

GTA | Guião de Trabalho Autónomo n.º 16

MATEMÁTICA A 10.º ANO

Tema 2: Estatística Subtema 2: Dados bivariados



PORQUÊ APRENDER SOBRE...?



O QUE VOU APRENDER?



COMO VOU APRENDER?



O QUE APRENDI?



COMO POSSO COMPLEMENTAR A
APRENDIZAGEM?



PORQUÊ APRENDER SOBRE...?

Dados bivariados

A estatística está em toda a parte, desde as páginas dos jornais até às discussões desportivas, passando pelas redes sociais ou as informações oficiais.

Contudo os números não falam sozinhos, temos de os *interpretar* e de ser capazes de perceber se as interpretações que os outros fizeram estão corretas ou se são realmente conclusões do estudo em vez de *fake news* ou opiniões dos responsáveis pela divulgação do estudo.

Por vezes, os dados resultam da observação de duas variáveis sobre cada unidade estatística selecionada para a amostra. Caso se pretendam estudar as duas características em conjunto, os dados aparecem sob a forma de pares de valores, isto é, cada indivíduo ou resultado experimental, contribui com um conjunto de dois valores.



O QUE VOU APRENDER?

- Representar os dados no diagrama de dispersão para interpretar a forma, direção e força da associação (linear) entre as duas variáveis.
- Obter o coeficiente de correlação linear, conhecer as suas propriedades e interpretá-lo geometricamente.
- Obter a equação da reta de regressão e o coeficiente de correlação e inferir o valor de y para um dado valor de x .



COMO VOU APRENDER?

GTA 16: Dados bivariados

GTA 17: Resolução de problemas sobre dados univariados e bivariados

Tema 2: Estatística

Subtema 2: Dados bivariados



GTA 16: Dados bivariados

Objetivo:

- Reconhecer que, para estudar a associação entre duas variáveis quantitativas de uma população, se observam essas variáveis sobre cada unidade estatística, obtendo-se uma amostra de pares de dados.
- Reconhecer a importância da representação dos dados no diagrama de dispersão, nuvem de pontos, para interpretar a forma, direção e força da associação (linear) entre as duas variáveis.

Modalidade de trabalho: individual, pares ou pequenos grupos.

Recursos e materiais: caderno diário, capítulo “Estatística (2.ª parte)” do **Manual NiuAleph**, manual escolar, calculadora gráfica ou folha de cálculo e *internet*.

TAREFA 1

Abre o teu manual escolar no tema “Estatística” e o teu caderno e **responde** às seguintes questões:

- O que é e como se constrói um diagrama de dispersão?
- O que é o coeficiente de correlação linear, quais são as suas propriedades e como se interpreta geometricamente?
- Como se pode obter a equação da reta de regressão e o coeficiente de correlação? E como se pode inferir o valor de y para um dado valor de x ?

TAREFA 2

Já sabes que o diagrama de dispersão é uma representação gráfica para os dados bivariados quantitativos, em que cada par de dados (x_i, y_i) é representado por um ponto de coordenadas (x_i, y_i) num sistema de eixos coordenados.

Este tipo de representação é muito útil, pois permite realçar algumas propriedades entre os dados, nomeadamente no que diz respeito ao tipo de associação entre os valores da variável independente e da variável dependente.

Responde agora às questões que se seguem. Depois, **compara** a tua resposta com as dos teus colegas.



A maratona é a última prova em diversos encontros desportivos. Existem diversos estudos estatísticos sobre esta prova, que envolvem dados como o tempo de conclusão da prova, a frequência cardíaca dos atletas no final da prova ou as condições climatéricas.

