PROBABILIDADES E ESTATÍSTICA E
Ano Lectivo 2011/2012

Exercícios sobre Estatística

Estatística Descritiva

1. No âmbito dos inquéritos que são efectuados por determinado organismo de obtenção de dados, é importante ter noção dos erros de digitação que os entrevistadores cometem ao anotarem informaticamente as respostas dos seus entrevistados. Assim, para um inquérito de 50 questões registaram-se, para 90 entrevistas, os seguintes números de erros:

N^o de erros	Frequência
0	5
1	17
2	29
3	23
4	11
5	5

   (a) Determine as frequências relativas e as frequências relativas acumuladas. Coloque-as em gráfico.
   (b) Que percentagem de entrevistas tiveram menos de 3 erros? Que número de erros é mais comum?
   (c) Determine a média do número de erros, o seu desvio padrão e o coeficiente de variação.

   Solução: –, 57%, 2 erros, 2.366667, 1.240243, 52.40463%

2. Suponha que os dados seguintes se referem ao número de palavras que constituem o vocabulário de crianças de 5 anos:

   205, 377, 292, 300, 179, 240, 300, 190, 680, 250, 180, 170, 211, 266, 303, 350, 375, 288, 360, 225

   (a) Estes dados são de natureza discreta ou contínua? Construa uma sua distribuição de frequências absoluta, absoluta acumulada, relativa e relativa acumulada. Esboce o histograma das frequências relativas e o correspondente polígono de frequências. Construa uma ogiva.
   (b) Existe algum *outlier* presente nos dados?
   (c) Determine a média, a mediana, a moda, e os 1^o e 3^o quartis dos dados. Construa o diagrama de caixa-e-bigodes dos dados.
   (d) Determine o desvio padrão e o coeficiente de variação dos dados.
   (e) O que pode dizer quanto à simetria dos dados? Justifique.

   Solução: –, –, 287.05, 277, 300, 205, 303, –, 113.9679, 39,70315%

3. Os chamados movimentos rápidos dos olhos durante o sono (REM - rapid eye movement) estão associados a períodos de sonho. A duração da actividade REM foi registada para 18 indivíduos (em segundos):

   7.00, 7.75, 9.50, 11.60, 10.55, 7.75, 12.00, 10.75, 12.51, 10.91, 8.30, 9.71, 10.50, 11.60, 6.25, 11.75, 9.75, 10.00

   (a) Construa uma distribuição de frequência dos dados usando uma amplitude de classe l de 1 segundo. Represente-a graficamente. Esboce uma ogiva.
   (b) Determine a média, a mediana e os 1^o e 3^o quartis dos dados.
   (c) Determine o desvio padrão e o coeficiente de variação dos dados.

   Solução: –, 9.898889, 10.25, 7.75, 8.30, 11.60, 1.826533, 18.4519%

4. Segue-se a distribuição por faixas etárias da população de uma certa cidade, com idades entre 5 e 39 anos, relativas ao ano de 1987:

Idade	Número
$[5, 10[$	30116
$[10, 15[$	14633
$[15, 20[$	29424
$[20, 25[$	40146
$[25, 30[$	29424
$[30, 35[$	44555
$[35, 40[$	40100

   (a) Construa um histograma de frequências.
   (b) Se o histograma tivesse sido calculado com base nas frequências relativas a sua forma diferiria do histograma desenhado na alínea anterior? Se não tiver a certeza construa-o para comparação.
   (c) Determine duas medidas de localização dos dados e duas medidas de dispersão.

   Solução: –, $\bar{x}$=24.465779, Me =22.876707, s=9.882474, cv =40,3930%

5. Os dados que se seguem dizem respeito aos salários mensais líquidos (Euro) de um conjunto de 36 pessoas de determinada cidade entrevistadas na rua ao acaso:

   1195, 660, 870, 1150, 1225, 2465, 1100, 2480, 1300, 2330, 2020, 1540, 685, 867, 1000, 1470, 1085, 1060, 1790, 2690, 1535, 3995, 1615, 1230, 670, 590, 1100, 1040, 4200, 1030, 1165, 3320, 1260, 1790, 2740, 1490

   (a) Construa uma distribuição de frequências relativas e ponha-a em gráfico. Construa ainda uma ogiva.
   (b) Qual a percentagem de pessoas que ganha menos que €1100 líquidos?
   (c) Determine a média, a mediana, a moda, os 1º e 3º quartis e o desvio padrão dos dados.
   (d) Comente a simetria da distribuição com base no coeficiente de enviesamento de Pearson e com base no diagrama de caixa de bigodes.

   Solução: –, 36 %, 1604.222, 1245, Duas modas: 1100 e 1790, 1040, 1790, 895.271263, 1.203732

6. Inserido num estudo antropólogo procura-se determinar algumas características fisiológicas de uma população. Os números seguintes representam os níveis de colesterol no sangue encontrado em 25 membros de uma tribo Africana, medido em miligramas de colesterol por decilitro de sangue:

   200, 241, 232, 177, 207, 181, 195, 182, 181, 233, 176, 170, 217, 164, 188, 164, 211, 204, 160, 172, 212, 186, 160, 203, 191

   (a) Construa as distribuições de frequências absolutas e relativas correspondentes.
   (b) Construa o histograma e o polígono de frequências das frequências relativas. Comente.
   (c) Determine a média, a mediana e a moda dos dados.
   (d) Determine o coeficiente de variação dos dados.

   Solução: –, –, 192.28, 188, Três modas: 160, 164, 181, 12.1064%

7. Um psicólogo desenvolveu uma técnica para ajudar as pessoas a melhorarem a sua memória. Certo material é dado a 30 pessoas para o memorizarem antes de aprenderem a técnica e semelhante material é dado às mesmas pessoas para o memorizarem depois de apreendida a referida técnica. A diferença de tempo que as pessoas demoraram a memorizar os materiais antes e depois de aprendida a técnica seguem-se (minutos):

   5, 40, 45, 11, 13, 20, 14, 5, 23, 18, 17, 4, 4, 5, 29, 18, 15, 21, 24, 16, 2, 15, 19, 30, 24, 21, 14, 18, 26, 40

   (a) Construa uma distribuição de frequências, o seu histograma e o seu polígono de frequências.
   (b) Tome uma classe e escreva por palavras exactamente o que ela lhe diz.
   (c) Calcule 3 medidas de localização dos dados e discuta a simetria da distribuição dos dados, construindo um diagrama de caixa de bigodes.

