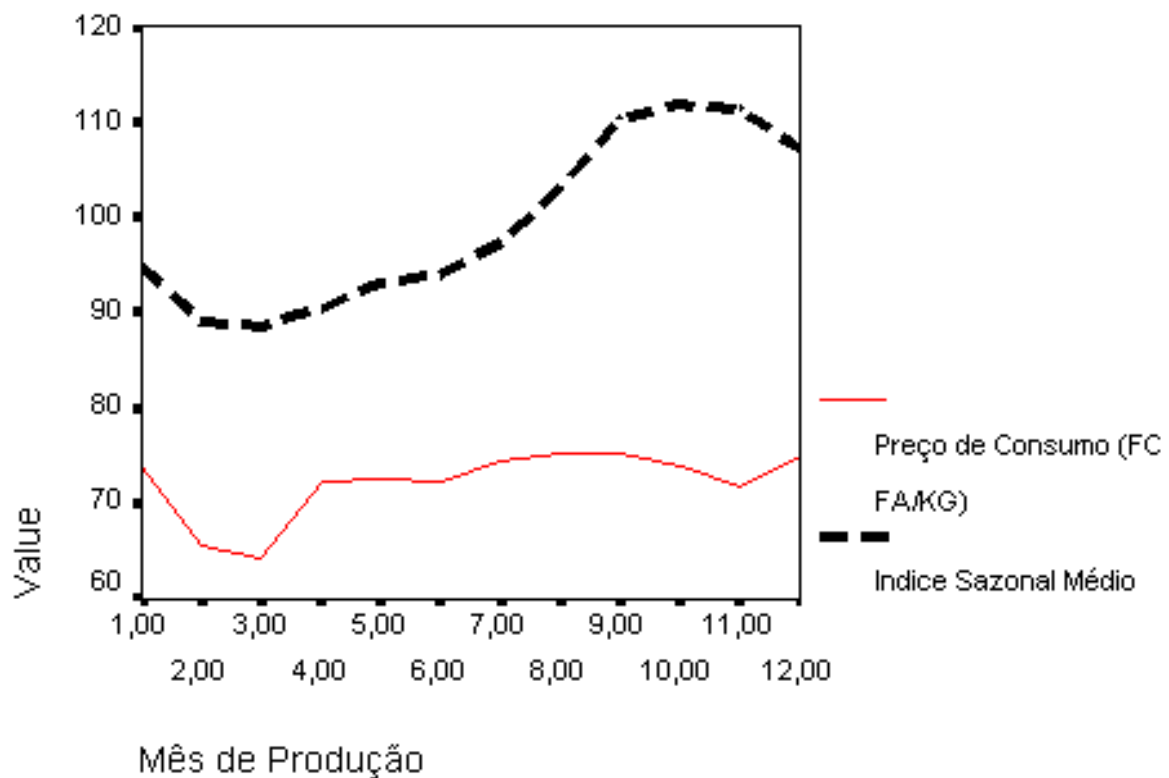
1. Active os dados sazonais abrindo o ficheiro EPOCA.sav (faça o **Paste** e **Run** o comando). Selecione os dados para **c_ano = 1994** e depois faça o merge dos dados do índice sazonal agregado como o ficheiro do desvio padrão, usando o comando **Merge**. Podemos fazê-lo seguindo os seguintes passos:
 2. Vá para **Data/Merge File/Add Variables. . .**
 3. Selecione o nome do ficheiro AGG_EPO2.sav e click no 
 4. Selecione **ano mes epoca epoca_2 dp_mais dp_menos** da lista em **New Working Data File:** e click no 
 5. Verifique o gráfico perto de **Match cases on key variables in sorted files**
 6. Click no botão circular perto de **External file is keyed table** (Ficheiro externo é tabela chave de referência)
 7. Selecione **c_mes** a partir da lista de **Excluded Variable:** (Variáveis excluídas)
 8. Click no  perto de **Key Variables:** (Variáveis chaves) (em baixo, á direita)
 9. Faça o **Paste** do comando
 10. Click no , depois 
 11. Selecione e **Run** o comando. Certifique-se no SYNTAX EDITOR de incluir EXECUTE. Grave o ficheiro como EPOCA94.SAV.

Agora apresentemos graficamente o índice sazonal agregado com preços nominais para o ano 1994. Tente apresentar graficamente as curvas, sozinho, mas aqui estão os passos caso tenha problemas:

1. **Graphs/Line. . .**
2. *Click* no quadrado perto de **Multiple**
3. Escolha o botão circular perto de **Values of individual cases**
4. *Click* no 
5. Insira **p4c** e **epoca_1** na caixa de **Lines represent**.
6. *Click* no botão circular perto de **Variable:** na caixa de **Category Labels**
7. Selecione **c_mes** e insira-o como a escolha da variável na caixa de **Category Labels**
8. Faça o **Paste** e **Run** o comando.

Comparação de Preços de Consumo 1994

e o Índice Sazonal




Observe o gráfico: veja como o índice sazonal é muito maior que o preço do consumidor para o ano de produção de 1994! Porque isso acontece? Guarde uma cópia do gráfico para análises mais aprofundadas. Não podemos realmente fazer qualquer comparação aqui, até colocar-mos ambas séries na mesma escala. Outra manipulação dos dados interessante para análises sazonais é conseguir-se modificar os preços nominais correntes sazonais dividindo o preço mensal pelo preço de Novembro (o mês principal de colheita) de modo que os valores resultantes estejam na mesma escala como o índice sazonal agregado com base para Novembro. Usando tal base, o preço da colheita pode capturar a valorização aos produtores ao longo do ano de comercialização. Este tipo de gráfico permite comparar os movimentos do preço para um ano contra um médio de médios sazonais (e desvios padrões)

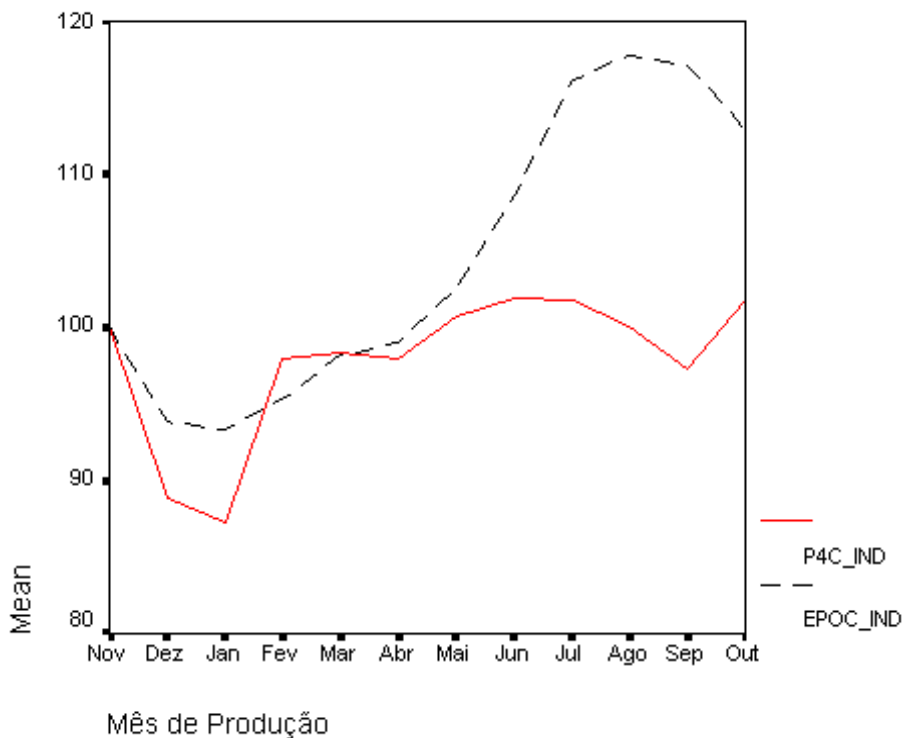
Quais são os valores do preço do consumidor e os preços sazonais para 1994? Se olharmos para a janela do **Data Editor**, podemos observar os valores para o mês de Novembro que são 73.85, e 94.91, respectivamente. Agora para obter o mesmo valor base ou 100 em Novembro devemos dividir todos os valores ao longo do ano de produção pelos valores de Novembro. Podemos fazer isso usando o comando **Compute**. Siga estes passos para calcular o índice:

1. A partir de **Transform** menu seleccione **Compute**. . .
2. Para **Target Variable**: insira **p4c_ind** e insira a etiqueta: “Índice de Preço de Consumidor 1994, Base de Novembro”
3. Na Numeric **Expression** box, insira **p4c/73.85*100**
4. Faça o **Paste** do comando
5. Volte ao menu **Transform/Compute**. . .
6. Para **Target Variable**: insira **epoc_ind** e insira uma etiqueta: “Índice Sazonal de Longo Prazo, Base de Novembro”

7. Na Numeric Expression box, insira **epoca_1/94.91*100**
8. Faça o **Paste** do comando e **Run** ambos comandos compute ao mesmo tempo

Queremos apresentar graficamente o índice sazonal agregado ajustado por mês com preços nominais ajustados. Tente apresentar graficamente essas curvas sozinho outra vez mas poderá seguir estes passos caso encontre problemas:

1. **Graphs/Line. . .**
2. *Click* no quadrado perto de **Multiple**
3. *Click* no botão circular perto de **Summaries of Separate variables**
4. *Click* no **Define**
5. Selecciona **p4c_ind** e **epoc_ind** a partir da coluna à esquerda e insira-os na caixa de **Lines represent**.
6. Agora escolha a variável **c_mes** na coluna à esquerda e *click* no  perto de **Category Axis:**
7. *Click* no **Options. . .**
8. *Click* no botão circular perto de **Display groups defined by missing values**
9. **Continue**



10. Faça o **Paste** e **Run** o comando

Olhando para gráfico podemos verificar que 1994 foi um ano distinto no sentido de não ter seguido a tendência sazonal de longo prazo. Os preços foram, para um mês particular, mais altos do que a média sazonal mas depois permaneceram muito mais próximos do mesmo preço. Poderá explicar esse aspecto? 1994 foi o período de desvalorização. Existia um medo inicial de os preços deverem aumentar em Janeiro (após o governo em Mali ter anunciado preços-tecto).

Uma vez que os comerciantes ultrapassaram o choque inicial, os outros agentes tomaram consciência de que a colheita foi boa, de que existiam bons estoques e preços reagindo de acordo com o fenómeno. Para melhor interpretar os resultados, seria bom trazer dos dados da produção, efectuar entrevistas formais e assim em diante. Escreva as suas ideias no seu SPSS VIEWER e compare com os seus colegas. Não esqueça de guardar seus resultados (em EPOCA.SPO).

Decomposição Sazonal

A Decomposição Sazonal estima factores ou componentes sazonais multiplicativos ou aditivos para séries temporais. Cria novas séries contendo factores sazonais ajustados ou *seasonal adjustment factors* (SAF), séries ajustadas sazonalmente ou *seasonally-adjusted series* (SAS), componentes de tendência cíclica ou *trend-cycle components* (STC), componentes de erro ou de “random” (aleatórios) ou *the error or random components* (ERR).

Qualquer série de preços mensais ou observações de quantidades ao longo do tempo podem ser decompostas nestas quatro partes conceptuais, sendo cada grupo de componentes unicamente relacionado com a observação actual.

Antes de efectuarmos qualquer análise de decomposição sazonal, precisamos de definir datas numa maneira pre-determinada. Para fazer isso, podemos usar o comando **Define dates** no Data menu.

1. Active o ficheiro de dados EPOCA.SAV .
2. Vá para **Data/Define dates. . .**
3. Seleccione **Years, months** (Anos, meses) da caixa de Cases Are: (Casos São)
4. Insira 1982 como o ano para a caixa de First case is: (Primeiro caso é)
5. *Click* no Ok. (Nota: O comando será posto no SPSS Viewer).

Para accionar as análises de decomposição sazonal, precisamos de usar a função **Time Series**.

1. Vá para **Analyze/Time Series/Seasonal Decomposition. . .**
2. Escolha o preço do consumidor **p4c** na coluna à esquerda e insira em Variable(s):
3. Na caixa de Moving Average Weight, seleccione o botão circular perto de **Endpoints weighted by .5** (veja explicação abaixo sobre a escolha de modelos)
4. Faça o **Paste e Run** o comando.

Dois tipos de modelos por escolher: Multiplicativo ou Aditivo.

Ajustamento Sazonal Multiplicativo.

A componente sazonal é o factor pelo qual a série ajustada sazonalmente é multiplicada para recriar a série original. Com efeito, **Trends** estima componentes sazonais que são proporcionais ao nível global de séries. Observações *sem* variações sazonais têm um componente sazonal de 1.

Ajustamento Sazonal Aditivo.

Os ajustamentos sazonais são acrescidos às séries ajustadas sazonalmente para obter os valores observados (originais). Estes ajustamentos tentam retirar o efeito sazonal de uma série de modo a ser capaz observar outras características de interesse que podem ser escondidas ou “masked” pela componente sazonal. Com efeito, **Trends** estima componentes sazonais que não dependem do nível global das séries. Observações *sem* variações sazonais têm um componente sazonal de 0.

