UNIVERSIDADE DE ÉVORA
ESCOLA DE CIÊNCIAS E TECNOLOGIA

DEPARTAMENTO DE MATEMÁTICA

Introdução à Probabilidade e Estatística (MAT0925)

Exercícios e Soluções

Dulce Gomes e Patrícia Filipe

6ECTS, Carga horária semanal: 2h Teóricas + 2h Práticas

3º semestre 2016/17: Matemática Aplicada e Geologia
Docente: Russel Alpizar-Jara (gab.232, alpizar@uevora.pt)

2º semestre 2016/17: Engenharia das Energias Renováveis, Engenharia Geológica, Engenharia Informática e Engenharia Mecatrónica
Docentes: Patrícia Filipe (gab.235, pasf@uevora.pt) e Ana Isabel Santos (gab.241, aims@uevora.pt)

Alguns dos exercícios apresentados foram retirados ou adaptados dos manuais constantes na bibliografia (ver programa da unidade curricular). A elaboração desta lista de exercícios propostos e respectivas soluções teve a colaboração dos vários docentes que têm leccionado a unidade curricular de Introdução à Probabilidade e Estatística nos últimos anos. Agradecemos o contributo das professoras Graça Carita, Marília Pires, Telma Santos e Ana Isabel Santos. Agradecemos também aos alunos que contribuíram com sugestões e na detecção de gralhas. Este é um manual em costante evolução, pelo que, solicitamos que caso detetem gralhas ou queiram fazer alguma sugestão nos contactem por e-mail para dmog@uevora.pt ou pasf@uevora.pt.

Índice

1 Exercícios

1.1 *Estatística Descritiva*

1. O número de chamadas telefónicas recebidas, por minuto, numa determinada central telefónica foi registado durante um período de 50 minutos, observando-se os seguintes valores:

Descriptives	
N	50
Median	**C**
Variance	1,452
Range	**D**
Minimum	0
Sum	88

Nº de chamadas	Frequência
0	6
1	20
2	**A**
3	10
4	5
	B

   (a) Determine **A**, **B**, **C** e **D** e complete a tabela de frequências.

   (b) Determine e interprete a média, a moda e o desvio-padrão do número de chamadas recebidas por minuto;

   (c) Complete as seguintes frases:

      i. Em 18% dos minutos foram recebidas __ chamadas;

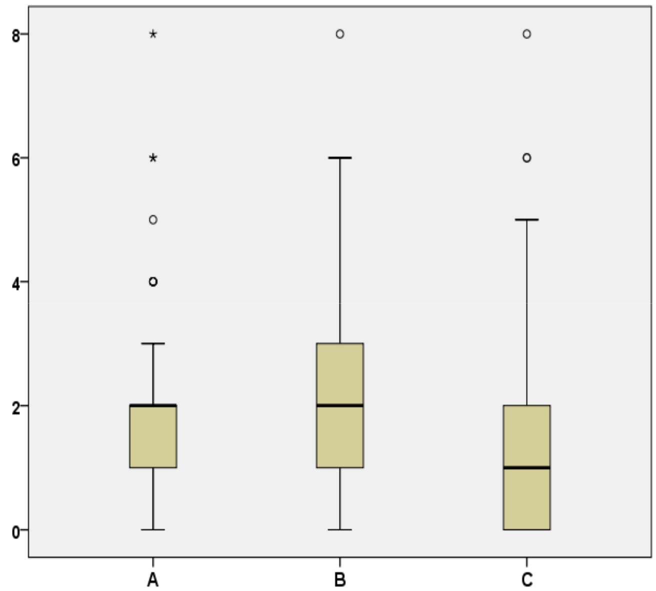
      ii. Em 30% dos minutos foram recebidas mais de __ chamadas;

      iii. Em __ minutos foi recebida no máximo 1 chamada.

   (d) Classifique a forma da distribuição dos dados quanto à assimetria e ao achatamento.

2. O responsável técnico de uma empresa pretende averiguar qual a origem das interrupções diárias do sistema. Entre outros dados, recolheu informação do número de interrupções diárias do sistema por avaria mecânica, ao longo de 70 dias. Alguns dos resultados obtidos foram os seguintes:

N		70
Mean		1,47
Median		1,00
Mode		0
Std. Deviation		1,759
Skewness		1,431
Std. Error of Skewness		,287
Kurtosis		2,140
Std. Error of Kurtosis		,566
Percentiles	25	,00
	50	1,00
	75	2,00
	85	3,00



(a) Interprete a média, a moda, a mediana e o desvio-padrão dos dados.

(b) Indique qual das boxplot (**A**, **B** ou **C**) corresponde à representação dos dados observados.

(c) Diga, justificando, se a seguinte afirmação é verdadeira: *"A percentagem de dias com mais de 4 interrupções do sistema por avaria mecânica é inferior a 15%"*

(d) Classifique a distribuição dos dados quanto ao tipo de assimetria e achamento. Justifique a sua resposta.

3. Um fabricante da indústria cerâmica pretende determinar se duas novas ligas *premium*, uma nacional e uma importada, possuem uma resistência ao calor superior à da liga *standard* já utilizada. Para tal foram realizados testes em que, para 20 fornadas de cada tipo de liga, se registou a temperatura máxima de resistência ao calor. Apresentam-se abaixo alguns dos resultados obtidos com a análise dos dados realizada com recurso ao *software* SPSS.

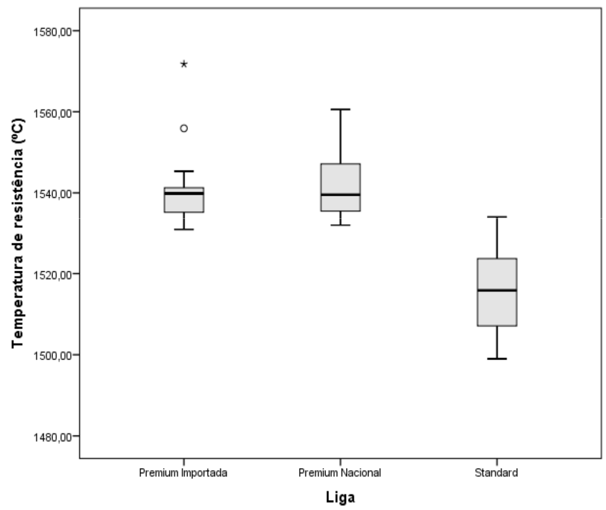
   (a) Interprete os valores da média, mediana e desvio-padrão observados para a temperatura de resistência da liga *premium* nacional.

   (b) O que pode dizer acerca da representatividade da temperatura de resistência média de cada uma das ligas?

   (c) O que sugere a representação gráfica acerca da forma (assimetria e achatamento) da distribuição das amostras correspondentes às três ligas?

Descriptives

	Liga		Statistic	Std. Error
Temperatura de resistência (ºC)	Premium Importada	Mean	1540,77	2,04
		Median	1539,85	
		Std. Deviation	9,10	
		Maximum	1571,77	
		Range	40,90	
		Skewness	2,38	,51
		Kurtosis	6,96	,99
	Premium Nacional	Mean	1541,48	1,68
		Median	1539,55	
		Variance	56,15	
		Minimum	1531,93	
		Maximum	1560,58	
		Skewness	0,94	,51
		Kurtosis	0,63	,99
	Standard	Mean	1516,05	2,26
		Median	1515,86	
		Std. Deviation	10,09	
		Minimum	1498,99	
		Range	35,06	
		Skewness	0,13	,51
		Kurtosis	-0,78	,99



4. Na seguimento do exercício anterior, são disponibilizados os seguintes dados da temperatura máxima de resistência ao calor registados nos testes das 20 fornadas da liga *premium* nacional:

Temperatura	f_i
$[1531.93; 1537.66[$	0.30
$[1537.66; 1543.39[$	0.40
$[1543.39; 1549.12[$	0.15
$[1549.12; 1554.85[$	0.10
$[1554.85; 1560.58]$	0.05

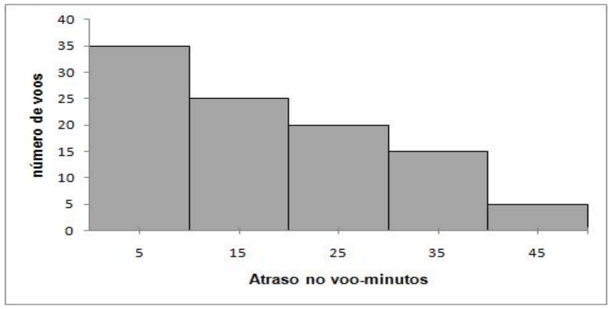
Com base nestes resultados,

(a) Calcule os valores da média, moda, mediana e desvio-padrão.

(b) Determine e interprete o 1^o quartil e o percentil 80.

(c) Através do cálculo de medidas adequadas, classifique a forma da distribuição dos dados quanto à assimetria e ao achatamento.

5. Um estudo sobre os atrasos nos 100 voos europeus durante um Verão, realizado em determinado aeroporto, conduziu à seguinte representação gráfica:



   (a) Construa a tabela de frequências a partir da informação fornecida pelo gráfico.
   (b) Calcule a variância referentes aos minutos de atraso dos voos.
   (c) O que poderá dizer quanto à assimetria dos dados? O que é que esta característica nos diz sobre os dados?

6. Foi feito um estudo de modo a avaliar quantas vezes durante uma tarde de estudo (5 horas) os alunos enviam e/ou recebem mensagens sms. Para tal, recolheu-se uma amostra aleatória de 100 alunos inscritos na disciplina de Introdução à Probabilidade e Estatística (IPE). O resultado deste estudo apresenta-se na tabela seguinte:

Nº de sms enviados	0	1	2	3	4	5	6	7	8
frequências relativas	0,01	0,2	0,25	0,03	0,28	0,12	0,05	0,04	0,02

   (a) Construa a tabela de frequências associada a estes dados.
   (b) Represente graficamente (usando a frequência relativa) estes dados.
   (c) Determine, para este conjunto de dados,:
      i. a média;
      ii. a moda e mediana
      iii. o desvio-padrão.
      iv. o 1º quartil.
      v. o percentil 75.
      vi. o 8º decil.
   (d) Através do cálculo do grau de assimetria de Bowley o que poderá dizer quanto à assimetria dos dados?
   (e) Qual a proporção de alunos que enviam
      i. pelo menos 4 sms?

ii. entre 2 a 5 sms?

7. Devido não só ao aumento dos preços dos preços combustíveis mas também a preocupações ambientalistas, muitos portugueses têm optado por transformar os seus automóveis e instalar GPL (este contribui para ajudar na preservação do meio ambiente, visto ser um combustível muito menos poluente, quando comparado com outros combustíveis). Uma grande empresa de instalação de GPL em veículos automóveis, com 1000 funcionários, pretende medir o tempo gasto na instalação de GPL. Para tal seleccionou uma amostra de 200 funcionários e registou os tempos gastos por cada um na instalação de GPL num veículo automóvel. Os resultados obtidos foram os seguintes:

Tempo (em horas)	X_i'	F_i
$[3, 10[$	6,5	0,015
$[10, 12[$	11	0,03
$[12, 15[$	13,5	0,07
$[15, 20[$	17,5	0,245
$[20, 30[$	25	0,695
$[30, 60[$	45	0,945
$[60, 120]$	90	1

(a) Complete a tabela de frequências associada a estes dados.

(b) Represente, como achar mais conveniente, esta colecção de dados.

(c) Determine a média e o desvio-padrão associados a esta colecção de dados.

(d) Determine ainda a classe modal e a classe mediana.

8. O encarregado do controlo de qualidade de uma fábrica de relógios digitais observou durante 200 dias amostras de 5 relógios, tendo registado o número de relógios defeituosos por amostra:

N.º de defeituosos, por amostra	0	1	2	3	4	5
Frequências	53	68	44	17	16	2

(a) Construa a tabela de frequências associada a este conjunto de dados.

(b) Calcule e interprete a média, a mediana, a moda e o desvio-padrão do número de relógios defeituosos, por amostra.

(c) Determine os 1º, 2º e 3º quartis. Interprete cada um dos valores obtidos.

(d) Determine e interprete P10 e P75 (percentis 10% e 75%).

(e) Numa outra fábrica de relógios (B) obteve-se uma média igual a 2,702 relógios defeituosos por amostra e uma variância igual a 1,6. Qual das duas fábrica apresenta resultados mais homogéneos?

9. Numa empresa recolheu-se uma amostra aleatória relativa à produção de energia eléctrica em KW/h em dois tipos de geradores, I e II. Admita que a produção de energia de cada gerador segue uma distribuição normal e que $\sigma_I^2 = \sigma_{II}^2$. Os resultados amostrais foram os seguintes:

Gerador tipo I (n=27)	15.01	3.81	2.74	16.82	14.30	13.45	8.75
	9.40	16.84	17.21	2.74	4.91	5.05	9.72
	9.02	12.31	14.10	9.64	10.21	10.34	9.04
	5.02	10.59	11.91	9.44	7.21	11.07	
Gerador tipo II (n=23)	10.87	8.07	10.31	11.08	10.84	6.34	10.05
	9.37	8.94	8.78	15.01	6.93	15.91	13.45
	6.84	9.37	10.04	10.94	2.04	16.89	14.04
	4.32	10.71					

(a) Complete a tabela de frequências associada à amostra do gerador I.

Classes	n_i
	6
	1
	11
	5
	4
	27

(b) Relativamente ao gerador I, determine a média considerando os dados agrupados e não agrupados. Compare os valores obtidos.

(c) Compare, através de uma caixa-de-bigodes, estas duas colecções de dados.

(d) Calcule, para os dados não agrupados:

i. duas medidas de localização e uma medida de dispersão que lhe pareçam ser as mais adequadas para ambos os conjunto de dados. Interprete estes valores.

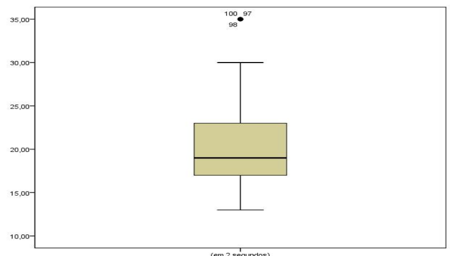
ii. Determine, apenas para o gerador I, o quantil 85% e interprete este valor.

10. De modo a estudar a performance dos atletas de alta competição do Eborense Futebol Clube foram seleccionados, ao acaso, 100 atletas e registado o número de metros percorridos em 2 segundos. Os resultados obtidos apresentam-se na seguinte tabela de frequências:

X_i'	F_i
14,6	0,24
17,8	0,53
21	0,71
24,2	0,89
27,4	0,92
30,6	0,96
33,8	1

(a) Complete a tabela de frequências associada a estes dados.

(b) Dada a representação abaixo, indique quais as medidas de localização e as medidas de dispersão adequadas para caracterizar esta colecção de dados.



(c) O que pode dizer sobre a forma (apenas em termos de assimetria) desta colecção de dados?

11. Tal como no caso do exercício anterior, pretendia estudar-se a performance dos atletas, agora não profissionais, do Eborense Futebol Clube. Para tal, foram seleccionados, ao acaso, 25 atletas e registado o número de metros percorridos em 5 segundos. Os resultados obtidos encontram-se resumidos no quadro seguinte:

N	Valid	25
	Missing	0
Mean		60,6400
Median		65,0000
Mode		57,00a
Std. Deviation		16,88411
Variance		285,073
Skewness		-,660
Std. Error of Skewness		,464
Kurtosis		-,078
Std. Error of Kurtosis		,902
Range		68,00
Minimum		20,00
Maximum		88,00
Percentiles	10	35,6000
	25	49,0000
	30	54,0000
	50	65,0000
	70	71,2000
	75	72,5000

a. Multiple modes exist. The smallest value is shown

(a) Determine o coeficiente de Bowley e interprete-o.

(b) O treinador destes atletas afirma que 30% dos atletas efectuam mais de x metros em 5 segundos. De quantos metros está o treinador a falar?

12. De modo a estudar a rapidez com que os alunos resolviam problemas de estatística, foi recolhida aleatoriamente uma amostra de 25 estudantes de várias licenciaturas (com a mesma matéria e os mesmos professores), aos quais foi apresentado um exercício para resolverem. Os dados que se seguem dizem respeito ao tempo (em minutos) que os alunos levaram a resolver o exercício proposto;

26	17	37	38	33	29	63	43	47	19	49	41	
21	53	59	42	60	58	40	48	44	54	47	31	32

(a) Construa a tabela de frequências associada a estes dados.

(b) Construa o histograma e o polígono de frequências correspondente à distribuição de frequências relativas.

(c) Determine a média e o desvio-padrão deste conjunto de dados.

(d) Através do coeficiente de assimetria (G_1) o que poderá dizer quanto à assimetria dos dados?

13. Foi feito um inquérito a um grupo de compradores de 40 carros novos para determinar quantas reparações ou substituições de peças foram feitas durante o primeiro ano de utilização dos carros. Obtiveram-se os seguintes resultados:

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

(a) Construa um quadro de distribuição de frequências absolutas.

(b) Calcule as frequências relativas.

(c) Construa um gráfico para as frequências absolutas.

(d) Calcule as frequências acumuladas.

(e) Construa um gráfico de frequências acumuladas.

14. Num estudo para analisar a capacidade de germinação de certo tipo de cereal foram semeadas cinco sementes em cada um dos vasos dum conjunto de vasos iguais, contendo o mesmo tipo de solo, e registou- se o número de sementes germinadas. Os resultados obtidos foram os seguintes:

N.º de sementes germinadas por vaso	0	1	2	3	4	5
Nº de vasos	16	32	89	137	98	25

(a) Calcule a média, a mediana e a moda do número de sementes germinadas.

(b) Represente graficamente os resultados.

(c) Calcule a proporção de vasos com mais de três sementes germinadas.

(d) Determine os 1º, 2º e 3º quantis. Interprete cada um dos valores obtidos.

(e) Determine e interprete P10 e P75 (percentis 10% e 75%).

(f) Calcule o desvio padrão e o coeficiente de variação.

15. Os dados que se seguem referem-se ao comprimento total (em cm) de uma colecção de dados de achigãs de uma barragem:

29.9	40.2	37.8	19.7	30.0	29.7	19.4	39.2	24.7	20.4
19.1	34.7	33.5	18.3	19.4	27.3	38.2	16.2	36.8	33.1
41.4	13.6	32.2	24.3	19.1	37.4	23.8	33.3	31.6	20.1
17.2	13.3	37.7	12.6	39.6	24.6	18.6	18.0	33.7	38.2

(a) Ordene os dados e calcule a média, mediana, desvio padrão, quartis e o quantil de ordem $2/3$. Encontre um valor tal que 70% dos peixes observados tenham comprimentos superiores a esse valor.

(b) Faça um agrupamento dos dados em classes, de forma conveniente, e represente-os graficamente.

(c) Calcule a média e variância para os dados classificados. Compare estes valores aproximados com os correspondentes valores exactos obtidos em (a). Indique a classe modal e interprete o seu significado.

16. O departamento de pessoal de uma certa firma fez um levantamento dos salários dos 120 funcionários do sector administrativo, obtendo os seguintes resultados:

Faixa Salarial (Nº de salários mínimos)	Frequência Relativa
$[0, 2[$	0,25
$[2, 4[$	0,40
$[4; 6[$	0,20
$[6, 10[$	0,15

(a) Esboce a representação gráfica apropriada a este conjunto de dados.

(b) Calcule a média e o desvio padrão.

(c) Se for concedido um aumento de 100% para todos os funcionários haverá alteração na média? E na variância? Justifique convenientemente a sua resposta.

(d) Se for concedido um abono de 2 salários mínimos para todos os 120 funcionários, haverá alteração na média? Justifique.

17. Foram medidas a altura e o peso de um grupo de homens e de um grupo de mulheres. O quadro seguinte contém a média e o desvio padrão dos dados observados:

	Altura (cm)		Peso (Kg)	
Grupos	$\overline{x}$	s	$\overline{x}$	s
Homens	173	3,2	71	3,2
Mulheres	158	3,2	54	2,7

(a) Calcule o coeficiente de variação da altura das mulheres. Suponha que a altura é medida em milímetros. Qual o novo coeficiente de variação? Justifique a resposta.

(b) Compare a dispersão da altura e do peso dentro de cada grupo e entre os dois grupos de indivíduos.

18. No quadro seguinte indicam-se os preços dum bem alimentar (em unidades monetárias) praticado durante 12 meses consecutivos e as quantidades vendidas.

Preço	110	90	80	76	74	71	70	65	63	60	55	50
Vendas	55	70	90	100	90	105	80	110	125	115	130	131

a) Represente graficamente a informação disponibilizada.

b) Através da análise gráfica, parece-lhe existir relação linear entre o preço e as quantidades vendidas observadas?

c) Calcule e interprete o valor do coeficiente de correlação amostral.

1.2 Exercícios de Probabilidades

1. Na empresa Sojoga foi decidido que nos próximos meses será desenvolvida uma aplicação para Smartphones que permita decidir qual o melhor caminho para que um indivíduo se desloque de um qualquer ponto A para um outro qualquer ponto B, tendo em conta o tráfego em tempo real. Para tal, o primeiro passo foi fazer um inquérito de grande dimensão para perceber qual a receptividade da respectiva aplicação. Nesse inquérito, 72% das pessoas afirmaram que iriam comprar a aplicação, e destas 27% afirmaram registar-se no site da empresa (para ter acesso a informações sobre as futuras actualizações da aplicação). Tendo em conta os resultados, e admitindo que 37% das pessoas se iriam registar no site, diga qual a probabilidade de uma pessoa, seleccionada ao acaso,

   (a) não comprar a aplicação?

   (b) comprar a aplicação e registar-se no site?

   (c) registar-se no site, sabendo que não comprou a aplicação?

