

APOSTILA PARA TRATAMENTO DE DADOS VIA *SPSS*

Ronaldo Pilati
Juliana B. Porto

ÍNDICE

<u>APRESENTACÃO</u>	<u>1</u>
<u>CAPÍTULO 1</u>	<u>1</u>
APRESENTAÇÃO DO SPSS	1
<u>CAPÍTULO 2</u>	<u>14</u>
DESCREVENDO DADOS NO SPSS	14
<u>CAPÍTULO 3</u>	<u>21</u>
EXPLORANDO E ADEQUANDO O ARQUIVO DE DADOS	21
<u>CAPÍTULO 4</u>	<u>34</u>
REALIZANDO ANÁLISES INFERENCIAIS: ANOVA UMA VIA, CORRELAÇÃO BIVARIADA, QUI- QUADRADO E REGRESSÃO LINEAR SIMPLES	34
<u>CAPÍTULO 5</u>	<u>39</u>
ANÁLISE FATORIAL EM AVALIAÇÃO DO TREINAMENTO	39
<u>CAPÍTULO 6</u>	<u>51</u>
GERAÇÃO E MANIPULAÇÃO DE SYNTAX	51

Apresentação

Nas últimas décadas o uso da microinformática tem facilitado enormemente o trabalho de organização, tratamento e análise de dados estatísticos de pesquisas em psicologia e outras áreas do conhecimento. Pelo fato do ferramental estatístico servir como uma auxílio para iluminar as informações advindas da realidade, esta tem sido uma estratégia secular de produção do conhecimento em psicologia e áreas afins. Boa parte do conhecimento que temos hoje sobre o comportamento e a psicologia humana se deve, em grande medida, ao uso da estatística na prática da pesquisa. A microinformática trouxe uma enorme capacidade de tratamento dos dados, graças ao incremento do processamento dos hardwares, a miniaturização dos computadores e ao desenvolvimento de programas cada vez mais fáceis de serem utilizados e poderosos em seus recursos e capacidades gráficas e de cálculo.

Atualmente há uma grande quantidade de softwares estatísticos que podem ser eficientemente utilizados para tratamento de dados nas pesquisas de psicologia e ciências humanas, como o Excel, o BMDP, Minitab, NCCSS, SYSTAT, Statplus, SAS, SPSS, entre centenas de outros. Há também softwares estatísticos gratuitos, vários com código livre, como EPINFO e o R. Há muitos programas que são especializados em algumas análises e outros, como a lista citada, que são gerais. Se você for curioso poderá encontrar dezenas de softwares estatísticos em uma simples busca na internet.

Se por um lado o avanço da microinformática gerou muitas possibilidades para a pesquisa em psicologia e áreas afins, também potencializou o mau uso da estatística. A máxima ‘lixo entra, lixo sai’ foi significativamente potencializada nos tempos modernos. Por este motivo, não é suficiente que o usuário saiba operar o software, mas sim que seja capaz de compreender e analisar os resultados que o programa oferece, além, é claro, de organizar as informações a serem inseridas no programa de análise de dados.

Esta apostila tem como objetivo ser um pequeno manual introdutório ao SPSS, um pacote estatístico que se tornou o principal software em uso entre psicólogos sociais, do trabalho e organizacionais no Brasil. Especificamente no departamento de Psicologia Social e do Trabalho da UnB este programa é utilizado há mais de 30 anos, o que significa que várias gerações de pesquisadores acabaram sendo formados em análise de

dados com o uso deste programa. Isto não quer dizer que ele seja o melhor daqueles tantos oferecidos no mercado, mas é um programa que oferece várias vantagens apreciadas pelos pesquisadores em ciências sociais, como menus fáceis de serem utilizados e a não necessidade de uso de linguagem de programação ou de conhecimento sofisticado de matemática para a realização das análises estatísticas. Além disto, o programa oferece uma ampla variedade de módulos, que permitem um grande número de análises estatísticas multivariadas, boa parte das quais freqüentes na pesquisa científica em psicologia. Dentro deste ‘zeitgeist’ é que este manual foi concebido. É também neste contexto que ele deve ser utilizado, pois pretende ser um pequeno manual introdutório àqueles que têm a pretensão de fazer pesquisa em psicologia social e áreas afins, de forma a abrir o vasto mundo de utilização da ferramenta estatística SPSS.

Mas atenção! Este manual não tem a pretensão de substituir a aprendizagem séria e comprometida do pensamento estatístico. Muito pelo contrário, este manual é apenas uma leitura complementar para uso do SPSS. Portanto, foi concebido como uma ferramenta extra para servir de apoio ao ensino de estatística, feito com, no mínimo, o uso de um livro texto de estatística voltado para cientistas sociais. Isto significa que o manual não pretende substituir a leitura de um bom livro introdutório ou de análise multivariada de dados, muito menos a realização de exercícios de aplicação de conhecimentos estatísticos, mas sim auxiliar o aluno na realização de procedimentos estatísticos via SPSS. Considera que o desenvolvimento destas competências é parte essencial da formação de pesquisa em psicologia.

Vale lembrar que o manual foi desenvolvido com base na versão 10.0 do SPSS, o que significa que várias dos aprimoramentos das versões posteriores não foram, necessariamente, implementadas no manual. De qualquer forma considera-se que a lógica geral do programa permanece a mesma para as versões posteriores, o que ainda o deixa atualizado. Uma boa leitura!

CAPÍTULO 1

Apresentação do SPSS

Conteúdo:

- 1. Apresentação das Janelas de Trabalho**
- 2. Descrição dos Menus de Comando**

Objetivos:

Ao final do capítulo o aluno deverá ser capaz de:

- 1. Identificar o SPSS**
- 2. Descrever cada uma das janelas de trabalho do programa**
- 3. Montar arquivos de dados**
- 4. Operar comandos básicos dos menus do SPSS**

JANELAS DE TRABALHO

O SPSS é um pacote estatístico, composto de diferentes módulos, desenvolvido para o uso em ciências humanas, sociais, biológicas, entre outras área do conhecimento. Está baseado no ambiente Windows, sendo de fácil operação e muito abrangente, pois permite realizar uma grande amplitude de análises estatísticas e gráficas (análises descritivas, análises inferenciais multivariadas, módulos gráficos, entre outros).

A principal janela de trabalho do software é apresentada abaixo. A Figura 1.1 mostra a janela de digitação (ou *input*) de dados **SPSS Data Editor**, na qual os bancos de dados são gerados e analisados. Na janela **SPSS Data Editor** as linhas são relativas aos casos, participantes ou respondentes e as colunas relativas as variáveis investigadas. Esta janela gera um dos três tipos de arquivos associados ao SPSS. Esse arquivo tem terminação *.sav* e armazena todas as informações relativas ao banco de dados, como definição de variáveis e os dados digitados.

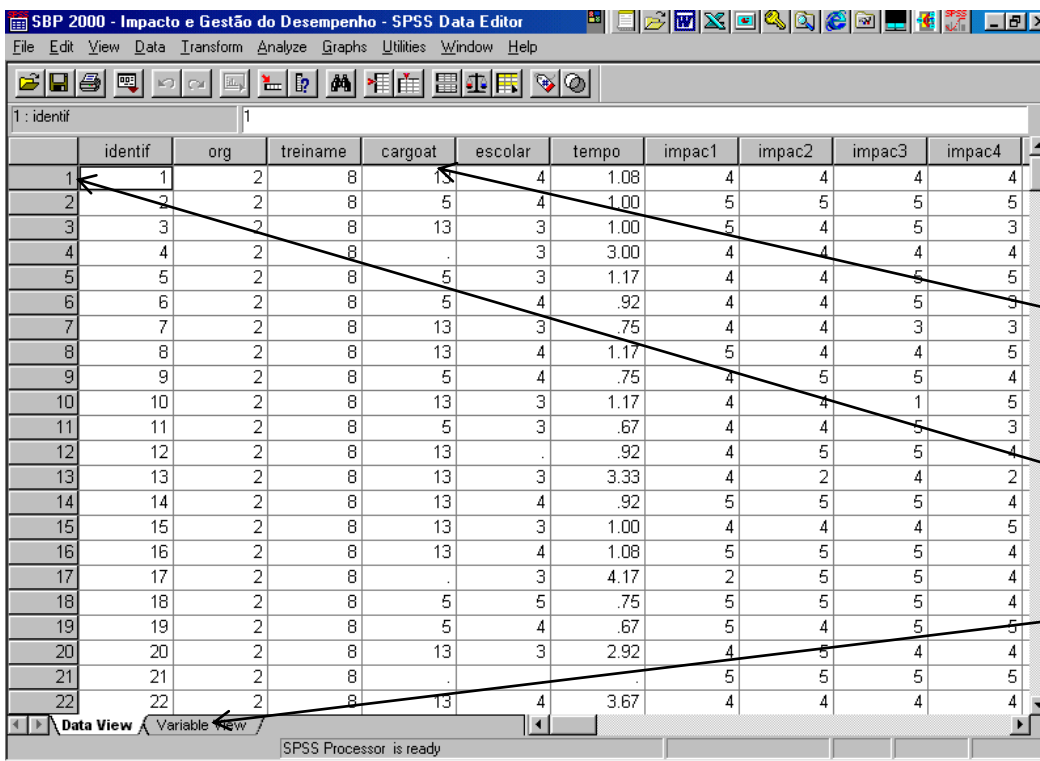


Figura 1.1 – Janela de input de dados do SPSS (v.10.0).

Colunas: representam as variáveis em um banco de dados.

Linhas: representam os casos em um banco de dados

Acesso a janela de definição dos parâmetros das variáveis.

Teoricamente, no SPSS, não existe limitação do número de colunas, muito menos do número de casos, possibilitando a análise de bancos de dados de qualquer tamanho.

Os bancos de dados são montados no próprio SPSS ou importados de outras planilhas. O SPSS importa praticamente todos os tipos de arquivos utilizados no mercado, além de trabalhar com arquivos de texto e *.dat*. Junto com a digitação dos dados é necessário definir cada uma das variáveis colocadas nas colunas da janela **SPSS Data Editor**.

Nesta versão do SPSS, a definição das variáveis do banco de dados é feita em outra janela também denominada **SPSS Data Editor** (Figura 1.2). Para tornar esta janela ativa, deve-se clicar na opção **Variable View**, disposta no lado inferior esquerdo da barra de rolagem da janela da Figura 1.1.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	
1	identif	Numeric	8	0	Identificação	None	None	8	Rigl
2	orga	Numeric	8	0	Organização	{1, UnB}...	None	8	Rigl
3	treiname	Numeric	8	0	Treinamento	{1, Treinament	None	8	Rigl
4	cargaot	Numeric	8	0		{1, Analista R	None	8	Rigl
5	escolar	Numeric	8	0	Escolaridade	{1, Primeiro in	None	8	Rigl
6	tempo	Numeric	8	2	Tempo de Em	None	None	8	Rigl
7	impac1	Numeric	8	0		{1, Discordo T	None	8	Rigl
8	impac2	Numeric	8	0		{1, Discordo T	None	8	Rigl
9	impac3	Numeric	8	0		{1, Discordo T	None	8	Rigl
10	impac4	Numeric	8	0		{1, Discordo T	None	8	Rigl
11	impac5	Numeric	8	0		{1, Discordo T	None	8	Rigl
12	impac6	Numeric	8	0		{1, Discordo T	None	8	Rigl
13	impac7	Numeric	8	0		{1, Discordo T	None	8	Rigl
14	impac8	Numeric	8	0		{1, Discordo T	None	8	Rigl
15	impac9	Numeric	8	0		{1, Discordo T	None	8	Rigl
16	impac10	Numeric	8	0		{1, Discordo T	None	8	Rigl
17	impac11	Numeric	8	0		{1, Discordo T	None	8	Rigl
18	impac12	Numeric	8	0		{1, Discordo T	None	8	Rigl
19	impac13	Numeric	8	0		{1, Discordo T	None	8	Rigl
20	cartra1	Numeric	8	0		{1, Discordo T	None	8	Rigl
21	cartra2	Numeric	8	0		{1, Discordo T	None	8	Rigl
22	cartra3	Numeric	8	0		{1, Discordo T	None	8	Rigl
23	cartra4	Numeric	8	0		{1, Discordo T	None	8	Rigl

Figura 1.2 – Janela de definição de variáveis.

Colunas: representam cada uma das características das variáveis a serem definidas.

Linhas: representam cada uma das variáveis que compõem um banco de dados.

A janela acima apresentada na Figura 1.2 possui várias colunas relativas aos principais parâmetros de cada uma das variáveis da planilha. São oito colunas:

Name: refere-se ao nome da variável, composto de até oito caracteres (a depender da versão do programa), que será colocado nas colunas da janela de *input* de dados. Deve-se clicar em qualquer cela dessa coluna para que se possa digitar o nome da variável;

Type: refere-se ao tipo de variável, ou seja, sua característica de notação (numérica, qualitativa, etc). Clicando-se na cela desta coluna abre-se a caixa de diálogo da Figura 1.3. Observe que a caixa de diálogo Variable Type permite a escolha de vários tipos de variáveis. A definição de cada uma pode ser acessada clicando-se no botão direito do mouse colocado em cima da opção (ver reprodução na Figura 1.3). Esta mesma caixa de diálogo também permite a escolha de outras características da variável (width e decimal places – ver definições abaixo).

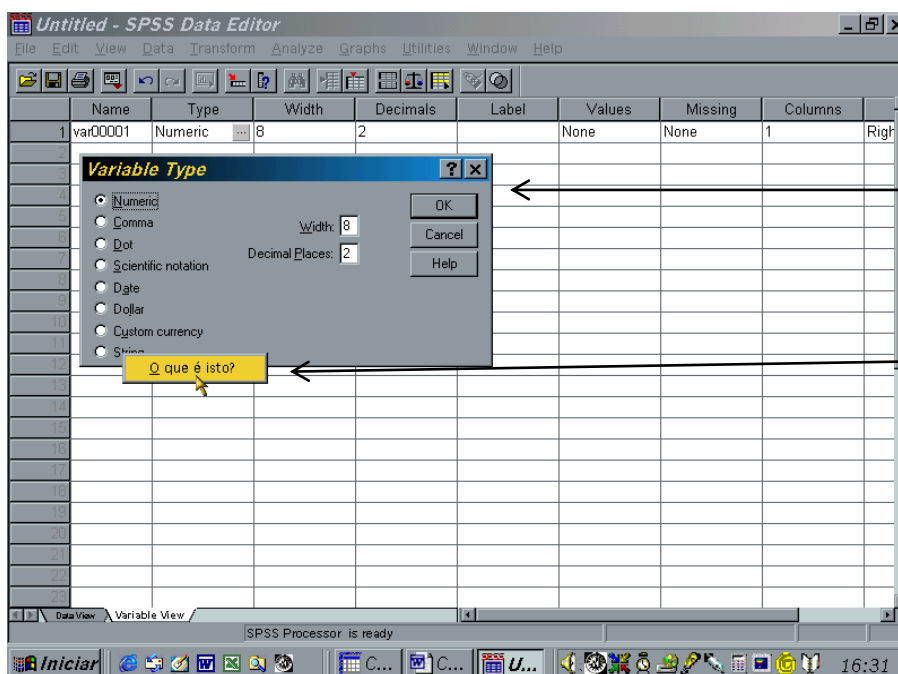


Figura 1.3 – Caixa de diálogo Variable Type..

Caixa de diálogo de definição de tipo da variável.

Clicando-se com o botão da direita do mouse abre-se esta janela explicativa.

Width: diz respeito ao número de caracteres da variável. Pode ser definido diretamente na cela ou através da caixa de diálogo **Variable Type** reproduzida na Figura 1.3.

Decimals define o número de casas decimais, a direita da vírgula, que serão apresentadas. Da mesma forma que o **Width**, pode ser definido diretamente na cela ou na caixa de diálogo **Variable Type**.

Label define o nome atribuído a uma variável e não possui restrição de número de caracteres (*e.g.*, idade, escolaridade, descrição do item de um questionário, entre outras);

Value refere-se aos valores que os *labels* podem assumir. Neste parâmetro o pesquisador deve definir todos os possíveis valores que uma variável pode assumir. Um tipo comum nas ciências sociais é a escala *Likert* (1 = discordo totalmente; 2 = discordo...);

Missing define o tipo de tratamento para os casos ausentes que o pesquisador quer adotar. O Default do programa geralmente é usado, mas outros valores podem ser definidos (*e.g.* 999 ou 99);

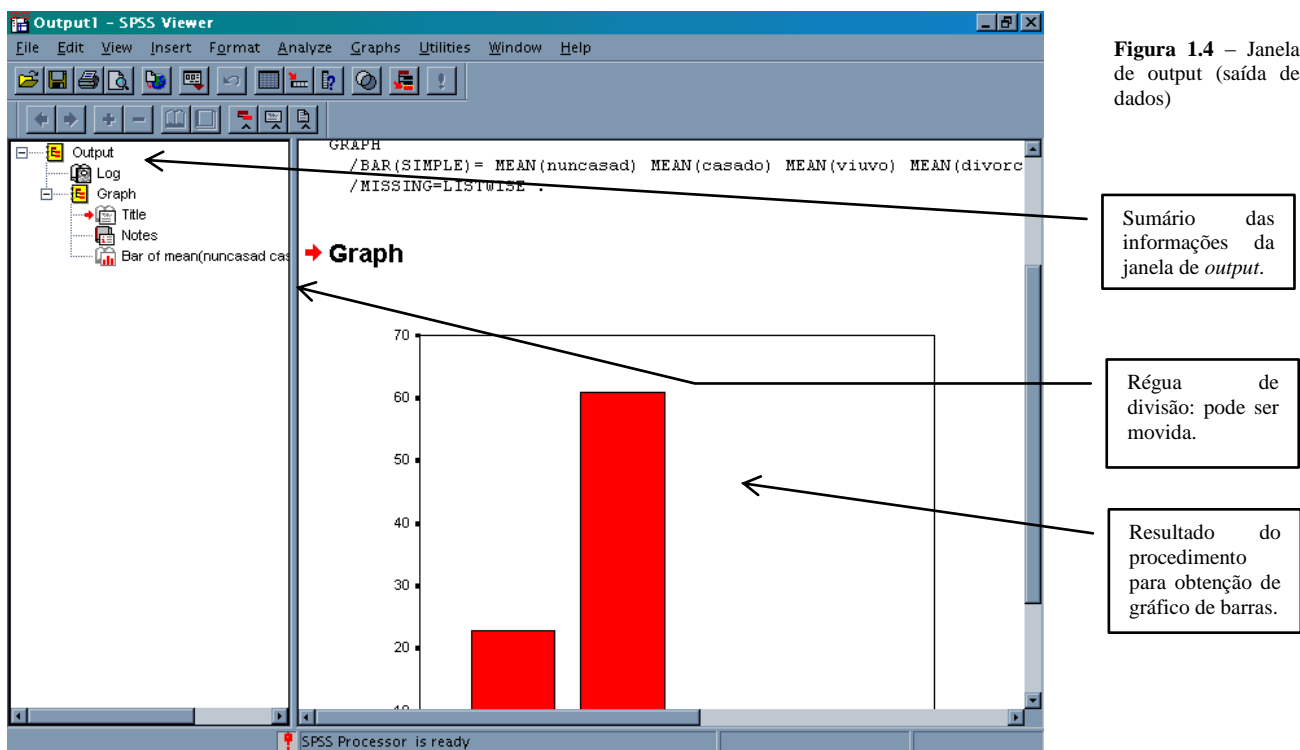
Columns define o tamanho da coluna da variável, que será apresentada na janela de *input* de dados;

Align define o alinhamento dentro de cada célula da planilha de dados;

Measure que informa o tipo de variável (contínua ou discreta) sendo estudada pelo pesquisador. Esta definição é fundamental, pois o software habilitará o uso das variáveis em certos procedimentos a partir do tipo de medida (*measure*) selecionada.