   Solução: –, –, $\bar{x} = 18.5333$, $Me = 18$, Duas modas: 5 e 18

8. O fluxo de tráfego num cruzamento, definido pelo número de veículos X que chegam ao cruzamento por minuto, tem uma distribuição que se supõe de Poisson. Efectou-se um conjunto de 100 observações independentes de X, com os seguintes resultados:

0	3	1	0	1	1	1	3	4	3	2	0	2	0	0	0	4	2	3	4
1	6	1	4	1	1	4	2	7	4	3	0	2	5	2	3	2	1	5	5
0	3	2	2	2	1	4	1	2	4	2	1	1	2	3	6	2	1	2	2
4	2	3	2	1	3	0	4	3	0	1	1	1	1	0	2	0	1	2	2
2	3	3	1	0	3	2	3	0	4	5	2	4	0	1	4	3	2	3	2

   (a) Construa a tabela de frequências desta amostra e esboce o respectivo diagrama de barras (histograma).
   (b) Determine a média, o desvio padrão amostral e o coeficiente de variação.
   (c) Calcule a mediana e os quartis amostrais. Desenhe a caixa de bigodes.
   (d) Verifique se existem outliers.

   Solução: –, –, 2.16, 1.561662, 72.29916%, 2, 1, 3

9. Considere-se o registo dos últimos 50 tempos (em anos) entre avarias consecutivas de uma máquina:

5,1	6,2	11,0	9,1	8,5	5,7	12,0	5,1	5,1	5,1
5,8	5,4	6,3	6,1	6,5	6,3	5,7	5,3	5,5	5,0
5,5	7,1	6,8	5,5	7,1	9,5	7,3	7,8	9,8	9,6
5,9	7,6	9,6	6,5	6,6	5,3	6,8	6,2	5,2	6,9
6,5	6,9	9,5	9,4	6,8	5,3	7,0	7,6	6,1	5,6

   (a) Determine a tabela de frequências, considerando as seguintes classes para agrupamento: $]0,6]$, $[6,7]$, $]7,8]$, $]8,9,6]$, $]9,6,11]$ e $]11,+\infty[$. Desenhe o histograma relativo àquela tabela de frequências. Que distribuição proporia para a v.a. X-tempo entre avarias (em anos)?
   (b) Determine a média, o desvio padrão amostral e o coeficiente de variação.
   (c) Calcule a mediana e os quartis amostrais. Desenhe a caixa de bigodes.
   (d) Verifique se existem outliers.

   Solução: –, –, –, 6.882, 1.694130, 24.61683%, 6.5, 5.5, 7.6

10. Numa empresa de engarrafamento de água mineral, procura saber-se se a máquina de enchimento de garrafas com 1.5 litros está bem calibrada (supõe-se que as garrafas têm capacidade superior a 1.5 litros). Para tal, o departamento de produção recolheu uma amostra de pesos de 50 garrafas, em gramas, cujos valores foram:

1503,50	1501,55	1503,37	1513,31	1497,78
1498,61	1496,55	1503,15	1496,94	1500,76
1506,22	1495,77	1511,71	1507,69	1497,39
1511,38	1500,11	1504,57	1509,51	1503,19
1510,99	1501,13	1504,07	1514,59	1504,84
1513,67	1494,41	1502,43	1504,58	1505,14
1494,08	1502,16	1514,86	1502,38	1503,39
1503,83	1502,98	1509,33	1508,38	1515,97
1510,48	1505,67	1516,88	1503,09	1496,29
1499,57	1503,17	1501,73	1508,79	1501,32

Suponha que o invólucro de plástico tem um peso fixo de 10 gramas e recorde que a água pesa um quilograma por litro.

Considere X-o volume (em litros) por garrafa.

(a) Determine a média, o desvio padrão e o coeficiente de variação da amostra dos volumes por garrafa.

(b) Agrupe os dados relativos ao peso das 50 garrafas de acordo com as seguintes classes $]0, 1498.08]$, $]1502.08, 1506.08]$, $]1510.08, 1514.08]$ e $]1518.08, +\infty[$. Esboce o respectivo histograma.

(c) Ainda para a amostra dos pesos por garrafa: calcule a mediana e os quartis amostrais, desenhe a caixa de bigodes e verifique se existem outliers.

Solução: 1494.465, 5.816772, 0.389221%, –, 1503.38, 1501.13, 1508.79

11. O quadro que se segue refere o valor das acções (em euros) das SAD do Sporting Clube de Portugal (SCP) e do Futebol Clube do Porto (FCP) de 3 de Janeiro a 27 de Março de 2000.

Dia	SCP	FCP	Dia	SCP	FCP	Dia	SCP	FCP
03-Jan	4,17	4,35	31-Jan	4,04	4,11	28-Fev	6,60	5,99
04-Jan	4,15	4,35	01-Fev	4,04	4,25	29-Fev	6,39	5,84
05-Jan	4,20	4,30	02-Fev	4,15	4,18	01-Mar	6,25	5,75
06-Jan	4,13	4,40	03-Fev	4,10	4,20	02-Mar	6,14	5,40
07-Jan	4,23	4,42	04-Fev	4,20	4,22	03-Mar	6,07	5,10
10-Jan	4,15	4,40	07-Fev	4,19	4,14	06-Mar	5,95	5,30
11-Jan	4,16	4,37	08-Fev	4,08	4,20	07-Mar	5,95	5,30
12-Jan	4,17	4,37	09-Fev	4,20	4,25	08-Mar	5,60	5,20
13-Jan	4,00	4,43	10-Fev	4,15	4,21	09-Mar	5,75	5,30
14-Jan	4,15	4,36	11-Fev	4,50	4,84	10-Mar	5,80	5,40
17-Jan	4,13	4,36	14-Fev	5,02	4,84	13-Mar	5,80	5,50
18-Jan	4,10	4,35	15-Fev	5,04	4,62	14-Mar	5,57	5,65
19-Jan	4,08	4,29	16-Fev	4,70	4,74	15-Mar	5,63	5,50
20-Jan	4,10	4,35	17-Fev	4,77	4,73	16-Mar	5,52	5,27
21-Jan	4,11	4,31	18-Fev	4,77	4,97	17-Mar	5,59	5,50
24-Jan	4,04	4,30	21-Fev	5,00	5,00	20-Mar	5,84	5,49
25-Jan	4,19	4,30	22-Fev	5,00	4,95	21-Mar	5,89	5,56
26-Jan	4,14	4,28	23-Fev	5,25	5,19	22-Mar	5,90	5,51
27-Jan	4,12	4,29	24-Fev	5,33	5,13	23-Mar	5,70	5,23
28-Jan	4,10	4,28	25-Fev	6,50	5,40	24-Mar	5,85	5,24

(a) Construa um histograma para o valor das acções da SAD do SCP utilizando as seguintes classes: $]0, 1.5]$, $]1.5, 3]$, $]3, 4.5]$, $]4.5, 6]$ e $]6, +\infty[$.

(b) Construa um histograma para o valor das acções da SAD do FCP com base nas seguintes classes: $]4, 4.2]$, $]4.2, 4.5]$, $]4.5, 5]$, $]5, 6]$ e $]6, +\infty[$. Parece-lhe que tais histogramas traduzem o comportamento dos activos em questão?

