

UFMG

MANUAL TREINAMENTO ***SPSS***

SPSS 10.1
for Windows

SPSS

www.spss.com

© 2000 SPSS Inc. All rights reserved

INICIANTE

Introdução

O **SPSS** é um *software* apropriado para a elaboração de análises estatísticas de matrizes de dados. Seu uso permite gerar relatórios tabulados, gráficos e plotagens de distribuições, normalmente utilizados na realização de análises descritivas e inferências a respeito de correlações entre variáveis. É muito vantajoso também no aspecto relativo à interação com arquivos gerados por outros softwares, como por exemplo: ACCESS, EXCELL, LOTUS e DBASE.

O objetivo desta apostila é fornecer noções básicas de manipulação do *software*. Ela trata de maneira simplificada de todas as etapas de uma análise: edição de matrizes de dados, manipulação de arquivos, transformação dos dados, edição de relatórios de análise e elaboração de análises descritivas e de correlação entre variáveis.

De modo geral, os arquivos de dados a serem analisados não estão organizados no formato ideal para sua utilização. Sendo necessário, por exemplo, combinar arquivos de dados, classificar os casos em diferentes ordens, selecionar um subgrupo dos dados, alterar a unidade de análise ou criar novas variáveis a partir de informações existentes em outras variáveis. A apostila foi organizada buscando dar soluções práticas a alguns problemas considerados freqüentes por aqueles que estão começando a se familiarizar com o *software*.

O que é uma Matriz de Dados

Antes de partirmos para a explicação da utilização de algumas das ferramentas disponíveis no SPSS, vamos dar um pequeno exemplo de como se processa a construção de uma matriz de dados. É essencial termos uma idéia bem clara do que é uma matriz de dados, para que possamos entender os resultados estatísticos fornecidos pelo sistema.

Uma associação de bairro tomou a providência de reunir algumas informações a respeito dos casos de roubo que haviam ocorrido em seu bairro para tomar algumas providências em relação a orientação do comportamento da população. Após algum tempo de pesquisa descobriram que, do total de quarteirões do bairro, cinco quarteirões pareciam ser os mais violentos. Foram reunidas, então, as seguintes informações a respeito destes quarteirões:

Quarteirão	Número Assaltos no Último Mês	Número de Agressões Físicas	Classificação Econômica das Moradias	Proximidade a Grandes Vias de Trânsito	Proximidade da Delegacia de Polícia no Bairro
1	30	3	classe média	próximo	próximo
2	35	22	classe média	próximo	próximo
3	50	17	classe alta	próximo	não próximo
4	15	30	classe baixa	não próximo	não próximo
5	45	18	classe alta	não próximo	próximo

Obs. Informações fictícias.

Esta tabela constitui uma matriz de dados. A construção desta simples tabela e de qualquer matriz de dados possui alguns requisitos fundamentais para

que possamos confiar nas suas informações e desenvolver análises relevantes. São eles:

- deve existir um corpo básico de questões que são submetidos a todos os casos da maneira mais uniforme possível, evitando problemas de interpretação
- cada uma das informações (variáveis) - **Número Assaltos no Último Mês; Número de Agressões Físicas; Classificação Econômica das Moradias; Proximidade a Grandes Vias de Trânsito; Proximidade da Delegacia de Polícia no Bairro** - deve ser arquivada sempre com a mesma unidade de medida
- a responsabilidade daquele que coleta as informações é essencial para garantir a confiabilidade das informações presentes na matriz de dados
- deve-se fazer um esforço enorme para não deixar questões sem resposta

1. Manipulação de Arquivos e Edição de Matrizes de Dados.....	6
1.1. - CRIAÇÃO DE MATRIZES DE DADOS	7
1.1.1. - Como Criar uma Variável.....	7
1.1.2. - Como Preencher a Matriz de Dados	10
1.1.3. - Como Excluir uma Variável ou Um Caso de Análise dentro da Matriz de Dados	10
1.1.4. - Como Inserir uma Nova Variável ou um Novo Caso dentro da Matriz de Dados	10
1.1.5. - Como Guardar uma Matriz de Dados na Forma de um Arquivo	11
1.1.5.1. - Salvando Matrizes de Dados pela Primeira Vez na Forma de Arquivos	11
1.1.5.2. - Salvando Arquivos já Existentes (Atualizando Informações)	12
1.1.5.3. - Salvando Arquivos Antigos com Novo Nome ou em Novo Diretório... ..	12
1.1.6. - Como Abrir uma Matriz de Dados já Existente o Formato SPSS	12
1.1.7. - Como Abrir uma Matriz de Dados Elaborada no Excel.....	13
1.2. – MANIPULAÇÃO DE ARQUIVOS	14
1.2.1. - Deslocamento em direção de um Caso Específico.....	14
1.2.2. - Como Ordenar os Casos Segundo Uma ou Mais Variáveis	14
1.2.3. - Como Agregar ou Reunir Subgrupos de Casos Produzindo uma Matriz de Dados com Nova Unidade de Análise	15
1.2.4. - Como transformar colunas em linhas e linhas em colunas dentro de uma Matriz de Dados	17
1.2.5. - Como Reunir Dois Arquivos Contendo Variáveis Diferentes ou Casos para Análise Diferentes.....	18
1.2.5.1. - Reunindo Dois Arquivos com Casos para Análise Diferentes e mesmas Variáveis	19
1.2.5.2. - Reunindo Dois Arquivos com Variáveis Diferentes e mesmos Casos para Análise.....	20

2.1. - CÁLCULO ENVOLVENDO VARIÁVEIS JÁ EXISTENTES EM UMA MATRIZ DE DADOS.....	23
2.2. - RECODIFICAÇÃO DE VARIÁVEIS.....	27
2.2.1. - Como Recodificar uma Variável substituindo a Variável Original.....	27
2.2.2. - Como Recodificar uma Variável Criando Nova Variável.....	29
3.1. - MANIPULAÇÃO DE ARQUIVOS DE RELATÓRIOS DE ANÁLISE	31
3.1.1. - Como Guardar Arquivos Contendo Relatórios de Análises.....	31
3.1.2. - Como Abrir Arquivos Contendo Relatórios de Análises	31
3.2. - EDIÇÃO DE RELATÓRIOS DE ANÁLISE	33
3.3. – SELEÇÃO DE CASOS PARA ANÁLISE.....	34
3.4. - ANÁLISE DESCRITIVA DOS DADOS	36
3.4.1. - Como Fazer uma Análise Descritiva Baseada em Medidas de Tendência Central e Dispersão (moda, média, mediana, desvio padrão, etc)36	
3.4.2. - Alternativa de Construção de Gráficos para Análises de Tendência Central.....	38
3.4.3. - Summaries for Groups of Cases.....	39
3.4.4. - Summaries of Separate Variables.....	40
3.4.5. - Como Fazer uma Análise Descritiva Baseada na Distribuição de Frequência	41
3.4.6. - Como Elaborar Análises Utilizando Gráficos de Controle de Processo	42
3.5. - ANÁLISE DE ASSOCIAÇÃO ENTRE VARIÁVEIS.....	45
3.5.1. - Como construir uma tabela para verificar a relação entre variáveis ..	45
3.5.2. - Como construir análises de associação entre variáveis em função de coeficientes de correlação	46
3.5.3. - Como Construir Análises de Regressão Linear	48

1. Manipulação de Arquivos e Edição de Matrizes de Dados

A seção do SPSS onde é feita a entrada, manipulação e exclusão de dados é denominada **SPSS Data Editor**. Como podemos observar na figura abaixo, sua estrutura é a de uma matriz (linhas x colunas). Podemos entender tal disposição da seguinte forma: cada coluna representa uma variável e cada linha representa um caso. De uma forma simplificada, enquanto as colunas corresponderiam às questões de um questionário, as linhas corresponderiam as informações de cada questionário aplicado.

Observe que a criação da matriz de dados envolve, não apenas o preenchimento das informações correspondentes a cada caso analisado, mas a discriminação precisa do nome, definição, tipo e outras características das variáveis com que se está trabalhando.

	id	cidade	pop	pesfogo	pesbranc
1	1	BETIM	169984	9	
2	2	BRUMADINHO	19246	9	
3	3	CAETÉ	33187	5	
4	4	CONTAGEM	448665	203	
5	5	ESMERALDAS	24232	2	
6	6	IBIRITÉ	92369	37	
7	7	MATEUS LEME	27015	5	
8	8	MOEDA	3885	2	
9	9	NOVA LIMA	52246	23	
10	10	NOVA UNIÃO	4863	1	
11	11	PEDRO LEOPOLDO	41493	2	
12	12	RAPOSOS	14191	1	
13	13	RIBEIRÃO DAS NEVES	143543	19	
14	14	RIO ACIMA	7066	0	
15	15	RIO MANSO	4457	0	

Na figura acima, temos o exemplo de uma matriz de dados preenchida. As seguintes informações são importantes nesta tela:

1. No cabeçalho encontramos o nome do arquivo com que estamos trabalhando (matriz de dados): **CR Reg. Metropolitana BH**
2. Na grande faixa de cor branca localizada abaixo da barra de ferramentas do programa encontramos a seguinte informação:
 - 1:id → casela correspondente ao caso 1 da variável id
 - 1 → conteúdo da casela
3. A faixa de cor cinza localizada na margem superior da tabela nos fornece os nomes das variáveis (cabeçalho de cada coluna)
4. A faixa de cor cinza localizada na margem esquerda da tabela nos fornece o número de cada caso
5. No interior da tabela, as linhas correspondem aos casos analisados e as colunas correspondem às variáveis trabalhadas.

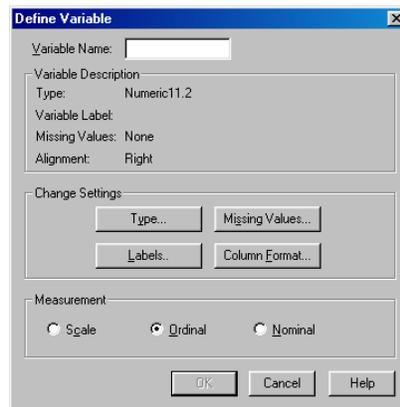
1.1. - Criação de Matrizes de Dados

A criação de uma matriz de dados envolve dois passos básicos: a criação e definição precisa das características de cada variável da matriz e o preenchimento dos dados para cada caso de análise existente.

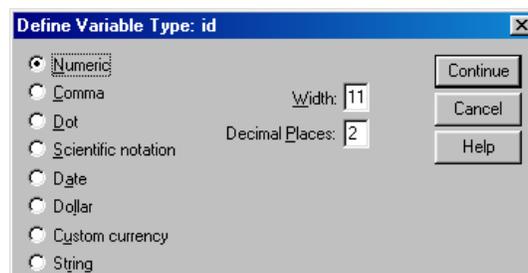
1.1.1. - Como Criar uma Variável

O passo mais importante na criação de uma matriz de dados é a definição das variáveis. Cada variável é criada separadamente, indicando seu nome, definição, tipo, categorias, formato da coluna na tabela e *missing values* (valores que por definição não entram nas análises estatísticas). Para definir uma variável, deve-se seguir os seguintes passos:

1. Selecione o local onde a variável será definida com um clique do *mouse* sobre o cabeçalho da coluna na margem superior da matriz
2. Selecione na barra de ferramentas a opção **DATA** e em seguida **DEFINE VARIABLE**. Chegaremos a seguinte figura:



3. No campo **Variable Name** devemos entrar com um nome para a variável. Este nome não pode ultrapassar 8 caracteres e não pode conter nenhum sinal algébrico ou espaço em branco no seu interior.
4. Os quatro campos abaixo - *type*, *labels*, *missing values* e *column format* - devem ser preenchidos em seguida, não importando a ordem com que são preenchidos. Ao final do preenchimento de cada quadro, deve-se selecionar a opção **CONTINUE**.
5. Ao selecionarmos a opção **TYPE** chegaremos a seguinte figura:

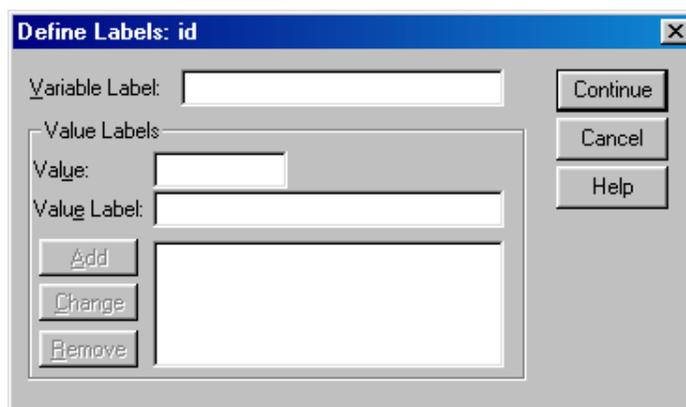


Entre as opções de tipo de variável acreditamos que as seguintes são importantes na formação de um conhecimento básico em SPSS:

- **Numeric**: estabelece que o campo será numérico → útil na definição de variáveis ordinais e categóricas
- **Date**: estabelece um formato de campo para a entrada de datas¹
- **Dollar**: estabelece um formato de campo para a entrada de valores monetários
- **String**: estabelece que o campo será alfa-numérico, podendo incluir qualquer tipo de informação desejada. Exemplo: nome de um município ou de uma pessoa.

Ainda neste campo - Define Variable Type - podemos selecionar o tamanho total de caracteres da variável (**Width**) e o número de caracteres correspondentes as casas decimais (**Decimal Places**), quando trabalhamos com o tipo numérico.

6. Ao selecionarmos a opção **LABELS** chegaremos a seguinte figura:



O campo **Variable Label** deve ser preenchido com uma descrição da variável.

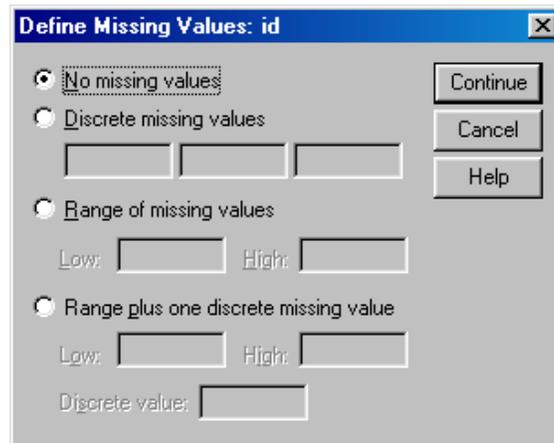
Para variáveis categóricas, o campo **Value Labels** permite a definição das diversas categorias de respostas. O valor a ser digitado na matriz deve ser inserido no campo **VALUE** e o significado corresponde inserido no campo **VALUE LABEL**. Para cada par de informações deve-se selecionar a opção **ADD** para adicioná-los a matriz de categorias. Caso algumas das categorias tenha sido definida de maneira errada, utilize as opções **CHANGE** ou **REMOVE** para fazer o seu acerto.

