



Introdução à análise estatística com SPSS

Guião nº2: Descrição de variáveis estatísticas

Abrir o ficheiro de dados com o SPSS

Abrir o SPSS e depois a opção **open an existing data source** e procurar pelo ficheiro *Ex3_Questionnaire_data.sav*¹. Alternativa será clicar duas vezes no ficheiro com os dados. Clique na janela **variable view** e observe as variáveis que estão no ficheiro. Compare os resultados apresentados com o ficheiro construído no guião nº1.

Descrever dados qualitativos nominais

1. Obter uma tabela de frequências.

Usando o comando **Analyse/Descriptive Statistics/Frequencies**, selecionar as variáveis qualitativas nominais para o quadro designado por **variables**. Clicar no **OK**. Inspeccione as tabelas de frequências e verifique se as categorias propostas estão todas representadas.

Questão 1: Preencha a tabela nº1 (próxima página) com os resultados obtidos.

Dica: O valor da frequência relativa é apresentado na coluna *valid percent*, enquanto os valores omissos (a existirem) são apresentados na linha *missing system*.

2. Medidas de estatística descritiva.

Com o comando **Analyse/Descriptive Statistics/Frequencies**, clique em **statistics**. Para este tipo de variável (qualitativa nominal) a única opção é a moda (**mode**). Clique em **continue** e depois em **ok**.

¹ O ficheiro *Ex3_Questionnaire_data.sav* está disponível no *website* da disciplina



Tabela 1- Resultados para as variáveis nominais selecionadas (sem nenhuma ordem em particular e não estão todas as variáveis nominais apresentadas)

Variável	Frequência absoluta (n)	Frequência relativa (%)	Frequência relativa acumulada (%)	Valores omissos (n, %)
Sexo				2 (0.6)
Masculino	202	60.8	60.8	
Feminino	130	39.2	100	
Faculdade				
Engenharia				
Ciências				
Saúde				
Outra				
Historial médico				
Obeso				0 (0)
Sim	65	19.5	19.5	
Não				
Diabetes				
Sim				
Não				
Fisicamente ativo				
Sim				
Não				
Fumador				
Sim				
Não				
Outro				
Sim				
Não				
Dor				
Sim				
Não				

Questão 2: Utilizando os resultados que encontraram para a moda, será que os poderiam obter de outra forma?

Dica: O valor da moda não é o código, mas sim a categoria que é representada pelo código e observe os resultados da tabela nº1.

3. Visualização gráfica

Voltando ao comando **Analyse/Descriptive Statistics/Frequencies**, clicar em **charts**. Para este tipo de variável (qualitativa nominal) as opções aconselháveis são



gráfico de barras (**Bar charts**) ou gráficos circulares (**Pie charts**). Os gráficos podem ser representados em valores absolutos (**Frequencies**) ou em valores relativos (**Percentages**). Escolha uma das opções para o tipo de gráfico, clique em **continue** e depois em **ok**. Inspeccione os resultados obtidos e confirme os resultados com a tabela nº1. Se pretender editar o gráfico para alteração dos títulos, das cores utilizadas, ou do tamanho e tipo de letra, clique duas vezes no gráfico. Uma janela designada por **Chart Editor** aparece e então é possível fazer as alterações que se pretendem. Quando terminar, basta fechar a janela do **Chart Editor** no seu canto superior direito. Proposta para modificação: alterar o fundo da figura para branco (clique na área do gráfico a cinzento correspondente á figura, seleccionar a opção branco na janela das **properties** e clicar em **apply**), alterar a cor das barras para um padrão a preto e branco (clique numa das barras, no menu **pattern** seleccionar o padrão axadrezado, no menu **fill** seleccionar a cor branca e clicar em **apply**), alterar o padrão apenas da barra correspondente ao sexo feminino (clique duas vezes da barra do sexo feminino, clique na opção **Fill & Border** e no menu **pattern** seleccionar o padrão diagonal, e clicar em **apply**). Para as opções relacionadas com texto, basta clicar nas legendas correspondentes e através o menu **Text Style** é possível alterar o tipo e o tamanho de letra. Clicando mais uma vez é possível editar o conteúdo das legendas.

Descrever dados qualitativos ordinais

1. Obter uma tabela de frequências.

Através do comando **Analyse/Descriptive Statistics/Frequencies**, seleccionar as variáveis qualitativas ordinais para o quadro designado por **variables**. Clique no **OK**. Inspeccione novamente as tabelas de frequência á procura de alguma inconsistência, como por exemplo, alguma categoria que não pertence á variável em questão.