(c) Trace o diagrama de dispersão do valor destes dois activos financeiros. Estime a covariância e o coeficiente de correlação amostrais entre estas duas variáveis. Teça comentários sobre o diagrama e o valor da última estimativa.

(d) Construa os cronogramas referentes a estes dois activos financeiros. Será que a análise gráfica permite confirmar os comentários feitos na alínea anterior?

Solução: –, –, –, 0.4419807, 0.960096

12. O quadro seguinte apresenta a duração (em segundos), o custo (em euros) e o tipo de chamada (L-local; LOP-local para outros operadores; M-móvel; INT-internacional) das chamadas efectuadas por determinado cliente da Portugal Telecom em Novembro de 2006.

Tipo	Duração	Custo	Tipo	Duração	Custo
L	55	0,041	M	68	0,496
M	60	0,248	LOP	20	0,07
LOP	116	0,096	M	20	0,248
M	70	0,496	L	97	0,075
L	91	0,083	LOP	36	0,07
M	72	0,496	M	33	0,248
L	27	0,041	LOP	93	0,085
L	49	0,07	M	171	0,744
L	200	0,131	INT	5	0,1
M	53	0,248	INT	502	2,992
L	58	0,07	M	27	0,248
L	27	0,07	L	137	0,081
M	394	1,735	M	190	0,992
L	395	0,216	M	62	0,496
LOP	39	0,07	L	85	0,074
M	26	0,248	M	41	0,248

(a) Trace o diagrama de dispersão para a duração e o custo das chamadas.

(b) Calcule o coeficiente de correlação amostral entre a duração e o custo das chamadas.

Solução: –, 0.7741977

13. Seja X e Y variáveis aleatórias normais que representam os tempos (em segundos) que dois algoritmos distintos demoram a resolver o mesmo problema num computador, e considere duas amostras, cada uma com 50 realizações de X e Y, respectivamente:

X	17	18	19	16	15	13	16	17	17	12	15	19	16	18	16	14	18	16	25	18
Y	17	13	11	11	12	14	15	13	11	8	16	12	15	16	13	15	17	12	5	16
X	14	17	18	14	13	20	13	14	17	16	16	15	19	18	11	18	15	17	15	22
Y	17	20	11	15	18	8	16	25	14	15	15	15	18	12	10	16	16	14	17	23
X	13	15	15	14	13	17	16	19	19	13										
Y	17	13	17	17	10	22	4	22	13	21										

(a) Para cada amostra, determine a média, o desvio padrão amostral e o coeficiente de variação.

(b) Construa as tabelas de frequências destas amostras e esboce os respectivos histogramas, considerando as seguintes classes para agrupamento de cada amostra:

X	$]0, 11.5]$	$]11.5, 13.5]$	$]13.5, 15.5]$	$[15.5, 17.5]$	$]17.5, 19.5]$	$]19.5, 21.5]$	$]21.5, 23.5]$	$]23.5, +\infty[$
Y	$]0, 4.5]$	$]4.5, 7.5]$	$]7.5, 10.5]$	$[10.5, 13.5]$	$]13.5, 16.5]$	$]16.5, 19.5]$	$]19.5, 22.5]$	$]22.5, +\infty[$

(c) Para cada amostra, calcule a mediana e os quartis amostrais. Desenhe a caixa de bigodes.

(d) Verifique se existem outliers em cada uma das amostras.

Solução: $\bar{x} = 16.22$, $\bar{y} = 14.66$, $s_X = 2.605254$, $s_Y = 4.192364$, $cv_X = 0.1606198$, $cv_Y = 0.285973$, –, –, $M_{e,X} = 16$, $M_{e,Y} = 15$, $Q_{1,X} = 14$, $Q_{1,Y} = 12$, $Q_{3,X} = 18$, $Q_{3,Y} = 17$

Estimação Pontual

14. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de uma população X com valor médio μ e variância σ^2. Seja $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ a média aritmética da amostra.

(a) Verifique que $E(\overline{X}) = \mu$ e que $E(X_i - \overline{X}) = 0$, para $i = 1, 2, \ldots, n$. Constate que $V(X_i - \overline{X}) = E\left[(X_i - \overline{X})^2\right]$, para $i = 1, 2, \ldots, n$.

(b) Prove que $V(\overline{X}) = \frac{\sigma^2}{n}$ e que $cov(X_i, \overline{X}) = \frac{\sigma^2}{n}$, para $i = 1, 2, \ldots, n$.

(c) Mostre que $V(X_i - \overline{X}) = \frac{n-1}{n}\sigma^2$, para $i = 1, 2, \ldots, n$, e que, usando a alínea (a), $E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = (n-1)\sigma^2$.

(d) Qual deverá ser o valor da constante a de modo a que $E\left[a\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \sigma^2$?

(e) Se considerarmos $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ um estimador da variância da população X, isto é um estimador de $V(X) = \sigma^2$, verifique que $E(S^2) = \sigma^2$.

(f) Se o estimador da variância da população X fosse $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, qual seria o seu valor médio?

Solução: $\frac{1}{n-1}$, $\frac{n-1}{n}\sigma^2$

15. Seja $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ a média de uma amostra aleatória de dimensão n extraída de uma população X com distribuição $P(\lambda)$ e $\overline{Y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} Y_i$ a média de uma amostra aleatória com a mesma dimensão, extraída de uma população Y com distribuição $P(2\lambda)$, independente de X. Considere $\widetilde{\lambda} = p\overline{X} + \dfrac{1-p}{2}\overline{Y}, \quad 0 < p < 1$, um estimador de λ.

(a) Mostre que $\widetilde{\lambda}$ é um estimador centrado de λ, ou seja, mostre que $E\left(\widetilde{\lambda}\right) = \lambda$.

(b) Determine a variância do estimador $\widetilde{\lambda}$ e diga para que valor de $\underline{p}$ ela atinge um valor mínimo.

Solução: $\frac{\lambda}{n}\left(p^2 + \frac{(1-p)^2}{2}\right)$, $1/3$

16. Seja X uma população com uma determinada distribuição de parâmetros (δ, α). Sabendo que $E(X) = \delta\alpha$ e que $V(X) = \delta^2\alpha$,

(a) Deduza os estimadores de momentos para os parâmetros δ e α.

(b) Admita que α tem um valor conhecido, nomeadamente $\alpha = 4$. Deduza o estimador de momentos para δ. Verifique se é um estimador centrado e determine o respectivo erro padrão.