2. Estudos recentes revelaram que, em Portugal, no 4º trimestre de 2012, do total de edifícios licenciados, 56% correspondiam a construções novas e, destas, 70% destinavam-se a habitação familiar.

   (a) Qual a probabilidade de um edifício licenciado ser antigo?

   (b) Qual a probabilidade de um edifício licenciado ser novo e não ser destinado a habitação familiar?

3. Estudos disponíveis em www.pordata.pt, acerca do tráfego telefónico nacional (rede móvel e rede fixa), revelaram que 75% dos minutos em chamadas têm origem em rede móvel, destes 2.7% têm como destino a rede fixa. Sabe-se ainda que 23% têm origem em rede fixa e como destino rede fixa.

   (a) Qual a probabilidade de um minuto de chamada ser realizado a partir da rede fixa?

   (b) Determine a probabilidade de um minuto em chamada, seleccionado aleatoriamente, ter como destino a rede fixa.

   (c) Sabendo que um minuto em chamada foi realizado para rede fixa, qual a probabilidade de ter tido origem em rede fixa também?

4. Uma fábrica dispõe de 3 máquinas para produção de peças electrónicas. A máquina 1 é responsável por 60% da produção e as máquinas 2 e 3 produzem em igual percentagem. Testes realizados às peças produzidas por cada uma das máquinas revelaram que, da produção da máquina 1, 1% das peças falham na verificação de todos os requisitos de qualidade, enquanto que da produção das máquinas 2 e 3, falham nos requisitos 2% e 4%, respetivamente.

(a) Qual a probabilidade de uma peça, seleccionada ao acaso de toda a produção, não falhar nos requisitos de qualidade?

(b) O proprietário da fábrica encontra-se numa situação em que tem de vender uma das máquinas. Qual a máquina que aconselharia o proprietário a vender?

5. Para estudarem para uma frequência, sabe-se que 40% dos alunos fizeram todos os exercícios propostos, cerca de 55% assistiram a mais de 9 aulas e de entre estes últimos cerca de 64% fizeram todos os exercícios propostos. Calcule a probabilidade de um aluno escolhido ao acaso:

   (a) não ter feito todos os exercícios propostos.

   (b) ter feito todos os exercícios propostos e ter assistido a mais de 9 aulas.

   (c) não ter feito todos os exercícios propostos nem ter assistido a mais de 9 aulas.

   (d) não ter feito todos os exercícios propostos e ter assistido a mais de 9 aulas.

   (e) ter assistido a mais de 9 aulas, sabendo que não realizou todos os exercícios propostos.

6. Numa determinada região 528 indivíduos foram classificados segundo duas características qualitativas. Sejam A_1, A_2, A_3 e B_1, B_2 e B_3 os atributos possíveis, respectivamente, na 1ª e 2ª classificação. Os resultados da dupla classificação dispõem-se a tabela:

	B_1	B_2	B_3
A_1	56		12
A_2	47	163	38
A_3	14	42	85

   Escolhido um indivíduo ao acaso, calcule a probabilidade de:

   (a) Preencha a célula vazia e explique o significado desse valor;

   (b) ter o atributo B_2;

   (c) ter os atributos A_3 e B_1;

   (d) ter os atributos A_1 ou B_3;

   (e) ter o atributo B_1, sabendo que possui o atributo A_3;

   (f) ter o atributo A_2, de entre os que possuem o atributo B_2.

7. Considere dois centros de Sondagens, centro XP e centro XPTO. Sabe-se que o centro XP acerta nas previsões em 75% dos casos e o centro XPTO em 55%. Em 20% dos casos nenhum dos centros acerta nas suas previsões.

(a) Calcule a probabilidade de ambos os centros acertarem nas suas previsões.

(b) Calcule a probabilidade de um centro acertar e o outro falhar.

(c) Calcule a probabilidade de pelo menos um dos centros acertar nas suas previsões.

(d) Sabendo que o centro XPTO acertou na sua previsão, qual a probabilidade do centro XP ter falhado?

8. Um conjunto de três atelier de Arquitectura (digamos atelier A, B e C) pretende elaborar conjuntamente um projecto para apresentar na Trienal de Arquitectura 2007, a realizar no Pavilhão de Portugal entre os dias 31 de Maio e 31 de Julho de 2007. A participação na Trienal de Arquitectura 2007 depende do cumprimento de todos os seguintes acontecimentos:

A="A tarefa do atelier A é executada a tempo"

B="A tarefa do atelier B é executada a tempo"

C="A tarefa do atelier C é executada a tempo".

Suponha que estes acontecimentos são independentes e com probabilidades iguais a 0.8, 0.7 e 0.9, respectivamente.

(a) Calcule a probabilidade do projecto final estar terminado a tempo de poder participar na Trienal de Arquitectura 2007.

(b) Calcule a probabilidade do atelier A ter cumprido a tempo a sua tarefa e pelo menos um dos outros ateliers o não ter conseguido.

(c) Sabendo que a tarefa do atelier C foi cumprida a tempo, qual a probabilidade do projecto final estar terminado a tempo de poder participar na Trienal de Arquitectura 2007?

9. Do conjunto de empresas que actuam num dado sector industrial, 25% possuem departamento de investigação, 50% realizam lucros e 20% possuem departamento de investigação e realizam lucros. Pretende-se calcular a probabilidade de uma empresa escolhida ao acaso estar nas seguintes condições:

(a) possuir departamento de investigação ou realizam lucros ou ambos;

(b) não possuir departamento de investigação;

(c) não possuir departamento de investigação nem realizar lucros;

(d) não possuir departamento de investigação ou não realizar lucros;

(e) possuir departamento de investigação e não realizar lucros;

(f) não possuir departamento de investigação e realizar lucros.

10. Em determinada população existem 60% mulheres das quais 6% têm mais de 170 cm de altura e existem 40% de homens dos quais 2% têm mais de 170 cm de altura. Seleccionou-se uma pessoa dessa população e verificou-se que tinha mais de 170 cm de altura. Qual a probabilidade de ser mulher?

11. Sejam A e B dois acontecimentos tais que $P(A) = \frac{3}{8}$, $P(B) = \frac{1}{2}$ e $P(A \cap B) = \frac{1}{4}$. Calcule:

    (a) $P(A \cup B)$;

    (b) $P(\overline{A})$;

    (c) $P(\overline{B})$;

    (d) $P(\overline{A} \cap \overline{B})$.

12. Sejam A e B dois acontecimentos tais que $P(A) = \frac{1}{3}$ e $P(A \cup B) = \frac{1}{2}$. Calcule $P(B \cap \overline{A})$.

13. Sabendo que $P[A \cap (B \cup C)] = 0,2$; $P(\overline{A}) = 0,6$ e que $P(\overline{A} \cap \overline{B} \cap \overline{C}) = 0,1$, calcule $P(B \cup C)$.

14. Considere os acontecimentos A, B e C de probabilidades não nulas. Sabendo que C é mutuamente exclusivo quer com A quer com B; dois dos acontecimentos são independentes; $P(A) = 0,2$; $P(C) = 0,15$ e $P(A \cap B) = 0,06$. Calcule $P(A \cup B \cup C)$.

15. Numa fábrica, um certo tipo de peças para um equipamento de laboratório é embalado em caixas e por uma das 3 linhas de produção diferentes: L1, L2, L3. Os registos mostram que as percentagens das caixas que não são embaladas em condições próprias sabendo que provêm de L1, L2 e de L3 são, respectivamente, de 3%, 3,5% e 2%. Sabe-se ainda que a percentagem de caixas embaladas por cada uma das linhas de produção é de 40%, 25% e 35%, respectivamente.

    (a) Qual a probabilidade de uma caixa, escolhida ao acaso, não estar em condições?

    (b) Sabendo que uma caixa foi devidamente embalada, qual a probabilidade de ser proveniente da linha de produção L2?

16. Num laboratório um investigador fez um ensaio com 3 classes de bactérias A, B e C, na proporção de 0.10, 0.30 e 0.60 de cada classe, respectivamente. As bactérias da classe A reagem na presença de sulfatos em 80% dos casos, as da classe B em 60% e as da classe C em 40%.

    (a) Qual a probabilidade de uma bactéria escolhida ao acaso ser da classe A e reagir na presença de sulfatos?

(b) Qual a probabilidade de uma bactéria escolhida ao acaso não reagir na presença de sulfatos?

(c) O investigador ao escolher uma bactéria verificou que ela não reagiu na presença de sulfatos. Qual a probabilidade desta pertencer à classe B?

17. O Nunes consegue à última da hora um bilhete para assistir às primeiras Meias-Finais do Campeonato Europeu de Futebol de 2008, que está a decorrer na Austria e na Suiça. Para se deslocar de casa até ao estádio de St.Jakob-Park, em Basileia, o Nunes tem apenas três meios de transporte distintos à sua escolha, T_1, T_2 e T_3. Sabe-se que a probabilidade de

    usar o meio de transporte T_1 é de 0,7:

    chegar atrasado, sabendo que utilizou o transporte T_1 é 0,8;

    chegar atrasado, sabendo que utilizou o transporte T_2 é 0,5;

    chegar a horas, sabendo que utilizou o transporte T_3 é 0,6;

    usar o meio de transporte T_2 é igual à probabilidade de o meio de transporte T_3.

    (a) Sabendo que o Nunes chegou atrasado ao jogo, calcule a probabilidade de ter utilizado o transporte T_2?

    (b) Qual a probabilidade do Nunes chegar atrasado ao jogo?

18. Numa determinada fábrica do Distrito de Évora, três inspectores o Sertório, o Geraldes e o Gastão, inspeccionam respectivamente, 20%, 30% e 50% dos artigos produzidos por essa fábrica. Sendo um artigo inspeccionado pelo inspector Sertório a probabilidade deste ser um artigo defeituoso é de 5%; é de 10% se tiver sido pelo inspector Geraldes e é de 15% se tiver sido pelo inspector Gastão.

    (a) Retirando ao acaso um artigo inspeccionado, qual a probabilidade deste ser defeituoso?

    (b) Sendo o artigo inspeccionado um artigo defeituoso, por qual dos três inspectores foi mais provável este artigo defeituoso ter sido inspeccionado? Apresente todos os cálculos.

19. Numa urbanização recente um inquérito aos moradores revelou que 5% viviam em moradias, 20% em prédios em banda e os restantes em torres. Alguns desses moradores foram aí instalados através de programas de realojamento. Dos moradores que vivem em moradias 2% são realojados, o mesmo acontecendo com 3% dos que vivem em prédios em banda e com 10% dos que vivem em torres. Escolhido ao acaso um dos habitantes dessa urbanização, qual a probabilidade de:

(a) Ele ter sido alvo do programa de realojamento?

(b) Ele viver numa moradia sabendo que se trata de um realojado?

20. A poluição do ar de certa cidade é causada essencialmente por gases industriais (75% dos casos) e por gases dos escapes de automóveis (25% dos casos). Nos próximos 4 anos prevê-se que a possibilidade de controlar com sucesso a poluição dado que provêem dessas duas fontes de poluição são de 70% e de 60%, respectivamente.

    (a) Qual a probabilidade de haver controlo bem sucedido da poluição do ar dessa cidade (i.e. controlar com sucesso pelo menos uma das fontes de poluição) nos próximos 4 anos?

    (b) Constatando-se que houve controlo bem sucedido de poluição do ar ao fim de 4 anos, qual a probabilidade de isso ser devido ao controlo de gases dos escapes de automóveis?

21. Durante a travessia do Canal da Mancha, a probabilidade de um velejador apanhar mau tempo é de $2/3$. Sabe-se ainda que, se estiver mau tempo, tem $1/4$ de probabilidade de ter uma colisão com um petroleiro, mas, não estando mau tempo, a probabilidade de atravessar o Canal da Mancha sem colidir é de $5/6$. Face a uma futura viagem, determine a probabilidade:

    (a) De vir a atravessar o Canal da mancha sem colidir.

    (b) Não apanhar mau tempo caso venha a colidir com um petroleiro.

1.3 Variáveis Aleatórias

1. Sejam X e Y duas variáveis aleatórias, em que X toma valores de 2, 3 e 4 e $Y = 3X$. Sabe-se que $P(X = 2) = P(X = 4) = 0.3$.

   (a) Determine as funções massa de probabilidade de X e de Y.

   (b) Construa a função distribuição de Y.

   (c) Determine $P(X \leq 3)$, $P(Y = 10)$, $P(Y > 7)$ e $P(Y \leq 9|X \geq 3)$

   (d) Construa a função massa de probabilidade conjunta de (X, Y).

   (e) Determine o valor esperado e variância para cada uma das variáveis.

   (f) Calcule e interprete o valor do coeficiente de correlação de (X, Y).

2. Sejam X e Y duas variáveis aleatórias. Sabe-se que X é uma variável aleatória, com valor esperado 2.21, cuja função de probabilidade é dada por:

x	1	2	3
$f(x)$	0,32	A	B

   A variável aleatória Y, pode assumir valor 2 ou 5, sendo que a P[$Y > 2$]= 0.85. O par aleatório (X, Y) possui a seguinte função massa de probabilidade conjunta:

$X \backslash Y$	2	5
1	**C**	0.27
2	0	3**C**
3	2**C**	0.43

   (a) Determine **A**, **B** e **C**. *Caso não consiga obter estes valores, use as letras para as alíneas seguintes.*

   (b) Determine a função distribuição da variável Y.

   (c) Calcule o valor médio e o desvio-padrão da variável Y.

   (d) Calcule a probabilidade de $Y = 5$ condicionada a $X = 2$.

   (e) Sabendo que $Y = 2$, calcule o desvio-padrão da $2X$.

   (f) As variáveis X e Y são independentes?

3. Sejam X e Y duas variáveis aleatórias independentes. Ambas as variáveis podem assumir valor 2, 3 ou 5. Diga, justificando convenientemente, se as seguintes afirmações são verdadeiras ou falsas:

   (a) Se $P[X = 2] = 0.7$ e $P[X = 2, Y < 3] = 0,35$ então $P[Y = 2] = 0.5$.

   (b) Se $P[X = 5|Y \geq 3] = 0.1$ então $P[X = 5] = 0.2$.

   (c) $E[X|Y = 3] = 2.5$ logo $E[4X + 12] = 22$.

(d) $Var[8X - Y] = 8Var[X] + Var[Y]$

(e) Se E[XY]=2,73 então $Cov[X, Y] = 3.47$.

4. Na tabela seguinte apresenta-se a função de probabilidade relativa ao número de exercícios (X) que um aluno tenta resolver durante a aula prática de uma determinada disciplina.

x	0	1	2	3	4
$f(x)$	0,2	0,1	0,4	0,1	0,2

(a) Determine a função distribuição do número de exercícios que os alunos tentam resolver numa aula.

(b) Qual a probabilidade de durante uma aula um aluno tentar resolver no máximo 2 exercícios?

(c) Sabendo que um aluno tentou resolver menos de 2 exercícios qual a probabilidade de não ter tentado resolver qualquer exercício?

(d) Considerando $Y = 15 - 4X$, calcule $E[Y]$ e $Var[Y]$.

5. Suponha que numa determinada cidade, o número de assoalhadas por casa apresenta a seguinte distribuição:

x	1	2	3	4
$f(x)$	0,4	A	0,2	B

(a) Tendo em conta que a probabilidade de uma casa ter no máximo 2 assoalhadas é de 0.7 determine A e B;

(b) Determine a função distribuição do número de assoalhadas por casa;

(c) Qual a probabilidade do número de assoalhadas por casa ser no máximo 3?

(d) Qual a probabilidade do número de assoalhadas por casa ser de pelo menos 2?

(e) Determine e interprete o valor esperado e o desvio-padrão do número de assoalhadas por casa.

6. Seja X uma variável aleatória à qual corresponde a seguinte função massa de probabilidade:

$$f(x) = \begin{cases} \dfrac{x+a}{20}, & x = 1, 2, 3, 4, 5 \\ 0, & \text{restantes casos} \end{cases}$$

(a) Mostre que, para que $f(x)$ seja uma função massa de probabilidade, $a = 1$.

(b) Qual a probabilidade da variável aleatória tomar quanto muito o valor 2?

(c) Calcule o valor esperado e a variância de $2X - 1$.

7. Seja X a variável aleatória que representa o número de vezes que cada aluno das várias Licenciaturas da Universidade de Évora se matriculam na disciplina de IPE. Admita que X tem a seguinte função massa de probabilidade:

$$f(x) = \begin{cases} \dfrac{10 - x^2}{16}, & x = 1, 2, 3 \\ 0, & \text{restantes casos} \end{cases}$$

(a) Qual a probabilidade de um aluno das várias Licenciaturas da Universidade de Évora fazer a disciplina de IPE à primeira?

(b) Calcule a probabilidade de um aluno se inscrever quanto muito 2 vezes à disciplina.

(c) Determine a função de distribuição da variável X e represente-a graficamente.

(d) Determine o número médio de vezes que um aluno se matricula a IPE.

(e) Prove que $Var[X - 1] = Var[X]$.

(f) Calcule $Var[2X - 1]$.

8. A fim de analisar a capacidade de germinação de sementes em certo cereal foram semeadas cinco sementes em cada vaso de um conjunto de vasos iguais, contendo o mesmo tipo de solo. Registou-se, então, o número de sementes germinadas. Obtiveram-se os seguintes resultados:

Nº de sementes germinadas por vaso	0	1	2	3	4	5
Nº de vasos	16	32	87	137	98	25

Seja X a variável aleatória que indica o número de sementes germinadas, por vaso.

(a) Defina a sua função massa de probabilidade.

(b) Calcule a probabilidade de terem germinado quanto muito três sementes.

(c) Calcule a probabilidade de terem germinado entre duas a cinco sementes, inclusive.

(d) Calcule a probabilidade de terem germinado mais de uma e menos de cinco sementes.

(e) Calcule a probabilidade de terem germinado menos de quinze sementes?

x_i	-2	-1	0	1	2
$f(x_i)$	12/60	15/60	10/60	6/60	m

9. Considere a variável aleatória discreta X que assume os valores $\{-2, -1, 0, 1, 2\}$ cuja função massa de probabilidade é:
   (a) Determine o valor de m de modo que $f(x_i)$ seja uma função massa de probabilidade;
   (b) Seja $Y = X^3 + 3X^2 + 2X$. Construa a função massa de probabilidade de Y e calcule $P(Y > 0)$.

10. Seja X uma dada variável aleatória contínua com a seguinte função densidade de probabilidade
$$f(x) = \begin{cases} ax, & 0 \leq x < 1 \\ a, & 1 \leq x < 2 \\ -ax + 3a, & 2 \leq x < 3 \\ 0, & \text{restantes casos} \end{cases}$$
   (a) Mostre que para $f(x)$ seja uma função de densidade, $a = 1/2$.
   (b) Calcule $P[X < 1]$ e $P[1 < X \leq 3]$.

11. Seja X uma variável aleatória que associa a cada indivíduo o tempo necessário para completar um determinado teste padrão. Constatou-se que X pode tomar valores entre 50 e 70 minutos e definiu-se a sua função densidade de probabilidade do seguinte modo:
$$f(x) = \begin{cases} \frac{1}{20}, & 50 \leq x \leq 70 \\ 0, & \text{caso contrário} \end{cases}$$
   (a) Determine a função de distribuição de X.
   (b) Calcule a probabilidade de um indivíduo demorar entre 61 a 70 minutos a completar o teste.
   (c) Determine o tempo médio necessário para completar o teste.

12. A distribuição conjunta das variáveis aleatórias, independentes, X e Y é dada pela seguinte tabela:

$X \backslash Y$	0	1	2	$f(x)$
1	0,06		0,02	0,2
2	0,15		0,05	
3				0,3
$f(y)$	0,3		0,1	

   (a) Complete a tabela.
   (b) Indique a função distribuição da variável aleatória X.
   (c) Calcule $E(X|Y = 2)$.
   (d) Seja $Z = 2X - 4Y$, calcule $E(Z)$ e $Var(Z)$.

13. Suponha que numa determinada cidade, o número de filhos por família (X) e o número de assoalhadas por casa (e por família) (Y) apresentam a seguinte distribuição conjunta:

$X \backslash Y$	1	2	3
2	0,08	0,05	0
3	0,15	0,15	0,05
4	0,17	0,10	0,25

   (a) Determine a função massa de probabilidade do número de filhos por família;
   (b) Calcule a função de distribuição do número de assoalhadas por casa;
   (c) Qual a probabilidade do número de assoalhadas por casa ser de pelo menos 2?
   (d) Calcule a probabilidade de uma família morar numa casa de 2 assoalhadas, sabendo que tem 3 filhos;
   (e) Se uma família vive numa casa com 3 assoalhadas, quantos filhos se espera que essa família tenha?

14. Considere que a editora "D. Quixote e Sr. Pança"possui duas máquinas impressoras. Cada uma destas máquinas trabalha em conjunto na impressão de uma determinada obra (quer isto dizer que cada máquina imprime, em separado mas em simultâneo, vários exemplares da mesma obra). No final da impressão, verifica-se que cada obra pode apresentar no máximo 3 pequenos defeitos.

   Considere duas variáveis aleatórias X e Y, onde X representa o número de defeitos em cada obra impressa e Y o tipo de máquina . Seja $f(x, y)$ a função massa de probabilidade conjunta do par (X, Y), dada por:

$Y\backslash X$	0	1	2	3
1	0.125	0.0625	0.1875	0.125
2	0.0625	0.0625	0.125	0.25

(a) Defina a função massa de probabilidade marginal da variável X.

(b) Calcule a função de distribuição conjunta no ponto $(1, 2)$.

(c) Seleccionou-se aleatoriamente uma obra do conjunto das obras impressas e verificou-se que esta não apresentava qualquer defeito. Qual a probabilidade desta obra ter sido impressa na máquina 1?

