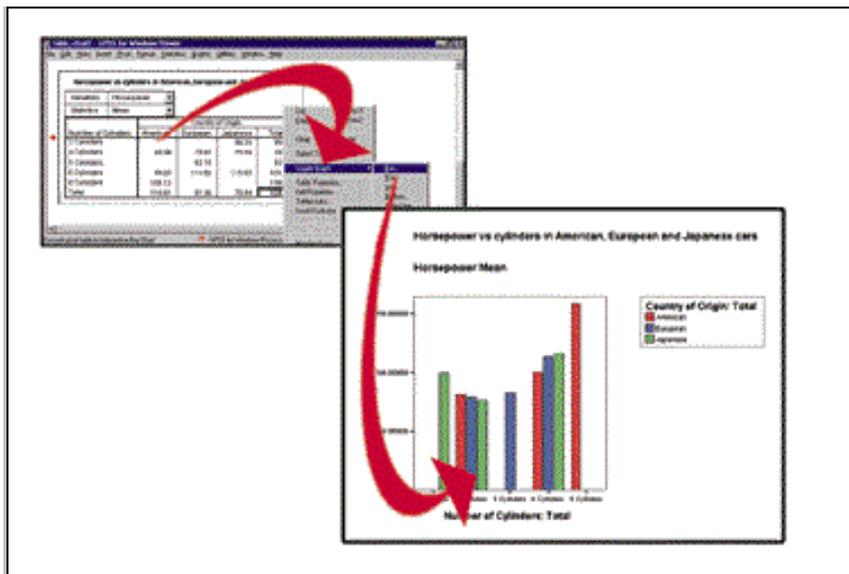


## Análise da pobreza com SPSS



*Manual do curso*

*compilado por Klaus Röder  
– Beira, Moçambique outubro 2004*

## CONTEÚDO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>4</b>
1.1	CONVENÇÕES DE ESTILOS DO TEXTO .....	5
1.2	COMO ESCOLHER AS OPÇÕES DOS MENUS .....	5
<b>2</b>	<b>UMA VISÃO GERAL SOBRE O SPS .....</b>	<b>5</b>
2.1	COMO EDITAR OS DADOS .....	8
2.2	ETIQUETAS DAS VARIÁVEIS E VALORES .....	9
2.3	EXERCÍCIOS .....	9
2.4	A PRIMEIRA ANÁLISE COM O SPSS .....	10
2.5	OS TIPOS DE VARIÁVEIS PARA A ANÁLISE ESTATÍSTICA .....	13
2.6	OS TIPOS DE VARIÁVEIS DO SPSS .....	13
2.7	EXERCÍCIOS .....	14
<b>3</b>	<b>A TRANSFORMAÇÃO DAS VARIÁVEIS.....</b>	<b>15</b>
3.1	DEFININDO AS VARIÁVEIS .....	15
3.2	NOMES DAS VARIÁVEIS NO SPSS .....	15
3.3	VALORES MISSING .....	16
3.4	CÁLCULO COM VARIÁVEIS.....	17
3.5	TRANSPONDO VARIÁVEIS .....	17
3.6	ORDENAR CASOS .....	17
3.7	SELECIONAR CASOS .....	18
3.8	EXERCÍCIOS .....	19
<b>4</b>	<b>CONSIDERAÇÕES PRELIMINARES SOBRE ÍNDICES DA POBREZA E A ANÁLISE .....</b>	<b>20</b>
4.1	A TRANSFORMAÇÃO DE ARQUIVOS DE DADOS .....	20
4.2	COMBINAÇÃO DE ARQUIVOS: ADICIONE CASOS.....	20
4.3	COMBINAÇÃO DE ARQUIVOS: ADICIONE VARIÁVEIS .....	21
4.4	A AGREGAÇÃO DOS DADOS .....	21
4.5	EXERCÍCIOS .....	22
<b>5</b>	<b>FUNÇÕES PARA A ANÁLISE INICIAL DOS DADOS.....</b>	<b>24</b>
5.1	FREQUÊNCIAS .....	24
5.2	ANÁLISE EXPLORATIVA DE DADOS.....	26
5.3	EXERCÍCIOS .....	27
<b>6</b>	<b>A APRESENTAÇÃO DOS DADOS.....</b>	<b>28</b>
6.1	APRESENTAR OS VALORES EM LISTA .....	28
6.2	O MÓDULO TABELAS.....	28
6.3	GENERAL TABLES .....	28
6.4	EXERCÍCIO:.....	30
6.5	GRÁFICOS.....	30
	<i>Criando gráficos .....</i>	<i>31</i>
	<i>Gráficos do menu dos gráficos para criar um gráfico.....</i>	<i>31</i>
	<i>Os gráficos que resultam das funções .....</i>	<i>31</i>
	<i>Gráficos do menu interactivo dos gráficos .....</i>	<i>32</i>
	<i>Usando o editor de gráficos.....</i>	<i>32</i>
	<i>O editor do gráfico para editar um objecto do gráfico.....</i>	<i>32</i>
6.6	EXERCÍCIO:.....	32
<b>7</b>	<b>A LINHA DE POBREZA .....</b>	<b>34</b>
7.1	O ÍNDICE DE INCIDÊNCIA DA POBREZA (POVERTY HEADCOUNT INDEX) .....	35

---

7.2	O ÍNDICE DE PROFUNDIDADE DA POBREZA (POVERTY GAP INDEX) .....	35
7.3	O ÍNDICE DE PROFUNDIDADE DA POBREZA AO QUADRADO (SQUARED POVERTY GAP INDEX) .....	35
7.4	EXERCÍCIO:.....	36
<b>8</b>	<b>ANÁLISE DA REGRESSÃO .....</b>	<b>37</b>
8.1	AS EQUAÇÕES DE REGRESSÃO.....	37
8.2	UM EXERCÍCIO:DADOS DO IAF DE SOFALA .....	37
8.3	AS CARACTERÍSTICAS DO MODELO DA REGRESSÃO .....	38
	<i>O exemplo de galinhas .....</i>	<i>39</i>
	<i>Um gráfico de Regressão.....</i>	<i>40</i>
8.4	MODELOS COM DOIS OU MAIS PREDICTORES .....	41
8.5	SUMÁRIO DO MODELO. ....	42
8.6	EXERCÍCIO:.....	43
<b>9</b>	<b>BIBLIOGRAFIA .....</b>	<b>44</b>
9.1	SPSS .....	44
9.2	ESTATÍSTICA, ANÁLISE DE POBREZA .....	44
9.3	WWW .....	44

## **1 Introdução**

Nas últimas décadas tem-se assistido a um grande desenvolvimento no domínio da utilização dos computadores e, no que aqui interessa, no desenvolvimento do software para tratamento e análise estatística, o que veio facilitar significativamente a aplicação da teoria estatística.

Este curso tem, assim, por objectivo familiarizar os participantes na utilização do programa SPSS, o software de análise estatística mais difundido do mercado. Uma vez que os participantes deste curso se encontram familiarizados com a utilização do EXCEL, serão apresentadas as diferenças entre este produto e o SPSS no tratamento e análise estatística, com a finalidade de fornecer os princípios básicos de utilização deste pacote de software, e, deste modo, permitir uma melhor compreensão sobre as vantagens e os benefícios da sua utilização em relação a um software mais geral, como é o caso do EXCEL.

Devido à circunstância deste curso focar alguns dos métodos estatísticos para a análise da informação pretende-se um duplo efeito: para alguns participantes será uma oportunidade para relembrarem a teoria já dada enquanto para outros será a oportunidade de se familiarizarem com alguns métodos estatísticos elementares.

Neste sentido, o curso foi desenhado para um leque abrangente de técnicos, com ou sem experiência no SPSS, exigindo-se, no entanto, uma formação básica sobre estatística.

A versão do SPSS a utilizar neste curso, versão 10, foi desenvolvida para o ambiente WINDOWS e para a utilização em rede, permitindo a troca de informação com as outras aplicações do WINDOWS, facilitando, deste modo, o processo de aprendizagem e de ajuda tanto do utilizador comum do WINDOWS (CUA = common user access) como dos utilizadores mais experientes.

A utilização deste manual, conjuntamente com o acesso ao tutorial “on-line”, será suficiente para atingir os objectivos do curso, permitindo aos participantes acompanharem o desenrolar da matéria e a resolução do conjunto de exercícios incluídos.

Por se tratar de um curso sobre a estatística da pobreza a informação estatística a utilizar no SPSS será a respeitante a esta área estatística.

## 1.1 Convenções de estilos do texto

Neste manual serão seguidas determinadas convenções, de modo a facilitar a compreensão do texto, como seja:

Alt]	Você é instruído para digitar a chave indicada
[Alt] + [Shift]	Você é instruído para digitar as chaves indicadas (juntas)
[End],[_]	Você é instruído digitar as chaves indicadas (uma após a outra)
[Files/Open/Data]	Escolhas do menu

## 1.2 Como escolher as opções dos menus

Você pode usar o rato ou o teclado para a escolha do menu

Há dois métodos de escolher itens de um menu:

Com um rato: Clique sobre ao nome do menu que Você quer. O nome do menu é destacado e a lista de itens do submenu estende-se para baixo. Você pode então ver os itens e clicar sobre o item particular que você quer. (ou estale sobre o nome de menu, prendendo seu dedo na tecla de rato. Sem liberar a tecla, arraste através dos artigos de menu em ordem. Tornam-se destacados na sucessão enquanto você arrasta através da lista. Quando o item que você quer é destacado, libere a tecla de rato.)

Para cancelar: clique fora do menu e da barra de menu.

Com o teclado: Pressione a chave [ Alt ] e a letra sublinhada (frequentemente a primeira) no nome de menu. Quando o menu deixa cair para baixo, pressione a letra sublinhada no nome do comando.

Para cancelar: imprensa [ Esc ]

## 2 Uma visão geral sobre o SPSS

✓ É objectivo deste capítulo fornecer-lhe uma visão geral do programa. Você deve aprender quais são os elementos principais de SPSS, os tipos de janelas e arquivos utilizados pelo SPSS e como iniciar e terminar o SPSS

- clique duas vezes no ícone do SPSS para iniciar o SPSS
- Escolhe [File/Exit] para terminar o SPSS

Os dois tipos mais importantes de janelas no SPSS são os seguintes:

**Editor dos dados:** Um arquivo rectangular, tipo folha de cálculo (spreadsheet), que contém os dados para serem trabalhados e analisados. Pode editar-se os dados da janela, adicionando ou suprimindo variáveis e alterando os seus atributos, excepto quando o processador de SPSS está modificando os dados. Utilize os comandos de menu:

[New] e [Open] para abrir um arquivo de dados existente no Editor de dados;  
[Save] ou [Save As] para salvar o conteúdo do **Editor de dados** quando o **Editor dos dados** está activo.

O editor dos dados não pode ser fechado, embora possa ser minimizar.

*A extensão do arquivo usado por SPSS para este tipo de arquivo é \*.sav.*

**Viewer:** Todos os resultados, tabelas e cartas estatísticos estão incluídos no **Viewer**, podendo o seu conteúdo ser editado e salvo para utilização futura. Uma janela do **Viewer** abre-se automaticamente a primeira vez que se executa uma função de SPSS que gere um resultado. Pode-se utilizar o visor para:

- Navegar nos resultados.
- Mostrar ou esconder as tabelas e as cartas seleccionadas.
- Alterar a ordem da apresentação dos resultados seleccionados.
- Alternar entre os resultados apresentados no ecrã e as outras aplicações.

O **Viewer** encontra-se dividido em duas janelas:

- A janela da esquerda do Viewer contém uma vista do esboço dos resultados, a exemplo do Explorador do Windows.
- A janela da direita contém tabelas de dados, apresentações gráficas e texto.

Pode utilizar-se as barras de scroll para visualizar os resultados, ou pode seleccionar-se um título para ir directamente à tabela ou ao gráfico correspondente.

SPSS e o sistema das janelas **alvo**

Se você tiver mais de uma janela aberta do **Viewer**, a saída está destinada à janelas **alvo** do **Viewer**.

