

Biomatemática e Bioestatística II

SPSS

Departamento de Produção e Sistemas

Escola Superior de Tecnologia da Saúde do Porto

Título:

Biomatemática e Bioestatística II - SPSS

Autores:

Ana Paula Nascimento, Agostinho Cunha, Carla Teixeira Lopes, Fátima Monteiro,
Liliana Pereira, Sandra Alves, Vítor Bento

Ano:

2005

Sumário

1. Introdução ao SPSS para Windows.	14
1.1. Opções gerais do programa	14
1.2. Janelas do SPSS	15
1.3. Menu principal. Barra de ferramentas. Barra de estado.....	17
Menu principal.....	17
Barra de ferramentas.	18
Barra de estado.....	18
1.4. Procedimento básico para uma análise de estatística.	19
1.5. Gráficos.....	19
2. Arquivo de Dados.....	22
2.1. Arquivo de Dados em SPSS	22
Criar um arquivo.....	22
Abrir um arquivo.....	23
Guardar um arquivo	25
Informação sobre arquivos de dados	26
2.2. Janela de edição de dados.....	26
Definição de variáveis	27
Entrada de dados.....	30
Edição de Dados.....	30
2.3. Transformações sobre os dados	31
Criação de variáveis.....	31
Recodificação de variáveis.....	34
Outras Transformações.....	38
2.4. Manipulação de arquivos.....	41
3. Estatística Descritiva.....	47

3.1. Medidas de Estatística Descritiva	47
3.2. Frequências	48
Tabela de Frequências.....	48
Estatísticas Descritivas	52
Gráficos	55
3.3. Cálculo de estatísticas básicas.....	60
3.4. Análise exploratória.....	63
Estatísticas.....	65
Diagrama de caixa e fio (Boxplots).....	68
Diagrama de tronco e folha (Steam-and-leaf).....	71
3.5. Tabelas	74
Modificação de estrutura	74
Modificação de formatação	78
4. Testes de hipóteses paramétricos.....	80
4.1. Condições de aplicação dos Testes Paramétricos	80
Teste de Kolmogorov-Smirnov	80
Teste de Shapiro-Wilk	80
Teste de Normalidade	81
4.2. Resultados ao Teste de Normalidade.....	83
4.3. Médias	84
4.4. Uma amostra.....	87
Testes-t simultâneos para a média de uma variável agrupada por outra variável	90
4.5. Duas amostras com dados independentes.....	93
4.6. Duas amostras com dados emparelhados.....	95
5. Testes de hipóteses não paramétricos.....	100
5.1. Teste do Qui-Quadrado.....	101
Qui- quadrado para a qualidade do ajuste.....	101

Exemplo Prático	104
Qui-Quadrado para independência de variáveis.....	106
Exemplo Prático	109
5.2. Binomial	111
Exemplo Prático	112
5.3. Duas amostras Independentes (Mann-Whitney).....	113
Exemplo Prático	115
5.4. Two Related Samples (Wilcoxon).....	116
Exemplo Prático	117
6. Regressão	120
6.1. Regressão Linear Simples	120
6.2. Regressão Linear Múltipla.....	123
6.3. Estimação da Curva	124
Teste aos coeficientes.....	125
R e R^2	126

Sumário de Figuras

Figura 2: Janela de entrada no SPSS.	14
Figura 3: Janela de Edição de Dados.	15
Figura 4: Janelas de Edição de Dados, de Resultados e de Sintaxe.	16
Figura 5: Janela de Edição de Gráficos.	16
Figura 6: Janela de Sintaxe.	17
Figura 7: Barras de menus e ferramentas.	17
Figura 8: Gráfico com a representação de duas variáveis.....	20
Figura 9: Gráfico com a representação de uma variável.	20
Figura 10: Gráfico Simples.....	20
Figura 11: Gráfico encaixado.	20
Figura 12: Menu para a criação de uma nova janela de edição de dados.....	22
Figura 13: Menu para abrir uma base de dados já existente.....	23
Figura 14: Ícone para abrir uma base de dados.....	23
Figura 15: Janela de diálogo para a indicação da localização de um ficheiro a ser aberto.....	24
Figura 16: Janela de diálogo com Wizard para a importação de dados para SPSS.	25
Figura 17: Menu para guardar um ficheiro.	25
Figura 18: Ícone para aceder à informação acerca de uma variável.	26
Figura 19: Janela de Data View.	26
Figura 20: Janela de Variable View.....	27
Figura 21: Janela de diálogo para definir o tipo de variável.....	28
Figura 22: Janela de diálogo para introduzir valores para a variável.....	28
Figura 23: Janela de diálogo para definir missings.....	29
Figura 24: Menus para inserir casos e variáveis.	30
Figura 25: Ícone para posicionar-se num determinado caso.	31
Figura 26: Ícone para encontrar dados.	31

Figura 27: Janela de diálogo para criar variáveis através de transformações noutras.	32
Figura 28: Janela com a informação acerca de uma variável do ficheiro Cars.	33
Figura 29: Janela de diálogo para criar variável accel_mi em função da variável accel presente no ficheiro Cars.	33
Figura 30: Janela de diálogo que permite seleccionar casos onde a transformação vai ser efectuada.	34
Figura 31: Janela de diálogo para a recodificação automática da variável cityname...	35
Figura 32: Resultado da recodificação automática da variável cityname.....	35
Figura 33: Janela de diálogo para a indicação do novo código.	36
Figura 34: Janela onde é exemplificada a recodificação dos missings.....	38
Figura 35: Janela de diálogo para criar categorias de forma automática.....	38
Figura 36: Janela de dados onde se mostra um exemplo de uma determinada base de dados.	39
Figura 37: Janela de diálogo que permite executar a função count.....	40
Figura 38: Janela de diálogo que permite dar a condição para proceder à contagem.	40
Figura 39: Janela de dados resultante do procedimento Transform Count.	41
Figura 40: Janela de diálogo que permite ordenar os dados de acordo com vários critérios.	42
Figura 41: Janela de dados resultante do procedimento Data Sort.	42
Figura 42: Janela de diálogo que permite executar o Split file.	43
Figura 43: Janela que exemplifica o Split file efectuado pela variável sexo (gender). .	44
Figura 44: Janela de diálogo para seleccionar dados.	45
Figura 45: Janela de diálogo para especificar a amostra aleatória.....	45
Figura 46: Janela de dados onde se mostra uma base de dados contendo a informação de altas verificadas num determinado hospital.	46
Figura 47: Tabela de frequências da variável dia da semana.	46
Figura 48 – Menu Analyze > Descriptive Statistics > Frequencies	48
Figura 49 – Janela de Frequências.....	49

Figura 50 – Tabela de frequências.....	49
Figura 51 – Tabela de Frequências com casos omissos.....	51
Figura 52 – Janela de Frequências – opção Format.....	52
Figura 53 - Menu Analyze > Descriptive Statistics > Frequencies.....	52
Figura 54 – Janela de Frequências.....	53
Figura 55 – Frequencies: Statistics.....	53
Figura 56 – Estatísticas descritivas.....	54
Figura 57 - Menu Analyze > Descriptive Statistics > Frequencies.....	56
Figura 58 – Janela de Frequências.....	56
Figura 59 – Frequencies: Charts.....	57
Figura 60 – Gráfico de barras por frequência absoluta.....	57
Figura 61 – Gráfico de barras por frequência relativa.....	58
Figura 62 - Gráfico circular.....	58
Figura 63 – Histograma.....	59
Figura 64 – Histograma com curva normal.....	59
Figura 65 – Janela de edição de Gráficos.....	60
Figura 66 – Analyze / Descriptive Statistics / Descriptives.....	61
Figura 67 – Janela Descriptives.....	61
Figura 68 – Janela Descriptives / Options.....	62
Figura 69 – Resultado do procedimento Descriptives.....	62
Figura 70 – Scores z.....	63
Figura 71 – Menu Explore.....	64
Figura 72 – Janela Explore.....	64
Figura 73 – Botão Statistics na janela Explore.....	65
Figura 74 – Explore: Statistics.....	66
Figura 75 – Estatísticas Descriptives do procedimento Explore para um grupo da categoria Primary Vehicle.....	66

Figura 76 – M-Estimators.....	67
Figura 77 - Outliers	67
Figura 78 – Percentiles	67
Figura 80 – Botão Plots na janela Explore	69
Figura 81 – Plots em Explore.....	70
Figura 82 - Diagrama caixa e fio	71
Figura 83 – Botão Plots na janela Explore	72
Figura 84 – Plots em Explore.....	72
Figura 85 – Tabela do SPSS	74
Figura 86- Menu de configuração de tabelas	75
Figura 87 – Pivoting Trays	76
Figura 88 – Inserir informação numa nova camada	76
Figura 89 – Tabela com uma nova camada	77
Figura 90 – Troca de linha para coluna.....	77
Figura 91 – Tabela com variável Primary vehicle em coluna.....	78
Figura 92 – Opção Toolbar	78
Figura 93 – Opção TableLooks	79
Figura 94 – Janela TableLooks	79
Figura 95: Menu a seguir para efectuar o teste à normalidade.....	81
Figura 96: Janela de diálogo no teste à normalidade onde se colocam as variáveis a testar e as que vão criar grupos.	82
Figura 97: Janela onde se selecciona a opção para efectuar o teste à normalidade... 83	
Figura 98: Tabela de resultado do teste à normalidade, que apresenta o número de casos válidos, missings e totais.	83
Figura 99: Tabela resultante do teste à normalidade, com os dois testes Kolmogorov-Smirnov e Shapiro-Wilk.....	84
Figura 100: Menu para efectuar o procedimento <i>Compare Means-> Means</i>	85

Figura 101: Janela de diálogo do procedimento <i>Compare Means-> Means</i> onde é visível onde se deverão colocar as variáveis dependentes e independentes.	85
Figura 102: Janela de diálogo onde é possível seleccionar as estatísticas a pedir.	86
Figura 103: Tabela resultante o procedimento <i>Compare Means-> Means</i> onde se mostram número de casos incluídos, excluídos e total.....	86
Figura 104: Tabela resultante do procedimento <i>Compare Means-> Means</i> onde se apresentam as estatísticas pedidas.	87
Figura 105: Menu para a execução do procedimento Teste t para uma amostra.	88
Figura 106: Janela de diálogo resultante da selecção do teste t para uma mostra onde é possível escolher qual a variável e o valor a testar.....	88
Figura 107: Janela de diálogo onde se pode indicar o nível de confiança para o intervalo.	89
Figura 108: Tabela resultante do teste t para uma amostra onde se apresentam algumas estatísticas da variável a testar.	90
Figura 109: Tabela resultante do teste t para uma amostra onde se apresenta entre outras informações o valor p, do respectivo teste.....	90
Figura 110: Janela onde se efectuará uma partição de todos os procedimentos em função da variável sexo.....	91
Figura 111: Exemplo de um teste t para uma amostra, mostra a variável e o valor a testar assim como o nível de significância.....	92
Figura 112: Tabela resultante de teste t simultâneos para uma variável agrupada por outra, onde se apresentam algumas estatísticas para os dois grupos (neste caso masculino e feminino).	92
Figura 113: Tabela resultante de teste t simultâneos para uma variável agrupada por outra, onde se apresentam entre outras informações o valor p para os dois testes.	92
Figura 114: Menu para a execução do teste t para amostras independentes.....	93
Figura 115: Janela de Diálogo onde se devem especificar os grupos.....	94
Figura 116: Tabela resultante da realização do teste t para duas amostras independentes na qual são visíveis algumas estatísticas.....	94

Figura 117: Tabela resultante da realização do teste t para duas amostras independentes na qual são mostrados os valores p assim como o intervalo de confiança para a diferença de médias.	95
Figura 118: Janelas de Diálogo para a realização do teste t para amostras emparelhadas.	97
Figura 119: Tabela resultante da realização do teste t para amostras emparelhadas, onde são visíveis algumas estatísticas.....	98
Figura 120: Tabela resultante da realização do teste t para amostras emparelhadas, onde é visível o coeficiente de correlação.	98
Figura 121: resultante da realização do teste t para amostras emparelhadas, onde é visível o valor p, assim como o intervalo de confiança para a média das diferenças.	98
Figura 122: Menu onde se encontram os testes não paramétricos.	101
Figura 123: Janela de Diálogo para a realização do teste do qui-quadrado onde se devem indicar as variáveis assim como os valores a testar.....	102
Figura 124: Janela de diálogo onde se pode optar por um teste exacto ou por uma aproximação à normal.	103
Figura 125: Janela de Diálogo onde se pode optar por “pedir” algumas estatísticas, assim como o tratamento aos missings.....	104
Figura 126: Tabela resultante do teste do qui-quadrado do ajuste onde se apresentam valores observados e esperados, assim como os resíduos.....	105
Figura 127: Tabela resultante do teste do qui-quadrado do ajuste onde se apresenta o	105
Figura 128: Janela de diálogo para proceder à realização do teste do qui-quadrado para a independência de variáveis, onde se deverão indicar quais as variáveis a testar.....	107
Figura 129: Janela de diálogo onde se podem “pedir” algumas estatísticas relativas às variáveis a testar no teste do qui-quadrado para a independência de variáveis	107
Figura 130: Janela de diálogo onde se pode optar por quais as informações a apresentar na célula da tabela de contingência.....	108

Figura 131: Janela de diálogo onde se pode optar qual a forma de ordenação dos dados.....	109
Figura 132: Tabela resultante do teste do qui-quadrado para independência, onde se mostra o número de casos válidos, missings e total.....	109
Figura 133: Tabela de contingência resultante da execução do teste do qui-quadrado.	110
Figura 134: Tabela resultante da execução do teste do qui-quadrado onde se mostram os valores prova.....	110
Figura 135: Janela diálogo para a execução do teste da binomial, onde se devem indicar as variáveis assim como as proporções a testar.....	112
Figura 136: Tabela resultante do teste da binomial onde são visíveis as proporções observadas, a proporção a testar assim como o valor p.....	113
Figura 137: Janela de diálogo para a execução de testes não paramétricos para amostras independentes, onde se deve indicar quais as variáveis e grupos a testar.....	114
Figura 138: Tabela resultante do teste de Mann-Whitney onde se apresentam os grupos em análise, a dimensão, a posição média do grupo e o somatório de posições.....	115
Figura 139: Tabela resultante do teste de Mann-Whitney, onde se apresenta o valor p.	115
Figura 140: Janela de diálogo para a execução dos testes não paramétricos para amostras emparelhadas.....	116
Figura 141: Tabela resultante do teste de Wilcoxon onde se apresentam a média e soma das posições.	118
Figura 142: Tabela resultante do teste de Wilcoxon onde se apresenta o valor p.	118
Figura 143: Representação das possíveis relações lineares entre duas variáveis. ...	120
Figura 144: Menu para a execução da regressão linear.	121
Figura 145: Janela de diálogo para a execução da regressão linear onde se deve indicar quais as variáveis dependente e independentes.....	122

Figura 146: Tabela resultante de uma regressão linear onde se apresentam os coeficientes.	122
Figura 147: Tabela resultante de uma regressão linear múltipla onde se apresentam os coeficientes.	123
Figura 148: Menu para a estimação da curva.	125
Figura 149: Tabela resultante da regressão linear onde se apresentam os coeficientes de determinação e correlação.	126

1. Introdução ao SPSS para Windows.

O SPSS (Statistical Package for Social Sciences –Pacote Estatístico para as Ciências Sociais) é uma “Solução modular para análise estatística de dados, incluindo poderosos sistemas de gestão de dados e de aplicação de procedimentos estatísticos num ambiente gráfico, utilizando menus e caixas de diálogo de utilização muito fácil”, segundo a empresa Produtos e Serviços de Estatística, Lda, que detém o direitos de comercialização em Portugal deste software.

Este software foi inventado, nos Estados Unidos, por Norman H. Nie, C. Hadlai (Tex) Hull e Dale H. Bent e desde a sua primeira versão, em 1968, já foram desenvolvidas doze novas versões. Este software, além da sua aplicação nas Ciências Sociais para onde foi inicialmente concebido, é também usado em pesquisas relacionadas com saúde e educação, pesquisas governamentais e de mercado e inúmeros outros sectores. O primeiro manual de usuário foi publicado em 1970 por Nie e Hall. A primeira versão do SPSS para computadores pessoais surgiu em 1984.

1.1. Opções gerais do programa

Para entrar no SPSS basta fazer um duplo clique no ícone SPSS. Aparecerá uma janela, como é exemplificado a seguir:

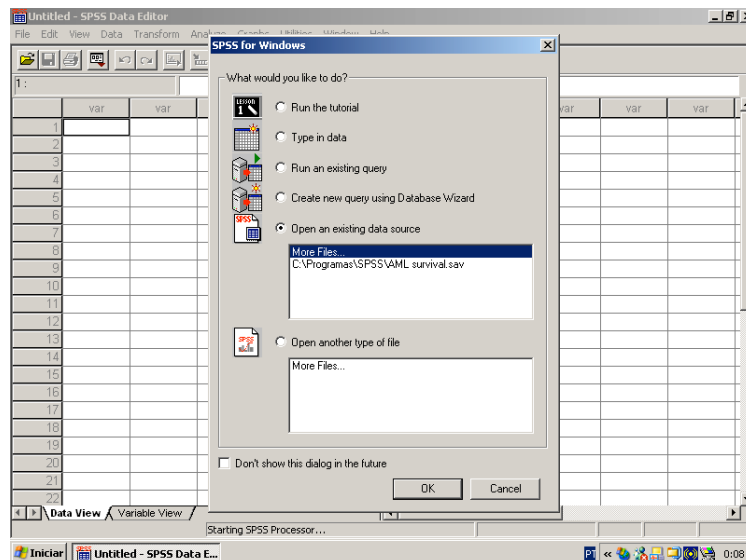


Figura 1: Janela de entrada no SPSS.

Onde se terá que escolher uma das opções.

Também podemos abrir uma sessão do SPSS fazendo duplo clique sobre o ícone correspondente o qualquer arquivo do tipo SPSS:

- **de dados:** têm extensão *.sav* e estão no formato SPSS
- **de gráficos:** têm extensão *.cht* (gráfico de barras, histograma, etc.)
- **de texto:** têm extensão *.sps* (*arquivo de sintaxe*) ou *.lst* (*arquivo de resultados*).

1.2. Janelas do SPSS

Basicamente a aplicação SPSS dispõe de três janelas onde os dados e os resultados estatísticos podem ser criados e manipulados ou exibidos. A janela de edição de dados, a janela de exibição de resultados e gráficos, e a janela de edição de gráficos.

O editor de dados activa uma janela com uma estrutura semelhante a uma folha de cálculo clássica que contém o arquivo de dados que estamos a utilizar. Esta janela abre-se automaticamente ao iniciar a sessão do SPSS.

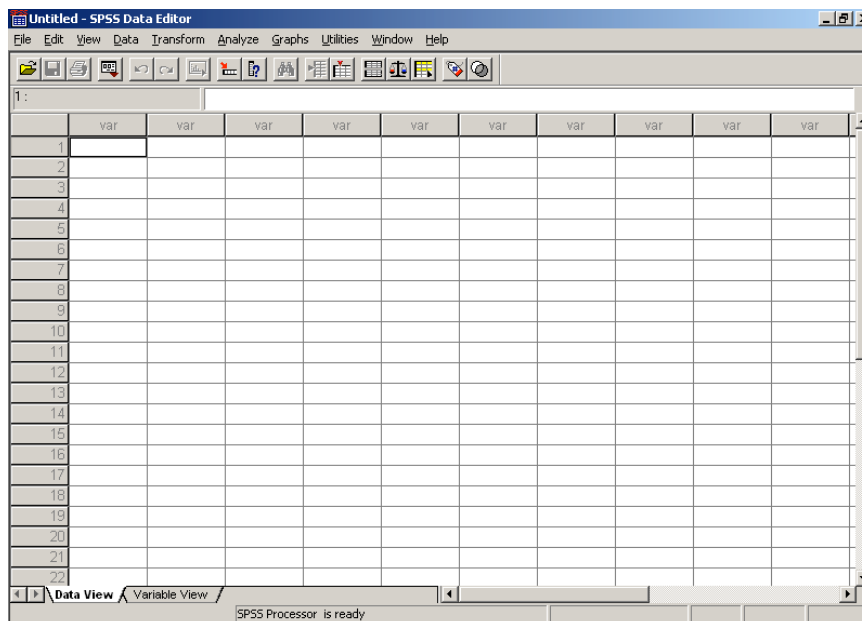


Figura 2: Janela de Edição de Dados.

A janela de exibição de resultados e gráficos é onde aparecem os resultados das análises realizadas com o programa. Podemos archivar os resultados para utilização posterior.

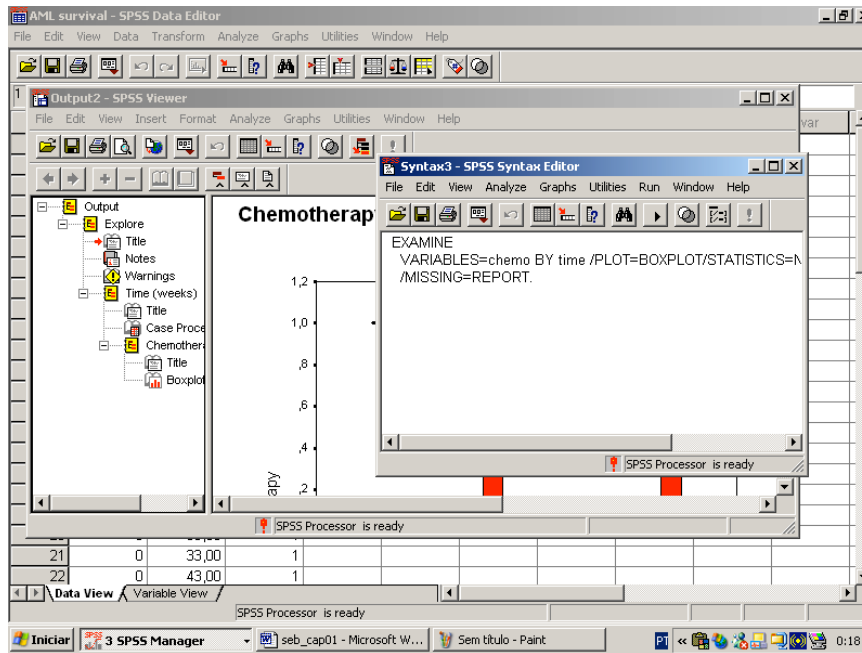


Figura 3: Janelas de Edição de Dados, de Resultados e de Sintaxe.

A janela de gráficos permite modificar e guardar gráficos e arquivos com extensão *.cht*.

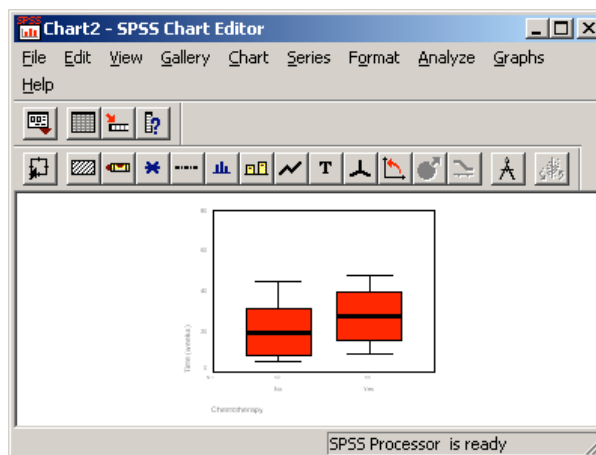


Figura 4: Janela de Edição de Gráficos.

Existe ainda uma outra janela que é a janela de sintaxe onde podemos colar os comandos seleccionados a partir de qualquer caixa de diálogo que corresponde à efectuação de uma ou mais análises estatísticas. Podemos guardar estes comandos num ficheiro de sintaxe (com extensão *.sps*) o que nos permite aplicar as mesmas análises estatísticas a ficheiros de dados com igual estrutura.

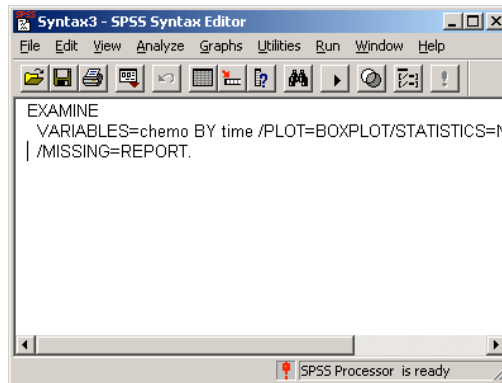


Figura 5: Janela de Sintaxe.

1.3. Menu principal. Barra de ferramentas. Barra de estado.

O conjunto de comandos disponível na barra de menus e na barra de ferramentas permite manipular ficheiros de dados (criar, editar, transformar, gravar e imprimir), realizar procedimentos estatísticos e criar gráficos.

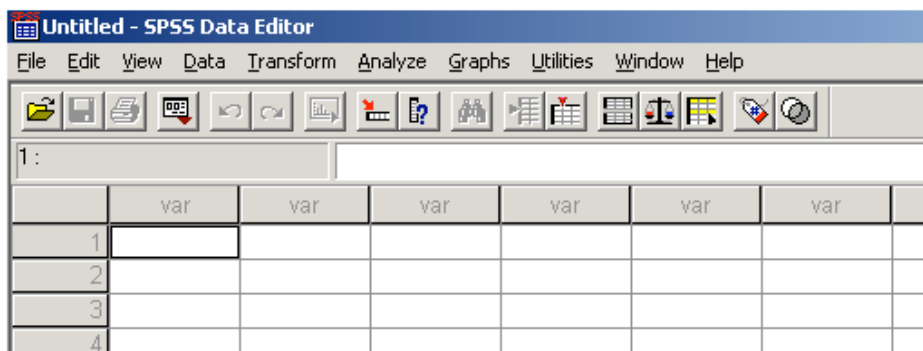


Figura 6: Barras de menus e ferramentas.

Menu principal.

A partir deste menu acedemos à maioria das funções do SPSS. O menu principal tem dez opções:

- **File**- criar um novo ficheiro SPSS, abrir um existente, guardar, ler dados criados com outra aplicações, etc.
- **Edit**- contém as habituais opções do Windows para copiar dados, procurar, etc.
- **View**- permite alterar o aspecto da janela de edição de dados.

- **Data-** contém opções para efectuar mudanças que afectam todo o arquivo de dados (unir arquivos, transpor variáveis e casos, criar subconjuntos de casos, etc.). Estas mudanças são temporárias enquanto não se guarda explicitamente o ficheiro.
- **Transform-** efectua mudanças sobre as variáveis seleccionadas, permite a criação de novas variáveis.
- **Analyze-** a partir desta opção são executados todos os procedimentos estatísticos.
- **Graphs-** gráficos de barras, histogramas, etc.
- **Utilities-** obter informação completa do arquivo de dados, aceder a um índice de comandos do SPSS, etc.
- **Window-** ordenar, seleccionar, controlar atributos das janelas abertas.
- **Help-** permite aceder à ajuda da aplicação.

Barra de ferramentas.

Situada abaixo da barra do menu principal, a barra de ferramentas exhibe um conjunto de ícones que, uma vez activados com o rato, realizam os procedimentos mais frequentes. Colocando o rato sobre o ícone, é exibida uma descrição sobre a tarefa activada.

Tanto a barra de ferramentas como a de estado se podem activar ou desactivar a partir da opção do menu View/Toolbar ou Status Bar. Também podemos mudar o tamanho dos ícones com a opção View/Toolbar.

Barra de estado.

Situada na base da janela da aplicação, indica o estado actual do processo e as seguintes informações complementares:

- O número de casos processados.
- Filtro de dados, se utilizarmos somente uma parte do arquivo de dados. É apresentada a mensagem *Filter on*.
- Variável ponderada, se utilizarmos o critério de ponderação. É apresentada a mensagem *Weight on*.

- Segmentação de dados, se trabalharmos com o arquivo de dados dividido em subgrupos a partir de uma ou mais variáveis de agrupamento. É apresentada a mensagem *Split File on*.

1.4. Procedimento básico para uma análise de estatística.

Para levar a cabo qualquer tipo de análise com SPSS é necessário efectuar estas três operações básicas:

1. Seleccionar uma base de dados
2. Seleccionar o procedimento estatístico pretendido (menu principal)
3. Seleccionar as variáveis a incluir na análise e outros parâmetros adicionais (que surgem nas caixas de diálogo)

1.5. Gráficos.

A representação em gráficos dos dados de tabelas, quadros e folhas de cálculo propicia uma leitura fácil e sedutora da informação. A informação estatística representada em gráficos é apreendida pelo destinatário de uma forma fácil e intuitiva.

Para além dos gráficos de utilização corrente, acessíveis em qualquer folha de cálculo, o aplicativo SPSS apresenta formatos particulares de gráficos de grande qualidade visual, adequados, quer à representação de informação específica, quer ao resumo de grandes quantidades de informação.

Em termos gerais, em SPSS, os gráficos são estruturados do seguinte modo:

1. Em função da natureza da variável estatística – nominal, dicotómica ou contínua.
2. Do número de variáveis a representar no mesmo gráfico.
3. Da natureza independente ou encaixada dos dados a representar.

Decorrentes dessa estrutura, em termos genéricos, os gráficos do SPSS podem ser tipificados de duas formas:

1. Considerando o número de variáveis a representar no gráfico:

Gráfico com medidas (de localização ou dispersão) de duas ou mais variáveis contínuas.

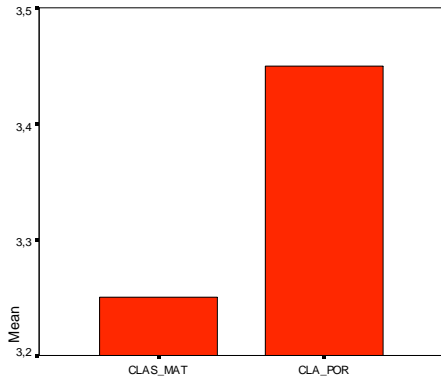


Figura 7: Gráfico com a representação de duas variáveis.

Gráfico de uma variável.

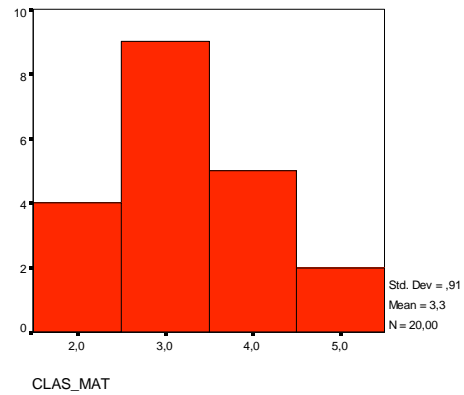


Figura 8: Gráfico com a representação de uma variável.

2. Considerando a natureza independente ou encaixada das variáveis representadas:

Gráfico simples – Alturas dos alunos de uma turma.

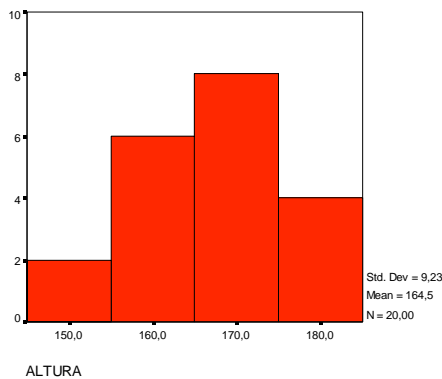


Figura 9: Gráfico Simples.

Gráficos encaixados – Notas em matemática dos rapazes e raparigas de uma turma.

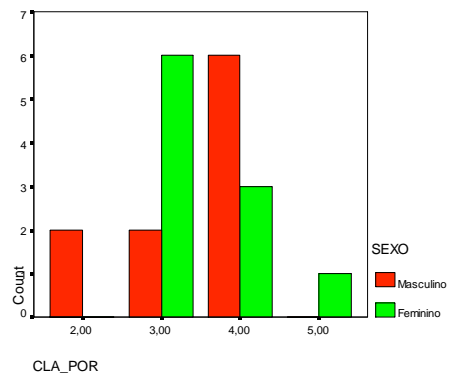


Figura 10: Gráfico encaixado.

Uma proposta de resumo dos gráficos mais populares em SPSS é a apresentada no quadro seguinte:

Tipos de variáveis	Tipos de gráficos
Resumo dos valores de uma variável nominal ou quantitativa discreta Casos individuais.	Gráfico de barras Gráfico de linhas Gráfico circular Gráfico de área Gráfico pareto
Estatísticas de duas ou mais variáveis contínuas	Gráfico de barras Gráfico de linha Gráfico circular Boxplot
Variáveis contínuas	Boxplot Histograma Diagrama de caule-e-folhas

A aplicação SPSS apresenta um conjunto de formatos de gráfico susceptíveis de ser configurados em função da natureza da variável ou variáveis que se pretendem representar. Quando se cria um gráfico no menu Graph, é exibida uma caixa de diálogo onde se realizam as selecções adequadas.

2. Arquivo de Dados

Os dados são uma parte fundamental do processo de análise estatística, o SPSS possibilita, para além da análise estatística dos dados, o armazenamento. Este armazenamento pode ser encarado sobre duas perspectivas; conforme os dados são organizados pela primeira vez no SPSS, e se cria a base de dados no programa, ou se a informação já foi de alguma forma organizada, e o armazenamento é apenas temporário, i.e., apenas do tempo de análise.

2.1. Arquivo de Dados em SPSS

Criar um arquivo

Criar um arquivo de dados no SPSS onde se faça o registo da informação, ou seja onde se crie uma nova base de dados, pode ser conseguido através do menu **File**→**New**→**Data**. Uma nova janela de dados (Data Editor) em branco é disponibilizada.

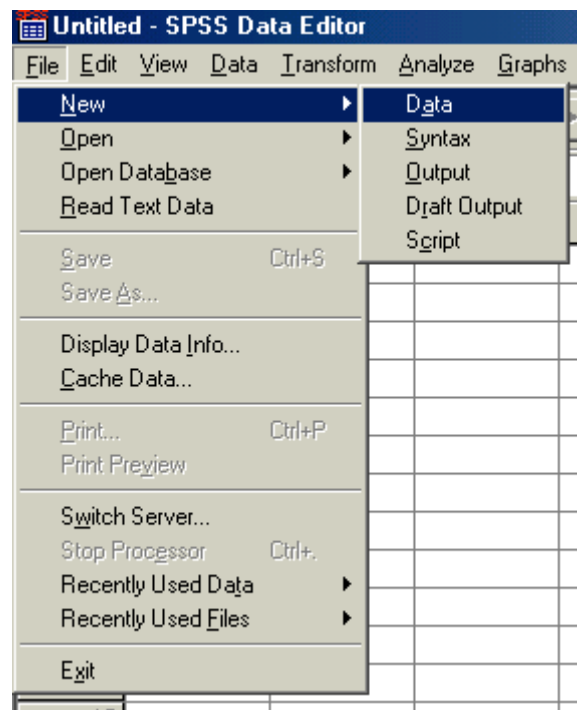


Figura 11: Menu para a criação de uma nova janela de edição de dados.