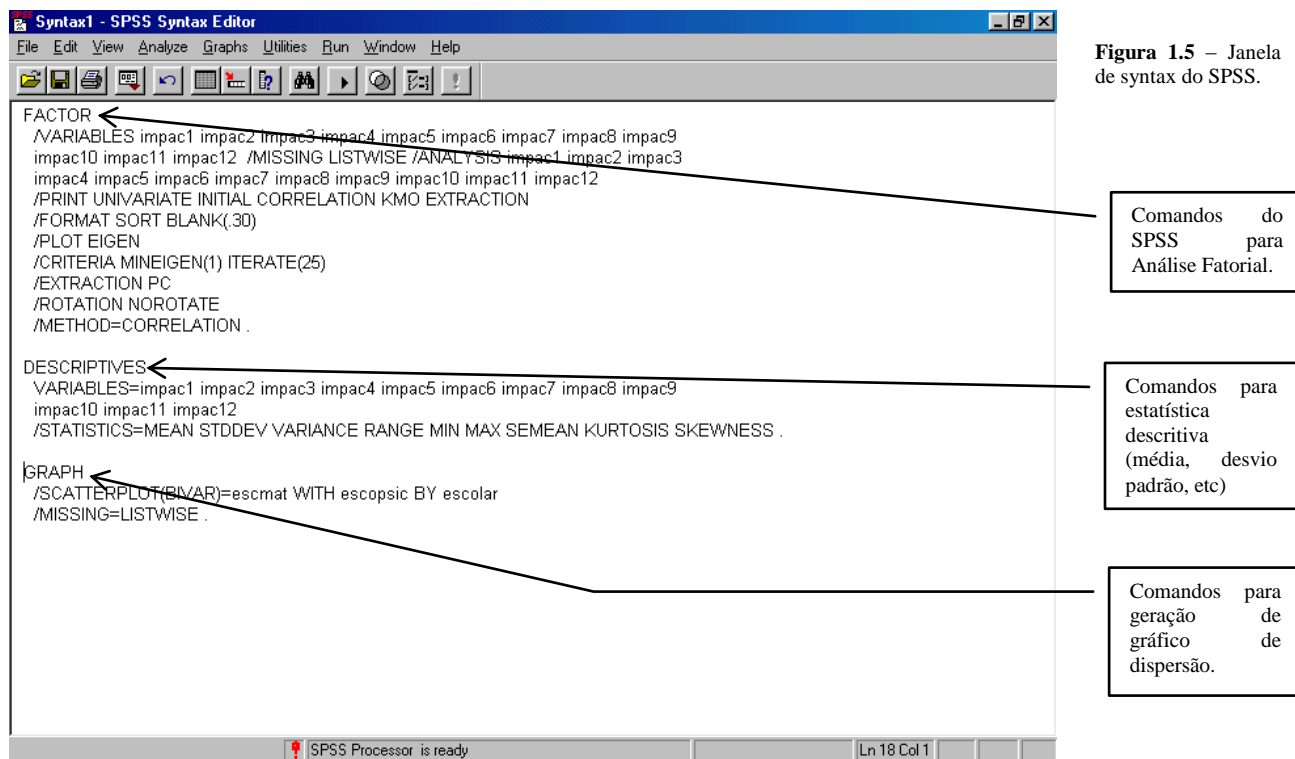
Na construção de um arquivo de dados para digitação de dados é importante criar como primeira variável uma coluna de identificação. A função desta é criar um rótulo (variável nominal) de identificação do questionário dentro do banco de dados. A Figura 1.2 apresenta, na primeira coluna, uma variável denominada *identif*. Este rótulo é importante para que o questionário possa ser identificado a qualquer momento, para correção de erros de digitação ou retorno ao questionário para coletar qualquer tipo de informação. A contagem automática do SPSS não é confiável, pois a exclusão de qualquer caso acarretará na modificação da numeração de todos os subsequentes.

O SPSS possui ainda mais janelas de trabalho. A próxima a ser apresentada refere-se ao *output* do programa. Nesta janela são geradas as tabelas e os gráficos processados a pedido do usuário. Quando qualquer tipo de procedimento é executado, mesmo na janela de *input*, os resultados são gerados na janela de *output*, que se abre automaticamente quando o procedimento de análise é solicitado. Essa janela é, também, um outro tipo de arquivo, com terminação *.spo*. É importante o usuário ter em mente que deve executar o procedimento de nomeação e salvamento desta janela, para que se torne uma janela de resultados do SPSS. A Figura 1.4 apresenta esta janela do programa.



A outra janela de trabalho do SPSS é a de *syntax*. Esta janela possibilita a execução de análises estatísticas e a geração de gráficos a partir dos códigos de comando do programa. Este tipo de janela é muito útil quando o padrão de análise de dados é extremamente complexo e/ou exige que as análises sejam feitas com frequência (como aquelas para relatórios periódicos) ou quando as mesmas análises devem ser feitas com diferentes variáveis, o que possibilita apenas que o nome das variáveis seja modificado, sem precisar repetir os procedimentos nos menus do programa. Também é importante ressaltar que, a exemplo da janela de *output*, a janela de *syntax* também é um arquivo

diferenciado, com terminação *.spss*. A Figura 1.5 apresenta esta janela de trabalho do SPSS.

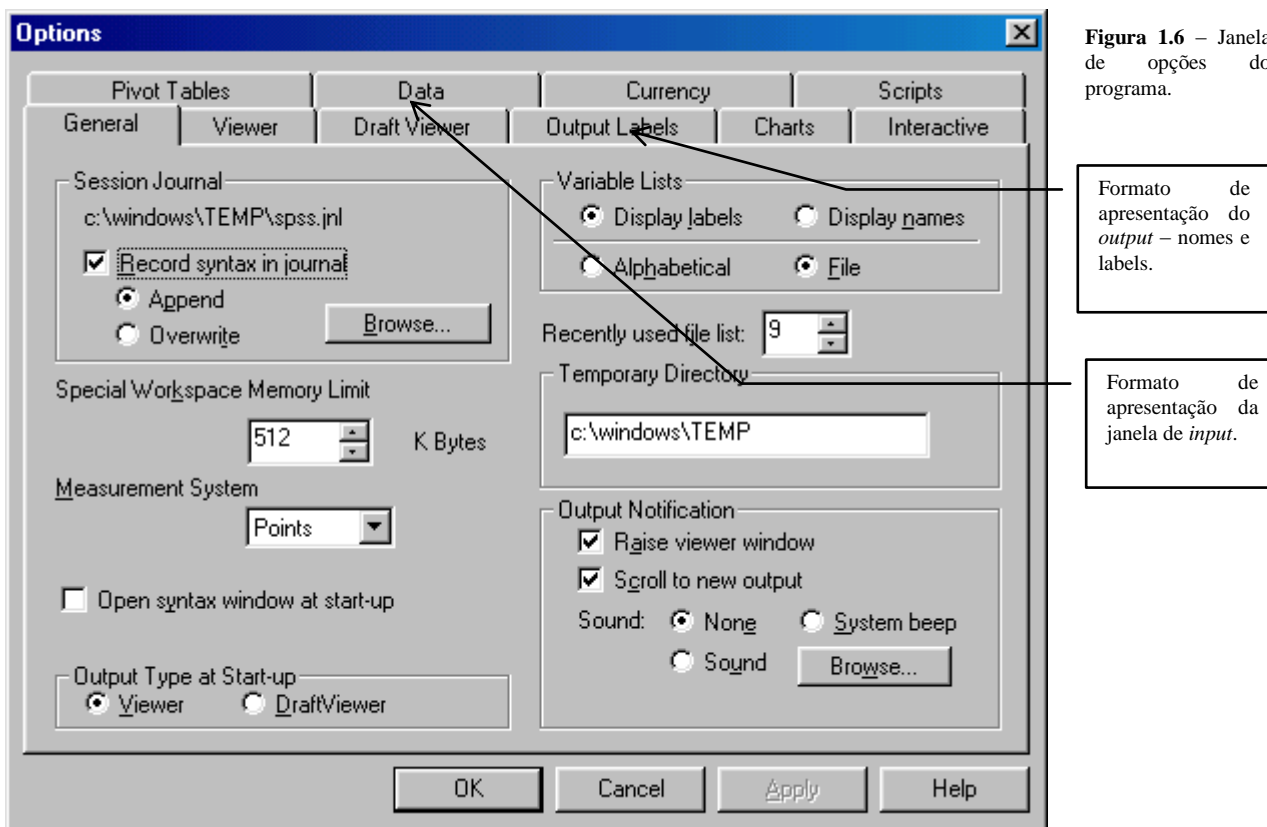


MENUS DE COMANDO

Na janela **SPSS Data Editor** o programa possui uma série de menus de comandos, que possibilitam a manipulação e análise dos dados, bem como os procedimentos do windows para trabalhar com arquivos.

O primeiro desses menus é denominado *File* e tem como principais funções a abertura, salvamento, importação, além de outras funções comuns aos diferentes programas da mesma plataforma operacional ou específicos do SPSS.

O segundo menu é o *Edit*, que possui as seguintes funções: voltar a última ação, copiar, colar, cortar, procurar (*find*) e opções. O menu opções tem como objetivo definir diferentes parâmetros do programa, como formato de apresentação dos dados digitados na SPSS Data Editor, o formato de geração das tabelas na janela de *output* de dados, entre outras funções. A Figura 1.6 apresenta a janela de diálogo do menu *edit/opções*.



Este menu de opções é importante para adaptar o padrão de apresentação do programa às características ou preferências de trabalho do usuário. Por exemplo, na pasta *output labels* é importante a definição de apresentação do nome das variáveis e também de seus *labels*, para que a tabela de resultados já possa fornecer todas as informações necessárias para a interpretação dos resultados das análises

O próximo menu da barra de comandos é denominada *Data*. Esse menu possibilita manipular o arquivo de dados de diferentes maneiras. A figura 1.7 apresenta a tela do SPSS com este menu aberto.

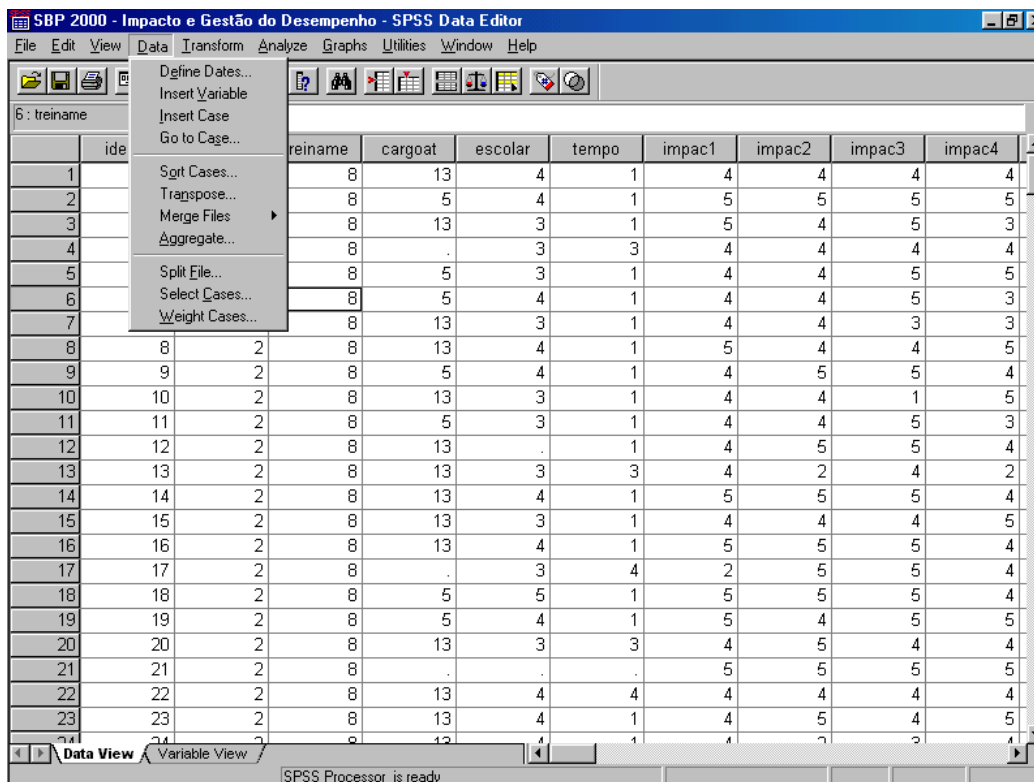


Figura 1.7 – Menu Data do SPSS.

Como pode ser observado na Figura 1.7 o menu Data possui vários comandos.

Estes serão apresentados em tópicos a seguir:

- O comando *insert variable* permite a inserção de uma nova coluna na planilha de dados, ao lado esquerdo da coluna onde se encontra a cela ativa da planilha;
- o comando *insert case* irá inserir uma nova linha em uma planilha de dados, acima da linha em que está a cela ativa da planilha;
- o *sort cases* realiza um ordenamento de todos os casos de uma planilha (linhas) a partir da seleção de uma das colunas do arquivo de dados; este ordenamento pode ser em ordem ascendente ou descendente, como a especificação apontada na janela de diálogo deste comando;
- o comando *merge files*, que possibilita a inserção de novos casos (linhas) ou variáveis (colunas) provenientes de um outro arquivo de dados;
- o *split files* possibilita a separação do banco de dados a partir de uma das variáveis deste banco. Quando este comando está acionado, por exemplo, as análises resultantes serão apresentadas para cada uma das categorias definidas por uma variável. Se for acionado o comando *split files* para a variável

escolaridade na Figura 1.7 serão apresentados os resultados para os sujeitos pertencentes a cada grupo de escolaridade;

- o comando *select cases* possibilita a seleção de um grupo de casos em um mesmo arquivo ou a criação de um outro arquivo a partir de um grupo de casos inicial.

O próximo menu da barra de menus do SPSS é denominado *Transform*. Este menu possibilita a transformação de variáveis do arquivo, assim como a criação de novas variáveis no arquivo de dados a partir de variáveis já presentes no mesmo. A Figura 1.8 apresenta o referido menu na janela de *input* do SPSS.

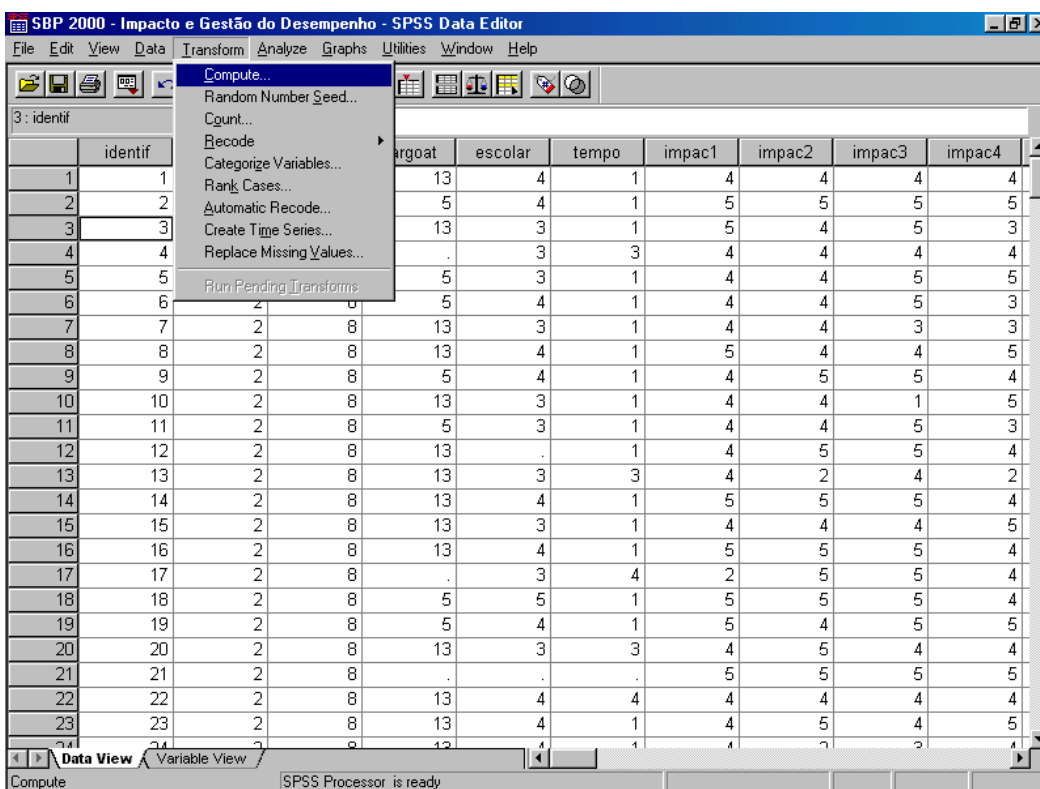


Figura 1.8 – Menu Transform do SPSS

A seguir são apresentados os comandos desse menu.