As janelas **alvo** são indicadas por um ponto da exclamação (!) na barra de ferramentas.

A janela **alvo** não deve ser confundida com a janela activa, que é a janela actualmente seleccionada.

Através do **Viewer** podem ser chamados vários outros editores (cliquando no objeto do **Viewer** e utilizando a escolha do menu respectivo)

*A extensão do arquivo usado pelo SPSS para este tipo de arquivo é \*.spo.*

**Editor de tabelas de pivot (Pivot Table Editor):** A saída indicada em tabelas do pivot pode ser modificada de várias maneiras com o editor de tabelas de pivot. Pode editar-se o texto, trocar dados nas linhas e nas colunas, alterar a cor, criar tabelas multidimensionais e, selectivamente, esconder e mostrar resultados.

**Editor de gráficos (Chart Editor):** Podem modificar-se os gráficos de alta-resolução – mudar cores e as características dos objectos gráficos.

**Editor de saída de texto (Text Output Editor):** As características do texto incluído nas tabelas podem ser modificadas com o editor de saída de texto (tipo, estilo, cor, tamanho).

**Editor da sintaxe:** Podem colar-se as escolhas da caixa de diálogo numa janela de sintaxe, surgindo estas selecções em formato de sintaxe de comando, correspondendo à linguagem de programação do SPSS. Pode editar-se a sintaxe de comando para incluir características especiais de caixas de diálogo directas não disponíveis no SPSS. Estes comandos podem ser guardados num arquivo para ser utilizado em sessões seguintes do SPSS.

Se estiver mais do que uma janela do Viewer aberta, o resultado está apresentado na janela alvo do Viewer. Se você tiver mais de uma janela aberta do editor da sintaxe, a sintaxe de comando está colada na janela alvo do editor da sintaxe. As janelas alvo são indicadas por um ponto da exclamação (!) na barra ferramentas.

*A extensão do arquivo usado por SPSS para este tipo de arquivo é \*.sps*

## 2.1 Como editar os dados

O editor de dados reflecte um arquivo básico de dados, em que cada linha representa um único caso (observação). Por exemplo cada pessoa entrevistada é um caso.

Cada variável contém informação sobre os casos: por exemplo uma variável contém informação sobre o sexo do indivíduo e a sua relação com o chefe da família.

No editor dos dados pode acrescentar-se e editar os dados mas não se podem executar cálculos ou incluir fórmulas.

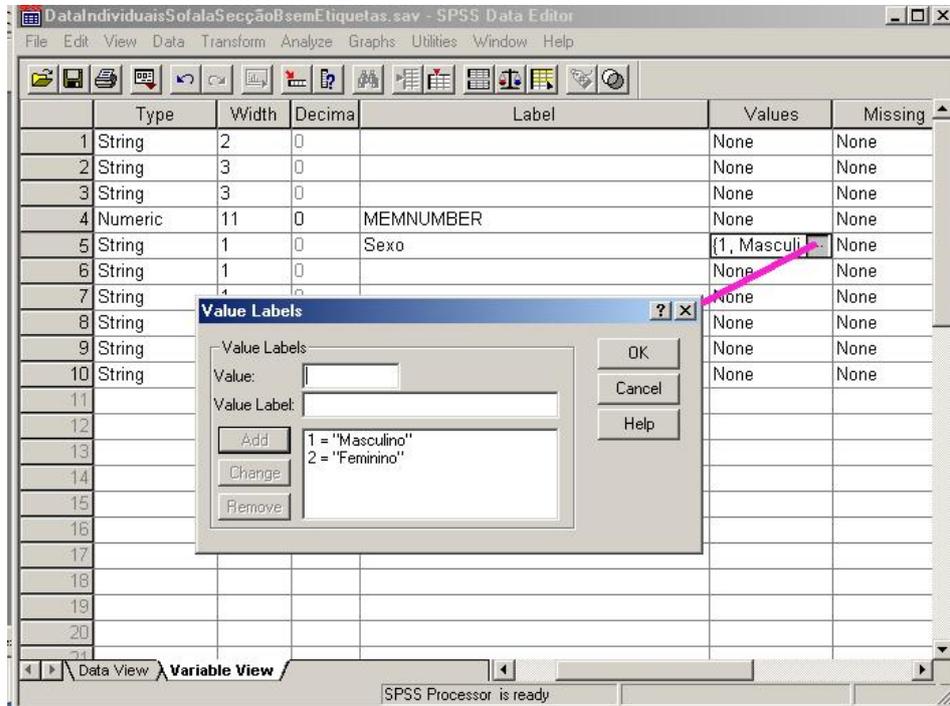
1 : region 07

	regio	a1	a2	memnumbe	b1	b2	b3	b4	b5	b6	var	var	v
1	07	469	001	1 1	1	1	1	51	4	.			
2	07	469	001	2 2	1	2	47	4	.				
3	07	469	001	3 2	1	3	30	6	.				
4	07	469	001	4 1	1	3	20	1	.				
5	07	469	001	5 2	1	3	18	1	.				
6	07	469	001	6 2	1	3	15	1	.				
7	07	469	001	7 1	1	3	10	.	1				
8	07	469	001	8 1	1	3	08	.	1				
9	07	469	001	9 2	1	3	11	.	1				
10	07	469	001	10 2	1	3	07	.	1				
11	07	469	003	1 1	1	1	58	4	.				
12	07	469	003	2 2	1	2	46	4	.				
13	07	469	003	3 1	1	3	16	1	.				
14	07	469	006	1 1	1	1	31	4	.				
15	07	469	006	2 2	1	2	19	4	.				
16	07	469	006	3 2	1	3	04	.	1				
17	07	469	008	1 1	1	1	31	6	.				
18	07	469	010	1 1	1	1	37	4	.				
19	07	469	010	2 2	1	2	22	4	.				

Data View Variable View / SPSS Processor is ready

## 2.2 Etiquetas das variáveis e valores

Para além de definir o tipo de dados, podem também ser definidas etiquetas para as variáveis e para os valores, para serem utilizadas em relatórios estatísticos e em gráficos. Por exemplo, atribuir a etiqueta de variável “Sexo” a uma variável e as etiquetas ' Masculino ' e ' Feminino



' aos valores numéricos 1 e 2.

## 2.3 Exercícios.

Observação geral. Os dados do IAF 2002/03 são propriedade do INE. Estes dados são confidenciais (não se conhece os dados pessoais, como endereço e localidade e os pormenores sobre as famílias e os indivíduos inquiridos). O uso indevido destes dados fora do INE consiste numa infracção da Lei da Estatística (7/96) e seria punido como tal.

- 1.) Use o arquivo de dados **DadosIndividuaisSofalaSecçãoBsemEtiquetas.xls**.  
Compare o arquivo com o questionário do IAF 2003
  - a. Quantas perguntas contem
  - b. Quantos indivíduos foram entrevistados
- 2.) Use o menu [File/Open/Data] para abrir um novo arquivo de dados. Salve o arquivo como **DadosIndividuaisSofalaSecçãoBsemEtiquetas.sav**
- 3.) Use o aceso “Variable View” do **Data Editor** para atribuir as seguintes etiquetas.:

	Variable Label		Value	Value Label
REGION	* No label *			
A1	Area Enumeração			
A2	Agregado Familiar			
MEMNUMBE	Número do Membro			
B1	Sexo		1	Masculino
			2	Feminino
B2	Nacionalidade		1	Moçambicana
			2	Outra
B3	Relação com chefe do AF		1	Chefe
			2	Cônjuge
			3	Filho/a
			4	Pai/mãe
			5	Outros parentes
			6	Sem parentesco
B4	Idade em Anos			
B5	Estado Civil		1	Solteiro/a
			2	Casado/a
			3	Casado/a em reg.poligamia
			4	Unido maritalmente
			5	Divorciado/a ou Separado/a
			6	Viúvo/a
B6	Mãe viva?		1	Sim
			2	Não

## 2.4 A primeira análise com o SPSS

### As principais etapas do SPSS

Existem poucas etapas básicas simples para a análise de dados com o SPSS:

Registrar os seus dados no **Editor de dados**.

Seleccionar um procedimento dos menus **criar tabelas**, **calcular estatísticas** ou **criar charts**.

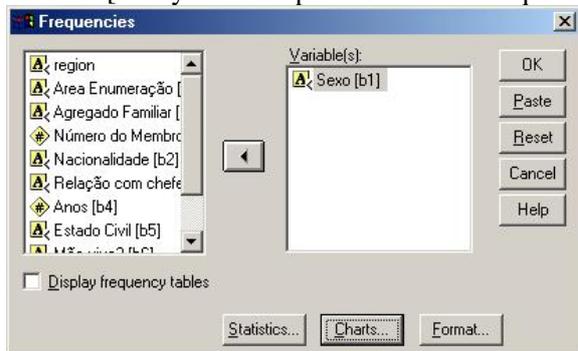
Seleccionar as variáveis que deseja utilizar nas análises.

Seleccionar os procedimentos respectivos e verificar os resultados.

Depois disso, comece com a primeira análise estatística simples:

Quantas mulheres e homens existem no grupo das pessoas do arquivo. Analise a frequência do sexo:

Escolha [Analyze/Descriptive Statistics/Frequencies..].



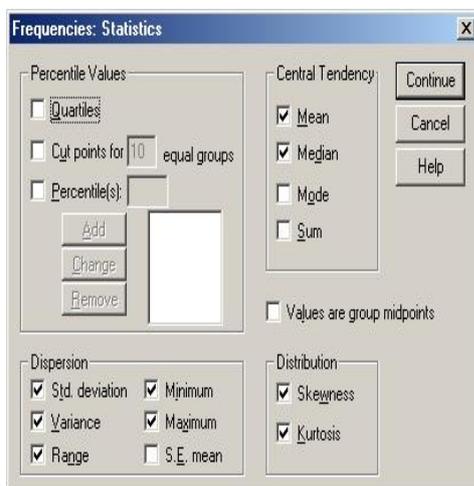
### Escolhendo Variáveis para frequências

Depois escolha as variáveis usando a caixa de diálogo e escolha a Opção [Statistics];

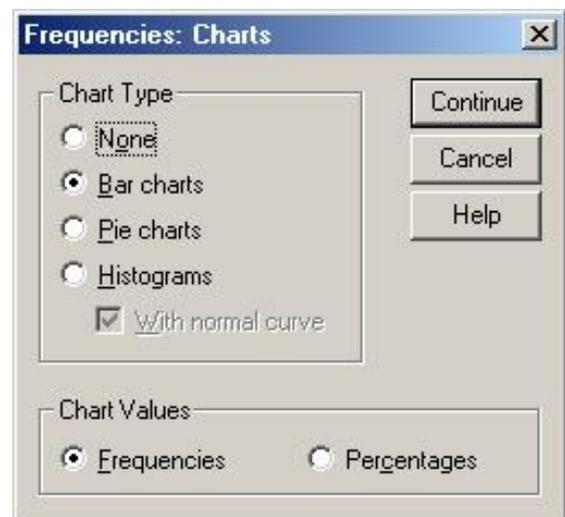
Escolha as estatísticas de tendência central e de distribuição e termine com [Continue];

Escolha a caixa de diálogo e a opção [Charts];

Com a escolha de *Bar Charts* e [Continue] na primeira janela e a escolha de [Paste] na segunda janela obtém-se os seguintes resultados:



### Escolhendo Estatísticas para frequências



### Escolhendo Charts para frequências

um arquivo de sintaxe, que você pode conservar (syn01.sps) e um outro resultado: tabelas e gráficos.

O comando para analisar a frequência será executado escolhendo [Run/All] ou pressionar a tecla [Run Current] na janela da sintaxe, se o cursor for posicionado em algum lugar na linha de comando do SPSS e todo o arquivo da sintaxe foi seleccionado com o mouse.



Embora não sendo necessário executar um arquivo de sintaxe em vez de utilizar os comandos do menu, a sua utilização será de grande utilidade no futuro, pelo que se recomenda a aprendizagem sobre o seu funcionamento.

