

1. Um artigo da revista *Wear* (1992) apresenta dados relativos à viscosidade do óleo e ao desgaste do aço macio. A relação entre estas duas variáveis revela-se bastante importante em muitas aplicações físicas. Os dados da revista são apresentados na tabela seguinte, em que X representa a viscosidade do óleo e Y o desgaste do aço macio (em  $10^{-4} \text{ mm}^3$ ):

X	1.6	9.4	15.5	20	22	35.5	43	40.5	33
Y	240	181	193	155	172	110	113	75	94

- Construa o diagrama de dispersão para os dados. Acha plausível uma relação linear entre as duas variáveis? Justifique.
- Calcule a equação da recta dada pelo método dos mínimos quadrados, e represente-a no diagrama construído na alínea anterior. Comente o ajustamento da recta obtida aos pontos  $(x_i, y_i)$ ,  $i=1,2,\dots,9$ .
- Preveja o desgaste sofrido no aço quando a viscosidade do óleo é de 30.

2. O volume de vapor de água despendido por mês numa fábrica química parece estar relacionado com a temperatura média ambiental (em °F) observada nesse mês. O volume de vapor de água e a respectiva temperatura média observadas na fábrica no passado ano são dados na seguinte tabela:

Mês	Temp(°F)	Vapor/1000	Mês	Temp(°F)	Vapor/1000
Jan	21	185.79	Julho	68	621.55
Fev	24	214.47	Ago	74	675.06
Mar	32	288.03	Set	62	562.03
Abr	47	424.84	Out	50	452.93
Mai	50	454.58	Nov	41	369.95
Junho	59	539.03	Dez	30	273.98

- Desenhe o diagrama de dispersão (Temperatura  $\times$  Volume) para os dados da tabela. O diagrama sugere-lhe uma relação linear entre as duas variáveis? Se sim, o grau de relacionamento (linear) parece-lhe forte ou fraco? Justifique.
- Assumindo que o modelo de regressão linear simples é apropriado para estes dados, calcule a equação de regressão estimada pelo método dos mínimos quadrados.
- Que alteração se espera no volume de vapor de água usado quando a temperatura média mensal aumenta 1°F?
- Qual é a estimativa esperada de volume de vapor usado quando a temperatura média mensal é de 47°F? E de 19°F?

3. Um artigo de um jornal de Engenharia Ambiente apresenta os resultados de um estudo acerca da presença de cloreto em várias zonas portuárias que circundam determinada ilha. Os dados apresentados a seguir dizem respeito à área destinada à navegação na bacia hidrográfica X (em %) e à concentração de cloreto Y (em mg/l) :

X	0.19	0.15	0.57	0.7	0.67	0.63	0.47	0.7	0.6	0.78	0.81	0.78	0.69	1.03	1.05
Y	4.4	6.6	9.7	10.6	10.8	10.9	11.8	12.1	14.3	14.7	15	17.3	19.2	23.1	27.4

- Determine os coeficientes de correlação e de determinação. Interprete-os.
- Determine a recta de regressão usando o método dos mínimos quadrados.
- Estime a concentração de cloreto para uma bacia que tem 1% da área para navegação.

4. Calcule e interprete os coeficientes de correlação e de determinação para os exercícios 1 e 2. Compare as suas conclusões com as obtidas anteriormente da observação do diagrama de dispersão.

5. Para um estudo sobre poluição atmosférica foram observados num certo local dum centro urbano o volume de tráfego V (nº carros/hora) e a concentração de monóxido de carbono C (Co). Os resultados de sete leituras são dados a seguir:

$$\sum_{i=1}^7 v_i = 1750; \sum_{i=1}^7 v_i^2 = 507500; \sum_{i=1}^7 c_i = 80.3; \sum_{i=1}^7 c_i^2 = 944.19; \sum_{i=1}^7 v_i c_i = 21330.$$

Comente a seguinte afirmação, apresentando todos os cálculos que achar necessários:

“Apenas 2% da concentração de monóxido de carbono é explicada por outros factores que não o volume de tráfego existente.”

6. Foi realizado um inquérito a dez alunos para tentar relacionar o número de horas de estudo X e a classificação obtida em certo exame Y. Obtiveram-se os seguintes valores:

Horas	8	5	15	13	10	5	18	15	2	8
Cotação	50	44	69	72	70	49	94	85	33	65

- Construa o diagrama de dispersão correspondente aos valores encontrados. Comente.
- Calcule a equação da recta de regressão da classificação obtida em função das horas de estudo. Faça a avaliação da recta, recorrendo aos coeficientes estudados.
- Cerca de \_\_\_\_\_% na variação das notas se deve a outros factores que não as horas despendidas no estudo. Complete a afirmação. Pode avançar outros factores importantes?