Solução: $\frac{(n-1)S^2}{n\overline{X}}$, $\frac{n\overline{X}^2}{(n-1)S^2}$, $\frac{\overline{X}}{4}$, $\frac{\delta}{2\sqrt{n}}$

17. Seja X uma população com distribuição Poisson de parâmetro λ e $(X_1, \ldots, X_n)$ uma amostra aleatória de X.
    (a) Deduza o estimador de momentos para o parâmetro λ.
    (b) Verifique que se trata de um estimador centrado e determine a respectiva variância. Mostre também que é um estimador consistente.
    (c) Verifique que a função log-verosimilhança é:
    $$l(\lambda) = -n\lambda + n\bar{x}\ln\lambda - \sum_{i=1}^{n} \ln x_i!, \;\; \lambda \in \mathbb{R}^+$$
    (d) Determine o estimador de máxima verosimilhança de λ.
    (e) Utilize a propriedade sobre a distribuição assintótica do estimador de máxima verosimilhança para justificar que
    $$\sqrt{n}\frac{\hat{\lambda} - \lambda}{\sqrt{\lambda}} \overset{a}{\sim} N(0,1)$$
    (f) Considere a informação e os dados do exercício 8. Apresente a estimativa de momentos e a estimativa de máxima versimilhança para λ. Calcule também a estimativa do respectivo erro padrão.

Solução: $\bar{X}$, –, $\frac{\lambda}{n}$, 2.16, 0.146969

18. Seja X uma população com distribuição exponencial de parâmetros (λ, δ) e uma amostra aleatória $(X_1, \ldots, X_n)$ desta população.
    (a) Deduza os estimadores de momentos para os parâmetros da distribuição $E(\lambda, \delta)$.
    (b) Suponha que o parâmetro δ tem um valor conhecido δ_0.
        i. Deduza o estimador de momentos para λ. Mostre que se trata de um estimador centrado e determine a respectiva variância.
        ii. Considere $\hat{\lambda} = \min(X_1, \ldots, X_n)$ outro estimador para λ. Sabendo que $\hat{\lambda}$ tem distribuição exponencial de parâmetros $(\lambda, \delta_0/n)$,
            A. Mostre que $\hat{\lambda}$ não é um estimador centrado para λ e indique o respectivo enviesamento.
            B. Determine o erro quadrático médio de $\hat{\lambda}$.
        iii. Compare o erro médio quadrático destes estimadores para λ e indique qual é o ´´melhor".
    (c) Considere a população X-tempo entre avarias consecutivas da máquina (em anos) apresentada no exercício 9.
        i. Admitindo que X tem distribuição $E(\lambda, \delta)$, apresente as estimativas de momentos para os seus parâmetros.
        ii. Admita agora que X tem distribuição $E(\lambda, 2)$. Apresente as estimativas para λ, resultantes da utilização do estimador de momentos e de $\hat{\lambda}$. Determine ainda o valor estimado do erro quadrático médio associado a cada estimativa.

Solução: $\bar{X} - \sqrt{\frac{n-1}{n}}S$, $\sqrt{\frac{n-1}{n}}S$, $\bar{X} - \delta_0$, $\frac{\delta_0^2}{n}$, $\frac{\delta_0}{n}$, $2\frac{\delta_0^2}{n^2}$, $\widehat{\lambda}$ é o ´´melhor"se $n > 2$, 1.676974, 5.205026, 4.9, 5, 0.08, 0.0032

19. Seja X uma população com distribuição $N(\mu, \sigma^2)$.
    (a) Determine os estimadores de momentos para os parâmetros μ e σ^2. Indique um estimador de momentos para σ.
    (b) Mostre que o estimador de momentos para μ é centrado e indique a respectiva variância.
    Mostre que o estimador para σ^2 não é centrado. Determine o seu enviesamento.
    (c) Considere a amostra apresentada no exercício 10. Admitindo que X-volume de água por garrafa (em litros) tem distribuição normal, apresente a estimativa de momentos para o volume médio de água por garrafa. Determine ainda uma estimativa centrada para a variância do volume de água por garrafa.

Solução: –, –, –, $\frac{\sigma^2}{n}$, $-\frac{\sigma^2}{n}$, 1494.465, 33.8348

20. Seja X uma população contínua com a seguinte função densidade de probabilidade:

$$f(x) = \begin{cases} \alpha x^{\alpha-1}, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}, \quad \alpha \in \mathbb{R}^+$$

   (a) Determine $E(X)$.
   (b) Deduza o estimador de momentos para α.
   (c) Mostre que a função verosimilhança para uma amostra $(x_1, \ldots, x_n)$ é:

$$L(\alpha) = \alpha^n \prod_{i=1}^{n} x_i^{\alpha-1}, \ \ \alpha \in \mathbb{R}^+$$

   (d) Deduza o estimador de máxima verosimilhança de α.
   (e) Verifique que $\sqrt{n}\,\dfrac{\hat{\alpha}-\alpha}{\alpha} \overset{a}{\sim} N(0,1)$.
   (f) Dada a amostra da população X,

0.23	0.69	0.62	0.09	0.68	0.05	0.69	0.32	0.87	0.05
0.74	0.17	0.57	0.93	0.67	0.45	0.21	0.86	0.70	0.41

   determine a estimativa de momentos para α e a estimativa de máxima verosimilhança de α.

Solução: $\frac{\alpha}{\alpha+1}$, $\frac{\bar{X}}{1-\bar{X}}$, 1

21. Considere a amostra aleatória $(X_1, X_2, \ldots, X_n)$ de uma população X com distribuição $B(1,p)$, em que p tem valor desconhecido.

$$P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}, \ \ x_i = 0, 1, \quad 0 < p < 1$$

   (a) Determine o estimador dos momentos de p.
   (b) Verifique que a função verosimilhança é:

$$L(p) = p^{n\bar{x}}(1-p)^{n-n\bar{x}}, \ \ 0 < p < 1$$

   e que a função log-verosimilhança é:

$$l(p) = n\bar{x}\ln p + n - n\bar{x}\ln(1-p), \ \ 0 < p < 1$$

   (c) Mostre que $\hat{p} = \bar{X}$ é o estimador de máxima verosimilhança de p.
   (d) Determine $E(\hat{p})$ e $V(\hat{p})$.
   Averigue a veracidade das seguintes afirmações:
   - $\hat{p}$ é estimador centrado de p.
   - $\hat{p}$ é estimador consistente de p.

   (e) Qual a distribuição exacta de $n\hat{p} = n\bar{X}$?
   (f) Invoque o T.L.C. para justificar o seguinte resultado:

$$\frac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \overset{a}{\sim} N(0,1)$$

Intervalos de confiança

22. (a) Determine um intervalo de confiança a 80% para o valor médio de uma variável aleatória com distribuição normal de variância igual a 4, com base numa amostra de valores $(9, 14, 10, 12, 7, 3, 11, 12)$.
    (b) Qual o coeficiente de confiança associado ao intervalo para o valor médio, de amplitude igual a 2.77?

    Solução: $[8.844903, 10.655097]$, 0.95

23. Certo equipamento automático encontra-se regulado para encher embalagens de um quilo de certo produto.

    O seu deficiente funcionamento origina prejuízos para a empresa: se a maioria das embalagens saírem com peso inferior ao estabelecido, haverá reclamações por parte dos clientes e perda de prestígio; se a maioria das embalagens saírem com peso excessivo, ter-se-á um fenómeno anti-económico.