Abrir um arquivo

Para abrir um arquivo de dados já existentes, basta aceder ao menu **File→Open→Data**

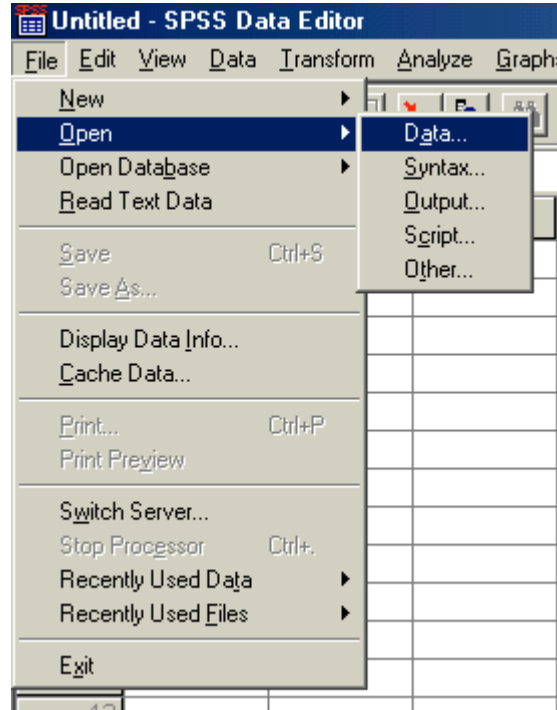


Figura 12: Menu para abrir uma base de dados já existente.

Ou aceder ao ícone disponível na barra de ferramentas



Figura 13: Ícone para abrir uma base de dados.

Em qualquer um dos casos aparecerá uma janela de diálogo onde deverá ser “indicada” a directoria, ou pasta onde se encontra o ficheiro pretendido, assim como o tipo de ficheiro que o arquivo é e finalmente o seu nome.

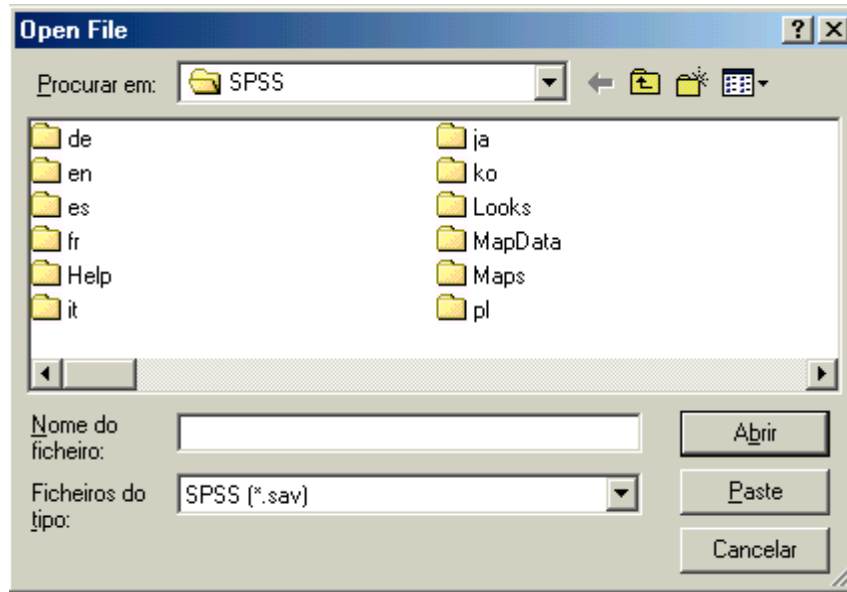


Figura 14: Janela de diálogo para a indicação da localização de uma ficheiro a ser aberto.

Tipos de arquivos que o SPSS reconhece

De entre os tipos de arquivos que o SPSS reconhece estão os ficheiros Excel (*.xls), os ficheiros gravados em dBase (*.dbf) e os ficheiros de texto (*.txt). A extensão dos ficheiros gravados em SPSS é sav.

O software irá ler dos ficheiros Excel e dBase, o nome das variáveis, que deverão cumprir certos requisitos (ver secção 2.2.1.), assim como abrirá casos e linhas conforme o n.º de colunas e linhas ocupadas. Os ficheiros de texto são importados com ajuda de um Wizard, que vai guiando cada passo e que possibilita a importação de ficheiros de dados delimitados por vírgulas, espaços, etc.

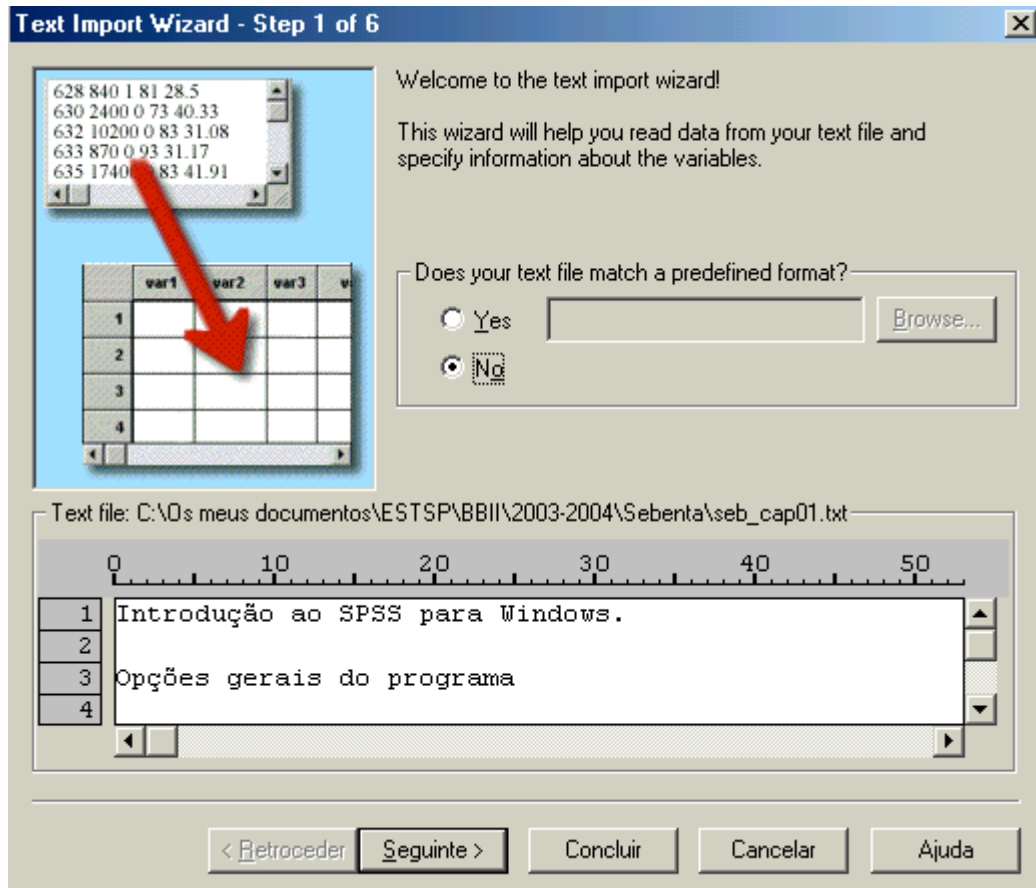


Figura 15: Janela de diálogo com Wizard para a importação de dados para SPSS.

Guardar um arquivo

As alterações feitas sobre um ficheiro só são válidas durante a sessão de SPSS em que são feitas, a não ser que o ficheiro seja gravado. Tal pode ser conseguido através do menu **File**→**Save**. O ficheiro alterado é guardado por cima do existente.

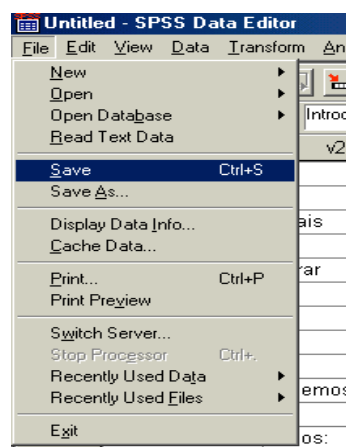


Figura 16: Menu para guardar um ficheiro.

Uma outra forma de guardar arquivos é através do menu **File**→**Save as**, onde se deve indicar o tipo de ficheiro e o nome que queremos dar ao ficheiro.

Informação sobre arquivos de dados

É possível aceder a informação sobre nome de todas as variáveis, formato das variáveis assim como informação acerca das “etiquetas” que as descrevem e os seus valores, de um ficheiro que se encontre aberto, através do menu **Utilities**→**File Info**. A “execução” deste menu produz um ficheiro de output contendo toda a informação acima referida.

Uma outra forma de aceder a informação sobre uma variável é através do ícone presente na barra de ferramentas.



Figura 17: Ícone para aceder à informação acerca de uma variável.

Finalmente é possível obter informação acerca de um ficheiro que não esteja aberto através do menu **File**→**Display Data Info**.

2.2. Janela de edição de dados

A janela de edição de dados é composta por duas janelas: a janela Data View e a janela Variable View. A primeira contém a informação sobre os dados dos casos em estudo a segunda a informação acerca das variáveis.

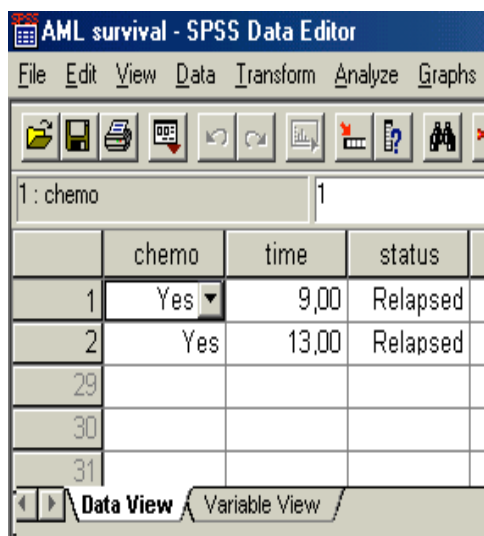


Figura 18: Janela de Data View.

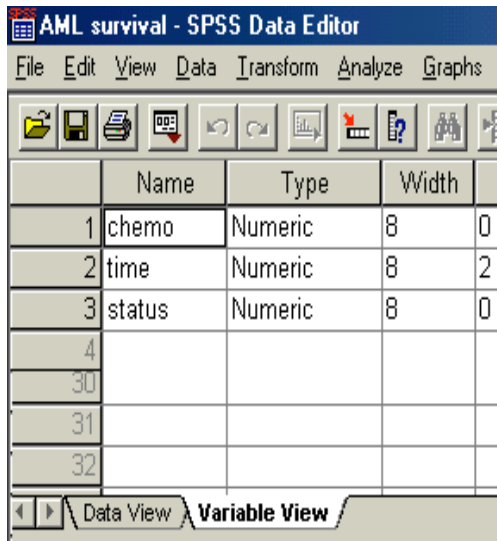


Figura 19: Janela de Variable View.

Definição de variáveis

Esta definição é feita na janela de edição de dados na janela variable view. Nesta janela cada linha contém informação de uma variável, informação essa que é definida através das colunas.

Name: deve ser indicada o nome da variável, este nome estará presente numa coluna da janela data view. Este nome deve seguir determinadas regras:

- Máximo 8 caracteres;
- Devem começar por uma letra e não acabar num ponto;
- Não serão aceites nomes em duplicado (o programa não distingue maiúsculas de minúsculas, o que significa que as variáveis exp e EXP têm o mesmo nome);

Type: deve ser indicado o tipo de variável, de uma escolha que o programa oferece na maioria dos casos os tipos utilizados são o tipo numérico (Numeric) e o data (Date), visto que mesmo que a natureza da variável seja string é possível codificá-la através de um código numérico.

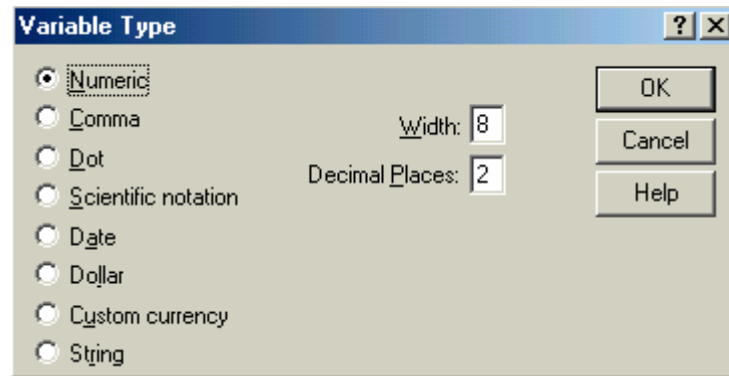


Figura 20: Janela de diálogo para definir o tipo de variável.

Width: este campo corresponde à dimensão da variável, define o n^o máximo de algarismos ou caracteres que o valor da variável pode tomar. Pode ser preenchido na janela variable view na coluna respectiva ou então na janela de diálogo do tipo.

Decimals: deve indicar-se o n.º de casas decimais que a variável pode tomar caso seja do tipo numérica. Do mesmo modo do campo width pode ser preenchido na janela variable view na coluna respectiva ou então na janela de diálogo do type.

Label: não é um campo obrigatório mas pode ser de extrema utilidade. Por exemplo, pode conter o nome por extenso da variável que devido ao facto de poder conter apenas 8 caracteres pode não ser explícito. Pode conter, por exemplo, a unidade na qual a variável está medida.

Values: onde é indicado o código utilizado para introduzir os valores da variável. Por exemplo: suponha-se que, para simplificar a introdução, se definia o valor 1 para o sexo masculino e 2 para o sexo feminino, então é aqui que se deve informar o sistema deste "código".

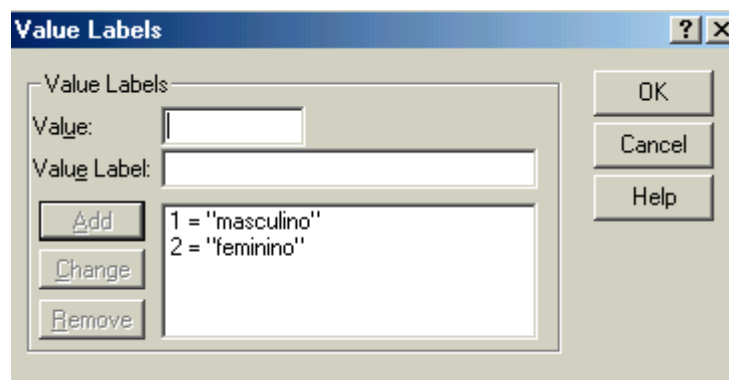


Figura 21: Janela de diálogo para introduzir valores para a variável.

Missing: onde se informa o sistema que um determinado valor representa uma situação anómala e não um valor aceitável para a variável. Este missing é designado user-missing, uma vez que o utilizador se dá conta e informa o sistema sobre essa possibilidade. Existem também system-missing, quando o sistema não é informado sobre a situação (qualquer célula em branco). Por exemplo, quando um indivíduo responde a um questionário e não responde a uma questão o utilizador pode deixar essa célula em branco (system-missing) ou informar o sistema que quando for introduzido um valor fora da gama admitida para a variável ele deve “interpretar” como uma não resposta.

Estes missings podem ser definidos de duas formas:

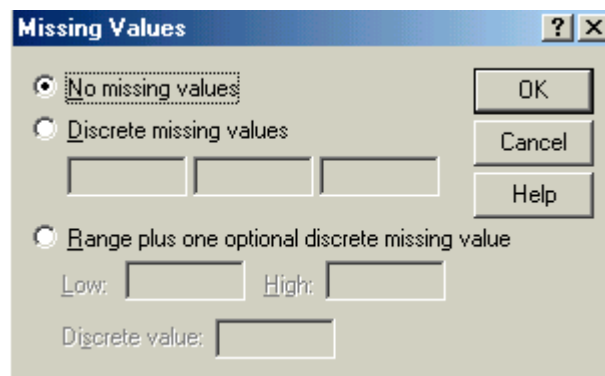


Figura 22: Janela de diálogo para definir missings.

- Através de pelo menos 3 valores discretos;
- Através de um intervalo de valores acrescido de um valor discreto.

Para se informar o sistema acerca do que significam tais valores deve ser usada a coluna values.

Columns: contém a definição do tamanho da coluna que conterà a informação da variável na janela data view.

Align: é onde se define qual o alinhamento da informação da coluna que conterà a informação da variável na janela data view. Pode ser alinhamento à esquerda, direita ou ao centro.

Measure: é onde se indica a escala de medida da variável, pode ser Nominal ou Ordinal para as variáveis qualitativas e Scale para as variáveis quantitativas.

NOTA (importante): A correcta definição das características das variáveis é essencial e pode comprometer, se não for feita correctamente, a análise a efectuar na variável.

Entrada de dados

A edição de dados é feita na janela Data View. Nesta janela:

As linhas correspondem a casos, i.e., suponha-se que queremos introduzir os dados de um questionário respondido por um certo n.º de indivíduos, cada linha corresponderá à informação contida num questionário.

As colunas correspondem as variáveis.

As células, que são o cruzamento dos casos com as variáveis, conterão a informação desse caso relativo a essa variável.

É necessário prudência para não ter casos e variáveis abertas que não contenham informação, pois serão consideradas para as variáveis numéricas, system-missings e para as variáveis string serão consideradas válidas e consideradas para análise.

Edição de Dados

É possível modificar total ou parcialmente o conteúdo de uma célula basta para isso seleccioná-la.

É possível apagar e copiar células, deve ter-se atenção quando se copia valores de células para que a célula destino tenha as mesmas características (type,...).

É possível inserir casos e variáveis através do menu **Data→Insert**.

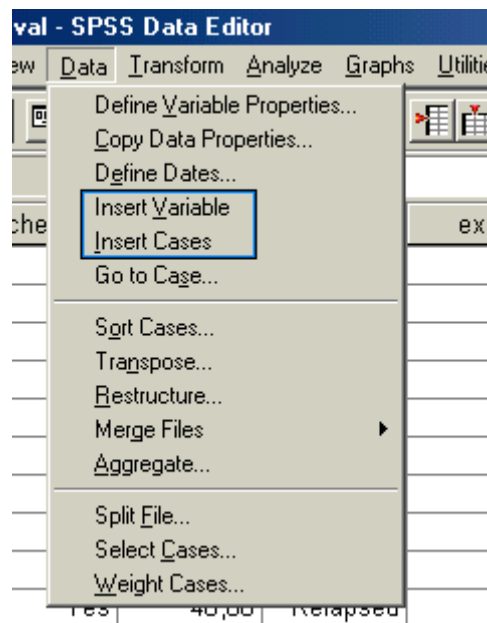


Figura 23: Menus para inserir casos e variáveis.

É possível retirar casos ou variáveis através do menu **Edit**→**Clear**.

É possível encontrar um caso sem ser sequencialmente, através do menu **Data**→**Go to Case**. Ou através do ícone que se segue, presente na barra de ferramentas.



Figura 24: Ícone para posicionar-se num determinado caso.

É possível encontrar dados através do menu **Edit**→**Find**, ou através do ícone que se segue, presente na barra de ferramentas.



Figura 25: Ícone para encontrar dados.

2.3. Transformações sobre os dados

Criação de variáveis

Para além da criação das variáveis através da janela **variable view**, o SPSS possibilita a criação de variáveis novas através de transformações nas já existentes.

O menu **Transform**→**Compute** atribui valores a uma variável nova, cujo nome deve ser indicado na *target variable*, através de transformações executadas noutras variáveis já existentes.

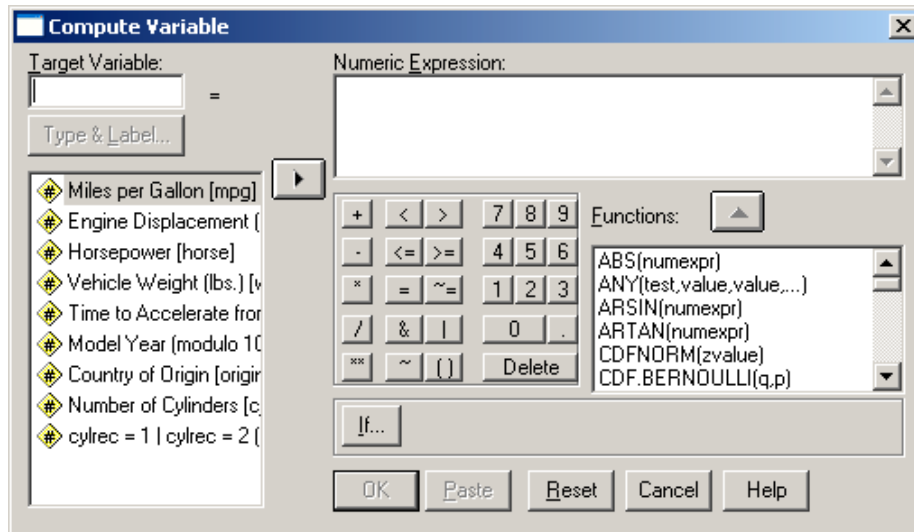


Figura 26: Janela de diálogo para criar variáveis através de transformações noutras.

Essas transformações devem ser definidas através da construção da expressão numérica (*numeric expression*). Nesta expressão numérica poderão constar variáveis existentes no ficheiro, constantes, operadores aritméticos e uma gama de funções, disponibilizadas pelo software, como o valor absoluto, o logaritmo, funções que transformam variáveis com valores em segundos em valores em minutos ou horas, como se pode verificar através do exemplo 1.

Exemplo1

No ficheiro Cars¹, encontra-se a variável *accel*, que guarda o tempo de aceleração das 0 às 60 milhas por hora, como é visível na janela que se segue.

¹ Este ficheiro é disponibilizado com o software SPSS, pode ser acedido (a não ser que se tenha procedido a alguma mudança de ficheiros) na pasta do SPSS que abre por defeito.

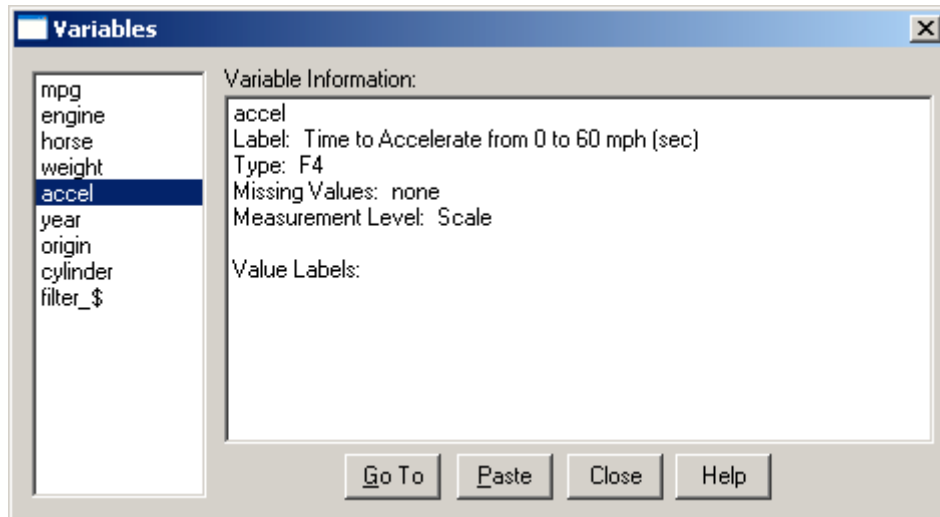


Figura 27: Janela com a informação acerca de uma variável do ficheiro Cars.

Suponha, agora que pretende alterar o tempo de aceleração para milhas por minuto, então o que deveria fazer era, por exemplo, criar uma variável nova, designá-la, por exemplo, *accel_mi*, e igualá-la à função `CTIME.MINUTES`, com argumento `accel`.

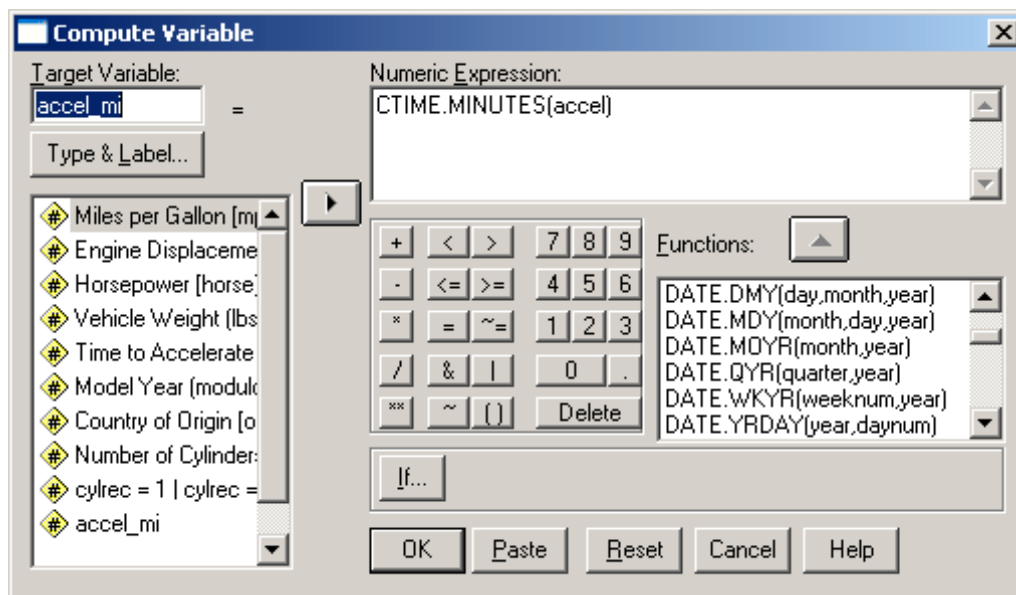


Figura 28: Janela de diálogo para criar variável *accel_mi* em função da variável *accel* presente no ficheiro Cars.

(Continua)

O botão *Type & Label* permite ao utilizador a especificação do tipo e atribuição de labels, de igual forma ao exposto no ponto 2.2. Uma opção importante nesta operação é a opção disponibilizada pelo botão *If...*, que permite efectuar a alteração apenas aos casos que verifiquem a condição indicada pela expressão.

Exemplo1 (Continuação):

Suponha que no exemplo, queremos apenas criar a nova variável *accel_mi*, quando o ano do carro for o ano 70. Então o procedimento a seguir seria na janela disponibilizada pelo botão *If*, inserir a condição *year=70* (*year*, variável que “guarda” o ano do automóvel)

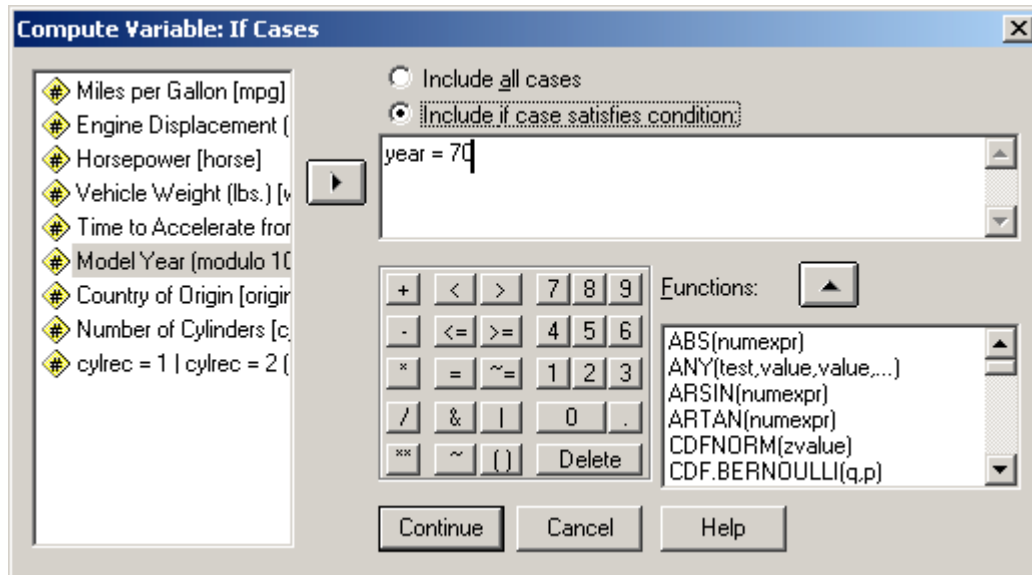


Figura 29: Janela de diálogo que permite seleccionar casos onde a transformação vai ser efectuada.

Recodificação de variáveis

Este procedimento possibilita a substituição dos valores de uma variável por outros. O SPSS possibilita esta transformação de duas formas uma das quais automática.

Através do menu **Transform**→**Automatic Recode**, o programa permite a transformação dos valores de uma variável noutros automaticamente. Um dos exemplos mais úteis na utilização deste menu, é a transformação de uma variável string em numérica, cujo nome deve ser explicitado na janela que resulta da execução desta opção. Automaticamente, o programa criará uma janela de output onde dará informação ao utilizador de qual a codificação que realizou.

Exemplo 2

Suponha que no ficheiro flying², que queria transformar a variável cityname, do tipo string numa variável numérica de nome city_2. Este nome deverá ser introduzido no rectângulo em branco mais pequeno e depois para que o nome seja aceite deve ser pressionado o botão *New Name*.

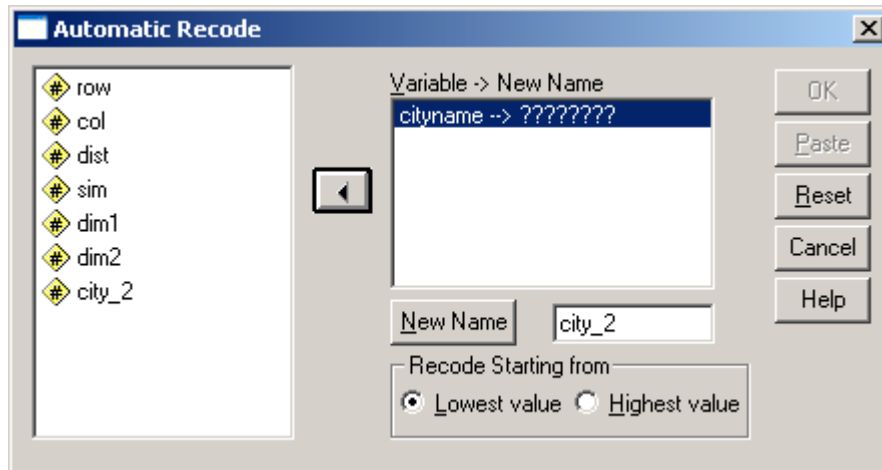


Figura 30: Janela de diálogo para a recodificação automática da variável cityname.

O output criado pelo SPSS é o seguinte, onde se podem ver os valores antigos, os novos valores e as suas “etiquetas”. Como se pode observar o software codificou também as células vazias, os missings.

CITYNAME	CITY_2	
Old Value	New Value	Value Label
	1	
ATLANTA	2	ATLANTA
CHICAGO	3	CHICAGO
DENVER	4	DENVER
HOUSTON	5	HOUSTON
LOS ANGELES	6	LOS ANGELES
MIAMI	7	MIAMI
NEW YORK	8	NEW YORK
SAN FRANCISCO	9	SAN FRANCISCO
SEATTLE	10	SEATTLE
WASHINGTON D.C.	11	WASHINGTON D.C.

Figura 31: Resultado da recodificação automática da variável cityname.

(Continua)

² De igual modo o ficheiro encontra-se (a não ser que se tenha procedido a alguma mudança de ficheiros) na pasta do SPSS que abre por defeito.

Através do menu **Transform**→**Recode**, o programa possibilita também, a recodificação, no entanto é o utilizador que deverá dar indicações de como essa recodificação será feita. Este menu possibilita, ele próprio, duas possibilidades: a recodificação *Into Same Variables* e a recodificação *Into Different Variables*. A diferença entre as duas possibilidades é apenas o “lugar” onde se guarda a nova codificação, na primeira possibilidade o novo código é guardado na variável original, na segunda o novo código é guardado numa nova variável, cujo nome deverá ser indicado na janela possibilitada pelo menu **Transform**→**Recode**→ **Into Different Variables**. Em ambas as opções existe a possibilidade de realizar as transformações apenas a alguns casos, aqueles que verificam uma determinada condição que deverá ser definida através do botão *If...*

É através do botão *Old and New Values* que o novo código deverá ser definido.

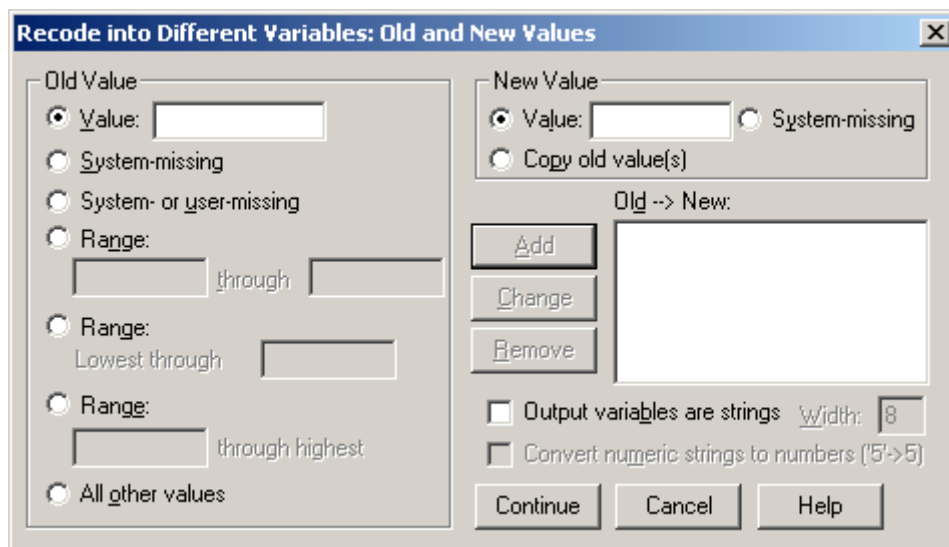


Figura 32: Janela de diálogo para a indicação do novo código.