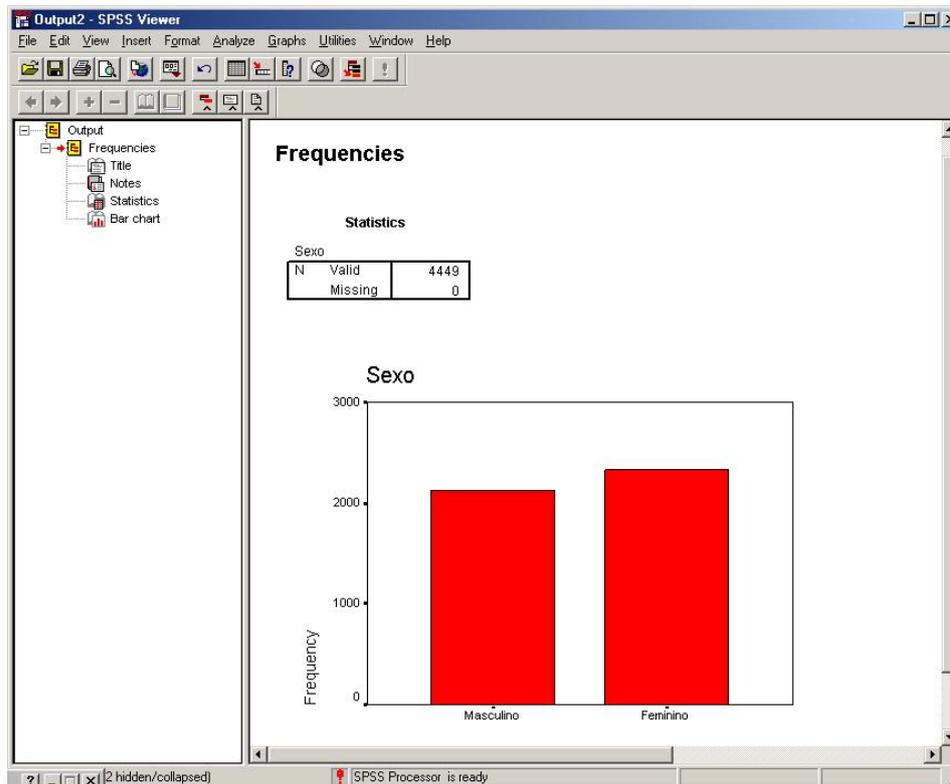
O conteúdo do arquivo de sintaxe:

## FREQUENCIES

```
VARIABLES=b1 /FORMAT=NOTABLE  
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN SKEWNESS  
SESKEW KURTOSIS SEKURT  
/BARCHART FREQ  
/ORDER= ANALYSIS .
```

O conteúdo do **Viewer**

Com a escolha duma variável alfanumérica e nominal Você não consegue calcular estatísticas de dispersão e distribuição nem de tendência central.



## 2.5 Os tipos de variáveis para a análise estatística

Existem diferentes níveis de medida de variáveis como seja a escala (dados numéricos numa escala de intervalo ou de relação), o ordinal ou o nominal. Os dados nominais e ordinais podem ser do tipo alfanuméricos ou numéricos. A especificação da medida é relevante para o tipo de análise a efectuar para cada variável:

Você pode diferenciar entre três níveis de medida:

### Scale (Escala).

Os valores dos dados são valores numéricos em uma escala de intervalo ou de relação (por exemplo, idade, renda). As variáveis da escala devem ser numéricas.

### Ordinal.

Os valores dos dados representam categorias com alguma ordem intrínseca (por exemplo, baixo, médio, elevado; concordo fortemente, concordo, discordo, discordo fortemente). As variáveis ordinais podem ser do tipo alfanumérico ou os valores numéricos que representam as categorias distintas (por exemplo, 1=low, 2=medium, 3=high). No geral, é mais seguro utilizar códigos numéricos para representar dados ordinais.

### Nominal.

Os valores dos dados representam categorias sem qualquer ordem intrínseca (por exemplo, categoria de trabalho ou divisão de companhia). As variáveis nominais podem ser do tipo alfanumérico ou os valores numéricos que representam as categorias distintas (por exemplo, 1=Male, 2=Female).

Por padrão:

As variáveis alfanuméricas são classificadas como do tipo nominal.

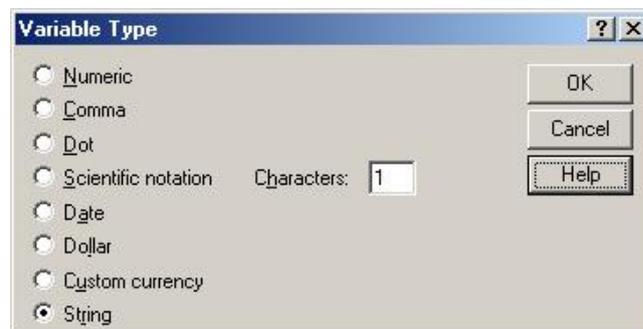
As variáveis numéricas com etiquetas de valor são classificadas como do tipo ordinal.

As variáveis numéricas sem etiquetas de valor são classificadas como do tipo ordinal.

As variáveis numéricas com etiquetas do valor e mais do que uns valores especificados são classificadas como do tipo escala.

## 2.6 Os tipos de variáveis do SPSS

Para definir o tipo de variáveis no SPSS, clique duas vezes no nome da variável no topo do nome da coluna ou na vista "Variabel View" na coluna "Type". Selecciona o tipo da variável da lista.



## 2.7 Exercícios.

Utilize o SPSS para analisar os dados do arquivo

**DadosIndividuaisSofalaSecçãoBcomEtiquetas.sav:**

1. Imprima a análise da frequência dos homens e das mulheres na amostra.
2. Modifique os tipos de variáveis, se necessário ou apropriado (Modifique B4 = numérico)
3. Analise a variável B4, imprima as estatísticas de tendência central, de dispersão e distribuição. Evite imprimir cada vez a tabela com todos os valores (Clique na janela principal desta análise e retire a seleção da opção “Display Frequency Tables”)
4. Imprima um histograma desta variável e comente as diferenças da distribuição desta variável em relação a uma distribuição normal

## 3 A transformação das variáveis

Antes de poder analisar os dados é, geralmente, necessário efectuar algum trabalho preparatório, como seja a transformação das variáveis e dos valores respectivos. Deve-se, por isso, conhecer as potencialidades das transformações e as razões pelas quais são utilizadas, bem como saber executá-las.

### 3.1 Definindo as variáveis

Para definir variáveis novas ou para mudar o formato de variáveis, dê um duplo clique numa coluna do editor de dados. Você pode também definir variáveis directamente numa coluna vazia especificada. Os nomes das variáveis serão fornecidos então pelo SPSS por defeito: a primeira variável nomeada será VAR00001, o segundo VAR00002 e assim sucessivamente.

### 3.2 Nomes das variáveis no SPSS

Os nomes das variáveis no SPSS podem conter até 8 caracteres. O primeiro caracter deve ser uma letra ou @ e cada caracter seguinte deve ser uma letra, um número, um ponto, um underscore (“\_”), ou um dos símbolos \$ # @, desde que o ultimo caracter não um ponto. As letras maiúsculas e minúsculas são equivalentes em nomes variáveis. Não pode ser utilizado como nome duma variável as seguintes palavras: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH.

Tipos de Variáveis

Conforme já referido são válidos os seguintes tipos variáveis :

- Numeric
- Comma
- Dot
- Scientific Notation
- Date
- Dollar
- Custom Currency
- String

Os dois tipos usados mais frequentemente são descritos a seguir:

**A variável numérica** define a totalidade dos valores numéricos cujos valores sejam indicados no formato numérico padrão, usando-se o delimitador decimal especificado no

painel de controle internacional. O editor de dados aceita valores numéricos no formato padrão ou na notação científica, contanto que o “E” e o sinal do expoente é incorporado (1E+1 mas não 1E1).

**A variável string** corresponde a uma variável alfanumérica, cujos valores possam conter todos os caracteres até ao comprimento definido.

As letras maiúsculas e minúsculas são consideradas de forma distinta. Alguns dos comandos do SPSS não podem usar variáveis alfanuméricas, desde que os cálculos com variáveis alfanuméricas não estejam definidos. As variáveis alfanuméricas definidas com mais de 8 caracteres de comprimento ("variáveis alfanuméricas longas") são restringidas no uso da análise estatística.

Existem funções para converter variáveis numéricas em variáveis alfanuméricas e vice-versa. Mesmo que você aprenda as funções do SPSS mais tarde, estas duas funções importantes são a seguir explicadas:

*STRING(numexpr,format) String* Devolve o valor alfanumérico deste resultado quando o *numexpr* é convertido a uma variável alfanumérica de acordo com o formato. O segundo argumento da função deve ser um formato para definir um valor numérico.

*NUMBER(strexp,format) Numeric* Devolve o valor da variável alfanumérica *strexp* como um número. O segundo argumento (*format*) é o formato numérico usado para converter o *strexp*, por exemplo o formato “f8” converte em valores numéricos de 8 caracteres sem decimais

### 3.3 Valores MISSING

Há dois tipos de valores em falta (missing) no SPSS: valores “user-missing” e valores “system-missing”.

A todas as células numéricas em branco no rectângulo dos dados é atribuído o valor system-missing, que é representado por um ponto.

Você pode também atribuir os valores que identificam a informação que falta para razões específicas e instruir então o SPSS para marcar estes valores como estando em falta.

Os valores System-missing e user-missing são suportados especialmente por procedimentos estatísticos do SPSS e por comandos de transformação de dados.

De agora em diante, todos os comandos do SPSS serão executados como comandos de sintaxe que tenham sido editados e executados. Você executará estas tarefas, geralmente as escolhas do menu, e colará as suas escolhas no editor da sintaxe. Você pode continuar a proceder deste modo em exercícios adicionais. Enquanto você prossegue neste curso, a

sintaxe do SPSS deve tornar-se mais familiar e ' a escrita ' de um programa no SPSS tornar-se-á a coisa a mais natural. Pelo menos, são estas as esperanças do autor e do formador deste curso.

### 3.4 Cálculo com variáveis

Menu: [Transform/Compute ]

Comando: *COMPUTE, IF*

Para calcular valores para todos os casos, digite o nome de uma variável do destino. Pode ser uma variável existente ou uma variável nova a ser adicionada a seu arquivo de dados de trabalho. Construa uma expressão para os valores que devem ser atribuídos a essa variável. Calcular selectivamente valores para os casos que satisfazem alguma condição lógica, seleccionar a tecla **[If]** e introduzir as condições. Se a variável do destino for uma variável alfanumérica você deve seleccionar **[Type] & [Label]** e indicar seu comprimento.

Se estiver calculando uma variável nova, você pode seleccionar **[Type] & [Label]** para definir-lhe uma etiqueta de variável.

### 3.5 Transpondo variáveis

Menu:[Data/Transpose ]:

Comando: *FLIP*

Esta função transpõe as linhas e as colunas no arquivo de dados de modo a que os casos se transformem em variáveis e as variáveis se transformem em casos. Selecciono ao menos uma variável. Estas variáveis são colunas no arquivo de dados de trabalhando, e serão transpostas em casos (linhas) no novo arquivo de dados. As variáveis que você não selecciona serão perdidas. O SPSS relata os nomes das variáveis no arquivo transposto e cria uma variável *CASE\_LBL* que contem os nomes das variáveis antigas.

Se o arquivo de dados de trabalho contiver um ID ou uma variável com valores únicos, você pode movê-la na caixa [Name Variable] e os seus valores serão convertidos em nomes de variáveis para o arquivo de dados transpostos.

### 3.6 Ordenar Casos

Menu:[Data/Sort Cases]:

Comando: *SORT CASES*

Selecione uma variável e mova-a pela caixa para sortear. Para ordenar por ordem descendente, selecione a variável para sortear e selecione na caixa “Sort Order” o tipo Ascending/descending (ascendente/descendente). Você pode selecionar variáveis adicionais incluídas na lista. Cada variável adicional especifica a ordem dos casos dentro do grupo que têm valores idênticos para todas as variáveis precedentes na lista. A ordem ascendente é representada por (A) na lista de variáveis incluídas na lista de ordem. As variáveis alfanuméricas são classificadas alfabeticamente. A ordem descendente é representada por (D) na lista de variáveis incluídas na lista de ordem.