- *compute* é um comando de geração de novas variáveis (ou sobreposição de antigas) a partir de variáveis já existentes no arquivo de dados. Este comando é que possibilitará, por exemplo, a geração dos escores fatoriais, gerados a partir da média aritmética de cada um dos itens componentes do fator. Ele permite a composição e manipulação de variáveis por meio de dezenas de funções matemáticas;

- o comando *recode* possibilita a recodificação de valores (*labels*) de uma variável nela mesma ou na criação de uma nova variável;
- *automatic recode* possibilita a recodificação de variáveis em novas variáveis, mas de forma automática, diferentemente do que possibilita o comando *recode*, onde a estipulação dos valores é manual;
- *create time series* possibilita a geração de uma variável de série temporal dentro de um arquivo de dados já existentes;
- o comando *replace missing values* realiza substituições automáticas de casos omissos nas variáveis selecionadas. Esta substituição pode ser feita pelo valor médio ou mediano, além de outros métodos. Também é importante ressaltar que o comando gera um novo conjunto de variáveis, onde a substituição é apresentada.

O próximo menu de comando é denominado *Analyze*. Neste menu todos os procedimentos estatísticos estão concentrados. Aqui são apresentadas todas as técnicas descritivas e inferenciais, univariadas e multivariadas. Vale uma ressalva: as técnicas descritivas são aquelas que sintetizam os dados brutos, apresentando os padrões de dispersão (*e.g.* variância, desvio padrão) e tendência central (*e.g.* média, moda e mediana), assim como as medidas de posicionamento (*e.g.* quartil, decil, percentil). As análises inferenciais são aquelas que aplicam modelos probabilísticos, estimando a chance do comportamento observado na amostra se repetir na população. Quanto a técnicas univariadas, estas levam em consideração apenas uma variável e as multivariadas levam em consideração múltiplas variáveis.

São comandos do menu *analyze: report cases, descriptive statistics, compare means, correlate, regression, data reduction, scale*, entre outros. Durante este curso serão abordadas técnicas estatísticas básicas que se encontram nos comandos *descriptives statistics, correlate e compare means*. Estes comandos serão explorados em detalhe no capítulo 3.

O próximo menu é denominado *graph*. Este é relativo aos módulos de geração de gráficos do SPSS, que apresenta uma grande quantidade de recursos de geração e

manipulação de dados a partir de gráficos. A Figura 1.9 apresenta a tela do SPSS com esse menu aberto.

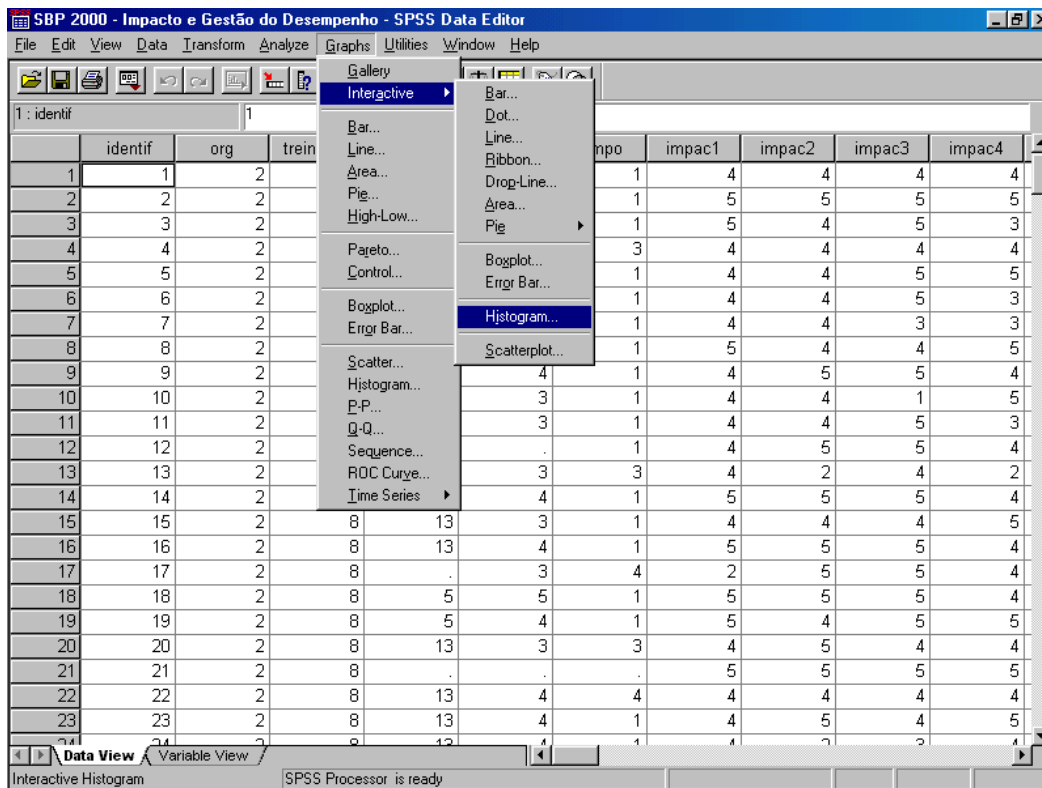


Figura 1.9 – Menu Graph do SPSS.

São tipos de gráficos gerados por este menu do SPSS:

- *bar* apresenta gráficos de barras, ideais para comparação de categorias de uma variável categórica, como escolaridade, sexo, entre outras. Esse tipo de apresentação também é importante para aquelas variáveis categóricas que não tem um somatório total de 100%;
- *pie* é um tipo de gráfico conhecido por pizza. Ele é ideal para variáveis discretas (categóricas) que possuem somatório total de categorias de 100%;
- *box-plot* é um gráfico que apresenta o tipo de dispersão de uma variável, apresentando o mínimo, o máximo, o primeiro e o terceiro quartil e a mediana em uma única representação gráfica. É direcionado para variáveis contínuas e ideal para observar casos extremos univariados. O SPSS trabalha a perspectiva de amplitude interquartílica para a identificação de casos extremos;
- *histogram* é o gráfico de barras direcionado para variáveis contínuas, onde todas as observações são encaixadas em intervalos e sua frequência contatada

(eixo X é o conjunto de intervalos e o Y é a contagem de casos dentro daquela amplitude). Este gráfico é ideal para observar a normalidade das distribuições;

- *scatter* é conhecido em português como gráfico de dispersão e apresenta o relacionamento entre duas variáveis, buscando captar o componente linear de correlação entre essas variáveis. É uma forma gráfica de captar um coeficiente de correlação bivariada ou um coeficiente de regressão simples;
- *interactive* é um módulo do SPSS para trabalhar com todos os gráficos descritos anteriormente e muitos outros, mas com a possibilidade de geração e manipulação tridimensional.

Os gráficos dos SPSS podem ser formatados na janela de output do programa. Para isso basta clicar duas vezes sobre o gráfico, na janela de output, que uma terceira janela será aberta e vários procedimentos de formatação podem ser executados, como nomear o gráfico, inserir notas de rodapé, modificar a cor de partes do gráfico, transpor os eixos, entre muitas outras opções. Esses procedimentos são importantes para que os gráficos possam sair do SPSS com a aparência desejada para serem inseridos no corpo de relatórios ou textos.

O menu *utilities* possui vários comandos que auxiliam na manipulação dos dados e das variáveis de um arquivo. Existem comandos para a visualização das variáveis e seus respectivos critérios de definição, estipulados na janela do **SPSS Data Editor** (*input*) chamada *variable view*.

O menu *Window* possibilita a mudança de janelas do SPSS, como em qualquer programa de ambiente windows. O último menu é denominado *Help*. Nele estão inseridas muitas informações sobre o programa, desde como executar determinados procedimentos até a explicação da função de diferentes técnicas estatísticas. Este menu possui um tutorial de procura de termos, que pode ser muito útil para a proficiência na operação do SPSS.

CAPÍTULO 2

Descrevendo Dados no SPSS

Conteúdo:

- 1. Tabelas de frequência**
- 2. Cruzamento de dados**
- 3. Medidas de tendência central**
- 4. Medidas de posição**
- 5. Medidas de dispersão**

Objetivos:

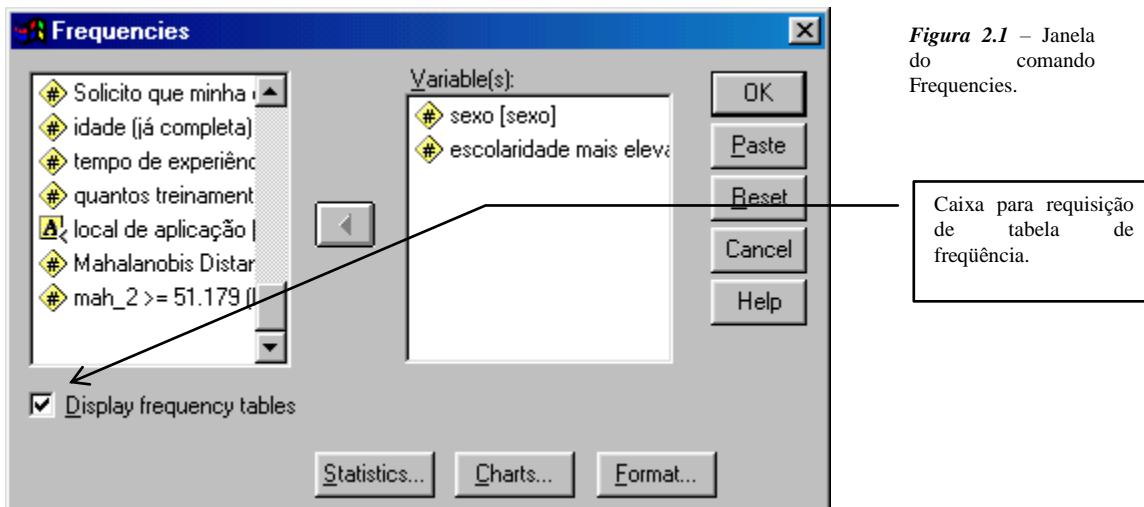
Ao final do capítulo o aluno deverá ser capaz de:

- 1. Gerar tabelas de frequência no SPSS.**
- 2. Gerar e interpretar tabelas de cruzamento de dados.**
- 3. Realizar procedimentos estatísticos de tendência central.**
- 4. Gerar medidas de posição.**
- 5. Calcular medidas de dispersão de dados.**

Uma das primeiras ações a serem feitas com dados estatísticos é sua descrição, por meio de estatísticas descritivas, como frequências, médias, medianas, variância, entre outras. Muitas vezes também é objetivo da análise de dados a realização de cruzamento de diferentes informações, para se ter uma noção da relação existente entre variáveis da pesquisa. Todos os procedimentos referentes a essa descrição de dados, e conseqüentemente todo o conteúdo desse segundo capítulo da apostila, estão concentrados no menu **Analyze** ⇒ **Descriptives Statistics** do SPSS.

Na presente apostila será iniciada a apresentação de técnicas de descrição de dados categóricos via SPSS. Para tanto inicialmente serão apresentadas a geração de tabelas de frequência e posteriormente a geração de tabelas de cruzamento de dados (*crosstabs*). Em seguida serão apresentadas as técnicas de descrição de dados contínuos no SPSS, primeiramente com as medidas de tendência central, em seguida com as de posição e por última as medidas de dispersão.

Para a obtenção de tabelas de frequência deve-se seguir os seguintes passos para a janela de diálogo apresentada na Figura 2.1: **Analyze** ⇒ **Descriptives Statistics** ⇒ **Frequencies**.



ESCOLAR escolaridade mais elevada

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 3 Segundo grau	144	50.7	50.7	50.7
4 Curso superior	90	31.7	31.7	82.4
5 Especialização	38	13.4	13.4	95.8
6 Mestrado	8	2.8	2.8	98.6
7 Doutorado	4	1.4	1.4	100.0
Total	284	100.0	100.0	

Tabela 2.1 – Resultado do procedimento *Frequencies*.

Colunas com frequência, percentagem, percentagem válida e percentagem acumulada.

Como pode ser observado na Figura 2.1 foi solicitada uma tabela de frequência para cada uma das variáveis do exemplo, *i.e.* sexo e escolaridade. A Tabela 2.1 apresenta a tabela de frequência para a variável escolaridade, com o resultado da análise solicitada na Figura 2.1.

Já a análise de cruzamento de dados (crosstabs) é obtida por meio do menu **Analyze** ⇒ **Descriptives Statistics** ⇒ **Crosstabs**. Esse comando resultará na janela apresentada na Figura 2.2.

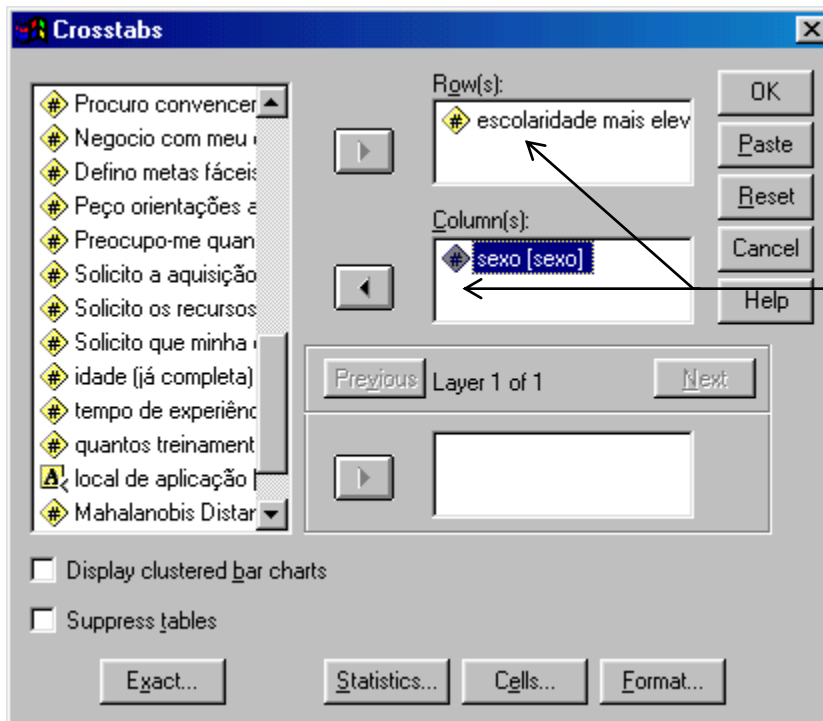


Figura 2.2 – Janela do comando Crosstabs.

Janelas de definição de variáveis que serão apresentadas nas linhas (*row*) e nas colunas (*column*).

ESCOLAR escolaridade mais elevada * SEXO sexo Crosstabulation

Count		SEXO sexo		Total
		1 Feminino	2 Masculino	
ESCOLAR	3 Segundo grau	69	75	144
escolaridade	4 Curso superior	45	45	90
mais elevada	5 Especialização	19	19	38
	6 Mestrado	4	4	8
	7 Doutorado	1	3	4
Total		138	146	284

Tabela 2.2 – Resultado do procedimento Crosstabs.

Valor que representa o número de pessoas do sexo feminino que possuem nível de escolaridade de segundo grau.

Como pode ser observado na Tabela 2.2, a tabela de cruzamento de dados do SPSS apresenta o cruzamento de informações entre variáveis, como sexo e escolaridade, como exemplificado. Dessa forma, é possível ter uma noção de covariação de variáveis categóricas, para uma exploração das características dos dados da pesquisa.

As medidas de tendência central são usualmente utilizadas em variáveis contínuas ou quantitativas. A Figura 2.3 apresenta a janela de diálogo para execução das principais medidas de tendência central no SPSS, média, mediana e moda. Essa janela é obtida por meio do caminho **Analyze** ⇒ **Descriptives Statistics** ⇒ **Frequencies**.

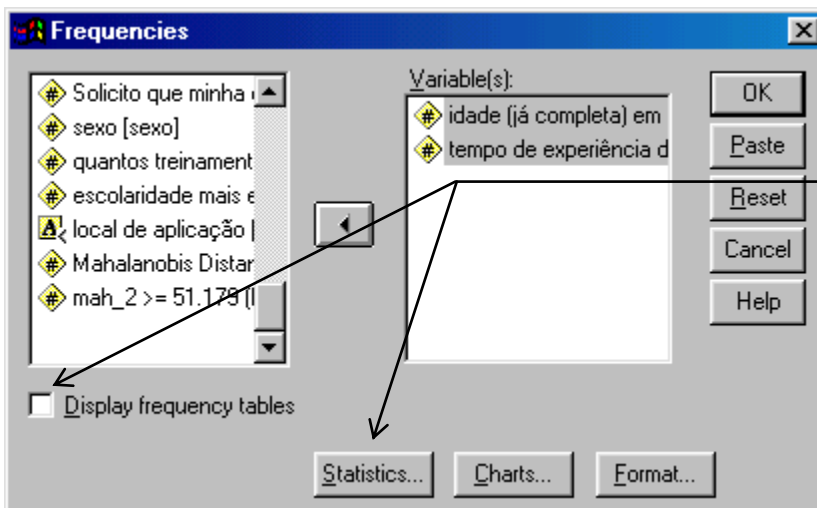
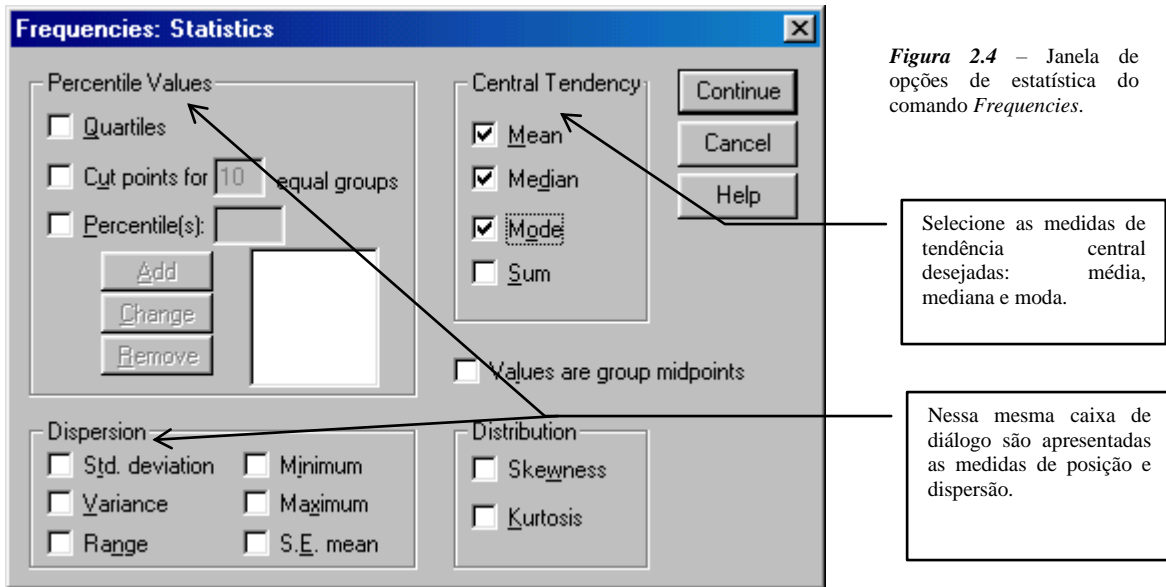


Figura 2.3 – Janela do comando Frequencies.

