

M.A.C.S. (10.º ano)

Análise de dados bivariados

Exercícios de Provas Nacionais - Propostas de resolução



1. Da observação do diagrama de dispersão e como a associação linear é positiva, podemos verificar que:

- o coeficiente de correlação também é positivo, ou seja, $r > 0$, pelo que apenas as opções (C) e (D) podem corresponder à distribuição apresentada;
- o declive da reta de regressão também é positiva, pelo que apenas as opções (A) e (D) podem corresponder à distribuição apresentada;

Resposta: **Opção D**

Exame – 2022, 1ª Fase

2. Inserindo na calculadora gráfica as listas com os dados apresentados, temos:

x	y
30	0,5
24	1
22,5	2
18	4
12	8
8	9
3	12

Determinando a equação da reta de regressão ($y = ax + b$), temos que os valores de a e b , usando valores aproximados com três casas decimais, são $a \approx -0,459$ e $b \approx 12,913$.

Desta forma a equação da reta de regressão é $y = -0,459x + 12,913$, pelo que uma pessoa que tenha estado seis horas na fila antes da abertura da bilheteira ($x = 6$), terá tido registado um tempo decorrido desde a abertura da bilheteira até à aquisição dos bilhetes (y), arredondado às unidades, de:

$$y \approx -0,459 \times 6 + 12,913 \approx 10 \text{ horas}$$

Exame – 2020, 1.ª Fase

3. Inserindo na calculadora gráfica as listas com os dados apresentados, temos:

Custo x	Espectadores y
435	99
379	84
65	16
60	13
276	75
59	12
43	9

Determinando a equação da reta de regressão, temos que os valores de a e b , usando valores aproximados com três casas decimais, são $a \approx 0,233$ e $b \approx 0,238$.

Desta forma a equação da reta de regressão é $y = 0,233x + 0,238$, e como 52,5 milhares de espectadores corresponde a um valor de $y = 52,5$, a estimativa do custo de produção associado é a solução da equação $52,5 = 0,233x + 0,238$. Resolvendo a equação e aproximando a solução às unidades, temos:

$$52,5 = 0,233x + 0,238 \Leftrightarrow 52,5 - 0,238 = 0,233x \Leftrightarrow \frac{52,262}{0,233} = x \Leftrightarrow x \approx 224$$

Assim a estimativa do custo de produção de um filme com 52,5 milhares de espectadores, é de 224 milhões de euros.

Exame – 2017, Ép. especial

4.

4.1. Determinando a frequência cardíaca do atleta vencedor em 2012, ou seja, o valor de P , sabendo que a média dos valores da tabela é 166,5, temos:

$$\frac{165 + 166 + 166 + 168 + 170 + 170 + P + 160 + 160 + 168}{10} = 166,5 \Leftrightarrow \frac{1493 + P}{10} = 166,5 \Leftrightarrow$$

$$\Leftrightarrow 1493 + P = 166,5 \times 10 \Leftrightarrow 1493 + P = 1665 \Leftrightarrow P = 1665 - 1493 \Leftrightarrow P = 172$$

Desta forma podemos constatar que a primeira afirmação é falsa, porque o atleta vencedor em 2012 terminou a maratona com uma frequência cardíaca superior a 171 pulsações por minuto.

Considerando o valor de $P = 172$ e ordenando os dados da tabela podemos verificar que os valores centrais, são 166 e 168:

$$\underbrace{160 \ 160 \ 165 \ 166 \ 166}_{50\%} \quad \underbrace{168 \ 168 \ 170 \ 170 \ 172}_{50\%}$$

Assim, temos que a mediana, \tilde{x} , das frequências cardíacas é: $\tilde{x} = \frac{166 + 168}{2} = 167$

Logo a segunda afirmação é falsa, porque a mediana não é 170.

Observando o diagrama de dispersão, podemos verificar que uma variação dos valores da temperatura ambiente estão corresponde a uma variação semelhante dos valores da frequência cardíaca dos atletas, ou seja, o aumento da temperatura está associado a um aumento da frequência cardíaca, o que representa uma correlação linear positiva, pelo que o respetivo coeficiente de correlação é positivo.