7. Ao selecionarmos a opção **MISSING VALUES** chegaremos a seguinte figura:

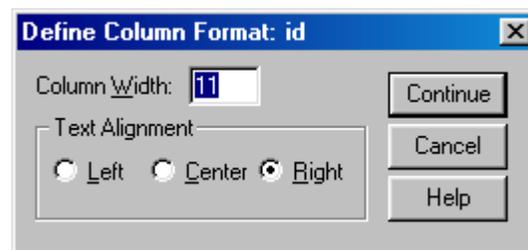
Como já foi indicado acima, aqui serão indicados todos os valores que não entrarão nas análises estatísticas que serão realizadas com a matriz de dados. É muito comum, por exemplo, estabelecer como missing os valores correspondentes

¹ Não é muito útil quando se pretende fazer contas de duração de tempo com as variáveis, pois ela não permite que se realize estes cálculos. Nesta situação é melhor criar três variáveis numéricas: uma com o ano, outra com o mês e a última com a dia da data a ser guardada.

às categorias: não respondeu, não sabe ou sem informação. Estes valores podem ser indicados de maneira precisa ou através de intervalos.



8. Ao selecionarmos a opção **COLUMN FORMAT** chegaremos a seguinte figura:



Este campo nos permite indicar a largura da coluna na tabela da matriz de dados e também o alinhamento do seu conteúdo dentro da casela. De um modo geral, este campo não é preenchido, utilizando o padrão que o próprio SPSS traz consigo.

9. O campo MEASUREMENT deve ser preenchida com o tipo de medida característica da variável. Estão disponíveis as seguintes opções:

- **Scale:** existe uma relação ordinal entre os valores que correspondem às categorias e o valor real da distância entre cada categoria é conhecida → ideal para variáveis que medem em números absolutos algum atributo dos casos como, por exemplo, idade, peso, altura, etc.
- **Ordinal:** existe uma relação ordinal entre os valores da variável mas a distância entre estes é desconhecida → ideal para variáveis que envolvem a avaliação de satisfação e que tem categorias como as seguintes: ótimo, bom, regular, ruim e péssimo. Não existe como definir qual é a distância precisa entre as posições ótimo e bom e ao mesmo tempo saber se esta distância é a mesma que existe entre as variáveis regular e ruim, por exemplo.
- **Nominal:** não existe nenhuma relação ordinal entre os valores da variável → ideal para variáveis que envolvem nomes de pessoas ou

lugares ou variáveis que contém informações como sexo, estado conjugal, etc

1.1.2. - Como Preencher a Matriz de Dados

Tendo definido todas as variáveis da matriz de dados, passamos para a entrada dos dados caso por caso; de um modo geral, recomenda-se que os dados sejam digitados por questionário, ou seja, linha por linha.

O preenchimento é feito digitando o valor atribuído à variável em cada caso seguido de (→ : o que fará com que se passe para a próxima variável do mesmo caso) ou (ENTER : o que fará com que se passe para o próximo caso na mesma variável). Para situações em que os valores se repetem muito, a utilização das opções RECORTAR(CUT) e COLAR (COPY) permite a agilização do trabalho.

1.1.3. - Como Excluir uma Variável ou Um Caso de Análise dentro da Matriz de Dados

Caso seja necessário excluir uma variável da matriz de dados, devemos selecionar a variável com o *mouse* clicando sobre o cabeçalho da coluna correspondente à variável que se deseja excluir. Tendo selecionado a variável, basta apertar a tecla **DELETE**. O mesmo procedimento deve ser usado em relação à exclusão de casos, selecionando-se a linha clicando sobre a margem esquerda da linha na matriz de dados.

1.1.4. - Como Inserir uma Nova Variável ou um Novo Caso dentro da Matriz de Dados

Caso seja necessário inserir uma nova variável no meio de variáveis já existentes numa matriz de dados, devemos inseri-la através dos seguintes passos:

1. Escolha o lugar onde a variável deve ser inserida e selecione a variável que estará a direita da nova variável a ser inserida clicando sobre o cabeçalho da coluna desta variável.
2. Selecione na barra de ferramentas o menu **DATA** e depois **INSERT VARIABLE**.
3. Em seguida seguiríamos todos os passos necessários para a definição da nova variável.

Se não for necessário inserir esta variável no meio de outras já existentes, basta ir até o fim do arquivo (à direita) e seguir os passos necessários para a definição da nova variável aproveitando os espaços em branco existentes após a última variável já existente.

Caso seja necessário inserir um novo caso para análise entre os casos já existentes numa matriz de dados, devemos seguir os seguintes passos:

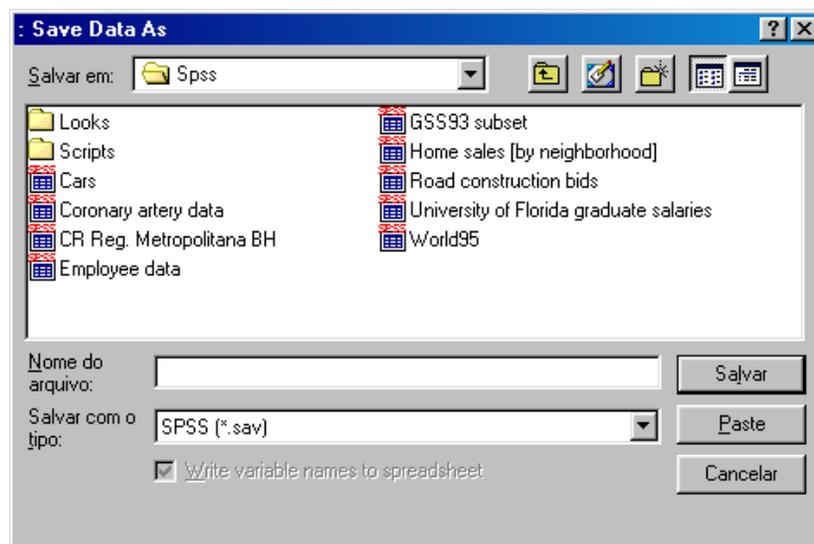
1. Escolha o lugar onde o caso deve ser inserido e selecione o caso que se localiza abaixo de onde este deve ser inserido clicando sobre a barra cinza que se localiza à esquerda do caso.
2. Seguiríamos, então, o seguinte caminho: na barra de ferramentas selecionaríamos **DATA** e depois **INSERT CASE**.
3. Em seguida seguiríamos todos os passos necessários para o preenchimento das informações do novo caso.

Se não for necessário inserir este novo caso entre os outros já existentes, basta ir até o fim do arquivo (abaixo) e seguir os passos necessários para o preenchimento de uma matriz de dados.

1.1.5. - Como Guardar uma Matriz de Dados na Forma de um Arquivo

1.1.5.1. - Salvando Matrizes de Dados pela Primeira Vez na Forma de Arquivos

Quando você está criando uma matriz de dados pela primeira vez, ou seja, ainda não existe um arquivo onde você vai guardá-la, é preciso criar este arquivo. O essencial nesta atividade, designada por **salvar** o arquivo, é dar nome ao arquivo e escolher o local ou diretório dentro no computador ou unidade de disco onde você vai guardá-lo. Você poderá salvar o arquivo antes de ter entrado com qualquer informação da matriz de dados que você deseja editar ou poderá editar uma parte ou toda a matriz e depois salvar o arquivo. Para salvar o arquivo você deverá selecionar **FILE** na barra de ferramentas e depois **SAVE**. Chegaremos na seguinte figura:



Através do campo **SALVAR EM** ou do retângulo abaixo deste campo você poderá selecionar o lugar (diretório) onde o arquivo será guardado. Tendo feito esta seleção, basta preencher o campo **NOME DO ARQUIVO** com o nome que se deseja dar ao arquivo. Lembre-se sempre de utilizar nomes que sejam claros na descrição do conteúdo da matriz de dados.

Caso você queira salvar o arquivo em outro formato diferente do padrão estabelecido pelo SPSS (*.sav), selecione o novo tipo desejado no campo **SALVAR COM O TIPO**.

1.1.5.2. - Salvando Arquivos já Existentes (Atualizando Informações)

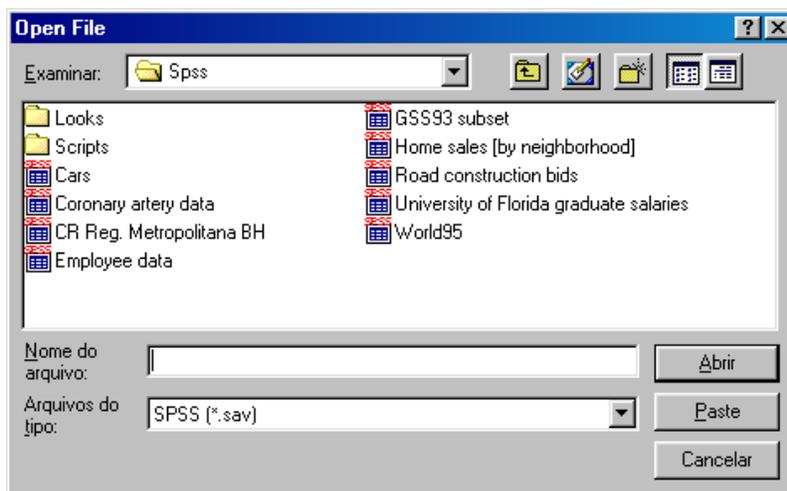
Para aqueles arquivos já criados anteriormente, basta selecionar **FILE** na barra de ferramentas e depois **SAVE** que o arquivo será salvo automaticamente pelo programa com o mesmo nome e no mesmo diretório onde ele se localiza.

1.1.5.3. - Salvando Arquivos Antigos com Novo Nome ou em Novo Diretório

Caso seja necessário fazer uma cópia do arquivo mudando seu nome ou localização dentro do computador, basta selecionar **FILE** na barra de ferramentas e depois **SAVE AS**. Esta opção lhe permitirá retornar para a primeira tela demonstrada nesta seção, onde será possível designar um novo nome e localização para o arquivo. Passará a existir duas cópias do arquivo, uma antiga com mesmo nome e localização da primeira e uma com novo nome e localização.

1.1.6. - Como Abrir uma Matriz de Dados já Existente o Formato SPSS

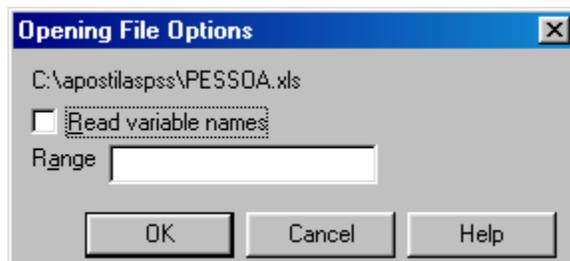
Cada matriz de dados é armazenada na forma de um arquivo próprio do SPSS, com terminação (*.sav) Para abrirmos uma matriz de dados já existente é essencial termos a informação precisa do local onde este arquivo se localiza. Seguiríamos, então, o seguinte caminho: na barra de ferramentas selecionaríamos **FILE** e depois **OPEN**. Chegaríamos a seguinte figura:



No campo **Examinar** devemos selecionar o diretório onde se localiza o arquivo que contém a matriz de dados com que queremos trabalhar. Em seguida selecione no quadro abaixo ao campo **Examinar** o arquivo ou digite no campo **Nome do Arquivo** o nome do arquivo a ser aberto. Tendo selecionado o arquivo, clique o *mouse* sobre a opção **Abrir**.

O padrão do SPSS é trabalhar com a terminação (*.sav) em seus arquivos. Caso você esteja querendo trabalhar com arquivos criados em outros sistemas, esta seleção padronizada precisará ser mudada. Teremos acesso aos diversos tipos de arquivos compatíveis com o SPSS, selecionando a seta localizada à

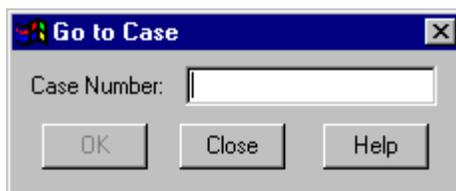
Estando o arquivo no formato apropriado para ser lido pelo SPSS, devemos selecionar na tela **OPEN FILE**, mostrada anteriormente, a opção (*.xls) no campo **Arquivo do Tipo**. Após a seleção do arquivo no diretório onde este está localizado, chegaremos à tela abaixo, onde deveremos selecionar a opção **Read variable names**.



Esta opção indicará para o programa que ele deverá ler a primeira linha do arquivo como o rótulo das variáveis (nome das variáveis). Bastará, então, selecionar a opção **OK** para que o arquivo seja aberto pelo programa SPSS.

1.2. – Manipulação de Arquivos

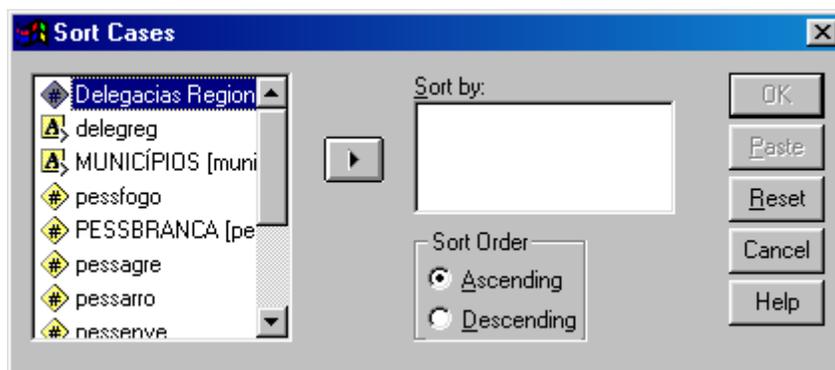
1.2.1. - Deslocamento em direção de um Caso Específico



Cada caso de uma matriz de dados está associada a um número específico na barra cinza que se localiza à esquerda da matriz de dados. Se for preciso localizar um caso específico para fazer alguma alteração de informações, deveremos utilizar o seguinte comando: na barra de ferramentas selecionaremos **DATA** e em seguida **GO TO CASE**. Chegaremos a seguinte tela, que deverá ser preenchida com o número do caso existente na barra cinza descrita.