Desmarque essa opção. Clique no botão *statistics* para selecionar as análises.

Quando selecionado o botão *statistics* da janela apresentada na Figura 2.3 tem-se a nova janela de diálogo de estatísticas, apresentada na Figura 2.4.



Depois de geradas as análises os resultados são apresentados na Tabela 2.3.

Statistics

		IDADE idade (já completa) em anos	EXPTRAB tempo de experiência de trabalho (em anos já completos)
N	Valid	274	282
	Missing	10	2
Mean		31.50	12.156
Median		29.00	10.000
Mode		21 ^a	2.0

Tabela 2.3 – Resultado das análises de tendência central.

As colunas apresentam as variáveis e as linhas as análises de tendência central.

a. Multiple modes exist. The smallest value is shown

Como salientado anteriormente a geração de medidas de dispersão e posição são selecionadas na mesma janela de diálogo, apresentada na Figura 2.4. O resultado da geração de uma análise de dispersão e de posição é apresentada na Tabela 2.4.

Statistics

		IDADE idade (já completa) em anos	EXPTRAB tempo de experiência de trabalho (em anos já completos)
N	Valid	274	282
	Missing	10	2
Std. Deviation		10.329	9.3293
Variance		106.698	87.0361
Percentiles	25	23.00	4.000
	50	29.00	10.000
	75	38.25	20.000

Tabela 2.3 – Resultado das análises de tendência central.

Medidas de dispersão (seta superior) e de posição (seta inferior).

Outra estratégia importante para descrição de dados é o uso de gráficos. Da mesma forma que os indicadores estatísticos os gráficos são específicos para cada tipo de variável em análise. Gráficos de barra e pizza são usados em variáveis categóricas e histogramas são representações de variáveis contínuas. A Figura 2.5 apresenta o menu *chart* da janela de diálogo *frequencies*, obtida pelo caminho **Analyze ⇒ Descriptives Statistics ⇒ Frequencies** e apresentada anteriormente na Figura 2.3.

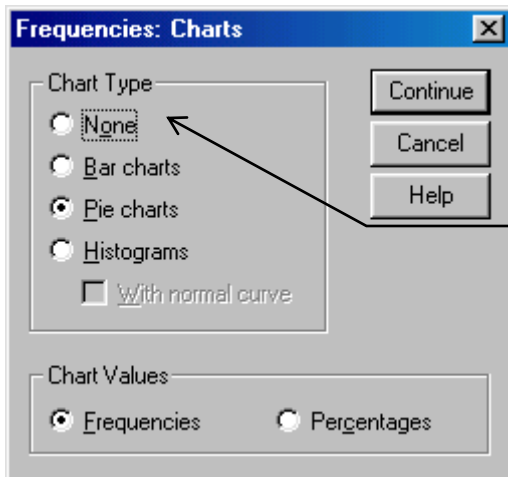


Figura 2.5 – Janela de gráficos do comando *Frequencies*.

Selecione o tipo de gráfico desejado de acordo com a variável a ser descrita.

O Gráfico 2.1 apresenta um gráfico de pizza, obtido por meio do comando descrito na Figura 2.5.

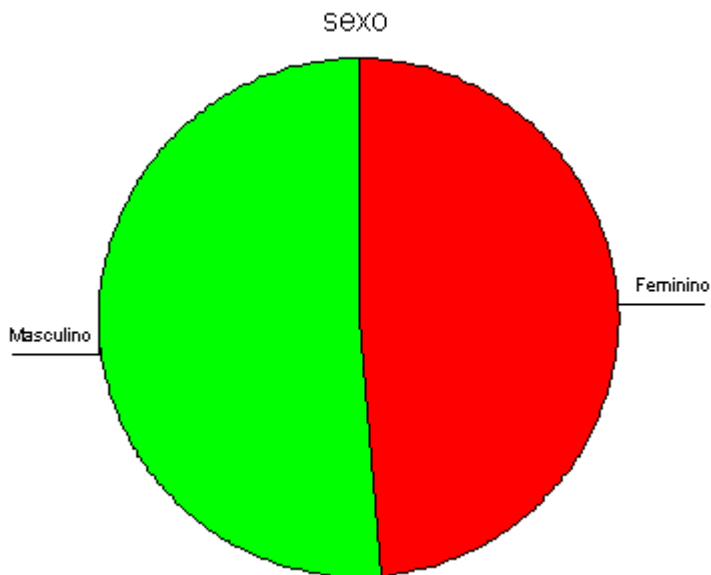


Gráfico 2.1 – Gráfico de pizza obtido por meio do menu *chart* do comando *Frequencies*.

Exemplos de histogramas, gráficos destinados à variáveis contínuas, são apresentados no capítulo 3 dessa apostila, destinado a obtenção de acuracidade de banco de dados após a digitação.

CAPÍTULO 3

Explorando e Adequando o Arquivo de Dados

Conteúdo:

- 1. Detecção e correção de erros de digitação**
- 2. Observação e tratamento de casos omissos**
- 3. Observação da normalidade de variáveis**
- 4. Estudo e eliminação de casos extremos univariados**
- 5. Estudo e eliminação de casos extremos multivariados**

Objetivos:

Ao final do capítulo o aluno deverá ser capaz de:

- 1. Detectar erros de digitação em arquivos de dados**
- 2. Produzir arquivos de dados acurados**
- 3. Tratar os casos omissos por meio dos procedimentos apresentados**
- 4. Identificar a normalidade/anormalidade das variáveis**
- 5. Identificar casos extremos univariados**
- 6. Tratar casos extremos univariados**
- 7. Tratar casos extremos multivaridos**

Após os dados terem sido digitados é fundamental que o primeiro passo do trabalho de análise de dados seja a limpeza e a promoção da acuidade do arquivo. Para tanto alguns cuidados e procedimentos devem ser seguidos. Pode-se apontar quatro procedimentos padrões, que serão apresentados no presente documento: (1) detecção e correção de erros de digitação, (2) observação e tratamento de casos omissos, (3) observação da normalidade das variáveis do arquivo de dados e (4) estudo e eliminação de casos extremos univariados.

DETECÇÃO E CORREÇÃO DE ERROS DE DIGITAÇÃO

É muito comum, durante o processo de inserção de dados no SPSS, ocorrerem erros de digitação, por exemplo, a duplicação de números em uma célula. Como os questionários em avaliação de treinamento geralmente são associados a escalas de resposta tipo Likert, a duplicação de valores em uma célula vai acarretar em inflacionamento das estatísticas daquela variável, prejudicando todo o procedimento de análise de dados. Para que isso não ocorra é fundamental, como primeiro passo para a análise de dados, verificar se não existem erros de digitação no arquivo de dados.

Esse procedimento pode ser realizado por meio do menu **Analyze** ⇒ **Descriptives Statistics** ⇒ **Explore**, que resultará na janela apresentada na Figura 3.1.

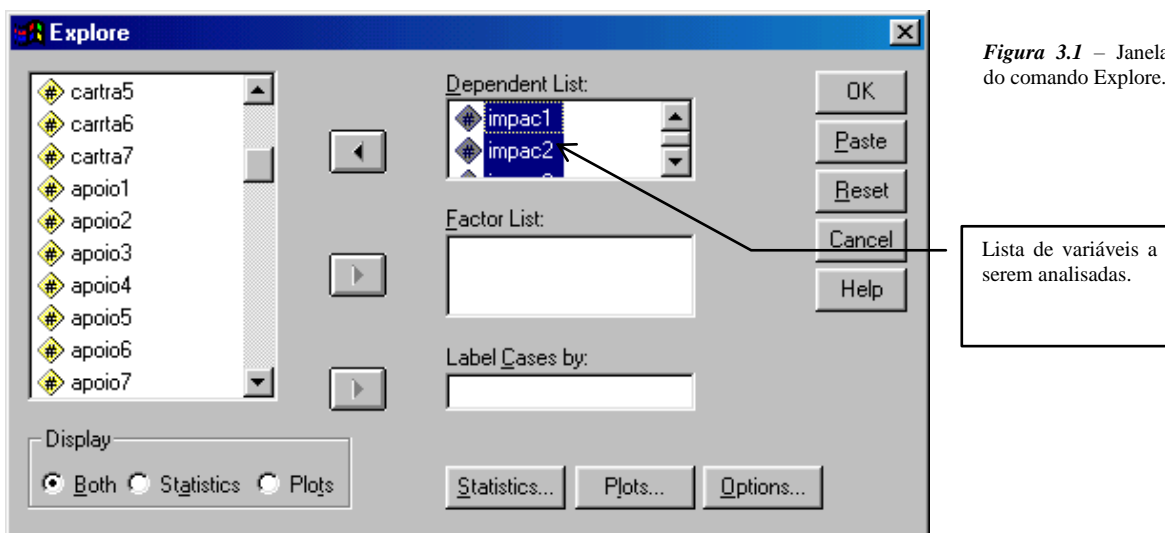


Figura 3.1 – Janela do comando Explore.

Após pressionado o botão *statistics* da janela da Figura 3.1 será aberta a janela apresentada na Figura 3.2.

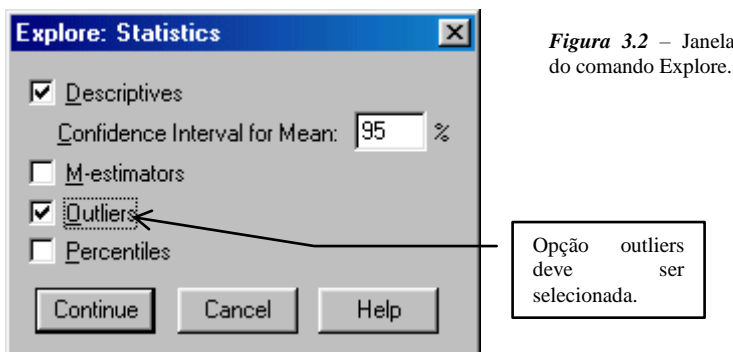


Figura 3.2 – Janela do comando Explore.

A seleção desta opção possibilitará a identificação de valores superiores a amplitude da escala de resposta dos itens do questionário. Assim, erros de digitação do tipo duplicação de valores em uma célula serão identificados. Além do valor ser apontado através deste procedimento, ele também possibilitará a identificação do caso onde se encontra esse valor errôneo. O resultado do procedimento obtido através da janela apresentada na Figura 3.1 é apresentado abaixo, na Tabela 3.1.

Extreme Values

			Case Number	Value
IMPAC1	Highest	1	2	5
		2	19	5
		3	217	5
		4	136	5
		5	40	a
	Lowest	1	279	1
		2	115	1
		3	182	1
		4	198	1
		5	190	1

Tabela 3.1 – Resultado do procedimento Explore.

Valores mínimos e máximos e respectivos

a. Only a partial list of cases with the value 5 are shown in the table of upper extremes.

É importante ressaltar que o procedimento explore apresenta várias informações na janela de output, mas apenas a tabela de *extreme values* possui as informações disponíveis na Tabela 3.1. Também é possível a identificação desses

erros de digitação por meio de outros módulos do SPSS, como **Analyze** ⇒ **Descriptives Statistics** ⇒ **Frequencies**. Após identificados os casos com erros de digitação o operador do arquivo de dados deve retornar a sua fonte bruta de informações (questionários respondidos pelos treinandos) e inserir a informações correta no local certo.

OBSERVAÇÃO E TRATAMENTO DE CASOS OMISSOS

O segundo passo a ser seguido é a identificação e tratamento de casos omissos (*missing data*). Um caso omissos é uma ausência de resposta, ou seja, por algum motivo o participante não atribuiu valor a uma pergunta do questionário. Essa ausência de resposta pode ser decorrente de diferentes fatores, como saturação de informações a serem processadas pelo respondente, inadequação da formatação do questionário, incapacidade do participante opinar sobre a asserção. A respeito dessa última possibilidade o código não se aplica (NA) de questionários buscam diminuir a ocorrência deste tipo de ausência de resposta.

É importante ressaltar que o excesso de casos omissos em determinado item de um questionário pode trazer várias indicações e pode, inclusive, ser considerado como escopo de um relatório, pois a incapacidade dos respondentes em avaliar questões pode indicar inadequação do instrumento, entre muitas outras possibilidades. Esse dado, por si mesmo, pode ser indicador um indicador relevante.

A identificação dos casos omissos pode ser obtida por meio do menu **Analyze** ⇒ **Descriptives Statistics** ⇒ **Frequencies**, que resultará na janela apresentada abaixo na Figura 3.3.

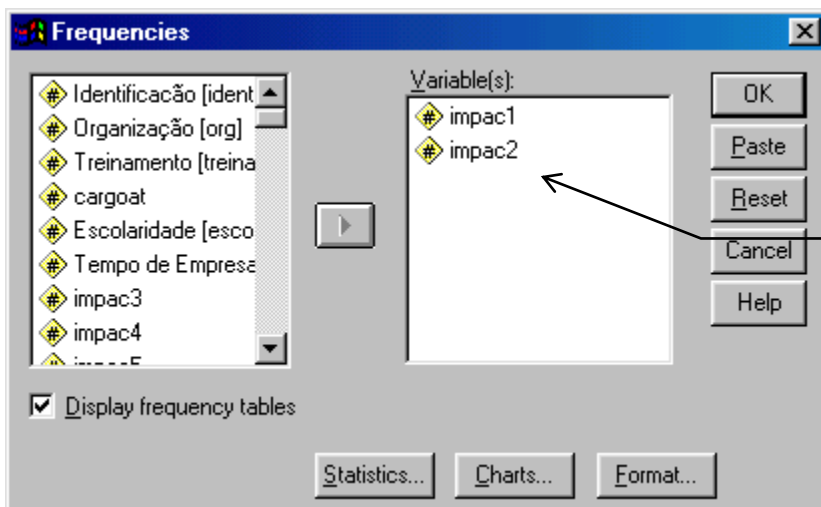


Figura 3.3 – Janela do comando Frequencies.

Variáveis selecionadas para análise.

Para a identificação dos omissos, basta que sejam deixadas as especificações padrões do SPSS. A Tabela 3.2 apresenta o resultado dos procedimentos executados através da Figura 3.3.

IMPAC 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Discordo Totalmente	5	1.8	1.8	1.8
	2 Discordo Pouco	26	9.3	9.5	11.3
	3 Não concordo, nem discordo	26	9.3	9.5	20.8
	4 Concordo	146	52.3	53.3	74.1
	5 Concordo Totalmente	71	25.4	25.9	100.0
	Total	274	98.2	100.0	
Missing	System	5	1.8		
	Total	279	100.0		

Tabela 3.2 – Resultado do procedimento Frequencies.

Porcentagem de casos omissos observados na variável.

A Tabela 3.2 apresenta os resultados obtidos por meio dos procedimentos executados no comando *frequencies*. É fundamental observar-se a percentagem de casos omissos observados em uma variável. Valores superiores a 3% podem ser considerados problemáticos e devem ser investigados com mais cuidado. Para variáveis com valores omissos inferiores ou iguais a 3%, estes podem ser substituídos pela média, por meio dos procedimento *Replace Missing Values*, do menu *Transform*, já apresentados no primeiro capítulo desta apostila.

A literatura sobre tratamento de dados omissos é ampla e bastante abrangente. De forma geral os pesquisadores e analistas de dados interessados no tema desenvolveram uma classificação para o tratamento destes casos. Um elemento central neste sistema de

classificação diz respeito à sistematicidade dos casos omissos. A presença de sistematicidade é crucial para o pesquisador, pois se isto ocorre a forma de tratamento destes casos deve ser diferenciada. A classificação pressupõe que os casos omissos de um banco de dados possam assumir as seguintes características: MCAR – omissos completamente aleatórios; MAR – omissos aleatórios; MNAR – omissos sistemáticos. O SPSS possui uma ferramenta útil para análise de casos omissos, o módulo *Missing Values Analysis* – MVA. Acesse o mesmo no menu **Analyze** ⇒ **Missing Values Analysis**. Será apresentada a janela da Figura 3.4.

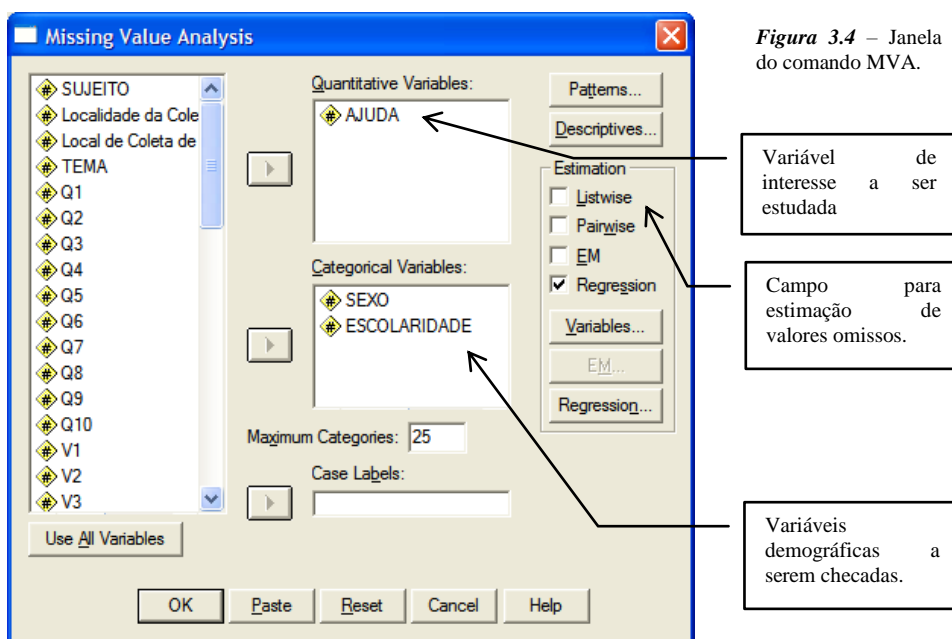


Figura 3.4 – Janela do comando MVA.