Observe os resultados no SPSS Viewer e depois no Data editor. O modelo sazonal é:

$$\text{PRICE} = \text{Trend} \times \text{Cyclical composition} \times \text{seasonal components} \times \text{error term.}$$

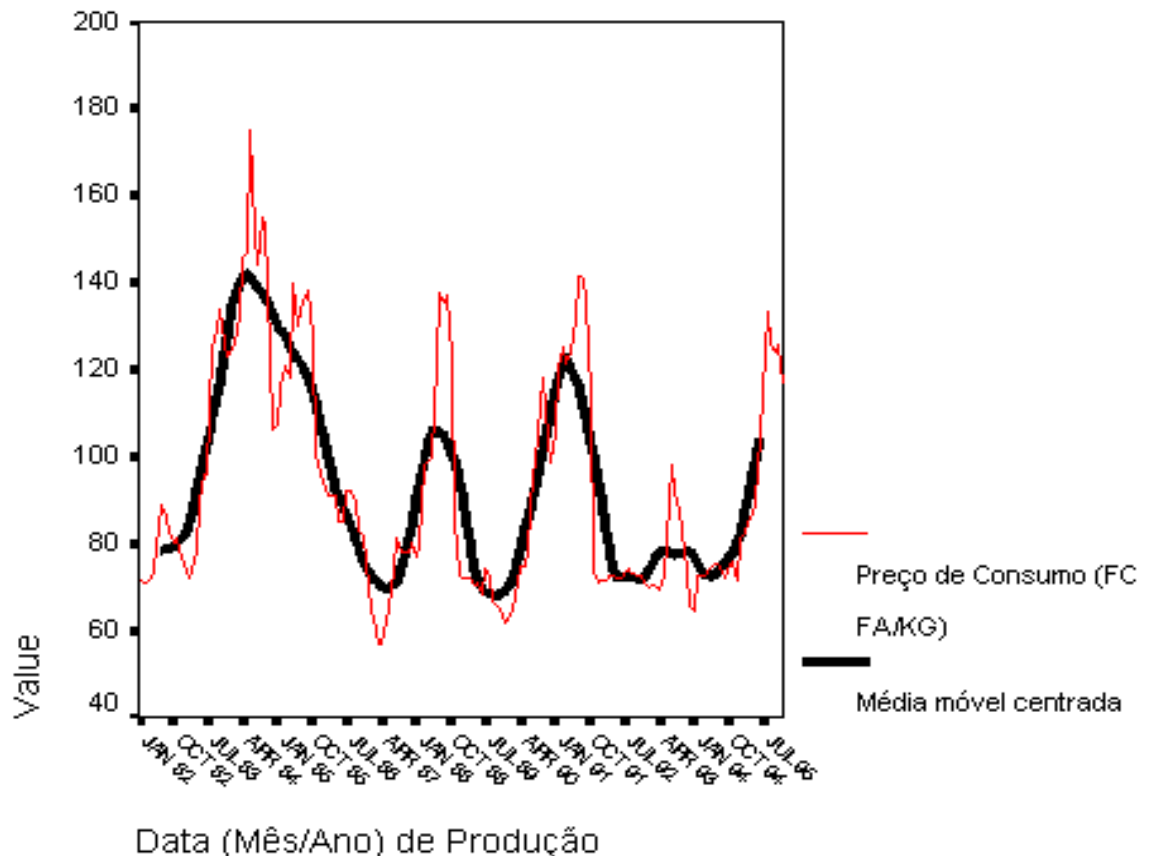
E por exemplo, para Janeiro de 1982 temos:


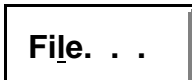

$$\begin{aligned} P(72) &= \text{STC}(80.101) \times \text{SAF}(.893) \times \text{ERR}(1.006) \text{ and} \\ \text{SAS}(80.593) &= P(72)/\text{SAF}(.893) \end{aligned}$$

As várias variáveis sazonais calculadas pelo SPSS aparentam corresponder ao modelo. Certifique-se de guardar o novo data EPOCA2.SAV file. Apresente graficamente o preço do consumidor (**p4c**) com a série ajustada sazonalmente (**sas_1**) por **date_m** durante todo o período e depois para cada ano de produção usando o comando **Split file**. Observe como as duas curvas variam de ano para ano!





Outra observação interessante a fazer é ao se apresentar graficamente a média dos SAF (factores ajustados sazonalmente para p4c) e os índices p4c. Podemos fazer isso dividindo o preço mensal pelo preço de Novembro. Para fazer isso, devemos primeiro criar uma variável para representar o valor para p4c. Denomine-o **nov**.

1. Vá para o menu **Transform/Compute. . .**
2. Insira **nov** como a target variable.
3. Na Numeric **E**xpression box, insira **p4c**
4. *Click* no **If. . .**
5. Seleccione o botão circular para Include **if** case satisfies condition:



6. Insira a declaração **mes = 11**
7. *Click* no 
8. Faça o **Paste** e **Run** o comando
9. No Syntax editor, tecle
Temporary.
Select if c_ano > 1982.
10. Agora use o comando **Aggregate** e vá para **Data/Aggregate. . .**
11. A partir da coluna à esquerda, seleccione **c_ano** e insira na caixa de **Break Variable(s)**
12. Seleccione a variável **nov** e insira na **Aggregate variable(s)** box.
13. *Click* no  e altere o nome do ficheiro para AGGNOV.SAV
14. 
15. Faça o **Paste** e **Run** o comando com o commando de o Temporary select if.

Agora precisamos fazer o **Merge** dos ficheiros de modo a calcular-mos os índices e usar os valores distribuídos para Novembro ao longo dos seus respectivos anos de produção.

1. Vá para **Data/Merge File/Add Variables. . .**
2. Seleccione o ficheiro AGGNOV.SAV e *click* no 
3. Verifica que **c_ano** está na lista de **Excluded Variables**:
4. Verifique a caixa perto de **Match cases on key variables in sorted files**
5. *Click* no radio perto de **External file is keyed table**
6. Seleccione **c_ano** a partir da lista **Excluded Variables**:
7. *Click* no  perto de **Key Variables: (bottom, right)**
8. Faça o **Paste** do comando
9. *Click* no , depois no 
10. Seleccione e **Run** o comando.

Agora para calcular os valores para os índices, use os seguinte passos:

1. A partir de **Transform** menu seleccione **Compute. . .**
2. Para a **Target Variable**: insira **p4c_ind** e uma etiqueta a sua escolha
3. Para a **Numeric Expression**: box, insira **p4c / nov_1*100**
4. Faça o **Paste** e **Run** o comando.
5. Repita os passos 1 até 4 com a target variable como **SAF_ind** e a numeric expression como **SAF_1*100**
6. Grave o ficheiro como EPOCA3.SAV

Agora estamos preparados para apresentar graficamente os factores ajustados sazonalmente como o índice dos preços do consumidor. Use o comando **Split file** por ano de produção e depois apresente graficamente tal como nos exercícios anteriores, as variáveis **SAF_ind** e **p4c_ind** por **c_mes**. Observe como as curvas variam de ano para ano e compare-as. Que interpretação poderemos fazer?

Guarde o seu syntax file (EPOCA.SPS) para acrescentar o sintaxe do exercício mais recente no módulo. Feche o *Viewer*, porque queremos abrir um novo para o módulo que se segue.

Exercício 6. 1

Este será um exercício que poderá demorar pelo facto de ser necessário repetir todos os diferentes tipos de análises sazonais efectuadas neste módulo. Deverá usar o milho (**cer = 9**) como o produto para este exercício. Repita cada uma das análises individuais alistadas abaixo da melhor forma que lhe for possível sem consultar o módulo, mas é claro que poderá retomar ao módulo quantas vezes precisar. Faça sempre o **Paste** dos comandos você mesmo. Não se esqueça de guardar os seus novos data files, o seu syntax window para documentação e qualquer gráfico ou output que poderá escolher para análises futuras. Algum dos dados irá preparar neste exercício sobre cereais, e serão usados em exercícios de módulos mais adiante.

- i) Primeiro deve agregar e fazer o merge dos dois data files **p_8289p. sav** e **PQ_8995P3.SAV**. Seleccione Bamako e milho tal como foi feito no módulo. Certifique-se de agregar usando as mesmas variáveis acima usadas e breaking pela variável do preço do consumidor (**p4c**). Lembre-se que terá que teclar parte do comando (Sugestão: /KEEP. . . /RENAME. . .)
Quando tiver preparado os dados, apresente graficamente o preço do consumidor por ano de produção.
- ii) Calcule a média móvel para o milho e apresente graficamente o preço de consumidor versus a média móvel. (Hint: you will need to create the moving average using the command **Transform/Create Time Series. . .**)
- iii) Calcule o índice sazonal e prepare o gráfico. (Sugestão: índice sazonal = preço nominal price/média móvel).
- iv) Crie um índice histórico mensal médio e calcule o desvio padrão usando o índice sazonal agregado e depois apresente graficamente várias curvas (Sugestão: Terá que fazer o **Merge** de ficheiros, e depois calcular e usar os valores para o mês de Novembro. . .)
- v) Prepare cálculos sazonais usando os comandos **define dates** e **seasonal decomposition** e depois apresente graficamente o mean seasonal adjusted contra os índices de preço (Sugestão: Irá precisar usar os comandos **aggregate, merge** e **compute**).

**SPSS para Windows “SAMPLE SESSION”
Módulo 7 - Tendências (Séries Temporais)**

**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
Março de 2000**

Uma série temporal pode apresentar uma tendência natural ou secular ao longo do tempo, uma tendência que poderá ser crescente ou decrescente. O factor de tendência é calculado através da elaboração de uma regressão dos mínimos quadrados ordinários (*ordinary least squares regression*) que será apresentada nesta sessão. É útil estimar a tendência como o primeiro passo na determinação da direcção na qual as séries tomaram no passado, na qual irão proceder no futuro, e permite-nos estimar os valores futuros. Para alcançar estas metas, iremos usar o ficheiro **epoca.sav** que contém séries sobre preços mensais do consumidor de mapira em Bamako de 1982 à 1995. Para o propósito deste módulo, iremos centrar-nos num período mais curto de quatro anos, de Janeiro de 1992 à Dezembro de 1995.

Assim irá precisar de criar um novo ficheiro das séries de preços para o período 1992-1995. Tal como fez anteriormente, faça o **Open** do ficheiro **Epoca.sav** (faça o **Paste** e **Run** o comando), usando o comando **Select Cases** para eliminar os dados antes de 1992 (Sugestão: ano \geq 1992), faça o **Paste** e **Run** o comando. Guarde este novo ficheiro com um novo nome (denomine-o **Tend1.sav**), de modo a manter os dados de 1982 para 1995 no ficheiro original. Agora podemos estudar a tendência do preço da mapira em Bamako de 1992 à 1995 usando análises de regressão. Use os seguintes passos:

1. **Analyze/Regression/Curve Estimation. . .**
2. Selecciona a variável do preço do consumidor **p4c** na coluna à esquerda e insira na caixa de **Dependent(s)**: (variável dependente).
3. **Click** no botão circular perto de **Time** (período) na caixa de **Independent** (variável independente)
4. Na caixa de **Models** (modelos), irá observar que **Linear** já está escolhido. Note que existem vários modelos para escolher; poderá também incluir um termo constante, dependendo dos dados (normalmente vai incluir um constante).
5. Na parte de baixo da caixa de diálogo, verifique a selecção do botão circular perto de **Display ANOVA table**
6. Faça o **Paste** e **Run** o comando.

De entre outros aspectos, os resultados na janela de **SPSS Viewer** irão apresentar-nos o seguinte:

- o R quadrado ajustado. Esta estatística dá-nos uma medida empírica da força da relação linear que pode existir entre as variáveis dependentes e independentes da amostra. Mede a proporção da variação da variável dependente que é explicada pela regressão. O coeficiente do R quadrado tem de ser corrigido pelo grau de liberdade para obter o *coeficiente de determinação* do R quadrado **ajustado** que pode ser comparado para as amostras que têm diferentes tamanhos e diferentes números de variáveis independentes. No nosso exemplo, o coeficiente de R quadrado ajustada é 43% que disse que 43% de variação no preço de consumidor é explicada pela passagem de tempo em intervalos regulares.

- os coeficientes de regressão (β) que nos dão a relação linear entre as variáveis dependentes e independentes. No nosso exemplo, irá verificar que o termo constante irá ser 61,58 FCFA e irá aumentar em 0,84 FCFA por mês (a unidade de tempo neste exemplo é um mês)..

- o nível de significância (ou probabilidade associada) de se testar as hipóteses nulas dos coeficientes de regressão (SIG T) e de toda a regressão (SIG F). Estes níveis dão-nos a probabilidade de aceitar ou rejeitar o facto de o coeficiente de regressão ser diferente de zero. Se o o coeficiente β for realmente igual a zero, não existe nenhuma relação linear entre as variáveis dependentes e as independentes. Quanto mais baixa for a probabilidade (SIG T < 0,05), melhor será a regressão em termos de dizer que o β é significativamente diferente do que zero. Nas nossas análises, as probabilidades são boas para os coeficientes de regressão individuais (SIG T) assim como para a regressão no seu todo (SIG F), na ordem de 0,0000.

