

**HOSPITAL DE CLÍNICAS DE PORTO ALEGRE
GRUPO DE PESQUISA E PÓS GRADUAÇÃO**

**INTRODUÇÃO À ANÁLISE ESTATÍSTICA
UTILIZANDO O SPSS 18.0**

**Marilyn Agranonik
Vânia Naomi Hirakata
Suzi Alves Camey**

Porto Alegre - Setembro de 2010

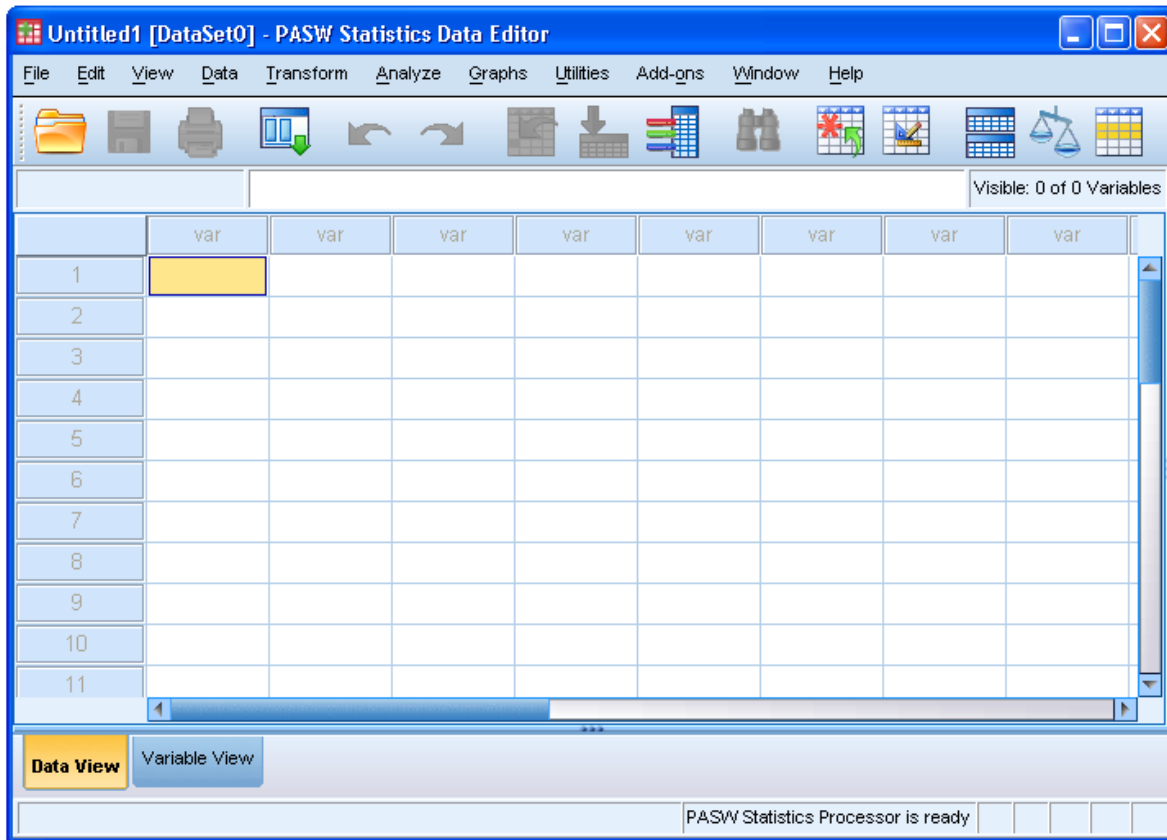
ÍNDICE

1. INTRODUÇÃO AO SPSS	3
1.1 - Banco de dados: definição	4
1.2 - Como criar um banco de dados	4
1.3 - Como dar nome a uma variável	6
1.4 - Como dar nome aos níveis de uma variável	6
1.5 - Como acessar um banco de dados já existente	7
1.6 - Como exportar um banco de dados do spss para o excel	7
2. INTRODUÇÃO AOS MÉTODOS ESTATÍSTICOS	8
2.1 - Tipos de variáveis: definição	8
2.2 - Descrição e exploração de dados	8
2.3 - Análise univariada	9
2.3.1 - Variáveis quantitativas	9
2.3.1.1 - Como obter as estatísticas descritivas	9
2.3.1.2 - Como obter um histograma	10
2.3.1.3 - Como obter um box-plot	11
2.3.2 - Variáveis categóricas (qualitativas)	13
2.3.2.1 - Como obter a distribuição de freqüências	13
2.3.2.2 - Como obter gráficos	14
2.3.2.2.1 - Como obter gráfico de barras	14
2.3.2.2.2 - Como obter gráfico de pizza	16
2.4 - Como exportar o output para word	18
3 - CATEGORIZAÇÃO DE VARIÁVEIS	18
3.1 - Como categorizar uma variável quantitativa	18
3.2 - Como criar uma variável a partir de datas	21
3.3 - Como criar uma variável através da combinação de outras duas	22
3.4 - Como criar uma variável através de um cálculo com outras variáveis	23
3.5 - Como criar uma variável a partir da contagem de outras	24
4. MANIPULAÇÃO DE DADOS	25
4.1 - Ordenação de casos: Sort Cases	25
4.2 - Seleção de casos: Select Cases	27
4.3 - Divisão de análises por grupos: Split File	30
4.4 - Juntar bancos de dados: Merge Files	32
4.4.1 - Juntando bancos com diferentes casos	32
4.4.2 - Juntando bancos com diferentes variáveis	32
4.5 - Manipulação de arquivos	33
4.6 - Edição do arquivo de resultados (*.spo): como apagar análises não desejadas no arquivo de resultados	34
5. ALGUMAS ANÁLISES BIVARIADAS	34
5.1 - Comparação de duas variáveis categóricas: Teste Qui quadrado	34
5.1.1 - Como obter os resíduos ajustados	36
5.2 - Como comparar médias entre dois grupos: Teste t para Amostras Independentes	37

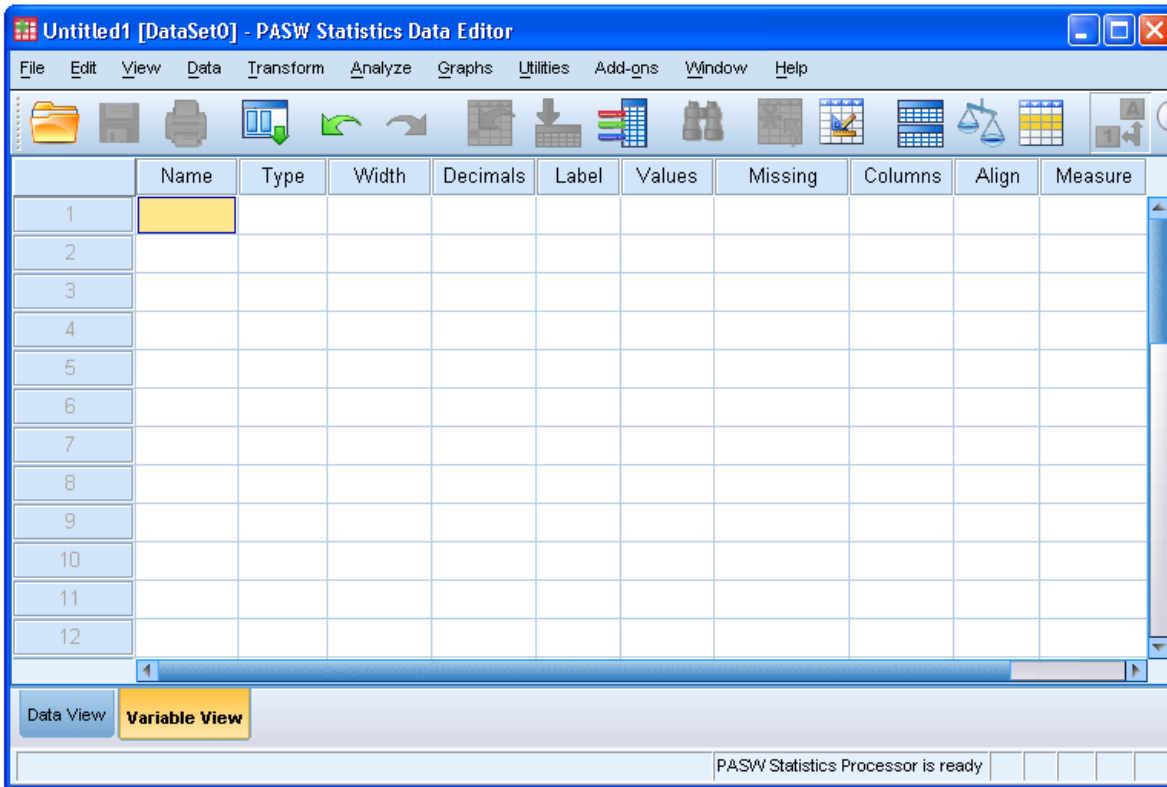
1. INTRODUÇÃO AO SPSS

O pacote estatístico **SPSS** (*Statistical Package for Social Sciences*) é uma ferramenta para análise de dados utilizando **técnicas estatísticas** básicas e avançadas. É um software estatístico de fácil manuseio e é internacionalmente utilizado há muitas décadas, desde suas versões para computadores de grande porte.

Tela inicial do *SPSS 18.0 for Windows*.



(a)



(b)

Figura 1: Tela inicial do *SSPS 18.0 for Windows*. 1(a): Planilha “Data View”; 1(b): Planilha “Variable View”.

1.1 - Banco de dados: definição

Banco de dados é um conjunto de dados registrados em uma planilha, em forma de matriz, com “n” linhas, correspondentes aos casos em estudo e “p” colunas, correspondentes às variáveis em estudo ou itens de um questionário.

O número de casos (número de linhas da matriz) deve ser, em geral, **maior** do que o número de variáveis em estudo (número de colunas).