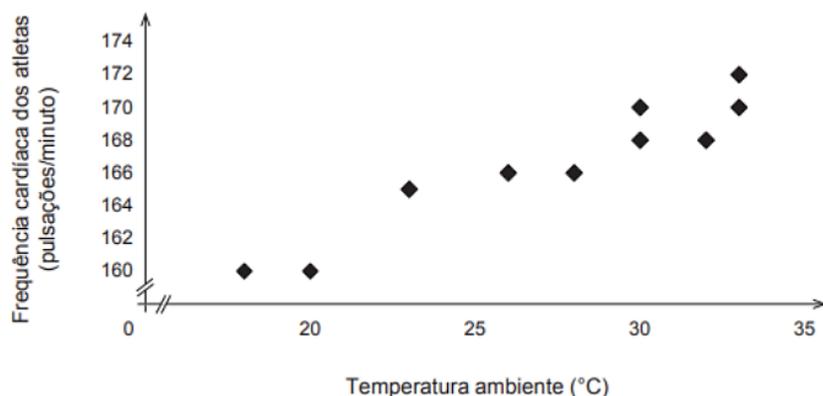
Na tabela seguinte, estão parcialmente registados os dados referentes à frequência cardíaca do atleta vencedor da maratona, no momento em que acaba a prova, nas edições do Encontro Desportivo Internacional dos últimos dez anos.

Sabe-se ainda que:

- a média dos valores constantes da tabela seguinte é 166,5;
- P representa as pulsações por minuto do atleta que venceu a maratona na edição de 2012.

Ano	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Frequência cardíaca (pulsações/minuto)	165	166	166	168	170	170	P	160	160	168

Na figura seguinte, está representado o diagrama de dispersão da temperatura ambiente, em graus Celsius ($^{\circ}\text{C}$), e da frequência cardíaca, em pulsações por minuto, dos atletas referidos na tabela anterior, no final da prova.



1. Com base no diagrama de dispersão, **o que se pode inferir** sobre a frequência cardíaca dos atletas à medida que aumenta a temperatura ambiente?

2. **Considera** as seguintes afirmações.

- O atleta vencedor em 2012 terminou a maratona com uma frequência cardíaca entre 169 e 171 pulsações por minuto.
- A mediana das frequências cardíacas é 170 pulsações por minuto.
- O coeficiente de correlação linear entre a temperatura ambiente e a frequência cardíaca dos atletas pode ser $-0,85$.

Elabora uma pequena composição na qual justifiques, com base nos dados apresentados, que as três afirmações são falsas.

Adaptado de: Exame de Matemática Aplicada às Ciências Sociais, 2.ª fase, 2016, IAVE



TAREFA 3

Habitualmente, um surto de gripe faz aumentar a procura dos serviços prestadores de cuidados de saúde.

Na tabela seguinte, apresentam-se os registos do número, x , de pessoas infetadas por vírus de gripe, e do respetivo número, y , de hospitalizações associadas a gripe, em diferentes momentos, numa certa região.

Número de pessoas infetadas (x)	31	857	11 973	12 000	86 000	86 123	320 083	616 545
Número de hospitalizações (y)	6	117	1000	1100	5426	6000	13 848	18 339

Admite um modelo em que a associação entre as variáveis número de pessoas infetadas por vírus de gripe (x) e número de hospitalizações (y) é, aproximadamente, linear.

Considera um modelo de regressão linear da forma $y = ax + b$, obtido a partir dos registos apresentados na tabela.

1. **Determina** os valores de a , de b e do coeficiente de correlação linear, r , recorrendo às ferramentas de estatística da calculadora.

Apresenta os valores pedidos com quatro casas decimais.

2. **Estima**, com base nesse modelo, o número de hospitalizações associadas a gripe, nessa região, caso sejam registados 250 000 infetados por vírus de gripe.

Apresenta o resultado final arredondado às unidades de milhar.