Dependent variable.. P4C
Listwise Deletion of Missing Data

Method.. LINEAR

Multiple R ,66540
R Square ,44276
Adjusted R Square ,43064
Standard Error 13,38002

Analysis of Variance:
DF Sum of Squares Mean Square

Regression 1 6543,2066 6543,2066
Residuals 46 8235,1530 179,0251

F = 36,54911 Signif F = ,0000

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
Time	,842788	,139406	,665399	6,046	,0000
(Constant)	61,580932	3,923632		15,695	,0000

A partir dos resultados das análises de regressão, podemos concluir que os preços do consumidor para mapira em Bamako aumentaram mensalmente na ordem de 0,84 FCFA para o período 1992-1995. Contudo os resultados dos mínimos quadrados somente são válidos sob certas condições (particularmente, a ausência de uma auto-correlação dos resíduos). Para as séries temporais, esta condição geralmente não é respeitada. Quando esta condição é violada, os diferentes coeficientes acima descritos não possuem um valor estatístico. O teste Durbin-Watson é um teste para verificar se os valores residuais em séries são correlacionados (ou auto-correlacionados). Uma das hipóteses das análises de regressão é que os valores residuais para





observações consecutivas **não** são correlacionados. Se isso for verdade, o valor esperado da estatística de Durbin-Watson é 2. Valores menores que 2 indicam uma auto-correlação positiva, um problema comum nos dados das séries temporais. Valores maiores que 2 indicam uma auto-correlação negativa.

É necessário testar este modelo em relação a auto-correlação dos resíduos. Isso pode ser feito através do comando **Curve Estimation** no SPSS para Windows conforme acima usado, mas aqui iremos usar o comando **Linear**. Mas antes disso, será necessário criar a variável **periodo** (o equivalente de **time** de Curve Estimation) como a variável independente que não se encontra automaticamente através do comando **Linear**. Iremos cria-lo como uma nova variável conforme se segue:

1. A partir de **Transform** menu seleccione **Compute. . .**
2. Para o **Target Variable**: insira **periodo** (com a etiqueta: Mês a partir de Jan 1992)
3. A partir de caixa de Numeric **Expression**: teclre: **(ano - 1992) * 12 + mes**
4. Faça o **Paste e Run** o comando

Podemos agora repetir as análises de regressão mas com a opção de se testar a auto-correlação dos valores residuais conforme se segue:

1. Vá para **Analyze/Regression/Linear. . .**
2. Insira **p4c** na caixa de **Dependent**:
3. Insira **periodo** na caixa de **Independent(s)**:
4.  . . .
5. *Click* na **Durbin-Watson** box na caixa de **Residuals**
6.  depois faça o **Paste e Run** o comando

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,665 ^a	,443	,431	13,380	,198

^a. Predictors: (Constant), PERIODO Número de Mes a partir de Jan 1992

^b. Dependent Variable: P4C Preço de Consumo (FCFA/KG)

Vá para o **SPSS Viewer**. Irá verificar que os coeficientes são os mesmos que aqueles dos valores anteriores e o coeficiente Durbin-Watson (D.W.) é igual a 0,198. Quando comparamos este coeficiente com os valores correspondentes na tabela dos valores teóricos de Durbin-Watson, concluimos que existe uma **auto-correlação positiva** entre os resíduos. Assim, devemos continuar a usar o comando de auto-regressão para considerar a auto-correlação, do seguinte modo:

1. Vá para **Analyze/Time series/Autoregression. . .**
2. Insira **p4c** na caixa de **Dependent**:
3. Insira **período** na **Independent(s)**: box
4. Na caixa de **Method**, seleccione o método **Prais-Winsten**. Nota: poderia ter usado o método **Cochrane-Orcutt** dado que apresenta as mesmas funções mas é preferível o primeiro método.
5. Faça o **Paste e Run** ambos comandos.

Nota: Observe que o comando inicia a análise através do método “ordinário” dos mínimos quadrados, o mesmo sintaxe que foi usado no princípio do módulo.

O **SPSS Viewer** revela que o coeficiente D.W. aumentou para 1,37 mas permanece no intervalo onde é considerada a existência de uma auto-correlação positiva. Contudo, podemos observar um redução nos coeficientes de regressão ajustados e dos níveis de significância dos testes das hipóteses nulas para os coeficientes de regressão.

Prais-Winsten Estimates

Multiple R	,33582303
R-Squared	,11277711
Adjusted R-Squared	,07334498
Standard Error	5,896952
Durbin-Watson	1,370633

O que acontece é que, longos períodos de tempo com muitas observações têm mais dificuldades de satisfazer os valores teóricos de D.W. que aumentam com o número de observações. Ainda, durante um longo período, somos tipicamente confrontados com mudanças de direcção em sentido contrário das tendências, tornando problemática a relação entre as variáveis dependente e independente. Assim, é necessário neste caso, subdividir o períodos em factos ou aspectos objectivos. Por exemplo, podíamos escolher o ano de desvalorização e considerar um período de pre-desvalorização (1992-93) e período de pós-desvalorização (1994-95).

Para apresentar a análise de regressão para o período 1992-93:

1. Seleccione os dados antes de 1994 ($\text{ano} < 1994$) fazendo o filtro (não eliminação) das observações não-seleccionadas que serão mais tarde usadas para o segundo período (acima descrito);
2. Calcule a regressão com o comando **Autoregression** executando os cinco passos acima descritos.

Prais-Winsten Estimates (1992-93)

Multiple R	,07528134
R-Squared	,00566728
Adjusted R-Squared	-,08903107
Standard Error	5,1612478
Durbin-Watson	,9808161

Irá observar que o coeficiente de regressão ajustado do primeiro período torna-se negativo (geralmente isso não é possível uma vez que esse coeficiente varia de zero a um. Isso é causado pelos procedimentos das análises que arredondam os valores para mais, o que podem tornar um número em cima de 1 -- muito estranho para um número que teoricamente só pode variar entre 0 e 1. Uma vez que este é um número que é subtraído por um, poderá acontecer que o coeficiente ajustado seja negativo. Encontramos esta situação quando o coeficiente não-ajustado está próximo de zero). Além disso, não obstante ocorrer um aumento (por 2 vezes) do coeficiente usando o procedimento de auto-regressão comparado com o método “ordinário” dos mínimos quadrados (OLS), o coeficiente D.W. ainda permanece no intervalo que nos leva a crer que existe uma auto-correlação positiva para os valores residuais. Além disso, verificamos que a tendência torna-se insignificante.

Para analisar o segundo período 1994-95, devemos:

1. Seleccionar o período adequado da mesma forma que o fizemos para o período de 1993;
2. Uma vez que as variáveis temporais para este período iniciam com o valor de 25, deve-se criar uma nova variável periodo2, começando por i:
 - i. A partir de **Transform** menu seleccione **Compute. . .**
 - ii. Para a **Target Variable:** insira **periodo2** (escolha uma etiqueta a seu gosto)
 - iii. A partir de caixa de Numeric **Expression:** insira: **periodo - 24**
 - e seleccione o botão circular perto de **Include if case satisfies condition:**
 - v. Insira: **ano > 1993**
 - vi.
 - vii. Faça o **Paste e Run** o comando
3. Run as análises de regressão tal como foi feito anteriormente mas use **periodo2** como a variável independente.

Prais-Winsten Estimates (1994-1995)

Multiple R	,64372575
R-Squared	,41438284
Adjusted R-Squared	,35860977
Standard Error	6,384782
Durbin-Watson	1,5360621

Contudo, o segundo período que começa a partir da desvalorização FCFA apresenta uma evolução do preço sem nenhuma auto-correlação com um coeficiente de regressão ajustado de 35%, e níveis de significância para os coeficientes de regressão, de cerca de 0,0000.

As análises que efectuou, consideram as tendências dos preços observados e numa extensão menos ampla, o ciclo. É importante aqui compreender que os preços do cereal no Sahel reflectem mais e mais a sua disponibilidade no contexto da liberalização do mercado. Esta disponibilidade está muito relacionada com os ciclos de secas que podem ser mais ou menos importantes. Além do mais, considerando o facto de a colheita realizar-se somente uma vez por ano, os preços estão sujeitos a variações sazonais. Seria também interessante aqui apresentar algumas análises sazonais (veja Módulo 6). Precisaríamos de, por exemplo, *de-sazonalizar* “deseasonalize” (tirar as influências de curto prazo) **p4c** e usar a nova variável **sas_1** como a variável dependente no comando **Autoregression** .

Ir  verificar que com estes dados, todos os per odos iriam beneficiar-se do coeficiente D.W. sem auto-correla o dos seus valores residuais. Contudo, para o per odo 92-93, a tend ncia demonstra que o aumento nos pre os n o foi significativo e com um coeficiente de regress o ajustado negativo. Podemos tamb m observar um coeficiente de determina o elevado de 57% para o per odo de p s-desvaloriza o, contra 12% para o per odo de 92-95. Tamb m verificamos que o aumento nos pre os   tamb m importante durante o per odo de p s-desvaloriza o: 2,64 F por m s contra 1,03 F por m s para o per odo 92-95.

Guarde o seu Syntax file para o TEND1.SPS e fa a o Close (fechar).

Exerc cio 7.1.

Abra o ficheiro com os dados sobre o pre o do milho que preparou no exerc cio 6.1 no  ltimo m dulo. Depois repita as v rias an lises sobre a tend ncia (atrav s da regress o) que acaba de aprender no m dulo 7. Interprete os resultados.

**SPSS para Windows “SAMPLE SESSION”
Módulo 8 - Preços Reais, Apresentação Gráfica de Preços e Tabelas
(Séries Temporais)**

**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
Abril de 2000**

Os preços reais permitem-nos analisar os dados em termos reais, considerando a inflação, significando na mesma base valor ao longo do tempo. Este passo importante nas análises de preços permitem-nos inflacionar/deflacionar dados sobre preços correntes de modo que todos os valores possuam um denominador comum. Não obstante ser menos importante para os produtores, os economistas preferem trabalhar com preços reais, relativos a preços de outros bens na economia, deflacionando os preços. Depois destas séries temporais correntes terem sido deflacionadas, aparentam ser mais irrealistas, mas, se existisse uma tendência significativa nos dados, é melhor saber se é causado pela inflação ou por uma tendência natural nos preços reais: assim, o procedimento deflacionário poderá ser usado para responder a esta questão. Usando o que aprendemos no módulo anterior em relação às tendências, seria eventualmente essencial, e eficaz, para as análises extrapolar a tendência com projecções lineares de modo a estimar o preço real para o mês em análise e estabelecer uma nova base mensal. Neste módulo, iremos criar preços reais importando dados sobre índices de preços de 1988 para 1995 do Quattro Pro (Spreadsheet software), conforme explicado no módulo 3.

Obtenção do índice de preço ao consumidor através do Quattro Pro para calcular preços reais ou deflacionários

Pelo facto de pretendermos obter dados do ficheiro Quattro pro, a extensão será *.wk1 que é equivalente a *.w*. Siga estes passos para obter os dados sobre o índice de preços:

1. **File/Open...**
2. Vá para **Files of type:** e seleccione **Lotus (*.w*)**
3. *Click* no ficheiro *indice.wk1* que irá aparecer na parte de baixo e *click* no **Paste**
4. Uma caixa *Opening File Options* (Opções para como abrir o ficheiro) vai aparecer. É possível importar através de nomes de variáveis (*field names*) ou em especificando a *Range* (a série) de colunas e linhas a importar (por exemplo, A1..C35). Neste caso, vamos utilizar o *default* que é simplesmente de *click* em cima de **OK**.
5. Será aberta uma janela de opções, *click* no **OK**
6. **Run** o comando
7. Gravar o ficheiro com o nome **PINDICE.SAV**, seleccionado **SPSS (*.sav)** no **Files of type:**

Certifique-se que o Data editor está activado. Irá observar três colunas: a primeira refere-se ao **ano**, a segunda ao **mês** e a terceira refere-se ao **píndice** de preços de consumidor (IPC) que queremos manter. Pode preencher as etiquetas dessas variáveis no **Variable View** e re-gravar, mais se quer programar as mudanças, faz o seguinte:


1. *Click* no **Data /Merge Files /Add Variables...**
2. Na caixa **Excluded Variables**: rename o variável **a** a **ano**, **b** a **mois** e **c** a **pindice**.
3. Põe os tres variáveis na caixa **New Working Data File**: e deve mudar **a**, **b** e **c** do caixa **New Working Data File**: à caixa **Excluded Variables**: .
4. Guarde o ficheiro a o nome **PINDICE.SAV**.
5. Abre o ficheiro **EPOCA.SAV**. Seleccione **File /Open /Data...**

Vamos juntar (“merge”) o ficheiro **pindex.sav** file com o ficheiro **EPOCA.SAV**. Para fazer isso, é preciso juntar em base aos variáveis chaves **ano** and **mes**. Ambos ficheiros deben ser ordenados em forma igual.