Existem, duas zonas na janela que decorre do botão *Old and New Values*; uma para a zona dos “antigos” valores e uma zona para os “novos” valores.

Old Value:

- **Value:** o valor “antigo” a ser codificado;
- **System-missing:** se seleccionarmos esta opção recodifica-se os missing de sistema;
- **System-or user-missing:** se seleccionarmos esta opção recodifica-se os missing de sistema e os do utilizador;

- **Range: through:** através desta opção pode ser recodificada toda uma gama de valores, indicando o valor mais baixo e o mais alto.
- **Range: Lowest through:** através desta opção pode ser recodificada toda uma gama de valores desde o valor mais baixo até a um valor indicado.
- **Range: through highest:** através desta opção pode ser recodificada toda uma gama de valores desde um valor indicado até o maior valor.
- **All other values:** nesta opção estão englobados todos os valores não incluídos nas situações anteriores.

New Value:

- **Value:** onde deverá ser indicado o novo valor;
- **System-missing:** os novos valores serão desconhecidos do sistema;
- **Copy old value:** os “valores” antigos vão ser iguais aos novos. É importante referir que se existirem valores que se mantêm eles devem ser referidos senão serão excluídos na nova variável.

Esta opção só está presente na opção *Into Different Variables*.

Exemplo 2 (Continuação):

Suponha que agora quer recodificar a variável city_2, numa nova variável city_3, do seguinte modo:

missing	99
Todos os outros valores	igual

Então deverá recorrer à opção *Into Different Variables*, nos *Old and New Values* deverá definir os novos valores usando os System-missing nos old values

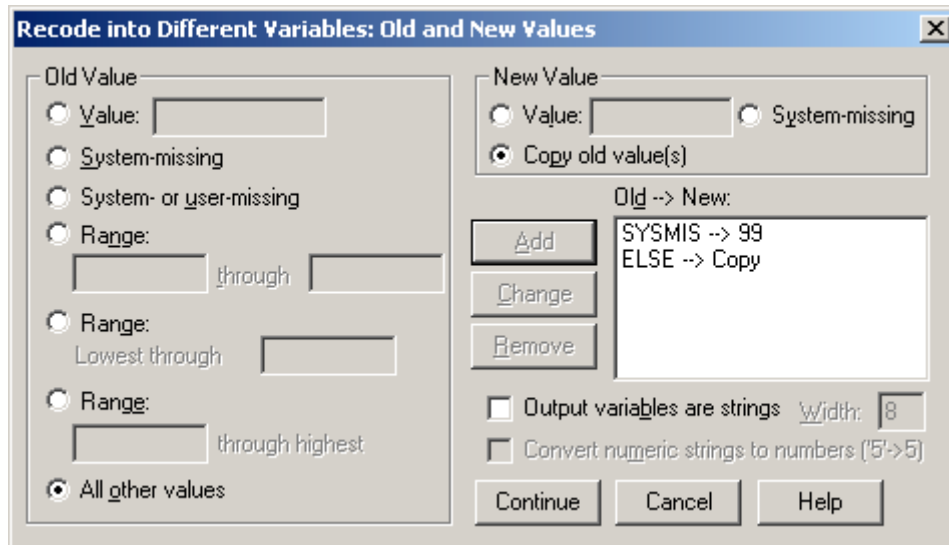


Figura 33: Janela onde é exemplificada a recodificação dos missings.

Outras Transformações

O menu **Transform**→**Categorize Variables** também possibilita a criação de categorias, mas de uma forma automática, o utilizador só tem a possibilidade de especificar o número de categorias a criar. Este procedimento é útil para agrupar dados de uma variável contínua, pois nem sempre é fácil neste tipo de variável a explicitação das categorias, assim basta ao utilizador dizer qual o número de categorias a criar e é criada uma variável contendo as categorias. O utilizador deve através de uma tabela de frequência, por exemplo, analisar quais os limites das classes criadas e criar rótulos.

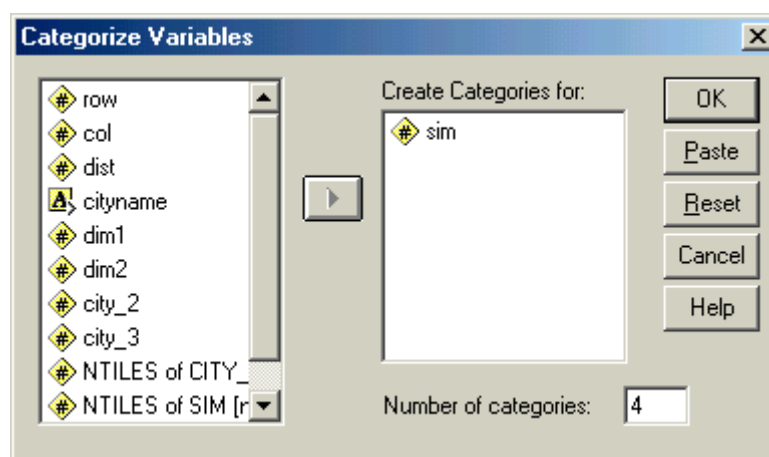


Figura 34: Janela de diálogo para criar categorias de forma automática.

Através do menu **Transform**→**Count** possibilita a criação de uma nova variável que contém, por cada caso, o número de vezes que um determinado valor apareceu nas variáveis seleccionadas.

Exemplo 3:

Suponha que tem um ficheiro de SPSS que contém para cada mês e para cada zona, o número de mortes por acidente, por causa natural e o número de acidentes devido a doença prolongada.

	mês	zona	acd	nat	doepro
1	1	1	231	564	45
2	1	2	23	565	21
3	1	3	32	154	245
4	2	1	4725	472	24
5	2	2	544	543	543
6	2	3	76	46	213
7	3	1	34	358	86
8	3	2	475	55	22
9	3	3	22	33	41
10	4	1	354	33	38
11	4	2	22	53	39
12	4	3	452	45	57
13					

Figura 35: Janela de dados onde se mostra um exemplo de uma determinada base de dados.

Suponha que quer saber quantas vezes o n.º de mortes devido a cada uma das causas foi superior a 100 em cada mês e em cada zona. Então poderá aceder ao menu **Transform**→**Count**, onde deverá explicitar o nome da nova variável, assim como o nome das variáveis onde pretende que o programa “procure” a condição que deverá ser definida através do botão *Define Values*.

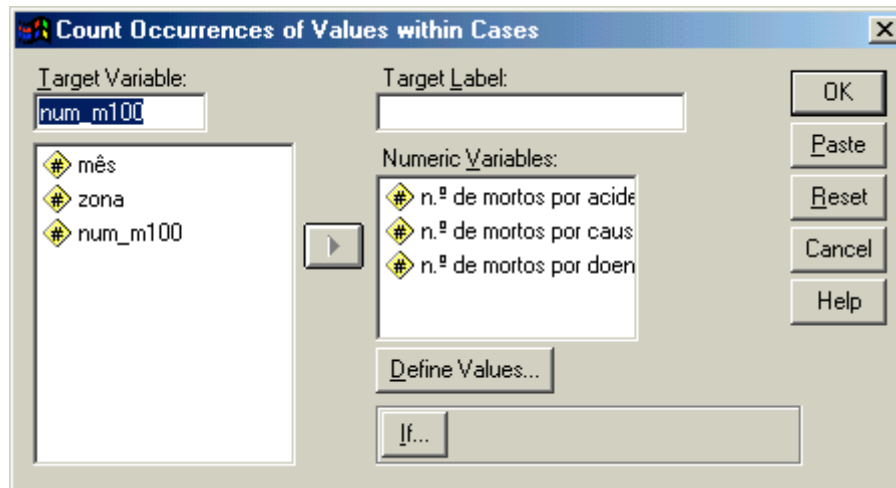


Figura 36: Janela de diálogo que permite executar a função count.

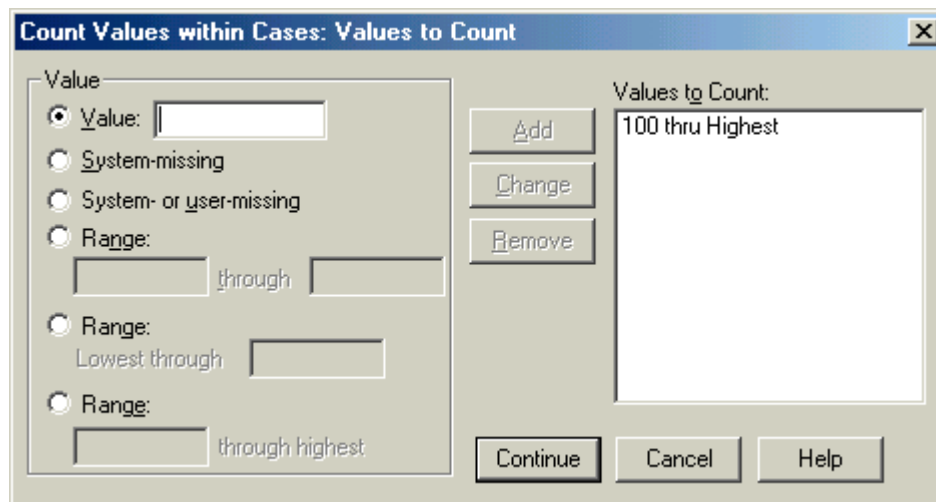


Figura 37: Janela de diálogo que permite dar a condição para proceder à contagem.

	mês	zona	acd	nat	doepro	num_m100
1	1	1	132	564	45	2
2	1	2	23	565	21	1
3	1	3	32	154	245	2
4	2	1	4725	472	24	2
5	2	2	544	543	543	3
6	2	3	76	46	213	1
7	3	1	34	358	86	1
8	3	2	475	55	22	1
9	3	3	22	33	41	0
10	4	1	354	33	38	1
11	4	2	22	53	39	0
12	4	3	452	45	57	1

Figura 38: Janela de dados resultante do procedimento Transform Count.

(Continua)

Nota:

Em todos estes procedimentos que resultam na criação de novas variáveis se deverá proceder à confirmação de todas as características da variável criada na janela *Variable View*.

2.4. Manipulação de arquivos

De seguida, expor-se-ão alguns procedimentos que permitem a manipulação de arquivos de dados, como é o caso do procedimento disponibilizado no menu **Data**→**Sort Cases**, que permite ordenar os casos através das características de uma ou mais variáveis. A primeira variável determina a primeira ordenação, as restantes determinam as ordenações dentro da primeira ordenação.

Exemplo 3 (Continuação):

Suponha que pretende organizar o ficheiro por ordem crescente de número de mortos por acidente e dentro desta ordenação pretende ordenar por ordem crescente de número de mortos devido a uma causa natural. Então deveria proceder da seguinte forma:

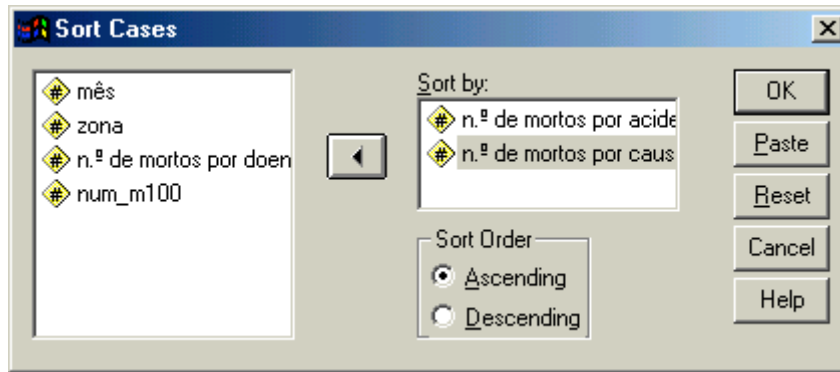


Figura 39: Janela de diálogo que permite ordenar os dados de acordo com vários critérios.

	mês	zona	acd	nat	doepro	num_m100
1	3	3	22	33	41	0
2	4	2	22	53	39	0
3	1	2	23	565	21	1
4	1	3	32	154	245	2
5	3	1	34	358	86	1
6	2	3	76	46	213	1
7	1	1	132	564	45	2
8	4	1	354	33	38	1
9	4	3	452	45	57	1
10	3	2	475	55	22	1
11	2	2	544	543	543	3
12	2	1	4725	472	24	2

Figura 40: Janela de dados resultante do procedimento Data Sort.

Através do menu **Data**→**Split File** é possível separar o ficheiro de dados em grupos definidos pelos valores de uma ou mais variáveis. Este procedimento é importante quando existe a suspeita da diferença entre determinados grupos, pois possibilita que toda a análise seja feita para cada grupo em separado.

Vejamos a janela disponibilizada pelo menu **Data**→**Split File**:

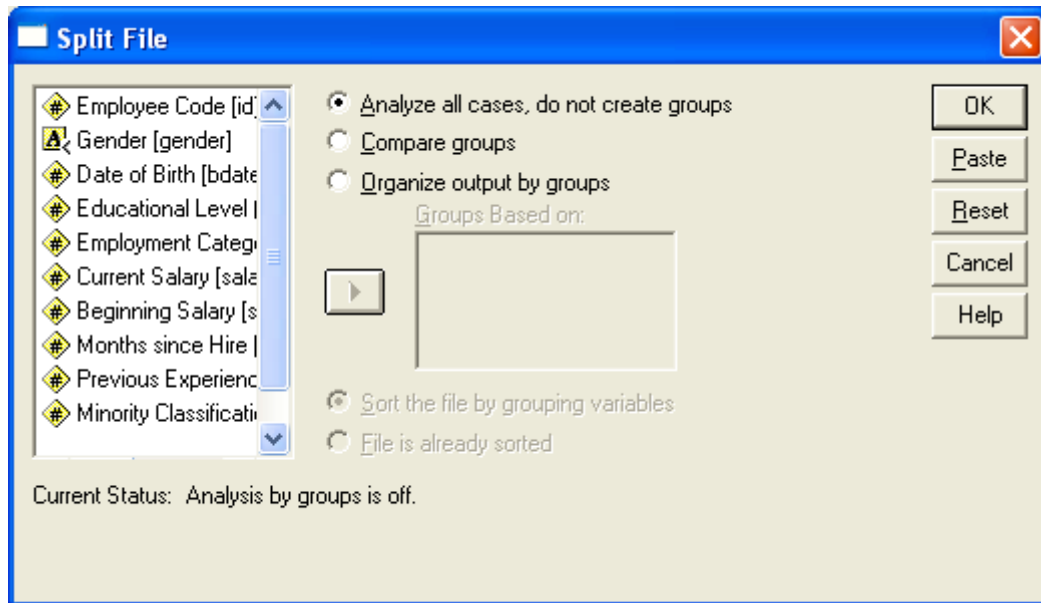


Figura 41: Janela de diálogo que permite executar o Split file.

Analyze all cases, do not create groups: desactiva qualquer separação feita anteriormente; esta opção está por defeito seleccionada, antes de qualquer divisão;

Compare groups: separa o ficheiro em grupos de modo a efectuar comparações;

Organize output by groups: qualquer procedimento produzirá outputs separados para cada grupo;

Groups Based on: é aqui que se deverá indicar quais as variáveis que servirão para criar os grupos, caso sejam introduzidas mais do que uma o programa agrupará pela ordem de introdução.

Sort the file by grouping variables: Ordena o ficheiro por grupos;

File is already sorted: dá-nos a indicação se o ficheiro já se encontra ou não ordenado;

Na parte inferior desta janela encontra-se uma indicação de como o ficheiro se encontra, **Current Status**.

Exemplo 4:

Suponha que pretende analisar as diferenças entre sexos dos empregados de uma determinada firma, cuja informação se encontra guardada no ficheiro Employee Data³. Então a variável que irá criar grupos será a variável Gender, como se pode ver na figura seguinte.

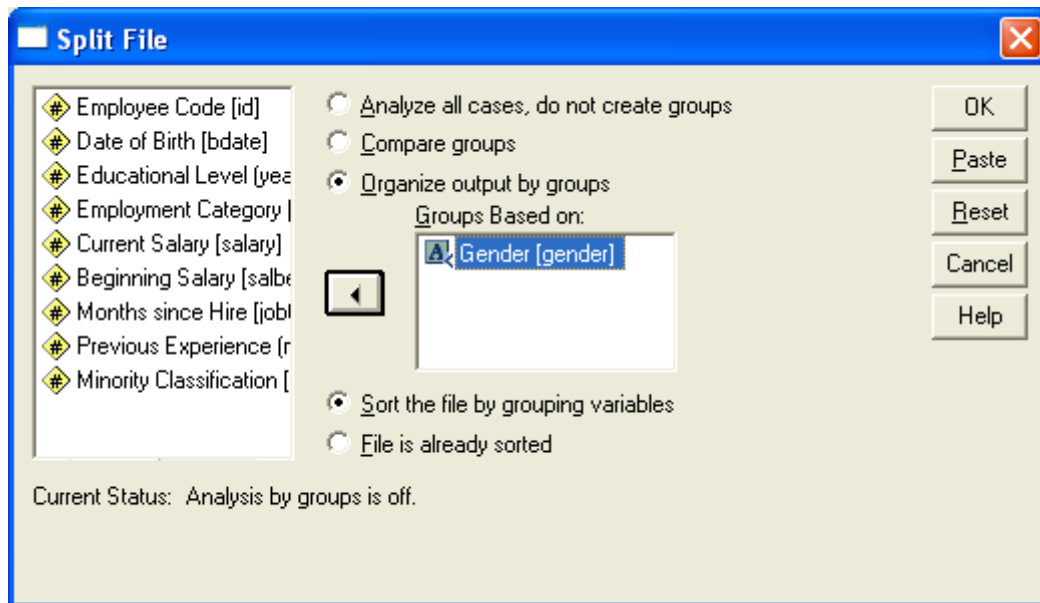


Figura 42: Janela que exemplifica o Split file efectuado pela variável sexo (gender).

Através do menu **Data**→**Select Cases** é possível seleccionar um grupo de casos a partir de um critério que pode incluir variáveis e expressões mais complexas. É possível através deste menu seleccionar uma amostra aleatória. Este procedimento possibilita a opção de apagar os casos que não satisfaçam as condições (**unselected cases are: deleted**) ou apenas não os incluir na análise (**unselected cases are: filtered**).

³ Este ficheiro também é disponibilizado pelo SPSS.

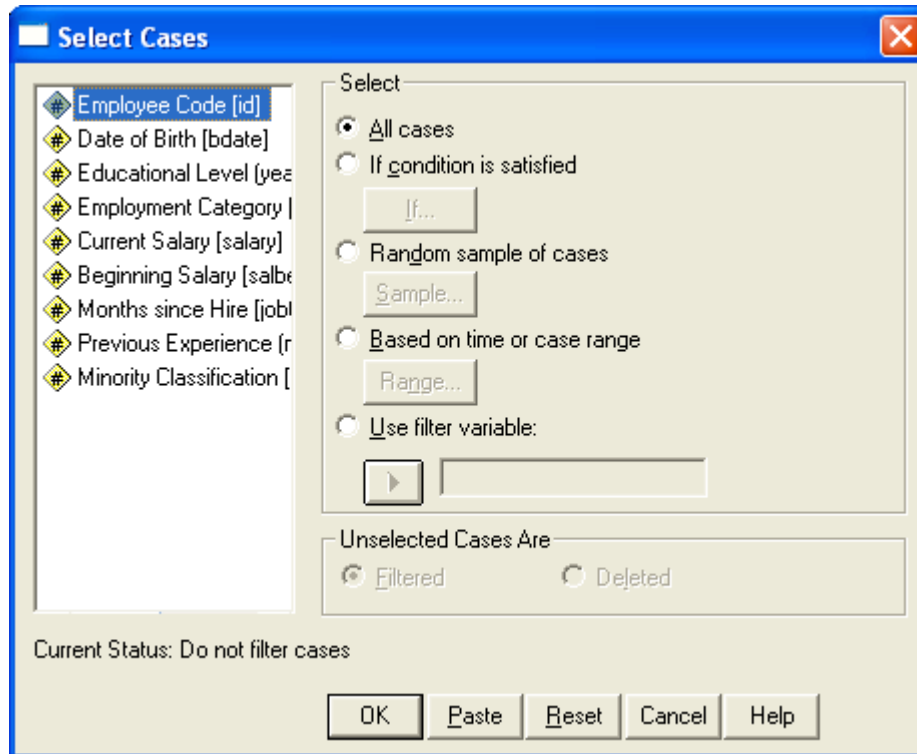


Figura 43: Janela de diálogo para seleccionar dados.

No caso da selecção de uma amostra aleatória então o utilizador tem a opção de indicar qual a percentagem de casos a incluir na amostra ou a possibilidade de indicar o número exacto de casos a escolher dos primeiros x (número também a indicar pelo utilizador).

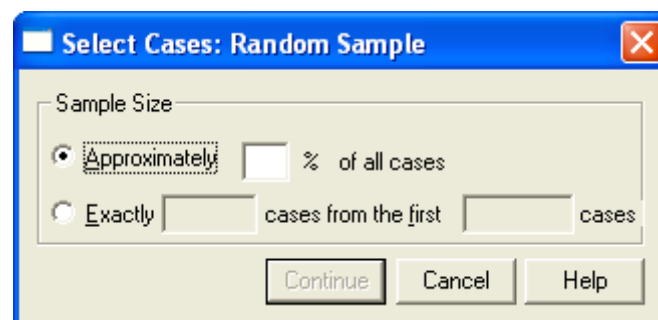


Figura 44: Janela de diálogo para especificar a amostra aleatória.

Sempre que este procedimento for efectuado é criada uma nova variável *filter* constituída apenas de 0 (zeros) e 1 (uns), conforme o caso não pertença ou pertença à amostra.

Através do menu **Data**→**Weight Cases** é possível criar uma ponderação cujos valores representam o número de casos observados.

Exemplo 5:

Suponha que possui um ficheiro de dados contendo a seguinte informação relativa às altas verificadas num determinado Hospital.

altas		655	
	dia_sema	altas	var
1	Domingo	20	
2	Segunda	175	
3	Terça	655	
4	Quarta	156	
5	Quinta	423	
6	Sexta	500	
7	Sábado	60	
8			

Figura 45: Janela de dados onde se mostra uma base de dados contendo a informação de altas verificadas num determinado hospital.

Se proceder à ponderação da variável dia da semana através da variável altas, então esta última passa a ser a frequência da primeira, como é visível se procedermos à execução de uma tabela de frequências.

		Dia da semana			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Domingo	20	1,0	1,0	1,0
	Segunda	175	8,8	8,8	9,8
	Terça	655	32,9	32,9	42,7
	Quarta	156	7,8	7,8	50,6
	Quinta	423	21,3	21,3	71,8
	Sexta	500	25,1	25,1	97,0
	Sábado	60	3,0	3,0	100,0
	Total	1989	100,0	100,0	

Figura 46: Tabela de frequências da variável dia da semana.

Note que se não procedermos à ponderação teríamos que introduzir o Domingo 20 vezes, a Segunda 175, a Terça 655, procedendo de igual forma para os restantes dias da semana.

3. Estatística Descritiva

A análise estatística univariada inicia o tratamento estatístico e deve ter em conta a tabela de frequências, a representação gráfica da distribuição da variável e as estatísticas apropriadas ao resumo dos dados em questão.

3.1. Medidas de Estatística Descritiva

As medidas de estatística descritivas apropriadas ao resumo dos dados dependem da escala de medida da variável a analisar. A seguinte tabela apresenta, resumidamente, as estatísticas apropriadas a cada tipo de variável.

	Medidas de tendência central	Medidas de dispersão	Medidas da forma global da distribuição
Escala intervalar ou escala racional	Média Moda Mediana	Desvio-padrão Variância Percentis Desvio interquartis Intervalo interquartis Intervalo absoluto Coeficiente de variação Score z	Coeficiente de assimetria de Pearson Medida de curtose
Escala ordinal	Moda Mediana	Percentis Desvio interquartis Intervalo interquartis Intervalo absoluto	
Escala nominal	Moda		

Tabela 1 – Medidas estatísticas apropriadas aos vários tipos de variáveis

3.2. Frequências

O procedimento *Analyze/Descriptive Statistics/Frequencies* permite obter informação sobre a distribuição dos valores da variável através de tabelas de frequências, de estatísticas descritivas e gráficos.

Este procedimento é útil para uma primeira análise de qualquer tipo de variável. A informação aqui obtida pode ajudar a determinar os valores típicos das variáveis (Que valores são mais frequentes? Qual a gama de valores que uma variável toma?), a verificar o cumprimento de determinadas assumpções necessárias para procedimentos estatísticos (Existe um número suficiente de observações? A distribuição da variável é adequada?) e a avaliar a qualidade dos dados (Existem valores omissos ou mal introduzidos? Existem valores que necessitam de ser recodificados?).

Tabela de Frequências

A tabela de frequências distribui os valores da variável em frequências simples e acumuladas, que podem ser absolutas ou relativas.

Para criar uma tabela de frequências no SPSS:

1. No menu *Analyze* escolher *Descriptive Statistics > Frequencies*.

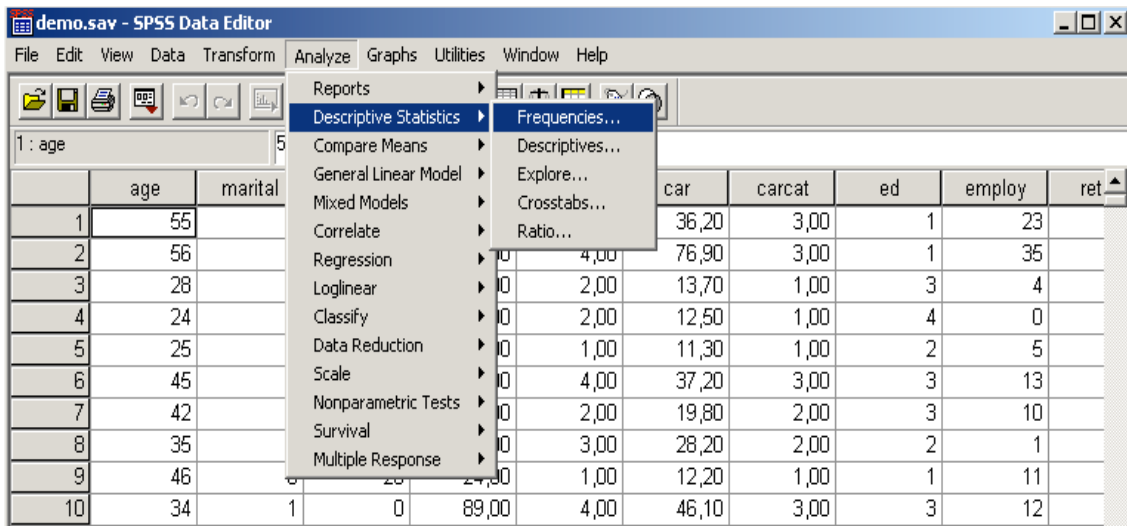


Figura 47 – Menu Analyze > Descriptive Statistics > Frequencies

2. Na janela que surge, seleccionar a(s) variável(is) que se pretende(m) analisar da lista de variáveis que aparece no lado esquerdo e através da seta, coloca-la no lado direito.

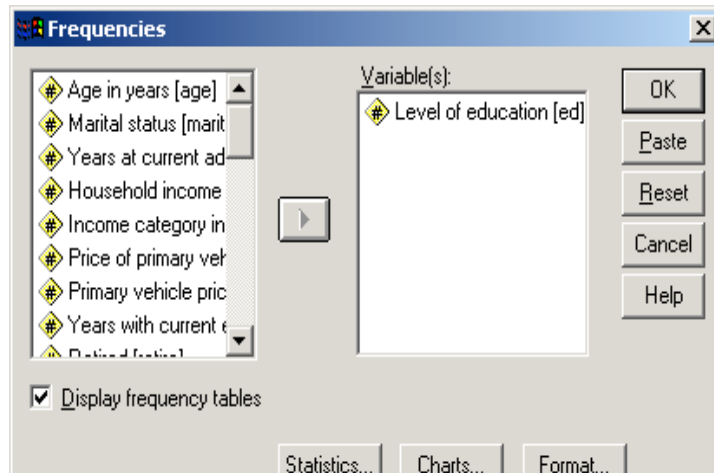


Figura 48 – Janela de Frequências

3. Selecciona a caixa de selecção *Display frequency table*. Por omissão, esta caixa encontra-se seleccionada.
4. Clicar em *OK*.

Na janela de saída serão apresentadas duas tabelas.

Statistics

Level of education

N	Valid	6400
	Missing	0

Level of education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	1390	21,7	21,7	21,7
	High school degree	1936	30,3	30,3	52,0
	Some college	1360	21,3	21,3	73,2
	College degree	1355	21,2	21,2	94,4
	Post-undergraduate degree	359	5,6	5,6	100,0
	Total	6400	100,0	100,0	

Figura 49 – Tabela de frequências

A primeira (*Statistics*), apresenta o número de casos válidos e casos omissos (*missing*). A segunda tabela é a tabela de frequências propriamente dita. Para os vários valores que a variável toma, são apresentados as seguintes frequências:

- *Frequency* (frequência simples e absoluta) – número de casos para um determinado valor (ou conjunto de valores) da variável. Por exemplo, no caso apresentado, 1390 pessoas não completaram o ensino secundário.
- *Percent* (frequência simples e relativa) - percentagem do número de casos em relação ao número total de casos (incluindo casos omissos). No caso apresentado, a % de pessoas que completou o ensino secundário é de 21,7%, valor que se obtém do quociente $1390/6400$, sendo 6400 o número de casos total (válidos + omissos).
- *Valid Percent* (frequência simples e relativa) – percentagem do número de casos em relação ao número de casos válidos. No caso apresentado, o número de casos válidos é igual ao número de casos total, pelo que a percentagem válida é igual à percentagem.
- *Cumulative Percent* (frequência acumulada e relativa) – percentagem acumulada até ao respectivo valor ou categoria, excluindo valores omissos.

A título de exemplo, apresenta-se de seguida uma tabela de frequências para uma situação onde existem casos omissos. Verifique-se na diferença entre a percentagem e a percentagem válida e numa nova linha no final da tabela para os casos omissos.

Statistics

Level of education

N	Valid	6386
	Missing	14

Level of education

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	1389	21,7	21,8	21,8
	High school degree	1934	30,2	30,3	52,0
	Some college	1353	21,1	21,2	73,2
	College degree	1351	21,1	21,2	94,4
	Post-undergraduate degree	359	5,6	5,6	100,0
	Total	6386	99,8	100,0	
Missing	System	14	,2		

Figura 50 – Tabela de Frequências com casos omissos

Na janela de frequências, na opção *Format* (ver figura 5), existem várias opções que permitem ordenar segundo diferentes critérios a tabela de frequências:

- Ascending Values – organiza a tabela de acordo com os valores das variáveis em ordem ascendente (do menor para o maior valor).
- Descending Values - organiza a tabela de acordo com os valores das variáveis em ordem descendente.
- Ascending Counts - organiza a tabela de acordo com as frequências de ocorrência dos valores das variáveis em ordem ascendente (do menos frequente para o mais frequente).
- Descending Counts - organiza a tabela de acordo com as frequências de ocorrência dos valores das variáveis em ordem descendente.

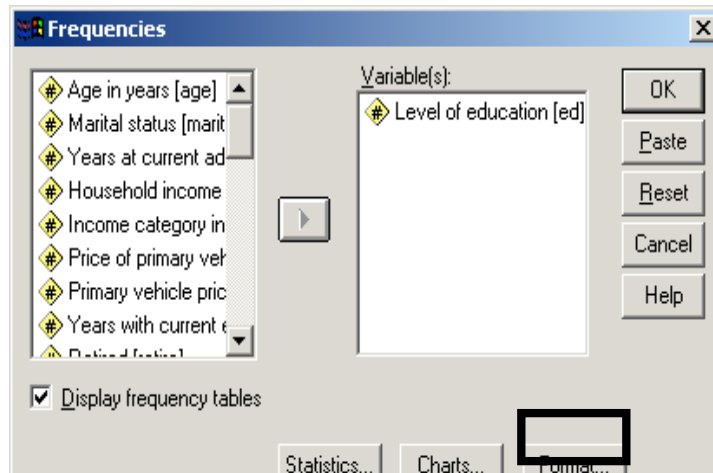


Figura 51 – Janela de Frequências – opção Format.

Estatísticas Descritivas

Através do procedimento *Analyze/Descriptive Statistics/Frequencies* é também possível calcular várias estatísticas de tendência central, de dispersão e de forma da distribuição.