Nas caixas de diálogo do menu [Data/Split File] pode também classificar o seu arquivo de dados conforme a ordem dos dados selecionados ou classificar os dados classificados em grupos de dados.

### 3.7 Selecionar casos

Menu: [Data/Select Cases]

Comando: *SELECT IF; SAMPLE; FILTER; USE*

Escolha uma das alternativas no grupo a selecionar.

*All cases* indica que você deseja processar todas os casos do seu arquivo de dados.

*If condition is satisfied* indica que você deseja processar somente as caixas que satisfazem algumas condições lógicas. Se você especificar uma condição, esta aparece ao lado do botão [If]. Esta alternativa ajusta a variável filtro filter\_\$, discutido a seguir.

*Random sample of cases* indica que você deseja selecionar aleatoriamente casos para processar. Se você incorporou especificações da amostra estas aparecem ao lado da tecla [**Sample**].

*Based on time or case range* indica que você deseja processar somente aqueles casos que caem dentro de uma escala especificada. Clique sobre [**Range**] para definir os dados a ser incluídos no processamento.

*Use filter variable* indica que você quer usar os valores de uma variável numérica existente para filtrar os casos. Selecione a variável na lista da esquerda. Os casos para que a variável do filtro tem o valor 0 são excluídos da análise.

Escolha uma das alternativas do grupo *Unselected cases are*:

*Filtered* exclui casos do processamento, mas mantêm-nos no arquivo de dados. Você pode voltar a esta caixa de diálogo e reverter esta alternativa para restaurar os casos.

*Deleted* suprime os casos do arquivo de dados. Para recuperá-los, você deve abrir o arquivo de dados originais outra vez (supondo que não substituíram o arquivo original com a versão reduzida).

Se você executar este comando com as alternativas **[If]** ou **[Sample]** e indique que os casos estejam filtrados, uma variável nomeada `filter_$` é calculada como 1 (seleccionado) ou 0 (não seleccionado). Você pode editar os valores de `filter_$` no editor dos dados; os casos com um valor de 0 para uma variável do filtro estão excluídos da análise, quando os casos com todos os outros valores numéricos válidos forem incluídos. Entretanto, esteja ciente que esta variável prende o status filtrando actual. A opção editar o seu valor determina que casos são filtrados.

## 3.8 Exercícios

Use SPSS para analisar os dados do arquivo  
DadosIndividuaisSofalaSecçãoBcomEtiquetas.sav:

Use SPSS para transformar os dados

1. Defina valores em falta para pelo menos uma variável (os candidatos são *Idade em Anos e Estado Civil*).
2. Defina uma variável nova `gr_age`
3. Compute valores para a variável `gr_age` para cada grupo com um intervalo de 10 anos (`gr_age = 1` para uma idade entre 0 e 10 anos, `gr_age = 2` para uma idade entre 10 e 20 anos etc.). Use o comando [Transform/Compute ]
4. Depois da primeira operação use a tecla [Paste] e use o arquivo de sintaxe para elaborar o resto das transformações
5. Qual é o grupo de idade com maior número de pessoas e qual é o respectivo numero?
6. Ordenar o arquivo conforme a idade
7. Ordenar o arquivo conforme a idade e o estado civil. Confira se têm crianças casadas, viúvo/viúvas etc.
8. Selecciona somente os chefes de família com mais que 50 anos de idade. Quantos são?
9. Selecciona somente os chefes de família do sexo feminino. Quantos são

## 4 Considerações preliminares sobre índices da pobreza e a análise

### 4.1 A transformação de arquivos de dados

Para iniciar a análise de dados os arquivos de dados têm que ser frequentemente preparados e transformados. Os arquivos podem ser importados de diferentes fontes, por exemplo as bases de dados como o ORACLE, EXCEL ou xBASE (dBASE, Clipper etc.) – às vezes têm que ser adicionados diversos arquivos outras vezes têm que ser combinados.

Neste capítulo você deve aprender como esta transformação é feita bem como as condições em que ela é necessária.

Combinando Arquivos: Considerações gerais

O menu [Data/Merge Files] permite combinar dois arquivos de dados de trabalho num único arquivo de dados, conservando o formato do SPSS (O arquivo actual chama-se "o arquivo interno", o arquivo no disco, que se junta é chamado "o arquivo externo"). Dois tipos de fusão são suportados:

1. [Add cases] adiciona todos os casos do arquivo externo no fim do arquivo de dados de trabalho. O arquivo novo inclui as variáveis partilhadas pelos dois arquivos, como padrão. Você pode optar por incluir as variáveis que estavam apenas num arquivo; os casos que vêm do outro arquivo terão valores system-missing ou em branco para tais variáveis. Você pode também excluir variáveis.
2. [Add variables] adiciona algumas ou todas as variáveis do arquivo externo aos casos correspondentes no arquivo de dados de trabalho. Você pode definir casos correspondentes de uma das seguintes formas:
  - [Case order]
  - [Key variable(s)]
  - [Table lookup]

### 4.2 Combinação de arquivos: Adicione Casos

Menu: [Data/Merge Files/Add Cases]

Comando: *ADD FILES*

Numa caixa de diálogo inicial solicita-se que seja identificado o arquivo ao qual se deseja adicionar casos. A lista da direita, variáveis no arquivo de dados de trabalho novo, contém todas as variáveis compartilhadas por ambos os arquivos. Se esta lista for satisfatória, pressione [OK. ] A lista das variáveis da esquerda, *Unpaired Variables*, contém as variáveis que não podem ser combinadas nos arquivos, porque são de tipos diferentes ou porque possuem nomes diferentes. As variáveis marcadas com (\*) estão no arquivo de dados de trabalho (interno). As variáveis marcadas com (+) estão no arquivo externo.

## 4.3 Combinação de arquivos: Adicione Variáveis

Menu: [Data/Merge Files/Add Variables]

Comando: *MATCH FILES*

Numa caixa de diálogo inicial solicita-se que seja identificado o arquivo ao qual se deseja adicionar variáveis.

Especifique o tipo de combinação (MATCH). Se você quiser combinar os casos nos dois arquivos só pela ordem sequencial deixe não seleccione a opção *Match cases*. Se ambos os arquivos forem identificados por uma ou mais "variáveis chave" e você quer combinar os casos de acordo com seus valores então seleccione a opção *Match cases* nas variáveis chaves em arquivos classificadas, escolha uma das três alternativas para o tipo de acomodamento (MATCH), e mova a variável chave da lista de variáveis excluídas na lista inferior das variáveis chaves.

Especifique as variáveis para o novo arquivo de dados de trabalho: A caixa de lista da direita, variáveis no novo arquivo de dados de trabalho, contém inicialmente todas as variáveis do arquivo de dados de trabalho, marcado (\*), seguido das variáveis no arquivo externo com nomes diferentes, marcada (+). Pode adicionar ou remover variáveis a esta lista.

A caixa de lista da esquerda, *variáveis excluídas*, contém as variáveis que não serão incluídas normalmente no novo arquivo de dados de trabalho, porque têm o mesmo nome que as variáveis do arquivo de dados de trabalho.

## 4.4 A agregação dos dados

Menu: [Transform/Aggregate ]

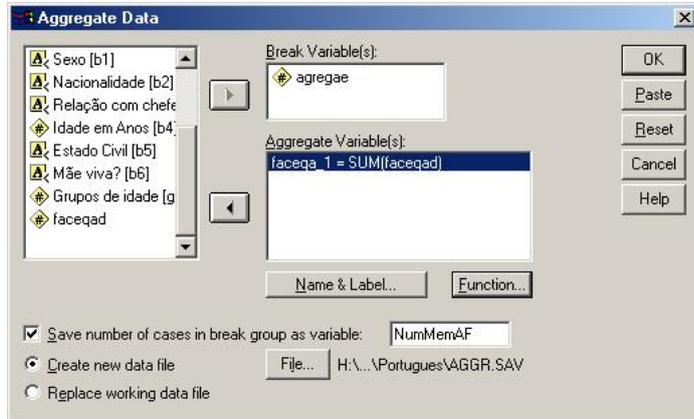
Comando: *AGGREGATE*

Os dados agregados combinam grupos dos casos num único caso sumário e criam um novo arquivo de dados de agregado. Os casos são agregados com base no valor de um ou de mais grupos de variáveis. O novo arquivo de dados contém um caso para cada grupo. Por exemplo, você poderia agregar dados dos Distritos para somar dados demográficos de cada distrito.

[Break Variable(s)]. Os casos são agrupados com base nos valores das *Break Variable(s)* (variáveis de quebra). Cada valor das variáveis de quebra define um grupo e gera um caso no novo arquivo agregado. Todas as variáveis de quebra são conservadas no arquivo com os nomes e informações existentes. A variável de quebra pode ter um formato numérico ou alfanumérico.

[Aggregate Variable(s).] são usadas para criar variáveis novas no arquivo agregado. Por defeito, os dados agregados criam nomes de variáveis agregadas novos usando os primeiros caracteres dos nomes das variáveis do arquivo original seguido por um underscore e por um

número sequencial de dois dígitos. O nome da variável agregado é seguido por uma etiqueta da variável opcional. As variáveis do arquivo original para funções agregadas devem ser numéricas.



Você pode substituir os nomes das variáveis do agregado de padrão por novos nomes de variáveis, fornecer etiquetas descritivas da variável e mudar as funções usadas para calcular os valores agregados dos dados. Você pode também criar uma variável que contenha o número dos casos para cada grupo de quebra.

## 4.5 Exercícios

Utilize o SPSS para analisar os dados no arquivo  
**DadosIndividuaisSofalaSecçãoBcomEtiquetas.sav:**

Use o SPSS para transformar os dados

1. Transforme as variáveis A1, A2, B1, B3, B4 e B5 em variáveis numéricas.
2. Use o comando [Transform/Compute ] para criar uma nova variável AgregAE (“Agregados Familiares na Área de Enumeração”) com o cálculo: `COMPUTE AgregAE = a1*1000+a2.`
3. Use o comando [Transform/Compute ] para criar uma nova variável FacEqAd (“Fator do Equivalente Adulto” com o cálculo:  
`IF (b4 >17) FacEqAd = 1.`  
`IF (b4 <= 17 & b4 > 10) FacEqAd = 0.7.`  
`IF (b4 <= 10 & b4 > 5) FacEqAd = 0.5.`  
`IF (b4 <= 5) FacEqAd = 0.2.`
4. Salve o arquivo como  
**DadosIndividuaisSofalaSecçãoBcomEtiquetasIndexed\_EqAd.sav**
5. Use o comando [Transform/Aggregate ] para agregar os dados do arquivo na variável AgregAE. Como *variável de quebra* inclua as funções seguintes para manter as outras variáveis:  
AGGREGATE

```
/OUTFILE='H:\Documents and'+  
'  
Settings\RDK\Training\CursoMS\MS_Kurs\SPSS_Poverty\Portugues\AGGR.SAV'  
/BREAK=agregae  
/memnum_1 = FIRST(memnumbe) /b1_1 = FIRST(b1) /b4_1 = FIRST(b4)  
/b5_1 =
```

6. Abre o arquivo novo AGGR.SAV e salve o arquivo como **DadosIndividuaisSofalaSecçãoBcomEtiquetasAggregated.sav**. Você tem agora o arquivo Individual agregado para ser combinado com os dados do AF
7. Abre o arquivo **DadosAggegadosFSofalaSecçãoF\_G.sav** Use o comando [Transform/Compute ] para criar uma nova variável *AgregAE* como no exercício 4.). As variáveis A1 e A2 tem que ser do tipo numérico.
8. Use o comando [Transform/Add Files/Match Variables] e adicione o arquivo **DadosIndividuaisSofalaSecçãoBcomEtiquetasAggregated.sav**. Escolhe *Match cases on key variables in sorted files (os arquivos tem que estar ordenados pela variável AgregAE)* e *Both file provide cases*. Escolhe a variável comum *AgregAE* como variável chave e integre os arquivos. Salve o arquivo como **DadosAgregadosFamiliaresComIndividuais.sav**. Este arquivo contém dados sobre as características dos AF e dos chefes dos AF para ser utilizado nos próximos capítulos

## 5 Funções para a análise inicial dos dados

Neste capítulo você aprenderá usar funções de SPSS. Haverá uma explanação muito pequena sobre a teoria estatística atrás das funções, conhecimento existente é suposto. Se você não for certo, o que o método faz, satisfizer consultar um livro de texto estatístico. O objectivo principal deste capítulo é explicar a manipulação das funções e a interpretação dos resultados. Você deve aprender ambos este neste capítulo.