Assim, a terceira afirmação é falsa, porque o coeficiente de correlação linear não pode ser negativo.



- 4.2. De acordo com a reta ajustada ao diagrama de dispersão, a temperatura ambiente registada no final da maratona, no ano de 2006, é o valor de x correspondente ao valor de $y = 165$, ou seja a frequência cardíaca do atleta vencedor nesse ano.

Assim, substituindo o valor de y na equação da reta e determinando o valor corresponde de x , temos:

$$165 = 0,71x + 147,1 \Leftrightarrow 165 - 147,1 = 0,71x \Leftrightarrow 17,9 = 0,71x \Leftrightarrow \frac{17,9}{0,71} = x$$

Como $\frac{17,9}{0,71} \approx 25,2$ podemos concluir que, de acordo com a reta ajustada do diagrama, $31,7^\circ\text{C}$ não é um valor admissível da temperatura ambiente registada no final da maratona, no ano de 2006, por ser bastante diferente do valor esperado ($25,2^\circ\text{C}$).

Exame – 2016, 2ª Fase

5.

- 5.1. Inserindo na calculadora gráfica as listas com os dados apresentados, temos:

Ano (x)	N.º de sócios (y)
0	10 585
10	23 150
20	30 981
30	39 125

Determinando a equação da reta de regressão, temos que os valores de a e b , usando valores aproximados com uma casa decimal, são $a \approx 934,5$ e $b \approx 11\,942,6$.

Desta forma a equação da reta de regressão é $y = 934,5x + 11\,942,6$, a que corresponde o modelo $N(t) = 934,5t + 11\,942,6$ e como o final de 2005 corresponde à passagem de 25 anos após o final de 1980, o número de sócios estimado, de acordo com o modelo, é:

$$N(25) = 934,5 \times 25 + 11\,942,6 \approx 35305$$

- 5.2. A opção correta é a Opção II (porque o diagrama evidencia uma correlação negativa forte, o declive da reta é negativo e a ordenada da origem é um valor compreendido entre 60 000 e 70 000).

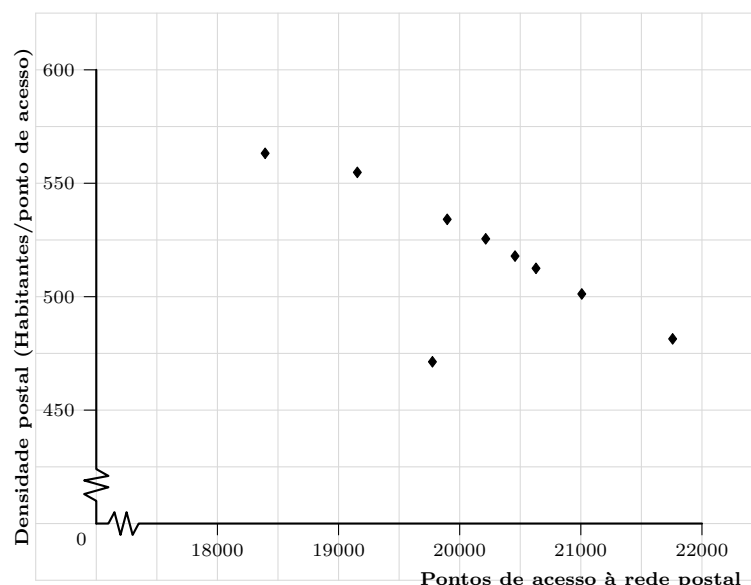
A opção I não pode estar associada ao diagrama de dispersão porque a reta de regressão deverá ter declive negativo (o aumento do número de sócios está associado a uma diminuição do número de lugares por vender) e na opção I, o valor indicado para o declive é positivo ($a = 1,744$).