Para o cálculo destas estatísticas, fazer:

1. No menu *Analyze* escolher *Descriptive Statistics > Frequencies*.

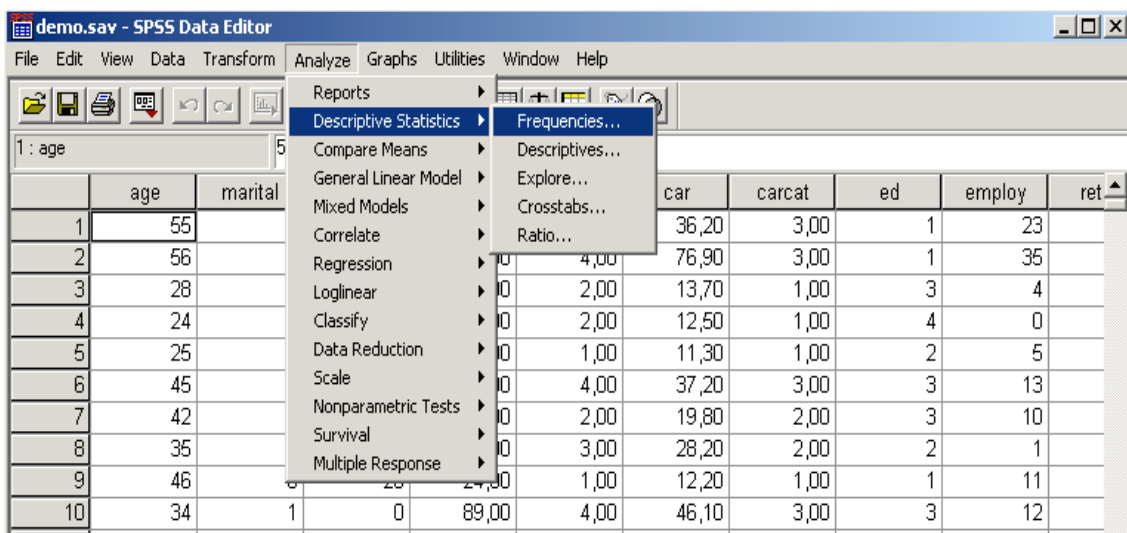


Figura 52 - Menu *Analyze > Descriptive Statistics > Frequencies*

2. Na janela que surge, seleccionar as variáveis pretendidas e seleccionar a opção *Statistics*.

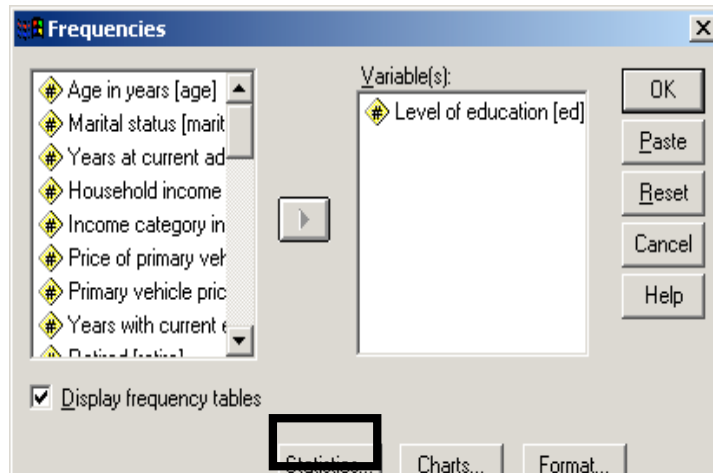


Figura 53 – Janela de Frequências

3. Surge então uma nova janela, onde se poderão escolher as estatísticas pretendidas.

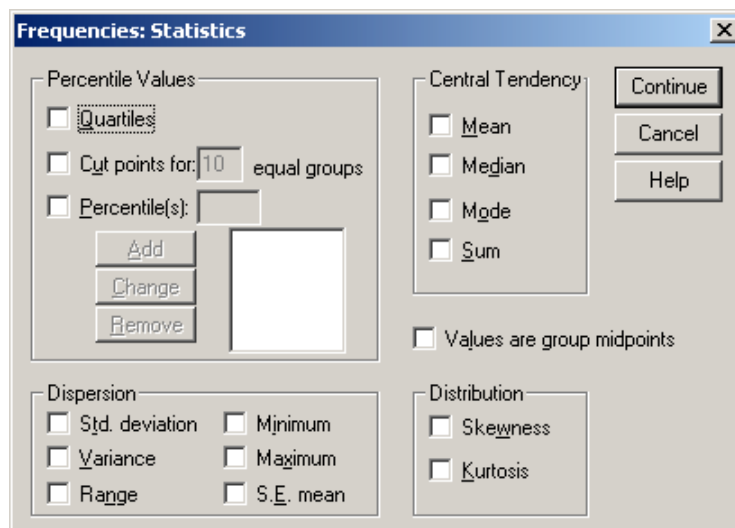


Figura 54 – Frequencies: Statistics

4. Quando estiverem seleccionadas todas as estatísticas pretendidas, deve-se clicar em *Continue*.
5. Na janela *Frequencies*, então visível, remover a selecção da opção *Display frequency table*, para não ser apresentada a tabela de frequências.
6. Clicar em *OK*.

Na janela de saída de dados irá surgir uma tabela semelhante a:

Statistics		
Level of education		
N	Valid	6400
	Missing	0
Mean		2,59
Std. Error of Mean		,015
Median		2,00
Mode		2
Std. Deviation		1,199
Variance		1,437
Skewness		,274
Std. Error of Skewness		,031
Kurtosis		-,982
Std. Error of Kurtosis		,061
Range		4
Minimum		1
Maximum		5
Sum		16557
Percentiles	25	2,00
	50	2,00
	--	---

Figura 55 – Estatísticas descritivas

As várias estatísticas disponíveis através deste procedimento são descritas de seguida:

- *quartiles* (Quartis) – calcula o percentil 25, percentil 50 (mediana) e o percentil 75.
- *Cut Points for n equal groups* (Pontos de corte para n grupos iguais) – apresenta os valores da variável que divide a amostra em n grupos com o mesmo número de casos. Depois de seleccionar esta estatística o utilizador deverá inserir o número (entre 2 e 100) de grupos pretendido. Por omissão, a distribuição é dividida em 10 grupos iguais.
- *Percentiles* (Percentis) – calcula os percentis inseridos pelo utilizador. Para inserir um percentil o utilizador deverá seleccionar a caixa de selecção associada à estatística, inserir o valor do percentil desejado e fazer **Add**.
- *Std. Deviation* (Desvio padrão) – Apresenta uma medida da dispersão em redor da média.
- *Variance* (Variância) – Calcula a soma do quadrado dos desvios em relação à média dividida pelo número de casos menos 1.
- *Range* (Intervalo absoluto) – Calcula a diferença entre o valor mais alto e o valor mais baixo.

- *Minimum* (Mínimo) – Apresenta o menor valor.
- *Maximum* (Máximo) – Apresenta o maior valor.
- *S.E. mean* (Erro padrão médio) – Apresenta uma estimativa da variabilidade amostral em relação à média. O seu cálculo obtém-se através do quociente entre o desvio padrão amostral e a raiz quadrada da dimensão da amostra.
- *Mean* (Média aritmética) – Devolve a soma de todos os valores dividido pelo número total de casos.
- *Median* (Mediana) – Calcula o valor acima e abaixo do qual se encontram metade dos valores da variável.
- *Mode* (Moda) – Calcula o valor que ocorre mais frequentemente na distribuição.
- *Sum* (Somatório) – calcula a soma de todos os valores que a variável toma.
- *Skewness* (Coeficiente de Assimetria) – Apresenta uma medida relativa de assimetria. Um sinal positivo indica um pico com uma cauda direita alongada e um sinal negativo uma cauda esquerda alongada. Se a distribuição for simétrica, esta medida tem o valor 0.
- *Kurtosis* (Medida de curtose) – apresenta uma medida que indica se a curva é achatada (valores negativos) ou alongada (valores positivos). Uma distribuição normal, tem uma medida de curtose igual a 0.

Para mais esclarecimentos acerca de cada medida estatísticas, deverá consultar a ajuda, clicando com o botão direito sobre o nome da medida estatística em questão.

Gráficos

Como já foi referido, o procedimento *Analyze/Descriptive Statistics/Frequencies* também permite visualizar o comportamento da(s) variável(is) graficamente, através de gráficos de barras, gráficos circulares e histogramas.

Para criar um gráfico utilizando este procedimento, fazer:

1. No menu *Analyze* escolher *Descriptive Statistics > Frequencies*.

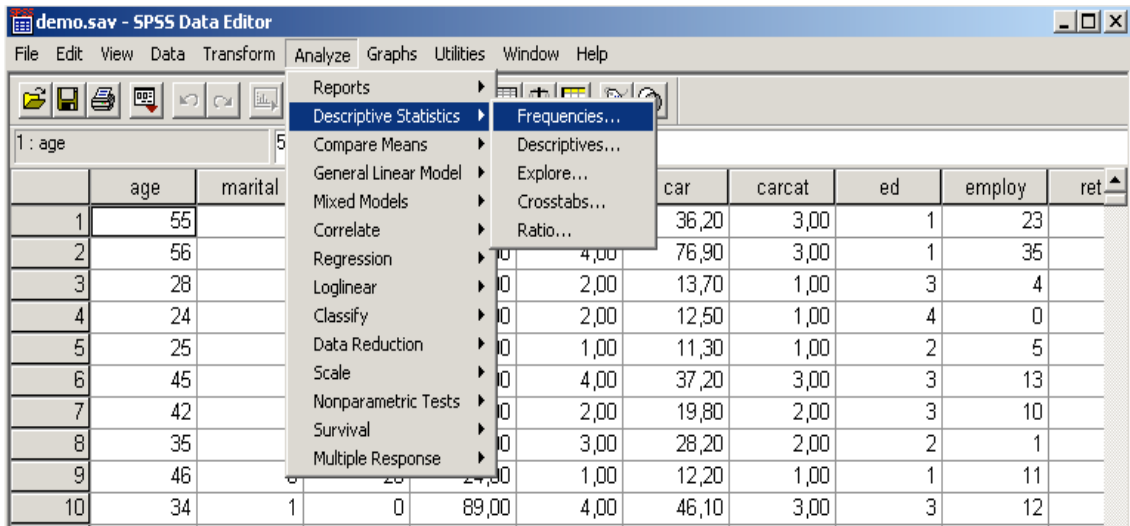


Figura 56 - Menu Analyze > Descriptive Statistics > Frequencies

- Na janela que surge, seleccionar as variáveis pretendidas e seleccionar a opção **Charts**.

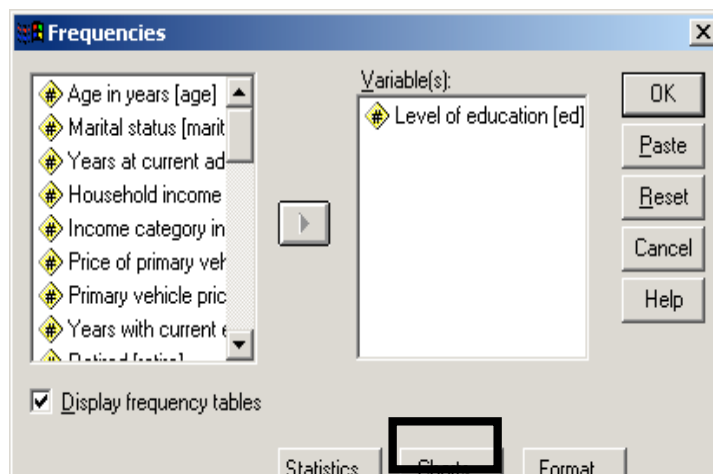


Figura 57 – Janela de Frequências

- Na janela que surge, deve seleccionar-se o tipo de gráfico pretendido. Nos gráficos de barras e nos gráficos circulares, existe a opção de elaborar o gráfico com base nas frequências absolutas (**Frequências**) ou com base nas frequências relativas (**Percentages**).

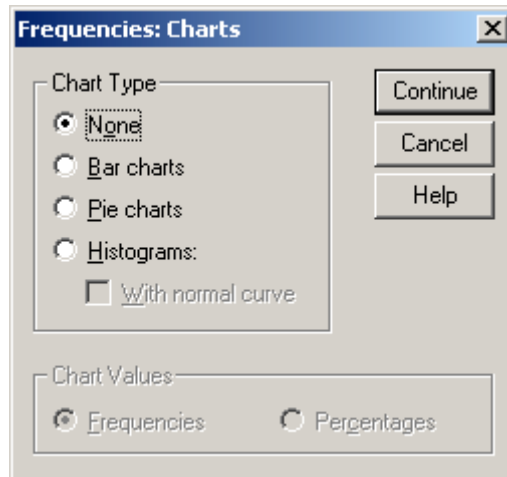


Figura 58 – Frequencies: Charts

4. Depois de seleccionado o gráfico pretendido, clicar em **Continue**.
5. Na janela **Frequencies**, então visível, remover a selecção da opção **Display frequency table**, para não ser apresentada a tabela de frequências.
6. Clicar em **OK**.

De seguida apresentam-se exemplos dos vários tipos de gráficos.

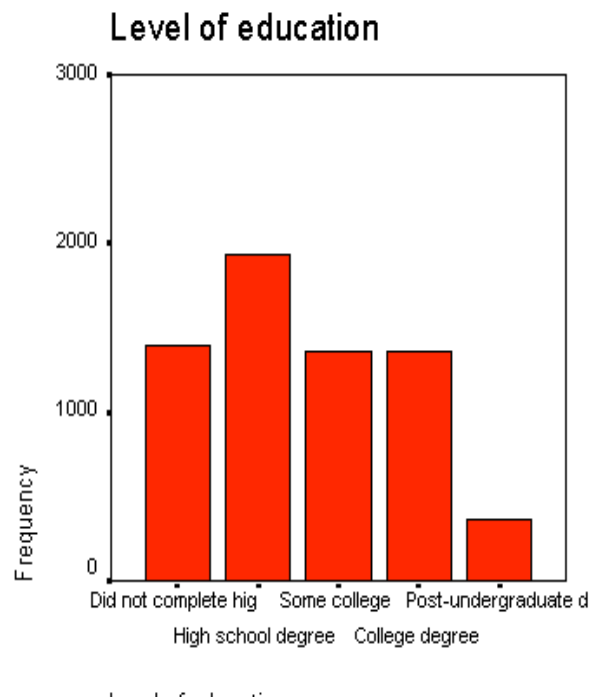


Figura 59 – Gráfico de barras por frequência absoluta

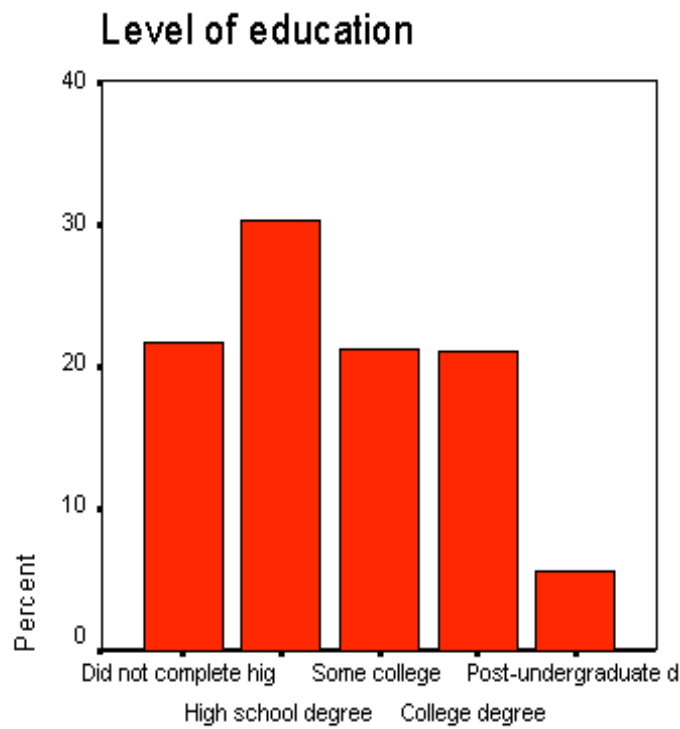


Figura 60 – Gráfico de barras por frequência relativa

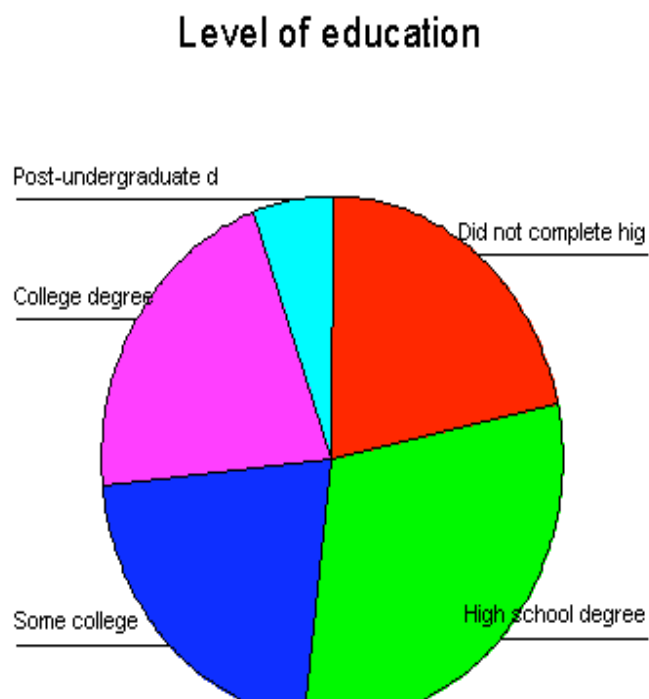


Figura 61 - Gráfico circular

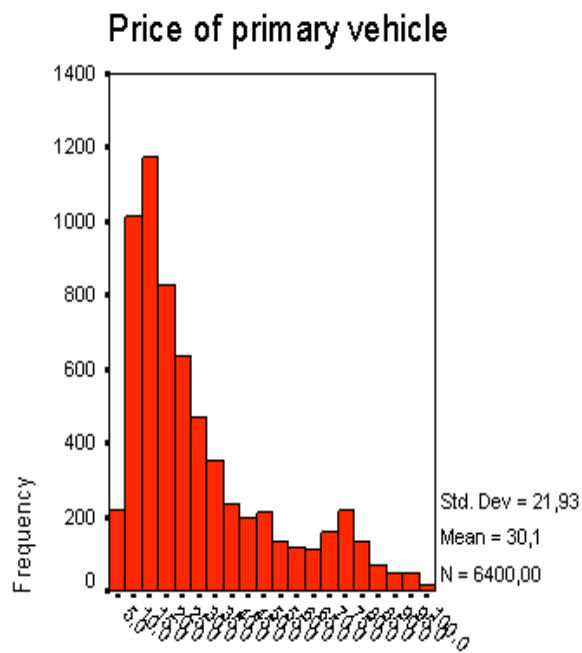


Figura 62 – Histograma

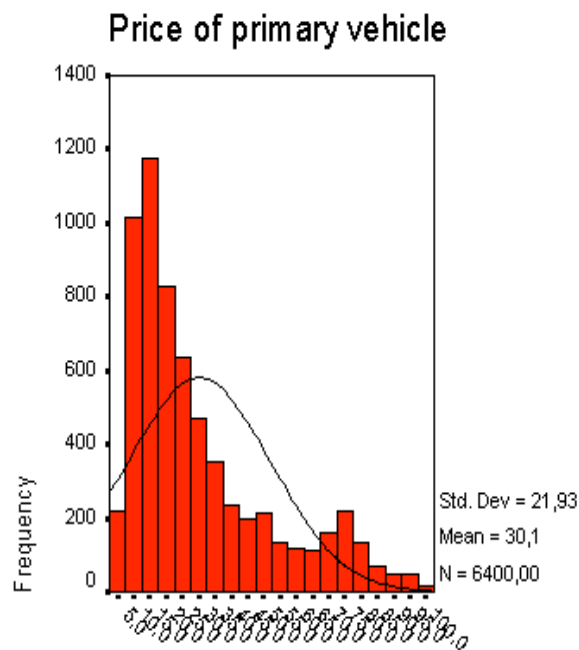


Figura 63 – Histograma com curva normal

Para edição dos gráficos gerados pelo SPSS, deve-se fazer um duplo-clique sobre o gráfico, que provocará a abertura da janela de edição de gráficos do SPSS:

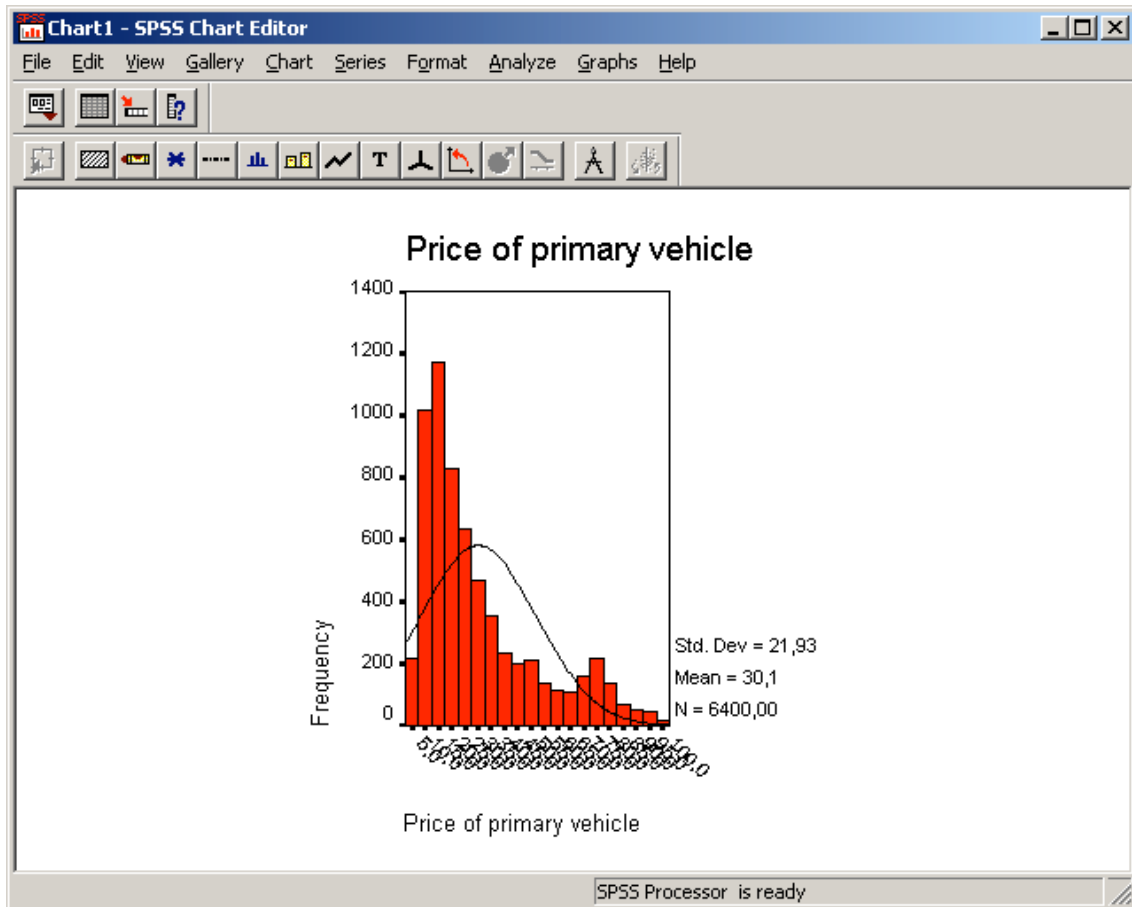


Figura 64 – Janela de edição de Gráficos

3.3. Cálculo de estatísticas básicas

O procedimento *Analyze/Descriptive Statistics/Descriptives* permite, tal como o procedimento *Analyze/Descriptive Statistics/Frequencies*, obter algumas estatísticas descritivas. As variáveis podem ser ordenadas pelo valor da sua média (ordem ascendente ou descendente), alfabeticamente ou pela ordem pela qual se selecciona as variáveis no procedimento (opção por omissão).

Para além destas estatísticas, permite também calcular valores padrão (scores z). Quando os valores padrão são guardados, são adicionados à janela de edição de dados e ficam disponíveis para qualquer procedimento estatístico.

Este procedimento é útil para obter comparações de variáveis quantitativas (intervalares ou racionais) e para identificar casos “anormais” entre as variáveis através do cálculo dos valores padrão.

Para executar este procedimento:

1. No menu *Analyze* escolher *Descriptive Statistics / Descriptives*

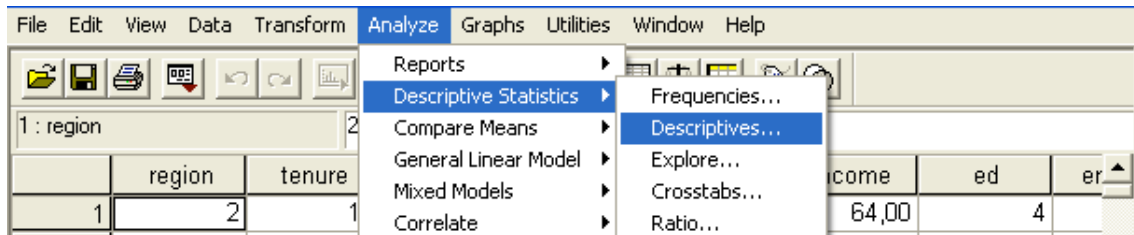


Figura 65 – Analyze / Descriptive Statistics / Descriptives

2. Na janela que surge, deverá seleccionar as variáveis que pretende analisar. Se pretender guardar os valores padrão das variáveis, deverá seleccionar a caixa *Save standardized values as variables*.

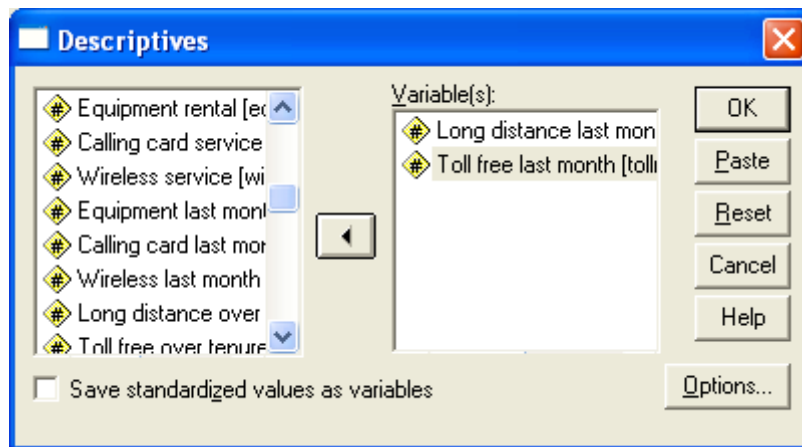


Figura 66 – Janela Descriptives

3. Para seleccionar as estatísticas clique em *options*. Na nova janela, poderá também escolher a ordem pela qual pretende que sejam apresentados os resultados.

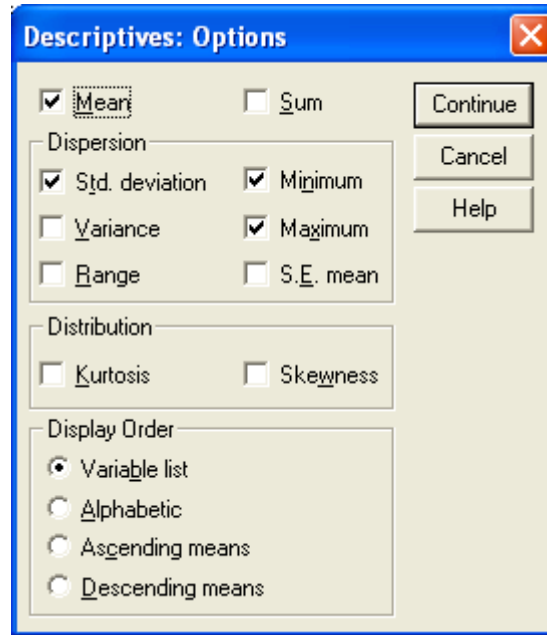


Figura 67 – Janela Descriptives / Options

4. Clique em *continue*. Na janela que surge clicar em *ok*. Na janela de saída aparecerá uma tabela com os valores de cada estatística seleccionada para cada variável.

Descriptive Statistics

	N	Range	Mean	Std. Deviation
Long distance last month	1000	99,05	11,7231	10,36349
Toll free last month	1000	173,00	13,2740	16,90212
Valid N (listwise)	1000			

Figura 68 – Resultado do procedimento Descriptives

5. Caso tenha seleccionado a opção *save standardized values as variables* serão criadas novas variáveis, que aparecerão nas últimas colunas da janela de edição de dados. Estas variáveis terão o nome das variáveis em análise, antecedido pela letra z. No caso apresentado, as variáveis em análise chamam-se *longmon* e *tollmon* e as variáveis criadas chamam-se *zlongmon* e *ztollmon*.

zlongmon	ztollmon
-,77417	Zscore: Toll free -,18339
-,70663	,44231
,62015	,27961
-,21934	-,78535
-,52329	-,78535
,00742	,35357
-,07942	-,78535
-,54741	1,87704
10000	00000

Figura 69 – Scores z

3.4. Análise exploratória

O procedimento *Explore* produz estatísticas sumárias e representações gráficas para todos os casos ou para grupos de casos separadamente. Através deste procedimento é possível:

- Examinar os dados e sua distribuição (detectar valores “anormais”, intervalos de dados vazios, outliers, variabilidades inesperadas e outras peculiaridades)
- Verificar o cumprimento de condições para a aplicação de procedimentos estatísticos (p.ex. para determinar se será necessário aplicar testes não paramétricos ou para determinar se será necessário transformar os dados para técnicas que requeiram distribuições normais)
- Caracterizar as diferenças existentes entre diferentes grupos de casos.

Para executar o procedimento deverá:

1. Escolher a opção *Analyze/Descriptive Statistics/Explore*.

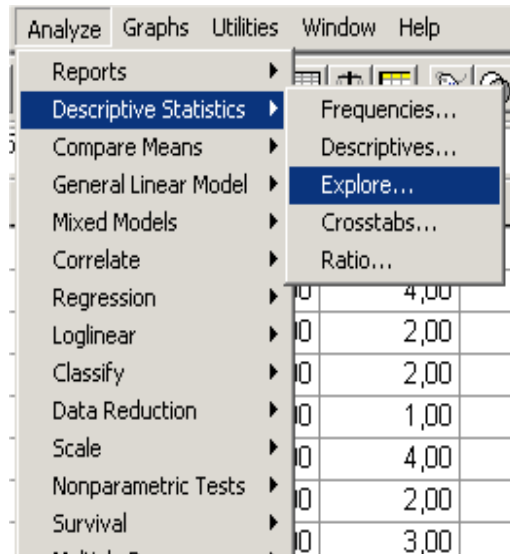


Figura 70 – Menu Explore

2. Na janela que surge, deverá escolher as variáveis a analisar e coloca-las na caixa *Dependent List*. As variáveis a analisar têm de ser numéricas. Para produzir análises separadas por grupos de casos, deverá introduzir na caixa *Factor List* as variáveis segundo as quais pretende fazer o agrupamento. Se introduzir mais do que uma variável na caixa *Factor List*, serão apresentados resultados separados para cada variável. Ainda nesta janela poderá escolher o tipo de representação que pretende: estatísticas, gráficos ou ambos.

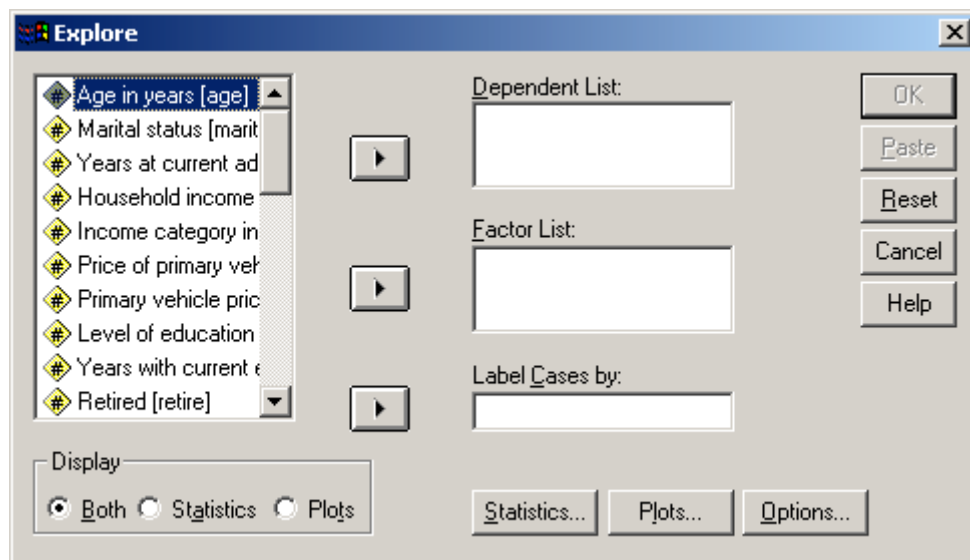


Figura 71 – Janela Explore

Estatísticas

Através do procedimento *Explore* é possível criar diversas estatísticas. Para tal, deverá, a partir da janela *Explore*:

1. Seleccionar as variáveis, seleccionar a opção *Both* ou *Statistics* em *Display* e clicar no botão *Statistics*.

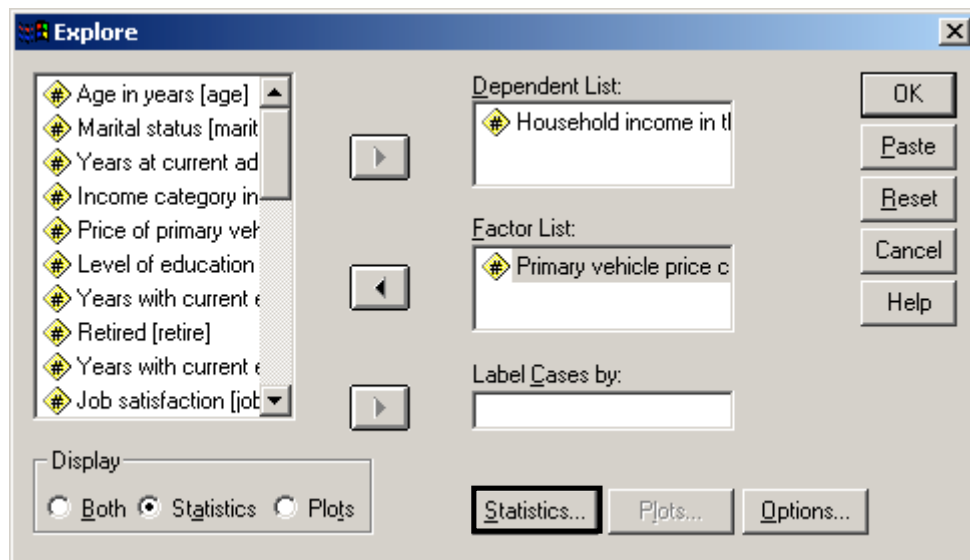


Figura 72 – Botão *Statistics* na janela *Explore*

2. Na janela que surge, poderá escolher uma das seguintes opções:
 - a. *Descriptives* – para calcular as seguintes estatísticas: média, moda, trimédia a 5% (média calculada eliminando 5% das observações mais baixas e 5% das observações mais altas), erro padrão, variância, desvio padrão, mínimo, máximo, intervalo absoluto, intervalo interquartil, coeficiente de assimetria, erro padrão de assimetria, coeficiente de curtose, erro padrão de curtose.
 - b. *M-estimators* – para calcular 4 alternativas de medidas de tendência central (Huber's M-estimator, Andrews' wave estimator, Hampel's redescending M-estimator e Tukey's biweight estimator). Estas quatro medidas utilizam critérios distintos de ponderação dos casos, cujo peso vai diminuindo à medida que os casos se afastam da média. Quando os dados provêm de uma distribuição com longas caudas, ou quando existem outliers, estas medidas são mais precisas do que a média ou mediana.
 - c. *outliers* – apresenta os casos com os 5 maiores valores e os casos com os 5 menores valores.

d. *Percentiles* – apresenta os seguintes percentis: 10, 25, 50, 75, 90, 95.