1.2.2. - Como Ordenar os Casos Segundo Uma ou Mais Variáveis

Podemos sempre ordenar os casos analisados (ordem crescente ou decrescente) segundo uma ou mais variáveis de um banco de dados. Esta opção nos permite ver de maneira direta qual a posição dos diversos casos num ranking constituído por todos os casos. Para tal, devemos selecionar na barra de ferramentas a opção **DATA** e depois **SORT CASES**. Chegaremos a seguinte tela:



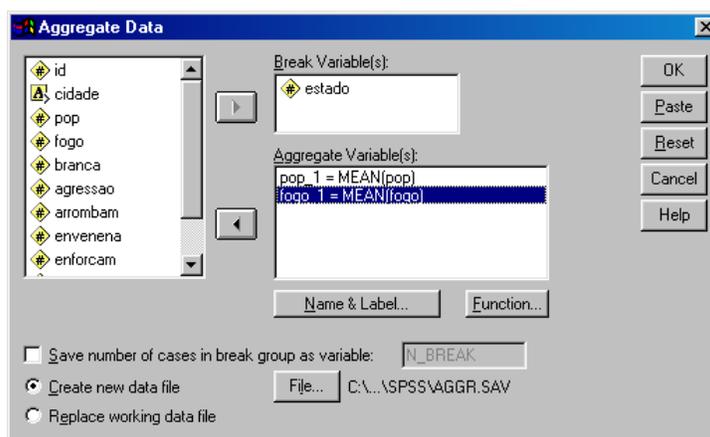
O campo **SORT BY** deverá ser preenchido com a variável(s) selecionada(s) no campo da direita, com base na qual se deseja fazer a classificação dos casos. No caso da inclusão de duas variáveis para ordenação, a primeira variável incluída no campo **SORT BY** definirá um primeiro padrão de classificação e a segunda definirá um padrão de classificação dentro da primeira já realizada.

O campo **SORT ORDER** deverá ser preenchido de modo a designar se a classificação dos casos seguirá ordem crescente ou decrescente. Estando estas duas opções selecionadas, bastará clicar na opção **OK**.

1.2.3. - Como Agregar ou Reunir Subgrupos de Casos Produzindo uma Matriz de Dados com Nova Unidade de Análise

O programa SPSS nos permite agregar casos criando uma nova matriz de dados onde a unidade de análise passa a ser subgrupos dos casos da primeira matriz. Por exemplo, podemos querer gerar uma matriz de dados contendo informações de alguns estados a partir de uma matriz onde temos informações sobre os municípios destes estados. Neste caso, o programa reunirá as informações dos municípios de cada estado separadamente e a unidade de análise passará a ser o estado. Estas informações poderão ser reunidas de várias maneiras a serem especificadas pelo usuário.

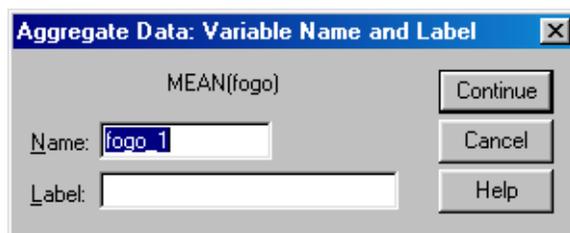
Primeiro, devemos abrir o arquivo a partir do qual se deseja fazer a agregação dos casos. Em seguida, devemos selecionar a opção **DATA** na barra de ferramentas e depois **AGGREGATE**. Aparecerá a seguinte tela, onde devemos estabelecer os critérios para a agregação dos casos.



O campo **BREAK VARIABLE(S)** deverá ser preenchido com as variáveis a partir das quais será realizada a agregação dos casos (sub-grupos). Na situação discutida no exemplo, desejamos fazer a agregação por estado. Este campo indica qual será a nova unidade de análise para a matriz de dados.

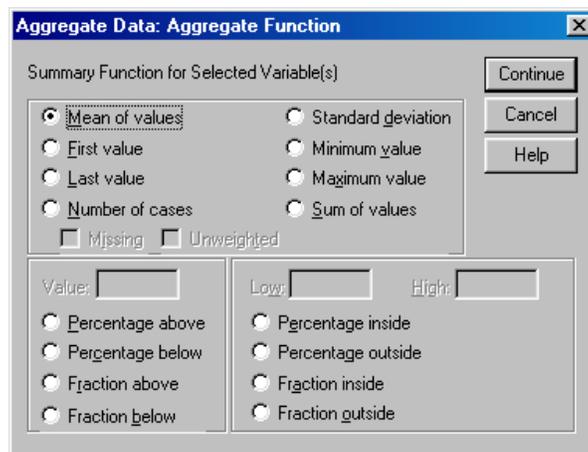
O campo **AGGREGATE VARIABLE(S)** deverá ser preenchido com as variáveis a serem agregadas e que farão parte da nova matriz de dados. Logo abaixo deste campo aparecem as opções **NAME & LABEL** e **FUNCTION**. Ambas deverão ser preenchidas para designar o tipo de função matemática utilizada durante a agregação de cada variável em particular e o nome a ser atribuído para cada nova variável a ser criada a partir desta agregação.

O tipo padrão de agregação das variáveis segue a opção MEAN (média dos valores) e atribui o seguinte nome a variável a ser criada no novo banco de dados: (antigo nome abreviado para 6 letras) seguido pelo sinal (_1) que indica o número de operações de agregação realizadas. Este número irá aumentando à medida em que se realizar agregações dentro de arquivos já agregados anteriormente. Estas duas opções podem ser mudadas de maneira específica para cada variável separadamente. Para tal, selecione a variável no campo **AGGREGATE VARIABLE(S)**. Ao selecionarmos a opção **NAME & LABEL** surgirá a seguinte tela:



O campo **NAME** deverá ser preenchido com o novo nome que se deseja dar a variável que contém o resultado da agregação. O campo **LABEL** deverá ser preenchido com a descrição da variável.

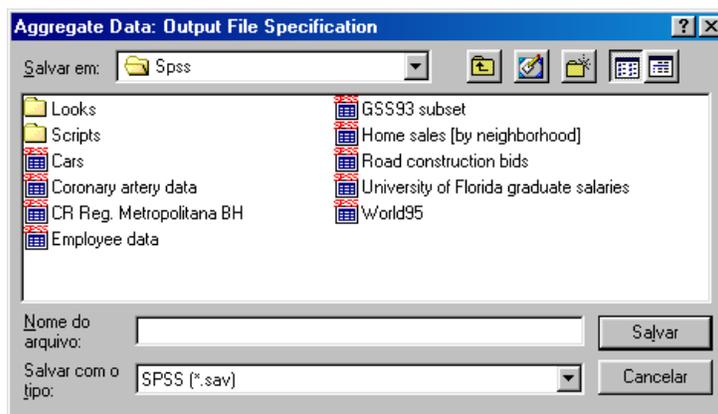
Ao selecionarmos a opção **FUNCTION** surgirá a seguinte tabela:



Nesta tela, poderemos selecionar a função matemática para agregação de cada variável separadamente. Como já foi dito anteriormente, a opção padrão do SPSS é a Mean of Values (Média dos Valores). Basta selecionar uma das outras opções existentes para que a função matemática de agregação das variáveis mude. Assim, cada variável poderá ter seus valores agregados por sub-grupos obedecendo uma certa função matemática, por exemplo: soma dos valores de todos os casos, media dos valores, maior valor, menor valor, etc.

Ainda poderemos selecionar a opção **SAVE NUMBER OF CASES IN BREAK GROUP AS VARIABLES**, se desejarmos criar uma variável na nova matriz de dados preenchida com o número total de casos agregados em cada um dos sub-grupos definidos pela **BREAK VARIABLE(S)**.

Por fim, poderemos ainda selecionar se a nova matriz de dados elaborada através da agregação será guardada substituindo a antiga matriz de dados (**REPLACE WORKING DATA FILE**), ou seja, com o mesmo nome e localização do antigo arquivo, ou guardada em um novo arquivo (**CREATE NEW DATA FILE**). Se escolhermos a última opção, deveremos indicar na tela abaixo o nome e localização do novo arquivo. O programa indica como padrão o nome (AGGR.SAV) para este arquivo e estabelece como localização o diretório (/SPSS) onde o programa foi instalado.

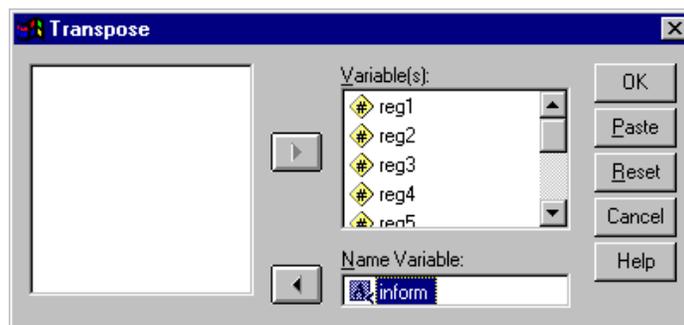


1.2.4. - Como transformar colunas em linhas e linhas em colunas dentro de uma Matriz de Dados

Quando estamos trabalhando com bases de dados elaboradas por outros pesquisadores, pode ser necessário utilizarmos do recurso de inverter as colunas e linhas em uma matriz de dados, para que seja possível aproveitar tal matriz dentro do programa SPSS.

	inform	reg1	reg2	reg3	reg4	reg5	reg6	reg7	var
1	IDADE	23,00	27,00	12,00	45,00	50,00	19,00	35,00	
2	POP	1500	1789	2870	4500	5634	8914	6475	

A tabela acima constitui um exemplo de um caso em que devemos fazer esta transformação. Como podemos perceber, os casos de análise da matriz estão na coluna e correspondem às regiões (1 a 7) e as variáveis estão na linha e correspondem à idade média e população total de cada região. Para utilizar este recurso, devemos selecionar **DATA** na barra de ferramentas e depois o comando **TRANSPOSE**. Chegaremos a seguinte tela:



O campo denominado VARIABLE deverá ser preenchido com o nome das variáveis a serem transformadas em casos. O campo NAME VARIABLE deverá ser preenchido com o nome da variável que contém a informação acerca do nome das variáveis na nova matriz de dados. Como resultado do preenchimento acima, chegaríamos à matriz de dados mostrada abaixo. Como podemos ver, cada coluna da primeira matriz de dados se transformou numa linha da nova matriz de dados. A única exceção desta transformação foi a primeira coluna, que forneceu as informações acerca da denominação das variáveis na nova matriz de dados.

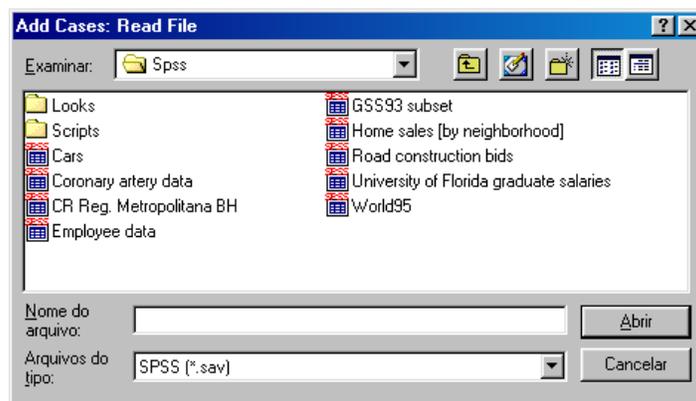
1:case_lbl		REG1			
	case_lbl	idade	pop	var	v
1	REG1	23,00	1500,00		
2	REG2	27,00	1789,00		
3	REG3	12,00	2870,00		
4	REG4	45,00	4500,00		
5	REG5	50,00	5634,00		
6	REG6	19,00	8914,00		
7	REG7	35,00	6475,00		

A única exigência para a execução deste comando pelo SPSS é que as variáveis a serem transformadas em casos na nova matriz de dados sejam definidas como variáveis numéricas.

1.2.5. - Como Reunir Dois Arquivos Contendo Variáveis Diferentes ou Casos para Análise Diferentes

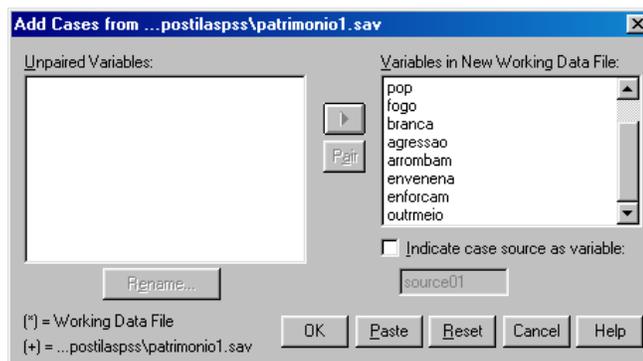
Com o SPSS, você pode reunir informações de dois arquivos diferentes. É possível agrupar informações a partir de dois arquivos que tenham as mesmas variáveis e casos para análise diferentes ou agregar informações de dois arquivos que tenham variáveis diferentes e mesmos casos para análise. Neste último caso podemos ter, por exemplo, um arquivo contendo o tamanho da população dos municípios de um estado e outro arquivo contendo algumas informações sócio-econômicas a respeito destes mesmos municípios. A variável chave município será utilizada para dar ordem a agregação de modo a não misturar as informações de municípios diferentes.

1.2.5.1. - Reunindo Dois Arquivos com Casos para Análise Diferentes e mesmas Variáveis



Para reunir as informações de dois arquivos de dados que contém as mesmas variáveis e casos para análise diferentes devemos inicialmente estar com algum arquivo aberto, onde serão adicionados os novos casos. Selecione na barra de ferramentas **Data** depois escolha a opção **Merge Files**, clique em **Add Cases...** Isto abrirá uma caixa de diálogo como a que se segue:

Nesta tela deveremos indicar para o programa qual o arquivo que contém os casos a serem adicionados no arquivo já aberto. Para tal, selecionaremos o diretório onde este arquivo se localiza e indicaremos o nome do arquivo. Podemos inclusive trabalhar com arquivos de tipos diferentes do padrão do SPSS (*.sav). Após indicado o nome do arquivo, clique com o mouse sobre o campo **ABRIR**. Chegaremos na seguinte figura.



O campo **Variables in New Working Data File** será preenchido automaticamente pelo computador com o nome de todas as variáveis comuns entre os dois arquivos e que poderão, deste modo, ter os casos agregados.

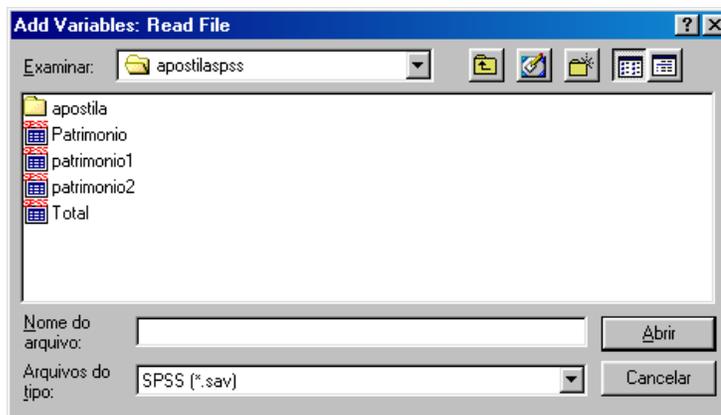
O campo **Unpaired Variables** será preenchido automaticamente pelo computador com o nome de todas as variáveis que não são comuns entre os dois bancos de dados. O usuário poderá também fazer uma seleção das variáveis que ele não deseja que tenham os casos agregados. Para isto, basta deslocar as variáveis do campo da direita (variáveis comuns ou compartilhadas pelos dois bancos) para o campo da esquerda (variáveis não comuns) utilizando a seta presente entre os dois campos.

