

Licenciatura em Engenharia Informática

Métodos Estatísticos

Exercícios

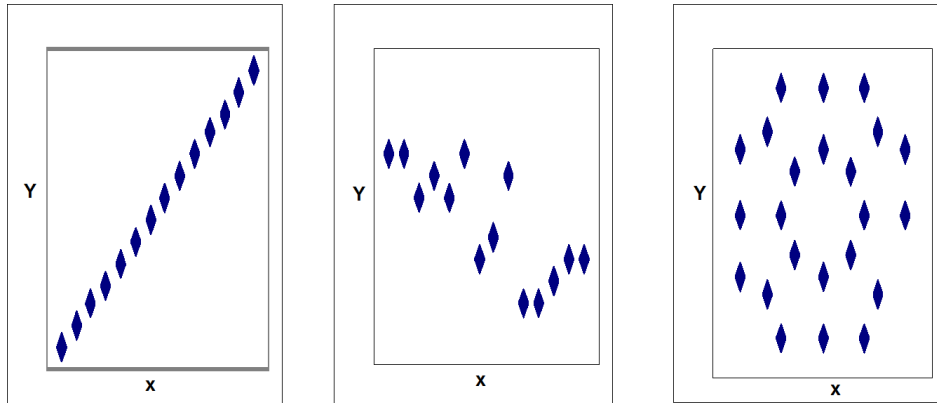
Regressão Linear Simples

Departamento de Matemática



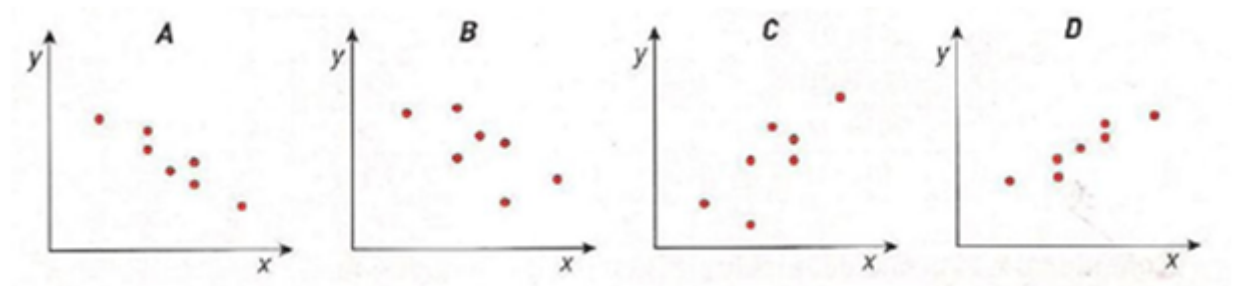
2 Regressão Linear Simples

Exercício 2.1 Indique, justificando, qual dos valores abaixo indicados se aproxima mais do coeficiente de correlação dos dados descritos nas seguintes nuvens de pontos,



1. $r_{xy} = 0$.
2. $r_{xy} = 1$.
3. $r_{xy} = -0.5$.
4. $r_{xy} = 2$.

Exercício 2.2 A cada uma das nuvens de pontos A, B, C e D representadas a seguir, faça corresponder um e um só dos seguintes coeficientes de correlação.



1. $r_{xy} = -0.70$.
2. $r_{xy} = 0.96$.
3. $r_{xy} = -0.94$.
4. $r_{xy} = 0.75$.

Exercício 2.3 Foi medida a várias profundidades (entre 1 e 5 metros), a percentagem de dióxido de urânio numa dada zona geológica, tendo-se obtido os seguintes dados:

$$\sum_{i=1}^8 x_i = 23 \quad \sum_{i=1}^8 x_i^2 = 81.5 \quad \sum_{i=1}^8 x_i y_i = 139.8 \quad \sum_{i=1}^8 y_i = 39.1 \quad \sum_{i=1}^8 y_i^2 = 247.05$$

onde X representa a profundidade e Y representa a percentagem de dióxido de urânio. Verifique se a profundidade explica a percentagem de dióxido de urânio e determine, comentando o resultado, a percentagem de dióxido de urânio, previsível a 2 metros de profundidade.

Exercício 2.4 Uma amostra de 10 alunos, selecionados ao acaso, de entre os inscritos nas unidades curriculares ME e Mat I na ESTSetúbal, produziu as seguintes classificações (0 a 100) num teste efetuado a meio do ano letivo (em %):

Mat I (x_i)	56	50	72	67	31	50	65	40	80	61
ME (y_i)	60	50	67	75	44	56	72	48	76	62

$$\sum_{i=1}^{10} x_i = 572 \quad \sum_{i=1}^{10} x_i^2 = 34716 \quad \sum_{i=1}^{10} y_i = 610 \quad \sum_{i=1}^{10} y_i^2 = 38394 \quad \sum_{i=1}^{10} x_i y_i = 36335$$

1. Represente os dados através de um diagrama de dispersão e comente-o com vista a um possível ajustamento de uma reta de regressão linear simples.
2. A Joana e o António obtiveram, respetivamente, as seguintes notas no teste de Mat I: 60% e 20%. Tendo ambos faltado ao teste de ME, sugira valores para as notas esperadas nos testes destes alunos. Indique, justificando, se os valores sugeridos são de confiança.
3. A Maria obteve 70% no teste de ME e ainda não conhece a nota do teste de Mat I. Indique, justificando, uma estimativa da nota do teste de Mat I da Maria.
4. Suponha que a amostra incluía mais um aluno, com as classificações de 12% e 78% respetivamente a Mat I e a ME. Represente novamente os dados através de uma nuvem de pontos e obtenha a correspondente reta dos mínimos quadrados, comentando a sua qualidade. Indique, justificando, como classifica a nova observação.