15. Sejam X e Y duas variáveis aleatórias discretas que representam, respectivamente, o número de avarias graves num dado veículo automóvel antes da instalação de GPL e depois da instalação de GPL. De um estudo prévio, sabe-se que a função de probabilidade conjunta do par aleatório (X, Y) é dada por

$X\backslash Y$	0	1	2
0	a	0.3	b
1	c	0.2	d

e que a função de distribuição marginal da variável Y dada por:

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ 0.3, & 0 \leq y < 1 \\ 0.8, & 1 \leq y < 2 \\ 1, & y \geq 2 \end{cases}$$

(a) Complete a tabela sabendo ainda que $E(XY) = 0,4$ e que $E(X^2) = 0,4$.

Nota: Se não resolveu a alínea (a) considere nos cálculos seguintes as letras a, b, c e d.

(b) Calcule a probabilidade de um dado veículo não ter tido nenhuma avaria grave antes da instalação de GPL e ter tido quanto muito uma avaria grave depois da instalação de GPL.

(c) Sabendo que um dado veículo teve duas avarias graves depois de ter instalado o GPL, calcule a probabilidade de antes não ter tido qualquer avaria grave.

(d) Será que o número de avarias graves num dado veículo automóvel antes e depois da instalação do GPL estão relacionados? Justifique e comente o valor obtido.

(e) Calcule $Var[Y|X = x]$.

16. Uma possível codificação de um jogo de computador é a seguinte:
O jogador escolhe ao acaso uma das três opções $(-1, 0, 1)$ e
O computador responde através da escolha, também aleatória, de outras três opções $(0, 1, 2)$.
O resultado do jogo é determinado pelo produto XY, onde X representa a escolha do jogador e Y a escolha do computador.

    (a) Escreva as funções massa de probabilidade marginais da variáveis X e Y.

    (b) Calcule o valor esperado da variável X.

    (c) Determine a função massa de probabilidade conjunta do par (X, Y), assumindo que
    $P[X = -1, Y = 1] = P[X = 1, Y = 2] = 1/6$
    e que
    $P[X = 1, Y = 1] = P[X = -1, Y = 2] = 1/9$

    (d) Calcule a variância de XY.

    (e) Determine a função massa de probabilidade de Y condicionada a $X = 0$.

17. Com o objectivo de analisar o tipo de precaução que os alunos de outras licenciaturas, que não a de Eng. Informática, tomam relativamente ao seu computador foi feito um estudo sobre não só o número de mensagens electrónicas marcadas por mês como *junk* antes de serem apagadas, como também o número de vezes, igualmente num mês, que foram feitas actualizações do anti-virus.

    Seja X a variável aleatória que representa o número de mensagens marcadas como *junk* e Y o número de vezes que o anti-virus foi actualizado.

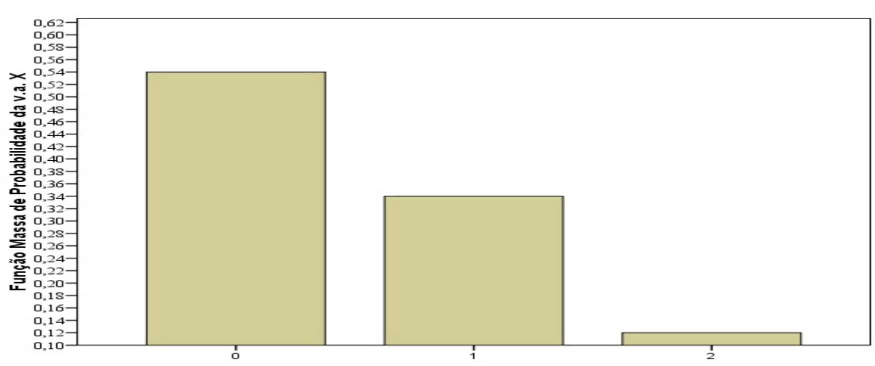


Figura 1: Função Massa de Probabilidade Marginal da v.a. X

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ 0.05, & 0 \leq y < 1 \\ 0.15, & 1 \leq y < 2 \\ 0.6, & 2 \leq y < 3 \\ 1, & y \geq 3 \end{cases}$$

Tendo em conta toda a informação fornecida anteriormente sobre as variáveis aleatórias X e Y e considerando que estas são independentes, responda às questões que se seguem.

Nota: Diz-se que duas variáveis aleatórias X e Y são independentes

se $f(x,y) = f_X(x) \times f_Y(y)$, $\forall (x,y) \in \mathbb{R}^2$
ou se $F(x,y) = F_X(x) \times F_Y(y)$, $\forall (x,y) \in \mathbb{R}^2$,

onde $f(x,y)$ representa a função massa de probabilidade conjunta do par (X,Y) e $F(x,y)$ representa a função de distribuição conjunta do par (X,Y).

(a) Determine a função de distribuição marginal da variável X e represente-a graficamente.

(b) Calcule a função de distribuição conjunta no ponto $(1,1)$.

(c) Qual a probabilidade da razão entre o número de mensagens marcadas como *junk* e o número de vezes que o anti-virus foi actualizado ser superior a 1?

(d) Calcule a variância da variável X.

(e) Calcule o valor esperado da variável $W = X + 2Y$.

(f) Calcule $E[XY]$.

18. Voltemos à questão colocada na frequência anterior, em que se pretendia analisar o tipo de precaução que os alunos de outras licenciaturas (que não a de Eng. Informática) tomavam relativamente ao seu computador. Para tal, foi feito um estudo sobre o número de mensagens electrónicas marcadas por mês como *junk* antes de serem apagadas e o número de vezes, igualmente num mês, em que foram feitas actualizações do anti-virus.

Seja X a variável aleatória que representa o número de mensagens marcadas como *junk* e Y o número de vezes que o anti-virus foi actualizado. Considere a seguinte função massa de probabilidade conjunta:

(a) Sabendo que um aluno escolhido ao acaso marcou durante um mês duas mensagens como *junk*, determine a função massa de probabilidade da variável Y.

$X\backslash Y$	0	1	2	3	
0	0,027	0,054	0,243	0,216	0,54
1	0,017	0,034	0,289		0,34
2		0,012	0,06	0,048	0,12

(b) Calcule $P(\frac{Y}{X} > \frac{1}{2}|X > 1)$.

(c) Calcule o valor esperado de Y^2, sabendo que $X = 2$.

(d) Sabendo que $E[XY] = 1,164$, qualifique a existência de relação, caso esta exista, entre as variáveis X e Y.

(e) Determine a covariância entre Z e W, sabendo que $Z = 2X$ e $W = -Y$.

19. Sejam X e Y duas variáveis aleatórias com função massa de probabilidade conjunta $f(x,y)$, da forma:

$$f(x,y) = k(2x+y),\ x = 0,1,2 \quad \text{e} \quad y = 0,1,2,3.$$

Calcule

(a) o valor da constante k.

(b) $P(X \geq 1, Y \leq 2)$.

(c) a função massa de probabilidade da v.a. X condicional a $Y = 0$.

(d) $Var[X|Y = 0]$.

20. Sejam X e Y duas variáveis aleatórias discretas. Sabe-se que X é uma variável aleatória discreta cuja função de distribuição é da forma:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 0,3, & 0 \leq x < 1 \\ 0,5, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

Considere ainda o par aleatório (X,Y) com a seguinte função massa de probabilidade conjunta

$X\backslash Y$	3	4	5
A	D	0,14	0,1
B	0,04	E	0,05
C	F	0,3	0,2

(a) Defina a função massa de probabilidade da variável X e represente-a graficamente.

(b) Complete a tabela. Ou seja, determine os valores de A, B, C, D, E e F.

(c) Sabendo que a variável X toma o valor C, calcule a probabilidade da variável Y tomar o valor 4.

(d) Calcule a variância da variável $Z = X - Y$.

(e) Sabendo que $X = 1$, calcule o valor esperado de Y^2.

21. O Evaristo pode encontrar-se num dos três estados de espírito: **1** - radiante; **2** - assim-assim; **3** - macambúzio. Seja X a variável aleatória que define o estado de espírito do Evaristo, cuja função massa de probabilidade é apresentada na tabela seguinte:

x_i	1	2	3
$f_X(x_i)$	0,2	0,3	0,5

Seja Y a variável aleatória que define o estado do tempo: **0** - Dia chuvoso com vento; **1** - Dia ameno de Outono; **2** - Dia de Primavera. Na tabela seguinte apresenta-se a função massa de probabilidade da variável X condicionada a $Y = 0$.

x_i	1	2	3
$f(x\|y=0)$	0,68	0,3	0,02

Sejam X e Y as duas variáveis aleatórias discretas anteriormente definidas, com função massa de probabilidade conjunta dada por:

$X \backslash Y$	0	1	2	
1	0,068	0,08	0,052	
2	**a**	0,11	**b**	
3	**c**	0,01	**d**	
	0,1	0,2	0,7	1

(a) Preencha a tabela da função massa de probabilidade conjunta.

Nota. A não resposta à alínea a) não implica a não realização das restantes alíneas. Caso não consiga responder à alínea a) use, sempre que necessário, as letras a, b, c, e d para responder às questões seguintes.

(b) Determine a função de distribuição da variável Y.

(c) Calcule $E[XY]$.

(d) Calcule a função de distribuição conjunta no ponto $(3, 1)$.

(e) Determine $Var[X|Y = 0]$.

1.4 Distribuições de Probabilidade

1. Estudos recentes revelaram que, em Portugal, no 4º trimestre de 2012, do total de edifícios licenciados, 56% correspondiam a construções novas e, destas, 70% destinavam-se a habitação familiar.

   (a) Qual a probabilidade de um edifício licenciado ser antigo? *(Ver Ficha n.º2)*

   (b) Qual a probabilidade de um edifício licenciado ser novo e não ser destinado a habitação familiar? *(Ver Ficha n.º2)*

   (c) Analisados ao acaso 8 licenciamentos, determine a probabilidade de pelo menos 6 corresponderem a construções novas.

2. A energia eólica representa o aproveitamento da energia cinética contida no vento para produção de energia mecânica (rotação das pás) que pode ser transformada em energia eléctrica por um gerador eléctrico. A energia eólica é produzida por um aerogerador, usualmente constituído por uma torre de 50 a 120 m de altura, que possui no topo um rotor, com 3 pás (de 25 m a 45 m cada), e a chamada "nacelle" que abriga o gerador e os sistemas de controlo da máquina (http://www.eneop.pt). Admita que o vento faz as pás darem em média 18 voltas por minuto.

   (a) Qual a probabilidade das pás darem 25 voltas num minuto?

   (b) Determine o desvio-padrão do número de voltas dadas pelas pás em 5 minutos.

   (c) Qual a probabilidade de, em 5 minutos, o número de voltas ser no mínimo 100?

3. Seja X uma variável aleatória que representa a produção anual de energia eléctrica a partir de fontes de energia eólica (em GWh). Considere que a probabilidade da produção anual de energia eléctrica proveniente de fontes de energia eólica exceder 7000 GWh é de 75.8% e que esta variável segue uma distribuição normal com uma variância de 4 000 000 GWh^2.

   (a) Determine o valor da produção média anual de energia eléctrica proveniente de fontes de energia eólica.
   (Caso não consiga resolver esta alínea considere para este valor 8400 GWh.)

   (b) Determine a probabilidade da produção anual de energia eléctrica proveniente de fontes de energia eólica:

      i. ser de pelo menos 7500 GWh;

      ii. estar entre 8500 GWh e 9200 GWh.

4. Uma determinada empresa pretende recrutar funcionários para uma nova área de negócio que pretende desenvolver. Para se candidatar aos lugares disponíveis é necessário realizar um teste. Sabe-se que os resultados do teste seguem uma distribuição normal de média 100 e variância 225. De entre os requisitos para contratação do funcionário é exigida a classificação mínima no teste de 115. Houve 60 candidatos aos lugares disponíveis pela empresa.

   (a) Qual a probabilidade de um candidato, seleccionado ao acaso, verificar o requisito exigido relativamente à classificação no teste?

   (b) Qual a probabilidade de um qualquer candidato obter uma classificação no teste entre 94 e 118?

   (c) Qual a percentagem de candidatos que se espera atingirem uma classificação acima de 100?

   (d) Determine a probabilidade de, entre os candidatos, em pelo menos 17 se verificar o requisito exigido relativamente à classificação no teste.

5. Uma fábrica produz discos de travão para automóveis de alta cilindrada. O diâmetro dos discos segue uma distribuição normal de média 322 mm e variância 0.0025 mm^2. O departamento de controlo de qualidade da fábrica avalia regularmente se o diâmetro dos discos produzidos estão dentro dos valores de qualidade exigidos.

   (a) Seleccionado aleatoriamente um disco de travão produzido na fábrica, calcule a probabilidade do seu diâmetro:

      i. exceder os 322.07 mm;

      ii. estar entre 321.95 mm e 322.02 mm;

   (b) Complete a frase: *"Em 93.7% dos casos o diâmetro do disco de travão não ultrapassa os __ mm"*.

   (c) Na fábrica existem 8 máquinas de produção de discos de travão. Uma máquina é inspeccionada sempre que produz discos que apresentem um diâmetro fora do intervalo [321.95; 322.05]. Calcule a probabilidade de pelo menos 2 máquinas serem inspeccionadas.

6. Uma conhecida fábrica de cerâmica produz azulejos que são embalados em caixas com quinze unidades. Sabe-se que o número de azulejos com defeitos segue uma distribuição de Poisson de variância igual a 2.

   (a) Qual a probabilidade de um qualquer azulejo ser defeituoso?

   (b) Qual a probabilidade de no máximo 2 azulejos serem defeituosos?

(c) Numa caixa escolhida aleatoriamente, qual a probabilidade de todos os azulejos se encontrarem em bom estado (ou seja, sem qualquer defeito)?

7. O número de viaturas que atestam o depósito de gasolina numa pacata vila alentejana numa manhã (5h), de um certo dia, é bem representado por uma distribuição de Poisson de média 5.1. Considere também que o número de viaturas que atestam o depósito de gasolina na mesma vila alentejana numa tarde (5h), de um certo dia, é bem representado por uma distribuição de Poisson de média 8.4. Considere ainda que as viaturas que atestam o depósito de manhã não o fazem à tarde.

   (a) Qual a probabilidade de, em duas horas da parte da manhã, quanto muito uma viatura atestar o depósito de gasolina na vila alentejana?

   (b) Em média, quantas viaturas atestam o depósito de gasolina na pacata vila alentejana durante todo o dia (10h)?

   (c) Qual a probabilidade de, num dia (10h), 15 viaturas atestarem o depósito de gasolina na vila alentejana?

8. Sabe-se que o número de alunos que chegam, por hora, para tirar dúvidas com a docente A duas semanas antes de uma frequência é uma v. a. X com distribuição Poisson com variância igual a 1.

   (a) Determine o número médio de alunos que chegam por "dia de atendimento"com a docente A. (Considere que o "dia de atendimento"é constituído por 4h).

   (b) Qual a probabilidade de chegarem no máximo 2 alunos por "dia de atendimento"com a docente A?

   (c) Considere agora que existem duas docentes (a A e a B) para tirar dúvidas e que o número de alunos que chegam, por "dia de atendimento"(4h), para tirar dúvidas com a docente B (duas semanas antes de uma frequência) é uma v. a. Y com distribuição Poisson com média igual a 2. Calcule a probabilidade do total de alunos que aparecem em 5 "dias de atendimento"seja de quanto muito 20 alunos.

9. Um dos elevadores dum grande edifício público transporta, no máximo 20 pessoas de cada vez. A carga máxima transportada pelo elevador é de 1300kg. Os utilizadores deste elevador pertencem a um largo extracto duma população em que se verificou que o peso duma pessoa é aproximadamente normal com média 61Kg e desvio-padrão 10Kg.

(a) Qual a probabilidade de haver nas 20 pessoas, que em certo momento viajam no elevador, quanto muito duas com peso superior a 40Kg e inferior a 85Kg?

(b) Calcule a probabilidade do peso médio destes 20 utilizadores exceder os 60Kg.

10. No âmbito de um estudo sobre a criação de um novo Centro Comercial na cidade de Évora, constatou-se que 40% dos cidadãos residentes na cidade de Évora concordam com esta criação. Inquiridos 10 cidadãos residentes na cidade de Évora, calcule:

    (a) a probabilidade de nenhum cidadão discordar com a criação de um novo Centro Comercial na cidade de Évora.

    (b) a probabilidade de pelo menos 2 cidadãos concordarem com a criação de um novo Centro Comercial na cidade de Évora

    (c) o número médio de cidadãos que discordam da criação de um novo Centro Comercial na cidade de Évora.

    (d) o desvio padrão do número de cidadãos que concordam com a criação de um novo Centro Comercial na cidade de Évora.

11. Um exemplo clássico da distribuição de Poisson envolve o número de militares do exército Prussiano mortos entre 1875 e 1894 devido a um coice de cavalo. Sabendo que a probabilidade de num ano ter morrido no mínimo um militar devido a um coice de cavalo é igual a 0.864 (i.e. $P(X \geq 1) = 0.864$), calcule:

    (a) a probabilidade de em 2 anos ter morrido 1 ou mais militares devido a um coice de cavalo.

    (b) o número esperado de militares mortos num mês devido a um coice de cavalo.

    (c) a probabilidade de entre 1875 e 1894 (20 anos) terem morrido menos de 60 militares devido a um coice de cavalo.

12. Numa pequena e média empresa de fornecimento de materiais de construção, sabe-se que as vendas diárias de areia (em toneladas) têm um comportamento aleatório, traduzido por uma distribuição normal, com média 2 e variância 0.04.

    (a) Qual a probabilidade de a empresa vender, diariamente, mais do que 2.5 toneladas de areia?

    (b) Qual a probabilidade de a empresa vender, diariamente, entre 1.5 e 2.5 toneladas de areia?

    (c) Determine a quantidade mínima de areia que é vendida diariamente em 50% casos.

(d) Qual a probabilidade de numa qualquer semana (6 dias) escolhida ao acaso, se ter pelo menos 2 dias em que as vendas de areia não tenham ultrapassado as 2.5 toneladas?

13. Através de estudos realizados, é possível admitir que ocorre no Japão (na província de Tóquio) uma média de 2 sismos por mês. Admitindo que o número de sismos que ocorre na província de Tóquio, por mês, segue uma distribuição de Poisson,

    (a) Determine:

        i. a probabilidade de ocorrer um sismo num mês.

        ii. a probabilidade de ocorrer quanto muito 5 sismos em 3 meses.

    (b) Admitindo agora que na província de Shizuoka ocorre uma média de 3 sismos de dois em dois meses e que os sismos registados em ambas as províncias (de Tóquio e de Shizuoka) são independentes entre si e ambos seguem uma distribuição de Poisson, determine a probabilidade de nos próximos 2 anos o número de sismos em ambas as zonas (ou seja, total) ter excedido os 90.

14. Considere que a quantidade de gasolina vendida numa manhã (5h) numa pacata vila alentejana é bem descrita por uma distribuição normal de média $30l$ e variância $3l^2$ e que a quantidade vendida numa tarde (5h), nessa mesma vila, é também bem descrita por uma distribuição normal de média $45l$ e variância $16l^2$. Considere ainda que quem se abastece de gasolina de manhã não o faz à tarde.

    (a) Qual a probabilidade da quantidade de gasolina vendida numa manhã na vila alentejana ser inferior a $25l$?

    (b) Qual a probabilidade de num dia serem vendidos mais de $80l$ de gasolina?

    (c) Em 10 viaturas escolhidas ao acaso das que abastecem durante o dia (10h), qual a probabilidade de pelo menos 2 delas abastecerem entre $65l$ e $80l$?

15. Um posto de transformação permite uma carga total de 2800kW. Sabe-se que esse posto de transformação alimenta uma fábrica com um consumo permanente de 2500kW. Além disso, alimenta uma população de 100 consumidores domésticos, gastando cada um, em média, 2kW (com desvio-padrão de 0.5kW) em electrodomésticos e, em média, 0.5kW (com desvio-padrão de 0.25kW) em iluminação, podendo estes dois tipos de consumo ser considerados independentes. Admita que ambos os consumos seguem uma distribuição normal.

    (a) Calcule a probabilidade de um consumidor (escolhido ao acaso de entre os 100) gastar em electrodomésticos entre 1.5 e 3 kW.

    (b) Calcule a probabilidade de um consumidor (escolhido ao acaso de entre os 100) gastar em iluminação mais de 0.95 kW.

(c) Sendo $W_i, i = 1, 2, ..., 1000$ o consumo total de electricidade (em kW) do consumidor doméstico, tal que $W_i = X_i + Yi$, $i = 1, 2, ..., 100$ (em que X_i representa o consumo em electrodomésticos e Y_i representa o consumo em iluminação), obtenha, para o total dos consumidores, a média e a variância do consumo total de electricidade.

(d) Calcule o valor de x tal que $P[X < x] = 0, 15$.

16. A resistência ao choque de certo tipo de mosaico cerâmico tem, em unidades convenientes, distribuição Normal de média igual a 3 e variância igual a 0.25.

    (a) Qual a probabilidade de um desses mosaicos ter uma resistência entre 2,5 e 3,5?

    (b) Qual a maior resistência que 20% dos mosaicos apresenta?

    (c) Em 10 mosaicos escolhidos ao acaso de um grande lote, qual a probabilidade de pelo menos 9 terem uma resistência superior a 3,5?

17. Numa fábrica existem 15 máquinas, 8 são novas e 7 são antigas. Sabe-se que a probabilidade de uma máquina antiga estar em mau estado de conservação é 0,5; enquanto que numa nova essa probabilidade é 0,1.

    Para avaliar o estado de conservação das máquinas escolheram-se, ao acaso, 4 máquinas para inspecção.