Em geral o estudo de sistematicidade de casos omissos está associada a observação de relação da variável de interesse com variáveis demográficas, para obtenção de indicações da ocorrência de sistematicidade entre a variável de interesse (no caso do exemplo a variável AJUDA) e estes indicadores. Para o estudo desta sistematicidade no botão ‘Patterns’ da caixa de diálogo você deve selecionar a opção ‘All cases optionally sorted by selected variables’. Esta opção trará na janela de saída de dados uma tabela apresentada todos os casos do banco e a ocorrência de caso omissos na variável de interesse. É uma ferramenta prática para a observação de algum tipo de

associação entre a variável de interesse e possíveis dados categóricos que possam estar associados a um padrão de omissão de resposta. Este módulo do programa também permite a estimação de valores para substituição de casos omissos. Como pode ser observado na Figura 3.4 é possível se selecionar estratégias de estimação de valores omissos, de acordo com a literatura da área. Cada opção tem um tipo de uso e o pesquisador deve estar consciente para a escolha da forma mais adequada para realizar a substituição de valores omissos. Esta estimação pode e deve ser utilizada em conjunto ao comando *Replace Missing Values*.

OBSERVAÇÃO DA NORMALIDADE DAS VARIÁVEIS

Um outro procedimento muito importante para testar a adequação dos dados para emissão de relatórios é o estudo da normalidade das variáveis. A ausência de normalidade na distribuição de uma variável pode indicar muitas coisas, como ausência de variabilidade (a grande maioria das pessoas concordam e/ou discordam de uma afirmativa de escala *Likert*). Essa informação, por si só, já é resultado de um relatório de avaliação de treinamento e pode ser utilizada como subsídio de um relatório. O que deve-se observar é a ausência de normalidade das variáveis que serão inseridas em análises inferenciais multivariadas, como regressão múltipla. Caso alguma variável anormal seja parte de uma análise multivariada, deve-se adotar procedimentos de transformação da mesma (*e.g.* logarítmica, raiz quadrada) através do comando *compute*, no menu *transform*.

Para a detecção da normalidade das variáveis deve-se executar o seguinte procedimento **Analyze ⇒ Descriptives Statistics ⇒ Frequencies**, que resultará na abertura da janela apresentada na Figura 3.5.

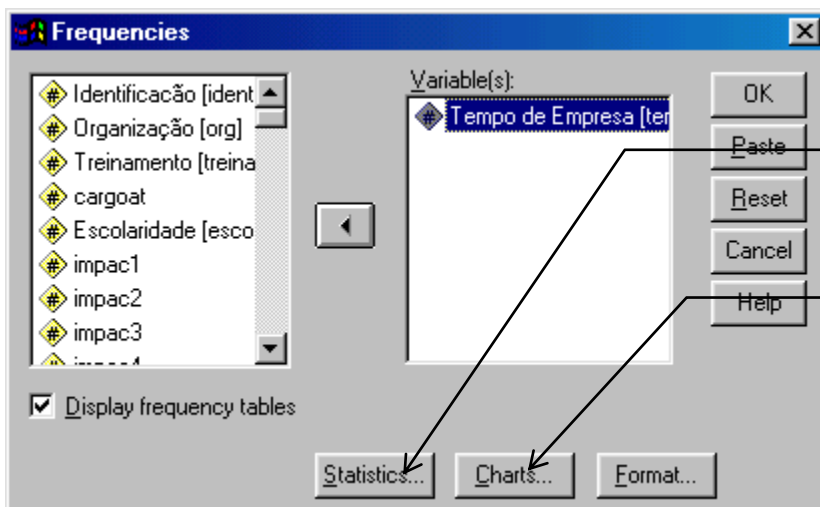


Figura 3.5 – Janela do comando Frequencies.

Selecionar análises neste botão.

Selecionar tipo de gráfico para estudo da normalidade.

Nesta janela de comando, a exemplo das outras janelas de comando estatístico do SPSS, o operador deve selecionar o conjunto de variáveis a serem analisados e movê-las para a coluna da direita. Para a observação da normalidade, o operador deve pressionar em dois botões da borda inferior da janela, o *statistics* e o *charts*. As janelas de diálogo destes botões estão apresentadas nas Figuras 3.6 e 3.7.

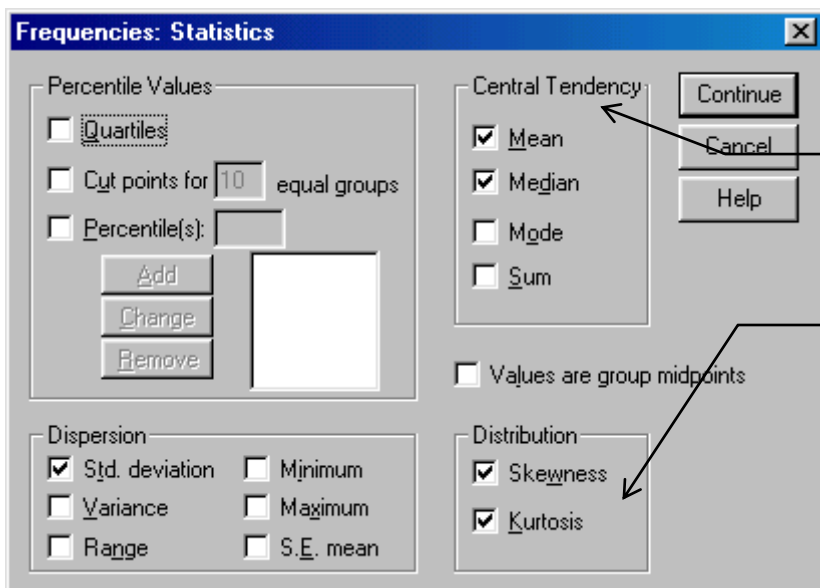


Figura 3.6 – Janela do comando Frequencies: Statistics.

Selecione medidas de tendência central.

É fundamental a seleção destas duas medidas de distribuição.

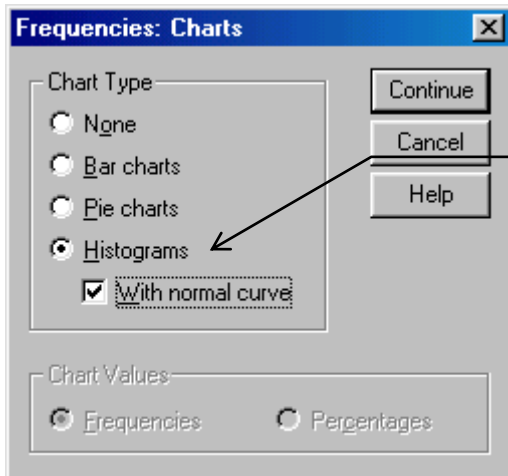


Figura 3.7 – Janela do comando *Frequencies: Charts*.

Selecione o histograma com curva densidade (normal).

Lembre-se que sempre após selecionar um procedimento dentro de uma caixa de diálogo deve-se clicar o botão continue, para que o programa execute os procedimentos solicitados. Os resultados dos procedimentos acima descritos são apresentados na Tabela 3.3 e no Gráfico 3.1.

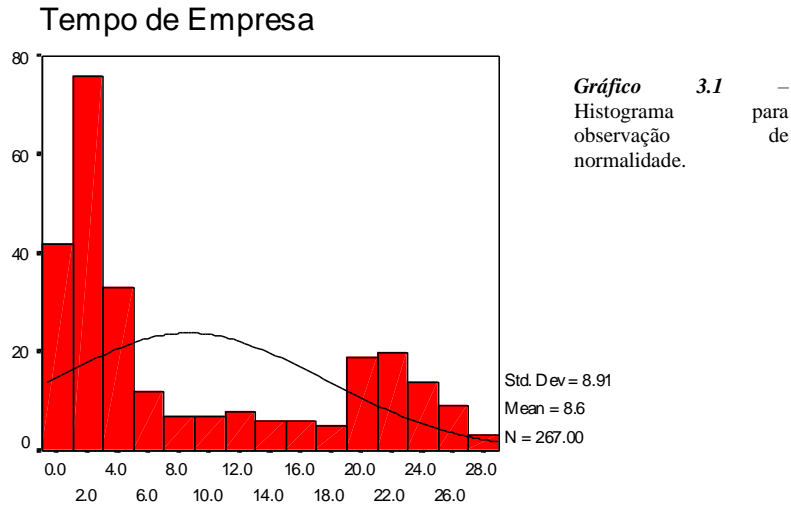
Statistics

TEMPO Tempo de Empresa

N	Valid	267
	Missing	12
Mean		8.6059
Median		4.0000
Std. Deviation		8.9123
Skewness		.782
Std. Error of Skewness		.149
Kurtosis		-1.036
Std. Error of Kurtosis		.297

Tabela 3.3 – Estatísticas descritivas do procedimento *frequencies*.

Valores de *skewness* e *kurtosis* e seus respectivos erros padrão.



Tempo de Empresa

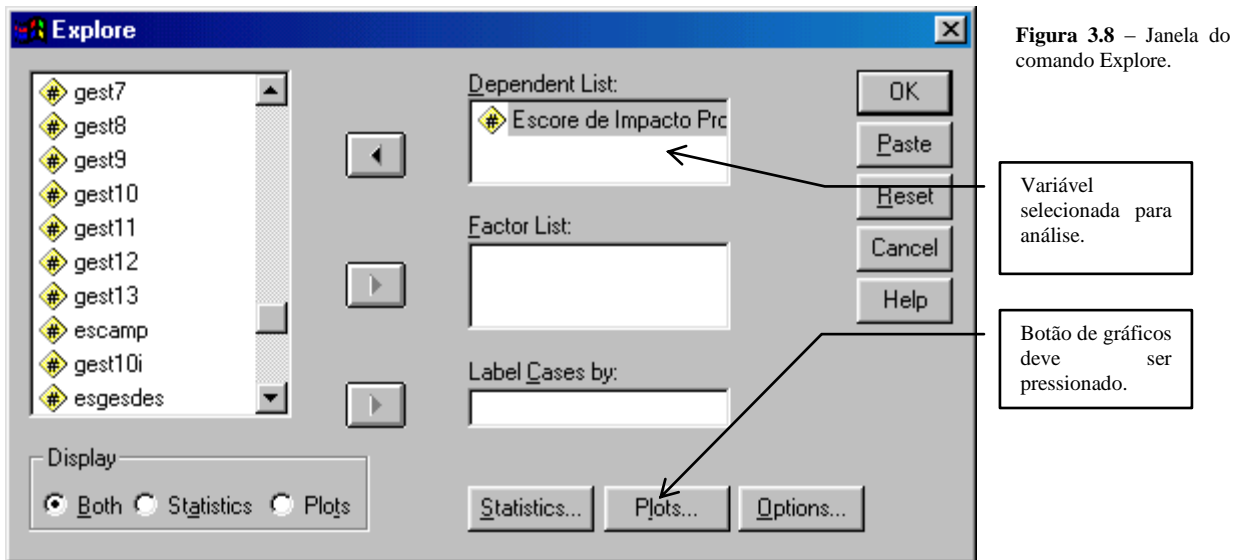
Para concluir se a distribuição de uma variável é normal deve-se observar para alguns indicadores apresentados na Tabela 2.3, como a *skewness* e a *kurtosis*. Quanto mais próximos de 0 estiverem esses valores, mais próximo da normalidade estará a distribuição da variável. Além de se observar o valor bruto dos indicadores de assimetria e achatamento deve-se dividir este indicador por seu erro padrão. Se o resultado da divisão for superior a 1,96 significa que aquele indicador é estatisticamente significativo. Também é importante observar na mesma tabela o valor do desvio padrão, que é superior a média, o que já indica ausência de normalidade na distribuição. Por fim pode-se comparar os resultados dos índices da Tabela 3.3 com o efeito gráfico do histograma, do Gráfico 3.1. Claramente a curva densidade (normal) não se aproxima da normalidade. Essa normalidade na curva poderia ser comparada a silhueta de um sino com a boca apontada para baixo.

ESTUDO E ELIMINAÇÃO DE CASOS EXTREMOS UNIVARIADOS E MULTIVARIADOS

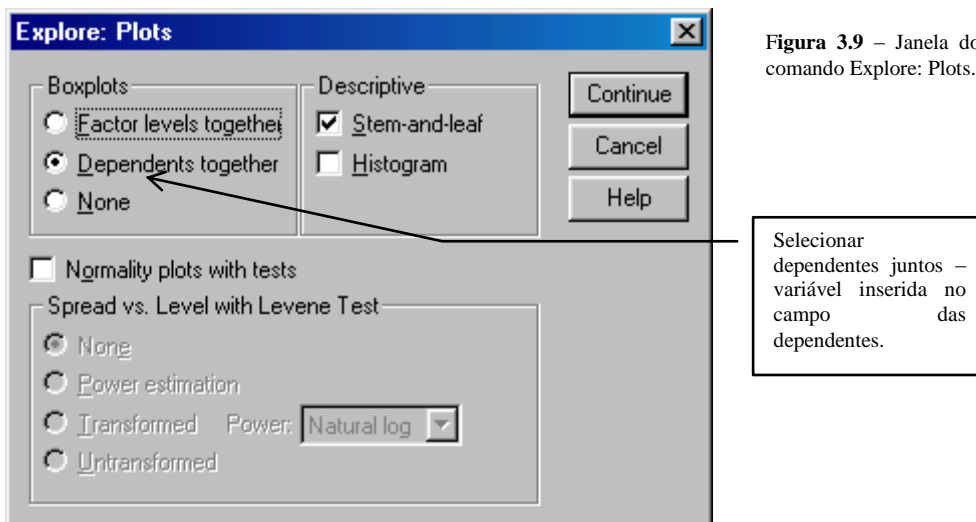
Casos extremos univariados são participantes que avaliaram de forma muito diferenciada do restante do grupo um determinado aspecto. Estes casos influenciam, por

exemplo, a normalidade de uma distribuição. Casos extremos podem ser provenientes de sujeitos que respondem sem atenção ao questionário, avaliando apenas um ponto da escala. Estes casos devem ser identificados, estudados e, se possível, eliminados do banco, para que os relatórios não sejam “mascarados” por avaliações severas ou lenientes em demasia.

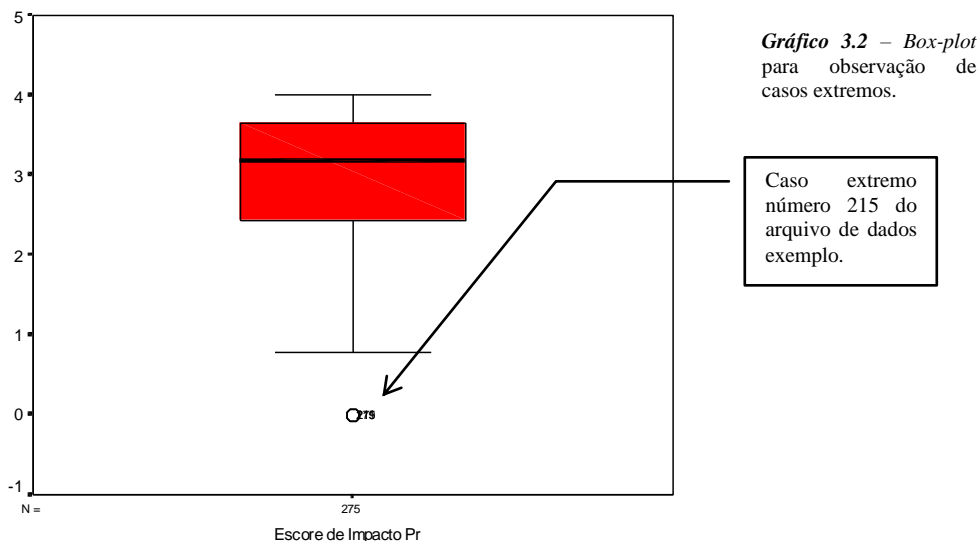
Para identificação de casos extremos univariados deve-se seguir **Analyze** ⇒ **Descriptives Statistics** ⇒ **Explore**, que resultará na janela apresentada na Figura 3.8.



Após pressionado o botão *plots* será apresentada a janela da Figura 3.9.



A melhor análise de casos extremos univariados pode ser obtida através do *box-plot* do SPSS, que apresenta os casos diretamente abaixo ou acima dos limites da distribuição. O Gráfico 3.2 é o resultado do procedimento descrito nas figuras anteriores.



O gráfico aponta qual(is) casos estão acima da amplitude interquartílica. O número disposto no caso indicado pelo gráfico refere-se ao número automático apresentado na planilha do programa. Caso o gráfico indique vários casos, inicie a deleção pelo número maior, para que a numeração dos mais baixos não seja alterada.

Há um outro tipo de caso extremo que também deve ser explorado pelo pesquisador, é o caso multivariado. Este caso é obtido a partir do estudo simultâneo de múltiplas variáveis de interesse. Existem alguns procedimentos no SPSS para a identificação dos casos extremos multivariados. A distância de *Mahalanobis* é a mais comumente utilizada. Você acessa este procedimento por meio do menu **Analyze** ⇒ **Regression** ⇒ **Linear**. Você obterá a janela de diálogo apresentada na Figura 3.10.