O campo **Indicate case source as variable** indica para o programa que seja criada uma nova variável neste arquivo que está sendo criado a partir da agregação. Esta variável será preenchida com o valores 0 e 1, onde 0 será associado aos casos do arquivo já aberto antes da agregação e 1 para os casos do arquivo aberto posteriormente.

1.2.5.2. - Reunindo Dois Arquivos com Variáveis Diferentes e mesmos Casos para Análise

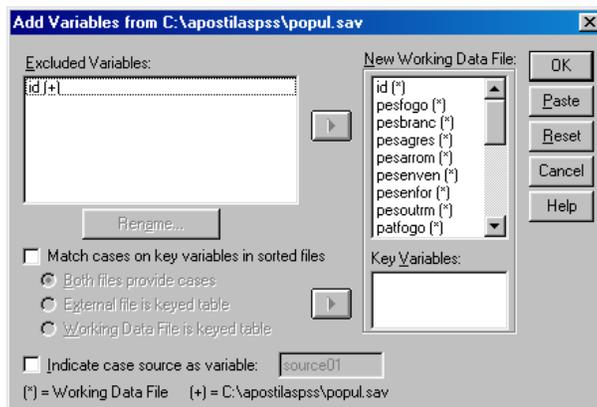
Para podermos reunir as informações de dois arquivos de dados que contém os mesmos casos para análise e variáveis diferentes, existe a necessidade de se atender a dois requisitos básicos: existir uma variável chave comum, ou seja, uma variável comum aos dois bancos de dados que servirá de referência para a junção das informações dos casos e os dois arquivos devem ter seus casos organizados em ordem crescente com base nesta variável chave.

Devemos inicialmente estar com algum arquivo aberto, onde serão adicionadas as novas variáveis. Selecione na barra de ferramentas **Data** depois escolha a opção **Merge Files**, clique em **Add Variables...** Isto abrirá uma caixa de diálogo como a mostrada a seguir.



Nesta tela deveremos indicar para o programa qual o arquivo que contém as variáveis a serem adicionados no arquivo já aberto. Para tal, selecionaremos o diretório onde este arquivo se localiza e indicaremos o nome do arquivo. Podemos inclusive trabalhar com arquivos de tipos diferentes do padrão do SPSS (*.sav).

Após indicado o nome do arquivo devemos selecionar a opção **ABRIR**. Chegaremos na seguinte figura.



O campo **Variables in New Working Data File** será preenchido automaticamente pelo computador com o nome de todas as variáveis que não são comuns entre os dois arquivos e que poderão, deste modo, ser agregadas.

O campo **Excluded Variables** será preenchido automaticamente pelo computador com o nome de todas as variáveis que são comuns entre os dois bancos de dados. Neste caso, a variável (**id**) é necessariamente comum entre os dois arquivos pois ela é a variável chave.

O campo **Match cases on key variables in sorted files** deverá ser selecionado para que o usuário indique qual é a variável a ser utilizada como chave durante a agregação dos arquivos. Logo em seguida deve-se selecionar a variável chave no campo **Excluded Variables** e inseri-la no campo **Key Variable** clicando sobre a seta à esquerda deste campo.

A função permite ainda a seleção de alguns padrões para a técnica de agregação dos arquivos segundo as “tabelas chave”. Por exemplo, podemos estar querendo agregar uma matriz de dados de vários municípios com uma matriz de vários Estados. O banco de dados contendo as informações dos estados poderá ser utilizado como “tabela chave”, para que as informações de cada estado sejam agregadas automaticamente em todos os municípios pertencentes aquele estado. Estes padrões são:

- Working Data File is Keyed Table: a “tabela chave” neste caso é o arquivo que já estava aberto antes do procedimento de agregação
- External File is Keyed Table: a “tabela chave” neste caso é o arquivo que foi aberto durante o processo de agregação
- Both Files Provide Cases: os dois arquivos são encarados pelo programa como “tabelas chave”.

O campo **Indicate case source as variable** indica para o programa que seja criada uma nova variável no arquivo a ser criado com a agregação dos dados. Esta nova variável será preenchida com o valores 0 e 1, onde 0 será associado aos casos do arquivo já aberto antes da agregação dos casos e 1 para os casos do arquivo aberto posteriormente.

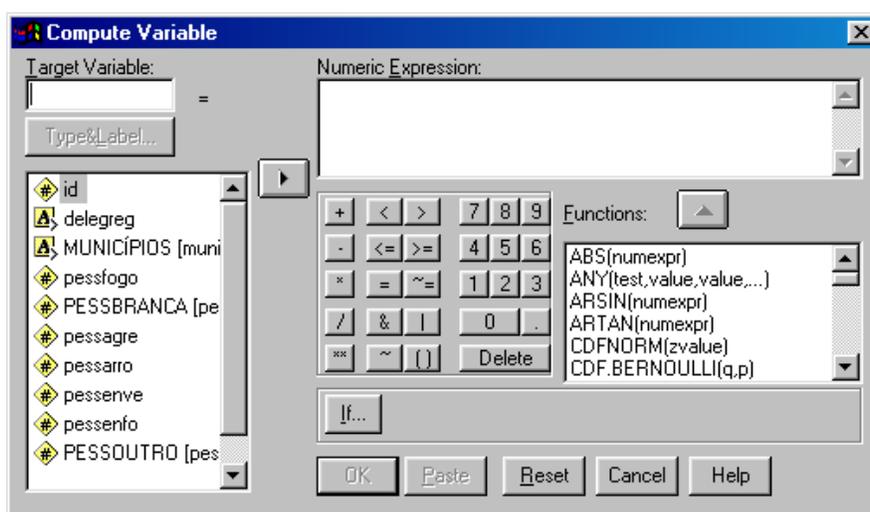
Como podemos ver na figura, todas as variáveis do banco de dados aberto antes do processo de agregação (working Data File) serão seguidas sempre por um (*). As variáveis do banco de dados aberto posteriormente serão seguidas sempre por um (+).

O programa exige que os dois arquivos a serem agregados tenham seus casos classificados em ordem crescente segundo a variável chave de agregação. Por isso ele sempre mostra uma mensagem informando esta exigência antes de fazer a agregação dos arquivos.

2. - Transformação de Dados

2.1. - Cálculo envolvendo Variáveis já Existentes em uma Matriz de Dados

Em muitas situações, obtém-se informações importantíssimas realizando cálculos a partir de variáveis presentes no banco de dados. Isto envolveria basicamente a criação de uma nova variável preenchida com o resultado da operação matemática com as outras variáveis envolvidas. Estando com o arquivo a partir do qual se deseja fazer tais cálculos aberto, devemos escolher na barra de ferramentas a opção **TRANSFORM** e depois **COMPUTE**. Em seguida aparecerá a tela:



Deve-se preencher o campo **TARGET VARIABLE** com o nome da nova variável, onde colocaremos o resultado da conta a ser realizada. É possível especificar o tipo e definição desta nova variável se selecionarmos o campo **TYPE & LABEL**.²

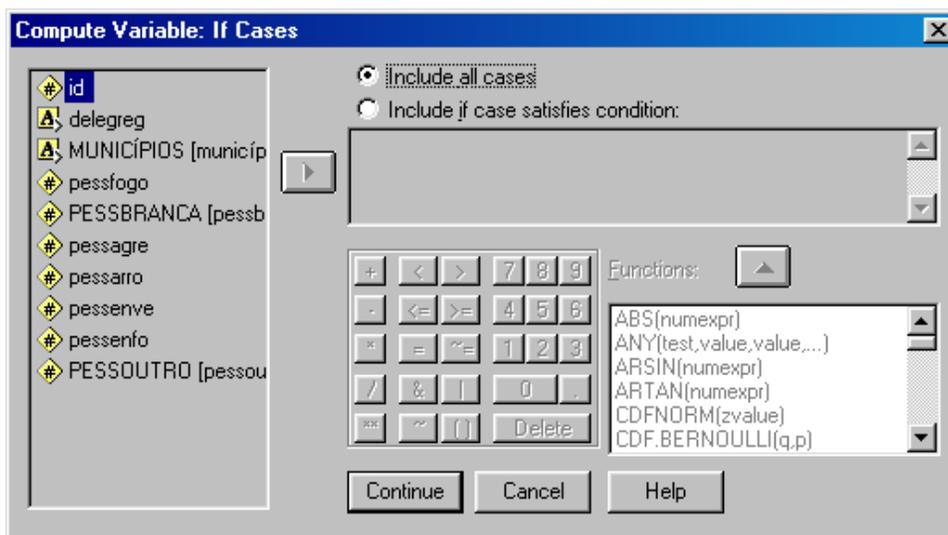
Esta nova variável será igual ao valor do resultado da operação matemática definida no campo **NUMERIC EXPRESSION**. A definição deste campo segue as regras básicas da matemática, como por exemplo a ordem de execução envolvendo parênteses, chaves e colchetes. O preenchimento do campo pode ser realizado através do clique do *mouse* sobre o quadro de sinais ou através do teclado digitando os mesmos sinais presentes na figura. Listamos, a seguir, alguns exemplos básicos para facilitar a compreensão da utilização do comando:

Operação	Expressão
Variável C é igual a soma de A e B	$C = A + B$
Variável C é igual a divisão de A por 100	$C = A / 100$

² Uma definição mais detalhada da nova variável pode ser realizada também utilizando a rotina indicada acima para a definição de variáveis.

A operação matemática descrita no campo **NUMERIC EXPRESSION** pode envolver também a utilização de algumas ferramentas matemáticas de maior complexidade associadas, por exemplo, à estatística ou à trigonometria. Para este caso, existe uma série de funções matemáticas definidas na caixa **FUNCTIONS**. Após escolhida a função matemática, insira a função no campo **NUMERIC EXPRESSION** clicando sobre a seta que está acima do campo **FUNCTIONS**.

Outra opção presente no comando **COMPUTE** é a possibilidade de fazer uma seleção dos casos em que a operação indicada será realizada. Isto dependerá dos valores encontrados em uma ou mais variáveis, presentes ou não na operação matemática descrita. Ao selecionar a opção **IF** chegaremos na seguinte figura:



A opção padrão é a **include all cases**, ou seja, a operação será realizada em todos os casos existentes no banco de dados. Podemos, no entanto, selecionar o caso em que esta operação se realizará ao clicarmos na opção: **include if case satisfies condition**. Indicaremos, então, uma nova expressão numérica que deve ser satisfeita para que a operação matemática indicada seja realizada. Também neste caso, a expressão numérica pode incluir funções matemáticas mais complexas. Listamos, a seguir, alguns exemplos básicos para facilitar a compreensão da utilização do comando:

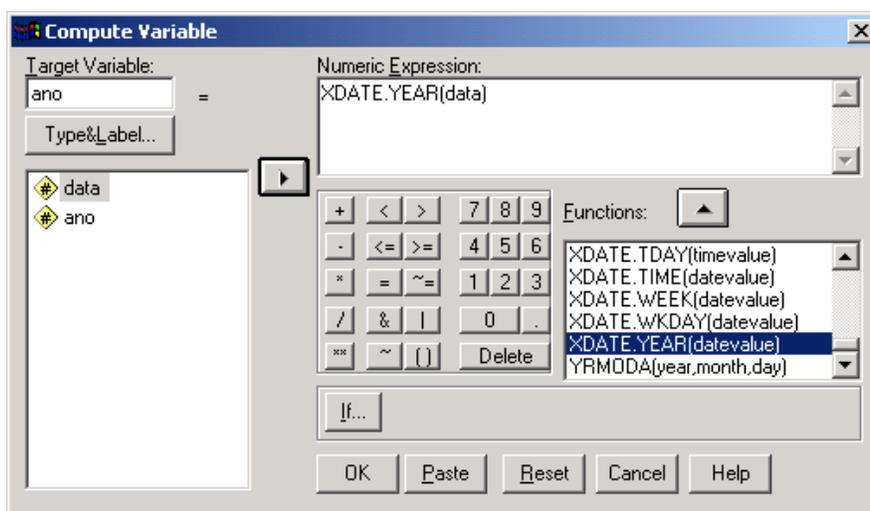
Condição	Expressão
Variável C é menor que 100	$C < 100$
Variável C é diferente de A	$C \neq A$
Variável C é menor que a soma de A e B	$C < A + B$

A opção **FUNCTIONS** pode ainda ser utilizada para retirar informações de tempo de variáveis definidas como tipo DATE. Estas opções permitem, por exemplo, retirarmos de uma variável preenchida como 22/12/2002 o ano da

ocorrência (2002). Seguindo os passos colocados acima para a utilização do comando COMPUTE devemos utilizar as opções apresentadas na tabela abaixo na digitação da fórmula para retirarmos cada um dos itens desejados.

Comando	Informação Desejada
XDATE.HOUR	HORA
XDATE.MINUTE	MINUTO
XDATE.TIME	NÚMERO SEGUNDOS DESDE A MEIA NOITE
XDATE.YEAR	ANO
XDATE.MONTH	MÊS
XDATE.MDAY	DIA DO MÊS
XDATE.QUARTER	TRIMESTRE
XDATE.SECOND	SEMESTRE
XDATE.WEEK	SEMANA DO MÊS
XDATE.WKDAY	DIA DA SEMANA
XDATE.JDAY	DIA DO ANO

O preenchimento do comando COMPUTE deverá ser feito conforme demonstrado na figura abaixo.



Untitled - SPSS for Windows Data Editor			
File Edit View Data Transform Analyze Graphs Utilit			
1:data			
	data	ano	var
1	22.02.2002		
2			
3			

Untitled - SPSS for Windows Data Editor			
File Edit View Data Transform Analyze Graphs Utilit			
1:ano			
	data	ano	var
1	22.02.2002	2002,00	
2			
3			

Conforme demonstrado acima, tínhamos um arquivo com uma variável data preenchida com o valor 22/02/2002 (22 de fevereiro de 2002) e executando o

comando `XDATE.YEAR` retiramos o ano, criando uma nova variável `ano` com o valor 2002.

2.2. - Recodificação de Variáveis

A necessidade de recodificação de variáveis envolve basicamente duas situações: a agregação de categorias de maneira a construir novas categorias mais apropriadas para a análise estatística desejada e a transformação de variáveis do tipo ordinal em variáveis categóricas. Por cautela, recomenda-se que as variáveis originais a serem recodificadas nunca sejam excluídas após a sua recodificação. Deve-se salientar, que para a execução de uma boa recodificação é essencial termos um conhecimento claro da distribuição de valores da variável original. As etapas para se chegar a este conhecimento serão explicadas posteriormente na seção **ANÁLISE DESCRITIVA**.