A opção III não pode estar associada ao diagrama de dispersão porque o diagrama de dispersão sugere a existência de uma correlação negativa forte, ou seja um valor do coeficiente de correlação próximo de -1 (os pontos estão dispostos num alinhamento próximo de uma reta de declive negativo) e na opção III, o valor indicado para o coeficiente de correlação indica a existência de uma correlação quase nula, ou seja, um valor próximo de zero ($r = -0,087$).

Exame – 2015, Ép. especial



6. Inserindo os dados relativos aos pontos de acesso da rede postal e da densidade postal em duas listas da calculadora gráfica, podemos obter o diagrama de dispersão seguinte:



Usando os mesmos dados, inseridos na calculadora, ou seja:

Pontos de acesso (x)	Densidade postal (y)
19 775	471,3
21 758	481,4
21 008	501,2
20 630	512,5
20 457	517,9
20 215	525,5
19 897	534,1
19 155	554,8
18 394	563,2

Podemos determinar o valor do coeficiente de correlação linear, incluindo os dados relativos ao ano 2001, cujo valor, arredondado às milésimas, é:

$$r \approx -0,728$$

Se eliminarmos os dados da primeira linha da tabela, ou seja os dados referentes a 2001, e calcularmos novamente o valor do coeficiente de correlação linear, este valor, arredondado às milésimas, é alterado para:

$$r \approx -0,992$$

Logo, podemos verificar que a supressão destes dados implica um maior ajustamento da reta de regressão à nuvem de pontos e uma alteração no valor do coeficiente de correlação linear, indicando uma correlação mais forte, ou seja, o seu valor aproxima-se de -1 .

Assim, podemos concluir que a exclusão do *outlier* permite a obtenção de uma reta de regressão mais ajustada aos dados restantes. Desta forma, a reta de regressão permite a obtenção de previsões mais ajustadas (e por isso, mais fiáveis) aos dados que são tomados em consideração.



7.

- 7.1. Inserindo na calculadora gráfica os dados relativos à Classificação interna numa lista, e os dados relativos à classificação externa noutra lista, e determinando o coeficiente de correlação linear, referente a estas duas listas, com aproximação às centésimas, obtemos o valor: $r \approx 0,439$

Apagando os registos relativos ao aluno número 14, ou seja, o valor 19 da primeira lista e o valor 4,2 da segunda lista, e voltando a determinar o coeficiente de correlação linear referente a estas duas listas, com aproximação às centésimas, obtemos o valor: $r \approx 0,913$

Assim podemos concluir que a aluna tem razão, porque o coeficiente de correlação linear obtido quando consideramos as classificações do aluno número 14, evidencia uma correlação fraca, e o coeficiente de correlação linear obtido com a supressão destes dados evidencia uma correlação forte.

- 7.2. A estimativa do valor da classificação no exame da disciplina de Matemática Aplicada às Ciências Sociais de um aluno da escola de Xisto que, no final do 3.º período de 2010, tenha obtido a classificação de 12, obtida a partir da reta de regressão, é o valor de y associado a $x = 12$.

Desta forma, substituindo o valor de x e calculando o valor de y , temos:

$$y = 1,0927 \times 12 - 1,8476 \Leftrightarrow y = 11,2648$$

Assim, a estimativa do valor da classificação de exame deste aluno, com arredondamento às décimas, é 11,3 valores.

Exame – 2012, 2ª Fase

8.

- 8.1. Inserindo na calculadora gráfica as listas com os dados relativos à TAA (x) e ao PIB (y), temos:

TAA (x)	PIB (y)
15,4	1839
74,8	7932
43,5	4251
17,8	2586
11,5	1524
89,6	12 674
61,2	6275

Desta forma, determinando a equação da reta de regressão, temos que os valores de a , de b e de r , usando valores aproximados com quatro casas decimais, são:

- $a \approx 123,7159$
- $b \approx -248,7235$
- $r \approx 0,9634$

- 8.2. Se eliminarmos os dados referentes ao país F, o valor do coeficiente de correlação linear ($r = 0,9937$) é maior do que se incluirmos os dados do país F ($r \approx 0,9634$).

