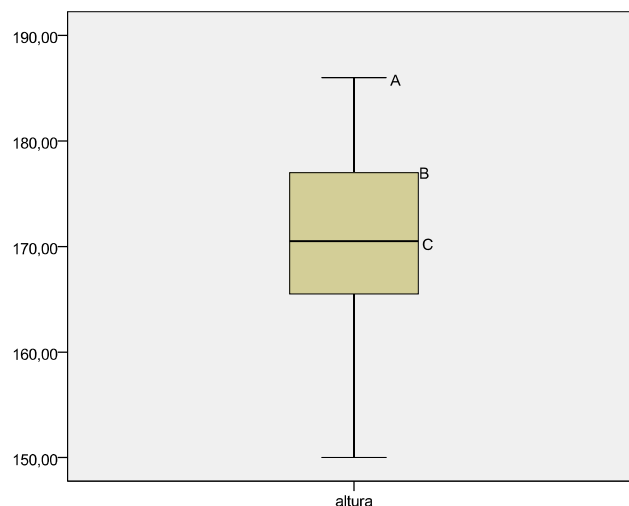


1. O ficheiro Ex1_dados_alunos.xlsx contém informação sobre os alunos de uma turma de uma determinada escola de música da cidade. O género encontra-se codificado, sendo 1- feminino; 0- masculino. Pretende-se representar a informação da amostra recorrendo a gráficos.
 - 1.1 Crie uma nova variável "ID" atribuído a cada registo um número de identificação, ordenados por ordem crescente iniciando em 1.
 - 1.2 Das características analisadas caracterize-as quanto à escala de medida.
 - 1.3 Elabore alguns gráficos que lhe pareçam elucidativos para representar variáveis nominais.
 - 1.4 Elabore o histograma do peso construído com classes de amplitude igual a 10. Altere a amplitude para 5 ou 8.
 - 1.5 Elabore o gráfico de extremos e quartis (diagrama de "caixa e bigodes") para a altura e posteriormente para a altura separado pelo género. Que pode concluir pela observação do mesmo.
 - 1.6 Determine as medidas de localização para a altura separada por género e insira texto no gráfico da alínea anterior indicando os valores extremos e os quantis representados no gráfico.
 - 1.7 Determine medidas de localização e/ou dispersão possíveis para a característica: cor de olhos. Confirme os seus resultados recorrendo ao Excel.
 - 1.8 Determine as medidas de localização e/ou dispersão possíveis para a característica altura.

" Cerca de 50% dos alunos da turma têm menos _____ cm de altura e 25% têm mais de _____ cm de altura"
2. Numa determinada família registaram-se as seguintes características dos seus elementos: cor dos olhos (1- azul, 2- verdes, 3 castanhos), altura(cm) peso(kg) e idade. Os resultados foram obtidos através de um software estatístico.
 - 2.1 Indique os valores representados em A, B e C do gráfico
 - 2.2 Interprete o valor D indicado na tabela.
 - 2.3 Comente a adequabilidade informação apresentada na coluna indicada com *.
 - 2.4 " Cerca de 50% dos elementos da família têm menos _____ anos de idade e 25% têm mais de _____ anos"

Statistics					
		altura	peso	cor_olhos(*)	idade
N	Valid	36	36	36	36
	Missing	0	0	0	0
Mean		170,56	69,33	1,61	48,00
Minimum		150,00	46,00	1	15,00
Maximum		186,00	100,00	3	81,00
Percentiles 25		165,25	60,25(D)	1,00	34,25
50		170,50	67,00	1,00	49,50
75		177,00	77,75	2,00	58,75



3. No sítio da internet do Instituto da Conservação da Natureza e das Florestas existe informação sobre Lista de incêndios florestais, ao nível do local. Responda às questões relativamente ao ano 2015 (<http://www.icnf.pt/portal/florestas/dfci/inc/estat-sgif>) e referindo-se apenas aos incêndios cuja área total ardida está compreendida entre 5 e 100 hectares. Estes valores estão disponíveis no ficheiro Ex3_incendios2015.xls.

3.1 Construa uma tabela onde seja possível analisar as frequências absolutas, as frequências relativas e as frequências relativas acumuladas do total de área ardida. Recorra a uma tabela em classes de amplitude igual a 10 unidades (recorra ao excel).

3.2 Construa um histograma de frequências relativas (recorra ao excel).

3.3 Recorrendo à tabela obtida na 1ª alínea determine:

- a percentagem de incêndios cuja a área total ardida excede os 55 hectares.
- A mediana e o 3º quartil. Interprete os valores obtidos.
- A média. Compare o valor obtido pelo valor obtido pelo excel. Que pode concluir?
- Esboce o diagrama de extremos e quartis, também denominado *box-plot*. Compare o seu esboço com o *box-plot* efetuado recorrendo ao excel.