2. Medidas de estatística descritiva.

Voltando ao comando **Analyse/Descriptive Statistics/Frequencies**, clique em **statistics**. Para este tipo de variável (qualitativa ordinal) as opções são a mediana (**median**) e os quartis (**quartile**). A moda (**mode**) pode também ser calculada, mas não



é muito usual. Para definir um determinado percentil, deve utilizar a opção **percentile(s)**, indicado o valor que pretende calcular numa escala [0;100], seguido de **add**. Clique em **continue** e depois em **ok**.

Questão nº3: Preencha a tabela nº2 com os resultados.

Dica: As medidas robustas são calculadas para a variável e não para as suas categorias. É importante indicar a categoria que o código representa e não o valor numérico desse código.

Tabela 2- Resultados para as variáveis ordinais

Variável	n (%)	1º Quartil	Mediana	3º Quartil
Percepção da saúde	328 (98.2)	poor	acceptable	good
Pobre				
Razoável				
Boa				
Excelente				
Quantifique a dor	334 (100)			
Sem dor	115 (34.4)			
Ligeira				
Moderada				
Severa	85 (25.4)			
Insuportável				

3. Visualização gráfica

A representação gráfica de variáveis ordinais pode ser feita através gráfico de barras (**Bar charts**) ou gráficos circulares (**Pie charts**). No entanto, para este tipo de variável pode-se construir caixas de bigodes (que é baseada na mediana e nos quartis). Através do comando **Analyse/Descriptive Statistics/explore**, selecione as variáveis ordinais para o quadro do *dependent list*. De seguida, clicar em **plots** e confirmar que a opção **factor levels together** na gravura *boxplot* está ativa e desativar a opção **Stem-and-leaf**². Clicar no **Continue** e depois no **ok**.

² Por opção dos autores, esta opção não irá ser discutida.



Questão nº4: Inspeccione as caixas de bigode. Existem *outliers* e extremos? **Dica:** O SPSS identifica os *outliers* com círculos e os extremos com asteriscos.

Cruzamento de duas variáveis qualitativas (tabela de contingência)

Uma tabela de contingência não é mais que uma tabela ou matriz que cruza informação de duas variáveis qualitativas. O objectivo é cruzar a informação sobre a variável *Sex* e variável *Faculty*. Com o comando **Analyse/Descriptive Statistics/Crosstabs**, escolha uma das variáveis para as linhas (**rows**) e a outra variável para as colunas (**columns**). A ordem é da escolha é arbitrária. Clique em **cells** e no quadro **percentages**, clique em **row**, **column**, e **total**. Clicar no **Continue** e depois no **ok**.

Questão nº5: Da inspeção da tabela resultante diga, quantas alunas existem na Faculdade de Saúde e qual a sua percentagem em relação aos alunos dessa faculdade? Calcule mais algumas percentagens.

Descrever dados quantitativos

1. Obter uma tabela de frequências.

Usando o comando **Analyse/Descriptive Statistics/Frequencies**, seleccionar as variáveis quantitativas: *Age*, *Kilos*, *Meters* e *NpDay* para o quadro **variables**. Clicar no **ok**. **Nota:** Tabelas de frequência muito compridas (com muitas classes ou com muitos valores absolutos baixos) não são muito interessantes para a análise. Será sempre preferível representar os dados por um método gráfico

2. Medidas de estatística descritiva.

Novamente através do comando **Analyse/Descriptive Statistics/Frequencies**, Clicar em **statistics** e seleccionar quartis (**quartile**), média (**mean**), mediana (**median**),



desvio padrão (**Std. Deviation**), mínimo (**minimum**) e máximo (**maximum**). Clique em **Continue** e depois no **Ok**.

Questão nº6: Preencha a seguinte tabela e faça um resumo, comentando as medidas de tendência central e de dispersão encontradas.

Tabela 3- Resultados para as variáveis idade, peso, altura, nº de cigarros por dia.

Variável	n	M	DP	Med	Q25%	Q75%
Idade (anos)	332	31.2	7.8	30.0	25.0	36.0
Peso (kg)						
Altura (m)						
Numero de cigarros/dia						

M: média; DP: desvio-padrão; Med: Mediana; Q25%: quartil 25%; Q75%: quartil 75%

3. Visualização gráfica

No caso de uma variável quantitativa, as melhores opções são a caixa de bigodes (**boxplot**) ou o histograma (**histogram**). Através do comando **Analyse/Descriptive Statistics/explore**, seleccione a variável *Age*. De seguida, clicar em **plots** e confirmar que a opção **boxplot** está activa, activar a opção **histogram** e desactivar a opção **Stem-and-leaf**. Clicar no **Continue** e depois no **Ok**.