Logo, podemos verificar que a supressão destes dados implica um maior ajustamento da reta de regressão à nuvem de pontos e uma alteração no valor do coeficiente de correlação linear, indicando uma correlação mais forte, ou seja, o seu valor aproxima-se de 1.

Assim, podemos concluir que a exclusão do *outlier* permite a obtenção de uma reta de regressão mais ajustada aos dados restantes. Desta forma, a reta de regressão permite a obtenção de previsões mais ajustadas (e por isso, mais fiáveis) aos dados que são tomados em consideração.

Exame – 2010, 2ª Fase



9.

9.1. Inserindo na calculadora gráfica as listas com os dados relativos ao rendimento mensal (x) e às despesas com a alimentação (y), temos:

Rendimento mensal (€) x	Despesas com a alimentação (€) y
1250	425
2800	600
1900	550
1650	425
1300	400
1800	575
1200	375
2500	600
1350	375
2100	550
1200	350
1500	400

Determinando os valores associados à correlação linear, temos que:

- o valor do coeficiente de correlação, com arredondamento às décimas é $r \approx 0,9$
- o que significa que a associação linear que é possível identificar entre as duas variáveis é do positiva e forte;
- podemos identificar este tipo de associação linear entre as duas variáveis porque o valor do coeficiente de correlação é positivo ($r > 0$) e é um valor próximo de 1, relativamente distante de zero, o que justifica o facto de ser uma associação forte.

9.2.

- Usando a correlação linear determinada no item anterior podemos observar ainda que os valores de a e b , arredondados com quatro casas decimais, são:
 - $a \approx 0,1656$
 - $b \approx 185,1833$
- Usando os valores do item anterior, e a equação $y = ax + b$ da reta de regressão que se ajusta ao diagrama de dispersão, ou seja, $y = 0,1656x + 185,1833$, temos que a estimativa do valor do valor das despesas mensais com a alimentação de um agregado familiar cujo rendimento mensal é de €1750, é o valor de y correspondente ao valor $x = 1750$

Desta forma, substituindo o valor de x e calculando o valor de y , temos:

$$y = 0,1656 \times 1750 + 185,1833 \Leftrightarrow y = 474,9833$$

Assim, a estimativa, em euros, com arredondamento às unidades, do valor das despesas mensais com a alimentação para o agregado familiar nas condições do problema, é de €475.

Exame – 2009, 1ª Fase



10. Inserindo na calculadora gráfica as listas com os dados relativos ao número de anos decorridos desde a criação da área protegida (x) e o número de animais existentes nessa área (y), temos:

Anos decorridos desde a criação da área protegida (x)	Número de animais existentes na área protegida (y)
0	8
1	9
2	13
3	18
4	24
5	30
6	38
7	45
8	62
9	75
10	84
11	88
12	100

Determinando a equação da reta de regressão, temos que os valores de a e b , com uma aproximação às décimas, são $a \approx 8,2$ e $b \approx -3,5$, pelo que o modelo de regressão linear, de equação $y = ax + b$, que se ajusta à nuvem de pontos apresentada, é:

$$y = 8,2x - 3,5$$

Exame – 2008, 2ª Fase



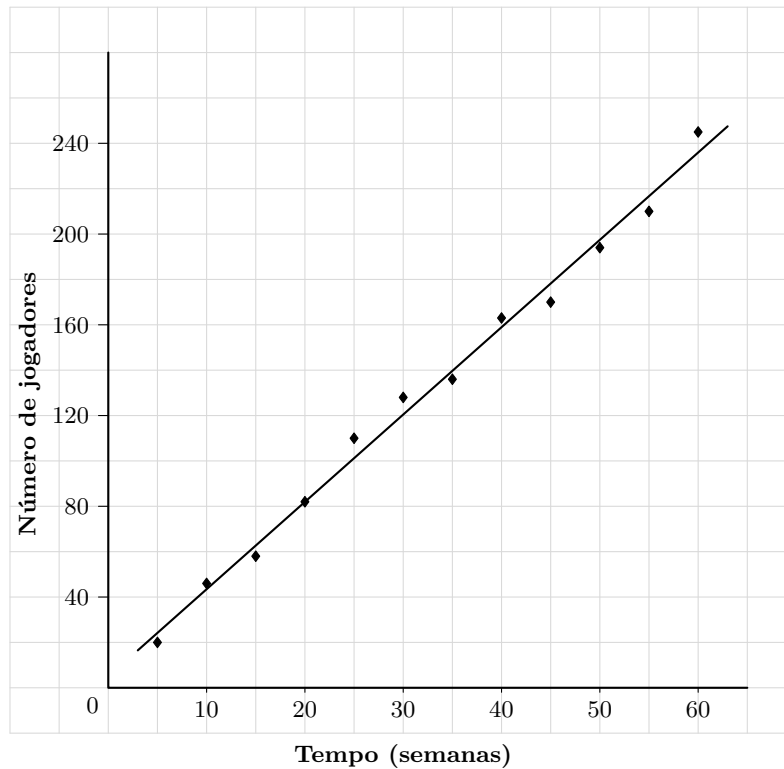
11. Inserindo na calculadora gráfica as listas com os dados relativos ao tempo decorrido desde o lançamento do site (x) e o número de jogadores (y), temos:

Tempo (em semanas) (x)	Número de jogadores (em milhares) (y)
5	20
10	46
15	58
20	82
25	110
30	128
35	136
40	163
45	170
50	194
55	210
60	245

Determinando a equação da reta de regressão, temos que os valores de a e b , com uma aproximação às centésimas, são $a \approx 3,85$ e $b \approx 4,94$, pelo que equação da reta de regressão linear, é:

$$y = 3,85x + 4,94$$

Visualizando o diagrama de dispersão e a reta de regressão linear, obtemos os seguintes gráficos:



12.

- 12.1. Inserindo na calculadora gráfica os valores dos anos numa lista e os valores correspondentes da população (em milhares) noutra lista, e determinando os valores associados à correlação linear, temos que o valor do coeficiente de correlação, arredondado às milésimas é $r \approx 0,988$

Como se pode observar no diagrama de dispersão apresentado, a reta de regressão é um modelo bem ajustado ao conjunto dos pontos que representam os dados da tabela. Este grau de ajustamento é confirmado pelo valor do coeficiente de correlação, que é próximo do valor 1, o que traduziria um ajustamento perfeito.

- 12.2. Apesar do modelo linear apresentado (reta de regressão) se ajustar bem ao conjunto dos pontos que representam os dados da tabela, podemos verificar que não é um modelo ajustado para:

- estimar o número de habitantes, em Portugal, há alguns séculos (três ou mais), porque neste caso, o modelo iria prever valores negativos para a população, o que não faria sentido no contexto da situação em estudo.

Por exemplo, para o valor de $a = 1700$, correspondente a, aproximadamente três séculos atrás, o valor da população correspondente, segundo o modelo apresentado, é:

$$p = 0,047 \times 1700 - 84,95 = -3,86 \text{ milhões}$$

- prever a evolução da população portuguesa, a muito longo prazo, porque neste caso, o modelo prevê um aumento ilimitado da população, o que não seria sustentável porque os recursos necessários para sustentar a população teriam que aumentar de forma também ilimitada, o que é impossível, porque estes recursos são, comprovadamente, finitos e limitados.

Por exemplo, para o valor de $a = 2100$, correspondente a, aproximadamente um séculos no futuro, o valor da população correspondente, segundo o modelo apresentado, é:

$$p = 0,047 \times 2100 - 84,95 = 15,52 \text{ milhões}$$

Neste caso significaria que a população teria um aumento de, aproximadamente 50% relativamente ao valor atual, sem que seja razoável assumir que os recursos disponíveis possam crescer na mesma proporção.

Exame – 2006, 2ª Fase