    Aceita-se da experiência passada que o peso das embalagens se comporta normalmente com uma dispersão dada por $\sigma = 12$ gramas.

    Para verificar a afinação do equipamento, seleccionaram-se em certo período, nove embalagens cujos pesos exactos foram anotados (em gramas):

    983 992 1011 976 997 1000 1004 983 998

    (a) Determine estimativas por intervalo de confiança para o efectivo peso médio de uma embalagem, usando coeficientes de confiança de: 90%, 95% e 99%. Como varia a precisão do intervalo com o coeficiente de confiança escolhido?
    (b) Suponha que, em vez da amostra de pesos de 9 embalagens, tinha sido obtida uma amostra de pesos de 100 embalagens, e que esta fornecia uma média de 994 gramas. Estime o peso médio de uma embalagem por intervalo de 95% de confiança. Que ilação retira do aumento da dimensão da amostra?
    (c) Qual deverá ser o tamanho mínimo da amostra a recolher, de modo a que a amplitude do intervalo a 95% de confiança para o peso médio de uma embalagem seja de, no máximo, 2 gramas?
    (d) Admita agora que o desvio padrão σ do peso das embalagens tem um valor desconhecido. Determine estimativas por intervalo de confiança para o efectivo peso médio de uma embalagem, usando coeficientes de confiança de: 90%, 95% e 99%. Compare a amplitude destes intervalos com a amplitude dos intervalos obtidos na alínea a).

    Solução: $\bar{x} \approx 993.8$, $s^2 \approx 127.44$, $[987.24, 1000.36]$, $[985.96, 1001.64]$, $[983.48, 1004.12]$, $[991.648, 996.352]$, 554, $[986.80, 1000.80]$, $[985.12, 1002.48]$, $[981.18, 1006.42]$

24. Foi enviado um questionário a 12 firmas seleccionadas ao acaso de um certo sector industrial. De entre os resultados das 10 que responderam, saliente-se os seguintes referentes ao montante dispendido com acções de formação profissional do pessoal (valores em u.m. constantes):

Firma	A	B	C	D	E	F	G	H	I	J
5 anos atrás	31	18	9	43	28	54	34	21	14	23
Ano passado	33	14	22	56	27	57	29	36	16	21

    Por intervalo de 95% de confiança, estime a evolução do montante médio dispendido em formação profissional, no período decorrido. Suponha que os montantes dispendidos com acções de formação profissional têm distribuição normal.

    Solução: $[-8.914733, 1.714733]$

25. Uma empresa tem 500 cabos armazenados em más condições há 15 anos. Um ensaio em 40 deles, escolhidos ao acaso, apresentou uma média de tensões de ruptura de 2400 kg e um desvio padrão amostral de 150 kg.

    (a) Quais os limites de confiança a 95% para a avaliação da tensão média de ruptura da totalidade dos cabos armazenados?

(b) Com que confiança podemos dizer que a tensão média de ruptura dos 500 cabos armazenados tem valores dentro dos limites 2400 ± 35 kg?

Solução: $[2353.514518, 2446.485482]$, 0.8612

26. Seleccionaram-se aleatoriamente 101 clientes de um supermercado, tendo-se registado os respectivos tempos de compras (em minutos). Verificou-se que a média e a variância amostrais desses tempos foram 35 e 225, respectivamente.

    (a) Estime por intervalo de 95% de confiança, o tempo médio de compras, μ.

    (b) Determine uma estimativa por intervalo de 90% de confiança para o desvio padrão, σ, do tempo de compras, admitindo que a população em causa tem distribuição normal.

    Solução: $[32.06, 37.94]$, $[13.45413232, 16.99178174]$

27. Considere a informação prestada no exercício 8 e os resultados do exercício 17.

    (a) Estime por intervalo de 95% de confiança, o número médio de veículos que chegam ao cruzamento (por minuto).

        i. Usando como estatística pivot $\sqrt{n}\dfrac{\bar{X}-\mu}{S} \overset{a}{\sim} N(0,1)$.

        ii. Usando como estatística pivot $\sqrt{n}\dfrac{\hat{\lambda}-\lambda}{\sqrt{\lambda}} \overset{a}{\sim} N(0,1)$ (Resultado da alínea e) do exercício 17).

        iii. Compare os valores da amplitude dos dois intervalos.

    (b) Estime por intervalo de 95% de confiança, o desvio padrão do número de veículos que chegam ao cruzamento (por minuto).

    Solução: $[1.85424, 2.46576]$, –, $[1.361705, 1.570274]$, –

28. Considere a amostra apresentada no exercício 9 sobre os tempos entre avarias consecutivas da máquina. Admitindo que X-tempo entre avarias consecutivas tem distribuição exponencial de parâmetros $(\lambda, 2)$, considere o estimador de momentos para o parâmetro λ já estudado no exercício 18. Determine um intervalo de confiança de 95% para λ.

    Solução: $[4.327628, 5.436372]$

29. Uma cadeia de supermercados tenciona abrir uma nova loja num determinado bairro. Antes de ser tomada uma decisão, foi feito um estudo no âmbito do qual se entrevistaram 400 pessoas. Destas, 310 indicaram poder vir a utilizar regularmente os serviços do novo supermercado. Estime a proporção de pessoas deste bairro que poderão vir a ser clientes habituais da nova loja, através de um intervalo de 95% de confiança.

    Solução: $[0.734077, 0.815923]$

30. A FNN decidiu comprar fatos novos para os atletas. Adquiriu 6 fatos da marca mais cara (Tipo A) e 7 da marca mais barata (TIPO B) e enviou-os para um laboratório, onde se registaram os tempos de duração até romperem. Os registos, em horas, aparecem na tabela que se segue:

Tipo A:	1400	1725	1610	1605	1950	1575	
Tipo B:	1615	1665	1730	1755	1632	1606	1790

    Admitindo que o tempo de duração dos fatos para cada marca têm uma lei normal com a mesma variância, determine um intervalo de 95% de confiança para a diferença entre as durações médias dos fatos de cada marca.

    Solução: $[-205.395881, 124.395881]$

31. O comprimento de uma pequena partícula foi medido várias vezes por dois aparelhos de precisão (A_1 e A_2), tendo sido obtidos os seguintes resultados:

A_1 :	100	101	103	98	97	98	102	101	99	101
A_2 :	97	102	103	96	100	101	100			

    Admitindo a normalidade da lei dos comprimentos registados pelos dois aparelhos,

(a) Estime por intervalo de 95% de confiança, a diferença entre os comprimentos médios registados pelos aparelhos em cada partícula, supondo que o desvio padrão do comprimento medido pelos aparelhos A_1 e A_2 é, respectivamente 1 e 2.