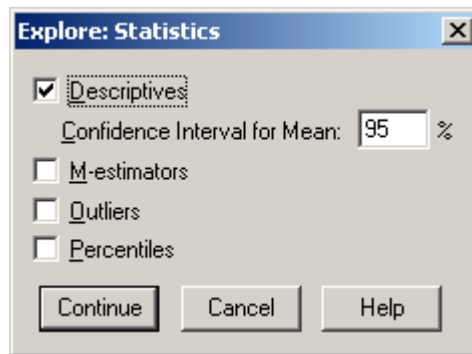


Figura 73 – Explore: Statistics

e. Depois de seleccionadas as estatísticas pretendidas, deverá fazer Continue e depois *OK* na janela *Explore*. Serão então apresentados os resultados na janela de saída do SPSS.

Descriptives

Primary vehicle			Statistic	Std. Error	
Household income in thousands	Economy	Mean	21,8876	,12215	
		95% Confidence Interval for Mean	Lower Bound	21,6480	
			Upper Bound	22,1271	
		5% Trimmed Mean	22,1027		
		Median	22,0000		
		Variance	27,471		
		Std. Deviation	5,24123		
		Minimum	9,00		
		Maximum	31,00		
		Range	22,00		
		Interquartile Range	8,0000		
		Skewness	-,479	,057	

Figura 74 – Estatísticas Descriptives do procedimento Explore para um grupo da categoria Primary Vehicle

M-Estimators

	Primary vehicle price category	Huber's M-Estimator ^a	Tukey's Biweight ^b	Hampel's M-Estimator ^c	Andrews' Wave ^d
Household income in thousands	Economy	22,2956	22,3468	22,2326	22,3435
	Standard	41,9249	42,0311	42,1576	42,0336
	Luxury	106,1755	98,3907	104,4940	98,3158

- a. The weighting constant is 1,339.
- b. The weighting constant is 4,685.
- c. The weighting constants are 1,700, 3,400, and 8,500
- d. The weighting constant is $1.340 \cdot n_i$

Figura 75 – M-Estimators

Extreme Values

Primary vehicle			Case Number	Value	
Household income in thousands	Economy	Highest	1	1929	31,00
			2	1963	31,00
			3	1984	31,00
			4	2006	31,00
			5	2008	31,00 ^a
	Standard	Lowest	1	25	9,00
			2	24	9,00
			3	23	9,00
			4	22	9,00
			5	21	9,00 ^b

Figura 76 - Outliers

Percentiles									
		Primary vehicle price category	Percentiles						
			5	10	25	50	75	90	95
Weighted Average(Definition 1)	Household income in thousands	Economy	12,0000	14,0000	18,0000	22,0000	26,0000	28,0000	29,0000
		Standard	31,0000	32,0000	35,0000	42,0000	49,0000	55,0000	58,0000
		Luxury	62,0000	65,0000	75,0000	101,0000	152,0000	244,0000	319,7500
Tukey's Hinges	Household income in thousands	Economy			18,0000	22,0000	26,0000		
		Standard			35,0000	42,0000	49,0000		
		Luxury			75,0000	101,0000	152,0000		

Figura 77 – Percentiles

Diagrama de caixa e fio (Boxplots)

Os diagramas de caixa e fio ou *boxplots* são uma representação gráfica útil para explorar e resumir os dados, permitindo analisar as frequências e identificar outliers, que tendem a distorcer a média e o desvio padrão. Dá-nos também informação importante sobre o centro, dispersão e assimetria dos dados.

Como se pode observar no diagrama abaixo, a base inferior do rectângulo (caixa) é o 1º quartil (percentil 25) e a base superior corresponde ao 3º quartil (percentil 75), ou seja, a caixa corresponde a 50% das observações totais. A mediana (2º quartil ou percentil 50) é representada pela linha dentro da caixa.

O comprimento máximo dos bigodes vai até 1,5 vezes o valor da amplitude interquartil, abaixo do 1º quartil e acima do 3º quartil. As observações que se situem para além destes limites são representadas com um círculo ou um asterisco, consoante sejam respectivamente outliers moderados ou severos. Os outliers são moderados quando se encontram entre 1,5 e 3 amplitudes interquartis abaixo do 1º quartil ou acima do 3º quartil. Os outliers são severos quando se encontram para além de 3 amplitudes interquartis para baixo do 1º quartil ou para cima do 3º quartil.

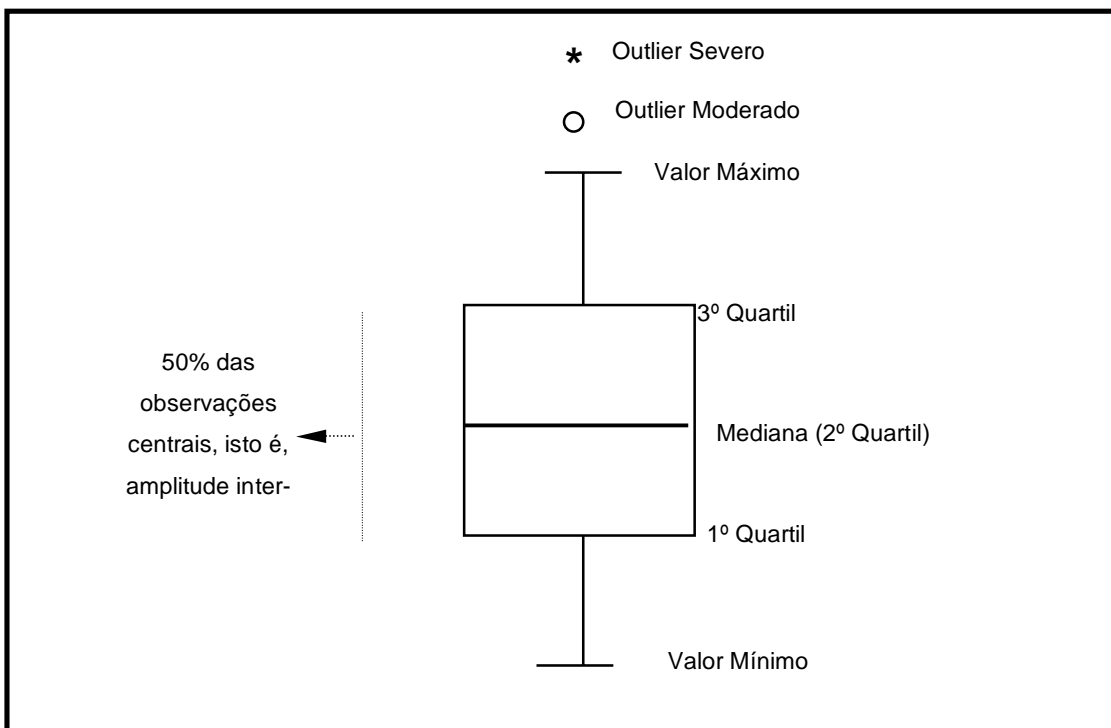


Figura 78 – Diagrama caixa e fio

Para criar um diagrama de caixa e fio no SPSS, a partir da janela Explore, deverá:

3. Seleccionar as variáveis, seleccionar a opção *Both* ou *Plots* em *Display* e clicar no botão *Plots*.

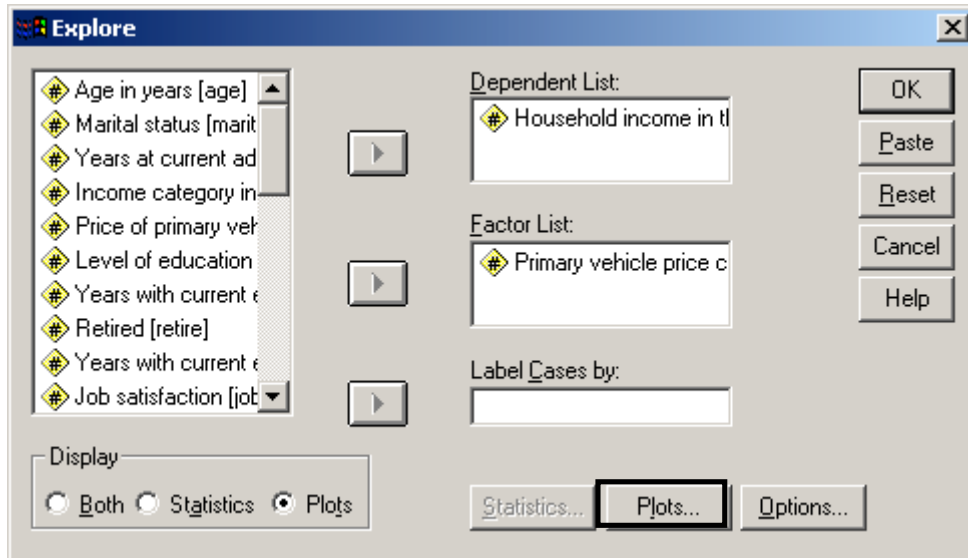


Figura 79 – Botão Plots na janela Explore

4. Na janela que surge, deverá escolher uma das seguintes opções:
 - a. *Factor levels together* – para cada variável dependente, os diagramas de caixa e fio de cada grupo são apresentados lado a lado. Desta forma, facilmente se compara como os valores da variável dependente varia nos vários grupos. Se não for introduzida nenhuma variável na caixa *Factor List* será apresentado apenas um único diagrama de caixa e fio para a totalidade dos casos.
 - b. *Dependents together* – para cada grupo, são apresentados os diagramas de caixa e fio de cada variável dependente lado a lado. Desta forma, facilmente se compara os valores das variáveis dependentes num grupo específico. Esta opção é útil quando as diferentes variáveis dependentes representam uma única característica medida em diferentes instantes temporais.

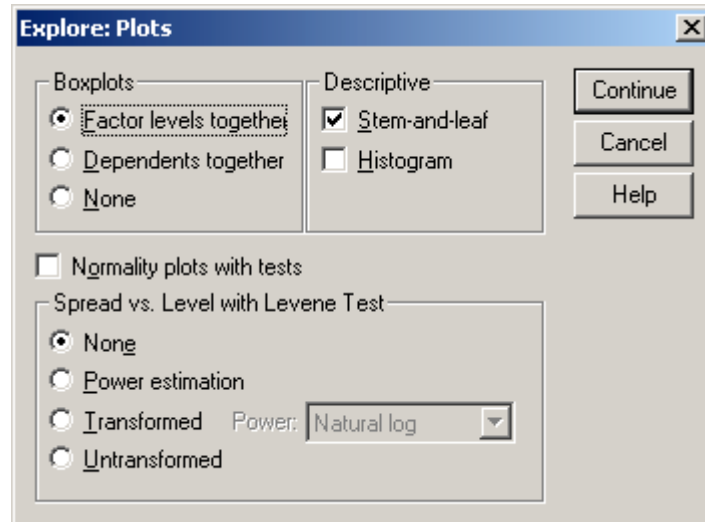


Figura 80 – Plots em Explore

5. Se seleccionar a opção *None*, não será criado nenhum diagrama de caixa e fio. Caso não pretenda a criação de um diagrama de tronco e folha, deverá remover a selecção da opção *Stem-and-Leaf*. Depois deverá clicar em *continue* e na janela *Explore* em *OK*.
6. Na janela de saída, será apresentado o diagrama de caixa e fio. O diagrama abaixo apresenta a relação entre o rendimento do agregado familiar e a categoria do principal veículo do agregado. Os outliers moderados são apresentados por um círculo e os outliers severos por um asterisco.

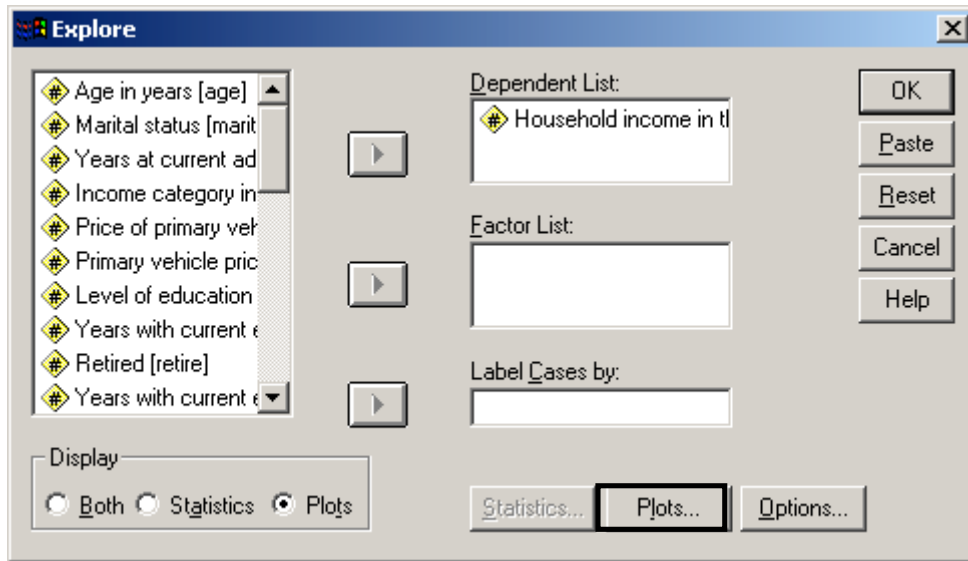


Figura 82 – Botão Plots na janela Explore

8. Na janela que surge, deverá seleccionar a opção *stem-and-leaf*. Caso não pretenda a criação de um diagrama caixa e fio, deverá seleccionar a opção *None*, em *Boxplots*. Se também pretender a criação de um histograma, deverá seleccionar a opção *Histogram*. Deverá então fazer *continue* e depois *ok* na janela *Explore*.

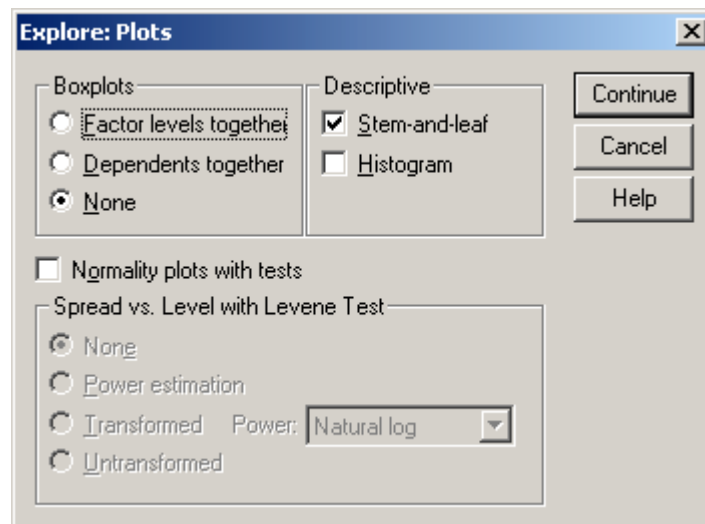


Figura 83 – Plots em Explore

9. Na janela de saída, será apresentado o diagrama de tronco e folha. O diagrama abaixo é relativo à variável que contém o rendimento do agregado familiar.

Household income in thousands Stem-and-Leaf Plot

Frequency	Stem &	Leaf
25,00	0 .	9
556,00	1 .	012345566778889999
1199,00	2 .	0000111122223333444455556666777788889999
1036,00	3 .	0000111122223333444455556666777888999
746,00	4 .	000112233445556667788999
529,00	5 .	0001123344556778899
436,00	6 .	00112234556789
278,00	7 .	0123456789
222,00	8 .	012345678&
211,00	9 .	023456789&
170,00	10 .	02369&&
116,00	11 .	&&&&
122,00	12 .	8&&&
89,00	13 .	&&&
73,00	14 .	&&
40,00	15 .	&
552,00	Extremes	(>=156)

Stem width: 10,00
Each leaf: 30 case(s)

Gráfico 1 – Diagrama tronco e folha

À esquerda do diagrama apresenta-se a frequência associada a cada intervalo. Cada caso é representado dentro da linha com o valor observado. Para tal, divide-se o valor observado em duas componentes: dígito(s) inicial(is) (tronco) e dígitos finais (folha).

O diagrama variará em função das unidades em que está medida a variável e do intervalo de valores. Na parte inferior do diagrama é explicado o modo como deve ser interpretado cada valor. O *stem width* indica o valor pelo qual deveremos multiplicar os valores para obter a unidade real e o *Each Leaf* indica o número de casos que cada elemento na folha representa.

No caso apresentado, o tronco corresponde à parte inteira das dezenas de milhares do rendimento (a unidade inicial era milhares e o *stem width* é 10) do agregado familiar e a folha à parte decimal. Na 1ª linha do diagrama, podemos verificar que existem 25 casos com rendimento familiar igual a 0,9 dezenas de milhar. No 2º caso,

podemos verificar que existem 30 casos (*Each Leaf* = 30 casos), com rendimento de 1,0; 30 casos com rendimento 1,1; 30 casos com rendimento 1,2; etc.

3.5. Tabelas

As tabelas produzidas pelo SPSS são apresentadas como tabelas pivot na janela de saída. Este tipo de tabelas tem uma grande flexibilidade na estruturação e formatação das tabelas.

Modificação de estrutura

Para modificar a estrutura de uma tabela, deverá:

1. Activar a tabela, fazendo um duplo clique em cima da tabela.


Descriptives				Statistic	Std. Error
Household income in thousands	Primary vehicle				
	Economy	Mean		21,8876	,12215
		95% Confidence Interval for Mean	Lower Bound	21,6480	
			Upper Bound	22,1271	
		5% Trimmed Mean		22,1027	
		Median		22,0000	
		Variance		27,471	
		Std. Deviation		5,24123	
		Minimum		9,00	
		Maximum		31,00	
		Range		22,00	
	Interquartile Range		8,0000		
	Skewness		-,479	,057	
	Kurtosis		-,470	,114	
	Standard	Mean		42,5600	,17967
95% Confidence Interval for Mean		Lower Bound	42,2077		
		Upper Bound	42,9123		
5% Trimmed Mean			42,3220		
Median			42,0000		
Variance			73,439		
Std. Deviation			8,56966		
Minimum			29,00		
Maximum			61,00		
Range			32,00		
Interquartile Range		14,0000			
Skewness		,337	,051		
Kurtosis		-1,018	,103		
Luxury	Mean		134,6410	2,14172	
	95% Confidence Interval for Mean	Lower Bound	130,4411		
		Upper Bound	138,8409		
	5% Trimmed Mean		120,1614		
	Median		101,0000		
	Variance		10476,661		
	Std. Deviation		102,35556		
	Minimum		58,00		
	Maximum		1116,00		
	Range		1058,00		
Interquartile Range		77,0000			

Figura 84 – Tabela do SPSS

2. Clicar com o botão direito em cima da tabela e escolher a opção *Pivoting Trays*.

Descriptives						
Household income in thousands	Primary vehicle	Statistic	Std. Error			
Economy	Mean	21,8876	,12215			
	95% Confidence Interval for Mean	Lower Bound	21,6480			
		Upper Bound	22,1271			
			22,1027			
			22,0000			
			27,471			
			5,24123			
			9,00			
			31,00			
			22,00			
			8,0000			
			-4,479	,057		
			-4,470	,114		
	Luxury	Mean	134,6410	2,14172		
95% Confidence Interval for Mean		Lower Bound	130,4411			
		Upper Bound	138,8409			
			120,1614			
			101,0000			
			10476,661			
			102,35556			
			58,00			
			1116,00			
			1058,00			
			77,0000			
		Skewness	,337	,051		
		Kurtosis	-1,018	,103		

Figura 85- Menu de configuração de tabelas

Aparecerá então a seguinte janela, onde cada conjunto de informação é representada por . Para saber a que se refere cada um destes símbolos, basta situar o rato por cima do símbolo durante um breve instante. Depois poderá alterar a localização de cada bloco de informação entre as linhas (*Row*), colunas (*Column*) ou entre uma nova camada de informação (*Layer*). Para alterar a localização, basta arrastar o símbolo para a localização pretendida.

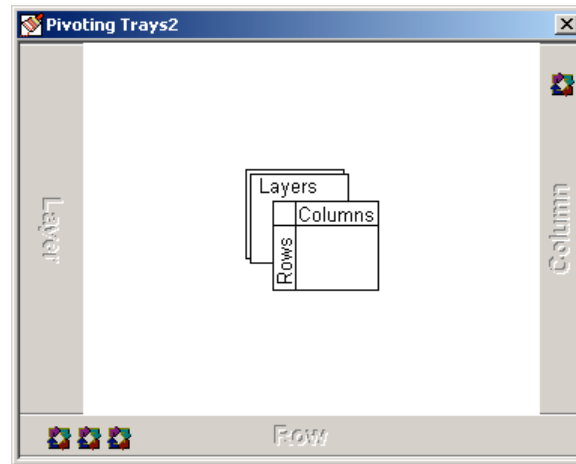


Figura 86 – Pivoting Trays

3. Se para o caso da tabela apresentada na figura 37, movimentar a variável *Primary vehicle* para uma nova camada, obtenho a tabela apresentada a figura 41, onde poderei escolher através de uma caixa de selecção qual o valor desta variável para o qual pretendo consultar as estatísticas.

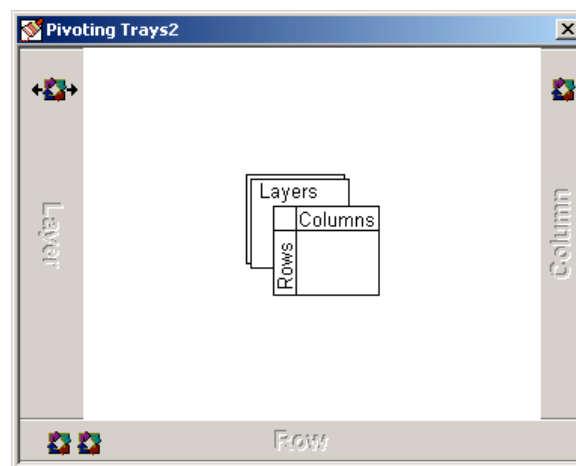


Figura 87 – Inserir informação numa nova camada

Descriptives					
Primary vehicle price category		Economy		Statistic	Std. Error
Household income in thousands	Mean	Standard		21,8876	,12215
	95% Confidence Interval for Mean	Luxury	Lower Bound	21,6480	
		Upper Bound		22,1271	
	5% Trimmed Mean			22,1027	
	Median			22,0000	
	Variance			27,471	
	Std. Deviation			5,24123	
	Minimum			9,00	
	Maximum			31,00	
	Range			22,00	
	Interquartile Range			8,0000	
	Skewness			-,479	,057

Figura 88 – Tabela com uma nova camada

- Se para o caso da tabela apresentada na figura 37, colocar a variável *Primary vehicle* em colunas, obtenho a tabela apresentada na figura 43.

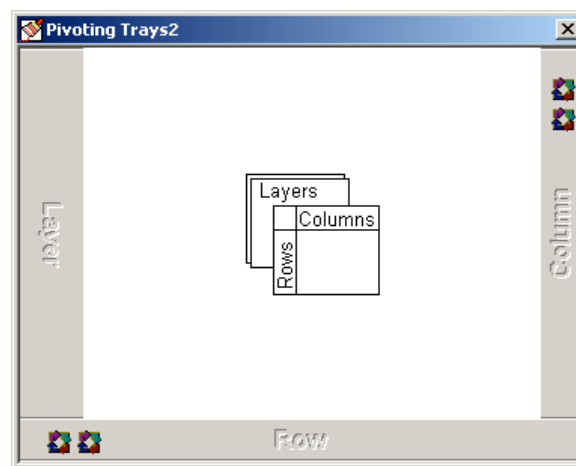


Figura 89 – Troca de linha para coluna

			Descriptives					
			Statistic			Std. Error		
			Primary vehicle price category			Primary vehicle price category		
			Economy	Standard	Luxury	Economy	Standard	Luxury
Household income in thousands	Mean		21,8876	42,5600	134,6410	,12215	,17967	2,14172
	95% Confidence Interval for Mean	Lower Bound	21,6480	42,2077	130,4411			
		Upper Bound	22,1271	42,9123	138,8409			
	5% Trimmed Mean		22,1027	42,3220	120,1614			
	Median		22,0000	42,0000	101,0000			
	Variance		27,471	73,439	10476,661			
	Std. Deviation		5,24123	8,56966	102,35556			
	Minimum		9,00	29,00	58,00			
	Maximum		31,00	61,00	1116,00			
	Range		22,00	32,00	1058,00			
	Interquartile Range		8,0000	14,0000	77,0000			
	Skewness		-,479	,337	3,694	,057	,051	,051
	Kurtosis		-,470	-1,018	21,142	,114	,103	,102

Figura 90 – Tabela com variável Primary vehicle em coluna

Como se pode verificar pelos exemplos apresentados, a estruturação de tabelas no SPSS tem bastante potencial e possibilidades.

Modificação de formatação

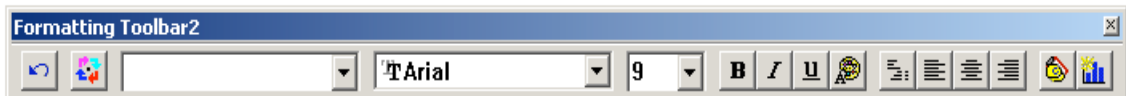
Para modificar a formatação associada a uma tabela, deverá:

1. Activar a tabela, fazendo um duplo clique em cima da tabela.
2. Clicar com o botão direito em cima da tabela e escolher a opção *Toolbar*.

			Descriptives					
			Statistic			Std. Error		
			Primary vehicle price category			Primary vehicle price category		
			Economy	Standard	Luxury	Economy	Standard	Luxury
Household income in thousands	Mean		21,8876	42,5600	134,6410	,12215	,17967	2,14172
	95% Confidence Interval for Mean	Lower Bound	21,6480	42,2077	130,4411			
		Upper Bound	22,1271	42,9123	138,8409			
	5% Trimmed Mean		22,1027	42,3220	120,1614			
	Median		22,0000	42,0000	101,0000			
	Variance		27,471	73,439	10476,661			
	Std. Deviation		5,24123	8,56966	102,35556			
	Minimum		9,00	29,00	58,00			
	Maximum		31,00	61,00	1116,00			
	Range		22,00	32,00	1058,00			
	Interquartile Range		8,0000	14,0000	77,0000			
	Skewness		-,479	,337	3,694	,057	,051	,051
	Kurtosis		-,470	-1,018	21,142	,114	,103	,102

Figura 91 – Opção Toolbar

3. Ficará visível a barra de formatação, com o que poderá formatar todo o aspecto visual da tabela. Para seleccionar a totalidade de uma coluna ou linha, deverá fazer Ctrl + Alt em cima do nome da coluna ou linha.



- Outra opção possibilidade de formatação é o acesso a modelos predefinidos. Para tal deverá clicar com o botão direito em cima da tabela e seleccionar a opção *tableLooks*.

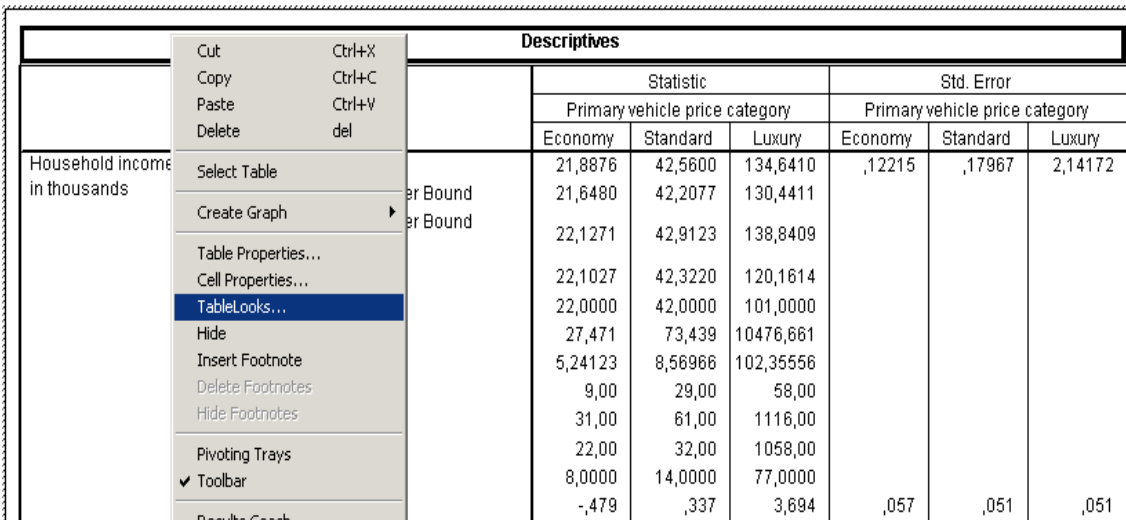


Figura 92 – Opção TableLooks

- Surgirá então uma janela, onde poderá escolher um dos modelos pré-definidos.

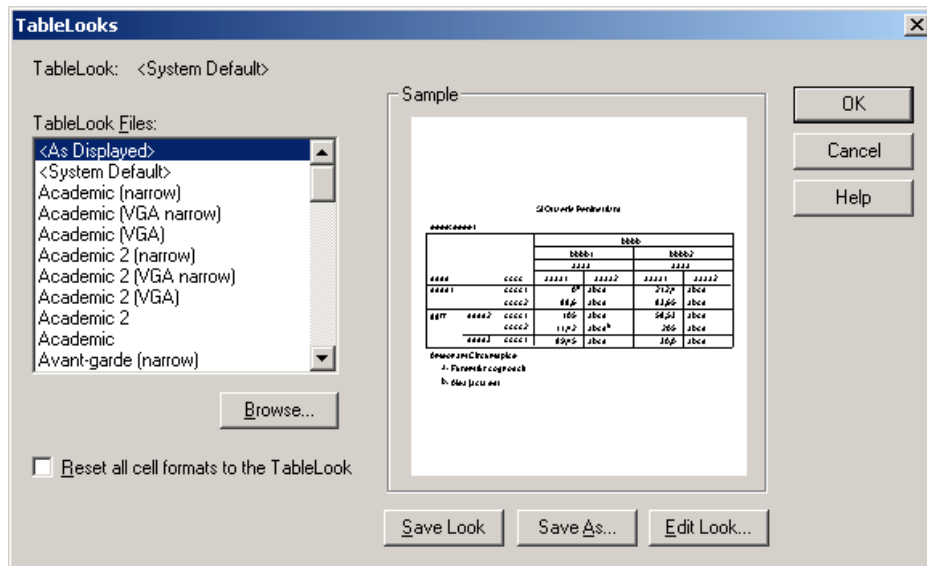


Figura 93 – Janela TableLooks

4. Testes de hipóteses paramétricos

Em análise estatística é necessário comparar parâmetros da população (média, variância, mediana, etc.) a partir de amostras aleatórias. Este tipo de procedimento inferencial é muito usado na avaliação dos efeitos de determinados tratamentos.

Este tipo de testes pode ser feito segundo duas metodologias:

- Testes Paramétricos – que exigem o conhecimento da distribuição amostral, geralmente a Normal;
- Testes Não Paramétricos – não exigem o conhecimento da distribuição amostral.

Convém destacar que a potência dos Testes Paramétricos é superior à dos Testes Não Paramétricos, isto é, a probabilidade de rejeitar correctamente H_0 é maior num Teste Paramétrico. Por este motivo, apenas se recorre aos Testes Não Paramétricos quando não é possível satisfazer as condições de aplicação dos Testes Paramétricos.

4.1. Condições de aplicação dos Testes Paramétricos

Para se poderem aplicar os Testes Paramétricos abordados neste capítulo é necessário verificar se a variável (dependente) segue uma distribuição normal.

Para testar a normalidade da variável em estudo, usa-se, normalmente, o Teste de Kolmogorov-Smirnov.

Teste de Kolmogorov-Smirnov

Neste teste pretende-se testar se a distribuição da variável é normal, com parâmetros μ e σ quaisquer:

$$H_0: X \sim N(\mu, \sigma) \quad \text{Vs} \quad H_1: X \sim N(\mu, \sigma)$$

Rejeita-se H_0 se o valor prova (p_value) for inferior ou igual ao nível de significância (α), em que o p_value é um valor produzido pelo SPSS usando a correcção de Lilliefors às tabelas com os valores críticos da distribuição de Kolmogorov-Smirnov.

Teste de Shapiro-Wilk

Este teste é apresentado, adicionalmente, pelo SPSS. É uma alternativa ao Teste de Kolmogorov-Smirnov no teste à normalidade da variável em estudo.

Teste de Normalidade

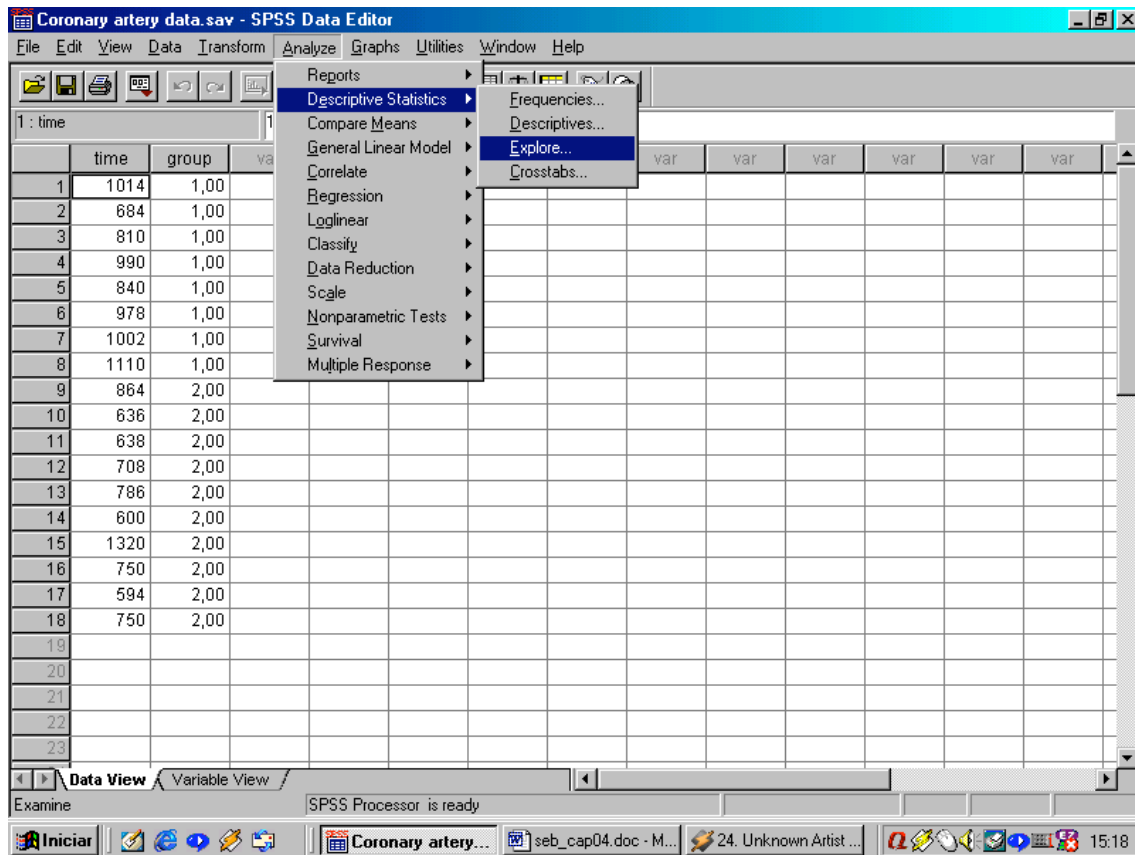


Figura 94: Menu a seguir para efectuar o teste à normalidade.