### 5.1 Frequências

Menu: [Analyse, Descriptive Statistics /Frequencies...]

Comando:FREQUENCIES

Movimento um ou mais variável na lista de variável(eis).

Uma tabela da frequência será produzida para cada um variável nesta lista, mostrando o número dos argumentos para cada valor.

Selecione [**Statistics**] para pedir estatísticas explicativas.

Selecione [**Charts**] para pedir gráficos de barra ou histogramas.

Selecione [**Format**] para controlar a ordem dos valores na tabela, seu arranjo na página, a exposição de um índice, e a exposição de etiquetas do valor.

Se você deseccione [Display frequency table], nenhuma tabela da frequência é indicada, mas todos as estatísticas ou gráficos que você pedir aparecerão. Tabela da frequência para arquivos de dados grande não é útil porque imprime todos os casos.

#### [Statistics]

<i>Quartiles</i>	divide os casos em quatro grupos do mesmo tamanho
<i>Mean</i>	Signifique a média aritmética; a soma dividida pelo número dos casos.
<i>Cut points for equal groups</i>	Corte pontos para os valores que dividem os casos em algum número do valor igual
<i>Median</i>	número médio, o 50th percentile.
<i>Percentiles</i>	Valores dos percentiles, os casos (observações) caem acima e abaixo destas porcentagens determinadas.
<i>Mode</i>	valor o mais frequentemente ocorrendo

---

<i>Sum</i>	soma ou o total dos valores, através de todos os casos com valores non-missing.
<i>Std. deviation</i>	Medida do desvio padrão da dispersão em torno do meio, igual à raiz quadrada da variância. O desvio padrão é medido nas mesmas unidades que a variável própria.
<i>Minimum</i>	Mínimo o valor menor de uma variável numérica
<i>Skewness</i>	Simetria de uma distribuição. O skewness positivo indica que os valores mais extremos são situados ao lado positivo da media; o skewness negativo indica que os valores mais extremos são situados ao lado negativo da media.
<i>Variance</i>	Variância A da dispersão em torno da media, igual à soma de desvios quadrados da media dividido por o número dos casos menos um. A variação é medida na unidade quadrada daquela da variável própria.
<i>Maximum</i>	Máximo o valor maior de uma variável numérica
<i>Kurtosis</i>	kurtosis (curtose) e a expressão que uma distribuição é “tail-heavy” comparada a uma distribuição normal. O kurtosis (curtose) positivo indica mais casos nas caudas extremas do que em uma distribuição normal com a mesma variação e os dados se aglomeram mais na vizinhança da media.
<i>Range</i>	a diferença entre os valores do máximo e do mínimo numérico
<i>S.E. mean</i>	A media do erro padrão da variável sumariada para todos os casos na amostra

## [Charts..]

**Bar Chart:** Um gráfico de barras indica a contagem da frequência para cada valor como uma barra separada. Os gráficos de barras são apropriados para as variáveis medidas no nível nominal, ou para variáveis com poucos valores distintos. Os valores para que as contagens em zero não aparecem.

**Histogram:** Um histograma indica a contagem da frequência para escalas dos valores como barras separadas. Os histogramas são apropriados para as variáveis para que é significativo

agrupar valores adjacentes. Uma curva normal pode ser sobreposta em um histograma em julgar se a variável está distribuída normalmente.

## [Format..]

A tabela da frequência pode ser arranjada de acordo com os valores reais nos dados, ou de acordo com a contagem (frequência da ocorrência) daqueles valores, e em ordem ascendente ou descendente.

## 5.2 Análise explorativa de dados

Como função *Frequencies* entrega alguma informação estatística geral sobre a série de dados, a função *Explore* permite de observar estatísticas sobre todas as observações como também sobre grupos das observações. Incorpora algumas técnicas gráficas que são especialmente úteis para uma primeira análise da série de dados, a comparação entre grupos das respostas e do teste da normalidade ou de uma comparação das variações. Como as outras funções neste capítulo, esta é uma primeira aproximação para examinar ou explorar os dados.

Menu: [Analyse, Descriptive Statistics, Explore]

Comando: EXAMINE

Mova ao menos uma variável na *Dependent List*. As estatísticas e os gráficos deste comando descreverão as distribuições das variáveis nesta lista. Você pode mover uma ou mais variáveis no *Factor List*. Os valores de variáveis do factor definem grupos dos casos, que serão descritos separadamente.

Você pode mover uma variável na caixa *Label Cases by*.

Esta opção é usada para identificar casos incomuns em alguns dos resultados.

[Plots] permite que você controle a saída dos gráficos.

[Statistics] permite que você peça estatísticas descritivas.

[Options] permite que você determine o tratamento de valores ausentes (missing).

Antes de começar a análise, deixe explicar momentaneamente as técnicas usadas: Os *Boxplots* mostram o número médio, quartiles e distinguem valores isolados. As “caixas” de *Boxplot* contêm o 50% dos valores que caem entre os 25% e 75% percentiles, e dos “bigodes”, as linhas que o estendem da caixa os valores os mais elevados e os mais baixos, excluindo *outliers* (valores que se encontram a um distancia de mais que a variação do resto da amostra).

O teste de normalidade compara os dados (do tipo escala) com uma distribuição normal. O Teste Kolmogorov-Smirnov mostra o nível de confiança de rejeitar a hipótese  $H^0 = A$  distribuição e normal. O gráfico *Q-Q normal* traça em gráfico a distribuição dos quantiles da variável (em pontos) em comparação aos quantiles da distribuição normal (em linha)

Esta funciona apresenta uma primeira estatística bivariante: a relação estatística entre duas variáveis ao contrario a estatísticas univariates: estatísticas analisando o comportamento duma variável somente.

## 5.3 Exercícios

Use SPSS para analisar os dados no arquivo

**DadosAgregadosFamiliaresComIndividuaisCompleto.sav**

1. Fazem uma análise de a frequência do grau do ensino mais alto obtido (C3)
2. Fazem uma análise explorativa do grau do ensino mais alto obtido (C3) entre chefes masculinos e femininos. Qual será um problema desta análise?
3. Fazem uma análise explorativa do tamanho da família (nummemaf) de o sexo do chefe (B1)
4. Conte a frequência de empregados do sector público, do sector privado e da conta própria.(na variável E8)
5. Crie uma variável nova do grupo de idade (4 valores somente). Analise pela análise explorativa de dados as relações entre estes grupos de idade e sexo(B1) e as variáveis: *grau do ensino mais alto obtido* (C3) e *Alguma vez frequentou a escola* (C2)
6. Analise a normalidade das variáveis acima mencionados com uma análise explorativa e de frequência, observando os dados extremos e a distribuição para cada grupo (grupos de idade e sexo(B1)) em comparação com a distribuição normal.
7. Analise as variáveis *Esteve doente ou ferido nas ultimas duas semanas* (D2) e *Quantos dias ficou sem trabalhar* (D3) das variáveis de grupos acima mencionadas.

## 6 A Apresentação dos dados

O objectivo da análise estatística é produzir resultados para serem comunicados aos utilizadores, que podem ser ou não ser estatísticos, e, por isso, devem ser apresentados de forma detalhada e atractiva, e tendo sempre em conta a audiência alvo.

Devem ser incluídas todas as explicações que permitam uma melhor compreensão dos resultados disponibilizados. Por outro lado, deve ser dada preferência à apresentação gráfica dos resultados ou em alternativa como complemento das formas de apresentação tradicionais.

Assim, neste capítulo abordam-se as possibilidades do SPSS para publicar dados estatísticos – listas, quadros e tabelas e representação gráfica –, as técnicas utilizadas na grande maioria das publicações estatísticas.

Você deveria aprender estes métodos e funções e você deveria saber quando e como usá-los.

### 6.1 Apresentar os valores em lista

Menu: [Analyse/Reports/Case Summaries]

Comando: LIST

Movimente uma ou mais variáveis na lista de variáveis. Os valores das variáveis seleccionadas serão indicados na lista apresentada.

Para listar um subconjunto dos casos que começam com o primeiro caso, selecione *Limit cases to ..* e incorpore então o número do último caso a ser listado.

Selecione o *Show case number* se desejar incluir o número de sequência de cada caso na lista.

### 6.2 O Módulo Tabelas

As tabelas do módulo suplementar são uma combinação de funções de apresentação em lista, de funções de relatório, de tabelas cruzadas e de estatísticas descritivas. Todas estas funções do SPSS permitem produzir as tabelas quase prontas para a publicação. Sem nenhuma dificuldade estas tabelas podem ser incorporadas em qualquer relatório usando as possibilidades comuns de WINDOWS para a troca de dados. Do elevado número de funções apenas vamos estudar o sub-modulo mais geral:

### 6.3 General Tables

As tabelas gerais (*General Tables*) permitem um maior controle sobre a aparência e o formato das tabelas, em troca de umas especificações um pouco mais complexas. Tipos de estatísticas diferentes podem ser indicadas para variáveis diferentes. Os comandos *General Tables* permitem que você desagregue algumas variáveis e agregue outras. Suporta-se também as

tabelas que envolvem jogos de dados múltiplo-resposta e múltiplo-dicotomia (Variáveis elementares dicotomias que têm duas categorias). Após ter seleccionado esta opção para criar um jogo múltiplo da dicotomia, incorpore um valor do inteiro para o valor contado. Cada uma variável tendo ao menos uma ocorrência do valor contado transforma-se uma categoria do jogo múltiplo da dicotomia..)

Menu: [Analyze/ Custom Tables/General Tables]

Comando: TABELAS

Especifique, em primeiro lugar, onde as etiquetas das estatísticas devem aparecer (*Edit Statistics Labels Appear*): nas colunas, nas linhas, nas camadas. (estas alternativas não se encontram disponíveis quando apenas uma única posição for possível).

Movimente uma ou mais variáveis em alguma ou em todas as caixas “linhas”, caixas “colunas”, ou caixas “camadas”(Layers).

Para que os botões **[OK]** e **[Paste]** estejam disponíveis:

1. Você deve pôr uma variável na lista da linhas.
2. Para as tabelas multi-páginas, etiquetas estatísticas devem estar *Across the top*. Se você puser uma variável na lista da camada, você deve também pôr uma na lista de colunas, a menos que especifique que as etiquetas das estatísticas devem aparecer *Across the top*.

Quando uma variável está destacada nas caixas das linhas, nas caixas das colunas ou nas caixas das camadas, você pode usar os controles *Selected Variables* na direita para modificar a sua função na tabela (*Define Cells* ou *Is summarized*).

Selecione **[Omit Label]** para suprimir a etiqueta da variável seleccionada.

As variáveis múltiplas nestas listas estão agrupadas dentro da tabela a menos que você as desagrupe. A tecla **[Nest]** permite que você agrupe uma variável seleccionada dentro das categorias da variável que a precede na lista (aninhar = agrupar variáveis dentro duma categoria).

A tecla **[Unnest]** permite que você remova uma variável agrupada da sua posição “aninhada”.

Até 4 níveis de agrupamento são permitidos. Você não pode agrupar uma variável "dentro" de uma variável sumariada ou de um total, e você não pode agrupar um jogo múltiplo de respostas dentro de um outro jogo de dados do múltiplo-resposta (variáveis que permitem repostas múltiplas).