(b) Estime por intervalo de 95% de confiança, o comprimento médio da partícula quando medido pela máquina A_1 e quando medido pela máquina A_2, admitindo que nada sabe sobre o desvio padrão das medições do comprimento.

Solução: $[-1.506039, 1.706039]$, $[98.609284, 101.390716]$, $[97.545642, 102.254358]$

32. Os dados abaixo indicados referem-se a duas amostras independentes da idade de aparecimento dos primeiros sintomas de cancro em homens e mulheres que sofrem desta doença.

Homens:	26	41	57	66	36	55	41	61	53	50	52	37	50
Mulheres:	58	52	50	49	56	52	54	48	41	37	67	70	

Supondo a normalidade das populações,

(a) Estime por intervalo de 90% de confiança, a variância da idade de aparecimento dos primeiros sintomas, para os homens e para as mulheres.

(b) Estime por intervalo de 90% de confiança, o desvio padrão da idade de aparecimento dos primeiros sintomas, para os homens e para as mulheres.

Solução: $[72.2419861, 290.6544202]$, $[49.38398983, 212.3781421]$, $[8.49952858, 17.048589977]$, $[7.02737432, 14.57319945]$

33. Considere o enunciado do exercício 13.

(a) Estime por intervalo de 95% de confiança, o valor médio de cada população.

(b) Determine uma estimativa por intervalo de 95% de confiança para o desvio padrão de cada população.

(c) Seja $p_X = P(X < 15)$ e $p_Y = P(Y < 15)$. Apresente estimativas pontuais e estimativas por intervalo de 90% de confiança para p_X e p_Y.

Solução: $15.479595 \leq \mu_X \leq 16.960405$, $13.468543 \leq \mu_Y \leq 15.851457$, $2.157934 \leq \sigma_X \leq 3.206003$, $3.472539 \leq \sigma_Y \leq 5.159088$, $\hat{p}_X = 0.26$, $\hat{p}_Y = 0.44$, $0.158267 \leq p_X \leq 0.361733$,$0.324872 \leq p_Y \leq 0.555128$

Exercício de desenvolvimento teórico

34. Admita que $(X_1, \ldots, X_n)$ é uma amostra aleatória de uma população X tal que $E(X) = \beta$ e $V(X) = \beta^2$.

(a) Considere a estatística $T = \frac{2}{\beta}\sum_{i=1}^{n} X_i$. Sabendo que T tem distribuição qui-quadrado com $2n$ graus de liberdade, deduza um intervalo de confiança $(1-\alpha)$ para β.

(b) Mostre, justificando cuidadosamente as suas deduções, que a estatística $E = \frac{\sum_{i=1}^{n} X_i - n\beta}{\sqrt{n}\,\beta}$ converge para uma variável aleatória Z com distribuição normal reduzida.

(c) Tendo em conta a estatística E da alínea anterior e a sua distribuição assintótica, deduza um intervalo assintótico de confiança $(1-\alpha)$ para β.

Teste de hipóteses

35. A temperatura média da água que é expelida de uma conduta da torre de refrigeração de uma central termo eléctrica não deve ultrapassar os 100^{o}F. Sabe-se que o desvio padrão da temperatura da água é de 2^{o}F. Foi medida a temperatura da água em 9 dias aleatoriamente seleccionados, e a média correspondente foi de 102^{o}F. A temperatura da água encontra-se dentro dos valores aceitáveis, ao nível de 5% de significância? Admita que a temperatura da água expelida tem distribuição normal.

    Solução: $z_o = 3$

36. A quantidade de lixo (toneladas) produzida no concelho do Xeisal, por dia, é uma variável aleatória com distribuição normal. De forma a avaliar o que se passa no concelho em relação a esta variável seleccionaram-se 15 dias ao acaso para os quais se registaram as correspondentes quantidades de lixo produzidas, resultando numa média de $\bar{x} = 100100$ toneladas e num desvio padrão amostral $s = 1117$ toneladas.

    (a) Assumindo que $\sigma = 1000$, teste a hipótese de que a quantidade média de lixo produzido por dia difere de 100000 toneladas. Use um nível de significância de 5%.

    (b) Teste a hipótese da alínea anterior, admitindo agora que não se conhece o valor de σ (Use um nível de 5% de significância).

    Solução: $|z_o| = 0.3873$, $|t_o| = 0.3467$

37. Uma empresa tenciona importar um grande número de instrumentos de precisão, de custo elevado. Os fabricantes garantem que o respectivo peso médio é de 100 g. Como a qualidade do produto depende essencialmente do seu peso, a administração da empresa solicita ao seu departamento técnico os esclarecimentos necessários sobre o peso médio, afim de decidir sobre a importação ou não. Para o efeito, o departamento técnico seleccionou aleatoriamente 49 instrumentos que forneceram os seguintes resultados referentes ao peso:

$$\sum_{i=1}^{49} x_i = 4410\, g \qquad \sum_{i=1}^{49} (x_i - \overline{x})^2 = 1080\, g^2$$

    Diga, justificando, qual a resposta que espera do departamento técnico? Considere o nível de significância de 5%.

    Solução: $|z_o| = 14.7573$

38. Um fabricante garante um mínimo de 5000 horas para a vida útil média de determinado componente. Ultimamente tem recebido reclamações de empresas clientes no sentido de que a referida duração média não tem atingido aquele valor. A fim de se certificar se as reclamações são de atender, o fabricante escolheu ao acaso 25 componentes e verificou que a sua vida útil média foi de 4940 horas com um desvio padrão de 200 horas.

    (a) Para $\alpha = 0.05$, que pode concluir o fabricante? Assuma que a população em causa tem distribuição normal.

    (b) E no caso de ter recolhido uma amostra de 40 componentes (com os mesmos resultados da anterior amostra para a média e desvio padrão)?

    Solução: $t_o = -1.5$, $z_o = -1.897367$

39. Considere a amostra tratada no exercício 8, relativa ao número de veículos que chegam ao cruzamento (por minuto). Admitindo que se vai decidir pela construção de uma rotunda caso o número médio de veículos que chegam ao cruzamento, seja superior a 3, diga qual a decisão a tomar? Considere $\alpha = 5\%$.

    Solução: $z_o = -5.378885$

40. Uma amostra aleatória de quantias gastas em prendas de Natal por 81 donas de casa (no último ano), apresenta uma média de 333 euros. Por estudos semelhantes realizados em anos anteriores, sabemos que esta população apresenta um desvio padrão de 94.5 euros. Tendo sido de 300 euros, o gasto médio registado em anos anteriores, podemos afirmar, para um nível de significância de 5%, que, neste último ano, se registou um aumento do valor desse gasto médio?

    Solução: $z_o = 3.142857$

41. Considere a amostra apresentada no exercício 9 sobre os tempos entre avarias consecutivas da máquina. Admitindo que X-tempo entre avarias consecutivas tem distribuição exponencial de parâmetros $(\lambda, 2)$, teste se o tempo médio entre avarias é inferior a 7.5 anos, com um nível de 10% de significância.