Agora que temos o índice de preço podemos calcular o preço real dividindo o preço do consumidor pelo índice, do seguinte modo:

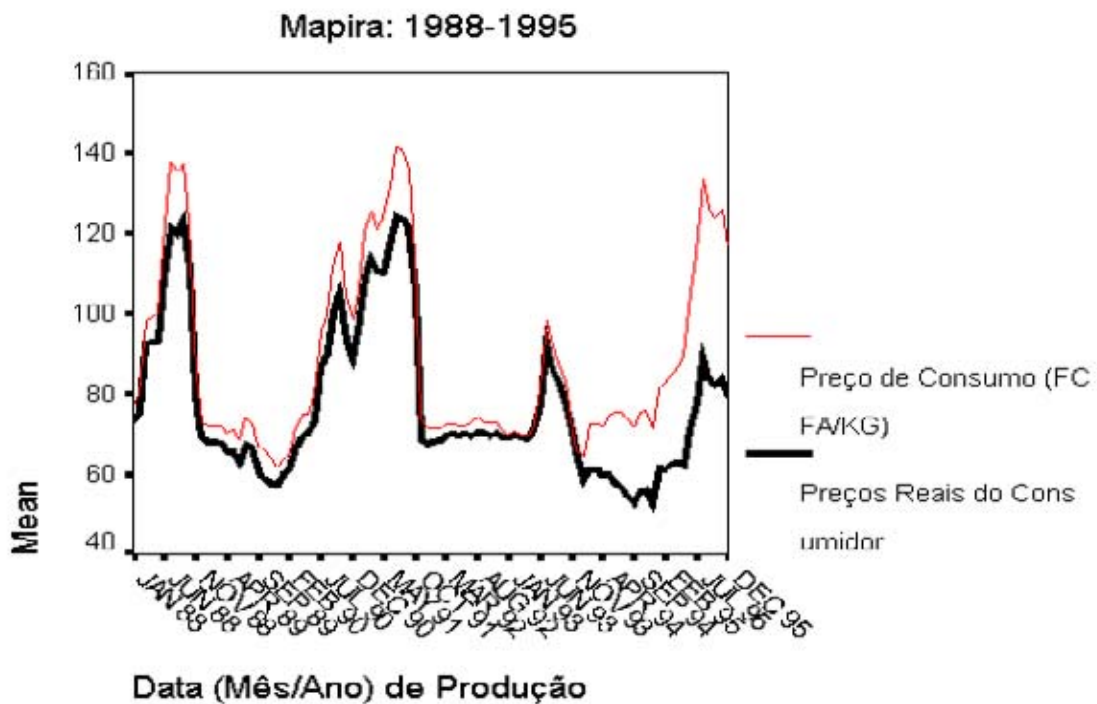
1. A partir de **Transform** menu seleccione **Compute...**
2. Para a **Target Variable**: insira **real** e uma etiqueta a sua escolha (ex. preços reais)
3. A partir de **Numeric Expression**: box, insira **p4c / pindice*100**
4. Faça o **Paste** e **Run** o comando
5. Se nada aparecer na coluna de real, vai ao **Transform/Run Pending Transformations** (Accionar todos os comandos pendentes)
6. Depois guarde o data file com o nome **PRECREAL.SAV** faça o **Paste** e **Run** o comando

Agora vejamos a comparação dos dois preços através da apresentação gráfica e conjunta dos preços nominal e real!

1. Como estamos a trabalhar com o índice de preços de e depois de 1998, devemos primeiro seleccionar os dados para o período de tempo correspondente. Tal como anteriormente feito, deverá teclar:
Temporary.
Select if ano >=1988.
2. **Graphs/Line...**
3. *Click* no quadrado perto de **Multiple**
4. *Click* no botão circular perto de **Summaries of separate variables**
5. *Click* no **Define**
6. Seleccione **p4c** e **real** da coluna à esquerda e insira os em **Lines represent**
7. Agora escolha a variável **c_mes** na coluna à esquerda e *click* no  perto de **Category Axis:**
8. *Click* no **Options...**
9. *Click* no botão circular perto de **Display groups defined by missing values** (Mostrar os grupos definidos pelos valores omissos)
10. **Continue**
11. Faça o **Paste** e **Run** o comando (iremos obter um gráfico com preços médios que incluem preços antes e depois da desvalorização)
12. Repita os passos 1 até 11 mas para **ano**>= 1993 (preços médios depois da desvalorização)
13. Repita o passo 12 para os anos 1988 em diante assim como para 1993 em diante mas com **data_m** como **Category Axis:** ao invés de **c_mes** (isso irá demonstrar a comparação de preços actuais)

Agora observe e compare todos os quatro gráficos. Olhando para o último gráfico, repare como os preços reais diferem dos preços do consumidor depois da desvalorização no CFA franc: veja como os preços nominais aumentam de forma acentuada enquanto que os preços reais somente aumentam ligeiramente e não seguem tanto a mesma tendência que os preços nominais como aconteceu antes da desvalorização. Quando comparamos os dois primeiros gráficos, observe como os preços reais médios são muito mais baixos e seguem menos a mesma tendência se excluirmos os preços antes da desvalorização.

Comparação dos Preços Médios Nominais e Preços Reais



Exercício 8.1

Use o mesmo índice do preço do consumidor para calcular os preços reais para o milho para o mesmo período (Sugestão: não precisa fazer a importação do ficheiro, somente faça o copy da variável **pindice** para os dados do preço no milho). Insira os preços no gráfico e depois guarde os seus novos dados sobre o preço do milho e os gráficos elaborados.

Apresentação Gráfica da Quantidade e Preços e Tabelas

Esta sub-sessão irá demonstrar como manipular ao mesmo tempo os dados sobre a quantidade e o preço, então devemos analisar as variações e relações existentes entre ambos. Como no módulo sobre Análises Básicas, iremos também apresentar estes dados em gráficos e tabelas.

Primeiro precisamos activar o ficheiro dos dados sobre o preço e quantidade e depois devemos seleccionar um diferente mercado que possui quantidades nos dados, por exemplo, Zangasso (mercado número 75). Como somente possuímos dados a partir de 1993, devemos também seleccionar os dados referentes a esse período. Siga os seguintes passos:

1. **File/Open...**
2. Selecciona PQ_8995P3.sav, faça o Paste e Run o comando
3. A partir de **Data** menu, seccione **Select Cases...** e seccione o botão circular perto de **If condition is satisfied**
4. *Click* no **If...** por baixo de **If condition is satisfied**
5. Insira **mer = 75 and cer = 8 and ano >= 1993** na caixa à esquerda
6. *Click* no **Continue**
7. Seccione o botão circular perto de **Deleted**
8. Faça o Paste e Run o comando

Para podemos ajustar preços e quantidades no mesmo gráfico devemos primeiro converter os dados sobre a quantidade de quilogramas para toneladas. Para transformar os dados em toneladas, siga os seguintes passos:

1. A partir de **Transform**, seccione **Compute...**
2. Para a **Target Variable**: insira **q6_t** e uma etiqueta: Quantidades compradas em toneladas
3. A partir de **Numeric Expression**: box, insira **q6 / 1000**
4. Faça o Paste e Run o comando
5. Guarde os dados com um novo nome: ZAN9395.SAV.

Agora está preparado para elaborar um número de gráficos, para observar e analisar os dados. Tentemos apresentar graficamente a quantidade e os preços em trimestres e por ano de produção (**c_ano**), seguindo os seguintes passos:


1. Como queremos apresentar um gráfico em trimestres devemos primeiro seleccionar quatro meses usando o comando **Temporary**. Tecla mas não execute o teste que segue no **Syntax Editor**:

Temporary.

Select if c_mes = 1 | c_mes = 4 | c_mes = 7 | c_mes = 10.

Para aqueles que se familiarizaram em fazer o *Copy* e o *Paste* dos comandos dentro do **Syntax editor**, poderão fazê-lo a qualquer momento. Convido a aqueles que não têm a certeza para tentar fazer o copy e o paste dos comandos. Executa-lo através do sistema do *menu* é mais rápido mas certifique-se de seleccionar o sintaxe correcto que corresponde ao sample session, conforme o seguinte gráfico:

2. **Graphs/Line...**
3. *Click* no quadrado perto de **Multiple**
4. O botão circular perto de **Summaries of separate variables** deve ser seleccionado

5. Click no **Define**
6. Seleccione **q6_t** e **p4p** (o preço do produtor) da coluna à esquerda e insira-os na **Lines represent** box.
7. Agora escolha a variável **data_m** na coluna à esquerda e click no  perto de **Category Axis:**
8. Click no **Options...**
9. Click no botão circular perto de **Display groups defined by missing values**
10. **Continue**
11. Faça o **Paste e Run** o comando a partir da declaração **Temporary**
12. Repita os passos 2 até 11 mas use o comando **Split File** de modo a observar melhor cada ano de produção (lembre-se de teclar/seleccionar do menu **SPLIT FILE OFF**). O sintaxe total para esta declaração deverá apresentar-se conforme o seguinte:

```

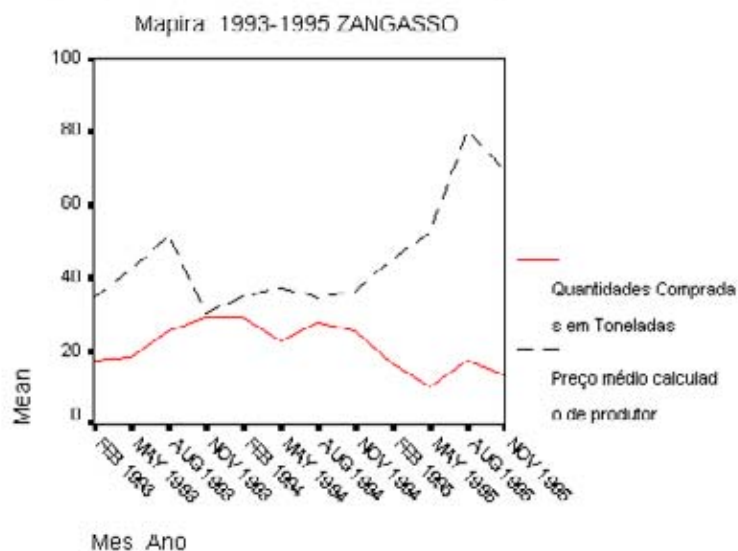
TEMPORARY
Select if c_mes = 1 | c_mes = 4 | c_mes = 7 | c_mes = 10.
SPLIT FILE BY c_ano.
GRAPH
  /LINE(MULTIPLE)=MEAN(q6_t) MEAN (p4p) BY data_m.
SPLIT FILE OFF.

```

O que podemos dizer ao observar-mos os gráficos? A relação entre o preço e a quantidade não é clara mas olhando para cada um dos anos de comercialização, parece que as quantidades decrescem e os preços aumentam. Podemos também verificar um certo decrescimento nas quantidades durante o verão/final do verão (a época chuvosa entre Junho e Setembro), a época 93/94 é um bom exemplo. Mas podemos também observar aumentos nas quantidades durante a colheita. Guarde os gráficos no **SPSS Viewer** para **PRECREAL.SPO**. Faz já algum tempo que o seu **Syntax editor** foi guardado, documentando todo o trabalho que você fez! Guarde o sintaxe para **PRECREAL.SPS**.

Como fizemos nos módulos anteriores, preparemos um gráfico e uma tabela básica para

Comparação dos Preços Médios Nominais e Preços Reais



apresentações de mudanças mensais de preços e para dados sobre quantidade e preço. Para este exemplo, iremos escolher o ano de 1994 (ano *calendário*) e iremos observar as mudanças mensais dos preços. Como está familiarizado com a utilização destes comandos, active o ficheiro **PQ_8995P3.SAV**, depois precisará de criar dois comandos **Select If** separados como: **ano = 1994 or c_ano = 1994** e o segundo **cer = 8** (certifique-se de seleccionar o botão circular perto de **Deleted**) e depois guarde os dados com o nome de **TAB94.SAVE** (faça sempre o **Paste** do seu ficheiro do sintaxe).


A razão pela qual estamos a incluir **c_ano = 1994** no comando **Select If** é para incluir o mês de Dezembro de 1993 para calcular a alteração do preço mensal para Janeiro para o mesmo mercado - --caso contrário, dada a estrutura dos dados iríamos calcular a alteração entre Janeiro e Dezembro de um mercado diferente, o que não estaria correcto. O sintaxe deve apresentar-se do seguinte modo:

```

GET
  FILE='C:\Sample\PQ_8995Pe.sav'.
EXECUTE .
FILTER OFF.
USE ALL.
SELECT IF(ano= 1994 or c_ano = 1994).
EXECUTE .
SELECT IF(mer = 8).
EXECUTE .
SAVE OUTFILE='C:\Sample\Tab94.sav'
  /COMPRESSED.

```

Depois de ter accionado o sintaxe, queremos agregar preços e calcular as alterações mensais dos preços:

1. Agora use o comando **Aggregate** e vá para **Data/Aggregate...**
2. A partir da coluna à esquerda, seleccione **loc mer cer ano mes** e insira na caixa de **Break Variable(s)** box
3. Seleccione a variável **p4c** e insira na caixa de **Aggregate variable(s)**
4. Atribua-a uma etiqueta como “Preço Médio do Consumidor” fazendo o *click* em **Name & Label...**
5. **Continue**
6. *Click* no **File...** e altere o nome do ficheiro para to AGGTAB94.SAV
7. **Save**
8. Faça o **Paste e Run** o comando
9. Agora active o ficheiro **aggtab94.sav**.
10. Crie uma nova variável usando o comando **Transform/Create time series...**
11. Seleccione **p4c_1** da lista à esquerda e *click* no  perto de caixa de **New Variables(s)**. Irá notar que **p4c_1=DIFF(p4c_1 1)** aparecerá na caixa
12. Altere o nome da variável para **p4c_1_1** to **p4c_p** (para mudança de preço) e *click* no **Change**. Faça o **Paste e Run** o comando.
13. É sempre recomendável, se pretender atribuir a uma nova variável uma etiqueta, por exemplo, **Mudança do preço**, poderá efectua-lo fazendo o *click* duas vezes no nome na janela de **Data editor**.