Exercício 2.5 Com a finalidade de elaborar um estudo sobre a produtividade dos trabalhadores de uma dada empresa por hora de trabalho (Y , em percentagem), face ao número médio de horas de trabalho, X , um sindicato registou num determinado dia, os seguintes resultados de 101 funcionários escolhidos aleatoriamente:

$$\sum_{i=1}^{101} x_i = 909 \quad \sum_{i=1}^{101} (x_i - \bar{x})^2 = 2000 \quad \sum_{i=1}^{101} x_i y_i = 62281 \quad \bar{y} = 76.3 \quad \sum_{i=1}^{101} y_i^2 = 614172$$

1. Caso se justifique, ajuste uma reta de regressão linear aos dados.
2. Tendo-se observado que $\min x_i = 6.5$ e $\max x_i = 13.5$ preveja a taxa de produtividade de um trabalhador que trabalhe uma média de 8 horas diárias, e de outro que trabalhe uma média de 20 horas diárias. Indique, justificando, se estas previsões têm boa qualidade.

Exercício 2.6 Com o intuito de verificar qual o impacto do investimento em publicidade (X , em milhares de euros) no volume de vendas das empresas (Y , em milhões de euros), recolheram-se dados de 15 empresas, tendo-se obtidos os seguintes resultados:

$$\sum_{i=1}^{15} x_i = 1322 \quad \sum_{i=1}^{15} x_i^2 = 117532 \quad \sum_{i=1}^{15} y_i = 78 \quad \sum_{i=1}^{15} y_i^2 = 460 \quad \sum_{i=1}^{15} x_i y_i = 7072$$

1. Determine a reta dos mínimos quadrados e comente o seu ajustamento aos dados.
2. Calcule uma estimativa do volume de vendas de uma empresa que tenha investido 90000 euros em publicidade. Comente o resultado, e determine o respetivo resíduo sabendo que o valor observado foi $y = 6$.

Exercício 2.7 A tabela seguinte constitui uma amostra aleatória referente às alturas, em centímetros, de 10 raparigas e respetivas mães:

Altura das mães	165	169	166	158	172	160	167	155	169	166
Altura das filhas	168	170	168	160	172	164	165	158	172	164

1. Defina, justificando qual a variável explicativa e explicada, sabendo que se pretende prever a altura das filhas.
2. Represente as observações num diagrama de dispersão e comente a possível existência de correlação linear.
3. Calcule o coeficiente de correlação linear empírico e comente-o.
4. Preveja a altura de uma rapariga cuja mãe tenha 150 cm de altura.
5. Calcule os resíduos, analise-os graficamente e conclua acerca da qualidade do ajustamento efectuado.

Exercício 2.8 As observações que se seguem dizem respeito à largura da estrada, em metros, em sete zonas da cidade de Setúbal e ao número de acidentes por ano nessas zonas:

zonas (i)	1	2	3	4	5	6	7
largura da estrada (x_i)	2.6	3.0	4.4	5.0	6.2	6.6	7.2
número de acidentes (y_i)	92	85	78	81	54	53	40

$$\sum_{i=1}^7 x_i = 35 \quad \sum_{i=1}^7 x_i^2 = 193.96 \quad \sum_{i=1}^7 y_i = 483 \quad \sum_{i=1}^7 y_i^2 = 35659 \quad \sum_{i=1}^7 x_i y_i = 2215$$

1. Recorrendo ao modelo de regressão linear simples, estime o número de acidentes se a largura da estrada for de 3.5 metros e comente a qualidade desta estimativa.
2. Sabendo que numa dada estrada da cidade se registaram 60 acidentes por ano, estime a largura da estrada e comente a qualidade da estimativa obtida.

Exercício 2.9 O quadro seguinte é o resultado de observações feitas num túnel rodoviário durante períodos de 5 minutos, para o estudo da fluidez do tráfego,

Densidade (veículos / km)	43	55	40	52	39	33	50	33	44	21
Velocidade (km / hora)	27	23	31	24	35	41	27	40	32	51

1. Calcule a variância de cada um dos conjuntos de dados observados. Qual dos conjuntos de dados apresenta maior dispersão? Justifique.
2. Calcule o coeficiente de correlação linear entre as duas variáveis. Que conclusões pode retirar?
3. Determine a equação da reta de regressão, caso se justifique.