1.2 - Como criar um banco de dados

Para se criar um BANCO DE DADOS novo procede-se da seguinte forma:

- a) Clicar em “**File**”; “**New**”; “**Data**”. Aparece a planilha “**data view**”. Na primeira linha estão indicadas as posições das variáveis (VAR001,

- VAR002,etc.), e uma margem vertical numerada a partir de 1 (como mostrado na Figura1a).
- b) Na primeira coluna, correspondendo à 1ª VAR, criar uma variável, por exemplo, “NumCaso” com o número do questionário ou do caso em estudo. Veja exemplo 1 abaixo.
 - c) Para serem registradas as características das variáveis em cada coluna, clique **duas vezes** sobre a coluna **var**. Aparece a planilha **variable view** na qual cada variável está definida em uma linha.
 - d) Na primeira coluna (**Name**), digitar o nome da variável desejada (Ex: **NumCaso**). Para o nome das variáveis utilize 8 dígitos no máximo e não utilize espaço em branco, nem os símbolos “-“, “. ” e ”/”. O SPSS não permite que o nome de uma variável comece com um número.
 - e) Clicar na coluna **Type** para definir o tipo de variável. Aparece uma janela **Variable Type** onde deve-se deixar a opção **Numeric** se a variável for numérica ou **String** se a variável for alfa-numérica (texto). Preferencialmente use sempre a modalidade **Numeric** para variáveis categóricas, como por exemplo, sexo, estado civil, município, etc. criando-se um código para as categorias.
 - f) No caso de não-resposta ou respostas que não se desejam considerar para o tratamento estatístico, como por exemplo, respostas não corretas, não se aplica, sem resposta, etc..., clicar na coluna **Missing**, abre-se uma janela (**Missing Values**), registrar, na opção **Discrete Missing Values**, o código de não-resposta, preferencialmente **9, 99, 999, etc**. Clicar em **OK**. Entretanto, a melhor opção para não resposta é deixar o espaço em branco no banco de dados.
 - g) Retornar à planilha **Data View** e passar a digitar, em cada linha da coluna identificada, o valor da variável.
 - h) À medida que as informações vão sendo registradas no BANCO DE DADOS é importante salvá-las. Para tanto, se procede da seguinte forma: clicar em **File → Save as...** (abre-se a janela na qual se escolhe o caminho desejado para salvar o arquivo) e criar um nome para o Banco de dados, que terá automaticamente a terminação **.sav**.

Exemplo 1: A tabela a seguir contém informações sobre 5 indivíduos para as seguintes variáveis:

- 1) Número de identificação (numcaso)
- 2) Nome (nome)
- 3) Sexo (sexo)
- 4) Data de nascimento (datanasc)
- 5) Altura em metros (altura)

Digite as informações no SPSS:

numcaso	nome	sexo*	datanasc	altura
1	A	1	05/01/1975	1,65
2	B	0	12/06/1984	1,80
3	C	0	24/09/1978	1,87
4	D	1	13/02/1981	1,68
5	E	1	30/04/1974	1,75

*0: masculino e 1: feminino

1.3 - Como dar nome a uma variável

No campo **label** é possível digitar um nome mais completo para a variável. Por exemplo, para a variável **NumCaso** é possível digitar em **label** a expressão “número do questionário”.

1.4 - Como dar nome aos níveis de uma variável

É conveniente registrar no banco de dados os nomes das categorias de variáveis categóricas. Por exemplo, para a variável **sexo** os códigos poderiam ser: 0: masculino e 1: feminino. Para registrar estes nomes, clicar **2 vezes** sobre a variável **sexo**, abrindo a planilha **Variable View** e proceder da seguinte forma:

- a) Clicar no campo **Values** da variável de interesse. Abre-se a janela **Value Labels**:
- b) Em **Value**, digitar **0**;
- c) Em **Label**, digitar masculino;
- d) Clicar em **ADD**;
- e) Proceder-se da mesma forma para os demais níveis de categorização: digitar **1** para **Value** e feminino para **Label**, seguindo-se por **ADD**;
- f) Clicar em **OK**.

OBSERVAÇÃO:

A manipulação do BANCO DE DADOS nos permite:

- Criar e recodificar variáveis;
- Realizar análise de dados através de estatísticas descritivas, gráficos, etc;
- Selecionar casos para análise, repetir a análise para grupos de casos diferentes.

É importante dar ao arquivo o nome mais claro possível para facilitar sua localização e acesso. Os arquivos de dados são do tipo **.sav**.

RECOMENDAÇÃO: A primeira coluna da matriz deve corresponder ao número do questionário, número do caso, ou ainda código do registro, pois facilita a localização de informações no caso de serem identificados equívocos de digitação.

1.5 - Como acessar um banco de dados já existente

Para acessar um banco de dados já existente, procede-se da seguinte maneira:

- a) Iniciar o programa **SPSS** (clique 2 vezes sobre o ícone);
- b) Clique em **File → Open → Data**, abra o arquivo que se deseja. Neste manual usaremos como exemplo o arquivo **World95.sav** que se encontra disponível junto com o programa SPSS.

1.6 - Como exportar um banco de dados do spss para o excel

Para se exportar um BANCO DE DADOS do SPSS para o EXCEL procede-se da seguinte forma:

- a) Clique em **File → Save as** (abra-se a janela do caminho desejado) e crie um nome para o Banco de dados;
- b) Em **Save as type** escolha **Excel**;
- c) Clique em **Save**. No local escolhido será salvo um novo arquivo com a terminação **.xls**.

Para exportar para outros programas procede-se da mesma maneira, apenas mudando em **“Save as type”** para o tipo de sua preferência.

2. INTRODUÇÃO AOS MÉTODOS ESTATÍSTICOS

2.1 - Tipos de variáveis: definição

Do ponto de vista estatístico, para decidirmos qual a análise estatística apropriada, devemos distinguir entre dois tipos básicos de variáveis:

- **Variáveis quantitativas:** são variáveis que podem ser mensuradas através de escalas quantitativas, isto é, escalas que têm unidades de medida. **Ex.: Renda Familiar** (medida em R\$ ou em salários mínimos); **Idade** (medida em anos, ou meses); **Faturamento de uma Empresa** (R\$, US\$); **Nº de Empregados** (Nº), **Peso** (em kg), **Altura** (em cm)...
- **Variáveis qualitativas ou categóricas:** são variáveis medidas originalmente em categorias. Ex: **Sexo, Profissão, Religião, Município, Região...**

2.2 - Descrição e exploração de dados

O objetivo básico deste procedimento é introduzir técnicas que permitam organizar, resumir e apresentar os dados, de tal forma que possam ser interpretados de acordo com os objetivos da pesquisa e o tipo de variável.

Um primeiro passo para analisar qualquer banco de dados é avaliar separadamente cada variável (o que será denominado de **análise univariada**). Se as variáveis são quantitativas usamos estatísticas descritivas (ex: média, desvio padrão, valor mínimo, valor máximo) ou gráficos (ex: histograma e box plot). Se as variáveis são qualitativas usaremos tabelas de freqüência ou gráficos (ex: gráfico de barras e gráfico de setores, também conhecido como gráfico de pizza).

OBSERVAÇÃO:

Não podemos calcular média, variância ou desvio-padrão de variáveis qualitativas ou variáveis categóricas.

2.3 - Análise univariada

2.3.1 - Variáveis quantitativas

2.3.1.1 - Como obter as estatísticas descritivas

A partir daqui será utilizado o banco **WORLD95.SAV**. Para calcular as estatísticas descritivas é possível proceder de 2 formas:

1 - Através do comando **Frequencies**

- Clicar em **Analyze → Descriptive Statistics → Frequencies**;
- Localizar na janela à esquerda a variável de interesse (por exemplo, **babymort: mortalidade infantil**) e clicar na →;
- Clicar em **Statistics**, e assinalar as opções desejadas (ex: quartiles, mean, standard deviation, minimum, maximum);
- Clicar em **Continue; OK**;
- Os resultados da análise estatística aparecem na janela de resultados (**OUTPUT**), que poderá ser salva, dando origem a um arquivo do tipo **.spo (SPSS output)**.

EXEMPLO: Frequencies

Statistics		
Infant mortality (deaths per 1000 live births)		
N	Valid	109
	Missing	0
Mean		42,313
Median		27,700
Std. Deviation		38,0792
Minimum		4,0
Maximum		168,0

2 - Através do comando **Descriptives**

- Clicar em **Analyze → Descriptive Statistics → Descriptives**;
- Localizar na janela à esquerda a variável de interesse (por exemplo, **mortalidade infantil (babymort)**) e clicar na →;
- Clicar em **Options**, e assinalar as opções desejadas;
- Clicar em **Continue; OK**;

- j) Os resultados da análise estatística aparecem na janela de resultados (**OUTPUT**), que poderá ser salva, dando origem a um arquivo do tipo **.spo (SPSS output)**.

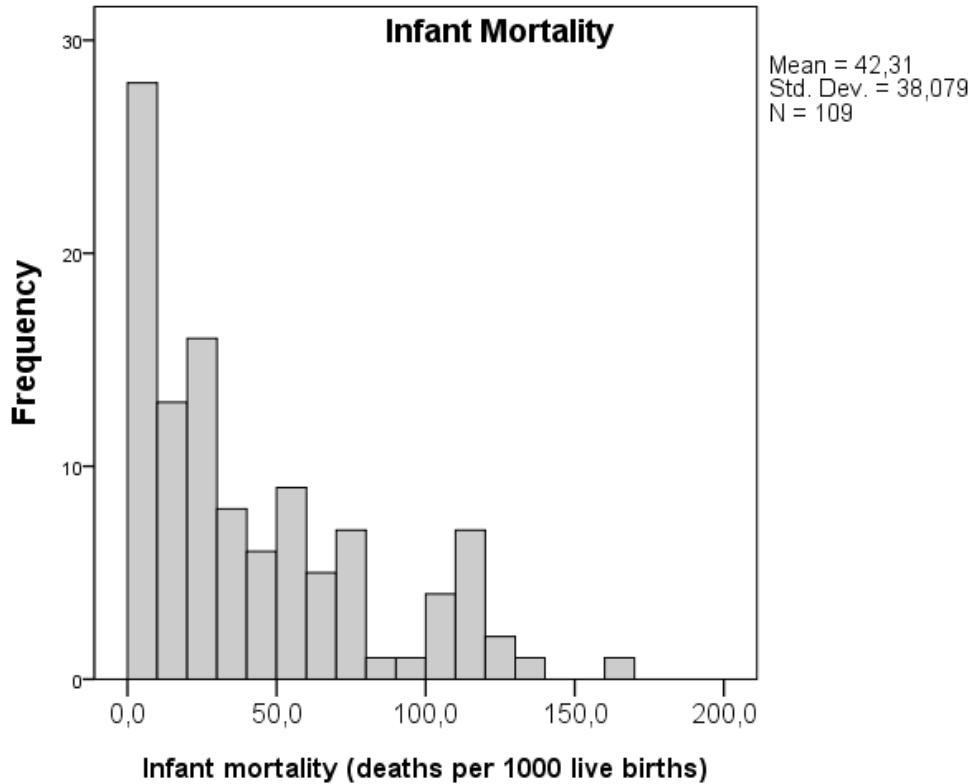
EXEMPLO: Descriptives

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Infant mortality (deaths per 1000 live births)	109	4,0	168,0	42,313	38,0792
Valid N (listwise)	109				

2.3.1.2 - Como obter um histograma

- a) Clicar em **Graphs, Histogram**
- b) Localizar na janela a variável desejada,
- c) Após ter selecionado a variável (neste caso, **babymort**), clicar na →;
- d) Pode-se clicar na opção **Titles** para dar um título ao histograma.
- e) Clicar em **OK**

EXEMPLO: Histograma da variável **Babymort (Infant Mortality)** :



2.3.1.3 - Como obter um box-plot

- Clicar em **Graphs** → **Legacy diaogs** → **Boxplot**;
- Selecione **Simple** → **Summaries of separate variables**;
- Clicar em **Define**;
- Em **Variable** selecionar uma variável quantitativa (por exemplo, **Babymort**);
- Clicar em **OK**.