Agora estamos preparados para apresentar graficamente as alterações do preço para qualquer mercado para o mapira durante o ano de 1994. Por exemplo, seleccionemos o mapira em Bamako teclando no Syntax editor.

```

Temporary.
Select if mer = 3 & ano = 1994.

```

Agora use o comando **Graph** para uma curva simples de modo a mostrar as alterações do preço **p4c_p** por mês **mes**. O syntax deverá apresentar-se do seguinte modo:

```

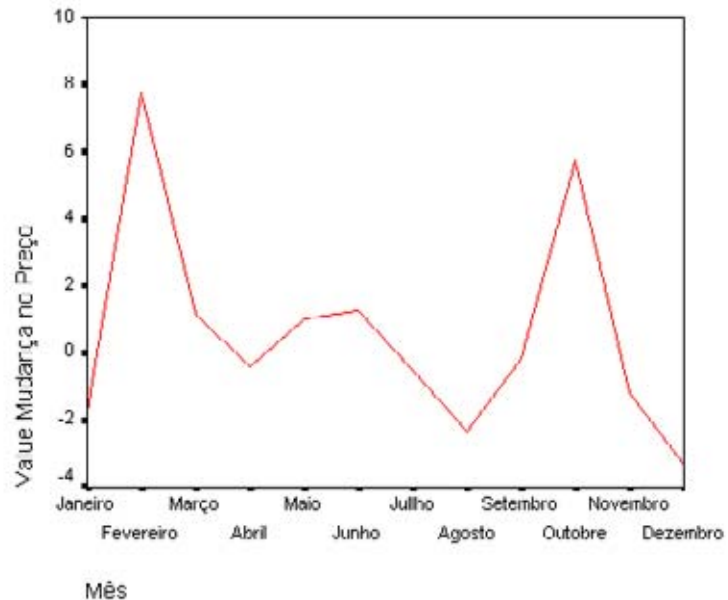
GRAPH
/LINE(SIMPLE)=MEAN(p4c_p) BY mes.

```

Execute ao mesmo tempo ambos comandos. O que observa? As colheitas ocorrem em Outubro e Novembro. Observe como as variações mais acentuadas começam antes da colheita (quando a oferta é mais baixa) até Fevereiro, alguns meses depois da colheita, e depois permanece na mesma área ao longo do resto do ano enquanto aumenta ou diminui no fim do ano de produção no momento da colheita! Porque isso ocorre? Escreva as suas ideias no SPSS Viewer e guarde.

Quando terminar de praticar nos diferentes gráficos, iremos preparar uma tabela com preços para dois meses e alterações percentuais. Podemos outra vez escolher qualquer mercado e produtos separadamente ou conjuntamente, mas para este exemplo, iremos observar o mapira em relação a três diferentes mercados para os primeiros dois meses de produção durante a colheita, isto é, Novembro e Dezembro. Utilizando o Select Cases./If, estabelece um filtro que atinge as seguintes condições:

(mer = 3 or mer = 10 or mer = 41) and (mes=11 or mes=12).



Usando **Basic Tables**, prepare uma tabela com preços para dois meses e alterações percentuais:

1. **Analyze/Custom Tables/Basic Tables...**
2. Mude **p4c_1** para **Summaries:** e mude **p4c_p** para **Summaries:**
3. Mude **mer** para **Down:**
4. Move **cer** and then **mes** to **Across:**
5. *Click* no **Layout...**
6. Na caixa de **Summary Variable Labels**, *click* o botão circular perto de **Across the top**
7. **Continue**
8. **Titles...**
9. Na caixa de **Title** tecle: **Mudança de preço do mapira para um mês por mercado**
10. **Continue**
11. Faça o **Paste** e **Run**

Observe a tabela no SPSS Viewer: veja a alteração do preço para Novembro e como a variável e os resultados estão apresentados. Este é uma excelente maneira para preparar relatórios ou boletins sobre mercados.

Mudança de preço do mapira para um mês por mercado

		Cereal			
		Mapira			
		Mês			
		Novembro		Dezembro	
		Preço Médio do Consumidor	Mudança no Preço	Preço Médio do Consumidor	Mudança no Preço
Mercado	BADALA	76,50	-1,20	70,88	-5,63
	DIBIDA	74,10	-,65	67,63	-6,48
	KOUTIALA	52,58	-3,00	47,40	-5,18

Exercício 8.2

Divirta-se modificando o output das tabelas a medida que vai variando o estilo (propriedades, cores, formato, extensão e etiqueta) para vários mercados do milho (**cer = 9**) e o mexoeira (**cer = 7**) para diferentes meses. Depois, tente fazer um gráfico sobre vários produtos para vários mercados usando o filtros diferentes de **Select Cases** conforme acima fizemos: elabora pelo menos um gráfico sobre mudanças percentuais para cereais e outro gráfico sobre quantidades e preços para o milho para cada ano de produção.

Taxas de Crescimento

Uma tarefa comum para economistas agrícolas e outros analistas, quanto aos sistemas de informação de mercados, é calcular as taxas de crescimento para produção e outros tipos de variáveis.

1. **File/Open...**, e vá para **Files of type:** e seleccione **Lotus (*.w*)**
2. *Click* no ficheiro **maliprod.wk1** que aparecerá na parte de baixo e faça o **Paste**
3. Irá aparecer a caixa de diálogo de **Options**, *click* **OK**. Depois **Run** o comando
4. No **Data Editor**, tecele nos nomes das diferentes variáveis tal como se segue: **Campanha**, **ano**, **tons_pr** (toneladas produzidas) - os produtos são mexoeira, mexoeira, milho, arroz e *fonio* (um cereal local).
5. Agora calculemos a taxa de crescimento usando o comando **Curve Estimation**. Vá para **Analyze/Regression/Curve Estimation...**
6. Seleccione a variável do preço do consumidor **tons_pr** na coluna à esquerda e insira-a na caixa de **Dependent(s)**:
7. *Click* no botão circular perto de **Time** na caixa de **Independent**
8. Na caixa de **Models**, irá verificar que **Linear** já está escolhido. Seleccione também os modelos **Growth** e **Exponential**
9. Na parte de baixo da caixa de diálogo, *click* no botão circular perto de **Display ANOVA table**
10. Faça o **Paste** e **Run** o comando.

O que observa? Os modelos de crescimento e exponencial apresentam os mesmos resultados, a diferença consiste no termo constante. A variável temporal coeficiente beta refere-se a percentagem anual da taxa de crescimento. Guarde o seu sintaxe para o PRECREAL.SPS e feche a janela.

Dependent variable.. TON_PR Method.. EXPONENT

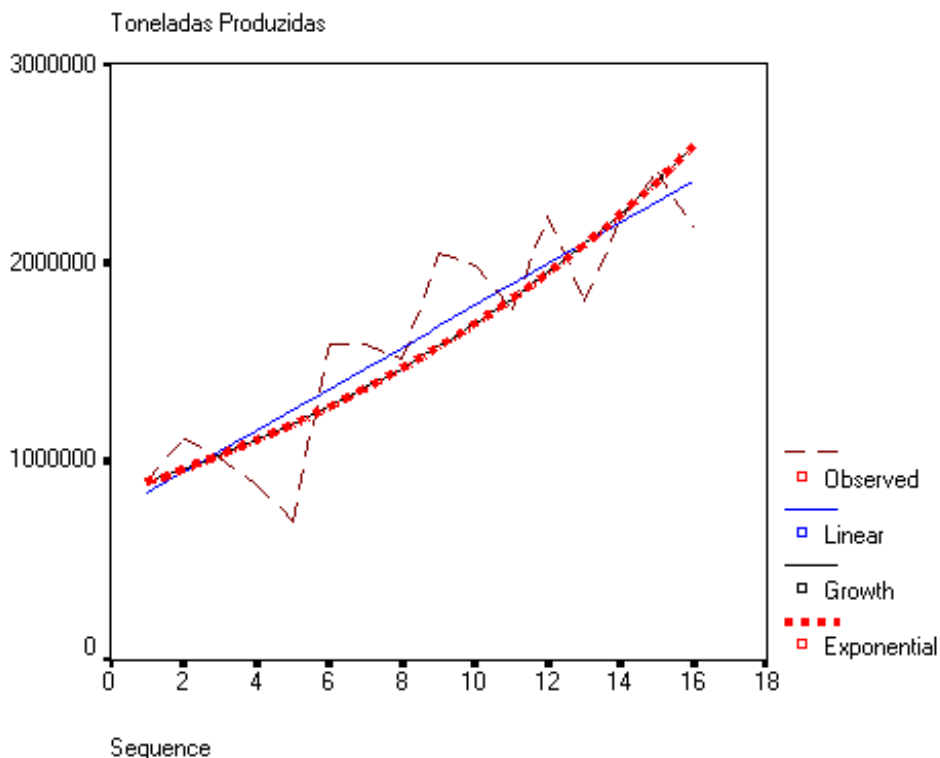
Multiple R ,86226
 R Square ,74350
 Adjusted R Square ,72518
 Standard Error ,20501

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	1	1,7054863	1,7054863
Residuals	14	,5883807	,0420272

F = 40,58055 Signif F = ,0000

Variables in the Equation						
Variable	B	SE B	Beta	T	Sig T	
Time	,070825	,011118	,862264	6,370	,0000	
(Constant)	833210,942117	89574,93202		9,302	,0000	



**SPSS para Windows “SAMPLE SESSION”
Módulo 9 - Análises Marginais (Séries Temporais)**

**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
Maio de 2000**

As margens de comercialização (*Marketed margins*) na sua forma mais simples, podem ser definidas como a diferença entre os preços pagos pelo consumidor por um produto, a nível retalhista e os preços recebidos pelos produtores quando eles vendem seus produtos para técnicos de montagem ou primeiros comerciantes ou intermediários (*first handlers*) (veja Goetz e Weber 1986, ID WP 29, Michigan State University para informações mais detalhadas). As margens reflectem o montante de serviços acrescidos ao produto quando este deixa a fábrica e está disponível para o consumidor de forma retalhista (i.e transformação espacial do produtor para o consumidor, e do *paddy* para o arroz sem casca). Esta margem bruta do consumidor-produtor será desagregada para margens mais precisas de modo a ilustrar os diferentes serviços acrescidos pelo sistema de comercialização.

É importante conceptualizar as várias dimensões em preços e visualizar o conjunto de funções de comercialização de actividades relacionadas de forma vertical do produtor para os técnicos de montagem, para o grossista, para o retalhista e finalmente para o consumidor. Os custos de produção e de insumos, compensação por esforços incorridos e custos de armazenamento, custos de transferência e de processamento, risco e por aí em diante. As margens comercializadas permitem assim que os analistas avaliem se os custos e os lucros são normais ou excessivos, eficiências relativas na implementação de serviços, comparação com o valor adicionado, estimativas do valor dos serviços e por aí em diante. Por isso, podemos ganhar perspicácia no desempenho do sub-sector. É um bom ponto de partida para análises mais aprofundadas sobre custos de comercialização e poderá ajudar a explicar as mudanças nos preços a diferentes níveis.

Antes de efectuarmos os cálculos das margens para diferentes mercados, precisamos medir o mercado do produtor, mercado de montagem, e dados sobre o preço a nível grossista e o preço do consumidor. Assim, primeiro precisamos de escolher um mercado para cada fase do canal de comercialização: para o nível do produtor iremos usar Zangasso (**loc=57**), para o nível de montagem será Koutiala (**loc=33**), e para o nível grossista/retalhista iremos usar Bamako (**loc=3**). Pelo facto de alguns preços terem sido recolhidos pelo SIMA no Mali a partir de 1993, iremos também seleccionar dados de 1993.

Manipulação de dados, ficheiros agregados e Merging para análises marginais

Para a lição que se segue, irá para cada mercado acima mencionado, preparar os dados seleccionando a mapira e a localidade correspondente (representando mercados numerosos dentro duma área específica) usando um filtro, e depois agregando-os num ficheiro mais pequeno para serem juntados com os outros ficheiros mais pequenos representando cada um dos diferentes mercados. Assim, iremos obter o nosso ficheiro do trabalho (*working file*) no qual iremos efectuar a análise marginal. Começemos por trabalhar no mercado do produtor.

MERCADO DO PRODUTOR

1. Active o ficheiro PQ_8995P3.sav (faça o Paste e Run o comando)
2. A partir de **Data** menu seleccione **Select Cases** e seleccione o botão circular perto de **If condition is satisfied**
3. *Click* no **If...** por baixo de **If condition is satisfied**
4. Insira **loc = 57 and cer =8 and ano >= 1993**
5. *Click* no **Continue**
6. Seleccione o botão circular perto de **Filtered**
7. Faça o Paste e Run o comando
8. Vá para **Data/Aggregate...**
9. A partir da coluna à esquerda, seleccione **cer ano mes c_ano c_mes sem** e insira-as na caixa de **Break Variable(s)**
10. Seleccione a variável **p4p** e insira na caixa de **Aggregate variable(s)**
11. Mude a etiqueta para “Preço Médio do produtor para Zangasso” e altere o nome para **p4p_zan** fazendo o *click* no **Name & Label...**
12. *Click* no **Continue**
13. Seleccione a variável **pr8** e insira na caixa de **Aggregate variable(s)**
14. Mude a etiqueta para “Preço Médio de Montagem para Zangasso” e altere o nome **pr8_zan** fazendo o *click* no **Name & Label...**
15. *Click* no **Continue**
16. *Click* no **File...** e altere o nome do ficheiro para MARGA.sav
17. **Save**
18. Faça o Paste e Run o comando.