Questão nº7: Inspeccione a caixa de bigode e o histograma e responda se existem *outliers* e extremos e se considera haver simetria?

No caso de duas ou mais variáveis quantitativas, a opção será fazer um gráfico de dispersão (**scatterplot**). Através do comando **Graphs/chart builder**³, e clicando no **ok**, têm-se acesso a todos os elementos gráficos existentes no SPSS. Escolher a opção **scatter/dot** na lista disponível **choose from**, clicar 2x no gráfico que está na primeira linha e na primeira coluna designado por **simple scatter** (alternativamente pode-se arrastar do quadro para a janela acima, através de *drag and drop*). Escolher a variável

³ Uma alternativa ao chart builder é utilizar os menus das opções gráficas através do comando **Graphs/legacy dialogs**



meters para **X-axis** e variável *kilos* para **Y-axis**, arrastando as variáveis para as caixas texto. Clicar em **Ok**.

Questão nº8: Inspeção o gráfico de dispersão e indique se existe alguma tendência entre as variáveis altura e peso.

No caso de mais do que duas variáveis (mas comparando apenas duas a duas) a opção a utilizar está localizado na segunda linha e na segunda coluna (**scatter matrix**). Arrastar as variáveis *Age*, *Meters*, *Kilos* (uma de cada vez) para a caixa de texto **scattermatrix** e depois clicar em **Ok**.

Questão nº9: Inspeção a matriz dos gráficos de dispersão e comente se existem algumas tendências entre as variáveis consideradas.

Cruzamento de uma variável quantitativa por uma ou mais variáveis qualitativas

1. Medidas de estatística descritiva.

Uma das situações mais frequentes numa análise estatística é cruzamento de uma variável quantitativa por uma variável qualitativa. Através do menu **Analyse/Descriptive Statistics/Explore** iremos calcular algumas estatísticas descritivas dos dados. A variável dependente ou a variável medida é sempre uma variável quantitativa, enquanto a variável independente ou factor é sempre uma variável qualitativa. Supondo que se pretendia estudar a idade (**Age**) em função da Faculdade (**Faculty**) e de como classifica a sua saúde (**Health**). Defina como variável dependente idade (**Age**) e como factors: Faculdade (**Faculty**) e saúde (**Health**). Clicando em **Statistics**, activar a opção **Percentiles** (mantendo as opções por defeito já seleccionadas) e clicar no **continue**. Se clicar em **Plots**, desativar a opção **Stem-and-leaf** e clicar em **Histogram** (mantendo as opções por defeito já seleccionadas).

Questão nº10: Preencha a tabela nº4.



Tabela 4- Resultados para as variáveis idade em função da Faculdade e de como classifica a sua saúde.

Variável	n	M	DP	Med	AI
Faculdade					
Engenharia					
Saúde	86	33.0	7.1	32.0	10.0
Ciências					
Percepção da saúde					
Pobre					
Razoável	55	28.7	7.9	27.0	6.0
Boa					
Excelente					

M: média; DP: desvio-padrão; Med: Mediana; AI: Amplitude inter-quartil;

2. Visualização gráfica

Graficamente, as opções mais utilizadas são as caixas de bigodes ou os gráficos de médias e desvios-padrão. Estas opções estão no menu **Graphs/chart buider**. Suponha que queria construir uma caixa de bigodes para mostrar graficamente o cruzamento de duas variáveis nominais (*sex* e *faculty*) e uma variável quantitativa (*Age*). Na lista disponível escolher a opção **boxplot** e depois a opção **clustered boxplot** (a opção do meio) e arrastar para a área do **chart preview**. O **X-Axis** deve ser ocupado com a variável nominal com mais classes (*Faculty*) e o **cluster** pela variável nominal com menos classes (*Sex*), embora a ordem seja arbitrária. O **Y-axis** é sempre preenchido pela variável quantitativa (*Age*).

Questão nº11: Inspeção a caixa de bigode e comente se existem *outliers* e extremos?

Para o gráfico de médias e desvios-padrão, o menu é o **Graphs/chart buider** e na lista disponível escolher a opção **bar**. Escolher a opção **clustered barot** (linha 2, coluna 4) e arrastar para a área do **chart preview**. A escolha das variáveis para as opções **X-Axis**, **cluster** e **Y-axis** são idênticas á construção da caixa de bigodes anterior.

Questão nº12: O que pode comentar sobre este gráfico? Compare com os resultados obtidos na caixa de bigodes anterior.