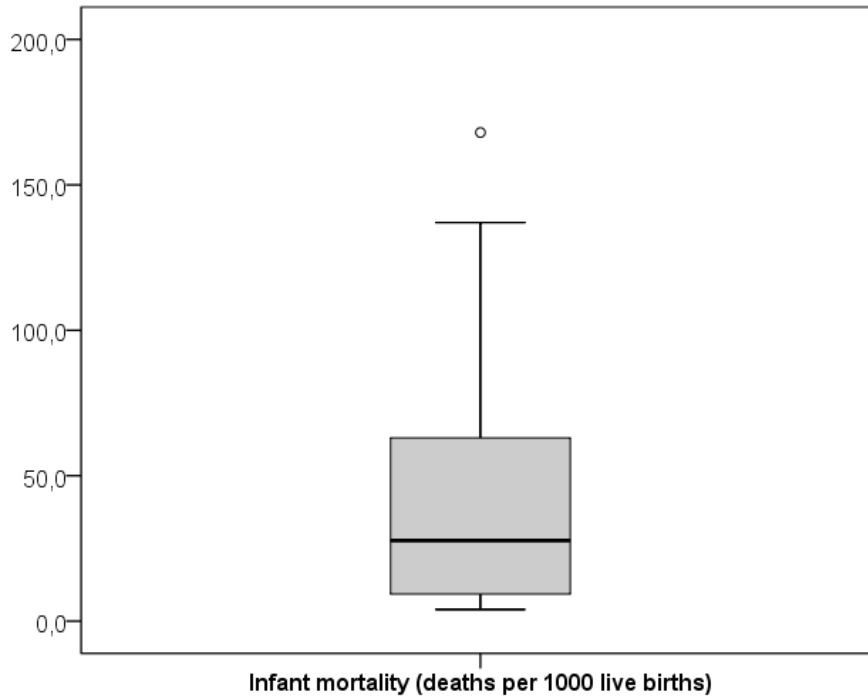
RESULTADO:

Explore

	Case Processing Summary					
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Infant mortality (deaths per 1000 live births)	109	100,0%	0	,0%	109	100,0%

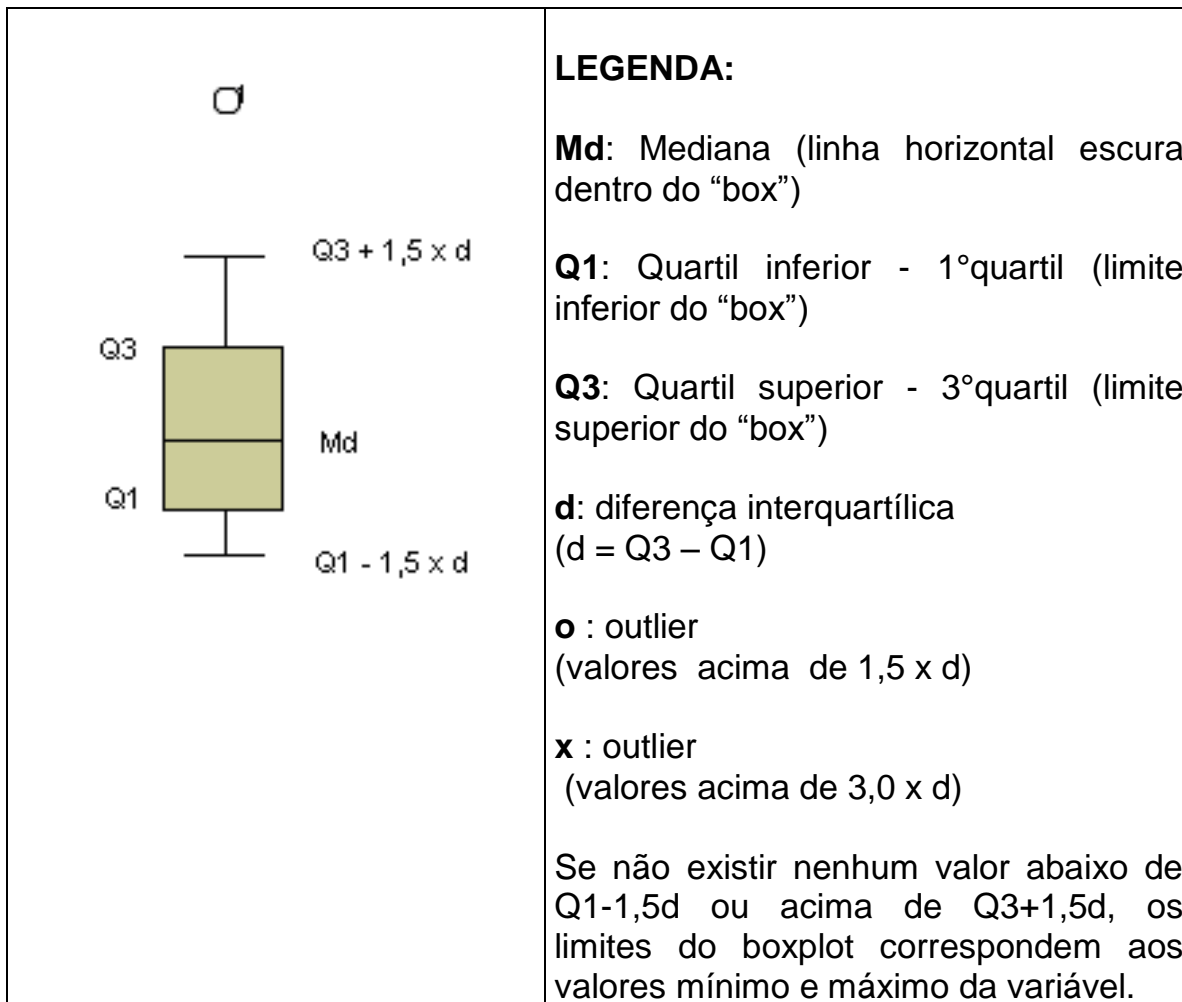
INTERPRETAÇÃO: A tabela acima apresenta o número de casos válidos (**valid**), o número de não respostas (**missing**) e o número total das observações.

Infant mortality (deaths per 1000 live births)



INTERPRETAÇÃO:

Através do Box-plot é possível observar como as variáveis estão distribuídas em relação à homogeneidade dos dados, valores de tendência central, valores máximos e mínimos e valores atípicos, se existirem. Quando a caixinha (box) é muito pequena, significa que os dados estão muito concentrados em torno da mediana, e se a caixinha for grande, significa que os dados são mais heterogêneos.



2.3.2 - Variáveis categóricas (qualitativas)

2.3.2.1 - Como obter a distribuição de freqüências

Para calcular as freqüências procede-se da seguinte forma:

- Clicar em **Analyze → Descriptive Statistics → Frequencies**;
- Selecionar a variável desejada (por exemplo, **region**), clicar na →;
- Selecionar **Display frequency tables**;
- Clicar em **OK**.

RESULTADO:

Statistics

Region or economic group

N	Valid	109
	Missing	0

Region or economic group

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid OECD	21	19,3	19,3	19,3
East Europe	14	12,8	12,8	32,1
Pacific/Asia	17	15,6	15,6	47,7
Africa	19	17,4	17,4	65,1
Middle East	17	15,6	15,6	80,7
Latin America	21	19,3	19,3	100,0
Total	109	100,0	100,0	

INTERPRETAÇÃO: A tabela acima apresenta na 1ª coluna a frequência simples de casos, na 2ª coluna a frequência relativa, na 3ª coluna a frequência relativa excluindo os missings e na 4ª coluna a frequência relativa acumulada.

2.3.2.2 - Como obter gráficos

Para se obterem os diferentes tipos de gráficos disponíveis no programa procede-se da seguinte forma:

- Clicar em **Graphs**, selecionar o gráfico desejado, que ao salvá-lo, dá origem a um arquivo do tipo **.cht** (Chart - arquivo de gráficos).

OBSERVAÇÃO:

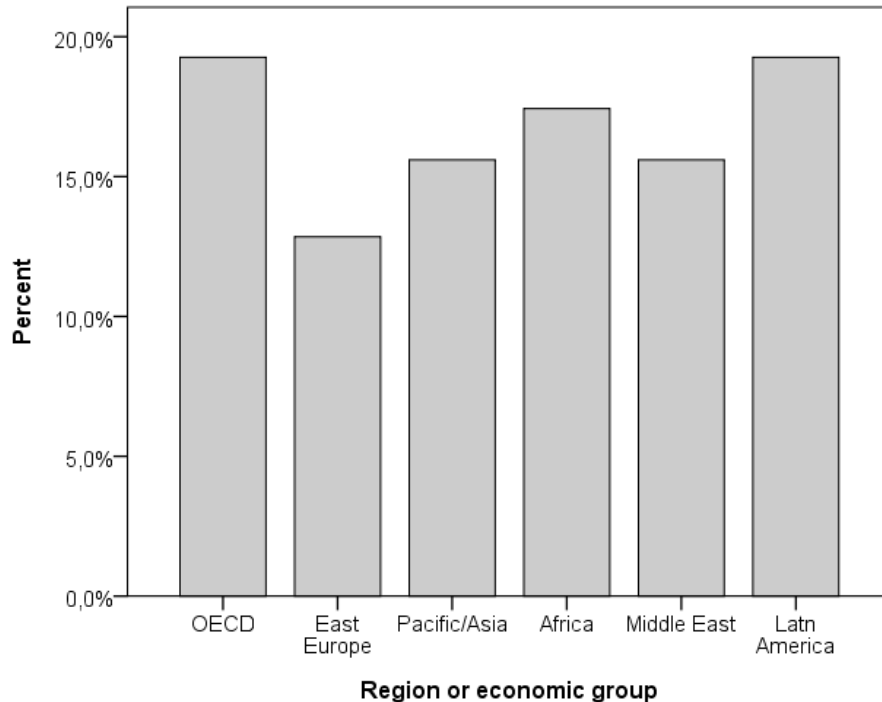
Com variáveis categóricas, o adequado é fazer gráfico de setores (Pie), de Colunas...