    (a) Qual a probabilidade de terem sido escolhidas pelo menos 2 máquinas antigas?

    (b) Qual a probabilidade da 1ª máquina escolhida para inspecção estar em mau estado de conservação?

18. Os portugueses são dos cidadãos europeus com menor envergadura física, tendo em média menos 4,4 centímetros e menos 3,2 quilos que o cidadão médio da União Europeia, revela um estudo encomendado pela Comissão Europeia.

    De acordo com os dados do Eurobarómetro sobre Saúde, Alimentação e Nutrição, os portugueses têm em média 1,65 metros de altura e pesam 69 quilos, o que os coloca no penúltimo lugar quer da tabela de *altura* (apenas à frente dos malteses), quer do *ranking* do *peso* (apenas com italianos atrás).

    Suponha que a variável *altura* dos portugueses segue uma distribuição normal.

    (a) Sabendo que a probabilidade de encontrar um português que meça menos de 1,70 metros é 0,75, determine a variância da distribuição.

    (b) Calcule a probabilidade de encontrar um português que meça mais de 1,49 metros mas menos de 1,73 metros?

(c) Qual a probabilidade de que, em 10 portugueses escolhidos ao acaso e com reposição, pelo menos 2 tenham mais de 1,70 metros?

19. Estudos tecnológicos para a qualificação de certo tipo de rochas naturais e/ou ornamentais, tendo em vista a sua aplicação industrial, usam um certo tipo de laser. Estudos efectuados permitiram aceitar com válida a hipótese do tempo de vida de um laser ter uma distribuição normal de média igual a 7000h e um desvio padrão igual a 600h.

    (a) Qual a probabilidade do tempo de vida de um laser, escolhido ao acaso, ser quanto muito de 5300 horas de funcionamento?
    (b) Qual a duração que 90% desses lasers excede?
    (c) Qual a probabilidade de em 10 lasers escolhidos ao acaso 4 deles apresentarem um tempo de vida superior a 5300h?

20. O número de fendas significativas numa auto-estrada, a ponto de exigirem reparação imediata, tem distribuição de Poisson com desvio padrão igual a 2 falhas por Km.

    (a) Qual é a probabilidade de não haver fendas que exijam reparação imediata em 3 Km de auto-estrada?
    (b) O estado de uma auto-estrada é considerado grave se num percurso de 3 Km forem encontradas 8 fendas que exijam reparação imediata. Foram escolhidas aleatoriamente 10 auto-estradas com tráfego semelhante. Calcule a probabilidade de 5 delas se encontrarem em estado grave.
    (c) Devido à má construção das auto-estradas verificou-se (passados 4 anos após a conclusão das mesmas) uma média de 25 fendas por Km a exigirem reparação imediata. Perante este cenário, calcule a probabilidade de num Km existirem menos de 18 fendas.

21. A pluviosidade anual, medida em cm^3/m^2, numa determinada região tem sido estudada e revelou um comportamento distribucional normal com média $65\,cm^3/m^2$ sendo possível concluir que em 14,92% dos anos estudados choveu mais de $85\,cm^3/m^2$.

    (a) Qual o desvio padrão da variável aleatória que mede a pluviosidade anual nessa região?
    (b) Calcule a probabilidade da pluviosidade anual se situar entre 65 e $85\,cm^3/m^2$?

22. O tempo de funcionamento (em horas) de um certo equipamento é uma v.a. X com distribuição exponencial de parâmetro $1/2$, ou seja, é uma v.a. com a seguinte função densidade de probabilidade

$$f(x) = \frac{1}{2}e^{-\frac{x}{2}},\ x \geq 0.$$

   (a) Calcule a probabilidade de que a 1ª avaria ocorra pelo menos 1 hora depois do inicio do funcionamento do equipamento.

   (b) Calcule a probabilidade de que a 1ª avaria não ocorra depois das 4 horas de funcionamento do equipamento.

   (c) Prove que a probabilidade de que o equipamento dure mais de 10h sabendo que já está a funcionar há 3 horas é igual à probabilidade de que o equipamento dure pelo menos 7 horas.

23. Considere a variável aleatória X com a seguinte função de distribuição:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-0{,}25x}, & x > 0 \end{cases}$$

   (a) Defina a função densidade da variável aleatória X.

   (b) Calcule o valor médio e a variância da variável aleatória X.

   (c) Calcule a probabilidade de em três observações independentes da variável aleatória X se obter, em todas elas, valores superiores a 1,5.

1.5 Intervalos de Confiança e Testes de Hipóteses

1. O departamento de recursos humanos da empresa que pretende recrutar funcionários para uma nova área de negócio, está a realizar um estudo para averiguar se o tipo de teste feito é adequado para o perfil de candidatos que procura. Para tal, seleccionou aleatoriamente a classificação no teste de 15 candidatos do sexo masculino e 15 candidatos do sexo feminino. De entre as várias análises estatísticas realizadas apresentam-se os seguintes resultados:

Group Statistics

	Sexo	N	Mean	Std. Deviation	Std. Error Mean
Resultados no teste	Masculino	15	94,40		4,775
	Feminino	15	98,20	15,167	3,916

Tests of Normality

	Sexo	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Resultados no teste	Masculino	,145	15	,200	,943	15	,425
	Feminino	,152	15	,200	,949	15	,505

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	Df	Mean Difference	Std. Error Difference	90% Confidence Interval of the Difference	
							Lower	Upper
Resultados no teste	Equal Variances assumed	,761	,390	28	-3,800	6,175		6,705
	Equal variances not assumed			26,97	-3,800	6,175		6,719

(a) Verifique, ao nível de significância de 1%, se existe evidência suficiente nos resultados para afirmar que, em média, os candidatos do sexo masculino verificam o requisito exigido de uma classificação mínima no teste de 115.

(b) Calcule o *p-value* associado ao teste da alínea anterior.

(c) Determine o intervalo a 90% de confiança para a variância da classificação no teste dos candidatos do sexo feminino.

(d) Diga, ao nível de significância de 1% se existem diferenças significativas na classificação média do teste entre os candidatos do sexo masculino e os candidatos do sexo feminino. Justifique convenientemente a sua resposta.

2. Um fabricante da indústria cerâmica pretende determinar se duas novas ligas *premium*, uma nacional e uma importada, possuem uma resistência ao calor superior à da liga *standard* já utilizada. Para tal foram realizados testes em que, para 20 fornadas de cada tipo de liga, se registou a temperatura máxima de resistência

ao calor. Apresentam-se abaixo alguns dos resultados obtidos com a análise dos dados realizada com recurso ao *software* SPSS.

Tests of Normality

	Liga	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Temperatura de resistência (ºC)	Premium Importada	,276	20	,000	,750	20	,000
	Premium Nacional	,156	20	,200*	,926	20	,129
	Standard	,106	20	,200*	,973	20	,825

a. Lilliefors Significance Correction

One-Sample Test

	Test Value = 1535			
	t	df	Sig. (2-tailed)	Mean Difference
Temperatura de resistência (ºC) Liga Premium nacional	3,869	19	**?**	6,48261

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means			
		F	Sig.	Mean Difference	Std. Error Difference	99% Confidence Interval of the Difference	
						Lower	Upper
Temperatura de resistência (ºC) Premium nacional vs Standard	Equal variances assumed	1,697	,200	25,42954	2,80969	?	33,04816
	Equal variances not assumed			25,42954	2,80969		

(a) As amostras recolhidas provêm de populações com distribuição normal?

(b) Podemos afirmar, ao nível de significância de 10%, que a temperatura de resistência média da liga *premium* nacional é diferente de 1535?

(c) Calcule o *p-value* associado ao teste da alínea anterior.

(d) Com que confiança, podemos dizer que a variância da temperatura de resistência da liga *standard* se encontra entre]71.12, 166.04[?

(e) Será que existe evidência estatística suficiente nos dados, ao nível de significância de 1%, para dizer que a temperatura de resistência média da nova liga *premium* nacional é significativamente diferente da liga *standard*? E significativamente melhor?

3. Foram retiradas 25 peças da produção diária de uma máquina, encontrando-se, para uma certa medida, uma média de 5.2 mm. Sabe-se que as medições têm distribuição normal. Construa intervalos de confiança para média populacional aos níveis de confiança de 90%, 95% e 99%.

(a) Com base num desvio-padrão populacional de 1.2 mm.

(b) Com base num desvio-padrão amostral de 1.2 mm.

(c) Justifique as diferenças obtidas.

4. Um conjunto de 40 condutores de camião, escolhidos aleatoriamente nas estradas nacionais, dispôs-se a participar numa experiência que tinha por objetivo medir os seus tempos de reação depois de almoço. A média e o desvio padrão dos tempos observados foram, respetivamente, 0.85 e 0.20 segundos. Admitindo que os tempos de reação seguem uma distribuição normal, determine:

   (a) o intervalo de confiança a 95% para o valor esperado do tempo de reação após o almoço.

   (b) o intervalo de confiança a 99% para a variância do tempo de reação após o almoço.

5. Pretende-se analisar os salários, por sexo, do pessoal em início de carreira duma determinada área. As tabelas seguintes foram obtidas com recurso ao SPSS e fornecem alguns elementos para descrever a amostra recolhida:

	N	Mean	Std. Deviation
salário	1100	2606.4205	696.79819

	Gender	N	Mean	Std. Deviation
salário	Feminino	A	2476.9510	689.57645
	Masculino	631	2702.6506	687.00971

Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	Df	Sig.(2-tailed)	Mean Difference	Std. Error Difference	90% Confidence Interval of the Difference	
								Lower	Upper
Resultados no teste	Equal Variances assumed	.034	.854	**B**	.000	**C**	41.95173	-308.014	-143.345
	Equal variances not assumed			1006.36	.000	-225.6996	41.97410	-308.068	-143.331

Com base nestes resultados:

(a) Calcule o intervalo a 90% de confiança para a média global dos salários do pessoal em início de carreira.

(b) Indique os valores de A, B e C.

(c) Existem ou não diferenças significativas entre os salários médios iniciais nos 2 sexos?

6. Registou-se o comprimento, em metros, dos saltos de 10 atletas portugueses do sexo masculino em provas de triplo salto em pista coberta:

14.97	15.32	15.02	15.33	15.10	14.75	15.66	15.05	15.05	15.25

Tenha em conta que $\sum_{i=1}^{10} x_i = 151.5$ e que $\sum_{i=1}^{10} x_i^2 = 2295.8$ e que o comprimento dos saltos dos atletas portugueses do sexo masculino (em provas de triplo salto) é uma variável aleatória com distribuição normal com valor médio μ e desvio padrão σ metros. Determine:

   (a) a confiança do intervalo [14.95, 15.35] para o comprimento médio dos saltos.

   (b) o intervalo de confiança a 99% para a variabilidade do comprimento dos saltos dos atletas portugueses do sexo masculino em provas de triplo salto.

   (c) Teste, para um nível de significância de 10%, se é razoável admitir que o comprimento médio dos saltos é superior a 15.25 metros.

7. O Serviço Nacional de Saúde (SNS) afirma que a proporção de asmáticos numa certa população masculina é inferior a 10%. Um médico, pensando que este valor é muito baixo, deseja testar esta hipótese e escolhe ao acaso 200 homens do ficheiro dos seus doentes tendo verificado que 31 doentes sofrem de asma.

   (a) Teste se o médico deve avisar o SNS de que a sua estimativa não está correcta?

   (b) Calcule o *p-value* associado a este teste.

   (c) Qual a potência de teste para p_1=0,18?

   (d) Calcule o intervalo de confiança para a verdadeira proporção de asmáticos.

8. Foram efetuados estudos em Lisboa com o objectivo de determinar a concentração de monóxido de carbono (CO) perto de vias rápidas. Para isso recolheu-se uma amostra de 20 pequenos volumes de ar, para os quais se determinaram a respectiva concentração de CO (em partes por milhão, ppm), usando um espectrómetro. Tais medições resultaram numa média de valores de 100,5ppm com variância de 27,5ppm2, tendo-se verificado que em 5 das medições a concentração observada ultrapassava os 110ppm. Obteve-se, ainda, para uma confiança de 95% que a concentração de CO segue uma distribuição normal.

   (a) Teste a hipótese de a variância da concentração esperada de CO ser inferior a 25ppm2, indicando os pressupostos que tenha de fazer. (Use um nível de significância de 1%.)

   (b) Deduza um intervalo de confiança a 90% para a concentração média de CO na população.

(c) Qual a dimensão da amostra que deveria considerar, para que o erro de estimativa não ultrapasse os 0,3ppm. (Considere nível de significância de 5%)

9. Num trabalho realizado há já algum tempo concluiu-se que 62% dos passageiros que entram na estação A do metro tem como destino o centro da cidade. Esse valor tem vindo a ser utilizado em todos os estudos de transportes realizados desde então. Tendo surgido dúvidas sobre a actualidade daquele valor pois crê-se que tem vindo a diminuir, acompanhando o declínio do centro, realizou-se um inquérito naquela estação. Dos 240 passageiros inquiridos, 126 indicaram o centro como destino. Com base nestes resultados construa um intervalo de confiança a 90% para a percentagem de passageiros entrados em A e que saem no centro. O que pode concluir?

10. Certa linha de fabrico está programada de modo a produzir uma percentagem de artigos defeituosos não superior a 3%. De modo a verificar se o processo está sob controlo, são recolhidas periodicamente amostras de 50 artigos. A determinada altura verificou-se que uma dessas amostras continha dois artigos defeituosos e o encarregado declarou o processo fora de controlo, dando ordens para a interrupção do mesmo.

    (a) Considerando um nível de significância de 5% diga se o encarregado agiu bem.
    (b) Calcule o *p-value* associado a este teste.
    (c) Calcule a potência do teste considerado na alínea a) para p_1=0,035 e para p_1=0,1. Que conclusões pode retirar sobre dois valores?

11. A poluição atmosférica é medida em dois locais distintos, um no centro de uma pequena cidade (Y) e outro numa zona rural, 15Km mais a sul, (X). As medições foram efectuadas 3 dias por semana, durante 4 meses e com idênticos instrumentos de medição. Em alguns dos dias não foram efectuadas medições devido a falhas de equipamento. Os resultados registados foram os seguintes:

$$\sum_{i=1}^{31} x_i = 158.218,\ \sum_{i=1}^{31} x_i^2 = 898.501.5,\ s_Y^2 = 3.8026 \text{ e } \sum_{i=1}^{21} y_i^2 = 735.734$$

    Considere que X e Y seguem uma distribuição normal.

    (a) Construa o intervalo de confiança a 97% para a comparação média da poluição atmosférica nos locais analisados. O que pode concluir?
    (b) Construa o intervalo de confiança a 98% para a variância da poluição atmosférica na zona rural.

12. Dois laboratórios (A e B) avaliam a quantidade de cloro de amostras de água recolhidas à mesma hora de cada dia. Considerando que as amostras recolhidas são independentes, e com base nos *outputs* abaixo,

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
LabA	,229	7	,200	,867	7	,176
LabB	,250	7	,200*	,856	7	,147

a. Lilliefors Significance Correction
*. This is a lower bound of the true significance

Group Statistics

	Grupo	N	Mean	Std. Deviation	Std. Error Mean
LabA	1	7	1,4671	0,45121	0,17054
	2	7	1,4929	0,44575	0,16848

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									90% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
LabA	Equal Variances assumed	0,004	0,952	0,107	12	0,916	-0,02571	0,23973	-0,54803	?
	Equal variances not assumed			0,107	11,998	0,916	-0,02571	0,23973	?	0,49661

(a) Construa o intervalo de confiança para a diferença entre as duas médias.

(b) Existe evidência suficiente nos resultados para afirmar, com uma confiança de 99%, que existem diferenças significativas entre os dois laboratórios?

13. Um investigador pretende estudar a capacidade de concentração dos alunos do ensino universitário antes e depois do almoço. Com esse objectivo, seleccionou aleatoriamente uma amostra de 10 alunos e mediu-se, utilizando um determinado coeficiente, a capacidade de concentração de cada aluno antes e depois do almoço:

Antes	52	45	55	50	55	47	50	46	56	53
Depois	50	49	51	48	53	43	49	47	55	50

Com base nos *outputs* apresentados responda às seguintes questões:

(a) Construa um intervalo a 98% de confiança para a diferença entre a capacidade média de concentração antes e depois do almoço.

(b) Verifique se existe evidência suficiente nos resultados para afirmar que existem diferenças significativas entre a capacidade de concentração antes e depois do almoço.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Antes	0,150	10	0,200	0,928	10	0,429
Depois	0,139	10	0,200	0,968	10	0,870

a. Lilliefors Significance Correction

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Antes-Depois	,234	10	,128	,879	10	,127

a. Lilliefors Significance Correction

Paired Samples Test

		Paired Differences			t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean			
Pair 1	Antes - Depois	1,400	2,413	0,763	1,835	9	0,100

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Antes	10	50,90	3,957	1,251

One-Sample Test

	Test Value = 10				
	t	df	Mean Difference	95% Confidence Interval of the Difference	
				Lower	Upper
Antes	?	9	40,900	?	?

(c) Teste se podemos admitir uma variabilidade da capacidade de concentração antes do almoço igual 10.

(d) Calcule o *p-value* associado ao teste de hipóteses da alínea anterior.

(e) Considere o teste de hipóteses $H_0 : \mu_X \leq k$ *vs* $H_1 : \mu_X > k$, onde μ_X representa o valor médio da capacidade de concentração antes do almoço. Sabendo que se rejeita a hipótese H_0, ao nível de significância de 5%, para um valor de $\bar{x} > 56$, determine o valor da constante k.

14. A uma eleição concorrem três candidatos A, B e C. Qualquer deles admite desistir se esperar obter menos de 10% dos votos. Uma sondagem a 150 potenciais eleitores revelou as seguintes intenções de voto:

	A	B	C
Nº de Intenções de Voto	12	74	64

(a) Determine um intervalo de confiança a 95% para a proporção de votos no candidato B.

(b) Teste, ao nível de significância de 5% se o candidato A deve desistir.

(c) Calcule o *p-value* associado ao teste da alínea anterior.

(d) Qual é o número máximo de intenções de voto que um dos candidatos poderia obter para que fosse levado a desistir? (Considere um nível de confiança de 99%)

15. Suponha que o teor de nicotina de duas marcas de cigarros foi analisado, não se tendo rejeitado a hipótese de serem provenientes de populações com distribuição normal, nem de possuírem homogeneidade de variâncias. Admita que, numa amostra de 40 cigarros da marca A, o teor médio é 2.65mg e o desvio padrão 0.23mg e, numa amostra de 35 cigarros da marca B, o teor médio é 2.30mg e o desvio padrão 0.22mg.

   (a) Teste, para uma confiança de 99%, se o teor médio de nicotina dos cigarros da marca A é superior ao teor médio de nicotina dos cigarros da marca B.

   (b) Calcule o *p-value* associado ao teste da alínea anterior.

   (c) Se o verdadeiro valor do teor médio de nicotina dos cigarros da marca A for de 2.1mg, determine a potência associada ao teste $H_0 : \mu_A = 2.5$ *vs* $H_1 : \mu_A \neq 2.5$.

16. Para comparar a resistência ao esforço físico de duas populações, A e B, submeteram-se dois grupos de indivíduos, um de cada população, a um exercício na passadeira rolante, medindo-se o tempo (em minutos) até ao consumo máximo de oxigénio.

   Considere que X e Y representam o tempo (em minutos) até ao consumo máximo de oxigénio para cada indivíduo da população A e da população B, respectivamente. Admite-se que X e Y têm distribuição normal. Caso necessite utilize os seguintes resultados:

   $$\sum_{i=1}^{10} x_i = 109.9,\ \sum_{i=1}^{10} (x_i - \bar{x})^2 = 49.41,\ \sum_{i=1}^{11} y_i = 174.7 \text{ e } \sum_{i=1}^{11} (y_i - \bar{y})^2 = 111.74$$

   (a) Indique estimativas pontuais para o valor médio e desvio padrão de Y.

   (b) Determine um intervalo a 90% de confiança para o tempo médio até ao consumo máximo de oxigénio dos indivíduos da população B.

   (c) Obteve-se o seguinte intervalo de confiança para a razão de variâncias das duas populações:]0.0995; 2.584[. Determine o nível de confiança desse intervalo.

17. Com o objectivo de estudar algumas características dos jogadores de futebol que participam no Campeonato Europeu de Futebol de 2008, que está a decorrer na Áustria e na Suíça, foi recolhida aleatoriamente uma amostra de 72 jogadores, para os quais foram registados os valores das seguintes variáveis: país de origem, posição em campo (guarda-redes, defesa, médio, avançado), altura (em cm), peso (em kg), idade (em anos) e número de internacionalizações. Na amostra recolhida, 12 jogadores são da República Checa e 12 são da Grécia.
   Com base nos resultados apresentados:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Altura dos jogadores Checos	0,141	12	0,200*	0,946	12	0,574
Altura dos jogadores Gregos	0,167	12	0,200*	0,930	12	0,376

a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Altura dos jogadores	Equal variances assumed	1,047	0,317

Group Statistics

	Grupo	N	Mean	Std. Deviation	Std. Error Mean
Altura dos jogadores	1	12	178,00	5,527	1,595
	2	12	180,00	4,390	1,267

(a) Determine um intervalo a 98% de confiança para a diferença entre as alturas médias dos jogadores checos e gregos. Justifique, convenientemente, a escolha do intervalo de confiança utilizado.

(b) Indique a margem de erro associada ao intervalo anterior.

(c) Para α=5%, diga se há diferença entre as alturas médias dos jogadores checos e gregos. Justifique, convenientemente, a sua resposta.

(d) Averigue se podemos admitir que a altura média dos jogadores checos é superior a 175 cm (use um nível de significância de 1%).