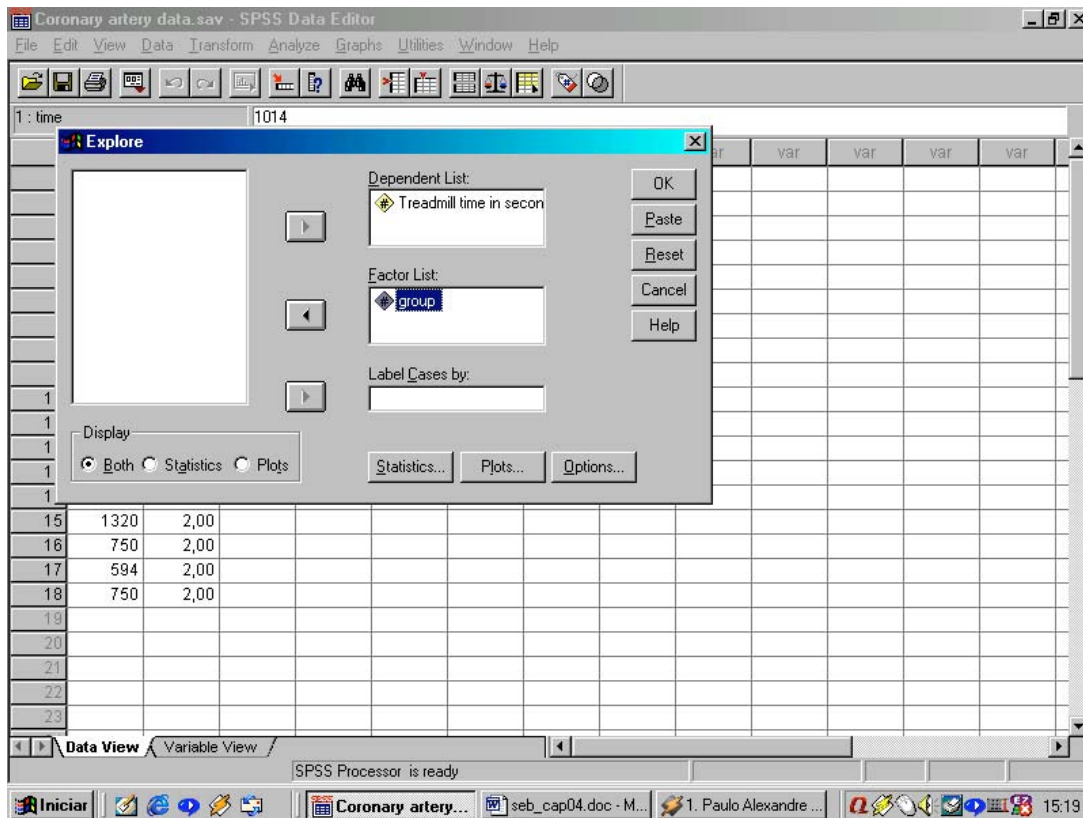


Figura 95: Janela de diálogo no teste à normalidade onde se colocam as variáveis a testar e as que vão criar grupos.

Seleccionar, como se mostra na figura seguinte,

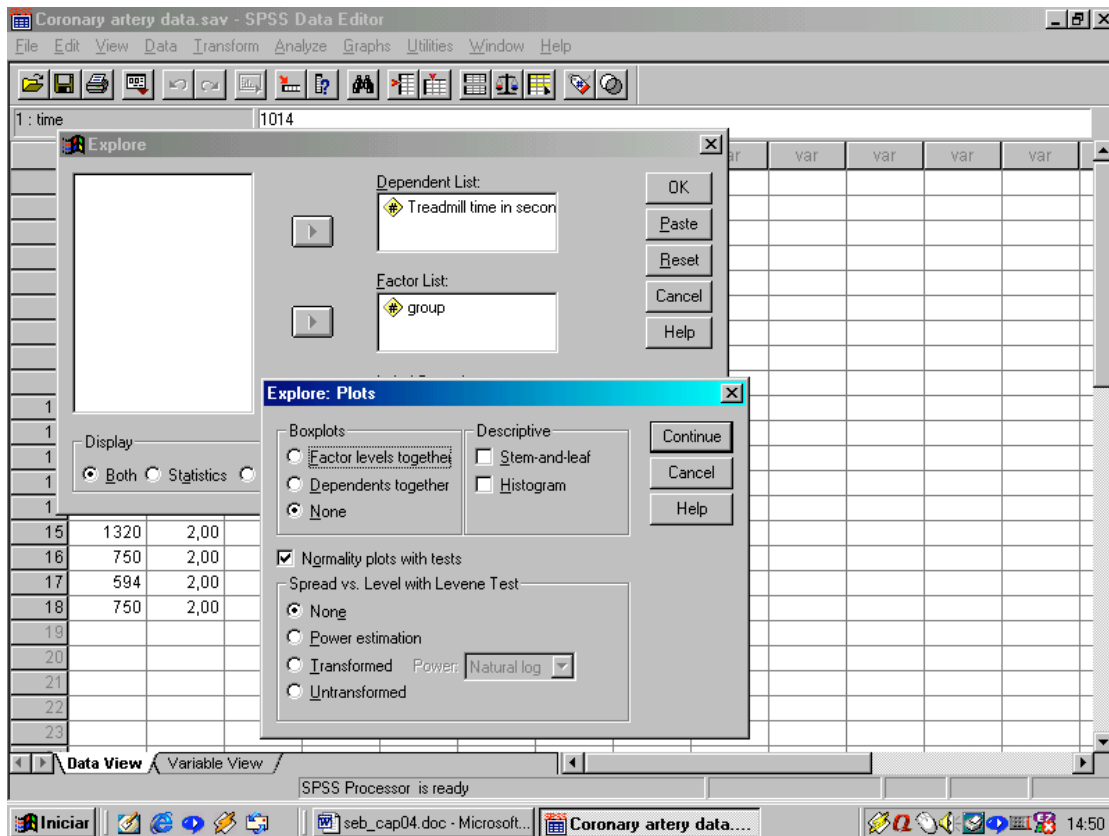


Figura 96: Janela onde se selecciona a opção para efectuar o teste à normalidade.

Fazer CONTINUE. Depois em DISPLAY seleccionar apenas PLOTS.

4.2. Resultados ao Teste de Normalidade

		b		d		Total	
		N	%	N	%	N	%
Valid	b	8	100%	0	0%	8	100%
Valid	d	0	0%	0	0%	0	0%

Figura 97: Tabela de resultado do teste à normalidade, que apresenta o número de casos válidos, missings e totais.

Neste primeiro quadro *output* é apresentado um pequeno resumo acerca da variável em estudo, para cada grupo.

y

		o		a		n	
		f _i		f _j		f _{ij}	
n	8	8	0	8	8	8	8
i	8	0	8	8	0	8	0

* 8
a 8

Figura 98: Tabela resultante do teste à normalidade, com os dois testes Kolmogorov-Smirnov e Shapiro-Wilk.

O segundo quadro apresenta o Teste de Kolmogorov-Smirnov com a correcção de Lilliefors para as duas amostras da variável *group*, com *p_value* de 0.103 e 0.054 respectivamente. Como o nível de significância é 0.05, não se rejeita a hipótese nula. Pode, por isso, aplicar-se o Teste de t- Student.

Adicionalmente é também apresentado o Teste de Shapiro-Wilk.

Como foi referido nas aulas teóricas a análise à normalidade será sempre efectuada a partir do Teste de Kolmogorov-Smirnov.

4.3. Médias

A média é uma Medida de Tendência Central muito usada na investigação, em áreas como as ciências biológicas.

Por este motivo, torna-se fundamental o estudo de métodos para analisar dados quantitativos (intervalares e racionais) cuja Medida de Tendência Central adequada seja a média.

No menu

>ANALYZE >COMPARE MEANS >MEANS

é possível calcular uma série de medidas estatísticas sumárias como média, desvio padrão, variância, etc., para um conjunto de variáveis em grupos, categorias ou variáveis distintas.

No ficheiro *Coronary artery data*,

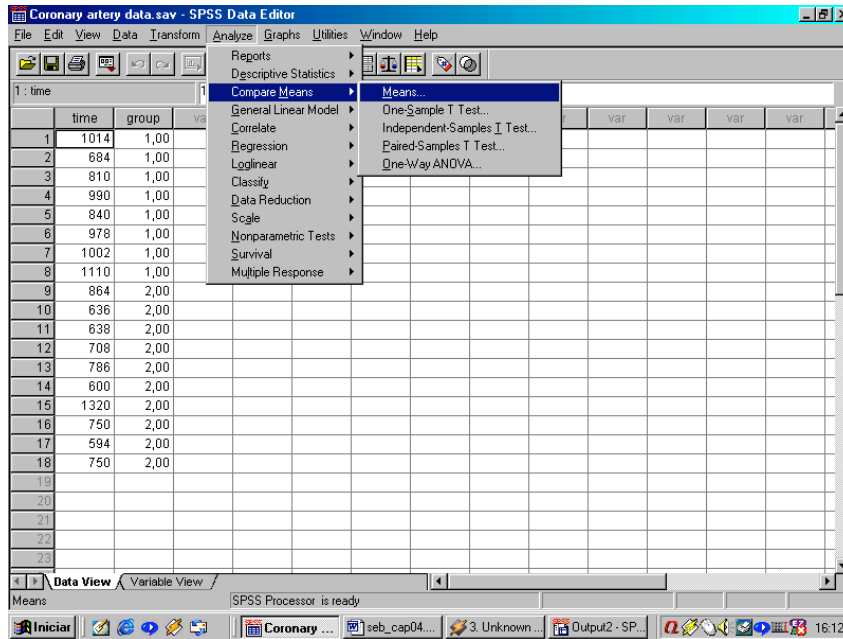


Figura 99: Menu para efectuar o procedimento *Compare Means*-> *Means*.

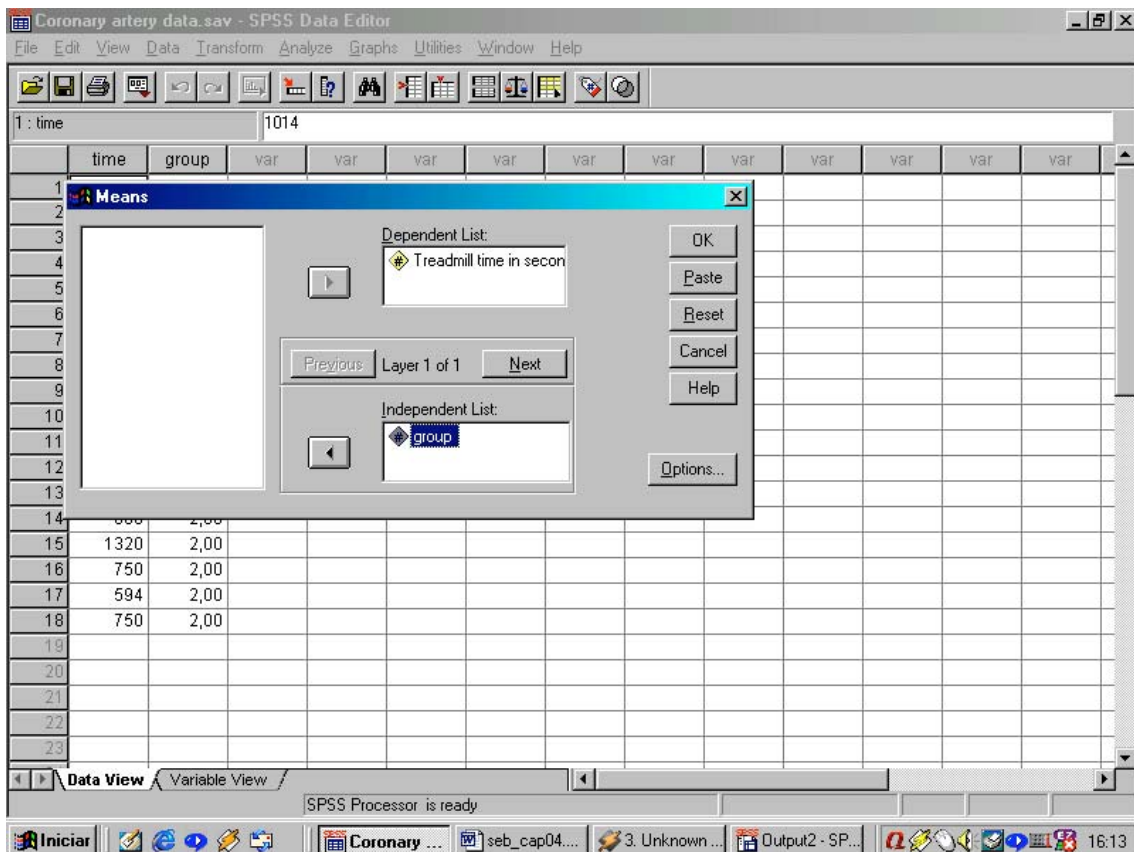


Figura 100: Janela de diálogo do procedimento *Compare Means*-> *Means* onde é visível onde se deverão colocar as variáveis dependentes e independentes.

Seleccionar *treadmil time in seconds* para a **Dependent List** e a *group* para a **Independent List**.

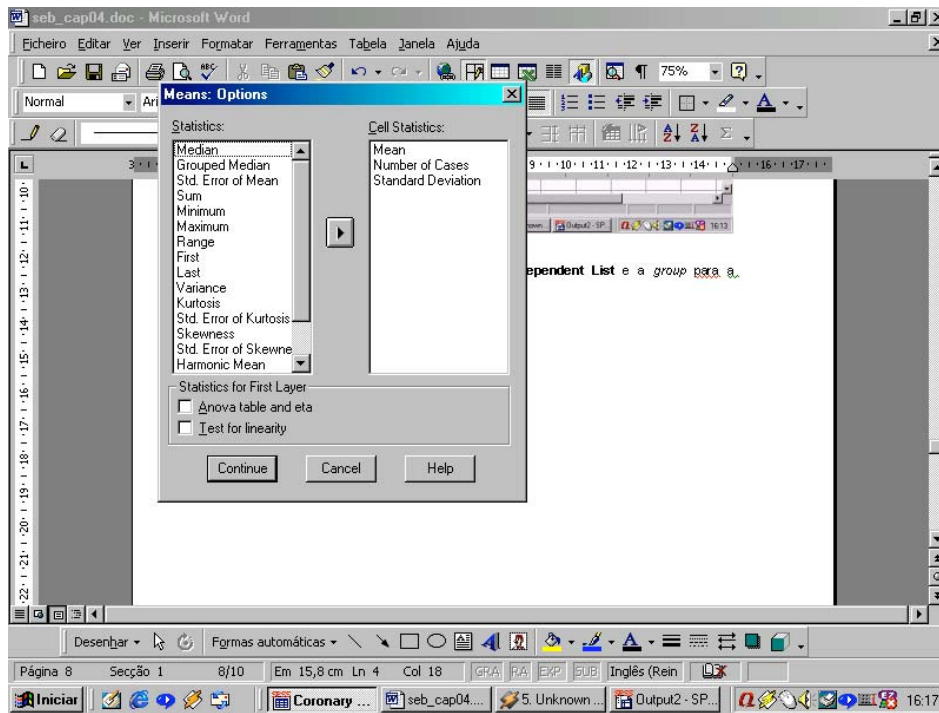


Figura 101: Janela de diálogo onde é possível seleccionar as estatísticas a pedir.

Na janela **OPTIONS** podem seleccionar-se as estatísticas que se pretendem calcular.

Como *output* obtêm-se dois quadros:

		e					
		h		h		h	
		N	P	N	P	N	P
e	h	8	%	0	0%	8	0%

Figura 102: Tabela resultante o procedimento **Compare Means-> Means** onde se mostram número de casos incluídos, excluídos e total.

O primeiro quadro apresenta o número de casos válidos, inválidos e total.

2

π		l	
θ	θ	N	θ
b	0	8	2
d	0	0	0
0	0	8	0

Figura 103: Tabela resultante do procedimento *Compare Means-> Means* onde se apresentam as estatísticas pedidas.

O segundo quadro diz respeito aos valores das estatísticas pretendidas de cada variável – *treadmill time in seconds* – em cada *group*.

4.4. Uma amostra

Quando se pretende verificar se a média de um determinado grupo de observações é igual a um valor padrão, usa-se um teste-t para uma amostra (**One sample t- test**).

Para aplicar este teste é necessário ter em conta a normalidade dos dados. Se a dimensão amostral for superior ou igual a 30, com base no Teorema do Limite Central, pode assumir-se que a distribuição das médias é normal, independentemente da distribuição da população. Se a dimensão da amostra for inferior a 30 é necessário efectuar um teste da normalidade (ver introdução ao Capítulo 4).

Abrir o ficheiro *saturaçãooex1*. Este ficheiro contém valores de um estudo efectuado em 31 homens e 29 mulheres saudáveis, sobre a saturação da bÍlis por colesterol.

Suponha-se que se realizou um estudo semelhante e que se obteve uma média de saturação da bÍlis de 87.3%.

De seguida, verifique-se se a estimativa da média da saturação da bÍlis do estudo realizado é significativamente diferente da média populacional do estudo anterior. Pretende-se testar:

$$H_0: \mu = 87.3 \quad \text{Vs} \quad H_1: \mu \neq 87.3$$

Fazer

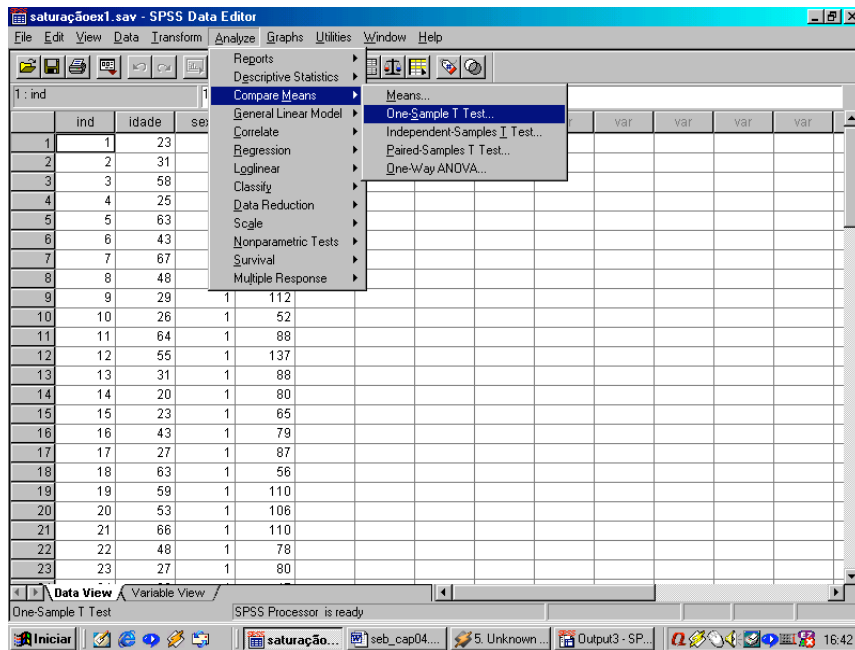


Figura 104: Menu para a execução do procedimento Teste t para uma amostra.

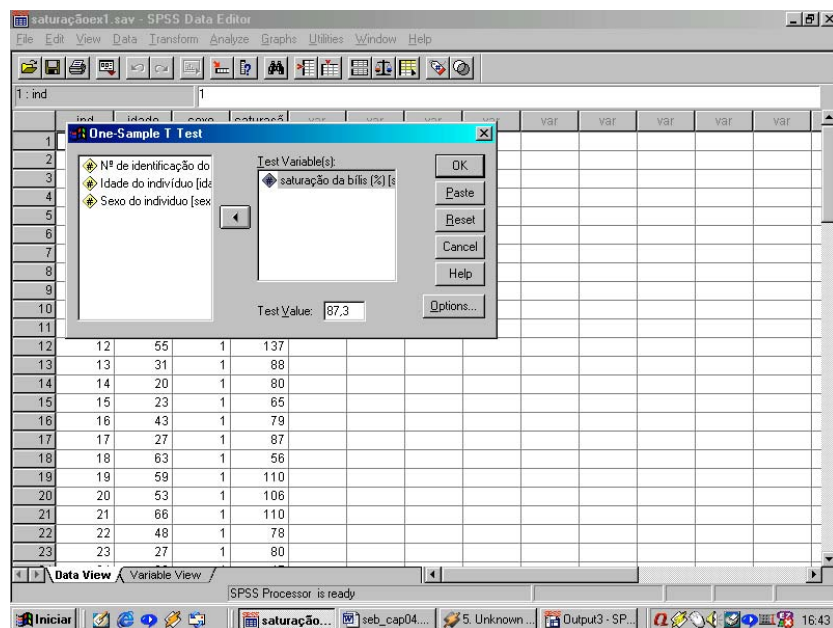


Figura 105: Janela de diálogo resultante da selecção do teste t para uma mostra onde é possível escolher qual a variável e o valor a testar.

Janela ONE-SAMPLE T TEST:

- ◆ Test Variable(s) – variáveis que se pretendem analisar;
- ◆ Test Value – valor numérico contra o qual se vai testar a média das variáveis seleccionadas.

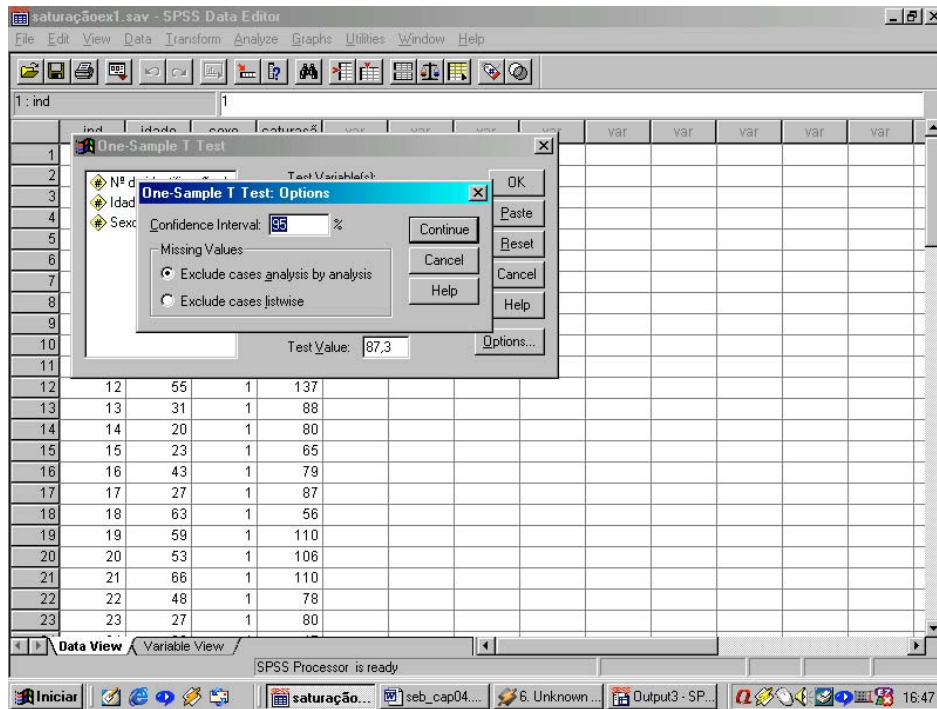


Figura 106: Janela de diálogo onde se pode indicar o nível de confiança para o intervalo.

Janela OPTIONS:

- ◆ Confidence interval – intervalo de confiança para a diferença das médias⁴;
- ◆ Missing values
 - Exclude cases analysis by analysis – são excluídos os casos com valores desconhecidos para a variável seleccionada;
 - Exclude cases listwise – são excluídos todos os valores desconhecidos da base de dados.

Mais uma vez surgem dois quadros.

⁴ Este intervalo de confiança é baseado na diferença da média da amostra e o valor de referência introduzido no teste. Note que é diferente do intervalo de confiança para a média da população apresentado na teórica.

	N	M		d
u	6		87.3	2

Figura 107: Tabela resultante do teste t para uma amostra onde se apresentam algumas estatísticas da variável a testar.

Este primeiro quadro apresenta o número de casos válidos (**N**), a média amostral (**Mean**), o desvio padrão (**Std. Deviation**) e o erro padrão da média (**Std. Error Mean**).

	7					
				M	D	
	t	df			u	l
u	9	9	87.3	87.3	87.3	87.3

Figura 108: Tabela resultante do teste t para uma amostra onde se apresenta entre outras informações o valor p, do respectivo teste.

No segundo quadro são apresentados o valor testado (**Test value**), o valor da estatística t, o número de graus de liberdade (**df**), o p_value (**Sig.(2-tailed)**), a diferença entre as médias (**Mean Difference**) e o intervalo de 95% de confiança para a diferença das médias.

Neste caso, como o p_value (=0.662) é superior ao nível de significância (=0.05), não se rejeita H₀, pelo que não se conclui, com uma confiança de 95%, que a média do estudo seja significativamente diferente de 87.3.

Testes-t simultâneos para a média de uma variável agrupada por outra variável

Considere-se o mesmo ficheiro, *saturaçãoex1*. Suponha que se pretende avaliar se a média do nível de saturação da bilis é ou não igual a 87.3% em cada sexo.

Para isso, era necessário começar por separar os dados de acordo com o sexo.

```
> DATA      > SPLIT FILE
> COMPARE GROUPS, BASED ON sexo
```

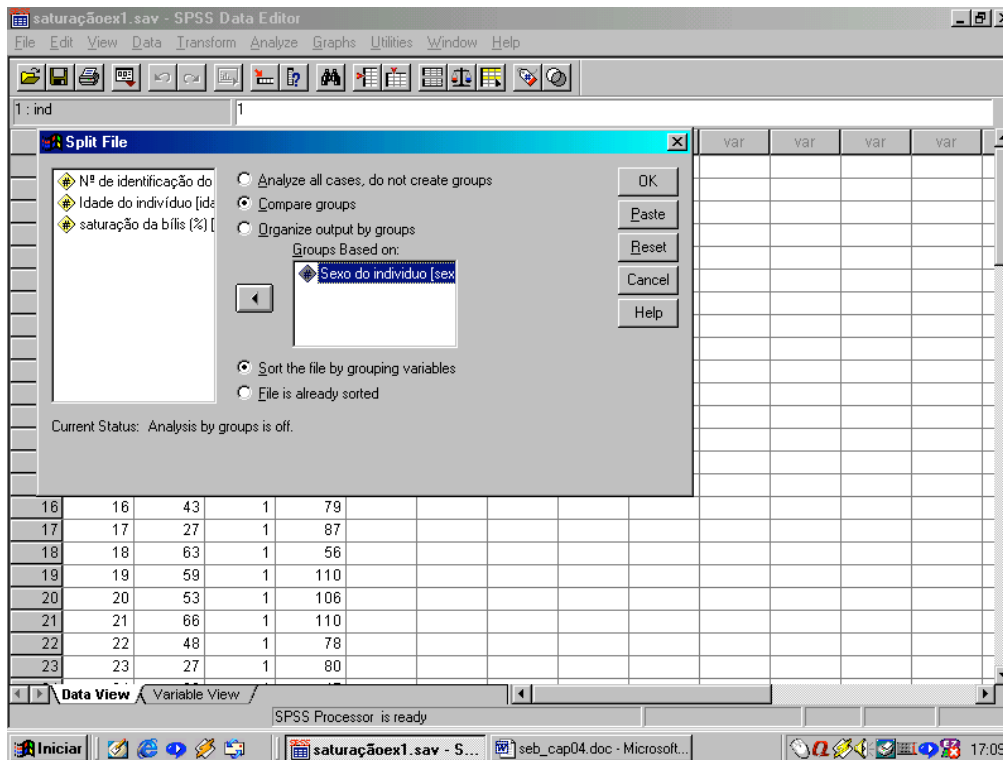


Figura 109: Janela onde se efectuará uma partição de todos os procedimentos em função da variável sexo.

De seguida, fazer como se descreveu anteriormente:

> ANALYZE >COMPARE MEANS >ONE SAMPLE T-TEST

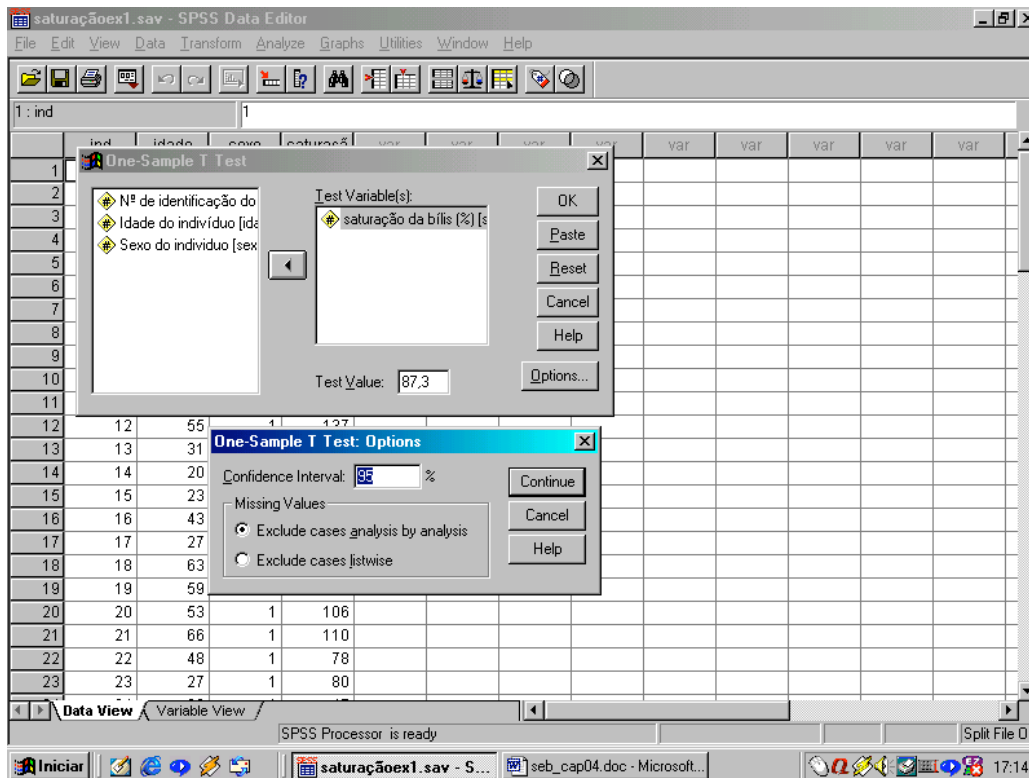


Figura 110: Exemplo de um teste t para uma amostra, mostra a variável e o valor a testar assim como o nível de significância.

	N	M	M
M a u	3		2
f i u	2	0	0

Figura 111: Tabela resultante de teste t simultâneos para uma variável agrupada por outra, onde se apresentam algumas estatísticas para os dois grupos (neste caso masculino e feminino).

	Z		W		f	
	t	d	z	0	De	
					b	l
M a u	6	0	2	0	0	2
f u	2	2	2	9	6	2

Figura 112: Tabela resultante de teste t simultâneos para uma variável agrupada por outra, onde se apresentam entre outras informações o valor p para os dois testes.

No primeiro quadro são apresentados, para a variável pretendida, tantos grupos de estatísticas quantas as categorias criadas. Neste caso, uma para o sexo feminino e outra para o masculino.

O p_value é superior ao nível de significância ($=0.05$) em ambos os casos, pelo que, não se rejeita H_0 e, com uma confiança de 95%, não se pode concluir que a média do estudo em cada sexo, seja significativamente diferente de 87.3.

4.5. Duas amostras com dados independentes

O teste t-student também pode ser usado para testar se as médias de dois grupos independentes são iguais, isto é, se são ou não significativamente diferentes.

Para aplicar este teste é necessário que as variáveis dependentes sigam uma distribuição normal e que as variâncias sejam homogêneas.

Abrir o ficheiro *Coronary artery data*.

Averiguar se o tempo médio (*treadmill time in seconds*) é igual para o grupo 1 (saúdável) e para o grupo 2 (doentes).

Pretende-se testar:

$$H_0: \mu_1 = \mu_2 \quad \text{Vs} \quad H_1: \mu_1 \neq \mu_2$$

Fazer

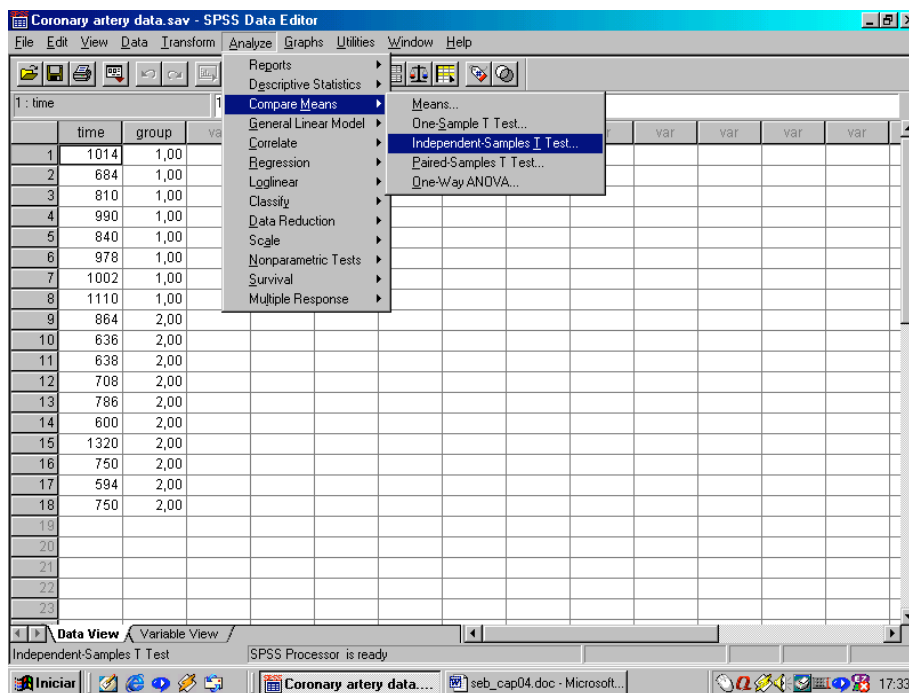


Figura 113: Menu para a execução do teste t para amostras independentes.