Tendo selecionado os parâmetros para a nova recodificação, deve-se seguir o seguinte caminho: na barra de ferramentas selecione **TRANSFORM** e depois **RECODE**. Em seguida, as seguintes opções para a realização da recodificação aparecerão: **into same variables** / **into different variables**. A primeira opção realiza a recodificação da variável sobre ela mesma, apagando o conteúdo da variável original. A segunda opção realiza a recodificação em uma variável diferente, permitindo que se mantenha a variável original intocada. Descreveremos a seguir o funcionamento das duas opções.

Na tabela abaixo sugerimos um exemplo típico de recodificação para permitir uma melhor compreensão do comando. Quando aplicamos questionários, é muito comum colhemos a informação idade na forma de valores absolutos. Na hora da análise, surge a necessidade de criar faixas etárias, pois para uma série de aspectos a análise por faixa facilita o trabalho. Surge, então, a seguinte situação:

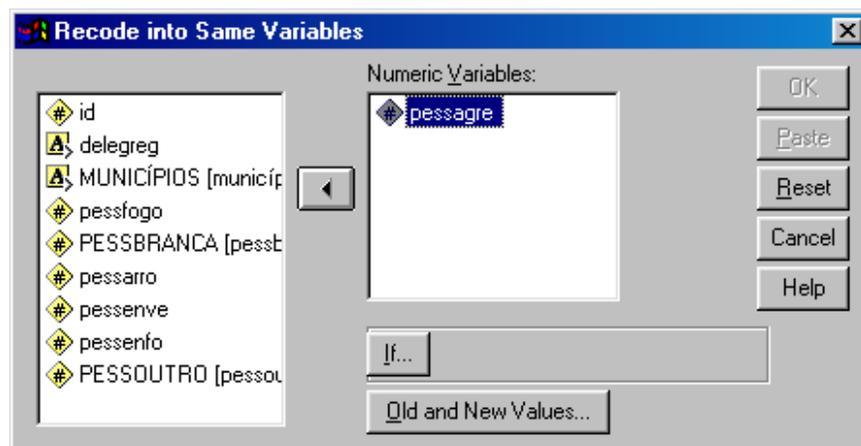
IDADE EM VALOR ABSOLUTO	COMANDO DE RECODIFICAÇÃO	NOVA VARIÁVEL RECODIFICADA
2	0 até 4 → 1	1
9	5 até 9 → 2	2
15	10 até 14 → 3	4
20	15 até 19 → 4	5
25	20 até 24 → 5	6
26	25 até 29 → 6	6

No comando definição da variável, todos estes valores da nova variável devem ser definidos como categorias (**value labels**). Assim, as análises serão reproduzidas pelo computador de uma forma que todos saberão qual faixa etária corresponde cada um destes valores. Deixamos de ter uma variável com valor absoluto para termos uma variável com categorias e com um significado específico para cada uma delas. O 1 corresponderá à faixa de 0 até 4 anos, o 2 à faixa de 5 até 9 anos e assim por diante. Apesar de perdermos um pouco de precisão na informação, ganhamos em agilidade para a análise dos dados.

2.2.1. - Como Recodificar uma Variável substituindo a Variável Original

Para fazermos uma recodificação substituindo a variável original sobre a qual foi feita a recodificação dos valores, devemos na barra de ferramentas

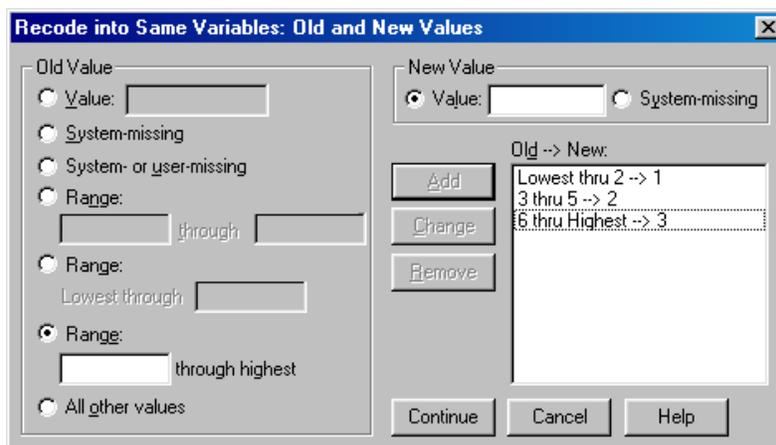
selecionar a opção **TRANSFORM**, depois **RECODE** e depois **INTO SAME VARIABLE**. Aparecerá a seguinte tela:



Devemos inicialmente selecionar no retângulo à esquerda a variável da matriz de dados a ser recodificada. Se existirem variáveis com mesmo padrão de recodificação, estas poderão ser feitas de uma só vez, selecionando todas ao mesmo tempo. Feita a seleção, clicamos na seta à direita deste retângulo para que a variável seja introduzida no campo **NUMERIC VARIABLE**.

Assim como para o comando **COMPUTE**, podemos selecionar através do **IF** os casos onde esta recodificação será realizada. O funcionamento desta opção é idêntica à encontrada para o comando **COMPUTE**, por isso volte a este comando para obter qualquer esclarecimento sobre esta opção.

O passo seguinte da recodificação será indicar os valores novos que substituirão os valores a serem recodificados. Selecionamos, para tal, a opção **Old and New Values**. Com esta seleção, encontraremos a seguinte figura:



O retângulo **Old Value** deverá ser preenchido com os valores a serem recodificados, enquanto que o retângulo **New Value** deverá ser preenchido com os valores que substituirão estes valores a serem recodificados. No campo **Old Value**, os valores podem ser preenchidos na forma de valores absolutos ou

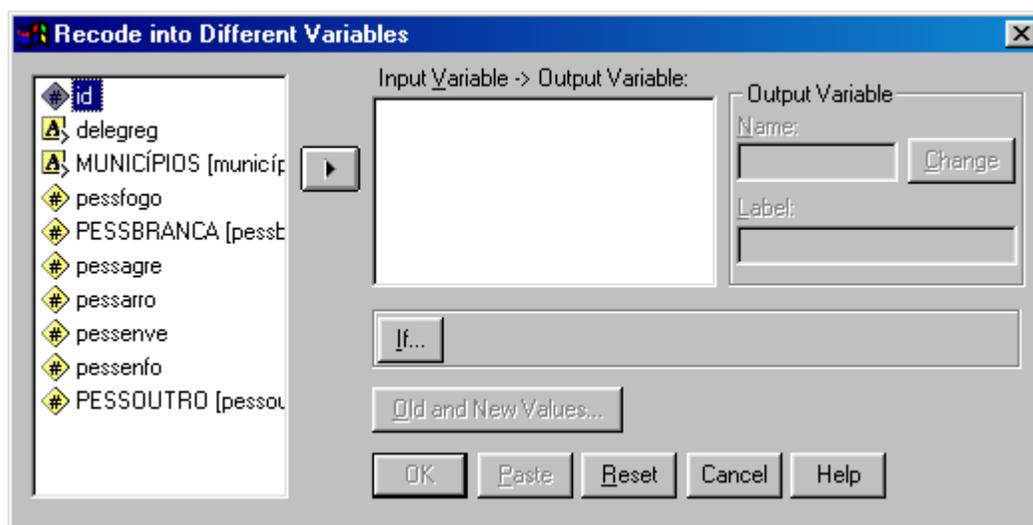
intervalos. No campo **New Value**, os valores só podem ser preenchidos na forma de valores absolutos.³

A cada par de **Old Value & New Value** deve-se clicar no campo **ADD** para inserir este par na lista de recodificações planejadas. Qualquer alteração neste par de valores a ser recodificado poderá ser realizada utilizando os recursos disponibilizados pelos campos: **CHANGE** e **REMOVE**. Após escolher todos os pares de valores a serem recodificados, basta apenas selecionar o campo **CONTINUE**.

Por fim, é bom deixar explícito o significado dos termos **system missing** e **user missing**. Ao preenchermos a nossa matriz de dados, os campos deixados sem informação (em branco) são preenchidos automaticamente pelo computador - **system missing** - com o seguinte símbolo ' '. Os valores denominados de **user missing** são definidos pelo próprio usuário através da definição de **missing values** no ato de definição das variáveis.

2.2.2. - Como Recodificar uma Variável Criando Nova Variável

Para fazermos uma recodificação criando uma nova variável a partir da variável recodificada, devemos na barra de ferramentas selecionar a opção **TRANSFORM**, depois **RECODE** e depois **INTO DIFFERENT VARIABLE**. Aparecerá a seguinte tela:



Para a opção Recode into Different Variables, devemos inicialmente selecionar no retângulo à esquerda a variável da matriz de dados a ser recodificada. Feita a seleção, clicamos na seta à direita deste retângulo para que a variável seja introduzida no campo **NUMERIC VARIABLE → OUTPUT VARIABLE**. Como estamos tratando de uma recodificação em diferentes variáveis, devemos em seguida preencher o campo **OUTPUT VARIABLE** com o nome da nova variável (**NAME**) e sua descrição (**LABEL**). Tendo preenchido estas

³ Estas limitações de preenchimento dos campos indica de maneira clara a utilidade do comando. Podemos construir, por exemplo, faixas etárias de modo que cada intervalo etário será substituído por uma única categoria (0 à 9 anos → 1 , 10 à 19 anos → 2 ,)

duas informações, basta clicar no campo **CHANGE** para que a nova variável seja também incluída no campo **NUMERIC VARIABLE → OUTPUT VARIABLE**.

Os comandos **IF** (seleção dos casos para recodificação) e **Old and New Value** (indicação dos valores a serem substituídos e dos valores novos) seguem as mesmas regras descritas para a opção **Recode into Same Variable (Recodificação sobre a mesma Variável)**.

3. - Relatórios de Análise de Matrizes de Dados

3.1. - Manipulação de Arquivos de Relatórios de Análise

A seção do SPSS onde é feita a criação, manipulação, exclusão e impressão dos resultados das análises estatísticas feitas pelo SPSS é denominada **SPSS Viewer**. Como podemos observar na figura abaixo, a tela se divide entre duas partes: a da esquerda apresenta em tópicos todas as tabelas e gráficos produzidos durante a análise e a da direita expõe a tabela ou gráfico selecionado na parte da esquerda. Podemos ver, a seguir, um exemplo da estrutura desta seção.

The screenshot shows the SPSS Viewer interface. On the left, a tree view shows the 'Output' folder containing 'Frequencies', 'Title', 'Notes', 'Statistics', and 'Delegacias Reg'. The 'Frequencies' item is selected with a red arrow. The main area displays the following statistics and table:

Statistics

Delegacias Regionais	
N	Valid 42
	Missing 0

Delegacias Regionais					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Uberlândia	12	28,6	28,6	28,6
	Uberaba	19	45,2	45,2	73,8
	Curvelo	11	26,2	26,2	100,0
	Total	42	100,0	100,0	

As análises contém sempre a mesma estrutura: título, anotações, um campo denominado *statistics* e a tabela ou gráfico produzida. Aquele item selecionado no campo da esquerda (marcado pela seta vermelha TM) aparecerá no campo da direita em evidência.

Toda vez que pedimos ao programa SPSS para gerar alguma tabela, gráfico, medidas estatísticas, etc, e que não existe nenhum arquivo de análise aberto, ele criará um novo arquivo dentro do **SPSS Viewer**. Devemos perceber que não se trata de uma arquivo de tipo (*.sav) como foi definido anteriormente para a guarda de matrizes de dados. Este novo arquivo terá a terminação (*.spo) e servirá apenas para guardar estas análises.

3.1.1. - Como Guardar Arquivos Contendo Relatórios de Análises

O SPSS gera automaticamente arquivos de relatórios com o nome **OUTPUT** seguido por um número (1, 2, 3, etc) de acordo com o número de arquivos de relatórios de análises já abertos. Se desejarmos alterar este nome ou direcionar a localização para o arquivo ser guardado, devemos seguir as mesmas instruções da opção **Como Guardar uma Matriz de Dados na Forma de um Arquivo**.

3.1.2. - Como Abrir Arquivos Contendo Relatórios de Análises

Se desejarmos abrir arquivos de relatórios de análise já guardados anteriormente, devemos seguir as instruções da opção **Como Abrir uma Matriz de Dados já Existente**.

3.2. - Edição de Relatórios de Análise

Uma vez produzido o relatório e análise, este pode ser alterado em alguns itens como título, texto dentro da tabela, formatação do gráfico, etc.

Se selecionarmos o item **Title** no espaço da esquerda, seu conteúdo será mostrado no espaço da direita. O título poderá, então, ser editado, se clicarmos duas vezes repetidas sobre ele no campo da direita.

Se selecionarmos o item **Notes** no espaço da esquerda, uma série de características da análise pedida será mostrada no espaço da direita: data em que foi criado o relatório, nome e localização do arquivo que contém a matriz de dados utilizada para elaborar o relatório, se foi utilizado algum filtro para selecionar os casos para a análise ou peso para atribuir importância diferente aos casos, número total de casos analisados, existência de *missing values*, o comando utilizado para gerar o relatório e o tempo total que o computador levou para fazer o relatório. Tais informações poderão ser editadas se clicarmos duas vezes repetidas sobre o quadro.

Notes

Output Created		11 May 00 17:14:44
Comments		
Input	Data	C:\apostilasps\apostila\DELEGA(A REGIONAL.sav
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	42
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics are based on all cases with valid data.
Syntax		FREQUENCIES VARIABLES=iddeleg /ORDER ANALYSIS .
Resources	Total Values Allowed	18724
	Elapsed Time	0:00:00.06

Se selecionarmos o item **Statistics** ou **Case Processing** no espaço da esquerda, seu conteúdo será mostrado no espaço da direita: o número total de casos considerados válidos para a análise e o número total de casos caracterizados como *missing values* e que por isso não foram computados na análise. Tal tabela poderá ser editada se clicarmos duas vezes com o mouse sobre ela no campo da direita.

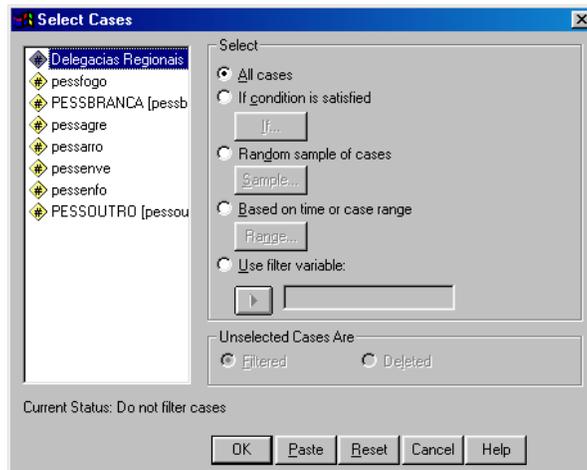
Por fim, o último item nos mostrará o relatório final da análise. Neste caso, a tabela de frequência ou o gráfico BOX PLOT. Tal tabela poderá ser editada se clicarmos duas vezes com o mouse sobre ela no campo da direita.

