

Licenciatura em Engenharia Informática

Métodos Estatísticos

Exercícios

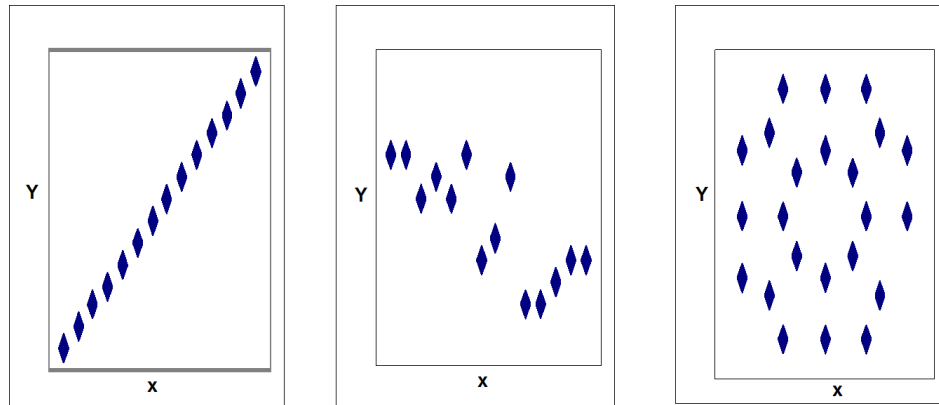
Regressão Linear Simples

Departamento de Matemática



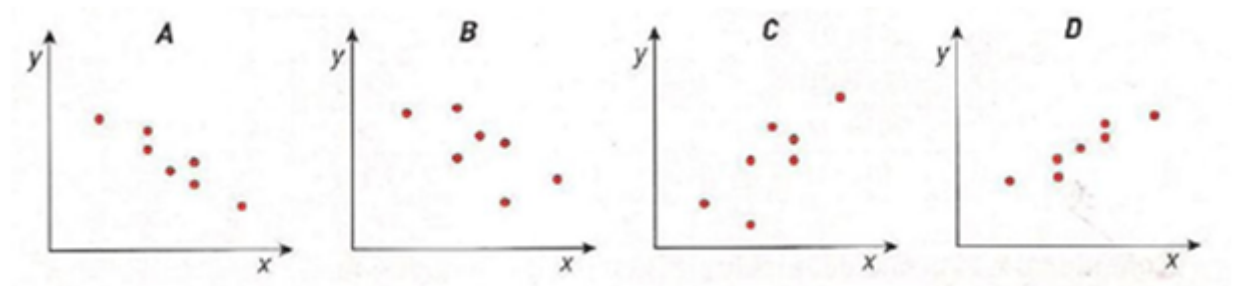
7 Regressão Linear Simples

Exercício 7.1 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 7.2 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 7.3 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} 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

1. Construa o diagrama de dispersão e calcule o coeficiente de correlação linear de Pearson. O que pode concluir quanto à existência de relação linear entre as variáveis "viscosidade do óleo" e "desgaste do aço"?
2. Obtenha a expressão da reta de regressão linear quando:

- (a) a variável independente é a "viscosidade do óleo";
 - (b) a variável dependente é a "viscosidade do óleo".
3. Preveja o desgaste sofrido no aço quando a viscosidade do óleo é de 30 e quando é de 75. Comente os valores obtidos.
 4. Calcule os resíduos do modelo utilizado na alínea anterior e construa o gráfico dos resíduos e comente.

Exercício 7.4 Um pesquisador deseja verificar se um instrumento para medir a concentração de determinada substância no sangue está bem calibrado. Para tal, considerou 15 amostras de concentrações conhecidas (X) e determinou a respectiva concentração através do instrumento (Y), obtendo os seguintes dados:

X	2	2	2	4	4	4	6	6	6	8	8	8	10	10	10
Y	2.1	1.8	1.9	4.5	4.2	4.0	6.2	6.0	6.5	8.2	7.8	7.7	9.6	10.0	10.1

1. Construa o diagrama de dispersão e calcule o coeficiente de correlação linear de Pearson. O que pode concluir quanto à existência de relação linear entre as variáveis?
2. Obtenha a reta de regressão linear que permite efetuar previsões para X com base em valores de Y . O que pode concluir quanto ao ajustamento deste modelo aos dados amostrais? Podemos dizer que o instrumento está bem calibrado? Porquê?
3. Preveja qual será o valor da concentração de determinada substância no sangue fornecido pelo instrumento para uma amostra sanguínea cuja concentração conhecida é 3. Comente.