    Solução: $z_o = -2.184960$

42. Considere de novo o enunciado do exercício 24. Poderá afirmar, com 5% de significância, que se registou uma subida no montante médio dispendido em formação, durante o período de 5 anos?

    Solução: $t_o = -1.532194$

43. Considere a situação e a amostra apresentadas no exercício 11. Considere a v.a. $D = X - Y$ onde X representa o valor das acções do SCP e Y representa o valor das acções do FCP. Para a amostra em estudo, $\bar{d} = 0.090$ e $s_D^2 = 0.1213$. Teste se existem diferenças significativas entre o valor das acções dos dois clubes. Considere $\alpha = 5\%$.

    Solução: $|z_o| = 2.001648$

44. A cadeia de supermercados referida no exercício 29, abrirá uma nova loja se a proporção de pessoas do bairro que poderão vir a ser clientes habituais, se cifrar em mais de 80%. Face aos dados obtidos na sondagem, qual a atitude a tomar? Considere $\alpha = 5\%$.

    Solução: $z_o = -1.25$

45. No registo do peso (em kg) de 25 sacos de adubo, verificou-se que $\sum_{i=1}^{25} x_i = 1255$ e que $\sum_{i=1}^{25} x_i^2 = 63252$. Apresente uma estimativa pontual da variância do peso dos sacos de adubo. Admitindo que o peso dos sacos tem distribuição normal, teste a hipótese do peso ter um desvio padrão inferior a 3.5 kg, com um nível de 10% de significância.

    Solução: $s^2 = 10.45833$, $x_o^2 = 20.489796$

46. Considere de novo o exercício 31.

    (a) Teste, ao nível de 5% de significância, a hipótese da variância do comprimento medido pelo aparelho A_1 ser superior a 4.

    (b) Teste, ao nível de 1% de significância, se os aparelhos facultam diferentes valores para o comprimento médio da partícula (suponha que o desvio padrão do comprimento medido pelos aparelhos A_1 e A_2 é, respectivamente 1 e 2).

    Solução: $x_o^2 = 8.505$, $|z_o| = 0.12204$

47. Considere a informação prestada no exercício 30. Ao nível de 5% de significância, o que poderá concluir acerca da igualdade da duração média dos fatos de cada tipo?

    Solução: $|t_o| = 0.540587$

48. Para comparar dois tipos de máquinas A e B existentes numa fábrica, 100 dos muitos operários que aí trabalham, operaram uma máquina numa semana e a outra na semana seguinte. No final das duas semanas, os operários deram informação acerca da máquina em que preferiram trabalhar e 58 mostraram preferência pela máquina A. Estime por intervalo de 95% de confiança, a proporção de operários que preferem a máquina A. Poderá dizer que os operários não têm a mesma preferência pelas duas máquinas?

    Solução: $[0.483263, 0.676737]$, para um nível de 0.05 de significância não existem razões para dizer que a preferência não é igual

49. Dada a amostra apresentada no exercício 10 e relativa ao peso de 50 garrafas de água mineral, teste, para um nível de 5% de significância, se é possível concluir que o volume médio de água por garrafa é inferior ao volume prescrito para essas garrafas?

    Solução: $|z_o| = 6.728293$

50. O tempo, em segundos, que uma máquina leva a executar uma certa operação em cada peça produzida está sujeito a variações. O desvio padrão para o tempo de execução é conhecido e tem o valor de 2 segundos. Contudo é necessário saber se esses tempos de execução se distribuem de acordo com uma distribuição normal. Para o testar, foi recolhida uma amostra de 680 tempos de execução, que, depois de agrupados em classes, originaram a seguinte repartição de frequências absolutas:

Tempo (em segundos)	≤ 65	$]65, 67]$	$]67, 69]$	$]69, 71]$	$]71, 73]$	$]73, 75]$	≥ 75
Total de tempos	15	26	151	293	167	17	11

    Sabendo que a média amostral foi de $\bar{x} = 70$, teste a normalidade dos tempos de execução com um nível de 10% de significância.

    Solução: $x_o^2 = 63.5325$

51. Considere de novo os dados do exercício 10. Para o agrupamento em classes aí proposto, teste se a população X-peso por garrafa, tem distribuição normal, com um nível de 10% de significância.

    Solução: $x_o^2 = 5.6138$

52. Para a informação prestada no exercício 8, teste se X-número de veículos que chegam por minuto ao cruzamento, tem distribuição de Poisson. Considere $\alpha = 5\%$.

    Solução: $x_o^2 = 2.6170$

53. Teste, para um nível de 5% de significância, o pressuposto até agora adiantado acerca da população X-tempo entre avarias consecutivas da máquina, referida no exercício 9, ter distribuição exponencial de parâmetros $(\lambda, 2)$.

    Solução: $x_o^2 = 3.7347$

54. Será de admitir, com base na amostra a seguir, que o número de clientes atendidos por hora em certo posto de venda segue uma distribuição normal?

Valores observados em 30 horas

41	30	28	41	28	26	28	41	30	34
40	36	30	20	43	36	36	20	42	43
42	40	32	26	28	41	34	24	42	40

    Considere $\alpha = 5\%$ e os dados agrupados nas seguintes classes: $]-\infty, 24]$, $]24, 28]$, $]28, 32]$, $]32, 36]$, $]36, 40]$ e $]40, \infty]$.

    Trata-se de uma amostra recolhida aleatoriamente? (Use $\alpha = 10\%$).

    Solução: $x_o^2 = 5.3444$, $|t_o| = 1.819435$

55. Teste ao nível de 20% de significância a aleatoriedade das amostras apresentadas no exercício 32.

    Solução: Homens $t_o = 4$, Mulheres $t_o = 6$

56. Considere o enunciado do exercício 13 e a sua continuação, exercício 33.

    (a) É normalmente aceite que o tempo médio de execução do problema pelo algoritmo X é inferior ou igual a 16 segundos. Existem razões para duvidar deste pressuposto, para um nível de 5% de significância?

    (b) Teste, ao nível de 10% de significância, as seguintes hipóteses relativas ao desvio padrão do tempo de execução do algoritmo Y:

$$H_0 : \sigma_Y \geq 4 \ \ vs \ \ H_1 : \sigma_Y < 4$$

    (c) A proporção p_X é significativamente menor que 0.265? Considere $\alpha = 0.05$.

(d) Estará em causa a normalidade da distribuição do tempo de execução dos algoritmos? Considere $\alpha = 0.1$.