3.3. – Seleção de Casos para Análise

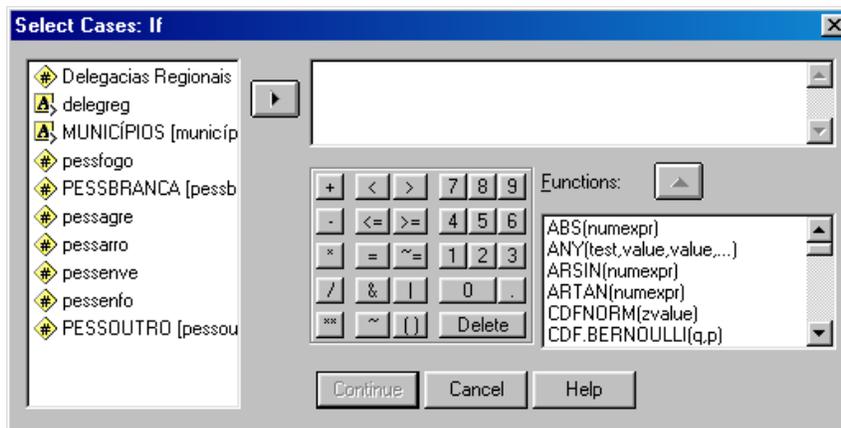
O recurso de seleção de casos para análise é muito utilizado, principalmente quando queremos restringir nossa análise a um grupo social específico dentre todos os presentes no conjunto total da amostra. Seu uso no desenvolvimento de análises comparativas é limitado pelo fato de que o *software* já traz uma série de recursos que possibilitam a efetivação desse uso de análise de uma maneira mais simplificada.

Para fazermos uma seleção de dados, devemos selecionar na barra de ferramentas **DATA** e depois **SELECT CASES**. O campo **Select** nos mostra 5 opções para seleção dos casos: **All Cases** (Todos os Casos), **If Condition is Satisfied** (Se Condição for Satisfeita), **Random Sample of Cases** (Amostragem Aleatória dos Casos), **Based on Case Range** (Baseado em Intervalo de Casos) e **User Filter Variable** (Uso de Variável Filtro).

A primeira opção - **All Cases** - nos permite trabalhar com todos os casos da amostra e é automaticamente definida pelo SPSS.



A segunda opção - **If Condition is Satisfied** - nos permite estabelecer uma condição em função de uma expressão matemática que deve ser satisfeita para que cada caso específico entre no grupo dos que serão analisados.

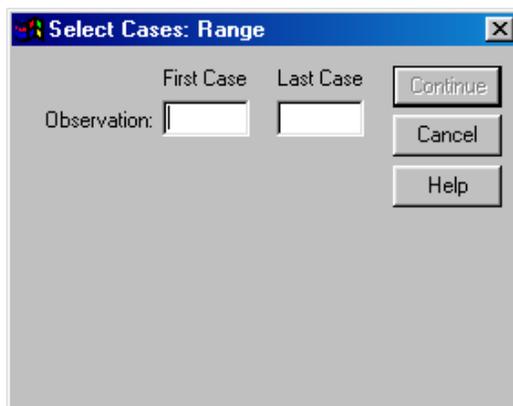


Esta tela é bastante semelhante àquela encontrada no comando **COMPUTE** na opção **IF** (condição). O retângulo à direita deverá ser preenchido com alguma expressão matemática que contenha pelo menos uma das variáveis presentes na lista à esquerda. Esta expressão agirá como um condicionante para que o caso seja inserido no grupo dos que serão analisados. Como exemplo de condição temos por exemplo: idade > 5 (a análise se restringirá às pessoas com mais de cinco anos de idade)

A terceira opção - **Random Sample of Cases** - nos permite escolher o número de casos a serem analisados em função de uma seleção aleatória simples. Poderemos indicar aproximadamente a porcentagem de casos a serem selecionados no total de casos ou o número exato de casos dentro de um número específico de primeiros casos; por exemplo: cinco casos dentro dos 100 primeiros.



A quarta opção - **Based on Case Range** - nos permite escolher os casos dentro de uma faixa específica de ordem de codificação.



Com base no código do caso - número do caso presente na margem esquerda da tabela de matriz de dados - indicaremos o intervalo de casos a serem selecionados.

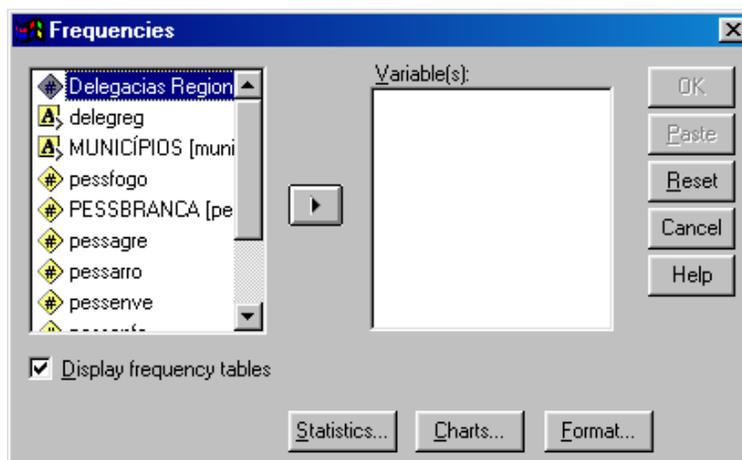
A quinta e última opção - **User Filter Variable** - nos permite selecionar os casos em função de uma variável filtro definida previamente. Esta opção exige uma variável de tipo especial (dummy) composta apenas de valores 0 e 1, onde os valores 1 serão selecionados e os valores 0 não serão selecionados.

3.4. - Análise Descritiva dos Dados

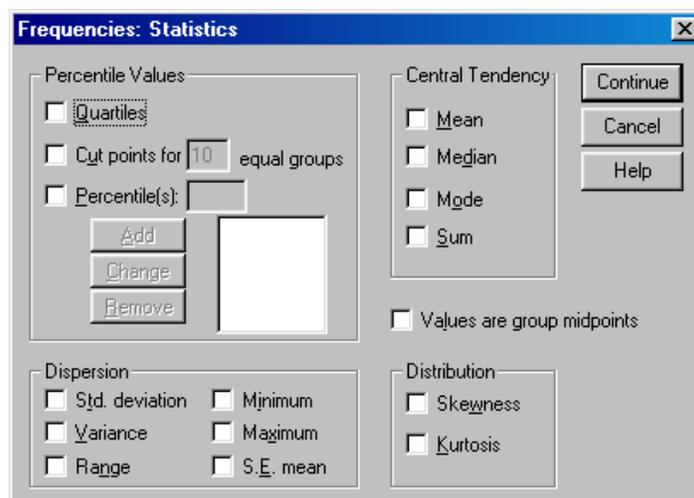
3.4.1. - Como Fazer uma Análise Descritiva Baseada em Medidas de Tendência Central e Dispersão (moda, média, mediana, desvio padrão, etc)

As análises de tendência central são muito importantes quando estamos trabalhando com variáveis ordinais. Os indicadores de tendência central são capazes de nos mostrar como uma certa variável ou característica do grupo estudado se distribui utilizando apenas um número. De um modo geral, dois fatores são importantes nas análises deste tipo: a avaliação da tendência central da distribuição e a avaliação da dispersão dos valores em torno desta tendência central.

Mostraremos aqui apenas um dos possíveis caminhos para se chegar a estas medidas de tendência central. Na barra de ferramentas selecione **STATISTICS**, depois **SUMMARIZE** e depois **FREQUENCIES**. Chegaremos, então, a seguinte figura:

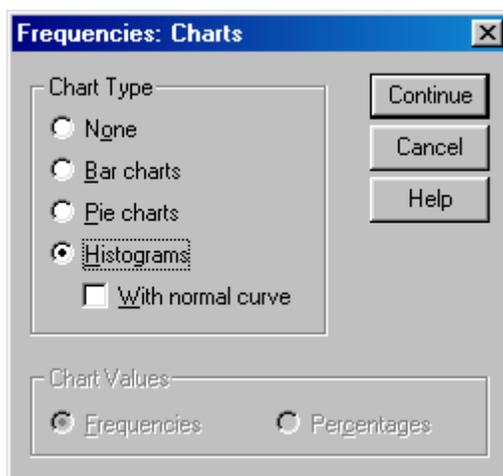


Este comando permite trabalharmos com a descrição da distribuição de valores de variáveis ordinais e categóricas. Seu padrão está direcionado para a análise de variáveis categóricas. Para conseguirmos indicadores de tendência central devemos selecionar no campo **STATISTICS** aqueles indicadores que consideramos importantes e desabilitar a opção **DISPLAY FREQUENCY TABLE** clicando sobre o campo. Ao clicarmos sobre o campo **STATISTICS** chegaremos a seguinte figura:



Tendo chegado a esta figura, basta selecionarmos entre as diversas opções existentes aquelas que desejamos. Entre as medidas de tendência central temos as seguintes opções: média, moda, mediana e soma; entre as medidas de dispersão dos valores temos as seguintes opções: desvio padrão, variância, intervalo, valor máximo e mínimo e média do erro padrão; em relação aos valores percentuais poderemos obter os quartis, os diversos percentis desejados e os valores que dividem a amostra no número de partes iguais desejadas⁴. Tendo feito a seleção das medidas desejadas, basta clicar em **CONTINUE**.

A opção **CHART** relaciona uma série de recursos para a visualização gráfica da distribuição de dados de variáveis. Dentre estas, a única apropriada para a visualização de variáveis ordinais é o **histograma**. Todas as outras opções são mais apropriadas para a análise de variáveis categóricas. Ao clicarmos na opção **CHART** chegaremos na figura abaixo, onde poderemos selecionar histogramas. Existirá ainda a alternativa de produzir a curva normal do gráfico, se selecionarmos a opção **With normal curve**.



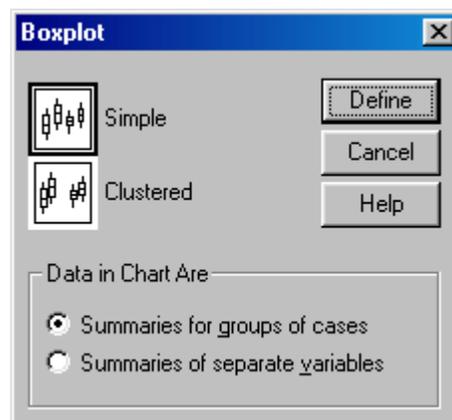
⁴ Os pontos de corte são obtidos a partir da partição da sequência, formada pelos valores ordenados da variável, no número de partes desejadas. Os valores obtidos como Cut Points são os valores que estão justamente no ponto em que essas partes são divididas.

A opção **FORMAT** relaciona uma série de recursos para a construção das tabelas de frequência, que como já foi dito antes se restringe a análise de variáveis categóricas. Estas duas opções serão descritas posteriormente quando tratarmos da análise da distribuição de variáveis categóricas.

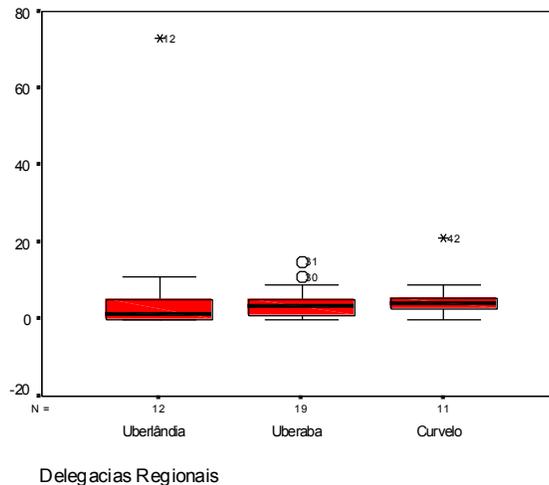
Tendo selecionado as opções de medidas centrais presentes no campo **STATISTICS** basta retornar a tela **FREQUENCIES** e selecionar as variáveis a serem analisadas, introduzindo-as no campo **VARIABLES**.

3.4.2. - Alternativa de Construção de Gráficos para Análises de Tendência Central

Além da opção **HISTOGRAMA** na tela **FREQUENCIES**, existe um segundo modo de chegarmos a uma representação gráfica de medidas de tendência central. Na barra de ferramentas escolheremos **GRAPHS**, depois **BOX PLOT** e chegaremos na seguinte figura:



Este gráfico nos permite fazer uma análise descritiva comparativa de distribuições de dados entre grupos de casos separados. Seus resultados são baseados na distribuição dos quartis e mediana. A definição do parâmetro de repartição da amostra é dado em função das categorias de uma segunda variável. Seria útil, como podemos ver abaixo, na comparação do nível de criminalidade em três delegacias diferentes.

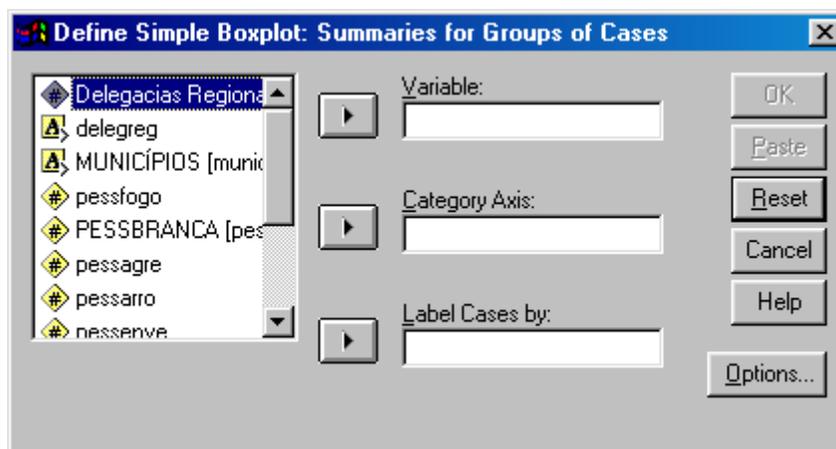


Estamos separando a análise da distribuição dos dados entre três delegacias: Uberaba, Uberlândia e Curvelo. O eixo Y do gráfico mostra os valores da variável PESSFOGO (Crimes contra pessoa utilizando arma de fogo). Dentro deste gráfico, a linha preta no interior do retângulo vermelho corresponde a mediana da distribuição em cada categoria de delegacia, os retângulos vermelhos compreendem a distribuição de 50 % dos casos e os traços pretos acima e abaixo deste retângulo compreendem os 50% restantes dos casos. Os pontos pretos acima e abaixo destes traços pretos são denominados *out liers* e podem ser definidos como os valores dos casos que se diferenciaram muito da distribuição central dos dados (estão fora da distribuição principal). Como podemos ver, temos neste gráfico não só uma idéia da tendência central da distribuição da variável, mas também, uma idéia da dispersão desta distribuição.