Exercício 7.5 Numa dada clínica foi feita uma pesquisa com 8 mulheres de 50 anos de idade. Nessa pesquisa foram feitas as seguintes perguntas:

- Qual é o nível de HDL-Colesterol no sangue?
- Quantas horas semanais pratica exercício físico?

e os resultados obtidos foram os seguintes:

HDL-Colesterol (mg/dl)	40	50	55	60	65	70	70	75
Exercício físico (número de horas semanais)	0	2	3	3	2	4	5	4

1. Construa o diagrama de dispersão e determine o coeficiente de correlação linear de Pearson. O que pode concluir quanto à existência de relação linear entre as variáveis "HDL-Colesterol" e "Número de horas de prática de exercício físico"?
2. Determine o modelo de regressão linear que permite prever o "HDL-Colesterol" de uma mulher com 50 anos de idade e que pratica 1 hora de exercício físico por semana. Comente o valor obtido.

Exercício 7.6 Uma amostra de alunos, selecionados ao acaso, de entre os inscritos em duas Unidades Curriculares, produziu as seguintes classificações (0 a 100) num teste efetuado a meio do ano letivo (em %):

Unidade Curricular A	56	50	72	67	31	50	65	40	80	61
Unidade Curricular B	60	50	67	75	44	56	72	48	76	62

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 da Unidade Curricular A: 60% e 20%. Tendo ambos faltado ao teste da Unidade Curricular B, 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 da Unidade Curricular B e ainda não conhece a nota do teste da Unidade Curricular A. Indique, justificando, uma estimativa da nota do teste da Unidade Curricular A da Maria.
4. Suponha que a amostra incluía mais um aluno, com as classificações de 12% e 78% respetivamente à Unidade Curricular A e à Unidade Curricular B. 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 7.7 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 efetuado.

Exercício 7.8 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	2.6	3.0	4.4	5.0	6.2	6.6	7.2
número de acidentes	92	85	78	81	54	53	40

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 7.9 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 7.10 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 7.11 Na tabela seguinte está registada a percentagem da população com mais de 65 anos e a taxa de mortalidade (em permilagem) por distritos (dados de 1990):

Distrito	População com mais de 65 anos (%)	Taxa de mortalidade (%)
Aveiro	11.0	8.3
Beja	18.2	13.7
Braga	9.1	7.4
Castelo Branco	19.8	13.3
Évora	16.3	12.2
Guarda	19.0	13.4
Portalegre	20.1	13.8
Setúbal	10.7	7.2
Vila Real	13.0	10.6
Viseu	14.6	10.9

1. Identifique a variável dependente e independente e represente os dados através de um diagrama de dispersão. Comente o diagrama com vista a um possível ajustamento de uma reta de regressão linear simples.
2. Ajuste uma reta de regressão aos dados.
3. Qual a taxa de mortalidade prevista para um distrito em que a percentagem de população com mais de 65 anos é igual a 20? Comente, justificando, a qualidade da previsão efetuada.

Exercício 7.12 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.

Exercício 7.13 A tabela seguinte inclui os registos referentes à esperança média de vida à nascença, para homens e mulheres de países da União Europeia. Nessa tabela, X designa o número médio de anos de vida esperados à nascença para as mulheres e Y designa o número médio de anos de vida esperados à nascença para os homens.