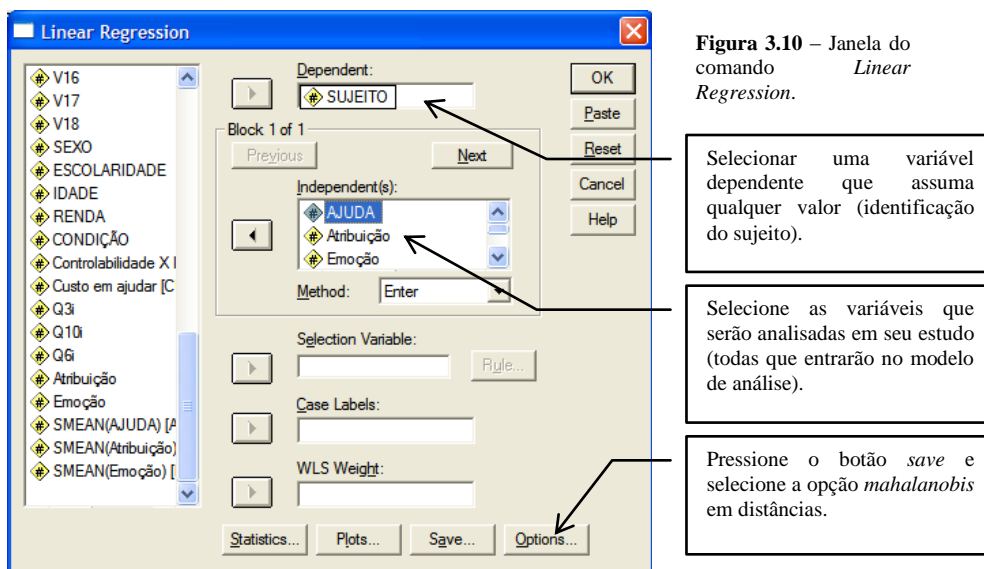


Figura 3.10 – Janela do comando *Linear Regression*.

Selecionar uma variável dependente que assuma qualquer valor (identificação do sujeito).

Selecione as variáveis que serão analisadas em seu estudo (todas que entrarão no modelo de análise).

Pressione o botão *save* e selecione a opção *mahalanobis* em distâncias.

O resultado desta análise gerará uma nova coluna de variável no seu banco de dados. Esta coluna conterá a informação da distância de *Mahalanobis*. Esta distância é uma distribuição de qui-quadrado e deve ser interpretada com auxílio de uma tabela de distribuição qui-quadrado. Valores superiores ao nível de significância determinado pelo pesquisador, considerando os graus de liberdade do modelo em teste (no exemplo $gl= 2$ pois são três variáveis em teste) determinam se o caso deverá ou não ser excluído do banco de dados. Para se automatizar o processo de eliminação de casos extremos multivariados recomenda-se o uso do comando *Select Cases* do menu *Data* do SPSS.

A seguir, no capítulo 4, serão apresentados os procedimentos para realização de análises inferenciais via SPSS, mais especificamente ANOVA uma via, para teste de diferença de médias, correlação para medidas paramétricas e não-paramétricas, teste de qui-quadrado para dados categóricos e análise de regressão linear simples.

CAPÍTULO 4

Realizando Análises Inferenciais: ANOVA uma via, Correlação Bivariada, Qui-quadrado e Regressão Linear Simples

Conteúdo:

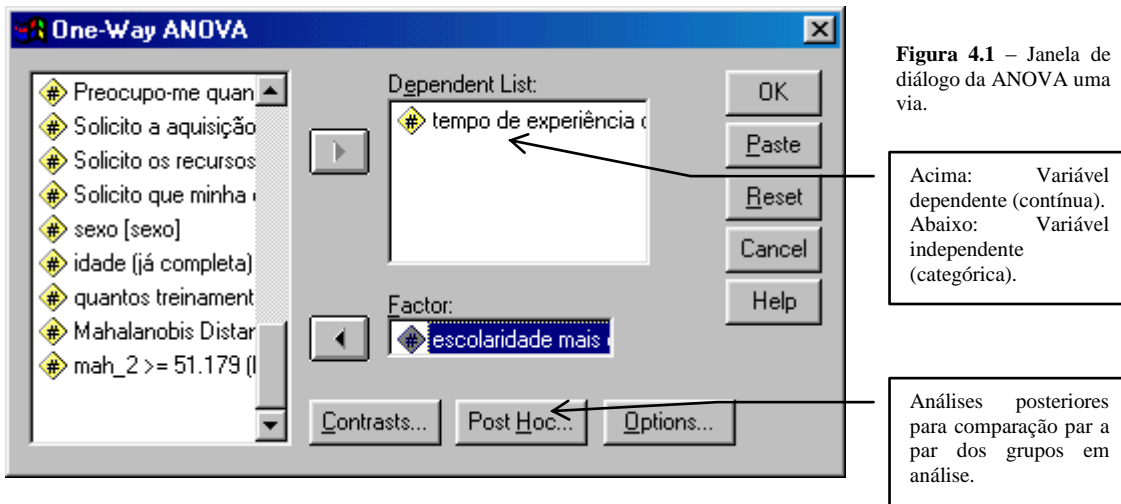
- 1. Comparação de Médias via ANOVA uma via**
- 2. Correlação bivariada para variáveis contínuas e categóricas**
- 3. Teste de Qui-quadrado**
- 4. Regressão linear simples**

Objetivos:

Ao final do capítulo o aluno deverá ser capaz de:

- 1. Realizar análises de comparação de médias de ANOVA uma via.**
- 2. Realizar análises de correlação bivariada.**
- 3. Realizar teste qui-quadrado.**
- 4. Realizar análise de regressão linear simples.**

Para realizar procedimentos de análises de teste da relação entre variáveis deve-se efetivar procedimentos específicos, chamados de inferenciais. Caso se deseje testar a diferença de média para uma variável dependente em diferentes grupos é necessária a realização de uma ANOVA uma via. A Figura 4.1 apresenta a caixa de diálogo da ANOVA uma via que é obtida por meio do caminho **Analyze** ⇒ **Compare Means** ⇒ **One-Way ANOVA**.



A análise solicitada na Figura 4.1 apresenta como resultado a Tabela 4.1, aonde são apresentados os índices de significância estatística do teste de F da ANOVA. Se o índice de significância (sig.) for menor que 0,05, há diferenças significativas entre os grupos.

ANOVA

EXPTRAB tempo de experiência de trabalho (em anos já completos)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6299.377	4	1574.844	24.025	.000
Within Groups	18157.758	277	65.551		
Total	24457.135	281			

Tabela 4.1 – Resultados da ANOVA uma via.

Para a realização de uma correlação bivariada deve-se seguir o caminho **Analyze** ⇒ **Correlate** ⇒ **Bivariate**. Aparecerá a janela de diálogo apresentada na Figura 4.2.

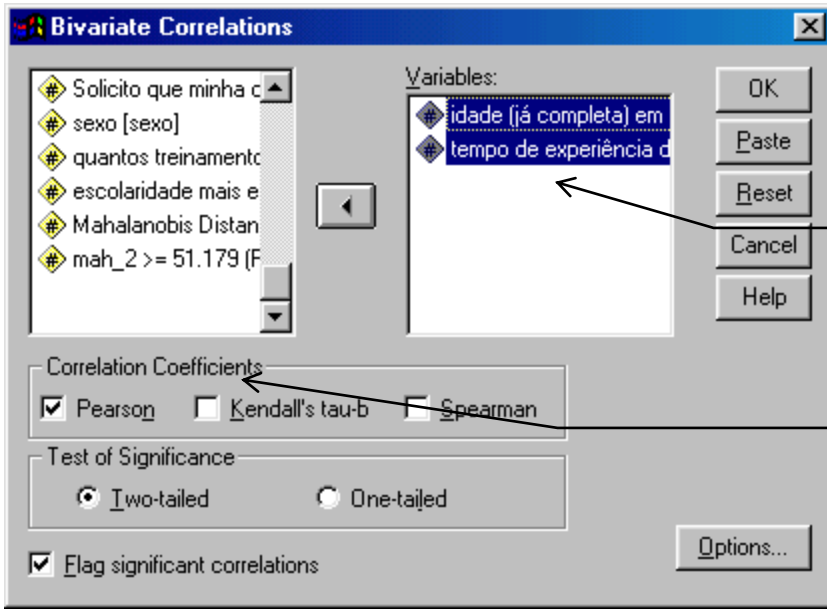


Figura 4.2 – Janela de diálogo de correlação bivariada.

Uma correlação bivariada exige ao menos duas variáveis para análise.

Faça a seleção dos coeficientes adequados para o tipo de variável: *Pearson* contínuas e demais para categóricas.

A Tabela 4.2 apresenta os resultados da análise de correlação. O coeficiente de correlação varia de -1 a $+1$ e quanto mais próximo de 0 menor é a associação entre duas variáveis.

Correlations

		IDADE idade (já completa) em anos	EXPTRAB tempo de experiência de trabalho (em anos já completos)
IDADE idade (já completa) em anos	Pearson Correlation	1	.909**
	Sig. (2-tailed)	.	.000
	N	274	272
EXPTRAB tempo de experiência de trabalho (em anos já completos)	Pearson Correlation	.909**	1
	Sig. (2-tailed)	.000	.
	N	272	282

Tabela 4.2 – Resultado da correlação bivariada.

Coeficiente de correlação. Asteriscos indicam o nível de significância da correlação.

** . Correlation is significant at the 0.01 level (2-tailed).

Para realizar comparação sobre a distribuição do número de casos em dois ou mais grupos distintos, pode-se utilizar o teste qui-quadrado. Este teste é obtido por meio do caminho **Analyze** ⇒ **Nonparametric tests** ⇒ **Qui-square**. A Figura 4.3 apresenta a caixa de diálogo do qui-quadrado.

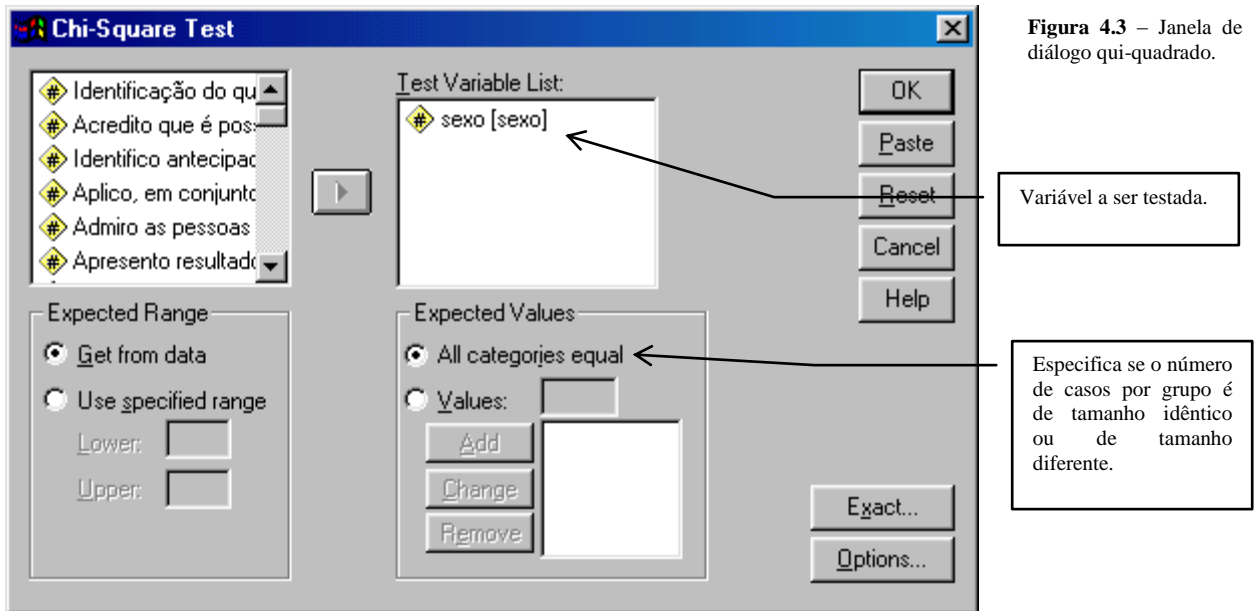


Figura 4.3 – Janela de diálogo qui-quadrado.

A Tabela 4.3 apresenta os resultados do qui-quadrado. A primeira tabela apresenta os valores observados no banco, os valores esperados e os resíduos. A segunda tabela apresenta o teste de significância. Se significativo, a distribuição dos casos é diferente da esperada.

ESCOLAR escolaridade mais elevada

	Observed N	Expected N	Residual
3 Segundo grau	144	56.8	87.2
4 Curso superior	90	56.8	33.2
5 Especialização	38	56.8	-18.8
6 Mestrado	8	56.8	-48.8
7 Doutorado	4	56.8	-52.8
Total	284		

Tabela 4.3 – Resultado do qui-quadrado.

Valores observados e esperados.

Test Statistics

	ESCOLAR escolaridade mais elevada
Chi-Square ^a	250.507
df	4
Asymp. Sig.	.000

Teste de significância.

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 56.8.

Para a realização de análises preditivas entre duas variáveis pode-se realizar uma análise de regressão simples. Esta análise é obtida por meio do caminho **Analyze** ⇒ **Regression** ⇒ **Linear**. A Figura 4.4 apresenta a caixa de diálogo da análise de regressão.

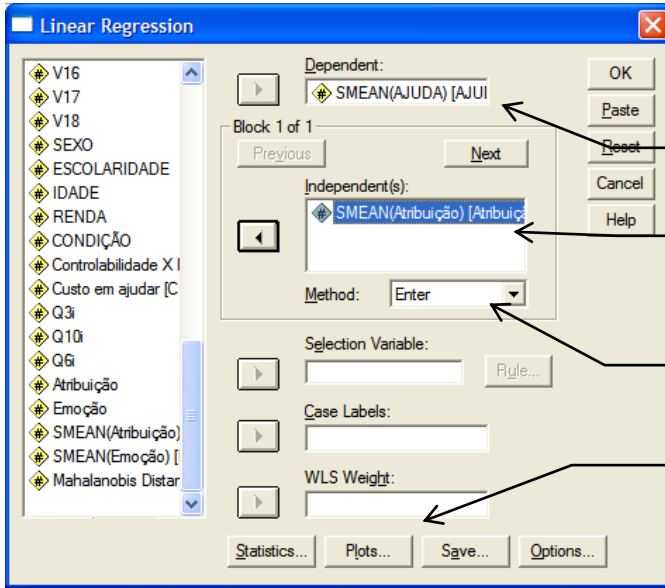


Figura 4.4 – Janela de diálogo da regressão linear.

- Defina a variável dependente de análise.
- Defina a variável independente de análise.
- Método de análise de regressão.
- Botões para seleção de procedimentos complementares de análise.

O resultado da análise é apresentada na Tabela 4.4.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,269 ^a	,072	,071	2,975

a. Predictors: (Constant), Atribuição_1 SMEAN(Atribuição)

Tabela 4.4 – Resultado da análise de regressão linear simples.

Sumário do modelo: R, R², e R² ajustado para o tamanho da amostra.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	376,768	1	376,768	42,569	,000 ^a
	Residual	4823,688	545	8,851		
	Total	5200,456	546			

a. Predictors: (Constant), Atribuição_1 SMEAN(Atribuição)
b. Dependent Variable: AJUDA_1 SMEAN(AJUDA)

Valores de F para regressão e nível de significância.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4,147	,316		13,141	,000
	Atribuição_1	,529	,081	,269	6,524	,000
	SMEAN(Atribuição)					

a. Dependent Variable: AJUDA_1 SMEAN(AJUDA)

Coefficiente beta da relação entre VI e VD e respectivos níveis de significância.

CAPÍTULO 5

Análise Fatorial em Avaliação do Treinamento

Conteúdo:

- 1. Análise Fatorial**
- 2. Procedimentos de Execução da Análise Fatorial via SPSS**

Objetivos:

Ao final do capítulo o aluno deverá ser capaz de:

- 1. Definir Análise Fatorial**
- 2. Identificar as funções da Análise Fatorial**
- 3. Descrever os principais conceitos associados à Análise Fatorial**
- 4. Realizar uma Análise Fatorial, seguindo todas as etapas, via SPSS**

ANÁLISE FATORIAL EM AVALIAÇÃO DE TREINAMENTO

A análise fatorial (AF) é uma técnica estatística multivariada desenvolvida no início do século XX por um psicólogo-estatístico chamado Spearman. Este pesquisador teve como objetivo estudar as capacidades humanas, por meio da identificação de construtos ou dimensões que são subjacentes à estas capacidades. Ele propôs que traços latentes determinam o desempenho de uma pessoa, por exemplo, a inteligência pode ser mensurada por meio de um conjunto de itens. A pressuposição é que existem habilidades correlacionadas que determinam uma dimensão subjacente, que permite que as pessoas desempenhem em situações que exijam o uso dessas habilidades. Diante desta pressuposição seria possível desenvolver um conjunto de itens para mensurar determinado traço latente. Se esses itens fizerem parte da mesma dimensão subjacente serão relacionados entre si e poderão ser considerados como indicadores de um determinado construto. Deve-se ressaltar que a construção de medidas deve respeitar alguns pressupostos da teoria psicométrica, como a elaboração de um número suficiente de itens (nem muito pouco, nem em exagero; cada habilidade ou atitude humana necessitam de diferentes números de itens) para mensuração de determinado traço latente.

No caso de instrumentos relativos a avaliação de treinamento a pressuposição é que os indicadores de modelos de avaliação constituem-se em habilidades e atitudes dos treinandos. Em um instrumento de reação, o traço subjacente é a atitude do treinando a respeito do evento instrucional, enquanto uma medida de aprendizagem procura mensurar as habilidades subjacentes desenvolvidas durante o evento instrucional.

A partir dessa pressuposição, a AF possibilita a redução de dados. Essa redução é feita com base no estudo das dimensões latentes que determinariam as atitudes e/ou habilidades das pessoas, que causariam a sua resposta a um conjunto de itens de determinada forma. A implicação prática deste tipo de procedimento está na redução de dados, tendo em vista que conjuntos muito extensos de itens podem ser reduzidos a poucos escores, graças a pressuposição dos modelos de avaliação de treinamento, de que as atitudes dos sujeitos são organizadas em dimensões subjacentes. Desta forma, pode-se

obter dois ou três índices para a mensuração de reação ao treinamento e, na maioria das vezes, um índice para o comportamento no cargo (impacto do treinamento no trabalho). É, ainda, importante ressaltar que os índices estatísticos dão subsídio para o técnico selecionar as dimensões, mas essa deve ser calcada, em última instância, no sentido teórico dos itens que se agregaram a um fator.

Esta técnica de análise de dados tem sido muito utilizada na avaliação de treinamento, tanto na pesquisa acadêmica quanto nos sistemas de avaliação das organizações. Pode-se considerar que cada um dos níveis de análise de um sistema de avaliação possui, no mínimo, uma dimensão latente, compartilhada pelo conjunto de treinandos envolvidos no processo instrucional. Para o nível de aprendizagem, supõe-se a observação de um conjunto de habilidades humanas para cada treinamento desenvolvido, mas os testes dimensionais para o nível de aprendizagem tem sido mais complicados, pela dificuldade na construção de múltiplas medidas (uma para cada treinamento).