No menu para definição deste gráfico, poderemos escolher entre trabalhar com apenas uma variável no eixo X (simple) ou trabalhar com duas variáveis agregadas (Clustered). Poderemos, ainda, selecionar por trabalhar com a separação por grupos de casos (groups of cases) ou por variáveis (separate variables).

3.4.3. - Summaries for Groups of Cases

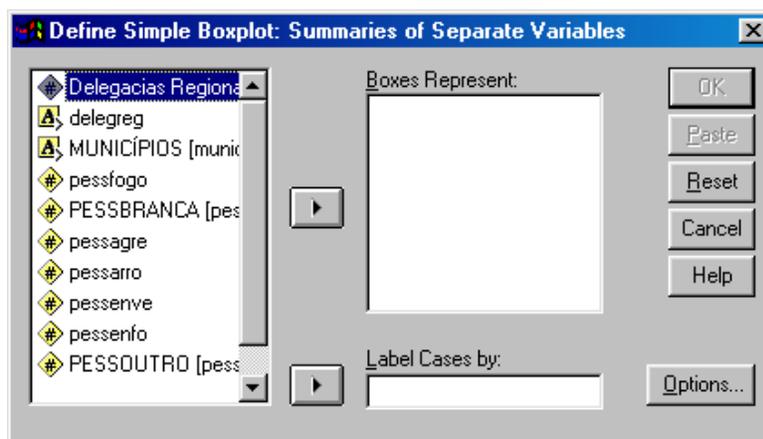
Se selecionarmos a opção **SIMPLE e Summaries for Groups of Cases**, chegaremos a seguinte tela:



Os campos **Variable**, **Category Axis** e **Label Cases By** devem ser preenchidos com as variáveis selecionadas presentes no retângulo da esquerda. Na construção do BOX PLOT, o campo **VARIABLE** deve ser preenchido com a variável em função da qual se deseja fazer a análise de tendência central comparando os diversos sub-grupos de casos e o campo **CATEGORY AXIS** deve ser preenchido com a variável em função da qual serão estabelecidos os sub-grupos de casos a serem comparados, por exemplo, Uberaba, Uberlândia e Curvelo. O preenchimento do campo **Label Cases By** não é obrigatório e a variável presente neste campo será usada para caracterização dos **outliers**.

Se continuarmos com a opção **Summaries for Groups of Cases** e selecionarmos a opção **Clustered**, além de preenchermos os campos indicados acima para a opção **Simple**, passaremos a ter que preencher o campo: **DEFINE CLUSTERS BY**. Este campo deverá ser preenchido com a variável em função da qual definiremos uma nova divisão dos casos dentro dos subgrupos já criados. Trabalharemos, deste modo, com grupos distintos dentro dos subgrupos já criados anteriormente.

3.4.4. - Summaries of Separate Variables



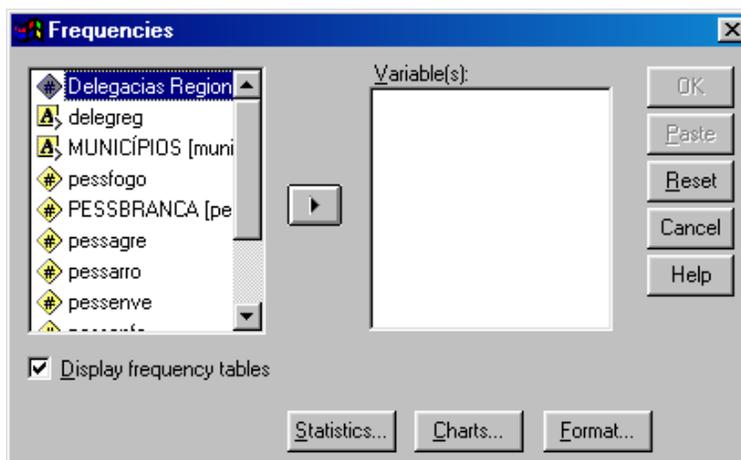
Se selecionarmos a opção **SIMPLE e Summaries of Separate Variables**, chegaremos a uma figura como a abaixo. Deveremos preencher o campo **BOXES**

REPRESENT com as variáveis a serem analisadas no gráfico. A análise passará a ser feita, então, para todo o conjunto de casos e para cada variável separadamente. Enquanto que a análise anterior era feita por grupos de casos diferentes e para apenas uma variável por vez. O campo **LABEL CASES BY** deverá ser preenchido como designado acima no caso do BOX PLOT por grupos de casos.

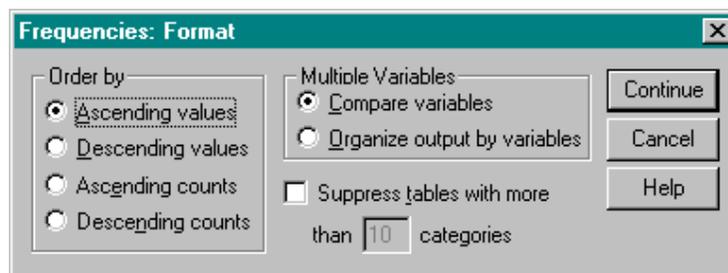
Se continuarmos com a opção **Summaries of Separate Variables** e selecionarmos a opção **Clustered**, além de preenchermos os campos indicados acima para a opção **Simple**, passaremos a ter que preencher o campo: **CATEGORY AXIS**. Este campo deverá ser preenchido com a variável em função da qual definiremos uma divisão dos casos dentro da análise das variáveis.

3.4.5. - Como Fazer uma Análise Descritiva Baseada na Distribuição de Frequência

Análises descritivas baseadas na distribuição de frequência são usadas para variáveis categóricas. Neste caso conseguimos saber a participação numérica e percentual de cada categoria no total de casos analisados por variável. Na barra de ferramentas devemos selecionar **STATISTICS**, depois **SUMMARIZE** e depois **FREQUENCIES**. O padrão do SPSS é conformar este comando para a análise de variáveis categóricas. Chegaremos a seguinte figura:



Após escolhermos a variável sobre a qual queremos construir a tabela de frequência e inseri-la no campo **VARIABLE(S)**, poderemos escolher no campo **FORMAT** algumas opções em relação ao formato desta tabela. Chegaremos a seguinte figura:



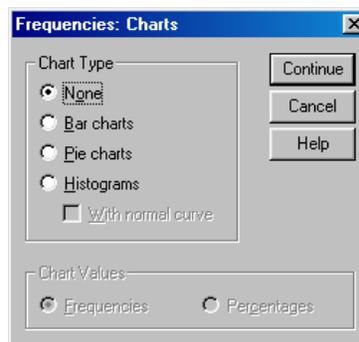
Poderemos escolher a ordem com que as categorias das variáveis serão mostradas na tabela:

- **ascending values:** categorias aparecem em ordem crescente do valor da categoria
- **descending values:** categorias aparecem em ordem decrescente do valor da categoria
- **ascending counts:** categorias aparecem em ordem crescente do valor do número total de casos por categoria
- **descending counts:** categorias aparecem em ordem decrescente do valor do número total de casos por categoria

Na opção **MULTIPLE VARIABLE** poderemos escolher entre incluir os resultados das análises de todas as variáveis trabalhadas em um mesmo relatório (*compare variable*) ou produzir um relatório por variável (*organize output by variable*).

Na opção **SUPRESS TABLES WITH MORE THAN** poderemos escolher por excluir do relatório aquelas tabelas que tenham um certo número definido de categorias.

Após escolhermos o formato de construção da tabela, poderemos também escolher por dar uma forma gráfica a estes resultados. Clicando na opção **CHART** chegaremos na figura abaixo, onde poderemos escolher por trabalhar com gráfico de barras, gráfico de *Pizza* ou histogramas. Poderemos ainda selecionar se os valores que aparecerão no gráfico serão os valores absolutos de casos das categorias (*frequencies*) ou os percentuais destes valores em relação ao número total de casos (*percentages*). Por fim, no caso de trabalharmos com o histograma, poderemos selecionar se queremos que apareça a curva da normalidade do gráfico (**With Normal Curve**).

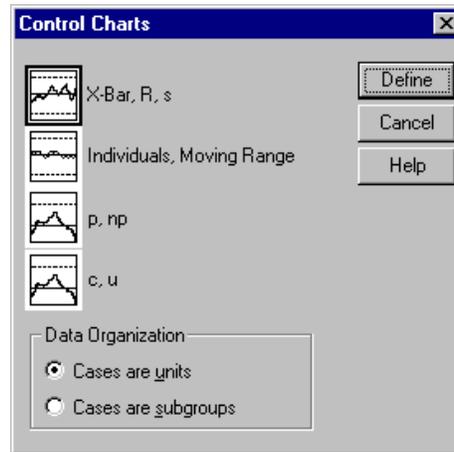


3.4.6. - Como Elaborar Análises Utilizando Gráficos de Controle de Processo

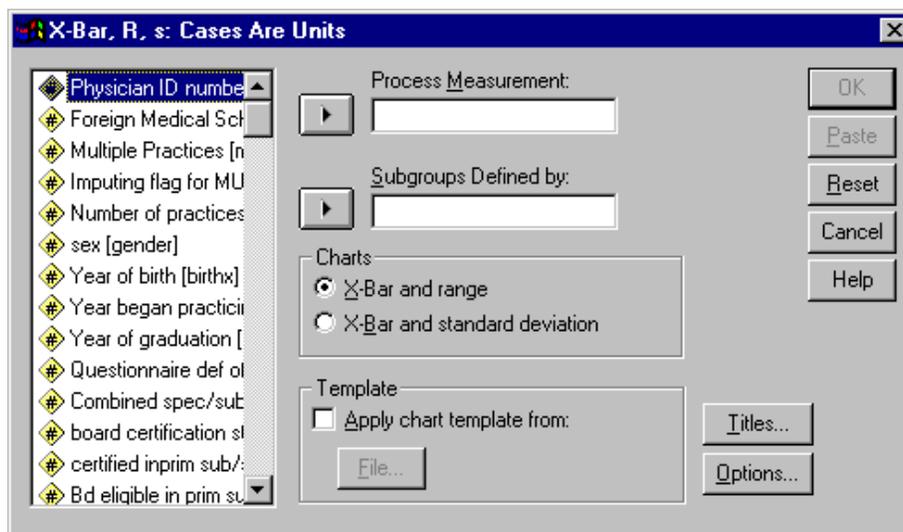
A análise de gráficos de controle de processo constitui um exemplo prático da utilização de medidas de tendência central e dispersão. Devido ao caráter introdutório desta apostila, vamos trabalhar apenas com a opção de análise: X-Bar, R, S. Na barra de ferramentas devemos selecionar **GRAPHS** e depois **CONTROL**. Chegaremos, então, à tela apresentada abaixo.

Selecionando a opção **DEFINE**, chegaremos à seguinte tela. Essa tela deverá ser preenchida da seguinte forma: o campo PROCESS MEASUREMENT

deverá ser preenchido com o nome da variável que estaremos acompanhando no processo de evolução e o campo SUBGROUPS DEFINED BY deverá ser preenchido com o nome da variável que utilizaremos para separar os instantes de observação dos nossos casos.



Na opção CHART, definiremos o tipo de gráfico que desejamos que seja apresentado pelo SPSS. Podemos seguir duas opções: X-Bar and Range e X-Bar and Standard Deviation.



Por fim, a opção TITLES nos permite editar os títulos a serem colocados no gráfico e a opção OPTIONS nos permitirá definir o intervalo máximo e mínimo esperado. Estes valores poderão ser estabelecidos através da definição do número de sigmas ou manualmente. Poderemos, ainda, definir o número de casos presentes em cada instante de observação através do MINIMUM SUBGROUP SIZE SAMPLE. Por fim, poderemos utilizar os valores contendo missing para definir um dos instantes de observação.

X-Bar, R, s: Options ✕

Number of Sigmas:

Specification Limits for X-Bar Chart

Upper:

Lower:

Minimum subgroup sample size:

Display subgroups defined by missing values

Continue

Cancel

Help

3.5. - Análise de Associação entre Variáveis

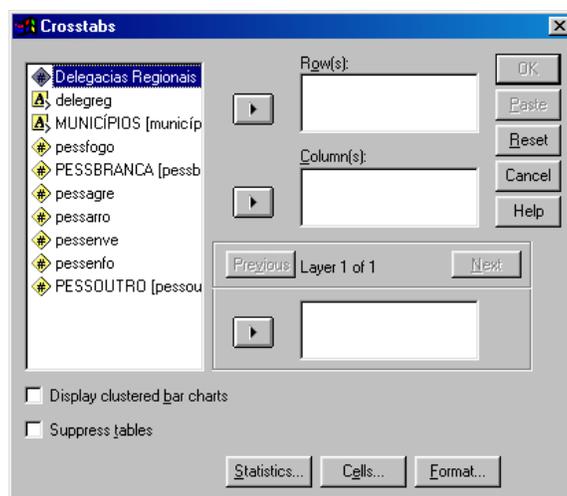
A análise de associação entre duas variáveis é um caso bastante simples de análise de associação entre variáveis, pois de um modo geral este fenômeno se estende por mais de duas variáveis. Para uma análise mais completa, levando várias variáveis em consideração ao mesmo tempo, devemos trabalhar com análises de regressão.

3.5.1. - Como construir uma tabela para verificar a relação entre variáveis

Um passo inicial para as análises de associação é a construção de tabelas de contingência, que tem o formato de (x) linhas por (y) colunas. O número de linhas e colunas é resultado do número de categorias das variáveis trabalhadas. Só se constrói tabelas que relacionem variáveis categóricas. Não é recomendado trabalhar com um número exagerado de linhas ou colunas, pois isso dificulta a análise da tabela.

Para realizar a construção da tabela devemos ir a barra de ferramentas e escolher a opção **STATISTICS**, depois **SUMMARIZE** e em seguida **CROSSTABS**. Chegaremos a uma figura como a abaixo.

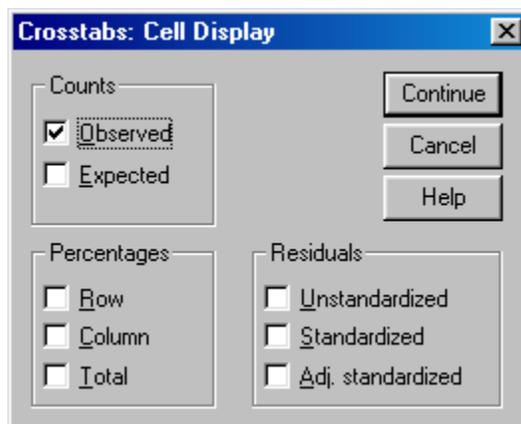
O campo **ROW** deverá ser preenchido com a variável a ser colocada na linha da tabela. O SPSS tem como padrão colocar sempre a variável independente neste campo. O campo **COLUMN** deverá ser preenchido com a variável a ser colocada na coluna da tabela. O SPSS tem como padrão colocar sempre a variável dependente neste campo. Note que estes campos poderão ser preenchidos com mais de uma variável. Neste caso, serão construídas, então, quantas tabelas forem necessárias envolvendo 2 variáveis para cumprir com as alternativas possíveis de cruzamento entre as variáveis indicadas. Por exemplo, se colocarmos a variável idade e educação na linha e sexo na coluna, chegaremos a duas tabelas: idade X sexo e educação X sexo.