Esperança média de vida à nascença para homens e mulheres

PAÍSES	MULHERES (x)	HOMENS (y)
Portugal	81,7	75,5
Espanha	85,0	78,9
França	84,3	77,5
Irlanda	81,6	76,8
Reino Unido	81,7	77,6
Bélgica	83,5	77,5
Holanda	82,3	78,3
Alemanha	82,4	77,2
Itália	84,1	78,8
Grécia	82,5	77,5

Fontes: INE e Eurostat

1. Construa o diagrama de dispersão. Comente.
2. Calcule o coeficiente de correlação linear empírico. Comente.
3. Na tabela anterior não constam os valores da Áustria. No entanto, sabe-se que os valores da esperança média de vida à nascença para homens e mulheres referentes à Áustria seguem o modelo de regressão linear obtido a partir dos dados da tabela. Determine um valor aproximado da esperança média de vida à nascença de um homem austríaco, sabendo que a esperança média de vida à nascença de uma mulher austríaca é 83 anos.

Exercício 7.14 No ficheiro EXCEL "Obesidade" tem as respostas a um inquérito efetuado num estudo sobre obesidade a um grupo de indivíduos obesos. No conjunto de variáveis disponíveis no ficheiro considere as seguintes:

- Género: Feminino, Masculino;
- Altura: em metros;
- Peso: em quilogramas;
- FAVC = Se come alimentos altamente calóricos habitualmente: 1=Sim, 0=Não;

1. Avalie o grau de relacionamento linear entre a altura e o peso dos indivíduos obesos do género feminino que comem habitualmente alimentos altamente calóricos.
2. Recorrendo ao modelo de regressão linear simples, estime o peso de um indivíduo obeso do género feminino que come habitualmente alimentos altamente calóricos e cuja a altura seja de 1.64 metros. Defina a variável independente e variável dependente, calcule a reta de regressão, efetue a previsão pretendida e comente o resultado obtido.

3. Recorrendo ao modelo de regressão linear simples, estime a altura de um indivíduo obeso do género feminino que come habitualmente alimentos altamente calóricos e pese 90 kg. Defina a variável independente e variável dependente, calcule a reta de regressão, efetue a previsão pretendida e comente o resultado obtido.

Soluções

7.1.1: 3ª nuvem de pontos. **7.1.2:** 1ª nuvem de pontos. **7.1.3:** 2ª nuvem de pontos. **7.1.4:** Valor impossível para r_{xy} .

7.2.1: B . **7.2.2:** D . **7.2.3:** A . **7.2.4:** C .

7.6.1: $r_{xy} = 0.94$. **7.6.2:** $\hat{y} = 19.681 + 0.722x$; Joana: $\hat{y} = 63.001$; António: $\hat{y} = 34.121$. **7.6.3:** $\hat{x} = -17.144 + 1.2187y$; $\hat{x} = 68.2$. **7.6.4:** $\hat{y} = 52.293 + 0.1931x$

7.7.1: $-$. **7.7.2:** $-$. **7.7.3:** $r_{xy} = 0.9196$. **7.7.4:** 154.1 cm. **7.7.5:** Bom.

7.8.1: $\hat{y} = 84.823$. **7.8.2:** $\hat{x} = 5.77$.

7.9.1: $s_X^2 = 104.8889$ e $s_Y^2 = 77.6556$; Velocidade (pois $CV_X = 25\%$ e $CV_Y = 26.6\%$). **7.9.2:** $r_{xy} = -0.9726$. **7.9.3:** $\hat{y} = 67.4114 - 0.83686x$.

7.10.1: Idade. **7.10.2:** Mulheres. **7.10.3a:** 57.75 kg. **7.10.3b:** 46 anos.

7.11.1: $-$. **7.11.2:** $\hat{y} = 1.5547 + 0.6275x$. **7.11.3:** $\hat{y} = 14.105$ e $r_{xy} = 0.97313$.

7.12.1: $-$. **7.12.2:** $r_{xy} = 0.1466$ e $\hat{y} = 0.01548 + 0.1002x$. **7.12.3:** $-$. **7.12.4:** $\hat{y} = 0.06019 + 0.033397x$. **7.12.5:** $r_{xy} = 0.8084$. **7.12.6:** $\hat{y} = 0.11029$ mil euros por mês.

7.13.1: $-$ **7.13.2:** $r_{xy} = 0.6728$ **7.13.3:** 78 anos.