2.3.2.2.1 - Como obter gráfico de barras

EXEMPLO: Gráfico de barras para a variável **region**

- a) Clicar em **Graphs, Legacy dialogs** selecionar **Bar**;
- b) Selecionar a opção **Simple** e, em **Data in chart are**, selecionar a opção **Summaries for groups of cases** e clicar em **Define**;
- c) Na opção **Bars represent** selecionar **% of cases**;
- d) Na opção **Category axes** selecionar a variável **region**, clicar **OK**.

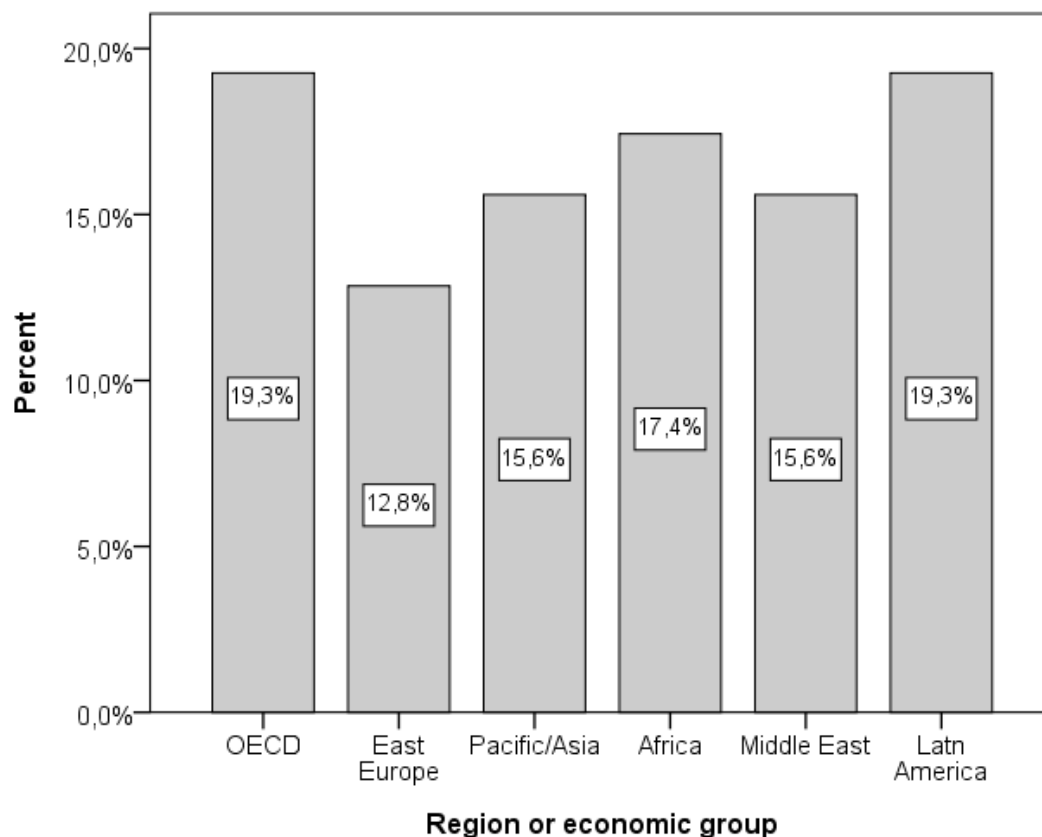
RESULTADO:



Para colocar o valor percentual de cada categoria no gráfico:

- a) Clicar duas vezes no gráfico;
- b) Abre o **SPSS Chart Editor**; clicar em **elements/show data labels**;
- c) Na janela **properties**, em **data value labels**, selecionar **percents**;
- d) Clicar **OK**.

RESULTADO:

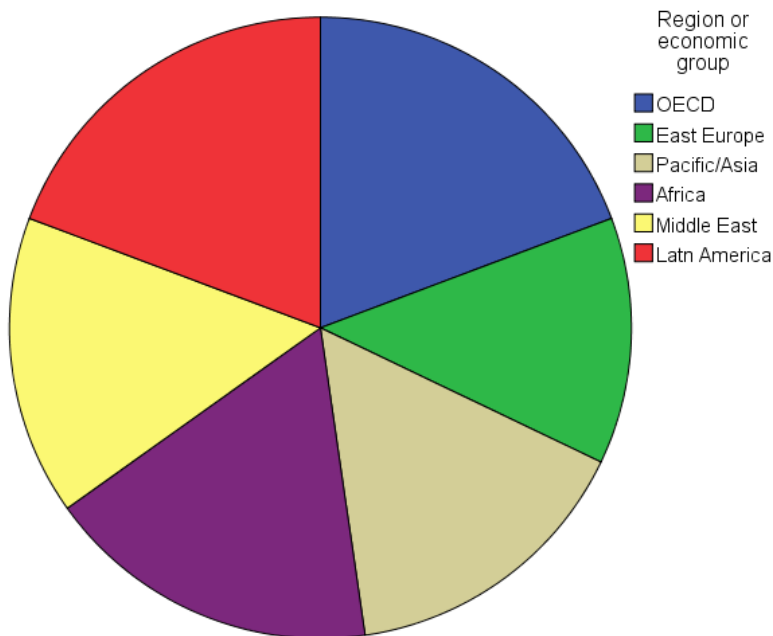


2.3.2.2.2 - Como obter gráfico de setores (pizza)

EXEMPLO: Gráfico de Setores (Pie) para a variável **region**

- Clicar em **Graphs**, selecionar **Pie**;
- Selecionar a opção **Summaries for groups of cases** e clicar em **Define**;
- Na opção **Define Slices by** selecionar a variável **region**.

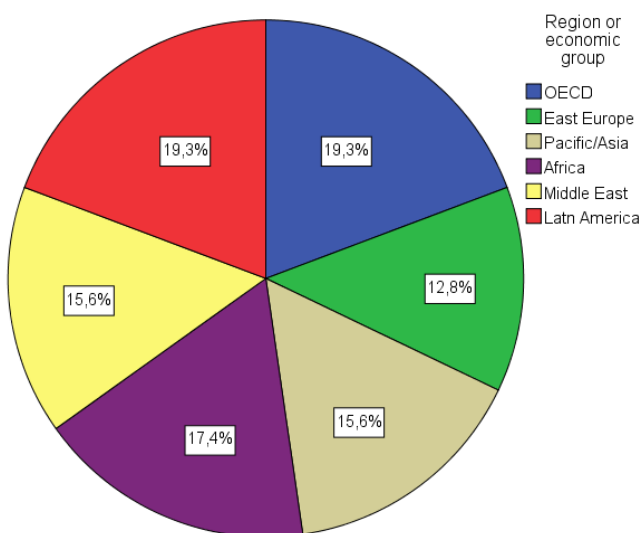
RESULTADO:



Para colocar no gráfico o valor percentual de cada categoria:

- Clicar duas vezes no gráfico;
- Abrirá a janela de edição de gráficos (**SPSS Chart Editor**);
- Nesta janela, clicar em **elements/show data labels**;
- Abrirá uma nova janela **“properties”**;
- Ir em **“data value labels”**, selecionar a opção **percents**;
- Clicar em **OK**.

RESULTADO:



2.4 - Como exportar o output para word

Para se exportar o output para Word procede-se da seguinte forma:

- a) No output, clicar em **File, Export**, abre-se a janela **Export output**;
- b) Em **Objects to export** é possível escolher exportar tudo, apenas uma seleção, ou tudo que está visível;
- c) Em **Document, Type**, escolha o tipo de arquivo para o qual se deseja exportar os resultados (neste caso **WORD**);
- d) Em **File name** digite o local e o nome do arquivo que será salvo;
- e) Clique **OK**.

No local escolhido será salvo um novo arquivo com a terminação **.doc**.

Para exportar para outros programas procede-se da mesma maneira, apenas mudando em **Type** para o tipo de sua preferência.

3 - CATEGORIZAÇÃO DE VARIÁVEIS

3.1 - Como categorizar uma variável quantitativa

Para exemplificar, usaremos uma variável categorizada utilizando quartis. Os quartis são pontos de corte na escala da variável de tal forma que cada grupo formado a partir destes pontos de corte terá um quarto dos casos, ou seja, 25% do tamanho total da amostra.

Os passos necessários para categorizar uma variável utilizando os “quartis” são os seguintes:

1ª OPÇÃO: utilizando o comando **Recode Into Different Variables**

1. Calcular os quartis da variável em questão, neste caso, População (**populatn**), do seguinte modo:
 - a) Clicar em **Analyze → Descriptive Statistics → Frequencies**;
 - b) Selecionar a variável que se deseja categorizar na janela esquerda e clicar →;
 - c) Retirar a opção de **Display Frequency Tables**, a fim de que não venha listada a totalidade de casos da variável (no estudo em pauta o número é de 109 casos);
 - d) Clicar em **Statistics** e assinalar **Quartiles**;
 - e) Clicar em **Continue**; **OK**.

RESULTADOS:

Frequencies

Statistics

Population in thousands

N	Valid	109
	Missing	0
Percentiles	25	5000,00
	50	10400,00
	75	37100,00

2. Criar uma variável com 4 categorias, definidas pelos quartis, da seguinte maneira:

<u>Categoria</u>	<u>Intervalo de valores</u>
1	Mínimo até 5000,00
2	5001,00 até 10400,00
3	10401,00 até 37100,00
4	37101,00 até o valor máximo da variável

Para categorizar a variável **populatn**, usando os limites dados pelos quartis procede-se da seguinte forma:

- a) Clicar em **Transform → Recode Into Different Variables**;
- b) Localizar, na janela à esquerda, a variável a ser categorizada (**populatn**) e clicar na →;
- c) Digitar um novo nome para a variável de saída (**Output Variable**) - por exemplo POPREC - e clicar em **Change**;
- d) Clicar em **Old and New Values**;
- e) Clicar em **Range (lowest through)** e digitar o valor obtido para o primeiro quartil, no caso 5000;
- f) Em **New Value**, digitar **1**;
- g) Clicar em **ADD**;
- h) Assinalar **Range**, colocando: 5001 até (Through) 10400 (segundo quartil);
- i) Na opção **New Value**, digitar **2**;
- j) Clicar em **ADD**;
- k) Assinalar **Range**, 10401 até (Through) o terceiro quartil 37100;
- l) Na opção **New Value**, digitar **3**;
- m) Clicar em **ADD**;

- n) Clicar em **Range (Through Highest)** e digitar o valor imediatamente superior ao 3º quartil, no caso 37101;
- o) Na opção **New Value**, digitar 4;
- p) Clicar em **ADD; Continue, OK.**

A nova variável **POPREC** corresponde à variável **populatn** categorizada, sendo esta automaticamente incluída no final banco de dados que estamos utilizando (**World95.sav** - Arquivo de dados).