7. Num estudo para determinar de que modo a habilidade para executar uma determinada tarefa complexa é influenciada pela quantidade de treino, foram usados 15 indivíduos aos quais foi dado um treino que variava de 3 a 12 horas. Depois do treino foram registados os tempos que cada um deles gastou a executar a tarefa. Representando por X a duração do treino (em horas) e por Y o tempo gasto na execução da tarefa (em minutos), os resultados obtidos resumem-se seguidamente:

$$\bar{x} = 7.2 \quad \bar{y} = 45.6 \quad \sum_{i=1}^{15} x_i^2 = 811.2 \quad \sum_{i=1}^{15} y_i^2 = 31350.6 \quad \sum_{i=1}^{15} x_i y_i = 4867.6$$

$$SSE=62.824$$

$$SST=160.2$$

- Determine a equação de regressão estimada pelo método dos mínimos quadrados.
- Interprete o coeficiente de regressão  $b_1$ .
- Calcule o coeficiente de correlação e interprete o valor encontrado.
- O que pode dizer relativamente ao tempo de execução da tarefa de um indivíduo que tenha sido sujeito a 35 horas de treino?

8. Querendo estudar a relação entre a quilometragem de um carro usado e o seu preço de venda, foram recolhidos os dados seguintes.

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

- Construa o diagrama de dispersão e conclua se um modelo linear é ou não adequado para descrever a relação entre as duas variáveis em estudo.
- Determine a equação de regressão linear estimada pelo método dos mínimos quadrados. Interprete o modelo estimado
- Calcule os coeficientes de correlação e determinação e interprete os seus valores.

## SOLUÇÕES

1.a) Por análise do diagrama de dispersão, parece razoável assumir um relacionamento linear entre X e Y, pois os pontos do diagrama de dispersão concentram-se com pequenos desvios à volta de uma recta imaginária.

1.b)  $\hat{y} = 234.07 - 3.51x$ ;

1.c) Quando a viscosidade do óleo é 30, prevê-se que o desgaste sofrido no aço seja, em média, 128.8 .

2.a) O diagrama de dispersão sugere uma relação linear entre as duas variáveis, já que os pontos se concentram bastante à volta de uma recta. Uma vez que a concentração dos pontos à volta da recta é bastante acentuada concluímos que o grau de relacionamento linear é forte.

2.b)  $\hat{y} = -6.34 + 9.21x$

2.c) Quando a temperatura média mensal aumenta 1°F, espera-se um acréscimo no volume de vapor de água usado de 9.21.

2.d) Para uma temperatura de 47°F, estima-se que, em média, o volume de vapor usado seja de 426.46. Para uma temperatura de 19°F, é arriscado usar a recta de regressão para estimar o volume de vapor usado, já que aquele valor de temperatura está fora do âmbito dos dados. De facto, fora do âmbito dos dados não há qualquer evidência de que a mesma relação linear seja ainda adequada para descrever o relacionamento entre X e Y.

3.a)  $r^2=0.76 \rightarrow 76\%$  da variação na concentração de cloreto está relacionada linearmente com a variação na área destinada a navegação;  $r=0.87 \rightarrow$  existe um relacionamento linear forte entre a concentração de cloreto e a área destinada a navegação.

3.b)  $\hat{y} = 0.16 + 20.93x$

3.c) Para uma bacia que tenha 1% de área de navegação, estima-se que a concentração de cloreto seja, em média, 21.09 mg/l.

4. **Para o exercício 1:**  $r^2=0.88$  e  $r=-0.94 \rightarrow$  sendo  $r^2$  muito próximo de 1 e  $r$  muito próximo de -1, podemos concluir que há um bom ajustamento dos dados à recta estimada, i.e., o modelo linear considerado é um modelo adequado. De facto, o coeficiente de determinação  $r^2$ , indica que cerca de 88% das variações no desgaste do aço macio são explicadas pelo modelo, i.e., por variações na viscosidade do óleo. Ficam por explicar apenas 12 % das variações no desgaste do aço macio, e

portanto, podemos dizer que o modelo considerado explica grande parte das variações na variável dependente. Estas conclusões estão de acordo com as obtidas por análise do diagrama de dispersão.

**Para o exercício 2:**  $r^2=0.9999$  e  $r=0.9999 \rightarrow$  Os valores muito próximos de 1 de ambos os coeficientes, indicam haver um relacionamento linear forte entre as duas variáveis e, portanto, o modelo linear considerado é bastante adequado para traduzir o relacionamento entre as duas variáveis. Estas conclusões estão de acordo com as obtidas por análise do diagrama de dispersão.

5.  $r^2=0.977 \cong 0.98 \rightarrow$  cerca de 98% das variações na concentração de monóxido de carbono são explicadas linearmente por variações no volume de tráfego. Ficam por explicar apenas 2% das variações na concentração de monóxido de carbono que resultam de outros factores não considerados no modelo.

6. b)  $\hat{y} = 29.06 + 3.44x$ ;  $r^2=0.89$ ;  $r= 0.95$

6.c) Cerca de 11% na variação das notas se deve a outros factores que não as horas despendidas no estudo. Outros factores que podem ter influência na classificação no exame: assiduidade às aulas, gosto pela matéria, ....

7.a)  $\hat{y}=57.857-1.70238x$ ;

7.b)  $b_1=-1.70238 \rightarrow$  estima-se que um acréscimo na duração do treino de uma hora, conduza, em média, a um decréscimo no tempo de execução da tarefa de 1.7 minutos;

7.c)  $r=-0.779$ ;

7.d) Não é sensato utilizar a equação de regressão obtida para predizer o tempo de execução da tarefa para as 35 horas de treino, uma vez que 35 está fora do âmbito dos dados.

8. *Resolvido nos Slides*