Exercício 2.10 Numa amostra de 21 homens e de 31 mulheres, escolhidas ao acaso numa dada região, observaram-se os seus pesos (em kg) e as suas idades (em anos), tendo-se obtido os valores que constam na tabela seguinte:

	Homens		Mulheres	
	Peso	Idade	Peso	Idade
média	70	45	60	40
variância	16	9	25	16
coeficiente de correlação linear empírico	0.75		0.9	

1. Em relação às mulheres, qual o conjunto de dados que apresenta maior dispersão, o peso ou a idade? Justifique a sua resposta.
2. Em que conjunto de dados, homens ou mulheres, o ajustamento do modelo de regressão linear parece ser melhor? Justifique a sua resposta.
3. Recorrendo a modelos de regressão linear adequados responda às seguintes questões.
 - (a) Qual é o peso previsto para uma mulher com 38 anos?
 - (b) Qual é a idade prevista de um homem que pesa 72 kg?

Exercício 2.11 Suponha que um médico está interessado em estudar a relação entre as despesas mensais com a saúde e os rendimentos dos portugueses. Recolheu-se informação (em milhares de euros) relativa a 15 famílias, tendo-se obtido os seguintes dados:

Rendimento	2.5	1.7	1.3	2.3	2.7	1.8	1.3	1.5
Despesa	0.14	0.106	0.1	0.12	0.15	0.11	0.106	0.1

Rendimento	2.1	2.8	1.9	1.4	1.7	2.1	1.8
Despesa	1.4	0.15	0.14	0.11	0.106	0.15	0.14

1. Represente as observações num diagrama de dispersão. Esta representação sugere a existência de alguma relação linear entre as variáveis em estudo?
2. Calcule o coeficiente de correlação linear empírico e obtenha a expressão da reta dos mínimos quadrados. Comente o resultado obtido.
3. Calcule os resíduos, analise-os graficamente e conclua acerca da qualidade do ajustamento efetuado.
4. Substitua na amostra a observação (2.1, 1.4) por (2.1, 0.14). Represente novamente os dados através de uma nuvem de pontos e obtenha a correspondente reta dos mínimos quadrados. Indique, justificando, como classifica a observação substituída.
5. Calcule, utilizando a nova observação, o coeficiente de correlação linear empírico e proceda a uma análise de resíduos.
6. Obtenha uma estimativa (com base no último modelo linear) para os gastos com a saúde de uma família cujos rendimentos mensais são de 1500 euros e comente a sua qualidade.

Soluções

2.1.1: 3^a nuvem de pontos. **2.1.2:** 1^a nuvem de pontos. **2.1.3:** 2^a nuvem de pontos. **2.1.4:** Valor impossível para r_{xy} .

2.2.1: B . **2.2.2:** D . **2.2.3:** A . **2.2.4:** C .

2.3: $r_{xy} = 0.9338$ e $\hat{y} = 3.329\%$.

2.4.1: $r_{xy} = 0.94$. **2.4.2:** $\hat{y} = 19.681 + 0.722x$; Joana: $\hat{y} = 63.001$; António: $\hat{y} = 34.121$. **2.4.3:** $\hat{x} = -17.144 + 1.2187y$; $\hat{x} = 68.2$. **2.4.4:** $\hat{y} = 52.293 + 0.1931x$

2.5.1: $r_{xy} = -0.97782$ e $\hat{y} = 108.14 - 3.5379x$. **2.5.2:** $x = 8 \Rightarrow \hat{y} = 79.837$ (tem boa qualidade); $x = 20 \Rightarrow \hat{y} = 37.382$ (não se pode aferir sobre a qualidade pois está fora do intervalo de valores analisados).

2.6.1: $\hat{y} = -11.863 + 0.1936x$; $r_{xy} = 0.8387$. **2.6.2:** $\hat{y} = 5.5619$; $e_i = 0.4381$.

2.7.1: $-$. **2.7.2:** $-$. **2.7.3:** $r_{xy} = 0.9196$. **2.7.4:** 154.1 cm. **2.7.5:** Bom.

2.8.1: $\hat{y} = 84.823$. **2.8.2:** $\hat{x} = 5.77$.

2.9.1: $s_X^2 = 104.8889$ e $s_Y^2 = 77.6556$; Velocidade (pois $CV_X = 25\%$ e $CV_Y = 26.6\%$). **2.9.2:** $r_{xy} = -0.9726$. **2.9.3:** $\hat{y} = 67.4114 - 0.83686x$.

2.10.1: Idade. **2.10.2:** Mulheres. **2.10.3a:** 57.75 kg. **2.10.3b:** 46 anos.

2.11.1: $-$. **2.11.2:** $r_{xy} = 0.1466$ e $\hat{y} = 0.01548 + 0.1002x$. **2.11.3:** $-$. **2.11.4:** $\hat{y} = 0.06019 + 0.033397x$. **2.11.5:** $r_{xy} = 0.8084$. **2.11.6:** $\hat{y} = 0.11029$ mil euros por mês.