(e) Calcule o *p-value* associado ao teste anterior. Qual a conclusão a retirar a partir deste valor?

(f) Qual deverá ser a altura média (amostral) dos jogadores checos de modo a poder considerar-se que a altura média (populacional) desses jogadores é superior a 175 cm, para um α=1%?

(g) Será que existe evidência estatística suficiente nos dados, ao nível de significância de 1%, para dizer que a variância das alturas dos jogadores checos é superior à variância das alturas dos jogadores gregos?

18. De 72 jogadores inquiridos, 36 jogam em equipas estrangeiras, 34 em equipas do país de origem e 2 não têm equipa. Um conhecido jornal desportivo tem na sua primeira página a seguinte notícia: “mais de 40% dos jogadores do Campeonato

Europeu de Futebol de 2008 jogam em equipas estrangeiras". Considerando um nível de significância de 5%, diga se concorda ou não com esta notícia.

19. Foi levado a cabo um estudo para averiguar se a ausência às aulas durante o semestre de Inverno é maior num centro urbano do norte ou do sul. Para tal, forma seleccionados aleatoriamente dois grupos de alunos: um grupo na cidade de Braga (X) e o outro na cidade de Faro (Y). De 300 estudantes de Braga, 64 faltaram pelo menos um dia e de 400 de Faro, 51 faltaram um ou mais dias.

    (a) Determine o intervalo de confiança a 99% para a diferença entre as proporções de estudantes que faltaram pelo menos um dia às aulas durante o Inverno.

    (b) Teste, ao nível de significância de 1%, se é possível afirmar-se que, durante o Inverno, se falta (pelo menos um dia) mais às aulas na região norte do que na região sul.

    (c) Calcule o *p-value* associado ao teste anterior. Qual a conclusão a retirar a partir deste valor?

    (d) Obteve-se o seguinte intervalo de confiança para a proporção de faltas na região sul:]0.0913; 0.1637[. Determine o nível (ou coeficiente) de confiança desse intervalo. Qual o desvio padrão da variável aleatória que mede a pluviosidade anual nessa região?

20. Tome-se o seguinte exemplo, relativo a dois tipos de geradores (I e II), e considere-se que

$$n_X = 27,\ \bar{x} = 10.0241,\ s_X = 4.20283$$
$$n_Y = 23,\ \bar{y} = 10.0496,\ s_Y = 3.53775$$

    com as variáveis X e Y a representarem a produção de energia eléctrica em KW/h do gerador do tipo I e II, respectivamente, e que a produção de energia eléctrica de ambos segue uma distribuição normal.

    (a) Teste se existe evidência estatística suficiente (para um ?=1%) para afirmar que o valor esperado da produção de energia eléctrica é igual nos dois geradores.

    (b) Calcule o *p-value* associado ao teste anterior. Qual a conclusão a retirar a partir deste valor?

    (c) O fabricante afirma que o desvio padrão da produção de energia eléctrica através do gerador II é de 4KW/h. Comente a afirmação do fabricante.

    (d) Determine um intervalo a 98% de confiança para a razão entre a variabilidade da produção de energia eléctrica dos dois geradores.

1.6 Testes Não-Paramétricos

1. Considere a seguinte amostra: 1.26; 0.34; 0.7; 1.75; 5.57; 1.55; 0.8; 0.42; 0.51; 3.2; 0.15; 0.49; 0.95; 0.24; 1.37; 0.17; 6.98; 0.13; 0.94; 0.38
   Com base no *output* apresentado, teste se é de admitir que a amostra seja proveniente de uma população com distribuição normal.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Ex1	,416	20	,000	,324	20	,000

a. Lilliefors Significance Correction

2. Considere duas amostras, ambas de dimensão 20, correspondentes a medições realizadas em dois grupos de pacientes. A um dos grupos foi administrado um placebo e ao outro administrado um tratamento específico. Os resultados registados foram os seguintes:

Dados de Placebo

1.26	0.34	0.7	1.75	50.57	1.55	0.8	0.42	0.51	3.2
0.15	0.49	0.95	0.24	1.37	0.17	6.98	0.13	0.94	0.38

Dados do tratamento

3,59	6,72	37,2	0,69	6,94	14,81	9,39	-10,95	13,28	12,84
4,65	18,59	30,54	2,28	1,54	1,41	-2,64	11,99	22,95	-16,03

Com base no output abaixo, teste se podemos admitir que ambas as amostras sejam provenientes de populações com distribuição normal.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Placebo	0,416	20	0,000	0,324	20	0,000
Tratamento	0,119	20	0,200*	0,971	20	0,773

a. Lilliefors Significance Correction
*. This is a lower bound of the true significance.

3. Um novo jogo de casino consiste em lançar 3 vezes um dado. O vencedor é aquele que obtiver o maior número de pintas 6. Um jogador, selecionado ao acaso, jogou 100 vezes e obteve os seguintes resultados:

Número de pintas 6	n_i
0	47
1	35
2	15
3	3

O jogador não aceita estes resultados e suspeita da honestidade do casino. Para tal, vai realizar um teste de hipóteses para avaliar a falsidade, ou não, do dado. O que conclui?

4. Foi realizado um estudo para determinar se a opinião pública era favorável à construção de uma barragem hidroeléctrica. Os resultados foram: 40% a favor da construção, 30% são indiferentes, 20% opõem-se e os restantes disseram não terem pensado no assunto. Uma amostra aleatória de 150 indivíduos da região afectada revelou que 42 eram a favor, 61 eram indiferentes e 33 eram contrários à construção.

   (a) Com base nos outputs seguintes, estarão estes dados de acordo com os resultados obtidos no referido estudo?

   (b) Determine o valor do p-value associado a este teste de hipóteses.

Opinião			
	Observed N	Expected N	Residual
1	42	60,0	-18,0
2	61	45,0	16,0
3	33	30,0	3,0
4	14	15,0	-1,0
Total	150		

Test Statistics	
	Opinião
Chi-Square	11,456[a]
df	3
Asymp. Sig.	P

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 15,0.

5. A procura diária de um certo produto foi, em 40 dias escolhidos ao acaso, a seguinte:

Número de unidades	0	1	2	3	4	5
Número de dias	6	14	10	7	2	1

Será que se pode admitir que tais observações foram extraídas de uma população com distribuição Poisson, isto é, será de admitir que a procura diária segue uma distribuição de Poisson?

6. Numa dada sala de cinema da região de Évora realizou-se um inquérito a 400 estudantes, escolhidos aleatoriamente, da Universidade de Évora, relativamente à sua preferência sobre 4 tipos de filmes: A, B C e D. Os resultados obtidos foram:

Filme			
	Observed N	Expected N	Residual
A	A	100	30
B	b	100	-10
C	c	100	-20
D	d	100	0
Total	e		

Test Statistics	
	Filme
Chi-Square	14,000[a]
df	3
Asymp. Sig.	0,003

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 100.

(a) Determine os valores de a, b, c, d e e.

(b) Ao nível de significância de 1%, poderá afirmar-se que não existe preferência por nenhum dos tipos de filmes? Efectue o teste estatístico que considere mais adequado.

7. Foi registado o número de nascimentos num hospital durante os quatro períodos do ano: Jan-Mar, Abr-Jun, Jul-Set, Out-Dez. Diz-se que durante o período de Jan-Mar nascem duas vezes mais crianças do que nos outros períodos. Verifique se os resultados obtidos na experiência contradizem a afirmação.

Trimestre			
	Observed N	Expected N	Residual
Jan-Mar	110	120,0	-10,0
Abr-Jun	57	60,0	-3,0
Jul-Set	53	60,0	-7,0
Out-Dez	80	60,0	20,0
Total	300		

Test Statistics	
	Trimestre
Chi-Square	8,467[a]
df	3
Asymp. Sig.	,037

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 60.

8. Com o objectivo de participarem numa determinada actividade social, os estudantes de uma escola foram submetidos a dois testes: um psicotécnico e um sobre regras de conduta. Obtiveram-se os seguintes resultados:

		Psicotécnico	
		Aprovado	Reprovado
Regras de conduta	Aprovado	54	73
	Reprovado	47	167

   Existe relação entre os resultados obtidos nos dois testes? Considere $\alpha = 5\%$.

9. Numa certa região do país foi feito um inquérito às preferências clubistas dos adeptos de futebol relativamente aos clubes de futebol Porto, Benfica e Sporting. Os inquiridos foram classificados por faixa etária em 35 anos ou menos e mais de 35 anos. De acordo com o seguinte output do SPSS responda às questões apresentadas.

Faixa etária * Clube Crosstabulation

			Clube			Total
			Porto	Benfica	Sporting	
Faixa etária	<=35 anos	Count	75	75	d	200
		Expected Count	a	80,0	60,0	200,0
	>35 anos	Count	75	125	100	300
		Expected Count	90,0	120,0	90,0	b
Total		Count	150	c	150	500
		Expected Count	150,0	200,0	150,0	500,0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	9,549[a]	2	,008
Likelihood Ratio	9,488	2	,009
Linear-by-Linear Association	8,663	1	,003
N of Valid Cases	500		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 60,00.

   (a) Determine os valores de a, b, c e d.

   (b) É de concluir que as preferências se podem considerar independentes da idade?

10. Foi conduzida uma experiência no âmbito da qual se procurou testar se existe alguma relação entre a qualidade da secagem de máquinas de lavar roupa de

um certo tipo e a velocidade de rotação a que se eleva o tambor da roupa na fase de centrifugação. Os resultados desta experiência, efectuada com base no comportamento de 90 máquinas, estão apresentados na seguinte tabela:

		Qualidade da secagem			
		Medíocre	Suficiente	Boa	Muito Boa
Velocidade de rotação (rpm)	600	12	8	7	3
	900	9	10	7	4
	1200	2	9	8	11

Teste, ao nível de significância de 5%, se as duas variáveis são ou não independentes.

11. Os dados seguintes são o resultado de um inquérito de opinião efetuado a residentes no litoral e no interior acerca do desenvolvimento da sua região. Para averiguar se a opinião acerca do desenvolvimento local é condicionada pela região de residência, procedeu-se a um teste adequado tendo-se obtido o seguinte quadro de resultados:

Região* Opinião Crosstabulation

			Opinião			Total
			Elevado	Suficiente	Insuficiente	
Região	Interior	Count	10	**B**	140	200
		Expected Count	**A**	90,0	80,0	**C**
	Litoral	Count	**D**	130	20	200
		Expected Count	30,0	90,0	**F**	200,0
Total			60	**E**	160	**G**

Chi-Square Tests

	Value	df
Pearson Chi-Square	152,222	2

(a) Diga qual o teste realizado, formule as hipóteses correspondentes e determine os valores de A a G.

(b) O que pode concluir, para 10% de significância?

1.7 Regressão Linear Simples

1. Pretende-se modelar a resistência de um determinado tipo de plástico em função do tempo que decorre a partir da conclusão do processo de moldagem até ao momento de medição de resistência (horas). Foram registadas as observações correspondentes a 12 peças construídas com este plástico, escolhidas aleatoriamente. Com base nos resultados apresentados abaixo, responda às seguintes questões:

Statistics

	Resistência	Tempo até medição da resistência
Mean	269,92	48,00
Variance	2125,356	349,091

Coefficients[a]

Model		Unstandardized Coefficients		Sig.	95,0% Confidence Interval for B	
		B	Std. Error		Lower Bound	Upper Bound
1	(Constant)	153,917	8,067		135,943	171,890
	Tempo até medição da resistência	2,417	,157	,000		

a. Dependent Variable: Resistência

(a) Indique a recta de regressão estimada e interprete os seus coeficientes.

(b) Determine e interprete os coeficientes de correlação e de determinação.

(c) Ao nível de significância de 10% pode concluir que a recta de regressão não passa pela origem?

(d) Diga se existe evidência de que o tempo desde a conclusão do processo de moldagem influencia linearmente de forma significativa a resistência.

(e) Decorridas 35 horas desde a conclusão do processo de moldagem, qual o valor da resistência previsto?

2. Uma empresa que avalia a qualidade de água das ETAR, analisa o nível de substâncias nocivas detectadas na água (mg/100ml) de modo a classificar a água nas duas categorias: água rejeitada e água em boas condições. Para tal, tenta analisar a possível relação linear entre o nível de substâncias nocivas detectadas na água (mg/100ml) e o nível de substâncias químicas utilizadas no tratamento da

mesma água (mg/100ml). Admita que existe relação linear entre as variáveis.

Descriptive Statistics

	N	Mean	Variance
Nível de substâncias químicas utilizadas no tratamento da água	100	4,20	2,800
Nível de substâncias nocivas detectadas na água	100	3,80	1,900

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	?[a]	,530	,520	,240

a. Predictors: (Constant), Nível de substâncias químicas utilizadas no tratamento da água

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	?	,188		?	,000
	Nível de substâncias químicas utilizadas no tratamento da água	?	,240	?	?	?

a. Dependent Variable: Nível de substâncias nocivas detectadas na água

(a) Qual a equação da recta de regressão estimada? Interprete os coeficientes de regressão estimados.

(b) Determine os coeficientes de correlação e de determinação e interprete-os.

(c) Ensaie, ao nível de significância de 5%, a hipótese da recta de regressão passar na origem.

(d) Ao nível de significância de 5%, teste se o nível de substâncias químicas utilizadas no tratamento da água influencia linearmente o nível de substâncias nocivas detectadas na água.

(e) Sabendo que um dos nível de substâncias nocivas detectadas na água foi de 3,21 mg/100ml e que o nível de substâncias químicas utilizadas no tratamento da água foi de 2,64 mg/100ml, calcule o resíduo de estimação.

(f) Por cada mg/100ml a mais no nível de substâncias químicas utilizadas no tratamento da água, quanto aumenta (ou diminui) o nível de substâncias nocivas detectadas na água? Page 1

3. O Presidente da Junta de Freguesia da pacata vila alentejana (e dono da única gasolineira) pretende averiguar se a quantidade de gasolina vendida por dia depende, ou não, do seu preço/l. Assim, com base nos últimos 25 meses (em que em cada mês se seleccionou um dia ao acaso para a recolha das amostras) e admitindo que existe relação linear entre as variáveis, colocaram-se as seguintes questões:

(a) Escreva a equação da recta de regressão estimada e interprete os coeficientes de regressão estimados.

(b) Qual a percentagem de variação da quantidade de gasolina vendida por dia em função do seu preço?

(c) Com base nos dados que conclusão transmitiria ao Presidente da Junta relativamente à sua questão? Ou seja, poderá transmitir-lhe que a associação entre o preço da gasolina e a quantidade vendida estão relacionados de modo significativo? Justifique adequadamente a sua resposta.

(d) Estime a quantidade de gasolina vendida por dia quando o seu preço por litro atingir os 2 euros.

Descriptive Statistics

	N	Mean	Variance
Preço da gasolina (por litro)	25	1,65	1,500
Quantidade de gasolina vendida (por dia)	25	37,00	3,500

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,96[a]	?	,520	,240

a. Predictors: (Constant), Preço da gasolina, por litro

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	?	,232		?	,000
	Preço da Gasolina (por litro)	?	,354	?	?	?

a. Dependent Variable: Quantidade de gasolina vendida por dia

4. Pretende-se obter um modelo que possa explicar a área foliar (em cm^2) através do comprimento da nervura principal (em cm) em folhas de videiras da casta Fernão Pires. Para tal recolheu-se uma amostra aleatória de 12 folhas, tendo-se obtido para cada uma delas os seguintes valores das 2 variáveis acima referidas:

Descriptive Statistics

	Mean	Std. Deviation	N
Área Foliar (cm^2)	175,42	30,776	12
Nervura principal (cm)	11,617	1,2298	12

Model Summary

Model	R Square	Adjusted R Square	Std. Error of the Estimate
1	0,940	0,934	7,887

Admita que existe relação linear entre as variáveis. Considere ainda que $Var[\hat{\beta}_0] = 509,77$ e que $Var[\hat{\beta}_1] = 3,74$.

(a) Qual a equação da recta de regressão estimada? Interprete os coeficientes de regressão estimados.

Page 1

(b) Qual o valor do coeficiente de determinação e interprete-o.

(c) Por cada 2 cm a mais no comprimento da nervura principal, quanto aumenta a área foliar da folha de videira?

(d) Ao nível de significância de 5%, teste se o comprimento da nervura principal influencia linearmente a área foliar.

(e) Obtenha o intervalo de confiança a 99% para o parâmetro que representa a ordenada na origem. (**Nota:** Caso não tenha determinado a equação da recta, considere valores hipotéticos para β_0 e β_1.)

(f) Sabendo que numa folha a nervura principal media 9,1 cm de comprimento e a área foliar era 126 cm^2, calcule o resíduo de estimação.

5. Para alguns países da Europa, foram registados alguns indicadores económicos, nomeadamente o PIBA (produto interno bruto, originado pela agricultura) e o PURB (percentagem de população urbana). Os valores apresentam-se na tabela seguinte:

Países	PIBA	PURB
Alemanha	2	85
Áustria	4	54
Bélgica	2	72
Dinamarca	4	84
Espanha	8	74
Finlândia	0	62
França	4	78
Grécia	16	62
Holanda	4	76
Itália	6	69
Noruega	5	53
Portugal	13	31
Reino Unido	2	91
Suécia	3	87

Considere que $\sum_{i=1}^{14} y_i = 73$, $\sum_{i=1}^{14} x_i = 978$, $\sum_{i=1}^{14} y_i^2 = 635$, $\sum_{i=1}^{14} x_i^2 = 71786$ e $\sum_{i=1}^{14} x_i y_i = 4591$.

(a) Estime da recta de regressão entre as variáveis PIBA e PURB.

(b) Ensaie a hipótese do declive da recta de regressão ser igual a -0,2 (ao nível de significância de 5%).

(c) Qual a percentagem de variabilidade da variável Y que é explicada pela variável X?

6. No gráfico de dispersão abaixo encontram-se registados os tempos (em segundos) do vencedor na final da corrida de 400 metros masculinos, em cada Olimpíada entre 1896 e 2004 (i.e., dos 25 anos em que as Olimpíada da era moderna se realizaram):

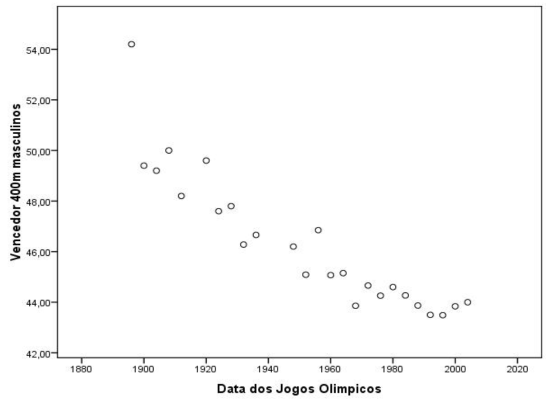


Figura 2: Gráfico de Dispersão (Scatter plot).

Considere os seguintes outputs que foram obtidos recorrendo ao software estatístico SPSS.

Vencedor 400m masculino

N	Valid	25
	Missing	0
Mean		46,306
Median		45,15
Std. Deviation		2,66223
Percentiles	25	44,13
	50	45,15
	75	48

Data dos Jogos Olímpicos

N	Valid	25
	Missing	0
Mean		1952
Median		1956
Std. Deviation		34,059
Percentiles	25	1922
	50	1956
	75	1982

Considere ainda que $\sum_{i=1}^{25} x_i y_i = 2257753$.

(a) Diga, destas duas variáveis, qual a variável dependente e qual a variável explicativa?

(b) A relação entre as duas variáveis será positiva ou negativa? Porquê?

(c) Determine o coeficiente de determinação e interprete-o.

(d) Construa o modelo de regressão linear simples e interprete as estimativas dos parâmetros do modelo.

(e) Obtenha uma previsão para o tempo do vencedor, na referida prova, para a Olimpíada de 2008.

7. Considerando a idade (em anos) e a capacidade pulmonar (em L) de 9 crianças, observou-se a seguinte tabela:

Idade (X)	4	5	6	7	8	9	10	11	12
Cap. Pulmonar (Y)	0,7	0,9	1,2	1,3	1,3	1,5	1,7	1,9	2,1

Com o auxílio do software estatístico SPSS obtiveram-se os seguintes resultados:

	Unstandardized	Coefficients	
Model	B	Std. Error	t
1 (constant)	0,093	0,081	1,147
Idade, em anos	0,163	0,010	16,878

(a) Calcule os coeficientes de correlação e determinação e interprete-os.

(b) Apresente a recta de regressão dos mínimos quadrados. Interprete o declive estimado.

(c) Determine o intervalo de 95% de confiança para o declive da recta.

(d) Ensaie a hipótese da recta de regressão passar na origem, ao nível de significância de 1%.

8. Pretende-se, se possível, modelar através de uma recta 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:

x_i	2	3	4	5	10	15	20	25
y_i	100	150	250	320	650	810	1040	1480

Dos quais resulta $\sum_{i=1}^{8} x_i y_i = 79700$.

Nº dias sem despejar o Ecoponto

N Valid	8
Missing	0
Mean	10,5
Median	7,5
variance	74,571

Quantidade de vidro num Ecoponto (Kg)

N Valid	8
Missing	0
Mean	600
Median	485
Variance	238285,7

(a) Escreva a recta de regressão estimada através do método dos mínimos quadrados.

(b) Diga, justificando convenientemente, se o modelo encontrado se ajusta bem ou não aos dados reais.

(c) Teste a hipótese de o declive da recta de regressão ser nulo, ao nível de significância de 10%.

(d) Qual o valor da quantidade de vidro produzida no Ecoponto que prevê ocorrer em 10 dias sem o despejar. Seria possível calcular o mesmo para um período de 35 dias?