OBS: você pode dar nome aos níveis desta nova variável. Para isso, proceda da seguinte forma:

- a) No banco de dados, clique **2 vezes** sobre a nova variável **Poprec**, obtendo-se a planilha **Variable View**.
- b) Na linha da variável **Poprec**, clicar em **Values**. Abre-se uma nova janela - **Value Labels**;
- c) Em **Value**, digitar 1;
- d) Em **Label**, digitar pop. inferior a 5000,00;
- e) Clicar em **ADD**;
- f) Proceder-se da mesma forma para os demais níveis de categorização: 2, 3 e 4;
- g) Clicar em **Continue; OK.**

2ª OPÇÃO: utilizando o comando **VISUAL BINNING**

- a) Clicar em **Transform → Recode → Visual Binning**;
- b) Localizar, na janela à esquerda, a variável a ser categorizada (**populatn**) e clicar na →;
- c) Clicar **continue**;
- d) Na caixa **Binned variable** incluir o novo nome da variável em quartis (ex: **popQuartis**);
- e) Em **make cutpoints**, escolher a opção **equal percentiles based on scanned cases**;
- f) Preencher uma das seguintes opções:
 - a. Number of cutpoints: nº de grupos - 1, no caso, 4 - 1 = 3; OU
 - b. Width (%): % de casos em cada grupo, no caso como são 4 grupos de mesmo tamanho, 100%/4 = 25%
- g) Clicar **apply**;
- h) Clicar em **make labels**;
- i) Clicar **OK.**

3.2 - Como criar uma variável a partir de datas

Para criar uma variável, p.ex. Idade, a partir da variável data de nascimento (**datanasc**) e da data atual (**hoje**), utilizamos a função DATEDIFF (datetime2,datetime1,"unit").

I. Para criar a variável **hoje**:

- a) Criar uma coluna no banco de dados chamada **hoje** com formato de data (não é necessário digitar um por um);
- b) Clicar no menu **Transform → Compute variable**;
- c) Em **Target Variable** digite o nome da variável **hoje** para preencher todos os casos com a data de hoje;
- d) Na janela **Functions** selecionar a opção DATE.Dmy(Day,month,year) e clicar na ↑;
- e) Na janela **Numeric Expression** digite o dia,mês,ano dentro do parênteses (separado por vírgulas);
- f) Clicar em **OK**;

II. Para calcular a **idade**:

- a) Criar a coluna **datanasc** com formato de data e digitar as datas de nascimento dos alunos;
- b) Selecionar **Transform, Compute variable**;
- c) Em **Target Variable** digite o nome da nova variável, p. exemplo **idade**;
- d) Na janela **Functions** selecionar a opção DATEDIFF(datetime2,datetime1,"unit") e clicar na ↑;
- e) Localizar na janela abaixo de **Target Variable** a variável **hoje** e clicar na → (a variável selecionada deve ficar entre os parênteses, no lugar de "datetime2");
- f) Na mesma janela, localizar a variável **datanasc** (a variável selecionada deve ficar entre os parênteses, no lugar de "datetime1");
- g) No lugar de "units" digitar a unidade de tempo que você deseja ("years" para anos, "months" para meses, "days" para dias, etc);
- h) Clicar em **OK**.

Exercício:

No banco hosp.xls, calcular o tempo de internação (em dias), a partir das variáveis DataInt (data de internação) e DataAlta (data da alta).
Fazer a análise descritiva apropriada.

3.3 - Como criar uma variável através da combinação de outras duas

Nesta seção, será utilizado o banco **GSS93.sav**, que também se encontra disponível junto com o programa SPSS.

Para criar uma variável a partir da combinação de outras duas, como por exemplo, combinar a variável sexo (**sex**) e a variável raça (**race**) utilizaremos o seguinte procedimento para criar a variável **SEXRACE**.

Sabendo que a variável **SEX** é categorizada da seguinte forma:

1-Male e 2-Female

e a variável **RACE** é categorizada da seguinte forma:

1- White, 2-Black e 3-Other

pode-se criar a variável **SEXRACE** com as seguintes categorias:

**1- White Male,
2- White Female,
3- Black Male
4- Black Female
5- Other Male
6- Other Female**

Então se procede da seguinte forma:

- a) Selecionar **Transform, Compute variable**;
- b) Em **Target Variable** digitar o nome da nova variável, por exemplo **SEXRACE**;
- c) Na janela **Numeric Expression** digitar **1**;
- d) Clicar em **If**;
- e) Selecionar a opção **Include if case satisfies condition**;
- f) Localizar na janela abaixo de **Include if case satisfies condition** a variável desejada,
- g) Após ter selecionado a variável (neste caso, **sex**), clicar na **→**;
- h) Digitar **=1 &** na janela ao lado da variável **sex**;
- i) Selecionar na janela ao lado a variável **race** e clicar na **→**;
- j) Na janela ao lado da variável **race** digitar **=1**;
- k) Após esse procedimento a expressão na janela deve ser a seguinte:
sex=1 & race=1;

- l) Clicar em **Continue** e **OK** ou **Paste**. A variável **SEXRACE** aparecerá no final do banco de dados.
- m) Para criar as demais categorias da variável **SEXRACE** procede-se de maneira análoga, alterando o código na janela **Numeric Expression** para 2, 3, 4, 5 e 6 e a expressão da janela **Include if case satisfies condition**.

Resultado das janelas:

Numeric Expression	Include if case satisfies condition
1	sex=1 & race=1
2	sex=2 & race=1
3	sex=1 & race=2
4	sex=2 & race=2
5	sex=1 & race=3
6	sex=2 & race=3

3.4 - Como criar uma variável através de um cálculo com outras variáveis

Nesta seção, será utilizado o banco **imc.sav**.

Para criar a variável **IMC**, a partir da combinação de outras duas (peso e altura), cuja fórmula é a seguinte: **IMC = peso(kg) / altura(m)²**, utilizaremos o seguinte procedimento para criar a variável **IMC**.

- a) Selecionar **Transform, Compute variable**;
- b) Em **Target Variable** digite o nome da variável **IMC**;
- c) Localizar na janela abaixo de **Target Variable** a variável **peso** e clicar na →;
- d) No “teclado virtual” localizado no centro da janela, clicar em “/” (função de divisão);
- e) Localizar na janela abaixo de **Target Variable** a variável **altura** e clicar na →;
- f) No “teclado virtual” localizado no centro da janela, clicar em “**” (função de exponenciação) e “2”;
- g) Clicar em **OK** ou **PASTE**.

Salve o banco de dados, pois o mesmo será utilizado em outro exercício.

Exercício:

- 1) Categorize a variável **IMC** (criada no banco **IMC**) nas seguintes categorias:

1: < 25

2: de 25 a 29,9 (sobrepeso)

3: ≥ 30 (obesidade)

- Realizar a análise descritiva adequada para a variável IMC em categorias.

2) Com o banco de dados **HAS.SAV**, criar a variável HAS (hipertensão arterial sistêmica) que é definida pela **PAS ≥ 140** ou **PAD ≥ 90**.

- Realizar a análise descritiva adequada para as variáveis PAS, PAD e HAS.

3.5 - Como criar uma variável a partir da contagem de outras

Muitas vezes nos deparamos com situações onde é necessário contar em quantas variáveis apareceu algum código. Por exemplo, quantas medicações o indivíduo toma, ou quantas comorbidades o paciente possui, ou ainda, quantas variáveis possuem dados perdidos (*missing values*). No comando **Count values within cases** (Menu **Transform**) é possível criarmos uma variável cuja informação seja uma contagem a partir de outras.

Para o nosso exemplo, utilizaremos o banco **GSS93.SAV**. A ideia é contar quantos estilos musicais (dentre os estilos: country, blues, clássica, jazz, rap e heavy metal) as pessoas gostam. Originalmente, as variáveis possuem 5 categorias válidas: 1-gosto muito, 2-gosto, 3-tanto faz, 4-não gosto, 5-não gosto muito. Para o exemplo, iremos considerar que as categorias 1 e 2 indicam quem gosta do estilo. Os passos são os seguintes:

- a) Clicar no menu **Transform**, e depois em **Count values within cases**;
- b) Na janela **Target variables**, digitar o nome da variável que vai armazenar a nossa contagem: **nestilosmusicais**;
- c) Na janela **Target label**, digitar um rótulo para a variável: **Quantidade de estilos musicais**;
- d) Na janela das variáveis, abaixo de **Target variables**, selecionar as variáveis que interessam: country, blues, classicl, jazz, rap, hvymetal, clicando em → para incluir na janela **Variables**;
- e) Clicar em **Define values**;
- f) Clicar em **Range** no quadro à esquerda e digitar **1** (gosto muito) na caixa de cima e **2** (gosto) na caixa de baixo;
- g) Clicar em **Add**;
- h) Clicar em **Continue**;
- i) Clicar em **Ok** ou **Paste**;

A tabela de frequências deve ficar assim:

Quantidade de estilos musicais					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,00	95	6,3	6,3	6,3
	1,00	345	23,0	23,0	29,3
	2,00	383	25,5	25,5	54,9
	3,00	375	25,0	25,0	79,9
	4,00	229	15,3	15,3	95,1
	5,00	54	3,6	3,6	98,7
	6,00	19	1,3	1,3	100,0
Total		1500	100,0	100,0	

Exercício:

- Com o banco **stroke.sav**, crie uma variável (p.ex: numeroFR) que seja a quantidade de fatores de risco (sedentarismo, DM, obesidade, fumo e hipercolesterolemia) de cada paciente.

- Realizar a análise descritiva adequada para a variável numeroFR.