MERCADO DE PREÇO DE MONTAGEM A NÍVEL GROSSISTA (preço bruto)

19. Desaccione o Filtro, depois mude o filter repetindo os passos 2 até 7 para seleccionar **loc = 33, cer =8 and ano >=1993**
20. Repita os passos 8 até 12 substituindo as variáveis na caixa de **Aggregate variable(s)** por **pr10**, dando o nome de **pr10_kou** e uma etiqueta como por exemplo “Preços médios mais comuns de montagem a nível grossistas (*Mean most common gross sale prices for Koutiala*)
21. *Click* no **File...** e altere o nome do ficheiro para MARGB.sav
22. *Click* no **Save**
23. Faça o Paste e Run o comando.

MERCADO RETALHISTA (preço do consumidor)

24. Desaccione o Filtro, depois altere o filtro repetindo o passo 19 para seleccionar **loc = 3, cer =8 and ano >=1993**

25. Repita o passo 20 substituindo a variável na caixa de **Aggregate variable(s)** por **p4c pga13 pgv16** dado-lhes os respectivos nomes **p4c_bko ga13_bko gv16_bko** e respectivas etiquetas: Preço Médio do consumidor para o Bamako (*Average consumer price for Bamako*), Preço médio de grossista no momento da compra para o Bamako (*Mean gross price at purchase*) e Preço médio de grossista no momento da venda para o Bamako (*Mean gross price at sale*)
26. *Click* no **File...** e altere o nome do ficheiro para MARGC.sav
27. *Click* no **Save**
28. Faça o **Paste** e **Run** o comando.

MERGING FILES

1. **File/Open...**
2. Selecciona **marga.sav**, faça o **Paste** e **Run** o comando
3. Vá para **Data/Merge File/Add Variables...**
4. Selecciona o ficheiro **margb.sav** e *click* no **Open**
5. Verifique a caixa perto de **Match cases on key variables in sorted files**
6. *Click* no botão circular perto de **Both files provide cases**
7. Selecciona **cer c_ano c_mes sem** da lista **Excluded Variables:**
8. *Click* no **▶** perto de **Key Variables:** (em baixo, à direita)
9. Faça o **Paste** do comando
10. *Click* no **OK**
11. Repita os passos 3 a 10 seleccionando o ficheiro **margc.sav**
12. Selecciona e **Run** ambos comandos.
13. Guarde o novo ficheiro com o nome de **margin.sav**

Irá verificar que no **Data Editor window** para alguns casos a mesma semana (**sem**) está repetida duas vezes pelo facto de os dois casos terem sido obtidos de ficheiros diferentes. Então precisamos de combinar estes ficheiros para eliminar o **sysmisses** e duplicações falsas. Podemos fazer isso accionando um comando **aggregate**. Tecla e **Run** no seu **syntax window** os seguintes comandos:

AGGREGATE

```

/OUTFILE=*
/BREAK=cer c_ano sem
/ano = MEAN(ano)
/mes = MEAN(mes)
/c_mes = First(c_mes)
/p4p_zan "Preço Médio do Produtor para Zangasso" = MEAN(p4p_zan)
/pr8_zan "Preço Médio de Montagem para Zangasso" = MEAN(pr8_zan)
/pr10_kou "Preço de montagem mais comum de grossista para Koutiala" =MEAN(pr10_kou)
/p4c_bko "Preço Médio do consumidor para Bamako" = MEAN(p4c_bko)
/ga13_bko "Preço médio de grossista no momento de compra" = MEAN(ga13_bko)
/gv16_bko "Preço médio de grossista no momento de venda" = MEAN(gv16_bko).

```

SORT CASES BY

```
c_ano (A) c_mes (A) sem (A) .
```

```
SAVE OUTFILE='D:\Datad\SPSS\SAMPLE\margin.sav'
```

```

/keep cer ano mes c_ano c_mes sem p4p_zan pr8_zan pr10_kou p4c_bko ga13_bko gv16_bko
/COMPRESSED.

```

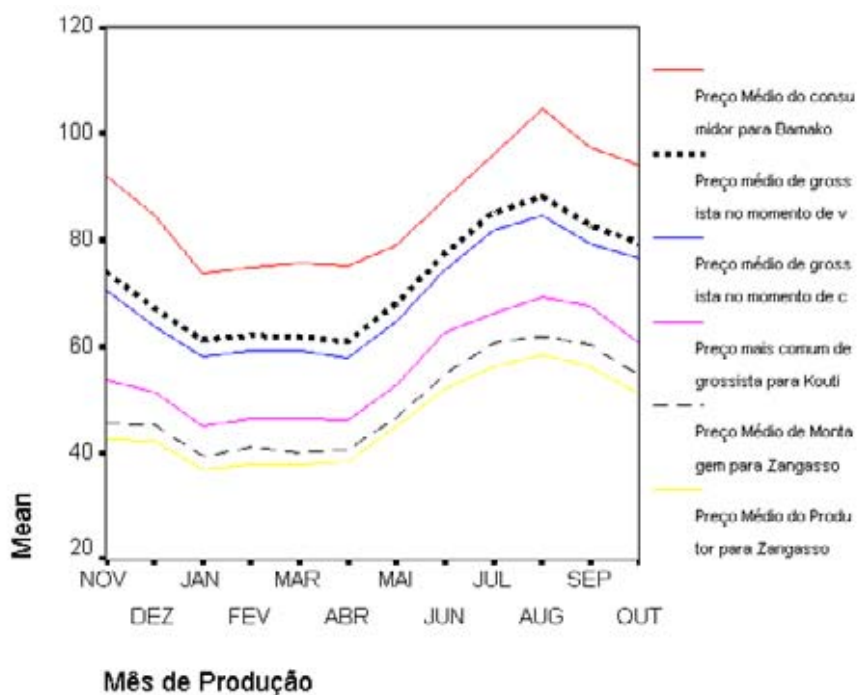
Gráfico para verificar a lógica dos preços

Como foi explicado no módulo 4, efectuar a limpeza dos dados é fundamental e deve ser executado antes de qualquer análise. Como somente verificamos alguns preços tais como do consumidor, de montagem e do produtor, agora precisamos de verificar os outros preços que

aqui introduzimos (**pr10**, **ga13** and **gv16**). Nesta ocasião, iremos usar uma técnica gráfica para verificar os preços. Abra o ficheiro **margin.sav** se ainda não o tiver feito.

1. Vá para **Graphs/Line...**
2. *Click* no quadrado perto de **Multiple**
3. O botão circular perto de **Summaries of separate variables** deve ser seleccionado
4. *Click* no **Define**
5. Selecciona **p4c_bko gv16_bko ga13_bko pr10_kou pr8_zan** e **p4p_zan** da coluna à esquerda e insira na caixa de **Lines represent** .
6. Agora escolha a variável **c_mes** na coluna à esquerda e *click* no **▶** perto de **Category Axis:**
7. *Click* no **Options...**
8. *Click* no botão circular perto de **Display groups defined by missing values**
9. **Continue**
10. Faça o **Paste** e **Run** o comando

Observe o gráfico, as curvas permitem-nos verificar se alguma delas cruza-se com outra curva. Isso iria significar que o preço lógico não foi respeitado, por exemplo, um preço do produtor mais elevado que um preço de montagem em Zangasso. Como não é o caso podemos aprofundar as nossas análises. O gráfico permite-nos verificar a lógica sequencial de cada preço (veja a mudança nas cores de cada linha e etiqueta correspondente). Se assim o desejássemos, pedíamos também verificar os preços de forma mais aprofundada, olhando para cada ano de produção usando o comando **Split file**.



Cálculo e elaboração de gráficos sobre as contribuições do canal de comercialização (Opcional)

Pelo facto de o módulo 9 ser um pouco longo esta secção foi escolhida como um exercício opcional. Refere-se as várias partes componentes do canal de comercialização até ao nível do consumidor.

O quociente do preço do produtor pelo preço do consumidor (ou preço mundial de referência) para um produto comercializável é a parte do preço que o produtor recebe pela venda de um produto no mercado. Poderá também indicar a possibilidade de se aumentar o preço do produtor no caso de não existirem distorções, por exemplo, políticas específicas que influem os preços.

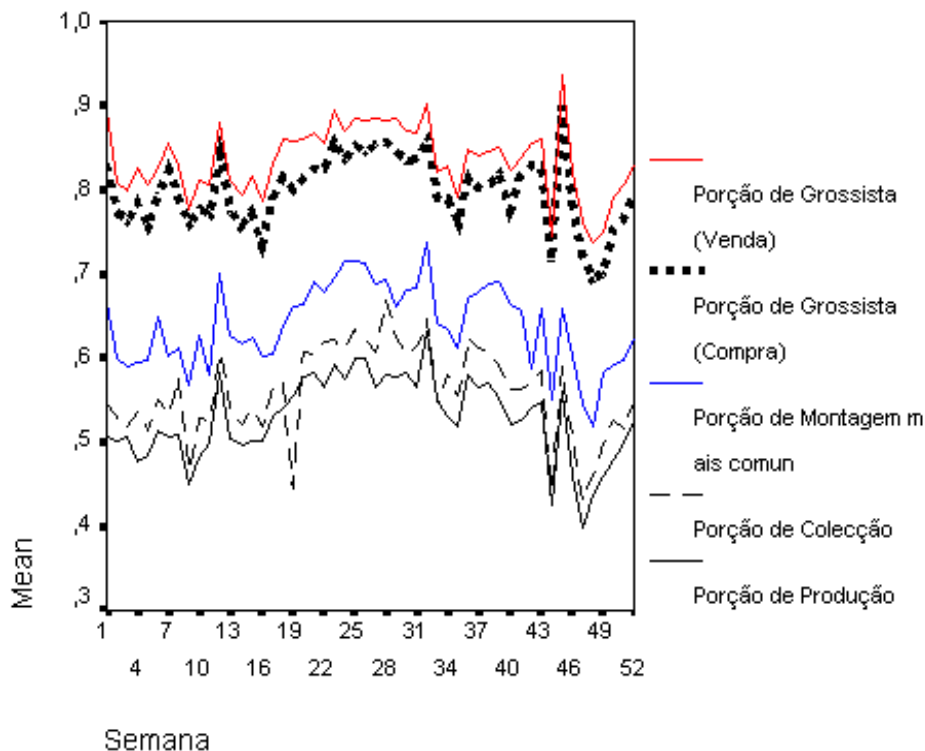
Por isso, podemos calcular a parte de cada nível dividindo cada preço pelo preço do consumidor em Bamako. Calcule estas partes usando os seguintes passos:

1. A partir do **Transform** seleccione **Compute...**
2. Para o **Target Variable**: insira **porcprod** e uma etiqueta a sua escolha (e.g. Porção de Produção)
3. A partir da caixa de **Numeric Expression**:(Expressões Numéricas), insira **p4p_zan / p4c_bko**
4. Faça o **Paste** e **Run** o comando
5. Repita os passos 1 a 4 substituindo a **target variable** e **numeric expressions** pela seguinte informação (dê a cada um uma etiqueta apropriada):

Target Variable	Numeric Expression	Etiqueta
mon1porc	pr8_zan / p4c_bko	Porção de colecção
mon2porc	pr10_kou / p4c_bko	Porção de montagem (preço bruto mais comum)
gr1porc	ga13_bko / p4c_bko	Porção de grossista (compra)
gr2porc	gv16_bko / p4c_bko	Porção de grossista (venda)

Demonstremos agora graficamente os quocientes dos preços relativos de modo a termos a ideia das variações ao longo do ano, e das percentagens para cada nível. Use o comando **Graph** e assegure-se de inserir **gr2_porc gr1_porc mon2porc mon1porc** e **porcprod** da coluna à esquerda na ordem em que os coloca na **Lines represent box**. Apresente graficamente as partes por semanas (**sem**). Este sintaxe deve apresentar-se da seguinte forma:

```
GRAPH
/LINE(MULTIPLE)=MEAN(gr2porc) MEAN(gr1porc) MEAN(mon2porc) MEAN(mon1porc)
MEAN(porcprod)
BY sem.
```



Cálculo e apresentação gráfica de margens brutas entre diferentes níveis de comercialização

Para calcular margens brutas siga as seguintes instruções:

1. A partir de **Transform** seleccione **Compute...**
2. Para o **Target Variable**: insira **marg1** e uma etiqueta a sua escolha (e.g. Margem de Colecção (Primeira Compra))
3. Para a caixa do Numeric **Expression**, insira **pr8_zan - p4p_zan**
4. Faça o **Paste** e **Run** o comando.
5. Repita os passos 1 a 4 substituindo a **target variable**, **label** e **numeric expressions** pela seguinte informação:

Target Variable	Label (Etiqueta)	Numeric Expression
marg2	Montagem	pr10_kou - pr8_zan
marg3	Margem bruto rural	ga13_bko - pr10_kou
marg4	Margem bruto urbano	gv16_bko - ga13_bko
marg5	Retalhista	p4c_bko - gv16_bko

Quanto tiver calculado todas as margens, tente apresentar graficamente os dados da margem por semana para todas as cinco variáveis marginais juntas. O que observa? As variações são muito numerosas e de compreensão complexa. Para se obter um bom entendimento dos dados, foi efectuada no módulo 1 uma boa análise para efectuar a sua estatística descritiva do seu desempenho.