Solução: $t_o = 0.597115$, $x_o^2 = 53.82625$, $z_o = -0.08011$, $x_o^2 = 9.0981$, $x_o^2 = 2.8196$

Exercícios de desenvolvimento teórico

57. Considere o enunciado do exercício 34.
Suponha o seguinte teste de hipóteses sobre o parâmetro β,

$$H_0 : \beta = 1 \;\; vs \;\; H_1 : \beta \neq 1,$$

definido pela regra de rejeição

$$\text{Rejeitar } H_0 \text{ se } \left|\overline{X} - 1\right| > a,$$

sendo $\underline{a}$ uma constante real positiva e $\overline{X}$ a média da amostra aleatória $(X_1, X_2, \ldots, X_n)$. Aplicando o resultado da alínea anterior e, para n "suficientemente grande", determine o valor de $\underline{a}$ de modo a conseguir estabelecer a regra de rejeição de um teste assintótico com nível de significância $\alpha = 5\%$.

Solução: $\frac{1.96}{\sqrt{n}}$

58. Seja X uma população com distribuição exponencial de parâmetros $(\lambda, 2)$ e uma amostra aleatória $(X_1, \ldots, X_n)$ desta população.

Considere $\hat{\lambda} = \min(X_1, \ldots, X_n)$ um estimador para λ e tenha em conta que $\hat{\lambda}$ tem distribuição exponencial de parâmetros $(\lambda, 2/n)$.

Para teste das hipóteses $H_0 : \lambda \leq 5 \;\; vs \;\; H_1 : \lambda > 5$, considere a seguinte regra de rejeição:

$$\text{Rejeitar } H_0 \text{ se } \hat{\lambda} > 5.2.$$

(a) Verifique que a função potência deste teste é:

$$Q(\lambda) = P(\text{Rejeitar } H_0 \,|\, \lambda) = \begin{cases} e^{\frac{n}{2}(\lambda - 5.2)}, & \lambda \leq 5.2 \\ 1, & \lambda > 5.2 \end{cases}$$

(b) Mostre que o nível de significância deste teste é $\alpha = e^{-0.1\,n}$.

(c) Para um teste com nível de significância $\alpha = 0.05$, qual deverá ser a dimensão da amostra?

Solução: –, –, $n = 30$

Regressão linear

59. Determinada empresa está interessada em contabilizar o tempo que o ar condicionado está ligado no verão, por dia, mediante a temperatura exterior (em °C). Assim, seleccionaram-se 14 dias ao acaso, para os quais se mediram as temperaturas (x) e se registaram o número de horas de utilização do ar condicionado (Y):

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x_i	29	28	29	35	26	25	32	31	34	27	33	33	32	28
y_i	10.5	9.0	10.4	18.6	5.5	5.2	11.6	10.4	17.8	9.9	13.7	14.2	12.3	8.7

(a) Disponha os dados em gráfico.

(b) Estime a recta de regressão de mínimos quadrados.

(c) Comente a qualidade da estimação efectuada, com base no coeficiente de determinação.

(d) Teste, ao nível de 5% de significância, a hipótese de o verdadeiro declive da recta de regressão ser nulo. Comente o resultado à luz da alínea anterior.

(e) Para uma temperatura exterior de 30ºC, qual o número de horas que estima que o ar condicionado esteja a trabalhar? E para uma temperatura de 40ºC?

Solução: $\widehat{Y} = -24.026734 + 1.171029\,x$, $R^2 = 0.884933$, $|t_o| = 9.606601$, $\widehat{Y}\,(30) = 11.104139$

60. Pretende-se, se possível, modelar através de um modelo de regressão linear simples, a quantidade de vidro Y produzido num ecoponto (kg), usando como variável independente x, o número de dias sem despejar o mesmo. Para tal, registaram-se os seguintes dados:

i	1	2	3	4	5	6	7	8
x_i	2	3	4	5	10	15	20	25
y_i	100	150	250	320	650	810	1040	1480

(a) Escreva a recta de regressão estimada através do método dos mínimos quadrados. Determine o coeficiente de determinação. Acha que conseguiu um bom ajuste?

(b) Teste a hipótese de o número de dias justificar significativamente ($\alpha = 5\%$) a quantidade de vidro depositado.

(c) Estime por intervalo de 95% de confiança, a ordenada na origem da recta de regressão.

(d) Qual o valor esperado da quantidade de vidro produzida no ecoponto durante 10 dias sem o despejar? Seria possível calcular o mesmo para um período de 35 dias?

(e) Estime por intervalo de 90% de confiança, a variância do erro associado ao modelo.

(f) Estime por intervalo de 95% de confiança, a quantidade de vidro depositado no ecoponto durante 2 semanas sem o despejar.

Solução: $\widehat{Y} = 10.632184 + 56.130268\,x$, $R^2 = 0.985981$, $|t_o| = 20.542655$, $-77.943208 \leq \beta_0 \leq 99.207576$, $\widehat{E}\,(Y\,|10\,) = 571.934866$, $1856.984 \leq \sigma^2 \leq 14301.62$, $632.7482 \leq Y\,(14) \leq 960.1636$

61. Considere a amostra apresentada no exercício 12 sobre a duração, tipo e custo das chamadas telefónicas. Se admitir um modelo de regressão linear para x-duração e Y-custo, de cada chamada,

   (a) Estime a recta de regressão de mínimos quadrados.

   (b) Determine o coeficiente de determinação.

   (c) Teste a significância ($\alpha = 5\%$) de x-duração para a explicação do custo Y das chamadas.

   Solução: $\widehat{Y} = -0.038343 + 0.003873\, x$, $R^2 = 0.599382$, $|t_o| = 6.699576$

62. Uma equipa médica observou 10 doentes para relacionar o consumo diário médio de gordura saturada x (gramas) com o nível de colesterol Y (miligramas por decilitro). Os resultados obtidos mostram-se na tabela que se segue:

x	48	66	44	39	37	66	53	69	61	57
y	192	209	185	172	166	215	186	225	210	195

   (a) Esboce o gráfico dos dados amostrais.

   (b) Estime a recta de regressão de mínimos quadrados.

   (c) Estime a variância do erro associado ao modelo de regressão linear.

   (d) Comente a qualidade da estimação efectuada, com base no coeficiente de determinação.

   (e) Será que uma função linear descreve adequadamente a relação probabilística entre x e Y? Comente o resultado. Use $\alpha = 1\%$.

   (f) Interprete as estimativas dos parâmetros β_0 e β_1.

   (g) Estime pontualmente e por intervalo de 95% de confiança, o nível médio de colesterol de uma pessoa que consuma diariamente 50 gramas de gordura saturada.

   (h) Determine um intervalo de previsão a 95% de confiança para o nível de colesterol de uma pessoa que consuma diariamente 50 gramas de gordura saturada. Compare a amplitude deste intervalo com o intervalo para o nível médio de colesterol obtido na alínea anterior e interprete estes valores.

   Solução: $\widehat{Y} = 110.301964 + 1.577741\, x$, $\hat{\sigma}^2 = 29.576821$, $R^2 = 0.927828$, $|t_o| = 10.141348$, $184.971550 \leq E(Y\,|50) \leq 193.406519$, $175.957762 \leq Y(50) \leq 202.4203067$