9. Mediu-se o número de pulsações por minuto antes e depois de uma determinada prova de esforço, num grupo de 10 fumadores com mais de 20 anos consecutivos de consumo de tabaco, tendo-se obtido os seguintes resultados:

Pulsação antes (x_i)	49	50	52	53	55	55	61	62	64	70
Pulsação depois (y_i)	98	95	97	91	96	100	110	109	115	120

Com base nos outputs da folha seguinte:

(a) Calcule os coeficientes de correlação e determinação e interprete-os.

(b) Apresente a recta de regressão. Interprete o declive estimado.

(c) Determine o intervalo de 95% de confiança para o declive da recta.

(d) Ensaie a hipótese da recta de regressão passar na origem, ao nível de significância de 1%.

(e) O que poderá dizer sobre a significância da regressão?

Descriptive Statistics

	Mean	Std. Deviation	N
Y Pulsação Depois	103,100	9,6891	10
X Pulsação Antes	57,10	6,839	10

Correlations

		Y Pulsação Depois	X Pulsação Antes
Pearson Correlation	Y Pulsação Depois	1,000	,934
	X Pulsação Antes	,934	1,000
Sig. (1-tailed)	Y Pulsação Depois	.	,000
	X Pulsação Antes	,000	.
N	Y Pulsação Depois	10	10
	X Pulsação Antes	10	10

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,934[a]	,872	,856	3,6752

a. Predictors: (Constant), X Pulsação Antes

b. Dependent Variable: Y Pulsação Depois

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	736,844	1	736,844	54,553	,000[a]
	Residual	108,056	8	13,507		
	Total	844,900	9			

a. Predictors: (Constant), X Pulsação Antes

b. Dependent Variable: Y Pulsação Depois

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	27,550	10,295		2,676	,028
	X Pulsação Antes	1,323	,179	,934	7,386	,000

a. Dependent Variable: Y Pulsação Depois

Coefficients[a]

Model		95,0% Confidence Interval for B	
		Lower Bound	Upper Bound
1	(Constant)	3,811	51,290
	X Pulsação Antes	,910	1,736

a. Dependent Variable: Y Pulsação Depois

2 Soluções

2.1 *Estatística Descritiva*

1. O número de chamadas telefónicas recebidas, por minuto, numa determinada central telefónica ...

   (a) **A**=9, **B**=50, **C**=1 e **D**=4.

X_i'	n_i	f_i	N_i	F_i
0	6	0,12	6	0,12
1	20	0,40	26	0,52
2	9	0,18	35	0,70
3	10	0,2	45	0,90
4	5	0,1	50	1

   (b) $\overline{x} = 1,76$: Em média, foram recebidas por minuto 1,76 chamadas;
   $\widehat{x} = 1$: Observou-se com maior frequência a recepção de 1 chamada por minuto.
   $s = 1,205$: O número de chamadas recebidas por minuto, apresentou um desvio típico em torno da média de cerca de 1,2 chamadas.

   (c) Complete as seguintes frases:

      i. Em 18% dos minutos foram recebidas 2 chamadas;
      ii. Em 30% dos minutos foram recebidas mais de 2 chamadas;
      iii. Em 26 minutos foi recebida no máximo 1 chamada.

   (d) A distribuição dos dados é assimétrica positiva e ligeiramente platicúrtica.

2. O responsável técnico de uma empresa pretende averiguar qual a origem das interrupções diárias do sistema. (...)

   (a) Em média, verificou-se 1,74 interrupções diárias do sistema. O que se verificou com mais frequência foi a ausência de interrupções diárias do sistema. Em 50% dos dias, observou-se no máximo 1 interrupção diária dos sistema. O desvio típico em torno da número médio de interrupções diárias do sistema é de cerca de 1,76 interrupções.

   (b) **C**.

   (c) Verdadeira, uma vez que $p_{85} = 3$

   (d) Assimétrica positiva e leptocúrtica.

3. Um fabricante da indústria cerâmica pretende determinar se duas novas ligas *premium*, uma nacional e uma importada, ...

   (a) Interprete os valores da média, mediana e desvio-padrão observados para a temperatura de resistência da liga *premium* nacional.

   (b) CV_1=0,591%; CV_2=0,486% e CV_3= 0,667%. Para as três ligas, verificamos que a temperatura de resistência média é bastante representativa (CV $<$ 50%).

   (c) As representações gráficas sugerem: para a liga premium importada uma distribuição assimétrica positiva e leptocúrtica; para a liga premium nacional uma distribuição assimétrica positiva e mesocúrtica (ou ligeiramente platicúrtica); para a liga standard uma distribuição simétrica e mesocúrtica.

4. Na seguimento do exercício anterior, são disponibilizados os seguintes dados da temperatura máxima de resistência ao calor...

   (a) $\overline{x} = 1541,68$; $\widehat{x} = 1539,30$; $\widetilde{x} = 1540,53$; $s = 6,6$;

   (b) $q_1 = 1536,705$: Em 25% dos testes, verificou-se uma temperatura de resistência ao calor inferior ou igual a 1536,705;
   $p_{80} = 1547,21$: Em 80% dos testes a temperatura de resistência ao calor foi no máximo de 1547,21.

   (c) A distribuição dos dados é assimétrica positiva e ligeiramente platicúrtica.

5. Um estudo sobre os atrasos nos 100 voos europeus durante um Verão...

   (a)

Classes	X_i'	n_i	f_i	N_i	F_i
$[0;10[$	5	35	0,35	35	0,35
$[10;20[$	15	25	0,25	60	0,6
$[20;30[$	25	20	0,20	80	0,8
$[30;40[$	35	15	0,15	95	0,95
$[40;50]$	45	5	0,05	100	1

   (b) $\overline{x} = 18$ minutos e $s^2 \simeq 152,5253$

   (c) Pela observação do histograma podemos dizer que os dados apresentam uma assimetria à esquerda (ou positiva). Pelo que a proporção de voos com pequenos atrasos é superior à proporção de voos com grandes atrasos.

6. Foi feito um estudo de modo a avaliar quantas vezes durante uma tarde de estudo (5 horas) os alunos enviam e/ou recebem mensagens sms...

(a)

X_i'	n_i	f_i	N_i	F_i
0	1	0,01	1	0,01
1	20	0,2	21	0,21
2	25	0,25	46	0,46
3	3	0,03	49	0,49
4	28	0,28	77	0,77
5	12	0,12	89	0,89
6	5	0,05	94	0,94
7	4	0,04	98	0,98
8	2	0,02	100	1

(b)

(c) i. $\overline{x} = 3.25$;

ii. $m_o = m_e = 4$;

iii. $s = 1.86$;

iv. $q_1 = 2$;

v. $p_{75} = 4$;

vi. $d_8 \equiv p_{80} = 5$.

(d) $g_B = -1$, a distribuição dos dados é assimétrica negativa.

(e) i. 0.51;

ii. 0.68.

7. Devido não só ao aumento dos preços dos preços combustíveis...

(a)

Tempo (em horas)	X_i'	F_i	N_i	n_i	f_i
$[3, 10[$	6,5	0,015	3	3	0,015
$[10, 12[$	11	0,03	6	3	0,015
$[12, 15[$	13,5	0,07	14	8	0,04
$[15, 20[$	17,5	0,245	49	35	0,175
$[20, 30[$	25	0,695	139	90	0,45
$[30, 60[$	45	0,945	189	50	0,25
$[60, 120]$	90	1	200	11	0,055

(b)

(c) $\overline{x} = 31.315$ e $s^2 = 317.29$ ($s = 17.813$).

(d) Classe modal e classe mediana: $[20, 30[$.

8. O encarregado do controlo de qualidade de uma fábrica de relógios digitais...

   (a) $\overline{x} = 1.405$, $m_e = m_o = 1$.

   (b)

X_i'	n_i	f_i	N_i	F_i
0	53	0,265	53	0,265
1	68	0,34	121	0,605
2	44	0,22	165	0,825
3	17	0,085	182	0,91
4	16	0,08	198	0,99
5	2	0,01	200	1

   (c) $q_1 = q_{0.25} = 0$, 25% das amostras de 5 relógios não tem nenhum relógio defeituoso.
   $q_2 = q_{0.5} = 1$, 50% das amostras de 5 relógios tem no máximo um relógio defeituoso.
   $q_3 = q_{0.75} = 2$, 75% das amostras de 5 relógios tem no máximo 2 relógios defeituosos.

   (d) $p10 = q_{0.1} = 0$, 10% das amostras de 5 relógios não tem nenhum relógio defeituoso. $p75 = q_{0.75} = 2$

   (e) $s = 1.244$ e $cv = 88.5\%$. $cv_B = 46.8\%$, pelo que os resultados da fábrica B são menos dispersos, logo mais homogéneos.

9. Numa empresa recolheu-se uma amostra aleatória relativa à produção de energia eléctrica...

   (a)

Classes	X_i'	n_i	f_i	N_i	F_i
$[2.74, 5.634[$	4.187	6	0.222	6	0.222
$[5.634, 8.528[$	7.081	1	0.037	7	0.259
$[8.528, 11.422[$	9.975	11	0.407	18	0.666
$[11.422, 14.316[$	12.869	5	0.185	23	0.851
$[14.316, 17.21]$	15.763	4	0.149	27	1
		27	1		

   (b) $\overline{x} = 9.975$ (dados agrupados), $\overline{x} = 10.027$ (dados não agrupados). Em geral, nunca são iguais. Contudo, quanto maior for o arredondamento na amplitude das classes, maior será esta diferença.

   (c)

(d) i. Gerador I: $\overline{x} = 10.027$, $m_e = 9.72$ e $s = 4.207$. Gerador II: $m_e = 10.05$; $q_3 = 11.08$ e $a_{IQ} = 3.01$.

ii. $q_{0.85} = 14.3$.

10. De modo a estudar a performance dos atletas de alta competição do Eborense Futebol Clube...

(a)

X_i'	n_i	f_i	N_i	F_i
14,6	24	0,24	24	0,24
17,8	29	0,29	53	0,53
21	18	0,18	71	0,71
24,2	18	0,18	89	0,89
27,4	3	0,03	92	0,92
30,6	4	0,04	96	0,96
33,8	4	0,04	100	1

(b) Moda e mediana para a localização e amplitude interquartílica para dispersão.

(c) Distribuição é assimétrica positiva.

11. Tal como no caso do exercício anterior, pretendia estudar-se a performance dos atletas agora não profissionais, do Eborense Futebol Clube.

(a) $q_B = -0,362$. Donde se conclui que a distribuição dos dados é assimétrica negativa.

(b) $q_{0,7} = 71,20$ metros.

12. De modo a estudar a rapidez com que os alunos resolviam problemas de estatística...

(a) $n = 25$. Número de classes: $k = [\frac{ln25}{ln2}] + 1 = 5$.

$\Delta = max - min = 63 - 17 = 46$. Amplitude das classes: $\frac{46}{5} = 9,2$.

(b)

(c)

$$\overline{x} = \frac{4 \times 21,6 + \cdots + 5 \times 58,4}{25} = 41,104 \quad \text{minutos}$$

$$s^2 = \frac{4 \times 21,6^2 + \cdots + 5 \times 58,4^2}{24} - \frac{25}{24}(41,104)^2 = 157,4304 \quad \text{minutos}^2$$

tal que $s = 12,547$ minutos.

Tabela 1: Tabela de frequências

Classes	Ponto Médio	n_i	f_i	N_i	F_i
$[17; 26,2[$	$21,6$	4	$0,16$	4	$0,16$
$[26,2; 35,4[$	$30,8$	4	$0,16$	8	$0,32$
$[35,4; 44,6[$	40	7	$0,28$	15	$0,6$
$[44,6; 53,8[$	$49,2$	5	$0,2$	20	$0,8$
$[53,8; 63]$	$58,4$	5	$0,2$	25	1
		25	1		

(d) $g_1 = -0.1916$, pelo que se pode dizer que a distribuição dos dados é assimétrica negativa.

13. Foi feito um inquérito a um grupo de compradores de 40 carros novos para determinar quantas reparações...

X_i'	n_i (a)	f_i (b)	N_i (d)	F_i (d)
0	3	0,075	3	0,075
1	12	0,3	15	0,375
2	10	0,25	25	0,625
3	8	0,20	33	0,825
4	4	0,10	37	0,925
5	2	0,05	39	0,975
6	0	0,00	39	0,975
7	1	0,025	40	1
	40	1		

14. Num estudo para analisar a capacidade de germinação de certo tipo de cereal...

(a) $\overline{x} = 2,866$, $m_e = 3$ e $m_0 = 3$.

(b)

(c) $\simeq 0,3098$.

(d) $q_1 = 2$, isto é, em 25% vasos germinaram 2 ou menos sementes. $q_2 = 3$, isto é, em 50% vasos germinaram 3 ou menos sementes. E $q_3 = 4$, isto é, em 75% vasos germinaram 4 ou menos sementes.

(e) $P75 \equiv Q_3$. $p_{10} = 1$, isto é, em 10% vasos germinaram 0 ou 1 sementes.

X'_i	n_i	f_i	N_i	F_i
0	16	0,0403	16	0,0403
1	32	0,0806	48	0,1209
2	89	0,2242	137	0,3451
3	137	0,3451	274	0,6902
4	98	0,2469	372	0,9370
5	25	0,0630	397	1
	397	1		

(f) $s = 1,182$ e $cv = 41,2\%$, donde se conclui que a média é pouco representativa da amostra.

15. Os dados que se seguem referem-se ao comprimento total (em cm) de uma colecção de dados de achigãs de uma barragem:

(a) $\overline{x} = 27,4475$, $m_e = 28,5$, $s = 8,9035$. $q_1 = 19,25$, $q_2 \equiv m_e$, $q_3 = 35,75$. $q_{2/3} = 33,3$ e $q_{0,3} = 19,55$.

(b)

Classes	X'_i	n_i	f_i	N_i	F_i
$[12,6;17,4[$	15	5	0,125	5	0,125
$[17,4;22,2[$	19,8	10	0,25	15	0,375
$[22,2;27[$	24,6	4	0,1	19	0,475
$[27;31,8[$	29,4	5	0,125	24	0,6
$[31,8;36,6[$	34,2	6	0,15	30	0,75
$[36,6;41,4]$	39	10	0,25	40	1
		40	1		

(c) $\overline{x} = 27,84$ e $s^2 = 77,258$. $[17,4;22,2[$ e $[36,6;41,4[$ representam as classes modais, ou seja são as classes com maior frequência absoluta.

16. O departamento de pessoal de uma certa firma fez um levantamento dos salários dos 120 funcionários...

(a)

(b) $\overline{x} = 3,65$ e $s = 2,274$ salários mínimos.

(c) Tanto a média como a variância sofrem alterações. Na nova situação, $\overline{x} = 7,3$, ou seja o dobro da anterior, e $s^2 = 20,68$, ou seja quatro vezes a variância anterior.

(d) Neste caso apenas a média sofre alteração. Na nova situação $\overline{x} = 5,65$.

17. Foram medidas a altura e o peso de um grupo de homens e de um grupo de mulheres...

    (a) $cv = 2\%$. Como esta medida é invariante quanto à escala o valor não se altera qualquer que seja a unidade de medida usada.

    (b) $cv = 2\%$ (para a altura das mulheres) e $cv = 5\%$ (para o peso das mulheres). $cv = 1,8\%$ (para a altura dos homens) e $cv = 4,5\%$ (para o peso dos homens). Tanto nos homens como nas mulheres o peso está mais disperso do que a altura, apesar de nas mulheres essa dispersão ser maior.

18. No quadro seguinte indicam-se os preços dum bem alimentar...

    $c)$ r=-0.926, associação linear negativa muito forte.

2.2 *Probabilidades*

1. Na empresa Sojoga foi decidido que ....
   (a) 0,28
   (b) 0,1944
   (c) 0,6271
2. Estudos recentes revelaram que, em Portugal, ...
   (a) 0,44
   (b) 0,168
3. Estudos disponíveis em www.pordata.pt, acerca do tráfego telefónico nacional ...
   (a) 0,25
   (b) 0,2503
   (c) 0,9191
4. Uma fábrica dispõe de 3 máquinas para produção de peças electrónicas...
   (a) 0,982
   (b) A máquina 3.
5. Para estudarem para uma frequência, sabe-se que...
   (a) 0,6
   (b) 0,352
   (c) 0,402
   (d) 0,198
   (e) 0,33
6. Numa determinada região 528 indivíduos foram classificados segundo duas características qualitativas...
   (a)
   (b) 0,523
   (c) 0,027
   (d) 0,496
   (e) 0,099

	B_1	B_2	B_3
A_1	56	71	12
A_2	47	163	38
A_3	14	42	85

(f) 0,591

7. Considere dois centros de Sondagens, centro XP e centro XPTO....

$A -$ "XP acerta na previsão" e $B -$ "XPTO acerta na previsão"

(a) $P(A \cap B) = 0.5$

(b) $P((\overline{A} \cap B) \cup (A \cap \overline{B})) = 0.3$

(c) $P(A \cup B) = 0.8$

(d) $P(\overline{A}|B) = 0.091$

8. Um conjunto de três atelier de Arquitectura (digamos atelier A, B e C) pretende elaborar conjuntamente um projecto para apresentar na Trienal de Arquitectura 2007...
Sejam:

A="A tarefa do atelier A é executada a tempo"

B="A tarefa do atelier B é executada a tempo"

C="A tarefa do atelier C é executada a tempo"

(a) $P(A \cap B \cap C) = 0,504$.

(b) $P(A \cap (\overline{B} \cup \overline{C})) = 0,296$.

(c) $P(A \cap B|C) = 0,56$.

9. Do conjunto de empresas que actuam num dado sector industrial, 25% possuem departamento de investigação, 50% realizam lucros...

(a) 0,55

(b) 0,75

(c) 0,45

(d) 0,8

(e) 0,05

(f) 0,3

10. Em determinada população existem 60% mulheres das quais 6% têm mais de 170 cm de altura...
R: 0,82

11. Sejam A e B dois acontecimentos tais que $P(A) = \frac{3}{8}$, $P(B) = \frac{1}{2}$ e $P(A \cap B) = \frac{1}{4}$...

 (a) 5/8
 (b) 5/8
 (c) 1/2
 (d) 3/8

12.
R: $P(B \cap \overline{A}) = 1/6$

13.
R: $P(B \cup C) = 0.7$

14. Considere os acontecimentos A, B e C de probabilidades não nulas...
R: $P(A \cup B \cup C) = 0.59$

15. Numa fábrica, um certo tipo de peças para um equipamento de laboratório...

 (a) 0,0278
 (b) 0,2481

16. Num laboratório um investigador fez um ensaio com 3 classes de bactérias A, B e C...

Seja A — "bactérias da classe A"; B — "bactérias da classe B"; C — "bactérias da classe C"; D — "Reacção à presença de sulfatos".

 (a) $P(A \cap D) = 0.08$
 (b) $P(\overline{D}) = 0.5$
 (c) $P(B|\overline{D}) = 0.24$

17. O Nunes consegue à última da hora um bilhete para assistir às primeiras Meias-Finais do Campeonato Europeu de Futebol de 2008, que está a decorrer na Austria e na Suiça...

(a) $P(T_2|A) = 0.1079$
(b) $P(A) = 0.695$

18. Numa determinada fábrica do Distrito de Évora, três inspectores o Sertório, o Geraldes e o Gastão, inspeccionam respectivamente, 20%, 30% e 50% dos artigos produzidos por essa fábrica...
    (a) $P(D) = P(D|S)P(S) + P(D|G)P(G) + P(D|Ga)P(Ga) = 0,05 \times 0,2 + 0,1 \times 0,3 + 0,15 \times 0,5 = 0,115$
    (b) $P(S|D) = 0,087$; $P(G|D) = 0,261$; $P(Ga|D) = 0,652$. Pelo foi mais provável este artigo defeituoso ter sido inspeccionado pelo inspector Gastão.

19. Numa urbanização recente um inquérito aos moradores revelou que 5% viviam em moradias, 20% em prédios em banda e os restantes em torres...
    (a) 0,082
    (b) 0,0122

20. A poluição do ar de certa cidade é causada essencialmente por gases industriais (75% dos casos) e por gases dos escapes de automóveis (25% dos casos)...
    (a) 0,675
    (b) 0,222

21. Durante a travessia do Canal da Mancha, a probabilidade de um velejador apanhar mau tempo é de 2/3...
    (a) 7/9
    (b) 1/4

2.3 *Variáveis Aleatórias*

1. Sejam X e Y duas variáveis aleatórias, em que X toma valores de 2, 3 e 4 e $Y = 3X$...

   (a)

x	2	3	4
$f(x)$	0,3	0,4	0,3

y	6	9	12
$f(y)$	0,3	0,4	0,3

   (b)

$$F_Y(y) = \begin{cases} 0, & y < 6 \\ 0.3, & 6 \le y < 9 \\ 0.7, & 9 \le x < 12 \\ 1, & x \ge 12 \end{cases}$$

   (c) $P(X \le 3) = 0,7$, $P(Y = 10) = 0$, $P(Y > 7) = 0,7$ e $P(Y \le 9|X \ge 3) = 0,5714$

   (d)

$X\backslash Y$	6	9	12
2	0,3	0	0
3	0	0,4	0
4	0	0	0,3

   (e) $\mu_X = E(X) = 3$, $\sigma_X^2 = 0,6$, $\mu_Y = E(Y) = 9$ e $\sigma_Y^2 = 5,4$.

   (f) $\rho_{(X,Y)} = 1$, as variáveis apresentam uma associação linear positiva perfeita.

2. Sejam X e Y duas variáveis aleatórias. Sabe-se que X é uma variável aleatória, com valor esperado 2.21...

   (a) **A**=0.15, **B**=0.53 e **C**=0.05.