1. A partir de **Analyze** seleccione **Descriptive Statistics/Descriptives..**
2. Seleccione **marg1 marg2 marg3 marg4 marg5** da lista à esquerda e insere na caixa das **Variable(s)**:
3. *Click* no **Paste** e **Run** o comando

Olhando para os resultados, podemos verificar que as margens médias são mais elevadas para

Descriptive Statistics						
	N	Minimum	Maximum	Mean	Std. Deviation	
MARG1 Margem de Colecção	146	-8,00	12,60	3,0148	2,3457	
MARG2 Margem de Montagem	139	-21,20	15,50	6,4748	3,9736	
MARG3 Margem bruto rural	143	5,43	44,77	13,5761	4,6971	
MARG4 Margem bruto urbano	150	,27	7,03	3,1444	1,0334	
MARG5 Margem retalhista	149	3,23	34,15	14,0222	4,8876	
Valid N (listwise)	136					

as margens rurais retalhistas e brutas e são mais baixas para as margens brutas urbanas e de colecção. Veja como as margens variam: observe os mínimos e os máximos. Como podemos interpretar essas variáveis? Anote e guarde as suas ideias no **SPSS Viewer** (*Margem.spo*) e discuta-os com os seus colegas. Antes de observarmos as margens médias anuais (*mean annual margins*) em vários gráficos, guarde os dados (*margin.sav*) pelo facto de já termos calculado muitas variáveis e gostaríamos de guarda-las para análises mais aprofundadas.

Exercício 9.1

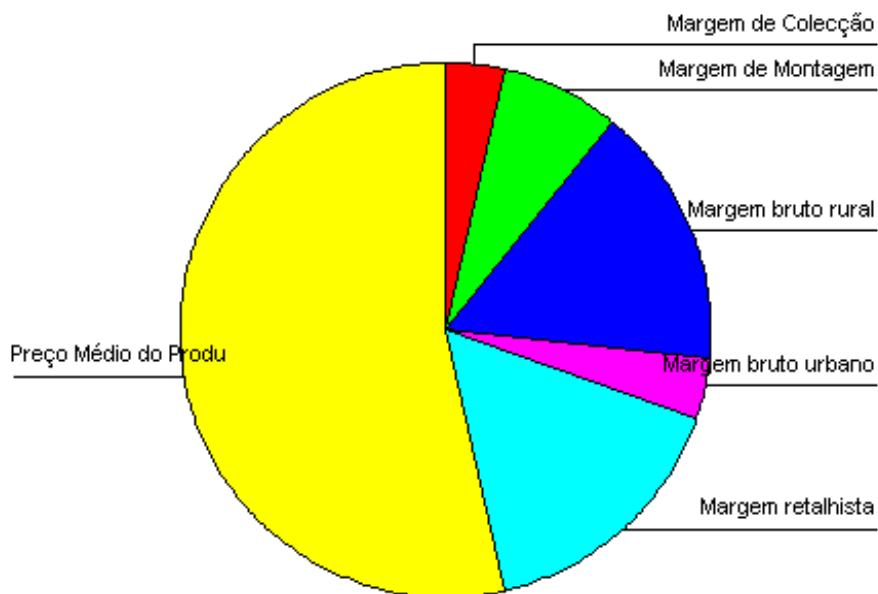
Poderá ser de interesse para os preços de “lag” ou retardamento calcular as margens “lagged” dependendo dos dias dos diferentes mercados para um dado canal, e do tempo para se transferir o produto do produtor para o comprador de montagem para o vendedor grossista para o consumidor, que poderá levar dias á semanas. Aplique o que aprendeu na primeira secção do módulo 5, utilizando os preços *lagged* por duas semanas. Depois, calcule e apresente graficamente as margens *lagged* com base no que aprendeu nas margens brutas. O que observa?

Elaboração do gráfico circular (“torta”) e de barras para margens de mercados à diferentes níveis

As margens médias anuais apresentado graficamente na forma circular.


1. Vá para **Graphs/Pie...**
2. O botão circular perto de **Summaries of separate variables** deve ser seleccionado
3. *Click* no **Define**
4. Seleccione **marg1 marg2 marg3 marg4 marg5** e **p4p_zan** da coluna à esquerda e insira na caixa de **Slices represent**.

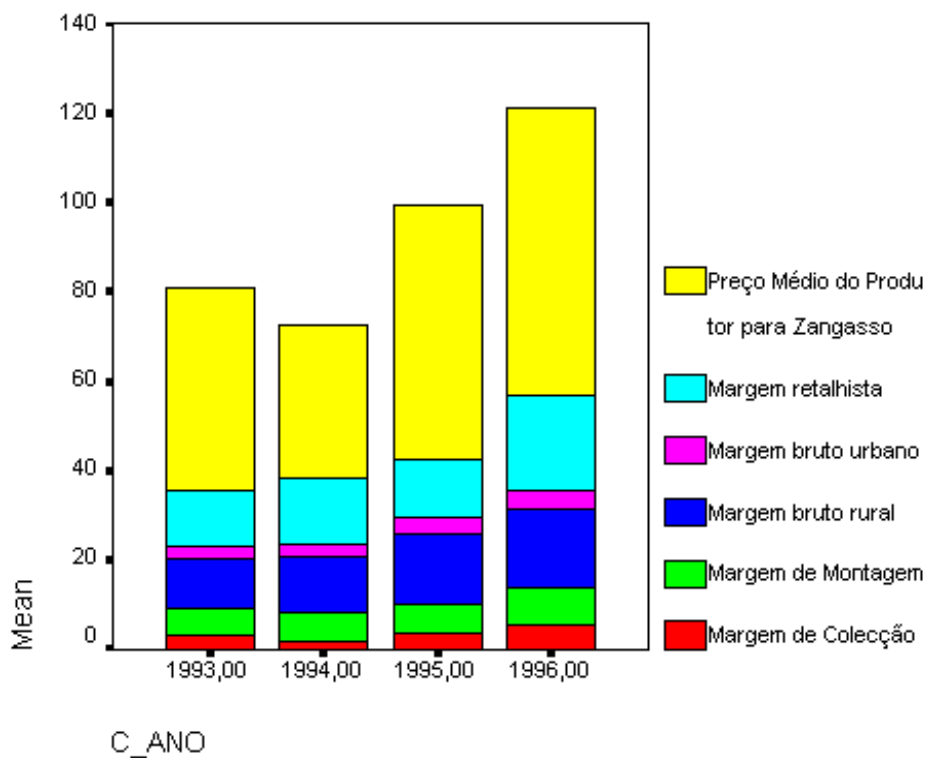
5. Click no **Change Summary**
6. Click no botão circular perto dos Mean of values
7. **Continue**
8. Faça o Paste e Run o comando.



A forma circular dá-nos uma ideia geral sobre a distribuição das margens mas observemos cuidadosamente a distribuição anual usando um gráfico de barras e determine se existem algumas mudanças:

Margens anuais apresentadas por ano num gráfico de barras.

1. Vá para **Graphs/Bar...**
2. Click no quadrado perto de **Stacked**
3. O botão circular perto de **Summaries of separate variables** deve ser seleccionado
4. Click no **Define**
5. Selecciona **marg1 marg2 marg3 marg4 marg5** e **p4p_zan** da coluna à esquerda e insira na caixa de **Bars represent**.
6. Agora escolha a variável **c_ano** na coluna à esquerda e click no  perto de **Category Axis**:
7. Click no **Options...**
8. Click no botão circular perto de **Display groups defined by missing values**
9. **Continue**
10. Faça o Paste e Run o comando.



Veja como as margens médias totais decrescem em 1994 e depois aumentam em 1995. O aumento absoluto mais elevado está no nível do produtor (**p4p_zan**) mas podemos ver um aumento relativo elevado também para o nível de montagem (**colecção**).

Agregar os Dados para níveis mensais e anuais e calcular coeficientes de variações em margens

Os coeficientes de variação (CV) são calculados através da divisão do desvio padrão pela média. Precisamos construir um ficheiro diferente de modo a podermos reorganizar os dados que nos permitam calcular os coeficientes de variação. Fazemos isso de modo a analisar a variabilidade *intra*-anual nas diferentes margens, que nos irá dar uma ideia sobre que níveis do sistema absorvem as mudanças dos preços. Por isso, antes de agregarmos os dados, active o ficheiro **margin.sav** se ainda não o tiver feito.

1. Para agregar, vá para **Data/Aggregate...**
2. A partir da coluna à esquerda, seleccione **c_ano** e insira na caixa de **Break Variable(s)**
3. Seleccione a variável **marg1** e insira na caixa de **Aggregate variable(s)**.
4. Seleccione outra vez a variável **marg1** e uma vez na caixa de **Aggregate variable(s)**, *click* outra vez, e depois *click* no **Function...**
5. *Click* no botão circular perto de **Standard deviation**
6. *Click* no **Continue**
7. Repita os passos 3 à 6 mas substitua a variável **marg1** com **marg2 marg3 marg4 marg5** e **p4p_zan**, mas mantenha preccionado o botão à esquerda do rato para copiar de todas as variáveis para a **Aggregate variable(s)** box ao

- mesmo tempo, mantenha-os destacados, depois *click* **Function...** para altera-los todos para **Standard deviation**.
8. *Click* no **File...** e altere o nome do ficheiro para **MARGPLUS.SAV**
 9. **Save**
 10. Faça o **Paste** e **Run** o comando.

Agora para fazer o cálculo, active o ficheiro **MARGPLUS.SAV** e depois siga os seguintes passos:

1. A partir de **Transform** seleccione **Compute...**
2. Para o **Target Variable**: teclé **cv_p4p** (e acrescente uma etiqueta a sua escolha)
3. A partir de caixa de **Numeric Expression**: , teclé **p4p_z2 / p4p_z1**
4. Faça o **Paste** e **Run** o comando.
5. Repita os passos 1 à 4 substituindo a target variable e numeric expression pela seguinte informação:

Target Variable	Numeric Expression
cv_marg1	marg1_2 / marg1_1
cv_marg2	marg2_2 / marg2_1
cv_marg3	marg3_2 / marg3_1
cv_marg4	marg4_2 / marg4_1
cv_marg5	marg5_2 / marg5_1

Com um simples gráfico, observemos a coeficiente média de variação para as várias margens de 1994 para 1995 (podemos também usar tabelas ou outras formas de apresentações). Apresente graficamente as seis variáveis acima calculadas por **c_ano**. Todas as curvas aparentam seguir a mesma tendência excepto a variável **cv_marg1** em 1994, que aumentou e tem um valor de 1,18! Poderia parecer que isto deve-se ao facto da existência de um erro (elevado desvio padrão). No seu todo, poderá parecer que as grossistas rurais (que são activos na **marg2** e **marg3**) estão a absorver grande parte da variação do mercado de preços.

Como iremos usar **margin.sav** para agregar a um nível mensal e deflacionar margens por taxas de inflação, e como estamos presentemente a um nível anual, guarde este ficheiro com um nome diferente – **margplus.sav** --de modo a utiliza-lo para análises posteriores mais aprofundadas.

Margens deflacionadas por taxas de inflação

Refira-se ao módulo 8 para uma breve explicação sobre preços deflacionários. Abra o ficheiro **margin.sav** (faça o **Paste** e **Run** o comando), depois faça a agregação por **ano** e **mes** seguindo estas instruções:

1. Vá para **Data/Aggregate...**
2. A partir da coluna à esquerda, seleccione **ano** e **mes** e coloque-os em **Break Variable(s)**
3. Seleccione a variável **marg1** e coloque na caixa de **Aggregate variable(s)**.
4. Repita os passos 3 para **marg2 marg3 marg4 marg5** e **p4p_zan** todos de uma

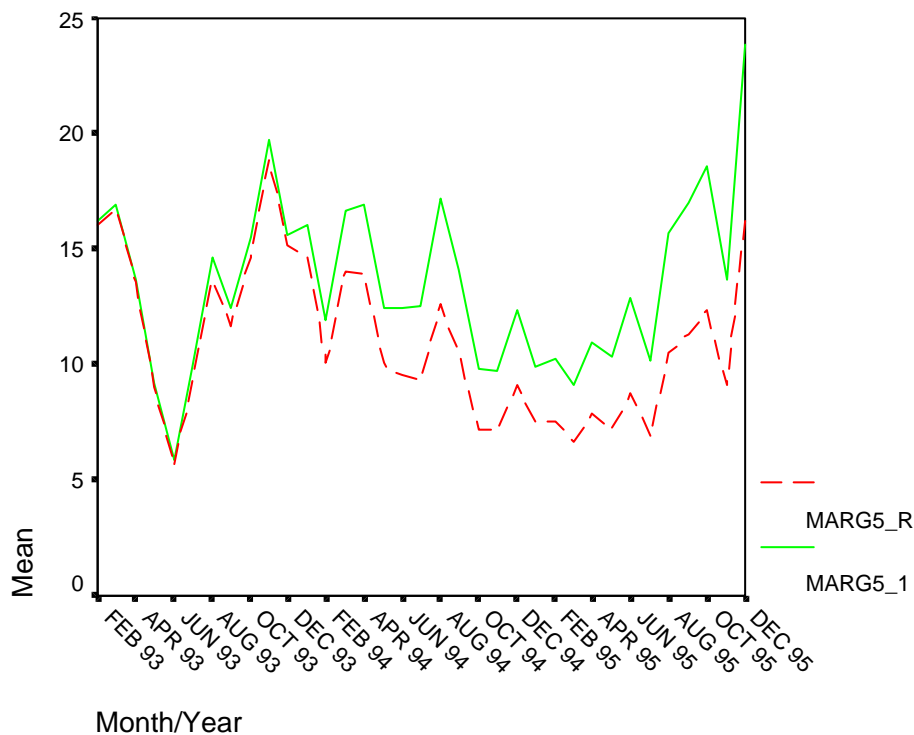
- só vez mantendo pressionada
5. *Click* no **File...** e altere o nome do ficheiro para MARGIND.SAV
 6. **Save**
 7. Faça o **Paste** e Run o comando
 8. Pretendemos abrir o consumer index file que guardamos para deflacionar a margem retalhista então precisamos activar o ficheiro **pindice.sav** (faça o **Paste** e Run o comando). Uma vez aberto, seleccione os dados para 1993 de cima para baixo (faça o **Paste** e Run o comando).
 9. Agora podemos fazer o **Merge** dos dois ficheiros.
 10. Vá para **Data/Merge File/Add Variables...**
 11. Seleccione o ficheiro **margind.sav** e *click* no **Open**
 12. Verifique a selecção de caixa perto de **Match cases on key variables in sorted files**
 13. *Click* no botão circular perto de **Both files provide cases**
 14. Seleccione **ano mes** da lista de **Excluded Variables**:
 15. *Click* no **▶** perto de **Key Variables**: (em baixo, à direita)
 16. Faça o **paste** do comando
 17. *Click* no **OK**
 18. *Click* no **NO** e depois Run o comando

Agora que temos o nosso ficheiro activo (*working file*), podemos calcular a margem retalhista real. Como anteriormente feito, é preciso usar **Transform/Compute**, depois para o **Target Variable**: tecle **marg5_r** e uma etiqueta correspondente, e na caixa de **Numeric Expression**: tecle **marg5_1 / pindex*100**. Guarde estes novos dados do ficheiro sob o nome de **margreal.sav**.