Para agrupar uma variável dentro das categorias de uma variável no fundo da lista, remova-a da lista e adicione-a outra vez no fundo. Para estatísticas sumárias do subgrupo, agrupe a variável sumariada dentro das categorias da variável em grupo.

**[Edit Statistics]** permite especificar as estatísticas indicadas para a variável seleccionada. Vocês não podem atribuir estatísticas a uma variável enquanto que uma outra variável é

agrupada. Uma vez que vocês atribuíram estatísticas a uma variável, transformam-se as estatísticas para as outras variáveis na mesma dimensão. Desta maneira é fácil de atribuir as mesmas estatísticas a diversas variáveis na mesma dimensão.

**[Insert Total]** permite que vocês adicionem estatísticas totais após as estatísticas do grupo da variável selecionada. Você pode modificar a etiqueta que será usada para as estatísticas totais na etiqueta *Total Label*. As mesmas estatísticas são indicadas normalmente selecionadas para um total como para a variável totalizada; mas selecionado o total e selecionando **[Edit Statistics]**, você pode especificar estatísticas diferentes para a linha ou para a coluna que contêm as estatísticas totais.

Selecione **[Format]** para ajustar margens, dimensões de página, margens, e o tratamento de colunas e de dados faltantes.

Selecione **[Titles]** para especificar linhas do texto para o cabeçalho da tabela, o rodapé e uma etiqueta no canto do título (alto-esquerdo).

## 6.4 Exercício:

Utilize o SPSS para analisar os dados no arquivo  
**DadosAgregadosFamiliaresComIndividuaisCompleto.sav**

Use SPSS para apresentar aos seguintes resultados:

Tabelas de:

1. Ocupação principal do chefe da família (frequência)
2. Ocupação principal do chefe da família segundo o sexo e por grupos de idade
3. Ocupação principal do chefe da família segundo grau de escolaridade obtido

## 6.5 Gráficos

Os gráficos são, certamente, os meios mais atractivos para comunicar com o utilizador da informação estatística. São recomendados, se pouco tempo ou espaço estiverem disponível para explicar a tese do analista. A publicação das estatísticas através de meios como a televisão ou os jornais é inconcebível sem a presença dos gráficos. No entanto, os gráficos, não são auto-explicativos, pelo que o analista deve ter o cuidado de explicar e comentar os gráficos de um modo apropriado, primeiro através das possibilidades oferecidas pelos próprios gráficos, como é o caso das legendas, ou simplesmente através de texto.

Nesta secção pretende-se que o formando aprenda a criar e a modificar gráficos e que compreenda a sua importância e os contextos em que devem ser utilizados.

## **Criando gráficos**

Existem fundamentalmente três maneiras de criar gráficos no SPSS:

1. Os comandos do menu dos gráficos na barra de menu principais produzem gráficos. A barra de menu principal está disponível quando o editor dos dados, uma janela da saída, ou uma janela da sintaxe estão activos.
2. Determinadas funções estatísticas produzem gráficos. Os exemplos incluem boxplots na função explorar ou scatterplots gerados na regressão linear.
3. Seleccionar gráficos interactivos no menu dos gráficos. Por exemplo, para criar um gráfico de barras, seleccione a barra do menu interactivo. Arraste e transferem variáveis da lista da origem na lista do destino.

Existem fundamentalmente duas maneiras de modificar gráficos no SPSS:

1. Usando o editor de gráfico
2. Usando os gráficos interactivos

## **Gráficos do menu dos gráficos para criar um gráfico...**

Do menu dos gráficos, na barra de menus, selecciona [Graphs].

Selecciona o tipo de gráfico desejado.

Selecciona o ícone para o tipo específico de gráfico que você quer.

As opções da barra do · do menu [Graphs] permitem que você crie um gráfico

[Bar] barra simples, aglomerada, ou empilhada de seus dados.

[Line] permite que você gere um gráfico de linha simples ou múltiplo dos seus dados.

[Area] permite que você gere um gráfico de área simples ou empilhada dos seus dados.

[Pie] permite que você gere uma gráfico de bolo simples ou uma gráfico de barra composta dos seus dados.

[High/Low] permite mostrar pares ou triplos de valores, por exemplo, no caso de acções da bolsa, mostrar os valores máximos, mínimos e preços de fecho.

Existem muitos mais gráficos, cujo uso se recomendado através da ajuda do SPSS ou do tutorial.

## **Os gráficos que resultam das funções**

Os exemplos são os gráficos de barras e os histogramas do procedimento das FREQUENCIES

## Gráficos do menu interactivo dos gráficos

Um conjunto limitado de tipos dos gráficos pode ser seleccionado com a opção do menu: [Graphs/Interactive..].

Os seguintes tipos estão também disponíveis como gráficos interactivos. .

- Bar
- Line
- Pie
- Histogram
- Boxplot
- Error Bar
- Scatterplot

## Usando o editor de gráficos

O editor do gráfico é onde se modifica, imprime, e armazena os gráficos criados noutra local. Aqui você pode alterar quase todos os atributos de um gráfico, bem como o tipo de gráfico (por exemplo, alterar um gráfico de barras para um gráfico da linha).

## O editor do gráfico para editar um objecto do gráfico

(tal como uma linha central, um título, um artigo da legenda etc.).

Primeiro seleccione esse objecto clicando nele uma vez.

Uma mão surge em cima do objecto seleccionado. (um retângulo surge à volta do texto seleccionado)

Seleccionar um objecto, tal como uma barra, significa de seleccionar todos os objectos dessa classe. Por exemplo, seleccionar uma barra num gráfico de barra simples selecciona todas as barras nesse gráfico. Seleccionar uma barra num gráfico de barra aglomerada selecciona todas as barras naquela série.

Após ter seleccionado o objecto, escolha a paleta apropriada para editar o atributo que deseja alterar. Selecciona de um dos seguintes modos:

1. Faça uma selecção do menu dos atributos.
2. Selecciona uma das teclas do atributo na barra do ícone da janela do gráfico. Pode também seleccionar objectos do gráfico através de duplo clique sobre eles, abrindo automaticamente a paleta apropriada.

## 6.6 Exercício:

Utilize o SPSS para apresentar os seguintes resultados:

Gráficos de:

1. Principais razões para não trabalhar nos últimos 7 dias (variável E5)?
2. Grau do ensino mais elevado do chefe da família (C3) por Ocupação principal (E9) do chefe da família e sexo (B1) (gráfico 3-D)
3. Ocupação principal (E9) e Tipo de habitação, Construção de paredes de..(F7)

## 7 A linha de pobreza

Existem diferentes abordagens para determinar as taxas e outras medidas de pobreza. Num estudo<sup>1</sup> - como foi o caso da primeira avaliação nacional da pobreza (MPF/UEM/IFPRI 1998) – seguiu-se a metodologia do custo das necessidades básicas (CBN) para construir linhas de pobreza regionais específicas (Ravallion 1994,1998). As linhas de pobreza resultantes são somas de despesas dos Agregados Familiares (AF), no Total para o AF e, mais importante, por pessoa e por dia.

De acordo com a abordagem do custo das necessidades básicas as linhas de pobreza alimentar estão ligadas à noção das necessidades alimentares básicas que, por seu turno, estão tipicamente associadas aos requisitos energéticos mínimos. Os requisitos energéticos dependem da idade, sexo, nível de actividade física, peso do corpo, estado de gravidez e o estado de amamentação. Visto que o IAF não inclui dados adequados sobre os níveis de actividades físicas ou o peso do corpo, estimamos os requisitos calóricos usando a idade e o sexo, bem como estimativas do Censo de 1997 sobre a proporção de mulheres grávidas e a amamentar. Neste sentido, utiliza-se um valor do Equivalente adulto para cada pessoa, de acordo com as características referidas para obter um valor realista do consumo. Obviamente o número dos Equivalentes adultos (EA) é inferior ao número de pessoas no AF. Dividindo as despesas totais pelo número dos EA obtém-se o valor das despesas per capita.

Na abordagem de CBN a linha de pobreza total é construída como a soma da linha da pobreza alimentar e a linha da pobreza não-alimentar. Uma vez construída a linha de pobreza, os agregados que apresentam valores per capita menores do que o valor indicado pela linha de pobreza são considerados pobres. No futuro chamaremos a este valor simplesmente linha de pobreza (LdP). Como qualquer linha de pobreza, a linha de pobreza dos produtos alimentares e não alimentares incorporam julgamentos de valor no que diz respeito às necessidades de base dos bens alimentares e não-alimentares. São especificadas em termos de níveis de gastos de consumo per capita que são consistentes com o alcance dessas necessidades básicas.

Neste curso não vamos calcular a linha de pobreza, pelo contrário, vai ser utilizado o valor da linha de pobreza existente no Quadro 1 do documento já referido<sup>1</sup>. Os valores existentes para Sofala são os seguintes:

	Linha de Pobreza Alimentar do Cabaz Flexível Não - Ajustado 2002-03 (Cabaz 2002-03 a Preços 2002-03) Valores em Meticais pro Pessoa por Dia
Sofala e Zambezia rural	5473
Sofala e Zambezia urbana	8775

### Resultados

À semelhança de 1996-97, foram usadas três medidas para medir a pobreza. Todas elas são membros da classe  $P_\alpha$  de índices de pobreza de Foster-Greer-Thorbecke (1984)<sup>2</sup>, e são usadas

<sup>1</sup> **Pobreza E Bem-Estar Em Moçambique: Segunda Avaliação Nacional** - Março de 2004 Direção Nacional do Plano e Orçamento, Ministério de Plano e Finanças Gabinete de Estudos, Ministério de Plano e Finanças Instituto Internacional de Pesquisa em Políticas Alimentares (IFPRI) Universidade de Purdue

<sup>2</sup> **Greer, J., J. Greer, and E. Thorbecke.** 1984. A class of decomposable poverty measures. *Econometrica* 52 (3): 761-765

de uma forma rotineira para medir a pobreza. Matematicamente, todos os índices desta classe têm a forma:

$$P_\alpha = \frac{1}{n} \sum_{y \leq z} \left(1 - \frac{y}{z}\right)^\alpha, \alpha \geq 0$$

Onde:  $n$  é a população,  $y$  é o consumo per capita,  $z$  é a linha de pobreza, e  $\alpha$  é um parâmetro não negativo.

Usamos as medidas com  $\alpha = 0, 1, \text{ e } 2$ , que correspondem à incidência da pobreza, ao índice de profundidade da pobreza, e ao índice de profundidade da pobreza ao quadrado, respectivamente.

## **7.1 O índice de incidência da pobreza (poverty headcount index)**

Significa a proporção da população cujo consumo per capita está abaixo da linha da pobreza. Este índice pode também ser expresso matematicamente como  $P_0 = q/n$ , onde  $q$  é o número de pessoas pobres numa dada região e  $n$  é a população total dessa região.

## **7.2 O índice de profundidade da pobreza (poverty gap index)**

Significa a distância média, em percentagem, em que o consumo medido se encontra abaixo da linha de pobreza, utilizando todos os agregados na amostra e onde os agregados que vivem acima da linha da pobreza recebem o valor zero. Matematicamente, significa a diferença média entre os níveis de consumo dos pobres e a linha de pobreza (expresso como uma proporção da linha de pobreza) multiplicada pela incidência da pobreza.

Uma vez que o índice de profundidade da pobreza mede “Quão pobres são os pobres?” este detecta mudanças na pobreza que o índice de incidência da pobreza não detecta. Por exemplo, se todos os pobres permanecessem abaixo da linha da pobreza e todos os não pobres permanecessem acima da linha de pobreza, mas os rendimentos dos pobres subissem, muitos diriam que a pobreza decresceu. A incidência da pobreza não mudaria para reflectir esta melhoria mas o índice de profundidade da pobreza irá decrescer, para mostrar que os pobres não são tão pobres como eram antes.