   (b)

$$F_Y(y) = \begin{cases} 0, & y < 2 \\ 0.15, & 2 \le y < 5 \\ 1, & x \ge 5 \end{cases}$$

   (c) $\mu_Y = E(Y) = 4.55$ e $\sigma_Y = 1.07$

   (d) $P(Y = 5|X = 2) = 1$

   (e) $\sigma_{2X|Y=2} = 1.88$.

   (f) As variáveis X e Y não são independentes.

3. Sejam X e Y duas variáveis aleatórias independentes. Ambas as variáveis podem assumir valor 2, 3 ou 5...

   (a) Verdadeira. Como X e Y são independentes,
   $P[X = 2, Y < 3] = P[X = 2]P[Y < 3] = 0,35$ e como $P[X = 2] = 0,7$, logo $P[Y < 3] = P[Y = 2] = 0.5$.

   (b) Falsa. Como X e Y são independentes,
   $P[X = 5|Y \geq 3] = P[X = 5] = 0.1 \neq 0.2$

   (c) Verdadeira. Como X e Y são independentes,
   $E[X|Y = 3] = E[X] = 2.5$ logo $E[4X + 12] = 4E[X] + 12 = 22$.

   (d) Falsa. $Var[8X - Y] = 64Var[X] + Var[Y]$

   (e) Falsa. Como X e Y são independentes, $Cov[X, Y] = 0$.

4. Na tabela seguinte apresenta-se a função de probabilidade relativa ao número de exercícios...

   (a)
   $$F_X(x) = \begin{cases} 0, & x < 0 \\ 0.2, & 0 \leq x < 1 \\ 0.3, & 1 \leq x < 2 \\ 0.7, & 2 \leq x < 3 \\ 0.8, & 3 \leq x < 4 \\ 1, & x \geq 4 \end{cases}$$

   (b) $P(X \leq 2) = 0.7$

   (c) $P(X = 0|X < 2) = 0.667$

   (d) $E(Y) = 7$ e $Var[Y] = 28.8$

5. Suponha que numa determinada cidade, o número de assoalhadas por casa...

   (a) A=0.3 e B=0.1

   (b)
   $$F_X(x) = \begin{cases} 0, & x < 1 \\ 0.4, & 1 \leq x < 2 \\ 0.7, & 2 \leq x < 3 \\ 0.9, & 3 \leq x < 4 \\ 1, & x \geq 4 \end{cases}$$

   (c) $P(X \leq 3) = 0.9$

   (d) $P(X \geq 2) = 0.6$

   (e) $\mu_X = E(X) = 2$ e $\sigma_X = 1$

6. Seja X uma variável aleatória à qual corresponde a seguinte função massa de probabilidade...
   (a)
   (b) $P(X \leq 2) = 0.25$
   (c) $E[2X - 1] = 6$ e $Var[2X - 1] = 7$

7. Seja X a variável aleatória que representa o número de vezes que cada aluno das várias Licenciaturas da Universidade de Évora...
   (a) $P(X = 1) = 0.5625$
   (b) $P(X \leq 2) = 0.9375$
   (c)
   $$F_X(x) = \begin{cases} 0, & x < 1 \\ 0.5625, & 1 \leq x < 2 \\ 0.9375, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$
   (d) $E(X) = 1.5$
   (e)
   (f) $Var[2X - 1] = 1.5$

8. A fim de analisar a capacidade de germinação de sementes em certo cereal foram semeadas cinco sementes em cada vaso de um conjunto de vasos iguais...
   (a)

x_i	0	1	2	3	4	5
$f(x_i)$	16/395	32/395	87/395	137/395	98/395	25/395

   (b) 0,6886
   (c) 0,8785
   (d) 0,8152
   (e) 1.

9. Considere a variável aleatória discreta X que assume os valores $\{-2, -1, 0, 1, 2\}$ cuja função massa de probabilidade é:

(a) 17/60

(b)

y_i	0	6	24
$f(y_i)$	37/60	6/60	17/60

$P(Y > 0) = 23/60.$

10. Seja X uma dada variável aleatória contínua com a seguinte função densidade de probabilidade

(a)

(b) $P[X < 1] = 1/4$ e $P[1 < X \leq 3] = 3/4$.

11. Seja X uma variável aleatória que associa a cada indivíduo o tempo ....

(a)

$$F(x) = \begin{cases} 0, & x < 50 \\ \frac{x}{20} - \frac{50}{20}, & 50 \leq x \leq 70 \\ 1, & x > 70 \end{cases}$$

(b) $P(61 \leq X \leq 70) = 0,45$

(c) $E(X) = 60$

12. A distribuição conjunta das variáveis aleatórias, independentes, X e Y ...

(a)

$X \backslash Y$	0	1	2	$f(x)$
1	0,06	0,12	0,02	0,2
2	0,15	0,30	0,05	0,5
3	0,09	0,18	0,03	0,3
$f(y)$	0,3	0,6	0,1	

(b)

$$F_X(x) = \begin{cases} 0, & x < 1 \\ 0.2, & 1 \leq x < 2 \\ 0.7, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

(c) $E(X|Y=2) = 2,1$.

(d) $E(Z) = 1$ e $Var(Z) = 7,72$.

13. Suponha que numa determinada cidade, o número de filhos por família (X) e o número de assoalhadas por casa (e por família) (Y) ...

(a)

x_i	2	3	4
$f_X(x_i)$	0,13	0,35	0,52

(b)

$$F_Y(y) = \begin{cases} 0, & y < 1 \\ 0.4, & 1 \le y < 2 \\ 0.7, & 2 \le y < 3 \\ 1, & y \ge 3 \end{cases}$$

(c) $P(Y \ge 2) = 0,6$

(d) $P(Y = 2|X = 3) = 0,4286$

(e) $E(X|Y = 3) = 3,83$

14. Considere que a editora "D. Quixote e Sr. Pança"possui duas máquinas impressoras...

(a)

x_i	0	1	2	3
$f(x_i)$	0.1875	0,125	0,3125	0,375

(b) $F(1,2) = P(X \le 1, Y \le 2) = 0.3125$

(c) $P(Y = 1|X = 0) = 0.667$

15. Sejam X e Y duas variáveis aleatórias discretas que representam, respectivamente, o número de avarias graves num dado veículo automóvel antes da instalação de GPL e depois da instalação de GPL...

(a) $a = 0.2$, $b = 0.1$, $c = 0.1$ e $d = 0.1$

(b) $P(X = 0, Y \le 1) = 0.5$

(c) $P(X = 0|Y = 2) = 0.5$

(d) $Corr(X,Y) = 0.1166$. O número de avarias graves num dado veículo automóvel antes e depois da instalação do GPL estão fracamente relacionados.

(e) $Var[Y|X = 0] = 0.4724$ e $Var[Y|X = 1] = 0.5$

16. Uma possível codificação de um jogo de computador é a seguinte: O jogador escolhe ao acaso uma das três opções $(-1,0,1)$ e...

(a)

x_i	-1	0	1
$f_X(x_i)$	$1/3$	$1/3$	$1/3$

y_i	0	1	2
$f_Y(y_i)$	$1/3$	$1/3$	$1/3$

(b) $E[X] = 0$.

(c)

$X \backslash Y$	0	1	2
-1	$1/18$	$1/6$	$1/9$
0	$4/18$	$1/18$	$1/18$
1	$1/18$	$1/9$	$1/6$

(d) $Var[XY] = 1.3858$.

(e)

y_i	0	1	2
$f_{Y\|X=0}(y_i)$	$4/6$	$1/6$	$1/6$

17. Com o objectivo de analisar o tipo de precaução que os alunos de outras licenciaturas, que não a de Eng. Informática, tomam relativamente ao seu computador...

(a)

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 0.54, & 0 \leq x < 1 \\ 0.88, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

(b) $F(1,1) = 0,132$

(c) 0,012

(d) $Var[X] = 0,4836$

(e) $E[W] = 4,98$

(f) $E[XY] = 1,276$

18. Voltemos à questão colocada na frequência anterior, em que se pretendia analisar o tipo de precaução que os alunos de outras licenciaturas (que não a de Eng. Informática) tomavam relativamente ao seu computador...

(a)

y_i	0	1	2	3
$f(y_i)$	0	0,1	0,5	0,4

(b) 0,9

(c) 5,7

(d) $Cov(X,Y) = -0,04008$. Conclusão?

(e) $Cov(Z,W) = 0,08016$

19. Sejam X e Y duas variáveis aleatórias com função massa de probabilidade conjunta $f(x,y)$, da forma:

$$f(x,y) = k(2x+y),\ x = 0,1,2 \quad \text{e} \quad y = 0,1,2,3.$$

(a) $k = 1/42$

(b) $P(X \geq 1, Y \leq 2) \approx 0,57143$

(c)

x_i	0	1	2
$f(x_i \mid y=0)$	0	1/3	2/3

(d) $\approx 0,222$

20. Sejam X e Y duas variáveis aleatórias discretas...

(a)

x_i	0	1	2
$f(x_i)$	0.3	0.2	0.5

$X \backslash Y$	3	4	5
0	0,06	0,14	0,1
1	0,04	0,11	0,05
2	0	0,3	0,2

(b)

(c) 0,6.

(d) 0,9275.

(e) 16,85.

21. O Evaristo pode encontrar-se num dos três estados de espírito: **1** - radiante; **2** - assim-assim; **3** - macambúzio...

(a) $a = 0,03$; $b = 0,16$; $c = 0,002$ e $d = 0,488$.

(b)

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ 0,1, & 0 \leq y < 1 \\ 0,3, & 1 \leq y < 2 \\ 1, & y \geq 2 \end{cases}$$

(c) $E[XY] = 3,81$

(d) $F(3,1) = 0,3$.

(e) $Var[X|Y = 0] = 0,2644$

2.4 Distribuições de Probabilidade

1. **(1ª Frequência 2013)** Estudos recentes revelaram que, em Portugal, no 4º trimestre de 2012...

   (c) 0.2377.

2. **(1ª Frequência 2013)** A energia eólica representa o aproveitamento da energia ...

   (a) 0.0237
   (b) 9.49
   (c) 0.1587

3. **(1ª Frequência 2013)** Seja X uma variável aleatória que representa a produção a...

   (a) 8400
   (b) i. 0.6736
       ii. 0.1355

4. **(2ª Frequência/Exame 2013)** Uma determinada empresa pretende recrutar funcionários...

   (a) 0.1587
   (b) 0.5403
   (c) 50%
   (d) 0.0068

5. **(Exame de recurso 2013)** Uma fábrica produz discos de travão para automóveis...

   (a) i. 0.0808
       ii. 0.4967
   (b) *"Em 93.7% dos casos o diâmetro do disco de travão não ultrapassa os 322.0765 mm".*
   (c) 0.776

6. **(1ª Frequência — 31 de Março de 2012)** Uma conhecida fábrica de cerâmica produz azulejos que são embalados ....

   (a) 0.2707
   (b) 0.6767

(c) 0.0088

7. **(Exame Época Normal — 14 de Junho de 2012)** O número de viaturas que atestam o depósito de gasolina numa pacata ...

   (a) 0.3953
   (b) 13.5
   (c) 0.0945

8. **(1ª Frequência — 30 de Abril de 2011)** Sabe-se que o número de alunos que chegam, por hora, para tirar dúvidas com a docente A.... Seja X="nº de alunos que chegam para tirar dúvidas com o docente A, numa hora", tal que $X \frown P(1)$.

   (a) $E[Y] = 4$.
   (b) $P(Y \leq 2) = f(0) + f(1) + f(2) \simeq 0.2381$.
   (c) Sejam Y_1="nº de alunos que chegam para tirar dúvidas com o docente A, em 5 dias", tal que $Y_1 \frown P(20)$ e Y_2="nº de alunos que chegam para tirar dúvidas com o docente B, em 5 dias", tal que $Y_2 \frown P(10)$. Então $Y_1 + Y_2 \frown P(30) \Rightarrow Y_1 + Y_2 \frown^a N(30, \sigma^2 = 30)$. Logo, $P(Y_1 + Y_2 \leq 20) = P(Z \leq \frac{20-30+0.5}{\sqrt{30}}) = 0.0418$.

9. **(Exame Época Normal — 4 de Julho de 2011)** Um dos elevadores dum grande edifício público transporta, no máximo 20 pessoas de cada vez...

   (a) Seja X="nº de pessoas com peso superior a 40kg e inferior a 85Kg", tal que $Y \frown B(20, 0.9739)$
   $P(Y \leq 2) \simeq 0$.
   (b) $P(\overline{X} > 60) = P(Z > -0.45) = 0.6736$.

10. **(2ª Frequência — 16 de Maio de 2009)** No âmbito de um estudo sobre a criação de um novo Centro Comercial na cidade de Évora...

    (a) 0.0001049
    (b) 0.9536
    (c) 6
    (d) 1.5492.

11. **(2ª Frequência — 16 de Maio de 2009)** Um exemplo clássico da distribuição de Poisson envolve o número de militares do exército Prussiano...

(a) 0.9817
(b) 1/6
(c) 0.9990

12. **(2ª Frequência — 16 de Maio de 2009)** Numa pequena e média empresa de fornecimento de materiais de construção...

    (a) 0.0062
    (b) 0.9876
    (c) 2
    (d) $\simeq 1$

13. **(3ª Frequência/Exame normal — 18 de Junho de 2009)** Através de estudos realizados, é possível admitir que ocorre no Japão...

    (a) i. 0.2707
        ii. 0.44568
    (b) 0.2389.

14. **(Exame de recurso — 27 de Junho de 2012)** Considere que a quantidade de gasolina vendida numa manhã...

    (a) 0.0019
    (b) 0.1251
    (c) ≈ 1

15. **(3ª Frequência/Exame normal — 18 de Junho de 2009)** Um posto de transformação permite uma carga total de 2800kW...

    (a) 0.8185
    (b) 0.0359
    (c) $E[Z] = 250$ e $Var[Z] = 31.25$
    (d) 1.48.

16. **(Exame de Recurso — 9 de Julho de 2009)** A resistência ao choque de certo tipo de mosaico cerâmico tem...

    (a) 0.6826

(b) 2.58

(c) $\simeq 0$.

17. **(Exame de Época Normal — 21 de Junho de 2008)** Numa fábrica existem 15 máquinas, 8 são novas e 7 são antigas...

    (a) $\simeq 0.6359$

    (b) $\simeq 0.1040$.

18. **(Exame de Recurso — 12 de Julho de 2008)** Os portugueses são dos cidadãos europeus com menor envergadura física...

    (a) 0,00549

    (b) 0,8445

    (c) $\simeq 0,756$.

19. **(2ª Frequência — 18 de Junho de 2007)** Estudos tecnológicos para a qualificação de certo tipo de rochas naturais e/ou ornamentais...

    (a) $P(X \leq 5300) = 1 - P(Z < 2,83) = 0,0023$.

    (b) $P(X > x) = 0,90 \Rightarrow P(Z \leq \frac{x-7000}{600}) = 0,1 \Rightarrow x = 6229$.

    (c) $P(Y = 4) = 3,08 \times 10^{-14}$ onde $Y \sim B(10; 0,9977)$.

20. **(Exame Recurso — 7 de Julho de 2007)** O número de fendas significativas numa auto-estrada...

    Seja X="número de fendas significativas numa auto-estrada, a ponto de exigirem reparação imediata"(por Km), tal que $X \frown P(\lambda = 4)$.

    (a) $P(X = 0) = e^{-12}$

    (b) Seja Y="O número de conjuntos de 8 fendas em 3 Km". Tal que
    $Y \frown B(10; 0,0655) \qquad (p = P(X = 8) = \frac{e^{-12}12^8}{8!})$.
    Logo $P(Y = 5) = 0,000217$.

    (c) $P(X < 18) = 1 - P(Z \leq 1,5) = 0,0668$.

21. **(Exame Especial — 27 de Setembro de 2007)** A pluviosidade anual, medida em cm^3/m^2, numa determinada região tem sido estudada...

    (a) $\sigma = 19,23$.

(b) 0,3508.

22. (a) 0,606
    (b) 0,865
    (c) 0,0302

23. (a) $f(x) = 0,25\, e^{-0,25x}, \quad x > 0$
    (b) $E[X] = 4$ e $Var[X] = 16$
    (c) 0,3247

2.5 Intervalos de Confiança e Testes de Hipóteses

1. O departamento de recursos humanos da empresa que pretende recrutar funcionários para uma nova área de negócio,...

   (a) $H_0: \mu_X \leq 115$ *vs* $H_1: \mu_X > 115$.
   Podemos admitir a normalidade da classificação no teste dos homens através da interpretação do *p-value* do teste de Shapiro-Wilk ($n < 50$): *p-value*=0,425, logo não rejeitamos a hipótese $H_0 : X \sim \mathcal{N}(\mu, \sigma)$ para 1% (5% ou 10%) de significância, uma vez que $0,01\,(0,05 \text{ ou } 0,1) \ngeq 0,425$.
   $t_{obs} = -4,314 \not> t_{14,0.99} = 2,624$, logo não rejeitamos a hipótese nula para $\alpha = 1\%$. Não existe evidência estatística que nos permita concluir que os candidatos do sexo masculino verificam o pressuposto.

   (b) *p-value* $\simeq$ 0,9995.

   (c) $]136,002; 489,814[$

   (d) $H_0: \mu_X = \mu_Y$ *vs* $H_1: \mu_X \neq \mu_Y$
   Para além da normalidade da classificação no teste dos homens, podemos admitir a normalidade da classificação no teste das mulheres através da interpretação do *p-value* do teste de Shapiro-Wilk ($n < 50$): *p-value*=0,505, uma vez que não rejeitamos a hipótese $H_0 : Y \sim \mathcal{N}(\mu, \sigma)$ para 1% (5% ou 10%) de significância, uma vez que $0,01\,(0,05 \text{ ou } 0,1) \ngeq 0,505$.
   Através da interpretação do teste de Levene ($H_0 : \sigma^2_X = \sigma^2_Y$ *vs* $H_0 : \sigma^2_X \neq \sigma^2_Y$), concluímos que não rejeitamos H_0, uma vez que *p-value*=0,390 (> qualquer um dos níveis de significância usuais). Logo podemos assumir a igualdade das variâncias.
   Como $|t_{obs}| = |\frac{-3,8}{6,175}| = 0,615 \not> t_{28,0.995} = 2,763$, não rejeitamos $H_0 : \mu_X = \mu_Y$ para $\alpha = 1\%$. Não existe evidência estatística que permita afirmar que as classificações médias no teste diferem significativamente entre sexos.
   (Uma forma alternativa que nos permitia concluir o mesmo seria através do intervalo a 90% de confiança para a diferença de médias:
   limite superior = 6,705 $\Rightarrow \bar{x} - \bar{y} + \varepsilon = 6,705 \Rightarrow \varepsilon = 6,705 - (-3,8) = 10,505$. Logo
   limite inferior = -3,8 - 10,505 = -14,305.
   Como $0 \in IC_{90\%}(\mu_X - \mu_Y) =]-14,305; 6,705[$, também irá pertencer ao $IC_{99\%}(\mu_X - \mu_Y)$, pelo que se chegaria à mesma conclusão a 1% de significância).

2. Um fabricante da indústria cerâmica pretende determinar se duas novas ligas...

   (a) Podemos admitir a normalidade da temperatura máxima de resistência ao calor da liga *premium* nacional e da liga *standard*, uma vez que para o teste de Shapiro-Wilk ($n < 50$): *p-value*$_{premium\ nacional}$=0,129 e *p-value*$_{standard}$=0,825, logo

não rejeitamos as hipóteses $H_0 : X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$ e $H_0 : X_3 \sim \mathcal{N}(\mu_3, \sigma_3)$ para 1% (5% ou 10%) de significância, uma vez que qualquer um dos $\alpha' s$ usuais $\ngeq 0,129$ ou $0,825$.
No caso da liga *premium* importada, já não podemos admitir a normalidade uma vez que *p-value*$_{premium\ importada}< 0,001$, o que conduz à rejeição de $H_0 : X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$, pois qualquer um dos $\alpha' s$ usuais $\geq$ p-value$_{premium\ importada}$.

(b) $H_0 : \mu_2 = 1535$ *vs* $H_1 : \mu_2 \neq 1535$.
$t_{obs} = 3,869 > t_{19,0.95} = 1,729$, logo rejeitamos a hipótese nula para $\alpha = 10\%$.
Existe evidência estatística de que a temperatura média de resistência ao calor da liga *premium* nacional é significativamente diferente de 1535.

(c) *p-value* $\simeq$ 0,001.

(d) (**Atenção:** Para a resolução desta alínea é necessário o valor da variância amostral da temperatura de resistência ao calor da liga *standard* que era apresentada neste exame numa primeira tabela que se pode encontrar no exercício 3 da Ficha n.º 1: $s_3^2 = 10,09^2$). Podemos dizer que $\sigma_3^2 \in]71,12; 166,04[$ com 80% de confiança.

(e) Através da interpretação do teste de Levene ($H_0 : \sigma_2^2 = \sigma_3^2$ *vs* $H_0 : \sigma_2^2 \neq \sigma_3^2$), concluímos que não rejeitamos H_0, uma vez que *p-value*=0,2 ($>$ qualquer um dos níveis de significância usuais). Logo podemos assumir a igualdade das variâncias.
Sabe-se relativamente ao intervalo a 99% de confiança para a diferença de médias:
limite superior = 33,04816 $\Rightarrow \bar{x_2} - \bar{x_3} + \varepsilon = 33,04816 \Rightarrow \varepsilon = 33,04816 - 25,42954 = 7,61862$. Logo
limite inferior = 25,42954 - 7,61862 = 17,81092.
Como o não está contido no $IC_{99\%}(\mu_2 - \mu_3) =]17,81092; 33,04816[$, podemos concluir que a temperatura média de resistência ao calor da liga premium nacional é significativamente diferente da temperatura média de resistência ao calor da liga standard, para 1% de significância.