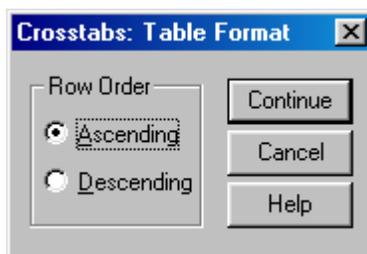
O retângulo abaixo do campo **COLUMN** poderá ser preenchido com uma nova variável que será utilizada na construção de uma tabela envolvendo três variáveis. Também este campo poderá ser preenchido com mais de uma variável.

Se a opção **Display Clustered Bar Charts** estiver selecionada, o relatório fornecido pelo SPSS fornecerá, além da tabela de cruzamento das variáveis, um gráfico de barras com o conteúdo da tabela.

Ao selecionarmos o campo **CELLS** chegaremos a figura abaixo, onde poderemos escolher se desejamos que a tabela mostre os valores absolutos observados no cruzamentos, os valores esperados em função da distribuição das marginais e os diversos percentuais de distribuição dos casos em função dos totais da linha, coluna ou da tabela e ainda os valores de resíduos padronizados e não padronizados.



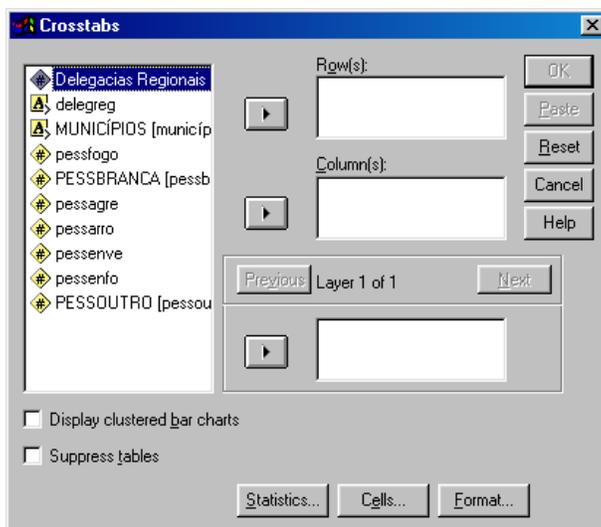
Se selecionarmos o campo **FORMAT** chegaremos a figura abaixo, onde poderemos selecionar a ordem de apresentação das categorias presentes na linha.



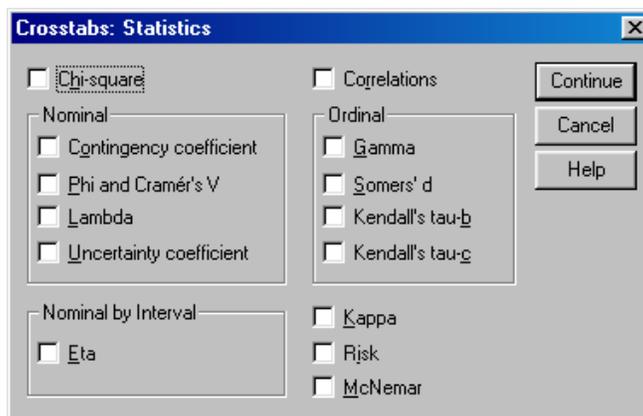
3.5.2. - Como construir análises de associação entre variáveis em função de coeficientes de correlação

A construção desta análise exige que se tenha um conhecimento mais aprofundado de estatística para que possamos selecionar de maneira mais adequada os coeficientes de correlação a serem analisados. Cada tipo de associação, dependendo do tipo de variável trabalhada, exige que se trabalhe com um coeficiente diferente. Toda análise desse tipo envolve, em geral, uma análise da dependência entre as variáveis e depois uma análise da intensidade da correlação. A análise da dependência lhe permite averiguar se a relação encontrada entre as variáveis decorre de uma simples coincidência dos casos analisados, ou seja, do processo de chance associado aos resultados. Já a análise da intensidade da correlação lhe indica a intensidade desta associação.

Para realizarmos este tipo de análise por coeficiente de correlação devemos selecionar a opção **STATISTICS** na barra de ferramentas, depois **SUMMARIZE** e depois **CROSSTABS**. Chegaremos na seguinte tela:



Se quisermos ver apenas os coeficientes de correlação, deveremos selecionar o campo **SUPPRESS TABLES**, o que fará com que a tabela de cruzamento das variáveis não seja produzida. Em seguida, deveremos selecionar o campo **STATISTICS** para podermos escolher o coeficientes com que queremos trabalhar. Surgirá, então, a seguinte figura:



A análise do **Chi-square** nos permite, como foi indicado acima, levantar em que medida a associação encontrada acima decorre do simples processo de chance relacionado aos resultados. O valor mostrado pelo programa para a associação investigada deverá, então, ser analisado dentro de uma tabela de Chi-square (presente em qualquer livro de estatística) para que possamos chegar a alguma conclusão a respeito da probabilidade do resultado ser decorrente de um processo de chance ou não.

Deveremos, então, selecionar os coeficientes de correlação desejados. Não se esqueça que cada um deles tem uma utilidade específica e os resultados de coeficientes diferentes não podem ser comparados entre si sem nenhum critério

sistemático. É recomendável que se faça toda a análise em função de apenas um coeficiente, para que seus valores sejam comparáveis com facilidade. O próprio SPSS trás na tabela acima, como podemos ver, algumas indicações para a escolha. Para trabalhar com variáveis nominais foram indicados os seguintes coeficientes: coeficiente de contingência, Phi, V de Cramer, coeficiente de incerteza e Lambda. Para trabalhar com variáveis ordinais foram indicados os seguintes coeficientes: Gamma, Somers e Kendall.

Os campos **ROW**, **COLUMN** e o retângulo abaixo do campo **COLUMN** deverão ser preenchido como especificado no item **Como construir uma tabela para verificar a relação entre variáveis**.

Enquanto as tabelas de cruzamento são restritas para a análise de variáveis categóricas, o uso de coeficientes de correlação pode ser desenvolvido para a análise de variáveis categóricas e variáveis ordinais. Existindo, até mesmo, coeficientes próprios para a análise de um e outro tipo de variável.

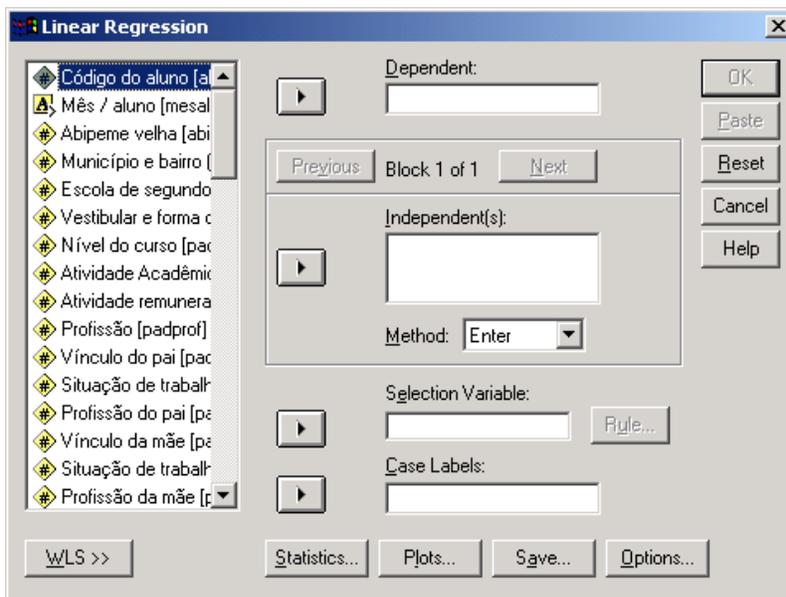
3.5.3. - Como Construir Análises de Regressão Linear

Análises de regressão são utilizadas em análises multivariadas. Existe uma série de pré-requisitos que devem ser observados antes da utilização desta técnica de análise. Nesta apostila temos como objetivo único apresentar os passos para elaboração da regressão através do SPSS.

O desenvolvimento de análises de regressão permite verificar a associação de mais de uma variável ao mesmo tempo com uma variável dependente final. A análise de regressão permite, assim, averiguar qual o grau de explicação que um certo conjunto de variáveis tem sobre um fenômeno estudado e o efeito separado de cada uma destas variáveis no acontecimento deste fenômeno.

Existem duas formas básicas para o desenvolvimento de regressões, uma adequada para situações onde as variáveis são escalares (Regressão Linear) e outra adequada para situações onde as variáveis são categóricas (Regressão Logística).

No menu **ANALYZE** devemos selecionar o menu **REGRESSION** e em seguida a opção **LINEAR**.

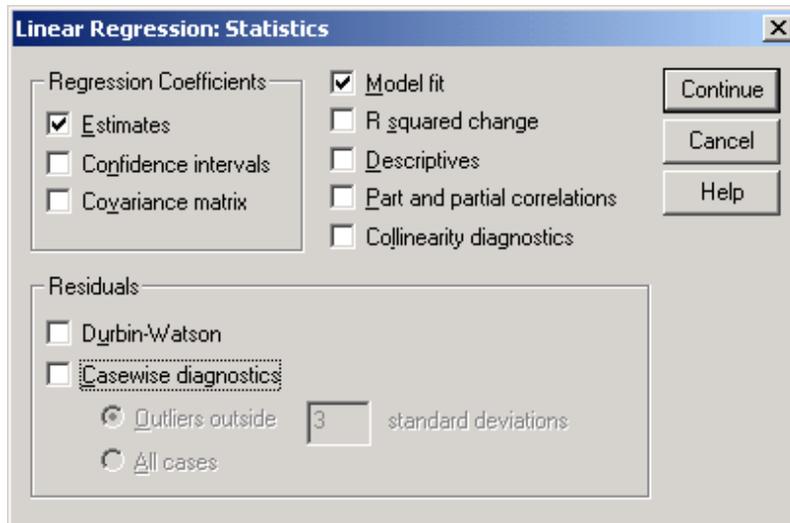


No campo **DEPENDENT** devemos indicar a variável a ser considerada como dependente (variável resposta) no modelo de regressão. No campo **INDEPENDENTS** devemos indicar as variáveis independentes no modelo de regressão. No campo **SELECTION VARIABLE** iremos indicar uma condição relacionada a uma variável para a seleção dos casos que serão avaliados para a realização da regressão. No campo **CASE LABELS** iremos selecionar o rótulo para ser atribuído aos casos destacados em todos os gráficos elaborados no relatório de análise.

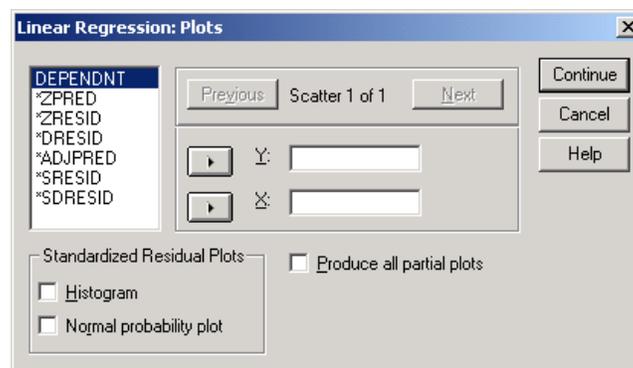
No campo **METHOD** devemos selecionar o método de inclusão das variáveis independentes na análise de regressão. O método **ENTER** faz a análise de apenas um modelo contendo todas as variáveis independentes de uma só vez. O método **FORWARD** começa pela elaboração da análise com um modelo que tem apenas 1 variável e é o mais explicativo e segue elaborando análises com modelos adicionando uma variável para cada modelo até incluir todas as variáveis independentes num modelo final. São tantos modelos, quantas variáveis existirem. O método **BACKWARD** faz o mesmo que o modelo **FORWARD** mas em sentido inverso, começa com um modelo completo até chegar a um modelo com apenas uma variável.

O botão **WLS** irá ativar a possibilidade de selecionarmos uma variável peso para ser inserida no processo de cálculo da regressão.

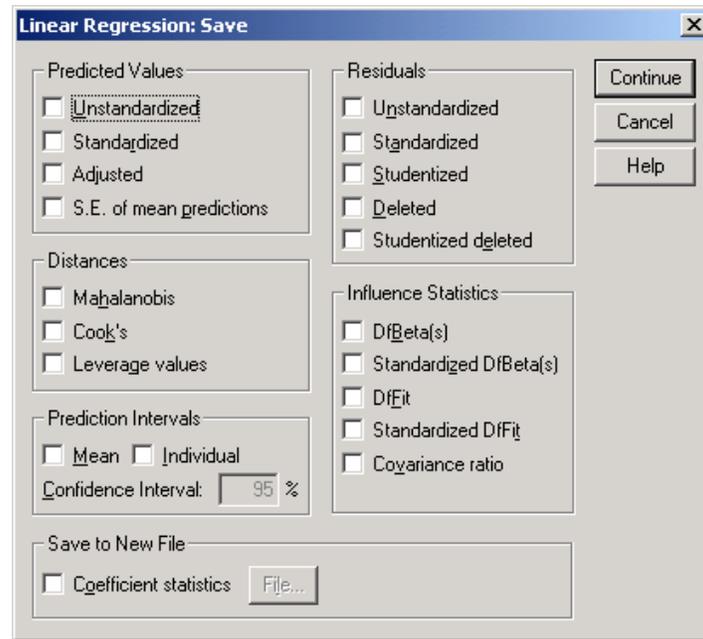
O botão **STATISTIC** permitirá a seleção de uma série de atributos a serem incluídos no relatório de análise relacionados à análise de resíduos, coeficientes de regressão e ajustes do modelo.



O botão **PLOTS** nos permitirá selecionar gráficos que serão incluídos no relatório estatístico da análise de regressão – resíduos, resíduos padronizados, valores preditos, variável dependente, etc.



O botão **SAVE** nos permitirá incluir na análise de regressão a geração de variáveis contendo os resíduos, valores preditos, distâncias, etc.



Por fim, o botão **OPTIONS** nos permitirá mudar os parâmetros de análise. Os valores de inclusão e exclusão em termos da estatística F, inclusão da constante na análise e opções para exclusão de casos em função da existência de missing values ou preenchimento com o valor da média dos outros casos.