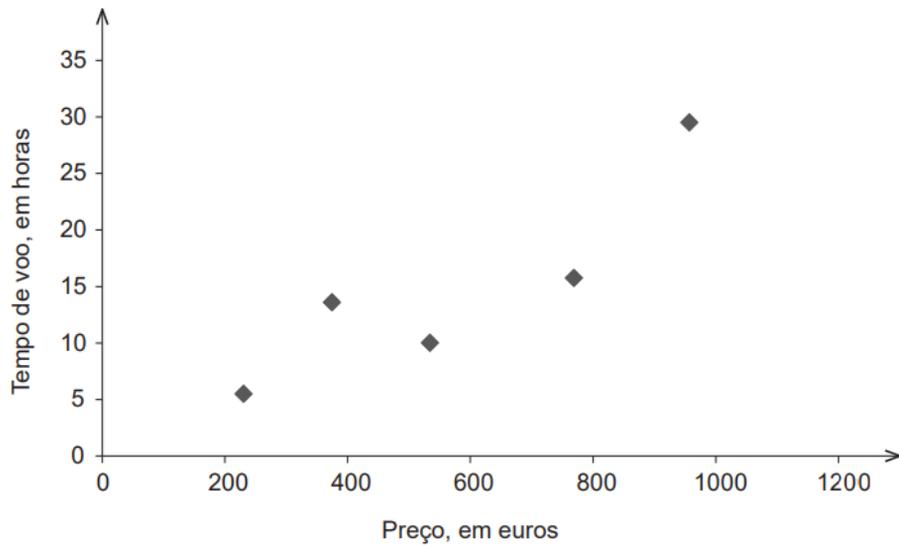
Adaptado de: Exame de Matemática B, 2.ª fase, 2016, IAVE

TAREFA 4

Autoavalia a tua aprendizagem, respondendo às questões seguintes.

Apresenta o teu raciocínio e **registra** as tuas dúvidas. No final, **compara** as tuas respostas com as dos teus colegas.

O diagrama de dispersão representado na figura seguinte mostra uma forte associação linear positiva entre o preço das viagens de avião vendidas na agência de viagens Ir&Voltar, num determinado período de tempo, e as horas de voo despendidas na viagem de ida e volta.



Em cada uma das opções seguintes, são dados um valor de r , coeficiente de correlação linear, e a equação de uma reta.

1. **Em qual** das opções poderão estar representados o valor de r e uma equação da reta de regressão linear da distribuição representada na figura anterior? **Explica** como pensaste.

(A) $r = -0,92$
 $y = 0,03x - 0,86$

(B) $r = -0,2$
 $y = -0,03x - 0,86$

(C) $r = 0,92$
 $y = -0,03x - 0,86$

(D) $r = 0,89$
 $y = 0,03x - 0,86$

2. De acordo com a opção escolhida em 1., e considerando y o tempo de voo, em horas, e x o preço das viagens, em euros, **estima**, o tempo de voo, em horas, para uma viagem vendida por 500 euros.

Apresenta o resultado final em horas e minutos.

Adaptado de: Exame de Matemática Aplicada às Ciências Sociais, 1.ª fase, 2022, IAVE



PROPOSTA DE RESOLUÇÃO

Consulta o manual escolar e **segue** as instruções sobre como podes usar a calculadora gráfica para responder a esta tarefa e às seguintes.

TAREFA 1

1. Pode-se inferir que à medida que a temperatura ambiente aumenta, aumenta também a frequência cardíaca dos atletas, em pulsações por minuto.
2. Começamos por determinar a frequência cardíaca do atleta vencedor em 2012, ou seja, o valor de P , sabendo que a média é 166,5:

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

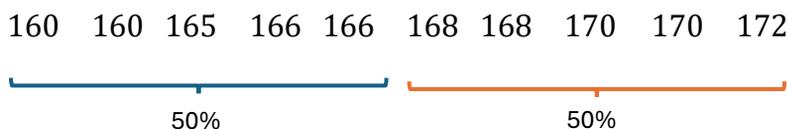
$$\Leftrightarrow \frac{11493 + P}{10} = 166,5$$

$$\Leftrightarrow 11493 + P = 1665$$

$$\Leftrightarrow P = 172$$

Assim, **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:



Como a mediana das frequências cardíacas é $\frac{166+168}{2} = 167$, **a segunda afirmação é falsa**.