3.4 Construa uma tabela de dupla entrada para a informação distrito e tipo causa.

- Indique os distritos com maior e menor percentagem de incêndios registados e indique as percentagens. Construa um gráfico que lhe permita visualizar esta informação.
- Indique a causa de incêndio mais frequente. Qual foi a sua percentagem?
- Qual o distrito com maior causa Negligente. Qual a sua percentagem?

iv. Para Viseu indique as percentagens de incêndios para as diferentes modalidades tipo de causa de incêndio.

3.5 Realize um estudo comparativo para a área ardida de mato, povoamento e agrícola. Recorra à medidas estatísticas de localização e dispersão. Interprete os seus valores. Qual a conclusão geral?

4. Uma transportadora pretende comparar a variabilidade do peso com a variabilidade do volume das caixas que transporta habitualmente. Sabe-se que as caixas têm um peso médio de 10 quilos com um desvio padrão de 3 kg. O volume médio é de 30 dm³ com um desvio padrão de 6 dm³. Ajude a transportadora neste estudo.
5. Compare a variabilidade do peso dos patos adultos e dos elefantes asiáticos adultos sabendo que os patos de uma determinada espécie de têm um peso medio de 4 kg com desvio padrão de 1,2 e que os elefantes asiáticos pesam em média 4200 kg com um desvio padrão de 400 kg.
6. Dentro de uma determinada gama de carros, pretende-se estudar a relação entre a quilometragem de um carro usado e o seu preço de venda. Para tal recolheu-se uma amostra de 14 observações tendo-se registado a quilometragem e o preço de venda.

Carros	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Quilometragem (1000 Km)	40	30	30	25	50	60	65	10	15	20	55	40	35	30	505
Preço de venda(x10 €)	1000	1500	1200	1800	800	1000	500	3000	2500	2000	800	1500	2000	2000	21600

- a) Construa o diagrama de dispersão para os dados. Acha plausível uma relação linear entre as duas variáveis? Justifique.
- b) Utilizando o método dos mínimos quadrados determine a equação de regressão estimada e interprete-a.
- c) Recorrendo ao modelo de regressão estime o preço médio de um carro usado com 20 000km.
- d) Recorrendo ao modelo de regressão estime o preço médio de um carro usado com 140 000km. Comente o resultado obtido.
- e) Determine o coeficiente de determinação e interprete o seu resultado.
7. Recorde o exemplo 1 desta ficha de trabalho onde se selecionaram-se aleatoriamente alunos de uma determinada escola de música. Pretende-se agora perceber se existe alguma relação entre o peso dos alunos e a sua altura.
8. Resolva o Exercício 2 dos slides de regressão.

Cofinanciado por:



UNIÃO EUROPEIA
Fundo Social Europeu

9. O Ipma - Instituto Português do Mar e da Atmosfera disponibiliza os dados de precipitação de Portugal desde 1950. No ficheiro `Precipitação_jan_Abril_ano 2000 e 2001.xlsx` encontram-se registos de precipitação dos meses de janeiro e abril dos anos 2000 e 2001. Os registos contêm valores registados nos diferentes pontos de observação cuja latitude, longitude é indicada.

10. (exercício prova 2016/17)

No ficheiro “dados_prova1”, a folha `Pordata`, contêm informação oficial retirada de <http://www.pordata.pt/> onde se registam os dias sem chuva entre o ano 2000 e 2015 para a região de Lisboa e Porto assim como o número e acidentes de viação com vítimas.

a) Compare as regiões de Lisboa e Porto relativamente ao número de dias sem chuva dos últimos 15 anos, calculando com auxílio do Excel:

i. A média e mediana para as duas regiões. Interprete os valores obtidos dizendo qual a zona que lhe parece ser mais seca.

ii. Esboce o diagrama de extremos e quartis, também denominado *box-plot* para o número de dias sem chuva da região de Lisboa. Indique claramente todas as medidas que teve de calcular.

iii. Complete a seguinte afirmação: “Nos últimos 15 anos, cerca de 25% dos anos o número de dias sem chuva da região de Lisboa foi superior ou igual a _____”.

iv. Sem recorrer ao Excel determine o terceiro quartil e mediana dos dias sem chuva para os registos do Porto. Apresente os cálculos e interprete o valor obtidos.

b) Calcule medidas de dispersão que permitam comparar a variabilidade dos dias sem chuva da região de Lisboa com os acidentes de viação nacionais dos últimos 15 anos. Qual das situações com maior variabilidade?

11. (exercício prova 2016/17)

No ficheiro “dados_prova1”, os dados da folha `PMEEExcelência 2014` contêm informação sobre as empresas nacionais de excelência registadas em 2014 (retirado de www.iapmei.pt/resources/download/PMEEExcelencia2014.xls).