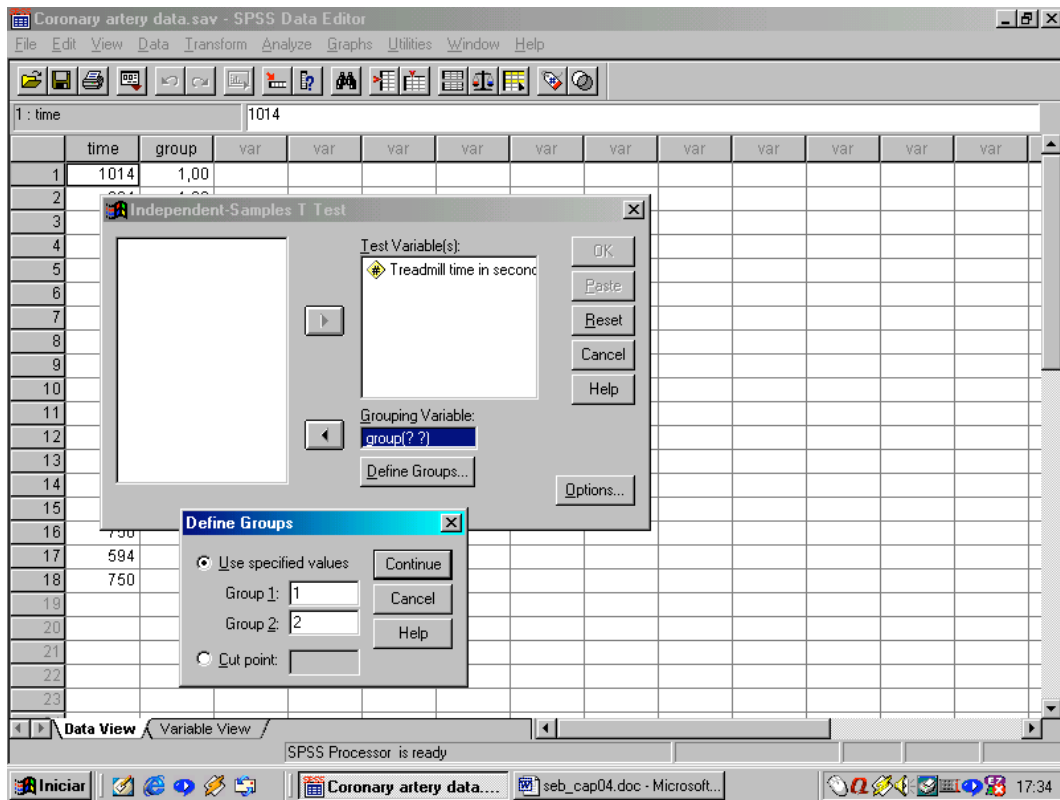


Figura 114: Janela de Diálogo onde se devem especificar os grupos.

Janela INDEPENDENT-SAMPLES T TEST:

- Test variable – variáveis escolhidas para a análise;
- Grouping variable – variáveis que dividem os casos pelos dois grupos teste.
- Define groups
 - Use specified values – se os grupos estão já codificados;
 - Cut point – seleccionar um ponto de corte para criar dois grupos: um abaixo e outro acima desse ponto.

		N	M	S	SE
1	2	8	0	2	0,7
3	4	0	0	2	0,7

Figura 115: Tabela resultante da realização do teste t para duas amostras independentes na qual são visíveis algumas estatísticas.

		t											
		F		t		df		Sig.		Mean Difference		Std. Error Difference	
1	Equal variances assumed	0.378	0.539	0.378	0.539	0.378	0.539	0.378	0.539	0.378	0.539	0.378	0.539
1	Equal variances not assumed	0.378	0.539	0.378	0.539	0.378	0.539	0.378	0.539	0.378	0.539	0.378	0.539

Figura 116: Tabela resultante da realização do teste t para duas amostras independentes na qual são mostrados os valores p assim como o intervalo de confiança para a diferença de médias.

Na primeira tabela aparecem algumas estatísticas descritivas.

Na segunda tabela, é apresentado um conjunto de informação acerca da variável pretendida.

Começa com o Teste de Levene para a homogeneidade de variâncias (**F** representa o valor da estatística de Levene e o respectivo **Sig.** ou **p_value**). Depois é apresentado o valor da estatística **t (t)**, o número de graus de liberdade (**df**), o **p_value (Sig.(2-tailed))**, a diferença de médias (**Mean Difference**), o erro amostral da diferença de médias (**Std. Error Difference**) e os limites inferior e superior do Intervalo de Confiança da diferença.

Neste caso, como o **p_value** para o Teste de Levene é superior ao nível de significância, não se rejeita H_0 e assume-se a igualdade das variâncias, deverá então ler-se o valor **p**, relativamente ao teste de igualdade das médias na 1ª linha (Equal variances assumed), como o **p_value** é superior a 0.05 (0.080) não se rejeita a hipótese nula e não se rejeita a igualdade das médias nos dois grupos ao nível de significância estabelecido.

4.6. Duas amostras com dados emparelhados

Em análise estatística é muito comum o recurso a amostras emparelhadas. Este tipo de amostras permite que os mesmos indivíduos sejam medidos em diferentes situações experimentais. O objectivo principal deste tipo de amostragem é controlar fontes de variabilidade que possam influenciar o efeito de um determinado tratamento.

Neste tipo de teste, é comparada a média das diferenças com um valor que pode ser ou não zero. No fundo, são calculadas as diferenças entre os valores das observações de cada variável das amostras.

Para aplicar este teste é necessário que as variáveis dependentes sigam uma distribuição normal.

Considere-se o exemplo seguinte:

Para testar a eficácia de um novo tratamento de emagrecimento, foram usados os dados relativos a 20 pacientes. Na tabela da página seguinte, apresentam-se os valores do peso de cada indivíduo antes e após o tratamento. Para analisar o efeito do tratamento efectua-se um teste para duas amostras com dados emparelhados.

Sujeito	Idade	Sexo	Altura	Peso inicial	Peso final
1	20	Masculino	1,77	99,8	85,3
2	45	Feminino	1,60	90,0	73,2
3	33	Feminino	1,65	100,7	68,0
4	51	Masculino	1,72	99,1	75,5
5	29	Feminino	1,69	92,4	74,2
6	18	Feminino	1,70	101,2	80,4
7	35	Feminino	1,59	79,1	60,1
8	44	Masculino	1,77	110,6	82,1
9	57	Feminino	1,62	85,3	62,4
10	61	Feminino	1,71	114,3	81,1
11	24	Masculino	1,76	101,2	74,2
12	36	Feminino	1,68	100,3	78,4
13	38	Masculino	1,81	99,5	85,3
14	46	Masculino	1,57	80,5	54,1
15	54	Masculino	1,86	116,9	90,8
16	49	Feminino	1,66	90,8	70,0
17	27	Masculino	1,69	100,5	68,1
18	21	Feminino	1,71	98,9	79,3
19	32	Feminino	1,72	103,5	83,4
20	41	Masculino	1,64	98,4	67,7

Tabela 2: Tabela contendo informação acerca de 20 indivíduos, relativa a avariáveis como idade, sexo, altura, peso inicial e peso final.

Pretende-se testar:

$$H_0: \mu_D = 0 \quad \text{Vs} \quad H_1: \mu_D \neq 0$$

Neste caso queremos comparar a média das diferenças com o valor zero, para analisar se o tratamento teve ou não efeitos.

Para isso fazer,

ANALYZE >COMPARE MEANS >PAIRED - SAMPLES T-TEST

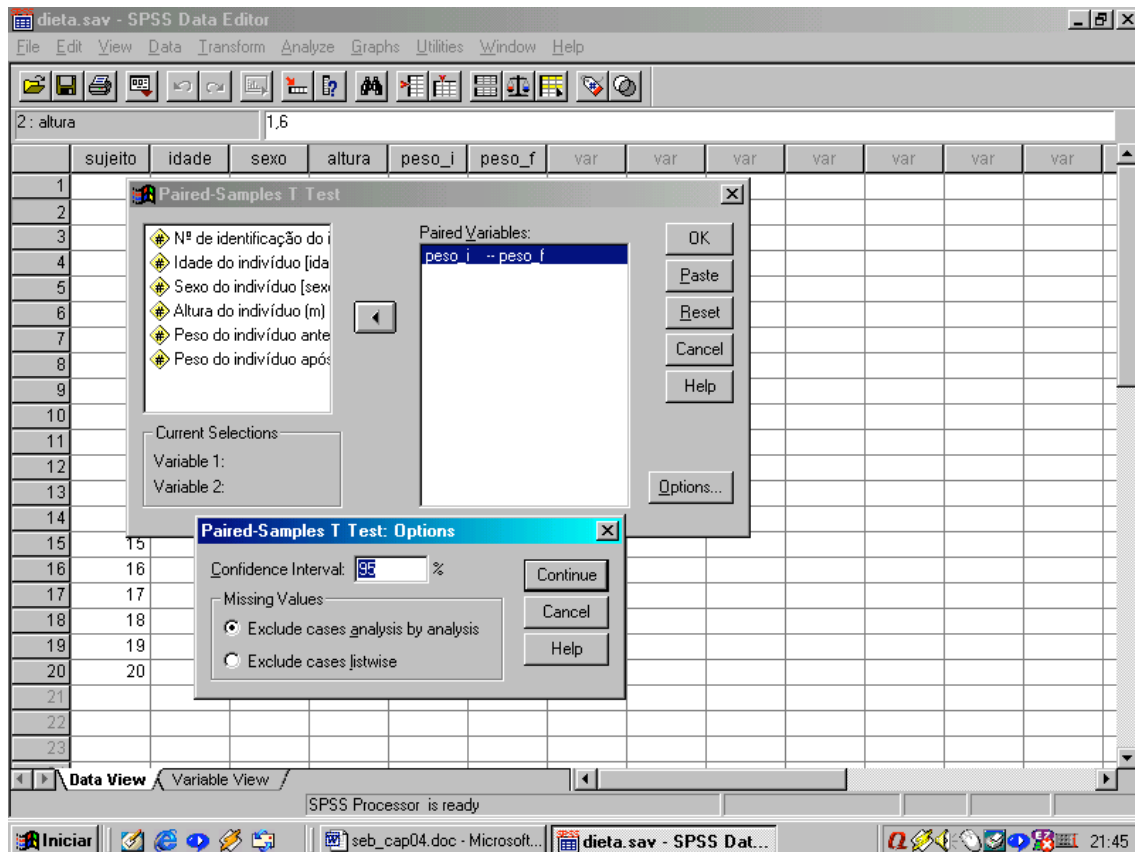


Figura 117: Janelas de Diálogo para a realização do teste t para amostras emparelhadas.

Ao seleccionar a primeira variável - peso_i – esta é assumida como variable 1; ao seleccionar a segunda variável – peso_f – esta aparece como variable 2. Só depois de seleccionadas estas duas variáveis é que o par é colocado na lista.

Em OPTIONS é seleccionado o nível para o Intervalo de Confiança. Neste caso 95%.

			N	o	M
1	d	a	0	0	0
			0	0	0

Figura 118: Tabela resultante da realização do teste t para amostras emparelhadas, onde são visíveis algumas estatísticas.

Num primeiro quadro é apresentado um conjunto de mediadas estatísticas (média, número de casos válidos, desvio padrão e erro amostral da média) para as variáveis seleccionadas.

		N	S
1	d	0	

Figura 119: Tabela resultante da realização do teste t para amostras emparelhadas, onde é visível o coeficiente de correlação.

Numa segunda tabela aparece informação que permite inferir acerca da correlação entre o par seleccionado (número de casos válidos para ambas as variáveis, coeficiente de correlação e nível de significância). Convém destacar que o coeficiente de correlação varia entre -1 e 1. Quanto mais próximo de -1 ou +1 estiver, mais linearmente relacionadas se encontram as variáveis. Se o valor prova (p_value) for inferior ao nível de significância as variáveis estão relacionadas.

1	d	a	0	0	0	0	0	0

Figura 120: resultante da realização do teste t para amostras emparelhadas, onde é visível o valor p, assim como o intervalo de confiança para a média das diferenças.

O último quadro contém informação acerca das variáveis seleccionadas: média, desvio padrão, erro amostral da média, os limites do Intervalo de Confiança, o valor da estatística t para o teste, o número de graus de liberdade e o nível de significância.

Neste caso, o p-value (**Sig. (2-tailed)**) tem o valor de 0.000 que é inferior ao nível de significância (0.05), pelo que se rejeita a hipótese nula. Por este motivo pode concluir-se que as médias são significativamente diferentes, isto é, a média no grupo mudou como resultado do tratamento.

5. Testes de hipóteses não paramétricos

Os testes não paramétricos são, geralmente, considerados uma alternativa aos testes paramétricos, nomeadamente quando as condições de aplicação dos paramétricos (normalidade da variável em estudo e homogeneidade de variâncias) não se verificam.

É geralmente aceite que a potência dos Testes Paramétricos é superior à dos Testes Não Paramétricos, isto é, a probabilidade de rejeitar correctamente H_0 é maior num Teste Paramétrico. No entanto, há autores que afirmam que os testes não paramétricos aplicados a amostras de pequena dimensão podem ser mais potentes.

Os testes paramétricos requerem geralmente variáveis quantitativas, enquanto que os não paramétricos podem ser aplicados a qualquer variável, com algumas limitações. Como tal, os Testes Não Paramétricos são aplicados sempre que não seja prático, ou possível, medir numa escala quantitativa.

É necessário considerar que alguns dos Testes Não Paramétricos foram desenvolvidos exclusivamente para variáveis ordinais, e que estes testes não são uma alternativa a utilizar numa má recolha de dados. Assim é necessário, como em qualquer inferência, assegurar a representatividade da amostra.

Os teste não paramétricos são requeridos no SPSS no menu **Analyze**, em **Nonparametrics Tests**.

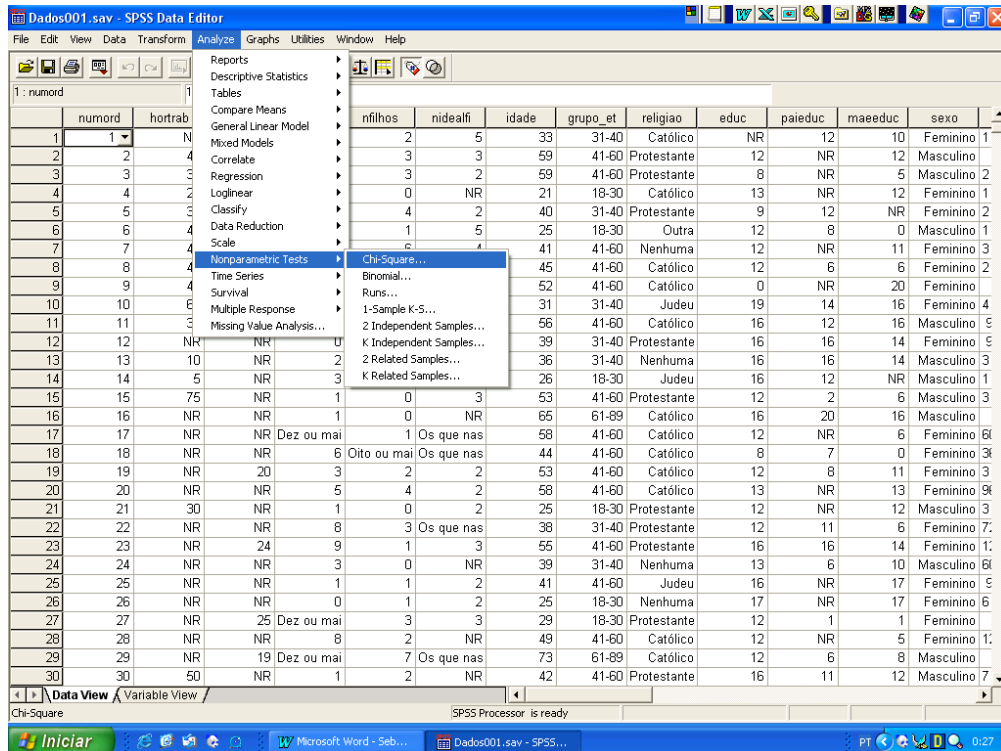


Figura 121: Menu onde se encontram os testes não paramétricos.

5.1. Teste do Qui-Quadrado

Qui-quadrado para a qualidade do ajuste

O teste do qui-quadrado permite verificar a aderência entre uma distribuição de frequências associada a uma amostra constituída por observações expressas em qualquer escala de medição e uma distribuição teórica. É um dos testes incluídos nos testes de qualidade do ajuste, ou seja permite verificar a representatividade de uma dada amostra relativamente à população.

O procedimento Qui-Quadrado executa-se *Analyze* → *Nonparametric Test* → *Chi-square*, obtendo-se a seguinte janela:

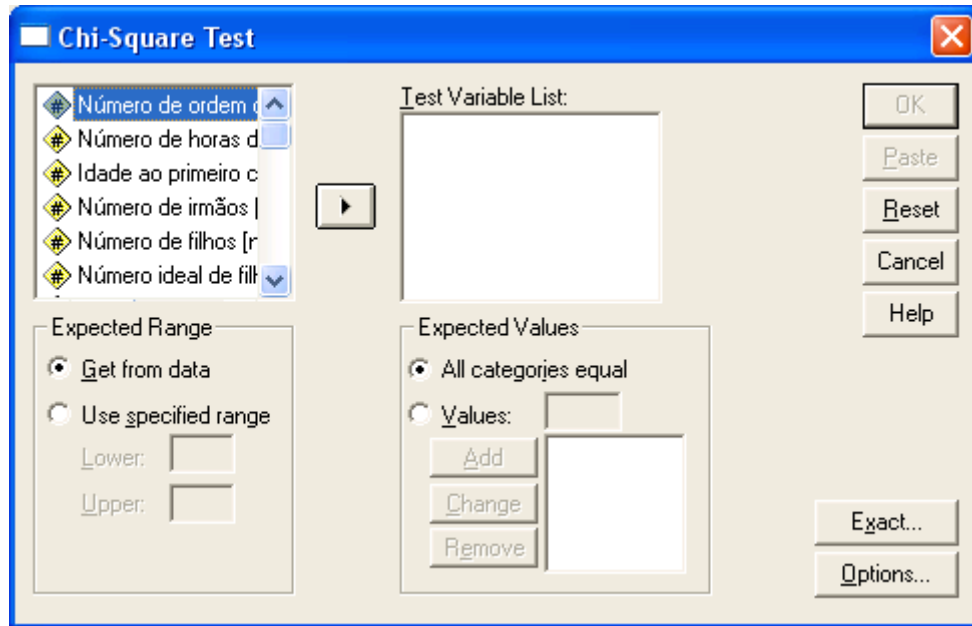


Figura 122: Janela de Diálogo para a realização do teste do qui-quadrado onde se devem indicar as variáveis assim como os valores a testar.

Na janela surge a localização da variável ou variáveis em estudo – *test variable list*. As variáveis que se pretendem analisar devem ser seleccionadas para esta lista.

Surgem igualmente opções para definir as categorias em análise – *Expected Range*:

- definida na variável – *Get from data*, (atribui a mediante a codificação, e todas estão em análise),
- definir as categorias mediante um intervalo – *Use specified range*, (neste caso apenas as categorias codificados pelo intervalo estão em análise).

A opção *Expected Values* corresponde ao valor que se pretende testar, ou seja à frequência esperada para cada categoria. Pode-se assumir como todas as categorias apresentam igual proporção (*All categories equal*), ou especificar os valores de proporção teóricos. Deve-se ter em atenção a introdução dos valores (*values*) pela ordem em que se encontram as categorias.

A Janela do Chi-Square apresenta também dois botões, *Exact* e *Options*.

Na janela *Exact Tests* existe a possibilidade de ajustar a exactidão do teste a realizar.

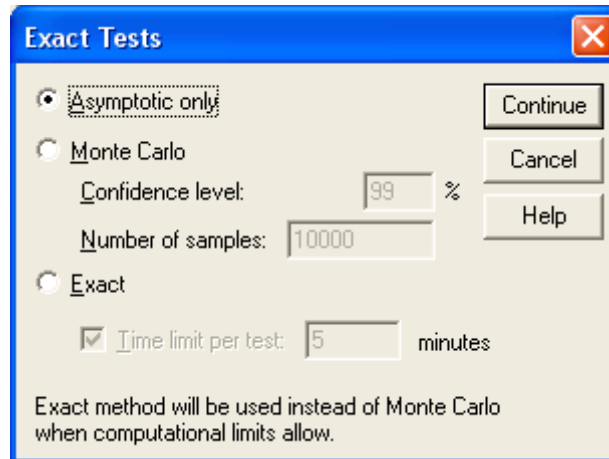


Figura 123: Janela de diálogo onde se pode optar por um teste exacto ou por uma aproximação à normal.

As opções são as mesmas que surgem na Janela *Exact Tests* que surge no procedimento *Analyze* → *Descriptives Statistics* → *Crosstabs*. Esta janela é igual para todos os testes não paramétricos.

Asymptotic only – apenas assintótica – o nível de significância baseado numa distribuição assintótica da estatística de teste é geralmente de 5%. Baseia-se na hipótese de o conjunto de dados ser elevado. Caso a amostra apresente uma dimensão pequenas, ou pobremente distribuída, pode não ser uma boa indicação para a significância do teste.

Monte Carlo – uma estimativa não enviesada do nível de significância exacto, calculado repetidamente por amostragens a partir de um conjunto de tabelas com as mesmas dimensões da tabela em estudo. Permite calcular significâncias exactas.

- *Confidence Level* – nível de confiança – introduzir um valor entre 0,1 e 99,9 para especificar a probabilidade com que o intervalo de confiança, baseado nos valores amostras, inclui a média da população.
- *Number of Samples* – número de amostras - introduzir um valor entre 1 e 1.000.000.000 para o número de amostras a utilizar no calculo da estatística. *Números mais elevados aumentam a fiabilidade, mas demoram mais tempo a serem executados.*

Exact – exacto – a probabilidade do resultado observado ou um resultado mais extremo ser calculado com exactidão. Tipicamente um valor menor que 0,05 é considerado significativo.

- *Time limit per Test* – limite de tempo por teste - introduzir um valor entre 1 e 9.999.999.999 como tempo máximo permitido para calcular cada teste.

Na janela de opções tem-se a possibilidade de requerer as estatísticas descritivas (média, desvio padrão, mínimo, máximo e dimensão da amostra) e os quartis.

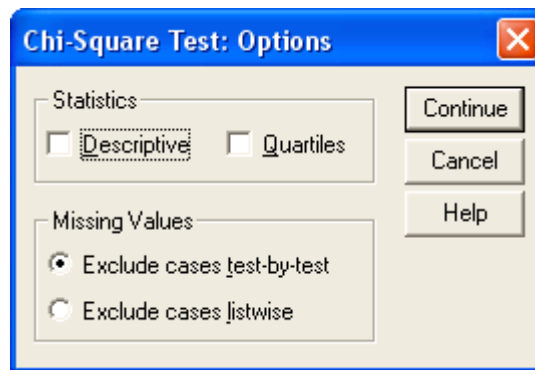


Figura 124: Janela de Diálogo onde se pode optar por “pedir” algumas estatísticas, assim como o tratamento aos missings.

Surgem igualmente as opções de tratamento de missings. O caso com um missing numa determinada variável pode ser excluído, apenas se a variável estiver envolvida no teste em questão (*Exclude cases test-by-test*) (tal significa que os dados desse caso serão usados em todos os procedimentos nos quais a variável não esteja envolvida) ou em todos os procedimentos mesmo que a variável não esteja envolvida (*Exclude cases listwise*).

Esta janela é igual para todos os testes não paramétricos.

Exemplo Prático

Num dado inquérito solicitava-se aos inquiridos que assinalassem a sua posição política, em termos de direita, centro e esquerda. Os resultados encontram-se na base de dados **Inquerito1**, na variável **política**.

Nas últimas eleições os eleitores votaram de acordo com a tabela seguinte:

	Direita	Centro	Esquerda
Proporção(n)	0,29	0,40	0,31

Tabela 3: Tabela contendo resultados de uma eleição.

Teste a hipótese desta amostra ser representativa da população em estudo para a variável em causa.

Esta primeira tabela contém os valores observados (Observed N) e os valores esperados (Expected N). Esta tabela foi ponderada pelos valores teóricos. Os resíduos correspondem à diferença entre os valores observados e os valores esperados.

a

	E	X	
O	0	2	2
O	5	0	
O	0	0	0
O	3		

Resultados:

Figura 125: Tabela resultante do teste do qui-quadrado do ajuste onde se apresentam valores observados e esperados, assim como os resíduos.

A tabela seguinte contém o valor da estatística do teste (**Chi-square**), o número de graus de liberdade (**df**), e o nível de significância (**Asymp. Sig.**) neste caso assintótica. Adicionalmente fornece informação sobre o próprio rigor do teste e da sua aplicação.

b

χ^2	a	0
df		2
p		0

a 0
0

Figura 126: Tabela resultante do teste do qui-quadrado do ajuste onde se apresenta o valor p.

Genericamente as hipóteses em estudo são:

H_0 : a população segue uma distribuição específica

H_1 : a população não segue essa distribuição específica

Conclusão:

No presente exemplo verificam-se os pressupostos para a aplicação do Qui-quadrado (a não existência de valores observados =0; apenas 20% dos valores esperados podem ser < 5; nenhum valor esperado poder ser <1)

O valor de prova (p) = 0,001, pelo que se conclui pela rejeição da hipótese nula, para um $\alpha=0,05$. Ou seja conclui-se pela não representatividade da amostra na variável posição política para a população.

Qui-Quadrado para independência de variáveis

Tabelas dupla entrada – Crosstabs

O procedimento *Crosstabs* permite a criação de tabelas de contingência de duas ou mais variáveis. Permite igualmente a realização de testes e de medidas de associação de variáveis, o teste a realizar é determinado pelo tipo de variáveis presentes na análise.

Neste procedimento as variáveis em análise devem ter um número limitado de categorias. Não devem ser utilizadas variáveis do tipo quantitativo.

Na janela *crosstabs* surgem as opções de visualizar gráficos de barras agrupados e a de suprimir as tabelas (neste caso apenas se obtêm medidas estatísticas). Nas linhas (*Row(s)*) e nas colunas (*Column(s)*) colocam-se as variáveis em análise, deve existir pelo menos uma variável em cada.

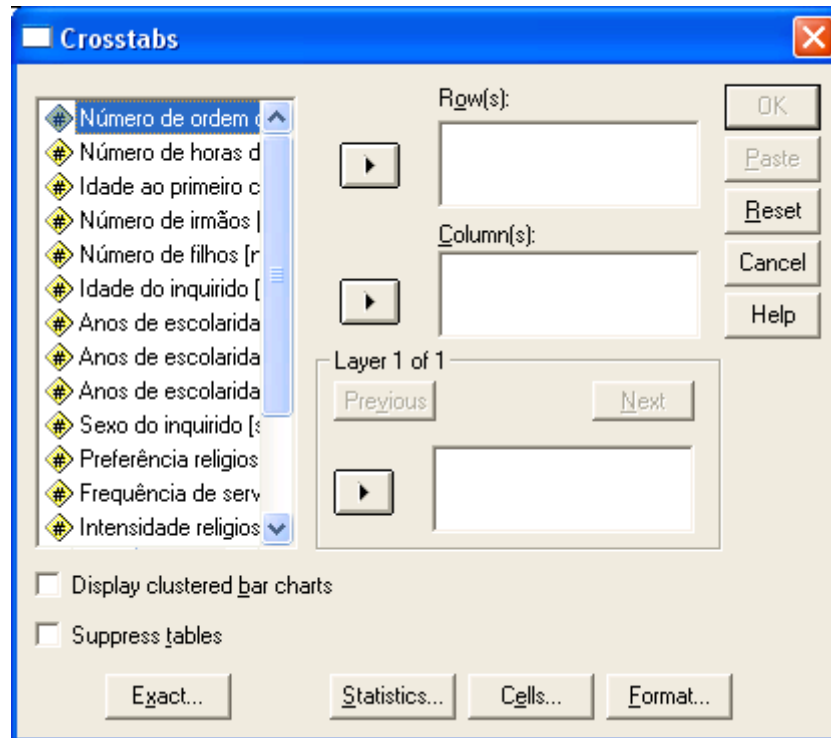


Figura 127: Janela de diálogo para proceder à realização do teste do qui-quadrado para a independência de variáveis, onde se deverão indicar quais as variáveis a testar.

Os botões *Statistics*, *Cells* e *Format*, abrem as janelas correspondentes.

Na janela *Statistics* pode-se proceder aos testes de independência de variáveis e de correlações.

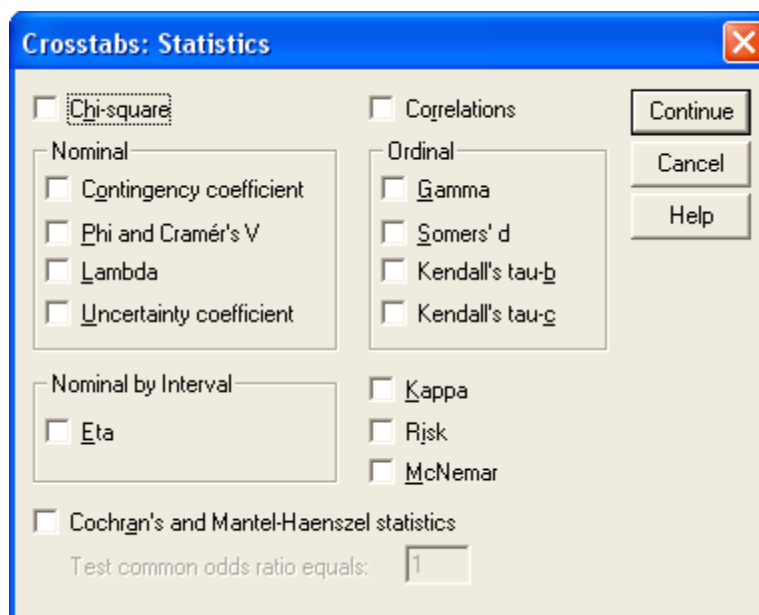


Figura 128: Janela de diálogo onde se podem “pedir” algumas estatísticas relativas às variáveis a testar no teste do qui-quadrado para a independência de variáveis

O qui-quadrado (*Chi-square*) testa apenas a hipótese de associação entre as variáveis, não indicando a intensidade ou direcção da associação.

As correlações (*Correlations*) apresentam o coeficiente de Pearson e o de Spearman, o primeiro para variáveis intervalares/racionais e o segundo para variáveis ordinais. Variam entre -1 e 1 , o 0 corresponde a uma ausência de relação, o sinal $-$ ou $+$ corresponde à direcção da associação.

É possível requerer outros coeficientes de correlação, tendo sempre em consideração o tipo de variável. A interpretação do resultado é idêntica à realizada para o coeficiente de Pearson ou de Spearman. As correlações nominais são isentas de sinal, a ordem de codificação é aleatória, não segue uma ordem, logo não existe uma direcção de associação.

Na janela *cell display* proporciona a opção de seleccionar os valores a visualizar na tabela de contingência, para além dos observados pode-se requerer os valores esperados, percentagem e resíduos. Os resíduos padronizados permitem uma comparação mais rápida dos desvios relativos aos valores esperados dos observados. Os resíduos padronizados são significativos quando superiores a 2 em termos absolutos. Resíduos padronizados superiores a 2 são suspeitos de existência de associação entre as variáveis.

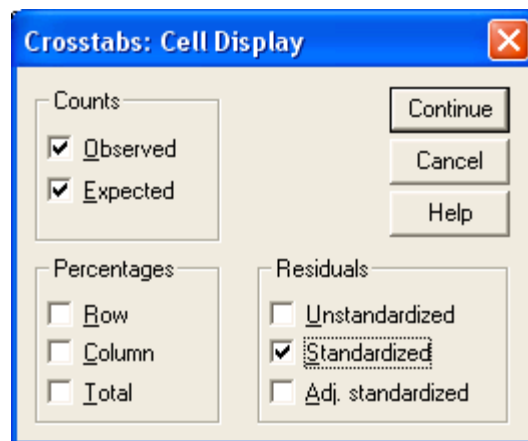


Figura 129: Janela de diálogo onde se pode optar por quais as informações a apresentar na célula da tabela de contingência.

A janela Format apresenta as opções de ordenação dos resultados.

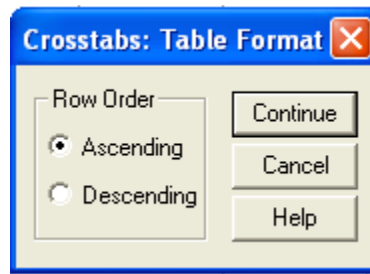


Figura 130: Janela de diálogo onde se pode optar qual a forma de ordenação dos dados.

Exemplo Prático

Num inquérito solicitava-se aos inquiridos que assinalassem a sua intensidade com que vivem a religião e a frequência de actos religiosos. Os resultados encontram-se na base de dados Inquerito1, na variável freqreli e intenrel. Teste a hipótese destas variáveis serem independentes.

Na primeira tabela temos um resumo da análise realizada em termos de número de casos válidos e *missings*.

	e		s			
	N	%	N	%	N	%
0	9	0%	7	2%		0%

Figura 131: Tabela resultante do teste do qui-quadrado para independência, onde se mostra o número de casos válidos, missings e total.