Kline (1994) aponta que existem alguns conceitos que são fundamentais para o entendimento e interpretação de resultados da AF. São esses:

- *Correlação*: conceito estatístico que busca captar o componente linear de variação conjunta de duas variáveis. Esse coeficiente varia de -1 até $+1$. Uma correlação perfeita possui o valor 1 (seja esta negativa ou positiva). Quanto maior a correlação entre duas variáveis mais estas possuem componentes de covariação em conjunto, ou seja, quanto mais uma varia (modifica de valor) mais a outra varia. A AF é baseada nas correlações entre os itens de um questionário de avaliação de treinamento. O primeiro passo para o cálculo de fatores em uma AF, realizado pelo computador, é o cálculo de uma matriz de correlação. Todos os itens de um instrumento de avaliação são correlacionados entre si, formando o ponto de partida para a AF.
- *Fatores*: busca-se obter, em uma análise fatorial, dimensões. Essas dimensões são denominadas fatores. Estas são determinadas pela correlação compartilhada por um conjunto de itens, ou seja, quanto mais correlacionados for um conjunto de itens eles têm uma grande chance de formar um fator. Essa

correlação é proveniente desta estrutura subjacente das pessoas, que determina a forma como elas respondem a esse conjunto de questões. Cada fator extraído de uma AF significa a identificação de uma dimensão latente.

- *Cargas Fatoriais*: após observada a correlação de todos os itens entre si e após a detecção de que alguns grupos desses itens podem ser organizados em fatores é necessário traçar qual a correlação de cada item com seu respectivo fator. Essa relação item-fator é dada por meio da carga fatorial. Esse é um coeficiente que varia de -1 a $+1$, como a correlação. Quanto mais próximo de 1 maior a relação entre o item e o fator.

É muito importante ter em mente a função prática da AF, que se traduz no conceito de validade. Em um modelo de avaliação, por exemplo no nível de reação, se pressupõe várias categorias diferentes, como habilidades do instrutor, qualidade do material didático, qualidade das instalações do ambiente de treinamento, entre outras. Mas apenas pode-se dizer que uma medida é válida quando submete-se o conjunto de instrumentos de avaliação em uma grande quantidade de treinandos e, posteriormente, realiza-se a redução de dados por meio da AF. Essa técnica fornece os parâmetros para os técnicos em avaliação de treinamento afirmarem que suas medidas conseguem mensurar o que eles propunham mensurar quando delinearão o modelo, ou seja, a validade da medida.

Após identificado o conjunto de estruturas subjacentes de um questionário de avaliação de treinamento, faz-se necessário extrair o índice de confiabilidade desses fatores. Esse índice de confiabilidade diz respeito a precisão da medida. Enquanto a AF apresenta a validade da medida, ou seja, o quanto aquele instrumento de avaliação é válido, por exemplo, para mensurar a reação dos treinandos perante o treinamento, o índice de confiabilidade informa qual a qualidade dessa medida.

Cuidado!

Como toda técnica estatística deve-se tomar alguns cuidados com a AF. Atualmente, com a sofisticação dos microcomputadores e também dos softwares estatísticos é muito fácil realizar procedimentos multivariados como o de AF. O técnico

sempre deve-se lembrar da máxima *garbage in, garbage out* (lixo entra, lixo sai). O técnico deve possuir uma pressuposição teórica, na qual hipotetize a existência de dimensões subjacentes. No caso de avaliação de treinamento, os modelos possuem tal pressuposição, pois são organizados em dimensões, constituídas por diferentes indicadores. Por este motivo é que não é coerente efetuar uma análise fatorial em um conjunto de itens onde a pressuposição anterior de categorias não exista, pois poderão ser extraídas dimensões, formadas apenas por correlações espúrias entre itens. Devido a possibilidade de correlações distorcidas é que o sentido teórico dos itens deve ser o principal critério para identificação de dimensões latentes.

PROCEDIMENTOS PARA EXECUÇÃO DA ANÁLISE FATORIAL

Para a realização de uma análise fatorial exploratória deve-se observar uma seqüência lógica de procedimentos para a detecção dos fatores latentes em um conjunto de itens. Inicialmente, o técnico deve realizar uma análise dos componentes principais e, posteriormente, uma análise fatorial propriamente dita. A análise dos componentes principais é uma técnica mais flexível, que tem como função dar indicações do número de fatores que devem ser extraídos no segundo passo do procedimento. A figura 3.1 apresenta a janela de diálogo da AF, obtida através de **Analyze ⇒ Data Reduction ⇒ Factor**.

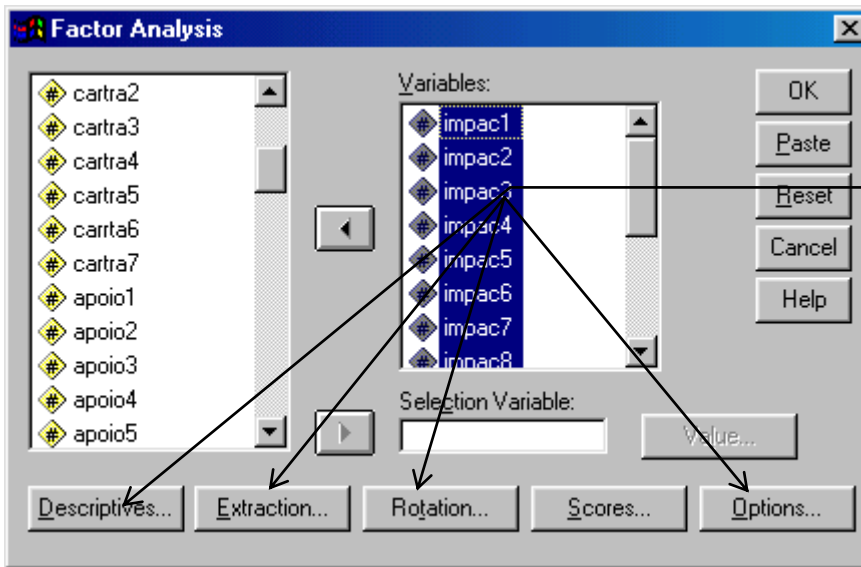


Figura 5.1 – Janela do comando Factor

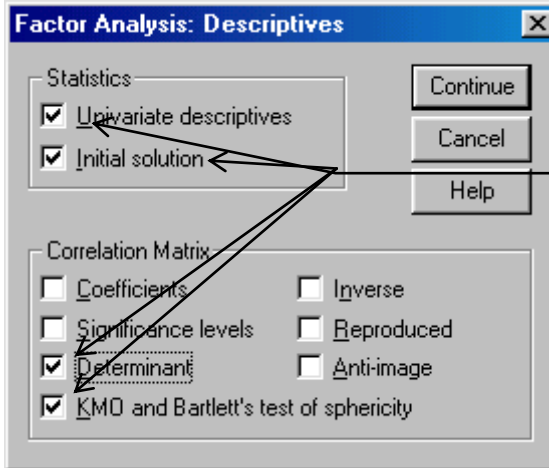


Figura 5.2 – Janela do comando Factor: Descriptives.

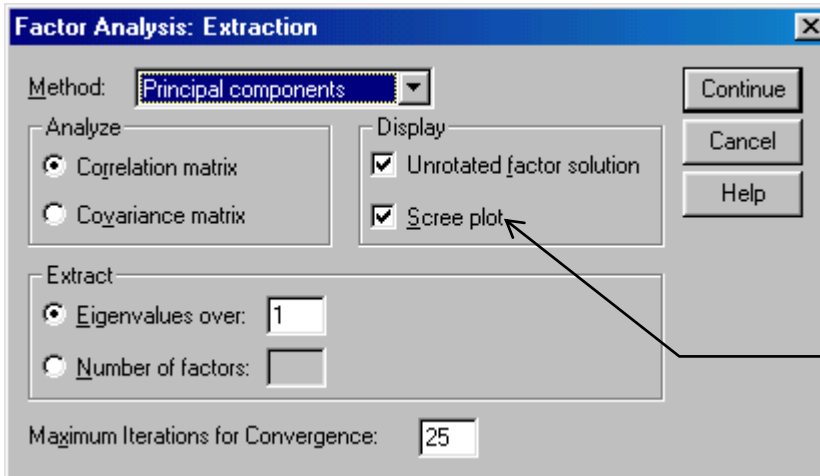


Figura 5.3 – Janela do comando Factor: Extrain.

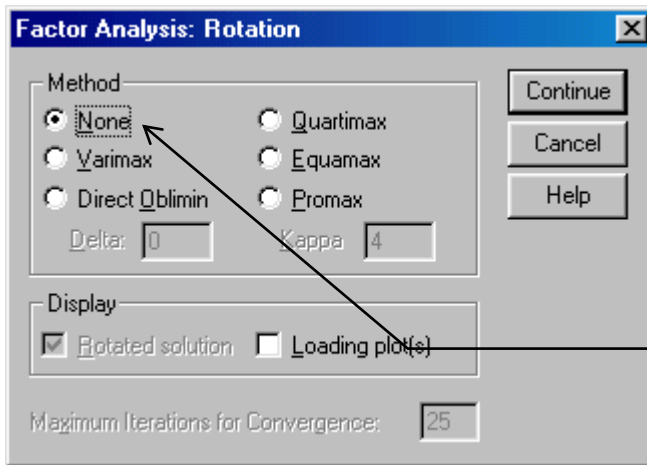


Figura 5.4 – Janela do comando *Factor: Rotation*.

Mantenha o default, que é sem indicação de rotação.

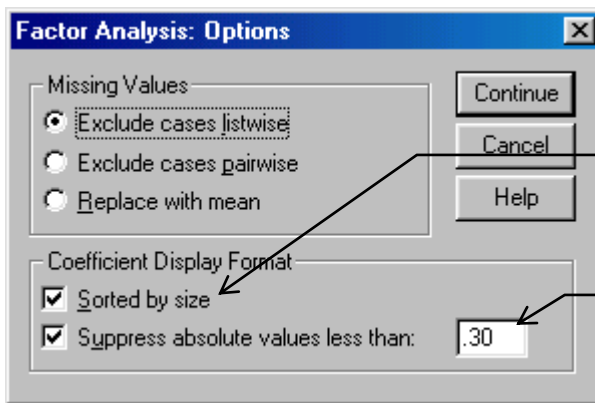


Figura 5.5 – Janela do comando *Factor: Options*

Selecione esta opção, que ordenará todos os coeficientes.

Selecione esta opção e modifique o valor para 0,30.

Todos esses procedimentos resultarão em tabelas e gráficos no output. Os principais, importantes para a tomada de decisão, serão apresentados a seguir. A tabela 3.1 apresenta o teste de KMO. Este é um indicador de fatorabilidade da matriz de dados. Quanto mais próximo de 1 melhor. Acima de 0,80 já existe indicativo de que os dados são submetidos a AF. No caso deste exemplo existe indicação de adequação dos dados para serem submetidos a AF.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy .		.889
Bartlett's Test of Sphericity	Approx. Chi-Square	971.990
	df	66
	Sig.	.000

Tabela 5.1 – Índice de KMO.

O próximo indicador, apresentado na tabela 3.2, aponta a quantidade de fatores apontado pela análise de componentes principais. Essa informação é obtida por meio dos *eigenvalues*. Valores superiores a 3,0 apontam para a existência de dimensões. Em seguida, no gráfico 3.1, são apresentados os *eigenvalues*. Quando a curva torna-se uma queda ordenada decrescente é que existe indicação da estrutura fatorial.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.756	39.634	39.634	4.756	39.634	39.634
2	1.251	10.424	50.058	1.251	10.424	50.058
3	.862	7.187	57.246			
4	.821	6.838	64.084			
5	.756	6.302	70.386			
6	.725	6.041	76.428			
7	.599	4.995	81.423			
8	.568	4.732	86.155			
9	.492	4.100	90.255			
10	.415	3.455	93.710			
11	.380	3.166	96.875			
12	.375	3.125	100.000			

Tabela 5.2 – Eigenvalues da solução.

Somente um valor é superior a 3, indicando apenas um fator.

Extraction Method: Principal Component Analysis.

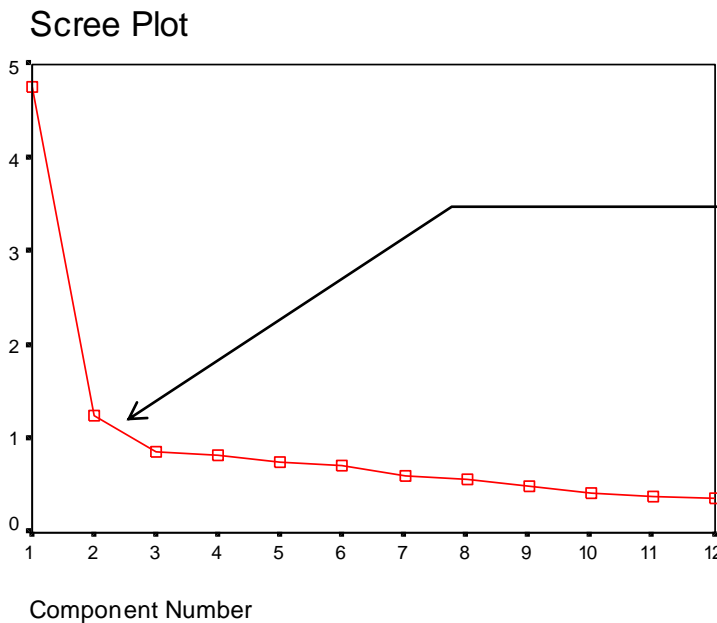


Gráfico 5.1 – Scree-plot – gráfico de eigenvalues.

Queda constante indica a presença de apenas um fator.

Como resultado da análise de principais componentes deste banco de dados observa-se que o conjunto de itens se agrega, adequadamente, em apenas um fator. A conclusão final pode ser extraída da observação do *scree plot*, que claramente indica a

presença de um fator, com *eigenvalue* muito elevado, indicando uma alta explicação da variância de todos os itens. Depois de concluído a estrutura fatorial a partir da análise de componentes principais, deve-se proceder a análise fatorial.

Para tanto, serão necessárias alterações em algumas das janelas (apresentadas nas figuras 3.1 a 3.5), como é apresentado a seguir:

- Janela *Factor* (figura 3.1): mantenha os mesmos itens selecionados para a análise de componentes principais;
- Janela *Factor: Descriptives* (figura 3.2): retire a marcação de todas as opções. Estas informações não são mais necessárias;
- Janela *Factor: Extraction* (figura 3.3): retire as marcações do *scree plot* e *unrotated factor solution*. Marque a opção número de fatores, informando o número observado na análise de componentes principais (um neste exemplo). Vá até o campo *Method* e selecione *principal axis factoring* (PAF).
- Janela *Factor: Rotation* (figura 3.4): mantenha o *default*, sem rotação. Isso deve ser feito para este exemplo, com indicação de apenas um fator. A ausência de rotação aplica-se apenas para soluções fatoriais de estrutura unifatorial.
- Janela *Factor: Options* (figura 3.5): mantenha as mesmas informações solicitadas na análise de componentes principais.

O arquivo de *output* desses procedimentos conterá um número consideravelmente menor de informações do que aquele gerado pela análise de componentes principais. É interessante observar algumas informações nestes resultados, como a tabela de porcentagem da variância explicada, a mesma apresentada na tabela 3.2. Mas a principal informação que deve ser observada é a matriz de cargas fatoriais da solução, apresentada na tabela 3.3. É bom ressaltar que esta matriz recebe este nome (*factor matrix*) quando a solução é unifatorial. Quando a solução for com mais de um fator a matriz que deve ser observada é denominada *pattern matrix*.

Factor Matrix^a

	Factor
	1
IMPAC6	.707
IMPAC1	.706
IMPAC9	.661
IMPAC3	.628
IMPAC5	.612
IMPAC2	.589
IMPAC8	.583
IMPAC12	.574
IMPAC4	.528
IMPAC7	.479
IMPAC10	.461
IMPAC11	.431

Tabela 5.3 – Cargas fatoriais dos itens da solução.

Extraction Method: Principal Axis Factoring.

a. 1 factors extracted. 4 iterations required.

Na tabela 3.3 podem ser observadas duas colunas. A primeira possui o nome de cada um dos itens do fator. Neste exemplo todos os itens do fator obtiveram valor de carga fatorial superior a 0,30 e todos os itens analisados (12 ao total) fizeram parte da solução final. A segunda coluna (direita) possui a carga fatorial de cada item, indicando o grau de correlação de cada um dos itens com o fator ao qual pertence.

Após finalizada a análise fatorial, o último procedimento diz respeito ao cálculo do índice de confiabilidade (precisão) do fator. Para tanto é necessário seguir o seguinte procedimento **Analyze ⇒ Scale ⇒ Reliability Analysis**, que resultará na abertura da janela apresentada na figura 3.6.

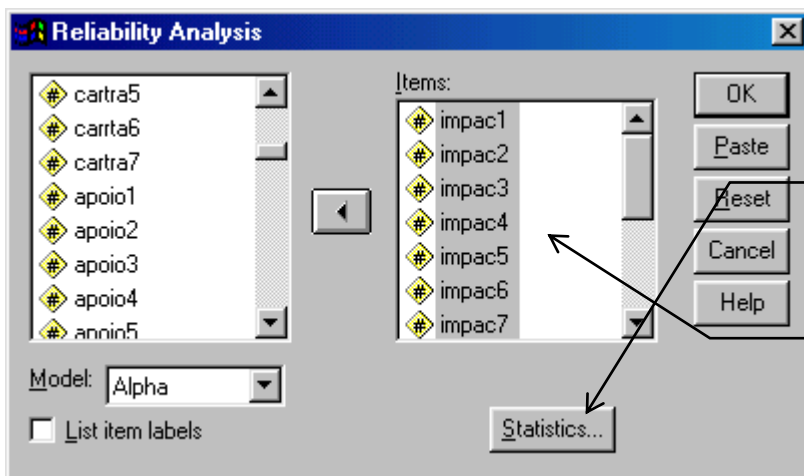


Figura 5.6 – Janela do comando Reliability.

Pressione este botão e escolha *scale if item deleted*.