Observando o diagrama de dispersão, verificamos que a uma variação dos valores da temperatura ambiente está associada a uma variação dos valores da frequência cardíaca dos atletas. A um aumento da temperatura está associado 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.

Adaptado de https://mat.absolutamente.net/compilacoes/mac3/10/dados_bivariados_resol.pdf



TAREFA 3

1. Depois de inserir na calculadora gráfica as listas com os dados relativos ao número de pessoas infetadas pelo vírus de gripe (x) e ao número de hospitalizações (y), determinamos os valores de a , de b e do coeficiente de correlação linear, r . Na Imagem 1, podes ver o ecrã de uma calculadora gráfica, com os valores pedidos:

- $a \approx 0,0304$
- $b \approx 1418,9582$
- $r \approx 0,9676$

	A pess...	B hosp...	C	D	E
=				=LinRegV	
2	857	117	RegEq...	m*x+b	
3	11973	1000	m	0.03042	
4	12000	1100	b	1418.96	
5	86000	5426	r ²	0.93624	
6	86123	6000	r	0.967595	
D6	=0.96759506051985				

Imagem 1 – Lista de dados

Repara que, como $r > 0$, existe uma associação linear positiva entre as variáveis. Já sabes que quanto maior for o valor absoluto de r , maior é a correlação. Assim, no que diz respeito às variáveis número de pessoas infetadas pelo vírus de gripe e número de hospitalizações, o facto de o coeficiente de correlação ser aproximadamente 0,9676 mostra que essas variáveis estão fortemente correlacionadas.

2. Usando os valores do item anterior, a equação da reta de regressão linear é

$$y = 0,0304x + 1418,9582$$

Substituindo na equação x por 250000 e arredondando às unidades de milhar,

$$y = 0,0304 \times 250000 + 1418,9582 = 9000$$

obtém-se, assim, uma estimativa de 9000 hospitalizações associadas à gripe, nessa região.



TAREFA 4

1. Opção D.

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 positivo, pelo que apenas as opções (A) e (D) podem corresponder à distribuição apresentada.

https://mat.absolutamente.net/compilacoes/mac3/10/dados_bivariados_resol.pdf

2. Usando os valores do item anterior, a equação da reta de regressão linear é

$$y = 0,03x - 0,86$$

Substituindo na equação x por 500, temos:

$$y = 0,03 \times 500 - 0,86 = 14,14 \text{ e } 0,14 \times 60 \approx 8 \text{ minutos}$$

Obtém-se, assim, uma estimativa de 14 horas 8 minutos de duração de uma viagem vendida por 500 euros pela agência Ir&Voltar.



O QUE APRENDI?

Já sabes interpretar e determinar as medidas de localização e de dispersão e interpretar propriedades dessas medidas?

És capaz de ...

- estudar a associação entre duas variáveis quantitativas de uma população, observando essas variáveis sobre cada unidade estatística, obtendo-se uma amostra de pares de dados?
- representar os dados no diagrama de dispersão para interpretar a forma, direção e força da associação (linear) entre as duas variáveis?
- obter o coeficiente de correlação linear, conhecer as suas propriedades e interpretá-lo geometricamente?
- obter a equação da reta de regressão e o coeficiente de correlação e inferir o valor de y para um dado valor de x ?

Procura no teu manual escolar os exercícios resolvidos sobre o tema “Estatística”. **Analisa-os** e **resolve-os** sozinho. Por fim, **compara** a tua resolução com a do manual e com as dos teus colegas.

Estuda, com um colega de turma, para consolidares a tua aprendizagem.



COMO POSSO COMPLEMENTAR A APRENDIZAGEM?

Se quiseres saber mais, podes consultar:

INE

https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_main



PORDATA

<https://www.pordata.pt/>



Para saber mais sobre medidas de dispersão:

Noções de Estatística, VI. Distribuições bidimensionais. ALEA



[Reta de regressão](#)



[Distribuições bidimensionais](#)