4. MANIPULAÇÃO DE DADOS

4.1 - Ordenação de casos: Sort Cases

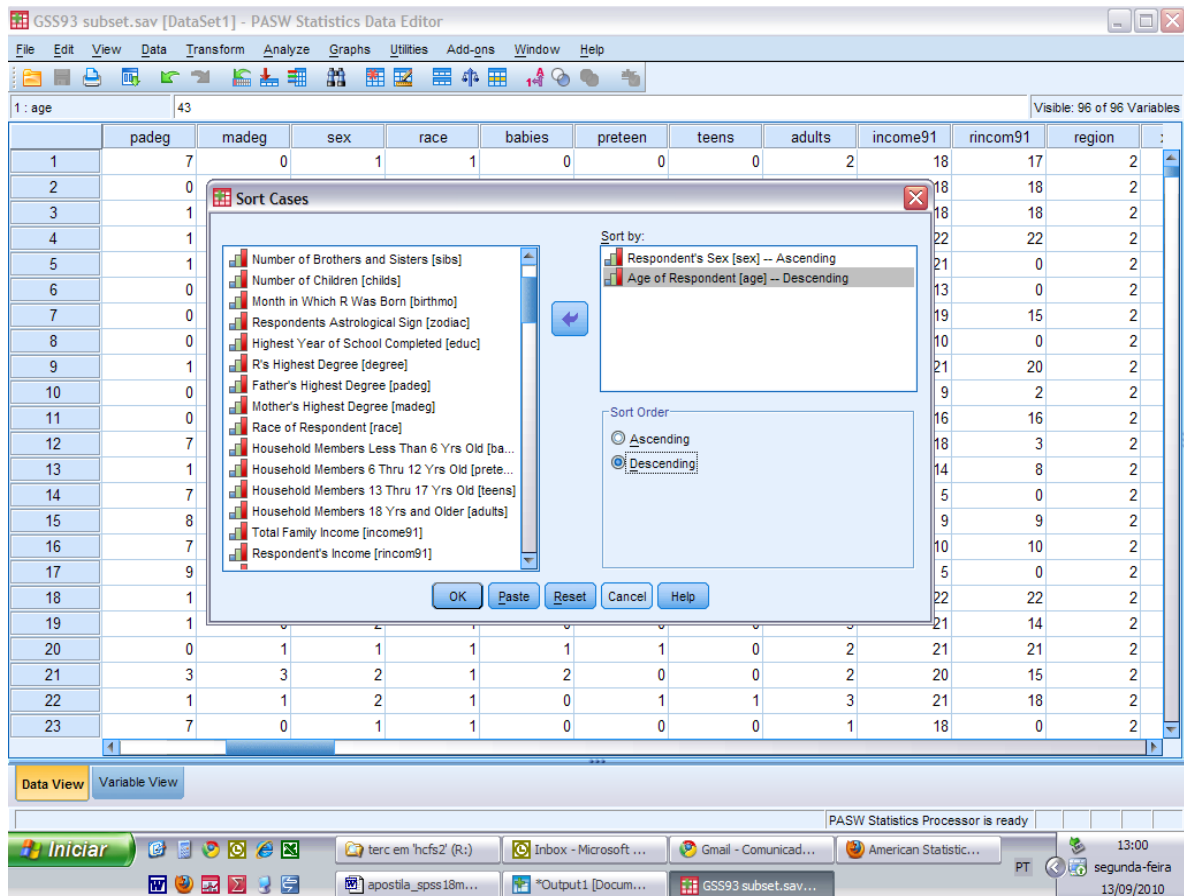
Uma das necessidades na hora da manipulação dos dados no dia-a-dia é a ordenação dos casos segundo uma ou mais variáveis. Para fazer isso no SPSS for Windows, usar o procedimento **Sort Cases** presente no menu **Data**.

Após clicar em **Data** opção **Sort Cases**, uma janela é aberta. Movemos para o quadro **Sort by** a variável segundo a qual o arquivo deve ser ordenado. Podemos mover para esse quadro mais do que uma variável. Nesse caso, o arquivo é ordenado, em primeiro lugar, pelos valores da primeira variável no quadro e, em segundo lugar, pela segunda variável no quadro; a segunda ordenação é feita para os valores comuns da primeira variável.

Podemos escolher também entre ordem crescente ou decrescente de ordenação para cada uma das variáveis. Isso é feito através do quadro **Sort Order** opções **Descending** (decrescente) ou **Ascending** (crescente).

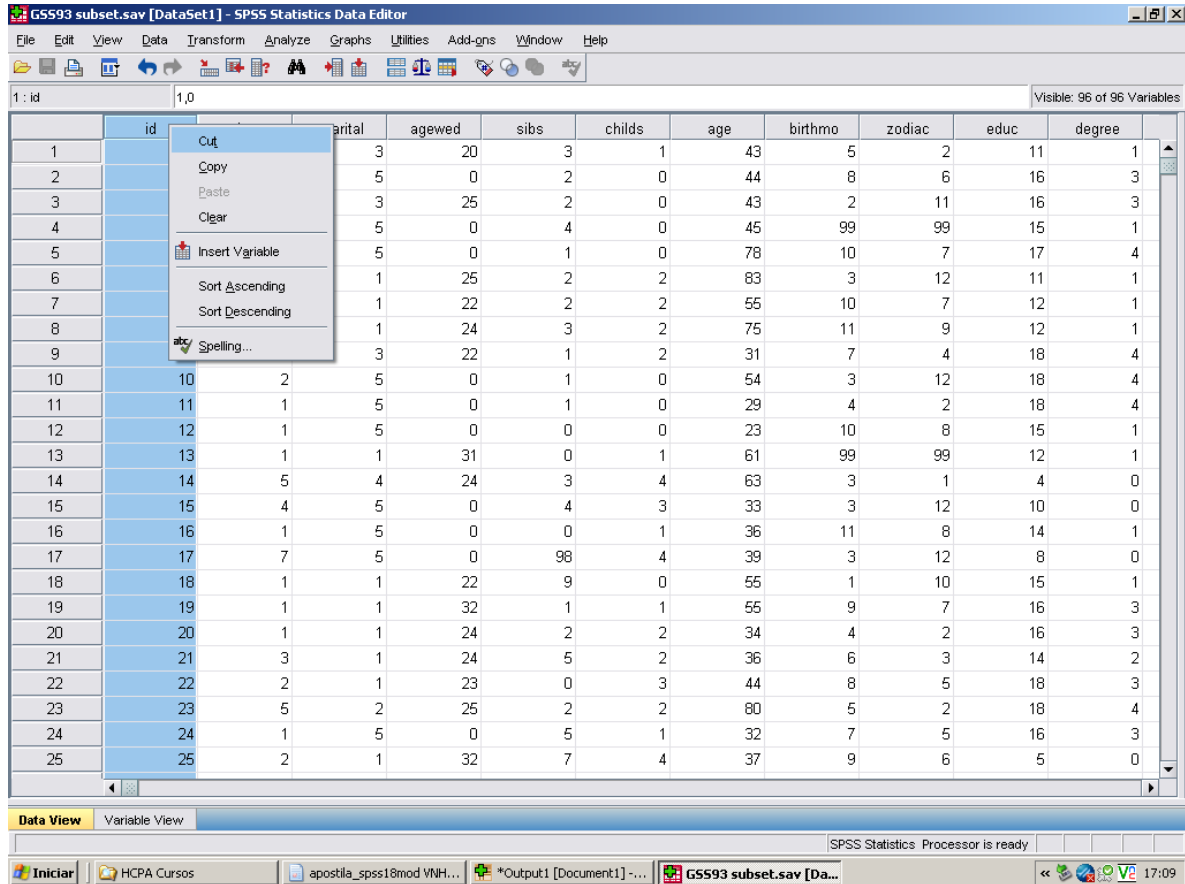
Vamos fazer uma ordenação segundo **age** (ordem decrescente) dentro dos códigos de **sex** (ordem crescente). Para isso procede-se da seguinte forma:

- Mover a variável **sex** para ao quadro **Sort Cases** e escolher a opção **Ascending** no quadro **Sort Order**.
- Mover em seguida a variável **age** para o quadro **Sort Cases** e escolher a opção **Descending** no quadro **Sort Order**.
- Clicar **OK** para executar a ordenação.



Note que após a execução deste comando a posição dos indivíduos nas linhas fica completamente alterada, pois o indivíduo na linha 1 do banco de dados após ordenado pode não ser o primeiro caso digitado. Para que esta informação não se perca é essencial que exista uma variável com o número do indivíduo.

Outro jeito de ordenar os casos é clicando com o botão direito do mouse no nome da variável (retângulo cinza). Um menu com alguns comandos é aberto. Vamos experimentar clicando na variável “ID”, colocar o banco de dados em ordem novamente.



4.2 - Seleção de casos: *Select Cases*

Uma outra necessidade é a seleção (temporária ou permanente) de parte do arquivo de dados. Digamos que estamos interessados em estudar um segmento específico da amostra. O SPSS possui várias formas de seleção de dados. Falaremos nessa seção de todas elas, mas discutiremos detalhadamente a mais usada de todas. Para maiores detalhes sobre as demais formas de seleção, recomenda-se que o leitor use o manual do *SPSS for Windows*.

Para fazer qualquer tipo de seleção, devemos clicar o menu **Data** opção **Select Cases**.

No quadro central **Select**, estão presentes cinco opções diferentes para seleção:

- **All cases** – opção usada por *default*, utiliza todas as observações do banco de dados;
- **If condition is satisfied** – através dessa opção, podemos definir expressões condicionais para seleção de casos;
- **Random sample of cases** – podemos selecionar uma porcentagem ou número exato de casos; a seleção é feita aleatoriamente;
- **Based on time or case range** – usamos essa opção quando estamos interessados em selecionar uma faixa específica de valores, por exemplo, os casos do número 100 ao 200; também utilizada para fazer seleções baseadas em datas;
- **User filter variable** – uma variável é escolhida no banco de dados e usada como filtro; todos os casos para os quais a variável filtro assume o valor 0 não serão selecionados.

Você tem duas opções para o tratamento dos casos que não serão selecionados. É através do quadro **Unselected Cases Are** que podemos fazer a escolha:

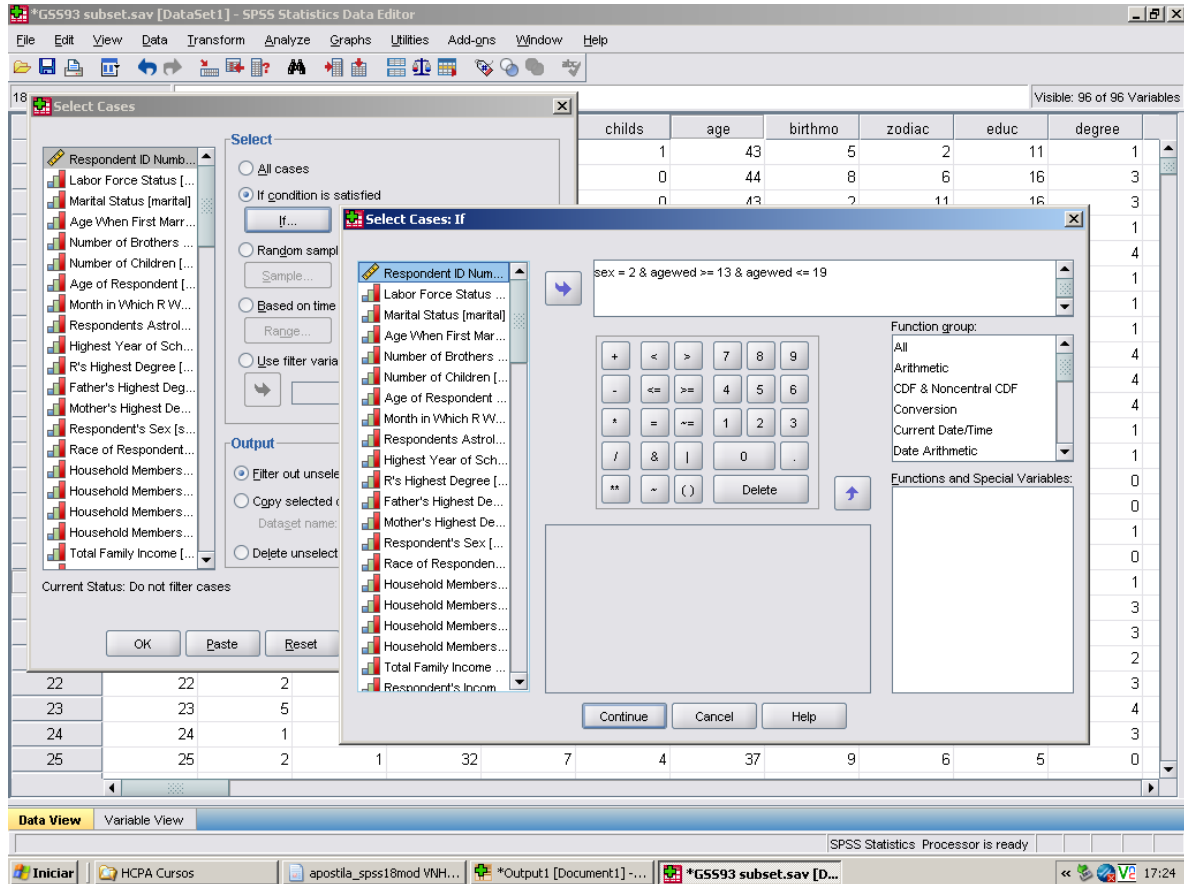
- **Filter...** – os casos (linhas) que não são selecionados não são incluídos nas análises posteriores, porém, permanecem na janela de dados; caso você mude de idéia e queira usar os casos não selecionados na mesma sessão do SPSS, basta “desligar” o filtro;
- **Copy...** - Nesta opção, o SPSS cria uma nova janela de dados com os casos selecionados. O nome da janela deve ser fornecido pelo usuário na lacuna “Dataset name”.
- **Deleted** – os casos (linhas) não selecionados são apagados da janela de dados; caso você mude de idéia e queira usar os casos não selecionados, você deverá ler novamente o arquivo de dados original. Neste caso deve-se tomar o cuidado de salvar o banco de dados com outro nome (**File...Save As**).

Suponha que estamos interessados em selecionar as mulheres que se casaram com menos de 20 anos (variável **agewed**) no banco de dados **GSS93.sav**. A função condicional para seleção nesse caso é dada por:

$$\text{sex} = 2 \ \& \ \text{agewed} \geq 13 \ \& \ \text{agewed} \leq 19$$

Portanto, o tipo de seleção de dados que faremos deve possibilitar a criação de sentenças matemáticas lógicas para seleção dos casos. Para isso, clicamos em **If condition is satisfied** e entramos no retângulo **If...**

Através da janela que é aberta, usamos o retângulo superior para escrever uma função lógica na qual a seleção vai ser baseada. Para a construção da função, podemos usar todas as variáveis que estão no quadro à esquerda e as funções disponíveis no quadro inferior direito.



Uma vez escrita a função que determina a regra de seleção dos casos, clique Continue e você voltará à janela anterior. No quadro inferior (Unselected cases are), vamos optar pelo modo Filtered (ou seja, os casos não selecionados permanecem na tela de dados, porém, não serão utilizados em análises futuras) e clicar OK.

Você pode perceber que, depois de feita a seleção, a janela de dados sofre algumas alterações. As linhas (casos) que não foram selecionadas apresentam uma listra no canto esquerdo da janela de dados. A barra localizada na parte inferior da janela apresenta a mensagem Filter On. Além disso, uma coluna de nome **filter_\$** é adicionada à janela de dados. Essa nova coluna apresenta valor 0 para as linhas que não foram selecionadas e valor 1 para as linhas que foram selecionadas.

Apesar de você conseguir ver os casos que não foram selecionados, qualquer análise efetuada daí para frente não leva em conta esses casos.

Podemos mudar de idéia e querer usar todas as observações para o cálculo das estatísticas. Temos duas maneiras de cancelar a seleção de casos, se a opção **Filtered** foi usada para efetuar a seleção. A primeira delas é ativar a opção **All Cases** da janela de seleção de casos (menu **Select Cases**) e clicar **OK**. A Segunda maneira é deletar a coluna **filter_\$** da janela de dados.

	id	wrkstat	marital	agewed	sibs	childs	age	birthmo	zodiac	educ	degree
412	437	3	1	20	3	2	47	3	12	14	2
413	438	2	5	0	1	0	21	11	8	13	1
414	439	7	1	31	4	2	59	8	5	12	1
415	441	1	1	18	4	1	23	9	6	11	0
416	442	5	1	23	10	2	76	9	6	7	0
417	443	1	1	19	2	2	40	9	6	12	1
418	444	7	1	16	9	4	38	9	7	12	1
419	445	7	2	18	11	2	66	1	10	12	1
420	446	5	2	23	9	2	73	5	2	7	0
421	447	5	5	0	9	0	80	7	5	8	0
422	448	7	3	16	3	1	57	7	4	6	0
423	449	5	2	38	0	0	78	3	12	11	0
424	450	5	5	0	6	0	66	12	9	12	1
425	451	5	2	25	2	2	88	2	11	12	1
426	452	7	1	21	4	5	58	7	4	12	1
427	453	4	5	0	3	0	39	5	2	12	1
428	454	1	3	18	2	0	31	9	7	12	1
429	455	1	1	22	1	3	51	8	5	16	3
430	456	1	1	22	2	3	37	10	7	17	3
431	457	2	1	18	2	2	32	2	12	12	1
432	458	5	2	22	3	1	73	3	12	8	0
433	459	1	1	22	1	2	47	1	11	16	3
434	460	7	2	27	2	4	71	1	11	16	3

4.3 - Divisão de análises por grupos: *Split File*

Vamos supor que, após uma série de análises, chegamos à conclusão de que o comportamento dos homens e das mulheres é completamente diferente com relação às preferências para horas de lazer. Não faz sentido, portanto, apresentar a análise do questionário de opinião sobre lazer com os homens e mulheres juntos. No fundo, o que pretendemos fazer, daqui para frente, são duas análises idênticas, uma para cada sexo.

Para esse tipo de situação, podemos utilizar o procedimento **Split File**, presente no menu **Data**. Por default sempre analisamos todos os casos juntos, sem separação por grupos. Por esse motivo, a opção selecionada na janela é **Analyze all cases**. Para repetir a análise para as categorias de uma determinada variável, clicamos em **Compare groups** ou **Organize output by groups**, e então o quadro **Groups Based on** fica disponível.

Moveremos para esse quadro a variável (ou variáveis) que definirão os grupos para os quais a análise deve ser repetida. Se mais do que uma variável for selecionada, os grupos serão definidos pela combinação das categorias de todas as variáveis. Podemos ainda escolher se o banco de dados deve ser ordenado pela variável que definirá os grupos (**Sort the file by group variables**) ou se o banco de dados já está ordenado pela variável que definirá os grupos (**File is already sorted**). Se o banco não estiver ordenado corretamente, o comando não será executado de forma adequada.

No nosso caso, selecionamos a variável **gender** e a movemos para o quadro **Groups Based on** e clicamos **OK**. A única mudança que acontece na janela de dados é a mensagem **Split File On** na barra inferior, ou a ordenação dos casos pela variável que definiu os grupos, caso o banco de dados ainda não estivesse ordenado. Porém, qualquer análise ou gráfico feitos de agora em diante vão gerar dois resultados, uma para os homens e outro para as mulheres.

Para exemplificar, vamos pedir uma análise descritiva (média, desvio-padrão, quartis, mínimo e máximo) da variável **age** separada por **sexo** do banco de dados **GSS93.sav**. A partir dos passos anteriores para dividir a amostra de acordo com o sexo, clicamos em **Analyze**, depois em **Descriptives** e depois em **Frequencies**. Selecionamos a variável **age** e, em **Statistics**, selecionamos as opções desejadas (quartiles, mean, std deviation, minimum, maximum). Após clicamos em **continue** e em **OK**.

Note que os resultados são apresentados em dois blocos, o primeiro para o sexo masculino e o segundo para o sexo feminino se a opção escolhida foi **Organize output by groups**.

Podemos mudar de idéia e querer usar todas as observações para o cálculo das estatísticas. Para cancelar o procedimento **Split File** basta ativar a opção **Analyze all cases** presente na janela de definição da opção **Split File** menu **Data**.

4.4 - Juntar bancos de dados: Merge Files

Muitas vezes os dados se encontram em diferentes bancos de dados e em algum momento precisamos uni-los para poder cruzar alguma variável que está num banco com outra de outro banco ou ainda unificar em um mesmo banco, pacientes de diferentes bancos de dados (com as mesmas variáveis).

No comando **Merge files**, no menu **Data**, o SPSS permite juntar bancos de dados que tenham as mesmas variáveis com diferentes sujeitos ou bancos que tenham variáveis diferentes dos mesmos sujeitos.

4.4.1 - Juntando bancos com diferentes casos

Os bancos de dados **GSS93 p1 casos.sav** e **GSS93 p2 casos.sav** possuem diferentes indivíduos cada um. Para juntar os bancos é IMPRESCINDÍVEL que eles possuam as mesmas variáveis (inclusive nos mesmos formatos). O procedimento é o seguinte.

- a) Abrir o banco de dados **GSS93 p1 casos.sav** (**File -> Open -> data**);
- b) Clicar no menu **Data -> Merge files -> Add cases**;
- c) Na janela **Add cases**, você deve escolher:
 - a. **An open dataset**: se o outro banco de dados já estiver aberto em alguma janela;
 - b. **An external PASW statistics data file**: se o banco estiver fechado, lembrando que o mesmo já deve estar em formato do SPSS, ou seja, ter sido salvo como .sav.
- d) Clicar em **continue**;
- e) As variáveis com os mesmos nome nos dois bancos de dados são automaticamente incluídas na janela **Variables in New Active Dataset**;
- f) Se você quiser identificar qual o banco de origem de cada caso no novo banco, clicar em **Indicate case source as variable**;
- g) Na janela **Unpaired Variables**, variáveis seguidas de * estão no primeiro banco de dados e aquelas seguidas de + estão no banco novo;
- h) Clicar em **OK** ou **Paste**.