A segunda tabela, tabela de dupla entrada, apresenta-nos os valores observados, os valores esperados e os resíduos padronizados. Pela análise desta tabela pode-se prever a dependência entre as variáveis. Os resíduos padronizados elevados, e o facto de seguirem uma linearidade na oblíqua leva-nos a pensar que tal se verifica.

g

			b	h	b	h
F e e	h p	C	3	9	9	9
		h	3	1	2	9
		h	6	6	5	
	h	C	0	9	2	3
		h	0	9	3	
		h	0	4	6	
	h	C	0	2	6	0
		h	3	0		
	h	C	6	3	2	9
		h	9	0		
	h	C	6	2	6	3
		h		0	0	
	e	C	9	2	5	7
		h	0	8	5	0
		h	8	1	0	
	h	C	5	0	9	9
		h		0		
		h	0	2	6	
	h n	C	3	4	7	2
		h	0	3	9	
		h	0	2	0	
h		C	2	6	6	9
		h				

Figura 132: Tabela de contingência resultante da execução do teste do qui-quadrado.

Resultados:

A terceira tabela apresenta-nos as estatísticas do teste e o valor de prova.

e

		e	f	R
a	b		4	0
	h		4	0
	h		1	0
	h	9		

a b (

Figura 133: Tabela resultante da execução do teste do qui-quadrado onde se mostram os valores prova.

Genericamente as hipóteses em estudo são:

H_0 : as variáveis são independentes

H_1 : as variáveis não são independentes

Neste exemplo podem ser concretizadas da seguinte forma:

H_0 : Os actos religiosos e a intensidade com que se vive a religião são independentes.

H_1 : Os actos religiosos e a intensidade com que se vive a religião não são independentes.

Conclusão:

No presente exemplo o valor de prova (p) $< 0,001$. Pelo que se conclui pela rejeição da hipótese nula, para um $\alpha=0,05$. Ou seja conclui-se que os actos religiosos e a intensidade com que se vive a religião não são independentes ao nível de significância estabelecido.

5.2. Binomial

O procedimento *binomial* permite testar se a proporção de uma categoria de uma variável dicotómica, ou tornada dicotómica, da amostra difere de uma determinada constante, por ex. a proporção do sexo na população.

Na janela surge a localização da variável em estudo – *test variable* – e o valor com que se vai proceder ao teste test Proportion, note que este valor a testar ficará associado ao 1º grupo, a proporção para o segundo grupo será 1 menos o valor introduzido.

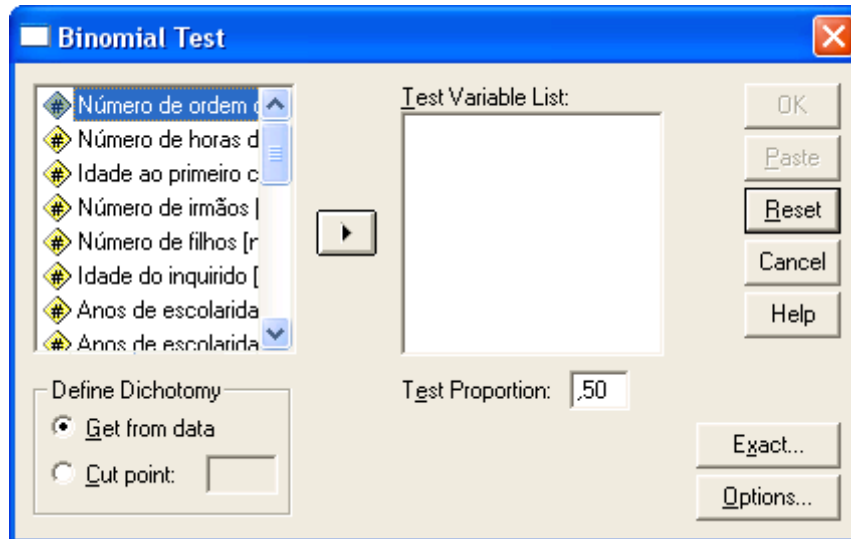


Figura 134: Janela diálogo para a execução do teste da binomial, onde se devem indicar as variáveis assim como as proporções a testar.

Surgem igualmente as opções para definir a dicotomia, pode ser a definida na variável (atribui os casos com a categoria mais baixa a um grupo e os de categoria mais alta a outro), ou definir a dicotomia por um ponto de corte, o valor a inserir deve corresponder ao value da categoria a ficar no segundo grupo, as categorias com value inferior ficam no grupo 1, o grupo 2 é constituído pelo valor inserido e superiores. A variável deve estar bem definida para não misturar diferentes itens.

A Janela do Binomial apresenta também dois botões, Exact e Options, já tratadas anteriormente.

Exemplo Prático

Num inquérito solicitava-se aos inquiridos que assinalassem a seu sexo. Os resultados encontram-se na base de dados **Inquerito1**, na variável **sexo**. Nos últimos censos realizados a proporção do sexo feminino era de 55%.

Teste a hipótese desta amostra ser representativa da população em estudo para a variável em causa.

Resultados:

A tabela de resultados apresenta a dicotomia em análise, a dimensão de cada categoria, a proporção observada. E o nível de significância do teste.

a

		N		p		p	
0	6	6	3	3	5	0	0
	6	6	3	3			
	6	6	3	0			

a B

Figura 135: Tabela resultante do teste da binomial onde são visíveis as proporções observadas, a proporção a testar assim como o valor p.

As hipóteses em estudo são:

Ho: a população apresenta uma proporção de femininos = 0,55

H1: a população apresenta uma proporção de femininos ≠ 0,55

É necessário ter em atenção se os resultados gerados são para um teste bilateral ou unilateral. O SPSS quando testa uma proporção diferente de 0,5 apresenta um teste unilateral. O nível de significância do teste, cujas hipóteses foram referidas anteriormente, obtém-se multiplicando o resultado Asymp. Sig. (1-tailed) por 2.

Conclusão:

No presente exemplo o valor de prova (p) = 0.188. Pelo que se conclui pela não rejeição da hipótese nula, para um α=0,05. Ou seja conclui-se pela não representatividade da amostra na variável sexo para a população.

5.3. Duas amostras Independentes (Mann-Whitney)

O procedimento two independent samples compara dois grupos de casos (amostras) independentes e verifica se provêm da mesma população.

Os testes não paramétricos utilizam as ordens para obter os resultados, ou seja, procedem ao ordenamento crescente dos valores das variáveis e calculam a sua posição média. É com este valor que procedem à comparação entre as amostras.

Na janela surge a localização da variável ou variáveis em estudo – *test variable* e a variável que vai dividir nos dois grupos em teste. A *grouping variable* deve ser uma variável dicotómica ou com os grupos que se querem testar.

Na definição dos grupos a comparar introduz-se os *values* correspondentes às categorias que se pretendem comparar.

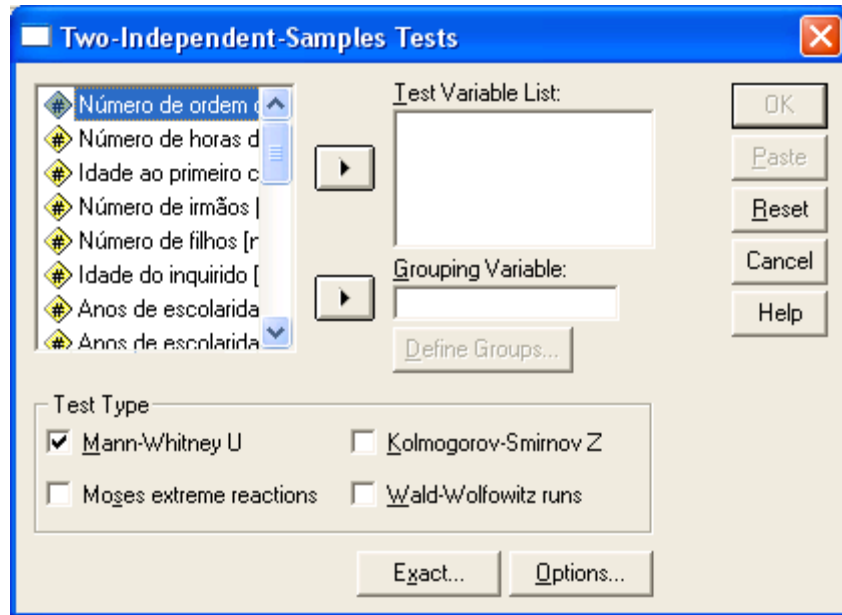


Figura 136: Janela de diálogo para a execução de testes não paramétricos para amostras independentes, onde se deve indicar quais as variáveis e grupos a testar.

Tipo de teste

Mann-Whitney U – o teste não paramétrico equivalente ao teste t, – testa a hipótese de duas amostras serem provenientes da mesma população. Requer que as variáveis se encontrem pelo menos numa escala ordinal. É mais potente que o teste de medianas por utilizar a ordem de casos. O valor de U é o número de vezes que um valor do primeiro grupo precede um valor no segundo grupo, quando ordenados de forma crescente.

Kolmogorov-sminov Z – testa a hipótese de duas amostras serem provenientes da mesma população. É sensível a qualquer tipo de diferença entre as duas distribuições. O teste é baseado na maior diferença entre as duas distribuições acumuladas.

Moses extreme reactios – concebido para testar hipóteses nas quais se esperam que a variável experimental afectará alguns casos numa direcção e outros casos na direcção oposta. Teste para respostas extremas comparadas com o grupo de controlo.

Wald-Wolfowitz runs – testa a hipótese de duas variáveis serem provenientes da mesma população. Os valores de ambas as amostras são combinados e ordenados de forma crescente. Se as variáveis forem provenientes da mesma população, os dois grupos deverão estar distribuídos aleatoriamente ao longa da ordenação.

A Janela do *Binomial* apresenta também dois botões, *Exact* e *Options*, já tratadas anteriormente.

Exemplo Prático

Num inquérito solicitava-se aos inquiridos que assinalassem o seu sexo e os anos de escolaridade completos. Os resultados encontram-se na base de dados **Inquerito1**, na variável **sexo e educ**. Teste a hipótese de os anos completos de escolaridade não diferirem entre os dois sexos.

Resultados

k

	6	N	7	
Á	M	0	3	7
α	E	0	0	
	B			

Figura 137: Tabela resultante do teste de Mann-Whitney onde se apresentam os grupos em análise, a dimensão, a posição média do grupo e o somatório de posições.

a

	d
MW	0
W	0
Z	4
pn	0

a 6

Figura 138: Tabela resultante do teste de Mann-Whitney, onde se apresenta o valor p.

A primeira tabela apresenta os grupos em análise, a dimensão, a posição média do grupo e o somatório de posições.

A segunda tabela apresenta os valores do teste estatístico, e a seu valor de prova.

Genericamente as hipóteses em estudo são:

Ho: A distribuição da variável nas duas populações é idêntica.

H₁: A distribuição da variável nas duas populações não é idêntica.

No exemplo podem ser concretizadas da seguinte forma:

H₀: Os anos de escolaridade completos são idênticos nos dois sexos.

H₁: Os anos de escolaridade completos não são idênticos nos dois sexos.

Conclusão:

No presente exemplo o valor de prova (p) = 0,909. Pelo que se conclui pela não rejeição da hipótese nula. Ou seja não se pode concluir que os anos de escolaridade nos dois sexos não são idênticos, ao nível de significância 0.05.

5.4. Two Related Samples (Wilcoxon)

O procedimento *Two Related-samples* compara as distribuições de duas amostras emparelhadas.

Na janela do comando procede-se à selecção do par de variáveis a testar, a visualização da selecção é projectada em *current selections*, deve-se marcar a primeira das variáveis e ela deverá surgir em *variable 1*, marca-se a segunda das variáveis e ela deverá surgir com *variable 2*. Só após seleccionar o par de variáveis é que é possível colocar o par na lista de variáveis a testar.

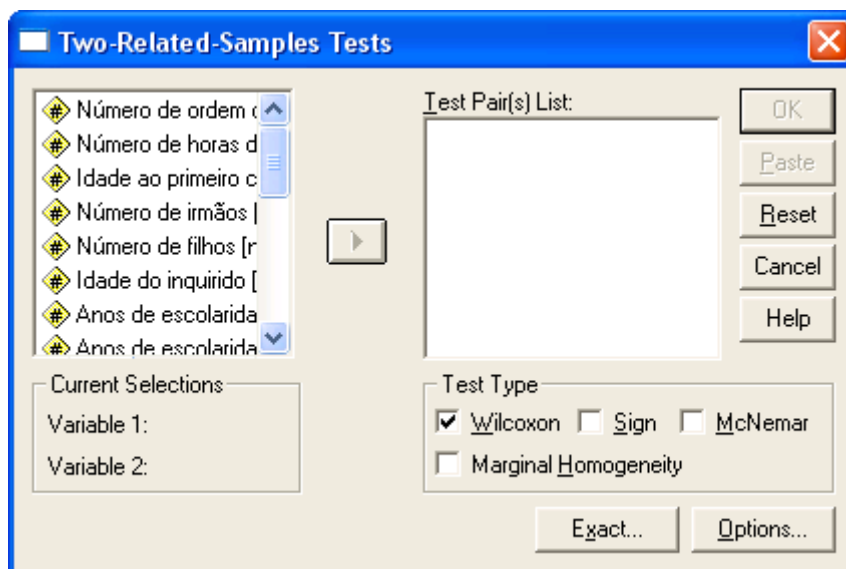


Figura 139: Janela de diálogo para a execução dos testes não paramétricos para amostras emparelhadas.

Tipo de teste

Wilcoxon – testa a hipótese de duas variáveis serem provenientes da mesma população, sem postular qualquer hipótese sobre a forma de distribuição das duas variáveis. A estatística do teste baseia-se nas ordens dos valores absolutos das diferenças entre as duas variáveis.

Sign – teste dos sinais – testa a hipótese de duas variáveis serem provenientes da mesma população. As diferenças entre as duas variáveis para todos os casos são calculados e classificados como positivos, negativos ou empate. Se duas variáveis forem similarmente distribuídas, as diferenças positivas e negativas não são estatisticamente significativas.

McNemar – Teste não paramétrico para duas variáveis emparelhadas dicotômicas. Testa as modificações nas respostas recorrendo a uma distribuição do Qui-Quadrado. É um teste útil para detectar as mudanças de opinião após uma intervenção (antes-depois).

Marginal Homogeneity – Teste semelhante ao McNemar, mas para variáveis multinomiais.

Exemplo Prático

Num inquérito solicitava-se aos inquiridos que assinalassem os anos de escolaridade completos do pai e da mãe. Os resultados encontram-se na base de dados **Inquerito1**, na variável **maeduc e paeduc**. Teste a hipótese de os anos completos de escolaridade não diferirem entre o pai e a mãe do inquirido.

Resultados

a

		N	\bar{x}	$\sum R$	
a	A	8	3	3	0
b	B	8	6	3	0
a	A	8	2		
b	B	8	6		

a A
 b B
 a A
 b B

Figura 140: Tabela resultante do teste de Wilcoxon onde se apresentam a média e soma das posições.

a b

	d e f
Z	$\hat{\sigma}$
$\hat{\mu}$	$\hat{\sigma}$

a B
 b W

Figura 141: Tabela resultante do teste de Wilcoxon onde se apresenta o valor p.

A primeira tabela apresenta os resultados das diferenças, a dimensão, a posição média dos grupos positivo e negativo e o somatório de posições para os mesmos.

A segunda tabela apresenta os valores do teste estatístico, e o seu valor de prova.

Genericamente as hipóteses em estudo são:

H₀: As distribuições das variáveis são idênticas.

H₁: As distribuições das variáveis não são idênticas.

No exemplo podem ser concretizadas da seguinte forma:

H₀: As distribuições das variáveis anos de escolaridade completos da mãe e anos de escolaridade completos do pai são idênticas.

H₁: As distribuições das variáveis anos de escolaridade completos da mãe e anos de escolaridade completos do pai não são idênticas.

Conclusão:

No presente exemplo o valor de prova (p) = 0,971. Pelo que se conclui pela não rejeição da hipótese nula. Ou seja não se pode concluir que distribuição das variáveis anos escolaridade do pai e da mãe sejam idênticas na população, ao nível de significância 0.05.

6. Regressão

A análise da regressão permite, através da relação entre variáveis, prever a evolução de uma em função da outra. Esta previsão é em termos médios ou de valores esperados. Por exemplo, podemos definir a nota média dos alunos à disciplina de BBII em função do número de horas de estudo que dedicaram à cadeira, supondo que existe entre estas duas variáveis uma relação neste sentido. Neste exemplo, estamos a considerar a nota de BBII como variável dependente e o número de horas de estudo como variável independente e $Y = f(X)$.

Esta relação funcional pode traduzir uma relação linear, exponencial, entre outras. Vamos considerar a regressão linear.

6.1. Regressão Linear Simples

Sendo Y variável dependente e X a variável independente podemos considerar a função mais simples, a equação linear $Y = a + bX$ que traduz um grande número de relações entre variáveis e é uma boa aproximação em relações matematicamente mais difíceis de descrever.

Nesta equação, a e b são constantes que se calculam com base em dados amostrais.

A partir do momento em que conhecemos as constantes a e b podemos prever o valor médio de Y para qualquer valor de X que pertença à região de validade do modelo.

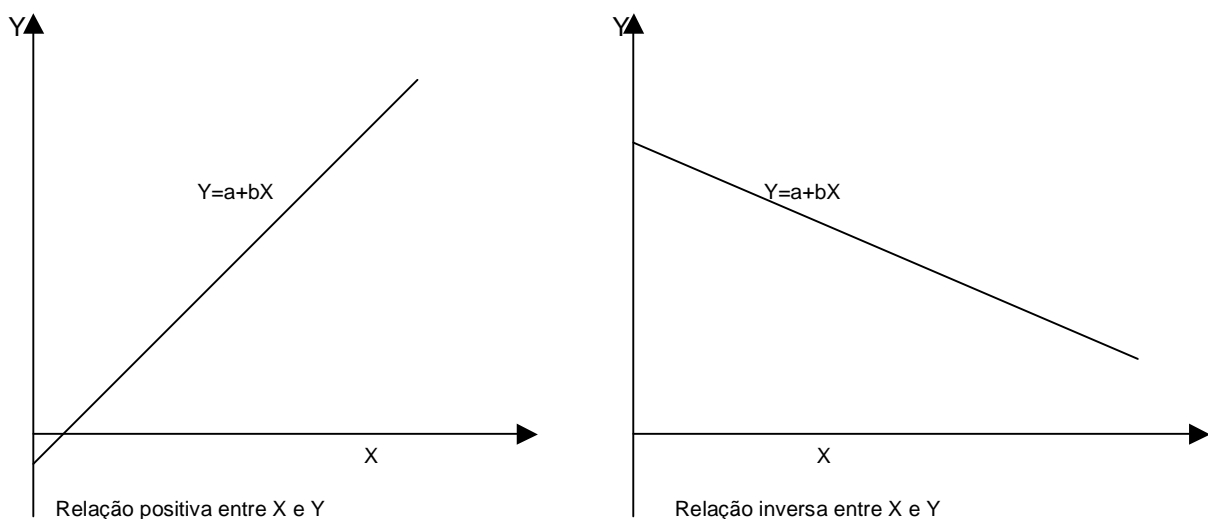


Figura 142: Representação das possíveis relações lineares entre duas variáveis.

O coeficiente a dá-nos o valor médio de Y quando X assume o valor zero. Só faz sentido proceder à sua análise quando o valor zero para X pertence à região de validade do modelo.

O coeficiente b dá-nos a variação média de Y quando X varia uma unidade. Se for um valor positivo traduz uma relação positiva entre as duas variáveis, ou seja, quando X aumenta, Y aumenta. Se for um valor negativo traduz uma relação inversa entre as duas variáveis, ou seja, quando uma aumenta a outra diminui.

No SPSS após introdução dos dados amostrais, para obter a análise da regressão linear entre duas variáveis devemos seguir os seguintes passos:

Analyse → Regression → Linear

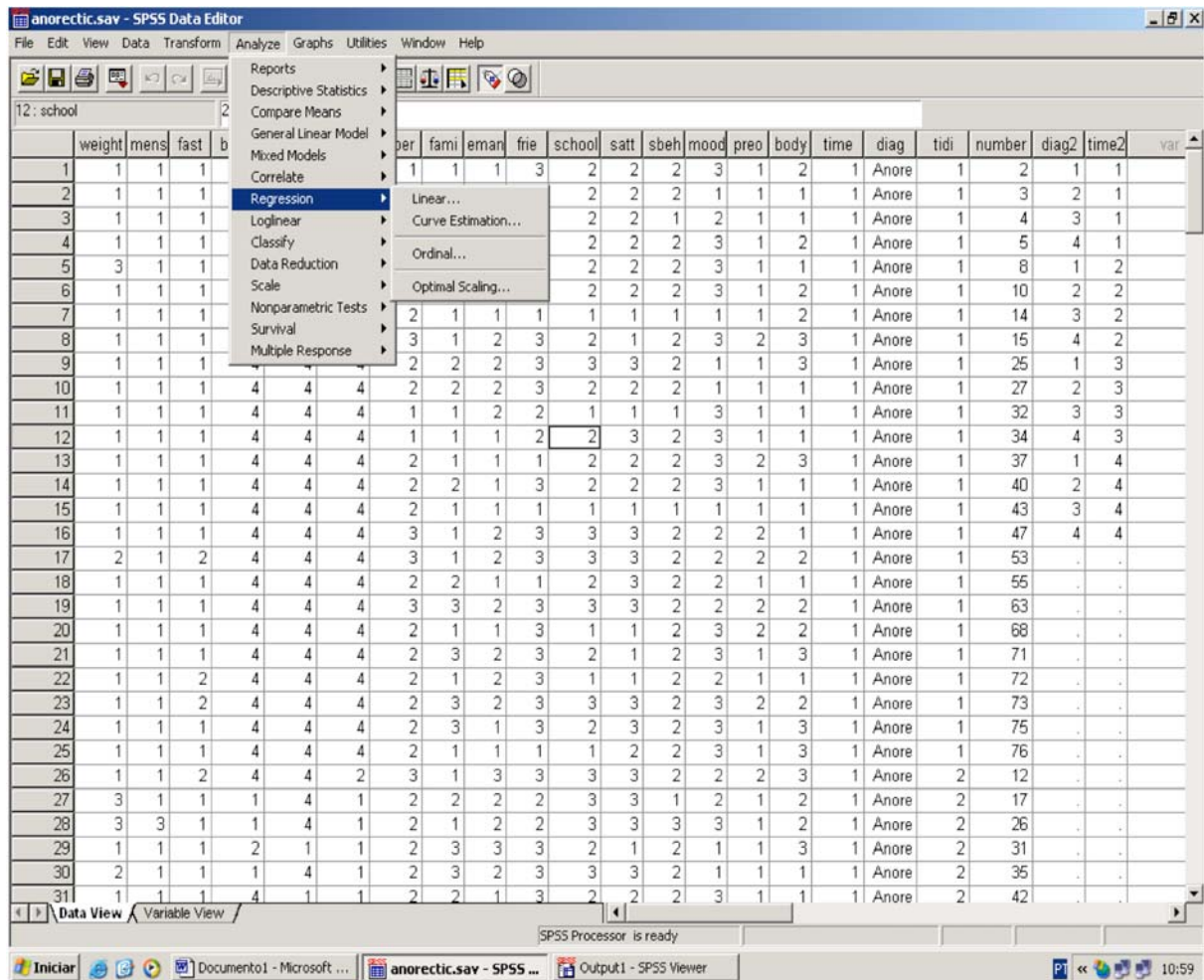


Figura 143: Menu para a execução da regressão linear.

Seleccionamos a variável dependente e independente e OK.

X Y

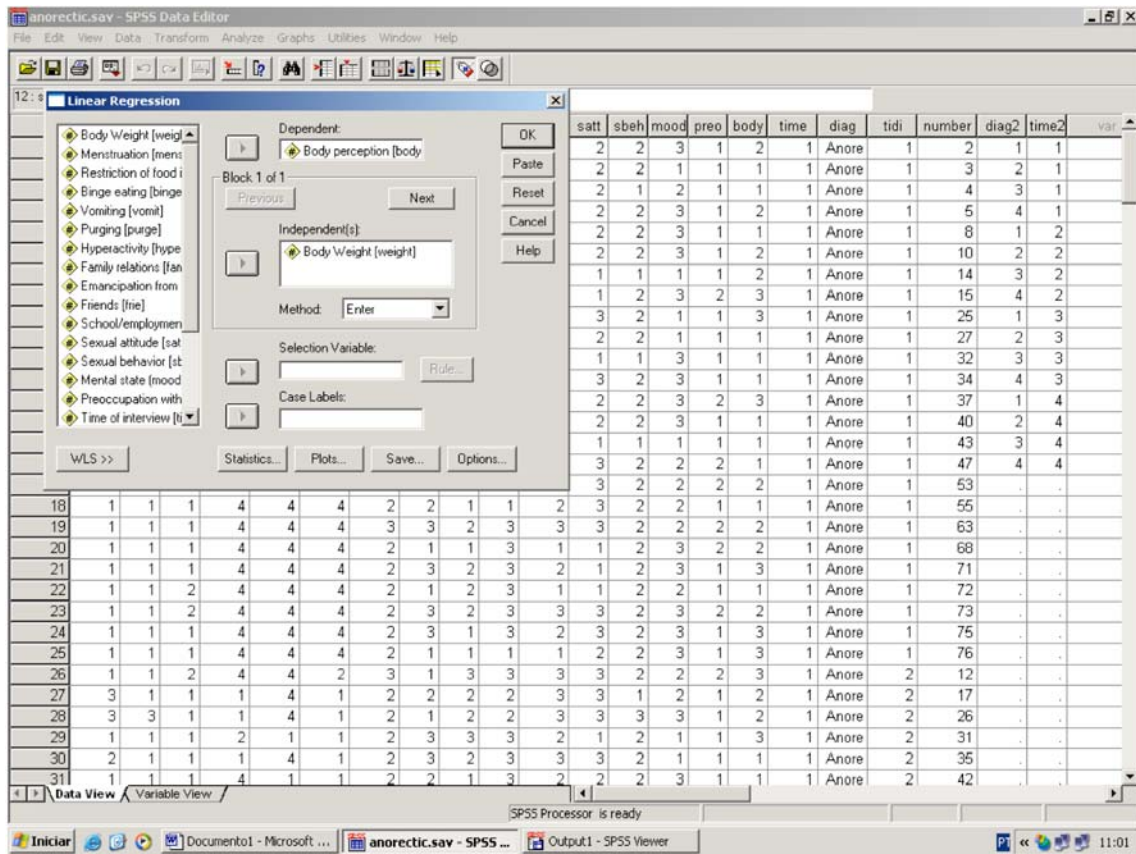


Figura 144:Janela de diálogo para a execução da regressão linear onde se deve indicar quais as variáveis dependente e independentes.

E obtemos, entre outros, o quadro com os coeficientes a e b.

a

M	B	t	
1	0	θ	4

a B B θ b

Figura 145: Tabela resultante de uma regressão linear onde se apresentam os coeficientes.

Neste caso, em que a nossa variável dependente era a Percepção Corporal (o SPSS indica qual a variável dependente em baixo do quadro) e a independente o Peso, obtemos a seguinte equação:

$$Y = 1,800 + 0,163X$$

Quando o peso é zero a percepção corporal é em média 1,800. Neste caso, não faz sentido a análise do coeficiente a porque zero não pertence à região de validade do modelo.

Quando o peso varia uma unidade (1Kg) a percepção corporal aumenta em média 0,163.

6.2. Regressão Linear Múltipla

A regressão linear simples considera apenas uma variável independente. No entanto, as situações reais mostram-nos muitas vezes que para explicar a variação de uma variável existe mais do que uma variável a explicar essa variação.

Assim, a equação da regressão linear múltipla é

$$Y = a + b_1X_1 + b_2X_2$$

onde o coeficiente a é o valor médio de Y quando X₁ e X₂ assumem simultaneamente o valor zero. Só faz sentido a sua análise quando o valor zero para X₁ e X₂ pertence à região de validade do modelo.

O coeficiente b₁ traduz a variação média de Y quando X₁ varia uma unidade e o coeficiente b₂ traduz a variação média de Y quando X₂ varia uma unidade, na região de validade do modelo.

No SPSS para obter os coeficientes da equação procedemos da mesma forma da considerada na regressão linear simples, definindo como variável independente não uma mas as n variáveis independentes que estivermos a considerar no modelo.

a

		B		t	B		
1	C				6	9	
	D		0	0	7	8	
	E	0	3	2	8	0	
	F	0	2	0	2	0	

a

Figura 146: Tabela resultante de uma regressão linear múltipla onde se apresentam os coeficientes.

A equação que obtemos é $Y = 0,242 - 0,053X_1 + 0,659X_2 + 0,398X_3$

6.3. *Estimação da Curva*

O modelo de regressão linear pode ser definido teoricamente das seguintes forma:

$$Y_i = \beta_0 + \beta_1 X_i + \xi_i$$

Onde β_0 e β_1 são parâmetros e ξ_i representa o erro aleatório de média zero e variância σ^2 . Assumimos que os erros das observações i e j não estão relacionados.

O valor esperado de Y pode ser definido como $E(Y_i) = \beta_0 + \beta_1 X_i$

No SPSS podemos pedir a estimação da curva de duas formas. Se pretendemos um modelo linear podemos seguir o definido nos pontos acima. Se o nosso objectivo for construir um modelo linear ou não linear podemos optar em

Analyse \Rightarrow Regression \Rightarrow Curve Estimation

onde temos diversos modelos para opção. Podemos igualmente optar por construir a análise de regressão através de uma série temporal escolhendo a opção Time na definição da variável independente. A série temporal permite-nos definir a evolução ao longo do tempo de uma determinada variável e prever o valor médio esperado na variável dependente para um determinado momento do tempo.

A estimação dos coeficientes da equação é baseada no método dos mínimos quadrados que nos permite encontrar estimadores não tendenciosos e com variância mínima. Este método considera para cada observação o desvio entre o valor Y_i e o valor esperado $E(Y_i)$, encontrando os estimadores a e b que minimizam o quadrado deste desvio.

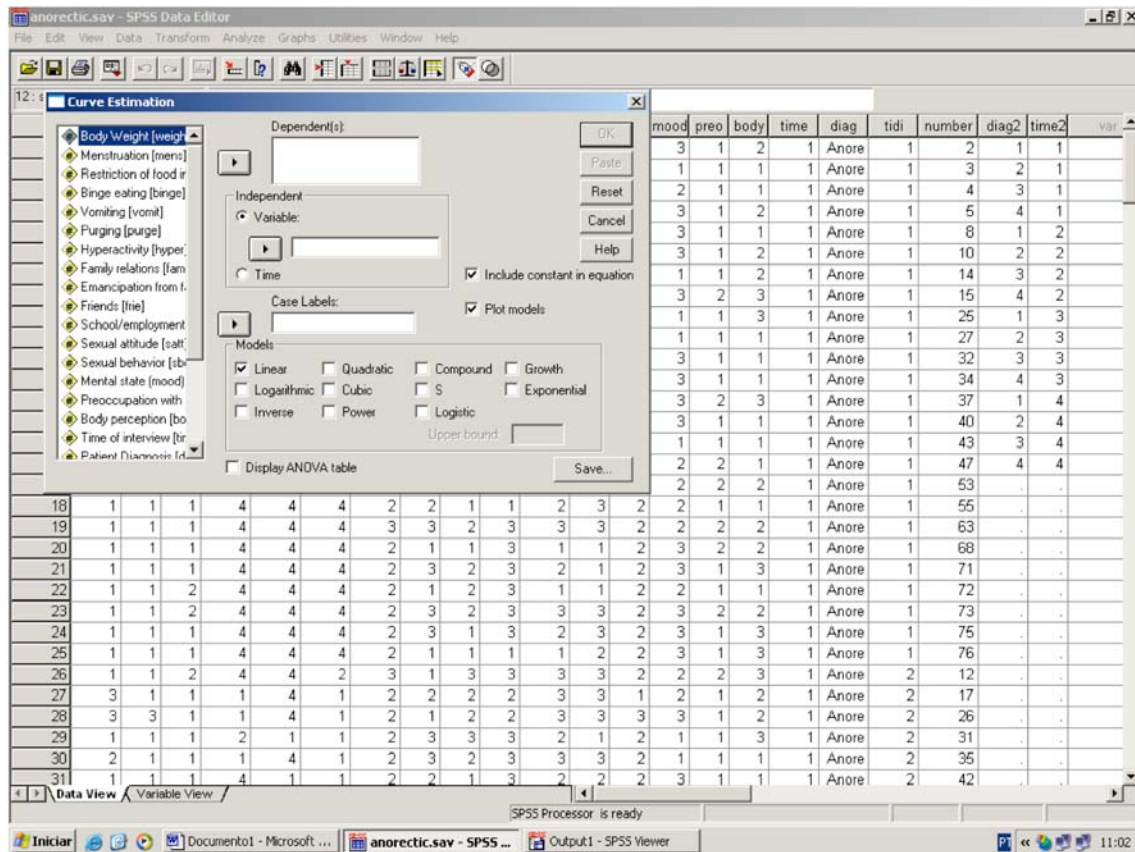


Figura 147: Menu para a estimação da curva.

Teste aos coeficientes

Podemos testar a hipótese de os coeficientes serem ou não iguais a zero, ou seja, avaliar se apesar do valor positivo ou negativo encontrado na estimação da recta para os coeficientes podemos considerar que na população eles apresentam um valor significativamente diferente de zero.

Para β_0

$$H_0: \beta_0 = 0$$

$$H_1: \beta_0 \neq 0$$

O valor p obtido é aproximadamente zero (ver quadro supra dos coeficientes), logo inferior ao nível de significância escolhido (0,05), pelo que rejeitamos H_0 e concluímos que para o nível de significância de 0,05 temos evidência estatística para considerar que o valor de β_0 é significativamente diferente de zero

O teste para β_1 obedece ao mesmo critério de análise.

R e R²

Na análise da regressão há duas medidas de associação entre a variável dependente e a(s) variáveis independentes:

- Coeficiente de Determinação (R²) indica-nos a proporção da variação total de Y que é “explicada” por X
- Coeficiente de Correlação (R) indica-nos o nível de associação entre as variáveis. Apresenta um sinal positivo ou negativo consoante o coeficiente b for positivo ou negativo.

O SPSS quando procede à análise da regressão sumaria o modelo utilizando o coeficiente de determinação e o coeficiente de correlação.

M	R			
1	a	θ	θ	θ
	a		b	

Figura 148: Tabela resultante da regressão linear onde se apresentam os coeficientes de determinação e correlação.

Neste caso, R²= 0,090, ou seja, a variável Peso “explica” apenas 9% da variação da Percepção Corporal e R=30% o que nos dá indicação de uma relação fraca entre as variáveis.