3. Foram retiradas 25 peças da produção diária de uma máquina...

   (a) $]4.805; 5.595[$; $]4.730; 5.670[$; $]4.582; 5.818[$

   (b) $]4.789; 5.611[$; $]4.705; 5.695[$; $]4.529; 5.871[$

4. Um conjunto de 40 condutores de camião, escolhidos aleatoriamente

   (a) $]0.786; 0.914[$

   (b) $]0.024; 0.078[$

5. Pretende-se analisar os salários, por sexo, do pessoal...
   (a) $]2571.86; 2640.98[$
   (b) A = 469; B = 1098; C = -225.69960

6. Registou-se o comprimento, em metros, dos saltos de 10 atletas portugueses do sexo masculino em provas de triplo salto em pista coberta...
   (a) $1 - \alpha = 0.9652$.
   (b) $]0.0244; 0.3315[$.
   (c) Dado que $t = -1.2511$, logo menor do que 1.383, não se rejeita a hipótese nula para uma significância de 10%.

7. O Serviço Nacional de Saúde (SNS) afirma que a proporção de asmáticos numa certa população masculina é inferior a 10%.
   (a) $H_0 : p \leq 0,1 \quad vs \quad H_1 : p > 0,1$. Como $z_{Obs} = 2,593$ existe evidência suficiente nos resultados para afirmar (ao nível de significância de 5%) que a proporção de asmáticos numa certa população masculina é superior a 10%. Pelo que o médico deve avisar o SNS de que a sua estimativa não está correcta.
   (b) *p-value* =0,0048
   (c) Potência de teste = 0,9515
   (d) IC a 95% para a proporção : $]0,1048; 0,2052[$.

8. Foram efectuados estudos em Lisboa com o objectivo de determinar a concentração de monóxido de carbono (CO) perto de vias rápidas.
   (a) Como $\chi^2_{obs} = 20.9$ não se rejeita H_0 para $\alpha = 1\%$.
   (b) $]98.473; 102.527[$.
   (c) $n \geq 1174$.

9. Num trabalho realizado há já algum tempo concluiu-se que 62% dos ...
   $]0.472; 0.578[$

10. Certa linha de fabrico está programada de modo a produzir uma percentagem de artigos defeituosos não superior a 3%.

(a) $H_0: \; p \leq 0,03 \quad vs \quad H_1: \; p > 0,03$

$$z_{Obs} = \sqrt{50}\frac{0,04-0,03}{\sqrt{0,03*(1-0,03)}} = 0,4145$$

Não se rejeita a hipótese nula, considerando um nível de significância de 5%. Ou seja, não existe evidência estatística suficiente para afirmar que o processo se encontrava fora de controlo, pelo que o encarregado agiu mal.

(b) $p - value = P(Z > 0,4145) = 0,3409.$

(c) $P(\text{Rejeitar}H_0|H_0\text{falsa}) = P(\overline{P} > 0,0697|p_0 = 0,035) = 0,0901.$

$P(\text{Rejeitar}H_0|H_0\text{falsa}) = P(\overline{P} > 0,0697|p_0 = 0,1) = 0,7611.$

11. A poluição atmosférica é medida em dois locais distintos, um no centro de uma pequena cidade (Y) e outro numa zona rural, 15Km mais a sul, (X)...

(a) $]-0.4992 - 2.206 * 0.51658; -0.4992 + 2.206 * 0.51658[=]-1.6389; 0.64027[.$ Existe evidência estatística suficiente para afirmar (com uma significância de 3%) que, em média, a poluição atmosférica é igual nos dois locais.

(b) $]1,7879; 6.0861[$

12. Dois laboratórios (A e B) avaliam a quantidade de cloro de amostras de água recolhidas à mesma hora de cada dia.

(a) Pelo 1º output não se rejeita (para um nível se significância de 5%) a hipótese nula da normalidade das população, quer para o Laboratório A (p-value=0.176) quer para o B (p-value=0.147). De igual modo, não se rejeita a hipótese nula de igualdade de variância pelo teste de Levene (p-value=0.952).

$]-0,5482; 0,4966[.$

(b) $H_0: \; \mu_A = \mu_B \quad vs \quad H_1: \; \mu_A \neq \mu_B$

Como p-value=0.916 não se rejeita H_0 para qualquer α=1%.

13. Um investigador pretende estudar a capacidade de concentração dos alunos do ensino universitário antes e depois do almoço.

(a) Como o p-value é igual 0.127 não se rejeita a hipótese de normalidade da diferença dos dados.
$]-0,7524; 3,5524[$

(b) Como $0 \in]-0,7524; 3,5524[$ não se rejeita (para uma significância de 2%) a hipótese nula. Ou seja, não existe evidência suficiente nos resultados para afirmar que existe diferença entre a capacidade média de concentração antes e depois do almoço.

(c) Como $\chi^2_{Obs} = 14,092$ não se rejeita (para uma significância de 5%) a hipótese nula. Ou seja, não existe evidência suficiente nos resultados para afirmar que a variabilidade da capacidade de concentração antes do almoço é diferente de 10.

(d) *p-value*=0,238

(e) Como $t_{Obs} = \frac{\overline{x}-k}{3{,}957}\sqrt{10}$ então rejeita-se H_0 se $\frac{\overline{x}-k}{3{,}957}\sqrt{10} > 1,833$. Donde se retira que $\overline{x} > k + 1,833\frac{3{,}957}{\sqrt{10}}$. Então, $k + 1,833\frac{3{,}957}{\sqrt{10}} = 56 \Rightarrow k = 53,706$.

14. A uma eleição concorrem três candidatos A, B e C.

(a) $]0,4130; 0,5730[$.

(b) $z_{obs} = -0,816 > -1,645$ não se rejeita H_0, com uma confiança de 95%.

(c) *p-value*=0,2061.

(d) Se $n_i = 23$.

15. Suponha que o teor de nicotina de duas marcas de cigarros foi analisado...

(a) Como $z_{obs} > 2,326$ rejeita-se H_0, pelo que existe forte evidência estatística para afirmar, com uma confiança de 99%, que o teor médio de nicotina da marca A é superior ao da marca B.

(b) *p-value*$\simeq 0$

(c) Potência do teste=$P(\overline{X} > 2,5713|\mu_1 = 2,1) + P(\overline{X} < 2,4287|\mu_1 = 2,1) \simeq 1$, para uma confiança de 95%.

16. Para comparar a resistência ao esforço físico de duas populações, A e B...

(a) $\overline{y} = 15,882$ e $s_Y = 3,343$.

(b) $]14,0557; 17,7083[$.

(c) $1 - \alpha = 0,98$.

17. Com o objectivo de estudar algumas características dos jogadores de futebol que participam no Campeonato Europeu de Futebol de 2008...

(a) Pelo 1º output não se rejeita (para um nível se significância de 5%) a hipótese nula da normalidade das população, quer para Altura dos jogadores Checos (p-value=0.574) quer para a Altura dos jogadores Gregos (p-value=0.376). De igual modo, não se rejeita a hipótese nula de igualdade de variância pelo teste de Levene (p-value=0.317).

$]-7.3374; 3.3374[$.

(b) 5.3374.

(c) Não se rejeita (para uma significância de 5%) a hipótese nula. Ou seja, não existe evidência suficiente nos resultados para afirmar que existe diferença entre as alturas médias dos jogadores.

(d) Não existe evidência suficiente nos resultados para (com $\alpha = 5\%$) afirmar que a altura média dos jogadores checos é superior a 175cm.

(e) *p-value*=0.0375. Conclusão?

(f) $\overline{x} = 179.34$.

(g) Não existe evidência suficiente nos resultados para afirmar (com $\alpha = 1\%$) que a variância das alturas dos jogadores checos é superior à variância das alturas dos jogadores gregos.

18. De 72 jogadores inquiridos, 36 jogam em equipas estrangeiras, 34 em equipas do país de origem e 2 não têm equipa.

    A notícia é verdadeira. Existe evidência estatística que permite concordar com a notícia, ao nível de significância de 5%.

19. Foi levado a cabo um estudo para averiguar se a ausência às aulas durante o semestre de Inverno é maior num centro urbano do norte ou do sul.

    (a)]0.0114; 0.1602[

    (b) Rejeitamos H_0, ao nível de significância de 1%. Existe uma forte evidência estatística de que no Inverno se falta mais às aulas na região do Norte.

    (c) 0.0012

    (d) 0.97

20. Tome-se o seguinte exemplo, relativo a dois tipos de geradores (I e II)...

    (a) Não existe evidência estatística suficiente (para um $\alpha = 1\%$) para afirmar que o valor esperado da produção de energia eléctrica é diferente nos dois geradores.

(b) *p-value*$\simeq 1$. Conclusão?

(c) Não existe evidência estatística suficiente (para um $\alpha = 5\%$) afirma que o desvio padrão da produção de energia eléctrica através do gerador II é diferente de 4KW/h.

(d) $]0.5077; 3.5989[$.

2.6 Testes Não-Paramétricos

1. $H_0 : X$ segue uma distribuição normal *vs* $H_0 : X$ não segue uma distribuição normal
   Interpretamos o teste de Shapiro-Wilk ($n < 50$)
   p-value<0.001, logo rejeitamos a hipótese nula para qualquer um dos níveis de significância usuais (1%, 5% ou 10%). Existe evidência estatística de que a amostra não provem de uma população com distribuição normal

2. X- medição do grupo Placebo; Y- medição do grupo Tratamento
   $H_0 : X$ segue uma distribuição normal *vs* $H_1 : X$ não segue uma distribuição normal
   Interpretamos o teste de Shapiro-Wilk ($n < 50$)
   p-value<0.001, logo rejeitamos a hipótese nula para qualquer um dos níveis de significância usuais (1%, 5% ou 10% $\geq$*p-value*). Existe evidência estatística de as medições do grupo placebo não verificam a normalidade.

   $H_0 : Y$ segue uma distribuição normal *vs* $H_0 : Y$ não segue uma distribuição normal
   Interpretamos o teste de Shapiro-Wilk ($n < 50$)
   p-value=0.0.773, logo não rejeitamos a hipótese nula para qualquer um dos níveis de significância usuais (1%, 5% ou 10% $\ngeq$). Não existe evidência estatística que permita concluir que as medições do grupo tratamento não verificam a normalidade.

3. $H_0 : p_1 = 0.579$, $p_2 = 0.347$, $p_3 = 0.069$, $p_4 = 0.005$
   $H_1 : p_1 \neq 0.579$, $p_2 \neq 0.347$, $p_3 \neq 0.069$, $p_4 \neq 0.005$
   $\chi^2_{obs} = 17.23 > \chi^2_{2,0.95} = 5.991$, logo rejeitamos a hipótese nula para $\alpha = 5\%$. Os dados evidenciam o dado é falso.

4. (a) $H_0 : p_1 = 0.4$, $p_2 = 0.3$, $p_3 = 0.2$, $p_4 = 0.1$
   $H_1 : p_1 \neq 0.4$, $p_2 \neq 0.3$, $p_3 \neq 0.2$, $p_4 \neq 0.1$
   $\chi^2_{obs} = 11.456 > \chi^2_{3,0.95} = 7.815$, logo rejeitamos a hipótese nula para $\alpha = 5\%$. Os dados evidenciam que os dados não estão de acordo com o estudo.

   (b) *p-value*$\approx$ 0.01

5. ver output do SPSS da aula Teórica.

6. (a) a=130, b=90, c=80, d=100, e=400

   (b) $H_0 : p_1 = p_2 = p_3 = p_4 = 0.25$
   $H_1 : \exists p_i \neq 0.25$, $i = 1, 2, 3, 4$

p-value= 0.003, logo rejeitamos a H_0, ou seja, podemos dizer que existe uma forte evidência estatística de preferência por algum dos tipos de filmes.

7. $H_0 : p_1 = 0.4, p_2 = p_3 = p_4 = 0.2$
   $H_1 : p_1 \neq 0.4$ ou $\exists p_i \neq 0.2$, $i = 2, 3, 4$
   p-value= 0.037, logo rejeitamos a H_0 para $\alpha = 5\%$ (5%≥p-value), ou seja, podemos dizer que para este nível de significância os dados contradizem a afirmação.

8. H_0 : X e Y são independentes *vs* H_1 : X e Y não são independentes
   Tabela 2×2, correcção de Yates, $\chi^2_{obs} = 15.186 > \chi^2_{1,0.95} = 3.841$, logo rejeitamos a hipótese nula para $\alpha = 5\%$. Os dados evidenciam existe relação entre os resultados obtidos nos dois testes.

9. (a) a=60, b=300, c=200, d=50
   (b) H_0 : X e Y são independentes *vs* H_1 : X e Y não são independentes
   p-value= 0.008 logo rejeitamos H_0 : para qualquer um dos níveis de significância usuais (1%, 5% ou 10%). Existe evidência estatística de que a preferência clubistica não é independente da idade.

10. ver output do SPSS da aula Teórica.

11. (a) Teste de independência do qui-quadrado. H_0 : X e Y são independentes *vs* H_1 : X e Y não são independentes
    A=30, B=50, C=200, D=50, E=180, F=80, G=400
    (b) Como $\chi^2_{obs} = 152.222 > \chi^2_{2,0.90} = 4,605$, rejeitamos a hipótese nula para $\alpha = 10\%$. Os dados evidenciam que existe relação entre os região de residência e a opinião acerca do desenvolvimento.

2.7 *Regressão Linear Simples*

1. Pretende-se modelar a resistência de um determinado tipo de plástico em função do tempo que decorre....

   (a) $\hat{Y}_i = 153.904 + 2.417X_i$ (no teste o valor de $\tilde{\beta}_0$ estava em falta e foi aqui calculado utilizando a fórmula). $\tilde{\beta}_0$ diz-nos que a resistência do plástico imediatamente após a conclusão da moldagem é, em média, de 153.904.
   $\tilde{\beta}_1$ significa que por cada aumento de 1 hora no tempo decorrido desde a conclusão da moldagem e a medição da resistência, verifica-se, em média, um aumento de 2.417 na resistência do plástico.

   (b) $R = 0.9796$: Existe uma associação linear positiva muito forte entre a resistência do plástico e o tempo decorrido desde a conclusão da moldagem e a medição da resistência. $R^2 = 0.9596$: cerca 96% da variabilidade da resistência do plástico é explicada pela relação linear que possui com o tempo decorrido desde a conclusão da moldagem e a medição da resistência

   (c) $H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 \neq 0.$ o quadro do SPSS apresenta o intervalo de confiança a 95% para β_0, uma vez que o 0 não está contido neste intervalo, também não estará contido no intervalo de confiança a 90% para β_0, pelo que ao nível de significância de 10% pode concluir que a recta de regressão não passa pela origem

   (d) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$, quadro do SPSS apresenta, para este este, o *p-value*$\approx$0, pelo que se rejeita a hipótese nula para qualquer um dos α usuais. Ou seja, é possível afirmar estatisticamente que o tempo desde a conclusão do processo de moldagem até à medição da resistência influencia linearmente de forma significativa a resistência do plástico.

   (e) 238.499.

2. Uma empresa que avalia a qualidade de água das ETAR, analisa o ...

   (a) $\hat{Y}_i = 1.2813 + 0.5997X_i$. $\tilde{\beta}_0$ diz-nos que para a ausência de substâncias químicas utilizadas no tratamento da água, o nível de substâncias nocivas é 1.2813 mg/ 100ml. $\tilde{\beta}_1$ significa que por cada aumento de 1 mg/100 ml no nível de substâncias químicas utilizadas no tratamento da água verifica-se, em média, um aumento de aproximadamente 0.6 mg/ 100 ml no nível de substâncias nocivas detectadas na água.

   (b) $R = \sqrt{0.53} = 0.728$: Existe uma associação linear positiva moderada entre o nível de substâncias químicas utilizadas no tratamento da água e o nível de substâncias nocivas detectadas água. $R^2 = 0.53$: 53% da variabilidade do nível de substâncias nocivas detectadas água é explicada pela relação linear que possui com o nível de substâncias químicas utilizadas no tratamento da água.

(c) $H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 \neq 0$. o quadro do SPSS apresenta o *p-value*≈*0* para este teste, logo rejeita-se a hipótese nula para $\alpha = 5\%$ ($\alpha = 5\% > p\text{-}value$). Evidência estatística de que a recta de regressão não passa na origem.

(d) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$. $t = 2.499 > z_{0.975} = 1.96$, pelo que se rejeita a hipótese nula para $\alpha = 5\%$. Ou seja, é possível afirmar estatisticamente que o nível de substâncias químicas utilizadas no tratamento da água influencia linearmente o nível de substâncias nocivas detectadas na água.

(e) 0.345

(f) 0.5997

3. O Presidente da Junta de Freguesia da pacata vila alentejana (e dono da única gasolineira) ...

(a) $\hat{Y}_i = 34.58 + 1.47X_i$. $\tilde{\beta}_0$ não faz sentido interpretar. $\tilde{\beta}_1$ significa que por cada aumento de 1 euro no preço da gasolina verifica-se, em média, um aumento de aproximadamente 1.47 l de gasolina vendida.

(b) $R^2 = 0.96^2 = 0.9216$: 92.16%

(c) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$. $t = 4.15 > z_{0.975} = 1.96$, pelo que se rejeita a hipótese nula para $\alpha = 5\%$. Ou seja, é possível afirmar estatisticamente o preço da gasolina e a quantidade vendida estão relacionados linearmente de modo significativo.

(d) 37.5

4. Pretende-se obter um modelo que possa explicar a área foliar (em cm^2) através do comprimento da nervura principal (em cm) em folhas de videiras da casta Fernão Pires.

(a) $\hat{Y}_i = -106.443 + 24.263X_i$. $\tilde{\beta}_0$ não tem interpretação, pois não faz sentido considerar que um comprimento da nervura principal igual a zero. $\tilde{\beta}_1$ significa que por cada aumento de 1 cm no comprimento da nervura principal a área foliar aumenta, em média, 24.263 cm^2.

(b) $R^2 = 0.94$. 94% da variabilidade da área foliar é explicada pelo comprimento da nervura principal.

(c) 48.526

(d) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$. $t = 12.550$, pelo que se rejeita a hipótese nula para $\alpha = 5\%$. Ou seja, é possível afirmar estatisticamente que o comprimento da nervura principal influencia linearmente a área foliar.

(e) $]-177.9927; -34.8933[$

(f) $\hat{Y}_{9.1} = 114.3519$, donde resulta um resíduo igual a 11.6497.

5. Para alguns países da Europa, foram registados alguns indicadores económicos, nomeadamente o PIBA...

   (a) $Y_i = 15,4623 - 0,1467X_i$.

   (b) $H_0 : \beta_1 = -0,2 \quad vs \quad H_1 : \beta_1 \neq -0,2$. Como $t_{obs} = 0,8417$ não se rejeita H_0 para $\alpha = 5\%$. Ou seja, não existe forte evidência estatística para afirmar (para $\alpha = 5\%$) que o declive da recta de regressão é diferente de -0,2.

   (c) Aproximadamente 29%.

6. No gráfico de dispersão abaixo encontram-se registados os tempos (em segundos) do vencedor na final da corrida de 400 metros masculinos...

   (a) A variável dependente corresponde aos tempos (em segundos) do vencedor na final da corrida dos 400 metros masculinos e a variável explicativa corresponde ao ano em que cada Olimpíada se realizou.

   (b) A relação entre as duas variáveis é negativa, pois à medida que os anos aumentam o tempo do vencedor na final da corrida dos 400 metros masculinos diminui.

   (c) $r^2 = 0,8277$. Aproximadamente 83% da variabilidade verificada nos tempos do vencedor na final da corrida dos 400 metros masculinos é explicada pelo tempo (ano da Olimpíada).

   (d) $\hat{Y}_i = 185,1177 - 0,07111X_i$. A interpretação de β_0 não faz sentido neste exercício, por correspondia ao tempo do vencedor para um ano zero. A interpretação de β_1 corresponde ao decréscimo médio, em segundos, para a Olimpíada seguinte.

   (e) $\hat{Y}_{2008} = 42,328$ segundos.

7. Considerando a idade (em anos) e a capacidade pulmonar (em L) de 9 crianças...

   (a) $r = 0,986$ e $r^2 = 0,972$.

   (b) $\hat{Y}_i = 0,093 + 0,163X_i$.

   (c) $]0,1394; 0,1866[$.

   (d) $t_{obs} = 1,147$. Não se rejeita a hipótese nula ao nível de significância de 1%. Ou seja, não existe forte evidência estatística para afirmar, com 99% de confiança, que a recta de regressão não passa na origem.

8. Pretende-se, se possível, modelar através de uma recta de regressão linear simples a quantidade de vidro...

(a) $\hat{Y}_i = 10,629 + 56,131X_i$.

(b) $r^2 = 0,986$.

(c) $t_{obs} = 21,972$. Rejeita-se a hipótese nula ao nível de significância de 10%.

(d) $\hat{Y}_{35} = 1975,214$ Kg.

9. Mediu-se o número de pulsações por minuto antes e depois de uma determinada prova de esforço...

(a) $r = 0,934$ e $r^2 = 0,872$

(b) $Y_i = 27,55 + 1,323X_i$

(c) $]0,910;1,736[$.

(d) $H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 \neq 0$. Como $t_{obs} = 2,676 < 3,355$ não se rejeita H_0 para $\alpha = 1\%$. Ou seja, existe forte evidência estatística para afirmar (para $\alpha = 1\%$) que a recta de regressão passa na origem.

(e) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$.
Como (pela alínea c) $0 \notin]0,910;1,736[$ rejeita-se H_0. Ou seja, existe forte evidência estatística para afirmar (para $\alpha = 5\%$) que a variável X explica linearmente o comportamento da variável Y.