## **7.3 O índice de profundidade da pobreza ao quadrado (squared poverty gap index)**

Significa a média do quadrado da profundidade da pobreza e mede a severidade da pobreza, tendo em conta as desigualdades entre os pobres. Por exemplo, se uma transferência é feita duma pessoa ligeiramente colocada abaixo da linha da pobreza para uma pessoa muito abaixo da linha de pobreza, o índice de profundidade da pobreza ao quadrado irá reduzir porque os padrões de vida dos mais pobres entre os pobres terão melhorado. Em contraste, tal transferência não afectaria nem o índice de incidência nem o índice de profundidade da pobreza.

Tornar-se-á mais claro através das fórmulas:

Para conhecer a proporção dos pobres em relação à população total calcula-se a proporção:

$$P_0 = q / n$$

onde  $n$  é o número das observações na população

$q$  é o número dos pobres considerados pobres (número de observações em baixo da linha de pobreza)

e  $z$  é sempre o valor da linha da pobreza

Embora este valor meça a percentagem dos pobres nada diz, no entanto, sobre as características dos dados da amostra relativamente aos pobres. A fórmula seguinte mede a profundidade da pobreza, significando que um valor maior representa um parte substancial da população se encontra mais afastada da LdP do que um valor menor (tem mais 'muito' pobres se este valor fosse elevado, tem menos 'muito' pobres se este valor fosse mais baixo). De modo simplificado, pode concluir-se que uma profundidade elevada da pobreza reflecte a existência de uma parte significativa da população extremamente pobre, enquanto uma profundidade baixa reflecte uma parte reduzida da população extremamente pobre, relativamente aos pobres "médios".

$$P_1 = q / n * (z - y^p) / z$$

em que  $y^p$  é a media do padrão de viver (no nosso caso são as despesas)  $y_i$  entre os pobres

A terceira formula:

$$P_2 = q / n * [(z - y^p) / z + (S_p / z)^2]$$

onde  $S_p$  é o desvio padrão do  $y_i$  entre os pobres

A característica mais importante destes indicadores talvez seja a sua possibilidade de decomposição, ou seja, que o índice pode ser aplicado aos grupos (conjuntos no senso matemático) disjuntos, por exemplo, urbano e rural, e os índices se agrupam e se decompõem consoante os grupos seleccionados.

## 7.4 Exercício:

Utilize o SPSS para analisar os dados no arquivo

**DadosAgregadosFamiliaresComIndividuaisCompleto.sav**

Utilizam o SPSS para apresentar os seguintes resultados:

1. Calcula as despesas per capita e as despesas por Equivalente Adulto. Cria duas variáveis novas e usa [Transform/Compute]
2. Usa o SPSS e o EXCEL para calcular os índices de pobreza ( $\alpha = 1, 2, 3$ ) em Sofala e para os grupos seguintes
  - a) Chefes de Família masculino e feminino
  - b) Agregados Familiares "grandes" > 6 pessoas e "pequenas" <=6
  - c) Zona rural e urbana

## 8 Análise da Regressão

**Fundamentos:** O que é regressão (linear): Calcula as coordenadas para uma linha recta usando "o método dos mínimos quadrados". Define uma linha recta que permite os melhores ajustamentos dos seus dados, e toma como base uma matriz de dados que descreva a linha. Porque esta função tem como base uma matriz de valores, deve ser incorporada como uma fórmula de matriz.

### 8.1 As equações de regressão

A equação para a linha é:

$y = m \cdot x + b + r$  para a regressão simples

ou

$y = m^1 \cdot x^1 + m^2 \cdot x^2 + \dots + b + r$  (se há valores múltiplos de  $x$ ) onde

$x$  e  $x^n$  são as variáveis independentes e

$y$  é a variável dependente

$b$  é a constante ou intercepto (a linha de regressão intercepta o eixo  $y$ )

$m^n$  são os coeficientes da regressão (os predictores ou estimadores) e

$r$  os residuais

Há diversos indicadores estatísticos que ilustram 'a qualidade' da relação entre as variáveis:

- Uma tabela Anova
- Os coeficientes de regressão, os erros padrão e os intervalos de confiança
- O erro padrão da estimativa de  $y$
- Os valores do  $R^2$  e da sua raiz quadrada  $R$

O indicador principal para a qualidade, como uma (ou muitas) variáveis explica(m) a variável dependente, é o  $R$  (correlação estatística) e o seu valor quadrado  $R$ .

Valores próximos de zero indicam a quase inexistência de relação entre as variáveis enquanto valores próximos de +1 indicam uma relação forte entre as variáveis.

A Pergunta: Existem indicações de que existe uma relação entre a pobreza e a idade do chefe do agregado familiar (AF)

### 8.2 Um Exercício: Dados do IAF de Sofala

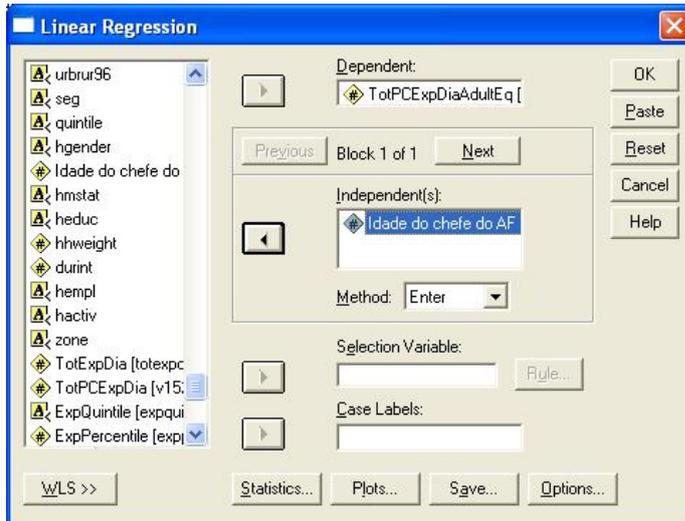
Relacionar a Despesa (a variável dependente e indicador de pobreza - Nome da variável: *totpcexp*) e a idade do chefe do agregado familiar. (Nome da variável: *hage*)

SPSS - Menu: [Analyze/ Regression/Linear]

Comando: REGRESSION

Selecione uma variável dependente numérica, e uma ou mais variáveis independentes numéricas.

Você pode controlar a entrada de variáveis independentes na análise, de duas formas: agrupando os respectivos dados em blocos, e escolhendo o método pelo qual as variáveis em cada bloco são processadas. Os métodos disponíveis são ENTER, REMOVE, STEPWISE, BACKWARD e FORWARD.



Começando com o primeiro bloco de variáveis, SPSS aplica o método seleccionando todas as variáveis no bloco; prossegue então para o bloco seguinte se houver mais do que um.

Selecione [Statistics..] para estatísticas adicionais. Selecione [Plots..] para scatterplots dos residuais, histogramas, gráficos dos outliers ou gráficos sobre a probabilidade da normalidade dos dados. Selecione [Save..] para criar as variáveis novas que contêm valores previstos,

residuais e estatísticas relacionados.

Selecione [Options..] para mudar os critérios usados nos métodos STEPWISE, pedir a linha (planície) da regressão que passa pela origem (quando Você deselectione a opção “Include Constant in equation”) ou controlar o tratamento de valores ausentes.

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-14839.1	85772.551		-.173	.865	-200139.382	170461.260
	Idade do chefe do AF	4297.214	2209.000	.493	1.945	.074	-475.040	9069.469
	HHSIZE	-17372.5	10108.764	-.436	-1.719	.109	-39211.127	4466.184

a. Dependent Variable: TotPCExpDiaAdultEq

## 8.3 As características do modelo da regressão

Os *Coefficients* são os coeficientes eles mesmos e

Os *Confidence Intervals* · são intervalos da confiança de 95% para os coeficientes

O *Covariance matrix* dá as variâncias e as covariâncias entre os coeficientes estimados.

O *Descriptives* fornece as medidas de tendência central e os desvios padrão de cada variável na análise, para além de uma matriz da correlação.

Model Summary <sup>b</sup>					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.334 <sup>a</sup>	.111	.048	94451	1.722

a. Predictors: (Constant), Idade do chefe do AF  
b. Dependent Variable: TotPCExpDiaAdultEq

**Modelo com uma variável independente**

O *Model summary* contém estatísticas do modelo e inclui o R, o R quadrado e a estimativa do R ajustado de R, o erro padrão da estimativa e uma tabela da análise da variância.

Adicionalmente: *Durbin Watson* indica o teste de Durbin-Watson para a correlação de séries residuais. (Descobre se as variáveis são ou não independentes ente si).

Model Summary <sup>b</sup>					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.525 <sup>a</sup>	.276	.165	88479	1.341

a. Predictors: (Constant), HHSIZE, Idade do chefe do AF  
b. Dependent Variable: TotPCExpDiaAdultEq

**Modelo com duas variáveis independentes**

Nota-se que o segundo modelo explica melhor a “Pobreza” de uma maneira mais satisfatória.

## O exemplo de galinhas

	Ovos	Idade (meses)
Gal1	0	2
Gal2	0	1
Gal3	1	4
Gal4	1	7
Gal5	1	10
Gal6	2	14
Gal7	2	12
Gal8	2	9
Gal9	2	6
Gal10	3	8
Gal11	3	10
Gal12	4	23
Gal13	4	8
Gal14	4	9
Gal15	5	12
Gal16	5	18

Para explicar melhor os conceitos da regressão utilizamos um exemplo muito simples. 16 galinhas numa capoeira põem ovos com uma frequência semanal diferente. O proprietário das galinhas suponha que a idade das galinhas tinha a ver com o número dos ovos postos. Fez uma tabela numa semana determinada com os ovos postos e a idade de cada galinha em meses.

Calcula-se a regressão com ovos como a variável dependente e idade como a variável independente.

Chegamos a uma equação de regressão simples da seguinte forma:

Model			
Coefficients	Unstandardized	95% Confidence Interval for E	
	Coefficients	Lower Bound	Upper Bound
	B		
(Constant)	0.523	-0.812	1.859
IDADE	0.200	0.079	0.322

Dependent Variable: OVOS

Que significa um equação da regressão:

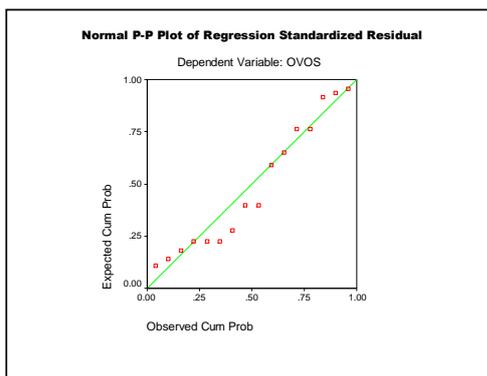
$$\text{Ovos} = 0.523 + 0.200 * \text{Idade}$$

Model Summary			
Model	R	R Square	Durbin-Watson
	0.6866	0.4714	0.7383
a	Predictors: (Constant), IDADE		
b	Dependent Variable: OVOS		

A qualidade da Regressão é boa mais não muito elevada.

Os residuais indicam um auto-correlação positiva, sendo o coeficiente Durban-Watson  $< 1$ . Isto porém não é um problema porque não se trata de uma série

cronológica.

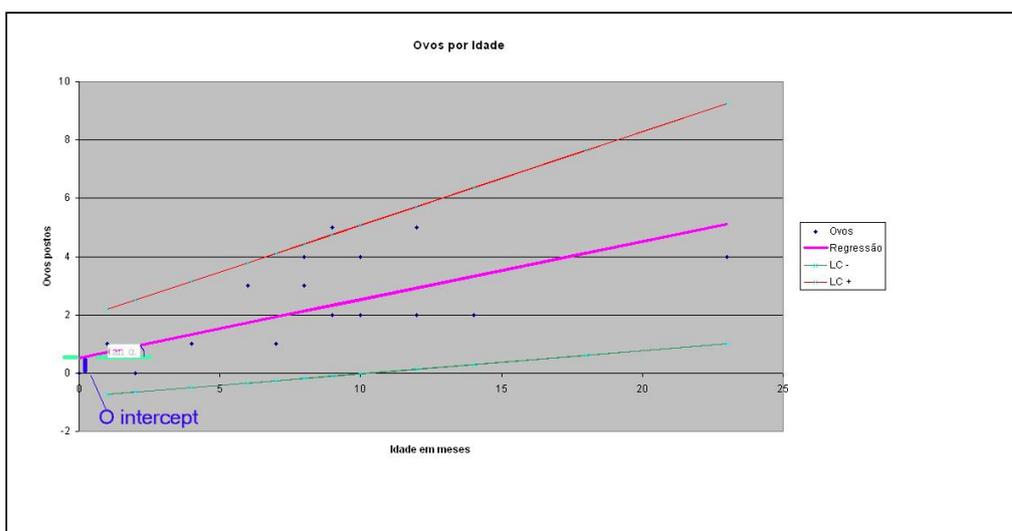


Os gráficos P-P são usados geralmente para determinar se a distribuição de uma variável está de acordo com um determinado tipo de distribuição (normal neste caso). Se a variável selecciona estiver de acordo com a distribuição do teste, os pontos aglomeram-se em torno de uma linha recta, como é o caso desta nossa regressão.