4.4.2 - Juntando bancos com diferentes variáveis

Os bancos de dados **GSS93 p1 var.sav** e **GSS93 p2 var.sav** possuem diferentes variáveis: no primeiro estão as variáveis demográficas

dos indivíduos e no segundo as respostas de um questionário de opiniões e preferências. Para juntar os bancos é IMPRESCINDÍVEL que eles NÃO possuam as mesmas variáveis e que os casos estejam ordenados. O procedimento é o seguinte.

- a) Abrir o banco de dados **GSS93 p1 var.sav (File -> Open -> data);**
- b) Ordenar o banco de dados pela variável **Id**: clicar com o botão da direita e em **Sort Ascending**;
- c) Clicar no menu **Data -> Merge files -> Add variables**;
- d) Na janela **Add variables**, você deve escolher:
 - a. **An open dataset**: se o outro banco de dados já estiver aberto em alguma janela e clique no **Dataset** que você deseja mesclar;
 - b. **An external PASW statistics data file**: se o banco estiver fechado, lembrando que o mesmo já deve estar em formato do SPSS, ou seja, ter sido salvo como .sav e clique em **Browse** para localizar o arquivo de dados.
- i) Clicar em **continue**;
- j) Clicar em **Match cases on key variables in sorted files**; ao clicar aqui, você deve optar:
 - a. **Both files provide cases**: se você sabe que cada banco possui uma linha para cada indivíduo da amostra;
 - b. **Non-active ou active dataset is keyed table**: se um dos bancos possuir mais de um registro (linha) por paciente, que devem estar identificados pela mesma variável chave (key variable);
- k) Clicar na variável que identifica o número do caso (deve estar na janela **Excluded Variables**), que no nosso exemplo é a variável **id** e depois clicar em → (a variável deverá aparecer na caixa **Key variables**);
- l) Na janela **Excluded Variables**, variáveis seguidas de * estão no primeiro banco de dados e aquelas seguidas de + estão no banco novo;
- m) Clicar em **OK** ou **Paste**.

4.5 - Manipulação de arquivos

Para retornar aos arquivos:

- *.sav (arquivo de dados)
- *.cht (arquivo com cada gráfico realizado)
- *.spo (arquivo de resultados)

Procede-se da seguinte maneira:

- a) Clica-se na opção de menu “Window”;
- b) Seleciona-se a janela de saída desejada que consta na lista de arquivos abertos ou disponíveis, clicando uma vez sobre sua indicação.

4.6 - Edição do arquivo de resultados (*.spo): como apagar análises não desejadas no arquivo de resultados

Quando inadvertidamente realiza-se um procedimento não desejado, para corrigir o equívoco, procede-se da seguinte forma:

- a) Clica-se no menu **Edit → Select → Output Block** a partir deste momento será selecionada a última saída executada, o que vai dar origem a uma “tarja preta”;
- b) Aperta-se o botão “Delete”, tornando sem efeito o último procedimento efetuado. Também podem-se apagar outros blocos de resultados, bastando para tal colocar o cursor sobre o bloco que se deseja apagar e repetir a operação explicada acima.

5. ALGUMAS ANÁLISES BIVARIADAS

5.1 - Comparação de duas variáveis categóricas: Teste Qui quadrado

Para avaliar a associação entre duas variáveis categóricas, ou ainda, para fazer uma comparação de uma variável categórica entre dois (ou mais) grupos, é utilizado o teste de χ^2 de associação ou de heterogeneidade, respectivamente. Na prática, ambos são obtidos no mesmo comando. Para exemplificar, vamos verificar se homens e mulheres diferem em suas opiniões sobre a permissão de porte de armas utilizando o banco **GSS93.SAV**.

- a) Entrar no menu **Analyze → Descriptive statistics → Crosstabs**;
- b) Como vamos comparar os sexos, incluir **gunlaw** nas colunas e a variável **sex** na linha: procurar a variável **sex** na lista de variáveis à esquerda e clicar em → ao lado da caixa **Row** e depois localizar a variável **gunlaw** e clicar em → ao lado da caixa **Column**;
- c) Clicar em **Exact** e depois marcar **exact**, – esta opção realiza os testes exatos ao invés dos assintóticos, que são uma aproximação – clicar em **continue**;

- d) Clicar em **Statistics**, depois em **Chi-square** – esta é a opção para pedir que ele realize o teste de χ^2 – e depois clique em **Continue**;
- e) Na opção Cells, podemos pedir diversos valores que aparecerão nas caselas da tabela. Clicar em **Cells**, depois em **Row** no quadrado das percentagens e depois em **continue**;
- f) Clicar em Ok ou Paste.

RESULTADO:

gunlaw Favor or Oppose Gun Permits * sex Respondent's Sex

Crosstab

			Favor or Oppose Gun Permits		Total
			Favor	Oppose	
Respondent's Sex	Male	Count	314	111	425
		% within Respondent's Sex	73,9%	26,1%	100,0%
	Female	Count	497	62	559
		% within Respondent's Sex	88,9%	11,1%	100,0%
Total		Count	811	173	984
		% within Respondent's Sex	82,4%	17,6%	100,0%

Podemos observar que, enquanto 73,9% dos homens opinaram a favor da permissão de uso de armas, 88,9% das mulheres manifestaram a mesma opinião.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	37,622 ^a	1	,000	,000	,000	
Continuity Correction ^b	36,592	1	,000			
Likelihood Ratio	37,417	1	,000	,000	,000	
Fisher's Exact Test				,000	,000	
Linear-by-Linear Association	37,584 ^c	1	,000	,000	,000	,000

N of Valid Cases	984				
------------------	-----	--	--	--	--

- a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 74,72.
- b. Computed only for a 2x2 table
- c. The standardized statistic is -6,131.

Pelo teste de χ^2 , podemos concluir que a diferença entre homens e mulheres é estatisticamente significativa ($P < 0,001$).

5.1.1 - Como obter os resíduos ajustados

Verificada a associação global entre as variáveis pode-se verificar se há associação local entre categorias, calculando-se os resíduos ajustados. O resíduo ajustado tem distribuição normal com média zero e desvio padrão igual a 1. Desta forma, caso o resíduo ajustado seja **maior que 1,96**, em valor absoluto, pode-se dizer que há evidências de associação significativa entre as duas categorias (p. ex. homem branco e salário alto) naquela casela. Quanto maior for o resíduo ajustado, maior a associação entre as categorias. Para obter os resíduos ajustados procede-se da seguinte maneira:

- a) Selecionar **“Analyze”**, **“Descriptive Statistics”**, **“Crosstabs”**;
- b) Clicar em **“Cells”**, abre-se a janela **“Crosstabs”**: **“Cell Display”**;
- c) Assinalar a opção **“Observed”** e **“Adj. standardized”**;
- d) Clicar em **“Continue”**; **“OK”**.

RESULTADO:

Crosstab

			Favor or Oppose Gun Permits		Total
			Favor	Oppose	
Respondent's Sex	Male	Count	314	111	425
		% within Respondent's Sex	-6,1	6,1	100,0%

	Female	Count	497	62	559
		% within Respondent's Sex	6,1	-6,1	100,0%
Total		Count	811	173	984

CONCLUSÃO: A associação entre **sex** (sexo) e **gunlaw** (permissão de porte de armas) já foi considerada significativa. Agora a pergunta é: Quais categorias estão associadas localmente? Olhando os resíduos ajustados vemos que os maiores valores (positivos) indicam forte associação entre homem e ser contra o porte de armas, bem como há forte associação entre mulher e ser a favor de porte de armas.

5.2 - Como comparar médias entre dois grupos: Teste t para Amostras Independentes.

O teste t é apropriado para comparar as médias de uma variável quantitativa entre dois grupos independentes.

EXEMPLO: Comparar a média de idade no primeira casamento entre homens e mulheres usando os dados do banco **GSS93.sav** .

- a) Sexo (**sex**) – variável categórica com dois grupos
- b) Idade no 1º casamento (**Agewed**) - Variável resposta quantitativa

Para a aplicação do teste t nesta situação procede-se da seguinte forma:

- a) Clicar em “**Analyze**”, “**Compare Means**”, “**Independent Samples t test**”;
- b) Clicar sobre a variável de teste (Test Variables): “**Agewed**” ou, conforme o caso em estudo, clicar na variável correspondente;
- c) Clicar sobre a variável de grupo (Grouping Variable) “**Gender**”;
- d) Clicar em: “**Define Group**”;
- e) Abre-se uma janela, na qual se define a categoria correspondente ao “**Group 1**” (no caso masculino) – digitando-se o código da categoria atribuída quando da construção do Banco de Dados,

nesse caso **1** e “**Group 2**” (no caso feminino) digitando-se o código **2**. (*Observação:* No caso de se desejar confirmar os valores atribuídos às variáveis, clicar sobre a variável e escolher a opção “**options**”)

g) Clicar em “**Continue**” e “**OK**”.

RESULTADO:

Group Statistics					
	Respondent's Sex	N	Mean	Std. Deviation	Std. Error Mean
Age When First Married	Male	492	24,16	4,867	,219
	Female	710	21,84	4,929	,185

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Age When First Married	Equal variances assumed	,342	,559	8,07	1200	,000	2,320	,288	1,756	2,885
	Equal variances not assumed			8,08	1064,66	,000	2,320	,287	1,757	2,883

INTERPRETAÇÃO: Uma das suposições do teste t é de que as variâncias dos dois grupos sejam iguais. Por isso é necessário testar esta hipótese antes (através do teste de Levene) e com base no resultado deste teste, escolhe-se o resultado do teste t supondo variâncias iguais ou supondo variâncias diferentes.

Ao serem analisados os dados do exemplo acima vemos o seguinte:

- a) Observa-se o resultado do teste de Levene. Neste exemplo, o valor de p para o teste Levene é 0,559, não se rejeita a hipótese de variâncias iguais.
- b) O teste t a ser utilizado é o que aparece na primeira linha (***Equal variances assumed***), considerando que $p < 0,001$ (Sig 2-tailed), rejeita-se a hipótese nula (H_0) de igualdade das médias dos dois grupos, logo, pode-se concluir que as médias da variável **agedwed** são significativamente diferentes entre os dois grupos de sexo.

As hipóteses do teste Levene de igualdade de variâncias são:

- Hipótese Nula (H_0): As variâncias dos dois grupos são iguais.
- Hipótese Alternativa (H_1): As variâncias dos dois grupos são diferentes.

As hipóteses do teste t para igualdade de médias entre amostras independentes são:

- Hipótese Nula (H_0): As médias dos dois grupos são iguais.
- Hipótese Alternativa (H_1): As médias dos dois grupos são diferentes