Selecione todos os itens que compõem o fator a ser examinado

O técnico deve pressionar o botão da janela da figura 3.6 e selecionar a opção, como indicada no texto explicativo ao lado da figura. Esse procedimento irá inserir na tabela o valor do *alpha* (coeficiente de confiabilidade) caso o item seja excluído da solução fatorial. Esse é mais um indicador para apontar a adequação do item na mensuração da estrutura latente. A tabela 3.4 apresenta as informações relativas ao procedimento de teste de precisão da escala.

S C A L E (A L P H A)

Tabela 5.4 – Resultados dos cálculos de confiabilidade.

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Alpha if Item Deleted
IMPAC1	41.3916	40.3384	.6428	.8383
IMPAC2	41.2548	42.6333	.5347	.8462
IMPAC3	41.3042	40.8537	.5732	.8430
IMPAC4	41.6350	42.4693	.4791	.8494
IMPAC5	41.5247	40.7542	.5624	.8438
IMPAC6	41.4068	40.8911	.6353	.8393
IMPAC7	41.7947	42.9424	.4398	.8519
IMPAC8	41.3764	41.1669	.5451	.8450
IMPAC9	41.4030	40.6156	.6196	.8399
IMPAC10	42.0570	41.6341	.4366	.8537
IMPAC11	41.6540	42.9905	.4133	.8538
IMPAC12	41.7529	41.1715	.5237	.8466

Coluna com valores de *alpha* caso o item fosse eliminado.

Índice geral da escala, com todos os itens selecionados.

Reliability Coefficients

N of Cases = 263.0

N of Items = 12

Alpha = .8570

O valor obtido nesta escala do exemplo pode ser considerado bom, pois o valor é superior a 0,80. Este coeficiente varia de 0 a 1 e os melhores índices são aqueles superiores a 0,90.

Todos os procedimentos apresentados neste capítulo serão suficientes para a extração de indicadores de validade e confiabilidade de questionários de avaliação de treinamento via SPSS. Caso o leitor esteja interessado em aprofundamentos na técnica de AF sugere-se recorrer a leitura recomendada ao final desta apostila.

Após definidos os itens que compõem cada fator do instrumento de avaliação de treinamento, é necessária a criação de escores fatoriais por caso no arquivo de dados, gerando, finalmente, a redução de um conjunto de itens a um único indicador. Esse procedimento é feito através do menu **Transform** ⇒ **Compute**, resultando na janela apresentada na figura 3.7.

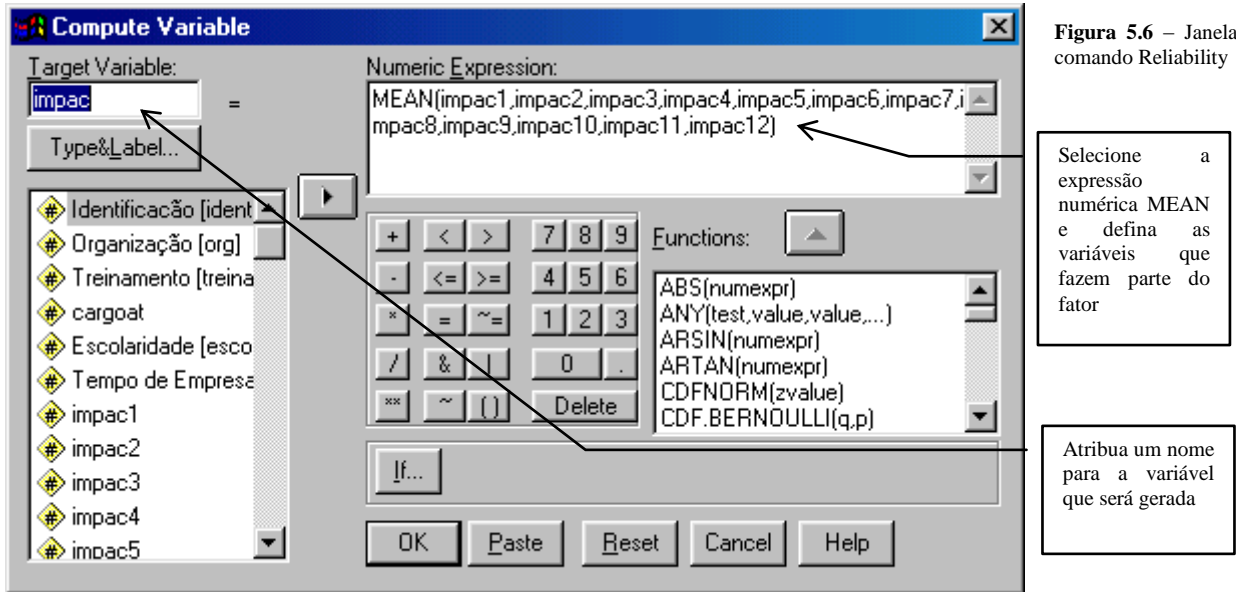


Figura 5.6 – Janela do comando Reliability

O campo *Functions* possui uma série de expressões. Para a criação do escore fatorial a expressão MEAN deve ser selecionada e “transportada” para o campo superior. Em seguida todas as variáveis que compõem o fator devem ser inseridas entre os parênteses que seguem a expressão numérica. Após esta seleção o botão OK deve ser pressionado. Isto irá gerar uma coluna ao final do arquivo de dados, atribuindo um escore para cada treinando que participou da avaliação do treinamento.

Esse mesmo procedimento pode gerar escores para qualquer estrutura fatorial observada. Após a execução desse procedimento todas as análises que irão subsidiar o relatório poderão ser feitas com base nesses escores, deixando de lado os itens que os constituíram. Os itens podem ser usados novamente, a depender do interesse dos técnicos em treinamento.

CAPÍTULO 6

Geração e Manipulação de *Syntax*

Conteúdo:

1. **Função e Uso da *Syntax* no SPSS**

Objetivos:

Ao final do capítulo o aluno deverá ser capaz de:

1. **Identificar quais os possíveis usos dos comandos da *syntax***
2. **Descrever alguns comandos da *syntax***
3. **Criar *syntax* de comandos para geração de rotinas de relatórios em avaliação de treinamento**

FUNÇÃO E USO DA SYNTAX NO SPSS

Como apresentado no primeiro capítulo desta apostila, o SPSS possui um tipo de arquivo que possibilita a geração e arquivamento de informações relativas a procedimentos de análise de dados. Este tipo de arquivo é muito útil quando o técnico necessita (a) realizar rotinas de análise de dados, como relatórios periódicos, onde os procedimentos não são diferenciados; (b) realizar análises complexas, onde um mesmo procedimento estatístico deve ser realizado, mas com diferentes variáveis de um mesmo arquivo de dados ou de arquivos de dados diferentes; e (c) trabalhar com múltiplos procedimentos de análise em um mesmo arquivo, como análise exploratória complexa. Na manutenção de um sistema de avaliação de treinamento é interessante a utilização de arquivos de *syntax* para a realização de rotinas periódicas de trabalho, como é o caso da elaboração de relatórios.

É importante chamar a atenção que a presente apostila tem como função informar sobre alguns tipos de comandos da *syntax* do SPSS. Existem uma quantidade muito grande de comandos via *syntax* e existem manuais específicos para a manipulação desses procedimentos. O menu *help* do SPSS traz muita informação sobre a *syntax*, podendo auxiliar na manipulação de dados por meio deste comando.

O arquivo de *syntax* pode ser obtido através do uso do botão *Paste*, presente na grande maioria das janelas de diálogo do SPSS. A figura 4.1 apresenta o procedimento para geração do comando GET em um arquivo de *syntax*. A janela é obtida por meio do **File ⇒ Open ⇒ Data**.

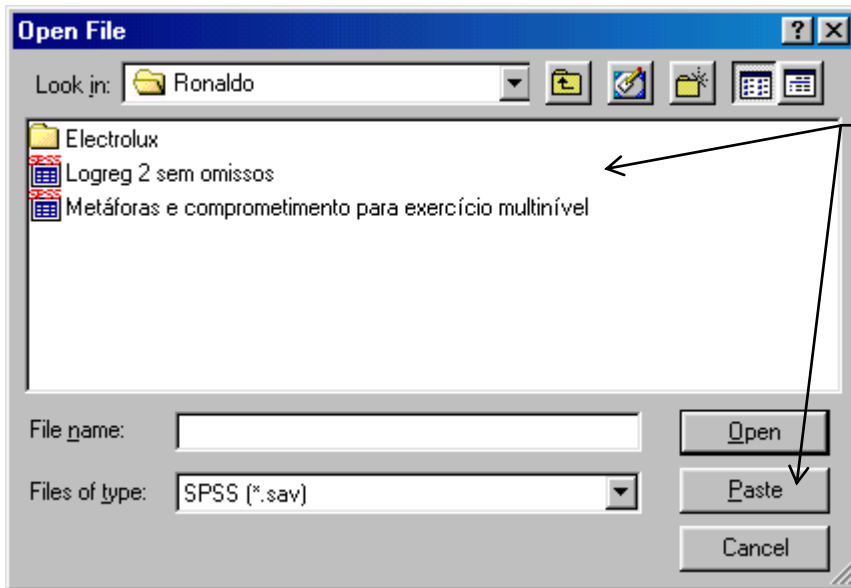


Figura 6.1 – Janela do comando *Open*

Escolha o arquivo desejado e pressione o botão indicado.

Esse mesmo procedimento, o botão *Paste*, pode ser utilizado em diversas caixas de diálogo do SPSS para geração dos arquivos *syntax* de comando. A figura 6.2 apresenta uma janela *syntax* com diferentes comandos já “copiados” em sua estrutura.

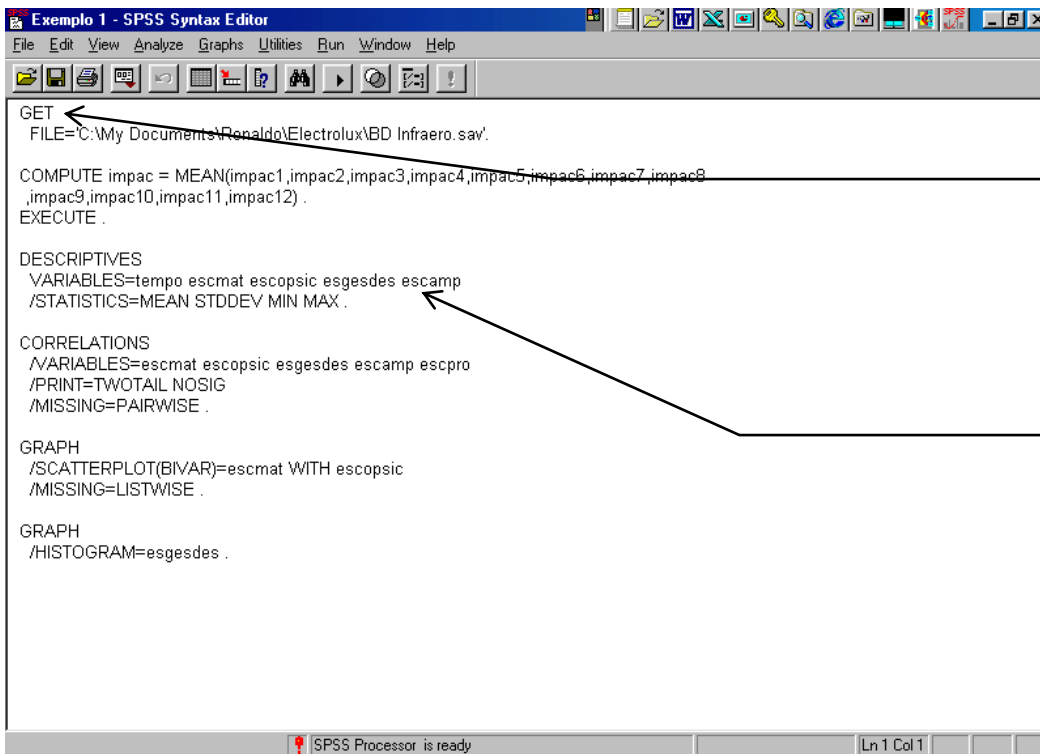


Figura 6.2 – Janela de arquivo *syntax*

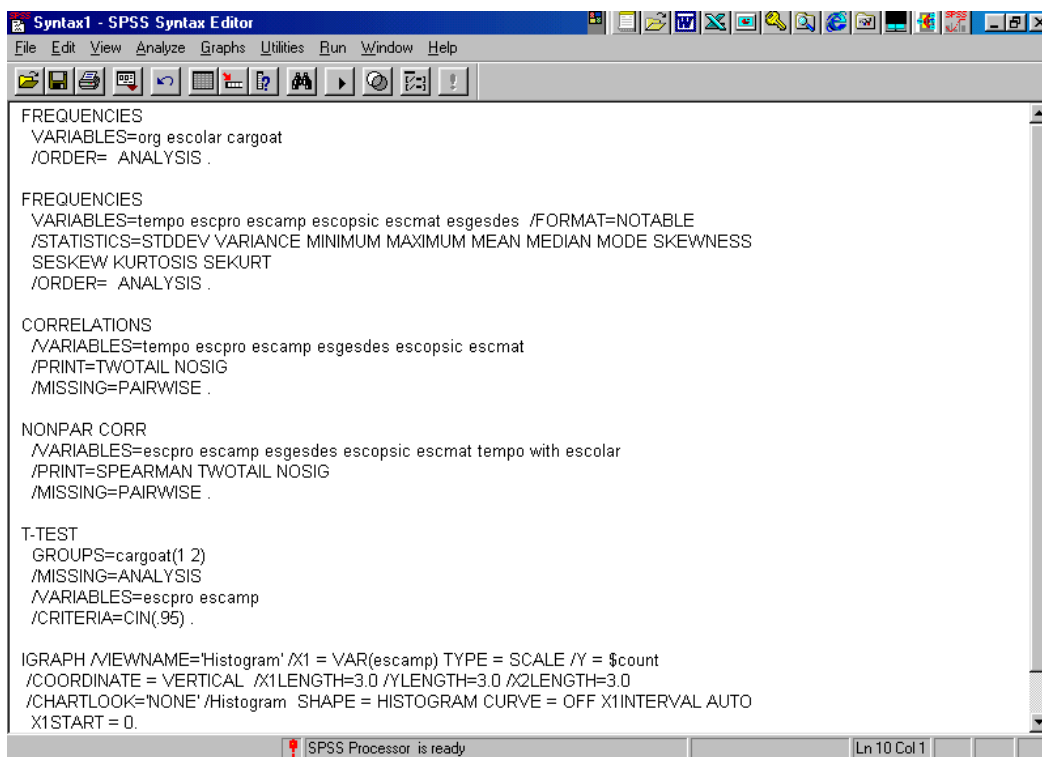
Resultado do comando executado na figura anterior

Variáveis selecionadas para análise descritiva.

Na mesma figura ainda são apresentados outros procedimentos de análise e transformação de dados. O comando *COMPUTE* faz a criação de escores fatoriais, como

apresentado na Figura 5.6. Dois comandos GRAPH, para geração de gráficos; no exemplo um gráfico de dispersão e um histograma. Também é apresentado no exemplo um comando CORRELATIONS.

Utilizando este procedimento é possível a geração de arquivos de dados com rotinas de trabalho, para geração de relatórios em avaliação de treinamento. A Figura 6.3 apresenta um exemplo de comandos para relatório de avaliação de treinamento, nos níveis de reação, aprendizagem, comportamento no cargo (impacto do treinamento) e mudança na organização.



```
Syntax1 - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Window Help
FREQUENCIES
  VARIABLES=org escolar cargoat
  /ORDER= ANALYSIS .

FREQUENCIES
  VARIABLES=tempo escpro escamp escapsic escmat esgesdes /FORMAT=NOTABLE
  /STATISTICS=STDDEV VARIANCE MINIMUM MAXIMUM MEAN MEDIAN MODE SKEWNESS
  SESKEW KURTOSIS SEKURT
  /ORDER= ANALYSIS .

CORRELATIONS
  /VARIABLES=tempo escpro escamp esgesdes escapsic escmat
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE .

NONPAR CORR
  /VARIABLES=escpro escamp esgesdes escapsic escmat tempo with escolar
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE .

T-TEST
  GROUPS=cargoat(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=escpro escamp
  /CRITERIA=CIN(95) .

IGRAPH /VIEWNAME='Histogram' /X1 = VAR(escamp) TYPE = SCALE /Y = $count
/COORDINATE = VERTICAL /X1LENGTH=3.0 /YLENGTH=3.0 /X2LENGTH=3.0
/CHARTLOOK='NONE' /Histogram SHAPE = HISTOGRAM CURVE = OFF X1INTERVAL AUTO
X1START = 0.

SPSS Processor is ready Ln 10 Col 1
```

Figura 6.3 – Janela de syntax com exemplo de rotina de relatório

Neste exemplo são solicitados vários tipos de análises que podem ser úteis em um relatório de avaliação de treinamento. Inicialmente foram solicitadas frequências das variáveis discretas do banco de dados. Neste comando a intenção é identificar a frequência de casos de cada uma das variáveis indicadas no comando. Em seguida foi solicitado o comando FREQUENCIES com especificação de índices relativos as variáveis contínuas (e.g. impacto do treinamento em amplitude e profundidade). Essas informações iniciais possibilitam identificar os índices descritivos do resultado do

treinamento, já indicando se o treinamento modificou o comportamento do treinando no cargo.

Em seguida, foram solicitadas duas correlações, para obter o grau de relação existente entre algumas variáveis. Neste caso a intenção é comparar se existe algum tipo e relação entre o resultado do treinamento no trabalho e variáveis de contexto e dos indivíduos. Também foi solicitada uma correlação não paramétrica, indicada para relacionar variáveis contínuas e categorias, neste exemplo escolaridade com o resultado do treinamento no trabalho.

Neste exemplo, também foi solicitado o cálculo de um teste T, para averiguação do nível de significância dos escores desta amostra em relação a população. Por último foi solicitado um gráfico interativo, através do comando IGRAPH, para observação do padrão de distribuição da variável impacto do treinamento no trabalho.

Este exemplo hipotético pode ser o escopo de informações de um relatório de apresentação dos resultados de impacto do treinamento no trabalho de um grupo de treinandos. Obviamente que a equipe de técnicos da organização deve definir quais são suas necessidades e também o interesse dos consumidores do resultado de um sistema de avaliação de treinamento.