Construa uma tabela (ou várias tabelas) para a informação por distrito e setor de atividade de modo a obter a extrair as seguintes informações:

i. Distritos com maior e menor percentagem empresas registados e indique as percentagens.

ii. Construa um gráfico que lhe permita visualizar a informação relativa às percentagens das empresas por distrito de modo a ficar bem evidente a resposta da alínea anterior.

iii. Indique o sector de atividade mais frequente. Qual foi a sua percentagem face ao total nacional?

iv. Complete a seguinte afirmação: O distrito com maior número de empresas no sector de atividade “Comercio” é _____ que corresponde a _____% de todas as empresas de comercio nacional.

Cofinanciado por:

12. (exercício prova 2016/17)

A tabela seguinte indica as idades (X) e os gastos semanais em Euros (Y) de 10 pessoas em produtos culturais.

Idade (X)	16	22	22	28	30	31	36	44	55	60
Gastos semanais (Y)	5	8	10	11	12	12	15	20	22	21

1. Copie a tabela para uma folha de cálculo e faça o gráfico – diagrama de dispersão onde se tenta obter a relação entre os gastos em produtos culturais em função da idade. No gráfico adicione a linha de tendência que melhor se ajusta aos dados, com a respectiva expressão e o valor de r^2 . Identifique claramente os eixos através de legenda adequada. Copie o gráfico para o espaço que se segue abaixo:

2. Complete as seguintes afirmações, apresentando os cálculos que entender necessários:

A percentagem de variação dos gastos semanais em produtos culturais que pode ser explicada pela idade é de _____%.

Recorrendo ao modelo obtido, espera-se que um indivíduo com 20 anos gaste semanalmente _____ euros em produtos culturais.

3. Recorrendo à regressão linear, obtivemos a equação $Y=0,3737+0,3845X$

Interprete significado do valor 0,3845

13 (exercício prova 2016/17) O ficheiro “dados_prova2”, a folha Pordata, contém informação oficial retirada de <http://www.pordata.pt/> onde se registam os dias sem chuva entre o ano 2000 e 2015 para a região de Lisboa e Porto assim como o número e acidentes de viação com vítimas.

a) Pretende-se estudar a evolução do número de acidentes com vítimas ao longo dos anos, para tal faça o gráfico – diagrama de dispersão- onde se tenta obter a relação entre número de acidentes com vítimas em função dos anos. No gráfico adicione a linha de tendência que melhor se ajusta aos dados, com a respetiva expressão e o valor de r^2 . Identifique claramente os eixos através de legenda adequada. Copie o gráfico para o espaço que se segue abaixo:

b) Complete as seguintes afirmações, apresentando os cálculos que entender necessários:

A percentagem de variação do número de acidentes com vítimas que pode ser explicada pelos anos é de _____%.

Recorrendo ao modelo obtido, estime o número de acidentes com vítimas de 2016.

c) Para estes dados, recorrendo à regressão linear, obtivemos a equação $Y=1\,891\,690-924X$

Interprete significado do valor -924.

Cofinanciado por:



UNIÃO EUROPEIA
Fundo Social Europeu

d) Complete os seguintes espaços que visam a comparar as regiões de Lisboa e Porto relativamente ao número de dias sem chuva dos últimos 15 anos:

- Em média, nos anos em análise, podemos afirmar que a cidade mais seca é _____ pois os valores médios número de dias sem chuva são de _____ para Lisboa e _____ para o Porto.

- Nos últimos 15 anos, cerca de 50% dos anos o número de dias sem chuva da região de Lisboa foi superior ou igual a _____, embora para o Porto cerca de 25% dos últimos anos o número de dias sem chuva foi inferior ou igual a _____.

Existe maior variabilidade no número de dias sem chuva para a região de _____ (Lisboa/Porto), pois a medida de variabilidade que compara os desvios em relação à média, denominada _____ indica que para o porto o valor é de _____ e para Lisboa é de _____.

e) Sem recorrer ao Excel determine o primeiro quartil e mediana dos dias sem chuva para os registos do Lisboa. Apresente os cálculos e interprete o valor obtidos.

14 -(exercício prova 2016/17) No ficheiro “dados_prova2”, a folha “cursos superiores” contem informações dos cursos superiores, obtidas em <http://infocursos.mec.pt/bds.asp> sobre classificações finais médias dos alunos relativamente aos cursos superiores existentes em Portugal em 2016.

a) Construa uma tabela e um histograma para a nota final média. As classes do histograma deverão ter classes de amplitude igual a 1 valor. Legende adequadamente o histograma. Copie o gráfico para o espaço abaixo.

b) Relativamente ao gráfico acima identifique a moda e interprete o seu valor.

c) Construa uma tabela e dupla entrada relativamente à natureza e tipo de ensino. Qual o tipo de ensino mais frequente em Portugal, Universitário ou Politécnico. Indique as percentagens de cada natureza.

d) Dentro do ensino publico, existe mais ensino politécnico ou universitário e indique as respetivas percentagens de cada uma das naturezas de ensino.

Cofinanciado por:



UNIÃO EUROPEIA
Fundo Social Europeu