Antes de preparar-mos os gráficos para observar as margens reais e nominais, para a apresentação, é fundamental ter-se uma data para a variável com um período de tempo combinado tal como mês ou ano. Podemos usar a função **DATE.MOYR(month,year)** no comando **Compute** conforme se segue (se ainda não estiver apresentado nos dados):

1. A partir de **Transform** menu seleccione **Compute...**
2. Para o **Target Variable**: tecle **date_m**
3. *Click* no **Type&Labels...** e tecla uma etiqueta (e.g. Mes/Ano).
4. *Click* no **Continue**
5. A partir de **Functions**: box, seleccione a função **DATE.MOYR(month,year)** e coloque na caixa de **Numeric Expression**.
6. Substitua ambas ? funções por **mes** e **ano**
7. Faça o **Paste** e Run o comando.

Agora, apresente graficamente a margem retalhista nominal e a margem retalhista real na forma de **date_m**. Observe como as duas curvas começam a divergir no princípio de 1994! Os resultados demonstram as margens em termos reais, tomando em consideração a elevada taxa de inflação após a devaluation. Se tivéssemos outros índices de preços, poderíamos ter criado as outras margens em termos reais. Poderia ter sido particularmente interessante se tivéssemos índices de preços a nível grossista mas eles são de difícil obtenção em África.



Margens Brutas

Nesta sub-sessão iremos demonstrar como avaliar o efeito dos índices de preços do consumidor nas margens, avaliar os efeitos de escala nos serviços de comercialização e avaliar o comportamento da determinação dos preços (*price setting behavior*).

A REGRESSÃO DE MARGEM BRUTA DE COMERCIALIZAÇÃO POR ÍNDICE DO PREÇO DE CONSUMIDOR PARA AVALIAR O EFEITO DOS ÍNDICES NAS MARGENS

Como estamos interessados na forma pela qual as margens de comercialização comporta-se ao longo do tempo, e a medida que as quantidades estabelecidas para o sistema num subsector particular aumentam e diminuem, as margens de comercialização poderão ser analisadas com mais pormenor usando técnicas sobre regressão linear. Assim, podemos examinar a relação existente entre uma variável dependente e outra independente tal como a margem real e o índice de preço. A regressão também produz algumas estatísticas (R^2 , F, ANOVA COEFF, e por aí em diante) dependendo da escolha da regressão feita.

Primeiro precisamos calcular a margem bruta. Mantenha o mesmo ficheiro do sub-sector acima usado (*margreal.sav*). A margem bruta poderá ser calculada da seguinte forma:

1. A partir de **Transform** menu seleccione **Compute...**
2. Para a **Target Variable**: teclie **margbrut** (etiqueta: margem bruta)
3. A partir de Numeric **Expression**: box, teclie **marg1_1 + marg2_1 + marg3_1 + marg4_1 + marg5_1**
4. Faça o **Paste** e **Run** o comando
5. Pretendemos eliminar os dados omissos para Janeiro de 1993. Então, use o comando **Select cases** para **margbrut > 0** e depois guarde os dados do ficheiro *margbrut.sav*.

Para usar uma regressão linear, siga os seguintes passos:

1. Vá para **Analyze/Regression/Linear...**
2. Coloque **margbrut** na Dependent: box
3. Coloque **pindice** na caixa de Independent(s):
4. Faça o **Paste** e **Run** o comando

Se existe uma mudança de um por cento na taxa de inflação, então a margem bruta aumenta ou diminui pelo valor do coeficiente Beta. Este é o efeito inflacionário na margem bruta. Os resultados demonstram um R Square relativamente bom que indica uma boa *fit* ou ajustamento e o termo do índice é significativamente diferente de 0 (como significante para os testes T e F abaixo de 0,05).

Coefficients^a


Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	14,436	6,401		2,255	,031
	PINDEX	,199	,050	,568	3,961	,000

a. Dependent Variable: Gross margin

A REGRESSÃO DE MARGEM BRUTA DE COMERCIALIZAÇÃO POR QUANTIDADE COMERCIALIZADA PARA AVALIAR OS EFEITOS DE ESCALA EM SERVIÇOS DE COMERCIALIZAÇÃO

Como temos dados insuficientes, será difícil diferenciar quantidades por localização do mercado. Para este exemplo, iremos usar o arroz como produto e iremos seleccionar Niono-Bamako como o canal de comercialização. Active o ficheiro PQ_8995P3.SAV. Depois seleccione e agregue para Bamako e Niono o seguinte:

1. A partir de **Data** menu seleccione **Select Cases** e seleccione o botão circular perto de **If condition is satisfied**
2. **Click** no **If...** por baixo de **If condition is satisfied**
3. Tecele **loc = 3 and cer =4 and ano >= 1993**
4. **Click** no **Continue**
5. Seleccione o botão circular perto de **Filtered**
6. Faça o **Paste** e **Run** o comando
7. Vá para **Data/Aggregate...**
8. A partir da coluna à esquerda, seleccione **ano mes** e **sem** e insira-os na caixa de **Break Variable(s)**
9. Seleccione a variável **p4c** e insira-a na caixa de **Aggregate variable(s)**.
10. **Click** no **File...** e altere o nome do ficheiro para **REARROZ.SAV**
11. **Save**
12. Faça o **Paste** e **Run** o comando

13. Repita os passos 1 à 6 e substitua **loc = 3** por **loc = 43** por baixo de **If condition is satisfied**
14. Repita os passos 7 à 9, substitua a variável agregada **p4c** por **p4p** e **q6** e depois **click** no botão circular perto de **Replace working data file**
15. Faça o **paste**, **click NO** e **Run** o comando
16. Vá para **Data/Merge File/Add Variables...**
17. Selecciono o ficheiro **rearroz.sav** e **click** no **Open**
18. Verifique a selecção da caixa perto de **Match cases on key variables in sorted files**
19. **Click** o botão circular perto de **Both files provide cases**
20. Selecciono **ano mes sem** da lista **Excluded Variables:**
21. **Click** no  perto de **Key Variables:** (em baixo, à direita)
22. Faça o **paste** do comando
23. **Click** no **OK**
24. **Click** no **NO** e depois **Run** o comando
25. A partir dos dados irá verificar que para a mesma semana (coluna **sem**), temos dois meses diferentes (ex. **sem = 13**, **mes = 4** **mes = 5**) o que nos dá a impressão de que existem dados omissos ou duplicados. Uma forma de resolver este aspecto é agregar outra vez por semana eliminando o mês do seguinte modo:
Faça o **open** do ficheiro **PQ_8995P3.SAV**. Substitua o mesmo **filter** como acima feito e vá para **Data/Aggregate...**
26. A partir da coluna à esquerda, seccione **ano** e **sem** e insira-os em **Break Variable(s)** box
27. Seccione as variáveis **p4c q6** e **p4p** insira-as em caixa de **Aggregate variable(s)**. Altere-as de **p4c_1 q6_1** e **p4p_1** para **p4c**, **q6** e **p4p**, e faça de novo as suas etiquetas
28. **Click** no botão circular perto de **Replace working data file**
29. Faça o **Paste** e **Run** o comando
30. Lembre-se de fazer o **save** do novo data file (com **escala.sav**)

Agora temos que calcular a margem para o arroz.

1. Calcule esta margem usando o comando **Compute** para a **target variable mararroz** tal como fez antes (a expressão numérica é **p4c - p4p**) e depois calcule a regressão linear do seguinte modo:
2. Faça o **Paste** do comando **Split file** por **ano** (certifique-se que os botões circulares **Compare groups** e **File is already sorted** estão seleccionados antes de fazer o **Paste**)
3. Vá para **Analyze/Regression/Linear...**
4. Insira **mararroz** na caixa de **Dependent:**
5. Insira **q6** na caixa de **Independent(s):**
6. Faça o **Paste** e **Run** o comando

Observe cada um dos três anos, poderemos verificar que os resultados não eram muito bons para os anos 1994 e 1995, e de algum modo são melhores para 1993 como o quadrado **R** demonstrou um ajustamento ligeiramente mais forte (estimativa optimista de como o modelo se ajusta bem com os dados) e o único resultado que é **significante**. As margens de comercialização aumentam ligeiramente para os primeiros dois anos com maior volume para o canal de comercialização (valor de regressão do coeficiente **B**).

A REGRESSÃO DO PREÇO DE UM NÍVEL POR OUTRO NÍVEL PARA AVALIAR O COMPORTAMENTO DA DETERMINAÇÃO DO PREÇO

Será que os preços retalhistas influenciam os preços do produtor, será que os preços retalhistas são dependentes dos preços do produtor? Podemos tentar responder estas questões usando a regressão linear. Use os seguintes passos para regressar os preços do consumidor por os preços do produtor:

1. Vá para Analyze/Regression/Linear...
2. Insira **p4c** na caixa de **D**ependent:
3. Insira **p4p** na caixa de **I**ndependent(s):
4. Faça o **P**aste e **R**un o comando

Os resultados demonstram um ajustamento muito bom já que o quadrado R é ligeiramente alto e os resultados também são muito significativos. O coeficiente para o preço do produtor (**p4p**) é significativamente diferente de 1 então não existe um aumento do preço *ou markup* constante no canal do preço do produtor para o preço retalhista. Os retalhistas e os intermediários não são somente tomadores de preços. Os preços retalhistas alteram em mais de uma unidade quando os preços a nível rural alteram.

Observemos a relação do preço do produtor/retalhista sob um ângulo diferente: será que o preço a nível do produtor altera em resposta das mudanças do preço retalhista?

1. Vá para Analyze/Regression/Linear...
2. Insira **p4p** na caixa de **D**ependent:
3. Insira **p4c** na caixa de **I**ndependent(s):
4. Faça o **P**aste e **R**un ambos comandos

Irá parecer que o ajustamento é razoavelmente bom e outra vez significativo --- e o valor de β também é elevado, perto de 1. Guarde o seu syntax para o **MARGEM.SPS** e faça o close.

Exercício 9.2

Este será um exercício que poderá levar tempo a ser feito uma vez que requer a repetição de todos os diferentes tipos de análises de margens efectuadas neste módulo. Deverá usar o milho (**cer = 9**) como o bem para este exercício. Repita cada uma das análises individuais que se referem a sua pessoa a medida que segue os passos do módulo para cada uma das análises. Tente interpretar os resultados e compare estes com os seus colegas. Faça sempre o **Paste** dos comandos para o Syntax editor e lembre-se que poderá teclar você mesmo em algum dos comandos. Não se esqueça de guardar os seus novos data files, o seu Syntax editor para documentação e qualquer gráfico ou resultado que poderá escolher para posteriores análises.

- i) Primeiro, você precisa de agregar e fazer o **Merge** dos files para as análises marginais. Use a mesma "localités" ou localidades (**loc**) e os mesmos períodos de tempo como no módulo. Os preços para o milho estão disponíveis nos vários níveis de comercialização.
- ii) Faça o gráfico para verificar a lógica do preço e modificar ou recodificar se necessário.

- iii) Calcule e faça o gráfico da contribuição do produtor assim como das margens brutas entre os níveis de comercialização. Pratique usando os gráficos circular e de barras.
- iv) Agregue em níveis mensais e anuais e calcule os coeficientes de variação.
- v) Deflacione as margens através da taxa de inflação. Use a variável **pindice** e aplique-a para os preços dos cereais.
- vi) Tente avaliar os efeitos dos índices do preço do consumidor nas margens, avaliar os efeitos de escala nos serviços de comercialização e avaliar o comportamento da determinação do preço usando a regressão linear conforme ilustrado neste módulo.