A inclinação é a relação entre a mudança vertical e a mudança horizontal da linha (planície) de regressão. O *intercept* ou a constante, chamada assim frequentemente, indica onde a linha (planície) de regressão intercepta o eixo Y;  $x = 0$  (isto é, quando a expectativa das variáveis independentes é 0, o intercept está à altura de 0 até à linha de regressão).

## Um gráfico de Regressão

Para explicar isto incluímos os dados e a regressão num gráfico de EXCEL



Os ovos aparecem como pontos de diamantes no gráfico. A equação da regressão leva uma linha recta grossa no gráfico. O intercept é a distância ente o ponto 0 e o ponto onde a linha da regressão cruza o eixo Y. O coeficiente da regressão ( $x=0.2$ ) está na tangente do ângulo entre a linha da regressão e o eixo X., muita vezes interpretado como:

Enquanto a variável dependente varia numa unidade a variável independente varia de acordo com o coeficiente da variável.

**No exemplo:** com um mês mais de idade, é esperado que a galinha ponha mais 0.2 ovos. Com mais 5 meses de idade, pode-se esperar que uma galinha ponha mais um ovo.

Para representar os erros no modelo, extraia uma linha vertical de cada ponto à linha de regressão. O comprimento destes segmentos entre a linha e os pontos do gráfico são chamados residuais e são estimativas para os erros no modelo. O SPSS usa o método dos mínimos quadrados para estimar a inclinação e o intercept. Este método minimiza a soma de residuais quadrados (isto é, a soma dos quadrados dos segmentos de linha verticais).

Na primeira equação,  $y$  é o dependente ou a variável do resultado, essa que você está tentando de prognosticar;  $x$  é a variável independente ou do predictor; e o intercept e a inclinação são coeficientes do modelo ou da equação da regressão.

Se o modelo for um bom descritor da relação entre as variáveis, você pode usar as estimativas dos coeficientes para prever o valor da variável dependente para casos novos.

**No exemplo:** Pode-se prever quantos ovos uma galinha põe com a idade de 36 meses.

Usa-se a fórmula:  $Ovos = 0.523 + 0.200 * Idade$  e o resultado é 8 ovos. Já se percebe a fraqueza das previsões da regressão desta forma simples. Não toma em conta o ciclo de vida duma galinha. A galinha velha não põe mais ovos, antes pelo contrário! Além disso, qualquer previsão estatística está sujeita a incerteza da indução estática. São os intervalos de confiança que permitem calcular a linha de confiança inferior (LC-) e superior (LC+) com os valores respectivos para o intercept e o coeficiente; então, a fórmula para linha de confiança inferior é

$$Ovos_{LC-} = -0.812 + 0.079 * Idade$$

e a fórmula para linha de confiança superior é

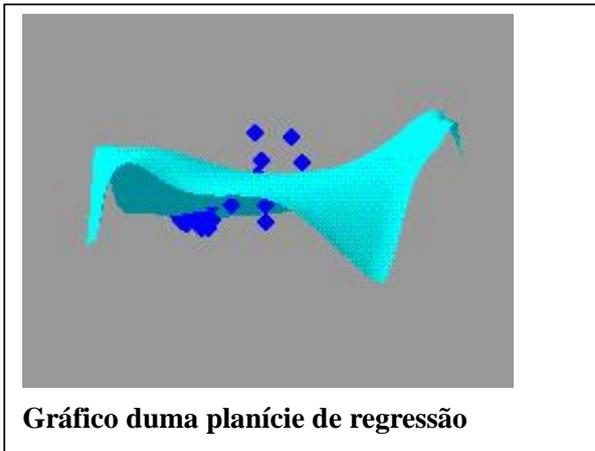
$$Ovos_{LC+} = 1.859 + 0.322 * Idade$$

No gráfico aparecem as Linhas de Confiança inferior e superior de 95% em forma mais fina. Dependendo desta amostra, pode-se esperar que, para uma outra galinha e em cada 19 de 20 casos, a relação entre ovos e idade indica que os ovos postos por esta galinha caem nesta faixa de confiança. Porém em um caso entre 20, este valor sai desta faixa. Mais um problema para uma previsão na base duma regressão

## 8.4 Modelos com dois ou mais predictores.

Adicionando uma segunda variável independente: A equação com uma variável independente é o modelo para a regressão linear simples; a equação com duas variáveis é um modelo para a regressão múltipla. O SPSS permite que você inclua mais de duas variáveis independentes em correlações de uma regressão múltipla. Se a correlação de Pearson entre as duas variáveis for significativa com um valor menos de 0.1 que indica a hipótese que a correlação é 0 (nenhuma relação linear entre as variáveis) está rejeitado. Quando O seu modelo inclui mais do que uma

variável independente e você escolhe *Colinearity diagnostics* na opção [Statistics..] SPSS indica correlações para todos os pares das variáveis



O gráfico deste modelo já não é uma linha com uma planície. Para mais variáveis ainda é mais difícil visualizar as planícies da regressão.

A normalidade não é requerida para as estimativas dos coeficientes. Para fazer testes e intervalos da confiança da estimativa, entretanto, estas suposições subsequentes são requeridas:

- os erros são distribuído com distribuição normal e media de 0
- os erros tem uma variância constante.
- os erros são independentes entre si.

Estas suposições são verificadas estudando os residuais do modelo.

A estatística de Durbin-Watson (disponível na regressão linear: na caixa do sub-diálogo dos [Statistics..]) pode ser usada para testar a correlação de séries de residuais de variáveis adjacentes.

## 8.5 Sumário do Modelo.

O valor de R (chamado também múltiplo R). Quando há somente uma variável independente, R é a correlação simples entre a variável independente e dependente. R<sup>2</sup> é o quadrado deste valor e é interpretado frequentemente como a proporção da variação total na expectativa da variável dependente explicado pelas variáveis independentes (o consumo do fertilizante " explica "???" % dos resultados da colheita), varia de 0 a 1. Se não houver nenhuma relação linear entre a variável dependente e independente, R<sup>2</sup> é 0 ou muito pequeno.

Se todas as observações caírem na linha de regressão, R<sup>2</sup> é 1. Esta medida da qualidade de ajuste ("goodness of fit") de um modelo linear é chamada também o **coeficiente da determinação**.

Mas cuidado! Tenha cuidado a concluir " se o consumo do fertilizante for aumentado, a colheita subirá na respectiva proporção" Há uma associação entre o consumo do fertilizante e a expectativa de colheita. Entretanto, estes dados vêm de uma observação, não duma experiência controlada; assim todas as indicações sobre relacionamentos do motivo-e-efeito podem ser enganadoras. A associação não é a mesma que o motivo ou a causa..

Se um investigador observar os valores das variáveis independentes e dependentes para um jogo dos sujeitos (casos), a associação não estabelece a causa. Se um investigador fizer uma experiência onde os valores da variável independente (para o exemplo, seis doses específicas de uma droga) foram ajustados pelo investigador e os efeitos na variável dependente foram observados, pode haver poucas dúvidas que os resultados estatísticos podem avaliar a causa dos efeitos na variável dependente pela variável independente. Mas o mundo dos estaticistas observando pessoas ou galinhas é diferente; dependemos das amostras e do mundo com todas as incertezas. É importante lembrar que um melhor conhecimento das ferramentas estatísticas minimiza os riscos duma interpretação errónea.

## 8.6 Exercício:

Use o SPSS para analisar os dados no arquivo do IAF de Sofala

**DadosAgregadosFamiliaresComIndividuaisCompleto.sav**

1. Use SPSS para a regressão linear e a série de dados de Sofala para relacionar-se: Nível de pobreza / despesa (a variável dependente) e sexo do chefe do AF e o grau mais elevado de ensino (C3),
2. Use mais variáveis explicativas para a pobreza, incluindo variáveis do emprego( E..)
3. Utilizem variáveis das características da habitação (F..). São variáveis independentes ou dependentes.

Discutam os resultados e a validade destes resultados

4. Calcule uma linha da pobreza na base do IAF, independente dos dados conhecidos. Discuta primeiramente os critérios para estabelecer esta medida. Que despesas devem ser consideradas e em que dimensão? Calculem três grupos de riqueza/pobreza – ricos, médios, pobres.
5. Usem as previsões para os grupos calculados no exercício da linha de pobreza com o uso dos “melhores” predictores (variáveis que expliquem melhor). Discutam e apresentem os resultados.
6. Calcule as previsões destes índices para várias sub-amostras:  
As AF com chefe feminino,  
Os grupos urbanos rurais,  
Grupos por Distrito (?)  
Os grupos de prosperidade.

Como será o impacto das variáveis explicativas escolhidas?

## 9 Bibliografía

### 9.1 SPSS

SPSS 10.0 Manuals - SPSS Inc. Headquarters, 233 S. Wacker Drive, 11th floor Chicago, Illinois 60606

SPSS 10.0 Regression Models ISBN 0130179043

SPSS 10.0 for Windows Student Version ISBN 0130280402

Discovering Statistics Using SPSS for Windows : Advanced Techniques for Beginners (Introducing Statistical Methods series) by Andy Field (Paperback )

SPSS/PC+ Basics and Graphics - Gerhard Brosius McGrawHill ISBN 3-89028-132-X

SPSS/PC+ Advanced Statistics and Tables Gerhard Brosius McGrawHill ISBN 3-89028-157-5

SPSS Base System und Professional Statistics - Gerhard Brosius, Felix Brosius International Thomson Publishing ISBN 3-929821-62-1

### 9.2 *Estatística, Analise de Pobreza*

How to Lie With Statistics by Darrell Huff, Irving Geis

Deaton, A., and S. Zaidi. 1999. *Guidelines for constructing consumption aggregates for welfare analysis*. Research Program in Development Studies Working Paper No. 192. Princeton, N.J.: Princeton University.

J.-L-Dubois, D.Blaizeau - Connaître les conditions de vie de ménages dans les pays en développement –Tome 3: Analyser les résultats- - Ministère de la Coopération et du développement- ISBN 2-11-884855-3

Foster, J., J. Greer, and E. Thorbecke. 1984. A class of decomposable poverty measures. *Econometrica* 52 (3): 761-765.

Ravallion, M. 1994. *Poverty comparisons*. Chur, Switzerland: HarwoodAcademic Publishers.

- ' 1998. *Poverty lines in theory and practice*. Living Standards Measurement Study Working Paper No. 133. Washington, D.C.: World Bank.

- ' 2001. *Growth, inequality and poverty: Looking beyond averages*. Policy Research Working Paper No. 2558. Washington, D.C.: World Bank.

Tarp, E, C. Arndt, H. T. Jensen, S. Robinson, and R. Heltberg. 2002a. *Facing the development challenge in Mozambique: An economywide perspective*. Research Report 126. Washington, D.C.: International Food Policy Research Institute.

Tarp, F., K. R. Simler, C. Matusse, R. Heltberg, and G. Dava. 2002b.. *Economic Development and Cultural Change* 51 (1): 77-108.

### 9.3 WWW

<http://www.spss.com/>

[http://www.ats.ucla.edu/stat/spss/library/sp\\_hist.htm](http://www.ats.ucla.edu/stat/spss/library/sp_hist.htm)